ORIGINAL PAPER

A new price index for multi-period and multilateral comparisons

Mario Faliva[1] · Consuelo Rubina Nava[2] · Maria Grazia Zoia[1]

Abstract
Within the stochastic approach, this paper establishes a closed-form solution to the price index problem for an arbitrary number of periods or countries. The index's reference basket merges the intersections of all couples of baskets in all periods/countries and provides an effective commodity coverage. Under spherical regression errors, the index satisfies the Geary–Khamis equation system and, as such, offers a general and compact representation of the latter as well as the inferential framework as a dowry. Furthermore, by relaxing sphericalness in favor of a more realistic assumption of commodity-dependent variances, a broader result is achieved. The solution to the price index problem thus obtained encompasses the Geary–Khamis formulation and sows the seeds to further advances.

1 Introduction

The paper works out a multi-period/multilateral price index to effectively compare sets of commodities over time and/or across countries. This is done by using a regression framework and a reference basket based on the set of all commodities that

✉ Maria Grazia Zoia
maria.zoia@unicatt.it

Mario Faliva
mario.faliva@unicatt.it

Consuelo Rubina Nava
consuelorubina.nava@unito.it

1 Faculty of Economics, Università Cattolica del Sacro Cuore, Milan, Italy

2 Department of Economics and Statistics "Cognetti de Martiis", Turin, Italy

appear at least in two periods/countries. A linear model in the deflator indexes and reference prices is specified, and estimation is performed by the method-of-averages argument. The resulting price index estimator follows as a by-product and enjoys fitting properties in between least squares and least absolute values. The price index constructed in this manner belongs to the family of stochastic indexes, which can be traced back to the works of Jevons (1863, 1869) and Edgeworth (1887, 1925). Its computation requires the knowledge of the quantities and values of the commodities included in the basket. Neither the model's specification nor the estimation method used explicitly calls for prices. This aspect becomes particularly useful when the prices of some commodities in specific periods/countries are not known. In the multi-period case, this can occur when some commodities enter/exit the basket due to a shift in consumer attitude, while in a multilateral perspective, it could happen when a commodity is not present or ceases to be present on the market in a given country. The reference basket of the index at hand is the result of the union of the intersections in pairs of the specific baskets for each period/country. The absence of a commodity in a given period/country simply means that both its quantity and value are set equal to zero in that period/country.

Thus, just like hedonic (Pakes 2003; Brachinger et al. 2018), GESKS (Balk 2012) and country/time-product-dummy (CPD/TPD) approaches with an incomplete price tableau (Rao and Hajargasht 2016; Weinand 2021), the index does not drop items which are not present in all countries/periods. A commodity contributes to the index's construction provided it occurs with non-null quantities in at least two periods/countries. The reference basket is, therefore, more inclusive and representative than baskets commonly referred to in the extant literature. Consequently, the index proves effective in a multi-period and/or multilateral framework and compares favorably to the Rao and Hajargasht (2016), GEKS, Ivancic et al. (2011) indexes.

Under sphericalness of errors in the regression model underlying the index estimator, the latter tallies with the Geary–Khamis (GK) index in a temporal setting and provides both a general closed-form representation of the latter and an inferential statistical apparatus as a dowry.

Moreover, the paper proposes an additional index under the assumption of commodity-dependent variances of the regression errors. This provides a more extended solution to the price index problem, which encompasses the Geary–Khamis Index as a special case and paves the way to further generalizations. In addition, the so-called reference prices, namely the prices expected to be paid for commodities in the base period/country, are easily obtainable as a spin-off of the regression estimation and can be conveniently used to evaluate the prices of those commodities that, for whatever reason, are not observable in a given period/country.

The paper is organized as follows. Section 2 formally states the problem of finding a multi-period/multilateral price index starting from the quantities and values of the commodities in each period/country in the basket/index. The regression model uses deflators and reference prices as parameters to be estimated. Section 3 shows how the deflators are estimated using the method of averages and how the estimator of the price indexes is obtained from the deflator estimates as a by-product. In Sect. 4, the hypotheses on the errors of the parent regression model are relaxed allowing the variance to possibly be commodity dependent. In Sect. 5, the novel

index formula and the TPD index are compared, in the case of a complete and an incomplete price tableau, using simulated data. Here we show the effectiveness of our method of estimating missing prices using reference prices in comparison with the standard techniques based on the imputation of missing values. Some conclusive remarks are made in Sect. 6. An appendix provides the proofs of the statements in Sect. 3.

2 Formulation of the price index problem

In order to set up the regression framework needed to work out the index, let us consider the $N \times T$ matrices $V = [v_{n,t}]$, $Q = [q_{n,t}]$ and $P = [p_{n,t}]$ whose entries are values, quantities and prices, respectively, of a basket of N goods in T periods (or countries). Such a basket is the union of the intersections in pairs of the baskets in the T periods (or countries), and we assume that it covers all (and only those) commodities which occur with non-null quantities in at least two periods (or countries). The matrices V, Q and P satisfy the equality

$$\underset{(N,T)}{V} = \underset{(N,T)}{Q} * \underset{(N,T)}{P} \tag{1}$$

where $*$ is the Hadamard-product symbol. The price index problem can be read as the problem of approximating the matrix P by the outer product $\pi \lambda'$ of a vector of reference prices, $\pi > \mathbf{0}_N$, and a vector of price indexes, $\lambda > \mathbf{0}_T$, that is,

$$\underset{(N,T)}{P} \approx \underset{(N,1)(1,T)}{\pi \ \lambda'}. \tag{2}$$

So, the equality in Eq. (1) can be reformulated as follows

$$V = Q * \pi \lambda' + H = D_\pi Q D_\lambda + H \tag{3}$$

where D_a denotes a diagonal matrix whose diagonal entries are the elements of the vector a and the matrix H accounts for discrepancies[1]. Post-multiplying Eq. (3) by $D_\phi = D_\lambda^{-1}$, yields

$$V D_\phi = D_\pi Q + E \tag{4}$$

where $E = H D_\phi$ is a matrix of disturbance terms. Henceforth, in the wake of Theil (1960), a sphericalness (i.e., constant variance and incorrelation) assumption will be made for the error components in (4), namely for the entries of the matrix E. The issue of possibly relaxing this assumption by dropping the hypothesis of constant variance across commodity errors will be addressed in Sect. 4. The vector ϕ, whose elements are the reciprocals of the elements of the price-index vector λ, is the deflator/exchange rate vector. Taking $t = 1$ as the base period (country), namely $\phi_1 = \lambda_1 = 1$, and partitioning V, Q, E, ϕ and D_ϕ as follows

---

[1] Use has been made of the relationship $A * (bc') = D_b A D_c$.

$$V = \begin{bmatrix} \underset{(N,1)}{v_1}, & \underset{(N,T-1)}{V_2} \end{bmatrix}, \; Q = \begin{bmatrix} \underset{(N,1)}{q_1}, & \underset{(N,T-1)}{Q_2} \end{bmatrix}, \; E = \begin{bmatrix} \underset{(N,1)}{\varepsilon_1}, & \underset{(N,T-1)}{E_2} \end{bmatrix}$$

$$\underset{(T,1)}{\phi} = \begin{bmatrix} 1 \\ \underset{(T-1,1)}{\varphi} \end{bmatrix}, D_\phi = \begin{bmatrix} 1 & \underset{(1,T-1)}{0} \\ \underset{(T-1,1)}{0} & \underset{(T-1,T-1)}{D_\varphi} \end{bmatrix}, \qquad (5)$$

we pass from the matrix Eq. (4) to the system

$$\begin{cases} \underset{(N,1)}{v_1} = \underset{(N,N)}{D_\pi} \underset{(N,1)}{q_1} + \underset{(N,1)}{\varepsilon_1} \\ \underset{(N,T-1)}{V_2} \underset{(T-1,T-1)}{D_\varphi} = \underset{(N,N)}{D_\pi} \underset{(N,T-1)}{Q_2} + \underset{(N,T-1)}{E_2} \end{cases}. \qquad (6)$$

