Statistics Netherlands

Report

Towards a generic price index method for scanner data in the Dutch CPI

Antonio Chessa[1]

Statistics Netherlands, CPI department

[1] E-mail: ag.chessa@cbs.nl

CBS Den Haag
Henri Faasdreef 312
2492 JP The Hague
P.O. Box 24500
2490 HA The Hague
+31 70 337 38 00

www.cbs.nl

project number

EOC
May 2015

remarks The views expressed in this paper are those of the author(s) and do not necessarily reflect the policies of Statistics Netherlands.

Abstract

In January 2013, scanner data fully replaced traditional price collection for supermarkets in the Dutch CPI. Meanwhile, scanner data from other providers have been acquired, which are considered for inclusion in the CPI in the coming years. These plans have motivated a search towards a more generic price index method, with the aim of applying the method to different types of consumer goods.

This paper presents an adaptation of the Geary-Khamis (GK) method, which was originally proposed for international price comparisons. When applied to a time setting, a price index is calculated by dividing a turnover index by a weighted quantity ("volume") index. A number of problems need to be resolved in order to apply the GK-method to the time domain: (1) The handling of new and disappearing goods; (2) The compilation of price indices for the publication month, as sales data for that month become available; (3) The choice of the length of the time window for price index calculation at elementary levels.

Statistical and empirical evidence has revealed a slight preference for a 1-year window, which also coincides with current practice in the Dutch CPI. In theory, this means that the product specific quantity 'weights' in the volume index are fixed for 1-year periods. In order to include all goods in the calculations, the product weights are calculated from price and quantity information of the current year. The product weights are therefore updated each month, which gives rise to a "real time index". Differences with the theoretical benchmark, with fixed weights, turn out to be negligible. The real time index does not drift from the benchmark, as both indices coincide at the end of each year. A fixed time window is preferred over a moving window; it is demonstrated here that the latter may lead to serious drift.

An iterative method, with a suitably chosen initial index, is proposed for calculating price indices. The method shows fast convergence rates. Other important issues are discussed, in particular the problem of defining "homogeneous" products. As the GK-method is known for the "substitution bias" in the PPP-context, an alternative formula for the product weights is considered and compared with the weighted arithmetic average of deflated prices. First results show negligible differences between the two formulas.

The index method is currently in a test phase for a Dutch department store and for mobile phones, with the aim of being included in the CPI in January 2016.

1. Introduction

The first use of scanner data in the Dutch CPI goes back to 2002. Since then, the use of such data has been extended through a number of phases and possibilities for widening its range of application are still being investigated. Scanner data have clear advantages over traditional survey data collection, notably because such data sets offer a better coverage of articles sold, sales data offer complete information (prices and quantities), and the data collection process is automised. In spite of their potential, scanner data are still used by a small number of statistical agencies in their CPI.[2]

In this paper, by scanner data we mean transaction data specifying turnover and numbers of articles sold by EAN (or GTIN, barcode). Scanner data were introduced in the Dutch CPI in 2002, which then involved two supermarket chains. In January 2010, the data were extended to six supermarket chains, as part of a re-design of the CPI (de Haan, 2006; van der Grient and de Haan, 2010; de Haan and van der Grient, 2011). At present, scanner data of 10 supermarket chains are used and surveys are not carried out anymore for supermarkets since January 2013. Scanner data are also used for do-it-yourself stores. Other forms of electronic data containing both price and quantity information are obtained from travel agencies, for fuel prices and for mobile phones. At present, more than 22% of the Dutch CPI is based on electronic data (in terms of Coicop weights).

The use of scanner data has thus grown considerably over the past 10 years and possibilities of using scanner data for other parts of the CPI are being investigated (e.g., see Chessa (2013)). Different price index methods are being used for different electronic data sets. Along with the search for new electronic data sources, the question has been put forward whether a generic index method could be developed that would be suitable for different consumer goods and data sets.

There are additional reasons for pursuing a generic method: (i) Some of the current methods take into account new products only in the year following their introduction into the market, (ii) price increases associated with re-codings of EANs, while the consumable part of an article remains unchanged ("re-launches"), are partially handled for supermarkets through the use of "dump price filters", (iii) system maintenance and monthly CPI validation would greatly benefit from using the same method for different data sets.

An index method is proposed, which appears to be an adaptation of the Geary-Khamis (GK) PPP-method to the time domain. The GK-method was originally proposed by Geary (1958) and popularised by Khamis (1972) (also, see Balk (1996, 2001, 2012)). When applied to a temporal setting, the method consists of dividing a turnover index by a weighted quantity, or "volume", index. The price index formula can be rewritten as the ratio of a unit value index and an index that measures changes in the 'product mix', so that the proposed method will be referred to as a "quality-adjusted unit value index method" (shortened to "QU-method").

The index method is described and motivated in Section 2. The transition from a spatial to a time domain brings along specific problems, which pertain to the dynamics of the temporal setting. The QU-method aims at incorporating price and quantity information of all goods, including new and disappearing goods. Section 2 discusses how the QU-method deals with assortment changes over time. It also describes how price

[2] The yearly organised scanner data workshops show gradual advances in the number of countries using scanner data. The 2014 workshop in Vienna evidenced that some 'new' countries are expecting their first data, while other countries made concrete steps towards acquiring their first scanner data.

indices are calculated for the publication month (direct or chained) and how additional monthly prices and quantities are incorporated. Other choices, in particular concerning the length of the time window for price index calculations at elementary levels, are also motivated. A computational method is presented in Section 2.3, which is illustrated with some results.

The QU-method is compared to other methods in Section 3, amongst which Krsinich's (2014) "window splice" method. The latter method calculates time product dummy indices for the publication period by making use of a moving, or rolling, time window. Part of Section 3 thus focuses on the comparison between a fixed and a moving time window.

Other important problems are discussed in Section 4. An increased use of electronic data implies that NSI's have to accept that information about individual items (EANs for scanner data) may be delivered in varying formats by different providers. Depending on the level of detail and organisation of the information delivered, text mining methods may have to be developed in order to extract information about article attributes from EAN-descriptions. This is necessary in order to address the essential question that precedes price index calculations: to determine appropriate levels of product differentiation, to define "homogeneous products". Which attributes need to be selected in order to consider combinations of attributes as homogeneous products? Or could each EAN be taken to represent a distinct homogeneous product (when problems with re-coding of EANs do not occur)? Some thoughts on how this problem could be addressed are expressed in Section 4.

A well-known issue in the CPI, which has also been related to the GK-method, is a phenomenon known as "substitution bias". The price structure for a group of countries tends towards the prices of the larger countries when applying the GK-method in international comparisons. For the QU-method this means that the prices from periods with the highest sales would tend to be 'over-represented' in the 'product weights' of the volume index. Alternative expressions for these weights are considered and the results between the corresponding alternative QU-method and the original method of Section 2 have been compared.

Section 5 summarises the main findings and briefly discusses current developments and future plans with regard to the CPI. Some remarks are also made on the use of internet prices for price index calculations.

2. The QU-method

2.1 Price index formulas

Some of the considerations that were at the basis for choosing a GK-type of method can be formulated as follows:

(1) The QU-index simplifies to a unit value index when all products are homogeneous;
(2) Imposing an additive structure on the weighted quantity index opens possibilities of making index numbers consistent on different time scales (e.g., monthly and yearly CPI figures).

For instance, the more well-known time-product dummy and hedonic index methods (e.g., see de Haan and Krsinich (2014)) do not satisfy (1) because of their geometric structure. With regard to point (2), it is not at all clear how one would make a transition from a monthly to a quarterly or a yearly time scale with such methods. In addition, these methods are flawed with regard to their use of turnover in constructing weights (see Section 3.2).

