Big Data in Economic Measurement: 
What a Billion Prices say about Inflation and Exchange Rates?

Roberto Rigobon
MIT Sloan, NBER, CSAC, CNStat
The world is not lacking of Data

It is lacking of Careful Empirical Analysis

It is lacking of Managerial Data Analysis
Different Types of Data

Data

- Designed
  - Survey
    - Aggregate
      - Public
      - Private
  - Administrative
  - Aspirational
  - Transactional

- Organic
## Types of Data

<table>
<thead>
<tr>
<th>Designed Data</th>
<th>Organic Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representative</td>
<td>Non-representative</td>
</tr>
<tr>
<td>Scope is limited</td>
<td>Volume, Velocity, Variety</td>
</tr>
<tr>
<td>Costly</td>
<td>Cheap</td>
</tr>
<tr>
<td>Difficult Access</td>
<td>Open</td>
</tr>
<tr>
<td>Intrusive</td>
<td>Non-intrusive</td>
</tr>
<tr>
<td>Problems</td>
<td>Survey</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Representativeness</td>
<td></td>
</tr>
<tr>
<td>Selection Bias</td>
<td></td>
</tr>
<tr>
<td>Reliability</td>
<td></td>
</tr>
<tr>
<td>Consistency</td>
<td></td>
</tr>
<tr>
<td>Transparency on Data Collection</td>
<td></td>
</tr>
<tr>
<td>Transparency on Data Treatment</td>
<td></td>
</tr>
<tr>
<td>Errors-in-variables</td>
<td></td>
</tr>
<tr>
<td>Aspirational (Transactional)</td>
<td></td>
</tr>
<tr>
<td>Private (as opposed to public)</td>
<td></td>
</tr>
<tr>
<td>Model Uncertainty</td>
<td></td>
</tr>
<tr>
<td>Behavioral Changes</td>
<td></td>
</tr>
<tr>
<td>Lead indicator</td>
<td></td>
</tr>
</tbody>
</table>
### Problems

<table>
<thead>
<tr>
<th></th>
<th>Survey</th>
<th>Estimation</th>
<th>Forecasting</th>
<th>Measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Representativeness</td>
<td>✓</td>
<td></td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Selection Bias</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Reliability</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Consistency</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Transparency on Data Collection</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Transparency on Data Treatment</td>
<td></td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Errors-in-variables</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Aspirational (Transactional)</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Private (as opposed to public)</td>
<td>✓</td>
<td></td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Model Uncertainty</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Behavioral Changes</td>
<td></td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Lead indicator</td>
<td></td>
<td></td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
The Possibilities!

• Inflation
  – Measuring CPI
  – Measuring Core

• Exchange Rates
  – Real Exchange Rate
  – Nominal Exchange Rate

• Economic Activity
  – Scarcity
  – Natural Disasters
Online Information and Indexes

Our Approach to Daily Inflation Statistics

1. Use scraping technology
2. Connect to thousands of online retailers every day
3. Find individual items
4. Store and process key item information in a database
5. Develop daily inflation statistics for ~20 countries

- Date
- Item
- Price
- Description
Countries covered
Argentina (http://www.inflacionverdadera.com)
UK
EuroZone
Alternative Measures

• Purpose of Core Inflation
  – Take out the effect of shocks that are not controlled by the monetary authority
  • Food prices
  • Oil prices
  • Taxes
  • Exchange Rates
Two effects

• Direct effect: exclusion procedure
  – Food, gasoline, taxes, etc. are part of the consumption basket.
  – The direct effect is eliminated by exclusion

• Indirect effect: pass-through procedure
  – Oil has an impact on gasoline, transportation, imported items, cost of energy, etc.
  – Exchange Rate has an impact on imported items, pricing of domestic competitors, and intermediate goods.
  – The indirect effect cannot be computed by exclusion procedures
US Core Price Index

USA: Core Inflation: 2014-01-01

[Graph showing the US Core Price Index from 2014 to 2016, with two lines representing Headline and Core inflation.]
Brazil Core Price Index

Brasil: Core Inflation: 2014-01-01

[Graph showing the Brazil Core Price Index with two lines: one for Headline and one for Core. The graph indicates a steady increase in both indices from 2014 to 2016.]
Purchasing Power Parity

Thousand BigMac Indexes
The International Comparison Program

• The International Comparison Program (ICP) is the world's largest statistical initiative.
  – Established in 1968
  – It is now the largest international data collection exercise involving five regions and 107 countries.
  – The results will be combined with the OECD/Eurostat PPP program for 43 countries, bringing the total to about 150 benchmark countries.

