

The Billion Prices Project

Using Online Data for Measurement and Research

Alberto Cavallo
MIT & NBER

Banco Central do Brasil - November 2017

Three related initiatives

INFLACION VERDADERA

Con el apoyo del Billion Prices Project @ MIT Sloan y Harvard Business School

www.inflacionverdadera.com

- Alternative inflation indices for Argentina (online) and Venezuela (crowdsourced)
- Started in 2007

THE BILLION PRICES PROJECT

MIT MANAGEMENT
SLOAN SCHOOL

www.thebillionpricesproject.com

- Academic initiative to use new data sources in macro and international research
- Started in 2008

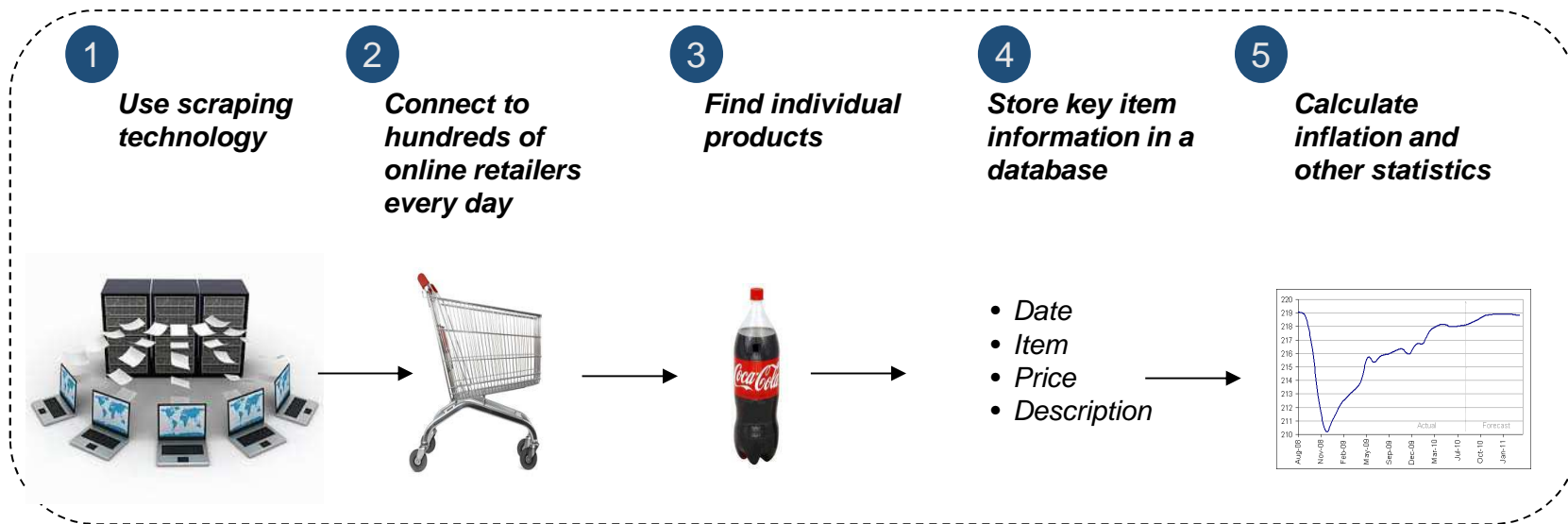
PriceStats[™]

www.pricestats.com

- Private company that collects daily price data from thousands of retailers that sell online, publishes daily inflation and PPP series, shares the micro data with the Billion Prices Project
- Started in 2011

The Billion Prices Project

- Academic initiative to use online price data for economic measurement and research
 - Co-founded with Roberto Rigobon
 - Daily price data since 2008
 - From hundreds of large multi-channel retailers
 - In over 60 countries

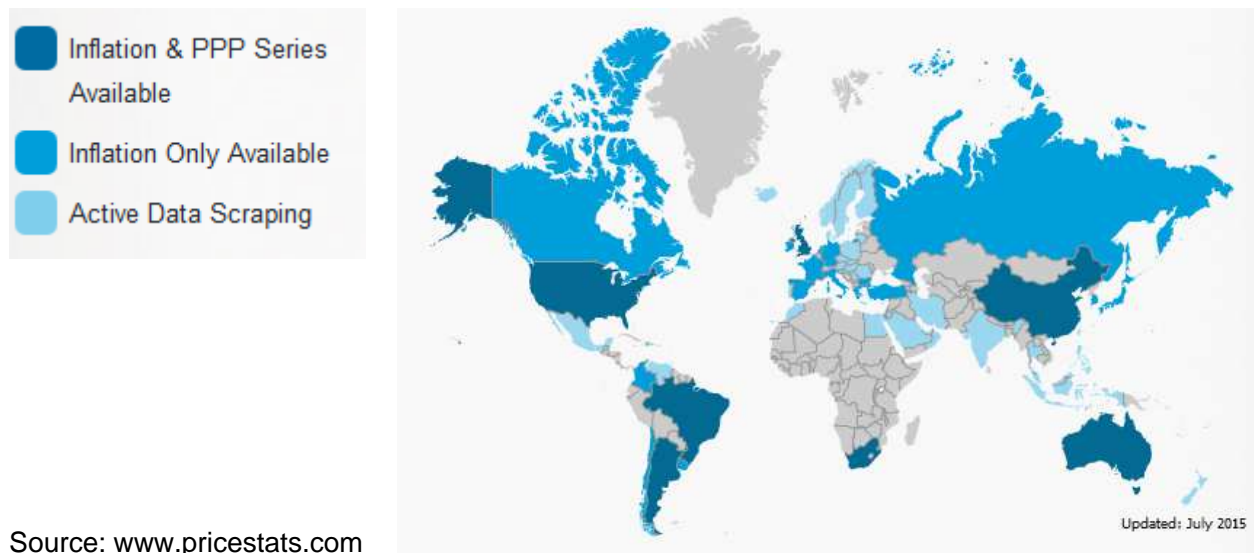


Uses and Research Applications

- Measurement of inflation and other price indicators
 - Online Data
 - Crowdsourced data
- Research in Macroeconomics
 - Price Dynamics & Stickiness
 - Consumer Inflation Expectations
- Research in International Economics
 - Real Exchange Rate Dynamics and Passthrough
 - High-frequency ICP → standards of living across countries

Daily Inflation Measurement in other countries

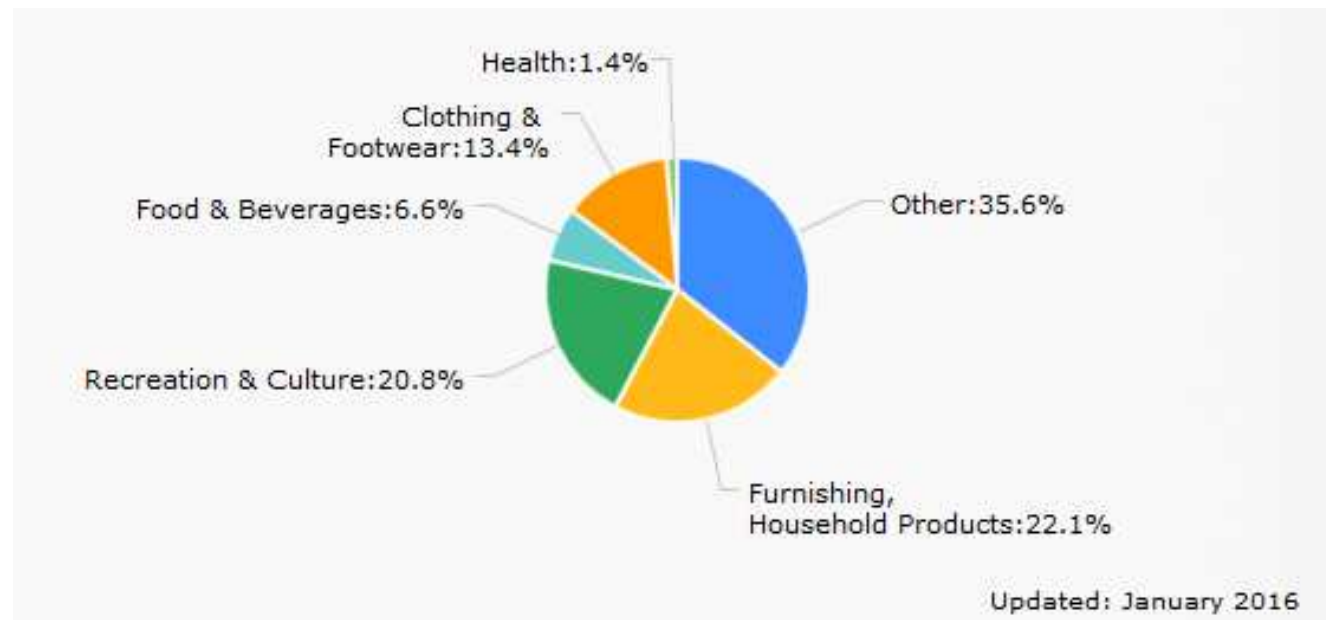
- In 2010 we started publishing a daily index for the US on the BPP website (bpp.mit.edu)
- In 2011 we founded PriceStats, which collects the data and publishes daily inflation and PPP indices in 22 countries in real-time. (www.pricestats.com)



