



Scanner Daten: Von der derzeitigen Praxis zu Empfehlungen

Dr. Berthold Feldmann

Eurostat C4

Preisstatistik; Kaufkraftparitäten; Wohnungsstatistik

Struktur des Vortrags

- Hintergrund
- Abgedeckte Produktgruppen
- Abdeckung der Supermärkte
- Abdeckung der Ausgaben der privaten Haushalte
- Zeitliche Abdeckung
- Die Verbindung zu Konsumsegmenten
- Indexberechnung

Hintergrund

- Scanner-Daten werden seit mehr als zehn Jahren als Quelle des VPI/HVPI genutzt
- Immer mehr nationale statistische Ämter verwenden Scanner-Daten oder planen sie zu nutzen
- Daneben werden auch zunehmend elektronische Transaktionsdaten und Web-Scraping als Quelle genutzt
- Eurostat unterstützt dies durch Finanzhilfen
- Viele NSÄ wünschen konkrete Empfehlungen, wie Scanner-Daten zu nutzen sind

Alle Empfehlungen in dieser Präsentation sind noch in der Diskussion und somit vorläufig

Rechtlicher Rahmen

- Gemäss den HVPI Rahmenverordnung ist die Nutzung von Scanner-Daten grundsätzlich nicht zu beanstanden
- Es gibt allerdings einige Spannungen mit den derzeitigen Durchführungsverordnungen (Preisnachlässe, Saisonale Produkte)
- Die volle Vergleichbarkeit der Ergebnisse mit den über traditionelle Methoden erzielte muss immer geprüft werden

Die Verwender von Scanner-Daten

- Derzeit nutzen sechs NSÄ Scanner-Daten:
NO (1995), NL (2002), CH (2008), SE (2012),
BE, DK (2016)
- Ziel des Einsatzes von Scanner-Daten:
 - **Verbesserung der Qualität** des HVPI und VPI
 - **Effizientere Verfahren** in der Verarbeitung
- Das Ziel von Eurostat:
 - Die **Vergleichbarkeit** der HVPI sicherstellen
 - **Einhaltung des Rechtsrahmens**
 - Verstärkte **Zusammenarbeit** zwischen den NSÄ

Definition von Scanner-Daten

Scanner-Daten =

Informationen zum Verkauf von Verbrauchsgütern durch 'Scannen' der Strichcodes an der Kasse für einzelne Erzeugnisse im Einzelhandel. Die Daten liefern detaillierte Informationen über Menge, Merkmale und Wert der verkauften Waren

Klärung der Begriffe

Angebotspreise (Regalpreise)

Traditionell
erhoben

Web-Scraping

Transaktionspreise

Scanner-Daten

Andere
Transaktions-
daten

Abgedeckte Produktgruppen

- 02, 01 Nahrungsmittel, Getränke, Tabak – alle 6 NSÄ
- *03 Bekleidung – NL (Web-Scraping)*
- 05.3-6 Erzeugnisse des täglichen Haushaltsbedarfs – BE, DK, NL (einschl. Drogeriemärkten), NO, CH
- 06.1 Arzneimittel – NO
- 07.2.2 Benzin - NL, SE, NO
- 09.3.5 Artikel für Haustiere – CH
- 09.6 Pauschalreisen – NL
- 12.1.3.2 Erzeugnisse für die Körperpflege - alle 6 NSÄ

Empfehlungen in der Diskussion

- Für Produktgruppen, in denen die Produkte verhältnismäßig **homogen** sind, ist die Verwendung von **Scanner-Daten** angebracht
- Für Produktgruppen mit eher **heterogenen Produktmerkmalen** wird eher die **traditionelle Preiserhebung** empfohlen
- Die Einführung von Scanner-Daten für weitere Produktgruppen wird gefördert, die Ergebnisse müssen jedoch nachweislich vergleichbar sein

