Real-Time Inflation Measurement

Alberto Cavallo
Harvard Business School

CFM - PER Data Initiative
Columbia University  March 2021
Big Data in Macro and International Economics

• Macro data have many problems (Eichengreen (2015), Summers (2016))

• We mostly rely on data collected by governments → Griliches (1985) The Uneasy Alliance

“... 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.”

• Can Big Data help? (Einav & Levin (2014))
Big Data in Macro and International Economics

- New data sources in recent years
  - Administrative data (eg. CPI and IPI micro data, 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)
  - Crowd-sourced data (web, mobile phones)
  - Online data (eg. Billion Prices Project)

- The 5 Vs and the origin of the BPP
Big Data

- Volume (Data Size)
- Velocity (Speed of Change)
- Variety (Different Forms of Data Sources)
- Veracity (Uncertainty of Data)
- Truth (Value Business Value)
DECEPTION IN ARGENTINA, 2007-2015

Government attempts to identify retailers sampled

National statistics office “intervened”

OFFICIAL CPI

PERCEIVED INFLATION

Source: Di Tella University
ONLINE PRICES PROVIDED AN ALTERNATIVE WAY TO COLLECT PRICE DATA

<html>
<descripcion> Leche Condensada </descripcion>
<br> <brand> Nestlé </brand>
<br> <td price> $1.199 Uni </td>
ONLINE INFLATION IN ARGENTINA

Online indices matched CPIs in other countries

Fig. 1. Online and official indexes in four Latin American countries: (a) Brazil; (b) Chile; (c) Colombia; and (d) Venezuela. Notes: The daily online supermarket index is constructed with an online prices and official CPI category weights. In Venezuela, the online data has no category information and therefore the online index is built as a geometric average of all price changes observed each day. The official supermarket index is an equivalent indicator constructed using a price-weighted double array. All weights are taken from the Office of Household Inflation Education Office.
IS IT ONLINE DATA?

Average annual inflation (%) between 2007–2011

- Venezuela: 27.4%
- Colombia: 27.4%
- Chile: 3.8%
- Brazil: 4.9%
- Argentina: 8.5%

Difference was not explained by data source, method, or type of goods

WHAT HAPPENS IF WE DIVIDE BY 3?

Statistical Uncertainty

- By underestimating inflation they overestimated growth
www.InflacionVerdadera.com
Un aporte a la transparencia en Argentina

Los Precios Hoy

Somos un grupo de economistas que todos los días registra los precios de un listado de productos en dos grandes supermercados de Buenos Aires y los publica en esta sitio.

En esta sección mostramos un resumen de nuestras principales estadísticas, basadas en la evolución diaria de dos índices propios de inflación:

- Un Índice de Alimentos y Bebidas, equivalente al 31% del IPC.
- Un Índice de la Canasta Básica Alimentaria, como el que el INDEC usa (o usaba) para calcular el nivel de inflación.

En la sección Noticias hacemos un análisis periódico de los datos. En las secciones de Precios, Aumentos y Gastos se pueden ver detalles de precios individuales. Finalmente, en Metodología y Preguntas explicamos cómo los hacemos y cómo pedir ayuda.

<table>
<thead>
<tr>
<th>AUMENTOS DE HOY (%)</th>
<th>AUMENTOS ESTA SEMANA (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIDEOS SECOS M</td>
<td>0%</td>
</tr>
<tr>
<td>POZO CONCORDA</td>
<td>2.74%</td>
</tr>
<tr>
<td>NABESA</td>
<td>2.65%</td>
</tr>
<tr>
<td>TAPA DE ASADO</td>
<td>3.48%</td>
</tr>
<tr>
<td>MARRAS DE JUGO</td>
<td>20.08%</td>
</tr>
<tr>
<td>PIZZA CONCORDA</td>
<td>2.54%</td>
</tr>
<tr>
<td>NABESA</td>
<td>2.45%</td>
</tr>
<tr>
<td>TAPA DE ASADO</td>
<td>3.23%</td>
</tr>
<tr>
<td>MARRAS DE JUGO</td>
<td>24.89%</td>
</tr>
<tr>
<td>PIZZA CONCORDA</td>
<td>2.12%</td>
</tr>
<tr>
<td>NABESA</td>
<td>24.23%</td>
</tr>
<tr>
<td>TAPA DE ASADO</td>
<td>129.03%</td>
</tr>
<tr>
<td>MARRAS DE JUGO</td>
<td>25.47%</td>
</tr>
</tbody>
</table>

Fuentes: www.InflacionVerdadera.com

(INCLUYE PRECIOS QUE ENTRAN Y SALEN DE OFERTA)
DON’T LIE TO ME, ARGENTINA

“…for the vast majority of Argentines, the indices compiled by INDEC…have been grossly manipulated.”

