N° 123 ÉCONOMIE ET STATISTIQUES
Mai 2021 WORKING PAPERS DU STATEC

Botir Radjabov and Marc Ferring (STATEC)*

The Implementation of a Multilateral Price Index Method for Scanner Data in the Luxembourg CPI

Abstract

Scanner data has been introduced in the Luxembourg CPI in 2018 using the "dynamic basket" method. The drawbacks of this method lied in its inability to incorporate all available products and in its inability to directly incorporate the selected products' turnover information into price index calculations. To avoid these drawbacks, while also avoiding a possible chain drift, multilateral price index methods can be used. In this research, several selected multilateral price index methods are analysed as possible replacement methods for the "dynamic basket" method. With this in mind, the decision of "GEKS HASP" on 25 months' window setup introduction in the Luxembourg CPI from January 2021 onwards is discussed and justified.

* The authors would like to thank Dr. Claude Lamboray and Mr. Lucien May for their useful comments and suggestions.

Contents

1 Introduction

STATEC has been using scanner data from participating retailers of Luxembourg in the CPI calculations since January 2018. The method used from 2018 to 2020 was the so-called "dynamic basket" method, which was recommended by EUROSTAT. The purpose of this research is to compare the selected multilateral price index methods with the method, which was used from 2018 to 2020. It is assumed that this comparison will justify the switching from the old method to one of the selected multilateral price index methods for Luxembourg CPI production purposes.

1.1 State of play in 2020

Since January 2018, STATEC has been including scanner data from several retailers of Luxembourg into the CPI calculations. The data from these retailers covers around 5 percent of the 2020 CPI basket. The current scanner data coverage is limited to COICOP division 01 (food and non-alcoholic beverages) products, excluding products of 0101060101 ("Fresh Fruits") and of 0101070101 ("Fresh Vegetables") COICOP sub-sub classes.

A typical scanner data file has a certain structure, which is presented in the Table 1:

Table 1: Scanner Data Structure

Retailer Category Code	Retailer Category Label	EAN	Product Label	Number of Units Sold	Turnover	Year	Month
19191	Prepared sauces, vinegar and other condiments	3596710400188	Natural Ketchup	19	57	2020	06

In the table above, EAN (European Article Number) represents the number, which is contained below the bar code of every product in any retailer. It can be viewed as a unique product identifier. However, this identifier is usually too detailed since the "same" product might be sold in a retailer under different EANs. That is why, a product is identified by a combination of its label and its EAN, which is represented by Product Category Code[1]. Importantly, turnover and the number of units sold values of products indicate the values of the first 14 days of months. Once turnover values are divided by the number of units sold values, an average (unit) price of products can be obtained. This is the price which enters the respective price index calculations.

1.2 The method used from 2018 to 2020

The CPI estimation method, which was used from 2018 to 2020 for Luxembourg CPI production purposes, can be summarized as follows. Scanner data indices were calculated using the so-called "dynamic basket" method (EUROSTAT 2017), which involved a creation of a dynamic basket of products and a usage of a monthly chained Jevons index. The sampling specification needed for dynamic baskets' creation was conducted in accordance with Van der Grient and de Haan (2010). The dynamic basket was obtained by analyzing products turnovers

[1] For more details about products` classification and matching, which have not changed from the beginning of scanner data introduction, please refer to Guerreiro et al. (2018).

shares of two adjacent months. More specifically, if an average of product's turnovers shares was above a certain threshold, the product was included in the sample. This means that a product was included in the sample if:

$$\frac{s_i^t + s_i^{t-1}}{2} \geq \frac{1}{n_t * \lambda}$$

where

- s_i^t is a turnover share of product *i* in month *t*
- s_i^{t-1} is a turnover share of product *i* in month *t-1*
- n_t is a number of products available in both months *t* and *t-1*
- λ is a threshold (= 1.25)

In order to be included in the above turnover share calculations, a product firstly needed to satisfy a minimum turnover and price requirements. With this is mind, dumping filter was applied. This filter excluded products, which experienced sharp decreases in their turnovers and in their prices (more than 75 % and 70 % of previous month values respectively). This filter was implemented to avoid the issue of stock clearances, presence of which might have introduced a downward bias in price index results. Products, which showed extreme price changes (an increase of 300 % or a decrease of 75% or more of previous month values), were also excluded with a help of an outlier filter. The prices of products, which were identified to belong to either dumping prices or outliers or which were temporary missing, were treated as missing prices. For these prices, price imputations were conducted using their previous month values and price change rate of products within the same sub-sub class of the same retailer. This was done to ensure that there was a price impact to the price index once the missing products became available again.

For each retailer and sub-sub class, separate elementary indices were calculated using a Jevons formula (a geometric mean of price ratios). Since December month of the previous year was regarded as a CPI reference month, the first price comparison was implemented by comparing January prices to those of December. Long-term indices were obtained by using a chained Jevons formula calculated as follows:

$$I_{c,r}^{m,y} = \prod_{i \epsilon S_{1,c,r}} \left(\frac{p_i^{1,y}}{p_i^{12,y-1}}\right)^{1/n_1} * \prod_{i \epsilon S_{2,c,r}} \left(\frac{p_i^{2,y}}{p_i^{1,y}}\right)^{1/n_2} \dots\dots\dots \prod_{i \epsilon S_{m,c,r}} \left(\frac{p_i^{m,y}}{p_i^{m-1,y}}\right)^{1/n_m}$$

where

- $I_{c,r}^{m,y}$ is elementary aggregate price index for month *m* and year *y* for sub-sub class *c* of a retailer *r*
- n_i is a number of products in the sample $\boldsymbol{S_{i,c,r}}$

