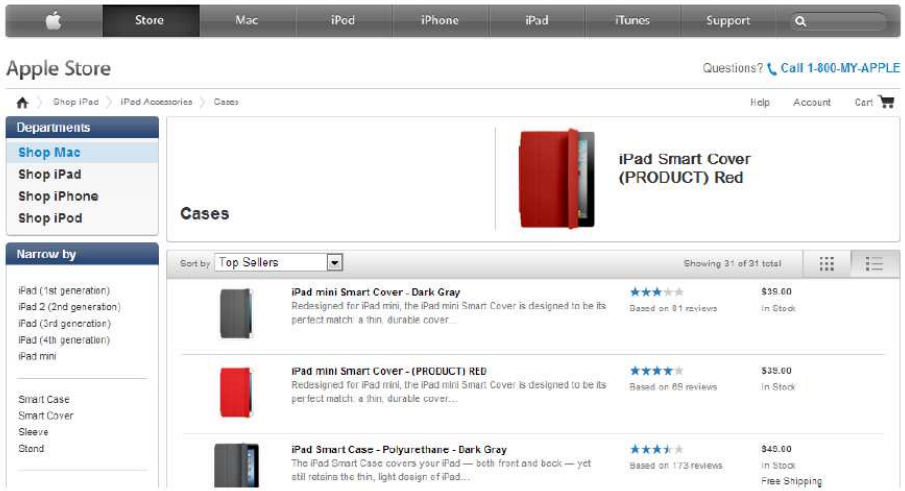
Publicly-Shared Datasets from The Billion Prices Project

Alberto Cavallo
MIT & NBER

IFM Data Project
NBER Summer Institute 2015

BPP Online Data

- Online prices collected
  - Directly from each retailer s website
  - Largest retailers by market share in each country
  - Multi-channel retailers → sell offline and online




```
<html>
<!-- START product -->
<ahref="productId=MD963LL"></a>
<p class="productname">Ipad Mini Smart Cover – Dark Grey</p>
<td class="Price">$39.00</td>
<!-- END product -->
.....
```

Every day, a software downloads a list of webpages, analyses the HTML code, extracts price data, and stores it in a database

Id	Product	Price	Date
MD963LL	Ipad Mini Smat Cover - Dark Grey	39	07/01/2012
MD955DD	IPad mini Smart Cover - (PRODUCT) RED	39	07/01/2012
MD48FSS	IPad Smart Case - Polyurethane - Dark Gray	49	07/01/2012

Online Price Data has Advantages and Disadvantages

Advantages	Disadvantages

• Cheaper to collect
• Frequency (daily)
• Granularity
  • All product details (brands, size, etc)
  • All goods and varieties available for sale (census)
  • New goods automatically sampled
• No time-averages, imputations, adjustments, or any kind of third-party (data collector) interference.
• Easier to compare internationally
	• Not all categories of goods and services are online
	• Fewer retailers and locations than CPI
	• Short time series
	• Online and Offline prices may behave differently

Publicly-Shared Datasets

1. Supermarket Data:

Daily prices for all goods sold by some of the largest Supermarkets in Latin America and the US

