
NUTZUNG VON SCANNERDATEN IN DER PREISSTATISTIK – EINE UNTERSUCHUNG ANHAND VON MARKTFORSCHUNGSDATEN

Matthias Bieg

↳ **Schlüsselwörter:** Verbraucherpreise – Scannerdaten – multilaterale Preisindizes – Digitalisierung – maschinelles Lernen

ZUSAMMENFASSUNG

Digital verfügbare Kassendaten des Einzelhandels, sogenannte Scannerdaten, sind eine vielversprechende Datenquelle für die Preisstatistik. Im Vergleich mit der traditionellen Preismessung treten bei der Nutzung von Scannerdaten jedoch teilweise völlig neue Herausforderungen auf. Um zu evaluieren, in welcher Form Scannerdaten in der deutschen Verbraucherpreisstatistik genutzt werden können, kaufte das Statistische Bundesamt einen Datensatz mit Kassendaten ausgewählter Lebensmittel und Getränke von einem Marktforschungsinstitut. Dieser Datensatz bildete die Basis für eine Untersuchung im Rahmen eines Projekts, bei der drei verschiedene Ansätze zur Indexberechnung mittels Scannerdaten getestet wurden. Die ermittelten Preisindizes wurden anschließend mit den Ergebnissen des Verbraucherpreisindex verglichen.

↳ **Keywords:** *consumer prices – scanner data – multilateral price indices – digitalisation – machine learning*

ABSTRACT

Digital point-of-sale (POS) data of the retail sector, also referred to as scanner data, are a promising data source for price statistics. However, compared to the traditional price collection methods, some completely new challenges arise when scanner data are used. In order to evaluate how scanner data can be employed in German consumer price statistics, the Federal Statistical Office bought a set of POS data on selected foods and beverages from a market research institute. As part of a project, this dataset served as basis for an analysis which focused on testing three different approaches to calculate price indices based on scanner data. Subsequently, the resulting price indices were contrasted with the consumer price index results.



Matthias Bieg

ist M. Sc. in International Economics and Public policy und als Referent im Referat „Methoden und Kommunikation in der Preisstatistik“ des Statistischen Bundesamtes tätig. Schwerpunktmäßig beschäftigt er sich mit der Implementierung von Scannerdaten in der Preisstatistik.

1

Einleitung

In der Verbraucherpreisstatistik wird die manuelle Preiserhebung vor Ort in den Geschäften beziehungsweise Dienstleistungsunternehmen bei einer Vielzahl von Statistikämtern weltweit mehr und mehr durch neue Formen der Preiserhebung ergänzt oder teilweise ersetzt. Zu nennen sind hier insbesondere die automatisierte Preiserhebung im Internet (Web Scraping) oder die Nutzung von Scannerdaten. Im Gegensatz zum Web Scraping werden Scannerdaten bislang noch nicht in der deutschen Verbraucherpreisstatistik genutzt.¹ Scannerdaten sind digitale Transaktionsdaten über Umsatz, Menge, Preis und Art der verkauften Artikel² auf Ebene der Global Trade Item Number (GTIN)³, die an den Kassen von Einzelhandelsgeschäften erfasst werden. Dabei sind für die Preisstatistik nicht Daten über einzelne Transaktionen relevant, sondern zusammengefasste Verkaufsdaten zu einzelnen Artikeln in großen, umsatzstarken Einzelhandelsketten beziehungsweise -geschäften. Insgesamt haben Scannerdaten das Potenzial, mit langfristig eher geringem Aufwand die Genauigkeit und Qualität der Messung der Verbraucherpreise für ausgewählte Produktgruppen zu verbessern. Die zunehmende Menge an Beobachtungen steigert die Genauigkeit und ermöglicht zusätzliche Auswertungen, zum Beispiel die Preisentwicklung von Bio-Produkten. Mit der Nutzung von Scannerdaten können Preise und Mengen zudem in den jeweiligen Geschäften für einen wesentlich längeren Zeitraum erfasst werden als derzeit. Bislang werden Preise in einem spezifischen Geschäft in der Regel nur an einem Tag im Monat erhoben.

Untersuchungen zur Nutzung von Scannerdaten in der deutschen Verbraucherpreisstatistik führte das Statistische Bundesamt erstmals im Jahr 2015 durch. Dabei wurden Scannerdaten zu Butter und Vollwaschmittel für den Zeitraum März bis Mai 2012 ausgewertet. Datenbasis dafür waren Testdatensätze, die verschiedene Marktforschungsinstitute bereitstellten. Der Test lieferte

erste Erkenntnisse über die Beschaffenheit von Scannerdaten. Aufgrund des kurzen Beobachtungszeitraums und der geringen Anzahl an Produkten war der Erkenntnisgewinn jedoch begrenzt. Für weitere Analysen war es notwendig, sowohl die Produktpalette als auch den Beobachtungszeitraum auszudehnen und dabei möglichst Scannerdaten direkt von den Einzelhandelsunternehmen zu erhalten.

Aus Gründen der schnelleren Verfügbarkeit entschied sich das Statistische Bundesamt, für eine zweite Testphase zunächst noch einen Scannerdatensatz von einem Marktforschungsinstitut einzukaufen. Ziel war es, parallel zu dieser zweiten Testphase, Einzelhandelsunternehmen für Datenlieferungen von Scannerdaten zu gewinnen, um in einer dritten Phase das zuvor aufgebaute Wissen auf „echte“ und unbearbeitete Scannerdatensätze anzuwenden. Der eingekaufte Scannerdatensatz beschränkte sich auf ausgewählte Produktgruppen aus den Bereichen Nahrungsmittel und Getränke. Bei diesen Produktgruppen fällt es im Vergleich zu Produktgruppen mit sehr stark wechselndem Sortiment wie Bekleidung oder Elektrogeräte leichter, valide Preisindizes zu erstellen, da Produktersetzungen oder Qualitätsbereinigungen seltener erforderlich sind.

Die zweite Testphase wurde im Rahmen eines vom Statistischen Amt der Europäischen Union (Eurostat) geförderten Projekts von Februar 2017 bis September 2018 durchgeführt und ist Gegenstand dieses Textes. Das Projekt zielte im Wesentlichen darauf ab, die Prozessschritte aufzubauen, die ab dem Dateneingang notwendig sind, um Preisindizes auf Basis von Scannerdaten zu berechnen. Dazu zählen neben der Datenaufbereitung und der eigentlichen Indexberechnung auch die Zuordnung der einzelnen Artikel zur Klassifikation des Verbraucherpreisindex (VPI)⁴ sowie die Suche nach Ersetzungsprodukten bei Produktausfällen. Insgesamt wurden auf Basis des Scannerdatensatzes drei verschiedene Preisindextypen berechnet, die mit den Ergebnissen des Verbraucherpreisindex verglichen wurden. Dabei wurden scannerdatenbasierte Indexreihen für Deutschland insgesamt ermittelt, die den entsprechenden Teilindizes des Verbraucherpreisindex gegenübergestellt wurden.

1 Zur Nutzung von Web Scraping im Verbraucherpreisindex für Deutschland siehe Blaudow (2018).

2 Mit dem Begriff „Artikel“ ist ein unverwechselbares, spezifisches Produkt gemeint, beispielsweise bei Butter die 200-g-Packung gesalzene Butter einer bestimmten Marke.

3 Bei der GTIN handelt es sich um eine Identifikationsnummer, mit der Artikel weltweit eindeutig zugeordnet werden können.

4 Die VPI-Klassifikation leitet sich aus der Systematik der Einnahmen und Ausgaben der privaten Haushalte (SEA) ab, welche wiederum auf der Klassifikation der Verwendungszwecke des Individualverbrauchs der Vereinten Nationen (Classification of Individual Consumption by Purpose – COICOP) basiert.

