



Ein sinnvoller Einstieg in die digitale Prüfungstechnik auf der Basis von Methodenverständnis und zielführender -vernetzung.

Andreas Wähnert

[Finanzamt für Zentrale Prüfungsdienste (ZPD) Kiel]

4. überarbeitete öffentliche Auflage (29.2.2024)

13. überarbeitete interne Auflage (Erstauflage v. 29.11.2007)

 andreas.waehnert@fa-zpd.landsh.de

Bei dem Handbuch für die Summarische Risikoprüfung (SRP) handelt es sich nicht um einen Erlass des Finanzministeriums Schleswig-Holstein sondern um eine Hilfestellung für Prüferinnen und Prüfer für den Einsatz moderner Prüfungsmethoden. Nicht alle Ausführungen des Handbuchs stellen daher zwingend die Position des Finanzministeriums Schleswig-Holstein dar.

Inhalt:	Seite
1. SRP-Kurzbeschreibung	6
2. Intention der SRP (Kernaspekte)	13
2.1. Heutige Vollzugssituation und Steuerausfallrisiken	17
2.2. Moderne Prüfungsmethodik in der digitalisierten Welt	22
2.3. Vorteile mehrperspektivischer Prüfungsnetze wie SRP	29
3. Einbindung der SRP in Prüfung und Risikomanagement	34
3.1. Die SRP zum Prüfungseinstieg und zur Prüffeldaufklärung	38
3.2. Die SRP im besonderen Kontext des § 158 Abs. 2 Nr. 1 AO	41
3.3. Voraussetzungen für quantitative Prüfungsmethoden	46
4. Die Verwendung der SRP-Vorlagen (Basis-SRP)	50
4.1. Formalprüfung: Ordnungsmäßige (prüfbare) Aufzeichnungen	55
4.2. Regelbasierte Ansätze (Zeitreihenanalysen)	66
4.2.1. Betriebsprofil	72
4.2.2. Prüfungssystematik der SRP-Vorlage BWA	78
4.2.3. Jahressummen	83
4.2.4. Quartalsrelationen	85
4.2.5. Monatszeitreihen	89
4.2.5.1. Doppelskaliert größenangepasster Zeitreihenvergleich	90
4.2.5.2. Einzelskalierter Zeitreihenvergleich	93

4.2.5.3. Verhältniszeitreihe (z.B. zum Aufschlagsatz)	95
4.2.6. Gleitschlittenzeitreihen	96
4.2.7. Jeweilige Entwicklungen der einzelnen Datenreihen	100
4.2.8. Rückrechnung der monatlichen Warenbestände	103
4.2.9. Interaktive Funktionen zur verbesserten Sachaufklärung	106
4.2.9.1. Einzelkorrektur von Datensätzen	108
4.2.9.2. Einschränkung des Betrachtungszeitraums	109
4.2.9.3. Verschieben der ersten Datenreihe	109
4.2.9.4. Beseitigung von (einseitigen konjunkturellen) Preiseinflüssen	110
4.2.9.5. Niveauanpassung eines Grafen	113
4.2.10. Kosten- und weitere Zeitreihen	113
4.2.11. Beweiskraft von (regelbasierten) Zeitreihenanalysen	118
4.3. Musterbasierte Analysen (Häufigkeitsauswertungen)	124
4.3.1. Ablauf von Häufigkeitsanalysen	133
4.3.1.1. Beobachtungen für Häufigkeitsanalysen	133
4.3.1.2. (Punkt-)Erwartungen für Häufigkeitsanalysen	134
4.3.1.3. Übereinstimmungsprüfung (Chi-Quadrat-Test)	136
4.3.1.4. Schlussfolgerungen zu Häufigkeitsanalysen	141
4.3.2. Ziffernanalyse	144
4.3.2.1. Anwendungshinweise zu SRP-Ziffernanalysen	146
4.3.2.2. Einerstellenuntersuchung	149
4.3.2.2.1. Die (Punkt-)Erwartung „Gleichverteilung“	149
4.3.2.2.2. Nicht-manipulative Ursachen für Einerstellenauffälligkeiten	151

4.3.2.2.3. Vorlagenunterstützte Abklärung von Einerstellenauffälligkeiten	151
4.3.2.3. Zweitziffernanalyse (Benford-Gesetz)	154
4.3.2.3.1. Die (Punkt-)Erwartung „Benford-Gesetz“	155
4.3.2.3.2. Nicht-manipulative Ursachen für Zweitziffernauffälligkeiten	157
4.3.2.3.3. Vorlagenunterstützte Abklärung von Zweitziffernauffälligkeiten	157
4.3.3. Klassifizierung (Schichtung) der Werte	161
4.3.3.1. Anwendungshinweise zu SRP-Wertklassifizierungen	163
4.3.3.2. Abklärung von Zweitziffernauffälligkeiten mit 99er-Schichtung	168
4.3.3.3. Vorbereitende Wertklassifizierungen für die Verteilungsanalyse	171
4.3.4. Verteilungsanalyse	174
4.3.4.1. Die (Punkt-)Erwartung „Logarithmische Normalverteilung“	176
4.3.4.2. Anwendungshinweise zur SRP-Verteilungsanalyse	180
4.3.4.3. Ergebnisverdeutlichung bei auffälligen Verteilungsanalysen	185
4.3.4.4. Andere Ursachen für auffällige Verteilungsanalysen	189
4.3.4.5. Abklärung von auffälligen Verteilungsanalysen	190
4.3.4.6. Quantifizierung von auffälligen Verteilungsanalysen	192
4.3.5. Beweiswert von (musterbasierten) Häufigkeitsanalysen	195
4.4. Gesamtergebnis einer plausibilitätsorientierten SRP	202
4.5. SRP im Verhältnis zu anderen Prüfungsmethoden	207
4.6. Welches Beweismaß fordert § 158 Abs. 2 Nr. 1 AO?	211
4.7. Einleitung des Strafverfahrens nach § 10 BpO	217
4.8. Schätzung im Besteuerungsverfahren (Quantilsschätzung)	221
4.8.1. Durchführung einer Quantilsschätzung	225

4.8.2. Anwendungshinweise zur Vorlagen-Quantilsschätzung	230
4.8.3. Rechtsprechung zur Quantilsschätzung	236
4.9. Strafrechtliche Schätzung	242
5. Freie SRP-Anwendung (systematische Datenvisualisierung)	247
5.1. Einsatzbereiche systematischer Visualisierung	252
5.1.1. Systematischer Überblick mit visualisierender Prüffeldsuche	256
5.1.2. Visualisierte Übereinstimmung (Schnittstellenprüfung)	265
5.1.3. Visualisierte Schlüssigkeit („sprechende Diagramme“)	269
5.2. Ausblick zur Prüfungsmethodik und -technik	280
6. (Ausgewählte) Literatur zur „Neuen Prüfungsmethodik/-technik“	288

1. SRP-Kurzbeschreibung

Die **Summarische Risikoprüfung (SRP)** ist

- eine vorlagenunterstützte (technisch entlastende und leitende) **Filterprüfung** für elektronische Betriebs-/Besteuerungsdaten (Buchführung, Vorsysteme etc.),
- um mit Hilfe mehrperspektivisch-vernetzter Gesamtbildverfahren (**Prüfungsnetz**)
- allgemein einzelfallbezogene **Risiko-Prüffelder** systematisch (effektiv und effizient) festzustellen und einzugrenzen
- sowie speziell bei Prüffeldern mit weniger eindeutigen Entscheidungssituationen wie insbesondere „Einnahmenvollständigkeit“ oder „Fremdüblichkeit von Verrechnungspreisen“ eine möglichst umfassende und abgesicherte **Schlüssigkeitsbeurteilung** – i.S. einer möglichst hohen Aussagegewissheit – vornehmen zu können;
- dadurch soll ein **ressourcenschonendes Risikomanagement** mit zeitgemäßen (digitalen) Prüfungsmethoden unterstützt werden.

Das Konzept der Summarischen Risikoprüfung (SRP) ist ein **Methodik-Konzept**¹, welches das Verständnis für die umfangreichen, interdisziplinären Grundlagen der zwei wichtigen Methodenbereiche „regelbasierte (Zeit-)Reihenanalysen“ und „musterbasierte Häufigkeitsauswertungen“ in den Mittelpunkt stellt, damit sie in zielführenden Prüfungsprozessen sinnvoll angewendet werden können – inklusive Validierung der Eignung von Methoden(-varianten) für bestimmte Prüfungsfragen und gegenseitiger Absicherung sich ergänzender Ansätze. Zur technischen Vereinfachung der Methodenanwendung leiten die SRP-Vorlagen Anwender*innen entlang logischer Prüfungspfade, nutzen die Funktionsvielfalt von Tabellenkalkulationsprogrammen, geben Bedien- und Warnhinweise, ermöglichen die interaktive Einarbeitung neuer Erkenntnisse z.B. aus konkreter Mitwirkung und erleichtern eine nachvollziehbare Dokumentation.

In Anbetracht der großen Anzahl an Unternehmen in Deutschland reichen die vorhandenen Personalkapazitäten nur für **unbefriedigende Prüfungsquoten**² im Bereich der Kleinst-

¹ Zwar wird ab der 3./12. Ausgabe des SRP-Handbuchs die Ausrichtung „Methodik-Konzept“ erstmalig explizit benannt, die entsprechenden Inhalte sind aber seit Begründung Bestandteil des Projekts.

² Monatsbericht des BMF für den Oktober 2023 mit der Betriebsprüfungsstatistik 2022, die Prüfungsquoten von 0,8 % für Kleinst-, 2,4 % für Klein- und 4,8 % für Mittelbetriebe ausweist; nachzulesen unter: https://www.bundesfinanzministerium.de/Monatsberichte/2023/10/Inhalte/Kapitel-3-Analysen/3-2-steuerliche-betriebspruefung-2022-pdf.pdf?__blob=publicationFile&v=4.

bis Mittelbetriebe, sodass infolge der (statistischen) Kurzlebigkeit viele von diesen nie intensiv geprüft werden. Deshalb könnte die Außenprüfung z.B. bei einem Kleinstbetrieb nach Literaturmeinung³ an sich schon als Akt staatlicher Willkür empfunden werden. Die Veranlagungsprüfung kann eine vergleichbare Tiefe nicht erzielen, weil der digitale Zugriff auf die Einzeldaten nach § 147 Abs. 6 AO auf die Außenprüfung beschränkt ist bzw. ein „automationsgestützte[s ...] Risikomanagementsystem“ (§ 88 Abs. 5 AO) an konsolidierten, informationsarmen Daten ansetzt und dabei voraussichtlich nur bekannte Risiken erfasst.

Zusätzlich wird die in der Konsequenz bereits begrenzte Aufdeckungswahrscheinlichkeit von Steuerdelikten durch die Verbreitung immer **komfortablerer digitaler Manipulationswerkzeuge** wie z.B. Zapper für Warenwirtschaftsprogramme von Apotheken⁴ weiter reduziert. Dementsprechend wird beim BFH „auf Seiten der Finanzverwaltung ein durchaus nachvollziehbares Bedürfnis, moderne Prüfungsmethoden [...] einzusetzen“⁵ gesehen. Der Bundesrechnungshof erkennt „durch immer komplexere Manipulationssoftware [...] bei der Besteuerung von Bargeschäften ein strukturelles Vollzugsdefizit“⁶. Die Kleine Anfrage „Schaden durch Betrug mit manipulierten Kassensystemen und mögliche Einführung der INSIKA-Lösung zur Betrugssicherung“⁷ zeigt, dass auch die Politik alarmiert ist. Entsprechend erreichte das Thema in den Populärmedien mit Titeln wie „Steuerhinterziehung: Zu Gast bei Betrügern“⁸ zwischenzeitlich eine starke Präsenz und wurde an prominenter Stelle vor einem gesellschaftsbedrohenden Verfall der Steuermoral gewarnt⁹.

Unabhängig von vorsätzlichen Eingriffen in Besteuerungsdaten offenbaren gezielte Prüfungsmaßnahmen z.T. gravierende **Datenmängel der elektronischen Betriebsverwaltungssysteme**¹⁰, die heutzutage zunehmend automatisiert ablaufen – tlw. mit unternehmensspezifischen Anpassungen und intransparent unvollständigen Zugriffsmöglichkeiten. Demnach scheinen die betriebsinternen Kontrollsysteme (IKS) die Risiken von Fehl-

³ Tipke, „Die Steuerrechtsordnung“, Verlag Dr. Otto Schmidt, Bd. III, 1407.

⁴ Nachzulesen unter: www.apotheke-adhoc.de/nachrichten/nachricht-detail/lauer-fischer-zapper-ist-nicht-von-uns.

⁵ Urt. v. 25.3.2015, BStBl. II 2015, 743, Rz. 54.

⁶ n-tv online v. 25.6.2015, nachzulesen unter: <http://www.n-tv.de/wirtschaft/Laender-kaempfen-gegen-Schummel-Kassen-article15373191.html>.

⁷ Nachzulesen unter: <http://dip21.bundestag.de/dip21/btd/18/061/1806118.pdf>.

⁸ Zeit online v. 20.2.2014, nachzulesen unter: <http://www.zeit.de/2014/09/steuerhinterziehung-gastronomie>.

⁹ Köhler bei der Verleihung des Max-Weber-Preises für Wirtschaftsethik 2008: „Wenn eine Mehrheit es für clever hält, das Finanzamt zu betrügen, dann gehen nicht nur Steuereinnahmen verloren, dann gerät auch der Rechtsstaat und das Vertrauen in die demokratische Ordnung in Gefahr“, nachzulesen unter: www.stern.de/politik/deutschland/manager-schelte-koehlers-neues-lieblingsthema-621740.html sowie „Steueroase Deutschland“, Süddeutsche.de v. 7.11.2013 und Spiegel 40/2013, 74.

¹⁰ Webel/Danielmeyer, „Schnittstellenverprobung elektronischer Betriebsverwaltungen – das kommende Standardprüffeld?“, StBp 2015, 353.

schlüsselungen, Datenschnittstellen etc. (derzeit noch) nicht ausreichend abzusichern. Entsprechend präsent werden (technische) Lösungsvorschläge wie „Blockchain-Technologie“ beworben und neue Standards¹¹ – inzwischen mit gesetzlicher Begleitung¹² – definiert.

Besonders im Bereich der **international agierenden Konzerne** kommen Anzeichen weiterer Risiken für dauerhafte Steuerausfälle hinzu. So moniert das Deutsche Institut für Wirtschaftsforschung (DIW) illegale Steuervermeidungsstrategien internationaler Unternehmensverbünde¹³, wogegen die groß angelegte OECD-Aktion „Base Erosion and Profit Shifting (BEPS)“¹⁴ sowie u.a. das „Gesetz zur Einführung einer Pflicht zur Mitteilung grenzüberschreitender Steuergestaltungen“¹⁵ ansetzen sollen. Ohne wirksame Überprüfungsmöglichkeiten der tatsächlichen Verhältnisse können der internationale Informationsaustausch und Regulierungen der Dokumentationsweise allerdings keine ausreichende Sicherung der Steueransprüche gewährleisten, wodurch sich die datenbasiert-transaktionale Ausrichtung der neuen „Verwaltungsgrundsätze Verrechnungspreise“¹⁶ erklärt.

Außerdem sind unter **digital agierenden Steuerpflichtigen** neue Geschäftsmodelle wie bspw. Werbeeinnahmen für Social-Media-Inhalte entstanden,¹⁷ bei denen die klassischen Kontroll- und Prüfmechanismen der Finanzverwaltungen nicht ausreichend greifen.

Zusammenfassend haben die von der Wirtschaft vorangetriebene Digitalisierung und Globalisierung unserer Gesellschaft bekannte Steuerrisiken verändert bzw. verschärft und neue geschaffen. Deshalb gibt es in vielen Finanzverwaltungen Bestrebungen, ein **wirksames Risikomanagement** aufzubauen.¹⁸ Schwerpunkte der Maßnahmen liegen in den

¹¹ IDW-Praxishinweis zur „Prüfung von Compliance Management Systemen“ (IDW PS 980), nachzulesen unter: <https://www.idw.de/idw/idw-verlautbarungen/idw-ps-980.html>. Liekenbrock/Danielmeyer, „Data Governance als Reaktion auf die zunehmend datengetriebene Betriebsprüfung?“, RET 2/2021, 32.

¹² Art. 97 § 38 EGAO mit der Systemprüfung von Steuerkontrollsystemen und der Möglichkeit, Beschränkungen der nächsten Außenprüfung zuzusagen.

¹³ n-tv online v. 28.5.2013, nachzulesen unter: www.n-tv.de/wirtschaft/Fiskus-entgehen-jedes-Jahr-Milliarden-article10717336.html.

¹⁴ Nachzulesen unter: <http://www.oecd.org/ctp/beps.htm>.

¹⁵ Zur Umsetzung der EU-Richtlinie 2018/822 des Rats bezüglich des verpflichtenden automatischen Informationsaustauschs im Bereich der Besteuerung über meldepflichtige grenzüberschreitende Gestaltungen vom 25. Mai 2018 (ABl. L 139/1 vom 5. Juni 2018) – sog. DAC 6 – s. z.B. unter: https://www.bundesfinanzministerium.de/Content/DE/Downloads/BMF_Schreiben/Weitere_Steuerthemen/Abgabenordnung/2021-03-29-Anwendung-Vorschriften-Pflicht-Mitteilung-grenzueberschreitende-Steuergestaltungen.html.

¹⁶ BMF-Schreiben v. 6.6.2023, „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093.

¹⁷ Vgl. z.B.: Danielmeyer, „Datenanalyse aus dem Blickwinkel der Außenprüfung“, beck.digitax 4/2023, 218.

¹⁸ S. z.B.: Risk Management Guide, Leitfaden – Risikomanagement für Steuerverwaltungen, Europäische Kommission, Generaldirektion Steuern und Zollunion, Indirekte Steuern und Steuerverwaltung, Verwaltungszusammenarbeit und Bekämpfung des Steuerbetrugs, Fiscalis-Projektgruppe Risikoanalyse, Version 1.02, Februar 2006.

Bereichen „Organisation“ (z.B. Neuaufnahmestellen zur frühzeitigen Risikoerkennung und Kassen-Nachschaufen¹⁹), „Informationsaustausch und Datenbankennutzung“²⁰, „automatisierte Filternetzwerke für die Steueranmeldungs- und -erklärungsdaten“ (§ 88 Abs. 5 AO) sowie „Beweislastumkehr bei Besteuerungsdaten, die gesetzlich vorgegebenen digitalen Schnittstellen nicht entsprechen“ (§ 158 Abs. 2 Nr. 2 AO). Zum einen sind Quartals- und Jahreswerte für eine wirksame Risikoerkennung jedoch zu sehr verdichtet (und die Risikospuren dadurch „geglättet“), weshalb z.B. Ungarn für das System RADAR differenzierte Rückmeldungen aus durchgeführten Außenprüfungen verkennziffert und in die Risikoskalierung einbezieht. Zum anderen verkennt der Einsatz „Künstlicher Intelligenz (KI)“²¹, dass die Konzentration auf selbstlernende Algorithmen die Prüfungskompetenzen des Personals „Intuition“ und „Erfahrung“ ausblendet und damit wichtige Ressourcen ignoriert.²²

Wegen des Vollzugriffs auf elektronische Besteuerungsunterlagen nach § 147 Abs. 6 AO dürften in der **Optimierung der Außenprüfung** die erfolgversprechendsten Chancen liegen, insbesondere auch die erhöhten (digitalen) Steuerausfallrisiken einzudämmen. So sieht die Wissenschaft einen Schlüssel für die erfolgreiche Durchführung von Prüfungsmaßnahmen in der Prüfungstechnik mit einem „hohen Maß an Methodenwissen, Flexibilität und Kreativität“²³. Dazu fordert der Bundesrechnungshof „den bundesweit einheitlichen Einsatz IT-gestützter quantitativer Prüfungsmethoden in der steuerlichen Betriebsprüfung [... denn t]raditionelle Methoden der steuerlichen Betriebsprüfung, beispielsweise Beleg- und Akteneinsicht, verlieren an Effektivität und Effizienz. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen“²⁴, was u.a. in dem BMF-Schreiben „[a]utomationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“²⁵ mündete. Der sinnvoll kombinierte Einsatz der Analysesoftware IDEA® mit Tabellenkalkulationsprogrammen wie MS Excel®, B(usiness-)I(n-

¹⁹ Anwendungserlass zu § 146b AO.

²⁰ Vgl. z.B.: EU-Projekt „VAT in the Digital Age (ViDA)“, nachzulesen unter: https://taxation-customs.ec.europa.eu/taxation-1/value-added-tax-vat/vat-digital-age_en.

²¹ Schwerpunktthema u.a. beim 14. Deggendorfer Forum zur digitalen Datenanalyse mit dem Titel „Die Welt ist komplex: Verstehen Computer sie besser?“, nachzulesen unter: <https://www.dfdda.de/index.php/de/veranstaltungen/forum/14-forum-2018>.

²² Folgerichtig wird in der Wirtschaftsprüfung ein sinnvoller Einsatzbereich der KI erkannt, indem „Künstliche Intelligenz (KI) und Robotics Routinetätigkeiten übernehmen [und] Wirtschaftsprüfer [...] neben ihrem Fachwissen Kenntnisse im IT-Bereich benötigen, um diese Prozesse zu verstehen“, früher nachzulesen unter: <https://www.der-betrieb.de/meldungen/die-zukunft-der-wirtschaftspruefung-ist-digital/>.

²³ Früher nachzulesen unter: www.wpunco.wiwi.uni-due.de/forschung/pruefungstechnik.

²⁴ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

²⁵ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594.

telligence)-Programmen und künftig voraussichtlich auch generativer KI²⁶ kann ein entsprechendes Maß an Funktionalität erreichen, um von einem Gesamtüberblick über Besteuerungsdaten zur gezielten Identifikation prüfungsrelevanter Sachverhalte vorzugehen. Hingegen sind die methodischen Grundlagen wirksamen Prüfens – insbesondere Statistik und Visualisierung – nicht trivial, worauf die Bildungswissenschaft hinweist.²⁷

Auf Basis dieser methodischen Grundlagen – also der notwendigen Lese- und Anwendungskompetenz²⁸ – nutzt die **SRP** mit ihrem zentralen Aspekt der „systematischen Visualisierung“ den bekannten wahrnehmungsphysiologischen Vorteil²⁹ der Verbildlichung von Zahlen und komplexen Sachverhalten in bedienfreundlich vorkonfektionierten Vorlagen, um den Einstieg in die digitale Prüfungstechnik zu erleichtern und dennoch die Entscheidungen bei den Prüfer*innen zu belassen; während die technischen Werkzeuge einem schnellen Wandel unterliegen, bildet das methodische Verständnis dabei das konstantere Gerüst. Durch die systematische Verknüpfung mehrerer geeigneter Analyseansätze werden die erreichbare Datentransparenz wesentlich verbessert und die Grundlagen für das sinnvolle Arbeiten mit B(usiness-)I(ntelligence)-Programmen gelegt. Ein bundesweites Austauschforum für Prüfungsideen³⁰ ermöglicht zudem die zeitnahe Verbreitung von Weiterentwicklungen und fördert damit die digitale Prüfungsweise insgesamt.

Zur einfacheren Erlernbarkeit der Datenanalyse-Methodik konzentriert sich die SRP auf zwei wesentliche Methodenbereiche, aus denen sich auch komplexere quantitative Verfahren wie z.B. multiple Regressionsanalysen weitgehend ableiten lassen: Erstens identifiziert eine mehrperspektivisch-chronologische Auswertung von (abhängigen) Betriebsgrößen (z.B. Warenfluss, umsatzbezogene Kosten, Datenschnittstellen oder Vertragsumsetzungen) in Zeitreihenanalysen Prüffelder als „Regelstörungen“ und kann zur Beurteilung der kaufmännischen Schlüssigkeit herangezogen werden. Zweitens suchen systematische Häufigkeitsuntersuchungen zu bestimmten Merkmalen oder Werteklassen nach „Musterstörungen“ in Form von auffälligen Häufungen oder Ausdünnungen als prüfungsrelevante Vorgänge (Gestaltungen und Fehler) und speziell in manipulationsgefährdeten Betriebsdaten (besonders

²⁶ Um das prüfungsmethodische und -technische Wissen angemessen nutzen zu können, wird auch die Finanzverwaltung nicht auf die Nutzung von generativer KI wie z.B. ChatGPT® verzichten können – unter Beachtung aller rechtlichen Einschränkungen.

²⁷ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 64 u. 89.

²⁸ Vgl. zur heutigen Bedeutung der Datenkompetenz z.B.: Stifterverband „Data Literacy Education“, nachzulesen unter: <https://www.stifterverband.org/data-literacy-education>.

²⁹ Barnard in einer US-amerikanischen Zeitschrift der Werbebranche 1921: „One look is worth a thousand words“, früher nachzulesen unter: www.oai.de/de/component/content/article/45-publikationen/spruchwort/811-ein-bild-sagt-mehr-als-tausend-woorte.html.

³⁰ Intranet: <http://wiki.ktz.testa-de.net/cgi-bin/twiki/view/PRUEFUNGSD/QuantitativePruefungsmethoden>.

Tageskasseneinnahmen) nach eindeutigen [stochastisch (mit Hilfe der Wahrscheinlichkeitstheorie) abgesicherten] Eingriffsspuren.

Beide Bereiche sind mit Hilfe der SRP-Vorlagen BWA und Stochastik schnell und variabel handhabbar. Sie setzen nur geringe technische Fähigkeiten und durchschnittliche mathematische Kenntnisse voraus, indem z.B. Anwendungsregeln in der Programmierung bzw. dem Prüfungspfad berücksichtigt werden. In der SRP-Vorlage BWA wird durch die multiple Betrachtung die Gefahr von Interpretationsfehlern (z.B. als Folge nicht beachteter betrieblicher Besonderheiten) reduziert. Vergleichbar konzentriert sich die SRP-Vorlage Stochastik im speziellen Teil für die Suche nach Manipulationsspuren zur Vermeidung von Fehldeutungen auf die praxisrelevantesten Hinterziehungsarten und unterstützt das Abklären betrieblicher Ursachen. Beide Vorlagen bieten Möglichkeiten für nachvollziehbare Korrekturen (Interaktivität), um eine zeitgemäße Prüfungstiefe und eine höchstmögliche Absicherung der Ergebnisse zu gewährleisten. Zudem erleichtern sie eine nachvollziehbare Dokumentation der Auswertungen. Zwei integrierte Schätzungsansätze („Quantilsschätzung“ und „Verteilungsanalyse“) sollen – *bei vorliegender Schätzungsnotwendigkeit* – zusätzliche Praxisvorteile für die Ermittlung einer einzelfallgerechten Schätzungshöhe bieten.

Die bisherigen **Erfahrungen zur SRP** aus den Prüfungen durch mehrere tausend geschulte Prüfer*innen in fast allen Bundesländern sowie von der Bundesprüfung, beim Zoll, in der niederländischen Finanzverwaltung und über Kooperationen mit dem Internationalen Währungsfond (IWF) haben gezeigt, dass

- ✓ nach der einwöchigen Basisfortbildung (und zeitnaher Anwendung an eigenen Fällen) üblicherweise ein ausreichendes Methodenverständnis bei den Anwender*innen für die vorlagenunterstützte Umsetzung erreicht wurde,
- ✓ der erforderliche Zeitaufwand für eine vorlagenunterstützte SRP als Orientierungshilfe zu Beginn einer Außenprüfung wenige Stunden beträgt,
- ✓ selbst bei nicht vollständiger Durchführbarkeit aller Einzelmethoden (z.B. infolge mäßiger Aufzeichnungsqualität) eine qualifizierte Einschätzung der Vertiefungsnotwendigkeit und Festlegung von Prüffeldern möglich ist (Risikomanagement),
- ✓ beim bargeldintensiven Einzelhandel in einer Vielzahl der Fälle Implausibilitäten angezeigt werden, was die zeitgemäße Aufdeckungswahrscheinlichkeit erkennen lässt,
- ✓ bei (offensichtlich) manipulierten Unterlagen i.d.R. diverse Auffälligkeiten auftreten, womit die Beweiswirkung der Indiziengesamtheiten gem. BGH-Rechtsprechung erheblich verstärkt wird,
- ✓ konventionelle Ergebniskontrollen z.B. mit Hilfe von (zeitaufwendigeren) Kalkulationen und Durchsuchungsmaßnahmen die Ergebnistendenz regelmäßig bestätigen und

- ✓ die (freie) Anwendung systematisch-mehrperspektivischer Visualisierung bei der Suche nach und Abarbeitung von Prüffeldern schneller zu vollständigeren Resultaten und damit auch zu besseren Entscheidungsgrundlagen für die Rechtsanwendung führt.

Von der Prüfungsbedeutung der Verbildlichung wird inzwischen medial berichtet.³¹ Sowohl die SRP als auch wichtige Teilaspekte wie das Betriebsprofil oder die systematische Vernetzung sind im **BMF-Schreiben zu „[a]utomationsgestützten quantitativen Prüfungsmethoden“**³² enthalten und von der **Kommentierung/Literatur** aufgenommen worden.³³ Ihre konzeptionelle Sorgfalt und die Aussagekraft entsprechend erarbeiteter Ergebnisse werden z.T. sehr positiv bewertet.³⁴

Trotz der systematischen Vorlagenunterstützung der SRP mit Maßnahmen zum Fehlerausschluss und zur Ergebnisabsicherung ist ein eingehendes Studium der angewandten Methoden (Begriffe, Funktionen, Eigenschaften etc.) inklusive kombinierter Beurteilung notwendig und zentrales Anliegen der SRP, damit die Anwendung bis hin zur Schätzung verantwortungsvoll und sachgerecht erfolgt. Dieses **Handbuch**³⁵ hat zwar den Charakter einer Bedienungsanleitung für die SRP-Vorlagen, wobei aber alle vorgestellten Auswertungen mit frei zugänglichen Programmen und deren originären Funktionen (ohne intransparente Programmierungen) nachvollzogen werden können. In den ersten beiden Kapiteln werden vorab die Intention der SRP aus Sicht des Autors und eine sinnvolle Einbindung in die einzelne Prüfung, das bestehende Risikomanagement sowie die Personalentwicklungskonzepte (Aus- und Fortbildungsaspekte) beschrieben.

³¹ Wirtschaftswoche v. 21.5.2015: „Fiskus 4.0: Mit Big Data auf Steuerfang“, früher nachzulesen unter: <http://www.wiwo.de/politik/deutschland/fiskus-4-0-mit-big-data-auf-steuerfang/11770850.html>.

³² BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594.

³³ Vgl. z.B.: Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 16 u. 20, Zugmaier/Nöcker, Abgabenordnung, 1. Aufl. 2022, § 162 AO, Rz. 155 ff., bzw. juris Lexikon Steuerrecht, 2. Aufl. 2021, Stand: 12.9.2023, Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstrafverfahren“, NZWiSt 1/2014, 24, sowie Brinkmann, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Aufl. 2023, u.a. Rz. 599.

³⁴ Vgl. z.B.: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 218: „Ein relevantes und wertvolles Werkzeug der Finanzbehörden ist dabei die Summarische Risikoprüfung (SRP-Handbuch 2020) geworden, die [...] den methodischen Standard der praktischen Nutzung darstellt.“ Pump/Heidl, „Hilfestellung bei der Kassenführung durch den steuerlichen Berater – Ärger wegen § 162 AO durch unzulängliche Kassenbuchführung“, StBp 2014, 163: „Der weitere Vorteil von SRP besteht darin, dass die finanzamtliche Beweisführung für die Hinzuschätzung auch im Steuerstrafverfahren gerichtsfest ist.“ SRP-kritische Stimmen wie bspw. Steinhauß [„Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388] erwähnen zwar tlw. die methodologischen Differenzierungen und Abwägungen der SRP, ziehen sich in ihrer Beurteilung aber wieder auf die nicht differenzierende Rechtsprechung zurück.

³⁵ Die z.T. häufigere Wiederholung von bestimmten Aussagen und insbesondere Fußnoten soll auch die auszugswise Nutzung als Handbuch gewährleisten.

2. Intention der SRP (Kernaspekte)

Leitmotive bei der (Weiter-)Entwicklung der SRP als Methodik-Konzept³⁶ für ein systematisch-mehrperspektivisches Prüfungsnetz sind bzw. waren es,

- die **methodischen Grundlagen** zu den beiden wichtigen **quantitativen**³⁷ Bereichen „regelbasierte (Zeit-)Reihenanalysen“ und „musterbasierte Häufigkeitsauswertungen“ so ausführlich darzustellen, dass ein zielführender Einsatz mit aussagekräftigen Ergebnissen gewährleistet werden kann,
- um die digitalen Vorteile „Rechengeschwindigkeit“, „Variabilität“ und „Interaktivität“ – in Verbindung mit den Stärken des Prüfungspersonals – dafür zu nutzen, zu Beginn einer Prüfungsmaßnahme eine **qualifiziertere Risikoeinschätzung** (mit zeitgemäßer Aufdeckungswahrscheinlichkeit betriebsspezifischer Prüffelder) sowie eine möglichst hochgradige Ergebnisklarheit von Plausibilitätsverprobungen zu erreichen, und
- die wahrnehmungsphysiologische Überlegenheit der Verbildlichung beim Erfassen, Bewerten und Kommunizieren komplexer Informationen für eine **schnellere, vollständigere und besser argumentierbare Abarbeitung einzelner Prüffelder** wie z.B. bei der Aufdeckung gesetzeswidriger Gestaltungen (s. [Pkt 5.1.3.](#)) einzusetzen.

Dazu setzt die SRP zentral auf die **Visualisierung von Besteuerungsdaten**, denn „ein Bild sagt mehr als tausend Worte [bzw. Zahlen]“³⁸. Das Sprichwort gilt zwar als Erkenntnis der Werbebranche, ist aber wahrnehmungsphysiologisch unumstritten³⁹ und lässt sich am Eigenver-

Warenflussdaten		
	Umsatz	Wareneinsatz
Jan.	97.791,65	25.159,28
Feb.	112.294,21	30.485,25
Mrz.	170.895,79	50.337,68
Apr.	226.075,89	77.657,37
Mai	216.656,23	64.092,84
Jun.	257.442,80	77.851,63
Jul.	234.162,84	71.635,76
Aug.	231.841,05	72.596,78
Sep.	181.955,20	55.737,06
Okt.	151.803,76	46.769,50
Nov.	122.264,11	38.390,08
Dez.	137.304,88	38.822,04

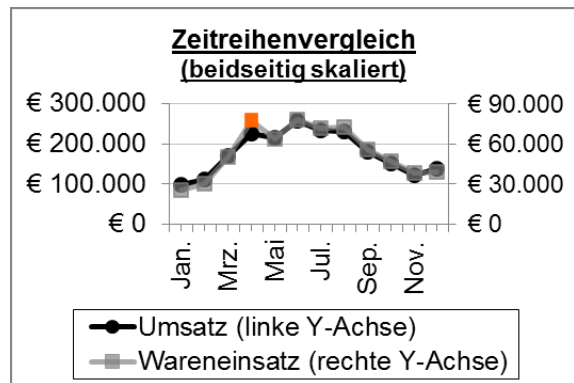
³⁶ Zwar wird ab der 3./12. Ausgabe des SRP-Handbuchs die Ausrichtung „Methodik-Konzept“ erstmalig explizit benannt, die entsprechenden Inhalte sind aber seit Begründung Bestandteil des Projekts.

³⁷ BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594.

³⁸ Barnard in einer US-amerikanischen Zeitschrift der Werbebranche 1921: „One look is worth a thousand words“, früher nachzulesen unter: www.oai.de/de/component/content/article/45-publikationen/spruchwort/811-ein-bild-sagt-mehr-als-tausend-woerter.html.

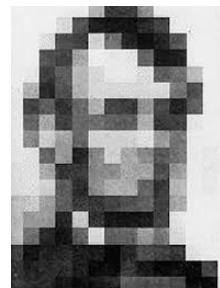
³⁹ Vgl. z.B.: Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. Die Analyse statistischer Daten beginnt häufig mit der Wahl einer geeigneten Darstellungsform des Erhebungsbefundes.“

sich einfach nachvollziehen: Während sich aus der Tabelle zu den Warenflussdaten eines gastronomischen Betriebs oberhalb kaum Informationen über mögliche Prüffelder oder die Gesamtplausibilität (direkt) ableiten lassen, offenbart eine geeignete Verbildlichung wie in dieser Abb. sowohl die strenge Abhängigkeit beider Betriebsfaktoren als auch ein hervorstechendes Einzelprüffeld mit dem relativ zu hohen **roten Wareneinsatzpunkt** im April.



In der Umsetzung der mehrperspektivisch (s. [Pkt. 2.3](#)) visualisierenden SRP wurden zudem **wichtige Erkenntnisse zur Datenanalyse** berücksichtigt wie insbesondere:

- ☒ Nach den **Fuzzy-Theorien** zur Bedeutung der (Un-)Schärfe ist die Aggregation von Daten entscheidend dafür, welche Informationen erkennbar sind. So lässt sich trotz grober Darstellung im nebenstehenden Bild⁴⁰ das Antlitz von Abraham Lincoln erkennen. Dieses Wissen ist für die Effizienz der Datenanalyse relevant, denn je verdichteter die zu prüfenden Daten sind, umso schneller und einfacher gestaltet sich die Untersuchung. Auf der anderen Seite wird dieser Effizienzgedanke von der abnehmenden Effektivität – also der Wahrscheinlichkeit des Erkennens – begrenzt. Zuletzt bewirkt eine zu große Informationsfülle i.d.R. auch eine Überforderung des Menschen, sodass zwar Details aber gröbere Regeln und Fehler nicht mehr gesehen werden.
- ☒ Die in der professionellen Datenanalyse⁴¹ geschätzten **Koinzidenzanalysen** untersuchen den Einfluss aller möglichen Faktorkombinationen auf Risiken bzw. Chancen und geben die Ergebnisse als komfortabel bedienbare „Rankings“ aus. Mit der (nahezu) vollständigen Verarbeitung der verfügbaren Informationen erfolgt die Datenanalyse ergebnisoffen und kann somit auch unerwartete Fehler und Manipulationen detektieren. Da die Datenvisualisierung allerdings maximal drei bis vier Faktorarten (Dimensionen) **gleichzeitig**⁴² erlaubt, werden multifaktorielle Ansätze häufig vollständig automatisiert, sodass das Prüfungspersonal seine Qualitäten „Intuition“ und „Erfahrung“ erst zu den



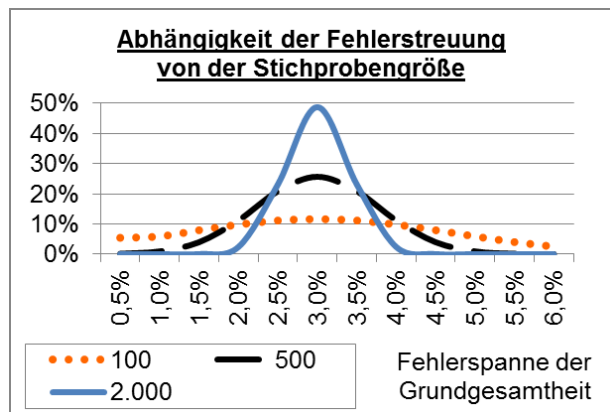
⁴⁰ Bildquelle: Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 54.

⁴¹ Vgl. z.B.: „HyperCube® [der Fa. BearingPoint] wird vom Pasteur-Institut als das derzeit leistungsstärkste Datenanalyse-Tool auf dem Markt angesehen. [...] ein neuartiger Datenbestimmungsalgorithmus, der die Kombinationen von Variablen identifizieren kann, die eine optimale Erklärung der zugrundeliegenden Fragestellung ermöglichen“, **früher** nachzulesen unter: <http://www.bearingpoint.com/de-de/expertise/enabling-technology/hypercube>.

⁴² Im Gegensatz zu einzelnen Komplexdiagrammen ermöglichen simulationsfähige Auswertungsmodelle mit B(usiness-)I(ntelligence)-Programmen auch umfangreichere Datenvisualisierungen, setzen dazu aber eine entsprechende systematische Anwendungskompetenz voraus.

Ergebnissen einbringen kann; das wiederum führt in Verbindung mit der meist komfortablen Bedienbarkeit i.d.R. dazu, dass weder der Informationsgehalt der Ausgangsdaten noch die vordefinierten Untersuchungsparameter kritisch hinterfragt werden.

- ☒ Als weltweit verbreitetes Lösungsmodell für komplexe Fragestellungen und Probleme orientiert sich die (biokybernetische) **Sensitivitätsanalyse** von Vester⁴³ an Mustern in der Datenstruktur, weil es „[s]imple Ursache-Wirkungs-Beziehungen [...] nur in der Theorie [gibt], nicht in der Wirklichkeit.“⁴⁴ Diese Idee eröffnet in der Übertragung auf die Prüfung eine Ausweitung der unter Wirtschaftsprüfer*innen üblichen ABC-Analyse mittels Ergebnisverbildlichung, wodurch wesentlich detaillierter Fehler und Manipulationen aufgedeckt werden können; gerade zu diesem leistungsfähigen Bereich der Datenanalytik befürchtet die Wissenschaft jedoch, „dass Abschlussprüfer häufig Defizite in der Fähigkeit von Mustererkennungen aufweisen“⁴⁵. Deshalb wird gerne mit Einfärbungen („Heatmap“ und „Traffic-light“) nachgeholfen, was als technische „Notlösung“ verstanden werden muss, solange den Anwender*innen die methodischen Grundlagen für Musteranalytik fehlen.
- ☒ Im Rahmen **statistischer Stichprobenverfahren** wie z.B. „Monetary Unit Sampling (MUS)“⁴⁶ wird die Unsicherheit der resultierenden Aussage mit Hilfe des mathematischen Spezialgebiets „Wahrscheinlichkeitslehre“ qualifiziert, weshalb sie gesetzlich zugelassen⁴⁷ und in der Wirtschaftsprüfung⁴⁸ verbreitet sind. So gibt die Abb.⁴⁹ die Übertragung eines 3 %-igen Stichprobenfehlers auf die Grundgesamtheit wieder: Während bei einer **großen Zufallsstichprobe von 2.000 Vorgängen** der tatsächliche Gesamtfehler nur geringfügig um die 3 % streut, kann bei einem **kleinen untersuchten Auszug von**



⁴³ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012.

⁴⁴ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 15.

⁴⁵ Velte/Drews, „Herausforderungen in der Wirtschaftsprüfung durch den Einsatz von Big-Data-Technologien – Welche Impulse muss die Prüfungsforschung geben?“, DB 43/2018, 2581, 2583.

⁴⁶ Giezek, „Monetary Unit Sampling“, Gabler Verlag, 2011.

⁴⁷ § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden.“

⁴⁸ IDW PS 310 „Repräsentative Auswahlverfahren (Stichproben) in der Abschlussprüfung“, nachzulesen unter: <https://www.idw.de/idw/verlautbarungen/idw-ps-310/43204>.

⁴⁹ Aus dem Grund der besseren Übersichtlichkeit wird die Fehlerstreuung mit der entsprechend angepassten stetigen Normalverteilung abgebildet, die approximativ den diskreten Verhältnissen der einschlägigen Hypergeometrischen Verteilung entspricht.

z.B. 100 Einzelsachverhalten nicht mit derselben Klarheit auf die Gesamtheit geschlossen werden. Die konkreten, reproduzier- und damit vergleichbaren Erkenntnisse der Statistik eignen sich besonders für die Einordnung von unsicheren Aussagen, wie sie u.a. immer vorkommen, wenn die Vergangenheit anhand unvollständiger Informationen nachvollzogen werden soll; deshalb weist die wissenschaftliche Literatur auch auf die große Bedeutung für die Rechtsprechung hin: „Die juristische Methodenlehre ist somit die Lehre von der richtigen und gerechten Entscheidung unter Unsicherheitslagen.“⁵⁰

Für die Ausgestaltung des SRP-Prüfungsduktus und der -Vorlagen, welche die technischen Umsetzungsanforderungen möglichst niedrig halten sollen, sind dieses Wissen und die umfangreichen Erfahrungen aus der täglichen Anwendung fundamental: In der Unterrichtung der SRP steht das umfassende Verständnis der Einzelmethoden im Vordergrund, weil bspw. die Varianten von Zeitreihenanalysen eine stark differierende Eignung für unterschiedliche Prüfungsfragen aufweisen (s. [Pkt. 2.2.](#)). Somit liefert die SRP auch wesentliche Voraussetzungen für eine sinnvolle Nutzung der immer stärker verbreiteten B(usiness-)I(ntelligence)-Programme oder analysefähiger „Künstlicher Intelligenz (KI)“.

Im Ergebnis soll die SRP zur Gleich- und Gesetzmäßigkeit der Besteuerung durch (zeitgemäß) **effektive Außenprüfungen** beitragen, denn „die Finanzbehörden haben alle notwendigen Maßnahmen zu ergreifen, um die entscheidungserheblichen Tatsachen aufzuklären“⁵¹. Art und Umfang sollen den „Grundsätzen der [...] Verhältnismäßigkeit“⁵² folgen und werden durch die Mitwirkungspflichten der Steuerpflichtigen beschränkt⁵³, bilden aber als *primäre* Aufgabe der Betriebsprüfungsstellen gemäß § 2 BpO die entscheidende Grundlage für die *anschließende* rechtliche Beurteilung – vergleichbar der sorgfältigen Diagnostik in der Medizin als Voraussetzung für eine erfolgreiche Therapeutik.

Entsprechend der Bedeutung der Sachverhaltsaufklärung definiert das Verfahrensrecht mit dem Verifikationsprinzip eine besondere Sphärenverantwortung im **Beweislastverhältnis des § 158 Abs. 2 Nr. 1 AO**.⁵⁴ Dieses gründet sich auf die Einsicht, „dass es

⁵⁰ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 145.

⁵¹ Anwendungserlass zu § 88 AO Nr. 1.

⁵² § 88 Abs. 2 AO.

⁵³ Anwendungserlass zu § 88 AO Nr. 5.

⁵⁴ BMF-Schreiben v. 3.12.2020 „Verwaltungsgrundsätze 2020“, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. [...] Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u.a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwor-

weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild einer Buchführung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann“⁵⁵. Danach führt eine „Verprobung [...] nach der es] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt“⁵⁶, zum Verlust der Vermutung sachlicher Richtigkeit – bis zur möglichen Konsequenz der ermessensfreien Schätzungspflicht⁵⁷. Insofern hat eine (zeitgemäße) wirksame Prüfungs^{methodik}⁵⁸ und -technik zur möglichst weitreichenden Sachverhaltsaufklärung eine grundlegende und direkte Bedeutung für den Schutz der Gleich- und Gesetzmäßigkeit der Besteuerung durch die Außenprüfung.

2.1. Heutige Vollzugssituation und Steuerausfallrisiken

Als Folge der Digitalisierung und Globalisierung der Wirtschaft haben sich die „**Herausforderungen der digitalen Betriebsprüfung**“⁵⁹ wesentlich erhöht: Rasant anwachsende Mengen an Betriebsdaten, spezielle Fehlerquellen der (weitgehend) automatisierten Datenverarbeitung und die nur schwierig verifizierbare Originalität (Unverändertheit) von elektronischen Aufzeichnungen⁶⁰ sowie neuartige Geschäftsmodelle ohne klassische Leistungswege und Zahlungsarten⁶¹ sind Hauptfaktoren für immer größere Probleme beim (vollständigen) Auffinden von Fehlern und Manipulationen in Besteuerungsdaten.

„Die heutigen Datenmengen und qualitativen Ansprüche übersteigen zunehmend die Möglichkeit einer manuellen Prüfung. Relevante Aussagen über ein Prüfungsobjekt können heute nicht mehr allein mit der Durchsicht von Listen, simpler Belegprüfung oder durch Stichproben getroffen werden.“⁶² Vom Prüfungspersonal wird ein „hohes Maß an Methoden-

tungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden.“

⁵⁵ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5.

⁵⁶ Anwendungserlass zu § 158 AO.

⁵⁷ § 162 Abs. 2 S. 2 AO.

⁵⁸ BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 4: „Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgabenordnung dazu führen, dass eine Schätzung durchgeführt werden muss.“

⁵⁹ Kulosa (RiBFH), „Herausforderungen der digitalen Betriebsprüfung, u.a. bei bargeldintensiven Betrieben“, FR 2017, 501.

⁶⁰ Huber/Reckendorf/Zisky, „Die Unveränderbarkeit der Kassenbuchführung § 146 Abs. 4 AO im EDV-Zeitalter und INSIKA – Teil 1: Historische Wurzeln und Realzustand“, BBK 12/2013, 567, „Teil 2: Probleme, Ursachen und Lösungen“, BBK 13/2013, 610 sowie „Teil 3: Konzept, Technik, Praxis“ BBK 14/2013, 663.

⁶¹ Mit einem Beispielsablauf eines Geschäftsvorfalles: Danielmeyer, „Der „digitale“ Steuerpflichtige wird virtuell – Die Evolution der Außenprüfung“, StBp 2022, 376, 377.

⁶² Ehlers, „Tax Compliance – Neue Beratungsfelder bei Betreuung kleiner und mittlerer Unternehmen“, NWB 18/2012, 1535, 1540.

wissen, Flexibilität und Kreativität“⁶³ gefordert: „Traditionelle Methoden der steuerlichen Betriebsprüfung, beispielsweise Beleg- und Akteneinsicht, verlieren an Effektivität und Effizienz. Nur IT-gestützte **quantitative** Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.“⁶⁴ So erkennt z.B. auch der BFH: „Elektronische Kassensysteme sind durch Umprogrammierung in nahezu beliebiger Weise manipulierbar; von derartigen Manipulationsmöglichkeiten machen Teile der betrieblichen Praxis nach dem Erkenntnisstand des Senats durchaus Gebrauch“⁶⁵. In der Konsequenz sieht der BFH „auf Seiten der Finanzverwaltung ein durchaus nachvollziehbares Bedürfnis [nach] moderne[n] Prüfungsmethoden“⁶⁶. Die (methodischen und technischen) Anforderungen an das Prüfungspersonal zur Erfüllung der vorrangigen Aufgabe „Sachverhaltsaufklärung“⁶⁷ haben sich in nahezu jedem Prüfungsbereich erhöht, wie z.B. auch mit der stärkeren Ausrichtung der Verrechnungspreisprüfung auf datenbasiert-transaktionale Arbeitsweise.⁶⁸ In einem neuen BMF-Schreiben werden die „[a]utomationsgestützte[n] quantitative[n] Prüfungsmethoden der steuerlichen Außenprüfung“, die – über das BpO-konforme⁶⁹ Aufdecken betriebsspezifischer Prüffelder hinaus – eine direkte Auswirkung auf Stpfl. haben können, als erste Informationsquelle aufgelistet.⁷⁰

Auf der anderen Seite eröffnet die digitale Außenprüfung durch die enorme Geschwindigkeit der Datenverarbeitung und die zielführende, flexible Vernetzbarkeit geeigneter Prüfungsmethoden (s. [Pkt. 2.3.](#)) verbesserte Möglichkeiten zur weitreichenden Sachaufklärung. Dadurch wird die Bearbeitung von diversen (alltäglichen) Prüffeldern schneller und vollständiger, z.T. sogar erstmalig über eine Stichprobenauswahl hinaus durchführbar.

Die **exponentielle Zunahme des Datenvolumens** gemäß der nachfolgenden Abb.⁷¹ geht darauf zurück, dass „[d]ie Buchführung [...] sich von der Durchschreibebuchführung hin zu einer umfassenden Datenverarbeitung entwickelt [hat], die elektronische Form ist nun-

⁶³ Internetseite des Lehrstuhls für Wirtschaftsprüfung, Unternehmensrechnung und Controlling von Prof. Mochty an der Universität Duisburg-Essen, früher nachzulesen unter: <https://www.wpunco.wiwi.uni-due.de/forschung/pruefungstechnik>.

⁶⁴ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

⁶⁵ Urt. v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

⁶⁶ Urt. v. 25.3.2015, BStBl. II 2015, 743, Rz. 54.

⁶⁷ § 2 BpO.

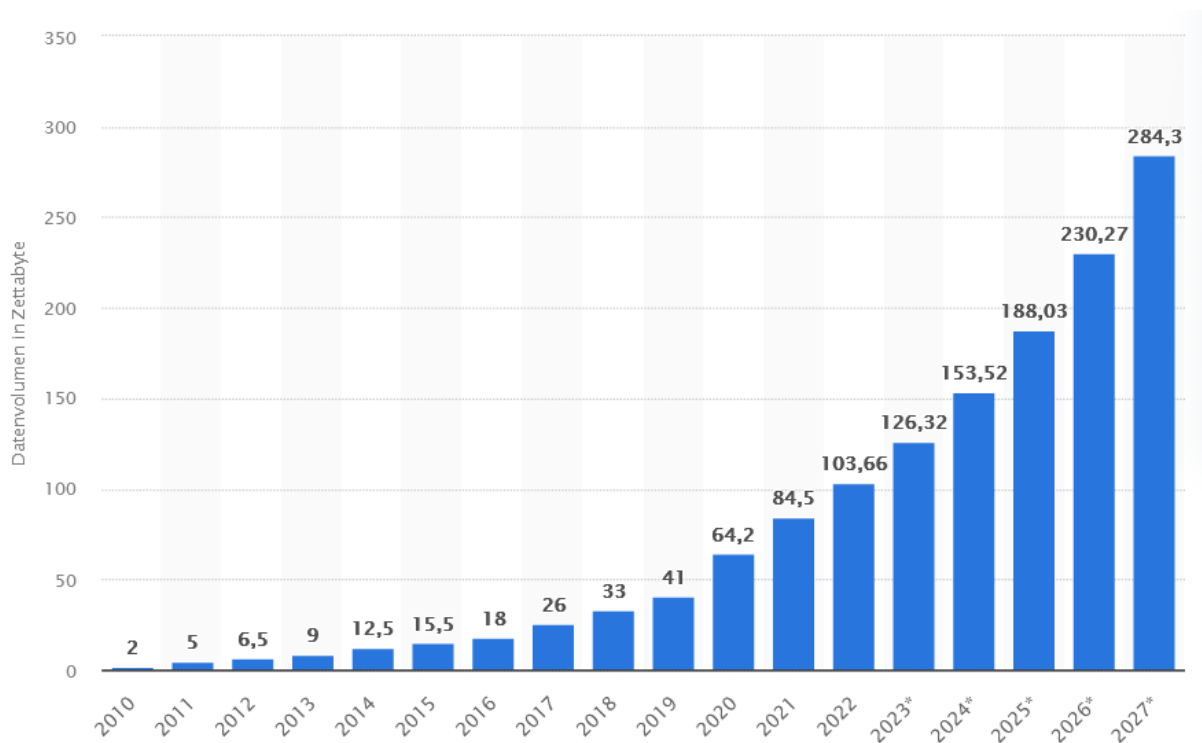
⁶⁸ BMF-Schreiben v. 6.6.2023 „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093.

⁶⁹ § 7 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen. Ihre Dauer ist auf das notwendige Maß zu beschränken.“

⁷⁰ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594.

⁷¹ Quelle der Grafik: „Statista 2023“, nachzulesen unter: <https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen>.

mehr die übliche. [...] Diese Entwicklung geschah auf Betreiben und im Interesse der Unternehmen, weil es betriebswirtschaftlich notwendig war und ist, diese Einzeldaten zu erzeugen, z.B. für ein aussagekräftiges Controlling oder zur Arbeitsprozesssteuerung. [...] Dabei steigt die Anzahl elektronisch überwachter Prozesse stetig.“⁷²



Weil zudem bekannt ist, dass automatisch generierte Betriebsdaten durchaus nennenswerte Fehlerquellen beinhalten (können),⁷³ wächst die Bedeutung des zielgenauen und zügigen Auffindens von Fehlern und Manipulationen ständig.

Systemprüfungen in den Groß- und Konzernbetriebsprüfungsstellen haben über die Datenmigration hinaus auch in Schlüsselungen, Zusammenführungen und Konsolidierungen von Betriebsdaten mitunter gravierende Abweichungen bzw. Fehler offenbart, weshalb bspw. die Untersuchung von Datenschnittstellen als wesentliches neues Prüffeld angesehen wird.⁷⁴ Als Reaktion auf dieses Problem werden in Wirtschaft und Beratung technische Lösungen wie die Vorgangsverkettung der „Blockchain-Technologie“⁷⁵ erprobt, die auch Bestandteil der

⁷² Graf, „Anforderungen an die Vorlage von Einzelaufzeichnungen – Ende der ‚Schonzeit‘ für Daten aus Vor-Systemen?“, BBK 18/2013, 875, 876 u. 877.

⁷³ Vgl. z.B.: Newsletter 4-2016 des Forums „Elektronische Steuerprüfung“ v. 22.4.2016: „(IT-)Management: Ein Drittel aller Unternehmen verlieren Daten bei der Migration: Ein Drittel (32 %) der Organisationen verlieren Daten, während sie Daten zwischen Geräten migrieren oder Betriebssysteme aktualisieren. Das ist das Ergebnis einer weltweiten Befragung unter fast 600 IT-Administratoren“, nachzulesen unter: http://elektronische-steuerpruefung.de/newsarchiv/news_16_04.htm?m=1450705466#Beitrag7.

⁷⁴ Webel/Danielmeyer, „Schnittstellenverprobung elektronischer Betriebsverwaltungen – das kommende Standardprüffeld?“, StBp 2015, 353.

⁷⁵ Vgl. z.B. „Betriebswirtschaftliche Prozesse in die Blockchain? – Machbarkeitsstudie von Telekom und DATEV zeigt Potenzial der Technologie auf“, nachzulesen unter:

österreichischen Registrierkassensicherheitsverordnung zur Manipulationsbekämpfung von elektronischen Aufzeichnungssystemen ist⁷⁶, oder es wird auf weitreichende Prozessautomatisierung⁷⁷ gesetzt. Parallel dazu sollen durch die Definition neuer Standards wie „Data Governance“⁷⁸ inklusive „[d]igitale[r] Operationalisierung [von] Tax Compliance Management-Systemen[en]“⁷⁹ Fehler in der betrieblichen Aufzeichnung und Aufbewahrung künftig verhindert werden; politisch wird dieser Ansatz dadurch unterstützt, dass „die Finanzbehörde im Benehmen mit dem Bundeszentralamt für Steuern dem Steuerpflichtigen [mit einem wirksamen Steuerkontrollsystem] auf Antrag unter dem Vorbehalt des Widerrufs für die nächste Außenprüfung nach § 193 Absatz 1 der Abgabenordnung Beschränkungen von Art und Umfang der Ermittlungen unter der Voraussetzung verbindlich zusagen [kann]“⁸⁰. Unabhängig von der Art und Größe eines zu prüfenden Unternehmens und in fast allen Prüfungsbereichen – einschließlich der Wirksamkeitsprüfung von Steuerkontrollsystemen – gewinnt somit die zeitgemäße Arbeitsweise der Außenprüfung eine zunehmende Bedeutung für die Gleich- und Gesetzmäßigkeit der Besteuerung. Die Literatur fordert schon länger eine den heutigen Risiken und Möglichkeiten angemessene Verifikation.⁸¹

Bereits seit 2003 warnt der Bundesrechnungshof vor „[d]rohende[n] Steuerausfälle[n] aufgrund moderner Kassensysteme“⁸². In einem Bericht hat die OECD die Prüfungssituation im Zeitalter **raffiniertes (digitaler) Manipulationswerkzeuge** mit beispielhaften „Techniken der Umsatzverkürzung, Aufdeckungsstrategien und [s]taatliche[n] Maßnahmen“⁸³ zusammengefasst.

Die systemimmanenten oder als Zusatzmodul käuflichen Eingriffsfunktionen nutzen in (komplexen) Softwarelösungen für Kasse(n), Warenwirtschaft, Rechnungslegung, Kundenverwaltung etc. die Intransparenz der Systeme und die leichte Veränderbarkeit digital gespeicherter Informationen aus; dabei wird z.T. offensiv mit den Möglichkeiten der Steuerhin-

<https://www.datev.de/web/de/m/presse/pressemitteilungen/meldungen-2018/betriebswirtschaftliche-prozesse-in-die-blockchain/>.

⁷⁶ Huber, „Reform der Kassenführung in Österreich – Teil 2 – Technische Umsetzung, Status und Erkenntnisse“, BBK 23/2016, 1148, 1150.

⁷⁷ „Robotic Process Automation (RPA)“.

⁷⁸ Liekenbrock/Danielmeyer, „Data Governance als Reaktion auf die zunehmend datengetriebene Betriebsprüfung?“, RET 2/2021, 32.

⁷⁹ Walz/Oehrle/Him, „Digitale Operationalisierung des Tax Compliance Management-Systems“, beck.digitax 4/2023, 199.

⁸⁰ Art. 97 § 38 EGAO „Erprobung alternativer Prüfungsmethoden“.

⁸¹ Fischer (Vors. RiBFH a.D.), „Das EuGH-Urteil Persche zu Auslandsspenden – die Entstaatlichung des Steuerstaates geht weiter“, FR 2009, 249.

⁸² Bundesrechnungshof, „Bemerkungen 2003 zur Haushalts- und Wirtschaftsführung des Bundes“, Pkt. 54, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/archiv>.

⁸³ OECD-Bericht „Umsatzverkürzung mittels elektronischer Kassensysteme – eine Bedrohung für die Steuereinnahmen“, 2013, nachzulesen unter: https://www.oecd.org/ctp/crime/Sales_suppression_German_website.pdf.

Diese Software unterstützt z.B. die fortlaufenden Rechnungsnummern. Um bei Bearbeitungsfehlern jedoch Korrekturen vornehmen zu können, ist das das Ändern von Rechnungsnummern jederzeit möglich! ?

Genauso ist es möglich jederzeit Daten unprotokolliert (also restlos) zu löschen ?

Bevor Sie diese Software benutzen, muss sich der Anwender beim Finanzamt, beim Steuerberater oder bei anderen zuverlässigen Quellen über diese Regelungen und Vorschriften informieren. !

terziehung geworben.⁸⁴ Ein Überblick über die weitreichenden Manipulationsoptionen im heutzutage üblichen e-Management der Betriebe findet sich bei *Teutemacher* mit Beispielen aus der Fahndungspraxis wie im abgebildeten Auszug

Mobile Kasse	Hauptkasse
Nummer : 1. Name : Kasse Vorhanden : 1 Rech.back.tab. send.: 1 Konsolidiert buchen : 0 IP Adr Byte 1 : 168 IP Adr Byte 2 : 0 IP Adr Byte 3 : 42 IP Adr Byte 4 : 00:80:C5:EA:6B:E9 MAC Adresse : 00:80:C5:EA:6B:E9 Nunmer : 3.	Kassen im Netz Vorgabe markiert : 1 Off. Tische senden : 0 Rech.back.tab. send.: 1 Konsolidiert buchen : 1 IP Adr Byte 1 : 168 IP Adr Byte 2 : 0 IP Adr Byte 3 : 42 IP Adr Byte 4 : 00:80:C5:EA:6B:E9 MAC Adresse : 00:80:C5:EA:6B:E9

zum System eingestellt: „Es wird so der Anschein erweckt, als ob alle Umsätze im Kassensystem erfasst werden. Dies ist jedoch ein Trugschluss, wie die Programmierung des Mobilgeräts [...] zeigt, denn dort wurde bei ‚konsolidiert buchen‘ eine ‚0‘ programmiert. Dies führt dazu, dass Umsätze, die über dieses Kassenterminal erfasst wurden, nicht im Tagesumsatz dokumentiert werden.“⁸⁵

Der mögliche Umfang von Steuerhinterziehungen mit Hilfe digitaler Aufzeichnungssysteme lässt sich aus einem Beschluss des FG Rheinland-Pfalz erahnen, nach welchem die hinterzogenen Steuern eines einzelnen Eiscafés für den erweiterten Prüfungszeitraum knapp € 2 Mio. betrugen.⁸⁶

Besonders im Bereich der digitalen Erfassung von Bareinnahmen wirkt sich der Wettlauf von Verbrechen und Aufdeckung insofern aus, dass das Prüfungspersonal zur effektiven Durchführung der Arbeit über technische Kenntnisse⁸⁷ sowie ein „hohes Maß an Methodenwissen, Flexibilität und Kreativität“⁸⁸ verfügen muss. Dabei kommt erschwerend der Eindruck hinzu, dass einige Vertreter*innen der steuerberatenden Berufe ihrer Vorprüfungsfunktion als

⁸⁴ Huber, „Gedanken zur Zukunft des steuerlichen Risikomanagements im Erlösbereich – Teil II“, StBp 2014, 153, 155 mit der Abb. und 156 mit einem Zitat aus der dortigen Abb. 10 „Werbung für Handwerkersoftware“: „Insbesondere von Vorteil ist, dass man jederzeit alle Einträge rückgängig machen oder ändern kann. Da speziell im Handwerk einiges anders läuft (z.B. Schwarzgeld) ist dies auch für Sie wichtig, denn Umbuchungen oder Löschungen sind nachträglich nicht mehr sichtbar.“

⁸⁵ Teutemacher, „Handbuch zur Kassenführung – Praxishandbuch für die rechtssichere Umsetzung“, NWB-Verlag, 2015, 152 mit der Abb. und dem Zitat.

⁸⁶ Beschl. v. 7.1.2015 – 5 V 2068/14; zitiert von: BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

⁸⁷ Vgl. z.B. die Mechanismen einer TSE-Prüfung durch TAR-File-Export sowie einer Kassenprüfung mittels DSFinV-K-Export: Till, „Kassenprüfung 2.0 – Auftakt zu mehr Steuergerechtigkeit“, RET 5/2021, 18.

⁸⁸ Internetseite des Lehrstuhls für Wirtschaftsprüfung, Unternehmensrechnung und Controlling von Prof. Mochty an der Universität Duisburg-Essen, früher nachzulesen unter: <https://www.wpunco.wiwi.uni-due.de/forschung/pruefungstechnik>.

Organ der Rechtspflege⁸⁹ – z.B. im Hinblick auf elektronische Vorsysteme – nicht (mehr) ausreichend nachkommen.⁹⁰ Ob die Umsetzung des „Gesetz[es] zum Schutz vor Manipulationen an digitalen Grundaufzeichnungen“⁹¹ dauerhaft Sicherheit in dieses Prüffeld bringen kann, bleibt abzuwarten.

Mit der „Besteuerung von **Social-Media-Akteuren**“ werden weitere neue Herausforderungen für die steuerliche (Außen-)Prüfung als Folge der Digitalisierung und Vernetzung erkannt.⁹² Sowohl durch die neuen „Meldepflichten digitaler Plattformbetreiber“⁹³ als auch mittels „Verordnung über Märkte für Kryptowerte“⁹⁴ soll diesen begegnet werden. Für die Betriebsprüfer*innen vergrößert das zusätzlich die Bedeutung der methodisch-technischen Datenanalysekompetenzen.⁹⁵

2.2. Moderne Prüfungsmethodik in der digitalisierten Welt

Schon die **Formalprüfung** zur Feststellung, ob Besteuerungsunterlagen infolge ihrer Ordnungsmäßigkeit eine inhaltliche Prüfung (in angemessener Zeit) zulassen, stellt in Verbindung mit elektronischer Aufzeichnung und Aufbewahrung höhere Anforderungen an das Prüfungspersonal. Während „[d]er sachverständige Dritte [...] über seine beruflich erlernten Buchhaltungskenntnisse hinaus in der Papier-/Tinte-/Bleistift-/Radierer-Zeit keine außergewöhnlich hohe technische Qualifikation [brauchte], um die Erfüllung der gesetzlichen Vorgaben feststellen zu können“⁹⁶, müssen heutzutage komplexe Betriebsverwaltungsprozesse verstanden werden: „Eine Systemprüfung hat das gesamte Verfahren, welches die Verarbeitung steuerlich relevanter Daten betrifft [...] zum Gegenstand. So soll festgestellt werden, ob und inwieweit eine zuverlässige, sichere, nachvollziehbare, ordnungsgemäße und vollständige Verarbeitung der steuerlich relevanten Daten durch den Steuerpflichtigen erfolgt.“⁹⁷ Neue,

⁸⁹ § 57 StBerG.

⁹⁰ Pump/Heidl, „Hilfestellung bei der Kassenführung durch den steuerlichen Berater – Ärger wegen § 162 AO durch unzulängliche Kassenbuchführung – Teil II“, StBp 2014, 162 u. 204, 163, 206: „Unverantwortliche Sorglosigkeit kann die Berater dazu verleiten, sich nicht um die Kassenführung ihrer Mandanten zu kümmern. Das wird damit begründet, dass dafür kein Prüfungsauftrag bestehe. [...] Die Berufspflichten für Berater sind streng. [...] Das Prinzip ‚Augen zu und ab mit der falschen Steuererklärung‘ ist töricht. Das würde die vom Berater zu erwartende Klugheit widerlegen, weil er sich durch das Mandat strafbar machen würde.“

⁹¹ BGBl. I 2016, 3152.

⁹² Kleine Anfrage 19/21307 v. 28.7.2020, nachzulesen unter: <https://dserver.bundestag.de/btd/19/213/1921307.pdf>.

⁹³ „Plattformen-Steuertransparenzgesetz“ v. 20.12.2022, BGBl. I 2022, 2730.

⁹⁴ Nachzulesen unter: https://ec.europa.eu/commission/presscorner/detail/de/qanda_22_7517.

⁹⁵ Vgl. z.B.: Danielmeyer, „Datenanalyse aus dem Blickwinkel der Außenprüfung“, beck.digitax 4/2023, 218.

⁹⁶ Huber/Reckendorf/Zisky, „Die Unveränderbarkeit der Kassenbuchführung § 146 Abs. 4 AO im EDV-Zeitalter und INSIKA – Teil 1: Historische Wurzeln und Realzustand“, BBK 12/2013, 567, 572.

⁹⁷ Wargowske/Werner, „Zur Fortentwicklung des Datenzugriffs gemäß § 147 Abs. 6 AO“, beck.digitax 2/2020, 66, 68 (u. 3/2020, 133).

den heutigen Aufzeichnungs- und Aufbewahrungsweisen angepasste Compliance-Standards⁹⁸ können dabei durch verbesserte Nachvollziehbarkeit hilfreich sein, weshalb die Gesetzgebung der Forderung des Bundesverbands der Deutschen Industrie e.V. (BDI), den Prüfungsprozess an „Tax Compliance Management Systemen (TCMS)“ zu orientieren⁹⁹, gefolgt ist.¹⁰⁰ Zum Verifikationsprinzip¹⁰¹ gehört grundsätzlich auch die Überprüfung der Ordnungsmäßigkeit (als Grundlage für die inhaltliche Prüfung) dazu. Insofern setzt eine „teleologisch[e]“¹⁰² Anwendung der Ordnungsmaßgaben auf die komplexen, veränderlichen elektronischen Aufzeichnungs- und Aufbewahrungsarten heutzutage digitales Verständnis voraus.

Die Rechtsprechung hat zwischenzeitlich in diversen Verfahren auf die grundlegend veränderten Bedingungen der digitalen Ordnungsmäßigkeitsprüfung reagiert, die gem. der Normenkette § 158 Abs. 1 i.V.m. § 162 Abs. 2 S. 2 AO direkt bis zur Schätzung der Besteuerungsgrundlagen führen kann¹⁰³: Im Allgemeinen sind die Ordnungskriterien „funktional“¹⁰⁴ auf den Einzelfall anzuwenden, sodass im Besonderen bspw. das Fehlen der Systemeinstellungen einer elektronischen Kasse einen gravierenden Buchführungsmangel darstellt¹⁰⁵, denn „der ‚Beweisverderber‘ oder ‚Beweisvereitler‘ [darf] aus seinem Verhalten keinen Vorteil ziehen (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO)“¹⁰⁶.

Seit Einführung der digitalen Außenprüfung zum 1.1.2002 werden regelmäßig neue¹⁰⁷ bzw. erweiterte Methoden für die zeitgemäße **materielle Prüfung** in Veröffentlichungen¹⁰⁸ vor-

⁹⁸ IDW-Praxishinweis zu „Tax Compliance Management Systemen“ (IDW PS 980), nachzulesen unter: <https://www.idw.de/idw/idw-verlautbarungen/idw-ps-980.html>.

⁹⁹ Bundesverband der Deutschen Industrie e.V. (BDI), „Steuerliche Betriebsprüfungen – Optionen für zeitgemäße Reformen“, 13, nachzulesen unter: <https://bdi.eu/publikation/news/steuerliche-betriebspruefungen/>.

¹⁰⁰ Art. 97 § 38 EGAO mit der Systemprüfung von Steuerkontrollsystemen und der Möglichkeit, Beschränkungen der nächsten Außenprüfung zuzusagen.

¹⁰¹ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

¹⁰² Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 12.

¹⁰³ Vgl. z.B.: Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, 3550: „Liegen keine materiellen Buchführungsmängel vor, entscheidet die Gewichtigkeit der formellen Buchführungsmängel darüber, ob eine Schätzungsbefugnis des Finanzamts vorliegt.“

¹⁰⁴ FG Hamburg v. 26.8.2016 – 6 V 81/16, 1. Leitsatz.

¹⁰⁵ BFH v. 25.3.2015, BStBl. II 2015, 743, 5. Leitsatz.

¹⁰⁶ BFH v. 15.2.1989, BStBl. II 1989, 462: „[...] Zur Vermeidung eines solchen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

¹⁰⁷ Bei etlichen als „neu“ angepriesenen Datenanalysemethoden handelt es sich nur um Abwandlungen bekannter, die sich unter einem englischsprachigen Namen besser verkaufen lassen. Die methodischen Grundlagen der „regelbasierten (Zeit-)Reihenanalysen“ und „musterbasierten Häufigkeitsauswertungen“ ermöglichen auch das (schrittweise) Nachvollziehen anderer, tlw. sehr komplexer Analyseverfahren.

¹⁰⁸ Vgl. z.B.: Sosna, „Einsatz statistischer Methoden zur Risikoanalyse, Recherche und Lokalisierung von Steuerausfällen“, StBp 2000, 41 u. 68, Huber, „Weiterentwickelte und neue Methoden der Überprüfung, Verpro-

gestellt. Durch den Vorteil der digitalen Prüfung, große Datenmengen in sehr hoher Geschwindigkeit und flexibel¹⁰⁹ aufbereiten bzw. analysieren zu können, sollen die infolge der elektronischen Betriebsverwaltung gestiegenen Anforderungen (s. [Pkt. 2.1.](#)) besser bewältigt werden. Dabei haben auch die Medien¹¹⁰ und professionellen Entwicklungsunternehmen¹¹¹ den wahrnehmungsphysiologischen Vorteil der Datenverbildlichung (s. [Pkt. 2.](#)) erkannt.

Wesentlicher Trend der Entwicklung in Wirtschaft und Beratung ist hierbei eine hochgradige Automatisierung der Prozesse: „Die an Systemschnittstellen erforderlichen manuellen Tätigkeiten werden immer mehr durch den Einsatz von Robotic Process Automation ersetzt, Entscheidungsprozesse werden dabei durch den Einsatz programmierter, selbstlernender Entscheidungsheuristiken, die auf Basis künstlicher Intelligenz arbeiten, unterstützt.“¹¹² Derweil die Schlüsselqualifikationen für zielführende Datenanalysen „Statistik“ und „Visualisierung“ nicht trivial sind,¹¹³ dürfte – neben dem Effizienzfaktor „Kosteneinsparung“ – der treibende Motor hierfür sein, dass zwar viel in Digitalisierung investiert wurde und wird, sich die Anwender*innen mit einem sinnvollen Einsatz jedoch schwer tun.¹¹⁴ In der Konsequenz hatte der Bundestag die Enquete-Kommission „Berufliche Ausbildung in der digitalisierten Arbeitswelt“ eingerichtet.¹¹⁵ Nach Erfahrungen des Autors muss sich jedoch zunächst die Einsicht verbreiten, wie interdisziplinär komplex es ist, aus einer Vielzahl von Methoden(-varianten) und Softwareprodukten mit einer Unzahl an Werkzeugen intelligente Lösungswege für bestimmte Prüfungsfragen zu entwickeln, damit Aus- und Fortbildungskonzepte so gelingen können, dass sie eine systematisch-zielführende Breitennutzung der digitalen Analyse- mög-

bung und Schätzung“, StBp 2002, 199, 233, 258 u. 293, *Gebbers*, „Die Strukturanalyse“, StBp 2007, 225, 257 u. 289, *Posch*, „Digitale Ziffernanalyse“, StBp 2010, 338, *Hentrich/Pyrcek*, „Compliance und Fraud Monitoring im Zeitalter von digitaler Transformation und Big Data“, BB 2016, 1451, u.w.

¹⁰⁹ Neben der Verarbeitungsgeschwindigkeit ist ein zentraler Vorteil der digitalen Datenanalyse, dass Ordnungsmaßnahmen, Berechnungen und Modellierungen jederzeit und umfangreich abgeändert werden können.

¹¹⁰ Bspw. Wirtschaftswoche online v. 21.5.2015: „Fiskus 4.0: Mit Big Data auf Steuerfang“, [früher](https://www.wiwo.de/politik/deutschland/fiskus-4-0-mit-big-data-auf-steuerfang/11770850.html) nachzulesen unter: <https://www.wiwo.de/politik/deutschland/fiskus-4-0-mit-big-data-auf-steuerfang/11770850.html>.

¹¹¹ Vgl. z.B.: Internetseite der Fa. Caseware zu IDEA V10: „IDEA V10. Neue Funktionen für intuitive Datenanalyse“, nachzulesen unter: <https://www.caseware.net/webinare/detail/idea-v10-neue-funktionen-fuer-intuitive-datenanalyse/> oder Visualisierungsprogramm „Visual Analytics“ des amerikanischen, weltweit operierenden Softwarehauses Statistical Analysis Systems, nachzulesen unter: http://www.sas.com/de_de/software/business-intelligence/visual-analytics.html.

¹¹² *Kompenhans/Wermelt*, „Big Data – Potenziale für innovative Abschlussprüfungen“, BB 2018, 299.

¹¹³ *Ihden*, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 64 u. 89.

¹¹⁴ Internetseite von McKinsey & Company, „Ten red flags signaling your analytics program will fail: [...] These days, it's the rare CEO who doesn't know that businesses must become analytics-driven. Many business leaders have, to their credit, been charging ahead with bold investments in analytics resources and artificial intelligence (AI). [...] But frustrations are beginning to surface; it's starting to dawn on company executives that they've failed to convert their analytics pilots into scalable solutions“, May 2018, nachzulesen unter: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail#>.

¹¹⁵ Nachzulesen unter: <https://www.bundestag.de/dokumente/textarchiv/2018/kw39-pa-enquete-berufliche-bildung-567960>.

lichkeiten erreichen (s. [Pkt. 3.3.](#)). Die Vorteile der „Künstlichen Intelligenz (KI)“ sind im Bereich der ermessensfreien Rechtsanwendung auf beschriebene Sachverhalte offensichtlich, wie Publikationen¹¹⁶ und werbende Internetauftritte¹¹⁷ unmissverständlich aufzeigen, **reichen aber darüber hinaus in datenanalytische Fähigkeiten hinein.**¹¹⁸ Trotzdem sind die erforderlichen Mindestkompetenzen für einen sinnvollen Praxiseinsatz „Lese- und damit Beurteilungsfähigkeit“ (s. [Pkt. 3.3.](#)), weshalb in der Wirtschaftsprüfung zu Recht die Grenzen des KIEinsatzes diskutiert werden.¹¹⁹

Die Erfahrungen von mehreren tausend Betriebsprüfer*innen mit Datenanalyse-Netzwerken wie der SRP lassen erahnen, dass Menschen **mit digitaler Analysekompetenz** Sachverhaltsaufklärung qualifiziert gelingt,¹²⁰ wobei Prüfungsabläufe i.d.R. nicht linear sind, sondern unter neuen Erkenntnissen ihre Ausrichtung und Fokussierung häufig ändern. Unter Begriffen wie „Dashboards“¹²¹ zeichnet sich allerdings bereits das nächste Analysemantra „viel hilft viel“ ab, indem auf Knopfdruck eine Vielzahl von Visualisierungen eingepflegter Daten ausgegeben werden. Hierin sollte die Gefahr erkannt werden, dass Datenverbildlichungen sehr unterschiedlich zur Aufklärung bestimmter Prüfungsfragen geeignet sind (s.a. *folgenden Abschnitt „SRP“*) und mitunter zu unzutreffenden Schlussfolgerungen führen können – **Methodik-Verständnis lässt sich durch Technik nicht ersetzen!**

¹¹⁶ Haufe online v. 11.3.2018: „Legal Tech und Algorithmen zur Vertragsanalyse: [...] ‚LawGeex‘ hat gemeinsam mit Professoren der Stanford University, der University of Southern California und der Duke University School of Law ein Experiment durchgeführt, bei dem sich ein Algorithmus mit Top-Anwälten messen musste. 20 Anwälte erhielten die Aufgabe, 5 ‚Non-disclosure agreements‘ (Geheimhaltungsverträge) zu überprüfen und darin 30 versteckte rechtliche Probleme zu identifizieren. Die Anwälte benötigten im Schnitt mehr als eineinhalb Stunden und erkannten durchschnittlich 85 % der Rechtsprobleme. Der Algorithmus erreichte eine Genauigkeit bei der Problemerkennung von 94 % und benötigte dafür exakt 26 Sekunden“, nachzulesen unter: https://www.haufe.de/recht/kanzleimanagement/legal-tech-steht-der-anwalts-roboter-schon-vor-der-tuer_222_445476.html.

¹¹⁷ Vgl. z.B.: Internetseite der Fa. Litax®: „Litax wurde bereits mit einem Algorithmus zur vollautomatischen Prüfung der grunderwerbsteuerlichen Konzernklausel ausgestattet. Gemeinsam mit einem deutschen Technikkonzern wurde ein Algorithmus entwickelt, der auf diese Klausel hin Umwandlungen und Umstrukturierungen im Konzern prüft und ein begründetes Gutachten ausgibt“, nachlesbar unter: <http://litax.net/> oder Internetseite der Fa. Taxy.io®: „Taxy.io KI Beta Club – Nutzen Sie jetzt die Gelegenheit, [...] die neuesten KI-Produkte testen [zu] können“, nachzulesen unter: <https://www.taxy.io/ki-beta-club>.

¹¹⁸ Vgl. z.B.: Webinarreihe „TAXPUNK meets ChatGPT [...] 4. BEYOND TEXTFILES [...] mit] Stefan Werner“, nachlesbar unter: <https://www.psp.eu/de/chatgpt-webinarreihe>.

¹¹⁹ DB online v. 30.4.2019 (Sponsored News v. EY), „Die Zukunft der Wirtschaftsprüfung ist digital: [...] Während Künstliche Intelligenz (KI) und Robotics Routinetätigkeiten übernehmen, konzentrieren sich die Prüfer von EY zunehmend auf die Interpretation der Daten. [...] Für Wirtschaftsprüfer heißt das, dass sie neben ihrem Fachwissen Kenntnisse im IT-Bereich benötigen, um diese Prozesse zu verstehen“, früher nachzulesen unter: <https://www.der-betrieb.de/meldungen/die-zukunft-der-wirtschaftspruefung-ist-digital/>.

¹²⁰ Giezek/Wähnert, „Künstliche Intelligenz vs. anwendergeleiteten Softwareeinsatz – Ist KI generell leistungsfähiger oder nur besser zu verkaufen?“, DB 31/2019, 1707.

¹²¹ Nachzulesen z.B. unter: https://docu.caseware.net/Audit/20.00/AuditTemplate/de/Content/User_Client_File_Setup/t_Using_the_Work_Flow_Panel.htm.

Die **SRP** verfolgt deshalb bei der Umsetzung der digitalen Außenprüfung folgende Strategie: Im Zentrum steht das **Verständnis der einzelnen Prüfungsmethoden**. Die Fortbildungen und Hilfsmittel – insbesondere die SRP-Vorlagen – sollen die technische Umsetzung erleichtern und beschleunigen, **ohne den Prüfungsprozess zu übernehmen**. Der Grund für diese Konzeptausrichtung liegt in den folgenden zwei Überlegungen:

- ! Viele Datenanalysemethoden sind für bestimmte Fragestellungen (sehr) gut geeignet, für andere hingegen nicht (s. *das Bsp. zum folgenden Abschnitt „Methodenwissen“*). Um die Gefahr von Ergebnisinterpretationen auf Basis ungeeigneter Analyseansätze zu minimieren, werden Methodengrundlagen und -unterschiede sowie zielführende Kombinationen anhand des SRP-Prüfungsdukts „vom groben Überblick zu den Details“ vermittelt und in den SRP-Vorlagen technisch so zur Verfügung gestellt, dass nur geringe IT-Kenntnisse erforderlich sind – besonders die Entscheidungsbildung verbleibt immer in der Hand der Prüfer*innen.
- ! Während die – vom Menschen entwickelte – Datenanalyse (**bisher¹²²**) auf der Grundlage von Ordnung funktioniert, entscheidet das menschliche Gehirn erfolgreich chaotisch-assoziativ.¹²³ Somit muss der optimale Einsatz der digitalen Datenverarbeitung diese zwei sehr unterschiedlichen Analyseformen verbinden. Demgegenüber würde die Konzentration auf nur eine Prüfungsressource „Mensch“ oder „Computer“ den Verzicht auf wichtiges Prüfungspotential bedeuten. **Außerdem zielt die aktuelle KI-Gesetzgebung in Europa auf die transparente Nachvollziehbarkeit von automatisierten Entscheidungsheuristiken ab.**¹²⁴

Am Anfang jeder (zielgerichteten) Datenanalyse steht eine zu klärende Frage. Für die steuerliche Außenprüfung ergibt sich diese aus den gesetzlichen Aufträgen der Plausibilitätsprüfung des Gesamtbilds¹²⁵ von (Teilen der) Besteuerungsaufzeichnungen gem. § 158 **Abs. 2 Nr. 1** AO sowie der effizienten¹²⁶ Aufdeckung von betriebsspezifischen Einzelprüffeldern.

¹²² Auf Lernprozesse von neuronalen KI-Netzwerken muss das nicht zutreffen.

¹²³ Intuitiv scheint dieser Unterschied den Führungskräften der Wirtschaft bewusst zu sein: „Manager vertrauen mehr auf ihr Bauchgefühl als auf B[usiness]I[ntelligence]“ zitiert CRN online v. 24.2.2010 eine entsprechende Studie.

¹²⁴ Europäisches Parlament: „KI-Gesetz: erste Regulierung der künstlichen Intelligenz“, nachzulesen unter: <https://www.europarl.europa.eu/news/de/headlines/society/20230601STO93804/ki-gesetz-erste-regulierung-der-kunstlichen-intelligenz>.

¹²⁵ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5: „Die gesetzliche Regelung des § 158 gründet sich insofern auf die Einsicht, dass es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild einer Buchführung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann.“

¹²⁶ § 7 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen. Ihre Dauer ist auf das notwendige Maß zu beschränken. Sie hat sich in erster Linie auf solche Sachverhalte zu erstrecken, die zu endgültigen Steuerausfällen oder Steuererstattungen oder -vergütungen oder zu nicht unbedeutenden Gewinnverlagerungen führen können.“

Im nächsten Schritt sollten alle¹²⁷ in Betracht kommenden Methoden und Werkzeuge zur Beantwortung der Prüfungsfrage auf ihre jeweilige Eignung hinterfragt werden, weshalb *Mochty* mit einem „hohe[n] Maß an **Methodenwissen**, Flexibilität und Kreativität“¹²⁸ große Anforderungen für das Prüfungspersonal sieht. *Ehlers* erwartet permanente „Maßnahmen der Qualitätssicherung“¹²⁹. Theoretische Überlegungen und praktische Erfahrungen sollten bei der Eignungsbewertung von Prüfungsansätzen zusammenkommen, wie es in der wissenschaftlichen Validierung üblich ist.

In Rechtsprechung¹³⁰ und Literatur¹³¹ finden sich zwar z.B. zur Zeitreihenanalyse umfangreiche theoretische Erörterungen, die jedoch oftmals von der inzwischen sehr umfangreichen Praxis (tlw.) widerlegt werden (*s. das nachfolgende Bsp. unter diesem Pkt. mit dem Vergleich zweier gebräuchlicher Zeitreihenvarianten*). „Prinzipiell sind zahlreiche Varianten eines Zeitreihenvergleichs denkbar“¹³², erkennt der BFH in seinem Zeitreihenurteil, ohne die Unterschiede zu ergründen, und lässt sich dennoch zu einem Generalurteil über alle Zeitreihenanalysen mit strengen Anwendungsbeschränkungen hinreißen¹³³. Demgegenüber nimmt das FG Hamburg zur unterschiedlichen Ausrichtung von Zeitreihenvarianten konkret Stellung: „Mit Hilfe der doppelskalierten Monatszeitreihe hat die Betriebsprüfung die Daten des Wareneinkaufs den Umsatzerlösen in zwei Einzelgrafiken gegenüber gestellt. Bei dieser Methode tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander. Aus kaufmännischer Sicht wird in einem Unternehmen auf jede Veränderung der Nachfragesituation (annähernd) adäquat mit einer Erhöhung bzw. Absenkung der Einkaufssumme reagiert“¹³⁴. Unglückli-

¹²⁷ Der Bundesrechnungshof fordert „den bundesweit einheitlichen Einsatz IT-gestützter quantitativer Prüfungsmethoden in der steuerlichen Betriebsprüfung [... denn n]ur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen“, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter:

<https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

¹²⁸ Internetseite des Lehrstuhls für Wirtschaftsprüfung, Unternehmensrechnung und Controlling von Prof. Mochty an der Universität Duisburg-Essen, früher nachzulesen unter: <https://www.wpunco.wiwi.uni-due.de/forschung/pruefungstechnik>.

¹²⁹ *Ehlers*, „Tax Compliance – Neue Beratungsfelder bei Betreuung kleiner und mittlerer Unternehmen“, NWB 18/2012, 1535, 1541.

¹³⁰ Vgl. z.B. BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 44 ff. mit Warnungen des BFH vor möglichen „Hebelwirkungen“ zuzüglich Kritik von *Wolenski* in „Das Zeitreihenurteil des X. Senats des BFH vom 25. 3. 2015“, Stbg 6/2016, 268 (mit einer beispielhaften Verdeutlichung auf Seite 269 und in Fußnote 5), weil die dazugehörigen Rechenbeispiele auf die besondere Zeitreihenvariante und den Einzelfall begrenzt sind.

¹³¹ S. z.B. die erheblichen Sorgfaltsanforderungen an den Einsatz von Zeitreihenanalysen in *Wiggen*, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168, die regelmäßig unreflektiert von den Finanzgerichten übernommen werden wie z.B. in BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 55 ff.

¹³² BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 37.

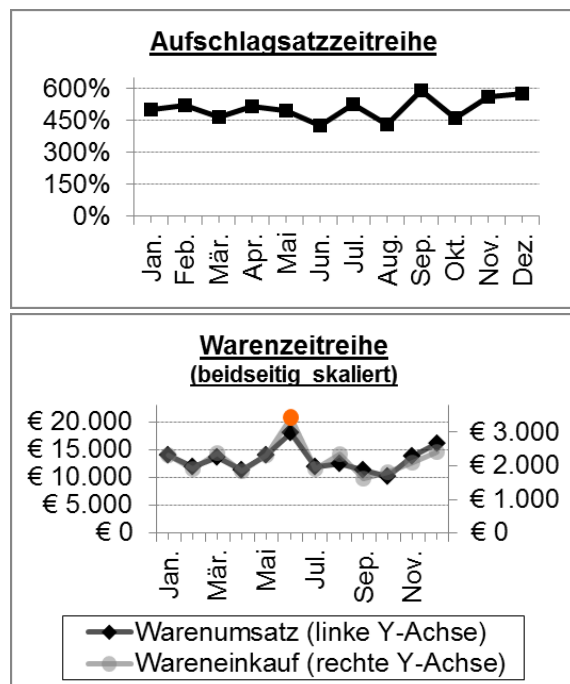
¹³³ BFH v. 25.3.2015, BStBl. II 2015, 743, 1. Leitsatz.

¹³⁴ Beschl. v. 31.10.2016 – 2 V 202/16, Rz. 48.

cherweise überwiegen selbst in der neueren Literatur die ersteren **undifferenzierten** Bewertungen.¹³⁵

In diesen zwei Abb. werden die Warenflussverhältnisse aus Einkauf und Umsatz eines gastronomischen Betriebs mit Einnahmen-Überschuss-Rechnung gem. § 4 Abs. 3 EStG anhand zweier Typen von Monatszeitreihen wiedergegeben. Losgelöst davon, dass diese Fallkonstellation mit länger haltbaren Getränken und Aufzeichnung nach Bezahlung laut finanzgerichtlicher Einschätzung nicht für Zeitreihenanalysen in Betracht kommt, offenbaren die beiden Diagramme vor allem **abweichende** Ergebnisse:

1. Die monatliche Aufschlagsatzentwicklung zeigt deutliche Schwankungen von unter 450 % bis fast 600 %. Diese Beobachtung könnte als implausibles Unternehmerverhalten i.S.d. § 158 **Abs. 2 Nr. 1** AO gedeutet werden.
2. Die mittels Doppelskalierung größenangepasste Gegenüberstellung derselben Ausgangsdaten (als Monatssummen) ist von gut erkennbarer Abhängigkeit des Warennachkaufs zum Umsatz geprägt. Zudem präsentiert die Datenverbildlichung z.B. in dem **relativ zu hohen Einkauf im April** ein konkretes Einzelprüffeld.



Als analytische Gegenkontrolle bestätigt das statistische Maß für die Abhängigkeit zweier Datenreihen „Korrelationskoeffizient“¹³⁶ mit einem sehr guten Wert von +0,90 das untere Resultat der **zentralen SRP-Zeitreihe für Plausibilitätsfragen**, während die obere Variante einer Aufschlagsatzzeitreihe¹³⁷ dem nicht entspricht.

Wegen dieser fundamentalen Unterschiede in der Eignung von Methodenvarianten, welche bisher weitgehend als einheitlich angesehen und „abgeurteilt“ wurden, liegt der Fokus

¹³⁵ Vgl. z.B.: *Steinhauß*, „Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388, worin zwar die methodologischen Differenzierungen und Abwägungen der SRP erwähnt werden, die Beurteilung aber wieder auf der nicht differenzierenden Rechtsprechung fußt.

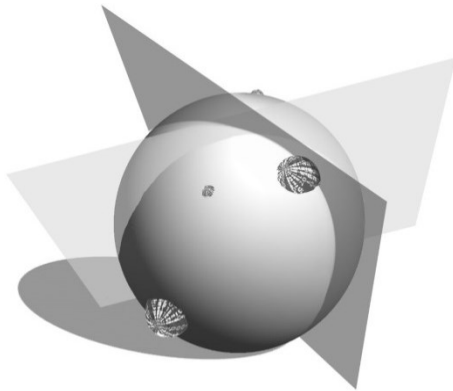
¹³⁶ Als lineares Abhängigkeitsmaß nimmt der Korrelationskoeffizient Größen von „-1“ für die 100 %-ige Gegenläufigkeit der Grafen über „0“ bei nicht vorhandener Beeinflussung bis zu „+1“ für eine vollständig (positive) Gleichrichtung ein.

¹³⁷ Auch dem BFH-Zeitreihenurteil (v. 25.3.2015, BStBl. II 2015, 743) liegt mit der dort eingesetzten „Gleitschlitzenzeitreihe“ [*Nöcker (RiBFH)*], „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548] eine Form der Aufschlagsatzzeitreihe zu Grunde.

der SRP auf einem ausreichenden Methodenverständnis, damit *zielführende* Datenvisualisierungen zu einer bestimmten Fragestellung verwendet werden.

2.3. Vorteile mehrperspektivischer Prüfungsnetze wie SRP

Dass **mehrperspektivische Betrachtungen** komplexer Objekte bzw. Situationen erheblich mehr Transparenz herstellen und damit eine wesentlich bessere Beurteilungsgrundlage liefern, ist z.B. aus der Medizin hinlänglich bekannt:



Einzelne Schichtaufnahmen von Körperteilen können nur (kleine) Ausschnitte der Gesamtinformation erfassen. So wird in nebenstehender Abb.¹³⁸ durch die zwei Schichten lediglich eine von vier Auffälligkeiten der Kugel – stellvertretend für ein komplexes Objekt – entdeckt. Seit die technischen Möglichkeiten vorhanden sind, wird deshalb in der Diagnostik und Operationsbe-

gleitung mit vielen Schichtaufnahmen gearbeitet, die entweder als Film abgespielt einen quasi-räumlichen Eindruck erzeugen oder computergestützt zu einer dreidimensionalen Rekonstruktion hochgerechnet werden.

Auch in der Datenanalytik führen mehrere *geeignete*¹³⁹ Verbirdlichungen als **Prüfungsnetz** zu einem besseren Verständnis der Informationen, weil jede

- Sortierung bspw. als Reihenlogik in einem Diagramm,
- Gruppierung wie bei klassifizierten Häufigkeitsauswertungen,
- Filterung vorzugsweise mit vernetzbaren Datenschnitten,
- Aufsummierung bzw. anderweitige statistische Zusammenfassung¹⁴⁰ zur vereinfachten Wiedergabe von umfangreichen Daten,
- Ausgestaltung regelbasierter Verbirdlichungen per (Zeit-)Reihenanalytik,
- Ausprägung musterbasierter Häufigkeitsauszählungen nach differenzierbaren Merkmalen wie z.B. dem USt-Schlüssel usw.

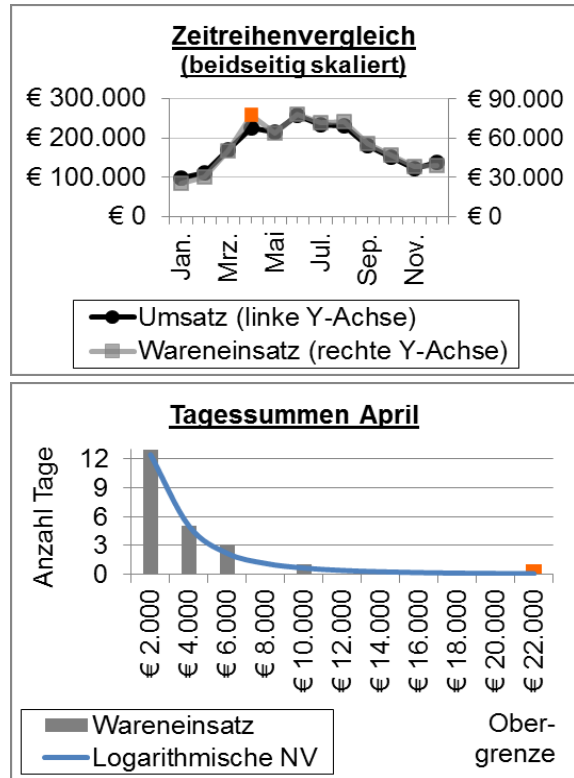
mit ihrer jeweiligen Fokussierung weitere Inhalte der Datenmenge offenlegen kann, sodass die Ergebniszusammenführung zu einem Gesamtbild eine viel vollständigere Beurteilungsgrundlage bildet als das Resultat eines einzelnen Prüfungsansatzes.

¹³⁸ Bildquelle: Mehret/Wähnert, „Prüfungsnetze vs. Einzelmethoden: Ein wichtiger, (bisher) verkannter Vorteil der Summarischen Risikoprüfung (SRP)“, DStR 2018, 314.

¹³⁹ Zusätzlich helfen mehrperspektivische Datenverbirdlichungen auch beim Erkennen (eher) ungeeigneter Analyseformen für eine bestimmte Fragestellung, indem die anderen bessere Antworten hinzufügen.

¹⁴⁰ Dem Zweck der beschreibenden (deskriptiven) Statistik folgend wird beim Aufsummieren bzw. Ermitteln von Kennzahlen die Gesamtinformation reduziert, um bestimmte (Kern-)Eigenschaften der Datenmenge herauszuarbeiten.

Folglich gelingt das Auffinden von betriebspezifischen Prüffeldern durch das zielführende Verknüpfen mehrerer regel- und musterbasierter Visualisierungen entlang des SRP-Prüfungsduktus „vom groben Überblick zu den Details“ auch in großen Datenmengen wirksam und zügig (s. [Pkt. 5.1.1.](#)). Als beispielhafter Prüfungspfad beleuchtet die mehrstufige Zeitreihenanalytik der SRP-Vorlage BWA die Entwicklungen und Abhängigkeit von Betriebsfaktoren wie in dieser Abb. u.a. mit dem monatlichen Warenfluss eines gastronomischen Betriebs als doppelskaliert größenangepasste Einzelgrafiken (s.a. [Pkt. 2.2.](#)). Zu dem hierdurch **aufgefallenen Monat „April“** ergänzt die SRP-Vorlage Stochastik wertgestaffelte Häufigkeitsauszählungen des Wareneinsatzes (*in der Abb. als Tagessummen*) – unter Gegenüberstellung der **spezifisch aus den Daten abgeleiteten Logarithmischen Normalverteilung** (s. [Pkt. 4.3.4.](#)). Wegen der zielführenden Kombination der beiden Verbildlichungsansätze lassen sich verursachende Vorgänge als risikoorientierte Prüffelder effektiv und effizient auf übersichtliche Teilbestände der Daten eingrenzen: **ein einzelner Tag mit außergewöhnlich hohem Wareneinsatz**. Die SRP-Vorlagen unterstützen solche systematischen Prüfungsprozesse mit ihrem Aufbau, der den Prüfungsduktus beinhaltet, und den zur Verfügung gestellten Werkzeugen; damit führen sie in methodisch-konzentrierter Form auch an die zunehmend in Wirtschaft und Beratung genutzten B(usiness-)I(ntelligence)-Programme¹⁴¹ heran.



Im Rahmen von indiziellen **Gesamtbildprüfungen** zu (Teilen von) Steuerungsdaten ist der positive Effekt mehrperspektivischer Analytik besonders groß. Während punktgenau entdeckte Einzelvorgänge anhand der Unterlagen i.d.R. weitgehend als Ursache-Wirkungs-Beziehungen wiederhergestellt werden können, gibt es diese einfachen (quasi-)¹⁴²kausalen

¹⁴¹ Microsoft wirbt für die eigene BI-Software „Power BI“ u.a. wie folgt: „Nutzen Sie die aktuellen Fortschritte bei der Microsoft-KI, um auch Laien das Aufbereiten von Daten, das Erstellen von Machine Learning-Modellen und das Auswerten strukturierter und nicht strukturierter Daten wie Text und Bildern zu ermöglichen“, **früher** nachzulesen unter: <https://powerbi.microsoft.com/de-de/why-power-bi/>. Mindestens die richtige Beurteilung multipler Analysen (Lesekompetenz) aus derartig funktionsreichen Programmen setzt aber Methodenverständnis und eine sinnvolle Prüfungssystematik voraus.

¹⁴² Da die Vergangenheit regelmäßig nicht vollständig und objektiv aufgezeichnet wurde, verbleiben auch bei der Beurteilung von Einzelsachverhalten Restunsicherheiten.

Zusammenhänge bei komplexen Sachverhalten nicht¹⁴³. So sind z.B. hinsichtlich der Vollständigkeit von Bareinnahmen vielfältige Szenarien für rechtskonforme bzw. -widrige Minderungen denkbar, die jeweils unterschiedliche Spuren in den Aufzeichnungen hinterlassen; als weiteres Beispiel gibt es bei konzerninternen Verrechnungspreisen aufgrund der vielen denkbaren Preisbildungseinflüsse häufig nicht den einen fremdüblichen Wert. Gesamtbild- bzw. Verprobungsverfahren zur Überprüfung leiden infolge der Komplexität des Prüfungssachverhalts folglich an wesentlichen Unklarheiten¹⁴⁴: bspw. können i.R. einer Kalkulation Einzelfaktoren wie der Warenverderb oder partielle Sonderpreise das Gesamtergebnis empfindlich verändern¹⁴⁵ bzw. preisbildungserhebliche Faktoren¹⁴⁶ nur phasenweise gewirkt haben. Dementsprechend müssen bei Schlüssigkeitsbeurteilungen eines Gesamtbilds Aussagen grundsätzlich unter erheblich größerer Unsicherheit getroffen werden.¹⁴⁷

Die Nutzung der digitalen Vorteile für eine systematisch-mehrperspektivische Verprobung eines komplexen Prüffelds kann die Aussagegewissheit über Plausibilitäten wesentlich verbessern, indem festgestellte Auffälligkeiten mit vertretbarem Aufwand hinsichtlich möglicher Besonderheiten des Einzelfalls hinterfragt werden. Das wirksame und zügige Auffinden außergewöhnlicher Sachverhalte (s.a. das Bsp. im Abschnitt „Prüfungsnetz“ oberhalb) ermöglicht es, diese in die Schlussfolgerungen angemessen einzubeziehen. Dabei hilft insbesondere die Interaktivität (Simulationsfähigkeit¹⁴⁸, s. [Pkt. 4.2.9.](#)), durch welche bspw.

¹⁴³ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 15: „Unser Dilemma im Umgang mit der Komplexität unserer Welt lässt sich darauf zurückführen, dass wir wohl darin ausgebildet wurden einfache logische Schlüsse zu ziehen und nahe liegende Ursache-Wirkungs-Beziehungen zu definieren. [...] Simple Ursache-Wirkungs-Beziehungen gibt es nur in der Theorie, nicht in der Wirklichkeit.“

¹⁴⁴ Mit einem Hinweis auf die unsichere Entscheidungssituation bei Bewertungen der rekonstruierten Vergangenheit: *Ihden*, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 145: „Die juristische Methodenlehre ist somit die Lehre von der richtigen und gerechten Entscheidung unter Unsicherheitslagen.“

¹⁴⁵ Vom BFH im Urf. v. 25.3.2015, BStBl. II 2015, 743, Rz. 44 ff., als „Hebelwirkungen“ im Zusammenhang mit Zuordnungsunschärfen in Zeitreihenanalysen bezeichnet.

¹⁴⁶ BMF-Schreiben v. 6.6.2023, „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093, Rz. 3.19.

¹⁴⁷ Da z.T. auch unter Autor*innen das Verständnis für die begrenzte Rekonstruierbarkeit einer (komplexen) Vergangenheit offenbar nicht ausreichend vorhanden ist, resultieren mitunter Aussagen wie: „Wahrscheinlichkeitstheoretisch geformte Algorithmen können Steueransprüche nicht begründen“ (*Prinz*, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4, M5). Eine realistische Einschätzung der Unvermeidbarkeit von Wahrscheinlichkeitsaussagen in Prüfungen und Gerichtsverhandlungen ist aber zwingende Voraussetzung für die (teleologisch) sachgerechte Einordnung des reduzierten Beweismaßes „Beanstandungsanlass“ gem. § 158 Abs. 2 Nr. 1 AO.

¹⁴⁸ Vgl. z.B.: PwC-Newsletter v. 31.8.2022: „Aus unserer Praxis: Der Einsatz von Data & Analytics im Verrechnungspreismfeld – quo vadis? [...] D&A-Lösungen finden aber auch vielfach Einsatz im Rahmen von Betriebsprüfungen, z. B. für die Simulation der Steuereffekte bei Diskussionen um fremdübliche Bandbreiten. Mittels BI-Lösungen können diese Effekte aus Diskussionen mit der Betriebsprüfung direkt und dynamisch simuliert werden, was die Entscheidungsfindung in der Diskussion möglicher Kompromissvorschläge deutlich verbessert und verkürzt“, nachzulesen unter: <https://www.pwc.de/de/newsletter/transfer-pricing-perspectives-dach/ausgabe-55-2022/aus-unserer-praxis-der-einsatz-von-data-and-analytics-im-verrechnungspreismfeld-quo-vadis.html>; dabei ist das systematisch-mehrperspektivische, interaktive Analysieren weniger eine Frage des Werkzeugs („BI-Lösungen“) als der zu Grunde liegenden Methodik.

- ☒ Untersuchungen weiter eingegrenzt,
- ☒ Einzeldaten nachvollziehbar korrigiert,
- ☒ Konjunkteinflüsse mittels Preisindizes bereinigt oder
- ☒ ganze Datenreihen im Datum verschoben werden können, um den jeweiligen Einfluss auf die Schlüssigkeit des Gesamtbilds zu konkretisieren – wie bspw. bei der Erfassung nach Bezahlung i.R. der Gewinnermittlung nach § 4 Abs. 3 EStG.

Nur durch diese Möglichkeiten kann die aus § 158 Abs. 2 Nr. 1 AO abgeleitete gemeinsame Sachaufklärungsverantwortung¹⁴⁹ im Prüfungs- bzw. Verprobungsprozess weitreichend und zeitsparend umgesetzt werden: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. Die Frage der Feststellungslast stellt sich, wenn ein entscheidungserheblicher Sachverhalt trotz Ausschöpfung aller zugänglichen und zumutbaren Ermittlungsmöglichkeiten nicht vollständig aufgeklärt werden kann. Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u.a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“¹⁵⁰ Von der wissenschaftlichen Literatur wird die Sorgfalt der SRP als allgemeinverständliches Qualitätsmerkmal bemerkt: „An die Beobachtungen aus Chi-Quadrat-Anpassungstest und Binomialtest schließen sich in der Praxis der Betriebsprüfung weitere Prüfungshandlungen an [...] Außerdem sind bei jeglicher Analyse die betrieblichen Besonderheiten des zu prüfenden Unternehmens einzubeziehen, um eine Verzerrung der Zahlenstruktur aufgrund von Preisgestaltungen oder bspw. Sitzplatzkapazitäten auszuschließen. Eine Übersicht der dafür durchgeführten Maßnahmen in der praktischen Anwendung der Betriebsprüfung bietet das Handbuch zur Summarischen

¹⁴⁹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen“.

¹⁵⁰ BMF-Schreiben v. 3.12.2020 „Verwaltungsgrundsätze 2020“, BStBl. I 2020, 1325, Rz. 68.

Risikoprüfung [...] Die Verteilung der Tageseinnahmen kann sich je nach betrieblicher Besonderheit bspw. in den einzelnen Geschäftsjahren, saisonal oder wochentagsspezifisch unterscheiden. Dies ist in den Analysen berücksichtigt. Ein solcher Umstand einer ‚uneinheitlichen Absatzsituation‘ wird in der Praxis ebenfalls regelmäßig und sehr tiefgehend geprüft¹⁵¹.

Neben der technischen Unterstützung liegt der Sinn der SRP-Vorlagen darin, das Risiko von Analysefehlern zu verringern und die systematisch-mehrperspektivische Prüfungsweise – soweit möglich bzw. sinnvoll – zu standardisieren sowie eine gute Nachvollziehbarkeit der Ergebnisausgaben¹⁵² zu gewährleisten.

Für die nach umfangreichen Praxiserfahrungen¹⁵³ typischen Gesamtbildergebnisse einer SRP mit redundanten¹⁵⁴ Auffälligkeiten bezüglich Ordnungsmäßigkeit und inhaltlicher Plausibilität (s.a. [Pkt. 4.4.](#) mit einem Bsp.) gilt zudem das beweisverstärkende Prinzip der Indizien-gesamtheiten¹⁵⁵.

¹⁵¹ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 139 u. 161.

¹⁵² Im Sinn „des Anspruchs der Steuerpflichtigen auf einen effektiven Rechtsschutz durch die Gerichte (Art. 19 Abs. 4, Art. 20 Abs. 3 GG)“ (BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 49).

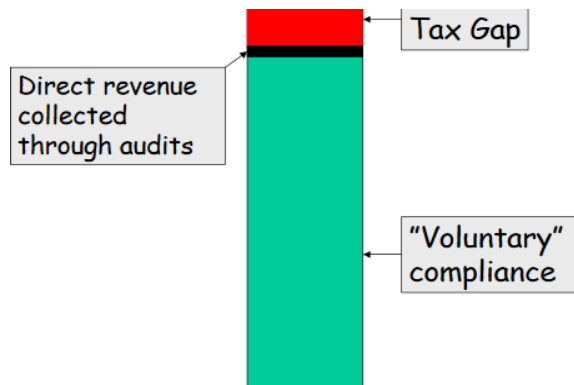
¹⁵³ Allein in den Pilotierungsjahren 2008 bis 2011 sind in Schleswig-Holstein ca. 1.000 Prüfungen mit der SRP auf die indizielle Ergebnislage hin abgefragt worden; vgl. a.: Landesrechnungshof Schleswig Holstein v. 6.2.2012, „Prüfungsmitteilung – Ergebnis der Prüfung ‚Organisation und Arbeitsweise der Betriebsprüfungsstellen‘“, LRH 13 – Pr 1667/2011, Pkt. 5.3. Auch in den anderen Bundesländern, in denen bis heute mehrere tausend Prüfer*innen Gesamtbildprüfungen mit der SRP durchführ(t)en, wird regelmäßig beobachtet, dass schätzungsrelevant unschlüssige Fälle gem. § 158 Abs. 2 Nr. 1 i.V.m. 162 Abs. 2 S. 2 AO mehrere unaufklär-bare Auffälligkeiten offenbaren – und nahezu immer auch gravierende Formalmängel.

¹⁵⁴ Vgl. z.B.: Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 24: „Allerdings entspricht es der Erfahrung, dass eine formell ordnungswidrige Buchführung i.d.R. auch sachliche Mängel aufweist“ oder Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstraßverfahren“, NZWiSt 1/2014, 24, 25: „In der Praxis wird sich die Frage, ob ein Zeitreihenvergleich allein die Beweisvermutung des § 158 AO widerlegt, [...] dadurch relativieren, dass die [...] aufgedeckten Unregelmäßigkeiten regelmäßig mit Unregelmäßigkeiten in der Kassenbuchführung einhergehen.“

¹⁵⁵ BGH v. 27.4.2010 – 1 StR 454/09, NStZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweismwürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreich[en].“

3. Einbindung der SRP in Prüfung und Risikomanagement

Ein wirksames **Risikomanagement** der Finanzverwaltung zur Sicherstellung der Gesetz- und Gleichmäßigkeit der Besteuerung muss bei über 8 Mio. registrierten Unternehmen und begrenzten Prüfungsressourcen¹⁵⁶ möglichst frühzeitig ansetzen, denn die Durchsetzung von (Steuer-)Gesetzen verlangt eine erfolgreiche Aufdeckung und spürbare Sanktionierung von Zuwiderhandlungen¹⁵⁷. Gemäß Erkenntnissen der Europäischen Union werden die allermeisten Steuern freiwillig entrichtet, wohingegen durch Prüfungsmaßnahmen lediglich ein Anteil im untersten Prozentbereich sichergestellt wird.¹⁵⁸ Deshalb kommt der Präventivwirkung verbreiteter, regelmäßiger und wirksamer Kontrollen eine überragende Bedeutung zu.



(Teil-)Automatisierte Maßnahmen wie RMS¹⁵⁹, RMS-BP¹⁶⁰, Textanalytik und Anomalieerkennung¹⁶¹ sowie erweiterter Meldepflichten¹⁶², ein zeitgemäßes Umsatzsteuersystem¹⁶³

¹⁵⁶ Laut Monatsbericht des BMF für den Oktober 2023 mit der Betriebsprüfungsstatistik 2022 waren 8.409.661 Betriebe in der Betriebskartei der Finanzämter erfasst und sollten von bundesweit 12.949 Prüfer*innen – also knapp 650 Betriebe pro Prüfer*in – intensiv geprüft werden; nachzulesen unter: https://www.bundesfinanzministerium.de/Monatsberichte/2023/10/Inhalte/Kapitel-3-Analysen/3-2-steuerliche-betriebspruefung-2022-pdf.pdf?__blob=publicationFile&v=4.

¹⁵⁷ Vgl. z.B.: Smettan, „Kriminelle Bereicherung in Abhängigkeit von Gewinn, Risiken, Strafen und Moral“, Kriminologische Forschungsreihe aus dem Max-Planck-Institut für ausländisches und internationales Strafrecht Bd. 63, 1992.

¹⁵⁸ European Commission, Directorate-General, Taxation and Customs Union, Indirect Taxation and Tax administration, Administrative cooperation and fight against fiscal fraud, „Risk Management Guide for Tax Administrations – Fiscalis Risk Analysis Project Group“, (Financial Code: FPG/11), Version 1.02 (February 2006), 60, nachzulesen unter: https://ec.europa.eu/taxation_customs/sites/taxation/files/resources/documents/taxation/tax_cooperation/general/overview/risk_management_guide_for_tax_administrations_en.pdf.

¹⁵⁹ Fokus Money online v. 26.3.2016: „Steuererklärung – Ihr Finanzamt stuft Sie jetzt nach Risiko ein – mit gravierenden Konsequenzen“, nachzulesen unter: https://www.fokus.de/finanzen/steuern/nicht-nur-die-schufa-ihr-finanzamt-stuft-sie-kuenftig-nach-risiko-ein-und-das-hat-konsequenzen_id_5257804.html. Die rechtlichen Grundlagen dafür sind mit der Ergänzung von Absatz 5 in § 88 AO durch das Gesetz zur Modernisierung des Besteuerungsverfahrens v. 18.7.2016 (anzuwenden ab 1.1.2017) geschaffen worden: „Die Finanzbehörden können zur Beurteilung der Notwendigkeit weiterer Ermittlungen und Prüfungen für eine gleichmäßige und gesetzmäßige Festsetzung von Steuern und Steuervergütungen sowie Anrechnung von Steuerabzugsbeträgen und Vorauszahlungen automationsgestützte Systeme einsetzen (Risikomanagementsysteme).“

¹⁶⁰ Vgl. z.B.: Danielmeyer, „Mit einem modernen Risikomanagement hin zur zeitnahen Betriebsprüfung?“, StBp 2021, 271.

¹⁶¹ Beide Analyseansätze werden in dem Kooperationsprojekt „Tax Defence Analytics (TaDeA)“ zwischen dem Landesamt für Steuern Niedersachsen und der Universität Oldenburg eingesetzt, um prüfungsrelevante Vor-

und die Kassen-Nachschau¹⁶⁴ sollen die digitalen Vorteile nutzen, um Steuerausfallrisiken in der Breite effektiver aufzudecken und damit die Prävention zu stärken. Im Vergleich zu diesen RMS-Lösungen setzt die SRP an (möglichst) unverdichteten Daten an, denn das Aufsummieren zu Periodenwerten (Anmelde- und Erklärungsdaten) vernichtet viele Spuren von Fehlern und Manipulationen (s.a. das Stichwort „Fuzzy-Theorien“ unter [Pkt. 2.](#) und das Zeitreihenbeispiel hierzu unter [Pkt. 4.2.2.](#)). Außerdem können die Qualitäten des Prüfungspersonals „Intuition“ und „Erfahrung“ bereits in den Risikoauswahlprozess (besser) eingebunden werden, indem trotz weitreichender technischer Entlastung weiterhin Menschen die Entscheidungen treffen.¹⁶⁵ Eine effektive Umsetzung der SRP ist in Deutschland nach derzeitigem Stand des Verfahrensrechts¹⁶⁶ jedoch erst i.R. von Außenprüfungen möglich.

Für die **Einführung der SRP** sind jeweils Schulungen (z.T. mit anschließenden Aufbauworkshops) zur vorlagengestützten und freien¹⁶⁷ Anwendung Voraussetzung. Hiermit werden die theoretischen Methodengrundlagen unterrichtet, um *geeignete* Verbildlichungsansätze zu bestimmten Fragestellungen zu wählen (s. den Abschnitt „Methodenwissen“ unter [Pkt. 2.2.](#)). Die SRP-Vorlagen unterstützen diesen Prozess, indem sie dem systematischen Prüfungsduktus „vom groben Überblick zu den Details“ folgen, Anwendungsregeln der verwendeten Methoden beachten, interaktive Optionen für das nachträgliche Einpflegen von neuen Erkenntnissen bzw. das Simulieren von Einlassungen (s. den Abschnitt „Gesamtbildprüfungen“ unter [Pkt. 2.3.](#) und das Beispiel unter [Pkt. 4.2.9.1.](#)) zur Verfügung stellen und eine ausreichende Dokumentation erleichtern. Der Einsatz der SRP zur zielführenden Prüffeldsuche (s. die Abschnitte „Prüfungsnetz“ unter [Pkt. 2.3.](#) und „systematisch-mehrperspektivische Musteranalytik“ unter [Pkt. 5.1.1.](#)) sowie zum Aufdecken und Abklären von Implausibilitäten gem. § 158 Abs. 2 Nr. 1 AO wird intensiv an Übungsbeispielen geschult; für spezielle Prüfungsthemen wie z.B. die datenbasiert-transaktionale Verrechnungspreisprüfung werden eigene Kurse angeboten. Die Bund-Länder-Unterarbeitsgruppe „Summarische Risikoprüfung (SRP) und visualisierende Datenanalyse“ koordiniert die Weiterentwicklung, die Aus- und Weiterbildungskonzepte sowie die Bereitstellung von technischen Hilfswerkzeugen. Zusätzlich be-

gänge in den umfangreichen strukturierten und unstrukturierten Besteuerungsdaten frühzeitig und vollständig erkennen zu können; nachzulesen unter: <https://tadea-ol.de/>.

¹⁶² Z.B. „Plattformen-Steuertransparenzgesetz“ v. 20.12.2022, BGBl. I 2022, 2730.

¹⁶³ Vgl.: EU-Projekt „VAT in the Digital Age (ViDA)“, nachzulesen unter: https://taxation-customs.ec.europa.eu/taxation-1/value-added-tax-vat/vat-digital-age_en.

¹⁶⁴ § 146b AO.

¹⁶⁵ Giezek/Wähnert, „Künstliche Intelligenz vs. anwendergelenkten Softwareeinsatz – Ist KI generell leistungsfähiger oder nur besser zu verkaufen?“, DB 31/2019, 1707.

¹⁶⁶ § 147 Abs. 6 AO: „Sind die Unterlagen nach Absatz 1 mit Hilfe eines Datenverarbeitungssystems erstellt worden, [...] hat die Finanzbehörde im Rahmen einer Außenprüfung das Recht, Einsicht in die gespeicherten Daten zu nehmen und das Datenverarbeitungssystem zur Prüfung dieser Unterlagen zu nutzen“.

¹⁶⁷ Die freie Anwendung des Methodenverständnisses und der Prüfungssystematik der SRP laufen als länderübergreifende Konzeptionierung auch unter dem Namen „Visualisierung“.

steht eine bundesweite Austauschplattform der Finanzverwaltung, über welche weiterführende Ideen zur Umsetzung der SRP-Prüfprinzipien in Echtzeit und thematisch sortiert geteilt werden können.

Damit ein reibungsloser Ablauf des dienststellenübergreifenden Besteuerungsverfahrens nicht behindert wird, gibt es angepasste Fortbildungsformate für alle involvierten Bereiche der Finanzverwaltung wie z.B. die Rechtsbehelfs- sowie Bußgeld- und Strafsachenstellen.

Ziel der SRP-Einführung ist es nicht, andere Herangehensweisen und Methoden der Prüfungsdienste zu verdrängen. Vielmehr ergänzt ein systematisch-mehrperspektivisches Gesamtbild über den Inhalt von Besteuerungsdaten zu Beginn einer Außenprüfung das Ergebnis einer Formalprüfung i.S. des § 158 Abs. 1 AO.¹⁶⁸ Danach ermöglicht die SRP als schnelle und leistungsfähige Orientierungsmaßnahme eine sachgerechte Einschätzung zur angemessenen Prüfungstiefe und zum Einsatz arbeitsaufwendigerer Prüfungsansätze wie z.B. einem Datenzugriff auf vorgelagerte Systeme („DavoS“¹⁶⁹) sowie eine möglichst weitreichende Aufklärung komplexer Sachverhalte innerhalb vertretbarer Zeit. I.R. von Plausibilitätsprüfungen nach § 158 Abs. 2 Nr. 1 AO können zeitintensivere Verprobungsverfahren wie Geldflussrechnungen oder Kalkulationen bedarfsgerecht eingesetzt werden, um den bisherigen Gesamteindruck am Maßstab der gesetzlich geforderten Gewissheit „Beanstandungsanlass“ (s. Pkt. 4.6.) ausreichend abzusichern.

Zur **Wahrung der Transparenz**¹⁷⁰ im Umgang der Außenprüfungsdienste mit den neuen digitalen Möglichkeiten¹⁷¹ sind zur SRP von Beginn an kontinuierlich aktualisierende Beiträge veröffentlicht worden¹⁷² – zuletzt auch das komplette Handbuch¹⁷³ und Kurzbeschreibungen

¹⁶⁸ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 24: „Allerdings entspricht es der Erfahrung, dass eine formell ordnungswidrige Buchführung i.d.R. auch sachliche Mängel aufweist.“

¹⁶⁹ Der gezielte Zugriff auf der Buchführung vorgelagerte Aufzeichnungssysteme wurde ursprünglich als Prüfungsprojekt „DavoS“ in Niedersachsen pilotiert und zwischenzeitlich auch in (fast) allen Bundesländern übernommen, weil sich die Erkenntnis durchgesetzt hat, dass viele prüfungsrelevante Auffälligkeiten in der häufig konsolidierten Buchführung nicht mehr ausreichend deutlich zu erkennen sind.

¹⁷⁰ Risse/Jotanovic/Krapp/Dallhammer, „Das steuerliche Kontrollumfeld: Katalysator für fortschrittliche Prüfsoftware“, RET 2023, 32, 35: „Allerdings legt die Finanzverwaltung ihre Aufbereitungsroutinen und Prüfungsalgorithmen sowie ihre Prüfungsmethodik nicht offen. Fußnote 18 dazu: Als Ausnahme Wähnert, Handbuch für die Summarische Risikoprüfung (SRP)“.

¹⁷¹ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 49: „Anspruch der Steuerpflichtigen auf einen effektiven Rechtsschutz [...] (Art. 19 Abs. 4, Art. 20 Abs. 3 GG)“. An dieser Stelle muss erwähnt werden, dass die SRP im Gegensatz zu vielen „Tax Compliance Management Systemen (TCMS)“ und Datenanalyselösungen der Wirtschaft und Beratung gerade *nicht* mit intransparenten Entscheidungsheuristiken und selbstlernenden Algorithmen (KI) arbeitet.

¹⁷² Erste Beschreibungen zur „Summarischen Risikoprüfung (SRP)“ finden sich in: Wähnert, „Summarische Risikoprüfung – Echtes Risikomanagement in der Finanzverwaltung mit Hilfe einer neuen Prüfungsmethode“, StBp 2008, 296. Aktuellere Darstellungen liefern neben dem Wikipedia-Eintrag unter https://de.wikipedia.org/wiki/Summarische_Risikoprüfung bspw.: Becker, „Außenprüfung digital – Prüfungsmethoden im Fokus (Teil II)“, DStR 2016, 1430, Nowotzin/Teutemacher, „Neue Wunderwaffe der BP: Summa-

(„Betriebsprüfung - SRP“ und „Betriebsprüfung - Quantilsschätzung“) im juris Lexikon Steuerrecht¹⁷⁴. Indem mit IDEA® und den auf MS Excel®¹⁷⁵ basierenden SRP-Vorlagen (sowie künftig voraussichtlich mit MS Power BI®-Lösungen¹⁷⁶) ausschließlich frei zugängliche Programme verwendet werden, ist die Reproduzierbarkeit der Prüfungsergebnisse gewährleistet. Zunehmende Datenmengen können hingegen nicht der Finanzverwaltung als Problem nachlassender Klarheit angelastet werden¹⁷⁷, weil dies eine zwangsläufige Folge der Umstellung auf die elektronische Verwaltung in der Wirtschaft ist (s. [Pkt. 2.1.](#)). Im Gegenteil sieht der Bundesrechnungshof hierin eine Situation, auf die mit einem stärkeren Einsatz „IT-gestützter quantitativer Prüfungsmethoden“ reagiert werden muss: „Mit der fortschreitenden Digitalisierung steigt das elektronische Datenvolumen in den Unternehmen beträchtlich. Traditionelle Methoden der steuerlichen Betriebsprüfung, beispielsweise Beleg- und Akteneinsicht, verlieren an Effektivität und Effizienz. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.“¹⁷⁸ Der offene Umgang der Finanzverwaltung mit ihren neuen Prüfungsansätzen fördert medial wahrnehmbar die Einsicht in Wirtschaft und Beratung, deren Aufgabe „Tax Compliance“ unter den hinzugewonnenen digitalen Optionen neu auszugestalten¹⁷⁹, wodurch der – *im Abschnitt „Risikomanagement“ unter diesem Pkt. benannte* – Hauptzweck der Besteuerungskontrolle „Prävention“ grundlegend gestärkt wird.

rische Risikoprüfung – Die neuen Prüfungstechniken und wie Steuerberater diese besser verstehen“, IWW BBP Sonderheft 2017, *Harle/Olles*, „Die moderne Betriebsprüfung“, NWB-Verlag, ab 3. Aufl. 2017, 221, NWB Betriebsprüfungs-Kartei, „Summarische Risikoprüfung (SRP)“, Stand: Februar 2017; Oberfinanzdirektion Nordrhein-Westfalen, in Zusammenarbeit mit der Finanzverwaltung Schleswig-Holstein, „Summarische Risikoprüfung (SRP) und Neue Prüfungstechnik (NiPt) – Wichtige Begriffe“, juris-Dokumenten-ID „bpnw-FMBP00000777“ (Pfad: „Fachinfosystem Bp > Allgemeine Prüfungshinweise > Arbeitspapiere und Leitfäden“), downloadbar auch unter: <https://taxtech.blog/2018/03/09/glossar-zur-summarischen-risikopruefung-srp-und-neuen-pruefungstechnik-nipt>.

¹⁷³ Downloadbar auf der Internetseite des Landes Schleswig-Holstein unter: https://www.schleswig-holstein.de/DE/fachinhalte/B/betriebspruefung_digital/Summarische_Risikopruefung.html.

¹⁷⁴ Ab 2. Auflage, Dez. 2021.

¹⁷⁵ Etwaige Übertragungen der SRP-Vorlagen auf andere Tabellenkalkulationsprogramme stammen aus den MS Excel®-Versionen.

¹⁷⁶ Vgl. z.B.: *Wargowske/Greil*, „Digitale Außenprüfung“, FR 2019, 608, sowie *Achilles/Jope*, „Corona-Dokumentation – Checkliste zur Vermeidung von Fehlinterpretationen in der Bp“, DB 2020, 1417.

¹⁷⁷ So einseitig geschehen in: BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 50.

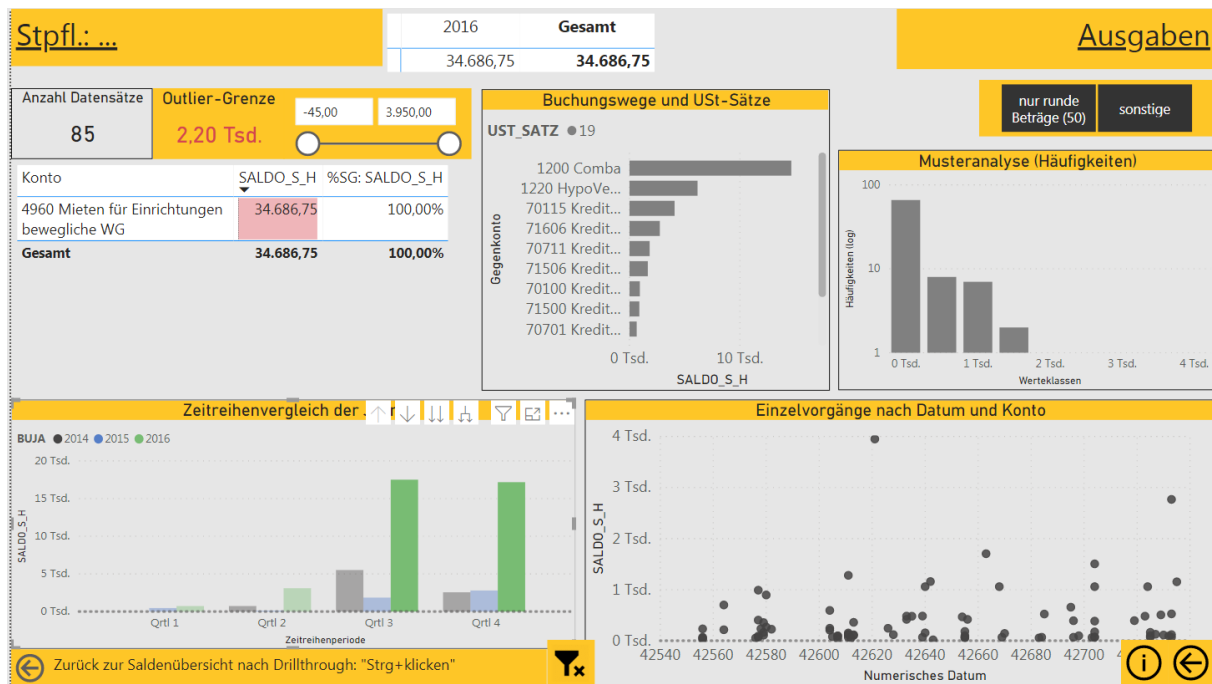
¹⁷⁸ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

¹⁷⁹ IDW-Praxishinweis zu „Tax Compliance Management Systemen“, nachzulesen unter: <https://www.idw.de/idw/idw-verlautbarungen/idw-ps-980.html>. Vgl. z.B. auch: *Liekenbrock*, „Digital Tax Compliance – Praxishinweise zur technischen Umsetzung von Tax Compliance Management Systemen“, Ubg 2018, 43, sowie *Ehlers*, „Tax Compliance – Neue Beratungsfelder bei Betreuung kleiner und mittlerer Unternehmen“, NWB 18/2012, 1535.

3.1. Die SRP zum Prüfungseinstieg und zur Prüffeldaufklärung

Haupteinsatzbereich der SRP ist die Erlangung eines aussagekräftigen Überblicks über und tiefen Einblicks in umfangreiche Datenbestände mit dem **systematischen Erkennen, Eingrenzen und Aufklären betriebspezifischer Prüffelder** wie zu Beginn einer Außenprüfung bzw. eines komplexen Einzelthemas.¹⁸⁰ Im Anschluss an die Formalprüfung, ob Besteuerungsunterlagen (in angemessener Zeit) prüfbar sind, ist diese zielgerichtete Arbeitsweise die Voraussetzung für eine effiziente Ressourcenverwendung.

Durch die systematisch-mehrperspektivische Datenverbildlichung kann in geringer Zeit ein qualifizierter Eindruck zu Vertiefungsnotwendigkeiten auch bei sehr großen Datenmengen bzw. datenintensiven Prüfungsfragen erarbeitet und über sog. „Drillthrough“-Funktionen der direkte Zugriff auf die betreffenden Einzelbuchungssätze – also die prüfungsrelevanten Vorgänge – unmittelbar erreicht werden:



So kombiniert das beispielhafte MS Power BI®-Modell¹⁸¹ zur Vertiefung eines – beim visualisierten Summen-Salden-Überblick aufgefallenen – Kostenkontos nach Filterung auf das letzte herausstechende Halbjahr in Leserichtung die Verteilung der Buchungswege inklusive USt-Sätzen, die wertbezogene Häufigkeitsauszählung, die (drillbare¹⁸²) Chronologie sowie

¹⁸⁰ B(usiness-)I(ntelligence)-Programme verfolgen denselben Hauptzweck aus der Selbstanalysesicht der Unternehmen, nachzulesen z.B. unter: <https://www.bigdata-insider.de/das-sind-die-beliebtesten-business-intelligence-loesungen-a-467954/> mit: „Wie entwickelt sich das eigene Unternehmen? In welchen Regionen bricht der Absatz ein? Was macht eigentlich der Mitbewerber? Auf Business Intelligence zu setzen, das heißt, bei der Entscheidungsfindung auf Fakten anstatt auf das Bauchgefühl zu vertrauen!“

¹⁸¹ Werner-Tutorials zur Erstellung solcher Analysemodelle finden sich auf dem KTZ-Server (Intranet).

¹⁸² In einigen Analyseprogrammen werden beim Zuordnen eines Datumsfelds zu einem Zeitreihendiagramm die wichtigsten Konsolidierungsperioden „Jahr“, „Quartal“, „Monat“ und „Tag“ automatisch angelegt, sodass ohne

die Streuung der Einzelbuchungen und bietet mit der Schaltfläche „nur runde Beträge (50)“ sowie dem „Outlier-Grenze“-Set¹⁸³ vorkonfektionierte Erweiterungen zum zügigen Auffinden prüfungsrelevanter Sachverhalte. Dabei legen Regel- und Musterabweichungen in den Entwicklungen, Abhängigkeiten und Häufigkeitsverteilungen der untersuchten Besteuerungsdaten aus Risikogesichtspunkten sinnvolle Intensivierungsbereiche offen. Auch der Aufbau und die Funktionalitäten der SRP-Vorlagen sind so ausgestaltet, dass aufgefundene Regel- und Musterstörungen unter wenig Aufwand bis zu den verursachenden (Einzel-)Sachverhalten nachverfolgt werden können.

Das Effizienzgebot aus § 7 BpO¹⁸⁴, „[d]ie Außenprüfung [...] auf das Wesentliche abzustellen [und i]hre Dauer [...] auf das notwendige Maß zu beschränken“, erfordert ein **systematisches Vorgehen und den Einsatz leistungsfähiger Methoden und Werkzeuge**. Unter den heutzutage üblichen Bedingungen der elektronischen Betriebsverwaltung mit rasant ansteigenden Datenmengen und zusätzlichen bzw. veränderten spezifischen Fehler- und Manipulationsrisiken (s. [Pkt. 2.1.](#)) sind digitale Hilfsmittel in der Außenprüfung unverzichtbar¹⁸⁵, wobei der Bundesrechnungshof die Wichtigkeit der „IT-gestützten *quantitativen* Prüfungsmethoden“ herausstellt.¹⁸⁶ Der Methodenreichtum der digitalen Datenanalyse sowie die extrem schnelle Verfügbarkeit per Mausklick bergen jedoch die Gefahr, dass (auch) ungeeignete Verfahren eingesetzt bzw. nicht zielführende Abläufe gewählt werden oder die Lesekompetenz für Analyseergebnisse unzureichend ist.

Die SRP stellt deshalb als Basis ein ausreichendes Verständnis wichtiger Einzelmethoden mit deren Anwendungsmöglichkeiten und Eignungen bereit (s. [Pkt. 2.2.](#)). Davon ausgehend wird der planvolle Prüfungsduktus „vom groben Überblick zu den Details“ vorgegeben, der zum einen – auf der Grundlage beobachteter Regeln und Muster – das Aufspüren von

weitere technische Vorbereitungsschritte zwischen groben und feinen Periodisierungen gewechselt werden kann, wie es im Zeitreihenvergleich anhand der „Drill-Pfeile“ ersichtlich ist.

¹⁸³ Derartige Ausreißer-Erkennungs-Ansätze nutzen Erfahrungen und statistische Kenngrößen wie die „Standardabweichung“ (typische Abweichung jedes einzelnen Datensatzes vom Mittelwert), um über deren dynamische Berechnung – also unter Beachtung von wählbaren Filterzuständen – das Auffinden von außergewöhnlichen Einzeldaten zu erleichtern.

¹⁸⁴ Durch das Gesetz zur Modernisierung des Besteuerungsverfahrens v. 18.7.2016 (anzuwenden ab 1.1.2017) ist das Effizienzgebot mit § 88 Abs. 2 S. 2 AO auch allgemein in den Untersuchungsgrundsatz aufgenommen worden: „Bei der Entscheidung über Art und Umfang der Ermittlungen können allgemeine Erfahrungen der Finanzbehörden sowie Wirtschaftlichkeit und Zweckmäßigkeit berücksichtigt werden.“

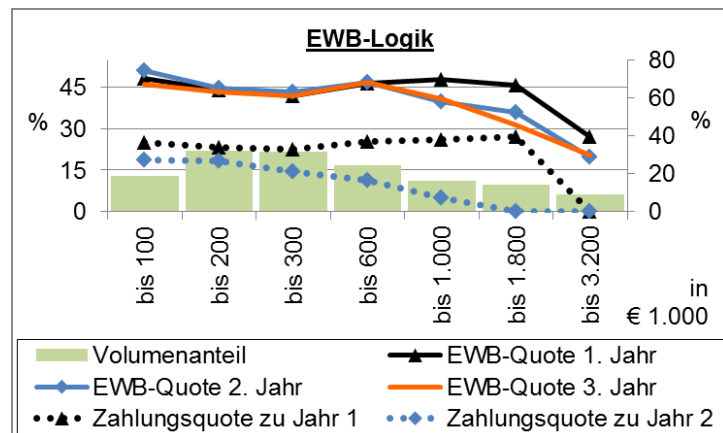
¹⁸⁵ Der Bundesverband der Deutschen Industrie e.V. (BDI) schlägt u.a. ebenfalls Maßnahmen im IT-Bereich für „schnellere, effizientere Betriebsprüfungen“ vor, nachzulesen unter:
<https://bdi.eu/publikation/news/steuerliche-betriebspruefungen/>.

¹⁸⁶ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter:
<https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

Systemfehlern sowie Einzelprüffeldern ermöglicht und zum anderen die zügige Eingrenzung und Abklärung unterstützt. Mit dieser Prüfungssystematik und den daran angelehnten SRP-Vorlagen wird das Prüfungspersonal in die Lage versetzt, seine Erfahrungen und Intuition zielführend in den komplexer gewordenen Prüfungssituationen zu nutzen.

Die vorlagenfreie Anwendung der SRP-Prüfungsprinzipien in Form von individuellen Lösungen für **komplizierte Einzelprüfungsfragen** wird in dem länderübergreifenden Methodik-Projekt „Visualisierung“¹⁸⁷ gebündelt und vermittelt prüfungstechnisch einen zielführenden Einsatz von IDEA®, Tabellenkalkulations- und B(usiness-)I(ntelligence)-Programmen (s. [Pkt. 5.](#)). So bewirkt eine fokussierte Verbildlichung unübersichtlicher Verhältnisse wie in

dem abgebildeten Beispiel¹⁸⁸ zur fortlaufenden Logik der Einzelwertberichtigungen (EWB) einer Bank i.d.R. ein besseres Sachverhaltsverständnis¹⁸⁹ – und folglich eine fundiertere rechtliche Entscheidung. In dem auf die wesentlichen Informationen konzentrierten Komplexdiagramm ist gut zu erkennen,



dass der – abgesehen von den höchsten Krediten – recht einheitliche EWB-Anteil im ersten Jahr durch die ebenso gleichmäßige – farblich zugeordnete – Zahlungsquote im Folgejahr beantwortet wird. Danach wurde die **EWB-Quote der drei höchsten Kreditklassen im zweiten Jahr deutlich abgesenkt**, woraufhin die **nachträgliche Begleichung in diesen Bereichen auf (nahezu) Null** abfiel. Daraus sind **im dritten Jahr offensichtlich keine Konsequenzen gezogen** worden, denn der EWB-Anteil wurde für die hohen Kreditklassen weiterhin niedriger gehalten. Demnach lässt sich im Überblick eindeutig ablesen, dass die Bewertung der Kredite unterschiedlicher Größenklassen nicht nach denselben Maßstäben vorgenommen wurde. Die Erfahrungen der Prüfungspraxis mit visualisierender Datenanalyse nehmen auch im Bereich komplexer Einzelprüffelder der Groß- und Konzernbetriebsprüfung wie bspw. der Fremdüblichkeit von Verrechnungspreisen¹⁹⁰ ständig zu, was sich im BMF-Schreiben zu

¹⁸⁷ Innerhalb der Bund-Länder-Unterarbeitsgruppe „Summarische Risikoprüfung (SRP) und visualisierende Datenanalyse“ kümmert sich ein Arbeitskreis um die vorlagenbasierte „Basis-SRP“, einer um die vorlagenfreie Datenvisualisierung und ein weiterer um die Erarbeitung bzw. Anpassung neuer sinnvoller Methodenansätze.

¹⁸⁸ Abb. abgeleitet von: Wolenski/Weiskopf, „Praxisfall: Einsatz indirekter Prüfungsmethoden in der Großbetriebsprüfung am Beispiel der Einzelwertberichtigung“, BBK 3/2013, 128, 131.

¹⁸⁹ Interessanterweise werden zur Darstellung von Beziehungen verbundener Unternehmen üblicherweise verbildlichende Organigramme verwendet, wohingegen die eigentliche Prüfung der viel komplexeren Unternehmensdaten häufig noch auf der Basis von Tabellen geschieht.

¹⁹⁰ Wähnert, „Die „Summarische Risikoprüfung“ in der Auslandsfachprüfung – Datenbasiert-transaktionale Verrechnungspreisprüfung“, StBp 2022, 408.

„[a]utomationsgestützte[n] quantitative[n] Prüfungsmethoden in der steuerlichen Außenprüfung“ widerspiegelt: „Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich“¹⁹¹; die neuen „Verwaltungsgrundsätze Verrechnungspreise“ lesen sich ebenfalls datenbasiert-transaktionaler¹⁹². Entsprechend wird in der Literatur für diesen Bereich von Beratungsseite zunehmend die Einbindung von IT-Konzepten gefordert, damit die komplexen betriebswirtschaftlichen Beurteilungen auf möglichst klaren (einzeldatenbasierten) Verhältnisbeschreibungen vorgenommen werden können.¹⁹³ Inwieweit diese Vorteile der Datenverbildlichung in die Wirksamkeitsprüfung von Steuerkontrollsystemen¹⁹⁴ verfahrenstechnisch einfließen (können), wo vom Ziel her gerade die Beschleunigung von Prüfungsmaßnahmen ohne ein Nachlassen der Prüfungsqualität erreicht werden soll, wird der Praxistest zeigen.

3.2. Die SRP im besonderen Kontext des § 158 Abs. 2 Nr. 1 AO

Gesamtbild-/Schlüssigkeitsverprobungen zur Vollständigkeit von (Bar-)Einnahmen mit der SRP stellen zwar nur einen kleinen Ausschnitt der Anwendungsbreite dar, werden von Kritiker*innen aber gerne in den Fokus gerückt¹⁹⁵, denn „wenn der Steuerpflichtige über seine Angaben keine ausreichenden Aufklärungen zu geben vermag“ oder die Beweiskraft von Besteuerungsaufzeichnungen wegen eines Beanstandungsanlasses nach § 158 Abs. 2 Nr. 1 AO – infolge der eingetretenen Beweislastumkehr und fehlender Nachweise – aberkannt werden muss, löst dies gem. § 162 Abs. 1 bzw. Abs. 2 S. 2 AO die Amtspflicht der Schätzung aus.

Derweil die Besteuerung einen schweren Eingriff in die Eigentumsrechte der Bürger*innen bedeutet, soll das **Verifikationsprinzip** die Gleichmäßigkeit in der Anwendung der Steuergesetze gewährleisten.¹⁹⁶ Danach führt das ordnungswidrige Aufzeichnen und Aufbe-

¹⁹¹ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 4.

¹⁹² BMF-Schreiben v. 6.6.2023, BStBl. I 2023, 1093.

¹⁹³ Vgl. z.B.: Braun/Köppe, „Digitalisierung in der Verrechnungspreisfunktion – Entwicklungen und Chancen“, DB 8/2019, 394, Schöneborn, „Verrechnungspreismanagement 4.0 – Zum Stand der Digitalisierung in der Praxis und welche Kompetenzen für ein erfolgreiches IT-Projekt entscheidend sind“, DB 14/2019, 737, Krüger, „Chemie 4.0: Digitalisierung, Verrechnungspreisbestimmung und Betriebsprüfung“, DStZ 17/2019, 659, Schindler/Totzek, „Anwendungsmöglichkeiten von Data & Analytics im Verrechnungspreiskontext“, IWB 20/2019, 827 etc.

¹⁹⁴ Art. 97 § 38 EGAO mit der Systemprüfung von Steuerkontrollsystemen und der Möglichkeit, Beschränkungen der nächsten Außenprüfung zuzusagen.

¹⁹⁵ Vgl. z.B.: IWW online, PStR Praxis Steuerstrafrecht v. 17.9.2015: „Summarische Risikoprüfung: Das neue Einfallstor für Steuerstrafverfahren?“, nachzulesen unter:
<https://www.iww.de/pstr/schwerpunktthema/betriebspruefung-summarische-risikopruefung-das-neue-einfallstor-fuer-steuerstrafverfahren-f88749>.

¹⁹⁶ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld

wahren von Besteuerungsunterlagen, falls es die Prüfung durch „eine[n] sachverständigen Dritten innerhalb angemessener Zeit“¹⁹⁷ verhindert, genauso zur Schätzung der Besteuerungsgrundlagen wie ein „Anlass [...]“, die sachliche Richtigkeit zu beanstanden“¹⁹⁸, sofern diese nicht doch nachgewiesen werden kann (Beweislastumkehr). Durch dieses außergewöhnliche Beweislastverhältnis soll sichergesellt werden, „dass der ‚Beweisverderber‘ oder ‚Beweisvereiteler‘ aus seinem Verhalten keinen Vorteil ziehen darf [...] Zur Vermeidung [...] sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt.“¹⁹⁹ Folgerichtig heben die neuen, allgemein betitelten „Verwaltungsgrundsätze 2020“ die Bedeutung der Beweisnähe und die Zumutbarkeit der Sachaufklärung für die Finanzverwaltung hervor.²⁰⁰

Besonders komplexe Verprobungen wie bspw. zur Vollständigkeit von (Bar-)Einnahmen lassen sich anhand der regelmäßig unvollständigen Dokumentationen und nicht objektiven Erinnerungen der Vergangenheit nur unter (großen) Unsicherheiten durchführen. Deshalb ergeben sich in der Prüfungspraxis – speziell für diesen Bereich noch durch das verbreitete Vorliegen ordnungswidriger Aufzeichnungen²⁰¹ verstärkt – unsichere²⁰² Gesamteindrücke,

einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

¹⁹⁷ § 145 Abs. 1 S. 1 AO.