Let G_0 and G_t denote sets of homogeneous products for distinct time periods 0 and t, and let $p_{i,t}$ and $q_{i,t}$ denote the prices and quantities sold for products $i \in G_t$, respectively.[3] According to the QU-method, the price index P_t for period t with respect to, say, a base period 0, is defined as follows:

$$(1) \qquad P_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}}.$$

The numerator is a turnover index, while the denominator is a weighted quantity ("volume") index. The unknown product specific parameters v_i in the denominator are given by:

$$(2) \qquad v_i = \sum_{z \in T} \varphi_{i,z} \frac{p_{i,z}}{P_z},$$

where

$$(3) \qquad \varphi_{i,z} = \frac{q_{i,z}}{\sum_{s \in T} q_{i,s}}$$

denotes the share of period z in the total amount of quantities sold for product i over some time interval T (the length of which is one of the factors to be determined).

The product quantities $q_{i,t}$ and $q_{i,0}$ in the volume index are multiplied by v_i, which, according to (2) and (3), is defined as a weighted average of deflated prices of product i over different time periods $z \in T$. The effect of price change is thus removed in order to yield product specific v_i in the volume index.

A straightforward rewriting of (2) gives:

$$(4) \qquad v_i = \sum_{z \in T} \frac{p_{i,z} q_{i,z}}{P_z} \Big/ \sum_{z \in T} q_{i,z}.$$

Expression (4) says that v_i is equal to turnover "in constant prices" of product i over period T, divided by the total number of products i sold in the same period. The numerator in (4) thus coincides with the notion of volume as used in national accounts. In this sense, v_i can be defined as volume per unit of product i sold.

Price index formula (1) can be written in the following compact form:

$$(5) \qquad P_t = \frac{\bar{p}_t / \bar{p}_0}{\bar{v}_t / \bar{v}_0},$$

where $\bar{p}_t$ and $\bar{v}_t$ denote weighted arithmetic averages of the prices and the v_i, respectively, over the set of goods in period t, that is,

$$(6) \qquad \bar{p}_t = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t}}{\sum_{i \in G_t} q_{i,t}},$$

[3] A different notation is used in this paper from the commonly accepted notation of time as a superscript in prices, quantities and indices. In this paper, preference is given to the notation of both product and time indices as subscripts. This is done in order to reserve the superscript for other purposes, as in Section 2.3.

$$(7) \qquad \bar{v}_t = \frac{\sum_{i \in G_t} v_i q_{i,t}}{\sum_{i \in G_t} q_{i,t}}.$$

Expression (5) shows how price index formula (1) essentially operates: When a set of goods is found to be homogeneous, in which case the v_i of all products have the same value, then the price index is equal to the unit value index, which is the numerator of (5). If a set of goods is not homogeneous, then the unit value index must be adjusted. The adjustment term is the denominator of (5), which captures shifts in the set of goods bought by consumers between two periods. A shift towards higher quality goods results in an upward effect on the volume index and, consequently, in a complementary downward effect on the price index.

Formulas (1)-(3) can be considered as a family of price indices. Special cases can be derived from this set of formulas, which are worth mentioning. Suppose that only prices and quantities from the publication period t would be used to derive the v_i. It is easily verified that the price index that results from (1) is a Laspeyres index. If price and quantity information from only the base period 0 would be used, then (1) turns into a Paasche index. The use of price and quantity information from both periods leads to a Lowe type of index:

$$(8) \qquad P_t = \frac{\sum_{i \in G_0 \cap G_t} p_{i,t}\, h(q_{i,0}, q_{i,t})}{\sum_{i \in G_0 \cap G_t} p_{i,0}\, h(q_{i,0}, q_{i,t})},$$

where h is the harmonic mean of the quantities sold in the two periods.

The expressions for the price indices obviously become more complex when information from more than two periods is included in the v_i. Formula (1) can then be expressed as a multilateral index, which encompasses all combinations of time paths when the v_i include prices and quantities of the entire period T. This makes the index method transitive. The index method will be applied in its original form in practice, as expressed by formulas (1)-(3). However, comparisons with Lowe index (8) may be very useful. For instance, one may want to show the contribution of new products to the price index, which are not captured when (8) is used as a direct index. Some examples are given in Section 3.1.

So far the basic formulas underlying the QU-method are given. In order to apply the method in practice, choices have to be made with regard to the length of the time window T, whether to keep T fixed or use it as a moving window, how to incorporate new goods and how to calculate the v_i. These issues are addressed in the next subsection.

2.2 Choices on window length and quantity weights

Statistical and empirical research carried out at the initial stages of this study has motivated the following choices with regard to the time window T:

- Preference is given to 1-year time windows over longer windows;
- Methods with a fixed window (i.e., with a fixed base month) are preferred over methods that calculate and update price index series by employing a monthly or quarterly moving window.

Scanner data of a large department store and of drug stores, and also electronic data on mobile phones, have been used to compare time windows that vary between 1 and 4 years in length. Methods with different window lengths were compared by calculating so-called "information criteria", which are a class of statistical fit measures that are useful for comparing methods and models with different numbers of parameters (Claeskens and Hjort, 2008).

A unique choice is not easy to make, as different results have been obtained for different types of goods. One-year windows turned out to give slightly better fits for the department store scanner data. Longer windows tend to show better fits for drug store scanner data, but the differences among the price indices for different window lengths are negligible in most cases. The same holds for mobile phones (Chessa, 2015). Based on the findings obtained so far, windows shorter than one year have not been considered. A 1-year window fits well with current practice in the Dutch CPI and is advantageous with regard to system maintenance compared to longer windows, as only items sold within one year have to be followed.

A fixed window is preferred over a moving window. The latter choice is proposed by Krsinich (2014). Both choices have been compared in the present study for a part of the aforementioned electronic data sets. The use of a moving window has evidenced different problems, which will be discussed in Section 3.2.

The choice for a 1-year window implies that the product specific quantity weights v_i are fixed during a year, but are allowed to vary across windows. But then the question arises how to determine the v_i. Should these be fixed based on data of the preceding year? In that case, new goods appearing in the current year would not enter the index. The following decisions were therefore made:

- The v_i are calculated from data of the current year;
- The v_i are updated each month with corresponding price and quantity information;
- Price indices are calculated with respect to the base month; that is, according to a direct index instead of monthly chaining.

The above choices imply that the theoretical 'benchmark' index with fixed v_i cannot be calculated, as the complete set of annual prices and quantities becomes available only in the final month of a year. However, the choice of calculating and monthly updating indices according to a direct index ensures that the resulting indices are identical with the theoretical benchmark at the end of each year. The obvious question is how the two price indices will compare in previous months. This will be illustrated in the next subsection with several examples. The choices made allow for the inclusion of all goods, both existing, disappearing and new ones.

2.3 Computation of a "real time index"

Inspection of formulas (1)-(3) learns that price indices cannot be calculated directly. The v_i depend on the price indices, which serve as deflators for product prices. Different methods can be proposed for computing the v_i and price indices. For instance, formulas (1)-(3) could be transformed to a system of linear equations, one for every price index.

However, given that CPI practitioners have to work with the method in the future, it is important to construct a computational method that is transparent to users. An own developed method is proposed, which can be implemented in different ways. A two-step method is suggested, which has the advantage that the first (approximate) step makes explicit how the v_i and the resulting price indices are calculated and updated from month to month, thus avoiding a sense of 'black box' among users. The two steps of the computational method can be described as follows, which are worked out in more detail afterwards:

1. An initial index is constructed for the publication month, by making use of the price indices calculated for the previous months and by adding the price and quantity information of the current month;
2. The initial index enters an iterative algorithm that is used for further optimisation, such that the price indices calculated until the current month satisfy formulas (1)-(3).

The notation of the previous sections is used below in order to describe the method in more detail. Period 0 is used to denote the base month, which in CPI practice is December of the previous year. The two steps of the method are described for an arbitrary month t, so that published indices up to month $t-1$ are available, which are denoted as $P_0^*, P_1^*, \dots, P_{t-1}^*$, with $P_0^* = 1$. Since 1-year time windows are used for calculating price indices, information of 13 months will eventually be used (as the base month is involved as well).

