10 SCANNER DATA

1. Introduction

The environment in which statistical agencies operate is changing. New opportunities to access and interrogate big data are becoming available, increasing the potential to provide new insights into matters of importance. The statistical landscape is becoming more complex, expectations of decision makers are growing, and National Statistical Offices (NSOs) are being challenged to deliver the best possible statistical program in more efficient and innovative ways.

The launch of barcode scanner technology during the 1970s, and its growth in the 20th century, has enabled retailers to capture detailed information on transactions at the point of sale. Scanner data is high in volume and contains information about individual transactions or summaries, date, quantities and values of products sold, and product descriptions. As such it is a rich data source to NSOs that can potentially be used to enhance their statistics, reduce provider burden, and reduce associated costs of physically collecting data.

This Annex discusses the opportunities and challenges presented when utilizing scanner data to compile the CPI and aims to provide guidance to NSOs. Section 2 outlines a number of practical considerations regarding the acquisition of scanner data sets, the assessment and preparation of the data, and implementation issues. New methods that have been developed to construct price indexes from scanner data sets, so-called multilateral methods, are presented in section 3. Section 4 discusses the assessment of the new methods and the empirical results, communication with users and stakeholders, and publication and dissemination of the ‘new’ price indexes.

2. Practical considerations

2.1 Introduction

The availability of scanner data provides a number of opportunities to improve the accuracy of the CPI. Scanner data sets typically contain complete coverage of items sold by a retailer at all their locations; as well as item quantities sold and revenue received by the retailer for these items. This information has the potential to: improve the accuracy of the prices used to compile the CPI by calculating unit values for homogenous products; improve the samples of items priced, with the potential to utilize a census of items sold to compile the CPI; and use quantity/revenue information to weight items according to their economic importance.

While scanner data sets present opportunities to improve the accuracy of the CPI, there are also challenges that need to be overcome before NSOs can utilize scanner data to compile the CPI.

This section describes some of the most important opportunities and challenges presented by scanner data sets; outlines practical considerations; and provides insight and advice to enable the use of scanner data to compile the CPI.

2.2 Obtaining scanner data sets

Scanner data have existed for several decades and their value in the compilation of official statistics has become evident over time. One challenge faced by NSOs is obtaining the scanner data sets. Two main options are available. NSOs may seek the supply of scanner data sets directly from retail businesses or from third party data providers. Both options present benefits and challenges.

A number of NSOs have successfully negotiated the supply of scanner data direct from retail businesses and used these data in the compilation of their CPI.[1] Direct collection of data sets from retail business by NSOs has a number of potential benefits. These include the ability to negotiate:

- the supply of the data set at no (or minimal) cost;
- the scope of items included in the data set;
- the level of item aggregation to ensure homogenous information;
- an agreed timetable for the supply of the data set to meet CPI processing requirements; and
- a contact officer within the retail business who is familiar with the data set to answer NSO data queries.

[1] Australia, the Netherlands, New Zealand, Sweden and Switzerland are some of the NSOs who have secured scanner data sets for use in compiling the CPI. A complete list of NSOs utilizing scanner data, along with the methods implemented, can be found in Attachment A of this Annex.

Negotiating the supply of scanner data sets directly with retail businesses presents challenges as well. The primary challenge is that the bilateral negotiation of scanner data sets with retail businesses is resource intensive. Experiences in the Netherlands, Switzerland and Australia suggest these negotiations will take approximately six months to complete. The negotiations relate to a wide range of topics: from IT systems and formats to confidentiality concerns. Any agreement reached between the NSO and retail business is typically formalised in a Memorandum of Understanding (or similar).[2]

An alternate approach to obtaining scanner data sets directly from retail businesses is to source these data sets from intermediaries or market research companies. Market research companies like Nielsen and GfK possess scanner data sets that have been obtained by some NSOs for CPI assessment and compilation purposes (Krsinich, 2014). The primary benefit of this approach is the ability to negotiate the supply of multiple data sets relating to a diverse set of products with a single or small number of data providers.

Obtaining scanner data sets from market research companies (or similar) does present some challenges. These data sets are generally purchased by the NSO. The costs may be offset by the reduced data collection costs as NSO staff is no longer required to personally visit retail businesses where scanner data sets have been secured.

The experience of NSOs using scanner data sets to compile their CPI suggests that obtaining data sets directly from retail businesses is preferred for the reasons outlined above. However, obtaining scanner data sets from market research companies is beneficial in contexts where securing data sets directly from retail businesses is not possible or resources are not available to negotiate bilateral data supply agreements.

2.3 Assessing and preparing scanner data for use

If the NSO is successful in securing scanner data sets, the challenge for NSOs is to then turn these data sets into information that can be effectively and efficiently used to compile the CPI. To achieve these objectives requires the NSO to overcome a number of challenges.

2.3.1 Developing an Information Technology (IT) system

Scanner data are by their very nature *big* data. The NSO requires an IT/computing system that can acquire, store and process the large scanner data sets if the information is to be used to compile the CPI. This IT system needs to be able to acquire and process data sets that have different classification structures, formats and contents. This is because retail businesses (and third party data providers) normally develop bespoke systems for their own internal reporting purposes. This can present challenges for the NSO as IT system development requires resources, including time and money. A number of NSOs have documented the challenges

[2] A Memorandum of Understanding (or similar) documents the roles and obligations of each party and aims to ensure an on-going supply of scanner data to the NSO to the agreed timetable.

presented by the need to develop an IT system, as well as their various approaches to address these challenges (Bird et al., 2014; Böttcher and Sergeev, 2014). The solution is heavily dependent on the local context.

What is clear is that NSO resources will be required for an IT system if the NSO is to utilize scanner data sets to compile the CPI, irrespective of the data provider.

2.3.2 Classifying scanner data

Scanner data sets generally possess product classifications that are unique to the individual retailer. The NSO will most likely receive data sets that contain different product classifications which need to be mapped to a single CPI classification. The classification of scanner data sets is likely to require significant NSO resources. The largest investment of resources is needed when the data sets are first received by the NSO; however there is a need to provide ongoing classification resources as new products enter the data set.

The challenge of classifying scanner data items to the CPI classification has been addressed by NSOs in various ways. All NSOs have attempted to find a solution in their local context. The Swiss NSO has classified scanner data items to the CPI classification by purchasing market research metadata (Muller, 2010). The Dutch NSO has employed a combination of undertaking classification of scanner data items themselves and utilizing market research information to aid the mapping process (de Haan et al., 2010). Some NSOs have, for various reasons, undertaken the entire classification of scanner data items to their CPI classification within the NSO (Howard et al, 2015).

The challenge of classifying scanner data items to a single CPI classification primarily arises when scanner data sets have been secured directly from retail businesses. Obtaining scanner data sets from market research companies may enable the NSO to negotiate the supply of scanner data that has already been classified to the NSO's CPI classification. This is viewed by some NSOs as a particular advantage to obtaining scanner data from market research companies.

2.3.3 Quality assurance of the scanner data sets

Scanner data sets are a new data source to compile the CPI. As is the case with any change in data source, the compilers of statistical series should undertake a range of 'checks' to ensure the new data source provides the foundation from which to produce fit-for-purpose statistics. These scanner data checks can be classified as either *global* checks or *detailed* checks.

