A Comparison of Weighted Time Dummy Hedonic and Time-Product Dummy Indexes

Jan de Haan[a], **Rens Hendriks**[b] **and Michael Scholz**[c]

15 August 2016

Abstract: This paper compares two model-based multilateral price indexes: the time-product dummy (TPD) index and the time dummy hedonic (TDH) index, both estimated by expenditure-share weighted least squares regression. The TPD model can be viewed as the saturated version of the underlying TDH model, and we argue that the regression residuals are "biased towards zero" due to overfitting. We decompose the ratio of the two indexes in terms of average regression residuals of the new and disappearing items (plus a third component that depends on the change in the matched items' normalized expenditure shares). The decomposition explains under which conditions the TPD index suffers from quality-change bias or, more generally, lack-of-matching bias. An example using scanner data on men's t-shirts illustrates our theoretical framework.

Keywords: hedonic regression, multilateral price indexes, new and disappearing items, quality change, scanner data.

JEL Classification: C43, E31.

---

[a] Corresponding author; Division of Corporate Services, IT and Methodology, Statistics Netherlands, and OTB, Faculty of Architecture and the Built Environment, Delft University of Technology; email: j.dehaan@cbs.nl.

[b] Statistics for Development Division, Pacific Community (SPC); email: rensh@spc.int.

[c] Department of Economics, University of Graz, Austria; email: michael.scholz@uni-graz.at.

The authors would like to thank Johan Verburg for excellent assistance. The views expressed in this paper are those of the authors and do not necessarily reflect the views of Statistics Netherlands or SPC.

1. Introduction

Several statistical agencies, including Statistics Netherlands, have been investigating the use of multilateral index number methods for the treatment of scanner data in the CPI. A range of methods is available; for an overview, see De Haan, Willenborg and Chessa (2016). The motivation behind these methods is that the amount of matches in the data is maximized without running the risk of introducing chain drift, which occurs in certain circumstances for period-on-period chained (weighted) price indexes. Unfortunately, the international CPI Manual (ILO et al., 2004) does not pay much attention to multilateral index number methods.

De Haan (2015a) proposed the use of two related model-based multilateral price indexes for incorporating scanner data, both estimated by expenditure-share weighted least squares regression: the Time Dummy Hedonic (TDH) index in case information on item characteristics is available, and the Time-Product Dummy (TPD) index when this information is lacking.[1] A potential problem with the TPD method is that the resulting price index is not explicitly adjusted for quality change. The aim of the present paper is to examine what drives the difference between the two methods.

Our paper builds on work by Aizcorbe, Corrado and Doms (2003), Aizcorbe and Pho (2005), and in particular Silver and Heravi (2005) and Krsinich (2016). Silver and Heravi (2005) compared TPD and TDH indexes but only in a period-on-period chained context, where the bilateral TPD index equals a matched-model index. Krsinich (2016) argued that the use of longitudinal price information makes the multilateral TPD index implicitly quality-adjusted. It is certainly true that long-term price differences between items that coexist provide us with information about the value of quality differences. As with any implicit quality-adjustment method, however, this does not necessarily imply proper treatment of new and disappearing items and does not rule out the possibility of quality-change bias or, more generally, lack-of-matching bias.

Krsinich (2016) pointed out that the TPD model can be viewed as, what we call, the saturated version of a hedonic model with only categorical characteristics: the TPD model implicitly includes all the first- and higher-order interactions along with the main effects whereas a typical hedonic model would include only main effects. In our view, this means the TPD model has too many parameters, fits the outliers and unduly raises R squared as compared with the true underlying hedonic model. That is, the TPD model

[1] The name TPD method was suggested by De Haan and Krsinich (2014a) since it adapts the multilateral Country-Product Dummy (CPD) method proposed by Summers (1973) for spatial comparisons to price comparisons across time.

suffers from overfitting and “biases the regression residuals towards zero”. Another way to describe the problem is that overfitting potentially gives rise to biased out-of-sample predictions. Because quality adjustment boils down to imputing the “missing prices” of new and disappearing items, i.e. to making out-of-sample predictions, the TPD index is susceptible to quality-change bias.

The remainder of the paper is structured as follows. Section 2 outlines a number of different expressions for the TDH and TPD indexes, derives a decomposition of the ratio of the two indexes in terms of the average regression residuals for unmatched new and disappearing items, explains in greater detail why the TPD model is likely to suffer from overfitting, and investigates potential bias in the TPD index. Section 3 presents an empirical illustration using scanner data on men’s t-shirts sold by a major Dutch chain of department stores. Section 4 discusses our findings and concludes.

2. Formulas and decompositions

2.1 Formulas for TDH and TPD indexes

The following notation will be used: p_i^0 and p_i^t denote the price of item i in the base period 0 and in comparison period t $(t=1,...,T)$, respectively; s_i^0 and s_i^t are the item’s expenditure shares. Let us consider the following log-linear hedonic regression model to be estimated on the pooled data of all periods $0,...,T$:

$$\ln p_i^t = \delta^0 + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{k=1}^{K} \beta_k z_{ik} + \varepsilon_i^t, \tag{1}$$

where z_{ik} denotes (the quantity of) characteristic k for item i and β_k the corresponding parameter; the β_k are constrained to be fixed across time. The time dummy variable D_i^t has the value 1 if the observation pertains to period t $(t=0,..,T)$ and 0 otherwise; it is assumed that the errors ε_i^t are independently distributed with zero mean. The estimated parameters are denoted by $\hat{\delta}^0$, $\hat{\delta}^t$ $(t=1,...,T)$, and $\hat{\beta}_k$ $(k=1,..,K)$.

Following Diewert’s (2005) proposal, we assume that a Weighted Least Squares (WLS) regression is run with the expenditure shares in each period serving as weights. The TDH index going from period 0 to period t, $P_{TDH}^{0t} = \exp(\hat{\delta}^t)$,[2] can then be written as (De Haan and Krsinich, 2014b)

[2] Because exponentiating is a non-linear transformation, time dummy indexes are not unbiased; for a bias-correction term, see Kennedy (1981). Unless the number of observations is extraordinary small, the bias can be ignored. We will not make any corrections in the empirical section 3.

$$P_{TDH}^{0t} = \frac{\prod_{i \in S^t} (p_i^t / \exp[\sum_{k=1}^{K} \hat{\beta}_k z_{ik}])^{s_i^t}}{\prod_{i \in S^0} (p_i^0 / \exp[\sum_{k=1}^{K} \hat{\beta}_k z_{ik}])^{s_i^0}} = \frac{\prod_{i \in S^t} (p_i^t)^{s_i^t}}{\prod_{i \in S^0} (p_i^0)^{s_i^0}} \exp\left[\sum_{k=1}^{K} \hat{\beta}_k \left(\bar{z}_k^0 - \bar{z}_k^t\right)\right], \tag{2}$$

where S^0 and S^t denote the sets of items sold in periods 0 and *t*. The first expression of equation (2) writes the index as the ratio of weighted geometric averages of estimated *quality-adjusted prices* $p_i^0 / \exp[\sum_{k=1}^{K} \hat{\beta}_k z_{ik}]$ and $p_i^t / \exp[\sum_{k=1}^{K} \hat{\beta}_k z_{ik}]$, while the second expression adjusts the ratio of weighted geometric average prices for the change in the weighted average characteristics $\bar{z}_k^0 = \sum_{i \in S^0} s_i^0 z_{ik}$ and $\bar{z}_k^t = \sum_{i \in S^t} s_i^t z_{ik}$.

Next, suppose there are *N* different items sold in one or more periods across the sample period and consider the following TPD model for the pooled data:

$$\ln p_i^t = \alpha + \sum_{t=1}^{T} \delta^t D_i^t + \sum_{i=1}^{N-1} \gamma_i D_i + \varepsilon_i^t, \tag{3}$$

where D_i^t is the time dummy defined earlier. D_i is a dummy variable that has the value of 1 if the observation relates to item *i* and 0 otherwise; the dummy for an arbitrary item *N* is excluded to identify the model. The parameters γ_i are *fixed effects*, with $\gamma_N = 0$. The WLS-based TPD index between period 0 to period *t*, with estimated fixed effects $\hat{\gamma}_i$ $(\hat{\gamma}_N = 0)$, can be written as (De Haan and Hendriks, 2013)

$$P_{TPD}^{0t} = \frac{\prod_{i \in S^t} \left(p_i^t / \exp(\hat{\gamma}_i)\right)^{s_i^t}}{\prod_{i \in S^0} \left(p_i^0 / \exp(\hat{\gamma}_i)\right)^{s_i^0}} = \frac{\prod_{i \in S^t} (p_i^t)^{s_i^t}}{\prod_{i \in S^0} (p_i^0)^{s_i^0}} \exp\left(\bar{\hat{\gamma}}^0 - \bar{\hat{\gamma}}^t\right), \tag{4}$$

with $\bar{\hat{\gamma}}^0 = \sum_{i \in S^0} s_i^0 \hat{\gamma}_i^0$ and $\bar{\hat{\gamma}}^t = \sum_{i \in S^t} s_i^t \hat{\gamma}_i^t$. The TPD model (3) is a in fact a special case of the TDH model (1) where the hedonic price effects $\sum_{k=1}^{K} \beta_k z_{ik}$ are replaced by fixed effects γ_i. The fixed effects estimates $\hat{\gamma}_i$ can therefore be viewed as approximations to the estimated hedonic price effects $\sum_{k=1}^{K} \hat{\beta}_k z_{ik}$, up to an additive scalar.

