GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

The present work was submitted to the Faculty of Mathematics, Computer, and Natural Sciences

Use Scanner Data in the Consumer Price Index

Bachelor Thesis

By

Senior Industrial Engineer

SENDER Bayanbaatar

Supervisor 1 / Examiner 1 Prof. Lkh. Altangerel

Supervisor 2 / Examiner 2 Mr. L. Bat-Uul

Ulaanbatar/Nalaikh,

16th of May, 2022

GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

Statutory Declaration

Bayanbaatar, Sender	15011365825168
Last Name, First Name	Student ID Number

I hereby affirm instead of an oath that I provided the submitted bachelor thesis

Use Scanner Data in the Consumer Price Index

I did not use any sources other than those stated. In case the work is additionally submitted on a data medium, I declare that the written and the electronic form are completely identical. The work was not submitted in the same or similar form to any examination authority.

Place, Date	Signature

GERMAN - MONGOLIAN INSTITUTE
FOR RESOURCES AND TECHNOLOGY

ACKNOWLEDGEMENTS

First and foremost, I would like to thank all the professors, lecturers, and GMIT staff who have supported me along the way during my bachelor years.

I would like to express my deepest gratitude to Professor Lkh. Altangerel to lead the thesis with wisdom and make it possible. Also, I thank the Head of the “Scopa alliance” company, which provided me with Scanner Data and a warm environment.

Lastly, I am grateful for my family for their continuous support and unconditional love. Without their encouragement, I would not be able to write final touches on my thesis work.

GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

LIST OF ABBREVIATION

CPI – Consumer Price Index

NSO – National Statistical Office

ILO – International Labor Organization

EAN code – European Article Number

HBS – Household Budget Survey

GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

LIST OF FIGURES

GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

LIST OF TABLES

GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

LIST OF EQUATION

GERMAN - MONGOLIAN INSTITUTE
FOR RESOURCES AND TECHNOLOGY

ABSTRACT

The consumer price index is the most important factor in the National Economy and most commonly used by the inflation rate of the Nation. The advance of the new technology as Scanner Data is to give more accuracy to the Consumer Price Index. Some European countries have already started to experience Scanner Data's opportunity and challenges to use.

Therefore, the goal of this bachelor thesis would be learning about the implementation of the International experiences of using Scanner Data and to search about challenges and opportunities to use Scanner Data in the Mongolian CPI. Also, to do calculation on price changes from March to April's yogurts on Ulaanbaatar's 16 retailer's partners of the "SCOP Alliance LLC" company using Tornqvist price index and Fischer price index and made analyses to identify how to change the categories of yogurts sold in different region. The price change index from March to April is around 1.05 that means price of yogurts of 16 retailers rose 5%. The result of analysis, the main categories of yogurts that purchased was „Amtlag" yogurt from SUU JSC, „Sain" yogurt from APU corporation, „Deej" yogurt from APU and „Goyo" yogurt from vitafit.

GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

Table of Contents

1. Introduction

Problem statement

Consumer Price Index measures the rate of changing price from period to period in households purchased basket of goods. It is a main indicator of macroeconomics which is responsible for controlling core inflation, setting interest rates, and calculating the minimum wage of the country. Therefore, even small accuracy of the CPI can affect those economic factors. (1)

Mongolian CPI is calculated by Traditional methodology which explained in Chapter 2 /2.1.1./, and the sampling price is collected by price collectors with the observation method.

With the advance in techniques, the NSO of some European countries (French, Germany, Switzerland etc.) began to use Scanner Data in their CPI. It contains following information such as each article sold in a store on a given day, the number of articles sold, and the sales price. Moreover, scanner data is a big data source which supports National CPI to be more accurate. Scanner Data reduces the cost of NSO and data processing while decreasing the number of physical data collectors.

Purpose

The study's main objective is to identify the challenges and opportunities to implement new data sources in Mongolian CPI based on Scanner Data of partnership retailers with ""SCOP Alliance LLC" in Mongolia. After considering historical data of the company, possible implementation of Scanner Data should be recommended on their CPI at the end of this paper.

Research Questions

1. How does CPI calculate internationally and in Mongolia?
2. What is the international practice of applications of scanner data to the improvement of CPI?
3. How can the structure of the Mongolian basket of goods be changed?

Structure of the thesis

According to the aim of the study, before analyzing the new data source, the theoretical background that implementation of the new data source as scanner data was studied. Then, research the background of Mongolia and another country that uses Scanner Data on their CPI. The last part is to processing and analyzing the CPI with “SCOP Alliance LLC” Scanner Data.

2. State of Art

2.1. Consumer Price Index

CPIs measure changes over time in the general level of prices of goods and services that households acquire for the purpose of consumption. (1)

Basket of goods

A basket of commodities is a group of consumer goods and services whose prices are assessed on a regular basis, usually monthly or annually. The basket of good is used to measure inflation in a particular market or country, so that if the price of the basket of goods rises by 2% in a year, inflation is stated to be 2%. The items in the basket are intended to be indicative of the larger economy and are updated on a regular basis to account for changes in consumer habits.

- A basket of products is a fixed collection of general items produced in an economy, the prices of which are tracked throughout time.
- The basket is used to track inflation over time, like the consumer price index (CPI).
- The basket components are updated and adjusted on a regular basis to reflect current consumer trends and to represent the larger economy. (2)

Weighted average

A consumer price index (CPI) is typically constructed as a weighted average of the relative price changes of the index's included products and services. The weights assigned to each product or service reflect its relative importance as determined by their percentage of total household consumption. The weight influences the impact of a change in price on the total index. Expenditure data may be gathered from a variety of sources, the most popular of which are household budget surveys. The scope and idea of the index will define which commodities and services should be covered in the CPI weights when using various data sources. (1)

The Usage of CPI

The indexation of wages, rents, contracts, and social security payments; the deflation of household final consumption expenditure in national accounts; and the use as a general

macroeconomic indicator, particularly for inflation targeting and setting interest rates are all examples of how a CPI can be used. Given the variety of applications for CPIs, one index is unlikely to perform equally well in all of them. As a result, some governments create multiple CPI variations for different objectives. To avoid confusion, each index should be correctly defined and designated, and a "headline" CPI measure should be expressly recognized. When only one CPI is issued, its kind and scope should be determined by its primary application. If there are numerous major applications, compromises may be necessary for terms of construction. All components of a CPI's construction should be guided by its goal. To ensure that their index serves its function, CPI makers need to understand how it can be used. In this regard, user consultation is essential. (1)

There are main three sectors that are most widely used by National CPI.

Indexation

A CPI can be used to index wages or contracts for any specified group, whether it's a population buying product or a subset of products. In either case, it should accurately reflect the group's coverage. For domestic indexation, the CPI should only include expenses incurred by residents. More broadly, whether the CPI should be a cost of living index (COLI) or a cost of goods index must be decided.

Inflation measurement

It might be claimed that central banks would benefit from a timely index of total inflation, rather than just consumer inflation. However, NSOs are often unable to create such indices, in part due to sampling challenges linked to government consumption. In the absence of such an index, most central banks rely on a CPI, which uses the domestic notion but is assessed on as broad a basis as possible, in terms of both items and geographical coverage.

National Accounts deflation

This application necessitates uniformity between the CPI pricing data and the national accounts spending data. Both data sets should include the same collection of products and services and employ the same ideas and categorization, in general, the Classification of Individual Consumption Based on Purpose (COICOP).

2.1.1. Traditional method of the CPI

Elementary aggregation structure

The methodology of calculating CPI has two stages that are price index estimated for simply elementary aggregates. Then the elementary price indexes are averaged to calculate higher-level indexes by using relative expenditure values of elementary price indexes as a weighted average.