¹⁹⁸ § 158 Abs. 2 Nr. 1 AO.

¹⁹⁹ BFH v. 15.2.1989, BStBl. II 1989, 462.

²⁰⁰ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. Die Frage der Feststellungslast stellt sich, wenn ein entscheidungserheblicher Sachverhalt trotz Ausschöpfung aller zugänglichen und zumutbaren Ermittlungsmöglichkeiten nicht vollständig aufgeklärt werden kann. Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u. a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“

²⁰¹ Westfälische Nachrichten online v. 16.3.2013: „Beamte verwerfen immer mehr Kassenbuchführungen – Betriebe bangen vor Steuerprüfern: [...] In 95 Prozent der Fälle wird die Kassenbuchführung verworfen“, nachzulesen unter: <https://www.wn.de/Muenster/2013/03/Beamte-verwerfen-immer-mehr-Kassenbuchfuehrungen-Betriebe-bangen-vor-Steuerpruefern>.

²⁰² Die grundsätzlich erheblichen Unsicherheiten bei aus der Vergangenheit rekonstruierten komplexen Sachverhalten wie insbesondere der Vollständigkeit von (Bar-)Einnahmen werden von einigen Autor*innen offenbar nicht anerkannt, weil (scheinbar) punktgenaue Berechnungen z.B. i.R. von Kalkulationen darüber hinwegtäuschen; nachvollziehen lässt sich das an Aussagen wie: „Wahrscheinlichkeitstheoretisch geformte Algorithmen können Steueransprüche nicht begründen“ (Prinz, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4, M5). Wenn die grundlegende Bedeutung von Wahrscheinlichkeitsaussagen für Prüfung und Rechtsprechung nicht verstanden wird, kann auch die Notwendigkeit einer (te-

deren Steuerausfallrisiken nach dem Verifikationsprinzip nicht von der Gemeinschaft der Steuerzahler*innen getragen werden sollen. In der Konsequenz sieht das Gesetz die „Beweiskraft der Buchführung (und Aufzeichnungen)“ nach § 158 Abs. 2 Nr. 1 AO bereits dann als widerlegt an, wenn es „nach [einer] Verprobung [...] *unwahrscheinlich* ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt“²⁰³, bzw. für eine ähnlich unsichere Beurteilungslage, wenn „der von der Finanzbehörde ermittelte Verrechnungspreis zumindest wahrscheinlicher ist“²⁰⁴; entsprechend wird im BMF-Schreiben zu „[a]utomationsgestützte[n] quantitativen Prüfungsmethoden in der steuerlichen Außenprüfung“ deren Bedeutung mit „kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit“²⁰⁵ festgestellt. Angelehnt an der Konkretisierung des (juristischen) Beweises gem. BGH²⁰⁶ in Verbindung mit den probabilistischen Theorien zur Vereinheitlichung von Aussagen unter Unsicherheit (Wahrscheinlichkeitsaussagen) aus der Gutachtenpraxis²⁰⁷ ließe sich das entsprechende gesetzliche Gewissheitsmaß des „**Beanstandungsanlasses**“ verallgemeinernd bei ca. 95 % einordnen (s. [Pkt. 4.6.](#)).

Genauso wie bei den herkömmlichen Verprobungen „Vermögenszuwachs- oder Geldverkehrsrechnung“ und „Kalkulation“ müssten quantitative Gesamtbilderergebnisse wie aus einer SRP ebenfalls an diesem Maßstab des § 158 Abs. 2 Nr. 1 AO gemessen werden, was derzeit noch nicht geschieht²⁰⁸; die Psychologie hat mit der sog. „Verfügbarkeitsheuristik“²⁰⁹

leologisch) möglichst konkreten, sachgerechten Einordnung des reduzierten Beweismaßes „Beanstandungsanlass“ gem. § 158 Abs. 2 Nr. 1 AO nicht gesehen werden.

²⁰³ Anwendungserlass zu § 158 AO.

²⁰⁴ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 73.

²⁰⁵ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 1.

²⁰⁶ Urt. v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 26.

²⁰⁷ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Reihe „Polizei + Forschung“, Bd. 26, herausgegeben vom Bundeskriminalamt (BKA), Kriminalistisches Institut, Luchterhand Verlag, 2004.

²⁰⁸ Unlängst hat eine wissenschaftliche Autorin bemerkt, dass eine gleichmäßige Beurteilung der neuen und herkömmlichen Prüfungsmethoden (am gesetzlichen Maßstab des § 158 AO) offensichtlich nicht stattfindet: „Der generelle Einsatz von mathematisch-statistischen Methoden ist zulässig, da sie in den Bereich der naturwissenschaftlichen Erkenntnisse gehören. Ein Kritikpunkt, der im Zusammenhang mit mathematisch-statistischen Methoden genannt wird, ist die unzureichende Sicherheit der Verfahren. [...] Wie vorher bereits erwähnt, ist eine Rekonstruktion der Vergangenheit nicht vollständig möglich. Bei genauerer Betrachtung der herkömmlichen Verfahren wird deutlich, dass für diese umfangreiche Informationen und Unterlagen des Stpfl. sowie gewisse Annahmen vonnöten sind. Dadurch sind die Methoden zunächst nicht nur zeitintensiv, sondern auch sehr fehleranfällig. Die herkömmlichen Methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl., ohne die Hilfe mathematisch-statistischer Methoden. Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten in den wiederhergestellten Daten. Erwähnenswert ist zudem, dass das Problem der nicht Anwendbarkeit der herkömmlichen Methoden regelmäßig vorliegt, da bspw. nicht genügend Unterlagen für eine Rekonstruktion zur Verfügung stehen. Schluss-

eine Erklärung für dieses „mit zweierlei Maß messen“, indem Ergebnisunsicherheiten einfacher Berechnungen auf der Basis unklarer Ausgangsgrößen wie bei Kalkulationen²¹⁰ unterschätzt werden, während konkrete Unschärfen quantitativ ermittelter Ergebnismöglichkeiten überbewertet werden, weil die einfachen Berechnungen „verfügbarer“ sind. Die Vorteile systematisch-mehrperspektivischer Prüfungsnetze liegen – neben einer höheren Aufdeckungsquote – vor allem darin, dass entdeckte Unschlüssigkeiten über ergänzende Datenverbildlichungen abgesichert werden, was auch in der wissenschaftlichen Literatur bemerkt wird.²¹¹ Zusätzlich lassen sich neue Erkenntnisse z.B. aus Einlassungen interaktiv (s.a. den Abschnitt „Gesamtbildprüfungen“ unter [Pkt. 2.3](#) u. [4.2.9](#)) einpflegen, wodurch eine qualifizierte Bewertung der Mitwirkungsbereitschaft und -qualität Stpfl. an der Sachaufklärung i.S. des § 162 Abs. 1 AO möglich wird. Zudem ist das resultierende Gesamtbild stets auf dem aktuellen Prüfungsstand und besteht nach umfangreichen Erfahrungen mit der SRP²¹² im Fall beanstandeter Plausibilität i.d.R. aus mehreren Indizien (s.a. das Beispiel unter [Pkt. 4.4](#)). Da-

folgernd stellt sich die Frage, ob das geforderte Sicherheitsmaß bei den herkömmlichen Methoden überhaupt erreicht werden kann“, Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324.

²⁰⁹ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Nomos, 2018, 56-57 u. 65: „Bei der Verfügbarkeitsheuristik schätzen Individuen die [...] Wahrscheinlichkeit für das Eintreten eines Ereignisses unter Unsicherheit danach ein, wie leicht bzw. schwer es ihnen fällt, damit verbundene Erscheinungen aus ihrem Gedächtnis abzurufen. Dabei werden Ereignisse für wahrscheinlicher erachtet, wenn sie sich leichter vorgestellt werden können und sich zu ihnen eine direkte Verfügbarkeit herstellen lässt. [...] Im Rahmen juristischer Entscheidungen scheint es unerlässlich, von diesen Entscheidungsverzerrungen Kenntnis zu haben, da eine verzerrte Entscheidung den Denkgesetzen der Logik widerspricht [...]“.

²¹⁰ Brinkmann, „Die Begründung des Sicherheitszuschlags“, StBp 2018, 212, 215 f.: „Doch auch wenn theoretisch eine Nachkalkulation der Umsätze möglich ist, so wird eine größere Genauigkeit [...] allenfalls dann erreichbar sein, wenn ausreichend detaillierte Informationen erlangt werden können. Aus dieser Prämisse ergeben sich so gut wie immer [...] Schwachstellen einer Nachkalkulation.“

²¹¹ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 139 u. 161: „An die Beobachtungen aus Chi-Quadrat-Anpassungstest und Binomialtest schließen sich in der Praxis der Betriebsprüfung weitere Prüfungshandlungen an [...] Außerdem sind bei jeglicher Analyse die betrieblichen Besonderheiten des zu prüfenden Unternehmens einzubeziehen, um eine Verzerrung der Zahlenstruktur aufgrund von Preisgestaltungen oder bspw. Sitzplatzkapazitäten auszuschließen. Eine Übersicht der dafür durchgeführten Maßnahmen in der praktischen Anwendung der Betriebsprüfung bietet das Handbuch zur Summarischen Risikoprüfung [...] Die Verteilung der Tageseinnahmen kann sich je nach betrieblicher Besonderheit bspw. in den einzelnen Geschäftsjahren, saisonal oder wochentagsspezifisch unterscheiden. Dies ist in den Analysen berücksichtigt. Ein solcher Umstand einer ‚uneinheitlichen Absatzsituation‘ wird in der Praxis ebenfalls regelmäßig und sehr tiefgehend geprüft“.

²¹² Aus dem (Flächen-)Einsatz der SRP seit 2008 in zwischenzeitlich fast allen deutschen Bundesländern durch mehrere tausend Prüfer*innen sind entsprechend viele Fälle hervorgegangen, die infolge indizieller Beurteilung der inhaltlichen Plausibilität zur Schätzung geführt haben. Dabei wurde die Einschätzung in Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 24: „Allerdings entspricht es der Erfahrung, dass eine formell ordnungswidrige Buchführung i.d.R. auch sachliche Mängel aufweist“ nicht nur regelmäßig dahingehend bestätigt, dass formelle und materielle Mängel – sofern letztere noch feststellbar sind – gemeinsam auftreten, sondern auch insofern, dass Indizien üblicherweise nicht einzeln vorkommen.

mit ist die Wahrscheinlichkeit, dass die Abweisungsschwelle des § 158 Abs. 2 Nr. 1 AO erreicht wird, deutlich größer, was die BGH-Rechtsprechung grundsätzlich mit der Bedeutung von Indiziengesamtheiten festgestellt hat²¹³.

Sofern die Schätzungsnotwendigkeit gem. § 162 Abs. 2 S. 2 AO gegeben ist, bieten die SRP-Vorlagen zwei **Schätzungsoptionen** an, um die bereits gewonnenen tiefreichenden Erkenntnisse zu den Besteuerungsdaten für die Ermittlung einer sachgerechten Schätzungshöhe zu nutzen:

- ⊗ Die **Quantilsschätzung** ist eine reine Schätzungsmethode, mit der intern auf Basis der regelbasierten Zeitreihenanalysen unter Verwendung statistischer und betriebswirtschaftlicher Standardinstrumente die objektivierte Leistungsfähigkeit abgeleitet werden soll, und folgt dazu dem Gebot des BFH, „in dem gegebenen Schätzungsrahmen an die oberste Grenze zu gehen“²¹⁴. Dabei nutzt sie die bekannten Vorteile dieser wissenschaftlichen bzw. Praxiswerkzeuge für ein verallgemeinerbares Schätzungsverfahren. Sie gründet sich nicht auf eine allgemeingültige Gesetzmäßigkeit, denn Steuerhinterziehungen erfolgen nicht einheitlich, sondern sie nutzt gängige, anerkannte Methoden aus Statistik und BWL für die notwendigen²¹⁵ Schritte, um gleichmäßig (aus Zeitreihenanalysen) ein sinnvolles Schätzungsergebnis gewinnen zu können (s. [Pkt. 4.8.](#)).
- ⊗ Mit der **Verteilungsanalyse** als Häufigkeitsauszählung von täglichen Bareinnahmen nach Größenklassen eröffnet sich *in geeigneten Fällen* eine weitere Möglichkeit, intern eine Schätzungshöhe zu ermitteln. Da Wirtschaftsdaten nach den Erkenntnissen der Wissenschaft i.d.R. der Logarithmischen Normalverteilung (s. [Pkt. 4.3.4.1.](#)) folgen, kann aus dem Vergleich der beobachteten Werte mit einer fallspezifisch rekonstruierten Logarithmischen Normalverteilung zumindest der Umfang eines – manipulationspsychologisch begründeten – stärkeren Eingriffs in die höheren Werte abgeleitet werden. Damit ermittelt diese Form der Schätzung allerdings eher einen Mindestschaden, der somit den Anforderungen einer strafrechtlichen Schätzung entspräche (s. [Pkt. 4.9.](#)).

²¹³ BGH v. 27.4.2010 – 1 StR 454/09, NSTZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreich[en].“

²¹⁴ Urt. v. 9.3.1967, BStBl. III 1967, 349; aktuell bestätigt von: Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitzen versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, 3549.

²¹⁵ Die i.R. der Quantilsschätzung bewusst vorgenommenen Arbeitsschritte und Entscheidungen sind bei jeder Schätzungsdurchführung unumgänglich, was aber tlw. unbewusst geschieht. Auf der Grundlage der „Zerlegung von Schätzungsmethoden in Einzelschritte“ lassen sich die Verfahren hinsichtlich der zu erwartenden Genauigkeit gut vergleichen.

Tlw. wird inzwischen auch in der Literatur die größtmögliche Realitätsnähe von Gesamtbildanalysen mittels SRP betont.²¹⁶ Nach derzeitigem Stand der Rechtsprechung werden vom X. Senat des BFH mit der „Drei-Stufen-Theorie“ jedoch die herkömmliche „Vermögenszuwachs- bzw. Geldverkehrsrechnung“ und „Kalkulation“ vorgezogen²¹⁷, wobei die aufgeführten Kritikpunkte an der Quantilsschätzung ausnahmslos unzutreffend²¹⁸ sind (s. den Abschnitt „*unzutreffende Kritikpunkte*“ unter [Pkt. 4.8.3.](#)). Dagegen haben einige Finanzgerichte sie als differenzierte griffweise Schätzungsmethode anerkannt.²¹⁹ Dementsprechend wird die Quantilsschätzung auch im BMF-Schreiben zu „[a]utomationsgestützte[n] quantitative[n] Prüfungsmethoden in der steuerlichen Außenprüfung“ als eine mögliche quantitative Schätzungsmethode aufgeführt.²²⁰

3.3. Voraussetzungen für quantitative Prüfungsmethoden

Die Digitalisierung und Vernetzung der Welt haben mit der Transformation der meisten Informationen von der analogen Form in die elektronische, der permanenten Bereitstellung eines Großteils des menschlichen Wissens – zuzüglich vieler Unwahrheiten und Meinungen – sowie unzähliger Analysefunktionen für strukturierte Daten sowie unstrukturierte Texte zu einem **grundlegenden Wandel des Wissensmanagements** geführt, indem sich die erforderlichen Schlüsselqualifikationen für Bildung geändert haben; der Stifterverband für die Deutsche Wissenschaft e.V. drückt es so aus: „Data Literacy gehört zu den wichtigen Future Skills, um in der digitalen Welt in Wissenschaft, Arbeitswelt und Gesellschaft bestehen und

²¹⁶ Vgl. z.B.: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 218: „Ein relevantes und wertvolles Werkzeug der Finanzbehörden ist dabei die Summarische Risikoprüfung (SRP-Handbuch 2020) geworden, die [...] den methodischen Standard der praktischen Nutzung darstellt.“ Pump/Heidl, „Hilfestellung bei der Kassenführung durch den steuerlichen Berater – Ärger wegen § 162 AO durch unzulängliche Kassenbuchführung“, StBp 2014, 162, 163: „Dabei werden durch umfassende Erfassung und Darstellung der betrieblichen Verhältnisse [...] die Besteuerungsgrundlagen extrem realitätsnah ermittelt.“ Negative Stimmen wie z.B. Beyer, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, beschreiben die SRP regelmäßig unzutreffend, was hier bspw. an der Fokussierung der Zeitreihenanalytik auf den Aufschlagsatz „RAS“ unschwer zu erkennen ist, und begründen ihre Ablehnung mit diesen unzutreffenden Beschreibungen.

²¹⁷ Urt. v. 25.3.2015, BStBl. II 2015, 743, 4. Leitsatz.

²¹⁸ Wichtige Annahmen über die Quantilsschätzung sind z.B. auch in dem Beitrag Kulosa, „Umgang mit Schätzungen in der Betriebsprüfung – Teil 2“, AgrB 1/2022, 13, nicht richtig, wie u.a.: „Dabei werden monatliche Werte für den Rohgewinnaufschlagsatz ermittelt; allerdings ohne Unterscheidung zwischen Wareneinkauf und Wareneinsatz (dies bewirkt eine erhebliche Ungenauigkeit)“; ausführlich dazu: Wähnert, „Missverständnisse zur Quantilsschätzung – Schätzungen mit IT-gestützten, quantitativen Methoden“, DStR 2022, 1470.

²¹⁹ Vgl. z.B.: FG Hamburg v. 26.8.2016 – 6 V 81/16, v. 31.10.2016 – 2 V 202/16 u. v. 18.7.2017 – 6 V 119/17 sowie FG Berlin-Brandenburg v. 9.1.2017 – 4 V 4265/15 u. v. 30.6.2017 – 10 V 10219/16: „Soweit das FG Berlin-Brandenburg [...] demgegenüber die Quantilsschätzung verworfen und eine pauschale Erhöhung der Umsätze der dortigen Antragstellerin um 10 % für vorzugswürdig gehalten hat, folgt der Senat dem nicht, weil die Quantilsschätzung eine deutlich differenziertere Schätzungsmethode darstellt.“

²²⁰ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 27 ff.

teilhaben zu können. Damit gehört sie zu den bedeutendsten Digitalkompetenzen der Zukunft. Data Literacy bezeichnet die Fähigkeit, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden. Sie ermöglicht die systematische Umwandlung von Daten in Wissen und ist damit unverzichtbarer Bestandteil der Allgemeinbildung²²¹ – oder kurz zusammengefasst: „Kein Bereich mehr ohne Datenanalysen“²²². An den Schulen ist trotz frühzeitigen Informatikunterrichts²²³ offenbar noch keine angemessene Neuausrichtung gelungen, wenn „[e]in Drittel der Schüler [in der Mittelstufe ...] ‚nur klicken und wischen‘“²²⁴ könne. Auch die Politik hatte mit der Einrichtung der Enquete-Kommission „Berufliche Bildung in der digitalen Arbeitswelt“ auf diese disruptive Veränderung der Aus- und Fortbildungssituation reagiert.²²⁵

Durch seine Forderung nach einem stärkeren und „bundesweit einheitlichen Einsatz IT-gestützter *quantitativer* Prüfungsmethoden in der steuerlichen Betriebsprüfung [..., denn n]ur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen“²²⁶ hat der Bundesrechnungshof klargestellt, dass dieser Wandel auch die Finanzverwaltung betrifft, zumal die Bedeutung von Rechtswissen durch immer bessere Assistenzsysteme²²⁷ und zunehmende Automatisierungen²²⁸ abnimmt. Im Gegensatz zu digitalen Ordnungsfunktionen sind die erforderlichen **Schlüsselqualifikationen** „Statistik“ und „Visualisierung“ für ein zielführendes Arbeiten mit quantitativen

²²¹ Nachzulesen unter: <https://www.stifterverband.org/data-literacy-education>.

²²² SpringerProfessional online v. 10.1.2019: „Diese Future Skills benötigt Deutschland“, nachzulesen unter: <https://www.springerprofessional.de/aus--und-weiterbildung/personalentwicklung/diese-future-skills-benoetigt-deutschland/16278042>.

²²³ Vgl. z.B.: Internetseite „Bildungsserver Berlin-Brandenburg“, nachzulesen unter: <https://bildungsserver.berlin-brandenburg.de/unterricht/faecher/mathematik-naturwissenschaften/informatik/unterrichtsmaterialien-und-fachthemen/1-informatik-in-der-grundschule>.

²²⁴ Deutschlandfunk online v. 5.11.2019: „Digitale Kompetenz in der Mittelstufe – Ein Drittel der Schüler könne ‚nur klicken und wischen‘“, nachzulesen unter: <https://www.deutschlandfunk.de/digitale-kompetenz-in-der-mittelstufe-ein-drittel-der-100.html>.

²²⁵ Nachzulesen unter: <https://www.bundestag.de/dokumente/textarchiv/2018/kw39-pa-enquete-berufliche-bildung-567960>.

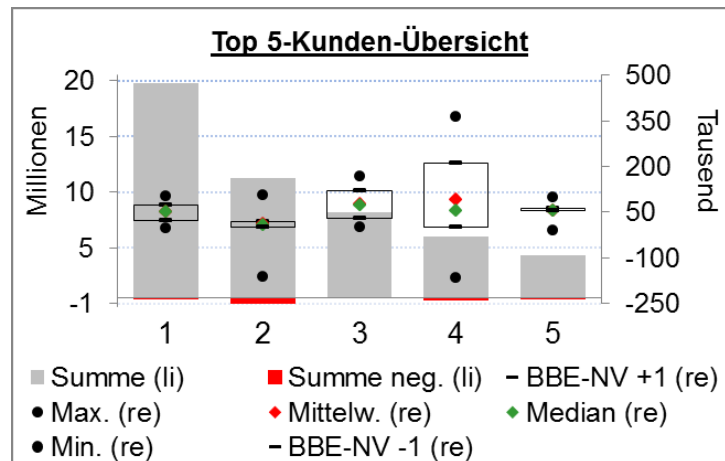
²²⁶ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

²²⁷ Vgl. z.B.: taxtech.blog v. 13.2.2019: „Steuergesetze in der semantischen Netzwerkanalyse – Taxy.io arbeitet an der semantischen Analyse von steuerrechtlichen Texten wie Gesetzen, Richtlinien und Urteilen, um ein maschinelles Verständnis dieser Texte aufzubauen und für individuelle Fragestellungen gezielte Lösungshinweise aus der Rechtslehre zu geben.“ Nachzulesen unter: <https://taxtech.blog/2019/02/13/steuergesetze-in-der-semantischen-netzwerkanalyse/>.

²²⁸ Vgl. z.B.: Internetseite der Fa. Litax®: „Litax wurde bereits mit einem Algorithmus zur vollautomatischen Prüfung der grunderwerbsteuerlichen Konzernklausel ausgestattet. Gemeinsam mit einem deutschen Technikkonzern wurde ein Algorithmus entwickelt, der auf diese Klausel hin Umwandlungen und Umstrukturierungen im Konzern prüft und ein begründetes Gutachten ausgibt“, nachlesbar unter: <http://litax.net/> oder Internetseite der Fa. Taxy.io®: „Taxy.io KI Beta Club – Nutzen Sie jetzt die Gelegenheit, [...] die neuesten KI-Produkte testen [zu] können“, nachzulesen unter: <https://www.taxy.io/ki-beta-club>.

Prüfungsmethoden nach den Erkenntnissen der Bildungswissenschaft nicht intuitiv oder trivial, sondern müssen gezielt unterrichtet werden.²²⁹ Allerdings finden sie sich bisher weder in der „Ausbildungs- und Prüfungsordnung für die Steuerbeamtinnen und Steuerbeamten – gehobener Dienst“²³⁰ noch in den Berufsexamina der Wirtschaftsprüfer*innen und Steuerberater*innen (in angemessener Form): „Beim Steuerberaterexamen fällt das vollständige Fehlen der Bereiche IT und Datenanalyse auf. [...] Defizite [innerhalb der berufsständischen Aus- und Weiterbildung des Wirtschaftsprüfers] bestehen vor allem im Bereich der Datenanalyse [...] Die Ausweitung des Lehrangebots – besonders an der Schnittstelle von Rechnungswesen, Steuerlehre und Steuerberatung zu IT(-Prüfung) und Datenanalyse – setzt eine angemessene Ressourcenausstattung, vor allem spezialisiertes Lehrpersonal voraus.“²³¹

Als Beispiel für die Andersartigkeit und interdisziplinäre Komplexität der Schlüsselqualifikationen zur quantitativen Datenlektur soll dieses Komplexdiagramm dienen: Durch die Kombination mehrerer Dateninhalte zu den Rechnungen einer großen²³² Anzahl an Personenkonten wird ein individualisierter Überblick für sinnvolle Vertiefungen erarbeitet. Dazu zeigen die Säulen sowohl die kundenbezogenen Rechnungssummen als auch den Gesamtwert der negativen Vorgänge – unter variabler Sortierbarkeit – an. Schwieriger zu lesen sind die Informationen zur Verteilung der Rechnungen pro Kunde mit den Stellvertreterwerten „Median“ und „Mittelwert“ u.a. zur Symmetrieeerkennung²³³, einer Bandbreiteneinengung anhand der Normalverteilungsverhältnisse²³⁴ sowie der Anzeige der jeweiligen Extrema. Mit so einer Diagrammgestaltung können Prüfer*innen verschiedene Kriterien zur eventuellen Vertiefung im vergleichenden Überblick erfassen und risikogerechte



²²⁹ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 64 u. 89.

²³⁰ Nachzulesen unter: https://www.gesetze-im-internet.de/stbapo_2022/anlage_11.html.

²³¹ Bravidor/Förster/Weißberger, „Berufsstand 4.0: Wirtschaftsprüfer und Steuerberater zwischen Fachexpertise, IT und Datenanalyse – Ansätze für die Ausbildung künftiger Wirtschaftsprüfer und Steuerberater an Hochschulen und in der Praxis“, WPg 2020, 287, 291, 292 u. 294.

²³² Zur Vereinfachung des Lesens wurde das Komplexdiagramm auf die wichtigsten fünf Kunden begrenzt.

²³³ Da der arithmetische Mittelwert im Gegensatz zum 50 %-Rang „Median“ von einseitigen Ausreißerwerten verzerrt wird, lässt sich aus der Lage zueinander die Symmetrie homogener Verteilungen ablesen.

²³⁴ Die Übertragung der (logarithmischen) Normalverteilungsverhältnisse auf zu untersuchende Daten sind als typische Zufallsstreuung ein guter Anhaltspunkt zur Abgrenzung von Ausreißern; je nach Betrachtung stehen dabei der Normbereich („Bandbreiteneinengung“) oder mögliche Ausreißer im Fokus, wozu arithmetische und Rangansätze wegen der unterschiedlichen Eignung wie im Box-Plot auch kombiniert werden.

Entscheidungen treffen, sofern die Darstellungslogik, die Begriffe und Funktionen der beschreibenden (deskriptiven) Statistik, die Eigenschaften von statistischen Kenngrößen und Diagrammarten, der variable Umgang mit diesen etc. hinreichend bekannt sind. Ohne diese Schlüsselqualifikationen bliebe nur das ziellose und nicht-BpO-konforme²³⁵ Scrollen, Sortieren und Filtern in den Daten. Infolge der sich rasant erweiternden Funktionalitäten „generativer Künstlicher Intelligenz (KI)“²³⁶ werden die technischen Fähigkeiten zur Erstellung solcher Auswertungen entbehrlicher werden, während die Lesefähigkeit – einschließlich kritischer Reflektion der Methodik – nicht an Bedeutung verliert.

Im **SRP-Methodik-Konzept** werden notwendige Begriffe und Funktionen beschrieben (s. z.B. zur Aufgabe von Wahrscheinlichkeitstests wie dem Chi-Quadrat-Test i.R. von Ziffern- oder anderen Musteranalysen unter [Pkt. 4.3.1.3.](#)), um Einzelmethoden sinnvoll einzusetzen. Zusätzlich wird deren Eignung hinterfragt (s. bspw. die Gegenüberstellung gebräuchlicher Zeitreihenvarianten für die Abhängigkeitsprüfung von Warenflüssen etc. inklusive Methodenvalidierung mittels statistischer Kenngrößen unter [Pkt. 4.2.](#)). Die Anwendung der Einzelmethoden erfolgt systematisch „vom groben Überblick zu den Details“ (s. beispielhaft für die regelbasierten Zeitreihenanalysen unter [Pkt. 4.2.2.](#)), damit keine entscheidungserheblichen Informationen übersehen werden. Außerdem werden die beiden Perspektiven „regelbasierte Zeitreihenanalysen“ und „musterbasierte Häufigkeitsauswertungen“ gemeinsam betrachtet, weil die unterschiedlichen Aspekte im Zusammenhang zu einem aussagekräftigeren Gesamtbild führen (s. [Pkt. 4.4.](#)) und mitunter bereits gegenseitige Erklärungen oder Konkretisierungen liefern (s. z.B. zur Trennung der Entgelte eines Betriebs mit mehreren Filialen im Bereich „systematische Kombination aus unterschiedlichen Datenverbildlichungen“ unter [Pkt. 5.1.1.](#)). Schlussendlich werden die methodisch-technisch erarbeiteten Ergebnisse der Sachaufklärung ins Verhältnis zu den steuerrechtlichen Vorschriften gesetzt (s. besonders [Pkt. 4.6.](#), [4.7.](#), [4.8.](#) u. [4.9.](#)). Für einen zentralen Teil der Prüfungsmethodik und -technik liefert die SRP somit exemplarisch die umfangreichen, interdisziplinären Grundlagen. Trotz der annehmbaren Akzeptanz der SRP unter mehreren tausend geschulten Prüfer*innen in fast allen Bundesländern sowie von der Bundesprüfung, beim Zoll, in der niederländischen Finanzverwaltung und über Kooperationen mit dem Internationalen Währungsfond (IWF) leidet auch dieses Methodik-Konzept am Fehlen propädeutischer Ausbildungsmaßnahmen, wie sie in (natur-)wissenschaftlichen Studiengängen üblich sind, denn ein für die freie Übertragung ausreichendes Verständnis komplexer Systeme setzt eine hohe Sicherheit bei den Begriffen, Funktionen, Zusammenhängen, Unterschieden etc. voraus.

²³⁵ § 7 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen. Ihre Dauer ist auf das notwendige Maß zu beschränken.“

²³⁶ Vgl. z.B.: Webinarreihe „TAXPUNK meets ChatGPT [...] 4. BEYOND TEXTFILES [...] mit] Stefan Werner“, nachlesbar unter: <https://www.psp.eu/de/chatgpt-webinarreihe>.

4. Die Verwendung der SRP-Vorlagen (Basis-SRP)

***Hauptzweck** der SRP ist die Vermittlung eines umfassenden Verständnisses „IT-gestützter quantitativer“²³⁷ Prüfungsansätze (Methodik-Konzept)²³⁸, um dadurch zu Beginn einer Außenprüfung zügig eine gute Orientierung in umfangreichen Steuerungsdaten zu erlangen und geeignete Methoden für die zielführende Bearbeitung unterschiedlichster Fragestellungen im Prüfungsalltag einzusetzen (s. Pkt. 2.). Hierfür stellen die SRP-Vorlagen einen abgestimmten Werkzeugkasten zur Verfügung, um die Vorteile systematisch-mehrperspektivischer Prüfungsnetze auch bei geringen analytischen Vorkenntnissen erfahren zu können, und erleichtern somit einen Einstieg in die Arbeitsweise der in Wirtschaft und Beratung zunehmend verbreiteten B(usiness-)I(ntelligence)-Programme.*

Die unter diesem Pkt. beschriebene Basisanwendung mit Hilfe der SRP-Vorlagen erfolgt exemplarisch für den Anwendungsbereich „Prüfung von (Bar-)Einnahmen auf Vollständigkeit“, weil hier der inhaltliche Zusammenhang der vielen regel- und musterbasierten Datenverbildlichungen entlang des SRP-Prüfungsduktes am allgemeinverständlichen erscheint und deshalb gut nachvollziehbar sein sollte. Für die freie Anwendung der SRP werden unter Pkt. 5. einige Beispiele vorgestellt.

Beiden SRP-Vorlagen ist als erstes Tabellenblatt eine **Kurzanleitung**²³⁹ vorgeordnet, welche jeweils den Versionsstand, im weißen Bereich eine kurze Beschreibung der Vorlage und generelle Bedienhinweise sowie im grünen (bis roten) Teil eine knappe Erläuterung der einzelnen Arbeitsblätter bzw. Analyseschritte entlang des SRP-Prüfungsduktes enthält. Auf den einzelnen Arbeitsblättern haben die ausfüllbaren Felder einen weißen Hintergrund (*Die anderen Felder sind mit einem Passwort schreibgeschützt.*) und werden die Funktionalitäten über Ankreuz- oder Auswahlelemente gesteuert. Über die roten Ecken oben rechts in den betreffenden Feldern sind Kommentare zu erkennen, die entweder in blauer Schrift als Erklärungs- oder in roter als Warnhinweise helfen sollen. U.a. dazu beinhalten viele Rechenfelder Plausibilitätsprüfungen, ob die eingegebenen Daten sinnvoll erscheinen und wichtige Überlegun-

²³⁷ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

²³⁸ Zwar wird ab der 3./12. Ausgabe des SRP-Handbuchs die Ausrichtung „Methodik-Konzept“ erstmalig explizit benannt, die entsprechenden Inhalte sind aber seit Begründung Bestandteil des Projekts.

²³⁹ In der SRP-Vorlage Stochastik ist die Kurzanleitung auf zwei Tabellenblätter aufgeteilt, um die Unterscheidung der allgemeinen Musteranalytik von der manipulationsorientierten deutlicher herauszustellen.

gen bzw. Arbeitsschritte beachtet wurden. Beim Einsatz von Wahrscheinlichkeitstests werden deren Anwendungsregeln berücksichtigt bzw. wird auf die Grenzen der Auswertbarkeit hingewiesen. Trotzdem können die Hilfestellungen in den SRP-Vorlagen nicht das ausführliche SRP-Handbuch ersetzen.

Für eine ausreichende **Dokumentation** der durchgeführten Prüfungshandlungen schließt sich an die (erste) Kurzanleitung jeweils ein Tabellenblatt „Datenquelle“ an, in welchem die verwendeten Ausgangsdateien, (berechneten) Datenfelder sowie etwaige Einschränkungen (Filter) der Datenreihe(n) tabellarisch eingegeben werden. Sämtliche Arbeitsblätter beider SRP-Vorlagen sind ausdruckvorbereitet und benennen dabei alle zur Reproduktion erforderlichen Informationen, den Dateinamen, das Ausdrucksdatum sowie etwaige Datenbeeinflussungen (bspw. Einzeldatenkorrektur in der SRP-Vorlage BWA oder Einschränkung der untersuchten Daten in der SRP-Vorlage Stochastik). Somit benötigen Stpfl. oder Berater*innen zur Überprüfung der Ergebnisse lediglich die verwendeten Daten z.B. als Export aus IDEA® oder durch Kopieren in eine leere Tabellenkalkulationsdatei. Zudem lassen sich auf jedem Auswertungsblatt in den Feldern „Bemerkung(en)“ freie Eintragungen ergänzen. Die Herausgabe der Vorlagendateien ist somit zur Erfüllung des „Anspruchs der Steuerpflichtigen auf einen effektiven Rechtsschutz durch die Gerichte (Art. 19 Abs. 4, Art. 20 Abs. 3 GG)“²⁴⁰ nicht erforderlich!

Neue Vorlagenversionen werden mit einer kurzen Übersicht über die Aktualisierungen ausgegeben, wodurch ein schnelleres Umgewöhnen ermöglicht werden soll.

Die **Datenanalyse in den SRP-Vorlagen** beginnt damit, die betreffenden Besteuerungsdaten mit ihren Spalten „Datum“ und „Betrag“ aus IDEA® oder z.B. einem Tabellenkalkulationsprogramm in die dafür vorgesehenen Felder der Arbeitsblätter „Vergleichsdaten“ (SRP-Vorlage BWA) bzw. „Daten“ (SRP-Vorlage Stochastik) hineinzukopieren:

-  eine oder zwei²⁴¹ Datenreihe(n) zur Untersuchung der Entwicklung(en und Abhängigkeit) in die SRP-Vorlage BWA: z.B. ein Kostenkonto bzw. -bereich mit auffallenden Jahres-salden, der Warenfluss aus -einkauf und -umsatz (zuzüglich Entnahmen und Beständen), operative Kosten(-stellen) versus Erfolgsgröße(n), Vertragssoll und -ist oder die Datenbestände vor und nach einer Datenschnittstelle bzw.
-  eine Datenreihe zur Analyse der Häufigkeitsverteilung(en) in die SRP-Vorlage Stochastik: z.B. eine Vielzahl an Kostenpositionen zum zügigen Auffinden von prüfungsrelevanten Ballungen, Ausdünnungen etc. oder Tagesbareinnahmen für die Prüfung auf Manipulationsspuren.

²⁴⁰ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 49.

²⁴¹ Durch die beispielhaften Überschriften in der leeren SRP-Vorlage BWA wird darauf hingewiesen, dass bei üblichen Anwendungen die höherwertige Datenreihe in den rechten Bereich eingefügt werden sollte.

Vorsicht: Beide SRP-Vorlagen haben nur ein begrenztes Fassungsvermögen der Eingabebereiche (s. die jeweiligen Haupthinweise überschriftennah auf den Tabellenblättern „Vergleichsdaten“ bzw. „Daten“). Sollen größere Datenbestände untersucht werden, gibt es die Möglichkeit, über die Arbeitsblätter „eigene Notizen 1[bzw. 2]“ mittels Pivot-Tabelle und Verlinkung zu arbeiten (s. den nachfolgenden Abschnitt „fortgeschrittene Anwendung der SRP-Vorlagen“ zu diesem Pkt.).

Zur **Kontrolle der korrekten Datenübernahme** wird in der SRP-Vorlage BWA nach dem Einfügen die **Datensatzanzahl der jeweiligen Datenreihe** angegeben. Auch die jährlichen Betragssummen auf dem Tabellenblatt „Jahreszahlen“ gewährleisten eine schnelle Überprüfung, dass die richtigen Daten vollständig in die Vorlage hineinkopiert wurden. Bei der SRP-Vorlage Stochastik werden auf dem Tabellenblatt „Daten“ im Bereich „statistische Kenngrößen“ die erforderlichen Parameter wie „Anzahl“, „Summe“ etc. für eine einfache Gegenkontrolle angezeigt.

Datum	Betrag
853	853
02.01.2002	105,16

I.R. von **Plausibilitätsprüfungen** mit der SRP-Vorlage BWA auf die kaufmännische Handlungslogik von Betriebsdaten wie bspw. hinsichtlich der Abhängigkeit von Wareneinkauf und -erlösen werden üblicherweise – entsprechend der betriebswirtschaftlichen Sicht – **Nettozahlen** verwendet. Je leistungsnäher dabei die Datumswerte z.B. infolge der Verwendung eines elektronischen Warenwirtschaftssystems sind, desto aussagekräftiger sind die Schlüssigkeitsauswertungen.

Die in der SRP-Vorlage Stochastik auf **Häufigkeitsmuster** zu untersuchenden Betriebsdaten sollten einen gleichartigen **Konsolidierungsgrad** („Granularität“) aufweisen: unverdichtete Einzelvorgänge, Tagessummen etc. Zur speziellen Untersuchung von (Bar-) Einnahmen auf **Manipulationsspuren** müssen diejenigen Daten herangezogen werden, die – nach dem jeweiligen Kenntnisstand zu den Aufzeichnungsprozessen – Ausgangspunkt menschlicher Eingriffe gewesen sein können: i.d.R. Bareinnahmen der konsolidierten Kasse(n) als **täglich²⁴² aufsummierte Bruttowerte²⁴³**.

Weitere Eingaben in die SRP-Vorlagen:

- ☒ Die Eingaben zu „Steuerpflichte/r“ und „Steuernummer“ werden vom Tabellenblatt „Datenquelle“ automatisch auf alle Auswertungsblätter übernommen.

²⁴² Die Überschrift „Datum“ weist auf dem Tabellenblatt „Daten“ nach dem Hineinkopieren darauf hin, falls mehrere Datensätze denselben Datumswert haben, da dies bei der manipulationsorientierten Untersuchung von Tagesbareinnahmen auf eine fehlerhafte Datenauswahl schließen lässt.

²⁴³ Sollen aufgrund eines konkreteren Verdachts einzelne Kassen oder die Einnahmen nach Umsatzsteuersatz getrennt untersucht werden, kann jeweils eine eigene Vorlage verwendet werden.

- ☒ In der SRP-Vorlage BWA lassen sich die Überschriften der beiden Datenreihen auf dem Tabellenblatt „Vergleichsdaten“ sowie des Verhältnisfaktors (z.B. „RGAS“) auf dem Tabellenblatt „Jahreszahlen“ überschreiben und werden in alle Auswertungen übertragen.
- ☒ Beginn und Ende des Prüfungszeitraums müssen in der SRP-Vorlage BWA auf dem Tabellenblatt „Vergleichsdaten“ eingegeben werden, damit die Auswertungen unproblematisch auf Teilzeiträume beschränkt werden können. Beide Vorlagen unterstützen nach entsprechender Eingabe auf dem Tabellenblatt „Vergleichsdaten“ (*SRP-Vorlage BWA*) bzw. „Daten“ (*SRP-Vorlage Stochastik*) das abweichende Wirtschaftsjahr.
- ☒ Neben den Eingabespalten „Datum“ und „Betrag“ zu beiden Vergleichsdatenreihen verfügt die SRP-Vorlage BWA jeweils über eine Korrektur- und eine Bemerkungsspalte, um punktuelle Berichtigungen der zeitlichen Zuordnungen oder Wertinhalte für einzelne Datensätze (Sachverhalte) interaktiv vorzunehmen (s. [Pkt. 4.2.9.](#)).
- ☒ Über das freie Feld oberhalb der Spalte „Datum“ zur ersten Datenreihe in der SRP-Vorlage BWA kann diese insgesamt um bis zu 30 Tage vor- oder rückdatiert werden, um mögliche Auswirkungen auf die Analyseergebnisse zu simulieren (s. [Pkt. 4.2.9.](#)).
- ☒ Auf dem Tabellenblatt „Jahreszahlen“ der SRP-Vorlage BWA können i.R. von Plausibilitätsprüfungen des Warenflusses die jährlichen Warenbestände, Entnahmen und **Gewinne** sowie Richtsatzwerte eingetragen werden. Die Entnahmewerte lassen sich *einheitlich* auf den Tabellenblättern „Quartalsrelationen“ und „Monatszeitreihen“ über Ankreuzfelder gleichmäßig auf die (vorhandenen) Quartale und Monate aufteilen; eine vergleichbare Funktion ist für die Bestandsveränderungen nicht²⁴⁴ vorgesehen, weil eine solche Verteilung nicht den üblichen operativen Verhältnissen entspricht. Punktuelle Umverteilungen können auf dem Tabellenblatt „Vergleichsdaten“ vorgenommen werden (s. [Pkt. 4.2.9.](#)).
- ☒ Für die retrograde Rekonstruktion des Warenbestands kann auf dem Tabellenblatt „Bestandsentwicklung“ ein (geschätzter) Anfangsbestand eingetragen werden.

Die **SRP-Vorlage Stochastik** bietet zur vertiefenden Untersuchung von Musterstörungen als risikoorientierte Prüffelder bzw. zur Absicherung entdeckter Manipulationsspuren eine **zweistufige Eingrenzung der untersuchten Daten** an: Die Einschränkungen auf Jahre, einen Teilzeitraum und/oder Wochentage des Tabellenblatts „Daten“ wirken grundsätzlich auf alle Arbeitsblätter, während die jeweiligen Beschränkungen einzelner Analysen wie z.B. auf dem Tabellenblatt „digZA“ nur für dort gelten. Einzelne Datensätze lassen sich zudem nachvollziehbar auf dem Tabellenblatt „Daten“ über die Ankreuzfelder links neben den importierten Datumswerten ausschließen. Alle Einschränkungen der untersuchten Daten

²⁴⁴ Dagegen ist i.R. der Quantilsschätzung auf dem Tabellenblatt „80 %-Quantile“ eine gleichmäßige Verteilung der Warenbestandsveränderungen auf die Monate per Ankreuzfeld möglich, falls keine konkreteren Verhältnisse (in angemessener Zeit) ermittelbar sind.

werden auf den einzelnen Auswertungsblättern angezeigt und dienen i.R. der Manipulationsanalyse einer verbesserten und abgesicherten Aufdeckung²⁴⁵.

Der **Prüfungsduktus** der vielen regelbasierten Einzelmethode in der SRP-Vorlage BWA folgt mit der Reiterabfolge dem vorgegebenen Ablauf „vom groben Überblick zu den Details“, um – anhand beobachteter Regeln – Systemauffälligkeiten und punktuelle Prüffelder zügig wie zielgenau aufzufinden bzw. die Plausibilität i.S. des § 158 Abs. 2 Nr. 1 AO erst nach Ausschluss (möglicher) Störeinflüsse zu bewerten. Die SRP-Vorlage Stochastik ist zweigeteilt: Für die allgemeine Musteranalytik orientieren sich die Tabellenblätter „Schichtung“ und „Schichtungsvertiefung“ ebenfalls an dieser Prüfungslogik, um aufgefundene „Musterstörungen“ als risikobezogene Prüffelder effektiv und effizient einzugrenzen. Weil eine derart stringente Reihenfolge der musterbasierten Manipulationsprüfung nicht möglich ist, enthält die SRP-Vorlage Stochastik auf dem Tabellenblatt „Kurzanleitung 2“ ein Prüfungsschema, welches sich auf die Aufdeckung zweier praxisrelevanter Manipulationsszenarien („erfundene Einnahmenwerte“ und „stärkere Verkürzung höherer Einnahmen“) konzentriert. Dadurch soll zur Ergebnisabsicherung ein weitreichender Ausschluss betrieblicher Sonderursachen gewährleistet werden.

Der beweisverstärkende Vorteil von Indiziengesamtheiten im Zusammenhang mit Plausibilitätsprüfungen gem. § 158 Abs. 2 Nr. 1 AO durch den Einsatz systematisch-mehrperspektivischer Prüfungsnetze wie SRP setzt die **zusammenführende Dokumentation** aller Einzelergebnisse von der Formalprüfung über die einzelnen Auswertungen bis zum interaktiven Umgang mit Einlassungen voraus.

Über die **fortgeschrittene Anwendung der SRP-Vorlagen** lassen sich einerseits deutlich größere Datenbestände als im jeweilig vordefinierten Datenbereich (*Hauptthinweise überschrittene auf den Tabellenblättern „Vergleichsdaten“ der SRP-Vorlage BWA und „Daten“ der SRP-Vorlage Stochastik*) untersuchen und andererseits Unterscheidungskriterien einbinden, indem mittels Pivot-Tabellen auf dem Tabellenblatt „eigene Notizen 1[und/oder 2]“ gearbeitet wird: Zuerst werden die zu untersuchenden Daten bei Bedarf inklusive zusätzlicher Differenzierungsmerkmale (in der Grafik: „Kontonummer“ und „-bezeichnung“) im Arbeitsblatt „eigene Notizen 1[und/oder 2]“ abgelegt. Anschließend werden die Einzeldaten als Pivot-Tabelle geordnet, wobei die Zuordnung des Datums als „Zeilenbeschriftung“ (per „drag and drop“) für eine tägliche Aufsummierung sorgt – es resultiert eine begrenzte Anzahl an Tageswerten. Nach Zuweisung der Wertinformation (hier: „SoHa“) mit automatischer Zu-

²⁴⁵ Der schnelle Wandel im Bereich der elektronischen Betriebsverwaltung lässt bei Manipulationen Veränderungen im Vorgehen (z.B. Umstieg von manueller Verkürzung zu Zappern) erwarten, was eine zeitliche Beschränkung möglicher Manipulationsspuren nach sich zieht. Außerdem können die Dateneingrenzungen eine Differenzierung zwischen Manipulation und Unterschlagung unterstützen.

sammenfassung als „Summe von SoHa“ werden die zusätzlichen Merkmale „Kontonummer“ und „Kontobezeichnung“ als Berichtsfilter (bzw. Datenschnitte²⁴⁶) gewählt.

Unformatierte, offene Seite für eigene Notizen und Ergebnisbeschreibungen					Kontonummer	(Alle)
					Kontobezeichnung	Aushilfslöhne
Kontonummer	Kontobezeichnung	Datum	SoHa		Zeilenbeschriftungen	Summe von SoHa
4100	Löhne und Gehälter	06.01.2003	269,05		04.01.2003	148,8
4100	Löhne und Gehälter	29.01.2003	-360		06.01.2003	489,1
4100	Löhne und Gehälter	04.02.2003	317,16		04.02.2003	758,8
4100	Löhne und Gehälter	04.02.2003	749,53		04.03.2003	500,8
4100	Löhne und Gehälter	05.02.2003	946,8		01.04.2003	258
4100	Löhne und Gehälter	20.02.2003	1.112,18		03.04.2003	527,7
4100	Löhne und Gehälter	28.02.2003	-360		05.05.2003	758,8
4100	Löhne und Gehälter	04.03.2003	39,03		05.06.2003	920
4100	Löhne und Gehälter	10.03.2003	946,8		04.07.2003	662
4100	Löhne und Gehälter	10.03.2003	749,53		05.08.2003	266,6
4100	Löhne und Gehälter	24.03.2003	1.112,18		06.08.2003	662
4100	Löhne und Gehälter	31.03.2003	-360		08.09.2003	269,7
4100	Löhne und Gehälter	02.04.2003	-732,87		12.09.2003	161,2
4100	Löhne und Gehälter	03.04.2003	231,1		06.10.2003	1139,2
4100	Löhne und Gehälter	03.04.2003	317,16		04.11.2003	780,9
4100	Löhne und Gehälter	07.04.2003	946,8		08.12.2003	870
4100	Löhne und Gehälter	07.04.2003	749,53		31.12.2003	-906,95
4100	Löhne und Gehälter	16.04.2003	1.112,18		05.01.2004	1295,9
4100	Löhne und Gehälter	28.04.2003	-360		06.02.2004	1278,8
4100	Löhne und Gehälter	05.05.2003	273,25		05.03.2004	928,8
4100	Löhne und Gehälter	06.05.2003	946,8		06.04.2004	928,8
4100	Löhne und Gehälter	06.05.2003	749,53		04.05.2004	928,8
4100	Löhne und Gehälter	19.05.2003	-180		03.06.2004	948,8
4100	Löhne und Gehälter	22.05.2003	1.112,18			

Nun werden die Datumswerte und Tagessummen der Pivot-Tabelle mittels Verweisformel²⁴⁷ auf das jeweilige Datenblatt der SRP-Vorlagen verlinkt (**Achtung: Die Spaltensummen der Pivot-Tabelle müssen vorher über die Funktion „PivotTable-Optionen.../Summen & Filter/Gesamtsummen“ ausgeblendet werden**). Die Zuweisung der zu untersuchenden Daten wird dank dieser Arbeitsweise über die Berichtsfilter (bzw. Datenschnitte) gesteuert und kann jederzeit umgestellt werden. Sollen also im abgebildeten Beispiel die Zeitreihenanalysen der Aushilfslöhne (eventuell im Vergleich zu den Umsätzen) auf die Festangestellten ausgeweitet werden, muss lediglich die Einstellung des Berichtsfilters über die Auswahlfläche ergänzt werden. Danach sind alle BWA-Auswertungen umgehend an die neue Prüfung der Entwicklungen (und Abhängigkeit) angepasst.²⁴⁸

4.1. Formalprüfung: Ordnungsmäßige (prüfbare) Aufzeichnungen

Als Anleitung für (systematisch-visualisierende) Prüfungsmethodik steht das umfangreiche Thema „(zeitgemäße) Formalprüfung“ – trotz seiner grundsätzlichen Bedeutung – nicht im

²⁴⁶ Während die älteren Berichtsfilter voneinander unabhängig sind, erkennen Datenschnitte automatisch Zusammenhänge: So zeigt der Berichtsfilter „Kontonummer“ im Beispielsfall keine Einschränkung an, obwohl über die Kontobezeichnung nur die Aushilfslöhne ausgewählt wurden. Bei Datenschnitten würde die Auswahl unter „Kontobezeichnung“ auf die Anzeige zur Kontonummer durchwirken.

²⁴⁷ In der SRP-Vorlage BWA reicht die einfache Verlinkung „=eigene Notizen 1[bzw. 2]!Feldadresse“, während in der SRP-Vorlage Stochastik eine Absicherung gegen leere Felder erfolgen muss: „=WENN('eigene Notizen 1[bzw. 2]!Feldadresse=""&"";'eigene Notizen 1[bzw. 2]!Feldadresse)“. Im Anschluss werden die ersten beiden Verlinkungen im vordefinierten Datenbereich der SRP-Vorlagen bis zur maximalen Anzahl an Tagen (bei 3 Prüfungsjahren: max. 1.096) nach unten kopiert [s.a. die Arbeitshilfe „Erweiterte SRP Vorlagennutzung für zusätzliche Unterscheidungsmerkmale“ auf dem KTZ-Wiki-Server (Intranet) im Bereich „KTZWiki > PRUEFUNGSD Web > Projekt_IDEA > Automationsgestützte quantitative Prüfungsmethoden (...)“].

²⁴⁸ Die Eilenbrock-„VISU Tools“ unterstützen die fortgeschrittene SRP-Vorlagen-Nutzung mit ihren Kapazitäts- und Flexibilitätsvorteilen durch ein intuitives Dialogverfahren zur Datenvorbereitung und -übernahme.

Vordergrund des SRP-Handbuchs; dafür sei hier auf Werke mit umfassenden, detaillierten Ausführungen verwiesen, wie insbesondere:

- 📖 Achilles, „Kassenführung – Aufzeichnung – Aufbewahrung – Vorlagepflichten“, Erich Schmidt Verlag, angekündigt: Dez. 2023.
- 📖 Achilles/Pump, „Lexikon der Kassenführung“, Verlag der DATEV eG, 2018.
- 📖 Brinkmann, „Schätzungen im Steuerrecht: Fälle – Methoden – Vermeidung – Abwehr (Steuerrecht und Steuerberatung, Band 49)“, Erich Schmidt Verlag, 2023 (6. Auflage).
- 📖 Danielmeyer, „Die Digitalisierung der Betriebsprüfung (Steuerrecht und Steuerberatung Band 58)“, Erich Schmidt Verlag, 2022.
- 📖 Teutemacher/Krullmann, „Handbuch zur Kassenführung: Praxishandbuch für die rechtssichere Umsetzung“, NWB-Verlag, 2023 (4. Auflage).
- 📖 Teutemacher, „Praxis-GoBD: Handbuch zur praktischen Umsetzung der Grundsätze zur ordnungsmäßigen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD)“, NWB-Verlag, 2020.

An dieser Stelle sollen lediglich einige grundsätzliche Erläuterungen zur **teleologischen²⁴⁹ bzw. funktionalen²⁵⁰ Auslegung** der abstrakten Ordnungsvorgaben erfolgen, denn durch die verbreitete Umstellung der Betriebsverwaltungen auf elektronische Arbeitsweisen²⁵¹ haben sich die Steuerausfallrisiken – und damit die notwendigen Maßnahmen für eine wirksame Ordnungsmäßigkeitsprüfung – fundamental verändert.



Ziel der Verknüpfung von „Beweiskraft der Buchführung (und Aufzeichnungen)“ nach § 158 Abs. 1 AO mit strengen Ordnungsregeln der (steuerlichen) Aufzeichnung und Aufbewahrung

²⁴⁹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 12.

²⁵⁰ FG Hamburg v. 26.8.2016 – 6 V 81/16, 1. Leitsatz.

²⁵¹ Graf, „Anforderungen an die Vorlage von Einzelaufzeichnungen – Ende der ‚Schonzeit‘ für Daten aus Vordaten“, BBK 18/2013, 875, 876 u. 877: „Die Buchführung hat sich von der Durchschreibebuchführung hin zu einer umfassenden Datenverarbeitung entwickelt, die elektronische Form ist nunmehr die übliche. [...] Diese Entwicklung geschah auf Betreiben und im Interesse der Unternehmen, weil es betriebswirtschaftlich notwendig war und ist, diese Einzeldaten zu erzeugen, z.B. für ein aussagekräftiges Controlling oder zur Arbeitsprozesssteuerung. [...] Dabei steigt die Anzahl elektronisch überwachter Prozesse stetig.“

gem. § 140 ff. AO ist es, das **Verifikationsprinzip**²⁵² umzusetzen. Danach ist die Prüfbarkeit von Besteuerungsgrundlagen durch einen „sachverständigen Dritten innerhalb angemessener Zeit“²⁵³ Grundvoraussetzung für das Vertrauen in die Steuererklärung, weil ein „Beweisverderber“ oder „Beweisvereitler“ aus seinem Verhalten keinen Vorteil ziehen darf (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO).²⁵⁴ Entsprechend müssen gravierende Formalmängel, die den Prüfungszugang entscheidend behindern²⁵⁵, sanktioniert werden und führen gem. § 162 Abs. 2 S. 1 u. 2 AO zur Amtspflicht der Schätzung.²⁵⁶

Bereits mit der Reichsabgabenordnung²⁵⁷ wurden hierzu als **abstrakte Ordnungsvorgaben** die Kriterien „fortlaufend, vollständig und richtig“²⁵⁸ sowie „[d]er ursprüngliche Inhalt einer Eintragung soll nicht mittels Durchstreichens oder auf andere Weise unleserlich gemacht, es soll nicht radiert, auch sollen solche Veränderungen nicht vorgenommen werden, deren Beschaffenheit es ungewiss lässt, ob sie bei der ursprünglichen Eintragung oder erst später vorgenommen sind [Gewissheit der Originalität]“²⁵⁹ definiert, welche die Nachvollziehbarkeit für einen „sachverständigen Dritten innerhalb angemessener Zeit“²⁶⁰ sicherstellen sollen. Die Merkmale zielen dabei nicht nur auf das retrograde Auffinden eines einzelnen Vorgangs mit den dazugehörigen Unterlagen ab, sondern gleichzeitig auf die Aussagefähigkeit des Gesamtbilds, weil es „weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild einer Buchfüh-

²⁵² BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

²⁵³ § 145 Abs. 1 S. 1 AO.

²⁵⁴ BFH v. 15.2.1989, BStBl. II 1989, 462: „[...] Zur Vermeidung eines solchen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

²⁵⁵ Aufgrund der erheblich schwierigeren Prüfungsbedingungen infolge der Digitalisierung mit umfangreichen, nicht selbsterklärenden Daten sieht der Gesetzgeber bereits dann den Prüfungszugang entscheidend behindert, wenn „die elektronischen Daten nicht nach der [gesetzlichen] Vorgabe der einheitlichen digitalen Schnittstellen [...] zur Verfügung gestellt werden“ (§ 158 Abs. 2 Nr. 2 AO).

²⁵⁶ Die „Drei-Stufen-Theorie“ aus dem BFH-Zeitreihenurteil [BStBl. II 2015, 743; ergänzend Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, 3553] widerspricht insofern mit der Abgrenzung von Stufe 2 („Buchführung formell nicht ordnungsgemäß, [...] aber materielle Unrichtigkeiten der Einnahmenerfassung nicht konkret nachgewiesen“, Rz. 63) auf 3 („Buchführung nicht nur formell, sondern auch materiell unrichtig“, Rz. 66) dem Verifikationsprinzip, weil gravierende Ordnungsmängel gerade das Nachweisen von materiellen Mängeln verhindern.

²⁵⁷ § 162 RAO.

²⁵⁸ § 162 S. 2 RAO; übernommen in: § 146 Abs. 1 S. 1 AO: „Die Buchungen und die sonst erforderlichen Aufzeichnungen sind einzeln, vollständig, richtig, zeitgerecht und geordnet vorzunehmen.“

²⁵⁹ § 162 S. 5 RAO; übernommen in: § 146 Abs. 4 AO: „Eine Buchung oder eine Aufzeichnung darf nicht in einer Weise verändert werden, dass der ursprüngliche Inhalt nicht mehr feststellbar ist. Auch solche Veränderungen dürfen nicht vorgenommen werden, deren Beschaffenheit es ungewiss lässt, ob sie ursprünglich oder erst später gemacht worden sind.“

²⁶⁰ § 145 Abs. 1 S. 1 AO.

rung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann²⁶¹. Die abstrakte Fassung des Ordnungsrahmens verfolgt die Wahlfreiheit der Aufzeichnungs- und Aufbewahrungsweise, was gerade im Hinblick auf die sich immer schneller wandelnde Technik realistisch ist. In der Konsequenz finden sich etliche Klarstellungen in der Rechtsprechung²⁶² oder in Erlassen²⁶³.

Zusammenfassen lassen sich die Ordnungsvorgaben in der **abstrakten Kernregel** „übersichtliche, zeitgerechte und unveränderliche Einzelaufzeichnung“. Auch in der jüngeren BFH-Rechtsprechung ist hierbei verstärkt eine „funktionale“²⁶⁴ Sichtweise zu erkennen.²⁶⁵ Schon nach dem Gleichheitsgrundsatz in Art. 3 GG muss dies auch für die Einnahmen-Überschuss-Rechnung gem. § 4 Abs. 3 EStG gelten²⁶⁶, damit das Prinzip der Verifizierbarkeit gewahrt bleibt²⁶⁷. An Bargeschäfte werden mit der Pflicht zur taggenauen Erfassung von Kasseneinnahmen und -ausgaben²⁶⁸ sowie neuerdings dem Schutz eines elektronischen Aufzeichnungssystems und der digitalen Aufzeichnungen durch eine zertifizierte technische Sicherheitseinrichtung²⁶⁹ deutlich strengere Ordnungsvorgaben angelegt, deren Nichteinhaltung nach ständiger Rechtsprechung die gesamte Buchführung infiziert.²⁷⁰

²⁶¹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5.

²⁶² Vgl. z.B.: BFH v. 2.10.1968, BStBl. II 1969, 157: „Länger als etwa zehn Tage darf ein Geschäftsvorfall grundsätzlich grundbuchmäßig nicht unerfaßt bleiben.“

²⁶³ Insbes. BMF-Schreiben v. 28.11.2019, „Grundsätze zur ordnungsgemäßen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD n.F.)“, BStBl. I 2019, 1269.

²⁶⁴ FG Hamburg v. 26.8.2016 – 6 V 81/16, 1. Leitsatz.

²⁶⁵ Vgl. z.B.: BFH v. 16.12.2014, BStBl. II 2015, 519, 2. Leitsatz: „Verwendet ein Einzelhändler, der in seinem Betrieb im allgemeinen Waren von geringem Wert an ihm der Person nach nicht bekannte Kunden über den Ladentisch gegen Barzahlung verkauft, eine PC-Kasse, die detaillierte Informationen zu den einzelnen Verkäufen aufzeichnet und eine dauerhafte Speicherung ermöglicht, so sind die damit bewirkten Einzelaufzeichnungen auch zumutbar“ sowie BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27: „Der erkennende Senat vertritt hierzu die Auffassung, dass das Fehlen einer lückenlosen Dokumentation zur Kassenprogrammierung in seinen Auswirkungen auf die Beurteilung der formellen Ordnungsmäßigkeit der Buchführung und der Eröffnung der Schätzungsbefugnis dem Fehlen von Tagesendsummenbons bei einer Registrierkasse bzw. dem Fehlen täglicher Protokolle über das Auszählen einer offenen Ladenkasse gleichsteht. In allen drei Fällen lässt der formelle Mangel zwar keinen sicheren Schluss auf die Verkürzung von Einnahmen zu. Gleichwohl gibt es systembedingt keine Gewähr mehr für die Vollständigkeit der Erfassung der Bareinnahmen, ohne dass eine nachträgliche Ergänzung der Dokumentation bzw. eine anderweitige Heilung des Mangels möglich wäre.“

²⁶⁶ Anwendungserlass zu § 146 AO, Pkt. 2.1.7.

²⁶⁷ Vgl. z.B.: BFH v. 7.2.2008 – X B 189/07: „Allerdings sind auch Steuerpflichtige, die den Gewinn nach § 4 Abs. 3 EStG ermitteln, verpflichtet, die ihrer Gewinnermittlung zugrunde liegenden Belege aufzubewahren. [...] Zutreffend hat das FG darauf abgestellt, dass gerade in Fällen, in denen Steuerpflichtigen eine Einzelaufzeichnungspflicht nicht zugemutet werden kann, die Einnahmeermittlung – z.B. bei Einsatz von Registrierkassen durch Erstellung und Aufbewahrung der Kassenendsummenbons – nachvollziehbar dokumentiert und überprüfbar sein muss.“

²⁶⁸ § 146 Abs. 1 AO.

²⁶⁹ § 146a Abs. 1 S. 2-3 AO; s.a.: OFD Karlsruhe v. 30.4.2021: „Informationen zum Thema ‚Ordnungsmäßigkeit der Kassenbuchführung‘“, S 0315 – St 42.

²⁷⁰ Vgl. z.B.: BFH v. 14.12.2011 – XI R 5/10 oder FG Sachsen v. 4.4.2008 – 5 V 1035/07.

Der **Anwendungsbereich** der Ordnungsvorschriften wird in den GoBD angegeben mit: „Somit sind alle Unternehmensbereiche betroffen, in denen betriebliche Abläufe durch DV-gestützte Verfahren abgebildet werden und ein Datenverarbeitungssystem [...] für die Erfüllung der [...] außersteuerlichen oder steuerlichen Buchführungs-, Aufzeichnungs- oder Aufbewahrungspflichten verwendet wird. [...] Die Erfassung der Geschäftsvorfälle in elektronischen Grund(buch)aufzeichnungen [...] und die Verbuchung im Journal [...] kann organisatorisch und zeitlich auseinander fallen [...]. In diesen Fällen gelten die Ordnungsvorschriften bereits mit der ersten Erfassung [...] und müssen über alle nachfolgenden Prozesse erhalten bleiben“²⁷¹. Diese Reichweite der Ordnungsmaßgaben findet sich regelmäßig in der Rechtsprechung²⁷² wieder und ist erforderlich, weil sich die Buchführung beim Vorliegen detaillierter Grundaufzeichnungen i.d.R. auf diese bezieht und alleine nicht den Formalanforderungen entspricht. Besonders trifft dieser erweiterte Anwendungsbereich der Ordnungsvorschriften auf Kassenslösungen zur Erstaufzeichnung zu, weil üblicherweise keine weiteren Nachweise der Geschäftsvorfälle vorliegen und das hohe Manipulationsrisiko bekannt ist²⁷³. Demzufolge wird in der Literatur die Beratungsverantwortung für die Kassensführung hervorgehoben.²⁷⁴

Die eine zentrale Frage der Formalprüfung ist: „Lässt sich die **Vollständigkeit der Aufzeichnungen nachvollziehen?**“ Nach den Ordnungsvorschriften der Reichsabgabenordnung sollte dies durch fest gebundene Bücher mit durchnummerierten Seiten (weitgehend) sichergestellt werden.²⁷⁵

Als Folge der Verbreitung elektronischer Betriebsverwaltungen stellen sich neue Praxisfragen hinsichtlich des Umfangs aufzeichnungs- und aufbewahrungspflichtiger Unterlagen.²⁷⁶ Wegen der digitalen Gefahren für die Vollständigkeit von Besteuerungsdaten (s. [Pkt. 2.1.](#)) ist durch das Gesetz zum Schutz vor Manipulationen an digitalen Grundaufzeichnungen vom

²⁷¹ BMF-Schreiben v. 28.11.2019 (GoBD n.F.), BStBl. I 2019, 1269, Rz. 9 u. 84.

²⁷² Vgl. z.B.: BFH. V. 26.2.2004, BStBl. II 2004, 599: „Die Schichtzettel sind Einnahmeursprungsaufzeichnungen; sie enthalten Angaben, aus denen sich die Höhe der Umsätze und damit auch der Betriebseinnahmen unmittelbar ergibt. Der Kläger war insoweit aufbewahrungspflichtig; die Vernichtung der Schichtzettel bedeutete einen Verstoß gegen diese Pflicht.“

²⁷³ Vgl. z.B.: BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27: „Elektronische Kassensysteme sind durch Umprogrammierung in nahezu beliebiger Weise manipulierbar; von derartigen Manipulationsmöglichkeiten machen Teile der betrieblichen Praxis nach dem Erkenntnisstand des Senats durchaus Gebrauch (zu einem solchen Fall z.B. Beschluss des FG Rheinland-Pfalz vom 7. Januar 2015 5 V 2068/14; vgl. zum Ganzen auch Tz. 54 der Bemerkungen des Bundesrechnungshofs 2003 zur Haushalts- und Wirtschaftsführung, BTDrucks 15/2020, 197 f.).“

²⁷⁴ Pump/Heidl, „Hilfestellung bei der Kassensführung durch den steuerlichen Berater“, StBp 2014 S 162, 163: „Unverantwortliche Sorglosigkeit kann die Berater dazu verleiten, sich nicht um die Kassensführung ihrer Mandanten zu kümmern. Das wird damit begründet, dass dafür kein Prüfungsauftrag bestehe. Wer am Fortbestand seiner Mandanten interessiert ist, muss sie vor fehlerhafter Kassensführung schützen.“

²⁷⁵ § 162 S. 4 RAO.

²⁷⁶ Wargowske/Werner, „Zur Fortentwicklung des Datenzugriffs gemäß § 147 Abs. 6 AO“, beck.digitax 2/2020, 66, u. 3/2020, 133.

22.12.2016 die Pflicht zur Einzelaufzeichnung explizit in § 146 Abs. 1 AO aufgenommen worden. Ein mögliches Kriterium zur Überprüfung der *vollständigen* Einzelaufzeichnung digitaler Betriebsdaten ist eine programmseitig mitgeschriebene Sequenznummer: „Die vollständige und lückenlose Erfassung und Wiedergabe aller Geschäftsvorfälle ist bei DV-Systemen durch ein Zusammenspiel von technischen [...] und organisatorischen Kontrollen sicherzustellen (z.B. [...] automatisierte Vergabe von Datensatznummern“²⁷⁷. In Wirtschaft und Beratung werden i.R. von „Tax Compliance Management Systemen (TCMS)“ technische Lösungen wie die Vorgangsverkettung der „Blockchain-Technologie“²⁷⁸ erprobt, die auch Bestandteil der österreichischen Registrierkassensicherheitsverordnung zur Manipulationsbekämpfung von elektronischen Aufzeichnungssystemen ist.²⁷⁹ Der Bundesverband der Deutschen Industrie e.V. (BDI) hat vorgeschlagen, die Dauer von Betriebsprüfungen am Bestehen derartiger Vorkehrungen auszurichten²⁸⁰, was in Form eines zu evaluierenden Tests Eingang in die Gesetzgebung erhalten hat²⁸¹; allerdings gilt für diese ebenso, dass sie von „einem sachverständigen Dritten innerhalb angemessener Zeit“²⁸² überprüfbar sein müssen.

Auch die regelmäßig als Grund für eine Ausnahme von der Einzelaufzeichnungspflicht angeführten Abwägungen des BFH zur Zumutbarkeit bei Barverkäufen geringen Werts an eine Vielzahl von (überwiegend unbekannten) Kunden²⁸³ werden inzwischen im Kontext der – heutzutage üblichen – digitalen Aufzeichnungs- und Aufbewahrungsweise bewertet: „Die Aufzeichnung jedes einzelnen Geschäftsvorfalles ist nur dann nicht zumutbar, wenn es technisch, betriebswirtschaftlich und praktisch unmöglich ist, die einzelnen Geschäftsvorfälle aufzuzeichnen [...]. Das Vorliegen dieser Voraussetzungen ist durch den Steuerpflichtigen nachzuweisen. Beim Verkauf von Waren an eine Vielzahl von nicht bekannten Personen gegen Barzahlung gilt die Einzelaufzeichnungspflicht nach § 146 Absatz 1 Satz 1 AO aus Zumutbarkeitsgründen nicht, wenn kein elektronisches Aufzeichnungssystem, sondern eine offene Ladenkasse verwendet wird [...]. Wird hingegen ein elektronisches Aufzeichnungssystem verwendet, gilt die Einzelaufzeichnungspflicht nach § 146 Absatz 1 Satz 1 AO unabhängig davon, ob das elektronische Aufzeichnungssystem und die digitalen Aufzeichnungen

²⁷⁷ BMF-Schreiben v. 28.11.2019 (GoBD n.F.), BStBl. I 2019, 1269, Rz. 40.

²⁷⁸ Vgl. z.B. „Betriebswirtschaftliche Prozesse in die Blockchain? – Machbarkeitsstudie von Telekom und DATEV zeigt Potenzial der Technologie auf“, nachzulesen unter: <https://www.datev.de/web/de/m/presse/pressemitteilungen/meldungen-2018/betriebswirtschaftliche-prozesse-in-die-blockchain/>.

²⁷⁹ Huber, „Reform der Kassenführung in Österreich – Teil 2 – Technische Umsetzung, Status und Erkenntnisse“, BBK 23/2016, 1148, 1150.

²⁸⁰ Bundesverband der Deutschen Industrie e.V. (BDI), „Steuerliche Betriebsprüfungen – Optionen für zeitgemäße Reformen“, 13, nachzulesen unter: <https://bdi.eu/publikation/news/steuerliche-betriebspruefungen/>.

²⁸¹ Art. 97 § 38 EGAO mit der Systemprüfung von Steuerkontrollsystemen und der Möglichkeit, Beschränkungen der nächsten Außenprüfung zuzusagen.

²⁸² § 145 Abs. 1 S. 1 AO.

²⁸³ Urt. v. 12.5.1966, BStBl. III 1966, 371.

nach § 146a Absatz 3 AO i. V. m. der KassenSichV mit einer zertifizierten technischen Sicherheitseinrichtung zu schützen sind. Die Zumutbarkeitsüberlegungen, die der Ausnahmeregelung nach § 146 Absatz 1 Satz 3 AO zugrunde liegen, sind grundsätzlich auch auf Dienstleistungen übertragbar (vgl. AEAO zu § 146, Nr. 2.2.6).²⁸⁴ Der BFH bestätigt dieses Abstellen auf die faktischen Verhältnisse des Einzelfalls.²⁸⁵

Die andere zentrale Frage der Formalprüfung ist: „Lässt sich die Unverändertheit der Aufzeichnungen nachvollziehen oder zumindest der ursprüngliche Inhalt erkennen (**Gewissheit der Originalität**)?“ Hierzu wurde in der Reichsabgabenordnung die Verwendung des unauslöschlichen Schreibmaterials „Tinte“ normiert.²⁸⁶

Derweil der Grundzustand elektronisch erfasster Informationen „veränderbar“²⁸⁷ ist, erfordert die Gewissheit der Originalität aktive Maßnahmen: „Das zum Einsatz kommende DV-Verfahren muss die Gewähr dafür bieten, dass alle Information (Programme und Datenbestände), die einmal in den Verarbeitungsprozess eingeführt werden (Beleg, Grundaufzeichnung, Buchung), nicht mehr unterdrückt oder ohne Kenntlichmachung überschrieben, gelöscht, geändert oder verfälscht werden können. [...] Die Unveränderbarkeit der Daten, Datensätze, elektronischen Dokumente und elektronischen Unterlagen [...] kann sowohl hardwaremäßig (z.B. unveränderbare und fälschungssichere Datenträger) als auch softwaremäßig (z.B. Sicherungen, Sperren, Festschreibung, Löschmerker, automatische Protokollierung, Historisierungen, Versionierungen) als auch organisatorisch (z.B. mittels Zugriffsberechtigungskonzepten) gewährleistet werden. Die Ablage von Daten und elektronischen Dokumenten in einem Dateisystem erfüllt die Anforderungen der Unveränderbarkeit regelmäßig nicht, soweit nicht zusätzliche Maßnahmen ergriffen werden, die eine Unveränderbarkeit gewährleisten.“²⁸⁸ Die jüngere Gesetzesaktualisierung zum Schutz elektronischer Aufzeich-

²⁸⁴ BMF-Schreiben v. 28.11.2019 (GoBD n.F.), BStBl. I 2019, 1269, Rz. 39.

²⁸⁵ BFH v. 16.12.2014, BStBl. II 2015, 519, Rz. 20: „Da die GoB indes nur eine Einzelaufzeichnung der Kassenvorgänge im Rahmen des nach Art und Umfang des Geschäftes Zumutbaren verlangen, hat der BFH die Einzelaufzeichnungspflicht für Einzelhandelsgeschäfte – in Betrieben, in denen Waren von geringerem Wert an eine unbestimmte Vielzahl nicht bekannter und auch nicht feststellbarer Personen verkauft werden – dahingehend eingeschränkt, dass die baren Betriebseinnahmen in der Regel nicht einzeln aufgezeichnet zu werden brauchen (BFH-Urteil in BFHE 86, 118, BStBl. III 1966, 371). Ausschlaggebend für den BFH war insoweit, dass es technisch, betriebswirtschaftlich und praktisch unmöglich war, an die Aufzeichnung der einzelnen zahlreichen baren Kassenvorgänge in Einzelhandelsgeschäften gleiche Anforderungen wie bei anderen Handelsgeschäften zu stellen, nämlich zur Identifizierung und zur Bestimmung des Inhalts des Geschäfts Namen und Anschrift des Kunden und den Gegenstand des Kaufvertrages festzuhalten.“

²⁸⁶ § 162 S. 6 RAO.

²⁸⁷ Vgl. z.B.: Tipke/Kruse, AO/FGO, 167. Lieferung 8/2021, § 146 AO, Rz. 59: „Moderne EDV-Buchführungssysteme [...] müssen programmmäßige Sicherungen und Sperren enthalten, die verhindern, dass einmal eingegeben Daten dem Zugriff zur Wiederaufbereitung (Änderung) preisgegeben sind. [...] Dies muss bereits vom Zeitpunkt der erstmaligen Speicherung an, nicht erst nach durchgeführter Verarbeitung gewährleistet sein“.

²⁸⁸ BMF-Schreiben v. 28.11.2019 (GoBD n.F.), BStBl. I 2019, 1269, Rz. 108 u. 110.

nungssysteme und der digitalen Aufzeichnungen durch eine zertifizierte technische Sicherheitseinrichtung²⁸⁹ – unter Einbindung des Bundesamts für Sicherheit in der Informationstechnik (BSI)²⁹⁰ – dient im Wesentlichen der Sicherstellung der Originalität von Besteuerungsdaten.

In der Literatur wird inzwischen zunehmend die „Ordnungsmäßigkeitsberatung“ als Möglichkeit gesehen, sich innerhalb der Konkurrenz durch ausreichende Kenntnisse der heutzutage üblichen komplexen Aufzeichnungs- und Aufbewahrungsprozesse hervorzuheben,²⁹¹ und es werden neue Qualitätsstandards wie „Data Governance“ ins Gespräch gebracht.²⁹²

Der praktische Umgang mit den Mindestanforderungen an die vereinfachte (papierene) Dokumentation von elektronischen Registrierkassen mittels Tagesendsummenbons (Z-Bons) laut Erlass²⁹³ hat jahrelang einen Teil der Prüfungen mitbestimmt. So leidet die Gewissheit der Originalität im abgebildeten Z-Bon-Beispiel²⁹⁴ an der **fehlenden Ausgabe von Korrekturen**. Da es beim Verkauf einer Vielzahl von Leistungen bzw. Waren äußerst unwahrscheinlich ist, dass – trotz rechtzeitiger Aufzeichnung – keine Erfassungsfehler geschehen sind, muss der nachteilige Schluss²⁹⁵ gezogen werden, dass Einnahmenverkürzungen verschleiert werden sollen. Dieser Eindruck wird noch dadurch bekräftigt, dass die drei Stornoebenen nicht die Tagessumme „0“ anzeigen, sondern offensichtlich in den Berichtseinstellungen deaktiviert wurden.

Z - Bericht			
04.01.2008 - 00:58			
Nr. 1			
Bereich: Alle Kassen			
Von: 03.01.2008 - 14:29			
Bis: 04.01.2008 - 00:40			
Brutto total	335	880.60	
Netto total	335	740.00	
Retoure			
Master Storno			
Storno			
Umsatz Brutto MwSt 19.00%		880.60	
Umsatz Netto MwSt 19.00%		740.00	
MwSt 19.00%		140.60	
Prozentanteil		100.00	
Durchschn. MWST Satz		19.00	
Zahlungsarten:			
in Bar	335	880.60	
Vorgangsdurchschnitt	96	9.17	
Kundendurchschnitt	96	9.17	
Umsatz		880.60	

²⁸⁹ § 146a Abs. 1 S. 2-3 AO; eingefügt durch das Gesetz zum Schutz vor Manipulationen an digitalen Grundaufzeichnungen vom 22.12.2016; anzuwenden spätestens ab 1.10.2020 gem. BMF-Schreiben v. 6.11.2019, BStBl. I 2019, 1010.

²⁹⁰ BMF-Schreiben v. 26.9.2017, „Verordnung zur Bestimmung der technischen Anforderungen an elektronische Aufzeichnungs- und Sicherungssysteme im Geschäftsverkehr (Kassensicherungsverordnung – Kassen-SichV)“, BStBl. I 2019, 1310, § 5 u. 7.

²⁹¹ Vgl. z.B.: Liekenbrock, „Digital Tax Compliance – Praxishinweise zur technischen Umsetzung von Tax Compliance Management Systemen“, Ubg 2018, 43.

²⁹² Vgl. z.B.: Liekenbrock/Danielmeyer, „Data Governance als Reaktion auf die zunehmend datengetriebene Betriebsprüfung?“, RET 2/2021, 32.

²⁹³ BFM-Schreiben v. 9.1.1996, „Elektronische Registrierkassen, Kassenstreifen“, BStBl. I 1996, 34.

²⁹⁴ Bildquelle: Wähnert, „Die Beweislast im Besteuerungsverfahren – Brennpunkt ‚Einnahmenvollständigkeit‘“, StC 11/2011, 16, 17.

²⁹⁵ BFH v. 15.2.1989, BStBl. II 1989, 462: „[...] der ‚Beweisverderber‘ oder ‚Beweisvereitler‘ aus seinem Verhalten keinen Vorteil ziehen darf (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO) [...] Zur Vermeidung eines solchen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

Eine „funktionale“²⁹⁶ **Übertragung der Hauptordnungskriterien** „nachvollziehbare Vollständigkeit“ und „Gewissheit der Originalität“ auf ein zeitgemäßeres Aufzeichnungssystem mit dauerhafter Speichermöglichkeit setzt zunächst grundsätzliches Verständnis der technischen Prozesse voraus: „Eine Systemprüfung hat das gesamte Verfahren, welches die Verarbeitung steuerlich relevanter Daten betrifft [...] zum Gegenstand. So soll festgestellt werden, ob und inwieweit eine zuverlässige, sichere, nachvollziehbare, ordnungsgemäße und vollständige Verarbeitung der steuerlich relevanten Daten durch den Steuerpflichtigen erfolgt.“²⁹⁷ Danach kann die **Lücke in den fortlaufenden Vorgangsnummern „RENr“** im Fall des nachfolgenden Auszugs aus den Daten eines Kassensystems folgende Ursachen haben:

1. Der Sachverhalt mit der RENr „7522“ ist aufgrund der Einstellung des Kassensystems im Datenexport nicht mit ausgegeben worden.
2. Durch einen Eingriff in die aufgezeichneten Verkaufsvorgänge ist jener mit der RENr „7522“ gelöscht worden.
3. Es sind mehrere Vorgänge nachträglich gelöscht worden. Bei der Neu Nummerierung der Datensätze ist jedoch die RENr „7522“ vergessen worden.

DATUM	ZEIT	RENr	TOTAL	BEDNR	KASSE
22.12.2007	23:31	7503	2,50	5	1
22.12.2007	23:40	7504	15,00	5	1
22.12.2007	23:42	7505	7,50	5	1
22.12.2007	23:42	7506	5,00	5	1
22.12.2007	23:44	7507	2,50	5	1
22.12.2007	23:46	7508	3,00	5	1
22.12.2007	23:49	7509	7,00	5	1
22.12.2007	23:50	7510	2,20	5	1
22.12.2007	23:51	7511	2,50	5	1
22.12.2007	23:52	7512	4,50	5	1
22.12.2007	23:55	7513	9,00	5	1
23.12.2007	00:00	7514	10,00	5	1
23.12.2007	00:02	7515	3,50	5	1
23.12.2007	00:05	7516	3,00	5	1
23.12.2007	00:07	7517	9,70	5	1
23.12.2007	00:08	7518	6,90	5	1
23.12.2007	00:08	7519	2,50	5	1
23.12.2007	00:08	7520	2,50	5	1
23.12.2007	00:09	7521	2,50	5	1
23.12.2007	00:14	7523	5,00	5	1
23.12.2007	00:16	7524	2,20	5	1
23.12.2007	00:16	7525	7,50	5	1

Weil der tatsächliche Grund für die fehlerhafte Datenausgabe ohne die vollständigen Systeminformationen inklusive aller Einstel-

lungsprotokolle nicht ermittelbar ist, sieht der BFH in dem „Fehlen einer lückenlosen Dokumentation zur Kassenprogrammierung in seinen Auswirkungen auf die Beurteilung der formellen Ordnungsmäßigkeit der Buchführung und der Eröffnung der Schätzungsbefugnis [einen Ordnungsmangel, der] dem Fehlen von Tagesendsummenbons bei einer Registrierkasse bzw. dem Fehlen täglicher Protokolle über das Auszählen einer offenen Ladenkasse gleichsteht.“²⁹⁸

Obwohl die beiden Voraussetzungen für die „Beweiskraft der Buchführung (und Aufzeichnungen)“ „Ordnungsmäßigkeit“ und „kein Beanstandungsanlass“ laut Gesetzestext von ei-

²⁹⁶ FG Hamburg v. 26.8.2016 – 6 V 81/16, 1. Leitsatz.

²⁹⁷ Wargowske/Werner, „Zur Fortentwicklung des Datenzugriffs gemäß § 147 Abs. 6 AO“, beck.digitax 2/2020, 66, 68 (u. 3/2020, 133).

²⁹⁸ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

inander unabhängig und gleichwertig sind²⁹⁹, erfährt nur die Formalprüfung in der aktuellen Rechtsprechung eine dem Gesetzeszweck folgende Auslegung³⁰⁰, während die für die Finanzverwaltung abgesenkte Beweisschwelle nach wie vor auf der Grundlage eines höchst-richterlich geforderten Übermaßes³⁰¹ (s. [Pkt. 4.6.](#)) umgesetzt und somit faktisch konterkariert wird. Der Zusammenhang zwischen formellen und inhaltlichen Mängeln von Besteuerungsaufzeichnungen wird ungeachtet dessen gesehen.³⁰² Folgerichtig führt „[d]ie Vorlage einer bereits formell ordnungswidrigen Buchführung [...] von vornherein zu einer erheblichen Absenkung des Beweismaßes für eine Schätzung“³⁰³.

Bei Beanstandungen der sachlichen Richtigkeit von Besteuerungsgrundlagen – insbesondere im Zusammenhang mit Gesamtbildprüfungen³⁰⁴ – ist es demzufolge wichtig, vorge-

²⁹⁹ § 158 AO „Beweiskraft der Buchführung: (1) Die Buchführung und die Aufzeichnungen des Steuerpflichtigen, die den Vorschriften der §§ 140 bis 148 entsprechen, sind der Besteuerung zugrunde zu legen. (2) Absatz 1 gilt nicht, 1. soweit nach den Umständen des Einzelfalls Anlass besteht, die sachliche Richtigkeit zu beanstanden [...]“

³⁰⁰ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27: „Der erkennende Senat vertritt hierzu die Auffassung, dass das Fehlen einer lückenlosen Dokumentation zur Kassenprogrammierung in seinen Auswirkungen auf die Beurteilung der formellen Ordnungsmäßigkeit der Buchführung und der Eröffnung der Schätzungsbefugnis dem Fehlen von Tagesendsummenbons bei einer Registrierkasse bzw. dem Fehlen täglicher Protokolle über das Auszahlen einer offenen Ladenkasse gleichsteht. In allen drei Fällen lässt der formelle Mangel zwar keinen sicheren Schluss auf die Verkürzung von Einnahmen zu. Gleichwohl gibt es systembedingt keine Gewähr mehr für die Vollständigkeit der Erfassung der Bareinnahmen, ohne dass eine nachträgliche Ergänzung der Dokumentation bzw. eine anderweitige Heilung des Mangels möglich wäre.“

³⁰¹ BFH v. 24.6.1997, BStBl. II 1998, 51: „Nach § 158 AO 1977 sind der Besteuerung die Buchführung und die Aufzeichnungen des Steuerpflichtigen, die den Vorschriften der §§ 140 bis 148 AO 1977 entsprechen, zugrunde zu legen, soweit nach den Umständen des Einzelfalls kein Anlaß besteht, ihre sachliche Richtigkeit zu beanstanden. Nur wenn die Würdigung des Sachverhalts ergibt, daß eine formell ordnungsmäßige Buchführung mit an Sicherheit grenzender Wahrscheinlichkeit sachlich unrichtig ist, kann das Ergebnis der Buchführung ganz oder teilweise verworfen werden.“ Einschätzung einer wissenschaftlichen Autorin zu dieser unrealistischen Gesetzesauslegung: „Die herkömmlichen [Verprobungs]methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...]. Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten in den wiederhergestellten Daten. [...] Schlussfolgernd stellt sich die Frage, ob das geforderte Sicherheitsmaß [mit an Sicherheit grenzender Wahrscheinlichkeit] bei den herkömmlichen Methoden überhaupt erreicht werden kann“, Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324.

³⁰² Vgl. z.B.: Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 24: „Allerdings entspricht es der Erfahrung, dass eine formell ordnungswidrige Buchführung i.d.R. auch sachliche Mängel aufweist“ oder Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstrafverfahren“, NZWiSt 1/2014, 24, 25: „In der Praxis wird sich die Frage, ob ein Zeitreihenvergleich allein die Beweisvermutung des § 158 AO widerlegt, [...] dadurch relativieren, dass die [...] aufgedeckten Unregelmäßigkeiten regelmäßig mit Unregelmäßigkeiten in der Kassenbuchführung einhergehen.“

³⁰³ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 9.

³⁰⁴ Gesamtbildprüfungen (Verprobungen) können die vom BFH geforderte Aussagegewissheit „mit an Sicherheit grenzender Wahrscheinlichkeit“ i.d.R. nicht erreichen, weil sie komplexe Sachverhalte auf Grundlage von vielen, eventuell unvollständigen Informationen aus der Vergangenheit wiederherzustellen versuchen. Mehrperspektivisch-systematische Prüfungsnetze wie die SRP erreichen zwar über die Absicherung von Einzelergeb-

fundene **Ordnungsmängel ausführlich zu dokumentieren**, wie in diesem Beispiel:
„Das beklagte Finanzamt (FA) führte [...] beim Kläger eine Außenprüfung [...] durch, in deren Verlauf es zu der Einschätzung kam, die Kassenbuchführung des Klägers sei nicht ordnungsgemäß. Die Prüferin hat im Verlauf der Prüfung diesbezüglich folgende Feststellungen getroffen (...):

- die Kasseneinnahmen wurden zum Teil nicht täglich festgehalten, sondern in einem Z-Bon, gleich für mehrere Tage, zusammengefasst,
- die Z-Zähler Nr. fängt mehrfach wieder mit Nr. 1 an, und zwar in 2010 am: 08.03., 12.03., 10.07., 04.08. und 21.12., in 2011 am: 26.02. und in 2012 am: 15.08., wobei dies auf die Anschaffung der neuen Kasse zurückzuführen ist,
- die Erlöse aus dem Z-Bon Nr. 160 vom 01.09.2011 wurden nicht als Einnahme erfasst (...),
- der Z-Bon Nr. 172 im September 2011 fehlt (...),
- die Erlöse für Sonntag, den 22.01.2012, fehlen (...),
- in allen Streitjahren wurden zum Teil hohe Stornobuchungen vorgenommen (...),
- für den Zeitraum 13.01.2010 bis 25.01.2010 liegt lediglich der Z-Bon mit der Nr. 97 vor (... €),
- für den Zeitraum vom 08.06.2010 bis zum 06.07.2010 liegt nur ein Z-Bon vom 07.07.2010 mit der Nr. 84 vor, wobei der darin ausgewiesene Betrag (... €) in den monatlichen Exceltabellen aufgeteilt wurde in Höhe von ... € auf den 30.06. und in Höhe von ... € auf den 06.07. (...),
- für den Zeitraum vom 07.01.2011 bis zum 14.02.2011 liegt lediglich ein Z-Bon mit der Nr. 16 (Einnahme: ... €) vor,
- für den Zeitraum 15.02.2011 bis 25.02.2011 liegen keine Z-Bons vor (...), ohne dass am 26.02.2011 ein entsprechend höherer Betrag erfasst worden wäre (... €),
- für den 12.01.2010 (...), 12.08.2012 (...) und den 21.08.2012 (...) liegen nur handschriftlich gefertigte Aufzeichnungen über die Tageseinnahmen vor,
- für den Zeitraum 16.09.2011 bis 29.09.2011 wurden keine Tageserlöse erklärt, dann am 30.09.2011 ein Betrag in Höhe von ... € (Z-Bon 174; ...),
- für den Zeitraum 01.10.2011 - 05.10.2011 wurden ebenfalls keine Tageserlöse erfasst, ohne dass aber am 06.10.2011 ein entsprechend höherer Betrag erfasst worden wäre (... € vgl. Z-Bon 175; ...) und

nissen und Indiziengesamtheiten ein qualifizierteres Gesamtbild, sind aber dennoch – wie alle Prüfungsmaßnahmen – von der Datenqualität abhängig.

- in 2012 wurden Kasseneinnahmen nur für 302 Tage erklärt, so dass bei einem Abzug von 52 Ruhetagen für 12 Tage keine Einnahmen erklärt wurden (...).³⁰⁵

Außerdem sollten die sich aus Formalmängeln ergebenden Bedenken konkret hinsichtlich möglicher verschleierte Fehler materieller Art ausgeführt werden, weil zuletzt wiederholt betont worden ist, dass Ordnungsmäßigkeitsprüfungen nicht dem Selbstzweck dienen³⁰⁶, sondern die Prüfbarkeit durch einen „sachverständigen Dritten innerhalb angemessener Zeit“³⁰⁷ sicherstellen sollen.

4.2. Regelbasierte Ansätze (Zeitreihenanalysen)

Regelbasierte Prüfungsansätze sollen entweder die **Entwicklung** einer bzw. die **Übereinstimmung** zweier Datenreihen (*ein Beispiel hierzu findet sich mit der Schnittstellenprüfung unter [Pkt. 5.1.2.](#)*) chronologisch nachverfolgen **oder** die **Plausibilität** von (Unternehmens-)Daten anhand der (betriebswirtschaftlich) erwarteten Abhängigkeit wiedergeben. Erstmalige Anwendungen werden dem 18. Jahrhundert zugeordnet³⁰⁸. Mit diesem scheinbar³⁰⁹ (*s.a. die nachfolgende Gegenüberstellung von Zeitreihenvarianten im Bereich „Methodenwissen“ unter diesem Pkt.*) einfachen Modell lassen sich die Visualisierungsvorteile für die Datenanalyse (*s. [Pkt. 2.](#)*) erfahren: Grundsätzlich bietet die zeitliche (i.d.R. periodisch-zusammengefasste) Ordnung einen guten und schnellen Prüfungszugang auch bei großen Datenmengen – insbesondere wenn die Bedeutung der unterschiedlichen Summationsebenen beachtet wird (*s. das Beispiel im Abschnitt „Fuzzy-Theorien“ unter [Pkt. 4.2.2.](#)*).

³⁰⁵ FG Hamburg v. 5.3.2018 – 3 K 205/15, Rz. 14.

³⁰⁶ Vgl. z.B.: Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, 3550: „Wie es § 158 i. V. mit § 162 Abs. 2 Satz 2 zweite Alternative AO bereits von seinem Wortlaut her vorsieht, ist eine Schätzung dem Grunde nach bei einem Steuerpflichtigen, der eine Buchführung (freiwillig oder gesetzlich verpflichtend) führt, vorzunehmen, wenn die Vorschriften der §§ 140 bis 148 AO verletzt sind. Die ständige höchstgerichtliche Rechtsprechung schränkt diese allein auf formellen Buchführungsmängeln beruhende Schätzungsbeurteilung aber insoweit ein, als sie allein Anlass geben, die sachliche Richtigkeit der Buchführungsergebnisse anzuzweifeln“.

³⁰⁷ § 145 Abs. 1 S. 1 AO.

³⁰⁸ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 25 mit Verweis auf Playfair, „The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure and Debts of England during the Whole of the Eighteenth Century“, London 1786.

³⁰⁹ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. Die Analyse statistischer Daten beginnt häufig mit der Wahl einer geeigneten Darstellungsform des Erhebungsbefundes. Die korrekte Interpretation solcher Grafiken ist allerdings nicht trivial und erfordert spezielle kognitive Fähigkeiten.“

Gesamtbildverprobungen, die sich auf die (kaufmännische) Schlüssigkeit als nachvollziehbare Abhängigkeit gründen, gehen bis auf die Reichsfinanzhof-Rechtsprechung³¹⁰ zurück und sind nach dem Verifikationsprinzip³¹¹ notwendig, weil „es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann“³¹². Anders als die im Revisionsbereich der Bargeschäfte verbreitete Zeitreihenverfolgung der Verhältnisgröße „Aufschlagsatz“³¹³ setzt die SRP zentral auf einen doppelskalierten Zeitreihenvergleich der auf einen schlüssigen Zusammenhang zu prüfenden Datenreihen, denn „[b]ei dieser Methode tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander.“³¹⁴ Während die Entwicklung des Aufschlagsatzes im Unternehmensalltag vielen Einflüssen ausgesetzt ist und deshalb in der Literatur entsprechende Unwegbarkeiten aufgelistet werden³¹⁵, gibt die zentrale SRP-Zeitreihe die Handlungslogik im Betrieb erheblich eindeutiger wieder (s. *das folgende Beispiel im Bereich „Methodenkenntnis“ unter diesem Pkt.*). Somit erreichen Schlüssigkeitsurteile gem. § 158 Abs. 2 Nr. 1 AO eine generell höhere Gewissheit.

Voraussetzung für die Auswahl **geeigneter Methoden** zur jeweiligen Prüfungsfrage ist eine ausreichende **Methodenkenntnis**, welche – neben der systematisch-mehrperspektivischen Vernetzung – den Fokus des Methodik-Konzepts „SRP“ bildet (s.a. *den Abschnitt „Methodenwissen“ unter Pkt. 2.2.*). Laut BFH „sind zahlreiche Varianten eines Zeitreihenvergleichs denkbar“³¹⁶. Im Gegensatz zum FG Hamburg³¹⁷ haben sich die RiBFH aber nicht mit

³¹⁰ Vgl. z.B.: RFH v. 13.1.1937, RStBl 1937, 317.

³¹¹ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

³¹² Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5.

³¹³ Vgl. z.B.: FG Münster v. 19.8.2004 – 8 V 3055/04, FG Düsseldorf v. 15.2.2007 – 16 V 4691/06 oder FG Köln v. 27.1.2009 – 6 K 3954/07. Auch das BFH-Zeitreihenurteil (v. 25.3.2015, BStBl. II 2015, 743) bezieht sich mit der verwendeten Gleitschlittenzeitreihe auf eine Zeitreihenvariante des Aufschlagsatzes.

³¹⁴ FG Hamburg v. 31.10.2016 – 2 V 202/16, Rz. 48. SRP-kritische Stimmen wie bspw. *Steinhauff* [„Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388] erwähnen zwar tlw. die methodologischen Differenzierungen und Abwägungen der SRP, ziehen sich in ihrer Beurteilung aber wieder auf die nicht differenzierende Rechtsprechung zu Zeitreihenanalysen des Aufschlagsatzes zurück.

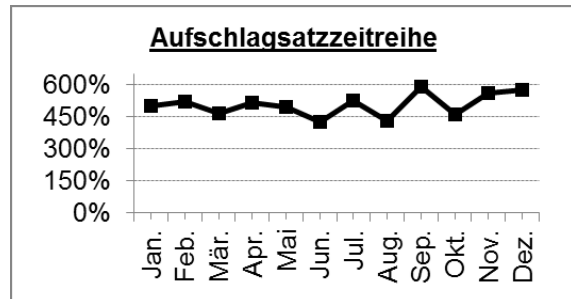
³¹⁵ Vgl. z.B.: *Wiggen*, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168. Dazu: *Sell*, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 27: „Dies ist keinesfalls ein Argument gegen die Methoden, sondern zeigt lediglich die bei jeder Verprobungsmethode obligatorische Verpflichtung zur sorgfältigen Anwendung ebendieser.“

³¹⁶ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 37.

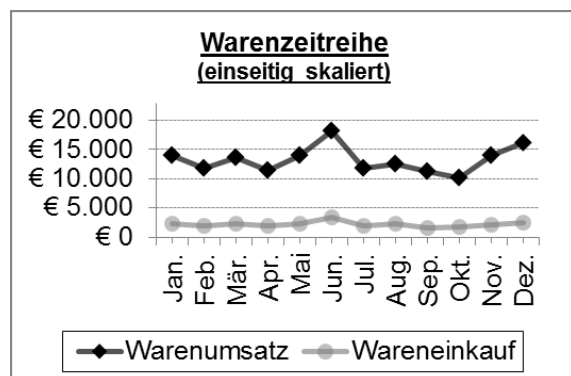
³¹⁷ FG Hamburg v. 31.10.2016 – 2 V 202/16, Rz. 48.

den Unterschieden³¹⁸ auseinandergesetzt, sondern Aufschlagsatzzeitreihen stellvertretend für alle Formen beurteilt³¹⁹. Weil sich dieses Vorgehen bei Kenntnis der eklatanten Modellunterschiede verbieten hätte,³²⁰ sollen hier drei verbreitete Varianten der Zeitreihenanalyse zu denselben Daten exemplarisch gegenübergestellt werden:

1. Das erste Diagramm zeigt die monatliche Entwicklung des Aufschlagsatzes aus Wareneinkauf³²¹ und -umsatz für eine kleine Kneipe, die nach § 4 Abs. 3 EStG aufzeichnet. Dabei ist eine Schwankungsbreite des Erfolgsfaktors von unter 450 % bis fast 600 % zu beobachten. Für die Prüfungsfrage der Gesamtschlüssigkeit ist beurteilungserheblich, ob die entdeckte „Regelstörung“³²² auf Zuordnungsunschärfen des Wareneinkaufs bzw. dessen Bezahlung, auf betriebliches Verhalten wie bspw. Lagerankäufe größerer Getränkemengen mit längerer Haltbarkeit oder auf sachlich falsche Besteuerungsdaten zurückgehen.



2. In der nächsten Auswertung werden die – der ersten Analyse zu Grunde liegenden – beiden Datenreihen mit dem monatlichen Wareneinkauf und -umsatz als Einzelgrafen an einer³²³ Y-Achse aufgetragen. So werden die Absatzveränderungen mit starken Anstiegen im Sommer und Winter ersichtlich. Die Veränderungen im Warennachkauf erscheinen deutlich moderater. Hinsichtlich der Schlüssigkeit der Daten ist erkennbar, dass der verstärkte Verkauf in den beiden Hochsaison-Zeiträumen auch eine



³¹⁸ Vgl. z.B. auch: Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 326: „Zusammengefasst liegen eine Reihe geeigneter Verfahren im Rahmen der Zeitreihenanalyse vor. [...] Jede Methode bietet sowohl Vor- als auch Nachteile.“

³¹⁹ Auch die im BFH-Zeitreihenurteil (v. 25.3.2015, BStBl. II 2015, 743) verhandelte Gleitschlittenzeitreihe bezieht sich auf eine chronologische Nachverfolgung des Aufschlagsatzes.

³²⁰ Bestätigend: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 28: „Der BFH vertritt hier eine distanziert-kritische Haltung [...], die im Grad ihrer Verallgemeinerung der Methode [Zeitreihenanalyse] nicht angemessen erscheint“.

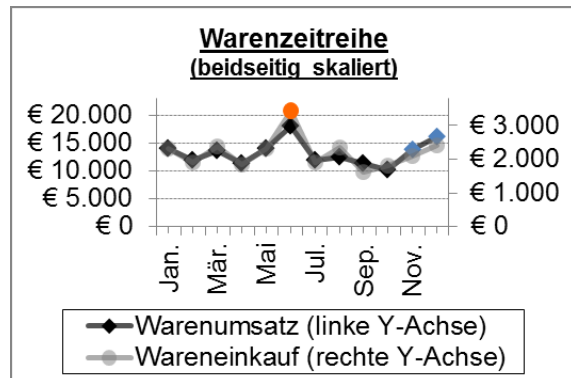
³²¹ Für die Gegenüberstellung dieser drei verbreiteten Zeitreihenvarianten sind die Wareneinkaufsdaten bewusst nicht zum „Wareneinsatz“ korrigiert worden, um die extrem unterschiedliche „Robustheit“ der einzelnen Verfahren gegenüber (alltäglichen) Störfaktoren noch deutlicher herauszustellen.

³²² Die Prüfungsregel bei dieser Datenverbildlichung ergibt sich mit der erwarteten (weitgehenden) Konstanz der Verhältnisgröße „Aufschlagsatz“.

³²³ Der einseitig skalierte Zeitreihenvergleich ist in diversen Datenanalyseprogrammen (zurzeit noch) die Standardauswertungsform.

Erhöhung des Einkaufs nach sich zieht; fraglich bleibt, ob diese Reaktionen in passender Höhe erfolgt sind oder auf mögliche Sachmängel hinweisen.

3. In der dritten Grafik werden die beiden Datenreihen mit *jeweils einer eigenen*³²⁴ Y-Achse abgebildet, um die unternehmerische Handlungslogik eines reagierenden Wareneinkaufs in Abhängigkeit vom aufgetroffenen Absatz mittels Datenvisualisierung nachvollziehbar wiederzugeben. Als Folge der Größenanpassung des Waren-



einkaufs an den Warenumsatz lässt sich bei dieser Zeitreihenanalyse unzweifelhaft ablesen, dass grundsätzlich ein enger Zusammenhang der beiden Betriebsfaktoren besteht. Im Zeitraum „August bis September“ liegt offenbar eine Zuordnungsunschärfe vor, indem der erhöhte Wareneinkauf aus dem August einen geringeren im nächsten Monat zur Folge hatte. Neben dem insgesamt plausiblen Verhältnis zwischen Marktfaktor „Absatz“ und unternehmerischem Verhalten „Wareneinkauf“ lassen sich der Auswertung zudem zwei klare Vertiefungsbereiche entnehmen: Erstens ist der **Wareneinkauf im Juni** (relativ) zu hoch, was als punktuelles Prüffeld untersucht werden sollte. Zweitens übertrifft der **Umsatzgraf ab November** den angepassten für den Einkauf dauerhaft. Hinter dieser Entwicklung ist eine Verkaufspreiserhöhung zu vermuten, die – nach Verifizierung – in der weiteren Gesamtbildprüfung zu berücksichtigen ist.

Da drei Zeitreihenvarianten zur Plausibilitätsprüfung *derselben* Warenflussdaten zu (stark) abweichenden Ergebnissen führen, muss die jeweilige Eignung mit Hilfe einer objektiven Gegenkontrolle hinterfragt werden (**Methodenvalidierung**). Die statistische Kenngröße für gleichgerichtete Abhängigkeiten zweier Datenreihen ist der Korrelationskoeffizient³²⁵. Zum Beispielsfall resultiert mit dem Wert „+0,90“ eine sehr gute Gleichrichtung. In der Konsequenz sind die ersten beiden Zeitreihentypen „Aufschlagsatzzeitreihe“ und „einseitig skalierten Zeitreihenvergleich“ für die Plausibilitätsprüfung „erwartete Abhängigkeit zweier Datenreihen“ nur begrenzt geeignet, während sich der **doppelskaliert größenangepasste Zeitreihenvergleich** davon positiv abgrenzt – *als eigenständige Methode*.

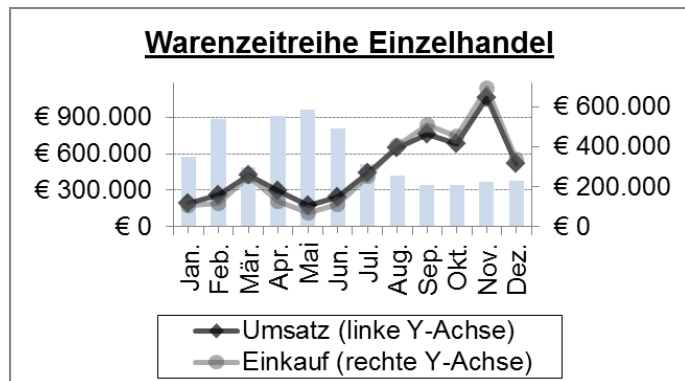
Deshalb wird i.R. der SRP der Einsatz von Zeitreihenanalysen streng nach dem **Verwendungszweck** differenziert:

³²⁴ Tabellenkalkulationsprogramme wie z.B. MS Excel® stellen die größenanpassende Doppelskalierung als Standardoption für Zeitreihenanalysen zur Verfügung.

³²⁵ Als Maß für den Zusammenhang zweier Datenreihen nimmt der Korrelationskoeffizient Größen von „-1“ für die 100 %-ige Gegenläufigkeit der Grafen über „0“ bei nicht vorhandener Abhängigkeit bis zu „+1“ für eine vollständig positive Gleichrichtung ein.

- ⊙ Soll die Entwicklung einer Betriebsgröße chronologisch nachverfolgt werden, geschieht dies durch eine Einzelgrößenzeitreihe wie beim Aufschlagsatz.
- ⊙ Sollen zwei Datenreihen auf Übereinstimmung geprüft werden wie bspw. bei Schnittstellen (s. [Pkt. 5.1.2.](#)), wird dies in einer einzelskalierten Gegenüberstellung vorgenommen, damit der Vergleich nicht durch eine automatische Größenanpassung verzerrt wird.
- ⊙ Soll die (chronologische) Abhängigkeit zweier Unternehmensfaktoren z.B. beim Warenfluss als logischer Handlungszusammenhang untersucht werden, erfolgt dies über einen doppelskaliert größenangepassten Zeitreihenvergleich.

Wegen der sehr guten Eignung des **doppelskaliert größenangepassten Zeitreihenvergleichs** für die Frage der Abhängigkeit zweier Datenreihen ist diese Zeitreihenform seit Beginn der SRP das zentrale Modell zur Plausibilitätsprüfung gem. § 158 **Abs. 2 Nr. 1** AO.³²⁶ Damit wird z.B. für die Warenflussdaten untersucht, ob der Nachkauf (möglichst) eng an dem vom Markt vorgegebenen Umsatz orientiert wird. Der Schlüssigkeitsansatz ist deutlich zwingender als die chronologische Nachverfolgung des Aufschlagsatzes, weil dieser in seiner Konstanz erfahrungsgemäß (s.a. *das Beispiel im Abschnitt „Methodenkenntnis“ oberhalb unter diesem Pkt.*) stärker von üblichen Betriebseinflüssen wie z.B. wechselnden Mengenverhältnissen mit unterschiedlichen Aufschlagsätzen oder konjunkturellen Preisveränderungen bestimmter Produkte bzw. Leistungen gestört wird³²⁷. Der nebenstehenden Abb. lässt sich die Robustheit des doppelskalierten Zeitreihenvergleichs abermals klar entnehmen: Selbst für die sehr lange haltbaren Waren eines Einzelhandels mit Elektroartikeln lässt sich die Logik des monatlichen Wareneinkaufs zum Absatz an den beiden korrelierenden Grafen einwandfrei ablesen. Demgegenüber können die außerordentlichen Schwankungen des Aufschlagsatzes³²⁸ leicht zur Fehldeutung betriebswirtschaftlicher Unschlüssigkeit verleiten. Abermals bestätigt



³²⁶ Insofern muss Kritik an der SRP in der Literatur, die sich auf Aufschlagsatzzeitreihen bezieht, als gegenstandslos betrachtet werden (vgl. z.B.: Beyer, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, 2933-2934: „Hierzu werden die einzelnen Quartals-RAS in einem Diagramm dargestellt. Schwanken die RAS in den Quartalen, kann dies mit Änderungen der Umsatzträger des Unternehmens zu tun haben“ oder Bleschick, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353: „Zunächst unternimmt er eine sog. betriebswirtschaftliche Auswertung: Hierzu ermittelt er im Wesentlichen mithilfe des Zeitreihenvergleichs einen Rohgewinnaufschlagsatz.“).

³²⁷ Ebenso: Wiggen, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168.

³²⁸ Die hellblauen Säulen zum Aufschlagsatz haben keinen direkten Wertbezug zu den Y-Achsen, sondern sind zur Methodengegenüberstellung über einen Faktor an die Relationen der linken Y-Achse angelehnt worden.

der Korrelationskoeffizient mit fast +1 den plausiblen Eindruck der zentralen SRP-Zeitreihe und unterstreicht den eklatanten Unterschied der beiden Zeitreihenvarianten.

Die bisherigen kritischen Aussagen zur eingeschränkten Anwendbarkeit der Zeitreihenanalytik in der Betriebsprüfung³²⁹ können deshalb nicht überzeugen. *Wolenski* hat an einem nachvollziehbaren Beispiel aufgezeigt, dass *geeignete* Zeitreihenformen für diverse Prüfungsfragen sinnvolle Analyseansätze darstellen³³⁰, welche tiefreichende und aussagekräftige Eindrücke zur betriebswirtschaftlichen Logik einer Vielzahl an Besteuerungsdaten liefern – einerseits um Prüffelder betriebsbezogen festzustellen und andererseits um dem Gebot der Gesamtbildprüfung³³¹ zu folgen.³³²

Durch die mehrperspektivische Absicherung (s.a. [Pkt. 2.3.](#)) von auffälligen Zeitreihenergebnissen sowie die Verwendung interaktiver Möglichkeiten, um neue Informationen z.B. aus Einlassungen umgehend in die Gesamtbildprüfung einzupflegen (s. [Pkt. 4.2.9.](#)), wird die Aussagegewissheit i.S. des § 158 Abs. 2 Nr. 1 AO allgemeinverständlich weiter verbessert. Darüber hinaus wird in der Literatur der aus Praxiserfahrungen bekannte Zusammenhang auffälliger Zeitreihenergebnisse mit Ordnungsmängeln der Aufzeichnungen angeführt³³³, wodurch sich Indiziengesamtheiten³³⁴ ergeben.

³²⁹ Als regelmäßige Zitiergrundlage für Kritiker*innen dient das BFH-Urt. v. 25.3.2015, BStBl. II 2015, 743, Rz. 56: „Bei bestimmten Betriebstypen oder in bestimmten betrieblichen Situationen scheidet der Zeitreihenvergleich schon dem Grunde nach als geeignete Schätzungs- und nach Ansicht des X. Senats wohl auch als Verprobungs-)methode aus.“

³³⁰ *Wolenski*, „Das Zeitreihenurteil des X. Senats des BFH vom 25. 3. 2015“, Stbg 6/2016, 268, 271.

³³¹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5: „Die gesetzliche Regelung des § 158 gründet sich insofern auf die Einsicht, dass es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild einer Buchführung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann.“

³³² So auch: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1 u. 4: „Bei einer steuerlichen Außenprüfung können mathematisch-statistische (quantitative) Prüfungsmethoden eingesetzt werden, um die Besteuerungsgrundlagen des Steuerpflichtigen zu verproben und auf Plausibilität zu überprüfen. Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich. [...] Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgabenordnung dazu führen, dass eine Schätzung durchgeführt werden muss.“

³³³ *Madauß*, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstrafverfahren“, NZWiSt 1/2014, 24, 25: „In der Praxis wird sich die Frage, ob ein Zeitreihenvergleich allein die Beweisvermutung des § 158 AO widerlegt, [...] dadurch relativieren, dass die [...] aufgedeckten Unregelmäßigkeiten regelmäßig mit Unregelmäßigkeiten in der Kassenbuchführung einhergehen.“

³³⁴ BGH v. 27.4.2010 – 1 StR 454/09, NStZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweismwürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreich[en].“

Mit der Quantilsschätzung (s. [Pkt. 4.8.](#)) bietet die SRP-Vorlage BWA zusätzlich eine [Option](#) an, *bei eingetretener Amtspflicht der Schätzung gem. § 162 AO* die systematisch erarbeiteten Erkenntnisse zu den Besteuerungsgrundlagen auch für die Festlegung einer möglichst sachgerechten Schätzungshöhe zu nutzen. Zwar haben Aufschlagsätze innerhalb der Schlüssigkeitsbeurteilung mittels SRP keine wesentliche Bedeutung, im Fall der Verwerfung der Besteuerungsgrundlagen nach § 158 Abs. 2 Nr. 1 AO erscheinen sie jedoch hilfreich, um unter angemessenem Arbeitsaufwand³³⁵ eine **Schätzung** durchzuführen.³³⁶

4.2.1. Betriebsprofil

Die Schlüssigkeitsbeurteilung von Betriebszahlen, -entwicklungen und -abhängigkeiten ist ohne Kenntnis der **wesentlichen Rahmenbedingungen des geprüften Unternehmens** i.d.R. nicht seriös möglich. Am Beispiel kalkulatorischer Kennzahlen wie dem Rohgewinnaufschlagsatz ist das mit Richtsatzspannen von z.T. weit über 100 %³³⁷, die auf regionale und einzelfallbezogene Verhältnisse zurückgehen, direkt erkennbar. Selbst bei den (scheinbar) klarsten Verprobungsmethoden „Vermögenszuwachs- oder Geldverkehrsrechnung“ lassen sich wichtige Größen wie bspw. der private Lebensaufwand nur anhand einzelfallspezifischer Informationen möglichst sachgerecht rekonstruieren.

Obwohl die SRP-Zeitreihenanalytik zentral auf das für Abhängigkeitsfragen besonders geeignete und gegenüber möglichen Störeinflüssen robuste Modell „doppelskaliert größenangepasster Zeitreihenvergleich“ (s. [Pkt. 4.2.](#)) setzt, sind zur richtigen Einordnung der grafischen Auswertungen grundlegende Informationen über den geprüften Betrieb wichtig: So können z.B. partiell einseitige Preisveränderungen oder punktuell gesamtbildrelevante Einzelsachverhalte wie betriebsbezogen teure Bestandseinkäufe zu Verzerrungen in der operativen Periodenzuordnung von Zeitreihenanalysen führen. Deshalb werden zu Beginn einer SRP – neben der Formalprüfung – die wesentlichen Rahmenbedingungen des geprüften

³³⁵ BFH v. 21.2.1990, BFH/NV 1990, 683, Rz. 31.: „Die Geldverkehrsrechnung ist eine arbeitsaufwendige Schätzungsmethode. Die vom Kläger erklärten Buchführungsergebnisse beruhen auf einer formell ordnungswidrigen Buchführung. Daher ist die Anwendung weniger arbeitsaufwendiger Schätzungsmethoden statthaft.“

³³⁶ Entsprechend: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 24-25 u. 27 ff.: „Besteht dem Grunde nach die Notwendigkeit, eine Schätzung durchzuführen, so können quantitative Prüfungsmethoden zur Auswahl der geeigneten Schätzungsmethode herangezogen werden. Die folgenden Schätzungsmethoden dienen dazu, eine möglichst genaue Schätzung bei einer Außenprüfung durchzuführen. [...] Quantilsschätzung“.

³³⁷ S. in der „Richtsatzsammlung für das Kalenderjahr 2022“ z.B.: Eisdielen mit einem Rohgewinnaufschlagsatz von 257 % bis 488 % bei einem Häufigkeitsschwerpunkt von 355 %; nachlesbar unter: https://www.bundesfinanzministerium.de/Content/DE/Downloads/BMF_Schreiben/Weitere_Steuerthemen/Betriebspruefung/Richtsatzsammlung/2023-08-10-download-richtsatzsammlung-2022.pdf?__blob=publicationFile&v=3.

Einzelfalls als Betriebsprofil (betriebswirtschaftliches Profil³³⁸) mittels Aktenstudium und Betriebsbesichtigung aufgenommen.

Mit dem Wissen über derartige Sonderbedingungen und -sachverhalte, die den Unternehmensleitenden aufgrund ihrer Bedeutung für den Erfolg bekannt sein dürften, lassen sich diese in den SRP-Vorlagen nachvollziehbar interaktiv bereinigen (s. [Pkt. 4.2.9.](#)). Dabei wird ihr Einfluss umgehend ersichtlich, und sie können gezielt i.R. der Plausibilitätsbeurteilung berücksichtigt bzw. analytisch auf die betrieblichen Normalverhältnisse korrigiert werden. Ein zusätzlicher Schutz der SRP, durch den das Übersehen gesamtbildrelevanter Besonderheiten verhindert werden soll, liegt im systematischen Prüfungsduktus „vom groben Überblick zu den Details“ (s. [Pkt. 4.2.2.](#)).

Auf der Grundlage eines ergebnisoffen³³⁹ protokollierten Zusammenstellens wichtiger Informationen aus der betrieblichen Sphäre zu einem **Betriebsprofil** werden von vornherein realistische (betriebsnahe) Erwartungen an Gesamtbildergebnisse gestellt – im Gegensatz zu generalisierten Branchenerfahrungen mit Übertragungsproblemen. Dabei ergeben sich die zwei **Kernfragen** zur (regelbasierten) Zeitreihenanalytik mit:

- ? Welche grundsätzliche Entwicklungstendenz von Erfolgsgrößen herrscht vor und für welche Betriebsfaktoren ist welche Intensität an Abhängigkeit im Unternehmen anzunehmen [z.B. Wareneinkauf zum Absatz, variable Kosten(stellen) zum Umsatz, Kostenveränderungen einzelner Unternehmensteile, Übereinstimmung von vertraglichen Soll- und Istwerten]?
- ? Welcher Grad an kaufmännischer Handlungslogik lässt sich aus der Vorbildung und Erfahrung ausführender Personen sowie aus indirekten Merkmalen wie bspw. dem Unternehmensauftritt, der Ablaufplanung oder den Absatzstrategien ableiten?

Bezogen auf die – in der Finanzrechtsprechung bisher verbreitetste – Zeitreihenanwendung „Aufschlagsatzentwicklung zum Warenfluss eines gastronomischen Betriebs“ hat sich auch der BFH zur **zeitlichen Intensität der Abhängigkeit von Betriebsfaktoren** ge-

³³⁸ Unter dem Begriff übernommen in: Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 16.

³³⁹ Insofern unzutreffend wiedergegeben von: Bleschick, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung“, DStR 2017, 353, 357: „Im ersten Schritt wird ein sog. betriebswirtschaftliches Profil erstellt. Hierzu identifiziert der Außenprüfer in der Buchführung des Steuerpflichtigen zwei Faktoren, von denen der eine *regelmäßig* der Umsatz und der andere der Wareneinkauf ist. Der Außenprüfer geht also von der Annahme aus, dass bei steigender Nachfrage und mithin steigendem Umsatz zunehmend auch Ware benötigt werde und umgekehrt eine verminderte Nachfrage einen verminderten Umsatz und einen verminderten Wareneinkauf zur Folge habe.“

äußert³⁴⁰; dass die Voraussetzungen unterschiedlicher Zeitreihenvarianten nicht ohne Weiteres vergleichbar sind, lässt sich an dem Methodenvergleich *unter Pkt. 4.2.* nachvollziehen und wird offenbar auch von den beteiligten RiBFH gesehen³⁴¹. Dennoch hat eine betriebs-spezifische Annahme der zu erwartenden Korrelationsintensität grundsätzliche Bedeutung für die Auswahl einer sachgerechten Vergleichsperiode. Um Fehler in diesem Bereich der Zeitreihenanalytik zu verhindern, arbeitet der systematische Prüfungsduktus der SRP von groben Aggregationsebenen der Daten zu feineren schrittweise hinunter (s. *Pkt. 4.2.2.*), so- dass widersprüchliche Teilergebnisse zum Ergründen der Ursachen anregen.

Folgende Aspekte bestimmen die Intensität betrieblicher Abhängigkeiten:

1. Ausgangspunkt eines chronologischen Zusammenhangs von betrieblichen Datenreihen ist die **Notwendigkeit der Faktorvorhaltung**: Für den reibungslosen Betriebsablauf und insbesondere zur Erlöserzielung sind die ausreichende Bereitstellung von Ware bzw. Leistung so- wie benötigter Kostenpositionen wie bspw. Personal Vo- raussetzung. Dabei wird die zeitliche Intensität der Korre- lation z.B. von der Güte der Einkaufsplanung, der Liquidi- tät im Unternehmen, der Haltbarkeit von Waren sowie der Lagerkapazität bestimmt.

Im Fall von variablen Kostenpositionen können vertragliche Vorgaben die Reaktionszeit erheblich verlängern (z.B. Festanstellungen). Zudem bestehen häufig Ersatzmöglichkei- ten (bspw. Substituierbarkeit von Festangestellten durch Aushilfen), sodass (zunächst) längere Vergleichsperioden wie Quartale anzuraten sind.

2. Die **Ablauforganisation** eines Betriebs kann Umstände mit sich bringen, die den operativen Zusammenhang zweier Betriebsfaktoren verzerren: Gewohnheitsmäßiges Einkaufsverhalten kann z.B. tatsächlichen Notwendigkeiten tlw. zuwiderlaufen. Auch Ab- rechnungsmodalitäten wie bspw. die Inrechnungstellung zum Monatsende können bei kurzen Vergleichsperioden wie Wochen relevante Verzerrungen hervorrufen. *Sowohl ei- ne allzu bedarfsferne Einkaufsplanung als auch eine sehr marktunübliche Zahlungswei- se gefährden jedoch erfahrungsgemäß die Unternehmensexistenz und müssen deshalb*

Abhängigkeiten
und deren Intensität:

- ? Art der Warenbedarfs-
analyse (*erfahrungs- vs.
erwartungsorientiert*)
- ? Einkaufsrhythmus (*preis-
vs. qualitätsorientiert*)
- ? Abrechnungsmodalitäten
- ? Bestandsstrategie und
Lagermöglichkeiten
- ? Bildung der Verkaufs-
preise (*Marktgleichgewicht
vs. Produktabgrenzung*)
- ? Substituierbarkeit von
Kostenfaktoren
- ? ...

³⁴⁰ BFH v. 25.3.2015, BStBl. II 2015, 743, Leitsatz 1: „Die Durchführung eines Zeitreihenvergleichs setzt voraus, dass im Betrieb das Verhältnis zwischen dem Wareneinsatz und den Erlösen im betrachteten Zeitraum weit- gehend konstant ist.“

³⁴¹ Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, mit der entsprechenden Ein- schränkung im Titel.

ohne ordnungsgemäße, nachvollziehbare Unterlagen eher als Indikatoren für unschlüssige Besteuerungsaufzeichnungen angesehen werden.

Punktuelle Sondereinflüsse können – nach Bestätigung durch entsprechende Dokumente wie Lieferscheine – interaktiv (s. [Pkt. 4.2.9.](#)) auf Normalverhältnisse korrigiert werden, um die tatsächlichen Auswirkungen festzustellen.

3. Weiterhin kann die **Erfassung der Geschäftsvorfälle** ursächlich für gesamtbildrelevante Störungen sein.

Nach den umfangreichen praktischen Anwendungserfahrungen zur SRP beeinträchtigen Erfassungsbesonderheiten wie z.B. die Buchung von Wareneinkäufen nach Bezahlung bei üblichen Zahlungszielen das robuste Zeitreihenmodell „doppelskaliert größenangepasster Zeitreihenvergleich“ nicht, was sich an den Extrembeispielen „Aufzeichnung nach Bezahlung“ und „lange haltbare Waren“ unter [Pkt. 4.2.](#) gut nachvollziehen lässt. Dennoch bietet die SRP-Vorlage BWA für den Sonderfall der Einnahmen-Überschuss-Rechnung gem. § 4 Abs. 3 EStG die interaktive Option der Vor- bzw. Rückdatierung der ersten Datenreihe um bis zu 30 Tage (s. [Pkt. 4.2.9.3.](#)). Dadurch lässt sich der tatsächliche Einfluss problemlos und zügig simulieren.

Im Verhältnis gravierender Ordnungswidrigkeiten zu einer Schlüssigkeitsprüfung verstärken sich die beiden unabhängigen Voraussetzungen der „Beweiskraft der Buchführung (und Aufzeichnungen)“ nach § 158 AO gegenseitig i.S. einer Indiziengesamtheit³⁴². Entsprechend schwächt auch die „Drei-Stufen-Theorie“ des BFH die Sorgfaltsanforderungen bei der Schätzung erheblich ab, wenn Besteuerungsaufzeichnungen formell und (offensichtlich) materiell unrichtig sind.³⁴³ Zudem beschreibt die AO-Kommentierung den aus der Prüfungspraxis hinlänglich bekannten Zusammenhang aus Ordnungswidrigkeit und sachlicher Unrichtigkeit.³⁴⁴ Bei substantieller³⁴⁵ Mitwirkungsbereitschaft ermöglichen

³⁴² BGH v. 27.4.2010 – 1 StR 454/09, NSZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreicht.“

³⁴³ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 62 ff.

³⁴⁴ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 24: „Allerdings entspricht es der Erfahrung, dass eine formell ordnungswidrige Buchführung i.d.R. auch sachliche Mängel aufweist.“

³⁴⁵ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. Die Frage der Feststellungslast stellt sich, wenn ein entscheidungserheblicher Sachverhalt trotz Ausschöpfung aller zugänglichen und zumutbaren Ermittlungsmöglichkeiten nicht vollständig aufgeklärt werden kann. Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u. a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann,

die interaktiven Funktionen der SRP-Vorlage BWA (s. [Pkt. 4.2.9.](#)) trotzdem in vielen Fällen die gemeinsame³⁴⁶ Herstellung einer ausreichenden Transparenz für eine sachgerechte Plausibilitätsbeurteilung gem. § 158 Abs. 2 Nr. 1 AO. Folglich müssen verbleibende Unklarheiten aufgrund von Erfassungsmängeln als Bestätigung beobachteter Unschlüssigkeiten gedeutet werden.

Entsprechend der generell begrenzten³⁴⁷ Beweismöglichkeiten aller Gesamtbildprüfungen stellt § 158 Abs. 2 Nr. 1 AO nur indizielle Anforderungen an eine sachliche Beanstandung (s. [Pkt. 4.6.](#)), und *erfahrungsgemäß* führt schlechtes Wirtschaften in gegebener Marktkonkurrenz zum Scheitern eines Unternehmens.³⁴⁸ Dennoch helfen betriebspezifische Kenntnisse über die vorhandene **unternehmerische Kompetenz** bei der Beurteilung von aufgefundenen Regelstörungen in Zeitreihenanalysen. Um fehlendes Geschick von dem Verdacht falscher Besteuerungsgrundlagen unterscheiden zu können, stellt die SRP-Vorlage BWA viele interaktive Funktionen zur Verfügung (s. [Pkt. 4.2.9.](#)), mit deren Hilfe Verprobungsauffälligkeiten (gemeinsam) ergründet und – bei nachvollziehbarer, belegter Erklärung – für die weitere Gesamtbildprüfung auf Normalverhältnisse korrigiert werden können.

Die unternehmerische Kompetenz im betreffenden Einzelfall lässt sich anhand folgender Faktoren einschätzen:

wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“

³⁴⁶ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen.“

³⁴⁷ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 323-324: „Prüfungen vergangener Aufzeichnungen können, nach dem logischen Menschenverstand, nur zu Wahrscheinlichkeitsaussagen führen. Der Grund dafür ist, dass es sich bei Aufzeichnungen um ein reduziertes Abbild komplexer Daten und Zusammenhänge handelt. [...] Schlussfolgernd kann bei der Prüfung von (steuerlichen) Daten keine (100 %) Sicherheit und somit kein Beweis vorliegen bzw. erfolgen. [...] Die herkömmlichen Methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...] Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten.“

³⁴⁸ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 28: „Es darf einem Gewerbetreibenden durchaus unternehmerisches Handeln unterstellt werden, sodass ein grober Verstoß gegen die betriebswirtschaftliche Handlungslogik nicht plausibel erscheint.“

- a. **Vorbildung und Erfahrung** handelnder Personen ergeben sich z.T. aus den Steuerakten und werden häufig im Eröffnungsgespräch (unaufgefordert) berichtet. Dabei ist zu bedenken, dass viele Ausbildungen auch kaufmännische Grundlagen vermitteln.

Praxiserfahrungen aus dem langjährigen Führen eines Unternehmens bzw. das Aufwachsen in einem Familienbetrieb dürften die wirtschaftliche Kompetenz i.d.R. mindestens genauso geschult haben wie ein Wirtschaftsstudium.

Wirtschaftlichkeit des Handelns:

- ? Vorbildung im Betrieb
(Unternehmer*innen oder leitende Angestellte)
- ? Erfahrung im Betrieb
(Unternehmer*innen oder leitende Angestellte)
- ? Marktpräsenz
(Bekanntheitsgrad und Corporate Identity)
- ? Marketingstrategien
(Absatzwege, innovative Produkte, Werbung etc.)
- ? ...

- b. Noch deutlicher lässt sich die unternehmerische Kompetenz aus **indirekten Beobachtungen** ableiten. So sprechen die folgenden Beispiele einer erfolgreichen Firmenvermarktung für ein darüber hinausgehendes Geschick mit demzufolge (weitgehend) schlüssigen Betriebsdaten und -entwicklungen: einheitlicher Marktauftritt mit hohem Wiedererkennungswert („Corporate Identity“), strukturierte Personal- oder Ablaufplanung, gelungene Produktabgrenzung vom Massenmarkt insbesondere bei hohem Preisdruck, Eröffnung alternativer Absatzwege, kreative Werbung etc.

Da Außenprüfungen – abgesehen von Nachschauen – i.d.R. kaufmännisch lange Zeiträume betrachten, ist eine regelbasierte Gesamtbildprüfung ohne Auffälligkeiten eher unrealistisch. In fast jedem Betrieb kommt es irgendwann zu unvorhersehbaren Markteinflüssen, die sich in den Unternehmenszahlen und -entwicklungen als Regelstörungen wiederfinden lassen. Umso wichtiger für die gesamte Plausibilitätsbeurteilung gem. § 158 Abs. 2 Nr. 1 AO sind die **Bereitschaft zur und Qualität der Mitwirkung bei der Sachaufklärung**³⁴⁹.

Sollten vorgefundene Unschlüssigkeiten der zentralen SRP-Zeitreihe (doppelskaliert größenangepasster Zeitreihenvergleich, s. [Pkt. 4.2.](#)) i.R. der Vorstellung im Unternehmen (offenbar) erstmalig wahrgenommen werden, sind Zweifel an der Richtigkeit der Besteuerungsdaten durchaus angebracht, denn das robuste Zeitreihenmodell arbeitet auf Quartals- und Monatsbasis ausschließlich betriebsbezogen existentielle Vorfälle heraus. Deren Ursachen sollten bereits bekannt sein und können anhand nachvollziehbarer Erklärungen und quantifizierender Unterlagen konkret interaktiv untersucht werden (s. [Pkt. 4.2.9.](#)). Außerdem sind nach einschneidenden Betriebsproblemen Kompensationsmaßnahmen wie bspw. Absatzaktionen, Kosteneinsparungen etc. zu erwarten. Insofern liefert die kooperative Ausei-

³⁴⁹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 9: „§ 162 II i.V.m. § 158 orientieren das Beweismaß an der Sphärenverantwortlichkeit [...] wie viel und in welcher Form der Stpfl. aus seiner Verantwortungssphäre zur Sachaufklärung beiträgt und wieweit die FinBeh. auf dessen Mitwirkung [...] angewiesen ist.“

nersetzung mit entdeckten Unschlüssigkeiten i.S. der gemeinsamen Sachaufklärungsverantwortung³⁵⁰ aussagekräftige Hinweise auf die Vertrauenswürdigkeit der Besteuerungsgrundlagen insgesamt.

4.2.2. Prüfungssystematik der SRP-Vorlage BWA

Im Gegensatz zu willkürlichen Methodenzusammenstellungen verfolgt die systematische Verknüpfung geeigneter Zeitreihenanalysen in der SRP-Vorlage BWA³⁵¹ den Prüfungsduktus „**vom groben Überblick zu den Details**“³⁵², um die Entwicklungen, Regeln und Zusammenhänge von Besteuerungsdaten zu verstehen. Prüffelder können so zügig und effektiv aufgedeckt bzw. die Gesamtplausibilität qualifiziert bewertet werden.³⁵³ Dabei wird die Unterscheidung zwischen systemischen³⁵⁴ Fehlern bspw. in Stammdaten und einzelfallbezogenen Besonderheiten verbessert, und es lassen sich punktuelle Prüffelder weitgehend eingrenzen, damit sie (rechtlich) geprüft und bei Bedarf interaktiv (s. [Pkt. 4.2.9.1.](#)) auf Normalverhältnisse korrigiert werden können. Zusätzliche Datenaufbereitungen der SRP-Vorlage BWA wie die jahresweise Gegenüberstellung von Monatssummen und Medianen der Wochentage dienen der erweiterten mehrperspektivischen Gesamtfallbetrachtung, um eine klarere Differenzierung „betrieblich veranlasst versus prüfungsrelevant“ sicherzustellen, soweit es die Datenqualität im Einzelfall zulässt.

Der SRP-Zeitreihensystematik liegt das Methodenwissen zu Grunde, dass Zeitreihenvarianten und -ausgestaltungen unterschiedliche Informationen aus zu untersuchenden Daten herausarbeiten (s. [Pkt. 4.2.](#)). Dabei haben die **Fuzzy-Theorien** (s.a. [Pkt. 2.](#)) eine wichtige Bedeutung, wonach der Grad der Aufsummierung einen wesentlichen Einfluss darauf hat,

³⁵⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen“.

³⁵¹ Der regelbasierte Prüfungsduktus der SRP-Vorlage BWA wird durch die Abfolge der Arbeitsblätter mit ihren jeweiligen Auswertungen und Funktionalitäten sowie die Kurzanleitung vorgegeben. Die Vorlagenbereiche zur Gesamtbildprüfung sind von denen für die Durchführung einer Schätzung – *nach eingetretener Amtspflicht zur Schätzung gem. § 162 AO* – farblich abgesetzt.

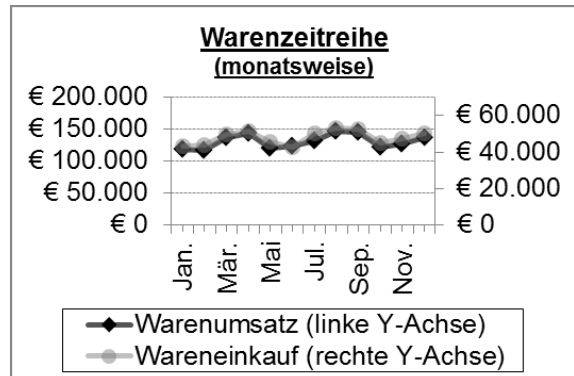
³⁵² Der Funktionsweise der in Wirtschaft und Beratung zunehmend verbreiteten B(usiness-)I(ntelligence)-Programme liegt bei sinnvollen Datenmodellen eine vergleichbare Prüflogik zu Grunde. Insofern erleichtert die SRP mit ihrer Prüfungssystematik den Einstieg in BI.

³⁵³ So auch: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1 u. 4: „Bei einer steuerlichen Außenprüfung können mathematisch-statistische (quantitative) Prüfungsmethoden eingesetzt werden, um die Besteuerungsgrundlagen des Steuerpflichtigen zu verproben und auf Plausibilität zu überprüfen. Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich.“

³⁵⁴ Systemische Fehler bzw. Besonderheiten der Datenerfassung und -verarbeitung wie falsche Schlüsselungen (z.B. USt-Satz) oder die verzögerte Aufzeichnung der Einnahmen-Überschuss-Rechnung (§ 4 Abs. 3 EStG) lassen sich häufig nur oder klarer auf stark verdichteten Ebenen wie Quartalen erkennen.

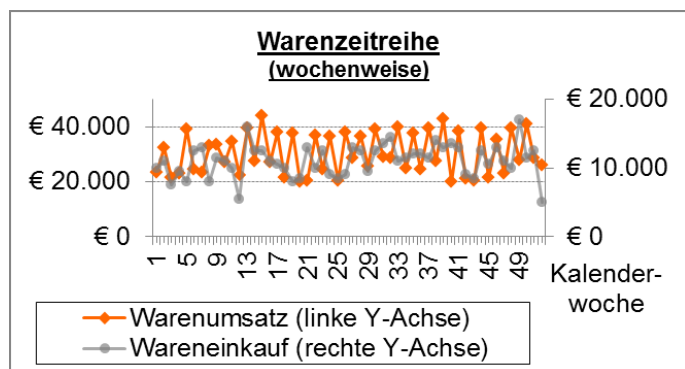
welche Dateninhalte erkennbar sein können. Dem folgenden Zeitreihenbeispiel eines Gemüsehändlers lassen sich die abweichenden Schwerpunkte der Verbildlichung mit ihren Konsequenzen u.a. für die Plausibilitätsbewertung klar entnehmen:

- © Auf der Monatsebene des doppelskalierten Zeitreihenvergleichs wird die wegen der geringen Haltbarkeit von Gemüse erwartete enge Abhängigkeit zwischen Wareneinkauf und -umsatz in Form fast perfekter Deckungsgleichheit (Korrelation) der beiden Grafen erfüllt.



Nach der SRP-Zeitreihensystematik beginnt die Untersuchung mit dem Tabellenblatt „Quartalsrelationen“ auf einer noch höher verdichteten Summationsstufe, um zunächst einen guten Überblick zu erlangen, bevor schrittweise zu den detailreicheren, leicht überfordernden Analysen kurzer Vergleichsperioden fortgesetzt wird.

- © Im tiefergehenden doppelskalierten Zeitreihenvergleich auf Wochenbasis offenbaren sich demgegenüber **regelmäßige Schwankungen des Absatzes ca. alle zwei Wochen**, welche die – erwartete und monatlich beobachtete – enge Abhängigkeit zwischen Wa-



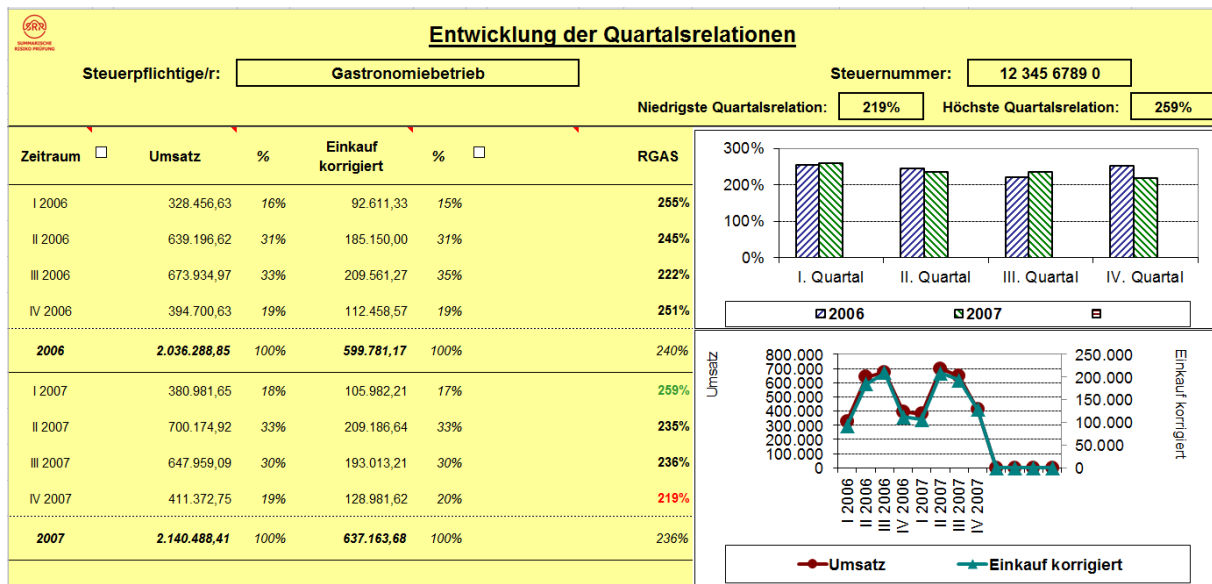
reneinkauf und -umsatz in Frage stellen. Der offensichtliche Widerspruch beider Zeitreihenanalysen führt dazu, sich die Einzeldaten zum Umsatz unter chronologischer Sortierung fokussiert auf besondere Ereignisse alle zwei Wochen anzuschauen. Mit diesem Vorwissen fällt sofort auf, dass im geprüften Unternehmen jeweils zu Monatsmitte und -ende Rechnungen an bestimmte Kunden gestellt werden. Somit erklärt ein eher untypischer Abrechnungsmodus für die betreffende Branche, dass eine wöchentliche Zeitreihenanalyse ohne periodengerechte Aufteilung der Rechnungsumsätze nicht für die Schlüssigkeitsbewertung in Betracht kommt. Die geeignete monatliche Warenflusszeitreihe legt mit ihrer sehr guten Handlungslogik keine weitere Intensivierung nahe.

Im Ergebnis werden durch das systematische Vertiefen der SRP-Zeitreihen vom groben Überblick langer zu den detaillierteren Perspektiven kürzerer Vergleichsperioden nicht nur betriebspezifische Prüffelder mit höherer Effektivität aufgedeckt, sondern Plausibilitätsbeurteilungen gem. § 158 Abs. 2 Nr. 1 AO gegen falsche Schlussfolgerungen abgesichert – auch

im Hinblick darauf, dass protokollierte Betriebsprofile (s. [Pkt. 4.2.1.](#)) unvollständig sein können.

Neben dem schrittweisen Arbeiten mit der Vergleichsperiode „vom groben Überblick zu den Details“ werden im SRP-Zeitreihensystem unterschiedliche Zeitreihenvarianten mit ihren jeweiligen Stärken für eine effiziente, weitest mögliche **Eingrenzung aufgefundener Regelstörungen zu ursächlichen Sachverhalten** genutzt. Im Anschluss können die Einzelprüffelder auf richtige (rechtliche) Behandlung untersucht werden. Zudem lassen sie sich bei Bedarf interaktiv (s. [Pkt. 4.2.9.1.](#)) auf Normalverhältnisse korrigieren, soweit sie eine vorzunehmende Schlüssigkeitsbewertung nach § 158 Abs. 2 Nr. 1 AO stören.

Am Beispiel des Warenflusses eines größeren gastronomischen Betriebs soll die SRP-Zeitreihensystematik vom Verstehen der Größenzusammenhänge über das Erkennen bis zum Eingrenzen von konkreten Prüffeldern beschrieben werden:



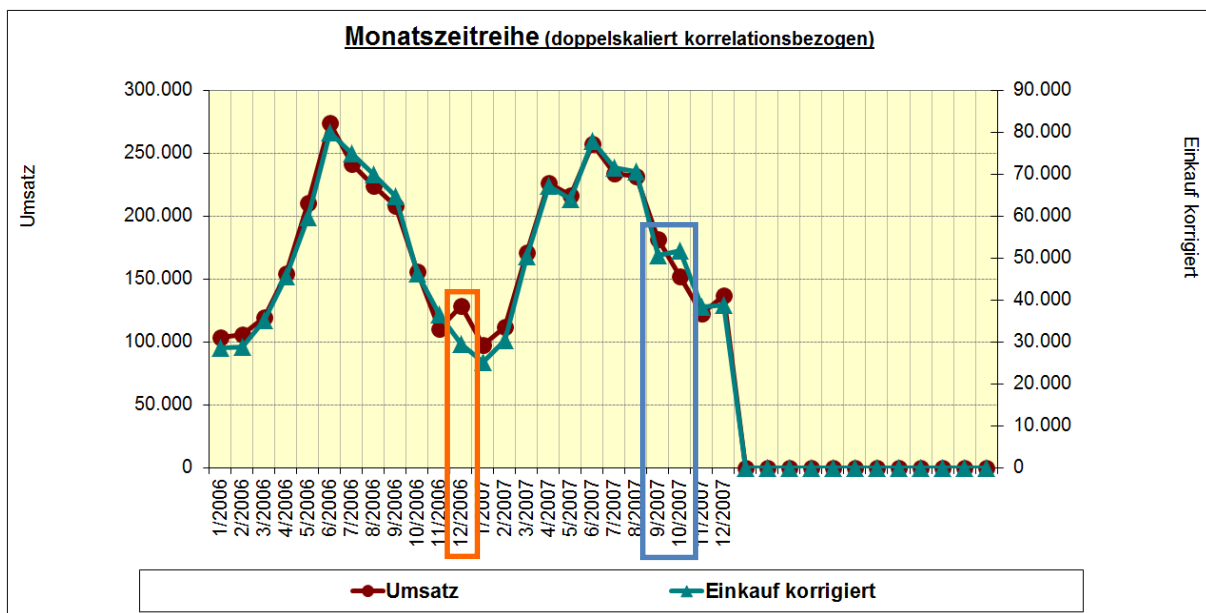
Zur groben Verbildlichung der quartalsweisen Entwicklungen und des Zusammenhangs beider Datenreihen werden zwei Ansichten zur Verfügung gestellt:

1. Die jahresweise Gegenüberstellung der Quartalssummen als Verhältnisgröße (hier: „RGAS“) soll besonders wiederkehrende Strukturveränderungen anzeigen.³⁵⁵ Im Beispielsfall ist keine gleichmäßige Erhöhung oder Absenkung in einzelnen Jahresteilern erkennbar, woraus sich schließen lässt, dass die veränderten Nachfragegewohnheiten der Kunden über das Jahr keinen relevanten Einfluss auf visualisierende Beurteilungen der Warenflüsse haben.

