

# **Bericht vom 16. Treffen der Ottawa Group**

## Rio de Janeiro, 08. - 10. Mai 2019

Elisabeth Wieland\*, Deutsche Bundesbank

# Inhalt

1. Hintergrund und Überblick
2. Schwerpunkt des 16. Treffens der Ottawa Group 2019: Einsatz neuer Datenquellen in der Verbraucherpreisstatistik
3. Ausblick und Diskussion

# 1. Hintergrund

## Die Ottawa Group

- Richtungsweisende **internationale Arbeitsgruppe zur Preisstatistik**.
- Forum zur **methodischen Fortentwicklung** der internationalen Preisstatistik und zum **Erfahrungsaustausch bezüglich der Messprobleme** von Preisveränderungen.
- **Teilnehmerkreis: Forscher** (Statistik/Indextheorie) sowie **Datenproduzenten** (insb. statistische Ämter) und **Nutzer** (internationale Organisationen, Zentralbanken).
- Praktische Relevanz der statistischen Themen und die Anwendbarkeit möglicher Problemlösungen stehen im Vordergrund.

# 1. Überblick

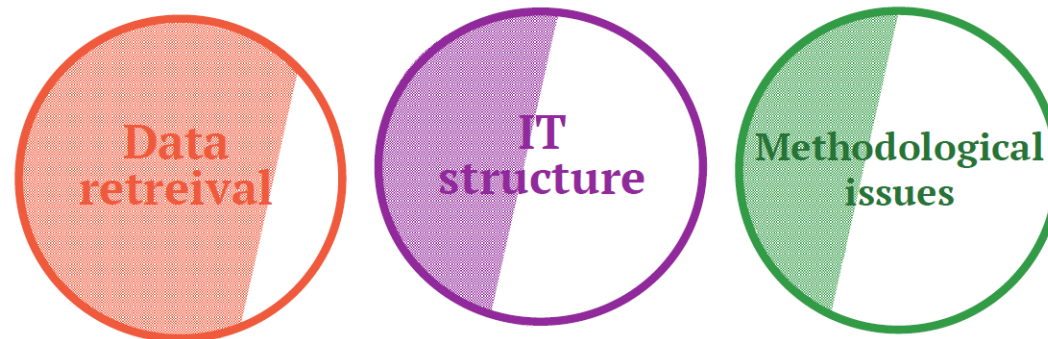
## Ottawa Group 2019: Themen und Teilnehmer

- Am **Treffen der Ottawa Group 2019** nahmen mehr als 80 Teilnehmer aus rund 40 Ländern teil.
  - „Invited Lecture“ von E. Diewert: [„Quality Adjustment and Hedonics: A unified approach“](#).
  - 26 Vorträge im Plenum und fast 30 Präsentationen in den Postersessions.
- Schwerpunkte des Treffens:
  - Der **Einsatz neuer Datenquellen** in der Verbraucherpreisstatistik (Scannerdaten, Webscraping) und die hieraus resultierenden **Herausforderungen an die Produktdefinition**.
  - Methoden der **Qualitätsbereinigung** und **Indexberechnung** (siehe hierzu auch die Vorträge von [L. von Auer](#) und [B. Goldhammer et al.](#)).
  - Messung von (Wohn-) **Immobilienpreisen**.

## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Nutzung von Webscraping und Scannerdaten

- Rapide Entwicklung in der amtlichen Preisstatistik seit Beginn der 2010er Jahre.
  - Vermehrt digitale Datenquellen wie Webscraping und Scannerdaten verfügbar.
  - Fortentwicklung von (multilateralen) Indexmethoden.
- Statistikämter stehen Herausforderungen bzgl. **Datenzugang**, Umbau der **IT-Infrastruktur** und **methodischer Fragestellungen** gegenüber.



Quelle: [Ståhl und Nilsson \(2019\)](#), S. 5.

## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Anwendungsbeispiele in der amtlichen Preismessung

- In einigen EU-Ländern finden Scannerdaten (S), Transaktionsdaten (T) und Webscraping (W) bereits Anwendung im Harmonisierten Verbraucherpreisindex (HVPI):

COICOP-Abteilung	Anwendungsbeispiele der amtlichen EU-Statistik
01 Nahrungsmittel und alkoholfreie Getränke	<i>Vollständige Produktabdeckung möglich (S)</i>
02 Alkoholische Getränke und Tabak	<i>Vollständige Produktabdeckung möglich (S)</i>
03 Bekleidung und Schuhe	<i>Vollständige Produktabdeckung möglich (W)</i>
04 Wohnung, Wasser, Elektrizität, Gas und andere Brennstoffe	Mieten für Studentenzimmer (W)
05 Hausrat und laufende Instandhaltung des Hauses	Möbel (W); Erzeugnisse des täglichen Haushaltsbedarfs (S/W)
06 Gesundheitspflege	Arzneimittel (W/T)
07 Verkehr	Benzin (T); Fernbus- und Bahnreisen, Gebrauch- und Mietwagen (W)
08 Nachrichtenübermittlung	Mobiltelefone (T)
09 Freizeit und Kultur	Notebooks (W); Pauschalreisen (T/W)
10 Erziehung und Unterricht	-
11 Restaurants und Hotels	Hotels (W)
12 Verschiedene Waren und Dienstleistungen	Erzeugnisse für die Körperpflege (S)

## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Webscraping und Transaktionsdaten

	Webscraping	Transaktionsdaten (Scannerdaten)
<b>Art des Preises</b>	Angebotspreise	Kaufpreise
<b>Mengenangaben?</b>	Nein	Ja
<b>Gemeinsame Herausforderungen</b>	<ul style="list-style-type: none"> <li>• Anpassung der <b>IT-Infrastruktur</b></li> <li>• Klärung von <b>Rechtsfragen</b></li> <li>• (Automatisierte) <b>Datenvalidierung</b></li> </ul>	
	<ul style="list-style-type: none"> <li>• <b>Produktdefinition</b></li> </ul>	
<b>Individuelle Herausforderungen in der regulären Datenproduktion</b>	<ul style="list-style-type: none"> <li>• <b>Dynamische bzw. personalisierte Preissetzung:</b> Preis ändert sich z.T. mehrmals am Tag</li> <li>• <b>Wechselnde Website-Architektur</b> und IP-Adressen</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Datenzugang</b> bzw. regelmäßige Datenlieferung sicherstellen</li> <li>• <b>Indexkonstruktion:</b> Nutzung der Mengenangaben</li> </ul>

## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Produktdefinition im HVPI

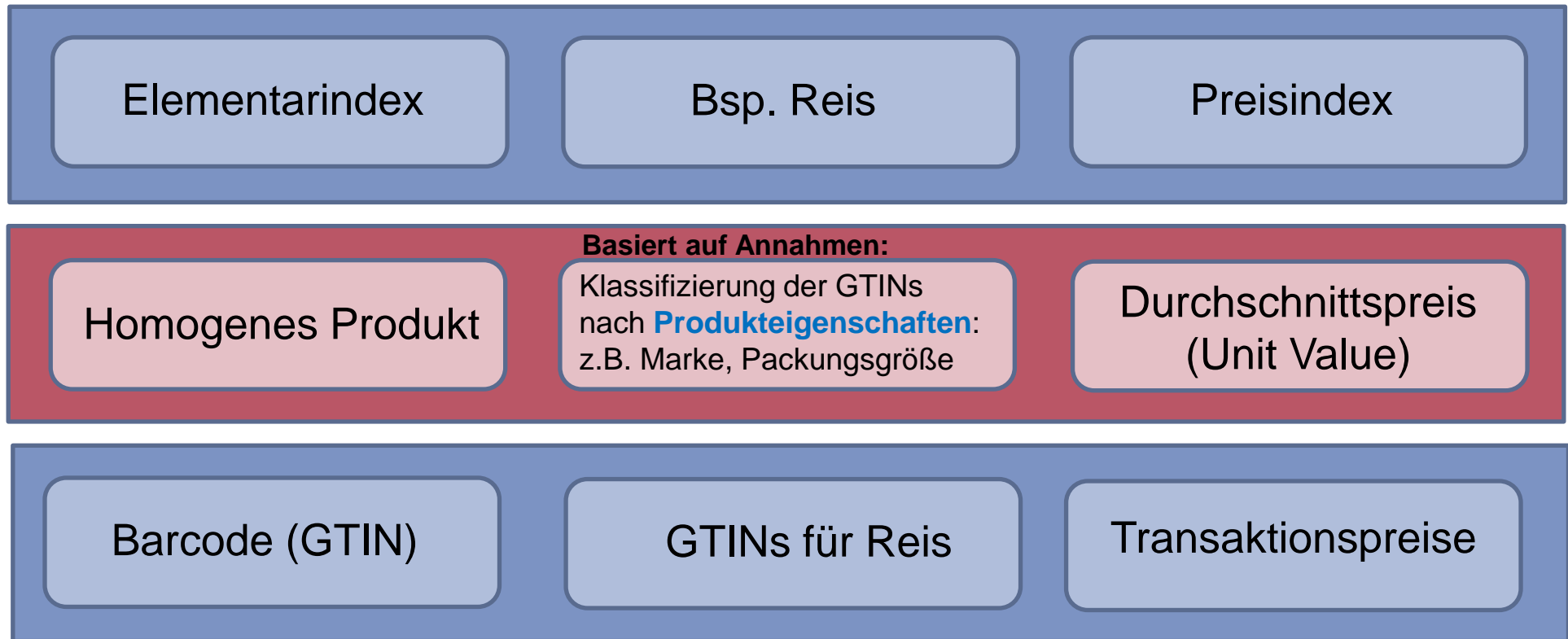
- HVPI: **Fixer Warenkorb**, d.h. was kosten die Produkte aus dem Referenzmonat 0 heute?
- Scannerdaten und Webscraping: **Produkt-ID** mittels
  - Barcode (Global Trade Item Number, GTIN), oder
  - händlerspezifischer ID (Stock-Keeping-Unit, SKU).
- Definition auf **Ebene der Produkt-ID** erweist sich i.d.R. als **zu detailliert**.
  - **Produktausfall** in folgenden Perioden, insb. bei hoher Produktfluktuation (“product churn”).
  - **Abwärtsdrift** im Preisindex, da Exit-Preis typischerweise niedriger ist als Preis bei Markteintritt.
- **Drei Gründe für Produktausfall** (vgl. [Nygaard et al., 2019](#)):
  - Stark saisonales Produkt (Bsp. Erdbeeren),
  - neue Produktgeneration (Bsp. Notebooks), oder
  - Relaunch/Update: Neue Produkt-ID z.B. aufgrund geänderter Verpackung oder Inhaltsstoffe.
- Lösung für letztgenannte Kategorie: **Homogene Produkte**, ähnlich der Einteilung in Konsumsegmente im Rahmen der traditionellen Preiserhebung.



## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Produktdefinition anhand von Supermarkt-Scannerdaten

**Homogene Produkte** als unterste Aggregationsebene:



In Anlehnung an [Chessa \(2019\)](#).



## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Ottawa Group: Produktdefinition als Herausforderung

- **Definition von Homogenen Produkten (HP)...**
  - erfordert eine (automatisierte) Klassifizierung der Produktcodes (GTIN/SKU).
  - ist abhängig von der Qualität der Metadaten.
- HP-Definition geeignet für **COICOP-Positionen mit hohem “product churn”** wie Bekleidung, dagegen problematisch bei Produkten mit unterschiedlicher Maßeinheit wie Scannerdaten ([Van Loon, 2019](#)).
- Trade-off zwischen Produkthomogenität und -kontinuität: **Maßzahl** zur Güte der HP-Definition ([Chessa, 2019](#)).
- **Sensitivitätstest** der vorgenommenen **HP-Definition**, z.B. Volatilität als Hinweis auf Fehlklassifikation ([Nygaard et al. 2019](#)).
- **Machine-Learning**-Ansätze zur Klassifizierung sind unumgänglich (Bsp. Webscraping/Bekleidung: [Rowland et al. 2019](#)).

## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Diskussionen im Rahmen der Ottawa Group 2019

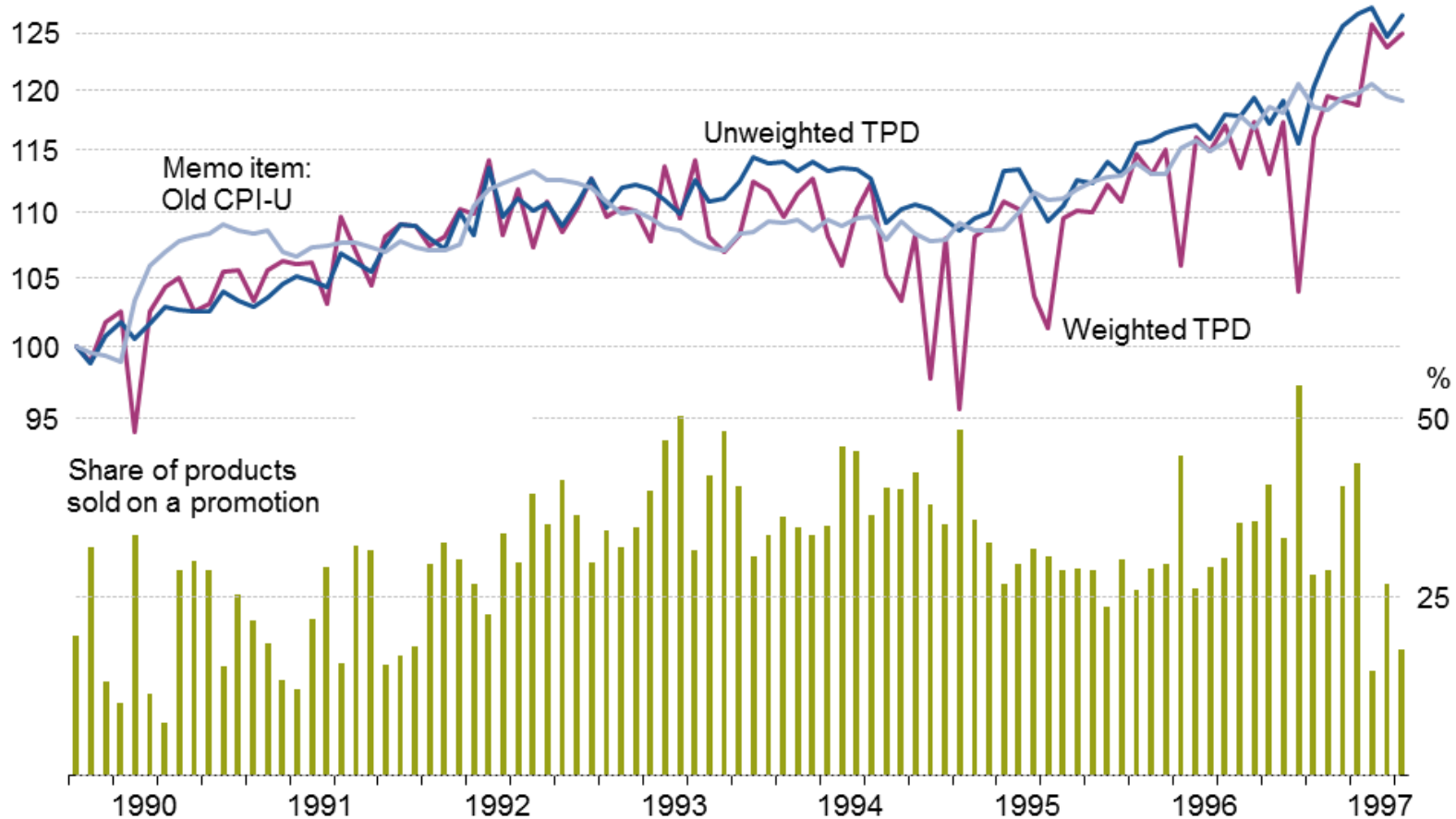
- **“Choices are many and answers are not obvious.”** ([Nygaard et al., 2019](#))
  - Definition von Homogenen Produkten sowie neue (multilaterale) Indexmethoden haben viele Stellschrauben.
  - **Systematische Auswertung** über verschiedene COICOP-Positionen hinweg oder anhand eines Benchmark-Datensatzes ([Mehrhoff, 2018](#))?
  - **Transparenz** und **Kommunikation** gegenüber Datennutzern ([Ståhl und Nilsson, 2019](#))?
- **“Leaving the fixed-basket approach for Electronic Transaction Data?”** ([Ståhl und Nilsson, 2019](#))
  - **Dynamischer Warenkorb** als Abkehr vom Laspeyres-Ansatz im HVPI?
- **“Scanner and web-scraped data are better in measurable terms – and worse, too!”** ([Mehrhoff, 2019](#))
  - Wird der Einsatz von digitalen Datenquellen die **Volatilität der HVPI-Komponenten** tendenziell erhöhen?
  - Scannerdaten spiegeln Abverkäufe wider: Ausgeprägteres Saisonprofil oder erratische Bewegungen?
  - **Parallelrechnung** vor Methodenumstellung wichtig!

