



OESTERREICHISCHE NATIONALBANK
EUROSYSTEM

Discussion of the Paper “Scraped Data and Sticky Prices” by Alberto Cavallo

Fabio Rumler
Oesterreichische Nationalbank

*Conference on Price Setting and Sticky Prices
Banque de France, Paris, 17 December 2015*

My reading of “Scraped Data and Sticky Prices”

- Most studies so far use scanner price data or CPI micro price data to document price stickiness, e.g. IPN and Nielsen price data
- This paper uses **daily online price data** retrieved from supermarket websites (scraped data) in 5 countries (Argentina, Brazil, Chile, Columbia, USA) to re-estimate frequency and size of price changes and hazard functions
- **Scraped data can avoid the bias** in the calculation of standard price statistics induced by time averaging and imputations present in other data sources; however they are **less representative**
- **Results:** relative to other data sources, prices from scraped data change less frequently, show a bi-modal distribution of the size of price changes and a hump-shaped hazard function

Main comments

- **Scraped data have drawbacks too:**
 - few products and retailers covered, contain few imputations**
- **Biggest advantage:**
 - daily frequency, easily available for researchers**
- **Micro CPI data without imputations still give different results:**
 - Austrian CPI data from 1996 to 2009, 641 products, 6.7 million obs.**
 - Food and beverages: average duration of prices: 7 months, av. size of price changes: 19%, initially downward-sloping hazard function**
- **Most important difference is daily observation frequency:**
 - Drawing monthly observations from daily online dataset: implied durations double, average size of price changes drops**

Pros and cons of alternative data sources - reloaded

	Scraped data	CPI data	Scanner data
Frequency	daily	monthly	weekly / monthly
Products / retailers covered	few	many	intermediate
Time series	long	long	short
Time averaging	no	no	yes
Imputations	no	some	no
Censored spells	few	potentially many	few
Sales	yes	in some countries	yes
Coupons and loyalty cards	no	no	yes
Available for research	Yes	limited	limited

- **Scraped data are not representative of the full CPI basket**
- **Scanner data are time aggregated and therefore no true micro data**
- **CPI data are very rich, but sometimes data quality is bad**

CPI micro data without imputations

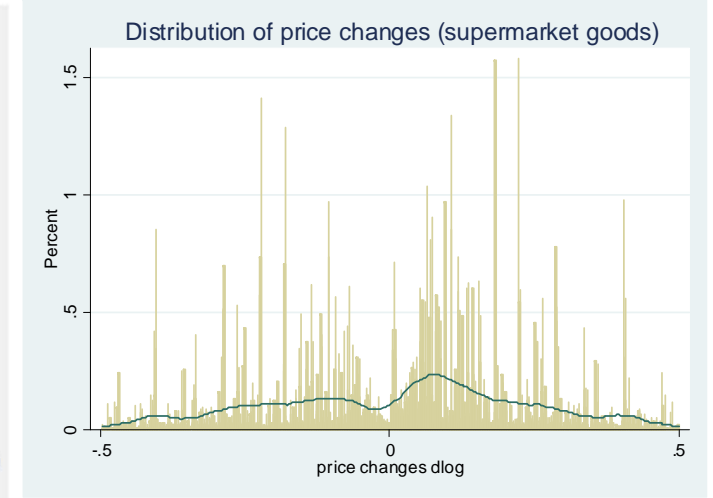
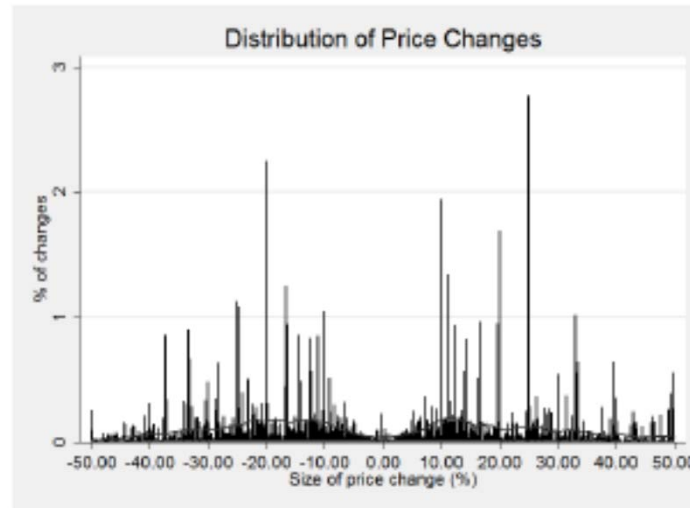
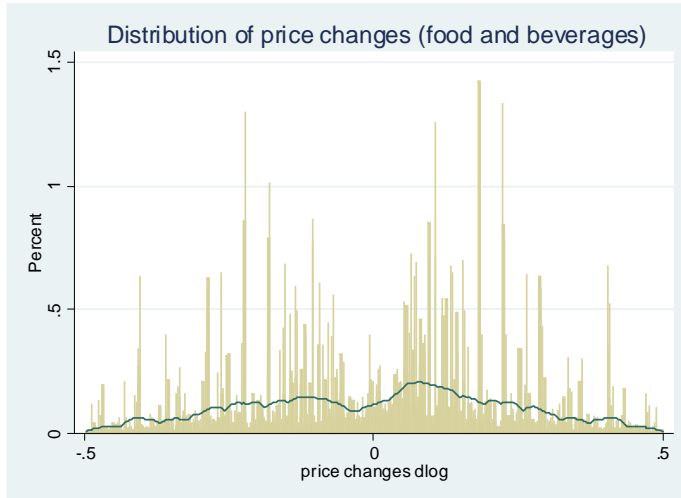
- In the Austrian CPI data all imputations in case of temporary unavailability are flagged, so they can be disregarded
- Dataset contains monthly price observations from Jan 1996 to Dec 2009 of 641 products and service items, total 6.7 million observations
- For comparability, use only (143) food and beverage items to calculate price stickiness statistics (2.5 million obs.) → see last column
- Average duration is about 7 months