Kapitel 2 beschreibt zunächst den verwendeten Datensatz. In Kapitel 3 wird untersucht, wie einzelne Artikel der VPI-Klassifikation zugeordnet werden können. Kapitel 4 widmet sich der Analyse der Produktstabilität, das heißt der Frage, wie häufig Artikel vom Markt verschwinden. Kapitel 5 erläutert dann die Formeln und Parameter, die zur Berechnung der scannerdatenbasierten Preisindizes angewandt wurden, gefolgt von einer Analyse der ermittelten Ergebnisse in Kapitel 6. Kapitel 7 gibt einen kurzen Ausblick auf die nächsten geplanten Schritte hin zur Nutzung von Scannerdaten in der deutschen Verbraucherpreisstatistik.

2

Der Datensatz

Der Datensatz für die vorliegende Untersuchung wurde beim Marktforschungsinstitut Nielsen Services Deutschland gekauft und enthält wöchentliche Scannerdaten über den Zeitraum Januar 2015 bis Dezember 2016. Das im Datensatz abgebildete Sortiment umfasst etwa 95 000 Artikel in 44 Produktgruppen (COICOP-10-Steller) der Bereiche Lebensmittel und Getränke.⁵ [↗ Übersicht 1](#) Für diese Produktgruppen decken die Daten im genannten Zeitraum den Großteil der Umsatz- und Absatzzahlen (verkaufte Stückzahlen) des Lebensmitteleinzelhandels in Deutschland ab. Der Quotient aus Umsatz und Absatz bildet den sogenannten Unit-Value, der sich als absatzgewichteter Durchschnittspreis eines Artikels interpretieren lässt.

Neben Umsatz und Absatz enthält der Datensatz weitere Informationen, beispielsweise zu Hersteller, Marke, Produktmerkmalen und GTIN. [↗ Übersicht 2](#) Bereitgestellt wurden die Daten auf unterschiedlichen Aggregations-ebenen. So wurden die Daten beispielsweise nach verschiedenen Geschäftstypen wie „Supermarkt“ (groß/klein), „Verbrauchermarkt“ (groß/klein), „kooperierende Discounter“, „nicht kooperierende Discounter“⁶, „Drogeriemärkte“ sowie „Getränkeabholmärkte“ (nur für Getränke) unterschieden. Zudem standen auch Scan-

⁵ Produktgruppe und COICOP werden im Folgenden synonym verwendet. Beide Begriffe bezeichnen in diesem Text den COICOP-10-Steller, die tiefste Gliederungsebene in der VPI-Klassifikation.

⁶ Die Zahlen für die nicht mit Nielsen kooperierenden Discounter wurden von Nielsen aufbauend auf Haushaltsbefragungen geschätzt.

Übersicht 1

Die 44 Produktgruppen im Datensatz

COICOP-Nummer	COICOP-Name
0114110100	Vollmilch
0114210100	Teilentrahmte Milch
0114300100	Kondensierte Milch
0114400200	Joghurt
0114507100	Quark
0114507200	Friskkäse
0114601100	Sahne
0114604100	Fertigdessert
0114606100	Buttermilch oder anderes Milchgetränk
0115100100	Butter
0115201100	Margarine
0115209100	Pflanzenfett zum Braten und Backen
0115300100	Olivenöl
0115400100	Sonnenblumenöl, Rapsöl oder Ähnliches
0115500100	Schmalz oder anderes tierisches Speisefett
0118100100	Zucker
0118201100	Marmelade, Konfitüre oder Gelee
0118203100	Honig
0118205100	Brotaufstrich auf Nougatbasis
0118301100	Schokoladentafel
0118309100	Riegel oder andere Erzeugnisse aus Schokolade
0118401100	Pralinen
0118405100	Bonbons
0118409100	Kaugummi, Gummibärchen oder Ähnliches
0118500100	Speiseeis
0118601100	Süßstoff
0121110200	Kaffeepads oder Kaffeekapseln
0121110300	Bohnenkaffee
0121121100	Instantkaffee oder Ähnliches
0121201100	Schwarzer Tee oder grüner Tee
0121203100	Früchtetee oder Kräutertee
0121300100	Kakaopulver oder Ähnliches
0122100100	Mineralwasser mit Kohlensäure
0122100200	Mineralwasser ohne Kohlensäure
0122211100	Colagetränk
0122219100	Koffeinfreie Limonade
0122311100	Apfelsaft oder ähnlicher Fruchtsaft
0122312200	Orangensaft
0122315100	Multivitaminsaft
0122320300	Gemüsesaft
0213100100	Pils, Lager, Schwarzbier oder anderes untergäriges Bier
0213200100	Weizenbier, Altbier oder Ähnliches
0213300100	Alkoholfreies Bier, Malzbier oder Ähnliches
0213400100	Biermixgetränk

Übersicht 2

Auszug aus dem Nielsen-Datensatz

GTIN ¹	Produktname	Hersteller	Marke	Absatz in Stück	Umsatz in EUR	Inhalt
SENSITIVE	EIGENMARKE WENIG CO2 NORMAL COFF FR 1 1.5 L EW PETFL	HANDEL	EIGENMARKE	15 357	3 993	1,5 L
NA	BECK S PILS OA AH 0.33 L RW LONGNECK 2 X 24	BECK'S	BECK S	98	2 055	0,33 L
4001686363263	HARIBO BRIXX FRUCHT 200 G BTL	HARIBO	HARIBO	7 104	5 126	200 G
4011800262013	SCHWARTAU FRUTTISSIMA HIMBEER 250 G BE	SCHWARTAU	SCHWARTAU	1 452	2 982	250 G

1 Global Trade Item Number.

nerdaten für verschiedene Handelsorganisationen separat zur Verfügung.

Im Zuge der Datenaufbereitung wurde eine erste Plausibilitätskontrolle in Form einer Boxplot-Ausreißerbereinigung vorgenommen.¹⁷ Darüber hinaus wurden nicht alle im Datensatz vorhandenen Wochen berücksichtigt, sondern je Monat jeweils nur die ersten drei vollständigen Wochen herausgefiltert. Diese Vorgehensweise ließe in der Praxis zeitlichen Spielraum, um das vorläufige Ergebnis des Harmonisierten Verbraucherpreisindex (flash estimate) rechtzeitig, das heißt bis zum Ende des Berichtsmonats, an Eurostat zu übermitteln. Die Variable „GTIN“ im Datensatz war als eindeutige Identifikationsnummer auf Artikelebene nicht geeignet, da in einigen Fällen keine GTIN vorlag („NA“) oder diese wurde von Nielsen als sensibel („SENSITIVE“) eingestuft und daher nicht bereitgestellt (siehe Übersicht 2). Die eindeutige Identifizierung der Artikel erfolgte stattdessen über hausinterne Produkt- und Marktkennungen von Nielsen, die ebenfalls im Datensatz vorhanden waren.

3

Klassifizierung der Artikel

Eine der größten Herausforderungen bei der Nutzung von Scannerdaten zur Preisindexberechnung ist die Zuordnung der einzelnen Artikel zur VPI-Klassifikation. Aufgrund der Vielzahl an verfügbaren Artikeln wäre eine manuelle Klassifizierung mit sehr hohem Zeitaufwand

7 Je COICOP und Einzelhandelsunternehmen wurden Beobachtungen gelöscht, wenn ihr Unit-Value um mehr als den 3,5-fachen Interquartilsabstand (IQR) über dem Unit-Value des dritten Quartils oder um mehr als den 1,5-fachen IQR unter dem Unit-Value des ersten Quartils lag.

und damit einhergehend mit entsprechend hohen Kosten verbunden. Daher werden alternative Möglichkeiten zur Klassifizierung von Artikeln getestet. Insbesondere die Verwendung von maschinellen Lernverfahren stellt einen vielversprechenden Lösungsansatz dar. So wurde von Eurostat Anfang 2018 ein Softwareprototyp zur Verfügung gestellt, der einzelne Artikel mittels ausgewählter maschineller Lernverfahren, nämlich (i) logistische Regression, (ii) Random Forests und (iii) Support Vector Machines, klassifiziert.¹⁸ Dabei dienen der Produktname, alle Wörter im Produktnamen und Buchstaben-trigramme des Produktnamens für jedes dieser Verfahren als Inputvariable.

Wie bereits erwähnt, deckte der Datensatz wöchentliche Um- und Absätze für zugehörige Artikel der in Übersicht 1 gelisteten Produktgruppen ab. Eine Klassifizierung der Artikel war nicht erforderlich, weil die entsprechende COICOP-Nummer im Datensatz als Variable gekennzeichnet war. Aufgrund der bereits erfolgten Zuordnung eignete sich der Datensatz hervorragend, um für künftige Verwendungen die Genauigkeit der im Softwareprototyp implementierten maschinellen Lernmethoden zu testen. Dazu wurde ein Trainingsdatensatz erstellt, der als zufällige Stichprobe gezogen wurde und 10 % des gesamten Datensatzes ausmachte.

In einem ersten Test wurden mit den drei genannten Verfahren jeweils nur jene Artikel klassifiziert, die mit einer geschätzten Eintrittswahrscheinlichkeit (predicted probability rate) von mehr als 80 % richtig zugeordnet werden konnten. Dabei fielen die Ergebnisse höchst unterschiedlich aus. Während mit dieser Einstellung bei der logistischen Regression nahezu alle Artikel (97 %)

8 Nähere Informationen zu maschinellen Lernverfahren finden sich beispielsweise bei Hastie und andere (2009).

klassifiziert wurden, waren es bei Random Forests 82 % und bei Support Vector Machines sogar weniger als jeder zweite Artikel (41 %). Betrachtet man bei diesem Test die Fehlerrate, das heißt den Prozentsatz der Abweichungen von der durch Nielsen vorgenommenen Zuordnung, dann zeigt sich, dass diese bei der logistischen Regression zwar am höchsten und dabei mehr als doppelt so hoch lag wie bei Support Vector Machines, jedoch immer noch deutlich unter 1 %. Da die logistische Regression bei diesem Test zudem mehr Artikel zuordnete als die beiden anderen Ansätze, war dieser Test nicht aussagekräftig, um zu beurteilen, welches Verfahren am besten klassifiziert. Aus diesem Grund wurde in einem zweiten Test die erforderliche Grenze für die geschätzte Eintrittswahrscheinlichkeit auf 0 % gesetzt, sodass alle Artikel im Datensatz zugeordnet wurden.⁹ Hier wies die logistische Regression mit 1,4 % die niedrigste, Support Vector Machines mit 4,6 % dagegen die höchste Fehlerrate auf.

↳ **Tabelle 1**

Tabelle 1
Maschinelle Lernverfahren zur Klassifizierung von Artikeln im Test

	Geschätzte Eintrittswahrscheinlichkeit	Klassifizierte Artikel	Abweichungen von der Nielsen-Zuordnung
	%		
Logistische Regression	> 80	97,2	0,63
Random Forests	> 80	82,4	0,30
Support Vector Machines	> 80	40,7	0,24
Logistische Regression	> 0	100	1,44
Random Forests	> 0	100	2,11
Support Vector Machines	> 0	100	4,58

Aus beiden Tests zusammen lässt sich schlussfolgern, dass die logistische Regression bei einem Trainingsdatensatz mit einer Abdeckungsquote von 10 % die beste Performance bei der Zuordnung von Artikeln zur COICOP liefert. Allerdings wurde bislang bei allen drei Methoden nur mit den Standardeinstellungen gearbeitet. Inwieweit Optimierungen verschiedener Parameter diese Rangfolge verändern, müssen künftige Untersuchungen zeigen. Darüber hinaus lässt sich nicht aus-

⁹ Zugeordnet wurde ein Artikel immer zur COICOP mit der höchsten geschätzten Eintrittswahrscheinlichkeit.

schließen, dass bei höheren Abdeckungsquoten des Trainingsdatensatzes die Ansätze Random Forests oder Support Vector Machines besser zur Produktklassifizierung geeignet sind als die logistische Regression.

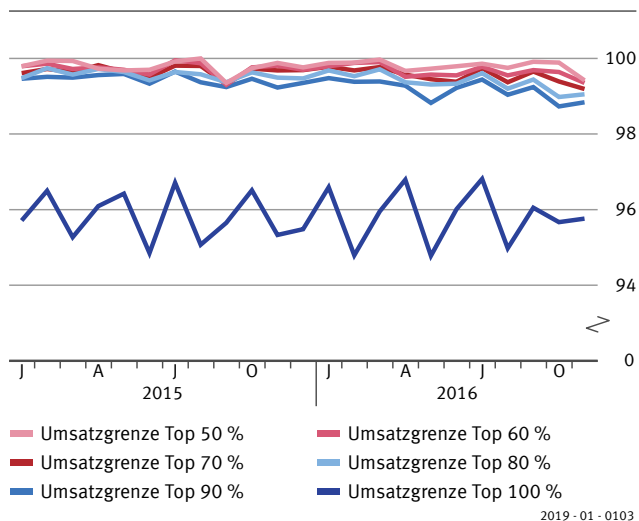
4

Analyse der Produktstabilität

Kenntnisse über die Produktstabilität sind bei der Nutzung von Scannerdaten zur Erstellung von Preisindizes wichtig. Hohe Produktstabilität bei Scannerdaten erleichtert die Berechnung von Preisindizes und reduziert das Risiko von fehlerhaften Berechnungen. Je nach Berechnungsart können Preisindizes bei Produktgruppen, deren Artikel insgesamt betrachtet eher kurze Lebenszyklen aufweisen, mehr oder weniger stark von nicht realen Abwärtstendenzen („Downward-Drifts“) betroffen sein. Dieser Downward-Drift kommt dadurch zustande, dass gerade bei innovativen Produkten mit schnellem Lebenszyklus zunächst ein hoher Markteinführungspreis angesetzt wird, der im weiteren Zeitverlauf kontinuierlich sinkt. Wenn dann Nachfolgeprodukte indexneutral in die Berechnung eingehen, ist die Gefahr groß, dass bei der Preismessung fälschlicherweise eine durchgehend fallende Entwicklung nachgewiesen wird.

Bei der Untersuchung der Produktstabilität des Datensatzes zu Nahrungsmitteln und Getränken wurde zunächst analysiert, ob merkliche Unterschiede für verschiedene Stichprobengrößen zu beobachten waren. In jeder der Stichproben waren jeweils die umsatzstärksten Artikel enthalten, wobei je COICOP Abschneidegrenzen von 50 %, 60 %, ..., 100 % angewendet wurden. Als Maß für die Produktstabilität wurde je Monat der Anteil an Artikeln, die auch im Folgemonat vorhanden waren, herangezogen. Nimmt man beispielsweise an, dass in Monat 1 vier verschiedene Artikel (A, B, C, D) und in Monat 2 fünf verschiedene Artikel (B, C, D, E, F) verkauft wurden, dann würde in diesem Fall für Monat 1 eine Produktstabilitätsrate von 75 % ausgewiesen. Diese Kennzahl zielt auf die Notwendigkeit von Ersetzungen ab (Artikel A) und lässt Artikel, die neu beziehungsweise erneut auf den Markt kommen (Artikel E und F) außen vor. ↳ **Grafik 1** zeigt jeweils je Abschneidegrenze die durchschnittliche monatliche Produktstabilitätsrate der

Grafik 1
Durchschnittliche monatliche Produktstabilität gruppiert nach Umsatzabschneidegrenzen in %



44 Produktgruppen gemessen anhand des arithmetischen Mittels. Für fast alle ausgewählten Stichproben lagen die durchschnittlichen Produktstabilitätsraten im gesamten Zeitverlauf über 98%. Lediglich bei der vollen Stichprobe fielen die Produktstabilitätsraten merklich ab, waren aber mit Werten von durchgehend über 94% immer noch auf einem hohen Niveau.

Die vorangehende Analyse war insbesondere für den später in Kapitel 5.2 beschriebenen „traditionellen“ Ansatz der scannerdatenbasierten Indexberechnung von Interesse, da dieser auf einer abgeschnittenen Stichprobe (cut-off sample) aufbaut, die im Zeitverlauf konstant gehalten wurde. Dazu musste im Vorfeld eine Abschneidegrenze für die Stichprobe ausgewählt werden. Aufbauend auf den Ergebnissen in Grafik 1 ließ sich schlussfolgern, dass sich alle Cut-off-Samples hinsichtlich der Produktstabilität nur geringfügig unterschieden und sich diesbezüglich gleichermaßen für den „traditionellen“ Ansatz eigneten.

Neben der Produktstabilität in Abhängigkeit von der Stichprobengröße wurde auch die Produktstabilität innerhalb der 44 Produktgruppen untersucht. Dabei wurde das gleiche Maß zur Messung der Produktstabilität angewandt wie zuvor und diesmal ausschließlich die volle Stichprobe betrachtet. In diesem Fall wurde jedoch nicht das arithmetische Mittel je Monat über alle 44 COICOP berechnet, sondern je COICOP über

alle 23 Monate (von Januar 2015 bis November 2016). Grafik 2 bestätigt die insgesamt hohe Produktstabilität der im Datensatz vorhandenen COICOP aus den Bereichen Nahrungsmittel und Getränke. Die Produktstabilitätsraten je COICOP lagen allesamt zwischen 93 und 98%, wobei bei Olivenöl und Zucker die niedrigsten, dagegen bei Schokoladentafel und Pralinen die höchsten Werte zu beobachten waren.

Insgesamt können die Produktgruppen im Datensatz als stabil angesehen werden. Die Gefahr eines Downward-Drift konnte aufgrund der gegebenen Produktstabilität als äußerst gering eingestuft werden. Empirische Untersuchungen (beispielsweise Chessa und andere, 2017) deuten darauf hin, dass ein Downward-Drift erst bei deutlich niedrigeren Produktstabilitätsraten auftritt.

5

Die Indexberechnung

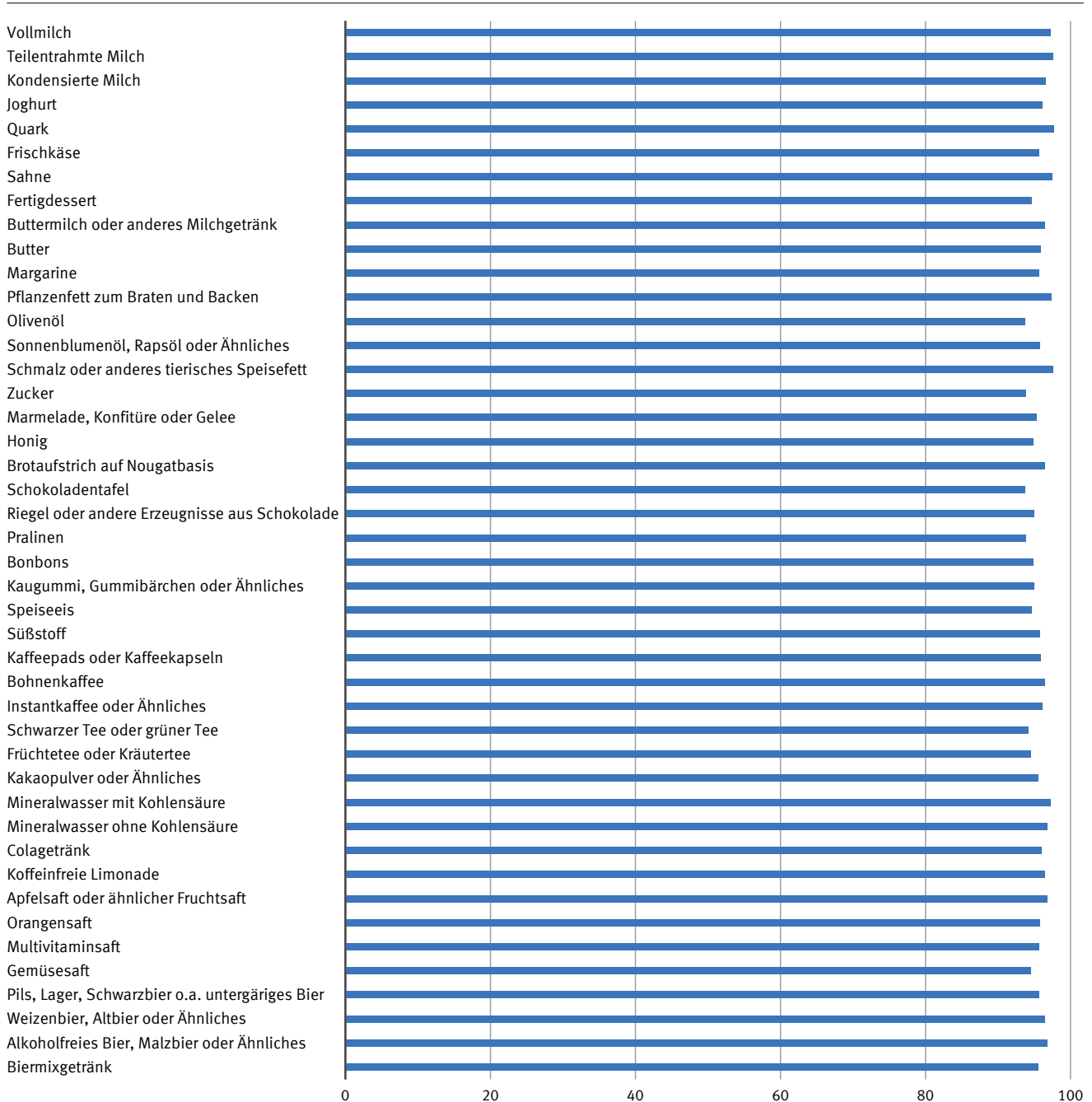
5.1 Vorbemerkungen

Im Rahmen der Untersuchung wurden mit dem „traditionellen“ Ansatz (SDTA, Abschnitt 5.2), dem QAUV-Index (Abschnitt 5.3) und dem GEKS-Index (Abschnitt 5.4) beispielhaft drei verschiedene scannerdatenbasierte Methoden zur Berechnung von Preisindizes getestet. Dabei wurde aus den wöchentlichen Umsatz- und Absatzzahlen ein Unit-Value gebildet, der als Preisrepräsentant diente. Der SDTA-Index ist konzeptionell an den Verbraucherpreisindex angelehnt, in dem Sinne, dass die Laspeyres-Indexformel verwendet wurde. Im Gegensatz zum Verbraucherpreisindex wurden die Elementarindizes jedoch nicht auf Basis (ungewichteter) Dutot-Indizes berechnet. Da im Scannerdatensatz die Umsätze je Artikel vorlagen, konnte stattdessen direkt mit dem Laspeyres-Index gerechnet werden. Ähnlich dem Verbraucherpreisindex wurde bei der Berechnung eine Stichprobe verwendet. Um den Stichprobenumfang zu erhalten, mussten Ersetzungsprodukte gesucht werden, wann immer ein Artikel für mehr als einen Monat nicht mehr im Datensatz vorhanden war.

