BIG DATA und Continuous Auditing

Betriebliche Daten und Betrug im Kontext prüferischer Urteilsbildung

Roger Odenthal, Köln

1 Wohin die Reise geht

„Oft hilft uns nur Kommissar Zufall“ gestehen Verantwortliche namhafter Wirtschaftsprüfungsgesellschaften, die sich häufig mit bilanziellen oder steuerlichen Folgen von Personaldelikten und Managerkriminalität auseinandersetzen müssen. Vergleichbar ratlos stehen interne Revisoren, steuerliche Betriebsprüfer und Mitarbeiter von Rechnungshöfen vor diesem Phänomen, dessen Eindämmung sie gleichwohl zu ihren zentralen Aufgaben zählen. Bei der Suche nach neuen Prüfungstechniken, die zur Lösung des aufgezeigten Dilemmas beitragen können, erfährt die Auswertung betrieblicher Massendaten mittels artifizieller Intelligenz besondere Aufmerksamkeit. Folgt man in diesem Zusammenhang den Offerten hierauf spezialisierter Berater und Software-Hersteller, dann können forensische Algorithmen sowie statistische Analysen die Spuren des Fehlverhaltens von Mitarbeitern mit hinreichender Sicherheit aufdecken.

Die aufgeführte Überzeugung hat sich zwischenzeitlich über zahlreiche Veröffentlichungen sowohl in einschlägigen Prüfungsstandards als auch in den Köpfen der Prüfenden fest verankert. Interne Revisoren analysieren unter dem Stichwort „Continuous Auditing“ regelmäßig kritische betriebliche Datenbereiche im Hinblick auf betrugsrelevante „Red Flags“, während Prüfer der Finanzverwaltung die Buchhaltungsdaten der Unternehmen automatisiert mittels „summarischer Risikoprüfung“ (SRP) zur Aufhellung von Steuerbetrug durchleuchten.

Wohin die Reise geht, können zahlreiche von solchen Prüfungen betroffene Bereiche und Unternehmen unschwer feststellen. Sie sehen sich, häufig ohne weitere Anhaltspunkte, allein auf der Basis „unplausibler“ statistischer Zahlenmuster oder -verteilungen mit Manipulationsvorwürfen konfrontiert, die sie mangels mathematischer Expertise kaum wirkungsvoll entkräften können. Eine zu Beginn des Jahres durchgeführte Podiumsveranstaltung der Bundesfinanzakademie zu dem Thema „Zeitgemäßer Umgang mit BIG DATA aus Sicht der Finanzverwaltung“ weist in die gleiche Richtung. Angesichts vielfältiger Steuerbetrügereien, so eine vielbeachtete Meinungsäußerung, sei es an der Zeit, Finanzgerichte davon zu überzeugen, dass nachteilige Steuerschätzungen und Pönalien sich im Streitfall ohne weitere Umstände mit ausreichender Zuverlässigkeit auf Anomalien betrieblicher Zahlenmuster stützen können.

Die aufgezeigten Entwicklungen erfordern angesichts hiermit verbundene Aufwendungen und Wirkungen eine kritische Begleitung. Was sagen praktische Prüfungserfahrungen? Bieten betriebliche Daten tatsächlich eine hinreichend aussagefähige Basis, um mittels regelbasierter Auswertungstechniken und statistischer Analysen sichere Betrugsindikatoren aufzuzeigen? Welche Rollen spielen Vernunft, Verstand und Erfahrung des Prüfers bei der prüferischen Urteilsfindung und lassen sich diese durch Computer ergänzen bzw. ersetzen? Mit diesen und weiteren Fragen beschäftigen sich die nachfolgenden Ausführungen.

2 „Normalität“ betrieblicher Zahlen?

In der kriminalistischen Praxis wird Betrug allgemein als „abweichendes Verhalten“ charakterisiert, welches sich, so die Vermutung, in einer Devianz bei Unternehmensdaten abbilden soll. Jedem hierauf bezogenen bewusstem oder statischem Auswertungsalgorithmus liegt somit die Vorstellung zugrunde, dass sich betriebliche Zahlen in vorhersehbaren Mustern, z.B. als Benford-Set, Normal- oder Pareto-Verteilung zusammenfinden, gegen die Betrugsfaktoren geprüft werden können.

Bereits diese einfach strukturierte Auffassung lässt sich mit praktischen Erkenntnissen kaum in Übereinstimmung bringen. Vielmehr ist der Entstehungsprozess solcher Zahlen in einen komplexen sozio-technischen Kontext eingebunden. Hierzu gehören u.a.:

- allgemeine geschäftliche Aktivitäten mit mehr kleinen als großen Geschäften
- Gesetzen und Vorschriften
  z.B. Aufschläge aus Umsatzsteuersätzen, vorgegebene Grenzwerte, Abgabemengen
- branchenbezogene Usancen
  mit Abrechnungsmodalitäten, Fallpauschalen, Transporteinheiten etc.
- regionale Besonderheiten
  unterschiedliche Finanzierungs-, Reise-, Kauf- oder Verzehrgewohnheiten
- betriebliche Vorgaben und Randbedingungen
  in Form von Konzernbeziehungen, Bestellrhythmen, Unterschriftsvollmachten, tätigen Organisationseinheiten, Losgrößen, sonstigen Optimierungen
- personelle Präferenzen
  als eigenständige Gestaltungen bei der Abwicklung von Arbeitsprozessen.

