



Julia Koch

ist Volkswirtin mit Abschlüssen der Universitäten Würzburg und Gießen und als wissenschaftliche Mitarbeiterin im Referat „Methoden und Kommunikation in der Preisstatistik“ des Statistischen Bundesamtes tätig. Sie beschäftigt sich vorwiegend mit der Integration von Scannerdaten in die Verbraucherpreisstatistik.



Baran Erdemsiz

hat Volkswirtschaftslehre an den Universitäten Mannheim und Jena studiert und ist seit 2017 wissenschaftlicher Mitarbeiter im Referat „Methoden und Kommunikation in der Preisstatistik“ des Statistischen Bundesamtes. Er beschäftigt sich schwerpunktmäßig mit der Weiterentwicklung des Verbraucherpreisindex unter der Verwendung von Scannerdaten.

EINSATZ VON SCANNERDATEN WÄHREND DER COVID-19-PANDEMIE

Julia Koch, Baran Erdemsiz

📌 **Schlüsselwörter:** Scannerdaten – Kaufverhalten – Verbraucherpreisindex – Digitalisierung – Coronakrise

ZUSAMMENFASSUNG

Scannerdaten sind digital verfügbare Kassendaten des Einzelhandels und eine neue vielversprechende Datenquelle zur Qualitätssicherung und Prozessoptimierung für die Preisstatistik. Im Zuge von Vorratskäufen und leeren Supermarktregalen zu Beginn der Coronakrise hat das Statistische Bundesamt die vorliegenden Scannerdaten auf das Kaufverhalten der Bürgerinnen und Bürger hin ausgewertet. Darüber hinaus konnten die Scannerdaten bei COVID-19-bedingten Erhebungsausfällen für die traditionelle Preiserhebung des Verbraucherpreisindex genutzt werden. Der Aufsatz beschreibt den Einsatz von Scannerdaten während der COVID-19-Pandemie, die angewendeten Methoden und die daraus resultierenden Ergebnisse.

📌 **Keywords:** scanner data – purchasing behaviour – consumer price index – digitalisation – corona crisis

ABSTRACT

Scanner data are digitally available data of the retail sector. They are a new and promising data source for purposes of quality assurance and process optimisation of price statistics. In the cause of stockpiling purchases and empty supermarket shelves at the beginning of the corona crisis, the Federal Statistical Office analysed the available scanner data to determine the purchasing behaviour of citizens. In addition, scanner data were used to replace the prices which the traditional price survey for the consumer price index could not collect because of COVID-19. This article describes the use of scanner data during the COVID-19 pandemic as well as the methods applied and the results obtained.

1

Einleitung

Der digitale Wandel verändert die Welt, in der wir leben: unsere Gesellschaft, unsere Wirtschaft, unsere Art der Kommunikation sowie unseren Konsum. Die amtliche Statistik soll die Realität unter Beachtung der Grundsätze der Neutralität, Objektivität und der fachlichen Unabhängigkeit abbilden, folglich bleibt sie von solchen Veränderungen nicht unberührt. Ganz im Gegenteil, die Statistischen Ämter des Bundes und der Länder treiben zur Erfüllung ihres gesetzlichen Auftrags die technische und methodische Weiterentwicklung voran. So hat die amtliche Statistik diverse Projekte initiiert, um das Potenzial von neuen digitalen Daten zu prüfen, neue Methoden zu entwickeln und diese in die offiziellen Berechnungen zu integrieren (Hauf und andere, 2020). Eine neue digitale Datenquelle mit großem Potenzial für die amtliche Statistik sind Scannerdaten von Einzelhandelsunternehmen (Bieg, 2019).

Der vorliegende Aufsatz erläutert zunächst in den Kapiteln 2 und 3, was Scannerdaten im Detail sind und wofür sie mittelfristig in der Preisstatistik verwendet werden sollen. Kapitel 4 beschreibt ihren Einsatz während der COVID-19-Pandemie: die zusätzliche Nutzung zur Auswertung des Kaufverhaltens und ihren Gebrauch bei COVID-19-bedingten Erhebungsausfällen für die traditionelle Preiserhebung des Verbraucherpreisindex. Der Artikel schließt mit einem Fazit und einem Ausblick.

2

Was sind Scannerdaten?

Scannerdaten sind Transaktionsdaten über Umsatz, Absatz und Art der verkauften Artikel, die an den Kassen von Einzelhandelsgeschäften erfasst werden. Die digital verfügbaren Kassendaten des Einzelhandels sind eine neue zukunftsorientierte Datenquelle für die amtliche Statistik, unter anderem für die Preisstatistik. Sie haben das Potenzial, die Genauigkeit und Qualität der Messung der Verbraucherpreisentwicklung zu sichern. Dabei sind für die Preisstatistik nicht Daten über einzelne Transaktionen relevant, sondern zusammengefasste Verkaufsdaten zu einzelnen Artikeln in großen, umsatzstarken Einzelhandelsketten und -geschäften.

➤ **Tabelle 1** stellt die Struktur von Scannerdaten anhand von Angaben, welche das Statistische Bundesamt bereits von Einzelhandelsunternehmen erhalten hat, exemplarisch und anonymisiert dar. Die Datenlieferungen enthalten darüber hinaus weitere Informationen, wie die interne Warengruppenstruktur des Einzelhändlers oder den jeweiligen Mehrwertsteuersatz des Artikels. Mithilfe der Global Trade Item Number (GTIN) sowie gegebenenfalls weiteren unternehmensinternen Artikelnummern können die Artikel eindeutig identifiziert werden. Weitere Informationen betreffen unter anderem den Artikelnamen sowie die Filiale. Der Artikelname ist eine wichtige Information, um die Scannerdaten zur Klassifikation des Verbraucherpreisindex (VPI) zuzuordnen. Anhand einer Filialidentifikationsnummer (Filial-ID)¹ ist es möglich, sowohl die Einzelhandelskette als auch den jeweiligen Standort der konkreten Filiale zu identifizieren. Diese Information ist zum Ermitteln regionaler Ergebnisse

¹ Den Identifikationsnummern der Unternehmen für ihre einzelnen Filialen wird im Statistischen Bundesamt für die jeweilige Einzelhandelskette noch eine Nummer hinzugefügt.

