THE DIFFERENCE BETWEEN HEDONIC IMPUTATION INDEXES AND TIME DUMMY HEDONIC INDEXES FOR DESKTOP PCs

Draft version 25 May 2004

Mick Silver* and Saeed Heravi*

* Cardiff University
Colum Drive
Cardiff CF10 3EU
UK
Email: Silver@cardiff.ac.uk

Acknowledgements are due to the U.K Office for National Statistics (ONS) for supplying the data for this study. This follows a well established research link with the ONS which has been of much benefit to the authors. Naturally the ONS are not responsible for any views, errors of omission or commission in this paper. We also acknowledge useful comments from Erwin Diewert (University of British Columbia) and other participants at a Centre for Applied Economic Research, Economic Measurement Group Workshop, UNSW, Sydney, December 2003.

ABSTRACT

This paper shows why the two main approaches to estimating hedonic indexes can produce quite different results. The proper measurement of inflation requires the use of hedonic indexes, rather than matched models, for product areas which have a high turnover of differentiated models. The two main approaches to hedonic indexes are *hedonic imputation* (HI) *indexes* and *dummy time hedonic* (DTH) indexes. HI indexes value a fixed period's basket of characteristics using both base period and current period hedonic coefficients and take the ratio of the latter to the former. HI index number formulas differ in their use of which period's characteristics are held constant for the valuation. DTH indexes estimate price change using the coefficient on a dummy variable for time in a hedonic regression which uses both base *and* current period's data. For DTH indexes the slope parameters are constrained to be the same for both periods to allow the intercept shift to measure quality-adjusted price change. For HI indexes the change in the parameters over time are, paradoxically, the essence of the measure.

The study provides a formal analysis of the difference between the two approaches. It shows the conditions under which the approaches will provide similar results which, surprisingly, may even be when parameters are unstable, and the factors governing differences between the results. It shows that differences between the methods can be substantial and discusses why this is the case and the issue of choice between these measures. An illustrative study for desktop PCs is provided. It demonstrates the importance of using hedonic indexes for product areas which have a high turnover of differentiated models, that the two approaches can differ and the factors underlying the difference.

KEYWORDS: HEDONIC REGRESSIONS; HEDONIC INDEXES; TIME DUMMY VARIABLE INDEX; CONSUMER PRICE INDEXES; PERSONAL COMPUTERS

JEL classification: C43, C82, E31.

INTRODUCTION

The purpose of this paper is to outline and compare the two[1] main and quite distinct approaches to the measurement of hedonic price indexes: hedonic dummy time variable indexes and hedonic imputation indexes. Both approaches not only correct price changes for changes in the quality of items purchased, but also allow the indexes to incorporate matched and unmatched models. However, they can yield quite different results and it is not immediately apparent which approach is preferable. The standard way price changes are measured by most national governments is through the use of the matched models method. In this method the details and prices of a representative selection of items are collected in a base reference period and their matched prices collected in successive periods so that the prices of 'like' are compared with 'like'. However, if there is a rapid turnover in available models, then the sample of product prices used to measure price changes becomes unrepresentative of the category as a whole as a result of both new unmatched models being introduced (but not included in the sample), and older unmatched models being retired (and thus dropping out of the sample). Hedonic indexes use matched and unmatched models and in doing so put an end to the matched models sample selection bias (Cole, *et al.* (1986), Silver and Heravi (2002) and Pakes (2003)).

Our interest in hedonic indexes arose from the need to reduce bias in the measurement of the U.S. consumer price index (CPI), which has been the subject of three major reports, the Stigler Committee (1961), Boskin Commission (1996) and the report by the Committee on National Statistics (2002): the Schultze panel. Each found the inability to properly remove the effect on price changes of changes in quality to be a major source of bias. Hedonic regressions were considered to be the most promising approach to control for such quality changes, though the Schultze panel cautioned for the need for further research on methodology.[2]

At first sight the two approaches to hedonic indexes appear quite similar and there is little to choose between them. Both rely on hedonic regression equations to remove the effects on price of quality changes. They can also incorporate a range of weighting systems, take specific functional forms for the aggregation (e.g. geometric and arithmetic) and can be formulated as chained or direct, fixed base comparisons. Yet they can give quite different results, even when using comparable weights, functional forms and the same periodic comparison. This is because they work on different principles. The dummy variable method constrains hedonic regression parameters to be the same over time while a hedonic imputation index paradoxically relies on parameter change as the essence of the measure. There has been some valuable research on the two approaches (see Berndt and Rappaport (2001), Diewert (2002b), Silver and Heravi (2003) and Pakes

---

[1] An alternative approach to hedonic indexes is to use a fixed effects panel estimator in which instead of hedonic characteristics on the right-hand-side of the regression, each model, other than a benchmark one, has a dummy variable (suggested by Diewert in (2002a)—see Aizcorbe (2003) for an application). In practice the panel estimation procedure is much simpler; the price deviations for each model from their mean over time are regressed on the deviations of the explanatory dummy variables for each model for all t with an adjustment for degrees of freedom (Davidson and MacKinnon, 1993: 323). For a direct fixed base comparison the estimate reduces to a matched models one.

[2] "Hedonic techniques currently offer the most promising approach for explicitly adjusting observed prices to account for changing product quality. But our analysis suggests that there are still substantial unresolved econometric, data, and other measurement issues that need further attention." (Committee on National Statistics, 2002: 6).

(2003)), though no formal analysis, to the author's knowledge, of the factors governing the differences between the approaches. Berndt and Rappaport (2001) and Pakes (2003) have highlighted the fact that the two approaches can give different results and both advise the use of HI indexes when parameters are unstable, though our analysis casts more light on this proposal.

This paper first, in section I, examines the alternative formulations of the two methods and then, in section II develops an expression for their differences. Section III provides an empirical study for desktop PCs and section IV discusses the issue of choice between the approaches in light of the findings in sections II and II; section V concludes.

I. HEDONIC INDEXES

A hedonic regression equation of the prices of i=1,...,N models of a product, p_i, on their quality characteristics z_{ki} , where z_k =1,….,K price-determining characteristics, is given in a linear form by:

$$p_i = \gamma_0 + \sum_{k=1}^{K} \beta_k z_{ki} + \varepsilon_i \tag{1}$$

The β_k are estimates of the marginal valuations the data ascribes to each characteristic (Rosen (1974), Griliches (1988) and Triplett (1988), see also Diewert (2003) and Pakes (2003)). Statistical offices use hedonic regressions for CPI measurement when a model is no longer sold and a price adjustment for quality differences is needed in order that the price of the original model can be compared with that of a non-comparable replacement model. Silver and Heravi (2001) refer to this as 'patching'. However, patching can only make use of data outside of the matched sample when a model is missing. It may be that several new models are introduced in a month when there are few, if any, models needing replacements. The likely atypical price changes of the new models will be ignored with patching, but not with hedonic indexes. The needs of quality adjustment, in dynamic markets, such as PCs, is to resample each month to cover a representative sample of what is purchased and *hedonic indexes* provide the required measures.

A. Hedonic imputation (HI) indexes

Hedonic imputation (hereafter—HI) indexes take a number of forms: first, as either equally-weighted or weighted indexes; second, depending on the functional form of the aggregator, say a geometric aggregator as against an arithmetic one; third, with regard to which period's characteristic set is held constant and finally, as direct comparisons between periods 0 and t, or as chained indexes, with individual links being calculated between periods 0 and 1, 1 and 2,….,t-1 and t, the results being combined by successive multiplication.

