

Alternative Macro Data: Some Challenges and Opportunities

Alberto Cavallo

Harvard Business School & NBER

ESCOE London June 2019

Motivation

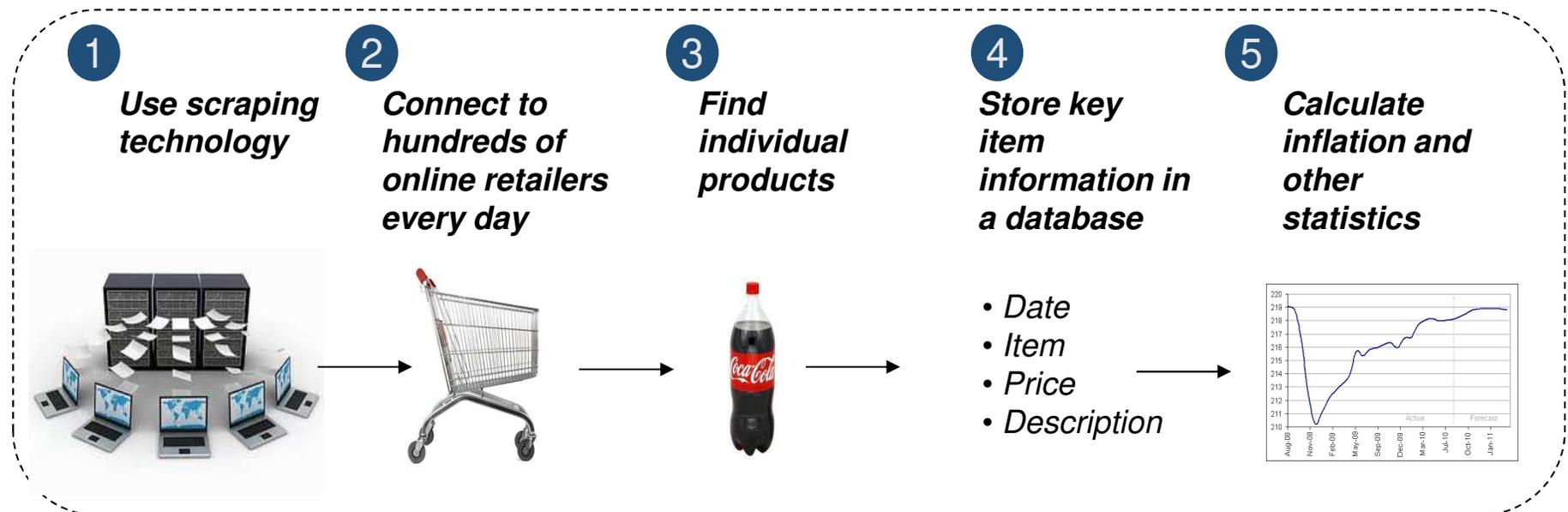
- Can Big Data help to improve macroeconomic measurement?
[Einav & Levin \(2014\)](#), [Eichengreen \(2015\)](#) , [Summers \(2016\)](#)
- A large amount of research in recent years, but still little impact on production within national statistical offices

Big Data in Macro Measurement

- New technologies for data collection
 - Not necessarily big
 - More accessible → private sector & academia
- Macro data sources
 - Surveys (National Statistical Offices)
 - Administrative data (eg. tax, property records)
 - Scanner data (eg. Nielsen)
 - Search data (eg. Google, Indeed)
 - Satellite data (eg. lights, parking lots, tanker and crop heights)
 - Sensor data (smart phones, smart watches, IOT devices)
 - Online scraped data (e.g. Billion Prices Project)
 - Crowd-sourced data

The Billion Prices Project

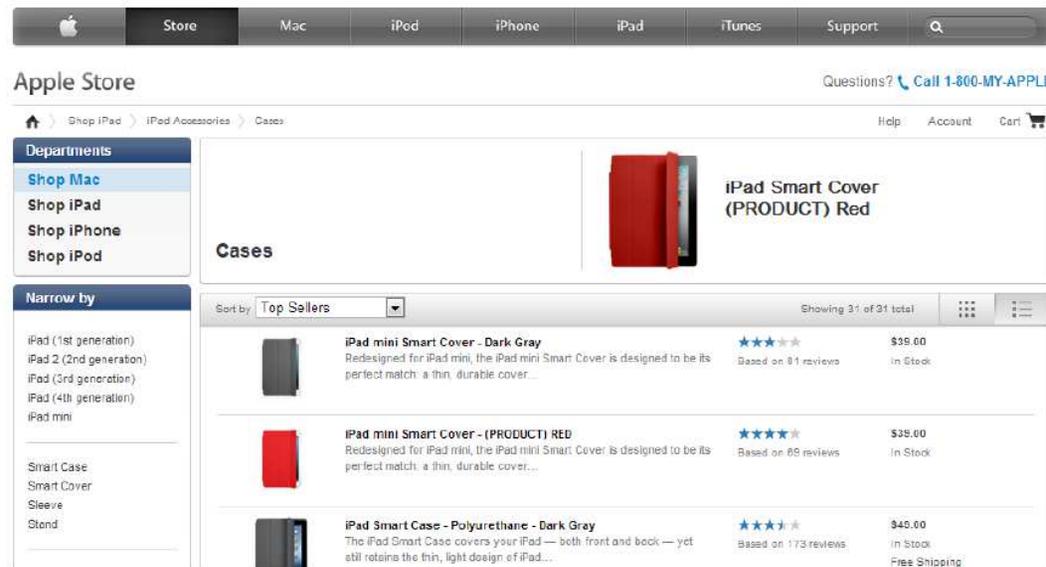
- Academic initiative to collect and use online price data for economic measurement and research
 - Daily prices since 2008 (currently ~25 million observations per day)
 - From hundreds of large multi-channel retailers
 - In over 60 countries



www.thebillionpricesproject.com

Web-Scraping Online Data

- Every day we use software to download public retail webpages, analyze the HTML code, extract product data, and store it in a database



