

Towards More Efficient Use of Transaction Data in the Swedish CPI

Peter Nilsson¹, Olivia Ståhl¹

Starting in 2012 with data on daily necessities, Statistics Sweden has over the last couple of years implemented transaction data for several different product groups in its monthly production of the Consumer Price Index, as well as in the HICP. Scanner (transaction) data now constitutes more than 20% av the Swedish CPI basket.

In the beginning of this development, it was sometimes the experience that each new transaction data source was associated with a whole new set of practical and conceptual implementation issues. However, many of the problems encountered have later proven to be recurrent and similar between subgroups. Up until now, however, transaction data in the Swedish CPI has not always been treated in a fully consistent way. As Statistics Sweden is soon about to take the next big steps in the scanner data area, implementing transaction data for services and clothing, we believe that now is a good time to intensify the work of reviewing methods, with the aim to make sure that principles are fully consistent between subgroups and that the best possible use is made of the data. This implies reviewing several different methodological issues, including index formulation, sampling methods and the editing process.

Regarding the question of appropriate index formulation, the aim must be a methodology which makes best possible use of the data; this includes taking the behaviour of consumers into account, perhaps (but not necessarily) by making use of monthly quantity information and/or by allowing for a dynamic basket. Currently, Statistics Sweden uses the so called “static method” for the computation of indices based on transaction data, either using a random sample of items or the full population. The static method was selected from start mainly based on a precautionary principle. Over the years, however, it has become increasingly clear that this approach is not fully optimal. Internationally, a lot of research has been conducted on this topic in recent years, and as a result the body of knowledge in the community as a whole has grown significantly. We therefore believe that it is possible, today, to make an

¹ Statistics Sweden. Contact information: peter.nilsson@scb.se, olivia.stahl@scb.se.

informed decision on the index formula issue, something which was not, in our opinion, possible in 2012.

When it comes to the discussion of sampling methods, the review should include an analysis contrasting sampling variance with measurement errors. The aim is to arrive at an optimal sample size, perhaps (but not necessarily) consisting of the whole population. Of course, the severity of the measurement error problem is also strongly related to the efficiency of the editing process in place.

In our presentation we will outline the preliminary schedule for our planned work within the transaction data area for the next coming years, including a list of what we feel should be the most prioritised methodological studies. These include, but are not limited to, (i) a review of the current definitions of homogeneous products in different product groups, i.e. of the level at which unit values and quantities are computed (this is to make sure that the same underlying principles apply for all sub-indices and, in particular, for goods and services); (ii) a literature review of empirical evidence produced internationally during the last couple of years when it comes to choice of index number formula for consumer price indices based on scanner data; and (iii) empirical studies of selected index number formulas and strategies, building on the knowledge acquired in phase (ii) and using the product definitions decided on in phase (i). Based on the results from the empirical studies, the final aim will be to be able to decide on a set of principles for the treatment of transaction data in the Swedish CPI. These principles should be clearly stated and in line with international recommendations, as well as approved by the national board of experts tied to the Consumer Price Index in Sweden. It is important to note, however, that the principles need not specify e.g. a specific index number formula or sampling design, but rather describe the main criteria's that should underlie each decision along the way. They will function as a road map for the next coming years, making sure that consistency between different sub-indices is upheld during this period of rapid development, and that important details are not missed out. Further details about our plans, including preliminary time schedules, will be given at the conference.