

# Measuring price dynamics of package holidays with transaction data\*

Elisabeth Wieland, Deutsche Bundesbank

\* Joint work with Karola Henn<sup>1</sup>, Chris-Gabriel Islam<sup>1</sup> and Patrick Schwind<sup>2</sup>

Ottawa Group, Rio de Janeiro, 10 May 2019

<sup>1</sup> Federal Statistical Office in Germany (Destatis). <sup>2</sup> 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](mailto:elisabeth.wieland@bundesbank.de).

# “Measuring price dynamics of package holidays with transaction data”

## Outline

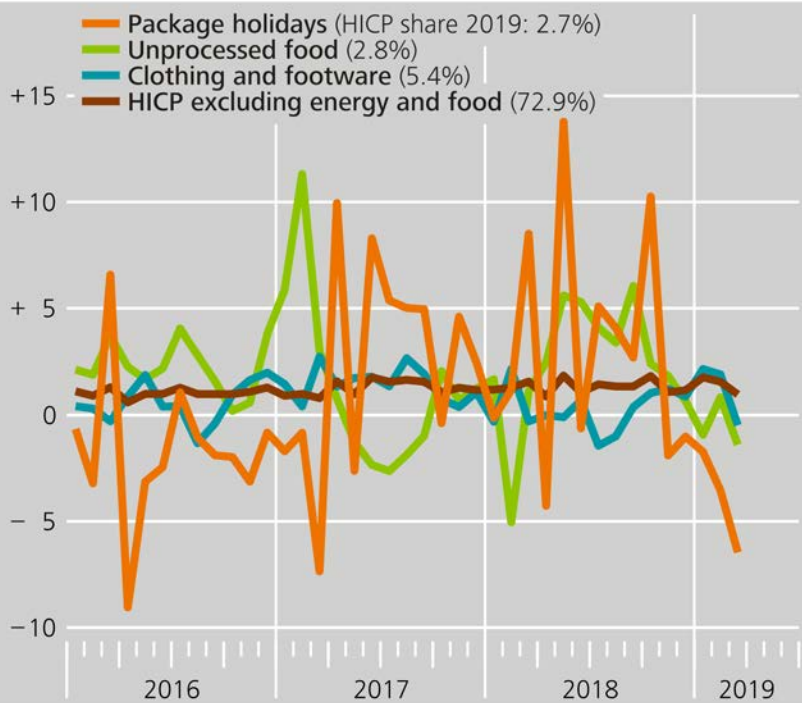
1. **Motivation**
2. **Current official practice**
3. **Description of AMADEUS data set**
4. **Different methods of price measurement**
  - 4.1 **Unit value price index**
  - 4.2 **Hedonic regression methods**
  - 4.3 **Stratification methods**
5. **Comparison of results**
6. **Summary and outlook**

# 1. Motivation

## HICP sub-index on package holidays in Germany

### HICP package holidays compared with other components

Year-on-year percentage change



Source: Eurostat.

Deutsche Bundesbank

S3PR0460.Chart

- **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).

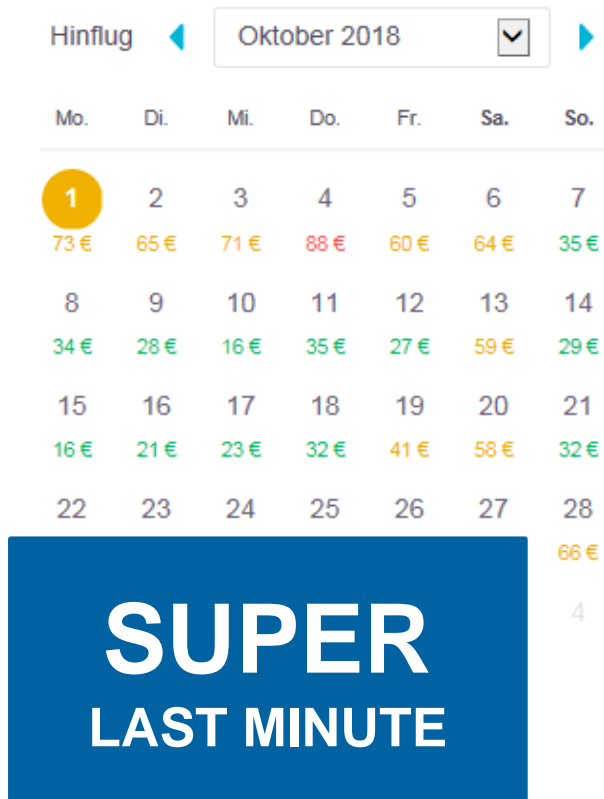
# 1. Motivation

## 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**.

## 2. Current official practice

### Sampling offer prices



- 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

### 3. Description of AMADEUS data set

#### How to deal with big data characteristics

#### Methodological challenges:

1. Relevance and incomplete or incorrect information
2. Uncategorised (text) information
3. Varying sample composition

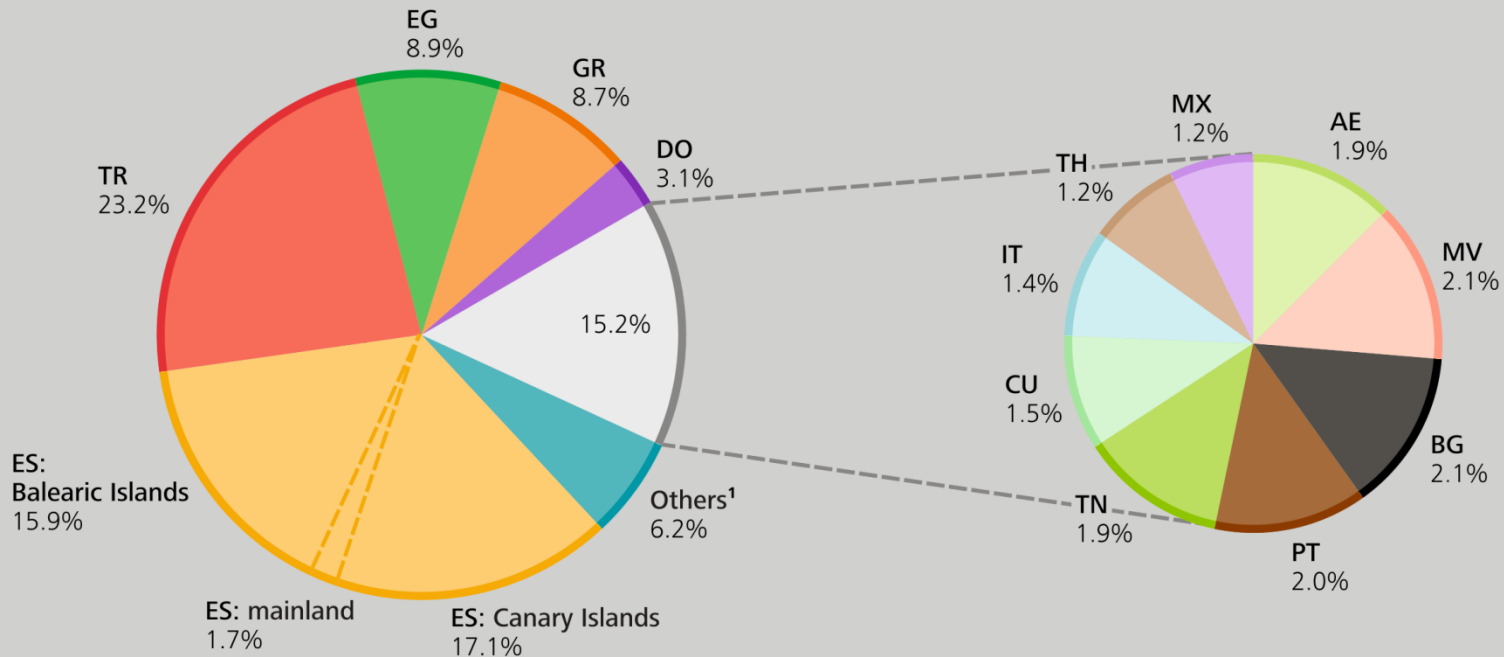
#### Prepare data set:

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. <sup>1</sup> Holiday regions with a transaction weight of less than 1%.

Deutsche Bundesbank

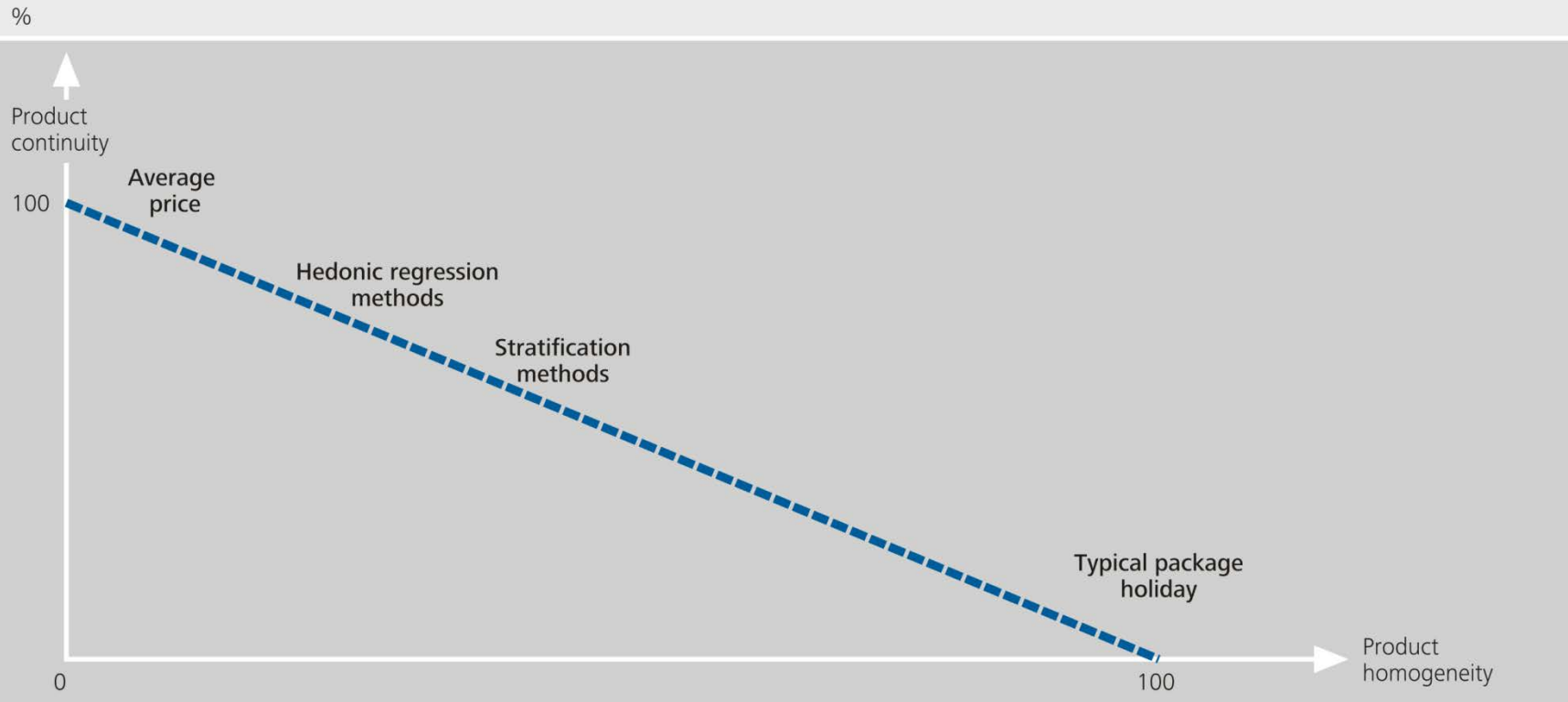
S3PR0464.Chart



## 4. Different methods of price measurement

Natural starting point: Compare prices of exactly the same item over time

Trade off between product continuity and product homogeneity\*



\* 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

S3PR0462.Chart

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):

$$\overline{PPD}_t = \frac{1}{N_t} \cdot \sum_{i=1}^{N_t} \frac{\text{totalPrice}_{i,t}}{\text{travellerCount}_{i,t} \cdot \text{duration}_{i,t}}$$

- 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

Period	Price A	Price B	Quantity A	Quantity B
0	100 €	120 €	700	700
1	100 €	120 €	<b>800</b>	<b>400</b>

→ 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_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

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

## 4. Different methods of price measurement

### Time Dummy Model

The REG Procedure  
Model: Linear\_Regression\_Model  
Dependent Variable: log\_totalPrice

#### Example:

Balearic Islands, July 2018

Number of Observations Read	64807
Number of Observations Used	62644
Number of Observations with Missing Values	2163

Variable	DF	Parameter-Schätzer	Standard Fehler	t-Wert	Pr >  t	Quadrat Semi-partiell Korr. Typ I	Quadrat Semi-partiell Korr. Typ II	Varianz Inflation
Intercept	1	4.84855	0.00820	591.11	<.0001	.	.	0
d_time_2018_7	1	0.54055	0.00522	103.57	<.0001	0.07329	0.02940	1.03466
log_travellerCount	1	0.83859	0.00237	353.20	<.0001	0.45782	0.34197	1.09778
log_duration	1	0.65769	0.00269	244.29	<.0001	0.21690	0.16360	1.08215
online	1	-0.08763	0.00191	-45.96	<.0001	0.01195	0.00579	1.06907
log_BookTime	1	0.05663	0.00117	48.58	<.0001	0.00423	0.00647	1.15411
D_star1	1	-0.59621	0.00913	-65.31	<.0001	0.00772	0.01169	1.01706
D_star2	1	-0.41474	0.00424	-97.92	<.0001	0.01633	0.02628	1.05338
D_star3	1	-0.21830	0.00189	-115.38	<.0001	0.03851	0.03650	1.07519
D_star5	1	0.18069	0.00760	23.79	<.0001	0.00155	0.00155	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

