The general class of multilateral indices and its two special cases

Jacek Białek*
Department of Statistical Methods, University of Lodz
Department of Trade and Services, Statistics Poland

17th Meeting of the Ottawa Group, Rome, 7 - 10 June, 2022

April 22, 2022

Abstract

Scanner data can be obtained from a wide variety of retailers (supermarkets, home electronics, Internet shops, etc.) and provide information at the level of the barcode, i.e. the Global Trade Item Number (GTIN) or its European version: European Article Number (EAN). One of advantages of using scanner data in the Consumer Price Index (CPI) measurement is the fact that they contain complete transaction information, i.e. prices and quantities for every sold item. One of new challenges connected with scanner data is the choice of the index formula which should be able to reduce the chain drift bias and the substitution bias. Multilateral index methods seem to be the best choice in the case of dynamic scanner data sets. These indices work on a whole time window and are transitive, which is a key property in eliminating the chain drift effect. Following the so-called identity test, however, one may expect that even when only prices return to their original values, the index becomes one. Unfortunately, the commonly used multilateral indices (GEKS, CCDI, GK, TPD, TDH) do not meet the identity test. The paper discusses the proposal of two multilateral indices, the idea of which resembles the GEKS index, but which meet the identity test and most of other tests. In an empirical and simulation study, these indices are compared, inter alia, with the SPQ index, which is relatively new and also meets the identity test. Analytical considerations as well as empirical studies confirm the high usefulness of the proposed indices.

JEL codes: C43, C63, C81, D12
Keywords: inflation measurement, Consumer Price Index (CPI), scanner data, multilateral indices, identity test

*This publication is financed by the *National Science Centre in Poland* (grant no. 2017/25/B/HS4/00387).

1 Introduction

Scanner data have numerous advantages compared to traditional survey data collection because such data sets are much bigger than traditional ones and they contain complete transaction information, i.e. information about prices and quantities. Scanner data contain expenditure information at the item level (i.e. at the retailer's code or the GTIN/EAN/SKU barcode level), which makes it possible to use expenditure shares of items as weights for calculating price indices at the lowest (elementary) level of data aggregation. Most statistical agencies use bilateral index numbers in the CPI measurement, i.e. they use indices which compare prices and quantities of a group of commodities from the current period with the corresponding prices and quantities from a base (fixed) period. A multilateral index is compiled over a given time window composed of $T + 1$ successive months (typically $T = 12$). Multilateral price indices take as input all prices and quantities of the previously defined individual products, which are available in a given time window, i.e. in at least two of its periods. These methods are a very good choice in the case of dynamic scanner data, where we observe a large rotation of products and strong seasonality (Chessa et al. 2017). Moreover, multilateral indices are transitive (see Appendix A), which means in practice that the calculation of the price dynamics for any two moments in the time window does not depend on the choice of the base period. By definition, transitivity eliminates the chain drift problem which may occur while using scanner data. The chain drift can be formalized in terms of the violation of the multi period identity test. According to this test, one can expect that when all prices and quantities in a current period revert back to their values from the base period, then the index should indicate no price change and it equals one. Thus, multilateral indices are free from the chain drift within a given estimation time window $[0, T]$. Although Ivancic et al. (2011) have suggested that the use of multilateral indices in the scanner data case can solve the chain drift problem, most statistical agencies using scanner data still make use of the monthly chained Jevons index (Chessa et al. 2017).

The Jevons index Jevons (1865) is an unweighted bilateral formula and it is used at the elementary aggregation level in the traditional data collection. As the scanner data provide information on consumption, it seems more appropriate to use weighted indices. Unfortunately, bilateral weighted formulas do not take into account all information from the time window, while the frequently chained weighted indices (even superlative) may generate chain drift bias (Chessa 2015) and therefore do not reflect a reasonable price change over longer time intervals. For this reason, many countries have experimented with multilateral indices or even implemented them for the regular production of price indices (Krsinich (2014), Inklaar & Diewert (2016), Chessa et al. (2017), Chessa (2019), Diewert

& Fox (2018), de Haan et al. (2021)).

Following the so-called identity test (International Labour Office 2004, von der Lippe 2007), however, one may expect that even when only prices return to their original values and quantities do not, the index becomes one. This test is quite restrictive for multilateral indices and causes some controversy among price statisticians. Nevertheless, it is mentioned among the axioms regarding multilateral indices both in the publications of the European Commission and in journals from the area of official statistics (Zhang et al. 2019). Unfortunately, the commonly used multilateral indices (GEKS, CCDI, GK, TPD, TDH) do not meet the identity test. The main aim of the paper is to present and discuss the proposition of two multilateral indices, the idea of which resembles the GEKS index, but which meet the identity test and most of other axioms. The proposed indices are compared with the multilateral SPQ index method, which is relatively new and also meets the identity test. Analytical considerations as well as empirical studies confirm the high usefulness of the proposed indices. We also compare how time-consuming all considered index methods are.

2 The list of considered multilateral price index methods

Multilateral index methods originate in comparisons of price levels across countries or regions. Commonly known methods include the GEKS method (Gini 1931, Eltetö & Köves 1964), the Geary-Khamis (GK) method (Geary 1958, Khamis 1972), the CCDI method (Caves et al. 1982) or the Time Product Dummy Methods (de Haan & Krsinich 2018). These indices work on the defined time window $[0, T]$. The idea of the SPQ multilateral price index is based on the relative price and quantity dissimilarity measure Δ_{SPQ} (Diewert 2020). The price dissimilarity measure is used to link together the bilateral Fisher indices according to the special algorithm, which extends the considered time window in each step.

Before we present the proposed multilateral price indices, let us denote sets of homogeneous products belonging to the same product group in months 0 and t by G_0 and G_t respectively, and let $G_{0,t}$ denote a set of matched products in both moments 0 and t. Although, in general, the item universe may be very dynamic in the scanner data case, we assume that there exits at least one product being available during the whole time interval $[0, T]$. Let p_i^τ and q_i^τ denote the price and quantity of the i-th product at time τ and $N_{0,t} = \operatorname{card} G_{0,t}$.

Since the indices proposed in the work are based on the idea of the GEKS index, let us recall its structure (see Section 2.1).

2.1 The GEKS method

Let us consider a time interval $[0, T]$ of observations of prices and quantities that will be used for constructing the GEKS index. The GEKS price index between months 0 and t is an unweighted geometric mean of $T + 1$ ratios of bilateral price indices $P^{\tau,t}$ and $P^{\tau,0}$, which are based on the same price index formula. The bilateral price index formula should satisfy the time reversal test, i.e. it should satisfy the condition $P^{a,b} \cdot P^{b,a} = 1$. Typically, the GEKS method uses the superlative Fisher (1922) price index, resulting in the following formula:

$$P_{GEKS}^{0,t} = \prod_{\tau=0}^{T} \left(P_F^{0,\tau} P_F^{\tau,t} \right)^{\frac{1}{T+1}} . \tag{1}$$

Please note that de Haan & van der Grient (2011) suggested that the Törnqvist price index formula (Törnqvist 1936) could be used instead of the Fisher price index in the Gini methodology. Following the article by Diewert & Fox (2018), the multilateral price comparison method involving the GEKS method based on the Törnqvist price index is called the CCDI method.

3 Axiomatic approach in the multilateral method selection

According to the axiomatic approach, desirable index properties (the so called "tests") are defined that a multilateral index may, or may not satisfy. The list of tests for multilateral indices can be found in the guide provided by the Australian Bureau of Statistics (2016) (see the chapter entitled: "CRITERIA FOR ASSESSING MULTILATERAL METHODS"). Interesting considerations concerning tests for price indices in the case of dynamic scanner data sets can be found in Zhang et al. (2019), where the authors - on the basis of the COLI (Cost of Living Index) and COGI (Cost of Goods Index) concepts - focus on five main test for a dynamic item universe (*identity test, fixed basket test, upper bound test, lower bound test* and *responsiveness test*).

Following the guidelines from the Australian Bureau of Statistics (2016) or the paper by Diewert (2020), we consider a wide set of tests for multilateral indices (see **Appendix A**) assuming that the conditions for their use are met (e.g. a set of matched products over a period of time is never empty).

Please note that the discussed multilateral index formulas (GK, GEKS, CCDI, TPD) meet most of the requirements at the same time, such as the *transitivity, multi-period identity test, positivity and continuity, proportionality, homogeneity in prices, commensurability, symmetry in the treatment of time periods* or *symmetry in treatment of products* tests. However, the discussed indexes differ in terms of the total set of tests they meet. For instance: the GEKS, CCDI and TPD indices do not satisfy the *basket test*, the Geary-Khamis and TPD indices do not satisfy the *responsiveness test to imputed prices* while the GEKS or CCDI can incorporate the imputed prices of missing products, and the *homogeneity in quantities* does not hold in the case of the Geary-Khamis formula. Please also note that the SPQ index is the only multilateral index that satisfies the *identity test*, which is a stronger requirement than the lack of chain drift.

