

Ingolf Boettcher

Messung der Preise
Berlin
8. Juni 2017

Scannerdaten

Vergleich unterschiedlicher Ansätze bei
der Preisindexberechnung mit Scanner-
daten

Spezialisierung des Konsumangebotes	Preisgestaltung des Einzelhandels
Höhere Sortimentsbreite	Steigender Promotionsanteil
Stärkere Segmentierung der Warengruppen	Mehr Promotionsarten
Neue Handelskanäle (z.B. Onlinehandel)	Häufigkeit der Promotionen



Qualität des VPI lässt sich nur durch die Verbesserung bzw. Vergrößerung der Datenbasis sichern

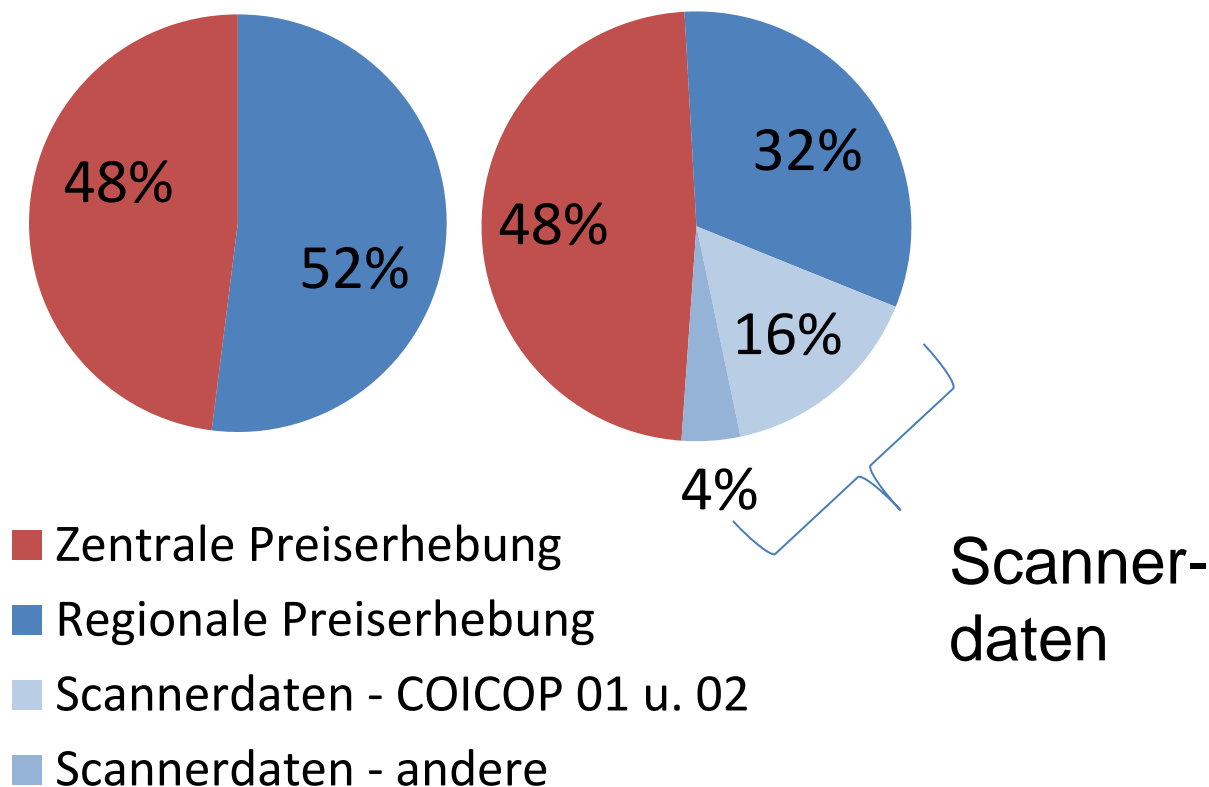
$$p_i^{t,f} = \frac{Umsatz_i^{t,f}}{Absatz_i^{t,f}}$$

Annahme:

Der Unit Value für das Produkt i ist näher am tatsächlichen Transaktionspreis als der Listenpreis des Produkts.

