



**Malte Hansen**

ist Entwicklungsökonom mit Abschlüssen der Universitäten Göttingen und Stellenbosch (Südafrika). Seit September 2017 arbeitet er im Statistischen Bundesamt, derzeit als wissenschaftlicher Mitarbeiter im Referat „Verbraucherpreise“. Er untersucht vorwiegend die dynamische Preissetzung im Onlinehandel zur Weiterentwicklung des Verbraucherpreisindex.

# DYNAMISCHE PREISSETZUNG IM ONLINEHANDEL: ZUR LANGFRISTIGEN ANWENDUNG VON AUTOMATISIERTER PREISERHEBUNG

Malte Hansen

📌 **Schlüsselwörter:** Verbraucherpreisindex – dynamische Preissetzung – Web Scraping – automatisierte Preiserhebung – Onlinehandel

## ZUSAMMENFASSUNG

Durch die Digitalisierung im Onlinehandel ist seit einigen Jahren ein neues Preissetzungsverhalten der Anbieter zu beobachten: dynamische Preissetzung. Dabei verwenden Anbieter maschinelle Algorithmen, um Preise von Produkten laufend an veränderte Marktsituationen anzupassen. Für die Verbraucherpreisstatistik stellt sich damit die Frage, inwieweit Anpassungen in der Datenerhebung für den Verbraucherpreisindex erforderlich sind. Das Statistische Bundesamt arbeitet deshalb seit einiger Zeit an der Automatisierung der Preiserhebung mittels Web Scraping. In diesem Zusammenhang wurde eine Studie durchgeführt, um optimale Handlungsempfehlungen für die langfristige Nutzung des Web-Scraping-Programms zu ermitteln. Der Aufsatz stellt das methodische Vorgehen und die Ergebnisse dieser Studie vor.

📌 **Keywords:** consumer price index – dynamic pricing – web scraping – automated price collection – online trade

## ABSTRACT

*As the result of digitalisation in online trade, a new price setting behaviour has been observed for several years: dynamic pricing. Vendors use automatic algorithms in order to adapt product prices continuously to changing market situations. Regarding consumer price statistics, the question arises whether adjustments are required in the process of collecting data for the consumer price index. For several years, the Federal Statistical Office has been working on the automation of price collection, using web scraping. A study has been conducted to provide recommendations for the optimal and long-term use of the web scraping program. This article presents the methodical approach and the findings of the study.*

## 1

### Einleitung

Seit vielen Jahren ist die im Tagesablauf variable Preissetzung von Treibstoffen an Tankstellen bekannt: Zu bestimmten Stoßzeiten sind die Preise höher als zu Standardzeiten. Würde an einem Tag beziehungsweise in einem Monat der Standardpreis erhoben und bei der nächsten Erhebung der erhöhte Preis zur Stoßzeit, würde die Preisstatistik fälschlicherweise einen Preisanstieg melden. Eine ähnliche Problematik ergibt sich durch die dynamische Preissetzung im Onlinehandel. Um diese und weitere Entwicklungen sowie ihre Auswirkungen auf die Preisstatistik zu untersuchen, führte das Statistische Bundesamt eine umfangreiche Studie durch. Deren Aufbau und Ergebnisse stellt der folgende Beitrag vor.

Durch die zunehmende Digitalisierung verlagert sich das Kaufverhalten in Deutschland verstärkt zum Onlinehandel. Die Wirtschaftsankunft Creditreform Boniversum GmbH (Boniversum) und der Bundesverband E-Commerce und Versandhandel Deutschland (bevhd) führten im Jahr 2017 eine Verbrauchermfrage durch. Demnach stieg der Umsatz im Onlinehandel im Jahr 2016 um über 12% gegenüber dem Vorjahr auf rund 52,7 Milliarden Euro. Sein Anteil am Gesamtumsatz im Einzelhandel belief sich damit auf 10,4% (Statistisches Bundesamt, 2020). Im Jahr 2017 erhöhte sich dieser Anteil auf 11,3%. Und immer mehr Käuferinnen und Käufer erwerben vermehrt im Internet die für den Verbraucherpreisindex relevanten Artikel. So berichtet der Spiegel (Wahnbaek, 2019), dass im Vorweihnachtsgeschäft 2019 mehr als ein Drittel der Deutschen beabsichtigten, ihre Geschenkkäufe online zu tätigen.

Der Verbraucherpreisindex (VPI) wird monatlich berechnet und veröffentlicht. Seine Veränderungsrate zum Vormonat gibt die durchschnittliche Preisentwicklung aller Konsumausgaben von privaten Haushalten im Inland an. Diese Veränderungsrate wird allgemein als Inflationsrate bezeichnet. Der Verbraucherpreisindex umfasst etwa 300 000 Artikel, die repräsentativ für alle in Deutschland zu erwerbenden Konsumgüter stehen (Statistisches Bundesamt, 2018). Etwa 10 000 dieser Artikel repräsentieren den Onlinehandel. Für die korrekte Berechnung der Entwicklung des Verbraucherpreisindex wird je Artikel mindestens einmal im Monat

ein Preis erhoben. Dabei wird angenommen, dass dieser Preis repräsentativ für den Monatspreis des Artikels ist. Vor dem Hintergrund der dynamischen Preissetzung im Onlinehandel muss diese Annahme hinterfragt und die Datenerhebung gegebenenfalls angepasst werden.

