

DYNAMISCHE PREISSETZUNG ALS HERAUSFORDERUNG FÜR DIE VERBRAUCHERPREISSTATISTIK

Christian Bladow, Florian Burg

↳ **Schlüsselwörter:** Verbraucherpreisindex – Harmonisierter Verbraucherpreisindex – dynamische Preissetzung – Internetpreiserhebung – Web Scraping

ZUSAMMENFASSUNG

Mithilfe von automatisierten Algorithmen können Internetgeschäfte die Preise von Gütern – abhängig von verschiedenen Parametern – in kurzen Zeitabständen ändern. Diese Form der Preisgestaltung wird als dynamische Preissetzung bezeichnet. Für den deutschen Verbraucherpreisindex (VPI) und den Harmonisierten Verbraucherpreisindex (HVPI) werden derzeit monatlich etwa 10 000 Preise im Internet erhoben, im Regelfall zu einem festgelegten Zeitpunkt. Dynamische Preissetzung im Internet stellt die Verbraucherpreisstatistik vor die Herausforderung, die Preisentwicklung weiterhin repräsentativ zu erfassen und dabei auch volatile Preise zu verarbeiten. Daher untersuchte im Jahr 2017 eine Studie zur dynamischen Preissetzung das Preissetzungsverhalten von Internetgeschäften. Es wurden ausschließlich Produkte aus der Stichprobe des VPI/HVPI beobachtet; als Technik für die Erfassung der Preise wurde Web Scraping eingesetzt.

↳ **Keywords:** consumer price index – harmonised index of consumer prices – dynamic pricing – online price collection – web scraping

ABSTRACT

The application of automatised algorithms allows online shops to change product prices at short intervals, depending on various parameters. This type of price-setting behaviour is referred to as dynamic pricing. At present, roughly 10,000 online prices are collected each month for the German Consumer Price Index (CPI) and the Harmonised Index of Consumer Prices (HICP), usually at a given point of time. Dynamic pricing on the internet presents consumer price statistics with the challenge of continuing to capture the representative price development and also of processing volatile prices. For this reason, a study on dynamic pricing was conducted in 2017 to examine the price-setting behaviour of online shops. The study only monitored products which were included in the CPI/HICP sample and used the technique of web scraping for collecting prices.



Christian Bladow

hat internationale Volkswirtschaftslehre an der Universität Göttingen studiert und ist seit März 2016 wissenschaftlicher Mitarbeiter im Referat „Zwischenörtliche Preisvergleiche, Parteienindex“ des Statistischen Bundesamtes. Schwerpunkt seiner Arbeit ist die Automatisierung von Preiserhebungen im Internet.



Florian Burg

ist Diplom-Volkswirt und leitet das Referat „Verbraucherpreise“ des Statistischen Bundesamtes. Er koordiniert die Zusammenarbeit mit den Statistischen Ämtern der Länder im Bereich der Verbraucherpreisstatistik und arbeitet seit einigen Jahren auf europäischer Ebene an der Weiterentwicklung der Methoden des Harmonisierten Verbraucherpreisindex mit.

1

Einleitung

1.1 Was bedeutet „dynamische Preissetzung“?

Allgemein wird von dynamischer Preissetzung gesprochen, wenn automatisierte Algorithmen für die kurzfristige Preisgestaltung genutzt werden. Dabei ändern sich Preise als Reaktion auf veränderte Marktbedingungen oder auf Änderungen der Zahlungsbereitschaft von Konsumentinnen und Konsumenten.

Preisänderungen als Reaktion auf veränderte Marktbedingungen sind grundsätzlich nicht neu. Vor dem digitalen Zeitalter war dynamische Preisgestaltung jedoch auf solche Waren und Dienstleistungen begrenzt, deren Preise ohne nennenswerte Kosten (sogenannte Menükosten) schnell geändert werden konnten. Zum Beispiel ändern sich die Preise von Kraftstoffen an Tankstellen üblicherweise mehrmals am Tag, da Tankstellen schon sehr lange elektronische Preisschilder verwenden. Dienstleistungen mit unregelmäßigen und kurzfristigen Preisänderungen finden sich vielfach in den Bereichen Transport und Tourismus, beispielsweise Flüge, Pauschalreisen und Mietwagen. Bei diesen Dienstleistungen sind sowohl der Zeitpunkt der Buchung als auch der Zeitpunkt der Inanspruchnahme der Leistung entscheidend für die Preisgestaltung. Vor dem digitalen Zeitalter wurde die Buchung solcher Dienstleistungen von Reisebüros übernommen. Inzwischen buchen Privatpersonen Reisen verstärkt über das Internet. Diese Art der Buchung erlaubt den Internetgeschäften unmittelbare Preisänderungen, je nach Auslastung und Saisonzeiten. Um solchen Preisveränderungen Rechnung zu tragen, werden für diese Dienstleistungen nicht nur einmal im Monat Preise erhoben (wie es für Produkte mit weniger volatilen Preisen üblich ist), sondern mehrere Male zu unterschiedlichen Zeitpunkten innerhalb eines Monats. Auch wenn Preiserheberinnen und Preiserheber mit den Preismustern der Güter vertraut sind, ist diese Vorgehensweise zeitaufwendig und erfordert eine entsprechende Personalausstattung. Darüber hinaus können sich Preismuster von Produkten verändern und Preismuster neu aufgenommener Produkte müssen analysiert werden. Aus diesen Gründen werden bereits

von einigen Produkten Preise automatisiert (mittels sogenanntem Web Scraping, siehe unten) erhoben, um unabhängig von Zeit und Personal in kurzen Abständen Preisbeobachtungen durchführen zu können.

Neben den oben genannten Güterbereichen, für die schon traditionell eine dynamische Preissetzung angewendet wird, scheint sich dieses Phänomen auf weitere Branchen und Produktgruppen im Internethandel auszuweiten. Von den Medien werden die dynamische Preissetzung und ihr Einfluss auf die Preisstatistik vermehrt aufgegriffen und kontrovers diskutiert (Jung, 2017; Fischer und andere, 2017; Klemm, 2017; Hoffmann, 2018).

Aus der Perspektive der Verbraucherpreisstatistik erschwert dynamische Preissetzung die Messung einer realitätsnahen Preisentwicklung von Produkten im Internethandel. Da dieser weiter an Bedeutung gewinnt, steht die Verbraucherpreisstatistik vor einer besonderen Herausforderung. Das Ausmaß dynamischer Preissetzung im Internethandel ist allerdings bisher wenig untersucht worden (Cavallo, 2018). Im vorliegenden Aufsatz wird eine erste Auswertung des Ausmaßes von dynamischer Preissetzung im Internethandel vorgestellt. Zu diesem Zweck wurden die Häufigkeit und die Höhe von Preisänderungen bei Produkten untersucht, die für den deutschen Verbraucherpreisindex (VPI) und den Harmonisierten Verbraucherpreisindex (HVPI) gemessen werden.

