Scanner Data with Limited Characteristics Information: A Comparison of Stratified Multilateral Indexes and the Time Dummy Hedonic Index

Jan de Haan[a]

11 October 2022

Abstract: Statistical institutes using scanner data to construct subcomponents of the CPI often have limited information on price-determining characteristics of the products at their disposal. While the most important characteristics may be available, typically in categorical form, there is usually some information on characteristics missing. This paper argues that (with sufficient price and quantity data across the sample period) a stratified approach with ratios of weighted geometric average prices at the strata level and upper-level Törnqvist aggregation is preferable to using the weighted Time Dummy Hedonic method. If transitivity is deemed useful, GEKS can be applied to the Törnqvist-type indexes. When product relaunches do not occur frequently, a further improvement might be the use of the Time Product Dummy method at the strata level, which adjusts for quality/compositional change within the strata.

Keywords: hedonic regression, multilateral index methods, scanner data, stratification.

JEL Classification: C43, E31.

---

[a] Australian Bureau of Statistics (ABS) and Delft University of Technology; email: jandehaan1@gmail.com. The views expressed in this paper are those of the author and do not necessarily reflect the views of the ABS.

1. Introduction

With scanner data, where both prices and quantities are known, the construction of superlative price indexes like Fisher and Törnqvist indexes is possible and preferred over non-superlative (typically non-symmetric) indexes. A potentially important problem is product churn; products disappear from the market and new products enter, the magnitude depending on the prevailing market circumstances and hence varying across product categories. Matched-model superlative indexes can be biased if churn is significant and quality change is important or when relaunches of products (identified for example by GTIN, Global Trade Item Number, or model number) occur frequently and are accompanied by "disguised" price changes.[1] High-frequency chaining is a straightforward way to deal with churn in scanner data, but it is well known that this can introduce chain drift. Transitivity of the price index would then be a useful/required property. Multilateral index methods, which produce transitive indexes, are worth considering, and these methods have been implemented by various national statistical institutes for subcomponents of the Consumer Price Index (CPI) calculated from scanner data.[2]

Hedonic regression has become a standard tool for dealing with quality change in price indexes. The expenditure-share weighted Time Dummy Hedonic method seems well suited at first glance: it is a multilateral method that takes the economic importance of the products into account and yields a transitive quality-adjusted index which also deals with the relaunch issue. Yet, there are two potential problems. The first problem is practical: while the most important price-determining characteristics – which are usually categorical – to be used as explanatory variables in the hedonic model may be available to the statistical institute, other characteristics are often unavailable. In practice, the hedonic index is therefore likely to suffer from omitted variables bias, though not necessarily to a large extent. The second problem is of a conceptual nature: the implicit aggregation of subindexes appears to be non-symmetric.

This paper explains why, with sufficient observations across the entire sample period, stratification based on the available characteristics using ratios of weighted geometric average prices at the strata level and upper-level Törnqvist aggregation may be a better approach than using the weighted Time Dummy Hedonic method. If transitivity is deemed useful, the GEKS

[1] The relaunch issue was already mentioned by Reinsdorf (1999).

[2] Ivancic, Diewert and Fox (2011) were the first to propose using multilateral index number methods to construct price indexes from scanner data. Van Kints, de Haan and Webster (2019) provided an overview of the different methods and an application to Australian scanner data. Fox, Levell and O'Connell (2022) discussed different ways to extend the time series when new data become available. Diewert and Fox (2022) focused on substitution bias in multilateral price indexes using the economic approach to index number theory. De Haan and Krsinich (2014) and de Haan and Daalmans (2019) examined hedonic quality adjustments in multilateral indexes.

(Gini, 1931; Eltetö and Köves, 1964; Szulc, 1964) procedure can be applied to the Törnqvist-type indexes. If product relaunches do not occur frequently, a further improvement could be the use of the multilateral Time Product Dummy method at the strata level, which adjusts for quality/compositional change within the strata.

The structure of the paper is as follows. Section 2 discusses the quality-adjusted unit value index. This is helpful to introduce two non-hedonic multilateral methods, Geary-Khamis and Time Product Dummy, and the Time Dummy Hedonic method that will be discussed later and also to explain the aggregation problem. Section 3 outlines the Geary-Khamis index, which is a version of the quality-adjusted index where the quality parameters are estimated without using characteristics information. Section 4 looks at the weighted Time Product Dummy index, which might be viewed as a (regression-based) geometric variant of the quality-adjusted unit value index where, like in the Geary-Khamis index, no information on characteristics is used to estimate the quality-parameters. Section 5 argues that stratification, superlative upper-level aggregation and using the Time Product Dummy or Geary-Khamis method at the strata level is preferable, data permitting, to applying these methods to the overall data set. Transitivity can be attained using GEKS. Section 6 goes into the Time Dummy Hedonic method, again focusing on aggregation issues and transitivity. Section 7 then shows why, with limited information on characteristics, stratification is preferable to using the Time Dummy Hedonic method. Section 8 provides some practical advice and points to stratification methods recently proposed in the literature which do not seem useful. Section 9 summarizes the main findings.