Obtained indices were then used in combination with weights, which corresponded to retailers' turnovers obtained from the structural business statistics survey, to achieve further aggregation. This was done by a Laspeyres-type price index as follows:

$$\boldsymbol{I_{c,SD}^{m,y} = \sum_{r} W_{c,r}^{y} * I_{c,r}^{m,y}}$$

where

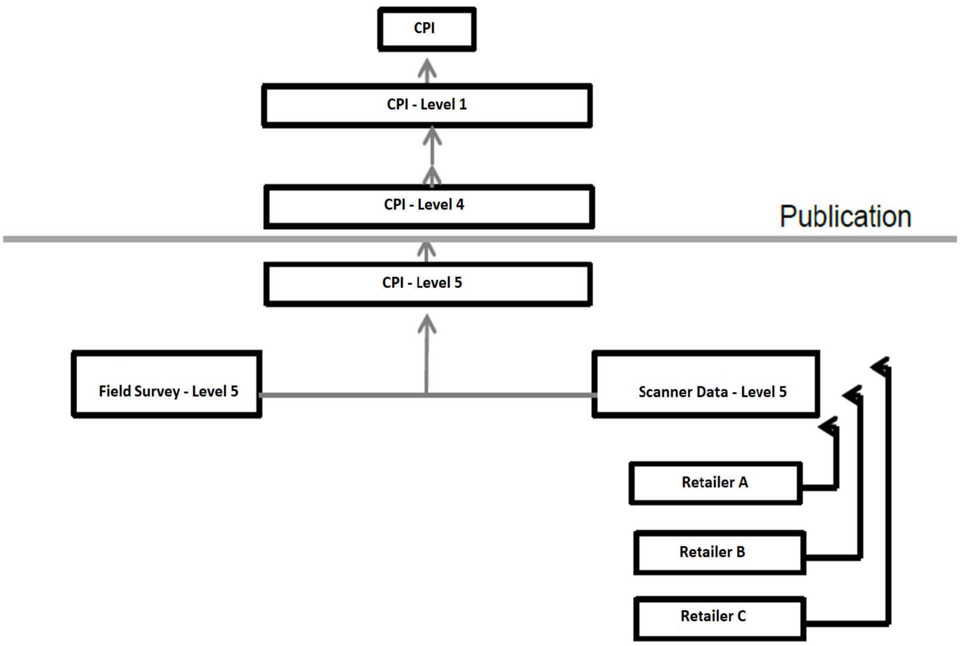
- $I_{c,SD}^{m,y}$ is a scanner data index for month *m* and year *y* for sub-sub class *c*
- $W_{c,r}^{y}$ is a relative weight of a retailer *r* during a year *y*

The resulting indices were then used in combination with analogous indices, obtained from traditional field survey methodology, and weights, to achieve even further aggregation. The weighting scheme was similar to the approach discussed for the previous aggregation level. The weight for scanner data index (W_{SD}) and the weight for field survey index (W_{FS}) corresponded to the turnover shares of the retailers covered by these two collection modes. The aggregation could be then summarized as follows:

$$\boldsymbol{I_{c}^{m,y} = W_{SD} * I_{c,SD}^{m,y} + W_{FS} * I_{c,FS}^{m,y}}$$

After this step, the aggregation to the higher level was performed by a usage of the usual CPI weights. At the higher levels, the indices were chain-linked into long-term time series of price indices (2015=100), always using December month as a linking month. The indices for level four and above were published. An overall CPI aggregation structure is presented in the Figure 1.

Figure 1: Overall CPI Aggregation Structure



2 Multilateral Price Index Methods

2.1 Coverage

One of the drawbacks of the old CPI estimation method lied in its inability to incorporate all available products into price index calculations. This is since the “dynamic basket” method used dumping and outlier filters as well as sampling for dynamic baskets’ creation. Importantly, to tackle this drawback, multilateral price index methods can be used due to the fact that these methods do not require a usage of any filters[2] or sampling for price index calculations. This is since all of the multilateral price index methods use all products’ turnover information in their calculations to automatically put more importance to price changes of products with larger turnover.

2.2 Chain drift

Another drawback of the old CPI estimation method lied in its inability to directly incorporate the selected products’ turnover information into price index calculations. That is why, alternative CPI estimation methods, such as superlative methods, could have been used instead. However, it is well-known that these latter methods’ usage in context of chained monthly price index calculations might lead to a chain drift. Chain drift occurs once chained price index, unlike its direct counterpart, is not equal to 100 once all prices of all products revert back to their original base month values. It has been shown by numerous studies that a chain drift is usually caused by activities of sales and discounts resulting in stock keeping behavior of consumers, and is usually of a downward nature (Diewert and Fox 2018).

Crucially, to tackle chain drift while still accounting for all available products’ information and maximizing the number of products’ matches in the data, multilateral price index methods can be used as well (Ivancic et al. 2011). While bilateral price index methods measure an aggregate price change between 2 months by comparing prices and quantities sold of products observed during a base and a comparison months, multilateral price index methods measure an aggregate price change between 2 months by comparing prices and quantities sold of products observed during multiple months. Noteworthy is the fact that multilateral price index methods have been applied over many years to make comparisons across space and have recently been adapted to make comparisons across time. Importantly, due to apparent advantages of multilateral price index methods discussed above, these methods possible usage is being particularly emphasized by international agencies such as by International Labor Organization, by International Monetary Fund and by EUROSTAT, which has created a Task Force dedicated to multilateral price index methods, among others during the recent years (ILO et al. 2020). Such emphasis has already triggered switches to multilateral price index methods usage for CPI production purposes in some countries such as in the Netherlands, in Australia, in Belgium and in Norway.

[2] Importantly, “GEKS HASP” on 25 months’ window setup, which is proposed for the Luxembourg CPI production purposes by this research, still uses an outlier filter and a dumping filter even if such a usage is not required. This is since an outlier filter, which compares a current month prices with an average of prices during the last available 25 months in that case, acts as a safeguard for a possible linking problem, which occurs if a product and a bundle of the same product are treated as the same product in price index calculations. This is also since a dumping filter, which compares a current month prices with an average of prices during the last available 25 months in that case as well, is necessary since GEKS multilateral price index method results are sensitive to dumping prices (Chessa et al. 2017).

2.3 The selected multilateral price index methods

A description of the selected multilateral price index methods such as Gini-Eltetö-Köves-Szulc (hereinafter GEKS), Geary-Khamis (hereinafter GK) and Weighted Time Product Dummy (hereinafter WTPD) is provided below. The reason for the selection of exactly these methods stems from their frequent usage in the context of price indices compilations worldwide.

