ONLINE PRICE DATA

for CPI Measurement in Indonesia

Chindy Saktias Pratiwi - Directorate of Price Statistics
Irien Kamaratih Arsiani - Directorate of Census and Survey Methodology Development

BPS-Statistics Indonesia

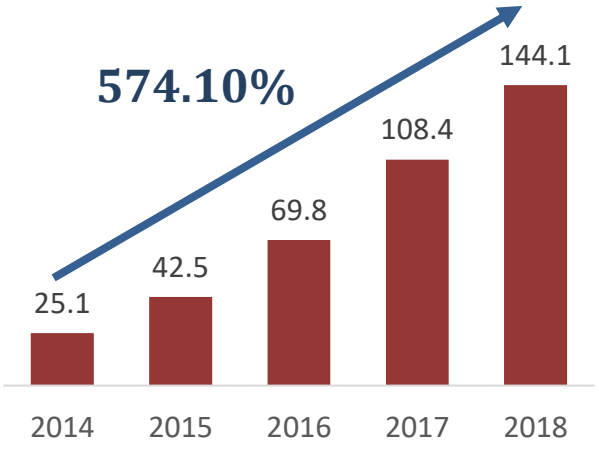
Badan Pusat Statistik
Statistics Indonesia

INTRODUCTION

Indonesia has the potential to become **the biggest digital economy country** in Southeast Asia, both in manufacture and **retail industry**

(BKPM-The Investment Coordinating Board of Indonesia, 2017)

E-Commerce Transaction in Indonesia (2014-2018)
(Trillion Rupiah)



Source: eMarketer (2016)

- Google and Temasek (2017) estimated that before 2025, e-commerce in Indonesia would reach US$81 billion
- The Institute for Development of Economics and Finance (Indef) has estimated that e-commerce contributes just 0.75 percent of GDP in Indonesia (The Jakarta Post, 2018).

AIM OF STUDY

utilize big data in CPI-inflation measurement and see the **feasibility to apply** it on constructing CPI

COLLECTING ONLINE PRICE DATA

FIELD — **C P I** — **ONLINE**

- ✓ CPI covers **82 cities** in Indonesia
- ✓ Classified into **7 major groups** and 34 sub-groups, 7 major groups are:
  - ❑ Food stuff
  - ❑ Prepared food
  - ❑ Beverage, cigarette and tobacco
  - ❑ Housing, water, electricity, gas, and fuel
  - ❑ Clothing
  - ❑ Health
  - ❑ Education, recreation and sports
  - ❑ Transportation, communication, and financial services
- ✓ Visiting and Interviewing all outlet samples

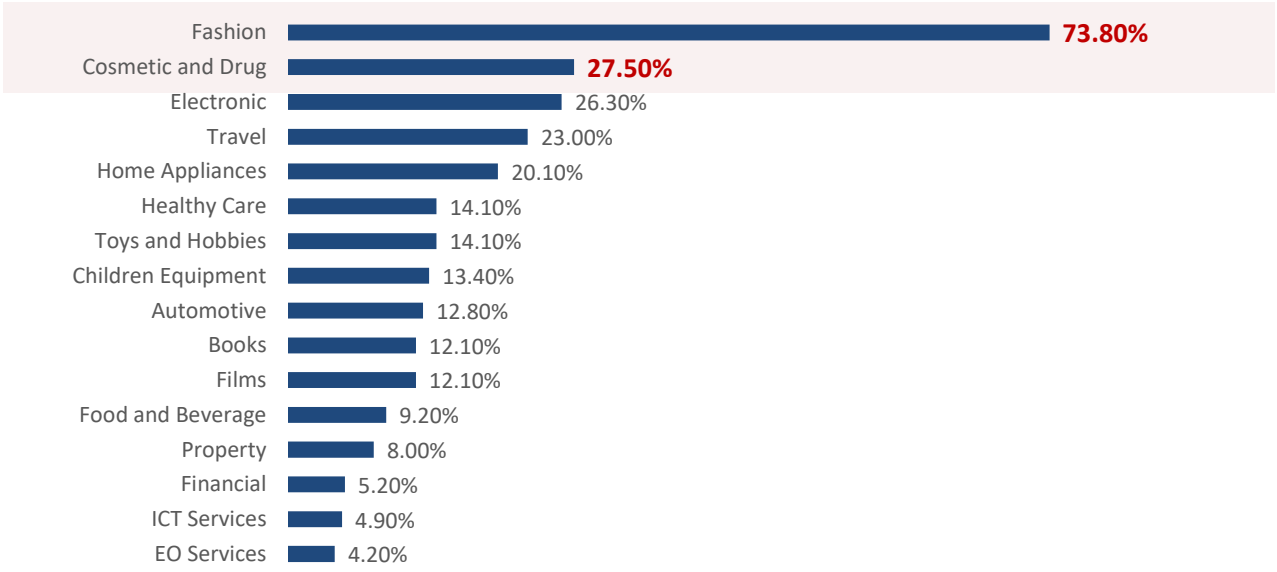
- ✓ Selected city is **Jakarta**
- ✓ Selected groups are:
  - ❑ **Clothing**
  - ❑ **Health**
- ✓ Automatically collected by system