Global checks relate to broad quality measures that are generally applied at the time the NSO receives the data set. These checks aim to ensure the data set is broadly consistent with data sets received by the NSO from the same data provider in previous periods. The checks may relate to the format of the data set; the total number of products within the data set; and the total revenue by outlet. These global checks should high-light significant errors with the data set.

Detailed checks are generally applied at the product or product group level. These checks aim to high-light significant changes in the quantities sold, revenue and the prices of the products within the data set. These detailed checks have traditionally been referred to as micro-editing of price data.

Both the *global* and *detailed* checks should be automated and reports prepared for analysis by NSO staff. The checks may require contact with the data provider, as well as confrontation with alternate price information sources (e.g. flyers and online prices).

2.3.4 The level of product data to support CPI compilation

Scanner data sets should ideally provide data for homogenous products. This is important to ensure the CPI reflects pure price change over time. Changes in the composition of products sold, and their qualities, should not be reflected in the CPI. NSOs should focus closely on this topic as part of the negotiation to secure scanner data sets from data providers, as well as the assessment of the quality of the scanner data sets.

This topic is discussed in more detail in subsection 2.4.3 and section 3 below.

2.4 Implementation – from confrontation to new methods

2.4.1 The benefits and challenges of utilizing scanner data sets

The use of information contained in scanner data sets to compile the CPI can represent quite a significant change to the data collection practices and the price index methods traditionally employed by NSOs. This suggests that these changes need to be carefully managed, both in terms of the statistical impacts as well as communication with users and key stakeholders. A detailed description of communication strategies associated with scanner data implementation is given in subsection 4.2.

This sub-section outlines scanner data collection practices and index number methods that progressively make greater use of the information contained in scanner data sets.

Scanner data potentially enable the accuracy of the CPI to be enhanced in a number of ways and at lower cost. The scanner data sets can be used to: (i) undertake data confrontation; (ii) replace field collected prices; (iii) expand pricing samples; (iv) weight products at the lowest levels of the CPI to reflect their economic importance; and (v) implement new methods[3] that possess desirable price index properties and enable process automation.

The enhancements listed above each improve the accuracy of the CPI. This will be explained below. A number of NSOs have utilized scanner data to achieve some of these enhancements, particularly (i), (ii) and (iii). While these enhancements are significant, implementing (iv) and (v) will maximize the use of scanner data to enhance the quality of the CPI. Of note, some NSOs have gradually implemented each of the enhancements listed above (ABS, 2017) while

[3] See Section 3 of this Annex for a detailed description of these methods.

other NSOs moved directly from enhancement (i) to (v) (Krsinich, 2015; Chessa, 2016). Both approaches are valid and often reflect the local context in which the NSO operates.

The next five subsections describe the benefits of utilizing scanner data sets for each of the enhancements listed above.

2.4.2 Using scanner data sets for data confrontation and quality assurance

The availability of scanner data provides NSOs with the opportunity to confront or quality assure the data currently used to construct the CPI.

Scanner data sets contain product quantities sold and revenue received by the retailer for these products for some period of time, usually a week or a month. This information enables NSOs to calculate a price for an individual product by dividing a product's revenue by the quantity sold. This price is referred to as a *unit value* and represents the average price experienced by consumers over a period of weeks or months.

For a homogeneous product, the unit value more accurately reflects prices paid by consumers over the whole period than point-in-time pricing (Balk, 1998).[4] *Unit values* contain discounts and the effects of these discounts on the quantity of products sold. The period for which unit values are calculated is important in terms of the accuracy of the unit value. Diewert, Fox and de Haan (2016) find that unit value prices used for constructing the CPI should be for the same period as the index to be constructed, rather than for a sub-period. The latter approach can lead to an upward bias in the CPI.

Price analysts are able to compare the prices collected in the field to those calculated from the scanner data sets. This analysis provides insight into any biases introduced to the CPI from point-in-time pricing compared with unit values. An analysis of the product revenue and quantities sold can be used by the NSO price analysts to high-light where CPI samples could be improved.

2.4.3 Using scanner data sets to replace field collected prices

In most countries the majority of prices used to compile the CPI are collected by personal visits to selected retail businesses. These personal visits are made by NSO field officers who observe point-in-time prices as well as discuss discounts, special offers and volume-selling items with the business operators. The NSO field officers record this information during the visit, often in handheld computers. The regular personal visits to businesses enable the NSO field officers to actively monitor market developments and observe product quality change.

Utilizing scanner data sets to replace field collected prices generally results in NSO resource savings. This is because NSO field officers are no longer required to visit businesses where prices were collected. The potential for NSO resource savings is influenced by the size of the

[4] Revenue data may not align perfectly with the purpose and concept of the national CPI because it may include expenditure by non-resident households and businesses (Fenwick, 2014).

field officer reductions and the increase in resources required by the NSO to manage and process scanner data sets.

Replacing field collected prices also presents challenges that need to be managed.

Unit values should relate to a single homogenous product whose specifications should remain constant over time because changes in the composition of products sold and the quality of products should not be reflected in price changes (ILO, 2004, p. 164). These requirements present some challenges when replacing field collected prices with information from scanner data sets. Negotiation between the NSO and data provider is often needed to determine the appropriate level of product aggregation (or disaggregation) needed to support the production of unit values for use in the compilation of the CPI.

Several NSOs have experience in producing unit value data from scanner data sets. In some countries the use of Stock Keeping Unit (SKU) has proven to be successful (Howard et al., 2015), while the use of Global Trade Item Number (GTIN) and European Article Number (EAN) may not always work well (Dalen, 2017). GTIN, for example, may be too low a level of detail, differentiating products by product aspects, such as packaging, which are considered irrelevant to consumers. While these detailed data reflect homogenous products, the problem of *item churn* or relaunches often occurs and impedes the calculation of the CPI.[5]

An essential part of price measurement is accounting for quality change and the introduction of new items (ILO, 2004). This has been achieved by most NSOs when field officers visit retail businesses with the aim of measuring price change for identical or equivalent items in successive periods, and identifying new items. As the characteristics of products are altered, the NSO field officers collect descriptive information that enables the effects of a quality change to be separated from the price change, so that the CPI measures pure price change.

Accounting for quality change is particularly challenging when using scanner data sets. Scanner data sets tend to exhibit a high level of churn in the products available from month to month. There are new models (and versions of models) of products becoming available in the market and old models dropping out of the market as they become obsolete. Calculating quality adjusted prices is therefore difficult.

There are broadly three scenarios where there is a need to quality adjust prices obtained from scanner data sets. They are:

a) where new items are brought into the price samples, including as replacements;
b) where there has been a quantity change (e.g. change in packet size) and the product identifier has changed; and
c) where there has been a quantity change and the product identifier has not changed.

The first scenario is the simplest case and requires calculating a previous period price for the new item.

[5] For example, when using barcode as item identifier, the price change of a homogenous product whose barcode changes at the same time will not be measured.

In the second and third scenario, a quality adjustment factor is calculated to account for the quantity change. The NSO will need to develop a method to link new and disappearing products. For example, if a product changes in packet size, the linking process could use information on the product description, price, revenues, timing (when products appear or disappear on sales listings) and quantity sold. This process identifies, as far as possible, that the new product is likely to be an appropriate replacement for the disappearing product (but with a different product identifier). Quality adjustment is then actioned by price analysts based on the product description.