GEKS-Törnqvist multilateral price index method

The GEKS-Törnqvist method uses all possible matching products to calculate the price index between *0* and *t* months as an unweighted geometric average of T+1 matched-model bilateral price indices P^{0l} and P^{lt} ratios, with l running through $[0, T]$, and can be defined as follows (Ivancic et al. 2011):

$$\boldsymbol{P_{GEKS}^{0,t}} = \prod_{l=0}^{T} \left(\boldsymbol{P^{0l}}/\boldsymbol{P^{tl}}\right)^{(1/T+1)} = \prod_{l=0}^{T} \left(\boldsymbol{P^{0l}P^{lt}}\right)^{(1/T+1)}$$

The bilateral price indices P^{0l} and P^{lt} are represented by Törnqvist price indices between *0* and l months and between l and t months respectively. A matched Törnqvist price index can be defined as follows:

$$\boldsymbol{P_{T}^{0,t}} = \prod_{i \in N_0 \cap N_t} \left(\frac{\boldsymbol{p_i^t}}{\boldsymbol{p_i^0}}\right)^{0.5\,(s_i^0 + s_i^t)}$$

where p_i^0 and p_i^t as well as s_i^0 and s_i^t denote prices and turnover shares of the matched products i in *0* and *t* months. The turnover shares can be calculated as $s_i^0 = \frac{p_i^0 q_i^0}{\sum_{j \in N_0 \cap N_t} p_j^0 q_j^0}$ and as $s_i^t = \frac{p_i^t q_i^t}{\sum_{j \in N_0 \cap N_t} p_j^t q_j^t}$.

GK multilateral price index method

The GK method (Chessa 2016) extensively uses unit value concept. Since aggregation of quantities sold values is cumbersome due to their non-homogeneous nature, the GK method suggests to use quality adjustment factors, v_i, to overcome this difficulty. More specifically, the quality adjustment factors make a transformation of quantity sold values of products to common units, $v_i q_i^t$, while also transforming prices of products to become quality adjusted prices, p_i^t / v_i. These transformations result in a quality adjusted unit value, $\tilde{p}^t$, in month *t* for a set of products U_t, which can be defined as follows:

$$\boldsymbol{\tilde{p}^t} = \frac{\sum_{i \in U_t} \boldsymbol{p_i^t q_i^t}}{\sum_{i \in U_t} \boldsymbol{v_i\, q_i^t}}$$

where p_i^t and q_i^t denote a price and a quantity of a product *i* in month *t*. Moreover, v_i denotes quality adjustment factor of a product *i*. With this in mind, GK price index between *0* and *t* months can be defined as follows:

$$\boldsymbol{P_{GK}^{0,t}} = \frac{\boldsymbol{\tilde{p}^t}}{\boldsymbol{\tilde{p}^0}} = \frac{\sum_{i \in U_t} \boldsymbol{p_i^t q_i^t} / \sum_{i \in U_0} \boldsymbol{p_i^0 q_i^0}}{\sum_{i \in U_t} \boldsymbol{v_i q_i^t} / \sum_{i \in U_0} \boldsymbol{v_i q_i^0}}$$

where p_i^0 and p_i^t as well as q_i^0 and q_i^t denote prices and quantities sold of a product *i* in *0* and t months. Moreover, U_0 and U_t denote set of products available in *0* and t months. Importantly, quality adjustment factors, v_i, can be defined as follows:

$$\boldsymbol{v_i = \frac{\sum_{z=0}^{T} q_i^z p_i^z / P_{GK}^{0,z}}{\sum_{z=0}^{T} q_i^z}}$$

Since GK price index is used to calculate quality adjustment factors, which are themselves used to calculate GK price index results, the last 2 equations shall be solved simultaneously, which can be done by utilizing iterative method.

WTPD multilateral price index method

The WTPD method uses a weighted least squares regression, with turnover shares of products acting as weights[3], which can be calculated as $s_i^t = p_i^t q_i^t / \sum_{i \in N_t} p_i^t q_i^t$, to estimate price indices taking into account all available products during a chosen time window. The model with *N* products within a time window of $[0, T]$ can be defined as follows (de Haan and Krsinich 2014):

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

The parameters δ^t and γ_i of the model denote time dummy and product dummy parameters. Time dummy variable, D_i^t, has a value of 1 if product i is available in month t and *0* otherwise. Product dummy variable, D_i, has a value of 1 if the observation relates to a product *i* and *0* otherwise. Due to possible multicollinearity issue, an arbitrary product *N* is excluded from the model. Conventionally, an estimated fixed effect of a product *i* is equal to $\exp(\hat{\gamma}_i)$ and an estimated WTPD price index is equal to $\exp(\delta^t)$.

2.4 Splicing methods

Since incorporation of a new month into the multilateral window may result in a revision of previously published price indices, the fact of which is not acceptable by statistical agencies, an overview of several splicing (extension) methods is also provided. To tackle a revisions problem, a rolling window approach is suggested. Rolling window approach functions in a way of shifting the estimation window (usually of 13 or 25 months) forward by one month and splicing the new price indices onto existing ones.

Movement splice

Movement splice involves a calculation of a price index for a new month t by chaining last month month-to-month price index of the shifted window to the price index of the previous month computed over the previous window. This can be expressed as follows[4] (de Haan and van der Grient, 2011):

$$\boldsymbol{P_{MS}^{0,t} = P_{MS}^{0,t-1} P_{t-T+1,t}^{t-1,t}}$$

[3] The fundamental difference between GEKS and WTPD turnover shares calculations lies in the fact that GEKS turnover shares are obtained only taking into account products, which are available in both *0* and t months, whereas WTPD turnover shares are calculated on month-by-month basis.

[4] In the following splicing methods formulae, the notation is as follows: the subscript shows the window period and the superscript shows the period for which a price index is calculated.