Consider the linear hedonic function $\hat{p}_i^0 = g^0\left(z_i^0\right)$ of (1) and a semi-logarithmic form of (1) $\hat{p}_i^0 = h^0\left(z_i^0\right)$ both estimated in period 0 with a vector of K quality characteristics $z_i^0 = z_{i1}^0,......,z_{iK}^0$ and N^0 observations and similarly for period t. Let quantities sold in period 0 be q_i^0 and relative sales value shares, $s_i^0 = p_i^0 q_i^0 / \sum_i p_i^0 q_i^0$ and again similarly for period t.

A hedonic Laspeyres index for *matched and unmatched period 0* models is given by:

$$P_{H-Las}=\frac{\sum_{i=1}^{N^0} g^t\left(z_i^0\right)q_i^0}{\sum_{i=1}^{N^0} g^0\left(z_i^0\right)q_i^0} \tag{2}$$

and a hedonic Paasche index for *matched and unmatched period t* models by:

$$P_{H-Pas}=\frac{\sum_{i=1}^{N^t} g^t\left(z_i^t\right)q_i^t}{\sum_{i=1}^{N^t} g^0\left(z_i^t\right)q_i^t}\approx\frac{\sum_{i=1}^{N^t} p^t q_i^t}{\sum_{i=1}^{N^t} g^0\left(z_i^t\right)q_i^t} \tag{3}$$

It is apparent from equations (2) and the first expression in (3) that a hedonic Laspeyres index holds characteristics constant in the *base* period and while a hedonic Paasche index holds the characteristics constant in the *current* period. Thus the differences between the hedonic valuations in Laspeyres and Paasche are dictated by the extent to which the characteristics change over time; i.e. $\left(z_i^0-z_i^t\right)$. The further the z_i values are apart, say due to greater technological change, the less justifiable is the use of an individual estimate and the less faith there is in a compromise geometric mean of the two indexes—a Fisher index.[3] Since neither hedonic Laspeyres and nor Paasche indexes can be considered superior, we focus in the analytical work on superlative indexes—the Fisher, Törnqvist and Walsh formulas—which make *symmetric* use of information in *both* periods, are supported by economic theory and axiomatic considerations and are preferred target indexes in the forthcoming international CPI Manual (Diewert, 2004: chapters 15-18).[4]
For our consideration of matched *and* unmatched new and old models in the indexes we consider four exhaustive sets of models: $i\in S^t\left(0\cap t\right)$ matched models in period *t*; $i\in S^0\left(0\cap t\right)$ matched models in period 0; $i\in S^t\left(t\neg 0\right)$ unmatched new models present in period *t*, but not in period 0; $i\in S^0\left(0\neg t\right)$ unmatched old models present in period 0, but not in period *t*. Let the number of models in these respective sets be denoted by $N^t\left(0\cap t\right), N^0\left(0\cap t\right), N^0\left(0\neg t\right)$ and $N^t\left(t\neg 0\right)$. We also denote the set of matched models with common characteristics $z_i^m=z_i^0=z_i^t$ in both periods as $i\in S\left(0\cap t\right)$ enumerated over $i=1,...,N^t\left(0\cap t\right)=N^0\left(0\cap t\right)=N^m$. A **generalized[5] hedonic Fisher index** (a geometric mean of (2) and (3)) is given by:

$$P_{H-Fisher}=\left[\left(\frac{\sum_{i\in S^0(0\cap t)}^{N^0(0\cap t)} g^t\left(z_i^m\right)q_i^0+\sum_{i\in S^0(0\neg t)}^{N^0(0\neg t)} g^t\left(z_i^0\right)q_i^0}{\sum_{i\in S^0(0\cap t)}^{N^0(0\cap t)} g^0\left(z_i^m\right)q_i^0+\sum_{i\in S^0(0\neg t)}^{N^0(0\neg t)} g^0\left(z_i^0\right)q_i^0}\right)\times\left(\frac{\sum_{i\in S^t(0\cap t)}^{N^t(0\cap t)} g^t\left(z_i^m\right)q_i^t+\sum_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} g^t\left(z_i^t\right)q_i^t}{\sum_{i\in S^t(0\cap t)}^{N^t(0\cap t)} g^0\left(z_i^m\right)q_i^t+\sum_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} g^0\left(z_i^t\right)q_i^t}\right)\right]^{\frac{1}{2}} \tag{4}$$

[3] As an estimate of a COLI index the spread is irrelevant since the need is to include substitution effects and Fisher meets this need. However, Laspeyres and Paasche answer meaningful question and act as bounds on models of economic behaviour that different consumer might pursue. A Fisher estimate with less 'spread' is more satisfactory
[4] Pakes (2003) advocates a Laspeyres hedonic as an upper bound to a target compensating index but notes that a Paasche hedonic index may be more feasible in real time, again raising issues as to the extent of the difference.

A generalized hedonic Törnqvist index is given by:

$$P_{H-Törnqvist} = \left[\frac{\prod_{i\in S^t(0\cap t)}^{N^m} h^t\left(z_i^m\right)}{\prod_{i\in S^0(0\cap t)}^{N^m} h^0\left(z_i^m\right)}\right]^{\tilde{s}_i} \times \left[\frac{\prod_{i\in S^0(0\neg t)}^{N^0(0\neg t)} h^t\left(z_i^0\right)}{\prod_{i\in S^0(0\neg t)}^{N^0(0\neg t)} h^0\left(z_i^0\right)}\right]^{\frac{s_i^0}{2}} \times \left[\frac{\prod_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} h^t\left(z_i^t\right)}{\prod_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} h^0\left(z_i^t\right)}\right]^{\frac{s_i^t}{2}} \tag{5}$$

where $\tilde{s}_i = (s_i^0 + s_i^t)/2$.

A **generalized hedonic Walsh index** is given by:

$$P_{HBC-Walsh} = \frac{\sum_{i\in S(0\cap t)}^{N^m} g^t\left(\sqrt{z_i^0 z_i^t}\right)\sqrt{q_i^0 q_i^t} + \sum_{i\in S^0(0\neg t)}^{N^0(0\neg t)} g^t\left(z_i^0\right)\sqrt{q_i^0} + \sum_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} g^t\left(z_i^t\right)\sqrt{q_i^t}}{\sum_{i\in S(0\cap t)}^{N^m} g^0\left(\sqrt{z_i^0 z_i^t}\right)\sqrt{q_i^0 q_i^t} + \sum_{i\in S^0(0\neg t)}^{N^0(0\neg t)} g^0\left(z_i^0\right)\sqrt{q_i^0} + \sum_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} g^t\left(z_i^t\right)\sqrt{q_i^t}}\,. \tag{6}$$

where $\sqrt{z_i^0 z_i^t} = z_i^m$ for the summation over N^m matched models.