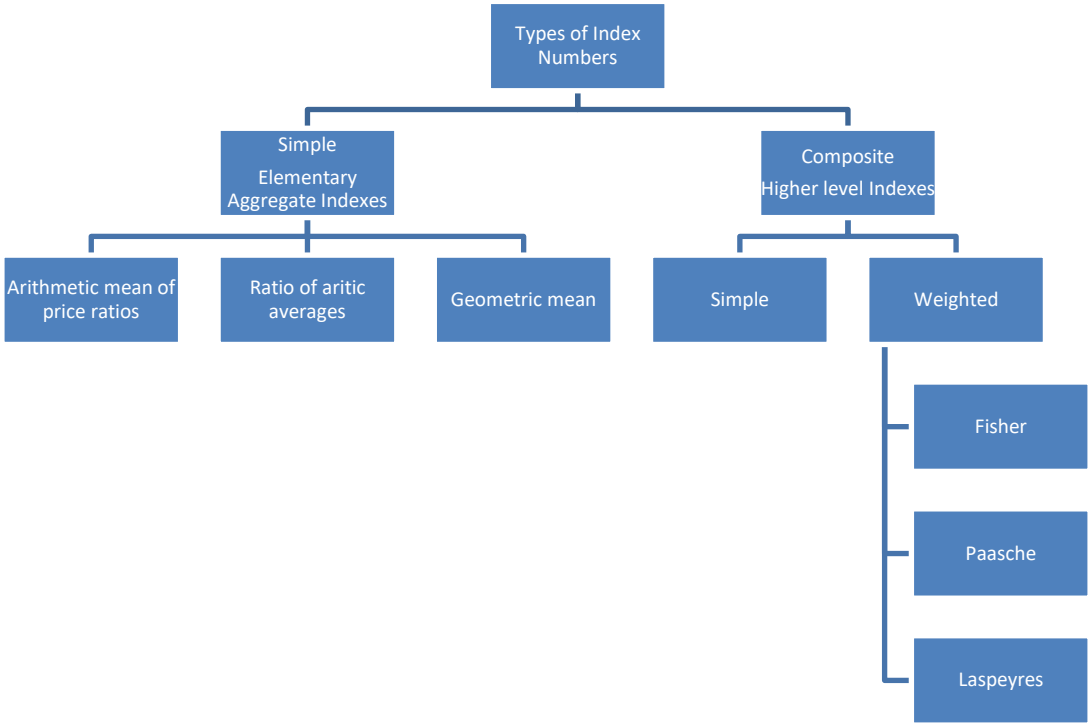

Figure 1: Index formula Classification

The Carli, Dutot, and Jevons formulae are the three most well-known elementary index formulas. All of them are based on unweighted averages of prices or price relatives, and each is connected with a variety of assumptions that will affect measured inflation.

Carli index

The first is the Carli index for i=1,., n items. It is defined as the simple, or unweighted, arithmetic mean. The arithmetic mean of the price ratios – *Carli* Indexle

$$I_C^{0:t} = \frac{1}{n}\sum\left(\frac{p_i^t}{p_i^0}\right)$$

Equation 1: Charli index

Dutot indethe

The Dutot index is defined as the ratio of unweighted arithmetic mean prices:

$$I_D^{0:t} = \frac{\frac{1}{n}\sum p_i^t}{\frac{1}{n}\sum p_i^0}$$

Equation 2: Dutot index

Jevons index

The third index is the Jevons index, which is defined as the unweighted geometric mean of the price relative or ratio that is the same as the ratio of the unweighted geometric mean prices:

$$I_J^{0:t} = \prod\left(\frac{p_i^t}{p_i^0}\right)^{1/n} = \frac{\prod(p_i^t)^{1/n}}{\prod(p_i^0)^{1/n}}$$

Equation 3:Jevons index

Higher index formulas

Laspeyre's price index

The Laspeyres price index is a method used in price statistics to calculate the price evolution of a basket of products and services consumed during the base period.

$$P_L = \frac{\sum_{i=1}^n p_i^t q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \equiv \sum_{i=1}^n (p_i^t / p_i^0) s_i^0$$

Equation 4:Laspeyre's price index

Paasche price index

The Paasche price index is a method for calculating the price development of a basket of products and services used in the present time in price statistics. It solves the issue of how much a basket purchased by customers in the present era would have cost in the base period.

$$P_P = \frac{\sum_{i=1}^n p_i^t\, q_i^t}{\sum_{i=1}^n p_i^0\, q_i^t} \equiv \left\{\sum_{i=1}^n (p_i^t / p_i^0)^{-1} s_i^t\right\}^{-1}$$

Equation 5:Paasche price index

Fischer price index

The Fischer-price index is a method used in price statistics to calculate the price development of products and services using baskets from both the base and current periods.

$$P_F \equiv \sqrt{P_L P_P}$$

Equation 6:Fischer price index

2.1.2. Scanner data Implementation

Scanner data, as defined by INSEE, is information collected by merchants when customers make purchases. They contain the number of products sold and the sales price for each piece sold in a store on a particular day. Retail chains provide them to Insee on a daily basis. These figures are used to compile the Consumer Price Index or to do research on the indicator. (3)

As technology advances, the availability of big data for statistical agencies is increasing. It is worth noting that the CPI mentioned here also increases the expectations of decision-makers as new and big data become available. Therefore, there is a need for the National Statistical Office to present statistical results more efficiently and innovative.

During the 1970s, barcode scanner technology was launched (the first introduction of barcode is in the US in 1974) (1) and it improved and became enabled to actively used retailers in the twentieth century. Scanner data are high in volume and contain information about individual transactions, including date, quantities and values, and detailed product characteristics. (2) The rich data source as scanner data is improved the accuracy of the CPI and reduced the cost of data collectors that collected physically, however this big data source needs different calculation than the traditional one and the cost of data analysis and the process would increase.

Calculating National CPI, International Labor Organization have huge role for CPI. ILO publishing the book about CPI in 2020. From the section of Scanner Data, following some challenges and experienes are mentioned.

EAN code (European Article Number)

Retail, distribution, and warehousing are the most common applications for EAN codes. Individual goods are identified using the EAN (European Article Number), which may

subsequently be connected to quantities or pricing in the store's database. EAN codes are not ideal for the type of complicated information seen in tickets or parcel shipment since they can only contain a limited amount of data.

- Despite the name, it is used everywhere around the world, with the exception of North America, which utilizes the UPC standard.
- Encodes a GTIN with either 13 or 8 digits (EAN-13 vs. EAN-8) (Global Trade Identification Number).
- EAN codes are specified per GS1 standards, and the last number is a mod10 checksum. (4)

Practical Considerations

The availability of scanner data opens up possibilities for improving the CPI. Scanner data sets often feature comprehensive coverage of products sold by a store at all of their locations, as well as amounts sold and income collected by the shop. This information has the potential to: increase the accuracy of the prices used to calculate the CPI by estimating unit values for homogenous products improve the samples of items purchased, with the potential to employ a census of items for the product categories and outlets covered by the scanner data, such as the product categories in the data given by a specific supermarket chain; and utilize quantity or revenue information to weight item samples. Scanner data does not often cover the whole world covered by the CPI. Scanning data, for example, does not include services, rentals, autos, restaurants, or cafe in most nations. Furthermore, this data may be provided only for huge retail chains, not for tiny independent businesses or other sorts of outlets.

Obtaining Scanner Data Sets

Scanner data have been around for decades, and their importance in the production of official statistics has grown over time. Obtaining scanner data sets is one of the challenges that NSOs encounter. There are two primary alternatives. NSOs may request scanner data sets from retailers or third-party data suppliers. Both solutions have advantages and disadvantages.