The screenshot shows the Apple Store website interface. At the top, there is a navigation bar with the Apple logo and links for Store, Mac, iPod, iPhone, iPad, iTunes, and Support. Below this, the page title is "Apple Store" with a search bar and a phone number "1-800-MY-APPLE". The main content area is titled "Cases" and features a large image of a red iPad Smart Cover. Below the image, there is a list of products with their names, descriptions, prices, and review counts. The products listed are:

Product Name	Description	Price	Reviews
iPad mini Smart Cover - Dark Gray	Redesigned for iPad mini, the iPad mini Smart Cover is designed to be its perfect match: a thin, durable cover...	\$39.00	Based on 81 reviews
iPad mini Smart Cover - (PRODUCT) RED	Redesigned for iPad mini, the iPad mini Smart Cover is designed to be its perfect match: a thin, durable cover...	\$39.00	Based on 69 reviews
iPad Smart Case - Polyurethane - Dark Gray	The iPad Smart Case covers your iPad — both front and back — yet still retains the thin, light design of iPad...	\$49.00	Based on 173 reviews

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

.....

Comparing Micro-Price Data Sources

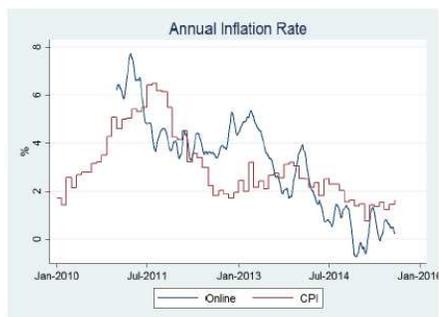
Table 1: Alternative Micro-Price Data Sources

	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

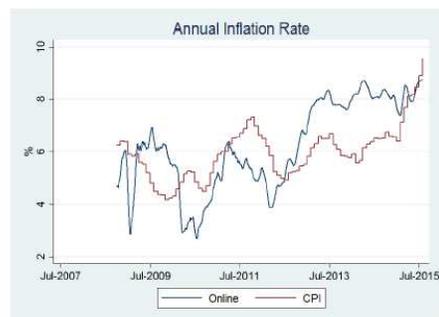
Note: Table from Cavallo (2015b). *Approximate numbers. The Billion Prices Project (bpp.mit.edu) datasets contain information from over 60 countries with varying degrees of sector coverage. Nielsen US scanner datasets are available at the Kilts Center for Marketing of the University of Chicago. Klenow 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.

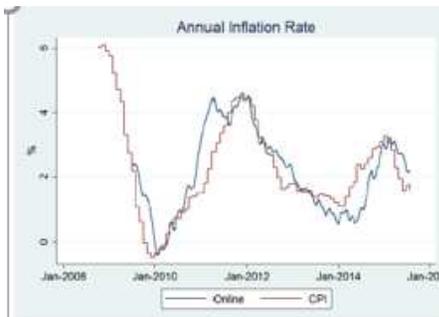
Countries and Sectors



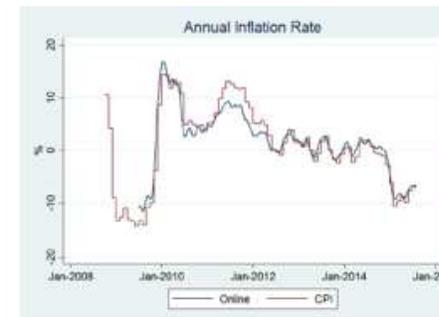
(a) China



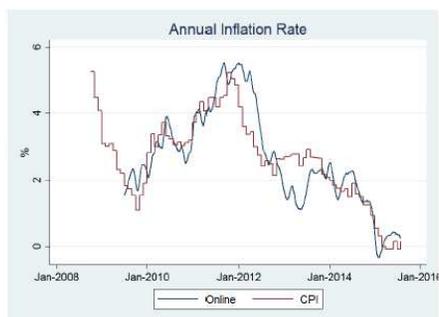
(b) Brazil



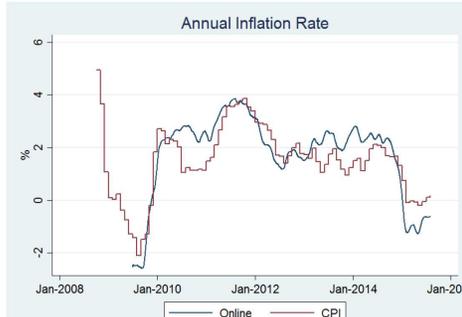
(i) USA Food



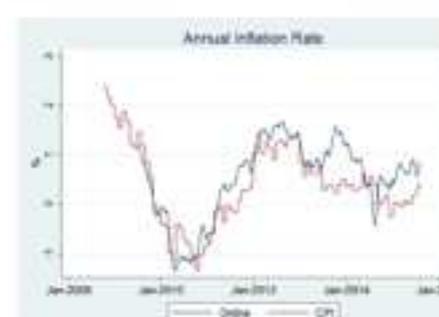
(j) USA Fuel



(e) UK



(f) USA



(k) USA Electronics



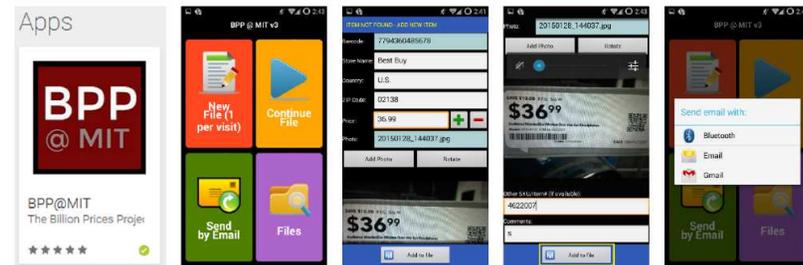
(l) USA Medical Car

Are online prices representative of retail transactions?

