



ORIGINAL ARTICLE

Citation: Białek, J. (2020). Proposition of a hybrid price index formula for the Consumer Price Index measurement. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 15(4), 697–716. doi: 10.24136/eq.2020.030


Contact: jacek.bialek@uni.lodz.pl, University of Lodz, Department of Statistical Methods, Re-wolucji 1905 r. nr. 41, 90-214 Lodz,

Received: 12.07.2020; Revised: 26.11.2020; Accepted: 7.12.2020; Published online: 20.12.2020

Jacek Białek

University of Lodz, Poland

Statistics Poland, Poland

 orcid.org/0000-0002-0952-5327

Proposition of a hybrid price index formula for the Consumer Price Index measurement

JEL Classification: C43; E31

Keywords: price index, Consumer Price Index, Lowe index, Young index, hybrid price index

Abstract

Research background: The Consumer Price Index (CPI) is a basic, commonly accepted and used measure of inflation. The index is a proxy for changes in the costs of household consumption and it assumes constant consumer utility. In practice, most statistical agencies use the Laspeyres price index to measure the CPI. The Laspeyres index does not take into account movements in the structure of consumption which may be consumers' response to price changes during a given time interval. As a consequence, the Laspeyres index can suffer from commodity substitution bias. The Fisher index is perceived as the best proxy for the COLI but it needs data on consumption from both the base and research period. As a consequence, there is a practical need to look for a proxy of the Fisher price index which does not use current expenditure shares as weights.

Purpose of the article: The general purpose of the article is to present a hybrid price index, the idea of which is based on the Young and Lowe indices. The particular aim of the paper is to discuss the usefulness of its special case with weights based on correlations between prices and quantities.

Methods: A theoretical background for the hybrid price index (and its geometric version) is constructed with the Lowe and Young price indices used as a starting point. In the empirical study, scanner data on milk, sugar, coffee and rice are utilized to show that the hybrid index can be a good proxy for the Fisher index, although it does not use the expenditures from the research period.

Findings & Value added: The empirical and theoretical considerations con-firm the hybrid nature of the proposed index, i.e. in a special case it forms the convex combination of the Young and Lowe indices. This study points out the usefulness of the proposed price index in the CPI measurement, especially when the target index is the Fisher formula. The proposed general hybrid

price index formula is a new one in the price index theory. The proposed system of weights, which is based on the correlations between prices and quantities, is a novel idea in the price index methodology.

Introduction

The Consumer Price Index (CPI) is a basic, commonly accepted and used measure of inflation. The index is a proxy for changes in the costs of household consumption and it assumes constant consumer utility, i.e. the Cost of Living Index (COLI). The CPI is used for indexing nominal values in the economy, which is important in price decision making by enterprises. The CPI is also important in the monetary policy if the central bank uses the direct inflation targeting strategy. It has been in force in Poland since 1999, and the CPI has been the reference indicator from the beginning.

In practice, in the case of the so-called “traditional data collection,” statistical agencies use the Laspeyres (1871) price index to calculate the CPI (see Clements and Izan (1987) or White (1999)). The Laspeyres index does not take into account movements in the structure of consumption which may be consumers' response to price changes during a given time interval. As a consequence, the Laspeyres index can suffer from the commodity substitution bias. To be more precise, please note that this kind of CPI bias results from changes in relative prices of individual goods included in the CPI basket. The substitution effect is that consumers respond to price changes by exchanging those goods or services that are relatively more expensive for relatively cheaper ones (Hałka & Leszczyńska, 2011). Although the substitution bias is not the only CPI bias, it is best recognized in the literature and can be crucial from the point of view of any financial decision which is based on the inflation rate. For example, the Boskin Commission (1996), when analyzing data from the USA for the period 1995-1996, determined the level of total CPI bias at 1.1 percentage point, split into: i) 0.5 p.p. — the substitution bias (0.4 p.p.) and the outlet bias (0.1 p.p.); ii), and the remaining 0.6 p.p. — the bias resulting from the change in the quality of goods along with the new goods bias. Today, sales markets are much more dynamic and technological changes are more rapid than 20 years ago, hence it can be expected that the CPI bias is not negligibly small. Please note that the CPI bias is always a cost for the economy. Apart from the fact that many central banks use the direct inflation targeting strategy, in many countries (including Poland), the CPI index is used for the valorization of pensions and for indexing financial contracts, including interbank ones. Therefore, the CPI estimation should be as accurate as

possible and the reduction of the CPI measurement bias by even a per mille has financial significance.

Most economists perceive superlative indices (such as the Fisher, Walsh or Törnqvist indices) as the best proxies for the COLI (Von der Lippe, 2007), CPI Manual (2004)). The difference between the Laspeyres index and any superlative index plays a role of an approximation of the CPI substitution bias. The Fisher index is treated as being the best proxy for the COLI but, similarly to other superlative indices, it needs data on consumption from both the base and research period. Obtaining consumption data for the current period is problematic from the practical point of view, because the household budgets survey always provides this information with a certain delay (usually a year or longer). This is due, among others, to the fact that this type of survey is very expensive. However, there are some ways to approximate the Fisher price index and to reduce the CPI substitution bias by using indices which require information about expenditures only from the base period, e.g. the Lloyd–Moulton price index, the AG Mean index, or the Lowe and Young indices (CPI Manual, 2004).

The paper proposes a hybrid and general price index the idea of which is based on the Young and Lowe indices. The aim of the paper is to discuss its special case with weights based on correlations between prices and quantities. The proposed system of weights is a new idea in the price index methodology, which is the added value of the paper. To confirm the usefulness of the proposed index method, real scanner dataset obtained from one supermarket chain in Poland is used, since scanner data contain full information about sold products, including current expenditures. The monthly aggregated scanner data on milk, sugar, coffee, and rice is used to show that a hybrid index can be a good proxy for the Fisher index, although it does not use the expenditures from the research period. The empirical and theoretical considerations confirm the hybrid nature of the proposed index, i.e. in a special case, it forms the convex combination of the Young and Lowe indices. The study points out the usefulness of the proposed price index in the CPI measurement.