1.2 Abgrenzung von dynamischer und individualisierter Preissetzung

Dynamische Preissetzung wird häufig als Sammelbegriff für Preisänderungen/-unterschiede im Internethandel verwendet, die durch die Anwendung von automatisierten Algorithmen hervorgerufen werden. Allerdings muss bei der Definition differenziert werden zwischen zeit- und personenabhängigen Preisunterschieden (Schleusener, 2016; Rimmel, 2016). Nur der erste Aspekt, Preisveränderungen im Zeitablauf, wird als dynamische Preissetzung definiert. Preisunterschiede für verschiedene Personen werden demgegenüber als individualisierte oder personalisierte Preissetzung bezeichnet. Bei individualisierter Preissetzung verlangen Internetgeschäfte entsprechend der jeweiligen Zahlungsbereitschaft ihrer Kundschaft verschiedene Preise.

Dieser Aufsatz beschäftigt sich ausschließlich mit dynamischer Preissetzung, deren Ausmaß und Auswirkungen auf die Verbraucherpreisstatistik analysiert werden. Eine Analyse von individualisierter Preissetzung wäre mit einem wesentlich größeren Aufwand verbunden.

1.3 Internetpreiserhebung für den VPI/HVPI

Der Warenkorb für den deutschen VPI/HVPI umfasst derzeit ungefähr 600 Arten von Waren und Dienstleistungen (Egner, 2013). Für die große Mehrheit dieser Güterarten werden im Rahmen der traditionellen Preiserhebung in stationären Geschäften und bei Dienstleistungsunternehmen monatlich mehr als 300 000 Einzelpreise erhoben. Die traditionelle Preiserhebung erfolgt in der Regel zu einem bestimmten Zeitpunkt des Monats. Das bedeutet: Für jedes Produkt einer Güterart fließt in der Regel genau ein Preis je Monat in die Berechnung von Durchschnittspreisen ein. Aus diesen werden dann Teilindizes und der gesamte VPI/HVPI errechnet. Für bestimmte Güter (insbesondere Dienstleistungen wie Flüge oder Pauschalreisen) werden Preise nach einem komplizierteren Muster von Buchungs- und Reisezeitpunkten erhoben; dieses Vorgehen erhöht die Beobachtungszahl deutlich. Die Preiserhebung im Internet wird aus Effizienzgründen zentral im Statistischen Bundesamt durchgeführt und findet zu einem bestimmten, immer gleich bleibenden Zeitpunkt des Monats statt. Die zentral erhobenen Preise dienen der Berechnung von Elementarindizes für den Internethandel; sie werden anschließend den Statistischen Ämtern der Länder zur weiteren Berechnung ihrer Verbraucherpreisindizes zur Verfügung gestellt.

Eine Geschäftstypengewichtung auf der Ebene der Waren und Dienstleistungen sowie der Bundesländer wurde 2008 in der deutschen Verbraucherpreisstatistik eingeführt (Sandhop, 2012). Die verwendeten Gewichte für Geschäftstypen ermöglichen es, die Anteile des Internethandels als eine eigenständige Geschäftstypenkategorie zu berechnen. Das Gewicht des gesamten Internethandels im deutschen VPI/HVPI beträgt derzeit etwas mehr als 5 % bezogen auf das Basisjahr 2010. Es ist zu erwarten, dass die Bedeutung in den nächsten Jahren zunehmen wird. Durch den wachsenden Internethandel werden Internetpreise mittlerweile für etwa 40 % aller im VPI/HVPI berücksichtigten Güterarten beobachtet und

für ungefähr 10 000 Produkte erhoben. Nicht enthalten sind hier im Internet erhobene Dienstleistungen, die jedoch nicht dem Geschäftstyp Internethandel zugeordnet werden. Allerdings ist zu beachten, dass die weitaus größere Anzahl an Preisen für die Berechnung des VPI/HVPI für Güter erhoben wird, die nicht oder nicht nur im Internet angeboten werden und deshalb aktuell noch nicht Ziel von dynamischer Preissetzung sein können.

2

Aufbau der Studie

2.1 Stichprobenbeschreibung

Der Stichprobenumfang für die vorliegende Studie umfasst ausschließlich Produkte im Internethandel, für die tatsächlich monatlich Preise für die Berechnung des VPI/HVPI erhoben werden. Jedes Produkt in der Stichprobe kann einer Güterart (sogenannte COICOP-10-Steller¹) zugeordnet werden. Insgesamt umfasst die Stichprobe Waren (keine Dienstleistungen) aus 242 COICOP-Positionen. Die Studie wurde begonnen mit einem Stichprobenumfang von 3 050 Produkten, verteilt auf 15 Internetgeschäfte. Die Auswahl der Internetgeschäfte erfolgte zum einen anhand ihrer Bedeutung für den deutschen Internethandel, zum anderen sollten sie ein Mindestmaß an Gütern anbieten, die für den VPI/HVPI von Bedeutung sind. Allerdings beschränkt sich die spätere Auswertung auf knapp 2 700 Waren und 14 Internetgeschäfte: Im Beobachtungszeitraum stellte ein Geschäft seine Tätigkeit ein, zudem waren einige Produkte nicht mehr verfügbar und konnten nicht ersetzt werden. Alle Preise wurden mithilfe von Web Scraping erfasst (siehe den folgenden Abschnitt). Dieses Verfahren wird seit einigen Jahren nicht nur vom Statistischen Bundesamt, sondern auch von anderen europäischen Statistikämtern für die Erhebung von Preisen im Internet verwendet (Griffioen und andere, 2014; Nygaard, 2015). Die Angebotsseite des jeweiligen Produkts im Internet wird dabei direkt und automatisiert angesteuert, vor allem um die Dauer der Erhebung zu minimieren.

1 COICOP: Classification of Individual Consumption of Purpose (Klassifikation des individuellen Verbrauchs nach Verwendungszwecken). Der COICOP-10-Steller ist die niedrigste Ebene der Klassifikation, die in der Verbraucherpreisstatistik verwendet wird, und stellt die Elementargruppen der Preiserhebung für den VPI/HVPI dar.

