Measuring price dynamics of package holidays with transaction data*

Elisabeth Wieland, Deutsche Bundesbank

* Joint work with Karola Henn¹, Chris-Gabriel Islam¹ and Patrick Schwind²

Ottawa Group, Rio de Janeiro, 10 May 2019

¹ Federal Statistical Office in Germany (Destatis). ² Deutsche Bundesbank. * This presentation represents the authors' personal opinions and does not necessarily reflect the views of Destatis, Deutsche Bundesbank or its staff. Corresponding author: elisabeth.wieland@bundesbank.de.
Outline

1. Motivation

2. Current official practice

3. Description of AMADEUS data set

4. Different methods of price measurement
   4.1 Unit value price index
   4.2 Hedonic regression methods
   4.3 Stratification methods

5. Comparison of results

6. Summary and outlook
1. Motivation
HICP sub-index on package holidays in Germany

- **Package holidays**, i.e. a combination of travel and accommodation services, are an important driver of the German inflation rate.

- **HICP sub-index** package holidays (ECOICOP 09.6) exhibits a **high volatility** (even higher than HICP Unprocessed food).

HICP package holidays compared with other components

<table>
<thead>
<tr>
<th>Year-on-year percentage change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Package holidays</strong> (HICP share 2019: 2.7%)</td>
</tr>
<tr>
<td><strong>Unprocessed food</strong> (2.8%)</td>
</tr>
<tr>
<td><strong>Clothing and footwear</strong> (5.4%)</td>
</tr>
<tr>
<td><strong>HICP excluding energy and food</strong> (72.9%)</td>
</tr>
</tbody>
</table>

Source: Eurostat.
Deutsche Bundesbank
1. Motivation
Research project on transaction-based price indices

- **Traditional price collection** uses offer prices from **pre-defined price representatives**.
  - Setting up a sufficiently large sample requires a lot of **manual work**, notably for seasonal and travel-related items.
  - Currently, the German HICP is only broken down by the two sub-indices “Domestic package holidays” (ECOICOP 09.6.0.1) and “International package holidays” (09.6.0.2).

- **Joint research project by Destatis and Bundesbank**
  - Chances and challenges in deriving a **transaction-based price index** for package holidays with actual bookings for the German travel market.
  - High number of observations enables a further **disaggregation by holiday destination**.
  - Compare several aggregation methods such as hedonic regression and stratification, with applications and challenges **similar to supermarket scanner data**.
Several challenges in measuring prices of package holidays arise:

- high price volatility within a specific travel month,
- the date of booking is a significant price determinant, and
- strong seasonality of bookable offers.

Destatis compiles offer prices based on pre-defined booking codes.

Main elementary aggregate „International flight package holidays“ covers approx. 300 price representatives (multiple offer prices by travel month and book time).
3. Description of AMADEUS data set
Transaction data on the German market

- **Commercial data set** by AMADEUS/TravelTainment, an IT provider for touristic services.
- Roughly **3.7 million transactions per year** on a daily basis from two separate booking channels (online and offline).
- For every transaction, information on the booking process and **important price determinants** are available:
  - Booking and travel **date**; duration
  - Departure and destination **airport**
  - Number and age of **travellers**
  - Location, name and star rating of the **hotel**
  - Online data only: **meal type** and **room category**
Methodological challenges:
1. Relevance and incomplete or incorrect information
2. Uncategorised (text) information
3. Varying sample composition

Prepare data set:
1. **Data cleansing:** Remove outliers in *duration* and *totalPrice*, *transactionDate < travelDate*, departure from a German airport, …
2. **Categorise** text information (e.g. *mealType* and *roomCategory*).
3. **Control** for or assign **weights** (e.g. according to revenue share).
3. Description of AMADEUS data set
Total revenue mostly covered by six holiday regions

Revenue share of package holidays by holiday region in 2015*

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH. * Without cruises. ES Spain, TR Turkey, EG Egypt, GR Greece, DO Dominican Republic, AE United Arab Emirates, MV Maldives, BG Bulgaria, PT Portugal, TN Tunisia, CU Cuba, IT Italy, TH Thailand, MX Mexico. 1 Holiday regions with a transaction weight of less than 1%.