Now consider the following model for the relation between $\hat{\gamma}_i$ and $\sum_{k=1}^{K} \hat{\beta}_k z_{ik}$:

$$\hat{\gamma}_i = a^t + b^t \sum_{k=1}^{K} \hat{\beta}_k z_{ik} + e_i^t \quad (t = 0,..., T; i = 1,..., N-1; \hat{\gamma}_N = 0), \tag{5}$$

where e_i^t is an error term with zero mean. The coefficients obtained from expenditure-share weighted regressions of model (5) separately for each time period t $(t = 0,..., T)$ are denoted by $\tilde{a}^t$ and $\tilde{b}^t$. Since the weighted residuals sum to zero, we have

$$\bar{\hat{\gamma}}^t = \tilde{a}^t + \tilde{b}^t \sum_{k=1}^{K} \hat{\beta}_k \bar{z}_k^t. \tag{6}$$

Substituting (6) for periods 0 and t $(t = 1,..., T)$ into (4), dividing the result by equation (2), and a bit of rearranging of terms yields a decomposition of the TPD to TDH ratio in terms of changes in the intercept terms of equation (6), changes in the slope coefficients and changes in the weighted average characteristics:

$$\frac{P_{TPD}^{0t}}{P_{TDH}^{0t}} = \exp\left(\tilde{a}^0 - \tilde{a}^t\right)\exp\left[\left(\tilde{b}^0 - \tilde{b}^t\right)\sum_{k=1}^{K}\hat{\beta}_k \bar{z}_k^0\right]\exp\left[\left(\tilde{b}^t - 1\right)\sum_{k=1}^{K}\hat{\beta}_k\left(\bar{z}_k^0 - \bar{z}_k^t\right)\right]. \tag{7}$$

Only when the intercept terms are the same in period 0 and period t $(\tilde{a}^0 = \tilde{a}^t)$ and the slope coefficients are equal to 1 $(\tilde{b}^0 = \tilde{b}^t = 1)$ will the TPD index be equal to the TDH index. In practice, these conditions are unlikely to hold. Firstly, WLS produces unstable coefficients if the errors in model (5) are homoscedastic. Secondly, and more importantly, systematic changes in the coefficients (or the average characteristics) may occur. For example, if $\tilde{a}^t$ increases over time and everything else remains the same, the first component of (7) becomes increasingly smaller than 1, causing downward bias in the TPD index relative to the TDH index.

2.2 A decomposition in terms of regression residuals

We denote the prices predicted by any method by $\hat{p}_i^0$ and $\hat{p}_i^t$. Using the least-squares property that the weighted residuals $u_i^0 = \ln(p_i^0) - \ln(\hat{p}_i^0)$ and $u_i^t = \ln(p_i^t) - \ln(\hat{p}_i^t)$ from the (WLS) TDH and TPD regressions sum to zero in each period, we have

$$\prod_{i \in S^0}\left(\frac{\hat{p}_i^0}{p_i^0}\right)^{s_i^0} = \prod_{i \in S^t}\left(\frac{\hat{p}_i^t}{p_i^t}\right)^{s_i^t} = 1. \tag{8}$$

Thus, an initial expression for the TDH and TPD indexes $P^{0t} = \hat{p}_i^t / \hat{p}_i^0$ is

$$P^{0t} = \prod_{i \in S^0}\left(\frac{\hat{p}_i^t}{p_i^0}\right)^{s_i^0} = \prod_{i \in S^t}\left(\frac{p_i^t}{\hat{p}_i^0}\right)^{s_i^t} = \prod_{i \in S^0}\left(\frac{\hat{p}_i^t}{p_i^0}\right)^{\frac{s_i^0}{2}}\prod_{i \in S^t}\left(\frac{p_i^t}{\hat{p}_i^0}\right)^{\frac{s_i^t}{2}}. \tag{9}$$

The first and second expression of (9) are a geometric Laspeyres-type and Paasche-type price index defined on a *dynamic universe*, where the period t prices for all $i \in S^0$ and the period 0 prices for all $i \in S^t$, respectively, are imputed by predicted values. The two indexes are constrained to be the same, hence equal to their geometric mean given by the third expression of (9), which defines a Törnqvist-type price index.

We now subdivide S^0 and S^t into matched and unmatched items: $S_M^{0t} = S^0 \cap S^t$ is the set of matched items between periods 0 and t; S_D^0 is the subset of S^0 consisting of disappearing items that are not sold in period t $(S_D^0 \cup S_M^{0t} = S^0)$; S_N^t is the subset of S^t

consisting of new items which were not yet sold in period 0 ($S_N^0 \cup S_M^{0t} = S^t$). The last expression of equation (9) then becomes

$$P^{0t} = \prod_{i \in S_M^{0t}} \left(\frac{p_i^t}{p_i^0} \right)^{\frac{s_i^0 + s_i^t}{2}} \prod_{i \in S_D^0} \left(\frac{\hat{p}_i^t}{p_i^0} \right)^{\frac{s_i^0}{2}} \prod_{i \in S_N^t} \left(\frac{p_i^t}{\hat{p}_i^0} \right)^{\frac{s_i^t}{2}} \left[\frac{\prod_{i \in S_M^{0t}} \left(\frac{\hat{p}_i^t}{p_i^t} \right)^{s_i^0}}{\prod_{i \in S_M^{0t}} \left(\frac{\hat{p}_i^0}{p_i^0} \right)^{s_i^t}} \right]^{\frac{1}{2}} . \tag{10}$$

Equation (10) provides some underpinning for the use of expenditure-share weighted regression to estimate the TDH and TPD models. The first three terms on the right-hand side define a *single imputation* Törnqvist price index, P_{SIT}^{0t}, where the "missing prices", i.e. the period t prices for $i \in S_D^0$ and the period 0 prices for $i \in S_N^t$, are imputed.[3] Single imputation price indexes typically apply predicted values based on regressions for each time period separately, but P_{SIT}^{0t} is based on predicted values from a pooled regression. P_{SIT}^{0t} is not transitive, hence dependent on the choice of base period; the fourth term of (10) turns P_{SIT}^{0t} into the transitive (TDH or TPD) index P^{0t}.

To gain more insight in what actually drives the difference between the TPD and TDH indexes, we will decompose the ratio of the TPD and TDH indexes (estimated on the same data set) in terms of the average regression residuals for the unmatched new and disappearing items. From the first expression of (8) it follows that

$$\prod_{i \in S_D^0} \left(\frac{\hat{p}_{i(TPD)}^0}{p_i^0} \right)^{s_i^0} \prod_{i \in S_M^{0t}} \left(\frac{\hat{p}_{i(TPD)}^0}{\hat{p}_{i(TPD)}^t} \right)^{s_i^0} \prod_{i \in S_M^{0t}} \left(\hat{p}_{i(TPD)}^t \right)^{s_i^0} = \prod_{i \in S_D^0} \left(\frac{\hat{p}_{i(TDH)}^0}{p_i^0} \right) \prod_{i \in S_M^{0t}} \left(\frac{\hat{p}_{i(TDH)}^0}{\hat{p}_{i(TDH)}^t} \right)^{s_i^0} \prod_{i \in S_M^{0t}} \left(\hat{p}_{i(TDH)}^t \right)^{s_i^0} ,$$

which, by defining $s_{iD}^0 = s_i^0 / s_D^0$, $s_{iM}^0 = s_i^0 / s_M^0$, $s_D^0 = \sum_{i \in S_D^0} s_i^0$, and $s_M^0 = \sum_{i \in S_M^{0t}} s_i^0$, and after some manipulation, yields

$$\frac{P_{TPD}^{0t}}{P_{TDH}^{0t}} = \left[\frac{\prod_{i \in S_D^0} \left(\frac{\hat{p}_{i(TPD)}^0}{p_i^0} \right)^{s_{iD}^0}}{\prod_{i \in S_D^0} \left(\frac{\hat{p}_{i(TDH)}^0}{p_i^0} \right)^{s_{iD}^0}} \right]^{\frac{s_D^0}{s_M^0}} \frac{\prod_{i \in S_M^{0t}} \left(\hat{p}_{i(TPD)}^t \right)^{s_{iM}^0}}{\prod_{i \in S_M^{0t}} \left(\hat{p}_{i(TDH)}^t \right)^{s_{iM}^0}} . \tag{11}$$

Similarly, from the second expression of (8) it follows that

---

[3] For the two period case and a particular set of regression weights, De Haan (2004) showed that the TDH index is equivalent to a single imputation Törnqvist index. For a comparison of time dummy hedonic and hedonic imputation indexes, see Diewert, Heravi and Silver (2009) and De Haan (2010).

$$\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TPD)}}{p_i^t}\right)^{s_i^t}\prod_{i\in S_M^{0t}}\left(\frac{\hat{p}^t_{i(TPD)}}{\hat{p}^0_{i(TPD)}}\right)^{s_i^t}\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TPD)}\right)^{s_i^t}=\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TDH)}}{p_i^t}\right)\prod_{i\in S_M^{0t}}\left(\frac{\hat{p}^t_{i(TDH)}}{\hat{p}^0_{i(TDH)}}\right)^{s_i^t}\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TDH)}\right)^{s_i^t},$$

which, using $s^t_{iN}=s^t_i/\sum_{i\in S_N^t}s_i^t$, $s^t_{iM}=s^t_i/\sum_{i\in S_M^{0t}}s_i^t$, $s^t_N=\sum_{i\in S_N^t}s_i^t$, and $s^t_M=\sum_{i\in S_M^{0t}}s_i^t$, and again after some manipulation leads to

$$\frac{P^{0t}_{TPD}}{P^{0t}_{TDH}}=\left[\frac{\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TPD)}}{p_i^t}\right)^{s^t_{iN}}}{\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TDH)}}{p_i^t}\right)^{s^t_{iN}}}\right]^{\frac{s^t_N}{s^t_M}}\frac{\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TPD)}\right)^{s^t_{iM}}}{\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TDH)}\right)^{s^t_{iM}}}. \tag{12}$$

Taking the geometric mean of (11) and (12) yields

$$\frac{P^{0t}_{TPD}}{P^{0t}_{TDH}}=\left[\frac{\prod_{i\in S_D^0}\left(\frac{\hat{p}^0_{i(TPD)}}{p_i^0}\right)^{s^0_{iD}}}{\prod_{i\in S_D^0}\left(\frac{\hat{p}^0_{i(TDH)}}{p_i^0}\right)^{s^0_{iD}}}\right]^{\frac{s^0_D}{2s^0_M}}\left[\frac{\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TDH)}}{p_i^t}\right)^{s^t_{iN}}}{\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TPD)}}{p_i^t}\right)^{s^t_{iN}}}\right]^{\frac{s^t_N}{2s^t_M}}\left[\frac{\prod_{i\in S_M^{0t}}\left(\hat{p}^t_{i(TPD)}\right)^{s^0_{iM}}}{\prod_{i\in S_M^{0t}}\left(\hat{p}^t_{i(TDH)}\right)^{s^0_{iM}}}\frac{\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TDH)}\right)^{s^t_{iM}}}{\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TPD)}\right)^{s^t_{iM}}}\right]^{\frac{1}{2}}. \tag{13}$$