Tabelle 1

Struktur von Scannerdaten (fiktive Angaben)

Global Trade Item Number (GTIN)	Artikel ¹	Filial-ID	Umsatz in EUR	Absatz in Stück	Gewicht (netto)	Maß	Kalenderwoche
400001122222	FRUCHTIGE HIMBEERMARMELADE GLAS 250 G	005_0286544	19,14	6	250	G	201932
400001122223	IRISCHE BUTTER SALZIG 250g	003_0057896	375,24	236	250,00	G	202001
400001122224	MARKEN PILSENER DOSE 0,33L MW	002_0009685	61,92	48	0,33	L	202005

¹ Der Begriff „Artikel“ bezeichnet ein unverwechselbares, spezifisches Produkt, beispielsweise bei Butter die 250-g-Packung gesalzene Butter einer bestimmten Marke.

notwendig. Mithilfe weiterer Informationen zu Umsatz, Absatz, Gewicht und Maßeinheit der Artikel kann ein Durchschnittspreis für einen bestimmten Zeitraum berechnet werden, der zur Ermittlung des Preisindex dient. Dieser Durchschnittspreis umfasst auch Rabatte, zum Beispiel aufgrund von Sonderangeboten. Die letzte Spalte enthält die jeweilige Woche, in der der Artikel zu angegebenem Umsatz und Absatz verkauft wurde.

3

Wofür sollen die Scannerdaten verwendet werden?

Aktuell werden monatlich für rund 650 Waren und Dienstleistungen Verbraucherpreisindizes ermittelt. Die Datenbasis hierfür stellen mehr als 300 000 Einzelpreise dar, die in Handels- und Dienstleistungsunternehmen manuell erhoben werden. Diese aktuell übliche Preiserhebung erfolgt einerseits dezentral durch etwa 600 Preiserheberinnen und Preiserheber in Geschäften und Dienstleistungsbetrieben in ganz Deutschland und andererseits für bestimmte Produkte zentral, hauptsächlich als Erhebung im Internet. Bei dieser traditionellen Preiserhebung notieren die Preiserheberinnen und Preiserheber die Preise häufig verkaufter Artikel in der Regel einmal je Monat. Um die Erhebungszeitpunkte in den verschiedenen Berichtsstellen zu variieren, erstreckt sich der gesamte Erhebungszeitraum jeden Monat über drei Wochen. Ein einmal für die Preisbeobachtung ausgewählter Artikel wird zur Sicherstellung der Vergleichbarkeit beibehalten und erst dann gegen einen anderen ausgetauscht, wenn er nicht mehr oder nur noch selten verkauft wird.

Im Zuge der Digitalisierung steht die Verbraucherpreisstatistik vor einer Reihe neuer Herausforderungen. Das Umfeld, für das die Verbraucherpreisstatistik konzipiert wurde, wandelt sich kontinuierlich. Die Handels- und Unternehmensstrukturen verändern sich. Die Bedeutung des Onlinehandels nimmt immer mehr zu, neue reine Internetgeschäfte entstehen und häufig bietet auch der stationäre Einzelhandel seine Produkte parallel online an. Zunehmend gibt es individualisierte Preise, also Preise für verschiedene Personengruppen, oder Internetgeschäfte nutzen mit dynamischen Preissetzungsstrategien automatisierte Algorithmen für die kurzfristige Preisgestaltung (Hansen, 2020).

Aber auch der stationäre Einzelhandel ändert sich. So werden zunehmend digitale Preisschilder verwendet, durch die sich die Unternehmen effizientere Prozesse in den Filialen und somit geringere Personalkosten erhoffen. Sie bieten jedoch auch die Möglichkeit, Preise minütlich zu ändern.

Zudem steigt die Anzahl von Produktinnovationen, was kürzere Produktlebenszyklen nach sich zieht. Stark von der Digitalisierung betroffene Güter sind insbesondere technische Güter wie Personenkraftwagen oder Smartphones.

Eine weitere Neuerung des Handels ist die Verlagerung der Leistungserbringung auf die Kundschaft. So wird zum Beispiel die frühere Dienstleistung eines Reisebüros immer häufiger von den Konsumentinnen und Konsumenten über eigenständiges Buchen der Reise im Internet übernommen.

Im Ergebnis steigt neben der Anzahl der Preisänderungen auch die Anzahl der verfügbaren Artikel, womit die traditionelle Preiserhebung an ihre Grenzen stößt. Die ergänzende Nutzung digitaler Datenquellen ermöglicht den statistischen Ämtern, die Zahl der monatlichen Preisbeobachtungen deutlich zu erhöhen und der gestiegenen Volatilität im Handel zu begegnen. So werden zum Beispiel für Güterbereiche mit besonders komplexer Preisgestaltung – wie Pauschalreisen (Henn und andere, 2020) oder Beiträge zu Kfz-Versicherungen – zusätzliche Datenquellen genutzt, unter anderem Datenbanken mit hoher Beobachtungszahl.

Zur Qualitätssicherung der amtlichen Verbraucherpreisstatistik gibt es weitere Maßnahmen. Die traditionelle Preiserhebung wird zunehmend durch neue Formen der Preiserhebung ergänzt oder teilweise ersetzt. Zu nennen sind hier insbesondere die automatisierte Preiserhebung im Internet (Web Scraping) und die Nutzung von Scannerdaten. Web Scraping wird bereits seit einigen Jahren in der amtlichen deutschen Verbraucherpreisstatistik genutzt (Blaudow/Seeger, 2019) und Scannerdaten wurden nun erstmals testweise als Ersatzerhebungsmethode in den VPI integriert (Mai/Kretzschmar, 2020).

Mit der Nutzung von Scannerdaten ist es möglich, Durchschnittspreise in den jeweiligen Geschäften für einen wesentlich längeren Zeitraum zu erfassen als derzeit. Die zunehmende Menge an Beobachtungen sichert die Genauigkeit der Erfassung von Preisschwankungen und

ermöglicht zusätzliche Auswertungen, zum Beispiel für die Preisentwicklung von Bio-Produkten. Durch den digitalen Transfer der Daten vom Einzelhandelsunternehmen direkt zum Statistischen Bundesamt reduziert sich der Erhebungsaufwand vor Ort. Grundsätzlich ermöglichen Scannerdaten einen Paradigmenwechsel in der Verbraucherpreisstatistik: von einer Statistik basierend auf einer repräsentativen Stichprobe hin zu einer umfassenden Erhebung.

Im Vergleich mit der bisherigen Vorgehensweise bei der Berechnung des Verbraucherpreisindex treten bei der Nutzung von Scannerdaten jedoch völlig neue Herausforderungen auf. Aktuell entwickelt das Statistische Bundesamt organisatorische, IT-technische und methodische Standards, um Scannerdaten in naher Zukunft in der laufenden Produktion der Preisstatistik verstärkt nutzen zu können.