Deutsche Bundesbank
4. Different methods of price measurement
Natural starting point: Compare prices of exactly the same item over time

---

*Product continuity refers to the degree of product match in terms of observations, when comparing a given month with a base period. Product homogeneity refers to the degree of similarity of items within a given product.

Deutsche Bundesbank

---

Measuring price dynamics of package holidays with transaction data
Henn/Islam/Schwind/Wieland
p. 9
4. Different methods of price measurement
Index class I: Unit Value Price Index

• Per holiday region: Average Price per Person per Day (PPD):

\[ PPD_t = \frac{1}{N_t} \cdot \sum_{i=1}^{N_t} \frac{totalPrice_{i,t}}{travellerCount_{i,t} \cdot duration_{i,t}} \]

• Problem: Requires identical, homogeneous items
→ Otherwise, Selection Bias due to compositional changes in the underlying bookings

• Example: Bookings for Hotel A and B in Turkey

<table>
<thead>
<tr>
<th>Period</th>
<th>Price A</th>
<th>Price B</th>
<th>Quantity A</th>
<th>Quantity B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100 €</td>
<td>120 €</td>
<td>700</td>
<td>700</td>
</tr>
<tr>
<td>1</td>
<td>100 €</td>
<td>120 €</td>
<td>800</td>
<td>400</td>
</tr>
</tbody>
</table>

→ No overall price change at all, but Unit Value Price Index would fall by 3 %.
4. Different methods of price measurement
Index class II: Hedonic regressions

- **Assumption:** The price of a given good or service is a function of several (observed) attributes, each having a marginal contribution to the overall price.

- Baseline: For each holiday region, \( \log_{\text{totalPrice}} \) is regressed on...
  - number of passengers (log)
  - duration (log)
  - booking days before departure (log)
  - online (dummy)
  - accommodation category (1 to 5 star) (dummy)
  - Double Imputation: dummy on Easter, Pentecost or Christmas during travel

<table>
<thead>
<tr>
<th>Model</th>
<th>Double Imputation</th>
<th>Time Dummy Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base period 0</td>
<td>Year 2015</td>
<td>January of a year</td>
</tr>
<tr>
<td>Comparison period</td>
<td>Current month ( t )</td>
<td>Current month ( t )</td>
</tr>
<tr>
<td>Regression techniques</td>
<td>OLS (separate for period 0 and ( t ))</td>
<td>OLS (pooling period 0 and ( t ))</td>
</tr>
<tr>
<td>Price index derived from...</td>
<td>Estimating <em>price relatives</em> for each transaction ( i ) in period ( t ): ( \hat{P}_i^t \rightarrow \hat{\beta}^t \cdot X_i^t ) and ( \hat{P}_i^0 \rightarrow \hat{\beta}^0 \cdot X_i^t )</td>
<td>Exponential of the <em>coefficient</em> on time dummy for current month ( t )</td>
</tr>
</tbody>
</table>
### 4. Different methods of price measurement

**Time Dummy Model**

**Example:**
Balearic Islands, July 2018

![Table with statistical data](image)