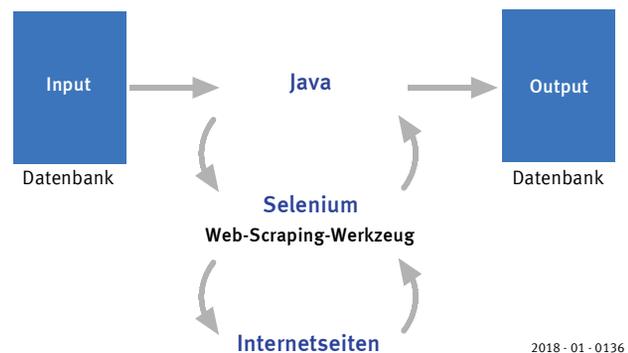
2.2 Technische Umsetzung des Web-Scraping-Programms^{1,2}

Die Eingangsdaten der Stichprobe [Produktname, Geschäft, COICOP-Nummer, COICOP-Name, Identifikationsnummer (ID), URL³ und Artikelnummer des Produkts] wurden in eine Tabelle einer relationalen Datenbank abgelegt. Des Weiteren wurde eine Tabelle für XPath angelegt. XPath geben die genaue Position von Informationen auf einer Internetseite (im HTML5-Format) an, welche dann extrahiert werden können. Die vom W3-Konsortium spezifizierte Abfragesprache XML Path Language (XPath) erlaubt es, Knoten genannte Teile eines HTML5- oder XML-Dokumentes zu adressieren und auszuwerten. Dabei kann ein XPath-Ausdruck genau einen oder auch mehrere Knoten auswählen. Ein Knoten kann ein Element (gegebenenfalls einschließlich Kinderelementen), einen Text (Wert) oder ein Attribut darstellen. Zusätzlich gibt es auch einen Knoten für das gesamte Dokument und Knoten für Namensräume, Verarbeitungsanweisungen und Kommentare. Im Normalfall besteht ein XPath aus einem eindeutigen Pfad zur jeweiligen Information in einem oder mehreren Knoten. Bei validen HTML5-Internetseiten (Dokumenten) identifiziert eine ID ein Element eindeutig. So kann statt eines kompletten absoluten Pfades vom Dokumentknoten der gewünschte Knoten gegebenenfalls auch direkt über seine ID identifiziert werden.

Für die automatisierte Erhebung wird die Programmiersprache Java eingesetzt. Für das tatsächliche Erheben („Scraping“) der Daten wird das Werkzeug Selenium verwendet, welches einen Internetbrowser steuert. Selenium kann als Plug-In für einen herkömmlichen Internetbrowser eingerichtet und dazu genutzt werden, Prozeduren im Internetbrowser aufzuzeichnen und abzuspielen. Die Instruktionen zur Navigation im Internet und zur Extraktion von Informationen werden in den Java-Code geschrieben und dazu die XPath aus der Datenbank verwendet. Die Nutzung eines herkömmlichen Internetbrowsers für die Navigation im Internet gestaltet sich effizient und hat mehrere Vorteile: Beispielsweise können bei der Verwendung eines herkömmlichen Internetbrowsers Skripte deaktiviert werden, damit Internetsei-

ten schneller geladen und innerhalb weniger Sekunden mehrere Internetseiten angesteuert werden können. Zusätzlich können mehrere Browserfenster gleichzeitig für die Preiserhebung eingesetzt werden. Dies führt zu einer großen Geschwindigkeitszunahme bei der Preiserhebung. → Grafik 1

Grafik 1
Technischer Aufbau der automatisierten Preiserhebung im Internet



2018 - 01 - 0136

2.3 Ziele und Grenzen der Untersuchung

Ziel der Studie war, die Frequenz und Höhe von Preisänderungen für im Internet angebotene Produkte festzustellen, die in die Berechnung des VPI/HVPI einfließen. Um ein detailliertes Bild für jedes Produkt und Informationen zur Preisgestaltung des jeweiligen Internetgeschäfts zu erhalten, wurden für die Studie stündlich Preise erhoben. Da eine Internetverbindung im Tagesverlauf nicht immer gleich stabil ist und Internetseiten unterschiedliche Ladezeiten haben, konnte es vorkommen, dass die Preiserhebungen für ein Produkt nicht genau 60 Minuten auseinander lagen. Die folgende Auswertung der Studie geht allerdings davon aus, dass alle Preiserhebungen mit dem exakt gleichen Zeitintervall erhoben wurden. Demnach können alle Produkte in der Stichprobe maximal eine Preisveränderung je Stunde aufweisen.

Für die automatisierte Preiserhebung wurden wechselnde IP-Adressen verwendet. Aus diesem Grund kann nicht untersucht werden, ob das vorherige Suchverhalten den Preis eines Produktes beeinflusst. Auch der Einfluss der Nutzung unterschiedlicher Geräte auf die Preiserhebung kann nicht analysiert werden, da alle Preise mit dem gleichen Gerät erfasst wurden.

² Für zusätzliche Erläuterungen siehe Blaudow (2018).

³ Uniform Resource Locator (Bezeichnungsstandard für Netzwerkressourcen).

3

Ergebnisse

3.1 Umgang mit Datenlücken

Die Preiserhebung für die Studie begann am 9. Dezember 2016. Preise, die stündlich bis zum 6. März 2017 erhoben wurden, fließen in den Hauptteil der Analyse ein. Dieser umfasst die Häufigkeit der Preisänderungen in allen Internetgeschäften und die Volatilität der Preisentwicklung. Das Programm sollte ursprünglich Preise für ein ganzes Kalenderjahr erfassen. Allerdings sind bei den meisten Geschäften im Laufe des Jahres viele Produkte ausgefallen, weil sie nicht mehr verfügbar waren. Für eine vergleichbare Auswertung hätten die ausgefallenen Produkte wie bei der tatsächlichen Preiserhebung ersetzt werden müssen. Ersetzungen konnten jedoch im Rahmen der automatisierten Erhebung nicht ohne Weiteres realisiert werden. Deshalb wurde der Untersuchungszeitraum auf drei Monate verkürzt. Beim Internetgeschäft, in dem der höchste Anteil an häufigen Preisänderungen beobachtet wurde, sind im Jahresverlauf nur wenige Produkte ausgefallen. Somit konnte hier eine zwölfmonatige Analyse der erhobenen Preise zwischen dem 9. Dezember 2016 und dem 9. Dezember 2017 durchgeführt werden (siehe Abschnitt 3.3).

In wenigen Fällen sind im Beobachtungszeitraum von drei Monaten bei der Preiserhebung Datenlücken aufgetreten. Ein Großteil solcher Datenlücken resultiert aus technischen Problemen. Um die Ergebnisse der Erhebung erfolgreicher auswerten zu können, wurden folgende Nachbearbeitungen vorgenommen:

- › Wenn die Preiserhebung eines Produkts eindeutig kurzfristig ausgefallen ist, also der Preis für kurze Zeit nicht erfasst wurde, füllte der letzte beobachtete Preis die Datenlücken. In den meisten Fällen kam es zu Ausfällen einzelner Preise, weil die Internetseite des Geschäfts gewartet wurde.
- › In die finale Analyse gingen nur die Produkte ein, für die mindestens in der Hälfte des Beobachtungszeitraums Preise vorlagen. Anderenfalls wäre die Vergleichbarkeit gegenüber Produkten, deren Preise im gesamten Beobachtungszeitraum beobachtet wurden, stark eingeschränkt. Dies ist einer der Hauptgründe,

warum die Studie mit über 3 000 Produkten begonnen wurde, aber nur etwa 2 700 Produkte für die Auswertung herangezogen werden.