Diese und weitere Faktoren wirken mit jeweils unterschiedlichem Anteil sowie divergierender Dynamik auf die unternehmerischen Daten einschließlich der sich hieraus ergebenden Muster. Betriebliche Zahlen spiegeln somit regelmäßig eine höchst *individuelle Konstellation*, in die Betrug lediglich mit einem homöopathischen Anteil eingeht.

Viele eher grobe statistische Auswertungsalgorithmen orientieren sich demgegenüber primär an den Erkenntnissen aus *allgemeinen geschäftlichen Aktivitäten*, mit der Folge, dass alleine daraus resultierende, unspezifische Abweichungen von Zahlenmustern und Verteilungen sich in erster Linie auf die vielfältigen weiteren Einflüsse der Zahlengestaltung zurückführen lassen, ohne dass hiermit ein Betrugsanzeichen verbunden wäre.

Der aufgeführte Sachverhalt erschwert nicht nur übergreifende datenanalytische Betrachtungen, sondern wirkt auch auf Zeit- und Organisationsvergleiche. Insbesondere verdachtsunabhängigen Betrugsuntersuchungen muss daher stets eine sorgfältige und prüffeldbezogenen Kalibrierung des Datenbestandes vorangestellt werden, ansonsten bleiben deren Ergebnisse erfahrungsgemäß blass.

3 Zahlen und Muster

Wenn bereits unser Wissen über die einer Musterbildung zugrundeliegenden betrieblichen Zahlen auf tönernen Füßen steht, wie verhält es sich dann mit den zu Prüfungszwecken herangezogenen Datenmustern, aus welchen zuverlässige Betrugsindikatoren herausgelesen werden sollen?

- Benford-Sets als Ziffernmuster

  Diesen Mustern liegen typische Verteilungen zur Häufigkeit einzelner Ziffern innerhalb *unbeeinflusster* betrieblicher Zahlenbestände zugrunde. Erwartet werden mehr niedrige als hohe Ziffern mit genau definierten Soll-Umfängen für Zifferpositionen und –kombinationen. Zu den theoretischen Verteilungsmodellen, -grundlagen und Ziffernanteilen finden sich zahlreiche Veröffentlichungen, auf die an dieser Stelle verwiesen werden kann. Interessant sind in diesem Zusammenhang die nachfolgend dargestellten Anwendungserfahrungen aus der Prüfungspraxis.

In aller Regel überdecken betriebliche und sonstige Faktoren die „allgemeinen“ Entwicklungstendenzen der Benford-Muster. Bereits hierdurch ergeben sich spezielle und absonderlich erscheinende Ziffernverteilungen mit hohen Abweichungen zu Benford-Werten, ohne dass daraus Betrugssignale abgeleitet werden können.

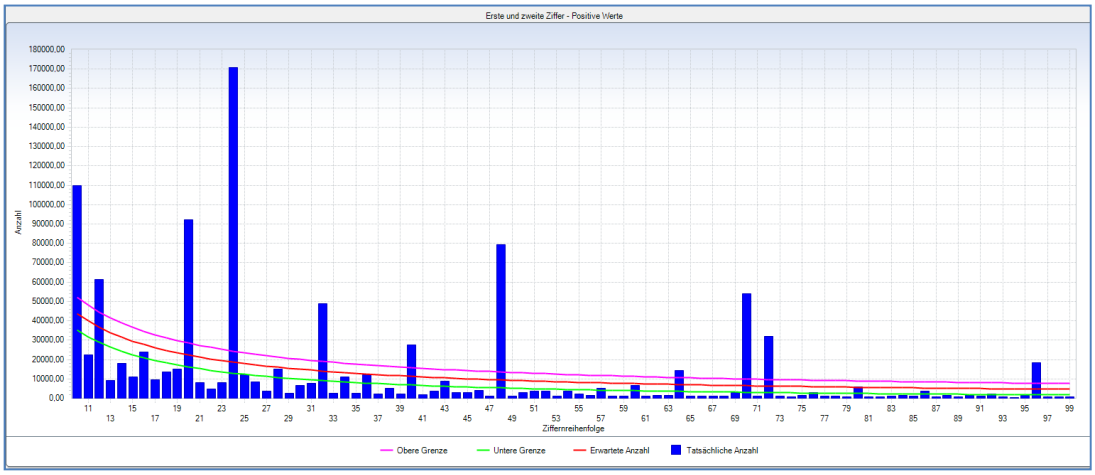


Abb. 1: Beispiel einer von Benford abweichenden Ziffernverteilung *ohne* Betrugsanzeichen

Selbst die exakte Einhaltung von Verteilungsmustern führt in der Praxis nicht weiter. So stützen Compliance-Verantwortliche ihre Hypothesen-Testverfahren als Negativtest häufig auf Benford-Analysen. Zeigen sich bei der Auswertung eines Datenbestandes keine über die Signifikanzgrenzen hinausreichenden Abweichungen, so gehen sie von betrugsfreien Verhältnissen aus.

