Index compilation with online prices for household appliances and consumer electronics

Meeting of the Group of Experts on Consumer Price Indices

7th of June 2023

Lucien May

Botir Radjabov

STATEC

Background

Introduction of Scanner Data in the CPI in 2018

Multilateral GEKS method introduced in 2021 (COICOP divisions 01 and 02)

Next step: *Household appliances* and *Consumer Electronics*

- Market research data for Luxembourg not available timely for CPI production
- No access to Scanner Data for a major retailer of consumer electronics

Bulk Scraping of the retailer's website for two years (prices and characteristics)

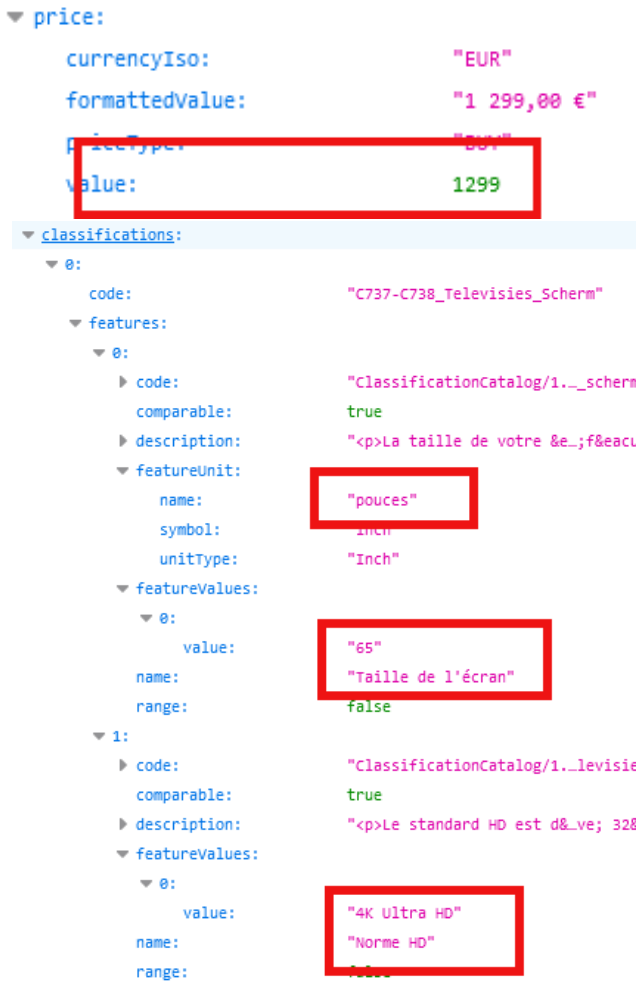
How to exploit this data source for the CPI?

Online data collection

- Access to the API of the retailer beginning of 2021
- Agreement with the retailer for scraping options (time, crawl-delay, frequency)

In practice:

- Retrieval of product sheets (.json files) for all products once per week
- Extractions of relevant information (internal classification, prices, product identifier, characteristics) in databases



Bulk dataset

Starting June 2021:

- Weekly data of (offer) prices and characteristics of all products for sale (>9000 products per week)
- We experienced one *temporary access problem* in two years, due to an internal change of the retailer (solved in two days)

Mapping of internal retailer categories into COICOP (updated every month)

Online prices and in shop prices are the same

No individualized pricing (price discrimination)

