STATBEL
Belgium in figures

Evaluating different index methods for short-term rentals

Ottawa Group on Price Indices
Ottawa, 13-15 May 2024

economie

STATBEL.FGOV.BE

.be

Contents

STATBEL

- Data collection
- Estimating a weight
- Index methods
- Final remarks

STATBEL

01 Data collection

Sample of cities

STATBEL

- In 2018, analysis carried out to collect data for short-term rentals/short-stay accommodation:
  - Airbnb deemed most important
  - Web scraping as a data collection method

- Sample selection based on traditional tourism statistics and # of accommodations available

- 3 cities were sampled with # of accommodations > 1000 per month:
  - Brussels: : data starting from 2018 (with significant data gaps)
  - Antwerp & Ghent: data from 2020

- Covid-19 pandemic: lockdowns and travel restrictions → useable data from November 2020 onwards

STATBEL

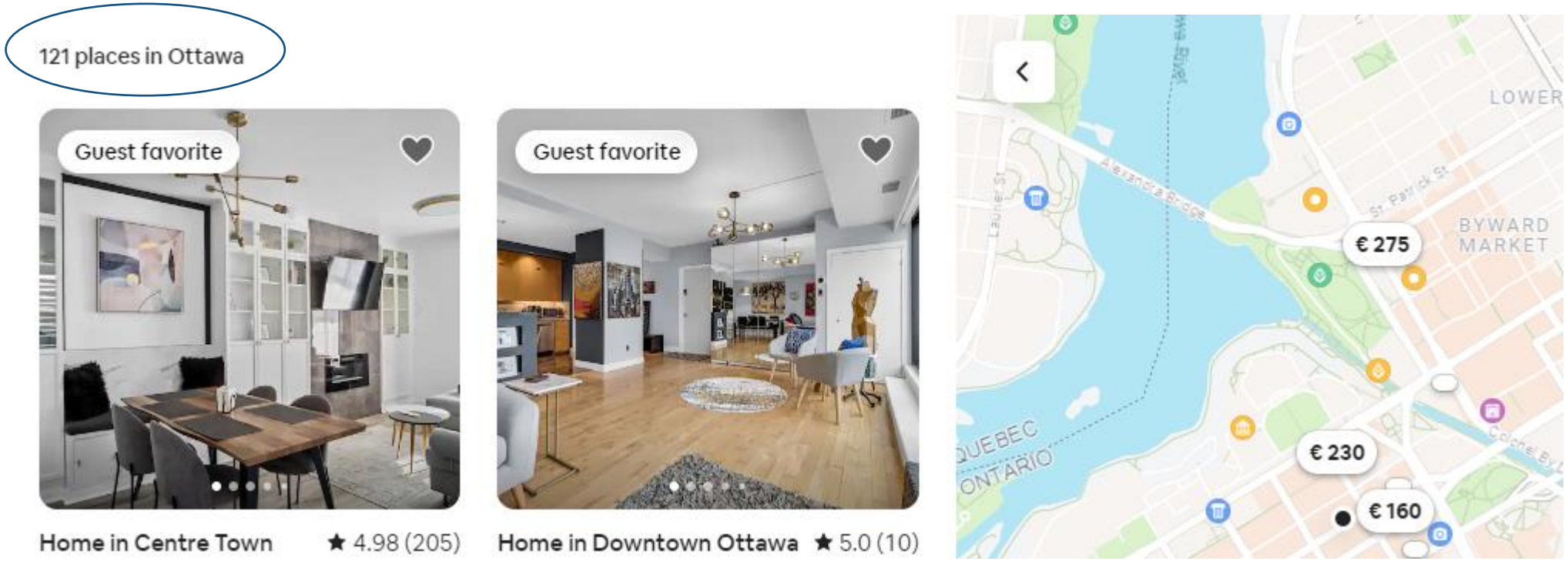
- Confirmed representativity of sampled cities in 2021 based on tourism statistics data
- Eurostat information on short-stay accommodation offered via online collaborative economy platforms (= platform data)
- Agreement for tourism statistics with Airbnb, Booking.com, Tripadvisor and Expedia Group

	Total number of stays (2019 data)	%
Belgium	820.703	100,0%
Brussels	290.756	35,4%
Antwerp	84.996	10,4%
Ghent	53.952	6,6%
Charleroi	6.101	0,7%
Liège	24.826	3,0%
Bruges	48.600	5,9%
Ostend	24.478	3,0%