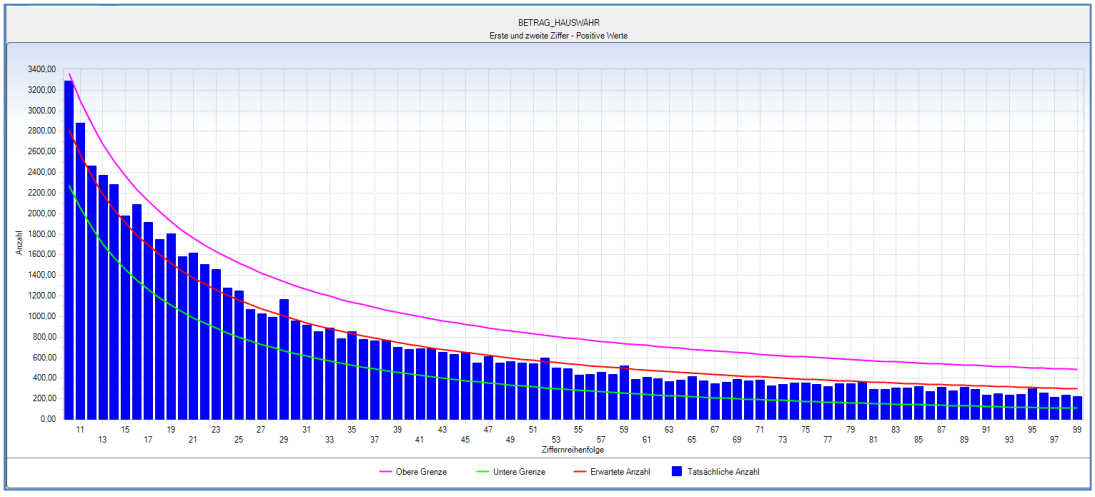


Abb. 2: Betrieblicher Datenbestand *mit* Betrug, der einer Benford-Verteilung folgt (Beispiel)

Vielfach handelt es sich um einen fehlerhaften Rückschluss. Einschlägige Untersuchungen zeigen, dass auch eine erhebliche betrügerische Zahlenkontamination (10%) in einem Prüffeld lediglich bei 60% der Analysen ein Überschreiten der Signifikanzschwelle nach sich zieht. Eine zwanzigprozentige Kontamination erhöht diese Entdeckungswahrscheinlichkeit auf ca. 85% der Betrachtungen. Selbst nachhaltige Betrügereien dürften allerdings selten entsprechende Größenordnungen erreichen, wenn nicht maßgebende Unternehmensteile hierin involviert sind und interne Kontrollen völlig versagen. Sie bleiben bei der undifferenzierten Benford-Analyse schlicht unentdeckt.

Als Trugschluss hat sich weiterhin die verbreitete Vermutung herausgestellt, dass sich, unabhängig von der Zahlengröße, in jedem vollständigen Ziffernbereich (z. B. 10 – 99, 100 – 999) eines benfordverteilten Zahlenbestandes wiederum ein Benford-Set entwickeln müsse. Sich hieraus ergebende Ziffernmuster haben vielmehr eine hohe beobachtbare Variationsbreite. Dieses muss bei einer Untersuchungskonzentration auf besonders hohe, risikobe-

haftete Unternehmenswerte berücksichtigt werden, um verlässliche Interpretationen von Abweichungen zu ermöglichen.

Überdies bieten Abweichungen von Benford-Verteilungen keine wirkungsvollen Ansatzpunkte für die Ursachenforschung bei unterrepräsentierten Ziffern. Aus diesen und weiteren Gründen lässt sich eine belastbare Betrugsvermutung kaum auf ein Benford-Set stützen.

- Normal- und Log-Normalverteilung als Zahlenmuster

  Insbesondere bei Suche nach „fehlenden" Werten helfen automatisierte Betrugsanalysen auf der Grundlage ziffernbasierter Benford-Muster selten weiter. Ein Problem, mit welchen sich steuerliche Betriebsprüfer angesichts des Manipulationspotentials im Umfeld unkontierter Barkassengeschäfte täglich auseinandersetzen müssen.

  Deren „Geheimwaffe" besteht in einer logarithmischen Umwandlung der Zahlen (z.B. Umsätze als Produkt von Menge und Preis) ihres Prüffeldes. Diese nimmt der Zusammenstellung heterogener Umsatzgrößen die Schiefe und führt im Folgenden *annähernd* zu einer Normalverteilung (oder Log-Normalverteilung) der logarithmischen Umsatzwerte. Damit ist die Grundlage für *eine rechnerische Bestimmung der Anzahl von Sollpositionen bei Umsätzen ausgewählter Größenklassen* gelegt. Diesen kann die tatsächliche Anzahl entgegengestellt werden, um auf unterlassene Einnahmeaufzeichnungen zu schließen.

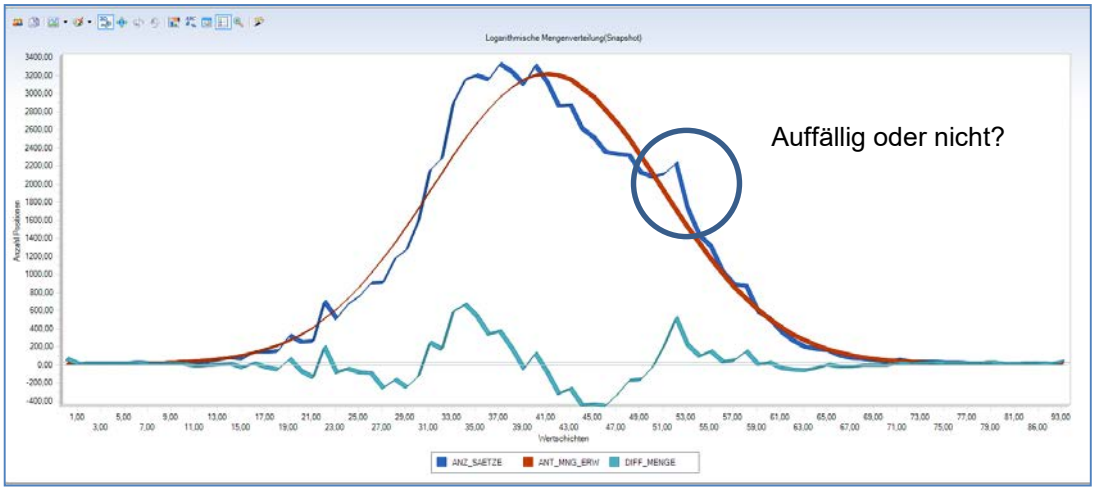


Abb. 3: Mengenverteilung logarithmierter betrieblicher Rechnungspositionen (blau) im Verhältnis zu einer errechneten Normalverteilung (rot) entsprechender Werte (Beispiel)

