Exploring the use of scanner data in the Norwegian CPI for products with high churn

**Keywords:** scanner data, CPI, product replacement, homogenous products, multilateral index methods.

# Introduction

Statistics Norway has a clear strategy of making increasingly use of new data sources in official statistics. The national statistical institute (NSI) has a long history of using scanner data in price statistics and in the Consumer Price Index (CPI) in particular. With scanner data we mean aggregated transaction data that provides information on turnover and quantity sold by certain article codes or barcodes. By increasing the use of scanner data in the price index the use of more traditional data sources like web questionnaires filled out manually by retailers can be reduced equivalently, and by that lowering the response burden and increasing index quality.

The calculation method mostly applied in the Norwegian CPI on scanner data is a matched model approach at article code level aggregated by a monthly chained unweighted geometric mean index (Jevons index) – referred to as the “dynamic method” according to Eurostat’s practical guide on scanner data. The method works well on relatively stable article codes like for most supermarket data, but is not appropriate for products with high churn i.e. more frequent changes in article codes, like clothing and consumer electronics for instance. The aim of the present scanner data development work in Statistics Norway is to implement a more generic method that is able to handle frequent changes in article codes. Secondly, a new calculation method should also preferably use both prices and quantities without causing index bias.

New calculation methods like multilateral index methods presented internationally during the last couple of years, do not by themselves solve the problem with product replacements. This issue must therefore be addressed separately. One crucial step is product definition. Defining the product at article code level may in many cases be too detailed especially for products with high churn as the match between new and old article code is lacking. A more appropriate approach could be to apply a broader definition of a product, for instance to combine different article codes of similar attributes of which the consumers are indifferent to. A practical solution is to create these homogenous products (HPs) by clustering together homogenous article codes and calculate a unit value. By calculating a unit value across homogenous article codes we allow for comparisons of new codes entering and old ones disappearing from the market. This paper presents work done in an ongoing grant agreement project and presents challenges related to HP definition and formation as well as effects on different multilateral index formulas.

# Methods

The results in this paper are mostly based on scanner data from clothing and sport equipment retailers. To facilitate choice of calculation method for future implementation, a total error framework has been developed that describes different diagnostics for that purpose. The aim of the framework is to provide a set of criteria for evaluating the empirical effects from the different price indices. The framework focuses on three necessary choices for unit value indices which are 1) direct vs indirect index, 2) bilateral vs multilateral reference universe and 3) formation of HPs. To motivate a practical choice concerning formation of HPs one needs to evaluate the effect of misclassification against missing replacement of unmatched products. These three choices are not new, but the relevant considerations differ when the product universe is dynamic and not fixed. Calculations show that the third choice, formation of HPs, may be the most important choice and therefore we focus mostly on that issue. In this paper we will look at HP misclassification and missing replacements as well as testing the sensitivity of the HPs. The idea is that with the framework diagnostics we get closer to a decision on index formula.

# Results

Based on data from a major sport equipment retailer we have calculated different indices, in figure 1 illustrated by the QU (quality adjusted unit value) indices (A. Chessa 2016), both at article code (blue line) and HP level (red line). In figure 1 we see the impact of using different product definition for a commodity group like socks for men. Calculations show that for this commodity group in particular, the price development in the series are quite similar indicating a certain stability in the product universe. For other commodity groups we see larger differences as in figure 2 with jackets for men. For most series tested we see a clear downward bias in the article code matched indices compared to HP matched indices. The effect of missing replacements is systematic since article codes leaving the market often have a downward trend, while new similar article codes often are introduced at a higher price.

Figure 1: Price development for socks, men. QU indices. 2016-2017



Figure 2: Price development for jackets, men. QU indices. 2016-2017



In this paper we will also look at the effects of using different index formulas both bilateral and multilateral price indices based on HPs. In figure 3 we show the price development for the commodity group sweaters/blouses for women. The development differs among the indices, but the effects seem to be smaller than the effects related to product definition and HP formation.

Figure 3: Different price indices of sweaters/blouses, women. 2016-2017



# Conclusions

Statistics Norway has an ongoing grant agreement with Eurostat for the period 2018 to 2020. The overall goal of the grant agreement project is to implement a more generic method on the use of scanner data compared to current practices and that works across different commodity groups. Preliminary test results achieved in light of a framework are important steps on the way to reach conclusion on method. It will also be important to keep up to date on international research and development work in other NSIs. Theoretical and empirical conclusions must also be balanced with legal requirements for the price indices through HICP regulations. Our preliminary results show that the HP formation is one of the most important issues to solve and the choice of HP formation seems to be of greater importance than the choice of index formula itself when it comes to products with high churn.

# References

1. A. Chessa, “A new methodology for processing scanner data in the Dutch CPI“. EURONA, 2016, <https://ec.europa.eu/eurostat/cros/system/files/euronaissue1-2016-art2.pdf>
2. L.C.Zhang, I. Johansen, R. Nygaard, “Testing unit value data price indices”. Ottawa group meeting, Germany, May 2017.