Scanner data sets are voluminous and can vary significantly in meta-data structure and formats from one retail business to the next. This can result in a significant amount of NSO resources being required to transform these raw datasets into a comprehensive database suitable for the analysis and creation of CPIs (Bird et al., 2014; Böttcher and Sergeev, 2014). Scanner data acquisition, storing, cleaning and coding are also resource intensive challenges that need to be considered by the NSO.

2.4.4 Using scanner data sets to update pricing samples

The collection of point-in-time prices by NSO field price collectors visiting retail businesses is resource intensive. A census of items cannot practically be priced each period resulting in the need for some sort of sampling approach. For example, sampled products may be selected for inclusion in the CPI basket by NSO field price collectors who discuss with the business operator which items are volume sellers, examine the shelf space of the products and make judgements about their relative importance. NSO field staff then aim to select a representative basket of items for pricing. This is a purposive sampling approach.

Purposive sampling has traditionally been used because sampling frames for items purchased were not available and detailed quantity or revenue data to measure the economic importance of the items was lacking. Purposive sampling can lead to biases when the selected items are not representative of the product population.

This traditional approach to sampling can be replaced by more scientific sampling methods due to the availability of scanner data. Since scanner data typically is a census of products, scanner data sets can be used as a sampling frame for updating pricing samples. A pricing sample is usually two-dimensional; it is a combination of a sample of outlets and a sample of items/product varieties. If all the stores from a retail chain are covered, the scanner data set can be used as sampling frame for both the outlet and item dimension; see also Chapter 5.

Revenue shares for each product (or product/outlet combination) can be used to determine the significance of each product within a product group. Products are then selected for inclusion in the CPI 'basket' based on revenue share either through sampling proportional to revenue or cut-off sampling (de Haan, Opperdoes and Schut, 1999).

Over time, however, products in the sample can lose relevance or even cease to exist. In these situations a replacement product is needed to maintain the relevance of the sample. Relevance tests can be used to highlight items in the samples that have become unsuitable and also highlight and rank suitable items as replacements.

The main principle behind these relevance tests is that the products should have a stable revenue share (i.e. consistent revenue share compared to other products) within the CPI product group. These product groups are referred to as the Elementary Aggregate or ‘EA’ in price index literature (Chapter 20 of ILO, 2004). The stable revenue share is particularly important, as items can have large sales when introduced into the market due to novelty or introductory sales prices, have insignificant revenue thereafter, and hence not be representative of the broader market.

To mitigate these problems, possible replacement products’ revenue must have been stable and significant for a specified period of time (e.g. three to six months) before they can be considered for inclusion into the price samples. CPI analysts can then manually check all items which are flagged for replacement and manually select items from a list ranked according to average monthly revenue share over the previous six months.

Many food and household items will have varieties of the same base item which have similar if not identical price evolution. A specific brand of canned tuna, for example, is available in many flavors and CPI compilers will be aware that prices for the different flavors from the same brand will behave similarly to one another, going on sale at the same time and changing price at the same time. Having a single flavor in our sample will hence represent the price movement for a much more significant portion of the market than that single flavor’s revenues would suggest.

The sampling process used to ensure product samples are representative is usually manually driven, requiring the CPI analysts to manually select a replacement from this ranked list of potential products that pass certain eligibility criteria. This scanner data sampling approach requires additional CPI analyst resources which, ideally, are offset by the reduced field collection resources.

2.4.5 Using scanner data to update index structures and to apply weights

Pricing samples have traditionally been small. When the additional CPI analyst resources are indeed offset by the reduced field collection resources, the NSO could decide to expand the pricing samples without changing the price index formula at the EA level or the sampling procedure.

It may be worthwhile, however, to reconsider the index structure and the sampling procedure, in particular when the NSO obtain scanner data directly from retail chains. Traditionally, an EA index is compiled from prices collected at outlets that belong to different retail chains (or independent stores). When the NSO wants to use much more price information from a retail chain than before, it seems preferable to treat EA-chain combinations as separate strata in the index compilation process.

In particular when the NSO decides to use the classification system provided by the retailer, it will be necessary to change the index structure: the lowest level in that classification should now be treated as a separate (chain-specific) EA. This raises several issues. The first issue is whether the stores belonging to the chain in question should be viewed as separate outlets, as is the traditional situation. In that case, unit values for the sampled items should be calculated at the store level. On the other hand, when the items are deemed homogeneous across stores, it may be useful to calculate unit values across all the stores belonging to the chain (Ivancic and Fox, 2013). Some NSOs do not have a choice, however, as they receive scanner data at the chain level.

The next issue is to what extent existing sampling procedures have to be changed. Suppose the NSO formerly used sampling of items proportional to revenue from the scanner data. This procedure can also be used to sample items from chain-specific EAs, where items are either defined (and unit values calculated) at the store level or the chain level. If the NSO wants to significantly increase the sample sizes in order to make use of a substantial part of the price information contained in the scanner data set, sampling procedures need to be reconsidered.

Another issue is how to integrate the chain-specific EA price indexes from scanner data with price information from other sources. Because these “EAs” are different from the EAs in the traditional index structure, the scanner data price indexes have to be aggregated up to a level – perhaps the lowest level of product aggregation the NSO publishes price indexes – where they can be combined with price indexes from other sources. In other words, two aggregation steps are required: aggregation of the chain-specific EA price indexes up to some higher level product category, and aggregation of the resulting scanner data indexes with price indexes at that level pertaining to other retail chains and independent stores.

The revenue data provides the opportunity for NSOs to weight price indexes more frequently using more timely data. This can be achieved in various ways, depending on the availability to the NSO of scanner data for multiple chains. It is suggested that the weights to combine the price indexes from scanner data be updated annually, using product revenue data from the previous 12 months. Combining the scanner data indexes with the price indexes compiled from other sources requires expenditure data for the latter indexes, which may be difficult to come by or estimate.

Without scanner data, detailed expenditure data by product (or product/outlet combination) are not available or available infrequently. The majority of NSOs therefore apply unweighted price index methods at the lowest levels of the CPI: the prices or price changes of the items sampled from an EA are combined without explicitly weighting the items according to their economic importance. In most cases the Jevons index formula is used by NSOs.

Scanner data sets contain revenue data at the most detailed level. These data can be used to sample items proportional to their revenue, as mentioned above, but that raises a few issues. The inclusion probabilities serve as implicit weights. That is, the EA price index will actually be an implicitly weighted index, and the inclusion probabilities should correspond with the target/population index aimed at (Balk, 2005). Moreover, the revenue distribution within a product category as observed in scanner data is often highly skewed. Consequently, sampling proportional to revenue is likely to select some high-revenue items with a probability of 1. Suppose the NSO tries to estimate a weighted geometric target index using the (unweighted) sample-based Jevons index. The small-revenue items will have an implicit weight of 1, but the high-revenue items will not be weighted at all, which is obviously not a good solution; the latter items should be explicitly weighted.

It seems preferable to reflect the items' economic importance *explicitly* via a weighted index number formula rather than implicitly via the inclusion probabilities in an unweighted index. Weighted methods for scanner data will be discussed in subsection 2.4.6 and in greater detail in section 3.

2.4.6 Using scanner data sets to implement new CPI compilation methods

The approaches outlined in subsections 2.4.2-2.4.5 enable the NSO to continue using sample-based methods to compile their CPI. Improvements to the accuracy of the CPI will be achieved because the prices (i.e. unit values) are more representative of those actually paid by consumers; the products sampled reflect volume sellers; and the weights used to produce aggregate measures of price change are based on more timely information and can be updated more frequently.