Bei Betrugsanalysen in bestimmten Prüffeldern (Rechnungen, Zahlungen, Kassengeschäfte) vermitteln die aufgeführten Überlegungen sinnvolle Ansatzpunkte für hierauf abgestimmte automatisierte Analysetechniken. Hinreichende Verdachtsmomente lassen sich ohne zusätzliche Erkenntnisse zu fehlenden oder manipulierten Belegen alleine hieraus jedoch nicht ableiten.

Zunächst entspricht nicht jede zufällige Verteilung einer Normal- oder Log-Normalverteilung. Betriebliche Usancen und individuelle Preismodelle wirken auch hier auf die Muster, welche einer Analyse zugrunde gelegt werden können. Für „normierte" Positionen ist weiterhin durchaus nicht gesichert, ob sie einer Normalverteilung oder logarithmierten Normalverteilung folgen. Häufig finden sich für beide Muster Argumente. Der Unterschied ist jedoch evident, da die Log-Normalverteilung für seltenere (höhere) Werte tendenziell größere Erwartungswerte ausweist, als die Normalverteilung[1]. Wie aber soll sich ein des Betrugs Verdächtigter zu vermuteten Schwarzgeld-Geschäften äußern, wenn nicht einmal klar ist, ob das zum Beweis herangezogene Zahlenmuster überhaupt eine praktische Evidenz aufweist?

[1] Vergleichende Untersuchungen zu diesen Themen finden sich u.a. bei Krehl/Strobel/Schaller, Stichproben und statistische Verfahren im Abschluss- und Betriebsprüfungsprozess

- Empirische Muster

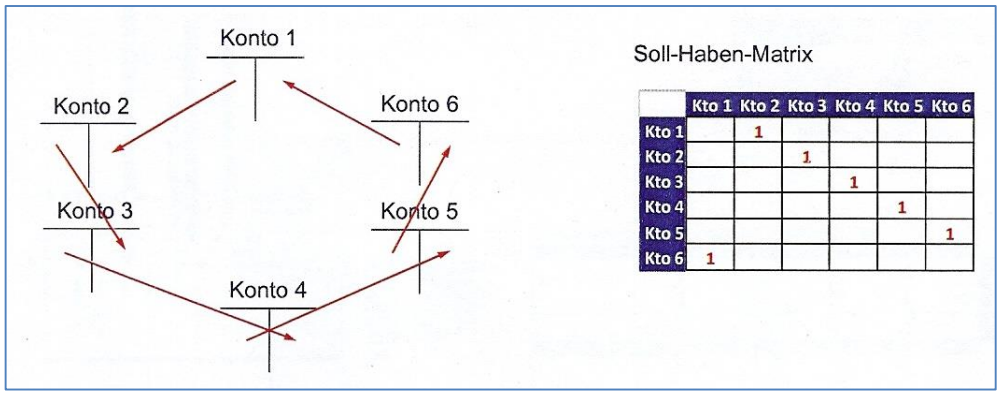
  Sollwerte für prüfungsbezogene Abweichungsanalysen lassen sich weiterhin aus nachvollziehbaren Zusammenhängen ableiten. Datenmuster aus Korrelationsanalysen oder Regressionsrechnungen können hier eingeordnet werden. Beliebt sind z.B. Plausibilitätsbetrachtungen zur korrespondierenden Entwicklung von Umsatz- und Wareneinsatzgrößen, der Sonnentage versus Außengastronomie-Umsätze oder zu Kraftstoffaufwendungen gegen Einnahmen im Taxigeschäft. Auf den ersten Blick auffällige Konstellationen verlieren aber auch hier rasch an Charme bei einem Blick auf die Details. Gegenläufige Zahlenmuster können häufiger auf geänderte Lagergewohnheiten, Mentalitätsunterschiede von Kunden oder unterschiedliche Verkehrsbedingungen zurückgeführt werden, als auf (Steuer-)Betrug.

- Kontierungszuordnungen als Handlungsmuster

  Zuletzt kann den statistisch determinierten Sollvorgaben ein Handlungsmuster beigestellt werden. Dabei werden z.B. Zahlenströme auf der Grundlage ihrer Buchungen als empfangend und abgebend (Soll / Haben) jeweils automatisiert betroffenen Konten-/ Gegenkonten zugeordnet.

  Hieraus lassen sich eine Reihe möglicher Betrugsindikatoren ableiten:

  - Ungewöhnliche Konten- / Gegenkonten Kombinationen
  - Fehlende übliche Kontenkombinationen
  - Karussell- und Umwegbuchungen
  - Bewegungen zwischen Geldverkehrskonten.