Online Data and Daily Inflation

- On average, prices from about 60% of CPI expenditure weights can be found online
 - All kinds of goods, including food and fuel
 - Relatively few services (improving)
 - Housing/rents information online, but not directly included in our indices

Micro data



Source: www.pricestats.com

Online Price Indices vs Official CPIs

- Measurement (not prediction) of inflation with an alternative source of data
- Three main characteristics compared to CPIs
 - Short-run discrepancies (differences in data, methods)
 - Medium/long-term co-movement
 - Anticipation of changes in inflation trends

Lehman Brothers Bankruptcy

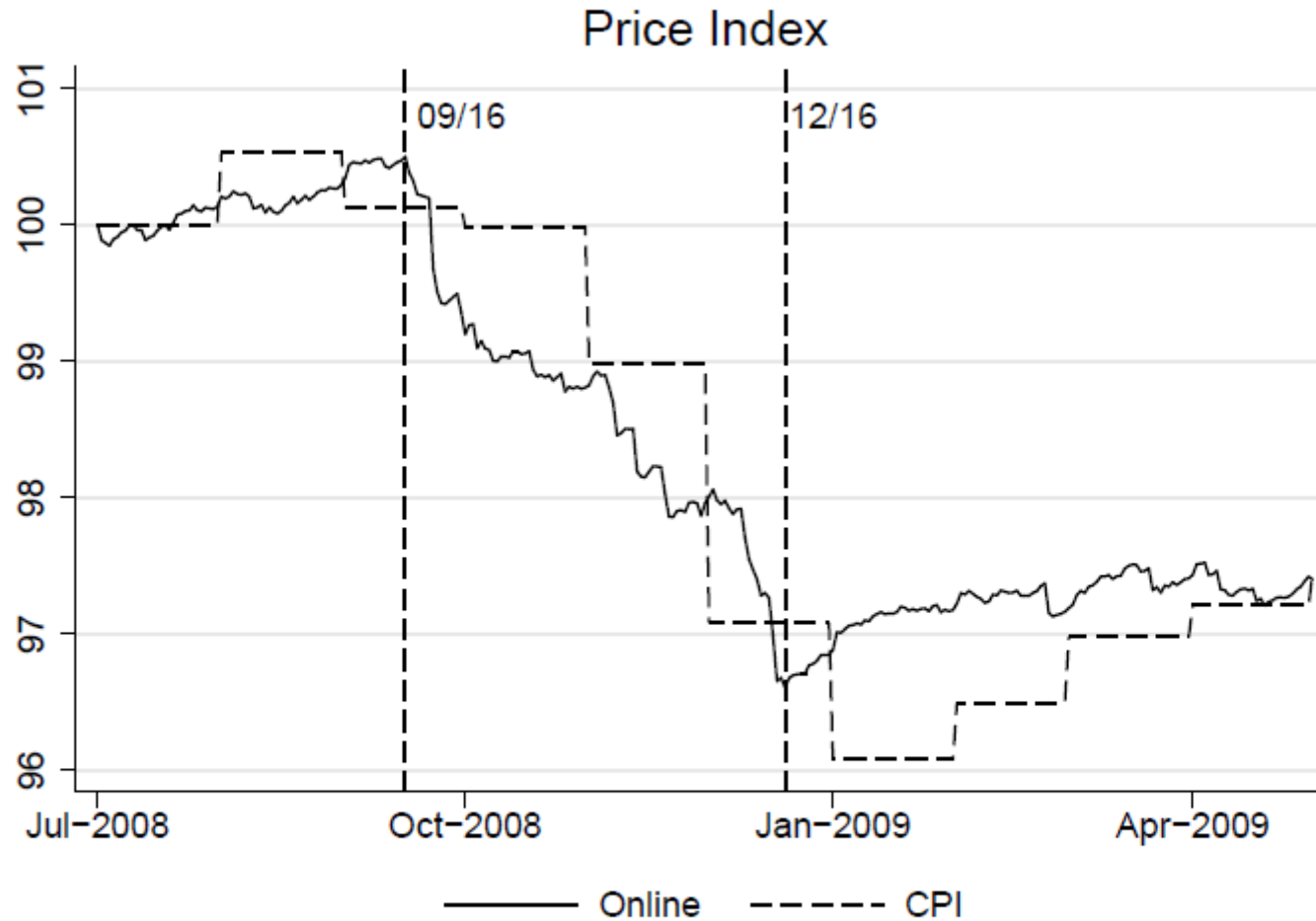
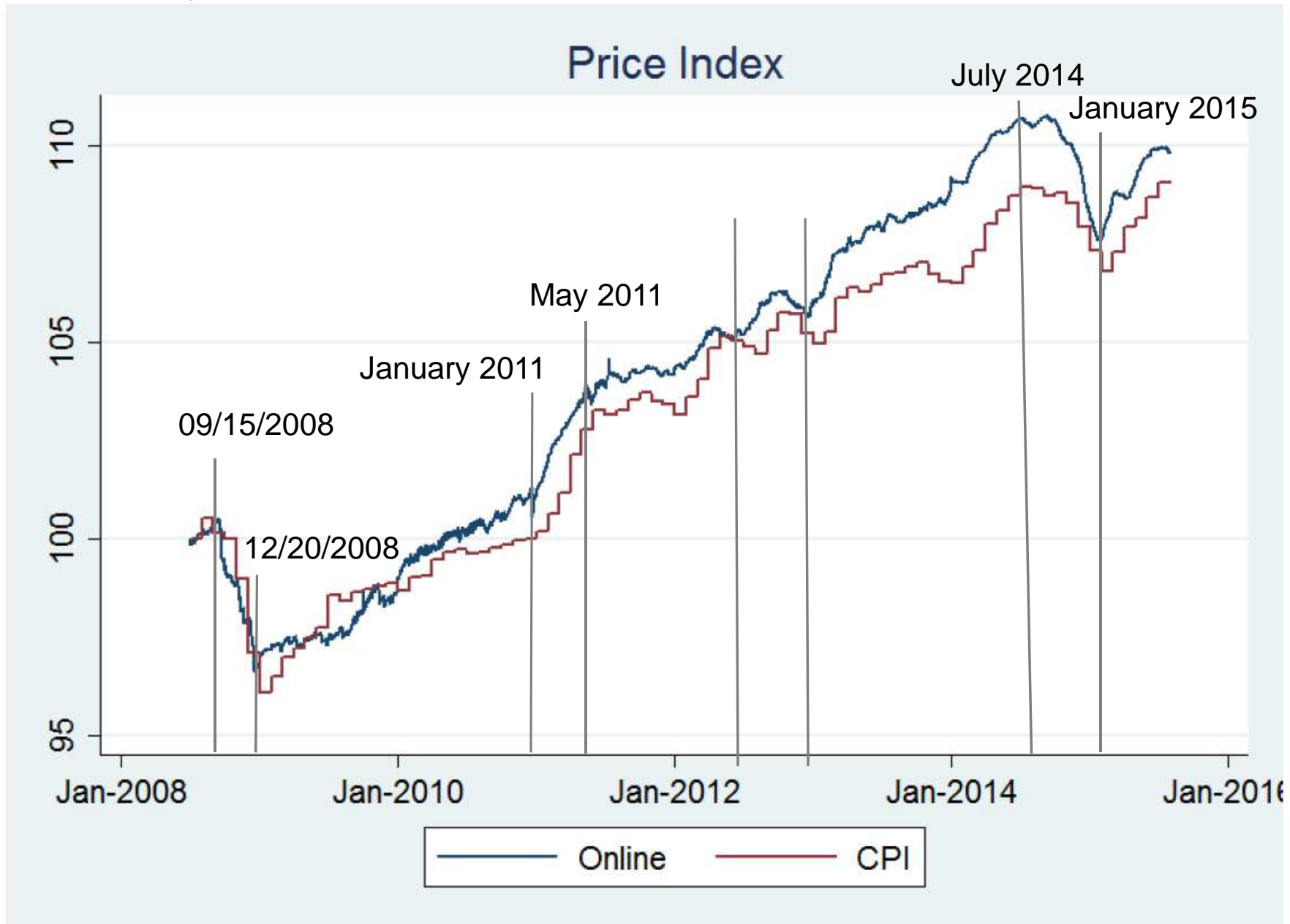