³⁵⁵ Für die eigentliche Plausibilitätsbeurteilung gem. § 158 Abs. 2 Nr. 1 AO hat der Aufschlagsatz innerhalb der SRP keine Bedeutung, was in der kritisierenden Literatur regelmäßig unzutreffend dargestellt wird (vgl. z.B.: Beyer, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, 2933-2934: „Hierzu werden die einzelnen Quartals-RAS in einem Diagramm dargestellt. Schwanken die RAS in den Quartalen, kann dies mit Änderungen der Umsatzträger des Unternehmens zu tun haben“).

2. Mit dem streng chronologischen, doppelskaliert größenangepassten Zeitreihenvergleich wird grob³⁵⁶ die Reaktion des Warennachkaufs im Betrieb auf die Absatzverhältnisse wiedergegeben. Im Beispielsunternehmen zeigt sich eine (nahezu) perfekte Anpassung an die stark veränderten Nachfragesituationen der einzelnen Jahreszeiten.

Als nächster Schritt wird die Prüfungsmaßnahme mit dem – für die Abhängigkeitsprüfung geeigneten (s. [Pkt. 4.2.](#)) – doppelskaliert größenangepassten Zeitreihenvergleich vertieft, indem der Monat als deutlich kürzeres Vergleichsintervall erheblich mehr Informationen aus den Daten erkennbar werden lässt:



Weiterhin ist der enge Zusammenhang beider Betriebsfaktoren unproblematisch abzulesen. Als Folge der wesentlich detaillierteren Analyse der Besteuerungsdaten sind nun aber auch eine klare **Ablösung des Wareneinkaufs 12/2006 vom ansteigenden Umsatz** sowie eine **Zuordnungsungenauigkeit**³⁵⁷ in den Monaten 9-10/2007 zu beobachten; während sich dabei

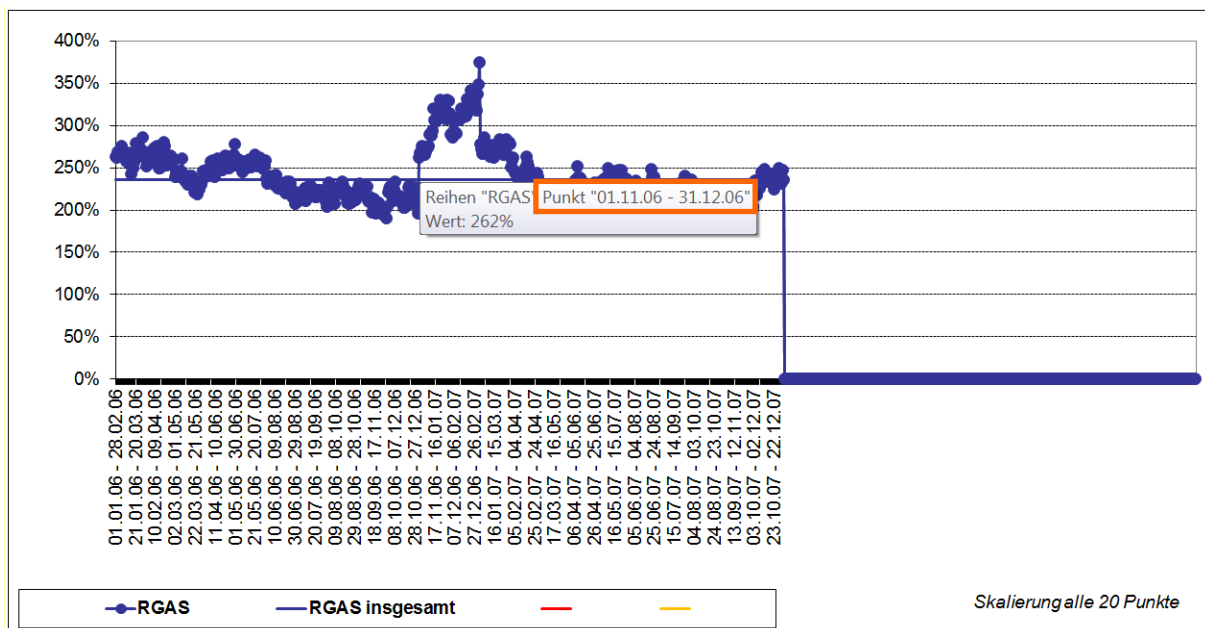
³⁵⁶ Infolge des statistischen Ausgleichs von einzelnen Ausreißerwerten durch die Aufsummierung über lange Quartalszeiträume ist nur eine sehr grobe Beurteilung der Handlungslogik möglich. Viele Einzelprüffelder sind auf dieser Ebene i.d.R. nicht deutlich genug sichtbar.

³⁵⁷ In Rechtsprechung und Literatur wird vor Zuordnungsunsicherheiten in Zeitreihenanalysen vehement gewarnt (vgl. z.B.: BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 47: „Dies zeigt, dass bereits eine Veränderung in der Zuordnung der Einkäufe nur einer einzigen Woche erhebliche Auswirkungen auf das Gesamtergebnis der Zehn-Wochen-Periode hat“ bzw. Wigger, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168, 171: „Von großer Bedeutung für die Qualität und Beweiskraft des ZRV ist, dass der Schritt zum zutreffenden ‚Periodenwareneinsatz‘ gelingt.“), obwohl die verbildlichte Schlüssigkeitsprüfung gerade nicht in der Beurteilung einer einzelnen Zahl bzw. eines Diagrammpunkts besteht, sondern der chronologische Gesamteindruck vor solchen Fehlschlüssen wie im Beispiel schützt; s.a.: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 27: „Die Qualität des Zeitreihenvergleichs wird maßgeblich dadurch bestimmt, wie sorgfältig die betriebswirtschaftlichen Größen erhoben werden und wie genau betriebliche Besonderheiten geprüft werden. [...] Dies ist keinesfalls ein Argument gegen die Methoden, sondern zeigt lediglich die bei jeder Verprobungsmethode obligatorische Verpflichtung zur sorgfältigen Anwendung ebendieser.“

der (relativ) überhöhte Wareneinkauf in 10/2007 zumindest tlw. durch den Vormonat aufhebt, offenbart sich in 12/2006 eine Regelstörung, die als Einzelprüffeld nachverfolgt wird.

Da Wochenzeitreihen in den meisten Branchen zu störanfällig für alltägliche Betriebsabläufe wie z.B. Bestandseinkäufe sind, nutzt die SRP zur weiteren Konkretisierung aufgedeckter Regelstörungen ein gleitendes³⁵⁸ Zeitreihenmodell. Hierbei wird die Entwicklung der Verhältnisgröße³⁵⁹ beider Datenreihen zwar für einen robust langen Zeitraum wiedergegeben, jedoch immer nur um einen Tag weitergeschrieben. In der Standardeinstellung der SRP-Vorlage BWA „2 Monate“ (s. [Pkt. 4.2.6.](#)) folgt dann auf den Aufschlagsatz zum Startintervall „1.1. bis 28.2.“ der für „2.1. bis 1.3.“ und danach für „3.1. bis 2.3.“ etc. Diese Auswertungsform bewirkt, dass ein besonderer Einzelsachverhalt, der ursächlich für die beobachtete Auffälligkeit ist, an zwei Punkten des Diagramms den Grafen abrupt verzerren muss: beim Ein- und Austreten in die bzw. aus der Vergleichsperiode.

Zum Beispielsfall mit der Regelstörung „12/2006“ ist ein plötzliches Ansteigen des Aufschlagsatzes im **hervorgehobenen Startintervall „1.11. bis 31.12.2006“** festzustellen:



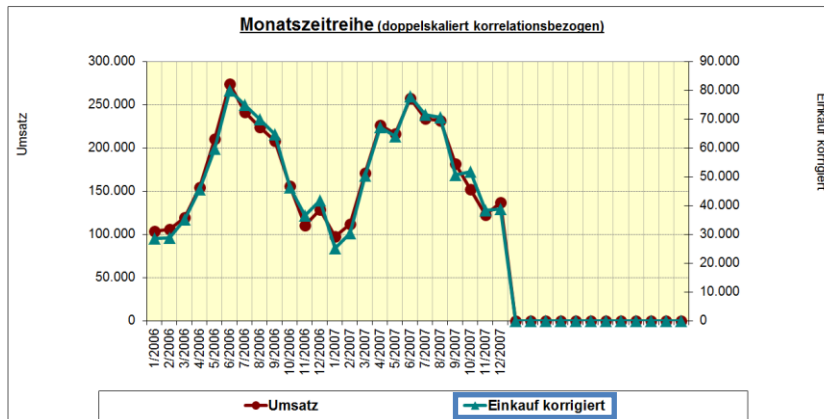
Nach diversen Punkten auf einem höheren Niveau fällt der Graf dann wieder abrupt ab. *Dieses typische Verhalten der Gleitschlittenzeitreihen erklärt sich im Beispiel dadurch, dass ein außergewöhnlich niedriger Tageswareneinkauf am 31.12.2006 erstmalig die betreffende Vergleichsperiode beeinflusst. Da die Folgeintervalle aber immer nur einen Tag weitergleiten, bleibt das kritische Datum etliche Punkte lang in den jeweiligen Aufschlagsätzen wirksam, bis sein Einfluss mit dem Intervall „1.1. bis 28.2.2007“ plötzlich verschwindet.*

³⁵⁸ Auch „rollierende Zeitreihe“, vgl.: Harle/Olles, „Die moderne Betriebsprüfung“, NWB-Verlag, 3. Auflage 2017, Rz. 1366 ff.

³⁵⁹ Das Gleitschlitten-Zeitreihenmodell wird in der SRP-Vorlage BWA zusätzlich für den doppelskalierten Zeitreihenvergleich zur Verfügung gestellt.

Durch eine gezielte Filterung beider Datenreihen auf den 31.12.2006 lässt sich zügig ein hoher Negativbetrag im Wareneinkauf finden. Nach Belegsichtigung wird deutlich, dass es sich um eine Bonuszahlung handelt, die nicht über ein Extrakonto transparent abgegrenzt worden ist. Somit hat die Aufklärung zur Konsequenz, dass die Behandlung etwaiger Boni für das Folgejahr intensiv geprüft wird, weil dort laut doppelskalierten Monatszeitreihenvergleich keine vergleichbare Ablösung des Wareneinkaufs vom -umsatz zu sehen ist.

Die Gegenkontrolle mit [interaktiv](#) (s. [Pkt. 4.2.9.1.](#)) [egalisiertem Warenbonus vom](#)



[31.12.2006](#), ob der identifizierte Sondersachverhalt vollumfänglich die beobachtete Zeitreihenauffälligkeit verursacht hat, führt zum schlüssigen Gleichlauf des Warennachkaufs zur aufoktroierten Marktsituation „Nachfrage“.

Insgesamt ermöglicht es die systematische SRP-Zeitreihenanalyse entlang des Duktus „vom groben Überblick zu den Details“ und unter planvoller Kombination unterschiedlicher Zeitreihenvarianten, in vielen tausend Einzelvorgängen die betriebsspezifischen Prüffelder zielgenau und zügig herauszuarbeiten. Auch das Ergebnis einer Plausibilitätsbeurteilung wird dadurch wesentlich qualifizierter, weil – neben der besseren Eignung des zentralen SRP-Modells „doppelskaliert größenangepasster Zeitreihenvergleich“ (s. [Pkt. 4.2.](#)) – z.B. die im Zusammenhang mit Zeitreihenanalysen stets problematisierten Bestandseinkäufe wirksam identifiziert und umverteilt werden können.

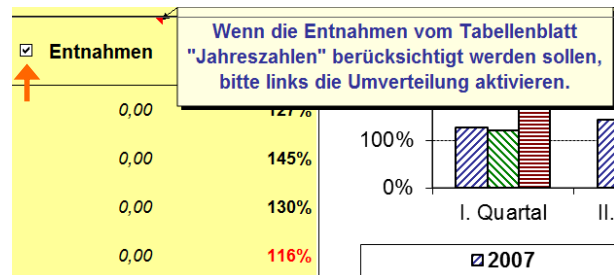
4.2.3. Jahressummen

Jahreszahlen				
Steuerpflichtige/r:	Café			Steuernummer: 12 345 6789 0
	2007	2008	2009	
oberer Richtsatzwert:	400%	400%	400%	
mittlerer Richtsatzwert:	257%	257%	257%	
unterer Richtsatzwert:	186%	186%	186%	
Warenumsatz:	72.273,01	73.121,67	81.548,09	
Wareneinkauf:	31.545,43	33.551,15	31.563,40	
Warenanfangsbestand:	2.716,50			
Bestandsveränderungen:	-198,50	-103,00	1.106,00	
Warenendbestand:	2.518,00	2.415,00	3.521,00	
Entnahmen:	0,00	2.895,00	3.108,00	
Einsatz:	31.743,93	30.759,15	27.349,40	
RGAS:	128%	138%	198%	
Gewinn:	11.254,00	13.188,00	6.631,00	
Gewinnsatz (Umsatzrentabilität):	16%	18%	8%	

Aus der jahresweisen Aufsummierung von Besteuerungsdaten können i.d.R. keine neuen Erkenntnisse gegenüber dem Veranlagungsverfahren gewonnen werden. Deshalb liegen die

wesentlichen Funktionen des Tabellenblatts „Jahreszahlen“ darin, die hineinkopierten Daten zügig auf vollständige Übernahme kontrollieren zu können (s.a. [Pkt. 4.](#)) und die eventuell zeitreihenrelevanten Informationen „Bestandsveränderungen“³⁶⁰ und „Entnahmen“ zu ergänzen. Sollen die jährlichen Bestandsveränderungen – entweder zur Simulation oder wegen genauerer Kenntnisse – punktuell in der Warenflussanalyse berücksichtigt werden, lässt sich die [Berechnung über das Ankreuzfeld unterdrücken](#), was die SRP-Vorlage BWA durch hellgraue Schrift anzeigt. Anschließend können die Anpassungen interaktiv auf dem Tabellenblatt „Vergleichsdaten“ eingegeben werden (s. [Pkt. 4.2.9.1.](#)).

Eingetragene jährliche Warenentnahmen lassen sich für die quartals- und monatsweisen Zeitreihenanalysen einheitlich und gleichmäßig über die **Anhakfelder** in den Tabellenblättern „Quartalsrelationen“ und „Monatszeitreihen“ auf die (vorhandenen) Perioden verteilen.³⁶¹ Eine entsprechende **Option** besteht für die Bestandsveränderungen nur



i.R. der Quantilsschätzung bei der Berechnung der 80 %-Quantile (s. [Pkt. 4.8.](#)), falls keine genauere Zuordnung möglich ist, weil diese Art Auf- und Abbau von Warenbestand nicht dem üblichen Geschäftsablauf entspräche; über die interaktiven Korrekturoptionen auf dem Tabellenblatt „Vergleichsdaten“ lassen sich – unter Mitwirkung der Stpfl. – problemlos die tatsächlichen Verhältnisse einpflegen (s. [Pkt. 4.2.9.1.](#)).

Die Verhältnisgröße beider Datenreihen „RGAS“ lässt sich auf dem Tabellenblatt „Jahreszahlen“ mit Übernahme in alle Auswertungen umbenennen und über das Ankreuzfeld rechts daneben von einer prozentualen Aufschlags- in eine direkte Verhältnisberechnung („Datenreihe 2 geteilt durch Datenreihe 1“) umstellen. Obwohl **Aufschlagsätze** weder im externen – als grösste³⁶² Verprobungsmethode – noch im internen Vergleich eine entscheidende Bedeutung i.R. der SRP haben (s.a. den Abschnitt „SRP-Zeitreihensystematik – Analysegrundsätze und Annahmen“ unter [Pkt. 4.2.11.](#)), lassen sich die Richtsatzwerte³⁶³ für den Rohge-

³⁶⁰ Die jährlichen Warenbestandsveränderungen werden aus den Eingaben für den Warenanfangsbestand und die Warenendbestände errechnet.

³⁶¹ Durch die Auftrennung von Gesamtbildprüfung (Verprobung) und Ermittlung der 80 %-Quantile für eine Schätzung in der SRP-Vorlage BWA kann der Fehler einer unterlassenen Aufteilung eingegebener Jahresentnahmen i.R. der Quantilsschätzung (s. [Pkt. 4.8.](#)) ab der Vorlagenversion 2019 nicht mehr passieren.

³⁶² Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 17: „Als grösste Methode zur Beanstandung der sachlichen Richtigkeit des Buchführungsergebnisses erweist sich der äußere Betriebsvergleich [...] Vergleich mit anderen Betrieben kommt regelmäßig ein geringerer Vergleichswert als inneren Betriebsvergleichen [...] zu“.

³⁶³ Die – für den Dreijahreszeitraum „2007 bis 2009“ bis zur Richtsatzsammlung 2021 unveränderten – Richtsätze für kleine Cafés können nachgelesen werden unter:
https://www.bundesfinanzministerium.de/Content/DE/Downloads/BMF_Schreiben/Weitere_Steuerthemen/Betr

winnaufschlagsatz zur Vollständigkeit des Schlüssigkeitseindrucks jahresweise einpflegen. Wird dieser Rahmen unterschritten, weist die rote Feldfarbe darauf hin (s. die Abb. zu Beginn dieses Pkt.).³⁶⁴ Ebenso können zur Groborientierung die jährlichen (Halb-)Reingewinne eingetragen werden, die entsprechend der Richtsatzsammlungen in Umsatzrentabilitäten („in v.H. des wirtschaftlichen Umsatzes“) umgerechnet werden.

4.2.4. Quartalsrelationen

Gemäß dem zentralen Zweck der Statistik, aus einer größeren Datenmenge *nur* die Kerninformationen herauszufiltern³⁶⁵, unterdrücken quartalsweise Aufsummierungen einer Vielzahl von betrieblichen Vorgängen die Einzelheiten zu Gunsten grundsätzlicher Größenordnungen, Tendenzen, Regeln und Muster. Diese grobe „Hubschrauberperspektive“ ist zu Beginn einer Prüfung als **Erstorientierung** wichtig, weil durch das Wissen über die grundsätzlichen Eigenschaften von Betriebsdaten die Unterscheidung zwischen (systemischen³⁶⁶) Unternehmensbesonderheiten und Aufzeichnungsfehlern verbessert wird. Andererseits darf von einer regelbasierten Gesamtbildprüfung bspw. zum Warenfluss *alleine* auf Quartalsebene keine ausreichende Gründlichkeit erwartet werden, denn die meisten einzelfallbezogenen Prüffelder bleiben unter der hochgradigen Verdichtung unsichtbar.

Die SRP verwendet zur Groborientierung **zwei zielgerichtete Ansichten**:

1. Das Zusammenfassen beider Datenreihen zu einem Verhältnisfaktor³⁶⁷ mit jahresweiser Gegenüberstellung der Quartalswerte vermittelt einen schnellen Überblick, ob unterjährige Strukturveränderungen gesamtbildrelevant sein können. So würde bspw. bei einer Kostenzeitreihe mit jahreszeitlichem Austausch wesentlicher, unterschiedlich teurer Kostenpositionen das Verhältnis wiederkehrend – quasi „wellenartig“ – schwanken. Trotz der Stabilität des doppelskaliert größenangepassten Zeitreihenvergleichs (s. [Pkt. 4.2.](#)) ist die Kenntnis starker struktureller Saisoneffekte, welche sogar auf robuster Quartalsebe-

[iebspruefung/Richtsatzsammlung/2022-11-28-download-richtsatzsammlung-2021.pdf? blob=publicationFile&v=2.](#)

³⁶⁴ Eine entsprechende bedingte Formatierung ist auf den (verprobenden) Tabellenblättern „Quartalsrelationen“ und „Monatszeitreihen“ nicht hinterlegt, weil die Plausibilitätsprüfung der SRP nicht an der Konstanz des (empfindlichen) Aufschlagsatzes ansetzt, sondern an der wesentlich robusteren Handlungslogik einer angemessenen Reaktion veränderbarer Betriebsfaktoren auf Markteinflüsse wie bspw. bei der Anpassung des Wareneinkaufs auf die Nachfrage (s. den Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.).

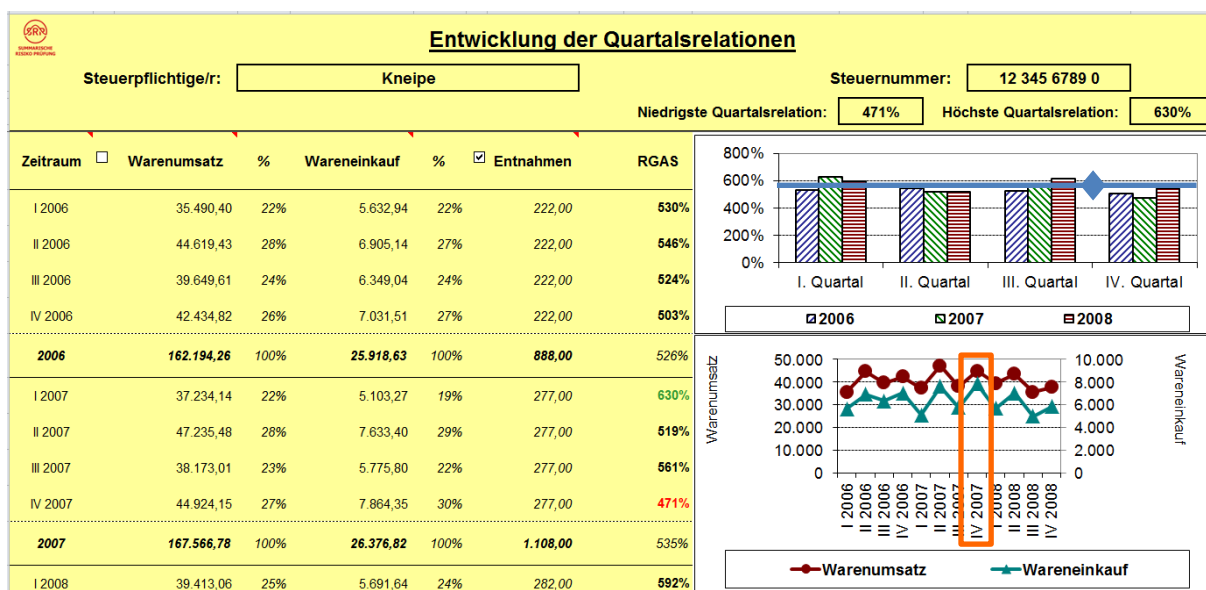
³⁶⁵ Auf diesem Zweck basiert auch die betriebswirtschaftliche Beratung und Planung mit der quartalsweisen und monatlichen Zusammenfassung umfangreicher entscheidungsrelevanter Daten.

³⁶⁶ Systemische Besonderheiten bzw. Fehler der Datenerfassung und -verarbeitung wie bspw. die verzögerte Aufzeichnung der Einnahmen-Überschuss-Rechnung (§ 4 Abs. 3 EStG) oder falsche Schlüsselungen (z.B. USt-Satz) lassen sich häufig nur oder besser auf stark verdichteten Ebenen wie Quartalen erkennen.

³⁶⁷ Die Berechnung des Verhältnisfaktors kann auf dem Tabellenblatt „Jahreszahlen“ per Ankreuzfeld als Aufschlagsatz mit „ $\frac{\text{Datenreihe 2} - \text{Datenreihe 1}}{\text{Datenreihe 1}}$ “ oder als Quotient $\frac{\text{Datenreihe 2}}{\text{Datenreihe 1}}$ ausgewählt werden; die Ergebnisse werden prozentual ausgegeben.

ne offensichtlich sind, für die sachgerechte Deutung insbesondere von Monatszeitreihen wichtig. Die Schlüssigkeit von Besteuerungsgrundlagen gem. § 158 Abs. 2 Nr. 1 AO wird i.R. der SRP *nicht* anhand von **Aufschlagsatzentwicklungen** beurteilt (s.a. den Abschnitt „SRP-Zeitreihensystematik – Analysegrundsätze und Annahmen“ unter [Pkt. 4.2.11.](#)); demzufolge wird die farbliche Hervorhebung von Werten, welche die unteren Richtsätze vom Tabellenblatt „Jahreszahlen“ unterschreiten, hier nicht fortgeführt.

- Darunter werden beide Datenreihen mit dem doppelskalierten Zeitreihenvergleich gegenübergestellt, weil dieser am besten die Reaktion der steuerbaren Größe auf die andere – also z.B. des Wareneinkaufs auf den Absatz – wiedergibt (s. den Vergleich von drei Zeitreihenvarianten im Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)). Damit wird bereits auf der hochverdichteten Quartalsebene der Fokus auf die sinnvollste Plausibilitätsfrage i.S. des § 158 Abs. 2 Nr. 1 AO gelegt: „Hat die Unternehmensleitung so erfolgsorientiert auf Markteinflüsse wie vor allem die Nachfrage *reagiert*, dass das Gesamtbild der Besteuerungsdaten keinen Anlass bietet, an ihrer sachlichen Richtigkeit zu zweifeln?“



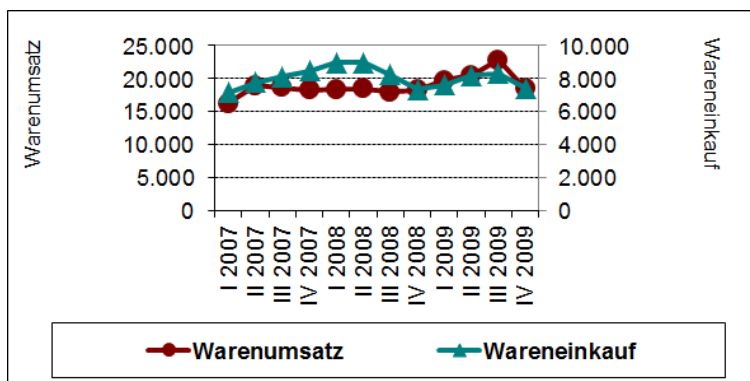
Für den Warenfluss³⁶⁸ der exemplarisch abgebildeten Kneipe sind der oberen Grafik keine jahreszeitlichen „Wellenbewegungen“ des Aufschlagsatzes zu entnehmen, die unterjährig auf erhebliche Veränderungen des Getränkesortiments mit abweichenden Margen hinweisen würden. In der Summenbetrachtung der Vielzahl von Warenumsätzen und -einkäufen pro Quartal, infolge welcher die Einzelheiten größtenteils „abgeschliffen“ werden, ergeben sich **Aufschlagsätze mit einer nur geringen Schwankungsbreite**.

Dem streng chronologischen Zeitreihenvergleich der zweiten Abb. (mit automatisch zuordnender Y-Achsenbeschriftung) sind jedes Jahr saisonartige Zu- und Abnahmen des Wa-

³⁶⁸ Auf dem Tabellenblatt „Jahreszahlen“ eingegebene Entnahmen können – wie im Beispiel – über die Ankreuzfelder auf den Tabellenblättern „Quartalsrelationen“ und „Monatszeitreihen“ einheitlich und gleichmäßig auf die (vorhandenen) Perioden umverteilt werden.

rendurchflusses zu entnehmen, die den Aufschlagsatz aber offensichtlich nicht erheblich beeinflussen. Dabei folgt der Getränkenachkauf dem Absatz weitgehend in Richtung und Intensität; auch die verzögerte Aufzeichnung des Wareneinkaufs der vorliegenden Einnahmen-Überschuss-Rechnung gem. § 4 Abs. 3 EStG hat den Zusammenhang der Betriebsfaktoren in dem robusten Zeitreihenmodell nicht stören können. Dagegen tritt die **Einzelablösung IV/2007** hervor, wo der (relativ) höhere Einkauf auch im Folgequartal offenbar nicht (vollständig) verbraucht worden ist³⁶⁹. Die singuläre Auffälligkeit kann nach dem Prüfungsduktus der SRP-Vorlage BWA systematisch nachverfolgt werden (s. den Abschnitt „Eingrenzung aufgefundener Regelstörungen zu ursächlichen Sachverhalten“ unter [Pkt. 4.2.2.](#)), um den verursachenden Vorgang (rechtlich) zu würdigen und gegebenenfalls für die weitere Gesamtbildprüfung interaktiv (s. [Pkt. 4.2.9.1.](#)) auf Normalverhältnisse zu berichtigen. *Durch die sich anschließende Intensivierung der Zeitreihenanalysen auf die empfindlichere Monatsbasis wird eine annehmbare Gründlichkeit bei der Aufdeckung betriebsspezifischer Prüffelder erreicht.*

Offensichtliche Regelablösungen eines doppelskaliert größenangepassten Quartals-



zeitreihenvergleichs wie im nebenstehenden Beispiel eines Cafés mit langfristig erheblich verstärktem Wareneinkauf begründen schwerwiegende Zweifel an der inhaltlichen Richtigkeit der untersuchten Besteuerungsdaten, weil die üblichen Erklärungsansätze³⁷⁰ zur Zerstreuung von Zeitreihen auffälligkeiten nicht geeignet sind:

! Einerseits ist das Modell des doppelskaliert größenangepassten Zeitreihenvergleichs gegenüber Zuordnungsunschärfen stabil, sodass ein bestehender Zusammenhang wie bei Warenflussdaten selbst für lange haltbare Waren oder trotz der Einnahmen-Überschuss-Rechnung gem. § 4 Abs. 3 EStG weiterhin problemlos zu erkennen bleibt (s. das Beispiel oben zu diesem Pkt. bzw. die Monatszeitreihe im Abschnitt „Robustheit des doppelskalierten Zeitreihenvergleichs“ unter [Pkt. 4.2.](#)).

! Einerseits ist das Modell des doppelskaliert größenangepassten Zeitreihenvergleichs gegenüber Zuordnungsunschärfen stabil, sodass ein bestehender Zusammenhang wie bei Warenflussdaten selbst für lange haltbare Waren oder trotz der Einnahmen-Überschuss-Rechnung gem. § 4 Abs. 3 EStG weiterhin problemlos zu erkennen bleibt (s. das Beispiel oben zu diesem Pkt. bzw. die Monatszeitreihe im Abschnitt „Robustheit des doppelskalierten Zeitreihenvergleichs“ unter [Pkt. 4.2.](#)).

³⁶⁹ Besonders in der Literatur (vgl. z.B.: Wigger, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168, 171: „Von großer Bedeutung für die Qualität und Beweiskraft des ZRV ist, dass der Schritt zum zutreffenden ‚Periodenwareneinsatz‘ gelingt.“) wird vor Zuordnungsunschärfen in Zeitreihenanalysen vehement gewarnt, obwohl die verbildlichte Schlüssigkeitsprüfung gerade nicht in der Beurteilung eines einzelnen Diagrammpunkts besteht, sondern der chronologische Gesamteindruck vor solchen Fehlschlüssen schützt.

³⁷⁰ Vgl. z.B.: BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 47: „Dies zeigt, dass bereits eine Veränderung in der Zuordnung der Einkäufe nur einer einzigen Woche erhebliche Auswirkungen auf das Gesamtergebnis der Zehn-Wochen-Periode hat.“

- ! Andererseits sind Quartalssummen sehr stark zusammengefasste – und damit robuste – Bezugsparameter, wodurch mögliche Zuordnungsfehler „kleingerechnet“ werden.³⁷¹
- ! Die – in der Literatur gefürchteten³⁷² – hochwertigen Bestandseinkäufe mit verzögertem Verbrauch bewirken höchstens punktuelle Ablösungen im Gleichlauf der Grafen „Warenumsatz“ und „Wareneinkauf“. Nach der SRP-Zeitreihensystematik werden derartige Einzelauffälligkeiten schrittweise eingegrenzt und bis zu den verursachenden Vorgängen nachverfolgt (s. [Pkt. 4.2.2.](#)). Damit können die aufgefundenen, gesamtbildrelevanten Sachverhalte nachvollziehbar interaktiv (s. [Pkt. 4.2.9.1.](#)) auf Normalverhältnisse korrigiert werden, sodass sie das Schlüssigkeitsurteil i.S. des § 158 Abs. 2 Nr. 1 AO nicht beeinflussen.
- ! Darüber hinaus zeigt die ständige Prüfungspraxis, dass offenkundige Implausibilitäten in Bestimmungsaufzeichnungen (nahezu) nie losgelöst von Formalmängeln und anderen materiellen Indizien vorkommen.³⁷³ Darum wird die Beurteilung der Aussagekraft einer so auffälligen Zeitreihenauswertung wie im Beispielsfall regelmäßig im Kontext von Indizien Gesamtheiten³⁷⁴ erfolgen müssen. Hierbei können die interaktiven (s. [Pkt. 4.2.9.](#)) Funktionen der SRP-Vorlage BWA praktikabel zur Einarbeitung substantiierter Einlassungen i.R. der gemeinsamen Aufklärungsverantwortung³⁷⁵ genutzt werden; auf die

³⁷¹ Wolenski, „Das Zeitreihenurteil des X. Senats des BFH vom 25. 3. 2015“, Stbg 6/2016, 268, mit einem anschaulichen Rechenbeispiel auf S. 269 zur Bedeutung der Größe der jeweiligen Bezugssummen (Wochen, Monate, 10-Wochen-Intervalle, Quartale etc.) auf die Störproblematik.

³⁷² Wiggen, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168, 171: „Von großer Bedeutung für die Qualität und Beweiskraft des ZRV ist, dass der Schritt zum zutreffenden ‚Periodenwareneinsatz‘ gelingt“; bezugnehmend darauf: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 27: „Die Qualität des Zeitreihenvergleichs wird maßgeblich dadurch bestimmt, wie sorgfältig die betriebswirtschaftlichen Größen erhoben werden und wie genau betriebliche Besonderheiten geprüft werden. [...] Dies ist keinesfalls ein Argument gegen die Methoden, sondern zeigt lediglich die bei jeder Verprobungsmethode obligatorische Verpflichtung zur sorgfältigen Anwendung ebendieser.“

³⁷³ Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstrafverfahren“, NZWiSt 1/2014, 24, 25: „In der Praxis wird sich die Frage, ob ein Zeitreihenvergleich allein die Beweisvermutung des § 158 AO widerlegt, [...] dadurch relativieren, dass die [...] aufgedeckten Unregelmäßigkeiten regelmäßig mit Unregelmäßigkeiten in der Kassenbuchführung einhergehen.“

³⁷⁴ BGH v. 27.4.2010 – 1 StR 454/09, NStZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreicht.“

³⁷⁵ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen.“

Weise lassen sich reine Schutzbehauptungen von tatsächlichen außergewöhnlichen Umständen effektiv und zügig unterscheiden.

4.2.5. Monatszeitreihen

Dass **monatliche Summen** zur Bewertung von betrieblichen Entwicklungen und Abhängigkeiten

- ✓ auf der einen Seite eine ausreichende Robustheit – also geringe Störanfälligkeit gegenüber eher alltäglichen Sondersituationen – haben und
- ✓ auf der anderen Seite empfindlich genug zum Aufzeigen wichtiger Veränderungen und Einflüsse sind,

wird aus der seit Jahrzehnten üblichen Praxis von Unternehmensplanung und -beratung ersichtlich, Monats-BWAs wichtigen Entscheidungen zu Grunde zu legen. Werden die monatlich aggregierten Betriebsdaten zusätzlich durch eine zielführende Zeitreihenvariante verbildlicht, können existentielle Schlussfolgerungen z.B. zu Investitionen verantwortungsvoll abgeleitet werden. Folgerichtig muss eine geeignete (s. [Pkt. 4.2.](#)) Zeitreihenanalyse der Monatssummen betrieblicher Daten grundsätzlich auch für eine Plausibilitätsbetrachtung i.S. des § 158 Abs. 2 Nr. 1 AO in Betracht kommen – insbesondere wenn sie gegen ungerechtfertigte Beanstandungen durch Maßnahmen wie die Prüfungssystematik „vom groben Überblick zu den Details“ (s. [Pkt. 4.2.2.](#)) und interaktive Optionen (s. [Pkt. 4.2.9.](#)) zur Bereinigung von gesamtbildrelevanten Sondersituationen abgesichert wird.

Auf dem Tabellenblatt „Monatszeitreihen“ der SRP-Vorlage BWA werden **drei Zeitreihenvarianten** entsprechend ihrer unterschiedlichen Vorzüge für bestimmte Zwecke inklusive zielgerichtet unterstützende Funktionalitäten zur Verfügung gestellt (s.a. *die jeweiligen Beispiele unter [Pkt. 4.2.5.1.](#), [4.2.5.2.](#) u. [4.2.5.3.](#)*):

1. Als zentrales SRP-Zeitreihenmodell gibt der doppelskaliert größenangepasste Zeitreihenvergleich die Abhängigkeit zweier Datenreihen auch unter schwierigen Bedingungen wie z.B. der Einnahmen-Überschuss-Rechnung nach § 4 Abs. 3 EStG oder bei längerer Haltbarkeit von Waren korrekt wieder (s. *die entsprechenden Beispiele unter [Pkt. 4.2.](#)*). Das statistische Maß der Gleichrichtung „Korrelationskoeffizient“³⁷⁶ kann zur Gegenkontrolle des visuellen Eindrucks für beliebige zusammenhängende Zeiträume ausgegeben werden. Zudem bietet die Funktion der Konjunkturanpassung die Möglichkeit, einseitige Preiseffekte mit gesamtbildrelevanter Bedeutung interaktiv auszugleichen, um zwischen extremen Markteinflüssen und unschlüssigem Handeln unterscheiden zu können.

³⁷⁶ Als lineares Abhängigkeitsmaß nimmt der Korrelationskoeffizient Größen von „-1“ für die 100 %-ige Gegenläufigkeit der Grafen über „0“ bei nicht vorhandener Beeinflussung bis zu „+1“ für eine vollständig (positive) Gleichrichtung ein.

2. Der anschließende einzelskalierte Zeitreihenvergleich wird besonders für Übereinstimmungsprüfungen (s. [Pkt. 5.1.2.](#)) benötigt, wenn bspw. Datenschnittstellen auf Fehlerfreiheit untersucht oder Steuerungsdaten anhand externer Datenbanken abgeglichen werden sollen. Dazu werden die monatlichen und jährlichen Differenzen der beiden Datenreihen zusätzlich in einer Tabelle dargestellt.
In einigen Fällen kann dieser Zeitreihentyp für eine Plausibilitätsprüfung ergänzend hilfreich sein, weil die unternehmerische Bedeutung einer Regelablösung klarer wird. Außerdem lässt sich dieser Zeitreihenvergleich auf relative³⁷⁷ Darstellung inklusive der Funktion umstellen, die erste Datenreihe in der Abbildungshöhe an die zweite manuell anzupassen, denn durch die sekundäre Y-Achse des doppelskalierten Zeitreihenvergleichs wird nicht immer automatisch die perfekte Größenanpassung erreicht.
3. Durch die abschließende (faktorbezogene) Zeitreihenanalyse der Verhältnisgröße³⁷⁸ aus den Monatssummen beider Datenreihen – unter Einbindung der Einzelgrafiken – soll das Gesamtbild abgerundet werden. Die Schlüssigkeit von Steuerungsgrundlagen gem. § 158 Abs. 2 Nr. 1 AO wird i.R. der SRP *nicht* anhand von **Aufschlagsatzentwicklungen** beurteilt (s.a. den Abschnitt „SRP-Zeitreihe systematik – Analysegrundsätze und Annahmen“ unter [Pkt. 4.2.11.](#)); demzufolge wird die farbliche Hervorhebung von Werten, welche die unteren Richtsätze vom Tabellenblatt „Jahreszahlen“ unterschreiten, hier nicht fortgeführt.

4.2.5.1. Doppelskaliert größenangepasster Zeitreihenvergleich

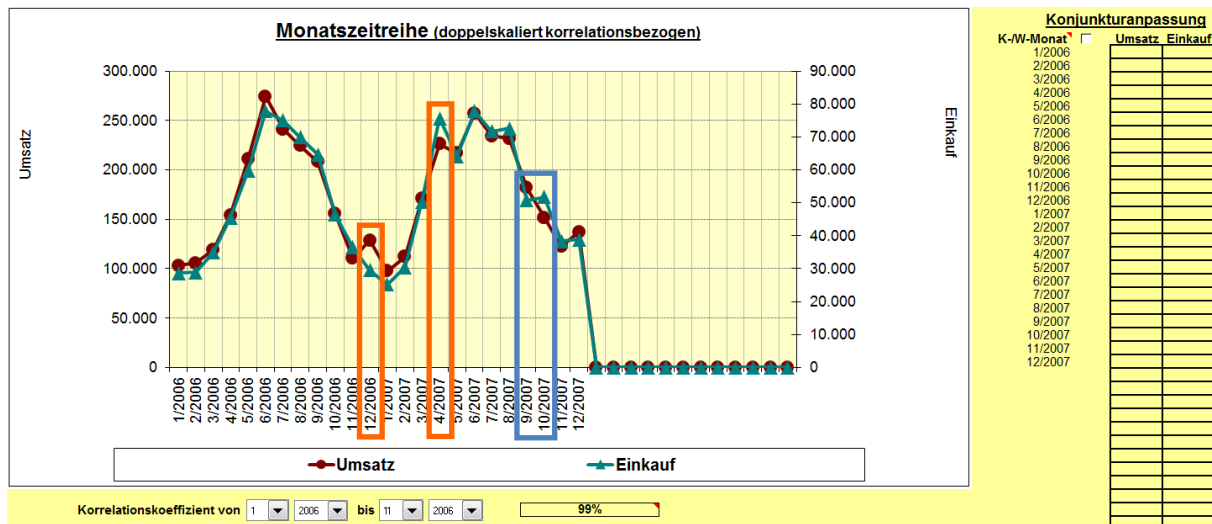
Aufgrund seiner **Robustheit gegenüber möglichen Störeinflüssen** (s. die Beispiele zur Einnahmen-Überschuss-Rechnung und zu lange haltbaren Waren unter [Pkt. 4.2.](#)) ist der doppelskaliert größenangepasste Zeitreihenvergleich i.R. der SRP das zentrale Mittel zur Plausibilitätsbeurteilung von Steuerungsgrundlagen. Dazu werden zunächst eventuelle betriebsspezifische Prüffelder mit ausreichender Gründlichkeit entlang der SRP-Zeitreihe systematik (s. [Pkt. 4.2.2.](#)) aufgedeckt, um die Beurteilungsgrundlage für die Schlüssigkeit nachvollziehbar von Störfaktoren zu säubern.

Dieser Gegenüberstellung von monatlichem Warenumsatz und -einkauf eines größeren gastronomischen Betriebs (mit automatisch zuordnender Y-Achsenbeschriftung) ist der enge Reaktionszusammenhang³⁷⁹ klar zu entnehmen:

³⁷⁷ Die Monatswerte werden als Anteile der jeweiligen Jahressumme berechnet.

³⁷⁸ Die Berechnung des Verhältnissfaktors kann auf dem Tabellenblatt „Jahreszahlen“ per Ankreuzfeld als Aufschlagsatz mit „ $\frac{(\text{Datenreihe 2} - \text{Datenreihe 1})}{\text{Datenreihe 1}}$ “ oder als Quotient $\frac{\text{Datenreihe 2}}{\text{Datenreihe 1}}$ ausgewählt werden; die Ergebnisse werden prozentual ausgegeben.

³⁷⁹ FG Hamburg v. 31.10.2016 – 2 V 202/16, Rz. 48: „Bei dieser Methode [doppelskalierte Monatszeitreihe] tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander. Aus kaufmännischer Sicht wird in einem Unternehmen auf jede Verände-



Durch den Korrelationskoeffizienten³⁸⁰ unterhalb der Grafik wird mit 99 % für den eingestellten Teilzeitraum „Januar bis November 2006“ (bzw. 98 % für den gesamten Zeitraum) das visualisierte Ergebnis bestätigt. Entgegen Warnungen aus der Literatur³⁸¹ sind Ablösungen infolge von **offenbaren Zuordnungsfehlern** wie im Beispiel der Monate „9-10/2007“ gut zu erkennen, weil die Interpretation von Zeitreihenanalysen gerade nicht auf einzelne Punkte sondern die (gemeinsame) Entwicklung der Grafen abzielt. Ebenso können dem Diagramm prüfungsrelevante **Regelstörungen der Monate „12/2006“ und „4/2007“** entnommen werden.

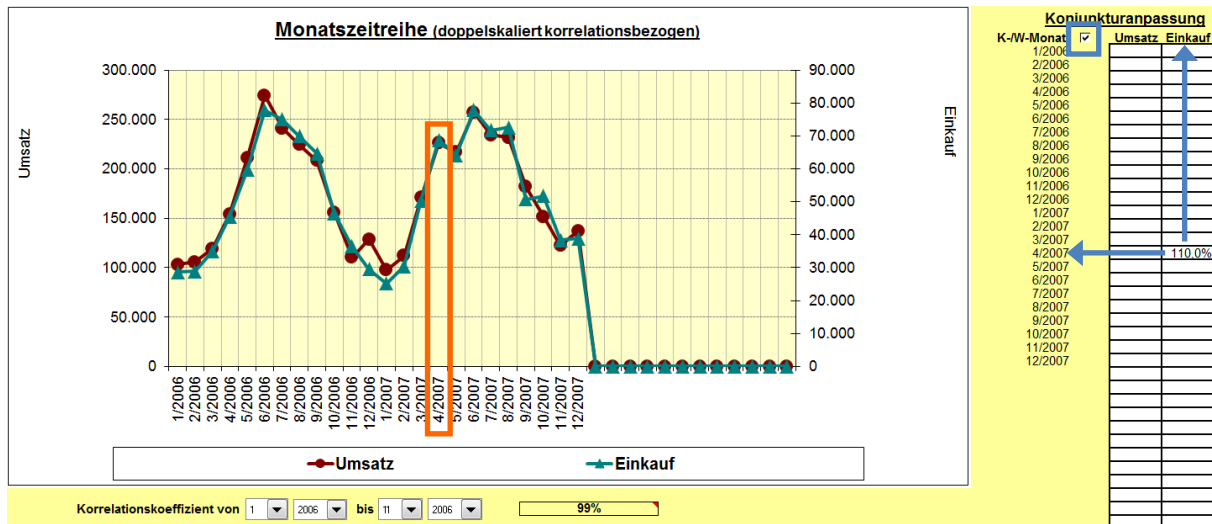
Die Monatszeitreihen haben – nach dem groben Überblick der Quartalszeitreihen für ein grundsätzliches Verständnis von Entwicklungen, Schwankungen und Abhängigkeiten in Besteuerungsdaten (s. [Pkt. 4.2.4.](#)) – eine Zwischenposition hinsichtlich der **erforderlichen Tiefe (Gründlichkeit)** einer Prüfung. Vor Beurteilung der Schlüssigkeit i.S. des § 158 Abs. 2 Nr. 1 AO werden entdeckte Einzelauffälligkeiten bis zu möglichen Verursachern zurückverfolgt. Dazu verfügt die SRP-Vorlage BWA über wichtige Funktionen (s.a. [Pkt. 4.2.9.](#)) wie die **Konjunkturbereinigung**: So hat im Beispielsfall die gezielte Untersuchung der **Regelstörung „4/2007“** ergeben, dass in dem Monat umfangreich Waren ausnahmsweise so viel teurer eingekauft werden mussten, dass dadurch ein **Gesamtpreisanstieg** von 10 % für den Mo-

nung der Nachfragesituation (annähernd) adäquat mit einer Erhöhung bzw. Absenkung der Einkaufssumme reagiert“.

³⁸⁰ Als Maß für den Zusammenhang zweier Datenreihen nimmt der Korrelationskoeffizient Größen von „-1“ für die 100 %-ige Gegenläufigkeit der Grafen über „0“ bei nicht vorhandener Abhängigkeit bis zu „+1“ für eine vollständig positive Gleichrichtung ein.

³⁸¹ Wigen, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168, 171: „Von großer Bedeutung für die Qualität und Beweiskraft des ZRV ist, dass der Schritt zum zutreffenden ‚Periodenwareneinsatz‘ gelingt“; bezugnehmend darauf: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 27: „Die Qualität des Zeitreihenvergleichs wird maßgeblich dadurch bestimmt, wie sorgfältig die betriebswirtschaftlichen Größen erhoben werden und wie genau betriebliche Besonderheiten geprüft werden. [...] Dies ist keinesfalls ein Argument gegen die Methoden, sondern zeigt lediglich die bei jeder Verprobungsmethode obligatorische Verpflichtung zur sorgfältigen Anwendung ebendieser.“

nat resultierte. Über einen entsprechenden [Preisindex](#) lässt sich dieser außergewöhnliche Sachverhalt umgehend auf Normalverhältnisse berichtigen. Dabei kann die Konjunkturbereinigung zur besseren Nachvollziehbarkeit – besonders von umfangreicheren Korrekturen (s. *das Beispiel „Kostenzeitreihe zum Kraftstoffverbrauch eines Taxiunternehmens“ unter Pkt. 4.2.9.4.*) – über das [Anhakfeld](#) komplett an- und abgeschaltet werden.



Nach der interaktiven Einarbeitung der neuen Erkenntnis zeigt sich im Beispiel umgehend, dass der besondere Umstand offenbar die (relative) Ablösung des Warennachkaufs vom Absatz bewirkt hat. Durch das Einpflegen in die Gesamtbildprüfung kann der Störeinfluss technisch problemlos **egalisiert** werden.

Bei **offensichtlichen Regelablösungen** in einem doppelskaliert größenangepassten Monatszeitreihenvergleich ergeben sich aufgrund der folgenden Methodenmerkmale (leicht abgeschwächt) ähnliche Zweifel an der sachlichen Richtigkeit wie bei der Quartalsvariante (*ausführlicher im Abschnitt „offensichtliche Regelablösungen“ unter [Pkt. 4.2.4.](#)*):

- ! die hohe Stabilität des verwendeten Zeitreihenmodells für die Prüfung von Datenreihen auf Abhängigkeit (s. [Pkt. 4.2.](#)),
- ! die ausreichende Robustheit von Monatswerten z.B. gegenüber Wochensummen³⁸² u.a. für übliche Planungen in der betriebswirtschaftlichen Beratung,
- ! die gründliche Aufdeckung, Nachverfolgung und mögliche Bereinigung von gesamtbildrelevanten Einzelstörfaktoren wie bspw. von Bestandseinkäufen i.R. des Warenflusses gem. SRP-Zeitreihensystematik (s. [Pkt. 4.2.2.](#)) sowie
- ! die ständigen Prüfungserfahrungen, dass gravierende Regelstörungen in Zeitreihenvergleichen nicht losgelöst von Formal- und anderen materiellen Mängeln vorkommen³⁸³,

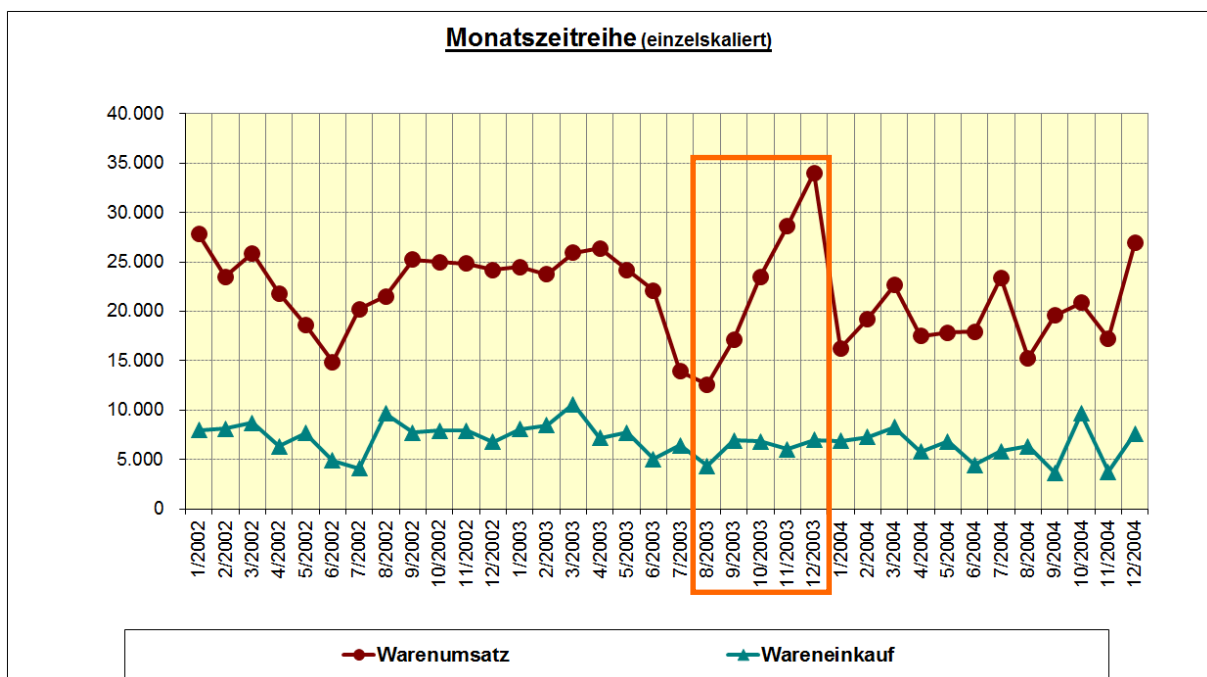
³⁸² Wolenski, „Das Zeitreihenurteil des X. Senats des BFH vom 25. 3. 2015“, Stbg 6/2016, 268, mit einem anschaulichen Rechenbeispiel auf S. 269 zur Bedeutung der Größe der jeweiligen Bezugssummen (Wochen, Monate, 10-Wochen-Intervalle, Quartale etc.) auf die Störproblematik.

zuzüglich der vielfältigen Optionen, konkrete Erklärungen zügig interaktiv (s. [Pkt. 4.2.9.](#)) daraufhin abzuklären, ob sie ursächlich oder reine Schutzbehauptungen sind.

4.2.5.2. Einzelskalierter Zeitreihenvergleich

Mit der zweiten Zeitreihenanalyse auf dem Tabellenblatt „Monatszeitreihen“ werden die beiden Datenreihen in einem einzelskalierten Zeitreihenvergleich gegenübergestellt. Diese Darstellungsform ist üblicherweise dann zielführend, wenn die **Übereinstimmung** der Periodengrößen wie bei der Überprüfung von Datenschnittstellen (s. [Pkt. 5.1.2.](#)) oder der Kontrolle von Steuerungsdaten z.B. mit Hilfe externer Informationen vom Zoll erwartet wird.

In einigen Fällen lässt sich das **Maß an Unschlüssigkeit** beobachteter Regelstörungen in einem zusätzlichen Zeitreihenvergleich unter gemeinsamer Skalierung der Unternehmensfaktoren – trotz der schlechteren Vergleichbarkeit der Einzelentwicklungen – noch deutlicher herauslesen, wie in diesem Beispiel zum Warenfluss eines Restaurants:

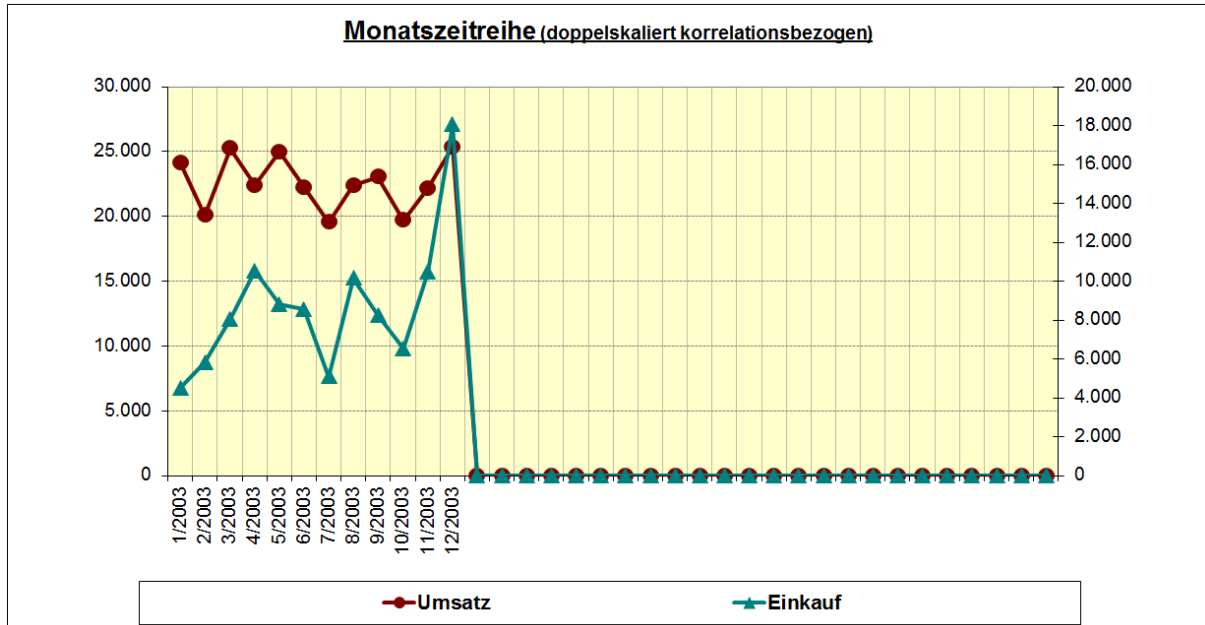


Besonders der Zeitraum „8-12/2003“ offenbart eine extreme Absatzzunahme, wofür jedoch keine nennenswerte Erhöhung des Warennachkaufs nötig gewesen sein soll. Da der Geschäftsablauf – infolge begrenzter Haltbarkeit von Lebensmitteln auch unter Einsatz von Warenbeständen – so unmöglich stattgefunden haben kann, ist es „unwahrscheinlich [...], dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt“³⁸⁴.

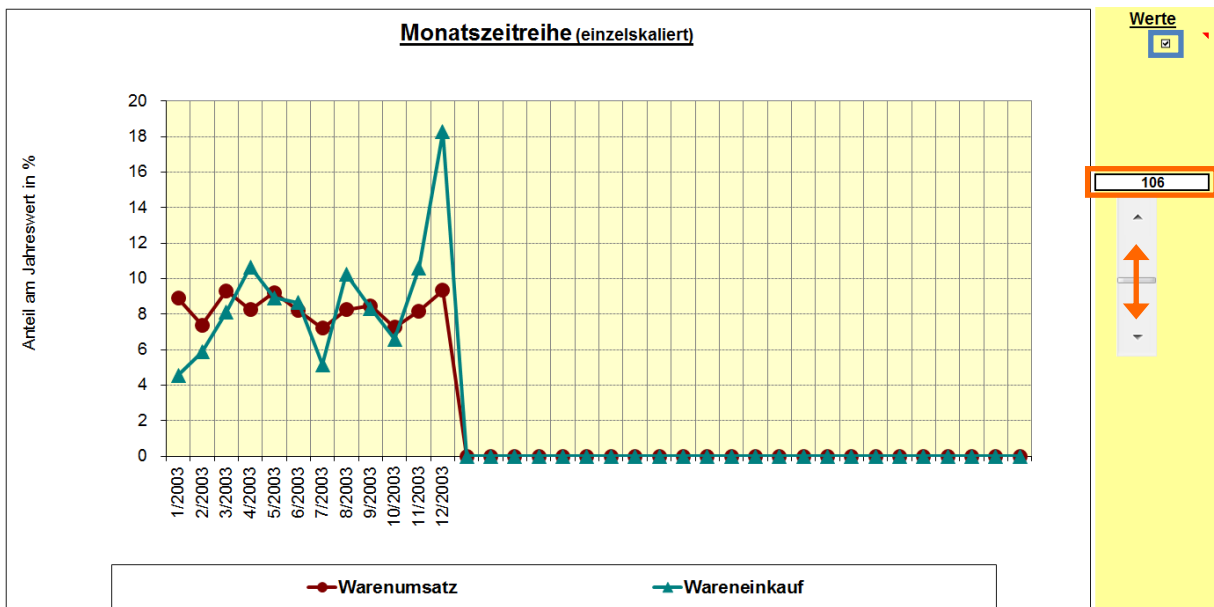
³⁸³ Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstraßverfahren“, NZWiSt 1/2014, 24, 25: „In der Praxis wird sich die Frage, ob ein Zeitreihenvergleich allein die Beweisvermutung des § 158 AO widerlegt, [...] dadurch relativieren, dass die [...] aufgedeckten Unregelmäßigkeiten regelmäßig mit Unregelmäßigkeiten in der Kassenbuchführung einhergehen.“

³⁸⁴ Anwendungserlass zu § 158 AO.

Mitunter können unzusammenhängende Ausreißer *einer*³⁸⁵ Datenreihe auch zu **verzerrten Y-Achsen-Skalierungen** wie in der folgenden Abb. durch den Wareneinkauf „12/2003“ eines Dönerimbisses führen:



Hierdurch wird die Vergleichbarkeit der Entwicklungen beider Datenreihen trotz Doppelskalierung verhindert. Neben der Nachverfolgung und Berichtigung der ursächlichen Sachverhalte gem. SRP-Prüfungssystematik (s. [Pkt. 4.2.2.](#)) bietet die SRP-Vorlage BWA zur schnellen Orientierung eine variabel-relative Form des einzelskalierten Zeitreihenvergleichs an:



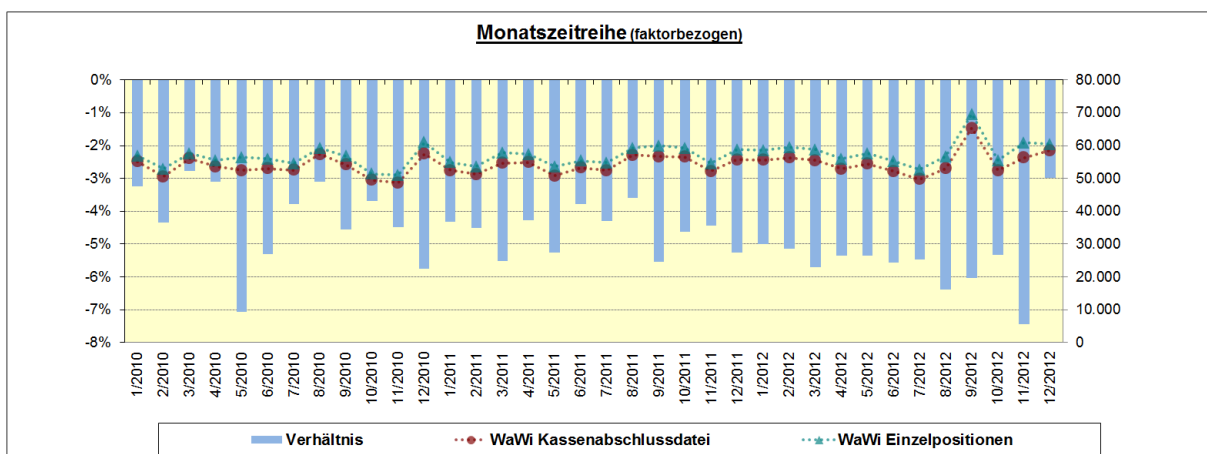
Über die [Umschaltfunktion](#) werden die Monatswerte für Warenumsatz und -einkauf als Anteile der jeweiligen Jahressummen ausgegeben. Nun kann die erste Datenreihe vom Tabellen-

³⁸⁵ Sofern beide Datenreihen in ihren Monatssummen stark schwanken, dabei aber die erwartete schlüssige Reaktion aufeinander zeigen, wird die Größenanpassung via Doppelskalierung nicht gestört.

blatt „Vergleichsdaten“ (für den Wareneinkauf) mittels Schieberegler auf das Größenniveau des zweiten Grafen gezogen (projiziert) werden: in feinen Stufen über die Pfeile und in größeren über die Schienen; im Feld über dem Schieberegler wird der aktuelle Anpassungsstand im Verhältnis zu 100 angezeigt. Für den Beispielsfall offenbart sich auf diese Weise ohne langwierige Vertiefungsmaßnahmen, dass die Regelstörungen (nahezu) das komplette Jahr betreffen. Insofern müsste eine erfolgreich aufklärende Begründung für den gebündelt erfassten Wareneinkauf in 12/2003 nachweisen, dass ein nicht unwesentlicher Anteil vom Jahresanfang stammt. Ein kompletter zeitlicher Versatz der Datenreihe „Wareneinkauf“ wegen verzögerter Erfassung der Einnahmen-Überschuss-Rechnung nach § 4 Abs. 3 EStG lässt sich weder in der Monats- noch in der vorgeschalteten, übersichtlicheren Quartalsansicht erkennen. Folglich kann die Unschlüssigkeit i.S. des § 158 Abs. 2 Nr. 1 AO nur auf materielle Unrichtigkeit der Besteuerungsgrundlagen oder so gravierende Ordnungsmängel zurückgehen, dass diese für sich bereits die Beweiskraft der Aufzeichnungen widerlegen.

4.2.5.3. Verhältniszeitreihe (z.B. zum Aufschlagsatz)

Die (faktorbezogene) Zeitreihenanalyse der Verhältnisgröße³⁸⁶ aus den Monatssummen der Datenreihen – einschließlich beider Einzelgrafien – kann zuvor gewonnene regelbasierte Eindrücke sinnvoll ergänzen, wie bei der visualisierten Systemprüfung eines Warenwirtschaftsprogramms unter [Pkt. 5.1.2.](#):



Zu den geringen, sehr gleichmäßigen Differenzen zwischen den Grafen „Einzelpositionen“ und „(systeminterne) Kassenabschlüsse“ berechnen sich dennoch keine konstanten Abschlagsverhältnisse, wie es bei einer prozentualen Manipulationseinstellung auf alle Verkäufe geschehen wäre. Diese – dank der Verbildlichungsweise offensichtliche – Erkenntnis hilft bei der Eingrenzung der Ursachen für die Regelstörung: sowohl im Hinblick auf mögliche Einstellungen im Programm als auch hinsichtlich anderer Hinterziehungsansätze.

³⁸⁶ Die Berechnung des Verhältnisfaktors kann auf dem Tabellenblatt „Jahreszahlen“ per Ankreuzfeld als Aufschlagsatz mit „ $\frac{\text{Datenreihe 2} - \text{Datenreihe 1}}{\text{Datenreihe 1}}$ “ oder als Quotient $\frac{\text{Datenreihe 2}}{\text{Datenreihe 1}}$ ausgewählt werden; die Ergebnisse werden prozentual ausgegeben.

I.R. der SRP wird die Schlüssigkeit von Besteuerungsgrundlagen gem. § 158 Abs. 2 Nr. 1 AO nicht³⁸⁷ anhand von **Aufschlagsatzentwicklungen** beurteilt (s.a. den Abschnitt „SRP-Zeitreihensystematik – Analysegrundsätze und Annahmen“ unter [Pkt. 4.2.11.](#)); demzufolge wird die farbliche Hervorhebung von Werten, welche die unteren Richtsätze vom Tabellenblatt „Jahreszahlen“ unterschreiten, hier nicht fortgeführt. Zum einen wird der externe Vergleich wegen seiner begrenzten Übertragbarkeit auf den Einzelfall in der AO-Kommentierung als „größte Methode“³⁸⁸ bezeichnet. Zum anderen zeigen Aufschlagsatzzeitreihen³⁸⁹ in der Prüfungspraxis im Gegensatz zu geeigneteren Modellen eine deutliche Anfälligkeit gegenüber eher alltäglichen Störeinflüssen (s. den *Modellvergleich im Abschnitt „Methodenkenntnis“* unter [Pkt. 4.2.](#)).

4.2.6. Gleitschlittenzeitreihen³⁹⁰

Gleitende Zeitreihen können die Robustheit längerer Vergleichsperioden mit der Detailliertheit von kürzeren bedingt³⁹¹ verbinden, indem sie ein für Störeinflüsse wenig anfälliges Intervall von bspw. 2 Monaten immer nur um einen Tag fortschreiben: Auf die Startperiode „1.1. bis 28.2.“ folgt „2.1. bis 1.3.“, „3.1. bis 2.3.“, „4.1. bis 3.3.“ etc. Einerseits werden dadurch nur Regelstörungen angezeigt, die so schwerwiegend sind, dass sie einen 2-Monatswert verzerren können. Andererseits bewirken sowohl der Einstieg als auch der Ausstieg einer ursächlichen Besonderheit jeweils eine abrupte Veränderung des betreffenden Grafen: So senkt z.B. ein hochwertiger Bestandseinkauf am 13.4. den 2-Monats-Aufschlagsatz erstmalig für die

³⁸⁷ Insofern muss Kritik an der SRP aus der Literatur, die sich auf Aufschlagsatzzeitreihen bezieht, als gegenstandslos betrachtet werden (vgl. z.B.: Beyer, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, 2933-2934: „Hierzu werden die einzelnen Quartals-RAS in einem Diagramm dargestellt. Schwanken die RAS in den Quartalen, kann dies mit Änderungen der Umsatzträger des Unternehmens zu tun haben“ oder Bleschick, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353: „Zunächst unternimmt er eine sog. betriebswirtschaftliche Auswertung: Hierzu ermittelt er im Wesentlichen mithilfe des Zeitreihenvergleichs einen Rohgewinnaufschlagsatz.“).

³⁸⁸ Tipke/Kruse, AO/FGO, [171. Lieferung 7/2022](#), § 158 AO, Rz. 17: „Als größte Methode zur Beanstandung der sachlichen Richtigkeit des Buchführungsergebnisses erweist sich der äußere Betriebsvergleich [...] Vergleichen mit anderen Betrieben kommt regelmäßig ein geringerer Vergleichswert als inneren Betriebsvergleichen [...] zu“.

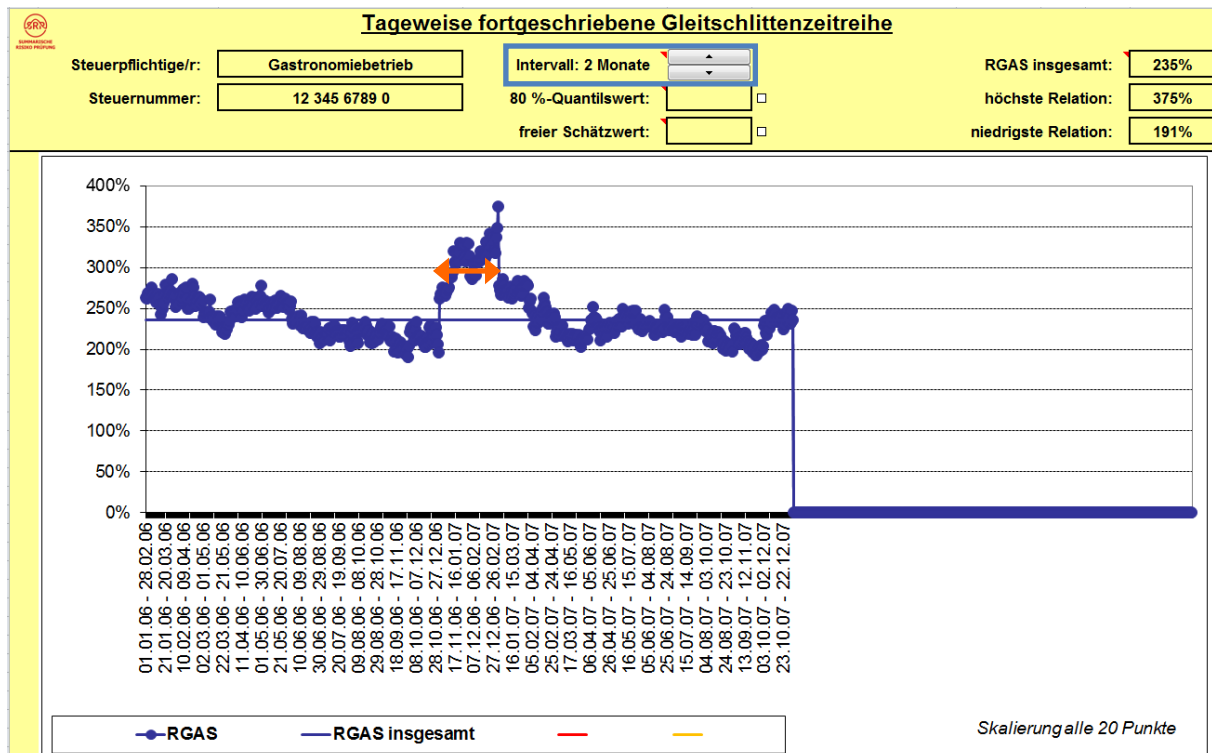
³⁸⁹ Auch dem BFH-Zeitreihenurteil (v. 25.3.2015, BStBl. II 2015, 743) liegt mit der betreffenden „Gleitschlittenzeitreihe“ eine Form der Aufschlagsatzzeitreihe zu Grunde, weshalb die erläuternde Literatur der betreffenden RiBFH z.T. die Anwendbarkeit auf entsprechende Zeitreihenvarianten beschränkt [Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548].

³⁹⁰ Auch „rollierende Zeitreihe“, vgl.: Harle/Olles, „Die moderne Betriebsprüfung“, NWB-Verlag, 3. Auflage 2017, Rz. 1366 ff.

³⁹¹ Die Empfindlichkeit für einzelfallbezogene Prüffelder (Detailliertheit) wird dadurch begrenzt, dass längere Zeiträume mit größeren Summen als rechnerische Bezugsgrößen robuster gegenüber außergewöhnlichen Einzelpositionen sind [vgl. a.: Wolenski, „Das Zeitreihenurteil des X. Senats des BFH vom 25. 3. 2015“, Stbg 6/2016, 268, 269 mit einem Beispiel, wie entscheidend die Vergleichsperiode (Wochen, Monate, 10-Wochen-Perioden, Quartale etc.) für den Einfluss von Störfaktoren ist].

Periode „14.2. bis **13.4.**“ (plötzlich) ab. In den Folgeintervallen „15.2. bis 14.4.“, „16.2. bis 15.4.“ usw. beinhaltet die 2-Monatsgröße weiterhin den Sondersachverhalt, sodass der Aufschlagsatz plateauartig niedrig bleibt, bis er sich mit dem Betrachtungszeitraum „14.4. bis 13.6.“ schlagartig erholt, denn der Bestandseinkauf wird nicht mehr mitgezählt.

Aufgrund dieser besonderen Eigenschaft der Gleitschlittenzeitreihen werden sie i.R. der SRP zur weiteren **Eingrenzung von Regelstörungen** eingesetzt, welche i.d.R. mit doppelskaliert größenangepassten Zeitreihenvergleichen entdeckt worden sind (s. die SRP-Zeitreihensystematik unter [Pkt. 4.2.2.](#)). Als **Standardeinstellung des Gleitschlittens** hat sich aus ständigen Prüfungserfahrungen das Intervall „2 Monate“ herauskristallisiert; über das **Drehfeld** sind jedoch auch die Alternativen „3 Monate“, „10 Wochen“, „6 Wochen“, „1 Monat“ und „14 Tage“ auswählbar.



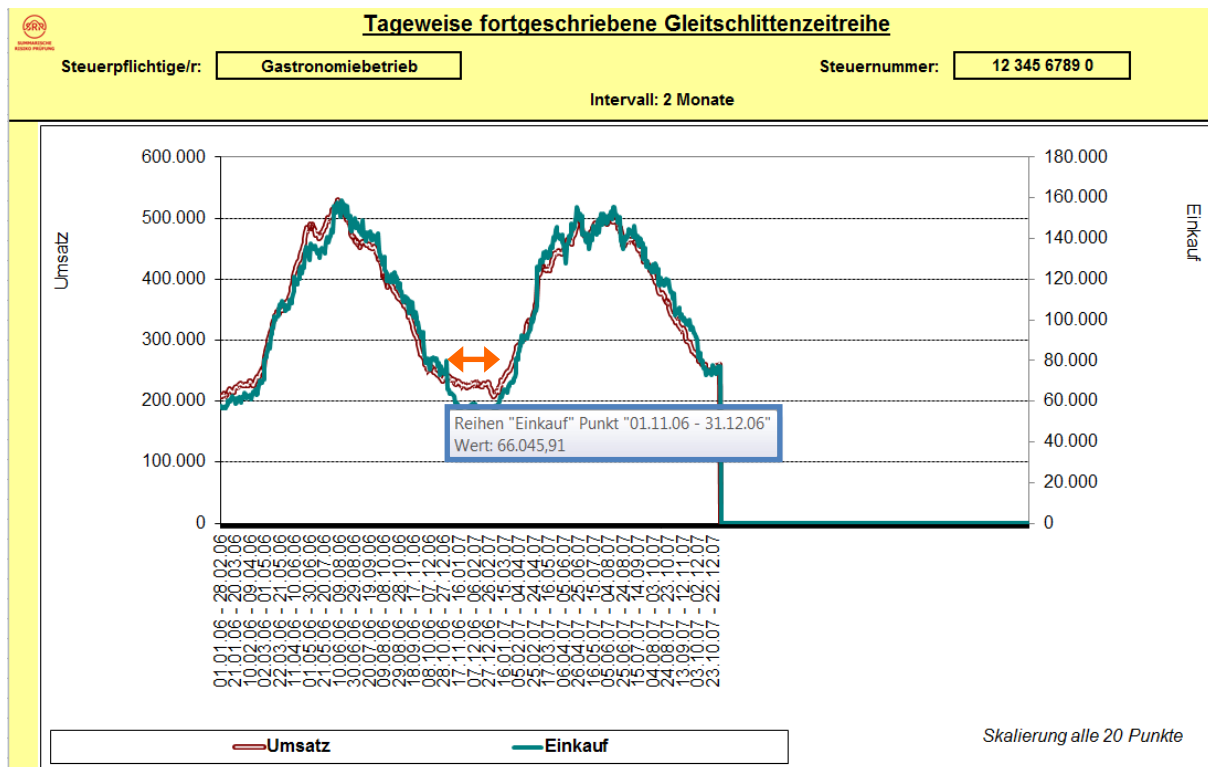
In der Abb. sind die **zwei abrupten Veränderungen der Verhältnissgröße**³⁹², die durch einen einzelnen Tag – und damit vermutlich einen Einzelsachverhalt – verursacht werden, gut zu erkennen. Wenn die Maus im Diagramm entweder auf den Plateaustart oder das -ende gelegt wird, erscheint ein Beschreibungsfähnchen für den jeweiligen Punkt. Damit lässt sich die Verzerrung hier auf das Erstintervall „1.11. bis 31.12.2006“ konkretisieren³⁹³, sodass der ursprüngliche Vorgang am 31.12.2006 zu suchen ist.

³⁹² Die Berechnung des Verhältnissfaktors kann auf dem Tabellenblatt „Jahreszahlen“ per Ankreuzfeld als Aufschlagsatz mit „ $\frac{\text{Datenreihe 2} - \text{Datenreihe 1}}{\text{Datenreihe 1}}$ “ oder als Quotient $\frac{\text{Datenreihe 2}}{\text{Datenreihe 1}}$ ausgewählt werden.

³⁹³ Das vorherige *unbeeinflusste* Intervall „31.10. bis 30.12.2006“ unterscheidet sich ausschließlich im ersten ausgeschlossenen Tag „31.10.2006“ und im letzten hinzugezogenen „31.12.2006“. Da keine abrupte Auf-

Über Filter- und Suchfunktionen können nun die betreffenden „Regelstörer“ zügig identifiziert und geprüft werden. In diesem Fall lässt sich in den verbleibenden Daten des einen Tages unproblematisch ein außergewöhnlich hoher Negativbetrag unter den Wareneinkäufen erkennen. Nach Belegprüfung stellt sich der Vorgang als Jahresbonus heraus, der nicht von den regelmäßigen Wareneinkäufen über ein Extrakonto abgegrenzt worden ist. In der Konsequenz sollte geprüft werden, ob im Folgejahr ebenfalls eine Erfolgsvergütung gezahlt wurde, denn die Gleitschlittenzeitreihe zeigt dort keinen (vergleichbaren) Anstieg. Außerdem muss der Vorgang für die Plausibilitätsbeurteilung des Warenflusses i.S. des § 158 Abs. 2 Nr. 1 AO interaktiv (s. [Pkt. 4.2.9.1.](#)) korrigiert werden, weil die zeitliche Zuordnung zum 31.12. operativ falsch und der Störeinfluss auf die Gesamtbildprüfung offensichtlich ist.

Der **doppelskaliert größenangepasste Zeitreihenvergleich per Gleitschlitten** zu beiden Datenreihen grenzt aufgefundene Regelstörungen auf einzelne Tage und die jeweilige Datenreihe ein. Damit führt er i.R. der SRP-Zeitreihensystematik (s. [Pkt. 4.2.2.](#)) noch direkter zu prüfungsrelevanten Sachverhalten, erfordert allerdings aufgrund seiner komplexen Darstellung mehr Analyseerfahrung.³⁹⁴



schlagsatzveränderung beim Einbeziehen des 31.10.2006 zu sehen ist, muss die Ursache folglich unter den Daten des 31.12.2006 zu finden sein.

³⁹⁴ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. [...] Die korrekte Interpretation solcher Grafiken ist allerdings nicht trivial und erfordert spezielle kognitive Fähigkeiten.“

Zu dem vorherigen Beispielsfall ergibt der differenzierende Zeitreihenvergleich (mit automatisch zuordnender Y-Achsenbeschriftung) in gleitender 2-Monatsform die Abb. oberhalb. Im betreffenden Zeitfenster (um den Monat „12/2006“), in welchem gemäß doppelskaliertem Monatszeitreihenvergleich eine Regelstörung aufgefallen ist, fällt nach dem Mauszeiger-Fähnchen die Datenreihe „Einkauf“ mit dem Intervall „1.11. bis 31.12.“ abrupt und ohne Reaktion des Umsatzes ab. Die korrespondierende Gegenbewegung beim Ausschleichen des 31.12.2006 mit dem Sondersachverhalt wird von dem Anzeigefähnchen verdeckt.

Die Konsequenzen aus der noch etwas konkreteren Identifikation des gesamtbildstörenden Einzelvorgangs erfolgen analog der Vorgehensweise bei der Gleitschlittenzeitreihe zum Aufschlagsatz (s. den Abschnitt „Eingrenzung von Regelstörungen“ unter diesem Pkt.).

Die Schlüssigkeit von Besteuerungsgrundlagen gem. § 158 Abs. 2 Nr. 1 AO wird i.R. der SRP nicht anhand von Aufschlagsatzentwicklungen beurteilt (s.a. den Abschnitt „SRP-Zeitreihensystematik – Analysegrundsätze und Annahmen“ unter Pkt. 4.2.11.).³⁹⁵ Aufschlagsatzvarianten der Gleitschlittenzeitreihe zeigen die übliche Anfälligkeit gegenüber alltäglichen Einflussfaktoren (s. den Vergleich von drei Zeitreihentypen im Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.), und die doppelskalierte Form ist schwer zu lesen. Dementsprechend wirkt die interaktive Konjunkturbereinigung des Tabellenblatts „Monatszeitreihen“ (s. Pkt. 4.2.5.1.) nicht auf die Gleitschlittenmodelle durch.

Die Anhak-Optionen, mit denen sich unterhalb der Intervalleinstellung die Schätzungsparameter „80 %-Quantil“ und „freier Schätzwert“ in der Aufschlagsatz-Gleitschlittenzeitreihe einblenden lassen (s. die erste Abb. unter diesem Pkt.), dienen ausschließlich der optischen Überprüfung, ob eine mittels SRP festgelegte Schätzungshöhe am Maßstab der innerbetrieblichen Entwicklung sachgerecht erscheint, und werden unter Pkt. 4.8. beschrieben.

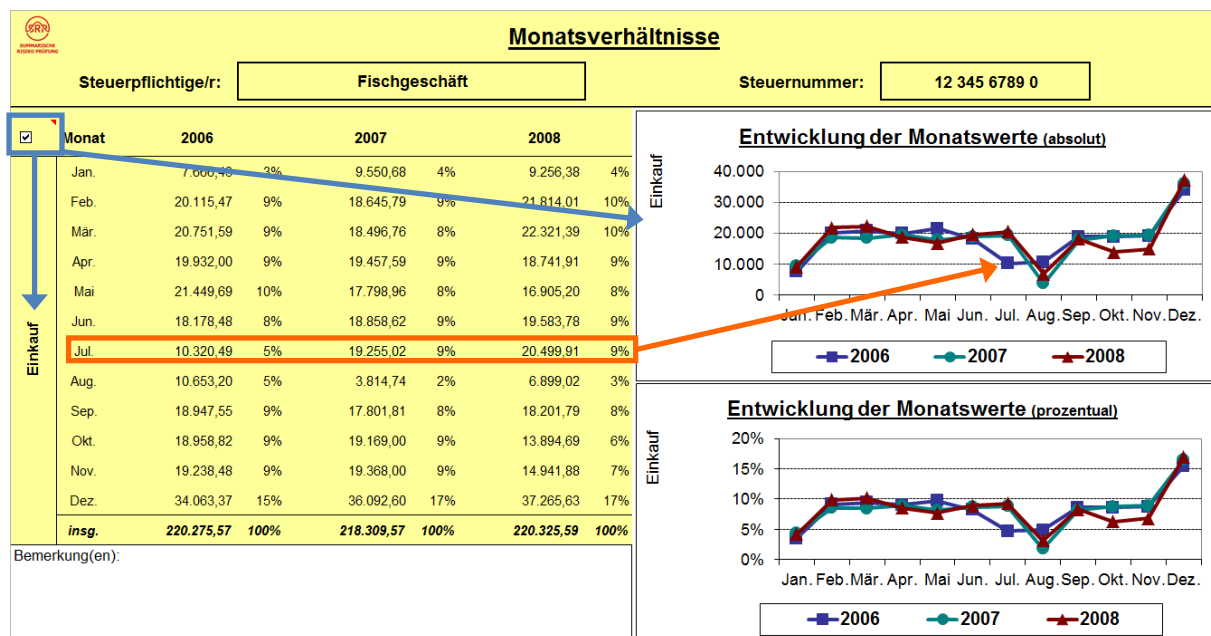
³⁹⁵ Weil in der bisherigen Rechtsprechung zu Zeitreihenanalysen die Gleitschlittenverfolgung des Aufschlagsatzes überwiegt (vgl. z.B.: BFH v. 25.3.2015, BStBl. II 2015, 743, FG Münster v. 19.8.2004 – 8 V 3055/04, FG Düsseldorf v. 15.2.2007 – 16 V 4691/06 oder FG Köln v. 27.1.2009 – 6 K 3954/07), werden die Beurteilungen von Kritiker*innen mitunter undifferenziert übernommen (s. z.B.: Beyer, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, 2933-2934: „Hierzu werden die einzelnen Quartals-RAS in einem Diagramm dargestellt. Schwanken die RAS in den Quartalen, kann dies mit Änderungen der Umsatzträger des Unternehmens zu tun haben“ oder Bleschick, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353: „Zunächst unternimmt er eine sog. betriebswirtschaftliche Auswertung: Hierzu ermittelt er im Wesentlichen mithilfe des Zeitreihenvergleichs einen Rohgewinnaufschlagsatz.“). Demgegenüber unterscheiden die RiBFH durchaus, wie bereits aus folgendem Titel zu entnehmen ist: Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548.

4.2.7. Jeweilige Entwicklungen der einzelnen Datenreihen

Eine **jeweilige Entwicklungsbetrachtung der einzelnen Datenreihen** als Gegenüberstellung von Jahren kann aus Besteuerungsdaten weitere Informationen zur zügigen Identifikation prüfungsrelevanter Sachverhalte bzw. Erkenntnisse für eine Plausibilitätsbeurteilung beisteuern. Dazu werden auf den Tabellenblättern „Monatsverhältnisse“ und „Wochentagswerte“ der SRP-Vorlage BWA die Jahresgrafen zur betreffenden Datenreihe als Monatssummen bzw. als Mediane für die Wochentage verglichen.

Für eine Vertiefung von beobachteten Auffälligkeiten bieten sich die Häufigkeitsanalysen in der SRP-Vorlage Stochastik auf den Tabellenblättern „Schichtung“ und „Schichtungsvertiefung“ an (s. [Pkt. 4.3.3.](#)).

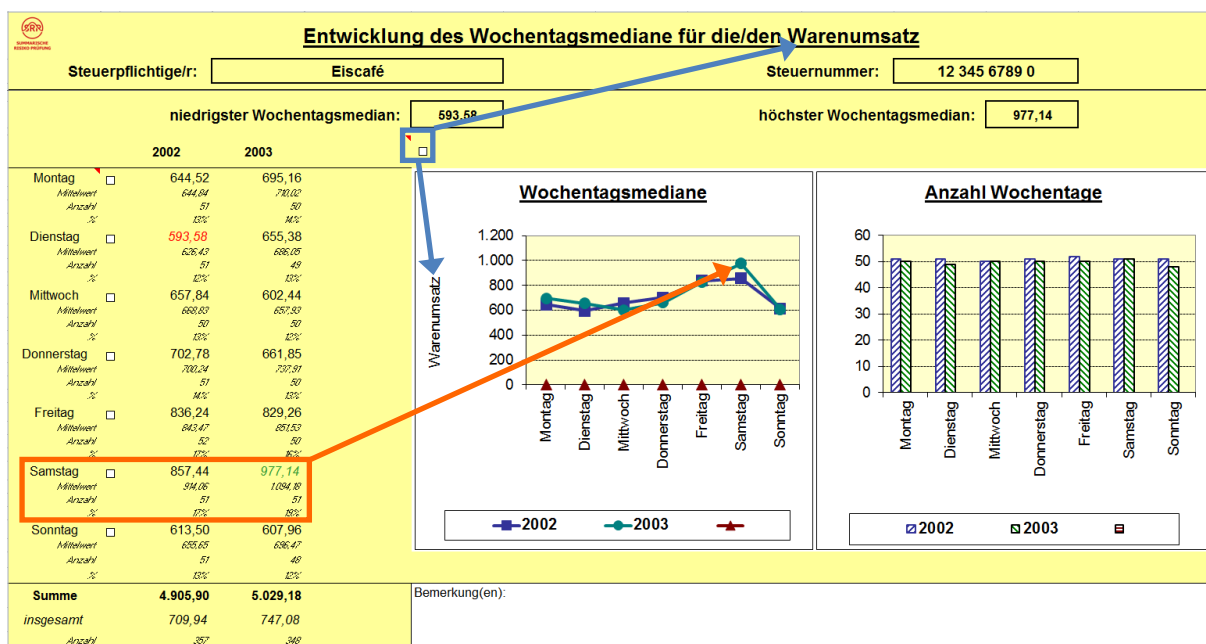
Die – in der Unternehmensplanung und -beratung geschätzte – Aussagekraft von betrieblichen Monatssummen (s. [Pkt. 4.2.5.](#)) kann für diverse Formen von Zeitreihenanalysen genutzt werden. Im Unterschied zur Prüfung, ob zwei Betriebsfaktoren eine handlungslogisch erwartete Abhängigkeit zeigen, stellt die **jahresweise Gegenüberstellung der monatlichen Entwicklung einer Betriebsgröße** häufig außerordentliche Veränderungen gegenüber Saisonschwankungen deutlicher heraus. Über ein [Anhakfeld](#) lässt sich die gewünschte Datenreihe auswählen:



Diese Grafik zu einem Fischgeschäft lässt den üblichen Wareneinkaufsrhythmus über das Jahr gut erkennen, indem die drei Grafen weitgehend deckungsgleich verlaufen. Vorhandene Ablösungen wie z.B. im Mai des ersten Jahres oder für die Phase „Oktober bis November“ des letzten können bspw. auf konjunkturelle Einflüsse zurückgehen und sollten sich dann entsprechend in abhängigen Betriebsfaktoren – also hier dem Umsatz – wiederfinden.

Noch auffallender ist die **andersartige Absenkung in den Sommermonaten „Juli bis August“**. Neben verzerrenden Sachverhalten kommen als mögliche Ursachen von Regelstörungen (kurzfristige) Strukturveränderungen³⁹⁶ im Unternehmen in Betracht. Mit Hilfe der SRP-Zeitreihensystematik lassen sich außergewöhnliche Sondervorgänge als Störeinflüsse identifizieren und bei Bedarf ausschließen (s. [Pkt. 4.2.2.](#)). Da zu diesem Beispiel keine derartigen Verursacher auffindbar sind, kann ein betrieblicher Auslöser in einer strukturellen Ablaufbesonderheit des ersten Jahres liegen. Eine entsprechende Nachfrage im Unternehmen ergibt, dass die Betriebsferien üblicherweise komplett im August stattfinden, in 2006 aber ausnahmsweise zur Hälfte in den Juli vorverlegt wurden. Diese Erkenntnis zu den betriebspezifischen Verhältnissen trägt in der Folge zu einer qualifizierteren Plausibilitätsbeurteilung i.S. des § 158 Abs. 2 Nr. 1 AO bei.

Insbesondere im Hinblick auf die dokumentierte Absatzsituation bei bargeldintensiven Betrieben sowie bei der Untersuchung einer Vielzahl von Kostenpositionen auf prüfungsrelevante Sonderfälle können aus dem **Jahresvergleich der Wochentagsmediane** ebenfalls wichtige Informationen hervorgehen.

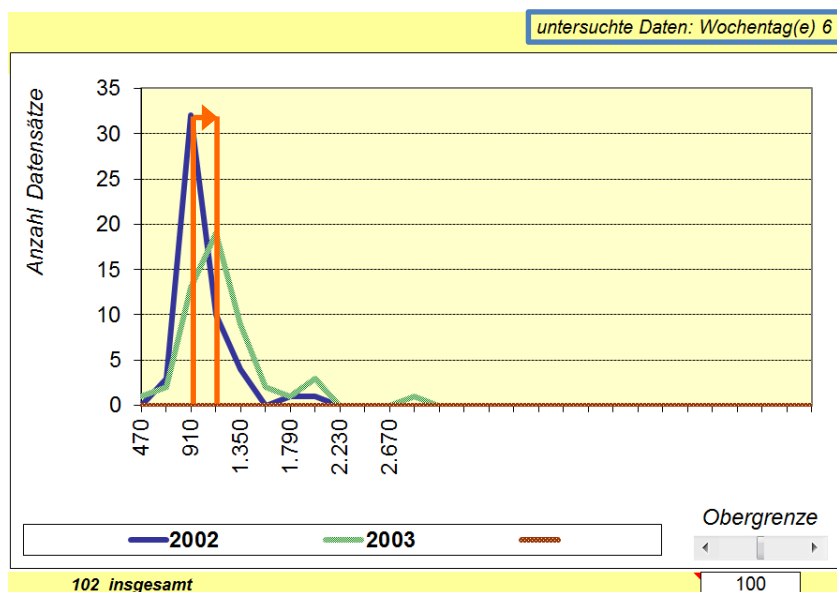


³⁹⁶ BFH v. 25.3.2015, BStBl. II 2015, 743, 1. Leitsatz: „Die Durchführung eines Zeitreihenvergleichs setzt voraus, dass im Betrieb das Verhältnis zwischen dem Wareneinsatz und den Erlösen im betrachteten Zeitraum weitgehend konstant ist. Es darf zudem im maßgebenden Zeitraum nicht zu solchen Änderungen in der Betriebsstruktur gekommen sein, die – nicht anderweitig behebbar – wesentliche Unsicherheiten bei der Aufstellung und Interpretation des Zahlenwerks mit sich bringen.“ Geeignete Zeitreihentypen wie hier im Beispiel oder beim doppelskaliert größenangepassten Zeitreihenvergleich für Abhängigkeitsuntersuchungen lassen solche Strukturänderungen als sinnvolle Vertiefungsfelder eindeutig erkennen (s.a. den Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.).

Weil arithmetische Mittelwerte sehr anfällig gegenüber Verzerrungen durch Ausreißer sind³⁹⁷, wird zur Stellvertretung der 50 %-Rang „Median“ ausgegeben – zuzüglich der jeweiligen Mittelwerte, der Anzahl an vorhandenen Wochentagen und der Anteile an der „typischen“ Wochensumme. Auch bei dieser Auswertung kann die zu untersuchende Datenreihe über ein [Anhakfeld](#) ausgewählt werden.

Für das oberhalb abgebildete beispielhafte Eiscafé ist die sehr ähnliche unterwöchige Nachfragesituation über die beiden Jahre in Form der Kongruenz (Deckungsgleichheit) bei der Grafen ablesbar. Allerdings sticht die einseitige **Verbesserung der Umsatzlage für die Samstage des zweiten Jahres** hervor: Der außerordentliche Anstieg der typischen Samstagserlöse um ca. € 120 betrifft einen Standardöffnungstag mit jeweils 51 Tageswerten und bedeutet eine Störung der erwarteten Regel „gleichmäßige Nachfrage“. Zusätzlich zu dieser insgesamt besseren Nachfragesituation im zweiten Jahr offenbart der größere Abstand zwischen Median und Mittelwert „€ 977,14 zu € 1.094,18“ gegenüber „€ 857,44 zu € 914,06“ einen weiterreichenden rechtsseitigen Auslauf (s. zur in Wirtschaftsdaten verbreiteten rechts-schiefen Häufigkeitsverteilung „Logarithmische Normalverteilung“ [Pkt. 4.3.4.1.](#)) mit mehr (starken) Ausreißern.

Trotz der vorlagenseitigen Trennung von regel- und musterbasierten Analysen können bei der Vertiefung derartiger Auffälligkeiten Häufigkeitsauszählungen helfen. In diesem Fall offenbart der Vergleich beider Jahresverteilungen zu den [Samstagserlösen](#) auf dem Tabellenblatt „Schichtungsverteilung“ in der SRP-Vorlage Stochastik (s. [Pkt. 4.3.3.](#)) gut erkennbar, dass die **gesamte**



Kurve der Samstage aus 2002 im Folgejahr auf der X-Achse nach rechts zu den höheren Werten gewandert ist. Der Niveauanstieg der Häufungszone von € 910 auf ca. € 1.100 ist

³⁹⁷ Arithmetische Kennzahlen wie Summen, Mittelwerte etc. sind anfällig für Ausreißerwerte: Wenn z.B. die Werte „10“, „10“, „11“, „12“, „12“, „13“, „15“, „15“, „18“ und „65“ durch ihren Mittelwert repräsentiert werden sollen, ergibt sich als wenig repräsentative Stellvertretung „18,1“. Rangkennzahlen, welche die Daten der Größe nach sortieren und dann einen absoluten bzw. relativen (Prozentränge = Quantile oder Bruchränge = Fraktile) Rang auswählen, bilden gerade bei asymmetrischen Verteilungen wie häufig in Wirtschaftsdaten (s. [Pkt. 4.3.4.1.](#)) aussagekräftigere Kenngrößen; so trifft der Median (50 %-Rang) in diesem Fall mit dem Wert „12,5“ [zwischen dem 5- und 6-kleinsten Wert wird interpoliert (zischengerechnet)] den Großteil der Daten besser.

folglich hauptverantwortlich für die in den Wochentagsmedianen erkannte Nachfrageverbesserung. Demgegenüber wird die festgestellte stärkere Verzerrung des Mittelwerts in 2003 im Wesentlichen von einem Extremausreißer der grünen Häufigkeitsverteilung hervorgerufen.

Aus Plausibilitätssicht bedeutet diese konkrete mehrperspektivische Beobachtung zum indiziellen Gesamtbild³⁹⁸ von Besteuerungsgrundlagen: Ist es glaubhaft, dass die Kundennachfrage von einem Jahr auf das nächste ausschließlich an einem Wochentag wesentlich zugenommen hat, oder ist es eher „unwahrscheinlich [...]“, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt³⁹⁹? Mögliche einzelfallbezogene Begründungen für diese Unschlüssigkeit wie bspw. regelmäßige Sonderverkaufsaktionen an Samstagen ab 2003 können i.R. der gemeinsamen Sachaufklärungsverantwortung⁴⁰⁰ gezielt hinterfragt und interaktiv (s. [Pkt. 4.2.9.](#)) eingearbeitet werden.

4.2.8. Rückrechnung der monatlichen Warenbestände

Die rechnerische Rekonstruktion nicht dokumentierter unterjähriger Warenbestände stellt in der SRP keine eigenständige Verprobung dar, sondern soll der Verdeutlichung von Regelstörungen dienen, die mit Hilfe von doppelskalierten Zeitreihenvergleichen aufgedeckt wurden (zur unterschiedlichen Eignung bestimmter Zeitreihenvarianten für die Abhängigkeits- und damit Plausibilitätsprüfung s. den Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) und trotz SRP-Zeitreihensystematik (s. [Pkt. 4.2.2.](#)) zuzüglich interaktiver Optionen zum Einpflegen von konkreten Erklärungsversuchen (s. [Pkt. 4.2.9.](#)) nicht aufgeklärt werden konnten.

Um die Auswirkungen eines unschlüssigen Wareneinkaufsverhaltens – als Beobachtung eines Zeitreihenvergleichs zum Warenfluss – greifbarer zu veranschaulichen, können die **unterjährigen Warenbestände anhand der (Perioden-)Daten rückgerechnet** werden, sofern sie nicht aufgezeichnet wurden.

Auf dem Tabellenblatt „Bestandsentwicklung“ der SRP-Vorlage BWA werden dazu die Datenreihen „Warenumsatz“ und „Wareneinkauf“ inklusive interaktiver (s. [Pkt. 4.2.9.1.](#)) Zuordnungskorrekturen sowie die Eingaben zu Warenanfangsbestand, jährlichen Warenendbeständen und Entnahmen (alle auf dem Tabellenblatt „Jahreszahlen“) verwendet. Unter An-

³⁹⁸ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5: „Die gesetzliche Regelung des § 158 gründet sich insofern auf die Einsicht, dass es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild einer Buchführung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann.“

³⁹⁹ Anwendungserlass zu § 158 AO.

⁴⁰⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen.“

nahme eines konstanten (Jahres-)Aufschlagsatzes⁴⁰¹ wird für jeden Monat der erforderliche Wareneinsatz⁴⁰² berechnet und dem Periodennachkauf gegenübergestellt. Mit der Differenz wird der Warenanfangsbestand monatlich fortgeführt.

Monat	Warenumsatz	Ent- nahmen	notwendiger Einsatz	Wareneinkauf	Bestands- veränderung	laufender Bestand
	2006	☐ → RGAS: 244,27%			Anfangsbestand:	22.620,00
Jan. 2006	103.316,36	795,61	30.806,27	28.669,66	-2.136,61	20.483,39
Feb. 2006	105.679,39	795,61	31.492,67	28.849,07	-2.643,60	17.839,79
Mär. 2006	119.460,88	795,61	35.495,82	35.092,60	-403,22	17.436,57
Apr. 2006	154.220,90	795,61	45.592,69	45.561,73	-30,96	17.405,61
Mai 2006	210.792,54	795,61	62.025,25	59.712,01	-2.313,24	15.092,38
Jun. 2006	274.183,18	795,61	80.438,55	77.790,74	-2.647,81	12.444,57
Jul. 2006	241.253,15	795,61	70.873,25	75.018,82	4.145,57	16.590,15
Aug. 2006	224.315,47	795,61	65.953,30	73.903,05	7.949,75	24.539,90
Sep. 2006	208.366,35	795,61	61.320,50	64.639,40	3.318,90	27.858,79
Okt. 2006	155.559,82	795,61	45.981,61	46.412,66	431,05	28.289,85
Nov. 2006	110.351,57	795,61	32.849,81	36.503,51	3.653,70	31.943,54
Dez. 2006	128.789,24	795,61	38.205,46	27.627,92	-10.577,54	21.366,00
insgesamt	2.036.288,85	9.547,30	601.035,17	599.781,17	-1.254,00	
	2007	RGAS: 246,53%				

In der exemplarischen Tabelle wurde die **Standard-Option „jährlich gesonderte Betrachtung“** ausgewählt, wodurch für jedes Jahr der eigene Aufschlagsatz zu Grunde gelegt wird. Wenn keine strukturellen Veränderungen der Jahre bekannt sind, kann durch Aktivierung des Ankreuzfelds für alle Jahre mit einem einheitlichen Gesamtaufschlagsatz⁴⁰³ gerechnet werden; in dieser Einstellung werden bei bilanzierenden Stpfl. die Jahresendbestände der Zwischenjahre nicht mit den ausgewiesenen übereinstimmen, sofern die jährlichen Aufschlagsätze voneinander abweichen.

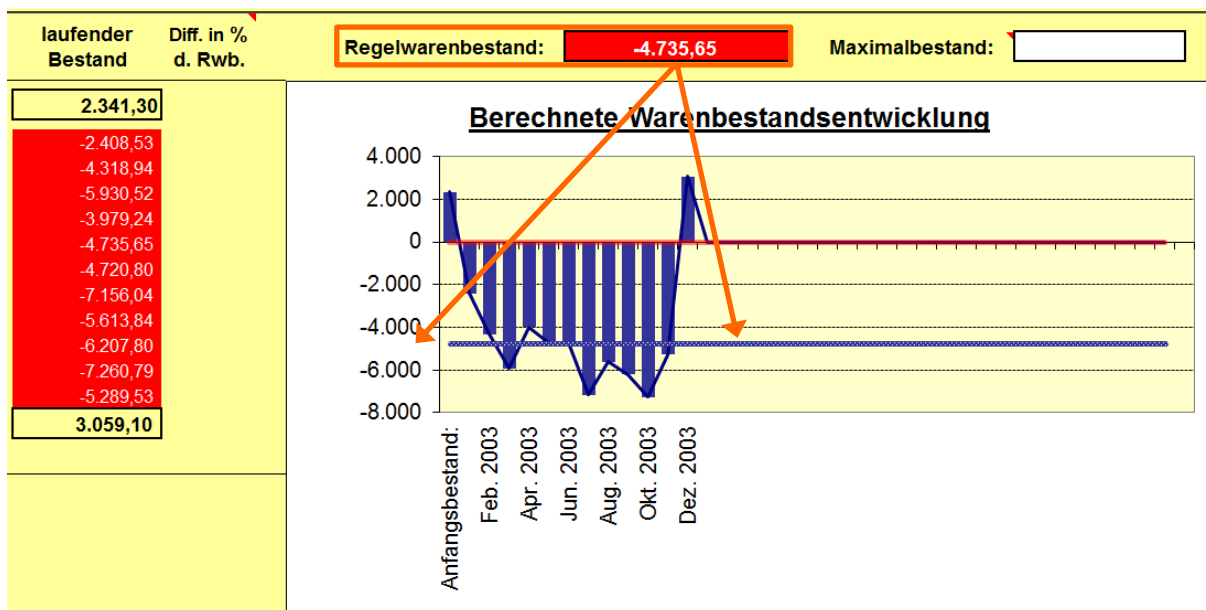
Als **mögliche Implausibilitäten der rückgerechneten Warenbestände** können (rechnerisch) negative oder unrealistisch hohe Zwischenstände resultieren; zur Verdeutlichung kann ein maximal möglicher Lagerbestand rechts oberhalb der ersten Grafik des Tabellenblatts „Bestandsentwicklung“ eingetragen werden, damit das zeitweise (rechnerische) Überschreiten im Diagramm offensichtlich wird.

⁴⁰¹ Damit teilt die rechnerische Rekonstruktion unterjähriger Warenbestände die größere Empfindlichkeit von Aufschlagsatzzeitreihen gegenüber Störeinflüssen im Vergleich zu den doppelskaliert größenangepassten Zeitreihenvergleichen (s. den direkten Vergleich im Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.). **Insofern trifft hier die Warnung aus dem BFH-Zeitreihenurteil (v. 25.3.2015, BStBl. II 2015, 743, 1. Leitsatz) zu:** „Die Durchführung eines Zeitreihenvergleichs setzt voraus, dass im Betrieb das Verhältnis zwischen dem Wareneinsatz und den Erlösen im betrachteten Zeitraum weitgehend konstant ist. Es darf zudem im maßgebenden Zeitraum nicht zu solchen Änderungen in der Betriebsstruktur gekommen sein, die – nicht anderweitig behebbar – wesentliche Unsicherheiten bei der Aufstellung und Interpretation des Zahlenwerks mit sich bringen.“

⁴⁰² Berechnung des monatlichen Wareneinsatzes nach der Formel: $\frac{\text{Warenumsatz}}{(\text{Aufschlagsatz} + 100\%)} + \text{Entnahmen}$.

⁴⁰³ Der Gesamtaufschlagsatz für den auf dem Tabellenblatt „Vergleichsdaten“ eingegebenen Prüfungszeitraum wird auf dem Tabellenblatt „Gleitschlittenzeitreihen“ oben rechts sowie in dem dazugehörigen Aufschlagsatzdiagramm ausgegeben (s.a. die erste Abb. unter Pkt. 4.2.6.).

Im nachfolgenden Beispiel eines Dönerimbisses resultieren innerhalb des Prüfungsjahres wegen sehr ungleichmäßiger Wareneinkäufe permanent rechnerisch negative Zwischenbestände des Warenlagers zu den Monatsenden, welche in der Tabelle farblich hervorgehoben werden. Deshalb ergibt sogar das Lagerrepräsentativ „Regelwarenbestand“, welches als Median (50 %-Rang) aller eingegebenen zuzüglich rückgerechneten Warenbestände ermittelt wird, einen **Negativwert**:



Diese ergänzende Perspektive zum aufgezeichneten Warenfluss veranschaulicht gegenständlicher, dass die in den doppelskalierten Zeitreihenvergleichen aufgefallenen Einkaufsdaten hinsichtlich ihrer Plausibilität nicht nachvollziehbar sind: So hatte dort der unzusammenhängende⁴⁰⁴ Extremwert „Dez 2003“ trotz größenanpassender Doppelskalierung die Vergleichbarkeit von Warenumsatz und -einkauf verhindert (s. [Pkt. 4.2.5.2.](#)). Durch die Umstellung der Monatswerte auf Anteile der jeweiligen Jahressummen im Zeitreihenvergleich mit geringfügiger Niveaueinstellung der Grafen ließ sich aus dem Diagramm entnehmen, dass der Zusammenhang der Warenflussdaten über das gesamte Jahr nicht schlüssig ist⁴⁰⁵. Aus Sicht der unterjährig rückgerechneten Warenbestände führt diese Auffälligkeit dazu, dass über das Jahr stets zu wenig Ware für den erzielten Umsatz vorhanden war, weil der Einkauf erst zum Jahresende konzentriert erfasst wurde.

Solange sich derart unschlüssige rechnerische Zwischenstände des Warenlagers nicht in gemeinsamer⁴⁰⁶ interaktiver (s. [Pkt. 4.2.9.](#)) Sachaufklärung über die Neuzuordnung von

⁴⁰⁴ Sofern beide Datenreihen in ihren Monatssummen stark schwanken, dabei aber die erwartete schlüssige Reaktion aufeinander zeigen, wird die Größenanpassung via Doppelskalierung nicht gestört.

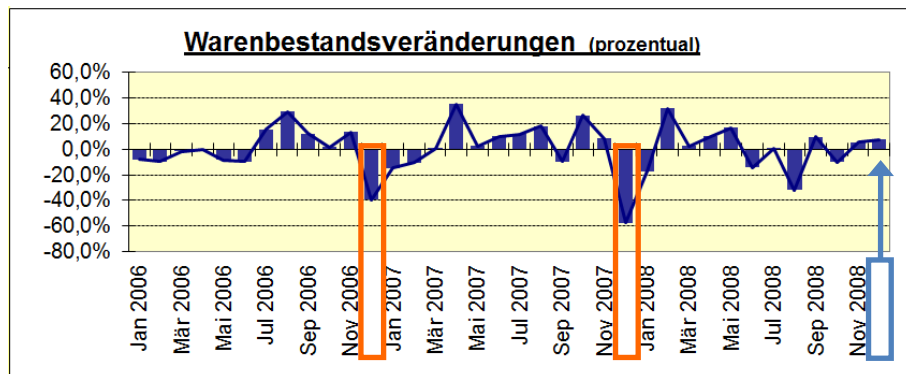
⁴⁰⁵ Mit Hilfe der interaktiven Optionen (s. [Pkt. 4.2.9.](#)) der SRP-Vorlage BWA konnte zudem ausgeschlossen werden, dass die Aufzeichnungsweise „Einnahmen-Überschuss-Rechnung“ nach § 4 Abs. 3 EStG (alleine) ursächlich für die Regelstörungen ist.

⁴⁰⁶ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert

Wareneinkäufen oder erklärliche Aufschlagsatzschwankungen (weitgehend) auflösen lassen, geben sie Anlass für berechtigte Zweifel an der sachlichen Richtigkeit der Besteuerungsgrundlagen. Gemäß umfangreicher Prüfungserfahrungen mit der SRP gehen unschlüssige rechnerische Warenzwischenstände normalerweise mit Regelstörungen des deutlich robusteren doppelskalierten Zeitreihenvergleichs (*zur unterschiedlichen Eignung bestimmter Zeitreihenvarianten für die Plausibilitätsprüfung s. den Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.*) einher, sodass die Schlüssigkeitsbeurteilung i.S. des § 158 Abs. 2 Nr. 1 AO **mehrperspektivisch abgesichert** erfolgt.

Sofern sich i.R. der Rückrechnung der unterjährigen Warenbestände ein vernünftiges (positives) Repräsentativ für den normalen Füllungszustand des Lagers ergibt, werden in einem zweiten Diagramm die (rechnerisch) **relativen Bestandsveränderungen** der einzelnen Monate greifbarer verbildlicht. Durch diese zusätzliche Sicht auf die Warenflussdaten fällt in

der Beispielsgrafik eines größeren Gastronomiebetriebs der jeweils **fehlende Wareneinkauf im Dezember der ersten beiden Jahre** auf; dass dies für das



letzte Jahr in der Rückrechnung nicht zu erkennen ist, liegt – unabhängig vom Vorhandensein der Ursache – an der Funktion des Warenendbestands⁴⁰⁷ als Fixparameter in der retrograden Berechnung. Die Nachverfolgung der entdeckten Regelstörungen im Beispiel führt nach der SRP-Zeitreihensystematik (s. Pkt. 4.2.2.) zu hohen Jahresrückvergütungen.

4.2.9. Interaktive Funktionen zur verbesserten Sachaufklärung

Ein fundamentaler Vorteil der digitalen Außenprüfung ist die Möglichkeit, neue Erkenntnisse jederzeit reversibel in die (Gesamtbild-)Prüfung einzupflegen und dadurch deren Auswirkungen auf die Ergebnisse umgehend zu simulieren⁴⁰⁸. Dazu beinhalten die SRP-Vorlagen di-

(§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen“.

⁴⁰⁷ Der Warenendbestand ergibt sich aus der Eingabe auf dem Tabellenblatt „Jahreszahlen“.

⁴⁰⁸ Vgl. auch: PwC-Newsletter v. 31.8.2022: „Aus unserer Praxis: Der Einsatz von Data & Analytics im Verrechnungspreismfeld – quo vadis? [...] D&A-Lösungen finden aber auch vielfach Einsatz im Rahmen von Betriebsprüfungen, z. B. für die Simulation der Steuereffekte bei Diskussionen um fremdübliche Bandbreiten“, nachzulesen unter: <https://www.pwc.de/de/newsletter/transfer-pricing-perspectives-dach/ausgabe-55-2022/aus-unserer-praxis-der-einsatz-von-data-and-analytics-im-verrechnungspreismfeld-quo-vadis.html>.

verse Funktionen, um die Prüfungsdaten bequem und nachvollziehbar an einen neuen Informationsstand bzw. mögliche Sachverhaltsabwandlungen anzupassen. Für die qualifizierte Beurteilung der Mitwirkungsqualität, ob „der Steuerpflichtige über seine Angaben [...] ausreichende Aufklärungen zu geben vermag“⁴⁰⁹, sowie der Plausibilität i.S. des § 158 Abs. 2 Nr. 1 AO ist es wichtig, alle bekannten Umstände – auch aus der „Beweisnähe [...] des Steuerpflichtigen“⁴¹⁰ – konkret auf die Bedeutung für die Schlüssigkeit hin untersuchen zu können.

Während die interaktiven Korrekturoptionen im Bereich der musterbasierten Analysen der SRP-Vorlage Stochastik wegen der speziellen Prüfsystematik nicht in einem gesonderten Kapitel zusammengefasst sondern zu den jeweiligen Einzeluntersuchungen bzw. Untersuchungsschritten beschrieben werden (s. [Pkt. 4.3.1. bis 4.3.4.](#)), gibt dieser Pkt. einen Überblick über die **Anpassungsoptionen in der SRP-Vorlage BWA**. Folgende Funktionen werden dazu in den nächsten Unterpunkten beschrieben:

- ☒ Einzelkorrektur von Datensätzen durch Ausblenden, Umdatierung oder Wertänderung,
- ☒ Einschränkung des Betrachtungszeitraums,
- ☒ Verschieben der ersten Datenreihe um eine feste Anzahl Tage,
- ☒ Bereinigung von (einseitigen konjunkturellen) Preiseinflüssen sowie
- ☒ Niveaueinstellung eines Grafen zur Verbesserung der Vergleichbarkeit von Datenreihen bei einer Zeitreihen gegenüberstellung.

Mit Hilfe dieser Möglichkeiten lässt sich die gemeinsame Pflicht zur Sachverhaltsaufklärung⁴¹¹ konkret und – nach umfangreichen Praxiserfahrungen – unter (deutlich) reduziertem Aufwand⁴¹² umsetzen. In der Konsequenz können die Substantiiertheit von Mitwirkungen

⁴⁰⁹ § 162 Abs. 2 S. 1 AO.

⁴¹⁰ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. Die Frage der Feststellungslast stellt sich, wenn ein entscheidungserheblicher Sachverhalt trotz Ausschöpfung aller zugänglichen und zumutbaren Ermittlungsmöglichkeiten nicht vollständig aufgeklärt werden kann. [...] Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“

⁴¹¹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen“.

⁴¹² Gelegentlich werden der SRP in der Literatur „überzogene Mitwirkungspflichten“ angelastet (vgl. z.B.: Steinhilff, „Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388, 389). Tatsächlich ist es durch die digitalen Möglichkeiten überhaupt erst möglich geworden, die unvermeidbaren Unsicherheiten von Verprobungen zügig auf das Minimalmaß zu reduzieren, welches wiederum durch die Ordnungsmäßigkeit der Aufzeichnungen bestimmt wird. Interaktive Funktionen wie in den SRP-Vorlagen ver-

Stpfl.⁴¹³ klarer eingeordnet und Schlüssigkeitsbeurteilungen von Besteuerungsgrundlagen gem. § 158 Abs. 2 Nr. 1 AO qualifizierter getroffen werden.

Sämtliche interaktiven Beeinflussungen untersuchter Besteuerungsdaten lassen sich mit den druckvorbereiteten SRP-Vorlagen nachvollziehbar dokumentieren. So werden vorgenommene Korrekturen auf allen betroffenen Auswertungsblättern automatisch angezeigt, und es können individuelle Kommentare in den Bemerkungsfeldern ergänzt werden.

4.2.9.1. Einzelkorrektur von Datensätzen

Für Einzelkorrekturen (tlw.) fehlerhafter Datensätze stehen zu den importierten Datenreihen in der SRP-Vorlage BWA jeweils zwei **Extraspalten** zur Verfügung, in denen **Wert-**

berichtigungen und dazugehörige Erklärungen eingetragen werden können

[Tipp: Zur besseren Nachvollziehbarkeit wird empfohlen, neu hinzugefügte Beträge z.B. wie in der Abb. zur (partiellen)

Einkauf				
-564,38				
Datum	Betrag	Korrektur	Bemerkung	Einkauf korrigiert
857	853	4	5	856
29.08.2002	2.257,50	-2.257,50	Rg. 647 mit Bestands-einkauf Getränke	0,00
15.09.2002		564,38	Rg. 647 Verteilung nach Verbrauch	564,38
15.10.2002		564,38	Rg. 647 Verteilung nach Verbrauch	564,38
15.11.2002		564,38	Rg. 647 Verteilung nach Verbrauch	564,38
15.12.2002			Rg. 647 Verteilung nach Verbrauch	

Umverteilung in die Korrekturspalte zu schreiben.]. Ein **Zählerkasten** protokolliert die Summe der bereits erfolgten Werteintragungen, sodass schrittweise Umdatierungen erleichtert werden. Über die **Autofilter** zu den Spalten lassen sich alle Anpassungen oder diejenigen zu einem besonderen Sachverhalt gesondert ausgeben und dokumentieren.

Auch wenn sich die vorgenommenen Einzelkorrekturen in ihrer Summe ausgleichen, erscheint unterhalb jeder betroffe- **Auswertung nach Anpassung von Einzeldaten (siehe Vergleichsdaten)** nen Auswertung automatisch der abgebildete Hinweistext. Zusätzlich wird bei insgesamt nicht neutralen Einzelkorrekturen auf dem Tabellenblatt „Schätzung“ mit einem roten Rahmen um das Indikatorfeld „80 %-Quantil RGAS“ daran erinnert (s. [Pkt. 4.8.](#)).

bessern dabei Komfort, Effektivität und Effizienz – auch für die Stpfl. in Form einer zielführenderen Einforderung der Mitwirkung. Dass Prüfungsmethoden auch zur unrechtmäßigen Verlagerung von Aufklärungspflichten eingesetzt werden können, ist kein Makel der SRP, sondern trifft alle und ist deshalb als Abwertungskriterium ungeeignet. (Technische) Möglichkeiten zur Einschränkung der unvermeidbaren Unsicherheiten bei der Aufarbeitung komplexer Prüffelder sind deshalb zeitgemäße Hilfsmittel zur Umsetzung der gesetzlichen Pflicht, „alle für den Einzelfall bedeutsamen [...] Umstände zu berücksichtigen“ (§ 88 Abs. 1 S. 2 AO).

⁴¹³ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 9: „§ 162 II i.V.m. § 158 orientieren das Beweismaß an der Sphärenverantwortlichkeit [...] Das Beweismaß hängt mithin davon ab, wie viel und in welcher Form der Stpfl. aus seiner Verantwortungssphäre zur Sachaufklärung beiträgt und wieweit die FinBeh. auf dessen Mitwirkung im konkreten Einzelfall angewiesen ist.“

4.2.9.2. Einschränkung des Betrachtungszeitraums

In einigen Fällen ist es bspw. zur Einschätzung einer beobachteten Verzerrung in Zeitreihen-
vergleichen durch einen außergewöhnlichen Teilzeitraum (s. das Beispiel zu „verzerrten Y-Achsen-Skalierungen“ unter [Pkt. 4.2.5.2.](#)) hilfreich, den Prüfungszeitraum zügig einschränken zu können. Hierfür lassen sich auf dem Tabellenblatt „Vergleichsdaten“ oben rechts auch **Teilzeiträume** frei eintragen, auf die dann alle Analysen begrenzt werden.

Ein gesonderter Hinweis hierzu ergeht auf den Auswertungsblättern nicht, weil den jeweiligen Diagrammen i.d.R. die Unvollständigkeit des Prüfungszeitraums zu entnehmen ist. Soll eine zeitliche Einschränkung dennoch gesondert aufgeführt werden, muss die Eingabe in den Bemerkungsfeldern manuell erfolgen. Demgegenüber wird auf dem Tabellenblatt „Schätzung“ durch einen roten Rahmen um das Indikatorfeld „80 %-Quantil RGAS“ darauf hingewiesen, falls Daten auf dem Tabellenblatt „Vergleichsdaten“ außerhalb des eingetragenen Prüfungszeitraums liegen.

Im Hinblick auf die sachgerechte Bestimmung einer Schätzungshöhe mit der Quantilschätzung (s. [Pkt. 4.8.](#)) – *nach eingetretener Amtspflicht der Schätzung gem. § 162 AO* – unter wesentlich abweichenden Bedingungen von Teilen des Prüfungszeitraums ist die Einschränkung des Analysezeitraums ebenfalls komfortabel, indem die SRP-Vorlage BWA mit allen Daten kopiert und für die jeweiligen Teilzeiträume eingestellt werden kann.

4.2.9.3. Verschieben der ersten Datenreihe

Trotz der guten Stabilität des doppelskaliert größenangepassten Zeitreihenvergleichs gegenüber eher alltäglichen Störeinflüssen (s. den Vergleich der gängigen Zeitreihenvarianten im Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) wird bei Fällen mit Einnahmen-Überschuss-Rechnung nach § 4 Abs. 3 EStG regelmäßig als Erklärungsansatz für Regelstörungen der **Zeitversatz zwischen operativem Wareneinkauf bzw. -verbrauch und dem Erfassungszeitpunkt** „Bezahlung“ angeführt. Um diese – nach ständigen Prüfungserfahrungen häufig *nicht* für die entdeckten Auffälligkeiten ursächliche – Besonderheit interaktiv in die SRP-Zeitreihensystematik einbinden zu können, lässt sich die erste Datenreihe um bis zu 30 Tage⁴¹⁴ komplett auf der Zeitachse nach hinten oder vorne verschieben.

⁴¹⁴ Eine generelle Zahlungsferne aller Wareneinkäufe von mehr als 30 Tagen erscheint im Wirtschaftsleben, in dem Liquidität ein kostbarer Vorteil ist, alleine schon implausibel i.S. des § 158 Abs. 2 Nr. 1 AO. Soll im Einzelfall dennoch eine weitreichendere Verschiebung einer Datenreihe simuliert werden, müssen die jeweiligen Daten entweder vor dem Hineinkopieren in die SRP-Vorlage BWA oder auf dem Tabellenblatt „eigene Notizen“ entsprechend vollständig vor- oder rückdatiert werden.

Für den exemplarischen Dönerimbiss, der in den Monatszeitreihen (s. [Pkt. 4.2.5.2.](#)) sowie bei der Rückrechnung des Warenbestands (s. [Pkt. 4.2.8.](#)) mit unschlüssigen Warenflussdaten aufgefallen ist, wird lt. Abb. eine simulierte Rückdatierung aller Wareneinkäufe um 14 Tage

Einkauf	
→ -14	Sollen alle Daten dieser Größe um ein bis maximal 30 Tage vor- oder rückdatiert werden, hier bitte die Anzahl Tage eintragen. Werden durch die allgemeine Umdatierung die Skalierungen der Diagramme gestört, sollte der Prüfungszeitraum angepasst werden.
Datum	
332	

eingetragen, um den Einfluss der verzögerten Aufzeichnungsweise auf die Zeitreihenanalysen zu prüfen. **Hinweis:** Durch diese Rückdatierung aller Positionen fallen i.d.R. einige Vorgänge aus dem eingegebenen Prüfungszeitraum heraus, wodurch insbesondere die Werte der Start- oder Endperiode erheblich verzerrt werden können. Falls dadurch einzelne Zeitreihenanalysen infolge automatisch angepasster Y-Achsenskalierungen unbrauchbar scheinen, lässt sich der (negative) Effekt am schnellsten durch eine Einschränkung des Prüfungszeitraums auf dem Tabellenblatt „Vergleichsdaten“ aufheben.

Eine zeitliche Verschiebung der ersten Datenreihe hat auf jedem betroffenen Auswertungsblatt automatisch die- **Auswertung nach Reihenverschiebung um -14 Tag(e) (siehe Vergleichsdaten)** sen Hinweis mit Benennung der eingetragenen Tagesanzahl und Richtung zur Folge.

Weil die Vor- oder Rückdatierung einer kompletten Datenreihe spätestens i.R. einer Schätzung – nach eingetretener Amtspflicht der Schätzung gem. § 162 AO – zurückgenommen werden sollte, erinnert das Indikatorfeld „80 %-Quantil RGAS“ mit roter Schrift an die bestehende Aktivierung dieser interaktiven Funktion (s. [Pkt. 4.8.](#)).

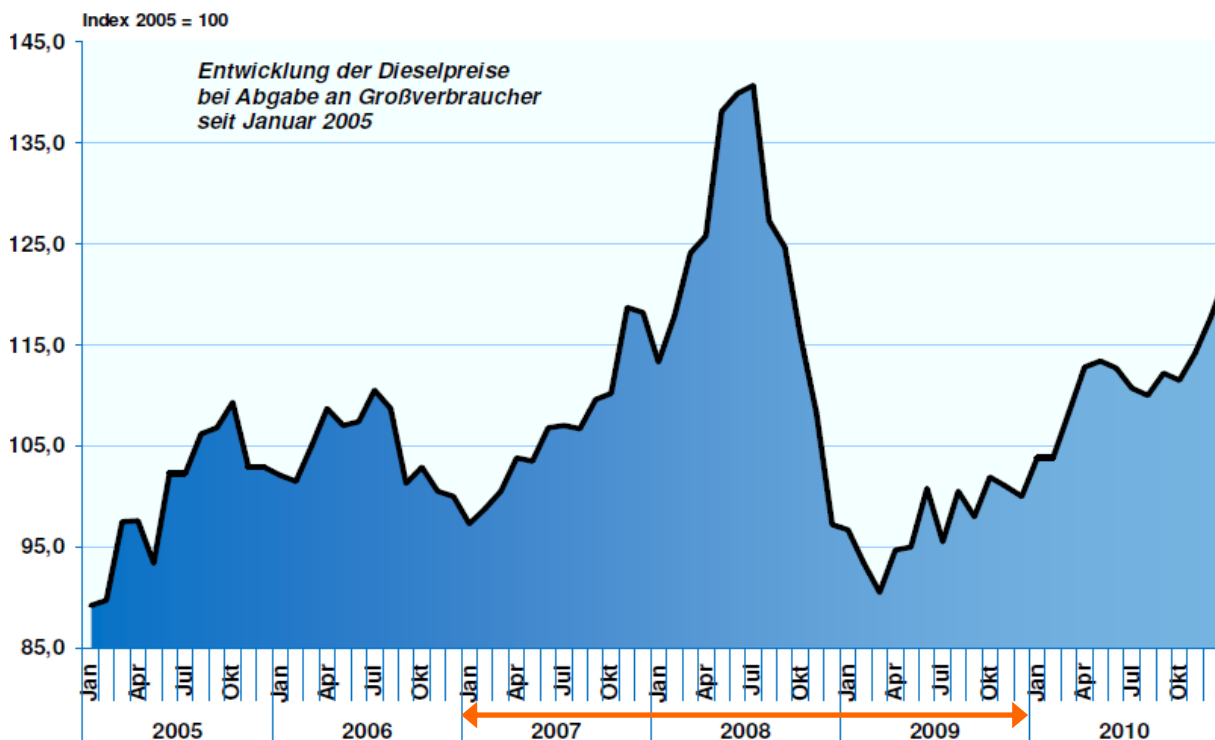
4.2.9.4. Beseitigung von (einseitigen konjunkturellen) Preiseinflüssen

Ein wichtiger Störfaktor für Zeitreihenergebnisse können einseitige Preiseinflüsse sein, auf die in Unternehmen nur bedingt reagiert wird bzw. werden kann: z.B. kurzfristige starke Preiserhöhungen im Einkauf (s. das Beispiel unter [Pkt. 4.2.5.1.](#)), die nicht an die Kunden weitergegeben werden sollen, oder (gesetzliche) Preisbindungen. Obwohl der doppelskaliert größenangepasste Zeitreihenvergleich stabil gegenüber vielen Störeinflüssen ist (s. den Vergleich gängiger Zeitreihenvarianten im Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)), bietet die SRP-Vorlage BWA mit der **Konjunkturanpassung** eine Funktion, um derartige (einseitige) Preiseffekte zu egalisieren⁴¹⁵; die vollständige Umrechnung aller Werte beider

⁴¹⁵ Die in der Volkswirtschaftslehre übliche Bereinigung von Preiseffekten durch Indizes ist vom BFH (Beschl. v. 12.7.2017 – X B 16/17, Rz. 128) bisher in ihrer Funktion nicht zutreffend erkannt worden: „Das FA hat hierzu in der Beschwerdeerwiderung ausgeführt, mit der ‚Konjunkturanpassung‘ hätten die Ergebnisse der Jahre 2009 und 2010 an das Preisniveau des Jahres 2008 angepasst werden sollen, um die Vergleichbarkeit der Schätzung herzustellen. Indes soll eine Schätzung nicht vergleichbar, sondern möglichst realitätsnah sein. Wenn im Prüfungszeitraum eine erhebliche Preiserhöhung stattgefunden hat, dann muss die Schätzung dies berücksichtigen, nicht aber die – nicht miteinander vergleichbaren – RAS der Jahre vor und nach der Preiserhöhung vergleichbar machen.“

Datenreihen mittels Indizes⁴¹⁶ auf ein einheitliches Preisniveau würde eine Betrags- (quasi) in eine Mengenverproben transferieren.

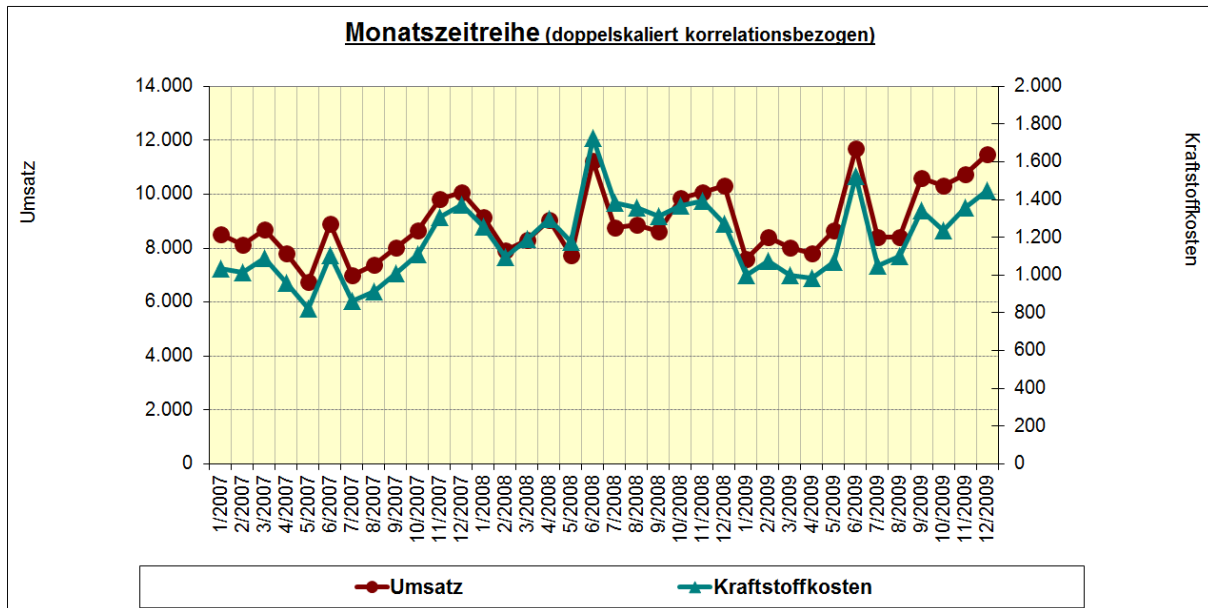
In der Taxibranche haben einseitige Konjunkturreffekte eine erhebliche Praxisbedeutung, weil den veränderlichen Kraftstoffpreisen als wesentlichem Ausgabenfaktor festgesetzte Tarife gegenüberstehen. Durch die regionale Preisbindung für Taxifahrten kann nur bedingt auf Marktschwankungen reagiert werden. So hat für den nachfolgenden Beispielsfall der Dieselpreis während des Prüfungszeitraums „2007 bis 2009“ lt. Abb. erheblich geschwankt⁴¹⁷:



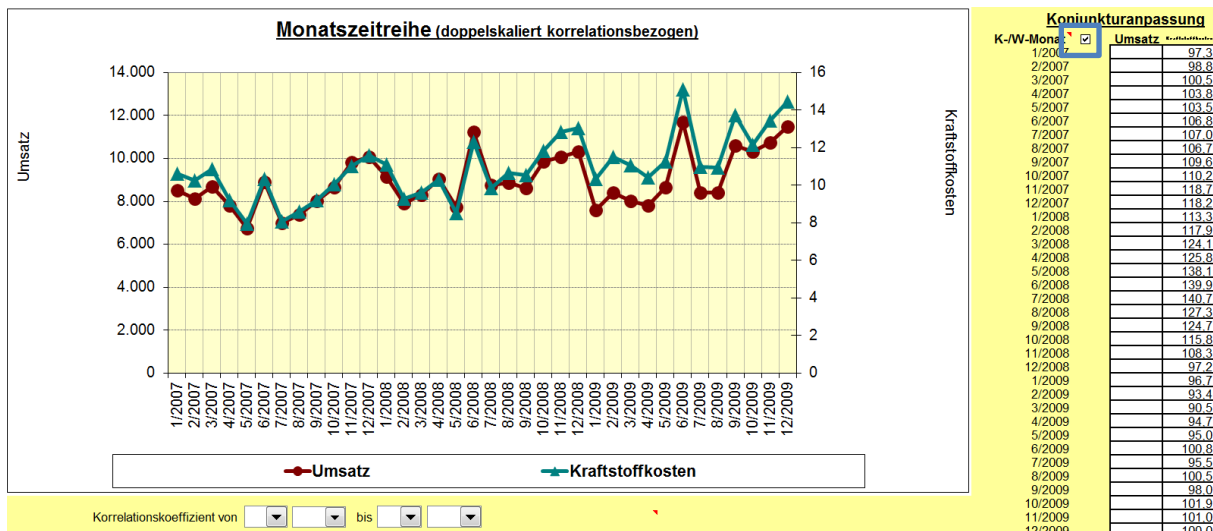
Aus dem nachfolgenden doppelskaliert größenangepassten Zeitreihenvergleich des geprüften Taxiunternehmens lässt sich dennoch grundsätzlich eine klare Abhängigkeit des Dieseleinkaufs zum Absatz erkennen: Auf (fast) jede Veränderung der Nachfrage reagieren die Kraftstoffkosten gleichgerichtet und mit ähnlicher Intensität. Auffällig ist hingegen die Niveauerhöhung der Ausgaben im mittleren Jahr, welches genau die Phase des dauerhaft starken Anstiegs der Dieselpreise markiert.

⁴¹⁶ Preisindizes rechnen Werte von „nominal“ auf „real“ um, wonach sie denselben Preisverhältnissen unterliegen und dadurch vergleichbar(er) werden.

⁴¹⁷ Quelle: Internetseite des Bundesverbands Güterkraftverkehr Logistik und Entsorgung (BGL) e.V., nachzulesen unter: <https://www.bgl-ev.de/images/downloads/dieselpreisinformation.pdf>. Auch wenn die tatsächlichen Preisschwankungen für das Beispielsunternehmen regional und als Endabnehmer von dieser Internetrecherche abweichen können, verdeutlicht die Gegenüberstellung der Zeitreihenvergleiche „nominal“ und „real“ den begrenzten Einfluss von Preiseffekten auf robuste Zeitreihenvarianten.



Im Tabellenbereich „Konjunkturanpassung“ neben dem Diagramm lassen sich in die SRP-Vorlage BWA die monatlichen Preisindizes für eine oder beide Datenreihen einpflegen, um anschließend per [Anhakfeld](#) zwischen der konjunkturbereinigten Realversion des Zeitreihenvergleichs und der nominalen hin- und herschalten zu können. Auch für dieses Extrembeispiel wird die Robustheit des doppelskaliert größenangepassten Zeitreihenvergleichs (s.a. *den Vergleich von gängigen Zeitreihenvarianten im Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.*) bestätigt, indem eine geringfügigere Veränderung resultiert, als aufgrund der eklatanten einseitigen Preisschwankungen zu erwarten gewesen wäre:



Neben der besseren Deckungsgleichheit beider Grafen in der ersten Hälfte trotz des starken Preisanstiegs für den Kraftstoff lässt sich eindeutig ablesen, dass ab Mitte 2008 der preisbereinigte Dieselaufwand *relativ* angewachsen ist. Diese – vorher unsichtbare – Regelstörung kann nun auf mögliche Verursacher wie z.B. die Anschaffung neuer Fahrzeuge mit einem höheren Verbrauch hinterfragt werden. Ergeben sich keine überprüfbaren betrieblichen

Erklärungen, besteht Grund, daran zu zweifeln, „dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt“⁴¹⁸.

Falls die Konjunkturanpassung aktiviert ist, wird dies zu den betroffenen Auswertungen der SRP wie für die „Quartals- *Auswertung nach aktivierter Preisbereinigung (siehe Monatszeitreihen)* relationen“ abgebildet angezeigt, wobei die Übertragung der Preisindizes vom Tabellenblatt „Monatszeitreihen“ auf die „Quartalsverhältnisse“ über das dortige Anhakfeld oben links bewusst angesteuert werden muss. Auf die Tabellenblätter „Gleitschlittenzeitreihen“, „Monatsverhältnisse“, „Wochentagswerte“ und „Bestandsentwicklung“ wird die Konjunkturanpassung nicht übernommen, weil diese Analyseschritte i.R. der SRP nicht einzeln für eine Plausibilitätsbewertung i.S. des § 158 Abs. 2 Nr. 1 AO herangezogen werden.

Bei der Ermittlung einer sachgerechten Schätzungshöhe – *für den Fall der eingetretenen Amtspflicht der Schätzung gem. § 162 AO* – hat die Konjunkturanpassung keinen Einfluss, weil eine Übertragung auf die entsprechenden Tabellenblätter „80 %-Quantile“ und „Schätzung“ nicht erfolgt (s. [Pkt. 4.8.](#)).

4.2.9.5. Niveaueinpassung eines Grafen

In einigen Fällen führt die automatische Größenanpassung durch eine zweite Y-Achse bei doppelskalierten Zeitreihenvergleichen nicht dazu, dass die beiden Grafen überwiegend aufeinander liegen und ihre Abhängigkeit somit besser beurteilt werden kann. Für diese Ausnahmen verfügt der einzelskalierte Zeitreihenvergleich auf dem Tabellenblatt „Monatszeitreihen“ über eine Umschaltfunktion auf quasi-prozentuale⁴¹⁹ Darstellung mit der zusätzlichen Möglichkeit, die **erste Datenreihe im Diagrammniveau anzupassen**, bis eine Bewertung der Gesamtschlüssigkeit i.S. des § 158 Abs. 2 Nr. 1 AO seriös möglich wird (s. *das Beispiel unter [Pkt. 4.2.5.2.](#)*).

Da die Niveaueinpassung der ersten Datenreihe i.R. des einzelskalierten Zeitreihenvergleichs keine Auswirkung auf die Größen anderer Auswertungen hat, ist ein warnender Hinweis auf anderen Analyseblättern bzw. für eine Schätzung – *unter eingetretener Amtspflicht der Schätzung gem. § 162 AO* – nicht erforderlich.

4.2.10. Kosten- und weitere Zeitreihen

Neben vielfältigen Einsatzmöglichkeiten zur Erstorientierung⁴²⁰ in einem Prüfungsfall sowie zur Bearbeitung von komplexen Einzelprüffeldern (s. [Pkt. 5.1.](#)) ergeben sich auch i.R. von Plausibilitätsprüfungen gem. § 158 Abs. 2 Nr. 1 AO praxisrelevante Ansätze für Zeitreihen-

⁴¹⁸ Anwendungserlass zu § 158 AO.

⁴¹⁹ Die Monatswerte werden dann als Anteile der jeweiligen Jahressummen angegeben.

⁴²⁰ So auch: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1: „Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich.“

analysen über die Verbildlichung des dokumentierten Warenflusses hinaus, wie insbesondere die Gegenüberstellung von Erfolgs- und (anderen) variablen Kostengrößen.

Die Fokussierung der Finanzrechtsprechung auf die gegenüber Störeinflüssen anfälligen Aufschlagsatzzeitreihen⁴²¹ (s. den Vergleich der gängigsten Zeitreihenvarianten im Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) ist untrennbare Grundlage für die festgelegten **Anwendungsbeschränkungen**⁴²². Dagegen hat Wolenski⁴²³ mit dem Beispiel eines bilanzierenden Unternehmens, welches Fremdleistungen auf Abruf einkauft und weiterveräußert, plastisch demonstriert, dass auch in anders gelagerten Fällen offensichtlich werden kann, ob „es nach [einer] Verprobung [...] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt“⁴²⁴. Nach der dortigen Abb. lassen sich in zwei Jahren zum jeweiligen Ende hin Fremdleistungseinkäufe (fast) in Millionenhöhe ablesen, obwohl keine – auch nur annähernd – gleichwertigen Verkäufe folgen. Da zudem jeweils keine „teilkünftigen Arbeiten“ bilanziert worden sind, verbleibt nur die logische Schlussfolgerung, dass die Erlöserfassung unvollständig sein muss. Folgerichtig merkt eine Autorin aus der wissenschaftlichen Betriebswirtschaftslehre zur „Anwendung von Zeitreihenvergleichen [...] im Hinblick auf die Pauschalisierung der Methode“ folgendes an: „Zusammengefasst liegen eine Reihe geeigneter Verfahren im Rahmen der Zeitreihenanalyse vor. [...] Jede Methode bietet sowohl Vor- als auch Nachteile. Daher ist es sinnvoll mehrere Methoden zu verwenden. Je nach vorliegendem Sachverhalt werden die entsprechenden Verfahren ausgewählt. [...] Durch die Vernetzung von Ergebnissen verschiedener Analysen kann eine Indizienverstärkung entstehen, um so eine höhere Sicherheit zu erreichen.“⁴²⁵

⁴²¹ In den ausführlichen Zeitreihenbeschreibungen der Finanzrechtsprechung dominiert mit der betreffenden 10-Wochen-Gleitschlittenzeitreihe eine Aufschlagsatzvariante wie z.B. in: BFH v. 25.3.2015, BStBl. II 2015, 743, FG Münster v. 19.8.2004 – 8 V 3055/04, FG Düsseldorf v. 15.2.2007 – 16 V 4691/06 oder FG Köln v. 27.1.2009 – 6 K 3954/07. Unter Beachtung der Verbreitung der SRP über fast alle Bundesländer und der Tatsache, dass die Gleitschlittenmodelle i.R. der SRP *nicht* zur Plausibilitätsbeurteilung herangezogen werden (s. [Pkt. 4.2.6.](#)), kann die kleine Auswahl an Finanzgerichtsverfahren nicht als repräsentativ für die Anwendung der Zeitreihenanalytik in der Finanzverwaltung angesehen werden. Das FG Hamburg (v. 31.10.2016 – 2 V 202/16, Rz 48) beschreibt entsprechend konkret die Andersartigkeit des doppelskaliert größenangepassten Zeitreihenvergleichs als zentrale SRP-Variante für die Schlüssigkeitsbeurteilung: „Mit Hilfe der doppelskalierten Monatszeitreihe hat die Betriebsprüfung die Daten des Wareneinkaufs den Umsatzerlösen in zwei Einzelgrafiken gegenüber gestellt. Bei dieser Methode tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander. Aus kaufmännischer Sicht wird in einem Unternehmen auf jede Veränderung der Nachfragesituation (annähernd) adäquat mit einer Erhöhung bzw. Absenkung der Einkaufssumme reagiert“.

⁴²² BFH v. 25.3.2015, BStBl. II 2015, 743, Leitsatz 1: „Die Durchführung eines Zeitreihenvergleichs setzt voraus, dass im Betrieb das Verhältnis zwischen dem Wareneinsatz und den Erlösen im betrachteten Zeitraum weitgehend konstant ist.“

⁴²³ Wolenski, „Das Zeitreihenurteil des X. Senats des BFH vom 25. 3. 2015“, Stbg 6/2016, 268, 271.

⁴²⁴ Anwendungserlass zu § 158 AO.

⁴²⁵ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 323 u. 326.

Gerade weil „[e]lektronische Kassensysteme [...] durch Umprogrammierung in nahezu beliebiger Weise manipulierbar [sind und] von derartigen Manipulationsmöglichkeiten [...] Teile der betrieblichen Praxis nach dem Erkenntnisstand des Senats durchaus Gebrauch [machen]“⁴²⁶, setzt eine effektive Prüfung von Besteuerungsgrundlagen auch Ansätze voraus, die deutlich schwieriger (systematisch) zu unterwandern sind [s.a. den Abschnitt „raffinierte (digitale) Manipulationswerkzeuge“ unter [Pkt. 2.1.](#)], wie bspw. **Kostenzeitreihen**, worauf auch der Bundesrechnungshof allgemein hinweist: „Nur IT-gestützte *quantitative* Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.“⁴²⁷. Der generelle Plausibilitätsansatz des § 158 Abs. 2 Nr. 1 AO ist nicht gesetzlich beschränkt und lässt sich mit derselben Schlüssigkeitsfrage übertragen: „Orientieren sich die Unternehmensleitenden mit dem Aufwand für variable Kosten an der Ertragssituation, um den Gewinn zu maximieren?“

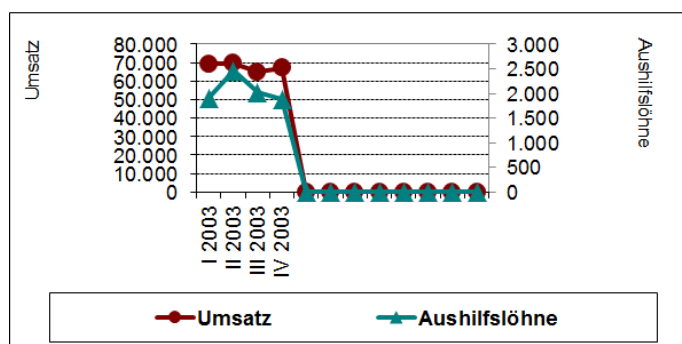
Wegen seiner Stabilität gegenüber alltäglichen Störeinflüssen (s. den Vergleich der gängigsten Zeitreihenvarianten im Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) ist der doppel-skaliert größenangepasste Zeitreihenvergleich auch für einen kostenbezogenen Einsatz gut geeignet. Bei der Auswahl der zur Gegenüberstellung heranzuziehenden Aufwandspositionen ist allerdings darauf zu achten, dass viele Kostenfaktoren (zeitweise) durch Alternativen ersetzt werden können (Substitution). Deshalb bietet sich für Kostenzeitreihen besonders die

Kontonummer	(Alle)
Kontobezeichnung	Aushilfslöhne
Zeilenbeschriftungen	Summe von SoHa
04.01.2003	148,8
06.01.2003	489,1
04.02.2003	758,8
04.03.2003	500,8

variable Arbeitsweise mittels Pivot-Tabelle auf einem der Tabellenblätter „eigene Notizen 1[bzw. 2]“ an (s. den Abschnitt „fortgeschrittene Arbeitsweise der SRP-Vorlagen“ unter [Pkt. 4.](#)). Über die [Berichtsfiler](#) bzw. Datenschnitte⁴²⁸ können die ausgewählten Kostenkonten

nun jederzeit ausgetauscht oder ergänzt werden.