Nutzung von Scannerdaten in der Preisstatistik – eine Untersuchung anhand von Marktforschungsdaten

Grafik 2

Produktstabilitätsraten nach Produktgruppen¹
 Durchschnitt über 23 Monate², in %



1 10-Steller der COICOP (Klassifikation der Verwendungszwecke des Individualverbrauchs).
 2 Januar 2015 bis November 2016.

Im Unterschied zum SDTA-Index wurde sowohl beim QAUV- als auch beim GEKS-Index die volle Stichprobe genutzt, die sich monatlich entsprechend der aktuellen Marktentwicklungen verändert. QAUV- und GEKS-Index

zählen beide zu den multilateralen Indizes, das heißt sie beziehen im Gegensatz zum bilateralen SDTA-Index mehr als nur die Basisperiode und die Beobachtungsperiode mit in die Berechnung ein. Auch Ersetzungen

Übersicht 3

Gewählte Parameter der verschiedenen Berechnungsmethoden

	SDTA	QAUV	GEKS
Umsatzabdeckung der Stichprobe	70 %	100 %	100 %
Ausreißerbereinigungs-Filter	<ul style="list-style-type: none"> • $p_t < 0.05$ oder $p_t > 2000$ • $p_t < 0.1 * p_{t-1}$ oder $p_t = > 10 * p_{t-1}$ 		
Dumping-Filter	<ul style="list-style-type: none"> • $p_t < 0.65 * p_{t-1}$ und $q_t < 0.65 * q_{t-1}$ 	nein	<ul style="list-style-type: none"> • $p_t < 0.65 * p_{t-1}$ und $q_t < 0.65 * q_{t-1}$
Preisbasis	Dezember 2015	Dezember 2015	–
Indexbasis	Dezember 2015	Dezember 2015	Dezember 2015
Ersetzungen	ja	nein	nein
Zeitfenster	–	13 Monate	13 Monate
Monatliche Verkettung	–	fixed based moving window	fixed based moving window

SDTA: ScannerDatenTraditioneller-Ansatz-Index; QAUV: Quality-Adjusted-Unit-Value-Index; GEKS: Gini-Eltető-Köves-Szulc-Index.

spielen keine Rolle, da die Stichprobe dynamisch und nicht statisch gehalten wird.¹⁰

Gleich waren bei dieser Untersuchung für alle drei Berechnungsarten zum einen die Wahl des Dezember 2015 als Indexbasis und zum anderen die Filter zur Ausreißerbereinigung. Preisrepräsentanten unter 5 Cent blieben ebenso unberücksichtigt wie Preisrepräsentanten über 2000 Euro. Darüber hinaus wurden auch Beobachtungen aus der Stichprobe entfernt, die im Vergleich zum Vormonat Preisveränderungen auf das Zehnfache oder mehr beziehungsweise auf ein Zehntel oder weniger aufwiesen. [↪ Übersicht 3](#)

Zusätzlich zur Ausreißerbereinigung wurde mit Ausnahme des QAUV-Index auch ein Dumping-Filter in den Berechnungsprozess einbezogen (siehe Übersicht 3).¹¹ Mit dem Dumping-Filter sollen Artikel aus der Berechnung ausgeschlossen werden, für die ein Abverkauf stattfindet, das heißt deren Absatz und Preis im Vergleich zum Vormonat stark zurückgeht. Solche Artikel sind nicht länger repräsentativ für das Konsumverhalten der privaten Haushalte in Deutschland und verursachen tendenziell einen (unangemessenen) Downward-Drift. Basierend auf den Erfahrungen anderer nationaler Statistikämter wurde der Dumping-Filter so gewählt, dass alle Beobachtungen entfernt wurden, bei denen gleichzeitig Preis und Absatz gegenüber dem Vormonat um mindestens 35 % zurückgingen.

10 Ausgenommen ist die Erkennung von Produkt-Relaunches, die im Rahmen dieser Untersuchung jedoch nicht gesondert berücksichtigt wurde.

11 Dumping-Filter werden laut Chessa (2016) für den QAUV-Index nicht benötigt und wurden daher auch bei dieser Untersuchung außen vor gelassen.

5.2 Der SDTA-Index

Bei der Ermittlung von scannerdatenbasierten Preisindizes nach „traditionellem“ Ansatz (SDTA – ScannerDaten-Traditioneller Ansatz) wurde je COICOP folgende Formel vom Typ Laspeyres angewandt:

$$I_t^L = \sum_{r=1}^R \sum_{i=1}^{N^r} \frac{p_t^{r,i}}{p_0^{r,i}} \cdot \frac{\sum_{z=1}^{12} p_{z,2015}^{r,i} \cdot q_{z,2015}^{r,i}}{\sum_{s=1}^R \sum_{j=1}^{N^s} \sum_{z=1}^{12} p_{z,2015}^{s,j} \cdot q_{z,2015}^{s,j}} = \sum_{r=1}^R \sum_{i=1}^{N^r} \frac{p_t^{r,i}}{p_0^{r,i}} \cdot w^{r,i}$$

Dabei ist:

- I_t^L der Laspeyres-Index für Monat t
- $p_t^{r,i}$ der Preis von Einzelhandelsunternehmen r für Artikel i in Monat t
- $p_{z,2015}^{r,i}$ der Preis von Einzelhandelsunternehmen r für Artikel i in Monat z im Jahr 2015
- $q_{z,2015}^{r,i}$ der Absatz von Einzelhandelsunternehmen r mit Artikel i in Monat z im Jahr 2015
- $w^{r,i}$ der Wägungsanteil für Artikel i von Einzelhandelsunternehmen r gemessen am Jahresumsatz 2015 der COICOP
- N^r die Anzahl der in der Stichprobe enthaltenen Artikel von Einzelhandelsunternehmen r
- R die Anzahl der in der Stichprobe enthaltenen Einzelhandelsunternehmen

Im Gegensatz zur „klassischen“ Laspeyres-Formel unterscheiden sich in der obenstehenden Formel Preisbasis und Wägungsbasis. Preisbasis war der Dezember 2015,

während als Wägungsbasis die Umsatzzahlen für das gesamte Jahr 2015 herangezogen wurden.

Die Stichprobe wurde je Einzelhandelsunternehmen und COICOP bei 70% des Gesamtumsatzes je Schicht abgeschnitten. Dabei wurden ausschließlich die umsatzstärksten Artikel des Jahres 2015 berücksichtigt. Im Falle von Artikelausfällen wurde der Preis fortgeschrieben, sofern er im Vormonat noch vorhanden war. War bereits im Vormonat kein Preis mehr vorhanden, so wurde nach einem Ersetzungsprodukt gesucht. Dabei kam ein automatisiertes Verfahren zum Einsatz, das aus einem Pool an potenziellen Ersetzungsprodukten den ähnlichsten Artikel herausfilterte. Auswahlkriterien waren zum einen die Ähnlichkeit des Artikelnamens und zum anderen der Übereinstimmungsgrad bei den produktbeschreibenden Merkmalen.