The major challenge faced by NSOs implementing these approaches relates to the increase in resources needed (Bird et al., 2014). Maintaining a sample-based approach, especially when the pricing samples are extended, requires significant manual intervention, primarily because *product turnover* can be large.[6]

Ideally, a NSO would use all the available information in scanner data sets rather than taking samples. Manually processing a census of products from scanner data sets is prohibitively expensive, however, and cannot be undertaken to meet the CPI production timeframes. That is, automating CPI compilation processes is required.

Also, when using a census of products, hence without sampling, a weighted index number formula should be used. Again, product turnover poses a significant problem. To maximize the number of matches in the data, chaining at high frequency will be needed. This, however, can lead to significant drift in the index. Multilateral price index number methods, which are drift-free by construction, are most suitable to handle a census of products from scanner data. These methods are discussed in section 3.

[6] When Statistics Netherlands first introduced scanner data for supermarkets into the CPI, a Lowe index was used (Schut et al., 2002). The idea was to mimic traditional methods and processes on a sample of about 10,000 items (barcodes) from each supermarket chain. This approach was very demanding in terms of manual selection of items as replacements of disappearing items and in terms of quality adjustments when deemed necessary.

3. Multilateral price index methods

3.1 Introduction

Scanner data can be implemented in the CPI using traditional sample-based methods: the prices formerly observed by price collectors visiting the stores can simply be replaced by unit values from scanner data without changing the sampling design and the price index number formula used. If the statistical agency decides to use all the available data rather than taking samples, which seems an obvious approach, multilateral price index number methods are most suitable. Multilateral methods were originally developed to compare price levels across countries, but they can be easily adapted to price comparisons over time. These methods are particularly useful for scanner data, where item turnover is often large and promotional sales occur frequently.

This section briefly describes the most important multilateral price index number methods; for a comprehensive discussion, see Chapter XX, Volume 2 of this Manual. For convenience, a short overview of traditional bilateral price indexes and chaining is provided.

3.2 Bilateral price indexes and chaining

Suppose first that the set of items sold is fixed over time, i.e. that we are dealing with a *static universe*. This fixed set of items is denoted by *S* and its size by *N*. The sample period exists of $(T+1)$ time periods $t = 0,...,T$. The prices (unit values) of item $i \in S$ $(i = 1,..., N)$ in periods 0 and t $(t = 1,...,T)$ are denoted by p_i^0 and p_i^t; q_i^0 and q_i^t are the corresponding quantities sold. The aim is to construct price indexes that compare period 0, the starting period of the time series, with each period *t*.

In the situation with no expenditure information, Chapter XX of this Manual recommends the use of the *Jevons price index*, the unweighted geometric mean of price relatives:

$$P_J^{0,t} = \prod_{i \in S} \left(\frac{p_i^t}{p_i^0} \right)^{1/N} = \frac{\prod_{i \in S} (p_i^t)^{1/N}}{\prod_{i \in S} (p_i^0)^{1/N}} . \tag{1}$$

The statistical agency traditionally draws a sample of items from the entire universe *S* to reduce CPI production cost. Without access to scanner data, *S* is unknown and a sampling frame is lacking. Most CPI samples have therefore been drawn purposively, with the risk of introducing bias in the index.

Since scanner data contains expenditure information for a census of items, the construction of superlative price indexes is possible on the entire set *S*. We focus on the Törnqvist rather than, for example, the Fisher formula; usually, the two indexes produce very similar results. The *Törnqvist price index* is given by

$$P_T^{0,t} = \prod_{i \in S} \left(\frac{p_i^t}{p_i^0} \right)^{(s_i^0 + s_i^t)/2}, \tag{2}$$

where $s_i^0 = p_i^0 q_i^0 / \sum_{i \in S} p_i^0 q_i^0$ and $s_i^t = p_i^t q_i^t / \sum_{i \in S} p_i^t q_i^t$ denote the expenditure shares in periods 0 and t.

In a *dynamic universe* there are new and disappearing items so that not all items can be matched over time. The sets of items in periods t $(t = 0,...,T)$ are denoted by S^t with size N^t . To maximize the number of matches in the data, chaining matched-model superlative price indexes seems useful, e.g. chaining period-on-period Törnqvist price indexes

$$P_T^{t-1,t} = \prod_{i \in S_M^{t-1,t}} \left(\frac{p_i^t}{p_i^{t-1}} \right)^{\frac{s_{iM}^{t-1} + s_{iM}^t}{2}}, \tag{3}$$

where s_{iM}^{t-1} and s_{iM}^t are the expenditure shares in the two periods with respect to the set $S_M^{t-1,t} = S^{t-1} \cap S^t$ of matched set of items that are available in both period *t*-1 and period *t*. However, empirical work showed that high-frequency chaining of superlative price indexes can lead to strong *chain drift*, which in scanner data is often due to promotional sales. Chain drift in superlative price indexes due to sales is typically downward.[7]

A simple solution to the chain-drift problem would be not to weight the items and construct a time series by chaining period-on-period matched-model Jevons price indexes:

$$P_J^{t-1,t} = \prod_{i \in S_M^{t-1,t}} \left(\frac{p_i^t}{p_i^{t-1}} \right)^{\frac{1}{N_M^{t-1,t}}}, \tag{4}$$

where $N_M^{t-1,t}$ is the number of matched items between periods *t*-1 and *t*. Item expenditures are usually highly skewed, and a crude form of implicit weighting can be attained by excluding many low-expenditure items, e.g. using a threshold based on the items' average expenditure shares in adjacent months.[8] Yet, the lack of explicit weighting is obviously not an optimal situation. A better solution would be to construct weighted multilateral price indexes.

3.3 Multilateral methods

Transitivity is desirable for price comparisons across countries because it means that the results will be independent of the choice of base country. Adapted to comparisons over time,

[7] Ivancic (2007), using ACNielsen scanner data on goods sold in Australian supermarkets, found downward drift in chained Fisher price indexes; see also Ivancic, Diewert and Fox (2009) (2011). Drift in chained matched-model superlative price indexes has been documented for durable goods as well. Here, the drift is likely due to seasonal fluctuations in prices and quantities. De Haan and Krsinich (2014), using scanner data from GfK, found downward drift in chained Törnqvist price indexes for consumer electronics goods sold in New Zealand. Silver and Heravi (2005) presented evidence of downward bias in chained Fisher indexes using UK GfK scanner data on televisions.

[8] This is what Statistics Netherlands did when they changed the method for the treatment of scanner data from supermarket chains in January 2010; for details, see van der Grient and de Haan (2010) (2011).

the use of multilateral price index methods as the results will then be independent of the choice of base period and therefore free from chain drift. Multilateral methods have in common that price indexes are constructed simultaneously for the entire sample period.

Two types of multilateral method can be distinguished. The first type starts from matched-model price comparisons between any pair of time periods across the entire sample period and then 'transitivizes' this set of bilateral price indexes. The best-known method is GEKS (Gini, 1931; Eltetö and Köves; 1964; Szulc, 1964).[9] The second type attains transitivity in another way, which will be explained below, and includes the Geary-Khamis method (Geary, 1958; Khamis, 1972) and the Country Product Dummy method (Summers, 1973).

3.3.1 GEKS method

The GEKS index between period 0 and period t is calculated as the geometric average of the ratios of the matched-model bilateral price indexes $P^{j,l}$ and $P^{k,l}$, constructed with the same index number formula, where each period l is taken as the base. Provided that the bilateral indexes satisfy the time reversal test, the GEKS index can be written as (Ivancic, Diewert and Fox, 2011; de Haan and Van der Grient, 2011)

$$P_{GEKS}^{0,t} = \prod_{l=0}^{T} \left[\frac{P^{0,l}}{P^{t,l}} \right]^{\frac{1}{T+1}} = \prod_{l=0}^{T} \left[P^{0,l} \times P^{l,t} \right]^{\frac{1}{T+1}} . \tag{5}$$

The time reversal test requires that when the base period and the comparison period are reversed, the result should be equal to the reciprocal of the original index. In its standard form, the GEKS method uses bilateral Fisher indexes, which satisfy the test, but other choices are possible, including bilateral Törnqvist indexes.

The choice of window length remains a point of concern. Ivancic, Diewert and Fox (2011) advocated a 13-month (or 5-quarter) window as this is the shortest window that can deal with strongly seasonal goods. Enlarging the window would lead to a *loss of characteristicity* in that recent price movements would be increasingly affected by prices and price changes in the distant past.[10]

3.3.2 Geary-Khamis method

The Geary-Khamis (GK) method, when applied to comparisons over time, gives rise to the following price index:

---

[9] Another method is based on spanning trees (Hill, 1999a; 1999b), where a spanning tree is a supplier of paths between countries. For a particular spanning tree, the bilateral indexes are chain-linked in order to construct price comparisons between any pair of countries or, adapted to our context, time periods. It is not clear, however, what the theoretical and practical advantages are over the easier-to-construct GEKS indexes.

[10] It is possible to construct weighted GEKS indexes, which may take into account the reliability of the bilateral price indexes (Rao, 2001). Melser (2016) proposed a weighted GEKS method where the weights depend on the degree of matching of the items, for example in terms of expenditure shares. Here, the choice of window length is less important since bilateral indexes with a lower degree of matching will be down-weighted.

$$P_{GK}^{0,t} = \frac{\sum_{i\in S^t} p_i^t q_i^t \Big/ \sum_{i\in S^t} \hat{p}_i q_i^t}{\sum_{i\in S^0} p_i^0 q_i^0 \Big/ \sum_{i\in S^0} \hat{p}_i q_i^0} = \frac{\left[\sum_{i\in S^t} s_i^t \left(\frac{p_i^t}{\hat{p}_i}\right)^{-1}\right]^{-1}}{\left[\sum_{i\in S^0} s_i^0 \left(\frac{p_i^0}{\hat{p}_i}\right)^{-1}\right]^{-1}}. \tag{6}$$

The numerator of (6) is a price index (using period t quantities) with "reference prices" $\hat{p}_i$ that are fixed across the sample period. The index should be equal to 1 in the starting period 0, so it will be necessary to normalize the index by dividing by its period 0 value, which is the numerator of (6).

The reference prices are given by

$$\hat{p}_i = \frac{\sum_{\tau\in S_i} q_i^\tau \left(\frac{p_i^\tau}{P_{GK}^{0,\tau}}\right)}{\sum_{\tau\in S_i} q_i^\tau}, \tag{7}$$

where S_i is the set of time periods in which item i is actually sold and for which prices are available. Equation (7) shows that $\hat{p}_i$ equals a weighted arithmetic average of the deflated observed prices, with each period's share in the total number of sales of the item across the entire sample period serving as weights.

Since the GK index acts as the deflator, (6) and (7) define a system of equations which must be solved simultaneously. This can be done iteratively, but there are other ways to solve the system (Balk, 2008).

3.3.3 Time Product Dummy method

This is a regression-based approach. Assuming N different items are observed in the entire sample period $0,...,T$ (most of which will typically not be sold in every time period), the Time Product Dummy (TPD) regression model for the pooled data is

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t, \tag{8}$$

where D_i is a dummy variable that has the value of 1 if the observation relates to item i and 0 otherwise, and D_i^t is a dummy variable with the value 1 if the observation relates to period t and 0 otherwise; dummies for item N and period 0 are excluded to identify the model.

Diewert (2004) proposed to estimate model (8) by Weighted Least Squares regression with the items' expenditure shares in each period serving as weights. Exponentiating the estimated time dummy parameter $\hat{\delta}^t$ yields the TPD index between periods 0 and t; $P_{TPD}^{0,t} = \exp(\hat{\delta}^t)$. The weighted TPD method can be written as a system of equations that is similar to the GK system defined by (6) and (7) (Rao, 2005):

$$P_{TPD}^{0,t} = \frac{\prod_{i \in S^t} \left(\frac{p_i^t}{\exp(\hat{\gamma}_i)} \right)^{s_i^t}}{\prod_{i \in S^0} \left(\frac{p_i^0}{\exp(\hat{\gamma}_i)} \right)^{s_i^0}}; \tag{9}$$

$$\exp(\hat{\gamma}_i) = \prod_{\tau \in S_i} \left(\frac{p_i^\tau}{P_{TPD}^{0,\tau}} \right)^{\frac{s_i^\tau}{\sum_{\tau \in S_i} s_i^\tau}}. \tag{10}$$

Equation (10) shows that the exponentiated item fixed effect estimates $\hat{\gamma}_i$, or reference prices, are equal to the expenditure-share weighted geometric averages of the deflated prices with the TPD index serving as deflator.

Notice that the TPD index (9) can be viewed as a normalized geometric Paasche index with imputed period 0 prices based on the reference prices (10). Similarly, the GK index (6) can be viewed as a normalized (ordinary) Paasche index with imputed period 0 prices based on the reference prices (7). In section 3.5 an alternative interpretation of the two indexes will be given.

3.4 Lack of matching and quality adjustment

3.4.1 Implicit quality adjustment

Like GEKS, GK and TPD are matched-model methods in the sense that items with a single observation in the entire sample period do not affect the index. This is easy to understand: items contribute to aggregate price change only when *price relatives* can be calculated from the prices observed in both periods compared, unless information on characteristics would be available to perform explicit quality adjustments; see subsection 3.4.3. One implication of the matched-model property is that items introduced in the most recent period T are ignored in any multilateral method.

In equations (6) and (9) for the GK index and TPD index, respectively, the base period prices for all the items sold in period t, including (unmatched) items that were not sold in period 0, are imputed by the reference prices. It can therefore be argued that the GK and TPD methods perform *implicit quality adjustment*, albeit partially as items observed only once in the sample period are ignored; see also Krsinich (2016).

Implicit quality adjustment can also be illustrated by using an alternate interpretation of the GK index. Dividing the value index of the product category by the ratio of 'quality-adjusted quantities' defines a *quality-adjusted unit value index* (de Haan 2004; 2015):

$$P_{QAUV}^{0,t} = \frac{\sum_{i \in S^t} p_i^t q_i^t \Big/ \sum_{i \in S^0} p_i^0 q_i^0}{\sum_{i \in S^t} \lambda_{i/b} q_i^t \Big/ \sum_{i \in S^0} \lambda_{i/b} q_i^0} = \frac{\sum_{i \in S^t} p_i^t q_i^t \Big/ \sum_{i \in S^t} \lambda_{i/b} q_i^t}{\sum_{i \in S^0} p_i^0 q_i^0 \Big/ \sum_{i \in S^0} \lambda_{i/b} q_i^0} = \frac{\left[\sum_{i \in S^t} s_i^t \left(\frac{p_i^t}{\lambda_{i/b}} \right)^{-1} \right]^{-1}}{\left[\sum_{i \in S^0} s_i^0 \left(\frac{p_i^0}{\lambda_{i/b}} \right)^{-1} \right]^{-1}}. \tag{11}$$

Notice that if $\lambda_{i/b} = 1$ for all *i*, then (11) simplifies to the ordinary unit value index.[11] The quality-adjustment factors $\lambda_{i/b}$ aim to express the quantities purchased of each item *i* in terms of quantities of an arbitrary item *b*. The ratios $p_i^t / \lambda_{i/b}$ and $p_i^0 / \lambda_{i/b}$ in the second expression of (11) are quality-adjusted prices.