Window splice
Window splice involves a calculation of a price index for a new month *t* by chaining the price indices to the price index calculated 24 months ago computed over the previous window in the case if the window length consists of 25 months. This can be expressed as follows (Krsinich, 2016):

$$P_{WS}^{0,t} = P_{0,T}^{0,1} P_{1,T+1}^{1,2} \cdots P_{t-T,t}^{t-T+1,t}$$

This is equivalent to:

$$P_{WS}^{0,t} = P_{WS}^{0,t-1} \frac{P_{t-T+1,t}^{t-T+1,t}}{P_{t-T,t-1}^{t-T+1,t-1}}$$

Half splice
Half splice involves a calculation of a price index for a new month *t* by chaining at the middle of the window length. Specifically, half splice takes place at $t = \frac{T+1}{2}$ in case T is odd and at $t = \frac{T}{2}$ in case T is even. If it is assumed that window length consists of 25 months, the splicing takes place at the 13th month of the window. This can be expressed as follows (de Haan, 2015):

$$P_{HS}^{0,t} = P_{HS}^{0,t-1} \frac{P_{t-T+1,t}^{t-\frac{T+1}{2}+1,t}}{P_{t-T,t-1}^{t-\frac{T+1}{2}+1,t-1}}$$

Mean splice
Mean splice involves a calculation of a price index for a new month *t* by using a geometric mean of all possible splicing months' options. This can be expressed as follows (Diewert and Fox, 2018):

$$P_{MS}^{0,t} = P_{MS}^{0,t-1} \prod_{l=t-T+1}^{t-1} \left(\frac{P_{t-T+1,t}^{l,t}}{P_{t-T,t-1}^{l,t-1}} \right)^{\frac{1}{T-1}}$$

Splicing on published indices
Splicing on published indices is similar to traditional splicing methods described above, with the only exception that published indices and not recalculated indices are used while splicing. With this in mind, window splice on published indices (hereinafter WISP) as well as half splice on published indices (hereinafter HASP) can also be considered[5].

3 Empirical Results

Scanner data of three retailers is used to conduct tests on the above-mentioned price index and splicing methods. This scanner data contains COICOP division 01 (food and non-alcoholic beverages) products. The "dynamic basket" and the selected multilateral price index methods are applied on sub-sub classes of products on retailer specific basis to obtain the respective price indices. These indices are then aggregated based on the standard CPI weights of each sub-sub class of products to obtain single, retailer specific, time series of price indices. Finally, these time series are again aggregated based on weights of each retailer to obtain a single

[5] Importantly, it should be noted that movement splice directly uses published indices.

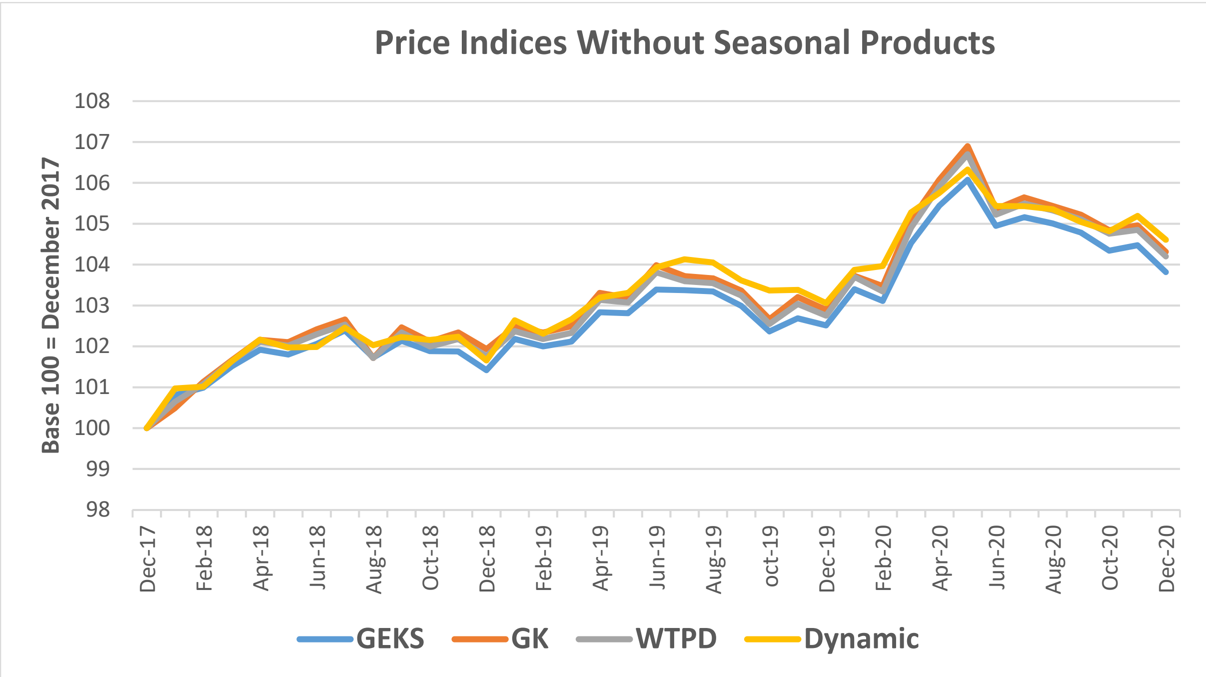
aggregated time series of price indices. All time series are starting with 100 in their first months.

3.1 Full Window Comparisons

Seasonal products (fresh fruits and fresh vegetables) are not considered at the first stage of comparisons presented below since these products were discarded once the "dynamic basket" method price indices were produced. However, these seasonal products are then added back and considered at the second stage of comparisons since it is an important goal of STATEC to include these products in Luxembourg CPI production in the near future.

The first stage comparisons of the "dynamic basket" and of the selected multilateral price index methods are conducted using the whole window of 37 months, which starts in December 2017 and ends in December 2020, excluding seasonal products. These comparisons results are presented in the Figure 2. Since full window price indices of multilateral price index methods are transitive, they can be used as benchmarks to make comparisons against price indices calculated with a help of different splicing methods. These comparisons will be shown in the next sub-section of this research.

Figure 2: The results for the "dynamic basket" and for the selected multilateral (GEKS, GK and WTPD) price index methods (full window – seasonal products excluded)



The Table 2 below presents mean, standard deviation and the end month differences among the selected multilateral and the "dynamic basket" methods' results. These results have been obtained by using full window and excluding seasonal products. The differences are calculated relatively to the "dynamic basket" method results (for instance, GEKS = GEKS - Dynamic).