2. The quality-adjusted unit value index

I start with some notation. The price of product *i* in period *t* $(t = 0,...,T)$ is denoted by p_i^t, the corresponding quantity sold by q_i^t, and the sample of products by S^t. Thus, $\sum_{i \in S^t} p_i^t q_i^t$ is the aggregate value in period *t*. The quality-adjusted unit value index is defined as

$$P_{QAUV}^{0t} = \frac{\sum_{i \in S^t} p_i^t q_i^t \Big/ \sum_{i \in S^t} \lambda_i q_i^t}{\sum_{i \in S^0} p_i^0 q_i^0 \Big/ \sum_{i \in S^0} \lambda_i q_i^0} = \frac{\sum_{i \in S^t} p_i^t q_i^t}{\sum_{i \in S^0} p_i^0 q_i^0} \left[\frac{\sum_{i \in S^t} \lambda_i q_i^t}{\sum_{i \in S^0} \lambda_i q_i^0} \right]^{-1} . \qquad (1)$$

The idea is that the quality parameters, λ_i, enable us to add up the quality-adjusted quantities $\lambda_i q_i^t$ across the various products. Because the quality parameters are assumed fixed across time, the quality-adjusted unit value index is transitive, i.e., independent of the choice of base period. Transitivity implies that the index can be expressed as a period-on-period chained index.

Suppose we subdivide the set of products into K non-overlapping subsamples or strata S_k^t $(k=1,...,K; t=0,...,T)$. What kind of upper-level aggregation is the quality-adjusted unit value index is implicitly assuming? For any stratification, equation (1) can be written as

$$P_{QAUV}^{0t} = \frac{\sum_{i \in S^t} p_i^t q_i^t}{\sum_{i \in S^0} p_i^0 q_i^0} \left[\frac{\sum_{k=1}^{K} \sum_{i \in S_k^t} \lambda_i q_i^t}{\sum_{l=1}^{L^0} \sum_{i \in S_k^0} \lambda_i q_i^0} \right]^{-1} = \left[\frac{\sum_{k=1}^{K} \frac{\sum_{i \in S_k^t} p_i^t q_i^t}{\sum_{i \in S^t} p_i^t q_i^t} \frac{\sum_{i \in S_k^t} \lambda_i q_i^t}{\sum_{i \in S_k^t} p_i^t q_i^t}}{\frac{\sum_{i \in S^0} \lambda_i q_i^0}{\sum_{i \in S^0} p_i^0 q_i^0}} \right]^{-1}$$

$$= \left[\frac{\sum_{k=1}^{K} \frac{\sum_{i \in S_k^t} p_i^t q_i^t}{\sum_{i \in S^t} p_i^t q_i^t} \frac{\sum_{i \in S_k^0} \lambda_i q_i^0}{\sum_{i \in S_k^0} p_i^0 q_i^0} \left[\frac{\sum_{i \in S_k^t} \lambda_i q_i^t \Big/ \sum_{i \in S_k^t} p_i^t q_i^t}{\sum_{i \in S_k^0} \lambda_i q_i^0 \Big/ \sum_{i \in S_k^0} p_i^0 q_i^0} \right]}{\frac{\sum_{i \in S^0} \lambda_i q_i^0}{\sum_{i \in S^0} p_i^0 q_i^0}} \right]^{-1} = \left[\sum_{k=1}^{K} s_k^t \frac{\tilde{p}_k^0}{\tilde{p}^0} \left[P_{QAUV,k}^{0t} \right]^{-1} \right]^{-1}, \qquad (2)$$

where $\tilde{p}_k^0 = \sum_{i \in S_k^0} p_i^0 q_i^0 / \sum_{i \in S_k^0} \lambda_i q_i^0$ and $\tilde{p}^0 = \sum_{i \in S^0} p_i^0 q_i^0 / \sum_{i \in S^0} \lambda_i q_i^0$ are the period 0 quality-adjusted unit values for stratum k and the total sample, respectively, $s_k^t = \sum_{i \in S_k^t} p_i^t q_i^t / \sum_{i \in S^t} p_i^t q_i^t$ is the period t expenditure share for stratum k, and

$$P_{QAUV,k}^{0t} = \frac{\sum_{i \in S_k^t} p_i^t q_i^t \Big/ \sum_{i \in S_k^t} \lambda_i q_i^t}{\sum_{i \in S_k^0} p_i^0 q_i^0 \Big/ \sum_{i \in S_k^0} \lambda_i q_i^0} = \frac{\sum_{i \in S_k^t} p_i^t q_i^t}{\sum_{i \in S_k^t} p_i^0 q_i^0} \left[\frac{\sum_{i \in S_k^t} \lambda_i q_i^t}{\sum_{i \in S_k^0} \lambda_i q_i^0} \right]^{-1} \qquad (3)$$

is the quality-adjusted unit value index for stratum k.

The last expression of equation (2) looks like an upper-level Paasche price index, with quality-adjusted unit value indexes at the lower level, but where the period t expenditure shares are multiplied by $\tilde{p}_k^0 / \tilde{p}^0$. In general, the "weights" $s_k^t (\tilde{p}_k^0 / \tilde{p}^0)$ do not sum to 1, which is not a useful property. For example, when all the lower-level indexes are equal to 1, the upper-level index will generally differ from 1. This reflects the fact that a quality-adjusted unit value index, like the ordinary unit value index, does not satisfy the (multi-period) identity test. The quality-adjusted unit value indexes for the strata, $P_{QAUV,k}^{0t}$, do not satisfy this test either, of course, but if we accept them as indicators of price change at the lower level, it makes more sense to use a superlative index number formula proper to aggregate them up.[3]

[3] Lamboray (2022) also suggested this.

3. Homogeneity, relative price change, and the Geary-Khamis method

We can address the aggregation issue from different angles. Equation (2) simplifies to an upper-level Paasche price index if $\tilde{p}_k^0$ is the same for all *k*. Because (2) holds for any stratification, including stratification down to the individual product level, the quality parameters must be the same for all products. In turn, this means that the products are essentially equivalent from the consumers' perspective, or in other words that the product category is fully homogeneous. The quality-adjusted unit value index then simplifies to the ordinary unit value index, which is what we expect to find.