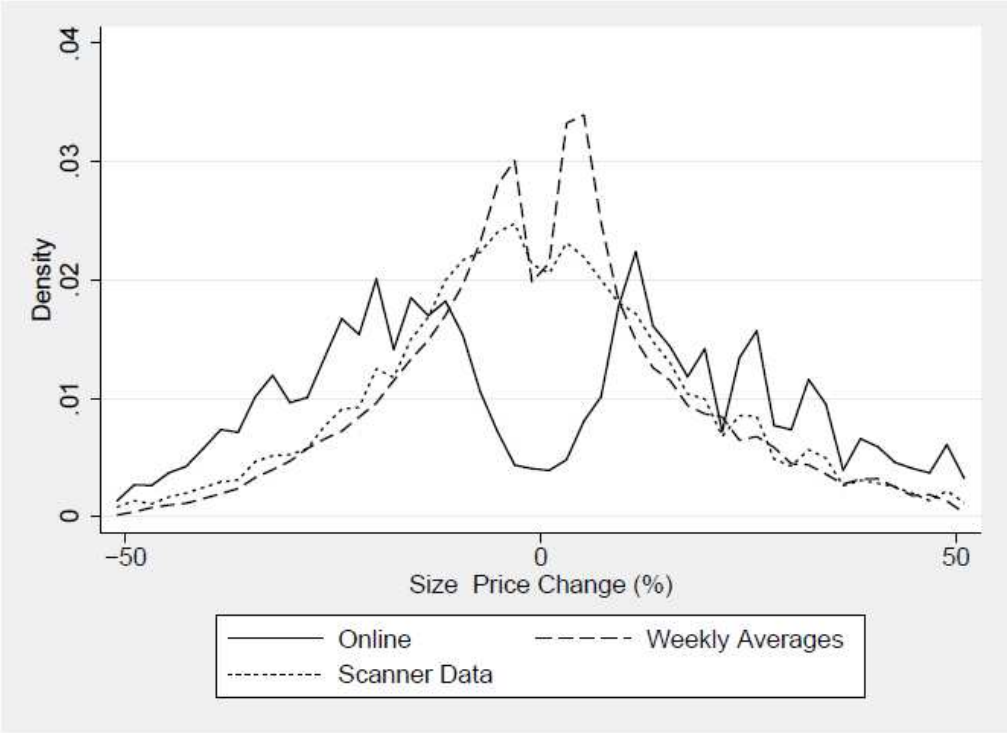
- Frequency: Daily
- Retailers: 7
- Countries: 6 (Argentina, Brazil, Chile, Colombia, Venezuela, and the US)
- Dates : 2007 to 2011
- Main variables: product id, date, price, sale indicator, category

id	date	price	sale	bppcat	cat_url	category	miss	fullprice	day	month	year	bulkquant	bulkprice	f
143477	17mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	17	3	2009	15	.182	
143477	18mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	18	3	2009	15	.182	
143477	19mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	19	3	2009	15	.182	
143477	20mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	20	3	2009	15	.182	
143477	21mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	21	3	2009	15	.182	
143477	22mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	22	3	2009	15	.182	
143477	23mar2009	1.99	1	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	1.99	23	3	2009	15	.1326667	
143477	24mar2009	1.99	1	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	1.99	24	3	2009	15	.1326667	
143477	25mar2009	1.99	1	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	1.99	25	3	2009	15	.1326667	
143477	26mar2009	1.99	1	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	1.99	26	3	2009	15	.1326667	
143477	27mar2009	1.99	1	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	1.99	27	3	2009	15	.1326667	
143477	28mar2009	1.99	1	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	1.99	28	3	2009	15	.1326667	
143477	29mar2009	1.99	1	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	1.99	29	3	2009	15	.1326667	
143477	30mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	30	3	2009	15	.182	
143477	31mar2009	2.73	0	133	1233	NONFROZEN NONCARBONATED JUICES AND DRINKS	.	2.73	31	3	2009	15	.182	

Example: Argentina. Cavallo (2013) "Online vs Official Price Indexes: Measuring Argentina's Inflation" - ***Journal of Monetary Economics*** *60(2), 152-165.*

Examples of Stylized Facts

- The Size of Price Changes



(e) USA

Figure 2: Effects of Weekly Averages and Scanner Data
Notes: The online and scanner data in the US was collected at the same retailer during the same time period. Scanner data collected by Nielsen and provided by the Kilts Center at Chicago Booth.

From Cavallo (2012) "Scraped Data and Sticky Prices"