Dieser Zeitreihenvergleich „Aushilfslöhne zum Umsatz“ offenbart im Beispielsfall „Dönerimbiss“, welcher bereits durch unschlüssige Aufzeich-



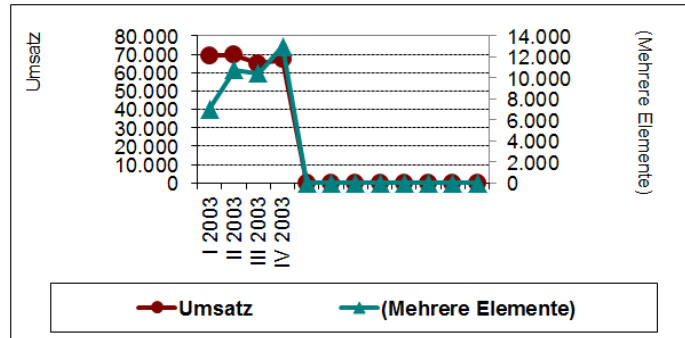
⁴²⁶ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

⁴²⁷ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

⁴²⁸ Während die älteren Berichtsfiler voneinander unabhängig sind, erkennen Datenschnitte automatisch Zusammenhänge. So zeigt der Berichtsfiler „Kontonummer“ im Beispielsfall keine Einschränkung an, obwohl über die Kontobezeichnung nur die Aushilfslöhne ausgewählt wurden. Bei Datenschnitten würde die Auswahl unter „Kontobezeichnung“ auf die Anzeige zur Kontonummer durchwirken.

nungen zum Warenfluss aufgefallen ist (s. [Pkt. 4.2.5.2.](#) sowie [4.2.8.](#)), dass sogar auf der unempfindlichen (s. die SRP-Zeitreihensystematik „vom groben Überblick zu den Details“ unter [Pkt. 4.2.2.](#)) Übersichtsebene der „Quartalsrelationen“ kein erkennbarer Zusammenhang zwischen der (recht) konstanten Absatzentwicklung und dem betriebsbezogen intensiven variablen Kostenfaktor besteht.

Zur Absicherung, ob diese Regelstörung auf eine Umstellung im Bereich der Personalanstellung zurückzuführen sein könnte, wird der Berichtsfilter mit der zusätzlichen Auswahl der Vergleichskosten „Löhne und Gehälter“ ausgeweitet. Der aktualisierte



doppelskalierte Zeitreihenvergleich klärt jedoch umgehend auf, dass diese mögliche Strukturveränderung⁴²⁹ im Betrieb nicht ursächlich für das unschlüssige Verhältnis der beiden Datenreihen ist.

Mit der Benennung der verwendeten Daten auf dem Tabellenblatt „Datenquelle“, den ausdrucksorientierten Auswertungen der SRP-Vorlage BWA und einem Hinweis auf den erweiterten kostenbezogenen Zeitreihenvergleich im Bemerkungsfeld lässt sich die ergänzende Untersuchung zur Schlüssigkeit der Bestimmungsaufzeichnungen i.S. des § 158 Abs. 2 Nr. 1 AO nachvollziehbar dokumentieren. Im Fall konkreter Einlassungen⁴³⁰ können die Argumente jederzeit interaktiv (s. [Pkt. 4.2.9.](#)) eingepflegt und ihre Auswirkungen auf die Zeitreihenergebnisse überprüft werden. Auf diese Weise lässt sich der Mitwirkungsgehalt⁴³¹ zwischen wichtigen neuen Informationen für das Gesamtbild der Bestimmungsaufzeichnungen einerseits und Schutzbehauptungen andererseits differenzieren, was als allgemeinverständliches Qualitätsmerkmal der Sachaufklärung indiziell entscheidungserheblich ist.

Eine weitere erfolgversprechende Option zur verbesserten Aufdeckung systematischer Einnahmenunterdrückungen [s.a. den Abschnitt „raffinierte (digitale) Manipulationswerkzeuge“

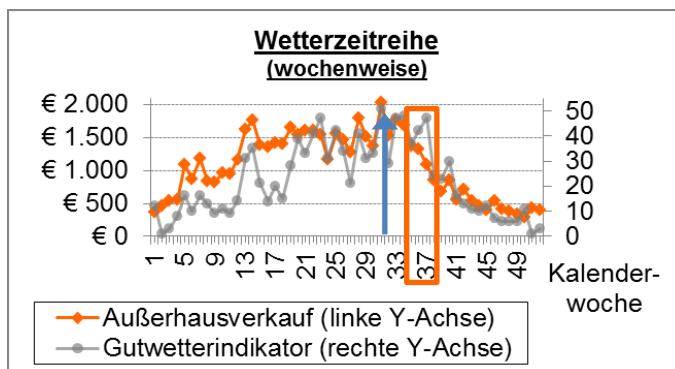
⁴²⁹ BFH v. 25.3.2015, BStBl. II 2015, 743, 1. Leitsatz: „Es darf zudem im maßgebenden Zeitraum nicht zu solchen Änderungen in der Betriebsstruktur gekommen sein, die – nicht anderweitig behebbar – wesentliche Unsicherheiten bei der Aufstellung und Interpretation des Zahlenwerks mit sich bringen.“

⁴³⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen.“

⁴³¹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 9: „§ 162 II i.V.m. § 158 orientieren das Beweismaß an der Sphärenverantwortlichkeit [...] Das Beweismaß hängt mithin davon ab, wie viel und in welcher Form der Stpfl. aus seiner Verantwortungssphäre zur Sachaufklärung beiträgt und wieweit die FinBeh. auf dessen Mitwirkung im konkreten Einzelfall angewiesen ist.“

unter [Pkt. 2.1.](#)] ergibt sich aus **Zeitreihenvergleichen mit externen Daten**. Während sich Unternehmensdaten innerhalb eines Betriebsverwaltungssystems zusammenhängend unauffällig manipulieren lassen, ist dies im Verhältnis zu externen Einflussfaktoren nur möglich, wenn diese Informationen eingespeist werden (können). Zudem sind heutzutage viele Daten zu Faktoren, die den Erfolg von Unternehmen und damit die Handlungslogik beeinflussen, frei im Internet zugänglich und können für Plausibilitätsuntersuchungen herangezogen werden. Ähnlich der interaktiven Bereinigung von einseitigen Dieselpreiseffekten bei einem Taxiunternehmen (s. [Pkt. 4.2.9.4.](#)) muss im Fall einer Beanstandung zwar die Anwendbarkeit der externen Größen auf den einzelnen Betrieb überprüft werden, der berechnete Ruf nach wirksameren Prüfungsmethoden gegen moderne Manipulationen⁴³² unterstützt aber diese Ansätze, die denklösig eine höhere Aufdeckungswahrscheinlichkeit erreichen.

In diesem Beispiel eines Zeitreihenvergleichs werden die Wetterdaten zur Schlüssig-



keitsprüfung der Erlöse eines Eiscafés eingesetzt. Dazu wurden die Außerhausverkäufe wochenweise aufsummiert, um die Informationsfülle der grafischen Auswertung zu begrenzen und dennoch die erwarteten schnellen Nachfragereaktionen auf Wetterveränderungen erkennen zu können. Au-

ßerdem sind die absatzsteigernden Wetterparameter zur einfacheren Interpretierbarkeit⁴³³ in einem Gutwetterindikator⁴³⁴ zusammengefasst und als Wochendurchschnitt berechnet worden. Dem Ergebnisdiagramm ist zu entnehmen, dass die dauerhafte Wetterverbesserung ca.

⁴³² BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 54: „Denn auch umgekehrt verschafft sich ein Teil der Steuerpflichtigen durch ausgefeilt geplante Doppelverkürzungen und/oder den Einsatz von Manipulationssoftware, die die Aufdeckungsmöglichkeiten herkömmlicher Prüfungsmethoden minimieren, technische Vorteile gegenüber der Finanzverwaltung [...] Hierauf darf und muss die Außenprüfung im Interesse der Wahrung der Gesetzmäßigkeit und Gleichmäßigkeit der Besteuerung (Art. 3 Abs. 1, Art. 20 Abs. 3 GG) reagieren, und zwar auch im Wege der Entwicklung und Anwendung neuartiger Prüfungsmethoden.“

⁴³³ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. Die Analyse statistischer Daten beginnt häufig mit der Wahl einer geeigneten Darstellungsform des Erhebungsbe-fundes. Die korrekte Interpretation solcher Grafiken ist allerdings nicht trivial und erfordert spezielle kognitive Fähigkeiten.“

⁴³⁴ Um den erwarteten stärkeren Einfluss einer zusätzlichen Sonnenstunde gegenüber einer Temperaturerhöhung um 1° C. auf die Wettergüte zu beachten, wird ein Gutwetterindikator mit folgender Formel angesetzt: $\text{Sonnenstunden} \times 3 + \text{Durchschnittstemperatur}$. Die Prüfungserfahrungen mit diesem Ansatz haben gezeigt, dass offensichtliche Unschlüssigkeiten bei einer anderweitigen Gewichtung im Gutwetterindikator gleichsam erhalten bleiben.

ab der 20. Kalenderwoche keine weitere Absatzsteigerung ausgelöst hat⁴³⁵, obwohl z.B. die 31. Kalenderwoche zusätzliches Steigerungspotential aufzeigt. Als punktuelle Implausibilität wird die **zweiwöchige Gutwetterperiode um die 37. Kalenderwoche** nicht von einer Zunahme des Außerhausverkaufs beantwortet.

Für eine qualifizierte Beurteilung derartiger Auffälligkeiten i.S. der Plausibilität von Besteuerungsgrundlagen gem. § 158 Abs. 2 Nr. 1 AO ist eine Aktualisierung des Zeitreihenvergleichs nach konkreten Einlassungen⁴³⁶ mit neuen gesamtbildrelevanten Informationen jederzeit interaktiv (s. [Pkt. 4.2.9.](#)) möglich, sodass echte Mitwirkung von reinen Schutzbehauptungen (weitgehend) unterschieden werden kann⁴³⁷. Diese Optionen führen – im Zusammenhang mit der gemeinsamen Sachaufklärungsverpflichtung mindestens indirekt – zu einer allgemeinverständlichen Verbesserung von Mitwirkungs- und Schlüssigkeitsbewertungen.

4.2.11. Beweiskraft von (regelbasierten) Zeitreihenanalysen

Im Gegensatz zum effektiven Erkennen und systematischen Eingrenzen von Prüffeldern⁴³⁸ (s. das Beispiel im Bereich „Eingrenzung aufgefundener Regelstörungen zu ursächlichen Sachverhalten“ unter [Pkt. 4.2.2.](#)) sollen (Gesamtbild-)Verprobungen wie Zeitreihenanalysen i.S. des § 158 Abs. 2 Nr. 1 AO die Vergangenheit anhand von Unterlagen, Auskünften und Eindrücken zur Beurteilung der Schätzungsnotwendigkeit nach § 162 Abs. 2 S. 2 AO wiederherstellen.⁴³⁹ Da Informationen zur Vergangenheit niemals vollständig und objektiv sein können, ergeben sich grundsätzlich **Sachverhaltsbeschreibungen mit Unsicherheiten**. Beim typischen Einsatzgebiet von Verprobungen „Vollständigkeit von (Bar-)Einnahmen“ sind diese Unsicherheiten besonders groß, weil aufgrund der Komplexität der entscheidungsrelevanten Gesamtsituationen keine einfachen Ursache-Wirkungs-Beziehungen be-

⁴³⁵ Nach den „Fuzzy-Theorien“, die in der SRP-Zeitreihensystematik berücksichtigt werden, wäre dieser Trend unter längeren Zeitreihenperioden besser zu erkennen (s. [Pkt. 4.2.2.](#)).

⁴³⁶ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen. Den Stpfl. seinerseits verpflichten die Bedenken etc. zugleich, durch die Erfüllung seiner Mitwirkungspflichten (§ 90) zu eben dieser Sachverhaltsaufklärung beizutragen“.

⁴³⁷ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 9: „§ 162 II i.V.m. § 158 orientieren das Beweismaß an der Sphärenverantwortlichkeit [...] Das Beweismaß hängt mithin davon ab, wie viel und in welcher Form der Stpfl. aus seiner Verantwortungssphäre zur Sachaufklärung beiträgt und wieweit die FinBeh. auf dessen Mitwirkung im konkreten Einzelfall angewiesen ist.“

⁴³⁸ BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1: „Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich.“

⁴³⁹ Anwendungserlass zu § 158 AO: „Die Vorschrift enthält eine gesetzliche Vermutung. Sie verliert ihre Wirksamkeit mit der Folge der Schätzungsnotwendigkeit nach § 162 AO, wenn es nach [einer] Verprobung [...] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt.“

stehen⁴⁴⁰. „Schlussfolgernd kann bei der Prüfung von (steuerlichen) Daten keine (100 %) Sicherheit und somit kein Beweis [...] erfolgen. [...] Nach dem logischen Menschenverstand [kann sie] nur zu Wahrscheinlichkeitsaussagen führen.“⁴⁴¹

Um dennoch in Gerichtsverfahren und Prüfungen Urteile bzw. ausreichend sichere Beurteilungen zu ermöglichen, werden an die volle richterliche Überzeugung geringere Anforderungen⁴⁴² als an den naturwissenschaftlichen Beweis gestellt.⁴⁴³ Für den Bereich des Besteuerungsverfahrens sieht das Gesetz mit **§ 158 Abs. 2 Nr. 1 AO (Verifikationsprinzip)** auf Seiten der Finanzverwaltung eine *darüber hinausgehende* Absenkung der erforderlichen Aussagegewissheit vor (s. [Pkt. 4.6.](#)), denn andernfalls ließe sich die (annähernde) Gleichbehandlung aller Stpfl. gem. Art. 3 GG in einem Massenverfahren nicht durchsetzen⁴⁴⁴. In den allgemein titulierten „Verwaltungsgrundsätze[n] 2020“ wird dieses reduzierte Beweismaß für das ähnlich komplexe – und damit vergleichbar unsichere – Verprobungsgebiet „Fremdüblichkeit von Verrechnungspreisen“ wie folgt konkretisiert: „Voraussetzung einer Einkünfteberichtigung ist vielmehr, [...] dass der von der Finanzbehörde ermittelte Verrechnungspreis zumindest wahrscheinlicher ist.“⁴⁴⁵

Die vom BFH⁴⁴⁶ zur Schätzung – und der Argumentation folgend ebenso zur Verprobung – bevorzugten **herkömmlichen Methoden** „Vermögenszuwachs-“, „Geldverkehrsrechnung“ und „Kalkulation“ setzen selbst bei ordnungsmäßigen Besteuerungsgrundlagen auf unvollständigen Ausgangsinformationen auf, indem bspw. private Geldzu- und -abflüsse

⁴⁴⁰ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 15: „Unser Dilemma im Umgang mit der Komplexität unserer Welt lässt sich darauf zurückführen, dass wir wohl darin ausgebildet wurden einfache logische Schlüsse zu ziehen und nahe liegende Ursache-Wirkungs-Beziehungen zu definieren. [...] Simple Ursache-Wirkungs-Beziehungen gibt es nur in der Theorie, nicht in der Wirklichkeit.“

⁴⁴¹ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324 u. 323.

⁴⁴² Vgl. z.B.: BFH v. 20.5.2010, BStBl II 2010, 1022, Rz. 13: „[...] wobei der Richter nicht eine von allen Zweifeln freie Überzeugung anstreben darf, sich in tatsächlich zweifelhaften Fällen vielmehr mit einem für das praktische Leben brauchbaren Grad von Gewissheit überzeugen muss“.

⁴⁴³ Tlw. werden die unvermeidbaren Unsicherheiten bei der Reproduktion der Vergangenheit in der Literatur offenbar grundsätzlich nicht akzeptiert (vgl. z.B.: Prinz, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4, M5: „Wahrscheinlichkeitstheoretisch geformte Algorithmen können Steueransprüche nicht begründen“), wodurch eine (teleologisch) sachgerechte Einordnung des Beweises im juristischen Sinn und des „Beanstandungsanlasses“ gem. § 158 AO unmöglich wird.

⁴⁴⁴ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

⁴⁴⁵ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 73.

⁴⁴⁶ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 63 u. 64 : „Allein die Ergebnisse eines Zeitreihenvergleichs lassen aufgrund der dieser Verprobungsmethode innewohnenden methodenbedingten Unsicherheiten aber noch keinen sicheren Schluss auf das Vorliegen und den Umfang auch materieller Unrichtigkeiten der Buchführung zu. In diesen Fällen sind andere Schätzungsmethoden, die auf betriebsinternen Daten aufbauen oder in anderer Weise die individuellen Verhältnisse des jeweiligen Steuerpflichtigen berücksichtigen (z.B. Vermögenszuwachs- oder Geldverkehrsrechnung, Aufschlagkalkulation) grundsätzlich vorrangig heranzuziehen.“

bzw. der Warenverderb nicht aufgezeichnet werden müssen. Zudem ist das Risiko ordnungswidriger Unterlagen in den betreffenden Fällen mit überwiegend Bareinnahmen besonders hoch.⁴⁴⁷ Das Modell „Schwarzeinnahmen zur Bezahlung von Schwarzlöhnen“ kann unter diesen Umständen mit den Mitteln einer Betriebsprüfung i.d.R. nicht ausgeschlossen werden. I.R. von Kalkulationen liegen häufig keine vollständigen, vertrauenswürdigen Informationen zu Verkaufspreisen, Sonderverkäufen, Wareneinkäufen etc. vor. Die entscheidungsrelevanten Größen „Geldverbrauch“ bzw. „Sollumsatz“ müssen in der Konsequenz aus unsicheren Zahlen rekonstruiert werden, sodass die berechneten Zielgrößen sicherer wirken, als sie es sein können⁴⁴⁸: „Die herkömmlichen Methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...]. Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten in den wiederhergestellten Daten.“⁴⁴⁹

Durch **sorgfältige Ermittlungen der Finanzbehörde**⁴⁵⁰ sollen die Unsicherheiten soweit möglich begrenzt werden. Nach dem Verifikationsprinzip haben aber auch die Mitwirkungsbereitschaft und -qualität der Stpfl. eine wesentliche Auswirkung auf das erforderliche Beweismaß für das Auslösen der Schätzungspflicht⁴⁵¹; diesbezügliche Warnungen vor

⁴⁴⁷ Teutemacher, „Handbuch zur Kassenführung – Praxishandbuch für die rechtssichere Umsetzung“, NWB-Verlag, 2. Auflage 2018, 3: „Betriebsprüfungen bei Betrieben, die überwiegend Bargeschäfte tätigen, führen in 95 % der Fälle zu dem Ergebnis, dass die Kassenführung verworfen wird, weil die vorgelegten Bücher bzw. Aufzeichnungen sowie Daten nicht den Ordnungsvorschriften entsprechen und damit die sachliche Richtigkeit in Zweifel gezogen wird“ mit Verweis auf Kalischke, „Betriebe bangen vor Steuerprüfern – Beamte verwerfen immer mehr Kassenbuchführungen“, Westfälische Nachrichten v. 15.3.2013.

⁴⁴⁸ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Nomos, 2018, 56-57 u. 65: „Bei der Verfügbarkeitsheuristik schätzen Individuen die [...] Wahrscheinlichkeit für das Eintreten eines Ereignisses unter Unsicherheit danach ein, wie leicht bzw. schwer es ihnen fällt, damit verbundene Erscheinungen aus ihrem Gedächtnis abzurufen. Dabei werden Ereignisse für wahrscheinlicher erachtet, wenn sie sich leichter vorgestellt werden können und sich zu ihnen eine direkte Verfügbarkeit herstellen lässt. [...] Im Rahmen juristischer Entscheidungen scheint es unerlässlich, von diesen Entscheidungsverzerrungen Kenntnis zu haben, da eine verzerrte Entscheidung den Denkgesetzen der Logik widerspricht [...]“.

⁴⁴⁹ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324.

⁴⁵⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Der Wortlaut des § 158 (Anlass zu beanstanden) mag dafür sprechen, dass bereits bloße Beweisanzeichen, Zweifel oder Bedenken ausreichen, um die Vermutung auszuräumen oder zu entkräften. Es ist jedoch nicht so, dass die FinBeh. nur irgendwelche Beanstandungen auszusprechen hätte und der Stpfl. dann gehalten wäre zu beweisen, dass die formell richtigen Buchungen sachlich nicht falsch sind. Ein solcher Negativbeweis kann unmöglich verlangt werden. Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen.“

⁴⁵¹ BMF-Schreiben v. 3.12.2020 „Verwaltungsgrundsätze 2020“, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw.

„überzogenen Mitwirkungspflichten“⁴⁵² laufen dem gesetzlichen Auftrag, „alle für den Einzelfall bedeutsamen [...] Umstände zu berücksichtigen“⁴⁵³ zuwider und verkennen, dass das gemeinsame interaktive (s. [Pkt. 4.2.9.](#)) Ergründen von Regelstörungen mit der SRP-Vorlage BWA erfahrungsgemäß wesentlich weniger aufwendig als bspw. das Nachvollziehen einer ausreichend detaillierten Ausbeutekalkulation ist. Bezogen auf Zeitreihenanalysen bedeutet das sorgfältige Erarbeiten eines Gesamtbilds, eine nachvollziehbare und möglichst *robuste*⁴⁵⁴ Schlüssigkeitsfrage voranzustellen, eine *geeignete* Zeitreihenvariante (s. den Vergleich der gängigsten Zeitreihenmodelle im Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) für die Untersuchung zu wählen sowie *mögliche Störeinflüsse*, welche das Zeitreihenergebnis verzerrt haben können, zu beachten.⁴⁵⁵

Im Hinblick auf die genannten Kriterien eines sorgfältig erarbeiteten – und damit aussagekräftigen – Gesamtbilds gründet sich die **SRP-Zeitreihensystematik** auf folgende **Analysegrundsätze und Annahmen**.⁴⁵⁶

erhöhen. [...] Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u.a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“ Tipke/Kruse, AO/FGO, **171. Lieferung 7/2022**, § 158 AO, Rz. 9: „§ 162 II i.V.m. § 158 orientieren das Beweismaß an der Sphärenverantwortlichkeit [...] Das Beweismaß hängt mithin davon ab, wie viel und in welcher Form der Stpfl. aus seiner Verantwortungssphäre zur Sachaufklärung beiträgt und wieweit die FinBeh. auf dessen Mitwirkung im konkreten Einzelfall angewiesen ist.“

⁴⁵² Vgl. z.B.: *Steinhauß*, „Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388, 389.

⁴⁵³ § 88 Abs. 1 S. 2 AO.

⁴⁵⁴ Je mehr betriebsinterne und externe Einflussfaktoren auf die Entwicklung oder Abhängigkeit von Unternehmensgrößen einwirken können, umso schwieriger ist eine aussagekräftige Plausibilitätsbeurteilung. So muss bspw. bei einem Zeitreihenvergleich aus Erlösen und einem variablen Kostenfaktor stets auch die Substituierbarkeit (Austauschbarkeit) beachtet werden, wie bei der Aushilfslöhne-Zeitreihe im Abschnitt „Kostenzeitreihen“ unter [Pkt. 4.2.10.](#)

⁴⁵⁵ *Sell*, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 27: „Die Qualität des Zeitreihenvergleichs wird maßgeblich dadurch bestimmt, wie sorgfältig die betriebswirtschaftlichen Größen erhoben werden und wie genau betriebliche Besonderheiten geprüft werden. [...] Dies ist keinesfalls ein Argument gegen die Methoden, sondern zeigt lediglich die bei jeder Verprobungsmethode obligatorische Verpflichtung zur sorgfältigen Anwendung ebendieser.“

⁴⁵⁶ Häufig werden diese Merkmale der SRP-Zeitreihensystematik in der kritisierenden Literatur gar nicht erwähnt (vgl. z.B.: *Beyer*, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, oder *Bleschick*, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353), oder sie werden zwar beschrieben, aber die Schlussfolgerungen gründen sich dann ausschließlich auf Zitate aus der bis-

- ☑ Die **Schlüssigkeitserwartung**, dass in Unternehmen auf veränderte Erfolgssituationen mit der Anpassung variabler Aufwandsfaktoren wie z.B. mit dem Warennachkauf auf den Absatz reagiert wird, stellt eine **nachvollziehbare und direkte Handlungslogik** dar.⁴⁵⁷ Infolge der Gegenüberstellung von zwei betrieblichen Datenreihen durch ein geeignetes (s. *den nächsten Aufzählungspkt.*) Zeitreihenmodell in der SRP-Vorlage BWA werden keine möglichen Störfaktoren vermischt, sondern es können beobachtete Regelauffälligkeiten gezielt vertieft werden. Demgegenüber wird in der Literatur mit Blick auf den Schlüssigkeitsansatz „konstanter Aufschlagsatz“, welcher in der Finanzrechtsprechung zurzeit noch dominiert⁴⁵⁸, vor vielfältigen möglichen Störeinflüssen gewarnt⁴⁵⁹.
- ☑ Die **zentrale Plausibilitäts-Zeitreihenvariante** „doppelskaliert größenangepasster Zeitreihenvergleich“ der SRP-Zeitreihensystematik ist im Gegensatz zu anderen gängigen Zeitreihenarten sehr **stabil gegenüber alltäglichen Sondereinflüssen**, wie die Praxisbeispiele zur Einnahmen-Überschuss-Rechnung gem. § 4 Abs. 3 EStG (s. *Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.*), zu lange haltbaren Waren (s. *Abschnitt „Robustheit des doppelskalierten Zeitreihenvergleichs“ unter Pkt. 4.2.*) und zu starken einseitigen Preiseffekten (s. *Pkt. 4.2.9.4.*) nachvollziehbar belegen: In all diesen für Zeitreihenanalysen schwierigen Fällen ist die Abhängigkeit der betrachteten Betriebsfaktoren auf monatlich aufsummierter Ebene dennoch problemlos ablesbar.
- ☑ Die **SRP-Zeitreihensystematik** verwendet mehrere Zeitreihenmodelle mit ihren jeweiligen Stärken in zielgerichteter Kombination, um – dem Prüfungsduktus „vom groben Überblick zu den Details“ folgend – den Zusammenhang der untersuchten Datenreihen zu verstehen (s. *Pkt 4.2.2.*). Der Einstieg über die hochgradig verdichtete Quartalsebene stellt mit ihrer Übersichtlichkeit das Erkennen systemtischer Besonderheiten wie einer permanent verzögerten Erfassung sicher (s. *Pkt. 4.2.4.*). Das schrittweise Vertiefen über den doppelskalierten Zeitreihenvergleich (s. *Pkt. 4.2.5.1.*), die punktgenau eingrenzenden Gleitschlittenzeitreihen (s. *Pkt. 4.2.6.*) und die Einzelperspektiven zu beiden Datenreihen (s. *Pkt. 4.2.7.*) sorgen im Anschluss für **größtmögliche Sicherheit im Erkennen aller gesamtbildrelevanten Sondersachverhalte**.

herigen nicht differenzierenden Finanzrechtsprechung [vgl. z.B.: *Steinhauff*, „Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388].

⁴⁵⁷ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 28: „Es darf einem Gewerbetreibenden durchaus unternehmerisches Handeln unterstellt werden, sodass ein grober Verstoß gegen die betriebswirtschaftliche Handlungslogik nicht plausibel erscheint.“

⁴⁵⁸ Vgl. z.B.: FG Münster v. 19.8.2004 – 8 V 3055/04, FG Düsseldorf v. 15.2.2007 – 16 V 4691/06 oder FG Köln v. 27.1.2009 – 6 K 3954/07. Auch das BFH-Zeitreihenurteil (v. 25.3.2015, BStBl. II 2015, 743) bezieht sich mit der verwendeten Gleitschlittenzeitreihe auf eine Zeitreihenvariante des Aufschlagsatzes.

⁴⁵⁹ Wiggen, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168, 171 f.: „2. Die Ermittlung des periodischen wirtschaftlichen Wareneinsatzes [...] 3. Das Fehlen von periodischen Inventurwerten [...] 4. Preisschwankungen und andere betriebliche Besonderheiten“.

- ☑ Schlussendlich können alle entdeckten bzw. von seiten der Stpfl. dargelegten **Sondervorgänge**, welche Zeitreihenergebnisse eventuell negativ beeinflusst haben können, **nachvollziehbar interaktiv (s. Pkt. 4.2.9.) korrigiert werden**, indem sich einzelne Datensätze (tlw.) löschen oder umdatieren, der Prüfungszeitraum (zeitweise) einschränken, die erste Datenreihe um eine feste Anzahl Tage verschieben, (einseitige) Preiseffekte über Indizes ausgleichen und/oder bei Versagen der automatischen Doppelskalierung eine Niveauanpassung des einen Grafen vornehmen lässt bzw. lassen.

Die hohe Aussagegewissheit einer Plausibilitätsprüfung i.S. des § 158 Abs. 2 Nr. 1 AO mit der SRP ergibt sich nicht nur aus diesen einzelnen allgemeinverständlichen Qualitätsmerkmalen, sondern aus der SRP-Zeitreihensystematik, die alle Vorteile durch einen zielführenden Prüfungsablauf verbindet. Darüber hinaus können die Ergebnisse einzelner Zeitreihenanalysen sowie das Gesamtbild zur Schlüssigkeit durch musterbasierte Häufigkeitsauswertungen unter [Pkt. 4.3.](#) weiter abgesichert werden. **Folgelogisch ordnet das BMF-Schreiben zu „[a]utomationsgestützten[n] quantitative[n] Prüfungsmethoden in der steuerlichen Außenprüfung“ deren Verprobungseignung wie folgt ein: „Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgabenordnung dazu führen, dass eine Schätzung durchgeführt werden muss.“⁴⁶⁰**

In seiner **Rechtsprechung** hat der BFH den Einsatz von Zeitreihenanalysen – scheinbar grundsätzlich – erheblich eingeschränkt⁴⁶¹ und dies mit systemimmanenten Methodenschwächen begründet⁴⁶². Zwar erwähnt der X. Senat die „zahlreichen Varianten eines Zeitreihenvergleichs“⁴⁶³, nimmt mit seiner allgemeinen Beurteilung aber doch keine Differenzierung vor⁴⁶⁴, was von wissenschaftlichen Autorinnen kritisiert wird⁴⁶⁵. Demgegenüber hat das

⁴⁶⁰ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 4.

⁴⁶¹ BFH v. 25.3.2015, BStBl. II 2015, 743, Leitsatz 1: „Die Durchführung eines Zeitreihenvergleichs setzt voraus, dass im Betrieb das Verhältnis zwischen dem Wareneinsatz und den Erlösen im betrachteten Zeitraum weitgehend konstant ist.“

⁴⁶² BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 39 u. 63: „Auch bei einer formell und materiell ordnungsmäßigen Buchführung führt ein Zeitreihenvergleich denklologisch immer zu einem Mehrergebnis gegenüber der Buchführung, da der höchste Rohgewinnaufschlagsatz aller Zehn-Wochen-Perioden des Jahres auf den Wareneinsatz für das gesamte Jahr angewendet wird. [...] Allein die Ergebnisse eines Zeitreihenvergleichs lassen aufgrund der dieser Verprobungsmethode innewohnenden methodenbedingten Unsicherheiten aber noch keinen sicheren Schluss auf das Vorliegen und den Umfang auch materieller Unrichtigkeiten der Buchführung zu.“

⁴⁶³ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 37.

⁴⁶⁴ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 54: „Aus diesen Befunden – so sensibel sie unter den aufgezeigten Aspekten der folgerichtigen Methodik und des effektiven Rechtsschutzes sein mögen – folgt für den Senat auf der Grundlage des derzeitigen Erkenntnisstandes allerdings nicht das Ergebnis, *die Methode* des Zeitreihenvergleichs grundsätzlich zu verwerfen.“

⁴⁶⁵ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 28: „Der BFH ver-

FG Hamburg die Notwendigkeit der Unterscheidung zutreffend beschrieben: „Mit Hilfe der doppelskalierten Monatszeitreihe hat die Betriebsprüfung die Daten des Wareneinkaufs den Umsatzerlösen in zwei Einzelgrafiken gegenüber gestellt. Bei dieser Methode tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander.“⁴⁶⁶ Mit den weiter ausführenden Beiträgen der urteilenden RiBFH ist z.T. der Bezug ihres Urteils zu der im betreffenden Fall verwendete 10-Wochen-Gleitschlitten-Aufschlagsatzzeitreihe klargestellt worden.⁴⁶⁷ Da weder die vorangestellte Plausibilitätsfrage, noch die Zeitreihenvariante, die mehrperspektivische Systematik oder die interaktiven Optionen zur Einarbeitung neuer Erkenntnisse im Fall des BFH-Zeitreihenurteils mit der Arbeitsweise der SRP übereinstimmen, lässt sich der Richtspruch nicht auf die SRP übertragen.

4.3. Musterbasierte Analysen (Häufigkeitsauswertungen)

„Komplexität hat sehr viel mit Vernetzung zu tun, ja kommt erst durch Vernetzung zustande. Komplexe Vorgänge verlangen daher zu ihrem Verständnis ein Denken in Zusammenhängen, das sich an der Struktur organisierter Systeme und ihrer speziellen Dynamik orientiert. [...] Damit treffen wir auf einen der Kardinalfehler in der Beurteilung, was für die Erfassung von Komplexität wichtig ist. Denn hier geht es um das ‚Gesicht‘ der Wirklichkeit, um das Erkennen von Mustern“⁴⁶⁸. Mustererkennung⁴⁶⁹ als Analysemittel setzt i.d.R. auf der **Auszählung von Informationen nach strukturierbaren**⁴⁷⁰ **Eigenschaften** auf. Diese Herangehensweise wird in der Wirtschaftsprüfung mit der ABC-Analyse seit langer Zeit zur Effizienzsteigerung genutzt, indem die Daten eines bestimmten Bereichs nach ihrem Wertinhalt in „sehr wichtige“, „wichtige“ und „weniger wichtige“ differenziert werden. Mit der Variabilität und Schnelligkeit der digitalen Datenanalyse ist die Möglichkeit hinzugekommen, derartige Häufigkeitsauswertungen verbildlicht darzustellen.

tritt hier ein distanziert-kritische Haltung [...], die im Grad ihrer Verallgemeinerung der Methode [Zeitreihenanalyse] nicht angemessen erscheint“. Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 325.

⁴⁶⁶ FG Hamburg v. 31.10.2016 – 2 V 202/16, Rz. 48.

⁴⁶⁷ Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548.

⁴⁶⁸ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 16 u. 21.

⁴⁶⁹ Die farblich hervorhebende Form der Musteranalytik zur einfacheren Wahrnehmung wird auch „Heatmap“ genannt; z.T. wird – anstatt der Logik „blau = normal vs. rot = auffällig“ – auch „grün vs. rot“ (Ampel-Logik) eingesetzt, was im Hinblick auf die verbreitete Rot-Grün-Sehschwäche jedoch problematisch ist.

⁴⁷⁰ Eigenschaften von Besteuerungsdaten sind strukturierbar, wenn sie entweder (bereits) in abzählbarer Unterscheidung vorliegen wie der USt-Schlüssel, die Zahlungsweise etc. oder mittels Einteilung in (Größen-)Klassen unterteilt werden können wie z.B. anhand ihres Wertinhalts.

Spätestens durch diese Optionen „[bilden e]inige der zentralen explorativen Verfahren die Data Mining-Ansätze, die prozessgestützt auffällige Muster und Beziehungen in den großen Datenvolumina ausfindig machen.“⁴⁷¹ Folglich fordert der Bundesrechnungshof „den bundesweit einheitlichen Einsatz IT-gestützter *quantitativer* Prüfungsmethoden in der steuerlichen Betriebsprüfung [... denn t]raditionelle Methoden der steuerlichen Betriebsprüfung, beispielsweise Beleg- und Akteneinsicht, verlieren an Effektivität und Effizienz. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige **Muster** und Beziehungen in großen Datenmengen ersichtlich machen.“⁴⁷² Im BMF-Schreiben zu „[a]utomationsgestützte[n] quantitative[n] Prüfungsmethoden in der steuerlichen Außenprüfung“ werden sowohl deren Eignung zum „Erkennen von Prüffeldern unter Risikogesichtspunkten“ als auch zur „Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung“ sowie als Schätzungsmethoden benannt.⁴⁷³ Allerdings sehen die zitierten wissenschaftlichen Autoren auch eine fehlende Vertrautheit mit diesen **leistungsfähigen Analysemethoden**: „Frühere Studien haben ergeben, dass Abschlussprüfer häufig Defizite in der Fähigkeit von Mustererkennungen aufweisen, [...] Insofern erwächst hierbei ein konkreter Handlungsbedarf, diese Fähigkeiten bei der Aus- und Fortbildung des Wirtschaftsprüfers zu adressieren.“ Die Bildungswissenschaft bestätigt sowohl die Leistungsfähigkeit der Musteranalytik als auch den Bildungsbedarf.⁴⁷⁴

Aus den Natur- und Wirtschaftswissenschaften ist bekannt, dass viele zusammengehörige Daten **gleichmäßige Muster der Häufigkeitsverteilungen** wie z.B. die Normalverteilung in der folgenden Abb.⁴⁷⁵ oder bei Wirtschaftsdaten eher die Logarithmische Normalverteilung⁴⁷⁶ bilden:

⁴⁷¹ Velte/Drews, „Herausforderungen in der Wirtschaftsprüfung durch den Einsatz von Big-Data-Technologien – Welche Impulse muss die Prüfungsforschung geben?“, DB 43/2018, 2581, 2582 u. 2583 (zweites Zitat).

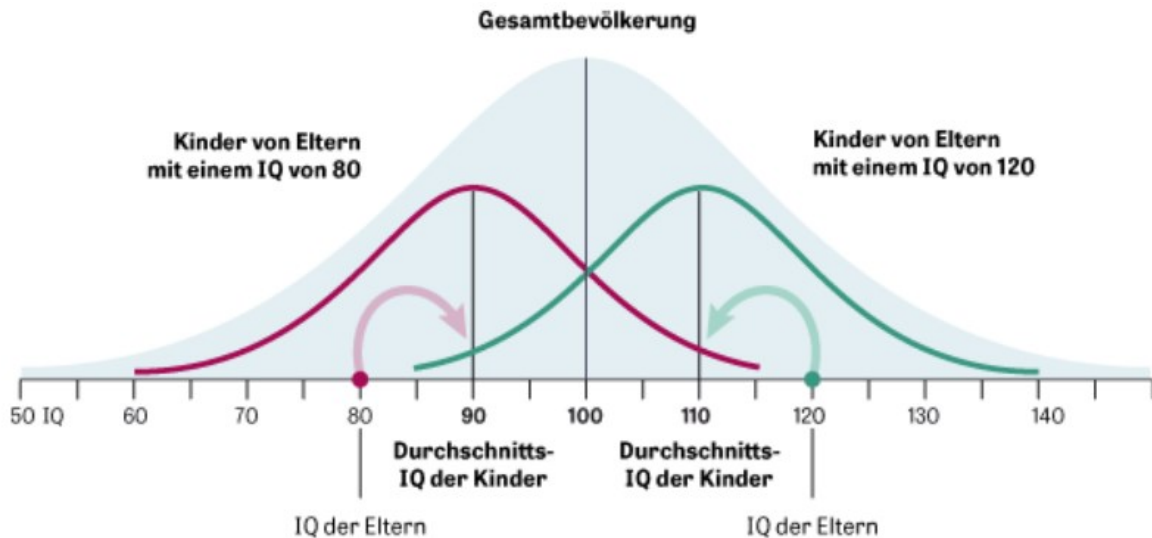
⁴⁷² Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

⁴⁷³ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 1, 4 u. 24 ff.

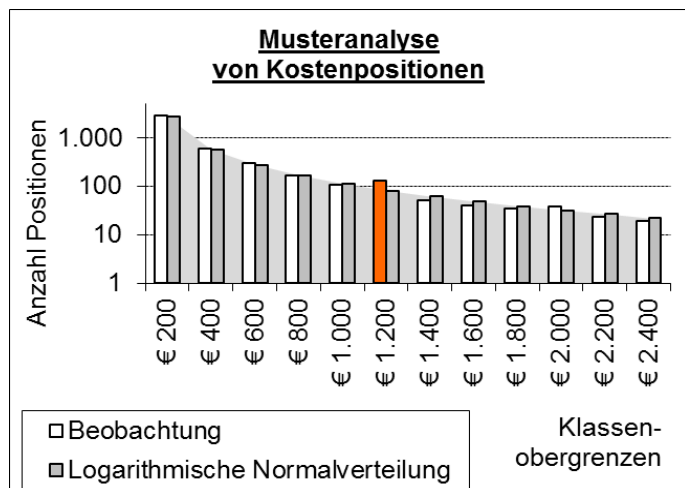
⁴⁷⁴ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. Die Analyse statistischer Daten beginnt häufig mit der Wahl einer geeigneten Darstellungsform des Erhebungsbefundes. Die korrekte Interpretation solcher Grafiken ist allerdings nicht trivial und erfordert spezielle kognitive Fähigkeiten.“

⁴⁷⁵ Zeit online v. 17.6.2015: „Acht Antworten – Warum kluge Eltern nicht unbedingt kluge Kinder bekommen – und sieben weitere Tatsachen über die Vererbung von Intelligenz“, nachzulesen unter: <https://www.zeit.de/2015/23/intelligenz-vererbung-iq>.

⁴⁷⁶ Vgl. z.B.: Limpert/Stahel/Abbt, „Lognormal distributions across the sciences: keys and clues“, BioScience Band 51, Nr. 5, 2001, 341, nachzulesen unter: https://de.wikipedia.org/wiki/Logarithmische_Normalverteilung.



Deshalb haben sich tabellarische Häufigkeitsauswertungen zu *visualisiert* musterbasierten Modellen weiterentwickelt, mit denen eine zügige und intensive Durchdringung großer Datenmengen möglich ist. So beschreibt bspw. die nebenstehende musterbasierte Häufigkeitsauswertung mehrerer tausend Kostenpositionen eines Betriebs deren Größenzusammenhänge. In Gegenüberstellung mit der – aus den Daten abgeleiteten – Logarithmischen Normalverteilung, die für Wirtschaftsdaten überwiegend typisch ist (s. [Pkt. 4.3.4.1.](#)), wird die Gleichmäßigkeit im gleitenden Auslauf der hinterlegten Fläche sichtbar. Durch die zielführende Darstellung mit logarithmischer⁴⁷⁷ Y-Achse werden die Größenklassen mit geringerer „Anzahl Positionen“ betont, wodurch die **nicht mustergetreue Häufung in der Klasse „> € 1.000 bis € 1.200“** auffällt. Mittels Filter- und Sortierfunktionen z.B. in IDEA® können die dort unerwartet geballt vorkommenden Kostenvorgänge unter geringem Aufwand herausgesucht und als betriebsbezogen sinnvolles Vertiefungsfeld abgeprüft werden.



Besonders Kombinationen⁴⁷⁸ von bspw. muster- mit regelbasierten Analysen können innerhalb kurzer Zeit eine hohe Transparenz trotz unprüfbar großer Datenmengen erzielen.

⁴⁷⁷ Eine logarithmische Skalierung belegt die gleichmäßigen Intervalle mit linear ansteigenden Exponenten wie in der Beispielsgrafik mit „10⁰ – 10¹ – 10² – 10³“ (entsprechend „1 – 10 – 100 – 1.000“) und erreicht dadurch eine zunehmende Betonung im Bereich der kleinen Zahlen.

⁴⁷⁸ Sowohl die „Dashboards“ in neueren Versionen von Datenanalyseprogrammen (vgl. z.B. auf der Internetseite des IDEA®-Vertriebs: https://docu.caseware.net/Audit/20.00/AuditTemplate/de/Content/User_Client_File_Setup/t_Using_the_Work_Flow_Panel.htm) als auch die – in Wirtschaft und Beratung immer verbreiteteren – B(usiness-)(ntelligence)-

Deshalb werden die Funktionalitäten beider SRP-Teile (SRP-Vorlagen BWA und Stochastik) in der Prüfungssystematik zusammengeführt (s. *das Beispiel zur Konkretisierung eines auffälligen Zeitreihenvergleichs im Abschnitt „Prüfungsnetz“ unter Pkt. 2.3.*).

Natürliche gleichmäßige Muster in großen Datenmengen werden jedoch dann gestört, wenn Menschen durch eine der folgenden Aktionen in die Verteilungsstruktur eingreifen:

-  Manipulationen⁴⁷⁹,
-  Fehler bei der Erfassung von Daten z.B. durch Verschreiben, falsche Zuordnungen oder Schlüsselungen etc. sowie
-  Gestaltungen i.d.R. in Form von vertraglichen Absprachen, wodurch die Eigenschaften der betroffenen Vorgänge bewusst festgelegt werden.

Auf Basis dieser Erkenntnis sind u.a. **musterbasierte Prüfungsansätze zur Aufdeckung von Manipulationen** entwickelt worden. Nachdem diese in der Bekämpfung von Wirtschaftsverbrechen tlw. schon länger bekannt waren⁴⁸⁰, hat wohl Nigrini⁴⁸¹ den entscheidenden Anstoß gegeben, derartige Verfahren auch in der Wirtschaftsprüfung einzusetzen. Inzwischen sind musterbasierte Prüfungsmethoden zur Aufdeckung von unerwünschten menschlichen Eingriffen in großen Datenmengen sowohl in der Wirtschaftsprüfung⁴⁸² als auch in der Wissenschaft⁴⁸³ vertreten. Da „[e]lektronische Kassensysteme [...] durch Umprogrammierung in nahezu beliebiger Weise manipulierbar [sind und] von derartigen Manipulationsmöglichkeiten [...] Teile der betrieblichen Praxis nach dem Erkenntnisstand des Senats

Programme mit ihrer Grundeinstellung verknüpfter Analysen zielen auf eine effektive Verbindung mehrerer Visualisierungsperspektiven ab; insofern erleichtert die SRP den Einstieg in diese Analyseinstrumente.

⁴⁷⁹ Hill, „The Difficulty of faking Data“, Chance 12/3/1999, 27.

⁴⁸⁰ Vgl. z.B.: Handelsblatt online v. 8.7.2012: „Volkswirte auf Verbrecherjagd: Ob illegaler Waffenhandel oder Steuerhinterziehung – Ökonomen stöbern immer häufiger Missetäter auf. Die forensische Ökonomie ist ein boomender Forschungszweig. Ermittlungsbehörden nutzen die Erkenntnisse der Forscher“, *früher* nachzulesen unter: <https://www.handelsblatt.com/politik/konjunktur/oekonomie/nachrichten/forensische-oekonomie-volkswirte-auf-verbrecherjagd/6838146.html> oder Handelsblatt online v. 30.11.2009: „Auf der Jagd nach Zahlen-Fälschern: Mit einem lange vergessenen mathematischen Gesetz wollen Ökonomen Kollegen überführen, die in wissenschaftlichen Studien Daten manipuliert haben. Denn längst nicht alle Wirtschaftswissenschaftler sind ehrlich – und die Anreize für Betrüger sind groß“, nachzulesen unter: <https://www.handelsblatt.com/politik/konjunktur/oekonomie/nachrichten/oekonomen-auf-der-jagd-nach-zahlen-faelschern/3315384.html>.

⁴⁸¹ Nigrini, „The detection of income tax evasion through an analysis of digital frequencies“, University of Cincinnati Ann Arbor Michigan, 1992; Internetseite von Nigrini: <http://www.nigrini.com/>.

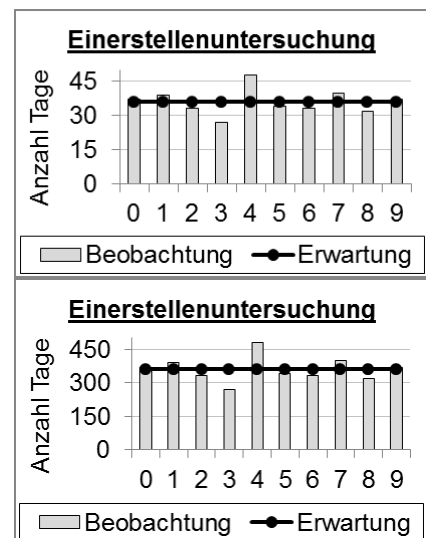
⁴⁸² Vgl. z.B. den Einsatz der gleichmäßigen Verteilungen gem. Benford's Law (s. Pkt. 4.3.2.3.1.) als „Fraud test“ i.R. des „Data analytics process“ der Wirtschaftsprüfungsgesellschaft EY, nachzulesen unter: <https://eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2013/11/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf>.

⁴⁸³ Vgl. z.B. die digitale Spurensuche mit modifizierter Benford-Verteilung (s. Pkt. 4.3.2.3.1.) am Fraunhofer-Institut für Sichere Informationstechnologie auf der Internetseite der Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V.: „Digitale Spurensuche – Wirtschaftsthemen Mai 2011“, nachzulesen unter: <https://www.fraunhofer.de/de/presse/presseinformationen/2011/mai/it-forensik-gegen-wirtschaftskriminalitaet.html>.

durchaus Gebrauch [machen]⁴⁸⁴ und dabei Steuerausfälle von knapp € 2 Mio. durch ein einzelnes Eiscafé entstehen können⁴⁸⁵ (s.a. [Pkt. 2.1.](#)), haben Begriffe wie „Chi-Quadrat-Test“, „Benford-Gesetz“, „Strukturanalyse“ usw. auch in den (Finanz-)Verwaltungen⁴⁸⁶ und dort speziell in der deutschen Betriebsprüfung⁴⁸⁷ Verbreitung⁴⁸⁸ erlangt – verstärkt durch die Einführung der digitalen Außenprüfung 2002.

Eine wichtige Besonderheit der Tendenz unbeeinflusster, zusammengehöriger Datenbestände, gleichmäßige Muster einzunehmen, ist, dass die **Güte der Übereinstimmung mit dem Idealmuster** wie z.B. der Gleichverteilung **von der Datenmenge abhängt**. Diese Verteilungseigenschaft ist aus der Stichprobenprüfung bekannt: So wird die Genauigkeit, mit der Zufallsstichproben ihre Grundgesamtheit richtig wiedergeben, unter zunehmender Stichprobengröße besser (s. die Abb. „*Abhängigkeit der Fehlerstreuung von der Stichprobengröße*“ im Abschnitt „*statistische Stichprobenverfahren*“ unter [Pkt. 2.](#)) – bei gleichmäßiger Stichprobenvergrößerung anfangs stark, dann immer schwächer.

Der Zusammenhang zwischen Datenmenge und Mustergenauigkeit lässt sich an den beiden Beispielsgrafiken zur Ziffernanalyse der Einerstelle aus Bareinnahmen (s. [Pkt. 4.3.2.2.](#)) gut verdeutlichen: Beide Vergleiche zur ausgezählten Ein-Euro-Stelle stimmen nicht mit der idealen Erwartung „Gleichverteilung aller Ziffern“ überein. Dabei zeigen die grafischen Auswertungen für die zwei Datenbestände ein offenbar identisches Abweichungsverhalten mit der deutlich zu häufigen „4“, der vergleichbar unterrepräsentierten „3“ und ansonsten kleineren Unterschieden zur erwarteten Gleichverteilung. An der Y-Achse lässt sich jedoch ablesen, dass im ersten Fall wesentlich weniger Werte bspw. aus einer abgekürzten Außenprüfung untersucht wurden, während im zweiten der zehnfache Datenumfang z.B. aus einer Durchsuchung verwendet werden konnte. Werden die visualisierten Gegenüberstellungen aus beiden Ziffernanalysen mit den Werkzeugen des mathematischen Spezialgebiets



⁴⁸⁴ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

⁴⁸⁵ FG Rheinland-Pfalz v. 7.1.2015 – 5 V 2068/14; zitiert von: BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

⁴⁸⁶ Vgl. z.B.: EU Science Hub – The European Commission's science and knowledge service: Cross-domain Conference on Benford's Law application "Benford's Law Conference, 10-12 July 2019 – Stresa, Italy", nachzulesen unter: <https://ec.europa.eu/jrc/en/event/workshop/benfords-law-conference>.

⁴⁸⁷ Vgl. frühe Beiträge zu dem Thema wie z.B.: Sosna, „Einsatz statistischer Methoden zur Risikoanalyse, Recherche und Lokalisierung von Steuerausfällen“, StBp 2000, 41 u. 68, oder Huber, „Weiterentwickelte und neue Methoden der Überprüfung, Verprobung und Schätzung“, StBp 2002, 199, 233, 258 u. 293.

⁴⁸⁸ BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 9 ff. u. 12 f.

„Wahrscheinlichkeitsrechnung (Stochastik)“ beurteilt, gibt der χ^2 -Test (Chi-Quadrat-Test)⁴⁸⁹ für die obere Abb. ca. 53 % und für die untere 0 % Übereinstimmung aus. Aufgrund der wesentlich größeren Datenmenge wird im zweiten Beispiel demnach eine erheblich bessere Mustertreue erwartet.

Aus der Gegenüberstellung der Beispiele wird deutlich, dass ein rein tabellarischer oder optischer Abgleich von beobachteten Mustern mit den jeweiligen Idealerwartungen zu falschen Schlussfolgerungen führen kann. Der mathematische Bereich der „**Wahrscheinlichkeitslehre (Stochastik)**“ bündelt die Theorie und Erfahrungen zum Streuverhalten von (zusammengehörigen) Daten und bietet komfortable Werkzeuge zur Übereinstimmungsprüfung (Wahrscheinlichkeitstests) an. Da Wahrscheinlichkeitstests über die Kombinatorik⁴⁹⁰ mittels Grundrechenarten theoretisch reproduziert werden können und die beste Beschreibung komplexen Streuverhaltens darstellen, sind sie weithin anerkannt: u.a. für die Absicherung von Studienergebnissen⁴⁹¹, i.R. der industriellen Qualitätssicherung⁴⁹², in der beweisfähigen Stichprobenprüfung⁴⁹³ und bei der automatischen Prozesssteuerung⁴⁹⁴. Diese Art der

⁴⁸⁹ Übliches stochastisches Mittel zur Übereinstimmungsprüfung zweier mehrklassiger Verteilungen mit unbekannter Verteilungscharakteristik ist der χ^2 -Test, was auch an der einheitlichen Zurverfügungstellung in Tabellenkalkulationsprogrammen ersichtlich ist. *Bachmann/Richter/Steinborn* fragen in StBp 2019, 38: „Führt ein anderer Anpassungstest in der Außenprüfung zu einer anderen Schlussfolgerung? – Chi-Quadrat- und Kolmogorov-Smirnov-Anpassungstest in der Summarischen Risikoprüfung“ und kommen dabei im entscheidenden Beispiel aus einer Veröffentlichung des Autors (*Wähnert*, „Die Ziffernanalyse als Bestandteil zeitgemäßer Prüfungsansätze“, StBp 2012, 241) zu einer abweichenden Übereinstimmung von 38,83 % gegenüber 0,02 %. Dabei wurde allerdings nicht beachtet, dass für den kumulierenden – und damit formbezogenen – Kolmogorov-Smirnov-Anpassungstest üblicherweise eine Klassensortierung vor der Auswertung empfohlen wird, weil sich ansonsten widersprüchliche Ergebnisse für vergleichbare Konstellationen ergeben (s. *mit ausführlichem Vergleich den Abschnitt „Chi-Quadrat-Test“ unter Pkt. 4.3.1.3.*). Zum selben Ergebnis kommt *Sell*, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 115, 141 u. 143: „Wie unüblich die Verwendung des KS-Tests auf diskrete Verteilungen ist, zeigt u.a. die Tatsache, dass eine entsprechende Erweiterung der ‚ks.test‘-Funktion in der Statistiksoftware R erst in 2011 [...] erfolgt ist und in der Mehrzahl anderer Software-Anwendungen gar nicht verfügbar ist [...] Der KS-Test ist [...] anfällig für die Sortierung der Klassen [...] Der KS-Test scheint nach diesen Einschränkungen kaum geeignet, um im Rahmen der Ziffernanalyse auf Gleichverteilung der Ziffernhäufigkeiten zu testen. Hierfür sollte der Chi-Quadrat-Anpassungstest angewendet werden.“

⁴⁹⁰ Die (mathematische) Kombinatorik beschäftigt sich u.a. mit den möglichen Ergebnisvariationen (Permutationen) in diskreten – im Gegensatz zu kontinuierlichen – Datenverteilungen wie bspw. mit den Wahrscheinlichkeitsverhältnissen beim Werfen einer Münze. Durch das Aufstellen eines Ergebnisbaums aus den denkbaren Resultaten (bei zwei Münzwürfen mit den Seiten „A“ und „B“: „AA“, „AB“, „BA“ und „BB“) können die Wahrscheinlichkeiten für Einzelresultate (z.B. „genau zweimal Seite ‚A‘ bei drei Münzwürfen“) oder Ergebnisbereiche (z.B. „mindestens zweimal Seite ‚A‘ bei drei Münzwürfen“) auf Basis der Wahrscheinlichkeit des einzelnen Vorgangs (für eine ideal geprägte Münze „50:50“) problemlos abgelesen und berechnet werden.

⁴⁹¹ Vgl. z.B.: *Bortz/Schuster*, „Statistik für Human- und Sozialwissenschaftler“, Springer-Verlag, 2010, 11.

⁴⁹² DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/de/norm/din-iso-2859-1/191697831>.

⁴⁹³ § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden“ sowie BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

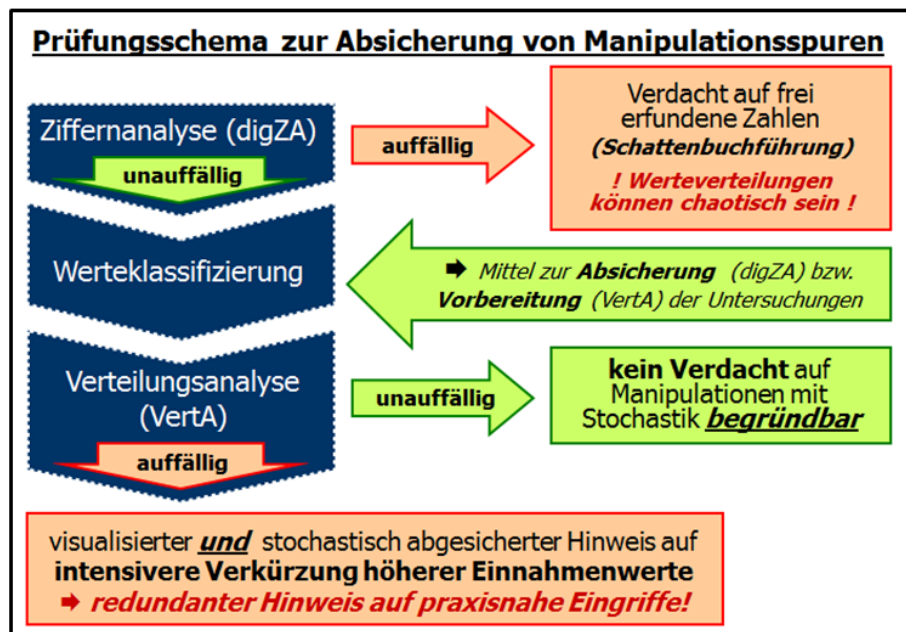
Übereinstimmungsprüfung von musterbasierten Häufigkeitsanalysen bedeutet somit die beste dem Menschen bekannte Form der Beurteilung, welche zudem die untersuchte Datenmenge im Ergebnis berücksichtigt⁴⁹⁵.

Sobald musterbasierte Häufigkeitsanalysen in der SRP über die Aufdeckung von Prüffeldern hinaus zur indiziellen Schlüssigkeitsbeurteilung eingesetzt werden, erfolgt grundsätzlich eine Übereinstimmungsbewertung mit Hilfe eines geeigneten Wahrscheinlichkeitstests.

Um i.R. der SRP die **Gefahr ungerechtfertigter Beanstandungen** auf Grundlage manipulationsorientierter Mustervergleiche wie Ziffernanalysen möglichst **gering zu halten**, beschränken sich die Anwendungen auf zwei praxistypische Manipulationsarten:

-  freies Erfinden von Steuerungsdaten, wodurch Ziffernvorlieben bzw. -abneigungen so deutlich entstanden sein können, dass normale Streutendenzen sie nicht mehr erklären können, sowie
-  Spitzeneingriffe als verstärkte Verkürzungen in (betriebsbezogen) hohen Tageskasseneinnahmen, was den manipulationspsychologischen Erfahrungen zu kriminellem Verhalten entspricht (s. [Pkt. 4.3.4.](#)).

Darüber hinaus leitet die SRP mit diesem „**Prüfungsschema zur Vermeidung von Fehldeutungen**“⁴⁹⁶ durch die manipulationsbezogenen Musteranalysen von Tageskasseneinnahmen, um einerseits vor Anwendungs- sowie In-



⁴⁹⁴ Vgl. z.B.: Richter, „Prinzipien der Künstlichen Intelligenz“. Teubner-Verlag, 1992.

⁴⁹⁵ Die regelmäßig angemerkte Skepsis bei manipulationsorientierten Mustervergleichen wie bspw. der Einerstellenprüfung auf die Gleichverteilung, ob genügend Daten für eine aussagekräftige Beurteilung vorlägen, ist insofern nur angebracht, solange *kein* Wahrscheinlichkeitstest zur Übereinstimmungsprüfung herangezogen wird. Für einzelne Wahrscheinlichkeitstests kann es wie beim χ^2 -Test (Chi-Quadrat-Test) Empfehlungen zu Mindestdatenmengen geben, die aber niedrig ausfallen und in der SRP beachtet werden (s. [Pkt. 4.3.1.3.](#)).

⁴⁹⁶ Das „Prüfungsschema zur Vermeidung von Fehldeutungen“ ist Bestandteil der SRP-Vorlage Stochastik und ergänzt deren Kurzanleitung (Tabellenblatt „Kurzanleitung 2“).

terpretationsfehlern zu bewahren und andererseits betriebliche Faktoren als Ursachen für aufgefallene Musterstörungen soweit möglich abzuklären; diese Sorgfalt wird von der wissenschaftlichen Literatur bemerkt⁴⁹⁷.

Dazu beginnt die Untersuchung von Tageskasseneinnahmen auf Manipulationsspuren in der SRP grundsätzlich mit den Zifferanalysen „Einerstelle auf die Gleichverteilung“ und „Zweitziffer auf die Benford-Verteilung“. Hiermit sollen mögliche Ziffernvorlieben und -abneigungen erkannt bzw. weitgehend ausgeschlossen werden. Sofern Ziffernauffälligkeiten entdeckt worden sind, werden denkbare betriebliche Ursachen (auch) mit Hilfe der speziellen Funktionen der SRP-Vorlage Stochastik hinterfragt, indem bspw. bei bestimmten Zweitzi fernhäufungen ein Dezimalsprung durch eine besondere Schichtung zügig und einfach als Ursache abgeklärt werden kann (s. *das entsprechende Beispiel unter [Pkt. 4.3.2.3.3](#)*). Bevor die (kompletten) Zahlen der Tageskasseneinnahmen fortfahrend auf stärkere Spitzeneingriffe (in die betriebsbezogen hohen Einnahmenwerte) untersucht werden, überprüft ein Vorbereitungsschritt, ob sie einer einzelnen Verteilung folgen. Hierzu stellt die SRP-Vorlage Stochastik ebenfalls zielführende Funktionen zur Verfügung, um bei fehlender Einheitlichkeit einen betrieblichen Grund in Nachfragebesonderheiten der Saison, der Wochentage etc. schnell aufzufinden (s. [Pkt. 4.3.3.3](#)). Durch den (weitgehenden) Ausschluss von frei erfundenen oder inhomogenen⁴⁹⁸ Daten kann eine seriöse Prüfung auf eine – kriminalpsychologisch typische – verstärkte Einnahmenverkürzung der (betriebsbezogen) hohen Tageskasseneinnahmen vorgenommen werden (s. [Pkt. 4.3.4](#)).

Wegen der Vorprogrammierung der SRP-Vorlage Stochastik und des begleitenden Prüfungsschemas können Fehler aufgrund von Missverständnissen zu manipulationsorientierten Musteranalysen, wie sie auch in der Literatur verbreitet sind⁴⁹⁹, verhindert werden. Im Ge-

⁴⁹⁷ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 139: „An die Beobachtungen aus Chi-Quadrat-Anpassungstest und Binomialtest schließen sich in der Praxis der Betriebsprüfung weitere Prüfungshandlungen an [...] Eine Übersicht der dafür durchgeführten Maßnahmen in der praktischen Anwendung der Betriebsprüfung bietet das Handbuch zur Summarischen Risikoprüfung“.

⁴⁹⁸ Wenn Tageskasseneinnahmen z.B. in der Saison oder an Wochenendtagen ein extrem abweichendes Niveau aufweisen, ergeben sich in der grafischen Häufigkeitsauswertung mehrere Häufungsspitzen, sodass eine uneinheitliche (inhomogene) Verteilung entsteht.

⁴⁹⁹ Vgl. z.B.: Bisle, „Chi-Quadrat-Test‘ und Zeitreihenvergleich: Keine Schätzung bei ordnungsgemäßer Buchführung“, PStR 1/2012, 15: „Der ‚Chi-Quadrat-Test‘ analysiert die Verteilung bestimmter Ziffern und basiert auf der Erkenntnis, dass jeder Mensch Lieblingsziffern hat und bei frei erfundenen Zahlen (z.B. nachgetragenen Tageseinnahmen oder -ausgaben) häufiger verwendet als andere.“ Tatsächlich ist der χ^2 -Test (Chi-Quadrat-Test) ein allgemeiner Wahrscheinlichkeitstest, der mehrklassige Verteilungen unter Berücksichtigung der Erkenntnisse der Wahrscheinlichkeitslehre auf Übereinstimmung prüft. Eine (zwingende) Verbindung zur Gleichverteilung von endständigen Ziffern in Wirtschaftsdaten besteht nicht. Vielmehr vermischt die Aussage die zwei Schritte „Erwartung“ und „Übereinstimmungsprüfung“ i.R. der musterbasierten Analyse und sorgt somit für Verständnisprobleme und Fehleinschätzungen; so auch: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtspre-

gensatz zur allgemeinen Akzeptanz von Musteranalysen bei der Aufdeckung von Manipulationen und von Wahrscheinlichkeitstests bei der Übereinstimmungsprüfung [s. die Abschnitte „musterbasierte Prüfungsansätze zur Aufdeckung von Manipulationen“ und „Wahrscheinlichkeitslehre (Stochastik)“ zu diesem Pkt. mit entsprechenden Verweisen] ist die derzeitige Zurückhaltung in der Finanzrechtsprechung⁵⁰⁰ offenbar darauf zurückzuführen, dass Fehler in der Anwendung befürchtet werden⁵⁰¹. In der Literatur wird hingegen zu Recht auf den qualifizierten wissenschaftlichen Hintergrund der Verfahren hingewiesen: „Erwähnenswert ist in diesem Zusammenhang eine Untersuchung der Universität Münster (Institut für Wirtschaftsinformatik) über die Aussagekraft des Chi-Quadrat-Tests. [...] Diese Untersuchung dokumentiert deutlich, dass diese statistische-mathematische Prüfungsmethode auf wissenschaftlich gesicherten Grundlagen beruht.“⁵⁰² Demgegenüber stellen zurückhaltende Stimmen von wissenschaftlicher Seite⁵⁰³ nicht auf das gesetzlich reduzierte Beweismaß für Verprobungen gem. § 158 Abs. 2 Nr. 1 AO (s. [Pkt. 4.6.](#)) sondern auf ein überzogenes ab, welches Verprobungen grundsätzlich nicht erreichen können.⁵⁰⁴

chung“, Diss., Verlag Dr. Kovač, 2021, 194: „[...] irritiert, dass die Verprobungsmethode ‚Ziffernanalyse‘ regelmäßig als ‚Chi-Quadrat-Test‘ geführt wird.“

⁵⁰⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit (s. FG Münster v. 14.8.2003 – 8 V 2651/03 E, U, EFG 2004, 9 m. Anm. Trossen; FG Münster v. 10.11.2003 – 6 V 4562/03 E, U, EFG 2004, 236 [237]; FG Münster v. 7.12.2005 – 1 K 6384/03 E, EFG 2006, 652 [653] m. Anm. Zimmermann [Fahrtenbuch]; FG Rh.-Pf. v. 24.8.2011 – 2 K 1277/10, EFG 2012, 10 [11] m. Anm. Kühnen; Sosna, StBp. 2000, 41 [Teil I], 68 [Teil II]; Meyer, DStR 2005, 2114; Watrin/Stuffert, DB 2006, 1748; Huber, StBp. 2009, 121 [124 ff.]“.

⁵⁰¹ Kohlmann, Steuerstrafrecht, 72. Lieferung 9.2021, § 397 AO, Rz. 12.3: „Die praktische Verwertbarkeit der statistischen Verfahren ist durchaus umstritten, und zwar bereits als Grundlage für steuerliche Schätzungen, aber erst recht für strafrechtliche Verurteilungen. Die FG lehnen bislang den Chi-Quadrat-Test als eine sichere Methode, die Ordnungsgemäßheit der Buchführung zu verwerfen, ab. Allerdings wird weniger die grundsätzliche Eignung der Analysen angezweifelt als vielmehr hinterfragt, unter welchen Voraussetzungen und mit welcher Anzahl von Ausgangswerten die Methode aussagekräftige Ergebnisse liefert.“

⁵⁰² Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstrafverfahren“, NZWiSt 1/2014, 24, 25.

⁵⁰³ Zuletzt z.B.: Pitz/Klapdor/Gardian/Haselau, „Das Benfordsche Gesetz als Werkzeug zur Aufdeckung von Umsatzmanipulationen in der Gastronomie?“, StuW 3/2021, 232, 245: „Eine der zentralen Fragestellungen, ob Testverfahren einen nicht manipulierten Umsatz auch wirklich als nicht manipuliert identifizieren, kann hier keinesfalls bejaht werden. [...] Auch wenn umfangreichere Untersuchungen noch ausstehen, so zeigen bei der Ziffernanalyse nach Benford bei dieser Fallstudie Betriebe deutliche Auffälligkeiten, bei denen eine Manipulation (eher) vermutet wird. Auch wenn eine Tendenz zu erkennen ist, ist eine statistisch signifikante Korrelation zwischen ‚Auffälligkeit‘ und ‚Zuverlässigkeit‘ der Datensätze (noch) nicht zwingend nachzuweisen.“

⁵⁰⁴ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324: „Die herkömmlichen [Verprobungs]methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...]. Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten in den wiederhergestellten Daten. [...] Schlussfolgernd stellt sich die Frage, ob das geforderte Sicherheitsmaß [mit an Sicherheit grenzender Wahrscheinlichkeit] bei den herkömmlichen Methoden überhaupt erreicht werden kann.“

4.3.1. Ablauf von Häufigkeitsanalysen

Musterbasierte Häufigkeitsanalysen folgen in ihrem Ablauf allgemein diesem **Schema**:



4.3.1.1. Beobachtungen für Häufigkeitsanalysen

Für Beobachtungen der Verteilungseigenschaften von (Besteuerungs-)Daten kommen alle **skalierbaren Informationen** in Betracht; damit sind Datenmerkmale⁵⁰⁵ gemeint, welche bereits wie der USt-Satz, die Zahlungsweise, die Kostenart etc. diskret⁵⁰⁶ differenzierbar sind oder als stetige⁵⁰⁷ Wertinhalte sinnvoll in Klassen unterteilt werden können. Nach dieser Eigenschaft werden die musterbasiert zu untersuchenden Daten ausgewertet. Das Ergebnis kann tabellarisch oder in einem Diagramm dargestellt werden: die unterschiedlichen Merkmalsausprägungen⁵⁰⁸ auf der X-Achse und die jeweiligen Häufigkeiten auf der Y-Achse (zur Überlegenheit von verbildlichten Analyseergebnissen s. den Abschnitt „Visualisierung von Besteuerungsdaten“ unter [Pkt. 2.](#)).

Bezogen z.B. auf Tageskasseneinnahmen können als *eine* Eigenschaft die **Ziffern** nach ihren Häufigkeiten ausgezählt werden. Hierbei ist die Betrachtungsrichtung wie in der



Abb. zu unterscheiden: **von vorne mit den Zahlenpositionen** „Erstziffer“, „Zweitziffer“, „Drittziffer“ usw. oder **mit vergleichbarer Größenordnung als „Einerstelle“, „Zehnerstelle“, „Hunderterstelle“** etc. Die erstgenannte Zahlenbeschreibung, auf der die Verteilungserfahrungen des Benford-Gesetzes (s. [Pkt. 4.3.2.3.1.](#)) basieren, kann dazu führen, dass Zahlenpositionen mit abweichender Wertigkeit gemeinsam ausgewertet werden: Die Zweitziffer der obigen Tageskasseneinnahme „€ 1.276,20“ meint die *Hunderterstelle*, während an einem absatzschwächeren Tag mit „€ 699,30“ die *Zehnerposition* die zweite Stelle von vorne bildet. Analysebesonderheiten, die sich aus der Überschneidung beider Beschrei-

⁵⁰⁵ Die Datenmerkmale befinden sich in einer Datenbank zu jedem Datensatz vereinheitlicht auf die Spalten verteilt.

⁵⁰⁶ Unter diskreten Verteilungen werden in der Mathematik solche verstanden, die keine Zwischenwerte einnehmen können. So kann es bei Münzwürfen aus Sicht einer bestimmten Münzseite nur zu ganzzahligen Erfolgs- bzw. Misserfolgshäufigkeiten kommen, nicht aber bspw. zu 3,437 Treffern. Der gegenteilige Verteilungstyp mit kontinuierlichen Ergebnismöglichkeiten wird „stetig“ genannt.

⁵⁰⁷ Vgl. die Erklärung zu „diskret“ unter der vorherigen Fußnote.

⁵⁰⁸ Zu jedem differenzierbaren Datenmerkmal – in der Datenanalytik auch Dimension genannt – gibt es die originären oder festgelegten (in Klassen unterteilten) Ausprägungen wie bspw. zum USt-Satz „0 %“, „7 %“, „19 %“ etc. bzw. bei Klassen „liegt im Wertebereich von ... bis ...“.

bungsperspektiven ergeben, werden von der SRP-Vorlage Stochastik automatisch beachtet (s. [Pkt. 4.3.2.2.1.](#)).

Anstelle der punktuellen Ziffernbetrachtung können dieselben Datensätze auch in **Größenklassen** eingeteilt werden. Um gleichmäßige Muster erkennen zu können, werden dabei i.d.R. konstante Klassenabstände wie mit dem € 100-Intervall („> 1.200 – 1.300“) im Beispiel gewählt. Die SRP-Vorlage Stochastik führt derartige Datenklassifizierungen (Schichtungen) automatisch auf Basis der praxiserprobten Wurzelmethode⁵⁰⁹ durch, bietet den Anwender*innen auf den Tabellenblättern „Schichtung“, „Schichtungsvertiefung“ und „Verteilungsanalyse“ jedoch diverse komfortable Optionen zur freien Anpassung der Schichtobergrenzen – und damit der Intervalle – sowie zur Umstellung von linear auf logarithmisch⁵¹⁰ gleichmäßige Klassen an.

Darüber hinaus liefert eine vorhandene Chronologie in zu untersuchenden Daten mit dem Wiederholungsverhalten (**Iteration**⁵¹¹) weitere Beobachtungsoptionen: Ausgewertet wird, wie viele Datensätze – also bei Tageskasseneinnahmen Tage – es dauert, bis bspw. die beobachtete Einerstelle „6“ des Beispiels erneut vorkommt. Dieser Ansatz ist insbesondere für moderne (digitale) Einnahmenverkürzungen [s. den Abschnitt „*raffinierte (digitale) Manipulationswerkzeuge*“ unter [Pkt. 2.1.](#)] von Bedeutung, weil die zeitliche Ziffernabfolge, welche von Zufallsgeneratoren erzeugt wird, nicht derjenigen entspricht, die natürlich entsteht.

4.3.1.2. (Punkt-)Erwartungen für Häufigkeitsanalysen

Jede (seriöse) Prüfungsschlussfolgerung zur Beobachtung von Datenmerkmalen braucht ein Vergleichsmaß. Diese Erwartung stellt als **Punktideal**⁵¹² die Grundlage für eine Beurteilung des beobachteten Häufigkeitsmusters im betriebswirtschaftlichen, rechtlichen oder in einem anderen Kontext dar.

⁵⁰⁹ Die Wurzelmethode berechnet die Klassenanzahl zur Datenauszahlung (Schichtung) nach folgender Formel: $\sqrt{(Anzahl\ Datensätze)}$. Das Ergebnis wird in der SRP-Vorlage Stochastik zur besseren Übersichtlichkeit auf max. 30 Intervalle begrenzt.

⁵¹⁰ Sofern ein asymmetrisch linkssteil-rechtsschiefes Häufigkeitsmuster wie die Logarithmische Normalverteilung (s. [Pkt. 4.3.4.1.](#)) erwartet wird, ist die Wahl von Klassengrenzen mit logarithmisch gleichen Abständen (z.B. „10²“ – „10³“ – „10⁴“ – „10⁵“ etc. entsprechend „100“ – „1.000“ – „10.000“ – „100.000“ usw.) oft zielführender, um die Gleichmäßigkeit der beobachteten Verteilung visuell zu beurteilen.

⁵¹¹ S. hierzu den Vortrag „Analyse des Iterationsverhaltens von Ziffern in Risikodaten“ von Rickert auf dem 15. Deggendorfer Forum zur digitalen Datenanalyse 2019, Kurzbeschreibung nachzulesen unter: <https://www.dfdda.de/index.php/de/veranstaltungen/forum/15-forum-2019/vortraege>, ausführlicher im Tagungsband „Künstliche Intelligenz – Chancen und Risiken für Wirtschaftsprüfung und Finanzwirtschaft“, IDW-Verlag, 2019, 132 ff.

⁵¹² Die ideale Punkterwartung kennt zunächst keine Toleranz wie z.B. ein exaktes Rechenergebnis aus einer Nachkalkulation. Erst i.R. der Übereinstimmungsprüfung werden Unsicherheiten der Berechnung oder aus unvollständigen Informationen als tolerierbarer Ergebnisrahmen (qualifiziert) einbezogen.

Erwartungsmuster können aus Erfahrungen mit dem geprüften Unternehmen (**interner Strukturvergleich**) wie z.B. die Häufigkeitsverteilung von Kostenpositionen aus vorherigen Prüfungszeiträumen oder bei der Gegenüberstellung mehrerer vergleichbarer Betriebe eines Konzerns stammen. Im Fall deutlich erkennbarer Unterschiede der Verteilungsmuster werden dadurch wesentlich veränderte Verhältnisse angezeigt, die auf umgestellte Handlungsabläufe, Umstrukturierungen oder menschliche Eingriffe hinweisen und somit einzelfallspezifische Prüffelder offenlegen.

Darüber hinaus sind Verteilungsmuster, die in (Wirtschafts-)Daten regelmäßig vorgefunden wurden, bekannt und können für Vergleiche herangezogen werden (**externer Strukturvergleich**). Das bekannteste Häufigkeitsmuster, welches wegen seiner weiten Verbreitung auch angewendet wird, wenn kein anderes bekannt ist, ist die Normalverteilung, wie sie z.B. die Verhältnisse zum Intelligenzquotienten in (Teilen) der Bevölkerung beschreibt (s. *die Abb. zum Abschnitt „gleichmäßige Muster der Häufigkeitsverteilungen“ unter [Pkt. 4.3.](#)*). Sie dient als Grundlagendverteilung für diverse Wahrscheinlichkeitstests (s.a. [Pkt. 4.3.1.3.](#)), weil sie als typisches natürliches Streuverhalten gilt. Mit ihrer ebenfalls symmetrischen Glockenform ist die Binomialverteilung⁵¹³ ein der Normalverteilung ähnliches Muster und stellt sich bei (dichotomen) ja-nein-Sachverhalten wie bspw. dem Verhältnis „Verzehr an Ort und Stelle versus Außerhausverkauf“ um ein Standardverhältnis⁵¹⁴ herum ein. Daneben beschreibt die Gleichverteilung mit identischen Anteilen für alle möglichen Merkmalsausprägungen das Abbild des absoluten, gleichgewichteten Zufalls und kommt als weitere allgemeine Erwartung vor (s. *die beiden Ziffernanalyse-Grafiken im Abschnitt „Güte der Übereinstimmung mit dem Idealmuster von der Datenmenge abhängig“ unter [Pkt. 4.3.](#)*). Nach empirischen Erkenntnissen weisen die führenden Zahlenpositionen in Wirtschaftsdaten besondere Häufigkeitsmuster der (möglichen) Ziffern auf (s. [Pkt. 4.3.2.3.1.](#))⁵¹⁵. In vielen Wissenschaftsbereichen wird zudem ein homogenes rechtsschiefes Muster für die nach Größenklassen gezählten Daten wiedergefunden, wenn diese in ihrer Entfaltung linksseitig begrenzt⁵¹⁶ sind: die

⁵¹³ Die Binomialverteilung kennt als diskrete Form im Gegensatz zur Normalverteilung keine Zwischenwerte und wird somit üblicherweise als Säulendiagramm wiedergegeben.

⁵¹⁴ Im Vergleich der Erlöstrennung „Verzehr an Ort und Stelle versus Außerhausverkauf“ mehrerer Filialen hat jede ein typisches Nachfrageverhältnis: Werden die einzelnen Tage nach den jeweiligen Anteilen ausgezählt, ergibt sich pro Filiale eine Binomialverteilung um den typischen Verhältniswert herum, sofern keine wesentlichen strukturellen Veränderungen eingetreten sind.

⁵¹⁵ Benford, „The Law of Anomalous Numbers“, Proceedings of the American Philosophical Society, 1938, 551.

⁵¹⁶ Als praxisnahes Beispiel seien hier die durchschnittlichen Kraftstoffverbräuche von Pkw-Typen genannt. Unterhalb eines minimalen Verbrauchs sind technisch keine Autos verfügbar, was die linksseitige Begrenzung bedeutet. Mit auf der X-Achse ansteigendem Kraftstoffverbrauch nehmen die Pkw-Typen bis zu einem (für den Beobachtungszeitpunkt) üblichen Normverbrauch zu. In den darüber hinausgehenden Verbrauchsklassen lassen die Häufigkeiten langsamer (als linksseitig des Normverbrauchs) nach und reichen erheblich weiter nach rechts bis zum Verbrauchsmaximum.

Logarithmische Normalverteilung⁵¹⁷; unter logarithmischer⁵¹⁸ Einteilung der X-Achse wird aus jeder Logarithmischen Normalverteilung eine Normalverteilung (s. [Pkt. 4.3.4.1.](#)).

Den aufgezählten verbreiteten Verteilungsmustern ist gemein, dass sie mit Hilfe weniger Informationen⁵¹⁹ vollständig beschrieben und deshalb problemlos auf die speziellen Verhältnisse untersuchter (Besteuerungs-)Daten angepasst werden können. In der SRP-Vorlage Stochastik werden die erwarteten Häufigkeitsmuster so automatisch und nachvollziehbar aus den eingepflegten Daten gebildet. Speziell für die manipulationsorientierte Analyse der Wertzusammenhänge in Tageskasseneinnahmen auf den Tabellenblättern „Verteilungsanalyse“ und „Verteilungsdiagramm“ – im Anschluss an die Vorbereitungsmaßnahmen gem. SRP-Prüfungsschema (s. *den Abschnitt „Prüfungsschema zur Vermeidung von Fehldeutungen“ unter [Pkt. 4.3.](#)*) – ist zudem eine gezielte Anpassung der Punkterwartung „Logarithmische Normalverteilung“ möglich. Auf diese Weise kann der Verdacht eines – manipulationspsychologisch typischen – stärkeren Spitzeneingriffs einzelfallbezogen genauer untersucht werden (s. [Pkt. 4.3.4.3.](#)).

4.3.1.3. Übereinstimmungsprüfung (Chi-Quadrat-Test)

Wie bei jedem prüfenden Soll-Ist-Vergleich schließt sich an eine musterbasierte Häufigkeitsbeobachtung zu (Besteuerungs-)Daten mit Zuordnung einer an den Einzelfall angepassten Punkterwartung eine Übereinstimmungsprüfung an.

Da eine exakte Übereinstimmung mit der Punkterwartung bei komplexen Gesamtbildern nur in sehr seltenen Fällen eintreten wird, umfasst eine Übereinstimmungsprüfung auch die **Festlegung eines Toleranzbereichs bis zur Abweisungsschwelle**. Vergleichbar wird bspw. beim externen Richtsatzvergleich der beobachtete Aufschlagsatz nicht am „Mittelsatz“⁵²⁰ sondern an der Richtsatzspanne gemessen und bei einer Nachkalkulation ein Ergebnisrahmen durch Hinzu- und Abrechnung eines bestimmten Betrags bzw. Prozentanteils definiert. Dieser übliche Unbedenklichkeitsbereich soll Informationslücken und Berechnungsunsicherheiten abdecken und dadurch falsche Beanstandungen vermeiden.

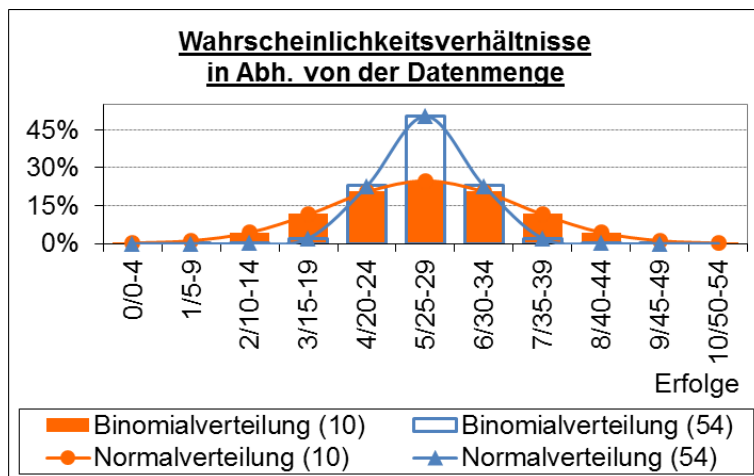
⁵¹⁷ Vgl. z.B.: *Limpert/Stahel/Abbt*, „Lognormal distributions across the sciences: keys and clues“, BioScience Band 51, Nr. 5, 2001, 341, nachzulesen unter: https://de.wikipedia.org/wiki/Logarithmische_Normalverteilung.

⁵¹⁸ Eine logarithmische Klasseneinteilung belegt die Intervallobergrenzen mit linear gleichmäßig ansteigenden Exponenten wie bspw. mit „ $10^0 - 10^1 - 10^2 - 10^3$ “ (entsprechend „1 – 10 – 100 – 1.000“).

⁵¹⁹ Die Berechnung der Häufigkeitsanteile für die Normal- und die Logarithmische Normalverteilung erfordert den Mittelwert und die Standardabweichung (vorzeichenneutrale typische Abweichung jedes einzelnen Datensatzes vom Mittelwert). Bei der Binomialverteilung lassen sich die Verhältnisse aus der Trefferanzahl, Trefferwahrscheinlichkeit und Datenanzahl errechnen. Eine Übertragung der Gleichverteilung auf die Daten eines Einzelfalls gelingt per Division durch die Klassenanzahl. Für die anteiligen Häufigkeiten nach den Benford-Verteilungen wird lediglich die betreffende Ziffer(nkombination) benötigt.

⁵²⁰ Vgl. z.B. die „Richtsatzsammlung für das Kalenderjahr 2022“, Rz. 6 der Vorbemerkungen, nachzulesen unter: https://www.bundesfinanzministerium.de/Content/DE/Downloads/BMF_Schreiben/Weitere_Steuerthemen/Betriebspruefung/Richtsatzsammlung/2023-08-10-download-richtsatzsammlung-2022.pdf?__blob=publicationFile&v=3.

Bei musterbasierten Häufigkeitsanalysen, die für indizielle Aussagen gem. § 158 Abs. 2 Nr. 1 AO gedacht sind, werden hierfür i.d.R. die Mittel des mathematischen Spezialgebiets für „Aussagen unter Unsicherheit“ verwendet: die **Wahrscheinlichkeitstests**. Im Gegensatz zu anderen Soll-Ist-Vergleichen der Betriebsprüfung ist dieses Vorgehen schon deshalb zu empfehlen, weil diese die Abhängigkeit der Mustertreue von der Anzahl untersuchter Daten (s. das Beispiel im Abschnitt „Güte der Übereinstimmung mit dem Idealmuster



von der Datenmenge abhängig“ unter [Pkt. 4.3.](#)) einbeziehen: So lässt sich z.B. an den Säulen der Binomialverteilungen in der Grafik problemlos ablesen, dass bei **zehn Würfeln mit einer ideal geprägten⁵²¹ Münze** ein sehr viel breiteres Ergebnisspektrum um die Punkterwartung „5-mal jede Seite“ herum als bei **erheblich**

mehr Versuchen (54) resultiert. Zur Übertragung lässt sich auch stetigen⁵²² Normalverteilungen, die als Grundlage diverser Wahrscheinlichkeitstests dienen, wie in der Abb. die Eigenschaft vermitteln, dass die Streuungsintensität von der Datenanzahl abhängt. In der Konsequenz berücksichtigt eine musterbasierte Häufigkeitsanalyse, bei welcher die Übereinstimmungsprüfung mit einem entsprechenden Wahrscheinlichkeitstest vorgenommen wird, die untersuchte Datenmenge: Werden z.B. wenige Tageskasseneinnahmen aus einer abgekürzten Außenprüfung auf die Übereinstimmung ihrer Einerstelle mit der Gleichverteilung untersucht, erwartet der χ^2 -Test (Chi-Quadrat-Test) eine wesentlich schwächere Deckungsgleichheit als bei den umfangreicheren Daten einer normalen Betriebsprüfung. Insofern sind die mitunter angeführten Forderungen nach hohen **Mindestdatenmengen** wie z.B. „ ≥ 500 “ unberechtigt. Vielmehr gibt es bspw. zum χ^2 -Test (Chi-Quadrat-Test) eine vorherrschende Empfehlung, wie viele Daten für eine seriöse Auswertung mindestens vorliegen sollten⁵²³, und das sind bei typischen Anwendungen deutlich weniger: Danach bedeutet ein Mindestdatenumfang von „ ≥ 5 erwartete Elemente pro Klasse“ für die Gleichverteilung zehn möglicher

⁵²¹ Die Ergebniswahrscheinlichkeit für die Seiten „A“ und „B“ ergibt sich dann bei jedem Wurf mit „50:50“.

⁵²² Verteilungen mit kontinuierlichen Ergebnismöglichkeiten werden „stetig“ genannt. Demgegenüber werden unter diskreten Verteilungen in der Mathematik solche verstanden, die keine Zwischenwerte einnehmen können. So kann es bei Münzwürfen aus Sicht einer bestimmten Münzseite nur zu ganzzahligen Erfolgs- bzw. Misserfolgshäufigkeiten kommen, nicht aber bspw. zu 3,437 Treffern.

⁵²³ Vgl. z.B.: Diller/Schmid/Späth/Kühne, „Zifferntests in der Betriebsprüfung – Chancen und Risiken“, DStR 6/2015, 311, 314: „Als Mindestumfang [für den χ^2 -Test] werden in der Literatur unterschiedliche Werte angegeben, wobei die am häufigsten vorgeschlagene Regel besagt, dass jede Klasse im Erwartungswert mit mindestens 5 Elementen besetzt sein sollte“ mit Verweis auf Rüger, „Test- und Schätztheorie“, 2002, 204.

Ziffern, dass mindestens 50 Datensätze zur Untersuchung vorliegen sollten. Bei Mustervergleichen mit den Benford-Verteilungen richtet sich die Mindestdatenmenge nach der Ziffer(nkombination) mit dem kleinsten Erwartungsanteil, was bei der Zweitzypher mit ca. 8,5 % für die „9“ eine Empfehlung von minimal 59 Datensätzen ergibt.

Die Ergebnisse (stochastischer) Übereinstimmungsprüfungen werden üblicherweise als **prozentuale Übereinstimmungs- bzw. Nichtübereinstimmungswahrscheinlichkeit** ausgegeben, wobei die Summe aus beiden grundsätzlich 100 % ist. Im Gegensatz zu bspw. „Kalkulationen“ oder „Zeitreihenanalysen“ resultiert aus der stochastischen Übereinstimmungsbeurteilung (eines beobachteten Häufigkeitsmusters mit der auf den Einzelfall angepassten Punkterwartung) ein konkretes Ergebnis, welches transparent und jederzeit reproduzierbar ist und sich somit – gerade für fein differenzierende Schlussfolgerungen wie im deutschen Steuerrecht – besser eignet als gefühlte Aussageklarheiten⁵²⁴; die dennoch verbreitet bevorzugte gefühlte Wahrscheinlichkeitsbewertung geht auf die sog. „Verfügbarkeitsverzerrung“ zurück⁵²⁵. Die stochastische Beurteilung unter Unsicherheit wird grundsätzlich anerkannt: z.B. bei der Absicherung von Studienergebnissen⁵²⁶, der industriellen Qualitätssicherung⁵²⁷, der beweisfähigen Stichprobenprüfung⁵²⁸, der automatischen Prozesssteuerung⁵²⁹ usw.

In der Praxis haben sich einzelne **Standard-Wahrscheinlichkeitstests** für bestimmte Fragestellungen durchgesetzt, bspw. indem ausschließlich diese in Kalkulationsprogrammen wie MS Excel®, OpenOffice® etc. enthalten sind: Sollen zwei Verteilungen mit mindestens drei Klassen wie bei der Ziffernanalyse zu neun bzw. zehn möglichen Ziffern auf Übereinstimmung geprüft werden, ist der Standardtest der χ^2 -Test (**Chi-Quadrat-Test**).

⁵²⁴ Tlw. werden die unvermeidbaren Unsicherheiten von Prüfungs- und Verprobungshandlungen in der Literatur offenbar grundsätzlich nicht akzeptiert (vgl. z.B.: Prinz, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4, M5: „Wahrscheinlichkeitstheoretisch geformte Algorithmen können Steueransprüche nicht begründen“), wodurch eine objektive Bewertung der Qualitätsunterschiede unterschiedlicher Prüfungsmethoden unmöglich wird.

⁵²⁵ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Nomos, 2018, 56-57 u. 65: „Bei der Verfügbarkeitsheuristik schätzen Individuen die [...] Wahrscheinlichkeit für das Eintreten eines Ereignisses unter Unsicherheit danach ein, wie leicht bzw. schwer es ihnen fällt, damit verbundene Erscheinungen aus ihrem Gedächtnis abzurufen. Dabei werden Ereignisse für wahrscheinlicher erachtet, wenn sie sich leichter vorgestellt werden können und sich zu ihnen eine direkte Verfügbarkeit herstellen lässt. [...] Im Rahmen juristischer Entscheidungen scheint es unerlässlich, von diesen Entscheidungsverzerrungen Kenntnis zu haben, da eine verzerrte Entscheidung den Denkgesetzen der Logik widerspricht [...]“.

⁵²⁶ Vgl. z.B.: Bortz/Schuster, „Statistik für Human- und Sozialwissenschaftler“, Springer-Verlag, 2010, 11.

⁵²⁷ DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/de/norm/din-iso-2859-1/191697831>.

⁵²⁸ § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden“ sowie BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

⁵²⁹ Vgl. z.B.: Richter, „Prinzipien der Künstlichen Intelligenz“. Teubner-Verlag, 1992.

Aus der Literatur wurde dazu die Frage „Führt ein anderer Anpassungstest in der Außenprüfung zu einer anderen Schlussfolgerung? – Chi-Quadrat- und Kolmogorov-Smirnov-Anpassungstest in der Summarischen Risikoprüfung“⁵³⁰ aufgeworfen. I.R. der vorgenommenen Gegenüberstellung zu einem veröffentlichten Beispiel des Autors⁵³¹ wird auf eine Abweichung der Übereinstimmung von 38,83 % gegenüber 0,02 % hingewiesen. Bei der Berechnung mit dem Kolmogorov-Smirnov-Test wurde jedoch die aufsteigende Sortierung der Daten bzw. Klassen unterlassen, welche üblicherweise empfohlen wird⁵³² und hochgradig entscheidungsrelevant ist. Da diese Sortierempfehlung in der Literatur nicht einheitlich zu finden ist, muss zur weiteren Klärung die unterschiedliche Arbeitsweise beider Wahrscheinlichkeitstests verdeutlicht werden:

Ziffer	Beobachtung		Erwartung		χ^2 -Test	K-S-Test $ F_0 - F_x $
	absolut	anteilig	absolut	anteilig		
2	11	0,05	20,6	0,10	4,4738	0,0466
8	15	0,13	20,6	0,20	1,5223	0,0738
0	18	0,21	20,6	0,30	0,3282	0,0864
6	18	0,30	20,6	0,40	0,3282	0,0990
7	18	0,39	20,6	0,50	0,3282	0,1117
1	20	0,49	20,6	0,60	0,0175	0,1146
5	20	0,58	20,6	0,70	0,0175	0,1175
3	21	0,68	20,6	0,80	0,0078	0,1155
9	22	0,79	20,6	0,90	0,0951	0,1087
4	43	1,00	20,6	1,00	24,3573	0,0000
	206		206		31,4757	0,1175
Übereinstimmungswahrscheinlichkeit:					0,02%	0,68%

Die fortlaufend aufsummierende Auswertung des Kolmogorov-Smirnov-Tests (Spalten „anteilig“ als Teile von „1“ in der Abb.) mit der anschließenden Bewertung (Umrechnung) der **größten Einzelabweichung** „ $|F_0 - F_x|$ “ führt zwangsläufig zur Egalisierung von Positiv- und Negativdifferenzen

der bis dahin gemeinsam ausgewerteten Ziffern, wodurch unter der Ziffernreihenfolge „0 bis 9“ – also ohne Häufigkeitssortierung – im abgebildeten Fall die Häufung der „4“ durch die zu seltenen Ziffern „0“ und „2“ (bei erwartungsgemäßer „1“ und „3“) tlw. ausgeglichen würde; das würde zu dem unsinnigen Ergebnis führen, dass alleine das Vertauschen der (offenbaren) Ziffernvorliebe „4“ gegen die „0“ (ohne möglichen Ausgleich vorheriger Ziffern) zu einer völlig anderen Übereinstimmung (ca. 1,5 % versus 38,83 %) käme, obwohl die beiden Konstellationen aus Sicht der Testfrage „signifikante Ziffernvorlieben oder -abneigungen?“ identisch sind. Sobald die Ziffern wie in der Abb. nach beobachteten Häufigkeiten sortiert werden, reduziert sich die Abweichung der beiden Wahrscheinlichkeitstests auf 0,68 % versus 0,02 % Übereinstimmung. Die verbleibende Differenz lässt sich für die vorliegende Anwendung mit Hilfe der punktuellen Betrachtung „Ziffer ,4‘ als Vorliebe?“ hinterfragen: Aus dieser Sicht liegt eine Zufallsverteilung mit nur zwei möglichen Ergebnissen „4“ und „≠ 4“ vor, für welche die exaktere Binomialverteilung die Wahrscheinlichkeit, dass eine Ziffer mindestens

⁵³⁰ Bachmann/Richter/Steinborn, „Führt ein anderer Anpassungstest in der Außenprüfung zu einer anderen Schlussfolgerung? – Chi-Quadrat- und Kolmogorov-Smirnov-Anpassungstest in der Summarischen Risikoprüfung“, StBp 2019, 38.

⁵³¹ Wähnert, „Die Ziffernanalyse als Bestandteil zeitgemäßer Prüfungsansätze“, StBp 2012, 241.

⁵³² Vgl. z.B. den Abschnitt „Vorgehensweise beim Einstichprobenproblem (Anpassungstest)“ unter: <https://de.wikipedia.org/wiki/Kolmogorow-Smirnow-Test>.

43-mal an einer Zahlenstelle unter Gleichverteilungsverhältnissen von 206 Datensätzen entstehen kann, mit 0,0003 % ausfällt; folglich ist die Aussage des Chi-Quadrat-Tests für die Ziffernanalyse zutreffender. Sell kommt in Ihrer Dissertation bezüglich der Eignung von Chi-Quadrat- versus Kolmogorov-Smirnov-Test zur selben Einschätzung: „Wie unüblich die Verwendung des KS-Tests auf diskrete Verteilungen ist, zeigt u.a. die Tatsache, dass eine entsprechende Erweiterung der ‚ks.test‘-Funktion in der Statistiksoftware R erst in 2011 [...] erfolgt ist und in der Mehrzahl anderer Software-Anwendungen gar nicht verfügbar ist [...] Der KS-Test ist [...] anfällig für die Sortierung der Klassen [...] Der KS-Test scheint nach diesen Einschränkungen kaum geeignet, um im Rahmen der Ziffernanalyse auf Gleichverteilung der Ziffernhäufigkeiten zu testen. Hierfür sollte der Chi-Quadrat-Anpassungstest angewendet werden.“⁵³³ Als Standard für (dichotome) Wahrscheinlichkeitsfragen mit nur zwei möglichen Ergebnissen wie z.B. beim Münzwurf oder zwei (zusammengefassten) Ergebnisklassen⁵³⁴ ist die **Binomialverteilung** einschlägig, welche sich über die Kombinatorik bis auf die Grundrechenarten nachvollziehen lässt⁵³⁵. Aufgrund der Verbreitung und z.T. kostenlosen Zugänglichkeit von Tabellenkalkulationsprogrammen können die Ergebnisse musterbasierter Häufigkeitsanalysen von „sachverständigen Dritten“⁵³⁶ problemlos reproduziert werden: Die Anwendung hat sich durch Komplettfunktionen wie bspw. „CHISQ.TEST“ in MS Excel®, wo nur noch die Datenreihen zugewiesen werden müssen, gegenüber den früher üblichen Ergebnistabellen erheblich erleichtert.

I.R. der **SRP** werden alle manipulationsorientierten Musteranalysen wegen ihres indiziellen Charakters mit einem für die jeweilige Fragestellung gängigen Wahrscheinlichkeitstest auf Übereinstimmung geprüft. Zur besseren Transparenz werden beim χ^2 -Test (Chi-Quadrat-Test) auch die Zwischenergebnisse „ χ^2 -Klassenwert“ und „ χ^2 -Gesamtwert“ ausgegeben. Die Anwendungsempfehlung zur Mindestdatenanzahl „ ≥ 5 Erwartungen pro Klasse“ wird automatisch und durchgehend in der SRP-Vorlage Stochastik umgesetzt: Sobald – insbesondere infolge von Dateneinschränkungen auf dem Tabellenblatt „Daten“ (s. den Abschnitt „zweistu-

⁵³³ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 115, 141 u. 143.

⁵³⁴ Sollen Ziffernanalysen im Hinblick auf eine einzelne zu häufig oder selten vorkommende Ziffer durchgeführt bzw. vertieft werden, resultieren daraus die möglichen Ergebnisse „Ziffer x“ und „alle anderen Ziffern“, also eine (dichotome) ja-nein-Verteilung mit nur noch zwei Klassen. Die Idealwahrscheinlichkeiten richten sich dann nach der jeweilig zutreffenden Punkterwartung mit z.B. „10 %“ zu „90 %“ bei der Gleichverteilung.

⁵³⁵ Die (mathematische) Kombinatorik beschäftigt sich u.a. mit den möglichen Ergebnisvariationen (Permutationen) in diskreten – im Gegensatz zu kontinuierlichen – Datenverteilungen wie bspw. mit den Wahrscheinlichkeitsverhältnissen beim Werfen einer Münze. Durch das Aufstellen eines Ergebnisbaums aus den denkbaren Resultaten (bei zwei Münzwürfen mit den Seiten „A“ und „B“: „AA“, „AB“, „BA“ und „BB“) können die Wahrscheinlichkeiten für Einzelresultate (z.B. „genau zweimal Seite ‚A‘ bei drei Münzwürfen“) oder Ergebnisbereiche (z.B. „mindestens zweimal Seite ‚A‘ bei drei Münzwürfen“) auf Basis der Wahrscheinlichkeit des einzelnen Vorgangs (für eine ideal geprägte Münze „50:50“) problemlos abgelesen und berechnet werden.

⁵³⁶ § 145 Abs. 1 S. 1 AO.

fige Eingrenzung der untersuchten Daten“ unter [Pkt. 4.\)](#) – die empfohlene Mindestdatenmenge unterschritten wird, werden entweder die Testergebnisse ausgegraut (s. [Pkt. 4.3.2.1.](#)) oder die betreffenden Klassen in der Auswertung nicht berücksichtigt⁵³⁷ (s. [Pkt. 4.3.4.2.](#)). Als Wahrnehmungserleichterung bei eindeutig (signifikant) auffälligen Häufigkeitsmustern erfolgt auf den Tabellenblättern „digZA“ und „Verteilungsanalyse“ der SRP-Vorlage Stochastik eine Hervorhebung von Testergebnissen in den zwei Stufen „ $\geq 99\%$ “ durch rote Hintergrundfarbe und „ $\geq 99,95\%$ “ zusätzlich durch rote Schrift. Während die erste Auffälligkeitsschwelle – nach erfolgloser Abklärung besonderer einzelfallbezogener Ursachen – auf das Vorliegen mehrerer Indizien als -gesamtheit⁵³⁸ (s. [Pkt. 4.5.](#)) ausgerichtet ist, gilt die zweite strengere für Einzelindizien.

4.3.1.4. Schlussfolgerungen zu Häufigkeitsanalysen

Als Ziel von Prüfungshandlungen mündet die festgestellte (Nicht-)Übereinstimmung von Beobachtung und Erwartung in einer (rechtlich) bewertenden Schlussfolgerung. Solange zu auffälligen (Besteuerungs-)Daten detaillierte, verlässliche Grundaufzeichnungen vorliegen, ergibt sich der Effizienzvorteil leistungsfähiger Musteranalysen aus der Risikoorientierung von aufwendigen Vertiefungsmaßnahmen.⁵³⁹ Besonders bei der Prüfung von Bareinnahmen sind jedoch häufig keine bzw. keine verlässlichen Grundaufzeichnungen (mehr) vorhanden (s. [Pkt. 4.1.](#)), weshalb Gesamtbildmethoden direkt zur Einschätzung herangezogen werden müssen, ob „es nach [einer] Verprobung [...] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt.“⁵⁴⁰

Dabei „[können] Prüfungen vergangener Aufzeichnungen [...], nach dem logischen Menschenverstand, nur zu **Wahrscheinlichkeitsaussagen** führen. [...] Die herkömmlichen Methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...] Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine

⁵³⁷ Kurvenverteilungen, die sich (asymptotisch) der X-Achse annähern, wie die Normal- und die Logarithmische Normalverteilung fallen zwangsweise zu den Seiten hin irgendwann unter die Klassengrenze „ ≥ 5 “. Deshalb wird der χ^2 -Test (Chi-Quadrat-Test) als Übereinstimmungsprüfung auf die Logarithmische Normalverteilung i.R. der Verteilungsanalyse auf die Klassen mit einer Erwartung „ ≥ 5 “ beschränkt (s. [Pkt. 4.3.4.2.](#)).

⁵³⁸ BGH v. 27.4.2010 – 1 StR 454/09, NStZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreich[en].“

⁵³⁹ Vgl. z.B. den Einsatz der gleichmäßigen Verteilungen gem. Benford's Law (s. [Pkt. 4.3.2.3.1.](#)) als übergeordneten „Fraud test“ i.R. des „Data analytics process“ der Wirtschaftsprüfungsgesellschaft EY vor der eigentlichen Datenanalyse, nachzulesen unter: <https://eyfinancialservicesthoughtgallery.ie/wp-content/uploads/2013/11/EY-Data-Analytics-The-role-of-data-analytics-in-fraud-prevention.pdf>.

⁵⁴⁰ Anwendungserlass zu § 158 AO.

vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten⁵⁴¹. Zur Gewährleistung einer dennoch guten Nachvollziehbarkeit und Vergleichbarkeit von Gerichtsurteilen – und analog Prüfungsergebnissen – fordern erfahrene Gutachter⁵⁴² eine numerische Vereinheitlichung von typischen Wahrscheinlichkeitsaussagen (s. *ausführlich im Abschnitt „Konkretisierung der Aussagegewissheit“ unter [Pkt. 4.6.](#)*), um die zu beurteilenden Situationen sachgerecht am Maßstab der (gesetzlich) geforderten Klarheit bewerten zu können. In logischer Konsequenz sieht die Bildungswissenschaft in der „juristische[n] Methodenlehre [...] die Lehre von der richtigen und gerechten Entscheidung unter Unsicherheitslagen.“⁵⁴³ Als Spezialdisziplin für „Wahrscheinlichkeiten“ stellt die mathematische Stochastik die besten⁵⁴⁴ Hilfsmittel für möglichst eindeutige Einordnungen von Wahrscheinlichkeitsaussagen bereit.

Unter vereinfachendem Außerachtlassen der Wahrscheinlichkeitslogik nach Bayes⁵⁴⁵ führt als konkretes Anwendungsbeispiel eine mittels χ^2 -Test festgestellte Nichtübereinstimmung von 99 % i.R. einer Ziffernanalyse der Einerstelle von Tageskasseneinnahmen zu der Schlussfolgerung, dass die untersuchten Besteuerungsdaten mit „sehr hoher Wahrscheinlichkeit“ (s. [Pkt. 4.6.](#)) nicht unbeeinflusst von menschlichen Eingriffen entstanden sind. Um diese Erkenntnis als entsprechend beweiskräftiges **Indiz** für frei erfundene Bareinnahmen heranziehen zu können, muss aber zunächst noch abgeklärt werden, ob die beiden anderen Formen menschlicher Beeinflussung „Fehler bei der Erfassung“ und „Gestaltungen“ ursächlich sein können.⁵⁴⁶ Dabei kann die Gefahr fehlerhafter Untersuchungsdaten wie z.B. tlw. für mehrere Tage zusammengefasste Kassenerlöse mit Hilfe von Datenbankenfunktionen wie Sortieren, Lückenanalyse etc. zügig abgeklärt und das Ergebnis unter Angabe der verwen-

⁵⁴¹ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 323 u. 324.

⁵⁴² Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Reihe „Polizei + Forschung“, Bd. 26, herausgegeben vom Bundeskriminalamt (BKA), Kriminalistisches Institut, Luchterhand Verlag, 2004.

⁵⁴³ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 145.

⁵⁴⁴ Mit Aussagen wie: „*Wahrscheinlichkeitstheoretisch* geformte Algorithmen können Steueransprüche nicht begründen“ (Prinz, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4, M5), die konsequent weitergeführt bedeuten „menschliche Bauchentscheidungen können Steueransprüche besser begründen“, scheinen vereinzelte Autor*innen diesen wissenschaftlichen Konsens tatsächlich anzuzweifeln.

⁵⁴⁵ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 143 ff.: Zur genaueren Bewertung eines Indizes müssen neben der fehlerhaften Beanstandung (hier also 1 %) auch die Tatverbreitung und Aufdeckungswahrscheinlichkeit zumindest schätzungsweise bekannt sein.

⁵⁴⁶ Diese Maßnahmen entsprechen nach der Wahrscheinlichkeitstheorie einer weitest möglichen Bereinigung des Vergleichspools „nicht manipulierte Daten“ um diejenigen Datenbestände, die aus nicht-manipulativen Gründen nicht die erwartete Verteilung ausbilden, wodurch die Aussagesicherheit einer Nichtübereinstimmung als Hinweis auf Manipulationen (erheblich) zunimmt.

deten Daten nachvollziehbar dokumentiert werden. Der Einfluss möglicher Gestaltungen lässt sich i.d.R. bereits deshalb unproblematisch hinterfragen, weil Tageskasseneinnahmen lediglich von den zwei Faktoren „Nachfragemenge(n)“ und „Einzelpreis(e)“ abhängig sind: So können ausschließlich auf „0“ und „5“ Euro endende Einzelpreise ohne Nachkommastellen an der Einerstelle in Tagessummen auch nur diese Ziffern hervorbringen. Als weiteres Beispiel werden Sondertagesveranstaltungen mitunter mit einem Pauschalpreis abgegolten, sodass der Nachfragezufall keine freie Entstehung der Ziffernhäufigkeiten bewirken kann. Insgesamt ist für die manipulationsorientierten Musteranalysen i.R. der SRP ein (weitest gehender) Ausschluss anderer Gründe bei eindeutig (signifikant) abweichender Beobachtung ohne mathematische Spezialkenntnisse möglich.⁵⁴⁷

Nach dem SRP-Prüfungsschema (s. den Abschnitt „Prüfungsschema zur Vermeidung von Fehldeutungen“ unter [Pkt. 4.3.](#)) werden derart auffällige Ergebnisse manipulationsorientierter Musteranalysen selbständig auf eine Verursachung durch die (wenigen) praxisrelevanten betrieblichen Faktoren hinterfragt.⁵⁴⁸ Dazu helfen bei schwierigen Fragen die Funktionalitäten der SRP-Vorlage Stochastik, wenn z.B. abzuklären ist, ob eine typische Zweitziiffernhäufung auf einen in den Daten vorliegenden Dezimalsprung zurückgeführt werden kann⁵⁴⁹ (s. das entsprechende Beispiel unter [Pkt. 4.3.2.3.3.](#)). Sofern auf diese Weise die anderen Formen menschlicher Einflüsse auf die Tageskasseneinnahmen (weitest gehend) ausgeschlossen werden können, erreicht eine eindeutig (signifikant) auffällige Beobachtung den Status eines aussagekräftigen Indizes gegen die sachliche Richtigkeit.

Bisher ist die Akzeptanz manipulationsorientierter Musteranalysen in der Finanzrechtsprechung (noch) sehr zurückhaltend⁵⁵⁰. Ursachen dafür dürften die Komplexität⁵⁵¹ der Me-

⁵⁴⁷ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Dabei ist allerdings die sphärenorientierte Mitwirkungspflicht des Stpfl. (s. § 88 Rz. 3 f.; § 90 Rz. 20) zu beachten. Mathematisch eindeutige Unschlüssigkeiten der Buchführungsergebnisse muss er durch betriebsspezifische Besonderheiten schlüssig erklären“.

⁵⁴⁸ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 139 u. 161: „An die Beobachtungen aus Chi-Quadrat-Anpassungstest und Binomialtest schließen sich in der Praxis der Betriebsprüfung weitere Prüfungshandlungen an [...] Außerdem sind bei jeglicher Analyse die betrieblichen Besonderheiten des zu prüfenden Unternehmens einzubeziehen, um eine Verzerrung der Zahlenstruktur aufgrund von Preisgestaltungen oder bspw. Sitzplatzkapazitäten auszuschließen. Eine Übersicht der dafür durchgeführten Maßnahmen in der praktischen Anwendung der Betriebsprüfung bietet das Handbuch zur Summarischen Risikoprüfung [...] Die Verteilung der Tageseinnahmen kann sich je nach betrieblicher Besonderheit bspw. in den einzelnen Geschäftsjahren, saisonal oder wochentagsspezifisch unterscheiden. Dies ist in den Analysen berücksichtigt. Ein solcher Umstand einer ‚uneinheitlichen Absatzsituation‘ wird in der Praxis ebenfalls regelmäßig und sehr tiefgehend geprüft“.

⁵⁴⁹ Die praxisrelevanten betrieblichen Besonderheiten, die ein auffälliges Häufigkeitsmuster der Einerstelle, Zweitziiffer oder Werteverteilung verursacht haben können, werden unter den jeweiligen Abhandlungen der Analysen beschrieben (s. [Pkt. 4.3.2.2.2.](#), [Pkt. 4.3.2.3.2.](#) sowie [4.3.4.4.](#)).

⁵⁵⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit (s. FG Münster v. 14.8.2003 – 8 V 2651/03 E, U, EFG 2004, 9 m. Anm. Trossen; FG Münster v. 10.11.2003 – 6 V 4562/03 E, U, EFG 2004, 236 [237]; FG Münster v. 7.12.2005 – 1 K 6384/03 E, EFG 2006, 652 [653] m. Anm. Zimmermann [Fahrtenbuch];

thoden und fragwürdige Veröffentlichungen⁵⁵² sein. Deshalb wurde in der SRP von Beginn an sehr großen Wert auf die Unterscheidung der einzelnen Analyseschritte und deren korrekte Beschreibung gelegt.

4.3.2. Ziffernanalyse

(Digitale) Ziffernanalysen verfolgen das Ziel, das **freie Erfinden von (Besteuerungs-) Daten** durch Spuren aufzudecken, die infolgedessen unbewusst hinterlassen wurden: Vorlieben für bzw. Abneigungen gegen bestimmte Ziffern oder Ziffernchaos. Untersuchungen zum Ausdenken von Zahlen haben gezeigt, dass dabei häufiger einzelne Ziffern so stark bevorzugt oder unterdrückt werden, dass sie außerhalb der zufällig möglichen Häufigkeiten liegen.⁵⁵³ Weil das natürliche Streuverhalten (um ein typisches Ergebnis herum) bekannt und im mathematischen Teilgebiet der Wahrscheinlichkeitslehre (Stochastik) zusammengefasst ist (s. [Pkt. 4.3.1.3.](#)), können konkrete (prozentuale) Aussagen zu auffälligen Ziffernkonstellationen darüber getroffen werden, wie sicher es ist, dass die untersuchten (Besteuerungs-)

FG Rh.-Pf. v. 24.8.2011 – 2 K 1277/10, EFG 2012, 10 [11] m. Anm. Kühnen; Sosna, StBp. 2000, 41 [Teil I], 68 [Teil II]; Meyer, DStR 2005, 2114; Watrin/Stuffert, DB 2006, 1748; Huber, StBp. 2009, 121 [124 ff.]; krit. gegenüber der zurückhaltenden FG-Rspr.: Wähnert, StBp. 2011, 269 [271]; Wähnert, StBp. 2012, 241 [244]; Wähnert, Stbg. 2014, 20 [22 f.).“

⁵⁵¹ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Nomos, 2018, 56-57 u. 65: „Bei der Verfügbarkeitsheuristik schätzen Individuen die [...] Wahrscheinlichkeit für das Eintreten eines Ereignisses unter Unsicherheit danach ein, wie leicht bzw. schwer es ihnen fällt, damit verbundene Erscheinungen aus ihrem Gedächtnis abzurufen. Dabei werden Ereignisse für wahrscheinlicher erachtet, wenn sie sich leichter vorgestellt werden können und sich zu ihnen eine direkte Verfügbarkeit herstellen lässt. [...] Im Rahmen juristischer Entscheidungen scheint es unerlässlich, von diesen Entscheidungsverzerrungen Kenntnis zu haben, da eine verzerrte Entscheidung den Denkgesetzen der Logik widerspricht [...]“.

⁵⁵² Vgl. z.B.: Bisle, „Chi-Quadrat-Test‘ und Zeitreihenvergleich: Keine Schätzung bei ordnungsgemäßer Buchführung“, PStR 1/2012, 15: „Der ‚Chi-Quadrat-Test‘ analysiert die Verteilung bestimmter Ziffern und basiert auf der Erkenntnis, dass jeder Mensch Lieblingsziffern hat und bei frei erfundenen Zahlen (z.B. nachgetragenen Tageseinnahmen oder -ausgaben) häufiger verwendet als andere.“ Tatsächlich ist der χ^2 -Test (Chi-Quadrat-Test) ein allgemeiner Wahrscheinlichkeitstest, der mehrklassige Verteilungen unter Berücksichtigung der Erkenntnisse der Wahrscheinlichkeitslehre auf Übereinstimmung prüft. Eine (zwingende) Verbindung zur Gleichverteilung von endständigen Ziffern in Wirtschaftsdaten besteht nicht. Vielmehr vermischt die Aussage die zwei Schritte „Erwartung“ und „Übereinstimmungsprüfung“ i.R. der musterbasierten Analyse und sorgt somit für Verständnisprobleme und Fehleinschätzungen; so auch: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 194: „[...] irritiert, dass die Verprobungsmethode ‚Ziffernanalyse‘ regelmäßig als ‚Chi-Quadrat-Test‘ geführt wird.“

⁵⁵³ Hill, „The Difficulty of faking Data“, Chance 12/3/1999, 27, sowie Hill, „Random-Number Guessing and the First Digit Phenomenon“, Psychological Reports 62/1988, 967. Derweil die natürliche Zufallsstreuung um die Idealerwartung herum bei der üblichen Anzahl an i.R. von Betriebsprüfungen untersuchten Tageskasseneinnahmen beträchtlich ist (s.a. [Pkt. 4.3.1.3.](#)), lassen sich auf diesem Weg nur starke Ziffernvorlieben und -abneigungen ausreichend sicher feststellen, weshalb diese Aussage im BFH-Urteil v. 16.9.2021 ([BStBl. II 2022, 101](#)) irritieren muss: „So hat den Feststellungen des FG zufolge auch der in der mündlichen Verhandlung vor dem FG anwesende Kassensystemprüfer glaubhaft angegeben, er selbst habe trotz genauer Kenntnis des [Chi-Quadrat-]Tests keine manipulierte Buchführung erstellen können, die den Test bestanden hätte.“

Daten menschlich beeinflusst worden sein müssen. Der Umkehrschluss, dass bspw. Tageskasseneinnahmen nicht frei erfunden wurden, lässt sich mit Ziffernanalysen jedoch *nicht* qualifiziert bestätigen.

Das wichtigste **Manipulationsszenario**, welches mit Ziffernanalysen aufgedeckt werden kann, ist die manuelle Aufzeichnung i.d.R.⁵⁵⁴ von Tageskasseneinnahmen. Darüber hinaus können z.B. durch die folgenden moderneren (digital unterstützten) Eingriffsformen in Barerlösen eindeutige Ziffernvorlieben oder -abneigungen als „Fingerabdrücke“ in die Ziffernverteilungen gelangen:

-  Beim „Zwei-Kassen-Modell“ („Double-Till“) wird zusätzlich zur Geschäftskasse mit den korrekten Erfassungen (für die Personalkontrolle) eine zweite desselben Typs im Verborgenen mit Fantasiiezahlen gespeist. Um den Arbeitsaufwand im erträglichen Rahmen zu halten, werden dabei oft nur Tagessummen für die relevanten Berichtspositionen eingegeben und verräterische Kassenberichtsangaben unterdrückt. Somit können sich (unbewusste) Ziffernvorlieben oder -abneigungen in die steuerlich aufgezeichneten Tageskasseneinnahmen hineinschleichen.
-  Die bei älteren Registrierkassenmodellen nach Prüfungserfahrungen verbreitetere Praxis, zunächst per X-Abschlag (ohne Nullstellung) die Bareinnahmen zu kontrollieren, anschließend mittels Manager-Storno zu löschen und einen Wunschbetrag für die Besteuerung einzugeben,⁵⁵⁵ kann ebenso auffällige Ziffernhäufigkeiten hervorrufen.
-  Selbst einige Varianten systematischer digitaler Manipulation beinhalten die Eingabe eines „gewünschten“ Betrags für die Tageskasseneinnahmen. So sahen vor allem frühere Zapperversionen⁵⁵⁶ für Warenwirtschaftsprogramme in Apotheken vor, dass die „neue“ Tageskasseneinnahme in ein bestimmtes Feld eingegeben wird. Danach erfolgte die intelligente Umverteilung auf die im Bericht auszugebenden Positionen wie die Aufteilung nach USt-Sätzen, Zahlungswegen etc.

Demgegenüber können Manipulationsweisen, die auf dem (tlw.) Löschen von Positionen oder auf einer prozentualen Reduktion von Werten beruhen, nicht durch Ziffernanalysen aufgedeckt werden, weil die verkürzten Tageskasseneinnahmen weiterhin ein Produkt zufälliger Entstehung sind.

⁵⁵⁴ Denkbar ist z.B. auch die (Vor-)Untersuchung (einer Vielzahl) von Bilanzen (in Konzernen) auf frei erfundene Zahlenmanipulationen bspw. zur Täuschung von Aktionär*innen oder Banken.

⁵⁵⁵ Auch wenn diese Kassentypen gem. § 146a Abs. 1 S. 2-3 AO nicht mehr verwendet werden dürfen, ist die Kenntnis solcher Manipulationsszenarien im Hinblick auf ordnungswidrige (Risiken für die inhaltliche Richtigkeit) Aufzeichnungen von Barerlösen und auf mögliche zukünftige Hinterziehungsweisen hilfreich.

⁵⁵⁶ Unter „Zappern“ werden externe Manipulationstools für elektronische Kassen(-systeme) verstanden, die als Programme z.B. auf einem USB-Stick liegen. Nach Ankopplung an das Aufzeichnungssystem gewähren sie den Lösch- und Änderungszugriff auf die gespeicherten Daten.

Tipp: Für das grundsätzliche Verständnis manipulationsorientierter Musteruntersuchungen sind die Ziffernanalysen hilfreich, weil hieraus die Übertragung der einzelnen Annahmen und Schritte des allgemeinen Ablaufs (s. [Pkt. 4.3.1.](#)) auf komplexere Ansätze wie besonders die wertbezogene Verteilungsanalyse (s. [Pkt. 4.3.4.](#)) leichter fällt.

4.3.2.1. Anwendungshinweise zu SRP-Ziffernanalysen

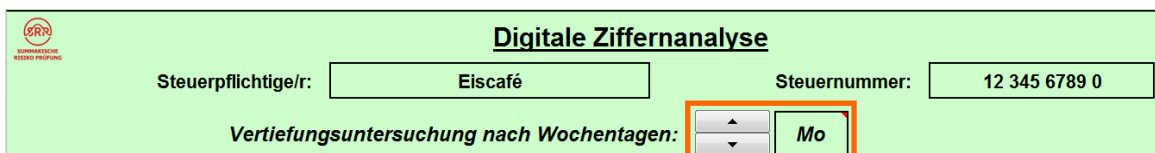
Als i.d.R. auf Tageskasseneinnahmen beschränkt angewendete Form der Musteranalytik (s. [Pkt. 4.3.2.](#)) findet sich das Tabellenblatt „digZA“ mit den Ziffernanalysen innerhalb der SRP-Vorlage Stochastik im zweiten Teil nach der „Kurzanleitung 2“. Um deren *sinnvollen* Einsatz zu unterstützen, *muss das Tabellenblatt bewusst über ein Anhakfeld mit entsprechendem Hinweis aktiviert werden.*

Folgende allgemeine Hinweise zur Durchführung von Ziffernanalysen mit der SRP-Vorlage Stochastik sind außerdem wichtig:

- ➔ Sofern **mehrere Datensätze mit demselben Datum** auf dem Tabellenblatt „Daten“ hineinkopiert werden, weist die Rotfärbung der Überschrift „Datum“ als Warnung darauf hin, weil das wertbezogene Benford-Gesetz (s. [Pkt. 4.3.2.3.1.](#)) eine vergleichbare Konsolidierung der Daten i.d.R. als Tageskasseneinnahmen voraussetzt; die Auswertungen werden dadurch – u.a. wegen der Untersuchung der Einerstelle auf die Gleichverteilung – nicht abgebrochen.
- ➔ Sobald zu untersuchende Daten in den beiden Spalten „Datum“ und „alle Werte“ auf dem Tabellenblatt „Daten“ enthalten sind und der Aktivierungshaken neben der Überschrift des Tabellenblatts „digZA“ gesetzt ist, werden die Einerstelle⁵⁵⁷ und die Zweitziffer automatisch nach ihren Mustern ausgewertet und den jeweiligen (Punkt-)Erwartungen (s. [Pkt. 4.3.1.2.](#)) „Gleich-“ bzw. „Benford-Verteilung“ gegenübergestellt. Der **redundante Ansatz** mit zwei untersuchten Zahlenstellen soll eine bessere Aufdeckungswahrscheinlichkeit für Ziffernauffälligkeiten erreichen, weil viele Manipulationsspuren im weiten Streuungsbereich des Zufalls untergehen [s.a. im Bereich „freie Erfinden von (Besteuerungs-)Daten“ unter [Pkt. 4.3.2.](#)].
- ➔ Daten mit **negativem Wert** werden nicht in die Ziffernanalysen einbezogen, weil dies im Widerspruch zu ihrem typischen beschränkten Einsatzgebiet „manipulationsorientierte Untersuchung von Tageskasseneinnahmen auf Ziffernvorlieben bzw. -abneigungen oder -chaos“ stünde; im Haupthinweis des Tabellenblatts „digZA“ wird diese Maßnahme angegeben.

⁵⁵⁷ Bei der Einerstelle werden Werte unter € 100 aus der Analyse ausgeschlossen, weil bei diesen die Einerstelle auf die beiden führenden Zahlenpositionen fällt und somit deren abweichende Benford-Verteilungen zu erwarten sind (s.a. [Pkt. 4.3.2.2.1.](#)).

- ➔ Beide (kompletten) Ziffernanalysen werden mit einem χ^2 -Test (**Chi-Quadrat-Test**) auf Übereinstimmung geprüft (s. [Pkt. 4.3.1.3.](#)). Wird die empfohlene Mindestdatenmenge „ ≥ 5 Elemente pro Erwartungsklasse“ unterschritten, verfärben sich die betroffenen Felder der Erwartungsspalte rot und wird das χ^2 -Testergebnis mit hellgrauer Schrift unterdrückt⁵⁵⁸. Mathematisch eindeutige (signifikante) Abweichungen vom jeweiligen Erwartungsmuster werden ab 99 % Nichtübereinstimmung mit roter Feldfarbe und ab 99,95 % zusätzlich mit roter Schrift hervorgehoben. Außerdem werden (per Anhakfeld ausblendbar) die Hinweise „Betriebliche Ursachen für Einerstellenauffälligkeiten wie die Preisgestaltung sollten abgeprüft werden.“ bzw. „Betriebliche Ursachen für Zweitziffernauffälligkeiten wie der Dezimalsprung sollten abgeklärt werden.“ angezeigt.
- ➔ Fallen in den Gegenüberstellungen der beobachteten und erwarteten Ziffernhäufigkeiten einzelne Ziffern als besonders häufig oder selten auf, werden sie in der betreffenden Vergleichstabelle durch rote Schriftfarbe betont. Die Untersuchung sollte aus Sicht dieser möglichen Vorlieben bzw. Abneigungen vertieft werden. Als entsprechende Untersuchungsoption findet sich die **Binomialverteilung**⁵⁵⁹ (s.a. [Pkt. 4.3.1.3.](#)) in einem eigenen Bereich jeweils unterhalb der kompletten Häufigkeitsauswertung (s. *das Beispiel unter [Pkt. 4.3.2.2.3.](#)*).
- ➔ Die SRP-Vorlage Stochastik bietet eine **zweistufige Eingrenzungsmöglichkeit** der untersuchten Daten, denn manipulative Eingriffe in Tageskasseneinnahmen erfolgen nicht (zwingend) ständig und entwickeln sich weiter, sodass Spuren wegfallen können.



Die Eingrenzungsoptionen für Jahre und **Wochentage** auf dem Tabellenblatt „digZA“ **mittels Drehfeld** sind für den schnellen Überblick gedacht und wirken nur auf die zugeordneten Ziffernanalysen. So kann insbesondere zügig nachvollzogen werden, ob beobachtete Ziffernauffälligkeiten in Teildatenbeständen oder deren Kombinationen zu einem intensiveren Manipulationsverdacht führen.⁵⁶⁰ Zur weiteren Abklärung einer solchen Mani-

⁵⁵⁸ Bei unterdrückten Ergebnissen des χ^2 -Tests (Chi-Quadrat-Test) ist die formatseitige Hervorhebung auffälliger Musterabweichungen ebenfalls deaktiviert.

⁵⁵⁹ Mit der Vertiefung einer Ziffernanalyse auf eine bestimmte Ziffer ändern sich die Beobachtungs- und Erwartungsverteilung in die (dichotome) ja-nein-Gegenüberstellung „Ziffer x versus alle anderen Ziffern“, für welche die Binomialverteilung die einschlägige Beschreibung der Wahrscheinlichkeitsverhältnisse bildet. Die Idealwahrscheinlichkeiten richten sich dann nach der jeweilig zutreffenden Punkterwartung mit z.B. „10 %“ zu „90 %“ bei der Gleichverteilung.

⁵⁶⁰ Jede Eingrenzung des Datenbestands bewirkt dabei entsprechend der stochastischen Grundlagen zu Mustervergleichen, dass die Erwartung an die Mustertreue abnimmt und somit der Wahrscheinlichkeitstest weni-

pulationsspur können die untersuchten Daten auf dem Tabellenblatt „Daten“ grundsätzlich **für alle Analysen** eingeschränkt werden, was auf jedem Auswertungsblatt mit einem entsprechenden Hinweis angezeigt wird.

In diesem Beispiel soll

Einschränkung auf Kalenderjahr(e):		
Teilzeitraum (z.B. Saison):	1 5 bis 30 9	
Wochentag(e):	5 6	

hinterfragt werden, ob die Freitage und Sonnabende der Sommersaison „1.5. bis 30.9.“ besonders auffällige Ziffernmuster aufweisen, denn es ist aus manipulationspsychologischer Sicht denkbar, dass an erfahrungsgemäß einnahmestärkeren Tagen eher verkürzt wird. Auf diese Weise können bspw. auch Unterschlagungen an bestimmten Wochentagen von Einnahmenhinterziehungen unterschieden werden.

- ➔ Ein **Ausschluss einzelner Datensätze** kann zusätzlich über die Spalte mit Ankreuzkästchen links neben den hineinkopierten Daten (Tabelleblatt „Daten“) vorgenommen werden. Auch diese Abwahl wirkt auf die kompletten Untersuchungen der SRP-Vorlage Stochastik und wird auf den Auswertungsblättern mit dem Hinweis „untersuchte Daten: Einzeldatenabwahl“ vermerkt. Sie ist besonders für die gemeinsame Sachaufklärung⁵⁶¹ geeignet, wenn z.B. bei entdeckten Ziffernauffälligkeiten einzelne Tageskasseneinnahmen ausgeschlossen werden sollen, weil sie zu pauschal abgerechneten Veranstaltungen gehören. Über den Autofilter zur Abwahlspalte kann der jeweilige Stand der gemeinsamen Aufklärung komfortabel dokumentiert werden.
- ➔ Auf dem Tabellenblatt „Schichtung“ besteht ferner die Möglichkeit, die Daten erst ab einem frei wählbaren Schichtungsstartwert (**Mindestwert**) zu untersuchen, was zur allgemeinen Anwendung der Musteranalytik mitunter hilfreich ist (s. *das Beispiel im Abschnitt „systematisch-mehrperspektivische Musteranalytik“ unter Pkt. 5.1.1.*). Während eine derartige wertabhängige Einschränkung der Daten für den absoluten Zufall „Gleichverteilung“ i.R. der Einerstellenuntersuchung unschädlich ist, steht sie im Widerspruch zur Entstehungstheorie des Benford-Gesetzes (s. *den Abschnitt „Mathematisch-theoretisch“ unter Pkt. 4.3.2.3.1.*); deshalb zeigt die SRP-Vorlage Stochastik die Eingabe eines Startwerts in allen Häufigkeitstabellen zur Zweitziffernuntersuchung über Rotfärbung der erwarteten Gesamtanzahl an.
- ➔ Die Auswertungsblätter der SRP-Vorlage Stochastik zu den Ziffernanalysen sind ausdrucksfähig und geben für eine nachvollziehbare **Dokumentation** folgende In-

ger streng bewertet (s. *das Beispiel im Abschnitt „Güte der Übereinstimmung mit dem Idealmuster von der Datenmenge abhängt“ unter Pkt. 4.3.*).

⁵⁶¹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit [...] Dabei ist allerdings die sphärenorientierte Mitwirkungspflicht des Stpfl. (s. § 88 Rz. 3 f.; § 90 Rz. 20) zu beachten. Mathematisch eindeutige Unschlüssigkeiten der Buchführungsergebnisse muss er durch betriebsspezifische Besonderheiten schlüssig erklären“.

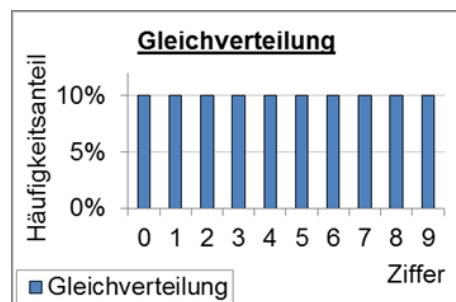
te wieder: Tabellen mit den beobachteten und den – gemäß Gleich- bzw. Benford-Verteilung – erwarteten Ziffernhäufigkeiten sowie allen (Zwischen-)Ergebnissen des jeweiligen χ^2 -Tests (Chi-Quadrat-Test), dazugehörige Grafiken mit den verbildlichten Muster-
vergleichen und – bei Aktivierung – die Auswahl der Einzelziffer mit den erforderlichen Parametern für die sowie dem Ergebnis der Binomialverteilung als Übereinstimmungstest. Damit werden alle Informationen ausgegeben, um die Ziffernanalysen mit allgemein zugänglichen Mitteln wie Tabellenkalkulationsprogrammen reproduzieren zu können; lediglich die verwendeten Daten müssen dazu z.B. auf dem Tabellenblatt „Datenquelle“ dokumentiert werden.

4.3.2.2. Einerstellenuntersuchung

Bei der Einerstellenuntersuchung wird die Ziffernverteilung an der Position vor dem Währungskomma in den zu untersuchenden Daten auf Übereinstimmung mit der „**Gleichverteilung**“ aller (möglichen) Ziffern geprüft.

4.3.2.2.1. Die (Punkt-)Erwartung „Gleichverteilung“

Die Annahme einer absolut zufälligen (gleichverteilten) Einerstelle besonders in Tageskasseneinnahmen gründet sich auf die **mathematisch-theoretische** Überlegung, dass der Nachfragezufall „Anzahl an verkauften Artikeln bzw. Leistungen“ durch die Multiplikation mit den jeweiligen Einzelpreisen nicht aufgehoben wird. Im Ergebnis würden sich die Ziffernverteilungen in den Tagessummen der Kasseneinnahmen genauso zufällig wie die täglichen Nachfragehäufigkeiten zusammensetzen.



Die **empirischen**⁵⁶² **Erkenntnisse** des Benford-Gesetzes⁵⁶³ bestätigen die Gleichverteilung der Ziffern annähernd ab der Drittziffer mit zunehmender Genauigkeit zu den endständigen Positionen hin (s. [Pkt. 4.3.2.3.1.](#)). Wegen der (erheblich) abweichenden Erfahrungen zu den ersten beiden Zahlenstellen gemäß Benford-Gesetz werden in der SRP-Vorlage Stochastik Einzeldatensätze, die kleiner als € 100 sind, automatisch für die Einerstellenuntersuchung ausgeschlossen, weil deren **Einerstelle** auf einer der **ersten beiden Zahlenpositionen** liegt.

1.276,20 Tageskasseneinnahme
→ ↑ Blickrichtung

tersuchung ausgeschlossen, weil deren **Einerstelle** auf einer der **ersten beiden Zahlenpositionen** liegt.

⁵⁶² Empirische Erkenntnisse stammen aus Studien, die den Untersuchungsgegenstand unter nachvollziehbaren konstanten Bedingungen beobachten, und sollen theoretische Annahmen entweder bestätigen oder verwerfen. Gesicherte empirische Erkenntnisse genießen einen höheren Stellenwert in der Wissenschaft als widerstreitende theoretische Annahmen, weil sie die Realität beschreiben.

⁵⁶³ Benford, „The Law of Anomalous Numbers“, Proceedings of the American Philosophical Society 1938, 551.

Obwohl die empirischen Beobachtungen zur Drittziffernverteilung gem. Benford-Gesetz eine Maximalabweichung der „0“ von knapp 1,8 % gegenüber der Gleichverteilung aufweisen, ist eine Anwendung dieser auch auf dreistellige⁵⁶⁴ Tageskasseneinnahmen im Gesamtzusammenhang sinnvoll und sachgerecht⁵⁶⁵:

-  Die Manipulationsform „frei erfundene Zahlen“, welche mit Ziffernanalysen aufgedeckt werden kann, kommt erfahrungsgemäß am ehesten bei kleineren Bargeschäftsbetrieben mit Tageskasseneinnahmen unter € 1.000 vor.
- ⊙ Der maximale Fehler von 1,8 % zur erwarteten Gleichverteilung muss mit zwei anderen Größen im Kontext bewertet werden: Erstens wird eine Einerstellenauffälligkeit erst ab einer Nichtübereinstimmungswahrscheinlichkeit von 99,95 % als Einzelindiz aufgegriffen; demgegenüber gilt zweitens das gesetzlich reduzierte Beweismaß für die Beanstandung gem. § 158 Abs. 2 Nr. 1 AO (s. [Pkt. 4.6.](#)). Zudem treten derart eindeutige (signifikante) Ziffernauffälligkeiten nach ständigen Prüfungserfahrungen nur im Verbund mit gravierenden Ordnungsmängeln auf (s. [Pkt. 4.4.](#)).
- ☑ Darüber hinaus werden i.R. der SRP festgestellte Ziffernauffälligkeiten so konkret nachverfolgt (s. *das Beispiel unter [Pkt. 4.3.2.2.3.](#)*), dass Erfassungsfehler und einzelfallbezogene Sonderfaktoren als nicht-manipulative Ursachen zügig und mit geringem Aufwand abgeklärt werden können⁵⁶⁶.

Die Übereinstimmungsprüfung des beobachteten Häufigkeitsmusters aller Ziffern an der Einerstelle in untersuchten Tageskasseneinnahmen mit der Gleichverteilung erfolgt durch den χ^2 -Test (Chi-Quadrat-Test; *zur Gegenüberstellung der Eignung von Chi-Quadrat- und Kolmogorov-Smirnov-Anpassungstest s. den Abschnitt „Chi-Quadrat-Test“ unter [Pkt. 4.3.1.3.](#)*); dadurch beinhaltet das Übereinstimmungsergebnis bereits die Anzahl untersuchter Daten (s. [Pkt. 4.3.1.3.](#)). Die übliche Empfehlung zur Mindestdatenmenge von 50 Datensätzen bei der Erwartung „Gleichverteilung aller Ziffern“ wird von der SRP-Vorlage Stochastik automa-

⁵⁶⁴ Gemeint sind Tageskasseneinnahmen im Bereich „≥ € 100 bis € 999,99“.

⁵⁶⁵ Bestätigend: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 191: „Die Einerstelle bezeichnet die Stelle vor dem Komma, die, sofern sie auf mindestens dreistellige Beträge angewendet wird, für einen Test auf Gleichverteilung nicht zu beanstanden ist.“

⁵⁶⁶ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Dabei ist [...] die sphärenorientierte Mitwirkungspflicht des Stpfl. (s. § 88 Rz. 3 f.; § 90 Rz. 20) zu beachten. Mathematisch eindeutige Unschlüssigkeiten der Buchführungsergebnisse muss er durch betriebspezifische Besonderheiten schlüssig erklären“. Vgl. a.: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 139: „An die Beobachtungen aus Chi-Quadrat-Anpassungstest und Binomialtest schließen sich in der Praxis der Betriebsprüfung weitere Prüfungshandlungen an [...] Außerdem sind bei jeglicher Analyse die betrieblichen Besonderheiten des zu prüfenden Unternehmens einzubeziehen, um eine Verzerrung der Zahlenstruktur aufgrund von Preisgestaltungen oder bspw. Sitzplatzkapazitäten auszuschließen. Eine Übersicht der dafür durchgeführten Maßnahmen in der praktischen Anwendung der Betriebsprüfung bietet das Handbuch zur Summarischen Risikoprüfung“.

tisch beachtet. Formatseitig werden die zwei Auffälligkeitsschwellen „99 %“ und „99,95 % Nichtübereinstimmung“ hervorgehoben (s. [Pkt. 4.3.2.1.](#)).

4.3.2.2.2. Nicht-manipulative Ursachen für Einerstellenauffälligkeiten

Einzelfallbezogene Besonderheiten als Ursache für mathematisch eindeutige (signifikante) Ziffernauffälligkeiten der Einerstelle können nur aus den Ursprungsgrößen der Tagesskasseneinnahmen „Nachfragehäufigkeit(en)“ und „Einzelpreis(e)“ stammen:

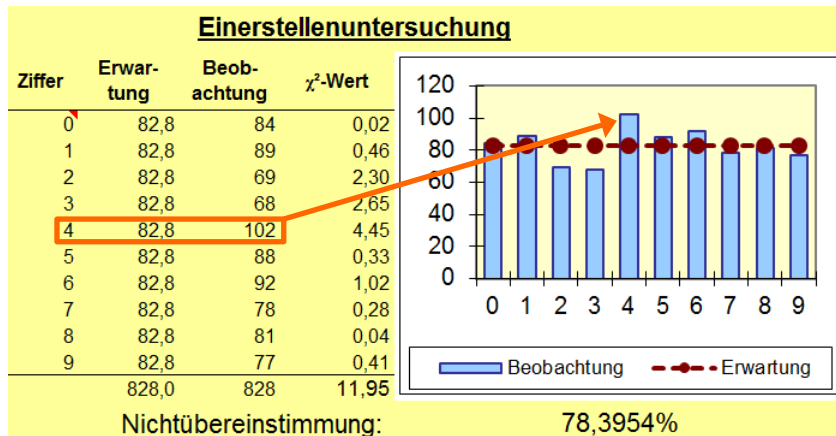
1. Praxisuntypische **besondere Einzelpreise** wie bspw. (fast) ausschließlich auf „0“ bzw. „5“ endende ohne Nachkommastellen können in der Multiplikation mit beliebigen Häufigkeiten ebenso nur die Einerstellen „0“ und/oder „5“ erzeugen. Diese Ausnahme lässt sich anhand der detaillierten Untersuchungsergebnisse sofort erkennen und durch Preislisten bestätigen.
2. Bei einer **täglich geringen Anzahl von verkauften Artikeln bzw. Leistungen** kann sich eine besondere Preisgestaltung in den Tagessummen tlw. wiederfinden: Wenn z.B. *sehr wenige* hochpreisige Produkte, welche alle auf „€ #.??9“ enden, die Tagesskasseneinnahmen bilden, werden die Endziffern „9“, „8“, „7“ etc. auf Kosten der kleineren gehäuft vorkommen. Diese Besonderheit lässt sich ebenfalls am dennoch gleichmäßigen Einerstellenmuster umgehend erkennen und durch die – dann zwingend erforderlichen – Einzelaufzeichnungen nachvollziehen.
3. Tagesveranstaltungen, die über **Pauschalpreise** abgerechnet werden, entstehen nicht durch den Nachfragezufall sondern durch Preisverhandlungen. Befinden sich derartige Sondervorgänge in den untersuchten Daten, können sie problemlos und nachvollziehbar durch Filtern und Ankreuzen auf dem Tabellenblatt „Daten“ der SRP-Vorlage Stochastik aus allen Analysen ausgeschlossen werden.

Hinweis: Die Untersuchung von Einzelvorgängen aus Kassendaten auf die Gleichverteilung der Ziffern an der Einerstelle ist nicht zielführend, weil nicht der kombinatorische Zufall „zufällige Mengen multipliziert mit Einzelpreisen“ wirkt, sondern die im Unternehmen festgelegte Preisstruktur die Ziffernverteilung bestimmt.

4.3.2.2.3. Vorlagenunterstützte Abklärung von Einerstellenauffälligkeiten

In der (manipulationsorientierten) **SRP-Musteranalytik** ist eine selbständige (vorlagenunterstützte) Nachverfolgung von Einerstellenauffälligkeiten vorgesehen. Einerseits können dadurch betriebliche Besonderheiten des Einzelfalls als nicht-manipulative Verursacher erkannt bzw. (weitest möglich) ausgeschlossen werden. Andererseits können partielle Manipulationen, wenn z.B. nur an erlösstärkeren Wochenendtagen eingegriffen wurde, aufgedeckt oder Unterschlagungen von Einnahmenverkürzungen besser unterschieden werden.

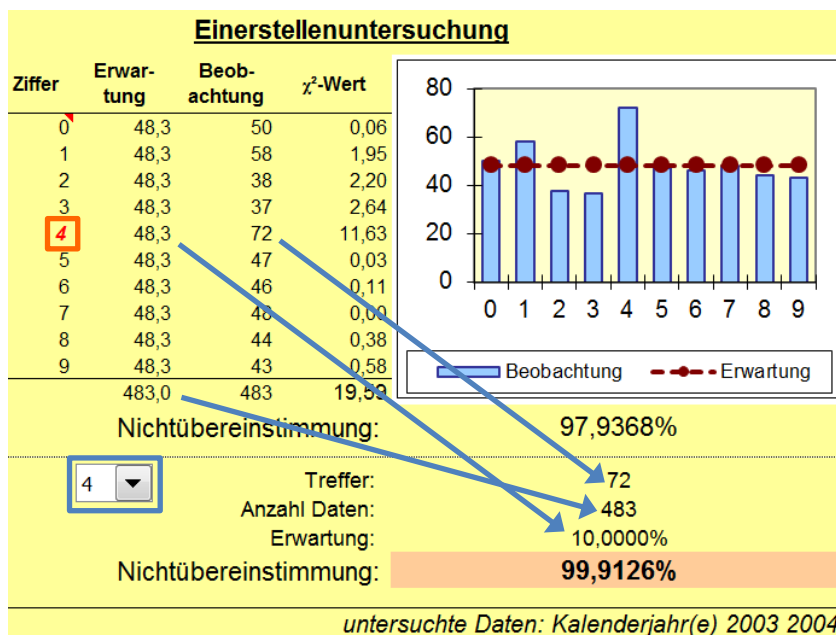
Diese beispielhafte Einerstellenverteilung zu den Tageskasseneinnahmen eines Eiscafés offenbart klare Unterschiede zwischen den Beobachtungssäulen und der dazugehörigen Gleichverteilung. Trotzdem ermittelt der χ^2 -Test (Chi-Quadrat-Test) nur eine Nichtübereinstimmungswahrscheinlichkeit von ca. 78 % in



Abhängigkeit von der untersuchten Datenmenge „828 Tage“. Damit werden die beiden Auffälligkeitsschwellen der SRP-Vorlage Stochastik „99 %“ bzw. „99,95 %“ (s. [Pkt. 4.3.2.1.](#)) nicht erreicht. Im verbildlichten Mustervergleich fällt allerdings die **Häufung der „4“** auf, was sich als punktuelle Spur einer Ziffernvorliebe in frei erfundenen Besteuerungsdaten herausstellen könnte.