	Argentina	Brazil	Chile	Colombia	USA	Austria
Inflation (% , average annual rate)	17.1%	5.1%	2.7%	4.2%	0.1%	1.9%
Price increases (% of price changes)	68%	57%	54%	55%	52%	54%
Price decreases (% of price changes)	32%	43%	46%	44%	48%	46%
Size of price increases (Mean*)	13%	12%	16%	11%	29%	18%
Size of price decreases (Mean*)	-13%	-12%	-14%	-10%	-22%	-20%
Daily Frequency	0.015	0.026	0.013	0.022	0.024	0.19 (monthly)
Implied Durations (days)	64	38	75	45	42	
Implied Durations (months)	2.1	1.3	2.5	1.5	1.4	6.9
Frequency of Increases (Freq+)	0.010	0.015	0.007	0.012	0.012	0.10 (monthly)
Frequency of Decreases (Freq-)	0.005	0.011	0.006	0.010	0.012	0.09 (monthly)
Freq+/Freq-	2.00	1.36	1.16	1.20	1.0	1.3

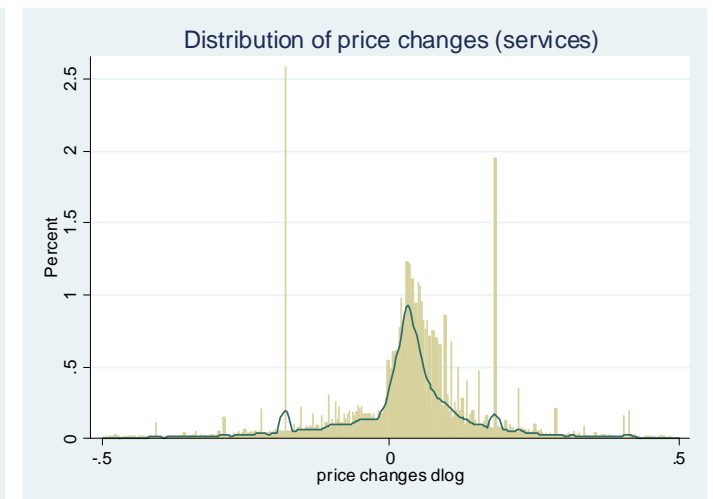
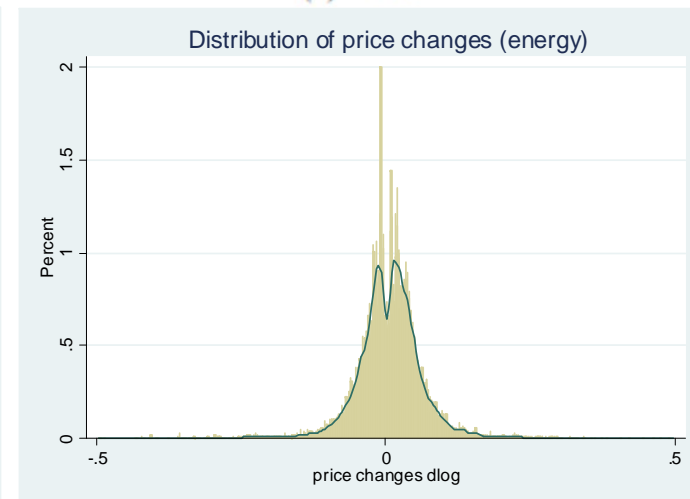
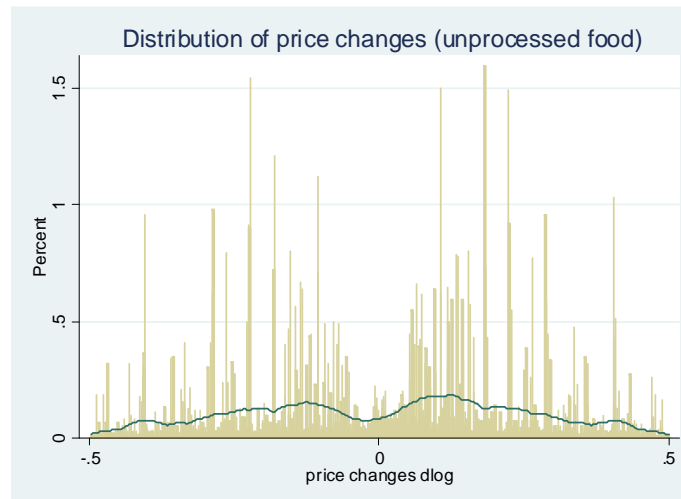
Notes: * Computed as the mean within categories, and then mean across all categories.

