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:**

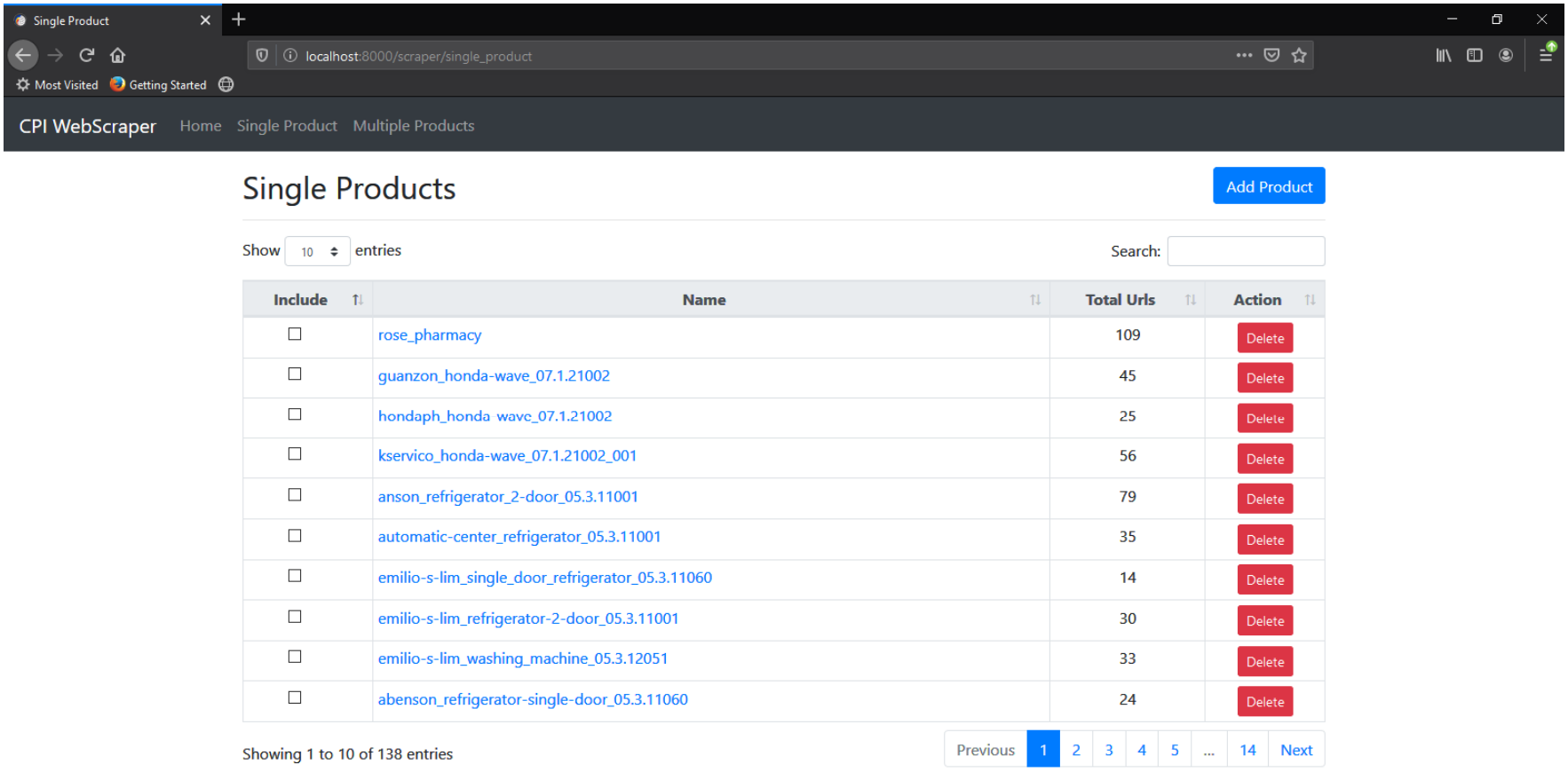
  ***1st collection period*: 1st to 5th day of the month**

