

The Billion Prices Project: Using Online data for Measurement and Research

Alberto Cavallo

Associate Professor

MIT

acavallo@mit.edu

Motivation

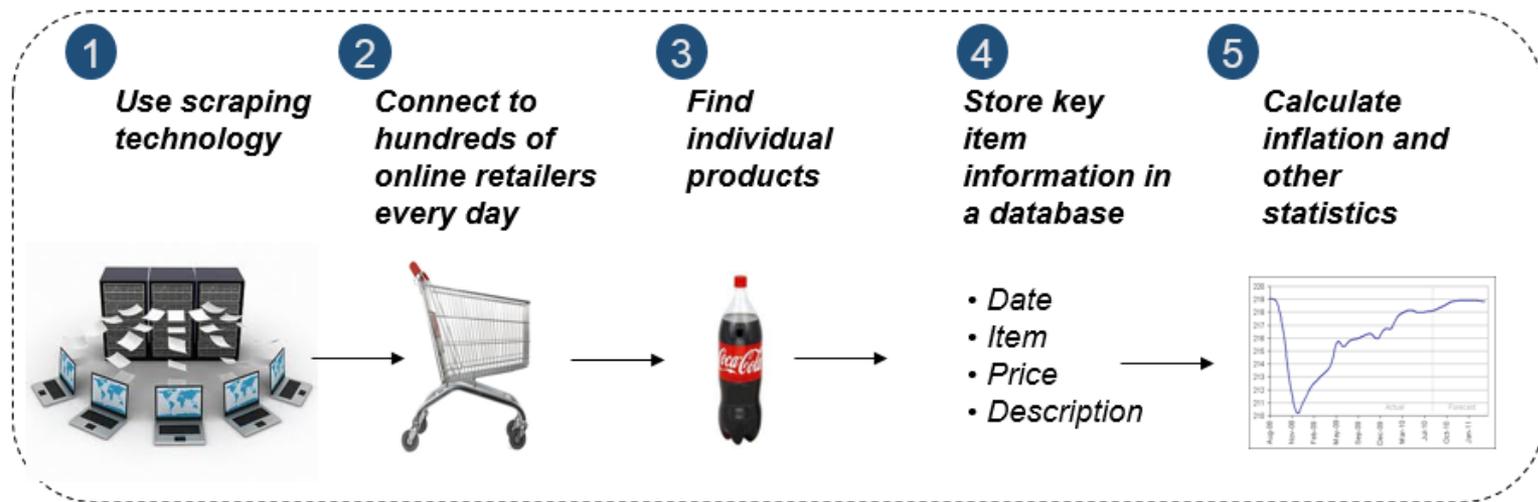
➤ Quote from Griliches (AEA 1985), on the “uneasy alliance” between economists and data

“... 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.”

- ## ➤ “Big Data” opportunity for macro/international economists:
- ❑ Get directly involved in data collection
 - ❑ Build datasets to fit specific measurement and research needs
 - ❑ Avoid biases and empirical issues common in traditional datasets

The BPP and PriceStats

- The Billion Prices Project at MIT: academic initiative started in 2008 to use online data for economic research
- PriceStats: company founded in 2011 to collect online data and compute daily price indices in over 20 countries