4

Einsatzfelder von Scannerdaten während der COVID-19-Pandemie

4.1 Absatzzahlen im Lebensmittel-einzelhandel

Zu Beginn der Coronakrise kam es verstärkt zu Vorratskäufen und leeren Supermarktregalen. Darauf hat das Statistische Bundesamt die vorliegenden Scannerdaten – ursprünglich zur Integration in die Verbraucherpreisstatistik gedacht – auf das aktuelle Kaufverhalten der Bürgerinnen und Bürger hin untersucht. Hierfür wurde die Entwicklung von Absatzzahlen gezielt nach Produkten mit besonders hoher Verkaufsbedeutung im Zuge der Coronakrise beobachtet. Die zugrunde liegenden Daten basieren auf einer geringen Anzahl von Verkaufsstätten aus dem gesamten Bundesgebiet. Sie sind daher eingeschränkt repräsentativ für das Kaufverhalten in Deutschland. Auswertungen dieser Art haben experimentellen Charakter und sind Teil eines Projekts im Bereich „Experimentelle Daten“, mit dem das Statistische Bundesamt neue Datenquellen und Methoden erprobt.²

² Ausgewählte Projektergebnisse werden in der Rubrik EXDAT – Experimentelle Daten auf www.destatis.de vorgestellt.

Konkret wurden die Absatzzahlen von Produkten des täglichen Bedarfs wie Nahrungsmittel und Hygieneartikel aggregiert und hieraus Indizes und Veränderungsraten berechnet. Den Startpunkt der Analyse stellt die 32. Kalenderwoche Anfang August im Jahr 2019 dar, da die neue Datenquelle seit diesem Zeitpunkt vorhanden war. Zur Berechnung der Veränderungsraten dienen als Bezugszeitraum die sechs Monate vor Beginn der Coronakrise von August 2019 bis Januar 2020. Die Ergebnisse waren zum Zeitpunkt der Veröffentlichung hochaktuell und wurden von der Presse, der breiten Öffentlichkeit und der Politik sehr häufig aufgegriffen.

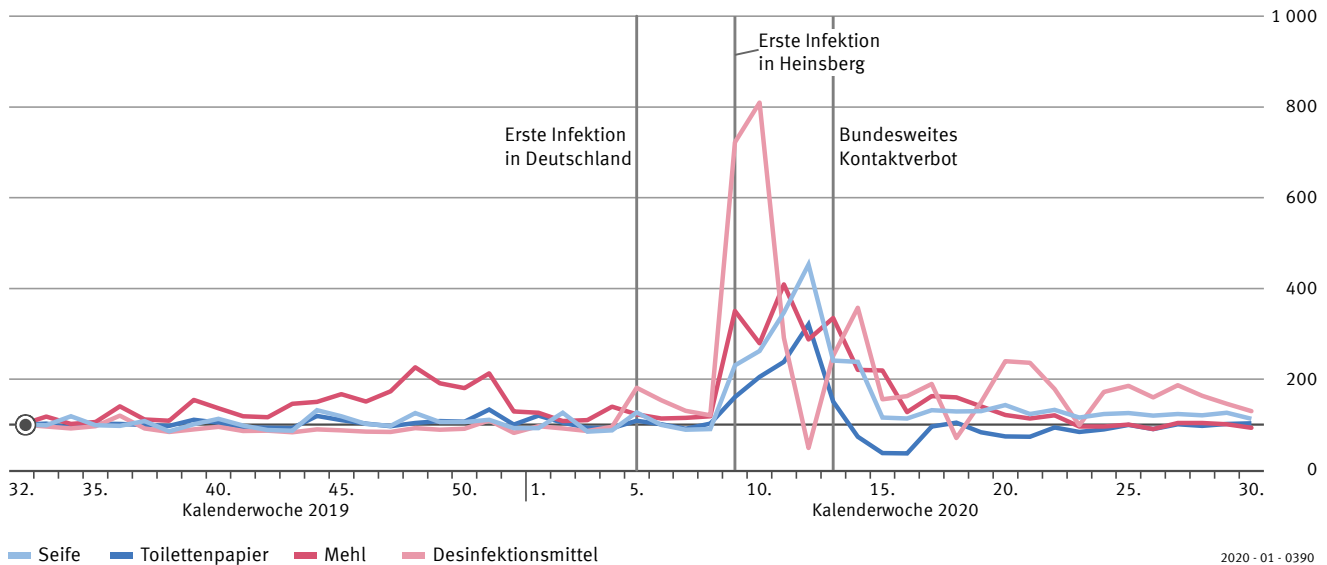
Zu Beginn der Coronakrise sah sich der Lebensmitteleinzelhandel einer ungebremsten Nachfrage nach bestimmten Gütern des täglichen Bedarfs gegenüber. Die Sonderauswertungen experimenteller Daten des Statistischen Bundesamtes zeigen, dass die Verkaufszahlen für ausgewählte Produkte ab Ende Februar bis Anfang April teils auf einem extrem hohen Niveau lagen. Ab Mitte April bis zum aktuellen Rand pendelten sich die Verkaufszahlen der meisten Produkte in etwa wieder auf dem Vorkrisenniveau ein. Eine dauerhafte Erhöhung der Absatzzahlen ist bei Seife und Desinfektionsmitteln festzustellen. [↘ Tabelle 2](#) am Ende des Abschnitts

Erstmals schnellten die Absatzzahlen der ausgewählten Lebensmittel und Hygieneartikel in der 9. Kalenderwoche vom 24. Februar bis zum 1. März in die Höhe. Bereits in dieser Woche waren die Absatzzahlen für Mehl (+ 150 %), Teigwaren (+ 109 %) und passierte Tomaten (+ 106 %) mehr als doppelt so hoch wie im Durchschnitt der sechs Monate zuvor. In der darauffolgenden Woche stiegen die Verkaufszahlen von Teigwaren um 150 %, danach gingen sie bis zur 14. Kalenderwoche (30. März bis 5. April) stark zurück. Grund dafür dürfte auch ein kurzzeitig fehlendes Angebot in diesem Warenssegment sein. Ab Kalenderwoche 16, also seit Mitte April, lagen die Absatzzahlen von Teigwaren und passierten Tomaten für einige Wochen niedriger verglichen mit dem Bezugszeitraum vor der Coronakrise. Dies kann zum einen darauf zurückzuführen sein, dass die Nachfrage infolge von Vorratskäufen gesättigt war. Zum anderen kann dies ein Indiz dafür sein, dass sich das Kaufverhalten der Verbraucherinnen und Verbraucher ab diesem Zeitpunkt normalisierte. [↘ Grafik 1](#)

Hygieneartikel wie Desinfektionsmittel, Toilettenpapier oder Seife verzeichneten eine noch stärker ausgeprägte

Grafik 1

Absatzindex von ausgewählten Verbrauchsgütern nach Kalenderwochen
32. Kalenderwoche 2019 = 100



