18th Meeting of the Ottawa Group on Price Indices at Ottawa, Canada, May 13-15, 2024.

Product Churn and Quality Adjustment: Using Scanner data of Laptop in Japan.*

W. Erwin Diewert† and Chihiro Shimizu‡

This Draft: May 12, 2024

Abstract:

High technology products are characterized by the rapid introduction of new models and the corresponding disappearance of older models. The paper addresses the problems associated with the construction of price indexes for these products. Several methods for the quality adjustment of product prices are considered: hedonic regressions that use either product characteristics (Time Dummy Characteristics regressions) or the product itself as the ultimate characteristic (Time Product Dummy regressions). The paper also considered regressions where the economic importance of products is taken into account (weighted versus unweighted regressions). The indexes which were generated by the hedonic regressions were compared to traditional index numbers that did not make any special adjustments for quality change. Finally, the Expanding Window variant of a Weighted Time Product Dummy regression was used to deal with the chain drift problem. The various approaches were implemented using Japanese price and quantity data on laptop sales in Japan for the 24 months over the years 2020-2021.

Key Words:

Quality adjustment, scanner data, hedonic regressions, predicted share similarity linking, expanding window approach to multilateral indexes, the chain drift problem, economic approach to index number theory.

Journal of Economic Literature Classification Codes

C32, C43, D20, D57, E31.

* The authors thank Naohito Abe, Jan de Haan, Kevin Fox, Alicia Rambaldi, Prasada Rao, and Paul Schreyer for their helpful comments. And the authors gratefully acknowledge the financial support from JSPS 24H00012.
† Vancouver School of Economics, University of British Columbia and the School of Economics, UNSW Australia, Sydney. Email address: erwin.diewert@ubc.ca.
‡ 2-1 Naka, Kunitachi, Tokyo, 186-8601, School of Social Data Science, Hitotsubashi University; Email address: c.shimizu@r.hit-u.ac.jp.

1. Introduction.

An increasing number of business firms are willing to share their price and quantity data on their sales of consumer goods and services to a national (or international) statistical office. These data are often referred to as scanner data.

Some scanner data involves high technology products which are characterized by product churn; i.e., the rapid introduction of new models and products and the short time that these new products are sold on the marketplace. This study will look at possible methods that statistical offices could use for quality adjusting this type of data. Our empirical example will use data on the sales of laptops in Japan.

A standard method for quality adjustment is the use of hedonic regressions. These hedonic regressions regress the price of a product (or a transformation of the price) on a time dummy variable and on either a dummy variable for the product or on the amounts of the price determining characteristics of the product. The first type of model is called a Time Product Dummy Hedonic regression while the second type of model is called a Time Dummy Characteristics Hedonic regression. The theory associated with these two classes of model will be discussed in sections 2 and 3 below. In particular, we will relate each hedonic regression to an explicit functional form for purchaser utility functions.

Section 4 discusses our laptop data for Japan which covers the 24 months in 2021 and 2022. The empirical hedonic regressions studied in this section are Time Dummy Characteristics type regressions. We used characteristics data on eight separate laptop characteristics in this section. We consider both unweighted (or more properly, equally weighted) least squares regression models with characteristics in this section. This section draws on the theory explained in section 3.

Section 5 draws on the theory explained in section 2; i.e., we consider weighted and unweighted Time Product Dummy hedonic regressions in this section. The models in this section use only a single product characteristic: the Japanese product code for each laptop sale. We consider a single panel regression versus a sequence of bilateral regressions that utilize the price and quantity data for two consecutive periods. The latter type of model can be implemented in real time and is called an Adjacent Period Time Product Dummy hedonic regression model.

Section 6 considers alternatives to hedonic regression models based on standard index number theory; i.e., maximum overlap chained Laspeyres, Paasche and Fisher indexes are computed in this section. We also compute the Predicted Share Similarity linked price indexes which have only been developed recently. The indexes calculated in this section are also "practical" indexes.

Unfortunately, the various real time indexes that are considered in sections 4-6 can suffer from a *chain drift problem*; i.e., an index for period t which is calculated by chaining together the results of adjacent period bilateral indexes is not equal to the corresponding index that directly compares the prices of period 1 to the prices of period t. In section 7, drawing on the work of Chessa (2016) (2021) an *Expanding Window* variant of the Implicit Weighted Time Product Dummy index is implemented which solves the chain drift problem (but may not be suitable if the products in scope are not close substitutes).

Section 8 lists some tentative conclusions that we can draw from this study. We also discuss the controversy between Krsinich (2016) and de Haan, Hendriks and Scholz (2021) on the merits of the Time Product Dummy and the Time Dummy Characteristics approaches to hedonic regressions.

The possible contributions of the present paper to the quality change literature include the following:

- A number of possible methods for dealing with the problem of quality change in the context of constructing price indexes are compared using the same data set. In particular, hedonic regression methods are compared to some multilateral index number methods.
- In section 2, we show that hedonic regression theory can be regarded as a special case of regular consumer theory where the preferences of purchasers are estimated.
- In section 3, we show how alternative assumptions about the functional form for the characteristics hedonic surface are related to the functional form for the purchaser's utility function.
- At the end of section 6, in the case where each product has its own unique characteristic and there are no missing observations in the two periods being compared, we show that hedonic imputation leads to the Fisher index using our economic approach to the measurement of quality change.
- In section 7, we show that the expanding window methodology pioneered by Chessa (2016) (2021) leads to indexes which are free from chain drift.

2. Hedonic Regressions and Utility Theory: The Time Product Dummy Hedonic Regression Model.

The problem of adjusting the prices of similar products due to changes in the quality of the products should be related to the usefulness or utility of the products to purchasers. Each product in scope has varying amounts of various *characteristics* which will determine the utility of the product to purchasers. A *hedonic regression* is typically based on regressing a product price (or a transformation of the product price) on the amounts of the various price determining characteristics of the product. An alternative hedonic regression model may be based on regressing the product prices on *product dummy variables*; i.e., each product has its own unique bundle of price determining characteristics which can be represented by a product dummy variable.[1] Each of these hedonic regression models can be related to specific functional forms for purchaser utility functions. In this section, we consider the second class of hedonic regression models and in the following section, we consider the first class of hedonic regression models that regress product prices on product characteristics.

Assume that there are N products in scope and T time periods. Let $p^t \equiv [p_{t1},...,p_{tN}]$ and $q^t \equiv [q_{t1},...,q_{tN}]$ denote the (unit value) price and quantity vectors for the products in scope for time periods $t = 1,...,T$.[2] Initially, we assume that there are no missing prices or quantities so that all prices and quantities are positive. We assume that each purchaser of the N products maximizes the following *linear function* $f(q)$ in each time period:

[1] This alternative class of models is more general than the first class so one could ask why should we consider estimating the characteristics model in place of the time product dummy variable model? Product churn may be so great that there are not enough degrees of freedom to accurately estimate the product dummy variables. Consider as a limiting case where every product is a new product in each period. The Time Product Dummy (TPD) regression model cannot be estimated in this case. Secondly, a new improved product loaded with useful characteristics may not cause older products to exit the market immediately due to incomplete information on the part of purchasers; i.e., consumers may not realize immediately how good the new product is until some time has passed. Krsinich (2016) following Diewert (2004), explained how the TPD hedonic regression model cannot accurately estimate a quality adjustment factor for a new product during the period of its introduction.

[2] The analysis in this section follows that of Diewert (2022; section 5).

(1) $f(q) = f(q_1,q_2,...,q_N) \equiv \Sigma_{n=1}^{N} \alpha_n q_n \equiv \alpha\cdot q$

where the α_n are positive parameters, which can be interpreted as quality adjustment factors. Under the assumption of utility maximizing behavior on the part of each purchaser of the N commodities and assuming that each purchaser in period t faces the same period t price vector p^t,[3] it can be shown that the aggregate period t vector of purchases q^t is a solution to the aggregate period t utility maximization problem, $\max_q \{\alpha\cdot q : p^t\cdot q = e^t ; q \geq 0_N\}$ where e^t is equal to aggregate period t expenditure on the N products. The first order conditions for an interior solution, q^t, λ_t to the period t aggregate utility maximization problem are the following N+1 equations, where λ_t is a Lagrange multiplier:

(2) $\alpha = \lambda_t p^t$;
(3) $p^t\cdot q^t = e^t$.

Take the inner product of both sides of equations (2) with the observed period t aggregate quantity vector q^t and solve the resulting equation for λ_t. Using equation (3), we obtain the following expression for λ_t:

(4) $\lambda_t = \alpha\cdot q^t/e^t > 0$.

Define π_t as follows:

(5) $\pi_t \equiv 1/\lambda_t$.

Divide both sides of equations (2) by λ^t and using definition (5), we obtain the *basic time product dummy estimating equations* for period t:[4]

(6) $p_{tn} = \pi_t\alpha_n$; $\quad t = 1,\ldots,T ; n = 1,\ldots,N.$

The period t aggregate price and quantity levels for this model, P^t and Q^t, are defined as follows:

(7) $Q^t \equiv \alpha\cdot q^t$;
(8) $P^t \equiv e^t/Q^t$
$\quad = \pi_t$

where the second equation in (8) follows using (4) and (5). Thus equations (6) have the following interpretation: the period t price of product n, p_{tn}, is equal to the period t price level π_t times a quality adjustment parameter for product n, α_n.[5]

[3] These are strong assumptions but strong assumptions are required in order to relate hedonic regression models to the utility of the products in scope.
[4] This model dates back to Court (1939; 109-111). He transformed these equations by taking logarithms of both sides of equations (6) and adding error terms. Diewert (2003b) (2023) considered the index number implications of making alternative transformations of the basic equations (6) and endorsed Court's transformation in the end.
[5] Note that α_n is the marginal utility to a purchaser of a unit of product n for n = 1,…,N. It can be shown that the period t price index π_t is equal to $c(p^t)$ where c(p) is the unit cost function that is dual to the utility function f(q); see Diewert (1974) on duality theory. Let $p \equiv [p_1,\ldots,p_N]$ be a vector of positive prices. The unit cost function that corresponds to the linear utility function $f(q) = \Sigma_{n=1}^{N} \alpha_n q_n$ is $c(p) = \min_n \{p_n/\alpha_n : n = 1,\ldots,N\}$. If utility maximizing purchasers buy

At this point, it is necessary to point out that our consumer theory derivation of equations (6) is not accepted by all economists. Rosen (1974) and Triplett (1987) (2004) have argued for a more general approach to the derivation of hedonic regression models that is based on supply conditions as well as on demand conditions. The present approach is obviously based on only consumer (or purchaser) preferences. This consumer oriented approach was endorsed by Griliches (1971; 14-15), Muellbauer (1974; 988) and Diewert (2003a) (2003b). Of course, the functional form assumptions which justify the present consumer approach are quite restrictive but, nevertheless, it is useful to imbed hedonic regression models in a traditional consumer demand setting.

Empirically, equations (6) are unlikely to hold exactly. Following Court (1939), we assume that the exact model defined by (6) holds only to some degree of approximation and so we add error terms e_{tn} to the right hand sides of equations (6). The unknown parameters, $\pi \equiv [\pi_1,...,\pi_T]$ and $\alpha \equiv [\alpha_1,...,\alpha_N]$, can be estimated as solutions to the following (nonlinear) least squares minimization problem:

$$(9)\ \min_{\alpha,\pi} \Sigma_{n=1}^{N} \Sigma_{t=1}^{T} [p_{tn} - \pi_t\alpha_n]^2 .$$

However, Diewert (2023) showed that the estimated price levels π_t^* that solve the minimization problem (9) had unsatisfactory axiomatic properties: for example, the estimated period t price levels π_t are not invariant to the units of measurement for the products. Thus we follow Court and take logarithms of both sides of the exact equations (6) and add error terms to the resulting equations. This leads to the following *least squares minimization problem*:[6]

$$(10)\ \min_{\rho,\beta} \Sigma_{n=1}^{N} \Sigma_{t=1}^{T} [\ln p_{tn} - \rho_t - \beta_n]^2$$

where the new parameters ρ_t and β_n are defined as the logarithms of the π_t and α_n; i.e., define :

$$(11)\ \rho_t \equiv \ln\pi_t\ ; \qquad t = 1,...,T;$$
$$(12)\ \beta_n \equiv \ln\alpha_n\ ; \qquad n = 1,...,N.$$

However, the least squares minimization problem defined by (10) does not weight the log price terms $[\ln p_{tn} - \rho_t - \beta_n]^2$ by their *economic importance* and so we consider the following *weighted least squares minimization problem*:[7]

$$(13)\ \min_{\rho,\beta} \Sigma_{n=1}^{N} \Sigma_{t=1}^{T} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2$$

positive amounts of all N products, then it must be the case that $p_1/\alpha_1 = p_2/\alpha_2 = \ldots = p_N/\alpha_N$ and purchasers get the same amount of utility from the purchase of one unit of each product.

[6] This model is an adaptation of Summer's (1973) country product dummy model to the time series context. See Aizcorbe, Corrado and Doms (2000) for an early application of this model in the time series context.

[7] Rao (1995) (2004) (2005; 574) was the first to consider this model using expenditure share weights; see also Diewert (2004). However, Balk (1980; 70) suggested this class of models much earlier using somewhat different weights. For the case of 2 periods, see Diewert (2004) (2005a) and de Haan (2004a).

where s_{tn} is the expenditure share of product n in period t. The first order necessary conditions for $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ to solve (13) simplify to the following T equations (14) and N equations (15):[8]

(14) $\rho_t^* = \Sigma_{n=1}^N s_{tn}[\ln p_{tn} - \beta_n^*]$; $t = 1,...,T;$

(15) $\beta_n^* = \Sigma_{t=1}^T s_{tn}[\ln p_{tn} - \rho_t^*]/(\Sigma_{t=1}^T s_{tn})$; $n = 1,...,N.$

Solutions to (14) and (15) are not unique: if $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ solve (14) and (15), then so do $[\rho_1^*+\lambda,...,\rho_T^*+\lambda]$ and $[\beta_1^*-\lambda,...,\beta_N^*-\lambda]$ for all λ. Thus we can set $\rho_1^* = 0$ in equations (15) and drop the first equation in (14) and use linear algebra to find a unique solution for the resulting equations.[9] Once the solution is found, define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* as follows:

(16) $\pi_t^* \equiv \exp[\rho_t^*]$; $t = 1,...,T$; $\alpha_n^* \equiv \exp[\beta_n^*]$; $n = 1,...,N.$

Note that since we have set $\rho_1^* = 0$, $\pi_1^* = 1$. The price levels π_t^* defined by (16) are called the *Weighted Time Product Dummy price levels*. Note that the resulting *price index* between periods t and τ is defined as the ratio of the period t price level to the period τ price level and is equal to the following expression:

(17) $\pi_t^*/\pi_\tau^* = \prod_{n=1}^N \exp[s_{tn}\ln(p_{tn}/\alpha_n^*)]/\prod_{n=1}^N \exp[s_{\tau n}\ln(p_{\tau n}/\alpha_n^*)]$; $1 \le t, \tau \le T.$

If $s_{tn} = s_{\tau n}$ for $n = 1,...,N$, then π_t^*/π_τ^* will equal a weighted geometric mean of the price ratios $p_{tn}/p_{\tau n}$ where the weight for $p_{tn}/p_{\tau n}$ is the common expenditure share $s_{tn} = s_{\tau n}$. Thus π_t^*/π_τ^* will not depend on the α_n^* in this case.