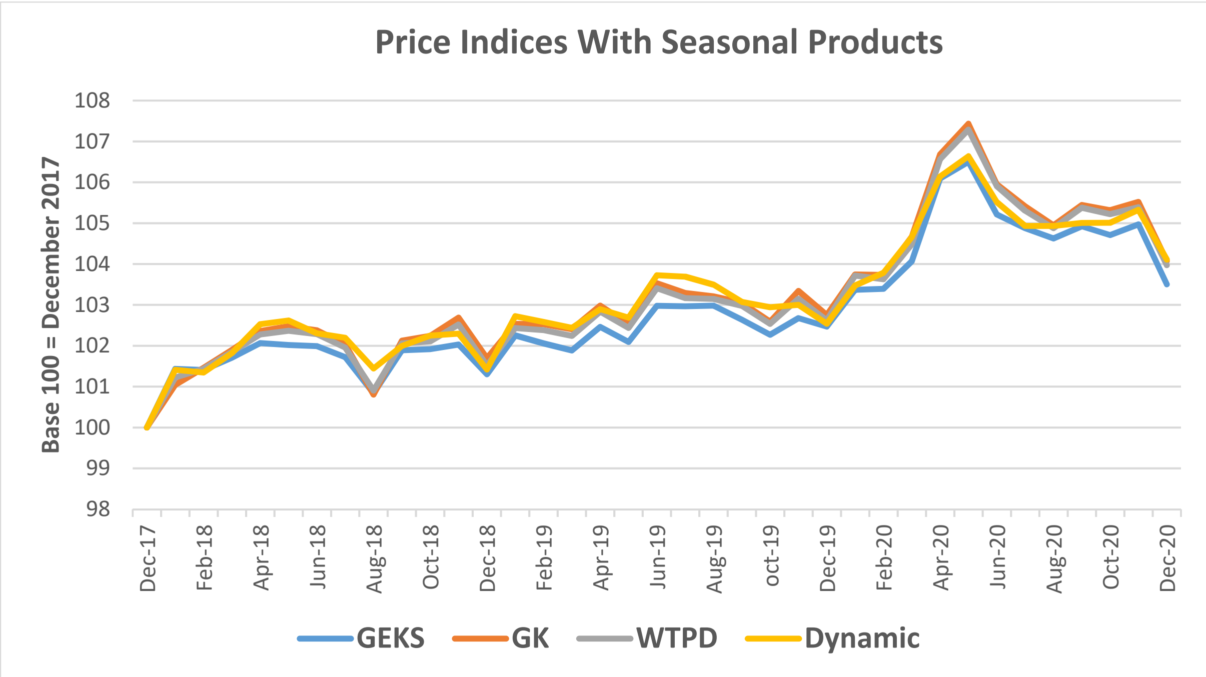
Table 2: The differences among the "dynamic basket" and the selected multilateral (GEKS, GK and WTPD) price index methods (full window – seasonal products excluded)

	METHOD	GEKS	GK	WTPD
Aggregated Results	MEAN	-0.40589	-0.04489	-0.15568
	SD	-0.11248	0.03666	0.001241
	END	-0.79381	-0.29401	-0.40872

It can be seen from the results of the Figure 2 and of the Table 2 that the aggregated results of the "dynamic basket" method and of the selected multilateral price index methods are not very different from each other.

The second stage comparisons of the "dynamic basket" and of the selected multilateral price index methods are conducted using the whole window of 37 months, which starts in December 2017 and ends in December 2020, including seasonal products. These comparisons results are presented in the Figure 3.

Figure 3: The results for the "dynamic basket" and for the selected multilateral (GEKS, GK and WTPD) price index methods (full window – seasonal products included)



The Table 3, which is analogous to the Table 2 and which differs only by an inclusion of seasonal products into results' calculations, is presented below.

Table 3: The differences among the "dynamic basket" and the selected multilateral (GEKS, GK and WTPD) price index methods (full window – seasonal products included)

	METHOD	GEKS	GK	WTPD
Aggregated Results	MEAN	-0.34183	0.005438	-0.6063
	SD	0.051969	0.173998	-0.03431
	END	-0.03717	0.148637	-0.13672

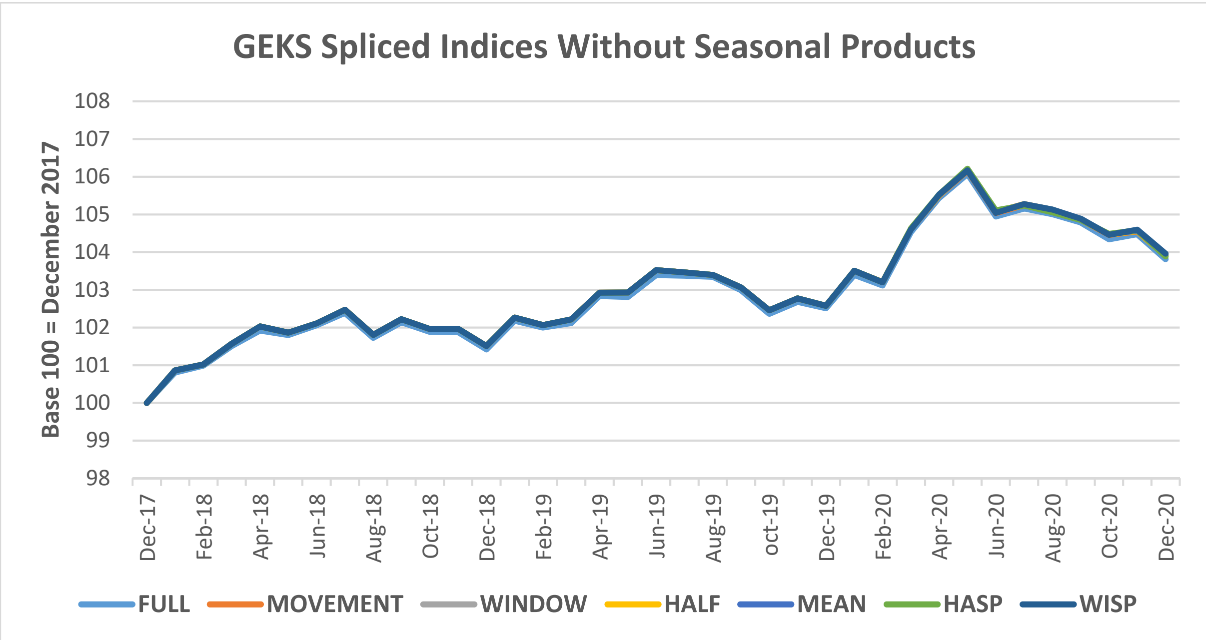
It can be seen from the results of the Figure 3 and of the Table 3 that the aggregated results of the "dynamic basket" method and of the selected multilateral price index methods are slightly more volatile than those presented in the Figure 2 and in the Table 2. However, these results are still not very different from each other.