- No
 - Online sales are still less than 10% of retail sales in most countries
 - Online retailers tend to have different pricing behaviors - Gorodnichenko et al (2017, 2018)
- Some evidence for Yes
 - Co-movement in online indices and CPIs
 - The 'online store' is effectively the *largest* store for most retailers. Eg: Last year Walmart had 4759 stores in the US. The median store had 0.02% of sales. The 'online store' had 8% of sales
- Are online prices different from what we would observe at a physical store?

Cavallo (2017) Are Online and Offline Prices Similar?

- Large-scale comparison of online and offline prices collected simultaneously in ~50 retailers in 10 countries.
- Crowdsourced workers scan random barcodes, enter prices, send emails with 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 then scraped the online price for the same good-retailer (within 7 days).

Which Online Prices ?



- Retailers the sell both online and offline (multi-channel).

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.

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, less similar in Drugstores and Office Supply retailers.
- Price changes have different timing, but similar frequency and size

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

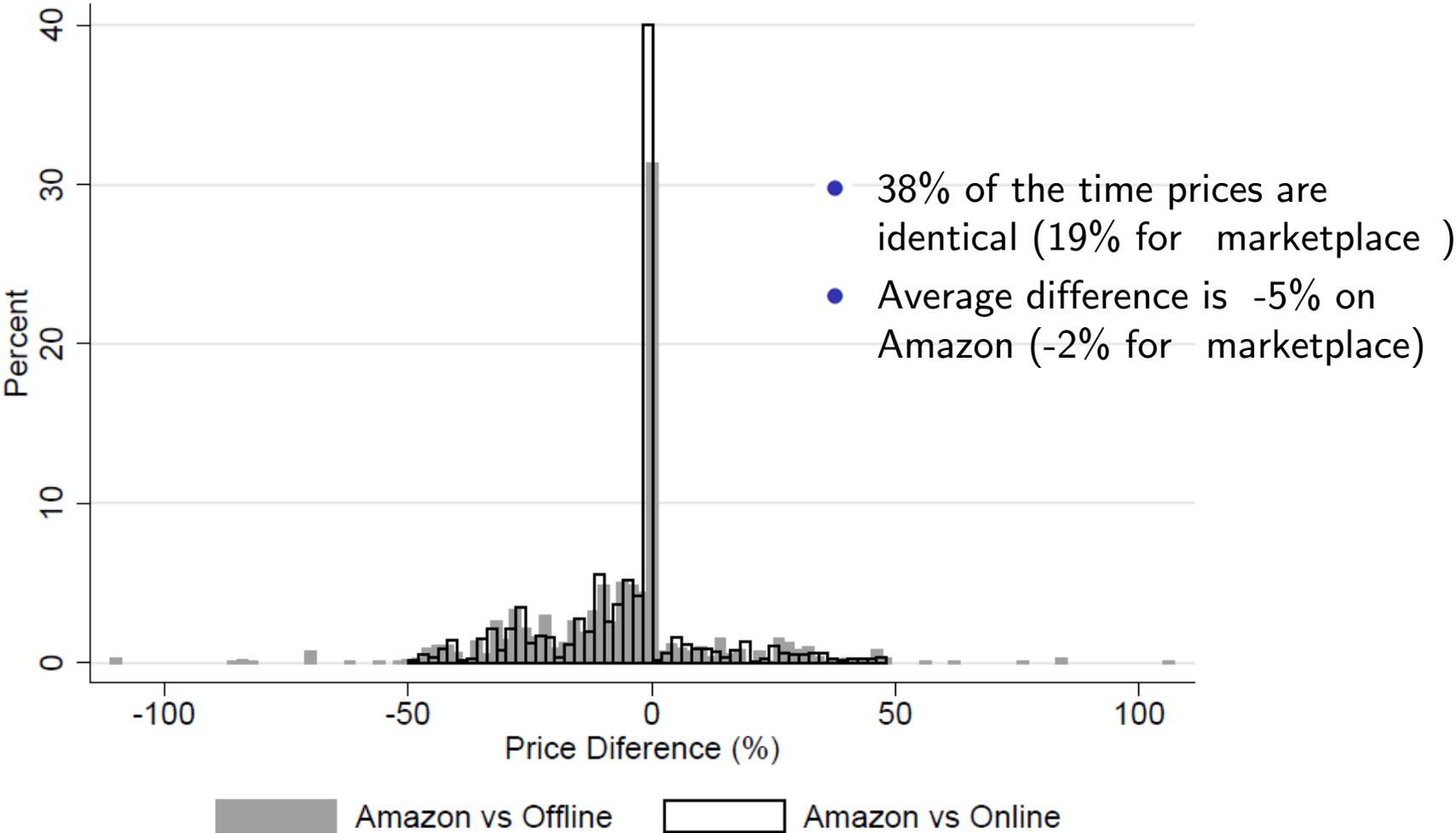


2. Persistent browsing habits: we scraped same products, every five minutes, for a full day.

- Online prices were always the same in my tests

More differences across retailers...but not so much

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

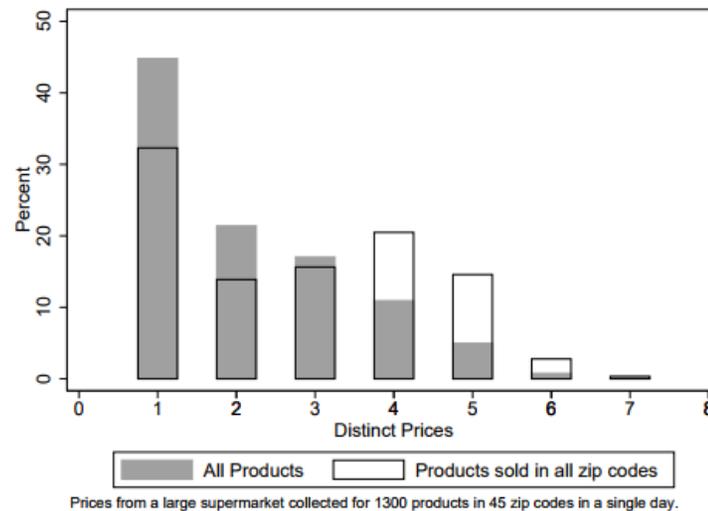


Source: Cavallo (2016) "Are Online and Offline Prices Similar?"