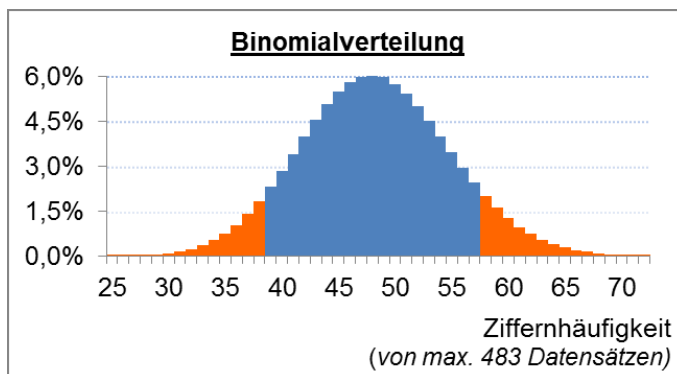
Um diese Beobachtung besser zu verstehen, werden zunächst die drei Prüfungsjahre mittels Drehfeld auf dem Tabellenblatt „digZA“ der SRP-Vorlage Stochastik einzeln betrachtet (s. den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)). Dadurch lässt sich zügig feststellen, dass die Häufung der „4“ lediglich in den ersten beiden Jahren auftritt, im letzten nicht mehr. Gerade bei einfachen Eingriffsarten wie dem freien Erfinden von Zahlen ist das plötzliche Verschwinden von Manipulationsspuren nicht ungewöhnlich und kann Folge einer Umstellung der bisherigen Hinterziehungsweise auf eine raffiniertere [s.a. den Abschnitt „raffinierte (digitale) Manipulationswerkzeuge“ unter [Pkt. 2.1.](#)] sein.

Deshalb wird die Einerstellenuntersuchung durch generelle Einschränkung der Daten im Tabellenblatt „Daten“ auf die ersten beiden Jahre begrenzt. In der Konsequenz gibt der χ^2 -Test (Chi-Quadrat-Test) nun eine Nichtübereinstimmungswahrscheinlichkeit von ca. 98 % aus. Zwar werden die beiden Auffälligkeitsschwellen weiterhin leicht unterschritten, die **Ziffer „4“** wird jetzt aber in



der Tabelle mit den Häufigkeitsverteilungen **betont**; das weist darauf hin, dass eine Einzelbetrachtung der „4“-Häufung als mögliche Ziffernvorliebe sinnvoll ist: Im unterhalb befindlichen **Auswahlfeld** wird die „4“ ausgewählt. Dadurch wird die Binomialverteilung aktiviert und berechnet auf Grundlage der übernommenen Parameter „**Beobachtung**“, „**Anzahl (untersuchter) Daten**“ und „**(prozentuale) Erwartung**“ die Wahrscheinlichkeit, dass eine Einzelziffer so außergewöhnlich häufig oder selten vorkommt. Im Beispielsfall resultiert danach eine Gewissheit von ca. 99,9 %, dass diese Häufung nicht ohne menschlichen Einfluss geschehen konnte. Die rote Feldfarbe zeigt an, dass die erste Auffälligkeitsschwelle der SRP-Vorlage Stochastik überschritten ist. Zusätzlich erscheint der Hinweis auf die Einschränkung der untersuchten Daten, damit die Ziffernanalyse nachvollziehbar bleibt.

Für diese Vertiefung der SRP-Ziffernanalysen auf eine Einzelziffer wird die **Binomialverteilung** in der *vorsichtigsten* Form angewendet. Das bedeutet die Gegenwahrscheinlichkeit einer scheinbar ungewöhnlich häufig (oder selten) beobachteten Ziffer wird aus der Summe aller Einzelwahrscheinlichkeiten errechnet, dass eine Ziffer so häufig oder noch häufiger und spiegelbildlich so selten bzw. noch seltener auftritt. Diese Abb. gibt das Beispiel der Ziffer „1“ wieder: Da sie in den Tageskasseneinnahmen an der Einerstelle 58-mal auftritt,



werden die **Wahrscheinlichkeitssäulen** für „58-mal“, „59-mal“, „60-mal“ bis „483-mal“ sowie die für „38-mal“, „37-mal“, „36-mal“ bis „0-mal“ **zusammengerechnet**. Die Summe beschreibt, wie wahrscheinlich es insgesamt ist, dass sich eine Ziffer so „außergewöhnlich“ zur (Punkt-)Erwartung verhält. Im Um-

kehrschluss drückt die Gegenwahrscheinlichkeit maximal vorsichtig aus, ob es sich um eine Unschlüssigkeit handelt, dass eine Ziffer so entfernt zur Erwartung liegt.

Abschließend ist für den konkreten Beispielsfall mit ca. 99,9 % unzufälliger Häufung der „4“ – *vor der indiziellen Beanstandung als Manipulationsspur* – noch zu klären, ob (nicht-manipulative) Erfassungsfehler oder betriebliche Besonderheiten (s. [Pkt. 4.3.2.2.2.](#)) zu der Ziffernauffälligkeit geführt haben können. Dabei können Preisgestaltungen ausgeschlossen werden, weil selbst die simulierende Annahme eines Eiscafé mit nur einem Produkt zu € 4 an der Einerstelle der Tageskasseneinnahmen (gleichmäßig) alle geraden Ziffern und keine ungeraden erzeugen müsste. Auch eine zu geringe Anzahl an Einzelverkäufen ist bei einem Eiscafé wirtschaftlich nicht denkbar. Sofern sich unter den untersuchten Tagen nachweisbar Einzeltage mit Pauschalpreisen befinden, lassen sich diese zügig und ohne großen Aufwand auf dem Tabellenblatt „Daten“ der SRP-Vorlage Stochastik aus den Analysen entfernen. Im

Ergebnis verbleibt nach Abklärung möglicher nicht-manipulativer Ursachen für die beobachtete Einerstellenauffälligkeit hier ein weitest gehend abgesicherter Manipulationsverdacht auf frei erfundene Tageskasseneinnahmen.

Die Beweiskraft von mathematisch eindeutigen (signifikanten) Einerstellenauffälligkeiten wird wegen der vergleichbaren Voraussetzungen aller manipulationsorientierten Musteranalysen in einem gemeinsamen Kapitel *unter Pkt. 4.3.5.* behandelt. Die derzeitige Beschränkung auf den Status eines „Zusatzindizes“⁵⁶⁷ lässt sich zumindest tlw. durch fehlerhafte Methodenbeschreibungen⁵⁶⁸ und resultierende Zweifel an der korrekten Anwendung⁵⁶⁹ erklären.

4.3.2.3. Zweitziffernanalyse (Benford-Gesetz)

Bei der Zweitziffernanalyse wird die *zweite schwebende*⁵⁷⁰ *Zahlenposition* in den zu untersuchenden Daten auf Übereinstimmung mit der Verteilung gemäß **Benford-Gesetz** geprüft.

1. 2 7 6 , 2 0 *Tageskasseneinnahme*
 **Blickrichtung**

⁵⁶⁷ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit (s. FG Münster v. 14.8.2003 – 8 V 2651/03 E, U, EFG 2004, 9 m. Anm. Trossen; FG Münster v. 10.11.2003 – 6 V 4562/03 E, U, EFG 2004, 236 [237]; FG Münster v. 7.12.2005 – 1 K 6384/03 E, EFG 2006, 652 [653] m. Anm. Zimmermann [Fahrtenbuch]; FG Rh.-Pf. v. 24.8.2011 – 2 K 1277/10, EFG 2012, 10 [11] m. Anm. Kühnen; Sosna, St-Bp. 2000, 41 [Teil I], 68 [Teil II]; Meyer, DStR 2005, 2114; Watrin/Stuffert, DB 2006, 1748; Huber, StBp. 2009, 121 [124 ff.]“.

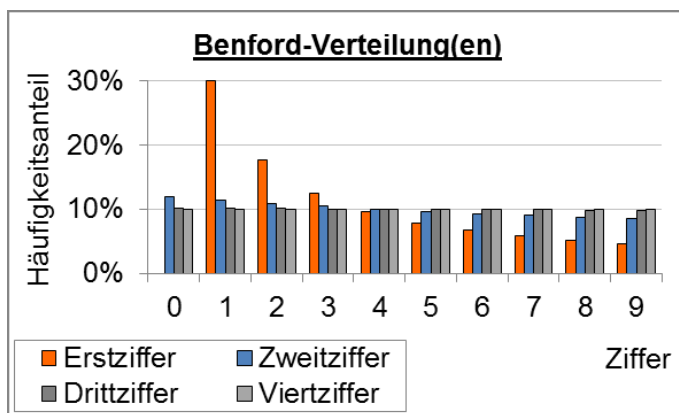
⁵⁶⁸ Vgl. z.B.: Bisle, „Chi-Quadrat-Test‘ und Zeitreihenvergleich: Keine Schätzung bei ordnungsgemäßer Buchführung“, PStR 1/2012, 15: „Der ‚Chi-Quadrat-Test‘ analysiert die Verteilung bestimmter Ziffern und basiert auf der Erkenntnis, dass jeder Mensch Lieblingsziffern hat und bei frei erfundenen Zahlen (z.B. nachgetragenen Tageseinnahmen oder -ausgaben) häufiger verwendet als andere.“ Tatsächlich ist der χ^2 -Test (Chi-Quadrat-Test) ein allgemeiner Wahrscheinlichkeitstest, der mehrklassige Verteilungen unter Berücksichtigung der Erkenntnisse der Wahrscheinlichkeitslehre auf Übereinstimmung prüft. Eine (zwingende) Verbindung zur Gleichverteilung von endständigen Ziffern in Wirtschaftsdaten besteht nicht. Vielmehr vermischt die Aussage die zwei Schritte „Erwartung“ und „Übereinstimmungsprüfung“ i.R. der musterbasierten Analyse und sorgt somit für Verständnisprobleme und Fehleinschätzungen; so auch: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 194: „[...] irritiert, dass die Verprobungsmethode ‚Ziffernanalyse‘ regelmäßig als ‚Chi-Quadrat-Test‘ geführt wird.“

⁵⁶⁹ Kohlmann, Steuerstrafrecht, 72. Lieferung 9.2021, § 397 AO, Rz 12.3: „Die praktische Verwertbarkeit der statistischen Verfahren ist durchaus umstritten, und zwar bereits als Grundlage für steuerliche Schätzungen, aber erst recht für strafrechtliche Verurteilungen. Die Finanzgerichte lehnen bislang den Chi-Quadrat-Test als eine sichere Methode, die Ordnungsgemäßheit der Buchführung zu verwerfen, ab. Allerdings wird weniger die grundsätzliche Eignung der Analysen angezweifelt als vielmehr hinterfragt, unter welchen Voraussetzungen und mit welcher Anzahl von Ausgangswerten die Methode aussagekräftige Ergebnisse liefert.“

⁵⁷⁰ In einem Datenbestand, der über einen Dezimalsprung hinausgeht wie bspw. „€ 800 bis € 1.500“, hat die Zweitziffer unterschiedliche Wertbedeutungen: bei „€ 800“ die Zehnerstelle und bei „€ 1.500“ die Hunderterposition.

4.3.2.3.1. Die (Punkt-)Erwartung „Benford-Gesetz“

Empirische⁵⁷¹ **Grundlage** für das Benford-Gesetz sind die übereinstimmenden Beobach-



tungen des Astronoms und Mathematikers Newcomb (1835-1909) sowie des Elektroingenieurs und Physikers Benford (1883-1948)⁵⁷². Danach zeigen die führenden Positionen (besonders) in wirtschaftsnahen Daten diese charakteristischen, von der Gleichverteilung abweichenden Ziffernhäufigkeiten⁵⁷³ in der Abb.: Für die **Erstziffer**

ergibt sich – neben dem Fehlen der „0“ – eine starke Betonung der kleinen Ziffern auf zunehmende Kosten der höheren. Die **Zweitziffer** nähert sich bereits sehr stark der Gleichverteilung an.

Mathematisch-theoretisch wird diese regelmäßige Erfahrung häufig mit der Dynamik der Verzinsung erklärt: Periodisch verzinstes Vermögen verweilt in jedem Wertebereich längere Zeit mit einer führenden „1“ als mit einer „2“ etc. – bevor die jeweils nächsthöhere Ziffer an der Frontposition erreicht wird. Dadurch vergrößert sich die Wahrscheinlichkeit, Werte mit einer „1“ als Erstziffer zufällig herauszugreifen, gegenüber denen mit höheren Führungsziffern. Eine allgemeinere Erklärung findet sich im Zusammenhang zwischen Benford-Gesetz und Logarithmischer Normalverteilung⁵⁷⁴: Die Benford-Verteilungen beschreiben innerhalb der Logarithmischen Normalverteilungen, welche in Wirtschaftsdaten überwiegend beobachtet werden (s. [Pkt. 4.3.4.1.](#)), die Ziffernmuster der Zahlenpositionen unter der Blickrichtung „von vorne“. Eine solche Verbindung zwischen Ziffernhäufigkeiten und (klassifizierter) Werteverteilung bildet auch die Basis für Ansätze der Wissenschaft, mit einer modifizier-

⁵⁷¹ Empirische Erkenntnisse stammen aus Studien, die den Untersuchungsgegenstand unter nachvollziehbaren konstanten Bedingungen beobachten, und sollen theoretische Annahmen entweder bestätigen oder verwerfen. Gesicherte empirische Erkenntnisse genießen einen höheren Stellenwert in der Wissenschaft als widerstreitende theoretische Annahmen, weil sie die Realität beschreiben.

⁵⁷² „Newcomb-Benford-Law (NBL)“ bzw. „Benford-Gesetz“, veröffentlicht in: Benford, „The Law of Anomalous Numbers“, Proceedings of the American Philosophical Society 1938, 551.

⁵⁷³ Das Benford-Gesetz wird durch die Formel „ $P_{(D)} = \log_{10}(1 + \frac{1}{D})$ “ für die Ziffer „D“ beschrieben.

⁵⁷⁴ Petersen, „Benford’s Law und die Betriebsprüfung – Eine Erwiderung auf Wähnerts ‚Interaktive Außenprüfung – Möglichkeiten der gemeinsamen Sachverhaltsaufklärung im digitalen Zeitalter‘ (Stbg 2014 S. 20 ff.) aus wissenschaftshistorischer Perspektive“, Stbg 12/2015, 506, 510: „Ist die Kongruenz der diskreten Benford-Verteilung mit der stetigen logarithmischen Normalverteilung wohl in der Tat existent [...]“ mit Verweis auf: Scott/Fasli, „Benford’s Law: An Empirical Investigation and a Novel Explanation“, CSM Technical Report 329, Essex 2001, 15.

ten Benford-Verteilung die Prüfung von Wirtschaftsdaten auf menschliche Eingriffe zu verbessern.⁵⁷⁵

Auf die Untersuchung der Erstziffern wurde in der SRP von Beginn an verzichtet.⁵⁷⁶ Diese wären zwar von – manipulationspsychologisch typischen (s. [Pkt. 4.3.4.](#)) – stärkeren Eingriffen in höhere Tageskasseneinnahmen am ehesten betroffen, das Risiko von Fehldeutungen einer Ziffernanalyse ist aber am größten: Da sich die Tageserlöse eines einzelnen Unternehmens, in welchem frei erfundene Besteuerungszahlen erfahrungsgemäß am ehesten vorkommen können, i.d.R. in einem begrenzten Zahlenbereich (auch) unterhalb € 1.000 bewegen, kann die Konzentration zur Betonung bestimmter Einzelziffern führen; so bewirkt ein üblicher Bereich an Tageskasseneinnahmen von „€ 500 bis € 800“ häufig die Erstziffern „5“, „6“ und „7“, und die Erstziffernkurve (s. *vorherige Abb.*) verschiebt sich demnach auf der Ziffernachse. Dieser störende Einfluss auf die Erstziffernanalyse tritt bei der Zweitziffer nur sehr begrenzt auf. Dennoch wird im Fall einer Zweitziffernauffälligkeit i.R. der SRP-Ziffernanalytik die mögliche Verursachung durch einen begrenzten Zahlenbereich gezielt abgeklärt (s. *das Beispiel unter* [Pkt. 4.3.2.3.3.](#)).

Ein größenabhängiger Ausschluss von Daten wie i.R. der Einerstellenuntersuchung mit Einschränkung auf Werte unter € 100 (s. [Pkt. 4.3.2.2.1.](#)) geschieht bei der Zweitziffernanalyse in der SRP-Vorlage Stochastik nicht, weil die besonderen Ziffernmuster der führenden Zahlenpositionen durch die Veränderungen an den Dezimalsprüngen verstärkt werden; somit kann es dazu kommen, dass auf dem Tabellenblatt „digZA“ unterschiedlich viele Datensätze in den beiden Ziffernanalysen der Einerstelle und Zweitziffer ausgewertet werden. Sollten die Daten durch Eingabe eines Schichtungsstartwerts auf dem Tabellenblatt „Schichtung“ wertabhängig eingeschränkt worden sein, weist die Rotfärbung der erwarteten Gesamtanzahl in allen Zweitzifferntabellen darauf hin.

Die Übereinstimmungsprüfung des beobachteten Häufigkeitsmusters aller Ziffern an der zweiten Position der untersuchten Daten mit der Benford-Verteilung erfolgt durch den χ^2 -Test (Chi-Quadrat-Test; zur Gegenüberstellung der Eignung von Chi-Quadrat- und Kolmogorov-Smirnov-Test s. den Abschnitt „Chi-Quadrat-Test“ unter [Pkt. 4.3.1.3.](#)); dadurch beinhaltet das Übereinstimmungsergebnis bereits die Datenanzahl (s. [Pkt. 4.3.1.3.](#)). Die übliche Empfeh-

⁵⁷⁵ Vgl. z.B. die digitale Spurensuche mit modifizierter Benford-Verteilung am Fraunhofer-Institut für Sichere Informationstechnologie auf der Internetseite der Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V.: „Digitale Spurensuche – Wirtschaftsthemen Mai 2011“, nachzulesen unter: <https://www.fraunhofer.de/de/presse/presseinformationen/2011/mai/it-forensik-gegen-wirtschaftskriminalitaet.html>.

⁵⁷⁶ Insofern verwundern tlw. umfangreiche Ausführungen zur Erstziffernverteilung gemäß Benford-Gesetz i.R. der kritisierenden Literatur an der SRP wie z.B. in: Petersen, „Benford’s Law und die Betriebsprüfung – Eine Erwiderung auf Wähnerts ‚Interaktive Außenprüfung – Möglichkeiten der gemeinsamen Sachverhaltsaufklärung im digitalen Zeitalter‘ (Stbg 2014 S. 20 ff.) aus wissenschaftshistorischer Perspektive“, Stbg 12/2015, 506, 507 (Kapitel „Genesis von Benford’s Law“).

lung zur Mindestdatenmenge von 59 Datensätzen⁵⁷⁷ bei der Erwartung „Benford-Verteilung für die Zweitziffer“ wird von der SRP-Vorlage Stochastik automatisch beachtet. Formatseitig werden die zwei Auffälligkeitsschwellen „99 %“ und „99,95 % Nichtübereinstimmung“ hervorgehoben (s. [Pkt. 4.3.2.1.](#)).

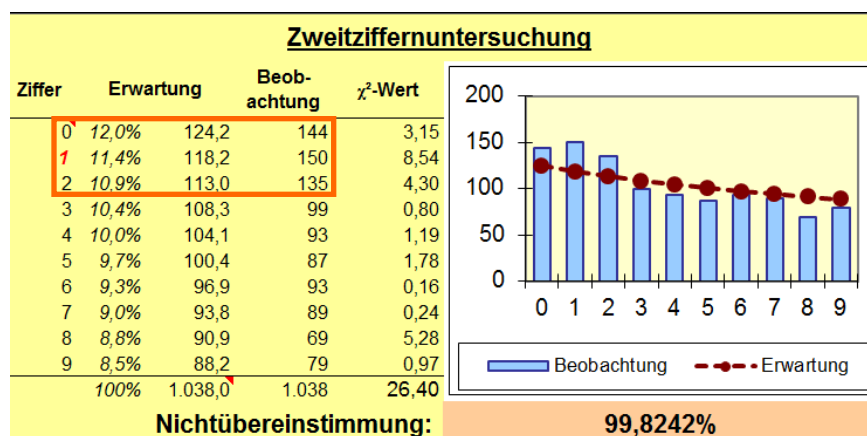
4.3.2.3.2. Nicht-manipulative Ursachen für Zweitziffernauffälligkeiten

Einzelfallbezogene Besonderheiten als Ursache für mathematisch eindeutige (signifikante) Ziffernauffälligkeiten der Zweitposition haben – im Gegensatz zur Einerstelle – i.d.R. einen Bezug zu den Größenverhältnissen der untersuchten Zahlen. Sofern besonders Tageskasseneinnahmen ihre Häufungszone entweder an einem **Dezimalsprung** oder in einem **begrenzten Bereich oberhalb von € 1.000** haben (s. [Pkt. 4.3.2.3.3.](#)), werden dadurch bestimmte Zweitziffern betont: So bringt ein unternehmenstypischer Schwerpunkt von Tageskasseneinnahmen im Bereich „€ 2.200 bis € 2.600“ überwiegend die Zweitziffern „2“, „3“, „4“ und „5“ hervor. Derartige Konstellationen sind eher selten, weil bei einem höheren Einnahmenniveau meistens auch die Streuung größer ist. Dennoch wird diese mögliche praxisnahe Besonderheit i.R. der SRP-Ziffernanalytik gezielt abgeklärt, wobei die Funktionalitäten der SRP-Vorlage Stochastik helfen.

4.3.2.3.3. Vorlagenunterstützte Abklärung von Zweitziffernauffälligkeiten

In der (manipulationsorientierten) **SRP-Musteranalytik** ist eine selbständige (vorlagenunterstützte) Nachverfolgung von Zweitziffernauffälligkeiten vorgesehen, um betriebliche Besonderheiten des Einzelfalls als Verursacher erkennen bzw. (weitest gehend) ausschließen zu können.

Die exemplarisch abgebildete Häufigkeitsverteilung zur Zweitziffer in den Tageskas-



seneinnahmen eines gastronomischen Betriebs stimmt nicht mit der (Punkt-)Erwartung gemäß Benford-Gesetz überein. Der χ^2 -Test (Chi-Quadrat-Test) bestätigt diesen visuellen Eindruck mit der Nicht-

übereinstimmungswahrscheinlichkeit von 99,8 % oberhalb der ersten Auffälligkeitsschwelle

⁵⁷⁷ Die Mindestdatenmenge von 59 Datensätzen ergibt sich durch Anwendung der Empfehlung „≥ 5 erwartete Elemente pro Klasse“ auf den Prozentanteil „8,5 %“ für die nach dem Benford-Gesetz am seltensten vorkommende Ziffer „9“.

der SRP-Vorlage Stochastik (s. *das Stichwort „Chi-Quadrat-Test“ unter [Pkt. 4.3.2.1.](#)*). Im offensichtlichen Widerspruch zum Einsatzzweck der Ziffernanalyse, punktuelle Vorlieben oder Abneigungen bzw. Chaos aufzudecken, ist allerdings eine recht gleichmäßige **Überbetonung der kleinen Ziffern „0“, „1“ und „2“** auf Kosten fast aller größeren abzulesen. Dabei weist die Rotfärbung der Ziffer „1“ darauf hin, dass diese aus der Einzelperspektive die Erwartung ebenfalls mathematisch eindeutig (signifikant) verfehlt (s. *das Beispiel im Bereich „Binomialverteilung“ unterhalb zu diesem Pkt.*). **Bei zwar eindeutigen aber in dieser Form gleichmäßigen Zweitziiffernauffälligkeiten muss die mögliche nicht-manipulative Ursache „Bündelung der untersuchten Zahlen an einem Dezimalsprung“ abgeklärt werden.**

Dazu geben die statistischen Kenngrößen auf dem Tabellenblatt „Daten“ in der SRP-Vorlage Stochastik mit dem „inneren“ Streuungsbereich einen ersten groben Eindruck: Nach dem **Mittelwert von € 852,18** und der (vorzeichenneutral) typischen Abweichung jedes einzelnen Datensatzes zum Mittelwert (**Standardabweichung**) von **€ 371,93** reicht der Normalbereich der Tageskasseneinnahmen „Mittelwert + Standardabweichung = € 1.224,11“ über den Dezimalsprung „Hunderter zu Tausender“ hinaus. Wie stark gehäuft die Daten zwischen € 1.000 und € 1.224,11 vorkommen, lässt sich aus den statistischen Kenngrößen nicht (direkt) ablesen, weil die Häufigkeitsverteilung „Logarithmische Normalverteilung“ asymmetrisch ist (s. [Pkt. 4.3.4.1.](#)).

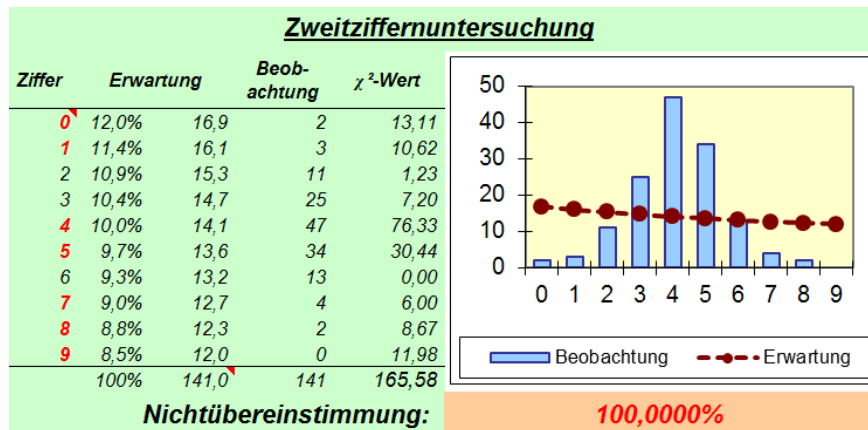
statistische Kenngrößen	
absolute Kenngrößen (lineare Schichtung)	
Anzahl:	1.038
Anzahl ab Schichtungsminimum:	
Summe:	884.561,93
Mittelwert:	852,18
Standardabweichung:	371,93

Da im Beispiel die betriebliche Besonderheit „Häufung der Tageskasseneinnahmen an einem Dezimalsprung“ als nicht-manipulative Ursache nicht ausgeschlossen werden kann, muss die Zweitziiffernanalyse weiter vertieft werden. Dazu bieten die Klassifizierungen der Tabellenblätter „Schichtung“ und „Schichtungsvertiefung“ in der SRP-Vorlage Stochastik mit ihren Unterteilungen der Daten in gleichmäßige Werteklassen schnelle und zielführende Funktionen an. Die weitere Beschreibung dieser Untersuchung eines auffälligen Zweitziiffernergebnisses auf eine mögliche nicht-manipulative Ursache erfolgt im betreffenden Abschnitt „Abklärung von Zweitziiffernauffälligkeiten“ unter [Pkt. 4.3.3.2.](#)

Als weiteres Beispiel für die vorlagenunterstützte Nachverfolgung eines mathematisch eindeutigen (signifikanten) Zweitziiffernmusters zeigen die Tageskasseneinnahmen eines Fischgeschäfts insgesamt ein zu häufiges Vorkommen der Ziffern „4“ und „5“. Eine zufällige Abweichung von der Benford-Verteilung kann bei einer Nichtübereinstimmungswahrscheinlichkeit von mehr als 99,99 % gemäß χ^2 -Test (Chi-Quadrat-Test) ausgeschlossen werden. Die Suche nach auslösenden Teilen in den Daten mit Hilfe der Auswahl-Drehfelder für Jahre und Wochentage auf dem Tabellenblatt „digZA“ (s. *den Bereich „zweistufige Eingrenzungs-*

möglichkeit“ unter [Pkt. 4.3.2.1.](#)) ergibt für mehrere Wochentage signifikant vom Benford-Gesetz abweichende Häufigkeitsmuster, die aber dennoch gleichmäßig aussehen.

Als wesentlicher Ursprung der insgesamt zu häufigen „4“ und „5“ fällt dabei umgehend



der Donnerstag mit der abgebildeten Zweitziffernverteilung auf. Trotz der mathematisch eindeutigen (signifikanten)

Nichtübereinstimmung mit der Erwartung gemäß Benford-Gesetz und diverser rot hervor-

gehobener (s. das Beispiel im Bereich „Binomialverteilung“ unterhalb zu diesem Pkt.) Ziffern deutet die glockenförmig gleichmäßige Häufigkeitskurve nicht auf konkrete Ziffernvorlieben oder -abneigungen bzw. -chaos hin. Vielmehr ist dieses Verteilungsmuster von Zweitziffern in Tageskasseneinnahmen typisch für Werte, die in einem **begrenzten Zahlenbereich oberhalb von € 1.000** liegen: Wenn die täglich aufsummierten Bareinnahmen eines Betriebs nicht allzu weit auf der Zahlenskala streuen, wirkt die Verzehnfachung⁵⁷⁸ der Trefferwahrscheinlichkeit für Zweitziffern oberhalb des Dezimalsprungs „Hunderter zu Tausender“ nicht auf alle Ziffern gleichermaßen. Für das abgebildete Beispiel würde eine Konzentration der Tageskasseneinnahmen im Bereich „€ 1.300 bis € 1.599,99“ als nicht-manipulative Erklärung in Betracht kommen. Auch die Abklärung dieser möglichen Ursache für die beobachtete Zweitziffernauffälligkeit gelingt problemlos mit Hilfe der Klassifizierungen der Tabellenblätter „Schichtung“ und „Schichtungsvertiefung“ in der SRP-Vorlage Stochastik und wird unter [Pkt. 4.3.3.2.](#) dargestellt.

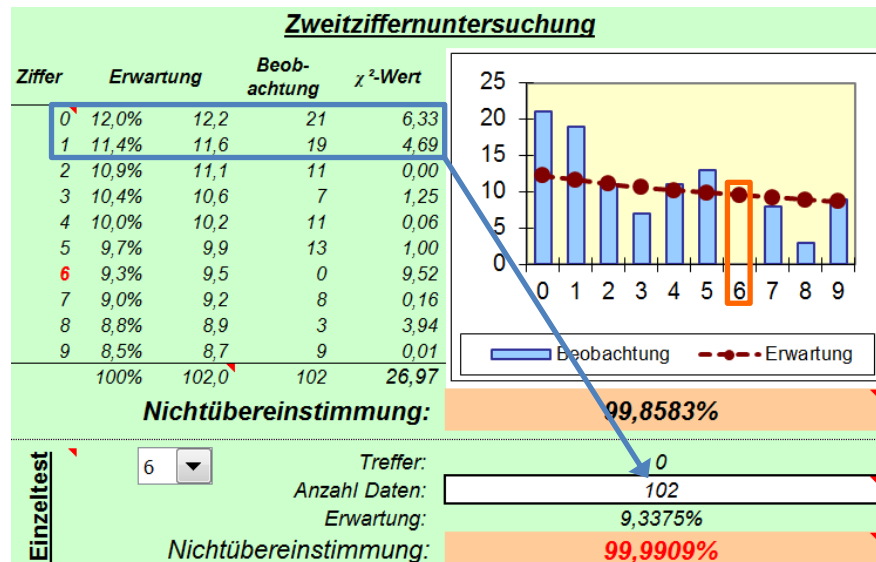
Ein Perspektivwechsel auf eine Einzelziffer, die an der Zweitposition besonders häufig oder selten beobachtet wird, ist auf dem Tabellenblatt „digZA“ der SRP-Vorlage Stochastik ebenfalls vorgesehen. Auffällige Ziffern werden in der Tabelle mit den Häufigkeitsverteilungen durch rote Schrift hervorgehoben. Sie lassen sich im Bereich „Einzeltest“ unterhalb über ein Drehfeld auswählen. Dadurch werden die Informationen „Treffer“ und „(prozentuale) Trefferwahrscheinlichkeit“⁵⁷⁹ zu dieser Einzelziffer sowie „Anzahl (untersuchter) Daten“ als erfor-

⁵⁷⁸ Während die Spanne für die Zweitziffer unterhalb des Dezimalsprungs „Hunderter zu Tausender“ mit z.B. „€ 950 bis € 959,99“ ein Intervall von € 10 umfasst, wächst sie oberhalb abrupt auf € 100 an: „€ 1.500 bis € 1.599,99“. Die Wahrscheinlichkeit, dass eine Tageskasseneinnahme in die jeweiligen Bereiche zufällig hineinfällt, verändert sich entsprechend.

⁵⁷⁹ Im Gegensatz zur Einerstellenuntersuchung sind bei der Zweitziffernanalyse die Trefferwahrscheinlichkeiten der Einzelziffern nicht gleich hoch, sondern müssen mit der Benford-Verteilung ermittelt werden. So ist z.B.

derliche Größen zur Berechnung der Nichtübereinstimmungswahrscheinlichkeit mit der einschlägigen **Binomialverteilung** übernommen.

Gegenüber der Einerstellenuntersuchung (s. unter [Pkt. 4.3.2.2.3.](#)) ist bei der Einzelbetrachtung einer Ziffer i.R. der Zweitziiffernanalyse eine Besonderheit zu beachten: Wie aus der Tabelle ersichtlich verschlimmern – neben dem auffälligen Fehlen der Ziffer „6“ – auch die zu häufigen Ziffern „0“ und „1“ das Ergebnis des χ^2 -Tests (Chi-Quadrat-Test) erheblich. Dazu kann mit Hilfe der Wertklassifizierungen der SRP-Vorlage Stochastik zügig herausge-



funden werden, dass sich diese beispielhaften Tageskasseneinnahmen eines Eiscafés am Dezimalsprung „Hunderter zu Tausender“ ballen (s. [Pkt. 4.3.3.2.](#)) und deshalb vermehrt Nullen und Einsen an der Zweitposition aufweisen. Dieser praxisrelevanteste Störeinfluss für die Zweitziiffernanalyse wirkt sich auch auf die Einzelziiffernbetrachtung aus, weil die Datensätze, welche durch den Dezimalsprung an „0“ und „1“ gebunden sind, nicht zu den erwarteten 9,5-mal „6“ beitragen können. **Ausschließlich im Hinblick auf diese Konstellation** besteht in der SRP-Vorlage Stochastik die Möglichkeit, die Datenanzahl für die Binomialverteilung anzupassen⁵⁸⁰, worauf der Kommentar zum Eingabefeld hinweist. Auf diese Weise kann die Wahrscheinlichkeit für das auffällige Fehlen der „6“ trotz Überlagerung mit einem Dezimalsprung *abgeschätzt* werden.

Die Beweiskraft von mathematisch eindeutigen (signifikanten) Zweitziiffernauffälligkeiten wird wegen der vergleichbaren Voraussetzungen aller manipulationsorientierten Musteranalysen in einem gemeinsamen Kapitel unter [Pkt. 4.3.5.](#) behandelt. Die derzeitige Beschränkung auf den Status eines „Zusatzindizes“⁵⁸¹ lässt sich zumindest tlw. durch fehlerhafte Me-

die Wahrscheinlichkeit für eine „0“ an der Zweitposition ca. 12 %. Somit ergibt sich als Summenwahrscheinlichkeit für alle anderen Ziffern der Gegenwert zu 100 %, also ca. 88 %.

⁵⁸⁰ Im Beispielsfall weist die Tabelle mit den Häufigkeitsverteilungen zur Zweitziiffer aus, dass die Ziffern „0“ und „1“ 16,2-mal [(21–12,2) + (19–11,6)] zu oft vorkommen. Als Annäherung kann diese Häufigkeit dem Dezimalsprung angelastet werden, sodass die verfügbare Datenanzahl für eine „6“ auf 102 – 16,2 = abgerundet 85 korrigiert wird.

⁵⁸¹ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Eine Ziiffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit (s. FG Münster v. 14.8.2003 – 8 V

thodenbeschreibungen⁵⁸² und resultierende Zweifel an der korrekten Anwendung⁵⁸³ erklären.⁵⁸⁴

4.3.3. Klassifizierung (Schichtung) der Werte

Als „zentrale explorative [...] Data Mining-Ansätze, die prozessgestützt auffällige Muster und Beziehungen in den großen Datenvolumina ausfindig machen“⁵⁸⁵, unterteilen Klassifizierungen (Schichtungen⁵⁸⁶) stetige⁵⁸⁷ Daten zunächst in Werteklassen und zählen sie danach aus, um durch Verbildlichung die Häufigkeitsmuster herauszuarbeiten, womit sie der Forderung des Bundesrechnungshofs entsprechen: „[D]as Bundesministerium der Finanzen (BMF) [muss] auf den bundesweit einheitlichen Einsatz IT-gestützter quantitativer Prüfungsmetho-

2651/03 E, U, EFG 2004, 9 m. Anm. *Trossen*; FG Münster v. 10.11.2003 – 6 V 4562/03 E, U, EFG 2004, 236 [237]; FG Münster v. 7.12.2005 – 1 K 6384/03 E, EFG 2006, 652 [653] m. Anm. *Zimmermann* [Fahrtbuch]; FG Rh.-Pf. v. 24.8.2011 – 2 K 1277/10, EFG 2012, 10 [11] m. Anm. *Kühnen*; *Sosna*, St-Bp. 2000, 41 [Teil I], 68 [Teil II]; *Meyer*, DStR 2005, 2114; *Watrin/Stuffert*, DB 2006, 1748; *Huber*, StBp. 2009, 121 [124 ff.]“.

⁵⁸² Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 136: „In der derzeitigen Praxis der Finanzämter hat sich der Chi-Quadrat-Anpassungstest als Überprüfungsverfahren der Verteilungsannahmen bei der Ziffernanalyse etabliert. Dies hat u.a. dazu geführt, dass die Ziffernanalyse anhand des Benfordschen Gesetzes, insbesondere in der Praktiker-Literatur, mitunter fälschlicherweise grobschlächtig als ‚Chi-Quadrat-Test‘ bezeichnet und somit als eigenständige Methode, und nicht als Analyseinstrument, aufgefasst wird. Hier zeigt sich ein häufig beobachtetes unzureichendes Verständnis der Methodik und der statistischen Testverfahren.“

⁵⁸³ Kohlmann, Steuerstrafrecht, 72. Lieferung 9.2021, § 397 AO, Rz 12.3: „Die praktische Verwertbarkeit der statistischen Verfahren ist durchaus umstritten, und zwar bereits als Grundlage für steuerliche Schätzungen, aber erst recht für strafrechtliche Verurteilungen. Die Finanzgerichte lehnen bislang den Chi-Quadrat-Test als eine sichere Methode, die Ordnungsgemäßheit der Buchführung zu verwerfen, ab. Allerdings wird weniger die grundsätzliche Eignung der Analysen angezweifelt als vielmehr hinterfragt, unter welchen Voraussetzungen und mit welcher Anzahl von Ausgangswerten die Methode aussagekräftige Ergebnisse liefert.“

⁵⁸⁴ Zwar werden in der (wissenschaftlichen) Literatur auch Bedenken bezüglich der Allgemeingültigkeit der Benford-Verteilungen geäußert (zuletzt z.B.: *Pitz/Klapdor/Gardian/Haselau*, „Das Benfordsche Gesetz als Werkzeug zur Aufdeckung von Umsatzmanipulationen in der Gastronomie?“, *StuW* 3/2021, 232, 245: „Auch wenn umfangreichere Untersuchungen noch ausstehen, so zeigen bei der Ziffernanalyse nach Benford bei dieser Fallstudie Betriebe deutliche Auffälligkeiten, bei denen eine Manipulation (eher) vermutet wird. Auch wenn eine Tendenz zu erkennen ist, ist eine statistisch signifikante Korrelation zwischen ‚Auffälligkeit‘ und ‚Zuverlässigkeit‘ der Datensätze (noch) nicht zwingend nachzuweisen.“), wobei aber offenbar beweismäßige Ansprüche an Verprobungen zu Grunde gelegt werden. Nach Ansicht des Autors realistischer: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 181 f.: „Sowohl Zweit-, Dritt- und Viertziffer sowie die Einerstelle als erste Stelle vor dem Komma zeigten insgesamt eine sehr gute Konformität mit der Benford-Verteilung. Diese Beobachtungen lassen darauf schließen, dass die Annahme der Benford-Verteilung als Verteilungsideal ordnungsmäßiger Daten als durchaus plausibel erscheint.“

⁵⁸⁵ *Velte/Drews*, „Herausforderungen in der Wirtschaftsprüfung durch den Einsatz von Big-Data-Technologien – Welche Impulse muss die Prüfungsforschung geben?“, DB 43/2018, 2581, 2582.

⁵⁸⁶ Für die Einteilung in Größenklassen werden unterschiedliche Begriffe wie bspw. „Schichtung“ im Datenanalyseprogramm IDEA® oder „Gruppierung“ in MS Excel® verwendet.

⁵⁸⁷ Unter diskreten Verteilungen werden in der Mathematik solche verstanden, die keine Zwischenwerte einnehmen können. So kann es bei Münzwürfen aus Sicht einer bestimmten Münzseite nur zu ganzzahligen Erfolgs- bzw. Misserfolgshäufigkeiten kommen, nicht aber bspw. zu 3,437 Treffern. Der gegenteilige Verteilungstyp mit kontinuierlichen Ergebnismöglichkeiten wird „stetig“ genannt.

den in der steuerlichen Betriebsprüfung hinwirken[, ... denn t]raditionelle Methoden [...], beispielsweise Beleg- und Akteneinsicht, verlieren an Effektivität und Effizienz. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.⁵⁸⁸ Dazu ist die Festlegung von i.d.R. gleich großen⁵⁸⁹ Klassen zielführend. Allgemein haben die hierfür i.R. der SRP vorgesehenen Tabellenblätter „Schichtung“ und „Schichtungsvertiefung“ die Funktion **Orientierung in größeren Datenmengen** (s. die Abschnitte „gleichmäßige Muster der Häufigkeitsverteilungen“ unter [Pkt. 4.3.](#) und „systematisch-mehrperspektivische Musteranalytik“ unter [Pkt. 5.1.1.](#)) und befinden sich deshalb im ersten Teil der SRP-Vorlage Stochastik mit der „Kurzanleitung 1“; i.R. der Verprobung „Vollständigkeit von (Bar-)Einnahmen“ dienen die Klassifizierungen zur individuell verstärkenden **Absicherung manipulationsorientierter Musteranalysen** gemäß SRP-Mustersystematik (s. den Abschnitt „Prüfungsschema zur Vermeidung von Fehldeutungen“ unter [Pkt. 4.3.](#)):

- © Das **Suchen nach prüfungsrelevanten Vorgängen in unprüfbar großen Datenmengen** anhand von Häufigkeitsmustern (s.a. das Beispiel zur Konkretisierung einer Zeitreihenauffälligkeit im Abschnitt „Prüfungsnetz“ unter [Pkt. 2.3.](#)) lehnt sich an eine zentrale Erkenntnis der Datenanalytik an: „Komplexität hat sehr viel mit Vernetzung zu tun, ja kommt erst durch Vernetzung zustande. Komplexe Vorgänge verlangen daher zu ihrem Verständnis ein Denken in Zusammenhängen, das sich an der Struktur organisierter Systeme und ihrer speziellen Dynamik orientiert. [...] Damit treffen wir auf einen der Kardinalfehler in der Beurteilung, was für die Erfassung von Komplexität wichtig ist. Denn hier geht es um das ‚Gesicht‘ der Wirklichkeit, um das Erkennen von Mustern“⁵⁹⁰.

⁵⁸⁸ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>. Mit dem BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1, 4, 9 ff. u. 12 f. schließt sich die Finanzverwaltung bzgl. der Wichtigkeit quantitativer Prüfungsmethoden und speziell der musterbasierten Häufigkeitsanalysen an: „Bei einer steuerlichen Außenprüfung können mathematisch-statistische (quantitative) Prüfungsmethoden eingesetzt werden, um die Besteuerungsgrundlagen des Steuerpflichtigen zu verproben und auf Plausibilität zu überprüfen. Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich. [...] Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgabenordnung dazu führen, dass eine Schätzung durchgeführt werden muss. [...] 2. Ziffernanalyse [...] 3. Struktur- und Verteilungsanalyse“.

⁵⁸⁹ Sofern ein asymmetrisch linkssteil-rechtsschiefes Häufigkeitsmuster wie die Logarithmische Normalverteilung erwartet wird, ist die Wahl von Größenklassen mit logarithmisch gleichen Abständen (z.B. bei „ $10^2 - 10^3 - 10^4 - 10^5$ etc.“ entsprechend „100 – 1.000 – 10.000 – 100.000 usw.“) oft zielführender, um die Gleichmäßigkeit der beobachteten Verteilung visuell zu beurteilen.

⁵⁹⁰ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 16 u. 21.

Da betriebsuntypische Sachverhalte häufig aus ansonsten gleichmäßigen Häufigkeitsmustern ausbrechen, bietet sich auf diesem Weg mit Hilfe digitaler Lösungen eine effektive Möglichkeit an, diejenigen Vorgänge effizient zu identifizieren, die ein besonders hohes Fehler- oder Manipulationsrisiko in sich tragen. Dazu verwendet die SRP-Vorlage Stochastik auf dem Tabellenblatt „Schichtung“ einen variablen und komfortabel bedienbaren *externen* Vergleich mit dem allgemeinen Häufigkeitsmuster in Wirtschaftsdaten „Logarithmische Normalverteilung“ (s. [Pkt. 4.3.4.1.](#)). Zusätzlich werden auf dem Tabellenblatt „Schichtungsverteufung“ bis zu drei Jahre und Wochentage per Auswahl über Drehfelder bei vergleichbarer Variabilität und Bedienbarkeit *intern* gegenübergestellt. Spezielle Funktionalitäten der SRP-Vorlage Stochastik wie bspw. die Umstellung von linearer auf logarithmische Klassifizierung (s. *den Abschnitt „Umstellung von linearer auf logarithmische Klassenbildung“ unter [Pkt. 4.3.3.1.](#)*) unterstützen die Musteranalytik für ein zielgenaues und zügiges Auffinden von betriebsspezifischen Prüffeldern.

- © Für möglichst eindeutige Indizien aus den beiden manipulationsorientierten Musteranalysen im zweiten Teil der SRP-Vorlage Stochastik werden zum einen **Zweitziffernauffälligkeiten** durch (weitest gehende) Abklärung denkbarer nicht-manipulativer Erklärungen mittels Klassifizierung **abgesichert** (s. *die beiden Beispiele im Abschnitt „Abklärung von Zweitziffernauffälligkeiten“ unter [Pkt. 4.3.3.2.](#)*), und es wird zum anderen die **Verteilungsanalyse** (s. [Pkt. 4.3.4.](#)) **vorbereitet**.⁵⁹¹ Auf diese Weise kann bei Zweitziffernanalysen die praxisrelevanteste nicht-manipulative Ursache für Auffälligkeiten bestätigt oder ausgeschlossen werden. I.R. der Untersuchung von Tageskasseneinnahmen auf – manipulationspsychologisch typische (s. [Pkt. 4.3.4.](#)) – stärkere Eingriffe in größere Werte (Verteilungsanalyse) wird bereits im Vorfeld der wichtigste Störfaktor „uneinheitlicher Datenbestand“ abgeklärt. Im Ergebnis erreichen verbleibende Musterauffälligkeiten allgemeinverständlich eine höhere indizielle Gewissheit (s. [Pkt. 4.3.5.](#)).

4.3.3.1. Anwendungshinweise zu SRP-Werteklassifizierungen

Folgende allgemeine Hinweise zur Durchführung von Datenklassifizierungen mit der SRP-Vorlage Stochastik sind wichtig:

- ➔ Alle auf dem Tabellenblatt „Daten“ vorgenommenen **generellen Dateneinschränkungen** (s. *den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)*) wirken auf die Werteklassifizierungen der SRP-Vorlage Stochastik; entsprechende Hinweise erfolgen auf den Auswertungsblättern „Schichtung“ und „Schichtungsverteufung“.

⁵⁹¹ Diese Sorgfalt wird in der wissenschaftlichen Literatur jüngst herausgestellt: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 139.

➔ Einheitlich für alle Werteklassifizierungen der Tabellenblätter „Schichtung“ und „Schichtungsvertiefung“ kann ein **Startwert** festgelegt werden, ab dem die musterbasierten Häufigkeitsanalysen ausgeführt werden. Hilfreich ist diese Funktion, um kleine Werte auszuschließen, die eine sinnvolle Musterabbildung behindern, oder um den Untersuchungsfokus auf einen bestimmten Wertebereich zu legen. Alle empfohlenen Schichtungsparameter (s. den nachfolgenden Auflistungspkt. „Klassenunterteilung“) werden automatisch auf die neuen Verhältnisse umgestellt. Zusätzlich zu den „statistischen Kenngrößen“ auf dem Tabellenblatt „Daten“ wird die „(Daten-)Anzahl ab Schichtungsminimum“ angegeben. Auf dem Tabellenblatt „Schichtungsvertiefung“, wohin der Startwert automatisch übertragen wird, lässt sich die Eingabe einer unteren Begrenzung an der roten Farbe des Felds „maximale Datenanzahl“ nachvollziehen.

➔ Um mittels **Klassenunterteilung** Häufigkeitsmuster gut in den zu untersuchenden Daten erkennen zu können, muss zunächst eine zielführende⁵⁹² Anzahl an Klassen festgelegt werden. Dann kann

diese auf den (relevanten) Wertebereich angewendet werden, womit der einheitliche Klassenumfang berechnet wird. Hierzu greift die SRP auf die prüfungserprobte **Wurzelmethode** zurück: Die Klassenanzahl ergibt sich als **Wurzel der**

statistische Kenngrößen	
<u>absolute Kenngrößen (lineare Schichtung)</u>	
Anzahl:	1.094
Anzahl ab Schichtungsminimum:	
Summe:	513.530,85
Mittelwert:	469,41
Standardabweichung:	260,27
Minimum:	63,30
Hilfsminimum:	
Maximum:	2.369,40
Hilfsmaximum:	
Vorschlag Intervallanzahl:	30
Vorschlag Schichtungsintervall:	80,00
Vorschlag erste Schichtobergrenze:	70,00
Vorschlag zweite Schichtobergrenze:	150,00

Datenanzahl, wird aber zur besseren Übersichtlichkeit auf maximal 30 Intervalle begrenzt; die Auflistung der statistischen Parameter auf dem Tabellenblatt „Daten“ neben den eingefügten Daten dokumentiert die automatische Klassenbildung. Im Beispiel führt die hohe Datenanzahl danach zu ca. 33 Intervallen ($\sqrt{1.094}$), sodass die Beschränkung auf 30 Klassen greift. Diese Klassenanzahl wird (gerundet) auf den Wertebereich vom Minimum bzw. Hilfsminimum (s. den nächsten Auflistungspkt. „Hilfsminimum“ und „Hilfsmaximum“) bis zum Maximum bzw. Hilfsmaximum angewendet, wodurch sich das feste Intervall gemäß Abb. berechnet. Nach jeder Dateneinschränkung auf dem Tabellenblatt „Daten“ (s. die zweite Abb. zu „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)) sowie bei Eingabe eines Schichtungsstartwerts auf dem Tabellenblatt „Schichtung“ (s. den Auflistungspkt. „Startwert“ oberhalb) wird die automatische Klassifi-

⁵⁹² Als Folge einer unvorteilhaften Klassenfestlegung verteilen sich die Daten entweder auf zu wenige Intervalle oder werden zu kleinteilig (unruhig) abgebildet, sodass jeweils das Streumuster schlecht sichtbar ist.

zierung umgehend angepasst. Anstelle der empfohlenen Klassenbildung können für jede Schichtung die ersten Obergrenzen bzw. die Intervallgröße komfortabel abgeändert werden (s. *den übernächsten Auflistungspkt. zur „freien Klassenbildung“*).

- ➔ Mit Blick auf zielführendere Werteklassifizierungen – also bessere Musterergebnisse – versucht die SRP-Vorlage Stochastik extreme Ausreißerwerte ohne engeren Verteilungsbezug zu den anderen untersuchten Daten selbständig zu erkennen und verwendet in dem Fall die Visualisierungsgrenzen „**Hilfsminimum**“ und „**Hilfsmaximum**“;

zusätzlich lassen sich automatisch identifizierte **potentielle Ausreißer**

Datum	alle Werte	Bemerkung	untersuchte Werte	pot. Ausr.	lg
24.12.2002	139,50		139,50	pot. Ausr.	2,145
08.04.2003	169,20		169,20	pot. Ausr.	2,228
24.12.2003	153,40		153,40	pot. Ausr.	2,186

tentielle Ausreißer, deren Vorhandensein durch Rotfärbung der Überschrift angezeigt wird, mittels **Filter der Spalte „pot. Ausr.“** auf dem Tabellenblatt „Daten“ zügig als sinnvolle Prüffelder ausgeben. Zur Ausreißer-Erkennung („Outlier-Mining“) werden die Erfahrungen nach *Grubbs*⁵⁹³ (abgewandelt) herangezogen: Datensätze, die mehr als den 2,2-fachen Interquartilsabstand⁵⁹⁴ vom 3. Quartil⁵⁹⁵ nach oben oder vom 1. Quartil nach unten abweichen, gelten als „Ausreißer“; i.R. der manipulationsorientierten Verteilungsanalyse (s. *Pkt. 4.3.4.*) sind die Grenzen mit dem 4,4-fachen Interquartilsabstand weiter gefasst, weil dort das Muster wichtiger als das Hervorheben von außergewöhnlichen Einzelsachverhalten ist. Stellen die Extremwerte der untersuchten Daten danach offenbare Ausreißer dar, werden sie durch die Hilfswerte als Visualisierungsgrenzen bei der Klassenfestlegung ersetzt; dies wird aus den „statistischen Kenngrößen“ auf dem Tabellenblatt „Daten“ ersichtlich, indem die Werte zu „Minimum“ und „Maximum“ hellgrau unterdrückt sind. Aus den Analysen werden die betreffenden Ausreißer-Datensätze dadurch **nicht** ausgeschlossen. Sollte dies gewünscht sein, müssen sie durch Eintragen eines „x“ in der Kästchenspalte links neben den Datensätzen einzeln abgewählt werden; daraufhin erscheint auf jedem Auswertungsblatt automatisch der Hinweis „untersuchte Daten: Einzeldatenabwahl“.

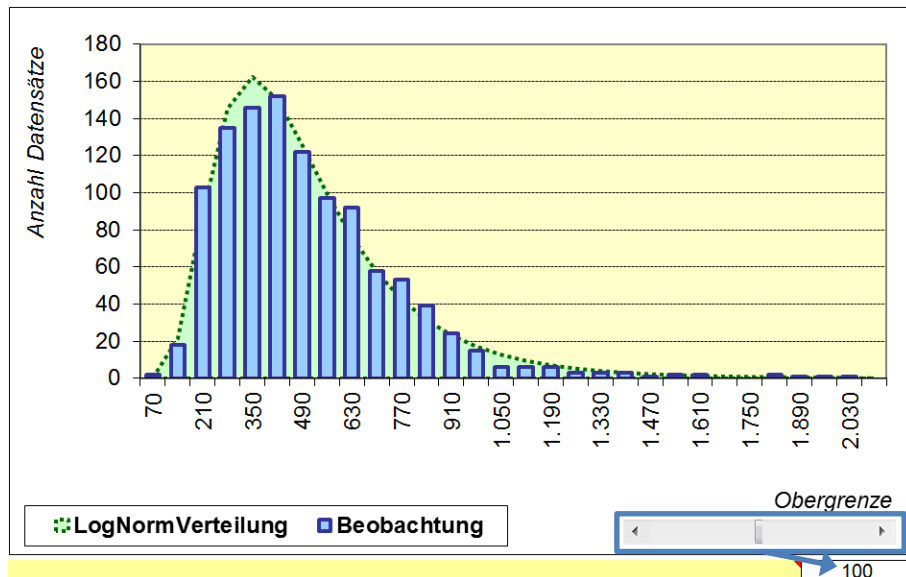
- ➔ Alle Datenklassifizierungen der Tabellenblätter „Schichtung“ und „Schichtungsvertiefung“ können auch mit **freier Klassenbildung** durchgeführt werden. Dazu lässt sich die zweite Schichtobergrenze – und damit das Intervall – komfortabel mit Hilfe des **Schie-**

⁵⁹³ Grubbs, „Engineering Statistics Handbook“, Kapitel 7.1.6. „What are outliers in the data?“, nachzulesen unter: <https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>.

⁵⁹⁴ Der Interquartilsabstand einer Datenmenge beschreibt den Wertumfang zwischen dem 25 %- und dem 75 %-Quantil (1. und 3. Quartil). Dazu wird die Differenz zwischen dem Wertinhalt des – aufsteigend sortiert – 75 %- und 25 %-größten Rangs ermittelt. Weil Wirtschaftsdaten i.d.R. asymmetrisch-rechtsschief verteilt sind (s. *Pkt. 4.3.4.1.*), erfolgt in der SRP-Vorlage Stochastik hierzu eine differenzierte Hochrechnung links- und rechtsseitig des Medians.

⁵⁹⁵ S. die vorherige Fußnote.

bereglers im jeweiligen Diagramm (prozentual) anpassen: in feinen Stufen über die Pfeile und in größeren über die Schienen; die **aktuelle Einstellung** wird im **Kasten unterhalb** ange-



zeigt. Die SRP-Vorlage Stochastik führt das dadurch entstehende Intervall automatisch fort, bis der höchste Wert der untersuchten Daten bzw. die maximale Klassenanzahl „30“ (Tabellenblatt „Schichtung“) resp. „28“ (Tabellenblatt „Schichtungsverteufung“) erreicht ist. Alternativ können die Werte der ersten Schichtobergrenzen einzeln oder gemeinsam in den beiden freien Feldern oberhalb der Tabellen abgeändert werden; **diese Form der manuellen Klassifizierung überschreibt die Justierung des Schiebereglers**. Steht eine manuell eingetragene Schichtobergrenze im Widerspruch zu einem eingetragenen Schichtungsstartwert (s. den vorherigen Aufzählungspkt. zum „Startwert“), färbt sich die Schrift der Zahl rot; dieser Hinweis hat keinen Einfluss auf die gewählte Klassifizierung. Für die Gegenüberstellungen der Häufigkeitsmuster nach Jahren bzw. Wochentagen auf dem Tabellenblatt „Schichtungsverteufung“ kann die (abgeänderte) Klassenbildung vom Tabellenblatt „Schichtung“ mittels Anhak-Kästchen neben den freien Eingabefeldern für die ersten beiden Schichtobergrenzen übernommen werden; **diese Klassenwahl überschreibt sowohl die Justierung des Schiebereglers als auch manuelle Eingaben**, worauf die ausgegrauten Eingabefelder hinweisen. Die jeweils gültigen Schichtobergrenzen lassen sich in den Tabellen und Diagrammen (X-Achsen-Beschriftung) ablesen.

- ➔ Falls infolge der (automatischen oder einer manuell abgeänderten) Klassenfestlegung **nicht alle hohen Spitzenwerte in den Auswertungen berücksichtigt** werden, verfärbt sich das Summenfeld unter der jeweiligen Häufigkeitsverteilung auf den Tabellenblättern „Schichtung“ und „Schichtungsverteufung“ rot.
- ➔ Sowohl für automatische als auch für abgeänderte Klassenbildungen bietet die SRP-Vorlage Stochastik auf den Tabellenblättern „Schichtung“ und „Schichtungsverteufung“ jeweils **per Anhakfeld** eine **Umstellfunktion von linearer auf logarithmische Klassifizierung** an; nach Aktivierung färbt sich die Schrift schwarz. Logarithmische

Klassen entwickeln sich mit gleichmäßig zunehmenden Exponenten: z.B. „ $10^1 - 10^2 - 10^3 - 10^4$ “ entsprechend „10 – 100 – 1.000 – 10.000“. Als Folge der dabei zunehmend größer werdenden Intervalle lassen sich Klassifizierungen einerseits zu sehr hohen Werten

Schicht- obergrenze	Beob- achtung	LogNorm- Verteilung
		Absolutwerte:
10,00	☒ log. Klassifizierung	
20,00		99er Schichtung
10	0	0
20	0	0
40	0	0
80	1	0
160	6	1

ausweiten. Andererseits erleichtert eine logarithmische Klassenbildung durch Symmetrierung die visuelle Übereinstimmungsprüfung, falls ein logarithmisches Häufigkeitsmuster wie die „Logarithmische Normalverteilung“ (s. [Pkt. 4.3.4.1.](#)) erwartet wird.

- ➔ Über ein weiteres **Anhakfeld** (Abb. unterhalb) lässt sich die automatische Klassifizierung direkt auf die **99er-Schichtung** umstellen: zur Abklärung einer entdeckten Zweitzielfernauffälligkeit (s. [Pkt. 4.3.3.2.](#)); durch Aktivierung wird die Beschriftung schwarz. Die Schichtungsvorschläge unter „absolute Kenngrößen (lineare Schichtung)“ auf dem Tabellenblatt „Daten“ werden berücksichtigt; eine Abänderung ist manuell möglich. Wegen der allgemeinen Anwendungsbreite der (logarithmischen) Musteranalyse ist die spezielle 99er-Schichtung gegenüber der Funktion „log. Klassifizierung“ nachrangig, sodass letztere nicht aktiviert sein darf.

- ➔ Die untersuchten Daten können auf dem Tabellenblatt „Schichtung“ per **Anhakfeld** auch als **vorzeichenneutrale Beträge** verarbeitet werden; die Anwahl wird durch schwarze Schrift ersichtlich. Diese Funktion betont u.a. häufige Korrekturbuchungen bzw. -aufzeichnungen in (gleichmäßigen) Häufigkeitsmustern, wodurch sie sich als Prüffelder leichter finden lassen. Die Aktivierung dieser *wertverändernden* Funktion wird zusätzlich durch Rotfärbung der Kenngrößen „Summe“, „Mittelwert“, „Standardabweichung“, „Minimum“ und „Maximum“ auf den Tabellenblättern „Daten“ („statistische Kenngrößen“) sowie „Schichtungsverteufung“ angezeigt. Die Berechnung eines Hilfsminimums und -maximums (s. den Auflistungspkt. „Hilfsminimum“ und „Hilfsmaximum“ oberhalb) wird dadurch unterdrückt.

Schicht- obergrenze	Beob- achtung	LogNorm- Verteilung
		Absolutwerte: ☒
	☐ log. Klassifizierung	
		99er Schichtung ☐
60	1	0
130	4	0
200	13	3
270	20	14

- Für die Gegenüberstellungen der Häufigkeitsmuster von bis zu drei auswählbaren Jahren oder Wochentagen auf dem Tabellenblatt „Schichtungsverteufung“ werden die Datenreihen als **Liniengrafen** wiedergegeben, obwohl es sich um diskrete⁵⁹⁶ Klassenver-

⁵⁹⁶ Unter diskreten Verteilungen werden in der Mathematik solche verstanden, die keine Zwischenwerte einnehmen können. So kann es bei Münzwürfen aus Sicht einer bestimmten Münzseite nur zu ganzzahligen Erfol-

teilungen handelt. Ebenso wird das externe Vergleichsmuster „Logarithmische Normalverteilung“ auf dem Tabellenblatt „Schichtung“ als Linie und Fläche abgebildet. Beides erfolgt aus dem Grund, die visuelle Übereinstimmung und außergewöhnliche Vorgänge durch eine bessere Übersichtlichkeit leichter wahrnehmen zu können.

- ➔ Keine der Klassifizierungen auf den Tabellenblättern „Schichtung“ und „Schichtungsverteilung“ wird mit einem **Wahrscheinlichkeitstest** auf Übereinstimmung geprüft. Da die Auswertungen die Funktionen „Orientierung in größeren Datenumfängen“ und „Absicherung manipulationsorientierter Musteranalysen (gegen nicht-manipulative Ursachen)“ haben und nicht indiziell verwendet werden, ist ein qualifizierter Abschluss unter Einbeziehung der untersuchten Datenmenge (s. [Pkt. 4.3.1.3.](#)) nicht erforderlich, denn herausstechende Einzelsachverhalte werden nach dem Auffinden als betriebsspezifische Prüffelder auf Grundlage der zugehörigen Unterlagen (rechtlich) beurteilt. Bei der Abklärung einer Zweitziiffernauffälligkeit gibt die 99er-Schichtung ausreichend Aufschluss darüber, ob ein begrenzter Zahlenbereich am Dezimalsprung „Hunderter zu Tausender“ oder oberhalb von € 1.000 ursächlich sein kann (s. *die Beispiele unter* [Pkt. 4.3.3.2.](#)); in diesem Fall wird sie als Indiz verworfen. Sollen Tageskasseneinnahmen auf die – manipulationspsychologisch typischen (s. [Pkt. 4.3.4.](#)) – stärkeren Eingriffe in höhere Werte untersucht werden, dienen die Wertklassifizierungen als Vorbereitung zum Erkennen der praxisrelevanten Besonderheit „uneinheitliche Mehrfachverteilungen“ (s. *das Beispiel unter* [Pkt. 4.3.3.3.](#)); eine qualifizierte Übereinstimmungsprüfung mit Wahrscheinlichkeitstest zur Berücksichtigung der untersuchten Datenmenge erfolgt i.R. der Verteilungsanalyse (s. [Pkt. 4.3.4.2.](#)).

4.3.3.2. Abklärung von Zweitziiffernauffälligkeiten mit 99er-Schichtung

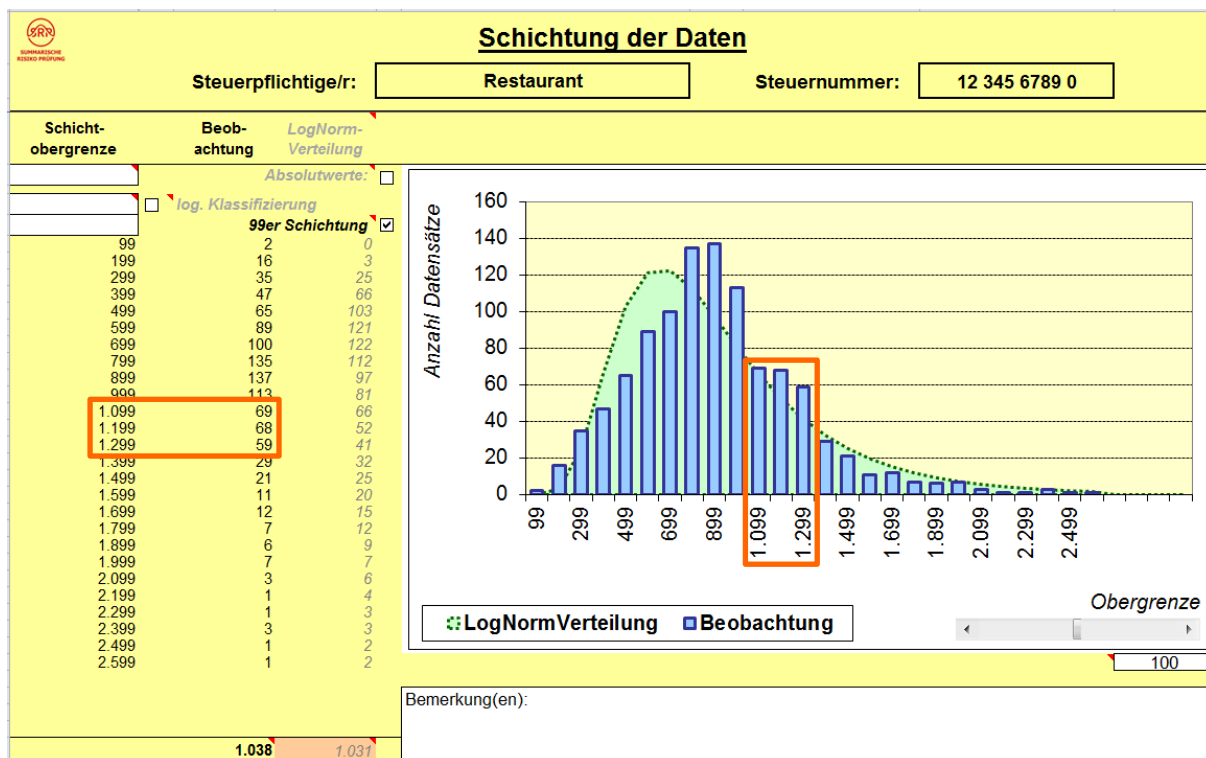
Entdeckte Ziiffernauffälligkeiten haben nur dann indiziellen Aussagewert gegen die sachliche Richtigkeit von Besteuerungsgrundlagen, wenn die anderen beiden Formen menschlicher Eingriffe (s. *oberhalb des Abschnitts „musterbasierte Prüfungsansätze zur Aufdeckung von Manipulationen“ unter* [Pkt. 4.3.](#)) (weitest gehend) ausgeschlossen werden können. Während dies bei Einerstellenuntersuchungen (mathematisch) einfache Überlegungen und (einigermaßen) ordnungsmäßige Betriebsunterlagen erfordert (s. [Pkt. 4.3.2.2.2.](#) sowie *zum Ende des Beispiels unter* [Pkt. 4.3.2.2.3.](#)), hilft bei der Abklärung von Zweitziiffernauffälligkeiten eine **besondere Form der Musteranalyse**: Mit Hilfe der **99er-Schichtung**⁵⁹⁷ lassen sich

gen (bzw. Misserfolgen) kommen, nicht aber zu bspw. 3,437 Treffern. Das Gegenteil kontinuierlicher Ergebnismöglichkeiten wird „stetig“ genannt.

⁵⁹⁷ Exakt müsste für eine gezielte Zweitziiffernbeobachtung oberhalb von € 1.000 mit Schichtobergrenzen gearbeitet werden, die auf „99,99“ enden (z.B. „€ 1.100 bis € 1.199,99“). Bei einer entsprechenden Eingabe auf dem Tabellenblatt „Schichtung“ in der SRP-Vorlage Stochastik ist darauf zu achten, dass im Diagramm auf

die zwei praxisrelevanten Sondergründe für auffällige Zweitziiffernmuster, dass die untersuchten Tageskasseneinnahmen konzentriert in einem begrenzten Zahlenbereich an einem Dezimalsprung – i.d.R. „Hunderter zu Tausender“ – oder besonders oberhalb von € 1.000 vorliegen (s. [Pkt. 4.3.2.3.2.](#)), einfach und schnell bestätigen bzw. ausschließen.

Das erste Beispiel mit auffälliger Zweitziiffernverteilung zu den Tageskasseneinnahmen eines gastronomischen Betriebs *unter* [Pkt. 4.3.2.3.3.](#) wies eine recht gleichmäßige Überbetonung der Ziffern „0“, „1“ und „2“ auf, was in der Gesamtbetrachtung der gegenübergestellten Ziiffernverteilungen „Beobachtung“ und „(Punkt-)Erwartung“ gemäß χ^2 -Test (Chi-Quadrat-Test) eine Nichtübereinstimmungswahrscheinlichkeit von 99,8 % ergab. Die vorgefundene Konstellation mit den zu häufigen kleinsten Ziiffern an der Zweitposition auf Kosten aller größeren ist das typische Bild eines **Dezimalsprungs**: Unterhalb von z.B. € 1.000 umfasst jede Zweitziiffernspanne € 10 wie bspw. in „€ 950 bis € 959,99“, während sie darüber abrupt auf € 100 anwächst (für die betreffende „5“: „€ 1.500 bis € 1.599,99“). Die Wahrscheinlichkeit, dass Datensätze zufällig in die Bereiche hineinfallen, erhöht sich entsprechend. Durch eine 99er-Schichtung auffälliger Tageskasseneinnahmen wie in diesem Diagramm wird die Zweitposition oberhalb von € 1.000 auf eine bestimmte Ziiffer fixiert:

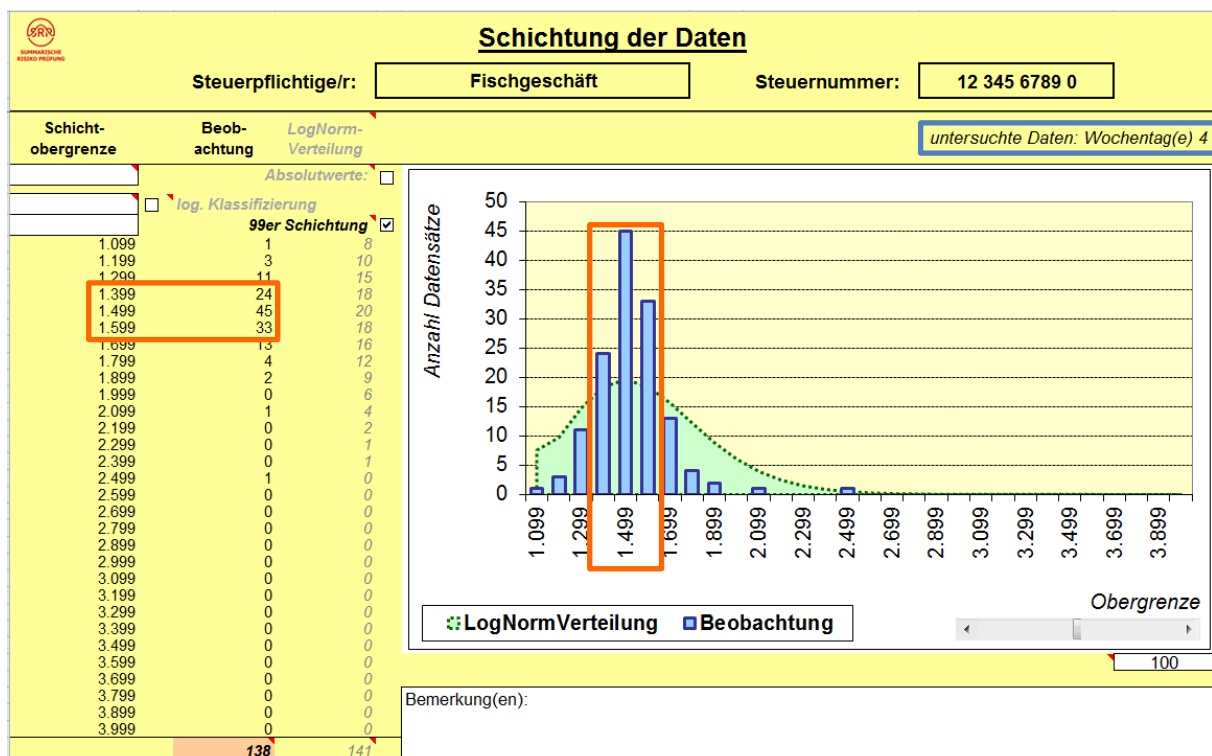


So ist dem Beispiel problemlos zu entnehmen, dass mit **fast 200 Datensätzen** ein **erheblicher** Teil der Tageskasseneinnahmen in jenen Zahlenbereich fällt, der **ab dem Dezimalsprung** **zwingend eine „0“, „1“ oder „2“ an der Zweitposition** bedingt. Da es sich hierbei nicht nur um

ganze Zahlen gerundet wird und somit jeweils die vorherige Zweitziiffer „klassenbestimmend“ ist: Für das Beispiel der Klasse „€ 1.100 bis € 1.199,99“ zeigt die X-Achse als Obergrenze „€ 1.200“ an, wobei alle Werte des Intervalls eine „1“ an der zweiten Position haben.

ein paar Werte im Verteilungsauslauf handelt, ist somit zumindest eine Mitverursachung der Zweitziiffernauffälligkeit bestätigt. **In der Konsequenz ist die Auffälligkeit als Indiz gegen die sachliche Richtigkeit entkräftet.**

Im zweiten Beispiel konnte die auffällige Häufung der „4“ und „5“ an der Zweitstelle der Tageskasseneinnahmen eines Fischgeschäfts (s. [Pkt. 4.3.2.3.3.](#)) mit Hilfe der vertiefenden Auswahl-Drehfelder für Jahre und Wochentage des Tabellenblatts „digZA“ auf den Donnerstag als Hauptverursacher konkretisiert werden. Die glockenförmig gleichmäßige Abnahme der Zweitziiffern von der Häufungszone „3, 4 und 5“ zu den kleineren und größeren hin passt jedoch nicht zu den erwarteten Manipulationsspuren frei erfundener Daten „Ziffernvorlieben, -abneigungen oder -chaos“. In solchen Fällen kann eine erfahrungsgemäß eher seltene Situation für kleinere Betriebe mit Bargeschäften vorliegen: Es werden **regelmäßig höhere Tageskasseneinnahmen oberhalb von € 1.000** erreicht, deren **Streuung** aufgrund sehr gleichmäßiger Nachfrage mit vorwiegend Stammkundschaft aber **begrenzt ausfällt**. Auch diese einzelfallbezogene Besonderheit kann als mögliche nicht-manipulative Ursache für eine beobachtete Zweitziiffernauffälligkeit mit der 99er-Schichtung problemlos abgeklärt werden:



Nachdem im Beispielsfall auf dem Tabellenblatt „Daten“ zunächst die (grundsätzliche) **Einschränkung der untersuchten Daten auf Donnerstage** vorgenommen wurde (s. den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)), bestätigt sich durch 99er-Schichtung die Annahme einer schmalbandigen Konzentration der donnerstäglichen Bareinnahmen oberhalb von € 1.000 als **Häufungszone „ab € 1.300 bis € 1.599“** mit den daraus hervor-

gehenden Zweitzißern „3“, „4“ und „5“. Davon haben sich die zu häufigen Zißern „4“ und „5“ als ursprüngliche Auffälligkeit aller Tageskasseneinnahmen offensichtlich von den Donnerstagen bis in den Gesamtdatenbestand durchgesetzt. **Somit ist der Verdacht frei erfundener Tageskasseneinnahmen anhand der beobachteten Zweitzißernauffälligkeit nach Vertiefung nicht aufrechtzuhalten.** *Auf der anderen Seite ergibt sich für ein Prüfungsnetz die wertvolle Erkenntnis, dass mit der sehr gleichmäßigen Nachfrage über alle Jahreszeiten hinweg eine ausgesprochen gute Planungssituation bestand.*

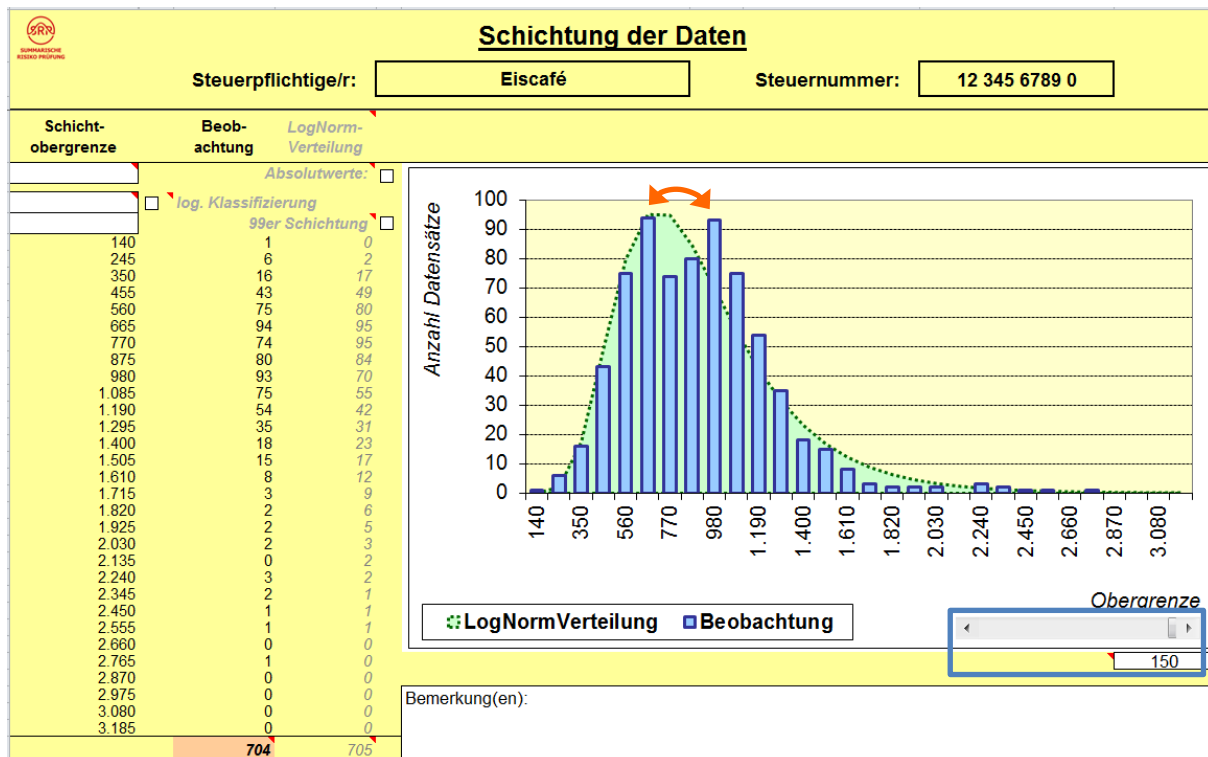
4.3.3.3. Vorbereitende Wertklassifizierungen für die Verteilungsanalyse

I.R. der SRP-Verteilungsanalyse soll das klassifizierte Häufigkeitsmuster von Tageskasseneinnahmen mit der Logarithmischen Normalverteilung (s. [Pkt. 4.3.4.1.](#)) verglichen werden, um die kriminalpsychologisch wahrscheinlichste Manipulationsspur „verstärkte Spitzeneingriffe in die höheren Werte“ besser aufdecken zu können (s. [Pkt. 4.3.4.](#)). Dazu wird aus den beiden Kennzahlen „Mittelwert“ und „Standardabweichung“⁵⁹⁸ der untersuchten Daten eine *einzelfallspezifische* Logarithmische Normalverteilung abgeleitet, deren asymmetrisches Auslaufen in Zahlenbereiche, die deutlich über dem mittleren Niveau liegen (s. [Pkt. 4.3.4.1.](#)), auch die (wenigen) außergewöhnlichen Nachfragetage eines Unternehmens erfasst. In besonderen Fällen kann allerdings insgesamt eine **uneinheitliche Absatzsituation** vorliegen, indem bestimmte Zeiten wie z.B. Wochenendtage oder eine Hochsaison nicht zur eigentlichen Häufigkeitsverteilung passen: „Bei der Lognormalverteilungsanalyse muss besonders auf die Homogenität des Datenbestandes geachtet werden. Die Verteilung der Tageseinnahmen kann sich je nach betrieblicher Besonderheit bspw. in den einzelnen Geschäftsjahren, saisonal oder wochentagsspezifisch unterscheiden.“⁵⁹⁹ Daraus kann eine Überlagerung von **zwei oder mehr** Verteilungen resultieren, welche nicht mit *einer* reproduzierten Absatzverteilung übereinstimmen können. Die variable Musterverbildlichung auf dem Tabellenblatt „Schichtung“ und die praxisorientierten Gegenüberstellungsoptionen von Jahren und Wochentagen auf dem Tabellenblatt „Schichtungsverteifung“ in Verbindung mit den Dateneinschränkungsfunktionen auf dem Tabellenblatt „Daten“ (s. *den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)*) erlauben ein effizientes Erkennen von Mehrfachverteilungen und die gezielte Auftrennung in homogene Teildatenbestände für deren anschließende separate Untersuchung.

⁵⁹⁸ Die Standardabweichung beschreibt die Streuungsintensität einer Datenmenge als vorzeichenneutrale typische Abweichung jedes einzelnen Datensatzes zum Mittelwert.

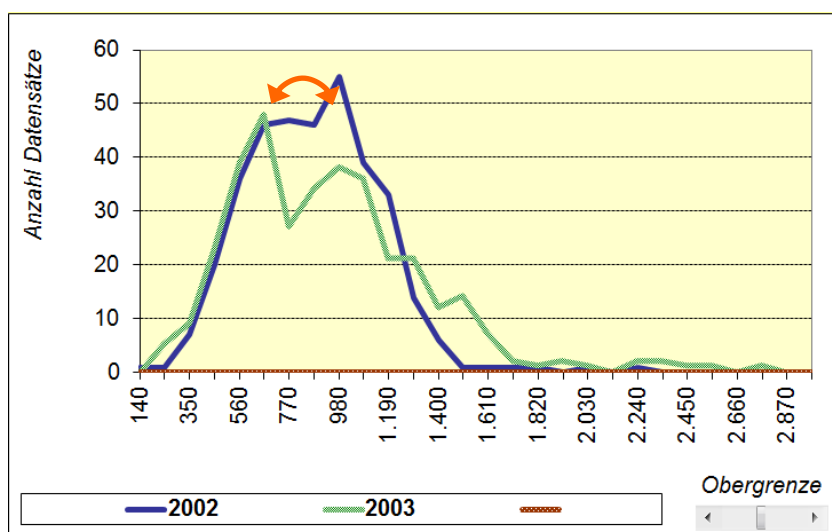
⁵⁹⁹ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 161.

Im abgebildeten Beispiel zu den Tageskasseneinnahmen eines Eiscafés wurde die automatische Klassifizierung der SRP-Vorlage Stochastik mit Hilfe des Schiebereglers größer⁶⁰⁰ eingestellt, um die Einheitlichkeit des Verteilungsmusters besser sehen zu können:



Im visualisierten Mustervergleich ist problemlos zu erkennen, dass eine **doppelgipflige Überlagerung von zwei Häufigkeitsverteilungen** einer einheitlichen (Punkt-)Erwartung gegenübersteht. Folglich kann ein Übereinstimmungstest nur zum Urteil der (nahezu) sicheren Nichtübereinstimmung kommen.

Vor einer indiziellen Untersuchung dieser Tageskasseneinnahmen auf verstärkte Spitzeneingriffe mit der Verteilungsanalyse (s. [Pkt. 4.3.4.](#)) müssen die zwei Absatzbereiche identifiziert und aufgetrennt werden. In einem ersten Schritt werden dazu die beiden Jahre des Prüfungszeitraums auf dem Tabellenblatt „Schichtungsverteufung“ mit aktivierter Übernahme der

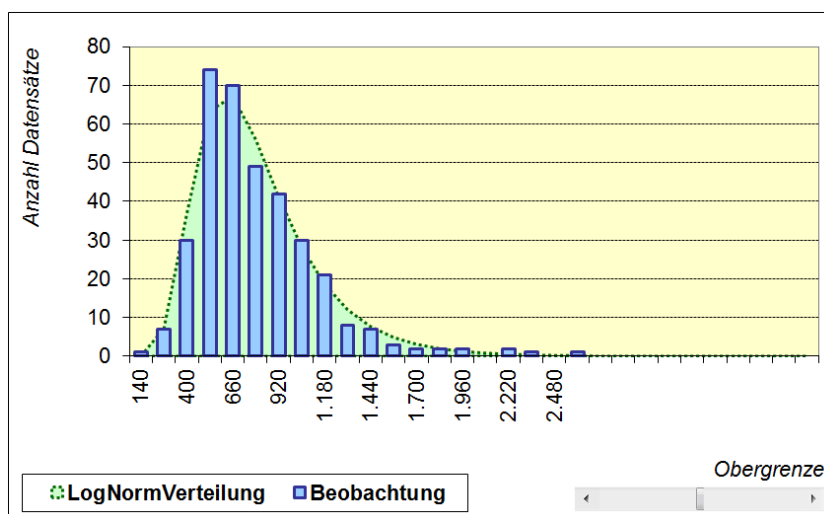


⁶⁰⁰ Wie bei der SRP-Zeitreihensystematik (s. den Abschnitt „Fuzzy Theorien“ unter Pkt. 4.2.2.) entscheidet die Intervallgröße bei Klassifizierungen darüber, ob entweder durch breitere Klassen der grobe Blick auf das Häufigkeitsmuster oder durch engere das Erkennen von Ausreißern in den Fokus gelangt.

Schichtobergrenzen vom Tabellenblatt „Schichtung“ gegenübergestellt. Diese Vertiefung löst das Problem nicht, denn beide Jahresgrafen weisen – neben Unterschieden – **eine Mischverteilung mit zwei Spitzen** auf.

Ein anschließender interner Vergleich der Häufigkeitsmuster für die einzelnen Wochentage auf dem Tabellenblatt „Schichtungsvertiefung“ legt offen, dass die Freitage und Sonnabende grundsätzlich ein höheres Einnahmeniveau erreichen⁶⁰¹. Durch die Einschränkung der Daten auf die Wochentage „1“, „2“, „3“, „4“ und „7“ bzw. „5“ und „6“ auf dem Tabellenblatt „Daten“ (s. den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)) kann umgehend geprüft werden, ob diese Beobachtung ursächlich für die zwei Gipfel im Häufigkeitsmuster ist. Auch diese Differenzierung der Tageskasseneinnahmen führt hier jedoch nicht zu einer einheitlichen Verteilung.

Als weitere praxistypische Option zur Aufteilung untersuchter Daten in uneinheitliche Teilbestände werden mit Hilfe der Einschränkungsoptionen auf dem Tabellenblatt „Daten“ eine Haupt- und Nebensaison unterschieden (s.a. das Beispiel im Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)). Im Beispielsfall löst sich die Überlagerung von



zwei Verteilungen für den abgebildeten – manipulationspsychologisch unkritischen – Zeitraum „1.10. bis 31.3.“ weitest gehend auf. Demgegenüber führt die Hauptsaison – selbst unter Einschränkung auf die recht kurze Phase „1.6. bis 30.9.“ – zu einer chaotischen Häufigkeitsverteilung.

Wegen der daraus resultierenden Schlussfolgerung, dass die Tageskasseneinnahmen (tlw.) frei erfunden⁶⁰² sein können, sind besonders quantitative Schlussfolgerungen aus der SRP-Verteilungsanalyse (s. [Pkt. 4.3.4.6.](#)) mit Vorsicht zu ziehen, weil jeglicher Verteilungs-

⁶⁰¹ Die Liniengrafen der Häufigkeitsverteilungen zu den Freitagen und Sonnabenden liegen im Diagramm insgesamt weiter rechts bei den höheren Werten auf der X-Achse als diejenigen der anderen Tage.

⁶⁰² Falls die Ziffernanalysen keinen mathematisch eindeutigen (signifikanten) Hinweis auf Spuren des freien Erfindens von Tageskasseneinnahmen ergeben haben, ist das dennoch keine Garantie für die gegenteilige Annahme. Als beispielhaftes Manipulationsmodell wird das regelmäßige Herausnehmen von glatten Hunderterbeträgen keine Ziffernvorlieben oder -abneigungen hinterlassen, kann das Häufigkeitsmuster jedoch bis zum Chaos zerstören. Insofern muss diese Aussage aus dem BFH-Urteil v. 16.9.2021 ([BSStBl. II 2022, 101](#)) verwirren: „So hat den Feststellungen des FG zufolge auch der in der mündlichen Verhandlung vor dem FG anwesende Kassensystemprüfer glaubhaft angegeben, er selbst habe trotz genauer Kenntnis des [Chi-Quadrat-]Tests keine manipulierte Buchführung erstellen können, die den Test bestanden hätte.“

zusammenhang zerstört sein kann (s. den Abschnitt „Prüfungsschema zur Vermeidung von Fehldeutungen“ unter [Pkt. 4.3.](#)). Insgesamt stellt Sell zur sorgfältigen Vorbereitung einer indiziellen Prüfung von Bareinnahmen auf typische manipulative Eingriffe anhand der Logarithmischen Normalverteilung fest: „Ein solcher Umstand einer ‚uneinheitlichen Absatzsituation‘ wird in der Praxis [...] regelmäßig und sehr tiefgehend geprüft (SRP-Handbuch“⁶⁰³.

4.3.4. Verteilungsanalyse

Die Verteilungsanalyse der SRP ist auf das Ziel ausgerichtet, speziell in Tageskasseneinnahmen statistisch eindeutige (signifikante) **Spuren von Spitzeneingriffen** – also verstärkte Verkürzungen der höheren Werte – aufzudecken. Im Gegensatz zu Ziffernanalysen können Gegenüberstellungen der (klassifizierten) Häufigkeitsverteilung von Besteuerungsdaten mit der dazu rekonstruierten Logarithmischen Normalverteilung (s. [Pkt. 4.3.4.1.](#)) grundsätzlich unterschiedlichste Formen von menschlichen Eingriffen anzeigen⁶⁰⁴, i.R. der SRP wird die Anwendung aber auf diese eine praxisrelevante Manipulationsart beschränkt, um das Risiko von Fehldeutungen⁶⁰⁵ zu minimieren.

Derartige stärkere Spitzeneingriffe sind auch unter Verwendung moderner Manipulationshilfsmittel [s. den Abschnitt „raffinierte (digitale) Manipulationswerkzeuge“ unter [Pkt. 2.1.](#)] das **wahrscheinlichste Eingriffsmuster**, denn nach neurophysiologischen Erkenntnissen werden die allermeisten menschlichen Handlungen nicht bewusst gesteuert, sondern unterbewusst entschieden und im Nachhinein (möglichst) rational erklärt⁶⁰⁶. Dabei lösen Gesetzesbrüche immer zuerst eine – nicht unbedingt bewusst wahrgenommene – (adrenerge⁶⁰⁷) Stressreaktion aus. Tritt durch die gesetzeswidrige Handlung ein persönlicher Erfolg ein, wird das Belohnungshormon „Dopamin“ ausgeschüttet. Bereits ein stark aktiviertes Be-

⁶⁰³ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 161.

⁶⁰⁴ Vgl. z.B.: Ernst, „Indirekte Methoden der steuerlichen Betriebsprüfung: klassische Verprobung, Systemprüfung, neuere Verfahren“, Schmidt-Verlag, 1986, oder Schumann, „Voraussetzungen und Verfahren der abgabenrechtlichen Schätzung nach § 162 Abs. 1 und 2 AO – insbesondere im Fokus der steuerlichen Außenprüfung“, Diss. an der Johannes Gutenberg-Universität Mainz, 2009, 105 f. mit diversen Abb.

⁶⁰⁵ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. [...] Die korrekte Interpretation solcher Grafiken ist allerdings nicht trivial und erfordert spezielle kognitive Fähigkeiten.“

⁶⁰⁶ Vgl. z.B.: Damasio, „Descartes' Irrtum. Fühlen, Denken und das menschliche Gehirn“, List-/Ullstein-Verlag, 2004, 25 ff., oder Ramachandran, „Eine kurze Reise durch Geist und Gehirn“, Rororo-Verlag, 2. Auflage 2006, 100 f. u. 119.

⁶⁰⁷ Der Nervenbotenstoff „Noradrenalin“ ist namensgebend für stressartige Reaktionen und den betreffenden Teil des (autonomen) vegetativen Nervensystems.

lohnungssystem allein kann unlogisches Verhalten bewirken⁶⁰⁸, für die wiederholte Abfolge von (Nor-)Adrenalin- und Dopamin-Ausschüttung wird sogar eine hohe Gefahr der Suchtentstehung gesehen⁶⁰⁹. In der Konsequenz dürften auch Steuerhinterziehungen der typischen Dynamik einer zunehmenden Intensivierung folgen. Dafür spricht außerdem der aus Studien bekannte Zusammenhang, dass Menschen, die ihr Geld (spielerisch) leicht verdienen, stärker zur Steuerhinterziehung neigen, als jene, die dafür hart arbeiten müssen⁶¹⁰.

Insbesondere folgende **Manipulationsszenarien**, die von physiologischen Abläufen oder operativen Notwendigkeiten⁶¹¹ beeinflusst sind, hinterlassen Spuren von Spitzeneingriffen und können somit durch die SRP-Verteilungsanalyse aufgedeckt werden:

-  Jede Form der Einnahmenverkürzung durch Positionslöschungen oder anteilige Minderung(en) kann zu stärkeren Spitzeneingriffen führen, wenn das Gewissen oder der Kassenbestand einen vergleichbaren Eingriff in einnahmeschwächere Tage behindern.
-  Eine systematische Einnahmenverkürzung bspw. durch prozentuale Positions-minderungen kann dann erkennbare Spuren in den (betriebsbezogen) höheren Tageskasseneinnahmen erzeugen, wenn die Gier innerhalb der langen Phase eines Prüfungszeitraums zunimmt und der Minderungsanteil erhöht wird.
-  Bei dem typischen Steuerhinterziehungsmodell „Schwarzlöhne aus Schwarzeinnahmen“ kann es zu stärkeren Spitzeneingriffen kommen, weil die schwarz bezahlten Kosten für Aushilfskräfte an Hocheinnahmetagen i.d.R. überproportional ansteigen.

Eine permanent systematische Einnahmenverkürzung unter Zuhilfenahme digitaler Manipulationsinstrumente kann auch durch Ansätze wie die Verteilungsanalyse nicht aufgedeckt werden, widerspricht allerdings den (kriminal-)psychologischen Erkenntnissen, weshalb Tenzer fragt: „Was können Psychologen den Finanzämtern raten?“⁶¹² Insofern erfüllt die Verteilungsanalyse genau das vom BFH beschriebene berechnete Interesse der Finanzverwaltung an wirksameren Verprobungsmethoden: „Elektronische Kassensysteme sind durch Umpro-

⁶⁰⁸ Für Funktionsstudien des Belohnungszentrums wurde Ratten per Schalter die Möglichkeit gegeben, die betreffende Zwischenhirnregion elektrisch zu stimulieren. Die Tiere haben daraufhin selbst sämtliche lebens- und arterhaltenden Handlungen (Nahrungsaufnahme, Interaktion und Vermehrung) komplett vernachlässigt, weil sie innerhalb kürzester Zeit süchtig nach dem Schalter wurden, vgl.: *Olds/Milner*, „Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain“, *Journal of Comparative Physiological Psychology* 47(6), 1954, 419.

⁶⁰⁹ Als Alltagsbeispiel für die Suchtgefahr durch gegenseitige Verstärkung von Adrenalin, Endorphinen und Dopamin kann Extremsport angeführt werden; vgl. hierzu z.B.: news.de online v. 28.7.2009, „Faszination Extremsport: Die Sucht nach der Leere im Gehirn“, nachzulesen z.B. unter: www.news.de/gesundheit/854672263/die-sucht-nach-der-leere-im-gehirn/1.

⁶¹⁰ Tenzer, „Warum Menschen Steuern zahlen – oder eben nicht“, *Psychologie Heute* 9/2013, 30, 32 mit Verweis auf: *Mühlbacher/Kirchler*, „Arbeitsaufwand, Anspruchsniveau und Steuerehrlichkeit“, *Zeitschrift für Arbeits- und Organisationspsychologie* 2/2008, 91.

⁶¹¹ Auch Dokumentationspflichten wie das taggenaue Führen einer Kasse können einen Einfluss auf Manipulationsweisen haben: bspw. um das Entstehen von Kassenfehlbeträgen zu vermeiden.

⁶¹² Tenzer, „Warum Menschen Steuern zahlen – oder eben nicht“, *Psychologie Heute* 9/2013, 30, 33.

grammierung in nahezu beliebiger Weise manipulierbar; von derartigen Manipulationsmöglichkeiten machen Teile der betrieblichen Praxis nach dem Erkenntnisstand des Senats durchaus Gebrauch [...] Vielmehr besteht auf Seiten der Finanzverwaltung ein durchaus nachvollziehbares Bedürfnis, moderne Prüfungsmethoden – zu denen in geeigneten Fällen auch mathematisch-statistische Methoden gehören können – einzusetzen.“⁶¹³ Noch deutlicher betont der Bundesrechnungshof die Bedeutung von musterbasierten Verfahren wie der Verteilungsanalyse für einen wirksamen Steuervollzug: „[D]as Bundesministerium der Finanzen (BMF) [muss] auf den bundesweit einheitlichen Einsatz IT-gestützter quantitativer Prüfungsmethoden in der steuerlichen Betriebsprüfung hinwirken[, ... denn t]raditionelle Methoden [...], beispielsweise Beleg- und Akteneinsicht, verlieren an Effektivität und Effizienz. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.“⁶¹⁴

4.3.4.1. Die (Punkt-)Erwartung „Logarithmische Normalverteilung“

Nach **empirischen**⁶¹⁵ **Erfahrungen** findet sich die Logarithmische Normalverteilung – anstelle der Normalverteilung – in vielen Wissenschaftsbereichen als regelmäßig beobachtetes Verteilungsverhalten von Daten wieder⁶¹⁶; besonders werden hierbei die Bereiche „Biologie und Wirtschaft“⁶¹⁷ herausgestellt. Letzteres passt dazu, dass das Benford-Gesetz⁶¹⁸ ebenfalls vornehmlich in Wirtschaftsdaten wiederzufinden ist, denn es besteht ein inhaltlicher

⁶¹³ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27 u. 54; nach Ansicht des Autors sollten die Finanzgerichte im Hinblick auf die vom BFH dargelegte erschwerte Beurteilungssituation bei Besteuerungsgrundlagen, die möglicherweise digital manipuliert worden sind, dasselbe Bedürfnis nach zeitgemäß effektiven Prüfungs- und Verprobungsmethoden haben, weil gem. § 76 Abs. 1 S. 1 FGO „[d]as Gericht den Sachverhalt von Amts wegen [erforscht].“

⁶¹⁴ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>. Mit dem BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1, 4 u. 12 f. schließt sich die Finanzverwaltung bzgl. der Wichtigkeit quantitativer Prüfungsmethoden und speziell der musterbasierten Häufigkeitsanalysen an: „Bei einer steuerlichen Außenprüfung können mathematisch-statistische (quantitative) Prüfungsmethoden eingesetzt werden, um die Besteuerungsgrundlagen des Steuerpflichtigen zu verproben und auf Plausibilität zu überprüfen. [...] Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgabenordnung dazu führen, dass eine Schätzung durchgeführt werden muss. [...] 3. Struktur- und Verteilungsanalyse“.

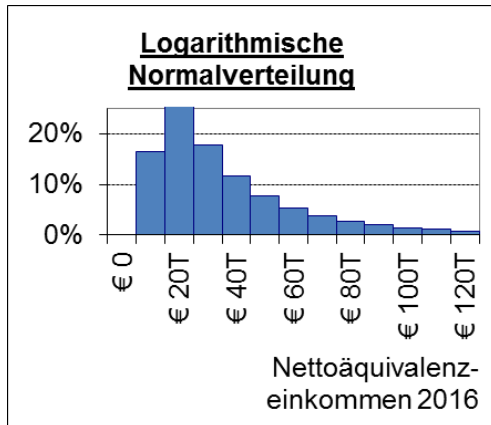
⁶¹⁵ Empirische Erkenntnisse stammen aus Studien, die den Untersuchungsgegenstand unter nachvollziehbaren konstanten Bedingungen beobachten, und sollen theoretische Annahmen entweder bestätigen oder verwerfen. Gesicherte empirische Erkenntnisse genießen einen höheren Stellenwert in der Wissenschaft als widerstreitende theoretische Annahmen, weil sie die Realität beschreiben.

⁶¹⁶ Vgl. z.B.: *Limpert/Stahel/Abbt*, „Lognormal distributions across the sciences: keys and clues“, BioScience Band 51, Nr. 5, 2001, 341, nachzulesen unter: https://de.wikipedia.org/wiki/Logarithmische_Normalverteilung.

⁶¹⁷ *Sachs*, „Angewandte Statistik“, Springer-Verlag, 10. Auflage 2002, 173.

⁶¹⁸ *Benford*, „The Law of Anomalous Numbers“, Proceedings of the American Philosophical Society 1938, 551.

Zusammenhang zwischen beiden (Punkt-)Erwartungen⁶¹⁹. Entsprechend berichtet *Rau*, dass „die logarithmische Normalverteilung [...] bei der Verteilung des Werts von Hauseigentum, von Industrien eines Landes nach Anzahl der Arbeitnehmer, Größe der Betriebe, Produktion, Kapital und Fabrikationskosten, bei der Verteilung von Einlagen bei Banken und Kassen, der Größe von landwirtschaftliche[n] Anwesen, von Löhnen und Gehältern in Berufsgruppen bei



einzelnen Großbetrieben sowie bei vielen wirtschaftlichen Zuständen auftritt⁶²⁰; die nebenstehende Logarithmische Normalverteilung⁶²¹ beschreibt als Beispiel die Einkommensverhältnisse 2016 in Deutschland auf Basis des Nettoäquivalenzeinkommens mit der sozialen Absicherung links und der weit auslaufenden Spreizung zu hohen Einkommen rechts. Zu Besteuerungsdaten sind dem Autor bisher keine umfangreichen Forschungen bekannt, aber *Sell* resü-

miert i.R. ihrer Dissertation: „Die Untersuchung der Tageseinnahmen auf Lognormalverteilung erfolgte visuell anhand von Histogrammen und QQ-Plots sowie rechnerisch anhand verschiedener Normalitätstests. [...] Die Ergebnisse der Gastronomiebetriebe deuten an, dass es bei Gastronomiebetrieben zumindest nicht unplausibel erscheint, von lognormalverteilten Tageseinnahmen als Verteilungsideal auszugehen. [...] Auch wenn die den Methoden zugrundeliegenden Verteilungsannahmen nicht auf mathematischer Sicherheit aufbauen, deuten die exemplarischen Darstellungen an, dass ihr Einsatz durchaus gerechtfertigt sein könnte. Auf dieser Basis lohnt es sich, die Bemühungen um empirische Daten zu verstärken und insbesondere verschiedene Branchen dabei zu berücksichtigen.“⁶²²

Als **mathematisch-theoretische** Erklärung für das asymmetrische Streuungsverhalten der Logarithmischen Normalverteilung wird ein multiplikatives Zusammenwirken von

⁶¹⁹ *Petersen*, „Benford's Law und die Betriebsprüfung – Eine Erwiderung auf Wähnerts ‚Interaktive Außenprüfung – Möglichkeiten der gemeinsamen Sachverhaltsaufklärung im digitalen Zeitalter‘ (Stbg 2014 S. 20 ff.) aus wissenschaftshistorischer Perspektive“, Stbg 12/2015, 506, 510: „Ist die Kongruenz der diskreten Benford-Verteilung mit der stetigen logarithmischen Normalverteilung wohl in der Tat existent [...]“ mit Verweis auf: *Scott/Fasli*, „Benford's Law: An Empirical Investigation and a Novel Explanation“, CSM Technical Report 329, Essex 2001, 15.

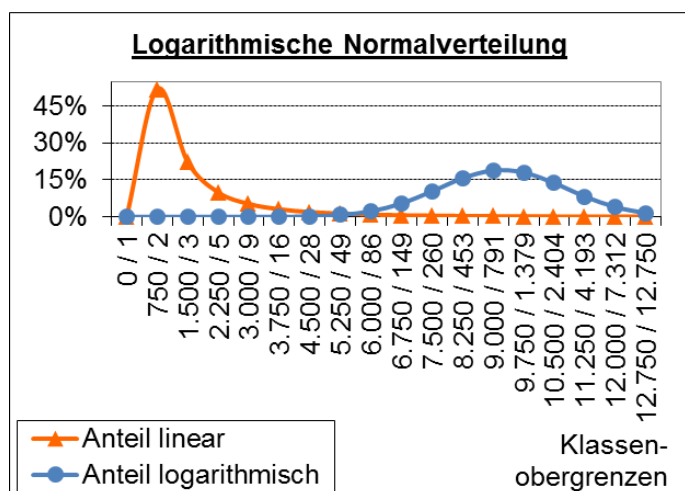
⁶²⁰ *Rau*, „Statistisch-mathematische Methoden der steuerlichen Betriebsprüfung und die Strukturanalyse als ergänzende Alternative“, Reihe: „Steuern, Wirtschaft und Recht“, Bd. 316, Eul-Verlag, 2012, 43 mit Verweis auf: *Daeves/Beckel*, „Großzahl-Methodik“, Verl. Chemie, 1958, 74 ff.

⁶²¹ Die abgebildete Logarithmische Normalverteilung ist auf Basis der Daten des Statistischen Bundesamtes rekonstruiert worden; die Daten können nachgelesen werden unter:
<https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Einkommen-Konsum-Lebensbedingungen/Lebensbedingungen-Armutsgefaehrung/inhalt.html#sprg475684>.

⁶²² *Sell*, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 182 u. 222.

Zufallsgrößen angeführt.⁶²³ M.E. dürfte jedoch der entscheidende Unterschied zur glockenförmig symmetrischen Zufallsstreuung gemäß Normalverteilung in einer linksseitigen Begrenzung der möglichen Datenausprägung liegen: Ein gut nachvollziehbares Beispiel im technischen Bereich ist hierfür der durchschnittliche Kraftstoffverbrauch von Fahrzeugtypen, indem sich unterhalb eines Minimalwerts (für den Betrachtungszeitpunkt) keine Autos herstellen lassen. Mit auf der X-Achse ansteigenden Verbräuchen nehmen die Häufigkeiten bis zu einem „Normverbrauch“ zu. In den darüber hinausgehenden Verbrauchsklassen gehen sie langsamer zurück und reichen erheblich weiter bis zum Verbrauchsmaximum. Ähnliche Verteilungsdynamiken kommen in der Biologie vor, wenn bspw. Lebensmittel konserviert werden: Die mikrobiologische Verunreinigung des Produkts wird durch eine annähernde Sterilität links auf der X-Achse begrenzt. Vom häufigen Standardzustand der Lebensmittel nach Konservierung reicht die Streuung zu höheren Restverunreinigungen deutlich weiter nach rechts hinaus. Diese Form der eigentlich glockenförmigen Streuung, die aber im linksseitigen Auslauf beschränkt wird, ist auch für das Wirtschaftsleben typisch: Hier stammt die untere Grenze entweder aus der sozialen Absicherung wie in der Abb. oberhalb zum Nettoäquivalenzeinkommen oder aus der Marktbereinigung, wenn erfolglose Unternehmen geschlossen werden (müssen), an absatzschwächeren Tagen nicht mehr geöffnet wird, schlecht verkäufliche Produkte aus dem Sortiment genommen werden usw. Insofern ist die regelmäßig beobachtete Häufigkeitsverteilung „Logarithmische Normalverteilung“ in Tageskasseneinnahmen – inklusive Benford-Gesetz für die führenden Zahlenpositionen und (annähernder) Gleichverteilung endständiger Zahlenstellen (ab der Drittziffer) – auch theoretisch schlüssig. **Beachtet werden muss dabei allerdings, dass die untersuchten Daten einen einheitlichen Bestand bilden.** Deshalb schützt die vorbereitende Wertklassifizierung der SRP-Musteranalytik davor, eine mögliche Uneinheitlichkeit zu übersehen und hilft bei der Ermittlung von Ursachen (s. [Pkt. 4.3.3.3.](#)).

Die besondere **Eigenschaft** von Logarithmischen Normalverteilungen gegenüber anderen asymmetrisch linkssteil-rechtsschiefen Häufigkeitsverteilungen ist, dass sie sich **unter logarithmischer Wertklassifizierung in symmetrische Normalverteilungen verwandeln**: In



⁶²³ Sachs, „Angewandte Statistik“, Springer-Verlag, 10. Auflage 2002, 173.

dieser Abb.⁶²⁴ ist die **asymmetrische Form einer Logarithmischen Normalverteilung** zu sehen, wie sie aus der Auszählung von entsprechenden Daten nach den linearen Klassen des unteren Teils der X-Achsenbeschriftung hervorgeht. Werden *dieselben* Daten entsprechend des oberen Teils der X-Achsenbeschriftung in logarithmische⁶²⁵ Intervalle aufgeteilt, ergibt sich das **glockenförmige Muster einer Normalverteilung**. Für die vollständige Beschreibung einer Logarithmischen Normalverteilung reichen die zwei Kenngrößen „Mittelwert“ und „Standardabweichung“⁶²⁶ aus dem dazugehörigen Datenbestand. Demnach lassen sich diese Verhältnisse problemlos auf den Einzelfall übertragen. Dabei hat die logarithmisch-klassifizierte symmetrierende Gegenüberstellung von „Beobachtung“ und „(Punkt-)Erwartung“ i.R. der SRP (s. [Pkt. 4.3.4.2.](#)) zwei Vorteile:

1. Das visuelle Ergebnis der Verteilungsanalyse kann infolge der Symmetrie der (Punkt-) Erwartung leichter beurteilt werden.
2. Wegen der permanent größer werdenden (logarithmischen) Häufigkeitsklassen kann die Übereinstimmungsprüfung mittels χ^2 -Test (Chi-Quadrat-Test) bis zu deutlich höheren Werten mit größerem Manipulationsrisiko (s. [Pkt. 4.3.4.](#)) ausgewertet werden, ohne die Mindestdatenempfehlung (s. [Pkt. 4.3.1.3.](#)) zu verletzen.

Die Übereinstimmungsprüfung des beobachteten Häufigkeitsmusters aus der Werteklassifizierung untersuchter Tageskasseneinnahmen mit der fallspezifisch rekonstruierten Logarithmischen Normalverteilung erfolgt durch den χ^2 -Test (Chi-Quadrat-Test)⁶²⁷; dadurch beinhaltet das Übereinstimmungsergebnis bereits die Datenanzahl (s. [Pkt. 4.3.1.3.](#)). Die übliche Empfehlung zur Mindestdatenmenge „ ≥ 5 erwartete Elemente pro Klasse“ wird von der SRP-Vorlage Stochastik automatisch in Abhängigkeit von der Klassenfestlegung beachtet, indem nur der Hauptverteilungsbereich (stochastisch) ausgewertet wird⁶²⁸, wobei sich die einbezogenen und die ausgeschlossenen Intervalle auf dem Tabellenblatt „Verteilungsanalyse“ leicht unterscheiden lassen (s. *das Beispiel im Abschnitt „Chi-Quadrat-Test“ unter [Pkt. 4.3.4.2.](#)*). Formatseitig werden die zwei Auffälligkeitsschwellen „99 %“ und „99,95 % Nichtübereinstimmung“ hervorgehoben (s. [Pkt. 4.3.2.1.](#)).

⁶²⁴ Aus Wahrnehmungsgründen werden beide klassifizierten Verteilungen als Liniengrafen dargestellt.

⁶²⁵ Die Klassenobergrenzen entsprechen ca. „ $10^{0,00} - 10^{0,24} - 10^{0,48} - 10^{0,72} - 10^{0,96}$ etc.“ mit gleichmäßiger Vergrößerung des Exponenten zur Basis „10“.

⁶²⁶ Die Standardabweichung beschreibt die Streuungsintensität einer Datenmenge als vorzeichenneutrale typische Abweichung jedes einzelnen Datensatzes zum Mittelwert.

⁶²⁷ Spezielle Form- und Symmetrietests zur Übereinstimmungsprüfung mit der Normalverteilung wie Anderson-Darling-, Shapiro-Wilk- oder Jarque-Bera-Test wurden deshalb nicht (für die symmetrierte Verteilungsanalyse) herangezogen, weil sie für verzerrende Einzelausreißer empfindlicher sind; kritisch hierzu: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 157 f.

⁶²⁸ Kurvenverteilungen, die sich (asymptotisch) der X-Achse annähern, wie die Normal- und die Logarithmische Normalverteilung fallen zwangsweise zu den Seiten hin irgendwann unter die Klassengrenze „ ≥ 5 “. Deshalb wird der χ^2 -Test (Chi-Quadrat-Test) als Übereinstimmungsprüfung für die Logarithmische Normalverteilung i.R. der Verteilungsanalyse auf die Klassen mit einer Erwartung „ ≥ 5 “ beschränkt.

4.3.4.2. Anwendungshinweise zur SRP-Verteilungsanalyse

Als speziell für die Anwendung auf Tageskasseneinnahmen konzipierte Form der stochastisch abgesicherten Musteranalytik (s. [Pkt. 4.3.4.](#)) finden sich die Tabellenblätter „Verteilungsanalyse“, „Verteilungsdiagramm“, „Referenzsumme“ und „Referenzdiagramm“ innerhalb der SRP-Vorlage Stochastik im zweiten Teil mit der vorgelagerten „Kurzanleitung 2“; um deren **sinvollen** Einsatz zu unterstützen, **müssen die Tabellenblätter „Verteilungsanalyse“ und „Referenzsumme“ jeweils bewusst über ein Anhakfeld mit entsprechendem Hinweis aktiviert werden.**

Folgende allgemeine Hinweise zur Durchführung einer Verteilungsanalyse mit der SRP-Vorlage Stochastik sind wichtig:

- ➔ Alle auf dem Tabellenblatt „Daten“ vorgenommenen **generellen Dateneinschränkungen** (s. den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)) wirken auf die Verteilungsanalyse der SRP-Vorlage Stochastik; entsprechende Hinweise erfolgen auf den Auswertungsblättern „Verteilungsanalyse“, „Verteilungsdiagramm“, „Referenzsumme“ und „Referenzdiagramm“.
- ➔ Die Festlegung eines „**Startwerts**“ für die Werteklassifizierungen auf dem Tabellenblatt „Schichtung“ (s. den Aufzählungspkt. „Startwert“ unter [Pkt. 4.3.3.1.](#)) hat auf die Verteilungsanalyse keine Auswirkung, weil Musterrekonstruktion und -bewertung hierdurch beeinträchtigt werden könnten.
- ➔ Die Funktion „**Absolutwerte**“ zu den Werteklassifizierungen auf dem Tabellenblatt „Schichtung“ (s. den Aufzählungspkt. „vorzeichenneutrale Beträge“ unter [Pkt. 4.3.3.1.](#)), mit welcher Korrekturbuchungen bzw. -aufzeichnungen hervorgehoben werden können, hat auf die Verteilungsanalyse keine Auswirkung, weil dies im Widerspruch zu deren Ausrichtung stünde.
- ➔ Enthalten die zu untersuchenden Daten **negative oder Nullwerte**, können diese nicht in die Verteilungsanalyse einbezogen werden, weil zu ihnen kein Logarithmus existiert; durch Rotfärbung des Felds „Anzahl“ unter den „logarithmischen Kenngrößen (Verteilungsanalyse)“ auf dem Tabellenblatt „Daten“ wird darauf hingewiesen. **Das Vorliegen von negativen und Nullwerten i.R. der Untersuchung von Tageskasseneinnahmen auf Vollständigkeit ist ein eigenes Prüffeld; hierzu können die entsprechenden Datensätze auf dem Tabellenblatt „Daten“ einfach und schnell über den Filter zur Spalte „alle Werte“ identifiziert werden.**

- Die automatische **Klassenunterteilung mit der Wurzelmethode**⁶²⁹ folgt derselben **Berechnung** (folgende Abb.) wie bei den Wertklassifizierungen (s. den *Aufzählungspkt. „Klassenunterteilung mit Wurzelmethode“* unter [Pkt. 4.3.3.1.](#)). Um jedoch die Vorteile einer logarithmischen⁶³⁰ Klassifizierung (s. [Pkt. 4.3.4.1.](#)) zu nutzen, werden die statistischen Kenngrößen aus den Logarithmen der Daten zur Basis „10“⁶³¹ berechnet. Nach jeder Dateneinschränkung auf dem Tabellenblatt „Daten“ (s. die zweite Abb. zu „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)) wird die automatische Klassifizierung der Verteilungsanalyse umgehend angepasst. Neben der empfohlenen Klassenbildung können die zwei ersten Obergrenzen bzw. die Intervallgröße abgeändert werden (s. den *übernächsten Aufzählungspkt. zur „freien Klassenbildung“*).

statistische Kenngrößen	
logarithmische Kenngrößen (Verteilungsanalyse)	
Anzahl:	705
Mittelwert:	2,900
Standardabweichung:	0,176
relativ hohe Werte (Std.-NV für +1):	3,062
Minimum:	2,145
Hilfsminimum:	
Maximum:	3,509
Hilfsmaximum:	
Vorschlag Intervallanzahl:	27
Vorschlag Schichtungsintervall:	0,060
Vorschlag erste Schichtobergrenze:	2,150
Vorschlag zweite Schichtobergrenze:	2,210

- Mit Blick auf zielführendere Werteklassifizierungen – also bessere Musterergebnisse – versucht die SRP-Vorlage Stochastik extreme Ausreißerwerte ohne engeren Verteilungsbezug zu den anderen untersuchten Daten selbständig zu erkennen und verwendet in dem Fall die Visualisierungsgrenzen „**Hilfsminimum**“ und „**Hilfsmaximum**“; in der Spalte „pot. Ausr.“ auf dem Tabellenblatt werden die potentiellen Ausreißer der linearen Werteklassifizierung markiert, die als Einzelprüffelder interessant sein können, während diejenigen zur Verteilungsanalyse i.R. der Rekonstruktion der betriebsspezifischen Logarithmischen Normalverteilung ausgeschlossen werden (s. den *Auflistungspkt. „visuelles Ergebnis“* weiter unten). Zur Ausreißer-Erkennung werden die Erfahrungen nach *Grubbs*⁶³² (abgewandelt) herangezogen: Datensätze, welche mehr als den 4,4-fa-

⁶²⁹ Die Wurzelmethode berechnet die Klassenanzahl zur Klassifizierung (Schichtung) nach folgender Formel: $\sqrt{(Anzahl\ Datensätze)}$. Das Ergebnis wird in der SRP-Vorlage Stochastik zur besseren Übersichtlichkeit auf max. 30 Intervalle begrenzt.

⁶³⁰ Eine logarithmische Klassifizierung bildet die gleichmäßigen Intervalle mit linear ansteigenden Exponenten wie bspw. mit „10⁰ – 10¹ – 10² – 10³“ (entsprechend „1 – 10 – 100 – 1.000“).

⁶³¹ Eine Tageskasseneinnahme von € 1.000 entspricht somit der Zahl „3“, denn es gilt: $\log_{10} 1.000 = 3$, also $10^3 = 1.000$.

⁶³² *Grubbs*, „Engineering Statistics Handbook“, Kapitel 7.1.6. „What are outliers in the data?“, nachzulesen unter: <https://www.itl.nist.gov/div898/handbook/prc/section1/prc16.htm>.

chen Interquartilsabstand⁶³³ vom 3. Quartil⁶³⁴ nach oben oder vom 1. Quartil nach unten abweichen, gelten als „Ausreißer“. Stellen Extremwerte danach offenbare Ausreißer dar, werden sie durch die Hilfswerte als Visualisierungsgrenzen der Verteilungsanalyse ersetzt; dies wird aus dem logarithmischen Teil der „statistischen Kenngrößen“ auf dem Tabellenblatt „Daten“ ersichtlich, indem die Werte „Minimum“ und „Maximum“ hellgrau unterdrückt sind. Die betreffenden Ausreißer-Datensätze werden dadurch **nicht** aus den Analysen ausgeschlossen; die Anzahl ausgewerteter Datensätze wird unter dem Diagramm auf dem Tabellenblatt „Verteilungsdiagramm“ angegeben (s. den Auflistungspkt. „visuelles Ergebnis“ weiter unten). Sollte dies dennoch gewünscht sein, weil es sich z.B. um nicht mit normalen Tageskasseneinnahmen vergleichbare Tagesveranstaltungen handelt, können sie per Kreuz in der Kästchenspalte auf dem Tabellenblatt „Daten“ abgewählt werden; in dem Fall erscheint auf jedem Auswertungsblatt automatisch der Hinweis „untersuchte Daten: Einzeldatenabwahl“.

- ➔ Die Datenklassifizierung der Verteilungsanalyse kann auf dem Tabellenblatt „Verteilungsanalyse“ auch mit **freier Klassenbildung** durchgeführt werden. Dazu lassen sich die ersten zwei Schichtobergrenzen einzeln oder zusammen in den beiden freien Feldern oberhalb abändern; die eingetragenen Exponenten zur Basis „10“ werden in der Spalte daneben in den Wert übersetzt. Das durch die Eingabe(n) entstehende Intervall

Schichtobergrenze	
lg	absolut
→ 1,950	
1,870	74
1,950	→ 89
2,030	107
2,110	129

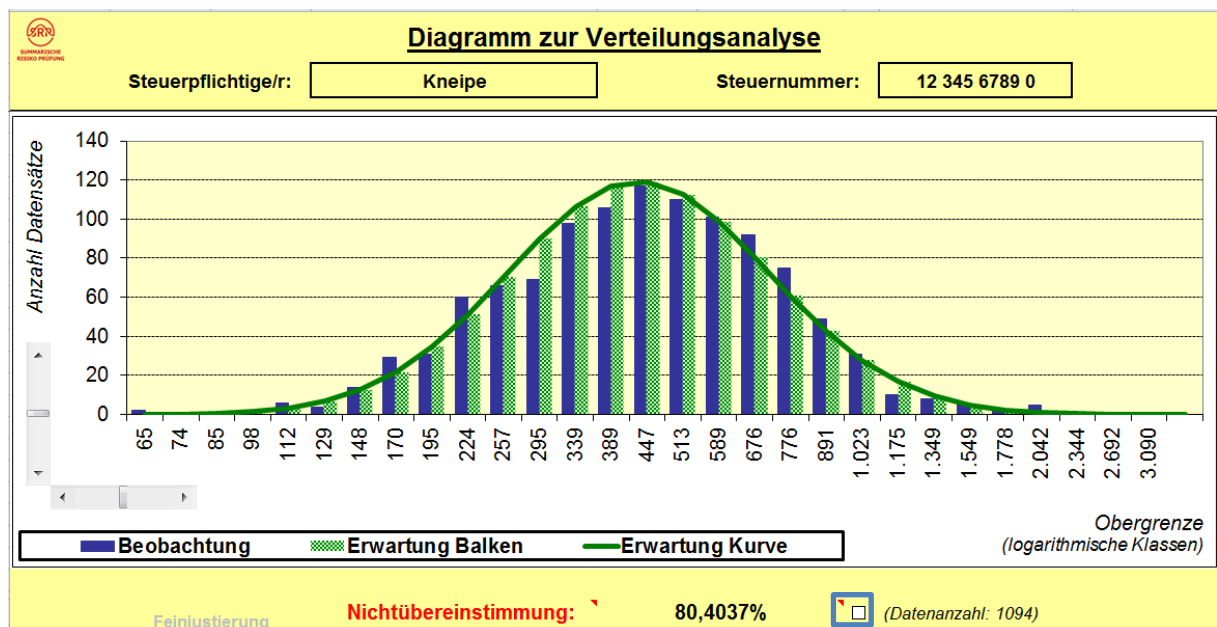
wird automatisch bis zum höchsten Wert der untersuchten Daten bzw. bis zur maximalen Klassenanzahl „30“ fortgesetzt.

- ➔ Falls infolge der (automatischen oder manuell abgeänderten) Klassenfestlegung **nicht alle hohen Spitzenwerte in der Auswertung berücksichtigt** werden können, verfärbt sich das Beobachtungssummenfeld unter der Häufigkeitsverteilung auf dem Tabellenblatt „Verteilungsanalyse“ rot.
- ➔ Zur besseren Übersichtlichkeit wird das **visuelle Ergebnis** der Verteilungsanalyse innerhalb der SRP-Vorlage Stochastik vom Übereinstimmungstest (s. nächster Aufzählungspkt.) getrennt auf dem eigenen Tabellenblatt „Verteilungsdiagramm“ ausgegeben.

⁶³³ Der Interquartilsabstand einer Datenmenge beschreibt den Wertumfang zwischen dem 25 %- und dem 75 %-Quantil (1. und 3. Quartil). Dazu wird die Differenz zwischen dem Wertinhalt des – aufsteigend sortiert – 75 %- und 25 %-größten Rangs ermittelt. Weil Wirtschaftsdaten i.d.R. asymmetrisch-rechtsschief verteilt sind (s. Pkt. 4.3.4.1.), erfolgt in der SRP-Vorlage Stochastik hierzu eine differenzierte Hochrechnung links- und rechtsseitig des Medians.

⁶³⁴ S. die vorherige Fußnote.

Durch die logarithmische⁶³⁵ Werteklassifizierung auf der X-Achse wird der optische Abgleich zur symmetrischen (Punkt-)Erwartung erleichtert; dass es sich um eine logarithmisch gleichmäßige Steigerung der Intervallgröße handelt, kann auf dem Tabellenblatt „Verteilungsanalyse“ nachvollzogen werden. Ebenso dient die doppelte Darstellung der erwarteten Logarithmischen Normalverteilung als Säulen- und Liniengraf⁶³⁶ dem verbesserten optischen Erkennen von Übereinstimmungen und Abweichungen; um die Erwartungsverteilung möglichst betriebsnah zu rekonstruieren, werden automatisch identifizierte offenbare Ausreißer (s. den Auflistungspkt. „Hilfsminimum“ und „Hilfsmaximum“⁶³⁷ weiter oben) nicht in die zu Grunde gelegten Kenngrößen „Mittelwert“ und „Standardabweichung“⁶³⁷ einbezogen, was die Rotfärbung von „Nichtübereinstimmung“ in dieser Abb. anzeigt:



Von der Übereinstimmungsprüfung des Tabellenblatts „Verteilungsanalyse“ wird nur das Nichtübereinstimmungsergebnis des χ^2 -Tests (Chi-Quadrat-Test) übertragen und lässt sich über ein [Anhakfeld](#) bei Bedarf ausblenden.

- ➔ Die qualifizierte **Übereinstimmungsprüfung** der Tageskasseneinnahmen auf die (einzelfallspezifische) Logarithmische Normalverteilung unter Berücksichtigung der untersuchten Datenmenge erfolgt in der SRP bewusst⁶³⁸ mit dem χ^2 -Test (Chi-Quadrat-

⁶³⁵ Eine logarithmische Klassifizierung bildet die gleichmäßigen Intervalle mit linear ansteigenden Exponenten wie bspw. mit „10⁰ – 10¹ – 10² – 10³“ (entsprechend „1 – 10 – 100 – 1.000“).

⁶³⁶ Die Darstellung der (diskreten) logarithmischen Werteklassifizierung als kontinuierlicher Liniengraf ist streng genommen nicht richtig, trägt aber zu einer korrekteren optischen Beurteilung bei.

⁶³⁷ Die Standardabweichung beschreibt die Streuungsintensität einer Datenmenge als vorzeichenneutrale typische Abweichung jedes einzelnen Datensatzes vom Mittelwert.

⁶³⁸ Symmetrie- und Formtests werten die Auslaufzonen mit (sehr) wenigen Elementen im Übereinstimmungsergebnis gleich stark wie den *einzelfalltypischen* Häufungsbereich, obwohl die Auswirkungen einzelner Daten-

Test). Das komplette Zahlenwerk findet sich auf dem Tabellenblatt „Verteilungsanalyse“. Die logarithmische Wertklassifizierung gibt die Schichtobergrenzen als Exponenten zur Basis „10“⁶³⁹ an und übersetzt sie in die „tatsächlichen“ Werte. In den beiden **freien Feldern oberhalb** kann die empfohlene Schichtung manuell abgeändert werden (s. *den Aufzählungspkt. „freie Klassenbildung“ weiter oben*).

Verteilungsanalyse							
Steuerpflichtige/r:		Kneipe		Steuernummer:		12 345 6789 0	
Schichtobergrenze		Erwartung (LogNormVerteilung)					
		kumuliert		schichtweise			
lg		allgemein		angepasst		Beob-	
absolut				Mittelwert Std.-Abw.		achtung	
						χ²-Wert	
						auffällig bzw.	
						sehr auffällig	

(Chi-Quadrat-Test); dabei werden zur Einhaltung der Mindestdatenempfehlung⁶⁴² die χ^2 -Einzelwerte zu Klassen mit einer Erwartung von weniger als 5 Elementen ausgegraut und in die Auswertung unterhalb *nicht* einbezogen, worauf die rote Hervorhebung der Klassenerwartung sowie der Kommentar zum χ^2 -Gesamtwert hinweisen. Ergänzend zur ermittelten **Nichtübereinstimmungswahrscheinlichkeit** wird zur besseren Orientierung ein „**Abweisungswert (99,95 %)**“ angegeben: Derweil im Gegensatz zu den Ziffernanalysen die Klassenanzahl⁶⁴³ bei der Verteilungsanalyse variabel ist, besteht kein starrer kritischer χ^2 -Gesamtwert mit einer bestimmten (Nicht-)Übereinstimmungswahrscheinlichkeit, weshalb der gesondert angegebene Abweisungswert den χ^2 -Gesamtwert benennt, welcher bei der Ziffernanalyse (mit zehn Ziffernmöglichkeiten) dem kritischen Wert „30“ entspricht. Dieser Grenzwert wird als Klassenabweisungswert auf die Anzahl ausgewerteter Intervalle gleichmäßig verteilt, um in der letzten Spalte die besonders kritischen Bereiche der untersuchten Tageskasseneinnahmen zu markieren: Klassen mit einem zu hohen χ^2 -Einzelwert, welche betriebsbezogen zu den hohen⁶⁴⁴ Einnahmen zählen und unterbesetzt sind – wie in der Abb. oberhalb das Intervall „bis € 1.175“.

4.3.4.3. Ergebnisverdeutlichung bei auffälligen Verteilungsanalysen

Um die in Tageskasseneinnahmen wahrscheinlichste, weil kriminalpsychologisch zu erwartende Manipulationsspur „Spitzeneingriffe“ (s. [Pkt. 4.3.4.](#)) deutlich herauszustellen, verfügt die Verteilungsanalyse der SRP-Vorlage Stochastik über eine **gezielte Anpassung der betriebsspezifischen Logarithmischen Normalverteilung**.

Im nachfolgend abgebildeten Beispielsfall eines gastronomischen Betriebs weicht die aus den Kennzahlen „Mittelwert“ und „Standardabweichung“⁶⁴⁵ der untersuchten Tageskasseneinnahmen rekonstruierte (betriebsspezifische) Logarithmische Normalverteilung von dem beobachteten Häufigkeitsmuster mathematisch eindeutig (signifikant) ab. Den zu großen Häufigkeitssäulen in einigen unteren Werteklassen und besonders im Bereich „> € 676 bis € 1.349“ stehen links- und rechtsseitig davon Unterbesetzungen gegenüber; die Rotfärbung der Schrift von „Nichtübereinstimmung“ zeigt an, dass die Vorlage automatisch potentielle Ausreißer entdeckt hat (s.a. den Auflistungspkt. „visuelles Ergebnis“ unter [Pkt. 4.3.4.2.](#)),

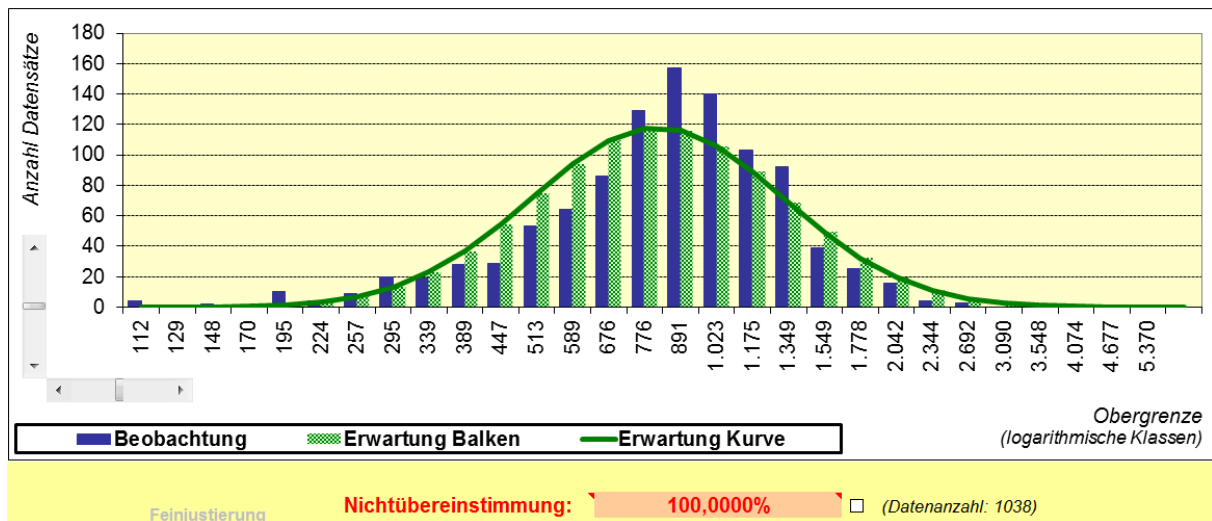
⁶⁴² Kurvenverteilungen, die sich (asymptotisch) der X-Achse annähern, wie Logarithmische Normalverteilungen fallen zwangsweise zu den Seiten hin irgendwann unter die Klassengrenze „ ≥ 5 “. Deshalb wird der χ^2 -Test (Chi-Quadrat-Test) i.R. der Verteilungsanalyse auf die Klassen mit einer Erwartung „ ≥ 5 “ beschränkt.

⁶⁴³ Die Klassenanzahl einer Verteilungsgegenüberstellung findet sich in der Übereinstimmungsprüfung mittels χ^2 -Test (Chi-Quadrat-Test) als erforderliche Information „Freiheitsgrade“ wieder: i.d.R. mit „Klassenanzahl – 1“.

⁶⁴⁴ Die Schwelle ergibt sich mit der Klassenobergrenze, die – unter logarithmischen Bedingungen – größer als „Mittelwert + Standardabweichung“ der untersuchten Tageskasseneinnahmen ist.

⁶⁴⁵ Die Standardabweichung beschreibt die Streuungsintensität einer Datenmenge als vorzeichenneutrale typische Abweichung jedes einzelnen Datensatzes zum Mittelwert.

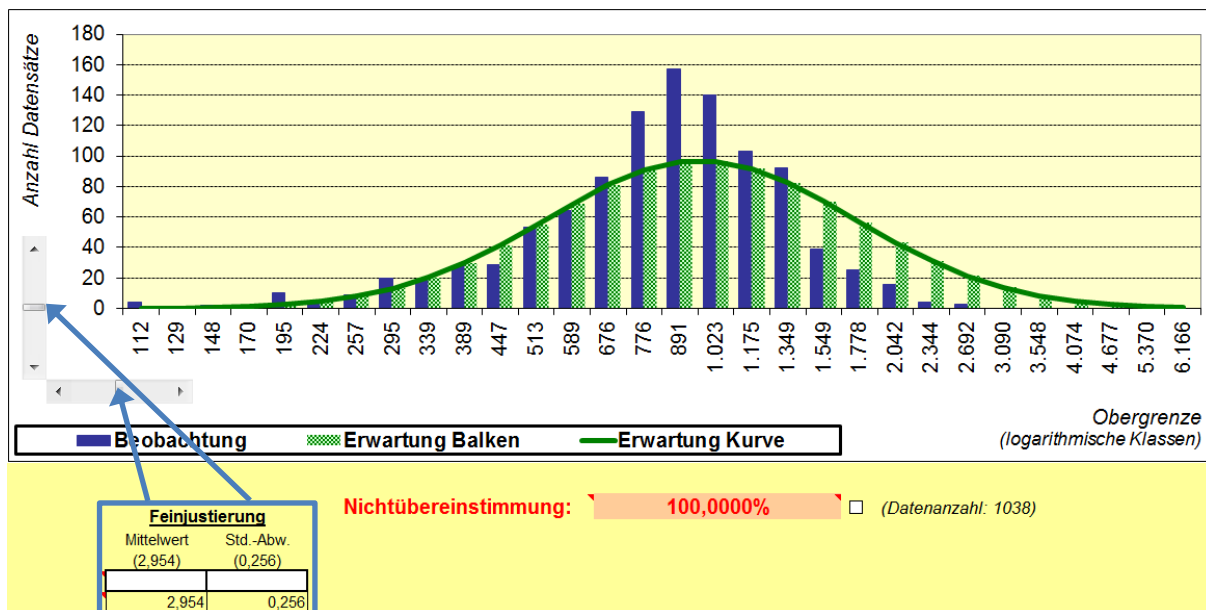
die bei der Rekonstruktion der erwarteten Logarithmischen Normalverteilung ausgeschlossen werden:



Falls es sich bei den festgestellten Unterschieden um stärkere Eingriffe in die betriebsbezogen höheren Einnahmen handelt, ist hiervon auch der Mittelwert betroffen⁶⁴⁶. Somit kann diese abgeleitete Logarithmische Normalverteilung nicht die ursprünglichen Absatzverhältnisse wiederherstellen. Da die Tageskasseneinnahmen in der Anstiegszone „> € 339 bis € 676“ nach kriminalistischen und Prüfungserfahrungen ein deutlich geringeres Manipulationsrisiko als höhere Werte tragen, ohne dabei unrepräsentativ niedrige Ausreißer darzustellen, bietet die SRP-Vorlage Stochastik auf dem Tabellenblatt „Verteilungsanalyse“ die Möglichkeit, per Haken bei „angepasst“ (s. die Abb. zum Aufzählungspkt. „Übereinstimmungsprüfung“ unter [Pkt. 4.3.4.2.](#)) aus einer aussagekräftigen Datenanzahl dieses Anstiegsbereichs⁶⁴⁷ – anstatt aus der Gesamtheit der Daten – die betriebsspezifische Logarithmische Normalverteilung abzuleiten. In der entsprechend angepassten Gegenüberstellung unterhalb ist nun gut zu erkennen, dass diese – weiterhin betriebsintern rekonstruierte – Logarithmische Normalverteilung die weniger hinterziehungsgefährdeten Verhältnisse des Einzelfalls mit den ansteigenden Häufigkeiten im Bereich „> € 339 bis € 676“ bereits weitgehend trifft. Um die Nachfragesituation des geprüften Betriebs aus den Tageskasseneinnahmen so genau wie möglich nachzubilden, können die beiden **Kenngößen „Mittelwert“ und „Standardabweichung“** **manuell feinjustiert** werden; die Ausgangsgrößen werden weiterhin angezeigt, um diesen Zustand zügig wieder herstellen zu können:

⁶⁴⁶ Da bei einer Einnahmenverkürzung die Summe der Einnahmen reduziert wird, sich die Anzahl der Tage aber i.d.R. dabei nicht verringert, wird der Mittelwert als $\frac{\text{Summe}}{\text{Anzahl}}$ infolge der Manipulation ebenfalls verkleinert.

⁶⁴⁷ Mit der Steigungszone zu den Werten „-0,75“ und „-0,5“ aus der Standardnormalverteilung werden ungefähr die Quantile „23%“ und „31 %“ gewählt, um die Verteilungscharakteristik der weniger manipulationsgefährdeten Tageskasseneinnahmen möglichst zutreffend zu erfassen. So verzerren weder mögliche Ausreißer im untersten Wertebereich noch starke Konzentrationen in der bereits betroffenen Haupthäufungszone den hochgerechneten Mittelwert.

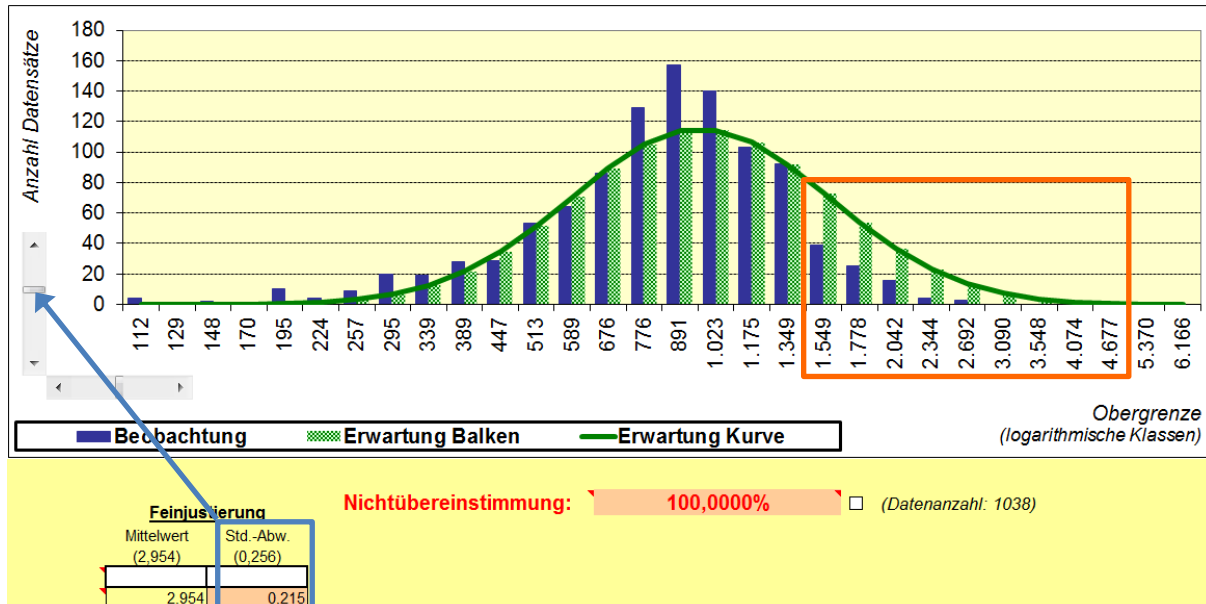


- ➔ Eine Erhöhung des Mittelwerts verschiebt die erwartete Logarithmische Normalverteilung insgesamt formstabil auf der X-Achse nach rechts, wohingegen eine Minderung eine Verschiebung der gesamten Kurvenstruktur nach links bedeutet.
- ➔ Durch eine Vergrößerung der Standardabweichung wird die Logarithmische Normalverteilung breiter auseinandergezogen, sodass die Häufigkeiten in der Mitte ab- und an den Rändern zunehmen. Entsprechend bewirkt eine Absenkung der typischen Abweichung vom Mittelwert eine Stauchung der Kurvenstruktur mit stärkerer Häufigkeitskonzentration in der Mitte.

Die Abänderung der Kenngrößen „Mittelwert“ und „Standardabweichung“ kann komfortabel über die Schieberegler erfolgen: in feinen Stufen über die Pfeile und in gröberen über die Schienen; befindet sich ein Schieberegler außerhalb des Neutralzustands färbt sich die Zelle mit der zugehörigen Kenngröße rot. Außerdem können beide Parameter direkt als Zahlen in die freien Eingabefelder eingetragen werden, *womit die jeweilige Einstellung der Schieberegler überschrieben wird.*

Im Beispielsfall wurde eine **Reduzierung der Standardabweichung über den Schieberegler** vorgenommen (*folgende Abb.*), um die stärkere Steigung der Häufigkeiten in der Zone „> € 389 bis € 676“ besser nachzubilden, also deren Klassensäulen soweit möglich anzugleichen. Auch unter dieser betriebsnahen Rekonstruktion der (Punkt-)Erwartung „Logarithmische Normalverteilung“ an einer größeren Anzahl Datensätze aus den weniger manipulationsgefährdeten Tageskasseneinnahmen bleibt es gemäß χ^2 -Test (Chi-Quadrat-Test) bei einer mathematisch eindeutigen (signifikanten) Nichtübereinstimmung für die auswertbaren Klassen. Diesem Diagramm lässt sich allerdings das manipulationstypische (s. [Pkt. 4.3.4.](#)) Szenario entnehmen: Falls die unteren Besteuerungsdaten richtig sind, müsste die Absatzsituation des Unternehmens nach den Erkenntnissen zur Verteilung von Wirtschaftsdaten

(s. [Pkt. 4.3.4.1.](#)) in den mittleren und hohen Einnahmeregionen der grünen Kurve weiter folgen. Tatsächlich ist eine unerwartet intensive Ballung im Standardbereich „> € 676 bis € 1.023“ auf Kosten aller besonders wertintensiven Intervalle zu beobachten:



Übertragen auf den Verdacht einer Einnahmenverkürzung würde eine derartige Abweichung von der Logarithmischen Normalverteilung resultieren, wenn regelmäßig (stärker) in die betriebsbezogen hohen Tageskasseneinnahmen eingegriffen wurde und dadurch häufig Werte im vertrauten Standardbereich entstanden sind. Im Hinblick auf den Zweck des Verifikationsprinzips⁶⁴⁸ sollte ein derart konkreter Verdacht Eingang in die rechtliche Gesamtwürdigung erhalten.

Im Gegensatz zu Ziffernanalysen, die lediglich einzelne Zahlenpositionen in untersuchten Daten betrachten, lässt sich aus einer auffälligen Verteilungsanalyse mit manipulationspsychologisch typischem Muster auch eine (annäherungsweise) Quantifizierung ableiten, was unter [Pkt. 4.3.4.6.](#) beschrieben wird.

Alle auf dem Tabellenblatt „Daten“ vorgenommenen **generellen Dateneinschränkungen** (s. den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)) wirken auf die angepasste Verteilungsanalyse der SRP-Vorlage Stochastik; ein Hinweis dazu erfolgt auf den Auswertungsblättern „Verteilungsanalyse“, „Verteilungsdiagramm“, „Referenzsumme“ und „Referenzdiagramm“ (s.a. [Pkt. 4.3.4.6.](#)).

