DEUTSCHE BUNDESBANK EUROSYSTEM

Statistical assessment of probabilistic data sets vs. scanner and web-scraped data sources

Jens Mehrhoff*, Directorate General Statistics

* This presentation represents the author's personal opinions and does not necessarily reflect the views of the Deutsche Bundesbank or its staff.

Preamble

- **How the average rate of change in consumer prices is derived differs in multiple ways** with scanner and web-scraped data compared to more traditional data sources such as the collection of prices in the field.

- First and foremost, scanner and web-scraped data **give access to a much broader continuum of products** than classical sampling allows.
- The supposedly better coverage of goods and services comes at a cost, though: churn due to **new and disappearing products**, i.e. a dynamic product universe.

- Moreover, quantities sold (with scanner data) or at least a popularity ranking (from websites) become available too, thus **allowing the calculation of weighted indices** rather than the need to rely on unweighted formulae.
- The cost here is **chain drift**, i.e. the index might show spurious trends over time.

Processing data

- Typical steps of **processing scanner and web-scraped data** include but are not limited to
  1. (the **automatic classification** of products);
  2. intermediate **aggregation of "homogeneous" products**;
  3. (rule-based **filtering of observations**); and
  4. the **calculation of the final index**.

- A first step in the calculation of elementary indices is the **definition of the *homogenous product***; that is, the level at which a **unit value** is calculated. A trade-off between **product homogeneity and product continuity** arises.
- In particular **high churn and seasonal products** need special attention; for consumer electronics, say, **hedonic quality adjustment might still be the best option**. Eventually, product continuity must not be bought at the expense of (unit value) bias. → See discussion on quality adjustment.

Index calculation: overview

- If a **multilateral approach** to index calculation is chosen, several decisions have to be taken:
  - **which particular multilateral approach** should be implemented;
  - using **how many months as the estimation window**; and
  - how should the disseminated **time series be extended in real time** without revisions?
- There is **no consensus on the "right" answers** here and it might be more straightforward to **search for robust methods** – those that produce reliable estimates even for challenging product groups – rather than some economic or statistical justifications.

- In no case should **weighted indices** such as Fisher or Tornqvist be chain-linked at a high frequency such as monthly. These have shown to be **subject to severe drift**.

Index calculation: multilateral approaches

- Plenty of methods have been suggested for interspatial comparisons but the following three emerged to be **preferred in the time domain**:
  - **time-product dummy** (TPD);
  - **Geary-Khamis** (GK); and
  - **Gini-Eltetö-Köves-Szulc** (GEKS).

- For example, the **TPD method** derives the price index from a log-linear regression framework ($\delta_0 = \gamma_N = 0$):

$$\ln p_{i,t} = \alpha + \underbrace{\delta_t}_{t=1,\ldots,T} + \underbrace{\gamma_i}_{i=1,\ldots,N-1} + \varepsilon_{i,t}, \text{ and } P_{0,t} = \exp \hat{\delta}_t.$$

- TPD model estimates $\hat{\delta}_t = \sum_{i=1}^{N} s_{i,t}\left(\ln p_{i,t} - \hat{\alpha} - \hat{\gamma}_i\right)$ as independent time dummies, i.e. uses **cross-section averaging**.

Index calculation: estimation window and extension

- When using any of these methods **revisions** are, unfortunately, unavoidable. To circumvent this problem, the estimation window is shifted forward while keeping its length fixed and the new index is spliced onto an already disseminated figure.
- There is a growing literature on **how long the estimation window** should be and **how exactly the extension** should be performed.
- Typically, the estimation windows should cover **no less than 13 months**; 25 months when **products are seasonally unavailable**.
- The splicing is performed onto the **previous month (movement splice)**, the **same month in the previous year (window splice)** or something similar.

- Evidence points to that **some kind of anchoring** mitigates path dependency of the index; the classical chain-linking approaches reflect this by either referring to the **average of the previous year (annual overlap)** or **the last quarter/month of the previous year (one-quarter/month overlap)**.