	Kto 1	Kto 2	Kto 3	Kto 4	Kto 5	Kto 6
Kto 1		1				
Kto 2			1			
Kto 3				1		
Kto 4					1	
Kto 5						1
Kto 6	1					

Abb. 4: Matrizendarstellung eines Buchungsjournals und Mustererkennung (Quelle: Mochty)[2]

Eine einfache Betrachtung, sollte man meinen, die sich alleine auf das vorhandene Buchungsmaterial stützt und artifizieller Intelligenz daher besonders zugänglich sein müsste. Zwar stehen auch hier einheitlichen „Sollvorgaben“ wiederum vielfältige betriebliche Handhabungen entgegen, praktisch scheitert eine Prüfung gegen solche Muster jedoch häufiger an technischen Randbedingungen.

Viele Buchungssysteme teilen computergestützte Belege in ein Kopf- und Belegsegment auf. Während ersteres den formalen Zweck der Buchung charakterisiert, wird der materiell und sachlich wirksame Buchungsinhalt in Buchungshalbsätzen (Splitbuchungen) unterschiedlicher Anzahl auf der Positionsebene abgebildet. In einer solchen Umgebung sind eindeutige computergestützte Zuordnungen für zusammengehörende Soll- / Habenpositionen bei mehr als zwei Buchungssegmenten ausgeschlossen[2]. Sie können lediglich interpretativ und mit menschlichem Sachverstand nachvollzogen werden.

Die Problematik dieser und weiterer für die automatisierte Betrugsaufklärung bereitstehender Muster liegt somit auf der Hand. Sie sind weder widerspruchsfrei noch gesichert. Ihre Anwendung beinhaltet weite Interpretationsspielräume, und hierauf basierende Urteile sind mit hohen Irrtumsrisiken verbunden.

[2] Mathematisch-methodische Beweisführung durch Prof. Ludwig Mochty, Universität Duisburg-Essen, Vortragsunterlagen „Visualisierung von Buchungsjournalen“, Bundesfinanzakademie, 04. März 2015

4 Forensische Analyselogik und Auswertungsregeln

Die *automatisierte Identifizierung* von Betrugsgeschehen mittels Unternehmensdaten setzt einen komplexen Transformationsprozess voraus. Dabei müssen für die prüferische Urteilsfindung erforderliche Analyseregeln auf das eingesetzte Softwareverfahren übertragen werden. Ein ambitioniertes Vorhaben, angesichts Milliarden möglicher Verschaltungen, die den Analysevorgang bereits in einem durchschnittlichen Prüferhirn begleiten können. Wirklich interessant sind in diesem Zusammenhang jedoch neurobiologische Erkenntnisse zu Optimierungsstrategien des Gehirns.

Einschlägige Untersuchungen zeigen, dass sich Urteile und Entscheidungen ganz überwiegend auf Erfahrungswissen stützen, welches sich in einer für das Bewusstsein nicht direkt zugänglichen Hirnregion, dem limbischen System, verankert. Problemorientierte Erkenntnisprozesse münden hierbei häufig bereits in Handlungsimpulsen, noch bevor sie den Neokortex und damit die Bewusstseinsebene erreichen[3]. Ein um Kohärenz bemühtes Bewusstsein bemüht sich vielfach erst *nach dem Handlungsimpuls* um Regeln, die mit dem voreilenden Entscheidungsablauf in Übereinstimmung zu bringen sind, um Handlungsvollmacht zu erringen.

Für die Automatisierung von Analysevorgängen sind die geschilderten Abläufe fatal, da hierbei lediglich Teile komplexer Entscheidungsregeln bewusst zugänglich sind. Intuition und Erfahrung finden hingegen kaum eine angemessene Berücksichtigung. Zwar bestehen automatisierte Lösungsstrategien zu unspezifischen Problemen, wie uns Schachcomputer zeigen. Diese wirken allerdings in einem fest definierten Rahmen von Figuren und Zugmöglichkeiten, der bei Betrug sowie Betrugsindikatoren nahezu völlig fehlt.

Unvollkommen automatisierte Analyseregeln führen bei der praktischen Anwendung computergestützten Monitorings regelmäßig zu Problemen. Diese äußern sich z.B. in:

- zu vielen Betrugsanzeichen

  Die Analyseregeln sind zu oberflächlich und grob, um betriebliche Faktoren, Fehler oder Betrug sorgfältig voneinander anzugrenzen,

- zu wenigen Betrugsanzeichen

  Kombinierte Analyseregeln zur Reduzierung des Ergebnisraums lassen wesentliche Handlungsmuster des Betrügers außer Acht.

Die vermeintlich einfache automatisierte Suche nach betrügerischen Doppelzahlungen im Kreditorenbereich vermittelt ein gutes Bild zu der aufgezeigten Problematik. Wenn ca. 10 Prozent aller Zahlungen erfahrungsgemäß mit identischen Beträgen (erste Suchstrategie) kontiert werden, ergibt sich ein kaum zu bewältigendes Untersuchungsvolumen. Begrenzt man diesen Umfang durch ein weiteres Merkmal, z.B. identische Kreditoren (zweite Suchstrategie), mit vergleichbarer Verteilung innerhalb des Zahlungsbestandes, so verbleiben kombiniert lediglich noch 1 Prozent Entdeckungswahrscheinlichkeit. Ein Glücksfall für Betrüger, die ihren Modus Operandi im Hinblick auf eines dieser Merkmale geringfügig variieren. Sie werden zu 99 Prozent unentdeckt bleiben.

Insgesamt vermittelt sich ein eher ambivalenter Eindruck zur Umsetzung heuristischer Suchregeln für forensische Prüfzwecke. Der Komplexität des Prüffeldes werden sie erfahrungsgemäß selten gerecht.