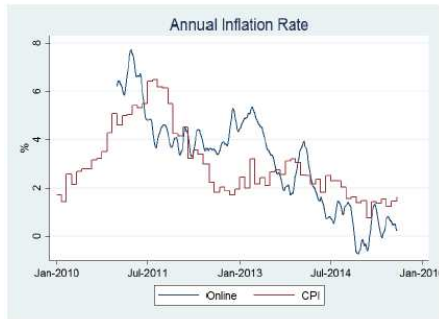
Figure 3: US Price Index around Lehman Brother's Bankruptcy

US Daily Price Index

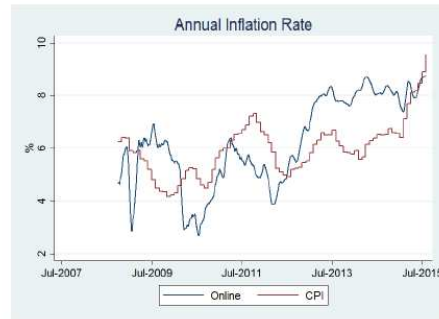


Source: Cavallo & Rigobon (2016) "The Billion Prices Project", Journal of Economic Perspectives, Spring 2016, Vol 30(2):151-78.

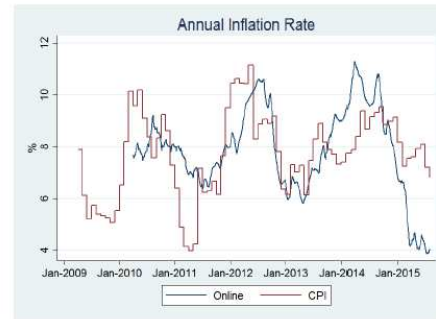
Developing vs Developed Countries



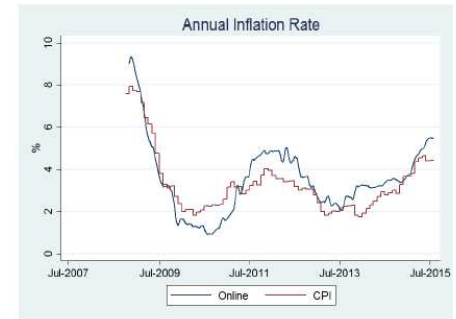
(a) China



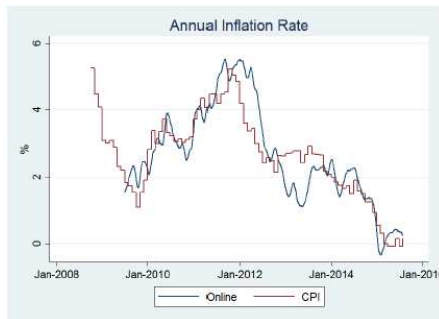
(b) Brazil



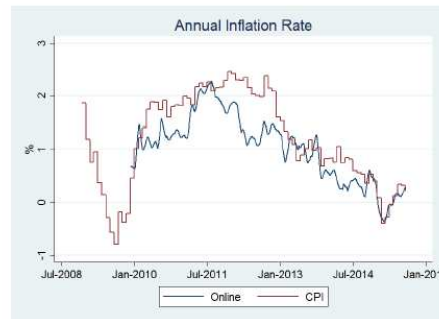
(c) Turkey



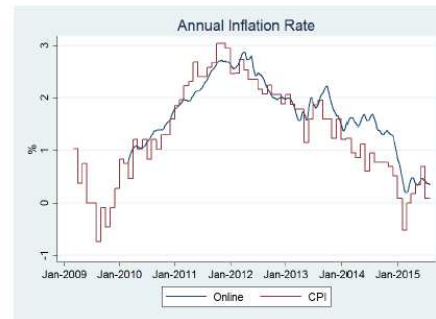
(d) Colombia



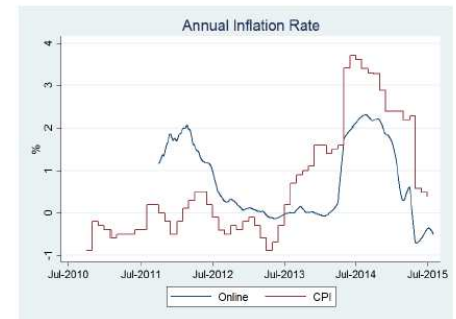
(e) UK



(f) France



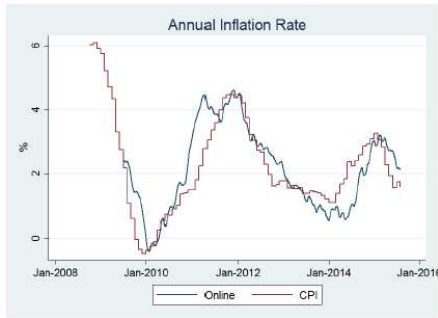
(g) Germany



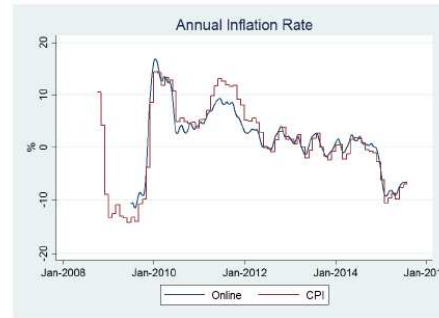
(h) Japan

Figure 5: Online vs CPI Annual Inflation Rates

Sectors vs Global Aggregates



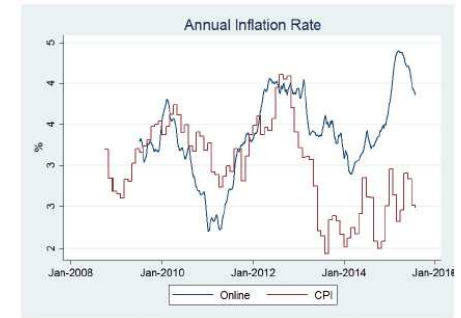
(i) USA Food



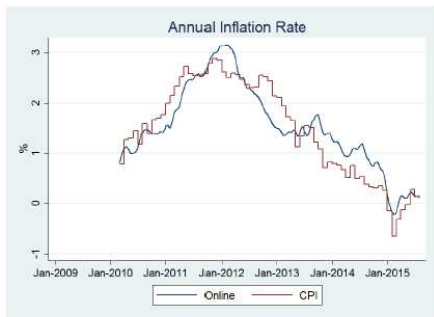
(j) USA Fuel



(k) USA Electronics



(l) USA Medical Care



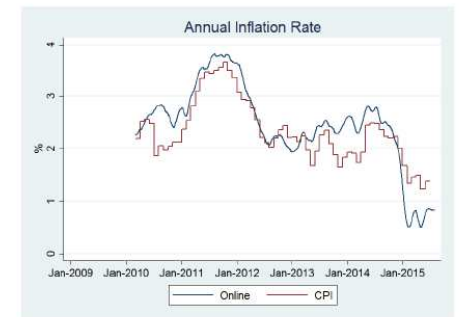
(m) Eurozone



(n) Developed M.