4 Proposition of the general class of semi-GEKS indices

In the "classical" approach to constructing the GEKS-type indices, the bilateral price index formula, which is used in the GEKS' body, is the superlative one. In other words, although the standard GEKS method uses the Fisher indices as inputs (Chessa et al. 2017), other superlative indices are possible choices as well, e.g. the Törnqvist or Walsh indices (van Loon & Roels 2018, Diewert & Fox 2018). Moreover, in the paper by Chessa et al. (2017), we can read that "the bilateral indices should satisfy the time reversal test". The choice of the superlative indices as an input for GEKS has its justification in the economic approach, since the superlative indices are considered to be the best proxies for the Cost of Living Index (International Labour Office 2004). However, please note that the concept of multilateral indices is not based on the COLI framework and requirements for multilateral methods differ from those dedicated to bilateral ones. The *time reversibility* requirement, which allows the GEKS index to be transitive, enables expressing the GEKS index in a more intuitive, quotient form:

$$P_{GEKS}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{P^{\tau,t}}{P^{\tau,0}} \right)^{\frac{1}{T+1}}. \tag{2}$$

where $P^{\tau,s}$ is the choosen bilateral price index formula (for $s = 0, t$).

Due to the above-presented remarks, in this paper we propose a general class of indices based on the idea of the GEKS method, where the base bilateral index breaks all "classical" assumptions: a) it is not superlative; b) it fails the *time reversal test*; c) it uses quantities only from one of two compared periods. Due to the c point, we will call it *the general class of semi-GEKS indices*, and it will be denoted here by $GS-GEKS$ (the General Semi-GEKS). Our proposal assumes that the formula $P^{\tau,s}$, which compares the current period s with the base period τ, can be written in the following form:

$$P^{\tau,s} = f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s}) \tag{3}$$

where a function $f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$ takes into account products from the $G_{\tau,s}$ set.

We have a list of minimal requirements for the function $f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$: **R1)** It must be a positive and continuous function of its arguments and $f_{G_{\tau}}(q^{\tau}, p^{\tau}, p^{\tau}) = 1$; **R2)** The *proportionality* in current prices and *inverse proportionality* in base prices must hold (*homogeneity* of degree +1 in current prices and homogeneity of degree -1 in base prices), i.e. we expect that $f_{G_{\tau,s}}(q^{\tau}, mp^{\tau}, kp^{s}) = \frac{k}{m} f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$; **R3)** The only possible reaction to identical quantity changes is no reaction, i.e. $f_{G_{\tau,s}}(kq^{\tau}, p^{\tau}, p^{s}) = f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$; **R4)** For any diagonal matrix $\mathbf{L} = diag(\lambda_1, \lambda_2, ..., \lambda_{N_{\tau,s}})$, it must hold that $f_{G_{\tau,s}}(\mathbf{L^{-1}}q^{\tau}, \mathbf{L}p^{\tau}, \mathbf{L}p^{s}) = f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$; **R5)** For two different data sets $G^{*}_{\tau,s}$ and $G^{**}_{\tau,s}$ being subsets of $G_{\tau,s}$, where one is contained in the other (e.g. $G^{*}_{\tau,s} \subset G^{**}_{\tau,s}$), we obtain, in general, two different function values, i.e. $f_{G^{*}_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s}) \neq f_{G^{**}_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$. Please note that conditions **R1, R2, R4** make the formula $f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$ satisfy all axioms from the system of minimum requirements of price index by Martini (1992) (*identity, commensurability, linear homogeneity*).

The requirements **R1-R5** are fundamental and they will be justified with respect to good axiomatic properties of the multilateral $GS-GEKS$ method in the next part of the paper (see Theorem 1). We will also show (see Section 5) that the proposed general class of semi-GEKS indices is not empty and, in fact, it includes a large number of potential price index formulas. We will also consider the stronger version of the **R5** condition (not obligatory), which can be written as: **R6)** For two different data sets $G^{*}_{\tau,s}$ and $G^{**}_{\tau,s}$ such as $G^{*}_{\tau,s} \subset G^{**}_{\tau,s} \subset G_{\tau,s}$, we obtain $f_{G^{*}_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s}) \geq f_{G^{**}_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s})$. Finally, let us also take into consideration an additional requirement of *monotonicity*: **R7)** If $p^{s}_{i} \leq p^{s*}_{i}$ for any $i-$th product from $G_{\tau,s}$, then it holds that $f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s}) \leq f_{G_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s*})$. As it will be shown (see Theorem 1), the requirements **R6** and **R7** are crucial with respect to the lower and upper bound tests.

Taking bilateral index formula (3) as an input in the GEKS body (2) we obtain the following form of the proposed general class of semi-GEKS indices:

$$P^{0,t}_{GS-GEKS} = \prod_{\tau=0}^{T} \left(\frac{f_{G_{\tau,t}}(q^{\tau}, p^{\tau}, p^{t})}{f_{G_{\tau,0}}(q^{\tau}, p^{\tau}, p^{0})}\right)^{\frac{1}{T+1}}. \tag{4}$$

The following theorem can be proved (see **Appendix B**)

Theorem 1 *Under restrictions R1-R5 each GS-GEKS index (4) satisfies the transitivity, identity, multi period identity, responsiveness, continuity, positivity and normalization, commensurability, price proportionality, homogeneity in prices and homogeneity in quantitites tests. If the requirements R6 and R7 are additionally fulfilled, this index also satisfies the lower and upper bound tests.*

5 Special cases of the class of GS-GEKS indices

This section presents propositions of two new multilateral price indices being special cases of the GS-GEKS class of indices.

5.1 Proposition based on the Laspeyres formula

Let us define the function $f_{G_{\tau,s}}(q^\tau, p^\tau, p^s)$ introduced in Section 4 as follows:

$$f^L_{G_{\tau,s}}(q^\tau, p^\tau, p^s) = \frac{\sum_{i \in G_{\tau,s}} q_i^\tau p_i^s}{\sum_{i \in G_{\tau,s}} q_i^\tau p_i^\tau}, \tag{5}$$

where the "L" subscript refers to the Laspeyres formula. Putting (5) in formula (4), we obtain

$$P^{0,t}_{GEKS-L} = \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^t}{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^\tau}}{\frac{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^0}{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^\tau}} \right)^{\frac{1}{T+1}}. \tag{6}$$

It is easy to verify that the function $f^L_{G_{\tau,s}}(q^\tau, p^\tau, p^s)$ satisfies the requirements R1-R5 and R7 described in Section 4 (the proof is omitted). As a consequence, the proposed formula (6) realizes the thesis of Theorem 1. However, we still do not know under what circumstances the condition R6 possibly holds (if any). In the further part of the work (see Section 5.3), we will make an additional assumption that will allow us to satisfy the requirement R6 and consequently the *upper* and *lower bound test*.

Please note that the GEKS-L index can be treated as the generalization of the Fisher price index formula ($P_F^{0,t}$) to the multi-period case. In fact, in a static item universe G observed over the two period time interval $[0, 1]$, we obtain

$$P^{0,1}_{GEKS-L} = \prod_{\tau=0}^{1} \left(\frac{\sum_{i \in G} q_i^\tau p_i^1}{\sum_{i \in G} q_i^\tau p_i^0} \right)^{\frac{1}{1+1}} = \left(\frac{\sum_{i \in G} q_i^0 p_i^1}{\sum_{i \in G} q_i^0 p_i^0} \times \frac{\sum_{i \in G} q_i^1 p_i^1}{\sum_{i \in G} q_i^1 p_i^0} \right)^{\frac{1}{2}} = P_F^{0,1}, \tag{7}$$

since $G_0 = G_1 = G_{0,1} = G$.

5.2 Proposition based on the qeometric Laspeyres formula

Let us define the function $f_{G_{\tau,s}}(q^\tau, p^\tau, p^s)$ introduced in Section 5 as follows:

$$f^{GL}_{G_{\tau,s}}(q^\tau, p^\tau, p^s) = \prod_{i \in G_{\tau,s}} \left(\frac{p_i^s}{p_i^\tau} \right)^{w_i^{\tau,s}(\tau)} \tag{8}$$

where

$$w_i^{\tau,s}(\tau) = \frac{q_i^\tau p_i^\tau}{\sum_{k \in G_{\tau,s}} q_k^\tau p_k^\tau}, \tag{9}$$

and the "GL" subscript refers to the geometric Laspeyres formula (von der Lippe 2007). Putting (8) in the formula (4), we obtain

$$P_{GEKS-GL}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{\prod_{i \in G_{\tau,t}} (\frac{p_i^t}{p_i^\tau})^{w_i^{\tau,t}(\tau)}}{\prod_{i \in G_{\tau,0}} (\frac{p_i^0}{p_i^\tau})^{w_i^{\tau,0}(\tau)}} \right)^{\frac{1}{T+1}}. \tag{10}$$

It is easy to verify that the function $f_{G_{\tau,s}}^{GL}(q^\tau, p^\tau, p^s)$ satisfies the requirements R1-R5 and R7 described in Section 4 (the proof is omitted). As a consequence, the proposed formula (6) realizes the thesis of Theorem 1. Please note that the GEKS-GL index can be treated as the generalisation of the Törnqvist (1936) price index formula ($P_T^{0,t}$) to the multi-period case. In fact, in a static item universe G observed over the two period time interval $[0, 1]$, we obtain

$$\begin{aligned} P_{GEKS-GL}^{0,1} &= \prod_{\tau=0}^{1} \left(\frac{\prod_{i \in G} (\frac{p_i^1}{p_i^\tau})^{w_i^{\tau,1}(\tau)}}{\prod_{i \in G} (\frac{p_i^0}{p_i^\tau})^{w_i^{\tau,0}(\tau)}} \right)^{\frac{1}{1+1}} = \prod_{\tau=0}^{1} \left(\prod_{i \in G} (\frac{p_i^1}{p_i^0})^{w_i(\tau)} \right)^{\frac{1}{2}} \\ &= \prod_{i \in G} (\frac{p_i^1}{p_i^0})^{\frac{w_i(0)}{2}} \prod_{i \in G} (\frac{p_i^1}{p_i^0})^{\frac{w_i(1)}{2}} = \prod_{i \in G} (\frac{p_i^1}{p_i^0})^{\frac{w_i(0)+w_i(1)}{2}} = P_T^{0,1}, \end{aligned} \tag{11}$$

since $G_0 = G_1 = G_{0,1} = G$, and consequently $w_i^{\tau,0}(\tau) = w_i^{\tau,1}(\tau) = w_i(\tau)$ for any τ.

