

# **The Billion Prices Project**

## **Research and Inflation Measurement Applications**

**Alberto Cavallo**

MIT & NBER

IMF Statistical Forum

November 2015

# Micro-Price Data in Macroeconomics

- Data Sources
  - Statistical Offices (CPI, PPI)
  - Scanner Data (eg. Nielsen)
  - Online Data (eg. Billion Prices Project)
- Uses
  - Inflation and other economic indicators
  - Research in Macroeconomics
    - Price Dynamics (Price Stickiness, Real Rigidities)
    - Market Segmentation
  - Research in International Economics
    - Pass-through and Border Effects
    - Law of One Price and PPP
    - Real Exchange Rates

# Each Data Source has Advantages and Disadvantages

## CPI Data

- Purpose: measure inflation

| Advantages  | Disadvantages   |
|---|---|
| <ul style="list-style-type: none"><li>• Representative sample<ul style="list-style-type: none"><li>• carefully-chosen goods</li><li>• many retailers and locations</li></ul></li><li>• Long Time Series</li><li>• Collection of posted prices in stores</li></ul> | <ul style="list-style-type: none"><li>• Very costly to collect and access</li><li>• Low frequency (monthly)</li><li>• Limited number of goods and varieties</li><li>• Some unit values and imputed prices</li><li>• Difficult international comparisons</li></ul> |

# Each Data Source has Advantages and Disadvantages

## Scanner Data

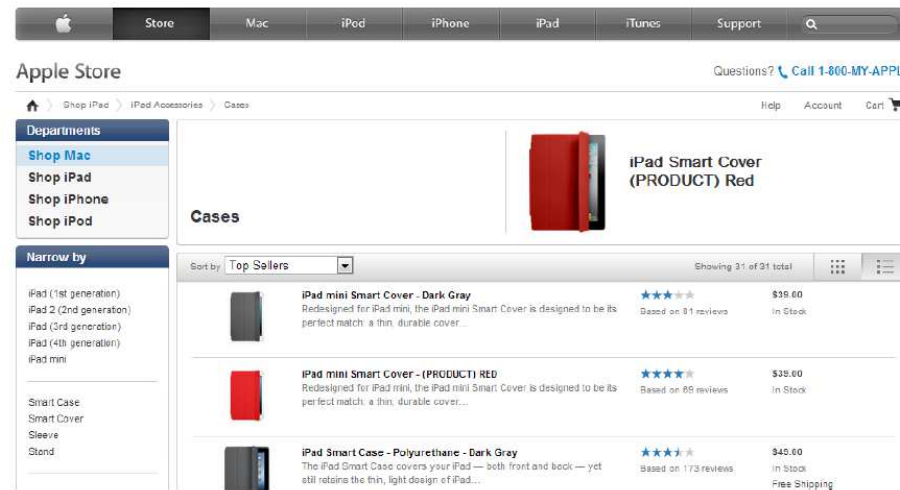
- Purpose: marketing analytics (eg. market shares)

| Advantages   | Disadvantages   |
|--|---|
| <ul style="list-style-type: none"><li>• Granularity<ul style="list-style-type: none"><li>• Some product details for all goods <i>sold</i></li></ul></li><li>• Transaction data<ul style="list-style-type: none"><li>• Contains quantities and sometimes costs</li></ul></li><li>• Frequency (weekly)</li></ul> | <ul style="list-style-type: none"><li>• High cost to collect/acquire</li><li>• Limited coverage (supermarkets, department stores)</li><li>• Data characteristics vary greatly depending on provider, location, time period, etc.</li><li>• Extremely difficult to compare internationally</li><li>• Unit values and time-averages (eg: prices are often calculated as sales/quantity in a week)</li></ul> |

# Each Data Source has Advantages and Disadvantages

## Online Data

- Can be collected using automated web-scraping software
- Every day, a *robot* downloads a public webpage, analyses its HTML code, extract price data, and stores it in a database



```
<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 -->
.....
```

# Each Data Source has Advantages and Disadvantages

## Online Data

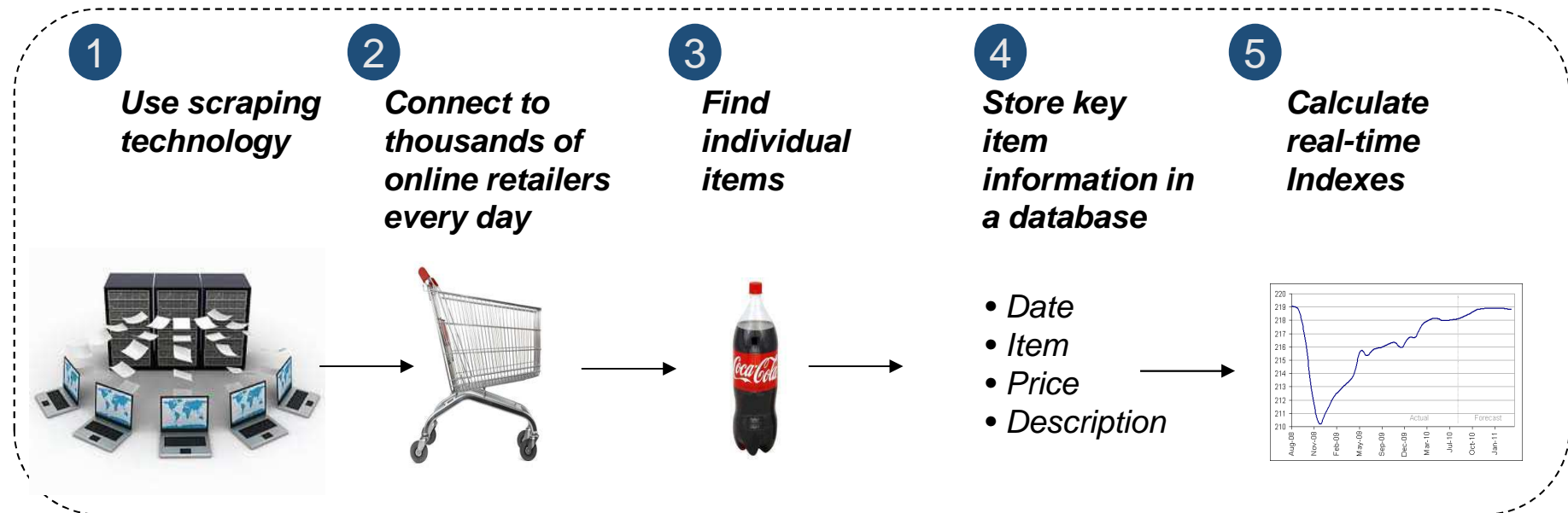
| Advantages  | Disadvantages  |
|---|--|
| <ul style="list-style-type: none"><li>• Frequency (daily)</li><li>• Cheap to collect (but complicated)</li><li>• Granularity<ul style="list-style-type: none"><li>• All product details (brands, size, anything shown online)</li><li>• All goods and varieties available for sale (census)</li><li>• New goods automatically sampled</li></ul></li><li>• Easier to compare internationally</li></ul> | <ul style="list-style-type: none"><li>• Fewer retailer and locations than CPI</li><li>• Short time series</li><li>• Not all categories of goods and services are online (not yet)</li><li>• Online and Offline prices may behave differently</li></ul> |