Steigerung der Absatzzahlen. In der ersten Märzwoche (Kalenderwoche 10 vom 2. bis 8. März 2020) stieg der Absatz von Desinfektionsmitteln auf mehr als das Achtefache im Vergleich zum Durchschnitt des vorherigen halben Jahres (+ 751 %). In den beiden Wochen danach brach der Absatz förmlich ein und lag Mitte März in der Kalenderwoche 12 nur noch bei der Hälfte des sonst üblichen Absatzes. Dies ist darauf zurückzuführen, dass das Produkt vorübergehend praktisch ausverkauft war. Die Verkaufszahlen von Seife und Toilettenpapier erreichten ihren Peak in der 12. Kalenderwoche. Sie lagen mehr als vier Mal beziehungsweise drei Mal so hoch wie in den sechs Monaten zuvor (Seife: + 337 %, Toilettenpapier: + 211 %) und gingen bis Anfang April in der 15. Kalenderwoche wieder stark zurück. Seit diesem Zeitpunkt liegen die Absatzzahlen von Desinfektionsmittel und Seife bis zum aktuellen Rand weitgehend konstant über dem Niveau des Vergleichszeitraums. Vermutlich hängt dies mit einem höheren Hygienebewusstsein der Verbraucherinnen und Verbraucher, aber auch mit den Hygienevorschriften in der Gastronomie zusammen.

Einige Nutzerinnen und Nutzer aus dem Medienbereich äußerten in Anfragen die Vermutung, dass frisches Obst und Gemüse so gut wie gar nicht mehr gekauft und in vielen Filialen regelrecht zum Ladenhüter würde. Daher nahm das Statistische Bundesamt diese Produktgrup-

pen ebenfalls in die Auswertung mit auf. Insgesamt war die Nachfrage nach Obst und Gemüse in den ausgewählten Filialen entgegen der genannten Erwartungen im Zeitraum Ende Februar bis Anfang April 2020 jedoch überdurchschnittlich hoch. Gleichwohl waren große Unterschiede im Konsumverhalten zwischen frischem Obst und frischem Gemüse sowie den jeweiligen Konserven beziehungsweise den tiefgekühlten Produkten zu beobachten.

Abhängig von der Haltbarkeit der Produkte und dem Lageraufwand änderten sich auch die Verkaufszahlen. Gemüsekonserven verzeichneten die größten Veränderungsraten, gefolgt von tiefgekühltem Gemüse.³ Die höchsten Steigerungen der Absatzzahlen von Gemüsekonserven wurden Mitte März in der 11. Kalenderwoche mit + 137 % verzeichnet, das tiefgekühlte Gemüse erreichte eine Woche später seinen Höhepunkt mit + 85 %. Aber auch der Absatz von frischem Gemüse lag leicht über dem Vergleichszeitraum vor der Coronakrise.

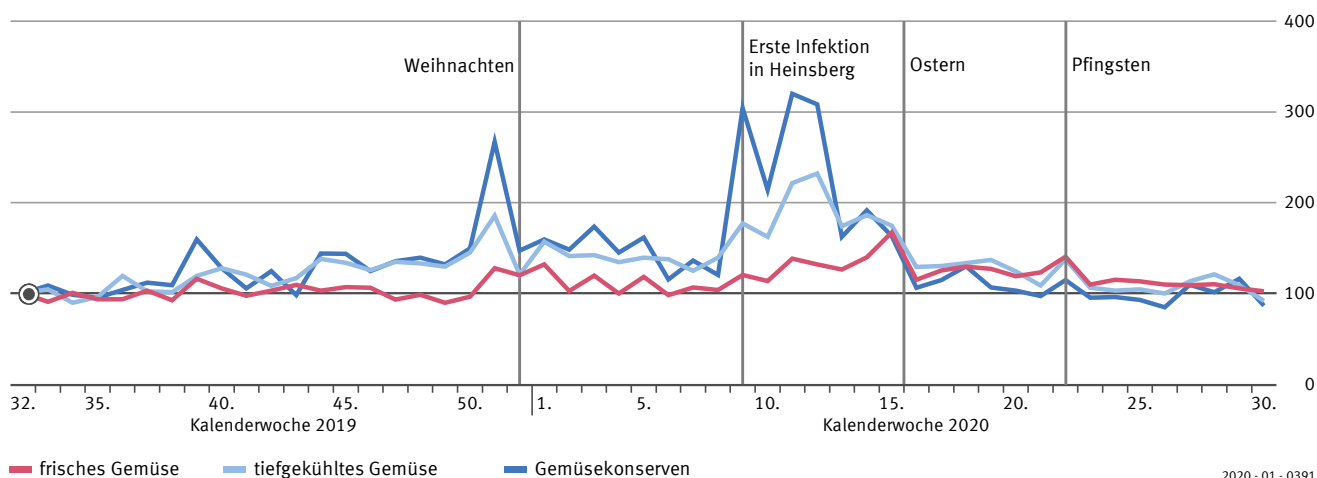
↳ Grafik 2

Zunehmend gab es Nachfragen von verschiedenen Seiten nach der Entwicklung des Konsums von alkoholischen Getränken während der Coronakrise. Daher wertete das

3 Konkret wird hier der Verlauf von Gemüse fokussiert. Der Verlauf von Obst zeigt ein ähnliches Bild (siehe Tabelle 2).

Grafik 2

Absatzindex von Gemüse nach Kalenderwochen
32. Kalenderwoche 2019 = 100



Statistische Bundesamt auch hierzu das Kaufverhalten genauer aus. Hintergrund für diese Nachfragen insbesondere von Seiten der Politik und Ministerien war die Vermutung, dass die Ausgangsbeschränkungen zu einem erhöhten Alkoholkonsum führen könnten. Diese Vermutung bestätigen die experimentellen Daten jedoch nicht.