The latter can be written in staked form as follows

$$y = X\beta + \varepsilon. \qquad (7)$$

by setting

$$\underset{(NT,1)}{y} = \begin{bmatrix} v_1 \\ 0 \end{bmatrix}, \underset{(NT,T-1+N)}{X} = \begin{bmatrix} 0 & D_{q_1} \\ -(I_{T-1} \otimes V_2)R'_{T-1} & (Q'_2 \otimes I_N)R'_N \end{bmatrix},$$

$$\underset{(T-1+N,1)}{\beta} = \begin{bmatrix} \varphi \\ \pi \end{bmatrix}, \underset{(NT,1)}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ vecE_2 \end{bmatrix} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}. \qquad (8)$$

Here $\otimes$ is the Kronecker product symbol, *vec* is the staking operator, and R_j is the following matrix (Faliva 1996)

$$\underset{(j \times j^2)}{R_j} = \begin{bmatrix} \underset{(1,j)}{e'_1} \otimes \underset{(1,j)}{e'_1} \\ \underset{(1,j)}{e'_2} \otimes \underset{(1,j)}{e'_2} \\ \dots \\ \underset{(1,j)}{e'_j} \otimes \underset{(1,j)}{e'_j} \end{bmatrix}, \qquad (9)$$

where e_i denotes the *i*th elementary vector[2]. The model in Eq. (7) together with a sphericalness hypothesis for the error terms is a classical linear regression model. The estimation of the parameters will be accomplished in the next section by using the Method of Averages, in short MA, (Kveǒtoň 1987). The estimator so obtained turns out to be an instrumental-variable (IV) estimator (Goldberger 1964) with a binary instrument matrix and enjoys the desirable properties of both MA and IV inferential procedures.

---

[2] Use has been made of the equalities

$$D_\pi q_1 = D_{q_1}\pi$$
$$vec(AD_bC) = (C' \otimes A)vec(D_b) = (C' \otimes A)R'_N b.$$

3 The solution of the index estimation problem and its meaning

In the following, the MA is applied to estimate the price-indexes via the intermediate estimation of the deflator indexes. To this end, let us look at

$$X\beta \approx y \tag{10}$$

as an over-identified system of linear equations and solve the derived system

$$L'X\beta = L'y, \tag{11}$$

where L is a binary matrix of the same dimensions as X which satisfies the rank condition

$$r(L'X) = N + T - 1. \tag{12}$$

If the binary matrix $X^b = [x_{ij}^b]$, $x_{ij}^b = 1$ if $x_{ij} \neq 0$ and $x_{ij}^b = 0$ otherwise, associated with X satisfies the rank condition (12), then this matrix provides a convenient choice for L. This leads to the linear system

$$(X^b)'X\beta = (X^b)'y \tag{13}$$

whose solution

$$\widehat{\beta} = ((X^b)'X)^{-1}(X^b)'y \tag{14}$$

gives the intended index estimator. Actually, the solution (14) occurs to be an instrumental variables (*IV*) estimator (Goldberger 1964), with the columns of X^b acting as instruments. Going back to the linear model in Eq. (7) and assuming sphericalness for the error terms, i.e.,

$$\mathbb{E}(\epsilon) = \mathbf{0}_{NT}, \mathbb{E}\{\epsilon\epsilon'\} = \sigma^2 I_{NT}, \tag{15}$$

the dispersion matrix of the estimator (14) is given by

$$\mathbf{\Sigma}(\widehat{\beta}) = \sigma^2((X^b)'X)^{-1}(X^b)'X^b(X'X^b)^{-1} \tag{16}$$

and the error variance σ^2 is estimated by

$$\widehat{\sigma}^2 = \frac{1}{NT - N - T + 1}(y - X\widehat{\beta})'(y - X\widehat{\beta}). \tag{17}$$

The estimator of the price index vector follows as a by-product of the estimator in Eq. (14), and the statistical properties of the former can be derived from those of the latter, accordingly. The estimator in Eq. (14) crucially rests on the following

Lemma 1 *The rank condition*

$$r((X^b)'X)) = N + T - 1 \tag{18}$$

holds true for X in Eq. (8).

Proof See Appendix.   □

On this premise, we can establish the following

Theorem 1 *The estimator* $\hat{\boldsymbol{\varphi}} = [\boldsymbol{I}_{T-1}, \boldsymbol{0}]\hat{\boldsymbol{\beta}}$ *of the deflator vector is given by*

$$\hat{\boldsymbol{\varphi}} = (\boldsymbol{D}_w - \boldsymbol{Q}_2'\boldsymbol{D}_{\bar{q}}^{-1}\boldsymbol{V}_2)^{-1}\,\boldsymbol{Q}_2'\boldsymbol{D}_{\bar{q}}^{-1}\boldsymbol{v}_1 = \tag{19}$$

$$= \begin{bmatrix} \sum_{i=1}^{N}\left(v_{i2} - \frac{q_{i2}v_{i2}}{\sum_{t=1}^{T}q_{it}}\right) & \cdots & -\sum_{i=1}^{N}\left(\frac{q_{i2}v_{iT}}{\sum_{t=1}^{T}q_{it}}\right) \\ \cdots & \cdots & \cdots \\ -\sum_{i=1}^{N}\left(\frac{q_{iT}v_{i2}}{\sum_{t=1}^{T}q_{it}}\right) & \cdots & \sum_{i=1}^{N}\left(v_{iT} - \frac{q_{iT}v_{iT}}{\sum_{t=1}^{T}q_{it}}\right) \end{bmatrix}^{-1}\begin{bmatrix} \sum_{i=1}^{N}\left(\frac{q_{i2}v_{i1}}{\sum_{t=1}^{T}q_{it}}\right) \\ \cdots \\ \sum_{i=1}^{N}\left(\frac{q_{iT}v_{i1}}{\sum_{t=1}^{T}q_{it}}\right) \end{bmatrix} \tag{20}$$

where $\boldsymbol{D}_w$ *and* $\boldsymbol{D}_{\bar{q}}$ *are diagonal matrices with the elements of the vectors*

$$\boldsymbol{w} = \boldsymbol{V}_2'\boldsymbol{u}_N \quad \text{and} \quad \bar{\boldsymbol{q}} = \boldsymbol{Q}\boldsymbol{u}_T \tag{21}$$

as diagonal entries, respectively; $\boldsymbol{u}_N$ *and* $\boldsymbol{u}_T$ *are vectors of 1's with N and T components, respectively. An estimator of the variance–covariance matrix of the estimator is*

$$\hat{\boldsymbol{\Sigma}}(\hat{\boldsymbol{\varphi}}) = \hat{\sigma}^2(\boldsymbol{D}_w - \boldsymbol{Q}_2'\boldsymbol{D}_{\bar{q}}^{-1}\boldsymbol{V}_2)^{-1}\,\boldsymbol{Q}_2'\boldsymbol{D}_{\bar{q}}^{-2}\boldsymbol{Q}_2\,(\boldsymbol{D}_w - \boldsymbol{Q}_2'\boldsymbol{D}_{\bar{q}}^{-1}\boldsymbol{V}_2)^{-1} \tag{22}$$

where

$$\hat{\sigma}^2 = \frac{1}{NT - N - T + 1}\left\{\boldsymbol{v}_1'\boldsymbol{v}_1 - 2[\boldsymbol{0}', \boldsymbol{v}_1'](\boldsymbol{X}'\boldsymbol{X}^b)^{-1}\begin{bmatrix}\boldsymbol{0}\\\boldsymbol{D}_{q_1}\boldsymbol{v}_1\end{bmatrix}\right.$$
$$\left. +[\boldsymbol{0}', \boldsymbol{v}_1'](\boldsymbol{X}'\boldsymbol{X}^b)^{-1}\boldsymbol{X}'\boldsymbol{X}((\boldsymbol{X}^b)'\boldsymbol{X})^{-1}\begin{bmatrix}\boldsymbol{0}\\\boldsymbol{v}_1\end{bmatrix}\right\} \tag{23}$$

is an estimator of the error variance.