Once the estimates for the π_t and α_n have been computed, we have two methods for constructing period by period price and quantity levels, P^t and Q^t for $t = 1,...,T$. The π_t^* estimates can be used to form the aggregates using equations (18) or the α_n^* estimates can be used to form the aggregate period t price and quantity levels using equations (19):[10]

(18) $P^{t*} \equiv \pi_t^*$; $Q^{t*} \equiv p^t \cdot q^t/\pi_t^*$; $t = 1,...,T;$

(19) $Q^{t**} \equiv \alpha^* \cdot q^t$; $P^{t**} \equiv p^t \cdot q^t/\alpha^* \cdot q^t$; $t = 1,...,T.$

Define the error terms $e_{tn} \equiv \ln p_{tn} - \ln\pi_t^* - \ln\alpha_n^*$ for $t = 1,...,T$ and $n = 1,...,N$. If all $e_{tn} = 0$, then P^{t*} will equal P^{t**} and Q^{t*} will equal Q^{t**} for $t = 1,...,T$.[11] However, if the error terms are not all equal to zero, then the

[8] If information on expenditures or quantities is not available, then the weighted least squares problem is replaced by the unweighted least squares problem (10). The first order conditions for the simplified problem (10) are given by (14) and (15) where the shares s_{tn} are replaced by the numbers 1/N for all t and n. In this unweighted case, the price index defined by (17) collapses down to a Jevons index.

[9] Alternatively, one can set up the linear regression model defined by $(s_{tn})^{1/2}\ln p_{tn} = (s_{tn})^{1/2}\rho_t + (s_{tn})^{1/2}\beta_n + e_{tn}$ for $t = 1,...,T$ and $n = 1,...,N$ where we set $\rho_1 = 0$ to avoid exact multicollinearity. This is the procedure we used in our empirical work below. Iterating between equations (14) and (15) will also generate a solution to these equations and the solution can be normalized so that $\rho_1 = 0$.

[10] Note that the price level P^{t**} defined in (19) is a quality adjusted unit value index of the type studied by de Haan (2004b).

[11] If all $e_{tn} = 0$, then the unweighted (or more accurately, the equally weighted) least squares minimization problem defined by (10) will generate the same solution as is generated by the weighted least squares minimization problem

statistical agency will have to decide on pragmatic grounds which option to use to form period t price and quantity levels: (18) or (19).

It is reasonably straightforward to generalize the weighted least squares minimization problem (13) to the case where there are missing prices and quantities. Assume that there are N products and T time periods but not all products are purchased (or sold) in all time periods. For each period t, define the set of products n that are present in period t as $S(t) \equiv \{n: p_{tn} > 0\}$ for $t = 1,2,...,T$. It is assumed that these sets are not empty; i.e., at least one product is purchased in each period. For each product n, define the set of periods t where product n is present as $S^*(n) \equiv \{t: p_{tn} > 0\}$. Again, assume that these sets are not empty; i.e., each product is sold in at least one time period. The generalization of (13) to the case of missing products is the following *weighted least squares minimization problem*:[12]

$$(20)\ \min_{\rho,\beta} \Sigma_{t=1}^{T} \Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2 = \min_{\rho,\beta} \Sigma_{n=1}^{N} \Sigma_{t\in S^*(n)} s_{tn}[\ln p_{tn} - \rho_t - \beta_n]^2.$$

Note that there are two equivalent ways of writing the least squares minimization problem; the first way uses the definition for the set of products n present in period t, S(t), while the second way uses the definition for the set of periods t where product n is present, $S^*(n)$. The first order necessary conditions for $\rho_1,...,\rho_T$ and $\beta_1,...,\beta_N$ to solve (20) are the following counterparts to (14) and (15):[13]

$$(21)\ \Sigma_{n\in S(t)} s_{tn}[\rho_t^* + \beta_n^*] = \Sigma_{n\in S(t)} s_{tn}\ln p_{tn}\ ; \qquad t = 1,...,T;$$
$$(22)\ \Sigma_{t\in S^*(n)} s_{tn}[\rho_t^* + \beta_n^*] = \Sigma_{t\in S^*(n)} s_{tn}\ln p_{tn}\ ; \qquad n = 1,...,N.$$

As usual, the solution to (21) and (22) is not unique: if $\rho^* \equiv [\rho_1^*,...,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,...,\beta_N^*]$ solve (21) and (22), then so do $[\rho_1^*+\lambda,...,\rho_T^*+\lambda]$ and $[\beta_1^*-\lambda,...,\beta_N^*-\lambda]$ for all λ. Thus we can set $\rho_1^* = 0$ in equations (22), drop the first equation in (21) and use linear algebra to find a unique solution for the resulting equations.[14] Define the estimated *price levels* π_t^* and *quality adjustment factors* α_n^* by definitions (11) and (12). Substitute these definitions into equations (21) and (22). After some rearrangement, equations (21) and (22) become the following equations:

$$(23)\ \pi_t^* = \exp[\Sigma_{n\in S(t)} s_{tn}\ln(p_{tn}/\alpha_n^*)]\ ; \qquad t = 1,...,T;$$
$$(24)\ \alpha_n^* = \exp[\Sigma_{t\in S^*(n)} s_{tn}\ln(p_{tn}/\pi_t^*)/\Sigma_{t\in S^*(n)} s_{tn}]\ ; \qquad n = 1,...,N.$$

Once the estimates for the π_t and α_n have been computed, we have the usual two methods for constructing period by period price and quantity levels, P^t and Q^t for $t = 1,...,T$. The counterparts to definitions (18) are the following definitions:

$$(25)\ P^{t*} \equiv \pi_t^* = \exp[\Sigma_{n\in S(t)} s_{tn}\ln(p_{tn}/\alpha_n^*)]\ ; \qquad t = 1,...,T;$$
$$(26)\ Q^{t*} \equiv \Sigma_{n\in S(t)} p_{tn}q_{tn}/P^{t*}\ ; \qquad t = 1,...,T.$$

---

defined by (13). This fact gives rise to the following rule of thumb: if the unweighted problem (10) fits the data very well, then it is not necessary to work with the more complicated weighted problem (13).

[12] If only price information is available, then replace the s_{tn} in (20) by $1/N(t)$ where N(t) is the number of products that are purchased in period t.

[13] The unweighted (i.e., equally weighted) counterpart least squares minimization problem to (20) sets all $s_{tn} = 1$ for $n\in S(t)$. The resulting first order conditions are equations (21) and (22) with the positive s_{tn} replaced with a 1.

[14] The resulting system of $T - 1 + N$ equations needs to be of full rank in order to obtain a unique solution. The solution can also be obtained by running a linear regression.

Thus P^{t*} is a weighted geometric mean of the quality adjusted prices p_{tn}/α_n^* that are present in period t where the weight for p_{tn}/α_n^* is the corresponding period t expenditure (or sales) share for product n in period t, s_{tn}. The counterparts to definitions (19) are the following definitions:

$$(27)\ Q^{t**} \equiv \Sigma_{n\in S(t)}\, \alpha_n^* q_{tn}\ ; \qquad t = 1,...,T;$$

$$\begin{aligned}(28)\ P^{t**} &\equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/Q^{t**} && t = 1,...,T;\\ &= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^* q_{tn} && \text{using (27)}\\ &= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_n^*(p_{tn})^{-1}p_{tn}q_{tn}\\ &= [\Sigma_{n\in S(t)}\, s_{tn}(p_{tn}/\alpha_n^*)^{-1}]^{-1}\\ &\leq \exp[\Sigma_{n\in S(t)}\, s_{tn}\ln(p_{tn}/\alpha_n^*)]\\ &= P^{t*}\end{aligned}$$

where the inequality follows from Schlömilch's inequality[15]; i.e., a weighted harmonic mean of the quality adjusted prices p_{tn}/α_n^* that are present in period t, P^{t**}, will always be less than or equal to the corresponding weighted geometric mean of the prices where both averages use the same share weights s_{tn} when forming the two weighted averages. The inequalities $P^{t**} \leq P^{t*}$ imply the inequalities $Q^{t**} \geq Q^{t*}$ for t = 1,...,T. The inequalities (28) are due to de Haan (2004b) (2010) and de Haan and Krsinich (2014) (2018; 763). The model used by de Haan and Krsinich is a more general hedonic regression model which includes the time dummy model used in the present section as a special case.

If the estimated errors $e_{tn}^* \equiv \ln p_{tn} - \rho_t^* - \beta_n^*$ that implicitly appear in the weighted least squares minimization problem turn out to equal 0, then the equations $p_{tn} = \pi_t\alpha_n$ for t = 1,...,T, $n\in S(t)$ hold without error and the hedonic regression provides a good approximation to reality. Moreover, under these conditions, P^{t*} will equal P^{t**} for all t. If the fit of the model is not good, then it may be necessary to look at other methods for quality adjustment.

The solution to the weighted least squares regression problem defined by (20) can be used to generate imputed prices for the missing products. Thus if product n in period t is missing, define $p_{tn} \equiv \pi_t^*\alpha_n^*$. The corresponding missing quantity is defined as $q_{tn} \equiv 0$. Some statistical agencies use hedonic regression models to generate imputed prices for missing prices and then use these imputed prices in their chosen index number formula.

One unsatisfactory property of the WTPD price levels π_t^* is the following one: a product that is available in only one period out of the T periods has no influence on the aggregate price levels π_t^*.[16] This means that the price of a new product that appears in period T has no influence on the price levels. The hedonic regression models in the next section that make use of information on the characteristics of the products do not have this unsatisfactory property of the weighted time product dummy hedonic regression models studied in this section.

3. The Time Dummy Hedonic Regression Model with Characteristics Information.

---

[15] See Hardy, Littlewood and Pólya (1934; 26).
[16] Diewert (2004) established this property.

In this section, it is again assumed that there are N products that are available over a window of T periods. As in the previous sections, we again assume that the quantity aggregator function for the N products is the linear function, $f(q) \equiv \alpha \cdot q = \Sigma_{n=1}^{N} \alpha_n q_n$ where q_n is the quantity of product n purchased or sold in the period under consideration and α_n is the quality adjustment factor for product n. What is new is the assumption that the quality adjustment factors are functions of a vector of K *characteristics* of the products. Thus it is assumed that product n has the vector of characteristics $z^n \equiv [z_{n1},z_{n2},...,z_{nK}]$ for n = 1,...,N. We assume that this information on the characteristics of each product has been collected.[17] The new assumption in this section is that the quality adjustment factors α_n are functions of the vector of characteristics z^n for each product and the same function, g(z) can be used to quality adjust each product; i.e., we have the following assumptions:

(29) $\alpha_n \equiv g(z^n) = g(z_{n1},z_{n2},...,z_{nK})$; $\qquad$ n = 1,...,N.

Thus each product n has its own unique mix of characteristics z^n but the *same function* g can be used to determine the relative utility to purchasers of the products. Define the period t quantity vector as $q^t = [q_{t1},...,q_{tN}]$ for t = 1,...,T. If product n is missing in period t, then define $q_{tn} \equiv 0$. Using the above assumptions, the aggregate quantity level Q^t for period t is defined as:

(30) $Q^t \equiv f(q^t) \equiv \Sigma_{n=1}^{N} \alpha_n q_{tn} = \Sigma_{n=1}^{N} g(z^n) q_{tn}$; $\qquad$ t = 1,...,T.

Using our assumption of (exact) utility maximizing behavior with the linear utility function defined by (30), equations (6) become the following equations:

(31) $p_{tn} = \pi_t g(z^n)$; $\qquad$ t = 1,...,T; $n \in S(t)$.

The assumption of approximate utility maximizing behavior is more realistic, so error terms need to be appended to equations (31). We also need to choose a functional form for the *quality adjustment function* or *hedonic valuation function* $g(z) = g(z_1,...,z_K)$. We will not be able to estimate the parameters for a general valuation function, so we assume that g(z) is the product of K separate functions of one variable of the form $g_k(z_k)$; i.e., we assume that g(z) is defined as follows:

(32) $g(z_1,...,z_K) \equiv g_1(z_1)g_2(z_2)\ldots g_K(z_K)$.

This is the assumption of *multiplicative separability*. The assumption that $g(z_1,\ldots,z_K)$ is *additively separable* in its characteristics is $g(z_1,...,z_K) \equiv g_1(z_1) + g_2(z_2) +\ldots+ g_K(z_K)$. Many other assumptions about the functional form for g are possible. If all $z_k > 0$, one could assume that $\ln g(z_1,...,z_K)$ is equal to $\gamma_0 + \Sigma_{k=1}^{K} \gamma_k \ln z_k + \frac{1}{2}\Sigma_{i=1}^{K}\Sigma_{k=1}^{K} \gamma_{ik} \ln z_i \ln z_k$ where $\gamma_{ik} = \gamma_{ki}$ for all i and k (translog functional form). Or one may try to estimate a general nonparametric approximation to $g(z_1,...,z_K)$. The problem with assuming very general functional forms for g(z) is that there will not be enough degrees of freedom to estimate all of the parameters that describe the general functional form. Thus typically investigators using the characteristics approach to the estimation of hedonic regressions assume that $\ln g(z_1,...,z_K)$ is equal to $\gamma_0 + \Sigma_{k=1}^{K} \gamma_k \ln z_k$, which is a first order Taylor series approximation (in logs) to a general g.

[17] Basically, we want to collect information on the most important price determining characteristics of each product; see Triplett (2004) and Aizcorbe (2014) for many examples of this type of hedonic regression and references to the applied literature on this topic. Of course, the fact that information on product characteristics must be collected is a disadvantage of the class of models studied in this section.

For our particular example, each characteristic takes on only a finite number of discrete values so in the empirical sections of this paper, we will assume that each $g_k(z_k)$ is a step function or a "plateaux" function which jumps in value at a finite number of discrete numbers in the range of each z_k. This assumption will eventually lead to a regression model where all of the independent variables are dummy variables.[18]

For each characteristic k, we partition the observed sample range of the z_k into N(k) discrete intervals which exactly cover the sample range. Let I(k,j) denote the jth interval for the variable z_k for k = 1,…,K and j = 1,…,N(k). For each product observation n in period t (which has price p_{tn}) and for each characteristic k, define the indicator function (or dummy variable) $D_{tn,k,j}$ as follows:

(33) $D_{tn,k,j} \equiv 1$ if observation n in period t has the amount of characteristic k, z_{nk}, that belongs to the jth interval for characteristic k, I(k,j) where k = 1,…,K and j = 1,…,N(k);
$\equiv 0$ if the amount of characteristic k for observation n in period t, z_{nk}, does not belong to the interval I(k,j).

We use definitions (33) in order to define $g(z^n) = g(z_{n1},z_{n2},\ldots,z_{nK})$ for product n if it is purchased in period t:[19]

(34) $g(z_{n1},z_{n2},\ldots,z_{nK}) \equiv (\Sigma_{j=1}^{N(1)}\, a_{1j}D_{tn,1,j})(\Sigma_{j=1}^{N(2)}\, a_{2j}D_{tn,2,j}) \ldots (\Sigma_{j=1}^{N(K)}\, a_{Kj}D_{tn,K,j})$.

Let P be the vector of all observed prices of the products in scope over the T periods. Substitute equations (34) into equations and we obtain the following system of possible estimating equations where the π_t and $a_{1j}, a_{2j}, \ldots, a_{Kj}$ are unknown parameters:

(35) $p_{tn} = \pi_t(\Sigma_{j=1}^{N(1)}\, a_{1j}D_{tn,1,j})(\Sigma_{j=1}^{N(2)}\, a_{2j}D_{tn,2,j}) \ldots (\Sigma_{j=1}^{N(K)}\, a_{Kj}D_{tn,K,j})$; $\quad t = 1,\ldots,T;\ n\in S(t)$.