Initial price index for month t

Initial price indices are calculated in two steps as well. The first step only makes use of the data up to month $t-1$, while information about the current month t is added in the second step:

i. In the first step, only information up to month $t-1$ is used in the calculation of the v_i and the price index for month t. (So, for January only information about the base month is used.) The notation $G_{0,t-1}$ is introduced to denote the union of the products sold between the base month and month $t-1$. For product $i \in G_{0,t-1}$, v_i is calculated as follows:

$$v_i = \sum_{z=0}^{t-1} \varphi_{i,z} \frac{p_{i,z}}{P_z^*}, \tag{9}$$

where $\varphi_{i,z} = \frac{q_{i,z}}{\sum_{s=0}^{t-1} q_{i,s}}$, for $z = 0, 1, \dots, t-1$. A first price index for month t is now calculated, which is denoted as $P_t^{(1)}$:

$$P_t^{(1)} = \frac{\sum_{i \in G_{0,t-1}} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_{0,t-1}} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}}. \tag{10}$$

ii. The second step takes into account all information of month t, so that new products introduced in month t are included in this step. The v_i of step (i) are updated, by including the prices and quantities of month t and price index $P_t^{(1)}$. A v_i for all products $i \in G_t$ can now be calculated:

$$(11) \qquad v_i = \sum_{z=0}^{t-1} \varphi_{i,z} \frac{p_{i,z}}{P_z^*} + \varphi_{i,t} \frac{p_{i,t}}{P_t^{(1)}},$$

where $\varphi_{i,z} = \frac{q_{i,z}}{\sum_{s=0}^{t} q_{i,s}}$, for $z = 0, 1, \ldots, t$. The price index for month t can now be updated, which is denoted as $P_t^{(2)}$:

$$(12) \qquad P_t^{(2)} = \frac{\sum_{i \in G_t} p_{i,t} q_{i,t} / \sum_{i \in G_0} p_{i,0} q_{i,0}}{\sum_{i \in G_t} v_i q_{i,t} / \sum_{i \in G_0} v_i q_{i,0}}.$$

Note that all products in month t now enter the calculation.

The iterative algorithm

The initial index method thus yields a price index $P_t^{(2)}$ for month t after step (ii). This index serves as input for a subsequent method, together with the (published) price indices $P_0^* = 1, P_1^*, \ldots, P_{t-1}^*$ for the preceding months. These price indices constitute the initial values for an iterative algorithm, which is used to find out whether the price index for the current month t can be improved.

The algorithm is very simple: every iteration step consists of updating the v_i by entering the price indices obtained so far in the right-hand side of expression (11) and by subsequently calculating updated price indices up to month t according to (12). Next, the updated price indices are compared with the price indices in the previous iteration step. A difference measure is calculated for the two price index vectors, which is compared with an arbitrarily set stopping rule. The algorithm terminates when the stopping rule is satisfied.[4]

The iterative algorithm uses the price indices $P_0^* = 1, P_1^*, \ldots, P_{t-1}^*, P_t^{(2)}$ as input, with $P_t^{(2)}$ being the result of step (ii) of the initial index method. The result of the iterative algorithm may be an improved price index P_t^* for month t. In order to obtain P_t^*, the price indices of the preceding months are re-calculated. However, the re-calculated indices will not replace the published price indices, since it is not possible to revise price indices of previous months (apart from exceptional situations). So, the only result for month t is the publication of a price index P_t^* for that month.

In figures 1 and 2, the initial index and the real time index are compared with the benchmark index. Figure 1 gives an impression of the accuracy of the initial index. In this case, the initial index is not even optimised. That is, the price indices $P_t^{(2)}$ from step (ii) of the initial index method are only shown. The iterative algorithm is not used to improve the price indices of the publication months, while the indices of the preceding months are not re-calculated. Figure 1 indicates fast convergence rates of the iterative algorithm, given the accuracy of the initial index. Numerous tests have shown that less than 10 iterations are needed in most cases, with the stop criterion set at 0.001 for the maximum absolute difference between the price index vectors of two successive iterations.

Figure 2 shows two examples from the first tests that have been done with the QU-method. In these examples, the real time indices are compared with the benchmark

[4] The algorithm is in fact a *fixed point iteration algorithm*, the "fixed point" being a vector of price indices, which is a solution of expressions (1)-(3). If a unique solution exists, the algorithm converges to that solution.

indices. The real time indices are obtained by applying the above full method. The price indices would thus be the indices as published in the CPI.

Figure 1. Benchmark indices and initial indices for menswear and ladies' clothing, based on scanner data of a department store (Feb. 2009 = 100).

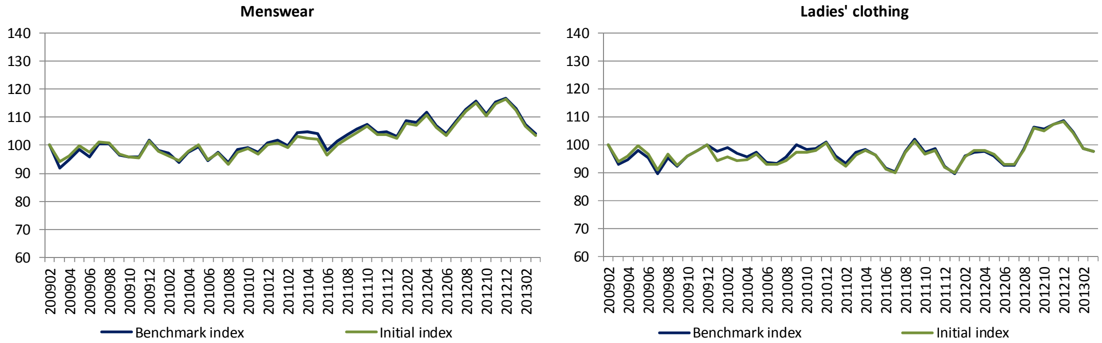
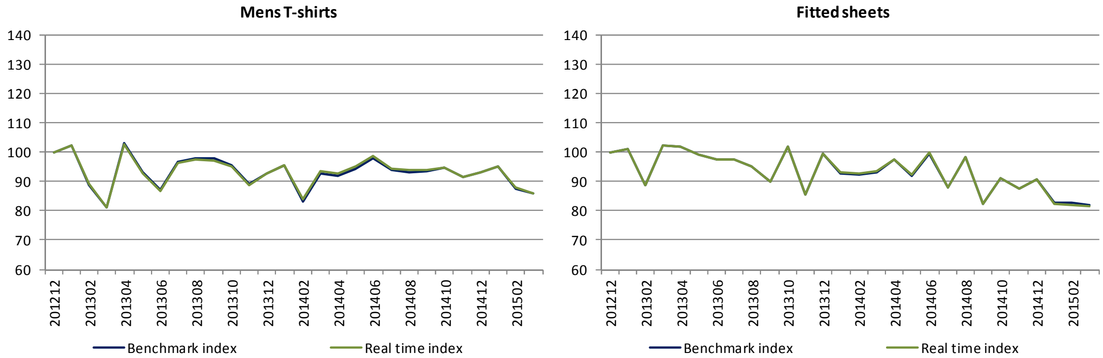


Figure 2. Real time and benchmark indices for T-shirts and fitted sheets, based on recent scanner data for the department store used in the first tests of the QU-method (Dec. 2012 = 100).



While Figure 1 makes comparisons at aggregate levels, Figure 2 shows results at a more detailed level (type of article). It can be noted that there is hardly any difference between the real time and the benchmark indices. This has been observed so far in the first results throughout the assortment of the department store. A few exceptional cases show some differences in the first few months of a year, but the two indices always move towards each other and eventually coincide, as is expected theoretically.