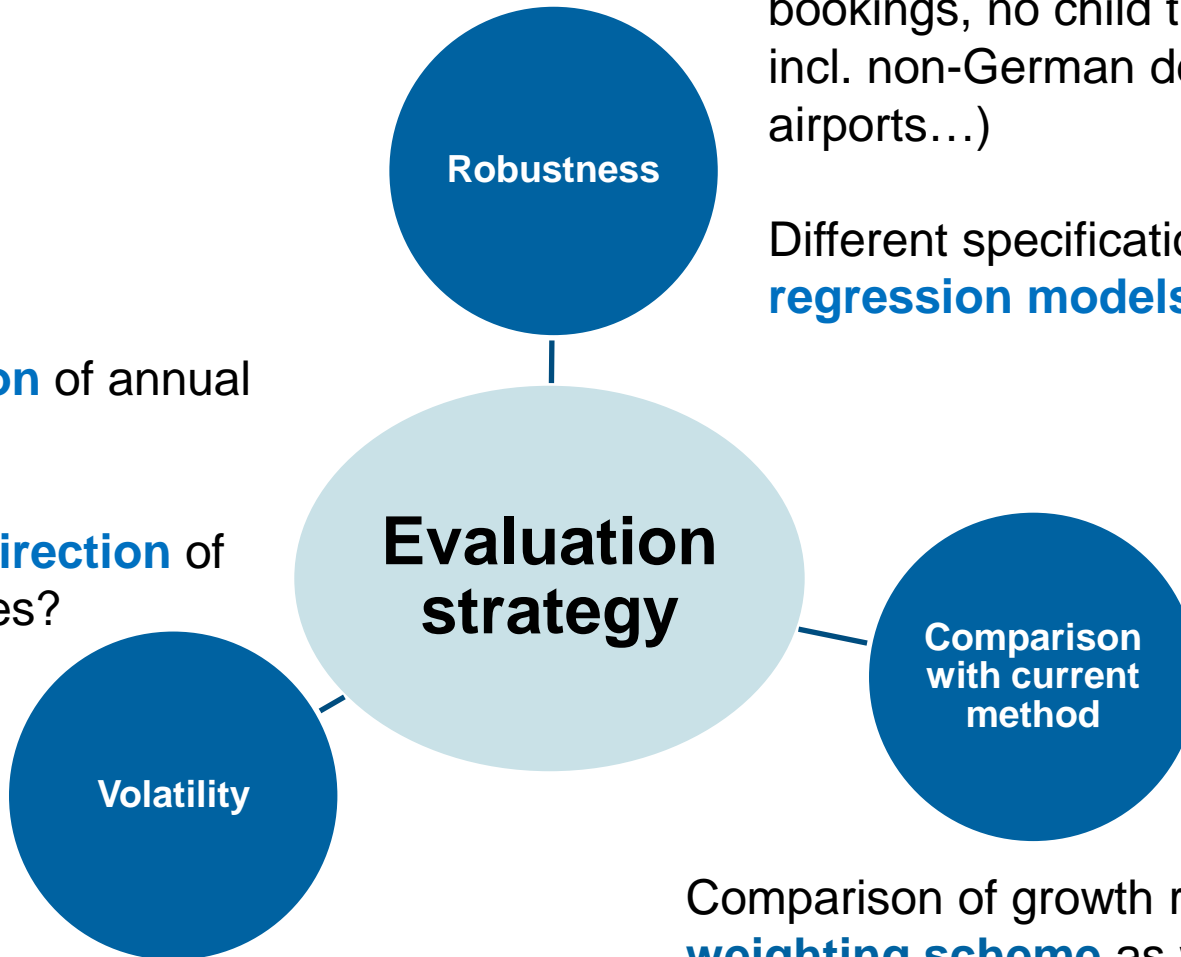
Model	Traditional Stratification	GEKS
Comparison	Bilateral price relatives	Multilateral price relatives (rolling window of 13 months)
Number of strata	24	24
Aggregation and weighting	Weighted arithmetic mean of PPD per stratum; based on the <i>average 2015</i> revenue share of the individual stratum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

## 5. Comparison of results

### Evaluation strategy

**Distribution** of annual rates

**Uniform direction** of growth rates?



**Different data filters** (only online data, incl. last minute bookings, no child travellers, incl. non-German departure airports...)

Different specifications of **regression models**

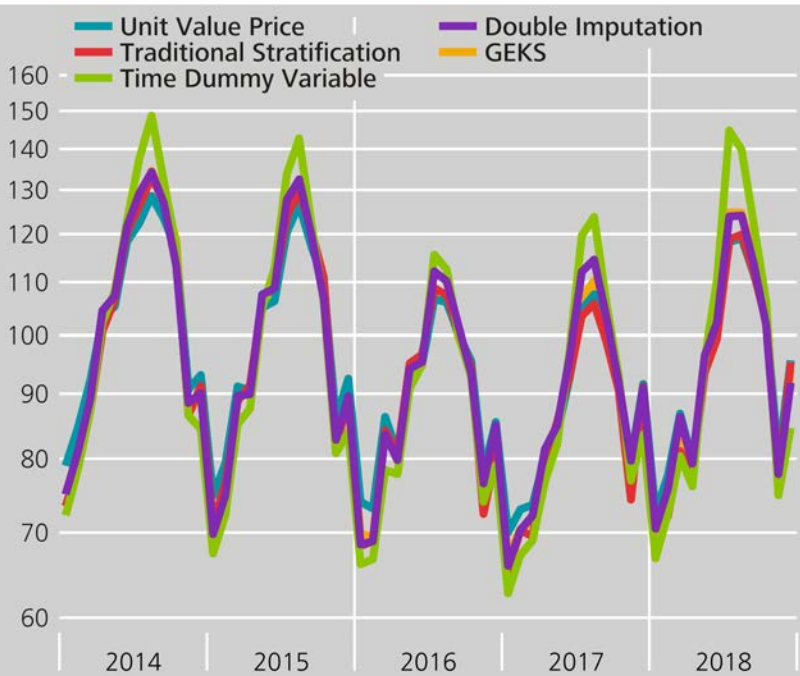
Comparison of growth rates using **same weighting scheme** as well as the official sub-indices „cruises“ and „city trips“