5 Vom Umgang mit Regeln

Wie wir bereits feststellen konnten, erweisen sich viele Gewissheiten, auf die wir uns weitgehend selbstverständlich bei der automatisierten forensischen Auswertung von BIG DATA stützen, bei näherer Betrachtung als Schimären. Einer *selbstreflektierenden, zurückhaltenden und verantwortungsvollen Interpretation hieraus resultierender Ergebnisse* kommt angesichts hiervon betroffener Menschen daher besondere Bedeutung zu. Worauf können wir in diesem Zusammenhang vertrauen?

Ohne Zweifel möchte sich jeder Prüfer, seinem Berufsethos folgend, überwiegend unvoreingenommen mit den Ergebnissen forensischer Datenanalysen auseinandersetzen. Allerdings setzt

[3] Vergleiche u.a. Gerd Gigerenzer, Bauchentscheidungen – Die Intelligenz des Unbewussten und die Macht der Intuition, Goldmann Verlag, 2008

ihm auch hier die menschliche Natur enge Grenzen, in dem sie assoziativen Denkvorgängen gegenüber Zufällen den Vorzug einräumt. Sobald wir vermeintliche oder tatsächliche „Muster“ in Unternehmensdaten zu erkennen glauben, erfolgt zwangsläufig eine Konstruktion von Zusammenhängen, auch wenn diese „Muster“ zufällig entstanden sind. Dieses irrationale Verhalten gegenüber zufälligen Aspekten lässt sich auf evolutionsbiologische Entwicklungen zurückführen. Die Annahme, dass bei Dunkelheit zwei nahe beieinanderliegende Lichtreflexe auf das Augenpaar eines Fressfeindes hinweisen und nicht auf sich im Liebesrausch umkreisende Glühwürmchen, vermittelte Selektionsvorteile, die uns bis heute begleiten. Auf zufällige Erscheinungen schließen wir in der Folge lediglich nachrangig und indirekt, wenn wir trotz aller gegenteiligen Mühen keine Muster mehr erkennen können.

Jedem erfahrenen Prüfer ist dieses Phänomen geläufig. Ein undifferenzierter Anfangsverdacht wird auch ohne weiteren Anlass bereits dann geboren, wenn sich Ähnlichkeitsassoziationen alleine durch die Betrachtung wertmäßig nahe beieinanderliegender Zahlungen oder durch Ziffernhäufungen einstellen.

Ein weiterer Gefahrenbereich verbirgt sich in unserem Erkenntnispotential. Wir können in erster Linie sowie unangestrengt ausschließlich das erkennen und interpretieren, was wir wissen. Fehlen uns Kenntnisse oder Erfahrungen, so bleiben selbst offensichtliche Sachverhalte unbeachtet. Ein gutes Beispiel vermitteln Vexierbilder mit unterschiedlichen Bedeutungsinhalten.

Abb. 5: Vexierbild mit unterschiedlichen Bedeutungsinhalten, je nach Betrachter (Beispiel)

Die vorstehenden Ausführungen verweisen darauf, dass wir in unseren zahlenbasierten Erkenntnis- und Beurteilungsprozessen wahrscheinlich weniger objektiv sein können, als wir möchten. Fehlerhafte Verdächtigungen sind ebenso inkludiert, wie unsachgemäße Rückschlüsse zu Prüffeldern, in welchen wir Zahlen mangels Wissen nicht zu deuten vermögen.

Die aufgeführten Folgen werden dort verstärkt, wo wir uns lediglich noch mit den *Ergebnissen automatisierter Analysen* auseinandersetzen dürfen und nicht mehr mit dem Weg dorthin. Gleiches gilt für die von der Fiskalverwaltung bei Betriebsprüfungen aktiv geförderte *„Verbildlichung“* unternehmerischer Zahlenbestände, welche *“Abweichungen“* in Grafiken besonders deutlich herausstellt. Die häufig laienhafte und fehlerhafte Einschätzung von Risiken zu fehlerhaften Ergebnis-Interpretationen leistet ebenfalls einen nachhaltigen Beitrag.

> Wenn eintausend Analysefälle zu unternehmerischen Daten einhundert signifikante Abweichungen bei erwarteten Zahlenmustern aufweisen, welche lediglich in zehn Fällen auf nachweisbaren Betrug zurückgeführt werden können, dann liegt der *Anteil fehlerhaft produzierter Verdachtsmomente bei neunzig* und nicht, wie oft kolportiert, bei einem Prozent.

Als Fußangel erweisen sich weiterhin komplizierte rechentechnische Skalierungen, die vermeintliche Objektivität suggerieren. Je aufwendiger (und undurchschaubarer) ein Analyse- und Bewertungsmodell, desto eher werden dessen Ergebnisse auch von skeptischen Prüfern unreflektiert übernommen und weitergetragen, wie zahlreiche Beispiele zeigen.

Einen möglichen Beitrag zu möglichen Missdeutungen leisten ebenfalls die zur Analyse eingesetzten Softwarewerkzeuge, wie das nachfolgende Beispiel des weitverbreiteten Programms IDEA (Version 9.2) zeigt.

