



POSTANSCHRIFT Bundesministerium der Finanzen, 11016 Berlin

Nur per E-Mail

Oberste Finanzbehörden
der Länder

HAUSANSCHRIFT

TEL

FAX

E-MAIL

DATUM 5. September 2023

Bundeszentralamt für Steuern

BETREFF **Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen
Außenprüfung**

ANLAGEN 1

GZ **IV D 3 - S 1445/20/10007 :006**

DOK **2023/0729678**

(bei Antwort bitte GZ und DOK angeben)

Nach Erörterung mit den obersten Finanzbehörden der Länder gilt die folgende Zusammenstellung zu den automationsgestützten quantitativen Prüfungsmethoden in der steuerlichen Außenprüfung. Eine Aufnahme dieser Zusammenstellung in die Betriebsprüfungsordnung ist nicht vorgesehen.

Dieses Schreiben wird im Bundessteuerblatt Teil I veröffentlicht. Es steht ab sofort für eine Übergangszeit auf der Internetseite des Bundesministeriums der Finanzen zum Herunterladen bereit: <http://www.bundesfinanzministerium.de> unter der Rubrik Themen - Steuern - Steuerverwaltung & Steuerrecht - Betriebsprüfung – BMF-Schreiben/Allgemeines.

Im Auftrag

Automationsgestützte quantitative Prüfungsmethoden in der steuerlichen Außenprüfung

A. Einführung	2
B. Quantitative Prüfungsmethoden.....	3
1. Zeitreihenanalyse (Zeitreihenvergleich).....	3
2. Ziffernanalyse	4
3. Struktur- und Verteilungsanalyse.....	4
4. Summarische Risikoprüfung.....	5
5. Stichprobenverfahren	6
a) Allgemeines	6
b) Monetary Unit Sampling	6
C. Schätzungsmethoden.....	7
1. Allgemeines	7
2. Zeitreihenbasierte Schätzung	7
3. Quantilsschätzung	8
4. Schätzung nach Monetary Unit Sampling (MUS).....	8

A. Einführung

- 1 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. Sofern Auffälligkeiten in den Angaben des Steuerpflichtigen auftreten, werden weitere Prüfungshandlungen veranlasst, um die Ursache zu klären. Gleichzeitig ist der Steuerpflichtige nach § 200 der Abgabenordnung bei der Sachaufklärung zur Mitwirkung verpflichtet.
- 2 Wenn betriebliche Gründe eine glaubhafte Ursache für Auffälligkeiten sind, werden diese im Rahmen der Auswertung entsprechend berücksichtigt. Sollte keine ausreichende Begründung vorliegen, liegt nach § 158 Absatz 2 Nummer 1 der Abgabenordnung ein Anlass vor, die sachliche Richtigkeit und die Buchführung des Steuerpflichtigen zu beanstanden.
- 3 Bei einer steuerlichen Außenprüfung können auch mehrere quantitative Prüfungsmethoden eingesetzt werden. Sie stehen dann nicht in einem Konkurrenzverhältnis zueinander, sondern

ergänzen sich idealerweise gegenseitig und werden durch systematisch visualisierende Betrachtungen unterstützt. Der (kombinierte) Einsatz (mehrerer) quantitativer Prüfungsmethoden hängt von der Datenlage und dem Analyseziel der jeweiligen steuerlichen Außenprüfung ab.

- 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. Quantitative Prüfungsmethoden sind von der Schätzung (siehe C.) zu unterscheiden.

Die folgende Darstellung gibt einen Überblick über automationsgestützte quantitative Prüfungs- und Schätzungsmethoden. Sie ist jedoch nicht abschließend.