Darüber hinaus kann eine angepasste Datenerhebung die Genauigkeit der Preisreihen sowie ihren Informationsgehalt erhöhen. Die adäquate Darstellung der monatlichen Preisentwicklung durch den Verbraucherpreisindex bleibt damit nicht nur gewährleistet, sondern wird sogar noch verbessert. Für die Studie zur dynamischen Preissetzung im Onlinehandel wurden Preise über einen Zeitraum von zwölf Monaten erfasst. Mit 8 624 Artikeln umfasst sie einen Großteil der im Verbraucherpreisindex hinterlegten 10 000 Artikel. Die Erkenntnisse führen zu Handlungsempfehlungen für die langfristige Nutzung des Web-Scraping-Programms, mit dem die automatisierte Datenerhebung durchgeführt wird.<sup>1</sup>

Neben dem Statistischen Bundesamt experimentieren auch andere nationale Statistikämter in Europa mit Web-Scraping-Daten oder haben diese bereits in ihre VPI-Berechnungen integriert. Das Statistikamt des Vereinigten Königreichs (ONS) benutzt Web-Scraping-Daten für mehrere Produktgruppen (Flower/Karachalias, 2019). Für die meisten Produktgruppen werden wöchentliche Erhebungen durchgeführt. Für einige Ausnahmen finden tägliche Erhebungen statt. Auch für den belgischen Verbraucherpreisindex werden die Preise wöchentlich oder täglich erhoben (van Loon/Roels, 2018).

Der weitere Aufsatz gliedert sich wie folgt: In Kapitel 2 wird die dynamische Preissetzung definiert und der Aufbau der Studie vorgestellt. Im Anschluss werden die Rohdaten auf ihre Verwertbarkeit analysiert und die angewandten Aufbereitungsschritte erläutert. Kapitel 3 beschreibt die Ergebnisse hinsichtlich der Anzahl der Preisänderungen, deren zeitliches Auftreten und gibt Handlungsempfehlungen zur automatisierten Preiserhebung. Dabei wird insbesondere auf Kalender- und Saisoneffekte eingegangen. Zudem werden die Ergebnisse zusammengefasst und die Umsetzbarkeit ihrer Implikationen diskutiert. Die Ausführungen schließen mit einem Fazit und Ausblick in Kapitel 4.

<sup>1</sup> Für eine detaillierte Beschreibung der automatisierten Datenerhebung mit dem Web-Scraping-Programm und dessen Funktionsweisen siehe Bladow/Seeger (2019).

## 2

### Methodisches Vorgehen

#### 2.1 Definition und Aufbau der Studie

Dynamische Preissetzung wird in dieser Studie wie folgt definiert: Dynamische Preissetzung liegt bei einem Artikel vor, wenn 3 % oder mehr der Preise eines Artikels über dessen Betrachtungszeitraum Änderungen zeigen. Dabei wurde von 8736 erhobenen Preisen ausgegangen, sofern der Preis eines Artikels über alle 52 Wochen eines kompletten Kalenderjahrs zu jedem Tag und jeder Stunde erhoben werden konnte. Ändern sich 3 % dieser Preise, liegen 262 Preisänderungen vor. Das ergibt – auf einzelne Wochen umgerechnet – fünf Preisänderungen im arithmetischen Mittel.

Einige Produktgruppen stehen zu Sonder- und Feiertagen im Fokus der Konsumenten und werden häufiger durch Rabattaktionen im Preis heruntergesetzt. Dazu gehören beispielsweise elektronische Unterhaltungsartikel rund um den Black Friday, den Tag nach Thanksgiving in den Vereinigten Staaten. Dadurch zählen diese Produktgruppen relativ viele Preisänderungen zu bestimmten Zeiten im Jahr. Im restlichen Verlauf des Jahres sind die Preisreihen jedoch stabil. Solch eine Dynamik soll nicht als dynamische Preissetzung verzeichnet werden, da in der manuellen Preiserhebung die Sonder- und Feiertage schon speziell betrachtet werden. Mit der 3-%-Regel wird versucht, dies abzudecken.

Neue Technologien wie der Onlinehandel vereinfachen das größtenteils automatisierte Preissetzungsverhalten der Warenanbieter. Die dynamische Preissetzung ist damit zu einem Phänomen geworden, welches Anpassungen in der Preiserhebung erfordern kann. Vor diesem Hintergrund wird in der Preisstatistik an der Einführung eines automatisierten Preiserhebungsprogramms gearbeitet. Durch die Untersuchung des Preissetzungsverhaltens kann festgestellt werden, welche Anbieter ihre Artikelpreise dynamisch setzen. Deren Preise sollten in erhöhter Frequenz abgefragt werden. Ziel dieser Studie ist es, diese Anbieter zu identifizieren und für sie entsprechende Erhebungsintervalle zu bestimmen. Darüber hinaus soll das Web-Scraping-Programm durch die hier gewonnenen Erkenntnisse zu bestimmten Tages-

und Wochenzeiten auch langfristig statistisch validierte Artikelpreise für den Verbraucherpreisindex liefern können. Für ein möglichst detailliertes Bild war deshalb die Erhebungsfrequenz, anders als bei der traditionellen, manuellen Preiserhebung, von monatlich auf stündlich erhöht.