There is another way of looking at the issue. Under optimal conditions, relative prices should reflect consumers' valuations of quality differences. Suppose we chose $\lambda_i = p_i^0 / p_r^0$ for all products, where *r* is an arbitrary reference product. All quality-adjusted base period prices $\tilde{p}_i^0 = p_i^0 / \lambda_i$ are then equal to p_r^0, and we have $\tilde{p}^0 = p_r^0$ and $\tilde{p}_k^0 = p_r^0$ in (2). Actually, we could choose any period and use the relative prices as indicators of the quality differences. Taking an average of p_i^t / p_r^t across all periods *t* seems a natural choice to measure the quality parameters. Also, since the choice of *r* is arbitrary, taking an average of p_r^t across all products seems rather natural. Essentially, this approach is equivalent to taking an average of deflated prices using a common deflator to estimate the quality parameters. However, averaging deflated prices across multiple periods implies that the period 0 quality-adjusted unit values in (2) will differ between the strata. Thus, this approach does not remove the unusual aggregation, unless there were no relative price changes.

Now let us explore the multilateral Geary-Khamis (GK) method. As mentioned before, the GK index is a version of the quality-adjusted unit value index where the quality parameters are estimated from the available price and quantity data, as follows:

$$\lambda_{i(GK)} = \sum_{t=0}^{T_i} w_i^t \frac{p_i^t}{P_{GK}^{0t}}; \qquad (4)$$

where $T_i (\leq T+1)$ denotes the number of periods that product *i* is purchased across the sample period and with $w_i^t = q_i^t / \sum_{t=0}^{T_i} q_i^t$. Equation (4) shows that each price observation is deflated by the GK price index to be determined, and so (1) and (4) together form a system of equations. In (4), a weighted average of all the deflated prices is taken, where the weights w_i^t are equal to the shares of the quantities purchased in each period with respect to the total quantity purchased across the entire sample period.

In particular if the price changes of the products differ in the longer run, i.e., if the price trends are different, the use of a single deflator (the GK index) is not appropriate; the parameter estimates will then be potentially biased. Having similar trends is highly unlikely for products belonging to a heterogeneous product category. Nevertheless, this is exactly what underlies the GK method: it is assumed that the price trends of all the products are the same and that short-term differences from the common trend are purely random and can be ignored.

GK has more problems. Promotional sales affect the precision of the quality parameter estimates. They are characterized by unusually low prices, but they typically also exhibit large increases in quantities sold so that sales periods have a bigger impact on the quality parameters than non-sales periods. As sales have little to with consumers' valuations of quality differences, this is an undesirable property. In the literature on spatial price indexes, the problem is known as the Gerschenkron effect. Furthermore, like any non-hedonic multilateral method, GK zeroes out products sold in a single period across the entire sample period. This means that products added in period T (as well as those exiting in period 1) do not affect the index.

To summarize: the GK approach is consistent with a homogeneous product (category) and/or when there are no relative price changes.[4] What we could try to do in practice is stratify the sample such that the strata are "relatively homogeneous". If the price trends of the products within each stratum are broadly similar, applying GK at the strata level and combining it with superlative upper-level aggregation might be useful.

4. The Time Product Dummy Method

The Time Product Dummy (TPD) method is a regression-based multilateral approach, adapted from Summers' (1973) CPD method. Instead of characteristics, product fixed effects are used as explanatory variables, measured by product dummy variables. Suppose there are N different products sold across the entire sample period. The TPD estimating equation then is

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t \quad (t = 0,...,T), \tag{5}$$

where D_i is a dummy variable for product i and γ_i is the corresponding parameter, i.e., the product fixed effect (on the log of price); D_i^t is a time dummy variable that has the value 1 if the observation pertains to period t and 0 otherwise, and δ^t is the time dummy parameter; α is the intercept term and ε_i^t is an error term with zero mean. Product N and period 0 are left out to identify the model.

[4] In spite of the problems with the GK method, Statistics Netherlands decided to implement this method for the treatment of scanner data in the CPI (Chessa, 2016). De Haan (2022) proposed using GK at a very detailed level of aggregation to estimate (stock-based) rent indexes on a very large Australian data set.

To take account of the products' economic importance, I follow Diewert (20XX) and assume that equation (5) is estimated by Weighted Least Squares (WLS) regression using the expenditure shares in the respective periods as weights. The WLS parameter estimates are $\hat{\alpha}$, $\hat{\delta}^t$ $(t=1,...,T)$ and $\hat{\gamma}_i$ $(i=1,...,N-1)$, with $\hat{\delta}^0=0$ and $\hat{\gamma}_N=0$. The TPD index is found by exponentiating $\hat{\delta}^t$: $P_{TPD}^{0t}=\exp(\hat{\delta}^t)$ with $P_{TPD}^{00}=1$.

