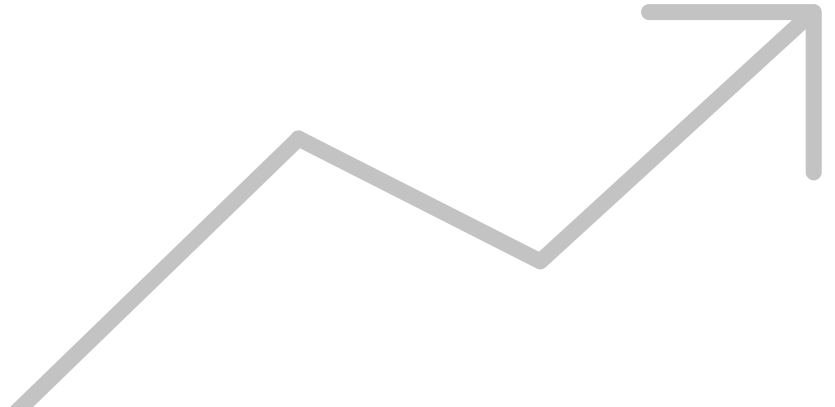
Transaction Data for the Price Index of Package Holidays in Germany

Amelie Blasius

Federal Statistical Office, Germany



Agenda

1) The German HICP

2) Data

3) Projects on German package holidays

4) Methodology: hedonic double imputation

5) Plausibility checks

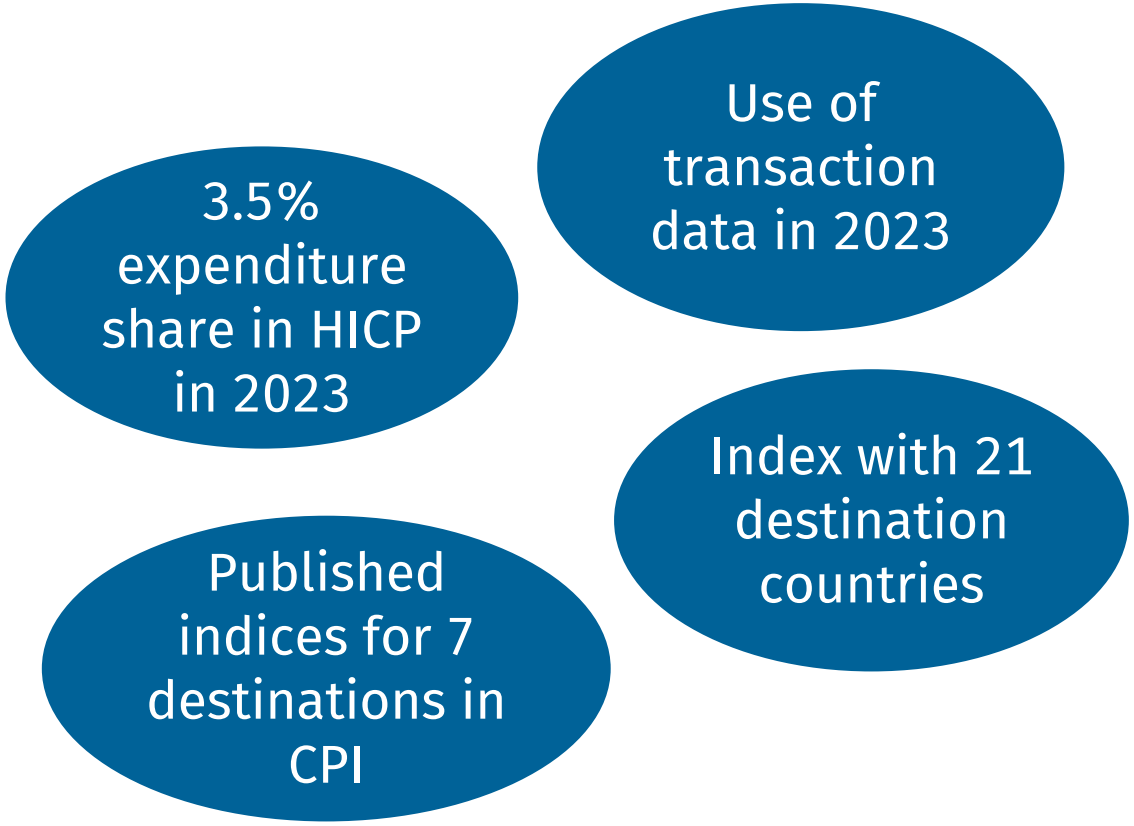
6) Index results

The German HICP and domestic CPI

Update of weights

- HICP: Update of consumption weights every year
- CPI: Update of consumption weights every 5 years
- Methodological changes are integrated with CPI weight updates