STATBEL

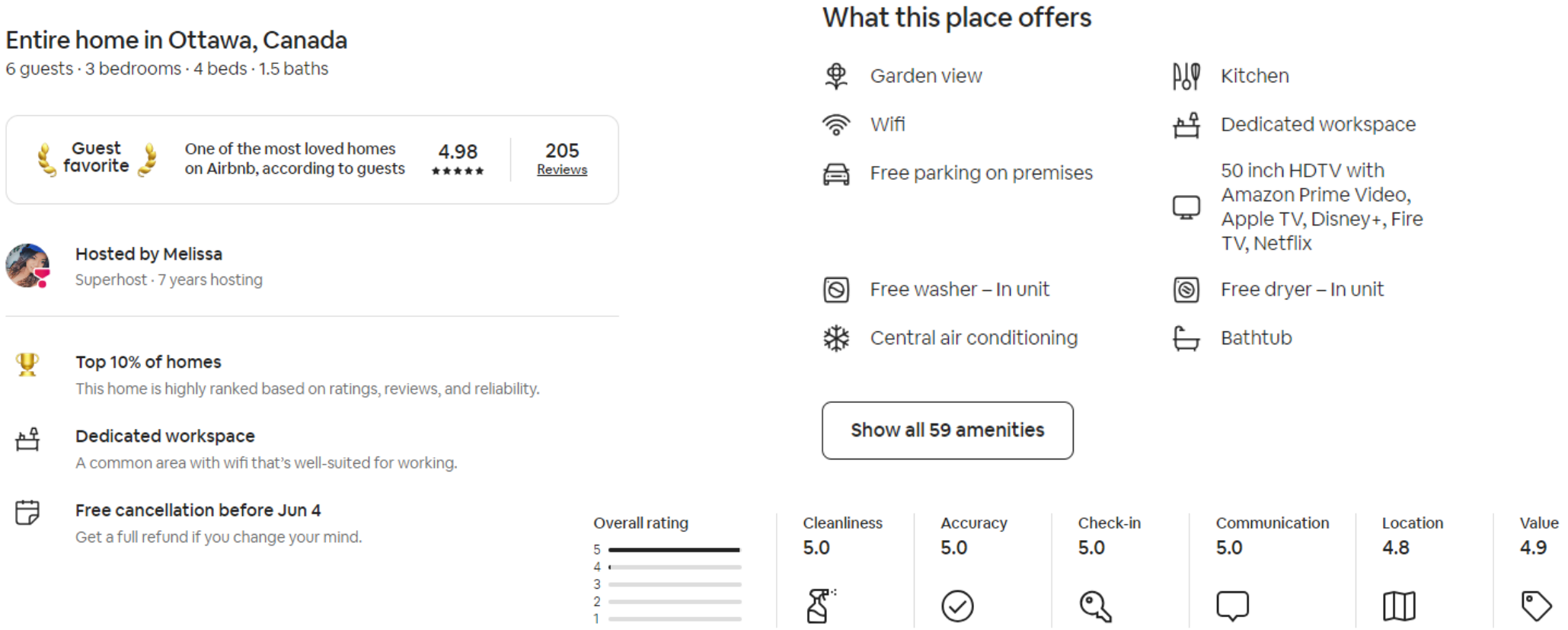
- Every search result on Airbnb is limited to a maximum of 270 accommodations (15 pages of 18 accommodations)