- › Bei Produkten mit sehr auffallenden Preissteigerungen von über 500% oder Preissenkungen von fast 100% erfolgte eine manuelle Prüfung auf Plausibilität. In den meisten Fällen resultierte eine starke Preisänderung aus einem Produktwechsel. Der Preis wurde fälschlicherweise trotzdem erhoben, weil dem neuen Produkt die Artikelnummer des weggefallenen zugeordnet wurde oder weil das Erhebungsprogramm auf die Angebotsseite eines vermeintlich ähnlichen Produkts umgeleitet wurde. Für beide Fälle wurde das Produkt für die Analyse aus der Stichprobe entfernt, weil nicht davon ausgegangen werden kann, dass ein Geschäft ein Ersatzprodukt nach den in der Verbraucherpreisstatistik üblichen Kriterien auswählt.
- › Ein Internetgeschäft hat nach etwa zwei Monaten des Beobachtungszeitraums seine Webseite eingestellt. Da dieses Geschäft somit nicht mehr für den VPI/HVPI relevant ist, wurden dessen Produkte auch nicht für die Analyse genutzt.

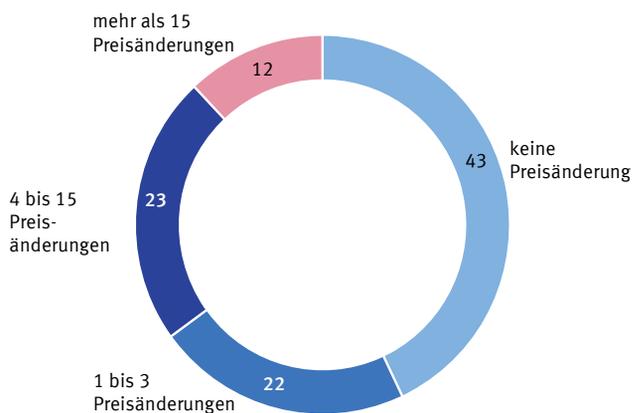
Nach der geschilderten Datenbereinigung gingen schließlich Preisreihen von 2 680 Produkten aus 14 Internetgeschäften in die Analyse ein. Die Geschäfte wurden in der Ergebnisdarstellung anonymisiert. Intern sind die Namen der Geschäfte natürlich von Bedeutung, da Handlungsempfehlungen für die Preiserhebung daraus abgeleitet werden können, zum Beispiel die Beobachtungszeitpunkte für bestimmte Produkte in bestimmten Geschäften auszuweiten.

3.2 Anzahl von Preisänderungen

In diesem Abschnitt wird die Häufigkeit von Preisänderungen bei den beobachteten Produkten während des Beobachtungszeitraums von drei Monaten analysiert. Für die Verbraucherpreisstatistik ist es wichtig zu erfahren, für wie viele Produkte sich die Preise so häufig ändern, dass eine traditionelle Preiserhebung die tatsächliche Entwicklung nicht mehr korrekt erfassen kann. Die zusammengefassten Ergebnisse zeigt [↗ Grafik 2](#).

Im Untersuchungszeitraum wiesen 43% der Produkte keine Preisänderungen auf. Für 22% der Produkte wurden die Preise ein- bis dreimal geändert. Diese Einteilung wurde gewählt, da 1 bis 3 Preisänderungen im

Grafik 2
Häufigkeit von Preisänderungen bei den beobachteten Produkten
in %



Beobachtungszeitraum: 3 Monate.

2018 - 01 - 0137

Quartal mit der traditionellen Preiserhebung im Internet zu einem Zeitpunkt im Monat grundsätzlich erfasst werden können. Erst wenn sich der Preis eines Produkts durchschnittlich mehr als einmal im Monat ändert, könnte das zu Verzerrungen in der Indexberechnung führen. Grund dafür ist, dass eine Preiserheberin oder ein Preiserheber nicht wissen kann, ob der beobachtete Preis repräsentativ für den betrachteten Monat ist. Die

Kategorie, die 4 bis 15 Preisänderungen zusammenfasst, enthält mit 23% fast ein Viertel aller Produkte in der Studie. Preisänderungen von Produkten, die in diese Kategorie fallen, werden durch die traditionelle Preiserhebung nur teilweise erfasst. Für etwa 12% der Produkte änderte sich der Preis mehr als 15-mal während des Beobachtungszeitraums. Das sind durchschnittlich mehr als fünf Preisänderungen im Monat. Generell sinkt die Wahrscheinlichkeit, im Rahmen der traditionellen Preiserhebung einen für den jeweiligen Monat repräsentativen Preis zu finden, je höher die Häufigkeit der Preisänderung im betrachteten Monat ist. Allerdings ist nicht nur die Häufigkeit der Preisänderungen relevant, sondern auch die Volatilität der Preisentwicklung.

Werden die Internetgeschäfte anhand der Häufigkeit der Preisänderungen ihrer Produkte verglichen, wird deutlich, dass bei den meisten Geschäften (Geschäfte 2, 3, 5, 7, 8, 10, 11, 12, 13) der Schwerpunkt der Häufigkeit der Preisänderungen zwischen 0 und 3 liegt. Wenn man dynamische Preissetzung als mehr als eine Preisänderung im Monat definiert, bedeutet das, dass dynamische Preissetzung im Beobachtungszeitraum nur von wenigen Geschäften in nennenswertem Umfang eingesetzt wurde. [↘ Tabelle 1](#)

Tabelle 1
Preisreihen je Geschäft, gruppiert nach der Anzahl von Preisänderungen

	Preisreihen mit ... Preisänderungen								
	0		1 bis 3		4 bis 15		mehr als 15		insgesamt
	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl	%	Anzahl
Geschäft 1	17	7,9	20	9,3	76	35,4	102	47,4	215
Geschäft 2	268	47,9	109	19,5	114	20,4	68	12,2	559
Geschäft 3	105	50,7	54	26,1	44	21,3	4	1,9	207
Geschäft 4	15	20,6	13	17,8	34	46,6	11	15,1	73
Geschäft 5	19	32,8	37	63,8	2	3,5	0	0,0	58
Geschäft 6	21	25,6	21	25,6	30	36,6	10	12,2	82
Geschäft 7	483	54,4	174	19,6	145	16,3	86	9,7	888
Geschäft 8	21	84,0	4	16,0	0	0,0	0	0,0	25
Geschäft 9	6	4,9	36	29,3	60	48,8	21	17,1	123
Geschäft 10	23	31,5	19	26,0	31	42,5	0	0,0	73
Geschäft 11	16	29,6	19	35,2	15	27,8	4	7,4	54
Geschäft 12	59	79,7	11	14,9	1	1,4	3	4,1	74
Geschäft 13	75	68,8	33	30,3	1	0,9	0	0,0	109
Geschäft 14	25	17,9	43	30,7	67	47,9	5	3,6	140
Insgesamt	1 153	43,0	593	22,1	620	23,1	314	11,7	2 680