# 5. Comparison of results

## Results for two major holiday regions

### Transaction-based price index of package holidays to Turkey

2015 = 100, log scale



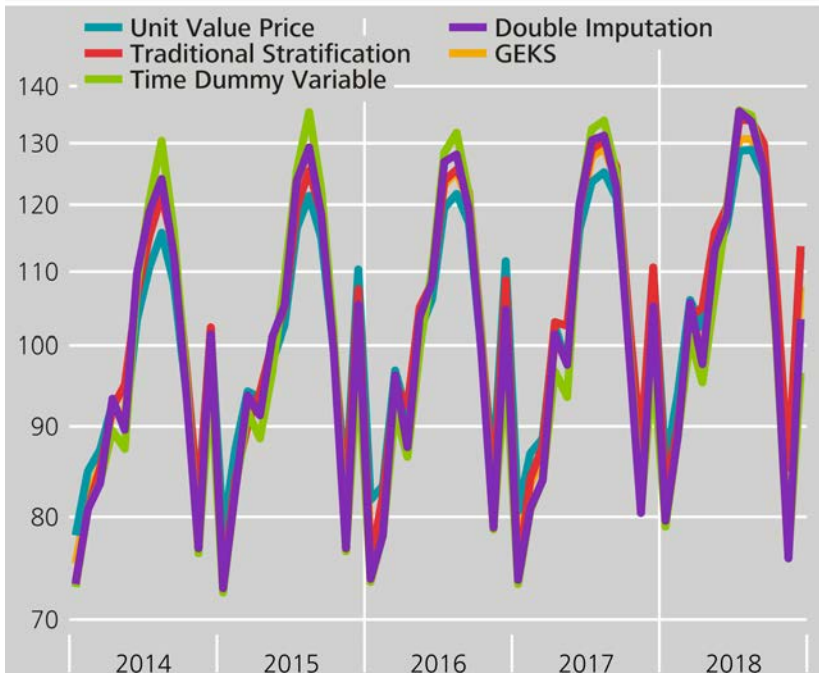
Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.

Deutsche Bundesbank

24 Apr 2019, 08:43:57, S3PR0471.Chart

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

2015 = 100, log scale



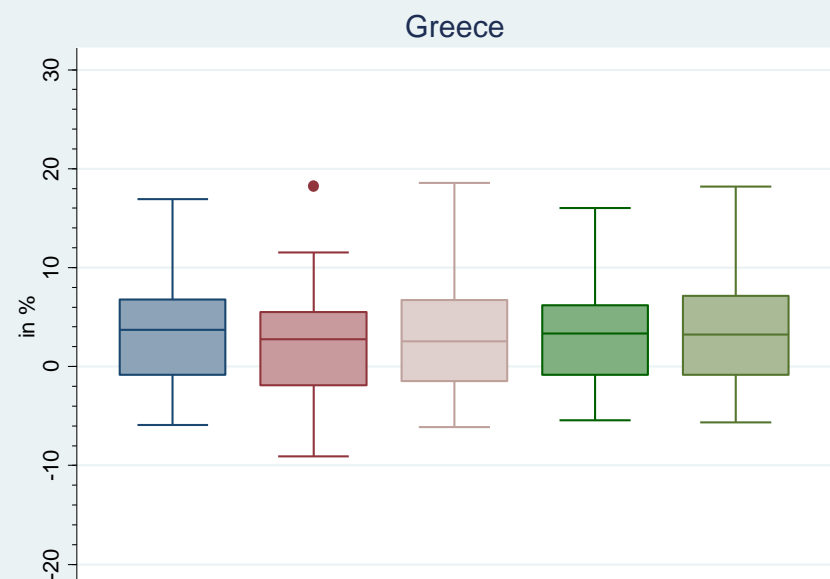
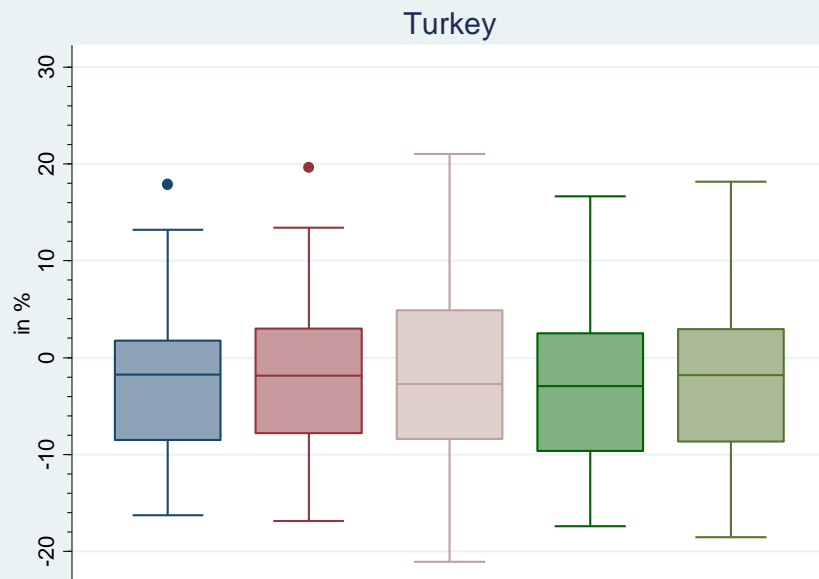
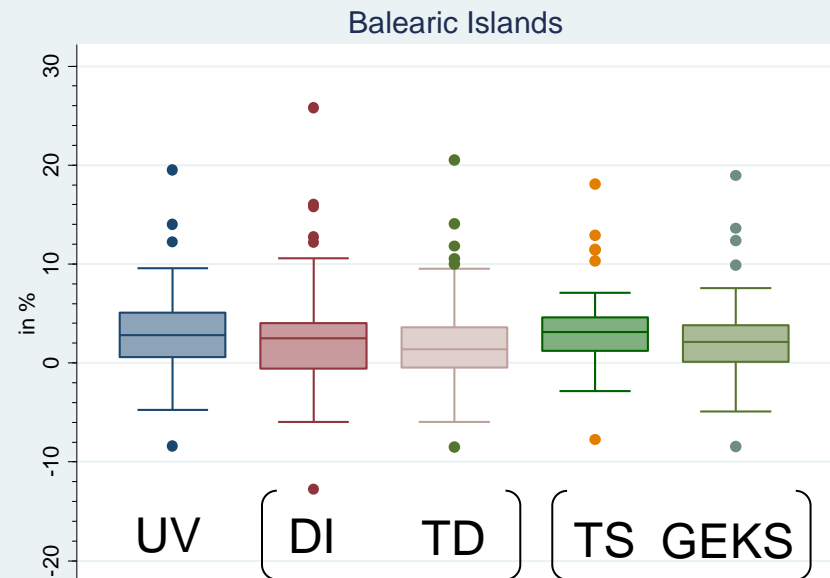
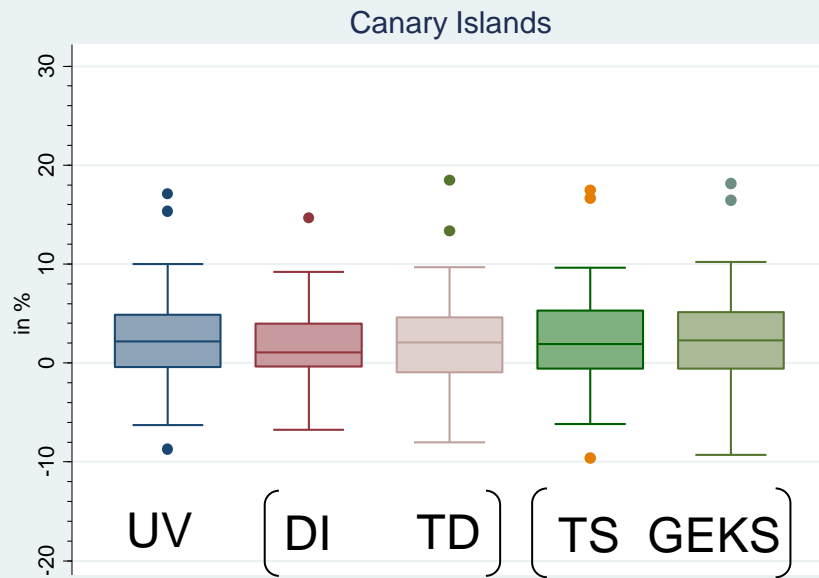
Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.