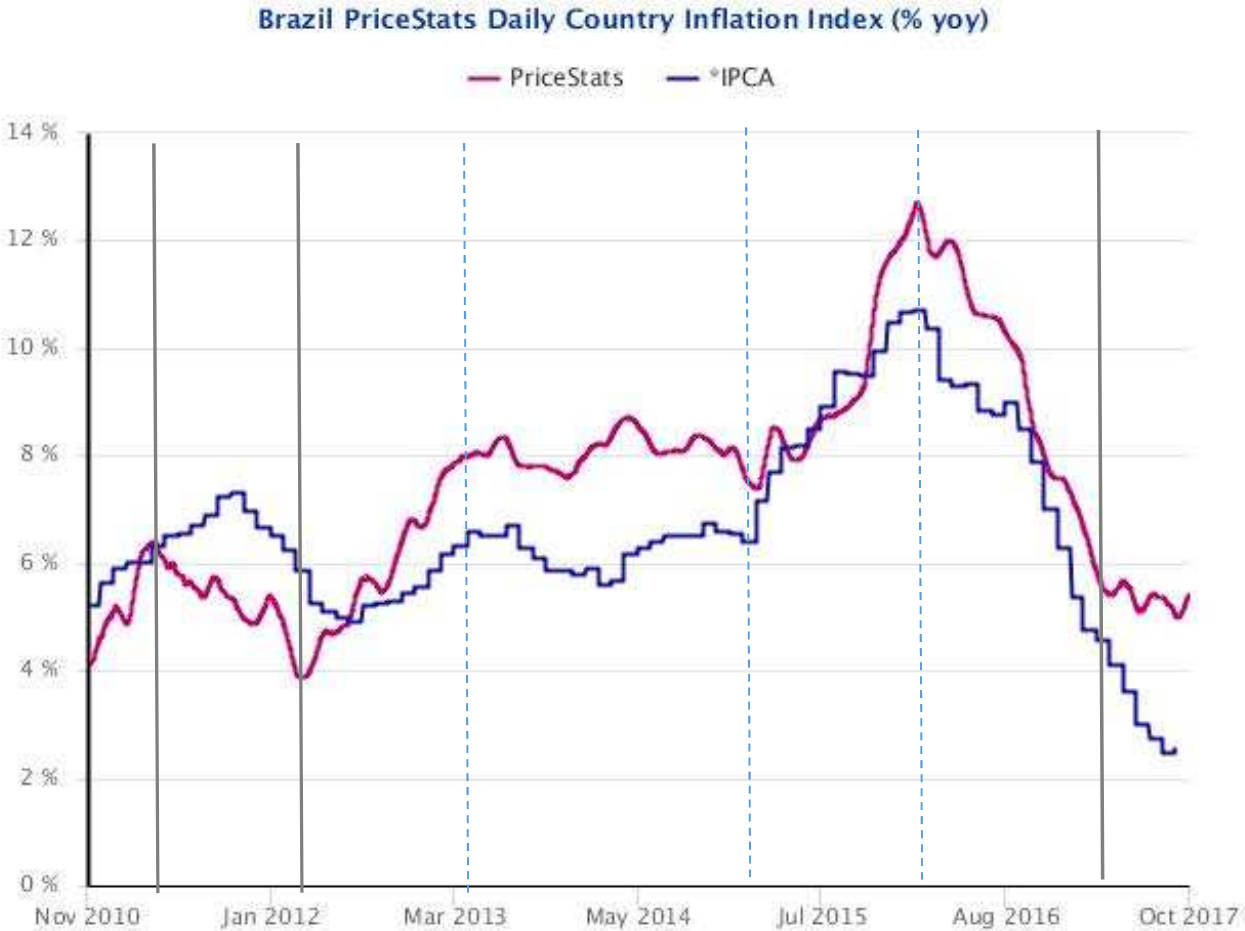
(o) Emerging M.



(p) World

Figure 5: Online vs CPI Annual Inflation Rates

Brazil Annual Inflation (updated last week)

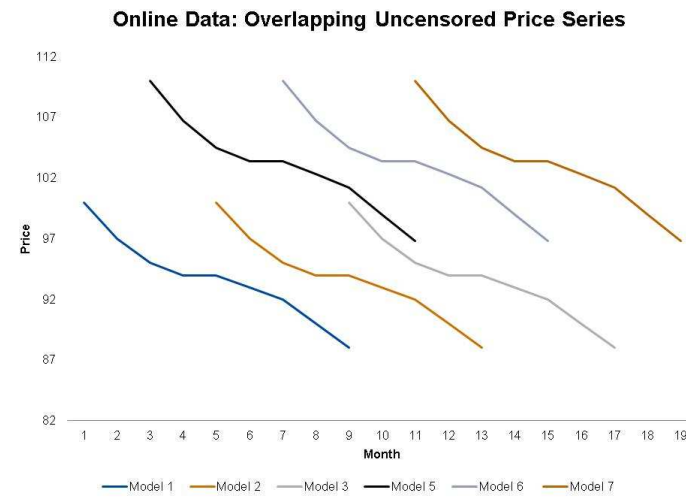
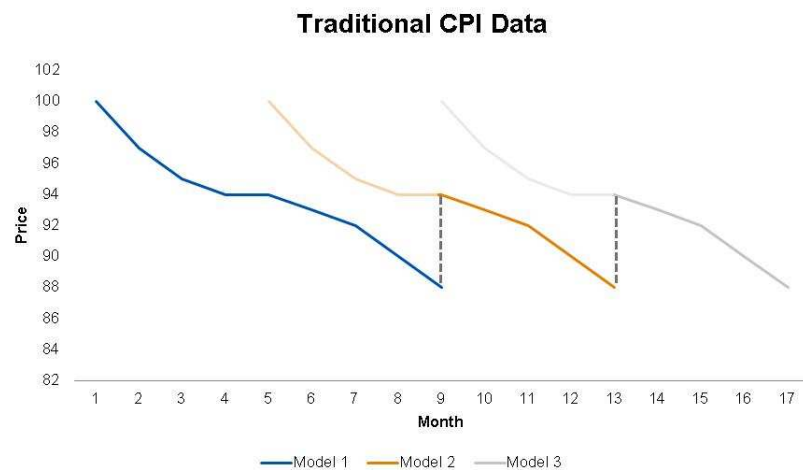


Source: State Street Global Markets Research

Date: 11/01/2010-10/29/2017

Differences with CPI : Quality Adjustments

- Many complex techniques applied in CPI methods, such as hedonic quality adjustments, are needed because the data has inherent limitations
- Online data has big data advantages:
 - uncensored spells (automatically included at introduction)
 - all varieties/models on display



Differences with CPI : Quality Adjustments

- Simple indices can approximate the level and trend of CPI inflation in hedonic-adjusted categories (as suggested in Silver & Heravi (99), Aizcorbe, Corrado & Doms (2003))