Publicly-shared Datasets

2. Global Retailer Data

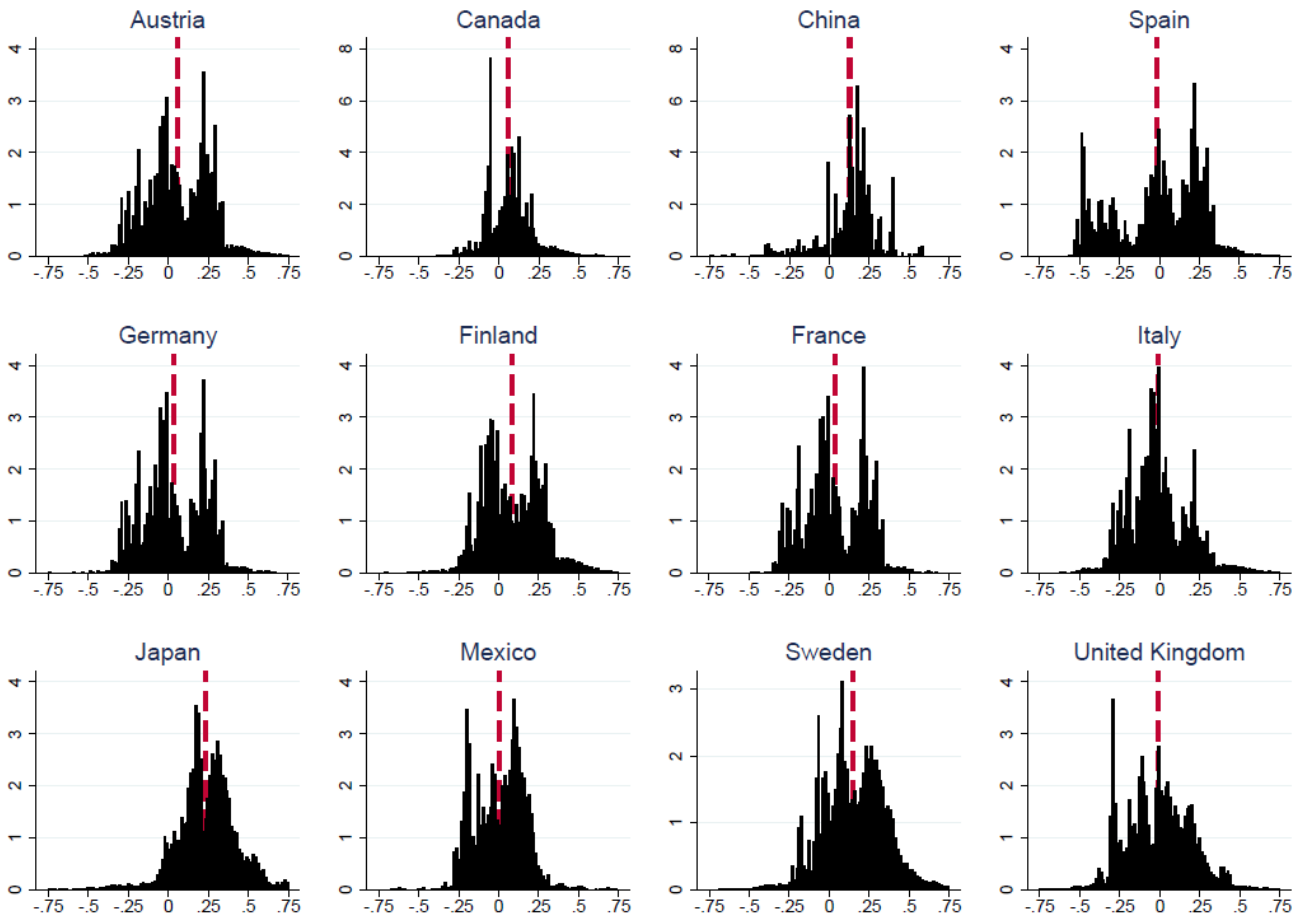
- Daily prices for all goods sold by APPLE, IKEA, ZARA, and H&M.
  - Frequency: Daily
  - Countries: 85 (coverage of countries varies by retailer and time)
  - Dates: 2008 to 2013
  - Main Variables: id, country, date, price, category

id	country	date	price	intro	category	categoryid
1193748	cn	01 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	02 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	03 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	04 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	05 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	06 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	07 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	08 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	09 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	10 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	11 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350
1193748	cn	12 Apr 13	139	0	-HOMBRE-CAMISETAS	1771303350

Example: Zara, China.