Several NSOs have negotiated the direct delivery of scanner data from retail firms, which they have utilized to compile their CPI. NSOs could benefit from collecting data sets directly from the retail industry. Negotiation skills are among them:

- The data set is provided for free (or at a low cost).
- The items included in the data set's scope.
- The degree of item aggregation required to ensure homogeneous data.
- The temporal range and level of detail (day, week, or month).
- An agreed-upon timeline for delivering the data collection in order to meet CPI processing requirements.
- To address NSO data requests, a contact officer within the retail business who is familiar with the data set.

Negotiating the direct provision of scanner data sets with retail firms is also difficult. The main difficulty is that the bilateral agreement of scanner data sets include data that could be considered confidential because it contains information on item turnover and quantities.

Another element is the legal and institutional framework that governs NSO-retailer relationships. A memorandum of understanding (or equivalent document) between the NSO and the retail firm normally defines the roles and obligations of each party and tries to secure a continued provision of scanner data to the NSO according to the agreed-upon schedule. Obtaining scanner data sets from intermediaries or market research companies is an alternative to getting them directly from retail shops. Market research companies have scanner data sets received from some NSOs for CPI assessment and compilation. However, obtaining Scanner Data diretly from retailer are prefered, if the company is securitized their data. (1)

Classifying Scanner Data

First, there is no duobt to needed high development of IT system to processing and analyzing the big data source as Scanner Data.

Scanner data sets typically include product categorization that are specific to the merchant. The NSO will almost certainly receive data sets including several product classifications that must be

translated to a single CPI categorization. The classification of scanner data sets may necessitate the use of large NSO resources. The most resources are required when the data sets are first received by the NSO. However, continuing classification efforts are required when new products are added to the data collection.

NSOs have addressed the classification of scanner data items to the CPI classification in a variety of ways, many of which use the classification of the specific retailer.

Such classifications give vital information and can be very valuable if they are at the same (or higher) level of detail than the Classification of Individual Consumption According to Purpose's lowest level (COICOP) When the correlation is 1:1 or n:1 (retailer:COICOP), scanner data can be automatically mapped. In other circumstances, scanner data must be categorized by the NSO or the data is rejected. The store may also change its classification from time to time. The IT system and categorization process should be built to be adaptable in order to accommodate changes in retailer categories. The IT system and categorization procedure should be flexible enough to manage changes in retailer classifications in a timely way. (1)

Multilateral price index changing

Traditional sample-based approaches can be used to integrate scanner data into the CPI. Prices previously observed by visiting price collectors can be substituted with unit values from scanner data without modifying the sampling design or the price index number calculation utilized. (1)

Bilateral price indexes and changing

Tornqvist price index

$$I_T^{0:t} = \prod_{i \in S} \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^0 + s_i^t}{2}}$$

$$s_i^0 = p_i^0 \, q_i^0 \Big/ \sum_{i \in S} p_i^0 q_i^0$$

$$s_i^t = p_i^t \, q_i^t \Big/ \sum_{i \in S} p_i^t q_i^t$$

Equation 7:Tornqvist price index

Multilateral methods

If the NSO wants to estimating all data of the scanner data rather than sampling, the multilateral methods are suitable.

GEKS

$$I_{GEKS}^{0:t} = \prod_{l=0}^{T} \left(\frac{I^{0,l}}{I^{t,l}} \right)^{\frac{1}{T+1}} = \prod_{l=0}^{T} \left(I^{0,l} \times I^{l,t} \right)^{\frac{1}{T+1}}$$

Equation 8: GEKS

Geary-Khamis method

$$I_{GK}^{0:t} = \frac{\frac{\sum_{i \in S^t} p_i^t q_i^t}{\sum_{i \in S^t} p_i q_i^t}}{\frac{\sum_{i \in S^0} p_i^0 q_i^0}{\sum_{i \in S^0} p_i q_i^0}} = \frac{\left[\sum_{i \in S^t} s_i^t \left(\frac{p_i^t}{p_i} \right)^{-1} \right]^{-1}}{\left[\sum_{i \in S^0} s_i^0 \left(\frac{p_i^0}{p_i} \right)^{-1} \right]^{-1}}$$

Equation 9:Geary-Khamis

2.2. Mongolian CPI

Evaluation of Mongolian CPI

The National Statistical Office (NSO) has been calculating CPI since September 1991 using the Laspeyres price index. Until 2008, Mongolian CPI described only Ulaanbaatar's calculation of CPI, however, the Mongolian CPI has been calculated using both Ulaanbaatar's and other Mongolian provinces' CPI. National Statistical Office has updated the consumer basket, weights, and the reference period 5 times. Since 2011 the number of items in the consumer basket increased to 329 (in provinces to 214) the new expenditure weights were derived from the 2010 HH Socio-Economic Survey results and prices of December 2010 were taken as the reference period. Since 2012 regional CPI has begun to be calculated and its 4 regions. Since 2013 CPI calculated by 3 digits of COICOP at the national level. Since 2013 NSO has begun to collect price data from the soums level. 112 out of 330 soums prices were collected every month and averaged in aimag's CPI calculation. Approximately 40-80 item prices were collected from every soums. In 2016, NSO has updated the base year, weights, and baskets. The basket of goods increased to 344 (in provinces to 238). CPI weights derived from the 2015 HBS (Household Budget Survey) results and geometric average prices for 2015 were taken as the reference periods. The geometric mean method is approved for the average price calculation. (5)

Source data collection of Mongolian CPI

Source data collected from selected market's of selected outlets. The prices are collected between the 22nd and 28th in Ulaanbaatar city, between 15th and 18th in provinces and soums, every month, in steady hours. Main food items and petrol's prices collected on a weekly basis and geometric average prices used are in monthly CPI calculation for those items. The national CPI is estimated on the basis of weighted average for 2010 HIES consumption expenditure of provinces and the capital city. Price data are collected mainly through price collectors' personal visits. In the case of the Ulaanbaatar CPI, about 400 shops, supermarkets and service centers are visited, and then checked. The price collected in monthly and some main food items and petrol's prices collected by weekly basis. (6)

Classification of Mongolian Basket of goods

In order to estimate CPI, Classification of Individual Consumption According to Purpose (COICOP) must be used. By using COICOP, the Mongolian Basket of goods is divided into 12 categories. Mongolian NSO takes HBS every 5 years to renew and weight each component of the Mongolian Basket of goods. From the Table 1, Mongolian Basket of goods Classification (6)

№	**Major components**	**Weights (%) /2010/**	**Weights (%) /2015/**
1.	Food and Non-alcoholic beverages	29.3	26.1
2.	Alcoholic beverages, tobacco	3.2	4.4
3.	Clothing footwear and cloth	16.1	16.6
4.	Housing, water, electricity, and fuels	12.3	9.3
5.	Furnishing, household equipment, and tools	4.7	4.9
6.	Health	3.6	3.6
7.	Transport	12.2	14.4
8.	Communication	4.4	4.4
9.	Recreation and culture	2.8	3.1
10.	Education	4.7	4.8
11.	Restaurant and Hotel	2.6	3.0
12.	Miscellaneous goods and service	4.1	5.4
13.	**Total**	**100**	**100**

Table 1: The main 12 components of the Mongolian basket of goods and the Weights percent of each component

2.3. International practice of Using Scanner Data

New Zealand

From the September 2014 quarter, Statistics New Zealand began using scanner data for consumer electronics devices in the New Zealand Consumers Price Index. GfK, a market research company, provides New Zealand Statistics with monthly aggregated scanning data for twelve consumer electronics product categories, including televisions, laptops, and cellular handsets. The Imputation Tornqvist rolling year GEKS (ITRYGEKS) method is used to construct price indexes from scanner data. The ITRYGEKS index is an expansion of Ivancic et al. (2011)'s RYGEKS index, which is built on time-dummy hedonic indexes to ensure that the implicit price movements of new items entering the market and old products leaving the market are reflected. Production procedures have been developed iteratively and are continually evolving. This allows for the flexible and efficient handling of unanticipated difficulties.(7)

New method

Consumer electronics is a continuously changing product category, thus it is critical that we employ procedures that will effectively adjust for changes in the quality of what is purchased.