Index calculation: estimation window and extension

- **Extending the time window** has the effect that the index loses what is known as **“characteristicity”**.
- The relative **differences in price levels** of the products are accounted for implicitly by multilateral methods. This adjustment is an average over the estimation window S; for example, in the TPD method:

$$\hat{\alpha} + \hat{\gamma}_i = \sum_{\tau \in S} \frac{s_{i,\tau}}{\sum_{\tau \in S} s_{i,\tau}} \ln \frac{p_{i,\tau}}{P_{0,\tau}}.$$

- However, should products within the elementary aggregate show differing trends, that **time average is just wrong** (it is not “stationary”).
- For strongly seasonal items expressly this can lead to obscure index numbers in the transitive benchmark index and different estimation windows can lead to **hugely divergent time series**.

Index calculation: summary

- **Multilateral approaches:** At least three different methods, **numerically very close** to each other empirically; no single method emerges as “best”.
- **Estimation windows: At least 13 months**, maybe rather 25 months – strongly seasonal items in particular and differing trends in general issues.
- **Extension methods:** Growing literature and ever new approaches but all data dependent and not generalizable; **biggest source of differences**.

- **Best path forward for BLS** in using scanner / web-scraped data: Assess all possible **combinations of these three dimensions** against criteria relevant for BLS, and **shortlist what works best** in most, if not even all, cases.
- Criteria can include **institutional arrangements** such as the current approach to consumer prices, e.g. by looking at the way the overall index is chain-linked.
- To cut a long story short: **Test new methods sufficiently** side-by-side to regular production to see and solve problems **before anything is put into production**.

Quality adjustment: introduction

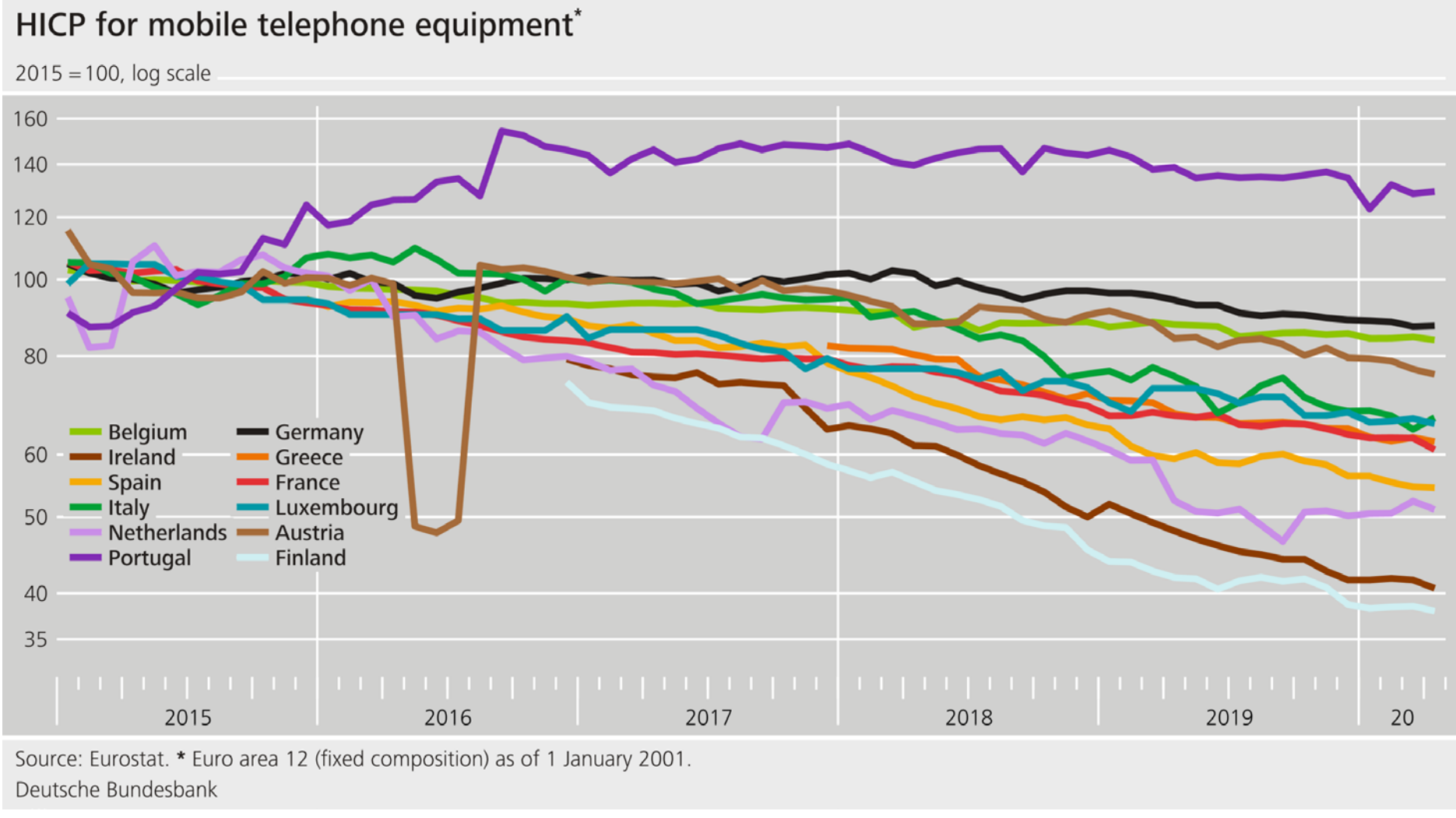
- **Multilateral methods** account implicitly for the relative differences in price levels of the products, only. However, for consumer electronics, say, hedonic **quality adjustment** might still be the best option.