Abdeckung der Supermärkte

Abdeckung der Supermarktketten

- 60 % – DK (bald 80 %)
- 75 % – BE, CH
- 90 % – NL, SE
- 99 % – NO

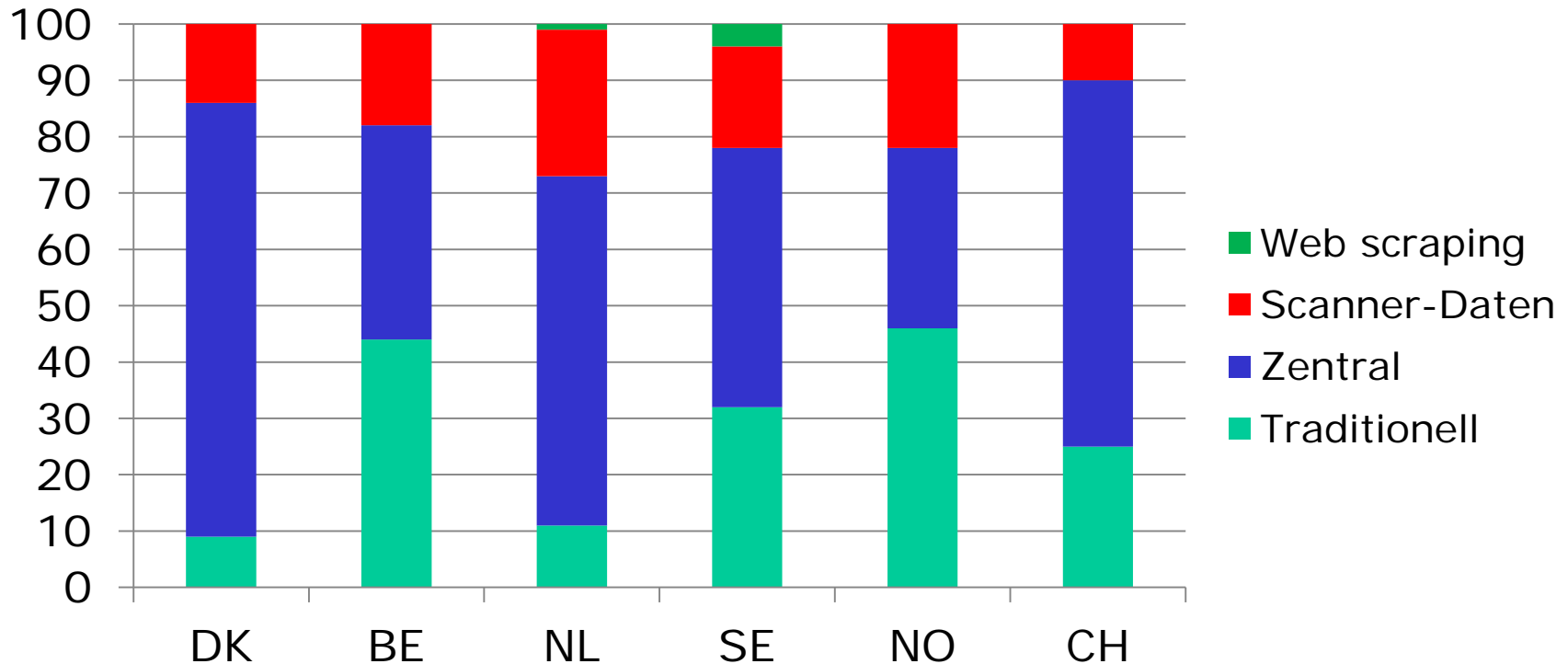
Stichprobe der Läden von Supermarktketten

- **Alle** Läden:
 - BE, DK, NL, CH
- Eine **Stichprobe** von Verkaufsstellen:
 - SE 60
 - NO 184

Empfehlungen in der Diskussion

- Scanner-Daten sollten **direkt von den Supermarktketten** eingeholt werden, nicht von einem Dritten
- **Alle Verkaufsstellen** der Supermarktketten sollten aufgenommen werden
- Die Daten sollten für alle Verkaufsstellen **aggregiert** sein, die im Hinblick auf ihre Leistungen homogen sind
- Ein Maximum von **Metadaten** sollte ermittelt werden

Abdeckung der Ausgaben der privaten Haushalte



Verträge mit Einzelhandelsketten

- Alle sechs NSÄ haben Verträge mit den Supermarktketten
- Kein NSA zahlt für die Daten

Rücksendung von Statistiken (als Belohnung)

- NL und CH

Notfallpläne, wenn keine Daten geliefert werden

- Drei NSÄ haben einen Notfallplan (DK, NO, CH)
- Er wurde nie genutzt

Empfehlungen in der Diskussion

- Eine **schriftliche Vereinbarung** (Vertrag) sollte mit den Einzelhandelsketten abgeschlossen werden
- Der Vertrag sollte eine **frühzeitige Warnung** im Falle von Änderungen bei der Übermittlung (Inhalt und Form) beinhalten