Deutsche Bundesbank

24 Apr 2019, 07:41:16, S3PR0469.Chart

## 5. Comparison of results

YoY-rates: Double Imputation (DI) and Traditional Stratification (TS) less volatile

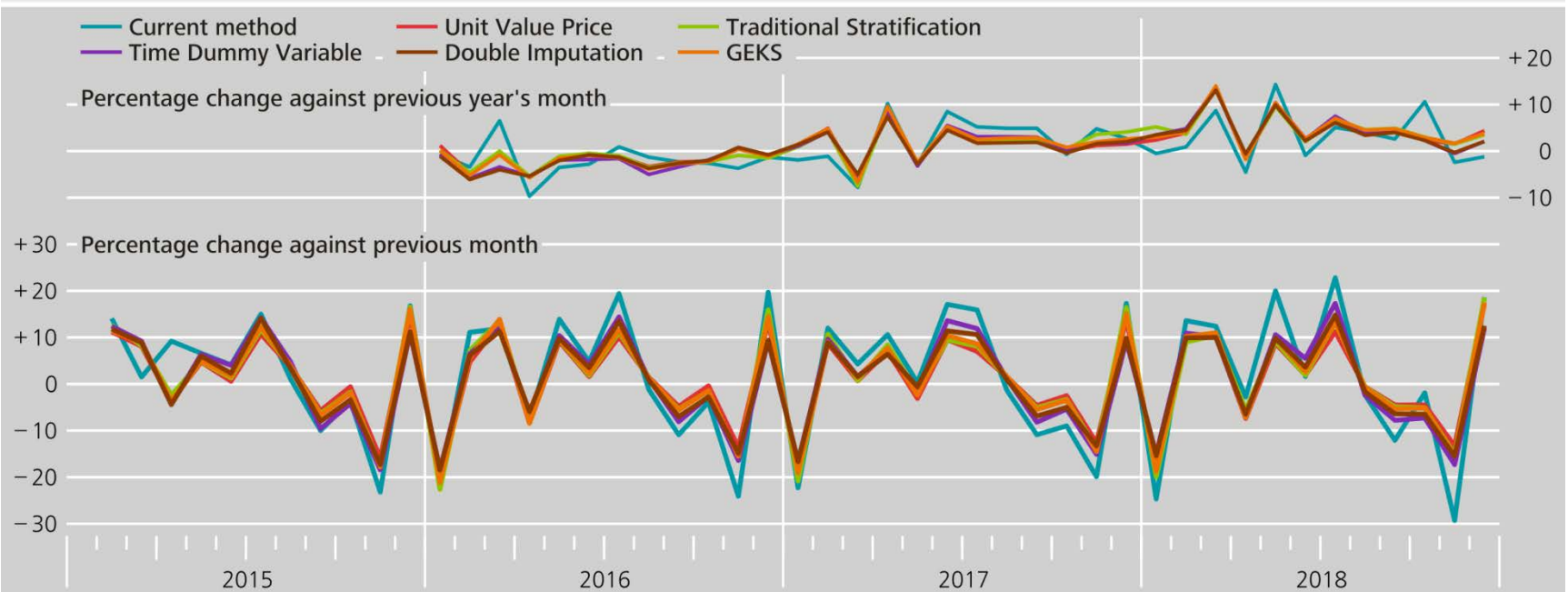




## 5. Comparison of results

### 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.

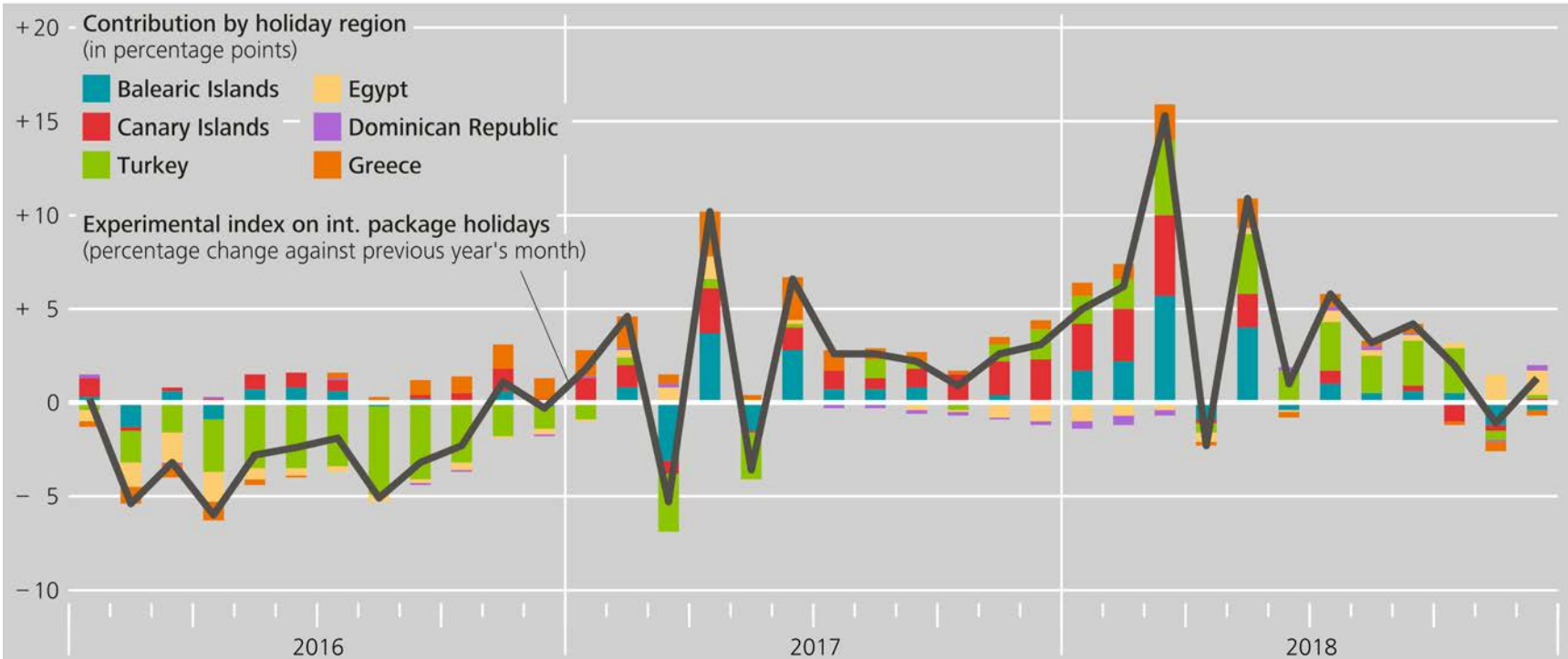
Deutsche Bundesbank

6 Mai 2019, 15:29:13, S3PR0477.Chart

## 5. Comparison of results

### Disaggregation allows for a detailed economic interpretation

#### Experimental index for international package holidays and contributions from holiday regions



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

17 Apr 2019, 14:18:55, S3PR0463.Chart

## 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.

## Selected References

Chessa, A.G. (2018), “Product definition and index calculation with MARS-QU: Applications to consumer electronics”, mimeo.