SELECTION



commodities

Product/Services Purchased Online in Last 3 Months



- Clothing
- Health

Source: Ministry of Communications and Information Technology,March 2018

SELECTION



retailers

Top 10 e-commerce players in Indonesia based on their number of visitors shown (at the 4th Quarter 2017)

	Merchant	Monthly Visits	AppStore Rank	PlayStore Rank	Twitter	Instagram	Facebook
1	Tokopedia	115,270,000	#2	#2	155,000	333,000	4,635,000
2	Bukalapak	80,089,000	#5	#4	128,000	233,000	1,949,000
3	BliBli	52,464,000	#6	#5	469,000	149,000	6,805,000
4	Bhinneka	6,863,000	#19	#15	105,000	18,900	1,044,000
5	Matahari Mall	5,629,000	#10	#8	106,000	264,000	1,563,000
6	Jakmall	2,101,000	n/a	n/a	2,400	14,700	104,000
7	Sociolla	1,515,000	n/a	n/a	n/a	162,000	3,000
8	Otten Coffee	1,195,000	#23	#21	7,900	222,000	227,000
9	Alfacart	1,071,000	#22	#16	5,300	22,200	938,000
10	Berrybenka	810,000	#8	#11	17,100	206,000	983,000

Source: iprice.co.id

The most visited multi-channel drug stores in Indonesia based on Alexa's Rank

	Drugstore	Alexa's Rank in Indonesia
1	K24	**2.609**
2	Kimia Farma	**10.419**
3.	Watsons	**11.292**
4	Century	**27.060**

Source: www.alexa.com

Practo.com The Asia's 1st doctor search engine used by over 1M patients every month

MATAHARI Mall .COM #1 eCommerce Try it! | Alfacart .com | K24 | Practo.com

SELECTION

online data extraction software



WebHarvy

1. simple “point-and”click” software solutions for web scraping
2. can be easily configured to perform browser interaction such as Clicking Links, Selecting List/Drop-down Options, Input Text to a field, Scrolling page etc
3. scrap data with various kinds of data display on the website, such as scrape from multiple pages, scrap from the category list, and scrap based keyword list
4. data extracted from web can be saved in a variety of formats, which are an SQL database, Excel, XML, CSV, JSON or TSV file

Online Data Extraction in Practice

defining the URLs

from which the online price data will be extracted

- ❑ Separated process to scrape price data
- ❑ Grouped process to scrape price data

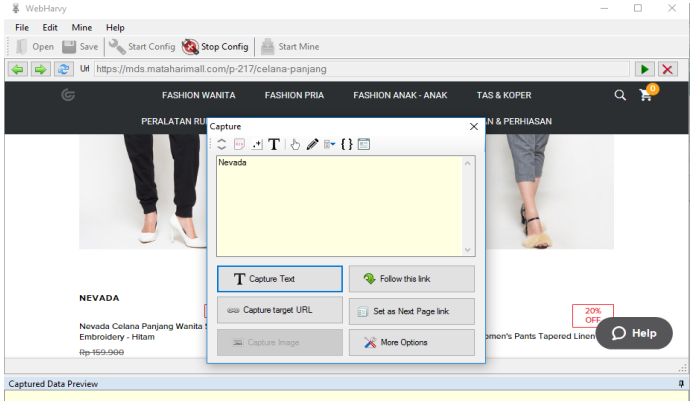
`keyword list`

teaching the software

what pieces of information that should be collected from the webpage

→ setting scraping configuration

- ❑ The software creates “robot” from the configuration



scraping process & extracting data

- ❑ WebHarvy gives options for the user to scrap all pages or just certain number of pages in the website
- ❑ User can choose to export the scraping results as an SQL database, Excel, XML, CSV, JSON or TSV file