Der Absatz von alkoholischen Getränken war in der Coronakrise bisher tendenziell unterdurchschnittlich im Vergleich zu den sechs Monaten von August 2019 bis Januar 2020. Dies liegt auch am stark überdurchschnitt-

lichen Absatz alkoholischer Getränke in der Weihnachtszeit, welcher den gesamten Durchschnitt im Vergleichszeitraum nach oben zieht. Eine erhöhte Nachfrage nach Alkoholika zeigte sich auch vor Feiertagen im Jahr 2020: Die Verkaufszahlen von Spirituosen lagen kurz vor Karneval (17. bis 23. März), kurz vor Ostern (6. bis 12. April) und in der Woche von Christi Himmelfahrt beziehungsweise „Vatertag“ (18. bis 24. Mai) leicht über den Verkaufszahlen der umliegenden Wochen. [↘ Grafik 3](#)

Grafik 3

Absatzindex von alkoholischen Getränken nach Kalenderwochen
32. Kalenderwoche 2019 = 100

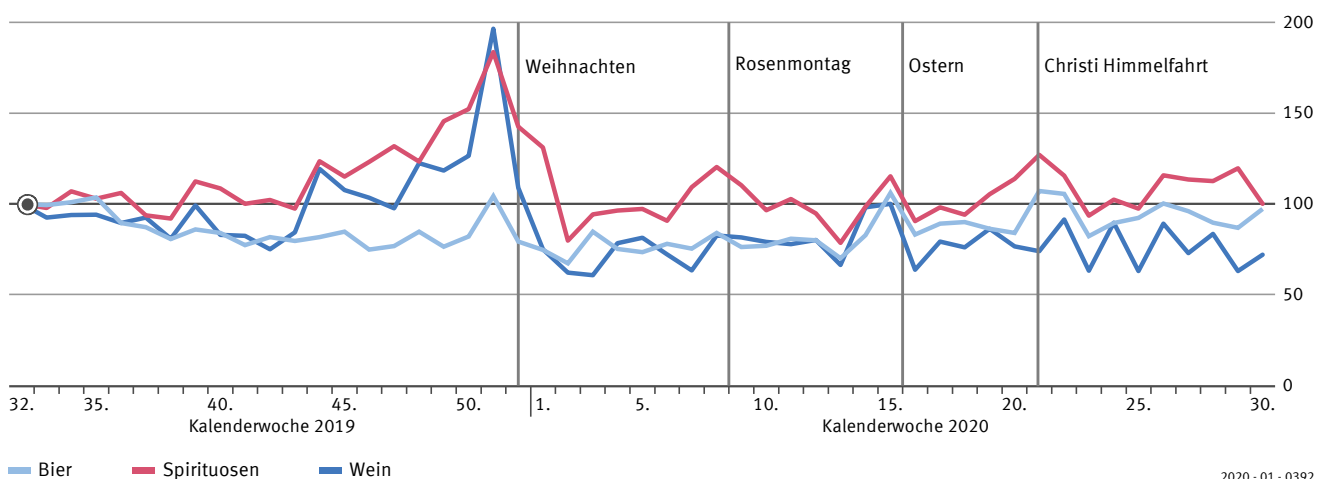


Tabelle 2

Absatz von ausgewählten Verbrauchsgütern nach Kalenderwochen 2020

	KW09	KW10	KW11	KW12	KW13	KW14	KW15	KW16	KW17	KW18	KW19
Veränderung gegenüber dem Zeitraum August 2019 bis Januar 2020 in %											
Seife	+ 122,4	+ 152,7	+ 235,0	+ 336,6	+ 132,7	+ 129,5	+ 11,6	+ 9,5	+ 27,0	+ 24,3	+ 25,1
Toilettenpapier	+ 55,9	+ 98,8	+ 130,7	+ 210,8	+ 45,9	- 28,9	- 64,2	- 64,7	- 7,5	+ 0,8	- 19,4
Teigwaren	+ 109	+ 149,8	+ 140,2	+ 116,8	+ 8,9	+ 2,0	+ 47,4	- 26,6	- 34,2	- 24,3	- 27,6
Mehl	+ 150,2	+ 99,3	+ 192,4	+ 105,3	+ 139,4	+ 57,4	+ 56,4	- 8,8	+ 16,0	+ 14,2	± 0,0
Zucker	+ 75,8	+ 39,1	+ 101,2	+ 100,1	+ 34,5	+ 9,7	+ 21,4	- 29,6	- 27,0	- 19,1	- 16,0
Reis	+ 205,6	+ 41,0	+ 163,2	+ 207,6	+ 26,5	+ 12,3	+ 39,1	- 27,4	- 25,0	+ 45,9	- 5,4
Desinfektionsmittel	+ 658,7	+ 750,5	+ 206,2	- 49,3	+ 164,7	+ 275,5	+ 63,4	+ 71,0	+ 98,8	- 26,6	+ 56,6
Äpfel	+ 45,9	+ 13,2	+ 64,7	+ 42,3	+ 26,0	+ 45,0	+ 28,9	- 2,2	+ 12,4	+ 11,3	+ 2,6
Passierte Tomaten	+ 106	+ 87,8	+ 205,5	+ 171,5	+ 47,4	+ 45,8	+ 11,4	- 18,4	- 8,0	+ 0,8	- 12,3
Hefe	+ 96,5	+ 50,7	+ 131,7	+ 51,5	- 36,5	+ 27,7	+ 38,7	+ 38,7	+ 29,6	+ 10,1	+ 55,1
Bier	- 9,5	- 8,7	- 4,1	- 5,1	- 16,8	- 1,7	+ 25,9	- 1,3	+ 5,8	+ 6,9	+ 2,6
Frisches Gemüse	+ 15,2	+ 8,8	+ 32,2	+ 26,3	+ 20,7	+ 33,8	+ 59,9	+ 10,2	+ 19,7	+ 24	+ 21,4
Tiefgekühltes Gemüse	+ 41,1	+ 29,2	+ 76,5	+ 84,8	+ 38,4	+ 48,3	+ 39,2	+ 2,8	+ 3,7	+ 6,2	+ 9,1
Gemüsekonserven	+ 125,1	+ 58,6	+ 137,0	+ 128,2	+ 20,1	+ 41,9	+ 20,7	- 21,3	- 15,0	- 3,8	- 21,2
Frisches Obst	+ 12,0	- 9,7	+ 24,0	+ 7,8	+ 8,5	+ 23,0	+ 27,9	+ 4,0	+ 2,1	+ 2,2	- 1,8
Tiefgekühltes Obst	+ 26,0	+ 32,1	+ 41,9	+ 47,6	+ 53,9	+ 52,6	+ 84,4	+ 52,5	+ 32,4	+ 27	+ 25,1
Obstkonserven	+ 137,8	+ 27,0	+ 126,1	+ 97,4	+ 35,8	+ 62,8	+ 42,1	+ 0,6	+ 39,4	- 2,9	- 3,9
Spirituosen	- 3,2	- 15,3	- 9,8	- 16,7	- 31,0	- 13,5	+ 1,2	- 20,6	- 13,8	- 17,3	- 7,5
Wein	- 16,0	- 18,6	- 19,8	- 17,5	- 31,7	+ 1,1	+ 3,0	- 34,4	- 18,4	- 21,8	- 11,1
Kondome	+ 6,3	- 3,6	+ 18,6	+ 56,0	- 0,5	+ 4,0	+ 6,1	- 7,9	- 17,0	- 7,2	+ 1,4
	KW20	KW21	KW22	KW23	KW24	KW25	KW26	KW27	KW28	KW29	KW30
Veränderung gegenüber dem Zeitraum August 2019 bis Januar 2020 in %											
Seife	+ 37,9	+ 18,8	+ 27,6	+ 11,3	+ 18,9	+ 20,7	+ 15,6	+ 18,6	+ 15,9	+ 21,6	+ 8,7
Toilettenpapier	- 28,1	- 28,9	- 9,4	- 18,6	- 13,2	- 3,5	- 12,6	- 2,4	- 5,7	- 1,8	- 0,6
Teigwaren	- 29,9	- 30,3	+ 1,0	- 23,9	+ 53,9	- 34,4	- 20,8	- 20,0	+ 40,2	- 32,8	- 23,2
Mehl	- 13,8	- 19,1	- 14,0	- 31,5	- 31,5	- 28,5	- 35,5	- 26,3	- 26,3	- 28,2	- 33,5
Zucker	- 22,2	- 16,8	- 6,4	- 18,3	- 18,6	- 16,4	- 23,3	+ 10,8	- 19,2	- 21,1	- 23,6
Reis	- 15,4	+ 37,5	- 16,0	- 15,0	- 1,3	+ 23,8	- 31,9	- 7,1	- 3,9	+ 19,4	- 28,0
Desinfektionsmittel	+ 152,1	+ 148,2	+ 87,3	+ 3,6	+ 80,5	+ 94,6	+ 67,8	+ 96,1	+ 72,0	+ 53,5	+ 36,4
Äpfel	- 6,1	- 10,7	+ 1,2	- 16,5	- 20,9	- 19,6	- 23,3	- 19,8	- 21,9	- 22,1	- 28,8
Passierte Tomaten	- 16,0	- 22,5	- 6,0	- 10,9	- 4,2	- 15,4	- 27,0	- 10,3	- 2,2	+ 37,8	- 21,3
Hefe	+ 52,3	+ 18,5	+ 29,4	+ 10,1	+ 41,5	+ 26,6	- 10,5	- 5,2	- 4,5	- 4,9	- 2,1
Bier	- 0,3	+ 27,2	+ 25,2	- 2,4	+ 6,3	+ 9,7	+ 19,0	+ 14	+ 6,5	+ 3,2	+ 15,5
Frisches Gemüse	+ 13,7	+ 17,7	+ 34,2	+ 5,1	+ 10,0	+ 8,3	+ 4,8	+ 4,4	+ 5,4	+ 1,0	- 2,0
Tiefgekühltes Gemüse	- 1,1	- 13,5	+ 9,6	- 15,5	- 18,1	- 17,0	- 20,7	- 9,8	- 3,6	- 13,2	- 27,0
Gemüsekonserven	- 23,6	- 28,2	- 15,1	- 29,6	- 29,0	- 31,4	- 37,4	- 19	- 25,1	- 14,2	- 36,1
Frisches Obst	- 0,5	- 7,0	- 2,8	- 10,3	- 1,4	- 4,6	- 7,1	- 7,8	- 0,1	- 4,8	+ 2,1
Tiefgekühltes Obst	+ 49,2	+ 13,8	+ 23,0	+ 3,6	+ 47,2	+ 3,4	+ 11,8	+ 11,9	+ 0,9	- 4,7	+ 39
Obstkonserven	- 18,0	- 12,3	- 5,9	- 26,0	- 27,5	- 5,8	- 28,3	- 25,2	- 26,8	- 29,3	- 0,5
Spirituosen	- 0,1	+ 11,6	+ 1,6	- 17,8	- 10,1	- 14,5	+ 1,7	- 0,4	- 1,1	+ 5,0	- 12,1
Wein	- 21,2	- 23,9	- 5,9	- 34,9	- 7,8	- 35,2	- 8,2	- 25,0	- 14	- 35,2	- 25,9
Kondome	- 13,4	- 13,9	- 3,2	- 12,9	- 13,0	- 2,4	- 10,9	+ 1,3	- 0,5	- 4,4	- 6,4