Destatis (2009): *Handbook on the application of quality adjustment methods in the Harmonised Index of Consumer Prices*, Statistics and Science Volume 13, Federal Statistical Office of Germany.

Deutsche Bundesbank (2017), “The volatility of the traditional core inflation rate in Germany”, in: Monthly Report, November 2017, pp. 49-51.

Eurostat (2018), *HICP Methodological Manual*, Luxembourg: Publications Office of the European Union, online at:

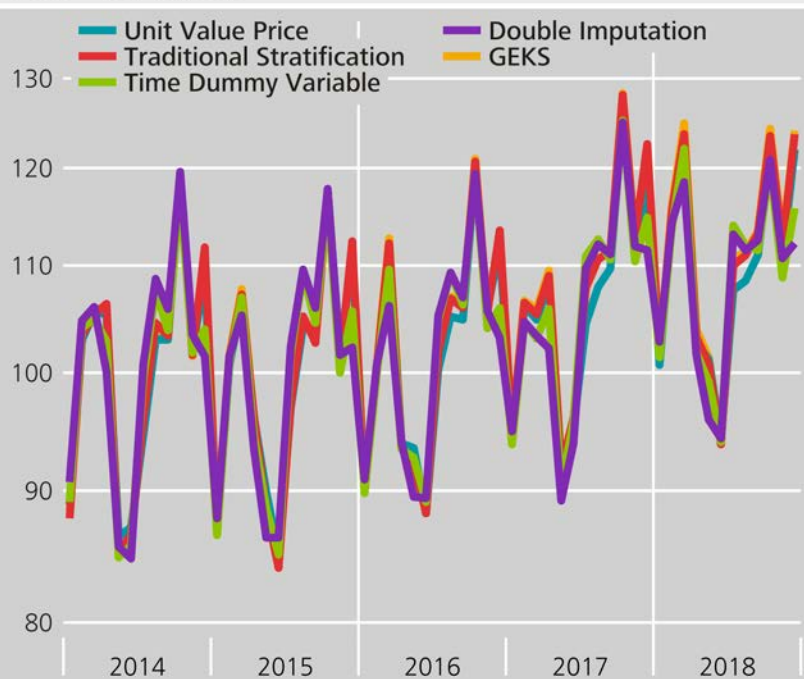
<https://ec.europa.eu/eurostat/documents/3859598/9479325/KS-GQ-17-015-EN-N.pdf>.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), “Scanner Data, Time Aggregation and the Construction of Price Indexes”, in: *Journal of Econometrics* 161, 24-35.

# Appendix: Results for individual holiday regions

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

2015 = 100, log scale



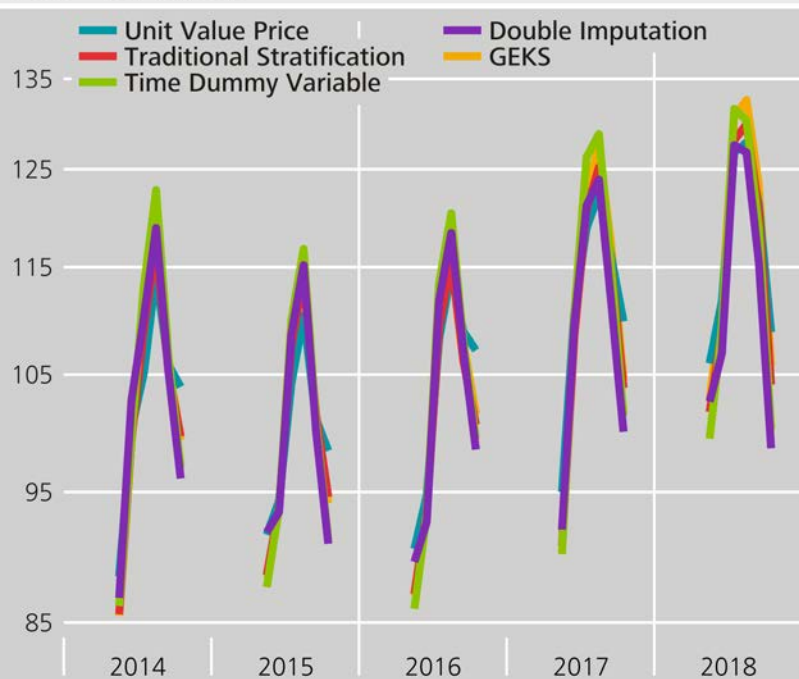
Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.