Take logarithms of both sides of equations (35) in order to obtain the following system of estimating equations:[20]

(36) $\ln p_{tn} = \ln\pi_t + \Sigma_{j=1}^{N(1)}\,(\ln a_{1j})D_{tn,1,j} + \Sigma_{j=1}^{N(2)}\,(\ln a_{2j})D_{tn,2,j} + \ldots + \Sigma_{j=1}^{N(K)}\,(\ln a_{Kj})D_{tn,K,j}$; $\quad t = 1,\ldots,T;\ n\in S(t)$.

Define the following parameters:

(37) $\rho_t \equiv \ln\pi_t$; $t = 1,\ldots,T$; $b_{1j} \equiv \ln a_{1j}$; $j = 1,\ldots,N(1)$; $b_{2j} \equiv \ln a_{2j}$; $j = 1,\ldots,N(2)$; …; $b_{Kj} \equiv \ln a_{Kj}$; $j = 1,\ldots,N(K)$.

Upon substituting definitions (37) into equations (36) and adding error terms e_{tn}, we obtain the following linear regression model:

---

[18] Instead of assuming that the $g_k(z_k)$ are step functions, one could assume that they are linear, quadratic or cubic spline functions of one variable. Our step function model is identical to Krsinich's (2016; 380) Main Effects Model which is a special case of her Fully Interacted Time Dummy Hedonic Model. Using the notation in (38) above, her fully interacted model would partition the characteristics space into $N(1)\times N(2)\times\ldots\times N(K)$ distinct cells and assign a parameter for each of these many cells. Of course, in any practical example, most of these cells would be empty; i.e., there would be no product that had the characteristics represented by any given finely defined cell.

[19] If product n is purchased in periods t and τ, then the expression on the right hand side of (34) remains the same.

[20] The hedonic price index which is generated by the model defined by equations (35) is not invariant to changes in the units of measurement of the characteristics; see Diewert (2023).

$$(38)\ \ln p_{tn} = \Sigma_{t=1}^{T} \ln \rho_t + \Sigma_{j=1}^{N(1)} b_{1j} D_{tn,1,j} + \Sigma_{j=1}^{N(2)} b_{2j} D_{tn,2,j} + \ldots + \Sigma_{j=1}^{N(K)} b_{Kj} D_{tn,K,j} + e_{tn}; \quad t = 1,...,T;\ n \in S(t).$$

There are a total of $T + N(1) + N(2) + \ldots + N(K)$ unknown parameters in equations (38). The least squares minimization problem that corresponds to the linear regression model defined by (38) is the following least squares minimization problem:

$$(39)\ \min_{\rho,\, b(1),\, b(2),\, \ldots,\, b(K)} \Sigma_{t=1}^{T} \Sigma_{n \in S(t)} \{\ln p_{tn} - \rho_t - \Sigma_{j=1}^{N(1)} b_{1j} D_{tn,1,j} - \Sigma_{j=1}^{N(2)} b_{2j} D_{tn,2,j} - \ldots - \Sigma_{j=1}^{N(K)} b_{Kj} D_{tn,K,j}\}^2$$

where ρ is the vector $[\rho_1, \rho_2, \ldots, \rho_T]$ and $b(k)$ is the vector $[b_{k1}, b_{k2}, \ldots, b_{kN(k)}]$ for $k = 1,2\ldots,K$. Solutions to the least squares minimization problem will exist but a solution will not be unique.[21] Using equations (35), it can be seen that components of the vectors π and $a(k) \equiv [a_{k1}, a_{k2}, \ldots, a_{kN(k)}]$ for $k = 1,2,\ldots,K$ are multiplied together to give us predicted values for the p_{tn}. Thus the parameters in any one of these K+1 vectors can be arbitrary but at least one component of each of the remaining vectors must be set equal to a constant. A useful unique solution to (39) is obtained by setting $\rho_1 = 0$ (which corresponds to $\pi_1 = 1$) and setting $b_{k1} = 0$ (which implies $a_{k1} = 1$) for $k = 2,\ldots,K$ (so b_{11} is not normalized).

However, in our empirical work, we used a slightly different parameterization for g(z) which treated each characteristic in a symmetric fashion. We replaced assumptions (35) with the following assumptions:

$$(40)\ p_{tn} = \pi_t a_0 (\Sigma_{j=1}^{N(1)} a_{1j} D_{tn,1,j})(\Sigma_{j=1}^{N(2)} a_{2j} D_{tn,2,j}) \ldots (\Sigma_{j=1}^{N(K)} a_{Kj} D_{tn,K,j})\ ; \quad t = 1,...,T;\ n \in S(t).$$

The new restrictions on the parameters are as follows: $\pi_1 = 1$; $a_{k1} = 1$ for $k = 1,\ldots,K$. This new parameterization of g(z) does not affect the π_t but in addition to treating the characteristics in a symmetric fashion, it allows the estimation of a constant term, a_0, in the regression model. Define $b_0 \equiv \ln a_0$ and take logarithms of both sides of (40) and we obtain the following linear regression model:

$$(41)\ \ln p_{tn} = \rho_t + b_0 + \Sigma_{j=2}^{N(1)} b_{1j} D_{tn,1,j} + \Sigma_{j=2}^{N(2)} b_{2j} D_{tn,2,j} + \ldots + \Sigma_{j=2}^{N(K)} b_{Kj} D_{tn,K,j} + e_{tn}; \quad t = 1,...,T;\ n \in S(t).$$

The restrictions that $b_{k1} \equiv \ln a_{k1} = 0$ for $k = 1,\ldots,K$ have been imposed in equations (41). An additional restriction that is required to obtain a unique solution to the counterpart to the least squares regression problem defined by (39) is $\rho_1 = 0$. The linear regression defined by (41) (with $\rho_1 = 0$) can be run and estimates for the unknown parameters $[\rho_2^*, \ldots, \rho_T^*]$, b_0^* and $[b_{k2}^*, \ldots, b_{kN(k)}^*]$ for $k = 1,2\ldots,K$ will be available. Use these estimates to define the logarithms of the quality adjustment factors α_n for all products n that were purchased in period t:[22]

$$(42)\ \beta_{tn}^* \equiv b_0^* + \Sigma_{j=2}^{N(1)} b_{1j}^* D_{tn,1,j} + \Sigma_{j=2}^{N(2)} b_{2j}^* D_{tn,2,j} + \ldots + \Sigma_{j=2}^{N(K)} b_{Kj}^* D_{tn,K,j}\ ; \quad t = 1,...,T;\ n \in S(t).$$

The corresponding estimated product n quality adjustment factors α_{tn}^* are obtained by exponentiating the β_{tn}^*:

$$(43)\ \alpha_{tn}^* \equiv \exp[\beta_{tn}^*]\ ; \quad t = 1,...,T;\ n \in S(t).$$

[21] Thus the X matrix that corresponds to the linear regression model defined by equations (36) will not have full column rank.

[22] If product n is available in multiple periods, the quality adjustment factors remain the same across periods.

Using the above α_{tn}^*, we can form a direct estimate for the aggregate quantity or utility obtained by purchasers during period t:

$$(44)\ Q^{t**} \equiv \Sigma_{n\in S(t)}\ \alpha_{tn}^* q_{tn}\ ; \qquad t = 1,...,T.$$

The corresponding period t price level obtained indirectly, P^{t**}, is defined by deflating period t expenditure by period t aggregate quantity:

$$(45)\ P^{t**} \equiv \Sigma_{n\in S(t)}\ p_{tn}q_{tn}/Q^{t**} = \Sigma_{n\in S(t)}\ p_{tn}q_{tn}/\Sigma_{n\in S(t)}\ \alpha_{tn}^* q_{tn}\ ; \qquad t = 1,\ldots,T.$$

In order to obtain a useful expression for the direct estimate for the period t price level, π_t, look at the first order conditions for minimizing the sum of squared errors with respect to ρ_t:

$$(46)\ 0 = \Sigma_{n\in S(t)}\ \{\ln p_{tn} - \rho_t^* - b_0^* + \Sigma_{j=1}^{N(1)}\ b_{1j}^*\ D_{tn,1,j} - \Sigma_{j=1}^{N(2)}\ b_{2j}^*\ D_{tn,2,j} - \ldots - \Sigma_{j=1}^{N(K)}\ b_{Kj}^*\ D_{tn,K,j}\} \qquad t = 2,\ldots,T$$
$$= \Sigma_{n\in S(t)}\ \{\ln p_{tn} - \rho_t^* - \beta_n^*\}$$

where we used definitions (42) to derive the second equality. Let N(t) be the number of products purchased in period t for t = 1,…,T. Using definitions (37) and (42), equations (46) imply that the direct estimate of the period t price level π_t^* is equal to:

$$(47)\ \pi_t^* = \Pi_{n\in S(t)}(p_{tn}/\alpha_{tn}^*)^{1/N(t)} \equiv P^{t*}\ ; \qquad t = 2,\ldots,T.$$

Thus the direct estimate for the period t price level P^{t*} is equal to the geometric mean of the period t quality adjusted prices (p_{tn}/α_{tn}^*) for the products that were purchased in period t. Note that this price level can be calculated using price information alone whereas the indirect measure P^{t**} requires price and quantity information on the purchase of products during period t.

A problem with the least squares minimization problem defined by (41) is that it does not take the economic importance of the products into account. Thus, we consider the corresponding weighted least squares problem defined below:

$$(48)\ \min_{\rho,\ b\text{'s}}\ \Sigma_{t=1}^{T}\ \Sigma_{n\in S(t)}\ s_{tn}\{\ln p_{tn} - \rho_t - b_0 - \Sigma_{j=2}^{N(1)}\ b_{1j}D_{tn,1,j} - \Sigma_{j=2}^{N(2)}\ b_{2j}D_{tn,2,j} - \ldots - \Sigma_{2=1}^{N(K)}\ b_{Kj}D_{tn,K,j}\}^2$$

where $s_{tn} = p_{tn}q_{tn}/\Sigma_{j\in S(t)}\ p_{tj}q_{tj}$ for t = 1,…,T and $n\in S(t)$ and we use the same definitions as were used in the unweighted (or more properly, the equally weighted) least squares minimization problem defined below (39).

The new weighted counterpart estimating equations to the linear regression model that was defined by equations (41) is given below:

$$(49)\ (s_{tn})^{1/2}\ln p_{tn} = (s_{tn})^{1/2}(\rho_t + b_0 + \Sigma_{j=2}^{N(1)}\ b_{1j}D_{tn,1,j} + \Sigma_{j=2}^{N(2)}\ b_{2j}D_{tn,2,j} + \ldots + \Sigma_{j=2}^{N(K)}\ b_{Kj}D_{tn,K,j}) + e_{tn};$$
$$t = 1,...,T;\ n\in S(t).$$

Use definitions (42)-(45) to define new β_{tn}^*, α_{tn}^*, Q^{t**} and P^{t**}. We rewrite P^{t**} in a somewhat different manner as follows:

$$(50)\ P^{t**} = \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_{tn}^{*}q_{tn} \qquad t = 1,\ldots,T$$
$$= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, (\alpha_{tn}^{*}/p_{tn})p_{tn}q_{tn}$$
$$= [\Sigma_{n\in S(t)}\, s_{tn}(p_{tn}/\alpha_{tn}^{*})^{-1}]^{-1}.$$

In order to obtain a useful expression for the direct estimate for the period t price level, π_t, look at the first order conditions for minimizing the new weighted least squares problem (48) with respect to ρ_t:

$$(51)\ 0 = \Sigma_{n\in S(t)}\, s_{tn}\{\ln p_{tn} - \rho_t^{*} - b_0^{*} + \Sigma_{j=2}^{N(1)}\, b_{1j}^{*}\, D_{tn,1,j} - \Sigma_{j=2}^{N(2)}\, b_{2j}^{*}\, D_{tn,2,j} - \ldots - \Sigma_{j=2}^{N(K)}\, b_{Kj}^{*}\, D_{tn,K,j}\}$$
$$= \Sigma_{n\in S(t)}\, s_{tn}\, \{\ln p_{tn} - \rho_t^{*} - \beta_n^{*}\}\ ; \qquad t = 2,\ldots,T$$

where we used definitions (42) to derive the second equality. Note that $\Sigma_{n\in S(t)}\, s_{tn} = 1$. Using definitions (37) and (43), equations (51) imply that the direct estimate of the period t price level π_t^{*} is equal to:[23]

$$(52)\ \pi_t^{*} = \Pi_{n\in S(t)}(p_{tn}/\alpha_{tn}^{*})^{s(t,n)} \equiv P^{t*}\ ; \qquad t = 2,\ldots,T$$

where $s(t,n) = s_{tn}$. The indirect period t quantity level is defined (as usual) as period t expenditure divided by P^{t*}:

$$(53)\ Q^{t*} \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/P^{t*}\ ; \qquad t = 1,\ldots,T.$$

Note that the direct period t price level defined by (52), P^{t*}, is a period t share weighted geometric mean of the period t quality adjusted prices p_{tn}/α_{tn}^{*} while the indirect period t price level P^{t**} defined by (50) is a period t share weighted harmonic mean of the period t quality adjusted prices and thus we have the de Haan inequalities: $P^{t**} \leq P^{t*}$ and $Q^{t**} \geq Q^{t*}$ for t = 1,…,T.

We turn to an empirical example where we estimate alternative hedonic regression models and make use of the above analysis.

4. Laptop Data for Japan and Hedonic Regressions Using Characteristics.

We obtained data from a private firm that collects price, quantity and characteristic information on the monthly sales of laptop computers across Japan. The data are thought to cover more than 60% of all laptop sales in Japan. We utilized the data for the 24 months in the years 2021 and 2022 for our regressions and index computations. There were 2639 monthly price and quantity observations on laptops sold in total over all months. Thus the prices and quantities are p_{tn} and q_{tn} where p_{tn} is the average monthly (unit value) price for product n in month t in Yen and q_{tn} is the number of product n units sold. The mean (positive) q_{tn} was 594.7 and the mean (positive) p_{tn} was 117640 yen. Over the 24 months in our sample, 366 distinct products were sold so n = 1,…,366. To save some space, we now set t = 1,2,…,24. If product n did not sell in month t, then we set p_{tn} and q_{tn} equal to 0. If each product sold in each month, we would have 366 x 24 = 8784 positive monthly prices and quantities, p_{tn} and q_{tn}, but on average, only 30.0% of the products were sold per month since 2639/8784 = 0.300. Thus there is tremendous product churn in the sales of laptops in Japan, with individual products quickly entering and then exiting the market for laptops.

[23] Our normalizations imply $\pi_1^{*} = 1$.

The positive prices p_{tn} and quantities q_{tn} are listed in Appendix B as the variables P and Q. This Appendix also lists the corresponding month of sale and the Japanese Product Code number (JAN) for each entry. This table also lists information on 9 additional characteristics of the laptop product, which are discussed below.

NEW is the number of months that the product has been available (or more precisely, sold) in Japan. NEW = 1 means the product was a new one. NEW ranges from 1 to 38 months in our sample. We did not use this characteristic in this paper.

CLOCK is the clock speed of the laptop. The mean clock speed was 1.94 and the range of clock speeds was 1 to 3.4. The larger is the clock speed, the faster the computer can make computations. There were 23 distinct clock speeds for the laptops in our sample.

MEM is the memory capacity for the laptop. The mean memory size was 8188.9. There were only 3 memory sizes in our sample: 4,096, 8,192 and 16,384.