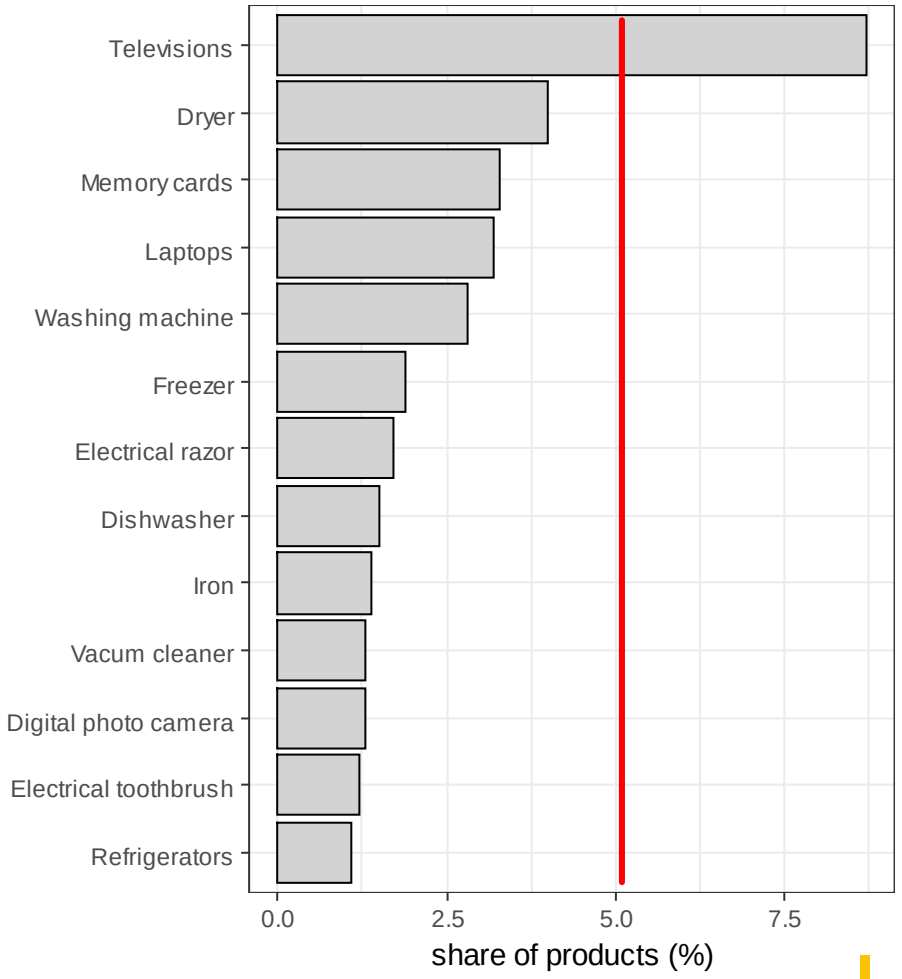
Price collection frequency

- For majority of products: No or one price change within a month
- Litte or no « dynamic pricing »
- **Weekly data collection seems to be sufficient**

- *Monthly aggregated price* p_i^t of a product i by

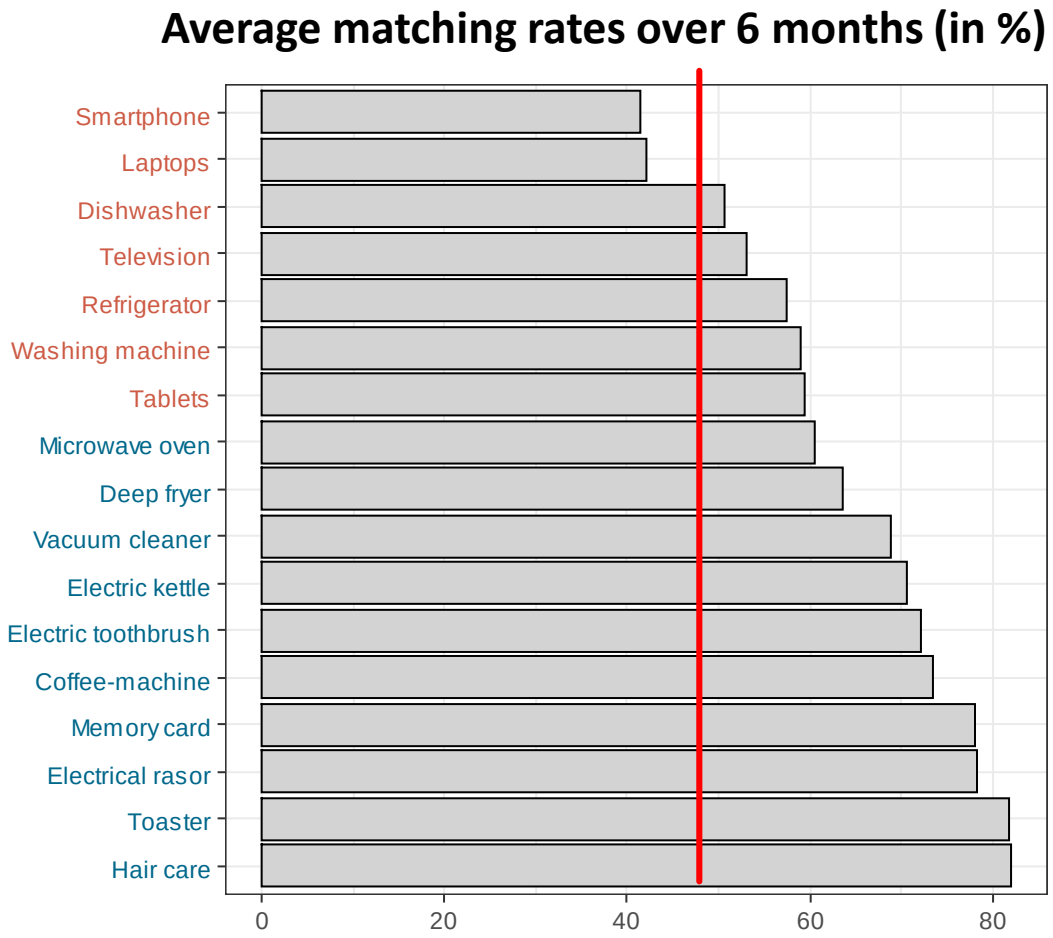
$$p_i^t = \frac{1}{L_{it}} \sum_{k=1}^{L_{it}} p_i^{t,k}$$