Für die stündliche Erhebung wurde in der Inputdatenbank des Web-Scraping-Programms für jeden Konsumartikel genau eine URL (Uniform Resource Locator) zugewiesen.<sup>12</sup> Bei jedem einzelnen Scraping-Vorgang passiert dann Folgendes: Automatisiert wird an jedem Tag und in jeder Stunde je Artikel eine URL in einem Web-Browser-Fenster abgefragt. Wird die URL erreicht, liest das Web-Scraping-Programm den jeweiligen Artikelpreis, dessen Name und die Artikelnummer heraus. Beim Artikelpreis wurde zwischen Normal- und Sonderpreis unterschieden. War für einen Artikel ein reduzierter Sonderpreis ausgewiesen, so wurde versucht, den entsprechenden Normalpreis ebenfalls zu erfassen, um anschließend eine Plausibilitätsprüfung durchzuführen.<sup>13</sup>

Ein potenzielles Problem bei der Online-Datenerhebung von Artikelpreisen ist das Auftreten von individualisierter Preissetzung. Dabei versuchen Anbieter, den einzelnen, potenziellen Kundinnen und Kunden aufgrund ihres Verhaltens auf der Internetseite einen möglichst hohen, noch von ihnen akzeptierten und individuellen Preis anzuzeigen. Das Verhalten der Kundinnen und Kunden ergibt sich unter anderem aus der Anzahl der Seitenaufrufe beim jeweiligen Onlinehändler, der Verweildauer auf der jeweiligen Seite und das Aufrufen der einzelnen Artikel beim jeweiligen Händler. Um diesen Effekt für die Studie ausschließen zu können, wurden drei Maßnahmen ergriffen: 1. Die Anfragen an die Artikelseiten wurden nicht von einer einzelnen IP(Internet-Protocol)-Adresse

- 2 Auf diese Weise konnte das Programm die einzelnen Artikelseiten direkt ansteuern und musste nicht durch die Internetseiten der einzelnen Onlinehändler navigieren. Ansonsten hätten die einzelnen Scraping-Vorgänge für eine stündliche Preiserhebung zu lange gedauert. Der Einsatz des Web-Scraping-Programms wurde somit vollständig auf dessen Geschwindigkeit optimiert.
- 3 Internetseiten sind meistens im sogenannten HTML-Format aufgebaut. Die einzelnen Elemente einer Internetseite im HTML-Format lassen sich über sogenannte XPath's genau ansteuern. Ein XPath gibt die Position und Funktion eines Objekts innerhalb eines HTML-Dokuments an. So ist zum Beispiel ein Artikelpreis durch seinen eigenen XPath als ein bestimmtes Objekt definiert und lässt sich auf der Internetseite genau eingrenzen und auslesen. Zur Lokalisierung der Positionen auf der Internetseite mussten für jeden Onlinehändler XPath's in der Inputdatenbank hinterlegt werden. Die XPath's konnten für alle Artikel eines Onlinehändlers genutzt werden, da die einzelnen Artikelseiten je Onlinehändler gleich aufgebaut sind.

gestartet, sondern es wurde zu jeder Stunde eine neue IP aus einer Liste von mehreren IPs zufällig ausgewählt. Auf diese Weise erschien das Web-Scraping-Programm nicht als einzelner Kunde mit einer IP-Adresse. 2. Vor jedem stündlichen Scraping-Vorgang wurden die Cookies im Internetbrowser gelöscht, um hinterlegte Daten zu Seitenaufrufen zu entfernen. 3. Die einzelnen URLs wurden zu jeder Stunde in zufälliger Reihenfolge aufgerufen. Der Internetbrowser öffnete mehrere Browserfenster, auf die die Seitenaufrufe aufgeteilt waren.

## 2.2 Datenanalyse und Datenaufbereitung

Der Rohdatensatz umfasst mehr als 90% der für den Verbraucherpreisindex relevanten 10 000 Artikel. Die Artikel werden zur Berechnung des Verbraucherpreisindex zu Produktgruppen der Classification of Individual Consumption by Purpose (COICOP) zusammengefasst. Die Artikel sind auf 324 COICOP-10-Steller, die unterste Gliederungsebene, aufgeteilt. Der Verbraucherpreisindex umfasst insgesamt etwas mehr als 600 COICOP-10-Steller (Statistisches Bundesamt, 2018). In der VPI-Datenbank sind rund 300 Onlinehändler vertreten. Zu 273 Onlinehändlern wurde mindestens ein Preis erhoben. Der Erhebungszeitraum sollte das komplette Jahr 2019 umfassen. Die Preiserhebung begann bereits Ende 2018, damit der reibungslose Ablauf der Datenerhebung möglichst zu Beginn des Jahres 2019 erreicht werden konnte. Aufgrund der dennoch eingetretenen

Anlaufschwierigkeiten in den ersten Wochen 2019 wurden auch Preise für die erste Januarwoche 2020 erhoben, um die Preisentwicklung nach Weihnachten und Neujahr mit besserer Datenlage analysieren zu können.

### ↘ Tabelle 1

Die erhobenen Daten sind Rohdaten, die aufbereitet und plausibilisiert werden müssen. Für die Studie wurden zwei Gruppen von Problemfällen identifiziert, die getrennt zu beachten sind: zu editierende Preise und unplausible Preisänderungen. Zu editierende Preise sind zum Beispiel solche, bei denen festgestellt wurde, dass das Web-Scraping-Programm einen Onlinepreis fehlerhaft ausgelesen hat. Dieser Fehler ließ sich durch die Korrektur der Kommastelle beheben. In anderen Fällen lag der Normalpreis unter dem Sonderpreis, sodass der Normalpreis anstatt des Sonderpreises eingesetzt wurde.