- Very much like the TPD method, the **time-dummy hedonic (TDH) approach** is estimated based on pooled data of all periods:

$$\ln p_{i,t} = \alpha + \delta_t + \sum_{k=1}^{K} \beta_k z_{i,k} + \varepsilon_{i,t} \text{ , and } P_{0,t} = \exp \hat{\delta}_t.$$

- The difference is that the time dummies γ_i are replaced by **price-determining characteristics** $z_{i,k}$.
- **Explicitly quality-adjusted** log prices are $\ln p_{i,t} - \sum_{k=1}^{K} \hat{\beta}_k z_{i,k}$; compared to the implicit variant derived from the TPD method $\ln p_{i,t} - \hat{\gamma}_i$.
- **Quality adjustment is as much an art as a science!**

Quality adjustment: examples



Source: Eurostat. * Euro area 12 (fixed composition) as of 1 January 2001.
Deutsche Bundesbank

Quality adjustment: examples

- Have **prices for smart phones** in five years (from 2015 to 2020 August) risen by 31% (Portugal), or fallen by 62% (Finland)? If we assume today's average smart phone would have cost €1,000 in 2015, it would now cost €380 in Finland and €1,310 in Portugal; a trade margin of €930 – per smart phone!
- Whether higher prices reflect quality growth or inflation is a matter of **debate and personal perception**; the iPhone XR is currently priced at $599, the iPhone 11 (Pro / Pro Max) at $699 ($999 / $1,099), the new iPhone SE at $399.
- Using **matched models or implicit methods** that fail to adjust for the higher average quality of new varieties (comparable replacement, overlap pricing, etc.) may cause the index to **overstate the product's price change**.
- Not to be overlooked is whether particularly **hedonic models** may systematically **overstate quality change**. Examples of quality declines that are not captured in price measures include the requirement to purchase new models of mobile phones in the absence of backwards compatibility of new software (e.g. operating system) with older hardware. **Hedonics per se is no panacea!**

Postscript

- Indices from scanner and web-scraped data have shown to be **more volatile than traditional indices**. While the traditional price collection of matched models shows little to no noise in the price developments, the new methods introduce a lot of **noise in the time series**.
- Furthermore, scanner and web-scraped data represent an admittedly “big” but **biased non-probabilistic sample** – not the population. There are transactions that are in the scope but are not recorded electronically, not available to the statistical office, deleted in the filtering step, cannot be matched or linked, and so forth. After all, not more data are better, **better data are better**.
- Scanner and web-scraped data can be very precise but at the same time may have **limited accuracy**. The danger lies in blindly trusting that these new data sources must give us better answers; in fact, big data are **not capturing all transactions**, just some, and we might not even know which ones are missing. That is why the combination of more traditional data with big data is the ticket to reducing **coverage bias**.