Average share of products with 2 price changes in a month or more

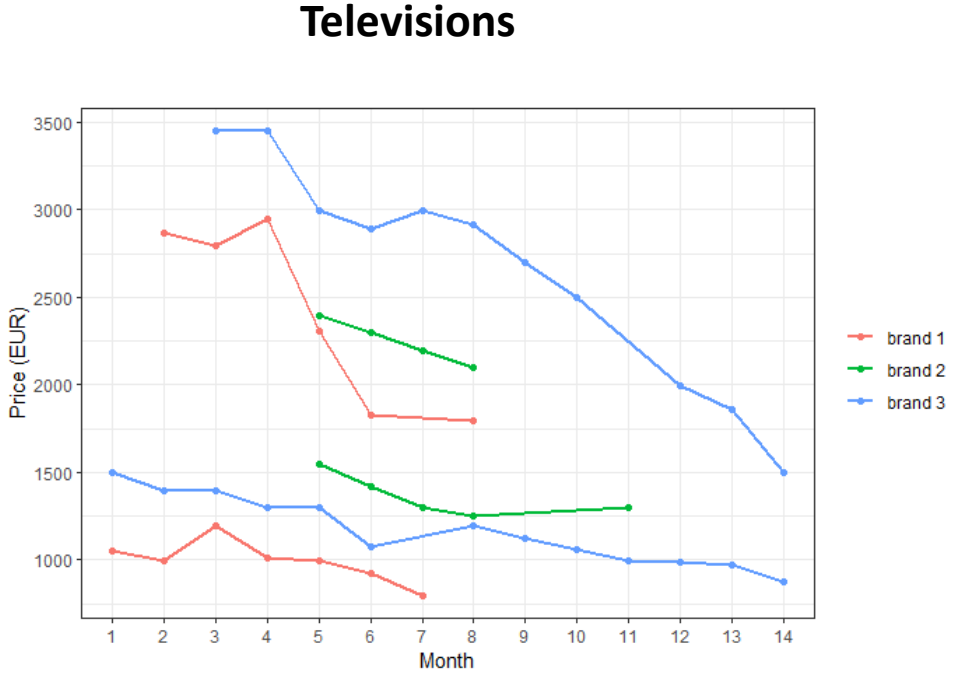


Assortment dynamics and lifecycle pricing

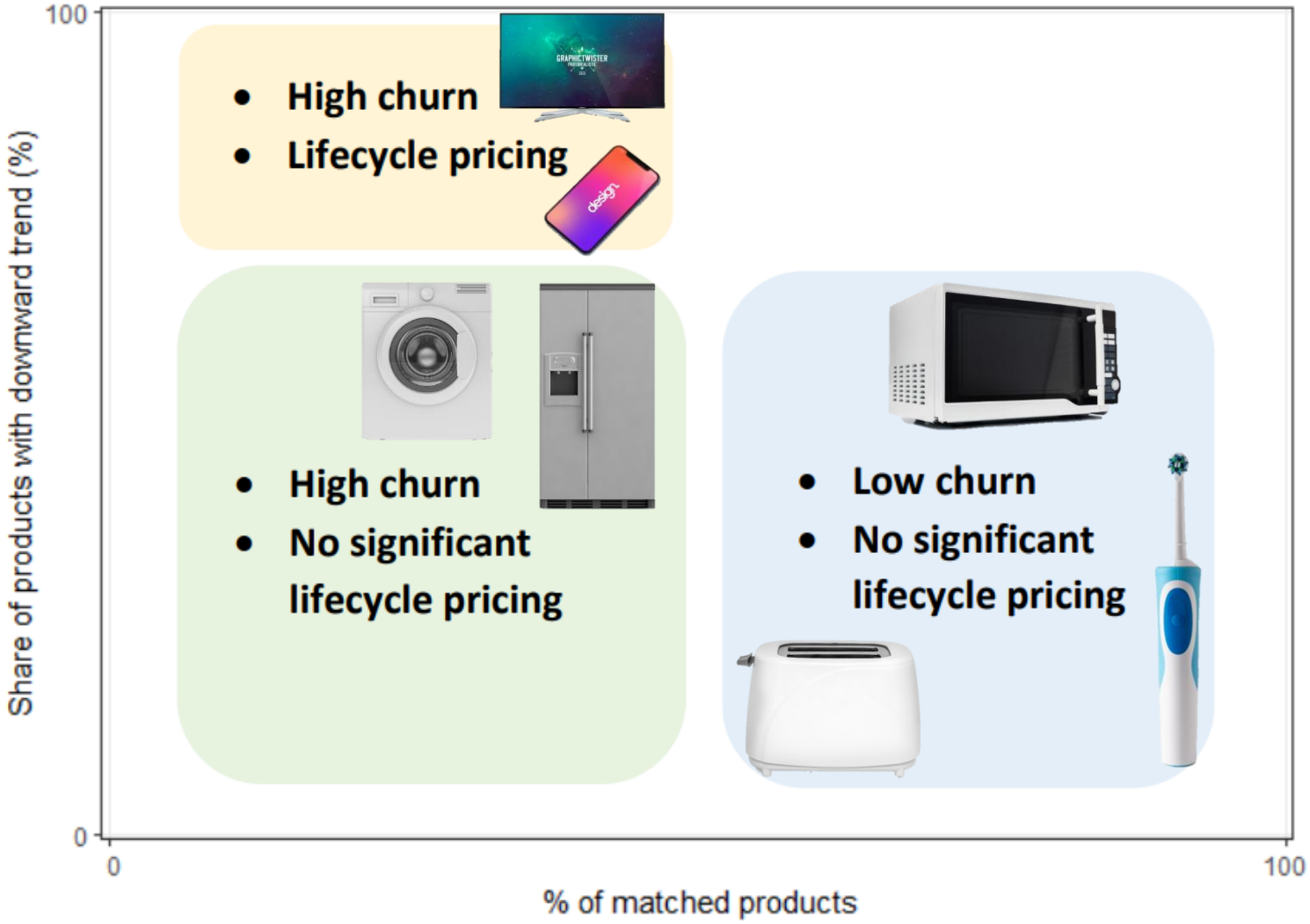
High churn for consumer electronics and big household appliances