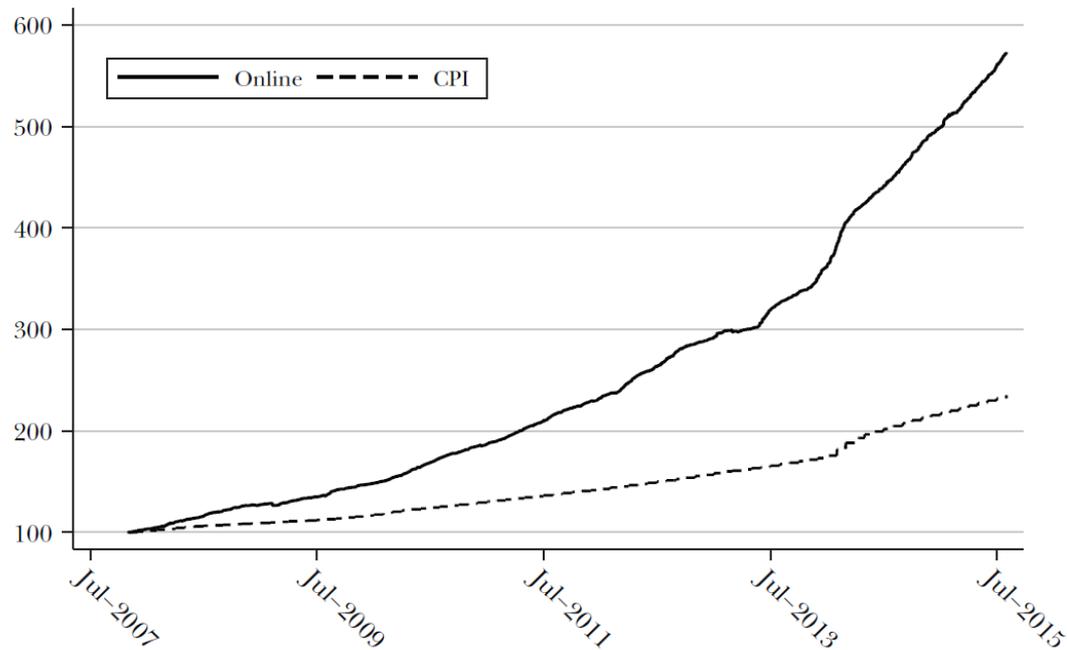
Our uses of online Data

- Measure inflation and other price statistics
- Research in Macro
 - ❑ Price dynamics
 - ❑ Market segmentation
- Research in International Economics
 - ❑ LOP and PPP
 - ❑ Pass-through and border effects
 - ❑ Real Exchange Rate dynamics

➤ 2007: Argentina's manipulation of inflation data

Figure 1
Argentina

A: Price index

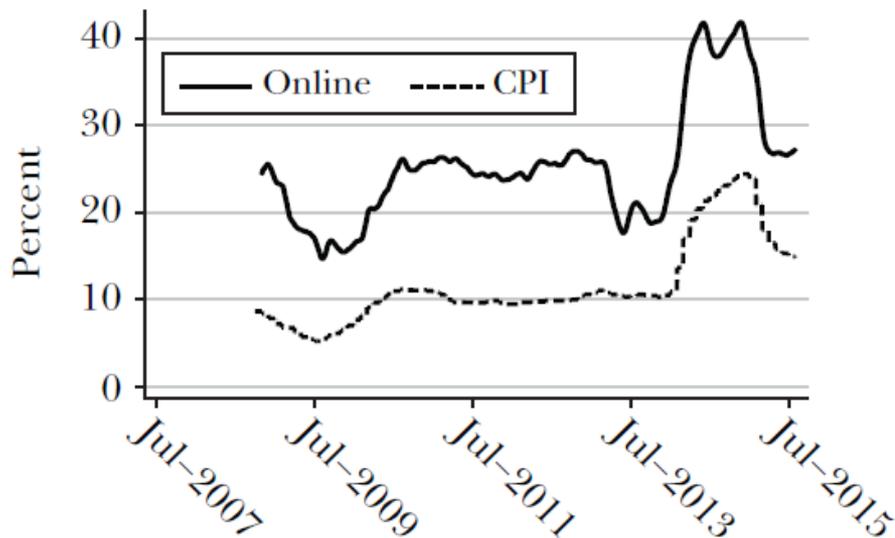


Source: Cavallo and Roberto (2016). "The Billion Prices Project: Using Online Data for Measurement and Research." *Journal of Economic Perspectives*.

