Jacek Białek[i], Anna Bobel[ii]

Comparison of Price Index Methods for CPI Measurement using Scanner Data

Abstract

Scanner data are a quite new data source for statistical agencies and the availability of electronic sales data for the calculation of the Consumer Price Index (CPI) has increased over the past 15 years. Scanner data can be obtained from a wide variety of retailers (supermarkets, home electronics, internet shops, etc.) and provide information at the level of the barcode, i.e. the Global Trade Item Number (GTIN, formerly known as EAN code). One of advantages of using scanner data is the fact that they contain complete transaction information, i.e. prices and quantities for every sold item. It makes that we may use expenditure shares of items as weights for calculating price indices at the lowest (elementary) level of data aggregation. One of new challenges connected with scanner data is the choice of the index formula, which should be able to reduce the chain drift bias and the substitution bias. In this paper we compare several price index methods for CPI calculations based on scanner data. In particular we consider bilateral index methods with chained versions of direct weighted and unweighted indices, and also selected multilateral index methods, i.e. the quality adjusted unit value method (QU method) and its special case (the Geary-Khamis method), the so called "real time index", the GEKS method and the CCDI method. We verify the impact of window updating methods and also different weighting schemes in the quantity weights on the price index, i.e. we consider alternatively the QU-TS method and the QU-EW method. We compare all these methods using artificial data set and the real scanner data sets obtained from *allegro.pl*.

[i] University of Lodz, jacek.bialek@uni.lodz.pl; Central Statistical Office in Poland;

[ii] Statistics Poland, a.bobel@stat.gov.pl;