# Macro and International Research

- Billion Prices Project at MIT
  - Academic initiative to collect and use micro-price data for economic research
  - Web scraping to collect price information from large retailers that sell online
  - Data collected since 2008 from hundreds of retailers in over 50 countries.
- Our Research is focused on
  - Inflation Measurement & Forecasting
  - Pricing Dynamics (eg. Stickiness)
  - Real Exchange Rate and PPP

# BPP and Daily Inflation Measurement

- In 2008 → daily price index for Argentina
- In 2010 → daily price index for the US on the BPP website
- Since 2011, PriceStats has been publishing daily inflation indices in 22 countries in real-time (3-day lag).

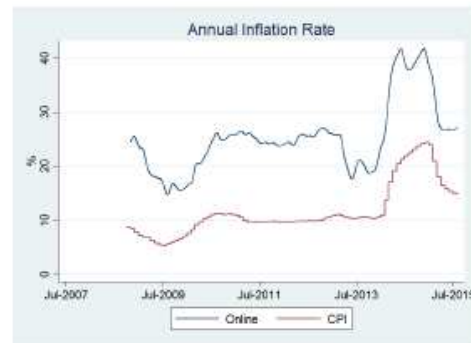




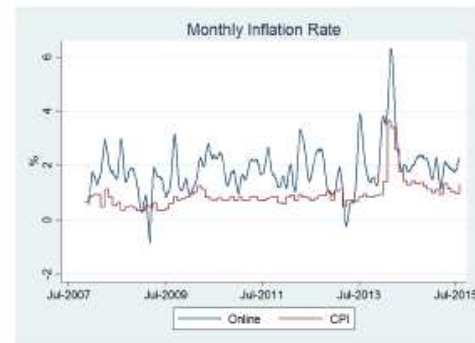
# The case of Argentina



(a) Price Index



(b) Annual Rate

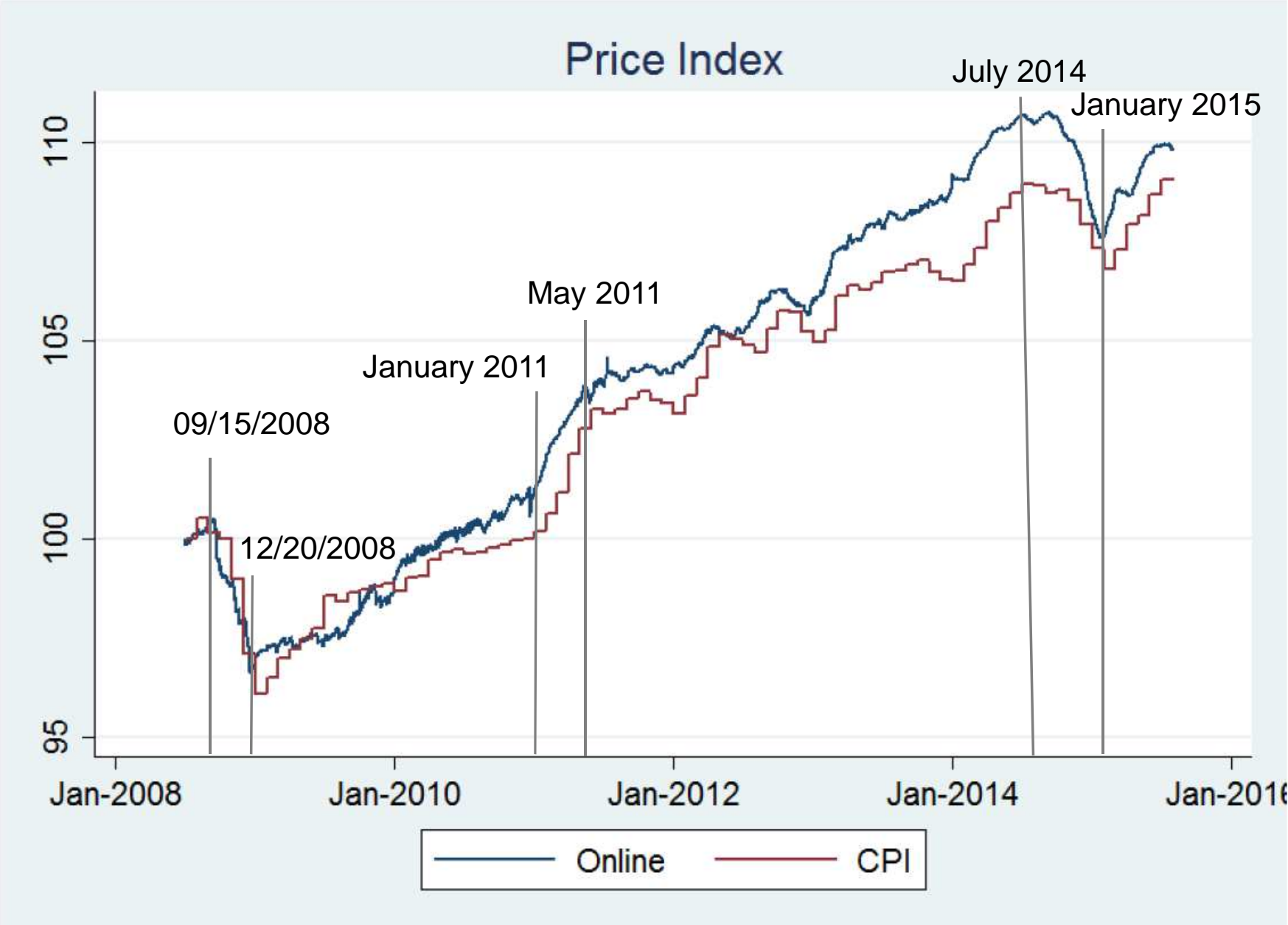


(c) Monthly Rate

Figure 2: Argentina

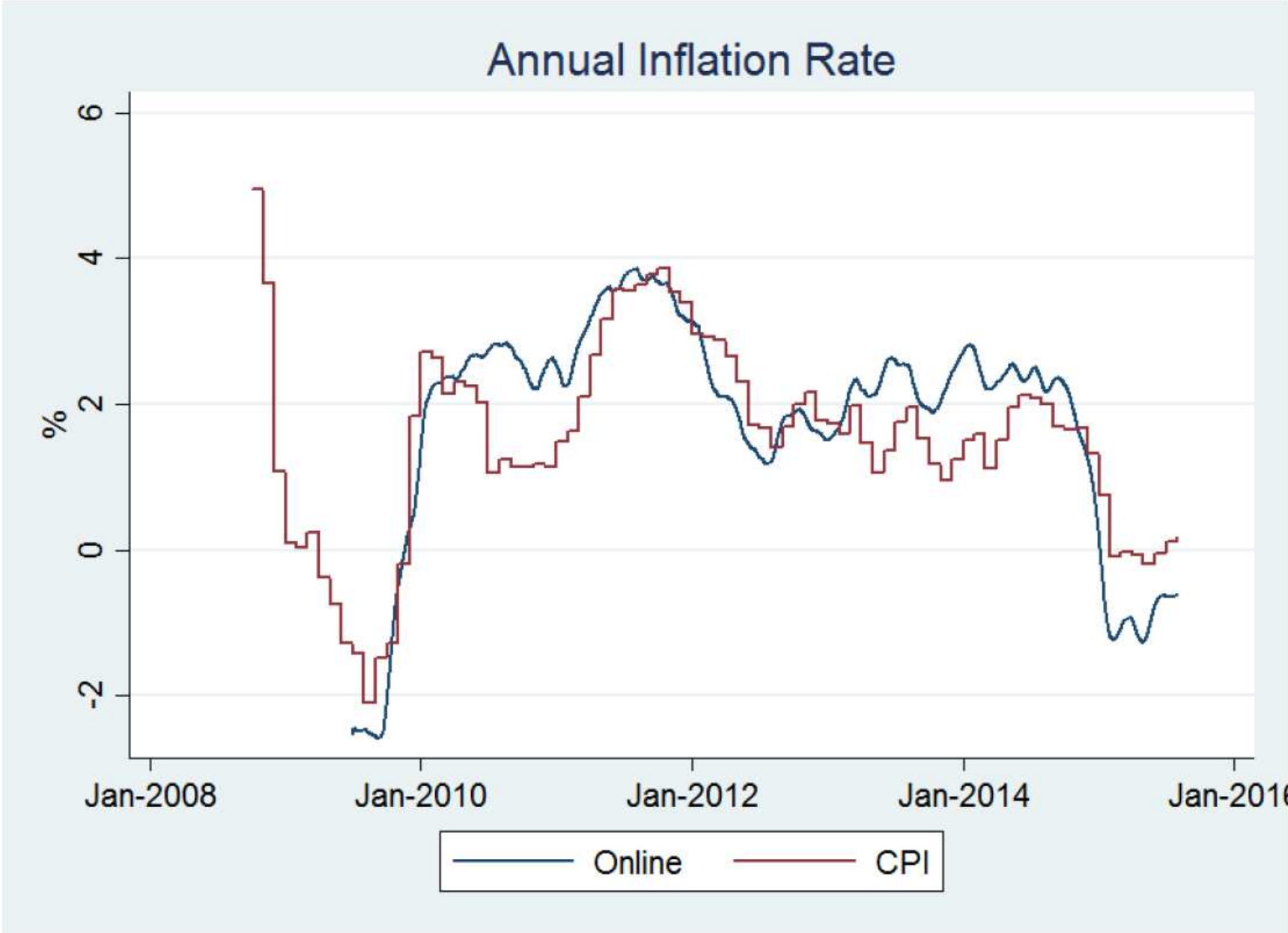
Source: Cavallo & Rigobon (2015) "The Billion Prices Project".