The predicted prices from the (weighted) TPD regression are $\hat{p}_i^0=\exp(\hat{\alpha})\exp(\hat{\gamma}_i)$ and $\hat{p}_i^t=\exp(\hat{\alpha})\exp(\hat{\gamma}_i)\exp(\hat{\delta}^t)$. We can view $\hat{\lambda}_{i(TPD)}=\exp(\hat{\gamma}_i)=\hat{p}_i^0/\exp(\hat{\alpha})$ as an estimate of the product's quality parameter. The ratio of the quality parameter estimates for products *i* and *j*, $\hat{\lambda}_{i(TPD)}/\hat{\lambda}_{j(TPD)}=\hat{p}_i^0/\hat{p}_j^0$, is again a measure of consumers' valuations of the quality difference. The quality parameters can be written as (see e.g., Rao, 2005)

$$\hat{\lambda}_{i(TPD)}=\prod_{t=0}^{T_i}\left(\frac{p_i^t}{P_{TPD}^{0t}}\right)^{s_i^t} \quad (i=1,...,N_l), \tag{6}$$

where $T_i(\leq T+1)$ is the number of periods that product *i* is purchased across the sample period, as before. That is, the TPD index acts as the deflator to express the observed prices in period 0 values, and then the expenditure-share weighted geometric average of all the deflated prices is taken. Since expenditure shares are usually more stable than relative quantities, the TPD index suffers less from the Gerschenkron effect than the GK index. Similar to GK, TPD zeroes out products purchased during a single period. The TPD index can also be expressed as the ratio of expenditure-share weighted geometric averages of quality-adjusted prices:[5]

$$P_{TPD}^{0t}=\frac{\prod_{i\in S^t}\left(p_i^t/\hat{\lambda}_{i(TPD)}\right)^{s_i^t}}{\prod_{i\in S^0}\left(p_i^0/\hat{\lambda}_{i(TPD)}\right)^{s_i^0}}. \tag{7}$$

As before, suppose we stratify the total sample of products in each period into *K* strata. Expression (7) for the TPD index can then be written as

$$P_{TPD}^{0t}=\frac{\prod_{k=1}^{K}\prod_{i\in S_k^t}\left(p_i^t/\hat{\lambda}_{i(TPD)}\right)^{s_i^t}}{\prod_{i\in S^0}\left(p_i^0/\hat{\lambda}_{i(TPD)}\right)^{s_i^0}}=\frac{\prod_{k=1}^{K}\left[\prod_{i\in S_k^t}\left(p_i^t/\hat{\lambda}_{i(TPD)}\right)^{s_{i(k)}^t}\right]^{s_k^t}}{\prod_{i\in S^0}\left(p_i^0/\hat{\lambda}_{i(TPD)}\right)^{s_i^0}}$$

[5] Equations (6) and (7) form a system which, like the GK system, can be solved iteratively. This means we do not have to run TPD regressions to obtain the TPD indexes and quality parameters. However, running regressions has some advantages: it is easy to perform and produces useful diagnostics such as goodness of fit (R squared) and standard errors of the coefficients.

$$= \prod_{k=1}^{K} \left[\frac{\prod_{i \in S_k^0} \left(p_i^0 / \hat{\lambda}_{i(TPD)} \right)^{s_{i(k)}^0}}{\prod_{i \in S^0} \left(p_i^0 / \hat{\lambda}_{i(TPD)} \right)^{s_i^0}} \frac{\prod_{i \in S_k^t} \left(p_i^t / \hat{\lambda}_{i(TPD)} \right)^{s_{i(k)}^t}}{\prod_{i \in S_k^0} \left(p_i^0 / \hat{\lambda}_{i(TPD)} \right)^{s_{i(k)}^0}} \right]^{s_k^t} = \prod_{k=1}^{K} \left[\frac{\prod_{i \in S_k^0} \left(p_i^0 / \hat{\lambda}_{i(TPD)} \right)^{s_{i(k)}^0}}{\prod_{i \in S^0} \left(p_i^0 / \hat{\lambda}_{i(TPD)} \right)^{s_i^0}} \right]^{s_k^t} \prod_{k=1}^{K} \left[P_{TPD*,k}^{0t} \right]^{s_k^t}$$

$$= \prod_{k=1}^{K} \left[\exp(\overline{e}_k^0) \right]^{s_k^t} \prod_{k=1}^{K} \left[P_{TPD*,k}^{0t} \right]^{s_k^t} = \exp\left[\sum_{k=1}^{K} s_k^t \overline{e}_k^0 \right] \prod_{k=1}^{K} \left[P_{TPD*,k}^{0t} \right]^{s_k^t}, \qquad (8)$$

where $s_k^t = \sum_{i \in S_k^t} p_i^t q_i^t / \sum_{i \in S^t} p_i^t q_i^t$ is the period t expenditure share of stratum k and $P_{TPD*,k}^{0t}$ is an implicit TPD-type index for stratum k with quality parameters (fixed effects) based on the full sample; $\overline{e}_k^0 = \sum_{i \in S_k^0} s_{i(k)}^0 e_i^0$ is a weighted average of the period 0 TPD regression residuals $e_i^0 = \ln(p_i^0 / \hat{p}_i^0)$ in stratum k, with normalized shares $s_{i(k)}^0 = s_i^0 s_k^0$ as weights. I made use of the fact that the weighted average of the period 0 residuals across all products equals 0.

The last expression in (8) is similar to equation (2) for the quality-adjusted unit value index, hence for the GK index. The second term explains the implicit upper-level aggregation: strata indexes are aggregated using the geometric Paasche formula. However, the first term in (8), $\exp[\sum_{k=1}^{K} s_k^t \overline{e}_k^0]$, generally differs from 1. So, the TPD index has the same issue as the GK index: if all the strata index numbers are the same, the aggregate (TPD) index will in general have a different value.