— La Nación Newspaper, March 10, 2008
DON’T LIE TO ME, ARGENTINA

“...we have decided to drop INDEC’s figures entirely...”

—The Economist, February 25, 2012
DON’T LIE TO ME, ARGENTINA

“… a declaration of censure… the Board called Argentina to... address the inaccuracy of the CPI.”

—IMF, February 2013
EVENTUALLY GOVERNMENT TELLS THE TRUTH

Source: PriceStats, INDEC, The Billion Prices Project
THE BILLION PRICES PROJECT: MEASUREMENT AND RESEARCH
The Billion Prices Project

- Academic initiative to collect and use online price data for economic measurement and research
  - Daily prices since 2008
  - From hundreds of large multi-channel retailers
  - In over 60 countries

1. Use scraping technology
2. Connect to hundreds of online retailers every day
3. Find individual products
4. Store key item information in a database
5. Calculate inflation and other statistics

Date
Item
Price
Description

www.thebillionpricesproject.com
### Micro-Price Data: Advantages and Disadvantages

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

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.

# The Billion Prices Project - Research

<table>
<thead>
<tr>
<th>Research Topic</th>
<th>Papers</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Macroeconomics</strong></td>
<td></td>
</tr>
<tr>
<td>Inflation Measurement</td>
<td>Cavallo (2020a)</td>
</tr>
<tr>
<td>Online Price Indices</td>
<td>Cavallo (2013), Cavallo &amp; Rigobon (2016)</td>
</tr>
<tr>
<td>Crowdsourcing with phones in Venezuela</td>
<td>Cavallo (2020b)</td>
</tr>
<tr>
<td>Online Pricing and Price Discrimination</td>
<td>Cavallo (2017), Cavallo (2018a)</td>
</tr>
<tr>
<td>Price Stickiness</td>
<td>Cavallo (2018b)</td>
</tr>
<tr>
<td><strong>International Economics</strong></td>
<td></td>
</tr>
<tr>
<td>LOP and market segmentation</td>
<td>Cavallo, Neiman, Rigobon (2014,2015)</td>
</tr>
<tr>
<td>Purchasing Power Parities</td>
<td>Cavallo, Diewert, Feenstra, Inklaar &amp; Timmer (2018)</td>
</tr>
<tr>
<td>International Wage Comparisons</td>
<td>Cavallo, Cravino, &amp; Drenik (2019)</td>
</tr>
<tr>
<td>Tariff passthrough</td>
<td>Cavallo, Neiman, Gopinath, &amp; Tang (2020)</td>
</tr>
</tbody>
</table>
Daily Inflation Measurement

- 2008 → Daily index for Argentina (www.inflacionverdadera.com)
- 2010 → Daily index for the US on the BPP website
- 2011 - now → PriceStats collects the micro data and publishes daily inflation in 23 countries in real-time (3-day lag).

Source: www.pricestats.com
Daily Inflation Measurement

Source: PriceStats - State Street Global Markets Research, and BLS
Daily Inflation Measurement

Figure 3
US Consumer Price Index around the Bankruptcy of Lehman Brothers

Source: Authors using online price index computed by PriceStats and the Consumer Price Index from the US Bureau of Labor Statistics.
Note: The figure highlights the events around the bankruptcy of Lehman Brothers, the fourth-largest investment bank in the United States, during September 2008.