  ***2nd collection period*: 15th to 17th day of the month**

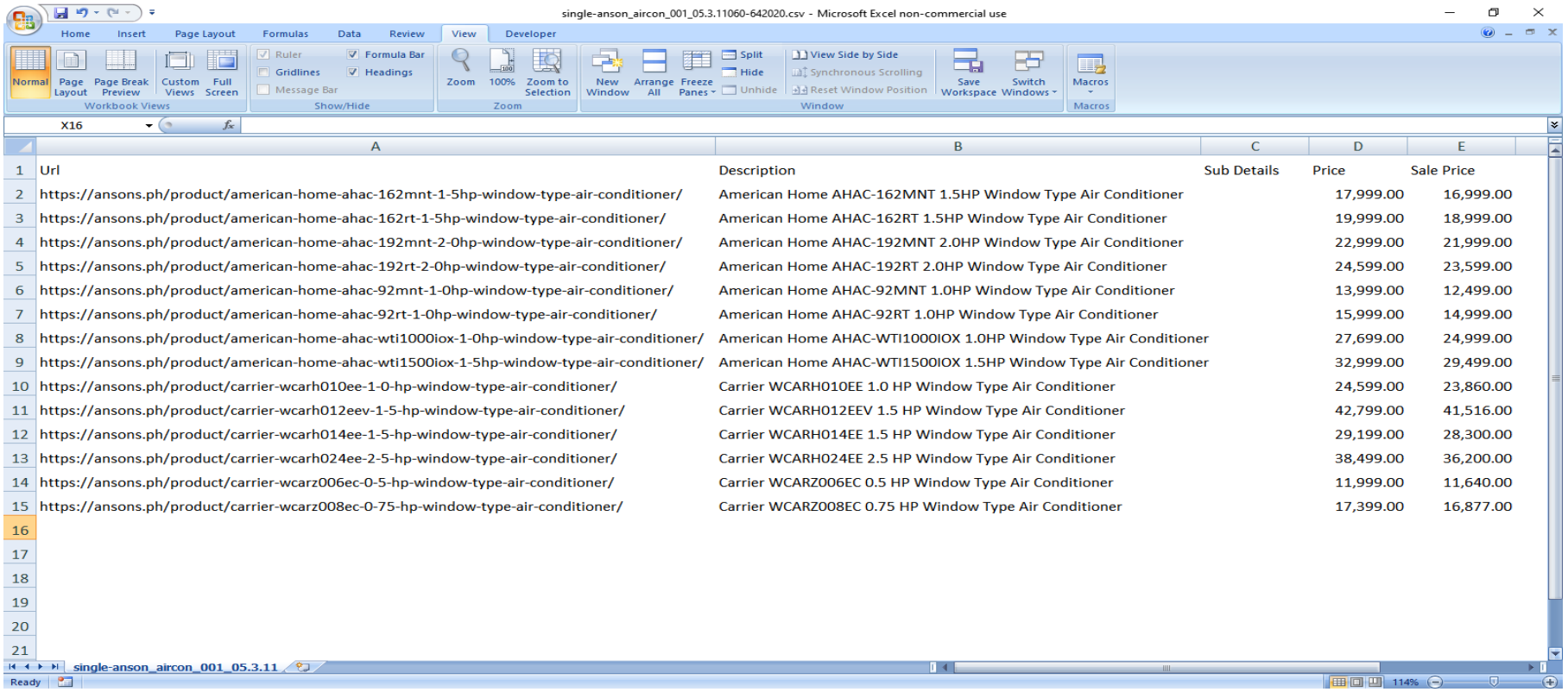
Eligible or *Scrapable* Websites

- 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



Scraper Program Output



single-anson_aircon_001_05.3.11060-642020.csv - Microsoft Excel non-commercial use

	A	B	C	D	E
1	Url	Description	Sub Details	Price	Sale Price
2	https://ansons.ph/product/american-home-ahac-162mnt-1-5hp-window-type-air-conditioner/	American Home AHAC-162MNT 1.5HP Window Type Air Conditioner		17,999.00	16,999.00
3	https://ansons.ph/product/american-home-ahac-162rt-1-5hp-window-type-air-conditioner/	American Home AHAC-162RT 1.5HP Window Type Air Conditioner		19,999.00	18,999.00
4	https://ansons.ph/product/american-home-ahac-192mnt-2-0hp-window-type-air-conditioner/	American Home AHAC-192MNT 2.0HP Window Type Air Conditioner		22,999.00	21,999.00
5	https://ansons.ph/product/american-home-ahac-192rt-2-0hp-window-type-air-conditioner/	American Home AHAC-192RT 2.0HP Window Type Air Conditioner		24,599.00	23,599.00
6	https://ansons.ph/product/american-home-ahac-92mnt-1-0hp-window-type-air-conditioner/	American Home AHAC-92MNT 1.0HP Window Type Air Conditioner		13,999.00	12,499.00
7	https://ansons.ph/product/american-home-ahac-92rt-1-0hp-window-type-air-conditioner/	American Home AHAC-92RT 1.0HP Window Type Air Conditioner		15,999.00	14,999.00
8	https://ansons.ph/product/american-home-ahac-wti1000iox-1-0hp-window-type-air-conditioner/	American Home AHAC-WTI1000IOX 1.0HP Window Type Air Conditioner		27,699.00	24,999.00
9	https://ansons.ph/product/american-home-ahac-wti1500iox-1-5hp-window-type-air-conditioner/	American Home AHAC-WTI1500IOX 1.5HP Window Type Air Conditioner		32,999.00	29,499.00
10	https://ansons.ph/product/carrier-wcarh010ee-1-0-hp-window-type-air-conditioner/	Carrier WCARH010EE 1.0 HP Window Type Air Conditioner		24,599.00	23,860.00
11	https://ansons.ph/product/carrier-wcarh012eev-1-5-hp-window-type-air-conditioner/	Carrier WCARH012EEV 1.5 HP Window Type Air Conditioner		42,799.00	41,516.00
12	https://ansons.ph/product/carrier-wcarh014ee-1-5-hp-window-type-air-conditioner/	Carrier WCARH014EE 1.5 HP Window Type Air Conditioner		29,199.00	28,300.00
13	https://ansons.ph/product/carrier-wcarh024ee-2-5-hp-window-type-air-conditioner/	Carrier WCARH024EE 2.5 HP Window Type Air Conditioner		38,499.00	36,200.00
14	https://ansons.ph/product/carrier-wcarz006ec-0-5-hp-window-type-air-conditioner/	Carrier WCARZ006EC 0.5 HP Window Type Air Conditioner		11,999.00	11,640.00
15	https://ansons.ph/product/carrier-wcarz008ec-0-75-hp-window-type-air-conditioner/	Carrier WCARZ008EC 0.75 HP Window Type Air Conditioner		17,399.00	16,877.00