The underlying problem with TPD and GK is that the prices of the different products are deflated by a common deflator, the TPD or GK index, for estimating the quality parameters. But if price trends actually differ, the use of a common deflator would distort the estimation of the parameters. Like GK, TPD assumes that the price trends for the different products are the same and that deviations from trend are purely random. This is not surprising, really, since we know that $\hat{p}_i^t / \hat{p}_i^0 = \exp(\hat{\delta}^t) = P_{TPD}^{0t}$ for all i, i.e., that the estimated price relatives are the same for all products.

5. Stratification

If we still want to apply TPD (or GK), it seems useful, data permitting, to stratify according to the most important price-determining characteristics and apply TPD (or GK) at the strata level. Because products within strata are broadly similar (the strata are relatively homogeneous), we would expect their price trends to be similar so that the quality parameter estimates at the strata level will be less biased. So, given the implicit aggregation, "improved versions" of equations (2) and (8) for GK and TPD, respectively, are stratified indexes using upper-level Paasche and geometric Paasche aggregation, respectively:

$$P_{P(GK)}^{0t} = \left[\sum_{k=1}^{K} s_k^t \left[P_{GK,k}^{0t} \right]^{-1} \right]^{-1}; \quad (9)$$

$$P_{GP(TPD)}^{0t} = \prod_{k=1}^{K} \left[P_{TPD,k}^{0t} \right]^{s_k^t}, \quad (10)$$

where $P_{GK,k}^{0t}$ and $P_{TPD,k}^{0t}$ are the GK and TPD indexes for stratum k (based on data from k only).

However, there is no reason for using asymmetric upper-level aggregation as in (9) or (10). Superlative formulae are symmetric. Thus, improved versions of (9) and (10) are stratified indexes based on GK or TPD indexes at the strata level combined with upper-level Fisher or Törnqvist aggregation:

$$P_{F(GK)}^{0t} = \left[\sum_{k=1}^{K} s_k^0 P_{GK,k}^{0t} \left[\sum_{k=1}^{K} s_k^t \left[P_{GK,k}^{0t} \right]^{-1} \right]^{-1} \right]^{1/2}; \quad (11)$$

$$P_{T(TPD)}^{0t} = \prod_{k=1}^{K} \left[P_{TPD,k}^{0t} \right]^{(s_k^0 + s_k^t)/2}. \quad (12)$$

A potential issue with equations (11) and (12), and also with (9) and (10), is that the aggregate indexes are not transitive. If transitivity is deemed important, we could apply GEKS to all possible bilateral upper-level Fisher or Törnqvist comparisons across the sample period. The GEKS counterparts to (11) and (12) are

$$P_{GEKS-F(GK)}^{0t} = \prod_{l=0}^{T} \left[P_{F(GK)}^{0l} P_{F(GK)}^{lt} \right]^{\frac{1}{T+1}}; \quad (13)$$

$$P_{GEKS-T(TPD)}^{0t} = \prod_{l=0}^{T} \left[P_{T(TPD)}^{0l} P_{T(TPD)}^{lt} \right]^{\frac{1}{T+1}}, \quad (14)$$

where l $(0 \le l \le T)$ is the link period; note that l can be greater than t. Taking the mean across all possible link periods ensures transitivity.

In many cases, non-transitivity is not a major issue at the upper aggregation level. Why might transitivity still be an important property in the present context? The importance depends on the detail of stratification. Suppose the characteristics used for stratification are categorical – which they typically are for products such as consumer electronics goods – and that we define the strata by full stratification, i.e., by crossing all the categories of the characteristics. Some strata will be empty across the whole sample period, but that is of course not a problem because these strata can be ignored. Other strata may be temporarily empty (and possibly have only few observations in periods when they are not empty). To make use of all the matches at the strata level across the sample period, applying GEKS might be useful.

The structure of (12) is simpler than that of (11). The three steps distinguished – TPD at the strata level, Törnqvist aggregation at the upper level, and GEKS – are all geometric, and this simplifies decomposition analysis. Importantly, the index given by (12) is better suited for a comparison with the Time Dummy Hedonic index when the same characteristics are used as explanatory variables in the hedonic model.

6. The Time Dummy Hedonic method

The Time Dummy Hedonic (TDH) method is a straightforward multilateral hedonic approach. The following log-linear regression model is run on the pooled data of the entire sample period:

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{m=1}^{M} \beta_m z_{im} + \varepsilon_i^t \quad (t = 0,...,T), \tag{15}$$

where z_{im} is characteristic m of product i and β_m the corresponding parameter $(m = 1,...,M)$. As before, D_i^t is the time dummy variable, δ^t is the corresponding parameter $(1,...,T)$, α is the intercept term, and ε_i^t is the error term. Again, I assume that the equation is estimated using expenditure-share weighted least squares regression. The weighted TDH index, P_{TDH}^{0t}, is equal to the exponential of the estimated time dummy parameter.

To compare the index with the index given by (14), I start by writing the TDH index as (see e.g., de Haan, 20XX)

$$P_{TDH}^{0t} = \frac{\prod_{i \in S^t} (p_i^t)^{s_i^t}}{\prod_{i \in S^0} (p_i^0)^{s_i^0}} \exp\left[\sum_{m=1}^{M} \hat{\beta}_m (\bar{z}_m^0 - \bar{z}_m^t)\right] = \frac{\prod_{i \in S^t} (p_i^t / \hat{\lambda}_{i(TDH)})^{s_i^t}}{\prod_{i \in S^0} (p_i^0 / \hat{\lambda}_{i(TDH)})^{s_i^0}}, \tag{16}$$

where $\bar{z}_m^t = \sum_{i \in S^t} s_i^t z_{im}$ is the expenditure-share weighted average of characteristic k in period t $(0,...,T)$. The second term in the first expression of (16) adjusts the ratio of expenditure-share weighted geometric average prices for changes in the weighted average characteristics. In that sense the TDH index is explicitly quality adjusted. The second expression pf (16) is similar to expression (7) for the TPD index and writes the TDH index as the ratio of expenditure-share weighted geometric average quality-adjusted prices, $p_i^t / \hat{\lambda}_{i(TDH)}$, where the quality parameter estimates $\hat{\lambda}_{i(TDH)} = \exp(\hat{\beta}_m z_{mi})$, are derived from the TDH regression.