SIZE is the screen size of the laptop. The mean screen size (in inches) was 14.49. There were 10 distinct screen sizes in our sample: 11.6, 12, 12.5, 13.3, 14, 15.4, 15.6,16, 16.1 and 17.3.

PIX is the number of pixels imbedded in the screen of the laptop. The mean number of pixels was 24.82. There were only 10 distinct number of pixels in our sample: 10.49, 12.46, 12.96, 20.74, 33.18, 40.96, 51.84, 55.30, 58.98 and 82.94.

HDMI is the presence (HDMI = 1) or absence (HDMI = 0) of a HDMI terminal in the laptop. If HDMI =1, then it is possible to display digitally recorded images without degradation.

WEIGHT is the weight of the laptop in kilograms. Laptop weights ranged from 0.747 to 2.9 kilos.

A priori, we expected that purchasers would value higher clock speed, memory capacity, screen size, the number of pixels and the availability of HDMI in a laptop, leading to increasing estimated coefficients for the dummy variables corresponding to higher values of the characteristic under consideration. We expected that purchasers would value a lighter laptop over a heavier one.

CPU is the type of Central Processing Unit that the laptop used. There were 12 types of CPU in our sample.

BRAND is the name of the manufacturer of the laptop. In the data file, BRAND takes on the values 1-12 but the second brand is not present in 2021-2022 so we have only 11 brands in our sample. BRAND is frequently used as an explanatory variable in a hedonic regression as a proxy for company wide product characteristics that may be missing from the list of explicit product characteristics that are included in the regression.

Table B1 in Appendix B lists the above variables in columns of dimension 2639.

The information in the column vectors TD and JAN were used to generate 24 time dummy variables, D_1, D_2, …, D_{24} and 366 product dummy variable vectors, D_{J1}, D_{J2}, …, D_{J366}.

In our regressions and calculation of price and quantity indexes, we transformed some of our units of measurement to make the mean value of the series closer to unity. Thus the p_{tn} were replaced by p_{tn}/100,000 so we are measuring prices in units of 100,000 Yen. Similarly MEM was replaced by MEM/1000, SIZE was replaced by SIZE/10 and PIX was replaced by PIX/10. The basic descriptive statistics for the above variables (after transformation) are listed in Table 1 below. The variables P and Q are the 2639 positive prices and quantities p_{tn} and q_{tn} stacked up into vectors of dimension 2639.

Table 1: Descriptive Statistics for the Variables

Variable	Mean	Std. Dev	Variance	Minimum	Maximum
JAN	195.75	103.94	10803.00	1.000	366
NEW	9.638	5.965	35.580	1.000	38
CLOCK	1.940	0.518	0.268	1.000	3.4
MEM	8.189	3.436	11.804	4.096	16.384
SIZE	1.449	0.138	0.019	1.160	1.730
PIX	2.482	1.289	1.662	1.049	8.294
HDMI	0.753	0.431	0.186	0.000	1.000
CPU	4.186	2.777	7.710	1.000	10
BRAND	9.153	2.209	4.880	1.000	12
WEIGHT	1.648	0.523	0.273	0.747	2.900
Q	595	736	541230	100	5367
P	1.176	0.492	0.242	0.174	2.873

Number of observations=2639

It is of interest to calculate the average price of a laptop that was sold in period t, PA^t, for each of the 24 months of data in our sample:

(54) $PA^t \equiv \Sigma_{n\in S(t)}\, p_{tn}/N(t)$; $\qquad t = 1,\ldots,24$

where N(t) is the number of laptops sold in period t and S(t) is the set of products sold in period t.

The average period t price of a laptop, PA^t, weights each period t laptop price equally and thus does not take the economic importance of each type of laptop into account. A more representative measure of average laptop price in period t is the period t *unit value price*, PUV^t, defined as follows:

(55) $PUV^t \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, q_{tn} = \Sigma_{n\in S(t)}\, e_{tn}/\Sigma_{n\in S(t)}\, q_{tn}$ $\qquad t = 1,\ldots,24$

where $e_{tn} \equiv p_{tn}q_{tn}$ is expenditure or sales of product n in period t for t = 1,…,24 and n = 1,…,366.

We convert the average prices defined by (54) and (55) into price indexes by dividing each series by the corresponding series value by the corresponding period 1 entry. Thus define the period t *average price index* $P_A{}^t$ and the period t *unit value price index* $P_{UV}{}^t$ as follows:

(56) $P_A{}^t \equiv PA^t/PA^1$; $P_{UV}{}^t \equiv PUV^t/PUV^1$; $\qquad t = 1,\ldots,24.$

The time series N(t), PA^t, PUV^t, $P_A{}^t$ and $P_{UV}{}^t$ are listed below in Table 2.

Table 2: Average Prices and Unit Values and Average Price and Unit Value Price Indexes

Month t	N(t)	PA^t	PUV^t	$P_A{}^t$	$P_{UV}{}^t$
1	146	1.23522	1.28422	1.000	1.000
2	134	1.27876	1.28041	1.035	0.997
3	147	1.27849	1.2967	1.035	1.010
4	133	1.2615	1.28001	1.021	0.995
5	110	1.31278	1.30992	1.063	1.020
6	95	1.31639	1.28645	1.066	1.002
7	103	1.26883	1.26349	1.027	0.984
8	94	1.26053	1.25112	1.020	0.974
9	83	1.24859	1.22112	1.011	0.951
10	78	1.27961	1.27247	1.036	0.991
11	71	1.25161	1.21663	1.013	0.947
12	72	1.17273	1.12868	0.949	0.879
13	124	1.11517	1.08334	0.903	0.844
14	136	1.12928	1.08597	0.914	0.846
15	150	1.11056	1.08594	0.899	0.846
16	135	1.15121	1.09629	0.932	0.854
17	105	1.10092	1.0304	0.891	0.802
18	109	1.06995	1.0154	0.866	0.791
19	107	1.05176	1.02634	0.851	0.799
20	101	1.02677	1.01863	0.831	0.793
21	100	1.04738	0.99001	0.848	0.771
22	91	1.1161	1.09602	0.904	0.853
23	96	1.06155	1.08657	0.859	0.846
24	119	1.1024	1.12772	0.892	0.878
Mean	109.96	1.177	1.1597	0.953	0.903

It can be seen that the equally weighted average price of a laptop, PA^t, is on average 1.5% higher than the average unit value price, PUV^t, since 1.1770/1.1597 = 1.01492. This means that on average, laptop models that have low sales have higher prices than high volume models. However, there are substantial fluctuations in average prices so that at times, $PA^t > PUV^t$, which happens when t = 1. When we convert the average prices PA^t and PUV^t into the price indexes $P_A{}^t$ and $P_{UV}{}^t$, it turns out that the mean of the $P_A{}^t$ is 0.95287 which is substantially higher than the mean of the $P_{UV}{}^t$ which is 0.90302. However, the two index number series end up fairly close to each other at month 24: $P_A{}^{24} = 0.89247$ while $P_{UV}{}^{24} = 0.87814$. We regard the unit value price index series, $P_{UV}{}^t$, as being more accurate than the average price series, $P_A{}^t$.

Note that the number of separate models sold in month t, N(t), ranges from a low of 71 in month 11 to a high of 147 in month 3. If each model sold in every month, then N(t) would equal 366 for each month.

Of course, the price indexes $P_A{}^t$ and $P_{UV}{}^t$ make no adjustments for changes in the average quality of laptops over time. Thus we now consider hedonic regression models of the type defined by equations (41) in the previous section. Initially, we consider various characteristics regressions that use the data for all 24 months in our panel. The resulting indexes are not real time "practical" indexes that could be produced by a national statistical office since typically Consumer Price Indexes cannot be revised. However, these panel data regressions are useful in that they will indicate how important it is to include a particular characteristic in the hedonic regression. Moreover, these panel type regressions are not subject to chain drift over the sample period.

Appendix A describes how these Time Dummy Characteristics Regressions were constructed in some detail. Our final Time Dummy Characteristics Hedonic regression of the type defined by equations (41) used all of the above characteristics except "newness". Our final model can be rewritten in vector notation as follows:

$$(57)\ \ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + \Sigma_{j=2}^{5} b_{Pj} D_{Pj} + b_{H2} D_{H2} + \Sigma_{j=2}^{11} b_{Bj} D_{Bj} + \Sigma_{j=2}^{10} b_{Uj} D_{Uj} + \Sigma_{j=2}^{7} b_{Wj} D_{Wj} + e$$

where lnP is a vector of log prices of dimension 2639, the D_t are vectors of time dummy variables, ONE is a vector of ones of dimension 2639, the D_{Cj}, D_{Mj}, D_{Sj}, D_{Pj}, D_{H2}, D_{Bj}, D_{Uj} and D_{Wj} are vectors of dummy variables for the step functions for the clock speed, memory size, screen size, number of pixels, the availability of HDMI, brand, type of computer chip and the laptop weight variables.

The R^2 between the observed price vector and the predicted price vector was 0.8926. The estimated coefficients on the time dummy variables in this regression are ρ_2^*, ρ_3^*, …, ρ_{24}^*. Define $\rho_1^* \equiv 0$ and the estimated period t price levels $\pi_t^* \equiv \exp[\rho_t^*]$ for t = 1,2,…,24. Define the month t *Time Dummy Characteristics Price Index*, $P_{TDC}^t \equiv \pi_t^*$ for t = 1,…,24. This index is listed in Table A2 of Appendix A and is plotted on Chart 1 below.

Recall that the expenditure share that corresponds to purchased product n in month t was defined as $s_{tn} = p_{tn}q_{tn}/\Sigma_{j\in S(t)}\, p_{tj}q_{tj}$ for t = 1,…,24 and $n\in S(t)$. To obtain the weighted counterpart to the hedonic regression model defined by (57) above, form a share vector of dimension 2639 that corresponds to the $\ln p_{tn}$ that appear in the vector lnP and then form a new vector of dimension 2639 that consists of the positive square roots of each s_{tn}. Call this vector of square roots SS. Now multiply both sides of (57) by SS to obtain a new linear regression model which again provides estimates for the unknown parameters that appear in (57). The R^2 for this new weighted regression model turned out to be 0.9152 which is substantially higher than the R^2 for the counterpart unweighted model which was 0.8926. The parameter estimates for this weighted hedonic regression model are listed in Table A1 of Appendix A.

The estimated coefficients on the time dummy variables in this regression are ρ_2^*, ρ_3^*, …, ρ_{24}^*. Define $\rho_1^* \equiv 0$ and define the estimated period t price levels $\pi_t^* \equiv \exp[\rho_t^*]$ for t = 1,2,…,24. Define the month t *Weighted Time Dummy Characteristics Price Index*, $P_{WTDC}^t \equiv \pi_t^*$ for t = 1,…,24. This index is listed in Table A2 in Appendix A and plotted in Chart 1 below and it is our preferred index thus far. The corresponding unweighted (or equally weighted) Time Dummy Characteristics Price Index P_{TDC}^t is also listed in Table A2 along with the simple average laptop price indexes P_A^t and P_{UV}^t defined by definitions (55) above.

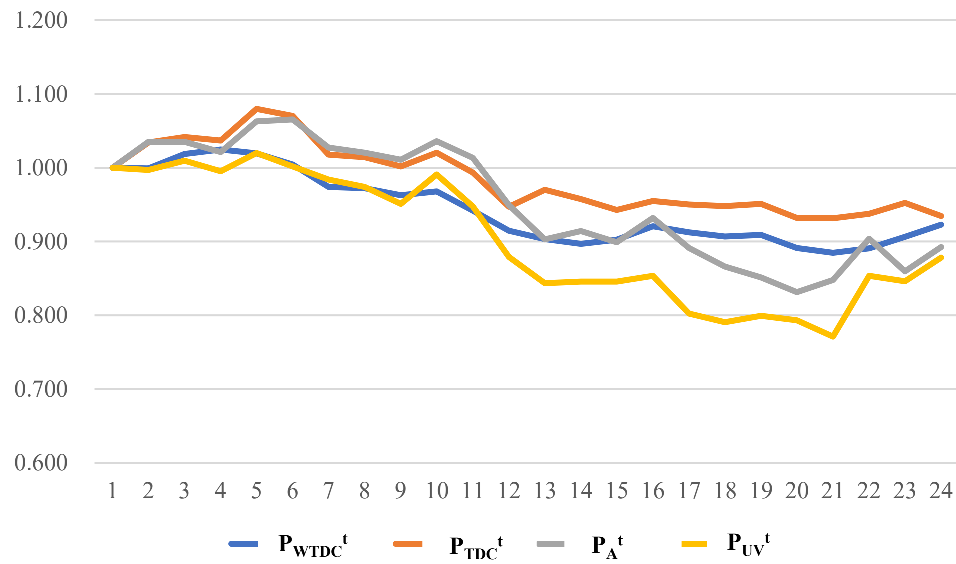

Chart 1. Time Product Dummy and Average Price Indexes.

The results in Table A2 and Chart 1 are not very plausible. Our preferred hedonic index, P_{WTDC}^t, ends up at 0.9229 when t = 24 which is well above the simple average price indexes P_A^t and P_{UV}^t for t = 24 (which ended up at 0.8925 and 0.8781). It seems unlikely that a quality adjusted price index for laptops could end up *higher* than a simple average price index for laptops. Note that the equally weighted hedonic price index P_{TDC}^t, ended up even higher at 0.93462 when t = 24, which shows that weighting matters. We also note that *missing characteristics* can greatly affect the resulting hedonic price: as we added characteristics to the various hedonic regressions, the resulting indexes changed significantly.

Although the weighted and unweighted time product characteristic indexes end up reasonably close to each other in month 24, there are substantial month to month differences between the two indexes. Moreover the mean of the weighted indexes P_{WTDC}^t (0.9436) is substantially below the mean of the unweighted indexes P_{TDC}^t (0.9842). Our conclusion here is that *weighting for laptops matters* and the weighted index should be produced by statistical agencies if price and quantity information is available.

It can be seen that our theoretically preferred index, P_{WTDC}^t, is the smoothest of the four indexes plotted on Chart 1 but in the latter half of the sample period, it lies well above the Unit Value price index, P_{UV}^t, which does not make any quality adjustments. Thus P_{WTDC}^t does not seem to be a plausible quality adjusted price index; i.e., it is unlikely that the quality of laptop computers declined over the two years.

There are two additional problems with our "best" directly defined weighted hedonic price index using characteristics, P_{WTDC}^t:

- It is not a real time index; i.e., it is a retrospective index that is calculated using the data covering two years;[24]
- It does not allow for gradual taste change on the part of purchasers.

[24] This difficulty can be overcome by using rolling window hedonic regressions. See de Haan (2015), Krsinich (2016) and Diewert and Fox (2022; 360-361) for discussions of the issues surrounding linking the results from a new panel of data with the results from a previous panel.

These difficulties can be avoided if we restrict the number of months T to be equal to 2. This restriction leads to *adjacent period hedonic regressions*.[25] Thus we can use the analytical framework presented in section 3 and simply apply it to the case where T = 2.