# US Daily Price Index



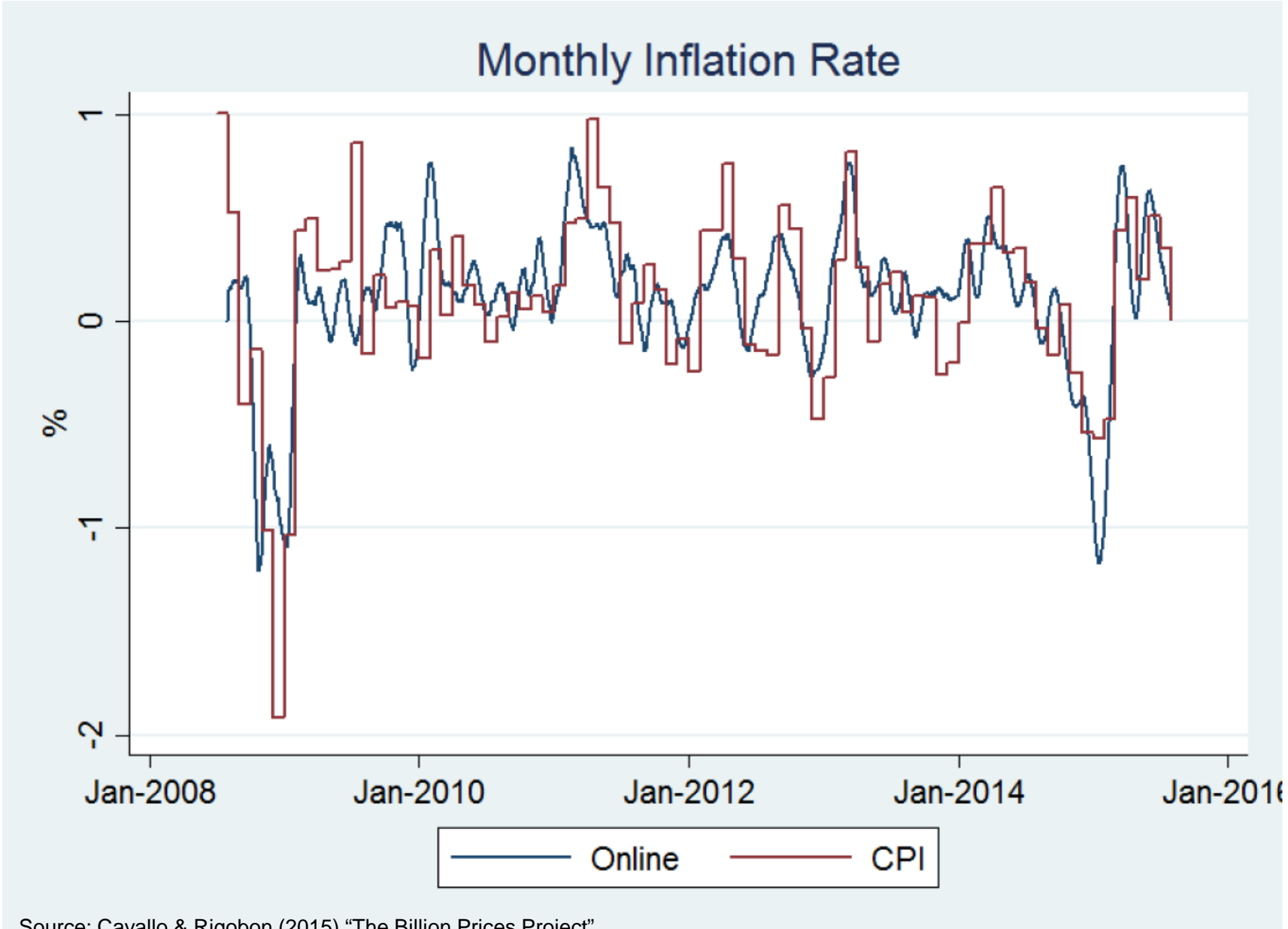
Source: Cavallo & Rigobon (2015) "The Billion Prices Project".

# US Annual Inflation



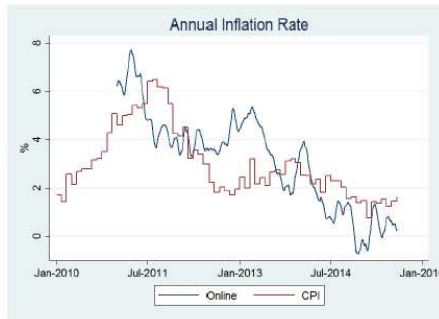
Source: Cavallo & Rigobon (2015) "The Billion Prices Project".

# US Monthly Inflation

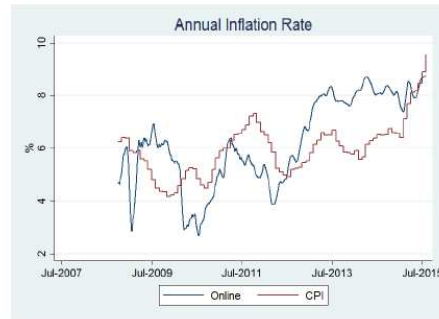


Source: Cavallo & Rigobon (2015) "The Billion Prices Project".

