Are Online and Offline Prices Different?

[preliminary results]

Alberto Cavallo
MIT & NBER

Ottawa Group Meeting
Tokyo - May 2015

Motivation

- Prices collected online are increasingly being used for price indices and academic research
- So far, online prices have only been validated by comparing price indices (BPP vs CPI)
- This paper tries to answer a simple question: is the price collected on a retailer's website the same that can be obtained in a physical store?

Simultaneous Online-Offline Data Collection

- Collect prices for:
  - a random set of goods
  - simultaneously (for now allowing a 7-day window)
  - in both the website and a (random) physical store of a given retailer

- Our Goal: 10 countries, 5-10 retailers in each

- Retailers must meet these conditions:
  1. Top 20 largest retailers by market share in each country (Euromonitor)
  2. Sell both online and offline
  3. Possible to link online and offline products (via UPC or some other product code)

List of Retailers

- Retailers (May 2015)

Table 1: List of Retailers

PRELIMINARY

Country	Retailers Included
Argentina	Carrefour, Coto, Easy, Sodimac, Walmart
Australia	Coles, Masters, Target, Woolworths
Brazil	Drogaraia, Lojas
Canada	Canadiantire, Homedepot, Thesource, Toysrus, Walmart
Germany	Galleriak, Obi, Real
Japan	Biccamera
Southafrica	Pnp, Woolworths
UK	Asda, Ms, Sainsburys, Tesco
USA	Bestbuy, Cvs, Homedepot, Lowes, Macys, Riteaid, Safeway, Target, Walgreens, Walmart