Uniform Pricing Online

- Even online supermarkets that require a zip code before showing prices tend to have similar prices 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)

Uniform Pricing Online

- Global retailers use uniform pricing in currency unions
- Cavallo et al (2014), using prices from Apple, Zara, Ikea, H&M

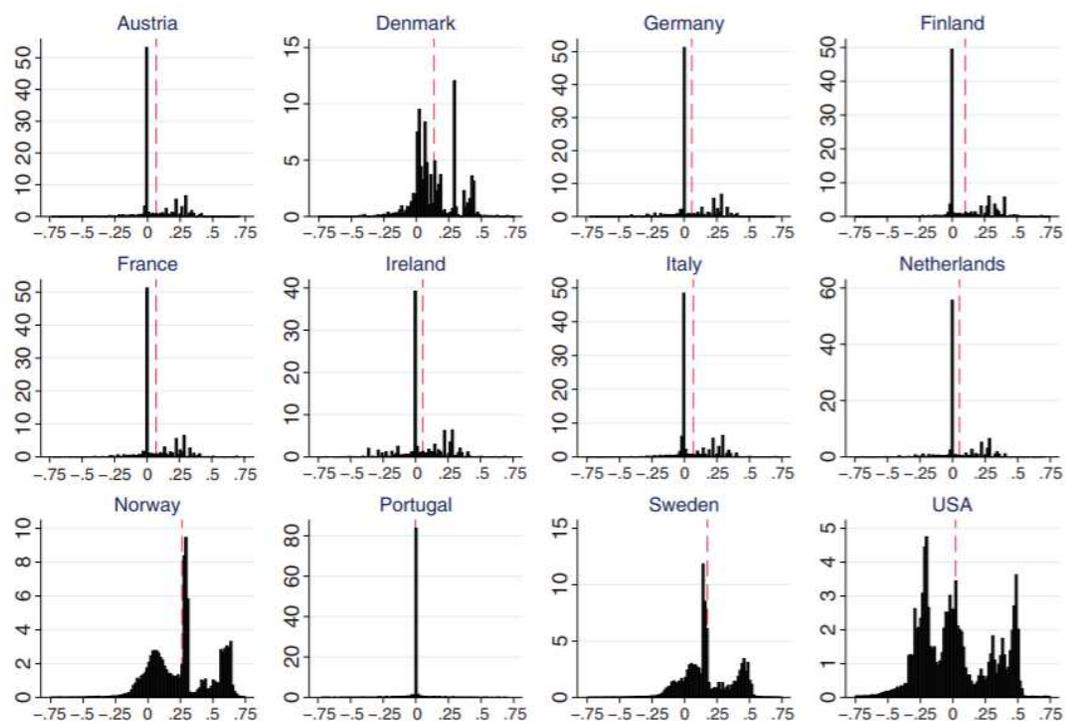


FIGURE IV

Good-Level RERs q_{ij} for Various Countries (i) with Spain (j)

Figure includes all goods z and all weeks t for which q_{ij} is defined, with Spain as country j and the other countries as country i . Histograms include frequency weights such that the contribution from each store is equalized for each country pair. We exclude the small number of observations where $|q_{ij}| > 0.75$. Dashed vertical lines indicate the weighted average RER. y-axes plot percents.

What about offline (different locations)?

- DellaVigna & Gentzkow (2017) find evidence of uniform pricing offline
 - Nielsen's scanner data in the US
 - Attributed mostly to high managerial decision costs
- Aten & Rigbi (2017) → use Israeli scanner data to show that online posting of prices in supermarkets reduces price dispersion across locations within the same retailer
- So is uniform pricing an Amazon Effect ? Is it driven by online transparency and competition?

Cavallo (2019) More Amazon Effects: Online Competition and Pricing Behaviors

- Online competition is affecting pricing behavior in large multi-channel retailers in two ways:
 - Increase in the frequency of price changes
 - Increase in uniform pricing
- Reasons?
 - Algorithmic pricing and matching
 - Online transparency and fairness constraints
- Online and offline pricing behaviors are becoming more integrated, in was that matter for measurement

Increase in the frequency of US price changes

- I use BPP data to measure the aggregate frequency of price changes in the US from 2008 to 2017 (with CPI expenditure weights)

