Índice de Preços ao Consumidor Web (IPC-W)
IPC-W Core Team

- EMAp researchers: Eduardo Mendes, Renato Souza
- IBRE researchers: André Braz, Pedro Ferreira, Vagner Ardeo
- Research assistants: Antônio Sombra, Delson Barros, Yasmin Cavaliere
- Crawlers’ engineering: Luiz Geron
Price Indices

Data Collection History:

- 1800’s -- $10^2$
  - Small scale, usually centered on industrial products.
- Mid-1900’s -- $10^3$
  - Civil servants collected prices.
- Today -- $10^5$
  - Trained staff with electronic devices collect prices.

What is next?

E.g. Measure Inflation

Prices at time $t$  
Prices at time $s$

Average changes in prices between $t$ and $s$
Price Indices

- Carefully select retailers and products.
- Tend to anticipate movements.
- May be used for nowcasting (high frequency changes).
- PriceStats (scrapped data), Adobe (scanner data), Premise (scrapped and collected).
Inspiration and basis: Billion Prices Project.

Aiming at constructing online prices index.

Advance research on macroeconomics and index literatures.

Use web-scrapped prices to nowcast CPI.

Goal: collect 1,000,000,000 prices a day in two years.
Billion Prices Project @ MIT

- Collect price data from retailers from outside the country and compute an index.
- 2011 started PriceStats “over 15 million products from over 900 retailers to build daily inflation indices for 60 countries with varying degrees of coverage.”
Scrapped Prices

- Low cost
- More frequent data
- Considers all products in each retailer
- Comparable across countries
- More information on products
- No censored price spells
- Real time availability

Manually Collected Prices

- Smaller set of retailers
- Smaller set of product categories
- Must be combined with expenditure weights
- Expensive infrastructure
Challenges

- Data requires use of NLP techniques to extract information.
- Required match of items descriptions with the CPI basket.
- Protected data / Restriction on amount of prices that can be accessed.
- Price engineering and other hidden biases.
- Counterintelligence - crawlers can be detected and “killed”.
- Lots of newcomers to the index market
Opportunities

- Improve hedonic quality adjustment methods.
- Test macroeconomic hypothesis.
- Quick response to demand and supply changes.
- Anticipate future changes in CPI.
- Be able to tell the difference between online and offline *prices*.
- Be able to tell the difference between online and offline *indices*.
Methodology

1. Identify retailers. Scrap price data.
2. Clean, homogenize store and categorize data.
3. Extract information for measurement and research.
IPC-W: Food & Basic Basket Prototype

- Scrapped prices from 12 websites...
- … of big national retailers (market share = approximately 20%).
- Over than 3.4 million prices per day.
- Basic basket prototype = 5 food types:
  - Rice + Milk + Coffee + Bread + Pasta.
- Calculation method strongly inspired on IBRE’s IPC methodology.
- Confidence intervals constructed using random sub-samples.
Results:

Basic Basket Prototype
Research Questions

● Are manually collected prices enough to assess food inflation?
● Is the market share a good criterion for choosing the informant set?
● Is there any bias from geographic unbalance (within the informants)?
● Are online and offline prices (and price indices) similar? Do they measure the same economic phenomenon?
● Does the IPC-W helps to forecast the IPC?
Next Steps:

Covering all Dimensions of the Survey of Family's Budgets (POF - IBGE) in two years

<table>
<thead>
<tr>
<th></th>
<th>ALIMENTAÇÃO</th>
<th>4</th>
<th>SAÚDE E CUIDADOS PESSOAIS</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>GÉNEROS ALIMENTÍCIOS</td>
<td>41</td>
<td>SERVIÇOS DE SAÚDE</td>
</tr>
<tr>
<td>12</td>
<td>ALIMENTAÇÃO FORA</td>
<td>42</td>
<td>PRODUTOS MÉDICO-ODONTOLÓGICOS</td>
</tr>
<tr>
<td>2</td>
<td>HABITAÇÃO</td>
<td>43</td>
<td>CUIDADOS PESSOAIS</td>
</tr>
<tr>
<td>21</td>
<td>ALUGUEL E ENCARGOS</td>
<td>5</td>
<td>EDUCAÇÃO, LEITURA E RECREAÇÃO</td>
</tr>
<tr>
<td>22</td>
<td>SERVIÇOS PÚBLICOS DE RESIDÊNCIA</td>
<td>51</td>
<td>EDUCAÇÃO</td>
</tr>
<tr>
<td>23</td>
<td>MOBILIÁRIO</td>
<td>52</td>
<td>LEITURA</td>
</tr>
<tr>
<td>24</td>
<td>Roupas de cama, mesa e banho</td>
<td>53</td>
<td>RECREAÇÃO</td>
</tr>
<tr>
<td>25</td>
<td>ELETRODOMÉSTICOS E EQUIPAMENTOS</td>
<td>6</td>
<td>TRANSPORTES</td>
</tr>
<tr>
<td>26</td>
<td>UTENSÍLIOS DIVERSOS</td>
<td>61</td>
<td>TRANSPORTE PÚBLICO</td>
</tr>
<tr>
<td>27</td>
<td>ARTIGOS DE CONSERVAÇÃO E REPARO</td>
<td>62</td>
<td>TRANSPORTE PRÓPRIO</td>
</tr>
<tr>
<td>28</td>
<td>SERVIÇOS DE RESIDÊNCIA</td>
<td>7</td>
<td>DESPESAS DIVERSAS</td>
</tr>
<tr>
<td>3</td>
<td>VESTUÁRIO</td>
<td>71</td>
<td>FUMO</td>
</tr>
<tr>
<td>31</td>
<td>Roupas</td>
<td>72</td>
<td>OUTRAS DESPESAS DIVERSAS</td>
</tr>
<tr>
<td>32</td>
<td>CALÇADOS</td>
<td>8</td>
<td>COMUNICAÇÃO</td>
</tr>
<tr>
<td>33</td>
<td>ACESSÓRIOS DO VESTUÁRIO</td>
<td>81</td>
<td>SERVIÇOS DE COMUNICAÇÃO</td>
</tr>
<tr>
<td>35</td>
<td>SERVIÇOS DO VESTUÁRIO</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Next Steps

● Increase number of retailers.
● Improve collection of prices.
● Construct Web Prices index covering online products and services.
● Collect (real) crowdsourced data:
  ○ Substitution.
  ○ Basket weights.
  ○ Customized indices.
Thank You!