5.3 Der QAUV-Index

Der Quality-Adjusted-Unit-Value(QAUV)-Index beruht auf einem Konzept nach Geary (1958) und Khamis (1972), das für interregionale Preisvergleiche angewandt wird. Im Gegensatz zum Geary-Khamis(GK)-Index vergleicht der QAUV-Index jedoch Preisniveaus nicht zwischen verschiedenen Ländern, sondern zwischen verschiedenen Zeitperioden, in diesem Fall Monaten. Als multilaterale Methode berechnet der QAUV-Index gleichzeitig Preisindizes für alle berücksichtigten Monate. Für Monate am aktuellen Rand muss jeweils eine neue Preisindexreihe berechnet werden und der jeweils neue Monat dann mit der bestehenden Preisindexreihe verkettet werden. Die Berechnung der QAUV-Indizes je COICOP basierte auf folgender Formel nach Chessa (2016):

$$I_t^{QAUV} = \frac{\sum_{r=1}^{R_t} \sum_{i=1}^{N_t^r} p_t^{r,i} \cdot q_t^{r,i}}{\sum_{r=1}^{R_t} \sum_{i=1}^{N_t^r} v^{r,i} \cdot q_t^{r,i}} \bigg/ \frac{\sum_{r=1}^{R_0} \sum_{i=1}^{N_0^r} p_0^{r,i} \cdot q_0^{r,i}}{\sum_{r=1}^{R_0} \sum_{i=1}^{N_0^r} v^{r,i} \cdot q_0^{r,i}}$$

mit

$$v^{r,i} = \sum_{z \in T} \varphi_z^{r,i} \frac{p_z^{r,i}}{I_z^{QAUV}}$$

und

$$\varphi_z^{r,i} = \frac{q_z^{r,i}}{\sum_{u \in T} q_u^{r,i}}$$

Dabei ist:

I_t^{QAUV} der QAUV-Index für Monat t

$p_t^{r,i}$ der Preis von Einzelhandelsunternehmen r für Artikel i in Monat t

$q_t^{r,i}$ der Absatz von Einzelhandelsunternehmen r mit Artikel i in Monat t

$v^{r,i}$ der Qualitätsanpassungsfaktor für Artikel i von Einzelhandelsunternehmen r

$\varphi_z^{r,i}$ der Wägungsanteil für Artikel i von Einzelhandelsunternehmen r in Monat z gemessen am Gesamtumsatz des Artikels im Berechnungszeitraum

N_t^r die Anzahl der in der Stichprobe enthaltenen Artikel von Einzelhandelsunternehmen r in Monat t

R_t die Anzahl der in der Stichprobe enthaltenen Einzelhandelsunternehmen in Monat t

T die Menge an Monaten, die bei der Berechnung der Preisindexreihe berücksichtigt werden

Da sich der Preisindex I_t^{QAUV} und die Qualitätsanpassungsfaktoren $v^{r,i}$ gegenseitig bedingen, musste für die Berechnung der Ergebnisse ein Preisindexwert I_t^{QAUV} ermittelt werden, aus dem Qualitätsanpassungsfaktoren $v^{r,i}$ resultieren, die wiederum den Preisindex I_t^{QAUV} nicht mehr oder zumindest nicht mehr wesentlich veränderten. Wie bei Chessa (2016) wurde dazu ein iteratives Verfahren verwendet. Sobald der Preisindex I_t^{QAUV} nach Aktualisieren der Qualitätsanpassungsfaktoren $v^{r,i}$ um weniger als 0,001 Indexpunkte vom letzten ermittelten Wert abwich, stoppte die Berechnung.

Das Zeitfenster für die Berechnung einer Indexreihe lag bei 13 Monaten. Dadurch wurden am aktuellen Rand auch immer die Preise des Vorjahresmonats mitberücksichtigt. Dies war für die Monate im Zeitraum Februar 2016 bis Dezember 2016 relevant, deren Indizes jeweils neu ermittelt wurden und anhand der Fixed-Based-Moving-Window-Methode nach Lamboray (2017) mit der bestehenden Indexreihe verkettet wurden. Bei Anwendung dieses Ansatzes wurde die Spanne des Zeitfensters durchgehend konstant gehalten und die Verkettung erfolgte analog zum Harmonisierten Verbraucherpreisindex immer über den Dezember des Vorjahres.¹²

12 Eine ausführliche Übersicht über die verschiedenen Verkettungsmethoden findet sich bei van Loon und andere (2018).

5.4 Der GEKS-Index

Genauso wie der QAUV-Index hat auch der GEKS-Index, benannt nach den Arbeiten von Gini (1931), Éltető und Köves (1964) sowie Szulc (1964), seinen Ursprung auf dem Gebiet der internationalen Kaufkraftvergleiche. Analog zum QAUV-Index werden beim GEKS-Index jedoch Preisniveaus zwischen verschiedenen Monaten miteinander verglichen. Als multilateraler Index bestehen auch beim GEKS-Ansatz die gleichen Herausforderungen wie bei der QAUV-Methode hinsichtlich der Verkettung der Ergebnisse. Daher wurden Zeitfenster (13 Monate) und Verkettungsart (fixed based moving window) genauso gewählt wie beim QAUV-Index (siehe Übersicht 3). Die Berechnung der GEKS-Indizes je COICOP orientierte sich an Ivancic, Diewert und Fox (2011):

$$I_t^{GEKS} = \prod_{z \in T} \left(\frac{I_{t,z}^{Fish}}{I_{0,z}^{Fish}} \right)^{\frac{1}{|T|}}$$

mit

$$I_{t,z}^{Fish} = \sqrt{I_{t,z}^L \cdot I_{t,z}^{Pa}}$$

wobei

$$I_{t,z}^L = \frac{\sum_{r=1}^{R_{t,z}} \sum_{i=1}^{N_{t,z}^r} p_t^{r,i} \cdot q_z^{r,i}}{\sum_{r=1}^{R_{t,z}} \sum_{i=1}^{N_{t,z}^r} p_z^{r,i} \cdot q_z^{r,i}}$$

und

$$I_{t,z}^{Pa} = \frac{\sum_{r=1}^{R_{t,z}} \sum_{i=1}^{N_{t,z}^r} p_t^{r,i} \cdot q_t^{r,i}}{\sum_{r=1}^{R_{t,z}} \sum_{i=1}^{N_{t,z}^r} p_z^{r,i} \cdot q_t^{r,i}}$$

Dabei ist:

I_t^{GEKS} der GEKS-Index für Monat t

$I_{t,z}^{Fish}$ der Fisher-Index für Monat t mit Monat z als Preisbasis

$I_{t,z}^L$ der Laspeyres-Index für Monat t mit Monat z als Preisbasis

$I_{t,z}^{Pa}$ der Paasche-Index für Monat t mit Monat z als Preisbasis

$p_t^{r,i}$ der Preis von Einzelhandelsunternehmen r für Artikel i in Monat t

$q_t^{r,i}$ der Absatz von Einzelhandelsunternehmen r mit Artikel i in Monat t

$N_{t,z}^r$ die Anzahl der in der Stichprobe sowohl in Monat t als auch in Monat z gleichzeitig enthaltenen Artikel von Einzelhandelsunternehmen r

$R_{t,z}$ die Anzahl der in der Stichprobe sowohl in Monat t als auch in Monat z gleichzeitig enthaltenen Einzelhandelsunternehmen

T die Menge an Monaten, die bei der Berechnung der Preisindexreihe berücksichtigt werden

Bei der Berechnung einer Preisindexreihe wurden für jeden Monat im Zeitfenster zunächst 13 verschiedene Fisher-Indizes erstellt und anschließend jeweils das geometrische Mittel dieser 13 Fisher-Indizes gebildet. Abschließend wurde für die Indexreihe das geometrische Mittel der Fisher-Indizes für Dezember 2015 als Indexbasis herangezogen. Wie beim QAUV-Index wurden für die Preisindizes von Februar 2016 bis Dezember 2016 jeweils neue Preisindexreihen berechnet und danach nur der Monat am aktuellen Rand mit der durchgängigen Preisindexreihe verkettet.