B. Dummy time hedonic (DTH) indexes

Dummy time hedonic (hereafter—DTH) indexes are a second approach which, as with HI indexes, do not require a matched sample.[6] The formulation is similar to equation (1) except that a single regression is estimated on the data in the two time periods compared, $i \in S(0 \cup t)$. The equation also includes a dummy variable D^t which is equal to 1 in period t, and zero otherwise:

$$\ln p_i^{0,t} = \delta_0 + \delta_1 D^{0,t} + \sum_{k=1}^{K} \beta_k z_{ki}^{0,t} + \varepsilon_i^{0,t} \tag{7}$$

The exponent of the estimated coefficient δ_1^* is an estimate of the quality-adjusted price change between period 0 and period t. A weighted version of (7) would use a weighted least squares (WLS) estimator and weights $\tilde{s}_i^0 = (s_i^0 + s_i^t)/2$ for matched models and s_i^0 or s_i^t for unmatched old and new models respectively (Diewert, 2002b).

The regression equation (7) constrains each of the β_k coefficients to be the same across the two months compared. In restricting the slopes to be the same, the (log of the) price change between periods 0 and t can be measured at any value of z, as illustrated by the difference between the dashed lines in Figure 1. For convenience it is evaluated at the origin as δ_1^*. Bear in mind the HI indexes outlined above estimate the differences between price surfaces with different slopes. As such the estimates have to be conditioned on particular values of z, which gives rise to the two estimates considered in (2) and (3): the base HI using z^0 and the current period HI using z^t, as shown in Figure 1. The very core of the DTH method is to constrain the coefficients to be the same, so there is no need to condition on particular values of z. The DTH estimates implicitly and usefully make symmetric use of base and current period data. As with hedonic imputation

[5] Diewert (2002b) referred to indexes that incorporated all unmatched and matched data as generalized forms.
[6] See De Haan (2003) for a variant that uses matched data when available and the time dummy only for unmatched data—his double imputation method.

indexes DTH indexes can take fixed and chained base forms, though they can also take a fully constrained form whereby a single constrained regression is estimated for say January to December with dummy variables for each month, though this is impractical in real time since it requires data on future observations.

II. WHY HEDONIC IMPUTATION AND DUMMY TIME HEDONIC INDEXES DIFFER

A. Algebraic differences

(i) a reformulation of the hedonic indexes

There has been little analytical work undertaken on the factors governing differences between the two approaches. To compare the HI approach to the DTH approach we first need to reformulate the HI indexes. We note that the HI approach relies on two estimated hedonic equations, $h^t\left(z_i^t\right)$ and $h^0\left(z_i^0\right)$ for periods 0 and *t* respectively:

$$ln\ p_i^t = \gamma_0^t + \sum_{k=1}^{K} \beta_k^t z_{ki}^t + \varepsilon_i^t \tag{8}$$

$$\ln\ p_i^0 = \gamma_0^0 + \sum_{k=1}^{K} \beta_k^0 z_{ki}^0 + \varepsilon_i^0 \tag{9}$$

We assume that the errors in each equation are similarly distributed, then phrase the two equations as a single hedonic regression equation with dummy time intercept and slope variables:

$$ln\ p_i^{0,t} = \gamma_0^0 + \gamma_1 D_0^{0,t} + \sum_{k=1}^{K} \beta_k^0 z_{ki}^{0,t} + \sum_{k=1}^{K} \beta_k D_{ki}^{0,t} + \varepsilon_i^{0,t} \tag{10}$$

where $D_0^{0,t} = 1$ if observations are in period *t* and 0 otherwise, $\gamma_1 = (\gamma_0^t - \gamma_0^0)$, $D_{ki}^{0,t} = z_{ki}^t$ if observations are in period *t* and 0 otherwise, and $\beta_k = (\beta_k^t - \beta_k^0)$. The exponent of the estimated γ_1^* is an estimate of the change in the *intercepts* of the two hedonic price equations and is thus a HI index evaluated at a particular value of $z_{ki}^{0,t}$, i.e., $\tilde{z}_k^{0,t} = 0$. A HI index evaluated at $\tilde{z}_k^{0,t} = 0$ has no economic meaning. We require a mean-value HI index evaluated at $\tilde{z}_k^{0,t} = \bar{z}_k^{0,t}$. The Walsh and Törnqvist indexes can be evaluated at mean values and, being superlative, should closely approximate each other and the Fisher index, all of which are recommended as target indexes for consumer price indexes (Diewert, 2004, chapters 15 and 17). The mean values that correspond to the generalized Törnqvist and Walsh HI indexes of (5) and (6) are respectively:[7]

$$\tilde{z}_k^{0,t} = \bar{z}_{k-Törn}^{0,t} = \prod_{i \in S(0 \cap t)}^{N^m} \left(z_i^m\right)^{\tilde{s}_i} \times \prod_{i \in S^0(0 \neg t)}^{N^0(0 \neg t)} \left(z_i^0\right)^{\frac{s_i^0}{2}} \times \prod_{i \in S^t(t \neg 0)}^{N^t(t \neg 0)} \left(z_i^t\right)^{\frac{s_i^t}{2}} \tag{11}$$

[7] Mean value weights that apply to price *changes* differ from those that apply to prices in the Walsh index. The former are $\sqrt{q_i^0 q_i^t}\, p_i^0 \Big/ \Sigma_i \sqrt{q_i^0 q_i^t}\, p_i^0$ for matched models, as opposed to $\sqrt{q_i^0 q_i^t} \Big/ \Sigma_i \sqrt{q_i^0 q_i^t}$, but it is the latter that concern us.

$$\tilde{z}_k^{0,t} = \bar{z}_{k-Walsh}^{0,t} = \frac{\sum_{i\in S(0\cap t)}^{N^m} \left(z_i^m\right)\sqrt{q_i^0 q_i^t} + \sum_{i\in S^0(0\neg t)}^{N^0(0\neg t)} \left(z_i^0\right) q_i^0 + \sum_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} \left(z_i^t\right) q_i^t}{\sum_{i\in S(0\cap t)}^{N^m} \sqrt{q_i^0 q_i^t} + \sum_{i\in S^0(0\neg t)}^{N^0(0\neg t)} q_i^0 + \sum_{i\in S^t(t\neg 0)}^{N^t(t\neg 0)} q_i^t} \tag{12}$$

We thus do not evaluate γ_1 at $\tilde{z}_k^{0,t} = 0$, but for Törnqvist and Walsh HI indexes at $\tilde{z}_k^{0,t} = \bar{z}_{k-Törn}^{0,t}$ and $\tilde{z}_k^{0,t} = \bar{z}_{k-Walsh}^{0,t}$ respectively. The required **Törnqvist HI index is thus estimated** (hereafter all references to indexes are after taking the exponents) as:

$$\gamma_1^* + \sum_{k=1}^{K} \beta_k^* \bar{z}_{k-Törn}^{0,t} \qquad \text{where } \beta_k^* \text{ is a WLS}^8 \text{ estimate of } (\beta_k^t - \beta_k^0) \tag{13}$$

Figure 1 illustrates for *k*=1 *z* variable how the estimate of the difference between the two hedonic equations at the intercept, γ_1^*, has an adjustment to allow it to be evaluated at the Törnqvist mean from (11), $\bar{z}_{k-Törn}^{0,t}$. Note that we do not estimate the equations using OLS, but an estimator whose weights correspond to a Törnqvist index, i.e., a WLS where the weights are those outlined after equation (7). We focus on the case of the Törnqvist HI index with similar principles applying to the Walsh HI index.

