

Evaluating unit-value price indices in a dynamic item universe *OTTAWA GROUP 2019*

Li-Chun Zhang¹²³, Ingvild Johansen³ and Ragnhild Nygaard³

1. University of Southampton, 2. University of Oslo, 3. Statistics Norway

Statistics Norway has a long history of using scanner data in price statistics and in the Consumer Price Index (CPI) in particular. The early research – in Norway as well as internationally -- was focused on supermarket data which consists largely of stable products. The attention has since gradually shifted towards the parts of consumption market that are characterized by high product churns, where the methodology initially introduced for supermarket data is no longer adequate. Several NSIs including Statistics Norway have been researching on a more generic scanner data methodology. The overall goal is to implement an approach that works well across different commodity groups including those with high product churns. A variety of methods and index formulas are currently being tested and implemented for CPIs in different parts of the world. At the same time, one cannot but notice a lack of standard regarding the process, by which the choice of index formula can be plausibly established in a given situation. We propose therefore a systematic approach to the investigation process, which has recently been developed at Statistics Norway. This consists mainly of two parts: a *Total Effect Framework (TEF)*, and a set of *generic diagnostics*.

The TEF is defined by the necessary choices required and the elements that affect these choices. We decompose the choice of an index formula into three parts: 1) index base -- direct vs. indirect index, 2) reference universe - bilateral vs multilateral index, and 3) formation of homogenous products (HPs) - instead of applying the index formulae directly to the observed items. We distinguish between three types of dynamic elements in a changing item universe: a) replacement items – relaunches and updates, b) regeneration items – new and outgoing items, and c) strongly seasonal items that are absent from time to time. The TEF is needed due to the unavoidable inter-dependence between the various choices and elements: e.g. the effects of a choice of reference universe can vary for different dynamic items, as well as it depends on the other necessary choices regarding the HPs and the index base.

Under the TEF, the different alternatives of a choice can be studied analytically wherever possible. Yet the complexity involved is often such that clear-cut conclusions cannot be reached *a priori* independent of the actual data. We therefore review, synthesize and develop a set of generic diagnostics, pertaining to the formation and misclassification of HPs, the sensitivity of a particular choice of index base or reference universe, etc. Most indices employed in such diagnostics are not genuine candidates for real production, but they are designed and introduced to generate useful empirical evidences, on which a plausible final choice of index formula can be based. We shall illustrate the generic diagnostics using scanner datasets mainly from the markets of sport equipment which have high product churns. We hope that over time the set of generic diagnostics will be expanded and refined, based on the joint efforts of the whole research community, such that they can form a standard toolset in practice.

Finally, we recognize that the choice of index formula depends not only on theoretical and empirical conclusions, but also the relevant legal requirements and international standards for price indices e.g. in terms of the HICP regulations must be taken into consideration. In addition to one's own results and

conclusions, it is important to align with and stay updated on international research carried out by others. To this end we shall summarize our own experiences and put forward some preliminary conclusions of a generic scanner data price index methodology.