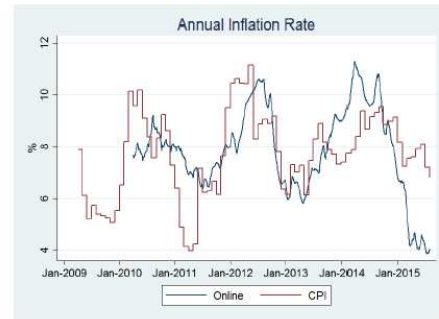
# Developing vs Developed Countries



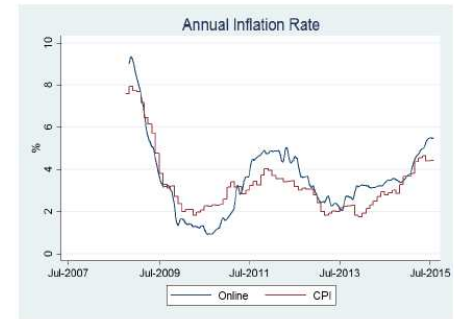
(a) China



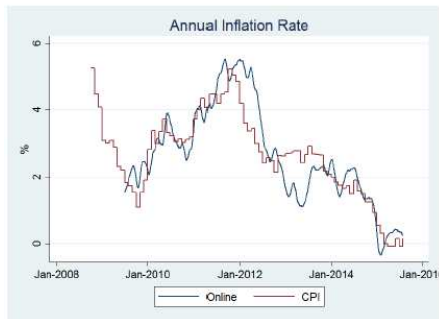
(b) Brazil



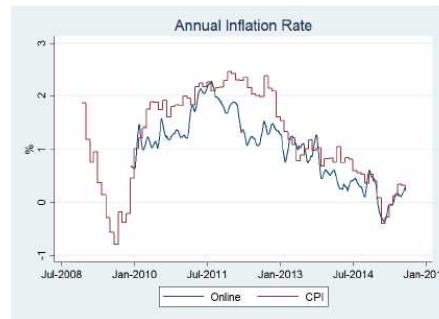
(c) Turkey



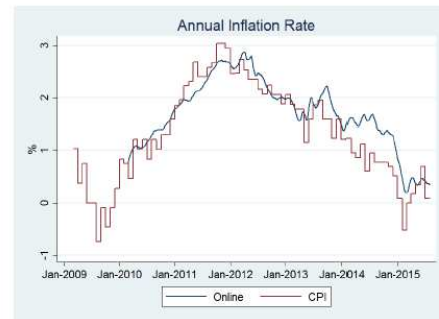
(d) Colombia



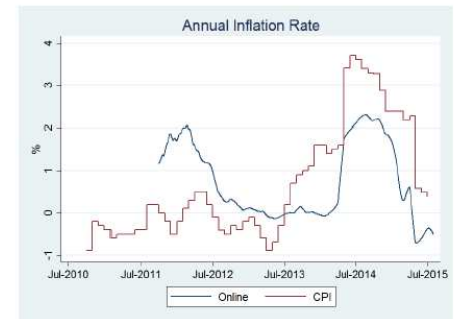
(e) UK



(f) France



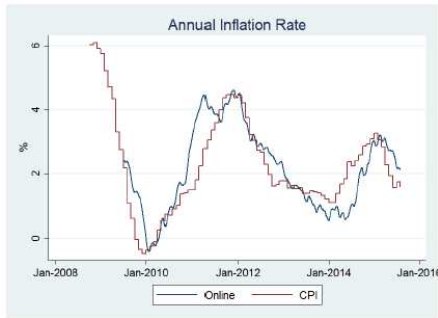
(g) Germany



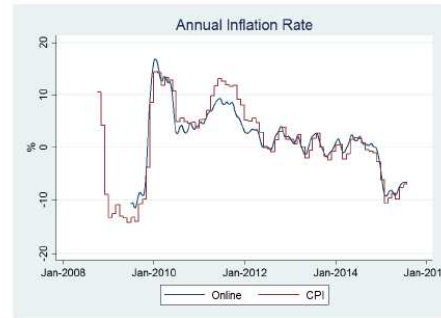
(h) Japan

Figure 5: Online vs CPI Annual Inflation Rates

# Sectors vs Global Aggregates



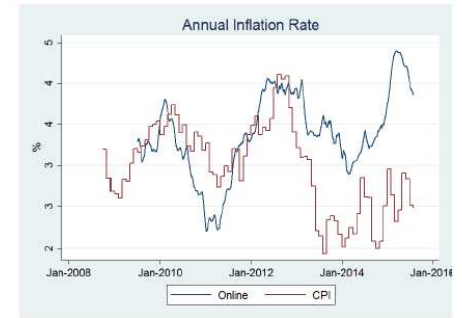
(i) USA Food



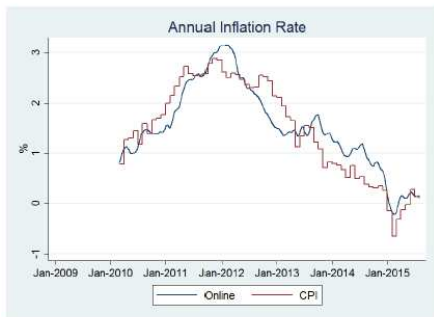
(j) USA Fuel



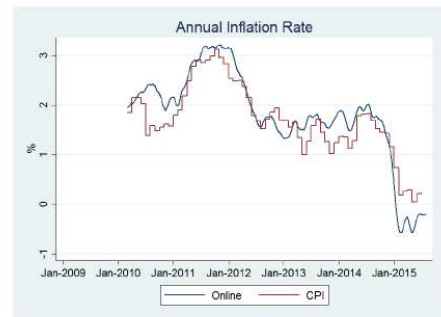
(k) USA Electronics



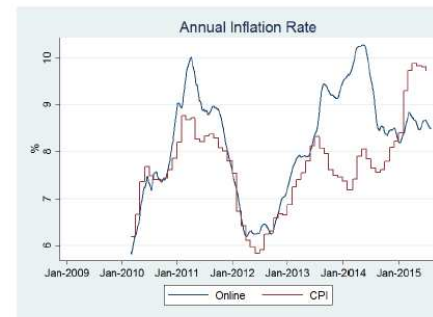
(l) USA Medical Care



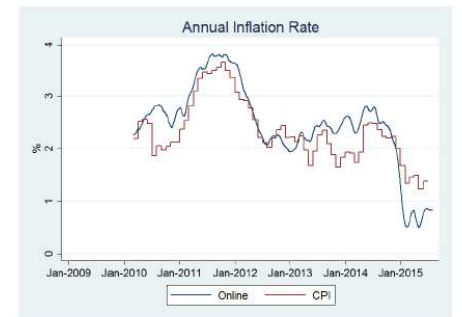
(m) Eurozone



(n) Developed M.



(o) Emerging M.

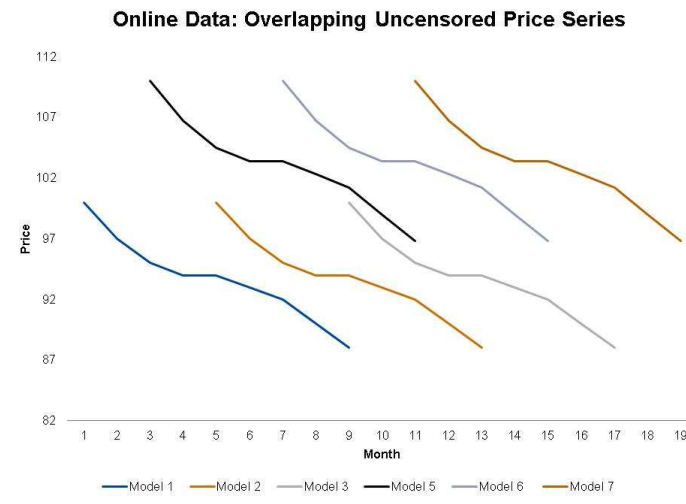
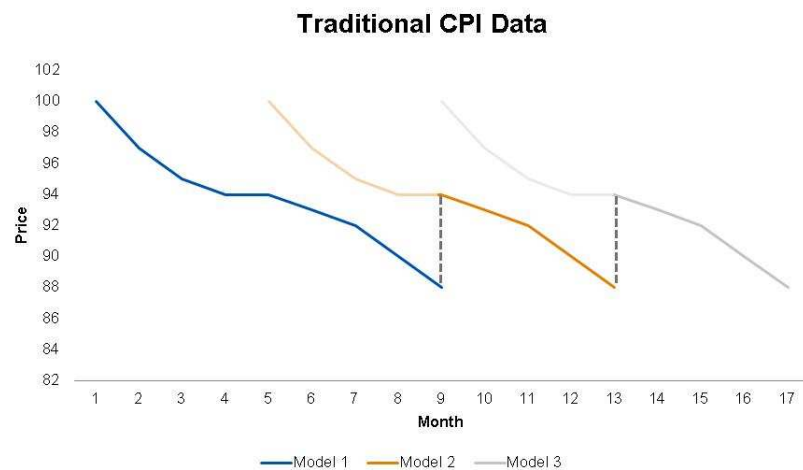