The paper is organized as follows: after the literature review concerning approaches to seeking the best price index in the CPI measurement, the next section presents methodology of the research, including a description of construction of the proposed hybrid index formula, the main hypothesis for the index and a summary of scanner data processing. The section with obtained results presents comparisons between the Fisher index and the new suggested price index formula, which is the special case of the general hybrid formula. Finally, the obtained results are compared to those known

from the literature, and the paper concludes with some recommendations for the CPI substitution bias reduction.

Literature review

The choice of the optimal price index formula in the CPI measurement is not easy and it depends on the data aggregation level and the data source. For instance, in the so-called “traditional” data collection (data collected by the interviewer in the field) if you do not have any information about the consumption level, we must choose among so-called elementary indices, e.g. the Dutot (1738), Carli (1804) or Jevons (1865) price indices. Having data on consumption, it is possible use weighted index formulas, i.e. in practice, the Laspeyres-type (1871) index is used, although the Fisher (1922) index seems to be a much better choice (von der Lippe, 2007). On the other hand, obtaining data from supermarkets, i.e. scanner data from electronic terminals located in sale points, allows you to have full information on products even at the elementary level, and the list of potential indices which can be used in this case is much longer. For instance, multi-lateral indices have recently gained popularity here, e.g. the GEKS (Gini, 1931; Eltetö & Köves, 1964; Szulc, 1983), the Geary-Khamis (Geary, 1958; Khamis, 1970) or Time Product Dummy (de Haan & Krsnich, 2017) indices. Finally, using web-scraped data in the inflation measurement, i.e. data from web-pages or electronic trading platforms, statisticians often use the mixture of unweighted bilateral indices and multilateral ones, such as the GEKS-Jevons formula (Van Loon & Roels, 2018; Białek & Bobel, 2019). More updated studies on price index methods dedicated to new data sources can be found here: Krsnich (2014), de Haan *et. al.* (2016), Chessa (2017), Diewert and Fox (2017), von Auer (2019), Mehrhoff (2019), Białek and Bobel (2019), Webster and Tarnow-Mordi (2019), Abe and Rao (2019), Zhang *et. al.* (2019), and Białek (2020a).

Although the history of the CPI has spanned 100 years, there are still open problems in the index methodology, and in general there are several main approaches when selecting the best price index formula. In the stochastic approach, the price index is treated as an unknown parameter in the regression model describing price changes (Selvanathan & Prasada Rao, 1992; Diewert, 2005). In the economic approach, correlations between prices and quantities are taken into considerations (Von der Lippe, 2007; CPI Manual, 2004), and we expect that the price index formula is as close to the true cost of living index as possible. The final report of the Boskin Commission begins with a recommendation that “the Bureau of Labor Sta-

tistics (BLS) should establish a cost of living index (COLI) as its objective in measuring consumer prices” (see Boskin *et al.*, 1996, p. 2). Further discussion on the COLI theory can be found in the papers of: Diewert (1993), Jorgenson and Slesnick (1983), and Pollak (1989). Finally, in the axiomatic approach, a well-constructed price index should satisfy a group of “tests” or axioms (Balk, 1995). Systems of minimum requirements of price indices were provided by Martini (1992), Eichhorn and Voeller (1976) and Olt (1996). Any new price index proposition should fulfill axioms from the system of minimal requirements (Białek, 2014a).

As mentioned above, superlative indices (Diewert, 1976) may be considered as best proxies for the COLI (Von der Lippe, 2007; CPI Manual, 2004). In an ideal case, the “traditional” CPI should be measured by using the Fisher index, but the problem is that the index needs data on consumption from the current period, which is out of range of any NSI (*National Statistical Institute*). One possible solution to this problem is using the price index method which is able to approximate the Fisher index despite the lack of weights from the current period. In the literature, one can encounter several interesting ideas for a Fisher index proxy, e.g. using the Constant Elasticity of Substitution (CES) framework, it is possible to approximate the Fisher price index if only you can estimate the elasticity of the substitution parameter. In particular, the Lloyd–Moulton price index (Lloyd, 1975; Moulton, 1996; Shapiro & Wilcox, 1997) and the AG mean indices (Lent & Dorfman, 2009) seem to be a good proxy for the Fisher index. Nevertheless, although there are lots of ways of estimating the elasticity parameter (Feenstra & Reinsdorf, 2007; Biggeri & Ferrari, 2010; Greenlees, 2011; Armknecht & Silver, 2012), the problem with this parameter is its instability in time (Białek, 2017a).

Another group of indices which can be applied for the approximation of the Fisher price index are the Young and the Lowe indices (Armknecht & Silver, 2012; CPI Manual, 2004), which use information about quantities from the arbitrary fixed prior period. It is empirically shown that the quality of their approximations is good enough (Białek, 2017b; Juszczak, 2020). However, it would be very interesting to provide a general, hybrid price index formula which: a) fulfills the most important requirements from the axiomatic approach; b) may serve as a good proxy for the Fisher index; c) includes the Lowe and Young indices as their special cases; and d) uses additional information about correlations between prices and quantities of goods and services. Some other general price index propositions can be found in the following papers: Białek (2012, 2015, 2020b), and the research on the impact of correlations between prices and quantities on price index results can be also encountered in the literature (Białek, 2019). Neverthe-

less, according to the author’s best knowledge, there is a lack of papers which consider all the above-mentioned postulates “a” — “d”. In the author’s opinion, the general hybrid price index formula which is proposed in the paper is a new one in the price index theory and practice, and may be considered as a good proxy for the Fisher price index.