Proof See Appendix.   □

In what follows, we will refer to $\hat{\lambda}_t$ as the reciprocal of the estimator of the parent deflator vector $\hat{\boldsymbol{\varphi}}_{t-1}$. In this connection, we establish the following

Corollary 1 *The estimator of the price index in period (or country) t,* $2 \leq t \leq T$*, is given by*

$$\hat{\hat{\lambda}}_t = (\boldsymbol{e}_{t-1}(\boldsymbol{D}_w - \boldsymbol{Q}_2'\boldsymbol{D}_{\bar{q}}^{-1}\boldsymbol{V}_2)^{-1}\,\boldsymbol{Q}_2'\boldsymbol{D}_{\bar{q}}^{-1}\boldsymbol{v}_1))^{-1} \tag{24}$$

where $\boldsymbol{e}_{t-1}$ *is the* $t - 1^{th}$ *elementary vector of* $T - 1$ *components and* $\hat{\boldsymbol{\varphi}}_{t-1}$ *is the* $t - 1$ *-th component of the estimator* (19)*. The estimated variance of* $\hat{\lambda}_t$ *is approximated by*

$$\widehat{var}(\hat{\lambda}_t) \approx \hat{\varphi}_{t-1}^{-4}\, \widehat{var}(\hat{\varphi}_{t-1}) = \hat{\varphi}_{t-1}^{-4}\, e'_{t-1}\hat{\Sigma}(\hat{\varphi})e_{t-1} \tag{25}$$

where $\hat{\Sigma}(\hat{\varphi})$ is the matrix (22).

Proof See Appendix. □

It can be shown that under sphericalness the analysis of the residuals of the regression model in Eq. (7) associated with the estimator in Eq. (14) leads to the systems of equations which determine the Geary–Khamis (GK) index. Thus, the estimator is a closed-form expression of the GK index. Although the GK index has been widely investigated (see, e.g., Diewert and Fox (2022, 2017); Balk (2012); Heston and Lipsey (2007)), a closed-form formula of the index for the general case of an arbitrary number of periods and/or countries is lacking up to now. The index formula devised in this paper provides the intended result within a regression model framework with the inherent inferential statistical toolkit as a dowry. In order to prove that $\hat{\lambda}_t$ represents a closed-form expression of the GK index, check that

$$(X^b)'e = \mathbf{0}_{N+T-1} \tag{26}$$

holds true for

$$\underset{(NT,1)}{e} = y - X\hat{\beta} \tag{27}$$

where

$$\hat{\beta} = \begin{bmatrix} \hat{\varphi} \\ \hat{\pi} \end{bmatrix} \begin{bmatrix} \lambda_2^{-1} \\ \cdots \\ \hat{\lambda}_T^{-1} \\ \hat{\pi}_1 \\ \cdots \\ \hat{\pi}_N \end{bmatrix} = ((X^b)'X)^{-1}(X^b)'y. \tag{28}$$

Simple computations, bearing in mind Eq. (50) and (61) in Appendix, show that

$$(X^b)'e = (X^b)'y - (X^b)'X\hat{\beta} = \tag{29}$$

$$= \begin{bmatrix} \underset{((T-1),1)}{\mathbf{0}} \\ \underset{(N,1)}{v_1} \end{bmatrix} - \begin{bmatrix} D_w & -Q'_2 \\ -V_2 & D_{\tilde{q}} \end{bmatrix} \begin{bmatrix} \underset{((T-1),1)}{\hat{\varphi}} \\ \underset{(N,1)}{\hat{\pi}} \end{bmatrix} = \begin{bmatrix} \underset{((T-1),1)}{0} \\ \underset{(N,1)}{\mathbf{0}} \end{bmatrix}. \tag{30}$$

The latter, together with Eq. (26), leads to the pair of equation systems

$$D_w\hat{\varphi} = Q'_2\pi \tag{31}$$

$$v_1 + V_2\hat{\varphi} = D_{\tilde{q}}\hat{\pi} \tag{32}$$

that can be rewritten as

$$\sum_{i=1}^{N} v_{it}\widehat{\lambda}_t^{-1} = \sum_{i=1}^{N} q_{it}\widehat{\pi}_i, t = 2, \ldots, T \tag{33}$$

$$v_{i1} + \sum_{t=2}^{T} v_{it}\widehat{\lambda}_t^{-1} = \sum_{t=1}^{T} q_{it}\widehat{\pi}_i\, i = 1, \ldots, N. \tag{34}$$

Solving Eqs. (33) and (34) for $\widehat{\lambda}_t$ and $\widehat{\pi}_i$ yield

$$\widehat{\lambda}_t = \frac{\sum_{i=1}^{N} v_{it}}{\sum_{i=1}^{N} q_{it}\widehat{\pi}_i} \tag{35}$$

$$\widehat{\pi}_i = \frac{\sum_{t=1}^{T} \frac{v_{it}}{\widehat{\lambda}_t}}{\sum_{t=1}^{T} q_{it}}. \tag{36}$$

Under $\lambda_1 = 1$, Eqs. (35) and (36) read as the equation systems of the GK index in the temporal setting and the index can be obtained, accordingly (as noticed by an anonymous referee we are indebted to).

4 Dropping the assumption of constant-variance errors

In the previous section, the estimation procedure of the deflator vector $\boldsymbol{\varphi}$ (and eventually of the price index λ) has been performed under the assumption of error sphericalness which embodies both uncorrelation and constant-variance of disturbances. Leaving apart the issue of dependence, in particular correlation, that we exclude from our analysis, let us investigate the assumption of constant variance. A hypothesis of constant variance over time for errors is tenable by virtue of the argument that the model specification is the outcome of a deflating transformation via $\boldsymbol{D}_\phi = \boldsymbol{D}_\lambda^{-1}$. No a-priori justification can be advanced for a constant-variance hypothesis for different commodities, if not computational convenience. So it is worth considering the issue more deeply. The analysis cannot but start from the residuals corresponding to the estimator $\widehat{\boldsymbol{\beta}}$, that is

$$\widehat{\boldsymbol{\varepsilon}}_{(TN,1)} = \boldsymbol{y} - \boldsymbol{X}((\boldsymbol{X}^b)'\boldsymbol{X})^{-1}(\boldsymbol{X}^b)'\boldsymbol{y}. \tag{37}$$