096 Package holidays



Projects on German package holidays

- Henn et al. (2019) tested transaction data for German package holidays for 2013-2018
  - Calculation methods performed well, none was clearly superior
  - Indices using transaction data differ from indices using offer pricesMissing variables: meal type and room category
- Feasibility study in 2021 of Destatis
  - Data source: tour operators or Amadeus?
  - Transaction data of 2015-2021
  - Variable handling, choice of index method, plausibility checks, imputation

Data

- Transaction data obtained by the global distribution system (GDS) Amadeus Germany
- Very high market share in Germany
- 21 destinations cover ~90% of bookings through Amadeus
- Bookings are heterogeneous in properties

Booking date	Travel date	Duration	Travelers	Depart. airport	Dest. airport	Hotel-name	City	Country	Hotel stars	Meal type	Room category	Price
01.03.2023	11.04.2023	10	2	HAM	AYT	Eftalia Hotel	Alanya - Türkler	Turkey	4	All Inclusive	DOUBLE-ROOM	662.90
06.02.2023	15.10.2023	8	4	CGN	FUE	allsun Hotel	Playa de Esquinzo	Canary Islands	5	All Inclusive	FAMILY-ROOM	2215.00

Methodology: hedonic double imputation

- Estimation of coefficients in current period t and base period b

$$(1)\ \ln(price_t) = \alpha_t + \beta_{1,t} * \ln(travellers_t) + \beta_{2,t} * \ln(duration_t) + \beta_{3,t} * star3_t + \ldots + \varepsilon$$

$$(2)\ \ln(price_b) = \alpha_b + \beta_{1,b} * \ln(travellers_b) + \beta_{2,b} * \ln(duration_b) + \beta_{3,b} * star3_b + \ldots + \varepsilon$$

- Prices are imputed by inserting bookings of both periods into equation (1)
- Price index is calculated via geometric mean

Methodology: hedonic double imputation

- Reliable indices in months with few bookings
- Takes into account many price-determining characteristics
- Method controls for omitted variable bias
- For each destination, month and over time, importance of variables differs
  - (forward) stepwise variable selection

Possible variables

- log no of travellers
- log duration
- booking time

Binary variables

- school holidays
- online/offline booking
- travel service included
- number of children
- hotel stars (2-5)
- meal types (5 categories)
- clusters of room category
- clusters of room features

Plausibility checks

Deletion of erroneous and implausible bookings

- Cancellations, duplicates, bookings with missing values
- Duration, Price, Number of travellers, booking time ≤ 0
- Cruises
- Bookings including a rental car
- Booking time > 366

Plausibility checks

Clearing of outliers by destination country

0.5% of bookings are cut from the upper (and lower) end of the spectrum

- price per day and person
- duration
- number of:
  - travellers
  - adults
  - children

