Thesis design

Lynette Joosten

March 8, 2017

Abstract

Since 2002 scanner data is used in the calculation in the Dutch CPI, but still the old methods and formulas are used. In this research the question "How can a simple method to calculate the CPI based on scanner data be created?" will be answered, by first conducting an extensive literature review and then using this information to construct a model. The impact of a new model would be to make the calculation of the CPI simpler and thus more efficient.

1 Personal details

My email `mailto:lynette.joosten@student.uva.nl`

My supervisor's email `mailto:e.kanoulas@uva.nl`

The wiki on my github account `https://github.com/LynetteJoosten/Thesis/wiki`

2 Research question

The Consumer Price Index (CPI) measures the price development of all consumption expenditure of consumers, and is the main measurement of inflation.[8] The calculation of the CPI was initially based on a fixed basket of goods. This basket of goods was collected through a telephone survey or visiting a shop to record the prices. In the 70s, barcode scanners were introduced at shops, which also meant that transaction data (from now on scanner data) were recorded electronically [4]. Scanner data of two supermarket chains has been used in the Dutch CPI since 2002, and since 2013 supermarket surveys are not carried out anymore in The Netherlands. The number of supermarket chains that supplied scanner data grew to ten supermarkets in 2015 [2]. Scanner data is not only provided by supermarket chains, but also by department stores, do-it-yourself stores, and many others [3]. There are multiple methods to use the scanner data in the calculation of the CPI [4]. These are based on the methods and formulas that were used in the time that the basket of goods was sampled in the supermarket itself. Since the first methods at statistics bureaus were devised, technologies have changed. The aim of this research is thus to see if it is possible to create a simpler model for the calculation of the CPI based on scanner data from supermarkets. A simpler model would lead to a more efficient calculation of the CPI. The model could then also be used as an international approach,

since a lot of statistics bureaus are not sure how to implement scanner data into their models [4].

The research question is:

How can we create a simple method to calculate the CPI based on scanner data?

To answer this research question, the following sub-questions have been defined:

- How is the Dutch CPI calculated at the moment?

- What are the attributes of the scanner data?

- What methods are currently used to calculate the CPI based on scanner data?

- What are the challenges of the current methods used to calculate the CPI?

3 Related Literature

Statistics Netherlands changed their method of calculating the CPI in 2010 to incorporate more scanner data. Van der Grient and De Haan described this thoroughly in their article.[7] This article will be the basis for answering the first sub-question, namely how the CPI is calculated at the moment.

Multiple papers have been written about the current methods used to calculate the CPI based on scanner data. One such paper is another paper by Van der Grient and de Haan in which they compare the proposed Dutch method (from 2010) to a method proposed by Invancic, Diewert and Fox.[5] In 2015, Chessa proposed yet another approach which is more generic to include scanner data in the CPI calculation. He also compares this method to other known methods at the time.[2] In addition to the paper by Chessa, a paper by De Haan was also published in 2015, in which he proposed a framework to use for scanner data. Next to his proposed framework, he also describes challenges of using scanner data in the calculation of the CPI.[3]

Triplett describes more challenges of using scanner data in his article. For example, he states that aggregation by a supermarket chain over all stores might not be the best input data, since prices can differ for each store.[6]

An unpublished paper by De Haan, Willenborg and Chessa contains an overview of all price index methods for scanner data. It also describes the attributes of scanner data.[4]

The Billion Prices Project of MIT has multiple good resources, among others a paper in which they propose a method to create an online price index based on scraped data. Although this research is not about scraped data, their proposed method is simple and elegant, and might well be a good starting point.[1]

4 Methodology

4.1 Data

This research will focus on a data set provided by a big supermarket chain. Weekly Statistics Netherlands gets a new data set from Albert Heijn consisting of all products, their price and the amount of products sold.

4.2 Methods

An extensive literature review will be done to answer the four subquestions. This background information will then be used to create a simple method to calculate the CPI based on scanner data.

4.3 Evaluation

The results will be evaluated in multiple ways:

- Can the new consumer price index show the same differences over time as the old consumer price index?

- Is the absolute number of the new consumer price index the same or relatively close to the old consumer price index?

- What is the difference in run time between the traditional CPI model and the newly created CPI model?

5 Risk assessment

The biggest risk of this research is that it might not be possible at all to create a simple model for the calculation of the CPI. However, this would also be a valid research result.

Another risk is the low amount of literature on creating a simple method for the calculation of the CPI. A solution would be to interview some of the experts of the CPI at Statistics Netherlands, who might be able to give insight in how to create such a model.

6 Project plan

Week	Deliverables
1	Literature research, answer subquestion 1
2	Literature research, answer subquestion 2
3	Literature research, answer subquestion 3 and 4
4	Start with creating a model
5	Create the model
6	Mid-term results
7	Create the model
8	Test and optimise the model
9	Evaluation of results
10	Conclusion and discussion
11	Hand in thesis
12	Defence

References

[1] Alberto Cavallo. "Online and official price indexes: Measuring Argentina's inflation". In: *Journal of Monetary Economics* 60.2 (2013), pp. 152–165.

[2] A. Chessa. "Towards a generic price index method for scanner data in the Dutch CPI". In: (May 2015).

[3] J. De Haan. *A framework for large scale use of scanner data in the Dutch CPI*. Ottawa Group, 2015.

[4] J. De Haan, L. Willenborg, and A. Chessa. "An Overview of Price Index Methods for Scanner Data". unpublished paper. 2016.

[5] Jan de Haan and Heymerik A Van der Grient. "Eliminating chain drift in price indexes based on scanner data". In: *Journal of Econometrics* 161.1 (2011), pp. 36–46.

[6] Jack E Triplett. "Using Scanner Data in Consumer Price Indexes. Some Neglected Conceptual Considerations". In: *Scanner data and price indexes*. University of Chicago Press, 2003, pp. 151–162.

[7] H.A. Van der Grient and J. De Haan. "The use of supermarket scanner data in the Dutch CPI". In: *Joint ECE/ILO Workshop on Scanner Data*. Vol. 10. 2010.

[8] J. Walschots. "Het mandje van de consumentenprijsindex, de bestedingen van consumenten en de meting van de CPI in 2016". In: (2016).