Unplausible Preisänderungen wurden durch zwei Berechnungsmethoden identifiziert. Für die erste Methode wurden für jeden Artikel prozentuale Preisänderungen zum vorigen Preis ermittelt. Unplausible Preisänderungen sind in dieser Studie definiert als Änderungen, die um 400 % über oder um 80 % unter dem vorangegangenen Preis liegen. Für die zweite Berechnungsmethode wurde für jede Preisreihe ein 95-%-Konfidenzintervall um das arithmetische Mittel der Preisänderungen über den gesamten Zeitraum errechnet. Liegt eine Preisänderung um 400 % über der Obergrenze oder um 80 % unter der Untergrenze des Konfidenzintervalls, wird diese als unplausibel markiert und aussortiert.

Nach der Datenaufbereitung waren für die verbliebenen 8 624 unterschiedlichen Artikel fast 42 Millionen Artikelpreise von 241 Onlinehändlern verwendbar. Die Preise wurden zur Indexberechnung auf 324 COICOP-10-Steller aufgeteilt. Zur Untersuchung des Preissetzungsverhaltens wurden ausschließlich erhobene Preise betrachtet. Fehlende Preise wurden nicht imputiert, um die Ergebnisse nicht zu verfälschen.<sup>4</sup> Aufgrund der vorliegenden Massendaten stellen fehlende Werte mit einem Anteil von über 25 % keinen Informationsverlust dar, der die Analyse der Zeitreihen beeinträchtigen würde. Die Ausfälle der Preise sind zufällig und verzerren deshalb nicht die vorhandenen Daten des Parameters von Interesse:

**Tabelle 1**

**Erhobene Preise nach dem Erhebungsmonat**

	Anzahl der Preise
Dezember 2018	342 562
Januar 2019	2 289 648
Februar 2019	2 668 073
März 2019	3 442 491
April 2019	2 697 063
Mai 2019	3 571 108
Juni 2019	3 214 571
Juli 2019	3 693 373
August 2019	4 025 614
September 2019	3 789 103
Oktober 2019	3 836 901
November 2019	3 473 676
Dezember 2019	3 457 040
Januar 2020	1 167 409
Insgesamt	41 668 632

4 In dieser Analyse werden vorrangig Veränderungen und Veränderungsraten von Artikelpreisen betrachtet. Je nach Wahl der Imputationsmethode wären diese Parameter „mutwillig“ verändert worden und hätten keine rein objektiven Beobachtungen mehr dargestellt.

den Artikelpreis. Generell gibt es keine Regel, ab welchem Anteil fehlender Werte eine Imputation notwendig ist und lückenhafte Zeitreihen nicht berücksichtigt werden dürfen (Little/Rubin, 2002).

▾ **Tabelle 2** zeigt den Aufbau des Datensatzes. Zu sehen sind unter anderem die Anzahl der plausiblen und unplausiblen

Preise und deren Anteile an allen Beobachtungen. Sowohl die berücksichtigten, editierten Preise als auch die aussortierten, unplausiblen Preise haben mit 1,25% beziehungsweise weniger als 1% jeweils nur einen sehr geringen Anteil an den insgesamt erfassten Preisen. Aufgrund der großen Datenmengen wurde davon abgesehen, die Abschneidegrenzen für unplausible Preisänderungen enger zu definieren – es war nicht abzuschätzen, wie viele Preise zu Unrecht durch diese Maßnahme aussortiert worden wären.<sup>15</sup>

### 3

## Ergebnisse

Die folgende Ergebnisanalyse berücksichtigt zwei Aspekte: die Anzahl der Preisänderungen (Abschnitt 3.1) und das zeitliche Auftreten der Preisänderungen (Abschnitt 3.2).

### 3.1 Anzahl der Preisänderungen

▾ **Grafik 1** gruppiert die Onlinehändler über den gesamten Erhebungszeitraum in sechs Volatilitätsgruppen nach Anzahl der anteiligen Preisänderungen. Volatili-

<sup>5</sup> Zum Vergleich der gewählten Abschneidegrenzen haben Blaudow/Burg (2018) ihre prozentualen Veränderungen deutlich weiter definiert. Eine Preisänderung gilt als auffällig, wenn ein gegebener Preis 500% über oder fast 100% unter dem vorigen Preis liegt. Außerdem beschreiben Blaudow/Seeger (2019), wie in Zukunft das Web-Scraping-Programm selbst Datenvalidierung beim Scraping-Vorgang betreiben soll, um die Erhebung unplausibler Preise zu reduzieren.

**Tabelle 2**  
Übersicht über die Datensatzstruktur

	Beobachtungen	Anteil an den Rohpreisen	Filter/Kommentar
	1 000	% <sup>1</sup>	
Alle Scraping-Vorgänge	rund 63 800		
Rohdaten	57 523		ohne leere Zellen beziehungsweise fehlende URL
Fehlende Preise	15 087		nicht berücksichtigt
Verbleibende Rohpreise	42 435	100	
Plausible und editierte Preise	41 669	98,19	verwendeter Datensatz; um unplausible Preise und Ausreißer bereinigt
Editierte Preise	528	1,25	
Unplausible Preisänderungen	260	0,63	nicht berücksichtigt

<sup>1</sup> Differenz zu 100% rundungsbedingt.

tätsgruppe 1 umfasst die Händler, die überhaupt keine Preisänderungen vornahmen, während die Händler in Volatilitätsgruppe 6 mehr als 3% ihrer Preise änderten. Einen ersten Überblick über die Anzahl der Preisänderungen auf der Ebene der Onlinehändler gibt die linke Seite von Grafik 1. Für einen Großteil der Onlinehändler wurden keine oder fast keine Preisänderungen (Gruppen 1 und 2) beobachtet. Diese haben einen Anteil von bis zu 0,2% der Preisänderungen an ihren jeweils erhobenen Preisen. Etwa 96% aller Onlinehändler zeigen nach der genannten Definition keine dynamische Preissetzung. Demnach scheint das derzeitige Erhebungsintervall von einmal im Monat ausreichend zu sein, um die monatliche Preisentwicklung genau darzustellen. Mehr als 3% anteilige Preisänderungen wurden bei zehn Onlinehändlern (4% der Gesamtheit) beobachtet, sie setzen Preise damit dynamisch.