B. Quantitative Prüfungsmethoden

1. Zeitreihenanalyse (Zeitreihenvergleich)

- 5 Eine Zeitreihenanalyse (auch Zeitreihenvergleich genannt) ist eine quantitative Prüfungsmethode, die dazu dient, historische Entwicklungen und Zusammenhänge von betriebswirtschaftlichen Kennzahlen darzustellen. Bei einer Zeitreihenanalyse werden die Daten in einem Diagramm periodenweise entlang einer Zeitachse aufgetragen. Es gibt verschiedene Varianten der Zeitreihenanalyse, je nachdem, welches Ziel verfolgt wird (vgl. BFH-Urteil vom 25. März 2015 – X R 20/13, BStBl. II 2015, 743). Es können beispielsweise Trends, Schwankungen und Ausreißer in Datenreihen identifiziert oder die erwartete Abhängigkeit (Korrelation) von verschiedenen Datenreihen miteinander überprüft werden.
- 6 Die Visualisierung der Daten in einem Diagramm hat den Vorteil, dass Entwicklungen und Muster besser als in einer Tabelle erkennbar sind. Auch Ausreißer lassen sich hierdurch wirksam identifizieren. Bei stark voneinander abweichenden Werten kann es sinnvoll sein, eine Größenanpassung des Datenbereichs vorzunehmen, zum Beispiel durch eine logarithmische Skalierung. Die Wahl der Vergleichsperiode kann einen großen Einfluss darauf haben, welche Zusammenhänge oder Einzelheiten erkennbar sind.
- 7 Die Kombination aus Diagrammen und Tabellen kann ein aussagekräftiges Gesamtbild der betriebswirtschaftlichen Daten des Steuerpflichtigen ergeben. Auf diese Weise lassen sich Abweichungen von den erwarteten Werten als Differenzen und Zeitversätze sichtbar machen.
- 8 Zur Beurteilung einer erwarteten Abhängigkeit zweier Datenreihen mit stark abweichenden Werten gibt es verschiedene Möglichkeiten, um die Vergleichbarkeit der Datenreihen zu verbessern, zum Beispiel durch größenanpassende Doppelskalierungen. Der Korrelationskoeffizient ist ein statistisches Maß, das dabei hilft, die erwartete

Abhängigkeit zwischen zwei Datenreihen zu bewerten. Ausreißer können zum Beispiel mit geeigneten Zeitreihenvarianten wie den Gleitschlittenansätzen (überlappende Zeitreihenperioden wie „01.01. bis 31.01.“, „02.01. bis 01.02.“, „03.01. bis 02.02.“ etc.) gezielt untersucht und korrigiert werden, um die Daten für die eigentliche Beurteilung der Schlüssigkeit anzupassen.

2. Ziffernanalyse

- 9 Eine Ziffernanalyse untersucht die Zahlenstruktur der zu prüfenden Daten, insbesondere von Bargeldumsätzen. Ziel ist es, durch die Analyse von Ziffernhäufigkeiten Manipulationen aufzudecken, die durch das Erfinden von Zahlen in manipulationsgefährdeten Daten entstehen können, beispielsweise durch Bilanzfälschungen oder „Schattenaufzeichnungen“.
- 10 Für die Ziffernanalyse wird die Häufigkeit der möglichen Ziffern oder Ziffernkombinationen ermittelt, wobei zu unterscheiden ist zwischen der Betrachtung der Zahlen von vorn (Erst-, Zweit-, Drittziffer etc.) und bei Geldbeträgen der Betrachtung der Zahlen von hinten (Ein-Euro-Stelle, Zehn-Euro-Stelle etc.). Empirischen Forschungen zur Folge weisen die „Erstziffer“ und „Zweitziffer“ in Wirtschaftsdaten eine Häufigkeitsverteilung auf, die deutlich von der zufälligen Gleichverteilung abweicht und sich in einer einheitlichen Funktion beschreiben lässt („Benford-Gesetz“ oder „Newcomb-Benford-Law“). Durch die Überprüfung der ausgewerteten Ziffernverteilungen kann aufgedeckt werden, ob einzelne Ziffern zu selten oder zu häufig vorkommen, was ein Hinweis auf Manipulation der Daten sein kann durch unbewusste Vorlieben oder Abneigungen des Steuerpflichtigen bei der Verwendung einzelner Ziffern. Die Übereinstimmung von beobachteten und erwarteten Häufigkeitsverteilungen kann mit Hilfe des Wahrscheinlichkeitstests „Chi-Quadrat-Anpassungstest“ (gesamte Ziffernverteilung) oder der „Binomialverteilung“ (einzelne Ziffern) bewertet werden.
- 11 Ziffernanalysen sind nur begrenzt einsetzbar und zielen ausschließlich darauf ab, Manipulationen durch das freie Erfinden von Zahlen aufzudecken. Einzelbetriebliche Besonderheiten lassen sich durch einfache mathematische Überlegungen oder Struktur- und Verteilungsanalysen berücksichtigen beziehungsweise ausschließen. Die Nachverfolgung von zu selten oder zu häufig auftretenden Ziffern in Teildatenbeständen erhöht die Wahrscheinlichkeit, Manipulationen in Teilen des Gesamtdatenbestandes aufzudecken.