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

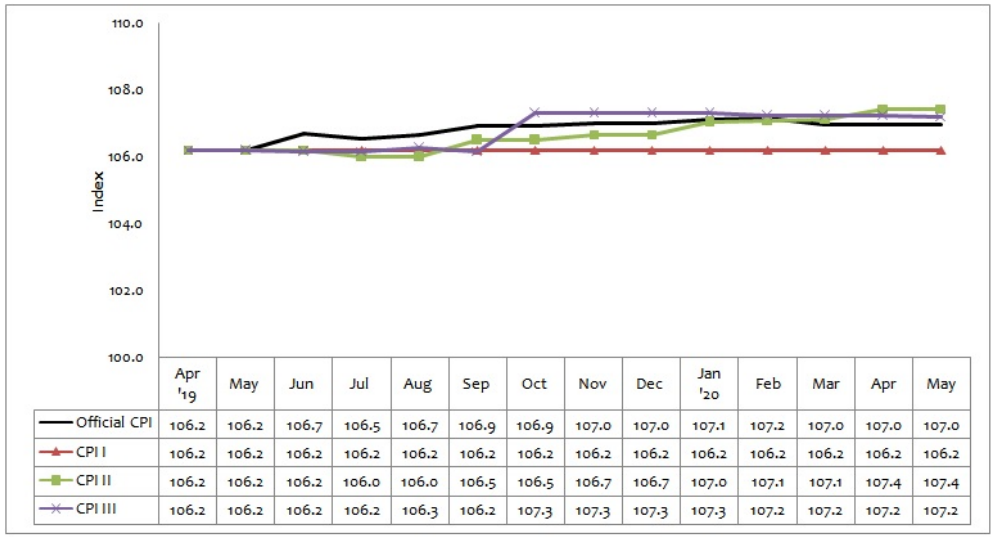
Other Observations

- **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 "*scrapable*" 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)**

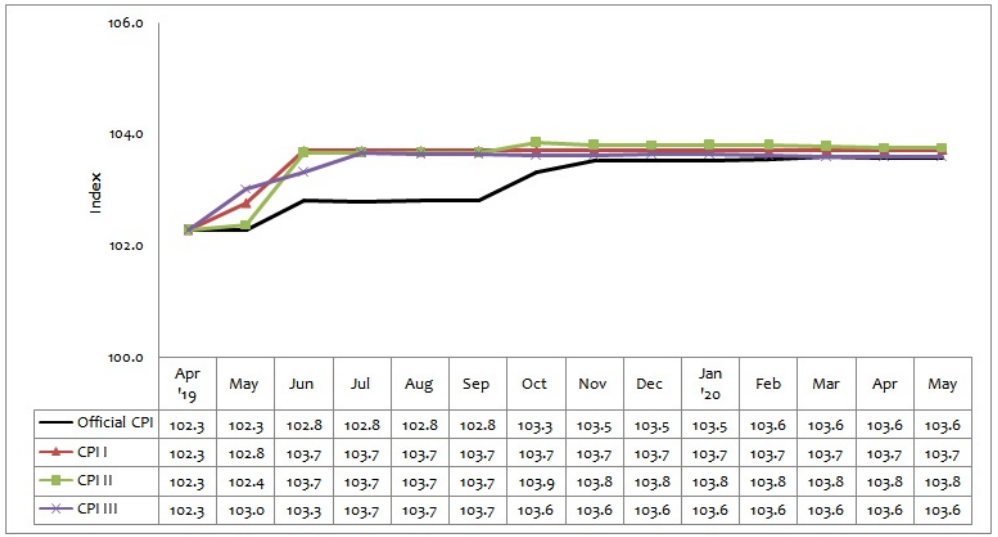
Results and Discussion



	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI	106.2	106.2	106.7	106.5	106.7	106.9	106.9	107.0	107.0	107.1	107.2	107.0	107.0	107.0
CPI I	106.2	106.2	106.2	106.2	106.2	106.2	106.2	106.2	106.2	106.2	106.2	106.2	106.2	106.2
CPI II	106.2	106.2	106.2	106.0	106.0	106.5	106.5	106.7	106.7	107.0	107.1	107.1	107.4	107.4
CPI III	106.2	106.2	106.2	106.2	106.3	106.2	107.3	107.3	107.3	107.3	107.2	107.2	107.2	107.2