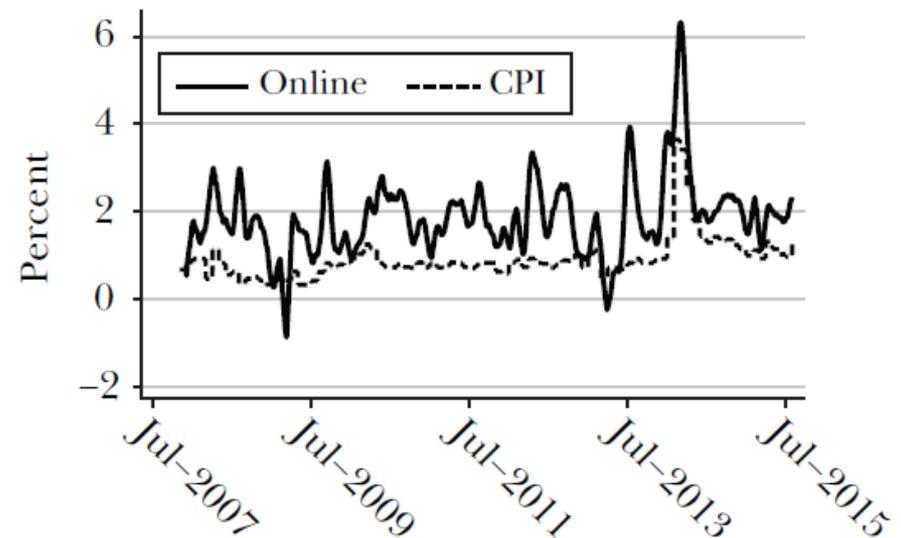
Timeline

➤ 2007: Argentina's manipulation of inflation data

B: Annual inflation rate



C: Monthly inflation rate



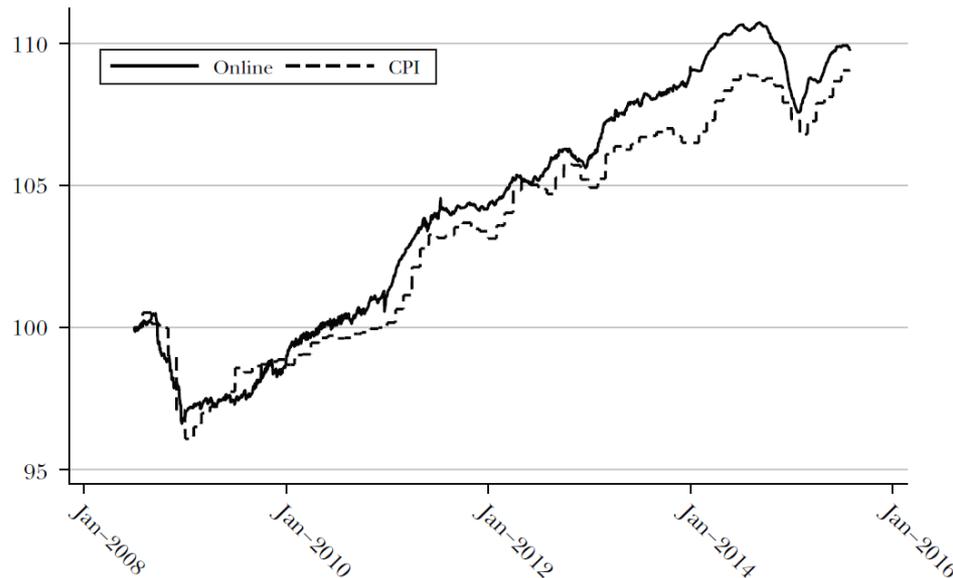
Source: Cavallo and Roberto (2016). "The Billion Prices Project: Using Online Data for Measurement and Research." *Journal of Economic Perspectives*.

Timeline

- 2007: Argentina's manipulation of inflation data
- 2008-9: BPP expands data collection to other countries
- 2010: Daily price index in the US