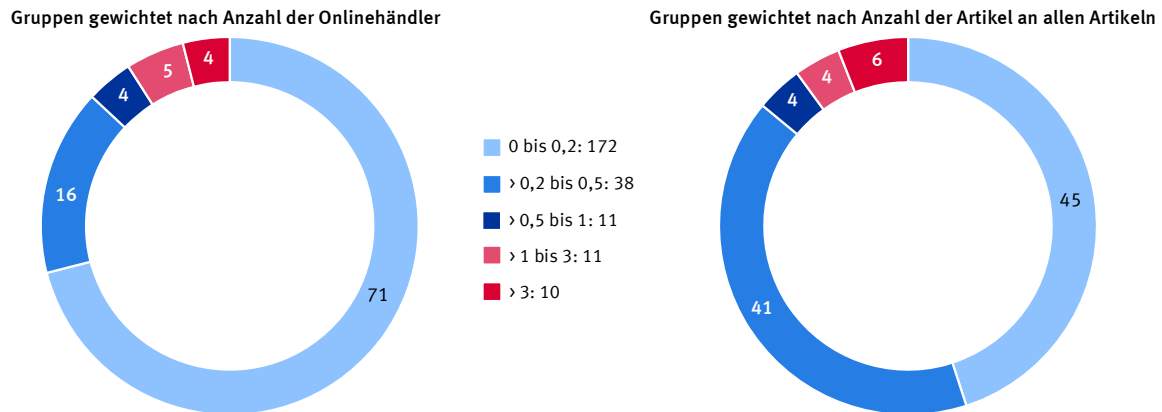
Der rechte Teil von Grafik 1 zeigt die Händler in den gleichen Volatilitätsgruppen über den gesamten Erhebungszeitraum gewichtet nach dem Anteil ihrer Artikel im Verhältnis zur Gesamtanzahl der Artikel. Die Händler der Gruppen 1 und 2, ohne oder mit sehr wenigen Preisänderungen, haben nun einen deutlich geringeren Anteil von weniger als 50% anstatt über 70% in der Gewichtung nach ihrer Anzahl. Die Gruppe 6, mit dynamischer Preissetzung, hat jetzt einen erhöhten Anteil von 6% anstatt 4%. Die Anpassung der Scraping-Intervalle wird für diese Händler vor dem Hintergrund der Datenqualität für den Verbraucherpreisindex empfohlen. Der Anteil ihrer Artikel an der Gesamtanzahl an Artikeln ist nunmehr statistisch signifikant. Mit der wachsenden Bedeutung des Onlinehandels kann davon ausgegangen werden, dass die Anzahl der VPI-relevanten Onlinehändler weiter stei-



# Dynamische Preissetzung im Onlinehandel: zur langfristigen Anwendung von automatisierter Preiserhebung

**Grafik 1**

Onlinehändler nach Volatilitätsgruppen<sup>1</sup> der Preisänderungen über den gesamten Erhebungszeitraum in %



1 Aus Darstellungsgründen wurden die Gruppen 1 (keine Preisänderungen) und 2 (bis 0,2 %) zusammengefasst.

2020 - 01 - 0211

gen wird, die mittels automatisierter Algorithmen ihre Artikelpreise ändern.

Zur genaueren Betrachtung der Onlinehändler mit den größten Anteilen an Preisänderungen sind diese in [Tabelle 3](#) anonymisiert gelistet. Die Onlinehändler sind nach der Anzahl abgedeckter Artikel absteigend sortiert. Die Tabelle listet alle Händler, die nach der 3%-Definition dynamisches Preissetzungsverhalten zeigen. Von den 241 erfassten Händlern erfüllen zehn diese Definition. Es fällt auf, dass acht der zehn Onlinehändler die 3%-Marke deutlich überschreiten. Von besonderem Interesse ist außerdem Händler 1, bei dem mehr als 1,2 Millionen Preise erhoben wurden. Die Arti-

kel von Händler 1 haben einen Anteil von über 4 % an der Gesamtanzahl an Artikeln. Sie sind in 83 der 324 COICOP-10-Steller vertreten, wodurch ihre Preisvolatilität potenziell einen signifikanten Einfluss auf die Indexberechnung haben kann.

