

Der digitale Detektiv: Wie Statistik helfen kann, Kassenbetrügern auf die Schliche zu kommen

Eine praxisrelevante Anwendung im Rahmen der Betriebsprüfung

Die Aufdeckung von Steuerhinterziehung in manipulationsanfälligen Kassendaten ist eine große Herausforderung. Dieser Artikel stellt den Chi-Quadrat-Anpassungstest als statistisches Prüfverfahren vor, mit dessen Hilfe Auffälligkeiten in Kassendaten identifiziert werden können. Die Methode legt den Grundstein für die Risikobewertung und macht Steuerprüfungen effizienter. Der Artikel erörtert die theoretischen Grundlagen, den rechtlichen Rahmen, detaillierte Anwendungsbeispiele und methodische Einschränkungen dieses praxisrelevanten Ansatzes.



Marcel Steinborn

ist Lehrbeauftragter für Wirtschaftsmathematik und Statistik. Bevorzugte Forschungsgebiete: Empirische Finanzmarktanalyse mit Copulas sowie Einsatz statistischer Verfahren in der Risikoprüfung.

Summary: The detection of tax evasion in manipulation-prone cash register data is a major challenge. This article introduces the Chi-Square goodness-of-fit test as a statistical procedure to identify anomalies in cash register data. The method lays the groundwork for risk assessment and makes tax audits more efficient. The article discusses the theoretical foundations, legal framework, detailed application examples, and methodological limitations of this practical approach.

Stichwörter: Datenforensik, Kassenprüfung, Statistische Tests, Steuerhinterziehung

1. Relevanz digitaler Prüfungsstrategien

Die Sicherstellung der **Steuerehrlichkeit** ist eine zentrale Aufgabe der modernen Finanzüberwachung. Während die enormen Datenmengen in heutigen Unternehmenssystemen prinzipiell neue analytische Möglichkeiten eröffnen, erschweren sie gleichzeitig eine manuelle Prüfung der

Transaktionsintegrität erheblich und verstärken damit die Notwendigkeit effizienter **digitaler Prüfstrategien**. Besonders in diesem Kontext gewinnen Kassendaten an Bedeutung. Trotz der zunehmenden **Digitalisierung** und verbesserter Sicherheitsstandards stellen gezielte Manipulationen von Kassendaten weiterhin eine Herausforderung für die Finanzverwaltung dar, was die Notwendigkeit von mathematisch-statistischen Prüfverfahren begründet (vgl. *Wähnert*, 2012, S. 241).

Dieser Artikel analysiert, wie der Chi-Quadrat-Anpassungstest als statistisches Werkzeug der Ziffernanalyse in der digitalen Betriebsprüfung und insbesondere im Rahmen der Summarischen **Risikoprüfung** (SRP) eingesetzt werden kann (vgl. *Wähnert*, 2009, S. 341 f.). Diese Methode ermöglicht eine fundierte Prüfung unterschiedlichster Datensätze. Dabei wird überprüft, ob signifikante Abweichungen von erwarteten Ziffernverteilungen vorliegen, die auf Unregelmäßigkeiten in Kassensystemen hindeuten könnten. Das Verständnis dieses statistischen Verfahrens ermöglicht nicht nur eine **praxisrelevante** Anwendung im steuerlichen Kontext, sondern ist zudem ein entscheidender Schritt zur Entwicklung unerlässlicher Kernkompetenzen in der modernen Wirtschaftsprüfung und Finanzanalyse. Ziel dieses Beitrags ist es, den Wert der Ziffernanalyse und speziell des Chi-Quadrat-Anpassungstests für die moderne Betriebsprüfung am Beispiel der Überprüfung von Kassensystemen zu unterstreichen. Durch konkrete Fallbeispiele soll die Relevanz für die Identifikation potenzieller Unregelmäßigkeiten greifbar gemacht werden.

2. Grundlagen und Einbezug statistischer Verfahren in der Summarischen Risikoprüfung

2.1. Erwartete Verteilungen von Kassendaten in der Ziffernanalyse

Spezielle Verteilungen in natürlichen Datensätzen können für Analysezwecke genutzt werden, was auch auf Kassendaten zutrifft. Insbesondere die Benford-Verteilung und die Gleichverteilung sind hier von Bedeutung. Beide dienen als Indikatoren für die Konformität der Daten.