Scannerdaten können einen Teil der regionalen, manuellen Preiserhebung ersetzen.

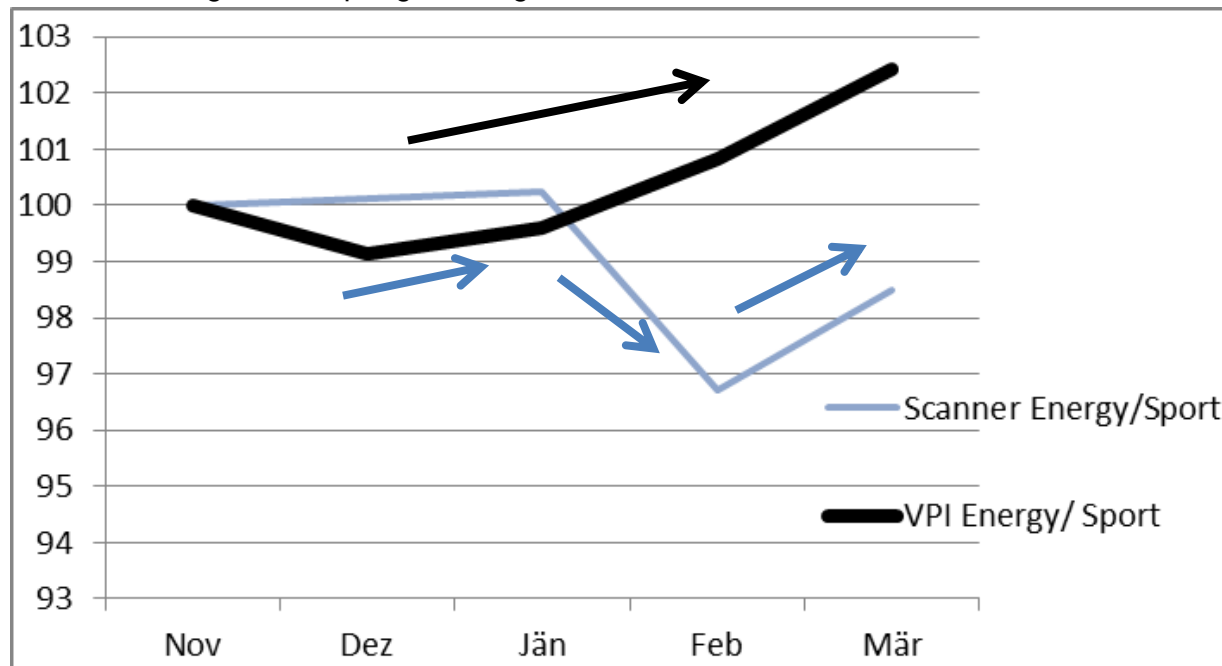
Gewichtsanteile Datengrundlage im VPI



Fallbeispiel: Energy- /Sportdrinks – Detailindex

Teilweise signifikante Unterschiede der gemessenen Preisentwicklungen

Chart 2 – Energie- und Sportgetränkegetränke November 2011 – März 2012



Quelle: Statistik Austria, Pilotprojekt ‚Scannerdaten‘ (vorläufige Werte)

Fallbeispiel: Energy- /Sportdrinks – Detailindex

Gründe für signifikante Unterschiede:

- Umsatzstarke Eigenmarken besitzen im Scannerdaten-Index ein höheres Gewicht als im VPI
- Preissenkungen der Eigenmarken werden durch Scannerdaten stärker berücksichtigt

Scannerdaten	Konventionelle Preiserhebung
Mehrere tausend Unit Values	Ca. 100 Preise
>50 einbezogene Produkte	Ca. 12 Produkte ausgewählt
4-wöchiger Erhebungszeitraum	Nur ein Erhebungstichtag