5.3 New multilateral price indices vs upper and lower bound tests

Before we make an important observation about the lower and upper bound test, let us assume that the supermarket assortment changes for two possible reasons: **(A1)** new products are introduced to store shelves, they are either cheaper equivalents of existing products, or are completely different products, however, introduced at a price not higher than the current average price in this product category; **(A2)** some of the products sold so far are withdrawn from sale and are characterized by a decreasing price and, at the same time, decreasing sales volume.

The assumption **A1** is reflected in the typical pricing strategies used by supermarket owners. Most often, new counterparts are sold discounted to generate an initiating demand (Krishnan et al. 1999). Sometimes supermarkets even use a "penetrative" strategy. This is the strategy in which the focus is on grabbing maximum market share. Hence, the price of the product is set very low initially so that it can penetrate the market and attract buyers of all segments (Rekettye & Liu 2018). The price strategy called "skimming", where the price for a new product is set very high initially, is used rather in the case of technologically innovative products and even if applied, it applies to a relatively small group of products (Hanif 2014).

The assumption **A2** is also firmly grounded in literature. As it is well known, in supermarkets, we can observe the so-called *clearance sales*, which concern products characterized by both a sharp decrease in prices and quantities compared to the previous period (van Loon & Roels 2018). There is an ongoing discussion in the current literature whether some indices suffer more or less from downward pressure from clearance prices (or "dump" prices) than others. Several analyses have revealed that especially the GEKS index reacts in a sensitive manner to product sell-offs (Chessa et al. 2017). Van Loon and Roels (2018) analyzed the effect of dump prices on multilateral methods without chaining (full window) whereas Chessa et al. (2017) applied the fixed base monthly expanding method (FBEW) to calculate the index series. It is strongly suggested in the literature to delete dump prices from the scanner sample before the CPI compilation, i.e. statisticians use the dump price filter which is supposed to detect and delete clearance sales (Białek & Beręsewicz 2021).

Summarizing, let us assume for a moment that we compare the current period s with a previous or next period τ. In the case of products which are available in both periods s and τ ($s < \tau$), we allow here for an increasing number of relatively new products, introduced in the period s at discounted prices, from the level $G^{*}_{\tau,s}$ to the level $G^{**}_{\tau,s}$. Please note that products introduced in the period τ and unavailable in the period s are not included in the set $G_{\tau,s}$. Similarly, if $\tau < s$, products observed in periods s and τ can be reduced by eliminating dump prices, i.e. from the level $G^{**}_{\tau,s}$ to the level $G^{*}_{\tau,s}$. In both scenarios, the assumptions **A1** and **A2** lead to the conclusion that after the change in the supermarket assortment, the average price change of products from the set $G^{*}_{\tau,s}$, which is a subset of the set $G^{**}_{\tau,s}$, is not smaller than the average price change of products from the set $G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}$. This conclusion can be expressed in the more formal way as follows:

$$\frac{\sum_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q_i^{\tau} p_i^{s}}{\sum_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q_i^{\tau} p_i^{\tau}} \leq \frac{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{s}}{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{\tau}}. \tag{12}$$

and

$$\prod_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} \left(\frac{p_i^{s}}{p_i^{\tau}}\right)^{\frac{q_i^{\tau} p_i^{\tau}}{\sum_{k \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q_k^{\tau} p_k^{\tau}}} \leq \prod_{i \in G^{*}_{\tau,s}} \left(\frac{p_i^{s}}{p_i^{\tau}}\right)^{\frac{q_i^{\tau} p_i^{\tau}}{\sum_{k \in G^{*}_{\tau,s}} q_k^{\tau} p_k^{\tau}}}, \tag{13}$$

where the equality is obtained if $\tau = s$. Let us note that the inequality (12) has the following implication

$$\begin{aligned} f^{L}_{G^{**}_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s}) &= \frac{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{s} + \sum_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q_i^{\tau} p_i^{s}}{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{\tau} + \sum_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q_i^{\tau} p_i^{\tau}} \\ &\leq \max\left\{\frac{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{s}}{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{\tau}}, \frac{\sum_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q_i^{\tau} p_i^{s}}{\sum_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q_i^{\tau} p_i^{\tau}}\right\} \\ &= \frac{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{s}}{\sum_{i \in G^{*}_{\tau,s}} q_i^{\tau} p_i^{\tau}} = f^{L}_{G^{*}_{\tau,s}}(q^{\tau}, p^{\tau}, p^{s}), \end{aligned} \tag{14}$$

since for any positive numbers a, b, c, d it holds that $\frac{a+b}{c+d} \leq \max(\frac{a}{c}, \frac{b}{d})$ (see Lemma from Appendix D in Białek (2019)).

Now we need to confirm that the condition R6 holds for the function $f^{GL}_{G^{**}_{\tau,s}}(q^\tau, p^\tau, p^s)$. Firstly please note that it can be written in the following form

$$
\begin{aligned}
f^{GL}_{G^{**}_{\tau,s}}(q^\tau, p^\tau, p^s) &= \prod_{i \in G^{**}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}} \\
&= \prod_{i \in G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}} \prod_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}} \\
&= \prod_{i \in G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k} \frac{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}} \prod_{i \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q^\tau_k p^\tau_k} \frac{\sum_{k \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q^\tau_k p^\tau_k}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}}.
\end{aligned} \tag{15}
$$

From (13) and (15), we conclude that

$$
\begin{aligned}
f^{GL}_{G^{**}_{\tau,s}}(q^\tau, p^\tau, p^s) &\leq \prod_{i \in G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k} \frac{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}} \prod_{i \in G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k} \frac{\sum_{k \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q^\tau_k p^\tau_k}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}} \\
&= \prod_{i \in G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k} \frac{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k + \sum_{k \in G^{**}_{\tau,s} \setminus G^{*}_{\tau,s}} q^\tau_k p^\tau_k}{\sum_{k \in G^{**}_{\tau,s}} q^\tau_k p^\tau_k}} = \prod_{i \in G^{*}_{\tau,s}} (\frac{p^s_i}{p^\tau_i})^{\frac{q^\tau_i p^\tau_i}{\sum_{k \in G^{*}_{\tau,s}} q^\tau_k p^\tau_k}} = f^{GL}_{G^{*}_{\tau,s}}(q^\tau, p^\tau, p^s)
\end{aligned} \tag{16}
$$

Please note that relations (14) and (16) mean that the requirements **R6** holds for both GEKS-L and GEKS-GL indices (obviously the requirement **R7** is also satisfied here). Taking into consideration Theorem 1, we can draw the final conclusion that, under assumptions **A1** and **A2**, the GEKS-L and GEKS-GL indices satisfy the *lower* and *upper bound tests*.

Remark

The assumptions **A1** and **A2** relate to the specific cause of changes in the product range, i.e. they only concern new or disappearing products. However, in a situation where we simply increase the range of already sold, typical products from the level $G^{*}_{\tau,s}$ to the level $G^{**}_{\tau,s}$, we can expect that the price index itself will not change significantly, i.e.

$$
\frac{\sum_{i \in G^{*}_{\tau,s}} q^\tau_i p^s_i}{\sum_{i \in G^{*}_{\tau,s}} q^\tau_i p^\tau_i} \approx \frac{\sum_{i \in G^{**}_{\tau,s}} q^\tau_i p^s_i}{\sum_{i \in G^{**}_{\tau,s}} q^\tau_i p^\tau_i}. \tag{17}
$$

Let us consider for a moment the following, normalized version of the $f^{L}_{G_{\tau,s}}$ function:

$$
f^{NL}_{G_{\tau,s}}(q^\tau, p^\tau, p^s) = \frac{1}{N_{\tau,s}} \frac{\sum_{i \in G_{\tau,s}} q^\tau_i p^s_i}{\sum_{i \in G_{\tau,s}} q^\tau_i p^\tau_i}. \tag{18}
$$

The $f^{NL}_{G_{\tau,s}}$ function satisfies the requirements **R1-R5** and **R7**. Moreover, since $N^{*}_{\tau,s} = card(G^{*}_{\tau,s}) < N^{**}_{\tau,s} = card(G^{**}_{\tau,s})$, from (17) and (18) we can expect in practice (especially for big differences in the considered product sets $G^{*}_{\tau,s}$ and $G^{**}_{\tau,s}$) that $f_{G^{*}_{\tau,s}}(q^\tau, p^\tau, p^s) \geq f_{G^{**}_{\tau,s}}(q^\tau, p^\tau, p^s)$. Thus, the normalized function $f^{NL}_{G_{\tau,s}}$ seems to satisfy also the requirement **R6** in the third scenario (17), which can be met in practice. The same considerations could be repeated for the normalized version of the $f^{GL}_{G_{\tau,s}}$ function, i.e. for the function defined as follows:

$$f^{NGL}_{G_{\tau,s}}(q^\tau, p^\tau, p^s) = \frac{1}{N_{\tau,s}} \prod_{i \in G_{\tau,s}} (\frac{p_i^s}{p_i^\tau})^{w_i^{\tau,s}(\tau)} \tag{19}$$