(p) World

Figure 5: Online vs CPI Annual Inflation Rates

# Differences with CPI : Quality Adjustments

- Many complex techniques applied in CPI methods, such as hedonic quality adjustments, are needed because the data has inherent limitations
- Online data has big data advantages:
  - uncensored spells (automatically included at introduction)
  - all varieties/models on display



# Differences with CPI : Quality Adjustments

- Simple indices can approximate the level and trend of CPI inflation in hedonic-adjusted categories (as suggested in Silver & Heravi (99), Aizcorbe, Corrado & Doms (2003))

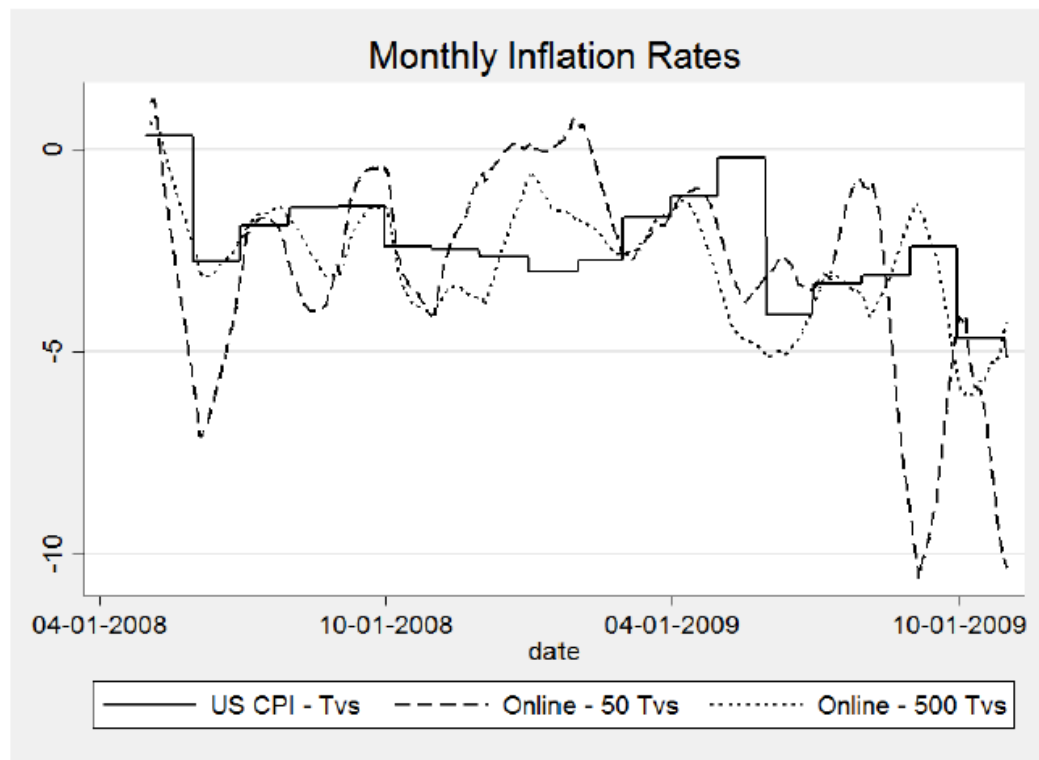


Figure 7: US CPI with Hedonics vs Online Jevons Index

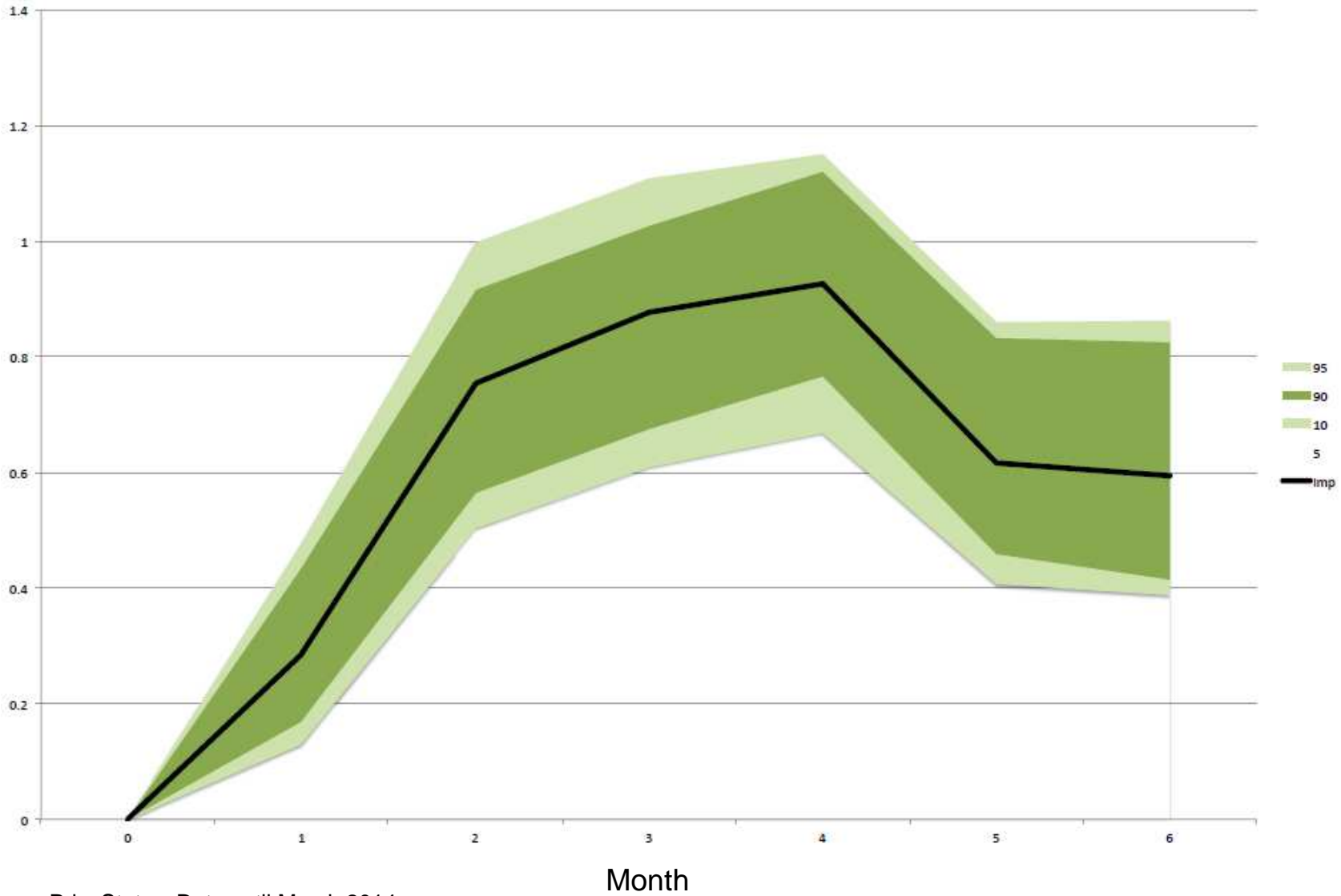


# Anticipation in Online Prices

- Online prices tend to anticipate changes in CPI inflation trends.
  - More than just quicker access to data
  - Online prices tend to *react faster* to shocks.
  - Why?
    - Lower adjustment (or menu) costs
    - Online shoppers may be less sensitive to price changes
    - More intense and transparent competition
- We can study the link between online data and CPIs using simple VARs.