The third term of decomposition (13) can be written as

$$\left[\frac{\prod_{i\in S_M^{0t}}\left(\hat{p}^t_{i(TPD)}\right)^{s^0_{iM}}}{\prod_{i\in S_M^{0t}}\left(\hat{p}^t_{i(TDH)}\right)^{s^0_{iM}}}\frac{\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TDH)}\right)^{s^t_{iM}}}{\prod_{i\in S_M^{0t}}\left(\hat{p}^0_{i(TPD)}\right)^{s^t_{iM}}}\right]^{\frac{1}{2}}=\left[\frac{P^{0t}_{TPD}}{P^{0t}_{TDH}}\right]^{\frac{1}{2}}\left[\frac{\prod_{i\in S_M^{0t}}\left(\frac{\hat{p}^0_{i(TPD)}}{\hat{p}^0_{i(TDH)}}\right)^{s^0_{iM}}}{\prod_{i\in S_M^{0t}}\left(\frac{\hat{p}^0_{i(TPD)}}{\hat{p}^0_{i(TDH)}}\right)^{s^t_{iM}}}\right]^{\frac{1}{2}}.$$

Substituting this result into (13) and solving for $P^{0t}_{TPD}/P^{0t}_{TDH}$ gives

$$\frac{P^{0t}_{TPD}}{P^{0t}_{TDH}}=\left[\frac{\prod_{i\in S_D^0}\left(\frac{\hat{p}^0_{i(TPD)}}{p_i^0}\right)^{s^0_{iD}}}{\prod_{i\in S_D^0}\left(\frac{\hat{p}^0_{i(TDH)}}{p_i^0}\right)^{s^0_{iD}}}\right]^{\frac{s^0_D}{s^0_M}}\left[\frac{\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TDH)}}{p_i^t}\right)^{s^t_{iN}}}{\prod_{i\in S_N^t}\left(\frac{\hat{p}^t_{i(TPD)}}{p_i^t}\right)^{s^t_{iN}}}\right]^{\frac{s^t_N}{s^t_M}}\frac{\prod_{i\in S_M^{0t}}\left(\frac{\hat{p}^0_{i(TPD)}}{\hat{p}^0_{i(TDH)}}\right)^{s^0_{iM}}}{\prod_{i\in S_M^{0t}}\left(\frac{\hat{p}^0_{i(TPD)}}{\hat{p}^0_{i(TDH)}}\right)^{s^t_{iM}}}. \tag{14}$$

Equation (14) can be written in terms of average regression residuals, as follows:

$$\frac{P^{0t}_{TPD}}{P^{0t}_{TDH}}=\exp\left[\frac{s^0_D}{s^0_M}\left(\bar{u}^0_{D(TDH)}-\bar{u}^0_{D(TPD)}\right)\right]\exp\left[\frac{s^t_N}{s^t_M}\left(\bar{u}^t_{N(TPD)}-\bar{u}^t_{N(TDH)}\right)\right]\prod_{i\in S_M^{0t}}\left(\frac{\hat{p}^0_{i(TPD)}}{\hat{p}^0_{i(TDH)}}\right)^{s^0_{iM}-s^t_{iM}}, \tag{15}$$

where $\bar{u}^0_{D(TPD)} = \sum_{i \in S^0_D} s^0_{iD} u^0_{i(TPD)}$, $\bar{u}^0_{D(TDH)} = \sum_{i \in S^0_D} s^0_{iD} u^0_{i(TDH)}$, $\bar{u}^t_{N(TPD)} = \sum_{i \in S^t_N} s^t_{iN} u^t_{i(TPD)}$ and $\bar{u}^t_{N(TDH)} = \sum_{i \in S^t_N} s^t_{iN} u^t_{i(TDH)}$. Equation (15) decomposes the ratio of the TPD index and the TDH index into three components. The first and second components are driven by the differences in the weighted average residuals for the disappearing and new items from the TPD and TDH regressions. The magnitude of these components also depends on the (relative) aggregate expenditure shares of the matched and unmatched items, s^0_M, s^t_M, $s^0_D = 1 - s^0_M$ and $s^t_N = 1 - s^t_M$.

The third component of (15) can be written in terms of the period 0 residuals for the matched items, for example as

$$\prod_{i \in S^{0t}_M} \left(\frac{\hat{p}^0_{i(TPD)}}{\hat{p}^0_{i(TDH)}} \right)^{s^0_{iM} - s^t_{iM}} = \exp\left[\left(\bar{u}^{0(t)}_{M(TPD)} - \bar{u}^0_{M(TPD)} \right) - \left(\bar{u}^{0(t)}_{M(TDH)} - \bar{u}^0_{M(TDH)} \right) \right], \tag{16}$$

with $\bar{u}^0_{M(TPD)} = \sum_{i \in S^{0t}_M} s^0_{iM} u^0_{i(TPD)}$, $\bar{u}^{0(t)}_{M(TPD)} = \sum_{i \in S^{0t}_M} s^t_{iM} u^0_{i(TPD)}$, $\bar{u}^0_{M(TDH)} = \sum_{i \in S^{0t}_M} s^0_{iM} u^0_{i(TDH)}$ and $\bar{u}^{0(t)}_{M(TDH)} = \sum_{i \in S^{0t}_M} s^t_{iM} u^0_{i(TDH)}$. This term thus depends on the normalized expenditure shares of the matched items. It generally differs from 1, even without any new or disappearing items, because WLS time dummy results are model dependent. This third term may be large when the matched-items' expenditures shares in periods 0 and *t* differ significantly and will be equal to 1 in the unlikely event that the shares remain constant over time. Note that this term would also be equal to 1 if unweighted (OLS) regressions had been run instead of weighted regressions.

2.3 A priori expectations

With a dynamic universe, the TPD and TDH indexes are most likely to differ. There are two issues at stake: variability (variance) and systematic difference (bias). Suppose the average residuals of the new and disappearing items in (15) fluctuate randomly around 0 across time, for both the TPD and TDH regressions, and the third component fluctuates around 1. While the two indexes may differ in each time period, they are expected to exhibit equal trends.

According to Krsinich (2016), the TPD index is implicitly quality-adjusted due to the use of longitudinal price information. But when quality change is important, there is no a priori reason to expect that the average residuals from the TPD regression for the new and disappearing items will be approximately equal to, or show the same trend as, those from the TDH regression. More specifically, the TPD residuals for the unmatched items tend to be "biased towards zero" as compared with the TDH residuals. Below, we will explain why this is the case.

In practice, items can be identified by a finite number of observable attributes, the range of possible values being discrete rather than continuous. In other words, a set of categorical variables (some of which will be ordinal) for each attribute can describe the different items belonging to the product category. Suppose we cross-classify all the categorical variables and know to which cell each item belongs. Obviously, many cells will be empty since not all combinations are feasible to produce or sell. Suppose further that we postulate a TDH model using additive dummy variables for the *main effects* and multiplicative dummies for all the *interaction terms*. As was shown by Krsinich (2016), this fully interacted or, as we will call it, saturated TDH model is essentially equivalent to the TPD model.[4]

A problem with the saturated model is that it includes many irrelevant variables. Importantly, the inclusion of interaction terms – and certainly the higher-order terms – in a hedonic model is difficult to justify because of interpretation problems. In a typical hedonic model, one would only find main effects for categorical variables and perhaps some first-order interaction terms. In addition, the TPD model will implicitly include all the variables that are incorporated into the key that is used to identify items, including attributes which are not deemed important from the consumers' point of view. One such key is GTIN (EAN, UPC), the usual key for exact matching of items and tracking prices over time.

The implicit inclusion of irrelevant variables in the TPD model is likely to lead to *overfitting*, in particular as compared to a TDH model that includes only main effects for attributes that are deemed important from the consumers' perspective. That is, the TPD model will fit the outliers, unduly raise R squared, and "bias the residuals towards zero". Note that items which are observed only once during the sample period lie on the regression surface so that their residuals are exactly equal to zero, but this is probably only a minor part of the problem. Econometrics textbooks tell us that the inclusion of irrelevant variables does not yield biased results. This result is, however, conditional on the sample data. If we want the exponentiated time dummy coefficient from a TPD or TDH regression to be a quality-adjusted price index, imputation of the "missing prices" for unmatched items is required; see equation (10). These imputations are *out-of-sample predictions* and can be biased.

[4] In Krsinich's (2016) data set on consumer electronics products, items are identified by cross classifying the available categorical characteristics. For laptop computers, one of the characteristics is weight. While laptop weight is positively correlated with price, it should not be included in a hedonic model. This is because, other things equal, consumers prefer lighter laptops over heavier ones and inclusion of weight can have adverse quality-adjustment effects.

While bias can also arise for the TDH model, the TPD model is likely to be more affected as overfitting makes out-of-sample prediction very problematic. Put differently, we expect significant differences in the imputed "missing prices" for unmatched items between the TPD and TDH methods. When the *average* imputations differ significantly, substantial differences between the TPD index and the TDH index can arise. As shown by equation (15), the ratio of the two indexes can be analyzed by comparing the average residuals for the unmatched new and disappearing items instead of the average imputed values.

Silver and Heravi (2005) argued that the difference in the average residuals for the matched and unmatched new and disappearing items is the driver of the difference between a hedonic index and the corresponding matched-model index. Using scanner data for several consumer electronics products, they found generally negative average residuals (i.e., relatively low observed prices, given their characteristics) for old models, or disappearing items in our language, and positive average residuals (relatively high observed prices) for new models. They attributed this result to the *pricing strategies* of retailers and manufacturers: inventory cleaning, or dumping, in case of old models and price skimming for new models.