3. Struktur- und Verteilungsanalyse

- 12 Die Struktur- und Verteilungsanalyse ist ein Verfahren zur Untersuchung von Daten, bei dem die Daten nach Klassen sortiert und ausgezählt werden, um die aufgetretene Verteilung mit der erwarteten Verteilung zu vergleichen. Ziel ist es, Unregelmäßigkeiten zu entdecken, die auf mögliche Fehler oder Manipulationen hinweisen.

Durch die Struktur- und Verteilungsanalyse werden vergleichbare Daten nach gleichmäßigen Werteklassen ausgewertet und als Diagramm dargestellt. Für stark asymmetrische Verteilungen werden logarithmisch gleichmäßig definierte Werteklassen verwendet. Die erwartete Häufigkeitsverteilung wird aus der Anzahl der Daten, dem Mittelwert und der Standardabweichung abgeleitet. Dabei lassen sich zum Beispiel durch die vorherige Aufteilung der Daten in Teildatenbestände betriebliche Besonderheiten ausschließen. Die Verfolgung von auffälligen Werteverteilungen in den Teilen des Gesamtdatenbestandes (etwa Wirtschaftsjahre, Haupt- und Nebensaisons oder Wochentage) erhöht die Wahrscheinlichkeit, partielle Manipulationen aufzudecken, und kann bei der Unterscheidung zwischen Manipulationen und Unterschlagungen helfen.

4. Summarische Risikoprüfung

- 14** Die Summarische Risikoprüfung (SRP) ist ein System von quantitativen Prüfungsmethoden, das verschiedene einzelne quantitative Prüfungsmethoden in systematischen Prüfungen miteinander verknüpft und technisch durch IT-gestützte Vorlagen unterstützt. Die SRP konzentriert sich besonders auf die Methodenbereiche der Zeitreihenanalysen, der Ziffernanalyse und der Struktur- und Verteilungsanalyse, die in diversen Varianten und Ausführungen zu einem mehrperspektivischen Prüfungskonzept zusammengeführt werden.
- 15** Der Prüfungspfad wird systematisch von einer groben Übersicht bis zu auffälligen Einzelsachverhalten verfolgt. Hierbei werden in der Regel statistisch-visualisierende Analysearten genutzt. Bei der Zeitreihenanalyse wird die Periodisierung schrittweise zu kürzeren, detailreicheren Intervallen verfeinert, wofür gezielt Varianten dieser Analysemethoden eingesetzt werden. In den Häufigkeitsauswertungen wie den Ziffernanalysen und den Struktur- und Verteilungsanalysen wird der Prüfungspfad über die Einstellung der Klassifizierung sowie die Dateneingrenzung auf Teilzeiträume verfolgt. Chronologische und verteilungsbezogene Teilergebnisse ergänzen sich in der Regel. Dadurch lassen sich Beobachtungen besser und einfacher beschreiben und Auffälligkeiten konkretisieren beziehungsweise eventuell bereits klären.
- 16** Durch das Zusammenführen der verschiedenen Perspektiven und der sich ergänzenden Beobachtungen und durch die technischen Möglichkeiten der SRP-Vorlagen können einzelfallbezogene Prüffelder festgestellt und eingegrenzt werden. Die Vorlagen fördern dabei den geeigneten Einsatz der einzelnen quantitativen Prüfungsmethoden, einen nachvollziehbaren Prüfungspfad sowie die Berücksichtigung neuer Erkenntnisse aus der Prüfung oder der Mitwirkung durch den Steuerpflichtigen. Auch die Ergebnisdokumentation wird durch die SRP-Vorlagen unterstützt.
- 17** Die in der SRP enthaltenen Berechnungsschritte können dem öffentlich zugänglichen SRP-Handbuch entnommen sowie mithilfe von gängigen Tabellenkalkulationsprogrammen nachvollzogen werden.