As a simple computation shows the (sub)vector of the residuals referable to the nth commodity over the time span $1 \leq t \leq T$ is given by

$$\widehat{\zeta}_n \underset{(T,1)}{=} J_n \underset{(T,TN)(TN,1)}{\widehat{\varepsilon}} = \begin{bmatrix} \widehat{\varepsilon}_n \\ \widehat{\varepsilon}_{n+N} \\ \vdots \\ \widehat{\varepsilon}_{n+(T-1)N} \end{bmatrix} \tag{38}$$

where J_n is the selection matrix

$$\underset{(T,TN)}{J_n} = \begin{bmatrix} \widehat{e}'_n \\ \widehat{e}'_{n+N} \\ \vdots \\ \widehat{e}'_{n+(T-1)N} \end{bmatrix} \tag{39}$$

with e_j denoting the jth elementary vector of TN components. Accordingly, an estimator of the variance σ_n^2 of the T errors referable to the nth commodity is given by

$$\widehat{\sigma}_n^2 = \frac{1}{T} \underset{(1,T)(T,1)}{\widehat{\zeta}' \widehat{\zeta}} = \widehat{\sigma}^2 \widehat{\varsigma}_n^2 \tag{40}$$

with $\widehat{\sigma}^2$ given by Eq. (17) and $\widehat{\varsigma}_n^2$ ensuing as a by-product. It follows that the former assumption of a scalar dispersion matrix for the errors no longer holds and it must be replaced by the following specification

$$E(\varepsilon\varepsilon') = \sigma^2(I_T \otimes D_{\widehat{\varsigma}}^2) \tag{41}$$

where $D_{\widehat{\varsigma}}$ is the $N \times N$ diagonal matrix whose diagonal entries are the squares of the scalars $\widehat{\varsigma}_1, \widehat{\varsigma}_2, \ldots, \widehat{\varsigma}_N$. Under Eq. (41), the model

$$y = X\beta + \varepsilon \tag{42}$$

is no longer a classical linear model. Nevertheless, it can easily be brought back to a classical model by premultiplying both sides of Eq. (42) by the matrix $(I_T \otimes D_{\widehat{\varsigma}}^{-1})$, which yields the specification

$$\widetilde{y} = \widetilde{X}\beta + \widetilde{\varepsilon} \tag{43}$$

where

$$\begin{aligned} \widetilde{y} &= (I_T \otimes D_{\widehat{\varsigma}}^{-1})y, \\ \widetilde{X} &= (I_T \otimes D_{\widehat{\varsigma}}^{-1})X, \\ \widetilde{\varepsilon} &= (I_T \otimes D_{\widehat{\varsigma}}^{-1})\varepsilon, \end{aligned} \tag{44}$$

with $\widetilde{\varepsilon}$ enjoying the sphericalness property

$$E(\widetilde{\varepsilon}) = 0, \quad E(\widetilde{\varepsilon}\widetilde{\varepsilon}') = \sigma^2 I_{NT}. \tag{45}$$

Noting that

$$\tilde{X}^b = X^b \tag{46}$$

the vector $\boldsymbol{\beta}$ can be newly estimated via the moving-average approach, with $X^b = \tilde{X}^b$ playing the role of the instrumental variable matrix. Eventually, we get the estimator

$$\tilde{\boldsymbol{\beta}} = [(\tilde{X}^b)'\tilde{X}]^{-1}(\tilde{X}^b)'\tilde{y} = [(X^b)'(I_T \otimes D_{\hat{\varsigma}}^{-1})X)]^{-1}(X^b)'(I_T \otimes D_{\hat{\varsigma}}^{-1})y. \tag{47}$$

This shows that the sphericalness assumption can be relaxed in the case of interest. This paves the way to further extensions, if required.

5 A simulation based on log-normal random draws

In this section we illustrate the performance of the index developed in Sect. 3, called MA index hereafter, through three simulated examples. The scope of this analysis is to investigate, in a comparative manner, the capability of both the MA and the time dummy product (TPD) (de Haan et al. 2020) index to reproduce the "true" index values, λ, in a multi-period perspective. To this aim, let us assume that quantities, Q, reference prices, $\boldsymbol{\pi}$, and price indexes, λ, of four commodities over six periods are specified as follows

$$Q = \begin{bmatrix} 5.00 & 7.00 & 7.00 & 8.00 & 10.00 & 12.00 \\ 15.00 & 20.00 & 21.00 & 24.00 & 25.00 & 27.00 \\ 25.00 & 22.00 & 20.00 & 23.00 & 23.00 & 25.00 \\ 5.00 & 6.00 & 6.00 & 8.00 & 10.00 & 15.00 \end{bmatrix} = \begin{bmatrix} \boldsymbol{q}_1, \cdots, \boldsymbol{q}_6 \\ {\scriptstyle(4,1)} \quad {\scriptstyle(4,1)} \end{bmatrix},$$

$$\lambda = \begin{bmatrix} 1.00 \\ 1.05 \\ 1.11 \\ 1.15 \\ 1.18 \\ 1.25 \end{bmatrix}, \quad \boldsymbol{\pi} = \begin{bmatrix} 3.5 \\ 2.5 \\ 1.7 \\ 2.9 \end{bmatrix}.$$

Then, with these data at hand, the values, V, have been computed as in (3) where, the random terms, H, without lack of generality, have been generated from a standard log-Normal distribution

$$V = \begin{bmatrix} 22.22 & 27.38 & 28.59 & 37.77 & 45.90 & 54.09 \\ 37.85 & 53.82 & 59.10 & 69.94 & 73.88 & 85.56 \\ 42.97 & 39.73 & 39.93 & 47.30 & 46.47 & 53.49 \\ 15.15 & 22.48 & 20.59 & 26.88 & 34.77 & 58.54 \end{bmatrix} = \begin{bmatrix} \boldsymbol{v}_1, \cdots, \boldsymbol{v}_6 \cdot \\ {\scriptstyle(4,1)} \quad {\scriptstyle(4,1)} \end{bmatrix}$$

The prices, P, needed to compute the TPD index, have been worked out as ratios between values and quantities:

$$P = \begin{bmatrix} 3.26 & 3.75 & 3.81 & 3.93 & 4.00 & 4.41 \\ 2.52 & 2.66 & 2.74 & 2.88 & 2.95 & 3.14 \\ 1.74 & 1.76 & 1.87 & 2.00 & 2.05 & 2.10 \\ 2.43 & 2.96 & 3.06 & 3.32 & 3.18 & 3.59 \end{bmatrix} = \begin{bmatrix} \underset{(4,1)}{p_1}, \cdots, \underset{(4,1)}{p_6} \cdot \end{bmatrix}$$

In this simulation, the matrices Q and V have been used to compute both the MA and the TPD indexes, in a multi-period perspective. In this regard, we have considered three different cases that cover three empirical scenario:

- **Case.1** Complete price tableau, implying a reference basket including a complete dataset for the four commodities;
- **Case.2** Incomplete price tableau, assuming missing the second and fourth commodity in the first and second period, respectively, (that is $q_{41} = v_{41} = 0$ and $q_{22} = v_{22} = 0$), with a "standard" reference basket that includes only the first and the third commodities;
- **Case.3** Incomplete price tableau assuming missing the second and fourth commodity in the first and second period, respectively, (that is $q_{41} = v_{41} = 0$ and $q_{22} = v_{22} = 0$), with the MA reference basket that includes commodities present in at least two periods, namely all the four commodities.