Figure 1

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

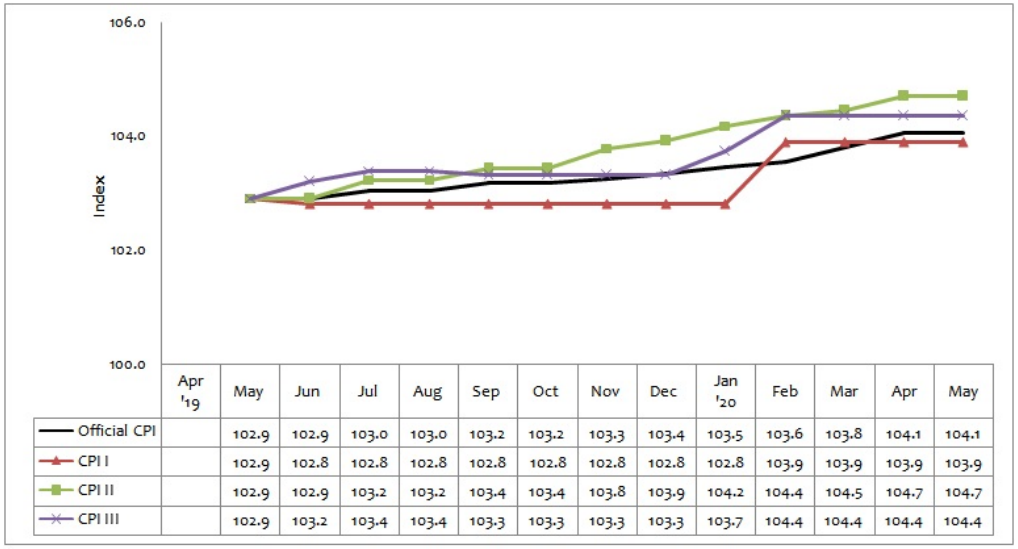


	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI	102.3	102.3	102.8	102.8	102.8	102.8	103.3	103.5	103.5	103.5	103.6	103.6	103.6	103.6
CPI I	102.3	102.8	103.7	103.7	103.7	103.7	103.7	103.7	103.7	103.7	103.7	103.7	103.7	103.7
CPI II	102.3	102.4	103.7	103.7	103.7	103.7	103.9	103.8	103.8	103.8	103.8	103.8	103.8	103.8
CPI III	102.3	103.0	103.3	103.7	103.7	103.7	103.6	103.6	103.6	103.6	103.6	103.6	103.6	103.6

Figure 2

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

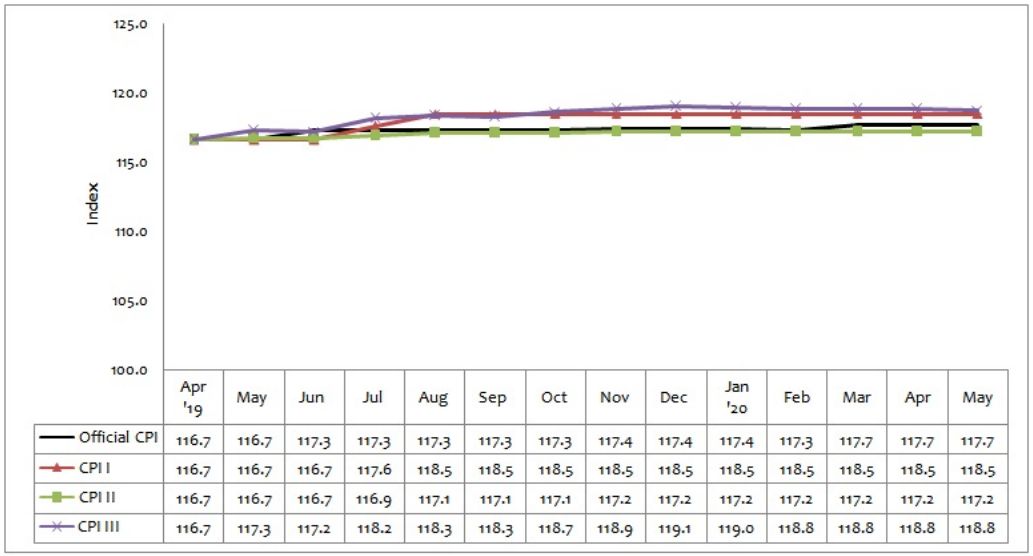
Results and Discussion



	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI		102.9	102.9	103.0	103.0	103.2	103.2	103.3	103.4	103.5	103.6	103.8	104.1	104.1
CPI I		102.9	102.8	102.8	102.8	102.8	102.8	102.8	102.8	102.8	103.9	103.9	103.9	103.9
CPI II		102.9	102.9	103.2	103.2	103.4	103.4	103.8	103.9	104.2	104.4	104.5	104.7	104.7
CPI III		102.9	103.2	103.4	103.4	103.3	103.3	103.3	103.3	103.7	104.4	104.4	104.4	104.4

Figure 3

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

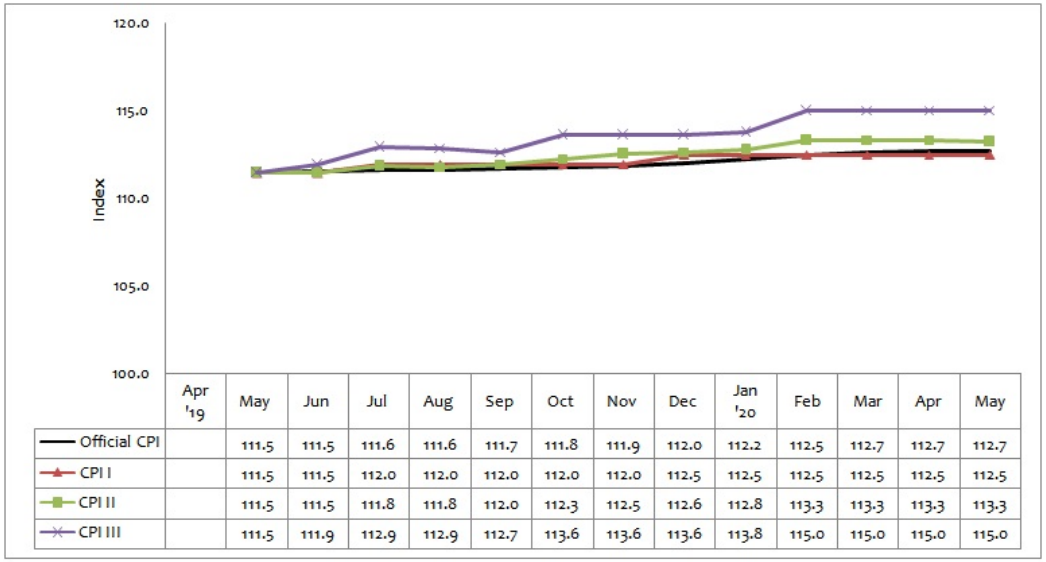


	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI	116.7	116.7	117.3	117.3	117.3	117.3	117.3	117.4	117.4	117.4	117.3	117.7	117.7	117.7
CPI I	116.7	116.7	116.7	117.6	118.5	118.5	118.5	118.5	118.5	118.5	118.5	118.5	118.5	118.5
CPI II	116.7	116.7	116.7	116.9	117.1	117.1	117.1	117.2	117.2	117.2	117.2	117.2	117.2	117.2
CPI III	116.7	117.3	117.2	118.2	118.3	118.3	118.7	118.9	119.1	119.0	118.8	118.8	118.8	118.8

Figure 4

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

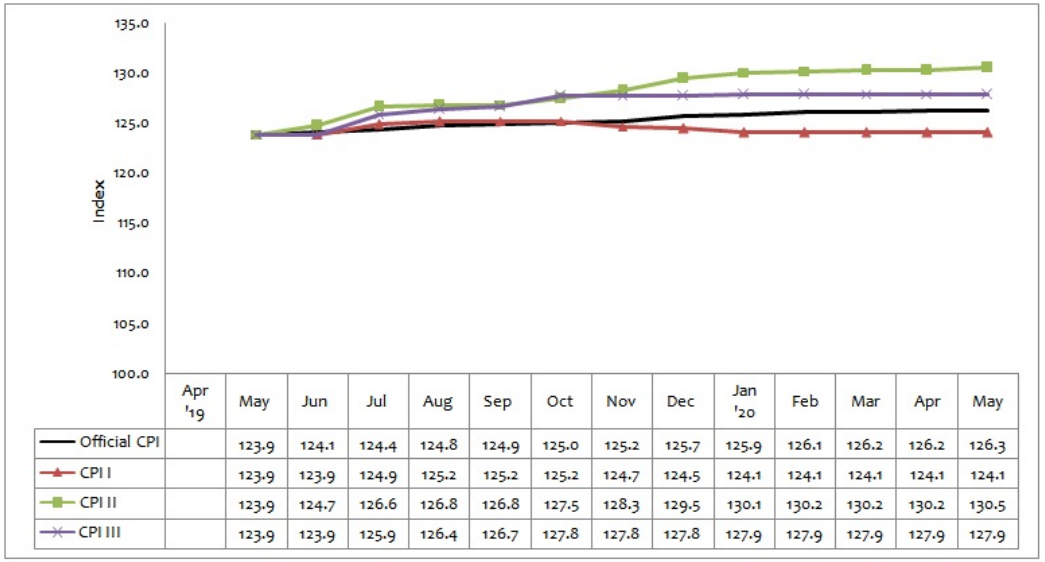
Results and Discussion



	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI		111.5	111.5	111.6	111.6	111.7	111.8	111.9	112.0	112.2	112.5	112.7	112.7	112.7
CPI I		111.5	111.5	112.0	112.0	112.0	112.0	112.0	112.5	112.5	112.5	112.5	112.5	112.5
CPI II		111.5	111.5	111.8	111.8	112.0	112.3	112.5	112.6	112.8	113.3	113.3	113.3	113.3
CPI III		111.5	111.9	112.9	112.9	112.7	113.6	113.6	113.6	113.8	115.0	115.0	115.0	115.0

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.)

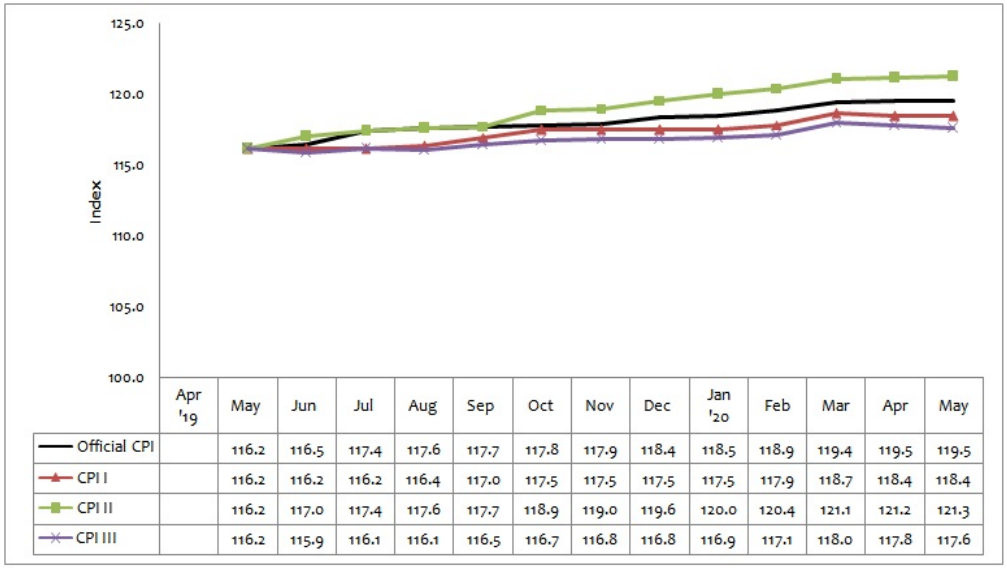


	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI		123.9	124.1	124.4	124.8	124.9	125.0	125.2	125.7	125.9	126.1	126.2	126.2	126.3
CPI I		123.9	123.9	124.9	125.2	125.2	125.2	124.7	124.5	124.1	124.1	124.1	124.1	124.1
CPI II		123.9	124.7	126.6	126.8	126.8	127.5	128.3	129.5	130.1	130.2	130.2	130.2	130.5
CPI III		123.9	123.9	125.9	126.4	126.7	127.8	127.8	127.8	127.9	127.9	127.9	127.9	127.9

Figure 6

Comparison of CPI Trends for Sub-Class 06.1.11
(Analgesic/Antipyretics and muscle relaxants)

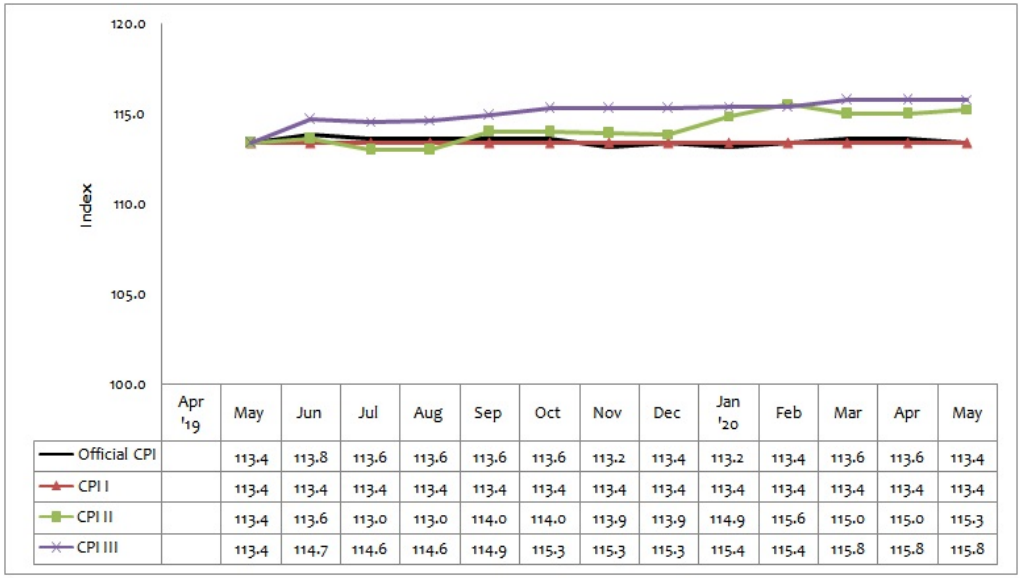
Results and Discussion



	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI		116.2	116.5	117.4	117.6	117.7	117.8	117.9	118.4	118.5	118.9	119.4	119.5	119.5
CPI I		116.2	116.2	116.2	116.4	117.0	117.5	117.5	117.5	117.5	117.9	118.7	118.4	118.4
CPI II		116.2	117.0	117.4	117.6	117.7	118.9	119.0	119.6	120.0	120.4	121.1	121.2	121.3
CPI III		116.2	115.9	116.1	116.1	116.5	116.7	116.8	116.8	116.9	117.1	118.0	117.8	117.6