Research methodology

The weighted bilateral (direct) price index formula is a function of a set of prices $\{p_i^\tau : i = 0, 1, \dots, N; \tau = s, t\}$ and quantities $\{q_i^\tau : i = 0, 1, \dots, N; \tau = s, t\}$ of the considered group of N commodities observed in the research (current) period t and the base period s . Thus, the Laspeyres price index can be written as follows:

$$P_{La} = \frac{\sum_{i=1}^N q_i^s p_i^t}{\sum_{i=1}^N q_i^s p_i^s} \quad (1)$$

while the Paasche formula can be expressed as

$$P_{La} = \frac{\sum_{i=1}^N q_i^t p_i^t}{\sum_{i=1}^N q_i^t p_i^s} \quad (2)$$

The Fisher price index is the geometric mean of the Laspeyres and Paasche indices, i.e.:

$$P_F = \sqrt{P_{La} P_{Pa}} \quad (3)$$

and thus, similarly to the Paasche formula, it uses quantities from the current period. The CPI is measured as the weighted arithmetic mean of price relatives where the weights are the expenditure shares in the base period. In practice, there are crucial differences between the expenditure shares from the survey period (τ) and the expenditure shares from the current period (t), but please note that the compilation of household expenditure data requires time. NSIs use prior period τ survey weights to calculate the CPI,

where $\tau < s < t$. As a consequence, the Young price index can be proposed (CPI Manual, 2004):

$$P_Y = \sum_{i=1}^N w_i^{\tau,\tau} \left(\frac{p_i^t}{p_i^s} \right) \quad (4)$$

where

$$w_i^{\tau,\tau} = \frac{p_i^\tau q_i^\tau}{\sum_{k=1}^N p_k^\tau q_k^\tau} \quad (5)$$

In the paper, the geometric Young index formula is also considered, which is as follows:

$$P_{GY} = \prod_{i=1}^N \left(\frac{p_i^t}{p_i^s} \right)^{w_i^{\tau,\tau}} \quad (6)$$

Armknacht and Silver (2012) state that: “More typically, weights are price-updated between period τ and the price reference period s to effect fixed period- τ quantities”. As a consequence, the Lowe price index can be written as follows (CPI Manual, 2004):

$$P_{Lo} = \frac{\sum_{i=1}^N p_i^t q_i^\tau}{\sum_{i=1}^N p_i^s q_i^\tau} = \sum_{i=1}^N w_i^{s,\tau} \left(\frac{p_i^t}{p_i^s} \right) \quad (7)$$

where

$$w_i^{s,\tau} = \frac{p_i^s q_i^\tau}{\sum_{k=1}^N p_k^s q_k^\tau} \quad (8)$$

The geometric version of the Lowe price index can be expressed as follows:

$$P_{GLo} = \prod_{i=1}^N \left(\frac{p_i^t}{p_i^s} \right)^{w_i^{s,\tau}} \quad (9)$$

Please note that both the Young and Lowe indices use quantities q_i^τ since more updated quantities (q_i^s or q_i^t) may not be available. However, information about prices p_i^τ , p_i^s and p_i^t is available, and thus the following weights can be also considered:

$$w_i^{t,\tau} = \frac{p_i^t q_i^\tau}{\sum_{k=1}^N p_k^t q_k^\tau} \tag{10}$$

Let us consider the following hybrid system of weights, namely

$$w_{Hi}^\gamma = \gamma_\tau w_i^{\tau,\tau} + \gamma_s w_i^{s,\tau} + \gamma_t w_i^{t,\tau} \tag{11}$$

where $\gamma_\tau, \gamma_s, \gamma_t$ are positive numbers that add up to one, i.e. $\gamma_\tau + \gamma_s + \gamma_t = 1$. The following general hybrid price index formula parametrized by the real 3-dimensional vector $\gamma = [\gamma_\tau, \gamma_s, \gamma_t]$ is proposed:

$$P_H^\gamma = \sum_{i=1}^N w_{Hi}^\gamma \left(\frac{p_i^t}{p_i^s} \right) \tag{12}$$

and its geometric version

$$P_{GH}^\gamma = \prod_{i=1}^N \left(\frac{p_i^t}{p_i^s} \right)^{w_{Hi}^\gamma} \tag{13}$$

It is easy to show that formulas (12) and (13) satisfy Martini’s system of axioms (the proof is upon request), according to which a properly constructed price index should meet the following postulates: identity, commensurability and linear homogeneity. In short, it means that (a) with identical prices in the base period and the current period, the index should be equal to one; (b) the index should not depend on the monetary unit in which the prices are expressed; (c) the index value should depend linearly on identical changes of all prices in the current period.

Please note that special cases of formulas (12) and (13) may cover the Young or Lowe indices or may lie between them. In fact, using weighting vectors $\gamma_Y = [1, 0, 0]$, $\gamma_{Lo} = [0, 1, 0]$ and $\gamma_{Y-Lo} = [0.5, 0.5, 0]$, the following

relations can be obtained: $P_H^{\gamma_Y} = P_Y$, $P_{GH}^{\gamma_Y} = P_{GY}$, $P_H^{\gamma_{Lo}} = P_{Lo}$, $P_{GH}^{\gamma_{Lo}} = P_{GLo}$ and the hybrid nature of proposed indices is observed: $P_H^{\gamma_{Y-Lo}} = (P_Y + P_{Lo})/2$, $P_{GH}^{\gamma_{Y-Lo}} = \sqrt{P_Y P_{Lo}}$. However, the author's previous results show that the choice between the Young and Lowe indices may depend on the correlation between price and quantity of sold products (Białek, 2017a, 2017b). Let $\rho_{a,b}$ denote the Pearson correlation coefficient between prices from the period a and quantities from the period b . Then three correlation coefficients can be considered: $\rho_{\tau,\tau}$, $\rho_{s,\tau}$ and $\rho_{i,\tau}$, and it is assumed that at least one of these coefficients is nonzero. Thus it seems reasonable to fix the γ -vector components in proportion to levels of correlation of prices and quantities in the following manner:

$$\gamma_{\tau}^0 = \frac{|\rho_{\tau,\tau}|}{|\rho_{\tau,\tau}| + |\rho_{s,\tau}| + |\rho_{i,\tau}|}, \quad \gamma_s^0 = \frac{|\rho_{s,\tau}|}{|\rho_{\tau,\tau}| + |\rho_{s,\tau}| + |\rho_{i,\tau}|}, \quad \gamma_i^0 = \frac{|\rho_{i,\tau}|}{|\rho_{\tau,\tau}| + |\rho_{s,\tau}| + |\rho_{i,\tau}|} \quad (14)$$

where $|x|$ denotes the absolute value of x number. For the vector $\gamma_0 = [\gamma_{\tau}^0, \gamma_s^0, \gamma_i^0]$, you obtain $\gamma_{\tau}^0 + \gamma_s^0 + \gamma_i^0 = 1$, and the empirical part of the paper considers hybrid indices (12) and (13) based on the system of weights γ_0 . On the one hand, the formulas need some additional calculations (correlation coefficients), which may be considered as their weakness, on the other hand, the completed empirical study proves that this additional information about correlations allows for a better approximation of the ideal Fisher price index.

The main hypothesis states that formulas $P_H^{\gamma_0}$ and $P_{GH}^{\gamma_0}$ are reasonable and better proxies for the Fisher index than the commonly used Laspeyres index. To prove this hypothesis, scanner data on milk, sugar, coffee, and rice obtained from one of the retail chains in Poland were used. The monthly data covered the period from Dec, 2018 to Dec, 2019. Before index calculations were carried out, data sets had been carefully prepared, i.e. the data set with over 500000 records had undergone all necessary processing stages (product classification, product matching and data filtration). One of the main problems encountered at this stage of data processing was the standardization of product descriptions. The purpose of the standardization process was to obtain a set of unique (identical) product names that could be used to match products over time. The original datasets from the retailer had been uncleaned, duplicated and contained errors and non-homogeneous names. As a result, labels such as "Milk" and "milk" (letter case) or "1L

drink" and "ILDDrink" would be treated as unique. To be more precise: after the manual classification of raw data, performed by the Statistical Office in Opole, there were over 400 different product names for which COICOP groups could not be easily matched. In order to classify products into COICOP 5 and local (Polish) COICOP 6 groups, text mining methods based on regular expressions were used to (1) extract measurement units (e.g. weight, volume) from product labels; (2) remove special signs and some typos and (3) separate key words and characteristic phrases. The LASSO regression, proposed by Santosa and Symes (1986) and Tibshirani (1996), was used for further product classification. After the classification process, products were matched over time by using the *reclin* R package separately for each outlet. The matching process took into consideration EAN product codes, retailer codes, and additional product labels. After the product matching, two data filters were used to exclude extreme price changes and products with relatively low sales (*PriceIndices* R package). To be more precise: in the case of extreme price filter, those products whose monthly price increase was above 200% or whose monthly price decrease was above 75% were eliminated. In the case of a low-sales filter, products whose relative share in monthly sales was less than 1% were eliminated. As a consequence, the sample size was reduced by over 30% when using these data filters. Finally, the Fisher price index and proposed hybrid indices were calculated for the considered group of products for the prior period set to Dec, 2017 and for the variant with and without filters.

Results

Tab. 1 and Tab. 2 include values of discussed parameters (also with correlations between prices and quantities) and the Laspeyres, Fisher, hybrid and geometric hybrid price indices calculated for datasets on milk, yoghurt, coffee, and rice. The needed calculations were done for the case without and with the above-mentioned data filtering and for τ =Dec, 2017; s =Dec, 2018; t =Dec, 2019. For all 4 product groups, the correlations between their prices and quantities are negative, which is typical of the CPI basket, in both cases they are the strongest for rice and quite weak for coffee. However, this does not affect the weights $\gamma_0 = [\gamma_\tau^0, \gamma_s^0, \gamma_t^0]$, which are quite similar for the analyzed product groups, which in turn results from the closely related correlation values $\rho_{\tau,\tau}$, $\rho_{s,\tau}$ and $\rho_{t,\tau}$ within each product group. In these tables, the values of the above-mentioned price indices were determined for the annual time segment using information about consumption

lagged by one year compared to the base period ($\tau = \text{Dec, 2017}$). In the vast majority of cases (the exception is the case of rice without the filtering option), the Laspeyres index turned out to be an inferior approximation of the Fisher index to the hybrid index or/and the geometric hybrid price index. Sometimes, hybrid indices were amazingly close to the Fisher index (e.g.: for filtered data, the following results were obtained for milk: $P_F = 1.0746$ and $P_H^{\%} = 1.0745$ as well as for rice: $P_F = 1.0408$ and $P_{GH}^{\%} = 1.0409$). A more accurate comparison of the indices, determined for each month from the annual time window, is shown in Fig. 1 and Fig. 2. It can be observed here that in the case of data for milk and yogurt, the hybrid formulas clearly better approximate the Fisher index than the Laspeyres index. The comparison of yearly indices together with calculated correlations and weights can be found in Tab. 1 and Tab. 2. It can basically be concluded that the hybrid indices and the Laspeyres index perform almost equally well at the end of the analyzed year. Let us add, however, that the largest differences between the Laspeyres index and the hybrid indices arose for coffee data, and in the case of rice, the corresponding differences are negligible and smaller after the data filtering. Detailed results, presenting simple statistics (Mean Absolute Error — MSE) comparing the quality of approximations of the Fisher index, are presented in Tab. 3. Please note that the use of the geometric hybrid index always leads to the reduction of the substitution effect, i.e. it is always a better proxy for the Fisher index than the Laspeyres formula, and almost always its use is more effective than the use of the basic hybrid price index. The basic hybrid index also works very well, i.e. it reduces the substitution bias effect almost in each case (e.g. the exception is the rice case). Finally, please note that the above-presented conclusions do not really depend on whether or not data filters were used. Interestingly, the use of filters had a clear impact on the values of the correlation coefficients (compare Tab. 1 and Tab. 2), but the proportion between them was kept, so the hybrid index itself changed only slightly (see Fig. 1 and Fig. 2).