## 2. Einsatz neuer Datenquellen in der Verbraucherpreisstatistik

### Mehr „Noise“ im Scannerdaten-basierten Preisindex

#### Prices for bottled juice, Dominick's Finer Foods

Oct 1989 = 100, log scale



## 3. Ausblick

- Alle Präsentationen und Papiere, die beim Treffen der Ottawa Group 2019 diskutiert wurden, werden auf folgender Internetseite zur Verfügung gestellt:  
<https://eventos.fgv.br/en/ottawa-group-meeting/publications>
- Informationen rund um die Ottawa Group, z.B. zu früheren Treffen, finden Sie unter folgendem Link: [www.ottawagroup.org](http://www.ottawagroup.org)
- Das **nächste Treffen der Ottawa Group** findet voraussichtlich im **Juni 2021 in Rom** statt.
- Update des **CPI Manuals**: [www.imf.org/en/Data/Statistics/cpi-manual](http://www.imf.org/en/Data/Statistics/cpi-manual)
  - Neue Kapitel: “Scanner data” & “Updating CPI weights and linking index series”
  - **Research Agenda**: Wo besteht Forschungsbedarf?  
→ Input für künftige Sitzungen der Ottawa Group sowie der Expertengruppe der Verbraucherpreisstatistik der UNECE (siehe [Präsentation](#) von Graf und Martins)

# Ausgewählte Präsentationen der Ottawa Group 2019

- Bertolotto, M. I. (2019): [“Online price index with product replacement: The closest match approach”](#).
- Chessa, A. G. (2019): [“MARS: A method for defining products and linking barcodes of item relaunches”](#).
- Flower, T. (2019): [“Using alternative data sources in consumer prices, UK”](#).
- Martindale, H., E. Rowland und T. Flower (2019), [“Machine learning for classification with big data in price statistics production pipelines”](#).
- Mehrhoff, J. (2019): [“Towards a new paradigm for scanner data price indices: applying big data techniques to big data”](#).
- Nygaard, R., L.-C. Zhang, I. Johansen (2019): [“Evaluating unit-value price indices in a dynamic item universe”](#).
- Ståhl, O., P. Nilsson (2019): [“Towards a roadmap for efficient use of electronic transaction data in the Swedish CPI”](#).
- Thomas, S. H., D. Ayoubkhani (2019): [“The use of approximate expenditure weights for web scraped data in consumer price indices”](#).
- Van Loon, K. (2019): [“Redefining what products are in the context of scanner data and web scraping, experiences from Belgium”](#).