Examples of Stylized Facts

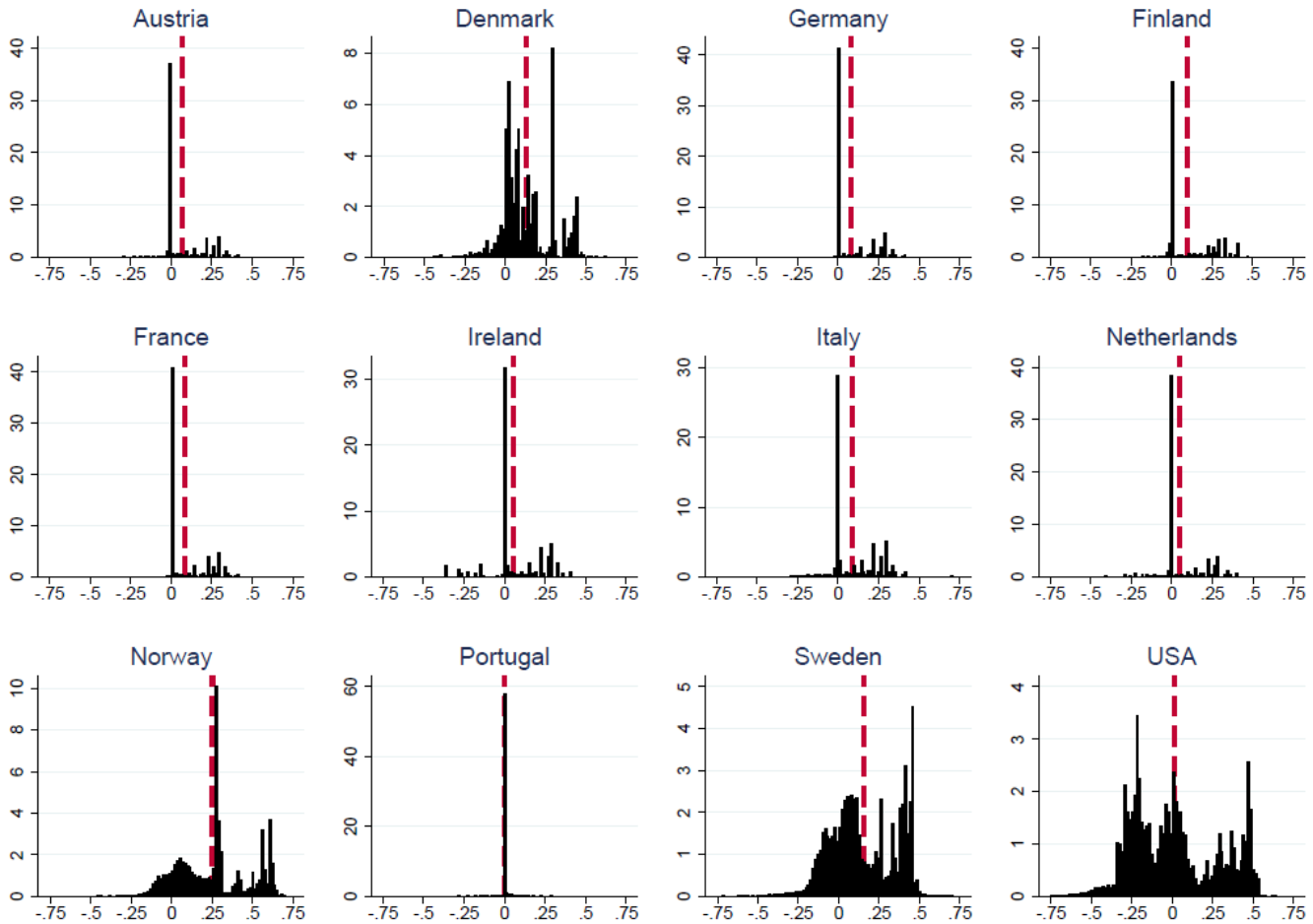
QJE 2014: Good-level RERs with the United States



Note: Log good-level RERs from Apple, Ikea, Zara, and H&M, 2008-2013
From Cavallo, A., Neiman, B., & Rigobon, R. (2014) "Currency Unions, Product Introductions, and the Real Exchange Rate" ***Quarterly Journal of Economics*** - *Vol. 129 (20), p.529-595.*