Consequently, from the point of view of the lower and upper bound test, it may be interesting to consider the following normalized versions of the GEKS-L and GEKS-GL indices:

$$P^{0,t}_{GEKS-NL} = \prod_{\tau=0}^{T} (\frac{\frac{1}{N_{\tau,t}} \frac{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^t}{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^\tau}}{\frac{1}{N_{\tau,0}} \frac{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^0}{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^\tau}})^{\frac{1}{T+1}}, \tag{20}$$

and

$$P^{0,t}_{GEKS-NGL} = \prod_{\tau=0}^{T} (\frac{\frac{1}{N_{\tau,t}} \prod_{i \in G_{\tau,t}} (\frac{p_i^t}{p_i^\tau})^{w_i^{\tau,t}(\tau)}}{\frac{1}{N_{\tau,0}} \prod_{i \in G_{\tau,0}} (\frac{p_i^0}{p_i^\tau})^{w_i^{\tau,0}(\tau)}})^{\frac{1}{T+1}}. \tag{21}$$

In analogy to the weighted GEKS index (Melser 2018), it would be also possible to consider the following weighted versions of the GEKS-L and GEKS-GL indices:

$$P^{0,t}_{WGEKS-L} = \prod_{\tau=0}^{T} (\frac{\frac{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^t}{\sum_{i \in G_{\tau,t}} q_i^\tau p_i^\tau}}{\frac{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^0}{\sum_{i \in G_{\tau,0}} q_i^\tau p_i^\tau}})^{v_\tau}, \tag{22}$$

and

$$P^{0,t}_{WGEKS-GL} = \prod_{\tau=0}^{T} (\frac{\prod_{i \in G_{\tau,t}} (\frac{p_i^t}{p_i^\tau})^{w_i^{\tau,t}(\tau)}}{\prod_{i \in G_{\tau,0}} (\frac{p_i^0}{p_i^\tau})^{w_i^{\tau,0}(\tau)}})^{v_\tau}, \tag{23}$$

where the weights concerning the period τ could be defined as follows:

$$v_\tau = \frac{\sum_{i \in G_\tau} q_i^\tau p_i^\tau}{\sum_{\tau=0}^{T} \sum_{i \in G_\tau} q_i^\tau p_i^\tau} \tag{24}$$

We will not take into further considerations formulas (20), (21), (22) and (23) in the paper, but the weighted versions (22) and (23) are implemented in our `PriceIndices` R package (see Białek (2021) and also Section 6 for more details about this package), i.e. they are available via package functions: `WGEKS-L()` and `WGEKS-GL()`.

6 Empirical illustration

In the following empirical study we use scanner data from one retail chain in Poland, i.e. monthly data on *long grain rice* (subgroup of COICOP 5 group: 011111), *ground coffee* (subgroup of COICOP 5 group: 012111), *drinking yoghurt* (subgroup of COICOP 5 group: 011441) and *white sugar* (subgroup of COICOP 5 group: 011811) sold in over 210 outlets during the period from December 2019 to December 2020 (352705 records, which means 210 MB of data). Before calculating the price indices, the data sets were carefully prepared. First, after deleting the records with the missing data and performing the deduplication process, the products were classified first into the relevant elementary groups (COICOP level 5) and then into their subgroups (local COICOP level 6). Product classification was performed using the `data_selecting()` and `data_classification()` functions from the `PriceIndices` R package (Białek 2021). The first function required manual preparation of dictionaries of keywords and phrases that identified individual product groups. The second function was used for problematic, previously unclassified products and required manual preparation of learning samples based on historical data. The classification itself was based on machine learning using random trees and the XGBoost algorithm (Tianqi & Carlo 2016). Next, the product matching was carried out based on the available GTIN (Global Trade Item Number) bar codes, internal retail chain codes and product labels. To match products we used the `data_matching()` function from the `PriceIndices` package. To be more precise: products with two identical codes or one of the codes identical and an identical description were automatically matched. Products were also matched if they had identical one of the codes and the Jaro-Winkler (1989) distance of their descriptions was smaller than the fixed precision value: 0.02. In the last step before calculating indices, two data filters were applied to remove unrepresentative products from the database, i.e. the `data_filtering()` function from the cited package was used. The *extreme price filter* (Białek & Beręsewicz 2021) was applied to eliminate products with more than three-fold price increase or more than double price drop from month to month. The *low sale filter* (van Loon & Roels 2018) was used to eliminate products with relatively low sales from the sample (almost 30% of products were removed). The results obtained for the GEKS-L, GEKS-GL, GEKS, Geary-Khamis, TPD and SPQ indices are presented in Figure 1.

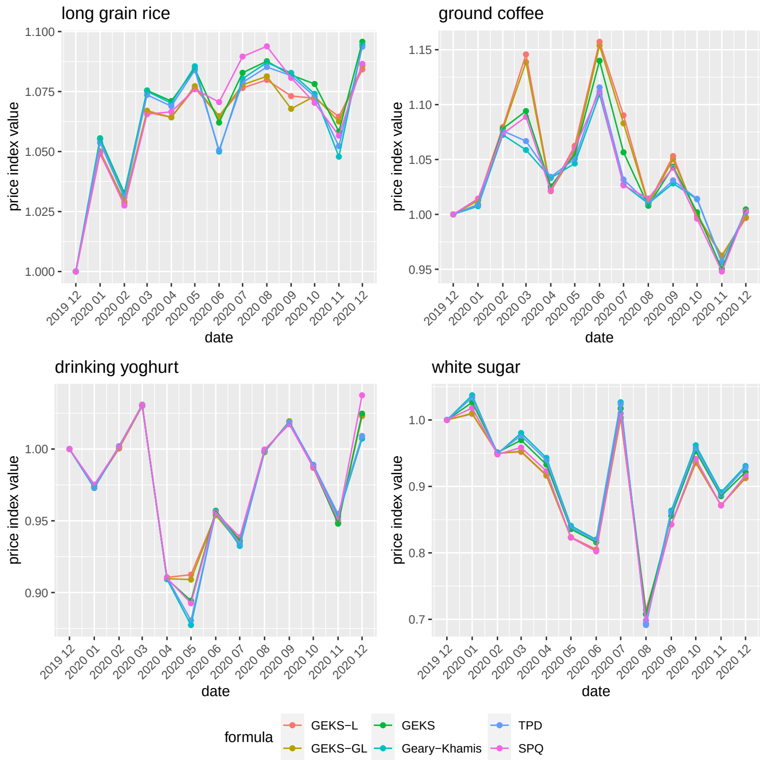

Figure 1: Comparison of selected multilateral indices for 4 homogeneous group of food products

Based on Fig.1, we can conclude that although the values of multilateral indices do not usually differ crucially, in the conducted study, noticeable differences between the studied indices were observed in the case of white sugar and ground coffee (see the scale on the Y-axis). First, an attempt was made to explain the reasons for the differences between the indices in the case of these two data sets. In order to determine the possible determinants of the differences in index indications, for each of the four scanner data sets, the following were examined (see Appendix C): *a*) monthly fractions of products remaining on sale since December 2019 (Fig.3); *b*) monthly values of Pearson's correlation coefficient between prices and quantities (Fig.4); *c*) monthly coefficients of variation of product prices (Fig.5); *d*) monthly coefficients of variation of product quantities (Fig.6). Observing the analyses *a* and *b*, a rather surprising conclusion was drawn that neither the level of price-quantity correlations and the *life expectancy* of the products differentiate the analyzed data sets (see Fig.3 and Fig.4), i.e. in all cases, we observe a weak or moderate, negative correlation between prices and quantities, and changes in the product assortment are similar to each other. And therefore these features did not contribute to the differences in the index indications, which seems to contradict the common opinion (Chessa et al. 2017) that high product churn (inflow and outflow of products) implies the differences between multilateral indices. Moreover, price volatility (measured by the coefficient of variation), which is the main cause of differences between bilateral price indices, also turned out not to differentiate the analyzed data sets (see Fig.5), and thus it was not price volatility that determined the differences between the values of the indices. Quite unexpectedly, the volatility of the quantity of products sold seems to have a clear impact on the differences between multilateral indices. Please note that the coefficients of variation of product quantities are clearly higher for the data sets for the white sugar and ground coffee (Fig.6). However, this thread requires further research.

As clear differences between the indices were observed in two of the four analyzed scanner data sets, the continuation of the graphical presentation of multilateral indices (Fig.1) was also the determination of exact, albeit averaged, differences between them. For this purpose, the average absolute differences between the indices on the basis of all monthly index values were determined by using the `compare_distances()` function from the `PriceIndices` package (see Tab.2 - Tab.5 in Appendix C). It was noted that the GEKS-L and GEKS-GL indices approximate each other and, moreover, their values are quite close to those of the GEKS index. As a rule, the values of the SPQ index are also the closest to the GEKS-L and GEKS-GL indices (in three cases, i.e. with the exception of the ground coffee data set). It seems that this observation confirms the separateness of indices that meet the *identity test*. The Geary-Khamis index is a good proxy for the Time Product Dummy (TPD) index, which confirms some previous results (Chessa et al. 2017, Białek & Beręsewicz 2021), but it always seems to be the most distant from the GEKS-L index.