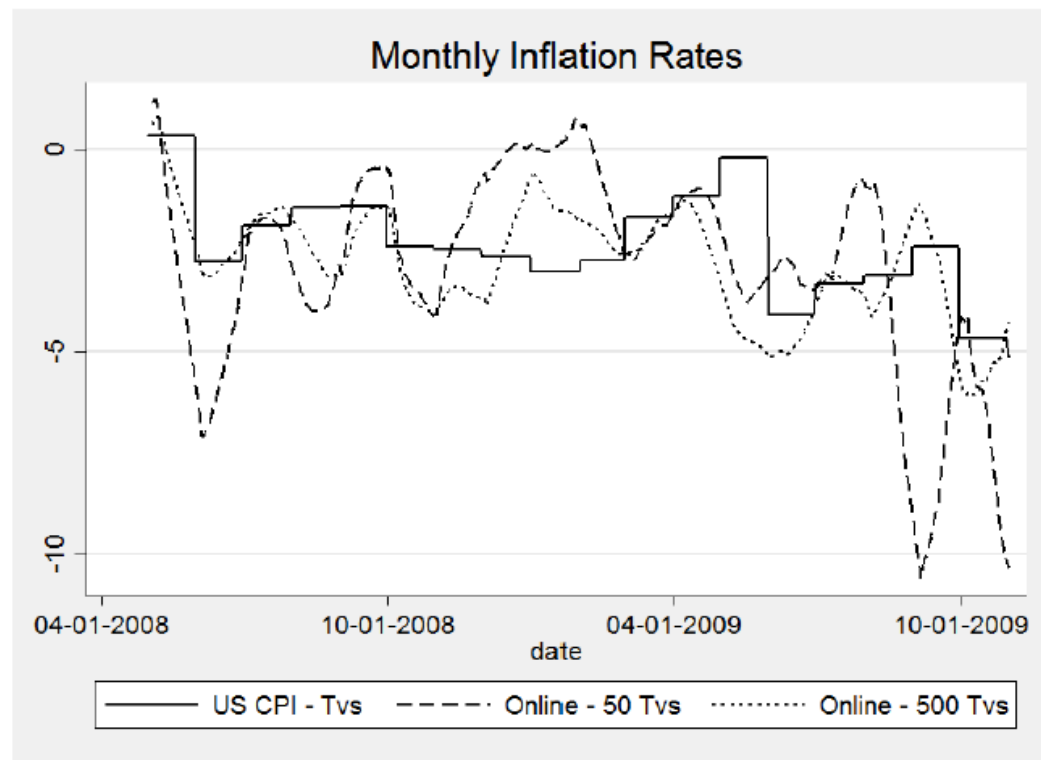


Figure 7: US CPI with Hedonics vs Online Jevons Index

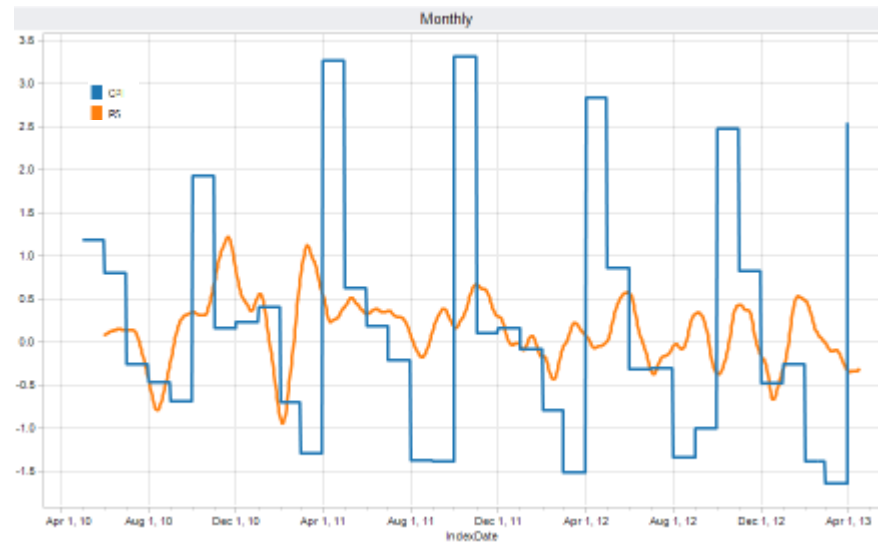
Differences with CPI : Quality Adjustments

- Simple indices can approximate the level and trend of CPI inflation in hedonic-adjusted categories (as suggested in Silver & Heravi (99), Aizcorbe, Corrado & Doms (2003))
- Implication: online series tend to be smoother than CPIs

Greece Price Index



Greece Monthly Inflation



Anticipation in Online Prices

- We can study the link between online data and CPIs using simple VARs and impulse responses.
- VAR regressions with Δ CPI on the LHS and lags of Δ CPI and Δ PS on the RHS (monthly data)

$$\Delta \ln(CPI_t) = \alpha + \beta \Delta \ln(Online_t) + \sum_{i \in [1,6]} \alpha_i \Delta \ln(CPI_{t-i}) + \sum_{i \in [1,6]} \beta_i \Delta \ln(Online_{t-i}) + \epsilon_t$$

- Impulse responses show the impact of a 1% shock in PriceStats on future CPI (reflecting additional information not contained in lagged CPI)

Anticipation in Online Prices

- Online prices tend to anticipate shocks in CPI inflation.

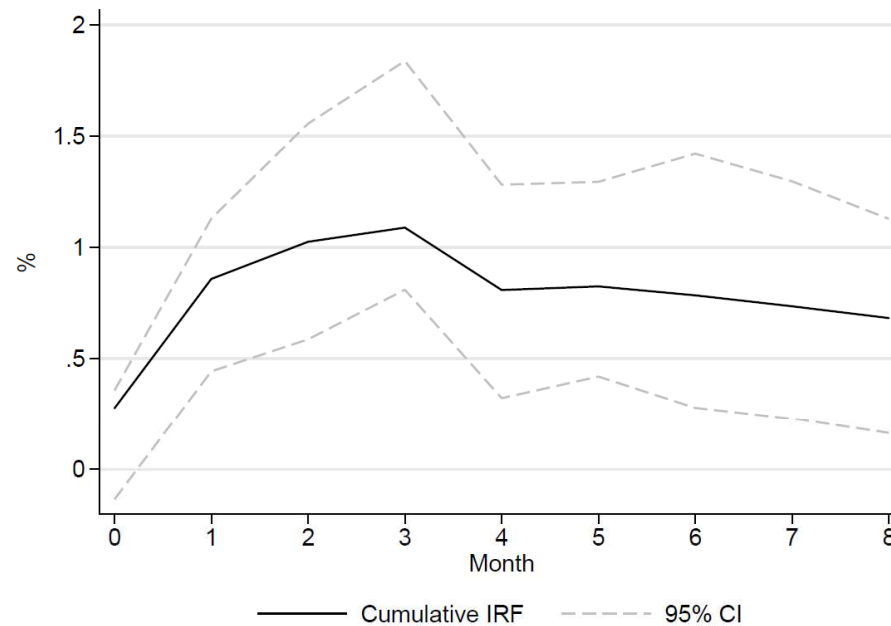


Figure 7: Impulse Response of the US CPI to a Online Price Index shock

Notes: Cumulative impulse response to a 1% shock in the online index. The online price index was computed by PriceStats. The CPI is a US city average, all items, non-seasonally adjusted from the Bureau of Labor Statistics. Data from July 2008 to January 2015.