Examples of Stylized Facts

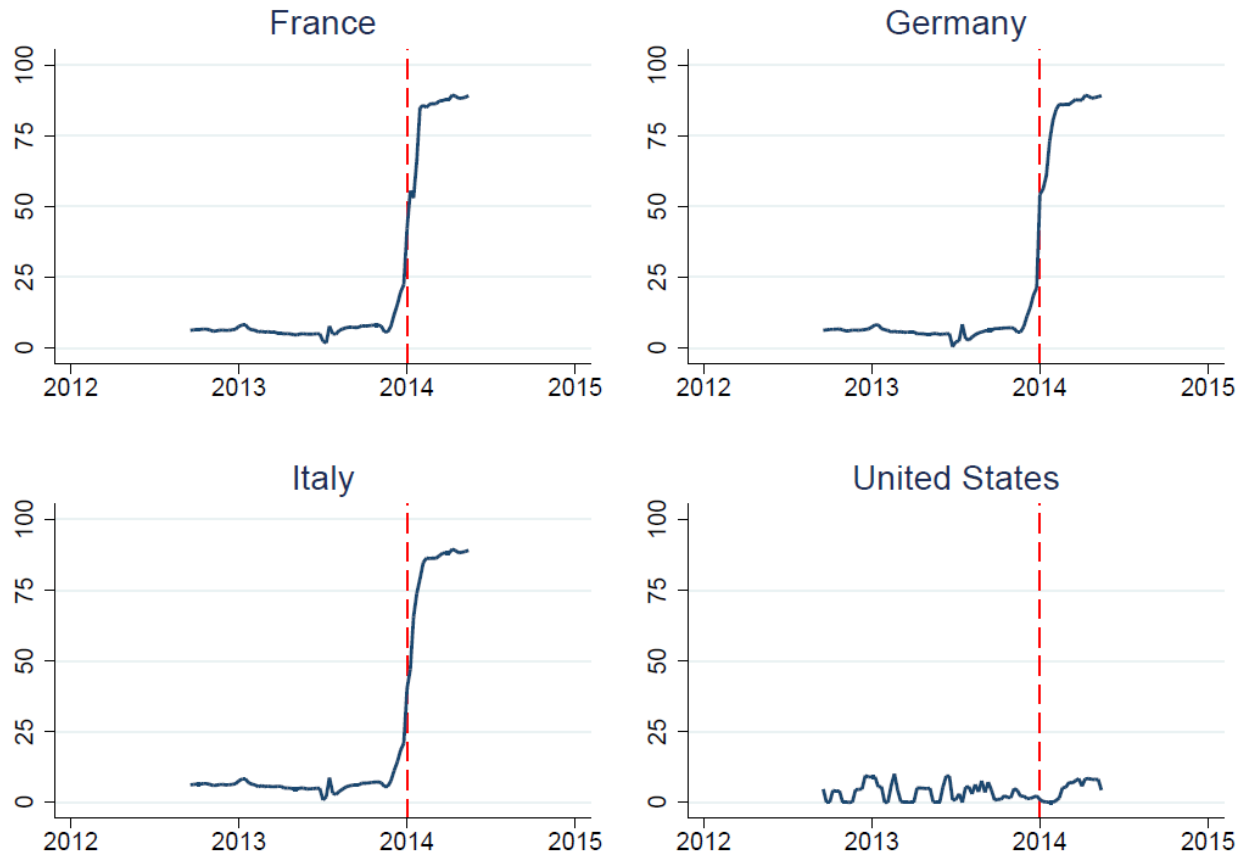
QJE 2014: Good-level RERs with Spain



Note: Log good-level RERs from Apple, Ikea, Zara, and H&M, 2008-2013
From Cavallo, A., Neiman, B., & Rigobon, R. (2014) "Currency Unions, Product Introductions, and the Real Exchange Rate" ***Quarterly Journal of Economics*** *- Vol. 129 (20), p.529-595.*

Examples of Stylized Facts

IMFER 2015: Share of Goods Satisfying LOP with Latvia

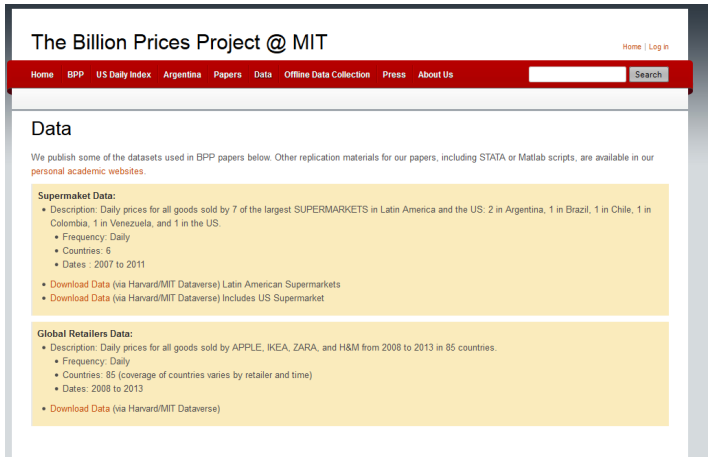


Notes: Plots show the percentage of observed good-level relative prices between Latvia and the listed country of absolute value less than 1 percent, measured each week.
From Cavallo, A., Neiman, B., & Rigobon, R. (2014) "The Price Impact of Joining a Currency Union: Evidence from Latvia" ***- IMF Economic Review*** *Forthcoming.*