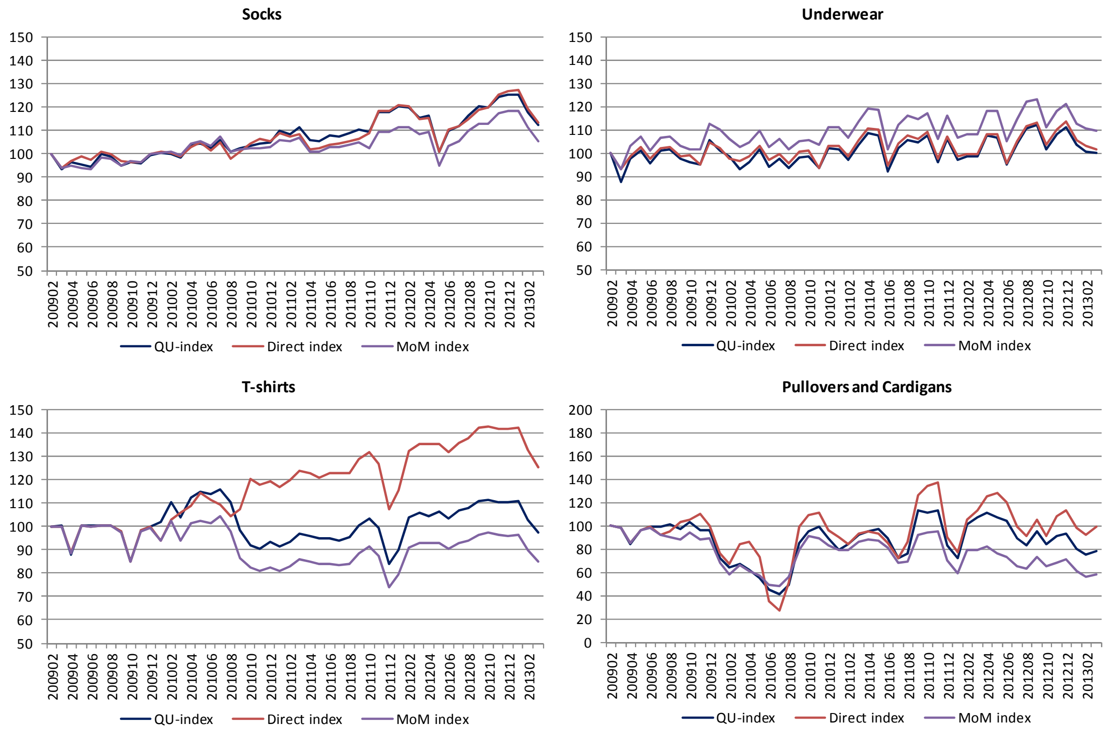
3. Comparisons with other methods

3.1 Comparisons with a monthly chained and a direct index method

In this section, the index method is compared with bilateral index (8), which is used both as a direct index and as a monthly chained index. In Section 3.2, comparisons are made with the rolling window method described in Krsinich (2014).

In Figure 3, QU-indices for four types of menswear are compared with the direct and monthly chained Lowe type price indices (8). The results show large differences, which have different causes. While the direct method and the QU-method give comparable results for socks and underwear, the direct method fails for T-shirts. The direct method does not capture the contribution of new products to price change in the year of introduction to the assortment. New types of T-shirts, made of organic cotton, were introduced in 2010 at high initial prices, which already started to decrease in 2010. The contribution of the new T-shirts is clearly captured by the QU-index, and also by the monthly chained index, but not by the direct index. The latter only evidences the price behaviour of the existing part of the assortment, which does not show a price decrease in 2010.

Figure 3. QU-indices compared with direct and monthly chained indices (MoM) for four types of menswear, based on scanner data of a department store (Feb. 2009 = 100).



The monthly chained indices do not come close to the QU-indices in none of the four cases. This can be partly explained by seasonal effects (e.g., articles returning into the assortment with price increases, which are missed). In addition, monthly chained methods do not account for the level of initial prices of new goods. This is accomplished in the QU-method by comparing the prices of new goods to their v_i, by interpreting the latter as imputed prices in the base period. Products that enter the stores at high prices, which decrease in subsequent months, tend to give a temporary upward effect on the price index (which follows by a simple rewriting of index formula (1)). In Figure 3, this is clearly illustrated by the QU-index for T-shirts in the first months of 2010.

The examples in Figure 3 show that the way in which the dynamics of an assortment are handled by different index methods may lead to completely different results. As a consequence, preference should be given to a method that is capable of dealing with assortment changes over time.

3.2 Comparisons with a rolling time window

The aim of this subsection is to compare the effects of using a rolling window for compiling price indices in the publication period to using a fixed window, that is, with a fixed base month. The method described by Krsinich (2014) uses a time window that is shifted each publication period (i.e., each quarter at Statistics New Zealand). The so-called "window splice method", which is abbreviated here to "WS-method", differs from the QU-method on two main aspects:

- The WS-method calculates a time-product dummy index ("TD-index" for short) for each shifted time window;
- A price index for the publication period is obtained by chaining the year-on-year index for the current window to the price index of one year ago (when using a 1-year time window).

Before comparing the implications of a moving window to a fixed window, the TD-index is first briefly explained. Let π_t denote the price index in period t with respect to some period 0 according to the TD-method, and let $s_{i,t}$ denote the turnover share of product i in period t. The turnover shares sum to 1 over the set of products, for every period t. The notation introduced in Section 2 will also be used below.

The expression for π_t can be written as follows:

$$\text{(13)} \qquad \pi_t = \frac{\prod_{i \in G_t} \left(\frac{p_{i,t}}{v_i}\right)^{s_{i,t}}}{\prod_{i \in G_0} \left(\frac{p_{i,0}}{v_i}\right)^{s_{i,0}}},$$

where

$$\text{(14)} \qquad v_i = \prod_{t \in T} \left(\frac{p_{i,t}}{\pi_t}\right)^{w_{i,t}},$$

and

$$\text{(15)} \qquad w_{i,t} = \frac{s_{i,t}}{\sum_{z \in T} s_{i,z}}.$$

Formulas (13)-(15) have a number of properties, which distinguish the TD-method from the QU-method of Section 2:

- If products are homogeneous, then (13) leads to an index of weighted geometric average prices instead of a unit value index;
- The weights $w_{i,t}$ of the deflated prices of product i are not well defined;
- It is not clear how to use (13) in order to derive price indices that are consistent for different time scales.

The first point needs no further mention. Regarding the second point, the weights $w_{i,t}$ of deflated prices are calculated as the share of period t in the sum of turnover shares over a time window T. However, a product's turnover is normalised per period, so that turnover shares cannot be compared in a meaningful way across periods. The common unit of measurement of the ratio scales, on which turnover is measured in each period, is lost by the normalisation, which may lead to curious and undesirable situations. For instance, according to (15) a period may receive the highest weight of all periods while having the smallest number of sales of all periods. This is counterintuitive and may lead to problems, for instance, with seasonal goods. For example, low, out-of-season prices may get a relatively large weight in the v_i. As this discussion is related to the so-called "substitution bias", it will be put aside now and resumed in Section 4.1.

Concerning the third point raised above, it is not clear how price indices, say, on a quarterly scale could be obtained from price indices on a monthly scale with the TD-method. Doing this by first calculating average quarterly prices per product, by dividing quarterly turnover by quarterly sales, and then applying formulas (13)-(15) with quarterly prices, may lead to large differences in price indices (for an impression, see de Haan and van der Grient (2011), fig. 13, p. 45). In contrast, the QU-method should be able to exploit the additivity properties of volume measures over both goods and time in order to deal with this problem.

As will be illustrated in Section 4.1, the QU- and TD-method have shown marginal differences in the price indices calculated with the scanner data considered in this and other studies (also, see Chessa (2015)). However, the above discussion shows that the two methods possess fundamentally different properties. The focus in the remainder of this section is on whether a moving window gives results that are comparable to a fixed window, with a fixed base month.

For this purpose, the TD-method is applied in both uses of the time window in order to quantify merely the difference between a fixed and a moving window. The WS-method has been applied to a large part of the scanner data of the department store. One-year time windows were shifted each quarter, as in the original method of Krsinich (2014). The results of the WS-method were then compared with those of a TD-method with a fixed base month. The results for four types of ladies' wear are shown in Figure 4. As differences with the WS-method only result after one year, the price indices for the first year are left out.

Figure 4. TD-indices for a fixed window and WS-indices for ladies' wear, based on scanner data of a department store (Feb. 2010 = 100).