Daily Inflation Measurement

Source: PriceStats - State Street Global Markets Research, and BLS
Applications in Forecasting

- Bertolotto (2019), Aparicio & Bertolotto (2020)

- Advantages for forecasting:
  - Frequency (daily) and speed of publication (3-day lag)
  - Provides information during crises and turning points, when traditional forecasting models based on lagged CPIs and low-frequency data tend to fail the most

- Real-time measurement helps to nowcast the present (1-2 months ahead)

- But for longer-term forecasting → we need to better understand shocks and inflation dynamics
Covid Inflation Dynamics

Source: PriceStats - State Street Global Markets Research, and BLS
Research related to Covid Inflation Dynamics

- Covid inflation was higher than measured by the CPI (Cavallo (2020) Inflation with COVID Consumption Baskets. NBER Working Paper No. 27352)


Updated results, all data and codes are available at) [https://projects.iq.harvard.edu/covid-cpi](https://projects.iq.harvard.edu/covid-cpi)
Research related to Covid Inflation Dynamics

- Covid inflation was higher than measured by the CPI (Cavallo (2020) Inflation with COVID Consumption Baskets, NBER Working Paper No. 27352)

- Online prices tend to react faster to aggregate shocks (Cavallo (2018) More Amazon Effects: Online Competition and Pricing Behaviors, Jackson Hole Symposium)
  - More frequent price changes + uniform pricing → more cost-shock passthrough

Table 6: Short-Run Pass-through into Walmart’s Prices (2016-2018)

<table>
<thead>
<tr>
<th></th>
<th>Found on Amazon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full Sample</td>
</tr>
<tr>
<td>Gas Prices (1 quarter)</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>Observations</td>
<td>191,690</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.17</td>
</tr>
<tr>
<td>Exchange Rate (2 quarters)</td>
<td>0.32</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td>Observations</td>
<td>191,690</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.17</td>
</tr>
</tbody>
</table>
Research related to Covid Inflation Dynamics

- Covid inflation was higher than measured by the CPI (Cavallo (2020) Inflation with COVID Consumption Baskets. NBER Working Paper No. 27352)

- Online prices tend to react faster to aggregate shocks (Cavallo (2018) More Amazon Effects: Online Competition and Pricing Behaviors, Jackson Hole Symposium)
  - More frequent price changes + uniform pricing → more cost-shock passthrough

- Retail cost-passthrough can be slow when shock is perceived to be temporary, and firms have other ways to adapt (Cavallo et al (2020) Tariffs Passthrough at the Border and at the Store: Evidence from US Trade Policy AER Insights)

- Covid supply disruptions are putting upward pressure on prices (Cavallo & Kryvtsov (2021) Stockouts and Prices during the Covid-19 Pandemic, Working Paper)
Supply Disruptions and Stockouts

- We measure product stock-outs in 7 countries for a subset of retailers showing out of stock information.

Stockouts and Prices

- Stockouts were positively correlated with US sectoral inflation in 2020

Product availability remains low

- Stockouts fell but the total number of varieties available for sale at these retailers is still ~20% below pre-pandemic levels


→ supply disruptions are still important in many retail categories and may continue to put upward pressure on prices for months
Conclusions

- Online data provide a unique measurement opportunity
  - Speed, frequency, details & customization, alternative to official sources

- Real-time price indices are useful for nowcasting inflation, particularly during times of crisis and shocks

- More importantly, better micro data can greatly improve our understanding of shock pass-through and longer-term inflation dynamics
Additional Slides
Are Online and Offline Prices Similar?

- Cavallo (2017) Are Online and Offline Prices Similar: Evidence from Large Multi-Channel retailers, American Economic Review

- 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.

- We then scraped the online price for the same good-retailer (within 7 days).

Figure 1: Screenshots from BPP App for Android Phones
Prices are identical ~70% of the time

### Table 3: Country - Level Differences

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

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.

### Table 4: Sector - Price Level Differences

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

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

Practice some web scraping

- It is not as hard as you think:
  - All webpages use an HTML code with \texttt{tags} that provide a stable structure to identify the data $\rightarrow$ you can teach a software to recognize them
  - Many alternative tools: R, python, specialized software, scraping services

- Similar Steps:
  1) Create a template for parsing the HTML code
  2) Create a list of URLs with the relevant data
  3) Run scraper, analyze data, debug

- Not just for retail prices:
  - Wage postings (Cavallo, Cravino & Drenik 2019), stockouts (Cavallo & Kryvtsov 2020), real estate listings, customs data, wholesale prices, or any structured data that can be found online
US Annual Inflation

US PriceStats Daily Country Inflation Index (% yoy)

- PriceStats
- *CPI-U

1m 3m 6m YTD 1y All

0% 2% 4% 6%

References

- Eichengreen, B., 2015. An Economics to Fit the Facts, Project Syndicate.