Importantly, the Figures 2 and 3 as well as the Tables 2 and 3 results all show that using unweighted price index method (the "dynamic basket" method) seems not to bias price index results. Analogously, these results also indicate that there is a possibility to substitute the "dynamic basket" method with one of the selected multilateral price index methods without an introduction of much of an impact caused by a change in price index methods.

3.2 Splicing Methods Comparisons

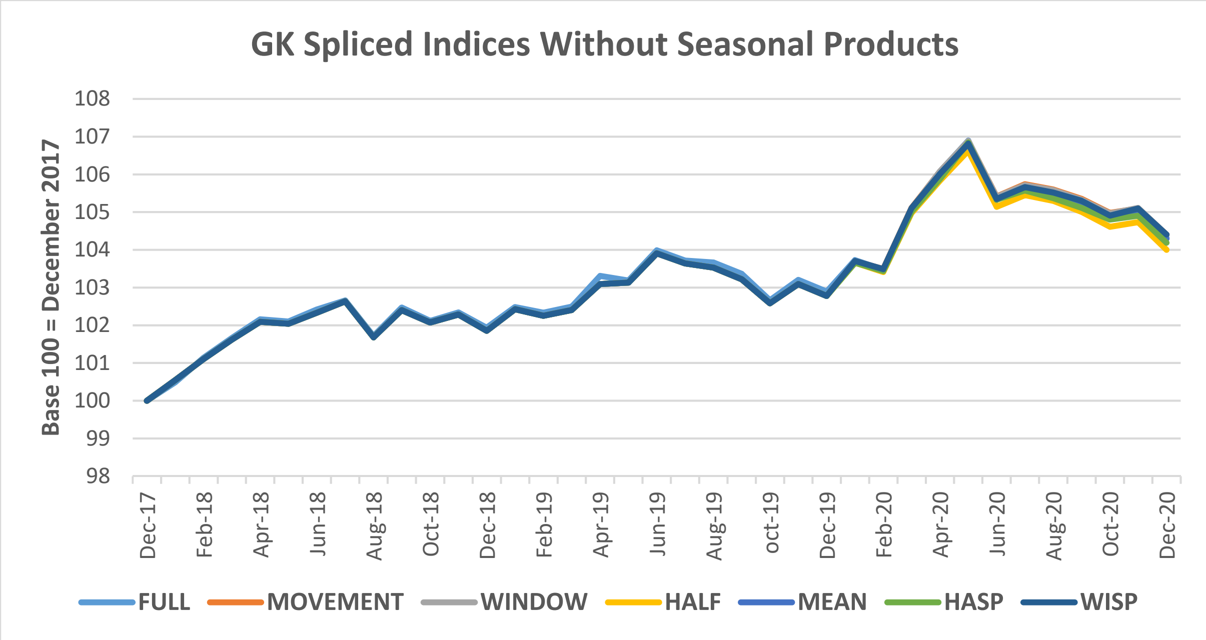
Six different splicing methods mentioned in the Section 2.4 of this research were tested for the selected multilateral price index methods[6]. The corresponding results are presented in the Figure 4 and in the Table 4 below. The differences for the Table 4 are calculated relatively to the full window price indices (for instance, GEKS (MOVEMENT) = GEKS (MOVEMENT) – GEKS (FULL)). The estimation window length has been selected to be equal to 25 months. This is due to the fact that, even though there is no wide consensus regarding an optimal estimation window length present, it seems that there is a convergence to prefer larger window lengths to smaller ones (Chessa 2019).

Figure 4: The full window and the (Movement, Window, Half, Mean, HASP and WISP) spliced indices[7] of the selected multilateral (GEKS, GK and WTPD) price index methods (seasonal products excluded)



[6] Importantly, it should be noted that splicing methods comparisons were conducted on COICOP division 01 (food and non-alcoholic beverages) products, excluding seasonal products, only since it is planned to add products of the seasonal COICOP sub-sub classes in Luxembourg CPI production starting from January 2022 onwards.

[7] Importantly, it should be noted that splicing methods begin to function only from January 2020 onwards since window length has been chosen to be equal to 25 months. That is why the spliced results of the Figure 4 begin to deviate only from that month onwards.