Discussion

First of all, it should be noted that due to the negative correlation between prices and quantities determined on the basis of the examined data sets, the Laspeyres index overestimated the actual inflation (if the Fisher index is the reference point), which results from so-called Bortkiewicz inequalities (Von der Lippe, 2007). The hybrid index (both basic and geometric) is not subject to this rule, i.e. in the study, their index values happened to be both above and below the Fisher index value, though in both cases they approx-

imated it quite accurately. Such in-plus and in-minus value oscillation of the hybrid indices throughout the year may have a beneficial effect if one wished to use chain versions of these indices, which has not been done in the paper. Chain versions of hybrid indices would not accumulate the distance from the Fisher index, while in the case of the Laspeyres chain index, its value would deviate from the Fisher index with each month. Another note concerns the update of weights in the CPI basket, the frequency of which varies from country to country. In Poland, the weight update is annual, so that the substitution bias (measured as the difference between the Laspeyres and Fisher index) is negligible (Hałka & Leszczyńska, 2011; Białek, 2014). In other countries, the frequency of updating CPI basket weights, and thus also the level of substitution bias, is far from ideal. The potential scale of this type of bias in measuring inflation can be found, for example, in: (Crawford, 1998). Nevertheless, given the high costs of the household budget survey which provides information on the level of consumption, perhaps an annual weight update is not necessary if the hybrid indices are used. The premise for this supposition is the fact that in the study the hybrid indices proved to be not inferior but usually superior approximations of the Fisher index to the Laspeyres index, and yet they used year-lagged consumption information (compared to the base period). The advantage of the hybrid indices is also that they use additional information about the level of correlation between prices and quantities. Earlier studies indicate that the accuracy of approximation of the Fisher index by indices using “lagged weights” (Young, Lowe) may depend on the level of the said correlation (Białek, 2017a, 2017b). Let us add, however, that the very measurement of correlations required by hybrid formulas may cause interpretation difficulties. We measure here, among others, correlations between prices and quantities from different periods (e.g. $\rho_{t,\tau}$ is the correlation coefficient determined for prices from the current period but quantities from the prior period).

Conclusions

To sum up, the proposed hybrid indices (or more precisely: the hybrid index in its basic (12) and geometric (13) version) seem to be useful in the practice of statistical offices in the framework of the so-called traditional data collection. In a situation where due to the cost of researching household budgets, the CPI basket update is annual or even less frequent, such indices can successfully compete with the Laspeyres index. In the completed study, these indices most often turned out to be a better approximation

of the Fisher index. The presented research confirms that especially the geometric hybrid price index is a very promising method leading to the reduction of the CPI substitution bias. These indices are also interesting from the theoretical point of view because they use additional knowledge about the correlation between prices and quantities of CPI basket components.

The only limitation here is the requirement that at least one of the three correlation coefficients used is nonzero, which in practice almost always occurs. Let us add, however, that the proposal of a system of weights based on correlations is only one of the possibilities and here the problem of optimal selection of these weights can still be considered as open.

Future research could also focus on taking into account all time moments from the entire interval $\langle \tau, t \rangle$. The author of this paper also plans to check how far the prior period can be removed in relation to the base period to continue to effectively approximate the Fisher index.

References

- Abe, N., & Prasada Rao, D. S. (2019). Multilateral Sato-Vartia index for international comparison of prices and real expenditures. *Economic Letters*, 183(C), 1-4. doi: 10.1016/j.econlet.2019.108535.
- Armknrecht, P., & Silver, M. (2012). Post-Laspeyres: the case for a new formula for compiling Consumer Price Indexes. *International Monetary Fund (IMF) Working Paper*, 12/105.
- Balk, M. (1995). Axiomatic price index theory: a survey. *International Statistical Review*, 63, 69-95. doi: 10.2307/1403778.
- Białek, J. (2012). Proposition of the general formula for price indices. *Communications in Statistics: Theory and Methods*, 41(5), 943-952. doi: 10.1080/03610926.2010.533238.
- Białek, J. (2014a). Simulation study of an original price index formula. *Communications in Statistics - Simulation and Computation*, 43(2), 285-297. doi: 10.1080/03610918.2012.700367.
- Białek, J. (2015). Generalization of the Divisia price and quantity indices in a stochastic model with continuous time. *Communications in Statistics: Theory and Methods*, 44(2), 309-328. doi: 10.1080/03610926.2014.968738.
- Białek, J. (2017a). Approximation of the Fisher price index by using the Lloyd–Moulton index: simulation study. *Communications in Statistics - Simulation and Computation*, 46(5), 3588-3598. doi: 10.1080/03610918.2015.1100737.
- Białek, J. (2017b). Approximation of the Fisher price index by using Lowe, Young and AG Mean indices. *Communications in Statistics - Simulation and Computation*, 46(8), 6454-6467. doi: 10.1080/03610918.2016.1205608.
- Białek, J. (2019a). Remarks on geo-logarithmic price indices. *Journal of Official Statistics*, 35(2), 287-317. doi: 10.2478/JOS-2019-0014.