Also, because consumer electronics goods might have short life cycles, it is critical that new products are included into the index in such a way that the implicit price movement associated with their introduction is reasonable. That instance, if a new product has a low introductory price relative to its set of features, this is a price decrease that the price index should reflect.

In addition to this necessity for adequate quality adjustment, the CPI is non-revisable. This adds another constraint to statistical methods for producing price indexes from scanner data.

GEKS

$$I_{GEKS}^{0:t} = \prod_{l=0}^{T}\left(\frac{I^{0,l}}{I^{t,l}}\right)^{\frac{1}{T+1}} = \prod_{l=0}^{T}\left(I^{0,l} \times I^{l,t}\right)^{\frac{1}{T+1}}$$

Equation 10: GEKS

Geary-Khamis method

$$I_{GK}^{0:t} = \frac{\frac{\sum_{i \in S^t} p_i^t q_i^t}{\sum_{i \in S^t} p_i q_i^t}}{\frac{\sum_{i \in S^0} p_i^0 q_i^0}{\sum_{i \in S^0} p_i q_i^0}} = \frac{\left[\sum_{i \in S^t} s_i^t \left(\frac{p_i^t}{p_i}\right)^{-1}\right]^{-1}}{\left[\sum_{i \in S^0} s_i^0 \left(\frac{p_i^0}{p_i}\right)^{-1}\right]^{-1}}$$

Aside from getting adequate data and determining an appropriate index methodology, some practical challenges stemming from data timeliness and appropriate levels of aggregation have to be solved. An iterative approach to processing system development has allowed the analysis and monitoring processes to evolve in response to unanticipated difficulties, such as the potential for coding of features to alter over time. A variety of forums were utilized to communicate the new development to various user groups at various levels of technical information.

The consumer electronics scanner data production processes in the New Zealand CPI are a prototype system that will inform the more formal development of a Prices 'big data' processing system that will include a variety of data sources such as supermarket scanner data, trade data, telecommunications bills data, and online data. (5)

3. Methodology

3.1. Data collection

As the mentioned above this thesis datas are provided by “Scopa allience” company’s partner retailers. The Scanner Data collected from 16 different retailers as following locations.

№	Number of retailers	Location
1.	Retailer 1	Bayangol district 21st khoroo
2.	Retailer 2	Khan-Uul district 11th khoroo
3.	Retailer 3	Songinokhairkhan district 8th khoroo
4.	Retailer 4	Bayanzurkh district 19th khoroo
5.	Retailer 5	Khan-Uul district 4th khoroo (1)
6.	Retailer 6	Bayanzurkh district 1st khoroo
7.	Retailer 7	Khan-Uul district 21th khoroo
8.	Retailer 8	Khan-Uul district 4th khoroo (2)
9.	Retailer 9	Bayangol district 8th horoo
10.	Retailer 10	Songinokhairkhan district 9th khoroo
11.	Retailer 11	Bayangol district 4th khoroo
12.	Retailer 12	Nalaikh district 8th khoroo
13.	Retailer 13	Sukhbaatar districs 11th khoroo
14.	Retailer 14	Nalaikh district 3rd khoroo
15.	Retailer 15	Khan-Uul district 4th khoroo (3)
16.	Retailer 16	Songinokhairkhan district 22nd khoroo

Table 2: given 16 retailers with location

Data are given as Merchant's name, date and time which the goods were sold, number and price of units sold, sales and barcode. For example:

Merchant's name	Year/Month/ Day/Time	Product label	Number of units sold	Price of units sold	Sales	Currency
Retailer 1	2022/05/16 12:30:56	Pepsi, Cola 500ml	2	1650	3300	MNT

Classification of Scanner Data

After receiving Scanner Data as written on the table from 16 retailers, classification was the next step.

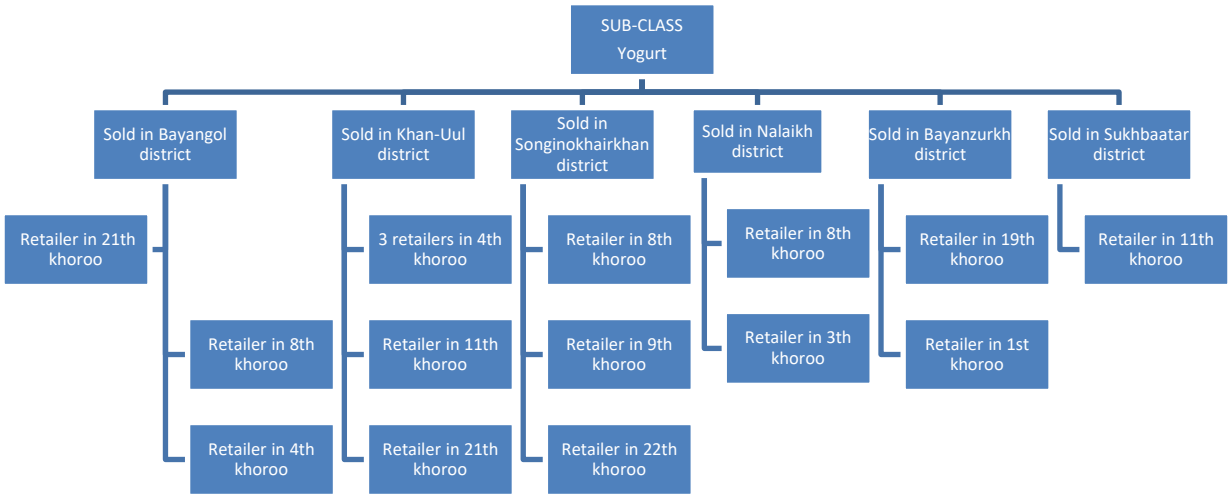

Figure 2: Classification of retailers sold Yogurt

Bar Code

13 digit EAN codes, two types of national barecode products account for the majority in 16 retailer's yogurt categories that are Mongolia /865/ and Russia /460/. Also, other two countries barecode were included are South Korea /880/ and Kazakhstans /487/. (8)

After the classifying yogurt data from whole datas, Yogurt Scanner Data are given from 19th of February to 02nd May 2022. However the dates are covering 4 months, only datas of March and April's whole monthly datas are provided. Therefore, those datas available for use of calculate CPI.

3.1. Used formulas of calculation

The experience of international use of Scanner Data and ILO book suggestion was using Multilateral price index was suitable for estimating Scanner Data. However, the Scanner Data from the company was only two months /March to April of 2022/. The data were enough to use Multilateral price index, if it is at least 13 months.

Therefore, from the Higher price indexes Fischer method and from bilateral price indexes Tornqvist indexes are used in the calculation.