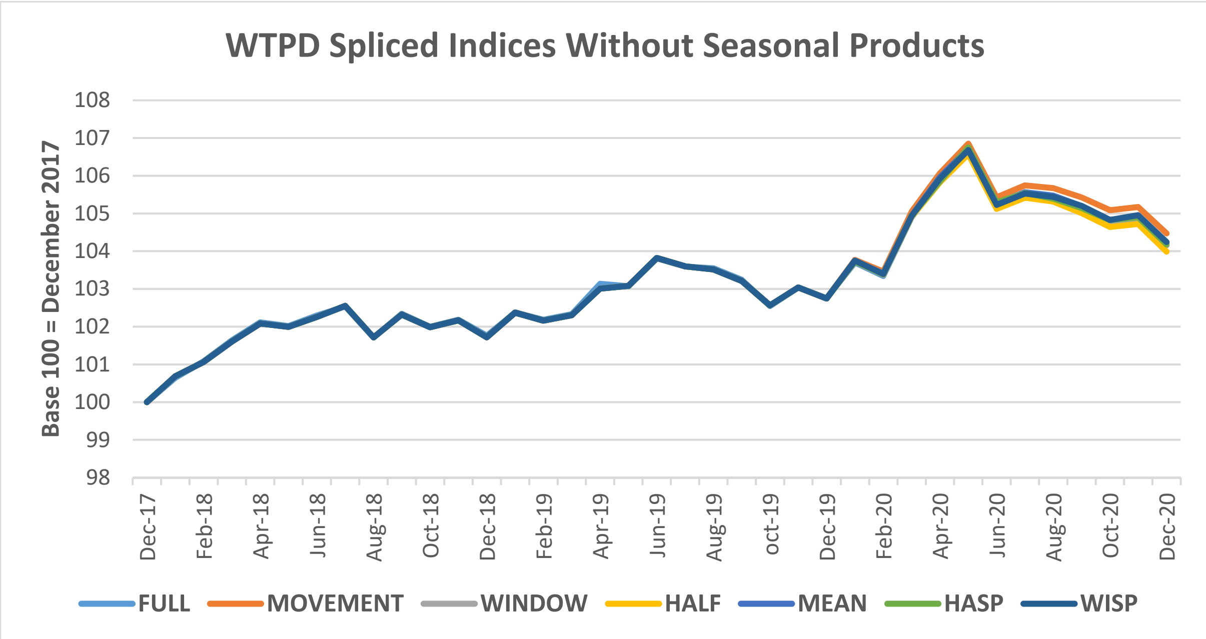


Table 4: The differences among the full window and the (Movement, Window, Half, Mean, HASP and WISP) spliced indices of the selected multilateral (GEKS, GK and WTPD) price index methods (seasonal products excluded)

	METHOD	MOVEMENT	WINDOW	HALF	MEAN	HASP	WISP
GEKS	MEAN	0.078823	0.082328	0.082878	0.084302	0.083358	0.08505
	SD	0.009499	0.014735	0.015553	0.017154	0.017205	0.017708
	END	0.094307	0.105062	0.089154	0.112737	0.0729	0.144763
GK	MEAN	-0.02728	-0.02921	-0.1107	-0.05466	-0.07388	-0.03994
	SD	0.038576	0.036085	-0.06705	0.003875	-0.01851	0.018971
	END	0.085856	0.087332	-0.31639	-0.01688	-0.12697	0.09351
WTPD	MEAN	0.062392	-0.00196	-0.0334	0.009021	0.001771	0.006408
	SD	0.093279	0.011307	-0.02916	0.025704	0.016754	0.01983
	END	0.272452	-0.04831	-0.21093	0.014514	-0.02185	0.04388

It can be seen from the results of the Figure 4 and of the Table 4 that GEKS method experiences relatively less volatility, comparing full window and spliced indices, than GK and WTPD methods. Since full window indices of multilateral price index methods are transitive, which implies that they are chain drift free by definition, the multilateral price index method, splicing indices of which are closer to full window indices, should be generally preferred.

3.3 The old and the new methods' comparisons

As it can be seen from the results of the previous subsection, GEKS method experiences relatively less volatility in terms of splicing choices in comparison with GK and WTPD methods. That is why and to produce price indices, which are consistent with the published annual rates, "GEKS HASP" on 25 months' window setup is proposed over the other multilateral price index methods setups for the Luxembourg CPI production purposes as a new method.

With this in mind, some more detailed comparison results of the old "dynamic basket" and of the new methods are presented below to justify the above proposition. For these results, there were, on average, 38 200 COICOP division 01 (food and non-alcoholic beverages) products available monthly (excluding seasonal products), from which only roughly 12 000 products were used for the "dynamic basket" method price index calculations. This was due to the fact that most of the products were excluded from these calculations due to their low turnover values. On the other hand, the "GEKS HASP" on 25 months' window method used roughly 34 000 products since it did not require a usage of sampling. This clearly shows an advantage of the new method over the old method mentioned in the Section 2.1 of this research.

Aggregated comparison results of COICOP division 01 (food and non-alcoholic beverages) products (seasonal products excluded) and of the selected COICOP sub-sub classes of the "dynamic basket" and of the "GEKS HASP" on 25 months' window price index methods are presented in the Figure 5 and in the Figure 6. Importantly, the results of the "GEKS HASP" on 25 months' window price index method presented below represent spliced indices starting from January 2020 onwards since the initial 25 months window, which starts in December 2017 and ends in December 2019, the results for which are not presented below, starts to shift from that month onwards.

Figure 5: The COICOP division 01 results for the "dynamic basket" and for the "GEKS HASP" on 25 months' window price index methods (seasonal products excluded)

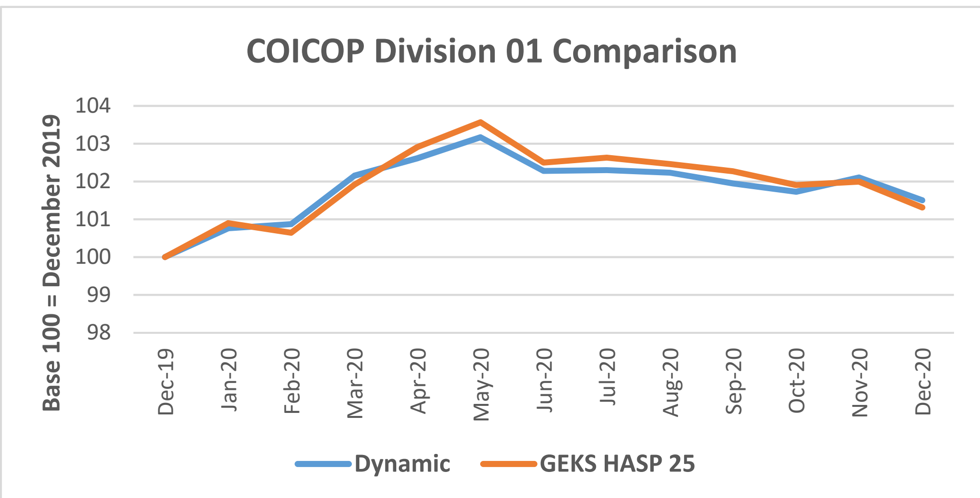
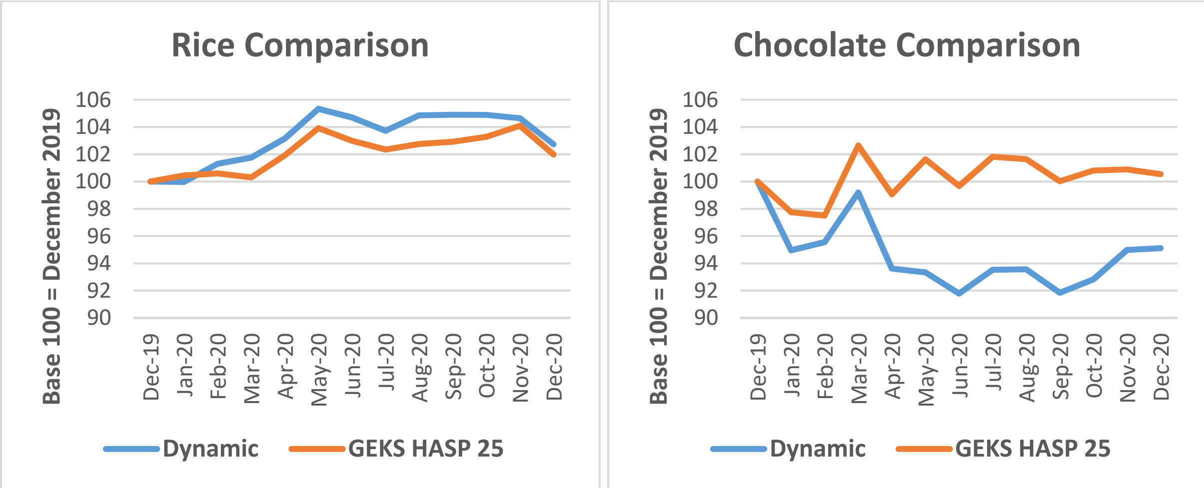