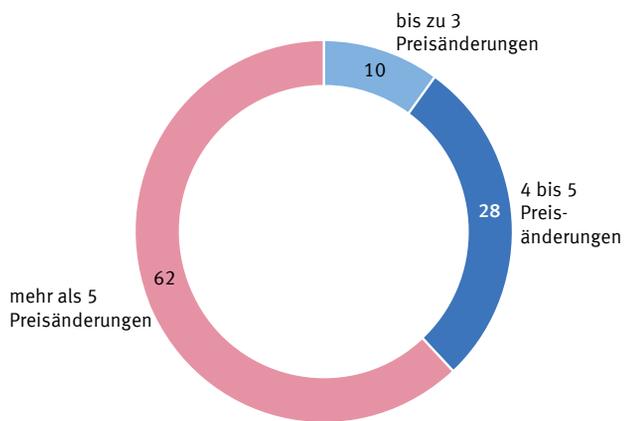
3.3 Detailbetrachtung von Produkten in Geschäft 1

Aus Tabelle 1 ist ersichtlich, dass in Geschäft 1 bei 47% der Produkte mehr als 15 Preisänderungen in drei Monaten gemessen wurden. Dieser Wert ist der bei Weitem höchste gemessene Wert für alle betrachteten Geschäfte. Da bei diesem Geschäft zusätzlich relativ wenige Produkte langfristig ausgefallen sind, wurden die Preisreihen der Produkte in diesem Geschäft genauer untersucht. Wird die Häufigkeit von Preisänderungen für den erweiterten Zeitraum von zwölf Monaten betrachtet, ist der Anteil an Produkten mit überdurchschnittlich vielen Preisänderungen noch höher als über den Zeitraum der ersten drei Monate. Für den Zeitraum von drei Monaten fielen 47% der Produkte in die Kategorie mit mehr als durchschnittlich fünf Preisänderungen im Monat, bei der Betrachtung von zwölf Monaten sind es sogar 62%.

↘ Grafik 3

Beispielhaft sei hier der Preisverlauf bei einem Produkt (Eau de Toilette) veranschaulicht, für das durchschnittlich fast jede zweite Stunde eine Preisänderung gemessen wurde. Ohne moderne Erhebungswerkzeuge sind

Grafik 3
Durchschnittliche Preisänderungen je Monat in Geschäft 1 in %



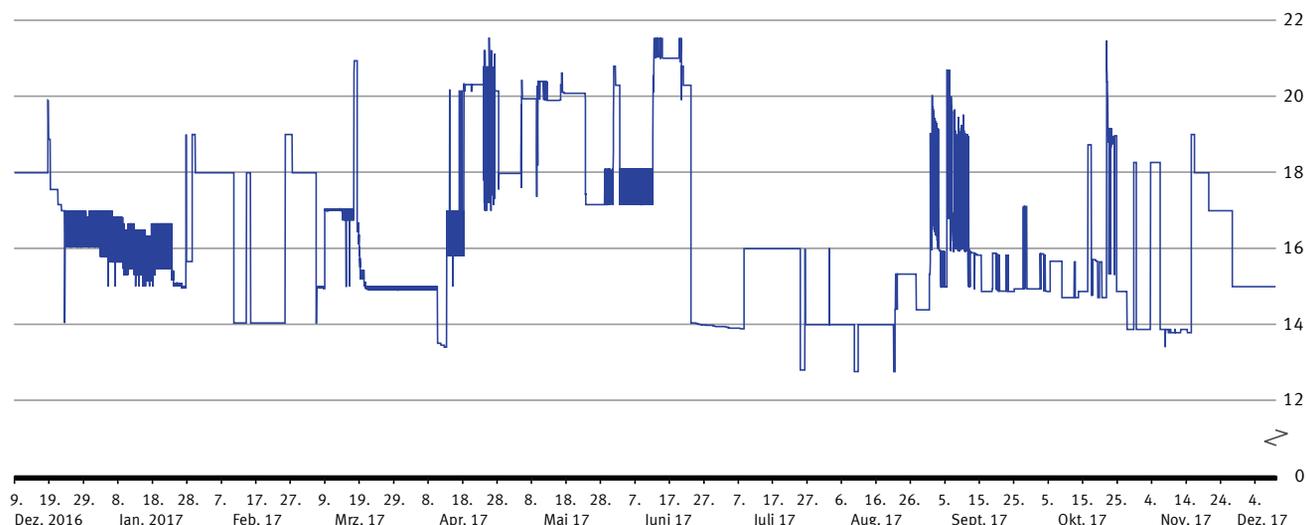
Beobachtungszeitraum: 12 Monate.

2018 - 01 - 0138

Preisänderungen solcher Art unmöglich zu erheben und nur schwer zu interpretieren. Saisonverläufe sind nicht zu erkennen, auch die Bandbreite der Angebotspreise ist kaum zu erklären. ↘ Grafik 4

Grafik 4

Extrembeispiel für Preisänderungen im Zeitraum von 12 Monaten: Eau de Toilette EUR



2018 - 01 - 0139

3.4 Volatilität der Preise

Betrachtet man die Fälle, bei denen sich Preise mehr als dreimal während des Beobachtungszeitraums geändert haben (durchschnittlich mehr als eine Preisänderung im Monat), ist auch die Volatilität zu berücksichtigen. Volatilität bezeichnet das Ausmaß der Preisänderung. Um die Volatilität von Preisen zu messen, wurde für jede Preisreihe der Variationskoeffizient berechnet, der die Standardabweichung ins Verhältnis zum Erwartungswert stellt. Der Variationskoeffizient wurde als Streuungsmaß gewählt, weil er ein relatives Maß ist und nicht vom Niveau des Preises abhängt. Diese Analyse hat einen wichtigen Einfluss auf die Beurteilung der traditionellen Preiserhebung. Ist die Volatilität von Preisen niedrig, ist dynamische Preissetzung kein großes Problem für die Inflationsmessung, da der Durchschnittspreis nicht durch große Preissprünge verzerrt wird. [↘ Tabelle 2](#)