## 3.2 Zeitliches Auftreten der Preisänderungen

Die Analyse des zeitlichen Auftretens der Preisänderungen erfolgt in den zwei Dimensionen Wochentag und Stunde. Zunächst wird die Anzahl der Preisänderungen gruppiert nach Wochentagen betrachtet. Es wird deutlich, dass – im Gegensatz zu den Arbeitstagen Montag bis Freitag – die Anteile der Preisänderungen an Samstagen und Sonntagen niedriger sind. Am Sonntag ist der Anteil der Preisänderungen mit 0,24 % am niedrigsten. In der Tendenz scheinen die Onlinehändler die Preise am Wochenende eher konstant zu halten. Die Anteile bleiben über die Arbeitswoche hinweg relativ konstant, liegen aber

**Tabelle 3**

Onlinehändler mit einem Anteil an Preisänderungen von jeweils mehr als 3 % an den erhobenen Preisen

	Preise		Preisänderungen		COICOP-10-Steller	Artikel	
	Anzahl	Anteil in %	Anzahl	Anteil in %		Anzahl	Anteil in %
Onlinehändler 1	1 230 803	57 908	4,70	83	369	4,28	
Onlinehändler 2	103 259	4 153	4,02	3	22	0,26	
Onlinehändler 3	96 711	5 412	5,60	5	19	0,22	
Onlinehändler 4	103 348	4 385	4,24	1	16	0,19	
Onlinehändler 5	59 687	1 909	3,20	1	16	0,19	
Onlinehändler 6	68 861	3 712	5,39	4	15	0,17	
Onlinehändler 7	63 981	2 170	3,39	2	13	0,15	
Onlinehändler 8	19 755	1 023	5,18	1	6	0,07	
Onlinehändler 9	27 922	1 121	4,01	1	5	0,06	
Onlinehändler 10	333	25	7,51	1	1	0,01	

**Tabelle 4**  
Preisänderungen nach Wochentagen

	Preis-änderungen	Preise	Preis-änderungen	Preis-steigerungen	Preis-senkungen	Verhältnis von Preissteigerungen zu Preissenkungen
	Anzahl		Anteil in %	Anzahl		
Montag	18 826	5 657 627	0,33	14 125	4 701	3,00
Dienstag	20 039	5 907 274	0,34	14 603	5 436	2,69
Mittwoch	21 343	6 043 378	0,35	15 652	5 691	2,75
Donnerstag	19 977	6 059 859	0,33	14 487	5 490	2,64
Freitag	21 843	6 246 878	0,35	15 855	5 988	2,65
Samstag	16 513	6 043 319	0,27	12 323	4 190	2,94
Sonntag	13 495	5 710 293	0,24	10 184	3 311	3,08

insgesamt auf einem höheren Niveau als am Wochenende. [↘ Tabelle 4](#)

An allen Wochentagen gibt es deutlich mehr Preissteigerungen als Preissenkungen. Im Median kommen auf eine Preissenkung 2,75 Preissteigerungen. Besonders Montage und Sonntage überschreiten diesen Wert deutlich. Anteilig gibt es am Wochenende deutlich weniger Preisänderungen im Vergleich zur Arbeitswoche. Bei diesen liegt aber das Verhältnis der Preissteigerungen zu -senkungen immer deutlich über dem Median der Woche.

Für eine optimierte Nutzung des Web Scrapers sollten Preise an unterschiedlichen Wochentagen erhoben werden. Es zeigen sich deutliche Unterschiede im Preissetzungsverhalten zwischen Wochenendtagen und Tagen innerhalb der Arbeitswoche. Werden Preise wöchentlich erhoben, sollte dies berücksichtigt und Preise abwechselnd in diesen Zeiträumen erhoben werden.

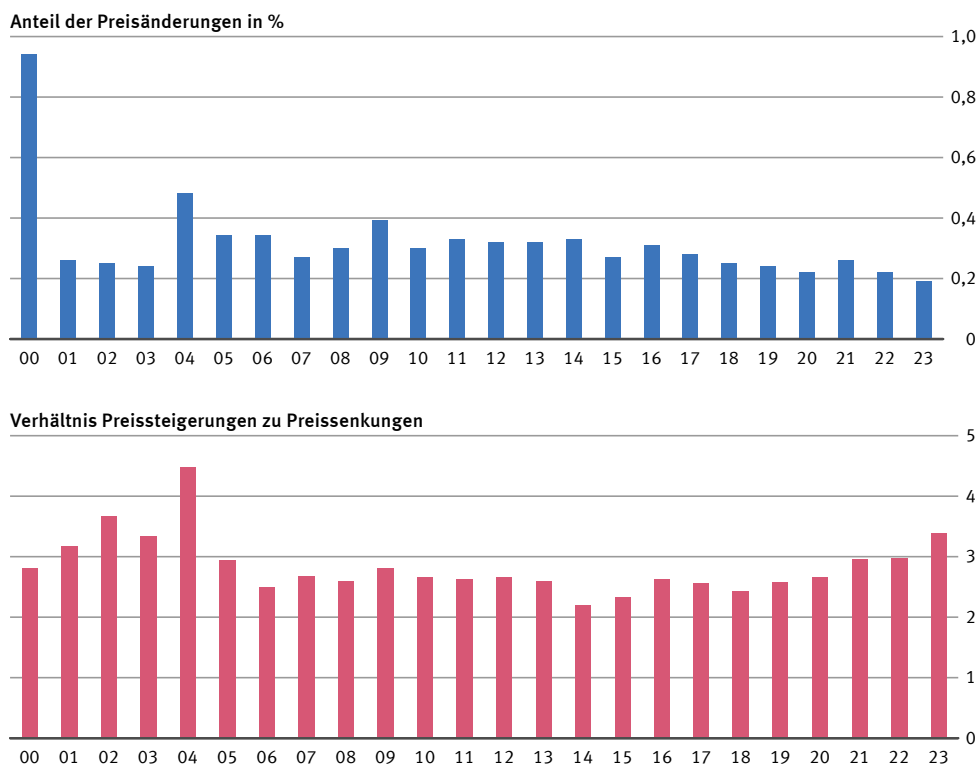
Abgesehen von den Wochentagen sollte die Erhebung auch zu unterschiedlichen Stunden erfolgen. Gerade zu Beginn

des Tages werden viele Preisanpassungen vorgenommen: 0,94 % in Stunde null, 0,48 % in Stunde vier. Ab Stunde 17 dagegen liegen die Anteile der Preisänderungen deutlich unter dem arithmetischen Mittel von 0,32%. Zwar sind keine Umsatzdaten für die täglichen Uhrzeiten verfügbar, es ist aber davon auszuge-

hen, dass die Konsumenten vermehrt abends (nach der Arbeit) online gehen, um Artikel zu kaufen. Auch die bereits erwähnte Verbraucherumfrage (Boniversum/bevh, 2017) untermauert dies: Die Befragten gaben an, dass sie 61 % ihrer online getätigten Einkäufe zwischen 18 und 24 Uhr durchgeführt haben. [↘ Grafik 2](#)