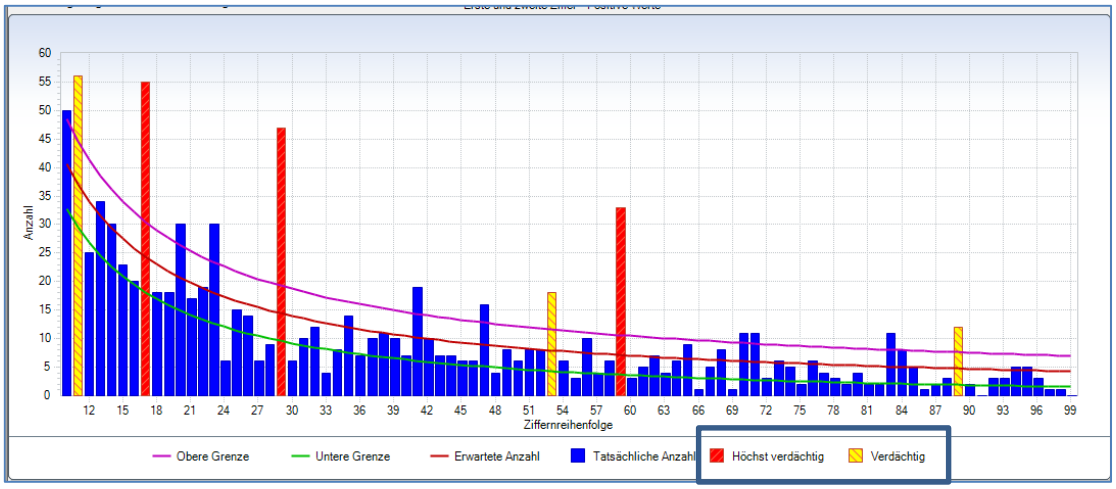


Abb. 6: Automatisierte forensische Muster-Auswertung in IDEA mit prüferischer Beurteilung

Abweichungen von einem erwarteten Zahlenmuster werden hier bereits softwaretechnisch und damit *a priori als „verdächtig" oder gar „höchst verdächtig"* bezeichnet. Die Analyse selbst impliziert gleichzeitig das prüferische Urteil und ist somit in den Augen vieler Revisoren handlungsleitend. Angesichts der dargestellten Unsicherheiten sicherlich eine sehr bemerkenswerte Programmfunktion!

Wie im vorliegenden Fall wird bei der Interpretation von Ergebnissen automatisierter forensischer Datenanalysen häufig auf einfache Erklärungsmodelle zurückgegriffen, um mit diesen „Billigvarianten" rezepthafte Aussagen bei komplexen Sachverhalten zu unterlegen. Ein Umstand, welcher keine optimistische Bewertung unseres Umgangs mit BIG DATA und Betrug ermöglicht.

6 Was leisten die Analyse-Werkzeuge?

Das kontinuierliche Monitoring zu Betrugsanzeichen in Unternehmensdaten erfolgt überwiegend mit Hilfe sogenannter „Prüfsoftware". Die hierbei gängigen Programme weisen einen mehr als 30ig-jährigen Entwicklungshorizont auf, welcher bis an die Anfänge des püferischen Einsatzes von Personalcomputern heranreicht. Der ursprünglichen Anwendung lag die Idee eines schrittweisen interaktiven Dialogs erfahrener Revisoren mit den Unternehmensdaten zugrunde, um hieraus Ansätze für die nachfolgende Sichtung ausgewählter Geschäftsvorfälle zu entwickeln. Die Software war mit nachfolgenden Schwerpunkten hierauf abgestimmt:

- Datenübernahmen aus unterschiedlichsten Vorsystemen,
- dialogorientierte und revisionsbezogene Auswertungswerkzeuge,
- vergleichsweise hohe Arbeitsgeschwindigkeiten auf dem Personalcomputer zur Verarbeitung größerer Datenmengen.

Weder Betrugsanalysen noch die Automatisierung von Auswertungen standen bei der Entwicklung solcher programmtechnischer Lösungen im Vordergrund.

Betrachtet man die aktuellen Programmversionen unter funktionalen Gesichtspunkten, so kommt man zu dem Ergebnis, dass sich dieses über unterschiedliche Software-Generationen hinweg kaum geändert hat. Weiterhin konzentriert sich das Anwendungsspektrum auf die bereits aufgeführten Programmmerkmale sowie wenige strukturelle Aufbereitungsmöglichkeiten. Funktionen mit besonderen Bezügen zu den Kennzeichen betrügerischer Handlungen, zur Vertuschung oder Täuschung, sucht man in den standardmäßig bereitgestellten Auswertungsroutinen weitgehend vergeblich. Wünschenswert wären z.B. die artifizielle Diagnose von Zusammenhängen und Abweichungen in Unternehmensdaten, Ähnlichkeitsanalysen bei Texten sowie bei Beträgen, Verprobungstechniken anhand von Quersummen oder Schlussziffernverfahren, Perspektivwechsel durch Dimensionsänderungen bei datentechnischen Zusammenstellungen,

sowie weitere Werkzeuge, die dem komplexen Prüffeld gerecht werden. Stattdessen liegen die Schwerpunkte softwaretechnischer Weiterentwicklungen erkennbar auf der Auswertung sogenannter „Massendaten“ mittels Serverkomponenten. Diese ist nicht nur aufwendig sondern *muss* auch automatisiert erfolgen, da sich ein solches Datenvolumen zufriedenstellend nicht mehr dialogorientiert handhaben lässt. Eine wesentliche Verbesserung des Erkenntnispotentials für Betrugsprüfungen ist hierbei allerdings kaum feststellbar.