Lifecycle pricing for consumer electronics



Conceptual classification of product categories



Price index methods

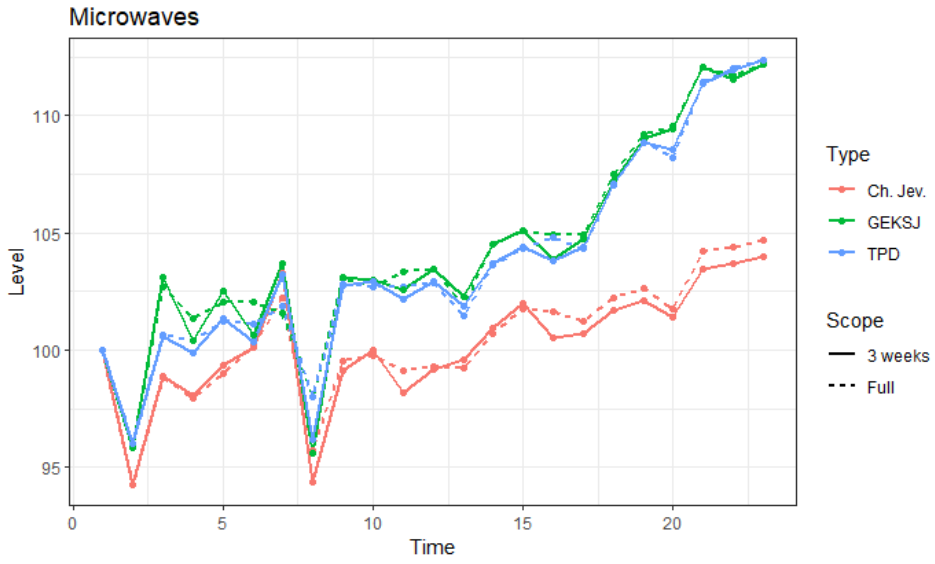
« Low » churn product categories without lifecycle pricing

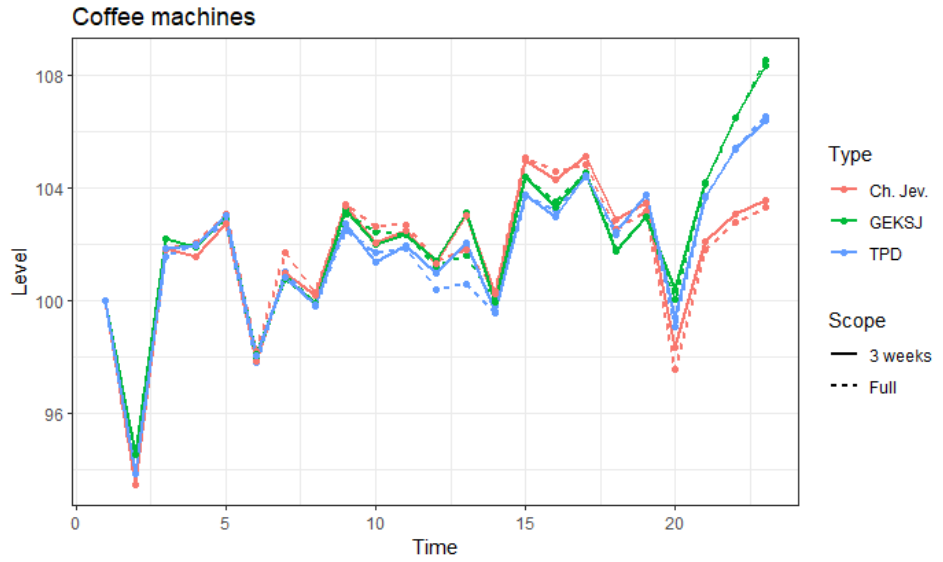
Small household appliances, electrical appliances for personal care, printers etc.