To see what can happen and to simplify matters somewhat, let us assume that the aggregate expenditure shares of the new and disappearing items in equation (15) are the same ($s_D^0 = s_N^t = s_{UM}^{0t}$) and the third component equals 1. Equation (15) then reduces to

$$\frac{P_{TPD}^{0t}}{P_{TDH}^{0t}} = \exp\left[\frac{s_{UM}^{0t}}{1-s_{UM}^{0t}}\left\{\left(\bar{u}_{N(TPD)}^{t} - \bar{u}_{D(TPD)}^{0}\right) - \left(\bar{u}_{N(TDH)}^{t} - \bar{u}_{D(TDH)}^{0}\right)\right\}\right]. \tag{17}$$

Suppose that the above inventory cleaning and price skimming strategies apply and we have $\bar{u}_{N(TDH)}^{t} > 0$ and $\bar{u}_{D(TDH)}^{t} < 0$, or $\bar{u}_{N(TDH)}^{t} - \bar{u}_{D(TDH)}^{0} > 0$. Because the residuals from the TPD regression for both the new and disappearing items are "biased towards zero", we would expect to find $\bar{u}_{N(TPD)}^{t} - \bar{u}_{D(TPD)}^{0} < \bar{u}_{N(TDH)}^{t} - \bar{u}_{D(TDH)}^{0}$, hence $P_{TPD}^{0t} < P_{TDH}^{0t}$. Under these pricing strategies, the TPD index is thus most likely downward biased compared to the TDH index.

Dumping and price skimming is not confined to products where quality change due to technical progress is important, such as consumer electronics goods; such pricing strategies can be found for groceries as well (Melser and Syed, 2015). An extreme case, referred to by Chessa (2015) (2016) as *re-launching*, arises when items are replaced by "new" items which are basically the same goods, apart perhaps from minor differences in packaging, but with different GTINs. This phenomenon seems to occur regularly for a number of product categories in the Netherlands. The prices of the replacement items are often higher than those of the replaced items; apparently, the retailers/manufacturers have some degree of market power and consumers are unable to substitute away from the replacements. If items are defined by GTIN, the TPD method and (other) matched-model methods cannot pick up disguised price increases due to re-launches,[5] in contrast to the TDH method.

We do not know the true underlying hedonic model, of course. Careful selection of characteristics and limited use of interaction terms is required, but some arbitrariness cannot be avoided. This is a disadvantage as compared with the TPD method. Another potential problem is the restrictive nature of the TDH model due to parameter fixity; the same applies to the fixed-effects TPD model. The least we can do is regularly update the coefficients and somehow chain link the short-term index series. We will return to these issues in the concluding section 4.

3. Empirical illustration

For an empirical illustration, we will use scanner data on men's t-shirts. The data runs from February 2009 to March 2013 and covers all the department stores belonging to a major Dutch retail chain. In addition to prices, i.e. monthly unit values, and quantities sold, we have information on six categorical attributes which has been extracted from the product descriptions:[6] shape of the neck (O or V), fabric (basic or organic), sleeve length (short or long); number of t-shirts per package (1, 2 or 3), color (white, black or other), and fit (normal or stretch).

3.1 EAN as item identifier

In our data set, items are identified by barcode or European Article Number (EAN), the European version of GTIN. Across the four-year sample period, 1953 different items were sold. The item turnover rate is high: from the more than 500 items that were sold in the first month, only 10% were still sold in the last month. The total number of items

---

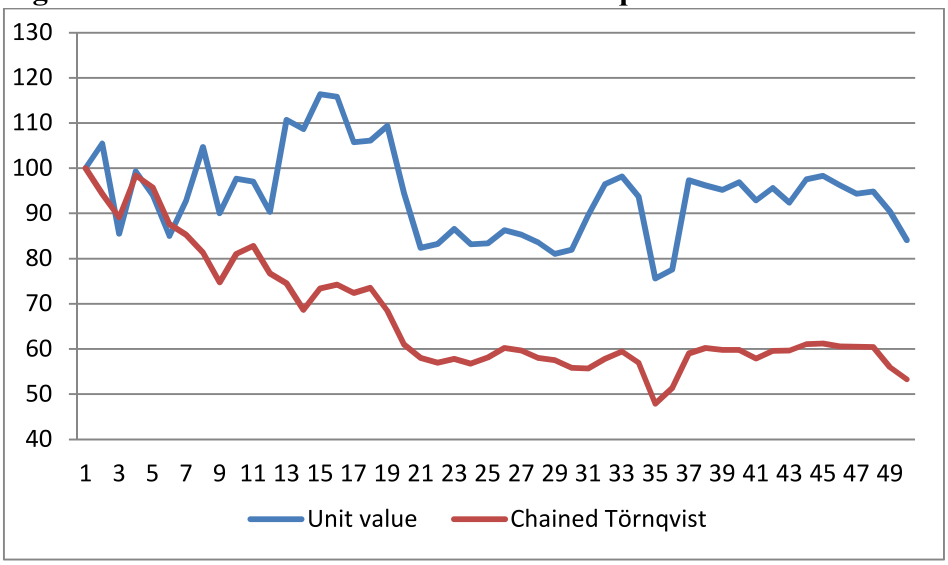
[5] It has been known for a long time that the GTIN level can sometimes be too fine for index construction; see e.g. Reinsdorf (1999) and De Haan (2002). Retailers' internal product codes or Stock Keeping Units (SKUs) are more stable and may be a more suitable choice. The Australian Bureau of Statistics uses SKU as item identifier for the treatment of scanner data (Howard et al., 2015).

[6] This was done manually by Antonio Chessa from Statistics Netherlands. The agency has recently started a project to automatically extract characteristics information from product descriptions using text mining or machine learning techniques. For first results, see De Boer (2016).

sold in each month is huge. Many packages with different EANs probably contain the same physical product or can be described by the same set of attributes so that the "true" rate of product churn may be overstated. In section 3.2, we will analyse what happens when items are defined by their attributes rather than at the barcode level.

Figure 1 plots two indexes based on EAN as item identifier: the unit value index and the monthly chained Törnqvist price index. Both indexes have their problems. The unit value index is defined as the ratio of total expenditure divided by total quantities sold in the periods compared. It is affected by compositional change, giving rise to a volatile time series and possibly also a wrong trend. Weighted price indexes, including superlative price indexes such as the Törnqvist, are prone to chain drift when consumers stock up goods during sales periods (Ivancic, Diewert and Fox, 2011, and De Haan and Van der Grient, 2011). The chained Törnqvist index does indeed have downward drift, especially during the first half of the sample period.

Figure 1. Unit value index and chained Törnqvist index



Although multilateral price indexes are free from chain drift by construction, this does not mean that they are necessarily unbiased. Figure 2 shows the expenditure-share weighted TPD and TDH indexes. The TDH index shows a plausible trend, but the TPD index seems to be severely downward biased. The bias in the TPD index mainly arises in months 12 and 13 when "organic" t-shirts were introduced in the stores and to a large extent replaced "basic" t-shirts. Surprisingly, the volatility of the TDH index is of the same order of magnitude as that of the unit value index, in spite of the fact that the TDH method controls for quality mix changes.

Figure 2. TPD index and TDH index

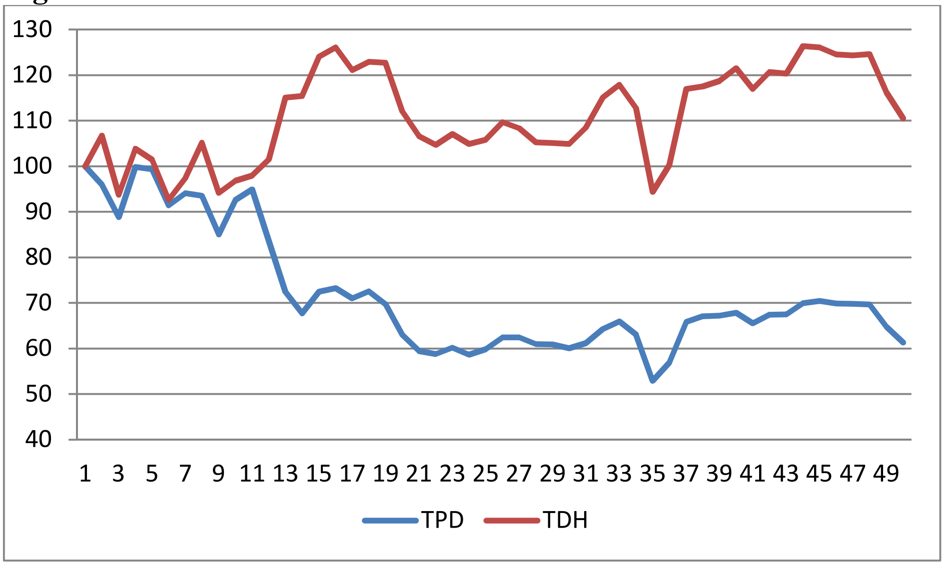


Table 1 contains the regression results for the weighted TDH model. “Organic” t-shirts are less expensive than “basic” t-shirts, other things equal. The signs of the other coefficients are as expected. Partly due to the large number of observations (24,797), the coefficients are highly significant, except for the attributes “shape of neck” (“V”) and “fit” (“stretch”). The R-squared value from the TDH regression (0.7607) is satisfactory, but nevertheless much lower than that from the TPD regression (0.9108). This confirms our suspicion that the TPD method unduly raises R-squared as compared with the TDH method.

Table 1. Regression results for TDH model (excluding time dummies)

Attribute	Dummy	Estimate	Std. error	t value	Signif.
Intercept		1.947536	0.009396	207.27	***
Shape of neck	V	0.006629	0.003154	2.10	*
Fabric	Organic	-0.237870	0.006555	-36.29	***
Sleeve length	Long	0.209994	0.004885	42.99	***
# T-shirts per package	2	0.489098	0.004065	120.31	***
	3	0.638112	0.005641	113.12	***
Color	White	-0.047670	0.003413	-13.97	***
	Black	-0.035280	0.005025	-7.02	***
Fit	Stretch	-0.021716	0.006636	-3.27	**

Significance codes: *** 0.001; ** 0.01; * 0.05
observations: 24,797
R-squared: 0.7607; Adjusted R-squared: 0.7601

Before turning to our main decomposition (15) of the TPD to TDH ratio, let us have a look at decomposition (7). The latter is based on a simple linear regression of the estimated TPD fixed effects against the estimated TDH hedonic price effects. Figure 3 plots the regression coefficients. As of month 14, the coefficients are quite stable, with the slope coefficient being close to the "desired" value of 1. Prior to month 14 though, the coefficients are extremely volatile; the TPD fixed effects are poor approximations of the hedonic price effects. The results of decomposition (7) in Figure 4 indicate that the change in the intercept term (the first component) mainly drives the change in the TPD to TDH ratio.