Neben den generellen Preisveränderungen ist auch die Anzahl der Preissteigerungen in den Morgenstunden

**Grafik 2**  
Preisänderungen nach Tagesstunden



2020 - 01 - 0213

hoch. Sie erreicht ihren höchsten Wert in der vierten Stunde mit durchschnittlich 4,47 Preissteigerungen je Preissenkung. Das arithmetische Mittel des Verhältnisses der Preisänderungen liegt über alle Tagesstunden hinweg bei 2,84. Von der vierzehnten bis zur zwanzigsten Stunden dagegen ist die Anzahl an Preissteigerungen vergleichsweise niedrig. Auch das lässt darauf schließen, dass Anbieter die Preise ihrer Onlineartikel tendenziell konstant halten, wenn mehr Konsumenten auf ihren Internetseiten einkaufen.

### 3.3 Handlungsempfehlungen zur automatisierten Preiserhebung

---

Insgesamt zeigt die Studie, dass die dynamische Preissetzung nach vorliegender Datenlage in einem überschaubaren Rahmen stattfindet und keine gravierenden Änderungen in der Datenerhebung für den Verbraucherpreisindex erfordert. Dynamische Preissetzung war bei zehn Onlinehändlern festzustellen; diese haben einen Artikelanteil von etwas mehr als 6 % an der Gesamtstichprobe.

Für die Datenerhebung ist vor allem der zeitliche Aspekt entscheidend. Speziell in den Morgenstunden, bis etwa 5 Uhr, werden seitens der Anbieter besonders viele Preise geändert. Ab 17 Uhr dagegen sind die Preisreihen sehr stabil. Nach Wochentagen betrachtet bleiben die Preise an den Wochenenden am stabilsten. Diese beiden Informationen kombiniert ergeben, dass vor allem Samstag- und Sonntagabende tendenziell geeignete Zeiträume für die Preiserhebung sind. Damit nicht nur die erhöhten Preise des Wochenendes erfasst werden, sollte das Scraping-Programm zumindest an einem zweiten Tag der Woche, zum Beispiel Mittwoch oder Donnerstag, zusätzlich Preise erheben. Dies sollte dann ebenfalls in den Abendstunden stattfinden, wenn die Preisreihen tendenziell unverändert bleiben.

Der Verbraucherumfrage (Boniversum/bevh, 2017) zufolge gaben 61 % der Befragten an, ihre Onlinekäufe bevorzugt zwischen 18 und 24 Uhr zu tätigen. Erfolgte die Preiserhebung also in diesem Zeitraum, würden vorrangig die Preise erfasst, die auch von den Verbraucherinnen und Verbrauchern bezahlt werden. In Bezug auf die Wochentage sind vor allem Samstag (30 %) und Freitag (16 %) beliebte Tage für Onlineeinkäufe, gefolgt von Sonntagen (15 %). Somit decken sich die identifi-

zierten Zeiten der Studie weitgehend mit den Erkenntnissen der Verbraucherumfrage.

Eine wöchentliche Erhebung könnte alle oben genannten Aspekte berücksichtigen und die Vorteile der automatisierten Datenerhebung nutzen, ohne unnötige Datenmengen zu erzeugen. Bei den zehn Onlinehändlern mit dynamischer Preissetzung könnte eine tägliche Erhebung sinnvoll sein. Die vorgestellten Ergebnisse sind Aggregationen über alle Onlinehändler hinweg. Es ist nicht auszuschließen, dass mehrere Onlinehändler leicht unterschiedliche Preissetzungsverhalten zeigen. Bei den täglichen Erhebungen könnten Scraping-Vorgänge deshalb zu unterschiedlichen Uhrzeiten angesetzt werden. Dies würde nicht nur die leichten Unterschiede der derzeit identifizierten Onlinehändler berücksichtigen, sondern wäre auch für eine sich verändernde Liste mit Anbietern geeignet. Wie die Daten gezeigt haben, sollte bei diesem Vorgehen zwischen null Uhr und fünf Uhr aufgrund der erhöhten Volatilität keine Preiserhebung erfolgen. Stattdessen sollten die täglichen Erhebungen ab fünf Uhr versetzt bis Mitternacht stattfinden.

## 4

---

### Fazit und Ausblick

---

Mit der voranschreitenden Digitalisierung und dem damit verbundenen Wachstum des Onlinemarkts steigt der Modernisierungsdruck auf die Preisstatistik. Die Einführung der automatisierten Preiserhebung erlaubt es, die Erhebungsfrequenz zu erhöhen sowie die Anzahl der erfassten Artikel und dadurch die Präzision des Verbraucherpreisindex und dessen Teilindizes zu verbessern. Zudem lassen sich durch ihren Einsatz Ressourcen einsparen und Fehlerquellen, die mit der manuellen Preiserhebung einhergehen, minimieren. Das Ziel der beschriebenen Studie war es, die optimale Nutzung des Web-Scraping-Programms zu gewährleisten und die Erhebungsfrequenz entsprechend zu bestimmen.