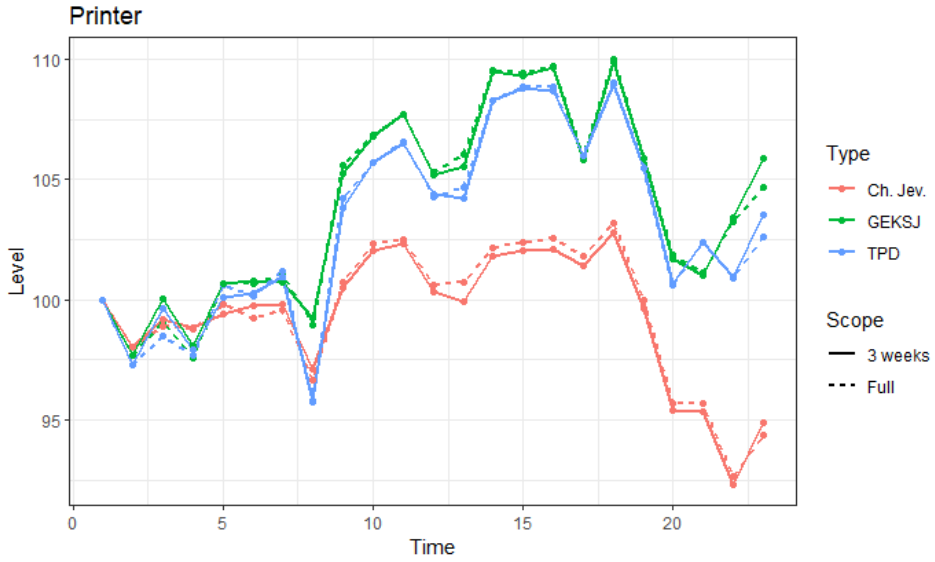
Chained Jevons Index	$I_{ChJ}^{0,t} = \prod_{j=1}^{t} I_J^{j-1,j}$
GEKS-Jevons Index	$I_{GEKS-J}^{0,t} = \prod_{j=0}^{T} \left(I_J^{0,k} I_J^{k,t}\right)^{\frac{1}{T+1}}$
Time product dummy Index	$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t$ *(OLS estimation)* $\qquad I_{TPD}^{0,t} = e^{\widehat{\delta^t}}$

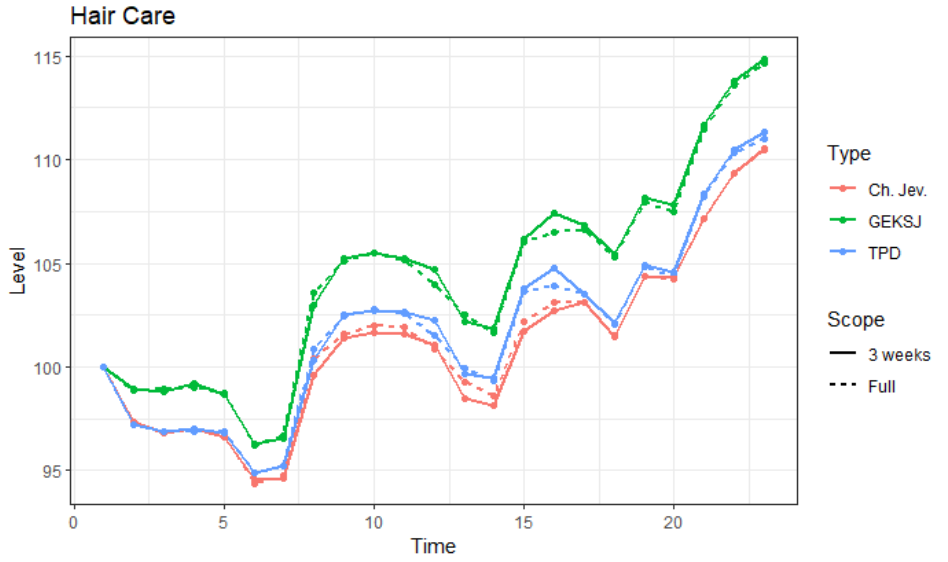
$I_J^{r,s}$ is the Jevons Index between periods r and s.

Results for some (low churn) product categories









Price index methods

« High » churn product categories with lifecycle pricing

Tablets, laptops, televisions, smartphones etc.