Consider now the DTH index in (7) which constrains $\beta_k = (\beta_k^t - \beta_k^0) = 0$ in (10) and thus $\sum_{k=1}^{K} \beta_k D_k^{0,t}$ to be zero. The DTH index in (7) corresponds to a **Törnqvist (DTH) index** if estimated using WLS where the weights are those outlined after equation (7) above. A natural question is how does the estimated DTH index δ_1^* in (7), which is invariant to values of $\boldsymbol{z}_{ki}^{0,t}$, differ from the HI index evaluated *at the means* of $\boldsymbol{z}_{ki}^{0,t}$ in (13)?

(ii) how does a Törnqvist HI index differ from a Törnqvist DTH index?

This difference is first considered by comparing $\left(\delta_1^* - \gamma_1^*\right)$, the difference *in the intercepts,* where $\tilde{z}_k^{0,t} = 0$, between the constrained and unconstrained regression equations (7) and (10) respectively. The difference is 'omitted variable bias' due to the omission of $\sum_{k=1}^{K} \beta_k D_k^{0,t}$ in (7). Consider the case of a single *k*=1 characteristic, the principles being readily extended to more variables. Consider further the regression, using data in periods 0 and *t*, of the omitted variable—the slope dummy variable, $D_{1i}^{0,t} = z_{1i}^t$ if period *t* and 0 otherwise, in (10)—on the remaining right-hand side variables, the intercept dummy $D_0^{0,t}$ and the $z_{1i}^{0,t}$ characteristics:

$$D_{1i}^{0,t} = \lambda_0^0 + \lambda_1 D_0^{0,t} + \lambda_2^0 z_{1i}^{0,t} + \omega_i^1 \tag{14}$$

Omitted variable bias is the product of the coefficient on the omitted variable, β_1 (for *k*=1 in (10)), and the coefficient λ_1^* which is estimated from the above regression (14) on $D_{1i}^{0,t}$ i.e. the multicollinearity between the missing and included variables (Davidson and McKinnon, 1993). Thus the difference (before taking exponents) DTH *minus* HI, *at the intercept* is $\left(\delta_1^* - \gamma_1^*\right) = \beta_1^* \times \lambda_1^*$ and the (log of the) DTH index is thus given by $\delta_1^* = \beta_1^* \times \lambda_1^* + \gamma_1^*$. The (log of the) Törnqvist HI index at $\bar{z}_{1-Törn}^{0,t}$ from (11) for one variable is estimated as:

$$\gamma_1^* + \beta_1^* \bar{z}_{1-Törn}^{0,t} \tag{15}$$

Thus the (log of the) Törnqvist DTH index *minus* the (log of the) Törnqvist HI index at $\bar{z}_{1-\text{Törn}}^{0,t}$ is:

$$\beta_1^* \times \lambda_1^* + \gamma_1^* - \left(\gamma_1^* + \beta_1^* \bar{z}_{1-\text{Törn}}^{0,t}\right) = \beta_1^* \left(\lambda_1^* - \bar{z}_{1-\text{Törn}}^{0,t}\right) = (\beta_1^{*t} - \beta_1^{*0})\left(\lambda_1^* - \bar{z}_{1-\text{Törn}}^{0,t}\right) \tag{16}$$

and the ratio of the DTH and HI indexes *at the intercept* is $exp\left(\delta_1^*\right) - exp\left(\gamma_1^*\right) = exp\left(\beta_1^* \times \lambda_1^*\right)$ and the DTH index is thus given by $exp\left(\delta_1^*\right) = exp\left(\beta_1^* \times \lambda_1^*\right) \times exp\left(\gamma_1^*\right)$. The Törnqvist HI index at $\bar{z}_{1-Törn}^{0,t}$ from (11) for one variable is estimated as: $exp\left(\gamma_1^* + \beta_1^* \bar{z}_{1-\text{Törn}}^{0,t}\right)$. Thus the ratio of the Törnqvist DTH index to the Törnqvist HI index at $\bar{z}_{1-\text{Törn}}^{0,t}$ is:

$$\left(\frac{\text{Törnqvist DTH index}}{\text{Törnqvist HI index}}\right)_{\mathbf{M}_{1-\text{Törn}}^{0,t}} = exp\left[(\beta_1^{*t} - \beta_1^{*0})\left(\lambda_1^* - \bar{z}_{1-\text{Törn}}^{0,t}\right)\right] \tag{17}$$

where β_1^{*t} and β_1^{*0} are WLS estimates.

If *either of the two terms* making up the product on the right–hand-side is close to zero then there will be little difference between the indexes. *Neither parameter instability nor a change in the mean characteristic is sufficient in itself to lead to a difference between the formulas*. The $\beta_1^* = (\beta_k^{*t} - \beta_k^{*0})$ from (10) is the estimated marginal valuation of the characteristic between periods 0 and *t*, which can be positive or negative, but may be more generally thought to be negative to represent diminishing marginal utility/cost of the characteristic.

We now consider the WLS estimated λ_1^*. Bear in mind that the left-hand-side of the regression in equation (14), $D_{1i}^{0,t}$, is 0 in period 0 and z_{1i}^t in period *t* and that D^t on the right-hand-side is 1 in period *t* and zero in period 0. If we assume quality characteristics are positive, λ_1 will always be positive as the change from 0 in period 0 to their values in period *t*. Consider the weighted (Törnqvist) means $\bar{z}_{1-Torn}^0$ for $i \in S^0(1 \cap 2)$ and $i \in S^0(0 \neg t)$ (matched period 0 and unmatched old period 0) and $\bar{z}_{1-Torn}^t$ for $i \in S^t(1 \cap 2)$ and $i \in S^t(t \neg 0)$ (matched period *t* and unmatched new models in period *t).*

If we assume for simplicity that $\bar{\boldsymbol{z}}_{\boldsymbol{1-Törn}}^{\boldsymbol{0}} = \bar{\boldsymbol{z}}_{\boldsymbol{1-Törn}}^{\boldsymbol{t}}$**,** then $\lambda_1^* \approx \bar{z}_{1-Torn}^t$ since it is an estimate of the change in $D_{1i}^{0,t}$ in (14) arising from changing from period 0, where $D_{1i}^{0,t} = 0$, to period *t* where it is $\bar{z}_{1-Torn}^t$ and has an expectation of $\bar{z}_{1-Törn}^t$. The λ_1^* estimate is conditioned in (14) on $\bar{z}_{1-\text{Törn}}^{0,t}$, the change from $\bar{z}_{1-Törn}^0$ to $\bar{z}_{1-Törn}^t$, but since we assume these to have not changed our estimate of $\lambda_1^* \approx \bar{z}_{1-Törn}^t$ holds true.[9] Thus the second part of the difference expression in (14), $\left(\lambda_1^* - \bar{z}_{1-Törn}^{0,t}\right)$, is simply $\left(\bar{z}_{1-Törn}^t - \bar{z}_{1-Törn}^{0,t}\right)$ which, given our assumption of $\bar{z}_{1-Törn}^0 = \bar{z}_{1-Törn}^t$, is equal to 0. Thus from the right-hand side of (17), for samples with negligible change in

---

[8] The weights used to correspond to a generalized Törnqvist index were outlined following equation (7) .

the mean values of the characteristics, the DTH and HI will be similar irrespective of any parameter instability. Diewert (2002b) and Aizcorbe (2003) have shown that the DTH and HI indexes will be the same for matched models and this analysis gives support to their finding. However, we find first, that it is not matching *per se* that dictates the relationship; for unmatched models all that is required is that $\bar{z}^{t}_{1-\text{Törn}} = \bar{z}^{0,t}_{1-\text{Törn}}$ which may occur without matching—it simply requires the means of the characteristics not to change. Second, that even when the means change the two approaches will be equal if $\beta^{*t}_{k} - \beta^{*0}_{k} = 0$, i.e. there is parameter stability. Finally, it follows that if either of the two right-hand side expressions in (17) are large, the differences between the indexes will be compounded.

