Using Online Price Data in Measuring the Consumer Price Index in the National Capital Region, Philippines

A Research Presented to the Faculty of Kobe University - GSICS

POLO GLEN
Philippine Statistics Authority
Introduction

• Consumer Price Index (CPI), one of the most important economic indicators, measures the change in the average price of products and services purchased by households, assuming that the quality of these commodities remain constant over time.

• Philippine Statistics Authority (PSA) is the agency responsible for compiling the CPI in the Philippines.

• Compilation of CPI involves processing around 360,000 price quotations per month.

• Data collection is done through personal visit to sample stores and recording the price data using pencil and paper.

• The use of Android-based data collection was introduced in 2015.
Background

• Advancement in technology made web scraping possible.
• Web scraping is the process by which data across the Internet are automatically extracted and saved as a database. It is an automated version of visiting the store, looking for the sample commodity, copying its price and compiling the data.”
• In the Philippines, online data is not yet utilized for the generation of official statistics
Problem Statement

• With the availability of the technology, up to what extent can online data replace the survey-based data of the PSA in compiling the current CPI series?
Related Studies

• Billion Prices Project by Alberto Cavallo of MIT; pioneering study on the use of web scraping for official statistics; involved collecting price data from 25 countries; used in compiling the CPI of Argentina

• Similar studies conducted by statistical offices of Italy, The Netherlands, United Kingdom, Norway, Austria and Singapore.
Benefits and Advantages

• Cheaper and more efficient
• Timeliness and provision of coverage expansion
• Allows for more frequent collection
• Provision for other studies due to the availability of large amount of data

Limitations and Disadvantages

• Legality and ethical issues
• Limited number of websites and commodities available
• Technical hurdles (e.g. the need for user ID)
• Dynamic pricing or price discrimination
• Competition with the official statistics
Data, Sources, Collection and Coverage

• Survey-based price data from the PSA
• Official CPI results from the PSA
• Online data gathered from the “eligible or scrapable” websites of CPI sample stores
• Covered 10 out of 222 sub-classes (5-digit PCOICOP) comprising 35 out of 720 commodities in the market basket

• Collection was done by a scraper program using Python language
• Web scraping was done twice a month from April 2019 to May 2020 according to PSA schedule:
  1st collection period: 1st to 5th day of the month
  2nd collection period: 15th to 17th day of the month
Eligible or Scrapable Websites

- Price data posted can be automatically collected using a web scraper program.
- The website does not have any technical hurdles, such as the need to enter user ID and password before showing the contents of the website or delivery address in the case of grocery stores before showing the price off the product, that prevent automated data collection.
# Scraper Program

## Single Products

<table>
<thead>
<tr>
<th>Include</th>
<th>Name</th>
<th>Total URLs</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>rose_pharmacy</td>
<td>109</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>guanzon_honda-wave_07.1.21002</td>
<td>45</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>hondaph_hondas_wave_07.1.21002</td>
<td>25</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>kservice_honda-wave_07.1.21002_001</td>
<td>56</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>anson_refrigerator_2-door_05.3.11001</td>
<td>79</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>automatic-center_refrigerator_05.3.11001</td>
<td>35</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>emilio-s-lim_single_door_refrigerator_05.3.11060</td>
<td>14</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>emilio-s-lim_refrigerator-2-door_05.3.11001</td>
<td>30</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>emilio-s-lim_washing_machine_05.3.12051</td>
<td>33</td>
<td>Delete</td>
</tr>
<tr>
<td></td>
<td>abenson_refrigerator-single-door_05.3.11060</td>
<td>24</td>
<td>Delete</td>
</tr>
</tbody>
</table>