5. Stichprobenverfahren

a) Allgemeines

- 18 Stichprobenverfahren sind statistische Methoden, die genutzt werden, um die Richtigkeit oder die Plausibilität von Daten zu überprüfen. Dabei wird eine zufällig ausgewählte Teilmenge eines Datenbestandes analysiert, um Rückschlüsse auf den Gesamtdatenbestand zu ziehen. Statistische Stichprobenverfahren sind objektiver als bewusste Auswahlprüfungen.

b) Monetary Unit Sampling

- 19 Monetary Unit Sampling (MUS) ist eine Methode, die auf der Grundlage mathematisch-statistischer Auswertung einer qualifizierten Zufallsstichprobe eines Prüffeldes Rückschlüsse auf die Grundgesamtheit zulässt. Im ersten Schritt erfolgt eine repräsentative Überprüfung des Prüffeldes. Je nach Ergebnis dieser Überprüfung kann in einem zweiten Schritt die Korrektur des Prüffeldes mittels MUS erfolgen (siehe C.4). Hierbei wird die Gesamtheit aller Elemente in einem Prüffeld in Geldeinheiten ausgedrückt. Zum Beispiel: 5.000 Rechnungen zu einem Gesamtbetrag in Höhe von 1.000.000 € führen zu einem Prüffeld bestehend aus 1.000.000 Geldeinheiten.
- 20 Die Anzahl der zu untersuchenden Elemente aus der Gesamtheit wird unabhängig von der Anzahl der in der Gesamtheit enthaltenen Elemente durch das Konfidenzniveau und durch die Wesentlichkeitsgrenze bestimmt. Das Konfidenzniveau gibt an, mit welcher Wahrscheinlichkeit Fehler in der Gesamtheit entdeckt werden, wenn in der Grundgesamtheit Fehler oberhalb der Wesentlichkeitsgrenze überschritten werden. Die Wesentlichkeitsgrenze gibt den Fehlerumfang an, der nicht übersehen werden soll. Die Festlegung der beiden Parameter zur Stichprobenziehung ist eine Abwägung zwischen dem Risiko, Fehler zu übersehen, und dem Bestreben, den Stichprobenumfang in einer realistischen Größe zu halten. Damit ist jedoch keine Akzeptanz einer Fehlerquote unterhalb der Wesentlichkeitsgrenze verbunden.
- 21 MUS wird vor allem bei großen Datenmengen und einer geringen Fehlererwartung eingesetzt. Exemplarisch anzuführen wären hier Bereiche wie
- der Vorsteuerabzug,
 - die Steuerfreistellung von Umsätzen,
 - die Bilanzierung von Rückstellungen und
 - die Abgrenzung von sofort abzugsfähigen Instandhaltungsaufwendungen zu aktivierungspflichtigen Aufwendungen.

In diesen Bereichen kann eine Überbewertung der Fehler eintreten, wenn höhere Werte verbucht werden, als zulässig sind. MUS nutzt eine Hypothese zum Prüffeld, auf deren Basis eine wertproportionale Stichprobe gezogen und anschließend analysiert wird. Im Rahmen der Analyse ist gegebenenfalls eine weitere Sachaufklärung durchzuführen. Erst im Anschluss können die Ergebnisse ausgewertet und die weiteren Schlussfolgerungen daraus gezogen werden.

- 23** Geschäftsvorfälle gelangen mittelbar über ihre Geldeinheiten in die Stichprobenziehung. Kommt es bei einer Stichprobe zu keiner Beanstandung, so kann mit der Aussagesicherheit des Konfidenzniveaus festgestellt werden, dass in der Grundgesamtheit keine Fehlerquote oberhalb der Wesentlichkeitsgrenze zu erwarten ist. Das Prüffeld kann in der Folge ohne weitere Beanstandungen akzeptiert werden. Kommt es bei einer Stichprobe zu einer Beanstandung, so kann mit der Aussagesicherheit des Konfidenzniveaus festgestellt werden, dass in der Grundgesamtheit eine nicht tolerierbare Fehlerquote oberhalb der Wesentlichkeitsgrenze vorhanden ist. Das Prüffeld kann in der Folge nicht ohne weitere Aufklärungsmaßnahmen als ordnungsgemäß bewertet werden. MUS bietet für Fälle, in denen das Prüffeld als nicht ordnungsgemäß zu bewerten ist, die Möglichkeit, in einem weiteren Schritt eine Korrektur des Prüffeldes durch Schätzung durchzuführen (siehe C.4).