Deutsche Bundesbank

24 Apr 2019, 07:43:45, S3PR0470.Chart

## 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.

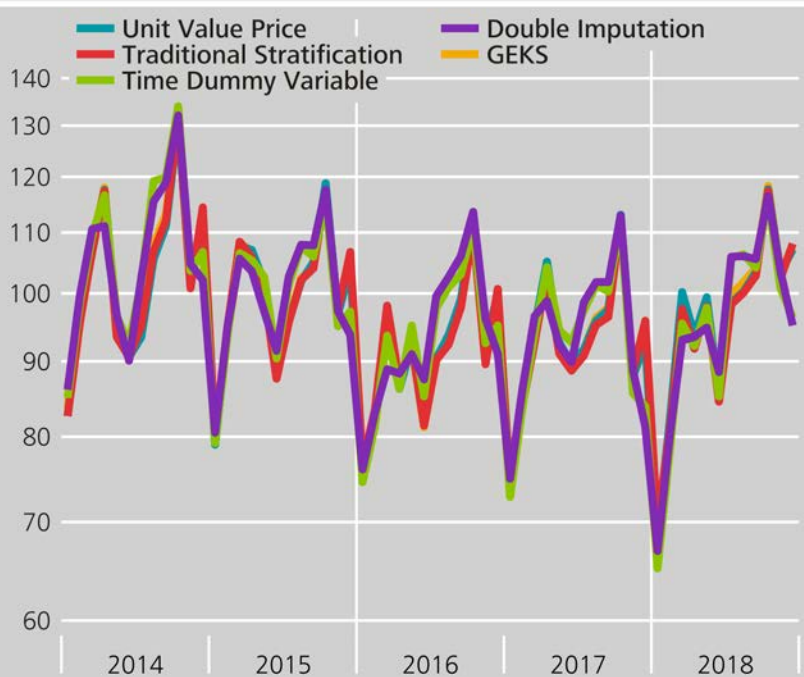
Deutsche Bundesbank

24 Apr 2019, 08:39:47, S3PR0474.Chart

# Appendix: Results for individual holiday regions

## Transaction-based price index of package holidays to Egypt

2015 = 100, log scale



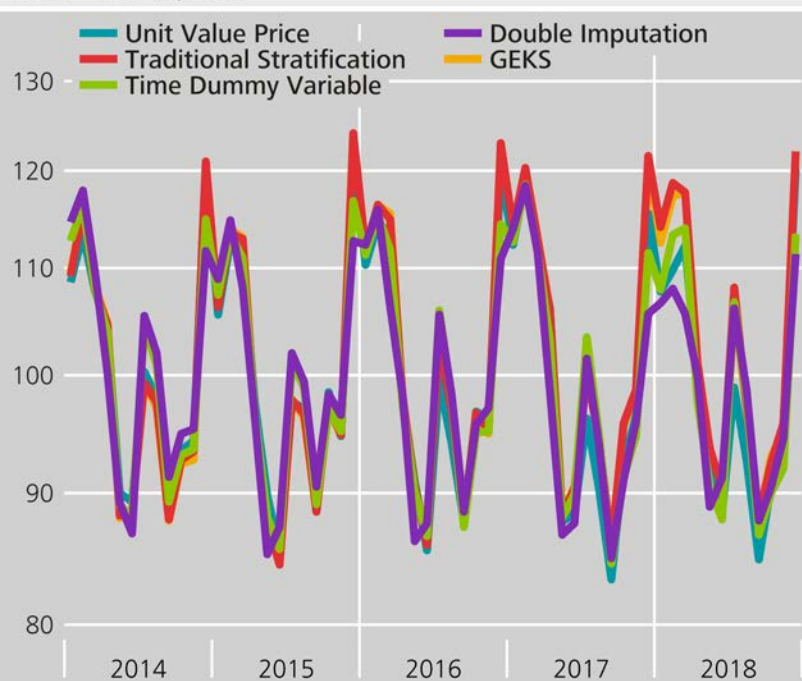
Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH.