  - VAR regressions with  $\Delta$  CPI on the LHS and lags of  $\Delta$  CPI and  $\Delta$ OPI on the RHS (monthly data)
  - Impulse responses show the impact of a 1% shock in online series (OPI) on future CPI (reflecting additional information not contained in lagged CPI)

# Impulse Response USA

## Cummulative IRF - 1% Shock to Online Aggregate Inflation



Source: PriceStats – Data until March 2014

# Online vs Offline Prices

- Are online prices representative ?
  - No
    - Online sales are still only about 10% of retail sales in developed countries
  - Yes
    - The `online store` is effectively the *largest* store for 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
    - Even if online transactions are rare, the online price can be a good proxy for the offline price
      - Close matching in price indices
      - Cavallo (2015) : simultaneous data collection for 50 large retailers in 10 countries shows they tend to have either identical online and offline price levels.

# Some Research Applications

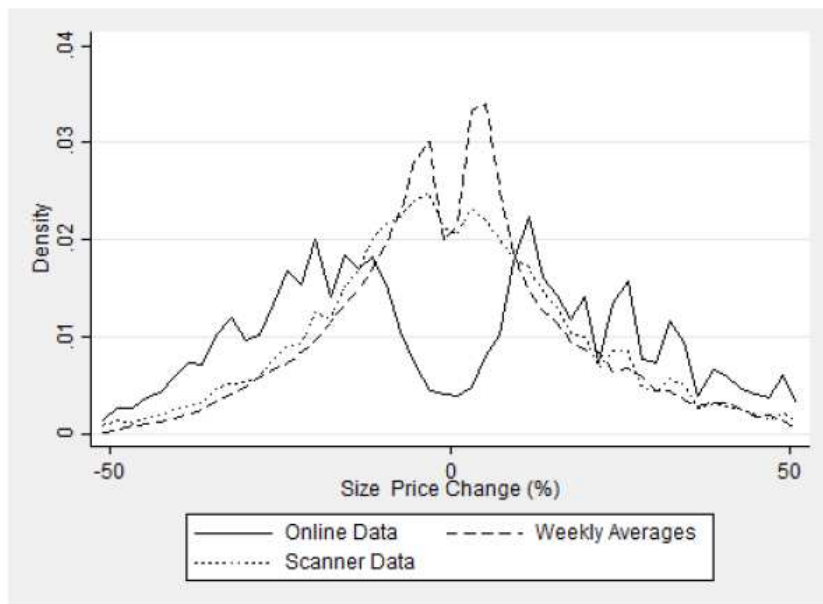
- Macro
  - Price Stickiness
- International
  - Law of One Price
  - Real Exchange Rates