To start the adjacent period methodology, we use the price data for products n that were sold in months 1 and 2. We also use data on the 8 characteristics of the products that were used in the last regression described above. The counterpart regression (linking month 1 to month 2) to the unweighted time dummy characteristic hedonic regression defined by (57) above is the following regression model:

$$(58)\ \ln P = \rho_2 D_2 + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + \Sigma_{j=2}^{5} b_{Pj} D_{Pj} + b_{H2} D_{H2} + \Sigma_{j=2}^{11} b_{Bj} D_{Bj} + \Sigma_{j=2}^{10} b_{Uj} D_{Uj} + \Sigma_{j=2}^{7} b_{Wj} D_{Wj} + e$$

where lnP is now the vector of log prices for the products which were sold only in months 1 and 2. Similarly, the vectors of independent variables on the right hand side of (58) are not of dimension 2639 but only of dimension equal to the number of products that were sold in months 1 and 2. Note that there is only a single time dummy variable D_2 on the right hand side of (58) and the nt component of D_2 takes on the value 1 for the products sold in month 2 and the value 0 for the products sold in month 1. The definitions for the other characteristic dummy variables on the right hand side of (58) are similar to our earlier panel wide definitions but now these characteristic dummy variables are only defined for products that were sold in months 1 and 2.[26]

Define $P^{1*} \equiv 1$ as the month 1 index level. Define ρ_2^* as the estimated month 2 time dummy coefficient for the bilateral regression defined by (58)[27] and define π_2^* as the exponential of ρ_2^*; i.e., define $\pi_2^* \equiv \exp[\rho_2^*]$. Define the month 2 direct price level as $P^{2*} \equiv \pi_2^*$.

Next, we restricted the definition of lnP to the products that were sold only in months 2 and 3. The new adjacent period hedonic regression was similar to the one defined by (58) except the time dummy term $\rho_2 D_2$ on the right hand side of (58) was replaced with the term $\rho_3 D_3$ where D_3 takes on the value 1 for the products

[25] For references to the history of this approach to hedonic regressions, see the many references in Diewert (2022) (2023).

[26] However, some complications occurred when implementing the above operations. When the data were restricted to 2 adjacent periods instead of the entire 2 years of data, some of the characteristic dummy variable vectors became zero vectors. To deal with this problem, some of our characteristic dummy variable vectors were aggregated together. Thus the clock speed dummy variables for groups 6 and 7 were aggregated together to form a new group 6. The terms $\Sigma_{j=2}^{7} b_{Cj} D_{Cj}$ on the right hand side of (58) were replaced by the terms $\Sigma_{j=2}^{6} b_{Cj} D_{Cj}$. The screen size dummy variables for groups 1 and 2 were aggregated together as were the dummy variables for groups 6 and 7. Thus the terms $\Sigma_{j=2}^{7} b_{Sj} D_{Sj}$ on the right hand side of (58) were replaced by the terms $\Sigma_{j=2}^{5} b_{Sj} D_{Sj}$. Groups 4 and 5 for the pixel groups were aggregated together so that the terms $\Sigma_{j=2}^{5} b_{Pj} D_{Pj}$ were replaced by $\Sigma_{j=2}^{4} b_{Pj} D_{Pj}$. Brands 5 and 11 had only sales of 4 and 3 units respectively over the two years in our sample so these brands were aggregated together with Brand 3, another low sales brand. Thus the terms $\Sigma_{j=2}^{11} b_{Bj} D_{Bj}$ on the right hand side of (58) were replaced by the terms $\Sigma_{j=2}^{9} b_{Bj} D_{Bj}$. The CPU cells 4, 5, 6 and 10 had a small number of observations so D_{U4} was replaced by $D_{U4} + D_{U5} + D_{U6} + D_{U10}$. Finally, even after making these reductions in the number of characteristic dummy variables, it turned out that occasionally, one or more of the consolidated characteristic dummy variable vectors in the 23 bilateral hedonic regressions was equal to a vector of zeros. These vectors were dropped from the applicable adjacent period regression.

[27] Taking into account the reduction in the number of cells for the various characteristics, (58) became: $\ln P = \rho_2 D_2 + b_0 ONE + \Sigma_{j=2}^{6} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{5} b_{Sj} D_{Sj} + \Sigma_{j=2}^{4} b_{Pj} D_{Pj} + b_{H2} D_{H2} + \Sigma_{j=2}^{9} b_{Bj} D_{Bj} + \Sigma_{j=2}^{7} b_{Uj} D_{Uj} + \Sigma_{j=2}^{7} b_{Wj} D_{Wj} + e$.

sold in month 3 and the value 0 for the products sold in month 2. Once ρ_3^* was estimated, we defined $\pi_3^* \equiv \exp[\rho_3^*]$ and the period 3 price level as $P^{3*} \equiv \pi_3^* P^{2*}$.

The above procedure was continued until we reached the final bilateral regression that used only the log product prices for products that were sold in months 23 and 24. The final bilateral hedonic regression gave us an estimate for ρ_{24}^*. Once ρ_{24}^* was estimated, we defined $\pi_{24}^* \equiv \exp[\rho_{24}^*]$ and the period 24 price level was defined as $P^{24*} \equiv \pi_{24}^* P^{23*}$. The *Adjacent Period Time Dummy* (Unweighted) *Characteristics Price Index* for month t, P_{ATDC}^t, was defined as follows:

$$(59)\ P_{ATDC}^t \equiv P^{t*}\ ; \qquad t = 1,\ldots,24.$$

The price index defined by (59) is not satisfactory because it does not take into account the economic importance of each product. The economic importance of product n sold in period t can be taken into account in the 23 bilateral regressions of the form given by (58) by multiplying the log price $\ln p_{tn}$ that appears in any of these bilateral hedonic regressions by the square root of the corresponding expenditure share $s_{tn}^{1/2}$. The term $s_{tn}^{1/2}$ is also applied to the corresponding components of the various dummy variable vectors that appear on the right hand sides of the estimating equations of the form given by (58). With the application of these multiplicative factors on both sides of the various estimating equations, we again obtain estimates for the logarithms of the various bilateral time dummy coefficients $\rho_2^*, \rho_3^*, \ldots, \rho_{24}^*$. Once these new estimates have been obtained, we took the exponentials of them to obtain the sequence of price levels π_t^* for $t = 2,3,\ldots,24$. Now follow the same steps as were made in the paragraphs above definitions (59) in order to define the *Weighted Adjacent Period Time Dummy Characteristics Price Index* for month t, P_{WATDC}^t, for $t = 1,2,\ldots,24$. This index along with its unweighted (or equally weighted) counterpart index, P_{ATDC}^t, are listed in Table A3 in Appendix A. For comparison purposes, Table A3 also lists the single regression weighted and unweighted Time Dummy Characteristics price indexes, P_{WTDC}^t and P_{TDC}^t, as well as the simple average and unit value price indexes, P_A^t and P_{UV}^t. See Chart 2 below for plots of the indexes listed in Table A3.

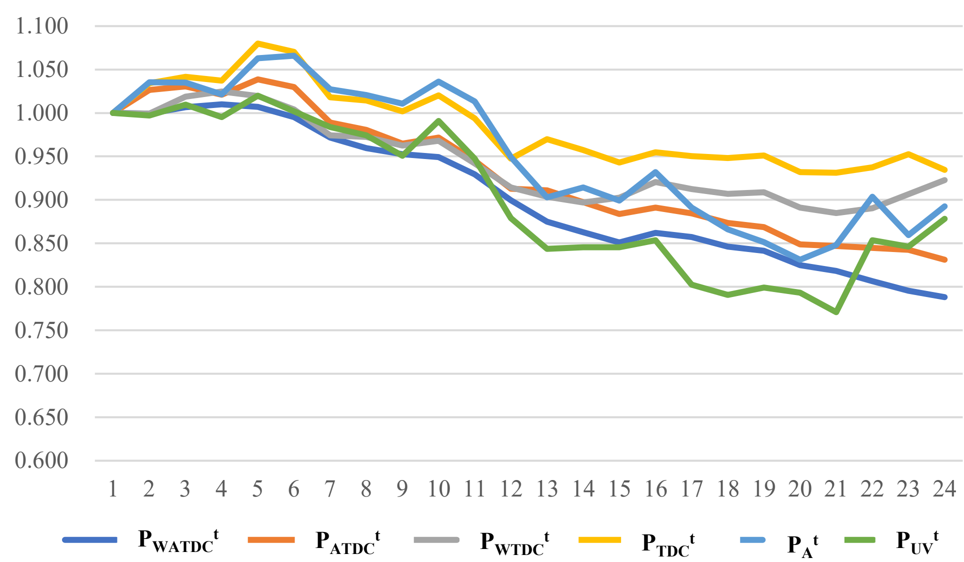

Chart 2. Sample Wide and Adjacent Period Time Dummy Characteristics Price Indexes.

Our new Adjacent Period Characteristics Price Indexes, P_{WATDC}^t and P_{ATDC}^t, finish well below their single regression counterpart indexes, P_{WTDC}^t and P_{TDC}^t when t = 24. More importantly, the new indexes finish below the Average Price index P_A^{24} and the Unit Value index P_{UV}^{24}, so that there was some positive quality improvement in laptops over our sample period. Thus the new adjacent period indexes are more plausible than the corresponding single regression based indexes.

Looking at the effects of weighting, it can be seen that the adjacent period equally weighted characteristics index P_{ATDC}^t finishes 4.3 percentage points above its weighted counterpart P_{WATDC}^t for t = 24 and on average, P_{ATDC}^t is 2.6 percentage points above the average for P_{WATDC}^t. Since this equally weighted index gives too much weight to unrepresentative products, we prefer the Weighted Adjacent Period Time Dummy Characteristics Index P_{WATDC}^t.

Here are some of the advantages and disadvantages of the Weighted Adjacent Period Time Dummy Characteristics indexes P_{WATDC}^t over the (sample wide) Weighted Time Dummy Characteristics indexes P_{WTDC}^t:

- The adjacent period indexes fit the data much better since each bilateral regression estimates a new set of quality adjustment parameters whereas the panel regression approach fixes the quality adjustment parameters over the entire window of observations.
- The adjacent period methodology that allows the quality adjustment parameters to change every month means that purchasers may not have stable consistent preferences over time and some economists may object to the resulting inconsistency of these indexes.
- The adjacent period indexes are chained indexes. If there are large fluctuations in the monthly product prices and quantities, then there is a danger that these indexes may be subject to the *chain drift problem*. Since there are large fluctuations in monthly prices and quantities in our data, there is some danger that our adjacent period indexes may be subject to some downward chain drift. We will revisit this point in section 7 below.

It is well known that missing characteristics can have a material effect on the price index.[28] A model that includes all possible product characteristics[29] is the Time Product Dummy model presented in section 2. Thus in the following section, we will consider weighted and unweighted time product dummy hedonic regression models.

5. Time Product Dummy Regression Models.

The Weighted Time Product Dummy least squares minimization problem was defined by (20). To obtain a unique solution to this problem, we added the normalization $\rho_1 = 0$. The corresponding Unweighted Time Product Dummy least squares minimization problem is defined by (20) with all expenditure shares s_{tn} set equal to 1.

---

[28] We found this to be the case as we worked through the various hedonic regression models described in Appendix A.

[29] There may be external environmental factors (that change over time) which affect the utility to purchasers of the products in scope. We are assuming that purchasers of the products in scope have preferences that are separable from other products which can only be a rough approximation to reality. Also, the "newness" or "oldness" of a product may affect purchaser utility.

In order to set up the unweighted regression problem for our particular application, we make use of the vectors of *time dummy* variables, $D_1, \ldots, D_{24}$ and the 366 *product dummy* variable vectors of dimension 2639, $D_{J1}, \ldots, D_{J366}$. Define the vector of the logarithms of observed laptop prices as lnP as was done in the previous section. Then the (sample wide) unweighted Time Product Dummy regression model can be expressed as the vector equation (60) where the unknown parameters are the log price levels $\rho_2, \rho_3, \ldots, \rho_{24}$ and the 366 product log quality adjustment factors $\beta_1, \beta_2, \ldots, \beta_{366}$:

$$(60)\ \ln P = \Sigma_{t=2}^{24}\ \rho_t D_t + \Sigma_{k=1}^{366}\ \beta_k D_{Jk} + e.$$

The R^2 for the above regression turned out to be 0.9836. Thus the fit for the model defined by (60) is substantially better than the fit for the unweighted characteristics regression defined by (57) in the previous section, which was 0.8926. We set ρ_t^* equal to one. The estimated ρ_t^* were exponentiated and the sequence of the $\pi_t^* \equiv \exp[\rho_t^*]$ are the *Time Product Dummy Price Indexes* P_{TPD}^t, which are listed in Table A4 in Appendix A.

To obtain the Weighted Time Product Dummy Price Indexes, multiply the vectors on both sides of (60) (excluding the error vector e) by the vector of positive square roots of the month by month expenditure shares s_{tn} on the products which were purchased in each period. The resulting linear regression in the same parameters $\rho_2, \rho_3, \ldots, \rho_{24}$ and $\beta_1, \beta_2, \ldots, \beta_{366}$ was run and the R^2 for this weighted time product dummy regression turned to be 0.9840. Again, set ρ_1^* equal to one. The estimated ρ_t^* were exponentiated and the new sequence of the $\pi_t^* \equiv \exp[\rho_t^*]$ are the *Weighted Time Product Dummy Price Indexes* P_{WTPD}^t which are listed in Table A4 and plotted on Chart 3 below.

The index P_{WTPD}^t is not a real time index. In order to obtain real time indexes, we can calculate adjacent period time product dummy regressions.

To start the adjacent period methodology, we use the price data for products n that were sold in months 1 and 2. Define S(1,2) as the set of products that were purchased in months 1 and 2. The counterpart regression to the unweighted time product dummy hedonic regression defined by (60) that links the prices of months 1 and 2 is the following regression model:

$$\begin{aligned}(61)\ \ln P^* &= \rho_2 D_2^* + \Sigma_{k=1}^{366}\ \beta_k D_{Jk}^* + e^{t*} \\ &= \rho_2 D_2^* + \Sigma_{k\in S(1,2)}\ \beta_k D_{Jk}^* + e^{t*}\end{aligned}$$

where the new log price vector $\ln P^*$, the new month 2 time dummy vector D_2^* and the new product dummy vectors $D_{J1}^*, \ldots, D_{J366}^*$ are restricted to the observations that correspond to the products n that were sold in periods 1 and 2. The first vector equation in (61) cannot be implemented using standard econometric packages because due to rapid product turnover, most of the product dummy variable vectors D_{Jk}^* will be vectors of zeros. Thus the second line in (61) sums over the products that actually sold in periods 1 and 2.[30] Using the results of the latter regression, we defined $P_{ATPD}^1 = 1$ and $P_{ATPD}^2 = \exp[\rho_2^*]$, the exponential of the estimated time dummy parameter ρ_2^*.

[30] It turned out to be somewhat difficult to go from line 1 in (79) to line 2 in (79). However, programs were constructed that overcome these difficulties.

The second bilateral regression has the form $\ln P^{**} = \rho_3 D_3^{**} + \Sigma_{k\in S(2,3)} \beta_k D_{Jk}^{**} + e^{t**}$ where the new vectors $\ln P^{**}$, $\rho_3 D_3^{**}$, D_{Jk}^{**} and e^{t**} are restricted to the products that were sold in periods 2 and 3. The estimated ρ_3^* was exponentiated and P_{ATPD}^3 was defined as P_{ATPD}^2 times $\pi_t^* \equiv \exp[\rho_3^*]$. A similar bilateral regression was run using the price data for periods 3 and 4 and the above process was continued. In all, 23 unweighted bilateral time product dummy variable regressions were run, the estimated ρ_t^* were converted into π_t^* and the π_t^* were chained into the Adjacent Period Time Product Dummy Price Indexes P_{ATPD}^t for t = 2, 3, …, 24. These indexes are listed in Table A4.