4.2 Nutzung für den Verbraucherpreisindex

Auch für die traditionelle Preiserhebung können Scannerdaten eine sinnvolle Ergänzung darstellen. Eine wichtige Säule der traditionellen dezentralen Preiserhebung besteht darin, dass Preiserheberinnen und Preiserheber Preise persönlich zum Beispiel in Geschäften des Einzelhandels oder in Dienstleistungsbetrieben der Gastronomie erheben. Im Zuge der Corona-Pandemie konnte diese Preiserhebung teilweise nur eingeschränkt oder gar nicht erfolgen. In einigen Geschäften war der Einlass reglementiert oder es kam zu kompletten Schließungen. Des Weiteren wurde teilweise von der Preiserhebung vor Ort abgesehen, um die Gesundheit der Preiserheberinnen und Preiserheber zu schützen. Für die Preisstatistik kam es dadurch zu Erhebungsausfällen.¹⁴

Grundsätzlich werden bei temporären Produkt- oder Berichtsstellenausfällen bestehende Fortschreibungsmechanismen angewendet, um die Berechnung des Preisindex auch in solchen Fällen zu gewährleisten. Im Zuge der Coronakrise entstand für die Preisstatistik jedoch eine vollkommen neue Situation: Zu dem vorübergehenden Ausfall einzelner Preise oder Geschäfte kam gleichzeitig der regionale Ausfall kompletter Produktbereiche hinzu.

Zusammen mit dem Statistischen Amt der Europäischen Union (Eurostat) und den übrigen nationalen Statistikämtern in der Europäischen Union hat das Statistische Bundesamt einheitliche Empfehlungen zum Umgang mit dieser neuen Situation entwickelt. Eine dieser Empfehlungen sieht vor, vorhandene Scannerdaten für coronabedingte Erhebungsausfälle der traditionellen Preiserhebung zu nutzen. Dieses Vorgehen bot sich insbesondere in Bereichen an, in denen die Geschäfte weiterhin geöffnet hatten, Scannerdaten also weiterhin generiert wurden, die Preiserhebung vor Ort aber dennoch eingeschränkt war, beispielsweise im Lebensmittel Einzelhandel.

Aufgrund der COVID-19-bedingten Krisensituation hat das Statistische Bundesamt zusammen mit denjenigen Statistischen Landesämtern, welche zeitweise stark von Erhebungsausfällen betroffen waren, damit begonnen,

den kurzfristigen Einsatz als Ersatzmethode im VPI zu testen.¹⁵

Scannerdaten haben als digitale Datenquelle großes Potenzial, stellen die Preisstatistik jedoch vor neue Herausforderungen. Eine Schwierigkeit bestand bei diesem Projekt in der derzeit noch eingeschränkten Scannerdaten-Datenbasis. Die gegenwärtig verfügbaren Scannerdaten decken nur einen kleinen Teil des Marktes ab und können auch nur einen Teil der traditionellen Erhebung ersetzen.