Tabelle 2

Preisreihen mit mehr als drei Preisänderungen in drei Monaten, gruppiert nach dem Niveau des Variationskoeffizienten

	Variationskoeffizient			
	kleiner als 0,05	0,05 bis unter 0,1	0,1 bis 0,25	größer als 0,25
Geschäft 1	99	46	27	6
Geschäft 2	67	44	52	19
Geschäft 3	26	19	2	1
Geschäft 4	38	7	0	0
Geschäft 5	0	0	0	2
Geschäft 6	24	13	3	0
Geschäft 7	75	40	67	49
Geschäft 8	0	0	0	0
Geschäft 9	47	24	9	1
Geschäft 10	2	18	11	0
Geschäft 11	11	5	3	0
Geschäft 12	0	1	1	2
Geschäft 13	0	0	1	0
Geschäft 14	5	6	57	4
Insgesamt	394	223	233	84
	%			
Insgesamt	42	24	25	9

Bei Geschäft 1 zum Beispiel änderten sich für mehr als 80% der beobachteten Produkte die Preise mehr als dreimal im Beobachtungszeitraum, für fast 50% sogar mehr als 15-mal im Beobachtungszeitraum (siehe

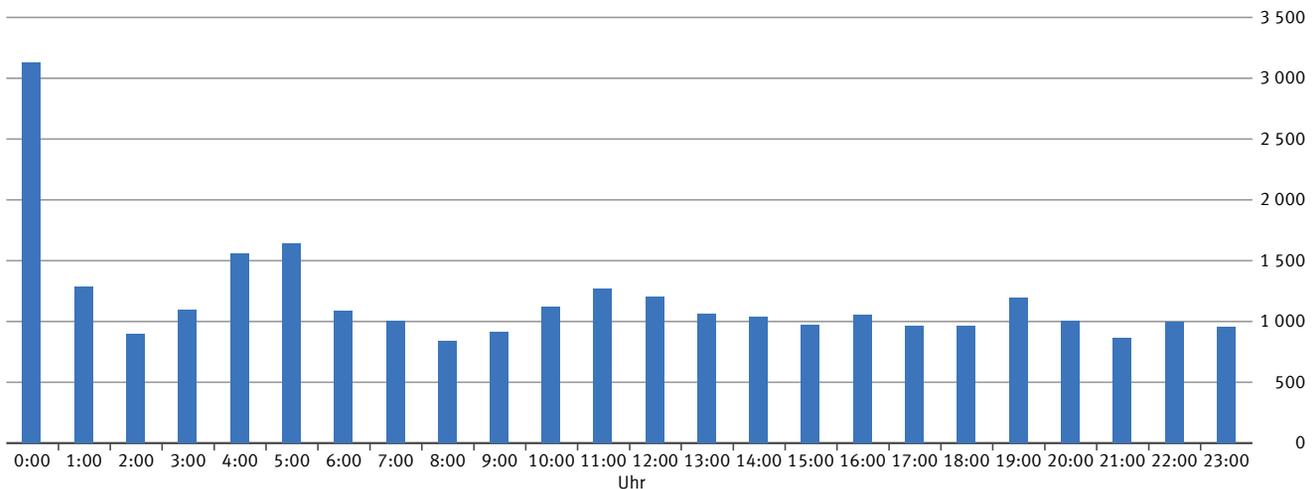
Tabelle 1). Für mehr als 80% der Preisreihen beträgt aber der Variationskoeffizient weniger als 0,1 und für 55% der Preisreihen sogar weniger als 0,05 (siehe Tabelle 2). Damit findet dynamische Preissetzung in diesem Geschäft zwar statt, aber die Volatilität für die meisten Preisreihen liegt in einem Bereich, der bei einer Ausweitung der Erhebungszeitpunkte der traditionellen Preiserhebung erfasst werden kann. Im Falle eines niedrigen Variationskoeffizienten ist es also unwahrscheinlich, dass durch die traditionelle Preiserhebung eine nennenswerte Verzerrung der Indexberechnung auftritt. Beispielhaft erwähnt sei ein Produkt, für das im Beobachtungszeitraum von drei Monaten 1 304-mal der Preis geändert wurde, allerdings in sehr geringer Höhe, was zu einem Variationskoeffizienten von lediglich 0,02 führte. Es können allerdings auch extreme Fälle auftreten, bei denen sich Produktpreise nahezu jede Stunde ändern und die gleichzeitig einen hohen Variationskoeffizienten aufweisen (siehe das Beispiel „Eau de Toilette“ in Grafik 4). Bei der Betrachtung der Preisreihen mit mehr als drei Preisänderungen in drei Monaten über alle Geschäfte sind 66% der Variationskoeffizienten geringer als 0,1 und 42% sogar geringer als 0,05. Es gab wenige Geschäfte in der Stichprobe, bei denen eine relativ hohe Anzahl an Preisänderungen mit relativ hohen Variationskoeffizienten einherging. Produktpreise des Geschäfts 14 änderten sich beispielsweise häufig (Tabelle 1) und wiesen gleichzeitig einen relativ hohen Variationskoeffizienten auf (Tabelle 2).

3.5 Uhrzeit von Preisänderungen

Betrachtet man die Zeitpunkte der Preisänderungen im Tagesverlauf, ist ein klares Muster mit einem Schwerpunkt zu Beginn des Tages, besonders zwischen Mitternacht und 1 Uhr, zu erkennen. Wahrscheinlich hat dies technische Gründe: Erstens ist bekannt, dass einige (eher einfache) Werkzeuge für die Durchführung von dynamischer Preissetzung Preisänderungen nur einmal am Tag erlauben. Des Weiteren könnten Internetgeschäfte Preisänderungen zu einem Zeitpunkt durchführen wollen, wenn sich möglichst wenige Kundinnen und Kunden auf ihren Internetseiten befinden. Morgens, nachmittags und abends finden weniger Preisänderungen statt. Grund hierfür ist möglicherweise, dass die meisten Menschen zu diesen Zeiten einkaufen und häufige Preisänderungen sie verärgern und abschrecken (Remmel, 2016). [↘ Grafik 5](#)

Grafik 5

Preisänderungen je Stunde im Tagesverlauf



Beobachtungszeitraum: 3 Monate.

2018 - 01 - 0140

4

Implikationen für die Verbraucherpreisstatistik

Informationen über Internetgeschäfte und ihre Preisgestaltung in Bezug auf Häufigkeit und Zeitpunkte von Preisänderungen sowie die Volatilität der Preisreihen sind sehr wichtig für die Steuerung der Preiserhebung und Indexberechnung. Diese Informationen ermöglichen eine effizientere Steuerung der Ressourcen für die Preiserhebung: Für die Preiserhebung in Internetgeschäften mit häufigen und hochvolatilen Preisänderungen sollten mehr Ressourcen eingesetzt werden als in Internetgeschäften mit wenigen und/oder wenig volatilen Preisänderungen.