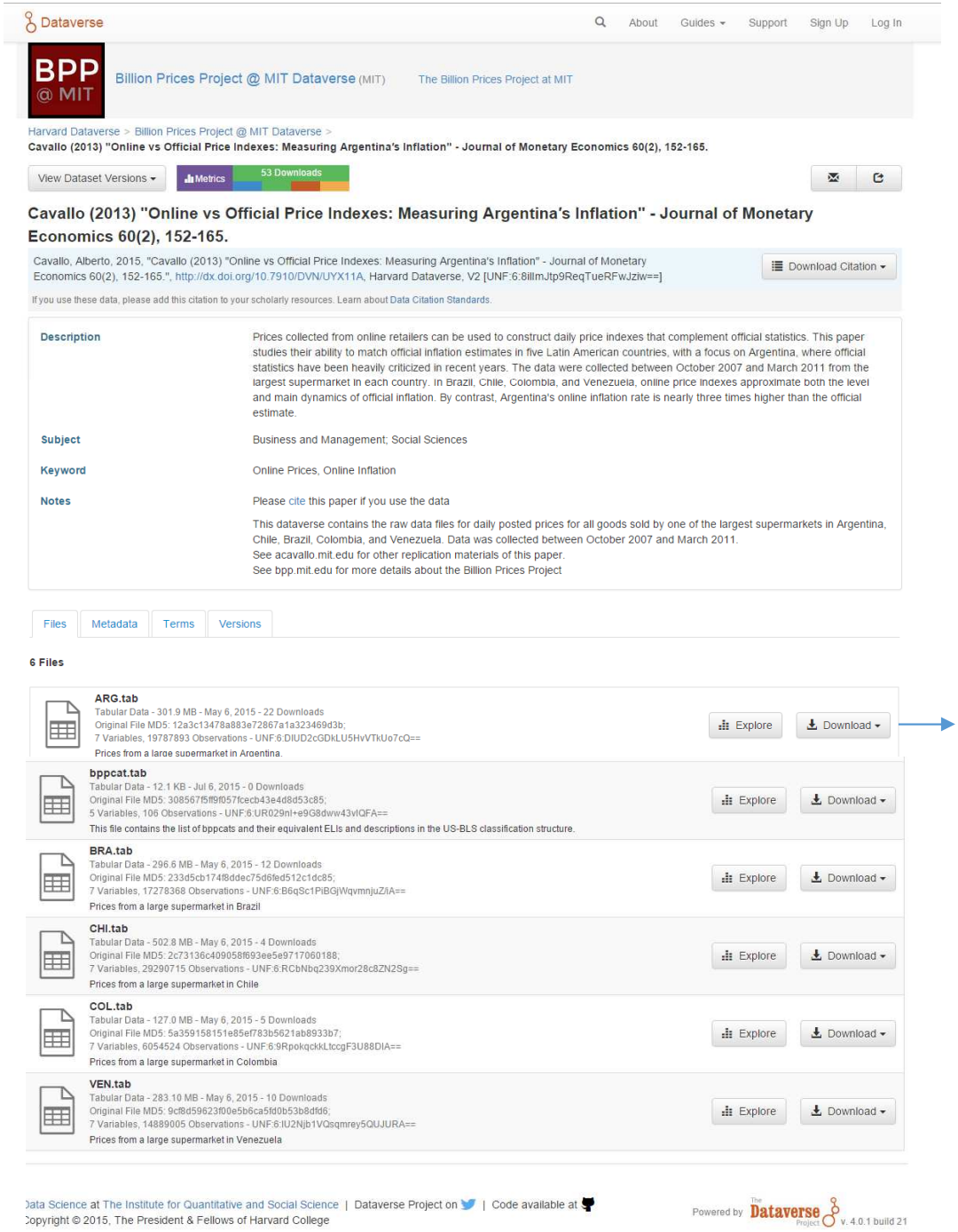
Publicly-shared Datasets

- Pre-cleaning:
  - De-identification of retailer and products names (except in global retailers data)
  - No other cleaning or adjustment → data contains errors and missing values

- Access is free:
  - Go to http://bpp.mit.edu/datasets/ for a list of available datasets
  - Data shared via Harvard s Dataverse Repository



Harvard's Dataverse Repository



Steps

1. Click Download
2. Accept "Terms of Use" (only for research purposes and will not seek/reveal retailer or brand identity)
3. Sign "Guestbook"

Are Online Prices representative ?