Figure 2
United States

A: Price index

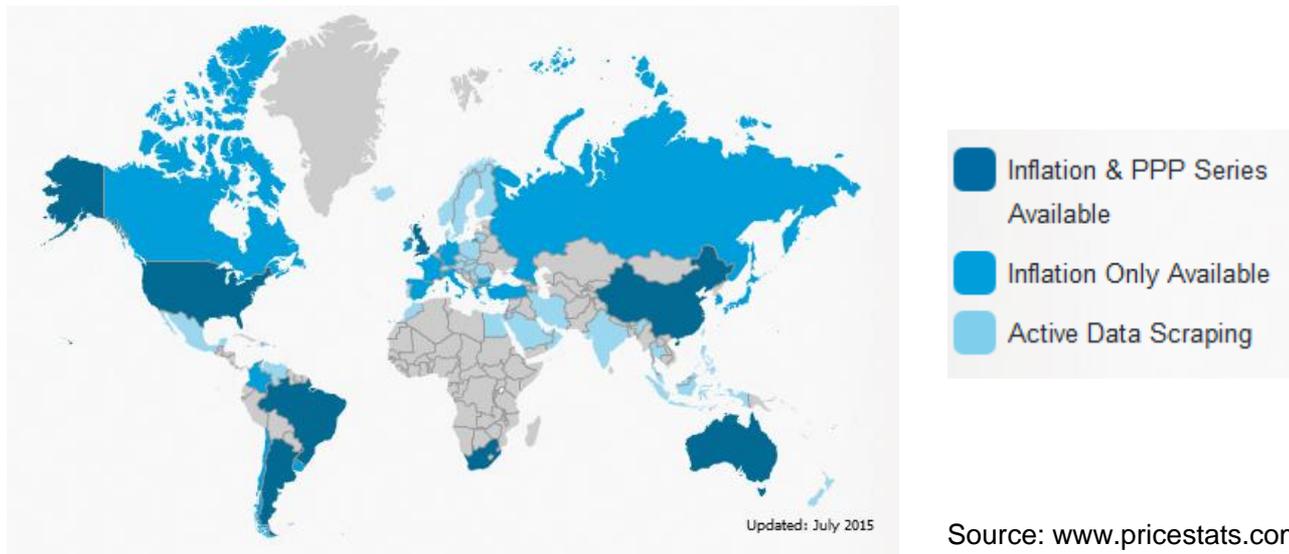


- ❑ Congruence
- ❑ Differences
- ❑ Anticipation

Source: Cavallo and Roberto (2016). "The Billion Prices Project: Using Online Data for Measurement and Research." *Journal of Economic Perspectives*.

Timeline

- 2007: Argentina's manipulation of inflation data
- 2008-9: BPP expands data collection to other countries
- 2010: Daily price index in the US
- 2011: PriceStats starts publishing daily price indexes in 22 countries (3-day lag, State Street)



Online Prices

➤ Advantages

- ❑ Cost
- ❑ Frequency (daily)
- ❑ All products for sale (census)
- ❑ Uncensored price spells
- ❑ Lots of countries, remotely collected
- ❑ Real-time

➤ Disadvantages

- ❑ Mostly goods, few services
- ❑ Fewer retailers than traditional CPI data
- ❑ No data in countries with no online retailers (eg Africa)
- ❑ No quantities, available in Scanner Data

Are Online and Offline Prices Similar?

- Cavallo (2016) “Are Online Prices Similar? Evidence from Multi-channel Retailers”, NBER
- Crowdsourced workers scan barcodes, enter prices, and email data files

amazon mechanical turk
Artificial Artificial Intelligence

Elance upwork™
formerly oDesk



Figure 1: Screenshots from BPP App for Android Phones

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

Online-Offline Price Comparison

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 17 Mar 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.

International Relative Prices

➤ PriceStats PPP Series

- ❑ Online prices
- ❑ Daily frequency
- ❑ 350 narrow product categories
- ❑ 30 thousand individually matched items
- ❑ 7 countries
- ❑ Food, fuel, and electronics
- ❑ Sector and Country-Level Indices