As in the previous section, to obtain Weighted Adjacent Period Time Product Dummy Price Indexes, P_{WAPD}^t we took the 23 bilateral regressions that were used to form the unweighted indexes and multiplied the dependent and independent variables in each of these regressions by the square root of the appropriate expenditure share.[31]

Table A4 in Appendix A lists the Adjacent Period Weighted and Unweighted Time Product Dummy price indexes, P_{WATPD}^t and P_{ATPD}^t, as well as the simple average and unit value price indexes, P_A^t and P_{UV}^t. Chart 3 below plots the indexes listed in Table A4.

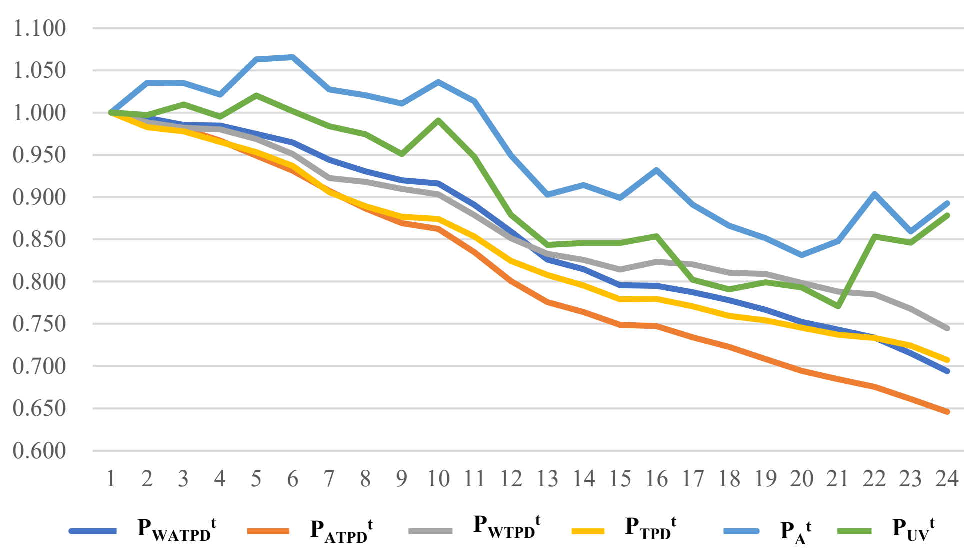

Chart 3. Sample Wide and Adjacent Period Weighted and Unweighted Time Product Dummy Indexes

There are large differences between the weighted and unweighted Time Product Dummy price indexes with the unweighted indexes generating lower rates of laptop inflation. As usual, we prefer the weighted estimates over their unweighted counterparts due to the unrepresentative nature of the unweighted indexes. However, the adjacent period indexes, P_{ATPD}^t and P_{WATPD}^t end up well below their panel data counterpart indexes, P_{TPD}^t and P_{WTPD}^t. Since there are large month to month fluctuations in laptop prices and quantities, it seems likely that the adjacent period indexes suffer from a chain drift problem. In the following section,

[31] The R^2 for the 23 bilateral Weighted Time Product Dummy regressions were as follows: 0.9993, 0.9985, 0.9979, 0.9988, 0.9991, 0.9988, 0.9991, 0.9976, 0.9987, 0.9980, 0.9980, 0.9985, 0.9974, 0.9980, 0.9989, 0.9993, 0.9987, 0.9989, 0.9980, 0.9986, 0.9990, 0.9988 and 0.9970. Needless to say, these regression fits are very good.

we will calculate some traditional indexes as well as Predicted Share Price Similarity linked indexes which may reduce chain drift.

6. Traditional Indexes and Similarity Linked Price Indexes for Laptops.

The indexes defined in the previous sections that made use of 23 adjacent period regressions were *chained* indexes; i.e., the index constructed for month t compared the prices for month t with the prices for month t − 1. However, it is not the case that all bilateral comparisons of prices between two months are equally accurate: if the relative prices for matched products in months r and t are very similar, then the Laspeyres and Paasche price indexes will be very close to each other and hence it is likely that the "true" price comparison between these two periods (using the economic approach to index number theory[32]) will be very close to the bilateral Fisher index that compares prices between the two periods under consideration. In particular, if the two price vectors are exactly proportional, then we would like the price index between these two months to be equal to the factor of proportionality (even if the associated quantity vectors are not proportional) and the direct Fisher price index between these two periods satisfies this proportionality test. This test suggests that a more accurate set of price indexes could be constructed if a bilateral comparison of prices was made between the two months that have the most *similar relative price* structures.[33] The *Predicted Share* method of linking months with the most similar structure of relative prices will be explained under the assumption that it is necessary to construct a price index P^t in real time.[34]

As a preliminary step, the price and quantity data that are listed in the Appendix need to be reorganized into 24 price and quantity vectors of dimension 366, $p^t \equiv [p_1{}^t,p_2{}^t,\ldots,p_{366}{}^t]$ and $q^t \equiv [q_1{}^t,q_2{}^t,\ldots,q_{366}{}^t]$, for $t = 1,\ldots,24$. If product k is not purchased during month t, then we set $p_k{}^t = q_k{}^t = 0$. For months r and t, define the set of products k that are present in both months as S(r,t). The *matched model Laspeyres and Paasche indexes*, $P_L(r,t)$ and $P_P(r,t)$, that relate the prices of month t to month r are defined as follows:

$$(62)\ P_L(r,t) \equiv \Sigma_{k\in S(r,t)}\, p_k{}^t q_k{}^r/\Sigma_{k\in S(r,t)}\, p_k{}^r q_k{}^r\ ; \qquad 1 \le r, t \le 24;$$
$$(63)\ P_P(r,t) \equiv \Sigma_{k\in S(r,t)}\, p_k{}^t q_k{}^t/\Sigma_{k\in S(r,t)}\, p_k{}^r q_k{}^t\ ; \qquad 1 \le r, t \le 24.$$

Note that the prices of the matched models for month t are in the numerators of definitions (62) and (63) and the corresponding prices of the matched models for month r in the denominators of definitions (62) and (63). The *matched model Fisher index* that relates the prices of month t to the prices of month r is defined as the geometric mean of $P_L(r,t)$ and $P_P(r,t)$:[35]

$$(64)\ P_F(r,t) \equiv [P_L(r,t)P_P(r,t)]^{1/2}\ ; \qquad 1 \le r, t \le 24.$$

The components $s_k{}^t$ of the 24 vectors of month t expenditure shares on the 366 products, $s^t \equiv [s_1{}^t,s_2{}^t,\ldots,s_{366}{}^t]$, are defined as follows:

$$(65)\ s_k{}^t \equiv p_k{}^t q_k{}^t/p^t\cdot q^t\ ; \qquad t = 1,\ldots,24\ ;\ k = 1,\ldots,366$$

[32] See Diewert (1976) for the relationship of the Fisher index to the economic approach to index number theory.
[33] In the context of making comparisons of prices across countries, the method of linking countries with the most similar structure of relative prices has been pursued by Hill (1997) (1999a) (1999b) (2009), Hill and Timmer (2006), Diewert (2009) (2013) (2023) and Hill, Rao, Shankar and Hajargasht (2017). Hill (2001) (2004) also pursued this similarity of relative prices approach in the time series context.
[34] This method is explained more fully in Diewert (2023).
[35] Note that there are 576 = 24x24 matched model bilateral Fisher (1922) indexes $P_F(r,t)$.

where the inner product of the vectors p^t and q^t is defined as $p^t \cdot q^t \equiv \Sigma_{k=1}^{366} p_k^t q_k^t$.

The choice of a measure of relative price similarity plays a key role in the similarity linking methodology. Various measures of the similarity or dissimilarity of relative price structures have been proposed by Allen and Diewert (1981), Kravis, Heston and Summers (1982; 104-106), Hill (1997) (2009), Sergeev (2001) (2009), Hill and Timmer (2006), Aten and Heston (2009) and Diewert (2009) (2023). A problem with most measures of relative price similarity is that they are not well defined if some products are missing. The following *Predicted Share measure of relative price dissimilarity*, $\Delta(p^r,p^t,q^r,q^t)$, is well defined even if some product prices in the two periods being compared are equal to zero:[36]

$$(66)\ \Delta(p^r,p^t,q^r,q^t) \equiv \Sigma_{k=1}^{366} [s_k^t - (p_k^r q_k^t/p^r \cdot q^t)]^2 + \Sigma_{k=1}^{366} [s_k^r - (p_k^t q_k^r/p^t \cdot q^r)]^2\ ; \qquad 1 \le r, t \le 24.$$

We require that $p^r \cdot q^t > 0$ for r = 1,...,24 and t = 1,...,24 in order for $\Delta(p^r,p^t,q^r,q^t)$ to be well defined for any pair of periods, r and t. Since the two summations on the right hand side of (66) are sums of squared terms, we see that $\Delta(p^r,p^t,q^r,q^t) \ge 0$. If $\Delta(p^r,p^t,q^r,q^t) = 0$, then the price vectors for months r and t are proportional. The closer $\Delta(p^r,p^t,q^r,q^t)$ is to 0, the closer prices are to being proportional between the two months. If prices are proportional for the two months, then any acceptable price index between the two months should equal the proportionality factor. If $p^t = \lambda p^r$ for some positive factor of proportionality λ, then the matched model Fisher index $P_F(r,t)$ defined by (64) will equal λ. Another very important property of the measure of relative price similarity defined by (66) is that the Predicted Share measure *penalizes* a lack of product matching across the two months r and t. Thus if the matched prices for months r and t are equal but there are some products that are only available in one of the two periods under consideration, then $\Delta(p^r,p^t,q^r,q^t)$ will be greater than 0.

The 24 by 24 matrix of Predicted Share measures of relative price similarity for our laptop data, $\Delta(p^r,p^t,q^r,q^t)$, are listed in Table A5 in Appendix A.

Table A5 can be used to construct the relative price similarity linked Predicted Share Price index, P_S^t, for t = 1,…,24. See Appendix A for the details of the construction.

It turns out that the relative price similarity linked indexes P_S^t, the Fisher chained maximum overlap indexes P_{FCH}^t and the Adjacent Period Weighted Time Product Dummy price indexes P_{WATPD}^t are all extremely close to each other for our laptop data set.[37] It can also the case that the chained Laspeyres and Paasche indexes, P_{LCH}^t and P_{PCH}^t, are very close to P_{FCH}^t for our particular data set. Table A6 in Appendix A lists these indexes and they are plotted in Chart 4 below.

[36] See Diewert (2023) for the axiomatic properties of this measure.
[37] The bilateral link Fisher indexes that were used to construct the similarity linked indexes were equal to adjacent period matched model bilateral Fisher indexes for 20 out of 23 bilateral links. This explains why the chained Fisher index, P_{FCH}^t, are so close to the relative price similarity linked indexes, P_S^t. The bilateral link indexes used to construct the Weighted Adjacent Period Time Product Dummy indexes, P_{WATPD}^t, are also numerically close to the corresponding matched model bilateral Fisher index, which explains why P_{WATPD}^t is close to P_{FCH}^t; see Diewert (2005b) on this approximation.

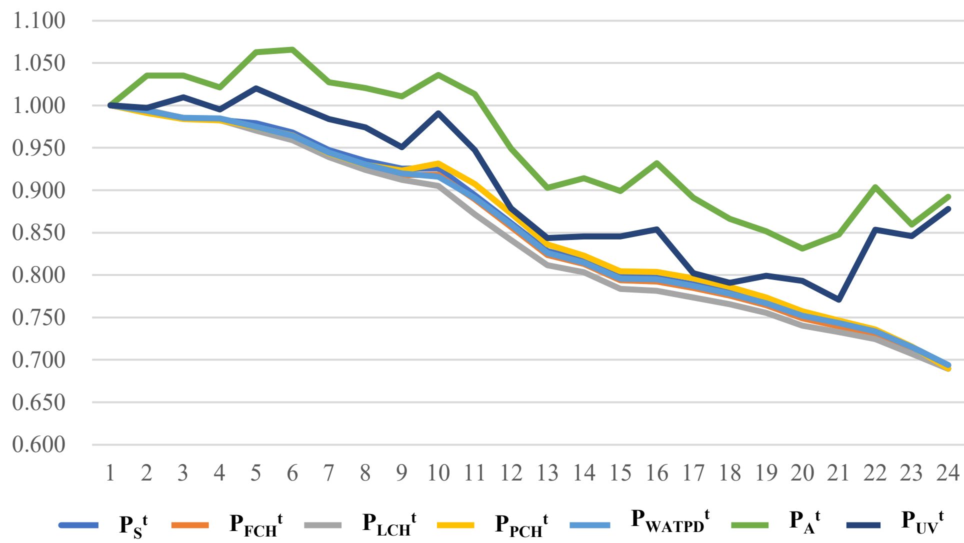

Chart 4: The Predicted Share Similarity Linked Index and Other Comparison Price Indexes

The chained Fisher indexes have the advantage that no complex hedonic regression methodology is required to implement these indexes. They are also relatively easy to explain to the public. However, in many applications where products go on sale or there are strongly seasonal products, chained Fisher indexes are likely to be subject to some chain drift. Since the similarity linked indexes are so close to the Fisher chained indexes, it is likely that they are also subject to some chain drift.[38] Thus it is likely that P_S^t, P_{FCH}^t, P_L^t, P_P^t and P_{WATPD}^t all suffer from some chain drift. We will address this potential chain drift problem in the following section.

We conclude this section by considering *Hedonic Imputation Indexes* and their relationship to Fisher indexes when each product has its own unique characteristic. Hedonic imputation indexes run a hedonic regression using the data for one period and then they use the results of this regression to impute prices or quantities in a subsequent period.[39]

We will first consider the case when there are N products that are available in both periods under consideration and there is a separate characteristic for each product. Thus the hedonic imputation regression for period 1 is given by (13) in the paper for a Weighted Time Product Dummy regression when there are T periods. When the number of periods T becomes T = 1, (13) becomes (67) below:

(67) $\min_{\rho,\beta} \{\Sigma_{n=1}^{N} s_{1n}[\ln p_{1n} - \rho_1 - \beta_{1n}]^2\}$.

[38] Thus the fact that the similarity linked indexes satisfy Walsh's Multiperiod Identity Test is not sufficient to eliminate possible chain drift. In order to be certain to eliminate chain drift, we need the indexes to satisfy the transitivity or circularity test in index number theory. The problem with the predicted share methodology is that it tends to pick the previous period as the preferred period to link the current period index to an earlier most similar in structure period. Fox, Levell and O'Connell (2023) noted this problem in their study of the chain drift problem for UK prices.
[39] The method dates back to Court (1939; 108) and Griliches (1971; 6). See also Diewert (2003b) for an extensive discussion of some of the issues.