- **Intercept**
  - DF: 1
  - Parameter-Schätzer: 4.84855
  - Standard Fehler: 0.00820
  - t-Wert: 591.11
  - Pr > |t|: <.0001
- **d_time_2018_7**
  - DF: 1
  - Parameter-Schätzer: 0.54055
  - Standard Fehler: 0.00522
  - t-Wert: 103.57
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.07329
  - Quadrat Semi-partiell Korr. Typ II: 0.02940
  - Varianz Inflation: 1.03466
- **log_travellerCount**
  - DF: 1
  - Parameter-Schätzer: 0.83859
  - Standard Fehler: 0.00237
  - t-Wert: 353.20
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.45782
  - Quadrat Semi-partiell Korr. Typ II: 0.34197
  - Varianz Inflation: 1.09778
- **log_duration**
  - DF: 1
  - Parameter-Schätzer: 0.65769
  - Standard Fehler: 0.00269
  - t-Wert: 244.29
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.21690
  - Quadrat Semi-partiell Korr. Typ II: 0.16360
  - Varianz Inflation: 1.08215
- **online**
  - DF: 1
  - Parameter-Schätzer: -0.08763
  - Standard Fehler: 0.00191
  - t-Wert: -45.96
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.01195
  - Quadrat Semi-partiell Korr. Typ II: 0.00579
  - Varianz Inflation: 1.06907
- **log_BookTime**
  - DF: 1
  - Parameter-Schätzer: 0.05663
  - Standard Fehler: 0.00117
  - t-Wert: 48.58
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.00423
  - Quadrat Semi-partiell Korr. Typ II: 0.00647
  - Varianz Inflation: 1.15411
- **D_star1**
  - DF: 1
  - Parameter-Schätzer: -0.59621
  - Standard Fehler: 0.00913
  - t-Wert: -65.31
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.00772
  - Quadrat Semi-partiell Korr. Typ II: 0.01169
  - Varianz Inflation: 1.01706
- **D_star2**
  - DF: 1
  - Parameter-Schätzer: -0.41474
  - Standard Fehler: 0.00424
  - t-Wert: -97.92
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.01633
  - Quadrat Semi-partiell Korr. Typ II: 0.02628
  - Varianz Inflation: 1.05338
- **D_star3**
  - DF: 1
  - Parameter-Schätzer: -0.21830
  - Standard Fehler: 0.00189
  - t-Wert: -115.38
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.03851
  - Quadrat Semi-partiell Korr. Typ II: 0.03650
  - Varianz Inflation: 1.07519
- **D_star5**
  - DF: 1
  - Parameter-Schätzer: 0.18069
  - Standard Fehler: 0.00760
  - t-Wert: 23.79
  - Pr > |t|: <.0001
  - Quadrat Semi-partiell Korr. Typ I: 0.00155
  - Quadrat Semi-partiell Korr. Typ II: 0.00155
  - Varianz Inflation: 1.01635
4. Different methods of price measurement
Index class III: Stratification methods

- **Assumption**: Similar items have a similar price. 
  → Divide the sample into **homogeneous strata** and compute the **average price per stratum**.

- **Product definition**: Results from hedonic regression and by a quantitative approach following Chessa (2018), the *Match Adjusted R Squared* (MARS), to find a balance between product continuity and product homogeneity.

- For each holiday region, PPD is stratified according to
  - booking days before departure: [15 – 30; 31 – 90; 91 – 180; > 180]
  - channel: online or offline
  - accommodation category: 3- to 5-star hotels

<table>
<thead>
<tr>
<th>Model</th>
<th>Traditional Stratification</th>
<th>GEKS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparison</strong></td>
<td>Bilateral price relatives</td>
<td>Multilateral price relatives (rolling window of 13 months)</td>
</tr>
<tr>
<td><strong>Number of strata</strong></td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td><strong>Aggregation and weighting</strong></td>
<td>Weighted arithmetic mean of PPD per stratum; based on the <em>average 2015 revenue share of the individual stratum</em></td>
<td>Product of Fisher index, weighted each month by the <em>current revenue share of the individual stratum</em></td>
</tr>
</tbody>
</table>
5. Comparison of results

Evaluation strategy

- Robustness
  - Different data filters (only online data, incl. last minute bookings, no child travellers, incl. non-German departure airports…)
  - Different specifications of regression models

- Distribution of annual rates

- Uniform direction of growth rates?