Die Benford-Verteilung stellt ein statistisches Phänomen dar, welches bei **führenden** Ziffern in natürlichen Datenbeständen, insbesondere Kassendaten, auftritt. Es besagt, dass kleinere Ziffern (z.B. 1 oder 2) als führende Ziffern häufiger vertreten sind als größere Ziffern (z.B. 8 oder 9). Die Ziffer 1 ist demnach die am häufigsten erwartete führende Ziffer, gefolgt von der 2, bis hin zur 9 als seltenste (vgl. Tödter, 2007, S. 93 f.). Für jede Ziffer i (als Kategorie) wird die theoretische Wahrscheinlichkeit nach folgender Formel berechnet (vgl. Benford, 1938, S. 554):

$$p_i = \log_{10}\left(1 + \frac{1}{i}\right) \quad (1)$$

Im Gegensatz zur Benford-Verteilung bei der führenden Ziffer in natürlichen Datensätzen beschreibt die Gleichverteilung das erwartete Muster der letzten Ziffer der Daten. Im Bereich von 0 bis 9 zeigt die Verteilung der **letzten** Ziffer eine annähernde Gleichverteilung. Dies bedeutet, dass jede Ziffer mit einer nahezu identischen Häufigkeit auftritt. Dieses Verhalten ist statistisch zu erwarten, wenn keine spezifischen Einflussfaktoren die Verteilung verzerren (bspw. **Manipulationen**, Preisgestaltung).

Beide Verteilungsmuster sind in Abb. 1 beispielhaft dargestellt: Die Benford-Verteilung der führenden Ziffern ist hier durch den schwarzen Balken gekennzeichnet, während der graue Balken die Gleichverteilung der letzten Ziffern illustriert.

2.2. Identifikation von Auffälligkeiten in Kassendaten im Rahmen der Summarischen Risikoprüfung

Die erwartete Verteilung der Kassendaten kann im Rahmen der SRP genutzt werden. In diesem Zusammenhang kommen **statistische** Verfahren zum Einsatz, um potenzielle Steuerumgehungen durch systematische Mustererkennung zu identifizieren. Als digitale Detektive ermöglichen sie das Aufspüren von Abweichungen gegenüber erwarteten Verteilungen und damit eine zielgerichtete Prüfung zur Aufdeckung potenzieller Manipulationen (vgl. Wähnert, 2012, S. 241).

Die **Relevanz** statistischer Verfahren, wie des Chi-Quadrat-Anpassungstests und anderer Zifferanalysen, wird sowohl durch die Rechtsprechung als auch die Finanzverwaltung untermauert. Als anerkannte Prüfungsinstrumente dienen sie primär als Indizien und Warnsignale für mögliche Manipulationen oder Mängel in der Buchführung. Die Finanzgerichtsbarkeit hat die Zulässigkeit bestätigt und ihre Indizwirkung hervorgehoben (vgl. FG Münster, 2002; FG Rheinland-Pfalz, 2011, S. 10). Dabei wird betont, dass sie in der Regel keine alleinige Grundlage für eine Schätzung der Besteuerungsgrundlagen darstellen, sondern stets im Kontext weiterer festgestellter Mängel zu betrachten sind. Das BMF-Schreiben vom 5. September 2023 bekräftigt diese Haltung und verankert quantitative Prüfungsmethoden explizit als Werkzeug für die Außenprüfung (vgl. BMF-Schreiben v.