⁶⁴⁸ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

4.3.4.4. Andere Ursachen für auffällige Verteilungsanalysen

Folgende **einzelfallbezogene Besonderheiten** bzw. andersartigen Eingriffe können ein Häufigkeitsmuster verursachen, welches in der Verteilungsanalyse ebenfalls mathematisch eindeutig (signifikant) von der (Punkt-)Erwartung abweicht und dabei einem manipulationspsychologisch typischen Spitzeneingriff (s. [Pkt. 4.3.4.](#)) ähneln kann:

1. Beim **freien Erfinden von Besteuerungsdaten** fehlt der wirtschaftliche Entstehungszufall, wie ihn i.d.R. die Nachfrage erzeugt. Insofern werden frei erfundene Tageskasseneinnahmen häufig chaotische Häufigkeitsverteilungen hervorbringen, in denen keine Gleichmäßigkeit erkennbar ist. Die doppelte Gegenüberstellung der betriebspezifischen (Punkt-)Erwartung „Logarithmische Normalverteilung“ als Säulen- und als Liniengraf (s. *den Aufzählungspkt. „visuelles Ergebnis“ unter [Pkt. 4.3.4.2.](#)*) soll – genauso wie die SRP-Prüfungssystematik (s. *den Abschnitt „Prüfungsschema zur Vermeidung von Fehldeutungen“ unter [Pkt. 4.3.](#)*) – bei der Unterscheidung helfen.
2. Falls die Absatzsituation eines Betriebs zeitweise so stark vom Standard abweicht, dass es sich quasi um ein anderes Unternehmen handelt, kann eine **zwei- oder mehrgipflige Häufigkeitsverteilung** (s. *die ersten beiden Abb. unter [Pkt. 4.3.3.3.](#)*) resultieren. Dieser Effekt wird aus den verbildlichten Wertklassifizierungen ersichtlich: Eine Häufigkeitskonzentration stellt das übliche Spektrum an Tageskasseneinnahmen dar, während entweder davor oder in deren Auslauf ein zweiter Verteilungsanstieg mit Häufigkeitsmaximum und -abfall erscheint. *Derartige Zweifachverteilungen werden bspw. häufig vorkommen, wenn – als möglicher Anwendungsfehler der SRP-Verteilungsanalyse – bei Betrieben mit Verzehr an Ort und Stelle sowie Außerhausgeschäften die beiden Tagessummen nicht vorher zusammengefasst sondern einzeln in einer Analyse verarbeitet werden; weil die beiden Absatzanteile in den meisten Fällen nicht dasselbe Standardniveau haben, ergeben sich zwei Häufigkeitsverteilungen, die sich überlagern. Aus dem Grund weist die SRP-Vorlage Stochastik auf dem Tabellenblatt „Daten“ mit Rotfärbung der Überschrift „Datum“ darauf hin, falls mehrere Datensätze mit demselben Datumswert hineinkopiert werden.*
3. Der rechtsschiefe, gleichmäßig sanfte Auslauf in den Häufigkeiten von Tageskasseneinnahmen zu den betriebsbezogen höchsten Werten hin kann durch eine (häufig erreichte) **Kapazitätsgrenze** gestört sein. Sollte es z.B. zum Geschäftsmodell gehören, eine besonders hohe Nachfrage nicht zu bedienen, wird die – unter logarithmischer⁶⁴⁹ Skalierung – glockenförmige Häufigkeitsverteilung ab einem bestimmten Intervall abbrechen.

⁶⁴⁹ Eine logarithmische Skalierung belegt die gleichmäßigen Intervalle mit linear ansteigenden Exponenten wie bspw. mit „ $10^0 - 10^1 - 10^2 - 10^3$ “ (entsprechend „1 – 10 – 100 – 1.000“).

Alle Tage mit noch besseren Einnahmen, die ansonsten im Auslauf verteilt wären, häufen sich folglich in der begrenzenden Klasse. Eine Sonderform an Kapazitätsgrenze ist zudem durch Absatzabhängigkeiten denkbar, wenn z.B. Apotheken eng mit Ärzten zusammenarbeiten und diese an Verschreibungsbeschränkungen gebunden sind.

4.3.4.5. Abklärung von auffälligen Verteilungsanalysen

Bereits die **Beschränkung auf ein praxistypisches Manipulationsszenario** bedeutet gegenüber einer vollständigen⁶⁵⁰ Interpretation des klassifizierten Häufigkeitsmusters von Tageskasseneinnahmen eine deutlich reduzierte Gefahr von fehlerhaften Anwendungen und Schlussfolgerungen. Verstärkte Spitzeneingriffe in die betriebsbezogen höheren Einnahmenwerte (s. [Pkt. 4.3.4.](#)) sind nicht nur die manipulationspsychologisch wahrscheinlichste Spur (digitaler) Einnahmenverkürzungen, sondern können auch am einfachsten und eindeutigsten identifiziert werden.

Eine Abgrenzung auffälliger Verteilungsanalyseergebnisse gegen die ersten zwei [unter Pkt. 4.3.4.4.](#) genannten Alternativursachen wird u.a. durch das Prüfungsschema zur Vermeidung von Fehldeutungen (s. [den Abschnitt „Prüfungsschema zur Vermeidung von Fehldeutungen“ unter Pkt. 4.3.](#)) gewährleistet:

1. Mit Hilfe der stets vorgeschalteten **redundanten**⁶⁵¹ **und variablen SRP-Ziffernanalysen** (s. [Pkt. 4.3.2.](#)) kann ein möglicher Verdacht auf frei erfundene Tageskasseneinnahmen weitgehend⁶⁵² ausgeräumt, wenn auch nicht mathematisch eindeutig (signifikant) ausgeschlossen werden.
2. Durch die **vorbereitende Werteklassifizierung zur Feststellung von Zwei- und Mehrfachverteilungen** (s. [Pkt. 4.3.3.3.](#)) wird dieser praxisrelevanteste⁶⁵³ Stör-

⁶⁵⁰ Vgl. z.B.: Ernst, „Indirekte Methoden der steuerlichen Betriebsprüfung: klassische Verprobung, Systemprüfung, neuere Verfahren“, Schmidt-Verlag, 1986, oder Schumann, „Voraussetzungen und Verfahren der abgabenrechtlichen Schätzung nach § 162 Abs. 1 und 2 AO – insbesondere im Fokus der steuerlichen Außenprüfung“, Diss. an der Johannes Gutenberg-Universität Mainz, 2009, 105 f. mit diversen Abb.

⁶⁵¹ Es werden mit der Einerstelle und Zweitziffer zwei unabhängige (s.a. [den Ausschluss von Werten unter € 100 bei der Einerstellenuntersuchung unter Pkt. 4.3.2.2.1.](#)) Zahlenpositionen gleichzeitig untersucht.

⁶⁵² Eine dahingehende Aussage aus dem BFH-Urteil v. 16.9.2021 (BStBl. II 2022, 101) muss aufgrund des weiten natürlichen Streuverhaltens zufälliger Ziffern- und Zahlenverteilungen verwirren: „So hat den Feststellungen des FG zufolge auch der in der mündlichen Verhandlung vor dem FG anwesende Kassensystemprüfer glaubhaft angegeben, er selbst habe trotz genauer Kenntnis des [Chi-Quadrat-]Tests keine manipulierte Buchführung erstellen können, die den Test bestanden hätte.“

⁶⁵³ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 161: „Bei der Lognormalverteilungsanalyse muss besonders auf die Homogenität des Datenbestandes geachtet werden. Die Verteilung der Tageseinnahmen kann sich je nach betrieblicher Besonderheit bspw. in den einzelnen Geschäftsjahren, saisonal oder wochentagsspezifisch unterscheiden. Dies ist in den Analysen berücksichtigt. Ein solcher Umstand einer ‚uneinheitlichen Absatzsituation‘ wird in der Praxis ebenfalls regelmäßig und sehr tiefgehend geprüft (SRP-Handbuch“.

faktor externer Vergleiche von Tageskasseneinnahmen mit (betriebsspezifischen) Logarithmischen Normalverteilungen im Vorfeld abgeklärt. Falls dabei ein gleichmäßiges, aber zwei- oder mehrgipfliges Häufigkeitsmuster beobachtet wird, stellt die SRP-Vorlage Stochastik zielführende Funktionen zur Verfügung, um die uneinheitlichen Teildatenbestände zu identifizieren (s. *das Beispiel unter [Pkt. 4.3.3.3](#)*). Anschließend können diese jeweils in einer eigenen SRP-Vorlage Stochastik separat untersucht werden.

Der letzte *unter [Pkt. 4.3.4.4](#)* genannte praxisrelevante Störeinfluss gegen die Übereinstimmung von Tageskasseneinnahmen mit der (betriebsspezifischen) Logarithmischen Normalverteilung „Kapazitätsgrenze“ ist nach ständigen Prüfungserfahrungen eher selten und lässt sich anhand der **grafischen Auswertung zur Verteilungsanalyse** auf dem Tabellenblatt „Verteilungsdiagramm“ gut nachvollziehen: Bricht das beobachtete Häufigkeitsmuster in der höchsten Werteklasse konsequent ab, ist sie als Ursache nicht auszuschließen, i.d.R. sind aber vereinzelte Tage in Intervallen mit noch höheren Einnahmen wie in der letzten Abb. *unter [Pkt. 4.3.4.3](#)* zu finden. Für diese problemlos herausfilterbaren Tage müssten abweichende Bedingungen wie bspw. erweiterte Öffnungszeiten, Veranstaltungs-Pauschalpreise oder ordnungswidrige Zusammenfassungen von mehreren Tagen vorliegen. Sollte ein bestimmter Wochentag mit andersartiger Geschäftsgrundlage als mögliche Ursache in Betracht kommen, kann er auf dem Tabellenblatt „Daten“ komfortabel ausgeschlossen werden (s. *den Aufzählungspkt. „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1](#)*), was zum Wegfall des Überhangs jenseits der Kapazitätsgrenze führen müsste. Insofern kann auch dieser nicht-manipulative Sonderfaktor eines Einzelbetriebs *ohne* größeren Aufwand⁶⁵⁴ konkret abgeklärt werden.

Die Beweiskraft von mathematisch eindeutig (signifikant) auffälligen Verteilungsanalysen wird wegen der vergleichbaren Voraussetzungen aller manipulationsorientierten Musteranalysen in einem gemeinsamen Kapitel *unter [Pkt. 4.3.5](#)* behandelt. Die derzeitige Beschränkung für stochastisch abgesicherte auffällige Verteilungsmuster auf den Status eines „Zu-

⁶⁵⁴ Gelegentlich werden der SRP in der Literatur „überzogene Mitwirkungspflichten“ vorgeworfen (vgl. z.B.: *Steinhauff*, „Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388, 389). Tatsächlich ist es durch die digitalen Möglichkeiten überhaupt erst möglich geworden, die unvermeidbaren Unsicherheiten von Verprobungen zügig auf das Minimalmaß zu reduzieren, welches durch die Ordnungsmäßigkeit der Aufzeichnungen vorgeben ist. Interaktive Funktionen wie in den SRP-Vorlagen verbessern dabei Komfort, Effektivität und Effizienz – auch für die Stpfl. in Form einer zielführenderen Einforderung der Mitwirkung. Dass Prüfungsmethoden z.T. auch zur unrechtmäßigen Verlagerung von Aufklärungspflichten eingesetzt werden können, ist kein Makel der SRP, sondern trifft jede und ist deshalb als Abwertungskriterium ungeeignet. (Technische) Möglichkeiten zur Einschränkung der unvermeidbaren Unsicherheiten bei der Aufarbeitung komplexer Prüffelder sind deshalb zeitgemäße Hilfsmittel zur Umsetzung der gesetzlichen Pflicht, „alle für den Einzelfall bedeutsamen [...] Umstände zu berücksichtigen“ (§ 88 Abs. 1 S. 2 AO).

satzindizes⁶⁵⁵ wie bei den Ziffernanalysen lässt sich zumindest tlw. durch fehlerhafte Methodenbeschreibungen⁶⁵⁶ und resultierende Zweifel an der korrekten Anwendung⁶⁵⁷ erklären.

4.3.4.6. Quantifizierung von auffälligen Verteilungsanalysen

Auf der Grundlage einer betriebspezifisch rekonstruierten Logarithmischen Normalverteilung (s. [Pkt. 4.3.4.1.](#)) als Erwartungsideal für untersuchte Tageskasseneinnahmen lässt sich i.R. der Verteilungsanalyse eine **Gesamtsumme bzw. -abweichung hochrechnen** – im Gegensatz zu Ziffernanalysen, bei denen einzelne Zahlenpositionen auf Manipulationsspuren untersucht werden. Als (möglichst) betriebsnah nachgebildete Absatzerwartung ist dieser Abweichungswert *in geeigneten Fällen* (s. [Pkt. 4.3.4.3.](#)) vergleichbar mit dem „Most Likely Error (MLE)“ einer mathematisch-statistischen Stichprobenprüfung (s. *den Aufzählungspkt. „statistische Stichprobenverfahren“ im Abschnitt „wichtige bestehende Erkenntnisse zur Datenanalyse“ unter Pkt. 2.*): Zwar ist die Wahrscheinlichkeit sehr gering, dass der Grundgesamtheitsfehler *exakt* dem MLE entspricht, dennoch ist sie für kein anderes Einzelergebnis höher – folglich entspricht der MLE einer vorgenommenen Stichprobenprüfung dem wahrscheinlichsten Gesamtfehler als Punktschätzung.

In der SRP-Vorlage Stochastik kann auf dem Tabellenblatt „Referenzsumme“ die (annäherungsweise) Berechnung des logarithmisch-normalverteilten Sollwerts für die untersuchten Tageskasseneinnahmen *per Haken ausgelöst* werden; *auch diese Funktion erfordert eine bewusste Aktivierung, weil sie nur dann anzuwenden ist, wenn das typische Manipulationsbild „Spitzeneingriff“ (s. [Pkt. 4.3.4.3.](#)) erkennbar ist und andere – besonders nicht-manipulative – Ursachen soweit möglich ausgeschlossen wurden (s. [Pkt. 4.3.4.5.](#)):*

⁶⁵⁵ Tipke/Kruse, AO/FGO, [171. Lieferung 7/2022](#), § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit (s. FG Münster v. 14.8.2003 – 8 V 2651/03 E, U, EFG 2004, 9 m. Anm. Trossen; FG Münster v. 10.11.2003 – 6 V 4562/03 E, U, EFG 2004, 236 [237]; FG Münster v. 7.12.2005 – 1 K 6384/03 E, EFG 2006, 652 [653] m. Anm. Zimmermann [Fahrtenbuch]; FG Rh.-Pf. v. 24.8.2011 – 2 K 1277/10, EFG 2012, 10 [11] m. Anm. Kühnen; Sosna, St-Bp. 2000, 41 [Teil I], 68 [Teil II]; Meyer, DStR 2005, 2114; Watrin/Stuffert, DB 2006, 1748; Huber, StBp. 2009, 121 [124 ff.]“.

⁶⁵⁶ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 136: „In der derzeitigen Praxis der Finanzämter hat sich der Chi-Quadrat-Anpassungstest als Überprüfungsmethode der Verteilungsannahmen bei der Ziffernanalyse etabliert. Dies hat u.a. dazu geführt, dass die Ziffernanalyse anhand des Benfordschen Gesetzes, insbesondere in der Praktiker-Literatur, mitunter fälschlicherweise grobschlächtig als ‚Chi-Quadrat-Test‘ bezeichnet und somit als eigenständige Methode, und nicht als Analyseinstrument, aufgefasst wird. Hier zeigt sich ein häufig beobachtetes unzureichendes Verständnis der Methodik und der statistischen Testverfahren.“

⁶⁵⁷ Kohlmann, Steuerstrafrecht, [72. Lieferung 9.2021](#), § 397 AO, Rz 12.3: „Die praktische Verwertbarkeit der statistischen Verfahren ist durchaus umstritten, und zwar bereits als Grundlage für steuerliche Schätzungen, aber erst recht für strafrechtliche Verurteilungen. Die Finanzgerichte lehnen bislang den Chi-Quadrat-Test als eine sichere Methode, die Ordnungsgemäßheit der Buchführung zu verwerfen, ab. Allerdings wird weniger die grundsätzliche Eignung der Analysen angezweifelt als vielmehr hinterfragt, unter welchen Voraussetzungen und mit welcher Anzahl von Ausgangswerten die Methode aussagekräftige Ergebnisse liefert.“

Einnahmensumme der angepassten Referenzverteilung					
☒	Steuerpflichtige/r:	Restaurant		Steuernummer:	12 345 6789 0
Schichtobergrenze	Referenzhäufigkeit (gerundet)			Beobachtung	Betrag (exakte Berechnung)
	absolut	kumuliert	schichtweise		
	200	1	1,2	18	60,60
	270	8	6,5	20	907,45
	340	25	17,8	32	3.700,97
	410	58	32,8	32	9.476,56
	480	106	47,9	48	17.961,44
	5.100	1.037	0,0	0	130,85
	5.170	1.037	0,0	0	118,57
	5.240	1.037	0,0	0	107,50
	5.310	1.037	0,0	0	97,52
	5.380	1.037	0,0	0	88,52
	optionales Maximum:				
	Rest	1.037	0,1	0	778,83
Bemerkung(en):				1.037	1.037
				1.086.317,79	
				erklärt / vorangemeldet:	882.061,93
				Abweichung:	204.255,86
				untersuchte Daten: Einzeldatenabwahl	

Anhand der Klassenmittelwerte und der ebenfalls gemittelten⁶⁵⁸ Häufigkeitserwartungen zu den Intervallen wird der Gesamtwertinhalt der betriebsspezifisch rekonstruierten Logarithmischen Normalverteilung nachvollziehbar im Annäherungsverfahren hochgerechnet; dass für die hohen Werteklassen trotz der Erwartung „0,0“ dabei geschätzte Einnahmen ermittelt werden, liegt an der Rundung der „schichtweisen Referenzhäufigkeit“ auf eine Stelle nach dem Komma, während für die Berechnung unter „Betrag“ die vollständigen Zahlen herangezogen werden. Rot umrandet wird abschließend die Differenz zur Summe der untersuchten Tageskasseneinnahmen ausgegeben.

Folgende Hinweise sind für die Anwendung der Funktion „Referenzsumme“ in der SRP-Vorlage Stochastik darüber hinaus wichtig:

- ➔ Alle auf dem Tabellenblatt „Daten“ vorgenommenen **generellen Dateneinschränkungen** (s. den Abschnitt „zweistufige Eingrenzungsmöglichkeit“ unter [Pkt. 4.3.2.1.](#)) wirken auf die Berechnung der erwarteten Gesamteinnahmen gemäß betriebsspezifischer Logarithmischer Normalverteilung; die **rot gefärbte Überschrift und der Hinweis „untersuchte Daten: Einzeldatenabwahl“** weisen in der beispielhaften Abb. oberhalb darauf hin, dass ein oder mehrere Datensätze zuvor abgewählt wurden, weil sie sich bspw.

⁶⁵⁸ Um die unterschiedlichen asymmetrischen Steigungen in den Abschnitten der Logarithmischen Normalverteilung zu berücksichtigen, wird der mittlere Wertinhalt jeder Schicht mit der gemittelten Häufigkeit aus dem betreffenden und dem vorherigen Intervall multipliziert.

- als außergewöhnliche Tageskasseneinnahmen pauschal abgerechneter Veranstaltungen herausgestellt haben und somit nicht vergleichbar sind.
- ➔ Die Referenzsumme arbeitet immer auf Basis der *angepassten* Logarithmischen Normalverteilung inklusive eventueller manueller Feinjustierung (s. [Pkt. 4.3.4.3.](#)), denn von einem Spitzeneingriff ist der Mittelwert der untersuchten Tageskasseneinnahmen ebenfalls betroffen⁶⁵⁹, sodass eine Hochrechnung ohne Ausrichtung der (Punkt-)Erwartung an diejenigen Daten mit einem geringeren Verkürzungsrisiko keinen Sinn ergeben würde.
 - ➔ Weil rechtsschiefe Logarithmische Normalverteilungen weit nach rechts zu den hohen Werten auslaufen, besteht die Möglichkeit, eine **Wertobergrenze für die Hochrechnung** festzulegen. In dem Fall werden alle noch nicht „verbrauchten“ Tage mit dem Maximalwert angesetzt; sollten Tageskasseneinnahmen tatsächlich oberhalb des optionalen Maximums liegen, werden diese nicht nach unten korrigiert.

Die Funktion „Referenzsumme“ soll primär dazu dienen, die großen Auswirkungen von Spitzeneingriffen zu verdeutlichen. Für eine **Schätzung** im Besteuerungsverfahren (s. [Pkt. 4.8.](#)) ist diese Hochrechnung i.d.R. nur bedingt geeignet, weil „[d]ie Schätzungsungewissheit [...] nicht dazu führen [darf], nur den Betrag anzunehmen, der auch im ungünstigsten Fall als sicher vereinnahmt angesehen werden kann“⁶⁶⁰. Vielmehr „ist das FA im allgemeinen nicht nur berechtigt, sondern verpflichtet, in dem gegebenen Schätzungsrahmen an die oberste Grenze zu gehen“⁶⁶¹, damit es „keinen Anreiz [gibt], sich weiterhin schätzen zu lassen“⁶⁶². Dennoch kann diese Auswertung in geeigneten Fällen – speziell auch i.R. von Strafverfahren (s. [Pkt. 4.9.](#)) – zur Ermittlung einer sachgerechten Schätzungshöhe beitragen.

Das Tabellenblatt „Referenzdiagramm“ der SRP-Vorlage Stochastik soll eine weitere Absicherung anbieten: Unter *linearer* Wertklassifizierung auf der X-Achse kann kontrolliert werden, ob die betriebsnah abgeleitete Logarithmische Normalverteilung den wichtigen Bereich mit stark ansteigenden Häufigkeiten der niedrigeren Tageskasseneinnahmen passend nachbildet, um daraus das gesamte Absatzverhalten hochzurechnen. Im nachfolgenden Beispielsdiagramm werden die Verhältnisse in der Zone „> € 340 bis € 690“⁶⁶³ optimal getroffen; der Hinweis auf vorgenommene Einschränkungen der Daten wird auch in dieser Auswertung ausgegeben.

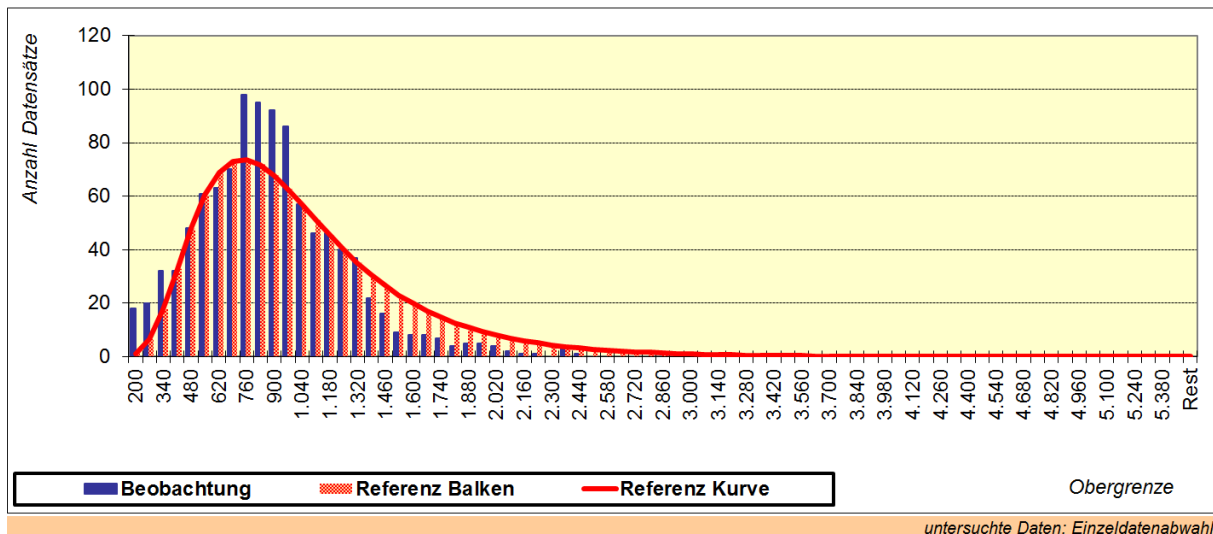
⁶⁵⁹ Da bei einer Einnahmenverkürzung die Summe der Einnahmen reduziert wird, sich die Anzahl der Tage aber i.d.R. dabei nicht verringert, wird der Mittelwert als $\frac{\text{Summe}}{\text{Anzahl}}$ infolge der Manipulation ebenfalls verkleinert.

⁶⁶⁰ Tipke/Kruse, AO/FGO, 170. Lieferung 5/2022, § 162 AO, Rz. 44.

⁶⁶¹ BFH v. 9.3.1967, BStBl. III 1967, 349, Leitsatz.

⁶⁶² FG Berlin-Brandenburg v. 30.6.2017 – 10 V 10219/16.

⁶⁶³ Die X-Achsenbeschriftung benennt nur die Obergrenzen jedes zweiten Intervalls.



4.3.5. Beweiswert von (musterbasierten) Häufigkeitsanalysen

Im Gegensatz zum effektiven Erkennen und systematischen Eingrenzen von Prüffeldern⁶⁶⁴ (s. das Beispiel im Bereich „systematisch-mehrperspektivische Musteranalytik“ unter Pkt. 5.1.1.) sollen (Gesamtbild-)Verprobungen wie Musteranalysen i.S. des § 158 Abs. 2 Nr. 1 AO die Vergangenheit anhand von Unterlagen, Auskünften und Eindrücken zur Beurteilung der Schätzungsnotwendigkeit nach § 162 Abs. 2 S. 2 AO wiederherstellen.⁶⁶⁵ Da Informationen zur Vergangenheit niemals vollständig und objektiv sein können, ergeben sich grundsätzlich **Sachverhaltsbeschreibungen mit Unsicherheiten**. Beim typischen Einsatzgebiet von Verprobungen „Vollständigkeit von (Bar-)Einnahmen“ sind diese Unsicherheiten besonders groß, weil aufgrund der Komplexität der entscheidungsrelevanten Gesamtsituationen keine einfachen Ursache-Wirkungs-Beziehungen bestehen⁶⁶⁶. „Schlussfolgernd kann bei der Prüfung von (steuerlichen) Daten keine (100 %) Sicherheit und somit kein Beweis [...] erfolgen. [...] Nach dem logischen Menschenverstand [kann sie] nur zu Wahrscheinlichkeitsaussagen führen.“⁶⁶⁷

Um dennoch in Gerichtsverfahren und Prüfungen Urteile bzw. ausreichend sichere Beurteilungen zu ermöglichen, werden an die volle richterliche Überzeugung geringere

⁶⁶⁴ BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1: „Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich.“

⁶⁶⁵ Anwendungserlass zu § 158 AO: „Die Vorschrift enthält eine gesetzliche Vermutung. Sie verliert ihre Wirksamkeit mit der Folge der Schätzungsnotwendigkeit nach § 162 AO, wenn es nach [einer] Verprobung [...] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt.“

⁶⁶⁶ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 15: „Unser Dilemma im Umgang mit der Komplexität unserer Welt lässt sich darauf zurückführen, dass wir wohl darin ausgebildet wurden einfache logische Schlüsse zu ziehen und nahe liegende Ursache-Wirkungs-Beziehungen zu definieren. [...] Simple Ursache-Wirkungs-Beziehungen gibt es nur in der Theorie, nicht in der Wirklichkeit.“

⁶⁶⁷ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324 u. 323.

Anforderungen⁶⁶⁸ als an den naturwissenschaftlichen Beweis gestellt.⁶⁶⁹ Für den Bereich des Besteuerungsverfahrens sieht das Gesetz mit **§ 158 Abs. 2 Nr. 1 AO (Verifikationsprinzip)** auf Seiten der Finanzverwaltung eine *darüber hinausgehende* Absenkung der erforderlichen Aussagegewissheit vor (s. [Pkt. 4.6.](#)), denn andernfalls ließe sich die (annähernde) Gleichbehandlung aller Stpfl. gem. Art. 3 GG in einem Massenverfahren nicht durchsetzen⁶⁷⁰. In den allgemein titulierten „Verwaltungsgrundsätze[n] 2020“ wird dieses reduzierte Beweismaß für das ähnlich komplexe – und damit vergleichbar unsichere – Verprobungsgebiet „Fremdüblichkeit von Verrechnungspreisen“ wie folgt konkretisiert: „Voraussetzung einer Einkünfteberichtigung ist vielmehr, [...] dass der von der Finanzbehörde ermittelte Verrechnungspreis zumindest wahrscheinlicher ist.“⁶⁷¹

Die vom BFH⁶⁷² zur Schätzung – und der Argumentation folgend ebenso zur Verprobung – bevorzugten **herkömmlichen Methoden** „Vermögenszuwachs- oder Geldverkehrsrechnung“ und „Kalkulation“ setzen selbst bei ordnungsmäßigen Besteuerungsunterlagen auf unvollständigen Ausgangsinformationen auf, indem bspw. private Geldzu- und -abflüsse bzw. der Warenverderb nicht aufgezeichnet werden müssen. Zudem ist das Risiko ordnungswidriger Unterlagen in den betreffenden Fällen mit überwiegend Bareinnahmen besonders hoch.⁶⁷³ Das Modell „Schwarzeinnahmen zur Bezahlung von Schwarzlöhnen“ kann unter diesen Umständen mit den Mitteln einer Betriebsprüfung i.d.R. nicht ausgeschlossen werden. I.R. von Kalkulationen liegen häufig keine vollständigen, vertrauens-

⁶⁶⁸ Vgl. z.B.: BFH v. 20.5.2010, BStBl II 2010, 1022, Rz. 13: „[...] wobei der Richter nicht eine von allen Zweifeln freie Überzeugung anstreben darf, sich in tatsächlich zweifelhaften Fällen vielmehr mit einem für das praktische Leben brauchbaren Grad von Gewissheit überzeugen muss“.

⁶⁶⁹ Tlw. werden die unvermeidbaren Unsicherheiten bei der Reproduktion der Vergangenheit in der Literatur offenbar grundsätzlich nicht akzeptiert (vgl. z.B.: Prinz, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4, M5: „Wahrscheinlichkeitstheoretisch geformte Algorithmen können Steueransprüche nicht begründen“), wodurch eine (teleologisch) sachgerechte Einordnung des Beweises im juristischen Sinn und des „Beanstandungsanlasses“ gem. § 158 Abs. 2 Nr. 1 AO unmöglich wird.

⁶⁷⁰ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

⁶⁷¹ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 73.

⁶⁷² BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 63 u. 64 : „Allein die Ergebnisse eines Zeitreihenvergleichs lassen aufgrund der dieser Verprobungsmethode innewohnenden methodenbedingten Unsicherheiten aber noch keinen sicheren Schluss auf das Vorliegen und den Umfang auch materieller Unrichtigkeiten der Buchführung zu. In diesen Fällen sind andere Schätzungsmethoden, die auf betriebsinternen Daten aufbauen oder in anderer Weise die individuellen Verhältnisse des jeweiligen Steuerpflichtigen berücksichtigen (z.B. Vermögenszuwachs- oder Geldverkehrsrechnung, Aufschlagkalkulation) grundsätzlich vorrangig heranzuziehen.“

⁶⁷³ Teutemacher, „Handbuch zur Kassenführung – Praxishandbuch für die rechtssichere Umsetzung“, NWB-Verlag, 2015, 2: „Betriebsprüfungen bei Betrieben, die überwiegend Bargeschäfte tätigen, führen in 95 % der Fälle zu dem Ergebnis, dass die Kassenführung verworfen wird, weil die vorgelegten Bücher bzw. Aufzeichnungen sowie Daten nicht den Ordnungsvorschriften entsprechen und damit die sachliche Richtigkeit in Zweifel gezogen wird“ mit Verweis auf Kalischke, „Betriebe bangen vor Steuerprüfern – Beamte werfen immer mehr Kassenbuchführungen“, Westfälische Nachrichten v. 15.3.2013.

würdigen Informationen zu Verkaufspreisen, Sonderverkäufen, Wareneinkäufen etc. vor. Die entscheidungsrelevanten Größen „Geldverbrauch“ bzw. „Sollumsatz“ müssen in der Konsequenz aus unsicheren Zahlen rekonstruiert werden, sodass die berechneten Zielgrößen sicherer wirken, als sie es sein können⁶⁷⁴: „Die herkömmlichen Methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...]. Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten in den wiederhergestellten Daten.“⁶⁷⁵

Durch **sorgfältige Ermittlungen der Finanzbehörde**⁶⁷⁶ sollen die Unsicherheiten soweit möglich begrenzt werden. Nach dem Verifikationsprinzip haben aber auch die Mitwirkungsbereitschaft und -qualität der Stpfl. eine wesentliche Auswirkung auf das erforderliche Beweismaß für das Auslösen der Schätzungspflicht⁶⁷⁷; diesbezügliche Warnungen vor

⁶⁷⁴ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Nomos, 2018, 56-57 u. 65: „Bei der Verfügbarkeitsheuristik schätzen Individuen die [...] Wahrscheinlichkeit für das Eintreten eines Ereignisses unter Unsicherheit danach ein, wie leicht bzw. schwer es ihnen fällt, damit verbundene Erscheinungen aus ihrem Gedächtnis abzurufen. Dabei werden Ereignisse für wahrscheinlicher erachtet, wenn sie sich leichter vorgestellt werden können und sich zu ihnen eine direkte Verfügbarkeit herstellen lässt. [...] Im Rahmen juristischer Entscheidungen scheint es unerlässlich, von diesen Entscheidungsverzerrungen Kenntnis zu haben, da eine verzerrte Entscheidung den Denkgesetzen der Logik widerspricht [...]“.

⁶⁷⁵ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324.

⁶⁷⁶ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Der Wortlaut des § 158 („Anlass zu beanstanden“) mag dafür sprechen, dass bereits bloße Beweisanzeichen, Zweifel oder Bedenken ausreichen, um die Vermutung auszuräumen oder zu entkräften. Es ist jedoch nicht so, dass die FinBeh. nur irgendwelche Beanstandungen auszusprechen hätte und der Stpfl. dann gehalten wäre zu beweisen, dass die formell richtigen Buchungen sachlich nicht falsch sind. Ein solcher Negativbeweis kann unmöglich verlangt werden. Wenn am Anfang einer Untersuchung Bedenken, Zweifel oder Beweisanzeichen gegen die sachliche Richtigkeit der Buchführung vorliegen, legitimiert (§ 158) und verpflichtet (§ 88) das die Behörde dazu, weitere Nachforschungen zur Sachverhaltsaufklärung anzustellen.“

⁶⁷⁷ BMF-Schreiben v. 3.12.2020 „Verwaltungsgrundsätze 2020“, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. [...] Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u.a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 9: „§ 162 II i.V.m. § 158 orientieren das Beweismaß an der Sphärenverantwortlichkeit [...] Das Beweismaß hängt mithin davon ab, wie viel und in welcher Form der Stpfl. aus seiner Verantwor-

„überzogenen Mitwirkungspflichten“⁶⁷⁸ laufen dem gesetzlichen Auftrag, „alle für den Einzelfall bedeutsamen [...] Umstände zu berücksichtigen“⁶⁷⁹ zuwider und verkennen, dass das gemeinsame interaktive (s. [Pkt. 4.3.2.2.3.](#), [4.3.2.3.3.](#) u. [4.3.4.5.](#)) Ergründen von Musterstörungen mit der SRP-Vorlage Stochastik erfahrungsgemäß wesentlich weniger aufwendig als bspw. das Nachvollziehen einer ausreichend detaillierten Ausbeutekalkulation ist. Bezogen auf manipulationsorientierte Musteranalysen bedeutet das sorgfältige⁶⁸⁰ Erarbeiten einer indiziellen Unschlüssigkeit, dass einem ausgezählten Häufigkeitsmuster (i.d.R. aus Tageskasseneinnahmen) eine *belastbare, betriebsspezifische (Punkt-)Erwartung* gegenübergestellt und die Übereinstimmung unter Berücksichtigung des vorliegenden Datenumfangs mit einem *geeigneten Wahrscheinlichkeitstest* beurteilt wird sowie bei Auffälligkeiten *denkbare nicht-manipulative Ursachen* weitest möglich abgeklärt werden.

Im Hinblick auf die genannten Kriterien eines sorgfältig erarbeiteten – und damit aussagekräftigen – Schlüssigkeitsbilds gründet sich die **manipulationsorientierte SRP-Musteranalytik** auf folgende **Analysegrundsätze und Annahmen**.⁶⁸¹

- ☑ Für **(Punkt-)Erwartungen** wie „Gleichverteilung“, „Benford-Gesetz“ und „Logarithmische Normalverteilung“ „ist eine statistisch signifikante Korrelation zwischen ‚Auffälligkeit‘ und ‚Zuverlässigkeit‘ der Datensätze (noch) nicht zwingend nachzuweisen“⁶⁸², weil zu den erforderlichen *realen* Testdaten „[nicht] garantiert [werden kann], einen ‚manipulationsfreien‘ Datensatz vorliegen zu haben.“⁶⁸³ Dennoch wird diesen Verteilungsmustern auch nach jüngeren wissenschaftlichen Untersuchungen eine *belastbare* Allgemeinbeschreibung von Wirtschaftsdaten zugetraut (s.a. [Pkt. 4.3.2.2.1.](#), [4.3.2.3.1.](#) u. [4.3.4.1.](#)): „Sowohl Zweit-, Dritt- und Viertziffer sowie die Einerstelle als erste Stelle vor dem Komma zeigten

tungssphäre zur Sachaufklärung beiträgt und wieweit die FinBeh. auf dessen Mitwirkung im konkreten Einzelfall angewiesen ist.“

⁶⁷⁸ Vgl. z.B.: *Steinhauff*, „Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 2019, 388, 389.

⁶⁷⁹ § 88 Abs. 1 S. 2 AO.

⁶⁸⁰ Kohlmann, *Steuerstrafrecht*, **72. Lieferung 9.2021**, § 397 AO, Rz 12.3: „Die praktische Verwertbarkeit der statistischen Verfahren ist durchaus umstritten, und zwar bereits als Grundlage für steuerliche Schätzungen, aber erst recht für strafrechtliche Verurteilungen. Die Finanzgerichte lehnen bislang den Chi-Quadrat-Test als eine sichere Methode, die Ordnungsgemäßheit der Buchführung zu verwerfen, ab. Allerdings wird weniger die grundsätzliche Eignung der Analysen angezweifelt als vielmehr hinterfragt, unter welchen Voraussetzungen und mit welcher Anzahl von Ausgangswerten die Methode aussagekräftige Ergebnisse liefert.“

⁶⁸¹ In der kritisierenden Literatur werden die Absicherungsmerkmale der SRP-Musteranalytik, die besonders die Übertragbarkeitsgrenzen von Verteilungserwartungen berücksichtigen, üblicherweise nicht erwähnt (vgl. z.B.: *Beyer*, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, oder *Bleschick*, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353).

⁶⁸² *Pitz/Klapdor/Gardian/Haselau*, „Das Benfordsche Gesetz als Werkzeug zur Aufdeckung von Umsatzmanipulationen in der Gastronomie?“, *StuW* 3/2021, 232, 245.

⁶⁸³ *Sell*, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 158.

insgesamt eine sehr gute Konformität mit der Benford-Verteilung [also inklusive Gleichverteilung für endständige Zahlenpositionen]. Diese Beobachtungen lassen darauf schließen, dass die Annahme der Benford-Verteilung als Verteilungsideal ordnungsmäßiger Daten als durchaus plausibel erscheint. [...] Die Ergebnisse der Gastronomiebetriebe deuten an, dass es bei Gastronomiebetrieben zumindest nicht unplausibel erscheint, von lognormalverteilten Tageseinnahmen als Verteilungsideal auszugehen. [...] Auch wenn die den Methoden zugrundeliegenden Verteilungsannahmen nicht auf mathematischer Sicherheit aufbauen, deuten die exemplarischen Darstellungen an, dass ihr Einsatz durchaus gerechtfertigt sein könnte.“⁶⁸⁴ I.R. der SRP werden diese Erwartungsmuster *einzelfallbezogen* an den Verhältnissen der zu untersuchenden Daten entwickelt (s. [Pkt. 4.3.1.2.](#)); dabei löst die SRP-Vorlage Stochastik das Problem widerstrebender Erwartungen gemäß Gleichverteilung versus Benford-Gesetz für die Einerstelle in Werten, die kleiner als „€ 100“ sind, automatisch (s. [Pkt. 4.3.2.2.1.](#)). Die Analyseschritte „(Punkt-)Erwartung“ und „Übereinstimmungsprüfung“ werden nicht vermengt⁶⁸⁵, sondern für eine nachvollziehbare Ergebnisbeurteilung konsequent getrennt.

- ☑ Bei der Suche nach Manipulationsspuren durch die SRP erfolgt die Übereinstimmungsprüfung von beobachteten Häufigkeitsmustern mit betriebspezifischen (Punkt-)Erwartungen immer unter Einsatz eines *geeigneten Wahrscheinlichkeitstests*: für mehrklassige⁶⁸⁶ Ziffernanalysen und die Verteilungsanalyse mit dem χ^2 -Test (Chi-Quadrat-Test, s. unter [Pkt. 4.3.2.1.](#) u. [Pkt. 4.3.4.2.](#); zur – aus der Wissenschaft bestätigten – besseren Eignung des Chi-Quadrat- gegenüber dem Kolmogorov-Smirnov-Anpassungstest s. den

⁶⁸⁴ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 181 f., 182 u. 222.

⁶⁸⁵ Vgl. z.B.: Bisle, „Chi-Quadrat-Test‘ und Zeitreihenvergleich: Keine Schätzung bei ordnungsgemäßer Buchführung“, PSIR 1/2012, 15: „Der ‚Chi-Quadrat-Test‘ analysiert die Verteilung bestimmter Ziffern und basiert auf der Erkenntnis, dass jeder Mensch Lieblingsziffern hat und bei frei erfundenen Zahlen (z.B. nachgetragenen Tageseinnahmen oder -ausgaben) häufiger verwendet als andere.“ Tatsächlich ist der χ^2 -Test (Chi-Quadrat-Test) ein allgemeiner Wahrscheinlichkeitstest, der mehrklassige Verteilungen unter Berücksichtigung der Erkenntnisse der Wahrscheinlichkeitslehre auf Übereinstimmung prüft. Eine (zwingende) Verbindung zur Gleichverteilung von endständigen Ziffern in Wirtschaftsdaten besteht nicht. Vielmehr vermischt die Aussage die zwei Schritte „Erwartung“ und „Übereinstimmungsprüfung“ i.R. der musterbasierten Analyse und sorgt somit für Verständnisprobleme und Fehleinschätzungen; so auch: Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 136: „In der derzeitigen Praxis der Finanzämter hat sich der Chi-Quadrat-Anpassungstest als Überprüfungsmethode der Verteilungsannahmen bei der Ziffernanalyse etabliert. Dies hat u.a. dazu geführt, dass die Ziffernanalyse anhand des Benfordschen Gesetzes, insbesondere in der Praktiker-Literatur, mitunter fälschlicherweise grobschlächtig als ‚Chi-Quadrat-Test‘ bezeichnet und somit als eigenständige Methode, und nicht als Analyseinstrument, aufgefasst wird. Hier zeigt sich ein häufig beobachtetes unzureichendes Verständnis der Methodik und der statistischen Testverfahren.“

⁶⁸⁶ Üblicherweise stellen Ziffernanalysen zunächst die Beobachtung und (Punkt-)Erwartung für alle möglichen Ziffern gegenüber, was z.B. für die Einerstelle i.d.R. alle zehn Ziffern bedeutet. Somit werden die Datensätze nach zehn Ziffern, also in zehn Klassen ausgezählt. Soll die Untersuchung im Hinblick auf eine einzelne zu häufig oder selten vorkommende Ziffer vertieft werden, resultieren daraus die möglichen Ergebnisse „Ziffer x“ und „alle anderen Ziffern“, also nur noch zwei Klassen.

Abschnitt „Chi-Quadrat-Test“ unter [Pkt. 4.3.1.3.](#) sowie für Ziffernanalysen aus der Sicht einer einzelnen Ziffer mit der Binomialverteilung (s. unter [Pkt. 4.3.2.1.](#) u. [Pkt. 4.3.2.2.3.](#)). Auf diese Weise wird die Anzahl untersuchter Datensätze in das Übereinstimmungsergebnis einbezogen, wobei die Empfehlung zur Mindestdatenmenge für den χ^2 -Test (Chi-Quadrat-Test) von der SRP-Vorlage Stochastik automatisch berücksichtigt wird (s. [Pkt. 4.3.1.3.](#)).

- ☑ Das *mögliche Vorliegen nicht-manipulativer Ursachen* für auffällige Häufigkeitsmuster (s. den Abschnitt *„musterbasierte Prüfungsansätze zur Aufdeckung von Manipulationen“* unter [Pkt. 4.3.](#)) wird sowohl bei den SRP-Ziffernanalysen als auch bei der SRP-Verteilungsanalyse *selbständig untersucht*.⁶⁸⁷ Dazu reichen bei Einerstellenauffälligkeiten die Einschränkungsoptionen der SRP-Vorlage Stochastik (s. den Abschnitt *„zweistufige Eingrenzungsmöglichkeit“* unter [Pkt. 4.3.2.1.](#)) in Verbindung mit einfachen mathematischen Überlegungen (s. [Pkt. 4.3.2.2.3.](#)). Für Zweitziffernauffälligkeiten stellt die SRP-Vorlage Stochastik zielführende Funktionen zum Ausschluss praxisrelevanter nicht-manipulativer Faktoren zur Verfügung (s. [Pkt. 4.3.2.3.3.](#)). Vorbereitende Maßnahmen (s. [Pkt. 4.3.3.3.](#)) und eine gezielte Nachbetrachtung sichern i.R. der Verteilungsanalyse ein beobachtetes typisches Manipulationsmuster ab. Sollten darüber hinaus konkrete Einlassungen von Stpfl.⁶⁸⁸ vorliegen, können diese interaktiv in der SRP-Vorlage Stochastik z.B. mit dem Ausschluss nicht vergleichbarer Sondersachverhalte durch Markieren der betreffenden Datensätze auf dem Tabellenblatt „Daten“ berücksichtigt werden.

Die hohe Aussagegewissheit einer Plausibilitätsprüfung i.S. des § 158 Abs. 2 Nr. 1 AO mit der SRP ergibt sich nicht nur aus diesen einzelnen allgemeinverständlichen Qualitätsmerkmalen, sondern aus der mehrperspektivischen SRP-Musteranalytik, die alle Vorteile über ein systematisches Prüfungsschema (s. den Abschnitt *„Prüfungsschema zur Vermeidung von Fehldeutungen“* unter [Pkt. 4.3.](#)) verbindet. **Folgelogisch ordnet das BMF-Schreiben zu „[a]u-**

⁶⁸⁷ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 139 u. 161: „An die Beobachtungen aus Chi-Quadrat-Anpassungstest und Binomialtest schließen sich in der Praxis der Betriebsprüfung weitere Prüfungshandlungen an [...] Außerdem sind bei jeglicher Analyse die betrieblichen Besonderheiten des zu prüfenden Unternehmens einzubeziehen, um eine Verzerrung der Zahlenstruktur aufgrund von Preisgestaltungen oder bspw. Sitzplatzkapazitäten auszuschließen. Eine Übersicht der dafür durchgeführten Maßnahmen in der praktischen Anwendung der Betriebsprüfung bietet das Handbuch zur Summarischen Risikoprüfung [...] Die Verteilung der Tageseinnahmen kann sich je nach betrieblicher Besonderheit bspw. in den einzelnen Geschäftsjahren, saisonal oder wochentagsspezifisch unterscheiden. Dies ist in den Analysen berücksichtigt. Ein solcher Umstand einer ‚uneinheitlichen Absatzsituation‘ wird in der Praxis ebenfalls regelmäßig und sehr tiefgehend geprüft“.

⁶⁸⁸ Tipke/Kruse, AO/FGO, **171. Lieferung 7/2022**, § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit [...] Dabei ist allerdings die sphärenorientierte Mitwirkungspflicht des Stpfl. (s. § 88 Rz. 3 f.; § 90 Rz. 20) zu beachten. Mathematisch eindeutige Unschlüssigkeiten der Buchführungsergebnisse muss er durch betriebspezifische Besonderheiten schlüssig erklären“.

tomationsgestützten[n] quantitative[n] Prüfungsmethoden in der steuerlichen Außenprüfung“ deren Verprobungseignung wie folgt ein: „Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indizienge-samtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgaben-ordnung dazu führen, dass eine Schätzung durchgeführt werden muss.“⁶⁸⁹

In der bisherigen **Rechtsprechung** werden manipulationsorientierte Musteranalysen (bis-her) nur als zusätzliches Indiz anerkannt.⁶⁹⁰ Diese Einschätzung muss vor allem deshalb verwundern, weil der BFH aufgrund moderner (digitaler) Steuerhinterziehungsmöglichkeiten und deren verbreiteter Nutzung (s. [Pkt. 2.1.](#)) ein „nachvollziehbares Bedürfnis, moderne Prü-fungsmethoden – zu denen in geeigneten Fällen auch mathematisch-statistische Methoden gehören können – einzusetzen“⁶⁹¹, sieht. Auch der Bundesrechnungshof weist deshalb expli-zit darauf hin, dass „[n]ur IT-gestützte *quantitative* Prüfungsmethoden [...] auffällige **Muster** und Beziehungen in großen Datenmengen ersichtlich machen [können]“⁶⁹². Im Widerspruch dazu können aber die „mathematisch-statistischen Methoden in der Außenprüfung [...] bis-her nicht umfänglich eingesetzt werden, da [von der Rechtsprechung] zu hohe Anforderun-gen bezüglich der Sicherheit bestehen. [...] Bei den herkömmlichen, viel verwendeten Me-thoden, stellt sich die Frage, ob das geforderte Sicherheitsmaß überhaupt erreicht werden kann.“⁶⁹³ Musteranalysen basieren auf wissenschaftlichen Grundlagen⁶⁹⁴ und werden z.B.

⁶⁸⁹ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 4.

⁶⁹⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit (s. FG Münster v. 14.8.2003 – 8 V 2651/03 E, U, EFG 2004, 9 m. Anm. Trossen; FG Münster v. 10.11.2003 – 6 V 4562/03 E, U, EFG 2004, 236 [237]; FG Münster v. 7.12.2005 – 1 K 6384/03 E, EFG 2006, 652 [653] m. Anm. Zimmermann [Fahrtbuch]; FG Rh.-Pf. v. 24.8.2011 – 2 K 1277/10, EFG 2012, 10 [11] m. Anm. Kühnen; Sosna, StBp. 2000, 41 [Teil I], 68 [Teil II]; Meyer, DStR 2005, 2114; Watrin/Stuffert, DB 2006, 1748; Huber, StBp. 2009, 121 [124 ff.]“.

⁶⁹¹ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27 u. 54; nach Ansicht des Autors sollten die Finanzgerichte im Hinblick auf die vom BFH dargelegte Beurteilungssituation bei Besteuerungsgrundlagen, die möglicherweise digital manipuliert worden sind, dasselbe Bedürfnis nach zeitgemäß effektiven Prüfungs- und Verprobungs-methoden haben, weil gem. § 76 Abs. 1 S. 1 FGO „[d]as Gericht den Sachverhalt von Amts wegen [er-forscht].“

⁶⁹² Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

⁶⁹³ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüch-lichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 327.

⁶⁹⁴ Madauß, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstraßverfahren“, NZWiSt 1/2014, 24, 25: „Erwähnenswert ist in diesem Zusammenhang eine Untersuchung der Universität Münster (Institut für Wirtschaftsinformatik) über die Aussagekraft des Chi-Quadrat-Tests. [...] Diese Untersuchung dokumentiert deutlich, dass diese statistische-mathematische Prüfungsmethode auf wissenschaftlich gesicherten Grundla-gen beruht.“

zur Aufdeckung von Wahlfälschungen oder Waffenhandel eingesetzt⁶⁹⁵. Folglich ist die finanzgerichtliche Zurückhaltung nur so zu erklären, dass eine unsachgemäße Anwendung oder Ergebnisdeutung befürchtet wird⁶⁹⁶. Die SRP-Musteranalytik mit ihrem Prüfungsschema zur Vermeidung von Fehldeutungen gewährleistet einen hohen Anwendungsstandard und nachvollziehbare Ergebnisse. Sollten trotz selbständiger Aufklärungsbemühungen und interaktiver Optionen zur Berücksichtigung konkreter Erklärungsversuche⁶⁹⁷ mathematisch eindeutige (signifikante) Musterauffälligkeiten bestehen bleiben, erreichen diese auch unter Beachtung bedingter Wahrscheinlichkeiten⁶⁹⁸ eine hohe indizielle Aussageklarheit und stellen demnach einen Beanstandungsanlass gem. § 158 Abs. 2 Nr. 1 AO (s. [Pkt. 4.6.](#)) dar. Deshalb wertet das FG Hamburg eine stochastisch eindeutige Musterauffälligkeit zu Recht als eigenständiges Indiz – wenn auch im Bereich der formellen Mängel.⁶⁹⁹

4.4. Gesamtergebnis einer plausibilitätsorientierten SRP

Nach einer plausibilitätsorientierten SRP werden Schlussfolgerungen gem. § 158 Abs. 2 Nr. 1 AO – gegebenenfalls i.V.m. § 162 Abs. 2 S. 2 AO – auf Basis des **Gesamtbilds** aus

- der Ordnungsmäßigkeitsprüfung
- und den systematisch-mehrperspektivischen Analysen

⁶⁹⁵ Vgl. z.B.: Handelsblatt online v. 8.7.2012: „Volkswirte auf Verbrecherjagd: Ob illegaler Waffenhandel oder Steuerhinterziehung – Ökonomen stöbern immer häufiger Missetäter auf. Die forensische Ökonomie ist ein boomender Forschungszweig. Ermittlungsbehörden nutzen die Erkenntnisse der Forscher“, [früher](https://www.handelsblatt.com/politik/konjunktur/oekonomie/nachrichten/forensische-oekonomie-volkswirte-auf-verbrecherjagd/6838146.html) nachzulesen unter: <https://www.handelsblatt.com/politik/konjunktur/oekonomie/nachrichten/forensische-oekonomie-volkswirte-auf-verbrecherjagd/6838146.html> oder Handelsblatt online v. 30.11.2009: „Auf der Jagd nach Zahlen-Fälschern: Mit einem lange vergessenen mathematischen Gesetz wollen Ökonomen Kollegen überführen, die in wissenschaftlichen Studien Daten manipuliert haben. Denn längst nicht alle Wirtschaftswissenschaftler sind ehrlich – und die Anreize für Betrüger sind groß“, nachzulesen unter: <https://www.handelsblatt.com/politik/konjunktur/oekonomie/nachrichten/oekonomen-auf-der-jagd-nach-zahlen-faelschern/3315384.html>.

⁶⁹⁶ Kohlmann, Steuerstrafrecht, 72. Lieferung 9.2021, § 397 AO, Rz 12.3: „Die praktische Verwertbarkeit der statistischen Verfahren ist durchaus umstritten, und zwar bereits als Grundlage für steuerliche Schätzungen, aber erst recht für strafrechtliche Verurteilungen. Die Finanzgerichte lehnen bislang den Chi-Quadrat-Test als eine sichere Methode, die Ordnungsgemäßheit der Buchführung zu verwerfen, ab. Allerdings wird weniger die grundsätzliche Eignung der Analysen angezweifelt als vielmehr hinterfragt, unter welchen Voraussetzungen und mit welcher Anzahl von Ausgangswerten die Methode aussagekräftige Ergebnisse liefert.“

⁶⁹⁷ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 21a: „Eine Ziffernanalyse durch einen sog. Chi²-Test besitzt aber allenfalls indiziellen Charakter für eine Unregelmäßigkeit [...] Dabei ist allerdings die sphärenorientierte Mitwirkungspflicht des Stpfl. (s. § 88 Rz. 3 f.; § 90 Rz. 20) zu beachten. Mathematisch eindeutige Unschlüssigkeiten der Buchführungsergebnisse muss er durch betriebsspezifische Besonderheiten schlüssig erklären“.

⁶⁹⁸ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 143 ff.: Zur genaueren Bewertung eines Indizes müssen neben der fehlerhaften Beanstandung (bei 99 % Testgenauigkeit also 1 %) auch die Tatverbreitung und Aufdeckungswahrscheinlichkeit zumindest schätzungsweise bekannt sein.

⁶⁹⁹ Urt. v. 5.3.2018 – 3 K 205/15, Rz. 66: „Schließlich hat die stochastische Analyse der erklärten Erlöse ergeben, dass die erklärten Zahlen der Samstage mit an Sicherheit grenzender Wahrscheinlichkeit nicht zufällig entstanden, sondern manipuliert worden sind.“

- sowie im Fall beobachteter Auffälligkeiten den selbständigen Aufklärungsmaßnahmen bzw. den eingepflegten Erklärungen der Stpfl. getroffen.

I.R. der systematisch-mehrperspektivischen Analysen stellt die **Kombination** aus *geeigneten* (s. die Abschnitte *Methodenkenntnis*“ u. *Methodenvalidierung*“ unter [Pkt. 4.2.](#)) regel- und musterbasierten Datenverbildlichungen – wie bei der (quasi) dreidimensionalen Bildgebung in der medizinischen Diagnostik – sicher, dass mögliche gesamtbildrelevante Informationen nicht übersehen werden (s. *„Prüfungssystematik der SRP-Vorlage BWA“* unter [Pkt. 4.2.2.](#)). Dadurch können sie für eine sachgerechte Beurteilung falls nötig interaktiv (s. [Pkt. 4.2.9.](#)) auf Normalverhältnisse korrigiert werden: bspw. bei erheblichen (einseitigen) Preiseffekten (s. *das entsprechende Beispiel* unter [Pkt. 4.2.5.1.](#)). Auffälligkeiten aus manipulationsorientierten Musteranalysen werden ebenfalls gezielt auf nicht-manipulative Ursachen hinterfragt und entsprechend angepasst bzw. bewertet (s. [Pkt. 4.3.2.2.3.](#), [4.3.2.3.3.](#) u. [4.3.4.5.](#)).

Im Ergebnis ist bereits jedes **einzelne verbleibende Beweisanzeichen** gegen die sachliche Richtigkeit untersuchter Besteuerungsgrundlagen so weit abgesichert, wie es die Datenqualität und Aufklärungsmitwirkung⁷⁰⁰ im Einzelfall zulassen.

Nach ständigen Prüfungserfahrungen besteht ein gem. § 158 **Abs. 2 Nr. 1** AO zu beanstandendes SRP-Gesamtbild – insbesondere bei Fällen mit überwiegend Bargeschäften – zudem aus **mehreren Indizien gegen die sachliche Richtigkeit**.

Die nachfolgende Abb.⁷⁰¹ listet die (wichtigsten) Ergebnisse einer typischen Beanstandung der Einnahmenvollständigkeit infolge einer SRP auf. Anhand der (knapp gefassten⁷⁰²)

⁷⁰⁰ BMF-Schreiben v. 3.12.2020 „Verwaltungsgrundsätze 2020“, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. Die Frage der Feststellungslast stellt sich, wenn ein entscheidungserheblicher Sachverhalt trotz Ausschöpfung aller zugänglichen und zumutbaren Ermittlungsmöglichkeiten nicht vollständig aufgeklärt werden kann. Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u. a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“ Tipke/Kruse, AO/FGO, **171. Lieferung 7/2022**, § 158 AO, Rz. 9: „§ 162 II i.V.m. § 158 orientieren das Beweismaß an der Sphärenverantwortlichkeit [...] Das Beweismaß hängt mithin davon ab, wie viel und in welcher Form der Stpfl. aus seiner Verantwortungssphäre zur Sachaufklärung beiträgt und wieweit die FinBeh. auf dessen Mitwirkung im konkreten Einzelfall angewiesen ist.“

Aufzählungspunkte zu festgestellten Ordnungswidrigkeiten entlang der drei Aufzeichnungsebenen „Einzelerfassung mit elektronischem Kassensystem“, „tägliche Zusammenfassung im Kassenbuch“ und „Buchführung“ lässt sich eindeutig schlussfolgern, dass keine ausreichenden Maßnahmen zur Absicherung der manipulationsgefährdeten Kassendaten gegen Eingriffe getroffen worden sind (s. den Abschnitt „Gewissheit der Originalität“ unter [Pkt. 4.1.](#)). Infolge der Schwere⁷⁰³ der Ordnungsmängel wird die Verifizierbarkeit wesentlicher Besteuerungsgrundlagen so stark beeinträchtigt, dass „systembedingt keine Gewähr mehr für die Vollständigkeit der Erfassung der Bareinnahmen“⁷⁰⁴ gegeben ist. Die „Beweiskraft der Buchführung (und Aufzeichnungen)“ ist damit bereits deshalb nach § 158 Abs. 1 AO erschüttert.

<u>Ordnungsmäßigkeit</u>	<u>SRP-Ergebnisse</u>
- Erfassung der Bareinnahmen mit einem (speicherfähigen) elektronischen Kassensystem aus Terminal und Ordermen ➔ keine Einstellungsprotokolle (Zusammenführung aller Ordermen?) - betriebsinternes Excel-Kassenbuch mit täglicher Eintragung der Bareinnahmen, Ausgaben, Entnahmen und Einlagen sowie fortlaufender Bestandsberechnung ➔ keine Absicherung (technisch bzw. organisatorisch) zur Unterbindung von nachträglichen spurlosen Änderungen - Weitergabe der Kassenberichte und Belege i.d.R. einmal monatlich zur Buchführung an das Steuerbüro ➔ keine sofortige Festschreibung der Kassenaufzeichnungen (frühestens mit Buchung im Steuerbüro)	- unschlüssige Entwicklung des Warenflusses (z.T. fehlende Einkaufsreaktionen auf den Absatz) ➔ Alltagseffekte beim robusten SRP-Modell als Ursache äußerst unwahrscheinlich ➔ keine Ausnahmevorgänge auffindbar bzw. nachgewiesen durchschnittliche Wochentagserlöse mit Anstieg um 20 % <u>nur</u> an den Freitagen ➔ Zufallsentwicklung äußerst unwahrscheinlich ➔ kein Nachweis einer regelmäßigen absatzerhöhenden Freitagsaktion Verteilungsanalyse der Freitags-Bareinnahmen mit typischem Bild des Spitzeneingriffs ➔ mathematisch eindeutige (signifikante) Abweichung von der (betriebsspezifischen) Logarithmischen Normalverteilung unter Beachtung der geringen Datenanzahl ➔ keine Doppel- bzw. Mehrfachverteilung und keine Kapazitätsgrenze erkennbar

Aus den auf der rechten Seite aufgeführten auffälligen Beobachtungen zu den (betroffenen) Besteuerungsdaten „Einnahmen“ und „Warenfluss“ lässt sich außerdem entnehmen, dass

⁷⁰¹ Quelle (überarbeitet): *Webel/Wähnert*, „Der Verdacht einer Steuerstraftat während der Außenprüfung – § 10 BpO im Kontext der Digitalisierung“, NWB 44/2014, 3324, 3332. Das Beispiel stammt aus der Zeit vor Einführung der zertifizierten technischen Sicherheitseinrichtung gem. § 146a Abs. 1 S. 2-3 AO und soll lediglich zur bildhaften Verdeutlichung der beweisverstärkenden Indiziengesamtheiten beitragen.

⁷⁰² In der Prüfungspraxis sollte die Darlegung von Ordnungswidrigkeiten ausführlicher erfolgen (s. das Beispiel im Abschnitt „Ordnungsmängel ausführlich zu dokumentieren“ unter Pkt. 4.1.).

⁷⁰³ Vgl. *Nöcker (RiBFH)*, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, 3550: „Liegen keine materiellen Buchführungsmängel vor, entscheidet die Gewichtigkeit der formellen Buchführungsmängel darüber, ob eine Schätzungsbefugnis des Finanzamts vorliegt.“

⁷⁰⁴ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

die mangelhafte Nachprüfbarkeit offenbar in einem Zusammenhang⁷⁰⁵ zu wesentlichen inhaltlichen Unrichtigkeiten steht. Dabei erreicht die SRP bereits für jedes einzelne der aufgelisteten Indizien gegen die sachliche Richtigkeit dank ihrer *speziellen Merkmale* eine hohe Aussagegewissheit:

1. das *robuste zentrale Zeitreihenmodell* (s. die Gegenüberstellung gängiger Zeitreihenmodelle im Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) i.V.m. der *SRP-Zeitreihen-systematik* zum Auffinden und Neutralisieren möglicher Störeinflüsse (s. [Pkt. 4.2.2.](#)),
2. die *vielfältigen und variablen (internen) Zeitreihenmöglichkeiten* der untersuchten Daten (s. [Pkt. 4.2.7.](#)) mit *interaktiven Korrekturoptionen* (s. [Pkt. 4.2.9.](#)) zur weitest möglichen Konkretisierung beobachteter Auffälligkeiten sowie
3. die *qualifizierte Beurteilung von manipulationsorientierten Musteranalysen* mit Hilfe von Wahrscheinlichkeitstests (s. [Pkt. 4.3.1.3.](#)) zuzüglich der *gezielten Abklärung möglicher nicht-manipulativer Ursachen* (für die Verteilungsanalyse des Beispiels s. [Pkt. 4.3.4.5.](#)).

Mindestens genauso entscheidend für die bessere Aussageklarheit der systematisch-mehrperspektivischen SRP ist außerdem die gegenseitige Verstärkung der einzelnen Indizien zu einer **Indiziengesamtheit**, wozu der BGH *grundsätzlich* feststellt: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreicht[en].“⁷⁰⁶ Diese beweisverstärkende Wirkung mehrerer Indizien ist mathematisch logisch: Ein einzelnes Indiz ist eine Feststellung zur möglichen Täterschaft mit einer i.d.R. nicht genau bekannten aber höheren Wahrscheinlichkeit. Liegen zwei (unabhängige⁷⁰⁷) Indizien vor, welche dieselbe Täterschaft vermuten lassen, ermittelt sich die Gesamtrestunsicherheit mit: „Restunsicherheit₁⁷⁰⁸ x Restunsicherheit₂“. Im Fall zweier unabhängiger Indizien von jeweils ca. 90 % Gewissheit reduziert sich damit die Restunsi-

⁷⁰⁵ Vgl. z.B.: Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 24: „Allerdings entspricht es der Erfahrung, dass eine formell ordnungswidrige Buchführung i.d.R. auch sachliche Mängel aufweist.“

⁷⁰⁶ BGH v. 27.4.2010 – 1 StR 454/09, NSZ 2011, 108, Rz. 18 u. 20. Vgl. auch: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 4: „Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit.“

⁷⁰⁷ Die einfache Berechnung von Gesamtwahrscheinlichkeiten unabhängiger Ereignisse ist aus der Spieltheorie bekannt. Soll z.B. beim Werfen eines Würfels die Möglichkeit, mehrere Male in Folge eine „6“ zu werfen, ermittelt werden, sind die Einzelwahrscheinlichkeiten zu multiplizieren. Für bspw. dreimal „6“ nacheinander ergibt sich somit: $\frac{1}{6} \times \frac{1}{6} \times \frac{1}{6} = \frac{1}{216}$, also ein seltenes Gesamt ereignis (ca. 0,5 %).

⁷⁰⁸ Restunsicherheit Indiz₁ = 100 % – Indizien gewissheit₁.

cherheit auf: „10 % x 10 % = 1 %“, und es steigert sich die Gesamtklarheit *exponentiell* auf 99 %; bei (möglicherweise) abhängigen Beweisanzeichen ist der verstärkende Effekt weniger extrem aber immer noch entscheidungserheblich.⁷⁰⁹ Sowohl zu den oberhalb *unter diesem Pkt.* aufgeführten methodischen Merkmalen der SRP als auch zu diesen mathematischen Grundlagen einer indiziellen Beweisführung gilt folgende Feststellung des BGH: „Die tatrichterliche Würdigung darf [...] den Gesetzen der Logik und dem gesicherten wissenschaftlichen Erfahrungswissen nicht widersprechen [...]. Soweit maßgeblich auf [...] statistische Wahrscheinlichkeitsberechnungen abgestellt wird, sind daher die zu Grunde liegenden mathematischen Denkgesetze zu beachten“⁷¹⁰. Wegen der Wichtigkeit dieser mathematischen Grundlagen zur Gesamtbeurteilung einzelner bzw. mehrerer Indizien beschäftigt sich auch die Wissenschaft mit „Bayessche[n Wahrscheinlichkeits]etze[n] zur Unterstützung juristischer Entscheidungen“⁷¹¹. Im Verhältnis zur SRP betont die wissenschaftliche Autorin Schütt hierzu: „Ein wesentlicher Aspekt, der bezüglich der Sicherheit eine wesentliche Rolle spielt, welcher ein Grundgedanke der SRP ist, ist die bereits erwähnte Indizienverstärkung. Werden mehrere Analysen durchgeführt, liegen als logische Schlussfolgerung mehrere Ergebnisse vor. Auch an dieser Stelle können Erkenntnisse aus der Mathematik bzw. der Wahrscheinlichkeitsrechnung genutzt werden.“⁷¹²

Bezüglich der gegenseitigen Verstärkung formeller und materieller Mängel von Besteuerungsaufzeichnungen wird die Wirkung von Indiziengesamtheiten auch in der Drei-Stufen-Theorie des BFH⁷¹³ verarbeitet. Indes sind die Folgen beim Zusammenfallen von zwei oder mehreren *materiellen* Einzelauffälligkeiten i.S. des § 158 Abs. 2 Nr. 1 AO genauso wichtig: In dem oberhalb abgebildeten Beispielsfall zu einem SRP-Gesamtbild kann eine Abhängigkeit der beiden unschlüssigen Beobachtungen zu den Freitagserlösen bzw. -bareinnahmen nicht ausgeschlossen werden. Demnach würde hier keine *vollständige* multiplikative Beweisverstärkung resultieren. Unzweifelhaft ist jedoch, dass die implausiblen Warenflussverhältnisse

⁷⁰⁹ Die Ausführungen der Gutachter Köller/Nissen/Rieß/Sadorf („Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Reihe „Polizei + Forschung“, Bd. 26, herausgegeben vom Bundeskriminalamt (BKA), Kriminalistisches Institut, Luchterhand Verlag, 2004, 22, s. Pkt. 4.6) beziehen sich auf die Bayes-Logik zur Verknüpfung abhängiger Wahrscheinlichkeiten (Beweisanzeichen) in Gerichtsverfahren.

⁷¹⁰ BGH v. 24.3.2016 – 2 StR 112/14, NStZ 2016, 490, Leitsatz Nr. 5.

⁷¹¹ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 145 ff.

⁷¹² Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 327. Demgegenüber werden die beweisverstärkenden Zusammenhänge bei Prüfungsnetzen wie der SRP in der kritisierenden Literatur entweder gar nicht erwähnt oder in der Beurteilung außenvorgelassen (vgl. z.B.: Beyer, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, oder Bleischick, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353).

⁷¹³ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 62 ff.

unabhängig von den beiden anderen Indizien sind. Im Ergebnis reduzieren sich die Restunsicherheiten der Plausibilitätsbeurteilung einmal stark (exponentiell) und zusätzlich einmal etwas weniger deutlich, also insgesamt wesentlich. In Wirtschaftsprüfung⁷¹⁴, Wissenschaft⁷¹⁵ und AO-Kommentierung⁷¹⁶ werden die Vorteile einer systematisch-mehrperspektivischen Prüfungsweise wie bei der SRP inzwischen durchaus gesehen.

4.5. SRP im Verhältnis zu anderen Prüfungsmethoden

Die SRP ist ein komplexes Prüfungsnetz zur schnellen und effektiven Orientierung in größeren Datenmengen. Sie verbindet dazu mit diversen regel- und musterbasierten Ansätzen zwei unterschiedliche Wege der Datenvisualisierung inklusive zielführender Anpassungsfunktionen, weil so die Stärken der menschlichen Wahrnehmung besser eingesetzt werden können (s. mit einem Beispiel dazu den Abschnitt *"Visualisierung von Besteuerungsdaten"* unter [Pkt. 2.](#)).⁷¹⁷ Trotzdem werden dadurch die vorgeschaltete Formalprüfung und **andere Methoden** wie z.B. Vermögenszuwachs- bzw. Geldverkehrsrechnungen oder Kalkulationen **nicht überflüssig**; zudem werden durch die Vorteile der Digitalisierung komplexere Prüfungsverfahren wie z.B. „Clustering“⁷¹⁸ oder iteratives⁷¹⁹ „Outlier-Mining“ (s.a. den Abschnitt *„systematisches Erkennen, Eingrenzen und Aufklären betriebsspezifischer Prüffelder“* mit einem herkömmlichen statistischen Verfahren unter [Pkt. 3.1.](#)) erstmalig allgemein zugänglich und dürften die Arbeitsweise der Betriebsprüfung erweitern.

⁷¹⁴ Internetseite der Wirtschaftsprüfungs-/Steuerberatungsgesellschaft Rödl & Partner GmbH, Veröffentlichung v. 2.5.2018: „Vorteile der Summarischen Risikoprüfung als digitale Prüfungsmethode – Schnell gelesen: 1.) Digitale Prüfungsmethoden bringen Vorteile mit sich und ermöglichen ein mehrperspektivisch erarbeitetes Gesamtbild. 2.) Durch eine sinnvolle Verknüpfung vieler Datenverbildlichungen lässt sich eine Schlussfolgerung im Sinne des § 158 AO ableiten.“

⁷¹⁵ Sell, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 218: „Ein relevantes und wertvolles Werkzeug der Finanzbehörden ist dabei die Summarische Risikoprüfung (SRP-Handbuch 2020) geworden, die [...] den methodischen Standard der praktischen Nutzung darstellt.“

⁷¹⁶ Zugmaier/Nöcker/Webel, Abgabenordnung, Februar 2022, § 162 AO, Rn. 157: „Die Ergebnisse einer technisch einwandfrei durchgeführten [S]ummarischen Risikoprüfung sind aufgrund ihrer umfassenden und systematischen mehrperspektivischen Absicherung ausreichend, um die Ordnungsmäßigkeit der Buchführung zu erschüttern.“

⁷¹⁷ Damit führt die SRP in konzentriert vereinfachter Form an die zunehmend in Wirtschaft und Beratung verbreiteten B(usiness-)I(ntelligence)-Programme heran. Zwar werben Hersteller wie Microsoft für ihre BI-Produkte mit: „Nutzen Sie die aktuellen Fortschritte bei der Microsoft-KI, um auch Laien das Aufbereiten von Daten, das Erstellen von Machine Learning-Modellen und das Auswerten strukturierter und nicht strukturierter Daten wie Text und Bildern zu ermöglichen“, früher nachzulesen unter: <https://powerbi.microsoft.com/de-de/why-power-bi/>. Tatsächlich setzt eine intuitiv-richtige Beurteilung multipler Analysen aus derartig funktionsreichen Programmen aber Methodenverständnis und eine sinnvolle Prüfungssystematik voraus.

⁷¹⁸ Beim „Clustering“ werden unzusammenhängende Teile einer Datenmenge statistisch bzw. visuell identifiziert.

⁷¹⁹ Durch die digitale Programmierbarkeit von wiederholten Prüfungs-Entscheidungs-Abläufen als automatisierte Routinen können aufwendige Ausreißer-Erkennungen z.B. mit mehrfacher Klassifizierung von Datenbeständen unter Beobachtung einzelner Datensätze gegenüber Häufungen komfortabel umgesetzt werden.

Am besten wirken sich auch die Vorteile der SRP aus, indem **alle zur Verfügung stehenden Prüfungsmittel** funktionsgerecht und nach Effektivitäts- wie Effizienzerwägungen⁷²⁰ in die jeweilige Prüfung **eingebunden** werden:

 Die **Formalprüfung** von Besteuerungsunterlagen bzw. -daten auf Ordnungsmäßigkeit ist zu Beginn einer Außenprüfung grundsätzlich erforderlich (s. [Pkt. 4.1.](#)). Elektronische Aufzeichnungen lassen ohne spezielle Merkmale bzw. Vorkehrungen alleine kaum seriöse Aussagen über ihre Vollständigkeit und Unverändertheit zu. Auch verprobende Gesamtbildprüfungen sind unter solchen Umständen nur bedingt aussagekräftig. Deshalb können gravierende Ordnungsmängel direkt die Amtspflicht der Schätzung auslösen.⁷²¹

 Mit dem Ziel eines sinnvollen Ressourceneinsatzes sollte bei jeder Außenprüfung daran anschließend die **Erarbeitung eines Überblicks** folgen, um gezielt einzelfallbezogene Prüffelder festzustellen und gegebenenfalls das Gesamtbild nach § 158 **Abs. 2 Nr. 1** AO auf Schlüssigkeit zu beurteilen. Besonders infolge der zunehmenden Verbreitung digitaler Betriebsverwaltungssysteme mit immer mehr aufgezeichneten Informationen⁷²² (s. [Pkt. 2.1.](#)) ist selbst bei kleinen Unternehmen eine vollständige Überprüfung aller steuerrelevanten Unterlagen nur mit einem unverhältnismäßigen Zeit- und Arbeitsaufwand möglich. Deshalb sind Übersichtsmethoden, welche betriebsbezogen risikobehaftete Sachverhalte entdecken können, von zentraler Bedeutung für ein funktionieren des Risikomanagement⁷²³. Dieser zielführende Überblick ist die Hauptfunktion der **SRP** (s. [Pkt. 2.](#)).

 **Vermögenszuwachs- bzw. Geldverkehrsrechnungen und Kalkulationen** sind arbeitsaufwendige Prüfungsmaßnahmen, worauf nur Stpfl. Anspruch haben, die ih-

⁷²⁰ § 7 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen. Ihre Dauer ist auf das notwendige Maß zu beschränken. Sie hat sich in erster Linie auf solche Sachverhalte zu erstrecken, die zu endgültigen Steuerausfällen oder Steuererstattungen oder -vergütungen oder zu nicht unbedeutenden Gewinnverlagerungen führen können.“

⁷²¹ Vgl. Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, 3550: „Liegen keine materiellen Buchführungsmängel vor, entscheidet die Gewichtigkeit der formellen Buchführungsmängel darüber, ob eine Schätzungsbefugnis des Finanzamts vorliegt.“

⁷²² Vgl. z.B.: Graf, „Anforderungen an die Vorlage von Einzelaufzeichnungen – Ende der ‚Schonzeit‘ für Daten aus Vor-Systemen?“, BBK 18/2013, 875, 876, 877: „Die Buchführung hat sich von einer Durchschreibebuchführung hin zu einer umfassenden Datenverarbeitung weiterentwickelt, die elektronische Form ist nunmehr die übliche. [...] In den Unternehmen [...] werden unterschiedlichste betriebliche Aufgaben und Anforderungen in Datenbanken protokolliert oder über Datensysteme gesteuert. Dabei steigt die Anzahl elektronisch überwachter Prozesse stetig.“

⁷²³ Bei weitgehend automatisierten Risikoselektionssystemen wie RMS-BP (vgl. z.B.: Danielmeyer, „Mit einem modernen Risikomanagement hin zur zeitnahen Betriebsprüfung?“, StBp 2021, 271) oder einer Systemprüfung von Steuerkontrollsystemen mit der Möglichkeit, Beschränkungen der nächsten Außenprüfung zuzusagen (Art. 97 § 38 EGAO), sollte jedoch kritisch hinterfragt werden, inwieweit die verfügbare Datengranularität Risiken aufzeigen kann bzw. ob Ergebnisoffenheit alleine durch regelbasierte Routinen erreichbar ist.

ren Mitwirkungspflichten nachkommen.⁷²⁴ Falls effizientere Übersichtsmittel wie die SRP trotz *Mitwirkungsbereitschaft von Stpfl.* nicht zu einer ausreichend klaren Beurteilungslage für § 158 Abs. 2 Nr. 1 AO führen, sollten sie partiell – und damit unter ressourcenschonendem Aufwand – zur Absicherung ergänzt werden; hierbei ist allerdings die in der Psychologie bekannte Wahrnehmungsverzerrung „Verfügbarkeitsheuristik“ zu beachten, wonach u.a. einfach berechenbare Ergebnisse auf der Basis unsicherer Ausgangsgrößen häufig zu Unrecht als sicherer und klarer angesehen werden als indirekte Verprobungsmethoden und wahrscheinlichkeitstheoretisch ermittelte Ergebnismöglichkeiten.⁷²⁵

 Auch **mathematisch-statistische Stichprobenverfahren**, wie sie in der Wirtschaftsprüfung z.B. in Form des "Monetary Unit Sampling (**MUS**)"⁷²⁶ üblich sind, verfolgen den Gedanken eines risikogerechten Prüfungsansatzes.⁷²⁷ Anwendungsbereiche und Konsequenzen unterscheiden sich jedoch von systematisch-mehrperspektivischen Datenvisualisierungen wie der SRP: MUS eignet sich für *Niedrigrisikobereiche* mit der Gefahr von (punktuellen) *Überbewertungen*⁷²⁸. Aufgefundene Mängel werden i.d.R. nicht nachverfolgt, sondern als Punktschätzung zum wahrscheinlichsten Fehler⁷²⁹ der Grundgesamtheit hochgerechnet. Eine zielführende Kombination aus MUS und SRP kann über erhebliche Effizienzgewinne wesentlich zur Gleichmäßigkeit der Besteuerungskon-

⁷²⁴ Vgl. z.B.: BFH v. 21.2.1990, BFH/NV 1990, 683, Rz. 31: „Der Anregung des Klägers, eine Geldverkehrsrechnung zu erstellen, braucht das FG nicht nachzukommen. Die Geldverkehrsrechnung ist eine arbeitsaufwendige Schätzungsmethode. Die vom Kläger erklärten Buchführungsergebnisse beruhen auf einer formell ordnungswidrigen Buchführung. Daher ist die Anwendung weniger arbeitsaufwendiger Schätzungsmethoden statthaft.“

⁷²⁵ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Nomos, 2018, 56-57 u. 65: „Bei der Verfügbarkeitsheuristik schätzen Individuen die [...] Wahrscheinlichkeit für das Eintreten eines Ereignisses unter Unsicherheit danach ein, wie leicht bzw. schwer es ihnen fällt, damit verbundene Erscheinungen aus ihrem Gedächtnis abzurufen. Dabei werden Ereignisse für wahrscheinlicher erachtet, wenn sie sich leichter vorgestellt werden können und sich zu ihnen eine direkte Verfügbarkeit herstellen lässt. [...] Im Rahmen juristischer Entscheidungen scheint es unerlässlich, von diesen Entscheidungsverzerrungen Kenntnis zu haben, da eine verzerrte Entscheidung den Denkgesetzen der Logik widerspricht [...]“.

⁷²⁶ Vgl. z.B.: Giezek, „Monetary Unit Sampling – Der Einsatz statistischer Verfahren im Rahmen der Jahresabschlussprüfung“, Gabler Verlag, 2011.

⁷²⁷ BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1, 4, 19 ff. u. 30: „Bei einer steuerlichen Außenprüfung können mathematisch-statistische (quantitative) Prüfungsmethoden eingesetzt werden, um die Besteuerungsgrundlagen des Steuerpflichtigen zu verproben und auf Plausibilität zu überprüfen. Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich. [...] Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgabenordnung dazu führen, dass eine Schätzung durchgeführt werden muss. [...] 5. Stichprobenverfahren [...] b) Monetary Unit Sampling [...] 4. Schätzung nach Monetary Unit Sampling“.

⁷²⁸ Für eine mathematisch-statistische Stichprobenprüfung im Hinblick auf mögliche *Unterbewertungen* muss vor allem das Ziehungsverfahren des „Monetary Unit Sampling (MUS)“ angepasst werden.

⁷²⁹ Most Likely Error (MLE).

trolle gem. § 85 AO beitragen. Demzufolge ist das Pilotprojekt "Monetary Unit Sampling (MUS)" in Rheinland-Pfalz von einer Kooperation mit der SRP begleitet worden.⁷³⁰

Demgegenüber ist der medial präsente⁷³¹ Trend einer (fast) vollständigen Automatisierung von Prüfungsprozessen durch „**Künstliche Intelligenz (KI)**“ oder vordefinierte „Dashboards“⁷³² kritischer zu sehen: *Entgegen der früheren Meinung des Autors werden die unzähligen Regeln ergebnisoffener, wirksamer Analysen mit ständigen Anpassungen an neue Erkenntnisse zwischenzeitlich z.T. vollständig durch selbstlernende Algorithmen umgesetzt*⁷³³, was die Frage aufwirft, ob sie von unabhängiger Seite in angemessener Zeit kontrolliert werden können. Außerdem schließen derartige Automatismen die menschlichen Stärken „Intuition“ und „Erfahrung“ (weitgehend) aus; da jedoch die Problemlösungen des menschlichen Gehirns chaotisch assoziativ und somit anders als die geordneter Datenbanken erfolgen, bliebe dadurch viel Prüfungskompetenz ungenutzt.⁷³⁴ Dieser Mangel wird scheinbar in den Führungsebenen der Wirtschaft zumindest „gefühl“ wahrgenommen.⁷³⁵ Entsprechend werden in der Wirtschaftsprüfung zwischenzeitlich die Grenzen des KI-Einsatzes diskutiert.⁷³⁶ *Unabhängig von den genannten Bedenken erhöht diese Entwicklung den Druck auf die Aus- und Fortbildung, eine angemessene Anwendungs- und Lesefähigkeit in der Breite der Arbeitswelt zu etablieren: „Data Literacy gehört zu den wichtigen Future Skills, um in der digitalen Welt in Wissenschaft, Arbeitswelt und Gesellschaft bestehen und teilhaben zu können. Damit gehört sie zu den bedeutendsten Digitalkompetenzen der Zukunft. Data Literacy bezeichnet die Fähigkeit, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzu-*

⁷³⁰ Bundesfinanzministerium, Bundesfinanzakademie, Quartals-Info I/2018, „Podiumsdiskussion: Monetary Unit Sampling (MUS) – mathematisch-statistische Stichproben in der Betriebsprüfung“, früher nachzulesen unter https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Ministerium/Geschaeftsbereich/Bundesfinanzakademie/Aktuelles/BFA_Quartalsinfo_2018_I.html.

⁷³¹ Kompenhans/Wermelt, „Big Data“ – Potenziale für innovative Abschlussprüfungen“, BB 2018, 299: „Die an Systemschnittstellen erforderlichen manuellen Tätigkeiten werden immer mehr durch den Einsatz von Robotic Process Automation ersetzt, Entscheidungsprozesse werden dabei durch den Einsatz programmierter, selbstlernender Entscheidungsheuristiken, die auf Basis künstlicher Intelligenz arbeiten, unterstützt.“

⁷³² Nachzulesen z.B. unter: https://docu.caseware.net/Audit/20.00/AuditTemplate/de/Content/User_Client_File_Setup/t_Using_the_Work_Flow_Panel.htm.

⁷³³ Vgl. z.B.: Webinarreihe „TAXPUNK meets ChatGPT [...] 4. BEYOND TEXTFILES [...] mit] Stefan Werner“, in YouTube eingestellt unter: https://www.youtube.com/watch?v=xT_L2wPXxac.

⁷³⁴ Giezek/Wähner, „Künstliche Intelligenz vs. anwender gelenkten Softwareeinsatz – Ist KI generell leistungsfähiger oder nur besser zu verkaufen?“, DB 31/2019, 1707.

⁷³⁵ CRN online v. 24.2.2010: „Management-Studie: Manager vertrauen mehr auf ihr Bauchgefühl als auf B[usiness] I[n]telligence“, früher nachzulesen unter: <https://www.crn.de/software-services/artikel-80249.html>.

⁷³⁶ DB online v. 30.4.2019 (Sponsored News v. EY), „Die Zukunft der Wirtschaftsprüfung ist digital: [...] Während Künstliche Intelligenz (KI) und Robotics Routinetätigkeiten übernehmen, konzentrieren sich die Prüfer von EY zunehmend auf die Interpretation der Daten. [...] Für Wirtschaftsprüfer heißt das, dass sie neben ihrem Fachwissen Kenntnisse im IT-Bereich benötigen, um diese Prozesse zu verstehen“, früher nachzulesen unter: <https://www.der-betrieb.de/meldungen/die-zukunft-der-wirtschaftspruefung-ist-digital/>.

wenden. Sie ermöglicht die systematische Umwandlung von Daten in Wissen und ist damit unverzichtbarer Bestandteil der Allgemeinbildung⁷³⁷.

4.6. Welches Beweismaß fordert § 158 Abs. 2 Nr. 1 AO?

Am Anfang jeder rechtlich zu beurteilenden (indiziellen) Situation muss eine möglichst klare Vorstellung von den gesetzlich geforderten Voraussetzungen stehen. Bezogen auf das Besteuerungsmassenverfahren soll das **Verifikationsprinzip** die Verfassungsmäßigkeit der Steuerbelastung dadurch sicherstellen, dass eine gleich- und gesetzmäßige Besteuerung **tatsächlich** umgesetzt werden kann⁷³⁸ – also „kein dem Gesetzgeber zuzurechnendes strukturelles Vollzugsdefizit“⁷³⁹ besteht, indem sich einzelne zu Lasten der Gemeinschaft der Steuerehrlichen (tlw.) aus der Besteuerung entziehen können. Mit § 158 AO stellt das steuerliche Verfahrensrecht hierfür zwei voneinander unabhängige (s. den Abschnitt „unabhängig von der Ordnungsmäßigkeit“ weiter unten) Voraussetzungen an die „Beweiskraft der Buchführung (und Aufzeichnungen)“, damit nicht die Beweislastumkehr eintritt und eventuell die Schätzungsnotwendigkeit nach § 162 Abs. 2 S. 2 AO ausgelöst wird:

1. die Ordnungsmäßigkeit gem. §§ 140 bis 148 AO und
 2. das Nichtvorhandensein eines „Anlass[es ...], die sachliche Richtigkeit zu beanstanden“.
- In der Konsequenz der *ersten* Vertrauensvoraussetzung löst eine unzureichende Mitwirkung im Besteuerungsverfahren – insbesondere „wenn der Steuerpflichtige Bücher oder Aufzeichnungen, die er nach den Steuergesetzen zu führen hat, nicht vorlegen kann“⁷⁴⁰ – die Amtspflicht der Schätzung aus, weil „der ‚Beweisverderber‘ oder ‚Beweisvereitler‘ aus seinem Verhalten keinen Vorteil ziehen darf (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO).“⁷⁴¹

Als *zweite* Vertrauensvoraussetzung hat die Gesetzgebung für das Besteuerungsverfahren mit dem „**Beanstandungsanlass**“ ein eigenes, abgemildertes (s. den Abschnitt „Konkretisierung der Aussagegewissheit“ weiter unten) Beweismaß bestimmt. Nach der allgemeinen Gesetzesformulierung gilt dieses ohne Einschränkungen für alle Sachverhalte, für welche die Finanzverwaltung grundsätzlich die Beweislast träge: Die „gesetzliche Vermutung [der sachlichen Richtigkeit von Besteuerungsgrundlagen ...] verliert ihre Wirksamkeit mit der Folge der Schätzungsnotwendigkeit nach § 162 AO, wenn es nach [einer] Verprobung [...] unwahr-

⁷³⁷ Nachzulesen unter: <https://www.stifterverband.org/data-literacy-education>.

⁷³⁸ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

⁷³⁹ BFH-Urteil v. 16.9.2021, BStBl. II 2022, 101.

⁷⁴⁰ § 162 Abs. 2 S. 1-2 AO.

⁷⁴¹ BFH v. 15.2.1989, BStBl. II 1989, 462: „[...] Zur Vermeidung eines solchen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

scheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt⁷⁴². Auf diese Weise wird der Gesetzauftrag des Verifikationsprinzips umgesetzt: Sowohl im Einzelfall als auch mit Blick auf die sehr große Anzahl an zu prüfenden Unternehmen in Deutschland⁷⁴³ wird die Feststellungslast der Finanzbehörde begrenzt, „wenn ein entscheidungserheblicher Sachverhalt trotz Ausschöpfung aller zugänglichen und zumutbaren Ermittlungsmöglichkeiten nicht vollständig aufgeklärt werden kann“⁷⁴⁴, damit die erreichbare Prüfungsbreite präventiv die Gleichmäßigkeit der Besteuerung gem. § 85 AO sicherstellen kann. Würden die fehlende Aufklärbarkeit des Einzelfalls in angemessener Zeit bzw. eine zu punktuelle Überprüfung einzelner Stpfl. dies in Frage stellen, könnte der intensive Eingriff einer Außenprüfung an sich bereits als staatliche Willkür empfunden werden.⁷⁴⁵ „Die gesetzliche Regelung des § 158 gründet sich insofern auf die Einsicht, dass es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild der Buchführung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann“.⁷⁴⁶

Das Beweismaß „Beanstandungsanlass“ gilt **unabhängig von der Ordnungsmäßigkeit** der Buchführung und Aufzeichnungen, denn es richtet sich laut Gesetz ausdrücklich an „[d]ie Buchführung und die Aufzeichnungen des Steuerpflichtigen, die den Vorschriften der §§ 140 bis 148 [AO] entsprechen“; dies ist durch die Neufassung⁷⁴⁷ des § 158 AO nochmals deutlicher geworden: „(2) Absatz 1 gilt *nicht* [...]“. Dennoch senkt das Zusammentreffen von Ordnungswidrigkeit und materiellen Mängeln das Beweismaß zur sachlichen Verwerfung erklärter Besteuerungsgrundlagen als Indiziengesamtheit⁷⁴⁸ zusätzlich ab (s. den Abschnitt „Indiziengesamtheit“ unter [Pkt. 4.4.](#)).

Eine nähere **Konkretisierung der Aussagegewissheit** „Beanstandungsanlass“ findet sich im Gesetz nicht. Durch die Normierung wird zunächst nur klar, dass der Bean-

⁷⁴² Anwendungserlass zu § 158 AO.

⁷⁴³ Laut Monatsbericht des BMF für den Oktober 2023 mit der Betriebsprüfungsstatistik 2022 waren 8.409.661 Betriebe in der Betriebskartei der Finanzämter erfasst und sollten von bundesweit 12.949 Prüfer*innen – also knapp 650 Betriebe pro Prüfer*in – intensiv geprüft werden; nachzulesen unter: https://www.bundesfinanzministerium.de/Monatsberichte/2023/10/Inhalte/Kapitel-3-Analysen/3-2-steuerliche-betriebspruefung-2022-pdf.pdf?__blob=publicationFile&v=4.

⁷⁴⁴ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 68.

⁷⁴⁵ Tipke, „Die Steuerrechtsordnung“, Verlag Dr. Otto Schmidt, Bd. III, 1407.

⁷⁴⁶ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5.

⁷⁴⁷ Anzuwenden ab 1.1.2023.

⁷⁴⁸ BGH v. 27.4.2010 – 1 StR 454/09, NStZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweismwürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreich[en].“

standungsanlass als eigenständiges Beweismaß vom „Beweis“ *eindeutig* abzugrenzen ist⁷⁴⁹, weil besonders über das Strafrecht unterschiedliche Rechtsfolgen ausgelöst werden (können). Der beschriebenen Gesetzesintention folgend, das „non liquet“ des Einzelfalls nicht zu Lasten der Gemeinschaft der Steuerehrlichen zu werten, wird das reduzierte Beweismaß in den allgemein titulierten „Verwaltungsgrundsätze[n] 2020“ so konkretisiert: „Voraussetzung einer Einkünfteberichtigung ist vielmehr, [...] dass der von der Finanzbehörde ermittelte Verrechnungspreis zumindest wahrscheinlicher ist.“⁷⁵⁰ Grundsätzlich plädieren die vier Autoren einer vom Bundeskriminalamt herausgegebenen Arbeit mit dem Untertitel „Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“ für eine *zahlenmäßige* Verdeutlichung von Aussagegewissheiten, welche in Gerichtsverfahren (und Prüfungen) regelmäßig verwendet werden: „Die numerische Schätzung bewirkt insbesondere eine Vermeidung von Missverständnissen oder möglichen Unterschieden in den scheinbar gleichen Größenordnungen verschiedener Individuen bei identischen verbalen Ausdrücken. [...] Sie erfordert [...] schlicht eine präzisere Einschätzung. Sie erhöht die intersubjektive Beurteilbarkeit dessen, was gemeint ist.“⁷⁵¹ Ohne numerische Konkretisierung ist also nach deren Erfahrungen keine Gleichmäßigkeit und Nachvollziehbarkeit von Urteilen gegeben. Auch der Untersuchungsgrundsatz für das Besteuerungsverfahren „§ 88 AO“ (analog § 76 FGO) und der Schätzungszusammenhang „§ 158 Abs. 2 Nr. 1 i.V.m. § 162 Abs. 2 S. 2 AO“ können ohne möglichst konkrete Einordnung des gesetzlichen Beweismaßes „Beanstandungsanlass“ nicht einheitlich und nachvollziehbar ausgelegt werden.

Als sinnvoller Ausgangspunkt zur Konkretisierung der Aussagegewissheit „Beanstandungsanlass“ ergibt sich die erforderliche (eindeutige) **Abgrenzung vom Beweis**, weil besonders über das Strafrecht unterschiedliche Rechtsfolgen vorgesehen sein können. Diese ist lediglich nach unten möglich, denn ein Beweis ist *im besten Fall* 100 % sicher; hierbei muss aber zwischen dem naturwissenschaftlichen und dem juristischen Verständnis des Beweises unterschieden werden:

- ! Beweise im naturwissenschaftlichen Sinn verlangen eine strenge Kausalität, also eine zweifelsfreie Ursache-Wirkungs-Beziehung. Durch sorgfältige Beobachtungen der *Gegenwart* kann diese (nahezu) 100 %-ige Klarheit erreicht werden.

⁷⁴⁹ Hätte der Gesetzgeber mit dem „Beanstandungsanlass“ einen Beweis gemeint, wäre die gesetzliche Formulierung redundant und damit unnötig, denn § 158 Abs. 2 Nr. 1 AO würde dann ausdrücken: „Bewiesenermaßen falsche Besteuerungsunterlagen haben keine Beweiskraft.“

⁷⁵⁰ BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 73.

⁷⁵¹ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Reihe „Polizei + Forschung“, Bd. 26, herausgegeben vom Bundeskriminalamt (BKA), Kriminalistisches Institut, Luchterhand Verlag, 2004, 13.

! Für einen **Beweis im juristischen Sinn** als volle richterliche Überzeugung muss ein geringerer Maßstab gelten, weil sich die Vergangenheit nicht vollständig und objektiv rekonstruieren lässt: „[...] wobei der Richter nicht eine von allen Zweifeln freie Überzeugung anstreben darf, sich in tatsächlich zweifelhaften Fällen vielmehr mit einem für das praktische Leben brauchbaren Grad von Gewissheit überzeugen muss“⁷⁵². Dokumentationen, Aussagen und beobachtende Rückschlüsse aus der Gegenwart auf die Vergangenheit können immer nur ein unvollständiges Abbild unter Unsicherheit herstellen, was laut Wissenschaft zu dem Schluss führt: „Die juristische Methodenlehre ist somit die Lehre von der richtigen und gerechten Entscheidung unter Unsicherheitslagen.“⁷⁵³ *In der Literatur werden diese unvermeidbaren Ungenauigkeiten mitunter generell abgelehnt, wenn z.B. gefordert wird: „Die Beweiskraft der Buchführung (§ 158 AO) kann [...] nur bei schweren [Ordnungs]mängeln oder klaren [100 % sicheren] Fehlern im Detail außer Kraft gesetzt werden“⁷⁵⁴, wodurch eine objektiv teleologische Eingrenzung des Beweises im juristischen Sinn und des Beanstandungsanlasses unmöglich wird; demgegenüber steht in der Wissenschaft die fehlende 100 %-Beweismöglichkeit für die Vergangenheit außer Frage: „Prüfungen vergangener Aufzeichnungen können, nach dem logischen Menschenverstand, nur zu Wahrscheinlichkeitsaussagen führen. Der Grund dafür ist, dass es sich bei Auf[ge]zeichnungen um ein reduziertes Abbild komplexer Daten und Zusammenhänge handelt. Steuerliche Unterlagen sind gleich zu setzen mit Zusammenfassung und Dokumentation vergangener Vorkommnisse und fallen daher in diesen Bereich. Schon durch den Begriff ‚Zusammenfassung‘ wird deutlich, dass die Vergangenheit nicht vollständig rekonstruiert werden kann. Schlussfolgernd kann bei der Prüfung von (steuerlichen) Daten keine (100 %) Sicherheit und somit kein Beweis vorliegen bzw. erfolgen.“⁷⁵⁵ Der BGH hat in einer strafrechtlichen Bestimmung des Schuldumfangs „die ermittelte Vertrauensuntergrenze [...] mit einer Wahrscheinlichkeit von 99,5 % „zutreffend“⁷⁵⁶ genannt. Diese konkretisierende Einordnung der „juristischen“ Beweisklarheit steht in Übereinstimmung mit wissenschaftlichen Bewertungen moderner Beweismittel wie DNA- und Fingerabdruckanalysen, welche nach (Meta-)Studien eine geschätzte Restunsicherheit von 0,5-1 % hinterlassen⁷⁵⁷.*

⁷⁵² BFH v. 20.5.2010, BStBl II 2010, 1022, Rz. 13.

⁷⁵³ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 145.

⁷⁵⁴ Prinz, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4, M5.

⁷⁵⁵ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 323-324.

⁷⁵⁶ Urt. v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 26.

⁷⁵⁷ Ausführlich hierzu: Giezek/Rupprecht/Wähnert, „Wie sicher muss ein Verprobungsergebnis sein? – Anleihen aus der Kriminalistik zur sinnvollen Einordnung des Beanstandungsanlasses“, BBK 2017, 236.

Auf Basis dieser umschreibenden und zahlenmäßigen Klarstellung des Beweises im juristischen Sinn (als volle richterliche Überzeugung) kann die gesetzlich geforderte Gewissheit „Beanstandungsanlass“ aus § 158 Abs. 2 Nr. 1 AO konkreter ein- und abgegrenzt werden.

Verbale Wahrscheinlichkeiten und deren numerische Zuordnungen

mit indifferenter Wahrscheinlichkeit (non liquet)	1–1/2	50 %
mit leicht überwiegender Wahrscheinlichkeit	1–1/4	75 %
mit überwiegender Wahrscheinlichkeit	1–1/10	90 %
mit hoher Wahrscheinlichkeit	1–1/20	95 %
mit sehr hoher Wahrscheinlichkeit	1–1/100	99 %
mit an Sicherheit grenzender Wahrscheinlichkeit	1–1/10.000	99,99 %

Nach der abgebildeten tabellarischen Übersetzung wichtiger Wahrscheinlichkeitsaussagen in numerische Wahrscheinlichkeiten durch erfahrene Gutachter⁷⁵⁸ lässt sich aus der höchstrichterlichen Rechtsprechung i.V.m. gutachterlichen Erfahrungen ableiten, dass das gesetzliche Beweismaß „**Beanstandungsanlass**“ nur als Aussagegewissheit von maximal **ca. 95 %** („mit hoher Wahrscheinlichkeit“) *eindeutig* von den beiden höchsten Stufen des Beweises im juristischen Sinn abgegrenzt werden kann.

Die Konkretisierung des Beanstandungsanlasses auf ca. 95 % deckt sich mit **anderen Lebens- und Rechtsbereichen**, in denen Wahrscheinlichkeitsaussagen notwendig sind und deshalb regelmäßig vorkommen: die industrielle Qualitätssicherung⁷⁵⁹, die Absicherung von wissenschaftlichen Studienergebnissen⁷⁶⁰, die Bewertung großer Datenmengen mit Hilfe von Stichproben⁷⁶¹ und die automatische Prozesssteuerung⁷⁶². Aus mathematischer Sicht beschreiben diese „ca. 95 %“ den Normalverteilungsbereich „Mittelwert ± doppelte Standardabweichung“⁷⁶³ des (natürlichen) Streuverhaltens, wodurch alle normalen Daten bzw.

⁷⁵⁸ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Reihe „Polizei + Forschung“, Bd. 26, herausgegeben vom Bundeskriminalamt (BKA), Kriminalistisches Institut, Luchterhand Verlag, 2004, 33; die Autoren begründen die nicht-lineare Abstufung mit „der Erkenntnis, dass eine logarithmische Beziehung zwischen Information und Wahrscheinlichkeit besteht“ (S. 32 f.), womit gemeint ist, dass für weitere Steigerungen der Klarheit einer beschriebenen Situation zunehmend mehr Informationen notwendig sind.

⁷⁵⁹ DIN ISO 2859-1:2014-08: Annahmestichprobenprüfung anhand der Anzahl fehlerhafter Einheiten oder Fehler (Attributprüfung), nachzulesen unter: <https://www.beuth.de/de/norm/din-iso-2859-1/191697831>.

⁷⁶⁰ Vgl. z.B.: Bortz/Schuster, „Statistik für Human- und Sozialwissenschaftler“, Springer-Verlag, 2010, 11.

⁷⁶¹ § 241 Abs. 1 HGB: „Bei der Aufstellung des Inventars darf der Bestand der Vermögensgegenstände nach Art, Menge und Wert auch mit Hilfe anerkannter mathematisch-statistischer Methoden auf Grund von Stichproben ermittelt werden“ sowie BGH v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 20 ff.

⁷⁶² Vgl. z.B.: Richter, „Prinzipien der Künstlichen Intelligenz“, Teubner-Verlag, 1992.

⁷⁶³ Die Standardabweichung beschreibt die Streuungsintensität einer Datenmenge als vorzeichenneutrale typische Abweichung jedes einzelnen Datensatzes zum Mittelwert.

Situationen *zuzüglich* der schwachen Ausreißer toleriert und ausschließlich starke Ausreißer abgelehnt werden.

Die **BFH-Auslegung** des gesetzlichen Beweismaßes „Beanstandungsanlass“ als Aussagegewissheit „**mit an Sicherheit grenzender Wahrscheinlichkeit**“⁷⁶⁴ passt nicht zu diesen logisch⁷⁶⁵ abgeleiteten Verhältnissen: Sowohl im allgemeinen als auch wissenschaftlichen Sprachgebrauch wird hierunter eine Aussageklarheit von nahezu 100 % verstanden, wie es auch die Tabelle „Verbale Wahrscheinlichkeiten und ihre numerischen Zuordnungen“ oberhalb wiedergibt. Nicht einmal *Beweise* im juristischen Sinn (als volle richterliche Überzeugung) können diese Anforderung regelmäßig erfüllen. Insofern wäre Rechtsprechung an dem Maßstab „mit an Sicherheit grenzender Wahrscheinlichkeit“ weitgehend unmöglich, weil sich die Vergangenheit mit den begrenzten Informationen nicht bis zu dieser Klarheit wiederherstellen lässt.

(Verprobende) Beurteilungen von Gesamtbildern wären unter diesem Beweismaß sinnlos, weil es zu komplexen Situationen „[s]imple Ursache-Wirkungs-Beziehungen [...] *nicht* in der Wirklichkeit [gibt]“⁷⁶⁶. In der Konsequenz dürften Verprobungen⁷⁶⁷ in der Prüfung von Besteuerungsgrundlagen *grundsätzlich* nicht anerkannt **und verwendet** werden, weil sie das erforderliche Beweismaß nicht erreichen **und somit den Prüfungsgrundsätzen**⁷⁶⁸ **nicht entsprechen** können. Tatsächlich finden sich aber zahllose bestätigende Finanzgerichtsurteile zu Vermögenszuwachs- bzw. Geldverkehrsrechnungen und Kalkulationen, welche aufgrund der unvollständigen und nur bedingt verifizierbaren Ausgangsgrößen *keine* Ergebnisklarheit „mit an Sicherheit grenzender Wahrscheinlichkeit“ (99,99 %) darlegen⁷⁶⁹.

⁷⁶⁴ BFH v. 24.6.1997, BStBl. II 1998, 51: „Nur wenn die Würdigung des Sachverhalts ergibt, daß eine formell ordnungsmäßige Buchführung mit an Sicherheit grenzender Wahrscheinlichkeit sachlich unrichtig ist, kann das Ergebnis der Buchführung ganz oder teilweise verworfen werden.“

⁷⁶⁵ BGH v. 24.3.2016 – 2 StR 112/14, NStZ 2016, 490, Leitsätze Nr. 5: „Die tatrichterliche Würdigung darf allerdings den Gesetzen der *Logik* und dem gesicherten wissenschaftlichen Erfahrungswissen nicht widersprechen (vgl. BGH NStZ 1995, 590, 592).“

⁷⁶⁶ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 15.

⁷⁶⁷ Anwendungserlass zu § 158 AO: „Die Vorschrift enthält eine gesetzliche Vermutung. Sie verliert ihre Wirksamkeit mit der Folge der Schätzungsnotwendigkeit nach § 162 AO, wenn es nach [einer] *Verprobung* [...] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt.“

⁷⁶⁸ § 7 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen. Ihre Dauer ist auf das notwendige Maß zu beschränken. Sie hat sich in erster Linie auf solche Sachverhalte zu erstrecken, die zu endgültigen Steuerausfällen oder Steuererstattungen oder -vergütungen oder zu nicht unbedeutenden Gewinnverlagerungen führen können.“

⁷⁶⁹ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324: „Die herkömmlichen [Verprobungs]methoden basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...]. Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten in den wiederhergestellten Daten. [...] Schlussfolgernd stellt sich die Frage, ob das geforderte Sicher-

Trotz dieses unverkennbaren Widerspruchs wird die offenkundig unzutreffende – weil im Gegensatz zur Gesetzesintention *beweisverschärfende* – Auslegung des Beanstandungsanlasses nach § 158 Abs. 2 Nr. 1 AO als Aussagegewissheit „mit an Sicherheit grenzender Wahrscheinlichkeit“ u.a. in Kommentaren⁷⁷⁰ übernommen. In Beiträgen wird hingegen jüngst zu einer korrekten Einordnung des Beanstandungsanlasses angehalten: „Widerlegen der Richtigkeitsvermutung [...] In der Literatur wird deshalb zu Recht ein Gewissheitsrahmen von 90-95 % für ausreichend gehalten. Das entspricht dem für das finanzgerichtliche Verfahren allgemein geltenden Grundsatz, dass der Richter keine von allen Zweifeln freie Überzeugung anstreben darf, sondern sich in tatsächlich zweifelhaften Fällen vielmehr mit einem für das praktische Leben brauchbaren Grad von Gewissheit begnügen muss.“⁷⁷¹

4.7. Einleitung des Strafverfahrens nach § 10 BpO

Nach § 152 Abs. 2 StPO muss jedes Strafverfolgungsorgan, wozu gem. § 386 Abs. 1 AO auch die Finanzbehörde gehört, beim Vorliegen eines Anfangsverdachts das Strafverfahren einleiten. Das **Legalitätsprinzip** lässt dahingehend keine Abwägung zu, sondern erteilt den Befehl, bei „zureichende[n] tatsächliche[n] Anhaltspunkte[n] für eine Straftat“ einzuschreiten, denn die Einleitung eines Steuerstrafverfahrens bewirkt eine Veränderung der rechtlichen Stellung der Stpfl. als nunmehr Beschuldigte eines Steuerstrafverfahrens.⁷⁷² Einerseits werden dadurch die Ermittlungsmöglichkeiten der Finanzverwaltung um strafprozessuale Maßnahmen erweitert.⁷⁷³ Andererseits „bestehen [gegenüber dem Besteuerungsverfahren] deutliche Unterschiede im Hinblick auf die Rechte, die dem Beschuldigten zukommen“⁷⁷⁴. Insbesondere muss er sich nicht selbst belasten, sodass die Einhaltung seiner Mitwirkungspflichten im Besteuerungsverfahren nicht mehr mit Zwangsmitteln durchsetzbar

heitsmaß [„mit an Sicherheit grenzender Wahrscheinlichkeit“] bei den herkömmlichen Methoden überhaupt erreicht werden kann.“

⁷⁷⁰ Vgl. z.B.: Hübschmann/Hepp/Spitaler, AO/FGO, 242. Lieferung 05.2017, § 158 AO, Rz. 5: „Nach der Rspr. kann eine formell ordnungsmäßige Buchführung ganz oder teilweise verworfen werden, wenn die Würdigung des Sachverhalts mit an Sicherheit grenzender Wahrscheinlichkeit ergibt, dass das Ergebnis der Buchführung sachlich unrichtig ist.“ Gosch, Abgabenordnung/Finanzgerichtsordnung, 161. Lieferung Juni 2021, § 158 AO, Rz. 12: „Eine formell ordnungsgemäße Buchführung kann somit letztlich nur dann ganz oder teilweise verworfen werden, wenn die Würdigung des Sachverhaltes mit an Sicherheit grenzender Wahrscheinlichkeit ergibt, dass das Buchführungsergebnis sachlich nicht richtig ist.“ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14: „Die Beanstandungen der sachlichen Richtigkeit durch die FinBeh. müssen mithin so konkret und glaubwürdig (fundiert) sein, dass sie das Ergebnis der Buchführung außerhalb der Wahrscheinlichkeit liegend erscheinen lassen“.

⁷⁷¹ Brinkmann, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Aufl. 2023, Rz. 559.

⁷⁷² § 393 Abs. 1 AO: „Die Rechte und Pflichten der Steuerpflichtigen und der Finanzbehörde im Besteuerungsverfahren und im Strafverfahren richten sich nach den für das jeweilige Verfahren geltenden Vorschriften.“

⁷⁷³ Wenzel, „Die Schätzung im Steuerstrafrecht – Teil 1: Grundlagen der fiskalischen Schätzung“, NWB 51/2023, 3505, 3506: „Aufgrund der differenzierten Ermittlungsrechte unterscheiden sich die Befugnisse und Möglichkeiten der Außenprüfung und der Steuerfahndung signifikant.“

⁷⁷⁴ Weber, „Steuerfahndung – Steuerstrafverteidigung“, NWB-Verlag, 3. Auflage 2016, 462.

ist. Deshalb führt ein „Verstoß gegen Belehrungspflichten über die Aussagefreiheit des Beschuldigten [...] im Strafprozeß grundsätzlich zu einem Verwertungsverbot der unmittelbar hierdurch erlangten Informationen“⁷⁷⁵. Für das Besteuerungsverfahren gilt dies jedoch nicht: „Ein Verwertungsverbot im Besteuerungsverfahren bei einer Verletzung des § 393 Abs. 1 Satz 4 AO [...] ergibt sich auch nicht aus allgemeinen verfassungsrechtlichen, insbesondere rechtsstaatlichen Grundsätzen. Vor allem folgt aus der Änderung der Rechtsprechung des BGH, die nunmehr ein strafprozessuales Verwertungsverbot bei unterlassener Belehrung bejaht [...], kein steuerrechtliches Verwertungsverbot.“⁷⁷⁶

Ein **Anfangsverdacht** kann sich zum einen auf die Mitwirkungsbereitschaft beziehen, weil „der ‚Beweisverderber‘ oder ‚Beweisvereitler‘ aus seinem Verhalten keinen Vorteil ziehen darf“⁷⁷⁷, indem mangels (prüfbarer) Besteuerungsaufzeichnungen ansonsten auf die Strafverfolgung verzichtet werden müsste. Zum anderen wird eine Gewissheit für die Annahme einer Steuerstraftat festgelegt: „[N]ach kriminalistischer Erfahrung [muss] die Möglichkeit besteh[en], daß eine verfolgbare Straftat vorliegt“⁷⁷⁸. „Trotz der Einbeziehung der ‚kriminalistischen Erfahrung‘ in die Definition des Anfangsverdachts müssen konkrete tatsächliche Erkenntnisse die Basis [...] bilden, die über statistisch gewonnene Erkenntnis[se] bzgl. der Häufigkeit von Straftaten in bestimmten Lebenszusammenhängen und/oder Örtlichkeiten hinausgeh[en].“⁷⁷⁹ Bezogen auf die Betriebsprüfung muss das Strafverfahren eingeleitet werden, wenn es aufgrund konkreter Erkenntnisse zum Einzelfall wahrscheinlicher ist, dass eine Straftat vorliegt als dass keine stattgefunden hat.

Der gleichlautende Erlass der obersten Finanzbehörden der Länder „Verdacht einer Steuerstraftat, Unterrichtung der Bußgeld- und Strafsachenstelle: Die obersten Finanzbehörden der Länder haben zu Anwendungsfragen zu § 10 Abs. 1 BpO Stellung genommen“⁷⁸⁰ konkretisiert den Anfangsverdacht gem. § 152 Abs. 2 StPO anhand **möglicher Betriebsprüfungserkenntnisse**. Dazu werden die Positionen „keine tatsächlichen Anhaltspunkte“ und „tatsächliche Anhaltspunkte“ für typische Prüfungssituationen hinsichtlich der Ordnungsmäßigkeit und des materiellen Gesamtbilds gegenübergestellt:

⁷⁷⁵ BGH v. 16.6.2005 – 5 StR 118/05.

⁷⁷⁶ BFH v. 23.1.2002, BStBl. II 2002, 328.

⁷⁷⁷ BFH v. 15.2.1989, BStBl. II 1989, 462: „[...] Zur Vermeidung eines solchen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

⁷⁷⁸ BGH v. 21.4.1988 – III RZ 255/86, NSTZ 1988, 510.

⁷⁷⁹ *Webel*, „Steuerfahndung – Steuerstraferverteidigung“, NWB-Verlag, 3. Auflage 2016, 440 mit Verweis auf *Radtke*, in *Radtke/Hohmann*, StPO, § 152 Rn. 19.

⁷⁸⁰ BMF-Schreiben v. 31.8.2009, BStBl. I 2009, 829.

„Grundsätzlich <i>keine tatsächlichen</i> Anhaltspunkte [...] liegen insbesondere für folgende beispielhaft aufgezählte Fälle vor:	<i>Tatsächliche</i> Anhaltspunkte [...] werden insbesondere in folgenden beispielhaft aufgezählten Fällen gegeben sein:
Bei Aufdecken bloß formeller oder auch kleinerer materieller Buchführungsmängel; [...]	Bei Feststellung schwerwiegender Buchführungsmängel, [...]
Das bloße Durchführen von Kalkulationen und Verprobungen, wie Geldverkehrsrechnungen, Richtsatzverprobungen, Chi ² -Test, Zeit-Reihen-Vergleich usw., [...]	Nach durchgeführter Kalkulation oder Verprobung verbleiben ungeklärte Differenzen von einige[m] Gewicht, [...]"

In der ergänzenden Klarstellung zu „**schwerwiegende[n] Buchführungsmängel[n]**“ wird auf das „Fehlen von sonst üblichen Belegen“⁷⁸¹ abgestellt, wodurch die Überprüfung der Vollständigkeit (weitest gehend) verhindert wird. Die Ausführungen zur Anwendung der abstrakten Ordnungsvorgaben auf die heutigen, überwiegend digitalen Aufzeichnungs- und Aufbewahrungsweisen unter [Pkt. 4.1.](#) gelten gleichermaßen: „Der erkennende Senat vertritt hierzu die Auffassung, dass das Fehlen einer lückenlosen Dokumentation zur Kassenprogrammierung in seinen Auswirkungen auf die Beurteilung der formellen Ordnungsmäßigkeit der Buchführung und der Eröffnung der Schätzungsbefugnis dem Fehlen von Tagesendsummenbons bei einer Registrierkasse bzw. dem Fehlen täglicher Protokolle über das Auszählen einer offenen Ladenkasse gleichsteht. In allen drei Fällen lässt der formelle Mangel zwar keinen sicheren Schluss auf die Verkürzung von Einnahmen zu. Gleichwohl gibt es systembedingt keine Gewähr mehr für die Vollständigkeit der Erfassung der Bareinnahmen, ohne dass eine nachträgliche Ergänzung der Dokumentation bzw. eine anderweitige Heilung des Mangels möglich wäre.“⁷⁸² Auf die besondere Bedeutung der formellen Ordnungsmäßigkeit von Besteuerungsaufzeichnungen für das Vertrauen in die sachliche Richtigkeit elektronischer Betriebsverwaltungen (s. [Pkt. 2.1.](#)) weist auch die AO-Kommentierung hin: „Benutzt der buchführende Stpfl. jedoch eine Manipulationssoftware [...], fehlt es bereits an einer formell ordnungsgemäßen Buchführung, die einen Vertrauensvorschuss verdient“⁷⁸³.

Eine *monetäre* Bewertung der „**ungeklärten Differenzen von einige[m] Gewicht**“ nach einer verprobenden Gesamtbildprüfung verbietet das im Strafrecht geltende

⁷⁸¹ BMF-Schreiben v. 31.8.2009, BStBl. I 2009, 829.

⁷⁸² BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 27.

⁷⁸³ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 14.

Legalitätsprinzip. Mit den (begrenzten) Mitteln einer Außenprüfung sind zuverlässige Aussagen zum Umfang möglicher Steuerhinterziehungen regelmäßig nicht möglich; hierbei sind die „ausgefeilt geplante[n] Doppelverkürzungen und/oder [der] Einsatz von Manipulationssoftware, die die Aufdeckungsmöglichkeiten herkömmlicher Prüfungsmethoden minimieren“⁷⁸⁴ zu bedenken. Deshalb muss der Gradmesser für einen materiellen Anfangsverdacht die *Aussagegewissheit* sein, dass die vorgelegten Besteuerungsunterlagen inhaltlich nicht richtig sein können. Die Vorteile systematisch-mehrperspektivischer Prüfungsnetze wie der SRP wirken diesbezüglich ebenso:⁷⁸⁵

- ✓ Einzelne Indizien gegen die sachliche Richtigkeit von untersuchten Besteuerungsgrundlagen sind hinsichtlich möglicher nicht-manipulativer Ursachen wie gesamtbildrelevanter Sondersachverhalte soweit abgesichert worden, wie es die Datenqualität und Aufklärungsmitwirkung im Einzelfall zulässt (s. den Abschnitt „*einzelne verbleibende Beweiszeichen*“ unter [Pkt. 4.4.](#)).
- ✓ I.d.R. besteht ein zu beanstandendes SRP-Gesamtbild aus der beweisverstärkenden⁷⁸⁶ Gesamtheit von (gravierenden) Ordnungsmängeln und/oder mehreren Indizien (s. den Abschnitt „*mehrere Indizien gegen die sachliche Richtigkeit*“ unter [Pkt. 4.4.](#)).

In der Literatur⁷⁸⁷ wird die Bedeutung multiindizieller Beanstandungen nach einer SRP für die Einleitung des Strafverfahrens und darüber hinaus tlw. festgestellt: „Summarische Risikoprüfung und Schnittstellenverprobung: Beide Prüfungsverfahren stellen auf betriebswirtschaftli-

⁷⁸⁴ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 54.

⁷⁸⁵ So auch: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1, 4 u. 14 ff.: „Bei einer steuerlichen Außenprüfung können mathematisch-statistische (quantitative) Prüfungsmethoden eingesetzt werden, um die Besteuerungsgrundlagen des Steuerpflichtigen zu verproben und auf Plausibilität zu überprüfen. [...] Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit. Die Ergebnisse können darüber hinaus nach § 162 Absatz 2 Satz 2 der Abgabenordnung dazu führen, dass eine Schätzung durchgeführt werden muss. [...] 4. Summarische Risikoprüfung“.

⁷⁸⁶ BGH v. 27.4.2010 – 1 StR 454/09, NSZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreichen.“

⁷⁸⁷ Pump/Heidl, „Hilfestellung bei der Kassenführung durch den steuerlichen Berater – Ärger wegen § 162 AO durch unzulängliche Kassenbuchführung“, StBp 2014, 162, 163: „Der weitere Vorteil von SRP besteht darin, dass die finanzamtliche Beweisführung für die Hinzuschätzung auch im Steuerstrafverfahren gerichtsfest ist.“ Demgegenüber sieht Beyer („Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931) die SRP kritisch, begründet dies aber auf der Grundlage unzutreffender Beschreibungen wie z.B. der Fokussierung der SRP-Zeitreihenanalytik auf Aufschlagsätze und einzelmethodischer Schlussfolgerungen aus der nicht differenzierenden Rechtsprechung (im Gegensatz dazu FG Hamburg v. 31.10.2016 – 2 V 202/16, Rz. 48: „Bei dieser Methode [doppelskalierte Monatszeitreihe] tritt das zahlenmäßige Verhältnis zu Gunsten der Abhängigkeit voneinander in den Hintergrund. Wesentlich ist die Reaktion der Faktoren aufeinander. Aus kaufmännischer Sicht wird in einem Unternehmen auf jede Veränderung der Nachfragesituation (annähernd) adäquat mit einer Erhöhung bzw. Absenkung der Einkaufssumme reagiert“).

che und stochastische Gesichtspunkte ab, die zur Verprobung genutzt werden. Ihre Ergebnisse sind aufgrund ihrer umfassenden und systematischen mehrperspektivischen Absicherung ausreichend, um die Ordnungsmäßigkeit der Buchführung zu erschüttern. Darüber hinaus können die Ergebnisse einer Summarischen Risikoprüfung auch durchaus ausreichen, um einen strafprozessualen Anfangsverdacht zu begründen.“⁷⁸⁸

4.8. Schätzung im Besteuerungsverfahren (Quantilsschätzung)

„Mit § 162 AO hat sich der deutsche Gesetzgeber für die Fälle nicht möglicher Ermittlungen [von Besteuerungsgrundlagen] für die **Schätzung** entschieden, weil durch das alternative Außerachtlassen nicht berechenbarer und nicht ermittelbarer Besteuerungsgrundlagen keine gerechte Besteuerung erreicht werden könnte.“⁷⁸⁹ Demnach münden Mitwirkungsverletzungen, infolge derer „die Finanzbehörde die Besteuerungsgrundlagen nicht ermitteln oder berechnen kann“⁷⁹⁰, sowie die Aberkennung der „Beweiskraft der Buchführung (und Aufzeichnungen)“⁷⁹¹ gem. § 162 AO in die Amtspflicht der Schätzung.

Intention einer Schätzung ist es, die Steuern so festzusetzen, als wären die Besteuerungsgrundlagen ordnungsgemäß erklärt worden. Dazu „[steht d]ie Auswahl zwischen verschiedenen Schätzungsmethoden [...] grundsätzlich im pflichtgemäßen Ermessen des FA bzw. FG“⁷⁹². Mit dem erklärten „Ziel, die Besteuerungsgrundlagen durch Wahrscheinlichkeitsüberlegungen so zu bestimmen, dass sie der Wirklichkeit möglichst nahe kommen“⁷⁹³, hat der BFH eine **Drei-Stufen-Theorie**⁷⁹⁴ zur Anwendung der gängigsten Schätzungsme-

⁷⁸⁸ *Webel*, „Steuerfahndung – Steuerstraftverteidigung“, NWB-Verlag, 3. Auflage 2016, 124. So auch: *Wenzel* in *Sievert/Wenzel*, „Reicht das SRP-Ergebnis für einen Anfangsverdacht?“, PStR 1/2023, 20: „Hierdurch sind die Ergebnisse sehr aussagekräftig. Denn sie sind wegen der umfassenden Ermittlung genauer als die manuelle, vom Außenprüfer durchgeführte Buchprüfung. Den Strafverfolgungsbehörden bieten sich dadurch tiefe Einblicke in das wirtschaftliche und steuerliche Handeln des Unternehmers. Richtig gedeutet kann unproblematisch ein Anfangsverdacht hergeleitet werden“; auch in diesem Fall eines „Kontrovers“-Beitrags stützt die Kritikerin ihre SRP-Ablehnung auf unzutreffende Aussagen wie „SRP ist lediglich eine automatisierte Verprobungsmethode“, wohingegen diesem Handbuch unproblematisch zu entnehmen ist, dass auch in den SRP-Vorlagen keine automatisierten *Entscheidungen* getroffen werden, und zudem offenbart sie ein unzureichendes wahr-scheinlichkeitstheoretisches Verständnis mit der Behauptung „[m]ithilfe der Stochastik erzeugte mathematische Wahrscheinlichkeiten dürfen [für einen Anfangsverdacht gem. § 152 Abs. 2 StPO] nicht als ausreichend betrachtet werden“, denn (nahezu) jeder juristische Beweis und alle indiziellen Sachverhaltsdarstellungen sind solche Wahrscheinlichkeitsaussagen ohne absolute Sicherheit.

⁷⁸⁹ *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Aufl. 2023, Rz. 1.

⁷⁹⁰ § 162 Abs. 2 S. 1 AO.

⁷⁹¹ § 158 Abs. 2 AO.

⁷⁹² BFH v. 3.9.1998, BFH/NV 1999, 290.

⁷⁹³ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 60.

⁷⁹⁴ *Nöcker (RiBFH)*, „Update Betriebsprüfung – Neues vom BFH und die Auswirkungen für die Praxis“, NWB 42/2016, 3157, 3165: „Stufe 1: Die Buchführung ist formell bis auf geringe Mängel korrekt. Folge: Eine Schätzung ist nur bei nachgewiesener materieller Unrichtigkeit der Buchführung möglich. Schätzungsmethoden: Erlösnachkalkulation (Ausbeute- und Aufschlagskalkulation)[,]Geldverkehrsrechnung[und]Vermögenszuwachsrechnung[.]Die aufgedeckten Differenzen können im Rahmen der Schätzung verwendet werden.“

thoden aufgestellt. Dabei stellt sich zum einen die Frage, wieso der BFH zwischen Stufe 2 „formell nicht korrekt, eine materiell fehlerhafte Buchführung ist aber (bislang) nicht konkret nachgewiesen worden“ und Stufe 3 „materiell nicht ordnungsgemäß und folglich auch formell nicht ordnungsgemäß“ unterscheiden will, wenn die betreffenden RiBFH selbst betonen: „Die ständige höchstrichterliche Rechtsprechung schränkt diese allein auf formellen Buchführungsmängeln beruhende Schätzungsbefugnis aber insoweit ein, als sie allein Anlass geben, die sachliche Richtigkeit der Buchführungsergebnisse anzuzweifeln“⁷⁹⁵; derweil formelle Mängel also nur dann schätzungsauslösend sind, wenn sie die materielle Prüfung (in angemessener Zeit) *verhindern*, kann das Auffinden konkreter materieller Unrichtigkeiten nur zufällig geschehen und darf nach Ansicht des Autors deshalb keine weitere Eskalationsstufe begründen. Zum anderen wird von einer wissenschaftlichen Autorin die Grundlage für die generelle Bevorzugung herkömmlicher Schätzungsmethoden hinterfragt: „Bei diesem [Drei-Stufen-]Modell stellt sich die Frage, auf welcher Grundlage die Reihen- bzw. Rangfolge basiert. Durch diese Rangfolge kann angenommen werden, dass die herkömmlichen Methoden anscheinend als sicherer angesehen werden, wenn als Rangfolgekriterium die naheliegende Sicherheit der Methoden gewählt wird. Allerdings liegt keine Beschreibung vor, bei der die Sicherheit der Methoden begründet wird bzw. es liegen keine Argumente vor, die für die sicherere Anwendung der herkömmlichen im Gegensatz zu den neuen, mathematisch-statistischen Methoden sprechen. Wie lässt sich demnach diese Rangfolge rechtfertigen?“⁷⁹⁶ Hier scheint die *direkt*ere Berechnung mit *scheinbar eindeutigerem* Ergebnis – allerdings auf der Basis von i.d.R. unvollständigen, unsicheren Ausgangsdaten – zur *gefühlsmäßigen* Bevorzugung zu verleiten,⁷⁹⁷ denn der X. Senat berücksichtigt die üblichen Ermittlungsprobleme

Stufe 2: Die Buchführung ist formell nicht korrekt, eine materiell fehlerhafte Buchführung ist aber (bislang) nicht konkret nachgewiesen worden. Folge: Eine Schätzung ist aufgrund der folgenden, auf betriebsinternen Daten oder individuellen Verhältnissen des Steuerpflichtigen basierenden Schätzungsmethoden möglich. Schätzungsmethoden: Erlösnachkalkulation (Ausbeute- und Aufschlagskalkulation)[,]Geldverkehrsrechnung[,]Vermögenszuwachsrechnung[und]Zeitreihenvergleich (wenn Steuerpflichtiger nicht mitwirkt oder komplexe finanzielle Verhältnisse vorliegen)[.]Die aufgedeckten Differenzen können im Rahmen der Schätzung verwendet werden, sind jedoch vom Finanzamt – wie auch vom Finanzgericht – von Amts wegen auf Plausibilität anhand der besonderen betrieblichen Verhältnisse des Steuerpflichtigen zu überprüfen, soweit diese bekannt sind. Bei verbleibenden Zweifeln können Sicherheitsabschläge geboten erscheinen. Stufe 3: Die Buchführung ist materiell nicht ordnungsgemäß und folglich auch formell nicht ordnungsgemäß. Folge: Eine Schätzung ist vorzunehmen, die zu einem möglichst genauen Ergebnis kommt, wobei die folgenden Schätzungsmethoden möglich erscheinen: Erlösnachkalkulation (Ausbeute- und Aufschlagskalkulation)[,]Geldverkehrsrechnung[,]Vermögenszuwachsrechnung[und]Zeitreihenvergleich[.]Folge: Die aufgedeckten Differenzen können im Rahmen der Schätzung verwendet werden.“

⁷⁹⁵ Vgl. Nöcker (RiBFH), „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548, 3550.

⁷⁹⁶ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 325.

⁷⁹⁷ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Nomos, 2018, 56-57 u. 65: „Bei der Verfügbarkeitsheuristik schätzen Individuen die [...] Wahrscheinlichkeit für das Eintreten eines Ereignisses unter Unsicherheit danach ein, wie leicht bzw. schwer

der Prüfungspraxis i.R. herkömmlicher, arbeitsaufwendiger⁷⁹⁸ Schätzungsmethoden nicht hinreichend; mindestens diese führen dazu, dass Geldverkehrs- bzw. Vermögenszuwachsrechnungen sowie Kalkulationen besonders in typischen Schätzungsfällen kein genau(er)es Schätzungsergebnis liefern können:

- ! Für die gängigen zwei Formen der Deckungsrechnung liegen mangels Aufzeichnungsverpflichtungen keine Informationen zu privaten Geldzuflüssen und -abflüssen vor. Außerdem ist zu bedenken: „Wenig verständlich ist in diesem Zusammenhang der Hinweis des BFH auf die Möglichkeit des Kontenabrufs. Denn erfahrungsgemäß wird ‚Schwarzgeld‘ in solche Länder verbracht, die an den entsprechenden Verfahren nicht teilnehmen, oder aber es wird mit Hilfe von Familienangehörigen oder Bekannten bzw. Familienmitgliedern versteckt.“⁷⁹⁹ Darüber hinaus kann die nach Prüfungserfahrungen verbreitete Verwendung von un versteuerten Einnahmen für die Bezahlung von „Schwarzlöhnen“ mit den Mitteln einer Außenprüfung in den meisten Fällen weder ausgeschlossen noch nachgewiesen werden.
- ! Die für Kalkulationen erforderlichen Ausgangsgrößen wie „Wareneinkauf“, „Warenverderb“, „Produktpreise bzw. -aufschlagsätze“, „Sonderpreise bzw. -aufschlagsätze“ und „Mengenanteile zu Sonderpreisen“ werden selbst in ordnungsmäßigen Besteuerungsaufzeichnungen nicht vollständig erfasst und finden sich erfahrungsgemäß in denen von Schätzungsfällen nur lückenhaft wieder. Durch die Multiplikation von Abweichungen beim Ansatz unrichtiger oder geschätzter Ausgangsgrößen führen Kalkulationen in einer Vielzahl der Fälle zu verstärkten „Hebelwirkungen“⁸⁰⁰. In Anbetracht dieser Unsicherheiten schließt *Schütt*: „Erwähnenswert ist zudem, dass das Problem der nicht Anwendbarkeit der herkömmlichen Methoden regelmäßig vorliegt, da bspw. nicht genügend Unterlagen für eine [seriöse] Rekonstruktion zur Verfügung stehen.“⁸⁰¹

Im Hinblick auf die generelle Festlegung einer Rangfolge von Schätzungsmethoden sollte außerdem bedacht werden, dass hierdurch gezielte Unterwanderungsstrategien der systematischen Steuerhinterziehung erleichtert werden.

es ihnen fällt, damit verbundene Erscheinungen aus ihrem Gedächtnis abzurufen. Dabei werden Ereignisse für wahrscheinlicher erachtet, wenn sie sich leichter vorgestellt werden können und sich zu ihnen eine direkte Verfügbarkeit herstellen lässt. [...] Im Rahmen juristischer Entscheidungen scheint es unerlässlich, von diesen Entscheidungsverzerrungen Kenntnis zu haben, da eine verzerrte Entscheidung den Denkgesetzen der Logik widerspricht [...].“

⁷⁹⁸ Vgl. z.B.: BFH v. 21.2.1990, BFH/NV 1990, 683, Rz. 31.: „Die Geldverkehrsrechnung ist eine arbeitsaufwendige Schätzungsmethode.“

⁷⁹⁹ *Brinkmann*, „Die Begründung des Sicherheitszuschlags“, StBp 2018, 212, 216.

⁸⁰⁰ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 44 ff. zu „Hebelwirkungen“ im Zusammenhang mit Zeitreihenanalysen.

⁸⁰¹ *Schütt*, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324.

Brinkmann merkt zum *neuen* Schätzungsziel „der Wirklichkeit möglichst nahe“⁸⁰² an: „Bei allem Streben nach ‚Einzelfallgerechtigkeit‘ scheint etwas in den Hintergrund zu geraten, dass die Schätzung ihrer Natur nach gerade keine genaue Berechnung ist und auch nicht sein kann, sondern in erster Linie dazu dient, die Besteuerungsgrundlagen mindestens in einer solchen Höhe festzusetzen, dass kein Vorteil gegenüber demjenigen entsteht, der seinen Verpflichtungen nachgekommen ist.“⁸⁰³ Dementsprechend wird in der Kommentierung darauf hingewiesen, dass „[n]aturgemäß [...] bei der Schätzung nach Wahrscheinlichkeitsgrundsätzen eine Bandbreite möglicher Wertansätze (sog. **Schätzungsrahmen**) [besteht.]“⁸⁰⁴ Für die Feststellung des Schätzungsrahmens im Einzelfall und für einen sachgerechten Ansatz innerhalb dieses sind folgende **allgemeine Maßgaben** bindend: „Der Schätzungsrahmen ist umso größer, je ungesicherter das Tatsachenmaterial ist, auf dem die Schätzung basiert.“⁸⁰⁵ Eine aus dem Schätzungsrahmen abgeleitete Schätzung muss vor allem „insgesamt in sich schlüssig [...] wirtschaftlich vernünftig und möglich sein.“⁸⁰⁶ Damit das Verifikationsprinzip (s. [Pkt. 4.6.](#)) nicht unterwandert wird, darf „der ‚Beweisverderber‘ oder ‚Beweisvereiteler‘ aus seinem Verhalten keinen Vorteil ziehen [...] (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO)“⁸⁰⁷ und müssen gem. § 158 **Abs. 2 Nr. 1** i.V.m. § 162 Abs. 2 S. 2 AO nicht (in angemessener Zeit) auflösbare Unschlüssigkeiten analog eine Schätzung auslösen⁸⁰⁸, damit es „keinen Anreiz [gibt], sich weiterhin schätzen zu lassen“⁸⁰⁹. Folglich „ist das FA im allgemeinen nicht nur berechtigt, sondern verpflichtet, in dem gegebenen Schätzungsrahmen an die oberste Grenze zu gehen.“⁸¹⁰ „[E]ine überzogene Schätzung [...] hat [...] nur die Rechtswidrigkeit [...] nicht aber bereits ihre Nichtigkeit zur Folge. Etwas anderes ist nach dieser Rechtsprechung, der der Senat folgt, allenfalls zu erwägen, wenn [...] das FA [...] bewusst zum Nachteil des Steuerpflichtigen geschätzt hat.“⁸¹¹ Diese zusammengefassten Anforderungen an eine (Voll-)Schätzung finden sich entsprechend in den allgemein titulierten „Verwaltungsgrundsätze[n] 2020“ wieder.⁸¹²

⁸⁰² BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 60.

⁸⁰³ *Brinkmann*, „Die Begründung des Sicherheitszuschlags“, StBp 2018, 212, 218.

⁸⁰⁴ Tipke/Kruse, AO/FGO, 170. Lieferung 5/2022, § 162 AO, Rz. 44.

⁸⁰⁵ Tipke/Kruse, AO/FGO, 170. Lieferung 5/2022, § 162 AO, Rz. 44.

⁸⁰⁶ BFH v. 18.12.1984, BStBl. II 1986, 226.

⁸⁰⁷ BFH v. 15.2.1989, BStBl. II 1989, 462.

⁸⁰⁸ Anwendungserlass zu § 158 AO: „Die Vorschrift enthält eine gesetzliche Vermutung. Sie verliert ihre Wirksamkeit mit der Folge der Schätzungsnotwendigkeit nach § 162 AO, wenn es nach [einer] Verprobung [...] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt.“

⁸⁰⁹ FG Berlin-Brandenburg v. 30.6.2017 – 10 V 10219/16.

⁸¹⁰ BFH v. 9.3.1967, BStBl. III 1967, 349; auch gemäß RiBFH Nöcker „zulässig“, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 2015, 3548, 3549.

⁸¹¹ BFH v. 2.12.2000, BStBl. II 2001, 381.

⁸¹² BMF-Schreiben v. 3.12.2020, BStBl. I 2020, 1325, Rz. 75-77.

4.8.1. Durchführung einer Quantilsschätzung

Die **Quantilsschätzung** ist eine eigenständige, optionale Methode zur verallgemeinerbaren Ermittlung einer angemessenen Schätzungshöhe⁸¹³ mit Hilfe von statistischen und betriebswirtschaftlichen Standardwerkzeugen – also kein zwangsläufiger Abschluss einer (un-aufgeklärt) auffälligen SRP (für die Plausibilitätsprüfung i.S. des § 158 Abs. 2 Nr. 1 AO). Sie bietet die Möglichkeit, die sorgfältig erarbeiteten umfangreichen Erkenntnisse aus der systematisch-mehrperspektivischen SRP für die Feststellung einer sachgerechten Schätzungshöhe zu nutzen.⁸¹⁴ Erfahrungsgemäß ist die informelle Ausgangslage in typischen Fällen, in welchen eine (Voll-)Schätzung der Bareinnahmen nach § 162 Abs. 2 AO zu erfolgen hat (s. *das Beispiel unter Pkt. 4.4.*), so lückenhaft bzw. ungesichert, dass auch aufwendige(re) Deckungsrechnungen und Kalkulationen keine genaueren Ergebnisse erzielen können. Der gewählte Lösungsansatz ist dabei i.S. der Gleichmäßigkeit der Besteuerung nach § 85 AO allgemein übertragbar und lässt sich dennoch variabel auf die Verhältnisse im Einzelfall anpassen⁸¹⁵.

Weil Steuerhinterziehungen weder einheitlichen Motiven noch Abläufen folgen, sucht die Quantilsschätzung nicht – in unrealistischer Erwartung – nach einer allgemeingültigen (Zeitreihen-)Regel für die wahrscheinlichste Punktschätzung. Sie orientiert sich vielmehr mit ihren **Annahmen und Arbeitsschritten** an den logischen Notwendigkeiten, um aus (sorgfältigen) Zeitreihenanalysen mit Hilfe anerkannter Werkzeuge der Mathematik und Wirtschaftswissenschaften sowie unter Beachtung der allgemeinen Anforderungen an eine Schätzung einen sachgerechten Schätzungsrahmen (**Bandbreitenschätzung**) abzuleiten und innerhalb dieses eine angemessene Schätzungshöhe festzusetzen:

⁸¹³ Die Quantilsschätzung ist keine Verprobungsmethode, wie es tlw. in der Literatur unzutreffend dargestellt wird; vgl. z.B.: Bleschick, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353: „Zunächst unternimmt er eine sog. betriebswirtschaftliche Auswertung: Hierzu ermittelt er im Wesentlichen mithilfe des Zeitreihenvergleichs einen Rohgewinnaufschlagsatz. Dieser wird auf den gesamten Wareneinkauf des Steuerpflichtigen angewandt. Unter Außerachtlassung von 20 % der höchsten und ggf. von 20 % der niedrigsten Rohgewinnaufschlagsätze ermittelt er den aus seiner Sicht möglichen Umsatz (sog. ‚Quantilsschätzung‘). Liegt der für den Außenprüfer mögliche Umsatz oberhalb des erklärten Umsatzes, sieht der Außenprüfer hierin aufgrund betriebswirtschaftlicher Auffälligkeiten einen Klärungsbedarf.“

⁸¹⁴ Insofern unzutreffend: Kulosa, „Umgang mit Schätzungen in der Betriebsprüfung“, Teil 2, AgrB 1/2022, 13, 16: „Dabei werden monatliche Werte für den Rohgewinnaufschlagsatz ermittelt; allerdings ohne Unterscheidung zwischen Wareneinkauf und Wareneinsatz (dies bewirkt eine erhebliche Ungenauigkeit).“

⁸¹⁵ Wenzel, „Die Schätzung im Steuerstrafrecht – Teil 1: Grundlagen der fiskalischen Schätzung“, NWB 51/2023, 3505, 3512: „Es verbietet sich eine standardisierte oder pauschale Anwendung von Schätzungsmethoden, vielmehr sind diese situations-, anwender- und normabhängig einzusetzen. Dies gilt im Übrigen auch für die Quantilsschätzung bei SRP.“ Durch die vielen interaktiven Optionen der SRP-Vorlage BWA (s. Pkt. 4.2.9.) können auf der einen Seite einzelfallbezogene Besonderheiten i.R. einer Schätzung berücksichtigt werden, während auf der anderen Seite dennoch ein verallgemeinerbares Verfahren zur besseren Gleichmäßigkeit des Steuervollzugs in vergleichbaren Konstellationen beiträgt.

1. Die Quantilsschätzung setzt auf den betriebseigenen Daten zum Rohgewinnaufschlagssatz und/oder Wareneinsatz auf, um aus diesen das eigentliche Schätzungsergebnis hochzurechnen, und ist somit eine **interne Schätzungsmethode**. Diese werden den externen allgemein vorgezogen: „Vergleiche mit anderen Betrieben haben regelmäßig einen geringeren Beweiswert als interne Betriebsvergleiche“⁸¹⁶. Zusätzlich werden i.R. der Quantilsschätzung gem. § 162 Abs. 1 S. 2 AO „alle Umstände [des Einzelfalls ...] berücksichtig[t]“, die durch Vertiefung oder Mitwirkung konkret bekannt geworden sind und somit in die SRP-Vorlage BWA eingepflegt werden können (s. [Pkt. 4.2.9.](#)).
2. Bei jeder zeitreihenbasierten Schätzung muss eine Periode gewählt werden. Eine sinnvolle Entscheidung ist zwischen den als „unzutreffend“ erkannten Jahreswerten, die den Grund für die Schätzung bilden, und empfindlichen kurzen Zeiträumen mit einem üblicherweise sehr hohen bis unmöglichen Berichtigungsbedarf für aussagekräftige Ergebnisse (s. *das Beispiel zum Einfluss der Zeitreihenperiode auf die Stabilität der Ergebnisse im Abschnitt „Fuzzy-Theorien“ unter [Pkt. 4.2.2.](#)*) zu treffen. Die Quartalsverhältnisse (s. [Pkt. 4.2.4.](#)) sind in solchen Fällen erfahrungsgemäß den unzutreffenden Jahreswerten sehr ähnlich, weshalb sie sich für eine sachgerechte Schätzung nicht eignen. **Monatsgrößen** werden hingegen in Form betriebswirtschaftlicher Auswertungen (BWAs) von der Wirtschaftsprüfung und Steuerberatung als aussagekräftig genug eingeordnet, um existenzrelevante Entscheidungen wie z.B. Investitionen, Umstrukturierungen, Festanstellungen etc. zu begründen. In der Konsequenz muss ihnen ebenso die notwendige Robustheit gegenüber alltäglichen⁸¹⁷ Störeinflüssen zugesprochen werden, wie sie für eine sachgerechte Schätzung wichtig ist.
3. Auch für die sorgfältig ausgewählten und bei Bedarf korrigierten internen Ausgangsdaten „Monatskennzahlen“ kann nicht vollständig ausgeschlossen werden, dass *einzelne* unrepräsentative Ausreißerwerte enthalten sind. Damit die Berechnung der Schätzgröße dadurch nicht verzerrt wird, müssen sie (möglichst) sicher erkannt und aus dem Schätzungsrahmen⁸¹⁸ entfernt werden. **Rangbetrachtungen** sind hierfür deutlich besser als arithmetische Verfahren geeignet, denn bereits *ein* außergewöhnlich hoher Monats-

⁸¹⁶ Tipke/Kruse, AO/FGO, 170. Lieferung 5/2022, § 162 AO, Rz. 56. Vgl. a.: Nöcker (RiBFH), „Update Betriebsprüfung – Neues vom BFH und die Auswirkungen für die Praxis“, NWB 2016, 3157, 3165: „Jedenfalls der X. Senat nennt [die Richtsatzschätzung] nur noch als eine von mehreren zur Durchführung einer Plausibilitätsprüfung der Ergebnisse des Zeitreihenvergleichs.“

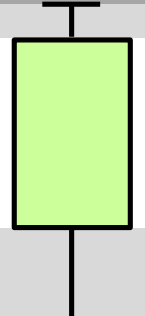
⁸¹⁷ Außergewöhnliche Störeinflüsse des Einzelfalls werden durch die systematisch-mehrperspektivische Arbeitsweise der SRP aufgedeckt (s. [Pkt. 4.2.2.](#)) und interaktiv auf Normalverhältnisse korrigiert (s. [Pkt. 4.2.9.](#)).