Figure 3. Coefficients from regression model (5)

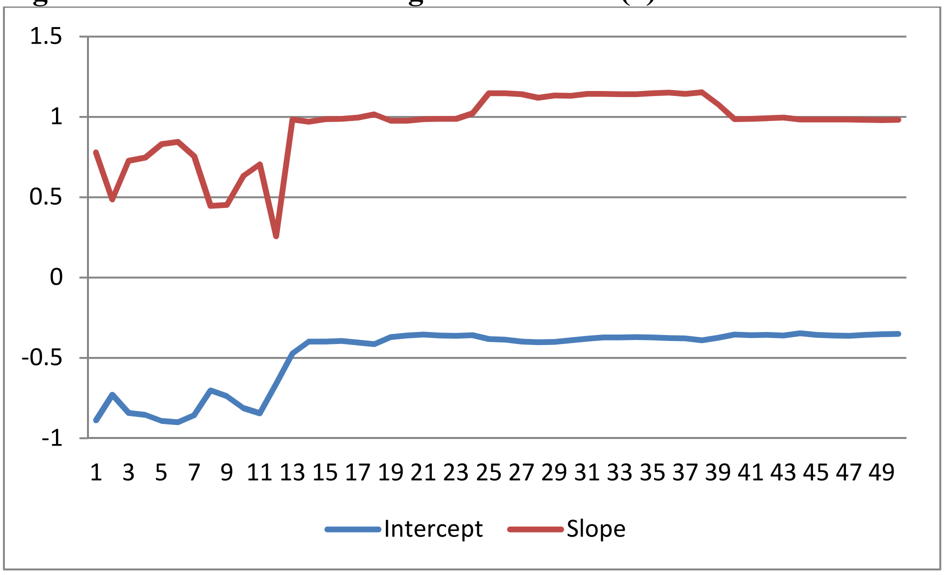
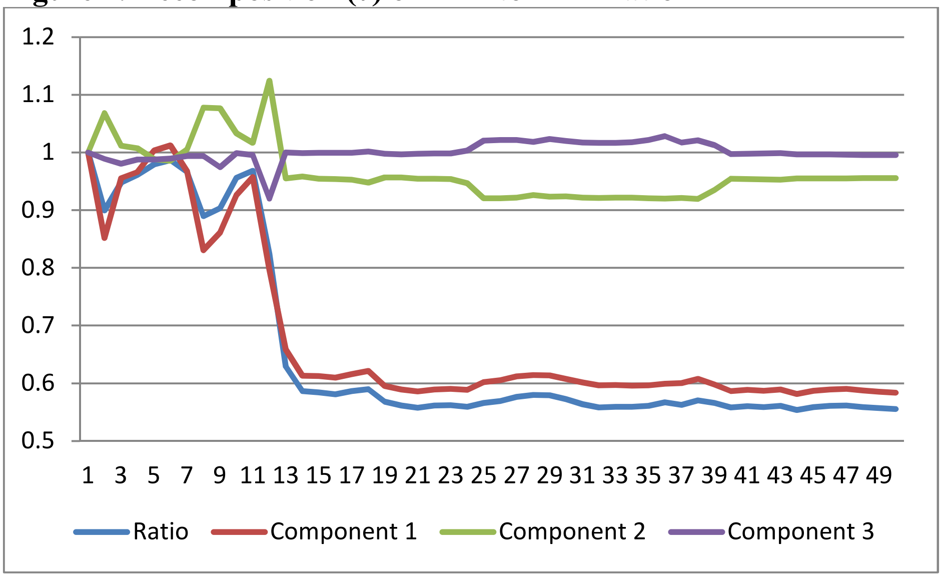


Figure 4. Decomposition (7) of TPD to TDH ratio



Decomposition (7), while instructive, says little about the causes of the change in the TPD to TDH ratio. This is where decomposition (15) comes into play. Here, the weighted average regression residuals of the new and disappearing items (with respect to the first or base month) are important drivers of the TPD to TDH ratio. As shown by Figure 5, the average TDH residuals tend to be positive for new items and negative for disappearing items during the first half of the sample period. The absolute values of the average TPD residuals are much smaller; they are "biased towards zero". These findings are consistent with the TPD index sitting below the TDH index. During the second half of the sample period, the average residuals for the new and disappearing items from the two regressions turn out to be very small.

Figure 5. Weighted average regression residuals

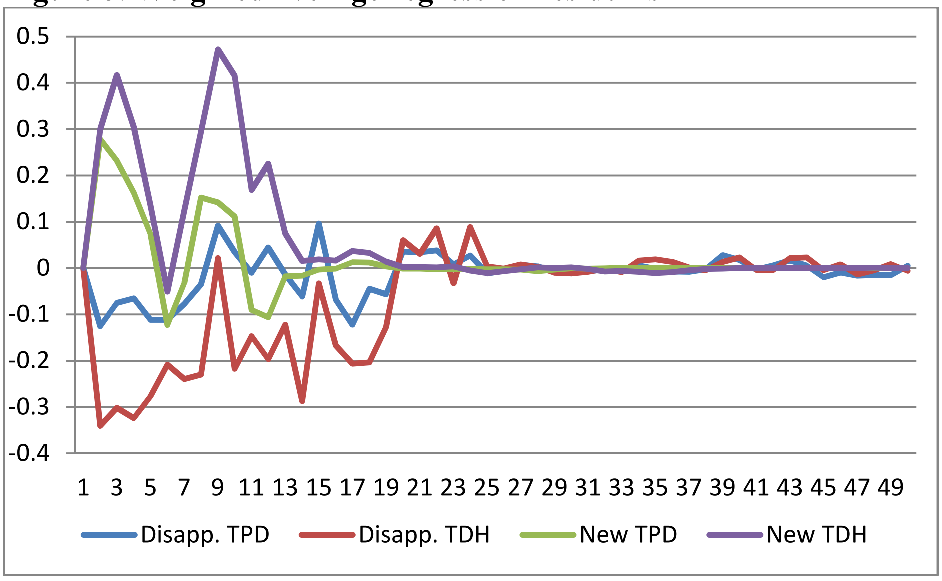


Figure 6 plots the aggregate expenditure shares of the unmatched and matched items. Two months witness dramatic changes with respect to the preceding months. In month 13, the expenditure share of new items rises from 0.08 to 0.84, and in month 25, the expenditure share of disappearing items rises from 0.09 to 0.62. The shares of the matched items in periods 0 and t drop accordingly. Note that the ratio of the period 0 aggregate expenditure shares for the disappearing and matched items and the ratio of the period t aggregate expenditure shares for the new and matched items cause *leveraging* in decomposition (15). The bigger these relative aggregate expenditure shares are, the more important the differences between the average residuals for the disappearing and new items from the TPD and TDH regressions become.

Figure 6. Aggregate expenditure shares

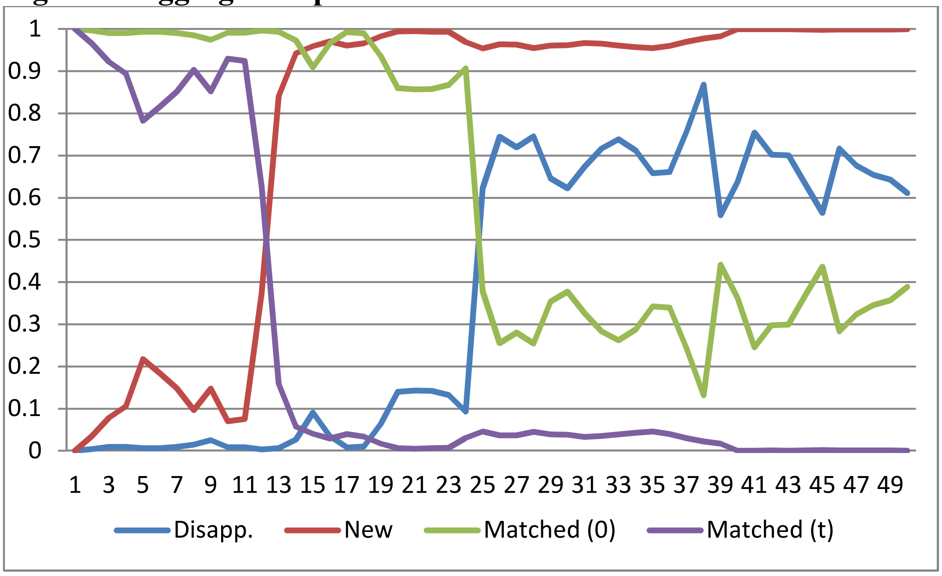
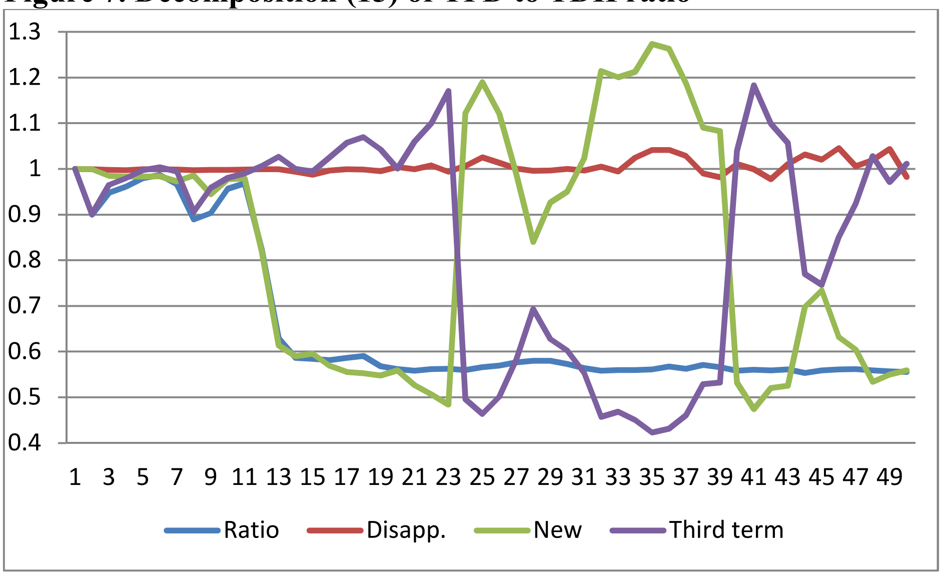


Figure 7 shows the results of decomposition (15). As we already saw, the TPD to TDH index ratio quite remains stable after the introduction of "organic" t-shirts. This suggests that the TPD method performs well unless a structural break in the assortment takes place, in the sense of a sudden introduction of many new items with a significant expenditure share. New items contribute most to the TPD to TDH ratio, except between months 22 and 38 when the third term takes over. In that period, the average residuals of the new items, although small, are positive and have high leverage. Interestingly, the contribution of the disappearing items is negligible.