Figure 1: Monthly Frequency of Price Changes, 2008 to 2017

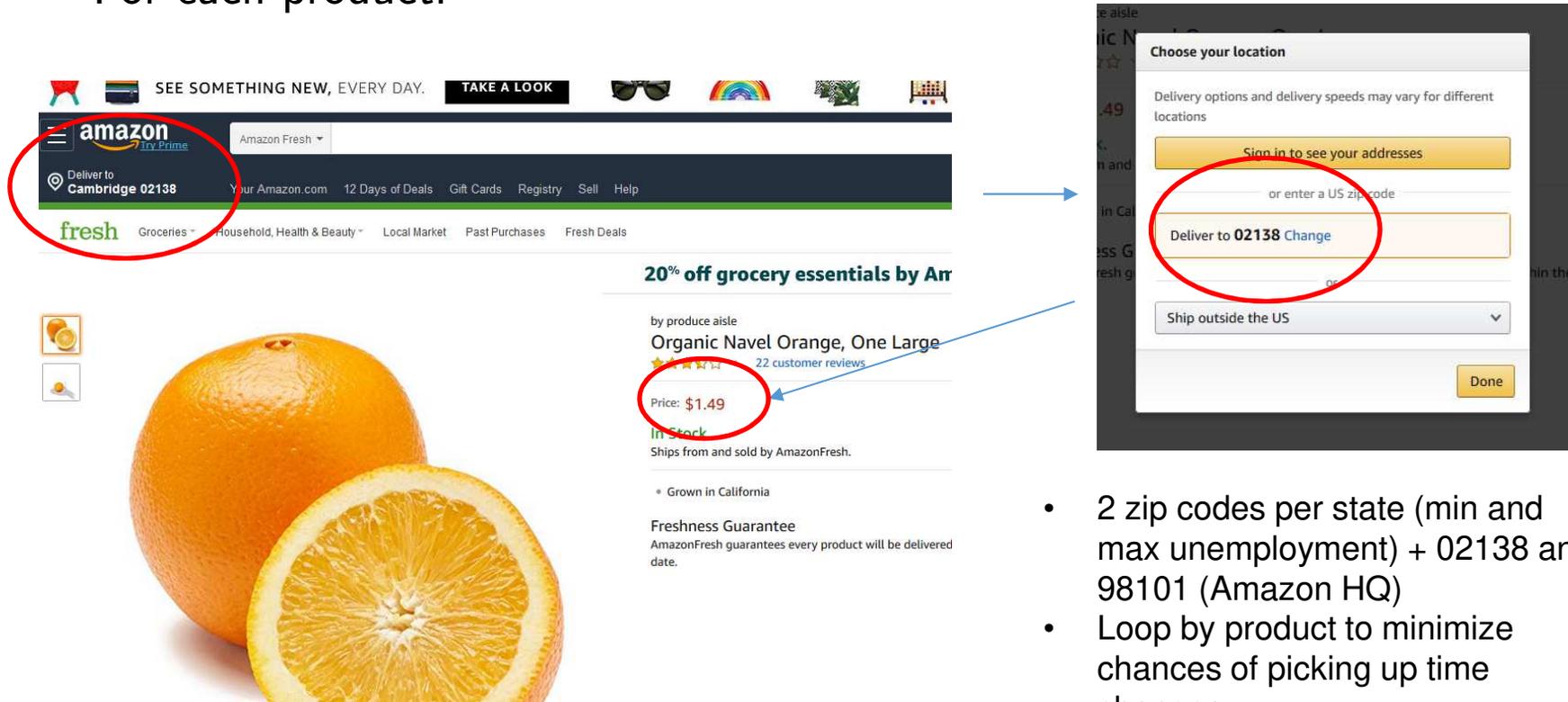


(a) Posted and Regular Price Changes

- The median monthly frequency has risen from approximately 14% in 2008 to 34% today → implied duration of prices has fallen from 7 to 3.5 months (closer to results with online-only retailers)

Comparing degrees of Uniform Pricing

- Data from 4 US retailers (Amazon & 3 traditional) in 102 zip codes.
- For each product:



The image shows a screenshot of the Amazon Fresh website. The top navigation bar includes the Amazon logo, a search bar, and a delivery location dropdown set to "Cambridge 02138". Below the navigation bar, there are promotional banners for "20% off grocery essentials by Amazon" and a product listing for "Organic Navel Orange, One Large". The price for the orange is listed as "\$1.49". To the right, a "Choose your location" modal is open, showing a "Sign in to see your addresses" button and a "Deliver to 02138 Change" button. A red circle highlights the "Deliver to 02138 Change" button in the modal, and another red circle highlights the price "\$1.49" on the product page. Blue arrows point from the modal to the product page and from the modal to the list of bullet points on the right.

- Two measures of price dispersion using all bilateral price comparisons: share of identical & average price difference

- 2 zip codes per state (min and max unemployment) + 02138 and 98101 (Amazon HQ)
- Loop by product to minimize chances of picking up time changes

Uniform Pricing in Amazon vs Traditional Retailers

Table 4: Evidence of Uniform Pricing in Large US Retailers

	Share of Identical Prices		Average Price Difference	
	Other Retailers	Amazon	Other Retailers (%)	Amazon (%)
Panel A: All Sectors				
Mean	0.78	0.91	5.49	1.61
Standard Deviation	(0.30)	(0.19)	(9.44)	(4.44)
Number of Products	9469	823		
Average Zip Codes	22	80		
Panel B: Major Sectors				
Food & Beverages				
Mean	0.76	0.84	6.33	2.97
Standard Deviation	(0.31)	(0.24)	(9.84)	(5.26)
Number of Products	6588	344		
Average Zip Codes	15	65		
Recreation & Electronics				
Mean	0.99	0.99	0.006	0.003
Standard Deviation	(0.16)	(0.05)	(0.22)	(0.04)
Number of Products	1578	191		
Average Zip Codes	42	100		

- Amazon has more uniform pricing, but traditional retailers are close
- Differences are concentrated in Food and Beverages

Online competition increases Uniform Pricing in Walmart

Table 5: Uniform Pricing for Walmart's Grocery Products Found on Amazon

	<u>Share of Identical</u>		<u>Average Price Difference</u>	
Found on Amazon	0.058 (0.008)	0.055 (0.008)	-1.979 (0.306)	-1.891 (0.309)
Zip Codes Sampled		0.002 (0.000)		-0.044 (0.017)
UE Rate Difference		-0.006 (0.002)		0.386 (0.071)
Constant	0.914 (0.004)	0.921 (0.009)	2.939 (0.152)	1.794 (0.386)
Observations	3,982	3,949	3,778	3,746
Obs. on Amazon	934	929	908	903
R-squared	0.022	0.031	0.014	0.024

- Walmart's products that are found on Amazon have less geographical price dispersion
- The impact of found on Amazon is equivalent to a 10% unemployment rate difference across locations

Uniform Pricing

- The same technology that enables faster price changes (algorithmic pricing) can be used to set different prices for different consumers (at the same time)
- So what can explain uniform pricing online?
 - Fairness / Customer Anger
 - Customers do not like to know that the exact same product is sold to someone else at a different price
 - particularly within same firm, at the same time
 - It can be legal, technologically feasible, and economically efficient to price discriminate, but fairness imposes a strong constraint on spatial price differences

Fairness concerns impose a constraint

- Online transparency → easier to detect/observe any price difference across locations or outlets



CNN Money

NEWS > Technology

Amazon pricing flap

September 28, 2000, 1:43 p.m. ET

amazon.com Web retailer apologizes for price test, refunds money to customers

NEW YORK (CNNfn) - The Web retailing giant Amazon.com apologized to its customers for conducting random price tests during which customers were charged different prices for the same item.