- The `online store` is effectively the *largest* store in 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
- Simultaneous online-offline data collection in 10 countries shows large retailers tend to have identical online and offline prices

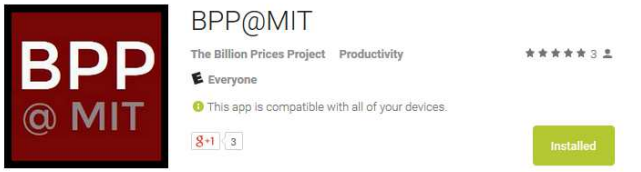
Preliminary results from Cavallo (2015) Are Online and Offline Prices Similar?

Table 1: Country - Price Level Differences

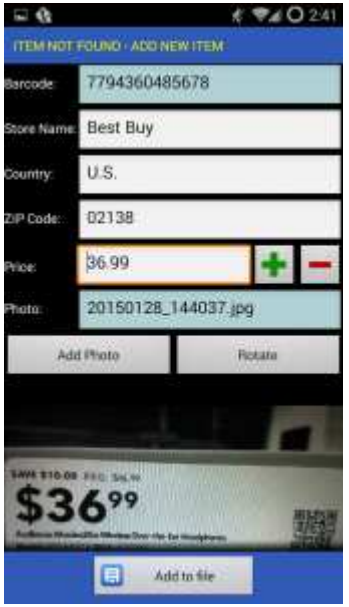
Country	(1) Ret.	(2) Days	(3) Workers	(4) Prod.	(5) Obs	(6) Ident. (%)	(7) High On (%)	(8) Low On (%)	(9) Differ. (%)	(10) On Mark. (%)
Argentina	5	73	20	2271	3560	55	29	16	2	4
Australia	4	61	18	2998	3603	72	21	7	1	4
Brazil	5	66	16	1014	1487	25	39	36	1	1
Canada	5	85	20	2400	3789	89	5	5	0	1
China	2	6	5	20	20	85	10	5	1	7
Germany	3	44	6	756	1077	82	3	16	-1	-6
Japan	3	16	4	428	500	37	11	52	-9	-14
Southafrica	3	38	15	1139	1277	83	8	8	0	1
UK	4	57	15	1813	2420	87	4	9	-1	-4
USA	9	125	243	3110	6379	71	8	21	-2	-7

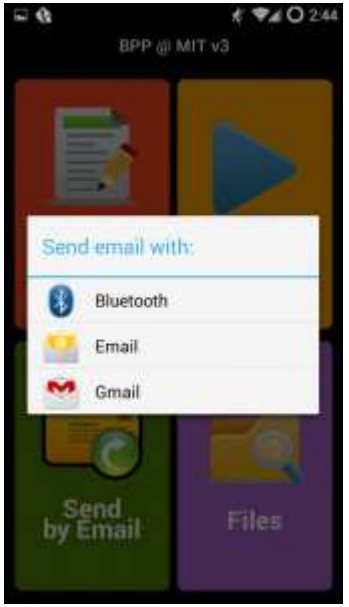
Note:"Difference" includes identical prices. "Online Markup" excludes identical prices. Update: 7/6/2015

A Free Offline Data Collection Tool



https://play.google.com/store/apps/details?id=com.mit.bpp&hl=en





- Download the free BPP app for android phones (bpp.mit.edu)
- Contact us to receive a custom Project Code → allows us to separate your data
- Steps:
  - You scan and email us the files (all within the app)
  - Every day, we clean and consolidate the data, and save everything (prices, text, photos) in a shared dropbox folder.

Future Data Releases

- We plan to periodically upgrade and publish new datasets:
  - See http://bpp.mit.edu/datasets/
  - Follow us on twitter to get notified (@bppmit).