Folgende Rückschlüsse können für die Preiserhebung der Verbraucherpreisstatistik gezogen werden:

Es gibt eindeutige Hinweise, dass dynamische Preissetzung derzeit von einigen Internetgeschäften in einem beachtlichen Umfang betrieben wird. Dies bedeutet eine Herausforderung für die Verbraucherpreisstatistik, der mit dem Einsatz neuer Erhebungswerkzeuge sowie der Nutzung neuer Datenquellen begegnet werden muss. Web Scraping ist ein passendes Werkzeug, um die monatliche Preiserhebung zu verbessern.

Die Studie hat ergeben, dass sich im Beobachtungszeitraum für zwei Drittel der beobachteten Produkte die Preise bis zu dreimal in drei Monaten (durchschnittlich eine Preisänderung im Monat) geändert haben. Diese Preisänderungen können durch die traditionelle Preiserhebung (eine Preisbeobachtung im Monat) in der Verbraucherpreisstatistik abgebildet werden. Für das restliche Drittel der Produkte in der Stichprobe müssen Wege gefunden werden, die Preise häufiger zu erfassen.

Für die Produkte, bei denen 4 bis 15 Preisänderungen in drei Monaten (durchschnittlich mehr als 1 bis 5 Preisänderungen im Monat) auftraten, könnte die traditionelle Preiserhebung ausgedehnt und Preise zu mehreren Zeitpunkten im Monat erfasst werden. Dagegen gilt für Produkte mit mehr als 15 Preisänderungen in drei Monaten (durchschnittlich mehr als 5 Preisänderungen im Monat): Es können Preiserhebungen nur mit modernen Werkzeugen wie Web Scraping in der Häufigkeit durchgeführt werden, die notwendig ist, um einen repräsentativen Preis für den jeweiligen Monat zu ermitteln.

Zu beachten ist dabei, dass sich häufig ändernde Preise dann kein größeres Problem für die Berechnung des VPI/HVPI darstellen, wenn der Variationskoeffizient niedrig ist. Erfahrungsgemäß kritisch einzuschätzen sind Produkte mit mehr als 3 Preisänderungen in drei Monaten (durchschnittlich mehr als eine Preisänderung im Monat), die gleichzeitig einen Variationskoeffizienten

von mehr als 0,05 aufweisen. Der Anteil dieser Produkte an allen beobachteten Produkten der Studie beträgt etwa 20%. Wird bei diesen Produkten die traditionelle Preiserhebung beibehalten, besteht die Gefahr, die Indexberechnung zu verzerren.

Ein weiterer Aspekt ist der Zeitpunkt von Preisänderungen. Die Preiserhebung sollte bevorzugt zu Zeiten erfolgen, wenn Konsumentinnen und Konsumenten Güter kaufen und die Preise vergleichsweise stabil sind, da davon auszugehen ist, dass diese Preise repräsentativ sind. Es sind jedoch weitere Analysen erforderlich, um die Haupteinkaufszeiten im Internet festzustellen.

Allerdings führt die Ausdehnung der automatisierten Preiserhebung auch zu methodischen Herausforderungen. Es ist schwierig bis unmöglich, Ersetzung und Qualitätsbereinigung zu automatisieren – insbesondere gilt dies für technische Produkte. Aus diesem Grund werden Ersetzungen und die Qualitätsbereinigung sehr wahrscheinlich weiterhin eine Aufgabe von Preiserheberinnen und Preiserhebern sowie von Produktexpertinnen und Produktexperten bleiben. Insgesamt wird sich die Arbeit der Preiserheberinnen und Preiserheber von der reinen Preiserhebung hin zu Plausibilitätsprüfungen und der Durchführung von Ersetzungen und Qualitätsbereinigungen verlagern. Darüber hinaus muss generell geprüft werden, ob die Berechnung von Durchschnittspreisen basierend auf sehr häufig erhobenen Preisen (zum Beispiel stündlich, wie in dieser Studie) fachlich sinnvoll ist (Mayhew, 2017; Breton und andere, 2016). Nicht alle Angebotspreise mögen repräsentativ für den Preisverlauf eines Produkts sein, zum Beispiel bei sehr hohen Preissprüngen oder in Zeiträumen, in denen typischerweise kaum eingekauft wird. Vor diesem Hintergrund erfordert die auf erhobenen Angebotspreisen basierende Berechnung von Durchschnittspreisen beispielsweise die Bereinigung von Ausreißern. Eine Lösung für dieses Problem wäre, Transaktionsdaten zu verwenden. Das skizzierte Problem bei der Durchschnittspreisberechnung würde sich dann nicht mehr stellen.

5

Fazit und Ausblick

Als wichtigstes Ergebnis dieser Studie ist festzuhalten, dass eine relevante dynamische Preissetzung derzeit bereits existiert, sich diese jedoch auf einige große Internetgeschäfte beschränkt. Dieses Ergebnis unterstützt eine differenzierte Vorgehensweise bei der Preismessung: eine Konzentration auf bestehende Werkzeuge (Web Scraping) und die Einführung zusätzlicher Prüfungen in Internetgeschäften mit häufigen Preisänderungen und zugleich einer hohen Volatilität der Preisreihen. Für die Preiserhebung in Internetgeschäften mit einer geringen Zahl von Preisänderungen und/oder einer geringen Volatilität der Preisreihen ist die traditionelle Preiserhebung zu einem Zeitpunkt im Monat weiterhin ausreichend.

Die Verwendung von Web Scraping ist eine sinnvolle und effiziente Methode, um in Zukunft die Preiserhebung im Internet auszudehnen, da mittelfristig weniger Ressourcen eingesetzt werden müssen als bei traditioneller Preiserhebung. Allerdings führt die Ausdehnung der automatisierten Preiserhebung auch zu methodischen Herausforderungen. Ersetzung und Qualitätsbereinigung sind schwer bis gar nicht zu automatisieren – insbesondere für technische Produkte.

Verschiedene stationäre Geschäfte in Deutschland haben in letzter Zeit, zumindest auf experimenteller Basis, digitale Preisschilder eingeführt. Sollte deren Einsatz in stationären Geschäften weiter zunehmen, wäre dynamische Preissetzung nicht nur eine Herausforderung für die Preiserhebung im Internet, sondern auch für die Preiserhebung im stationären Handel. Eine mögliche Lösung für diese Herausforderung könnte die Verwendung von Scannerdaten sein. Scannerdaten zeichnen die tatsächlichen Käufe eines Gutes durch Konsumentinnen und Konsumenten auf. Dieser Ansatz ist sehr vielversprechend und wird aktuell auch in Deutschland geprüft, war aber nicht Gegenstand der vorliegenden Studie.