Zeitliche Abdeckung

- Wöchentliche Daten - alle NSÄ außer CH
- Die erste Woche, 2 Wochen, ganzer Monat – CH

Für den VPI / HVPI verwendet

- Erste drei Wochen des Monats – BE, NL, SE
- Zwei Wochen in Monatsmitte – DK
- Ersten zwei Wochen (1. bis 14. Tag) – CH
- Eine Woche (die Woche um den 15.) - NO

Empfehlungen in der Diskussion

- Die Scanner-Daten sollten **zu einer Woche aggregiert** sein, da tägliche Daten einen sehr großen Datensatz erzeugen, aber keinen weiteren Informationsgewinn bringen
- **Für jede Woche** des Monats sollten Daten eingeholt werden

Verknüpfung der Scanner-Daten mit den Konsumsegmenten

- Die NSÄ verknüpfen zunächst auf der Grundlage einer supermarkteigenen Klassifikation
- Verknüpfung mithilfe von supermarkteigenen Codes (SEK)
 - BE, CH (Beide erhalten auch GTIN (=EAN))
- Verknüpfung mit GTIN - DK, NL, SE, NO

Empfehlungen in der Diskussion

- Die Nutzung von SEK ist häufig überlegen, da diese stabiler sind als GTIN
- Beides (SEK und GTIN) sollte übermittelt werden

Der Verknüpfungsprozess

- Automatisch – NL, NO (Lebensmittel)
- Warenkorb von Hand – BE, DK, SE, CH

Empfehlungen Eurostats

- Eine **manuelle Verknüpfung** ist der Ausgangspunkt, wenn Scanner-Daten eingeführt werden (??)
- Die **automatische Verknüpfung** ist langfristig überlegen

Indexberechnung auf Elementarebene

- Alle sechs NSÄ berechnen einen **nicht gewichteten Jevons-Index** auf Elementarebene
- Unterschiede: Anzahl der verwendeten GTIN
 - **Automatische Zuordnung** und automatische Ausreißer Erkennung (GTIN oberhalb einer bestimmten Umsatzschwelle) → **hohe Anzahl**
 - Impliziert monatliche Verkettung
 - Verknüpfung **von Hand** → **weniger GTIN**

Empfehlungen in der Diskussion

- Die Methode der automatischen Verknüpfung ist dem Warenkorb-Ansatz vorzuziehen
 - *Der Warenkorb-Ansatz imitiert das Vorgehen der traditionellen Preiserhebung, nutzt aber nicht das volle Potenzial von Scanner-Daten*
- Für die erste Phase der Einführung von Scanner-Daten ist der Warenkorb-Ansatz angemessen (??)

Vergleich der Ansätze

Methoden	Vorteile	Nachteile
Traditionell	Die Methode hat sich bewährt	Kleine Stichprobe; Regalpreise; Abhängigkeit von Entscheidungen der Preiserheber
<i>Web-Scraping</i>	<i>Potenziell sehr große Stichprobe (Big Data)</i>	<i>Angebotspreis; Keine Umsätze</i>
Scanner-Daten – Warenkorb Ansatz	Große Stichprobe; tatsächliche Verkaufspreise; Zeitliche Abdeckung	Arbeitsintensiv; nutzt nicht das Potenzial
Scanner-Daten – dynamisches Konzept	Sehr große Stichprobe; tatsächliche Verkaufspreise; Zeitliche Abdeckung	leicht veränderter Warenkorb von Monat zu Monat

Erforderliches Personal der monatlichen Produktion (VZÄ)

beinhaltet

- Datenimport
- Monatliche Aktualisierung des Warenkorbes
- Verarbeitung von Scanner-Daten
- Überprüfung der Daten
- Transfer auf das allgemeine System des VPI/HVPI

- 1.5 Tage – DK
- 4-5 Tage – SE, NO
- 10 Tage – BE, NL, CH

Die Beteiligung Eurostats

- Eurostat unterstützt weiterhin die Entwicklung des **Warenkorb-Ansatzes** als Einstieg zu Scanner-Daten
- Eurostat unterstützt die Entwicklung von Methoden zur **probabilistischen Zuordnung** von Produktcodes zu Konsumsegmenten
- Eurostat unterstützt Forschungsarbeiten zur Verwendung von Transaktionsdaten für Produktgruppen außerhalb ECOICOP 1 und 2



**Vielen Dank für
Ihre
Aufmerksamkeit!**

**Fragen oder
Anregungen?**