Equation (16) can be written in a similar stratified form as (8), hence

$$P_{TDH}^{0t} = \exp\left[\sum_{k=1}^{K} s_k^t \bar{e}_{k(TDH)}^0\right] \prod_{k=1}^{K} \left[P_{TDH*,k}^{0t}\right]^{s_k^t}, \tag{17}$$

where $P_{TDH*,k}^{0t} = \left[\prod_{i \in S_k^t} (p_i^t / \hat{\lambda}_{i(TDH)})^{s_{i(k)}^t} \Big/ \prod_{i \in S_k^0} (p_i^0 / \hat{\lambda}_{i(TDH)})^{s_{i(k)}^0} \right]$ is the implicit TPD-type index for stratum k with quality parameters based on the overall sample; $\overline{e}_k^0 = \sum_{i \in S_k^0} s_{i(k)}^0 e_{i(TDH)}^0$ is the expenditure-share weighted geometric average of the TDH residuals $e_{i(TDH)}^0 = \ln(p_i^0 / \hat{p}_{i(TDH)}^0)$ in period 0 for stratum k. With full stratification (complete crossing of the available categorical characteristics), however, the quality parameters within each stratum are the same so that the strata indexes are equal to the ratios of expenditure-share weighted geometric average prices, $P_{TDH*,k}^{0t} = \left[\prod_{i \in S_k^t} (p_i^t)^{s_{i(k)}^t} \Big/ \prod_{i \in S_k^0} (p_i^0)^{s_{i(k)}^0} \right] = P_{GM,k}^{0t}$. Thus, equation (17) becomes

$$P_{TDH}^{0t} = \exp\left[\sum_{k=1}^{K} s_k^t \overline{e}_{k(TDH)}^0 \right] \prod_{k=1}^{K} \left[P_{GM,k}^{0t} \right]^{s_k^t} . \tag{18}$$

The second term in equation (18) is an upper-level geometric Paasche price index with geometric mean indexes at the strata level. The first term adjusts the Paasche index but, since it does not contain changes in characteristics, it does not adjust for quality change. This is not surprising because full stratification already adjusts for quality change in terms of the available characteristics information. What the adjustment term does is preserving transitivity.

Note, by the way, that (18) can also be written as

$$P_{TDH}^{0t} = \prod_{k=1}^{K} \left[\frac{\tilde{p}_k^t}{\tilde{\hat{p}}_k^0} \right]^{s_k^t} , \tag{19}$$

where $\tilde{\hat{p}}_k^0 = \prod_{i \in S_k^0} (\hat{p}_i^0)^{s_{i(k)}^0}$ is the expenditure-share weighted geometric mean of estimated base period prices for stratum k.

7. Do we need hedonics?

When the transitivity preserving term in equation (18) for the TDH index has a trend, i.e., when it differs systematically from 1, and if transitivity is not a requirement, ignoring this term seems to make sense. We then obtain an upper-level geometric Paasche price index with geometric mean indexes at the strata level:

$$P_{GP(GM)}^{0t} = \prod_{k=1}^{K} \left[P_{GM,k}^{0t} \right]^{s_k^t} . \tag{20}$$

Because upper level Törnqvist aggregation is preferred over geometric Paasche aggregation, a better option would be

$$P_{T(GM)}^{0t} = \prod_{k=1}^{K} \left[P_{GM,k}^{0t} \right]^{(s_k^0 + s_k^t)/2} . \tag{21}$$

And again, if transitivity is not a requirement but still useful, e.g., when some of the strata are temporarily empty, GEKS could be applied to all of the bilateral indexes, giving rise to

$$P_{GEKS-T(GM)}^{0t} = \prod_{l=0}^{T} \left[P_{T(GM)}^{0l} P_{T(GM)}^{lt} \right]^{\frac{1}{T+1}} . \tag{22}$$

The index given by (22) is transitive, quality adjusted, and makes use of the same information in terms of characteristics as well as prices and quantities as the (transitive and quality adjusted) TDH index. Due to improved upper-level aggregation, I believe (22) is a better choice than the TDH index. In other words, in case some price-determining characteristics are missing but the most important ones are available (in categorical form), we do not need hedonics and perhaps even try to avoid it.

The next question is: would (22) also be a better choice than the GEKS-Törnqvist index given by (14)? The only difference is that in (14) TPD rather than geometric mean indexes are used at the strata level. Two issues are heterogeneity of the strata and relaunching of products. When the strata are quite heterogeneous due to unavailability of some important characteristics, in which case the resulting indexes tend to have omitted variables bias, (14) seems preferable. This is because the TPD method implicitly adjusts for compositional/quality change within the strata by using ratios of quality-adjusted rather than unadjusted geometric mean prices, which will in general be an improvement.

