Central Bank of Armenia
Statistics department

The use of Big Data in Central Bank of Armenia

G.Lazyan, T.Baghdasaryan, G.Aghajanyan

Abstract

In this modern world, where the technology develops in a very high speed, more and more data are generated, especially by web and electronic devices. The concept of "Big data" arose in order to describe the huge amount of generated data in the world, which could be characterized with the "3 V's": volume, velocity and variety. The use of big data by policy makers and researchers became more popular and acceptable in recent years. The Central Bank of Armenia does not fall behind the other policy makers in the world and have started the collection of Big data since 2016. The main source is Internet data sets. The sources of the data are online markets, supermarkets, service providers, realty agencies, employment agencies, etc. At this moment the main goals of using this information is to follow real-time price dynamics of goods and services in the market (food, non-food, services) and get flash estimates of the CPI, track housing prices in order to explore real estate market and compute housing price index and finally estimate the demand in labor market by industries of economy and type of occupation. The technology used to access the data is web scraping.

Keywords: Big Data, price statistics, real estate market, labor market, central banks

Contents

Introduction

In this pace of technological progress, more and more data are generated from web and various devices and sensors. This vast amount of generated data is known as "Big Data". It could be described by "3V"s: volume, variety, velocity. Every single second and almost everything around us is generating data: mobile phones, computers, smart watches, cameras, Facebook and twitter posts and so on. These vast amounts of data are very different from each other and generally unstructured. Some authors may add another "V"s, like veracity which refers to the accuracy of the data, valence – the connectedness of the data, variability, value, etc.

Nowadays big data is widely used in both business and public policy decisions. Many governments and central banks around the world already use big data statistics in their policies.

The Central Bank of Armenia (CBA) have started the collection of big data since 2016. The main objectives are to generate flash estimate of consumer price index (CPI), examine real estate market and use the outcome in the estimation of Housing price index in the future, and to get estimates of labor demand. The main sources of data are supermarkets', realty agencies' and labor organizations' websites.

We will first discuss the main challenges of CBA, and then will turn to the sources of data, methodology and the results. Lastly, we will talk about main limitations that the CBA met and future contributions to this paper.

The main challenges for Central Bank of Armenia

As we have mentioned in the introduction one of the main objectives of big data use is to get a flash estimate of CPI. The National Statistical Service of the Republic of Armenia (NSSRA) publishes preliminary CPI in the 3rd working day after the reference month.[1] As an authority, responsible for price stability, CBA needs the estimates of CPI as soon as possible in order to implement an effective forward-looking policy. Understanding this and taking into account the wide coverage of modern statistics and data availability, having resources and an opportunity as well the Statistics department took up the matter to collect price data from supermarkets' websites and compute flash estimates of CPI.

The next challenge is important from the point of view of both monetary policy and financial stability. Housing price index (HPI) is widely used in the world, although it is not available yet in Armenia. Thus, we decided to try to analyze real estate market in order to get some trends and characteristics of housing market that will also help in the more precise estimation of the HPI.

Finally, the last, but not least, challenge is the estimation of labor market demand. It could be used as a leading indicator for business sectors. NSSRA collects a good statistics on labor market, and do have indicators describing the demand of labor market, but it takes one month to get the data and it is grouped by industries, not by profession.

These are only the challenges the Statistics department set in 2016 for the first time, and in case we succeed, the scope of the challenges will get larger in the future.

[1] National statistical service of the Republic of Armenia

Data and methodology

In this section, we will talk about data and the main results of our analysis based on this data. We have used only data from internet datasets, using web-scraping tools with different intervals (generally daily data) depending the data availability and update interval.