The outcome of the Breusch-Pagan test has led to rule out the presence of heteroschedasticity in all these three scenarios. In what follows the sum of the squares of the differences between the estimated MA and TPD indexes, $\hat{\lambda}$, and the "real" index, λ have been worked out for the three said cases:

- **Case.1** Complete price tableau: 0.0032 (MA) and 0.0066 (TPD);
- **Case.2** Incomplete price tableau ("standard" basket): 0.004 (MA) and 0.010 (TPD);
- **Case.3** Incomplete price tableau ("novel" basket): 0.001 (MA) and 0.003 (TPD).

It is worth noting that the MA index provides always the best fit to the index λ compared to the TPD one. As expected the index coincides, given the absence of heteroschedasticity across commodities, the MA index turns out to tally with the GK one. In all cases, the MA estimates turn out to be more efficient (see Fig. 1), as they have lower variances and, consequently, they are always included in a 2σ confidence band of the TPD index, as shown in Fig. 1. Looking at Fig. 2, we see that the values of the MA index provide the best fit to the real price index λ, avoiding the TPD overestimation issue present in all cases and, in particular, when there are missing prices.

These examples also highlight the role played by the reference prices, $\tilde{p}$, which are the prices that consumers are expected to pay for the commodities in a given period/country. Reference prices prove useful in obtaining estimates of the prices of those commodities which, being missing in the basket, can not be determined. Indeed, the price of a commodity, say i, missing in a period, say t, is undetectable, but can be determined as $\hat{\pi}_i' \hat{\lambda}_t$, where $\hat{\pi}_i$ and $\hat{\lambda}_t$ are the estimate of the reference price of the commodity i and the MA index at time t, respectively. This strategy has been used to estimate the
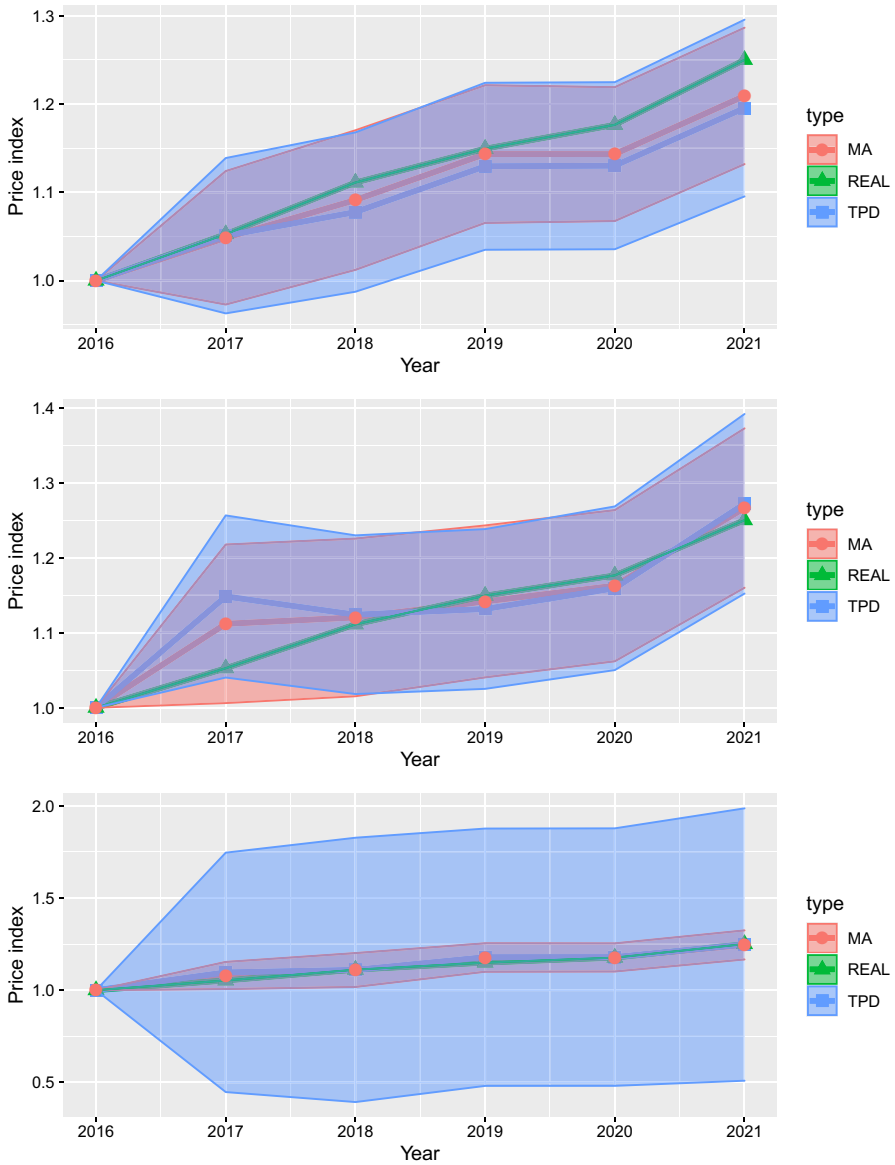
Fig. 1 Comparison of the MA and the TPD indexes (with 2σ confidence bands, given (22) for the MA index) for Case.1, Case2. and Case.3, respectively

prices of the second and fourth commodity in Case.3. According to (2), the prices $p_{4,1}$ and $p_{2,2}$ can be estimated as follows

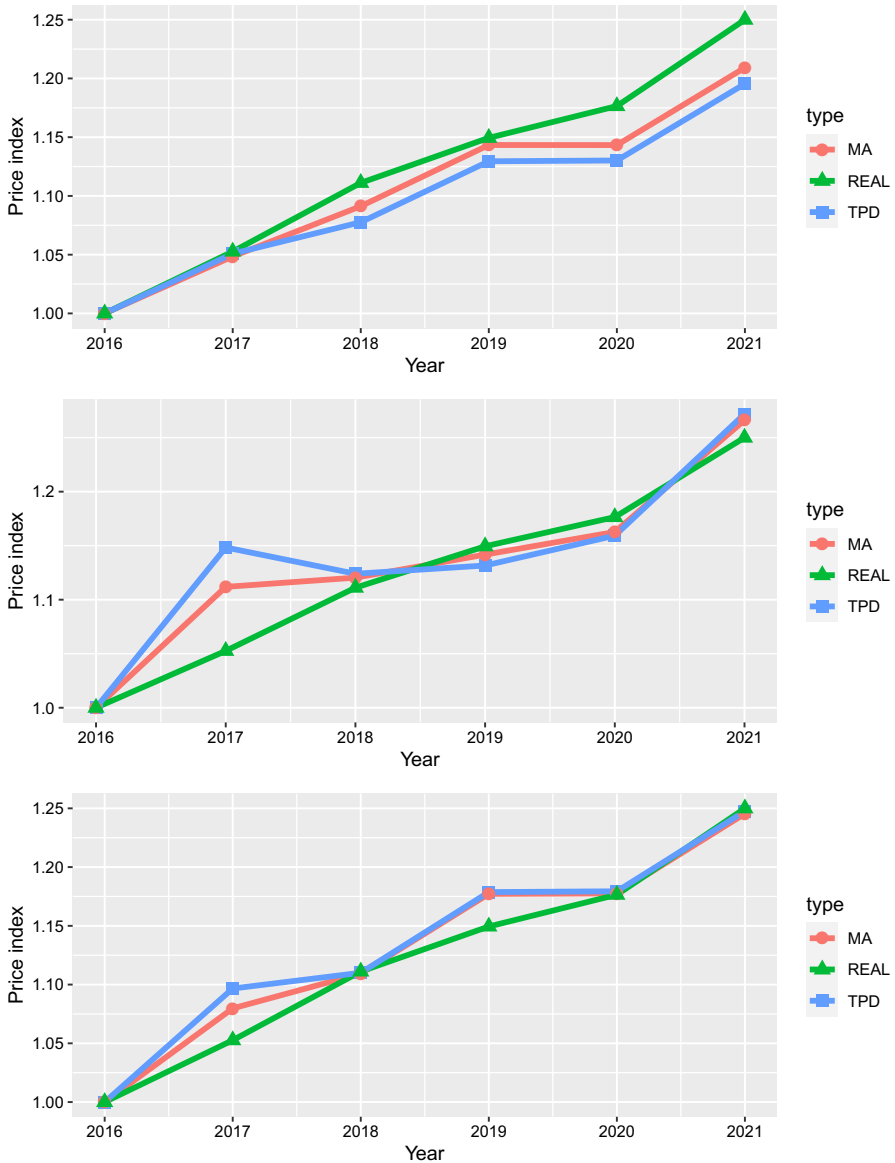$$\widehat{p}_{i,t} = \hat{\pi}_i \hat{\lambda}_t. \tag{48}$$