Anticipation in Online Prices

- We consistently find that online prices tend to anticipate changes in inflation trends.
 - This goes beyond the ability to collect and publish the data faster
 - Online prices tend to *react faster* to shocks.
 - Why?
 - Lower adjustment (or menu) costs
 - Online shoppers may be less sensitive to price changes
 - More intense and transparent competition

Online vs Offline Prices

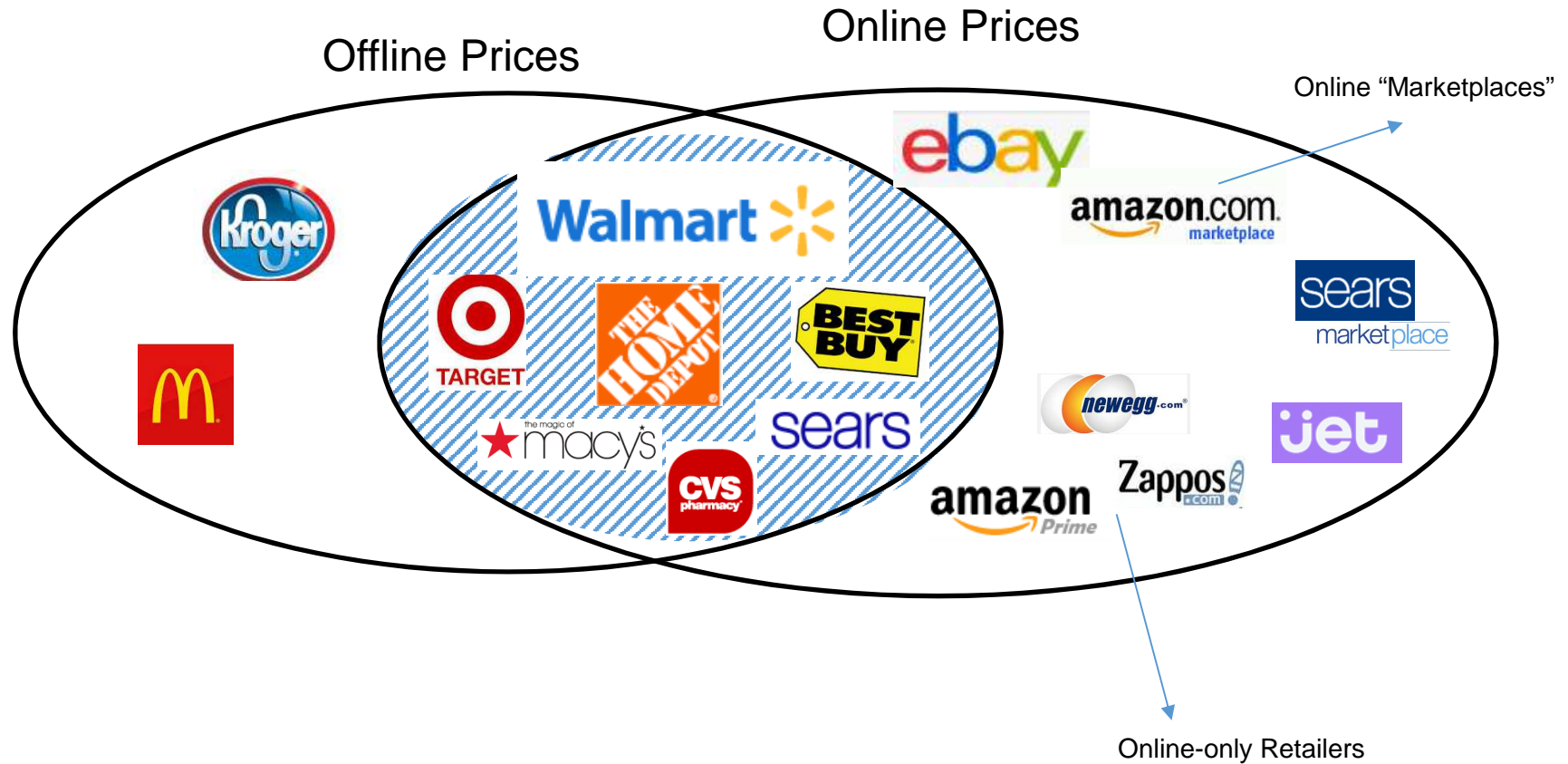
- Are online prices representative ?
 - Online sales are still less than 10% of retail sales in developed countries
- Suggestive evidence for yes :
 - Co-movement in online and CPI indices
 - The `online store` is effectively the *largest* store for most retailers. Eg: Walmart has 4759 stores in the US. The median store has 0.02% of sales. The `online store` has 8% of sales
- Still, for some applications, online prices may introduce their own measurement biases → what are they?

Cavallo (2017) Are Online and Offline Prices Similar? , AER forthcoming.

- Large-scale comparison of online and offline prices collected simultaneously in ~50 retailers in 10 countries.
- Key questions:
 - Are the price *levels* similar?
 - Are price *changes* similar in timing, frequency, and size?
 - Are offline goods also available online?
- Implications for NSOs, macro/international papers, IO and price dispersion.

Which Online Prices ?

Price Aggregators or Shopbots
Google product search pricegrabber



- I study retailers the sell both online and offline (multi-channel).

Online-Offline Data Collection

- Collect prices for:
 - a random set of goods
 - at the website and any physical store of a given retailer
 - at the same time (7-day window)
- 10 countries, 5-10 retailers in each
- Retailer included:
 1. Be a top 20 largest retailer by market share
 2. Sell both online and offline
 3. Possible for us to link online and offline products (via UPC or product id)

List of Retailers Included

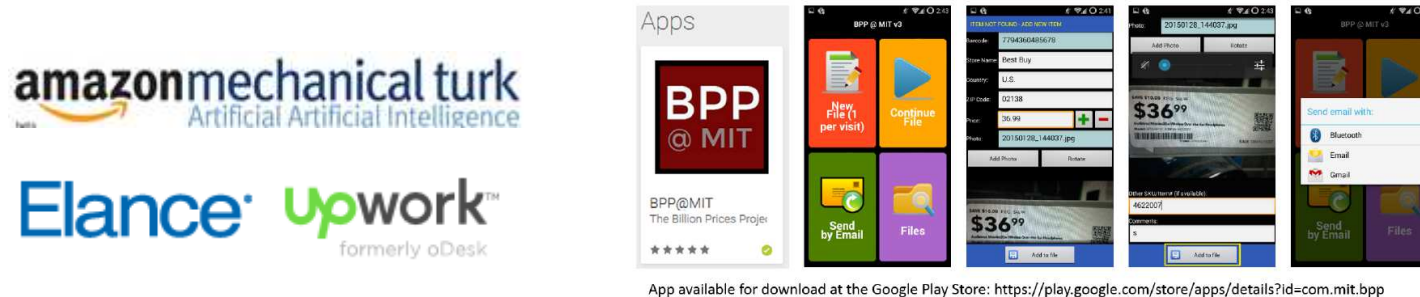
Table 1: List of Retailers

Country	Retailers Included
Argentina	Carrefour, Coto, Easy, Sodimac, Walmart
Australia	Coles, Masters, Target, WoolWorths
Brazil	Droga Raia, Extra, Magazine Luiza, Pao de Azucar, Renner
Canada	Canadian Tire, Home Depot, The Source, Toys R Us, Walmart
China	Auchan Drive, Sams Club
Germany	Galeria Kaufhof, Obi, Real, Rewe, Saturn
Japan	Bic Camera, K's Denki, Lawson, Yamada
South Africa	Clicks, Dis-Chem Pharmacy, Mr Price, Pick n Pay, Woolworths
UK	Asda, Mark and Spencer, Sainsburys, Tesco
USA	Banana Republic, Best Buy, CVS, Forever 21, GAP, Home Depot, Lowe's, Macys, Nike, Office Max/Depot, Old Navy, Safeway, Staples, StopandShop, Target, Urban Outfitters, Walmart

Note: Results updated on 23 Mar 2016.