Consumer price index

In order to get flash estimates of CPI, we started the collection of price statistics in 2016 using supermarkets data, which have an online web store. After the research both on internet and business register, we found only two supermarkets, which have listed all its products in its website. One of them has about 4000 listed items and another one about 9000.[2] For our analysis, we took only one supermarket data, which has more listed items. The interval in which we scrape the data is 10 days, so that we have price data for every 10th, 20th and 30th day of a month. We tried to get close to the methodology used by NSSRA for the computation of CPI as much as possible. Based on this methodology, 470 goods and services are included in the basket to compute the CPI, which divided into 3 main groups: food, non-food and services, with corresponding weights. Seasonal products (food) have a large share in Armenian consumer basket and usually play a "key role" in CPI fluctuations. Therefore, we took only food products and created indices of each subgroup of products to the previous period, as we only had a one-year data. In the first step, we have created indices to the December of previous year for each product, based on the daily price data. Then weights (set by the NSSRA) were applied in order to get the price index of each group of products and divided by the corresponding index in order to get price indices over the previous month. The graphs in Appendix A show the comparison of the official (published by the NSSRA) and the computed indices based on our analysis.

As we can see from the graphs, although errors from the official index are sometimes larger, the overall trends are pretty much the same.

Further contribution to this challenge would be a computation of CPI using different weighting methods like principal components analysis or factor analysis, data envelopment analysis etc.[3]

Real estate market

Monitoring the trends in real estate market plays a vital role from central banks perspectives. Changes in real estate market may have an important impact on the monetary policy transmission mechanism and may have influence on aggregate demand and inflation. However, there is no reliable and representative indicator of housing price changes in Armenia. Developing such a measure is very challenging and one of the main problems is the availability of data.

Starting from January 2017, we scrape the web page of one of the largest realty agencies in the market, which includes more than 8000 announcements. The main advantage of the collected data is that it is very detailed; there are more than 40 properties of apartments:

- Full address, including city, region and district
- Number of Rooms and bathrooms
- Area in sq.m.
- Ceiling height
- Floor/total floors

[2] We scrape the information about the name (including the item name, type and weight) and price.
[3] "Handbook on constructing composite indicators: methodology and user guide", OECD, 2008

- Building type (monolit, panel, stone, other)
- Condition (newly repaired, good, zero condition)
- Available facilities (gas, water, heating, view, close to a bus station, etc.)
- Price

This will enable us to construct a hedonic price index. However, indices based on advertised prices have a major drawback. Houses can be withdrawn from market and the agreed selling price may not equal the seller's asking price.[4] Taking into account the mentioned drawbacks the data collected might have some use at least being considered as an indicator of housing supply price.

The main idea of hedonic regression method is that heterogeneous goods can be described by their attributes or characteristics. That is, a good is essentially a bundle of (performance) characteristics. In the housing context, this bundle may contain attributes of both the structure and the location of the properties. There is no market for characteristics, since they cannot be sold separately, so the prices of the characteristics are not independently observed. Price index is calculated using dummy variable hedonic model.

$$ln\ p_n^t = \beta_0 + \sum_{t=1}^{T} \delta^t D_n^t + \sum_{k=1}^{K} \beta_k z_{nk}^t + \varepsilon_n^t$$

Where β_0 and β_k are the intercept term and the characteristics parameters to be estimated.

The time dummy variable D_n^t has the value 1 if the observation comes from period t and 0 otherwise.

According to the results obtained, the price index of February 2017 compared to the previous month is equal **100.15%**.[5] In comparison, the preliminary estimations of price index computed based on the State Committee of Real Estate Cadastre of RA data is 103.5%.

Labor market

Labor statistics is very developed in Armenia, and NSSRA collects really a wide range of information and has many indicators describing labor market. However, sometimes for policy makers it is necessary to get the data much earlier than it is published. That is why we started the collection of job announcements in a daily basis. We use web-scraping technics again to get the data of online job announcements. The information include the position to which a person is applying, the industry the company is in, the starting date and deadline for application, and approximate salary (not always shown). This data could be useful in creating a leading indicator of each business sector.

[4] "Handbook on Residential Property Prices Indices (RPPIs)" European Union, International Labor Organization, International Monetary Fund, Organization for Economic Co-operation and Development, United Nations Economic Commission for Europe, The World Bank, 2013

[5] See full Stata output table in Appendix B

Limitations

On the way of achieving our goals, we faced some limitations, which delayed or created obstacles, which need to be solved or avoided. Here are some of them:

- Not too much online supermarkets**.** The price sources are not too much, and this fact has a negative impact on reliability of the data. It is very difficult to follow the price dynamics of a particular good or a group, since it always changing: you may not find the item in next two months or six months later, and it is very difficult to find a substitute one. The problem may also be behind the website or web scraping tool.