Eine weitere Herausforderung bestand in der Verknüpfung der traditionell, manuell erhobenen Daten mit den Scannerdaten. Die traditionellen Erhebungsdaten des VPI enthalten in der Regel für eine Güterart (COICOP-10-Steller)¹⁶ aus einem Geschäft genau einen Preis von einem Artikel. Die Scannerdaten enthalten jedoch umfangreiche Angaben über alle verkauften Artikel aus dem gleichen Geschäft. Ziel war es, eine artikelgenaue 1:1-Verknüpfung von Erhebungs- und Scannerdaten herzustellen, um so die fehlenden Preise aus der Vor-Ort-Erhebung durch Preise aus den Scannerdaten ersetzen zu können. Die konkreten einzelnen Erhebungsartikel aus der traditionellen Erhebung auch in den Scannerdaten wiederzufinden, ist allerdings mit erheblichem Aufwand verbunden: Unter anderem können Bezeichnungen abweichen und in der traditionellen Erhebung wurde die GTIN bisher nicht als Merkmal aufgenommen.

Wenn beispielsweise in einem Geschäft üblicherweise der Preis des Artikels „STREICHZARTE BUTTER RAHM 250g“ für die Güterart „Butter“ erhoben wird, ist es aufgrund der großen Menge an Scannerdaten nicht trivial, diesen Artikel entweder automatisiert oder manuell in den Scannerdaten zu finden. Die manuelle Zuordnung stellte sich als sehr zeitintensiv heraus. Um diese Suche zu erleichtern, wurden je Filiale Ranglisten von möglichst passenden und absatzstarken Artikeln je Güterart aus den Scannerdaten ermittelt. Diese Auswertung berücksichtigt, dass auch bei der traditionellen Erhebung die Preiserheberinnen und Preiserheber angehalten sind, bei der Preiserhebung die absatzstärksten Artikel je Güterart vor Ort auszuwählen. Es werden daher in den Scannerdaten je Berichtsstelle alle Verkaufsartikel einer

¹⁴ Zur Inflationsmessung während der Corona-Pandemie siehe auch Mai/Kretzschmar (2020).

¹⁵ Die traditionelle Erhebung vor Ort wird grundsätzlich von den Statistischen Ämtern der Länder durchgeführt.

¹⁶ Die Klassifikation der Preisstatistik ist die COICOP (Classification of Individual Consumption by Purpose – Klassifikation der Verwendungszwecke des Individualverbrauchs).

Tabelle 3

Scannerdatenauswertung am Beispiel Butter (fiktive Angaben)

Berichts- stelle	Global Trade Item Number (GTIN)	VPI-Klassifikation (COICOP)	Artikel ¹	Wochendurch- schnittspreis in EUR	Umsatz in EUR	Absatz in Stück	Gewicht (netto)	Maß	Kalender- woche	Rang (Absatz)
0005	400001122223	0115100100	IRISCHE BUTTER SALZIG 250g	1,59	375,24	236	250,00	G	202024	1
0005	400001122225	0115100100	STREICHZARTE BUTTER RAHM 250g	2,49	283,86	114	250,00	G	202024	2
0005	400001122226	0115100100	LEICHTE BUTTER 33% 250g	1,69	174,07	103	250,00	G	202024	3
0005	400001122227	0115100100	MARKENBUTTER PREMIUM 250g	3,89	194,50	50	250,00	G	202024	4
0005	400001122229	0115100100	JOGHURT BUTTER 69% 250g	2,49	59,76	24	200,00	G	202024	5
0005	400001132225	0115100100	BUTTER GENTECHNIK- FREI REFORM 250g	2,69	61,87	23	250,00	G	202024	6

1 Der Begriff „Artikel“ bezeichnet ein unverwechselbares, spezifisches Produkt, beispielsweise bei Butter die 250-g-Packung gesalzene Butter einer bestimmten Marke.

bestimmten Güterart überprüft und nur die absatzstärksten Artikel ausgegeben. Absatzangaben sind in den Scannerdaten auf Artikelebene enthalten und können daher zur Filterung genutzt werden. [↘ Tabelle 3](#) enthält einen exemplarischen Auswertungsauszug. Je Güterart wird für jeweils eine Woche eine Liste der absatzstärksten Artikel mit den errechneten durchschnittlichen Wochenpreisen je Artikel und Berichtsstelle abgebildet. Für den COICOP-10-Steller „Butter“ werden hier die aus den Scannerdaten errechneten Preise der sechs absatzstärksten Artikel der jeweiligen Berichtsstelle eines Bundeslands dargestellt.

Derzeit beträgt die Anzahl der ausgewerteten Güterarten bis zu 60 je Bundesland. Die Auswahl erfolgte zum Großteil aufgrund einer relativ engen Güterbeschreibung. Eine enge Güterbeschreibung bedeutet eine überschaubare, dem COICOP-10-Steller zugrunde liegende Artikelvielfalt und somit eine hohe Genauigkeit der Zuordnung von Artikeln zur VPI-Klassifikation.¹⁷

Anhand der Auswertungsrangliste prüft anschließend das Statistische Landesamt, ob genau der bisher durch die traditionelle Preiserhebung erhobene Artikel enthalten ist, um dessen ausgefallenen Erhebungspreis ersetzen zu können. Auf diese Weise wird beispielsweise der bisher in der traditionellen Erhebung in Berichtsstelle 0005 für den COICOP-10-Steller „Butter“ erho-

bene Artikel „STREICHZARTE BUTTER RAHM 250g“ identifiziert. Anschließend kann der in Tabelle 3 einsehbare Preis von 2,49 Euro ersatzweise als Erhebungspreis gewählt werden. Eine solche eindeutige Identifizierung eines Artikels ermöglicht es zudem, ohne Mehraufwand diesen Artikel in der nächsten Periode gezielt auszuwerten. Da dann GTIN und Artikelnummer vorliegen, ist der Artikel in den Scannerdaten einfach wiederzufinden.

Die Scannerdaten enthalten grundsätzlich Angaben über den Umsatz und den Absatz eines Artikels. Ein expliziter Preis, der direkt als Ersatzpreis genutzt werden könnte, ist nicht vorhanden. Das Statistische Bundesamt errechnet aus den verfügbaren Angaben einen Wochendurchschnittspreis, den sogenannten Unit-Value, aus dem Quotienten von Umsatz und Absatz (siehe Tabelle 3). Dieser Preis stellt einen gemittelten Transaktionspreis der aggregierten Umsätze und Absätze des Artikels einer Filiale in einer Woche dar.