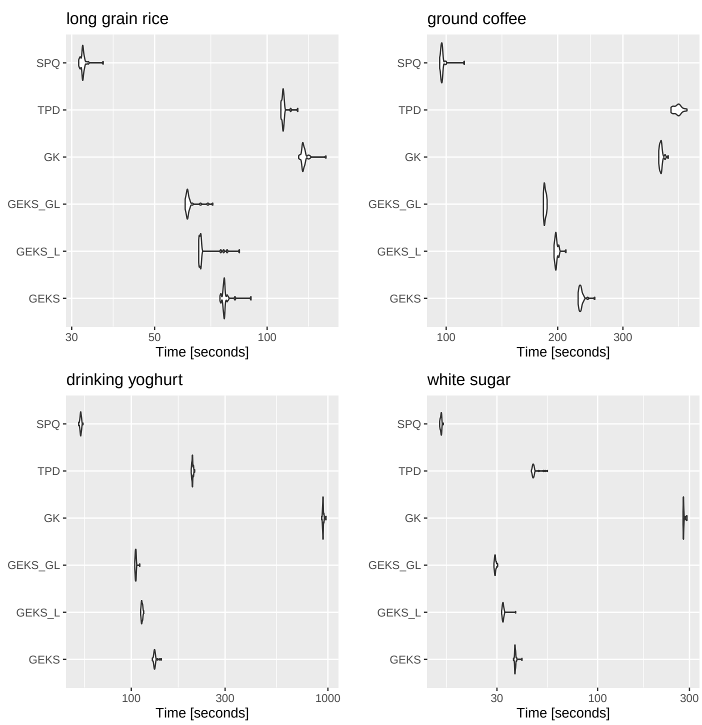

Figure 2: Comparison of computation times of selected multilateral price indices

The SPQ, GEKS-L and GEKS-GL indices require shorter calculation times compared to other index methods. This fact is not surprising for the SPQ index, as it does not work on the traditional time window. Meanwhile, the GEKS-L and GEKS-GL indices do not take into account the quantities from the current period, and thus they save on calculation time. The longest computation time recorded in the study was for the TPD and GK indices (Fig. 2).

7 Conclusions

7.1 Concluding remarks

The proposed, general class of multilateral indices (GS-GEKS), on the one hand, is based on the idea of the GEKS method, and, on the other hand, it differs crucially from this method due to the assumptions concerning the base formula of the index. Although the GS-GEKS class indices do not require the base index used in their body to be superlative (or even symmetric), as it is shown (Theorem 1), multilateral indices of this type have almost all the required properties, including the restrictive *identity test*. Moreover, the two specific cases of this general class of multilateral indices proposed in the paper, i.e. the GEKS-L and GEKS-GL indices (see Section 5), with certain assumptions regarding the circumstances of increasing or reducing the range of products sold (Section 5.3), meet the *lower* and *upper bound test*. Both the empirical and simulation studies confirmed that the two proposed indices behave rationally, and any differences in relation to the other considered multilateral indices appear only with large variability of quantity in homogeneous groups of products. Quite surprisingly, the price volatility, price-quantity correlation and product life expectancy did not play a significant role in the empirical study as determinants of differences between multilateral indices (see Section 6). Moreover, in the empirical study (see Section 6), we found that the computation times needed for the GEKS-L and GEKS-GL indices are noticeably shorter compared to most other multilateral indices (the exception is the SPQ index, which, however, does not take into account the entire time window). We also emphasize that although the base formula for the GEKS-L and GEKS-GL indices is the Laspeyre index and the geometric Laspeyres index, respectively (neither of them is superlative), the GEKS-L and GEKS-GL indices can be treated as a generalization of the Fisher and Törnqvist indices, which are superlative (see Section 5).

The paper shows that the general nature of the GS-GEKS class allows the construction of further, theoretically interesting formulas of multilateral price indices (see Remark in Section 5.3). Although these indices were not the subject of the study in this article, they retain the properties of GS-

GEKS class indices and are therefore an interesting research direction for the future. It should also be noted that both the previously known multilateral indices and the new indices proposed and discussed in the paper are implemented in the PriceIndices R package (Białek 2021), and thus the reader can verify their usefulness on their own data sets.

7.2 Further works and plans

Some aspects of the behavior of the proposed multilateral indices still remain unexplored. From a practical point of view, it seems interesting how big the sensitivity of these methods to changing the window updating methods is or if the selection of filter thresholds has a crucial influence on the index values. From a theoretical point of view, it would be interesting to define the conditions for the GS-GEKS class under which the *fixed basket test* is likely (if at all) to be fulfilled. It is planned to investigate these aspects in the near future.

Another, possible variants of indices with a structure derived from the GEKS index idea seem to be an interesting direction of research. At this point, we will only indicate the directions of further research on this type of indices.

7.2.1 Proposition based on the asynchronous quality-adjusted unit value

In the unit value concept, prices of homogeneous products are equal to the ratio of expenditure and quantity sold (International Labour Office 2004, Chessa et al. 2017). However, quantities of different products cannot be added together as in the case of homogeneous products. That is why the idea of quality-adjusted unit value assumes that prices p_i^s of different products $i \in G_s$ in month s are transformed into "quality-adjusted prices" $\frac{p_i^s}{v_i}$ and quantities q_i^s are converted into "common units" $v_i q_i^s$ (Chessa et al. 2017). Thus, the "classical" quality adjusted unit value $QUV_{G_s}^s$ of a set of products G_s in month s can be expressed as follows

$$QUV_{G_s}^s = \frac{\sum_{i \in G_s} q_i^s p_i^s}{\sum_{i \in G_s} v_i q_i^s} \tag{25}$$

The term “Quality-adjusted unit value method” (QU method for short) was introduced by Chessa (Chessa 2015, 2016). The QU method is a family of unit value based index methods and its general form can be expressed by the following ratio:

$$P_{QU}^{0,t} = \frac{QUV_{G_t}^t}{QUV_{G_0}^0} \tag{26}$$

In practice, consumer response to price changes can be delayed or even accelerated as consumers not only react to current price changes but also use their own "forecasts" or concerns about future price increases. For example, consumption of thermophilic (seasonal) fruit is likely to be higher in summer because they are cheaper than in winter, when the season is almost over. For instance, some interesting study on "unconventional" consumer behaviour, such as stocking and delayed quantity responses to price changes, and its impact on chain drift bias can be found in the paper by von Auer (2019). Since in practice we often observe prices and quantities that are not perfectly synchronised in time, the following form of the "asynchronous quality-adjusted unit value" is proposed:

$$AQUV_{G_{\tau,s}}^{\tau,s} = \frac{\sum_{i \in G_{\tau,s}} q_i^{\tau} p_i^s}{\sum_{i \in G_{\tau,s}} v_i q_i^{\tau}}, \tag{27}$$

where τ is any period from the considered time interval $[0, T]$. Obviously it holds that $AQUV_{G_{s,s}}^{s,s} = QUV_{G_s}^{s}$. Let us define now the function $P^{\tau,s}(q^{\tau}, p^{\tau}, p^s)$ as follows:

$$P^{\tau,s}(q^{\tau}, p^{\tau}, p^s) = \frac{AQUV_{G_{\tau,s}}^{\tau,s}}{AQUV_{G_{\tau,\tau}}^{\tau,\tau}}. \tag{28}$$

Putting (28) in formula (2) we obtain:

$$P_{GEKS-AQU}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G_{\tau,t}} q_i^{\tau} p_i^t}{\sum_{i \in G_{\tau,t}} v_i q_i^{\tau}}}{\frac{\sum_{i \in G_{\tau,0}} q_i^{\tau} p_i^0}{\sum_{i \in G_{\tau,0}} v_i q_i^{\tau}}} \right)^{\frac{1}{T+1}}. \tag{29}$$

Please note, that the proposed index behaves like a GEKS index based on the Laspeyres index in the case of static item universe G. In fact, if the item universe is static, we obtain

$$\begin{aligned} P_{GEKS-AQU}^{0,t} &= \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G} q_i^{\tau} p_i^t}{\sum_{i \in G} v_i q_i^{\tau}}}{\frac{\sum_{i \in G} q_i^{\tau} p_i^0}{\sum_{i \in G} v_i q_i^{\tau}}} \right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T} \left(\frac{\sum_{i \in G} q_i^{\tau} p_i^t}{\sum_{i \in G} q_i^{\tau} p_i^0} \right)^{\frac{1}{T+1}} \\ &= \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G} q_i^{\tau} p_i^t}{\sum_{i \in G} q_i^{\tau} p_i^{\tau}}}{\frac{\sum_{i \in G} q_i^{\tau} p_i^0}{\sum_{i \in G} q_i^{\tau} p_i^{\tau}}} \right)^{\frac{1}{T+1}} = P_{GEKS-L}^{0,t}. \end{aligned} \tag{30}$$

Finally please also note, that theoretically the class of the $GEKS - AQU$ indices is infinite, since different choices of v_i factors lead to different index values. We could, for instance, consider v_i factors defined in the GEKS body resulting a new, hybrid index, which would be a mixture of the GEKS and Geary-Khamis ideas. That would, however, be probably a slow solution. In this paper, we adopt the system of weights v_i corresponding to the augmented Lehr index (Lamboray 2017, van Loon & Roels 2018), where

$$v_i = \frac{\sum_{t=0}^{T} p_i^t q_i^t}{\sum_{t=0}^{T} q_i^t}. \tag{31}$$

The proposed multilateral price index GEKS-AQU has good axiomatic properties, i.e. the following theorem can be proved (the proof is omitted):

Theorem 2 *The GEKS-AQU index (29) satisfies the following tests: the transitivity, identity, multi period identity, responsiveness, continuity, positivity and normalisation, price proportionality and weak commensurability. If the item universe is the same at compared periods* 0 *and* t *then the GEKS-AQU index satisfies also the homogeneity in prices and homogeneity in quantitites test.*

7.2.2 Proposition based on the asynchronous quality-adjusted price index

Let us note that formula (27) can be expressed by using quality-adjusted prices and quantities:

$$AQUV_{G_{\tau,s}}^{\tau,s} = \frac{\sum_{i \in G_{\tau,s}} v_i q_i^{\tau} \frac{p_i^s}{v_i}}{\sum_{i \in G_{\tau,s}} v_i q_i^{\tau}}. \tag{32}$$

If we place all the adjusted prices ($\frac{p_i^s}{v_i}$) with the relative prices ($\frac{p_i^s}{p_i^{\tau}}$), then we obtain an "asynchronous quality-adjusted price index" (AQI), i.e.

$$AQI_{G_{\tau,s}}^{\tau,s} = \frac{\sum_{i \in G_{\tau,s}} v_i q_i^{\tau} \frac{p_i^s}{p_i^{\tau}}}{\sum_{i \in G_{\tau,s}} v_i q_i^{\tau}}. \tag{33}$$

This means that the AQI formula can be treated as a weighted arithmetic mean of partial indices $\frac{p_i^s}{p_i^{\tau}}$, where the weights are proportional to the relative share of the product's adjusted quantities (from the base period τ) in the sum of all adjusted quantities.