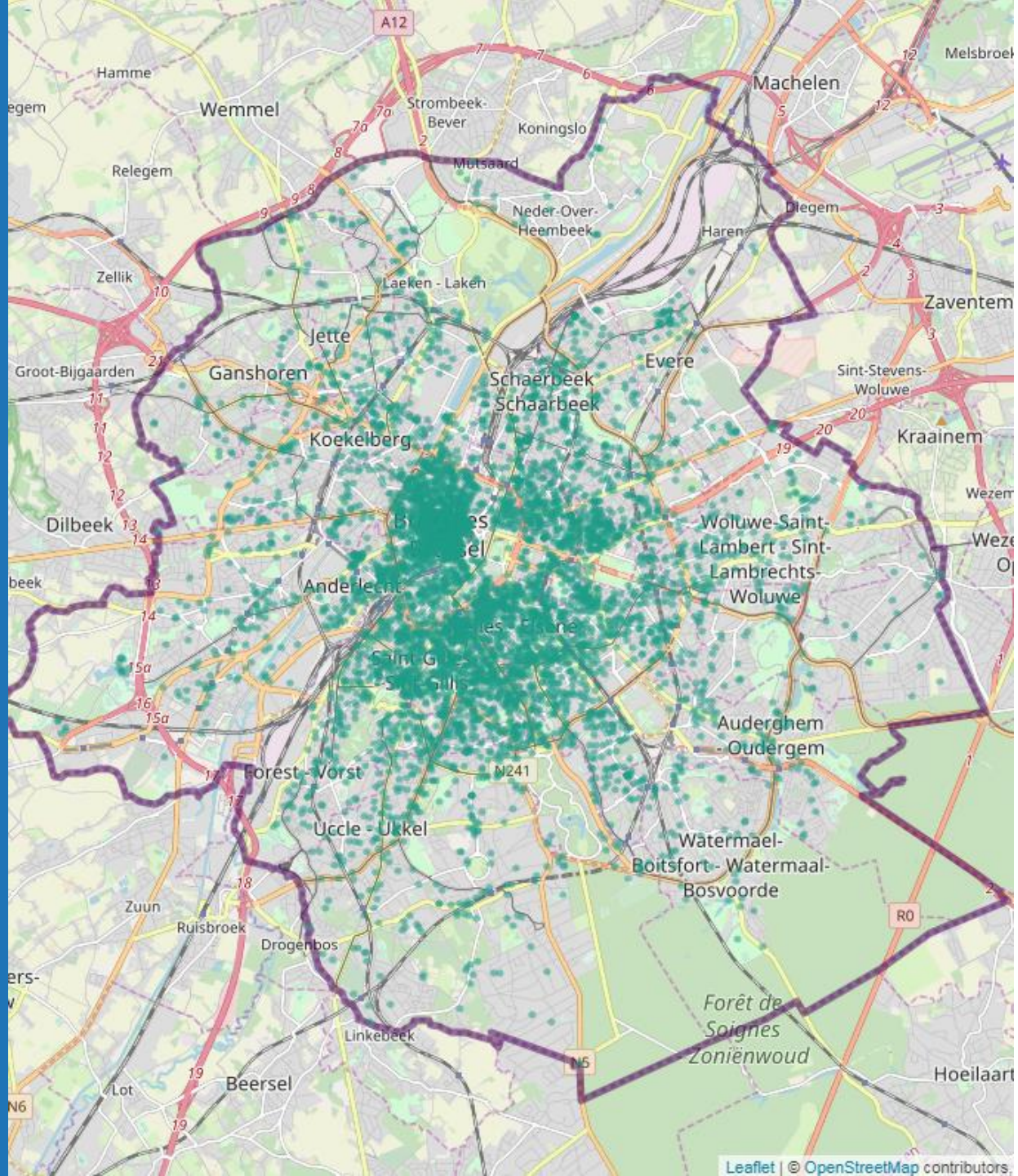
- By focusing on areas on the map or by using filters we can apply an iterative procedure to get all available listings in a city



STATBEL

- Detailed information available per accommodation





Scraped data

- **Type of property**: house, apartment, …
- **Characteristics** : max number of guests, bedrooms, bathrooms, …
- **Amenities**: wifi, dryer, free parking, …
- **Calendar information**: price, minimum number days for a stay
- **Host information**: response time, superhost, …
- **Rating information**: cleanliness, accuracy, …
- **Location information**: neighbourhood, geo-coordinates

STATBEL

02 Estimating a weight

STATBEL

- Airbnb transactions for Belgium are carried out through their Luxembourg and Irish subsidiaries:
  - No VAT number, no tax representative, no annual accounts declared

- No other Airbnb related administrative data available in Belgium:
  - Rental income is in most cases untaxed
  - No registration required for a host with administrative authority

- Not captured as a separate accommodation services category in Belgian national accounts (implicitly captured in imputed rents for OOH)

- Own estimation using price X volume approach

STATBEL

- Price for a stay = ((price per night + fee for additional guest) x number of nights + cleaning fee) x service charge

	2020		
	Price	Stays	Expenditure = price * stays
Brussels	338 €	97.653	33.013.608 €
Antwerp	427 €	32.165	13.723.995 €
Ghent	434 €	27.255	11.835.124 €
Charleroi	254 €	3.245	823.937 €
Liège	327 €	13.581	4.436.355 €
Bruges	491 €	19.095	9.383.845 €
Ostend	475 €	20.538	9.756.546 €
Total (excl. cities)	389 €	223.706	86.928.636 €
Total (incl. cities)			169.902.046 €

- Deduct +-25% for business stays → depending on the year: consumption expenditure is around 0,1% to 0,15% of total household consumption.
- Experimental results, since not everything has been analyzed (e.g., intermediate consumption)