Illustration: Coke Prices



Compare prices for a bottle of Coke across countries

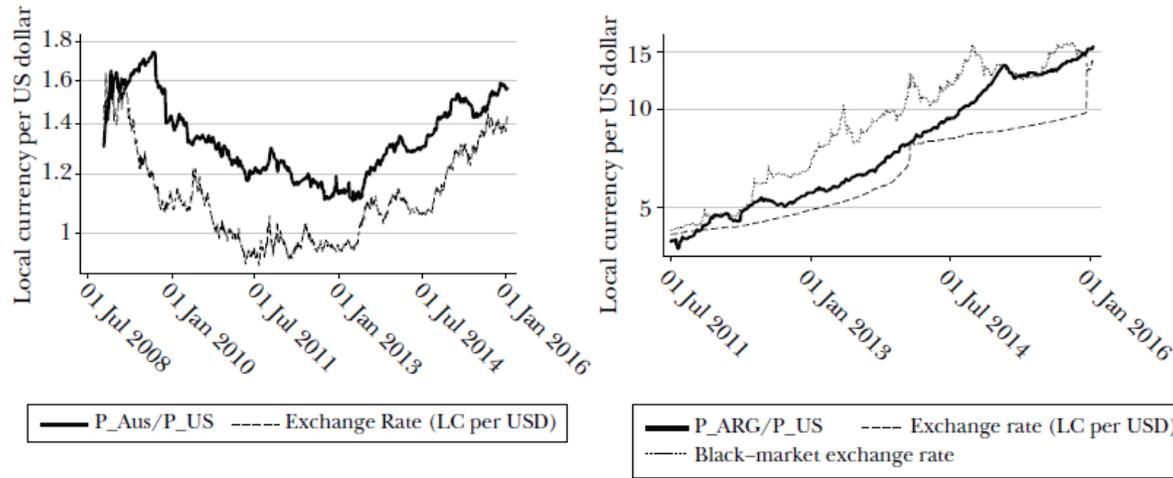
Repeat for hundreds of products

Compute daily RERs at different levels of aggregation

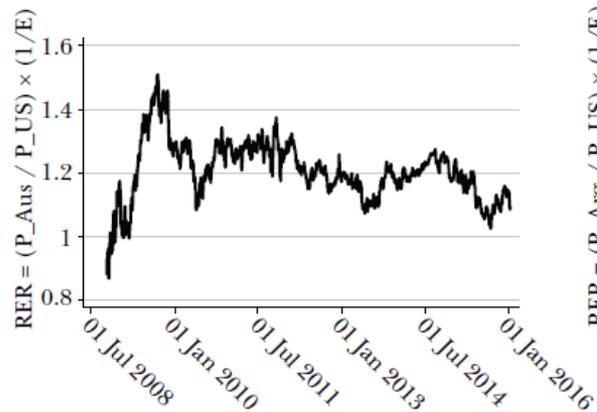
Relative Prices and Exchange Rates

Figure 9
Relative Prices and Exchange Rates

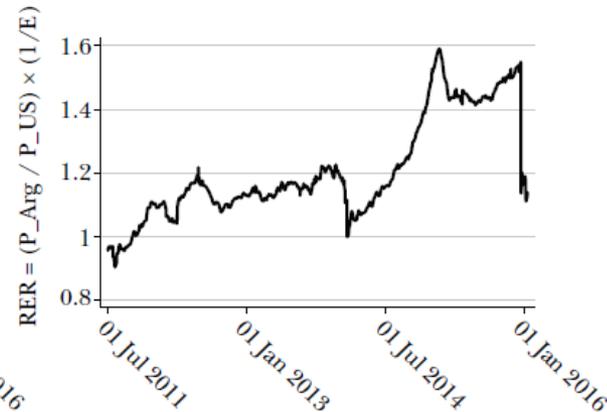
A: Australia–Relative prices and nominal exchange rate B: Argentina–Relative prices and nominal exchange rate



C: Australia–Real exchange rate



D: Argentina–Real exchange rate



Comparison to ICP 2011

➤ 3-digit COICOP

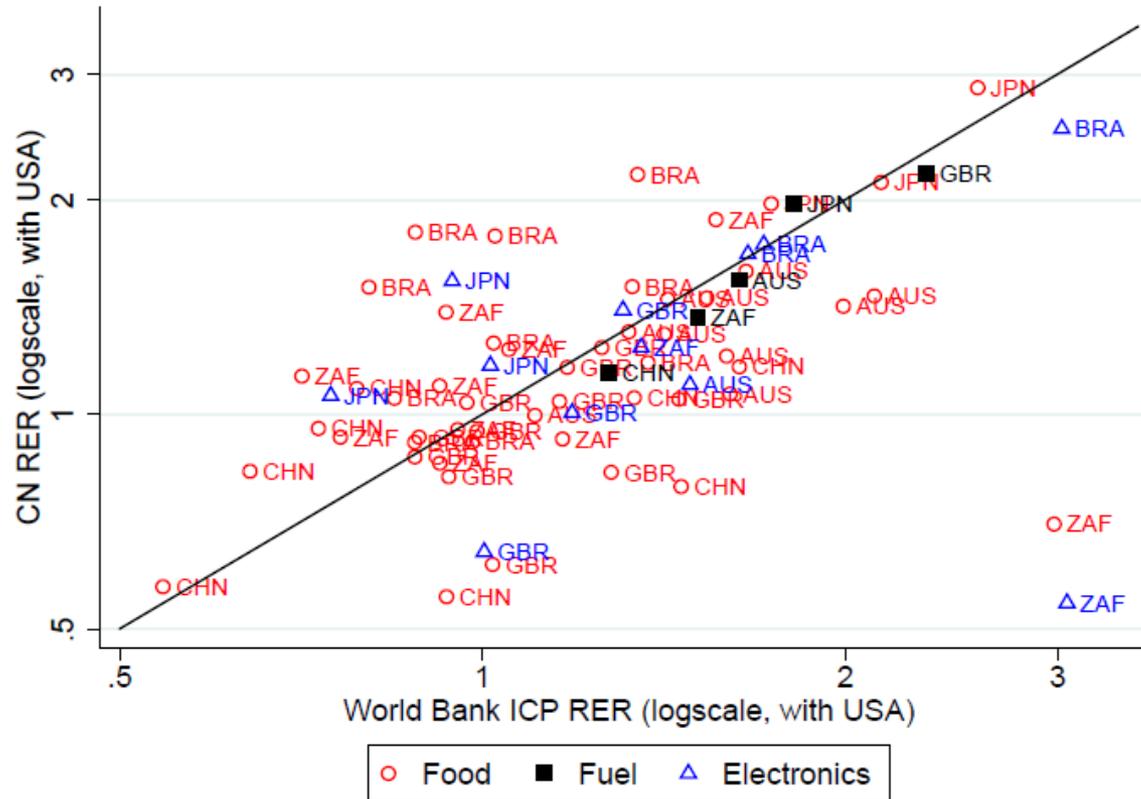


Figure 1: 3-Digit Real Exchange Rate Levels Relative to the United States, ICP vs. CN