But what if $\bar{z}^{0}_{1-Törn} \neq \bar{z}^{t}_{1-Törn}$? The estimated coefficient λ^{*}_{1} from (14) regression is given by:[10]

$$\lambda^{*}_{1} = \frac{\bar{z}^{t}_{1}(N - N^{t})\sigma^{2}_{z} - N\,cov(D^{0,t}_{1i}, z^{0,t}_{1i})(\bar{z}^{t}_{1} - \bar{z}^{0,t}_{1})}{(N - N^{t})\sigma^{2}_{z} - (\bar{z}^{t}_{1} - \bar{z}^{0,t}_{1})^{2} N^{t}} \tag{18}$$

for an unweighted regression where N^{t} and N are the respective number of observations in period *t* and both periods 0 and *t*, σ^{2}_{z} is the variance of *z* and $cov(D^{0,t}_{1i}, z^{0,t}_{1i}) = \sum\left(z^{t}\right)^{2} - N^{t}\bar{z}^{t}\bar{z}$ is the covariance of $D^{0,t}_{1i}$ and $z^{0,t}_{1i}$ from (14). Readers are reminded that from (17):

$$\left(\frac{\text{Törnqvist DTH index}}{\text{Törnqvist HI index}}\right)_{\mathbf{M}^{0,t}_{1-\text{Törn}}} = exp\left[(\beta^{*t}_{1} - \beta^{*0}_{1})\left(\lambda^{*}_{1} - \bar{z}^{0,t}_{1-\text{Törn}}\right)\right] \tag{19}$$

First, as noted, if there is *either* negligible parameter instability *or* a negligible change in the mean of the characteristic, then there will be little difference between the formulas. However, as parameter instability increases *and* the change in the mean characteristic increases, the multiplicative effect on the difference between the indexes is compounded. The likely direction and magnitude of any difference is not immediately obvious. Assume diminishing marginal valuations of characteristics, so that $(\beta^{*t}_{1} - \beta^{*0}_{1}) \langle 0$. Second, even assuming a positive technological advance, $(\bar{z}^{t}_{1} - \bar{z}^{0,t}_{1}) \rangle 0$, and given $(N - N^{t})$, $\text{cov}(D^{0,t}_{1i}, z^{0,t}_{1i})$, $\bar{z}^{t}_{1}$ and σ^{2}_{z} are positive, it remains difficult to establish from (18) the effect on λ^{*}_{1} of changes in its constituent parts. However, third, as N^{t} becomes an increasing share of N, and at the limit if N^{t} takes up all of N (i.e. $(N - N^{t}) = 0$), then $\text{cov}(D^{0,t}_{1i}, z^{0,t}_{1i}) \rightarrow \sigma^{2}_{z}$ and more importantly, $\bar{z}^{t}_{1} \rightarrow \bar{z}^{0,t}_{1}$ and the difference between the formulas is dictated by $(\beta^{*t}_{1} - \beta^{*0}_{1})\left(-\bar{z}^{0,t}_{1}\right)$; i.e. a DTH index will exceed a HI index. Note that (18) is based on an OLS estimator and for a WLS Törnqvist estimator similar principles apply, though the determining factor for a DTH index to exceed a HI index is for the weights of new models to be increasing, a much more reasonable scenario. Thus other things being equal, the nature and extent of any differences between the two indexes will depend on (i) changes in the mean quality of models $(\bar{z}^{t}_{1} - \bar{z}^{0,t}_{1})$, (ii) the

---

[9] We are not requiring $\bar{z}^{0}_{1-Törn} = \bar{z}^{t}_{1-Törn}$; unconditioned estimates might be evaluated by examining omitted variable bias in (12).

[10] The estimated coefficient on x_1 of a regression of *y* on x_1 and x_2 is given by: $\dfrac{\sum yx_1 \sum x_2^2 - \sum yx_2 \sum x_1x_2}{\sum x_1^2 \sum x_2^2 - \left(\sum x_1x_2\right)^2}$

relative number of models in each period, $(N-N^t)$, (iii) the dispersion in *z*, σ_z^2, (iv) the absolute mean of the characteristics in period *t*, $\bar{z}_1^t$, (v) the $\mathrm{cov}(D_{1i}^{0,t}, z_{1i}^{0,t})$ which as $(N-N^t)\to 0$, i.e. N^t takes up all of N, then $\mathrm{cov}(D_{1i}^{0,t}, z_{1i}^{0,t})\to\sigma_z^2$ and $\bar{z}_1^t = \bar{z}_1$ and $\lambda_1^* = 0$, and (vi) parameter instability, $(\beta_1^{*t}-\beta_1^{*0})$.

B. Treatment of unmatched observations

Diewert (2002b) and Aizcorbe (2003) show that while the DTH and HI indexes will be the same for matched models, they differ in their treatment of unmatched data. Consider hedonic functions $\boldsymbol{h}_i^t\left(\boldsymbol{z}_i^t\right)$ and $\boldsymbol{h}_i^0\left(\boldsymbol{z}_i^0\right)$ for periods *t* and 0 respectively as in (8) and (9) and a (constrained) time dummy regression equation (7). Consider an unmatched observation only available in period *t*. A base period HI index such as (2) would exclude it, while a current period HI index such as (3) would include it and a geometric mean of the two (4) would give it half the weight in the calculation of that of a matched observation. A Törnqvist hedonic index (5) would also give an unmatched model half the weight of a matched one. For a DTH index such as (7) an unmatched period *t* model would appear only once in period *t* in the estimation of constrained parameters, as opposed to twice for matched data. We would therefore expect superlative HI indexes, such as (4), to be closer to DTH indexes than their constituent elements, (2) and (3) because they make symmetric use of the data.

C. Observations with undue influence

Silver (2002) has shown that while HI indexes such as (2)-(6) explicitly incorporate weights, they are only implicitly incorporated in the OLS or WLS estimator used for DTH indexes in a manner that may not be fully representative of the weights used. Adverse influence and leverage effects are shown to be generated by observations with unusual characteristics and above average residuals.

D. Chaining

Chained base HI indexes are preferred to fixed base ones, especially when matched samples degrade rapidly and their use reduces the spread between Laspeyres and Paasche. However, caution is advised in the use of chained monthly series when prices may oscillate around a trend (i.e. ‘bounce’) and as a result, chained indexes can ‘drift’ (Forsyth and Fowler, 1981 and Szulc, 1983).