Note: Preliminary Results

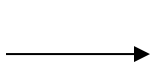
Data Collection

- Offline→ use crowdsourcing platforms + custom BPP app

amazon mechanical turk
Artificial Artificial Intelligence

Elance

oDesk



BPP @ MIT

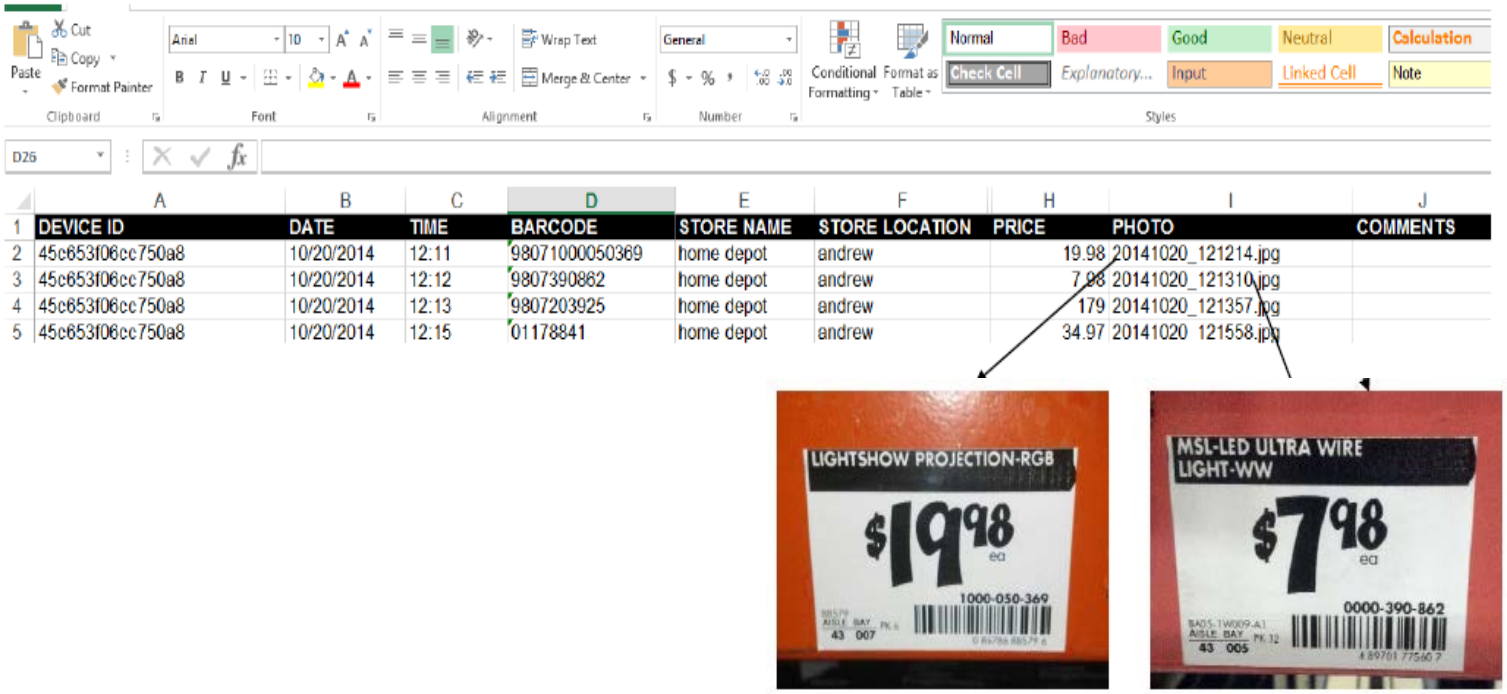
Android App

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

- Task:
  1. Scan 10-50 random offline products in any physical store
     - Product Id is obtained by barcode scanning if possible
     - Price is manually entered
     - Photo of price tag is automatically attached for validation
  2. Upload the data to our central server using the app

Data Collection

- Variables: project id, date, retailer, zip code, upc, price, and photo



	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	

- Every day we process and consolidate the offline data, and then use ids for online collection

Data Collection

- Online: BPP Web Scraping
  - Scrape only the products collected offline each day
  - Search by UPC on the website, not product description → goal is perfect matching

- Some Potential Differences:
  - Sales and coupons
  - Zip Codes
  - We allow the online price to be collected up to 7 days later (to maximize comparable observations)

Price Levels – Country

Table 2: Country - 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	54	17	1668	2682	49	33	18	2	4
Australia	4	41	15	1912	2335	66	25	9	1	4
Brazil	3	46	12	469	761	15	37	48	0	0
Canada	5	73	19	1664	2924	90	5	5	0	1
Germany	3	29	5	346	574	87	3	10	-1	-5
Japan	1	3	1	50	50	22	24	54	-7	-9
Southafrica	2	22	11	542	605	80	9	11	0	1
UK	4	56	15	1810	2317	87	4	9	-1	-4
USA	10	109	239	3053	6055	72	8	19	-2	-7

Note:"Difference" includes identical prices. "Online Markup" excludes identical prices.

Price Levels – Country

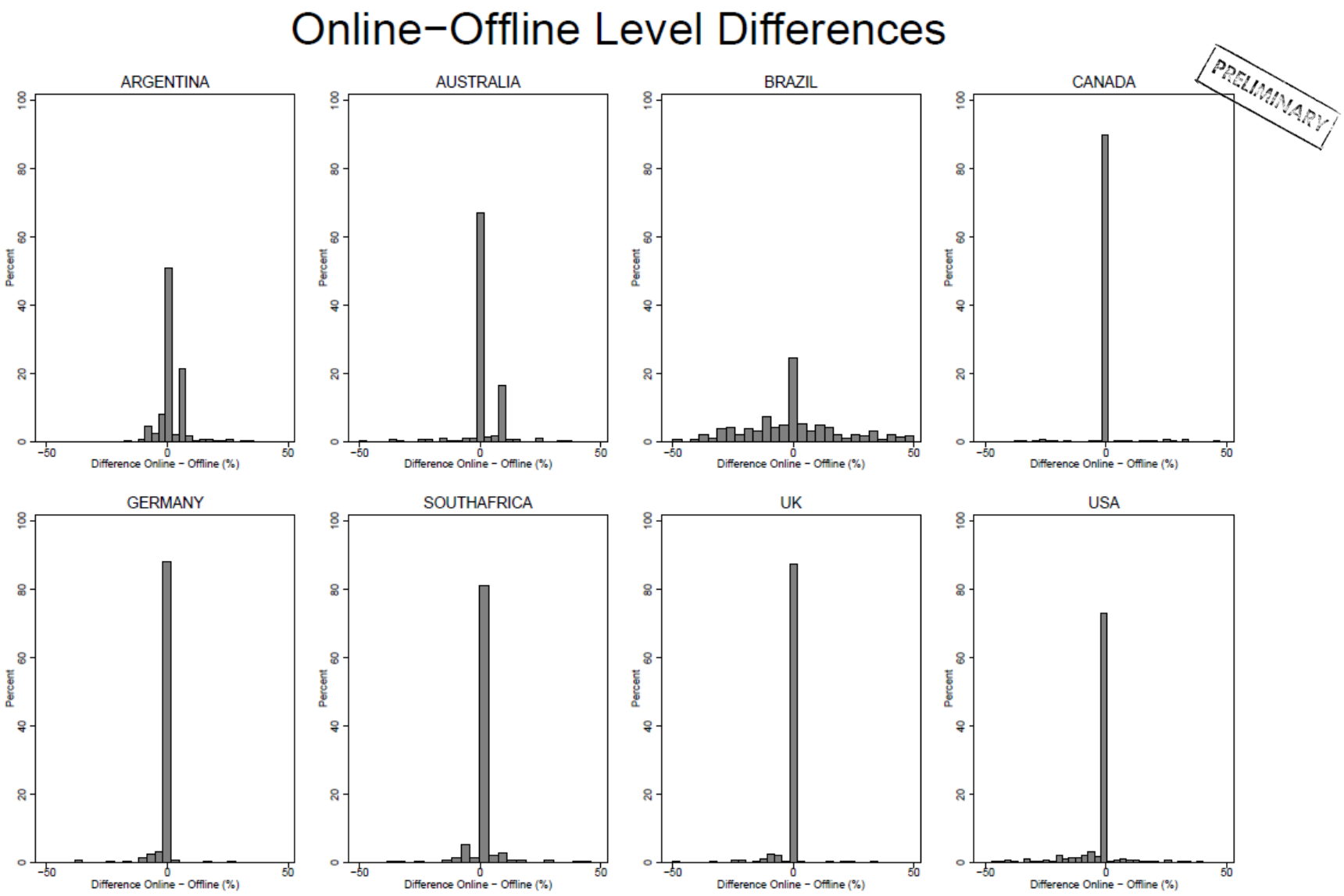
Table 3: Country - Level Differences (cont.)