Höhere Indexqualität durch Scannerdaten

Index- Berechnungsmethode für VPI-Elementaraggregate	T1 Formel $\pi_a^{t1/t0} =$	T2 Formel $\pi_a^{t2/t0} =$
1) unweighted fixed base geometric mean of prices (HVPI Standard)	$\prod_{i=1}^n \left(\frac{p_i^{t1}}{p_i^{t0}} \right)^{\frac{1}{n}}$	$\prod_{i=1}^n \left(\frac{p_i^{t2}}{p_i^{t0}} \right)^{\frac{1}{n}}$
2.) unweighted monthly chained Jevons Index (Niederländische Methode)	$\prod_{i=1}^n \left(\frac{p_i^{t1}}{p_i^{t0}} \right)^{\frac{1}{n}}$	$\prod_{i=1}^n \left(\frac{p_i^{t1}}{p_i^{y,t0}} \right)^{\frac{1}{n}} \times \prod_{i=1}^n \left(\frac{p_i^{t2}}{p_i^{t1}} \right)^{\frac{1}{n}}$
3) Monthly Chained Superlative Index – (Törnqvist)	$\prod \left(\frac{p_i^{t1}}{p_i^{t0}} \right)^{\frac{(s_i^{t0} + s_i^{t1})}{2}}$	$\prod \left(\frac{p_i^{t1}}{p_i^{t0}} \right)^{\frac{(s_i^{t0} + s_i^{t1})}{2}} \times \prod \left(\frac{p_i^{t2}}{p_i^{t1}} \right)^{\frac{(s_i^{t1} + s_i^{t2})}{2}}$
<p><u>Wobei:</u></p> <p>p_i^t :Preis für Artikel i zum Zeitpunkt t</p> <p>$\pi_a^{t1/t0}$:die Preisniveauentwicklung eines Elementaraggregates zwischen den Zeitpunkten t0 und t1</p> <p>n :Anzahl der Warenkorbpositionen die einen für ein Elementaraggregat zugeordnet werden</p> <p>s_i^t :Marktanteil des Artikels i zum Zeitpunkt t</p>		

“When estimating price indices on the whole population of items in scanner data rather than on relatively small samples, a multilateral method seems preferable in order to avoid chain drift.” (UNECE 2016)



Rolling Year GEKS (RYGEKS); Ivancic, L., W.E. Diewert and K.J. Fox (2009)

Quality Adjusted Unit Value (QAUV); Chessa (2016)

Fixed Effects with a Window Splice (FEWS) ; *Krsinich, F. (2016)*

2017 Ottawa Group: Weighted Time-Product Dummy; Time Dummy
Hedonic Indexes; Geary-Khamis index;
(siehe bundesbank.de/ottawagroup/)