III. EMPIRICAL STUDY: DESKTOP PCs

The empirical study is of the measurement of quality-adjusted monthly prices of UK desktop PCs in 1998. The data were monthly scanner data from the bar-code readers of PC retailers. There were 7,387 observations (models in a month sold in a specialized or non-specialized PC store-type) representing a sales volume of 1.5 million models worth £1.57 billion pounds over the year. Table 1 shows that in the January to February price comparison there were 584 matched models available in both months for the price comparison. However, for the January to December price comparison only 161 matched models were available with 509 unmatched ‘old’ models (available in January, but unmatched in December) and 436 unmatched ‘new’ models (available in December but unavailable in January for matching).[11]

The calculation of hedonic indices first require the estimation of hedonic regressions. To simplify the illustration only one variable was included, the speed in MHz. The regressions were for each month for the HI indexes and over January and the current month, including a dummy variable for the latter, for the DTH indexes. The estimated coefficients for speed in the hedonic regressions were statistically significant coefficients with the expected positive signs.[12]

The selectivity bias inherent in a *matched models* Törnqvist index is shown in Table 1 by the fall of 24% which understates the fall of around 50% for the two Törnqvist *hedonic* indexes which use all of the data. The particular concern of this paper is that the HI and DTH Törnqvist indexes give different results: falls of 55% and 50% respectively—a difference of nearly 10%—with no immediate explanation as to which is better. Having shown the need for hedonic indices we now turn to an explanation in Table 2 of the differences between the results from the two methods.

Table 2 column (1) shows the ratio of the two estimates, finding the DTH estimate to consistently fall at a slower rate. The early months have very little quality-adjusted price change compared with January and the difference between the formulas are less amenable to analysis. However, from later months it is clear that while the parameter instability $\left(\beta_1^{*t} - \beta_1^{*0}\right)$ in column (2) is an important driver of the differences between the hedonic indexes, other factors are at work. For example, in September and December the differences between the parameters were about the same, yet the differences between the HI and DTH indexes in column (1) are higher in December than September. Furthermore, in October and November the parameter instability falls compared with September, yet the differences between the indexes increase. The other factor at work and affecting the difference, as shown in equation (17), is of course $\left(\lambda_1^{*} - \bar{z}_{1-T\&n}^{0,t}\right)$ which in column (3) of Table 2 is seen to increase (in absolute values) in October, November and December. This increase explains that part of the differences in hedonic indexes unexplained by the parameter instability. The product of the two terms in column (3) shows an overall increase in October, November and December and, indeed, the exponent of column (3) equals column (1) by equation (17).

A driver of the substantial fall in $\left(\lambda_1^{*} - \bar{z}_{1-T\&n}^{0,t}\right)$ can be seem from columns (5) and (6) to be relatively large increase in $\bar{z}_{1-T\&n}^{0,t}$, the average speed of PCs over the two periods compared, an increase from 208 MHz for the January to February comparison to 273 MHz. for the January to December comparison. Bear in mind that the indexes were estimated using Törnqvist weights as outlined following equation (7). The overall sales value weight given to December models compared with January models increased by

[11] Bear in mind that the indexes estimated in this paper are weighted by shares in sales values and that the fall-off in the coverage of the matched sample by sales is even more dramatic: for the January to December comparison matched models made up only 71% of the January sales value and a mere 12% of the December sales value.

[12] The F-statistics for the null hypothesis of coefficients being equal to zero averaged 34.2 for HI indexes and 53.4 for DTH indexes, consistently rejecting the null at a 0.01% level and lower. The explanatory power of the estimated equations were naturally low for this specification with a single explanatory variable, especially since they did not include dummy variables on brand. Details of estimates from a fully specified model are available from the authors.

$\left[1-\left(0.70736-2\right)=0.29\right]$ 29% and the resulting fall in the statistic in column 9 of Table 2 is at least one of the reasons why the DTH index exceeded the HI index. Equation (18) and columns (7) to (10) of Table 2 show how λ_1^* can be decomposed, though the complexity of equation (18) makes it difficult to explain the factors that dictate its change, unless $\bar{z}_1^t = \bar{z}_1^{0,t}$ which is clearly not the case.

IV. CHOICE BETWEEN HEDONIC INDEXES AND TIME DUMMY HEDONIC INDEXES

The question of choice between these approaches has mainly been considered[13] from a perspective of concern over parameter instability. Berndt and Rappaport (2001) found, for example, from 1987 to 1999 for desktop PCs, the null hypothesis of adjacent-year equality to be rejected in all but one case. For mobile PCs the null hypothesis of parameter stability was rejected in eight of the 12 adjacent-year comparisons. Berndt and Rappaport's (2001) preferred the use of HI indexes if there was evidence of parameter instability. Pakes (2003) using quarterly data for hedonic regressions for desktop PCs over the period 1995 to 1999 rejected just about any hypothesis on the constancy of the coefficients. He also advocated HI indexes on the grounds that "....since hedonic coefficients vary across periods it [the DTH index approach] has no theoretical justification." Pakes (2003: 1593). However, equation (19) showed how the ratio of DTH and HI indexes was not solely dependent on parameter instability. It depended on the exponent of the product of two components: the change over time in the (WLS estimated) hedonic coefficients *and* the difference in (statistics that relate to) the (weighted) mean values of the characteristic. Even if parameters were unstable, the difference between the indexes may still be small due to a minimal change in the other component.

Consider two forms of 'spread'. The first arises because the basis of a DTH index is the constraining of parameters over time to be the same. Constraining the estimated coefficients to either β_1^{*t} or β_1^{*0} may give quite different results and this difference is the spread. A DTH index constrains the parameters to be the same, an average of the two. There is a sense in which we have less confidence in an index based on constraining similar parameters than one based on constraining two disparate parameters. This is the (implicit) stance taken by Berndt and Rappaport (2001) in their advice to avoid DTH indexes when there is parameter instability and to use HI indexes.

But there is a second type of spread and this relates to HI indexes. This is the difference between base and current period hedonic indexes, between (2) and (3), and arises from the use of a constant z_i^t as against z_i^0. The difference or spread between the indexes is dictated by the choice between which period's characteristics are held constant and there is a sense in which we have more confidence in an index based on an average of z_i^0 and z_i^t that has relatively small spread, than one based on an average of z_i^0 and z_i^t in which they were very different. There is thus a case to avoid HI indexes and use DTH indexes when

---

[13] Diewert (2002b) points out that the main advantage of hedonic imputed indexes is that it they are more flexible; i.e., changes in tastes between periods can readily be accommodated. HI indexes are argued to have a *disadvantage* that *two* distinct estimates will be generated and it is somewhat arbitrary how these two estimates are to be averaged to form a single estimate of price change. Diewert (2002b) further identifies the main advantages of the DTH method are that it conserves degrees of freedom and is less subject to multicollinearity problems.

characteristics change. If neither $(\beta_1^{*t} - \beta_1^{*0})$ nor $\left(\lambda_1^* - \bar{z}_{1-\text{Törn}}^{0,t}\right)$ is particularly large[14] relative to the other, then a symmetric average, say geometric mean, of the two indexes is to be preferred. If both differences are significant, but differ, a more appropriate estimate might be a weighted mean such as:

$$\frac{exp\left(\beta_1^{*t} - \beta_1^{*0}\right) \times HI + exp\left(\lambda_1^* - \bar{z}_{1-\text{Törn}}^{0,t}\right) \times DTH}{exp\left(\beta_1^{*t} - \beta_1^{*0}\right) + exp\left(\lambda_1^* - \bar{z}_{1-\text{Törn}}^{0,t}\right)} \tag{20}$$