Simultaneous Offline and Online Data Collection

- Crowdsourced workers scan barcodes, enter prices, and email data files



App available for download at the Google Play Store: <https://play.google.com/store/apps/details?id=com.mit.bpp>

Figure 1: Screenshots from BPP App for Android Phones

- We automatically process and consolidate the offline data

The screenshot shows an Excel spreadsheet with the following data:

	A	B	C	D	E	F	H	I	J
1	DEVICE ID	DATE	TIME	BARCODE	STORE NAME	STORE LOCATION	PRICE	PHOTO	COMMENTS
2	45c653f06cc750a8	10/20/2014	12:11	98071000050369	home depot	andrew	19.98	20141020_121214.jpg	
3	45c653f06cc750a8	10/20/2014	12:12	9807390862	home depot	andrew	7.98	20141020_121310.jpg	
4	45c653f06cc750a8	10/20/2014	12:13	9807203925	home depot	andrew	179	20141020_121357.jpg	
5	45c653f06cc750a8	10/20/2014	12:15	01178841	home depot	andrew	34.97	20141020_121558.jpg	

- We use the barcodes to check prices online using customized web scrapers

Data Collected

Table 2: Data by Country

Country	(1) Ret.	(2) Start	(3) End	(4) Days	(5) Workers	(6) Products	(7) Obs.
Argentina	5	02/15	08/15	71	18	2324	3699
Australia	4	03/15	08/15	64	13	3073	3797
Brazil	5	05/15	03/16	53	18	1437	1915
Canada	5	12/14	07/15	88	15	2658	4031
China	2	07/15	03/16	16	5	410	513
Germany	5	03/15	03/16	50	9	1215	1604
Japan	4	04/15	03/16	66	7	1127	2186
South Africa	5	03/15	03/16	80	21	2336	3212
UK	4	03/15	05/15	39	12	1661	2094
USA	17	12/14	03/16	222	206	7898	15332
ALL	56	12/14	03/16	383	323	24132	38383

Note: Results updated on 5 Apr 2016. Column 1 has the number of retailers. Columns 2 and 3 have the start and end months of data collection. Columns 4 and 5 report the number of days with data and workers that collected offline prices. Columns 6 and 7 provide the number of products and price observations that could be matched with both online and offline information.

Price Levels Country

Table 3: Country - Level Differences

Country	(1) Ret.	(2) Obs	(3) Identical (%)	(4) High On (%)	(5) Low On (%)	(6) Markup (%)	(7) Difference (%)
Argentina	5	3699	60	27	13	3	1
Australia	4	3797	74	20	5	5	1
Brazil	5	1915	42	18	40	-7	-4
Canada	5	4031	91	3	5	-5	0
China	2	513	87	7	6	3	0
Germany	5	1604	74	4	23	-8	-2
Japan	4	2186	48	7	45	-13	-7
South Africa	5	3212	85	6	9	-3	-1
UK	4	2094	91	2	7	-8	-1
USA	17	15332	69	8	22	-5	-1
ALL	56	38383	72	11	18	-4	-1

Note: Results updated 5 Apr 2016. Column 3 shows the percentage of observations that have identical online and offline prices. Column 4 has the percent of observation where prices are higher online and column 5 the percentage of price that are lower online. Column 6, is the online markup, defined as the average price difference excluding cases that are identical. Column 7 is the average price difference including identical prices.

Sector Differences

Table 4: Sector - Price Level Differences

Sector	(1) Ret.	(2) Obs	(3) Identical (%)	(4) High On (%)	(5) Low On (%)	(6) Markup (%)	(7) Difference (%)
Food	10	5953	52	32	15	3	1
Clothing	7	2534	92	5	3	3	0
Household	9	7875	79	5	16	-8	-2
Drugstore	4	3053	38	11	52	-5	-3
Electronics	5	3712	83	4	13	-9	-1
Office	2	1089	25	37	38	1	1
Multiple/Mix	18	14149	80	5	15	-9	-2

Note: Results updated 5 Apr 2016. Markup excludes identical prices. Difference includes identical prices.

- Prices more similar in Clothing and Electronics.
- Prices less similar in Drugstores and Office Supply retailers.

Online-Offline Price Comparison

- Price changes have different timing, but similar frequency and size

Table 6: Country - Price Change Frequency and Size

	(1) Price Changes	(2) Mean Freq. Online	(3) Mean Freq. Offline	(4) t-test p-val	(5) Mean Abs Size Online	(6) Mean Abs Size Offline	(7) t-test p-val
Argentina	247	.137	.147	.52	13.61	12.74	.67
Australia	73	.056	.092	.06	45.76	42.18	.62
Brazil	85	.167	.138	.36	10.55	9.36	.53
Canada	120	.077	.068	.48	31.11	21.71	.06
Germany	16	.035	.041	.74	27.08	15.86	.26
Japan	98	.074	.014	0	12.1	8.2	.34
South Africa	108	.1	.075	.13	23.33	16.75	.1
UK	26	.046	.072	.25	47.68	39.99	.58
USA	562	.052	.047	.34	23.78	21.38	.22
ALL	1330	.076	.068	.07	22.01	19.99	.11

Note: Results updated 18 Mar 2016.

Is there online price discrimination?

- Do prices vary online based on ip-location or browsing habits?
- Two simulations with US retailers:



1. IP-based pricing: we scraped the same products from a group of 12 different ip-address locations



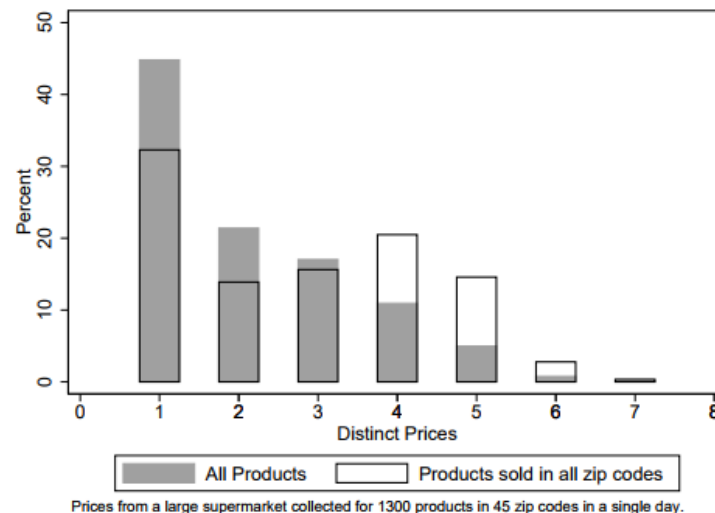
2. Persistent browsing habits: we scraped same products, every five minutes, for a full day.

- Result: Prices were always the same

Uniform Pricing *within* Retailers

- Most retailers sell online at a single price, regardless of the location of the buyer
 - Shipping costs are also becoming the same
 - Also evidence that large firms use uniform-pricing offline [DellaVigna & Gentzkow (2017), Aten & Rigbi (2017)]
- Even grocery retailers, which ask for zip codes before showing prices, tend to price identically across locations