Fischer price index /Higher index formulas/	**Tornqvist price index /Bilateral price index/**
$\boldsymbol{P_F \equiv \sqrt{P_L P_P}}$	$\boldsymbol{I_T^{0:t} = \prod_{i \in S} \left(\frac{p_i^t}{p_i^0}\right)^{\frac{s_i^0 + s_i^t}{2}}}$
Where: $P_L = \frac{\sum_{i=1}^n p_i^t q_i^0}{\sum_{i=1}^n p_i^0 q_i^0} \equiv \sum_{i=1}^n (p_i^t / p_i^0) s_i^0$ $P_P = \frac{\sum_{i=1}^n p_i^t q_i^t}{\sum_{i=1}^n p_i^0 q_i^t} \equiv \left\{ \sum_{i=1}^n (p_i^t / p_i^0)^{-1} s_i^t \right\}^{-1}$	Where: $s_i^0 = p_i^0 q_i^0 \Big/ \sum_{i \in S} p_i^0 q_i^0$ $s_i^t = p_i^t q_i^t \Big/ \sum_{i \in S} p_i^t q_i^t$

4. Results

In this section, following result will be shown

1. Estimation of the price changes percent of overall yogurts of 16 retailers with using Fischer and Tornqvist price index
2. Define the quantities of best sold yogurt each location of retailers

4.1. Calculation of Yogurts of 16 retailer's CPI

Filtering yogurts from the whole Scanner Data there are 167 number of unique name of yogurts.

As mentioned on the Methodology part the main two calculations are used. One is the Fischer price index and another one is the Tornqvist price Index.

The overall calculation of yogurts of Tonqvist price index was calculated first. The formula was given as:

Fischer-price Index

P_F **= 1.050683**

Tornqvist price Index

$I_T^{0:t}$ **was equal to 1.052718005**

This means, the average price of Yogurt from the 16 retailers raised around 5%.

The substitution of the components to calculating the Tornqvist price index and Fischer price index of yogurt were:

$I_T^{0:t}$ Tornqvist price index of yogurts

p_i^t is defined as price of each categories of yogurts during the April

p_i^0 is defined as price of each categories of yogurts during the March

q_i^0 is quantities of each categories of yogurts sold during March

q_i^t is quantities of each categories of yogurts sold during April

All of the categories of sold yogurts can be included, unless one price of the March or April were missed.

Also, the same principles were used in the Fischer price index. Before finding Fischer price index, Laspeyres and Paasche price index should be calculated. Also, the substitution of the formula is the same as the Tonqvist price index's substitution.

Missing price solution

Some of the product's prices were missed from the data of March or April. /More detailed data, especially price changes of each retailer are attached in Appendix/. Consequently, when the price is missed, it should be imputed by the average price of the product to the CPI. However, there are two main reasons to remove from the calculation the missing price rather than averaging. First reason is the duration of the data was two months, the more datas are needed for averaging the missing price. Second, if the missing price datas is removed, the total turnover changes. The total turnliminated. Besides, the total sales of April was changed from 9062490MNT to 8657390MNT. The average percentage of those eliminated missing prices contained 4.8%. As, having low duration of data /two months/ and the contents of missing price were low, the prices are removed from the overall Yogurts.

Following those methods, yogurts of 16 retailers price index were calculated, Fischer price index and the Tornqvist price indexes were calculated as follows. over of the 16 retailers in March of 7340500MNT became 6955400MNT when missing prices were eliminated.

Also, following four categories of products are estimated.

Products	Tornquist price index	Fischer price index
Yogurt	1.0527	1.0506
Potato	1.0683	1.0684
Oil	1.0339	1.0337
Egg	1.0832	1.0922

Table 3: Index of price of change of four products

4.2. Product catergory changed by region

In this section, the type of yogurts, sold quantities and total sales of yogurts are shown in the graphs. From those information's that given from Scanner Data make it possible to identify the categories of most sold yogurts of each retailer. Also, this part will concern with the Sales of yogurts of the 6 retailers to finding what categories of yogurts are sold and how it can be changed by the district.

Out of 16 retailers, 10 of them overall sold quantities of yogurt was below than 150. If the whole sold quantities were lower than 150, the data will be too small to make observation. Therefore, this part will make observations from the remaining 6 stores. These are:

Retailer 1: Bayangol district 21st khoroo

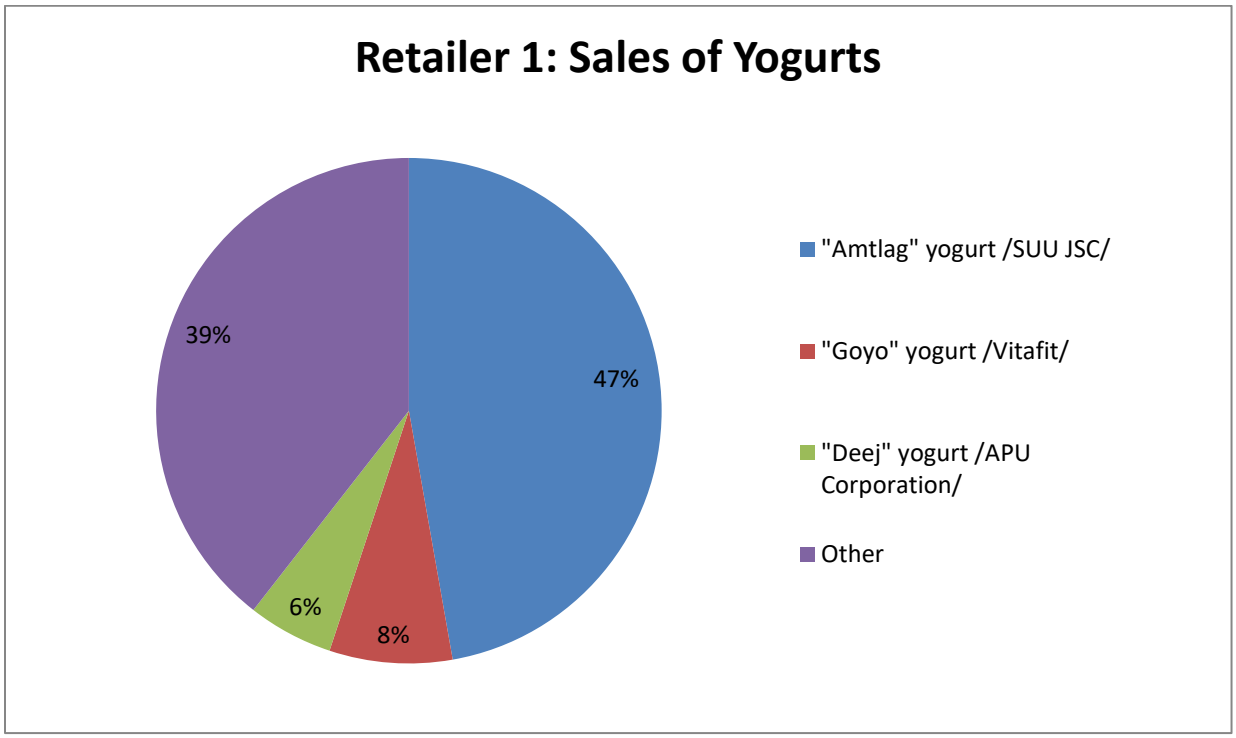

Figure 3: Total sales of Yogurts /Bayangol district 21st khoroo/

In Figure 6, the Bayangol district 21st khoroo's retailer "Amtlag" yogurts from ,"SUU JSC" company were sold the highest in yogurts categories that are 47%. Second, the „Goyo" yogurt

from Vitafit /8%/ and „Deej“ yogurt of APU /6%/ corporation have highest sales in the other categories of product. At last, 39% of total sales of yogurts are contained.