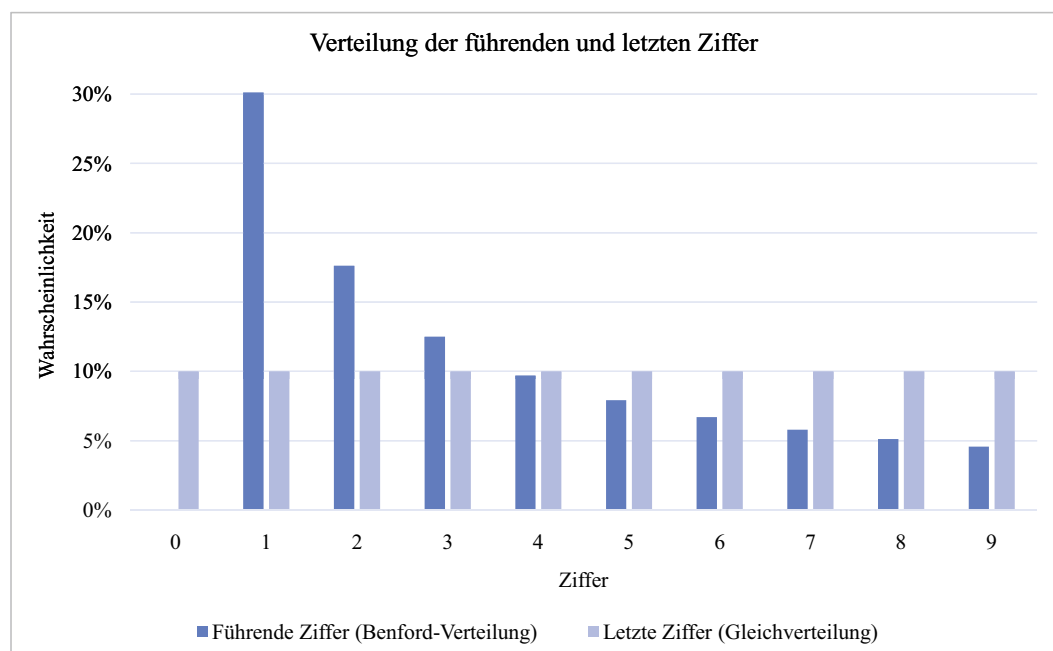


Abb. 1: Verteilung der führenden und letzten Ziffer

5.9.2023, BStBl. I S. 1546). Diese rechtliche Einschätzung verdeutlicht die große **Bedeutung** des Chi-Quadrat-Anpassungstests und anderer Ziffernanalysen bei der Überprüfung digitaler Finanzaufzeichnungen.

2.3. Verwendung des Chi-Quadrat-Anpassungstests

Als quantitative Prüfungsmethode kommt im Folgenden der Chi-Quadrat-Anpassungstest zum Einsatz. Dieses **fundamentale** statistische Werkzeug wird verwendet, um die Übereinstimmung zwischen einer beobachteten und einer erwarteten theoretischen Verteilung zu überprüfen (vgl. *Sosna*, 2004, S. 252 f.; *Wähnert*, 2009, S. 341 f.). Im Kontext der Betriebsprüfung dient er insbesondere dazu, Abweichungen von der Benford-Verteilung bei führenden Ziffern oder von der Gleichverteilung bei letzten Ziffern in Kassendaten zu identifizieren. Derartige Abweichungen können Hinweise auf systematische Anpassungen oder Manipulationen liefern, bspw. durch Auf- oder Abrundungen (vgl. *Sosna*, 2004, S. 250 f.). Für diesen Test werden, wie allgemein bei statistischen Prüfverfahren, zwei Hypothesen formuliert: Die Nullhypothese (H_0) besagt, dass die beobachtete Verteilung der erwarteten Verteilung entspricht und etwaige Unterschiede rein zufällig sind. Die Alternativhypothese (H_1) postuliert hingegen eine signifikante Abweichung von der erwarteten Verteilung (vgl. *Sosna*, 2004, S. 251).

Zur Quantifizierung dieser Abweichung wird die Chi-Quadrat-Teststatistik (χ^2) berechnet. Die Formel hierfür lautet:

$$\chi^2 = \sum_{i=1}^k \frac{(B_i - E_i)^2}{E_i} \quad (2)$$

Hierbei steht k für die Anzahl der Kategorien (bspw. Anzahl der Ziffern), B_i repräsentiert die beobachtete Häufigkeit in der jeweiligen Kategorie i , und E_i bezeichnet die erwartete Häufigkeit der Kategorie i . Die erwartete Häufigkeit berechnet sich aus der theoretischen Wahrscheinlichkeit (p_i) für die Kategorie i , multipliziert mit der Gesamtzahl der Beobachtungen (N).

Die **Interpretation** des Ergebnisses erfolgt durch den Vergleich des berechneten χ^2 -Wertes mit einem kritischen Wert aus einer Chi-Quadrat-Verteilungstabelle oder durch die Bestimmung des p-Wertes. Ist der berechnete χ^2 -Wert größer als der kritische Wert, oder ist der p-Wert kleiner als das festgelegte Signifikanzniveau α (typischerweise 5 %), wird die Nullhypothese abgelehnt. Dies bedeutet, dass die beobachtete und erwartete Verteilung mit hoher Wahrscheinlichkeit voneinander abweichen und die Unterschiede als nicht zufällig angesehen werden können. Im Kontext der Betriebsprüfung wird eine solche signifikante Abweichung als ein Indiz für Unregelmäßigkeiten oder gar Manipulationen interpretiert.