⁸¹⁸ Tipke/Kruse, AO/FGO, 170. Lieferung 5/2022, § 162 AO, Rz. 44: „Naturgemäß besteht bei der Schätzung nach Wahrscheinlichkeitsgrundsätzen eine Bandbreite möglicher Wertansätze (sog. Schätzungsrahmen). Soweit sich die Schätzung innerhalb des Rahmens bewegt, den der durch die FinBeh. unter Mitwirkung des Stpfl. festgestellte Sachverhalt vorgibt, ist sie nicht zu beanstanden.“

aufschlagsatz würde bspw. den arithmetischen Mittelwert rechnerisch stark beeinflussen – im Gegensatz zum mittleren Rang „Median“. Deshalb verwendet die Quantilsschätzung mit dem 80 %-Quantil einen Rang; relative⁸¹⁹ (Quantils-)Ränge sind dabei immer dann vorzuziehen, wenn eventuell unterschiedlich viele Daten wie bei abweichenden Prüfungszeiträumen bewertet oder verglichen werden sollen.

4. Eine übersichtlichere Darstellung von Rangverhältnissen liefert das statistische Instrument „**Box-Plot**“. Die nebenstehende Abb.

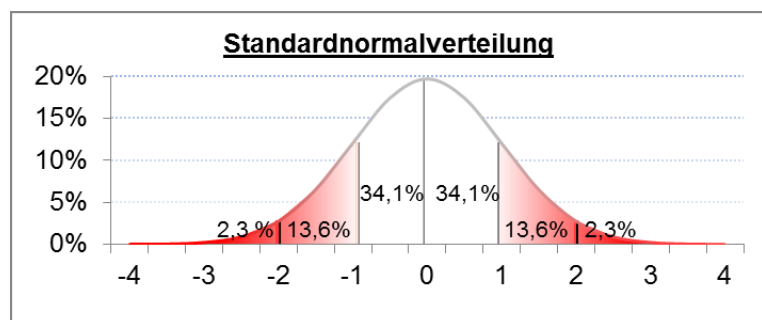
zeigt einen solchen zu zwölf zusammengehörigen Beispelsdatensätzen: Die Werte werden zunächst sortiert und mit den absoluten und relativen Rängen versehen. Nun werden die Bestandteile „Box“ mit den „normalen“ Daten und „Whisker“ mit den „schwachen Ausreißern“ bestimmt; außerhalb befinden sich – soweit vorhanden – „extreme Ausreißer“. Diese übliche Form der Ordnung von Datenmen-

Box-Plot			
Rang	Wert	Prozentrang (Quantil)	Box-Whisker-Plot
12	125	100%	
11	95	91%	
10	55	82%	
9	52	73%	
8	51	64%	
7	49	55%	
6	48	45%	
5	47	36%	
4	46	27%	
3	42	18%	
2	41	9%	
1	38	0%	

gen wird z.B. auch bei der „Interquartilmethode“ des § 1 Abs. 3a S. 3 ASTG⁸²⁰ zur Bandbreitenbetrachtung des Fremdvergleichsgrundsatzes verwendet.

5. Anstelle der mittleren 50 % beim Interquartilbereich wird die Unterscheidung von Normwerten und Ausreißern i.R. der Quantilsschätzung aus den **Normalverteilungen** (mit

ihrer glockenförmigen Verteilungscharakteristik) entnommen. Diese beschreiben das typische Streuverhalten natürlich entstandener Daten und werden wegen ihrer starken Verbrei-



tung auch dann angenommen, wenn keine anderen belastbaren Erkenntnisse vorliegen – wie bei *uneinheitlich manipulierten* (Bar-)Einnahmen. Jede Normalverteilung lässt sich aus den abgebildeten Rang-Wert-Verhältnissen der Standardnormalverteilung⁸²¹ ablei-

⁸¹⁹ Quantile geben die Ränge als Prozentposition in der Datenmenge wieder und werden mit ihrem Namen aus dem Lateinischen abgeleitet: z.B. Quartile (4 Stufen à 25 %), Quintile (5 Stufen à 20 %), Dezile (10 Stufen à 10 %), Perzentile (100 Stufen à 1 %). Fraktile sind Bruchteilsränge, sodass das ¼-Fraktile dem 25 %-Quantil entspricht. Tabellenkalkulationsprogramme unterstützen die einheitliche Ermittlung von Quantilen inklusive Zwischenrechnung (Interpolation) bei nicht ganzzahligen Zielrängen.

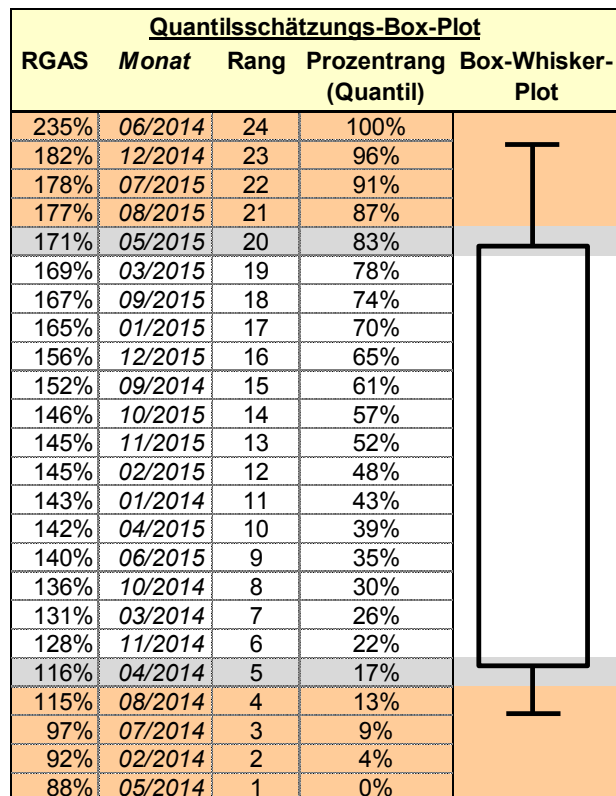
⁸²⁰ Vgl. a.: BMF v. 6.6.2023, „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093, Rz. 3.30.

⁸²¹ Die Standardnormalverteilung ist definiert mit den beiden Kenngrößen „Mittelwert = 0“ und „Standardabweichung (typische vorzeichenneutrale Abweichung jedes einzelnen Datensatzes vom Mittelwert) = 1“.

ten: Den Normalbereich der Daten „Mittelwert $\pm$ Standardabweichung“ bilden jeweils die ca. 34 % der Datensätze links und rechts vom Mittelwert, also insgesamt ca. 68 %. Etwa 27 % stammen als schwache Ausreißer aus den Bereichen „-1 bis -2“ sowie „1 bis 2“ und reichen bis zur doppelten Standardabweichung. Jenseits dieser verbleiben knapp 5 % starke Ausreißer.

6. Als **Übertragung der Normalverteilungsverhältnisse** auf einzelfallbezogene

Zahlen wie in nebenstehender Abb. wird der Normalbereich „16 %- bis 84⁸²² %-Quantil“ i.R. der Quantilsschätzung seit ihrer erstmaligen Beschreibung⁸²³ vorsichtshalber mit „20 %- bis 80 %-Quantil“ etwas enger gefasst, damit unrepräsentative Ausreißer (möglichst) sicher und vollständig ausgeschlossen werden. Dabei ergibt sich für die Beispelsdaten laut Box-Plot: Die niedrigsten und höchsten vier⁸²⁴ Aufschlagsätze werden eindeutig als Ausreißer gewertet. Die Ränge „5“ und „20“ liegen tlw. innerhalb und außerhalb des Intervalls „20 %- bis 80 %-Quantil“, weshalb zwischen den benachbarten



Rängen interpoliert wird; die Quantilsschätzung verwendet hierfür die Standardfunktion „QUANTIL.INKL“ von MS Excel^{®825}. Bei einigermaßen homogenen Daten wie im Beispiel liegen die Werte des Normalbereichs so eng beieinander, dass die Interpolation kaum eine Auswirkung auf das Ergebnis hat. Dem Verifikationsprinzip⁸²⁶ folgend wird „in dem

⁸²² Da der Schätzungsrahmen aus der Mitte der Ausgangsdaten gebildet werden soll, müssen bei der Festlegung des oberen Rands für den Normalbereich die unteren Ausreißer mitgezählt werden, also: 16 % untere Ausreißer + 68 % Normalwerte = 84 % obere Grenze des Normalbereichs.

⁸²³ Wähnert, „Die Quantilsschätzung – Schätzungsmöglichkeiten mit summarischen Methoden“, BBK 20/2008, 1083.

⁸²⁴ Bei 24 Monatswerten ergeben sich die beiden Grenzen des Normalbereichs unter der Rangzuordnung „1“ bis „24“ mit: 24 x 20 % bzw. 80 % = 4,8 resp. 19,2 (zu den Interpolationsregeln für die standardisierte Rangberechnung s. folgende Fn.).

⁸²⁵ Die beiden unterschiedlichen Interpolationsverfahren (MS Excel[®]: „QUANTIL.INKL“ und „QUANTIL.EXKL“) folgen den allgemeinen Berechnungsregeln (vgl. z.B.: <http://www.statistiker-blog.de/archives/excel-quantile-berechnen/6316.html>) und lassen sich somit z.B. auch in anderen Tabellenkalkulationsprogrammen nachvollziehen.

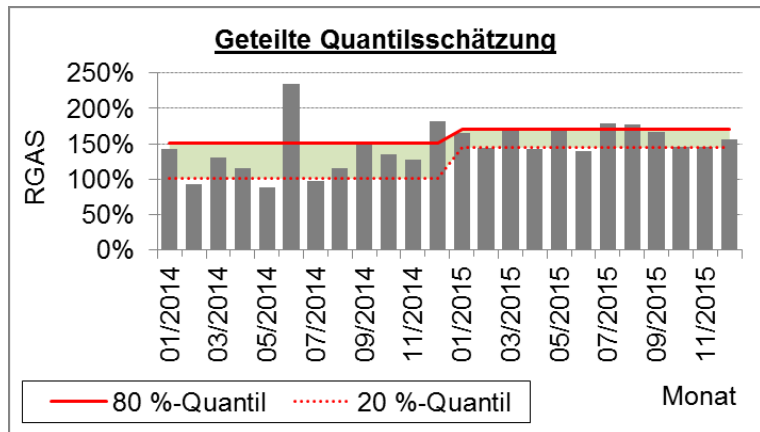
⁸²⁶ BFH v. 15.2.1989, BStBl. II 1989, 462: „der ‚Beweisverderber‘ oder ‚Beweisvereitler‘ aus seinem Verhalten keinen Vorteil ziehen darf (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO) [...] Zur Vermeidung eines sol-

gegebenen Schätzungsrahmen [...] die oberste Grenze⁸²⁷ für die Berechnung der Schätzungshöhe angesetzt.

7. Sofern Informationen vorliegen, wonach der Prüfungszeitraum aufgrund gravierender struktureller Veränderungen nicht homogen ist und deshalb nicht einheitlich geschätzt werden sollte, ist auch eine

geteilte Anwendung der Quantilsschätzung

möglich⁸²⁸. In der Beispielsgrafik zeigen die andersartigen Schwankungen der Aufschlagsätze beider Jahre Unterschiede im Geschäftsbetrieb an. Deshalb werden



für beide Prüfungsjahre getrennte Normalbereiche ermittelt; durch die Verwendung *relativer* Ränge ergibt sich auch für Teilzeiträume abweichender Länge ein vergleichbarer Schätzungsrahmen.

8. Abschließend ist das Schätzungsergebnis überschlägig daraufhin zu prüfen, ob es „insgesamt in sich schlüssig [...] wirtschaftlich vernünftig und möglich“⁸²⁹ ist: Eine **Gegenkontrolle** kann z.B. grob über das Preisgefüge erfolgen. Zudem bietet die SRP-Vorlage BWA dazu auf dem Tabellenblatt „80 %-Quantile“ zwei Verbildlichungen der betreffenden Monatsverhältnisse an, die durch chronologische und Häufigkeitsdarstellungen ein problemloses Erkennen von strukturellen Veränderungen gewährleisten und somit zur Absicherung eines Schätzungsergebnisses – i.S. einer „Sensitivitätsanalyse“⁸³⁰ – beitragen sollen (s. den Auflistungspkt. „Diagramme“ unter [Pkt. 4.8.2.](#)), sowie die plausi-

chen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

⁸²⁷ BFH v. 9.3.1967, BStBl. III 1967, 349; auch gemäß RiBFH Nöcker „zulässig“, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 2015, 3548, 3549.

⁸²⁸ Die Umsetzung der Quantilsschätzung für uneinheitliche Teilzeiträume lässt sich in der SRP-Vorlage BWA z.B. einfach dadurch erreichen, dass zwei (oder mehrere) Vorlagen mit dem kompletten Dateninhalt und jeweils eingeschränktem Prüfungszeitraum verwendet werden. Auch eine vorherige Ermittlung der 80 %-Quantile für Teilzeiträume und anschließende Eingabe im jährlichen Schätzungsfeld „frei geschätzter RGAS“ bzw. „frei geschätzter Einsatz“ ist denkbar (s. den Auflistungspkt. „freie Schätzung“ unter Pkt. 4.8.2.).

⁸²⁹ BFH v. 18.12.1984, BStBl. II 1986, 226.

⁸³⁰ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 69; vgl. a.: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 26: „Bei der zeitreihenbasierten Schätzung können betriebliche Besonderheiten berücksichtigt und verbleibende Unsicherheiten in einer Sensitivitätsanalyse, unter Umständen mit Sicherheitsabschlägen, berücksichtigt werden.“

bilisierende Fortschreibung der auf dem Tabellenblatt „Jahreszahlen“ eingetragenen (Halb-)Reingewinne nach Schätzung (s. [Pkt. 4.2.3.](#)).

4.8.2. Anwendungshinweise zur Vorlagen-Quantilsschätzung

Folgende Hinweise zur Durchführung einer (Quantils-)Schätzung – *nach eingetretener Amtspflicht zur Schätzung gem. § 162 AO* – inklusive Absicherung des Ergebnisses mit Hilfe der SRP-Vorlage BWA sind wichtig:

- ➔ Das Tabellenblatt „80 %-Quantile“ mit der **Ermittlung der Schätzgrößen für eine Quantilsschätzung** (s. den Aufzählungspkt. 6 „Übertragung der Normalverteilungsverhältnisse“ unter [Pkt. 4.8.1.](#)) **muss über das Anhakfeld neben der Überschrift bewusst aktiviert werden**, weil es nicht für Verprobungen⁸³¹ sondern *ausschließlich* für Schätzungen vorgesehen ist. Nach dem Anhaken werden oberhalb der Tabelle mit den Monatswerten die 80 %-Quantile zu den

80 %-Quantil der Monatswerte:	317%	7.874,10
--------------------------------------	-------------	-----------------

Aufschlagsätzen und Wareneinsatzsummen angezeigt. Für deren Ermittlung werden auf dem Tabellenblatt „Jahreszahlen“ eingetragene Entnahmen (s. [Pkt. 4.2.3.](#)) gleichmäßig auf die Monate umverteilt.⁸³² Sofern dort errechnete Warenbestandsveränderungen ebenfalls zu gleichen Anteilen auf die Monate umgelegt werden sollen, weil keine genaueren Informationen vorliegen, kann dies über das Ankreuzfeld neben der Überschrift „Einsatz“ des ersten Jahres bewirkt werden.⁸³³ Die interaktiven Berichtigungseingriffe in die untersuchten Daten „Einzelkorrektur von Datensätzen“ (s. [Pkt. 4.2.9.1.](#)), „Einschränkung des Betrachtungszeitraums“ (s. [Pkt. 4.2.9.2.](#)) und „Verschieben der ersten Datenreihe“ (s. [Pkt. 4.2.9.3.](#)) werden automatisch berücksichtigt (s.a. die Warnhinweise auf dem Tabellenblatt „Schätzung“ unter dem Auflistungspkt. „Warnfunktion“ weiter unten), die anderen unter [Pkt. 4.2.9.](#) genannten interaktiven Korrekturoptionen *nicht*.

⁸³¹ Insofern unzutreffend: Bleschick, „Der kalkulierte Beanstandungsanlass: Kein Nachweis von Mehrergebnissen durch die Summarische Risikoprüfung – Darstellung der Summarischen Risikoprüfung“, DStR 2017, 353: „Zunächst unternimmt er eine sog. betriebswirtschaftliche Auswertung: Hierzu ermittelt er im Wesentlichen mithilfe des Zeitreihenvergleichs einen Rohgewinnaufschlagsatz. Dieser wird auf den gesamten Wareneinkauf des Steuerpflichtigen angewandt. Unter Außerachtlassung von 20 % der höchsten und ggf. von 20 % der niedrigsten Rohgewinnaufschlagsätze ermittelt er den aus seiner Sicht möglichen Umsatz (sog. ‚Quantilsschätzung‘). Liegt der für den Außenprüfer mögliche Umsatz oberhalb des erklärten Umsatzes, sieht der Außenprüfer hierin aufgrund betriebswirtschaftlicher Auffälligkeiten einen Klärungsbedarf.“

⁸³² Eine differenziertere Einarbeitung von Entnahmen kann auf dem Tabellenblatt „Vergleichsdaten“ über taggenaue Korrekturen der untersuchten Daten (s. [Pkt. 4.2.9.1.](#)) erreicht werden.

⁸³³ Sollen Warenbestandsveränderungen punktuell eingearbeitet werden, können die automatische Verarbeitung der eingegebenen Warenbestände auf dem Tabellenblatt „Jahreszahlen“ per dortigem Anhakfeld verhindert (s. [Pkt. 4.2.3.](#)) und auf dem Tabellenblatt „Vergleichsdaten“ eine taggenaue Korrektur der untersuchten Daten (s. [Pkt. 4.2.9.1.](#)) vorgenommen werden.

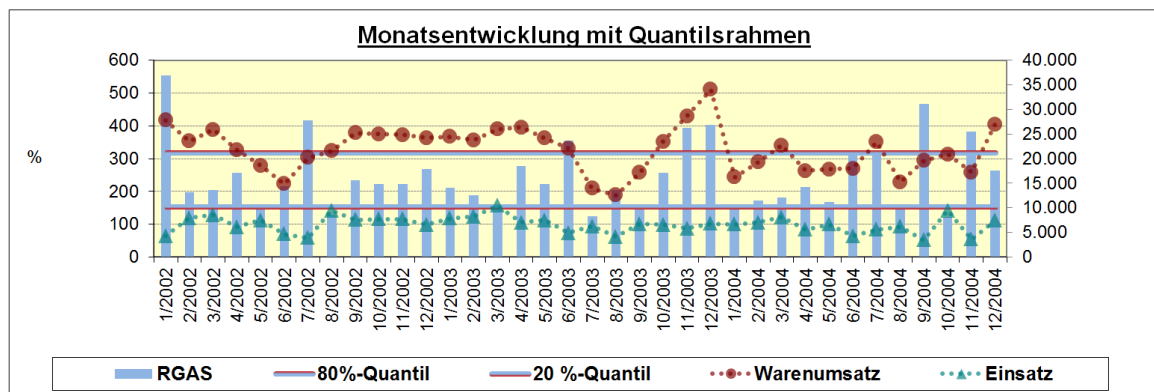
- ➔ Zusätzlich zur Anzeige der beiden 80 %-Quantile erscheint neben dem oberen Diagramm auf dem Tabellenblatt „80 %-Quantile“ der **Quantilsschätzungs-Box-Plot** mit dem **Normalbereich** und den **formatseitig hervorgehobenen Ausreißern** (s. die Aufzählungspkte. 4-6 unter [Pkt. 4.8.1.](#)); durch **Ankreuzen** kann zwischen den Monatsgrößen für die Aufschlagsätze und den Wareneinsatz gewechselt werden.
- Sollen bei der Ermittlung der 80 %-Quantile einzelne Monate ausgeschlossen werden, bspw. weil es sich um Vorbereitungszeiten von Saisongeschäften ohne relevanten operativen Betrieb handelt, lassen sich diese **nach Aktivierung einzeln löschen**; die

Ranking RGAS	
Rang	Wert
36	466 %
35	417 %
34	403 %
33	393 %
32	383 %
31	356 %
30	322 %
29	317 %
28	277 %
27	268 %
26	263 %
25	258 %

		2002					
Monat	Warenumsatz	Einsatz		Entnahmen		RGAS	
Jan.	27.803,15	10%	4.255,85	5%	205,00		553%
Feb.	23.521,00	9%	7.913,02	10%	205,00	x	197%
Mär.	25.866,62	9%	8.521,19	10%	205,00	x	204%
Apr.	21.804,03	8%	6.128,11	7%	205,00	x	256%
Mai	18.626,54	7%	7.489,32	9%	205,00	x	149%
Jun.	14.843,98	5%	4.688,12	6%	205,00	x	217%

grau-kursive Unterdrückung der betroffenen Werte weist auf deren Abwahl hin.

- ➔ Über die Angabe der 80 %-Quantile und die tabellarische Sortierung der Monatswerte hinaus wird der Schätzungsrahmen auf dem Tabellenblatt „80 %-Quantile“ in zwei **Diagrammen** ausgegeben, um die Beurteilung zu erleichtern, ob ein daraus resultierender Schätzungsansatz „insgesamt in sich schlüssig [...] wirtschaftlich vernünftig und möglich“⁸³⁴ erscheint (s. den Aufzählungspkt. 8 „Gegenkontrolle“ unter [Pkt. 4.8.1.](#)); die beiden Diagramme werden bei einem Wechsel zwischen „Aufschlagsätzen“ und „Wareneinsatz“ (s. die Anzeigeeoption „Ankreuzen“ unter dem Auflistungspkt. „Quantilsschätzungs-Box-Plot“ oberhalb) automatisch angepasst.

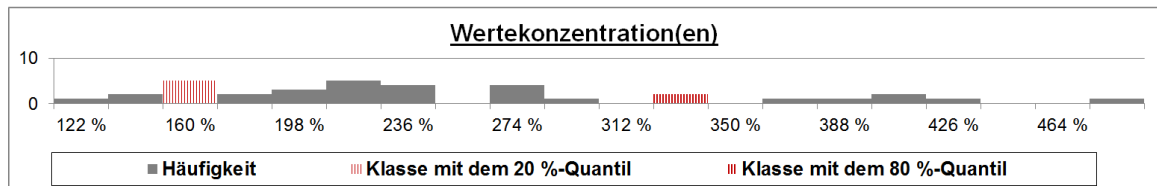


Die obere Grafik bildet die chronologische Entwicklung der Monatswerte ab, wobei die Datenreihen zu den Einzelgrößen „(Waren-)Umsatz“ und „(Waren-)Einsatz“ über ein An-

⁸³⁴ BFH v. 18.12.1984, BStBl. II 1986, 226.

kreuzfeld links neben dem Diagramm ein- und ausgeblendet werden können. Durch diese Verbildlichung lassen sich sowohl längere Teilzeiträume, die offenbar strukturell andersartige Bedingungen aufweisen (s. den Aufzählungspkt. 7 „geteilte Anwendung der Quantilsschätzung“ unter [Pkt. 4.8.1.](#)) als auch Einzelmonate erkennen, die als inoperative Phasen ausgeschlossen werden sollten (s. die Option „nach Aktivierung einzeln löschen“ unter dem Auflistungspkt. „Quantilsschätzungs-Box-Plot“ weiter oben).

Eine zweite Visualisierung der Verteilungsverhältnisse mit dem Ziel einer besseren Beurteilung, ob eine (Bandbreiten-)Schätzung mit einem 80 %-Quantil (für die Aufschlagsätze und/oder die Wareneinsätze) im Einzelfall sachgerecht ist, wird auf dem Tabellenblatt „80 %-Quantile“ unterhalb der ersten bereitgestellt; die Grenzen des Schätzungsrahmens „20 %-Quantil“ und „80 %-Quantil“ werden formatseitig hervorgehoben:



Mit Hilfe dieser Häufigkeitsdarstellung der Monatsgrößen „Aufschlagsätze“ bzw. „Wareneinsatz“ an einer linearen Klassifizierung (X-Achse) kann – besonders i.V.m. der chronologischen Abb. – gut nachvollzogen werden, ob sich die Werte als einheitliche Verteilung darstellen, oder ob offenbar unvergleichbare Teilzeiträume zu beachten sind. Im Beispielsfall offenbart sich eine inhomogene Streuung der monatlichen Aufschlagsätze mit einzelnen stark nach oben abweichenden Extremwerten, die sich auch durch die vorherige chronologische Verbildlichung nicht als saisonale oder strukturelle Veränderung erklärt. Deshalb sollte hier mit zumutbaren Hilfsmitteln vertieft werden, ob eine Schätzung innerhalb der Klasse „bis 274 %“ sachgerechter wäre.

- ➔ Als weitere Möglichkeit zur Absicherung des Ergebnisses aus einer Quantils- oder einer freien Hinzuschätzung des Aufschlagsatzes (s. den Aufzählungspkt. 8 „Gegenkontrolle“ unter [Pkt. 4.8.1.](#)) können das betreffende 80 %-Quantil bzw. der auf dem Tabellenblatt „Schätzung“ eingegebene Aufschlagsatz im oberen Diagramm der **Gleitschlittenzeitreihen** (s. [Pkt. 4.2.6.](#)) angezeigt werden; unabhängig von der Auswahl der Gleitschlittenperiode wird hierbei immer das 80 %-Quantil der *Monatsaufschlagsätze* angezeigt, weil diese nach logischen Erwägungen die Grundlage der Quantilsschätzung bilden (s. den Aufzählungspkt. 2 „Monatsgrößen“ unter [Pkt. 4.8.1.](#)).
- ➔ Die eigentliche Quantils- bzw. andersartige Hinzuschätzung findet auf dem **Tabellenblatt „Schätzung“** statt und wird über das Anhakfeld neben der Überschrift bewusst aktiviert (das Anhakfeld zur „Ermittlung der Schätzgrößen für eine Quantilsschätzung“

weiter oben ist mit diesem synchronisiert), damit dieser Bereich der SRP-Vorlage BWA ausschließlich für Schätzungen genutzt wird:

☒ **Interaktive Schätzungsvorlage**

Steuerpflichtige/r: **Restaurant** Steuernummer: **12 345 6789 0**

	2002	2003	2004
Warenumsatz:	273.498,86	276.713,18	234.729,54
Wareneinkauf korrigiert:	87.392,44	84.048,45	75.983,33
Bestandsveränderung:	4.904,47	-5.575,55	2,75
Entnahmen:	2.460,00	2.460,00	2.592,00
weitere Einsatzminderungen (ermäßigter USt-Satz): ☐			
weitere Einsatzminderungen (voller USt-Satz): ☐			
weitere Einsatzminderungen (ohne USt): ☐			
Einsatz:	80.027,97	87.164,00	73.388,58
RGAS:	242%	217%	220%
80 %-Quantil RGAS: ☐			
frei geschätzter RGAS: ☐			
80 %-Quantil Einsatz: ☐			
frei geschätzter Einsatz: ☐			
ermittelter Umsatz: ☐			
freie Umsatzhinzuschätzung: ☐			
Umsatzschätzung:			
Mehrerlös durch Schätzung:			
(RGAS nach Schätzung): ☐			
Mehrerlös insgesamt:			
Gewinn nach Schätzung:			
Gewinnsatz nach Schätzung:			
Anteil ermäßigter Steuersatz:			
Anteil Mehrerlös (ermäßigter USt-Satz): ☐			
freie Aufteilung Mehrerlös:			
Anteil Mehrerlös (ermäßigter USt-Satz): ☐			
Anteil Mehrerlös (voller USt-Satz): ☐			
Mehr-USt durch Schätzung:			
Mehr-USt insgesamt:			

Auswertung nach Reihenverschiebung um -10 Tag(e) (siehe Vergleichsdaten)
Auswertung nach Anpassung von Einzeldaten (siehe Vergleichsdaten)

Im oberen Bereich zwischen den durchgezogenen Linien werden die Jahressummen inklusive der in der SRP-Vorlage BWA vorgenommenen interaktiven Korrekturen nach [Pkt. 4.2.9.1.](#), [4.2.9.2.](#) und [4.2.9.3.](#) (s. den Auflistungspkt. „Ermittlung der Schätzgrößen für eine Quantilsschätzung“ weiter oben, die Hinweise unten in der Abb. sowie den folgenden Auflistungspkt. „Warnfunktion“) analog dem Tabellenblatt „Jahreszahlen“ übernommen. Alle ausfüllbaren weißen Felder sind ausschließlich für die (gemeinsame) Erarbeitung einer sachgerechten Schätzungshöhe vorgesehen. Sofern eine Einigung über die Hinzuschätzung erreicht werden kann, lässt sich das Vorlagenblatt über das [Ankreuzfeld oben links](#) in „Anlage zur Tatsächlichen Verständigung“ umbenennen; auf diese Weise kann der letzte gemeinsame Verhandlungsstand unproblematisch dokumentiert und der Verfahrensablauf nach einer Tatsächlichen Verständigung somit erleichtert werden.

- ➔ Wie in der Abb. zum Auflistungspkt. „Tabellenblatt ‚Schätzung‘“ oberhalb zu erkennen ist, erfüllt das zentrale Schätzungsfeld „80 %-Quantil RGAS:“ eine **Warnfunktion**:
- Durch eine **rote Umrandung des Felds** wird darauf hingewiesen⁸³⁵,

⁸³⁵ Der frühere Warnhinweis, dass auf dem Tabellenblatt „Jahreszahlen“ eingetragene Entnahmen nicht auf die Monate umverteilt worden sind und somit bei der Berechnung der 80 %-Quantile unberücksichtigt bleiben, ist ab der SRP-Vorlagenversion 2019 unnötig geworden, weil dies auf dem neuen Tabellenblatt „80 %-Quantile“ – unabhängig von der Gesamtbildprüfung (Verprobung) – automatisch geschieht.

1. dass sich Einzeldaten-Korrekturen auf dem Tabellenblatt „Vergleichsdaten“ in ihrer Summe nicht zu einer reinen Umverteilung ausgleichen (s. [Pkt. 4.2.9.1.](#)) oder
2. Datensätze außerhalb des auf dem Tabellenblatt „Vergleichsdaten“ eingegebenen Prüfungszeitraums liegen und somit nicht einbezogen werden können.

Die **Rotverfärbung der Schrift** „80 %-Quantil RGAS:“ erinnert daran⁸³⁶, falls die erste Datenreihe auf dem Tabellenblatt „Vergleichsdaten“ durch eine entsprechende Eingabe komplett vor- oder rückdatiert wurde (s. [Pkt. 4.2.9.3.](#)).

*Die Quantilsschätzung wird jedoch infolge dieser erkannten Eingriffe in die Betriebsdaten nicht unterbunden, weil gerade hierdurch „alle Umstände [...], die für die Schätzung von Bedeutung sind“⁸³⁷ im Einzelfall Berücksichtigung finden können. Vielmehr sollen die Warnhinweise nach einem umfangreichen Prüfungsprozess verhindern, dass derartige Datenveränderungen **ungewollt** die Schätzung beeinflussen.*

➔ *Im Gegensatz zu Kalkulationen bewirkt eine Erhöhung der Entnahmen bei der Schätzung mit einem bestimmten Quantil für den Periodenaufschlagsatz keine sichere Ergebnisminderung, weil die Verringerung des Wareneinsatzes zu durchgehend höheren Aufschlagsätzen führt und damit i.d.R. auch das betreffende Quantil vergrößert wird.*

2002			
Warenumsatz:			273.498,86
Wareneinkauf:			87.895,17
Bestandsveränderung:			4.904,47
Entnahmen:			2.460,00
weitere Einsatzminderungen (ermäßigter USt-Satz):	☒	7,0%	1.000,00
weitere Einsatzminderungen (voller USt-Satz):	☒	16,0%	1.500,00
weitere Einsatzminderungen (ohne USt):			
Einsatz:			78.030,70
RGAS:			251%
80 %-Quantil RGAS:	☒		317%

Um die Möglichkeit einvernehmlicher Schätzungen dadurch nicht zu verschlechtern, bietet die SRP-Vorlage BWA auf dem Tabellenblatt „Schätzung“ eine zielführende Funktion an, **zusätzliche Entnahmen im Schätzungsweg** zu ergänzen: Eingaben, welche im Bereich „weitere Einsatzminderungen (...)“ eingetragen werden, haben aus verhandlungstechnischen Gründen keine Auswirkungen auf die Ermittlung der **80 %-Quantile**. Für die Berechnung der jeweiligen Schätzungsauswirkungen (s. *den Auflistungspkt. „Berechnung des Schätzungsergebnisses“ unterhalb*) lassen sich die Werte durch **Aktivie-**

⁸³⁶ Der frühere Warnhinweis, dass die untersuchten Daten mittels Konjunkturanpassung auf dem Tabellenblatt „Monatszeitreihen“ (tlw.) von einem nominalen auf ein reales Preisniveau umgerechnet werden, ist ab der SRP-Vorlagenversion 2019 unnötig geworden, weil diese interaktive Korrektur i.R. einer Gesamtbildprüfung (Verprobung) nicht auf das neue Schätzungs-Tabellenblatt „80 %-Quantile“ durchwirkt.

⁸³⁷ § 162 Abs. 1 S. 2 AO.

ung in die Umsatzbesteuerung einbeziehen; die eingegebenen **Umsatzsteuersätze**⁸³⁸ werden automatisch in den Berechnungsbereich „Mehr-USt durch Schätzung“ übernommen.

- ➔ Soll (auch) der **Wareneinsatz mit dem 80 %-Quantil** geschätzt werden⁸³⁹, weil eindeutige Hinweise auf eine Doppelverkürzung vorliegen, kann dies ebenfalls per Anhaken auf dem Tabellenblatt „Schätzung“ der SRP-Vorlage BWA geschehen:

80 %-Quantil RGAS:	☐	
frei geschätzter RGAS:	☐	
80 %-Quantil Einsatz:	☒	12
		95.644,31

Es wird automatisch die **Anzahl an Monaten** mit einem Wareneinkauf größer Null gemäß Tabellenblatt „Monatszeitreihen“ vorgegeben; dieser Multiplikator für den 80 %-höchsten Monatswareneinsatz kann aber manuell – auch mit Nachkommastellen – abgeändert werden, wenn in einzelnen Monaten kein oder nur ein deutlich reduzierter Geschäftsbetrieb z.B. aufgrund von Urlaub stattgefunden hat.

- ➔ Die SRP-Vorlage BWA unterstützt mit ihrer transparenten Darstellung einer Schätzungsverhandlung ebenfalls **freie Schätzungen** des Aufschlagsatzes und/oder Wareneinsatzes. Dazu können die betreffenden Werte pro Jahr einzeln eingegeben oder per Anhakfeld vom ersten Jahr auf die folgenden übertragen werden. Sobald widersprüchliche Eingaben wie bspw. die aktivierte Quantilsschätzung trotz eines manuell eingetragenen Werts entstehen, unterbindet die SRP-Vorlage BWA den Auswertungsprozess mit einem entsprechenden Fehlerhinweis. Das abschließende Feld im Auswahlblock zur Erlösschätzung „**freie Umsatzhinzuschätzung**“ überschreibt in der Ergebnisberechnung (s. den Auflistungspkt. „Berechnung des Schätzungsergebnisses“ unterhalb) alle anderen Eingaben, weil dieser Schritt in Schätzungsverhandlungen üblicherweise den letzten (eilvernehmlichen) Stand darstellt, der zu dokumentieren ist.
- ➔ Sobald auswertungsfähige Daten in die SRP-Vorlage BWA hineinkopiert, ein (anteiliger) Prüfungszeitraum eingegeben, die Tabellenblätter zur Schätzung infolge festgestellter Schätzungsnotwendigkeit *bewusst* aktiviert und eine Schätzungsform ausgewählt wurden, erfolgt die **Berechnung des Schätzungsergebnisses** pro Jahr und insgesamt. Kommt es durch die Schätzungseingaben zu **Minderschätzungen**, wird wie in der folgenden Abb. formatseitig mit roter Schrift darauf hingewiesen; der Berechnungspro-

80 %-Quantil RGAS:
frei geschätzter RGAS:
80 %-Quantil Einsatz:
frei geschätzter Einsatz:
ermittelter Umsatz:
freie Umsatzhinzuschätzung:

⁸³⁸ Bei nicht ganzzahligen oder uneinheitlichen Umsatzsteuersätzen für einzelne Jahre z.B. infolge eines abweichenden Wirtschaftsjahres oder einer unterjährigen Änderung können Nachkommastellen verwendet werden, wovon die erste (gerundet) angezeigt wird.

⁸³⁹ Hierzu wird der 80 %-höchste Rang der monatlichen Wareneinsatzsummen ermittelt (s. den Auflistungspkt. „Ermittlung der Schätzgrößen für eine Quantilsschätzung“ oberhalb) und mit der Anzahl an Monaten multipliziert.

zess wird allerdings nicht abgebrochen, weil ein partielles Negativergebnis zielführender Bestandteil einer Gesamtschätzung sein kann.

80 %-Quantil RGAS:	☐				
frei geschätzter RGAS:	☐				
80 %-Quantil Einsatz:	☐				
frei geschätzter Einsatz:	☐				
ermittelter Umsatz:	☐				
freie Umsatzhinzuschätzung:	☐				
Umsatzschätzung:	☐				
Mehrerlös durch Schätzung:	☐				
(RGAS nach Schätzung):	☐				
Mehrerlös insgesamt:	☐				-7.747,55
Gewinn nach Schätzung:	☐				10.907,45
Gewinnsatz nach Schätzung:	☐				4%
festе Aufteilung Mehrerlös:	☐				
Anteil Mehrerlös (ermäßigter USt-Satz):	☐	7,0%			-2.324,26
freie Aufteilung Mehrerlös:	☐				
Anteil Mehrerlös (ermäßigter USt-Satz):	☐	7,0%			-5.423,28
Anteil Mehrerlös (voller USt-Satz):	☐	16,0%			-1.030,42
Mehr-USt durch Schätzung:	☐				
Mehr-USt insgesamt:	☐				-1.030,42

Im unteren Abschnitt des Tabellenblatts „Schätzung“ können die Auswirkungen der Hinzuschätzung auf die Umsatzsteuer berechnet werden (zur Eingabe der Umsatzsteuersätze s. den Auflistungspkt. „zusätzliche Entnahmen im Schätzungsweg“ oberhalb); dazu wird der Mehrerlös entweder einheitlich per Prozentsatz oder durch jahresweise Eingabe des Betrags zum ermäßigten Umsatzsteuersatz aufgeteilt. Wareneinsatzminderungen, die i.R. der Schätzung ergänzt wurden, lassen sich durch Anhaken berücksichtigen (s. den Auflistungspkt. „zusätzliche Entnahmen im Schätzungsweg“ oberhalb).

4.8.3. Rechtsprechung zur Quantilsschätzung

Als Folge der länderübergreifenden Verbreitung der (SRP und) Quantilsschätzung haben einige Finanzgerichte (mehrfach) zu der Methode Stellung genommen⁸⁴⁰; dabei wird sie allerdings nur z.T. ausführlich und korrekt beschrieben.⁸⁴¹ In den überwiegenden Fällen wurde die Quantilsschätzung von der **Finanzrechtsprechung** positiv gesehen und vor allem gegenüber „griffweisen“ Schätzungen als „deutlich differenziertere Schätzungsmethode“⁸⁴² anerkannt. Das FG Berlin-Brandenburg stellt folglich fest: „Ohne Erfolg macht die Antragstellerin hier ‚erhebliche Zweifel an der Schätzmethode‘ geltend. Sie verkennt, dass die Methode der Quantilsschätzung nicht grundsätzlich unzulässig ist. Sie wird vielmehr auch in der neueren finanzgerichtlichen Rechtsprechung als zulässig angesehen, wenn sie nicht dazu genutzt wird, um die Buchführung zu widerlegen, sondern um eine [...] Schätzung durchzuführen“⁸⁴³.

⁸⁴⁰ Dem Autor sind folgende Entscheidungen mit Ausführungen bekannt: FG Hamburg v. 4.9.2019 – 6 K 14/19, FG Berlin-Brandenburg v. 2.8.2018 – 10 V 10039/18, FG Hamburg v. 5.3.2018, 3 K 205/15 (EFG 2018, 1081), FG Hamburg v. 18.7.2017 – 6 V 119/17, FG Berlin-Brandenburg v. 30.6.2017 – 10 V 10219/16, FG Berlin-Brandenburg v. 9.1.2017 – 4 V 4255 u. 4265/15 (EFG 2017, 537), FG Hamburg v. 31.10.2016 – 2 V 202/16, FG Hamburg v. 26.8.2016 – 6 V 81/16, FG Berlin-Brandenburg v. 24.8.2016 – 5 V 5089/16 (EFG 2017, 12), FG Berlin-Brandenburg v. 25.5.2015 – 1 V 1039/15, FG Schleswig-Holstein v. 30.9.2014 – 1 V 83/14, FG Berlin-Brandenburg v. 31.1.2013 – 14 V 14316/11 u. FG Sachsen v. 10.1.2012 – 6 V 269/11.

⁸⁴¹ Vgl. z.B.: FG Hamburg v. 26.8.2016 – 6 V 81/16, Rz. 85 ff.

⁸⁴² FG Berlin-Brandenburg v. 30.6.2017 – 10 V 10219/16.

⁸⁴³ FG Berlin-Brandenburg v. 2.8.2018 – 10 V 10039/18.

Nach ergänzenden Anmerkungen zu einer aktuelleren BFH-Entscheidung betrachtet der RiBFH Nöcker die höchstrichterliche Meinungsbildung – trotz der kritischen Ausführungen aus 2017⁸⁴⁴ – als nicht abgeschlossen: „Insbesondere Fragestellungen rund um die Quantilsschätzung [...] harren insoweit einer weitergehenden Klärung durch den BFH“⁸⁴⁵; hilfreich dafür wäre allerdings, dass diese auch korrekt wahrgenommen und wiedergegeben wird.⁸⁴⁶

Sowohl der BFH⁸⁴⁷ als auch der 5. Senat des FG Berlin-Brandenburg⁸⁴⁸ haben ihre skeptische Haltung gegenüber der Quantilsschätzung ausschließlich auf **unzutreffende Kritikpunkte** gestützt, die deshalb hier einzeln kommentiert werden:

- ! „Die besonderen Risiken [einer Zeitreihenschätzung] ([...], erhebliche mathematische Hebelwirkungen, [...]) bestehen hier ebenso.“⁸⁴⁹
 - ➔ Die kritischen Ausführungen des BFH zu möglichen „Hebelwirkungen“ infolge von Zuordnungsfehlern in Zeitreihenanalysen gründen sich auf punktuelle Betrachtungen des **maximalen** 10-Wochen-Aufschlagsatzes.⁸⁵⁰ Derartige stöempfindliche Extremwerte werden i.R. der Quantilsschätzung sorgfältig ausgeschlossen (s. die *Aufzählungspkte. 4-6 „Box-Plot“, „Normalverteilungen“ und „Übertragung der Normalverteilungsverhältnisse“ unter Pkt. 4.8.1.*). In BWL, Wirtschaftsprüfung und Steuerberatung werden **normale**⁸⁵¹ Monatskennzahlen als so **robust** gegenüber Alltagseinflüssen angesehen, dass sie die Grundlage für existentielle Entscheidungen bilden (s. den *Aufzählungspkt. 2 „Monatsgrößen“ unter Pkt. 4.8.1.*). Außerdem werden eventuell verzerrende Sondersachverhalte vor einer Quantilsschätzung mit Hilfe der SRP-Prüfungssystematik gezielt gesucht und korrigiert (s. *Pkt. 4.2.2.*).
- ! „Die besonderen Risiken [einer Zeitreihenschätzung] ([...], Ausgabe großer und kaum vollständig überprüfbarer Datenmengen) bestehen hier ebenso.“⁸⁵²
 - ➔ Im Gegensatz zu den kritisierten „ca. 1 100 Zeilen zu je 10 Spalten, d.h. insgesamt ca. 11 000 Eintragungen“⁸⁵³ setzt die Quantilsschätzung i.d.R. auf 36 Monatswerten

⁸⁴⁴ Beschl. v. 12.7.2017 – X B 16/17.

⁸⁴⁵ Beschl. v. 8.8.2019 – X B 117/18, HFR 2020, 325, Anmerkung, Pkt. 1.

⁸⁴⁶ In wichtigen Merkmalen wie dem zitierten Beispiel unzutreffend: *Kulosa*, „Umgang mit Schätzungen in der Betriebsprüfung“, Teil 2, AgrB 1/2022, 13, 16: „Dabei werden monatliche Werte für den Rohgewinnaufschlag ermittelt; allerdings ohne Unterscheidung zwischen Wareneinkauf und Wareneinsatz (dies bewirkt eine erhebliche Ungenauigkeit).“

⁸⁴⁷ Beschl. v. 12.7.2017 – X B 16/17; da nach Auffassung des X. Senats gem. Rz. 86 im betreffenden Fall „bei summarischer Betrachtung keine Befugnis zur Vollschatzung abgeleitet werden [kann]“, wäre die Stellungnahme zur durchgeführten Quantilsschätzung als reine Schätzungsmethode nicht erforderlich gewesen.

⁸⁴⁸ FG Berlin-Brandenburg v. 24.8.2016 – 5 V 5089/16.

⁸⁴⁹ BFH v. 12.7.2017 – X B 16/17, Rz. 111.

⁸⁵⁰ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 44 ff.

⁸⁵¹ Bandbreiteneinengung mit Ausschluss von offensichtlichen unrepräsentativen Ausreißern.

⁸⁵² BFH v. 12.7.2017 – X B 16/17, Rz. 111.

⁸⁵³ BFH v. 25.3.2015, BStBl. II 2015, 743, Rz. 50.

auf. Diese Zahlen resultieren aus der monatlichen Aufsummierung der zur Prüfung überlassenen Daten inklusive möglicher interaktiver Korrekturen (s. [Pkt. 4.2.9.](#)), welche durch die SRP-Vorlagen nachvollziehbar dokumentiert werden und z.T. das Ergebnis aus der Mitwirkung Stpfl. sind. Zudem ist die verringerte Transparenz bei der Prüfung von digitalen Betriebsunterlagen eine Konsequenz daraus, dass „die Anzahl elektronisch überwachter Prozesse stetig [steigt]“⁸⁵⁴ und somit kaum der Betriebsprüfung anzulasten.

! „Das FA hat hierzu in der Beschwerdeerwiderung ausgeführt, mit der ‚Konjunkturanpassung‘ hätten die Ergebnisse der Jahre 2009 und 2010 an das Preisniveau des Jahres 2008 angepasst werden sollen, um die Vergleichbarkeit der Schätzung herzustellen. Indes soll eine Schätzung nicht vergleichbar, sondern möglichst realitätsnah sein.“⁸⁵⁵

➔ Die Bereinigung von Preisschwankungen durch Preisindizes ist das übliche Werkzeug der Volkswirtschaftslehre, um Zahlen über die Jahre miteinander vergleichen zu können. Auch i.R. der SRP dient dieses Mittel dazu, besonders einseitige Preiseffekte zu korrigieren (s. [Pkt. 4.2.5.1.](#) u. [4.2.9.4.](#)), damit die Ergebnisse von Zeitreihenanalysen hierdurch nicht verzerrt werden. Im betreffenden Fall hat der Prüfer die Konjunkturanpassung entsprechend tlw. für die Gesamtbildprüfung nach § 158 Abs. 2 Nr. 1 AO verwendet. Zur eigentlichen Quantilsschätzung der konjunkturbereinigten Jahre hat er sie laut Beschreibung dann wieder aufgelöst und somit die tatsächlichen Verhältnisse zu Grunde gelegt.⁸⁵⁶

! „Dieser Einwand verkennt, dass die Methode der Quantilsschätzung gerade auf der Heranziehung eines der höchsten RAS beruht. [...] Tatsächlich rekurriert die Quantilsschätzung nicht etwa auf den ‚normalen Geschäftsverlauf‘, sondern stützt sich für die vorgenommene Vollschatzung auf einen Wert, der in 80 % der Zeitabschnitte gerade nicht erreicht wird.“⁸⁵⁷

➔ Mittlere Werte in einem Schätzungsrahmen würden nicht zu einem wahrscheinlicheren Schätzungsergebnis führen⁸⁵⁸, denn diese sollen nach ihrer statistischen Aufgabe die als „unzutreffend“ erkannten Besteuerungsgrundlagen möglichst gut repräsentieren; vielmehr würde auf diese Weise dem Schätzungsziel entgegengewirkt,

⁸⁵⁴ Graf, „Anforderungen an die Vorlage von Einzelaufzeichnungen – Ende der ‚Schonzeit‘ für Daten aus Vor-Systemen?“, BBK 18/2013, 875, 877.

⁸⁵⁵ BFH v. 12.7.2017 – X B 16/17, Rz. 128.

⁸⁵⁶ Vgl. BFH v. 12.7.2017 – X B 16/17, Rz. 10 ff.

⁸⁵⁷ BFH v. 12.7.2017 – X B 16/17, Rz. 135 u. 144.

⁸⁵⁸ Durch einen mittleren Ansatz im Schätzungsrahmen würde das Ergebnis den erklärten Besteuerungsgrundlagen nahekommen und sie somit nachträglich anerkennen, obwohl sie nach der Gesamtbildprüfung gem. § 158 Abs. 2 Nr. 1 AO zu verwerfen sind.

dass es „keinen Anreiz [gibt], sich weiterhin schätzen zu lassen“⁸⁵⁹. Deshalb wählt die Quantilsschätzung innerhalb des sorgfältig festgestellten Schätzungsrahmens *vorsichtig* die „oberste Grenze“⁸⁶⁰ (s. die Aufzählungspkt. 6 „Übertragung der Normalverteilungsverhältnisse“ unter [Pkt. 4.8.1.](#)), weil gemäß Verifikationsprinzip (s. [Pkt. 4.6.](#)) „der ‚Beweisverderber‘ oder ‚Beweisvereiteler‘ aus seinem Verhalten keinen Vorteil ziehen darf (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO)“⁸⁶¹.

! „In mathematischer Hinsicht setzt die Anwendung [...] zuvörderst voraus, dass die RAS überhaupt der Normalverteilung folgen und die erhobene Grundgesamtheit (hier: 36 Einzelgrößen) groß genug ist.“⁸⁶²

➔ Die Ableitung eines sachgerechten Schätzungsergebnisses aus betriebseigenen (Zeitreihen-)Größen setzt vor allem den (möglichst sicheren) Ausschluss von eventuell enthaltenen unrepräsentativen Ausreißern voraus. Zu dieser Bandbreiteineengung nutzt die Quantilsschätzung das bekannteste Streuverhalten der Normalverteilungen (s. die Aufzählungspkt. 5 „Normalverteilungen“ unter [Pkt. 4.8.1.](#)) als *verallgemeinerbaren wissenschaftsbezogenen* Ansatz – vergleichbar der „Interquartilsmethode“ i.R. der Fremdvergleichsprüfung von Verrechnungspreisen⁸⁶³; insofern ist es hier genauso unerheblich, ob Aufschlagsätze – besonders nach unterschiedlichen Hinterziehungsweisen – typischerweise normalverteilt sind, wie es für die Verrechnungspreisprüfung unwichtig ist, ob die externen Vergleichspreise einer bestimmten Verteilung folgen. Auch die Anzahl an vorliegenden Monatswerten ist unbeachtlich, weil im Quantilsschätzungs-Box-Plot mit Quantilen gearbeitet wird (s. die Aufzählungspkt. 6 „Übertragung der Normalverteilungsverhältnisse“ unter [Pkt. 4.8.1.](#)), denn das Ziel von relativen Prozenträngen ist gerade die Übertragbarkeit auf unterschiedlich viele und auch wenige Daten.

! „Das FG Hamburg (Beschluss vom 26. August 2016 6 V 81/16) hatte in einem AdV-Beschluss einen Sachverhalt zu beurteilen, in dem Schwarzeinkäufe konkret nachgewiesen waren, die Gewinnermittlung sich also schon ohne den Zeitreihenvergleich als materiell unrichtig erwiesen hatte. Es hat hier die Quantilsschätzung dem Grunde nach nicht beanstandet, der Höhe nach aber erhebliche Korrekturen vorgenommen.“⁸⁶⁴

⁸⁵⁹ FG Berlin-Brandenburg v. 30.6.2017 – 10 V 10219/16.

⁸⁶⁰ BFH v. 9.3.1967, BStBl. III 1967, 349; auch gemäß RiBFH Nöcker „zulässig“, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 2015, 3548, 3549.

⁸⁶¹ BFH v. 15.2.1989, BStBl. II 1989, 462.

⁸⁶² BFH v. 12.7.2017 – X B 16/17, Rz. 139.

⁸⁶³ § 1 Abs. 3a S. 3 AStG; vgl. a.: BMF v. 6.6.2023, „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093, Rz. 3.30.

⁸⁶⁴ BFH v. 12.7.2017 – X B 16/17, Rz. 115.

- ➔ Tatsächlich wurden die Hinzuschätzungen durch die Quantilsschätzung mit einem Aufschlagsatz von 199 % im betreffenden Beschluss *vollständig* anerkannt.⁸⁶⁵ Die herausgestellten „erheblichen Korrekturen“ beziehen sich auf andere Sachverhalte.
- ! „Insoweit ist *Kulosa*⁸⁶⁶ zuzustimmen, wenn dieser ausführt, dass ‚rechnerische oder grafische Ergebnisse selbst noch keine zwingenden Besteuerungsgrundlagen darstellen, sondern zunächst in den tatsächlichen Lebenssachverhalt eingeordnet, auf Plausibilität verprobt und rechtlich gewürdigt werden müssen‘.“⁸⁶⁷
- ➔ Diese Generalkritik richtet sich gleichermaßen gegen alle indirekten Prüfungsverfahren, also auch gegen die höchstrichterlich anerkannten Vermögenszuwachs- und Geldverkehrsrechnungen sowie Kalkulationen, weshalb es fragwürdig ist, sie als Ablehnungsgrund für ein *spezielles* Schätzungsverfahren einzusetzen. Das Verifikationsprinzip⁸⁶⁸ sieht gerade die Anwendung von verprobenden⁸⁶⁹ Gesamtbildmethoden⁸⁷⁰ vor, um nicht nur im Einzelfall sondern auch in der Gesamtheit der Stpfl. die Gesetzmäßigkeit der Besteuerung nach § 85 AO umzusetzen.
- ! „Die Quantilsermittlung stellt nach Auffassung des Senats lediglich eine Verprobungs-, aber keine zulässige Schätzungsmethode dar. Denn ebenso wie andere Verprobungsmethoden [...] zeigt die Quantilsermittlung lediglich sog. Ausreißerwerte an, die sodann zu weiteren Ermittlungen Anlass geben, [...]“⁸⁷¹.
- ➔ Von Beginn an⁸⁷² ist die Quantils*schätzung* mit dem „objektivierten Leistungsfähigkeitsmaß“⁸⁷³ als reine Schätzungsmethode vorgestellt worden. Dabei hat das (möglichst sichere und vollständige) Erkennen und Ausschließen von unrepräsentativen Ausreißerwerten im Umkehrschluss die Bereinigung der Betriebsdaten auf den „Normalbereich“ als sachgerechten Schätzungsrahmen zur Folge. Dieser Schritt hat

⁸⁶⁵ FG Hamburg v. 26.8.2016 – 6 V 81/16, Rz. 19 u. 97.

⁸⁶⁶ *Kulosa (RiBFH)*, „Mathematisch-statistische Schätzungsmethoden in der Betriebsprüfung – Erste BFH-Rechtsprechung zum Zeitreihenvergleich“, DB 2015, 1797, 1799.

⁸⁶⁷ FG Berlin-Brandenburg v. 24.8.2016 – 5 V 5089/16.

⁸⁶⁸ BFH v. 16.5.2013, BStBl. II 2014, 225: „Der Gesetzgeber ist demgemäß von Verfassungs wegen verpflichtet, zur Vermeidung der Verfassungswidrigkeit des materiellen Steuergesetzes dieses in ein normatives Umfeld einzubetten, das die tatsächliche Lastengleichheit der Steuerpflichtigen gewährleistet, insbesondere auch durch die Ergänzung des Deklarationsprinzips durch das Verifikationsprinzip.“

⁸⁶⁹ Anwendungserlass zu § 158 AO: „Die Vorschrift enthält eine gesetzliche Vermutung. Sie verliert ihre Wirksamkeit mit der Folge der Schätzungsnotwendigkeit nach § 162 AO, wenn es nach [einer] *Verprobung* [...] unwahrscheinlich ist, dass das ausgewiesene Ergebnis mit den tatsächlichen Verhältnissen übereinstimmt.“

⁸⁷⁰ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5: „Die gesetzliche Regelung des § 158 gründet sich insofern auf die Einsicht, dass es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild einer Buchführung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann.“

⁸⁷¹ FG Berlin-Brandenburg v. 24.8.2016 – 5 V 5089/16.

⁸⁷² *Wähnert*, „Die Quantilsschätzung – Schätzungsmöglichkeiten mit summarischen Methoden“, BBK 20/2008, 1083.

⁸⁷³ *Schumann/Wähnert*, „Die Quantilsschätzung – Schätzung anhand eines objektivierten Leistungsfähigkeitsmaßes“, Stbg 12/2012, 535.

eine überragende Bedeutung dafür, dass ein berechnetes Schätzungsergebnis nicht verzerrt wird (s. die Aufzählungspkte. 4-6 „Box-Plot“, „Normalverteilungen“ und „Übertragung der Normalverteilungsverhältnisse“ unter [Pkt. 4.8.1.](#)).

! „Vielmehr stellt es nach Auffassung des Senats einen Fehlschluss dar, dass zwischen der Eliminierung von Ausreißerwerten sowie der sich daraus ergebenden Ermittlung eines Normalmittels einerseits und den tatsächlichen betrieblichen Gegebenheiten andererseits ein sachlicher Zusammenhang besteht, [...]“⁸⁷⁴.

➔ Der „[Zusammenhang] zwischen der Eliminierung von [unrepräsentativen] Ausreißerwerten [...] und den tatsächlichen betrieblichen Gegebenheiten“ besteht darin, dass jeder vernünftige aus Betriebsdaten abgeleitete Schätzungsrahmen (s. den Abschnitt „Schätzungsrahmen“ unter [Pkt. 4.8.](#)) als Ausgangspunkt für eine Schätzung⁸⁷⁵ die Einzelfallverhältnisse so realistisch wie möglich und damit *ohne verzerrende Extremwerte* abbilden sollte.

! „Zudem ist die Anwendung der Quantilsermittlung je nach Länge des Zeitraums [...] als willkürlich anzusehen. Erfasst die Quantilsermittlung nur einen kurzen Zeitraum und liegen in diesem Zeitraum zufällig besonders hohe oder besonders niedrige Werte, verzerrt dieser Umstand in erheblicher Weise die Schätzung.“⁸⁷⁶

➔ Der Quantilsschätzungs-Box-Plot bewertet die Monatsgrößen des gesamten Prüfungszeitraums (s. den Aufzählungspkte. 6 „Übertragung der Normalverteilungsverhältnisse“ unter [Pkt. 4.8.1.](#)) – es ist also nicht vorgesehen, dass ein kurzer Teilzeitraum zur Beurteilung eines viel größeren Prüfungszeitraums übertragen wird. Handelt es sich um „einen kurzen Zeitraum [...] mit] zufällig besonders hohe[n] oder besonders niedrige[n] Werte[n]“, ist eine entsprechende Festlegung der Schätzungsgrößen für diesen Zeitraum sachgerecht. Sollten die „besonders hohe[n] oder besonders niedrige[n] Werte“ nur Ausreißer darstellen, werden sie über den Box-Plot aus dem Schätzungsrahmen ausgeschlossen. Gerade durch die Einordnung der (betriebseigenen) Monatswerte mit *relativen* Prozenträngen (Quantilen) in Normalwerte und Ausreißer wird erreicht, dass das Ergebnis auch für unterschiedlich lange Prüfungszeiträume vergleichbar ist (s. den Aufzählungspkte. 3 „Rangbetrachtungen“ unter [Pkt. 4.8.1.](#)).

⁸⁷⁴ FG Berlin-Brandenburg v. 24.8.2016 – 5 V 5089/16.

⁸⁷⁵ Tipke/Kruse, AO/FGO, [170. Lieferung 5/2022](#), § 162 AO, Rz. 44: „Naturgemäß besteht bei der Schätzung nach Wahrscheinlichkeitsgrundsätzen eine Bandbreite möglicher Wertansätze (sog. Schätzungsrahmen). Soweit sich die Schätzung innerhalb des Rahmens bewegt, den der durch die FinBeh. unter Mitwirkung des Stpfl. festgestellte Sachverhalt vorgibt, ist sie nicht zu beanstanden (BFH BStBl. 93, 259; 01, 381, 382; BFH/NV 99, 741 f.; 06, 1801, 1803; FG BW EFG 97, 45 f.; FG Hess. EFG 01, 798 f. m. Anm. Fumi). Der Schätzungsrahmen ist umso größer, je ungesicherter das Tatsachenmaterial ist, auf dem die Schätzung basiert.“

⁸⁷⁶ FG Berlin-Brandenburg v. 24.8.2016 – 5 V 5089/16.

Die Quantilsschätzung kann **genauso wenig wie jede andere Schätzungsmethode** die wahren Besteuerungsgrundlagen wiederherstellen. Sie ist aber ein **verallgemeinerbares wissenschaftsbasiertes** Verfahren, um aus den Betriebsdaten einen Schätzungsrahmen abzuleiten, in welchem sehr wahrscheinlich die richtigen Bemessungsgrundlagen für die Steuern liegen. Dem stimmt die wissenschaftliche Autorin *Schütt* zu: „Bei den herkömmlichen Methoden besteht häufig das Problem der Anwendbarkeit aufgrund fehlender Aufzeichnungen. Die Quantilsschätzung hat den Vorteil, dass diese bei lückenhaften oder manipulierten Unterlagen angewandt werden kann und die Aufzeichnungen des Stpfl. verwendet. Dadurch werden betriebspezifische Besonderheiten des Unternehmens berücksichtigt.“⁸⁷⁷ „[F]ür den Zeitreihenvergleich und die Quantilsschätzung [ist] zu berücksichtigen, dass sie – im Gegensatz zum äußeren Betriebsvergleich – bereits auf betriebsinterne Daten abstellen. Wird überdies durch ein ‚Gleitschlitten‘-Verfahren auch den betrieblichen Ursachen für Rohgewinnaufschlagsatz-Schwankungen nachgegangen, so sind die Ergebnisse eines Zeitreihenvergleichs i.S. einer Quantilsschätzung als Ausgangspunkt einer Schätzung des Ergebnisses einer formell ordnungswidrigen Buchführung ausreichend.“⁸⁷⁸

Die **Erfahrungen mit dem 80 %-Quantil** aus mehreren tausend Schätzungsfällen seit 2008 haben ergeben: Durch sorgfältige Quantilsschätzungen resultieren vernünftige, eher vorsichtige Schätzungsergebnisse. Entsprechend haben auch etliche Überprüfungen durch Fahndungsmaßnahmen gezeigt, dass die tatsächlichen Steuerhinterziehungen in (nahezu) allen Einzelfällen vergleichbar oder häufig höher waren. **Folgelogisch wird die Quantilsschätzung auch im BMF-Schreiben zu „[a]utomationsgestützte[n] quantitative[n] Prüfungsmethoden in der steuerlichen Außenprüfung“ als eine mögliche quantitative Schätzungsmethode aufgeführt.**⁸⁷⁹

4.9. Strafrechtliche Schätzung

Das Verifikationsprinzip des § 158 Abs. 2 Nr. 1 AO (s. [Pkt. 4.6.](#)) mit seinem abgeschwächten Beweismaß „Beanstandungsanlass“ gilt **nicht** für das (Steuer-)Strafverfahren.⁸⁸⁰ Vielmehr muss die Finanzbehörde (Steuer-)Delikte für eine strafrechtliche Verurteilung beweisen. Dem Zweifelsgrundsatz „in dubio pro reo“ folgend setzt dies die volle richterliche Gewissheit vor-

⁸⁷⁷ *Schütt*, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 326.

⁸⁷⁸ Zugmaier/Nöcker/Webel, Abgabenordnung, Februar 2022, § 162 AO, Rn. 154.

⁸⁷⁹ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 27 ff.

⁸⁸⁰ Unzutreffend wird in der Literatur behauptet, dass i.R. der SRP die Unterschiede zwischen Besteuerungs- und Strafverfahren nicht beachtet würden: *Beyer*, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, 2939: „Das Handbuch zur SRP geht unkritisch von einer Anwendbarkeit der SRP im Steuerstrafverfahren aus [...] Auch geht das Handbuch nicht auf den Widerspruch zur strafgerichtlichen Rechtsprechung ein, nach der ein Beschuldigter schweigen darf und im Strafverfahren keine Minderung des Beweismaßes zulässig ist.“

raus, weshalb „in solchen Fällen von der für den Angeklagten günstigsten Möglichkeit auszugehen [ist], die nach den gesamten Umständen in Betracht kommt“⁸⁸¹.

Die Ausführungen des BGH zur ausreichenden Aussagegewissheit i.R. einer strafrechtlichen Bestimmung des Schuldumfangs „die ermittelte Vertrauensuntergrenze [...] mit einer Wahrscheinlichkeit von 99,5 % zutreffend festgestellt“⁸⁸² können als Anhaltspunkt für eine verallgemeinernde⁸⁸³ Einstufung des **Beweises** i.S. der vollen richterlichen Überzeugung herangezogen werden (s.a. [Pkt. 4.6.](#)). (Meta-)Studien zu Folge können auch moderne Beweismittel wie DNA- und Fingerabdruckanalysen keine höhere Gewissheit als 99-99,5 % gewährleisten.⁸⁸⁴ In Anbetracht der Tatsache, dass sich insbesondere komplexe(re) Vorgänge nicht vollständig und objektiv aus der Vergangenheit rekonstruieren lassen⁸⁸⁵, weil Dokumentationen, Aussagen etc. nur in sehr seltenen Fällen vollständig und objektiv sind, erscheint diese Wahrscheinlichkeitsstufe zur Erlangung der vollen richterlichen Überzeugung ausreichend. Folgende Hinweise des BGH konkretisieren zudem die realistische Einordnung des strafrechtlich erforderlichen Beweismaßes: „Liegen mehrere Beweisanzeichen vor, so genügt es nicht, sie jeweils einzeln abzuhandeln; erforderlich ist vielmehr eine Gesamtwürdigung. Auch wenn keine der jeweiligen Indiztatsachen für sich allein zum Nachweis der Täterschaft des Angekl. ausreicht, besteht die Möglichkeit, daß sie in ihrer Gesamtheit dem Gericht die entsprechende Überzeugung vermitteln können [...] Außerdem ist der Tatrichter nicht verpflichtet, entlastende Angaben des Angekl., für deren Richtigkeit oder Unrichtigkeit es keine Beweise gibt, ohne weiteres als unwiderlegt hinzunehmen. Vielmehr muß er sich

⁸⁸¹ BGH v. 5.2.2003 – 2 StR 321/02.

⁸⁸² UrT. v. 14.12.1989 – 4 StR 419/89, BGHSt 36, 320, Rz. 26.

⁸⁸³ Köller/Nissen/Rieß/Sadorf, „Probabilistische Schlussfolgerungen in Schriftgutachten – Zur Begründung und Vereinheitlichung von Wahrscheinlichkeitsaussagen in Sachverständigengutachten“, Reihe „Polizei + Forschung“, Bd. 26, herausgegeben vom Bundeskriminalamt (BKA), Kriminalistisches Institut, Luchterhand Verlag, 2004, 13: „Die numerische Schätzung bewirkt insbesondere eine Vermeidung von Missverständnissen oder möglichen Unterschieden in den scheinbar gleichen Größenordnungen verschiedener Individuen bei identischen verbalen Ausdrücken. [...] Sie erfordert vom Sachverständigen schlicht eine präzisere Einschätzung. Sie erhöht die intersubjektive Beurteilbarkeit dessen, was gemeint ist.“

⁸⁸⁴ Ausführlich hierzu: Giezek/Rupprecht/Wähnert, „Wie sicher muss ein Verprobungsergebnis sein? – Anleihen aus der Kriminalistik zur sinnvollen Einordnung des Beanstandungsanlasses“, BBK 2017, 236.

⁸⁸⁵ Vester, „Die Kunst vernetzt zu denken“, dtv Verlag, 2012, 15: „Unser Dilemma im Umgang mit der Komplexität unserer Welt lässt sich darauf zurückführen, dass wir wohl darin ausgebildet wurden einfache logische Schlüsse zu ziehen und nahe liegende Ursache-Wirkungs-Beziehungen zu definieren. [...] Simple Ursache-Wirkungs-Beziehungen gibt es nur in der Theorie, nicht in der Wirklichkeit.“ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 323-324: „Prüfungen vergangener Aufzeichnungen können, nach dem logischen Menschenverstand, nur zu Wahrscheinlichkeitsaussagen führen. Der Grund dafür ist, dass es sich bei Aufzeichnungen um ein reduziertes Abbild komplexer Daten und Zusammenhänge handelt. [...] Schlussfolgernd kann bei der Prüfung von (steuerlichen) Daten keine (100 %) Sicherheit und somit kein Beweis vorliegen bzw. erfolgen.“

auf der Grundlage des gesamten Beweisergebnisses entscheiden, ob diese Angaben geeignet sind, seine Überzeugungsbildung zu beeinflussen“⁸⁸⁶.

Aus den Ergebnissen von Wahrscheinlichkeitstests i.R. von entsprechend abgesicherten Musteranalysen geht hervor, dass Aussagegewissheiten von „≥ 99 bzw. 99,5 %“ zur (Nicht-) Übereinstimmung von untersuchten Besteuerungsdaten mit ihren erwarteten Punktidealen erreichbar sind (s. [Pkt. 4.3.1.3.](#)) und anhand weiterer (geschätzter) Informationen zur indizialen Beweiskraft für einen Manipulationsverdacht umgerechnet werden können⁸⁸⁷; u.a. durch zielgerichtete Funktionalitäten der SRP-Vorlage Stochastik können dabei andere Ursachen (weitest gehend) ausgeschlossen werden (s. [Pkt. 4.3.2.2.3.](#), [4.3.3.2.3.](#) u. [4.3.4.5.](#)). Diese Klarheit passt zur Beurteilung der Belastbarkeit von Musteranalysen durch die Steuerstrafrechtskommentierung.⁸⁸⁸ Zusätzlich sind bei Ergebnissen aus Prüfungsnetzen wie der SRP die **allgemeinen Beweisprinzipien**⁸⁸⁹ zu beachten:

- ➔ Durch systematisch-mehrperspektivisches Nachverfolgen von Auffälligkeiten in Besteuerungsdaten (s. [Pkt. 4.2.2.](#)) werden mögliche nicht-manipulative Faktoren grundsätzlich (weitest gehend) abgeklärt und gegebenenfalls interaktiv berücksichtigt (s. [Pkt. 4.2.9.](#)); dazu können alle konkreten Erkenntnisse aus der Mitwirkung⁸⁹⁰ und aus strafprozessualen Maßnahmen herangezogen werden. Verbleibende Unschlüssigkeiten im Gesamtbild erlangen somit eine höhere Aussagegewissheit (s. *den Abschnitt „einzelne verbleibende Beweisanzeichen“ unter [Pkt. 4.4.](#)*).

⁸⁸⁶ BGH v. 25.11.1982 – 4 StR 564/82, NSTZ 1983, 133.

⁸⁸⁷ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 143 ff.: Zur genaueren Bewertung eines Indizes müssen neben der fehlerhaften Beanstandung (bei 99 % Testgenauigkeit also 1 %) auch die Tatverbreitung und Aufdeckungswahrscheinlichkeit zumindest schätzungsweise bekannt sein.

⁸⁸⁸ Kohlmann, Steuerstrafrecht, 72. Lieferung 9.2021, § 397 AO, Rz 12.3: „Die praktische Verwertbarkeit der statistischen Verfahren ist durchaus umstritten, und zwar bereits als Grundlage für steuerliche Schätzungen, aber erst recht für strafrechtliche Verurteilungen. Die Finanzgerichte lehnen bislang den Chi-Quadrat-Test als eine sichere Methode, die Ordnungsgemäßheit der Buchführung zu verwerfen, ab. Allerdings wird weniger die grundsätzliche Eignung der Analysen angezweifelt als vielmehr hinterfragt, unter welchen Voraussetzungen und mit welcher Anzahl von Ausgangswerten die Methode aussagekräftige Ergebnisse liefert.“

⁸⁸⁹ BGH v. 24.3.2016 – 2 StR 112/14, NSTZ 2016, 490, Leitsatz Nr. 5: „Die Anforderungen, die gemäß § 261 StPO an die Überzeugungsbildung des Tatrichters zu stellen sind, werden nicht von – tatsächlich oder vermeintlich – unterschiedlichen wissenschaftlichen Konzepten bestimmt; maßgeblich ist allein, dass die Beweiswürdigung auf einer rationalen, verstandesmäßig einsichtigen und intersubjektiv diskutierbaren Grundlage beruht. Kann eine Feststellung nur mit Hilfe naturwissenschaftlicher Mittel getroffen werden, ist der Tatrichter zwar nicht gezwungen, sich insoweit nur auf allgemein anerkannte Methoden zu stützen. Die tatrichterliche Würdigung darf allerdings den Gesetzen der Logik und dem gesicherten wissenschaftlichen Erfahrungswissen nicht widersprechen.“

⁸⁹⁰ Hierzu bemängelt Beyer („Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931, 2939), „das [SRP]-Handbuch [gehe] nicht auf den Widerspruch zur strafgerichtlichen Rechtsprechung ein, nach der ein Beschuldigter schweigen darf“, was hingegen im Abschnitt „Legalitätsprinzip“ unter Pkt. 4.7. explizit erfolgt; außerdem wird dort festgestellt, dass die erweiterten Ermittlungsmöglichkeiten der Finanzverwaltung um strafprozessuale Maßnahmen diese Lücke häufig schließen.

➔ Zusätzlich besteht ein beanstandetes Gesamtbild eines Prüfungsnetzes wie der SRP erfahrungsgemäß aus mehreren (unabhängigen) Indizien, die sich als Indiziengesamtheiten⁸⁹¹ in ihrer Aussageklarheit (erheblich) verstärken (s. den Abschnitt „mehrere Indizien gegen die sachliche Richtigkeit“ mit einem Beispiel unter [Pkt. 4.4.](#)).

Entsprechend wird in der Literatur zwischenzeitlich erkannt: „Der weitere Vorteil von SRP besteht darin, dass die finanzamtliche Beweisführung für die Hinzuschätzung auch im Steuerstrafverfahren gerichtsfest ist.“⁸⁹²

Auch bei der Feststellung des **Umfangs von (Steuer-)Straftaten** als Ausgangspunkt für das Strafmaß gilt der Zweifelsgrundsatz: „Im Strafverfahren genügt eine Wahrscheinlichkeit nicht, stattdessen wird die sichere Überzeugung über die Höhe der geschätzten Besteuerungsgrundlagen gefordert.“⁸⁹³ Andererseits geht „[d]er ‚In-dubio-pro-reo‘-Grundsatz [...] aber nicht soweit, dass eine mathematische Gewissheit für eine Verurteilung erforderlich wäre.“⁸⁹⁴ Deshalb sind Schätzungen der hinterzogenen Steuern auch im Strafverfahren zulässig⁸⁹⁵ und „notwendig, wenn keine oder nur mangelhafte Aufzeichnungen vorhanden sind [...] Denn ansonsten müsste der Täter freigesprochen werden, obwohl seine Schuld als solche feststeht“⁸⁹⁶. Demzufolge sieht *Brinkmann* eine „[insgesamt deutlich verringerte] Zurückhaltung der strafgerichtlichen Rechtsprechung gegenüber finanzamtlichen Schätzungen“⁸⁹⁷.

Mit der **Quantilsschätzung** (s. [Pkt. 4.8.](#)) wird die Maßgabe einer Schätzung im Besteuerungsverfahren verfolgt, dass „der ‚Beweisverderber‘ oder ‚Beweisvereitler‘ aus seinem Verhalten keinen Vorteil ziehen darf (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO)“⁸⁹⁸. Danach „ist das FA im allgemeinen nicht nur berechtigt, sondern verpflichtet, in dem gege-

⁸⁹¹ BGH v. 27.4.2010 – 1 StR 454/09, NStZ 2011, 108, Rz. 18 u. 20: „Rechtsfehlerhaft ist es auch, wenn sich das Tatgericht bei seiner Beweiswürdigung darauf beschränkt, die einzelnen Belastungsindizien gesondert zu erörtern und auf ihren jeweiligen Beweiswert zu prüfen, ohne eine Gesamtabwägung aller für und gegen die Täterschaft sprechenden Umstände vorzunehmen [...] Denn die Indizien können in ihrer Gesamtheit [...] die entsprechende Überzeugung vermitteln, auch wenn [sie] jeweils für sich allein nicht zum Nachweis der Täterschaft des Angeklagten ausreicht.“

⁸⁹² *Pump/Heidl*, „Hilfestellung bei der Kassenführung durch den steuerlichen Berater – Ärger wegen § 162 AO durch unzulängliche Kassenbuchführung“, StBp 2014, 163. Demgegenüber setzt sich die widersprechende Literatur (vgl. insbesondere: *Beyer*, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931) mit diesen wahrscheinlichkeits-theoretisch beweiskraftverstärkenden Eigenschaften der SRP gar nicht auseinander, sondern beschränkt sich auf das Zitieren der methodisch nicht differenzierenden Rechtsprechung.

⁸⁹³ *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Aufl. 2023, Rz. 1218.

⁸⁹⁴ *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Aufl. 2023, Rz. 1218.

⁸⁹⁵ BGH v. 4.2.1992 – 5 StR 655/91, wistra 1992, 147 u. v. 24.9.1985 – 1 StR 313/85, wistra 1986, 65.

⁸⁹⁶ *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Aufl. 2023, Rz. 1219.

⁸⁹⁷ *Brinkmann*, „Strafrechtliche Aspekte der Schätzung“, StBp 9/2013, 250, 253.

⁸⁹⁸ BFH v. 15.2.1989, BStBl. II 1989, 462: „[...] Zur Vermeidung eines solchen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

benen Schätzungsrahmen an die oberste Grenze zu gehen.⁸⁹⁹ **Durch diese Ausrichtung ist sie nicht (direkt) für eine strafrechtliche Schätzung geeignet.**

Demgegenüber rekonstruiert die **Verteilungsanalyse** mit ihrer quantitativen Auswertung (s. [Pkt. 4.3.4.6.](#)) eine wahrscheinliche Nachfragesituation des geprüften Stpfl. aus dessen (betriebseigenen) Tageskasseneinnahmen auf der Grundlage wissenschaftlicher Verteilungserkenntnisse (s. [Pkt. 4.3.4.1.](#)). Sofern im Einzelfall das typische Manipulationsbild „Spitzeneingriff“ (s. [Pkt. 4.3.4.3.](#)) erkennbar ist und nicht-manipulative Ursachen (weitest gehend) ausgeschlossen werden können (s. [Pkt. 4.3.4.5.](#)), beziffert das Ergebnis dabei allerdings nur den Mindestschaden. Deshalb könnte diese Methode auch als strafrechtliches Schätzungsverfahren in Betracht kommen, falls keine konkreteren Informationen vorliegen: „Wenn der BGH – unter den von ihm vorgegebenen Voraussetzungen – im Rahmen der tatrichterlichen Beweiswürdigung selbst so generalisierende Schätzungsmethoden wie die Schätzung nach der Richtsatzsammlung zulässt, dann sollten auch die Ergebnisse der statistischen-mathematischen Methoden für die richterliche Überzeugungsbildung von Relevanz sein, da der Beweiswert eines inneren Betriebsvergleichs [...] sicherlich höher einzuschätzen ist als ein äußerer Betriebsvergleich (Richtsatzsammlung).“⁹⁰⁰

⁸⁹⁹ BFH v. 9.3.1967, BStBl. III 1967, 349.

⁹⁰⁰ *Madauß*, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstrafverfahren“, NZWiSt 1/2014, 24, 25. Die diesbezügliche Kritik „Zudem hinkt der Vergleich mit der Richtsatzsammlung. [...] Meines Erachtens ist die Richtsatzsammlung kein geeignetes Verprobungs- und Schätzungsinstrument“ gründet *Beyer* („Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter“, NWB 40/2019, 2931) lediglich auf seine persönliche Meinung, während sich andernorts in der Literatur differenzierter damit auseinandergesetzt wird wie z.B. in *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Aufl. 2023, Rz. 1227: „Der BGH hat dennoch die Anwendung der Richtsätze wegen der im Urteilsfall fehlenden Buchführung für möglich gehalten und darüber hinaus das Fehlen ordnungsgemäßer Aufzeichnungen sogar als Strafschärfungsgrund betrachtet. Es bleibt aber trotz bzw. gerade wegen der o.g. Entscheidung dabei, dass ein Rückgriff auf die amtlichen Richtsätze im Strafverfahren voraussetzt, dass keine konkretere Schätzungsmöglichkeit besteht und dass zur konkreten Situation des Betriebs Feststellungen getroffen und evtl. betriebliche Besonderheiten im Vergleich zu den ‚Normbetrieben‘ der Richtsatzsammlung ausreichend berücksichtigt werden.“

5. Freie SRP-Anwendung (systematische Datenvisualisierung)

Die **SRP-Vorlagen sind technische Einstiegshilfen** in eine systematisch-mehrperspektivische Datenvisualisierung für ein (möglichst) vollständiges, zügiges Erkennen von betriebspezifischen Prüffeldern und das zielführende Abarbeiten dieser mit aussagekräftigen, nachvollziehbar dokumentierten Ergebnissen. Sie sollen besonders Kolleg*innen mit geringer IT-Affinität einen Eindruck von den Vorteilen sinnvoll verbildlichender Prüfungsweisen vermitteln, ohne dass dazu umfangreiche Kenntnisse datenanalytischer (Tabellenkalkulations-)Programme im Vorfeld angeeignet werden müssen.

Voraussetzung alleine für die Lesekompetenz von Ergebnissen datenbasierter Prüfung – selbst bei Nutzung technisch vereinfachender Hilfsmittel wie Programm-Tools, Makros, Vorlagen und generativer „Künstlicher Intelligenz (KI)“⁹⁰¹ – sind jedoch **methodische Grundlagen**, weil weder die (Fach-)Begriffe, die statistischen und Visualisierungsfunktionen noch deren Unterschiede sowie Zusammenhänge intuitiv, also trivial sind⁹⁰² – geschweige denn das fallbezogene Erarbeiten eines planvollen effektiven Prüfungsablaufs (s. zum „*Prüfungsprozess*“ weiter unten unter diesem Pkt.). Trotz tlw. Umstrukturierungen ist es den Schulen bisher offensichtlich (noch) nicht gelungen, die heutigen Schlüsselqualifikationen zur (datenbasierten) Wissensbeschaffung, -validierung und -verknüpfung ausreichend zu vermitteln, und zumindest in den nicht-technischen rechtsnahen Berufsausbildungen werden diese Defizite auch nicht ausgeglichen (s. [Pkt. 3.3.](#)). Für ein BpO-konformes⁹⁰³ Intensivprüfen unter sinnvoller Nutzung aller zur Verfügung stehenden Hilfsmittel durch ein „hohes Maß an **Methodenwissen**, Flexibilität und Kreativität“⁹⁰⁴ müssen diese Fähigkeiten jedoch vorhanden sein, weshalb der Bundesrechnungshof fordert: „[D]as Bundesministerium der Finanzen (BMF) [muss] auf den bundesweit einheitlichen Einsatz IT-gestützter *quantitativer* Prüfungsmethoden in der steuerlichen Betriebsprüfung hinwirken [... denn t]raditionelle Methoden der steuerlichen Betriebsprüfung, beispielsweise Beleg- und Akteneinsicht, verlieren an Effektiv-

⁹⁰¹ Vgl. z.B.: Webinarreihe „TAXPUNK meets ChatGPT [...] 4. BEYOND TEXTFILES [...] mit] Stefan Werner“, in YouTube eingestellt unter: https://www.youtube.com/watch?v=xT_L2wPXxac.

⁹⁰² *Ihden*, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 64 u. 89.

⁹⁰³ § 7 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen. Ihre Dauer ist auf das notwendige Maß zu beschränken. Sie hat sich in erster Linie auf solche Sachverhalte zu erstrecken, die zu endgültigen Steuerausfällen oder Steuererstattungen oder -vergütungen oder zu nicht unbedeutenden Gewinnverlagerungen führen können.“

⁹⁰⁴ Internetseite des Lehrstuhls für Wirtschaftsprüfung, Unternehmensrechnung und Controlling von Prof. Mochty an der Universität Duisburg-Essen, früher nachzulesen unter: <https://www.wpunco.wiwi.uni-due.de/forschung/pruefungstechnik>.

tät und Effizienz. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.“⁹⁰⁵ Folgelogisch weist die Finanzverwaltung auf deren Bedeutung zwischenzeitlich grundsätzlich hin: „Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich.“⁹⁰⁶ Als **Methodik-Konzept**⁹⁰⁷ steht bei der Summarischen Risikoprüfung (SRP) das Verständnis der umfangreichen, interdisziplinären Grundlagen für die zwei wichtigen Methodenbereiche „regelbasierte (Zeit-)Reihenanalysen“ und „musterbasierte Häufigkeitsauswertungen“ im Mittelpunkt, damit diese in zielführenden Prüfungsprozessen sinnvoll angewendet werden können; insofern sind die SRP-Vorlagen als „Stützräder“ zu verstehen, um sich anfangs auf die Methodik konzentrieren zu können, während mittelfristig eine freie Anwendung der SRP-Prinzipien anzustreben ist. Auf Basis des methodischen Wissens lassen sich dabei auch deutlich komplexere, heutzutage häufig weitgehend automatisierte Analyseansätze wie z.B. multiple Regressionsanalysen (schrittweise) nachvollziehen.

Ausgangspunkt einer zeitgemäß effektiven und effizienten Prüfungsweise ist die Erkenntnis der **überlegenen Leistungsfähigkeit einer (systematisch-mehrperspektivischen) Datenvisualisierung** gegenüber dem Arbeiten mit Texten, Tabellen, Datenbanken etc.; diese lässt sich nicht nur aus den Erfahrungen der Werbebranche⁹⁰⁸ entnehmen, sondern wird ebenso von der Bildungswissenschaft kommuniziert: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. Die Analyse statistischer Daten beginnt häufig mit der Wahl einer geeigneten Darstellungsform des Erhebungsbefundes“⁹⁰⁹ und kann an dem Beispiel *im Abschnitt „Visualisierung von Steuerungsdaten“ unter Pkt. 2.* sowie an verbreiteten Diagrammen wie dem nachfolgenden zu den demografischen Verhältnissen in Deutschland⁹¹⁰ problemlos selbst nachvollzogen werden:

⁹⁰⁵ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

⁹⁰⁶ BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1.

⁹⁰⁷ Zwar wird ab der 3./12. Ausgabe des SRP-Handbuchs die Ausrichtung „Methodik-Konzept“ erstmalig explizit benannt, die entsprechenden Inhalte sind aber seit Begründung Bestandteil des Projekts.

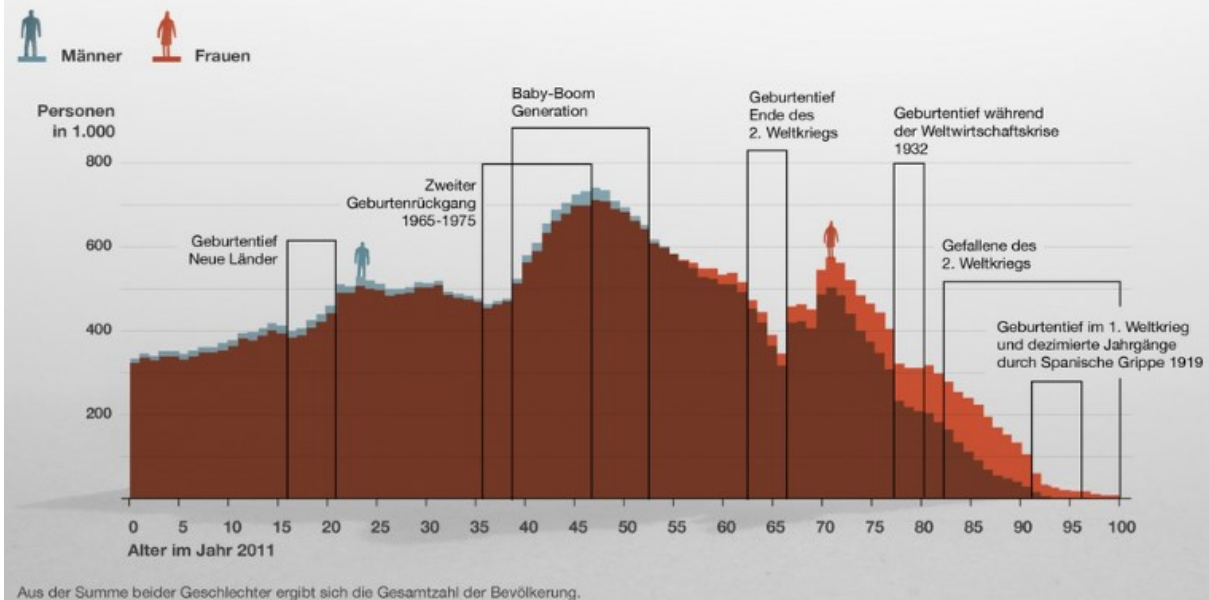
⁹⁰⁸ Barnard in einer US-amerikanischen Zeitschrift der Werbebranche 1921: „One look is worth a thousand words“, früher nachzulesen unter: www.oai.de/de/component/content/article/45-publikationen/sprichwort/811-ein-bild-sagt-mehr-als-tausend-woerter.html.

⁹⁰⁹ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89.

⁹¹⁰ Bildquelle: Internetseite der Bundeszentrale für politische Bildung (bpb), Bereich „Politik/Innenpolitik/Demografischer Wandel“, nachzulesen unter: <http://www.bpb.de/politik/innenpolitik/demografischer-wandel/209140/infografiken>; Lizenz: [cc by-nc-nd/3.0/de/](https://creativecommons.org/licenses/by-nc-nd/3.0/de/).

Die Altersstruktur Deutschlands nach Geschlecht 2011

Die jüngsten Jahrgänge sind zahlenmäßig kleiner als viele der älteren, obwohl die Zahl der Angehörigen eines Jahrgangs mit zunehmendem Alter abnimmt. Ursachen sind die gesunkene Geburtenrate und auch die verringerte Sterblichkeit der Älteren. Zudem veranschaulicht die Anzahl der Frauen und Männer in den jeweiligen Jahrgängen die Auswirkung historischer Ereignisse auf die Zusammensetzung der Bevölkerung.



Diese Besonderheit der (menschlichen) Gehirnphysiologie ist in den prüfenden Berufen tlw. offenbar seit langem bekannt, wenn Unternehmensverbünde in Form von Organigrammen anstatt als Tabelle wiedergegeben werden. Umso erstaunlicher ist es, dass die viel umfangreicheren und komplexeren Steuerungsdaten mitunter noch in Datenbank-/Tabellenform geprüft werden.

Inzwischen werben auch Hersteller von Datenanalysesoftware, die zuvor auf ein datenbankenbasiertes Arbeiten gesetzt haben, mit ihren Visualisierungsmöglichkeiten.⁹¹¹ Derweil außerdem in der Wirtschaftsprüfung jüngst erkannt wurde, dass der dominierende Trend der Datenanalytik in den letzten Jahren „Künstliche Intelligenz (KI)“⁹¹² die kostbaren menschlichen Stärken „Intuition“ und „Erfahrung“ (weitgehend) ungenutzt lässt⁹¹³, gewinnt die zielführende (systematisch-mehrperspektivische) Datenvisualisierung auch in steuernahen Berufen

⁹¹¹ Vgl. z.B.: Internetseite der Fa. Caseware zu IDEA V10: „IDEA V10. Neue Funktionen für intuitive Datenanalyse“, nachzulesen unter: <https://www.caseware.net/webinare/detail/idea-v10-neue-funktionen-fuer-intuitive-datenanalyse/>.

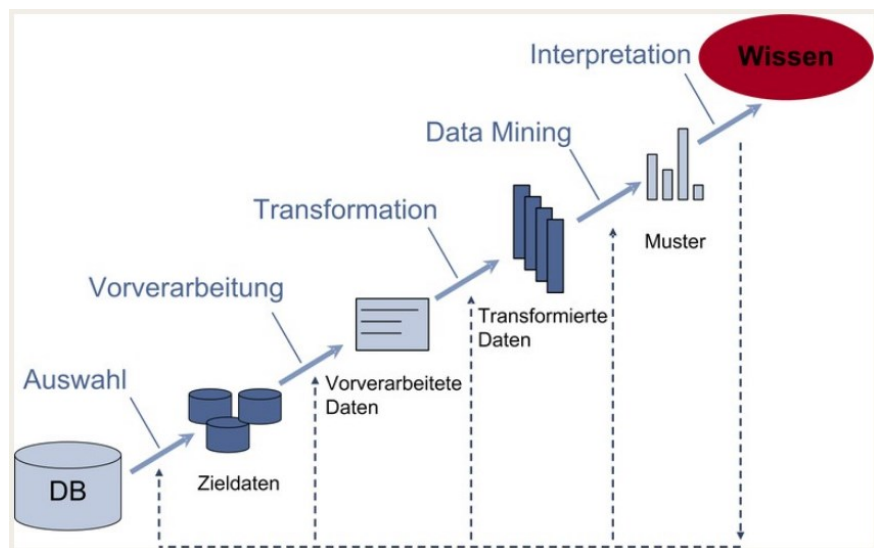
⁹¹² Kompenhans/Wermelt, „Big Data“ – Potenziale für innovative Abschlussprüfungen“, BB 2018, 299: „Die an Systemschnittstellen erforderlichen manuellen Tätigkeiten werden immer mehr durch den Einsatz von Robotic Process Automation ersetzt, Entscheidungsprozesse werden dabei durch den Einsatz programmierter, selbstlernender Entscheidungsheuristiken, die auf Basis künstlicher Intelligenz arbeiten, unterstützt.“

⁹¹³ DB online v. 30.4.2019 (Sponsored News v. EY), „Die Zukunft der Wirtschaftsprüfung ist digital: [...] Während Künstliche Intelligenz (KI) und Robotics Routinetätigkeiten übernehmen, konzentrieren sich die Prüfer von EY zunehmend auf die Interpretation der Daten. [...] Für Wirtschaftsprüfer heißt das, dass sie neben ihrem Fachwissen Kenntnisse im IT-Bereich benötigen, um diese Prozesse zu verstehen“, früher nachzulesen unter: <https://www.der-betrieb.de/meldungen/die-zukunft-der-wirtschaftspruefung-ist-digital/>.

zunehmend an Bedeutung; entsprechend bauen die in Wirtschaft und Beratung immer verbreiteteren B(usiness-)I(ntelligence)-Programme ihre überlegene Funktionalität wahrnehmungsfreundlich u.a. auf diversen miteinander verknüpften Verbildlichungen auf⁹¹⁴, was die Anforderungen an die Lesekompetenz abermals erhöht.⁹¹⁵

Zwar nimmt die Datenvisualisierung demnach i.R. von gleichermaßen effektiven wie effizienten Prüfungsmaßnahmen unter Einbindung von Menschen eine entscheidende Rolle ein, sie ist dabei aber Bestandteil in einem komplexeren systematischen **Prüfungsprozess**, wie er in der nachfolgenden Abb. schematisch dargestellt wird⁹¹⁶:

1. Auf Grundlage einer möglichst konkreten Prüfungsfrage werden zunächst die Daten aus dem Gesamtdatenbestand „DB“ eines Betriebs ausgewählt, welche die benötigten Informationen enthalten (sollten).



2. Damit die Visualisierungen nicht von Summen, Korrekturen, Steuerungsinformationen etc. verzerrt werden, müssen die „Zieldaten“ vor der eigentlichen Analyse von derartigen Bestandteilen bereinigt werden.
3. Häufig haben die bereinigten „[v]orverarbeitete[n] Daten“ nicht bereits die Tabellenform, die sich für die Umsetzung in eine möglichst aussagekräftige Grafik eignet. In diesen

⁹¹⁴ B(usiness-)I(ntelligence)-Programme verfolgen denselben Hauptzweck wie die SRP aus der Selbstanalyse des Unternehmens, nachzulesen z.B. unter: <https://www.bigdata-insider.de/das-sind-die-beliebtesten-business-intelligence-loesungen-a-467954/> mit: „Wie entwickelt sich das eigene Unternehmen? In welchen Regionen bricht der Absatz ein? Was macht eigentlich der Wettbewerb? Auf Business Intelligence zu setzen, das heißt, bei der Entscheidungsfindung auf Fakten anstatt auf das Bauchgefühl zu vertrauen!“

⁹¹⁵ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. [...] Die korrekte Interpretation solcher Grafiken ist allerdings nicht trivial und erfordert spezielle kognitive Fähigkeiten.“

⁹¹⁶ Abb. aus Chamoni, „Informations-, Daten- und Wissensmanagement/Data Mining“ der „Enzyklopädie der Wirtschaftsinformatik (Online-Lexikon)“ mit Verweis auf: Fayyad/Piatetsky-Shapiro/Smyth, „From data mining to knowledge discovery: an overview“, in: Fayyad/Piatetsky-Shapiro/Smyth/Uthurusamy (Hrsg.), „Advances in knowledge discovery and data mining“, Menlo Park et al., 1996, **früher** nachzulesen unter: <https://www.enzyklopaedie-der-wirtschaftsinformatik.de/lexikon/daten-wissen/Business-Intelligence/Analytische-Informationssysteme--Methoden-der-/Data-Mining/index.html>.

Fällen muss sich eine „Transformation“ des Datenextrakts in einen diagrammfähigen Aufbau der Tabelle anschließen.⁹¹⁷

4. Für die transformierten Daten lässt sich dann ein Diagrammtyp wählen, welcher die (betrieblichen) Informationen zur Ausgangsfrage anhand von „Mustern“ und Regeln möglichst übersichtlich aufbereitet, damit i.R. der „Interpretation“ Intuition und Erfahrung eingebracht werden können.

Die (gestrichelten) Pfeile in der Abb. zeigen an, dass ein systematischer Prüfungsprozess an jedem Schritt wieder zurückgeführt werden kann bzw. muss⁹¹⁸, wenn die (Zwischen-)Ergebnisse keine befriedigende Antwort auf die Prüfungsfrage geben (können). Sollte dies jedoch an Mängeln der Aufzeichnungen liegen, wird im Besteuerungsverfahren das besondere (geminderte) Beweismaß des § 158 Abs. 1 AO (s. [Pkt. 4.6.](#)) wirksam, damit „der ‚Beweisverderber‘ oder ‚Beweisvereitler‘ aus seinem Verhalten keinen Vorteil ziehen [...] (Rechtsgedanke des § 444 ZPO i.V.m. § 155 FGO)“⁹¹⁹ kann.

Da Prüfungsfragen und die dazu vorgehaltenen Betriebsdaten i.d.R. einzelfallspezifisch sind – abgesehen von wenigen Prüffeldern mit weitgehend ähnlichen Ausgangsbedingungen wie „Vollständigkeit von (Bar-)Einnahmen“ –, lassen sich auch die Aufgaben „Datentransformation“ und „Datenvisualisierung“ innerhalb systematischer Prüfungsprozesse nur bedingt automatisieren. Deshalb kann eine zeitgemäße Qualität von Prüfungen nur dadurch sichergestellt werden, dass den Prüfer*innen die *umfangreichen interdisziplinären* Methodik-Grundlagen der **freien Datenvisualisierung** nachhaltig als „echte“ Digitalkompetenz vermittelt werden.⁹²⁰ Es wird schon länger festgestellt⁹²¹, dass diese Herausforderung nicht durch die übli-

⁹¹⁷ Zur besseren Bedienfreundlichkeit hat sich als Ausgangspunkt der (visualisierenden) Datenanalytik insbesondere in Tabellenkalkulationsprogrammen die Pivot-Tabelle durchgesetzt, wodurch sich etliche derartige Transformationen erübrigen; dies gelingt jedoch nicht in allen Fällen. Alleine das Erkennen einer Transformationsnotwendigkeit setzt ein Grundverständnis für Datenmodelle voraus.

⁹¹⁸ Diese Variabilität eines Datenanalyseprogramms ist somit ein wichtiger Aspekt der notwendigen Funktionalität für ergebnisoffene, effektive und effiziente Prüfungsmaßnahmen.

⁹¹⁹ BFH v. 15.2.1989, BStBl. II 1989, 462: „[...] Zur Vermeidung eines solchen Ergebnisses sind auch belastende Unterstellungen [...] oder nachteilige Schlüsse im Rahmen der Beweiswürdigung gerechtfertigt“.

⁹²⁰ „Data Literacy gehört zu den wichtigen Future Skills, um in der digitalen Welt in Wissenschaft, Arbeitswelt und Gesellschaft bestehen und teilhaben zu können. Damit gehört sie zu den bedeutendsten Digitalkompetenzen der Zukunft. Data Literacy bezeichnet die Fähigkeit, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden. Sie ermöglicht die systematische Umwandlung von Daten in Wissen und ist damit unverzichtbarer Bestandteil der Allgemeinbildung“; nachzulesen unter: <https://www.stifterverband.org/data-literacy-education>.

⁹²¹ Internetseite von McKinsey & Company, „Ten red flags signaling your analytics program will fail: [...] These days, it's the rare CEO who doesn't know that businesses must become analytics-driven. Many business leaders have, to their credit, been charging ahead with bold investments in analytics resources and artificial intelligence (AI). [...] But frustrations are beginning to surface; it's starting to dawn on company executives that they've failed to convert their analytics pilots into scalable solutions“, May 2018, nachzulesen unter: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail#>.

che Strategie „ein paar zusätzliche Kurse“ bewältigt werden kann, was wohl auch (mit) zur Einrichtung einer Enquete-Kommission des Bundestags⁹²² geführt hat. M.E. wird diesbezüglich in Entscheidungsebenen häufig weiterhin nicht der Unterschied zwischen der Bedienung einer Abfragesoftware und der Erarbeitung von komplexen Lösungen mit Hilfe unzähliger digitaler Werkzeuge in diversen Programmen verstanden: Bei digitaler Analysekompetenz handelt es sich um einen Anspruch, der ein breites Wissen, geistige Anpassungsfähigkeit und die Bereitschaft zu ständigem Lernen voraussetzt.⁹²³

Ergänzend zum Methodik-Konzept der Summarischen Risikoprüfung (SRP) ist als länderübergreifendes Projekt ein umfangreiches, variables Fortbildungsmodul zum Erlernen systematischer digitaler Datenanalyse unter Einsatz sinnvoller Werkzeuge erarbeitet worden. Als angestrebte Erfahrungsstufe soll in möglichst großer Anwendungsbreite erreicht werden, dass die einzelfallabhängige Kombination aus vorlagengestützter und freier Datenvisualisierung i.V.m. anderen Ansätzen wie z.B. mathematisch-statistischen Stichproben (s. *den Abschnitt „statistische Stichprobenverfahren“ unter Pkt. 2.)* im Prüfungsalltag umgesetzt wird. Zur (Weiter-)Entwicklung von Schulungen, Material und technischen Werkzeugen ist die Bund-Länder-Unterarbeitsgruppe „Summarische Risikoprüfung (SRP) und visualisierende Datenanalyse“ eingesetzt worden. Mittelfristig werden sich allerdings auch Einstellungskriterien und Ausbildung stärker an die veränderten Arbeitsbedingungen anpassen müssen, denn die technischen Entwicklungen gehen rasant weiter und stellen immer höhere Anforderungen alleine an die Lesekompetenz der Ergebnisse, wobei die Verfahren an sich dadurch i.d.R. noch nicht auf ihre Sinnhaftigkeit hinterfragt werden.

5.1. Einsatzbereiche systematischer Visualisierung

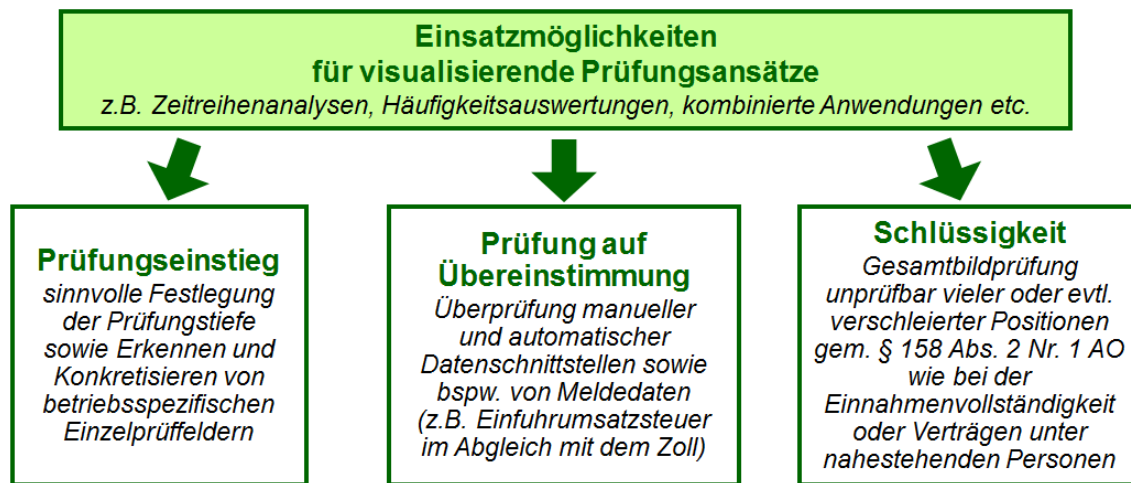
Weil gerade anfänglich bei fehlender Erfahrung mit neuartigen Prüfungsansätzen noch viel Unklarheit darüber besteht, in welchen Bereichen sie erfolversprechend eingesetzt werden können, soll die nachfolgende **nicht abschließende Übersicht** dabei helfen. Nur durch regelmäßig wiederholtes Anwenden (mindestens) in diesen Bereichen kann die notwendige Erfahrung und Übung mit systematisch(-mehrperspektivisch)er Datenvisualisierung (Analysekompetenz) resultieren, um die ganzen Vorteile in der Prüfung zu erfahren. Mit der Zeit entstehen die „[erforderliche] Flexibilität und Kreativität“⁹²⁴ (s. *Pkt. 5.*), um auf Basis der ver-

⁹²² Nachzulesen unter: <https://www.bundestag.de/dokumente/textarchiv/2018/kw39-pa-enquete-berufliche-bildung-567960>.

⁹²³ Wie sehr digitale Analysekompetenz noch immer unterschätzt wird, lässt sich bereits am Namen der geplanten Weiterqualifizierung laut Bundessteuerberaterkammer „Fach*assistent* für Digitalisierung und IT-Prozesse“ (Laubach, „Berufsstand braucht Kopfstand“, JUVE Steuermarkt Jan./Feb. 2021, 47.) erkennen, weil neben den methodischen und technischen Fähigkeiten *zusätzlich* für einen zielführenden Einsatz ein guter Überblick über das Steuerrecht vorhanden sein muss.

⁹²⁴ Früher nachzulesen unter: www.wpunco.wiwi.uni-due.de/forschung/pruefungstechnik.

fügbaren (digitalen) Informationen zu komplexen oder schwierigen Prüfungsfragen mit Hilfe der zur Verfügung stehenden Methodik und Technik eigene aussagekräftige Verbildlichungen zu erstellen; dabei ist der (zeitliche) Aufwand i.d.R. verhältnismäßig gering.



Derweil sich „[d]ie Buchführung [...] von der Durchschreibebuchführung hin zu einer umfassenden Datenverarbeitung entwickelt [hat], die elektronische Form [...] nunmehr die übliche [ist, ...] steigt die Anzahl elektronisch überwachter Prozesse stetig.“⁹²⁵ In der Folge ist es selbst bei kleinen Betrieben mit (tlw.) digitaler Aufzeichnung häufig nicht mehr möglich, Vollprüfungen durchzuführen (s. [Pkt. 2.1.](#)). Unsystematisches „Hineinprüfen“ alleine kann „die [nach BpO geforderte] Ermittlung und Beurteilung [all]er steuerlich bedeutsamen Sachverhalte“⁹²⁶ nicht gewährleisten: „Traditionelle Methoden der steuerlichen Betriebsprüfung, beispielsweise Beleg- und Akteneinsicht, verlieren an Effektivität und Effizienz. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.“⁹²⁷ Deshalb ist es inzwischen unerlässlich, dass für zielführende, ressourcenschonende Prüfungsmaßnahmen zu Beginn ein **systematischer Überblick** über die (gesamten) Besteuerungsdaten erarbeitet wird; auch die im Test befindliche Systemprüfung von Steuerkontrollsystemen mit der Möglichkeit, Beschränkungen der nächsten Außenprüfung zuzusagen,⁹²⁸ wird zumindest eine entsprechende Lesefähigkeit voraussetzen müssen. Hierbei ist die Datenvisualisierung aufgrund wahrnehmungsphysiologischer Vorteile (s. [Abschnitt „Visualisierung von Besteuerungsdaten“ unter Pkt. 2.](#)) unverzichtbarer Bestandteil (s. [Pkt. 5.1.1.](#)), wenn Risikoselektion nicht ausschließlich vollautomatisiert (RMS-

⁹²⁵ Graf, „Anforderungen an die Vorlage von Einzelaufzeichnungen – Ende der ‚Schonzeit‘ für Daten aus Vor-Systemen?“, BBK 18/2013, 875, 876 u. 877.

⁹²⁶ § 2 BpO: „Aufgaben der Betriebsprüfungsstellen: Zweck der Außenprüfung ist die Ermittlung und Beurteilung der steuerlich bedeutsamen Sachverhalte, um die Gleichmäßigkeit der Besteuerung sicherzustellen.“

⁹²⁷ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

⁹²⁸ Art. 97 § 38 EGAO.

BP⁹²⁹) geschehen sondern das Prüfungspersonal mit seinen Stärken „Intuition“ und „Erfahrung“ einbezogen werden soll.

Wesentliches Argument für die zunehmende Digitalisierung und Automatisierung von Betriebsverwaltungen⁹³⁰ war und ist die Einsparung von Personal. Dass die Lösungen (bisher) nicht immer so fehlerfrei arbeiten, wie es das allgemeine Vertrauen in die digitale Datenverarbeitung vermuten lassen würde, bestätigen sogar die Spezialist*innen in den Unternehmen.⁹³¹ Gemäß den GoBD „[hat der Steuerpflichtige f]ür die Einhaltung der Ordnungsvorschriften des § 146 AO [...] Kontrollen einzurichten, auszuüben und zu protokollieren.“⁹³² Aus regelmäßigen Prüfungserfahrungen ist bekannt, dass Unstimmigkeiten in Betriebsdaten nicht nur infolge von Migration sondern auch durch die Einstellung und Bedienung der komplexen Systeme und offenbar sogar innerhalb der automatisierten Prozesse auftreten. Darum wird schon länger in der Literatur darauf hingewiesen, dass die **Untersuchung von Datenschnittstellen** ein heutiges Schwerpunktprüffeld sein sollte.⁹³³ In Wirtschaft und Beratung werden technische Lösungen zur Verkettung zusammengehöriger Abläufe und Aufzeichnungen („Blockchain-Technologie“⁹³⁴) getestet, wie sie bspw. bereits in der Kassensicherungseinrichtung in Österreich im Einsatz sind⁹³⁵, und es werden neue Compliance-Standards⁹³⁶ inklusive „[d]igitale[r] Operationalisierung [von] Tax Compliance Management-

⁹²⁹ Vgl. z.B.: *Danielmeyer*, „Mit einem modernen Risikomanagement hin zur zeitnahen Betriebsprüfung?“, StBp 2021, 271.

⁹³⁰ *Graf*, „Anforderungen an die Vorlage von Einzelaufzeichnungen – Ende der ‚Schonzeit‘ für Daten aus Vor-Systemen?“, BBK 18/2013, 875, 876: „Die Buchführung hat sich von der Durchschreibebuchführung hin zu einer umfassenden Datenverarbeitung entwickelt, die elektronische Form ist nunmehr die übliche. [...] Diese Entwicklung geschah auf Betreiben und im Interesse der Unternehmen, weil es betriebswirtschaftlich notwendig war und ist, diese Einzeldaten zu erzeugen, z.B. für ein aussagekräftiges Controlling oder zur Arbeitsprozesssteuerung.“

⁹³¹ Vgl. z.B.: Newsletter 4-2016 des Forums „Elektronische Steuerprüfung“ v. 22.4.2016: „(IT-)Management: Ein Drittel aller Unternehmen verlieren Daten bei der Migration: Ein Drittel (32 %) der Organisationen verlieren Daten, während sie Daten zwischen Geräten migrieren oder Betriebssysteme aktualisieren. Das ist das Ergebnis einer weltweiten Befragung unter fast 600 IT-Administratoren“, nachzulesen unter: http://elektronische-steuerpruefung.de/newsarchiv/news_16_04.htm?m=1450705466#Beitrag7.

⁹³² „Grundsätze zur ordnungsgemäßen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD n.F.)“, BMF-Schreiben v. 28.11.2019, BStBl. I 2019, 1269, Rz. 100: „6. Internes Kontrollsystem (IKS)“.

⁹³³ Vgl. z.B.: *Webel/Danielmeyer*, „Schnittstellenverprobung elektronischer Betriebsverwaltungen – das kommende Standardprüffeld“, StBp 2015, 353, sowie *Danielmeyer/Neubert/Unger*, „Praxiserfahrungen zu Vorsystemen – Prüffeld: Visualisierte Schnittstellenverprobung“, StBp 2016, 322.

⁹³⁴ Vgl. z.B. „Betriebswirtschaftliche Prozesse in die Blockchain? – Machbarkeitsstudie von Telekom und DATEV zeigt Potenzial der Technologie auf“, nachzulesen unter: <https://www.datev.de/web/de/m/presse/pressemeldungen/meldungen-2018/betriebswirtschaftliche-prozesse-in-die-blockchain/>.

⁹³⁵ *Huber*, „Reform der Kassensicherung in Österreich – Teil 2 – Technische Umsetzung, Status und Erkenntnisse“, BBK 23/2016, 1148, 1150.

⁹³⁶ IDW-Praxishinweis zu „Tax Compliance Management Systemen“ (IDW PS 980), nachzulesen unter: <https://www.idw.de/idw/idw-verlautbarungen/idw-ps-980.html>. *Liekenbrock/Danielmeyer*, „Data Governance als Reaktion auf die zunehmend datengetriebene Betriebsprüfung?“, RET 2/2021, 32.

Systemen[en]⁹³⁷ angeführt; diese (technischen) Lösungen müssen allerdings ebenfalls verifizierbar sein. Auch hier bieten visualisierende Gegenüberstellungen der Daten an einer Schnittstelle eine schnelle und wirksame Prüfungslösung (s. [Pkt. 5.1.2.](#)).

Die **Schlüssigkeitsprüfung** von Gesamtbildern hat in der Besteuerungsverifikation seit jeher eine wichtige Bedeutung⁹³⁸, weil i.d.R. keine zweifelsfreie Wiederherstellung vergangener Vorgänge möglich ist: Prüfungen „basieren auf der Rekonstruktion der Vergangenheit anhand der Dokumentation der Stpfl. [...]“. Das bedeutet, selbst bei vollständig vorliegenden Daten und somit einer (vollständigen) Rekonstruktion sind Unsicherheiten enthalten. Liegen darüber hinaus keine vollständigen Unterlagen vor und die Daten werden mit Hilfe der herkömmlichen Methoden rekonstruiert, birgt das zusätzliche Unsicherheiten in den wiederhergestellten Daten.⁹³⁹ Besonders bei datenintensiven und eventuell verschleierte Sachverhalten müssen rechtliche Schlussfolgerungen *immer* unter größeren Unsicherheiten getroffen werden, worauf auch die Wissenschaft hinweist⁹⁴⁰. Deshalb gilt im Besteuerungsverfahren nach § 158 Abs. 2 Nr. 1 AO das reduzierte Beweismaß „Beanstandungsanlass“ (s. [Pkt. 4.6.](#)) für die Beweislastumkehr und eventuelle Amtspflicht der Schätzung. Hiervon ist nicht nur die Überprüfung erklärter (Bar-)Einnahmen auf Vollständigkeit betroffen, sondern bspw. auch, ob Verträge zwischen nahestehenden Personen steuerlich anzuerkennen sind. Um die Unsicherheiten so gering wie möglich zu halten, ziel(t)en die rechtlichen Anpassungen zur Dokumentation konzerninterner Verrechnung gerade in den letzten Jahren stets darauf ab, die Prüfbarkeit der häufig komplizierten Verhältnisse durch mehr Informationen sicherzustellen⁹⁴¹; dazu passend ist die Ausrichtung der neuen „Verwaltungsgrundsätze[n] Verrechnungspreise“ zur datenbasiert-transaktionalen Verrechnungspreisprüfung nachge-

⁹³⁷ Walz/Oehrle/Hirn, „Digitale Operationalisierung des Tax Compliance Management-Systems“, beck.digitax 4/2023, 199.

⁹³⁸ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 5: „Die gesetzliche Regelung des § 158 gründet sich insofern auf die Einsicht, dass es weder möglich noch wünschenswert ist, eine Unzahl von Geschäftsvorfällen lückenlos nachzuprüfen, und deshalb das äußere Gesamtbild einer Buchführung als ausreichende Grundlage für die Sachverhaltsfeststellung genügen kann.“

⁹³⁹ Schütt, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323, 324.

⁹⁴⁰ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag, Nomos, 2018, 145: „Die juristische Methodenlehre ist somit die Lehre von der richtigen und gerechten Entscheidung unter Unsicherheitslagen.“

⁹⁴¹ Vgl. z.B.: „BEPS 2015 Final Reports: Final BEPS package for reform of the international tax system to tackle tax avoidance (BEPS = Base Erosion and Profit Shifting)“, nachzulesen unter: <http://www.oecd.org/ctp/beps-2015-final-reports.htm> oder Umsetzung der EU-Richtlinie 2018/822 des Rats bezüglich des verpflichtenden automatischen Informationsaustauschs im Bereich der Besteuerung über meldepflichtige grenzüberschreitende Gestaltungen vom 25. Mai 2018 (ABl. L 139/1 vom 5. Juni 2018) – sog. DAC 6 – bspw. unter: https://www.bundesfinanzministerium.de/Content/DE/Downloads/BMF_Schreiben/Weitere_Steuerthemen/Abgabenordnung/2021-03-29-Anwendung-Vorschriften-Pflicht-Mitteilung-grenzueberschreitende-Steuer-gestaltungen.html.

schärft worden⁹⁴². Zusätzliche Beschreibungen und Daten nützen jedoch nur, wenn sie im Zusammenhang zielführend geprüft werden können. Gerade in der (betriebswirtschaftlichen) Betrachtung der Fremdüblichkeit konzerninterner Verrechnung hilft die Verbildlichung der dazugehörigen Einzeldaten **zuzüglich Simulationsmöglichkeiten**⁹⁴³ regelmäßig deutlich, um eine sachgerechte Einschätzung zu erlangen (s. [Pkt. 5.1.3.](#)).

5.1.1. Systematischer Überblick mit visualisierender Prüffeldsuche

Für das zügige Erlangen eines aussagekräftigen Überblicks über die (gesamten) Daten eines Prüfungsfalls zur *verhältnismäßigen*⁹⁴⁴ Feststellung einzelfallbezogener Prüffelder ist u.a. die chronologische Nachverfolgung von Periodensummen im Hinblick auf Entwicklungen und Abhängigkeiten (Zeitreihenanalytik) bestens geeignet. Diese Idee liegt auch den betriebswirtschaftlichen Analysen (BWA) zu Grunde und beachtet dabei die wahrnehmungsphysiologischen Vorteile der Verbildlichung [s. den Abschnitt „überlegene Leistungsfähigkeit einer (systematisch-mehrperspektivischen) Datenvisualisierung“ unter [Pkt. 5.](#)]. Werden bei derartigen **regelbasierten Ansätzen** zudem die Fuzzy-Theorien (s. den Abschnitt „Fuzzy-Theorien“ unter [Pkt. 2.](#)) berücksichtigt, wonach der Grad der Aufsummierung von Daten (Summationsebene) das mögliche Erkennen unterschiedlicher Auffälligkeiten stark beeinflusst, kann mit geringem Aufwand ein ergebnisoffener und tiefreichender Einblick in die Verhältnisse eines zu prüfenden Betriebs gewonnen werden.

Anhand solcher Zeitreihenanalysen wie der nachfolgend beispielhaft abgebildeten wurde der Einstieg in die Prüfung bei einem Großhandelsunternehmen gewählt.⁹⁴⁵ Die gute Übersicht über die umfangreichen Informationen hat für einen der Warenbereiche, welche vorab über die Kostenstellen aufgetrennt wurden, die im Diagramm markierten Auffälligkeiten ergeben: Der scheinbar **fehlende Erlös** konnte zügig als Erfassungsunregelmäßigkeit auf einem anderen Konto wiedergefunden werden. Demgegenüber führte die vertiefende Auseinandersetzung mit der offensichtlichen **Durchbrechung der typischen Umschlagsfrequenz** zu

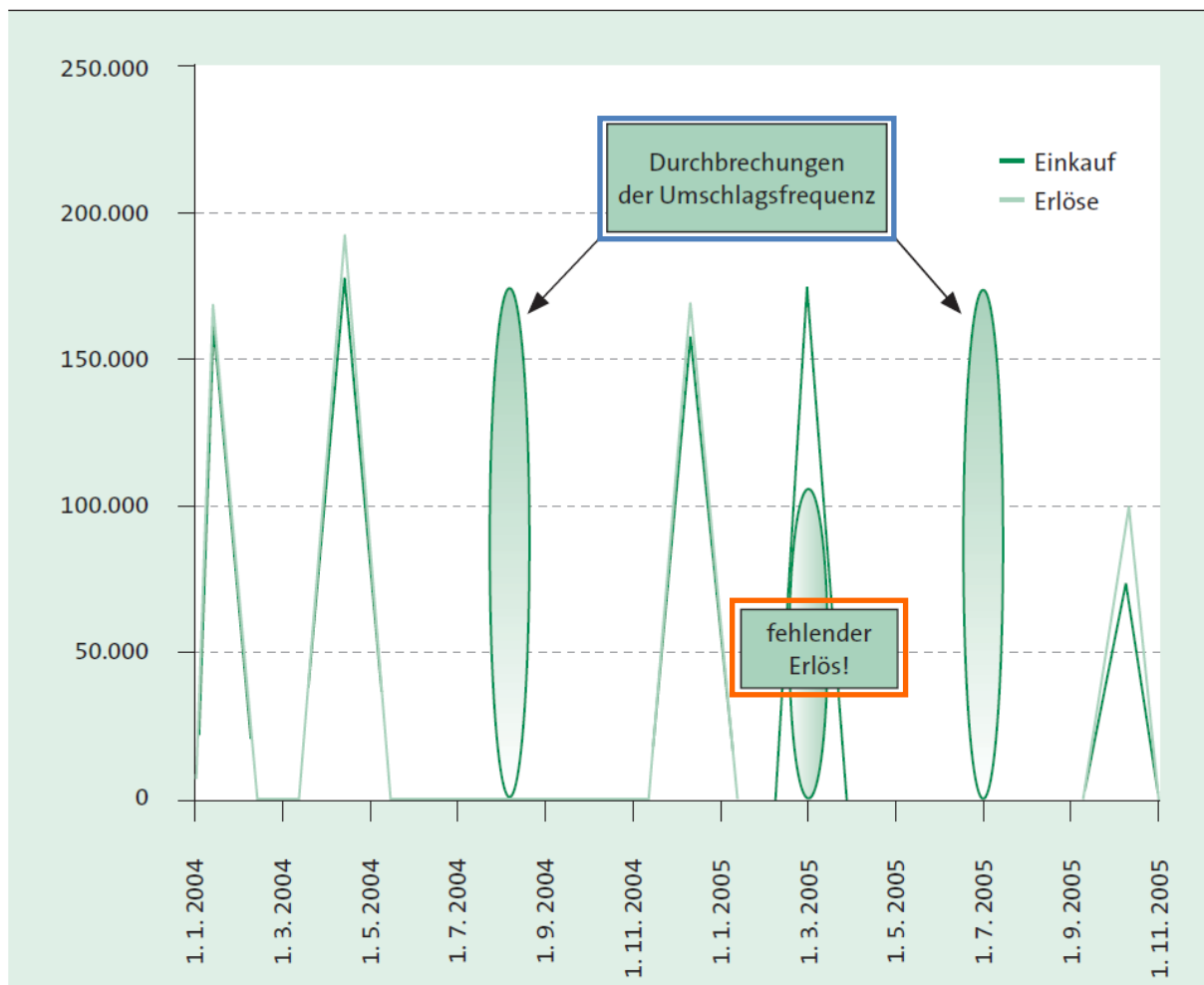
⁹⁴² BMF-Schreiben v. 6.6.2023, BStBl. I 2023, 1093.

⁹⁴³ Vgl. z.B.: PwC-Newsletter v. 31.8.2022: „Aus unserer Praxis: Der Einsatz von Data & Analytics im Verrechnungspreisumfeld – quo vadis? [...] D&A-Lösungen finden aber auch vielfach Einsatz im Rahmen von Betriebsprüfungen, z. B. für die Simulation der Steuereffekte bei Diskussionen um fremdübliche Bandbreiten. Mittels BI-Lösungen können diese Effekte aus Diskussionen mit der Betriebsprüfung direkt und dynamisch simuliert werden, was die Entscheidungsfindung in der Diskussion möglicher Kompromissvorschläge deutlich verbessert und verkürzt“, nachzulesen unter: <https://www.pwc.de/de/newsletter/transfer-pricing-perspectives-dach/ausgabe-55-2022/aus-unserer-praxis-der-einsatz-von-data-and-analytics-im-verrechnungspreisumfeld-quo-vadis.html>; dabei ist das systematisch-mehrperspektivische, interaktive Analysieren weniger eine Frage des Werkzeugs („BI-Lösungen“) als der zu Grunde liegenden Methodik.

⁹⁴⁴ § 88 Abs. 2 S. 1 AO: „Die Finanzbehörde bestimmt Art und Umfang der Ermittlungen nach den Umständen des Einzelfalls sowie nach den Grundsätzen der Gleichmäßigkeit, Gesetzmäßigkeit und *Verhältnismäßigkeit*“.

⁹⁴⁵ Bildquelle: Weiskopf/Wähnert, „Neue interaktive Prüfungstechnik – Transfereffekte moderner Verprobungsansätze auf die Großbetriebsprüfung“, BBK 3/2012, 127, 129.

einem unerwarteten Prüffeld: „Darüber hinaus sind die oszillierenden Einkaufs- und Verkaufsamplituden von Interesse, die das Produkt der betrieblichen Rahmenverhältnisse sind, zu denen sowohl die Lieferkapazitäten der Erzeuger als auch das Nachfrageverhalten der Kunden zählen. Bei dieser Produktgruppe ist die Tatsache bedeutsam, dass sie neben dem Wareneinsatz mit ergänzendem Provisionsaufwand belastet ist. Weitere durch das Verprobungsergebnis veranlasste Untersuchungen lieferten die Erkenntnis, dass es in der Belieferung künstliche Engpässe gab, um den Provisionsforderungen Nachdruck zu verleihen. Somit kann diese Art der Analyse auch für das Fraud-Management hilfreich sein.“⁹⁴⁶



Aus steuerlicher Sicht hat sich herausgestellt, dass im Zusammenhang mit den Provisionsstreitigkeiten erhebliche nicht abzugsfähige Betriebsausgaben gewinnwirksam erfasst worden waren.

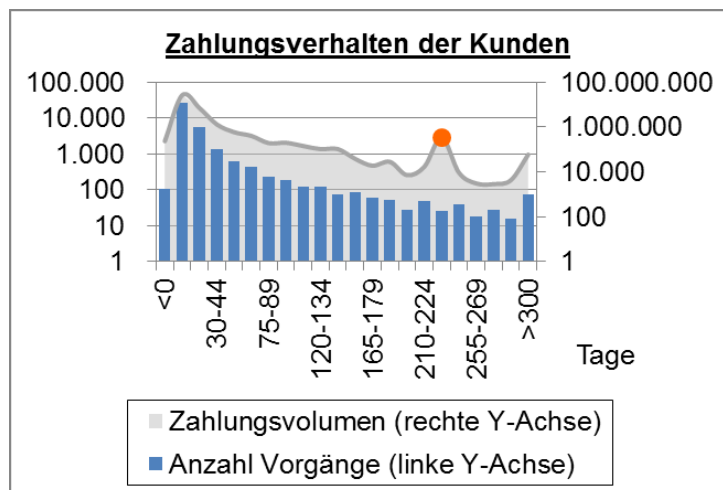
Werden **geeignete** (s. den Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) regelbasierte Untersuchungen an unterschiedlichen Unternehmensdaten zu Waren- bzw. Leistungsfluss, Kostenarten und -stellen, regelmäßigen Bewertungsfragen, Vertragsdurchführungen usw. systematisch vom groben Überblick zu Detailperspektiven – entsprechend der **SRP-Zeit-**

⁹⁴⁶ Weiskopf/Wähnert, „Neue interaktive Prüfungstechnik – Transfereffekte moderner Verprobungsansätze auf die Großbetriebsprüfung“, BBK 3/2012, 127, 129.

reihensystematik – durchgeführt (s. den Abschnitt „Eingrenzung aufgefundener Regelstörungen zu ursächlichen Sachverhalten“ unter [Pkt. 4.2.2.](#)), können zielführend und zeitsparend generelle wie punktuelle Prüffelder des Einzelfalls identifiziert werden.

Als weitere geeignete Verbildlichungsform zur Erlangung eines guten Überblicks über die (gesamten) Daten eines Prüfungsfalls setzen **musterbasierte Häufigkeitsauswertungen**⁹⁴⁷ auf der – in der Wirtschaftsprüfung üblichen – ABC-Klassifizierung auf, beachten dabei aber ebenfalls die wahrnehmungsphysiologischen Vorteile der Verbildlichung [s. den Abschnitt „überlegene Leistungsfähigkeit einer (systematisch-mehrperspektivischen) Datenvisualisierung“ unter [Pkt. 5.](#)]. Mit ihren merkmalsbezogenen Häufigkeitsmustern zu umfangreichen Unternehmensdaten (s. den Abschnitt „Auszählung von Informationen nach strukturierbaren Eigenschaften“ unter [Pkt. 4.3.](#)) arbeiten sie eine andere Information bzw. Ordnung als die regelbasierten Ansätze heraus; insofern konkurrieren regel- und musterbasierte Analysen bei der Gewinnung eines Überblicks über größere Datenmengen nicht, sondern ergänzen sich. Durch den Vergleich beobachteter Verteilungsmuster entweder mit betriebsinternen Vergleichsdaten oder mit allgemeinen (externen) Erfahrungen wie der Logarithmischen Normalverteilung (s. [Pkt. 4.3.4.1.](#)) können ungewöhnliche Werte und Häufungen effektiv aufgedeckt werden. Hierbei sollte ebenso an die Fuzzy-Theorien (s. den Abschnitt „Fuzzy-Theorien“ unter [Pkt. 2.](#)) gedacht werden, indem die Intervallgröße bei Klassifizierungen (s. den Abschnitt „Klassenunterteilung mit Wurzelmethode“ unter [Pkt. 4.3.3.1.](#)) darüber mitentscheidet, welche Auffälligkeiten erkennbar sind.

In dieser beispielhaften (mehrperspektivischen) Häufigkeitsauswertung wurden die Debitorendaten eines Unternehmens in einem Konzern auf das Zahlungsverhalten aller Kunden hin untersucht, u.a. weil mögliche Vorzugsbehandlungen verbundener Unternehmen steuerlich als verdeckte Gewinnausschüttungen zu behandeln sein können.⁹⁴⁸ Dazu wurden alle bezahlten Rechnungen in 15-Tages-Intervallen ausgezählt und



⁹⁴⁷ Die farblich hervorhebende Form der Musteranalytik zur einfacheren Wahrnehmung wird auch „Heatmap“ genannt; z.T. wird – anstatt der Logik „blau = normal versus rot = auffällig“ – auch „grün versus rot“ (Ampel-Logik) eingesetzt, was im Hinblick auf die verbreitete Rot-Grün-Sehschwäche jedoch problematisch ist.

⁹⁴⁸ Vgl. Weiskopf/Wähnert, „Neue interaktive Prüfungstechnik – Transfereffekte moderner Verprobungsansätze auf die Großbetriebsprüfung“, BBK 3/2012, 127, 131.

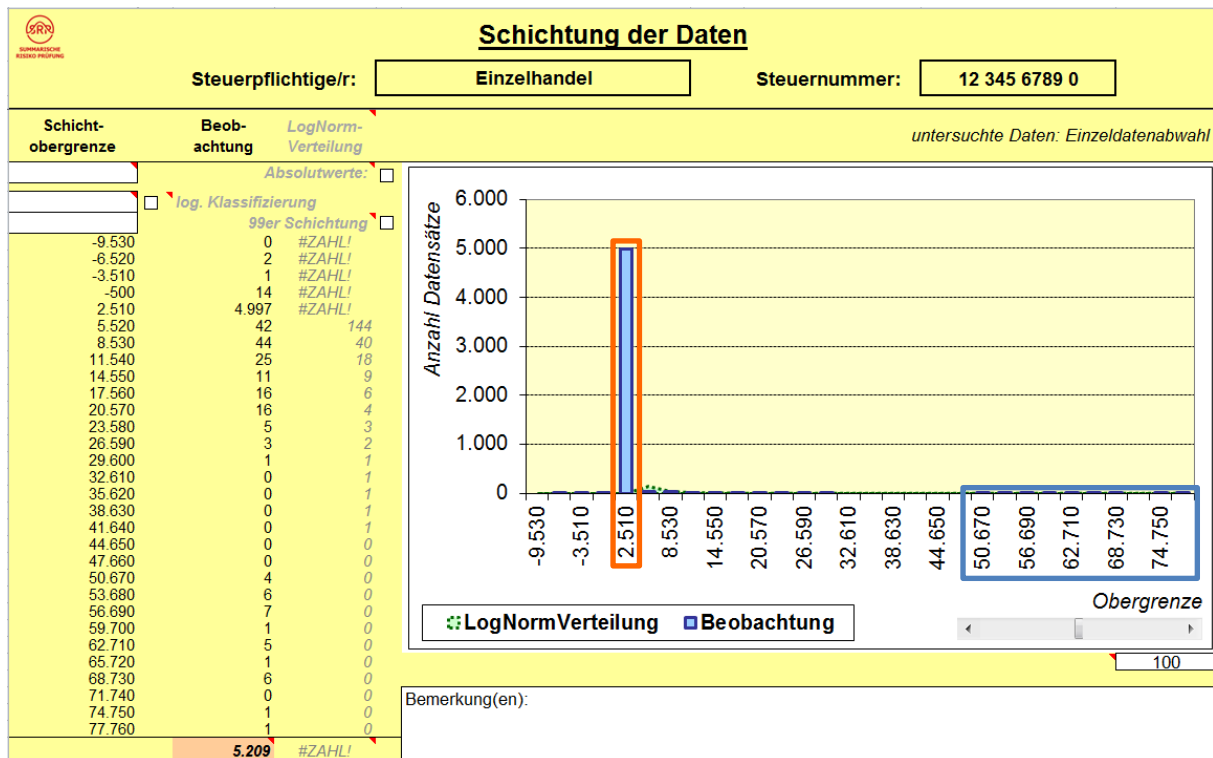
zusammen mit den Forderungsvolumina klassenweise im Diagramm abgebildet; für die beiden Y-Achsen wurde jeweils die logarithmische Form gewählt, um den Fokus auf einzelne Ausreißer als betriebspezifische Prüffelder zu lenken. Die Erwartung einer gleichmäßigen Häufigkeitsabnahme unter zunehmender Zahlungsdauer ist bis zur vorletzten Klasse weitgehend gegeben, denn die leichten Unruhen nach „165 bis 179 Tage“ bewegen sich im empfindlichen⁹⁴⁹ unteren Skalenbereich „Anzahl Vorgänge (linke Y-Achse) = 10 bis 100“. Die letzte Säule fasst alle Zahlungen ab einer Rechnungsferne von mehr als 300 Tagen zusammen, weshalb der gemäßigte Wiederanstieg nicht verwundert. Aus der zusätzlichen Information der klassenweise aufsummierten Forderungsvolumina sticht – im Gegensatz zur bisher festgestellten Gleichmäßigkeit der Häufigkeiten – das (nicht beschriftete) Intervall „225 bis 239 Tage“ hervor: Trotz einer (passend) geringen Anzahl an Vorgängen ergibt sich ein vergleichsweise hoher Forderungsbetrag von mehr als € 300.000; demnach müssen eine oder wenige werthaltige Rechnungen entsprechend lange unbezahlt geblieben sein. Die gezielte Filterung der Daten aus der betreffenden Klasse und die anschließende Begutachtung der dazugehörigen Unterlagen hat ergeben, dass es sich um wenige Rechnungen an ein verbundenes Unternehmen handelte, und dass die Langzeitforderung nicht verzinst wurde. Die geforderte⁹⁵⁰ Effizienz von Prüfungsansätzen ergibt sich zusätzlich daraus, dass neben diesem Einzelaufgriff keine weiteren (stichprobenartigen) Prüfungen der bezahlten Forderungen für das betrachtete Unternehmen vorgenommen werden müssen, weil alle anderen konzerninternen Verrechnungen im fremdüblichen Zahlungsverhalten unauffällig „untergehen“ – insofern bestärkt das Beispiel die Einschätzung des Bundesrechnungshofs: „Nur IT-gestützte quantitative Prüfungsmethoden können auffällige **Muster** und Beziehungen in großen Datenmengen ersichtlich machen.“⁹⁵¹

Im Hinblick auf eine **systematisch-mehrperspektivische Musteranalytik**, die den Fuzzy-Theorien (s. den Abschnitt „Fuzzy-Theorien“ unter [Pkt. 2.](#)) folgt, bietet die SRP-Vorlage Stochastik mit den beiden Tabellenblättern „Schichtung“ und „Schichtungsvertiefung“ praxisgerechte Unterstützungen, um große Mengen an Besteuerungsdaten gezielt auf Häufigkeitsmuster und prüfungsrelevante Abweichungen zu untersuchen:

⁹⁴⁹ Die gleich großen Abstände der beiden logarithmischen (zur Basis „10“) Y-Achsen nehmen in ihrer Bedeutung von oben nach unten pro Intervall (gleitend) um den Faktor „10“ ab. Somit werden bspw. die Schwankungen um „210 bis 224 Tage“ und „255 bis 269 Tage“ herum 10-mal intensiver als Unruhen im Häufigkeitsintervall „100 bis 1.000“ dargestellt.

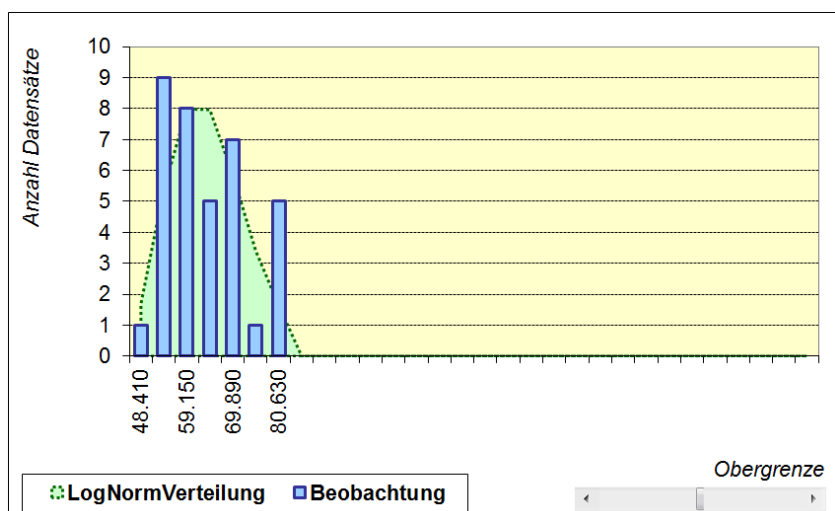
⁹⁵⁰ § 7 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen. Ihre Dauer ist auf das notwendige Maß zu beschränken. Sie hat sich in erster Linie auf solche Sachverhalte zu erstrecken, die zu endgültigen Steuerausfällen oder Steuererstattungen oder -vergütungen oder zu nicht unbedeutenden Gewinnverlagerungen führen können.“

⁹⁵¹ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.



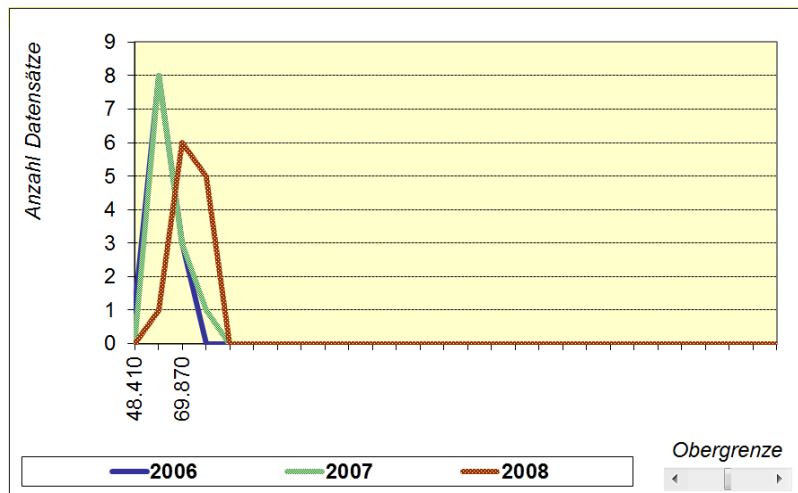
Nach der SRP-Prüfungssystematik „vom Groben ins Detail“ ist in dieser Übersichtsgrafik zu einem Einzelhandelsbeispiel auf dem Tabellenblatt „Schichtung“ der wertabhängige Verteilungszusammenhang für die gesamten Kostenpositionen zu sehen; dabei weist der Hinweis „untersuchte Daten: Einzeldatenabwahl“ rechts oberhalb des Diagramms darauf hin, dass einige Vorgänge (Summenumbuchungen) aus der Musteranalyse ausgeschlossen wurden (s. den Auflistungspkt. „Ausschluss einzelner Datensätze“ unter [Pkt. 4.3.2.1.](#)): **Nahezu alle Sachverhalte fallen in die Klasse „> € -500 bis € 2.510“**, weil eine **Gruppe von ungewöhnlichen Höchstbeträgen** bewirkt hat, dass die automatischen Auszahlungsintervalle unverhältnismäßig groß gebildet wurden; gemäß Warnfarbe des Summenfelds werden dennoch nicht alle Datensätze angezeigt (s. den Auflistungspkt. „nicht alle hohen Spitzenwerte in den Auswertungen berücksichtigt“ unter [Pkt. 4.3.3.1.](#)). Außerdem können aus der Tabelle drei Kostenpositionen mit erheblichen Negativbeträgen abgelesen werden, welche als Einzelprüffelder gefiltert und rechtlich geprüft werden können.

In diesem ersten Vertiefungsschritt werden die außergewöhnlich hohen Aufwendungen durch Ein-



gabe eines entsprechenden Startwerts (s. den Auflistungspkt. „Startwert“ unter [Pkt. 4.3.3.1.](#)) isoliert betrachtet: Insgesamt 36 Datensätze stellen einen offenbar eigenständigen Bestand an Kostenpositionen dar, zu denen sich durch eine gezielte Datenfilterung der Buchführung mit (exemplarischer) Belegsichtung umgehend herausfinden lässt, dass es Monatsabrechnungen eines Großhändlers sind. Trotzdem ergibt sich für diese kein gleichmäßiges Häufigkeitsmuster.

Mittels interner Gegenüberstellung der drei Prüfungsjahre wird die Ursache hierfür auf dem Tabellenblatt „Schichtungsverteufung“ untersucht. Dabei offenbart sich, dass die ersten beiden Jahre eine fast identische Häufigkeitsverteilung aufweisen, während die zwölf Monats-



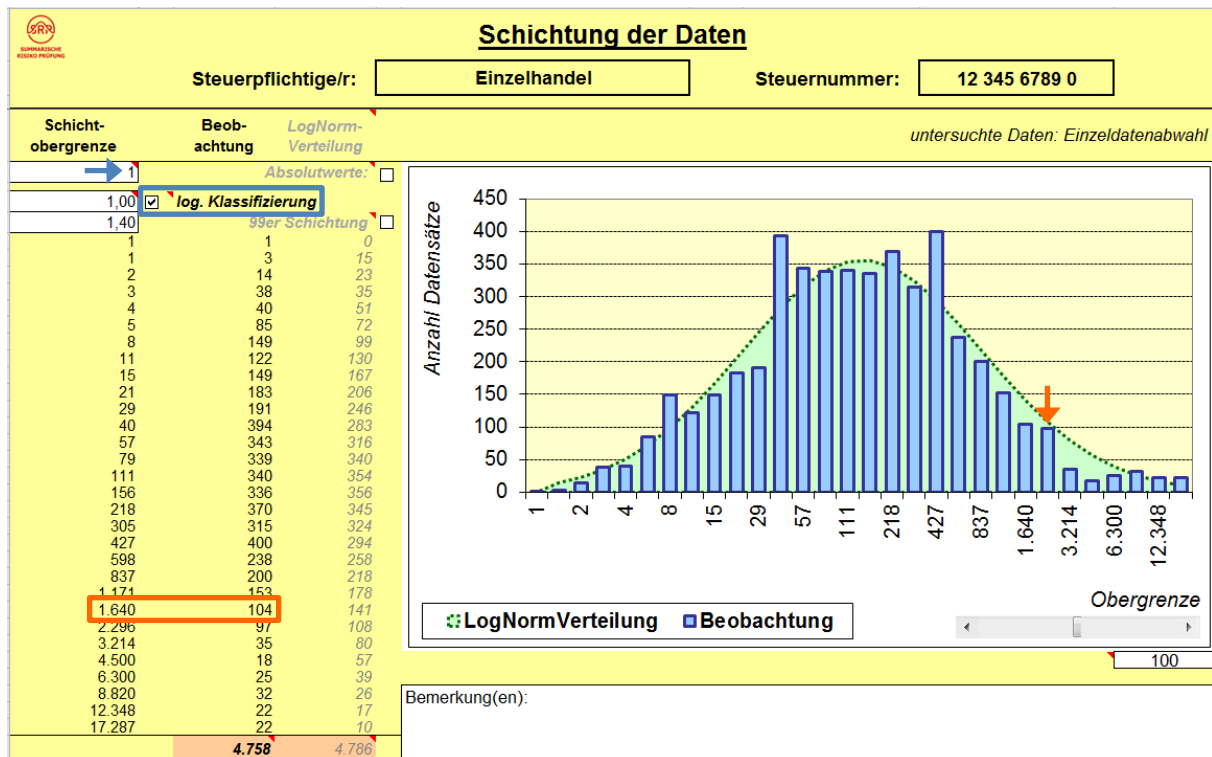
rechnungen des letzten Jahres insgesamt ein deutlich höheres Niveau erreichen. Sofern sich diese Veränderung nicht durch einen höheren Gesamtumsatz oder den Umstieg von einem anderen Lieferanten nachvollziehbar begründen lässt, sollte z.B. stichprobenartig nachgeprüft werden, ob (steuerlich) kritische Vorgänge die Auslöser sind.

Die niedrigeren Aufwendungen werden als zweite Vertiefung in der Abb. unterhalb durch eine schrittweise verfeinerte Schichtung (s. den Auflistungspkt. „freie Klassenbildung“ unter [Pkt. 4.3.3.1.](#)) auf deren Häufigkeitsmuster und prüfungsrelevante Datenkonzentrationen untersucht; infolge der extremen Intervallverzerrung durch die – bereits untersuchten – Höchstbeträge reicht die Spannweite des Schiebereglers zur zielführenden Kleinerstellung der Klassen nicht aus, weshalb diese über eine manuell austestende Eingabe der ersten beiden Schichtobergrenzen (im Beispiel: „1,00“ und „1,40“) vorgenommen werden muss. Durch die [logarithmische](#)⁹⁵² [Klassifizierungsform](#) wird der Betrachtungsfokus auf die niedrigen Werte gelegt, während die Festlegung des [Startwerts](#) „1“ die Analyse auf positive Datensätze einschränkt, sodass als grobe Orientierung die – unter logarithmischen Bedingungen symmetrische – Logarithmische Normalverteilung vollständig berechnet werden kann⁹⁵³. Als Ergebnis zeigt sich z.B. im rechtsseitigen Auslauf eine (relative) [Vorgangshäufung](#) des Intervalls „> € 1.640 bis € 2.296“ im Vergleich zu den benachbarten. Ob es sich dabei um eine vertrag-

⁹⁵² Eine logarithmische Klassenbildung belegt die gleichmäßigen Intervalle mit linear ansteigenden Exponenten wie bspw. „10² – 10³ – 10⁴“ (entsprechend „100 – 1.000 – 10.000“).

⁹⁵³ Da es keinen Logarithmus zu Null und negativen Zahlen gibt, kann die Herleitung der betriebspezifischen Logarithmischen Normalverteilung nur für den positiven Wertebereich erfolgen.