However, when product relaunches occur frequently, the choice becomes less clear. A product relaunch is the situation where a product, as identified by a detailed product identifier such as GTIN, is no longer available and replaced by a product that is deemed the same from the consumers' perspective but has a different identifier. Relaunches can go together with what are sometimes called "disguised" price changes: when product identifiers change, e.g., due to a slight change in packaging, prices may change as well. The TPD method is unable to pick up such price changes (and neither is GK) since it relies on tracking prices of individual products over time. Thus, when relaunches are accompanied by price increases, which has been found for product categories such as personal care, some of the TPD strata indexes will be downward biased as will be the index given by (14).

The geometric mean strata indexes include "disguised" price changes and so does the index given by (22). The choice between (14) and (22) is mainly a practical matter and depends on the frequency of relaunches and the magnitude of the price changes. The situation is likely to differ across product categories; for consumer electronics goods, for example, I do not think relaunches are a big issue. As we do not know the "true" aggregate price change, not much can be said about the trade-off between heterogeneity and relaunches.

8. Practical advice

This section provides some practical advice in terms of what statistical institutes might do if they have access to scanner data with sufficient observations across the sample period but when not all (categorical) product characteristics are available to them. These conditions may already be met in scanner data sets which are currently available to statistical institutes. Some institutes, including Statistics Netherlands and Statistics Norway, enriched scanner data with information from alternate sources, in particular by extracting it from retailers' or manufacturers' websites using web-scraping.[6]

Applying GEKS at the upper level is not necessary if none of the strata are empty; the direct or chained Törnqvist index will suffice. However, some strata may become empty in the future and to make use of all the matches (without running the risk of introducing chain drift) it seems worthwhile using GEKS from the start. If there is evidence of product relaunches and significant "disguised" price changes, then the index given by equation (22) can be used. This would also be a useful choice for fashion goods such as clothing and footwear. There are four steps involved.

The first step is full stratification by crossing all available categorical characteristics; strata which are empty across the sample period are redundant. The second step is calculating (expenditure-share weighted geometric) average prices for each period for all the strata. Using these average prices and the corresponding strata weights, the third step is constructing bilateral Törnqvist price indexes to be used in the GEKS procedure. The fourth step is applying GEKS to the bilateral Törnqvist indexes.

If there is no evidence of product relaunches with a substantial impact on the results, it would be preferable to use the index given by equation (14) instead. The second step becomes a bit more complicated because we need to run TPD regressions for all of the strata. The third step is similar to the one above except that the bilateral Törnqvist price indexes are now based on the bilateral comparisons from the TPD regressions. The fourth step is again applying GEKS to the bilateral Törnqvist indexes.

In many cases we are probably uncertain about the importance of relaunches so that we cannot make an informed choice between (14) and (22). Taking the geometric mean could then be considered:

$$P^{0t}_{GEKS-T(TPD/GM)} = \left[P^{0t}_{GEKS-T(TPD} P^{0t}_{GEKS-T(GM)} \right]^{1/2} = \prod_{l=0}^{T} \left[(P^{0l}_{T(TPD)} P^{0l}_{T(GM)})^{1/2} (P^{lt}_{T(TPD)} P^{lt}_{T(GM)})^{1/2} \right]^{\frac{1}{T+1}} . \qquad (23)$$

[6] See for example de Haan and Daalmans (2019) and Nyborg Hov and Nygaard (2022).

It should be noted that (14), (22) and (23) require a large number of observations. These methods may be labelled "big data" methods. For relatively small data sets on the other hand, the TDH method could still be worthwhile. Before implementing a method in CPI production, I would recommend also estimating TDH (and TPD) indexes during the development stage to examine how they compare to the GEKS-Törnqvist indexes (14), (22) and (23). For comparison purposes, it might even be interesting to calculate GEKS-Fisher with GK strata indexes, given by (13), as well.

There are a couple of things that have recently been proposed which I do not find useful. If relaunches are deemed important, Chessa (2016) proposed applying the GK method to the strata indexes, where the latter are based on unit values. I have no problem with the use of unit values for relatively homogeneous strata (noting that when the variance of the prices is modest, expenditure-share weighted geometric mean prices will be similar to unit values). But applying GK to the strata indexes to attain transitivity does not make much sense as that implies Paasche-type upper-level aggregation while better aggregation and transitivity preserving methods are possible with scanner data.

In a follow-up paper, Chessa (2021) proposed a method called MARS (Match Adjusted R Squared) to derive the optimal stratification scheme by trading off increased matching over time of the resulting strata and increased heterogeneity. That is, MARS "defines products […] by balancing two measures: one measure quantifies the "homogeneity" […] within products, while the second measure expresses the degree to which products can be "matched" each month with respect to a fixed comparison period". The first problem I have is the language used. Strata with heterogeneous products should not be called products. Products cannot be defined; they are what they are, and stratification is simply classifying individual products into, hopefully relatively homogeneous, strata. Less detailed stratification simply defines away a low matching rate.

My main criticism is that there is no benchmark index involved. Chessa took the limited information on characteristics as the starting point, and his MARS method essentially compares all the possible upper level (GK) indexes using the variance of the time series as the criteria to select the optimal stratification. MARS is a method – and perhaps not the best one – that seeks to reduce variability of the upper-level index but ignores the fact that the resulting trend may differ from the unknown "true" trend. Diewert (2022) explains the issue as follows: the "MARS measure of product homogeneity focuses only on the variance of current period prices rather than focusing on how proportional the prices in the current period are relative to the base period prices."

9. Summary of main findings

The key takeaway from this paper is: with limited information on categorical characteristics, we do not need hedonics to adjust for quality changes and perhaps try to avoid it. Importantly, most of the methods discussed are "big data" methods in the sense that they require a sufficient number of observations for each stratum across the entire sample period. For small data sets, the TDH method can still be useful.