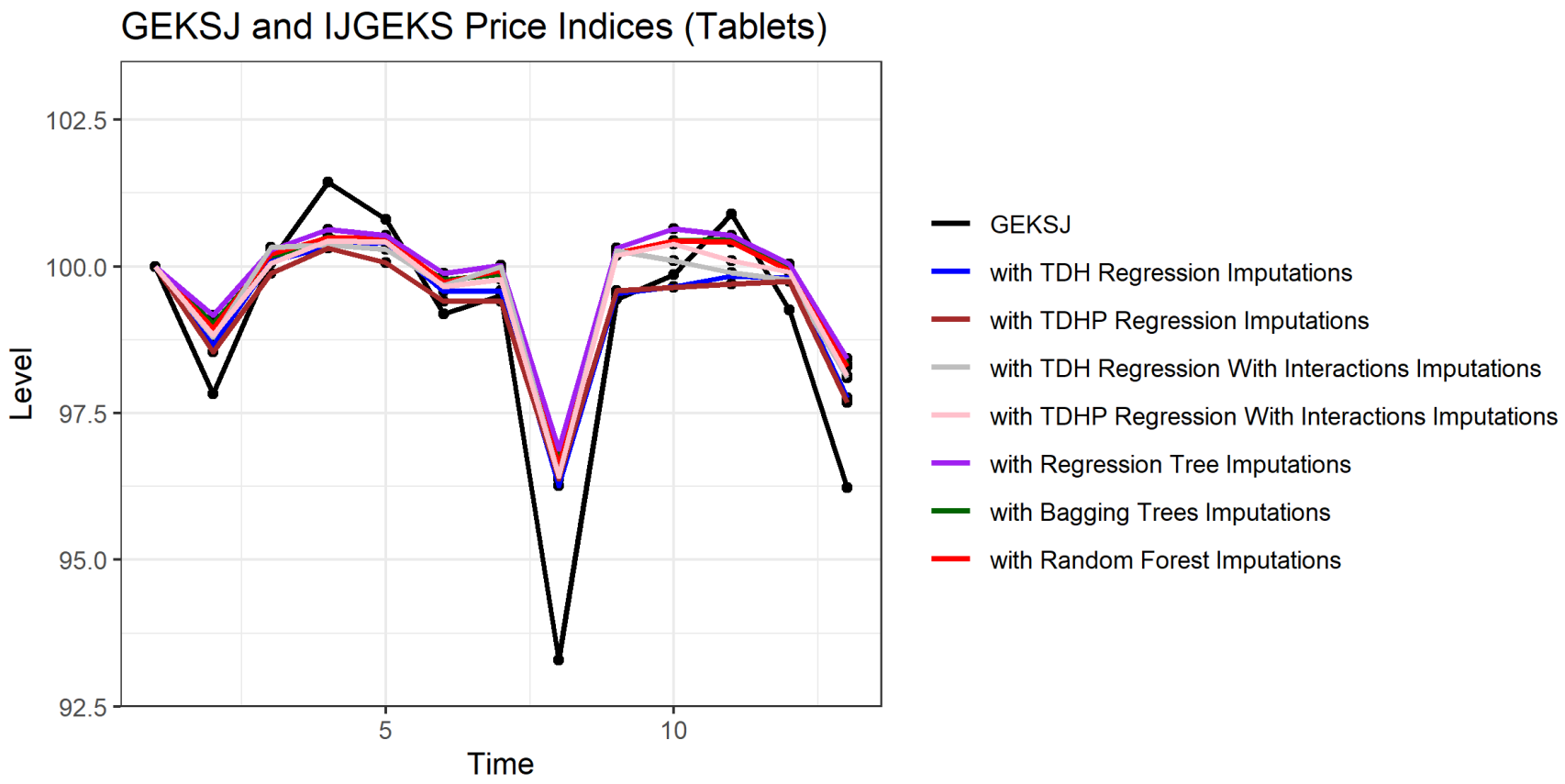
Imputation Jevons Index	$I_{IJ}^{0,t} = \prod_{i \epsilon U_M^{0,t}} \left(\frac{p_i^t}{p_i^0}\right)^{\frac{1}{2N_0}+\frac{1}{2N_t}} \prod_{i \epsilon U_D^{0,t}} \left(\frac{\hat{p}_i^t}{p_i^0}\right)^{\frac{1}{2N_0}} \prod_{i \epsilon U_N^{0,t}} \left(\frac{p_i^t}{\hat{p}_i^0}\right)^{\frac{1}{2N_t}}$
Imputation Jevons GEKS (IJGEKS) Index	$I_{IJGEKS}^{0,t} = \prod_{j=0}^{T} \left(I_{IJ}^{0,k} I_{IJ}^{k,t}\right)^{\frac{1}{T+1}}$

$\hat{p}_i^0$ and $\hat{p}_i^t$ are imputed prices of new and disappearing products