STATBEL

03 Index methods

STATBEL

- Traditional method of matching items to compile a price index

- Airbnb accommodation have unique identifiers → can be used to match same accommodation

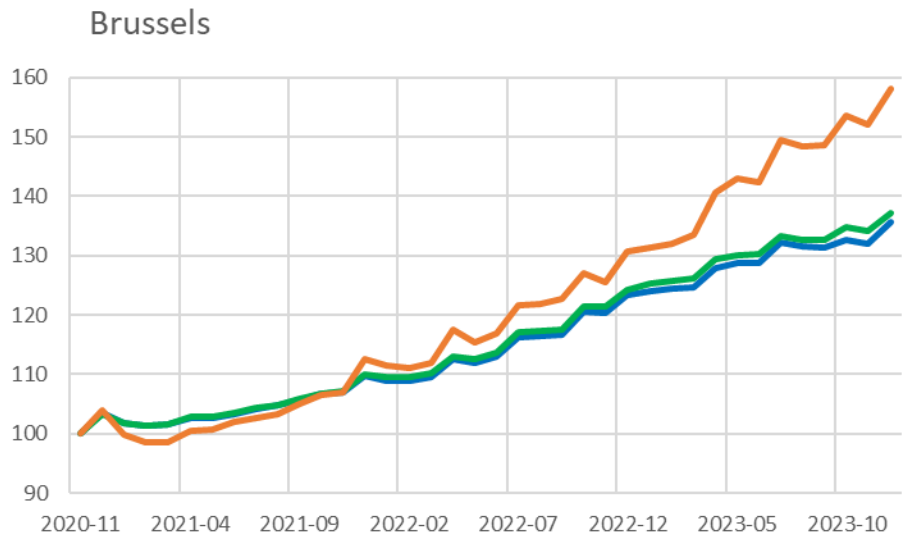
- Time product dummy index: $\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t$

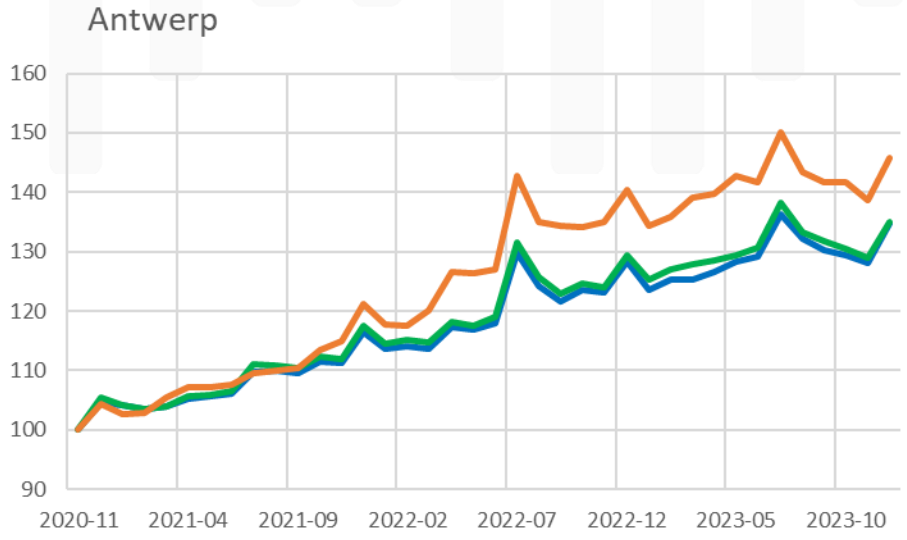
- GEKS-Jevons index: $P_{GEKS-J}^{0,t} = \prod_{l=0}^{T} \left(P_J^{0l} P_J^{lt}\right)^{(1/T+1)}$