A comparison of equations (6) and (11) shows that the GK index can be viewed as a quality-adjusted unit value index where the quality-adjustment factors are measured by the reference prices (7). Similarly, the TPD index (9) can be viewed as its geometric counterpart where the quality-adjustment factors are measured by the reference prices (10). Whether the reference prices in the GK and TPD indexes properly reflect quality differences is likely to depend on the market circumstances (Silver and Heravi, 2005).

Data on item characteristics permitting, *explicit quality adjustment* is preferable, in particular using hedonic regression.

3.4.2 Explicit quality adjustment

A useful starting point is the multilateral Time Dummy Hedonic (TDH) model

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k z_{ik} + \varepsilon_i^t, \tag{12}$$

where z_{ik} is the quantity of characteristic k $(k = 1,...,K)$ for item *i*. Notice that, as pointed out by Aizcorbe, Corrado and Doms (2003) and Krsinich (2016), the TPD model (8) arises from the TDH model (12) by replacing the hedonic effects $\exp(\sum_{k=1}^{K} \beta_k z_{ik})$ by item-specific fixed effects $\exp(\gamma_i)$. Similar to the estimation of the TPD model it is assumed that (12) is estimated by expenditure-share weighted regression on the pooled data of all time periods $t = 0,...,T$.

The resulting weighted TDH index, $P_{TDH}^{0,t} = \exp(\hat{\delta}^t)$, can be expressed in a similar way as the TPD index (9), with the estimated hedonic effects $\exp(\sum_{k=1}^{K} \hat{\beta}_k z_{ik})$ instead of the estimated item fixed effects $\exp(\hat{\gamma}_i)$ now acting as reference prices. The $\exp(\sum_{k=1}^{K} \hat{\beta}_k z_{ik})$ can also be used to estimate the quality-adjustment factors $\lambda_{i/b}$ in equation (11). The resulting explicitly quality-adjusted unit value index – or 'hedonic variant' of the GK index – is expected to be very close to the TDH index (de Haan and Krsinich, 2017).

[11] In the static universe case (without new and disappearing items), $\lambda_{i/b} = p_i^0 / p_b^0$ turns the quality-adjusted unit value index into the Paasche price index. Von Auer (2014) showed that many conventional price indexes can be viewed as, what he called, a *generalized unit value index*.

De Haan, Hendriks and Scholz (2016) compared the TPD and TDH methods. They argued that the TPD model suffers from overfitting because it has too many parameters) and “distorts the regression residuals towards zero”. Under certain pricing strategies of manufacturers or retailers, such as price skimming (new items) and dumping (old items), the TPD index can be quite different from the TDH index. Similarly, the GK index can be quite different from its hedonic counterpart. If relaunches of homogeneous products with different barcodes or SKUs are a major part of the problem, then defining items by their characteristics is likely to reduce the differences to a large extent.

The GEKS method does not aim at implicitly adjusting for quality change. It is possible to estimate explicitly quality-adjusted GEKS indexes by replacing the bilateral matched-model Törnqvist price indexes by bilateral hedonic imputation Törnqvist indexes, as proposed by de Haan and Krsinich (2014).

3.5 Revisions in multilateral indexes

When new data becomes available, previously estimated multilateral indexes change. This is problematic because the CPI is not revisable. Two types of method have been proposed to extend a multilateral time series without revising published index numbers: rolling window methods and an annually-chained direct method.

Rolling window methods estimate multilateral indexes on a window with fixed length, which is shifted forwards each period. The results of the latest window are then spliced onto the existing time series, for example by splicing the most recent movement onto the latest index number. An alternative to this *movement splice* is a *window splice*, which splices the most recently estimated movement across the entire window onto the index level of T-1 periods ago. The movement splice was proposed by Ivancic, Diewert and Fox (2011) in the context of the GEKS method and the window splice was proposed by Krsinich (2016) in the context of TPD.

These two extension methods splice price movements onto a single link period. Since all link periods are equally valid, Diewert and Fox (2017) proposed using a *mean splice* by taking the geometric mean of the price indexes obtained from using every possible link period. This makes the result independent on the choice of link period.

The annually chained *direct extension method* (Chessa, 2016) constructs multilateral index series of, say, 13 months, starting in e.g. December and ending in December of the next year, and chain links them in December of each year to obtain a long-term time series. The length of the estimation window for the short-term indexes is extended each month – the index for January in the short-term series is estimated on two months of data (which is actually a bilateral rather than multilateral comparison), and so forth, until in December thirteen months of data is used.

A potential weakness of the direct extension method is that the price indexes for the first couple of months of each year are based on sparse data and expected to be volatile. Also, December acts as the short-term index reference period and is given special importance. If, for some reason, December is an 'unusual' month, the results may be adversely affected.[12]

[12] Lamboray (2017) suggested combining the annually chained direct extension method with a rolling window approach to mitigate these problems.

4. Implementation of multilateral methods

4.1 Assessing multilateral methods

The implementation of new data sources and methods in any statistical series requires careful consideration of the statistical impacts, as well as the benefits and costs. Only a handful of NSOs have actually implemented multilateral price indexes in the CPI, displaying caution in implementation and a divergence in methods and practices.[13]

ABS (2016) proposed criteria to assess multilateral methods considering a broad concept of statistical quality. The framework includes seven dimensions of statistical quality:

- Institutional Environment – pertains to the institutional and organizational context in which a statistical producer operates;
- Relevance – pertains to how well a statistic meets user needs;
- Timeliness – pertains to how quickly and frequently the statistic is published;
- Accuracy – pertains to how well a statistic measures the desired concept;
- Coherence – pertains to how consistent the statistic is with sources of related information;
- Interpretability – pertains to the information available to provide insight into the statistic; and
- Accessibility – pertains to ease of access to the statistic.

This framework can be used by the NSO to determine the benefits and challenges of using multilateral index methods in the local context.

Multilateral index methods to compile the CPI can also be assessed from a theoretical perspective. The assessment can utilize approaches previously applied to bilateral and spatial indexes. This Manual assesses bilateral price indexes both from axiomatic/test approaches (Chapter 16) and economic approaches (Chapters 17-18). Similar approaches to assessing multilateral indexes in a spatial context have been developed and presented in several papers, especially Diewert (1999) and Balk (2001).

ABS (2016) provides a detailed description of the theoretical assessments of multilateral price index methods in the present temporal context using the axiomatic/test and economic theory approaches. This assessment can be used as a basis for NSOs to undertake similar assessments in their local context. A comprehensive discussion of the various multilateral methods using the economic approach to index number theory can be found in Chapter XX of Volume 2 of this Manual; see also Diewert and Fox (2017). The most important result is that GEKS deals appropriately with substitution effects whereas GK and TPD are appropriate only under restrictive assumptions about consumer preferences.

[13] See Attachment A of this Annex.

In addition, expert peer review of the proposed multilateral methods may be appropriate in circumstances where users of the CPI may expect NSOs to demonstrate broader endorsement of the proposed changes.

4.2 Calculating indexes

4.1.1 Operational choices

The matched-model property of (non-hedonic) multilateral indexes implies that without any manual intervention, the results depend on the *choice of item identifier*. For example, when using barcode as item identifier, the price change of a homogenous product whose barcode changes at the same time – a 'relaunch' – will not be measured. As mentioned earlier, the use of Stock Keeping Unit (SKU) mitigates the problem since SKU generally consists of multiple barcodes for similar items and is more stable than barcode. Nevertheless, even SKU may be too detailed.

If a relatively small number of observable attributes with discrete values suffice to define homogeneous products, items could be defined by cross-classifying the sets of categorical variables for each attribute and prices calculated as unit values across all the barcodes/SKUs. Most likely there will still be new and disappearing items (cells) across the sample period. To maximize the degree of matching without introducing chain drift, a multilateral method could be applied (Chessa, 2016). A potential issue is that the available characteristics information may be limited, especially when the characteristics are extracted from product descriptions in scanner data, which are often rather broad. In this case, unit value bias is likely to arise. Also, if the characteristics information is deemed sufficient, it may be better to construct hedonic indexes.

While taking a (cut-off) sample that ignores items, however defined, with small expenditure shares would in many cases not significantly affect the results, it is not necessary when using a weighted multilateral method. Most of the issues discussed in section 2.4, like the choice of calculating unit values at the store or chain level, and the need to have an index structure that facilitates the use of scanner data, apply here as well. GK and TPD in particular require rather detailed chain-specific Elementary Aggregates because their (approximate) additivity is not supported by economic theory and (implicit) quality adjustment is only useful for broadly similar items. Note that the estimation of quality-adjusted GEKS indexes would also require such detailed Elementary Aggregates.

4.1.2 Producing empirical results

The purpose of producing empirical results of multilateral methods is twofold: to examine the performance of various methods in local contexts; as well as demonstrating to users of the CPI the likely impacts of moving from *current* CPI data sources and methods to new approaches.

Ideally these multilateral methods should be examined against each other as well as in comparison to the official CPI. These comparisons should be undertaken at the lowest level of the published CPI as well as at various aggregation levels, including the *Total CPI*.

A number of insights can be obtained from producing empirical results. Often these insights further reinforce the theoretical arguments for utilizing multilateral index methods to compile the CPI. This may include the impact of using contemporaneous information for weighting purposes that capture consumer behavior, including substitution, over time.[14] The empirical results should be communicated to CPI users and stakeholders.

4.2 Communicating with users and stakeholders

The use of scanner data to compile the CPI can represent quite a significant change to the data sources and methods employed by NSOs for many years. These changes need to be carefully communicated to CPI users and stakeholders.

A suggested set of activities includes:

- publishing information papers that outline the proposed new methods and data sources;
- conducting face-to-face meetings with key stakeholders (e.g. Central Banks, Treasury and Finance Departments) and other interested parties, including members of the public;
- using media releases and briefing of economic journalists to help inform the public of proposed changes;
- encouraging stakeholders and the public to provide submissions to the NSO for consideration; and
- engaging with leading academics to both review the proposed changes and to encourage their support.

Following this consultation, which could take a couple of years, the NSO should publish a position paper that both responds to the topics raised as part of the consultation process; and clearly articulates how the NSO will proceed with the use of scanner data to compile the CPI, including the rationale and empirical results that support this approach. The positon paper should clearly state the data sources and methods to be employed, and a timetable for the implementation of changes.

4.3 Publication and dissemination

Following the publication of the position paper, it is suggested the NSO compile in parallel the CPI using the current and new data sources and methods for a period of approximately six

[14] For empirical evidence on multilateral price index methods, see Chapter 5 of ABS (2016), Chapter 3 of ABS (2017), and Chessa, Verburg and Willenborg (2017). The evidence indicates that GEKS, GK and TPD can yield very similar results, despite the theoretical differences.

months. This *transition* period allows the NSO to refine processes and procedures to compile the CPI using the new methods, as well as to confront the empirical results of the two approaches. This transition period is often the first opportunity for the NSO to utilize the new data sources and methods in real time, i.e. in accordance with the CPI processing and publication timetable. It is at the NSO's discretion whether the results of the parallel processing are made public.

The first period for which the new data sources and methods are used to compile the CPI should be accompanied by a significant amount of communication with media and users. This will ensure the changes to the CPI are well understood.

Further information relating to the dissemination of the CPI can be found in chapter X of this manual.

Attachment A:

A summary of NSOs using scanner data [To be updated]

Country	Source of scanner data		Use of scanner data			Multilateral methods			
	Direct from businesses	Secondary sources e.g. Market research companies	Confrontation	Price replacement	Sample enhancement	GEKS	GK	TPD	

References

Aizcorbe, A., C. Corrado and M. Doms (2003), "When Do Matched-Model and Hedonic Techniques Yield Similar Price Measures?", Working Paper 2003-14, Federal Reserve Bank of San Francisco.

von Auer, L. (2014), "The Generalized Unit Value Index Family", *Review of Income and Wealth* 60, 843-861.

Australian Bureau of Statistics (ABS) (2013), "Feature Article: The Use of Transactions Data to Compile the Australian Consumer Price Index", 6401.0, 23 October 2013, Canberra: ABS.

Australian Bureau of Statistics (ABS) (2016), "Making Greater Use of Transactions Data to Compile the Consumer Price Index", Information Paper 6401.0.60.003, 29 November 2016, Canberra: ABS.

Australian Bureau of Statistics (ABS) (2017), "An Implementation Plan to Maximise the Use of Transactions Data in the CPI", Information Paper 6401.0.60.004, 16 June 2017, Canberra: ABS.

Balk, B.M. (1998), "On the Use of Unit Value indices as Consumer Price Subindices", in: *Proceedings of the Fourth Meeting of the International Working Group on Price Indices*, pp. 112-120, Washington DC, U.S. Bureau of Labor Statistics.

Balk, B.M. (2005). "Price Indexes for Elementary Aggregates: The Sampling Approach", *Journal of Official Statistics* 21, 675-699.

Balk, B.M. (2008), *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*. New York: Cambridge University Press.

Bird, D., R. Breton, C. Payne and A. Restieaux (2014), "Initial Report on Experiences with Scanner Data in ONS", Report, Office for National Statistics, Newport, UK.

Böttcher, I. and S. Sergeev (2014), "Austrian Scanner Data Project – Report Multipurpose Consumer Price Statistics: The Use of Scanner Data", Paper presented at the ECE/ILO meeting of the Group of Experts on Consumer Price Indices, 26-28 May 2014, Geneva, Switzerland.

Chessa, A.G. (2016), "A New Methodology for Processing Scanner Data in the Dutch CPI", *Eurona* 1, 49-69.

Chessa, A.G., J. Verburg and L. Willenborg (2017), "A Comparison of Price Index Methods for Scanner Data", Paper presented at the 15th meeting of the Ottawa Group, 10-12 May 2017, Altville am Rhein, Germany.