PRELIMINARY

Country	(1) Obs	(2) Ident. (%)	(3) $< \lvert 1\% \rvert$ (%)	(4) $< \lvert 5\% \rvert$ (%)	(5) $< \lvert 10\% \rvert$ (%)
Argentina	2682	49	50	64	92
Australia	2335	66	67	70	82
Brazil	761	15	23	32	45
Canada	2924	90	90	91	92
Germany	574	87	88	92	95
Japan	50	22	32	50	66
Southafrica	605	80	81	86	93
UK	2317	87	87	88	93
USA	6055	72	73	76	82

Note: Preliminary Results. $\lvert 1\% \rvert$ means that the absolute price difference is less than 1%

Price Levels – Country



Levels – Country

- On average, about 70% of prices are identical
  - Highest in developed countries

- Conditional on a difference, the average "online markup" is typically negative and small (-5%)

Price Levels – Retailers

PRELIMINARY

Table 4: Retailer - Level Differences

Retailer	(1) Days	(2) Workers	(3) Prod.	(4) Obs	(5) Ident. (%)	(6) High On (%)	(7) Low On (%)	(8) Differ. (%)	(9) On Mark. (%)
Argentina_1	17	8	249	484	88	10	2	1	10
Argentina_2	20	6	433	725	18	78	4	5	7
Argentina_3	21	7	390	609	52	12	36	-1	-1
Argentina_4	3	4	91	90	91	0	9	-1	-16
Argentina_5	23	8	507	774	48	24	28	1	3
Australia_6	16	4	521	611	14	81	6	8	10
Australia_7	8	4	277	279	98	2	0	0	4
Australia_8	14	7	639	708	92	1	7	-2	-24
Australia_9	16	7	476	737	72	11	16	-1	-3
Brazil_10	12	4	170	234	39	26	35	-3	-5
Brazil_11	37	8	304	527	4	42	54	1	1
Canada_12	16	6	293	424	89	10	0	3	29
Canada_13	16	8	590	672	88	4	8	0	-3
Canada_14	25	5	192	701	90	3	7	-1	-12
Canada_15	29	7	416	711	88	6	6	0	-2
Canada_16	32	5	172	416	95	4	1	1	13
Germany_17	12	4	104	208	92	6	2	0	4
Germany_18	7	2	71	98	55	2	43	-3	-6
Germany_19	13	4	171	268	96	1	3	-1	-14
Japan_20	1	1	50	50	22	24	54	-7	-9
Southafrica_21	20	8	483	544	78	10	11	0	1
Southafrica_22	3	4	61	61	92	2	7	0	-2
UK_23	25	7	629	840	92	3	5	-1	-8
UK_24	21	6	405	496	96	2	3	0	-6
UK_25	15	3	352	420	58	7	35	-1	-4
UK_26	24	8	457	561	93	4	2	0	2
USA_27	51	54	504	920	75	6	19	-2	-9
USA_28	48	55	182	494	25	19	56	-2	-3
USA_29	65	54	513	1008	87	6	7	0	-3
USA_30	26	19	49	394	92	6	2	1	11
USA_31	36	51	386	622	65	11	23	-4	-11
USA_32	25	20	72	88	95	3	1	0	10
USA_33	15	18	43	43	5	77	19	12	12
USA_34	77	105	791	1146	68	8	25	-3	-9
USA_35	18	23	18	14	43	0	57	-4	-7
USA_36	60	89	598	1379	78	6	16	-2	-9

Note: Preliminary Results. "Difference" includes identical prices. "Online Markup" excludes identical prices.