Die Tests zur Verwendung von Scannerdaten für den VPI wurden erstmals im April 2020 zusammen mit neun Statistischen Landesämtern durchgeführt. Die Tests waren erfolgreich, für die Produktion des VPI genutzt wurden Scannerdaten von fünf Statistischen Landesämtern. Aufgrund der beschränkten Datenbasis ist der Umfang des Einsatzes aber noch begrenzt: Scannerdaten konnten bisher nur wenige hundert ausgefallene Erhebungspreise ersetzen.

Statistische Landesämter, welche die Auswertung erhalten, aber nicht für die Produktion verwendet haben, nutzten die Daten beispielsweise, um erste Erfahrungen

7 Die Zuordnung der einzelnen Artikel in den Scannerdaten zur VPI-Klassifikation erfolgt mithilfe von Machine Learning, zum Beispiel Zuordnung Artikel IRISCHE BUTTER SALZIG 250 G zum COICOP-10-Steller 0115100100 Butter.

mit Scannerdaten zu sammeln oder Vergleiche mit Erhebungspreisen durchzuführen. So zeigt eine erste Testauswertung, dass es einen hohen Anteil an Fällen gibt, in denen eine exakte Übereinstimmung von Erhebungspreis und Scannerdatenpreis vorliegt. In anderen Fällen waren Unterschiede zu beobachten. So verzeichnete beispielsweise ein Artikel (alkoholisches Getränk) einen Erhebungspreis von 8,99 Euro, aber einen Scannerdatenpreis in Höhe von 5,99 Euro in derselben Filiale. Dafür kann es verschiedene Gründe geben. Ein Grund kann sein, dass Scannerdaten- und Erhebungspreis aus zwei unterschiedlichen Wochen stammen und der Artikel in einer dieser beiden Wochen zumindest zeitweise im Angebot war. Derartige Differenzen können aber auch zustande kommen, wenn Erhebungs- und Scannerdatenpreis aus derselben Woche stammen und sich der Artikelpreis im Wochenverlauf geändert hat. Dann liegt der Grund darin, dass der Scannerdatenpreis einen gemittelten Transaktionspreis darstellt. Er berücksichtigt alle Transaktionen eines Artikels in einer Filiale über eine oder mehrere Wochen, so auch Transaktionen von reduzierten Angebotspreisen. Dies stellt einen großen Vorteil der Scannerdaten dar.


5

Fazit und Ausblick

Scannerdaten können für unterschiedliche statistische Zwecke genutzt werden. Sie bieten den Vorteil, hochaktuell und hochfrequent verfügbar zu sein. Die Sonderauswertungen von Scannerdaten im Zuge der COVID-19-Pandemie zur Untersuchung des Kaufverhaltens hatte eine sehr große Resonanz in der Öffentlichkeit und der Politik. Auch die Medien haben die Untersuchungsergebnisse breit aufgegriffen, was für einen hohen Nachrichtenwert spricht. Diese hohe Reichweite resultiert unter anderem aus der sehr zeitnahen Veröffentlichung zu einer aktuellen gesellschaftlichen Debatte der Vorratskäufe. Dies unterstreicht gleichzeitig die wichtige Rolle der amtlichen Statistik in Krisenzeiten.

Zudem war die Verwendung von Scannerdaten bei ausgefallenen Erhebungspreisen des Verbraucherpreisindex aufgrund der zeitnahen Verfügbarkeit der Scannerdaten hilfreich, um die traditionelle Preiserhebung zu unterstützen. Hierdurch entstanden auch Lerneffekte für die

künftige Verwendung von Scannerdaten. Beispielsweise wurden die Gründe für die Unterschiede zwischen den in der traditionellen Preiserhebung erhobenen Preisen und den anhand der Scannerdaten errechneten Preisen erörtert.

Insgesamt zeigen diese positiven Ergebnisse das große Potenzial von Scannerdaten. Sie bieten grundsätzlich weitere Verwendungsmöglichkeiten auch für andere Teilbereiche der amtlichen Statistik. Da die Auswertung von Scannerdaten noch auf einer geringen Anzahl von Filialen basiert, ist einer der wichtigsten nächsten Schritte, diese Datenbasis weiter auszubauen. Hier soll eine sichere und belastungsarme Datenlieferung im Dialog mit den Einzelhandelsunternehmen entwickelt werden. Weiterhin arbeitet das Statistische Bundesamt an organisatorischen, IT-technischen und methodischen Voraussetzungen, um Scannerdaten in naher Zukunft in der laufenden Produktion der Preisstatistik zu nutzen. 

LITERATURVERZEICHNIS

Bieg, Matthias. [*Nutzung von Scannerdaten in der Preisstatistik – eine Untersuchung anhand von Marktforschungsdaten*](#). In: WISTA Wirtschaft und Statistik.

Ausgabe 2/2019, Seite 25 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.

Hansen, Malte. [*Dynamische Preissetzung im Onlinehandel: zur langfristigen Anwendung von automatisierter Preiserhebung*](#). In: WISTA Wirtschaft und Statistik.

Ausgabe 3/2020, Seite 14 ff.

Hauf, Stefan/Stehrenberg, Shari/Zwick, Markus. [*EXDAT – experimentelle Daten und Methoden für eine innovative Statistik*](#). In: WISTA Wirtschaft und Statistik.

Ausgabe 4/2020, Seite 51 ff.

Henn, Karola/Islam, Chris-Gabriel/Schwind, Patrick/Wieland, Elisabeth. *Measuring price dynamics of package holidays with transaction data*. In: EURONA – Eurostat

Review on National Accounts and Macroeconomic Indicators. Ausgabe 2/2019, Seite 95 ff. Luxemburg 2020.

Mai, Christoph-Martin/Kretschmar, Marco. [*Inflationsmessung in Zeiten der Corona-Pandemie*](#). In: WISTA Wirtschaft und Statistik. Ausgabe 4/2020, Seite 107 ff.

Herausgeber

Statistisches Bundesamt (Destatis), Wiesbaden

Schriftleitung

Dr. Daniel Vorgrimler

Redaktionsleitung: Juliane Gude

Redaktion: Ellen Römer

Ihr Kontakt zu uns

www.destatis.de/kontakt

Erscheinungsfolge

zweimonatlich, erschienen im August 2020

Das Archiv älterer Ausgaben finden Sie unter www.destatis.de

Artikelnummer: 1010200-20004-4, ISSN 1619-2907

© Statistisches Bundesamt (Destatis), 2020

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.