

Jacek Białekⁱ,

Scanner data – advantages, challenges, and their processing in the *PriceIndices* package

Abstract

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) or its European version: European Article Number (EAN). One of advantages of using scanner data in the Consumer Price Index (CPI) measurement is the fact that they contain complete transaction information, i.e. prices and quantities for every sold item. 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. The main purpose of the presentation is to discuss a broad spectrum of benefits and challenges related to the use of scanner data in measuring inflation.

An additional purpose of the work is to present the utility of the *PriceIndices* R-package in the field of analysing the dynamics of scanner prices. The presentation of this R package is divided into the following areas: scanner data preparing, data set characteristics, bilateral index calculations, multilateral index calculations, extensions of multilateral indices, aggregation of index results, and comparison of price indices. In particular, to demonstrate the package, the entire data processing (from the preparation of the row scanner data to the calculation of the price indices) and some empirical study will be performed.

Key words: scanner data, scanner data classification, product matching, price indices, multilateral indices, *PriceIndices* package

ⁱ University of Lodz, Department of Statistical Methods, jacek.bialek@uni.lodz.pl;
Statistics Poland, Department of Trade and Services, J.Bialek@stat.gov.pl;