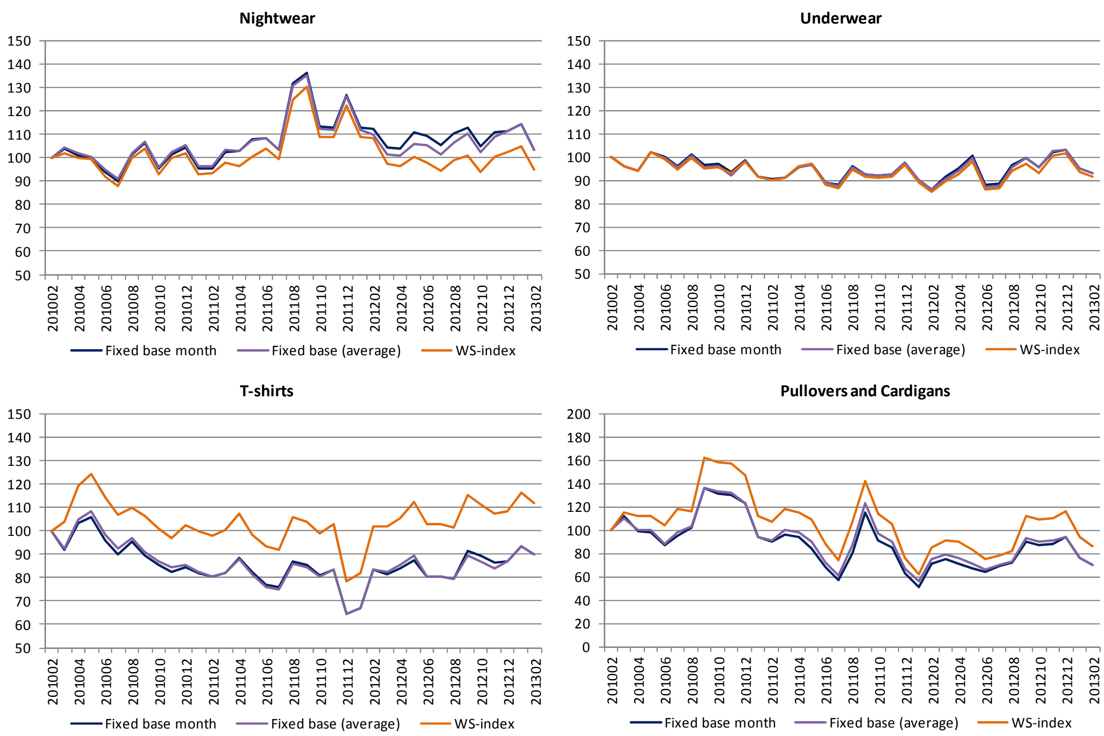


Figure 4 shows that the differences between a TD-index with a fixed window and the WS-method can become very large. Of course, price indices calculated for a fixed window are dependent on the choice of the base month. The shifted windows used in the

WS-method were therefore also used to calculate price indices for four different base months in order to quantify the effects of the choice of the base month as well. The price indices were averaged over the four base months and included in Figure 4. The results indicate that the choice of base month has a small or even negligible influence on the price indices. This supports the claim that the use of a rolling window as in the WS-method is responsible for the large differences shown in Figure 4.

The main findings after applying the WS-method can be summarised as follows:

- Price indices are continuously influenced by price and quantity information of one year ago, which enter the v_i at each shift of the time window. Such information may turn out to be dated, as will be argued below for T-shirts;
- This characteristic of the WS-method also becomes apparent in the consistently poorer weighted least squares measures, when compared with a fixed window. The model fits were even found to be poorer when ignoring the larger number of free parameters in the WS-method;
- The WS-method leads to 'chain breaks'.

Figure 4 shows large differences in 3 out of 4 article groups. The four article groups cover more than half of turnover for ladies' fashion of the department store, so that the differences should not be taken lightly also in terms of market importance. A closer look into the differences for T-shirts is now taken, which are extremely large.

The WS-index starts to rise above the TD-indices with a fixed window already in the course of 2010. At the beginning of 2010, a new assortment of organic cotton T-shirts entered the stores at high introduction prices, which promptly generated significant turnover. The high prices in the first months of 2010 still have a significant influence on the v_i in the first months of 2011, a period in which the prices have already dropped considerably. As a consequence, the WS-method assigns larger values to the v_i of organic cotton T-shirts than the TD-indices with a fixed base month tend to do. The opposite occurs with standard cotton T-shirts, which increased in price. The differences between the values of the v_i for a moving and a fixed window, combined with shifts in sales among different types of T-shirts, lead to the differences shown in Figure 4.

One of the elegant features of the WS-method turns out to be a weakness of the method as well. In general, it can be stated that the WS-indices depend more on price information of about one year ago than methods with a fixed base month. Using a fixed base month helps to keep the v_i up to date. This view is also supported by the fact that the model underlying the WS-method gives poorer fits to the scanner data of the department store. In the case of the T-shirts, the model typically fails in the first quarter of 2011, precisely because the prices of the first months of 2010 are no longer representative of how consumers value T-shirts one year later.

It should also be noted that the WS-method has been applied with a quarterly moving window. It is expected that a monthly moving window will produce even more deviating results. This is a finding that should not be underestimated in practice. For instance, the problems that arise with T-shirts could be representative for other types of goods with high introduction prices for new articles, such as electronics (Chessa, 2015).

One would expect that a rolling window approach intends to produce results that are similar to the averaged indices in Figure 4, where the base month is varied. This is not the case, so that it can be concluded that the WS-method does not serve the purpose for which it should be constructed. Both the indices with a fixed base month and the ones averaged over different base months are obtained by chaining price indices for 1-year windows. These price indices are thus 'chain preserving' by construction. The WS-method does not possess this property, which provides an alternative explanation of the drifting behaviour shown in Figure 4.

Apart from giving theoretically and statistically more justifiable and consistently better results, methods with a fixed base month are also much more transparent to CPI practitioners and, in addition, fit perfectly within current CPI routine. When using such methods, a QU-type of method is preferred to TD-methods for reasons discussed in the first part of this subsection.

4. Other issues

4.1 Substitution effects

As was stated in Section 3.2, the QU-method differs from the TD-method in several respects. Yet, both methods have given similar results for the scanner data sets considered so far. Calculations for the department store and drug stores usually show differences in the order of several tenths of a percentage point per year and sometimes differences are negligible. In Figure 5, the QU- and TD-indices are compared for the four types of ladies' fashion shown in Figure 4.

Figure 5. QU- and TD-indices for ladies' wear, based on scanner data of a department store (Feb. 2010 = 100).

The results show very small differences for three types of clothing, but are larger for T-shirts. A part of the differences between the two methods can be traced back to the use of different weighting systems for the deflated prices in expressions (2) and (14) for the v_i. As was stated previously, the use of the weights (15) in the TD-method is flawed because turnover shares over different periods are compared. This may distort the relative importance of prices over different periods. It is easy to imagine that this may typically occur with seasonal goods.

Long sleeve T-shirts are sold in smaller amounts during the summer season, and usually at lower (out-of-season) prices. The scanner data of the department store exhibit this price behaviour. The data also show that weight formula (15) tends to assign too large a weight to lower out-of-season prices than to regular winter prices, which leads to smaller values of v_i for long sleeve T-shirts than one would expect. The QU-method assigns weights to the deflated prices $p_{i,t}/P_t$ according to the quantities sold in different periods. If products are hardly sold out of season, then the prices in the corresponding months should be weighted accordingly, meaning that the v_i would practically be based on their regular prices, which is more defensible.

Speaking in more general terms, the weights (15) imply that the TD-method is insensitive to certain types of consumer behaviour. Certain forms of interaction between prices and quantities are not captured by TD-indices. An overall increase or decrease of numbers sold, which may be caused by factors such as goods being sold out of season or by sales offers, tends to leave the v_i unaffected.

The interplay between prices and quantities is incorporated in the QU-method. At the same time, one should be critical and ask whether a linear function for the v_i is appropriate, which implies perfect substitution of a good over time. This property is often mentioned in one breath with a phenomenon known as the “Gerschenkron effect” or “substitution bias” (Balk (1996), p. 214). Relative price and quantity changes tend to be negatively correlated in a consumer dominated market.