Es ergibt sich also, dass Preiserhebungen an bestimmten Tagen und zu gegebenen Uhrzeiten sinnvoll sind. Die Onlinehändler zeigen ein klares Preissetzungsverhalten, das nach Wochentagen und Tagesstunden zu unterscheiden ist. Preise sollten erhoben werden, wenn diese möglichst konstant gehalten werden und die Konsumenten vermehrt die Produkte nachfragen. Das Wochenende



und die Abende an jedem Tag der Woche sind zur Erhebung besonders geeignet. Neben diesen Wochentagen und Uhrzeiten sollten aber auch zu anderen Zeiträumen Erhebungen stattfinden, um verfälschte Preisänderungen zu vermeiden, wie am Beispiel der Treibstoffpreise an Tankstellen verdeutlicht. Onlinehändler haben zudem ein besonderes Preissetzungsverhalten vor und um Sonder- und Feiertage gezeigt, das allerdings bereits durch die traditionelle Preiserhebung erfasst wird.

Daraus folgt, dass für die Mehrheit der Preisreihen wöchentliche Erhebungen empfohlen werden. Bei den Onlinehändlern mit sehr volatilen Preisreihen könnten tägliche Erhebungen die Präzision des Verbraucherpreisindex weiter erhöhen. Allerdings sollte das Web-Scraping-Programm möglichst ressourceneffizient eingesetzt werden. Übermäßige Datenmengen könnten sonst zu einer Überforderung der Datenanalysten führen.

Im Zuge dieser Analysen wurde auch das Auftreten der dynamischen Preissetzung untersucht. Ein vorläufiges Ergebnis der Studie ist, dass die derzeitige Praxis der monatlichen Preiserhebung eine adäquate Darstellung der monatlichen Preisentwicklung durch den Verbraucherpreisindex zu gewährleisten scheint. Abschließende Untersuchungen zu diesem Thema stehen aber noch aus und werden in einem folgenden Aufsatz vorgestellt werden.

Der Onlinehandel ist im stetigen Wandel begriffen und so sollte sich auch die Methodenforschung für den Verbraucherpreisindex entsprechend weiterentwickeln und die automatisierte Datenerhebung stärken. Der Einsatz des Web-Scraping-Programms erlaubt es, die Entwicklung auf dem Onlinemarkt gezielter zu beobachten und Onlinehändler zu identifizieren, die ihre Preise mittels Algorithmen automatisiert anpassen. 

## LITERATURVERZEICHNIS

---

Blaudow, Christian/Burg, Florian. *Dynamische Preissetzung als Herausforderung für die Verbraucherpreisstatistik*. In: WISTA Wirtschaft und Statistik. Ausgabe 2/2018, Seite 11 ff.

Blaudow, Christian/Seeger, Daniel. *Fortschritte beim Einsatz von Web Scraping in der amtlichen Verbraucherpreisstatistik – ein Werkstattbericht*. In: WISTA Wirtschaft und Statistik. Ausgabe 4/2019, Seite 19 ff.

Creditreform Boniversum GmbH (Boniversum)/Bundesverband E-Commerce und Versandhandel Deutschland (bevh). *Verbraucherumfrage – Bevorzugte Tageszeiten und Wochentage für Online-Einkäufe*. 2017. [Zugriff am 13. Mai 2020]. Verfügbar unter: [www.boniversum.de](http://www.boniversum.de)

Flower, Tanya/Karachalias, Eleftherios. *Using alternative data sources in consumer price indices: May 2019*. 2019. [Zugriff am 13. Mai 2020]. Verfügbar unter: [www.ons.gov.uk](http://www.ons.gov.uk)

Little, Roderick J.A./Rubin, Donald B. *Statistical Analysis With Missing Data*. Wiley Series in Probability And Statistics. 2. Auflage. New Jersey 2002.

Statistisches Bundesamt. *Qualitätsbericht – Verbraucherpreisindex für Deutschland*. 2018. Verfügbar unter: [www.destatis.de](http://www.destatis.de)

Statistisches Bundesamt. *GENESIS-Online*. 2020. Verfügbar unter: <https://www-genesis.destatis.de>

van Loon, Ken/Roels, Dorien. *Integrating big data in the Belgian CPI*. Vortrag beim UNECE-ILO Meeting of the Group of Experts on Consumer Price Indices. 2018. [Zugriff am 13. Mai 2020]. Verfügbar unter: [www.unece.org](http://www.unece.org)

Wahnbaeck, Carolin. *Ein Drittel der Deutschen kauft Geschenke vor allem online*. 2019. [Zugriff am 2. Dezember 2019]. Verfügbar unter: [www.spiegel.de](http://www.spiegel.de)

**Herausgeber**

Statistisches Bundesamt (Destatis), Wiesbaden

---

**Schriftleitung**

Dr. Daniel Vorgrimler

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

---

**Ihr Kontakt zu uns**

[www.destatis.de/kontakt](http://www.destatis.de/kontakt)

---

**Erscheinungsfolge**

zweimonatlich, erschienen im Juni 2020

Das Archiv älterer Ausgaben finden Sie unter [www.destatis.de](http://www.destatis.de)

---

Artikelnummer: 1010200-20003-4, ISSN 1619-2907

---

© Statistisches Bundesamt (Destatis), 2020

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.