- No Application-programming interface (API) available**.** During this time, no any supermarket or agency provided an API in order to access the data easily. This is one of the main technical issues in our way.

- Sometimes data is not much reliable**.** This mainly refers to online announcements. Anyone could create an announcement on the website and moderators not always concerned about it and just move forward into the website. Sometimes one could see an apartment for only 10$. These types of data should be considered during the analysis.

Further contribution

As a further contribution for big data usage in CBA, we can mention the followings:

- Further and deep analysis of price data in order to get an indicator comparable with CPI

- Implementation of the methodology of Housing price index in Armenia based on the data from realty agencies

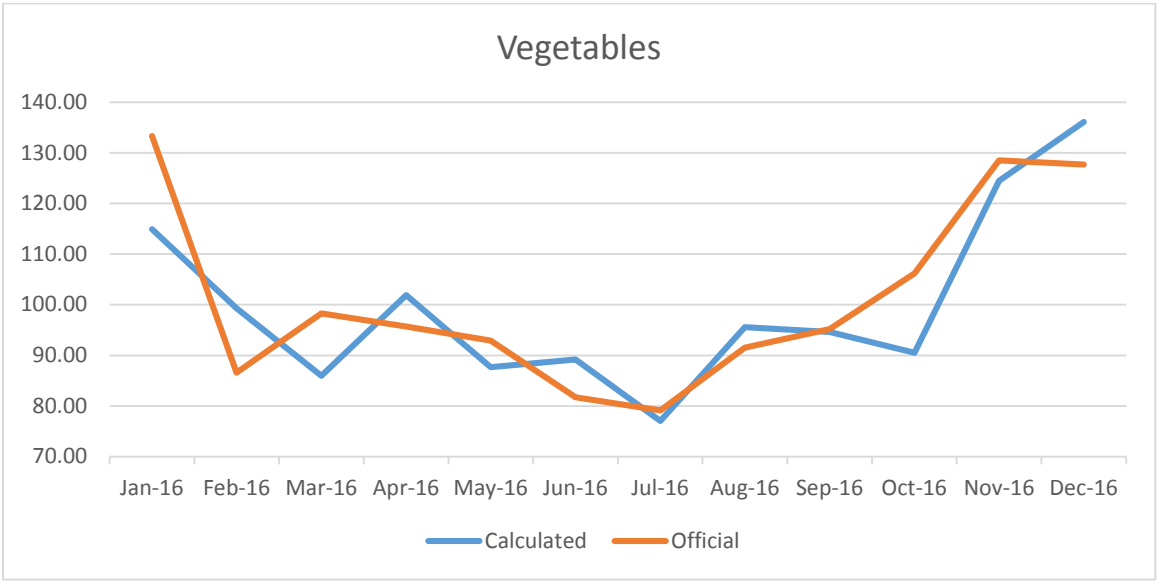
- Widen the range of the usage of big data in Armenia in terms of administrative and social media sources

