From Data to Information

Big Data in Finance

Roberto Rigobon

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

Lacking of Careful Empirics

Lacking of Managerial Data Analysis
Official Data Sources

Data Collected
Big Data

Data Collected
## Data Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Official</td>
<td>• Representative&lt;br&gt;• High Quality</td>
<td>• Limited&lt;br&gt;• Collection is Costly</td>
</tr>
<tr>
<td>Big Data</td>
<td>• Size, details, frequency&lt;br&gt;• Might solve some identification problems&lt;br&gt;• Might improve measurement</td>
<td>• Statistics are harder&lt;br&gt;• Not representative&lt;br&gt;• Sample Selection</td>
</tr>
</tbody>
</table>
New Home Sales

Search → P&S → Mortgage → Purchase → NHS
New Home Sales

Search > P&S > Mortgage > Purchase > NHS

Open Houses and Phones
New Home Sales

Search > P&S > Mortgage > Purchase > NHS

- Open Houses and Phones
- Contract Realtor
- Search for Mortgage
Inflation

• Official Consumer Price Index
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**
Argentina (http://www.inflacionverdadera.com)
Catching up!
UK
What is next?

• Economic Activity
• Purchasing Power Parity
• Labor Market Conditions
• Real Estate
Scarcity
Gauging shortages in Japanese goods

Japanese Industrial Production
Goods availability in Japan (20-day average)
PPP indicators

- The objective of the PPP series is to help determine whether currencies are under or overvalued over time.

- 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

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</td>
</tr>
<tr>
<td></td>
<td>(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</td>
</tr>
<tr>
<td></td>
<td>(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</td>
</tr>
<tr>
<td></td>
<td>(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</td>
</tr>
<tr>
<td></td>
<td>(including Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Ketchup_All Other-Regular</td>
<td>Television_Sony_LED 44-47inch All Other</td>
</tr>
<tr>
<td></td>
<td>(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</td>
</tr>
<tr>
<td></td>
<td>Full HD, Smart, 3D)</td>
</tr>
<tr>
<td>Soy Sauce_All Other_Low Sodium, no salt,</td>
<td>Television_LG_LED 40-43inch basic</td>
</tr>
<tr>
<td>light</td>
<td></td>
</tr>
</tbody>
</table>

**Define a “Product”**

- 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
PPP-Exchange Rate (Adjusted)

Eppp Adjusted and E

Food UK/US

Eppp Adjusted by subtracting mean difference Eppp and E
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!!!!