Retailer Results

- There is significant heterogeneity across retailers, even within countries

- Three main categories of retailers
  - Identical price levels
  - Differences with no online markup (online is just ‘another store’)
  - Differences with online markup

Price Changes

- Price changes do not have identical timing

Table 5: Country - Price Change Differences

PRELIMINARY

Country	(1) Prod.	(2) Obs.	(3) No PC (%)	(4) Same PC (%)	(5) Different PC (%)
Argentina	491	1019	74	3	23
Australia	224	436	86	2	12
Brazil	175	365	37	0	63
Canada	435	1357	85	4	11
Germany	121	262	92	1	6
Southafrica	55	62	71	8	21
UK	277	518	86	2	12
USA	531	3057	89	2	9

Note: Preliminary Results. -Same- means the price changes are simultaneous and with identical size.

Price Changes

- But the average frequency and size of changes is similar

Table 6: Country - Price Change Frequency and Size

PRELIMINARY

Country	(1) Freq. Online	(2) Freq. Offline	(3) Size Online	(4) Size Offline
Argentina	.151	.174	8.13	13.48
Australia	.108	.076	24.75	28.3
Brazil	.51	.216	23.85	21.17
Canada	.089	.105	37.62	36.01
Germany	.061	.031	31.51	18.96
Southafrica	.145	.242	18.93	14.2
UK	.069	.102	34.58	36.56
USA	.07	.074	22.1	27.98

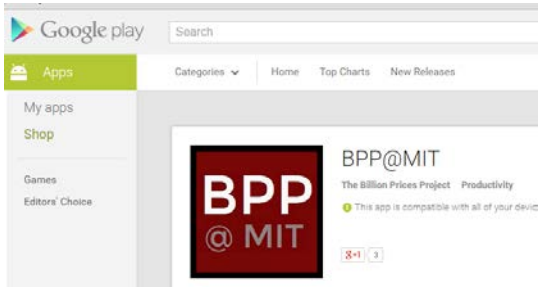
Preliminary Conclusions

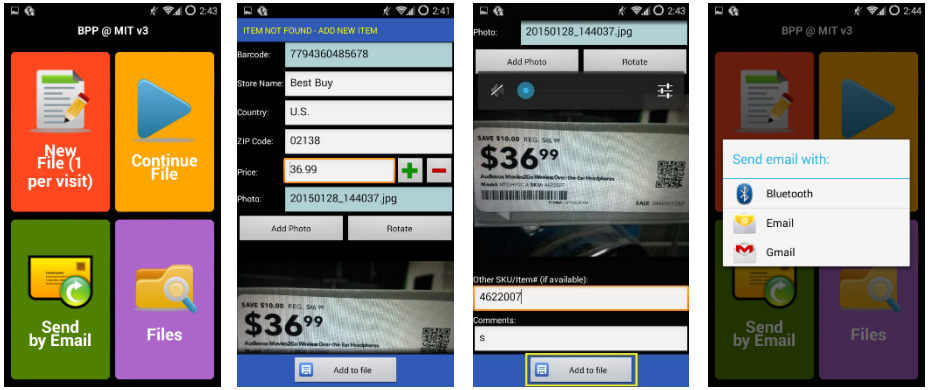
- Prices collected online and offline are very similar in most countries
  - Levels → aprox. 70% are identical prices
  - Changes → different timing, but same frequency and mean size
- Results in this paper are lower-bounds for similarity (imperfect matching, zip code differences, etc)
- Differences across retailers → important to test for potential biases, particularly in papers/ applications with few retailers

Help us collect more offline data

- BPP app is available as a free download
  - We will assign you a 'project code', needed for installation, which allows us to separate your data
  - For online-offline validation or just regular offline-data collection

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





- Steps
  1. You scan
  2. We process the offline data (and check online prices)
  3. Share all the raw data and findings with you

Contact: acavallo@mit.edu , mariabf@mit.edu


MITSloan
MANAGEMENT