Figure 7. Decomposition (15) of TPD to TDH ratio



3.2 Some results at group level

A straightforward way of getting rid of re-launches and disguised price changes is to identify items by cross classifying the (categorical) attributes rather than by barcode. In doing so, any difference between the TPD and TDH indexes can be entirely attributed to the implicit use of first-order and higher-order interaction terms in the TPD regression. A full cross-classification of the six categorical attributes available in or scanner data set yields $2 \times 2 \times 2 \times 3 \times 3 \times 2 = 144$ possible combinations or "groups", as we will call them, but only 37 of those are actually found in the data. Prices are calculated as unit values across all the EANs belonging to the respective groups.

A comparison of Figure 8 with Figure 2 reveals that the group-based TPD index differs substantially from the EAN-based TPD index. This shows how sensitive to the choice of item identifier the TPD method can be. Assuming the groups can be viewed as homogeneous products, there must have been disguised price increases the EAN-based TPD index was unable to pick up, which is tantamount to saying there must have been a lack of matching. In accordance with our expectations, the group-based TDH index is very similar to its EAN-based counterpart. Although the group-based TPD index comes close to the group-based TDH index, a gap remains.

Figure 8. TPD index and TDH index, group level

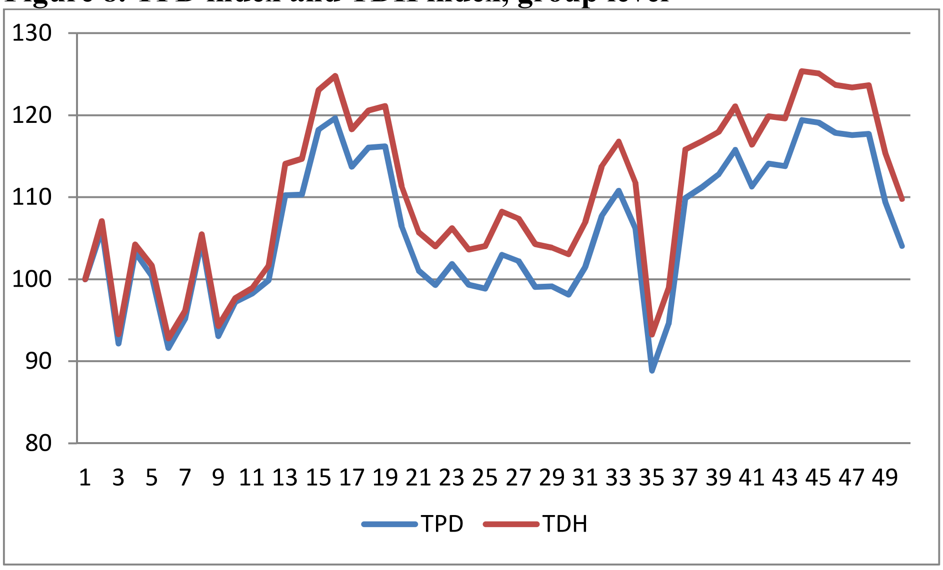


Table 2 contains the regression results for the expenditure-share weighted TDH model at the group level. At this level, the number of observations is considerably lower (1,289) than at the EAN level. The coefficients for "shape of neck" and "fit" have now become insignificant. The difference between the R-squared values from the TPD and

TDH regressions is rather small at the group level, 0.8844 versus 0.8664, which is not surprising since group churn is modest compared with EAN churn. Because aggregation of EANs into groups reduces noise in the prices data, R-squared is higher at the group level than at the EAN level.

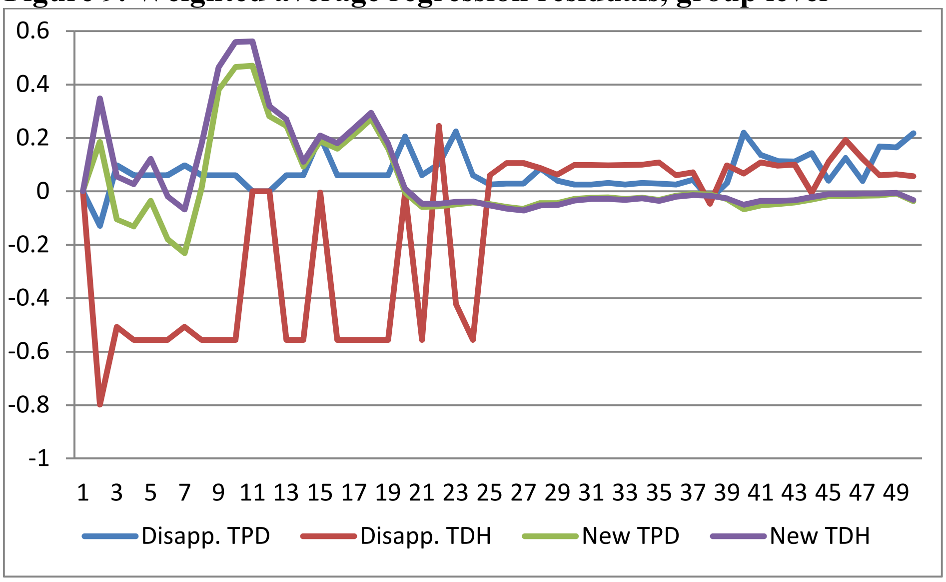
Table 2. Regression results for TDH model, group level (excluding time dummies)

Attribute	Dummy	Estimate	Std. error	t value	Signif.
Intercept		1.906238	0.029731	64.12	***
Shape of neck	V	0.008876	0.009980	0.89	
Fabric	Organic	-0.231588	0.020739	-11.17	***
Sleeve length	Long	0.210602	0.015455	13.04	***
# T-shirts per package	2	0.529237	0.012863	41.15	***
	3	0.678165	0.017848	38.00	***
Color	White	-0.038794	0.010799	-3.59	***
	Black	-0.024664	0.015900	-1.55	
Fit	Stretch	0.012907	0.020998	0.62	

Significance codes: *** 0.001; ** 0.01; * 0.05
observations: 1,289
R-squared: 0.8664; Adjusted R-squared: 0.8602

The weighted average regression residuals from the TPD and TDH regressions at the group level in Figure 9 exhibit a similar pattern as those at the EAN level. There are a few noticeable differences though. The average TPD residuals for the new items do not differ much any longer from the average TDH residuals. Also, the average TPD residuals for the disappearing are now mostly positive.

Figure 9. Weighted average regression residuals, group level



As can be concluded from Figure 6, the ratio of the period *t* expenditure shares for the new and matched items after month 13 is huge when items are defined by EAN, up to more than 2400 in month 40. This is very different for the group-based items. The patterns of the group-based aggregate expenditure shares shown in Figure 10 are similar to those in Figure 6, but much less pronounced. For example, the share of new items never gets above 0.77, which implies that the ratio of the period *t* expenditure shares for new and matched items never exceeds 3.3. Put differently, the degree of leveraging is much less at the group level than at the EAN level.

Figure 10. Aggregate expenditure shares, group level

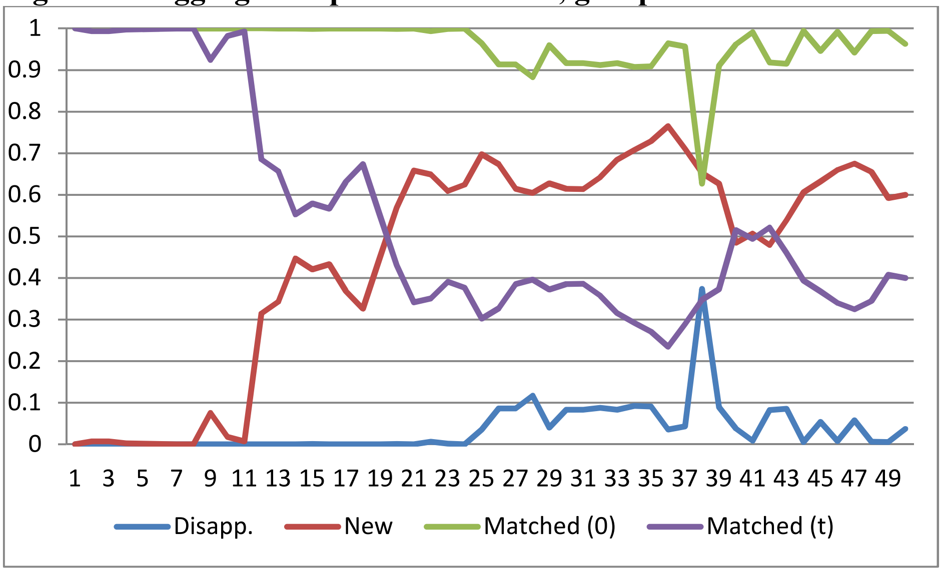
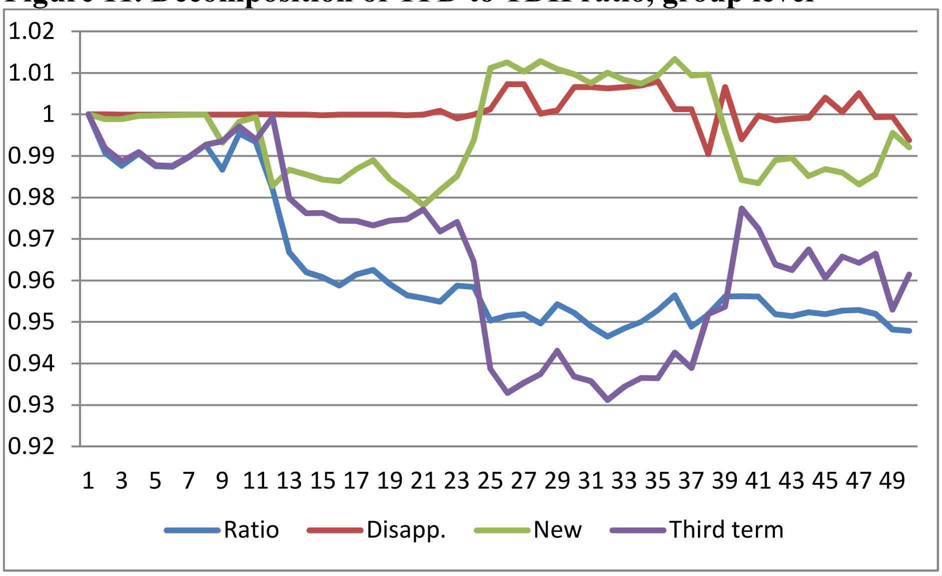