C. Schätzungsmethoden

1. Allgemeines

- 24** An die Verprobung mit quantitativen Prüfungsmethoden kann sich gemäß § 162 der Abgabenordnung eine Schätzung anschließen. Besteht dem Grunde nach die Notwendigkeit, eine Schätzung durchzuführen, so können quantitative Prüfungsmethoden zur Auswahl der geeigneten Schätzungsmethode herangezogen werden.
- 25** Die folgenden Schätzungsmethoden dienen dazu, eine möglichst genaue Schätzung bei einer Außenprüfung durchzuführen. Die Wahl der geeigneten Schätzungsmethode hängt vom jeweiligen Einzelfall ab. Der Steuerpflichtige hat keinen Anspruch auf eine bestimmte Schätzungsmethode. Die Auswahl der Schätzungsmethode liegt im pflichtgemäßen Ermessen der Finanzbehörde.
- Die folgende Zusammenstellung der Schätzungsmethoden ist nicht abschließend.

2. Zeitreihenbasierte Schätzung

- 26** Die zeitreihenbasierte Schätzung nutzt Ergebnisse von Zeitreihenanalysen als Grundlage, um die Schätzungshöhe festzulegen. Diese Schätzungsmethode basiert auf betriebsinternen Daten und ist deshalb externen Ansätzen vorzuziehen. Bei der zeitreihenbasierten Schätzung können betriebliche Besonderheiten berücksichtigt und

verbleibende Unsicherheiten in einer Sensibilitätsanalyse, unter Umständen mit Sicherheitsabschlägen, berücksichtigt werden.

3. Quantilsschätzung

- 27 Die Quantilsschätzung ist eine Schätzungsmethode zur internen Ermittlung einer sachgerechten Schätzungshöhe mit Hilfe von betriebswirtschaftlichen und statistischen Standardwerkzeugen. Dazu werden Prozentränge (Quantile) als Ausgangswerte genutzt.
- 28 Die monatlichen Werte für den Rohgewinnaufschlagssatz beziehungsweise den Wareneinsatz werden als Grundlage verwendet. Für die Ermittlung einer sachgerechten Schätzungshöhe können in den Datenbestand alle konkreten Erkenntnisse eingepflegt werden, die durch Prüfungsmaßnahmen beziehungsweise durch die Mitwirkung des Steuerpflichtigen gewonnenen worden sind.
- 29 Vor der Feststellung des Schätzungsrahmens werden nicht repräsentative Ausreißer großzügig aus den Grundlagenwerten ausgeschlossen (Bandbreiteneinengung). Es werden Ranggrößen verwendet, um eine Verzerrung durch Extremwerte zu vermeiden. Zur verallgemeinerbaren Einschränkung der Ausgangsgrößen auf „Normalwerte“ werden die Rang-Wert-Verhältnisse der Normalverteilungen als typisches zufälliges Streuverhalten herangezogen. Der Schätzungsansatz wird mit dem 80%-Quantil in dem eingengten Schätzungsrahmen der internen Normalwerte „16%- bis 84%-Quantil“ an der oberen Grenze (BFH-Urteil vom 9. März 1967 – IV 184/63, BStBl. III 1967, 349) angesetzt. Je nach Sachlage kann dieses verallgemeinerbare Schätzungsmaß auf die monatlichen Rohgewinnaufschlagssätze, auf die Wareneinsatzsummen oder auf beide Ausgangsgrößen angewandt werden.

4. Schätzung nach Monetary Unit Sampling

- 30 Wenn bei der Überprüfung einer Stichprobe in den Daten (materielle) Mängel festgestellt werden, kann die ursprüngliche Hypothese zum Prüffeld nicht mehr aufrechterhalten und das Prüffeld nicht als ordnungsgemäß anerkannt werden. Wenn die im Prüffeld vorhandenen Fehler nicht vollständig identifiziert werden können und eine Behebung der Fehler nicht möglich ist, muss der geschätzte Betrag der Fehler berechnet werden. MUS bietet hierzu die Möglichkeit, den Anteil der Fehler in der Stichprobe auf die Gesamtheit zu extrapolieren.