- Białek, J. (2020a). Remarks on price index methods for the CPI measurement using scanner data. *Statistika – Statistics and Economy Journal*, 100(1), 54-69.
- Białek, J. (2020b). Comparison of elementary price indices. *Communications in Statistics - Theory and Methods*, 49(19), 4787-4803, doi: 10.1080/03610926.2019.1609035.
- Białek, J., & Bobel, A. (2019). *Comparison of price index methods for CPI measurement using scanner data*. Paper presented at the 16th Meeting of the Ottawa Group on Price Indices, Rio de Janeiro, Brazil.
- Biggeri, G., & Ferrari (Eds.) (2010). *Price indexes in time and space*. Berlin Heidelberg: Springer-Verlag. doi: 10.1007/978-3-7908-2140-6.
- Boskin, M. J., Dulberger, E. R., Gordon, R. J., Griliches, Z., & Jorgenson, D. (1996). Toward a more accurate measure of the cost of living. Final report to the Senate Finance Committee from the Advisory Commission to study the Consumer Price Index.
- Carli, G. (1804). Del valore e della proporzione de' metalli monetati. *Scrittori Classici Italiani di Economia Politica*, 13, 297-336.
- Chessa, A. G. (2017). *Comparisons of QU-GK indices for different lengths of the time window and updating methods*. Paper prepared for the second meeting on multilateral methods organized by Eurostat, Luxembourg, 14-15 March 2017. Statistics Netherlands.
- Clements, K. W., & Izan, H. Y. (1987). The measurement of inflation: a stochastic approach. *Journal of Business and Economic Statistics*, 5, 339-350. doi: 10.1080/07350015.1987.10509598.
- Consumer Price Index Manual. Theory and practice (2004). ILO/IMF/OECD /UNECE/Eurostat/The World Bank. International Labour Office (ILO), Geneva.
- Crawford, A. (1998). Measurement biases in the Canadian CPI: an update. *Bank of Canada Review*, Spring, 38–56.
- de Haan, J., & Krsinich, F. (2017). Time dummy hedonic and quality-adjusted unit value indices: do they really differ? *Review of Income and Wealth*, 64(4), 757-776. doi: 10.1111/roiw.12304.
- de Haan, J., Willenborg, L., & Chessa, A. G. (2016). *An overview of price index methods for scanner data*. Paper presented at the Meeting of the Group of Experts on Consumer Price Indices, 2-4 May 2016, Geneva, Switzerland.
- Diewert, W. E. (1976). Exact and superlative index numbers. *Journal of Econometrics*, 4, 114-145. doi: 10.1016/0304-4076(76)90009-9.
- Diewert W. E. (1993). The economic theory of index numbers: a survey. In W. E. Diewert & A. O. Nakamura (Eds.). *Essays in index number theory, vol. 1*. Amsterdam, 177-221.
- Diewert, W. E. (2005). Weighted country product dummy variable regressions and index number formulae. *Review of Income and Wealth*, 51, 561-570. doi: 10.1111/j.1475-4991.2005.00168.x.
- Diewert, W. E., & Fox, K. J. (2017). Substitution bias in multilateral methods for CPI construction using scanner data. *School of Economics, University of British Columbia, Discussion Paper*, 17-02.

- Dutot, C. F., (1738). Reflexions politiques sur les finances et le commerce. In *The Hague: Les Freres Vaillant et Nicolas Prevost*, Vol. 1.
- Eichhorn, W., & Voeller J. (1976). *Theory of the price index. Fisher's test approach and generalizations*. Berlin, Heidelberg, New York: Springer-Verlag.
- Eltető, Ö., & Köves, P. (1964). On a problem of index number computation relating to international comparisons. *Statisztikai Szemle*, 42, 507-518.
- Feenstra, R. C., & Reinsdorf, M. B. (2007). Should exact index numbers have standard errors? Theory and application to Asian growth. In E. R. Berndt & C. R. Hulten (Eds.). *Hard-to-measure goods and services: essays in honor of Zvi Griliches*. National Bureau of Economic Research.
- Fisher, I. (1922). *The making of index numbers*. Boston: Houghton Mifflin.
- Geary, R. G. (1958). A note on comparisons of exchange rates and purchasing power between countries. *Journal of the Royal Statistical Society Series A*, 121, 97-99. doi: 10.2307/2342991.
- Gini, C. (1931). On the circular test of index numbers. *Metron*, 9(9), 3-24.
- Greenlees, J. S. (2011). Improving the preliminary values of the chained CPI-U. *Journal of Economic and Social Measurement*, 36, 1-8. doi: 10.3233/jem-2011-0341.
- Hałka, A., & Leszczyńska, A. (2011). The strengths and weaknesses of the Consumer Price Index: estimates of the measurement bias for Poland. *Gospodarka Narodowa*, 9, 51-75. doi: 10.33119/gn/101086.
- Jevons, W. S. (1865). The variation of prices and the value of the currency since 1782. *Journal of the Statistical Society of London*, 28, 294-320.
- Jorgenson, D. W., & Slesnick D. T. (1983). Individual and social cost of living indexes. In W. E. Diewert & C. Montmarquette (Eds.). *Price level measurement: proceedings of a conference sponsored by Statistics Canada*. Ottawa: Statistics Canada, 241-336.
- Juszczak, A. (2020). Estimation of the optimal parameter of Delay in Young and Lowe indices in the Fisher index approximation. *Statistika – Statistics and Economy Journal*, 1, 32-53.
- Khamis, S. H. (1972). A new system of index numbers for national and international purposes. *Journal of the Royal Statistical Society Series A*, 135, 96-121. doi: 10.2307/2345041.
- Krsinich, F. (2014). *The FEWS index: fixed effects with a window splice – non-revisable quality-adjusted price indices with no characteristic information*. Paper presented at the meeting of the group of experts on consumer price indices, 26-28 May 2014, Geneva, Switzerland.
- Laspeyres, E. (1871). Die Berechnung einer mittleren Waarenpreissteigerung. *Jahrbücher für Nationalökonomie und Statistik*, 16, 296-314.
- Lent, J., & Dorfman, A. H. (2009). Using a weighted average of base period price indexes to approximate a superlative index. *Journal of Official Statistics*, 25(1), 139-149.
- Lloyd, P. J. (1975). Substitution effects and biases in non true price indices. *American Economic Review*, 65, 301-313.