Fig. 2 Comparison of the MA and the TPD indexes (without confidence bands) for Case.1, Case.2 and Case.3, respectively

Note that $\hat{\pi}_i$ represents the price of the ith commodity in the base period/country, here assumed to be $t = 1$. According to (48), the price estimates $\hat{p}_{i,t}$ at times $t = 2, 3, \dots, T$ are obtained by updating $\hat{\pi}_i$ by means of the values of the index $\hat{\lambda}_t$ at these periods (or for the countries $t = 2, 3, \dots, T$). To assess the goodness of the $\hat{p}_{i,t}$ estimates, the sum of the squares between observed, p_i, and estimated prices, $\hat{p}_i$,

have been computed for all the commodities in the three cases under study. These are

$$\sum_{t=1}^{6} \sum_{i=1}^{4} (\hat{p}_{i,t} - p_{i,t})^2 = 0.258 \text{ for Case.1}$$

$$\sum_{t=1}^{6} \sum_{i=\{1,3\}} (\hat{p}_{i,t} - p_{i,t})^2 = 0.028 \text{ for Case.2} \qquad (49)$$

$$\sum_{t=1}^{6} \sum_{i=1}^{4} (\hat{p}_{i,t} - p_{i,t})^2 = 0.102 \text{ for Case.3}$$

and confirm the satisfactory performance of the MA index in reproducing missing prices by using reference ones. Clearly, this result hinges on the ability of this approach to provide good estimates of the reference prices. In this regard, Fig. 3 that compares the estimates of the MA reference prices with the "real" ones provides evidence of the goodness of the estimated reference prices for all commodities, also in the presence of missing prices. Indeed, all the points are close to the bisector of each panel. Unfortunately, given that the TPD approach does not provide reference prices as spin off, a comparison between the MA and the TPD indexes under this latter aspect, that is the "quality" of missing prices constructed by using reference prices, is not possible.

6 Concluding remarks

The paper provides the solution to the multi-period and/or multilateral price index in closed form, under proper error assumptions, taking a regression model specification as the frame of reference and the method of averages as the estimation approach. Two specifications are assumed in turn: sphericalness and commodity-dependent variances of the error terms. The former leads to an expression in compact form of the price indexes for an arbitrary number of periods and/or countries. The regression inferential apparatus applies accordingly. The price-index expression thus obtained proves to tally with the already known Geary–Khamis (GK) index, which is eventually endowed of the inferential heritage of the former. The second error specification drops the homoskedastic assumption in favour of a commodity-dependent hypothesis for the error variances, which leads to a new and more general price-index formula with a significance that extends beyond the GK case and opens up pathways for further research.

A Appendix

A.1 Proof of Lemma 1

Let X be defined as in Eq. (8) and X^b the binary matrix associated with X and set $B = V_2^b = Q_2^b$. Then, simple computations yield
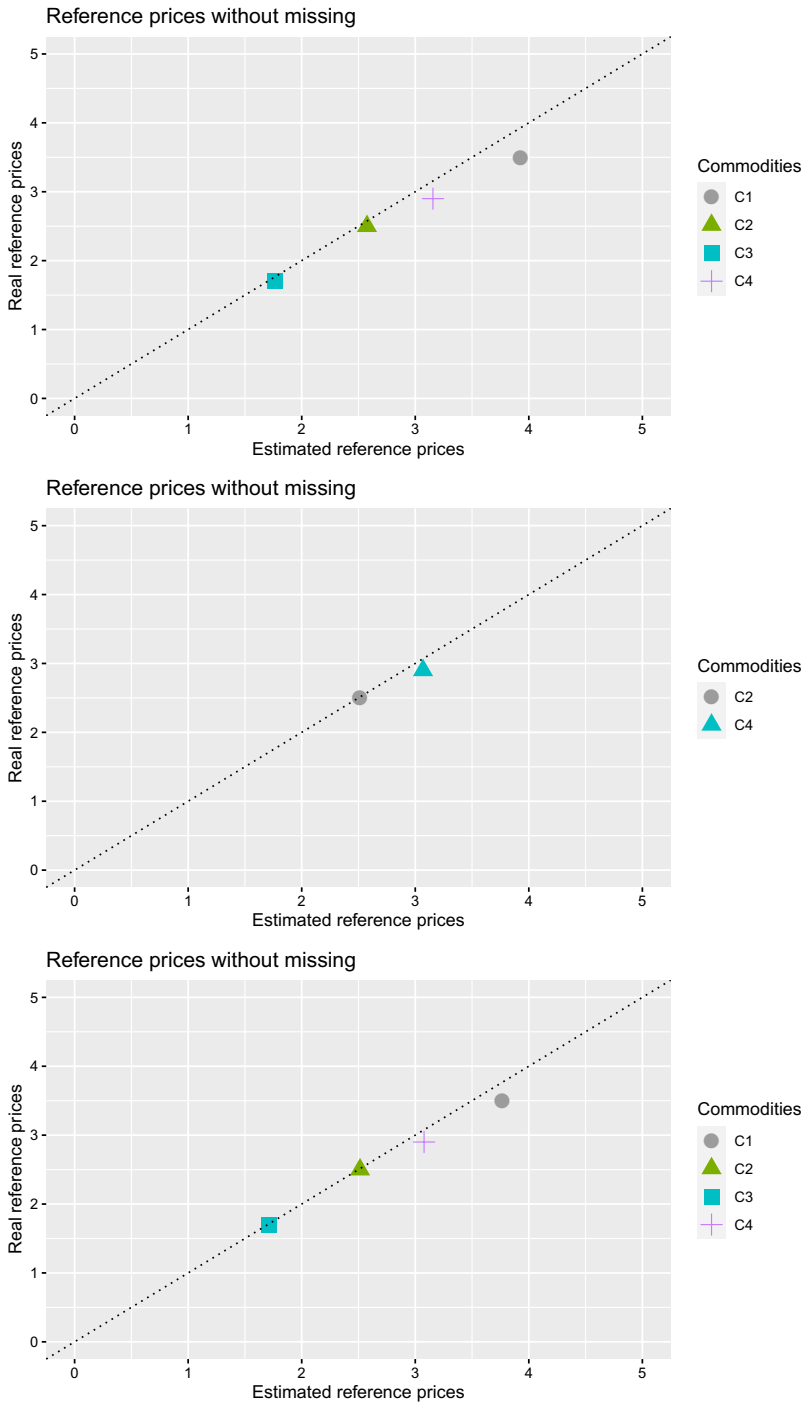
Fig. 3 MA estimates of the reference prices compared with the "real" ones for Case.1, Case.2 and Case.3, respectively. The dotted line represents the bisector

