

MARS: A method for defining products and linking barcodes of item relaunches

Antonio G. Chessa¹

Abstract

An essential choice that has to be made prior to index calculation is how to define the products within each aggregate. Transaction data sets specify sales data by GTIN (barcode), which represents the most detailed level of product differentiation. However, items may leave and re-enter stores with different GTINs and typically at higher prices than the exiting GTINs. Old and new GTINs of these so-called item “relaunches” have to be linked in order to capture these price increases. This is an example where GTINs are unsuitable as products.

The processing of large electronic data sets calls for methods to define products and link barcodes in a highly automated way. This study presents a method that defines products (i.e., groups of one or more GTINs) by balancing two measures: one measure quantifies the “homogeneity” of GTINs within products, while the second measure expresses the degree to which products can be “matched” each month with respect to a fixed comparison period. The two measures have opposite effects in a nested stratification scheme: tighter defined products are more homogeneous but have the same or worse product match over time compared to broader defined products.

The method MARS, which stands for (product) “match adjusted R squared”, has been applied to a broad range of product types. Different GTIN stratification schemes were evaluated and ranked with MARS. GTINs are suited as products for food and beverages, but not for product types with higher rates of churn, such as clothing, pharmacy products and electronics. In cases with high churn and relaunches, products are defined as combinations of characteristics, such that GTINs with the same characteristics are grouped into the same product. The mechanism behind MARS will be illustrated by various examples. The results will also focus on the impact of different stratification levels on a price index.

MARS searches for the optimal degree of GTIN stratification and relaunches are detected automatically. The method therefore facilitates the processing of complete transaction data sets and increases the efficiency of monthly CPI production tasks. Besides its primary objective of defining products, MARS is also useful as a data monitoring tool, as it is able to detect inconsistencies in item descriptions and classifications over time. Future applications of MARS will cover more service-oriented product types like package holidays and web scraped data. Meanwhile, Statistics Netherlands will start implementing MARS in its CPI in the near future.

Keywords: CPI, transaction data, GTIN, relaunch, product homogeneity, stratification.

¹ Division of Economic and Business Statistics and National Accounts, CPI department, Statistics Netherlands, email: ag.chessa@cbs.nl