Fischer-price Index

P_F = **1.0412**

Tornqvist price Index

$I_T^{0:t}$ = **1.0402**

Retailer 2: Khan-Uul district 11th khoroo

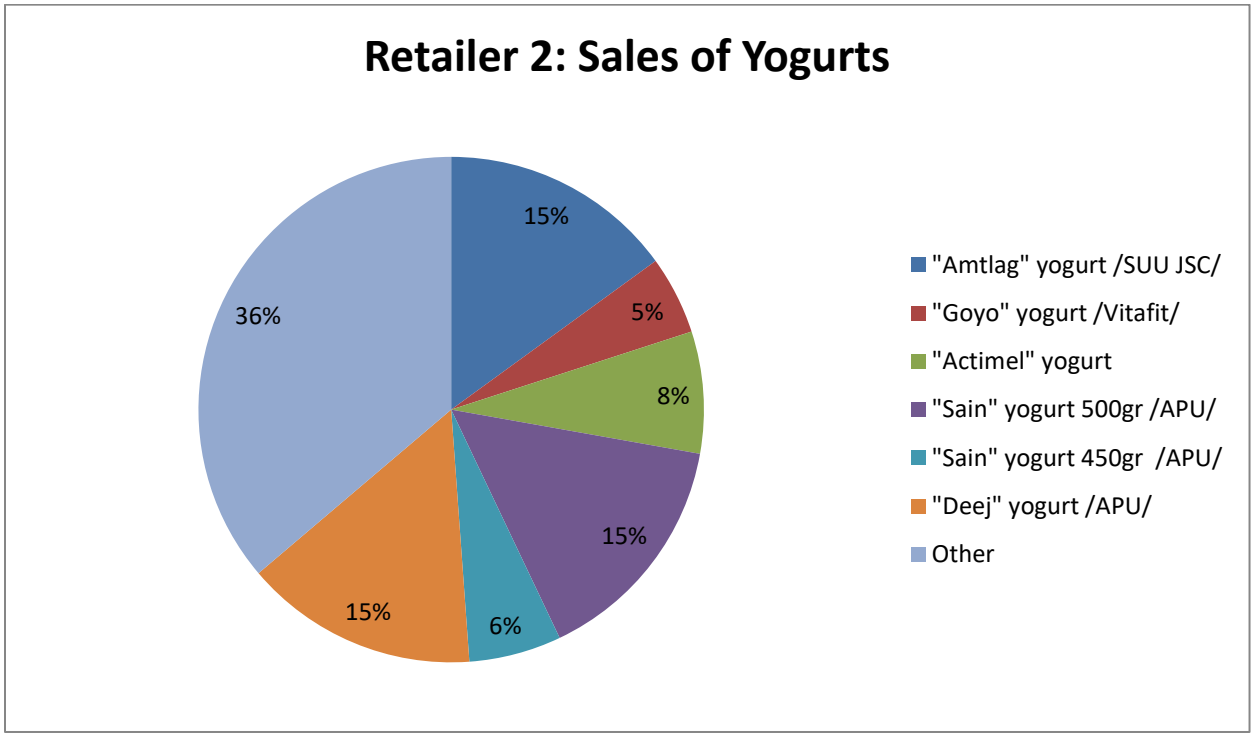

Figure 4: Total sales of sold Yogurt /Khan-Uul district 11th khoroo/

From Figure 8, the main 6 categories of yogurts are dominated in the sales in Retailer 2. 3 of them from APU corporation and it is around 36%. Other categories of yogurts are sold lower percentage as 36%. This means the retailer 2 has not sold many type of products or the sales of other yogurts are high. Also the „Amtlag“ yogurt and „Deej“ yogurts has the same amount of sales.

Fischer-price Index

P_F = **1.0451**

Tornqvist price Index

$I_T^{0:t}$ = **1.0451**

Retailer 3: Khan-Uul district 21st khoroo

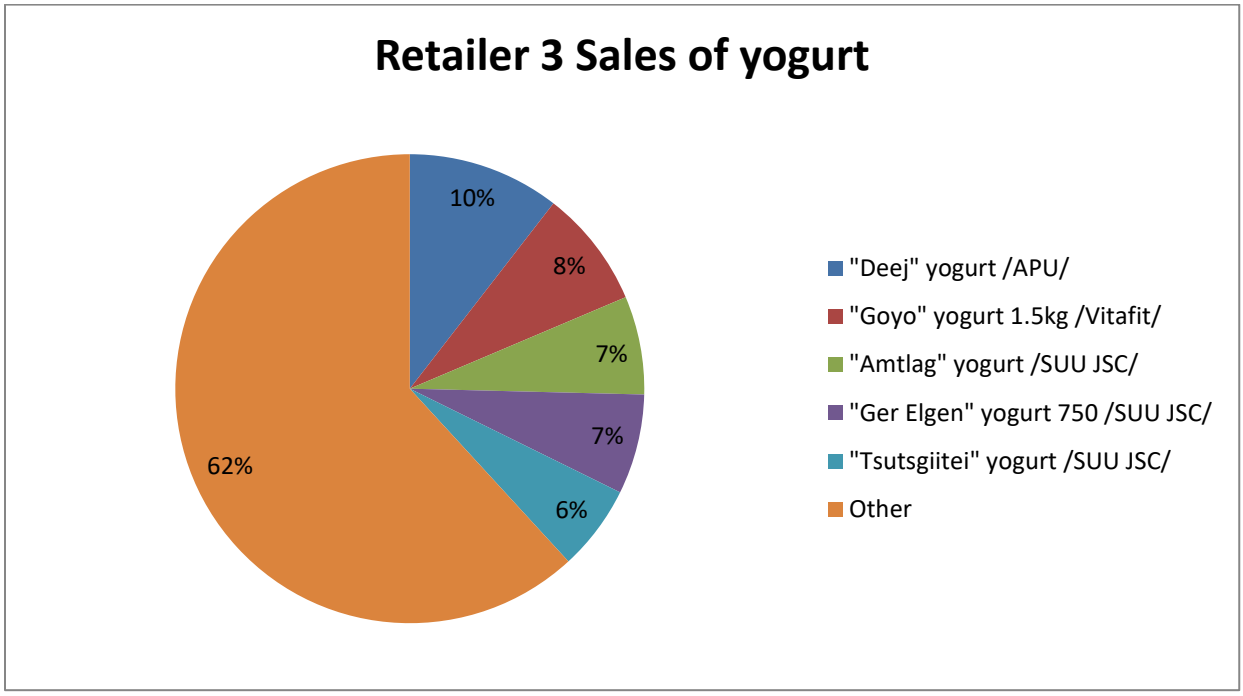

Figure 5: Total sales of yogurts /Khan-Uul district 21st khoroo/

The graph shows us, the main 5 yogurts are most sold in Khan-Uul district 21st khoroo's retailer. "Deej" yogurt from APU JSC was sold the most from others as 10%. Different from the previous graph of retailer 2, "Ger Elgen" yogurt from SUU JSC is included. The highest percentage is 62%, the sales of other categories of yogurts are increased which means retailer of classification of yogurts is high.