# Price Stickiness

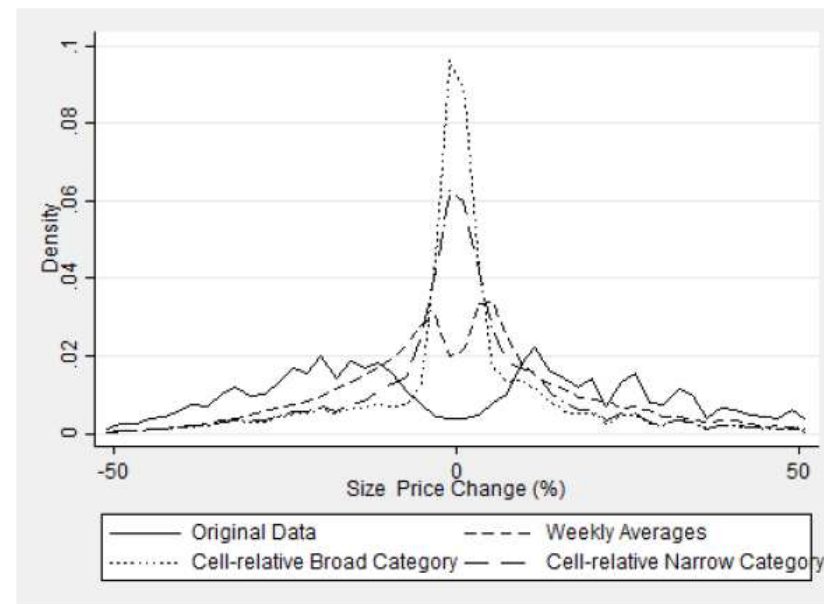
- I use online data to show that many of the previous findings in the literature were driven by measurement biases in CPI and scanner data:
  - Weekly average prices in scanner data
  - Cell-relative imputation for temporary missing prices in CPI data
- Effects:
  - Reduce the duration (stickiness) of price changes
  - Create spurious small changes → altering the shape of distribution
  - Cause downward sloping hazard functions

# Distribution of the Size of Price Changes

- The distribution is bimodal, with little mass near 0% → more consistent with adjustment/menu cost models



(a) Online vs Scanner



(b) Online vs CPI-simulation

Figure 9: The Distribution of the Size of Price Changes in the US

# Advantages of Online Data to Measure Stickiness

- Free from measurement bias
  - High frequency
  - Posted prices
  - All products
  - Uncensored price spells
- Available in multiple countries → no differences in methods, time periods or type of data.
- Real-time data, with no delays → policy applications

# Online data and International Relative Prices

- We use online prices to improve (i) coverage of countries, (ii) quality of matches, and (iii) frequency of observations relative to other proxies of the real exchange rate.
- QJE 2014 : RERs for goods sold by Apple, IKEA, H&M, Zara, and other global retailers  $\rightarrow$  RER = 1 in currency unions but not in pegs or floats.

Table 3: Absolute Value of Good-Level Log RER

|                     | All Stores | Apple | IKEA  | H&M   | Zara  |
|---------------------|------------|-------|-------|-------|-------|
| (i) Currency Unions | 0.076      | 0.023 | 0.129 | 0.020 | 0.102 |
| (ii) NER Pegs       | 0.116      | 0.085 | 0.145 | 0.119 | 0.115 |
| (iii) Floats        | 0.187      | 0.143 | 0.216 | 0.145 | 0.207 |

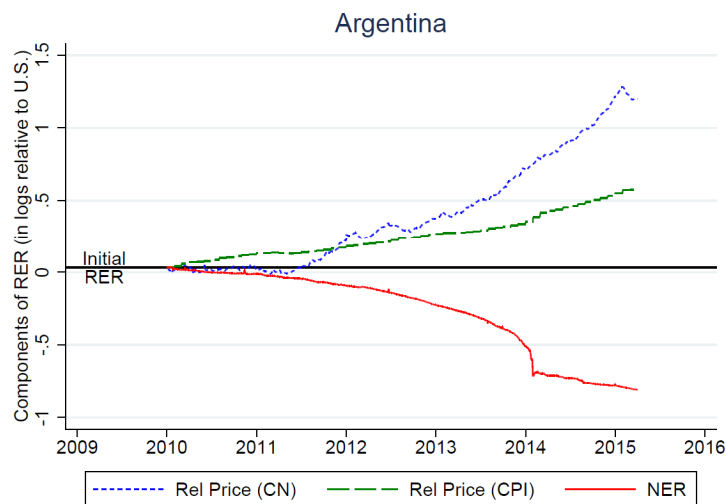
Notes: Extracted from Cavallo et al. (2014a). Unconditional means of the average (across weeks in the data) of the absolute value of each good's log RER, separated by the currency regime. We exclude the small number of observations where  $|q_{ij}| > 0.75$ . Currency regime definitions closely follow Ilzetzki et al. (2008) and are described in Cavallo et al. (2014a).



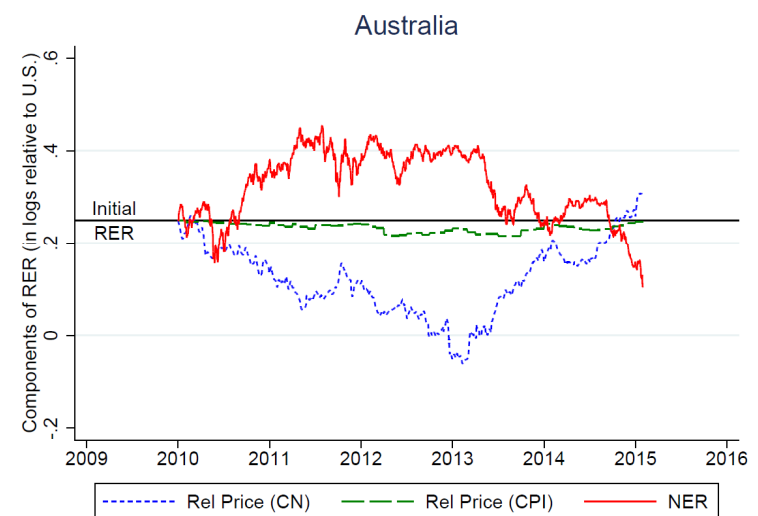
# Online data and International Relative Prices

- We are currently constructing high frequency RERs in levels → broader set of closely-matched across in the largest retailers in 7 countries

Components of RER, Comparison with CPI – Argentina



Components of RER, Comparison with CPI – Australia



- We find much faster adjustment of relative prices to nominal exchange rate fluctuations than observed with CPI-based RERs

# Conclusions

- Online data can also be a reliable source of information for inflation measurement
  - There are similarities, differences, and heterogeneity
  - ``Big data`` characteristics can greatly simplify measurement
  - Online prices tend to anticipate changes in inflation trends
- Online prices dramatically increase the quality of micro price data available for research.
  - Re-evaluate old empirical puzzles or address questions that could not be answered before

# Conclusions

- Quote from Griliches (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 : a revolution in *data collection*
  - Both opportunities and challenges