6

Analyse der Ergebnisse

Mit den in Kapitel 5 beschriebenen Indexformeln wurden für die 44 im Datensatz enthaltenen Produktgruppen je drei scannerdatenbasierte Preisindizes für den Zeitraum von Januar 2015 bis Dezember 2016 berechnet. Beispielfhaft werden in [Grafik 3](#) die Indizes für die vier COICOP-Positionen Margarine, Honig, Bohnenkaffee und Orangensaft oder ähnlicher Fruchtsaft gezeigt und diese den jeweiligen Teilindizes des Verbraucherpreisindex als Vergleichsmaßstab gegenübergestellt. Dabei wurden die Indexreihen jeweils auf den Jahresdurchschnitt 2015 umbasiert. Tendenziell skizzieren die scannerdatenbasierten Preisindizes in Grafik 3 ähnliche Entwicklungen im Zeitverlauf wie der Verbraucherpreisindex.

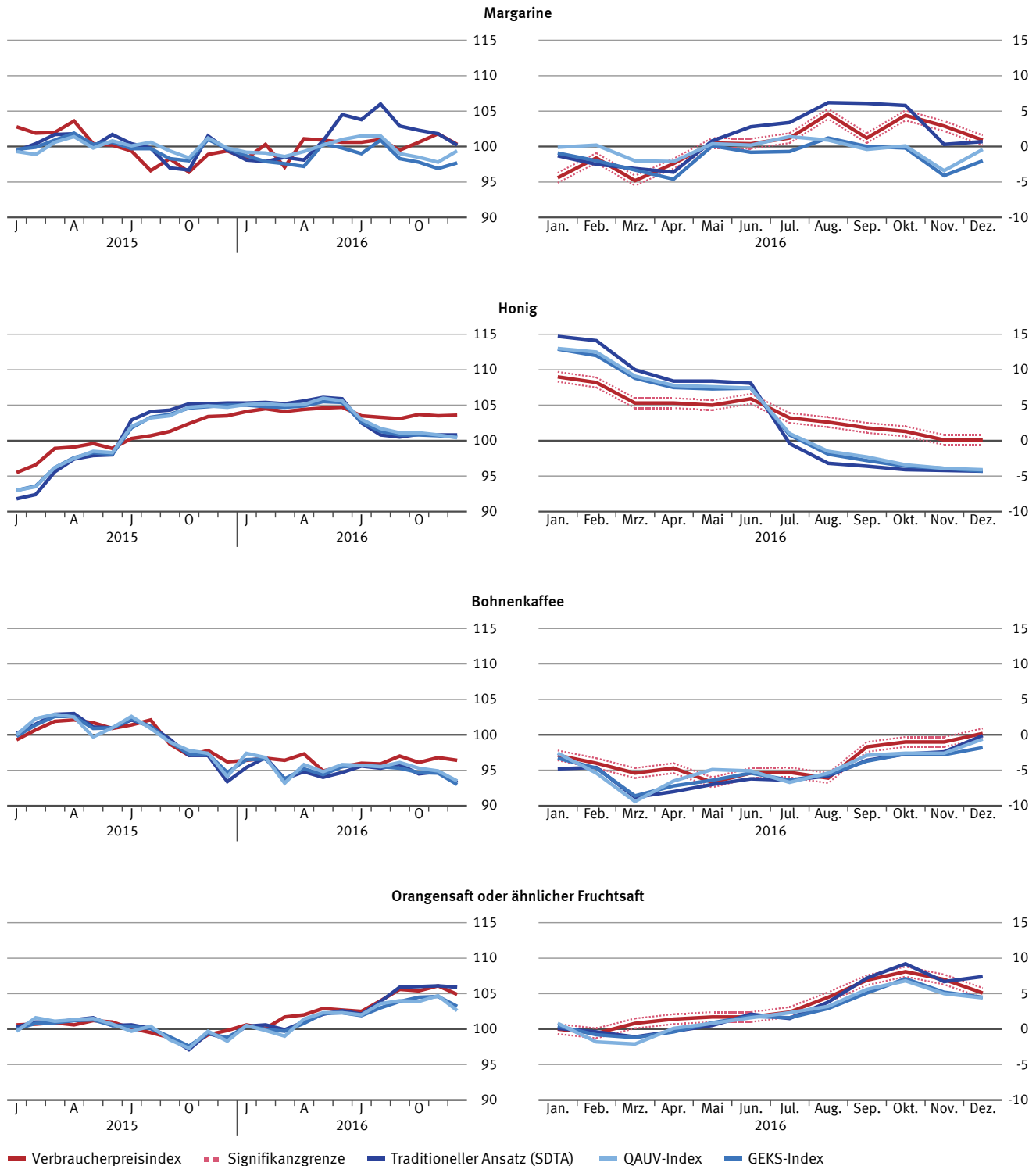
Es fällt auf, dass bei Margarine die Indexreihe des „traditionellen“ Ansatzes (SDTA) in der zweiten Jahreshälfte 2016 etwas stärker abweicht. Dies liegt daran, dass bei der in Abschnitt 5.2 beschriebenen automatisier-

Nutzung von Scannerdaten in der Preisstatistik – eine Untersuchung anhand von Marktforschungsdaten

Grafik 3
Preisentwicklungen ausgewählter Produktgruppen

Preisindizes
2015 = 100

Veränderungsraten zum Vorjahresmonat
in %



2019 - 01 - 0105

ten Suche nach Ersetzungsprodukten in einem Fall ein nicht adäquater Artikel ausgewählt wurde. Für künftige Untersuchungen müsste daher das Verfahren zur automatisierten Suche nach Ersetzungsprodukten optimiert werden.

Bei Bohnenkaffee sowie Orangensaft oder ähnlicher Fruchtsaft sind für die scannerdatenbasierten Indizes insbesondere Ende 2015 bis Anfang 2016 stärkere Preisschwankungen zu beobachten als beim entsprechenden Teilindex des Verbraucherpreisindex. Dies lässt sich möglicherweise darauf zurückführen, dass Sonderangebote durch den erweiterten Erhebungszeitraum (von wenigen Tagen auf drei Wochen je Monat) besser abgedeckt werden und diese durch die Verwendung von Unit-Values auch stärker zur Geltung kommen.

Um abzuschätzen, ob sich die scannerdatenbasierten Preisindizes signifikant von den jeweiligen Teilindizes des Verbraucherpreisindex unterscheiden, wurden die Veränderungsdaten zum Vorjahresmonat herangezogen, die in Grafik 3 ebenfalls dargestellt sind. Gemäß EU-Verordnung 2016/792¹³ ist eine methodische Änderung als signifikant zu deklarieren, wenn in einem beliebigen Zeitraum die Veränderungsrate des Harmonisierten Verbraucherpreisindex insgesamt zum Vorjahresmonat um mehr als 0,1 Prozentpunkte, eines COICOP-2-Stellers um mehr als 0,3 Prozentpunkte, eines COICOP-3-Stellers um mehr als 0,4 Prozentpunkte, eines COICOP-4-Stellers um mehr als 0,5 Prozentpunkte und eines COICOP-5-Stellers um mehr als 0,6 Prozentpunkte beeinflusst wird. Diesem Konzept folgend wurde für diese Untersuchung die Grenze für die COICOP auf 10-Steller-Ebene bei 0,7 Prozentpunkten gesetzt. Trotz der augenscheinlich relativ ähnlichen Indizes würde basierend auf den Ergebnissen dieser Untersuchung der Umstieg von traditioneller Preiserhebung hin zur Nutzung von Scannerdaten eine signifikante methodische Veränderung bedeuten, denn bei allen drei scannerdatenbasierten Indizes weichen die Veränderungsdaten zum Vorjahresmonat in über 60% der Fälle um mehr als 0,7 Prozentpunkte im Vergleich zur entsprechenden Veränderungsrate beim Verbraucherpreisindex ab. In Grafik 3 wird speziell bei Margarine und Honig deutlich, dass sich die Verände-

rungraten zum Vorjahresmonat zum Teil erheblich vom Verbraucherpreisindex unterscheiden.