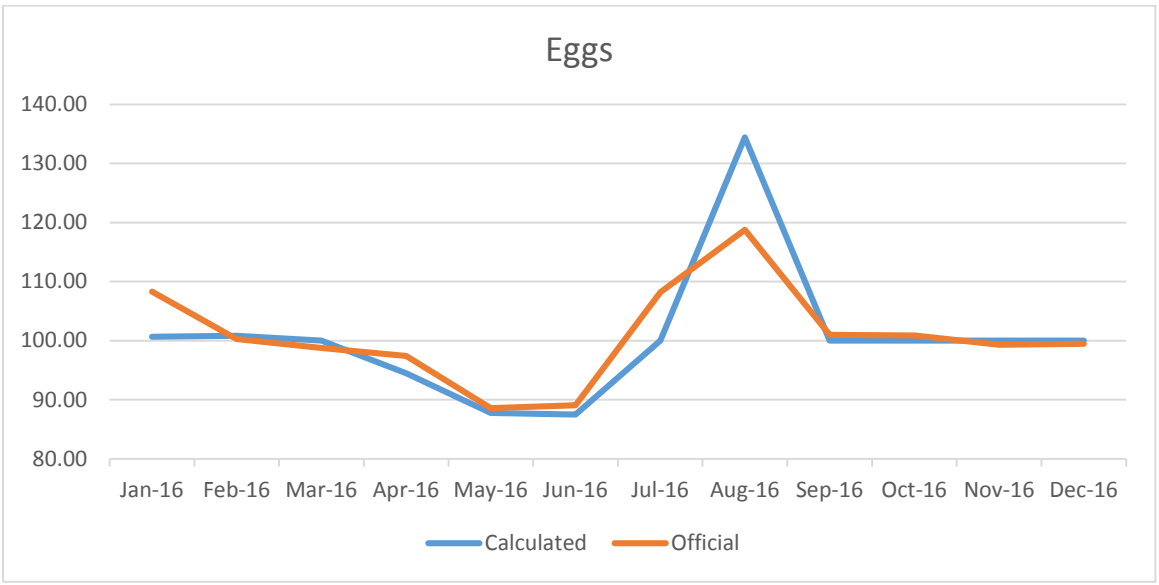
Conclusion

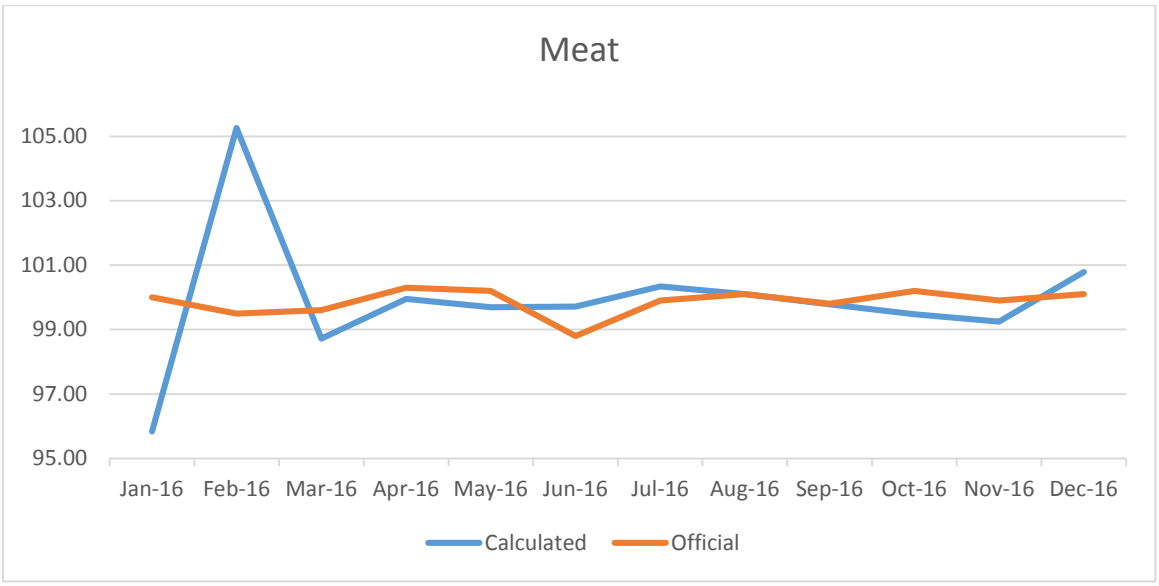
Concluding we can say that nowadays the use of big data is not only useful but it is already necessary for central banks, governments and other authorities and national statistical services, of course parallel with the official statistics. Big data is not yet substitute but complementary for the official statistics and it requires new technics and methodology other than traditional. It could be used in different aspects and for different purposes: in tourism statistics, traffic and transport statistics, price statistics and also in consumer behavior via social media.