- Martini, M. (1992). A general function of axiomatic index numbers. *Journal of the Italian Statistics Society*, 1(3), 359-376. doi: 10.1007/bf02589086.
- Mehrhooff, J. (2019). *Towards a new paradigm for scanner data price indices: applying big data techniques to big data*. Paper presented at the 16th Meeting of the Ottawa Group on Price Indices, Rio de Janeiro, Brazil.
- Moulton, B. R. (1996). *Constant elasticity cost-of-living index in share-relative form*. Washington DC: U. S. Bureau of Labour Statistics, Mimeograph.
- Olt, B. (1996). *Axiom und Struktur in der statistischen Preisindex theorie*. Frankfurt: Peter Lang.
- Pollak, R. A. (1989). *The theory of the cost-of-living index*. Oxford: Oxford University Press.
- Santosa, F., & Symes, W. W. (1986). Linear inversion of band-limited reflection seismograms. *Journal on Scientific and Statistical Computing*, 7(4), 1307-1330. doi: 10.1137/0907087.
- Selvanathan, E. A., & Prasada Rao, D. S. (1994). *Index numbers: a stochastic approach*. Ann Arbor: University of Michigan Press.
- Shapiro, M. D., & Wilcox, D. W. (1997). Alternative strategies for aggregating prices in the CPI. *Federal Reserve Bank of St. Louis Review*, 79(3), 113-125.
- Szulc, B. (1964). Indices for multiregional comparisons. *Przegląd Statystyczny*, 3, 239-254.
- Tibshirani, R., Bien, J., Friedman, J., Hastie, T., Simon, N., Taylor, J., & Tibshirani, R. J. (2012). Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society: Series B*, 74(2), 245-266. doi: 10.1111/j.1467-9868.2011.01004.x.
- White, A. G. (1999). Measurement biases in Consumer Price Indexes. *International Statistical Review*, 67(3), 301-325. doi: 10.1111/j.1751-5823.1999.tb00451.x.
- Van Loon, K. V., & Roels, D. (2018). *Integrating big data in the Belgian CPI*. 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*. Paper presented at the 17th Meeting of the Ottawa Group on Price Indices, 8–10 May 2019, Rio de Janeiro, Brasil.
- Von der Lippe, P. (2007). *Index theory and price statistics*. Peter Lang, Frankfurt, Germany. doi: 10.3726/978-3-653-01120-3.
- Zhang, L., Johansen, I., & Nygaard, R. (2019). Tests for price indices in a dynamic item universe. *Journal of Official Statistics*, 35(3), 683-697. doi: 10.2478/jos-2019-0028.

Acknowledgments

The author would like to thank the National Science Center in Poland for financing this publication (grant no. 2017/25/B/HS4/00387).

Annex

Table 1. Values of discussed parameters and the Laspeyres, Fisher, hybrid and geometric hybrid price indices calculated for datasets on milk, yoghurt, coffee, and rice (without filtering for: τ =Dec, 2017; s =Dec, 2018; t =Dec, 2019)

parameter	milk	yoghurt	coffee	rice
$\rho_{\tau,\tau}$	-0.1149	-0.4105	-0.2038	-0.4446
$\rho_{s,\tau}$	-0.1234	-0.3763	-0.1911	-0.4203
$\rho_{t,\tau}$	-0.1177	-0.4088	-0.1820	-0.4004
γ_{τ}^0	0.3227	0.3433	0.3532	0.3513
γ_s^0	0.3466	0.3147	0.3312	0.3322
γ_t^0	0.3307	0.3420	0.3156	0.3165
P_F	1.0699	1.0648	0.9792	1.0284
P_{La}	1.0713	1.1087	1.0058	1.0468
$P_H^{\gamma_0}$	1.0707	1.0578	0.9846	1.0514
$P_{GH}^{\gamma_0}$	1.0655	1.0449	0.9783	1.0479

Source: own calculations in the *PriceIndices* package.

Table 2. Values of discussed parameters and the Laspeyres, Fisher, hybrid and geometric hybrid price indices calculated for datasets on milk, yoghurt, coffee, and rice (with filtering for: τ =Dec, 2017; s =Dec, 2018; t =Dec, 2019)

parameter	milk	yoghurt	coffee	rice
$\rho_{\tau,\tau}$	-0.4414	-0.4314	-0.3221	-0.7050
$\rho_{s,\tau}$	-0.4375	-0.3823	-0.3353	-0.6281
$\rho_{t,\tau}$	-0.4456	-0.4042	-0.3310	-0.5593
γ_{τ}^0	0.3332	0.3542	0.3258	0.3725
γ_s^0	0.3303	0.3139	0.3392	0.3319
γ_t^0	0.3365	0.3319	0.3350	0.2956
P_F	1.0746	1.0615	0.9926	1.0408
P_{La}	1.0799	1.1023	1.0119	1.0488

Table 2. Continued

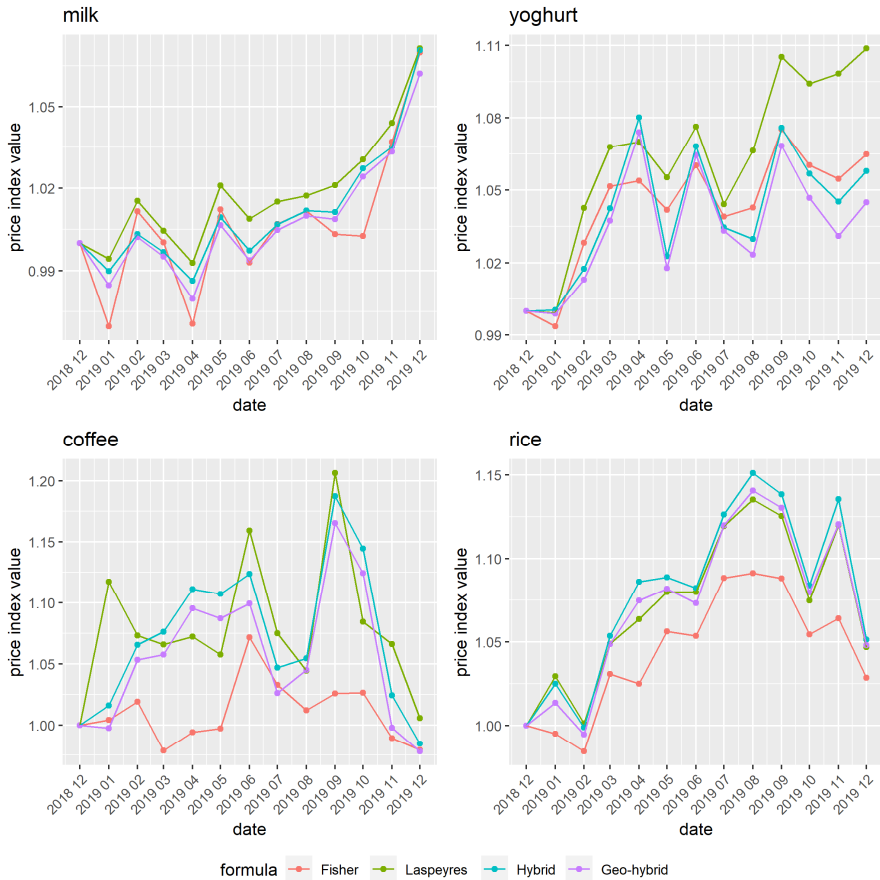
parameter	milk	yoghurt	coffee	rice
$P_H^%$	1.0745	1.0521	0.9738	1.0444
$P_{GH}^%$	1.0737	1.0398	0.9706	1.0409