Source: Cavallo and Neiman (2016) "Real Exchange Rate Behavior: Evidence from Online Retailers in Nine Countries." *Working Paper*.

Comparison to ICP 2011

➤ 1-digit COICOP

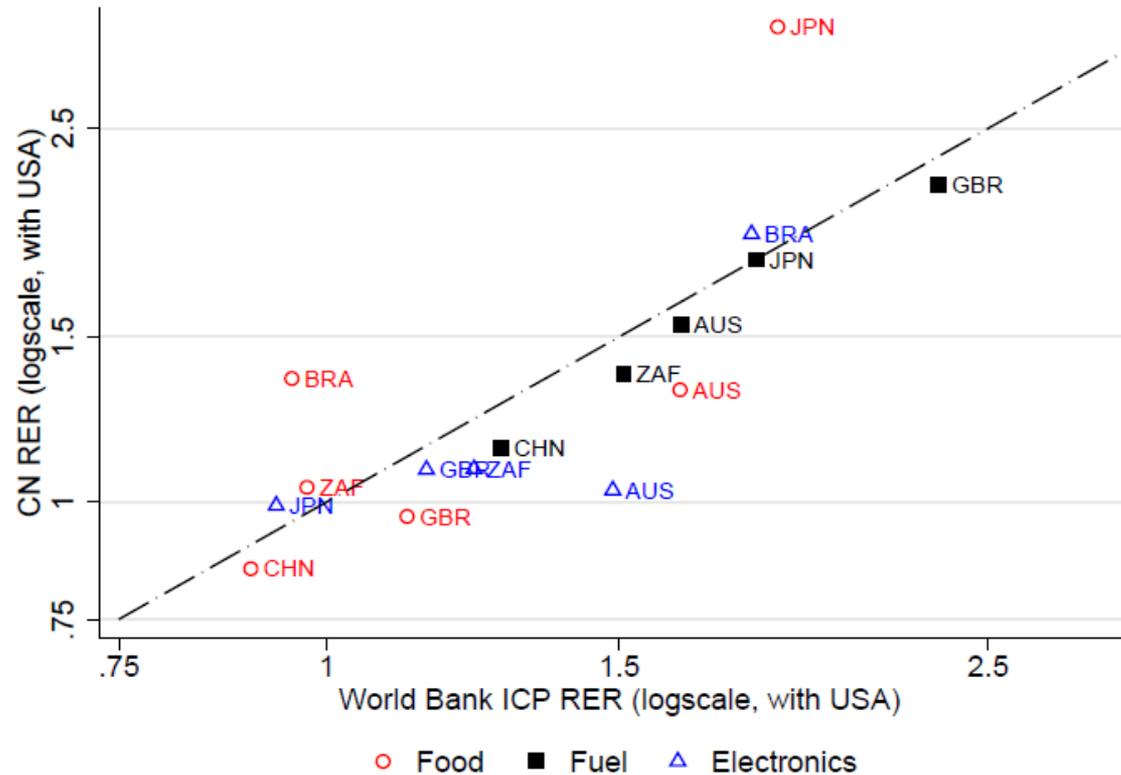


Figure 2: 1-Digit Real Exchange Rate Levels Relative to the United States, ICP vs. CN

Source: Cavallo and Neiman (2016) "Real Exchange Rate Behavior: Evidence from Online Retailers in Nine Countries." *Working Paper*.

Conclusions

- “Big Data” in economics → *measurement* opportunity
 - ❑ New data collection tools (web, sensors, phones, gps, satellites)
 - ❑ Build customized datasets that fit specific measurement and research needs

- Online Price data
 - ❑ Advantages for inflation and PPP measurement
 - ❑ Increase the quantity and quality of micro price data available for research, helping us re-evaluate old empirical puzzles or address questions that could not be answered before

THANK YOU