V. CONCLUSIONS

It is recognised that extensive product differentiation with a high model turnover is an increasing feature of product markets (Triplett, 1999). The motivation of this paper lay in the failure of the matched models method to adequately deal with price measurement in this context and the need for hedonic indexes as the most promising alternative (Schultze and Mackie, 2002). The paper first developed in section I *generalized superlative hedonic indexes*, i.e. index number formulas which are *generalized* to deal with matched and unmatched models, use *hedonic* regressions to control for quality changes and are *superlative* in that they have a good foundations in economic theory and axiomatic considerations (Diewert, 2004). The paper second, considered the two main approaches to hedonic index numbers, HI, and DTH indexes. Their commonalities in terms of functional form of aggregator, weighting and periodicity of comparison, makes it difficult to justify one approach against the other, yet the two approaches can yield quite different results. This is of concern if they are to be the principle tools for dealing with quality adjustment in price measurement for product markets with high model turnover. In section II the paper provided a formal exposition of the factors underlying the difference between the two approaches. It was shown that differences between the two approaches may arise from both parameter instability and changes in the characteristics and such differences are compounded when both occur. It further showed that similarities between the two approaches resulted if there was little difference in either component. The empirical study of desktop PCs showed the superiority of hedonic indexes over matched model indexes, that DTH and HI indexes can differ, and that parameter instability need not be the main factor dictating such differences.

Consideration of the issue of choice between the two approaches was based in section IV on minimising two concepts of spread. The analysis led to the advice that (i) Either the DTH or HI index approach is acceptable if either the parameters are relatively stable or the characteristics do not change over time, otherwise (ii) DTH indexes be avoided and HI indexes used when there is evidence of parameter instability, (iii) HI indexes be avoided and DTH indexes used when there is evidence of changes in quality characteristics of the form given in the last term of (19), and (iv) that when both characteristics and parameters change substantially, an average, such as (20), of the two approaches be used.

[14] Note that statistical significance tests should not dictate whether a difference is large or small, it is also the magnitude of the differences that matters.

References

Aizcorbe, A. (2003) The Stability of Dummy Variable Price Measures Obtained from Hedonic Regressions, *Mimeo*, Federal reserve Board, Washington, D.C.

Berndt, E.R. and Rappaport N.J. (2001). Price and Quality of Desktop and Mobile Personal Computers: A Quarter-Century Historical Overview, *American Economic Review*, 91, 2, 268-273.

Boskin MS (Chair) Advisory Commission to Study the Consumer Price Index (1996), Towards a More Accurate Measure of the Cost of Living, *Interim Report to the Senate Finance Committee*, Washington DC.

Cole, R., Chen, Y.C., Barquin-Stolleman, J.A., Dulberger, E., Helvacian, N. and Hodge, J.H., (1986), Quality-Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment, *Survey of Current Businesses*, 66, 1, January 41-50.

Committee on National Statistics (2002) At What Price? Conceptualising and Measuring Cost-of-Living and price indexes, Panel on Conceptual, Measurement and Other Statistical Issues in Developing Cost-of-Living Indexes, Charles Schultze and Chris Mackie (eds.), Committee on National Statistics, Washington D.C., National Academy Press.

Davidson, J.R. and Mackinnon, J.G. (1993) *Estimation and Inference in Econometrics*, Oxford University Press.

Diewert, W.E. (2002a), Weighted Country Product Dummy Variable Regressions and Index Number Formulae, *University of British Columbia Department of Economics Discussion Paper Series*, 02-15, September, Vancouver, B.C., Canada.

Diewert W.E. (2002b). Hedonic Regressions: A Review of Some Unresolved Issues, *Mimeo*, Department of Economics, University of British Columbia.

Diewert, W.E. (2003), Hedonic Regressions: A Consumer Theory Approach. In Mathew Shapiro and Rob Feenstra (eds.), *Scanner Data and Price Indexes*, National Bureau of Economic Research, Studies in Income and Wealth, vol.61 Chicago: University of Chicago Press, 317-348.

Diewert, W.E. (2004), Chapters 15 to 18 in: The Manual on Consumer Price Indices: Theory and Practice, International Labour Office, Geneva: ILO.

Forsyth, FG, and Fowler, RF (1981), The Theory and Practice of Chain Price Index Numbers, *Journal of the Royal Statistical Society*, A, 144, 2, 224-47.

Griliches, Zvi [1988], "Postscript on Hedonics," in Zvi Griliches, *Technology, Education, and Productivity*, New York: Basil Blackwell, Inc., 119-122.

Haan, J. de (2003), Time Dummy Approaches to Hedonic Price Measurement, *Paper presented at the Seventh Meeting of the International Working Group on Price Indices*, 27-29 May 2003, INSEE, Paris.

Pakes A. (2003), A Reconsideration of Hedonic Price Indexes with an Application to PCs, *The American Economic Review*, 93, 5, December, 1576-93.

Rosen, S. (1974), Hedonic Prices and Implicit Markets: Product Differentiation and Pure Competition, *Journal of Political Economy*, 82, 34-49.

Schultze,C. and Mackie C. (2002) (editors), see Committee on National Statistics (2002) *op. cit.*

Silver, M. and Heravi, S. (2001), Scanner Data and the Measurement of Inflation, *The Economic Journal*, 11 June, 384-405.

Silver, M and Heravi, S. (2002), Why the CPI Matched Models Method May Fail Us: Results from an Hedonic and Matched Experiment Using Scanner Data, European *Central Bank (ECB) Working Paper* No. 144, Frankfurt: ECB.

Silver, M. (2002), The Use of Weights in Hedonic Regressions: the Measurement of Quality-Adjusted Price Changes. *Mimeo*, Cardiff Business School, Cardiff University.

Silver, M. and Heravi, S. (2003), The Measurement of Quality-Adjusted Price Changes. In Mathew Shapiro and Rob Feenstra (eds.), *Scanner Data and Price Indexes*, National Bureau of Economic Research, Studies in Income and Wealth, vol. 61, Chicago: University of Chicago Press, 277-317.

Stigler, G. (1961) The Price Statistics of the Federal Government, *Report* to the Office of Statistical Standards, Bureau of the Budget, New York, NBER.

Szulc, BJ (1983) Linking Price Index Numbers, in *Price Level Measurement*, 537-566, Ottawa, Statistics Canada.

Triplett, J.E. (1988), Hedonic Functions and Hedonic Indexes, in *The New Palgraves Dictionary of Economics*, pp. 630-634.

Triplett, J.E. (1999), “The Solow Productivity Paradox: What do Computers do to Productivity?”, *Canadian Journal of Economics*, 32, 2, April , 309-334.