Below, I list a range of detailed findings.[7]

- The multilateral GK index, which is an instance of the quality-adjusted unit value index, essentially assumes that the price changes of all products are the same.
- The multilateral TPD index is based on a similar assumption.
- GK and TPD implicitly aggregate lower-level indexes using the Paasche or geometric Paasche formula, respectively.
- Stratifying according to a set of price-determining characteristics, applying GK or TPD at the strata level and using upper-level Fisher or Törnqvist aggregation is preferable to applying GK or TPD to the overall data set.
- When transitivity is deemed useful, the GEKS procedure can be applied to the bilateral Fisher or Törnqvist indexes.
- Like the TPD method, the TDH method (which is multilateral too) implicitly aggregates lower-level indexes using the geometric Paasche formula.
- Stratifying according to the most important characteristics, applying ratios of weighted geometric mean prices at the strata level, using upper-level Törnqvist aggregation, and finally applying GEKS to preserve transitivity is generally preferable to applying TDH to the overall data set.
- Data permitting, a further improvement could be to apply the TPD method at the strata level to adjust for compositional change within the strata. This approach should not be used when there is evidence of frequent product relaunches with significant "disguised" price changes (and neither for fashion goods).
- In practice, it is likely that we are often uncertain about the impact of relaunches. Taking the geometric mean of the two options, i.e., upper-level GEKS-Törnqvist indexes which rely either on ratios of geometric mean prices or TPD indexes at the strata level, could then be considered.

[7] Admittedly, not all the results are new and can also be found in Diewert (2021a; 2021b). Furthermore, some of the methods discussed in this paper have not yet been applied empirically, and so their practical relevance remains somewhat unclear at this point.

References

Chessa, A.G. (2016), "A New Methodology for Processing Scanner Data in the Dutch CPI", *Eurona* 1, 49-69.

Chessa, A.G. (2021), "A Product Match Adjusted R Squared Method for Defining Products with Transaction Data", *Journal of Official Statistics* 37, 411-432.

Diewert, W.E. (2004), "On the Stochastic Approach to Linking the Regions in the ICP", Discussion Paper 04-16, Department of Economics, The University of British Columbia, Vancouver, Canada.

Diewert, W.E. (2021a), "Scanner Data, Elementary Price Indexes and the Chain Drift Problem", Chapter 7 in Consumer Price Index Theory (CPI Manual), available from Update of the Consumer Price Index Manual (imf.org).

Diewert, W.E. (2021b), "Quality Adjustment Methods", Chapter 8 in Consumer Price Index Theory (CPI Manual), available from Update of the Consumer Price Index Manual (imf.org).

Diewert, W.E. and K.J. Fox (2022), "Substitution Bias in Multilateral Methods for CPI Construction", *Journal of Business & Economic Statistics* 40, 355-369.

Eltetö, Ö. and P. Köves (1964), "On an Index Computation Problem in International Comparisons" [in Hungarian], *Statiztikai Szemle* 42, 507-518.

Fox, K.J., P. Levell and M. O'Connell (2022), "Multilateral Index Number Methods for Consumer Price Statistics", ESCoE Discussion Paper 2022-08, Economic Statistics Centre of Excellence, London, United Kingdom.

Geary, R.C. (1958), "A Note on the Comparison of Exchange Rates and Purchasing Power between Countries", *Journal of the Royal Statistical Society A* 121, 97-99.

Gini C. (1931), "On the Circular Test of Index Numbers", *International Review of Statistics* 9, 3-25.

de Haan, J. (2022), "Adjusting for Compositional Change in Monthly Rent Indexes from "Big Data"", Unpublished paper, Australian Bureau of Statistics.

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

de Haan, J. and J. Daalmans (2019), "Scanner Data in the CPI: The Imputation CCDI Index Revisited", Paper presented at the 16th meeting of the Ottawa Group, 8-10 May 2019, Rio de Janeiro, Brazil.

Ivancic, L., W.E. Diewert and K.J. Fox (2011), "Scanner Data, Time Aggregation and the Construction of Price Indexes", *Journal of Econometrics* 161, 24-35.

Khamis, S.H. (1972), "A New System of Index Numbers for National and International Purposes", *Journal of the Royal Statistical Society A* 135, 96-121.

van Kints, M., J. de Haan and M. Webster (2019), "Utilizing Big Data and Multilateral Index Methods to Produce the Australian CPI: Theory, Implementation and Empirical Results", *Statistical Journal of the IAOS* 35, 387-405.

Lamboray, C. (2022), "What Impact Does Product Specification Have on a Fisher Price Index?", Paper presented at the 17th meeting of the Ottawa Group, 7-10 May 2022, Rome, Italy.

Nyborg Hov, K. and R. Nygaard (2022), "New Ways of Measuring Price Development on Consumer Electronics", Paper presented at the 17th meeting of the Ottawa Group, 7-10 May 2022, Rome, Italy.

Rao, D.S.P. (2005), "On the Equivalence of Weighted Country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Price Comparisons", *Review of Income and Wealth* 51, 571-580.

Reinsdorf, M.B. (1999), "Using Scanner Data to Construct CPI Basic Component Indexes", *Journal of Business & Economic Statistics* 17, 152-160.

Summers, R. (1973), "International Price Comparisons Based Upon Incomplete Data", *Review of Income and Wealth* 19, 1-16.

Szulc, B. (1964), "Index Numbers of Multilateral Regional Comparisons" [in Polish], *Przeglad Statystyczny* 3, 239-254.