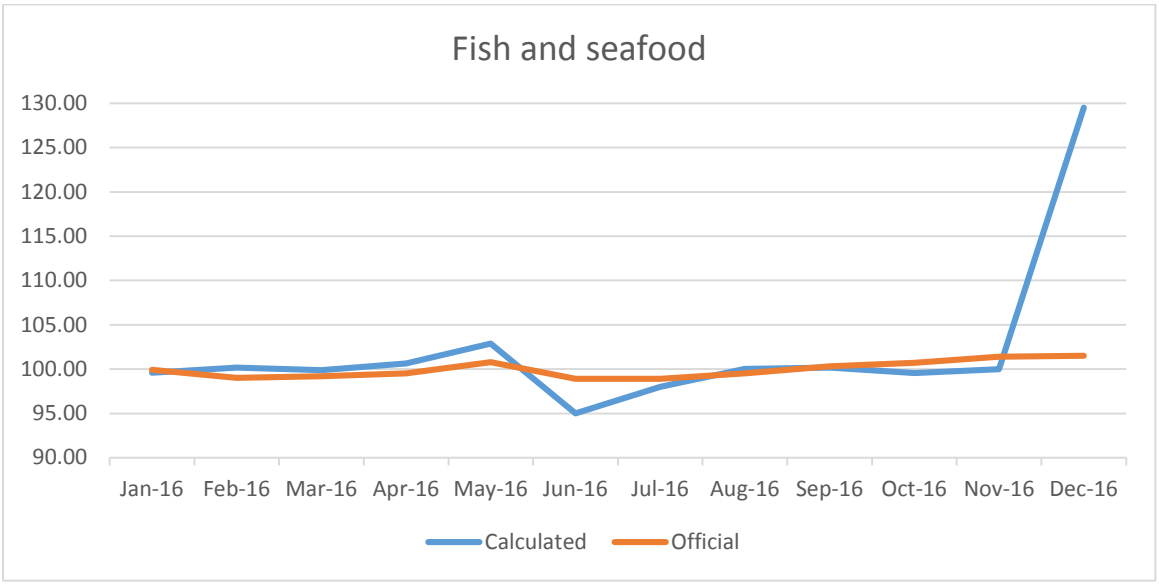
Appendix

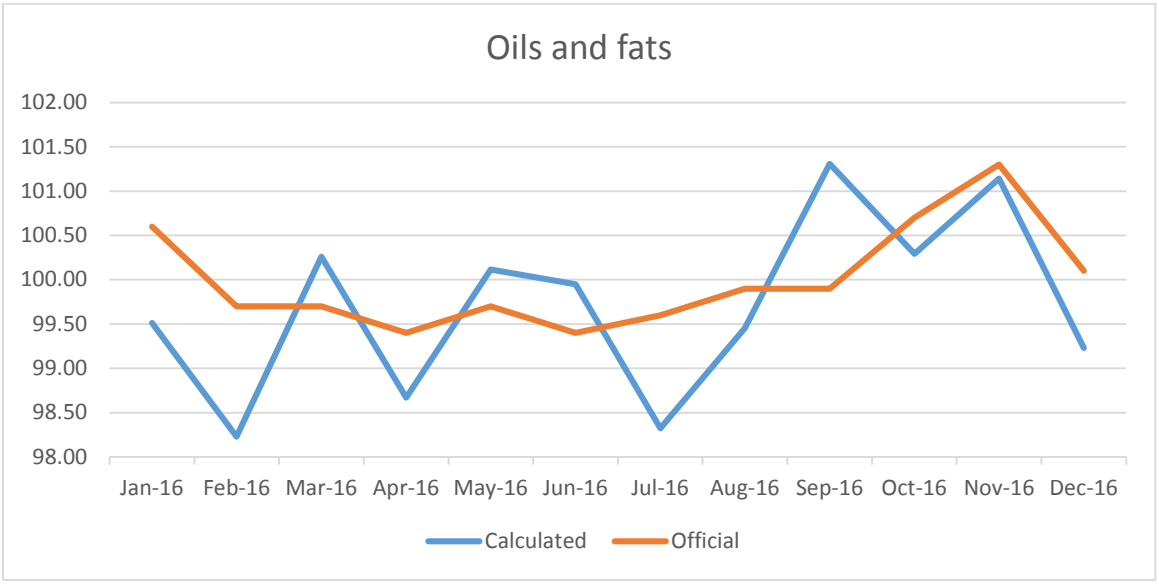
Appendix A

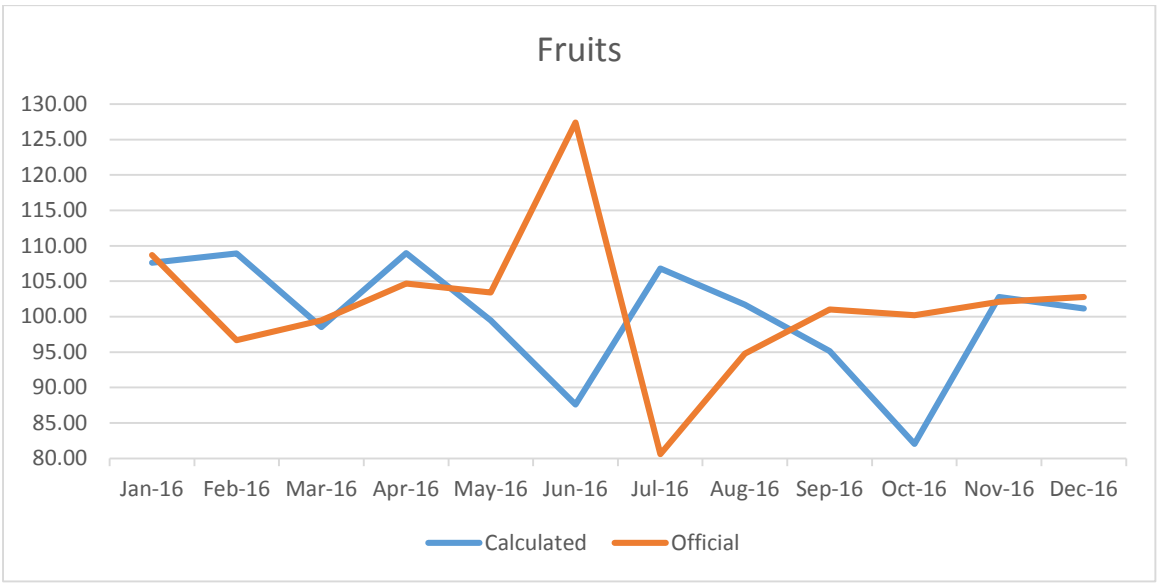


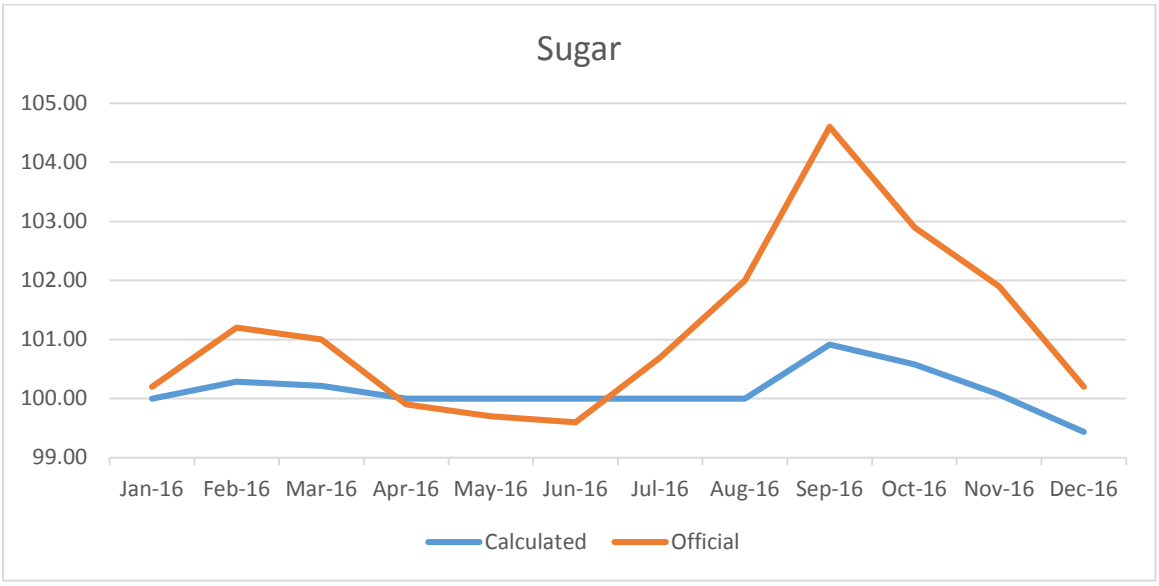


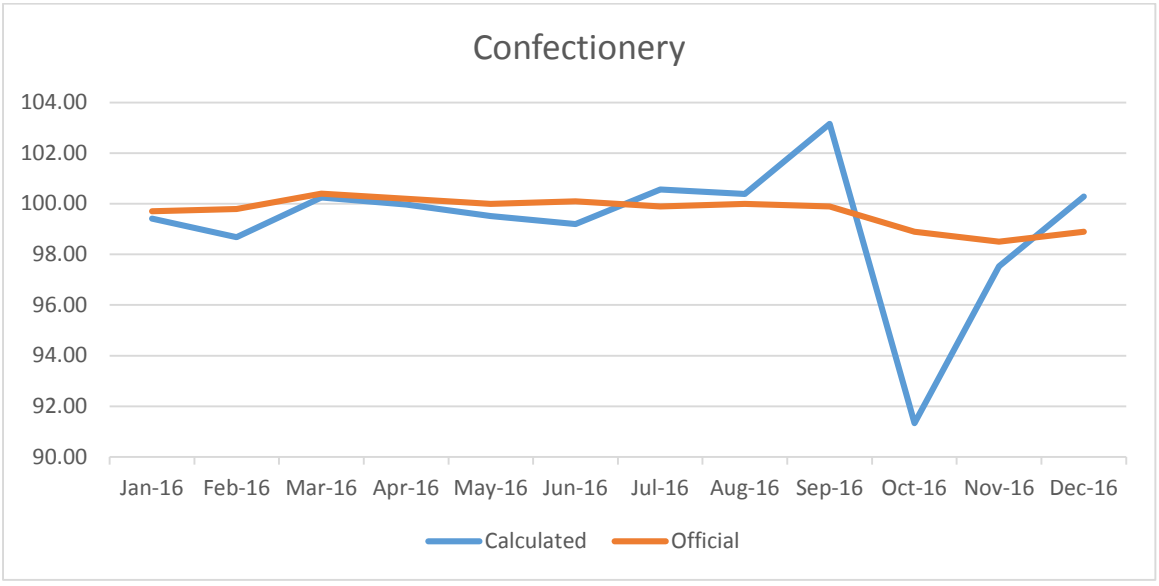

Meat
105.00
103.00
101.00
99.00
97.00
95.00
Jan-16
Feb-16
Mar-16
Apr-16
May-16
Jun-16
Jul-16
Aug-16
Sep-16
Oct-16
Nov-16
Dec-16
Calculated
Official


Fish and seafood
130.00
125.00
120.00
115.00
110.00
105.00
100.00
95.00
90.00
Jan-16
Feb-16
Mar-16
Apr-16
May-16
Jun-16
Jul-16
Aug-16
Sep-16
Oct-16
Nov-16
Dec-16
Calculated
Official


Oils and fats
102.00
101.50
101.00
100.50
100.00
99.50
99.00
98.50
98.00
Jan-16
Feb-16
Mar-16
Apr-16
May-16
Jun-16
Jul-16
Aug-16
Sep-16
Oct-16
Nov-16
Dec-16
Calculated
Official