Obviously, not all of the parameters in (67) can be uniquely determined so we set $\rho_1 = 0$ and then the β_{1n} are identified and we obtain the following solution to the weighted least squares problem defined by (67):[40]

(68) $\rho_1 = 0$; $\beta_{1n} = \ln p_{1n}$; $\quad n = 1,\ldots,N.$

Exponentiating the β_{1n} gives us the α_{1n} parameters for the purchasers' period 1 utility function:

(69) $\alpha_{1n} \equiv \exp[\beta_{1n}] = p_{1n}$; $\quad n = 1,\ldots,N.$

Using equations (19) give us estimates for aggregate quantity (or utility) for period t, Q^{t*}, and the aggregate price level for period t, P^{t*}, for t = 1,2, using the preferences estimated using the period 1 data. Let $p^1 \equiv [p_{11},\ldots,p_{1N}]$ be the observed period 1 price vector and let $\alpha^1 \equiv [\alpha_{11},\ldots, \alpha_{1N}] = p^1$ be the vector of period 1 marginal utility parameters. Thus we obtain the following period 1 and period 2 aggregates using the period 1 hedonic regression:

(70) $Q^{1*} \equiv \alpha^1\cdot q^1 = p^1\cdot q^1$; $P^{1*} \equiv p^1\cdot q^1/\alpha^1\cdot q^1 = p^1\cdot q^1/p^1\cdot q^1 = 1$; $Q^{2*} \equiv \alpha^1\cdot q^2 = p^1\cdot q^2$; $P^{2*} \equiv p^2\cdot q^2/\alpha^1\cdot q^2 = p^2\cdot q^2/p^1\cdot q^2$.

Thus the price index relating the prices of period 2 to the prices of period 1, P^{2*}/P^{1*}, can be calculated using (70):

(71) $P^{2*}/P^{1*} = p^2\cdot q^2/p^1\cdot q^2 = P_P$

where P_P is the ordinary Paasche price index between periods 2 and 1. Thus the period 1 hedonic regression in the special case where each product is given a separate characteristic leads to the regular Paasche price index, provided that there are no missing products in each period.

The hedonic imputation regression for period 2 is given by (72) below:

(72) $\min_{\rho,\beta} \{\Sigma_{n=1}^N s_{2n}[\ln p_{2n} - \rho_2 - \beta_{2n}]^2\}$.

Again, not all of the parameters in (72) can be uniquely determined so we set $\rho_2 = 0$ and then the β_{2n} are identified and we obtain the following solution to the weighted least squares problem defined by (72):

(73) $\rho_2 = 0$; $\beta_{2n} = \ln p_{2n}$; $\alpha_{2n} \equiv \exp[\beta_{2n}] = p_{2n}$; $\quad n = 1,\ldots,N.$

Using equations (19) gives us estimates for aggregate quantity (or utility) for period t, Q^{t**}, and the aggregate price level for period t, P^{t**}, for t = 1,2, using the preferences estimated using the period 2 data. Let $p^2 \equiv [p_{21},\ldots,p_{2N}]$ be the observed period 2 price vector and let $\alpha^2 \equiv [\alpha_{21},\ldots, \alpha_{2N}] = p^2$ be the vector of period 2 marginal utility parameters. Thus we obtain the following period 1 and period 2 aggregates using the period 2 hedonic regression:

(74) $Q^{1**} \equiv \alpha^2\cdot q^1 = p^2\cdot q^1$; $P^{1**} \equiv p^1\cdot q^1/\alpha^2\cdot q^1 = p^1\cdot q^1/p^2\cdot q^1$;

$Q^{2**} \equiv \alpha^2\cdot q^2 = p^2\cdot q^2$; $P^{2**} \equiv p^2\cdot q^2/\alpha^2\cdot q^2 = p^2\cdot q^2/p^2\cdot q^2 = 1$.

[40] Since the fit in the regression that corresponds to the least squares problem (67) is perfect, we obtain the same solution if we use the unweighted least squares minimization problem.

Thus the price index relating the prices of period 2 to the prices of period 1, P^{2**}/P^{1**}, using the preferences of period 2 can be calculated using (74):

(75) $P^{2**}/P^{1**} = 1/[p^1\cdot q^1/p^2\cdot q^1] = p^2\cdot q^1/p^1\cdot q^1 = P_L$

where P_L is the ordinary Laspeyres price index between periods 2 and 1. Thus the period 2 hedonic regression in the special case where each product is given a separate characteristic leads to the regular Laspeyres price index, provided that there are no missing products in each period.

The two equally valid indexes of price change, P_L and P_P, should be averaged in order to obtain a final estimate of price change. Taking the geometric average of P_L and P_P leads to the Fisher index which satisfies more tests than competing indexes. Thus we have shown that in the case where each product has its own characteristic (and there are no missing observations in the two periods being compared), *hedonic imputation leads to the Fisher index using our economic approach to the measurement of quality change*.

What happens if there are missing observations? The weighted least squares minimization problems (67) and (72) are replaced by the following two weighted least squares minimization problems:

(76) $\min_{\rho,\beta} \{\Sigma_{n\in S(t)} s_{tn}[\ln p_{tn} - \beta_{tn}]^2\}$; $t = 1,2$.

In (76), S(t) is the set of products that are available in period t. It can be seen that our previous methodology breaks down when there are products that are available in only one of the two periods. The period 1 regression will give us estimates for β_{1n} and hence estimates $\alpha_{1n} = \exp[\beta_{1n}]$ for products n that are available in period 1 but we will have no estimated α_{1k} for products k that are available only in period 2. Thus all we can do is restrict n in each period to products that are present in both periods being compared. If we do this, *we will end up with maximum overlap Fisher indexes*.

7. Expanding Window Weighted Time Product Dummy Indexes.

We can determine whether a given price index suffers from a chain drift problem by comparing it to a "reasonable" index that does not suffer from chain drift. But how exactly can we find a "reasonable" target index that is not subject to chain drift?

To answer this question, consider the sample wide Weighted Time Product Dummy price index, P_{WTPD}^t, that was defined in section 5. The linear regression that defined this index gave rise to estimated parameters ρ_2^*, ρ_3^*, …, ρ_{24}^* and β_1^*, β_2^*, …, β_{366}^* and we exponentiated the ρ_t^* to obtain estimated period t price levels, $\pi_t^* = P_t^* = P_{WTPD}^t$, for t = 1,…,24. In this section, we exponentiate the β_n^* to obtain the linear utility function parameters α_n^* for n = 1,…,366. Let $\alpha^* \equiv [\alpha_1^*,…, \alpha_{366}^*]$. Now use equations (19) to define the month t aggregate quantity levels Q^{t**} and the Implicit Weighted Time Product Dummy price levels P^{t**}; i.e., define:

(77) $Q^{t**} \equiv \alpha^*\cdot q^t$; $P^{t**} \equiv p^t\cdot q^t/\alpha^*\cdot q^t$;   t = 1,…,24.

It can be seen that the Q^{t**} are transitive; i.e., $Q^{3**}/Q^{1**} = (Q^{2**}/Q^{1**})(Q^{3**}/Q^{2**})$. Expenditures $e^t \equiv p^t\cdot q^t$ are also transitive so that $e^{3**}/e^{1**} = (e^{2**}/e^{1**})(e^{3**}/e^{2**})$. Hence the indirectly determined price levels P^{t**} are also transitive. Define the month t *Implicit Weighted Time Product Dummy price index* P_{IWTPD}^t as the following normalization of the month t price level P^{t**}:

(78) $P_{IWTPD}{}^t \equiv P^{t**}/P^{1**}$; $t = 1,\ldots,24.$

The implicit price index $P_{IWTPD}{}^t$ is also *transitive*; i.e., it satisfies the *circularity test* and thus it is free from chain drift. It is not always a "reasonable" target index (since the products in scope may not be close substitutes or there may not be a sufficient amount of product matching over time) but in our context, it probably is a reasonable index, since the R^2 for the underlying hedonic regression was 0.9840. Thus the assumption of common constant linear preferences by all purchasers is satisfied to a "reasonable" degree of approximation. The direct and indirect price indexes, $P_{WTPD}{}^t$ and $P_{IWTPD}{}^t$ generated by this weighted panel regression are listed in Table 3 below and plotted in Chart 5. There are only small differences between these two indexes.

However, $P_{IWTPD}{}^t$ is not a real time index. Thus we have to find a suitable approximation to this target index that can be calculated in real time.

Our first suggested approximation constructs *Expanding Window Weighted Time Product Dummy price indexes*, $P_{EW}{}^t$, for $t = 1, 2, \ldots, 24$.[41] We show how this can be done.

Step 1: define $P_{EW}{}^1 \equiv 1$.

Step 2: Run the weighted Time Product Dummy regression that links months 1 and 2 as in section 5 above. Thus run the weighted version of equations (61) to get estimates for the β_k that correspond to products that were purchased in months 1 or 2. For products k that were not purchased in months 1 or 2, set $\beta_k{}^* = 0$. Exponentiate these $\beta_k{}^*$ to get estimated $\alpha_k{}^*$ for $k = 1,\ldots,366$. Now use definitions (77) for $t = 1,2$ to define P^{1**} and P^{2**} indirectly. Define $P_{EW}{}^2 \equiv P^{2**}/P^{1**}$. Note that the quantities associated with the missing prices for months 1 and 2 are equal to zero.

Step 3: Run a weighted Time Product Dummy regression using the data for months 1, 2 and 3 [42] to get estimates for the β_k that correspond to products that were purchased in months 1, 2 or 3. For products k that were not purchased in months 1, 2 or 3, set $\beta_k{}^* = 0$. Exponentiate these $\beta_k{}^*$ to get estimated $\alpha_k{}^*$ for $k = 1,\ldots,366$. Now use definitions (77) for $t = 1$ and 3 to define P^{1**}, $P^{2**}$and P^{3**} indirectly. Define $P_{EW}{}^3 \equiv P^{3**}/P^{1**}$. Note that P^{1**}, $P^{2**}$and P^{3**} are fully transitive price levels based on the information that is available at the end of month 3. However, $P_{EW}{}^3$ will not (in general) be equal to $P_{IWTPD}{}^3$ since $P_{IWTPD}{}^3$ is based on more complete data that is available at the end of month 24 (and thus $P_{IWTPD}{}^3$ uses more data to estimate the underlying linear utility function).

Step 4: Run a weighted Time Product Dummy regression using the data for months 1, 2, 3 and 4 [43] to get estimates for the β_k that correspond to products that were purchased in months 1, 2, 3 and 4. For products k that were not purchased in months 1, 2, 3 or 4, set $\beta_k{}^* = 0$. Exponentiate these $\beta_k{}^*$ to get estimated $\alpha_k{}^*$ for

---

[41] Using an expanding window (instead of using a rolling window) to construct multilateral indexes was suggested (and implemented) by Chessa (2016) (2021). The idea of using an ever expanding window was suggested by Diewert (2023) in the context of the predicted share price similarity methodology.

[42] The unweighted counterpart to the weighted regression has the form $\ln P^* = \rho_2 D_2{}^* + \rho_3 D_3{}^* + \Sigma_{k\in S(1,2,3)}\, \beta_k D_{Jk}{}^* + e^{t*}$ where S(1,2,3) is the set of products which were sold in months 1, 2 or 3.

[43] The unweighted counterpart to the weighted regression has the form $\ln P^* = \rho_2 D_2{}^* + \rho_3 D_3{}^* + \rho_4 D_4{}^* + \Sigma_{k\in S(1,2,3,4)}\, \beta_k D_{Jk}{}^* + e^{t*}$ where S(1,2,3,4) is the set of products which were purchased in months 1, 2, 3 or 4.

k = 1,…,366. Now use definitions (77) for t = 1 and 4 to define P^{1**}, P^{2**}, P^{3**} and P^{4**}. Define $P_{EW}^{4} \equiv P^{4**}/P^{1**}$.[44]
…

Step 24: The final step simply sets P_{EW}^{24} equal to the month 24 Implicit Weighted Time Product Dummy price index P_{IWTPD}^{24}.

The *Expanding Window price indexes*, P_{EW}^{t}, are listed in Table 3 below and plotted on Chart 5. It can be seen that the Expanding Window price indexes P_{EW}^{t} are reasonably close to their transitive counterpart indexes, the Implicit Weighted Time Product Dummy indexes, P_{IWTPD}^{t}, and they are very close near the end of the sample period.

If the statistical agency is able to collect price and quantity data on the products in scope on a retrospective basis, then Modified Expanding Window price indexes P_{MEW}^{t} could be used.[45] To construct these indexes, start off with a window of 12 months of data and construct Implicit Weighted Time Product Dummy indexes for this 12 month window. Then simply switch over to the Expanding Window price indexes for months 13 to 24. Thus the resulting indexes will be real time indexes over months 13-24. The *Modified Expanding Window price indexes*, P_{MEW}^{t}, are also listed in Table 3 below and plotted on Chart 5. It can be seen that the Modified Expanding Window price indexes P_{MEW}^{t} are generally closer to their transitive counterpart indexes, the Implicit Weighted Time Product Dummy indexes, P_{IWTPD}^{t}, than the unmodified indexes P_{EW}^{t} for t = 2, 3, …, 12.

Table 3: Expanding Window, Modified Expanding Window and Weighted Time Product Dummy Price Indexes

Month	P_{EW}^{t}	P_{MEW}^{t}	P_{IWTPD}^{t}	P_{WTPD}^{t}
1	1.000	1.000	1.000	1.000
2	0.994	0.988	0.988	0.988
3	0.985	0.979	0.982	0.982
4	0.984	0.976	0.980	0.980
5	0.985	0.966	0.969	0.969
6	0.973	0.950	0.951	0.951
7	0.943	0.928	0.922	0.923
8	0.934	0.920	0.918	0.918
9	0.921	0.913	0.910	0.910
10	0.908	0.906	0.903	0.903
11	0.884	0.884	0.878	0.879
12	0.855	0.855	0.851	0.851
13	0.822	0.822	0.833	0.833

[44] Note that we always link our price level estimate for the last period in our current regression back to the price level in period 1 because the period 1 price level set equal to 1 is the "true" period 1 price level and never changes whereas our new price levels for periods 2 to t−1 are changing as we add another period of data to the regression. Thus we are following Krsinich (2016) in our choice of "best" linking period.
[45] Krsinich (2016; 401) noted that estimates for the quality adjustment parameters α_n are not reliably determined until the products have been present in the marketplace for several periods. This observation helps to explain why the Modified Expanding Window (EW) estimates will be more accurate than the simple EW estimates for the first few periods.

14	0.811	0.811	0.825	0.826
15	0.797	0.797	0.814	0.814
16	0.806	0.806	0.823	0.823
17	0.811	0.811	0.820	0.820
18	0.797	0.797	0.810	0.810
19	0.799	0.799	0.809	0.809
20	0.792	0.792	0.798	0.798
21	0.784	0.784	0.788	0.788
22	0.786	0.786	0.784	0.785
23	0.766	0.766	0.768	0.768
24	0.745	0.745	0.745	0.745
Mean	0.870	0.866	0.870	0.870

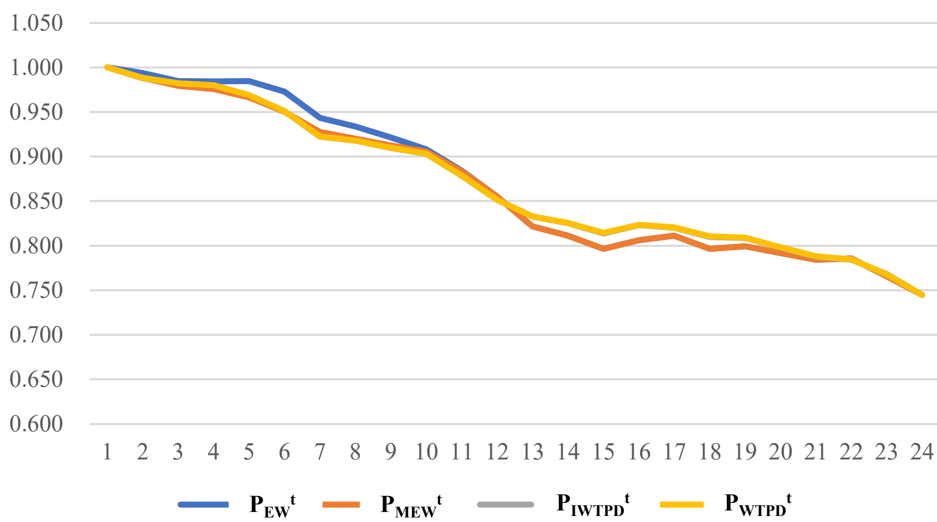

Chart 5: Expanding Window and Weighted Time Product Dummy Price Indexes

The Modified Expanding Window indexes P_{MEW}^t are in general closer to our target transitive indexes P_{IWTPD}^t indexes for months 1-12 than the Expanding Window indexes P_{EW}^t. Of course, P_{MEW}^t coincides with P_{EW}^t for months 13-24 by construction. The first 3 indexes listed in Table 3 all coincide at month 24. Finally, the directly estimated weighted TPD indexes, P_{WTPD}^t are slightly different from the final transitive indexes P_{IWTPD}^t.