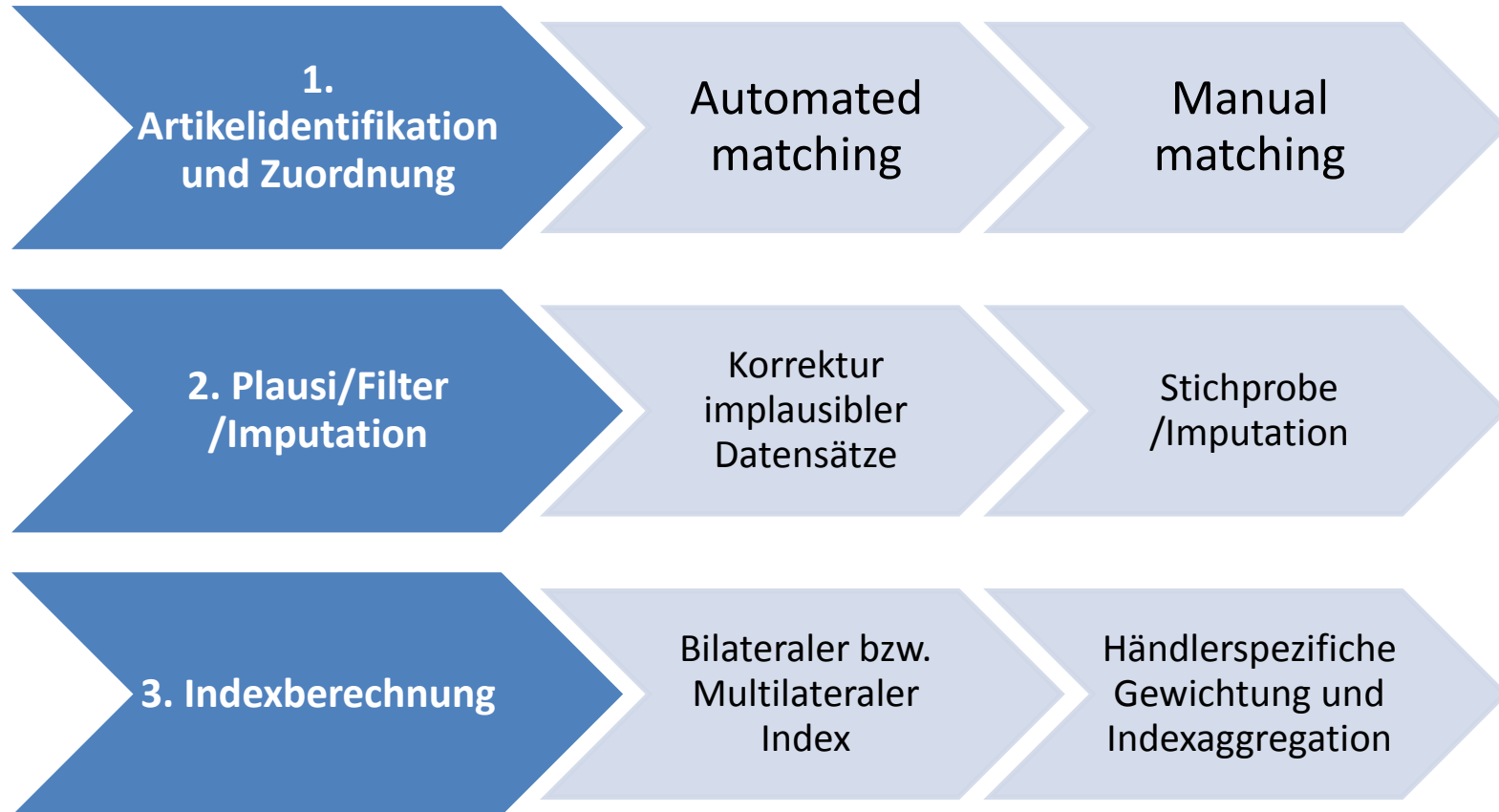
Rolling Year GEKS (RYGEKS); Ivancic, L., W.E. Diewert and K.J. Fox (2009)

$$\prod_{s=1}^{13} [P_{t,s}/P_{t-1,s}]^{1/13}$$

Schema of the RYGEKS with the window of 13 months

		GEKS for months: 0-12											GEKS _{12,0}				
1st 13 months		M ₀	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂			
			GEKS for months: 1-13											GEKS _{13,12}			
14 months	M ₀	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	M ₁₃	=> P _{13,0} = GEKS _{12,0} * GEKS _{13,12}		
				GEKS for months: 2-14											GEKS _{14,13}		
15 months	M ₀	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	M ₁₃	M ₁₄	=> P _{14,0} = GEKS _{12,0} * GEKS _{13,12} * GEKS _{14,13}	
					GEKS for months: 3-15											GEKS _{15,14}	
16 months	M ₀	M ₁	M ₂	M ₃	M ₄	M ₅	M ₆	M ₇	M ₈	M ₉	M ₁₀	M ₁₁	M ₁₂	M ₁₃	M ₁₄	M ₁₅	=> P _{15,0} = GEKS _{12,0} * GEKS _{13,12} * GEKS _{14,13} * GEKS _{15,14}