In the further part of the work, the GEKS index based on the AQI formula will be marked as GEKS-AQI, i.e. by inserting (33) into the formula (2), we obtain:

$$P_{GEKS-AQI}^{0,t} = \prod_{\tau=0}^{T} \left(\frac{\frac{\sum_{i \in G_{\tau,t}} v_i q_i^{\tau} \frac{p_i^t}{p_i^{\tau}}}{\sum_{i \in G_{\tau,t}} v_i q_i^{\tau}}}{\frac{\sum_{i \in G_{\tau,0}} v_i q_i^{\tau} \frac{p_i^0}{p_i^{\tau}}}{\sum_{i \in G_{\tau,0}} v_i q_i^{\tau}}} \right)^{\frac{1}{T+1}}. \tag{34}$$

Note that the GEKS-AQI index takes into account prices and quantities directly from all time window periods, while the GEKS-AQU index takes into account all quantities but only prices from the reference and base period. However, both formulas indirectly need information about the prices (and quantities) of products from each period in the time window to determine the factors v_i defined by formula (31). In this way, each new product in the analyzed time window has an impact on the final value of the proposed indices. The following theorem holds (the proof is omitted):

Theorem 3 *The GEKS-AQI index (34) satisfies the following tests: the transitivity, identity, multi period identity, responsiveness, continuity, positivity and normalisation, price proportionality and weak commensurability. If the item universe is the same at compared periods* 0 *and* t *then the GEKS-AQI index satisfies also the homogeneity in prices and homogeneity in quantitites test.*

Appendix A Tests for multilateral indices

Let P and Q denote all prices and quantities observed in the time interval $[0, T]$, i.e. $P = [p^0, p^1, ..., p^T]$, $Q = [q^0, q^1, ..., q^T]$, where p^t and q^t mean the vector of prices and the vector of quantities of products sold at time t, respectively. Let us denote by $P^{0,t}(P, Q)$ the considered multilateral price index defined for the entire time window $[0, T]$. The list of potential tests for that index is as follows:

Transitivity

The transitivity means that $P^{0,t}(P, Q) = P^{0,s}(P, Q)P^{s,t}(P, Q)$ for any $0 \leq s < t \leq T$.

Identity

This property means that the index equals identity if all prices revert back to their initial level, i.e. if it holds that $p_i^t = p_i^0$ for $i \in G_{0,t}$ then $P^{0,t}(P, Q) = 1$. We assume here that the item universe is the same at periods 0 and t.

Multi period identity test

This property means that if all prices and quantities revert back to their initial level, the chained index will equal the unity, i.e. if it holds that $p_i^t = p_i^0$ and $q_i^t = q_i^0$ for $i \in G_{0,t}$ then we obtain $P^{0,1}(P, Q) \times P^{1,2}(P, Q) \times .... \times P^{t-1,t}(P, Q) = 1$. We assume here that the item universe is the same at periods 0 and t.

Fixed basket test

If $G_0 = G_t$ and $q_i^0 = q_i^t = q_i$ for $i \in G_{0,t}$, then $P^{0,t}(P, Q) = \frac{\sum_{i \in G_{0,t}} p_i^t q_i}{\sum_{i \in G_{0,t}} p_i^0 q_i}$.

Upper bound test

If $G_0 \subset G_t$ and $p_i^t \leq p_i^0$ for all $i \in G_0$, then $P^{0,t}(P, Q) \leq 1$.

Lower bound test

If $G_t \subset G_0$ and $p_i^t \geq p_i^0$ for all $i \in G_t$, then $P^{0,t}(P, Q) \geq 1$.

Responsiveness test

For $G_0 \neq G_t$, if $p_i^t = p_i^0$ for all $i \in G_{0,t}$, then $P^{0,t}(P, Q)$ cannot always equal one, regardless of sets: $G_0 \setminus G_t$ and $G_t \setminus G_0$.

Continuity, positivity and normalization

$P^{0,t}(P, Q)$ is a positive and continuous function of prices and quantities, $P^{0,0}(P, Q) = 1$.

Price proportionality

If all prices are proportional in compared periods 0 and t, i.e. $p_i^t = kp_i^0$ for all $i \in G_{0,t}$ and some positive k, then the price index depends only on this proportion: $P^{0,t}(P, Q) = k$. We assume here that the item universe is the same at periods 0 and t.

Homogeneity in quantities

Rescaling the quantities in any s-th period does not influence on the price index, i.e. for any positive k it holds that $P^{0,t}(P, q^0, ..., kq^s, ..., q^t) = P^{0,t}(P, q^0, ..., q^s, ..., q^t)$.

Homogeneity in prices

Rescaling the prices in the current period changes the price index by the same proportion, i.e. for any positive k it holds that $P^{0,t}(p^0, p^1, ..., kp^t, Q) = kP^{0,t}(p^0, p^1, ..., p^t, Q)$.

Commensurability

Changing the units in which prices and quantities are expressed does not change the price index. In other words, if for each time moment $s \in [0, T]$ we have $\tilde{p}_i^s = \lambda_i p_i^s$ and $\tilde{q}_i^s = \frac{q_i^s}{\lambda_i}$ for all $i \in G_s$ ($\lambda_i > 0$), then $P^{0,t}(\tilde{P}, \tilde{Q}) = P^{0,t}(P, Q)$.

Appendix B Proof of Theorem 1

B.1 Transitivity

Let us consider such periods s and t from the time window $[0, T]$ that $0 \leq s < t \leq T$. We obtain

$$P_{GS-GEKS}^{0,s} \times P_{GS-GEKS}^{s,t} = \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,s}}(q^\tau,p^\tau,p^s)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}\right)^{\frac{1}{T+1}} \times \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,s}}(q^\tau,p^\tau,p^s)}\right)^{\frac{1}{T+1}} =$$
$$= \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,s}}(q^\tau,p^\tau,p^s)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,s}}(q^\tau,p^\tau,p^s)}\right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}\right)^{\frac{1}{T+1}} = P_{GS-GEKS}^{0,t}.$$

B.2 Identity

Let us assume that $G_0 = G_t = G_{0,t}$ and $p_i^t = p_i^0$ for $i \in G_{0,t}$. We have

$$P_{GS-GEKS}^{0,t} = \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}\right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}\right)^{\frac{1}{T+1}} = 1.$$

B.3 Multi period identity test

Let us assume that that $p_i^t = p_i^0$ and $q_i^t = q_i^0$ for $i \in G_{0,t} = G_0 = G_t$. We obtain

$$P_{GS-GEKS}^{0,1} \times P_{GS-GEKS}^{1,2} \times ... \times P_{GS-GEKS}^{t-1,t} = \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,1}}(q^\tau,p^\tau,p^1)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)} \times \frac{f_{G_{\tau,2}}(q^\tau,p^\tau,p^2)}{f_{G_{\tau,1}}(q^\tau,p^\tau,p^1)} \times \frac{f_{G_{\tau,3}}(q^\tau,p^\tau,p^3)}{f_{G_{\tau,2}}(q^\tau,p^\tau,p^2)} \times\right.$$
$$\left.... \times \frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,t-1}}(q^\tau,p^\tau,p^{t-1})}\right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}\right)^{\frac{1}{T+1}} = \prod_{\tau=0}^{T}\left(\frac{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}\right)^{\frac{1}{T+1}} = 1.$$

Please note that this proof does not require the condition $q_i^t = q_i^0$.

B.4 Responsiveness test

Let us assume that $G_0 \neq G_t$ and $p_i^t = p_i^0$ for all $i \in G_{0,t}$. Since $G_0 \neq G_t$, we know that for at least one period τ_0 we have $G_{\tau_0,t} \neq G_{\tau_0,0} \cap G_{\tau_0,t}$ and, from our initial assumption (see Section

2), we have that $G_{\tau,0} \cap G_{\tau,t} \neq \emptyset$ for any τ. From the assumption R5 (see Section 4), we get that $f_{G^{*}_{\tau_0,t}}(q^{\tau_0}, p^{\tau_0}, p^t) \neq f_{G^{**}_{\tau_0,t}}(q^{\tau_0}, p^{\tau_0}, p^t)$, where $G^{*}_{\tau_0,t} = G_{\tau_0,0} \cap G_{\tau_0,t}$ and $G^{**}_{\tau_0,t} = G_{\tau_0,t}$. In a similar way, it can be shown that $f_{G_{\tau_0,0}}(q^{\tau_0}, p^{\tau_0}, p^0) \neq f_{G_{\tau_0,0} \cap G_{\tau_0,t}}(q^{\tau_0}, p^{\tau_0}, p^0)$. Thus, in general, it holds that $P^{0,t}_{GS-GEKS} = \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \neq \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}}$.
Since we assume that prices at compared time moments are identical, i.e. $p_i^t = p_i^0$, we obtain finally that: $P^{0,t}_{GS-GEKS} \neq \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = 1$.