Deutsche Bundesbank

24 Apr 2019, 07:52:30, S3PR0472.Chart

## 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.

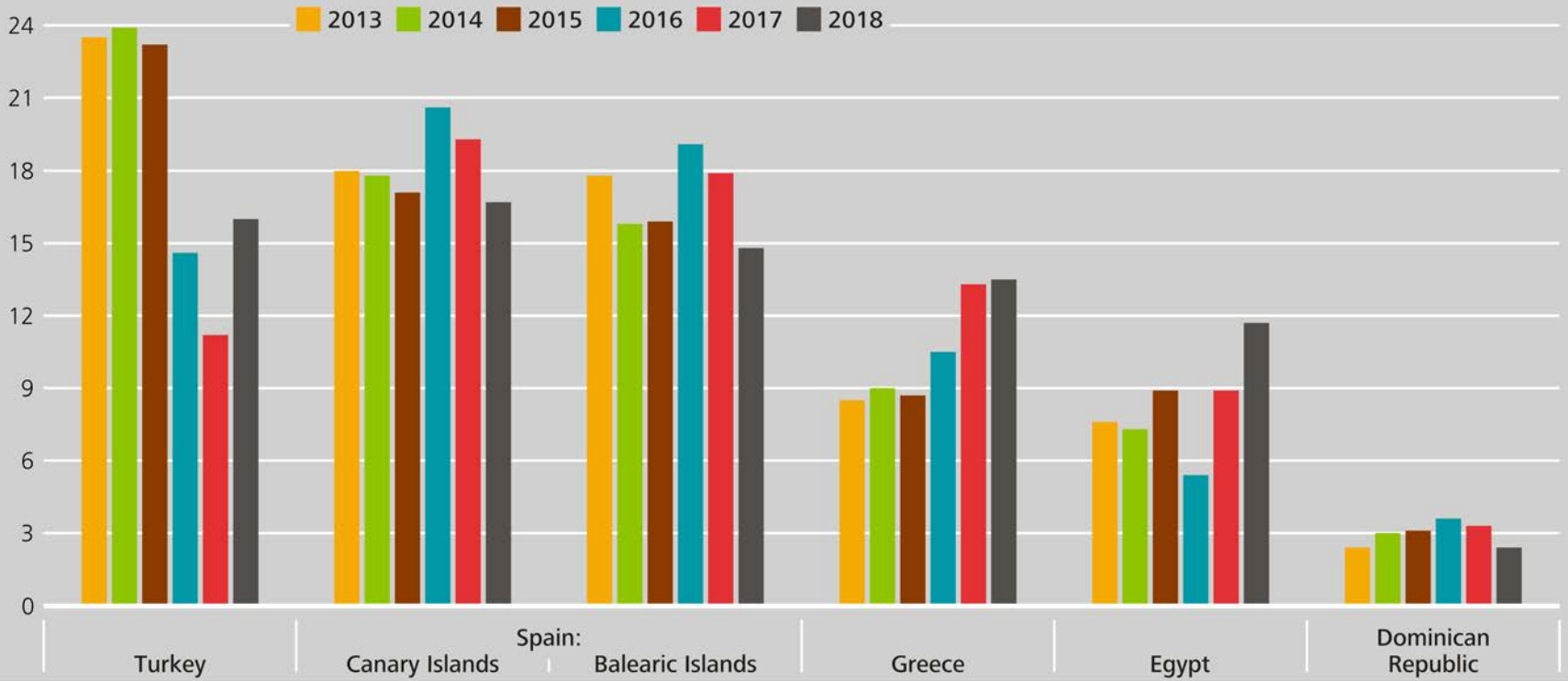
Deutsche Bundesbank

24 Apr 2019, 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.

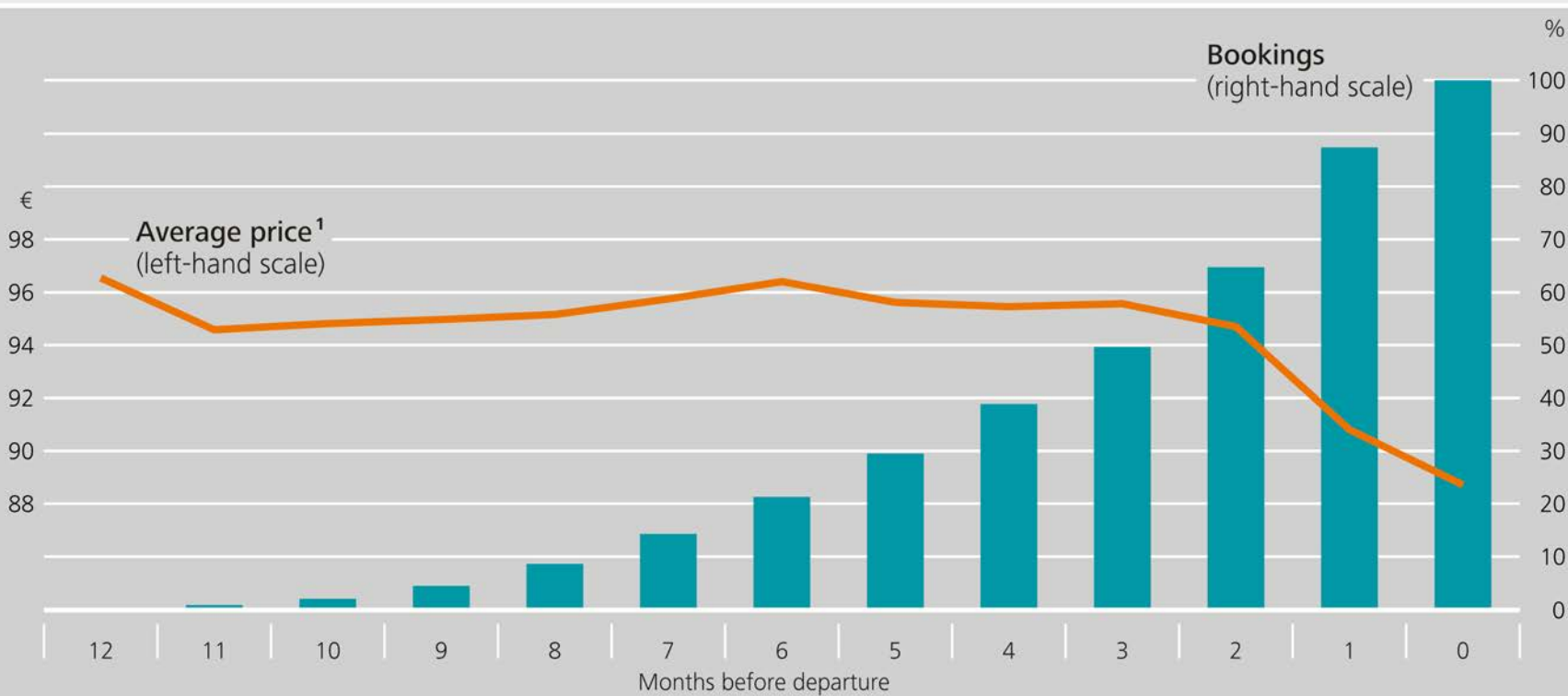
Deutsche Bundesbank

S3PR0466.Chart

# Appendix: Booking time as an important price determinant

## Bookings and average price by months before departure\*

2013 – 2018



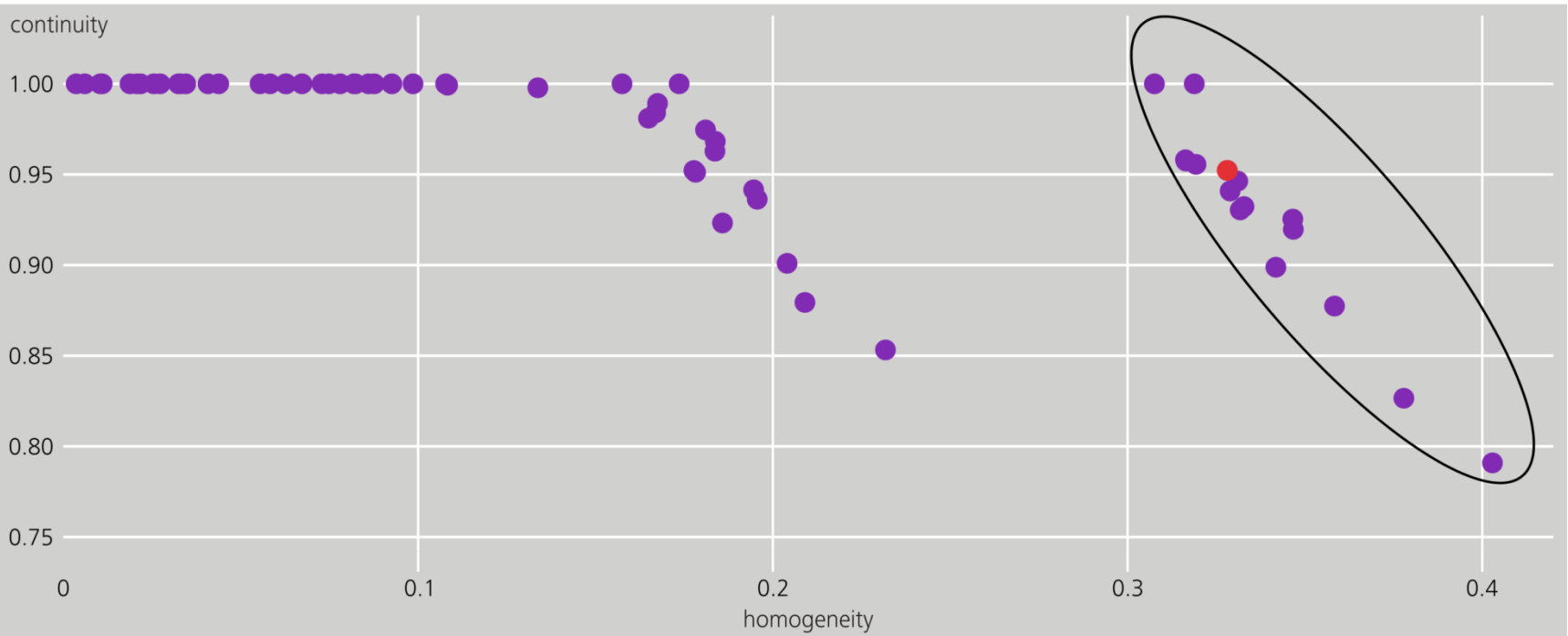
Source: Bundesbank calculations on the basis of booking data from TravelTainment GmbH. \* Without cruises. 1 € per person and day.  
Deutsche Bundesbank

S3PR0465.Chart



# 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.