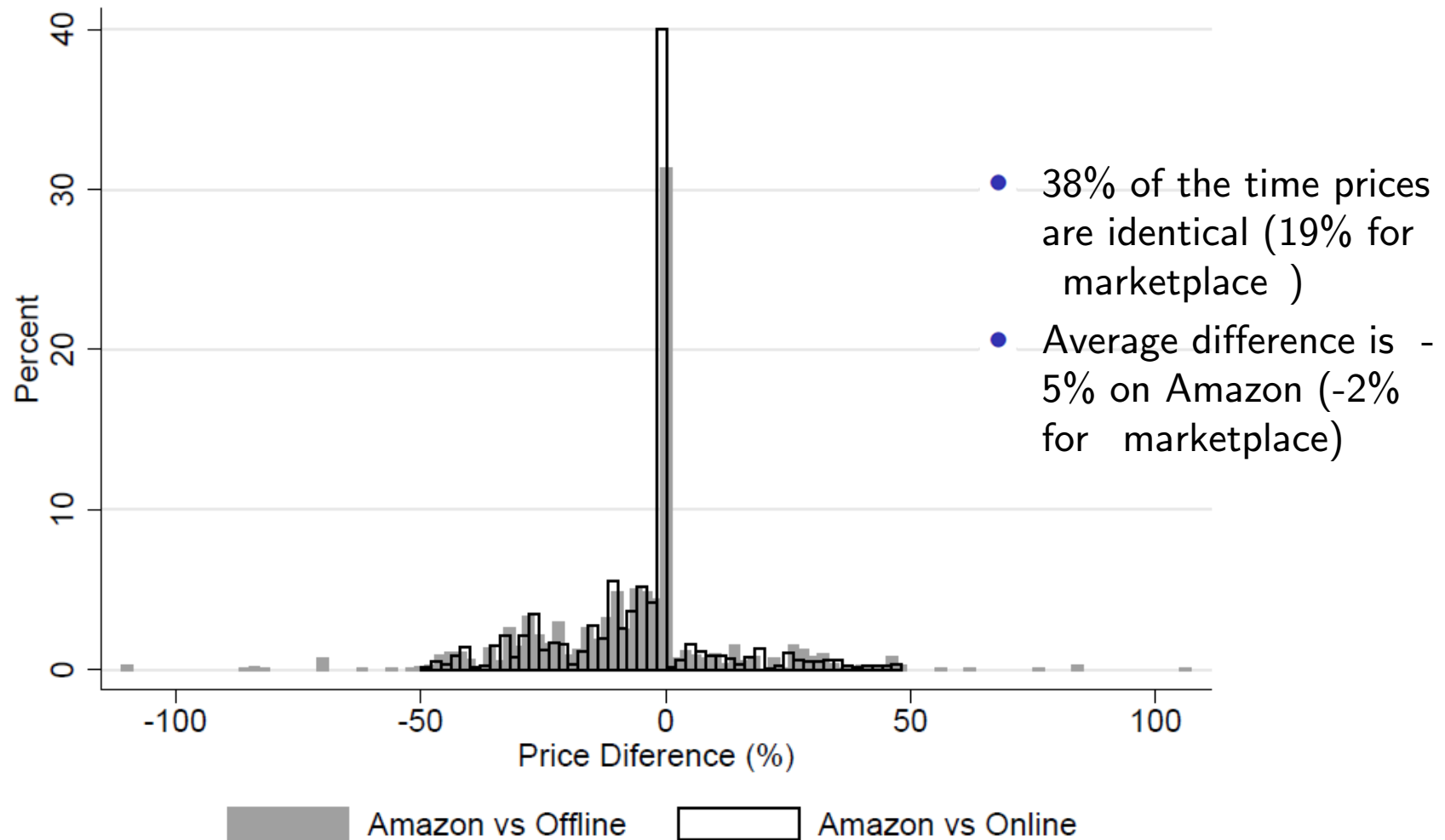
Figure A2: Supermarket Products with Distinct Prices



Source: Cavallo (2017) "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers", *Online Appendix*, American Economic Review Vol 107(1)

Compared to Amazon.com?

FIGURE 3. PRICE DIFFERENCES WITH AMAZON.COM (US ONLY)



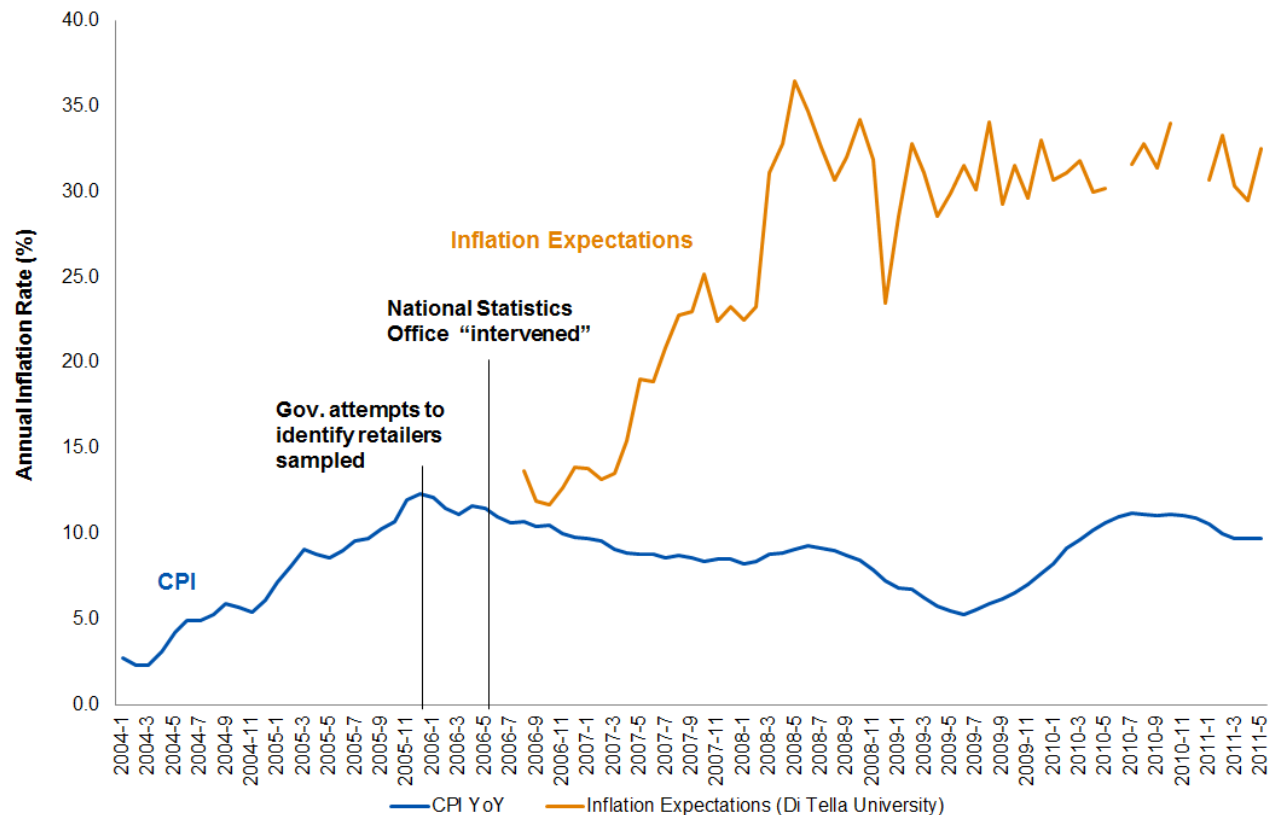
Other Research Applications

- International Relative Prices (this afternoon and tomorrow)
 - Real Exchange Rates and PPP dynamics
 - High-frequency standard of living
- Macro (tomorrow)
 - Price Stickiness
 - Inflation expectations
 - Crowdsourced inflation



How we got started: Argentina's inflation mystery

- Argentina's inflation data (2007-2015) was widely questioned
 - National Statistical Office (INDEC) intervened in 2007
 - CPI inflation remained below 10% for many years
 - Survey inflation expectations consistently above 25%



Each Data Source has Advantages and Disadvantages

Table 1: Alternative Micro-Price Data Sources

	Online Data	Scanner Data	CPI Data
Cost per observation	Low	Medium	High
Data Frequency	Daily	Weekly	Monthly
All Products in Retailer (Census)	Yes	No	No
Uncensored Price Spells	Yes	Yes	No
Countries with Research Data*	~60	<10	~20
Comparable Across Countries	Yes	Limited	Limited
Real-Time availability	Yes	No	No
Product Categories Covered	Few	Few	Many
Retailers Covered	Few	Few	Many
Quantities or Expenditure Weights	No	Yes	Yes

Note: Table from Cavallo (2015b). *Approximate numbers. The Billion Prices Project (bpp.mit.edu) datasets contain information from over 60 countries with varying degrees of sector coverage. Nielsen US scanner datasets are available at the Kilts Center for Marketing of the University of Chicago. Klenow and Malin (2010) provide stickiness results with CPI data sourced from 27 papers in 23 countries. See Cavallo (2013) for more details.

Source: Cavallo & Rigobon (2016) "The Billion Prices Project", Journal of Economic Perspectives, Spring 2016, Vol 30(2):151-78.