Fischer-price Index

P_F = **1.0504**

Tornqvist price Index

$I_T^{0:t}$ = **1.0503**

Retailer 4: Bayangol district 8th khoroo

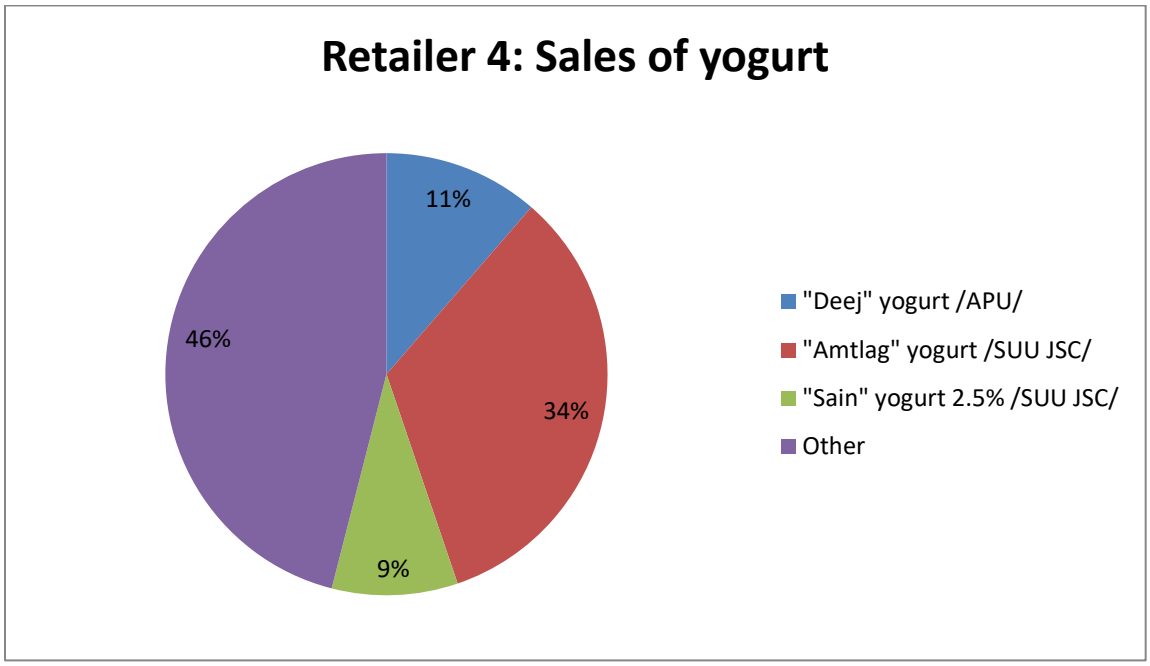

Figure 6: Total sales of Yogurt /Bayangol district 8th khoroo/

From the graph, 34% of the whole sales of yogurt is „Amtlag“ yogurt from SUU JSC. Other categories of yogurts are lower than compared with perviuos retailer. Other two categories which is „Sain“ yogurt from SUU JSC and „Deej“ yogurt from APU has higher percentage /20%/ in sales.

Fischer-price Index

P_F **= 1.0565**

Tornqvist price Index

$I_T^{0:t}$ **= 1.0568**

Retailer 5: Songinokhairkhan district 9th khoroo

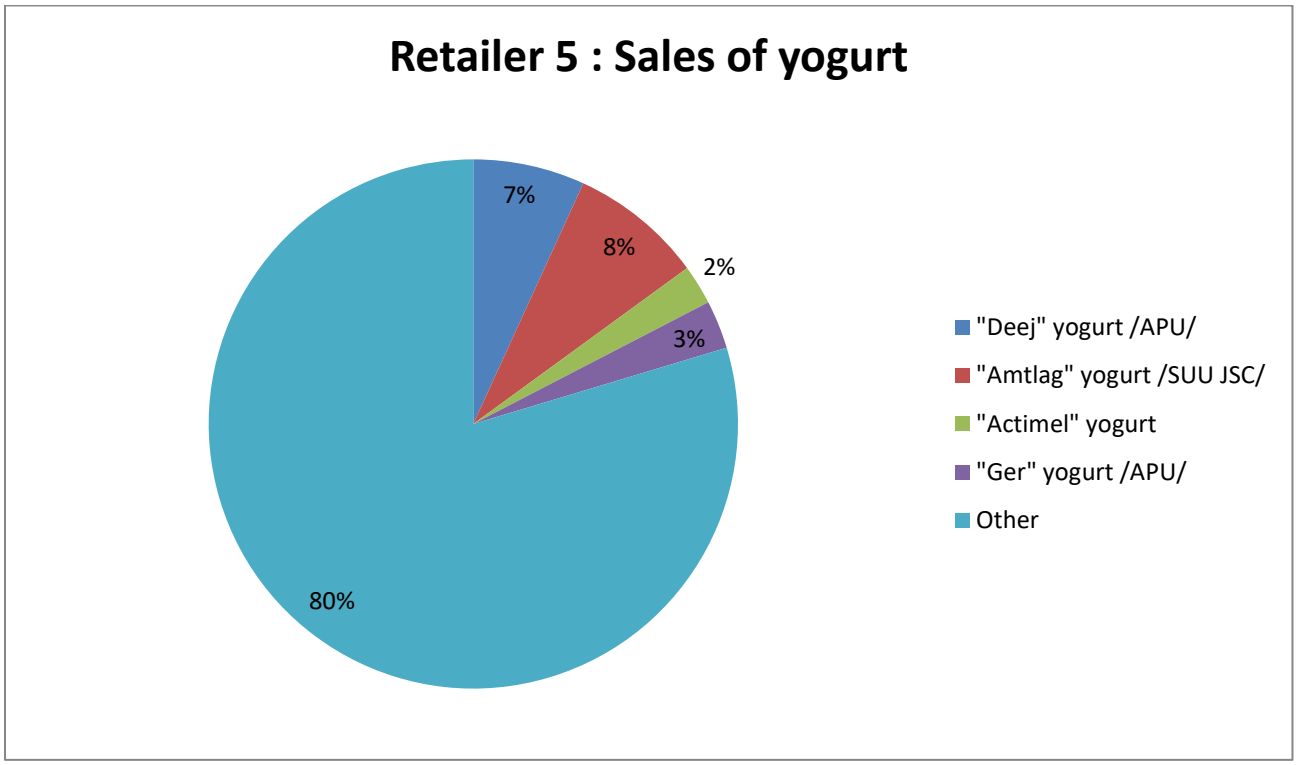

Figure 7: Total sales of Yogurts /Songinokhairkhan district 9th khoroo/

The highest part of the sales is other categories of yogurts. The main yogurts sold in retailers the most is just 20% of whole sales. 20% of the sales contain the 4 categories of yogurt which is „Deej“ yogurt from APU, „Amtlag“ yogurt from SUU JSC, „Actimel“ and „Ger“ yogurt from APU corporation.

Fischer-price Index

P_F **= 1.0490**

Tornqvist price Index

$I_T^{0:t}$ **= 1.0490**

Realler 6: Bayangol district 4th khoroo

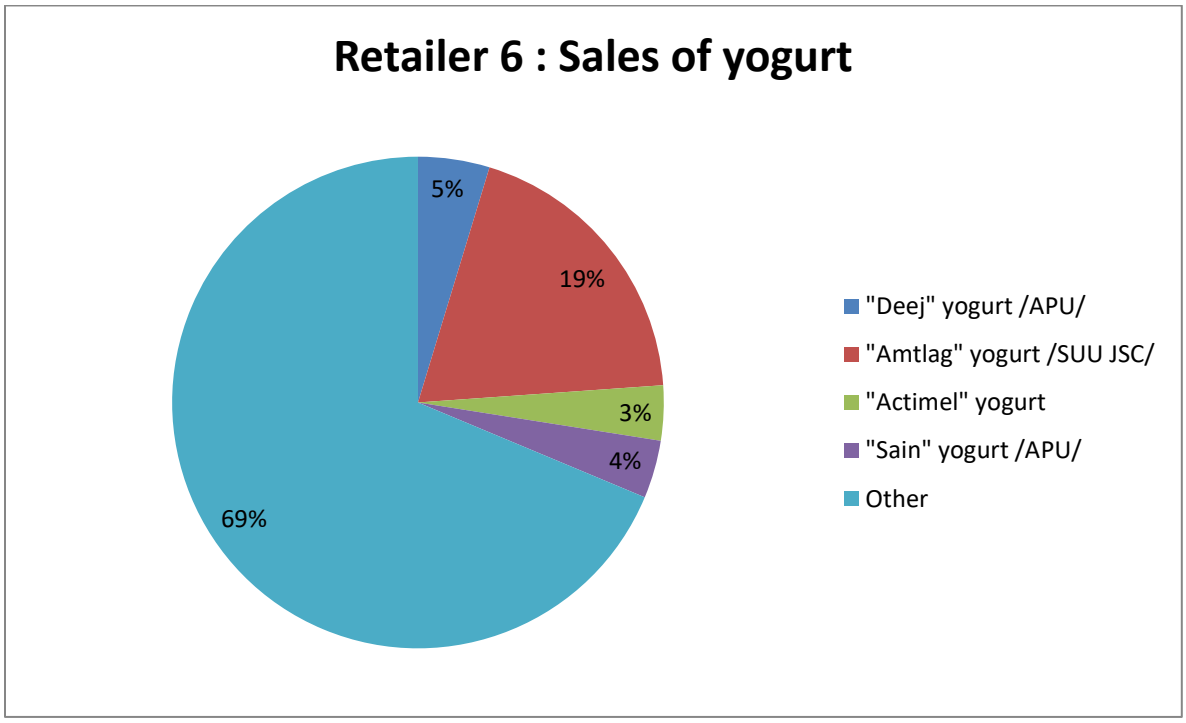

Figure 8: Total sales of Yogurts /Bayangol district 4th khoroo/

The graph shows that 69% of the whole sales contain the other categories of yogurts. Other 31% of the sales contain the 4 categories of yogurt which is „Deej“ yogurt from APU, „Amtlag“ yogurt from SUU JSC, „Actimel“ and „Sain“ yogurt from APU corporation.

The main categories of yogurts that purchased was „Amtlag“ yogurt from SUU JSC, „Sain“ yogurt from APU corporation, „Deej“ yogurt from APU and „Goyo“ yogurt from vitafit from the observation.

5. Conclusion

With this thesis, Scanner Data on CPI on the local example of “SCOP Alliance LLC” while studying the international case of New Zealand was performed. This was achieved by market research in Mongolia, especially 16 retailers from “SCOP Alliance LLC”.

The panel of the study aimed to answer the questions “How does CPI calculate internationally and in Mongolia?”, “What is the international practice of applications of scanner data to the improvement of CPI?”, “How can the structure of the Mongolian basket of goods be changed?”

Using Scanner Data on national CPI has a lot of benefits even though it was challenging. If Scanner Data started to use in Mongolian CPI, the staffs will need who can analyze the high amount of data, potential of IT development will also need.

5.1. The benefits and challenges using Scanner Data

Challenges

The several challenges can occurred when using Scanner Data in CPI. Throughout the reading ILO book about Consumer Price Index, it mentioned that the negotiating and contracting with the retailers to obtaining the data is the challenging part. However, in this thesis Scanner Data are provided by “SCOP Alliance LLC”. The one of the challenging part of my research was solved by the company. Second, Classification of Data was most challenging part of the processing and analyzing the data. Scanner Data is the huge data with lot of churns. In order to classify the data, separating each retailer was the first thing should do. After that using COICOP classification, it would be easier. After, the classifying the Scanner Data, the data should be process, analyzing and implemented into the equation.

Benefits

From this research Scanner Data is a huge data can strongly increase the accuracy of the CPI. The understanding about the using Scanner Data would changes that Scanner data not only to using price changes, but also can use the basket of goods of the Nation and can impacted on

weights. One of the biggest information that can Scanner Data give is Sales of the purchased products, therefore it can be used for the see economic side as market research.

5.2. Further Recommendation

The advance of the technics, international CPI using Scanner Data will increases. Therefore, the following issues to concern for the future.

- Data classification would needed more advanced technologies such as machine learning
- The staffs needed who can analyze the advanced data
- At least 13 months of data needed when using Multilateral price index
- Negotiation and the make contract with the retailers are important

GERMAN - MONGOLIAN INSTITUTE FOR RESOURCES AND TECHNOLOGY

5. References

1. International Labour Office. | International Monetary Fund. | Organisation for Economic Co-operation and Development. | European Union. | United Nations. | World Bank. CONSUMER PRICE INDEX MANUAL:Concept and methods. Washington, D.C; 2020.
2. Монголын Үндэсний Статистикийн Газар. ХЭРЭГЛЭЭНИЙ ҮНИЙН СТАТИСТИКИЙН МЭДЭЭНИЙ ТУХАЙ, ТООЦОХ АРГА ЗҮЙ. Ulaanbaatar; 2018.
3. EAN Barcode Scanner | Scanbot SDK [Internet]. [cited 2022 May 17]. Available from: https://scanbot.io/en/sdk/scanner-sdk/barcode-scanner-sdk/ean-code
4. 3 EXPENDITURE WEIGHTS AND THEIR SOURCES.
5. Definition - Scanner data | Insee [Internet]. [cited 2022 May 17]. Available from: https://www.insee.fr/en/metadonnees/definition/c2159
6. ҮСХ-ны даргын 2016 оны 12 сарын 22-ны өдрийн А/71 тоот тушаалын хавсралт. “Хэрэглээний үнийн индекс” тооцох аргачлал .
7. workshop 4: Scanner data 26 May. Expert group meeting on CPI . 2014.
8. EAN and EAN codes - What is the difference? | Weber Marking Systems [Internet]. [cited 2022 May 17]. Available from: https://www.weber-marking.com/blog/ean-code-is-not-ean/

6. Appendix

Product label	Sales of Yogurt /MNT/
"Amtlag" yogurt /SUU JSC/	941220
"Goyo" yogurt /Vitafit/	156600
"Deej" yogurt /APU Corporation/	108900
Other	786300
Total	**1993020**

Appendix 1: Data of sales of Yogurt /Bayangol district 21th horoo/

Product label	Sales of yogurt /MNT/
"Deej" yogurt /APU/	232900
"Goy" yogurt 1.5kg /Vitafit/	180300
"Amtlag" yogurt /SUU JSC/	150400
"Ger Elgen" yogurt 750 /SUU JSC/	153000
"Tsutsgiitei" yogurt /SUU JSC/	130600
Other	702100
Total	**1549300**

Appendix 2: Data sales of Yogurt /Khan-Uul district 21st khoroo/

Product label	Sales of yogurt /MNT/
"Deej" yogurt /APU/	603600
"Amtlag" yogurt /SUU JSC/	1772600

"Sain" yogurt 2.5% /SUU JSC/	488600
Other	2441400
Total	**5306200**

Appendix 3: Data sales of Yogurt /Bayangol district 8th khoroo/

Product label	**Sales of yogurt /MNT/**
"Deej" yogurt /APU/	371200
"Amtlag" yogurt /SUU JSC/	443000
"Actimel" yogurt	131600
"Ger" yogurt /APU/	160600
Other	4331400
Total	**5306200**

Appendix 4: Data of sales Yogurts /Songinokhairkhan 9th khoroo/

Product label	**Sales of yogurt /MNT/**
"Deej" yogurt /APU/	258750
"Amtlag" yogurt /SUU JSC/	1055800
"Actimel" yogurt	198900
"Sain" yogurt /APU/	209850
Other	3781800
Total	**5306200**

Appendix 5: Data of sales Yogurts /Bayangol district 4th khoroo/

Number of retailer	Tornquist price index	Fischer price index
Retailer of **Bayangol district 21st khoroo**	1.0412	1.0402
Retailer of **Khan-Uul district 11th khoroo**	1.0451	1.0451
Retailer of **Khan-Uul district 21th khoroo**	1.0504	1.0503
Retailer of **Bayangol district 8th khoroo**	1.0565	1.0568
Retailer of **Songinokhairkhan district 9th khoroo**	1.0490	1.0490

Appendix 6. Index of price changes of five retailers