Between Aug. 31 and Sept. 5, Amazon (AMZN: Research, Estimates) varied the prices on 68 DVDs to test how price changes would affect the unit volume of DVD sales and the total amount of revenue received. In technical terms, the retailer was trying to determine the elasticity of the demand curve for DVDs.

Amazon said that it varied the discount levels on a totally random basis, not with respect to customer demographic information. When customers detected that different prices were being charged for the same item, they became annoyed, causing Amazon to abandon the test and

SAVE THIS
EMAIL THIS
PRINT THIS
MOST POPULAR



Date	Price
1/3	30
3/8	45
5/11	60
7/17	75
9/18	85

Source: Screenshot taken on 4/9/2018, <https://money.cnn.com/2000/09/28/technology/amazon/>



WIRED

LUKE ORSHAN SECURITY 02.05.07 11:55 AM

BEST BUY PUSHES WORST BUY

TECHIE MEGA-STORE BEST Buy has a secret intranet site that employees used to rip off customers, according to a consumer watchdog column in the Hartford Courant.



Employees at Best Buy stores repeatedly failed to honor discounted prices listed on BestBuy.com, putting the burden of proof on consumers. When questioned about the cheaper online prices, Best Buy employees flashed customers the intranet site, which often failed to display the discounted prices.

Source: Screenshot taken on 4/9/2018, <https://www.wired.com/2007/03/best-buy-pushes/>



Popular Latest Sections Magazine More Subscribe



How Online Shopping Makes Suckers of Us All

Will you pay more for those shoes before 7 p.m.? Would the price tag be different if you lived in the suburbs? Standard prices and simple discounts are giving way to far more exotic strategies, designed to extract every last dollar from the consumer.

JERRY USSEM | MAY 2017 ISSUE BUSINESS

Source: Screenshot taken on 4/9/2018, <https://www.theatlantic.com/magazine/archive/2017/05/how-online-shopping-makes-suckers-of-us-all/521448/>

Measurement Implications

- Higher frequency of price changes → monthly sampling will produce more volatile estimates
- With uniform pricing → one store is all we need!
 - No need to visit multiple locations for each retailer
 - Can do it online (since online-offline prices similar)
 - Use resources instead to improve measurement in other categories (or increase frequency)

Other BPP Projects with measurement implications

- Relative Prices and Purchasing Power Parities (PPPs)
 - Cavallo, A., Diewert, W.E., Feenstra, R.C., Inklaar, R., Timmer, M.P., 2018. Using Online Prices for Measuring Real Consumption Across Countries. AEA Papers and Proceedings.
- Scraped Wages (from freelancer websites)
 - Cavallo, Cravino, Drenik (2019) Using online data for international wage comparisons
- Crowdsourcing & phones to measure inflation in Venezuela

Can we crowd-source inflation measurement?

The case of Venezuela

- No CPI data since 2015 → Central Bank stopped publishing it
- No online data (the last retailer in the BPP dataset disappeared in 2014)
 - Until that time, online estimates close to CPI
- IMF uses the black-market exchange rate and PPP to estimate an annual inflation rate above 1.3 million % for 2018
- Local private efforts to send people to collect prices face fines or even criminal charges.

- Can we crowd-source the data collection?

Venezuela s Crowdsourced Inflation

INFLACION VERDADERA

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



INFLACIÓN ARGENTINA

En 2007 creamos InflationVerdadera.com para proveer índices de precios alternativos en Argentina, donde las estadísticas oficiales no fueron creíbles entre 2007 y 2015.



INFLACIÓN VENEZUELA

En 2017 comenzamos a estimar la inflación de Venezuela usando precios recolectados con teléfonos móviles y crowdsourcing.

We tried to enlist volunteers



- We failed..
- So we hired 50 workers in Upwork (freelancer website)
- Currently collecting ~500 prices every week in 10 cities

Methodology

- Customized app (Android)
- Data sent via email
- IFTTT / Zapier → transfer records and files from gmail to Dropbox
- Stata loops through files, cleans data, computes an index
- Public download of all micro data and code

InflacionVerdadera.com Argentina ▾ Venezuela ▾ Blog ▾

Micro Data

Los precios que usamos en las estadísticas pueden bajarse en formato Excel o Stata. Estos archivos sólo incluyen los precios que han sido validados. Los nombres de las tiendas/establecimientos y otros datos sensibles han sido codificados en estos archivos por razones de confidencialidad.

[Bajar Data](#)

Código de Replicación

El código fuente con el que computamos todas las estadísticas puede bajarse y leerse abriendo los archivos .do que están en este archivo zip. Para correr el código es necesario usar Stata 14 y colocar al archivo .dta (la data en formato Stata) en el mismo directorio.