B.5 Continuity, positivity and normalization

Continuity, *positivity* and *normalization* are a consequence of the requirement R1 (Section 4).

B.6 Price proportionality

Assumption that the item universe is the same at periods 0 and t means that $G_0 = G_t = G_{0,t}$ and also $G_{\tau,0} = G_{\tau,t}$ for any τ. Let us assume that $p_i^t = kp_i^0$ for all $i \in G_{0,t}$ and some positive k. As a consequence, we obtain
$P^{0,t}_{GS-GEKS} = \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0}}(q^\tau,p^\tau,kp^0)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}}$.
From the requirement R2 (see Section 4), we have that
$P^{0,t}_{GS-GEKS} = \prod_{\tau=0}^{T} (\frac{kf_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = (k^{T+1})^{\frac{1}{T+1}} = k$.

B.7 Homogeneity in quantities

The requirement R3 (Section 4) leads to the conclusion that

$$
\begin{aligned}
P^{0,t}_{GS-GEKS}(P, q^0, ..., kq^s, ..., q^t) &= (\frac{f_{G_{s,t}}(kq^s,p^\tau,p^t)}{f_{G_{s,0}}(kq^s,p^\tau,p^0)})^{\frac{1}{T+1}} \times \prod_{\tau=0,\tau\neq s}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \\
&= (\frac{f_{G_{s,t}}(q^s,p^\tau,p^t)}{f_{G_{s,0}}(q^s,p^\tau,p^0)})^{\frac{1}{T+1}} \times \prod_{\tau=0,\tau\neq s}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \\
&= \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = P^{0,t}_{GS-GEKS}(P, Q).
\end{aligned}
$$

B.8 Homogeneity in prices

The requirement R2 (Section 4) leads to the conclusion that

$$
\begin{aligned}
P^{0,t}_{GS-GEKS}(p^0, ..., kp^t, Q) &= \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,kp^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = \prod_{\tau=0}^{T} (k\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \\
&= (k^{T+1})^{\frac{1}{T+1}} \times \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = k \times P^{0,t}_{GS-GEKS}(P, Q).
\end{aligned}
$$

B.9 Upper bound test

Let us assume that the requirement R6 is additionally satisfied (see Section 4). Let us also assume that $G_0 \subset G_t$, which leads to the conclusion that $G_{\tau,0} = (G_{\tau,0} \cap G_{\tau,t}) \subset G_{\tau,t}$ for any τ. From R6, we obtain

$P_{GS-GEKS}^{0,t} = \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \leq \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}}$.

Due to the assumption that $p_i^t \leq p_i^0$ for all $i \in G_0$, from R7 we have that

$\prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \leq \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}}$

$= \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = 1,$

which means that $P_{GS-GEKS}^{0,t} \leq 1$.

B.10 Lower bound test

Let us assume that the requirement R6 is additionally satisfied (see Section 4). Let us also assume that $G_t \subset G_0$, which leads to the conclusion that $G_{\tau,t} = (G_{\tau,0} \cap G_{\tau,t}) \subset G_{\tau,0}$ for any τ. From R6, we obtain

$P_{GS-GEKS}^{0,t} = \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \geq \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}}$.

Due to the assumption that $p_i^t \geq p_i^0$ for all $i \in G_t \subset G_0$, from R7, we have that

$\prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^t)}{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} \geq \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,t}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}}$

$= \prod_{\tau=0}^{T} (\frac{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)}{f_{G_{\tau,0} \cap G_{\tau,t}}(q^\tau,p^\tau,p^0)})^{\frac{1}{T+1}} = 1,$

which means that $P_{GS-GEKS}^{0,t} \geq 1$.

Appendix C Additional analysis concerning the *Empirical illustration*

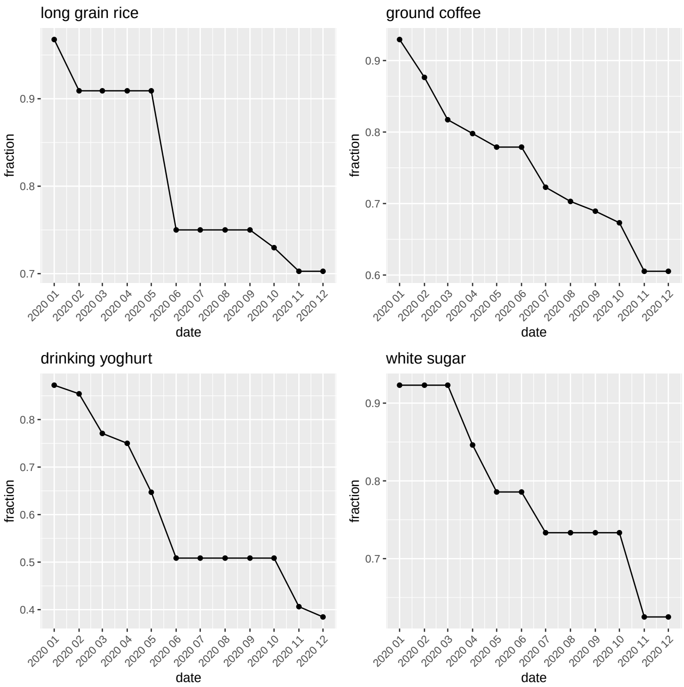

Figure 3: Monthly fractions of products remaining on sale since Dec 2019

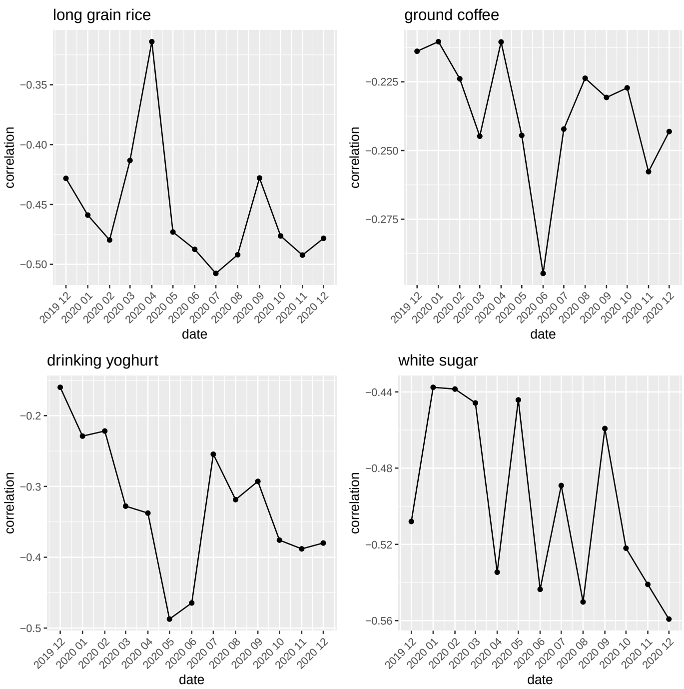

Figure 4: Monthly values of Pearson's correlation coefficient between prices and quantities