Showing 1 to 10 of 138 entries

Previous: 1, 2, 3, 4, 5, ..., 14, Next
## Scraper Program Output

<table>
<thead>
<tr>
<th>URL</th>
<th>Description</th>
<th>Sub Details</th>
<th>Price</th>
<th>Sale Price</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.162mnt-1.5hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.162mnt-1.5hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC 162MNT 1.5HP Window Type Air Conditioner</td>
<td></td>
<td>17,999.00</td>
<td>16,999.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.162rt-1.5hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.162rt-1.5hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC 162RT 1.5HP Window Type Air Conditioner</td>
<td></td>
<td>19,999.00</td>
<td>18,999.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.192mnt-2.0hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.192mnt-2.0hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC 192MNT 2.0HP Window Type Air Conditioner</td>
<td></td>
<td>22,999.00</td>
<td>21,999.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.192rt-2.0hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.192rt-2.0hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC 192RT 2.0HP Window Type Air Conditioner</td>
<td></td>
<td>24,599.00</td>
<td>23,599.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.92mnt-1.0hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.92mnt-1.0hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC 92MNT 1.0HP Window Type Air Conditioner</td>
<td></td>
<td>13,999.00</td>
<td>12,499.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.92rt-1.0hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.92rt-1.0hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC 92RT 1.0HP Window Type Air Conditioner</td>
<td></td>
<td>15,999.00</td>
<td>14,999.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.wti1000ox-1-0hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.wti1000ox-1-0hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC-WTI1000OX 1.0HP Window Type Air Conditioner</td>
<td></td>
<td>27,699.00</td>
<td>24,999.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/american-home-ahac.wti1500ox-1-5hp-window-type-air-conditioner/">https://ansons.ph/product/american-home-ahac.wti1500ox-1-5hp-window-type-air-conditioner/</a></td>
<td>American Home AHAC-WTI1500OX 1.5HP Window Type Air Conditioner</td>
<td></td>
<td>32,999.00</td>
<td>29,499.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/carrier-wcrah010ree-1-0-hp-window-type-air-conditioner/">https://ansons.ph/product/carrier-wcrah010ree-1-0-hp-window-type-air-conditioner/</a></td>
<td>Carrier WCRAH010RE 1.0 HP Window Type Air Conditioner</td>
<td></td>
<td>24,599.00</td>
<td>23,860.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/carrier-wcrah012ree-1.5-hp-window-type-air-conditioner/">https://ansons.ph/product/carrier-wcrah012ree-1.5-hp-window-type-air-conditioner/</a></td>
<td>Carrier WCRAH012REE 1.5 HP Window Type Air Conditioner</td>
<td></td>
<td>42,799.00</td>
<td>41,516.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/carrier-wcrah014ree-1.5-hp-window-type-air-conditioner/">https://ansons.ph/product/carrier-wcrah014ree-1.5-hp-window-type-air-conditioner/</a></td>
<td>Carrier WCRAH014REE 1.5 HP Window Type Air Conditioner</td>
<td></td>
<td>29,199.00</td>
<td>28,300.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/carrier-wcrah024ree-2-5-hp-window-type-air-conditioner/">https://ansons.ph/product/carrier-wcrah024ree-2-5-hp-window-type-air-conditioner/</a></td>
<td>Carrier WCRAH024REE 2.5 HP Window Type Air Conditioner</td>
<td></td>
<td>38,499.00</td>
<td>36,200.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/carrier-wcrah006ehc-0-5-hp-window-type-air-conditioner/">https://ansons.ph/product/carrier-wcrah006ehc-0-5-hp-window-type-air-conditioner/</a></td>
<td>Carrier WCARH006EHC 0.5 HP Window Type Air Conditioner</td>
<td></td>
<td>11,999.00</td>
<td>11,640.00</td>
</tr>
<tr>
<td><a href="https://ansons.ph/product/carrier-wcarh008lec-0-75-hp-window-type-air-conditioner/">https://ansons.ph/product/carrier-wcarh008lec-0-75-hp-window-type-air-conditioner/</a></td>
<td>Carrier WCARH008ECL 0.75 HP Window Type Air Conditioner</td>
<td></td>
<td>17,399.00</td>
<td>16,877.00</td>
</tr>
</tbody>
</table>
Index Computation and Comparison

- Followed the official 2012-based CPI index computation of the PSA using the official CPI weights

- Three types of indices were computed and compared with the official CPI
  - CPI I: online price data only
  - CPI II: combination of online and survey-based price data
  - CPI III: expanded coverage for online price data
Limitation of the Study

- Excludes aggregator sites and websites of stores not included in the CPI sample stores
- Sample websites may not be representative of the actual number
- Limited commodity coverage
Results and Discussion

Availability of Websites for Scraping

• 48% of sample stores have websites

• 280 unique websites and only 112 scrapable

• Number of price quotations vary per month depending on the availability of the products on the websites

Computed CPI

• Trends of computed CPIs generally resemble the Official CPI trends at the sub-class level

• Among the three computed CPIs, CPI I had the lowest mean and range of absolute deviation in 5 of the 10 sub-classes despite having the lowest number of price quotations
Results and Discussion

Other Observations

- No price change in two sub-classes for the whole duration of research, while prices changed only once or twice for the other five sub-classes.
- Trend lines at the class level (4-digit PCOICOP) index are closer to that of the official index.

Challenges Encountered

- Technical difficulties were experienced during scraping.
- Limited number of “scrapeable” websites due to technical hurdles introduced by the target websites (e.g. Websites need information on user ID and password, accepting the TOC, needs address for delivery).
Results and Discussion

Figure 1
Comparison of CPI Trends for Sub-Class 05.3.11
(Refrigerators, freezers and fridge-freezers)

Figure 2
Comparison of CPI Trends for Sub-Class 05.3.12
(Washing-machines, dryers, drying cabinets, dishwashers, ironing and pressing machines)
Results and Discussion

Figure 3
Comparison of CPI Trends for Sub-Class 05.3.13
(Cookers, spit roaster, hobs, ranges, ovens and micro-wave ovens)

Figure 4
Comparison of CPI Trends for Sub-Class 05.3.14
(Air conditioners, humidifiers, space heaters, water heaters, ventilators and extractor hoods)
Results and Discussion

**Figure 5**
Comparison of CPI Trends for Sub-Class 05.3.21
(Small electric household appliances, such as electric fans, rice cooker, flat iron, toasters, etc.)

**Figure 6**
Comparison of CPI Trends for Sub-Class 06.1.11
(Analgesic/Antipyretics and muscle relaxants)
Results and Discussion

Figure 7
Comparison of CPI Trends for Sub-Class 06.1.15
(Antiallergics, cough and colds medicines)

Figure 8
Comparison of CPI Trends for Sub-Class 06.1.16
(Anti-infectives)
Results and Discussion

Figure 9
Comparison of CPI Trends for Sub-Class 07.1.21 (Motorcycle)

Figure 10
Comparison of CPI Trends for Sub-Class 07.2.43 (Driving lessons, driving tests and driving licenses)
Conclusion

• Online price data can be used to compile the CPI, especially if the number of sample websites are increased, or be incorporated in the current CPI series for applicable subclasses

• Web scraping is more efficient and allows for expanded commodity coverage

• Programming knowledge is needed in creating the scraper program, at least for Python language

• Needs supplemental study with expanded commodity coverage and a longer series

• Inclusion of inquiry on online stores in the Commodity and Outlet Survey
References:


References:


References:


References:


