



## Implementation of Marketplace Data in The Calculation of The Consumer Price Index (Case Study of CPI cities in Java, Indonesia)

Muhammad Khozy Al Haqqoni<sup>1</sup>, Setia Pramana<sup>1</sup>

<sup>1</sup>BPS-Statistics Indonesia

### Abstract

Digital Economy in recent years, especially in Southeast Asia, including Indonesia, is growing rapidly. E-commerce is one part of the Digital Economy. BPS-Statistics Indonesia as a Non-ministerial Government Agency responsible directly to the president has conducted an E-commerce Survey in 2019. From this publication it is concluded that the interest of Indonesian traders using the internet in selling in recent years has increased. So, the urgency of using e-commerce data in its application in official statistics is increasingly needed. Several studies have carried out the application of e-commerce data in the calculation of CPI. In this research, e-commerce data is applied with a case study using data from one of the online marketplaces in Indonesia in calculating CPI at the city level CPI in Java. The purpose of this study is to compare the marketplace CPI data and BPS CPI. Data collection is done through web scraping techniques and followed by preprocessing data and analyzed descriptively. Web scraper that is built can be used in obtaining data. Commodity-level CPI with marketplace data tends to have relatively large prices which results in higher CPI being compared to BPS CPI.

### Keywords

Consumer price index (CPI); Marketplace; Web scraping

## 1. Introduction

The digital economy is defined as a part of economic output that solely originates or primarily from digital technology with a business model based on digital goods or services (Bukht & Heeks, 2017). In Southeast Asia, including Indonesia, the digital economy has developed rapidly in recent years. Anandan et al. (2016) predicted an internet economy of 200 billion USD in Southeast Asia in 2025. This means that the role of the digital economy as business transactions on the internet is now starting to be evenly distributed and will continue to grow.

The U.S. Bureau of Economic Analysis classifies the digital economy into three main types of goods and services, namely infrastructure where digitization is needed so that it can be connected, e-commerce transactions, and digital media. According to the Indonesian E-commerce Association (IdEA) in Wijaya & Mariyah's research (2019), e-commerce is divided into three categories, namely (1) Classified ads, e-commerce where there is no online transaction facility from the website. (2) Marketplace, e-commerce where the website facilitates online transactions for online merchants. (3) Online retail, e-commerce where the website domain belongs to the online store itself and there is a stock of goods / services to be sold.

BPS-Statistics Indonesia has conducted the 2019 E-commerce Survey. The survey was conducted on a sample of 3,504 Census Blocks spread across 101 cities / regencies in Indonesia. The survey results show that only 15.08 percent of respondents do e-commerce businesses. However, of the 15.08 percent of e-commerce businesses, 45.03 percent of businesses started using the internet from 2017 to 2018 and in 2019 as many as 25.11 percent, the rest started from 2016 onwards. From this, it can be seen that in recent years, the interest of Indonesian traders by using the internet in selling has increased. So, the urgency of using e-commerce data in its application to official statistics is increasingly needed.

Several studies have taken a big data approach on official statistics. Rizkika, et al. (2019) shows that in general, there is a similar movement of official price data patterns between commodities that collected by BPS-Statistics Indonesia and data collected through big data approach. Wijaya & Mariyah's research (2019) applies data from several online retailers in Indonesia and calculates the consumer price index with this

data, which then compares the results of the e-commerce-based CPI and BPS. Nonetheless, the research locus is assumed to be in the DKI Jakarta area due to limited online retail data which does not have information on sales locations. The results of this study indicate that the e-commerce-based CPI calculation can be done using the BPS calculation approach. Also, in early 2020, the BPS CPI calculation uses a method based on the results of the 2018 Cost of Living Survey.

Therefore, this study will implement e-commerce data from one of the online marketplaces in Indonesia in calculating the CPI. Researchers build a web scraper to obtain data from one of the online marketplaces in Indonesia. From the data obtained, a comparison is made between the CPI marketplace data approach and the BPS CPI.

## 2. Methodology

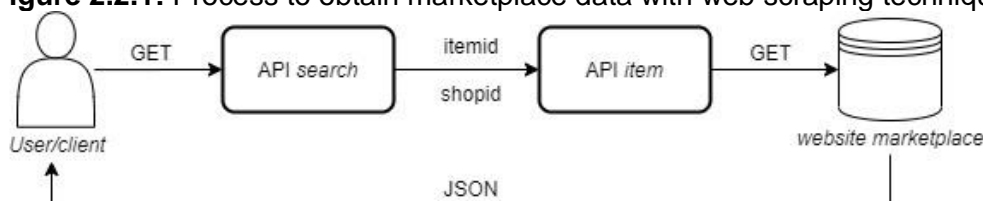
### 2.1 Data Sources

- a) Marketplace data obtained from one of the online marketplaces in Indonesia using web scraping technique during April to June 2020.
- b) Official Consumer Price Data Index obtained from BPS-Statistics Indonesia.

### 2.2 Web Scraping and Data Preprocessing

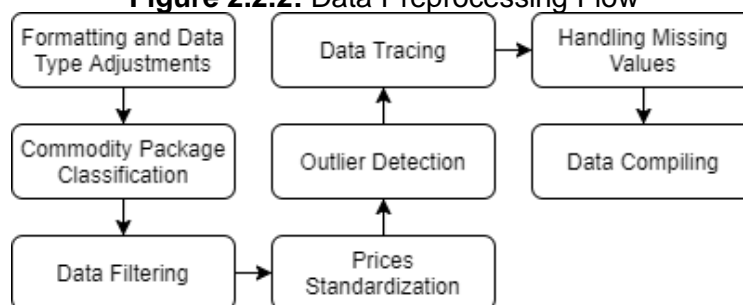
A simple web scraper application scrapes web pages, extracts the required data from the response and then saves the extracted data to a data store such as a flat file (Patel & Patel, 2014). In this study, the web scraper request on the marketplace website with the GET method, this can be done because the marketplace provides an API that makes it easier for researchers to obtain data.

**Figure 2.2.1:** Process to obtain marketplace data with web scraping technique



The web scraper that has been built will be run by searching for commodity data based on BPS commodity packages, whereas a list of commodity package to be included in the search is required. The BPS-Statistics Indonesia commodity package according to the 2018 SBH (Cost of living Survey) consists of 11 Expenditure Groups which consist of commodity goods and services. The data that has been obtained through scraping is then entered in the preprocessing stage as follow in figure below.

**Figure 2.2.2:** Data Preprocessing Flow



### 2.3 Consumer Price Index

In the 2020 CPI Processing Book, Book 5, BPS-Statistics Indonesia uses a modified Laspeyers index method, called the Modified Laspeyers method. In general, the formula used is as follows:

$$I_n = \frac{\sum_{i=1}^k \frac{P_{ni}}{P_{(n-1)i}} P_{(n-1)i} \cdot Q_{0i}}{\sum_{i=1}^k P_{0i} \cdot Q_{0i}} \tag{1}$$

Where  $I_n$  denotes index in period n;  $P_{ni}$  denotes commodity price  $i$  in period n;  $P_{(n-1)i}$  denotes commodity price  $i$  in period n-1;  $P_{0i} \cdot Q_{0i}$  denotes consumption value of commodity  $i$  in base year;  $P_{(n-1)i} \cdot Q_{0i}$  denotes consumption value of commodity  $i$  in period n-1; and  $k$  denotes the sum of commodity in a pack.

In this study, several adjustments were made from the calculation of BPS-Statistics Indonesia. The differences in the counting process are described in Table 2.3.

**Table 2.3:** Differences in BPS CPI Processing and Marketplace CPI

No	Differentiator	BPS-Statistics	Marketplace
(1)	(2)	(3)	(4)
1	Market Weight	Location consists of the name of the market within a market type. However, current weighting has only been applied to the rice commodity	Market location listing was not carried out because the selection of respondents on marketplace data was different from the selection of BPS respondents. Therefore, the weight of the location is neglected.
2	Weighted by Type of Market	Based on the weight of traditional markets and modern markets.	Because it only consists of one type of market, namely the online marketplace, the weight of location type is 1 which in other words can be ignored..
3	Quality Weight	calculating the average commodity price for each quality.	Calculation of the average commodity price is by calculating the average price of all qualities with similar commodities.
4	Commodity Weights	Used in processing CPI at the national level..	This study it did not cover all CPI cities according to the 2018 SBH. So, the calculation of the CPI is limited to the city level of the CPI,
5	Consumption Value	The Basic Consumption Value uses the BPS Base NK for the 2018 SBH results.	The value of total marketplace and BPS consumption by expenditure group is adjusted to the number of commodities covered in the research. This consumption value is used in calculating the CPI for the expenditure group.

### 3. Result and Discussion

In this section, sample results are selected that represents the research. BPS-Statistics commodity packages consist of 835 types of commodities. The results of commodity listings used for search on the web scraper amounted to 535 types of commodities. Meanwhile, the number of commodities obtained was 487 commodities. Where Jakarta has the highest commodities, which is 221 commodities and Sumenep has the lowest commodities, only 4 commodities.

Differences in the concept and definition of BPS, such as the weight of quality and consumption value for calculating the CPI and the unavailability of marketplace supporting data are the main limitations in this study. However, after going through a process of adjustment and approach, this research was able to produce CPI with a

marketplace data approach. Furthermore, a descriptive analysis was carried out regarding the CPI that had been produced in the study. The analysis was carried out in each city at the commodity level The CPI that was compared in the study was the CPI for May 2020.

**Commodity Level**

At the commodity level, a case is analyzed based on the BPS CPI value, namely commodities that generally have the highest BPS CPI. Based on the BPS CPI for May 2020, the commodity that generally has the highest CPI value in each CPI city is the shallot commodity. Which is described as figure below.

**Figure 3.1:** BPS CPI and the marketplace for shallot commodities

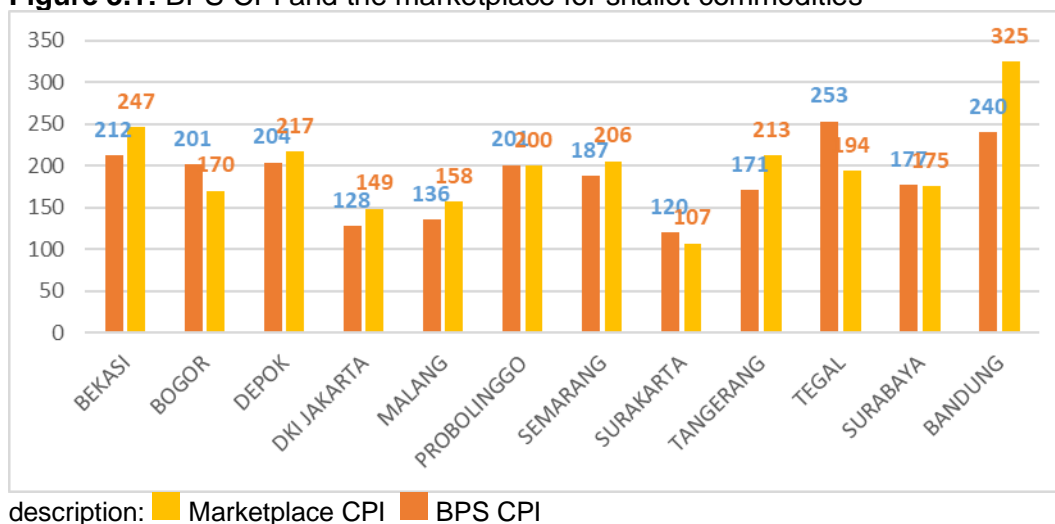


Figure 3.1 shows that the marketplace CPI at the commodity level generally has a higher index than the BPS CPI. Five (5) cities that had lower index scores than BPS CPI were Bogor, Probolinggo, Surakarta, Tegal, and Surabaya. Based on several online news publications, the demand for shallots before Eid al-Fitr has experienced a rapid increase in prices. This incident caused a high CPI for both the BPS CPI and the CPI with a marketplace data approach.

**4. Conclusion**

Based on the results and discussion that has been done, the following conclusions can be obtained:

- a) The construction of a web scraper in one of the marketplaces in Indonesia to obtain marketplace data has been successfully built. The web scraper created can extract data directly into CSV format after receiving a response from the marketplace website. However, it is necessary to pay attention to internet connection when obtaining data.
- b) The CPI with a marketplace data approach at the city level tends to have a higher index value than the BPS CPI. This indicates a difference in relative prices for BPS commodities and marketplace data. The BPS CPI and the marketplace in general have quite a small difference, meaning that the CPI calculation with marketplace data is quite good.

**References**

1. Anandan, R., Sipahimalani, R., Bharadwaj, Alap., Jhangiani J., Kim D., Ramesh, S. (2016). *e-conomy SEA: Unlocking the \$200B Digital Opportunity*. Asia Pasific: Google-Temasek.

2. Badan Pusat Statistik. (2013). *Buku Pedoman Survei Harga Konsumen*. Jakarta: BPS.
3. Badan Pusat Statistik. (2020). *Buku Pengolahan Indeks Harga Konsumen, Buku 5*. Jakarta: BPS.
4. Badan Pusat Statistik. (2020). *Diagram Timbang Indeks Harga Konsumen Hasil Survei Biaya Hidup 2018, Buku 2*. Jakarta: BPS.
5. Badan Pusat Statistik. (2019). *Statistik E-Commerce 2019*. Jakarta: BPS.
6. Bukth, R. & Heeks, R. (2017). Defining, Conceptualising and Measuring the Digital Economy. *Manchester Centre for Development Informatics Working Paper 68*. Manchester, UK.
7. International Monetary Fund. (2020). Consumer Price Index Manual Concepts and Methods. Diakses melalui <https://www.imf.org/~media/Files/Data/CPI/cpi-manual-concepts-and-methods.ashx>
8. Rizkika, T., Siddhikara P,W., Yuniarto, B, Yordani, R., Pramana, S. (2019) Nowcasting modeling of volatile and non-volatile food prices using crowdsourcing data (case study of some food commodities prices on Lombok island in 2015). Kuala Lumpur. *ISI World Statistics Congress, contributed paper session volume 4*, pp. 432-440.
9. U.S. Bureau of Economic Analysis. (March 2018). Digital Economy. Diakses melalui <https://www.bea.gov/data/special-topics/digital-economy>
10. Wijaya, A. & Mariyah, S. (2019). Study of Consumer Price Index based on E-Commerce in Indonesia. *Asia-Pacific Economic Statistics Week. UN ESCAP*. Bangkok, Thailand.