• Measure disequilibrium two ways:
  – Compare the purchasing power across nations for similar goods or basket of goods
  – Compare the time series or international relative prices for similar goods or basket of goods
Thousands Big Mac’s Project

• Online prices represent an effective tool to measure PPP fluctuations
  – Identical items sold around the world
  – Detailed descriptions to achieve a nearly perfect matching
  – Daily Prices
• PPP indices:
  – More than 300 narrow product categories
  – With thousands individually matched items
  – In food, fuel, and electronics: we are missing clothing, personal care, household products.
  – Cars we will never match

Compare prices for a bottle of Coke across countries
Methodology

EXAMPLES OF PRODUCT CATEGORIES

<table>
<thead>
<tr>
<th>FOOD</th>
<th>ELECTRONICS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee_Illy_Ground (excluding decaf)</td>
<td>Television_Samsung_LED 32inch basic</td>
</tr>
<tr>
<td>Coffee-Regular_Ground (excluding decaf)</td>
<td>Television_Samsung_LED 32inch All Other (including Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Coffee-Regular_Beans (excluding decaf)</td>
<td>Television_Samsung_LED 40-43inch basic</td>
</tr>
<tr>
<td>Coffee_Decaf</td>
<td>Television_Samsung_LED 40-43inch All Other (including Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Coffee_All Other</td>
<td>Television_Sony_LED 32inch basic</td>
</tr>
<tr>
<td>Ketchup_Heinz-Regular</td>
<td>Television_Sony_LED 32inch All Other (including Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Ketchup_Heinz_Low Sodium, no salt</td>
<td>Television_Sony_LED 40-43inch basic</td>
</tr>
<tr>
<td>Ketchup_Heinz_All Other (e.g., flavored)</td>
<td>Television_Sony_LED 40-43inch All Other (including Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Ketchup_All Other-Regular</td>
<td>Television_Sony_LED 44-47inch All Other (including Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Ketchup_All Other_All Other (e.g., flavored)</td>
<td>Television_LG_LED 32inch basic</td>
</tr>
<tr>
<td>Soy Sauce_All Other-Regular</td>
<td>Television_LG_LED 32inch All Other (including Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Soy Sauce_All Other_Low Sodium, no salt, light</td>
<td>Television_LG_LED 40-43inch basic</td>
</tr>
</tbody>
</table>

…

• Very narrow product definition
• Must be available in multiple countries
• Branded and Unbranded categories
Methodology

1. Define a “Product”
   - Very narrow product definition
   - Must be available in multiple countries
   - Branded and Unbranded categories

2. Select, clean, size individual items
   - Dozens of items per “product” in each country
   - Different retailers, brands, and sizes

3. Product index
   - Product availability varies across countries and time
   - If a good is not available in the US, it will not appear in our series

4. Repeat for hundreds of Products

5. Sector Index (E.g. Food)
   - Some sectors are cheaper, others more expensive

6. Country Level Index (weighted)
   - Use to compare Eppp and E
Two exchange rates

• **E**: Nominal Exchange Rate
  – Determined in financial markets
    • Uncovered and covered interest rate parity
    • Order Flows

• **Eppp**: Implied Exchange Rate
  – Computed from retailer’s relative prices

\[ E_{ppp} = \sum \frac{P_{Brazil,i}}{P_{USA,i}} w_i \]
From Eppp to E?

• Why are the two related?
  – From E to Eppp
    • Theory of Pass-through
  – From Eppp to E?
    • Can we learn something about the nominal exchange rate by paying attention to relative retail prices?
      – Demand for tradeables / importables
      – Hedging demand

• Eliminating average level differences...
  – Differences in
    • Taxation
    • Degree of competition
    • Levels of development
Predictive Model

• Generate Depreciations and Appreciations signals
  – Machine Learning – Random Forest
  – Training set
    • Predict the change in the nominal exchange rate the next (60 or 120) business days
    • Use past information on the nominal exchange rate (E) and the implied exchange rate (Eppp) (typically 1000 observations)
    • Re-estimate the model everyday
  – Signal
    • Create a signal out of sample (the day after the training set)
    • Prediction is either appreciation (up) or depreciation (down)
  – Evaluate using an event study
    • Align all the “devaluation signals” and track the actual exchange rate for 120 days
    • Compute the 90 and 10 percent bands of the actual events
    • Compute the probability of depreciation or the probability of appreciation
Appreciations

UK: Event Study: Appreciation

Appreciation vs. Day
Appreciations
What is next?

• Economic Activity
• Labor Market Conditions
• Real Estate
Scarcity in Venezuela
Japan

Gauging shortages in Japanese goods
What is my (our?) objective?

• Change the way statistical offices measure economic indicators
  – Price is the first step
  – Scarcity and PPP is the second one
  – International Trade, Economic activity, Labor Markets, Consumer Confidence, Real Estate... and GDP....

  – Well, I just need... one billion prices... Well, actually...
What is my (our?) objective?

• Change the way statistical offices measure economic indicators
  – Price is the first step
  – Scarcity and PPP is the second one
  – International Trade, Economic activity, Labor Markets, Consumer Confidence, Real Estate... and GDP....

  – Well, I just need... one billion prices... Well, actually...

One trillion things!!!!
What is my (our?) objective?

• Change the way statistical offices measure economic indicators
  – Price is the first step
  – Scarcity and PPP is the second one
  – International Trade, Economic activity, Labor Markets, Consumer Confidence, Real Estate... and GDP....

  – Well, I just need... one billion prices... Well, actually...

One trillion things!!!!!!