In international price comparisons, the GK-method assigns relatively high weights to the (higher) prices of larger and richer countries. Similarly, periods with high prices could be assigned large weights in the QU-method. Departures from linear function (2) can be conveniently studied through a CES-type parametrisation:

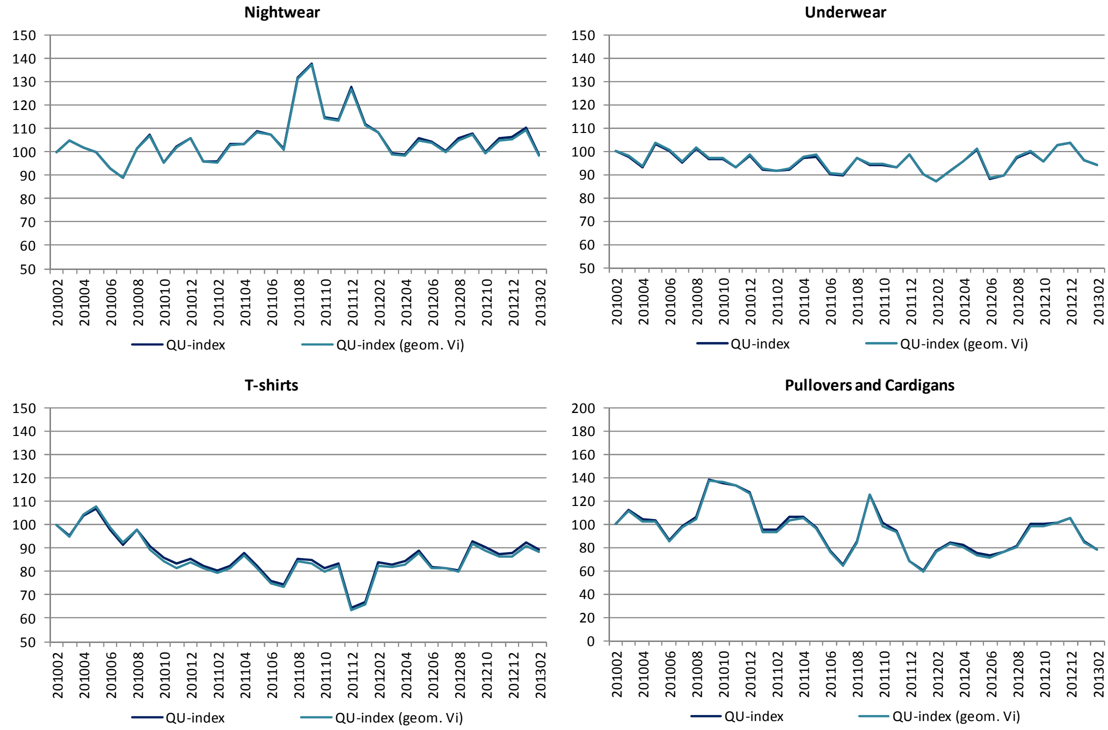
$$(16) \qquad v_i = \left(\sum_{z \in T} \varphi_{i,z} \left(\frac{p_{i,z}}{P_z} \right)^{\mu} \right)^{\frac{1}{\mu}},$$

with real-valued μ, which, initially, could be assumed to have the same value for different products. If $\mu = 1$, then expression (16) simplifies to linear function (2). Other values of μ lead to some well-known forms. If $\mu \to 0$, then (16) tends to a weighted geometric average of deflated prices (with different weights than in the TD-method, obviously). If $\mu \to -\infty$, then v_i tends to $\min\{p_{i,z}/P_z : z \in T\}$, while v_i tends to $\max\{p_{i,z}/P_z : z \in T\}$ when $\mu \to +\infty$. Using price index formula (1) in combination with (16) has the nice feature of preserving the unit value index property under product homogeneity.

Expression (16) thus allows to study a range of different product valuation functions. It would be interesting to explore to what extent the value of μ differs among different types of consumer goods. Another interesting question is whether the base QU-method given by formulas (1)-(3) can be considered a suitable approximation to an undoubtedly more complex underlying form of consumer behaviour.

As an initial exercise, the results in Figure 5 were compared with those for a QU-method with weighted geometric averages for the v_i. The results are shown in Figure 6. There are hardly any differences between the two QU-indices. In addition, the differences for T-shirts appear to be much smaller than those with the TD-index in Figure 5 and are negligible. Clearly, the different weights applied to the deflated prices, and also the fact that the TD-method does not simplify to a unit value index under product homogeneity, are at the basis of the different results in Figure 6.

Figure 6. QU-indices with weighted arithmetic and weighted geometric average v_i for ladies' wear, based on scanner data of a department store (Feb. 2010 = 100).



Also in other cases the two functions for the v_i were found to give negligible differences, such as for menswear, the cosmetics and the restaurant departments of the department store, and also in a study on mobile phones (Chessa, 2015). The first results thus appear to be robust under different forms of the v_i.

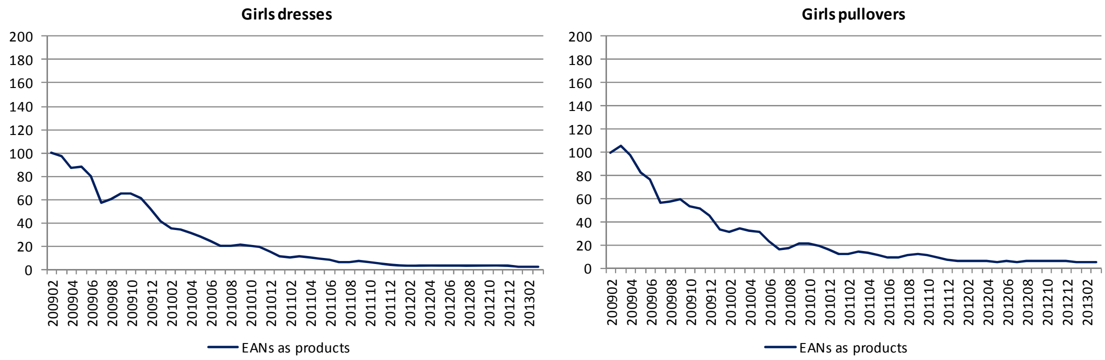
4.2 Product characterisation

An essential problem to be resolved before price indices can be actually computed is how the sets of goods G_t are composed. The problem is to search for an appropriate partition of individual articles (i.e., EANs in scanner data sets), where each element of the partition is a set of one or more EANs that are found to be 'comparable' on some grounds in order to be labelled as a "homogeneous product". These are the elements of G_t, which have to be constructed in some way.

A natural choice would be to identify each EAN as a separate homogeneous product. However, each modification of an article leads to assigning a new EAN. This also holds when only characteristics are changed without modifying the consumable part of an article. For instance, if the colour or the shape of a bottle of shampoo are slightly modified, then a new EAN is assigned. A retailer could do this from marketing considerations, for example, in order to fit an existing article within a new product line. If such a re-coding or "re-launch" of an existing article is accompanied by a higher price of the follow-up EAN, then a product differentiation at EAN-level will result in price indices that do not capture these price increases.

Such “hidden” price increases are not uncommon in the drug store world (Chessa, 2013). Fashionable and trendy consumer good sectors seem to be more sensitive to re-launches than other sectors. Another example is clothing. Figure 7 shows what may happen to price indices when each EAN is labelled as a homogeneous product. For both girls dresses and pullovers, the price index drops to almost zero after about three years.

Figure 7. QU-indices for girls dresses and pullovers for a department store, when every EAN is taken to be a product (Feb. 2009 = 100).



The assortment renewals for dresses and pullovers resemble very much the re-launch phenomenon: articles in the current assortment may drop considerably in price when the new assortment is about to be introduced in the stores. The remaining articles are sold at low prices, while the new articles are introduced at regular, higher prices. The EANs in Figure 7 have a rather short lifetime, with prices rapidly decreasing within several months. (Also see Greenlees and McClelland (2010) for similar data.)