Dalen, J. (2017), "Unit Values and Aggregation in Scanner Data – Towards a Best Practice", Paper presented at the 15th meeting of the Ottawa Group, 10-12 May 2017, Altville am Rhein, Germany.

Diewert, W.E. (2004), "On the Stochastic Approach to Linking the Regions in the ICP", Discussion Paper no. 04-16, Department of Economics, The University of British Columbia, Vancouver, Canada.

Diewert, W.E., K.J. Fox and J. de Haan (2016), "A Newly Identified Source of Potential CPI Bias: Weekly versus Monthly Unit Value Price Indexes", *Economics Letters* 141, 169-172.

Diewert, W.E. and K.J. Fox (2017), "Substitution Bias in Multilateral Methods for CPI Construction Using Scanner Data", Discussion Paper 17-02, Vancouver School of Economics, The University of British Columbia, Vancouver, Canada.

Eltetö, Ö. and P. Köves (1964), "On an Index Computation Problem in International Comparisons" [in Hungerian], *Statiztikai Szemle* 42, 507-518.

Fenwick, D. (2014), "Exploiting New Technologies and New Data Sources – The Opportunities and Challenges Associated with Scanner Data", Paper presented at the ECE/ILO meeting of the Group of Experts on Consumer Price Indices, 26-28 May 2014, Geneva, Switzerland.

Geary, R.C. (1958), "A Note on the Comparison of Exchange Rates and Purchasing Power between Countries", *Journal of the Royal Statistical Society A* 121, 97-99.

Gini C. (1931), "On the Circular Test of Index Numbers", *International Review of Statistics* 9, 3-25.

van der Grient, H.A. and J. de Haan (2010), "The Use of Supermarket Scanner Data in the Dutch CPI", Paper presented at the ECE/ILO Workshop on Scanner Data, 10 May 2010, Geneva, Switzerland.

van der Grient, H. and J. de Haan (2011), "Scanner Data Price Indexes: The "Dutch" Method versus Rolling Year GEKS", Paper presented at the 12th meeting of the Ottawa Group, 4-6 May 2011, Wellington, New Zealand.

de Haan, J. (2004), "Estimating Quality-Adjusted Unit Value Indexes: Evidence from Scanner Data", Paper presented at the SSHRC International Conference on Index Number Theory and the Measurement of Prices and Productivity, 30 June – 3 July 2004, Vancouver, Canada.

de Haan, J. (2015), "A Framework for Large Scale Use of Scanner Data in the Dutch CPI", Paper presented at the 14th meeting of the Ottawa Group, 20-22 May 2015, Tokyo, Japan.

de Haan, J., E. Opperdoes and C.M. Schut (1999), "Item Selection in the Consumer Price Index: Cut-Off versus Probability Sampling", *Survey Methodology* 25, 31-41.

de Haan, J. and H.A. van der Grient (2011), "Eliminating Chain Drift in Price Indexes Based on Scanner Data", *Journal of Econometrics* 161, 36-46.

de Haan, J. and F. Krsinich (2014), "Scanner Data and the Treatment of Quality Change in Nonrevisable Price Indexes", *Journal of Business & Economic Statistics* 32, 341-358.

de Haan, J., R. Hendriks and M. Scholz (2016), "A Comparison of Weighted Time-Product Dummy and Time Dummy Hedonic Indexes", Graz Economics Papers 2016-13, Department of Economics, University of Graz, Austria.

de Haan, J. and F. Krsinich (2017), "Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ?", *Review of Income and Wealth,* DOI: 10.1111/roiw.12304.

Hill, R.J. (1999a), "Comparing Price Levels across Countries using Spanning Trees", in: *International and Inter-Area Comparisons of Income, Output and Prices,* A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth, Volume 61, University of Chicago Press.

Hill, R.J. (1999b), "Comparing Price Levels across Countries Using Minimum-Spanning Trees", *The Review of Economics and Statistics* 81, 135-142.

Howard, A., K. Dunford, J. Jones, M. van Kints, K. Naylor and R. Tarnow-Mordi (2015), "Using Transactions Data to Enhance the Australian CPI", Paper presented at the 14th meeting of the Ottawa Group, 20-22 May 2015, Tokyo, Japan.

Ivancic, L. (2007), *Scanner Data and the Construction of Price Indices*, PhD thesis, University of New South Wales, Sydney, Australia.

Ivancic, L., W.E. Diewert and K.J. Fox (2009), "Scanner Data, Time Aggregation and the Construction of Price Indexes", Discussion Paper no. 09-09, Department of Economics, University of British Columbia, Vancouver, Canada.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), "Scanner Data, Time Aggregation and the Construction of Price Indexes", *Journal of Econometrics* 161, 24-35.

Ivancic, L. and K.J. Fox (2013), "Understanding Price Variation across Stores and Supermarket Chains: Some Implications for CPI Aggregation Methods", *Review of Income and Wealth* 59, 629-647.

Khamis, S.H. (1972), "A New System of Index Numbers for National and International Purposes", *Journal of the Royal Statistical Society A* 135, 96-121.

Krsinich, F (2014), "The FEW Index: Fixed effects with a window splice; Non revisable quality-adjusted price indexes with no characteristic information", Paper presented at the meeting of the group of experts on consumer price indices, 26-28 May 2014, Geneva, Switzerland

Krsinich, F (2015), "Implementation of Consumer Electronics Scanner Data in the New Zealand CPI", Paper presented at the 14th meeting of the Ottawa Group, 20-22 May 2015, Tokyo, Japan.

Krsinich, F. (2016), "The FEWS Index: Fixed Effects with a Window Splice", *Journal of Official Statistics* 32, 375-404.

Lamboray, C. (2017), "The Geary Khamis Index and the Lehr Index: How Much Do They Differ?", Paper presented at the 15th meeting of the Ottawa Group, 10-12 May 2017, Altville, Germany.

Melser, D. (2016), "Scanner Data Price Indexes: Addressing Some Unresolved Issues", *Journal of Business & Economic Statistics*, DOI: 10.1080/07350015.2016.1218339.

Muller, R. (2010), "Scanner Data in the Swiss CPI: An Alternative to Price Collection in the Field", Room document at the ECE/ILO meeting of the Group of Experts on Consumer Price Indices, 10-12 May 2010, Geneva, Switzerland.

Rao, D.S.P. (2001), "Weighted EKS and Generalized CPD methods for Aggregation at the Basic Heading Level and Above Basic Heading Level", Joint World Bank – OECD Seminar on purchasing Power Parities, February 2001, *Recent Advances in Methods and Applications*, Washington DC, U.S.

Rao, D.S.P. (2005), "On the Equivalence of Weighted Country-Product Dummy (CPD) Method and the Rao-System for Multilateral Price Comparisons", *Review of Income and Wealth* 51, 571-580.

Schut, C. et al. (2002), "The Use of Scanner Data from Supermarkets in the Consumer Price Index" [in Dutch], Statistics Netherlands, Voorburg, The Netherlands.

Silver, M. and S. Heravi (2005), "A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment Using Scanner Data", *Journal of Business & Economic Statistics* 23, 269-281.

Summers, R. (1973), "International Price Comparisons Based Upon Incomplete Data", *Review of Income and Wealth* 19, 1-16.

Szulc, B. (1964), "Index Numbers of Multilateral Regional Comparisons" [in Polish], *Przeglad Statystyczny* 3, 239-254.