Number of booking after outlier clearing

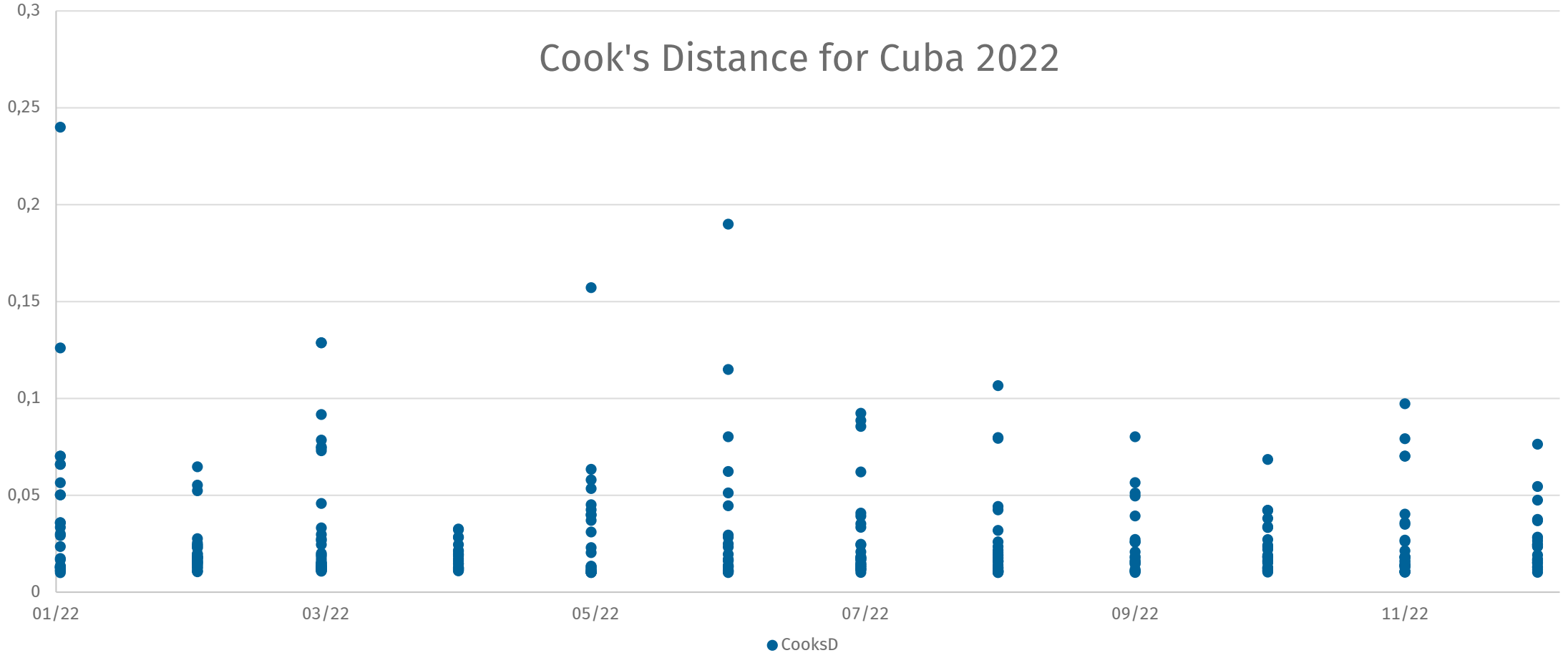
Dataset	Bookings 2019	Bookings 2020	Bookings 2021	Bookings 2022
Without erroneous bookings	1,656,985	1,029,711	740,073	1,349,654
Without 0.5%- and 99.5%-quantils	1,630,078	1,006,402	720,424	1,312,000
Difference	1.62 %	2.26 %	2.66 %	2.89 %

Plausibility checks

Influential bookings

- Cook‘s Distance: Indicator for the influence of a data point
- Cutoff value for outliers?
  - Commonly used rules of thumb ($D > 1$ or $D > 4/N$ and others) do not apply well
  - Little insight through analysis of values of single bookings
  - Graphical analysis seems best applicable

Plausibility checks



Plausibility checks

Minimum number of observations

Framework derived from own robustness checks:

- More than 140 bookings: Index is robust
- Less than 140 bookings: common rule of thumb

$$min\ N\ = N\ explanatory\ variables\ *\ 10$$

- Insufficient number of bookings: Index is compensated

Plausibility checks

Test of multicollinearity

Multicollinearity: two explanatory variables correlate

Ex.: most bookings to 5-star-hotels in Egypt include all-inclusive meals

- Biased index
- Indicator: variance inflation factor (VIF)
- Model improvement by deleting selected variables

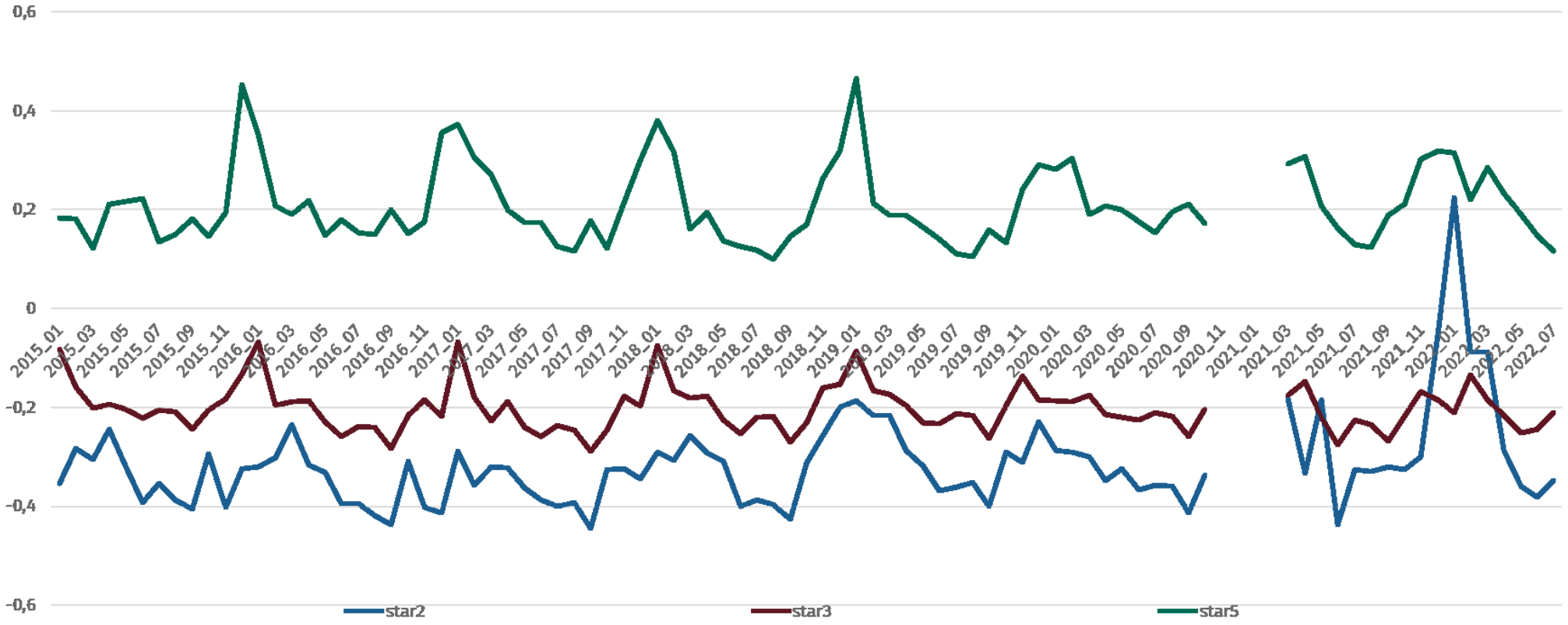
Plausibility checks

Further model examination

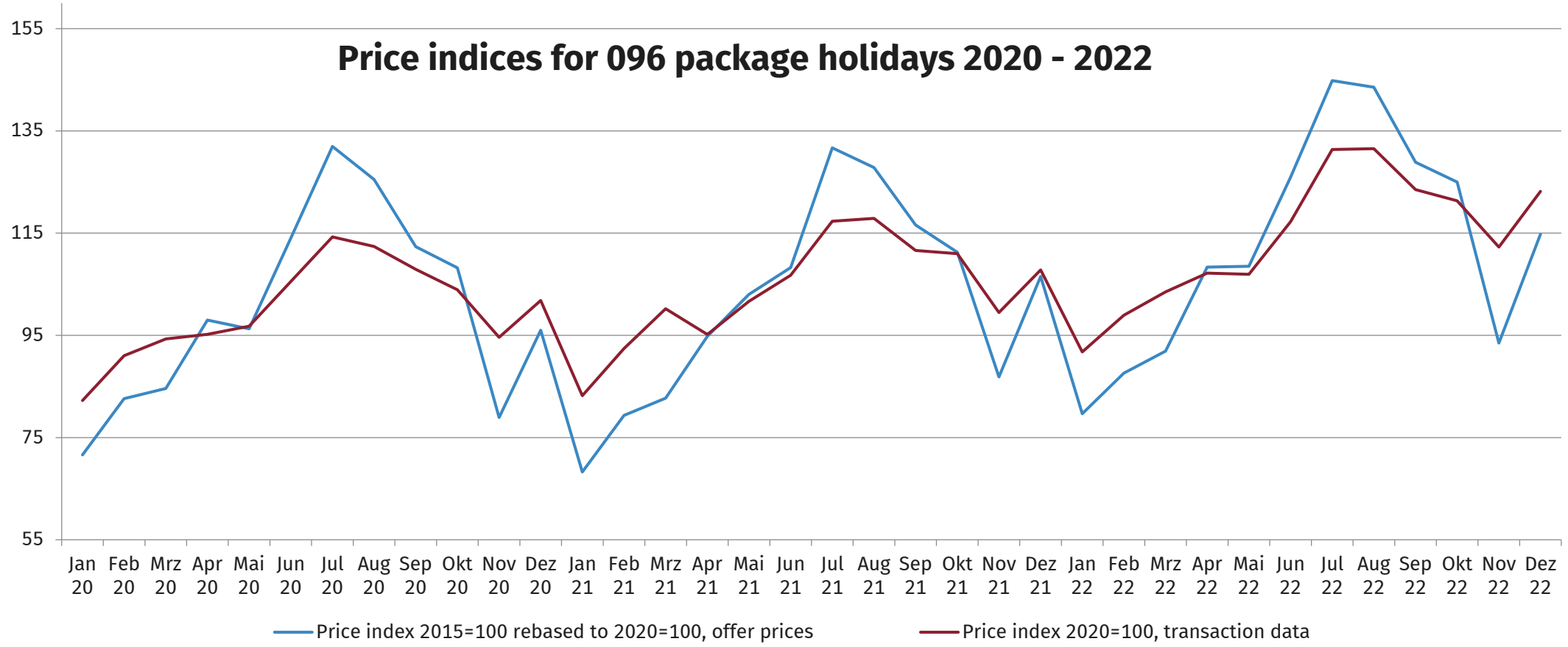
- Graphical analysis of β-coefficients and adjusted R^2
- Regular pattern over time
- Some coefficients display a seasonal pattern
- Significant inconsistencies point to a problem in the model
- Appear for dummy variables with little variation
- Bookings might have to be deleted

Plausibility checks

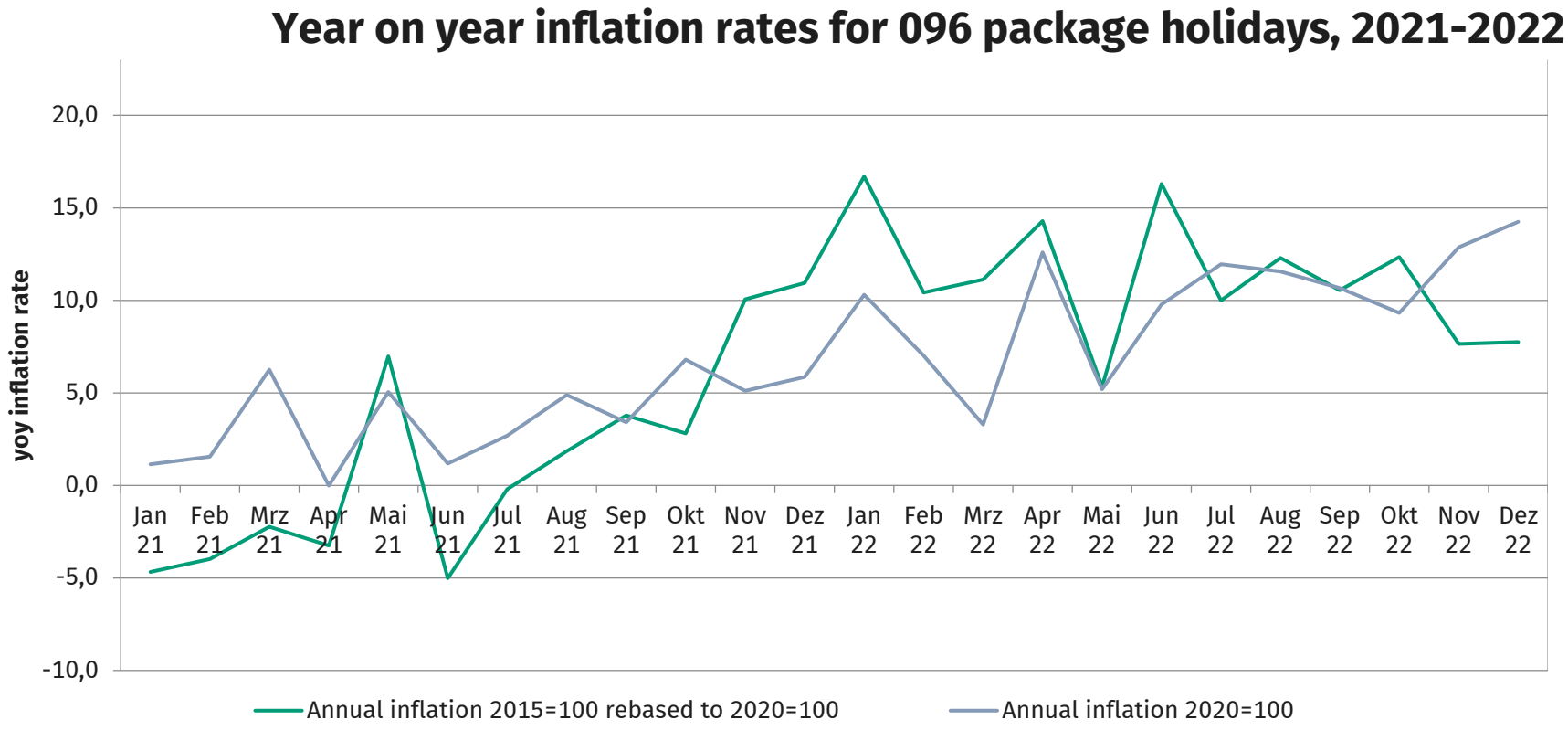
Regression coefficients for Balearic Islands



Index results



Index results



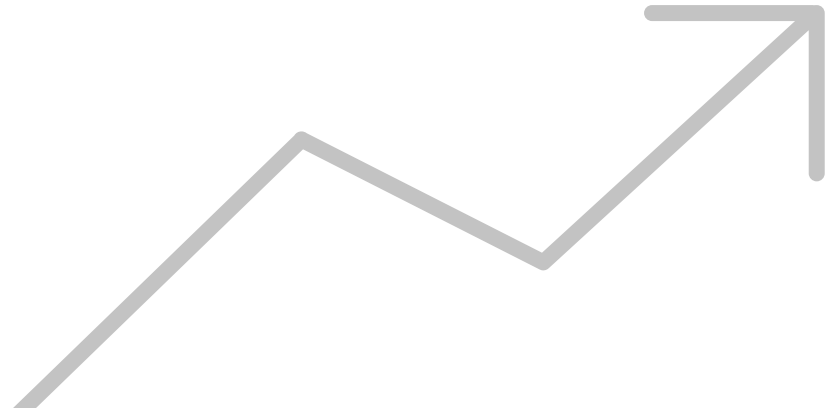
Contact

Statistisches Bundesamt
65180 Wiesbaden
Germany

Contact Person
Amelie Blasius
amelie.blasius@destatis.de
Phone +49 611 75-4732

www.destatis.de

www.destatis.de/kontakt



References

Henn, K., Islam, C.-G., Schwind, P., and Wieland, E. (2019). Measuring Price Dynamics of Package Holidays with Transaction Data. EURONA, 2/2019:95–132.

Holiday destinations covered by transaction data

Published indices for domestic CPI

- Balearic islands
- Canary islands
- Egypt
- Greece
- Turkey

Further indices

- Bulgaria
- Cuba
- Cypris
- Dominican Republic
- Italy
- Maledives
- Mauritius
- Mexico
- Portugal
- Spanish mainland
- Thailand
- Tunisia
- United Arab Emirates

- City trips
- Cruises