[Bajar Código](#)

We use Amazon Mechanical Turk to validate prices

InflacionVerdadera.com

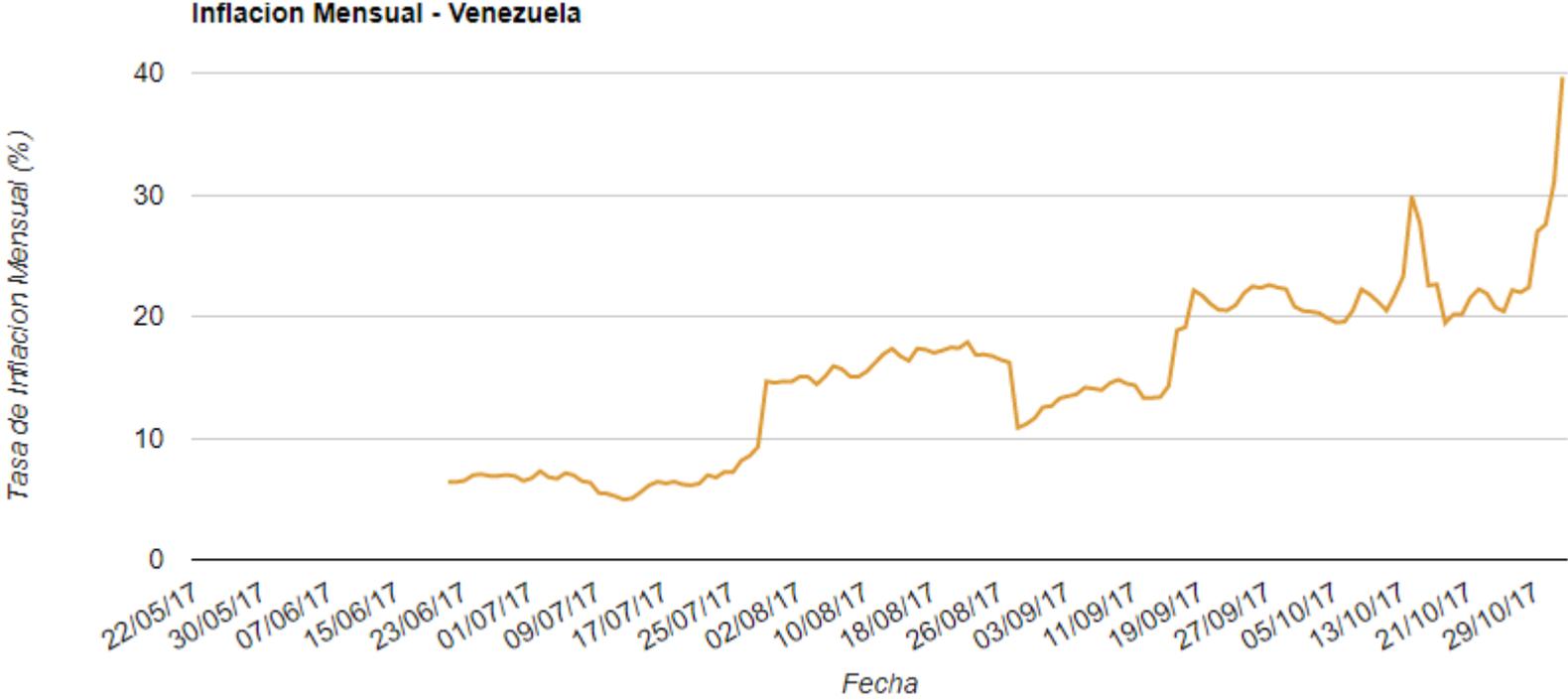


Cual es el precio que se ve en la imagen?

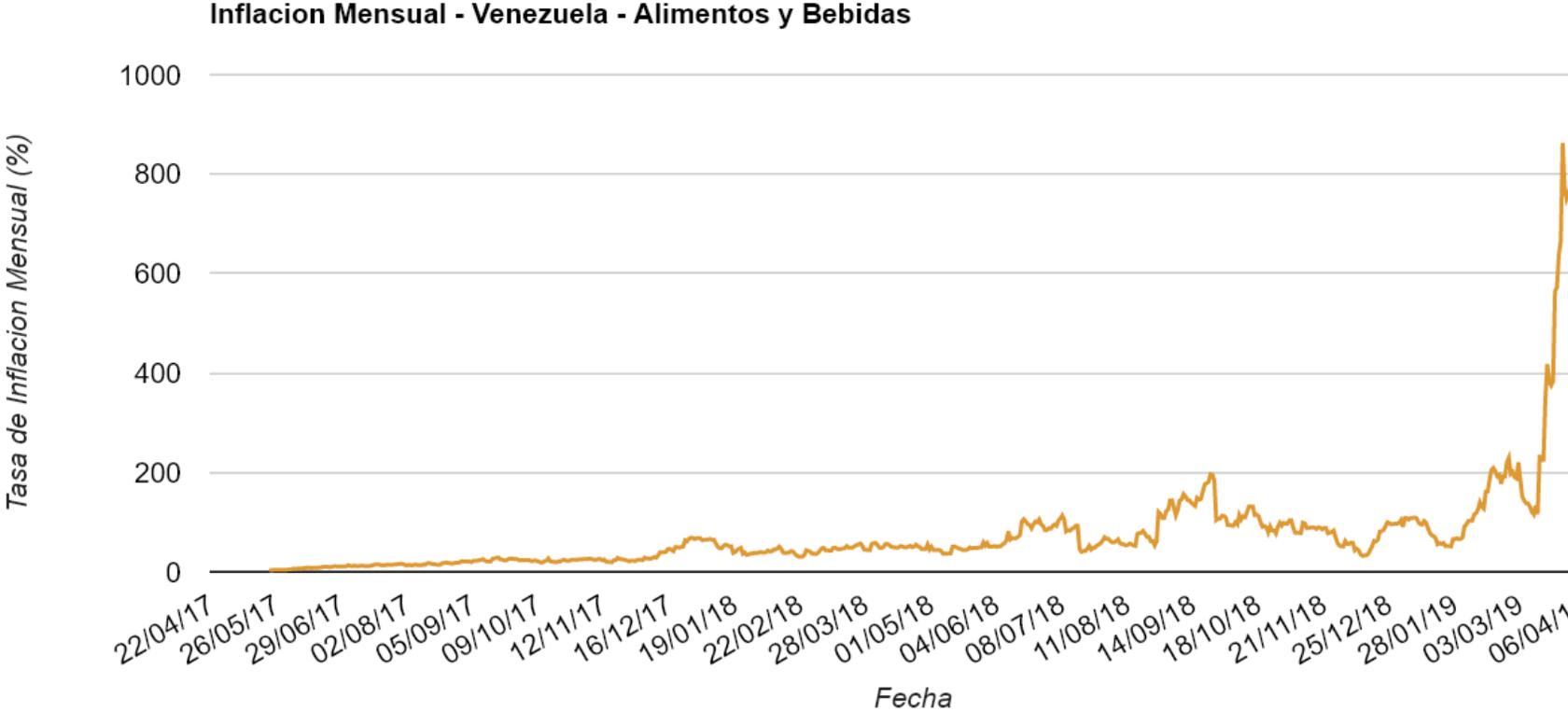
El total a pagar, incluyendo el IVA (PMVP + IVA)