- Volatility

- Comparison with current method

- Comparison of growth rates using **same weighting scheme** as well as the official sub-indices „cruises“ and „city trips“

Measuring price dynamics of package holidays with transaction data
Henn/Islam/Schwind/Wieland
p. 14
5. Comparison of results
Results for two major holiday regions

Transaction-based price index of package holidays to Turkey

2015 = 100, log scale

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.
Deutsche Bundesbank

Transaction-based price index of package holidays to the Balearic Islands

2015 = 100, log scale

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.
Deutsche Bundesbank

Measuring price dynamics of package holidays with transaction data
Henn/Islam/Schwind/Wieland
p. 15
5. Comparison of results
YoY-rates: Double Imputation (DI) and Traditional Stratification (TS) less volatile
5. Comparison of results
Transaction-based indices comove strongly

Comparison of transaction-based pseudo indices with HICP sub-index “International Package Holidays” (ECOICOP 09.6.0.2)

Source: Destatis; Bundesbank calculations on the basis of booking data from TravelTainment GmbH. For each transaction-based method, the elementary indices on six holiday regions were aggregated together with the official (confidential) elementary indices “city trips” and “cruises” by using the official fixed-weighting scheme.
Deutsche Bundesbank
5. Comparison of results
Disaggregation allows for a detailed economic interpretation

Experimental index for international package holidays and contributions from holiday regions

- Contribution by holiday region (in percentage points)
  - Balearic Islands
  - Egypt
  - Canary Islands
  - Dominican Republic
  - Turkey
  - Greece

Experimental index on int. package holidays (percentage change against previous year's month)

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH. The experimental index is based on the “double imputation” method and consists of six holiday regions, which are weighted together by their respective revenue share (2015-2016 average).
Deutsche Bundesbank
6. Summary and Outlook

• The AMADEUS data set allows several approaches to derive a transaction-based price index for package holidays, with a sample size large enough to provide a disaggregation by major holiday regions.

• All methods under consideration provide similar results in terms of price dynamics.
  – Same seasonal pattern.
  – Sign of y-o-y (m-o-m) growth rate diverges in only 4 (0) periods out of 36 (47) periods.

• Transaction data provide weight information (e.g. by revenue share) at a very detailed level.
  → Statistical challenge: Set up an appropriate price index.

• Open question: Do currently applied transaction-based methods perform sufficiently well in terms of mixing and quality adjustment (notably regarding room type)?
  → Offer prices: aim to ensure pure price comparison.
  → Further research at the micro level on booking characteristics needed.


Measuring price dynamics of package holidays with transaction data
Henn/Islam/Schwind/Wieland
p. 20
Appendix: Results for individual holiday regions

Transaction-based price index of package holidays to the Canary Islands

2015 = 100, log scale

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.
Deutsche Bundesbank

Transaction-based price index of package holidays to Greece

2015 = 100, log scale

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.
Deutsche Bundesbank
Appendix: Results for individual holiday regions

Transaction-based price index of package holidays to Egypt
2015 = 100, log scale

Transaction-based price index of package holidays to the Dominican Republic
2015 = 100, log scale

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.
Deutsche Bundesbank
24 Apr 2019, 07:52:30, S3PR0472.Chart
Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.
Deutsche Bundesbank
24 Apr 2019, 08:45:58, S3PR0473.Chart
Appendix: Revenue shares over time

Transaction weights of selected holiday regions*
As a percentage of total transaction volume

Source: Bundesbank calculations based on booking data by TravelTainment. * Without cruises.
Deutsche Bundesbank
Appendix: Booking time as an important price determinant

Bookings and average price by months before departure*

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH. * Without cruises. 1 € per person and day.
Appendix: Product definition in the context of package holidays

Continuity and homogeneity of several product definitions following Chessa (2018)

Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH; average values of 2015. The product definition highlighted in red was selected for the subsequent analysis. For illustration purpose, one outlier combination with homogeneity = 1 (i.e., every package holiday represents an own product class) is excluded.

Deutsche Bundesbank