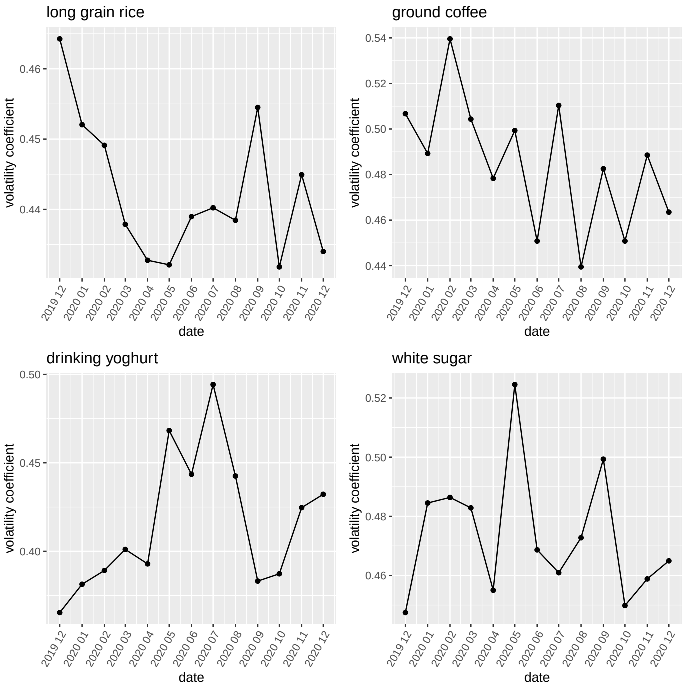

Figure 5: Monthly coefficients of variation of prices

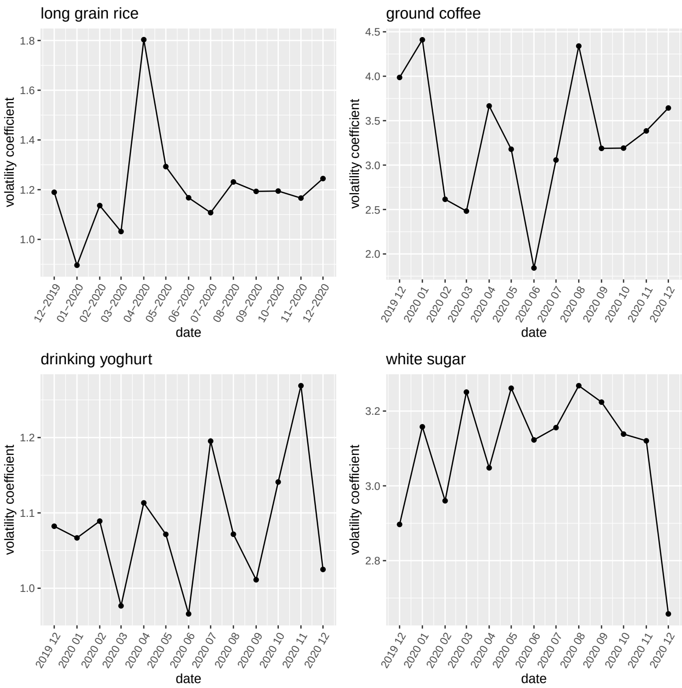

Figure 6: Monthly coefficients of variation of quantities

Table 1: Mean absolute differences between considered price indices (*long grain rice*) [p.p]

distances	GEKS-L	GEKS-GL	GEKS	GK	TPD	SPQ
GEKS-L	0.000	0.124	0.644	0.751	0.609	0.445
GEKS-GL	0.124	0.000	0.605	0.713	0.571	0.457
GEKS	0.644	0.605	0.000	0.270	0.306	0.580
GK	0.751	0.713	0.270	0.000	0.143	0.707
TPD	0.609	0.571	0.306	0.143	0.000	0.617
SPQ	0.445	0.457	0.580	0.707	0.617	0.000

Table 2: Mean absolute differences between considered price indices (*ground coffee*) [p.p]

distances	GEKS-L	GEKS-GL	GEKS	GK	TPD	SPQ
GEKS-L	0.000	0.262	1.189	2.227	2.006	1.634
GEKS-GL	0.262	0.000	0.965	2.003	1.783	1.494
GEKS	1.189	0.965	0.000	1.203	0.984	0.773
GK	2.227	2.003	1.203	0.000	0.256	0.869
TPD	2.006	1.783	0.984	0.256	0.000	0.808
SPQ	1.634	1.494	0.773	0.869	0.808	0.000

Table 3: Mean absolute differences between considered price indices (*drinking yoghurt*) [p.p]

distances	GEKS-L	GEKS-GL	GEKS	GK	TPD	SPQ
GEKS-L	0.000	0.060	0.237	0.513	0.459	0.370
GEKS-GL	0.060	0.000	0.215	0.480	0.437	0.360
GEKS	0.237	0.215	0.000	0.361	0.337	0.207
GK	0.513	0.480	0.361	0.000	0.082	0.483
TPD	0.459	0.437	0.337	0.082	0.000	0.437
SPQ	0.370	0.360	0.207	0.483	0.437	0.000

Table 4: Mean absolute differences between considered price indices (*white sugar*) [p.p]

distances	GEKS-L	GEKS-GL	GEKS	GK	TPD	SPQ
GEKS-L	0.000	0.114	1.153	1.891	1.660	0.469
GEKS-GL	0.114	0.000	1.087	1.820	1.592	0.361
GEKS	1.153	1.087	0.000	0.757	0.523	0.914
GK	1.891	1.820	0.757	0.000	0.264	1.495
TPD	1.660	1.592	0.523	0.264	0.000	1.277
SPQ	0.469	0.361	0.914	1.495	1.277	0.000

References

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

Białek, J. (2021), 'Priceindices – a new R package for bilateral and multilateral price index calculations', *Statistika – Statistics and Economy Journal* **36**(2), 122–141.

Białek, J. & Beręsewicz, M. (2021), 'Scanner data in inflation measurement: from raw data to price indices', *The Statistical Journal of the IAOS* **37**, 1315–1336.

Białek, J. (2019), 'Remarks on geo-logarithmic price indices', *Journal of Official Statistics* **35**(2), 287–317.

Caves, D. W., Christensen, L. R. & Diewert, W. E. (1982), 'Multilateral comparisons of output, input, and productivity using superlative index numbers', *Economic Journal* **92**(365), 73–86.

Chessa, A. (2015), Towards a generic price index method for scanner data in the dutch cpi, *in* '14th meeting of the Ottawa Group, Tokyo', pp. 20–22.

Chessa, A. (2016), 'A new methodology for processing scanner data in the dutch CPI', *Eurostat review of National Accounts and Macroeconomic Indicators* **1**, 49–69.

Chessa, A. (2019), A comparison of index extension methods for multilateral methods, *in* 'Paper presented at the 16th Meeting of the Ottawa Group on Price Indices, Rio de Janeiro, Brazil'.

Chessa, A. G., Verburg, J. & Willenborg, L. (2017), A comparison of price index methods for scanner data, *in* '15th meeting of the Ottawa Group, Eltville', pp. 10–12.

de Haan, J., Hendriks, R. & Scholz, M. (2021), 'Price measurement using scanner data: Time-product dummy versus time dummy hedonic indexes', *Review of Income and Wealth* **67**(2), 394–417.

de Haan, J. & Krsinich, F. (2018), 'Time dummy hedonic and quality-adjusted unit value indexes: Do they really differ?', *Review of Income and Wealth* **64**(4), 757–776.

de Haan, J. & van der Grient, H. A. (2011), 'Eliminating chain drift in price indexes based on scanner data', *Journal of Econometrics* **161**(1), 36–46.

Diewert, W. E. (2020), The chain drift problem and multilateral indexes, Technical report, Discussion Paper 20-07, Vancouver School of Economics.

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

Eltetö, O. & Köves, P. (1964), 'On a problem of index number computation relating to international comparison', *Statisztikai Szemle* **42**(10), 507–518.

Fisher, I. (1922), *The making of index numbers: a study of their varieties, tests, and reliability*, Vol. xxxi, Houghton Mifflin.

Geary, R. C. (1958), 'A note on the comparison of exchange rates and purchasing power between countries', *Journal of the Royal Statistical Society. Series A (General)* **121**(1), 97–99.

Gini, C. (1931), 'On the circular test of index numbers', *Metron* **9**(9), 3–24.

Hanif, M. (2014), 'Market skimming pricing: An examination of elements supporting high price for new products in pakistan', *European Journal of Business and Management* **6**(23), 180–187.

Inklaar, R. & Diewert, W. E. (2016), 'Measuring industry productivity and cross-country convergence', *Journal of Econometrics* **191**(2), 426–433.

International Labour Office (2004), 'Consumer price index manual: Theory and practice.', Geneva.

Ivancic, L., Diewert, W. E. & Fox, K. J. (2011), 'Scanner data, time aggregation and the construction of price indexes', *Journal of Econometrics* **161**(1), 24–35.

Jaro, M. (1989), 'Advances in record-linkage methodology as applied to matching the 1985 census of tampa, florida', *Journal of the American Statistical Association* **84**(406), 414–420.

Jevons, W. S. (1865), 'On the variation of prices and the value of the currency since 1782', *Journal of the Statistical Society of London* **28**(2), 294–320.

Khamis, S. H. (1972), 'A new system of index numbers for national and international purposes', *Journal of the Royal Statistical Society: Series A (General)* **135**(1), 96–121.

Krishnan, T., Bass, F. & Jain, D. (1999), 'Optimal pricing strategy for new products', *Management Science* **45**(12), 1650–1663.

Krsinich, F. (2014), The fews index: Fixed effects with a window splice–non-revisable quality-adjusted price indexes with no characteristic information, *in* 'meeting of the group of experts on consumer price indices', pp. 26–28.

Lamboray, C. (2017), The geary khamis index and the lehr index: how much do they differ, *in* 'Paper to be presented at the 15th meeting of the Ottawa Group', pp. 10–12.

Martini, M. (1992), 'A general function of axiomatic index numbers', *Journal of the Italian Statistical Society* **3**(1), 359–376.

Melser, D. (2018), 'Scanner data price indexes: Addressing some unresolved issues', *Journal of Business and Economic Statistics* **36**(03), 516–522.

Rekettye, G. & Liu, J. (2018), *Pricing: The New Frontier*, 1st edn, Transnational Press, London.

Tianqi, C. & Carlo, G. (2016), Xgboost: A scalable tree boosting system, *in* 'Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016. ACM', p. 785–794.

Törnqvist, L. (1936), 'The bank of finland's consumption price index', *Bank of Finland Monthly Bulletin* pp. 1–8.

van Loon, K. V. & Roels, D. (2018), Integrating big data in the Belgian CPI, *in* 'Paper presented at the meeting of the group of experts on consumer price indices, 8-9 May 2018, Geneva, Switzerland'.

von Auer, L. (2019), The nature of chain drift, *in* 'Paper presented at the 17th Meeting of the Ottawa Group on Price Indices, 8–10 May 2019, Rio de Janerio, Brasil'.

von der Lippe, P. (2007), *Index Theory and Price Statistics*, Peter Lang, Berlin, Germany.

Zhang, L.-C., Johansen, I. & Nyagaard, R. (2019), 'Tests for price indices in a dynamic item universe', *Journal of Official Statistics* **35**(3), 683–697.