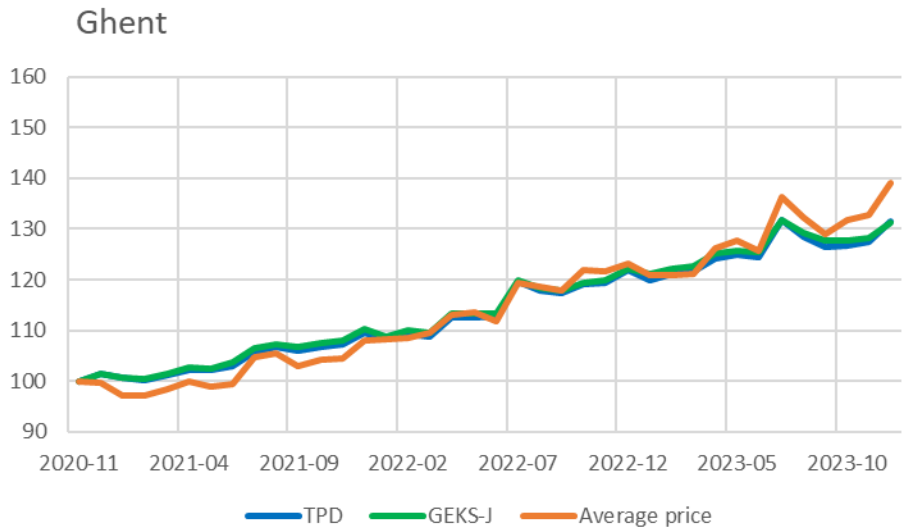
- Compared with a simple average price

Matched model indices vs average price

STATBEL







STATBEL

- However, new listings appear, and others disappear

- Not taken into account in matched model indices

- Examined 3 methods to “solve” this problem:
  - Hedonic methods
  - Combining a matched model index with hedonic imputations
  - Stratification

STATBEL

- Detailed characteristics can be used for hedonics

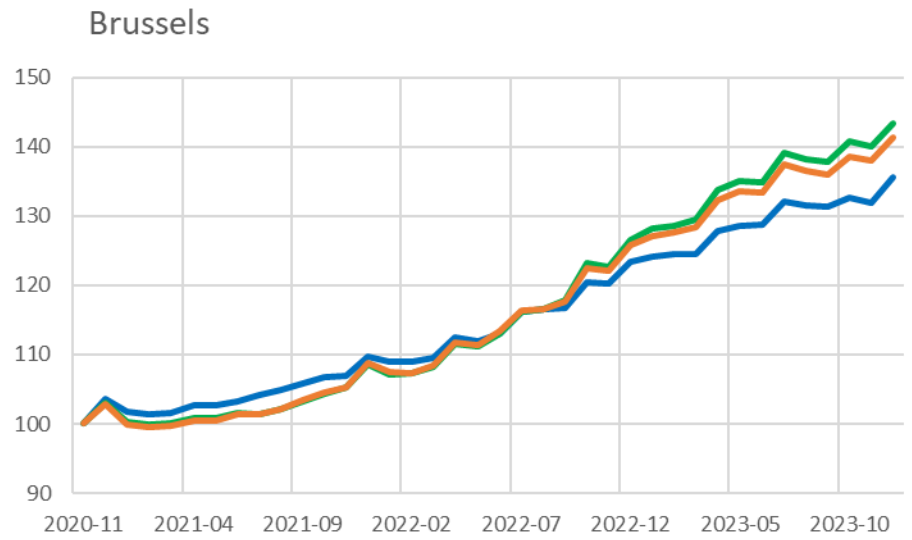
- R^2 was on average between 0.76 and 0.82 with and did not differ that much from city to city

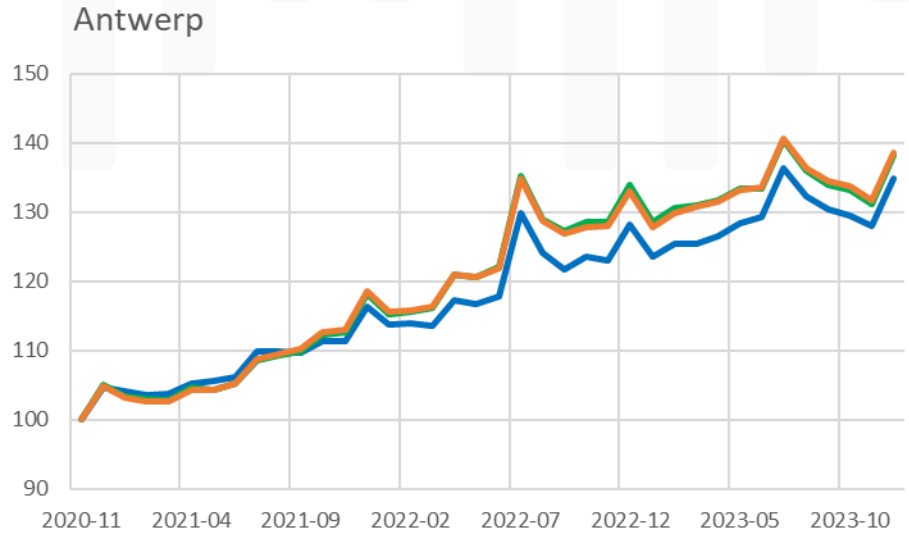
- 2 time dummy hedonics methods:
  - TDH with all periods pooled in the same regression → drawback: fixity of the parameters and without product churn the index does not equal a matched index
  - Chained TDH where two adjacent periods are pooled together → drawback: fewer observations are used which might cause unstable coefficients