2.4. Einschränkungen des Tests

Die Anwendung des Chi-Quadrat-Anpassungstests in der **Ziffernanalyse** von Kassendaten erfordert die Berücksichtigung mehrerer wesentlicher Einschränkungen, um valide Ergebnisse und korrekte Interpretationen sicherzustellen:

- **Stichprobengröße und Verteilungsannahmen:** Eine ausreichend große Stichprobe ist für verlässliche Testergebnisse unerlässlich; zu kleine Stichproben können verzerrte Resultate liefern. Ebenso setzt der Chi-Quadrat-Anpassungstest voraus, dass die erwartete Verteilung korrekt spezifiziert ist, da inkorrekte Annahmen (z. B. über die Verteilung) zu fehlerhaften Schlussfolgerungen führen. Darüber hinaus ist die Unabhängigkeit der Beobachtungen entscheidend; das heißt, einzelne Kassentransaktionen dürfen sich nicht gegenseitig beeinflussen (vgl. *Wähnert*, 2008, S. 314–315).
- **Interpretation der Ergebnisse:** Es ist wichtig, die Ergebnisse **kritisch** zu hinterfragen. Nicht jede signifikante Abweichung bedeutet zwangsläufig Manipulation. Sie kann auch durch legitime Geschäftsprozesse, komplexe Markt- und Preisstrukturen oder systematische Rundungsmechanismen der Kassensoftware verursacht werden (vgl. *Tödter*, 2007, S. 96; *Wähnert*, 2008, S. 315).

Diese Punkte unterstreichen die **Bedeutung** eines ganzheitlichen Ansatzes bei der Prüfung von Kassendaten mittels statistischer Methoden. Der Chi-Quadrat-Anpassungstest ist ein initialer Ansatz in der Ziffernanalyse. Um jedoch aussagekräftiges Feedback für Steuerprüfungen zu liefern, muss er stets im Kontext einer umfassenden Daten- und Geschäftsanalyse betrachtet werden.

3. Anwendung auf Kassendaten

3.1. Detaillierte Praxisbeispiele

Um die Anwendung des Chi-Quadrat-Anpassungstests zur Überprüfung von Kassendaten genauer zu **veranschaulichen**, folgen hier detaillierte Beispiele. Da reale Kassendaten aus Datenschutzgründen schwer zugänglich sind, wurde ein realistischer, synthetischer Datensatz von Tageseinnahmen erstellt.

a) Beispiel 1: Test der führenden Ziffern von Kassendaten

Ausgangssituation: Das Unternehmen analysiert die führenden Ziffern von 1.000 Transaktionsbeträgen (siehe *Tab. 1*). Eine Benford-Verteilung würde zeigen, dass kleinere Ziffern häufiger als größere Ziffern auftreten.

Um die Analyse durchzuführen, wird die Teststatistik für den Chi-Quadrat-Anpassungstest aus Formel (2) und die Berechnung der Wahrscheinlichkeit der Ziffer i mit Hilfe der Benford-Verteilung aus Formel (1) benötigt. Beispiels-

Tab. 1: Verteilung der beobachteten Häufigkeiten je führender Ziffer von Transaktionsbeträgen

Führende Ziffer	1	2	3	4	5	6	7	8	9
Beobachtete Häufigkeit	250	160	150	90	80	70	60	70	70

weise berechnet sich der Beitrag zur **Teststatistik** für die Ziffer 1 durch

$$\frac{(B_1 - E_1)^2}{E_1} = \frac{\left(250 - \log_{10}\left(1 + \frac{1}{1}\right) \cdot 1000\right)^2}{\log_{10}\left(1 + \frac{1}{1}\right) \cdot 1000} = 8,65.$$

Analog wird für die weiteren Ziffern vorgegangen (siehe Tab. 2).

Die Addition der Beiträge ergibt die Teststatistik von 35,64.

b) Beispiel 2: Test der Endziffern von Kassendaten

Ausgangssituation: Ein Unternehmen führt 1.000 Transaktionen durch (siehe Tab. 3). Um einen möglichen **Manipulationsverdacht** zu überprüfen, wird die Verteilung der letzten Ziffer jedes Transaktionsbetrags analysiert. Theoretisch sollte jede Ziffer (0 bis 9) gleich häufig vertreten sein. Für die Endziffern vereinfacht sich die Bestimmung der Teststatistik des Chi-Quadrat-Anpassungstests aus Formel (2) erheblich. Die Wahrscheinlichkeit der Ziffer i beträgt jeweils 10 %. Bspw. berechnet sich der Beitrag zur Teststatistik für die Ziffer 0 durch

$$\frac{(B_1 - E_1)^2}{E_1} = \frac{\left(95 - \frac{1}{10} \cdot 1000\right)^2}{\frac{1}{10} \cdot 1000} = 0,25.$$

Das Vorgehen ist für die weiteren Ziffern analog (siehe Tab. 4). Die Beiträge summieren sich zur Teststatistik von 9,50.