In situations with a stable assortment, the EAN-level works fine with regard to product differentiation. But it cannot be taken as a general rule. A less detailed level of product differentiation is then needed, so the obvious question is what level to pick and on the basis of what information. Two possible routes could be thought of:

- First of all, NSI’s should ask retailers whether they have tables for linking old and follow-up EANs. Such information would be expected to be available, as retailers need to monitor and analyse a shelf’s turnover in time;
- In any case, NSI’s should ask information about article characteristics, such as brand, content and unit of measurement, size, colour, seasonality (e.g., sleeve length) and fabric.

Availability of an EAN linking table would be ideal, subject to the condition that NSI’s are aware of the rules behind the links. If retailers use their own article numbers for linking old and new EANs, then it is important to know whether, and if so, how often and in which situations article numbers are re-used.

If EAN linking tables cannot be obtained, then NSI’s should establish the links themselves through information about article characteristics. In this case, it is important to receive the information in a structured way, preferably each type of characteristic (“attribute”) in a separate field. The information in the Dutch department store scanner data is bundled in the article description, so that text mining had to be done in order to filter out the characteristics and put these in separate fields. This has turned out to be a time consuming task (Chessa, 2014).

The question then is which attributes are relevant in order to define homogeneous products. Should all available attributes be selected? Or is a limited set sufficient? This is a reformulation of the problem of setting up article descriptions, which is a standard part of traditional surveys. The problem needs to be handled efficiently in the realm of 'big data' when one aims to process such data integrally for price index calculations.

There seems to be little guidance in the literature in order to support the selection process (for some considerations on the subject, see Dalén (2014)). One approach could be to view the attribute selection problem as a statistical problem. The idea is to set up measures for assessing the fit of the price model underlying an index method, in which $v_i P_t$ is considered as the expectation of the (random) price of product i in period t. Products can be characterised as combinations of article characteristics. Different selections of attributes give rise to different sets of products and thus different numbers of parameters v_i.

The statistical problem behind the selection of article attributes can thus be stated as a problem of selecting between models with different numbers of parameters. Increasing the number of attributes leads to a more detailed product differentiation and will improve the fit of a model to the data, but may result in overfitting as parameters are continuously added. Model fit and model complexity should therefore be balanced in some way. This can be achieved by using *information criteria*, which consist of a likelihood function and a penalty term for the number of model parameters (Claeskens and Hjort, 2008).

The *Akaike Information Criterion* (AIC) and the *Bayesian Information Criterion* (BIC) are well known in the field of model selection, and therefore both measures have been applied in the present research. The AIC and BIC differ in the severity of the penalty term, with the BIC having the highest penalty for the number of parameters (for sample sizes larger than 7). Both are calculated by maximising the likelihood function for all models being compared, and the model with the highest AIC or BIC is eventually selected.

For our purposes, the BIC has proven to be more appropriate. It is calculated as follows:

$$\text{(17)} \qquad \text{BIC} = 2L_{\max} - \ln(n)\,k,$$

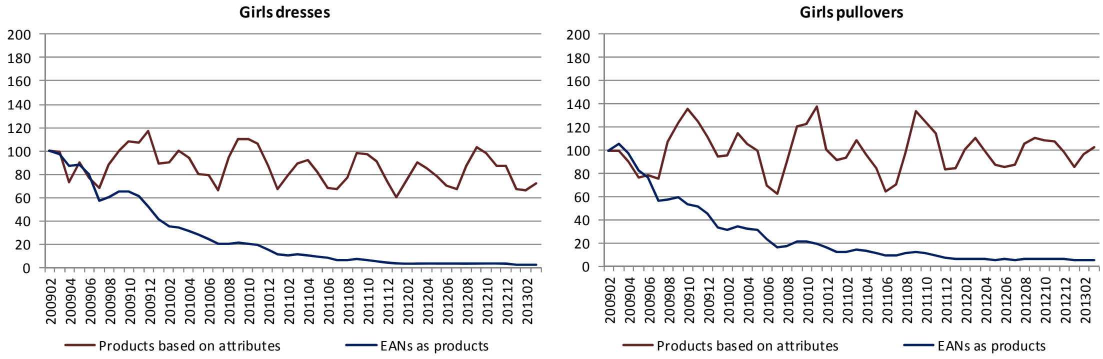
where $L_{\max}$ denotes maximum log-likelihood, n is the sample size and k the number of model parameters. The AIC has a penalty term equal to $2k$.

The AIC favours models with larger numbers of parameters, which in our studies has consistently led to considering every EAN as a separate product. This does not only occur for article groups where the assortment is stable in time, but also in situations where the article assortment is almost completely renewed each year, such as in the case of Figure 7.

Ideally, a fit measure should reject such levels of product differentiation in these cases. For this reason, information criteria with more severe penalty terms are preferred. Also the BIC does not always give plausible results. However, when applied in a real time setting it seems to give promising results, as first results suggest. In fact, an average BIC is then used, which results in a higher penalty term due to the fact that the ratio k/n tends to be larger in the initial stages of a time window.

First results indicate that this approach works well, but more tests are needed in order to gain more feeling with its performance. Figure 8 compares the price indices from Figure 7 with the price indices, in which homogeneous products for clothing are differentiated according to the following attributes, for each type of article: fabric, colour, seasonality and content (number of items in a package).[5] Product differentiation at this level gives a better (average) BIC than at EAN level.

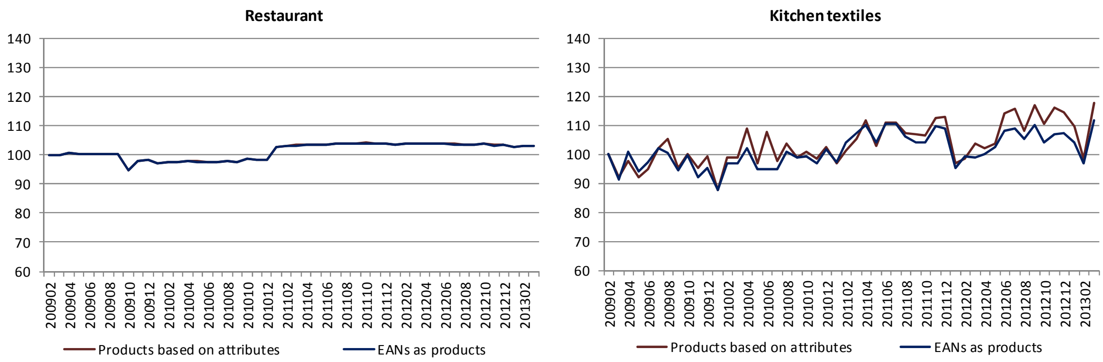
Figure 8. QU-indices for girls dresses and pullovers for a department store, for two different levels of product differentiation (Feb. 2009 = 100).



In practice, consumers pick an article from the new assortment, which they consider to be a suitable replacement when the relevant characteristics of the old and new article are the same. In the case where products are characterised in terms of a limited set of attributes, price increases associated with the introduction of a new assortment are captured.

Figure 9 shows the price indices at two levels of product differentiation for the restaurant part of the department store and for kitchen textiles. The BIC indicates that the EAN level is more suitable in both cases. The article assortments are quite stable over time, so that calculating price indices at EAN level does not lead to problems caused by re-launches. Product differentiation according to a limited set of attributes (article type, size and taste for the restaurant, and only article type and colour for kitchen textiles) gives comparable results, meaning that a limited set of attributes is capable of capturing product homogeneity quite well.

Figure 9.QU-indices for the restaurant and for kitchen textiles of the department store, for two levels of product differentiation (Feb. 2009 = 100).



To conclude, it should be emphasised that the approach described in this section is a first attempt towards a more comprehensive view on the essential problem of product homogeneity. The approach is in a test phase. It is therefore also important to stress that product characterisation should not be exclusively the result of a statistical exercise, but it should rather be the result of a combined effort, in which the roles of consumer

[5] The price indices shown previously in this paper use the same attributes for defining homogeneous products.

specialists, and also of retailers, are indispensable. Any statistical outcome should therefore be checked and discussed with information from specialists in the field.