Table 1, Number of matched and unmatched models and matched models and hedonic indexes

	Number of matched models	Number of unmatched old models in January of the comparison	Number of unmatched new models in the current month of the comparison	Matched models Törnqvist index	**HI** Törnqvist index	**DTH** Törnqvist index
	Figures are for comparisons between January and each current month			Fixed base, January 1998=1.000		
February	584	86	104	0.945	0.939	0.939
March	577	93	181	0.887	0.877	0.877
April	346	324	191	0.825	0.802	0.802
May	315	355	227	0.767	0.738	0.737
June	297	373	265	0.736	0.656	0.658
July	282	388	301	0.765	0.618	0.623
August	276	394	351	0.769	0.616	0.626
September	247	423	382	0.777	0.599	0.629
October	193	477	402	0.772	0.518	0.551
November	164	506	435	0.765	0.458	0.495
December	161	509	436	0.764	0.452	0.496

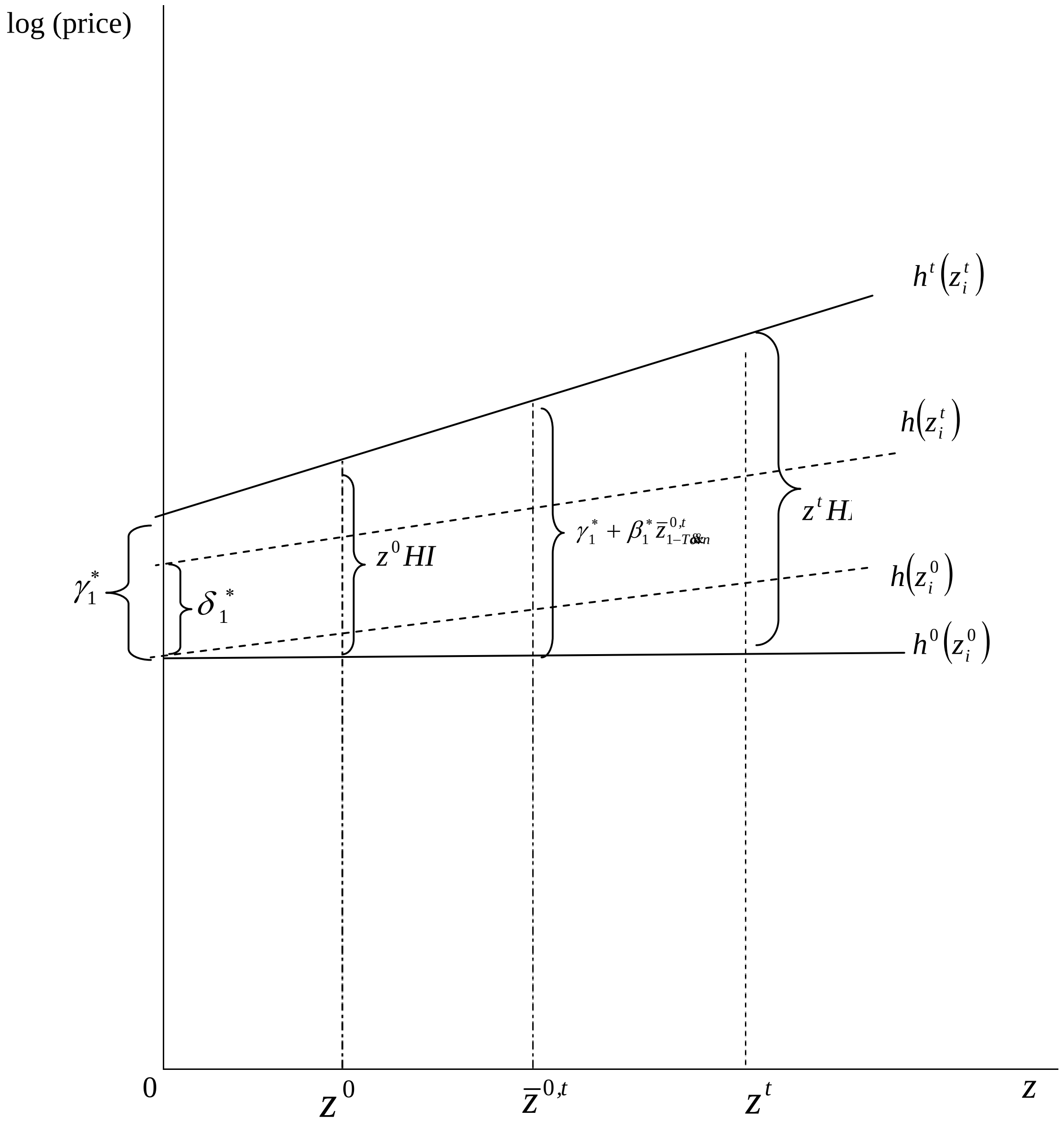
Table 2, Decomposition of differences between HI and Törnqvist indexes for desktop PCs, January 1998=1.000*

	HI ÷ DTH Törnqvist indexes (1)	$(\beta_1^{*t}-\beta_1^{*0})$ (2)	$(\lambda_1^{*}-\bar{z}^{0,t}_{1-\text{Törn}})$ (3)	$(\lambda_1^{*}-\bar{z}^{0,t}_{1-\text{Törn}}) \times(\beta_1^{*t}-\beta_1^{*0})$ (4)	$\bar{z}^{0,t}_{1-\text{Törn}}$ (5)	λ_1^{*} (6)	$(\bar{z}_1^{t}-\bar{z}_1^{0,t})$ (7)	$\text{cov}(D_{1i}^{0,t}, z_{1i}^{0,t})$ (8)	$(2-\sum_{i\in S(t} s_i^{0,t})$ (9)†	σ_z^2 (10)	$\bar{z}_1^{t}$ (11)
Fixed base, January 1998=1.000											
February	0.999999	-0.00042	0.00292	-0.0000012	208.0586	208.0616	-0.15937	1193.948	0.9975	2383.568	207.8993
March	1.000001	-0.0000021	-0.05901	0.0000013	211.0463	210.9873	0.961408	1323.364	0.98284	2339.425	212.0077
April	1.000121	0.000588	0.20158	0.0001186	219.3599	219.5615	7.047697	1952.174	0.993321	2431.949	226.4076
May	0.998721	0.00069	-1.86554	-0.0012871	225.0111	223.1456	11.53865	2911.7	0.957149	2822.441	236.5498
June	1.002505	-0.00033	-7.59551	0.0025079	231.0368	223.4413	16.03635	4207.419	0.879115	3395.21	247.0731
July	1.008229	-0.00078	-10.56785	0.0082167	236.5226	225.9548	18.31796	4804.071	0.830078	3640.906	254.8406
August	1.015499	-0.00081	-19.01683	0.0154110	244.4779	225.4611	23.22285	6481.596	0.796006	4599.436	267.7008
September	1.050295	-0.00172	-28.49892	0.0491427	250.8653	222.3664	26.35361	8272.112	0.754089	5888.89	277.2189
October	1.063765	-0.00154	-40.14595	0.0619327	262.9021	222.7562	34.28005	10189.09	0.732207	6729.035	297.1822
November	1.079882	-0.00166	-46.48257	0.0770125	269.5905	223.1079	37.8419	11185.42	0.717218	7220.941	307.4324
December	1.097709	-0.00177	-52.66507	0.0933989	273.4374	220.7723	39.68194	12243.07	0.70736	7995.446	313.1193

- All statistics are weighted using the weights commensurate with a Törnqvist index as outlined in the text following equation (7).

† The expression $(2-\sum_{i\in S(t} s_i^{0,t})$ replaces $(N-N^t)$ in (17) since the statistics are weighted. The difference is that between the sum of the shares over the two periods, i.e. 2, and the sum of that share arising from observations in period t . With new models taking on an increased share of sales the statistic falls.

Figure 1


log (price)
$h^t(z_i^t)$
$h(z_i^t)$
$z^t HI$
$\gamma_1^* + \beta_1^* \bar{z}^{0,t}_{1-T\&n}$
$z^0 HI$
$h(z_i^0)$
$h^0(z_i^0)$
γ_1^*
δ_1^*
0
z^0
$\bar{z}^{0,t}$
z^t
z