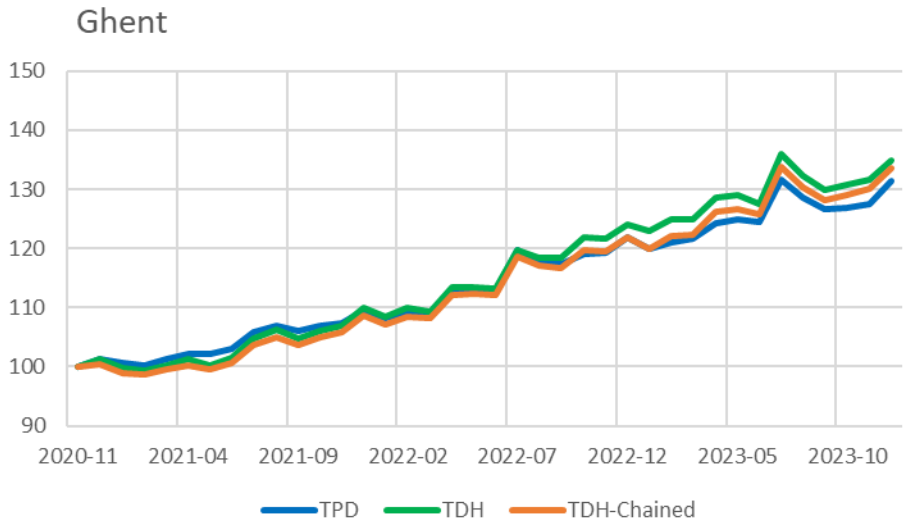
- Log-linear specification (OLS estimated): $\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k z_{ik} + \varepsilon_i^t$

Hedonic indices vs TPD

STATBEL







Imputation Jevons GEKS

STATBEL

- Combining matching and imputations, imputation Jevons GEKS: $P_{GEKS-IJ}^{0,t} = \prod_{l=0}^{T}\left(P_{IJ}^{0l} P_{IJ}^{lt}\right)^{(1/T+1)}$

- With imputation Jevons index: $P_{IJ}^{0,t} = \prod_{i \in U_M^{0,t}} \left(\frac{p_i^t}{p_i^0}\right)^{0.5(N_0+N_t)} \prod_{i \in U_D^{0,t}} \left(\frac{\hat{p}_i^t}{p_i^0}\right)^{0.5(N_0)} \prod_{i \in U_N^{0,t}} \left(\frac{p_i^t}{\hat{p}_i^0}\right)^{0.5(N_t)}$

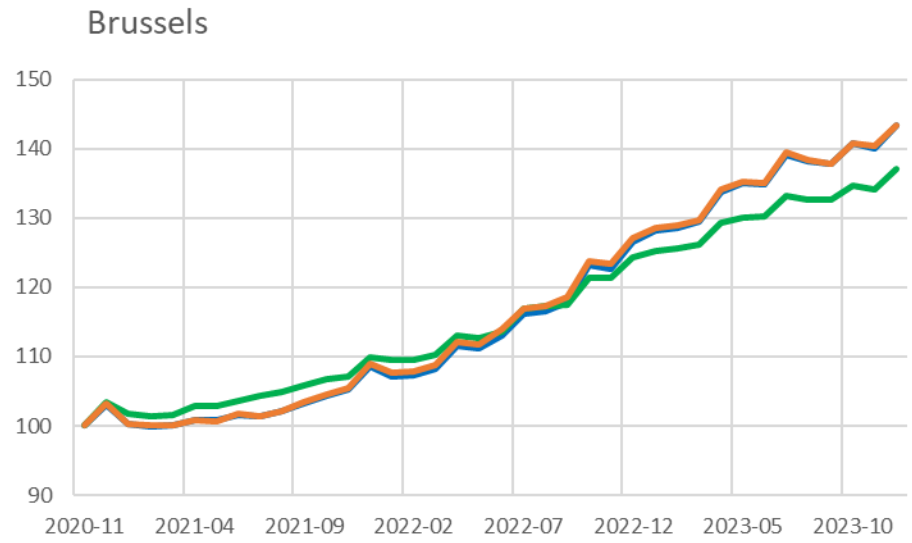
- With $\hat{p}_i^t$ and $\hat{p}_i^0$ as imputed prices for new and disappeared accommodations → with bilateral time dummy hedonic method