7 Der feine Unterschied

Unabhängig von den bisher vorgestellten Kritikpunkten stellt sich die Frage nach der Nutzung von Erfolgsfaktoren des BIG DATA-Konzeptes. Schließlich sind mit diesem Ansatz milliardenschwere Konzerne entstanden. Unternehmensdaten werden als unerschlossene Goldgruben bezeichnet, die es lediglich zu heben gilt, um mit Hilfe dieser Schätze sowie neuen Erkenntnissen Differentialvorteile im wirtschaftlichen Wettbewerb zu generieren. Warum kann das nicht gleichermaßen für die automatisierte Aufhellung von Betrugsverhalten gelten?

Zunächst einmal sind computergestützte Prüfungstechniken ohne jeden Zweifel eine wertvolle Bereicherung des revisorischen Werkzeugkastens. Hierbei dürfen allerdings wesentliche Unterschiede zu BIG DATA Ansätzen nicht übersehen werden. Diese verfolgen in erster Linie einen *rekursiven Ansatz*. Weitgehend ohne Prädisposition ermittelt man durch vielfältige Verknüpfungen und Strukturierungen von Daten zunächst mögliche Muster, um diese anschließend einer Verifizierung und Einordnung hinsichtlich ihrer Bedeutung zu unterziehen. Dabei ist die im ersten Schritt erfolgende Zusammenstellung primär ein artifizieller Prozess, die nachfolgende Einordnung und Klassifizierung der Muster aber nur teilweise. Hier spielen der menschliche Faktor mit Wissen, Erfahrung und Intension wieder eine herausragende Rolle.

Der Unterschied zum prüferischen Vorgehen, bei dem in der Regel bereits eine Sollerwartung sowie die Analyse von Abweichungen im Vordergrund stehen, ist offensichtlich. Zudem treten bei normalen geschäftlichen Aktivitäten die hieraus resultierenden Daten weitgehend unverborgen zu Tage, während es sich bei Betrug um Heimlichkeitsdelikte und höchst individuelle Erscheinungen handelt. Hier bestimmen persönliche Dispositionen, Fantasie oder organisatorische Einbindung die Aktivitäten des Betrügers einschließlich der hieraus resultierenden Daten. Sie weisen eine weite, kaum fassbare Variationsbreite auf. Es bestehen somit entscheidende Unterschiede zwischen BIG DATA, der Auswertung von Massendaten und automatisierten Betrugsanalysen.

8 Alles Nichts oder?

Welche Optionen eröffnen sich angesichts der zahlreichen Kritikpunkte aus praktischer Sicht für die computergestützte Betrugsaufhellung?

Zunächst gilt es, die durch zahlreiche Veröffentlichungen sowie von Beratern geschürte Hybris bei der automatischen Betrugsdetektion zu überwinden. Das verdachtslose Fahnden nach Betrugsanzeichen bei geschäftlichen Aktivitäten ist und bleibt hartes Handwerk in einem schwierigen prüferischen Umfeld. Letztendlich geht es um Menschen und dementsprechend sind menschliche Faktoren wie Empathie, Erfahrung, Intention sowie die andauernde kritische Auseinandersetzung mit Zwischen- und Gesamtergebnissen von herausragender Bedeutung. Dieser Prozess kann nicht ohne weiteres in automatisierter Form auf Maschinen übertragen werden.

Besinnt man sich allerdings zurück auf den ursprünglichen Ansatz computergestützter Prüfung, den Mensch-Maschine-Dialog, so ergibt sich ein differenzierteres Bild. Der Prüfer appliziert hierbei sein prüferisches Wissen auf die innerhalb der Software bereitgestellten Unternehmensdaten. Er erarbeitet schrittweise Ergebnisse, in dem er Strukturen erstellt, begutachtet, verwirft und verfeinert. Der Computer ist ihm zuhanden und er verfolgt, was entscheidend ist, den Weg seiner Analysen einschließlich der Faktoren, die er bis zum endgültigen Ergebnis außer Acht lassen musste. Insgesamt stützt er seine prüferische Betrugsvermutung auf ein wesentlich breiteres Fundament, als bei automatisierten Auswertungen. Hieraus resultieren Erfahrungen, die qualitative Verbesserungen in weiteren Prüfungen ermöglichen. Zudem wird die Fragilität von Aussagen wieder erfahrbar und trügerische Formen unangemessener Selbstgewissheit treten in den Hintergrund.

Für diese Art der Auswertung sollten ambitionierte Revisoren dann allerdings nachhaltig eine Verbesserung der Funktionen bei Softwareherstellern von ihnen eingesetzter Produkte einfordern.

9 Zusammenfassung

Abschließend gilt es noch einmal die eingangs geschilderten Entwicklungstendenzen zu bewerten. Betrug, in allen Spielarten, ob als Steuerbetrug, Bilanzfälschung, Wettbewerbsdelikt oder Kassenunterschlagung wird in Zukunft, abseits aller gegenteiligen Erwartungen, zufällig zutage treten. Hieran wird auch die eingesetzte Computertechnik nichts Wesentliches ändern. Dies liegt nicht nur an den besonderen Merkmalen des Betrugsgeschehens, sondern auch daran, dass sich kaum jemand die Tätigkeit in einem Unternehmen vorstellen möchte, welches jeden Betrug zuverlässig durch den Computer angezeigt.

Gleichwohl zeigen sich Ansatzpunkte für Verbesserungen bei computergestützten Betrugsprüfungen. Hier steht allerdings wieder verstärkt der Prüfer mit seinen kognitiven Fähigkeiten und erforderlichen Werkzeugen zur sachkundigen Analyse im Vordergrund. Für trügerische Selbstgewissheiten sowie Strafen, die sich alleine auf statistische Verdachtsmomente stützen möchten, bleibt hiernach nur noch wenig Raum.