Weiterführende Studien könnten sich zum einen auf die Untersuchung von individualisierter Preissetzung konzentrieren. Für die Analyse dieses Phänomens wäre es notwendig, verschiedene Nutzerprofile für die automatisierte Preiserhebung zu verwenden. Beispielfhaft zu

nennen wäre, verschiedene Geräte für die exakt gleiche Preiserhebung zu benutzen, oder unterschiedliche Suchverhalten entsprechend der vorher besuchten Internetseiten einzubeziehen.⁴ Zum anderen wäre eine langfristige Studie zu dynamischer Preissetzung mit anschließendem Vergleich der Ergebnisse aus automatisierter und traditioneller Preiserhebung wichtig. Aus diesem Grund wird derzeit an einer weiteren Studie mit über 10 000 Produkten bei etwa 400 Internetgeschäften gearbeitet. Hierbei sollen wieder stündlich Preise über einen Zeitraum von zwölf Monaten erhoben und anschließend analysiert werden. Die Preiserhebung wird erneut mit der Hilfe von Web Scraping umgesetzt, nach Möglichkeit sollen aber auch manuelle und automatisierte Ersetzungen bei Produktausfällen durchgeführt werden.

Zusammenfassend zeigt die vorliegende Studie, dass dynamische Preissetzung im Internet einerseits grundsätzlich eine Herausforderung für die Verbraucherpreisstatistik darstellt. Andererseits können derzeit immer noch fast zwei Drittel der im Rahmen der Studie beobachteten Preisänderungen von Produkten im Internethandel von der traditionellen Preiserhebung gut erfasst werden. Den häufigeren Preisänderungen für die restlichen Produkte wird entweder durch eine Ausdehnung der traditionellen Preiserhebung begegnet oder durch den Einsatz von modernen Werkzeugen wie Web Scraping. Neue wirtschaftliche Entwicklungen, etwa die dynamische Preissetzung, können durch eine Modernisierung des preisstatistischen Instrumentariums aufgefangen werden, sodass VPI und HVPI weiterhin als qualitativ hochwertige Statistikprodukte verfügbar sind. 

4 Es könnte den Preis beeinflussen, wenn eine Person die Internetseite eines Produkts mehrere Male besucht. Das Geschäft nimmt daraufhin verstärktes Interesse für dieses Gut an. Die Navigation zum Produkt könnte ebenfalls Auswirkungen auf die Preisgestaltung haben. Personen erreichen ein Geschäft, indem sie entweder die URL eintragen oder eine Suchmaschine verwenden. Letzteres kann als preisbewusstes Konsumverhalten interpretiert werden, welches zu einem geringeren Preis führen kann. Dieses Phänomen wurde während der traditionellen Preiserhebung für den VPI/HVPI festgestellt.

LITERATURVERZEICHNIS

Bladow, Christian. *Fortschritte und Herausforderungen beim Web Scraping – Automatisierung von Preiserhebungen im Internet*. In: Methoden – Verfahren – Entwicklungen. Ausgabe 1/2018, Seite 3.

Bladow, Christian/Burg, Florian. *Dynamic pricing as a challenge for Consumer Price Statistics*. 2017. [Zugriff am 2. März 2018]. Verfügbar unter: www.bundesbank.de

Breton, Robert/Flower, Tanya/Mayhew, Matthew/Metcalf, Elizabeth/Milliken, Natasha/Payne, Christopher/Smith, Thomas/Winton, Joe/Woods, Ainslie. *Research indices using web scraped data: May 2016 update*. Office for National Statistics (ONS). 2016.

Cavallo, Alberto. *Scraped Data and Sticky Prices*. In: Review of Economics and Statistics. Jahrgang 100. Ausgabe 1/2018, Seite 105 ff.

Egner, Ute. [Verbraucherpreisstatistik auf neuer Basis 2010](#). In: Wirtschaft und Statistik. Ausgabe 5/2013, Seite 329 ff.

Fischer, Malte/Goebel, Jacqueline/Hielscher, Henryk. *Der Preis ist heiß*. In: Wirtschaftswoche. Nr.10/2017, Seite 18 ff.

Griffioen, Robert/De Haan, Jan/Willenborg, Leon. *Collecting clothing data from the Internet*. Den Haag 2014.

Hoffmann, Catherine. *Schwer zu fassen*. In: Süddeutsche Zeitung. 4. Januar 2018, Seite 18.

Jung, Alexander. *Die Preis-Frage*. In: Der Spiegel. Nr. 9/2017, Seite 76 f.

Klemm, Thomas. *Misstraut der Inflation!* In: Frankfurter Allgemeine Sonntagszeitung. Nr. 34/2017, Seite 29.

Mayhew, Matthew. *ONS methodology working paper series number 12 – a comparison of index number methodology used on UK web scraped price data*. 2017.

Nygaard, Ragnhild. *The use of online prices in the Norwegian Consumer Price Index*. Tokio 2015.

Rommel, Johannes. *Die verbraucherpolitische Perspektive: aktuelle Entwicklungen im Online-Handel*. In: ZBW/Wirtschaftsdienst 2016/12, Seite 875 ff.

Sandhop, Karsten. [Geschäftstypengewichtung im Verbraucherpreisindex](#). In: Wirtschaft und Statistik. Ausgabe 3/2012, Seite 266 ff.

Schleusener, Michael. *Dynamisch und personalisiert: Wie entwickelt sich die Preissetzung im Online-Handel?* In: ZBW/Wirtschaftsdienst 2016/12, Seite 868 ff.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Sabine Bechtold

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im April 2018

Das Archiv aller Ausgaben ab Januar 2001 finden Sie unter www.destatis.de/publikationen

Print

Einzelpreis: EUR 18,- (zzgl. Versand)

Jahresbezugspreis: EUR 108,- (zzgl. Versand)

Bestellnummer: 1010200-18002-1

ISSN 0043-6143

ISBN 978-3-8246-1069-3

Download (PDF)

Artikelnummer: 1010200-18002-4, ISSN 1619-2907

Vertriebspartner

IBRo Versandservice GmbH

Bereich Statistisches Bundesamt

Kastanienweg 1

D-18184 Roggentin

Telefon: +49 (0) 382 04 / 6 65 43

Telefax: +49 (0) 382 04 / 6 69 19

destatis@ibro.de

Papier: Metapaper Smooth, FSC-zertifiziert, klimaneutral, zu 61% aus regenerativen Energien

© Statistisches Bundesamt (Destatis), 2018

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.