Figure 11. Decomposition of TPD to TDH ratio, group level



The results of decomposition (15) at the group level are plotted in Figure 11. The third term, which depends on the change in the matched items' normalized expenditure shares, now contributes most. This is simply because at the group level there are few new and disappearing items; if all items were matched, this term would be exactly equal to the TPD to TDH index ratio.

4. Discussion and conclusions

The TPD model is essentially a pooled regression model with fixed-effects for all items sold during the sample period and with no time-varying variables other than dummies for time. In general, the inclusion of fixed effects in a regression model estimated on a panel data set controls for unobservable characteristics. The scanner data sets used for price measurement are not fixed panels though; there is usually substantial turnover of items, i.e. the data is characterized by many entries (new items) and exits (disappearing items). If the item universe was static rather than dynamic, a matched-model approach would suffice and modeling was not required.

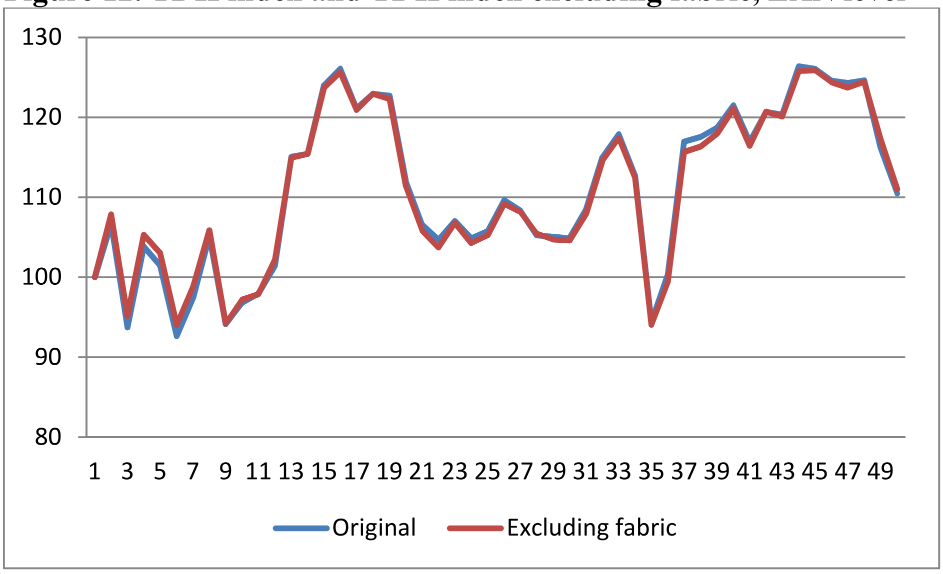
In this paper, we have discussed some of the issues that arise when applying the TPD method in a dynamic-universe context in order to construct quality-adjusted price indexes. Our main point is that the TPD model likely suffers from overfitting because it includes all first- and higher order interactions as compared with the "true" TDH model that is restricted to main effects and possibly some first-order interactions. We derived a decomposition explaining the ratio of the TPD to TDH index in terms of the weighted average regression residuals for unmatched new and disappearing items and applied the decomposition to scanner data for t-shirts.

Our results pointed to downward bias in the TPD index for t-shirts, especially if items are identified by EAN (barcode). Clothing is well known for its lack of matching at the EAN level due to seasonality. For example, summer clothes disappear in autumn and then re-appear in spring but often with different EANs. These (strongly) seasonal items are re-launched, and potentially exhibit disguised price changes. Seasonality was not the major cause of the downward bias of the TPD index in our example, however. While some (weak) seasonal effects in sales occur as more short-sleeve t-shirts are sold in spring and summer then in autumn and winter, the problem was rather a sudden sharp decline in the number of matches in the data as a result of the introduction of "organic" t-shirts which immediately took up a very large share of expenditure at the expense of "basic" t-shirts.

There are a number of practical issues we have not addressed so far. One issue is the treatment of *revisions*. The problem here is that when the sample period is extended, new data is added and the regression models are re-estimated, previous TPD and TDH estimates will change. Statistical agencies do not accept such revisions. In the Appendix we discuss some of the solutions that have been put forward in the literature to deal with revisions, including two moving-window approaches, and apply them to the EAN-based TDH index for t-shirts.

There are also issues regarding the specification of our hedonic model, such as omitted variables. For example, an indicator for quality of fabric other than "basic" or "organic" was not included, because we are constrained by the relatively broad product descriptions given in the scanner data sets.[7] Omitting variables in hedonic regressions can lead to bias in the resulting price indexes, but the bias does not necessarily have to be large. Suppose we would omit "fabric" from the time dummy hedonic model. This is an interesting case because of the big impact the introduction of "organic" t-shirts had on the difference between the TDH and TPD indexes, particularly at the EAN level. The coefficient for "organic" in the original EAN-based TDH regression was negative and highly significant. Figure 12 shows what happens to the TDH index when this variable is omitted: almost nothing.

Figure 12. TDH index and TDH index excluding fabric, EAN level



[7] "Brand" often serves as a proxy for unobserved quality characteristics in empirical hedonic models, but this particular retailer only sells t-shirts under a house brand.

A comparison of the new regression results in Table 3 with the old ones in Table 1 reveals that the downward effect of "organic" is now largely being picked up by the dummy for "stretch" due to a high correlation between these variables. This example also reminds us that multicollinearity is not a big issue when estimating TDH indexes, where we are interested in the predicted prices rather than the estimated characteristics parameters.

Table 3. Regression results for TDH model without fabric, EAN level (excluding time dummy variables)

Attribute	Dummy	Estimate	Std. error	t value	Signif.
Intercept		1.929611	0.009630	200.38	***
Shape of neck	V	0.015069	0.003228	4.67	***
Sleeve length	Long	0.312975	0.004080	76.71	***
# T-shirts per package	2	0.496930	0.004166	119.28	***
	3	0.664429	0.005741	115.74	***
Color	White	-0.056324	0.003494	-16.12	***
	Black	-0.066130	0.005083	-7.021	***
Fit	Stretch	-0.202044	0.004515	-44.57	***

Significance codes: *** 0.001; ** 0.01; * 0.05
observations: 24,797
R-squared: 0.7480; Adjusted R-squared: 0.7474

The multilateral TDH method is not the only way to estimate transitive quality-adjusted price indexes. De Haan and Krsinich (2014a) proposed a (rolling-year) GEKS approach called ITRYGEKS, where the "missing prices" of the new and disappearing items in the bilateral comparisons – in their case measured by bilateral Törnqvist price indexes – are imputed using bilateral TDH regressions.[8] A strong point of (ITRY)GEKS is its reliance on a superlative index number formula; it is grounded in standard index number theory. A practical disadvantage of the ITRYGEKS approach is its complexity, in particular because many models have to be estimated each month.

The above is not to say that weighted TDH has no theoretical underpinning. As shown by equation (10) in section 2.2, the weighted TDH index (and the weighted TPD index) can be written as an imputation Törnqvist price index times a factor that induces transitivity. Moreover, De Haan and Krsinich (2014b) shave hown that the expenditure-share weighted TDH index will be an accurate approximation of a *quality-adjusted unit value index*. The latter is a modified unit value index in which the observed prices and quantities have been replaced by quality-adjusted (standardized) prices and quantities.

[8] They used De Haan's (2004) result mentioned in footnote 3. Statistics New Zealand implemented the ITRYGEKS method in the CPI for many consumer electronics products using scanner data from market research company GfK (Statistics New Zealand, 2014).

The quality-adjusted unit value approach is appealing for products existing of broadly comparable items. Irrespective of the interpretation, the TDH model should be restricted to broadly comparable items, i.e. at a detailed elementary level of aggregation, because different products typically have different sets of characteristics or different parameters for the same characteristics.

To summarize: the use of the TPD method can lead to biased results when there is insufficient matching due to re-launches of "the same" items with different barcodes or a sudden introduction of new items that account for a large share of expenditure. The first problem is basically a data problem: the barcode/EAN may be too detailed a level to compare like with like. Data permitting, we could identify items by cross classifying the categorical attributes and apply the TPD method. This does not necessarily resolve the second problem since the TPD method wrongly treats all interaction terms as quality characteristics.[9] The TDH method would be the preferred choice. Moreover, if sufficient information on characteristics is available, there is no need to form "groups": the TDH method can be directly applied to scanner data at the EAN level.