Figure 7

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

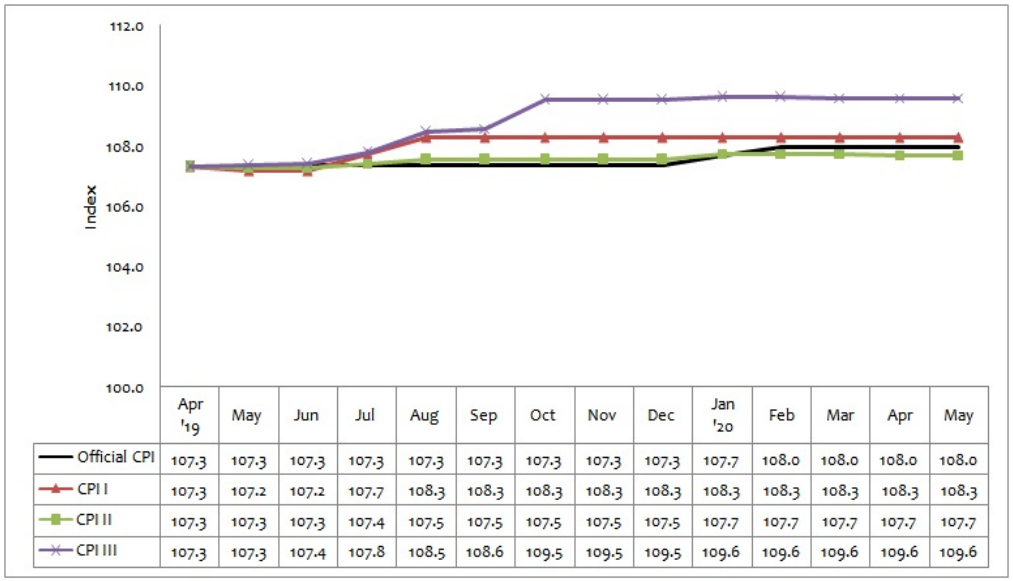


	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI		113.4	113.8	113.6	113.6	113.6	113.6	113.2	113.4	113.2	113.4	113.6	113.6	113.4
CPI I		113.4	113.4	113.4	113.4	113.4	113.4	113.4	113.4	113.4	113.4	113.4	113.4	113.4
CPI II		113.4	113.6	113.0	113.0	114.0	114.0	113.9	113.9	114.9	115.6	115.0	115.0	115.3
CPI III		113.4	114.7	114.6	114.6	114.9	115.3	115.3	115.3	115.4	115.4	115.8	115.8	115.8

Figure 8

Comparison of CPI Trends for Sub-Class 06.1.16 (Anti-infectives)

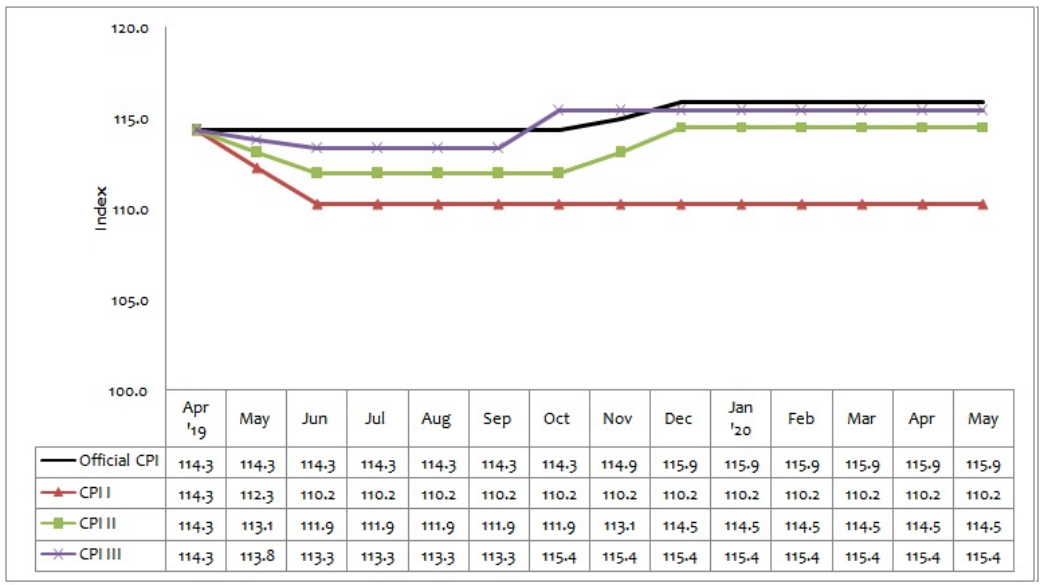
Results and Discussion



	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI	107.3	107.3	107.3	107.3	107.3	107.3	107.3	107.3	107.3	107.7	108.0	108.0	108.0	108.0
CPI I	107.3	107.2	107.2	107.7	108.3	108.3	108.3	108.3	108.3	108.3	108.3	108.3	108.3	108.3
CPI II	107.3	107.3	107.3	107.4	107.5	107.5	107.5	107.5	107.5	107.7	107.7	107.7	107.7	107.7
CPI III	107.3	107.3	107.4	107.8	108.5	108.6	109.5	109.5	109.5	109.6	109.6	109.6	109.6	109.6

Figure 9

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



	Apr '19	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan '20	Feb	Mar	Apr	May
Official CPI	114.3	114.3	114.3	114.3	114.3	114.3	114.3	114.9	115.9	115.9	115.9	115.9	115.9	115.9
CPI I	114.3	112.3	110.2	110.2	110.2	110.2	110.2	110.2	110.2	110.2	110.2	110.2	110.2	110.2
CPI II	114.3	113.1	111.9	111.9	111.9	111.9	111.9	113.1	114.5	114.5	114.5	114.5	114.5	114.5
CPI III	114.3	113.8	113.3	113.3	113.3	113.3	115.4	115.4	115.4	115.4	115.4	115.4	115.4	115.4

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 sub-classes
- 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:

- Albert, J.A. (2014). Big data: Big threat or big opportunity for official statistics. *Paris21*. https://paris21.org/news-center/news/big-data-big-opportunity-or-big-threat-official-statistics
- Armknecht, P., & Weyback, D. (1989). Adjustment for quality change in the U.S. consumer price index. *Journal of Official Statistics, 5*(2), 107-123.
- Bhardwaj, H., Flower, T., Lee, P., & Mayhew, M. (2017). Research indices using web scraped data: August 2017 update. *Office of the National Statistics, United Kingdom.* https://www.ons.gov.uk/economy/inflationandpriceindices/articles/researchindicesusingwebscrapedpricedata/august2017update
- Boettcher, I. (2015). Automatic data collection on the Internet (web scraping). *Statistics Austria.* https://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s2p6_pap.pdf
- Breton, R., Clews G., Metcalfe, L., Miliken, N., Payne, C., Winton, J., & Woods, A. (2015). Research indices using web scraped. *Office of the National Statistics, United Kingdom.* http://inflationmatters.com/wp-content/uploads/2015/10/ONS-web-scraped-data-article-01092015.pdf
- Brynjolfsson, E., & Smith, M. (2000). Frictionless commerce? A comparison of Internet and conventional retailers. *Management Science, 46*(4), 563-585.

References:

- Cavallo, A. (2012). Online and official price indexes: Measuring Argentina's inflation. *Journal of Monetary Economics (2012), 60*(2), 152–165.
- Cavallo, A. (2017). Are online and offline prices similar? Evidence from large multi-channel retailers. *American Economic Review 2017, 107*(1), 283-303.
- Cavallo, A., & Rigobon, R. (2016). The billion prices project: Using online prices for measurement and research. *Journal of Economic Perspectives*, *30*(2), 151–178.
- Chuanyang, F., & Lee Wen Hao, J. (2016). Experiences with the use of online prices in consumer price index. *Statistics Singapore Newsletter, September 2016*. https://www.singstat.gov.sg/-/media/files/publications/reference/newsletter/ssnsep2016.pdf
- Edelman, B. (2012). Using Internet data for economic research. *Journal of Economic Perspective*, *26*(2), 189-206.
- Ehrestein, V., Nielsen, H., Pedersen, A., Johnsen, S., & Pedersen L. (2017). Clinical epidemiology in the era of big data: New opportunities, familiar challenges. *Clinical Epidemiology,* 2017(9), 245-250.

References:

- Griffioen, R., de Haan, J., & Willenborg, L. (2014). Collecting clothing data from the Internet. *Department of Methodology, Statistics Netherlands*. https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2014/UNECE-ILO_2014_Griffioen_deHaan_Willenborg.pdf
- Griffioen, R., & ten Bosch, O. (2018). On the use of Internet data for the Dutch CPI. *Statistics Netherlands*. https://www.researchgate.net/publication/324896024_On_the_use_of_Internet_data_for_the_Dutch_CPI
- Harchaoui, T., & Janssen, R. (2018). How can big data enhance the timeliness of official statistics? The case of the U.S. consumer price index. *International journal of Forecasting, 34*, 225-234.
- Hoekstra, R., ten Bosch, O., & Harteveld, F. (2012). Automated data collection from web sources for official statistics: First experiences. *Statistical Journal of the IAOS, 28*, 99-111.
- Krotov, V., & Silva, L. (2018). Legality and ethics of web scraping. *Conference Paper, Twenty-fourth Americas Conference on Information Systems, New Orleans*. https://www.researchgate.net/publication/324907302_Legality_and_Ethics_of_Web_Scraping/citation/download

References:

- Nygaard, R. (2015). The use of online prices in the Norwegian CPI. *14th meeting of the Ottawa Group*, Tokyo, Japan, 20–22 May 2015. https://www.stat.go.jp/english/info/meetings/og2015/pdf/t1s2p5_pap.pdf
- Pedraza, P., Visintin, S., Tijdens, K., & Kismihók, G. (2017). Survey vs scraped data: Comparing time series properties of web and survey vacancy data. *University of Amsterdam*, ISSN online: 2213-4980.
- Philippine Statistics Authority (2006). Consumer Price Index User's Manual. Unpublished.
- Philippine Statistics Authority (2013). Implementing rules and regulations of Republic Act No. 10625, otherwise known as the Philippine Statistical Act of 2013. https://psa.gov.ph/sites/default/files/attachments/gad/article/RA%2010625_IRR.pdf
- Philippine Statistics Authority (2018). PSA updates the base year for consumer price index for all income households to 2012 from 2006. https://psa.gov.ph/sites/default/files/attachments/gad/article/RA%2010625_IRR.pdf
- Philippine Statistics Authority (2019). 2015 Survey on Information and Communication Technology - non-core ICT industries - Final Results. http://www.psa.gov.ph/content/2015-survey-information-and-communication-technology-non-core-ict-industries-final-results

References:

- Philippine Statistics Authority (n.d.). 2009 Philippine Classification of Individual Consumption According to Purpose (PCOICOP). Retrieved 18 May 2020. https://psa.gov.ph/classification/pcoicop/technical-notes
- Polidoro, F., Giannini, R., Lo Conte, R., Mosca, S., & Rossetti, F. (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. *Statistical Journal of the IAOS*, *31*(2), 165–176.
- Struijs, P. (2016). Big data for official statistics. *Department for Methodology and Process Development, Statistics Netherlands*. http://www.eustat.eus/productosServicios/datos/58_Big_Data_for_Official_Statistics_Peter_Struijs.pdf
- Zhao, B. (2017). Web Scraping. *Researchgate*. https://www.researchgate.net/publication/317177787_Web_Scraping