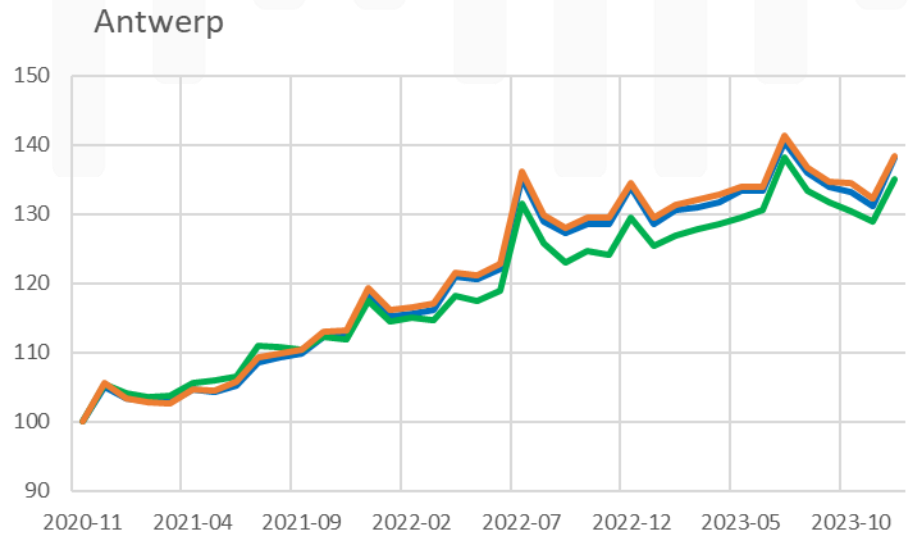
- Drawbacks: many bilateral regressions, fewer observations might cause unstable coefficients

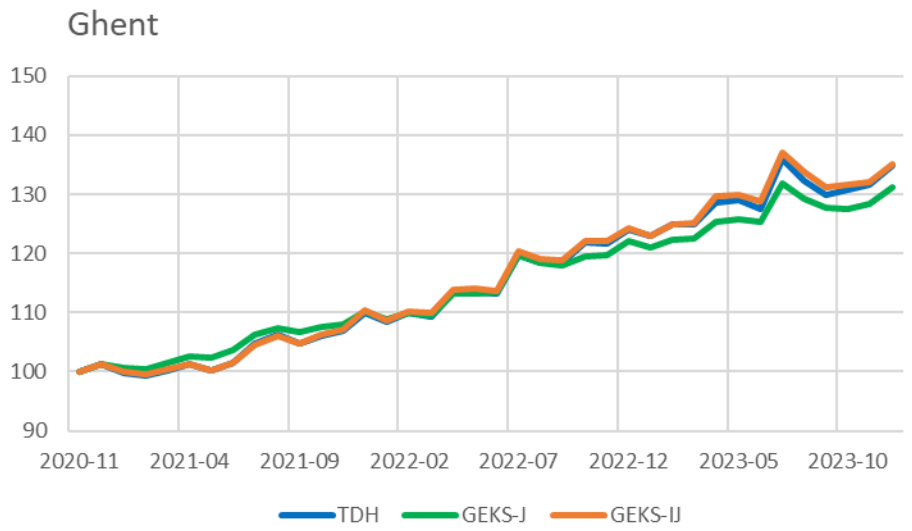
- Advantage compared to (chained) TDH: without product churn, the index is equal to a matched index

Imputation Jevons GEKS

STATBEL







STATBEL

- Stratification:
  - accommodations are attributed to a strata based on characteristics
  - the price of new and disappeared accommodations is directly compared within the same strata