Preisindexerstellung mit Scannerdaten neuartige Arbeitsschritte



Scannerdaten müssen über verschiedene Dimensionen hinweg aggregiert werden:

- **Zeit**
Tag vs. Woche(n) vs. Monat
- **Raum**
Markt vs. PLZ vs. Region vs. Bundesland vs. Nation
- **Produkt**
GTIN/EAN vs. PLU vs. Homogenes Produkt
- **Sektor**
Hypermarkt vs. Supermarkt vs. Express

„How much aggregation is desirable as turn-over and quantities of GTIN/Products/Homogenous-goods are aggregated over outlets?“

Outlet A

$$P = \frac{\sum(p_{c,t_n} \cdot q_c)}{\sum(p_{c,t_0} \cdot q_c)}$$

Outlet B

$$P = \frac{\sum(p_{c,t_n} \cdot q_c)}{\sum(p_{c,t_0} \cdot q_c)}$$

Outlet C

$$P = \frac{\sum(p_{c,t_n} \cdot q_c)}{\sum(p_{c,t_0} \cdot q_c)}$$

VS.

Σ

Outlet A
Outlet B
Outlet C

$$P = \frac{\sum(p_{c,t_n} \cdot q_c)}{\sum(p_{c,t_0} \cdot q_c)}$$

Scannerdaten werden vor der Verwendung zur Indexberechnung bereinigt, eventuell imputiert und ausgewählt, z.B. durch:

$$\frac{P_i^{t1}}{P_i^{t0}} \geq 4 \text{ ODER } \frac{P_i^{t1}}{P_i^{t0}} \leq 0,25$$

→ Nicht-Berücksichtigung von P_i^{t1}

$$\frac{P_i^{t1}}{P_i^{t0}} \leq 0,8 \text{ UND } \frac{\text{Umsatz}_i^{t1}}{\text{Umsatz}_i^{t0}} \leq 0,2$$

→ *Nicht-Berücksichtigung von P_i^{t1}*

Scannerdaten werden vor der Verwendung zur Indexberechnung bereinigt, eventuell imputiert und ausgewählt:

- **Inkonsistente Datensätze**
z.B. Missings, falsche Gruppenzuordnung, unlogische Umasatz- und Absatzangaben, unlogische Änderung zum Vormonat (zu hohe Preisunterschiede)
- **Unerwünschte Produktangebote**
z.B. Abverkäufe, Mengenaktionen (4 zum Preis von 3), etc.
- **Marktanteil**
in den Index sollen Waren eingehen die ?% des Umsatzes zum Zeitpunkt t? repräsentieren

Scannerdaten werden vor der Verwendung zur Indexberechnung den einzelnen Elementaraggregaten und COICOP zugeordnet, durch:

- **Manuelle Zuordnung**
nur für Festbasenansatz denkbar (Verlinkung Dezember t-1)
und einer relativ restriktiven Cut-Off-Stichprobe (z.B. nur Top 5)
- **Händlerklassifikation**
meistens hilfreich, aber häufige Änderungen und im allg. sehr handelsorientiert (z.B. Klassifikation nach Verpackungsart)
- **Machine learning**
derzeit starke Bemühungen Zuordnung und Verlinkung zu automatisieren
- **Stammdaten externer Anbieter**
Marktforschung , GS1 (erstellt GTINs)

Scannerdaten werden vor der Verwendung zur Indexberechnung mit den jeweiligen Scannerdaten einer früheren Periode verlinkt

- **Verlinkung über GTIN/EAN**
viele Missings, Produkt Re-Launches werden nicht erkannt
- **Verlinkung über PLU**
gut, steht aber nicht immer zur Verfügung
- **Verlinkung über Produkteigenschaften**
Definition eines homogenen Produktes, unabhängig von GTIN/EAN/PLU (z.B. Coca Cola, PET, 1 Liter → homogen)
- **Verlinkung über Produktgruppen**
zu heterogen
- **Machine learning**
derzeit starke Bemühungen Zuordnung und Verlinkung zu automatisieren
- **Stammdaten externer Anbieter**
Marktforschung, GS1 (erstellt GTINs)