3.2. Interpretation der Testergebnisse: Was die Zahlen bedeuten

Um die Hypothesen zu überprüfen, muss die berechnete Teststatistik mit dem entsprechenden **Quantil** der Chi-Quadrat-Verteilung (kritischer Wert) verglichen werden. Für ein gängiges Signifikanzniveau α von 5 % ergeben sich

spezifische kritische Werte, die von den **Freiheitsgraden** (df) abhängen. Diese ergeben sich aus $k-1$, wobei k die Anzahl der Kategorien (z. B. der Ziffern) darstellt.

Im ersten Beispiel, bei dem die führenden Ziffern von Kassendaten hinsichtlich der Benford-Verteilung untersucht werden, basieren die Berechnungen auf neun Ziffernkategorien (1 bis 9). Dies führt zu $df = 9 - 1 = 8$ Freiheitsgraden. Der zugehörige kritische Wert für α von 5 % beträgt 15,507. Die Teststatistik für dieses Beispiel ergibt einen Wert von 35,64 (siehe Tab. 2). Dieser **übersteigt** den kritischen Wert von 15,50. Das führt zur Ablehnung der Nullhypothese, was auf eine signifikante Abweichung der beobachteten Verteilung von der erwarteten Benford-Verteilung hindeutet.

Für das zweite Beispiel, bei dem die letzten Ziffern von Kassendaten hinsichtlich der Gleichverteilung überprüft werden, liegen zehn Ziffernkategorien (0 bis 9) vor. Dies resultiert in $df = 10 - 1 = 9$ Freiheitsgraden. Der entsprechende kritische Wert für α von 5 % beläuft sich auf 16,919. Die Teststatistik ergibt 9,50 (siehe Tab. 4). Da diese den kritischen Wert von 16,919 nicht übersteigt, konnte die Nullhypothese nicht widerlegt werden. Die beobachteten Häufigkeiten der Endziffern zeigen somit keine signifikante Abweichung von der Annahme einer Gleichverteilung und können mit hoher **Wahrscheinlichkeit** als zufällig erachtet werden.

3.3. Fazit und Handlungsempfehlungen für die Prüfungspraxis

Diese konkreten Anwendungsbeispiele veranschaulichen, wie der Chi-Quadrat-Anpassungstest zur Plausibilitätsprüfung von Kassendaten eingesetzt werden kann.

Die Ergebnisse zeigen den unterschiedlichen Aussagegehalt des Tests: Im ersten Fall überstieg die Teststatistik bei der Benford-Analyse den kritischen Wert deutlich, was auf eine **signifikante** Abweichung und damit auf mögliche Unregelmäßigkeiten hindeutet. Solche Diskrepanzen stellen

Führende Ziffer	1	2	3	4	5	6	7	8	9	Summe
Beitrag zur Teststatistik	8,65	1,47	5,03	0,49	0,01	0,14	0,07	6,94	12,84	35,64

Tab. 2: Beitrag zur Teststatistik je führender Ziffer von Transaktionsbeträgen (auf zwei Nachkommastellen gerundet)

Tab. 3: Verteilung der beobachteten Häufigkeiten je letzter Ziffer von Transaktionsbeträgen

Letzte Ziffer	0	1	2	3	4	5	6	7	8	9
Beobachtete Häufigkeit	95	110	85	105	95	105	110	90	115	90

Letzte Ziffer	0	1	2	3	4	5	6	7	8	9	Summe
Beitrag zur Teststatistik	0,25	1,00	2,25	0,25	0,25	0,25	1,00	1,00	2,25	1,00	9,50

Tab. 4: Beitrag zur Teststatistik je letzter Ziffer von Transaktionsbeträgen

jedoch keinen unmittelbaren Beleg dar, sondern erfordern stets eine vertiefte Untersuchung zur Klärung der Ursachen (vgl. Wähnert, 2008, S. 315; Wähnert, 2012, S. 244).