CALCULATING CPI FROM ONLINE PRICE DATA

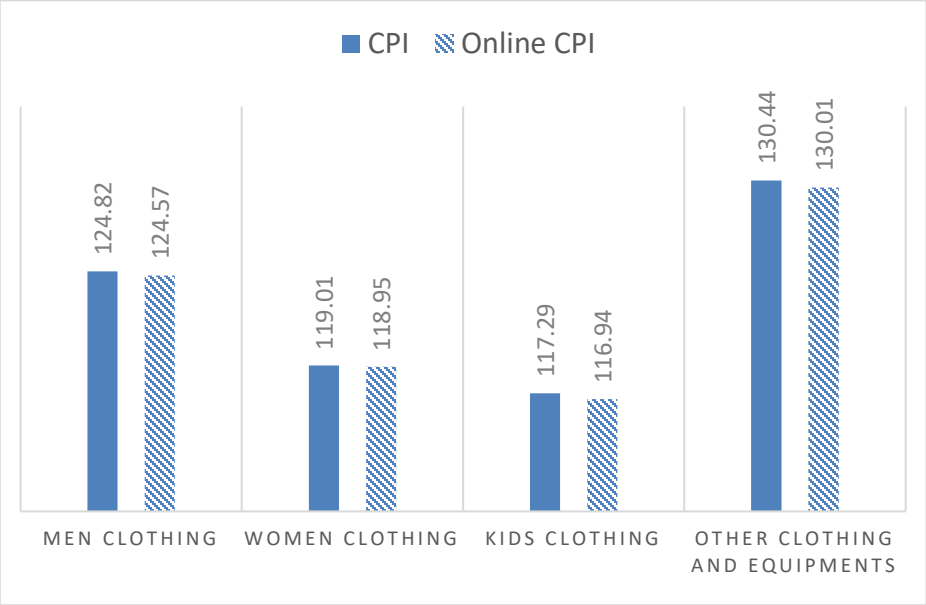
- ✓ This study takes the **weights for commodity basket** by using the available expenditure weights from **2012 Cost of Living Survey of standard CPI**.
- ✓ The first step in data processing is **matching price quote** data by its quality across time. (semi-manually using Ms.Excel)
- ✓ Construct indices using **Modified Laspeyres** Index
- ✓ The collected online price data is produced into **2 indices** of commodity groups and 7 indices of sub-groups.
  - ❑ **Clothing** (4 of 4 sub-groups with 58 commodities – 80.56% online price data)
  - ❑ **Medical Care** (3 of 4 sub-groups with 28 commodities – 62,22% online price data)

The online price data has so many discount prices on its products. Statistics Indonesia has not yet calculated CPI from discount price. Therefore, the discount price of commodities in this study is excluded.

RESULT

Clothing

Using a significance level of 5%, there is no difference between online and offline indices of men clothing; women clothing; kids clothing; and other clothing and equipment.