Fruits
130.00
125.00
120.00
115.00
110.00
105.00
100.00
95.00
90.00
85.00
80.00
Jan-16 Feb-16 Mar-16 Apr-16 May-16 Jun-16 Jul-16 Aug-16 Sep-16 Oct-16 Nov-16 Dec-16
Calculated
Official


Sugar
105.00
104.00
103.00
102.00
101.00
100.00
99.00
Jan-16 Feb-16 Mar-16 Apr-16 May-16 Jun-16 Jul-16 Aug-16 Sep-16 Oct-16 Nov-16 Dec-16
Calculated
Official


Confectionery
104.00
102.00
100.00
98.00
96.00
94.00
92.00
90.00
Jan-16 Feb-16 Mar-16 Apr-16 May-16 Jun-16 Jul-16 Aug-16 Sep-16 Oct-16 Nov-16 Dec-16
Calculated
Official

Appendix B

Source	SS	df	MS	Number of obs	=	31,415
				F(32, 31382)	=	743.55
Model	5320.17087	32	166.25534	Prob > F	=	0.0000
Residual	7016.86535	31,382	.223595225	R-squared	=	0.4312
				Adj R-squared	=	0.4307
Total	12337.0362	31,414	.392724143	Root MSE	=	.47286

lnsqm	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
2.data	.0014942	.0053385	0.28	0.080	-.0089694	.0119578
district						
2	-.2996772	.0068163	-43.96	0.000	-.3130375	-.286317
3	-.5214722	.0142232	-36.66	0.000	-.5493503	-.4935942
4	-.6678232	.0127125	-52.53	0.000	-.6927402	-.6429061
5	-.6026666	.0138618	-43.48	0.000	-.6298363	-.5754969
6	-.6690586	.0203225	-32.92	0.000	-.7088915	-.6292258
7	-.6502159	.0159638	-40.73	0.000	-.6815055	-.6189262
8	-.4849368	.0138351	-35.05	0.000	-.5120541	-.4578196
9	-.6556698	.0136613	-47.99	0.000	-.6824465	-.628893
10	-.6757284	.0145046	-46.59	0.000	-.7041581	-.6472988
11	-1.139305	.0896023	-12.72	0.000	-1.314929	-.9636813
room	-.0247698	.0030124	-8.22	0.000	-.0306742	-.0188654
area	-.0000609	6.79e-07	-89.68	0.000	-.0000622	-.0000596
floor	-.0055479	.0010425	-5.32	0.000	-.0075912	-.0035046
totfloors	.0016496	.0011056	1.49	0.136	-.0005175	.0038167
ceilingheight	.2227033	.0206595	10.78	0.000	.1822098	.2631968
buildingtype						
buildingtype						
2	-.1238673	.0093867	-13.20	0.000	-.1422656	-.105469
3	.106891	.0104752	10.20	0.000	.0863593	.1274228
4	.0127545	.0112388	1.13	0.256	-.0092739	.0347829
condition						
2	-.1387202	.0068987	-20.11	0.000	-.152242	-.1251985
3	-.2217118	.0106944	-20.73	0.000	-.2426733	-.2007504
centralheating	.0007537	.0089093	0.08	0.933	-.0167089	.0182163
closetothebusstation	-.0246399	.0062108	-3.97	0.000	-.0368133	-.0124666
electricity	-.0213074	.0099196	-2.15	0.032	-.0407502	-.0018646
elevator	.0160751	.0071133	2.26	0.024	.0021326	.0300175
eurowindows	.0499275	.0070852	7.05	0.000	.0360403	.0638147
furniture	.0645036	.0066094	9.76	0.000	.0515489	.0774583
garage	.0636425	.0107786	5.90	0.000	.042516	.084769
gas	.0299009	.0090089	3.32	0.001	.0122431	.0475588
hotwater	-.0115161	.0075187	-1.53	0.126	-.026253	.0032207
irondoor	.0153218	.0065242	2.35	0.019	.0025341	.0281096
water247	.0404765	.0110352	3.67	0.000	.018847	.0621059
_cons	6.54568	.0626492	104.48	0.000	6.422885	6.668475