Im zweiten Fall hingegen, der Analyse der Endziffern auf Gleichverteilung, überstieg die Teststatistik den kritischen Wert nicht. Folglich konnte keine signifikante Abweichung festgestellt werden. Für eine **vertiefte** Überprüfung der Kassendaten liegen somit keine hinreichenden Indizien vor.

4. Zusammenfassung

Dieser Beitrag untersucht die Anwendung des Chi-Quadrat-Anpassungstests bei der Integritätsprüfung von Kassendaten in der modernen Finanzüberwachung. Insbesondere liegt der Fokus auf dessen Verwendung in Verbindung mit der Benford-Verteilung sowie der Annahme einer Gleichverteilung. Dieser statistische Test ermöglicht eine zielgerichtete Identifikation von Ziffernverteilungen, die auf Manipulationen oder Unregelmäßigkeiten hindeuten können. Die Anwendungsbeispiele zeigten exemplarisch den praktischen Wert dieser Analysemethode: Eine signifikante Abweichung von der Benford-Verteilung signalisiert Untersuchungsbedarf, während die Konsistenz mit der Gleichverteilung keine Hinweise auf Auffälligkeiten liefert und somit die Plausibilität der Daten stützt. Es ist jedoch wichtig zu betonen, dass die Ergebnisse dieses Tests **Indizien** und keine abschließenden Beweise darstellen, und stets im Kontext weiterer Prüfungsaktivitäten zu bewerten sind. Jedoch können sie zu einer ressourcenschonenden, effizienten Prüfung beitragen.

Obwohl methodische **Einschränkungen** wie Stichprobengröße und die Notwendigkeit der passenden Verteilungsannahmen zu beachten sind, unterstreichen die gericht-

liche Anerkennung und die wissenschaftliche Literatur die Relevanz der Ziffernanalysen für die Betriebsprüfung. Damit erweist sich der Chi-Quadrat-Anpassungstest als digitaler Detektiv in der modernen Finanzüberwachung.

Der Chi-Quadrat-Anpassungstest, angewendet auf Ziffernverteilungen, bildet demnach eine fundierte Basis für die **Risikobewertung** von Kassendaten. Er trägt maßgeblich dazu bei, Prüfern ein präzises Werkzeug zur Aufdeckung potenzieller Ungereimtheiten an die Hand zu geben und die Ordnungsmäßigkeit der Kassenführung besser beurteilen zu können, was wiederum die Steuerehrlichkeit fördern dürfte.

Literatur

- Benford, F., The Law of Anomalous Numbers, in: Proceedings of the American Philosophical Society, Vol. 78 (1938), No. 4, S. 551–572.
- BMF-Schreiben v. 5.9.2023, IV D 3 – S 1445/20/10007 :006, BStBl. I 2023, S. 1546.
- FG Münster v. 5.12.2002, 8 V 5774/02 E,G,U, BeckRS 2002, 16213.
- FG Rheinland-Pfalz v. 24.8.2011, 2 K 1277/10, EFG 2012, 10.
- Sosna, C., Statistische Ziffernanalyse, in: Die Steuerliche Betriebsprüfung, 34. Jg. (2004), Nr. 9, S. 249–253.
- Tödter, K. H., Das Benford-Gesetz und die Anfangsziffern von Aktienkursen, in: WiSt – Wirtschaftswissenschaftliches Studium, 36. Jg. (2007), Nr. 2, S. 93–98.
- Wähnert, A., Anwendbarkeit, Aussagekraft und Grenzen stochastischer Manipulationstests, in: Die steuerliche Betriebsprüfung, 38. Jg. (2008), Nr. 11, S. 312–318.
- Wähnert, A., Einsatzmöglichkeiten neuer Prüfverfahren für die Finanzverwaltung – Integration der neuen Prüfungsmethoden in ein systematisches Risikomanagement, in: BBK, 15. Jg. (2009), Nr. 15, S. 738–747.
- Wähnert, A., Die Ziffernanalyse als Bestandteil zeitgemäßer Prüfungsansätze, in: Die Steuerliche Betriebsprüfung, 42. Jg. (2012), Nr. 8, S. 241–246.
- Wähnert, A., Zeitgemäße Datenanalyse der Betriebsprüfung – Das Datenprüfungsnetz „Summarische Risikoprüfung (SRP)“, in: DER BETRIEB, 71. Jg. (2016), S. 2627–2631.