Künftige Untersuchungen müssen zeigen, inwieweit sich diese Einschätzung auch mit „echten“ Scannerdaten bestätigt. Der vom Marktforschungsinstitut Nielsen eingekaufte Datensatz war nicht perfekt für die Berechnung von Preisindizes geeignet, weil Nielsen Zahlen für einige Einzelhandelsunternehmen schätzen musste und für manche Einzelhandelsunternehmen Ergebnisse nicht separat zur Verfügung stellen konnte. Zudem wurden für die Berechnungen national aggregierte Daten verwendet, was ursächlich dafür sein kann, dass die ermittelten Indizes nicht wie vorgesehen die „reine“ Preisveränderung messen. Stattdessen werden beispielsweise auch regionale Preisunterschiede in Kombination mit wechselndem Einkaufsverhalten als Preisveränderungen ausgewiesen.

7

Fazit und Ausblick

Im Zuge der vorliegenden Untersuchung konnten umfangreiche Erfahrungen bei der Berechnung von Preisindizes basierend auf Scannerdaten gewonnen werden. Hervorzuheben sind die Klassifizierung von Artikeln, die Suche nach Ersetzungsprodukten oder die Frage, welche Formel zur scannerdatenbasierten Indexberechnung herangezogen werden soll. Durch die Arbeiten an diesem Projekt konnte im Statistischen Bundesamt Know-how aufgebaut werden, das zur Überwindung der Hürden bei der Berechnung von Preisindizes aus Scannerdaten unabdingbar ist.


Da das Statistische Bundesamt inzwischen „echte“ Transaktionsdaten im Bereich Nahrungsmittel einwerben konnte, können die in diesem Projekt getesteten Ansätze in den nächsten Schritten auf „echte“ Scannerdaten angewandt und weiterentwickelt werden. Neben dem QAUV- und dem GEKS-Index sollen in diesem Zusammenhang auch noch weitere international diskutierte Berechnungsverfahren wie die Time-Product-Dummy (TPD)-Methode getestet werden.

Neben der Anwendung muss außerdem eine Bewertung der verschiedenen Ansätze erfolgen, sodass eine fundierte Entscheidung zur Methodik getroffen werden

13 Verordnung (EU) 2016/792 des Europäischen Parlaments und des Rates vom 11. Mai 2016 über harmonisierte Verbraucherpreisindizes und den Häuserpreisindex sowie zur Aufhebung der Verordnung (EG) Nr. 2494/95 des Rates (Amtsblatt der EU Nr. L 135, Seite 11).

kann, mit der Scannerdaten in den Produktionsprozess des Verbraucherpreisindex integriert werden sollten. Unabhängig vom Berechnungsansatz deuten die Ergebnisse in Kapitel 6 darauf hin, dass die Implementierung von Scannerdaten eine erhebliche methodische Änderung für die betroffenen Produktgruppen des Verbraucherpreisindex bedeuten würde.

Zur Implementierung von Scannerdaten in der deutschen Verbraucherpreisstatistik müssen jedoch nicht nur methodische Fragestellungen evaluiert, sondern es muss auch eine geeignete IT-Infrastruktur geschaffen werden. Diese umfasst neben der in diesem Aufsatz beschriebenen Datenaufbereitung den Dateneingang, die Datenhaltung sowie eine Schnittstelle mit dem Verbundprogramm, mit dem monatlich die Berechnung der Verbraucherpreisindizes durchgeführt wird. Zudem wird an einer Anpassung der nationalen Gesetzesgrundlage gearbeitet, um regelmäßige Lieferungen von Scannerdaten an die Statistischen Ämter des Bundes und der Länder sicherzustellen.

Zusammengefasst lässt sich festhalten, dass die vorliegende Untersuchung wichtige Erkenntnisse zur Nutzung von Scannerdaten in der deutschen Verbraucherpreisstatistik liefert. Generell ist die Verwendung von Scannerdaten mit neuen, teilweise großen Herausforderungen verbunden, deren Bewältigung jedoch bisher als machbar erscheint. Allerdings sind noch eine Vielzahl an Arbeiten auf unterschiedlichsten Ebenen erforderlich, um Scannerdaten als neue Datenquelle zu erschließen und in den preisstatistischen Produktionsprozess zu integrieren. 

LITERATURVERZEICHNIS

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

Chessa, Antonio G. *A new methodology for processing scanner data in the Dutch CPI*. In: Eurostat review of National Accounts and Macroeconomic Indicators. Ausgabe 1/2016, Seite 49 ff.

Chessa, Antonio G./Verburg, Johan/Willenborg, Leon. *A Comparison of Price Index Methods for Scanner Data*. Aufsatz vorgestellt beim 15. Treffen der Ottawa-Gruppe. Eltville am Rhein 2017.

Éltető, Ödön/Köves, Pál. *On an Index Computation Problem in International Comparisons*. Statiztikai Szemle. 1964. Band 42, Seite 507 ff. [auf Ungarisch].

Gini, Corrado. *On the Circular Test of Index Numbers*. In: Metron. 1931. Band 9, Seite 3 ff.

Geary, Robert Charles. *A Note on the Comparison of Exchange Rates and Purchasing Power between Countries*. In: Journal of the Royal Statistical Society A. Band 121. Nummer 1/1958, Seite 97 ff.

Hastie, Trevor/Tibshirani, Robert/Friedmann, Jerome. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. Auflage. New York 2009.

Ivancic, Lorraine/Diewert, Walter Erwin/Fox, Kevin J. *Scanner Data, Time Aggregation and the Construction of Price Indices*. In: Journal of Econometrics. Band 161. Ausgabe 1/2011, Seite 24 ff.

Khamis, Salem Hanna. *A New System of Index Numbers for National and International Purposes*. In: Journal of the Royal Statistical Society A. Band 135. Nummer 1/1972, Seite 96 ff.

Lamboray, Claude. *The Geary Khamis index and the Lehr index: how much do they differ?* Aufsatz vorgestellt beim 15. Treffen der Ottawa-Gruppe. Eltville am Rhein 2017.

Szulc, Bohdan. *Index Numbers of Multilateral Regional Comparisons*. In: Przegląd Statystyczny. Band 3/1964, Seite 239 ff. [auf Polnisch].

van Loon, Ken/Roels, Dorien. *Integrating big data in the Belgian CPI*. Aufsatz vorgestellt beim Treffen der Expertengruppe zu Verbraucherpreisindizes. Genf 2018.

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 2019

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

Print

Einzelpreis: EUR 19,- (zzgl. Versand)

Jahresbezugspreis: EUR 114,- (zzgl. Versand)

Bestellnummer: 1010200-19002-1

ISSN 0043-6143

ISBN 978-3-8246-1081-5

Download (PDF)

Artikelnummer: 1010200-19002-4, 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: Design Offset, FSC-zertifiziert

© Statistisches Bundesamt (Destatis), 2019

Vervielfältigung und Verbreitung, auch auszugsweise, mit Quellenangabe gestattet.