5. Main findings and future plans

The increased use of electronic transaction data in the Dutch CPI motivated a search towards a more generic index method that could be considered for application to different data sets and types of consumer goods. The ongoing transition from traditional surveys to electronic data brings along a number of changes:

- A shift from samples of goods towards full sets of transaction data, with both prices and numbers sold;
- A shift from a fixed basket to the dynamics of an assortment in the real world;
- A different treatment of the entire process from textual and data analysis until index calculation and system maintenance.

Given these changes in the current landscape, the aim of this study was to find an index method that could process entire data sets and deal with assortment changes in time. An adaptation of the Geary-Khamis method to the time domain has been proposed, which updates the product weights in the volume index each month. The main findings with the resulting QU-method can be summarised as follows:

- The proposed real time index hardly differs from the theoretical benchmark with annually fixed weights v_i for each product;
- The iterative algorithm converges rapidly, given the accuracy of the constructed initial index;
- The method does not require data from previous years and can thus be applied from scratch, since the v_i are calculated from data of the current year;
- It is highly recommended to use the QU-method with a fixed 1-year window and a fixed base month, instead of using a rolling window like in the WS-method. The latter has shown clear signs of a high risk of drift and does not meet its purpose;
- In addition, the WS-method makes use of a TD-index, which does not simplify to the unit value index under product homogeneity, while the v_i are not well defined in terms of the weighting scheme used. The QU-method does not suffer from these shortcomings;
- The comparisons with a direct and a monthly chained index in Section 3.1 show that a method should be used that deals with the dynamics of an assortment, and does not exclude articles without clear reasons;
- Departures from the base method, with a linear function (2) for the v_i, have indicated that the results are robust when abandoning the implicit assumption of perfect substitution.

The index method should be considered as a part of an integrated framework, which combines different aspects related to price index calculation. Foremost, the problem of product characterisation needs to be incorporated. The way in which this is achieved depends on the information about articles obtained by retailers. If the data contain information that links old EANs to follow-up EANs, then such linking information could efficiently resolve the problem of product homogeneity.

If such information is not available, then NSI's have to recur to information about article characteristics. The experiments with the BIC measure in this respect look promising and have given plausible results so far. Future work with such a measure needs to be combined in some way with the expertise of consumer specialists.

The methodology has been implemented and is now in a test phase for the department store mentioned throughout this paper. The process that ranges from the linking of article characteristics to EANs until price index calculation has been automised. The experiences so far are positive: the first results are checked and found to be correct, while the linking of article characteristics to EANs operates through a short list of key words that has remained stable over time.

The QU-method is being tested by applying it on the most detailed level of article type. For instance, this means that price indices are calculated for each type of sock or pullover separately, rather than for types of socks or pullovers combined. On the one hand, the article type level is purer in a sense, but on the other hand it has a higher probability of imputing price indices. So far, the number of imputations has turned out to be small. The index method will be applied at a higher level of aggregation as well in order to compare the results.

The process cannot be fully automised. A part of monthly maintenance will be reserved for controlling new EANs on whether they contain new types of articles and attributes. The consumer specialist will deal with this and subsequently update the lists of article types and the key words with possibly new attributes. In the latter case, the relevant attributes have to be selected before the new EANs can be included in price index calculations.

The QU-method is also being tested on data of mobile phones (Chessa, 2015). The aim is to take the method into CPI production in January 2016, for both the department store and the mobile phone data. In the course of this year, the QU-method will also be tested on drug store scanner data.

First of all, these plans serve to replace the current sample-based methods for the three aforementioned data sets by the QU-method. Statistics Netherlands is also setting up a research program for the coming years, which aims at studying possibilities for further improvement of the methodology, ranging from text mining and data analysis/exploration to price index calculation. This research will be extended to other scanner data sets, amongst which data for do it yourself stores and supermarkets.

Two points for further research that arise from the present study are worth mentioning:

- Aggregation to higher product levels and over time;
- Departures from the base method, in particular concerning the implicit assumptions about substitution.

The index method proposed offers the possibility to re-think some traditional choices made in the compilation of the CPI. The additivity properties of quantity and volume measures offer an alternative for aggregating price and volume indices over sets of goods and over time. Aggregation over sets of goods means that one could simply extend the way of aggregation over products within article types to higher levels.

Concerning the second point, it would be interesting to extend the initial analysis of Section 4.1 and consider different versions of the QU-method according to expression (16) for the v_i. An interesting question could be whether the value of the parameter μ differs substantially for different types of consumer goods. Expression (16) also offers the possibility to study the robustness of the results obtained with the base method. The first signs in this respect look promising. The idea behind expression (16) could also be extended to volume measures across products, thus obtaining a nested type of CES-function.

To conclude, Statistics Netherlands is putting a lot of effort into collecting internet prices through web scraping and is planning to study possibilities of using the index method for calculating price indices from internet prices. However, considerable care is needed in using such data, as numbers of articles sold are needed for constructing weights of products. Methods that assign equal weights to articles (EANs) generally give poor statistical fits to price data and the resulting price indices may differ considerably from price indices in which articles are weighted according to turnover shares (Chessa, 2014).

Additional information is therefore needed before internet prices can be used in a meaningful way. In this respect, it could be useful to ask ourselves to what extent we can still learn from the long-term experiences built up with traditional surveys.

Acknowledgement

The author wants to thank various colleagues for their continuous support and discussions. The author is especially grateful to Professor Bert Balk and Jan Walschots for their comments on the previous version of this paper, and to Stefan Boumans for implementing the index method and for his useful suggestions during that process.

References

Balk, B.M. (1996). A comparison of ten methods for multilateral international price and volume comparison. *Journal of Official Statistics*, 12, 199-222.

Balk, B.M. (2001). Aggregation methods in international comparisons: What have we learned? Paper originally prepared for the Joint World Bank - OECD Seminar on Purchasing Power Parities, 30 January - 2 February 2001, Washington DC.

Balk, B.M. (2012). Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference. Cambridge University Press.

Chessa, A.G. (2013). Comparing scanner data and survey data for measuring price change of drugstore articles. Paper presented at the *Workshop on Scanner Data for HICP*, 26-27 September 2013, Lisbon.

Chessa, A.G. (2014). An index method for a Dutch department store. Internal report, Statistics Netherlands. (In Dutch)

Chessa, A.G. (2015). A new price index method for mobile phones based on GfK-data. Internal report, Statistics Netherlands.

Claeskens, G., Hjort, N.L. (2008). *Model Selection and Model Averaging*. Cambridge University Press, UK.

Dalén, J. (2014). The use of unit values in scanner data. Written in the framework of the network of methodological experts for HICP. Eurostat, version of September 2014. (A version of this paper was also presented at the Scanner data workshop held in Vienna, 2-3 October 2014.)

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.

Greenlees, J.S., McClelland, R. (2010). Superlative and regression-based consumer price indexes for apparel using U.S. scanner data. Paper presented at the *IARIW 31st General Conference*, 22-28 August 2010, St-Gallen, Switzerland.

van der Grient, H.A., de Haan, J. (2010). The use of supermarket scanner data in the Dutch CPI. Paper presented at the *Joint ECE/ILO Workshop on Scanner Data*, 10 May 2010, Geneva. Also available on: www.cbs.nl

de Haan, J. (2006). The re-design of the Dutch CPI. *Statistical Journal of the United Nations Economic Commission for Europe*, 23, 101-118.

de Haan, J., van der Grient, H.A. (2011). Eliminating chain drift in price indexes based on scanner data. *Journal of Econometrics*, 161, 36-46.

Haan, J. de, Krsinich, F. (2014). Time dummy hedonic and quality-adjusted unit value indexes: Do they really differ? Paper presented at the *1st Conference of the Society for Economic Measurement*, 18-20 August 2014, University of Chicago.

ILO et al. (2004). *Consumer price index manual: Theory and practice*. A publication of ILO/IMF/OECD/UNECE/Eurostat/The World Bank. Geneva, International Labour Office.

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 FEWS 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.