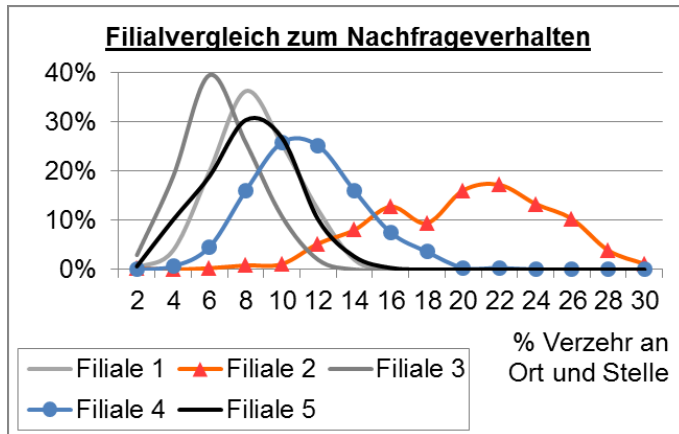
liche Festlegung von (Perioden-)Werten oder ein sinnvolles Vertiefungsfeld handelt, kann über die Filterung der betreffenden Datensätze auf dem Tabellenblatt „Vergleichsdaten“ und die anschließende Sichtung der Belege zielführend ermittelt werden.



Am effektivsten lässt sich ein aussagekräftiger Überblick über die (gesamten) Daten eines Prüfungsfalls durch **systematische Kombinationen aus unterschiedlichen Datenverbildlichungen** erzielen. Da regel- und musterbasierte Ansätze als chronologische bzw. Verteilungsperspektive auf Besteuerungsdaten unterschiedliche Informationen herausarbeiten, ergänzen diese sich i.S. eines mehrperspektivischen Sehens bei der Prüffeldaufdeckung und -eingrenzung (s. das Beispiel im Abschnitt „Prüfungsnetz“ unter [Pkt. 2.3.](#)).

Als Beispiel einer systematisch-mehrperspektivischen Datenverbildlichung zu einem konkreten *Einzelpüffeld* soll die Trennung der Entgelte in einem Großbetrieb mit vielen Filialen dienen. Um über Millionen von Einzeldaten aus den einheitlichen Kassensystemen einen ersten Überblick für gezielte Vertiefungen zu gewinnen, werden die prozentualen Anteile „Verzehr an Ort und Stelle“ pro Tag mit Hilfe des Erfassungsschlüssels als spezifische Häufigkeitsmuster der einzelnen Filialen gegenübergestellt. Infolge der ja-nein-Entscheidung, ob im Geschäft oder außerhalb verzehrt wird, müsste sich pro Filiale eine Binomialverteilung um den typischen Anteil herum ausbilden. Zur besseren Übersichtlichkeit werden in diesem Diagramm nur fünf der Filialen mit ihren Häufigkeitsmustern als prozentuale Liniengrafen⁹⁵⁴

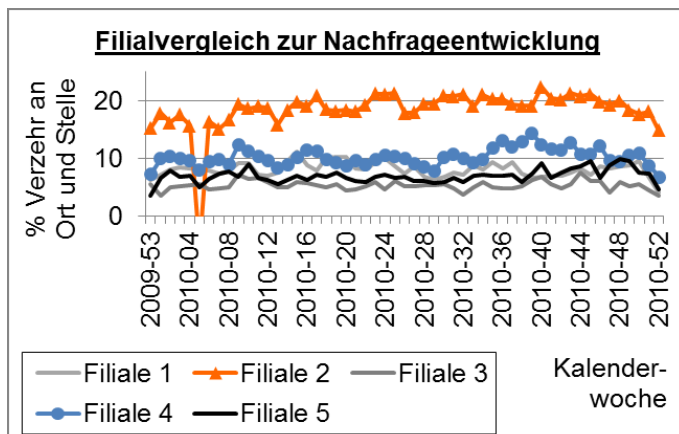
⁹⁵⁴ Bei der Binomialverteilung für ja-nein-Vorgänge handelt es sich um eine diskrete Verteilung mit ausschließlich ganzzahligen Ergebnissen ohne Zwischenwerte, sodass Säulen eigentlich die richtige Darstellungsform wä-



wiedergegeben. Dabei zeigen die Verteilungen der Verkaufsstellen „1“ und „3“ das erwartete symmetrische Verhalten und auf der X-Achse mit „ca. 8 %“ und „ca. 6 %“ den jeweils üblichen Tagesanteil „Verzehr an Ort und Stelle“. Die Filialen „4“ und „5“ haben weitgehend glockenförmige Verteilungen mit **leichten Auffälligkeiten im Aus-**

lauf bzw. Anstieg. Als eindeutig prüfungsrelevante Unternehmenseinheit sticht **Filiale „2“ mit einer unruhigen, zweigipfligen Häufigkeitskurve** hervor, weshalb gem. § 7 BpO⁹⁵⁵ der Vertiefungsschwerpunkt besonders auf diese gelegt wird.

Zunächst sollen durch die gegenüberstellende Ergänzung der regelbasierten Zeitreihendarstellung zur Entwicklung der Anteile „Verzehr an Ort und Stelle“ mögliche Ursachen für



die auffällige Häufigkeitsverteilung von **Filiale 2** hinterfragt werden. Der von anfänglich ca. 15 % tendenziell ansteigende Graf mit Stabilisierung um 20 % ab etwa der 15. Kalenderwoche legt den Gedanken nahe, dass dies das Häufigkeitsmuster beeinflusst haben könnte. *Nebenbei lässt sich dem Diagramm noch entnehmen, dass **Fili-***

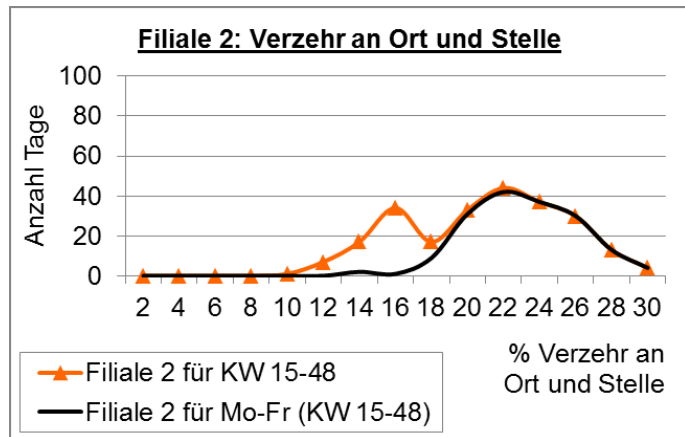
ale 4 vor allem zwischen der 35. und 46. Kalenderwoche eine deutliche Absatzveränderung mit höherem Anteil hatte. Das erklärt zwar deren langsameren (rechtsschiefen) Auslauf der Häufigkeitsverteilung (Musteranalyse der vorherigen Abb.), wirft aus Prüfungssicht allerdings folgende Frage auf: Ist der Auslöser für den zeitweisen Effekt ein verändertes Verzehrverhalten der Kunden oder eine tlw. falsche Systembedienung?

Die weitere Prüfungsvertiefung zur Trennung der Entgelte in Filiale 2 setzt auf der neuen Information auf, dass deren typischer Anteil „Verzehr an Ort und Stelle“ erst im Laufe des Jahres erreicht wurde. Deshalb werden die Daten auf den **Zeitraum „15. bis 48. Kalenderwo-**

ren. Zudem wird durch die (stetig) prozentuale Wiedergabe die Streuungsabhängigkeit von der Anzahl Daten (s. die Abb. im Abschnitt „Chi-Quadrat-Test“ unter Pkt. 4.3.1.3.) ausgeblendet. Derweil zur Auswahl der besonders prüfungsrelevanten Filialen jedoch Abweichungen von der Verteilungssymmetrie am interessantesten sind, wurde die abgebildete Darstellungsvariante gewählt.

⁹⁵⁵ § 7 S. 1 BpO: „Prüfungsgrundsätze: Die Außenprüfung ist auf das Wesentliche abzustellen.“

che“ beschränkt.⁹⁵⁶ Da hierdurch aber die Doppelstruktur mit zwei Gipfeln nicht aufgelöst wird, kann der anfänglich geringere Anteil nicht die Erklärung liefern. In einem anschließenden Schritt werden die Daten für den Betrachtungszeitraum „15. bis 48. Kalenderwoche“ nach Wochentagen differenziert. Mittels Ein- und Ausblenden



einzelner Wochentage unter ständiger Beobachtung des Häufigkeitsmusters lässt sich zügig erkennen, dass die Wochenendtage für den ersten Häufungshochpunkt mit einem typischen Anteil „Verzehr an Ort und Stelle“ von nur ca. 16 % verantwortlich sind; der Ausschluss der Wochenendtage aus dem Häufigkeitsmuster löst die Doppelstruktur mit zwei Gipfeln folgerichtig auf. Somit hat sich durch systematisch-mehrperspektivische Verbildlichung einer unprüfbar großen Anzahl an Steuerungsdaten eine weitere klare Prüfungsfrage herauskristallisiert: Warum weisen die Wochenendtage **ausschließlich** in Filiale 2 deutlich weniger Verzehr an Ort und Stelle auf als die anderen Wochentage? Kann es sich dabei um einen Bedienfehler des Wochenendpersonals handeln?

Dieses Beispiel zeigt, wie die gezielte Verknüpfung von verbildlichenden Untersuchungen in großen Datenmengen eine leistungsstarke sowie zügige Risikoorientierung mit konkret herausgearbeiteten Prüfungsfragen erzielen kann⁹⁵⁷ – genauso wie durch das Datenmodell *im Bereich „Haupteinsatzbereich“* unter [Pkt. 3.1.](#) inklusive Einbindung weiterer „IT-gestützter quantitativer Prüfungsmethoden“⁹⁵⁸. Bereits für die Lesekompetenz von (Zwischen-)Ergebnissen und besonders für das zielführende Navigieren in geeigneten Hilfsmitteln ist allerdings ein fundiertes methodisches und prozessuales Verständnis Voraussetzung (s. [Pkt. 3.3.](#) und den Abschnitt „Prüfungsprozess“ unter [Pkt. 5.](#)).

⁹⁵⁶ Die Umstellung der Y-Achse von Prozentanteilen auf Tage dient der anschließenden Untersuchung nach Wochentagen; die Y-Achsenkalierung ist dabei so gewählt, dass das Diagramm mit der Filialgegenüberstellung vergleichbar ist.

⁹⁵⁷ So auch: BMF-Schreiben v. 5.9.2023, „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, BStBl. I 2023, 1594, Rz. 1: „Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich.“

⁹⁵⁸ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20: „Gleichmäßige Besteuerung gefährdet: BMF soll auf einheitliche Maßstäbe für den Einsatz IT-gestützter Prüfungsmethoden in der Betriebsprüfung hinwirken“, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

5.1.2. Visualisierte Übereinstimmung (Schnittstellenprüfung)

Im Vorfeld punktueller inhaltlicher Prüfungshandlungen sollte eine Überprüfung der vorliegenden Unterlagen auf die Einhaltung von Ordnungsvorgaben (s. [Pkt. 4.1.](#)) – also auf die *Prüfbarkeit* – und auf die rechnerische Schlüssigkeit im Überblick erfolgen. Während diese vorgelagerten Prüfungsschritte in Zeiten überwiegend manueller Betriebsverwaltung selbstverständlich waren, besteht gegenüber elektronischen Datenerfassungs-, -verarbeitungs- und -speicherprozessen nicht dieselbe professionelle Skepsis. Nachdem i.R. von Betriebsprüfungen Mängel in der Datenkonsistenz bei diversen Unternehmen aufgefallen sind und das Thema „Schnittstellenprüfung“ publiziert wurde⁹⁵⁹, entwickelt sich ein neues (zeitgemäßeres) Verständnis für „Tax-Compliance-Management-Systeme (TCMS)“ heraus: „Es müssen die Bereitschaft, die personelle Kapazität und das Budget vorhanden sein, damit die Softwarelösungen optimal in die Unternehmensprozesse integriert werden können, was insbesondere eine laufende Pflege der Programme und insbesondere der (unvermeidbaren) Schnittstellen erfordert“⁹⁶⁰ – mit Festlegung neuer Compliance-Standards⁹⁶¹. Korrespondierend fordern die GoBD mit einem internen Kontrollsystem (IKS) „[...] Erfassungskontrollen (Fehlerhinweise, Plausibilitätsprüfungen), Abstimmungskontrollen bei der Dateneingabe, Verarbeitungskontrollen [und] Schutzmaßnahmen gegen die beabsichtigte und unbeabsichtigte Verfälschung von Programmen, Daten und Dokumenten“⁹⁶² als Voraussetzung für die Ordnungsmäßigkeit. Um mit der Personaleinsparung einen Hauptvorteil der elektronischen Betriebsverwaltung dennoch aufrecht erhalten zu können, wird in Wirtschaft und Beratung zwischenzeitlich verstärkt an technischen Sicherungslösungen wie der „Blockchain-Technologie“⁹⁶³ gearbeitet. Trotz der Forderung des BDI⁹⁶⁴ „[d]as Bestehen eines TCMS mit oder ohne Testat eines Wirtschaftsprüfers sollte zu einer beschleunigten Betriebsprüfung führen“, gehören **Schnittstellenprüfungen** – unabhängig von der Größe und Art des zu prüf-

⁹⁵⁹ Vgl. z.B.: *Webel/Danielmeyer*, „Schnittstellenverprobung elektronischer Betriebsverwaltungen – das kommende Standardprüffeld“, StBp 2015, 353, sowie *Danielmeyer/Neubert/Unger*, „Praxiserfahrungen zu Vorsystemen – Prüffeld: Visualisierte Schnittstellenverprobung“, StBp 2016, 322.

⁹⁶⁰ *Liekenbrock*, „Digital Tax Compliance – Praxishinweise zur technischen Umsetzung von Tax Compliance Management Systemen“, Ubg 2018, 43, 49.

⁹⁶¹ IDW-Praxishinweis zu „Tax Compliance Management Systemen“, (IDW PS 980), nachzulesen unter: <https://www.idw.de/idw/idw-verlautbarungen/idw-ps-980.html>. *Liekenbrock/Danielmeyer*, „Data Governance als Reaktion auf die zunehmend datengetriebene Betriebsprüfung?“, RET 2/2021, 32.

⁹⁶² „Grundsätze zur ordnungsgemäßen Führung und Aufbewahrung von Büchern, Aufzeichnungen und Unterlagen in elektronischer Form sowie zum Datenzugriff (GoBD n.F.)“, BMF-Schreiben v. 28.11.2019, BStBl. I 2019, 1269, Rz. 100.

⁹⁶³ Vgl. z.B. „Betriebswirtschaftliche Prozesse in die Blockchain? – Machbarkeitsstudie von Telekom und DATEV zeigt Potenzial der Technologie auf“, nachzulesen unter: <https://www.datev.de/web/de/m/presse/pressemitteilungen/meldungen-2018/betriebswirtschaftliche-prozesse-in-die-blockchain/>.

⁹⁶⁴ Bundesverband der Deutschen Industrie e.V. (BDI), „Steuerliche Betriebsprüfungen – Optionen für zeitgemäße Reformen:“, 13, nachzulesen unter: <https://bdi.eu/publikation/news/steuerliche-betriebspruefungen/>.

den Betriebs – zur Sorgfaltspflicht dazu, solange Programme zur (steuerlichen) Aufzeichnung und Aufbewahrung nicht standardmäßig über *wirksame, zügig verifizierbare* Absicherungen gegen Dateninkonsistenzen verfügen.⁹⁶⁵

Als Beispiel wird hier schematisch ein (Auszug aus einem) **Warenwirtschaftssystem** dargestellt, welches die vom Handel gewünschten Effizienz- und Transparenzfunktionen erfüllen soll; dafür müssen die benötigten Informationen über Stammdaten oder Eingaben beigesteuert werden und diverse Verarbeitungsprozesse automatisch ablaufen:



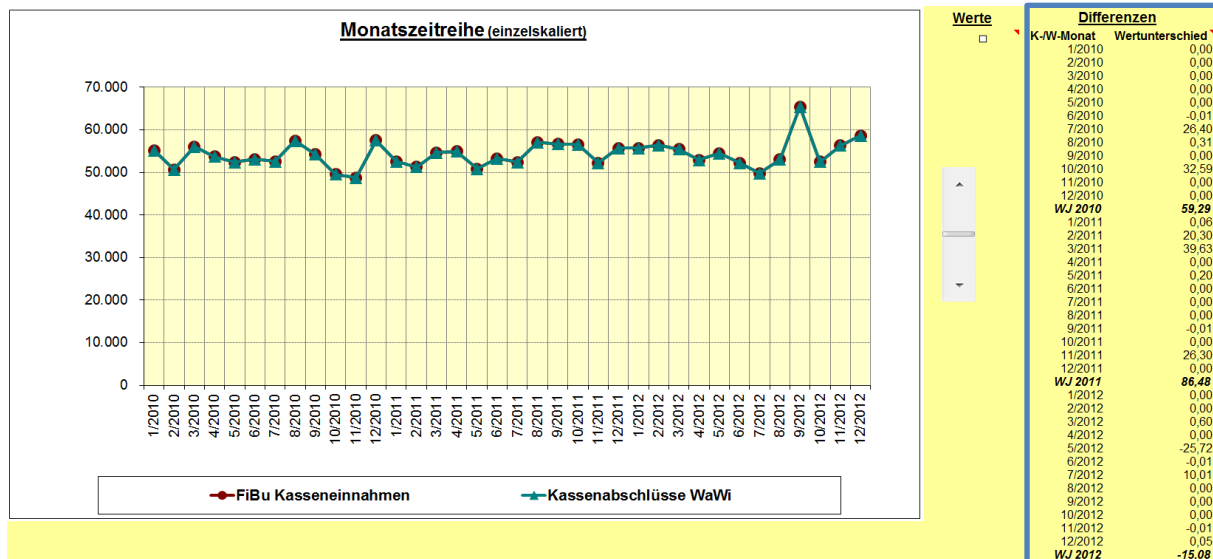
Die **Kassenfunktion des Systems** versteht die einzelnen Verkaufsvorgänge bei der Erfassung mit allen Merkmalen zur protokollierenden Aufsummierung über den einzelnen Beleg bis zum täglichen Kassenabschluss. Dadurch werden die Abgrenzungen für die Umsatzbesteuerung, für die unterschiedlichen Zahlungswege etc. vorgenommen. Parallel werden die Daten der Lagerhaltung bzw. Inventur aktualisiert. Für die wertmäßige Inventursicht müssen zudem Bewertungen der Einzelpositionen über die Stammdaten oder manuell ergänzt werden. Alle abgebildeten Abläufe können durch Falscheingaben oder fehlerhaft programmierte Prozesse zu unrichtigen Auswertungsergebnissen führen. Ziel der Schnittstellenprüfung ist es, in möglichst effizienter Weise Fehler der Verwaltungsprozesse (weitest gehend) auszuschließen oder aufzudecken und einzugrenzen.

Da i.R. von Schnittstellenprüfungen üblicherweise (betriebliche) Fortlaufdaten mit einer chronologischen Abfolge geprüft werden sollen, bietet sich die Arbeitsweise mit **visualisierenden Zeitreihenanalysen** an. Im Gegensatz zu Schlüssigkeitsprüfungen, bei denen die

⁹⁶⁵ Art. 97 § 38 EGAO mit der Systemprüfung von Steuerkontrollsystemen und der Möglichkeit, Beschränkungen der nächsten Außenprüfung zuzusagen.

Abhängigkeit zweier Datenreihen untersucht wird, ist hierbei allerdings eine (nahezu) exakte Übereinstimmung zu erwarten. Deshalb müssen sich beide Datenreihen in der Zeitreihengrafik an derselben Y-Achse orientieren, weil jede automatische Größenanpassung einer Doppelskalierung das Ergebnis verzerren würde (s.a. *den Vergleich der Zeitreihenvarianten im Abschnitt „Methodenkenntnis“ unter Pkt. 4.2.*). Besteht die Gefahr zeitlich abweichender Erfassungen einer Datenreihe gegenüber der anderen wie z.B. bei später (aufsummiert) erklärten Meldedaten im Verhältnis zur Buchführung, sollte nach der SRP-Zeitreihensystematik „vom groben Überblick zu den Detailperspektiven“ vorgegangen werden (s. *Pkt. 4.2.2.*), um echte Differenzen von Versätzen unterscheiden zu können.

In diesem Beispiel eines Einzelhandelsbetriebs soll die Übereinstimmung der im Warenwirtschaftssystem erfassten Verkaufsvorgänge mit der Buchführung verifiziert werden:

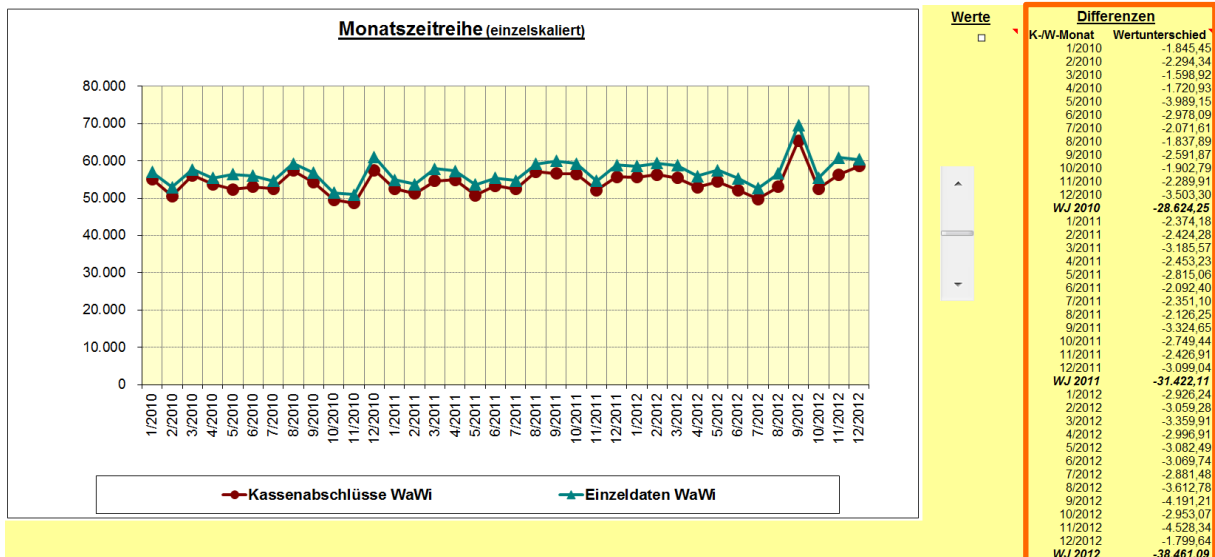


Dazu werden zunächst die täglichen Kassenabschlüsse des Warenwirtschaftssystems den Buchführungsdaten mit dem einzelskalierten Zeitreihenvergleich vom Tabellenblatt „Monatszeitreihen“ in der SRP-Vorlage BWA (s. *Pkt. 4.2.5.2.*) gegenübergestellt, weil es keine Hinweise auf zeitliche Verschiebungen über die Monatsgrenzen hinaus gibt. Als Ergebnis stimmen die in Papierform übergebenen Tageskassenabschlüsse mit den im Steuerbüro gebuchten Zahlen weitest gehend überein, wie auch die *tabellarische Ausgabe mit geringfügigen Monatsdifferenzen* bestätigt. Insoweit lässt sich zügig feststellen, dass die Schnittstelle „Warenwirtschaftssystem des Betriebs ⇔ Buchführung im Steuerbüro“ ordnungsgemäß ist.

Im zweiten Schritt werden die programminternen Abläufe des Warenwirtschaftssystems auf Schlüssigkeit überprüft. Hierzu werden die knapp 300.000 einzelnen Vorgangsdaten gemäß fortgeschrittener Anwendung der SRP-Vorlagen zuerst auf dem Tabellenblatt „eigene Notizen“ abgelegt,⁹⁶⁶ tageweise per Pivot-Tabelle aufsummiert und anschließend auf das

⁹⁶⁶ Der Vorteil der verknüpfenden Arbeitsweise über das Tabellenblatt „eigene Notizen“ liegt in der Möglichkeit, die Zeitreihenanalysen nach weiteren Unterscheidungskriterien der Daten einzugrenzen (s. *den Abschnitt*

Tabellenblatt „Vergleichsdaten“ verknüpft (s. den Abschnitt „fortgeschrittene Anwendung der SRP-Vorlagen“ unter [Pkt. 4.](#)). Aus der Gegenüberstellung mit den täglichen Kassenabschlussdaten des Warenwirtschaftssystems, die mit der Buchführung übereinstimmen, resultiert dieses verbildlichte und tabellarische Ergebnis:



Während der Konsolidierung der einzelnen Verkaufsvorgänge zu Tagessummen kommt es im Warenwirtschaftsprogramm zu einer **Wertminderung**. Diese **erstreckt sich ziemlich regelmäßig über den gesamten Prüfungszeitraum** und beschreibt dadurch die Systembesonderheit bzw. den -fehler. Zur Abgrenzung der in Rechnung gestellten Verkäufe von den Tageskassenerlösen ergibt eine gezielte Vertiefung anhand des betreffenden Programmschlüssels, dass sie in der Fortlaufdatenbank der Einzelvorgänge mit dem Wert „0“ geführt werden und deshalb die Differenzen nicht ausgelöst haben können. Somit ist für einen „sachverständigen Dritten“⁹⁶⁷ kein Grund erkennbar, warum die erfassten Einzelverkäufe die Tagesabschlüsse übersteigen. Gerade systematische Manipulationen mit digitalen Hilfsmitteln [s. den Abschnitt „raffinierte (digitale) Manipulationswerkzeuge“ unter [Pkt. 2.1.](#)] können derartige Datenverfälschungen hervorrufen. Deshalb wirft das Ergebnis der (Schnittstellen-)Prüfung des Betriebsverwaltungssystems in diesem Fall den Verdacht auf, dass die Zahlen der Buchführung nicht die korrekten Besteuerungsgrundlagen wiedergeben. Die Aufklärungsverantwortung liegt auf Seiten der Stpfl., weil sie sich für diese Art der Aufzeichnung und Aufbewahrung sowie für das Warenwirtschaftsprogramm entschieden haben.⁹⁶⁸

„fortgeschrittene Anwendung der SRP-Vorlagen“ unter Pkt. 4.). Sofern die Datenanzahl an Einzelvorgängen eines (Erst-)Erfassungssystems den möglichen Zeilenumfang von MS Excel® übersteigt, muss die tageweise Aufsummierung als Vorbereitungsschritt in IDEA® durchgeführt werden. Die Tagessummen können dann direkt in das Tabellenblatt „Vergleichsdaten“ der SRP-Vorlage BWA kopiert werden.

⁹⁶⁷ § 145 Abs. 1 S. 1 AO.

⁹⁶⁸ Tipke/Kruse, AO/FGO, 171. Lieferung 7/2022, § 158 AO, Rz. 9: „Das Beweismaß [für eine Schätzung] hängt mithin davon ab, [...] wie weit die FinBeh. auf dessen Mitwirkung im konkreten Einzelfall angewiesen ist.“

Grundsätzlich ist eine Übereinstimmungsprüfung bspw. von Schnittstellen auch mit Hilfe von **musterbasierten Häufigkeitsanalysen** möglich. Bei dieser Herangehensweise überlagern sich im Diagramm zwei Häufigkeitsverteilungen wie z.B. in der Grafik zur „Altersstruktur Deutschlands nach Geschlecht 2011“ unter [Pkt. 5](#). Mögliche Unterschiede der beiden gegenübergestellten Datenbestände lassen sich – in Abhängigkeit von der Unterteilbarkeit des Klassifizierungsmerkmals (s. den Abschnitt „Auszählung von Informationen nach strukturbaren Eigenschaften“ unter [Pkt. 4.3](#)) – entsprechend ablesen: für die Altersstruktur bspw. die Erhöhung des Frauenanteils durch die Gefallenen im 2. Weltkrieg. Insbesondere wenn viele ähnliche Daten vorliegen wie bei Erlösen oder Kosten, führt jedoch die (zusätzliche) Einbindung der chronologischen Information mittels verbildlichender Zeitreihenanalytik i.d.R. schneller zu konkreten Übereinstimmungsergebnissen.

5.1.3. Visualisierte Schlüssigkeit („sprechende Diagramme“)

Neben der Verprobung von (Bar-)Einnahmen auf Vollständigkeit hat die Schlüssigkeitsprüfung i.R. der Besteuerung besonders für den **Fremdvergleich** wie bspw. bei der konzerninternen Verrechnung eine überragende Bedeutung: Für eine möglichst sachgerechte Beurteilung, ob Vereinbarungen zwischen nahestehenden Personen fremdüblich und somit steuerlich anzuerkennen sind, müssen „alle für den Einzelfall bedeutsamen [...] Umstände [...] berücksichtig[t]“⁹⁶⁹ und im Zusammenhang bewertet werden. Dabei ist zum einen zu beachten, dass herkömmliche⁹⁷⁰ Verrechnungspreisdokumentationen häufig eine (subjektive) Zusammenfassung der realen Vorgänge darstellen, die ohne Einbeziehung der dazugehörigen Einzeldaten i.d.R. nicht seriös beurteilt werden können. Zum anderen sind die Verhältnisse häufig so komplex (gestaltet), dass sie alleine auf Basis von Texten und Tabellen nicht ausreichend erfasst werden können. Deshalb wird in der Literatur schon seit einigen Jahren immer häufiger festgestellt, dass im Bereich „**Verrechnungspreise**“ digitale Planungs- und Kontrolllösungen unverzichtbar sind.⁹⁷¹

⁹⁶⁹ § 88 Abs. 1 S. 2 AO.

⁹⁷⁰ Neue Formen der datenbasierten Verrechnungspreisdokumentation erobern zunehmend den Markt; vgl. z.B.: https://universalunits.com/en/products/transfer-pricing?gclid=EAlalQobChMI-L-J1sSX9AIVbY9oCR20ZwaEEAAYASAAEgK5s_D_BwE.

⁹⁷¹ Vgl. z.B.: Braun/Köppe, „Digitalisierung in der Verrechnungspreisfunktion – Entwicklungen und Chancen“, DB 8/2019, 394, Schöneborn, „Verrechnungspreismanagement 4.0 – Zum Stand der Digitalisierung in der Praxis und welche Kompetenzen für ein erfolgreiches IT-Projekt entscheidend sind“, DB 14/2019, 737, Krüger, „Chemie 4.0: Digitalisierung, Verrechnungspreisbestimmung und Betriebsprüfung“, DStZ 17/2019, 659, Schindler/Totzek, „Anwendungsmöglichkeiten von Data & Analytics im Verrechnungspreiskontext“, IWB 20/2019, 827 etc.

Die von der OECD vorgeschlagenen Maßnahmen zur Bekämpfung illegaler Steuervermeidung multinationaler Konzerne „BEPS“⁹⁷² mit den erweiterten Meldepflichten für grenzüberschreitende Gestaltungen⁹⁷³ zielen auf eine transparentere Offenlegung aller Grundlagen zur internen Verrechnung ab. Dabei entstehen zwangsläufig noch mehr Daten zu internen Vorgängen oder externen Vergleichsmaßstäben, welche bei der betriebswirtschaftlichen Beurteilung der Fremdüblichkeit einbezogen werden sollen; so ist den neuen „Verwaltungsgrundsätze[n] Verrechnungspreise“ an diversen Stellen die Ausrichtung auf datenbasiert-transaktionale Prüfung zu entnehmen.⁹⁷⁴ Wegen der meist komplexen Verhältnisse z.B. in Form von umfangreichen Vorgängen bzw. Einflussfaktoren ist die wahrnehmungsphysiologische Überlegenheit flexibler Datenvisualisierungen entscheidend [s. die Abschnitte „Visualisierung von Besteuerungsdaten“ unter [Pkt. 2.2](#) und „überlegene Leistungsfähigkeit einer (systematisch-mehrperspektivischen) Datenvisualisierung“ unter [Pkt. 5](#)].

Insbesondere in Fällen, in welchen direkte Vergleiche „interne versus externe Preise“ nicht bzw. nur unter schwierig quantifizierbaren Anpassungen möglich sind, kann die Verbildlichung der chronologischen Verrechnungsverhältnisse mit Hilfe **systematischer Zeitreihenanalytik** einen guten Überblick über die Schlüssigkeit vermitteln, weil (steuerlich) gestaltete Verrechnungspreise bei Veränderungen der Preisbildungsbedingungen oft nicht sorgfältig genug nachjustiert werden.

Als Beispiel für zeitreihenbasierte Schlüssigkeitsprüfungen dient hier die interne Verrechnung an ein deutsches Fertigungsunternehmen durch die Schwestergesellschaft in Singapur, die Millionen von einfachen Kunststoffkleinteilen herstellt und zur weltweiten Distribution nach Deutschland liefert; demgegenüber erfolgt die Herstellung aufwendigerer Produkte in europäischen Fabriken. Innerhalb der Lieferkette „Produktion in Singapur ⇒ Lieferung nach Deutschland ⇒ Vermarktung weltweit“ ist die (beträchtliche) Roh-Marge besonders zwischen Singapur und Deutschland sachgerecht aufzuteilen, denn wegen der sehr unterschiedlichen Unternehmensbesteuerung beider Länder ergibt sich für die Eigentümer*innen ein klares Motiv zur Gewinnverlagerung.⁹⁷⁵

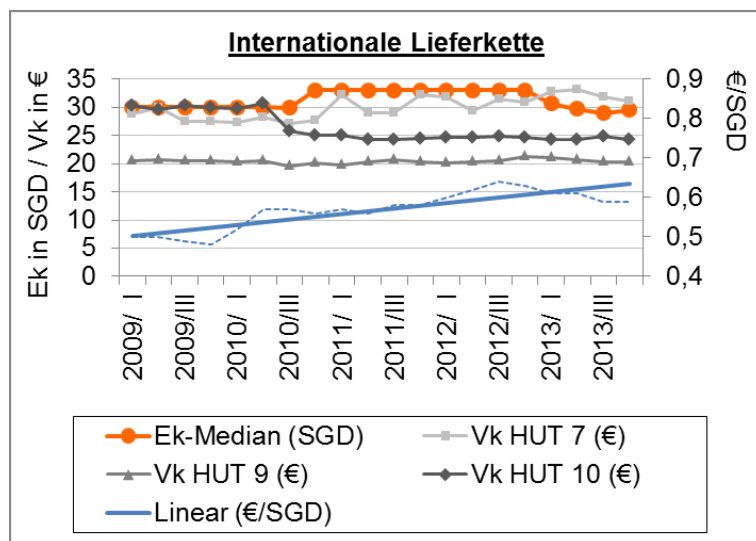
⁹⁷² Internetseite der OECD „BEPS 2015 Final Reports: Final BEPS package for reform of the international tax system to tackle tax avoidance (BEPS = Base Erosion and Profit Shifting)“, nachzulesen unter: <http://www.oecd.org/ctp/beps-2015-final-reports.htm>.

⁹⁷³ Zur Umsetzung der EU-Richtlinie 2018/822 des Rats bezüglich des verpflichtenden automatischen Informationsaustauschs im Bereich der Besteuerung über meldepflichtige grenzüberschreitende Gestaltungen vom 25. Mai 2018 (ABl. L 139/1 vom 5. Juni 2018) – sog. DAC 6 – s. z.B. unter: https://www.bundesfinanzministerium.de/Content/DE/Downloads/BMF_Schreiben/Weitere_Steuerthemen/Abgabenordnung/2021-03-29-Anwendung-Vorschriften-Pflicht-Mitteilung-grenzueberschreitende-Steuer-gestaltungen.html.

⁹⁷⁴ BMF-Schreiben v. 6.6.2023, BStBl. I 2023, 1093.

⁹⁷⁵ Neue Zürcher Zeitung online v. 14.2.2017: „Singapur geht den pragmatischen Weg: Im Stadtstaat erfolgt eine graduelle Anpassung der Unternehmensbesteuerung an international geltende Standards. Man will kein Paria

Deshalb sollte eine Verbildlichung der chronologischen Preisentwicklungen für die relevantesten Produkte inklusive des wichtigen Einflussfaktors „Währungskurs“ möglichst viel



Klarheit über die Abläufe herstellen. In diesem auf Übersichtlichkeit ausgelegten Diagramm⁹⁷⁶ zu den internen Verrechnungsverhältnissen (aus Sicht des deutschen Unternehmens) wurden die vielen **Einkäufe aller zehn Hauptumsatzträger (HUT) als repräsentativer Median der durchschnittlichen Quartalspreise**⁹⁷⁷ zusammengefasst, weil die vor-

ab durchgeführte Untersuchung der internen Einkaufspreise für alle Produkte eine gleichartige Entwicklung auf unterschiedlichem Preislevel offenbart hat. Demgegenüber weisen die Verkaufspreisentwicklungen der exemplarischen Hauptumsatzträger „7“, „9“ und „10“ unterschiedliche Tendenzen auf. Somit kommt der (jeweilige) Absatzpreis der Produkte als Festlegungsgrundlage für die internen Einkaufspreise *nicht* in Betracht; i.V.m. dieser Erkenntnis verwundert die über zwei Jahre anhaltende ca. 10 %-ige Erhöhung der internen Verrechnung besonders. Als zusätzliche Information erklärt die **permanente Währungsverteuerung**⁹⁷⁸ diese beobachtete Unschlüssigkeit nicht; vielmehr verstärkt sie die Erhöhung aller Einkaufspreise indirekt noch. Unter fremden, unabhängigen Geschäftspartner*innen wäre zu vermuten, dass das deutsche Unternehmen seine Verhandlungsposition „große Abnahmemengen“ dazu genutzt hätte, um den schwächer werdenden Euro über einen reduzierten Einkaufspreis zumindest tlw. ausgleichen zu können. Die aus den unschlüssigen Verrechnungsverhältnissen resultierende Gewinnminderung für den deutschen Konzernteil kann gem. § 158 **Abs. 2 Nr. 1** AO nur auf der Grundlage einer *nachprüfbaren* Begründung anhand weiterer Konzern-

sein“; nachzulesen unter: <https://www.nzz.ch/wirtschaft/unternehmensbesteuerung-singapur-geht-den-pragmatischen-weg-ld.145555>. Im alten Erlass zur „internationale[n] Einkunftsabgrenzung nach dem Maßstab des Fremdvergleichs“ v. 23.2.1983, BStBl. I 1983, 218, wurde die besondere Situation gegenüber niedrig besteuerten Ländern explizit aufgegriffen: „1.5. Geschäftsbeziehungen zu Nahestehenden in niedrig besteuerten Gebieten“.

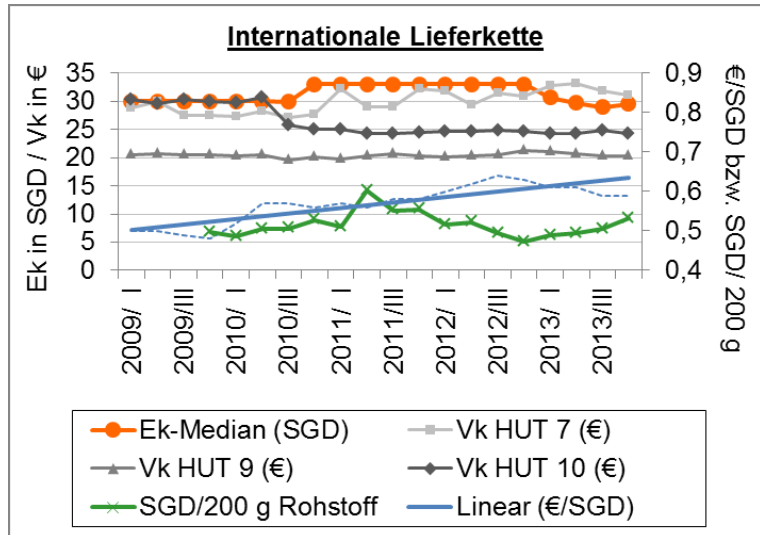
⁹⁷⁶ Bildquelle (überarbeitet): Wolenski/Wähnert, „Systematische Visualisierung von internationalen Gewinnverschiebungen“, IWB 3/2015, 105, 110.

⁹⁷⁷ Die Einkaufspreise werden in der Grafik mit der Hauswährung „SGD“ verarbeitet und orientieren sich an der linken Y-Achse.

⁹⁷⁸ Über eine Skalierungsverschiebung der sekundären Y-Achse für die Währungskursentwicklung wird die Darstellung als Zwei-Ebenen-Diagramm (unterschiedliche Informationen in getrennten Diagrammbereichen) erreicht, damit alle chronologischen Größenentwicklungen gut im Zusammenhang zu erkennen sind.

bzw. Marktfaktoren oder einer besonderen Unternehmensstrategie steuerlich anerkannt werden.⁹⁷⁹

Entsprechend führte die Wirtschaftsprüfungsgesellschaft als Grund für die kommunizierte Unschlüssigkeit einen „permanenten Preisanstieg des Hauptrohstoffs“ an. Da prüfbare Nachweise zu dieser Begründung z.B. in Form auszugsweiser Rechnungen der Schwestergesellschaft aus Singapur fehlten, sollte eine Erweiterung der Sachverhaltsverbildlichung mit zugänglichen Mitteln das Argument hinterfragen. Dazu wurden Roh-



stoffkurse des betreffenden Kunststoffes von einer Internetseite heruntergeladen und in das Diagramm einbezogen.⁹⁸⁰ Obwohl für die ersten drei Quartale keine Daten vorliegen, erweist sich der Einwand als nicht tragfähig: Eine relevante⁹⁸¹ Verteuerung des Rohstoffs kann erst im II. Quartal 2011 und somit ein halbes Jahr nach der vertraglichen Erhöhung aller Einkaufspreise festgestellt werden; außerdem fiel der Rohstoffpreis innerhalb eines Jahres wieder auf sein altes Niveau zurück, während die Einkaufspreiserhöhung noch deutlich länger Bestand hatte. Infolge dieser effektiven und effizienten Aufbereitung einer prüfungsbedürftigen Vereinbarung zwischen nahestehenden Personen kehrt sich die Beweislast um: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören [...] Die Vorlage verwertbarer Aufzeichnungen schließt Einkünfteberichtigungen nicht aus. [...] Verwertbare Auf-

⁹⁷⁹ BMF-Schreiben v. 6.6.2023, „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093, Rz. 3.19 mit einer Aufzählung wesentlicher „Faktoren [...], die sich auf die Preisgestaltung auswirken könnten.“

⁹⁸⁰ Die offiziellen Rohstoffpreise wurden auf die Bezugsmenge „200 g“ umgerechnet, damit die Preisentwicklung über die Sekundärachse gut nachvollziehbar im Diagramm platziert wird.

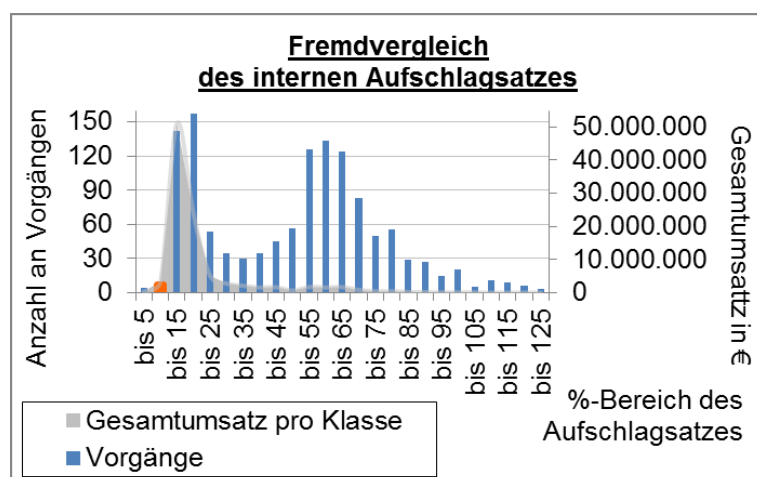
⁹⁸¹ Die Anpassung der nachträglich eingebundenen Rohstoffpreisentwicklung an die sekundäre Y-Achse für die Währungskursentwicklung und an die Zwei-Ebenen-Aufteilung des Diagramms führt zu einer gewissen Verzerrung der Wiedergabe, wohingegen relativ stärkere Veränderungen trotzdem erkennbar bleiben.

zeichnungen sind lediglich Ausgangspunkt für die Prüfung. Die Vorlage verwertbarer Aufzeichnungen schließt daher auch die Schätzungsbefugnis der Finanzbehörde gemäß § 162 Abs. 1 und 2 AO nicht aus. [...] Voraussetzung einer Einkünfteberichtigung ist vielmehr, dass die vom Steuerpflichtigen angewandten Verrechnungspreise mit hoher Wahrscheinlichkeit nicht dem Fremdvergleich entsprechen und dass der von der Finanzbehörde ermittelte Verrechnungspreis zumindest wahrscheinlicher ist.⁹⁸²

Auch in Schlüssigkeitsfragen können merkmalsbezogene Häufigkeitsauswertungen als andere Betrachtungsperspektive neben der Chronologie einen guten Überblick für sachgerechte Schlussfolgerungen geben. Deshalb kann die **visualisierende Musteranalytik** (zusätzlich) die Klarheit komplexer bzw. datenintensiver Sachverhalte insbesondere in den folgenden Konstellationen unterstützen:

- ✓ Eine Vielzahl von uneinheitlichen Vergleichsvorgängen soll übersichtlich strukturiert werden (s. das nachfolgende Beispiel).
- ✓ Mehrere uneinheitliche Prüfungssachverhalte sollen zur Beurteilung an unklaren Maßstäben systematisch überblickt werden (s. das nachfolgende Beispiel).
- ✓ Für einen in Klassen unterteilbaren (s. den Abschnitt „Auszählung von Informationen nach strukturierbaren Eigenschaften“ unter [Pkt. 4.3.](#)) Sachverhalt bestehen Teilerwartungen, sodass aufgrund der (Nicht-)Erfüllung Schlüssigkeitsbeurteilungen vorgenommen werden können, wie bspw. bei der Verpflichtung zur gleichmäßigen Einzelbewertung von Forderungen unabhängig vom Ausgangswert⁹⁸³ (s. den Abschnitt „komplizierte Einzelprüfungsfragen“ unter [Pkt. 3.1.](#)).

In einem Beispielsfall für musterbasierte Schlüssigkeitsprüfungen soll die Fremdüblichkeit



der Verrechnung eines deutschen Unternehmens an ein verbundenes in Spanien auf der Grundlage mehrerer tausend Datensätze zum Rechnungsausgang beurteilt werden. Die kalkulatorische Gesamtgegenüberstellung der Informationen „Umsatz“ und „Kosten“ ergibt eine erhebliche Margendifferenz

⁹⁸² BMF-Schreiben v. 3.12.2020 „Verwaltungsgrundsätze 2020“, BStBl. I 2020, 1325, Rz. 68 u. 73.

⁹⁸³ Vgl. Wolenski/Weiskopf, „Praxisfall: Einsatz indirekter Prüfungsmethoden in der Großbetriebsprüfung am Beispiel der Einzelwertberichtigung“, BBK 3/2013, 128.

mit 8,1 % intern zu 22,3 % extern. Durch diese klassifizierte Häufigkeitsauswertung⁹⁸⁴ der Aufschlagsätze nach Vorgängen wird jedoch die Uneinheitlichkeit der Verhältnisse ersichtlich. Zu den zwei Konzentrationsbereichen „> 10 % bis 20 %“ sowie „> 50 % bis 65 %“ muss offensichtlich eine Differenzierung vorgenommen werden; über die ergänzende Hintergrundinformation „Gesamtumsatz pro Aufschlagsatzklasse“ wird dazu deutlich, dass die höheren Margen zwar viele Kunden mit allerdings geringen Rechnungsbeträgen betrifft. Die daraufhin eingeforderte Erklärung ergibt, dass es sich bei den Sachverhalten mit niedrigerer Marge um Neuwaren und bei denen mit höherer um Ersatzteile handelt, und ermöglicht einen konkreten sachgerechteren Fremdvergleich: Mit Hilfe von Filtern kann die Datenanalyse zügig auf passende Vergleichsprodukte eingeschränkt werden.

Da (Zeit-)Reihen- und Musterverbildlichungen unterschiedliche Aspekte aus (Betriebs-)Daten herausarbeiten, die sich in der Aufklärung zu beurteilender Sachverhalte i.d.R. gut ergänzen, erzielt deren Verknüpfung fast immer eine bessere Entscheidungsgrundlage und dadurch sachgerechtere Schlussfolgerungen. Im Gegensatz zur schrittweisen Suche nach betriebs-spezifischen Prüffeldern (s. die Abschnitte „*Prüfungsnetz*“ unter [Pkt. 2.3.](#) und „*systematische Kombination aus unterschiedlichen Datenverbildlichungen*“ unter [Pkt. 5.1.1.](#)) ist es aber für das vollständige Erfassen komplizierter Schlüssigkeitsbeurteilungen häufig von Vorteil, die beiden Ansichten z.B. in einem **mehrperspektivischen (Komplex-)Diagramm** zu vernetzen, damit die Erkenntnisse im Zusammenhang verstanden werden.

Im nachfolgenden Beispiel soll ein möglichst aussagekräftiger Überblick zum Einstieg in die Verrechnungspreisprüfung bei einem deutschen Unternehmen mit einer ausländischen Tochtergesellschaft mittels Preisvergleichsmethode erlangt werden, um bei Bedarf (punktuell) tiefer einsteigen zu können. Dazu wird die Reihenlogik der X-Achse eingesetzt, um die mögliche steuerliche Bedeutung für den Prüfungszeitraum anhand des absteigend sortierten IC⁹⁸⁵-Volumens der Hintergrundfläche ablesen zu können; für weniger Information und somit einen besseren Überblick im Diagramm kann die zu Grunde liegende Tabelle per Filter auf ein Mindest-IC-Volumen beschränkt werden. Der produktbezogene Preisvergleich erfolgt an dem gem. § 1 Abs. 3a S. 3 AStG⁹⁸⁶ eingeeengten Box-Plot „1. Quartil – Median – 3. Quartil“⁹⁸⁷ für die externen Verkaufspreise, damit nicht bspw. eine außergewöhnliche kostenlose Pro-

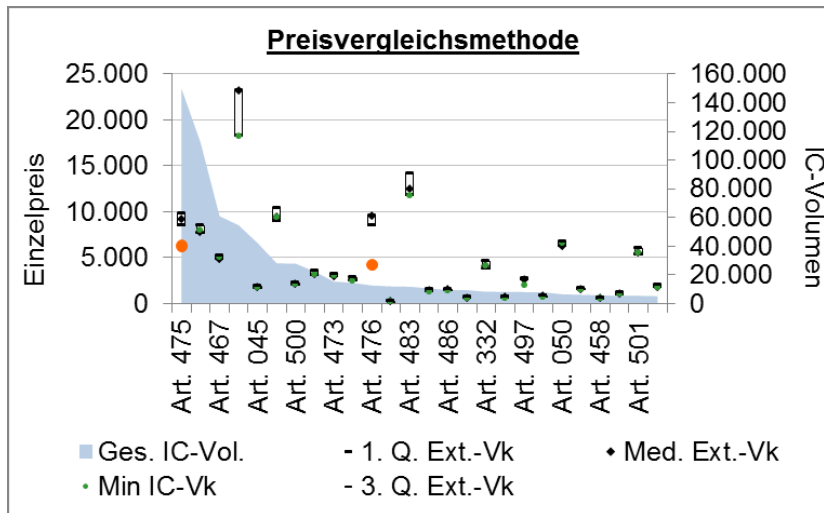
⁹⁸⁴ Bildquelle (überarbeitet): Werner/Wähnert, „Verbildlichende Prüfung am Beispiel der Gewinnverlagerung – Die Überlegenheit der Visualisierung von Datenregeln und -verteilungen gegenüber Kennzahlen“, IStR 19/2016, 811.

⁹⁸⁵ „IC = Inter Company“ für die Beziehung zum verbundenen Unternehmen.

⁹⁸⁶ Vgl. a.: BMF-Schreiben v. 6.6.2023, „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093, Rz. 3.30.

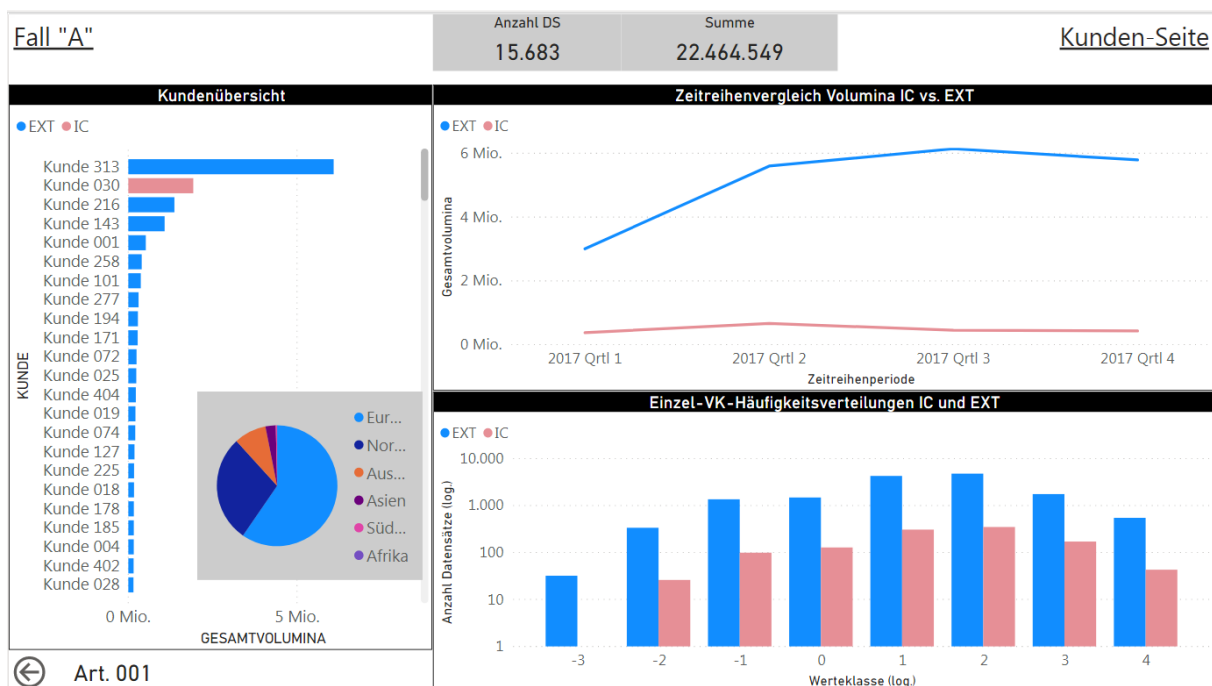
⁹⁸⁷ Der Box-Plot beschreibt somit den Wertumfang der externen Preise mit dem Rahmen vom 25 %- bis zum 75 %-Quantil (1. und 3. Quartil) und zeigt zusätzlich im Hinblick auf § 1 Abs. 3a S. 4 AStG den 50 %-Rang „Median“ als Raute innerhalb dessen an.

duktbeigabe als minimaler Produktpreis an Fremdkunden diese Gegenüberstellung verzerrt. Der niedrigste⁹⁸⁸ IC-Preis ist als grüner bzw. orangener Punkt dem betreffenden Artikel zu-



geordnet. Auf diese Weise lässt sich problemlos feststellen, dass **zwei relevante Produkte mindestens einmal intern zu einem deutlich niedrigeren Preis verrechnet** wurden, wovon besonders „Art. 475“ in der Prüfungsperiode ein nennenswertes Volumen erreicht hat. Durch eine Zeit-

reihenanalyse kann nun – auch wegen der notwendigen Übersicht für die betreffenden Produkte in Einzeldarstellung – systematisch konkretisiert werden, ob es sich dabei um einen Vorgang, eine Phase oder ein generelles Prozedere gehandelt hat, damit eine nachprüfbare Begründung i.R. der Fremdüblichkeitsprüfung konkret eingefordert werden kann.

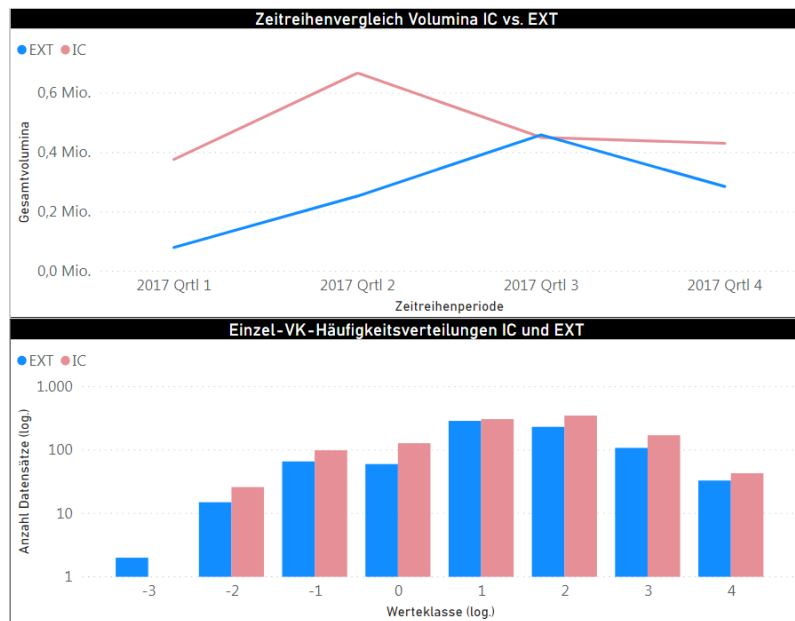


Eine andere technische Umsetzung für die übersichtliche Verarbeitung vieler entscheidungserheblicher Informationen bieten **B(usiness-)I(ntelligence)-Lösungen**, indem die sich ergänzenden ordnenden, visualisierenden und statistischen Auswertungen nicht in ei-

⁹⁸⁸ Im Gegensatz zum Vergleichs-Box-Plot wird intern der niedrigste Verkaufspreis genommen, weil bereits ein Einzelsachverhalt (bei entsprechendem Gesamtvolumen und fremdunüblicher Verrechnung) ein Prüffeld begründen kann.

nem Komplexdiagramm zusammengefasst, sondern als Datenanalysemodell wie in der MS Power BI®-Abb. oberhalb miteinander verknüpft sind. So beinhaltet die abgebildete Kunden-Seite⁹⁸⁹ zum vorherigen Beispiel – neben den statistischen Abstimmungsfeldern in der Kopfzeile – ein Kundenranking inklusive farblicher Trennung „intern versus extern“, eine Unterteilung nach Kontinenten sowie Gegenüberstellungen der Chronologie und Einzelpreishäufigkeiten, die alle gemeinsam auf (kombinierte) Filterungen reagieren.

Zur Auswahl sinnvoller Vergleichsregionen bzw. -kunden lassen sich über diese Seite umgehend die Verhältnisse des Falls feststellen: Auf dem Kontinent der Tochtergesellschaft gibt es keinen relevanten Vergleichskunden. Durch ausprobierendes Filtern ist aber zügig ein Unternehmen identifiziert, das in einer vergleichbaren Region ansässig ist und ein offenbar ähnliches Leistungsspektrum bezieht, denn die klassifizierten Häufigkeiten der Einzelpreise ähneln sich im Säulendiagramm rechts. Zusätzlich zeigt der Zeitreihenvergleich darüber mit abweichenden Abläufen an, dass die Geschäftszeitpunkte spätestens i.R. einer Vertiefung



fremdunüblicher Vorgänge hinsichtlich möglicher betrieblicher Begründungen⁹⁹⁰ beachtet werden sollten.

So hilfreich solche systematisch-mehrperspektivischen Analysemodelle in allen Phasen der datenbasierten Prüfung sind, muss bspw. beim Säulendiagramm zum Geschäftsbeziehungsvergleich aus methodischer Sicht verstanden werden, dass dieses nur in doppelt logarithmischer Ausführung zielführend ist, weil zum einen eventuell sehr unterschiedliche Vorgangszahlen gegenübergestellt werden sollen (Y-Achse) und zum anderen asymmetrisch-rechtsschiefe Häufigkeitsverteilungen (s. [Pkt. 4.3.4.1.](#)) erwartet werden, bei denen eine lineare Intervallbildung zur Konzentration in wenigen Einzelklassen führen würde (X-Achse). Gerade bei der Anwendung von Komplexdiagrammen oder vernetzten Modellen zur Schlüssigkeitsprüfung mit der möglichen Konsequenz der Schätzungspflicht gem. § 162 Abs. 2 S. 2 AO ist zu bedenken, dass dies zwar zu wesentlich eindeutigeren Entscheidungsgrund-

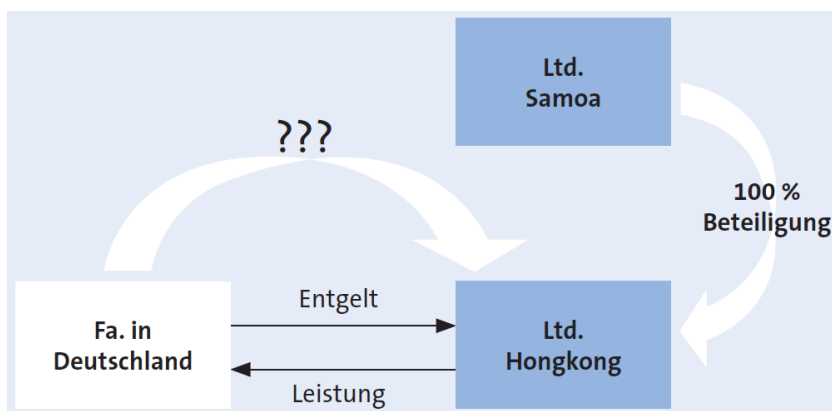
⁹⁸⁹ Werner-Tutorials zur Erstellung solcher Analysemodelle finden sich auf dem KTZ-Server (Intranet).

⁹⁹⁰ BMF-Schreiben v. 6.6.2023, „Verwaltungsgrundsätze Verrechnungspreise“, BStBl. I 2023, 1093, Rz. 3.19 mit einer Aufzählung wesentlicher „Faktoren [...], die sich auf die Preisgestaltung auswirken könnten.“

lagen führen kann, aber bereits für die Lesekompetenz von (Zwischen-)Ergebnissen und besonders für das zielführende Erarbeiten ein fundiertes methodisches und prozessuales Verständnis voraussetzt (s. [Pkt. 3.3.](#) und den Abschnitt „Prüfungsprozess“ unter [Pkt. 5.](#)).

Neben Verbesserungen bei der Sachverhaltsermittlung bewirkt eine systematisch visualisierende Schlüssigkeitsprüfung auch eine allgemeinverständlichere Grundlage für Planungs- und Prüfungsbesprechungen („**sprechende Diagramme**“). So hat sich inzwischen in vielen Prüfungen gezeigt, dass fokussierte Verbildlichungen der Daten, die einen auf Fremdüblichkeit zu beurteilenden Sachverhalt betreffen, zu Ergebnissen führen, welche weniger vom Verhandlungsgeschick der Parteien und dafür mehr von der Gesetzmäßigkeit der Abläufe geprägt sind.

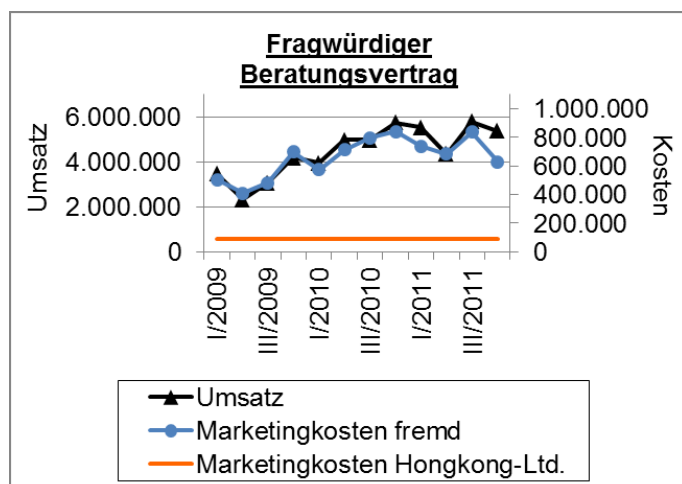
Als Beispiel für den Visualisierungsvorteil in Gesprächen über prüfungswürdige Gestal-



tungen dient dieses einfache Konstrukt: „Fraglich ist, ob Beteiligungen der Gesellschafter der deutschen D-GmbH an der Ltd. in Hongkong über eine Ltd. in Samoa indirekt gehalten werden. Samoa erteilt an deutsche Finanzbehörden

zumindest nur sehr schleppend Auskünfte, sodass der Standort bewusst gewählt sein könnte, um die Unternehmensverbindung der D-GmbH und der Hongkong Ltd. im Verborgenen zu lassen. Auf diese Weise können über einen Leistungsvertrag mit fingierten oder überteuerten Betriebsausgaben Gewinne aus Deutschland in einen Staat mit erheblich niedrigerer Unternehmen[s]steuerquote verschoben werden.“⁹⁹¹

Da zu den **pauschalen Marketingkosten der Ltd. aus Hongkong** keine nachprüfbaren Tätigkeitsnachweise vorgelegt worden sind, wird deren Entwicklung mit erwarteter Abhängigkeit vom Erfolg der D-GmbH im Diagramm



per Zeitreihe **vergleichbaren Beratungskosten** gegenübergestellt. Unter der groben Ver-

⁹⁹¹ Wolenski/Wähnert, „Systematische Visualisierung von internationalen Gewinnverschiebungen“, IWB 3/2015, 105, 106 mit Text u. Bild.

gleichsperiode „Quartal“ ist ein klarer Zusammenhang zwischen dem tendenziell zunehmenden Erfolg der D-GmbH als Beratungsadressat und den berechneten Fremdkosten zu erkennen; selbst die zwischenzeitlichen Umsatzenschwankungen werden durch Anpassungen der Marketingkosten beantwortet. Demgegenüber bleiben die Aufwendungen für die Ltd. in Hongkong unverändert. Das visualisierte Ergebnis bedeutet für den Fortgang der Prüfung, dass die Eigentumsverhältnisse der Ltd. aus Samoa unbedingt aufzuklären sind. Im „Beispielsfall hat dies [während einer Besprechung] zur Offenlegung [...] geführt, [wonach] die Gesellschafter der D-GmbH über die Ltd. in Samoa auch die Anteile an der Hongkong-Ltd. hielten und sich somit der Verdacht nahestehender Unternehmen bestätigte.“⁹⁹² Demnach hat die unmissverständliche Aufbereitung der Sachverhaltsdaten mindestens eine deutliche Beschleunigung der ausreichenden Aufklärung für eine sachgerechte steuerliche Beurteilung erzielt.

„D[ata]&A[nalytics]-Lösungen finden aber auch vielfach Einsatz [...] für die **Simulation** der Steuereffekte bei Diskussionen um fremdübliche Bandbreiten. Mittels BI-Lösungen können diese Effekte aus Diskussionen mit der Betriebsprüfung direkt und dynamisch simuliert werden, was die Entscheidungsfindung in der Diskussion möglicher Kompromissvorschläge deutlich verbessert und verkürzt.“⁹⁹³ Diese direkte Anpassbarkeit von Prüfungs- bzw. Verprobungsauswertungen an aktuell veränderte Informationsstände trotz umfangreicher Daten wird von den SRP-Vorlagen durch deren interaktive Optionen (s. [Pkt. 4.2.9.](#) sowie [4.3.2.2.3.](#), [4.3.2.3.3.](#), [4.3.3.2.](#), [4.3.3.3.](#) u. [4.3.4.5.](#)) vorkonfektioniert bereitgestellt, kann mit Hilfe einer ausreichenden technischen Unterstützung bei entsprechenden Programmbedienfähigkeiten jedoch auch manuell hergestellt werden, um gesetzeskonformere Schlussfolgerungen zu ermöglichen.

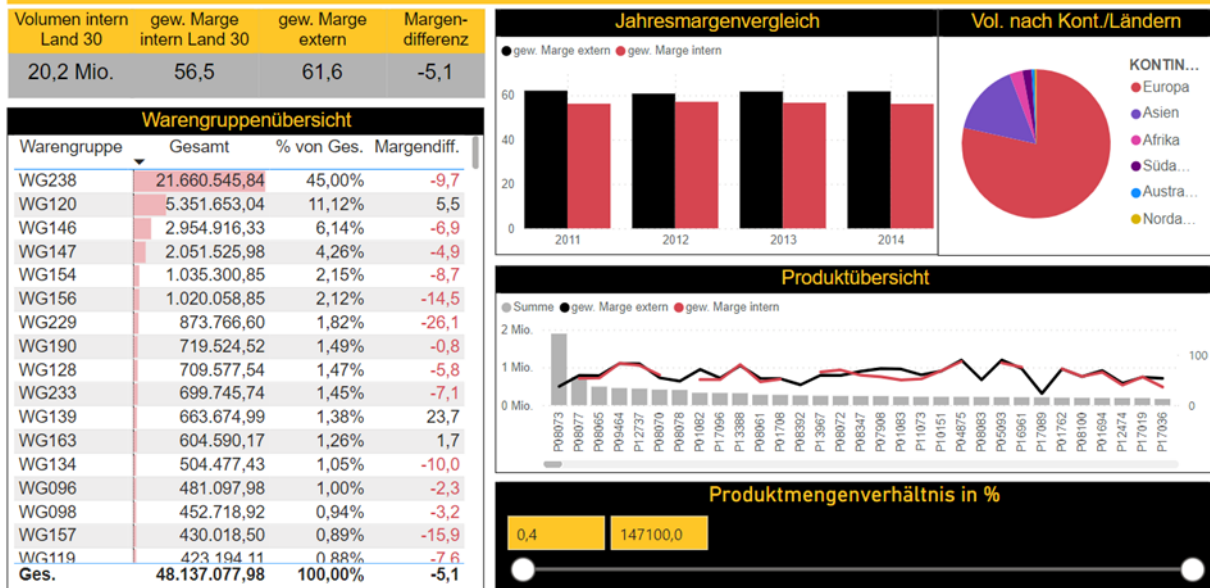
Zur beispielhaften Fremdüblichkeitsprüfung von Produktmargen eines deutschen Unternehmens gegenüber einem von vielen verbundenen Auslandsbetrieben soll durch das nachfolgend abgebildete MS Power BI®-Modell der Einfluss unterschiedlicher Faktoren auf die Erfolgsgröße simuliert werden; die dynamischen⁹⁹⁴ Kennzahlen reagieren dazu auf jede Filterung(-skombination) in den ordnenden und visualisierenden Auswertungen nach Warengruppen, Jahren, Regionen und Produkten sowie auf originäre Programm- und den angelegten Produktmengenfilter:

⁹⁹² Wolenski/Wähnert, „Systematische Visualisierung von internationalen Gewinnverschiebungen“, IWB 3/2015, 105, 109.

⁹⁹³ PwC-Newsletter v. 31.8.2022: „Aus unserer Praxis: Der Einsatz von Data & Analytics im Verrechnungspreismfeld – quo vadis?“, nachzulesen unter: <https://www.pwc.de/de/newsletter/transfer-pricing-perspectives-dach/ausgabe-55-2022/aus-unserer-praxis-der-einsatz-von-data-and-analytics-im-verrechnungspreismfeld-quo-vadis.html>

⁹⁹⁴ Durch die Filterzusammenhänge in BI-Programmen, aber auch über bedingte Formeln in Tabellenkalkulationsprogrammen können Berechnungen in Abhängigkeit von Vorbedingungen wie Filterungen ausgeführt werden; die SRP-Vorlagen nutzen diese Option auch für ihre Interaktivität.

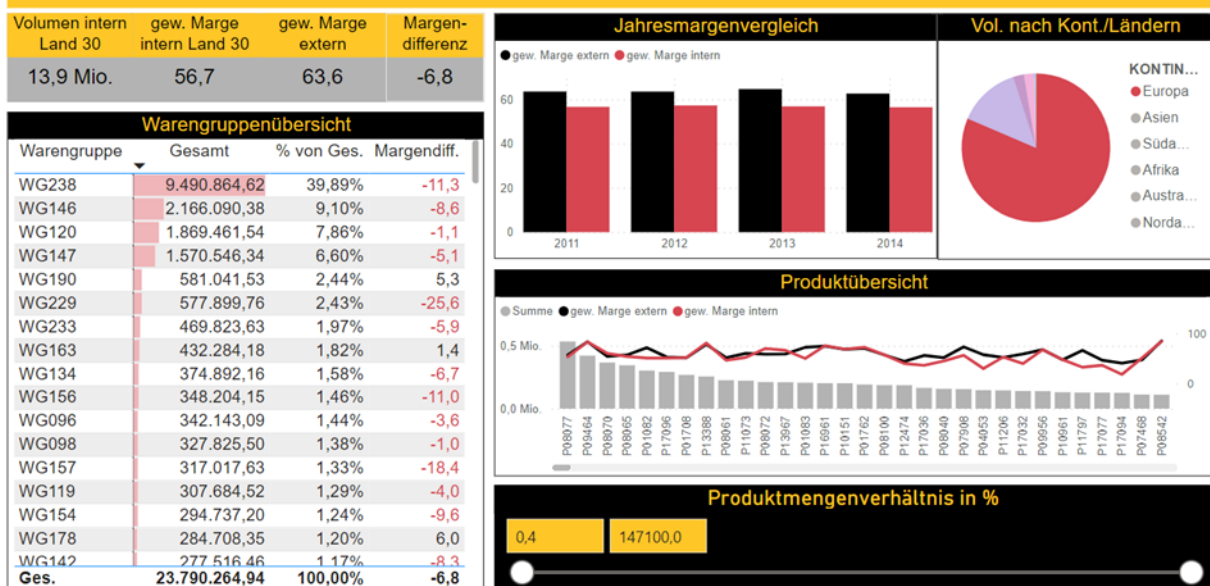
Margenvergleich "intern Land 30 vs. externe Kunden"



An den gewogen berechneten Werten oben links lässt sich insgesamt eine deutlich niedrigere Marge bei der Verrechnung mit dem verbundenen Unternehmen in „Land 30“ ablesen. Während die Verhältnisse im „Jahresmargenvergleich“ sehr konstant sind, unterscheiden sich die Warengruppen stark, was zu Saldierungseffekten führt. Im Diagramm „Produktübersicht“ ist zudem anhand der unterbrochenen Margen-Grafen ersichtlich, dass in den Margenvergleich offenbar Leistungen einbezogen werden, die nicht in beide Richtungen erbracht wurden und die Fremdüblichkeitsprüfung wesentlich verzerren können.

Um diesen Grundsatz der Vergleichbarkeitsanalyse zu erfüllen und andere mögliche Einflussfaktoren wie bspw. die Region darauf aufbauend zu simulieren, wird das Modell entsprechend nach vergleichbaren Produkten und dem Kontinent gefiltert:

Margenvergleich "intern Land 30 vs. externe Kunden"



Die Margendifferenz für danach verbleibende ca. zwei Drittel der unmittelbar vergleichbaren internen Verrechnungen vergrößert sich infolge dieser Konkretisierungen sogar. Dabei bleiben die deutlichen Unterschiede der Warengruppen erhalten, was weiterhin die Frage nach betrieblichen Begründungen für Saldierungen aufwirft; durch einen Warengruppenfilter könnte die Saldierung auf dieser Ebene problemlos unterbunden werden. Z.B. mit dem Datenschnitt „Produktmengenverhältnis in %“ unten rechts können weitere Simulationen dazu vorgenommen werden, ob zusätzliche bisher unberücksichtigte Faktoren für die Fremdüblichkeitsprüfung einbezogen werden sollten; genauso kann auf konkrete Mitwirkung⁹⁹⁵ umgehend reagiert werden.

Zusammenfassend lassen sich mit den heutigen technischen Mitteln unter entsprechender methodischer und Anwendungskompetenz (s. [Pkt. 3.3.](#) und den Abschnitt „Prüfungsprozess“ unter [Pkt. 5.](#)) auch komplexeste Prüfungs- bzw. Verprobungssituationen so weit aufklären, dass (rechtliche) Schlussfolgerungen eine klare Grundlage haben. Offenbar aus Gewohnheit heraus werden anstatt solcher transparenter Lösungen mitunter weiterhin komplex-statistische Methoden wie z.B. multiple Regressionsanalysen⁹⁹⁶ eingesetzt, die – aus der analogen Zeit stammend – wenig interaktiv sind und nur schwer nachvollziehbare Ergebnisse liefern.

5.2. Ausblick zur Prüfungsmethodik und -technik

Nachdem die **Textanalytik** ihre Leistungsfähigkeit und weitreichenden Möglichkeiten in anderen (Rechts-)Bereichen eindrucksvoll bewiesen hat⁹⁹⁷, werden zwischenzeitlich auch für

⁹⁹⁵ BMF-Schreiben v. 3.12.2020 „Verwaltungsgrundsätze 2020“, BStBl. I 2020, 1325, Rz. 68: „Grundsätzlich trägt die Finanzbehörde die Feststellungslast für diejenigen Tatsachen, die einen Steueranspruch begründen bzw. erhöhen. [...] Aus der gemeinsamen Verantwortung des Steuerpflichtigen einerseits und der Finanzbehörde andererseits für die vollständige Sachaufklärung folgt u.a., dass sich dann, wenn ein Steuerpflichtiger ihm obliegende allgemeine oder besondere Mitwirkungs-, Informations- oder Nachweispflichten verletzt, grundsätzlich die Ermittlungspflicht der Finanzbehörde entsprechend reduziert. [...] Die Verletzung der Mitwirkungspflichten kann insbesondere dann, wenn sie Tatsachen oder Beweismittel aus dem alleinigen Verantwortungsbereich des Steuerpflichtigen betrifft, dazu führen, dass aus seinem Verhalten für ihn nachteilige Schlüsse gezogen werden. Besondere Bedeutung kommt in diesem Zusammenhang dem Gedanken der Beweisnähe zu: Die Verantwortung des Steuerpflichtigen für die Aufklärung des Sachverhalts ist umso größer, je mehr Tatsachen oder Beweismittel der von ihm beherrschten Informations- und/oder Tätigkeitssphäre angehören (BFH-Urteil vom 15.2.1989, X R 16/86, BStBl II S. 462). Die Finanzbehörde hat dennoch auch in diesen Fällen den Sachverhalt aufzuklären, soweit ihr dies zumutbar ist.“

⁹⁹⁶ Vgl. dazu bspw. die Internetseite zum kostenlosen Programm „gretl“, nachzulesen unter: <http://gretl.sourceforge.net/>.

⁹⁹⁷ Vgl. z.B. zur Anwendung von Textanalytik auf die Vertragsprüfung: Haufe online v. 11.3.2018: „Legal Tech und Algorithmen zur Vertragsanalyse: [...] ‚LawGeex‘ hat gemeinsam mit Professoren der Stanford University, der University of Southern California und der Duke University School of Law ein Experiment durchgeführt, bei dem sich ein Algorithmus mit Top-Anwälten messen musste. 20 Anwälte erhielten die Aufgabe, 5 ‚Non-disclosure agreements‘ (Geheimhaltungsverträge) zu überprüfen und darin 30 versteckte rechtliche Probleme zu identifizieren. Die Anwälte benötigten im Schnitt mehr als eineinhalb Stunden und erkannten durchschnittlich 85 % der Rechtsprobleme. Der Algorithmus erreichte eine Genauigkeit bei der Problemerkennung von

einige Steuerrechtsanwendungen vollautomatisierte Lösungen angeboten: „Litax wurde bereits mit einem Algorithmus zur vollautomatischen Prüfung der grunderwerbsteuerlichen Konzernklausel ausgestattet[, ... der] Umwandlungen und Umstrukturierungen im Konzern prüft und ein begründetes Gutachten ausgibt.“⁹⁹⁸ Der damit verbundene Trend regelbasiert programmierbarer Gesetze („Legal Tech“) wird auch in Deutschland vor allem aus der Wirtschaft zunehmend gefördert⁹⁹⁹, weil frühzeitige Rechtssicherheit einen großen Wert hat. Parallel dazu werden Rechtsassistentensysteme entwickelt, die den Nutzer*innen zu einer Sachverhaltsbeschreibung wie bspw. einem Vertrag permanent und mit punktgenauem Zugriff alle aktuellen Entscheidungsgrundlagen von Gesetzen über Rechtsprechung und Verwaltungsanweisungen bis zu Literaturquellen bereitstellen.¹⁰⁰⁰ Dieser technische Fortschritt richtet sich in seiner Bedienfreundlichkeit an alle potentiellen Rechtsanwender*innen und bedeutet einen grundlegenden Umbruch der erforderlichen und gleichsam sinnvollen Qualifikationen: Wie schon in anderen Berufen¹⁰⁰¹ wird dadurch das aufwendige Erlernen von (rechtlichem) *Detailwissen* ineffizient, weil mit Hilfe eines strukturellen Überblicks und von technischen Bedienfähigkeiten sowie systematischer Vorgehensweise inklusive punktueller Kontrollen ein erheblich schnelleres und ergebnisrichtigeres Arbeiten möglich wird¹⁰⁰², wobei sich zudem

94 % und benötigte dafür exakt 26 Sekunden“, nachzulesen unter:

<https://www.haufe.de/recht/kanzleimanagement/legal-tech-steht-der-anwalts-roboter-schon-vor-der-tuer-222-445476.html>.

⁹⁹⁸ Internetseite der Fa. Litax®, nachlesbar unter: <http://litax.net/>. S.a.: Internetseite der Fa. Taxy.io®: „Taxy.io KI Beta Club – Nutzen Sie jetzt die Gelegenheit, [...] die neuesten KI-Produkte testen [zu] können“, nachzulesen unter: <https://www.taxy.io/ki-beta-club>, oder ein „ChatBot [...] rund um das Thema GoBD“ auf der Internetseite von TAXPUNK, nutzbar unter: <https://www.taxpunk.de/chatbot/gobd-plus/>.

⁹⁹⁹ Vgl. z.B.: Internetseite des Instituts für Digitalisierung im Steuerrecht e.V. (IDSt), Fachausschuss I: „Mission: Diesem Fachausschuss obliegt die Entwicklung einer auf alle Steuerarten anwendbaren Methode, die es ermöglicht, bereits bestehende und neue steuerliche Rechtsnormen möglichst weitgehend digital abzubilden.“ Nachlesbar unter: https://idst.tax/fachausschuss_eins/.

¹⁰⁰⁰ Vgl. z.B.: taxtech.blog v. 13.2.2019: „Steuer Gesetze in der semantischen Netzwerkanalyse – Taxy.io arbeitet an der semantischen Analyse von steuerrechtlichen Texten wie Gesetzen, Richtlinien und Urteilen, um ein maschinelles Verständnis dieser Texte aufzubauen und für individuelle Fragestellungen gezielte Lösungshinweise aus der Rechtslehre zu geben.“ Nachzulesen unter: <https://taxtech.blog/2019/02/13/steuer-gesetze-in-der-semantischen-netzwerkanalyse/>.

¹⁰⁰¹ So wird bspw. ein sehr umfangreiches Detailwissen über (Abgrenzungs-)Symptome in der Krankheitsdiagnostik bzw. über (Kontra-)Indikationen und Wechselwirkungen zur Medikation benötigt, um Patient*innen vor Fehlentscheidungen zu schützen. Da mittels ausgereifter Programme dieses Wissen inklusive der Zusammenhänge vollständiger und fehlerfreier als von den allermeisten Menschen verwaltet werden kann, sollten sie in der täglichen Medizinpraxis mindestens zur Absicherung eingesetzt werden. Diese Erkenntnis hat auch Auswirkungen auf die im Studium zu erlernenden Mindestqualifikationen.

¹⁰⁰² MIT News v. 14.7.2023: „Study finds ChatGPT boosts worker productivity for some writing tasks – A new report by MIT researchers highlights the potential of generative AI to help workers with certain writing assignments[:] Access to the assistive chatbot ChatGPT decreased the time it took workers to complete the tasks by 40 percent, and output quality rose by 18 percent“, nachzulesen unter: <https://news.mit.edu/2023/study-finds-chatgpt-boosts-worker-productivity-writing-0714>.

Einspareffekte in der Ausbildung ergeben, die für zeitgemäßere¹⁰⁰³ Wissensvermittlung verwendet werden können.

Demgegenüber schien der Hype um **vollautomatisierte Datenanalytik** offenbar seinen Zenit überschritten zu haben, indem die Funktionen von Technik und Mensch selbstbewusster abgegrenzt wurden: „Die Zukunft der Wirtschaftsprüfung ist digital: [...] Während Künstliche Intelligenz (KI) und Robotics Routinetätigkeiten übernehmen, konzentrieren sich die Prüfer [...] zunehmend auf die Interpretation der Daten.“¹⁰⁰⁴ Tatsächlich zeigen jüngere Einsatzbeispiele von generativer KI, dass zumindest ein verbreitetes Produkt inzwischen datenanalytische Funktionen inklusive Visualisierungen und Interpretationen bereitstellt.¹⁰⁰⁵ Auch werden neben Robotik¹⁰⁰⁶ z.B. Dokumentationen auf Basis hochkomplexer statistischer Modelle wie multipler Regressionsanalysen¹⁰⁰⁷ sowie Dashboards¹⁰⁰⁸ mit Zusammenstellungen diverser Datenvisualisierungen auf Knopfdruck immer häufiger zur Anwendung kommen, weil Verwaltungssysteme mit wenig Personal- und Einarbeitungsbedarf Anklang finden. In logischer Konsequenz zu Berichten über hochgradig automatisierte Steuerprüfungen im Ausland¹⁰⁰⁹ wird zudem auch in (bzw. für) Deutschland verstärkt an Maßnahmen wie RMS¹⁰¹⁰, RMS-BP¹⁰¹¹, Textanalytik und Anomalieerkennung¹⁰¹², zeitgemäßes Umsatzsteuersystem¹⁰¹³

¹⁰⁰³ SpringerProfessional online v. 10.1.2019: „Diese Future Skills benötigt Deutschland“, nachzulesen unter: <https://www.springerprofessional.de/aus--und-weiterbildung/personalentwicklung/diese-future-skills-benoetigt-deutschland/16278042>.

¹⁰⁰⁴ DB online v. 30.4.2019 (Sponsored News v. EY), früher nachzulesen unter: <https://www.der-betrieb.de/meldungen/die-zukunft-der-wirtschaftspruefung-ist-digital/>.

¹⁰⁰⁵ Vgl. z.B.: Webinarreihe „TAXPUNK meets ChatGPT [...] 4. BEYOND TEXTFILES [...] mit] Stefan Werner“, in YouTube eingestellt unter: https://www.youtube.com/watch?v=xT_L2wPXxac.

¹⁰⁰⁶ *Kompenhans/Wermelt*, „Big Data“ – Potenziale für innovative Abschlussprüfungen“, BB 2018, 299: „Die an Systemschnittstellen erforderlichen manuellen Tätigkeiten werden immer mehr durch den Einsatz von Robotic Process Automation ersetzt, Entscheidungsprozesse werden dabei durch den Einsatz programmierter, selbstlernender Entscheidungsheuristiken, die auf Basis künstlicher Intelligenz arbeiten, unterstützt.“

¹⁰⁰⁷ Vgl. dazu bspw. die Internetseite zum kostenlosen Programm „gretl“, nachzulesen unter: <http://gretl.sourceforge.net/>.

¹⁰⁰⁸ Nachzulesen z.B. unter: https://docu.caseware.net/Audit/20.00/AuditTemplate/de/Content/User_Client_File_Setup/t_Using_the_Work_Flow_Panel.htm.

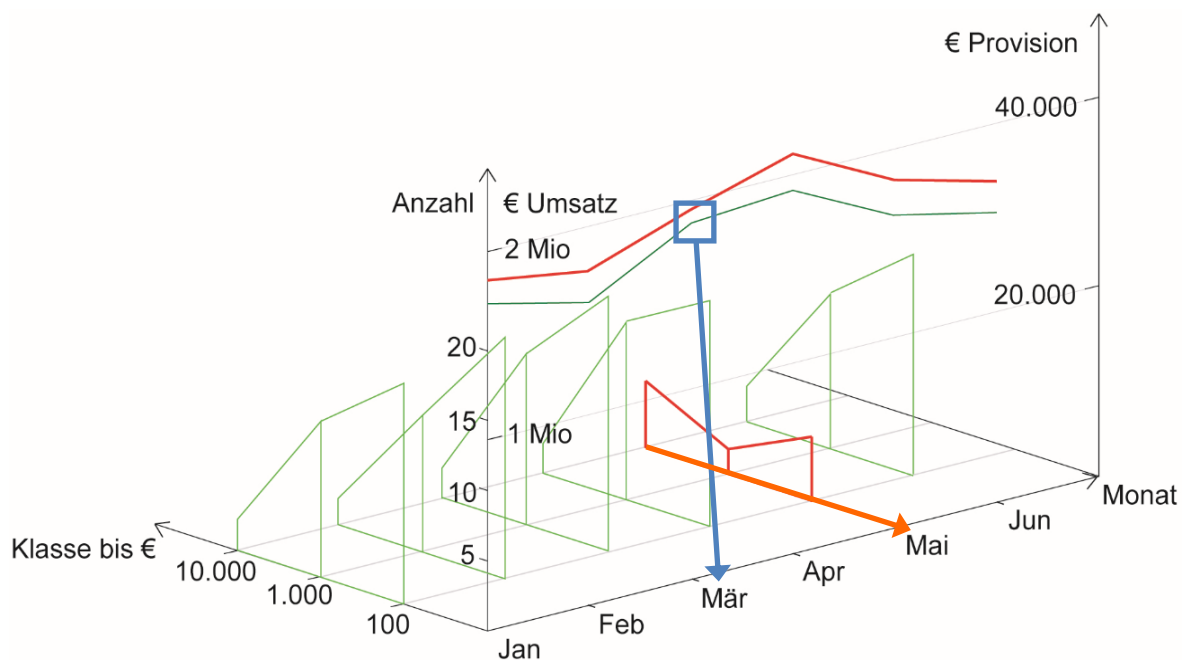
¹⁰⁰⁹ *Kowallik*, „Indien implementiert mit dem Faceless Assessment Scheme eine digitale Außenprüfung ohne Personen-Interaktion – Eine Blaupause für Deutschland?“, RET 5/2021, 49.

¹⁰¹⁰ Fokus Money online v. 26.3.2016: „Steuererklärung – Ihr Finanzamt stuft Sie jetzt nach Risiko ein – mit gravierenden Konsequenzen“, nachzulesen unter: https://www.focus.de/finanzen/steuern/nicht-nur-die-schufa-ihr-finanzamt-stuft-sie-kuenftig-nach-risiko-ein-und-das-hat-konsequenzen_id_5257804.html. Die rechtlichen Grundlagen dafür sind mit der Ergänzung von Absatz 5 in § 88 AO durch das Gesetz zur Modernisierung des Besteuerungsverfahrens v. 18.7.2016 (anzuwenden ab 1.1.2017) geschaffen worden: „Die Finanzbehörden können zur Beurteilung der Notwendigkeit weiterer Ermittlungen und Prüfungen für eine gleichmäßige und gesetzmäßige Festsetzung von Steuern und Steuervergütungen sowie Anrechnung von Steuerabzugsbeträgen und Vorauszahlungen automationsgestützte Systeme einsetzen (Risikomanagementsysteme).“

¹⁰¹¹ Vgl. z.B.: *Danielmeyer*, „Mit einem modernen Risikomanagement hin zur zeitnahen Betriebsprüfung?“, StBp 2021, 271.

etc. gearbeitet bzw. geforscht. Fraglich ist dabei allerdings, ob nicht eine bessere Einbindung der Menschen in die Digitalisierung und Vernetzung durch umfangreichere und zielführendere Bildungsmaßnahmen (s. [Pkt. 3.3.](#)) – neben ethischen Erwägungen – auch aus Qualitätssicht gerechtfertigt wäre¹⁰¹⁴; ab einem gewissen Automatisierungsgrad wird dies kaum mehr jemand seriös beurteilen können.

Nach Einschätzung des Autors gibt es in der visualisierenden Datenanalytik (als unterstützendem Werkzeug) durchaus noch Entwicklungspotential, welches aber nur im interdisziplinären Zusammenhang mit der Wahrnehmungsphysiologie zu erkennen ist. Danach sollten die heutigen technischen Möglichkeiten dazu genutzt werden, **visualisierende Prüfungsansätze** so **wahrnehmungsfreundlich und variabel auszugestalten**, dass sie trotz Komplexität von einem Großteil der potentiellen Anwender*innen mit etwas Übung und dem nötigen Methodenwissen richtig eingesetzt werden können:



So kombiniert diese abgebildete Zeichnung¹⁰¹⁵ als Beispiel regel- und musterbasierte Verbildlichungen zur Untersuchung von umsatzabhängigen Provisionen in *einer* Auswertung, wobei die Frontalebene aus der X- und den zwei Y-Achsen die Abhängigkeit beider Betriebsgrößen per doppelskaliertem Zeitreihenvergleich verfolgt; mit dem hierfür besonders

¹⁰¹² Beide Analyseansätze werden in dem Kooperationsprojekt „Tax Defence Analytics (TaDeA)“ zwischen dem Landesamt für Steuern Niedersachsen und der Universität Oldenburg eingesetzt, um prüfungsrelevante Vorgänge in den umfangreichen strukturierten und unstrukturierten Besteuerungsdaten frühzeitig und vollständig erkennen zu können; nachzulesen unter: <https://tadea-ol.de/>.

¹⁰¹³ Vgl.: EU-Projekt „VAT in the Digital Age (ViDA)“, nachzulesen unter: https://taxation-customs.ec.europa.eu/taxation-1/value-added-tax-vat/vat-digital-age_en.

¹⁰¹⁴ Giezek/Wähnert, „Künstliche Intelligenz vs. anwendergelenkten Softwareeinsatz – Ist KI generell leistungsfähiger oder nur besser zu verkaufen?“, DB 31/2019, 1707.

¹⁰¹⁵ Bildquelle: Wähnert, „Das große Potenzial von Strukturvergleichen in der Revision“, StC 4/2012, 18, 23.

geeigneten (zur unterschiedlichen Eignung bestimmter Zeitreihenvarianten für die Plausibilitätsprüfung s. den Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) Zeitreihenmodell ist umgehend zu erkennen, dass die Provisionssumme im März verhältnismäßig höher als in allen anderen Monaten ausfällt. Zusätzlich werden die monatlichen Provisionen über die Z-Achse i.V.m. der sekundären Y-Achse in die Anzahl an Einzelvorgängen pro (logarithmischer¹⁰¹⁶) Größenklasse zerlegt; auf diese Weise lässt sich neben der Höhe an monatlichen Provisionskosten auch feststellen, ob diese zeitweise anders abgerechnet wurden. Im Beispiel legt diese Kombi-Analyse offen, dass für den Monat „Mai“ doppelt so viele (betriebsbezogen) sehr große Provisionsrechnungen und folglich weniger mittlere und kleine vorliegen.

Natürlich setzt das richtige Lesen solcher komplex-grafischen Datenauswertungen erweiterte „spezielle kognitive Fähigkeiten“¹⁰¹⁷ voraus, die in Aus- und Fortbildungen angelegt werden müssen, aber durch die heutigen Möglichkeiten der Filter- und Drillbarkeit¹⁰¹⁸ sowie dreidimensionalen Animation dürfte eine bessere Durchdringung von großen Datenmengen erreichbar sein. Im Gegensatz zu sequentiellen „Durchstiegs-Funktionen“¹⁰¹⁹ in B(usiness-)I(ntelligence)-Programmen¹⁰²⁰ liegt bei der Beispielsgrafik der Vorteil in der gleichzeitigen, zusammenhängenden Darstellung von chronologischen und Verteilungseigenschaften. Dadurch wird eine Beurteilung im Kontext zwingender; Prüfungsfragen werden so häufig konkreter, weil Regel- und Musterstörungen einen inhaltlichen Zusammenhang haben (s. den Abschnitt „systematische Kombination aus unterschiedlichen Datenverbildlichungen“ unter [Pkt. 5.1.1.](#)).

¹⁰¹⁶ Eine logarithmische Klassenbildung belegt die gleichmäßigen Intervalle mit linear ansteigenden Exponenten wie in der Beispielsgrafik mit „ $10^2 - 10^3 - 10^4$ “ (entsprechend „100 – 1.000 – 10.000“) und lenkt dadurch den Fokus auf die selteneren höherwertigen Rechnungen.

¹⁰¹⁷ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 89: „Paper 2 betont die Eignung von Abbildungen, Informationen auf einen Blick zu erfassen und diese dem Betrachtenden [...] somit leichter zugänglich zu machen. Die Analyse statistischer Daten beginnt häufig mit der Wahl einer geeigneten Darstellungsform des Erhebungsbefundes. Die korrekte Interpretation solcher Grafiken ist allerdings nicht trivial und erfordert spezielle kognitive Fähigkeiten.“

¹⁰¹⁸ Über Drillfunktionen in (visualisierten) Auswertungen lässt sich einfach zwischen Konsolidierungsstufen wie z.B. bei Zeitreihenanalysen den Perioden „Jahr“, „Quartal“, „Monat“, „Tag“ etc. hin- und herschalten, wodurch insbesondere der Fuzzy-Gedanke (s. [Pkt. 2](#)) komfortabel umgesetzt werden kann.

¹⁰¹⁹ Die Verknüpfung unterschiedlicher Analyseperspektiven wird in vielen Datenanalyseprogrammen über sog. Drilldown-Optionen hergestellt, indem durch das Anklicken eines Diagrammpunkts dieser Teilbereich detaillierter bzw. in einer anderen Darstellungsform vertieft werden kann.

¹⁰²⁰ B(usiness-)I(ntelligence)-Programme sollen eine leistungsfähige, effiziente Selbstanalyse von Unternehmen ermöglichen, nachzulesen z.B. unter: <https://www.bigdata-insider.de/das-sind-die-beliebtesten-business-intelligence-loesungen-a-467954/> mit: „Wie entwickelt sich das eigene Unternehmen? In welchen Regionen bricht der Absatz ein? Was macht eigentlich der Mitbewerber? Auf Business Intelligence zu setzen, das heißt, bei der Entscheidungsfindung auf Fakten anstatt auf das Bauchgefühl zu vertrauen!“

Für B(usiness-)I(ntelligence)-Programme wird zwar mit der intuitiven Allgemeinverständlichkeit der (visualisierten) Auswertungen geworben¹⁰²¹, tatsächlich zeigt sich in der betrieblichen Praxis nach Feststellungen aus der Unternehmensberatung aber ein anderes Bild: „These days, it’s the rare CEO who doesn’t know that businesses must become analytics-driven. [...] Many CEOs have dedicated a lot of their own time to implementing analytics programs, appointed chief analytics officers (CAOs) or chief data officers (CDOs), and hired all sorts of data specialists. [...] But *frustrations* are beginning to surface; it’s starting to dawn on company executives that they’ve failed to convert their analytics pilots into scalable solutions.“¹⁰²² Auch von wissenschaftlichen Autoren wird fehlendes Wissen über die (digitale) Datenanalytik beim Beratungs- und Prüfungspersonal bemängelt: „Einige der zentralen explorativen Verfahren bilden die Data Mining-Ansätze, die prozessgestützt auffällige Muster und Beziehungen in den großen Datenvolumina ausfindig machen. [...] Frühere Studien haben ergeben, dass Abschlussprüfer häufig *Defizite* in der Fähigkeit von Mustererkennungen aufweisen, welche sich sowohl auf finanzielle als auch auf nichtfinanzielle Informationen erstrecken. Insofern erwächst hierbei ein konkreter Handlungsbedarf, diese Fähigkeiten bei der Aus- und Fortbildung [...] zu adressieren.“¹⁰²³ Diese defizitären Fähigkeiten werden vom Stifterverband für die Deutsche Wissenschaft e.V. konkretisiert: „Data Literacy gehört zu den wichtigen Future Skills, um in der digitalen Welt in Wissenschaft, Arbeitswelt und Gesellschaft bestehen und teilhaben zu können. Damit gehört sie zu den bedeutendsten Digitalkompetenzen der Zukunft. Data Literacy bezeichnet die Fähigkeit, Daten auf kritische Art und Weise zu sammeln, zu managen, zu bewerten und anzuwenden. Sie ermöglicht die systematische Umwandlung von Daten in Wissen und ist damit unverzichtbarer Bestandteil der Allgemeinbildung.“¹⁰²⁴ Bezüglich möglicher Ursachen wird festgestellt, dass „[b]eim Steuerberaterexamen [...] das vollständige Fehlen der Bereiche IT und Datenanalyse auffällt. [...] Defizite [innerhalb der berufsständischen Aus- und Weiterbildung des Wirtschaftsprüfers] bestehen vor allem im Bereich der Datenanalyse [...] Die Ausweitung des Lehrangebots – besonders an der Schnittstelle von Rechnungswesen, Steuerlehre und Steuerberatung zu IT(-Prüfung) und Datenanalyse – setzt eine angemessene Ressourcenausstattung, vor allem spezialisiertes

¹⁰²¹ Vgl. z.B.: Microsoft über das eigene BI-Produkt: „Nutzen Sie die aktuellen Fortschritte bei der Microsoft-KI, um auch Laien das Aufbereiten von Daten, das Erstellen von Machine Learning-Modellen und das Auswerten strukturierter und nicht strukturierter Daten wie Text und Bildern zu ermöglichen“, *früher* nachzulesen unter: <https://powerbi.microsoft.com/de-de/why-power-bi/>.

¹⁰²² Internetseite von McKinsey & Company, „Ten red flags signaling your analytics program will fail“, May 2018, nachzulesen unter: <https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/ten-red-flags-signaling-your-analytics-program-will-fail#>.

¹⁰²³ Velte/Drews, „Herausforderungen in der Wirtschaftsprüfung durch den Einsatz von Big-Data-Technologien – Welche Impulse muss die Prüfungsforschung geben?“, DB 43/2018, 2581, 2582 u. 2583.

¹⁰²⁴ Nachzulesen unter: <https://www.stifterverband.org/data-literacy-education>.

Lehrpersonal voraus.“¹⁰²⁵ In der Wirtschaftsprüfungspraxis werden diese Herausforderungen wahrgenommen: „Die Zukunft der Wirtschaftsprüfung ist digital: [...] Für Wirtschaftsprüfer heißt das, dass sie neben ihrem Fachwissen *Kenntnisse im IT-Bereich* benötigen, um diese Prozesse zu verstehen.“¹⁰²⁶ Jüngst hat der Bundesrechnungshof auch für die Finanzverwaltung darauf hingewiesen, dass „[t]raditionelle Methoden der steuerlichen Betriebsprüfung, beispielsweise Beleg- und Akteneinsicht, [...] an Effektivität und Effizienz [verlieren]. Nur IT-gestützte quantitative Prüfungsmethoden können auffällige Muster und Beziehungen in großen Datenmengen ersichtlich machen.“¹⁰²⁷ Entsprechend wird im BMF-Schreiben zu „[a]utomationsgestützte[n] quantitativen Prüfungsmethoden in der steuerlichen Außenprüfung“ deren Bedeutung formuliert: „Bei einer steuerlichen Außenprüfung können mathematisch-statistische (quantitative) Prüfungsmethoden eingesetzt werden, um die Besteuerungsgrundlagen des Steuerpflichtigen zu verproben und auf Plausibilität zu überprüfen. Auch das Erkennen von Prüffeldern unter Risikogesichtspunkten ist mit quantitativen Prüfungsmethoden möglich. [...] Den Ergebnissen aus der Anwendung quantitativer Prüfungsmethoden kann bei der Widerlegung der gesetzlichen Beweisvermutung des § 158 Absatz 1 der Abgabenordnung eine Beweiswirkung zukommen, besonders als Indiziengesamtheit.“¹⁰²⁸ In der „Ausbildungs- und Prüfungsordnung für die Steuerbeamtinnen und Steuerbeamten – gehobener Dienst“¹⁰²⁹ finden sich die dafür benötigten (heutigen) Schlüsselqualifikationen zur datenbasierten Sachaufklärung „Statistik“ und „Visualisierung“ jedoch genauso wenig, obwohl diese nach Auffassung der Bildungswissenschaft nicht intuitiv sind, sondern ausgiebig erlernt werden müssen¹⁰³⁰.

Dass (teil-)automatisierte digitale Datenanalyse z.B. mit Hilfe von fertigen Programmtools oder vorkonfektionierten Prüfmakros nicht gleichbedeutend mit zielführenden Prüfungsabläufen und richtiger Ergebnisinterpretation ist, lässt sich bspw. an den eklatanten Unterschieden gängiger Zeitreihenvarianten (s. den Abschnitt „Methodenkenntnis“ unter [Pkt. 4.2.](#)) sowie den abweichenden (Nicht-)Übereinstimmungswahrscheinlichkeiten scheinbar aussageidentischer Mustervergleiche (s. das Beispiel im Abschnitt „Güte der Übereinstimmung mit

¹⁰²⁵ Bravidor/Förster/Weißenberger, „Berufsstand 4.0: Wirtschaftsprüfer und Steuerberater zwischen Fachexpertise, IT und Datenanalyse – Ansätze für die Ausbildung künftiger Wirtschaftsprüfer und Steuerberater an Hochschulen und in der Praxis“, WPg 2020, 287, 291, 292 u. 294.

¹⁰²⁶ DB online v. 30.4.2019 (Sponsored News v. EY), „Die Zukunft der Wirtschaftsprüfung ist digital“, früher nachzulesen unter: <https://www.der-betrieb.de/meldungen/die-zukunft-der-wirtschaftspruefung-ist-digital/>.

¹⁰²⁷ Bundesrechnungshof, „Bemerkungen 2020“, Allgemeine Finanzverwaltung, Pkt. 20, nachzulesen unter: <https://www.bundesrechnungshof.de/de/veroeffentlichungen/produkte/bemerkungen-jahresberichte/jahresberichte/2020-hauptband/einzelplanbezogene-pruefungsergebnisse/finanzverwaltung/2020-20>.

¹⁰²⁸ BMF-Schreiben v. 5.9.2023, BStBl. I 2023, 1594, Rz. 1 u. 4.

¹⁰²⁹ Nachzulesen unter: https://www.gesetze-im-internet.de/stbapo_2022/anlage_11.html.

¹⁰³⁰ Ihden, „Die Relevanz statistischer Methoden in der Rechtsprechung und mögliche Implikationen für die juristische Ausbildung“, Diss., Verlag Nomos, 2018, 64 u. 89.

dem Idealmuster von der Datenmenge abhängig“ unter [Pkt. 4.3.](#)) nachvollziehen; ebenso geht dies aus verbreiteten Mängeln in der Beschreibung von Datenanalysemethoden hervor.¹⁰³¹ Deshalb ist ein ausreichendes methodisches und prozessuales Verständnis (inklusive der Grundlagen wie Begriffe, Funktionen, Zusammenhänge, Unterschiede etc.) Voraussetzung dafür, geeignete Mittel für eine Prüfungsfrage zu wählen und sinnvoll einzeln oder in Kombination anzuwenden. Die Einrichtung der Bundestags-Enquete-Kommission „Berufliche Ausbildung in der digitalisierten Arbeitswelt“¹⁰³² ließ vermuten, dass es bisher noch zu wenige zufriedenstellende Konzepte gab und gibt. Demgegenüber zeigen die Erfahrungen aus der Fortbildung tausender Prüfer*innen in SRP und dem länderübergreifenden Visualisierungs-Konzept, dass hierdurch ein gelungener Einstieg in moderne Arbeitsweisen wirksam unterstützt wird; eine Bund-Länder-Unterarbeitsgruppe „Summarische Risikoprüfung (SRP) und visualisierende Datenanalyse“ arbeitet permanent an der Weiterentwicklung von Aus- und Weiterbildungskonzepten sowie der Bereitstellung von technischen Hilfswerkzeugen. Mittelfristig muss aber auch in den nicht-technischen rechtsnahen Berufsausbildungen erkannt werden, wie wichtig die propädeutische Vermittlung zeitgemäßer Schlüsselqualifikationen für die spätere Berufsausübung ist, damit die umfangreichen, leistungsfähigen Hilfsmittel der digital-vernetzten Arbeitswelt i.S. von mehr Qualität und Effizienz genutzt werden können.

¹⁰³¹ Vgl. z.B.: *Bis/e*, „Chi-Quadrat-Test‘ und Zeitreihenvergleich: Keine Schätzung bei ordnungsgemäßer Buchführung“, PStR 1/2012, 15: „Der ‚Chi-Quadrat-Test‘ analysiert die Verteilung bestimmter Ziffern und basiert auf der Erkenntnis, dass jeder Mensch Lieblingsziffern hat und bei frei erfundenen Zahlen (z.B. nachgetragenen Tageseinnahmen oder -ausgaben) häufiger verwendet als andere.“ Tatsächlich ist der χ^2 -Test (Chi-Quadrat-Test) ein allgemeiner Wahrscheinlichkeitstest, der mehrklassige Verteilungen unter Berücksichtigung der Erkenntnisse der Wahrscheinlichkeitslehre auf Übereinstimmung prüft. Eine (zwingende) Verbindung zur Gleichverteilung von endständigen Ziffern in Wirtschaftsdaten besteht nicht. Vielmehr vermischt die Aussage die zwei Schritte „Erwartung“ und „Übereinstimmungsprüfung“ i.R. der musterbasierten Analyse und sorgt somit für Verständnisprobleme und Fehleinschätzungen; so auch: *Sell*, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung“, Diss., Verlag Dr. Kovač, 2021, 194: „[...] irritiert, dass die Verprobungsmethode ‚Ziffernanalyse‘ regelmäßig als ‚Chi-Quadrat-Test‘ geführt wird.“

¹⁰³² Nachzulesen unter: <https://www.bundestag.de/dokumente/textarchiv/2018/kw39-pa-enquete-berufliche-bildung-567960>.

6. (Ausgewählte) Literatur zur „Neuen Prüfungsmethodik/-technik“¹⁰³³

-  *Achilles*, „Kassen-Nachschau nach § 146b AO“, DB 1-2/2018, 18.
-  *Achilles*, „Kassenführung – Bargeschäfte sicher dokumentieren“, DATEV eG, 2. Auflage 2018.
-  *Achilles*, „Kassenführung in bargeldintensiven Unternehmen“, BoD Verlag, 2014.
-  *Achilles/Danielmeyer*, „Die Kassenführung im Spannungsfeld von TSE, Corona und Mehrwertsteuersenkung, RET 4/2020, 18.
-  *Achilles/Jope*, „Corona-Dokumentation – Checkliste zur Vermeidung von Fehlinterpretationen in der Bp“, DB 2020, 1417.
-  *Achilles/Pump*, „Lexikon der Kassenführung“, DATEV eG, 2018.
-  *Anders/Gärtner*, „Manipulationssicherheit in der Bargeldbranche durch Einzelaufzeichnungen“, Stbg 2/2016, 49.
-  *Anders/Rühmann*, „Aufbewahrungspflicht für Protokolle zur Kassenprogrammierung – Nichtaufbewahrung als schwerer Ordnungsmangel der Kassenführung“, BBK 13/2013, 627.
-  *Bachmann/Richter/Steinborn*, „Führt ein anderer Anpassungstest in der Außenprüfung zu einer anderen Schlussfolgerung? – Chi-Quadrat- und Kolmogorov-Smirnov-Anpassungstest in der Summarischen Risikoprüfung“, StBp 2019, 38.
-  *Barthel*, „Skalierte Außenprüfung“, Stbg 3/2018, 115.
-  *Barthel*, „Betriebsprüfung: Die griffweise Schätzung“, Stbg 7-8/2017, 315.
-  *Baselt*, „Einführung in die Strukturanalyse“, Lehrgangsskript der Bundesfinanzakademie.
-  *Becker*, „Außenprüfung digital – Prüfungsmethoden im Fokus“, DStR 24-25/2016, 1386 u. 1430.
-  *Becker*, „Verprobung der Pfandverbindlichkeiten von Mineralbrunnen mit Hilfe der Visualisierung“, Bonner Bp-Nachrichten 12/2015, 19.
-  *Becker*, „Textbausteinsystem Schätzung“, juris (finanzverwaltungsintern): Fachinfosystem Bp/Allgemeine Prüfungshinweise/Arbeitspapiere und Leitfäden/TBS Schätzung.
-  *Becker/Danielmeyer/Neubert/Unger*, „Digitale Offensive‘ der Finanzverwaltung: Die Schnittstellen-Verprobung (SSV)“, DStR 51/2016, 2983.
-  *Becker/Giezek/Webel/Wähnert*, „Der Beanstandungsanlass nach § 158 AO“, DStR 32/2016, 1878.

¹⁰³³ Grau gedruckte Quellen sind nach Kenntnisstand des Autors nur innerhalb der Finanzverwaltung erhältlich.

-  *Becker/Schumann/Wähnert*, „Aktuelle Rechtsprechung zur Neuen Prüfungstechnik: Summarische Risikoprüfung (SRP) und Quantilsschätzung“, DStR 23/2017, 1243.
-  *Becker/Wiethölter*, „Aufzeichnungspflichten beim Einnahmen-Überschuss-Rechner gemäß § 4 Abs. 3 EStG – Der falsch verstandene Beschluss des BFH vom 16. 2. 2006 X B 57/05“, StBp 2009, 377.
-  *Becker/Wiethölter*, „Aufbewahrung der Belege und Einzelaufzeichnung von Betriebseinnahmen bei Gewinnermittlern nach § 4 Abs. 3 EStG - oder: die Ruhe nach dem durch Beschluss des BFH vom 16.2.2006 verursachten Sturm“, StBp 2009, 239.
-  *Bendisch*, „Die Strukturanalyse als Instrument der Betriebsprüfung“, wissenschaftliches Gutachten im Auftrag des Bundesministeriums für Finanzen, Mai 1980 (Az. BFA – P 3220 – 5/80).
-  *Benzel*, „Der Weg zur Digitalisierung der Betriebsprüfung“, RET 3/2022, 30.
-  *Beyer*, „Kritische Anmerkungen zur Summarischen Risikoprüfung der Finanzämter – Praxishinweise für Berater und Unternehmer bei Betriebsprüfung und Steuerstraßverfahren“, NWB 40/2019, 2931.
-  *Beyer*, „FinMin Schleswig-Holstein veröffentlicht SRP-Handbuch – erste kritische Hinweise zur Anwendbarkeit im Steuerstraßverfahren“, NWB 17/2019, 1209.
-  *Bundesfinanzakademie*, „Podiumsdiskussion: Monetary Unit Sampling (MUS)“ sowie „Zehnjähriges Jubiläum der Summarischen Risikoprüfung“, Quartals-Info I/2018, downloadbar unter:
https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Ministerium/Geschaeftsbereich/Bundesfinanzakademie/Aktuelles/BFA_Quartalsinfo_2018_I.html.
-  *Bundesfinanzakademie*, „Podiumsveranstaltung ‚Prüfung und Verprobung von Big Data‘“, QuartalsInfo I/2015, 11, downloadbar unter:
https://www.bundesfinanzministerium.de/Content/DE/Standardartikel/Ministerium/Geschaeftsbereich/Bundesfinanzakademie/Aktuelles/BFA_QuartalsInfo_2015_I.html.
-  *Brähler/Bensmann/Emke*, „Der Einsatz mathematisch-statistischer Methoden in der digitalen Betriebsprüfung, Ilmenauer Schriften zur Betriebswirtschaftslehre 4/2010, Verlag proWiWi e.V., 2010.
-  *Braun/Köppe*, „Digitalisierung in der Verrechnungspreisfunktion – Entwicklungen und Chancen“, DB 8/2019, 394.
-  *Brinkmann*, „Die Begründung des Sicherheitszuschlags“, StBp 2018, 212.
-  *Brinkmann*, „Schätzungen im Steuerrecht“, Erich Schmidt Verlag, 6. Auflage 2023.
-  *Brinkmann*, „Der so genannte Sicherheitszuschlag“, StBp 2014, 29 u. 69.
-  *Brinkmann*, „Strafrechtliche Aspekte der Schätzung“, StBp 2013, 250 u. 291.

-  *Danielmeyer*, „Datenanalyse aus dem Blickwinkel der Außenprüfung“, *beck.digitax* 4/2023, 218.
-  *Danielmeyer*, „Der ‚digitale‘ Steuerpflichtige wird virtuell – Die Evolution der Außenprüfung“, *StBp* 2022, 376.
-  *Danielmeyer*, „Die Digitalisierung der Betriebsprüfung“, Erich Schmidt Verlag, 2022.
-  *Danielmeyer*, „Die Anforderungen an die digitale Betriebsprüfung im Jahre 2022“, *StBp* 2021, 323.
-  *Danielmeyer*, „Cybercrime und andere elementare Datenverluste in der Steuerwelt“, *RET* 6/2021, 47.
-  *Danielmeyer*, „Mit einem modernen Risikomanagement hin zu zeitnahen Betriebsprüfung?“, *StBp* 2021, 271.
-  *Danielmeyer*, „Die digitale Betriebsprüfung im Jahr 2021 – mit dem digitalen Datenabdruck zur Anomalieerkennung“, *RET* 3/2021, 39.
-  *Danielmeyer*, „Blockchain in der Steuerwelt“, *beck.digitax* 2/2021, 93.
-  *Danielmeyer*, „Der Anwendungserlass für elektronische Aufzeichnungssysteme (§ 146a AO) - Anmerkungen zum BMF-Schreiben vom 17.6.2019“, *StuB* 15/2019, 89.
-  *Danielmeyer/Neubert/Unger*, „Die Betriebsprüfung von A–Z“, *StBp* 2017, 291.
-  *Danielmeyer/Neubert/Unger*, „Praxiserfahrungen zu VORSYSTEMEN-Prüffeld – Visualisierte Schnittstellenverprobung“, *StBp* 2016, 322.
-  *Danielmeyer/Neubert/Unger*, „Schnittstellenverprobung im Rahmen der SRP“, *Bonner Bp-Nachrichten* 8/2016, 25.
-  *Danielmeyer/Neubert/Unger*, „Schnittstellen-Verprobung – Das Prüffeld mit Zukunft!“, *Bonner Bp-Nachrichten* 4/2016, 17.
-  *Deggendorfer Forum zur digitalen Datenanalyse e. V.*, „Tagungsband des 15. Deggendorfer Forum zur digitalen Datenanalyse 2019: Künstliche Intelligenz – Chancen und Risiken für Wirtschaftsprüfung und Finanzwirtschaft“, IDW-Verlag, 2019.
-  *Deggendorfer Forum zur digitalen Datenanalyse e. V.*, „Tagungsband des 14. Deggendorfer Forum zur digitalen Datenanalyse 2018: Digitale Geschäftsmodelle – Auswirkungen zunehmender Komplexität auf Prüfung, Revision und Controlling“, IDW-Verlag, 2018.
-  *Diller/Schmid/Späth/Kühne*, „Zifferntests in der Betriebsprüfung – Chancen und Risiken“, *DStR* 6/2015, 311.
-  *Diverse*, „Kassenprüfungen und Zeitreihenvergleich in der Praxis“, *BBK-Sonderheft* 2016.
-  *Ehlers*, „Tax Compliance – Neue Beratungsfelder bei Betreuung kleiner und mittlerer Unternehmen“, *NWB* 18/2012, 1535.

-  *Ettinger/Frielinghaus*, „Massendatenanalyse und Benchmarkstudien – einfach Gewinn bringend“, Bonner Bp-Nachrichten I u. II/2019, 44 u. 86.
-  *Ettinger/Papenfuß*, „Visualisierung und Verrechnungspreise oder ‚ein Bild sagt mehr als tausend Worte‘“, Bonner Bp-Nachrichten II/2018, 26.
-  *Franke-Roericht*, „‚Wenn die Kasse klüngelt‘: Beihilfe durch Verkauf von Kassenmanipulationssoftware“, PStR 3/2015, 66.
-  *Frielinghaus/Melsa/Wähnert*, „Digitale Verrechnungspreisprüfung: vorgangsbezogen-interaktive Visualisierung mit Power BI (Bp) vs. (Multi-)Regressionsanalyse (Wp/Stb)“, Bonner Bp-Nachrichten I 2021, 45.
-  *Gebbers*, „Durchführung eines z-Tests zur Prüfung der Umsätze“, StBp 2012, 308, 340 u. StBp 2013, 13.
-  *Gebbers*, „Analyse der Rechtsprechung zum Chi-Quadrat-Test in der Außenprüfung – Anwendung zur Prüfung von Erlösen und Fahrtenbüchern“, StBp 2008, 209 u. 290.
-  *Gebbers*, „Die Strukturanalyse“, StBp 2007, 225, 257 u. 289.
-  *Giezek*, „Monetary Unit Sampling“, Gabler Verlag, 2011.
-  *Giezek/Rupprecht/Wähnert*, „Wie sicher muss ein Verprobungsergebnis sein? – Anleihen aus der Kriminalistik zur sinnvollen Einordnung des Beanstandungsanlasses“, BBK 5/2017, 236.
-  *Giezek/Wähnert*, „Anwendung des Monetary Unit Sampling in der Betriebsprüfung – Ein falsches Verständnis des Konfidenzniveaus erschwert den Zugang zu objektiven Stichprobenverfahren“, NWB 14/2022, 980.
-  *Giezek/Wähnert*, „Digitalisierung braucht Kompetenz“, Creditreform 2/2021, 23.
-  *Giezek/Wähnert*, „Künstliche Intelligenz vs. anwendergelenkten Softwareeinsatz – Ist KI generell leistungsfähiger oder nur besser zu verkaufen?“, DB 31/2019, 1707.
-  *Giezek/Wähnert*, „Wahrscheinlichkeitstheorie in der Betriebsprüfung – Von der qualifizierten Stichprobenprüfung zur Ziffernanalyse“, DB 9/2018, 470.
-  *Giezek/Wähnert*, „Schätzungsmethoden der Betriebsprüfung im Vergleich – Einordnung der wichtigsten Schätzungsverfahren durch Praxis und Wissenschaft“, BBK 21/2017, 998.
-  *Giezek/Wähnert/Becker*, „Anwendung des Monetary Unit Sampling (MUS) in der steuerlichen Betriebsprüfung – Qualifizierte Untersuchung und Bewertung von Stichproben“, StBp 12/2016, 347.
-  *Graf*, „Anforderungen an die Vorlage von Einzelaufzeichnungen - Ende der ‚Schonzeit‘ für Daten aus Vor-Systemen“, BBK 18/2013, 875.

-  *Graf*, „Paradigmenwechsel durch die Datenanalyse – Datenanalyse und Mitwirkungsverpflichtung bei steuerlichen Betriebsprüfungen – eine Symbiose?“, BBK 7/2012, 331.
-  *Härtl*, „Einzelaufzeichnungspflicht und Datenzugriff auf digital aufgezeichnete Geschäftsvorfälle von Bargeschäften – (Kein) Datenzugriff bei Registrierkassen, Apotheken?“, StBp 2014, 1 u. 45.
-  *Härtl*, „Der Datenzugriff bei Überschussrechnern nach § 4 Abs. 3 EStG im Licht des BFH-Urteils vom 24.6.2009 (VIII R 80/06), im Besonderen auf gesetzlich nicht vorgeschriebene digitale Aufzeichnungen im Sinne des § 146 Abs. 6 AO“, StBp 2009, 337.
-  *Härtl/Schieder*, „Ordnungsmäßigkeit digital geführter Erlösaufzeichnungen – Elektronische Registrierkassen und digitale Erlöserfassungssysteme im Brennpunkt des Steuerrisikos Erlösverkürzung“, StBp 2011, 33, 68 u. 97.
-  *Harle/Nüdling/Olles*, „Die moderne Betriebsprüfung“, NWB-Verlag, 5. Auflage 2023.
-  *Hauschildt/Wähnert*, „Vollständigkeitsprüfung mit IDEA – Was bedeutet die elektronische Betriebsprüfung für Unternehmer und Berater“, NWB 40/2007, 3539.
-  *Henn*, „Verfahrensdokumentationen nach den GoBD“, DB 5/2016, 254.
-  *Huber*, „Die Zukunft des steuerlichen Risikomanagements im Einnahmehereich“ – Der Vertrauensvorschuss nach § 158 AO im historischen Zusammenhang und im Umfeld der Aufzeichnungstechnik der Gegenwart“, Protokoll zum 130. Bochumer Steuerseminar für Praktiker und Doktoranden vom 14.11.2014 (Wilms), downloadbar unter: http://www.fachanwalt-fuer-steuerrecht.de/pdf/Protokoll_20100091.pdf.
-  *Huber*, „Die Verwertung von Echtdaten aus Kassennachschauen mit Datenzugriff – Praxisbeispiele zur interaktiven Prüfungstechnik bei Bargeschäften“, BBK 16/2014, 757.
-  *Huber*, „Gedanken zur Zukunft des steuerlichen Risikomanagements im Erlösbereich“, StBp 2014, 121, 153, 185, 221, 245, 277 u. 317.
-  *Huber*, „Praktische Erfahrungen mit der Kassenführung in Österreich – Umsetzung der Kassenrichtlinie 2012 und Kassennachschau durch die Finanzpolizei“, BBK 6/2014, 286.
-  *Huber*, „Steueraufsicht und Betriebsprüfung in der Zeit der Kassenandroiden und ohne INSIKA – Grundgedanken, Ziele, Risiken und zweitbeste Lösungen“, StBp 2012, 301, 333, StBp 2013, 5, 35, 71 u. 93.
-  *Huber*, „Risikomanagement in der Außenprüfung“, Protokoll zum Bochumer Steuerseminar für Praktiker und Doktoranden vom 19. Februar 2010 (Gabert), downloadbar unter: http://www.fachanwalt-fuer-steuerrecht.de/pdf/protokoll_20100002.pdf.
-  *Huber*, „Die neue Prüfungstechnik – Mathematische Weiterentwicklungen – Theorie der Prüfungsmethoden & praktische Revisionsbeispiele“, LexisNexis-Verlag ARD Orac, 2009.

-  *Huber*, „Über Registrierkassen, Phantomware, Zapping und Fiskallösungen aus Deutschland und Österreich“, StBp 2009, 153, 185, 217, 253, 286, 317 u. 342.
-  *Huber*, „Digitale Ziffernanalyse versus Strukturanalyse und die logische Herleitung von Benford (NBL)“, StBp 2008, 241 u. 273.
-  *Huber*, „Weiterentwickelte und neue Methoden der Überprüfung, Verprobung und Schätzung“, StBp 2002, 199, 233, 258 u. 293.
-  *Huber/Reckendorf/Zisky*, „Die Unveränderbarkeit der (Kassen-)Buchführung nach § 146 Abs. 4 AO im EDV-Zeitalter und INSIKA – Teil 1: Historische Wurzeln und Realzustand, Teil 2: Probleme, Ursachen und Lösungen, Teil 3: Konzept, Technik, Praxis“, BBK 12/2013, 567, BBK 13/2013, 610 u. BBK 14/2013, 663.
-  *Huber/Seer*, „Steuerverwaltung im 21. Jahrhundert: Risikomanagement und Compliance“, StuW 4/2007, 355.
-  *Huber/Wähnert*, „Der systematische Einsatz neuer Prüfverfahren in der Finanzverwaltung – Neue interaktive Prüfungstechnik (NiPt)“, NWB 36/2009, 2814.
-  *Huber/Wähnert*, „Das Kölner Zeitreihenurteil und das Projekt ‚NiPt‘“, StBp 2009, 207.
-  *Jansen*, „Begründungstiefe in Fällen der Hinzuschätzung – Aktuelle Schwachstellen in der Kassenführung, Handlungsempfehlungen und das Verbot der Prämierung von Mitwirkungspflichtverletzungen“, StBp 2021, 59.
-  juris Lexikon Steuerrecht, „Betriebsprüfung – Quantilsschätzung“, 2. Auflage Dez. 2021.
-  juris Lexikon Steuerrecht, „Betriebsprüfung – Summarische Risikoprüfung (SRP)“, 2. Auflage Dez. 2021.
-  *Kläne*, „Registrierkassen im Fokus der Außenprüfung – Aktuelle Entwicklung und Rahmenbedingungen“, NWB 13/2013, 923.
-  *Klingebiel*, „Ordnungsgemäße Kassenführung und sachliche Richtigkeit der Buchführung – Voraussetzungen für Hinzuschätzungsberechtigungen bei Außenprüfungen“, NWB 41/2008, 3849.
-  *Knief*, „Eine BWA-Mindestanalyse - Eine logische Konsequenz der Digitalisierung des Rechnungswesens und der verschiedenen gesetzlichen Anforderungen“, StBp 1/2017, 21.
-  *Knief*, „Die grafische Betrachtung der unterjährigen Unternehmensentwicklung für mittelständische Unternehmen – wider die Stichtagsanalyse“, DStR 43/2018, 2080.
-  *Kompenhans/Wermelt*, „‘Big Data‘ – Potenziale für innovative Abschlussprüfungen“, BB 2018, 299.
-  *Krüger*, „Chemie 4.0: Digitalisierung, Verrechnungspreisbestimmung und Betriebsprüfung“, DStZ 17/2019, 659.

-  *Krumm*, „Rechtsfragen der digitalen Betriebsprüfung – (Summarische Risikoprüfung)“, DB 20/2017, 1105.
-  *Kulosa*, „Herausforderungen der digitalen Betriebsprüfung, u.a. bei bargeldintensiven Betrieben“, FR 2017, 501.
-  *Kulosa*, „Replik zum Beitrag von Wolenski zum Zeitreihenurteil des X. Senats des BFH“, Stbg 6/2016, 271.
-  *Liekenbrock*, „Digital Tax Compliance – Praxishinweise zur technischen Umsetzung von Tax Compliance Management Systemen“, Ubg 2018, 43.
-  *Liekenbrock/Danielmeyer*, „Data Governance als Reaktion auf die zunehmend datengetriebene Betriebsprüfung?“, RET 2/2021, 32.
-  *Liekenbrock/Ewel/Braul*, „GoBD-Compliance als Fundament für ein stabiles Tax Compliance Management System“, Ubg 2022, 439.
-  *Liekenbrock/Wähnert*, „Digitale Betriebprüfung – Effiziente Vorbereitung und planvolle Begleitung“, IDW-Verlag, 2021.
-  *Mack*, „Überprüfung und Schätzung im so genannten bargeldintensiven Handel – Problematik der Prüfsysteme von elektronischen Kassensystemen“, AO-StB 1/2016, 17.
-  *Madauß*, „Statistische-mathematische Methoden in Besteuerungs- und Steuerstraßverfahren“, NZWiSt 1/2014, 24.
-  *Mehret/Wähnert*, „Prüfungsnetze vs. Einzelmethoden: Ein wichtiger, (bisher) verkannter Vorteil der Summarischen Risikoprüfung (SRP)“, DStR 6/2018, 314.
-  *Neubert*, „Die digitalen Methoden der modernen Betriebsprüfung“, AO-StB 9/2018, 291.
-  *Nöcker*, „Quantile in der BFH-Rechtsprechung – Eine erste Analyse der beiden BFH-Beschlüsse X B 16/17 und IV B 4/17 zur sog. Quantilsschätzung“, NWB 40/2017, 3050.
-  *Nöcker*, „Update Betriebsprüfung – Neues vom BFH und die Auswirkungen für die Praxis“, NWB 42/2016, 3157.
-  *Nöcker*, „Zeitreihenvergleich im Gleitschlitten versus Programmierprotokolle der Registrierkasse – Anmerkung zum BFH-Urteil X R 20/13 und einiges mehr“, NWB 48/2015, 3548.
-  *Nordhoff*, „Methodische Hinweise zum Umgang mit nicht behebbaren Messunsicherheiten im Rahmen der statistischen Datenanalyse und Steuerschätzung“, DB 9/2020, 417.
-  *Nowotzin/Teutemacher*, „Die neue Wunderwaffe der Betriebsprüfung: Die Summarische Risikoprüfung (SRP) Das müssen Sie als Steuerberater wissen“, IWW BBP-Sonderheft 2017.
-  *Pitz/Klapdor/Gardian/Haselau*, „Das Benfordsche Gesetz als Werkzeug zur Aufdeckung von Umsatzmanipulationen in der Gastronomie?“, StuW 3/2021, 232.

-  *Posch*, „Digitale Ziffernanalyse – Eine Erweiterung des Newcomb-Benford-Verfahrens“, StBp 2010, 338.
-  *Pump*, „Die Kassenführung im bargeldintensiven Betrieb am Beispiel des Friseurs (§§ 158, 162 AO) – Kassenführung mit Barerlösen im bargeldintensiven Betrieb“, StBp 2016, 131, 169, 199, 262, 289, 327 u. 364.
-  *Pump*, „Die Einzelaufzeichnungspflicht von Barerlösen bei offenen Ladenkassen bei Umsatzsteuer und Einkommensteuer – Kryptische Rechtsprechung und der Einsatz von Registrierkassen“, StBp 2015, 1.
-  *Pump*, „Der teure Verzicht auf Verwendung einer Registrierkasse – Beratungsfragen für steuerliche Berater und ihre Mandanten“, DStZ 18/2014, 648.
-  *Pump*, „Die Sicherung des Steuer- und Sozialversicherungsaufkommens bei Taxiunternehmen durch § 30 AO – Integrierte Sicherheitslösung (INSIKA) versus Manipulationssoftware und andere Betrügereien“, VR 11/2013, 361.
-  *Pump*, „Rechtsfolgen bei Verwendung von Manipulationssoftware (Zappersoftware) – Steuerrecht, Steuerstrafrecht und besonderes Strafrecht“, DStZ 9/2013, 299.
-  *Pump/Heidl*, „Hilfestellung bei der Kassenführung durch den steuerlichen Berater – Ärger wegen § 162 AO durch unzulängliche Kassenführung“, StBp 2014, 162 u. 204.
-  *Pump/Wähnert*, „Das BFH-Urteil zum Zeitreihenvergleich als Verprobungs- und Schätzungsmethode“, NWB 39/2015, 2869.
-  *Prinz*, „Modernisierung der steuerlichen Bp – Der Laptop allein kann es sicher nicht richten!“, DB 45/2019, M4.
-  *Rau*, „Statistisch-mathematische Methoden der steuerlichen Betriebsprüfung und die Strukturanalyse als ergänzende Alternative“, Reihe: „Steuern, Wirtschaft und Recht“, Bd. 316, Eul-Verlag, 2012.
-  *Risse/Jotanovic/Krapp/Dallhammer*, „Das steuerliche Kontrollumfeld: Katalysator für fortschrittliche Prüfsoftware“, RET 5/2023, 32.
-  *Schindler/Totzek*, „Anwendungsmöglichkeiten von Data & Analytics im Verrechnungspreiskontext“, IWB 20/2019, 827.
-  *Schlegel*, „Monetary Unit Sampling (MUS) – Eine verfahrensrechtliche Einordnung zur Beweiskraft der Buchführung und steuerlichen Schätzungsbefugnis der Stichprobenauswertung“, StBp 2020, 350.
-  *Schmidt*, „Monetary Unit Sampling (MUS) – ein kurzer Überblick – Ein mathematisch-statistisches Verfahren als weiteres wirksames und zeitgemäßes Tool im „Werkzeugkasten“ der digitalen Betriebsprüfung!“, NWB 5/2022, 328.

-  *Schmitz/Gebbers/Huber*, „Nachruf auf Wilhelm Baselt, ‚Vater‘ der modernen steuerlichen Prüfungstechnik“, StBp 2011, 93.
-  *Schöneborn*, „Verrechnungspreismanagement 4.0 – Zum Stand der Digitalisierung in der Praxis und welche Kompetenzen für ein erfolgreiches IT-Projekt entscheidend sind“, DB 14/2019, 737.
-  *Scholz*, „Betriebsprüfung: SRP – jetzt allgemein zugänglich“ AO-StB 7/2019, 226.
-  *Scholz*, „Der Quantensprung für die Schätzung – Die Quantilsschätzung: Ein Bestandteil der Summarischen Risikoprüfung SRP“, WP Praxis 9/2018, 281.
-  *Scholz*, „Überlegungen des BFH zu Schätzmethoden – Die Quantilsschätzung nach dem BFH-Beschluss vom 12.7.2017“, NWB-BB 11/2017, 351.
-  *Scholz*, „Einsatz der Zeitreihenanalytik in der Prüfungspraxis – heute – BFH-Rechtsprechung im Kontext zur Summarischen Risikoprüfung SRP“, BBK 12/2016, 603.
-  *Schütt*, „Mathematisch-statistische Methoden in der Außenprüfung – Über Pauschalisierungen, Widersprüchlichkeiten und zu hohe Anforderungen im Kontext des Zeitreihenvergleichs“, StBp 2018, 323.
-  *Schützeberg*, „Die Schätzung im Besteuerungs- und im Strafverfahren“, StBp 2009, 33.
-  *Schumann*, „Urkundenvorlagepflicht und Datenzugriff in Bezug auf IFRS-Abschlüsse und Aufzeichnungen“, Ubg 2018, 400.
-  *Schumann*, „Voraussetzungen und Verfahren der abgaberechtlichen Schätzung nach § 162 Abs. 1 und 2 AO – insbesondere im Fokus der steuerlichen Außenprüfung“, Dissertation im Fachbereich Rechts- und Wirtschaftswissenschaften der Johannes Gutenberg-Universität Mainz, 2009.
-  *Schumann/Wähnert*, „Die Quantilsschätzung – Schätzung anhand eines objektivierten Leistungsfähigkeitsmaßes“, Stbg 12/2012, 535.
-  *Schwenkert*, „Erkennen Sie den Unterschied? – Ersetzendes Scannen“, DATEV magazin 2/2015, 16.
-  *Sell*, „Quantitative Verprobungsmethoden in der Betriebsprüfung – Evaluation mathematisch-statistischer Verfahren und ihrer Verwertbarkeit in der Rechtsprechung, Verlag Dr. Kovač, 2021.
-  *Sievert/Wenzel*, „Reicht das SRP-Ergebnis für einen Anfangsverdacht?“, PStR 2023, 20.
-  *Sosna*, „Statistische Ziffernanalyse“, StBp 2004, 249 u. StBp 2005, 97.
-  *Sosna*, „Einsatz statistischer Methoden zur Risikoanalyse, Recherche und Lokalisierung von Steuerausfällen“, StBp 2000, 41 u. 68.
-  *Steinhauff*, „Kritische Anmerkungen zur Summarischen Risikoprüfung (SRP) der Finanzämter“, AO-StB 12/2019, 388.

-  *Teutemacher*, „Handbuch zur Kassenführung – Praxishandbuch für die rechtssichere Umsetzung“, NWB-Verlag, 2. Auflage 2018.
-  *Teutemacher*, „Die Kassennachschau kommt! – Die neue Form der ‚Feldüberwachung‘ durch die Finanzverwaltung“, BBK 24/2017, 1160.
-  *Teutemacher*, „Zulässigkeit einer Schätzung mittels Zeitreihenvergleichs – Auswirkungen der jüngsten BFH-Rechtsprechung auf die Betriebsprüfung“, BBK 11/2016, 544.
-  *Teutemacher*, „Aufbewahrung digitaler Unterlagen bei Bargeschäften – Checkliste für eine ordnungsgemäße Kassenführung“, BBK 23/2012, 1073.
-  *Velte/Drews*, „Herausforderungen in der Wirtschaftsprüfung durch den Einsatz von Big-Data-Technologien – Welche Impulse muss die Prüfungsforschung geben?“, DB 43/2018, 2581.
-  *Wähnert*, „BMF-Schreiben „Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung“, RET 1/2024, 40.
-  *Wähnert*, „Die „Summarische Risikoprüfung“ in der Auslandsfachprüfung – Datenbasiert-transaktionale Verrechnungspreisprüfung“, StBp 2022, 408.
-  *Wähnert*, „Missverständnisse zur Quantilsschätzung – Schätzungen mit IT-gestützten, quantitativen Methoden“, DStR 29/2022, 1470.
-  *Wähnert*, „Bundesrechnungshof: ‚[E]inheitliche Maßstäbe für den Einsatz IT-gestützter quantitativer Prüfungsmethoden‘ – Widersprüche bei der Anerkennung von digitalen Prüfungskonzepten wie der ‚Summarischen Risikoprüfung (SRP)‘“, StBp 2022, 3.
-  *Wähnert*, „Das Beweismaß des Besteuerungsverfahrens – Warum eine Konkretisierung des ‚Beanstandungsanlasses‘ aus § 158 AO unverzichtbar ist“, StBp 2018, 199.
-  *Wähnert*, „SRP-Glossar: Wichtige Begriffe zur Summarischen Risikoprüfung (SRP) und zur Neuen Prüfungstechnik (NiPt)“, Stand: Dezember 2017, downloadbar unter: <https://taxtech.blog/2018/03/09/glossar-zur-summarischen-risikoprufung-srp-und-neuen-prufungstechnik-nipt/>.
-  *Wähnert*, „Schätzungen mit Hilfe von Zeitreihenanalysen – Schrittweise Beschreibung der Quantilsschätzung“, StBp 2017, 323.
-  *Wähnert*, „Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz ‚Summarische Risikoprüfung (SRP)‘“, DB 45/2016, 2627.
-  *Wähnert*, „Engmaschige Prüfungsnetze – Prüfmethode der Finanzverwaltung“, DATEV magazin 4/2016, 15.
-  *Wähnert*, „Die Unterschiedlichkeit von Zeitreihenmodellen – Wie sehr sich Zeitreihenvarianten in der Prüfungsanwendung unterscheiden“, StBp 2016, 61.
-  *Wähnert*, „Logarithmische Normalverteilung als Prüfungsmittel für Wirtschaftsdaten – Die Logik eines leistungsfähigen Verteilungsmusters“, StBp 2016, 1.

- ☞ *Wähnert*, „Eine Antwort auf Petersens ‚wissenschaftshistorische‘ Betrachtung des Benford’s Law – Sachgerechte Maßstäbe für die Neue Prüfungstechnik“, Stbg 12/2015, 511.
- ☞ *Wähnert*, „Verteilungsbasierte Schätzungen im Steuer(straf)recht – Die Überlegenheit von Verteilungen gegenüber Kennzahlen“, StBp 2015, 92.
- ☞ *Wähnert*, „Auch digital ist legal – GoBD“, DATEV magazin 2/2015, 14.
- ☞ *Wähnert*, „Bildgebende Verfahren‘ in der (Groß- und Konzern-)Betriebsprüfung – Gläserne Massendaten für korrekte rechtliche Bewertungen“, StBp 2014, 217.
- ☞ *Wähnert*, „Der Beweiswert der Neuen interaktiven Prüfungstechnik (NiPt) – Chi-Quadrat-Test und vergleichbare Verfahren in der Betriebsprüfung“, BBK 8/2014, 380.
- ☞ *Wähnert*, „Deutung von Zeitreihenauffälligkeiten – Die Stabilität wirtschaftlicher Abhängigkeiten“, StBp 2014, 97.
- ☞ *Wähnert*, „Interaktive Außenprüfung – Möglichkeiten der gemeinsamen Sachverhaltsaufklärung im digitalen Zeitalter“, Stbg 1/2014, 20.
- ☞ *Wähnert*, „Aussagekraft und Belastbarkeit von Zeitreihenvergleichen – Erkenntnisse der Neuen interaktiven Prüfungstechnik (NiPt)“, BBK 9/2013, 420.
- ☞ *Wähnert*, „Manipulationspsychologie – Kann Manipulation dauerhaft unsichtbar bleiben?“, StBp 2013, 102.
- ☞ *Wähnert*, „Die Ziffernanalyse als Bestandteil zeitgemäßer Prüfungsansätze“, StBp 2012, 241.
- ☞ *Wähnert*, „Beweiswert neuer Verprobungsmethoden – Gegenüberstellung alter und neuer Ansätze der Betriebsprüfung“, NWB 34/2012, 2774.
- ☞ *Wähnert*, „Das große Potenzial von Strukturvergleichen in der Revision“, StC 4/2012, 18.
- ☞ *Wähnert*, „Betriebsprüfung im digitalen Zeitalter – Untersuchungsgrundsatz versus Mitwirkungsverpflichtung“, BBK 7/2012, 317.
- ☞ *Wähnert*, „Neue interaktive Prüfungstechnik (NiPt)“, Protokoll zum Bochumer Steuerseminar für Praktiker und Doktoranden vom 27. Januar 2012 (Klesen), downloadbar unter:
http://www.fachanwalt-fuer-steuerrecht.de/pdf/Protokoll_20100036.pdf.
- ☞ *Wähnert*, „Mehrfachindizien in der Betriebsprüfung – Die unterschätzte Bedeutung von Indizienketten in der Prüfungspraxis“, BBK 24/2011, 1187.
- ☞ *Wähnert*, „Die Beweislast im Besteuerungsverfahren – Brennpunkt ‚Einnahmenvollständigkeit‘“, StC 11/2011, 16.
- ☞ *Wähnert*, „Die Bedeutung von Indizienketten im Besteuerungsverfahren“, StBp 2011, 269.
- ☞ *Wähnert*, „Neue Prüfmethoden – Ausfallrisiko mindern“, DATEV magazin 3/2011, 36.
- ☞ *Wähnert*, „Beweislast und Schätzung im Zusammenhang mit modernen Verprobungsnetzen“, StBp 2011, 107.

-  *Wähnert*, „Moderne Einnahmenhinterziehung – (ein) Hauptproblem der Besteuerung“, StC 4/2010, 25.
-  *Wähnert*, „Aktuelle Fragen aus der Praxis der Außenprüfung: Digitale Manipulation – Bedrohung für das Steueraufkommen“, StBp 2010, 56.
-  *Wähnert*, „Von zeitgemäßer Risikofilterung bis zu Quantilsschätzung“, StBp 2010, 1.
-  *Wähnert*, „Integration der neuen Prüfungsmethoden in ein systematisches Risikomanagement – Einsatzmöglichkeiten neuer Prüfverfahren für die Finanzverwaltung“, BBK 15/2009, 738.
-  *Wähnert*, „Die Quantilsschätzung – Schätzungsmöglichkeiten mit summarischen Methoden“, BBK 20/2008, 1083.
-  *Wähnert*, „Anwendbarkeit, Aussagekraft und Grenzen stochastischer Manipulationstests“, StBp 2008, 312.
-  *Wähnert*, „Summarische Risikoprüfung – Echtes Risikomanagement in der Finanzverwaltung mit Hilfe einer neuen Prüfungsmethode“, StBp 2008, 296.
-  *Wähnert*, „Die betriebswirtschaftliche und statistische Analyse als Mittel der Außenprüfung“, Stbg 2007, 289.
-  *Wähnert*, „Die Anwendung von Wahrscheinlichkeitstests in der Außenprüfung und die richtige Interpretation der Ergebnisse“, StBp 2007, 65.
-  *Wähnert/Pump*, „Problemfragen der Schätzung nach § 162 AO“, Protokoll zum 140. Bochumer Steuerseminar für Praktiker und Doktoranden vom 4.3.2016 (Wilms), downloadbar unter: http://www.fachanwalt-fuer-steuerrecht.de/pdf/Protokoll_20100118.pdf.
-  *Wargowske/Greil*, „Digitale steuerliche Außenprüfung“, FR 2019, 608.
-  *Wargowske/Werner*, „Der Zugriff auf die elektronischen Daten in der steuerlichen Außenprüfung – Ein Leitfaden für die Praxis“, RET 5/2022, 71.
-  *Wargowske/Werner*, „Fortentwicklung des Datenzugriffs der Verwaltung“, beck.digitax 2/2020, 66 u. 3/2020, 133.
-  *Webel*, „Steuerfahndung – Strafverteidigung“, NWB-Verlag, 3. Auflage 2016.
-  *Webel/Danielmeyer*, „Schnittstellenverprobung elektronischer Betriebsverwaltungen – das kommende Standardprüffeld?“, StBp 2015, 353.
-  *Webel/Wähnert*, „Summarische Risikoprüfung: Einleitung von Strafverfahren in der digitalen Welt“, PStR 3/2015, 70.
-  *Webel/Wähnert*, „Der Verdacht einer Straftat während der Außenprüfung – § 10 BpO im Kontext der Digitalisierung“, NWB 44/2014, 3324.
-  *Weiskopf/Wähnert*, „Neue interaktive Prüfungstechnik – Transfereffekte moderner Verprobungsansätze auf die Großbetriebsprüfung“, BBK 3/2012, 127.

-  *Weiskopf/Wolenski*, „Prüfungsansätze moderner, interaktiver Verprobungsverfahren in der Konzern- und Großbetriebsprüfung“, StBp 2013, 33.
-  *Wenzel*, „Die Schätzung im Steuerstrafrecht – Teil 1: Grundlagen der fiskalischen Schätzung“, NWB 51/2023, 3505.
-  *Werner*, „ChatGPT im Steuerrecht oder: Hilfe, ich schaffe mich ab!“, RET 4/2023, 14.
-  *Werner/Wähnert*, „Verbildlichende Prüfung am Beispiel der Gewinnverlagerung – Die Überlegenheit der Visualisierung von Datenregeln und -verteilungen gegenüber Kennzahlen“, IStR 19/2016, 811.
-  *Wiggen*, „Die Beweiskraft des Zeitreihenvergleichs“, StBp 2008, 168.
-  *Wilke*, „Monetary Unit Sampling in der Außenprüfung - Zur Frage der verfahrensrechtlichen Zulässigkeit der Hochrechnung“, StBp 2022, 365 u. StBp 2023, 18.
-  *Wolenski*, „Das Zeitreihenurteil des X. Senats des BFH vom 25. 3. 2015“ u. „Antwort auf Kulosas Replik zur Kritik am Zeitreihenurteil des X. Senats des BFH“, Stbg 6/2016, 268 u. 272.
-  *Wolenski/Wähnert*, „Systematische Visualisierung von internationalen Gewinnverschiebungen – Beispiele für verbesserte Möglichkeiten durch Neue interaktive Prüfungstechnik (NiPt)“, IWB 3/2015, 105.
-  *Wolenski/Weiskopf*, „Praxisfall: Einsatz indirekter Prüfungsmethoden in der Großbetriebsprüfung am Beispiel der Einzelwertberichtigung“, BBK 3/2013, 128.