- El precio es:
- No se ve el precio / No se ve la imagen

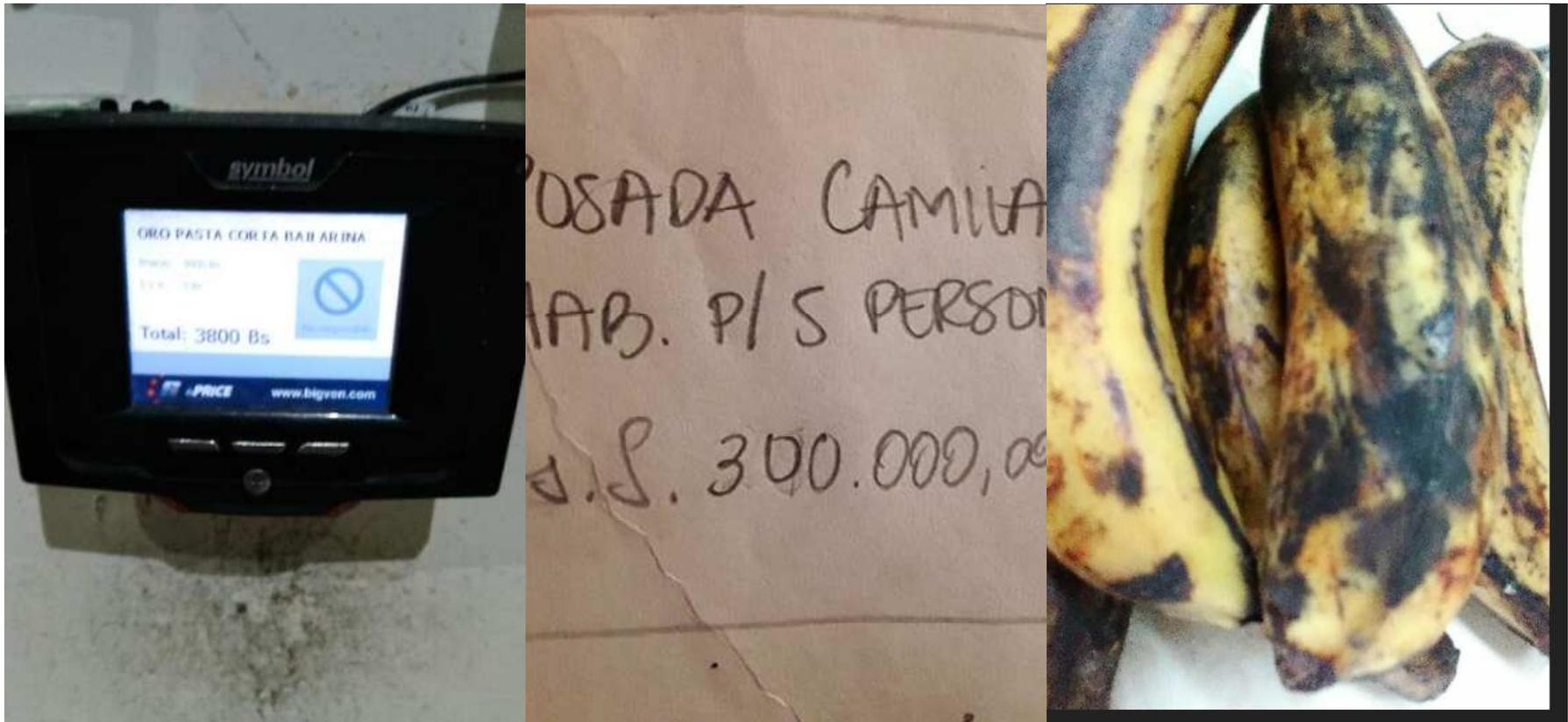
Monthly Inflation when we started



Monthly Inflation now..



Price-tag photos are changing...



But results are promising

- Possible to use off-the-shelf technology
- Low cost (US\$ 0.1 per validated observation)
- Automated cross-validation of data with other freelancers (mechanical turk)
- Freelancers are careful and reliable

Other Challenges

- Technological skills → use toolkits / simple platforms
 - Point-and-click scraping software (not python)
 - Crowdsourcing:
 - Freelancers: Upwork, mechanical turk
 - Automation tools: IFTTT, dropbox
- Data Messiness
 - You don't have to use it all
 - Know the questions beforehand
 - Improve the data collection!
- Move to production/experimental soon
 - Academic research environment is a poor substitute for real production settings → active learning

Final Remarks

- Big Data → huge measurement opportunity
 - New technologies for data collection (web, sensors, phones, gps, satellites)
 - Customized datasets that fit specific measurement and research needs
- [Griliches \(1985\)](#) *The Uneasy Alliance* between economists and data

“... we have shown little interest in improving it [the data], in getting involved in the grubby task of designing and collecting original data sets of our own. Most of our work is on “found” data, data that have been collected by somebody else, often for quite different purposes... “They” collect the data and are responsible for all their imperfections. “We” try to do the best with what we get, to find the grain of relevant information in all the chaff.”

- Better data collection → better measurement

Final Remarks

- Online Price data
 - Advantages for inflation and PPP measurement
 - Improves the quantity and quality of micro price data available for research (<http://www.thebillionpricesproject.com>)
- Crowd-sourced data
 - Only option for some countries (e.g. Venezuela), viable alternative for other countries / categories of goods or services