$$
\underset{(N+T-1,N+T-1)}{(X^b)'X} = \begin{bmatrix} \mathbf{0} & -R_{T-1}(I \otimes B') \\ D_{q_1}^b & R_N(B \otimes I) \end{bmatrix} \begin{bmatrix} \mathbf{0} & D_{q_1} \\ -(I \otimes V_2)R'_{T-1} & (Q'_2 \otimes I)R'_N \end{bmatrix}
$$

$$
= \begin{bmatrix} R_{T-1}(I_{T-1} \otimes B'V_2)R'_{T-1} & -R_{T-1}(Q'_2 \otimes B')R'_N \\ -R_N(B \otimes V_2)R'_{T-1} & D_{q_1} + R_N(BQ'_2 \otimes I_N)R'_N \end{bmatrix} \quad (50)
$$

$$
= \begin{bmatrix} I_{T-1} * (B'V_2) & -Q'_2 * B' \\ -B * V_2 & Dq_1 * I_N + (BQ'_2) * I_N \end{bmatrix} = \begin{bmatrix} D_w & -Q'_2 \\ -V_2 & D_{\bar{q}} \end{bmatrix}
$$

with $w > \mathbf{0}_{T-1}$ and $\bar{q} > \mathbf{0}_N$ defined in Eq. (21). Here use has been made of the following formula (see Faliva 1996)

$$
R_N(A \otimes B)R'_M = A * B \quad (51)
$$

which holds for any couples of matrices A and B of order $N \times M$. As (see Guttman 1946)

$$
r((X^b)'X) = r(D_{\bar{q}}) + r(D_w - Q'_2 D_{\bar{q}}^{-1} V_2) = N + r(D_w - Q'_2 D_{\bar{q}}^{-1} V_2) \quad (52)
$$

and

$$
D_w - Q'_2 D_{\bar{q}}^{-1} V_2 = (I_{T-1} - Q'_2 D_{\bar{q}}^{-1} V_2 D_w^{-1})D_w, \quad (53)
$$

it follows that the matrix $(X^b)'X$ is of full rank if and only if $C = I_{T-1} - Q'_2 D_{\bar{q}}^{-1} V_2 D_w^{-1}$ is non-singular, which occurs if the maximum eigenvalue of the non-negative matrix $\Gamma = Q'_2 D_{\bar{q}}^{-1} V_2 D_w^{-1}$ is lower than one. According to Perron-Frobenius theorem (Lancaster 1968), this is the case if the following

$$
u'_{T-1}\Gamma \le u'_{T-1} \quad (54)
$$

holds with some equality sign. Setting

$$
\underset{(T-1,N)}{Q'_2} = [\underset{(T-1,1)}{k_1}, \underset{(T-1,1)}{k_2}, \dots \underset{(T-1,1)}{k_N}], \quad \underset{(N,T-1)}{V_2} = [\underset{(N,1)}{v_2}, \underset{(N,1)}{v_3}, \dots \underset{(N,1)}{v_T}]. \quad (55)
$$

simple computations show that

$$
u'_{T-1}Q'_2 D_{\bar{q}}^{-1} = [k_1, k_2, \dots, k_N] \begin{bmatrix} \sum_{j=1}^T q_{1j} & 0 & \dots & 0 \\ 0 & \sum_{j=1}^T q_{2j} & 0 & 0 \\ 0 & & \dots & 0 \\ 0 & 0 & \dots & \sum_{j=1}^T q_{Nj} \end{bmatrix}^{-1} = \gamma'_N \quad (56)
$$

$$
= \left[\frac{\sum_{j=2}^T q_{1j}}{\sum_{j=1}^T q_{1j}}, \frac{\sum_{j=2}^T q_{2j}}{\sum_{j=1}^T q_{2j}}, \dots, \frac{\sum_{j=2}^T q_{Nj}}{\sum_{j=1}^T q_{Nj}} \right] \le u'_N
$$

$$u_N' V_2 D_w^{-1} = u_N' [v_2, v_3, \ldots, v_T] \begin{bmatrix} \sum_{i=1}^{N} v_{i2} & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \sum_{i=1}^{N} v_{i3} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ldots & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ldots & \sum_{i=1}^{N} v_{iT} \end{bmatrix}^{-1} = \tag{57}$$

$$= \left[\frac{\sum_{i=1}^{N} v_{i2}}{\sum_{i=1}^{N} v_{i2}}, \frac{\sum_{i=1}^{N} v_{i3}}{\sum_{i=1}^{N} v_{i3}}, \ldots, \frac{\sum_{i=1}^{N} v_{iN}}{\sum_{i=1}^{N} v_{iN}} \right] = u_{T-1}'.$$

and Eq. (54) follows accordingly. Thus, $I - \Gamma$ is non-singular and

$$r(X^{b'} X) = N + T - 1 \tag{58}$$

as requested. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

A.2 Proof of Theorem 1

From Eqs. (8) and (14), it follows that

$$\hat{\varphi} = \begin{bmatrix} I_{T-1}, & \mathbf{0} \end{bmatrix} \underset{(N+T,1)}{\hat{\beta}} = \begin{bmatrix} I_{T-1}, & \mathbf{0} \\ & {\scriptstyle (T-1,N)} \end{bmatrix} \underset{(N+T,N+T)}{((X^b)'X)^{-1}} \underset{(N+T-1,1)}{(X^b)'y} \tag{59}$$

with $(X^{b'} X)$ defined as in Eq. (50). A well known partitioned-inversion formula (Faliva and Zoia 2009) gives

$$\underset{(N+T,N+T)}{((X^b)'X)^{-1}} = \begin{bmatrix} D_w & -Q_2' \\ -V_2 & D_{\bar{q}} \end{bmatrix}^{-1} = \begin{bmatrix} S^{-1} & S^{-1} Q_2' D_{\bar{q}}^{-1} \\ D_{\bar{q}}^{-1} V_2 S^{-1} & D_{\bar{q}}^{-1} + D_{\bar{q}}^{-1} V_2 S^{-1} Q_2' D_{\bar{q}}^{-1} \end{bmatrix} \tag{60}$$

where $S = D_w - Q_2' D_{\bar{q}}^{-1} V_2$. Besides, the following holds

$$\underset{(N+T-1,1)}{(X^b)'y} = \begin{bmatrix} \mathbf{0} \\ {\scriptstyle (T-1,1)} \\ R_N (u_N \otimes I_N) v_1 \\ {\scriptstyle (N,1)} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ {\scriptstyle (T-1,1)} \\ R_N (u_N \otimes v_1) R_1' \\ {\scriptstyle (N,1)} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ {\scriptstyle (T-1,1)} \\ u_N * v_1 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ {\scriptstyle (T-1,1)} \\ v_1 \\ {\scriptstyle (N,1)} \end{bmatrix}. \tag{61}$$