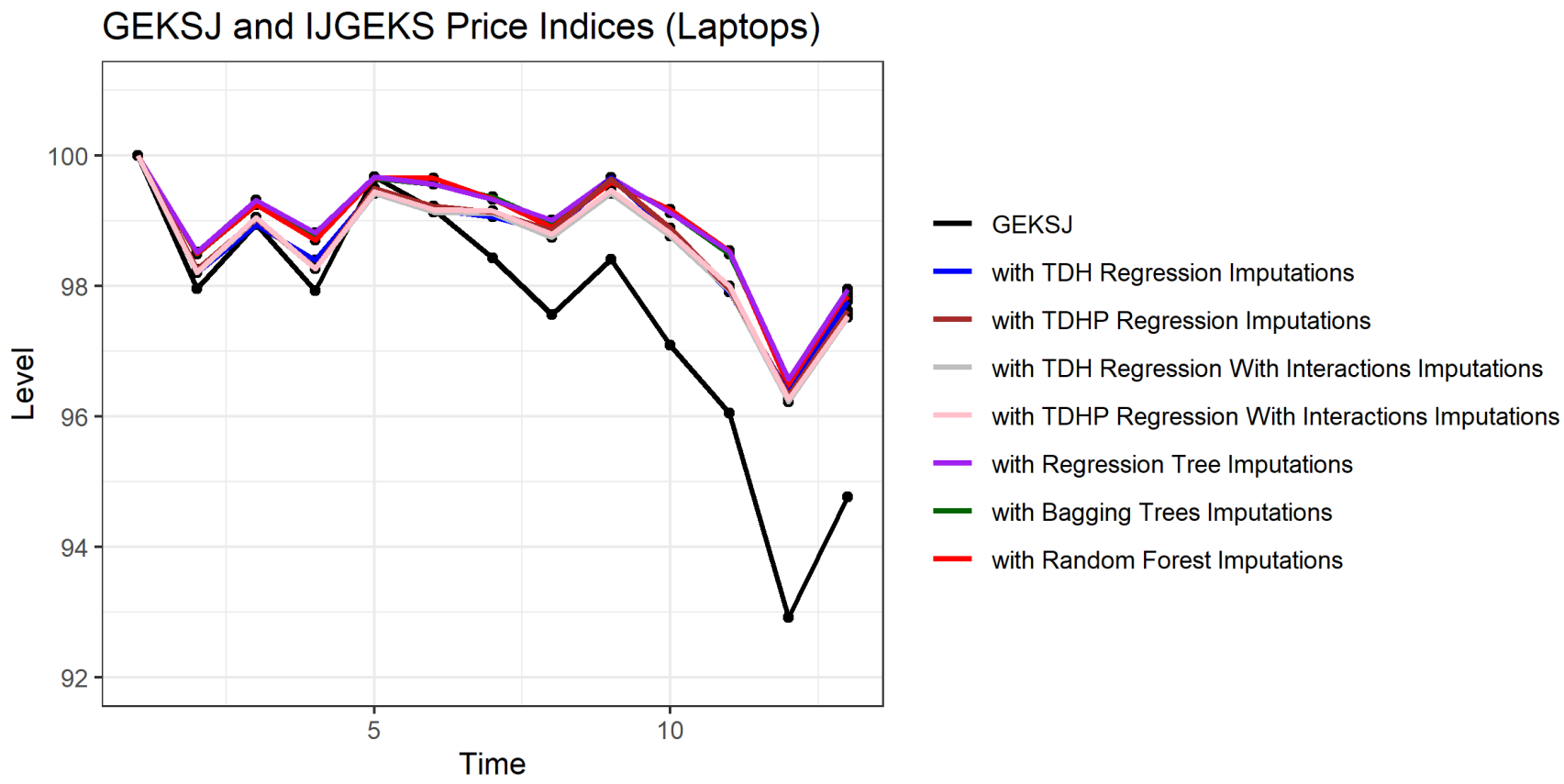
The chosen price imputation methods

Hedonic Linear Regression Based Price Imputation Methods	Tree Based Price Imputation Methods
Time Dummy Hedonic Regression	Regression Tree
Time Dummy Hedonic Polynomial Regression	Bagging Trees
Time Dummy Hedonic Regression with Interactions	Random Forest
Time Dummy Hedonic Polynomial Regression with Interactions	