Figure 6: The “Rice” and the “Chocolate” COICOP sub-sub classes’ results for the “dynamic basket” and for the “GEKS HASP” on 25 months’ window price index methods



It can be seen from the results of the Figure 5 that the aggregated results of the old “dynamic basket” and of the new “GEKS HASP” on 25 months’ window price index methods are not very different from each other once the whole COICOP division 01 (food and non-alcoholic beverages) products, excluding seasonal products, are considered. However, it can be seen from the results of the Figure 6 that the aggregated results of the old “dynamic basket” and of the new “GEKS HASP” on 25 months’ window price index methods are becoming more volatile and are becoming sometimes significantly different from each other once COICOP sub-sub classes are considered separately. Such significant differences can usually be seen once COICOP sub-sub classes, which contain non-identifiable seasonal products, are considered. For instance, if Chocolate sub-sub class, which contains Christmas and Easter products, some of which are not identifiable as seasonal products by their labels and hence not excluded from the “dynamic basket” price index calculations, is considered, it can be seen that the aggregated results of the “dynamic basket” and of the “GEKS HASP” on 25 months’ window price index methods are significantly different from each other. Importantly, these non-identifiable seasonal products are driving the “dynamic basket” price index results down in January and in May months since these products experience sales after Christmas and Easter, but not driving the “GEKS HASP” on 25 months’ window price index results down since that method accounts for both price and turnover information of products unlike the old method, which accounts only for price information of products. With this in mind, it can be claimed that the new “GEKS HASP” on 25 months’ window price index method is more suitable for handling seasonal products than the old “dynamic basket” price index method.

4 Conclusion

As it can be seen from the results of this research, the old “dynamic basket” method had several drawbacks. These drawbacks were an inability of this method to incorporate all available products and the selected products’ turnover information into price index calculations. Importantly, this research shows that these drawbacks can be eliminated by a usage of multilateral price index methods. For instance, coverage of products, which enter price index calculations, can be increased from roughly 12 000 to roughly 34 000 monthly, if a multilateral price index method is used instead of the “dynamic basket” method.

As it can be also seen from the results of this research, GEKS method generally provides more favorable results in terms of splicing methods' usage comparing to GK and WTPD methods. This might serve as a first reason as of why GEKS method should be used for Luxembourg CPI production purposes. It should be also noted that GEKS method, rather than GK or WTPD methods, is generally consistent with the so-called economic approach to index number theory (Diewert and Fox 2018), which suggests that consumers' utility functions are generally not linear. This might serve as a second reason as of why GEKS method should be used for Luxembourg CPI production purposes.

The results of this research also show that the decision regarding a preferred splicing method to use for Luxembourg CPI production purposes cannot be easily made due to the fact that all splicing methods considered provide similar results. However, WISP and HASP splicing methods can be regarded as the most favorable candidates to be chosen as splicing methods for Luxembourg CPI production purposes. Crucially, HASP splicing method conducted on 25 months' window has an additional advantage in favor to be used for Luxembourg CPI production purposes since it provides price indices, which are consistent with the published annual rates. With this in mind, the new "GEKS HASP" on 25 months' window setup is used for Luxembourg CPI production purposes from January 2021 onwards for COICOP division 01 (food and non-alcoholic beverages) products, excluding products of 0101060101 ("Fresh Fruits") and of 0101070101 ("Fresh Vegetables") COICOP sub-sub classes. It is planned to add products of these sub-sub classes in Luxembourg CPI production from January 2022 onwards.

References

Chessa, A. (2019) A comparison of index extension methods for multilateral methods.

Chessa, A. G., Verburg, J., and Willenborg, L. (2017) A comparison of price index methods for scanner data. In 15th Meeting of the Ottawa Group on Price Indices

Chessa, A. (2016) A new methodology for processing scanner data in the Dutch CPI. - Eurostat review of National Accounts and Macroeconomic Indicators, 1/2016, 49-69.

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

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

de Haan J. and Krsinich, F. (2014) Scanner Data and the Treatment of Quality Change in Non-Revisable Price Indexes. - Journal of Business and Economic Statistics 32(3), 341-358.

Diewert, W. E. and Fox, K. J. (2018) Substitution bias in multilateral methods for CPI construction using scanner data. UNSW Business School Research Paper, (2018-13).

EUROSTAT (2017) Practical Guide for Processing Supermarket Scanner Data.

Guerreiro, V., Walzer, M. and Lamboray, C. (2018) The use of supermarket scanner data in the Luxembourg consumer price index. Economie et Statistiques.

ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2020) Consumer Price Index Manual: Theory and Practice. ILO Publications, Geneva.

Ivancic L., Diewert, E. W. and Fox, K.J. (2011) Scanner data, time aggregation and the construction of price indexes. - Journal of Econometrics 161, 24-35.

Krsinich F. (2016) The FEWS Index: Fixed Effects with a Window Splice. - Journal of Official Statistics 32 (2), 375-404.

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

van der Grient, H. and 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