Source: own calculations in the *PriceIndices* package.

Table 3. Values of mean absolute errors (MAEs) calculated for the Laspeyres, hybrid and geometric hybrid price indices compared with the Fisher index (the target index) for datasets on milk, yoghurt, coffee, and rice (without and with filtering for: τ =Dec, 2017; s =Dec, 2018; t =Dec, 2019)

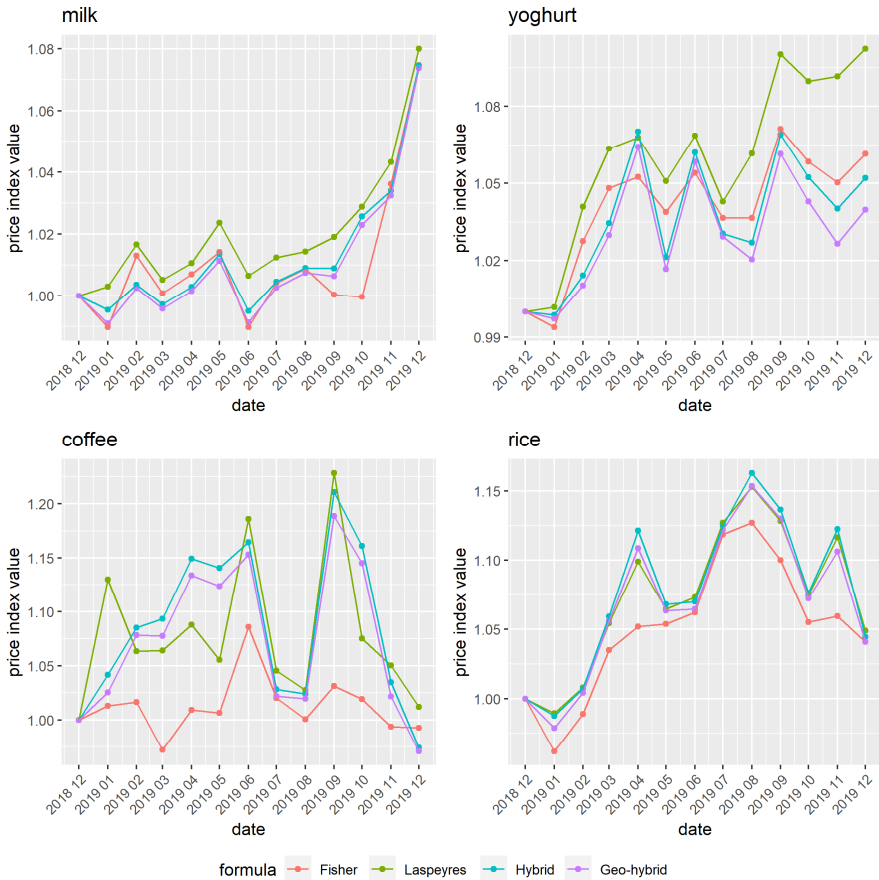
Datasets are not filtered		Datasets are filtered	
milk			
MAE (Laspeyres)	0.0118	MAE (Laspeyres)	0.0099
MAE (Hybrid)	0.0072	MAE (Hybrid)	0.0051
MAE (Geometric Hybrid)	0.0070	MAE (Geometric Hybrid)	0.0049
yoghurt			
MAE (Laspeyres)	0.2725	MAE (Laspeyres)	0.0172
MAE (Hybrid)	0.1943	MAE (Hybrid)	0.0083
MAE (Geometric Hybrid)	0.1992	MAE (Geometric Hybrid)	0.0081
coffee			
MAE (Laspeyres)	0.0690	MAE (Laspeyres)	0.0664
MAE (Hybrid)	0.0624	MAE (Hybrid)	0.0757
MAE (Geometric Hybrid)	0.0481	MAE (Geometric Hybrid)	0.0648
rice			
MAE (Laspeyres)	0.0266	MAE (Laspeyres)	0.0266
MAE (Hybrid)	0.0311	MAE (Hybrid)	0.0326
MAE (Geometric Hybrid)	0.0237	MAE (Geometric Hybrid)	0.0266

Figure 1. Comparison of the Laspeyres, Fisher, hybrid and geometric hybrid price indices made for scanner datasets on milk, yoghurt, coffee, and rice (the case without data filtering)



Source: own calculations in the *PriceIndices* package.

Figure 2. Comparison of the Laspeyres, Fisher, hybrid and geometric hybrid price indices made for scanner datasets on milk, yoghurt, coffee, and rice (the case with data filtering)



Source: own calculations in the *PriceIndices* package.