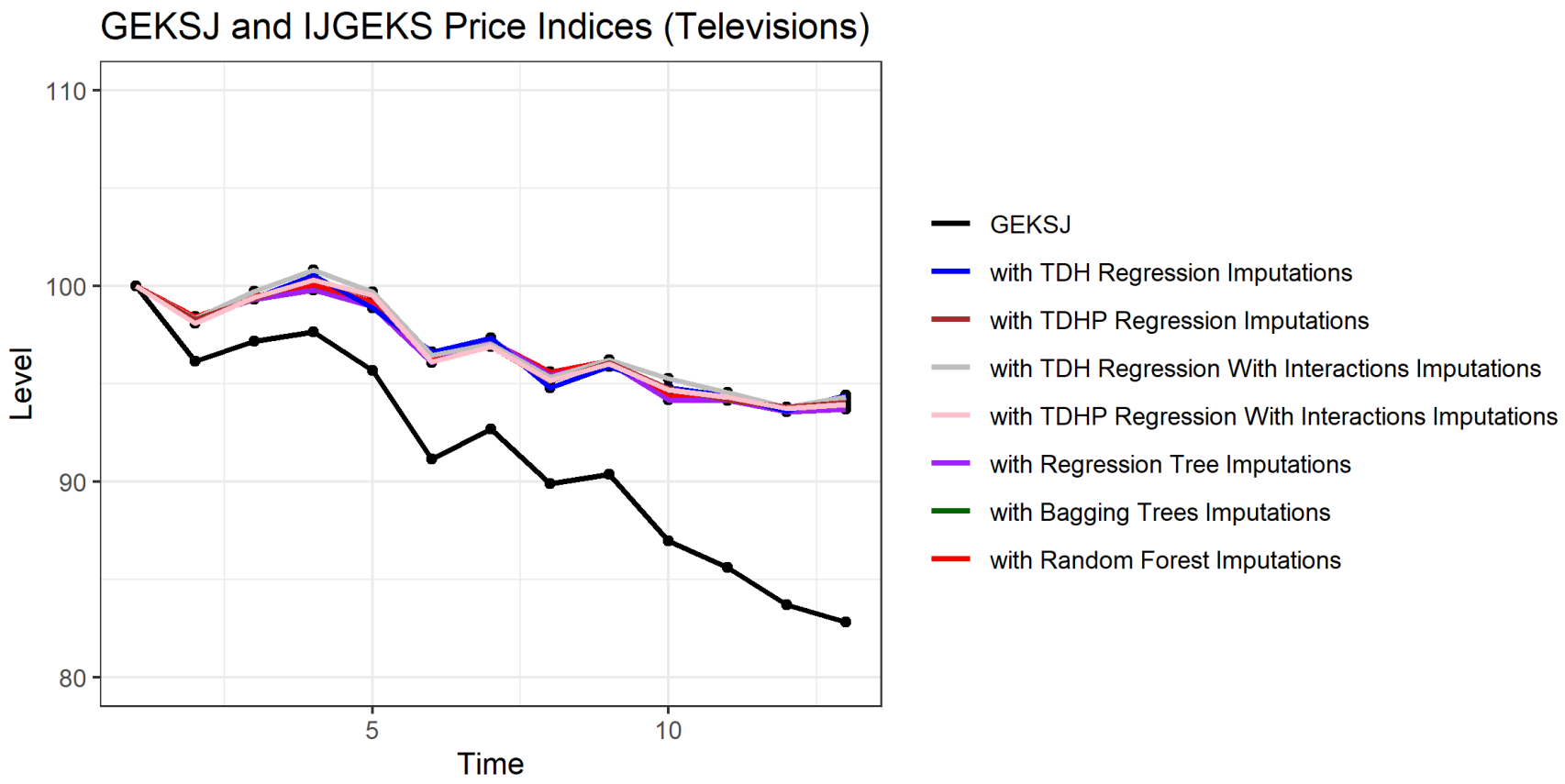
Results for some (high churn) product categories



Results for some (high churn) product categories



Results for some (high churn) product categories



Which price imputation method is more accurate?

R^2 values of the Price Imputation Methods (average for 20 samples)

	Tablets	Laptops	Televisions
Time Dummy Hedonic Regression	0.9727873	0.8898543	0.8913327
Time Dummy Hedonic Polynomial Regression	0.9820377	0.9004979	0.9254533
Time Dummy Hedonic Regression with Interactions	0.9828617	0.9208489	0.9155837
Time Dummy Hedonic Polynomial Regression with Interactions	0.9883839	0.9184804	0.9342877
Regression Tree	0.9879692	0.9528261	0.9693565
Bagging Trees	0.9902542	0.9658194	0.9765517
Random Forest	0.9950604	0.9769152	0.9832115

Conclusion

New datasource of online data for household appliances and consumer electronics

Different price index methods need to be used for different product categories

« Low » churn product categories: GEKS-Jevons seems to be appropriate

« High » churn product categories with lifecycle effects require explicit quality adjustment methods

- Different price imputation methods give similar results
- Resource demanding

Future work: Analyse other product categories and splicing methods

STATEC

Institut national de la statistique
et des études économiques

Thank you!

13, rue Erasme
L-1468 Luxembourg

(+352) 247-84219

info@statec.etat.lu

Lucien May
lucien.may@statec.etat.lu

Botir Radjabov

botir.radjabov@ext.statec.etat.lu

statistiques.public.lu

@Statec Luxembourg **/STATEC** **@STATEC** **Statec Luxembourg**