„Pragmatischer Statischer Ansatz“

- **Feste Basis**
Preisrelationen zwischen t0 und Dezember des Vorjahres
- **Manuelle Verlinkung über EAN bzw. Stammdaten externer Anbieter**
- **Bilaterale Indizes** (normalerweise Jevons)
- **Restriktive Filter**

Geeignet für:

- **Schnelle SD-Einführung**
- **Geringe Ressourcen für mtl. quantitative statistische Auswertungen**
- **Hohen Transparenzbedarf**
- **Länder mit hoher Marktkonzentration und geringen regionalen Preis- und Sortimentsunterschieden**
- **Wenige Marktführer in einzelnen Produktsegmenten**

„Pragmatischer Dynamischer/Quantitativer Ansatz“

- **Monthly Chaining**
Preisrelationen zwischen t_0 und Vormonat
- **Größtenteils automatisierte Verlinkung über EAN /PLU**
- **Bilaterale Indizes** (normalerweise Jevons)
- **Cut-Off Filter**

Geeignet für:

- **Länder mit vielen SD-Lieferanten / regionalen Preis- und Sortimentsunterschieden**
- **SD mit detaillierten Produktcharakteristika**
- **Ressourcen für mtl. quantitative statistische Auswertungen vorhanden**
- **Viele Marktführer in einzelnen Produktsegmenten**

„Umfassender Ansatz“

- **Monthly Chaining**
Preisrelationen zwischen t0 und Vormonat
- **Größtenteils automatisierte Verlinkung über EAN /PLU/sowie Produkteigenschaften**
- **Multilaterale Indizes**
- **Nur Datenbereinigung, keine/kaum Filter**

Geeignet für:

- **Länder mit vielen SD-Lieferanten / regionalen Preis- und Sortimentsunterschieden**
- **quantitative statistische Auswertungen**
- **Viele Marktführer in einzelnen Produktsegmenten**
- **Flexiblen Revisionsstandards**
- **SD mit detaillierten Produktcharakteristika**

Contact:
Ingolf Boettcher

Guglgasse 13, 1110 Wien

Tel: +43 (1) 71128-7917

Fax: +43 (1) 7180718

ingolf.boettcher@statistik.gv.at

Scannerdaten

Vergleich unterschiedlicher Ansätze bei
der Preisindexberechnung mit
Scannerdaten

Chessa (2016) Processing scanner data in the Dutch CPI: A new methodology and first experiences https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Session_1_Netherlands_Processing_scanner_data_in_the_Dutch_CPI.pdf

Eurostat (2017) Practical Guide for Processing Supermarket Scanner Data, Draft Version May 2017

Ivancic, L., W.E. Diewert and K.J. Fox (2009) Scanner Data, Time Aggregation and the Construction of Price Indexes”, Discussion Paper 09-09, Department of Economics, University of British Columbia, Vancouver, Canada. http://econ.sites.olt.ubc.ca/files/2013/06/pdf_paper_erwin-diewert-09-9-scanner-data.pdf

de Haan, J. (2015) A Framework for Large Scale Use of Scanner Data in the Dutch CPI. Paper presented at the fourteenth Ottawa Group meeting, 20-22 May 2015, Tokyo, Japan. http://www.stat.go.jp/english/info/meetings/og2015/pdf/t6s11p33_pap.pdf

Krsinich, F. (2016) The FEWS Index: Fixed Effects with a Window Splice. Journal of Official Statistics, 32, 375-404. <https://www.degruyter.com/downloadpdf/j/jos.2016.32.issue-2/jos-2016-0021/jos-2016-0021.pdf>

UNECE (2016) Economic Commission for Europe Conference of European Statisticians Group of Experts on Consumer Price Indices Thirteenth session Geneva, 2 – 4 May 2016 Report https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2016/Report_CPI_2016_of_7_June.pdf