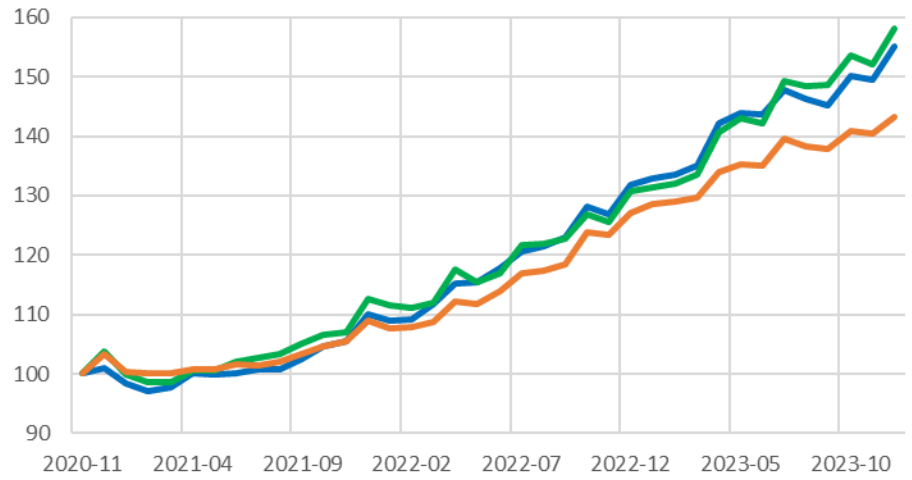
- Aggregation of the strata indices gives the global index

- Drawback of stratification is a potential unit value bias, because:
  - Limited # of variables can be used
  - All variables used for stratification are treated as categorical

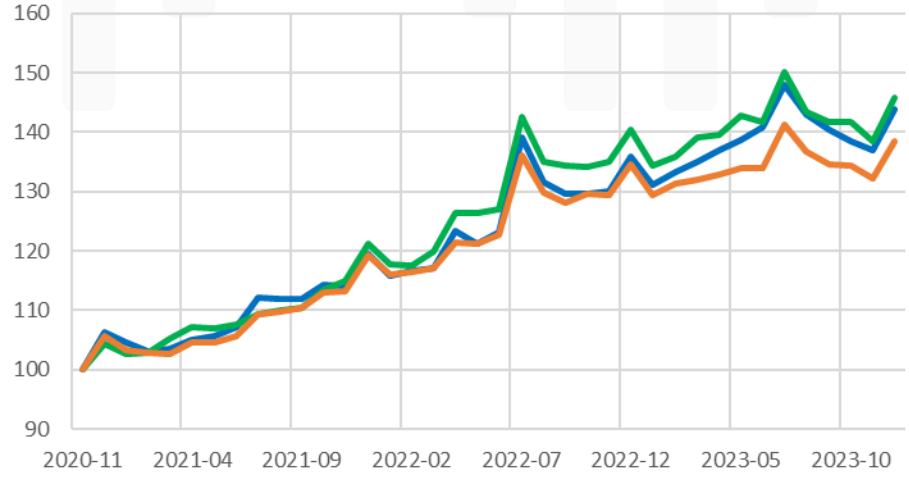
- To determine the variables for stratification we used MARS

STATBEL

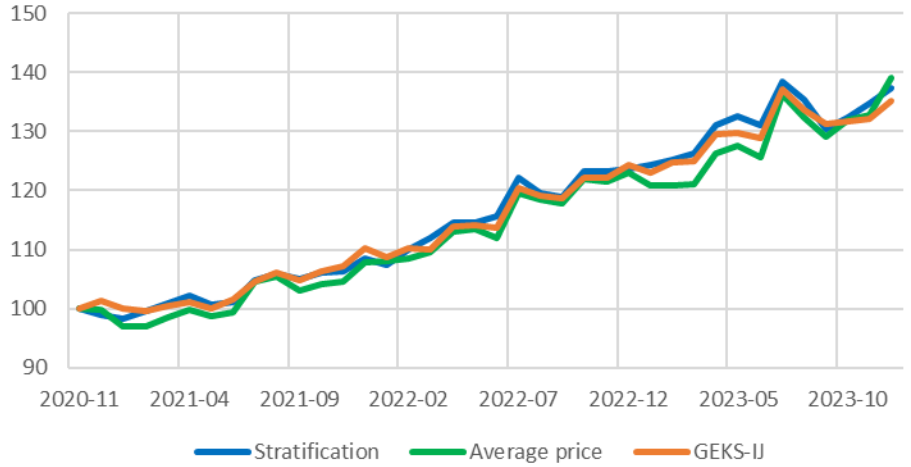
Brussels



Antwerp



Ghent



STATBEL

04 Final remarks

STATBEL

- Experimental estimation of consumption expenditures > 0,1%
- Sufficient characteristics information can be scraped from the Airbnb website
- Index methods
  - Matched indices : downward bias in our sample
  - TDH performs a bit better than a bilateral chained TDH
  - Imputation Jevons GEKS index quite similar results to TDH, but might be preferred
  - Stratification: unit value bias
- It is possible to compile a reliable index
- Future work: impact of splicing or extension methods

STATBEL
Belgium in figures

Thank you!

Ken Van Loon
ken.vanloon@economie.fgov.be