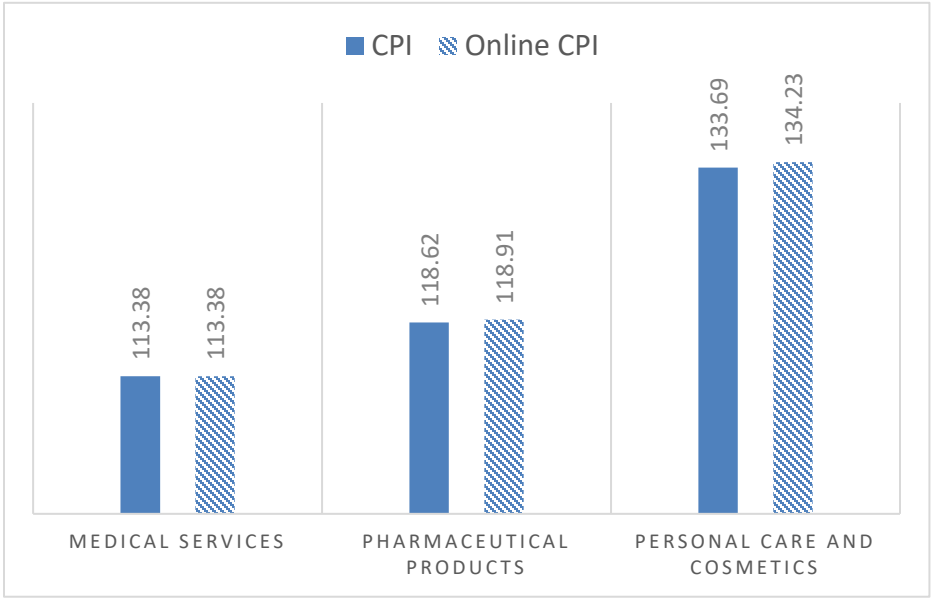
- ✓ The online product has **more variation** type of qualities rather than the offline products.
- ✓ Using web scraping, the collected product has many kinds of qualities with **more detailed specification** comparing with offline data

Source of offline CPI: Official Press Release of Statistics Indonesia 1st March 2018

RESULT

Health

Using a significance level of 5%, there is no difference between online and offline indices of medical services and pharmaceutical products; significantly different between personal care and cosmetics (face cleaner and facial wash (at 5% significance level), and liquid body wash (at 10% significance level)).



Source of offline CPI: Official Press Release of Statistics Indonesia 1st March 2018

- ✓ First, Online CPI of **Medical Services** has the almost the same value with offline CPI. **The pattern of online data in medical services are more stagnant.** The variation is not in the services but in the goods provided in medical services (the prescription drugs).
- ✓ Second, **Pharmaceutical Products available online are more diverse than offline products**. These variability of prices and its changes will affect the construction of index.
- ✓ Last, **both online and offline Personal Care and Cosmetics data had diverse qualities**. The combination of both data will be considered in calculation. Therefore, it needs further examination on how to treat the data.

RESULT

Index Biases and Web-Scraping Online Price Data in Indonesia

1. Quality Change Bias

- Online data gives more detailed specification comparing with offline data.
- all price information are available so that using one method introduced by ILO (2004), the overlapping quality, will be very helpful

2. Elementary Index Bias

This study tried to minimize the elementary index bias by choosing retailer that has both online and offline store.

RESULT

Index Biases and Web-Scraping Online Price Data in Indonesia

3. Substitution Bias

This bias happened when consumers choose to substitute one good for another after its price becomes cheaper than the good they normally buy. Mostly, the online retailer offers a lower price of the same product available offline. Therefore, this behavior will not affect the quantity of product.

4. Outlet Substitution Bias

This study try to minimalize this bias by choosing the most visited online retailer that has offline store. However, there is still high risk for consumer trying to move to discount price available in many online stores.

5. New Product Bias

This study finds that more new products are easily discovered online than offline.
By using web-scrapper, new products will be easily monitored as the data has been formed so that it is easier to review the market

ADVANTAGES

of web-scraping online price data



budget efficiency

more complete data without visiting the stores



time efficiency

Quick data collecting and minimized tme for data processing



minimized quality change

more complete specification for each product quality



more complete data

data inputted for processing is more complete than offline data



minimize the new product bias

CHALLENGES

of web-scraping online price data



The retailers referred to CPI is not easy to be identified in internet



deeper study to decide the 'territory limits' of online retailers



Limit numbers and types of online retailers rather than the offline stores



Most of the online stores in Indonesia does not facilitate user to access historical data



Not all websites provide information about the quantity of product sold by online retailers

CONCLUSION

- Using online price data will be very useful for BPS-Statistics Indonesia to capture a more representative data as the behavior of consumer started to shift.
- Constructing the CPI of Clothing Group and Health Group, this study shows that in order to minimalize the bias, it is important not to randomly choose the online retailers and the available online price data.
- Online price data does not only give obvious advantages but also some limitation for Statistics Indonesia to develop more representative price index.
- Supporting program is needed to be built in order to make cleaning and matching data process more efficient.
- This study gives feedback to analyze a deeper research on online price quote data in Indonesia.