From Eqs. (59), (60) and (61), the closed-form of the MA deflator is easily obtained, that is,

$$\hat{\varphi} = \begin{bmatrix} I_{T-1}, & \mathbf{0} \\ & {\scriptstyle (T-1,N)} \end{bmatrix} \begin{bmatrix} S^{-1} & S^{-1} Q_2' D_{\bar{q}}^{-1} \\ D_{\bar{q}}^{-1} V_2 S^{-1} & D_{\bar{q}}^{-1} + D_{\bar{q}}^{-1} V_2 S^{-1} Q_2' D_{\bar{q}}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ {\scriptstyle (T-1,1)} \\ v_1 \\ {\scriptstyle (N,1)} \end{bmatrix} \tag{62}$$

$$= S^{-1} Q_2' D_{\bar{q}}^{-1} v_1.$$

The more explicit representation in Eq. (20) follows from the latter by noting that

$$S = \begin{bmatrix} \sum_{i=1}^{N} \left\{ v_{i2} - \frac{q_{i2}v_{i2}}{\sum_{t=1}^{T} q_{it}} \right\} & -\sum_{i=1}^{N} \frac{q_{i2}v_{i3}}{\sum_{t=1}^{T} q_{it}} & \cdots & -\sum_{i=1}^{N} \frac{q_{i2}v_{iT}}{\sum_{t=1}^{T} q_{it}} \\ -\sum_{i=1}^{N} \frac{q_{i3}v_{i2}}{\sum_{t=1}^{T} q_{it}} & \sum_{i=1}^{N} \left\{ v_{i3} - \frac{q_{i3}v_{i3}}{\sum_{t=1}^{T} q_{it}} \right\} & \cdots & -\sum_{i=1}^{N} \frac{q_{i3}v_{iT}}{\sum_{t=1}^{T} q_{it}} \\ \cdots & \cdots & \cdots & \cdots \\ -\sum_{i=1}^{N} \frac{q_{iT}v_{i2}}{\sum_{t=1}^{T} q_{it}} & \cdots & \cdots & \sum_{i=1}^{N} \left\{ v_{iT} - \frac{q_{iT}v_{iT}}{\sum_{t=1}^{T} q_{it}} \right\} \end{bmatrix} \tag{63}$$

and

$$Q_2' D_{\bar{q}}^{-1} v_1 = \begin{bmatrix} \sum_{i=1}^{N} \frac{q_{i2}v_{i2}}{\sum_{t=1}^{T} q_{it}} \\ \cdots \\ \sum_{i=1}^{N} \frac{q_{iT}v_{iT}}{\sum_{t=1}^{T} q_{it}} \end{bmatrix} \tag{64}$$

which proves (20). The estimator in Eq. (22) of the dispersion matrix of the MA estimator $\boldsymbol{\varphi}$ follows from the equality

$$\widehat{\Sigma}(\widehat{\boldsymbol{\varphi}}) = S^{-1} Q_2' D_{\bar{q}}^{-1} \widehat{\Sigma}(v_1) D_{\bar{q}}^{-1} Q_2 S^{-1} \tag{65}$$

as $\widehat{\Sigma}(v_1) = \widehat{\Sigma}(\epsilon_1) = \widehat{\sigma}^2 I_N$, according to Eq. (15). The estimator in Eq. (23) of the variance σ^2 follows from Eq. (17) by working out the sum of the squared residuals

$$\begin{aligned} (y - X\widehat{\beta})'(y - X\widehat{\beta}) &= v_1' v_1 - 2[0', v_1'](X'X^b)^{-1} \begin{bmatrix} 0 \\ D_{q_1} v_1 \end{bmatrix} \\ &+ [0', v_1'](X'X^b)^{-1}(X'X)(X^{b'}X)^{-1} \begin{bmatrix} 0 \\ v_1 \end{bmatrix}, \end{aligned} \tag{66}$$

bearing in mind Eqs. (8), (14), (59), and noting that

$$X'y = \begin{bmatrix} 0 \\ D_{q_1} v_1 \end{bmatrix}. \tag{67}$$

$\square$

A.3 Proof of Corollary 1

The estimator of the price index at time t is the reciprocal of the estimator of the $(t-1)^{th}$ component of $\widehat{\boldsymbol{\varphi}}$, namely $\widehat{\varphi}_{t-1}$, given in (20), that is

$$\widehat{\lambda}_t = (e_{t-1}' \widehat{\boldsymbol{\varphi}})^{-1} = \widehat{\varphi}_{t-1}. \tag{68}$$

The variance of the estimator of the price index can be approximated following Benaroya et al. (2005) $\square$

Data availability Simulated data are directly contained in the paper while empirical data are publicly available on the Eurostat website.

Code availability All code written in R and available on request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

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

Benaroya, H., Han, S.M., Nagurka, M.: Probability Models in Engineering and Science. Taylor and Francis, Boca Raton (FL) (2005)

Brachinger, H.W., Beer, M., Schöni, O.: A formal framework for hedonic elementary price indices. AStA Adv. Stat. Anal. **102**(1), 67–93 (2018)

Diewert, W.E., Fox, K.J.: Decomposing productivity indexes into explanatory factors. Eur. J. Oper. Res. **256**(1), 275–291 (2017)

Diewert, W.E., Fox, K.J.: Substitution bias in multilateral methods for cpi construction. J. Business Econom. Stat. **40**(1), 355–369 (2022)

Edgeworth, F.Y.: Report of the committee etc. appointed for the purpose of investigating the best methods of ascertaining and measuring variations in the value of the monetary standard. memorandum by the secretary. Rep. British Assoc. Adv. Sci. **1**, 247–301 (1887)

Edgeworth FY (1925) Memorandum by the secretary on the accuracy of the proposed calculation of index numbers. Papers Relating to Political Economy 1

Faliva, M.: Hadamard matrix product, graph and system theories: motivations and role in econometrics. In: Camiz, S., Stefani, S. (eds.) Matrices and Graphs. World Scientific, Singapore (1996)

Faliva, M., Zoia, M.G.: Dynamic Model Analysis. Springer, Berlin (2009)

Goldberger,: Econometric Theory. Wiley, New York (1964)

Guttman, L.: Enlargement methods for computing the inverse matrix. Ann. Math. Stat. **17**(3), 336–343 (1946)

de Haan, J., Hendriks, R., Scholz, M.: Price measurement using scanner data: Time-product dummy versus time dummy hedonic indexes. Rev. Income Wealth **67**(2), 394–417 (2020). https://doi.org/10.1111/roiw.12468

Heston, A., Lipsey, R.E.: International and Interarea Comparisons of Income, Output, and Prices, vol. 61. University of Chicago Press, Chicago (2007)

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

Jevons, W.S.: A Serious Fall in the Value of Gold Ascertained: And Its Social Effects Set Forth. E. Stanford, London (1863)

Jevons, W.S.: The depreciation of gold. J. Stat. Soc. Lond. **32**, 445–449 (1869)

Kvečtoň, K.: Method of averages as an alternative to l1-and l2-norm methods in special linear regression problems. Comput. Stat. Data Anal. **5**(4), 407–414 (1987)

Lancaster, K.: Mathematical Economics. Mcmillan, New York (1968)

Pakes, A.: A reconsideration of hedonic price indexes with an application to pc's. Am. Econom. Rev. **93**(5), 1578–1596 (2003)

Rao, D.P., Hajargasht, G.: Stochastic approach to computation of purchasing power parities in the international comparison program (icp). J. Econom. **191**(2), 414–425 (2016)

Theil, H.: Best linear index numbers of prices and quantities. Econometrica **28**(2), 464–480 (1960)

Weinand S (2021) Measuring spatial price differentials at the basic heading level: a comparison of stochastic index number methods. AStA Adv. Stat. Anal. pp 1–27

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.