Appendix: The treatment of revisions

One way to deal with revisions is to use a *rolling-window approach*.[10] Rolling-window approaches shift the estimation window of, say, 13 months, forward each month and then splice the new indexes onto the existing time series. De Haan (2015b) compared two rolling-year versions: standard (movement) splicing and Krsinich's (2016) window splicing. The standard method splices the most recent month-on-month index movement onto the existing time series, while the other method splices the entire newly estimated 13-month series onto the index number of 12 months ago.

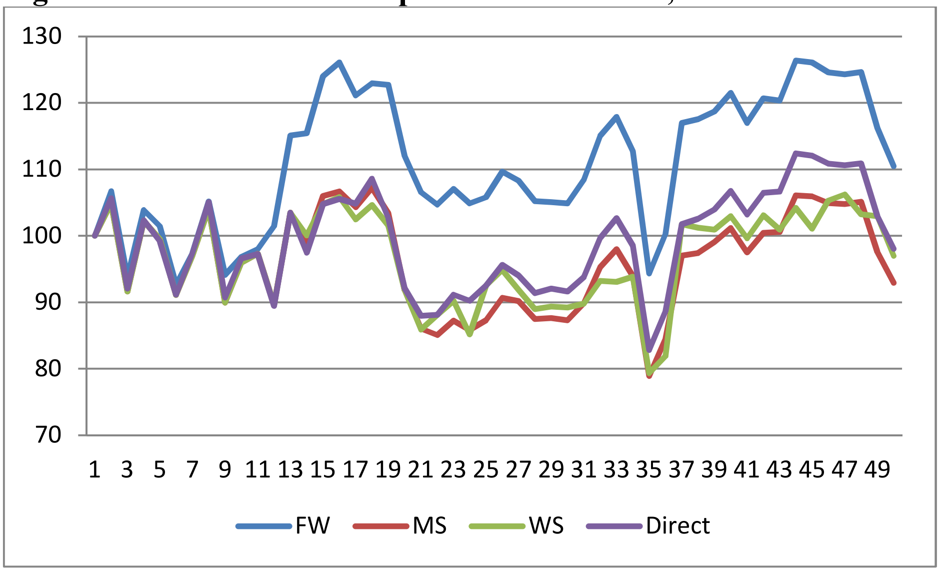
---

[9] If information on quality characteristics is lacking, the TPD model could perhaps be improved by using a life cycle function, providing that stable product life cycles exist and can be accurately estimated from the data. The extended model would control for time, unobserved characteristics, and "age". Melser and Syed (2015) applied such a model to a large U.S. scanner data set. Bils (2009) and Abe et al. (2016) also estimated life cycle functions, albeit not in a TPD framework. Given the complexity of estimating life-cycle functions, it is doubtful whether this approach is fit for real-time CPI production. Also, a life-cycle approach is unlikely to resolve the problem of re-launches and disguised price changes because both their timing and magnitude are rather unpredictable.

[10] Ivancic, Diewert and Fox (2009) applied a rolling-year approach to TPD indexes and GEKS indexes. In the published version of their paper (Ivancic, Diewert and Fox, 2011), however, they did not present the rolling-year TPD indexes.

In Figure A.1, the EAN-based TDH index from Figure 2, estimated on the full window (FW) of 50 months, is copied and compared to the rolling-year indexes with a movement splice (MS) and a window splice (WS). The choice of splicing method does not matter much, which is reassuring.[11] However, as from month 13 when "organic" t-shirts were introduced, a disturbing gap between the full-window index and the rolling-year indexes arises. Assuming the full-window index is the appropriate benchmark, the two rolling-year splicing methods seem to cause downward drift.

Figure A.1. TDH index and spliced TDH indexes, EAN level



The issue of potential drift in rolling-window indexes has led researchers to look for other ways to extend the time series without revising previously published indexes. Chessa (2016) constructed short-term index series, starting in December and ending in December of the next year, and chain linked the short-term series in December of each year.[12] Figure A.1 also depicts the results for this "direct" extension method, but with February instead of December as the link month. The index is only slightly above the

---

[11] The evidence on movement versus window splicing is not conclusive. Using New Zealand scanner data on consumer electronics products, De Haan (2015b) found some significant differences for TPD indexes. Melser (2015), using a large U.S. supermarket scanner data set, did not find any systematic differences, but he did observe slightly higher variability of the price index series with a movement splice.

[12] Chessa (2016) not only used an alternative extension method but also used another multilateral method to construct price indexes, the Geary (1958) – Khamis (1972) method. For reasons of symmetry, De Haan (2015a) proposed a rolling-window approach with a so-called half window splice. Here, the length of the estimation window is 25 instead of 13 months, with the middle month serving as link month.

rolling-window indexes, in particular during the last year, and so this method does not seem to alleviate the downward drift much.

Because the length of the window to estimate the short-term indexes in Chessa's method increases over time, from 2 up to 13 months, the models are initially estimated on sparse data. We therefore expected the indexes initially to be volatile, but there is no sign of this. A potential disadvantage is that the link month is given special importance, which conflicts with the idea behind multilateral methods of making results independent of the choice of base or link period.

References

Abe, N., Y. Ito, K. Munakata, S. Ohyama and K. Shinozaki (2016), "Pricing Patterns over Product Life-Cycle and Quality Growth at Product Turnover: Empirical Evidence from Japan", Paper presented at the Meeting of the Group of Experts on Consumer Price Indices, 2-4 May, Geneva, Switzerland.

Aizcorbe, A., C. Corrado and M. Doms (2003), "When Do Matched-Model and Hedonic Techniques Yield Similar Price Measures?", Working paper 2003-14, Federal Reserve Bank of San Francisco, San Francisco.

Aizcorbe, A. and Y. Pho (2005), "Differences in Hedonic and Matched-Model Price Indexes: Do the Weights Matter?", Working paper 2005-06, Bureau of Economic analysis, Washington DC.

Bils, M. (2009), "Do Higher Prices for New Goods Reflect Quality Growth or Inflation?", *Quarterly Journal of Economics* 124, 637-675.

de Boer, G. (2016), "Automated Classification for Consumer Price Indexes", Research paper, Statistics Netherlands, The Hague, The Netherlands.

Chessa, A.G. (2015), "Towards a Generic Price Index Method for Scanner Data in the Dutch CPI", Paper for the fourteenth Ottawa Group Meeting, 20-22 May 2015, Tokyo, Japan.

Chessa, A.G. (2016), "Processing Scanner Data in the Dutch CPI: A New Methodology and First Experiences", Paper presented at the Meeting of the Group of Experts on Consumer Price Indices, 2-4 May, Geneva, Switzerland.

Diewert, W.E. (2005), "Weighted Country Product Dummy Variable Regressions and Index Number Formulae", *Review of Income and Wealth* 51, 561-570.

Diewert, W.E., S. Heravi and M. Silver (2009), "Hedonic Imputation versus Time Dummy Hedonic Indexes", pp. 87-116 in W.E. Diewert, J. Greenlees and C. Hulten (eds.), *Price Index Concepts and Measurement*, Studies in Income and Wealth, Vol. 70. Chicago: University of Chicago Press.

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.

de Haan, J. (2002), "Generalised Fisher Price Indexes and the Use of Scanner Data in the Consumer Price Index", *Journal of Official Statistics* 1, 61-85.

de Haan, J. (2004b), "The Time Dummy Index as a Special Case of the Imputation Törnqvist Index", Paper presented at the eighth meeting of the Ottawa Group, 23-25 August 2004, Helsinki, Finland.

de Haan, J. (2010), "Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and 'Re-pricing' Methods", *Jahrbücher fur Nationalökonomie und Statistik* 230, 772-791.

de Haan, J. (2015a), "A Framework for Large Scale Use of Scanner Data in the Dutch CPI", Paper presented at the fourteenth Ottawa Group Meeting, 20-22 May 2015, Tokyo, Japan.

de Haan, J. (2015b), "Rolling Year Time Dummy Indexes and the Choice of Splicing Method", Research paper, Statistics Netherlands, The Hague, The Netherlands.

de Haan, J. and H.A. van der Grient (2011), "Eliminating Chain Drift in Price Indexes Based on Scanner Data", *Journal of Econometrics* 161, 36-46.

de Haan, J. and R. Hendriks (2013), "Online Data, Fixed Effects and the Construction of High-Frequency Price Indexes", Paper presented at the Economic Measurement Group Workshop, 28-29 November 2013, Sydney, Australia.

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

de Haan, J. and F. Krsinich (2014b), "Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ?", Paper presented at the Society for Economic Measurement Conference, 18-20 August 2014, Chicago, U.S.

de Haan, J., L. Willenborg and A.G. Chessa (2016), "An Overview of Price Index Methods for Scanner Data", Paper for the UNECE-ILO Meeting of the Group of Experts on Consumer Price Indices, 2-4 May 2016, Geneva, Switzerland.

Howard, A., K. Dunford, J. Jones, M. van Kints and K. Naylor (2015), "Using Transactions Data to Enhance the Australian CPI", Paper presented at the fourteenth Ottawa Group Meeting, 20-22 May 2015, Tokyo, Japan.

Ivancic, L., W.E. Diewert and K.J. Fox (2009), "Scanner Data, Time Aggregation and the Construction of Price Indexes", Discussion Paper no. 09-09, Department of Economics, University of British Columbia, Vancouver, Canada.

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.

ILO/IMF/OECD/UNECE/Eurostat/The World Bank (2004), *Consumer Price Index Manual: Theory and Practice*. ILO Publications, Geneva.

Kennedy, P.E. (1981), "Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations", *American Economic Review* 71, 801.

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.

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

Melser, D. (2015), “Scanner Data Price Indexes: Addressing Some Unresolved Issues”, Unpublished paper, School of Economics, Finance and Marketing, RMIT University, Melbourne, Australia.

Melser, D. and I.A. Syed (2015), “Life Cycle Price Trends and Product Replacement: Implications for the Measurement of Inflation”, *Review of Income and Wealth*, DOI: 10.1111/roiw.12166.

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

Silver, M. and S. Heravi (2005), “A Failure in the Measurement of Inflation: Results from a Hedonic and Matched Experiment Using Scanner Data”, *Journal of Business & Economic Statistics* 23, 269-281.

Statistics New Zealand (2014), “Measuring Price Change for Consumer Electronics Using Scanner Data”. Available from www.stats.govt.nz.

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