CPI micro data without imputations

- Distribution of the size of price changes very similar in Austrian CPI data without imputations and scraped data



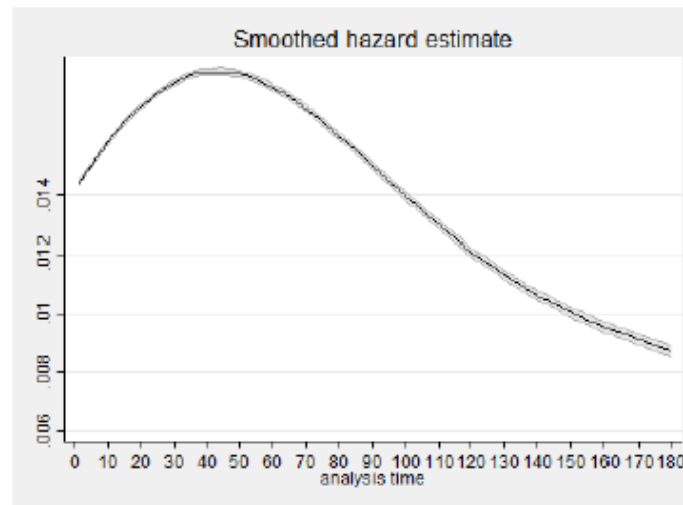
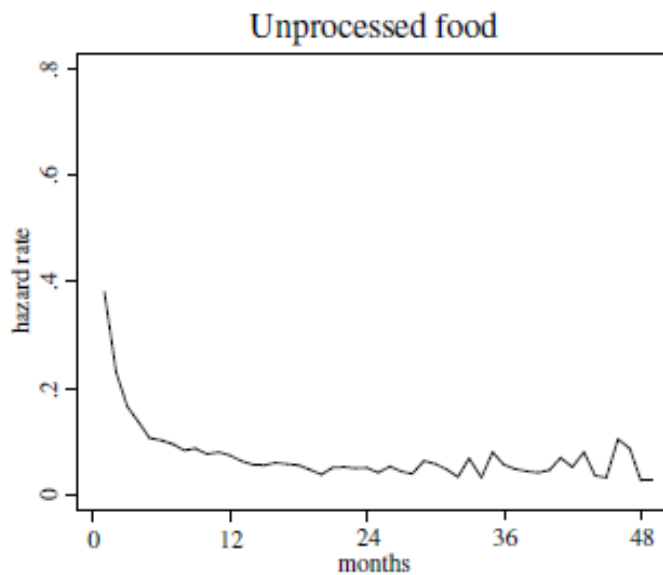
(e) USA



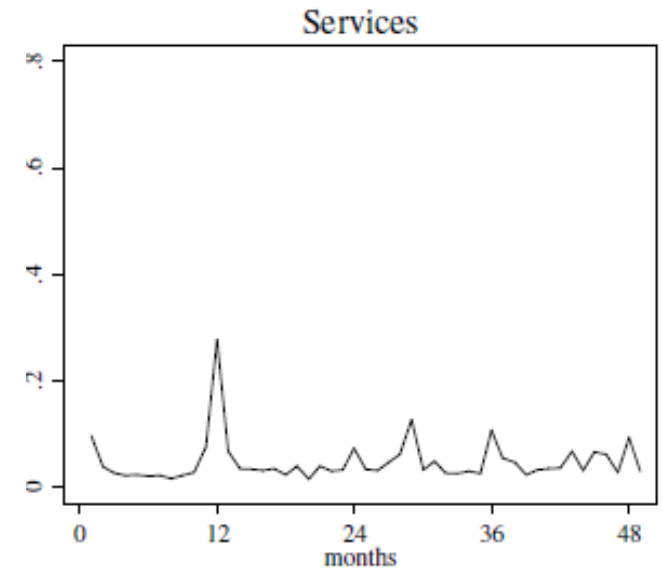
CPI micro data without imputations

- Hazard functions look very different:

Initially downward-sloping hazard function based on monthly food prices in Austria vs. hump-shaped hazard function for daily online prices in the US



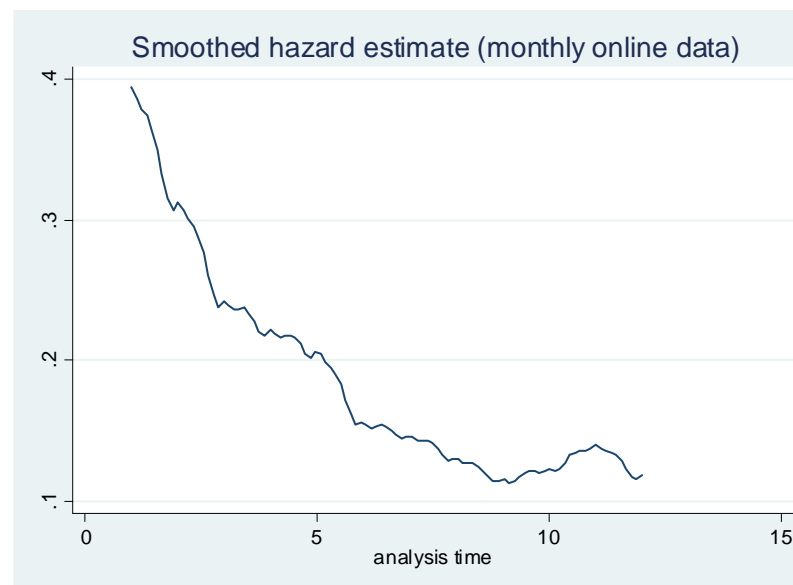
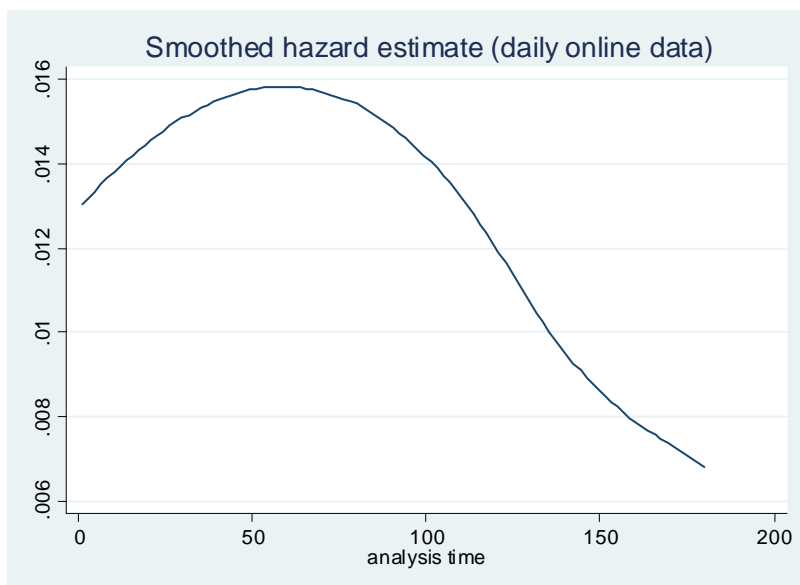
(e) USA



Frequency of observation is key

- Switching from daily to monthly observation in the same dataset (US online prices): implied duration of prices almost double, size of price changes declines

	Daily online data	Monthly online data
Frequency of price changes	0.024 (daily)	0.34 (monthly)
Implied durations (months)	1.4	2.4
Av. size of price increases (dlog)	25%	22%
Av. size of price decreases (dlog)	-27%	-25%



Conclusions

- **Main advantage of scraped data is daily frequency**
Frequency of price changes and durations can be measured more exactly from daily observations, hazard functions are more dynamic
- **Scraped data are useful to calculate high frequency price indices**
However scraped data have drawbacks too: low representativeness
- **Scraped data can abolish one source of measurement bias**
The bias induced by imputations is absent in scraped data, but not the bias of exclusive discounts (coupons, loyalty cards) → underestimate the frequency of price changes
- **For macro modelling: calibrated price stickiness in macro models depends on the periodicity of the model**
If the model is monthly, use monthly price durations, if it is quarterly, use quarterly durations
- **To assess real effects of monetary policy we need price stickiness measured from a representative goods basket**