At each month t, the current Weighted TPD generates our "best" estimates for the α_n parameters and the "best" transitive set of indirectly determined price levels for months 1 to t. However, the *fixity* (or no revisions) *constraint* on consumer price indexes prevents Statistical Offices from publishing new "improved" estimates of inflation for months 2 to t−1: all that can be done is to publish the new estimate for the period t price level. Thus there will be some differences or *biases* between the newly generated price levels for months 2 to t−1 and the historical published price levels for months 2 to t−1. These biases could be termed *fixity biases*. However, the newly estimated price levels will satisfy the circularity test for months 1 to t. Hence the period t price level should be free from any chain drift bias using these new price levels.

As the number of periods in the sample increase, the complexity of the Weighted TPD regression increases. Thus one might ask: when should we start dropping the prices that correspond to the early months in the sample? In situations *where there is a rapid product turnover*, what will happen is that the α_n estimates for products that were purchased at the beginning of the sample period stabilize. Moreover, the q_{tn} that correspond to these disappeared products will be 0 and so these obsolete products will no longer enter the quantity aggregate of the current month. Thus at some point, the ever expanding window can be replaced by a (probably long) rolling window.[46] This point will be reached when the current price level estimate for the expanding window regression for period t is very close to the corresponding rolling window estimate.

What are the limitations of the Expanding Window Weighted TPD model or more generally, of a TPD hedonic regression?

- New and existing products must compete in the marketplace for more than one period. It is this competition between products that will enable us to estimate the relative values of the products to purchasers.[47]
- The Time Product Dummy model relies on the assumption that purchasers have linear preferences over the products in scope, at least to some degree of approximation. This assumption implies that the products are highly substitutable. The validity of this assumption can be determined by looking at the R^2 for the TPD model (provided that there is an adequate amount of matching of products over time): if the R^2 is low, it may be best to turn to a model that allows for more flexible purchaser preferences such as Rolling Year GEKS or CCDI.[48]
- The Expanding Window Weighted TPD model does not allow for preference changes. The Rolling Window TPD model does allow for gradual preference changes at the cost of introducing some possible chain drift.

8. Conclusion.

The following tentative conclusions emerge from our study:

- When price and quantity data for the products in scope are available, it is best to use weighted hedonic regressions that take into account the economic importance of the products. We found substantial differences between our weighted and unweighted (or more accurately, equally weighted) hedonic regressions.
- The Time Dummy Characteristics approach to hedonic regressions did not work well for our particular example. This approach requires data on characteristics (which is expensive) and it is subject to the missing characteristics problem. We found that the indexes changed substantially as

[46] For an early application of the Rolling Window Weighted TPD model, see section 7 of Ivancic, Diewert and Fox (2009). For their particular scanner data set, the authors found little difference between their Rolling Window GEKS estimates and their Rolling Window TPD estimates. However, their monthly data series covered only 15 months.
[47] This limitation of the TPD methodology was recognized by Krsinich (2016; 400-401) and de Haan, Hendricks and Scholz (2021; 395). Again consider the extreme example where a new product enters the marketplace every period but exits after only one month. There is an extreme *lack of matching bias*.
[48] On the Rolling Year GEKS methodology, see Ivancic, Diewert and Fox (2011). On the Rolling Year CCDI methodology, see Diewert and Fox (2022) and Fox, Levell and O'Connell (2022) (2023). The CCDI multilateral price index is due to Caves, Christensen and Diewert (1982) and Inklaar and Diewert (2016).

we added additional characteristics to the regressions.[49] However, there are situations where the TPD approach does not work well and other approaches should be used.[50]

- There was a chain drift problem with all of our models that used chaining to link adjacent periods. The chain drift problem was not cured by the use of the multilateral predicted share method because most of the bilateral links chosen by the method were chain links.[51]
- A satisfactory solution to the chain drift problem for our example was provided by the use of the Expanding Window methodology explained in section 7. This method should work well for many product classes where substitution between the competing products is high and each product is available on the marketplace for a number of consecutive periods.

An interesting question for further research is a comparison of the Expanding Window Weighted TPD method explained above with the Expanding Window Geary (1958) Khamis (1970) index that was introduced by Chessa (2016) (2021). Both indexes are exact for linear preferences but the GK index can be implemented without econometric estimation.

There are many other problems for further research that could be explored such as determining what is the "best" approach to aggregation of microeconomic data at the individual product level: is it simple unit value aggregation, a Rolling Window method based on a bilateral superlative index like the Rolling Window GEKS and CCDI methods or is it a regression based approach like the Expanding Window Weighted Time Product Dummy method?

[49] Thus we tend to agree with Krsinich (2016) in her defense of the Time Dummy Product approach to hedonic regressions as opposed to de Haan, Hendricks and Scholz (2021), who favoured the Time Dummy Characteristics approach because the TPD approach led to too many parameters and an overfitting problem. If a product is only on the market for a few periods, then the quality adjustment parameter for that product will not be accurately estimated by the TPD method and in that sense, there will be overfitting. But if products are present for say five consecutive periods and the regression fit is high, then the TPD method should work well.

[50] The construction of residential and commercial property price indexes cannot use the TPD method because each property is unique in its location. If the property has a structure, then the structure changes over time due to depreciation and possible renovations so the property is unique over time as well. Thus each property at each point in time has its own unique characteristic and the TPD method fails. Thus the TPC model should be used in this context.

[51] The recent paper by Fox, Levell and O'Connell (2023) found the same result for many product classes.

18th Meeting of the Ottawa Group on Price Indices at Ottawa, Canada, May 13-15, 2024.

References.

Aizcorbe, A., 2014. A Practical Guide to Price Index and Hedonic Techniques. Oxford University Press, Oxford, UK.

Aizcorbe, A., Corrado, C., Doms, M., 2000. Constructing Price and Quantity Indexes for High Technology Goods. Industrial Output Section, Division of Research and Statistics, Board of Governors of the Federal Reserve System, Washington District of Columbia.

Allen, R.C., Diewert, W.E., 1981. Direct versus Implicit Superlative Index Number Formulae. Review of Economics and Statistics 63, 430–435. https://doi.org/10.2307/1924361.

Aten, B., Heston, A., 2009. Chaining methods for international real product and purchasing power comparisons: Issues and alternatives, in: Prasada Rao, D.S. (Ed.). Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications. Edward Elgar Publishing, Cheltenham UK, pp. 245–273.

Balk, B.M., 1980. A method for constructing Price indices for seasonal commodities. Journal of the Royal Statistical Society, Series A 143, 68–75. https://doi.org/10.2307/2981770.

Caves, D.W., Christensen, L.R., Diewert, W.E., 1982. Multilateral comparisons of output, input, and productivity using superlative index numbers. Economic Journal 92, 73–86. https://doi.org/10.2307/2232257.

Chessa, A.G., 2016. A new methodology for processing scanner Data in the Dutch CPI. EURONA. 1, 49–69.

Chessa, A.G., 2021. 'Extension of Multilateral Index Series Over Time: Analysis and Comparison of Methods', paper written for the 2021 Meeting of the Group of Experts on Consumer Price Indices. Geneva. https://unece.org/sites/default/files/2021-05/Session_1_Netherlands_Paper.pdf. (Accessed 7 May 2021).

Court, A.T., 1939. Hedonic Price indexes with automotive examples, in: The *Dynamics of Automobile Demand*. General Motors Corporation, New York, pp. 99–117.

de Haan, J., 2004a. The time dummy index as a special case of the imputation Törnqvist index, " paper presented at The Eighth Meeting of the International Working Group on Price Indices (the Ottawa Group), Helsinki, Finland.

de Haan, J., 2004b. Estimating quality-adjusted unit value indices: Evidence from scanner Data, Paper presented at the Seventh EMG Workshop, Sydney, Australia, December 12–14.

de Haan, J., 2010. Hedonic Price indexes: A comparison of imputation, time dummy and repricing methods. Jahrbücher für Nationökonomie und Statistik 230, 772–791.

de Haan, J., 2015. A framework for large scale use of scanner Data in the Dutch CPI, Paper Presented at the 14th Meeting of the Ottawa Group, Tokyo.

De Haan, J., Hendriks, R., Scholz, M., 2021. Price Measurement using Scanner Data: Time Product Dummy versus Time Dummy Hedonic Indexes. Review of Income and Wealth 67, 394–417. https://doi.org/10.1111/roiw.12468.

de Haan, J., Krsinich, F., 2014. Scanner Data and the treatment of quality change in nonrevisable Price indexes. Journal of Business and Economic Statistics 32, 341–358. https://doi.org/10.1080/07350015.2014.880059.

de Haan, J., Krsinich, F., 2018. Time dummy hedonic and quality-adjusted unit value indexes: Do they really differ? Review of Income and Wealth 64, 757–776. https://doi.org/10.1111/roiw.12304.

Diewert, E., 2005a. Weighted country product dummy variable regressions and index number formulae. Review of Income and Wealth 51, 561–570. https://doi.org/10.1111/j.1475-4991.2005.00168.x.

Diewert, W.E., 1974. Applications of duality theory, in: Intriligator, M.D., Kendrick, D.A. (Eds.), *Frontiers of Quantitative Economics*, Vol. II. North-Holland, Amsterdam, pp. 106–171.

Diewert, W.E., 1976. Exact and superlative index numbers. Journal of Econometrics 4, 115–145. https://doi.org/10.1016/0304-4076(76)90009-9.

Diewert, W.E., 2003a. Hedonic regressions: A consumer theory approach, in: Feenstra, R.C., Shapiro, M.D. (Eds.). Scanner Data and Price Indexes, Studies in Income and Wealth, Vol. 61. University of Chicago Press, Chicago, pp. 317–348.

Diewert, W.E., 2003b. Hedonic regressions: A review of some unresolved issues, Paper presented at the Seventh Meeting of the Ottawa Group, Paris, 27–29 May.

Diewert, W.E., 2004, On the Stochastic Approach to Linking the Regions in the ICP, Discussion Paper No. 04-16. Department of Economics, the University of British Columbia, Vancouver, Canada.

Diewert, W.E., 2005b. Adjacent period dummy variable hedonic regressions and bilateral index number theory. Annales *D'Économie et de Statistique* 79/80, 759–786. https://doi.org/10.2307/20777596.

Diewert, W.E., 2009. Similarity indexes and criteria for spatial linking, in: Prasada Rao, D.S. (Ed.). Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications. Edward Elgar Publishing, Cheltenham, UK, pp. 183–216.

Diewert, W.E., 2013. Methods of aggregation above the basic heading Level within regions, in: Measuring the Real Size of the World Economy: The Framework, Methodology and Results of the International Comparison Program—ICP. The World Bank, Washington District of Columbia, pp. 121–167.

Diewert, W.E., 2022, Chapter 8. Quality adjustment methods. https://www.imf.org/en/Data/Statistics/cpi-manual, in: Consumer Price Index Theory. International Monetary Fund, Washington District of Columbia, published online at.

18th Meeting of the Ottawa Group on Price Indices at Ottawa, Canada, May 13-15, 2024.

Diewert, W.E., 2023. Scanner Data, elementary Price indexes and the chain drift problem, in: Chotikapanich, D., Rambaldi, A.N., Rhode, N. (Eds.). Advances in Economic Measurement. Palgrave MacMillan, Singapore, pp. 445–606.

Diewert, W.E., Fox, K.J., 2022. Substitution bias in multilateral methods for CPI construction. Journal of Business and Economic Statistics 40, 355–369. https://doi.org/10.1080/07350015.2020.1816176.

Fisher, I., 1922. The Making of Index Numbers. Houghton Mifflin, Boston.

Fox, K.J., Level, P., O'Connell, M., 2022, 'Multilateral Index Number Methods for Consumer Price Statistics', ESCoE Discussion Paper No. 2022-08. Economic Statistics Centre of Excellence, National Institute of Economic and Social Research, London.

Fox, K.J., Levell, P., O'Connell, M., 2023, Inflation Measurement with High Frequency Data, Working Paper 23/29. Institute for Fiscal Studies, London, U.K.

Geary, R.G., 1958. A note on comparisons of exchange rates and purchasing Power between countries. Journal of the Royal Statistical Society Series A 121, 97–99.

Griliches, Z., 1971. Introduction: Hedonic Price indexes revisited, in: Griliches, Z. (Ed.). Price Indexes and Quality Change. Harvard University Press, Cambridge Massachusetts, pp. 3–15.

Hardy, G.H., Littlewood, J.E., Pólya, G., 1934. Inequalities. Cambridge University Press, Cambridge.

Hill, R.J., 1997. A taxonomy of multilateral methods for making international comparisons of prices and quantities. The Review of Income and Wealth 43, 49–69. https://doi.org/10.1111/j.1475-4991.1997.tb00200.x.

Hill, R.J., 1999a. Comparing Price levels across countries using minimum spanning trees. The Review of Economics and Statistics 81, 135–142. https://doi.org/10.1162/003465399767923881.

Hill, R.J., 1999b. International comparisons using spanning trees, in: Heston, A., Lipsey, R.E. (Eds.), Studies in Income and Wealth Volume 61. International and Interarea Comparisons of Income, Output and Prices. National Bureau for Economic Research, Chicago. The University of Chicago Press, pp. 109–120.

Hill, R.J., 2001. Measuring inflation and growth using spanning trees. International Economic Review 42, 167–185. https://doi.org/10.1111/1468-2354.00105.

Hill, R.J., 2004. Constructing Price indexes across space and Time: The case of the European Union. American Economic Review 94, 1379–1410. https://doi.org/10.1257/0002828043052178.

Hill, R.J., 2009. Comparing per capita income levels across countries using spanning trees: Robustness, prior restrictions, hybrids and hierarchies, in: Prasada Rao, D.S. (Ed.). Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications. Edward Elgar Publishing, Cheltenham UK, pp. 217–244.

Hill, R.J., Prasada Rao, D.S., Shankar, S., Hajargasht, R., 2017. Spatial chaining as a way of improving international comparisons of prices and real incomes, paper presented at the Meeting on the International Comparisons of Income, Prices and Production, Princeton University, May 25–26.

Hill, R.J., Timmer, M.P., 2006. Standard errors as weights in multilateral Price indexes. Journal of Business and Economic Statistics 24, 366–377. https://doi.org/10.1198/073500105000000270. https://ifs.org.uk/publications/inflation-.

Inklaar, R., Diewert, W.E., 2016. Measuring industry productivity and cross-country convergence. Journal of Econometrics 191, 426–433. https://doi.org/10.1016/j.jeconom.2015.12.013.

Ivancic, L., Diewert, W.E., Fox, K.J., 2009, Scanner Data, Time Aggregation and the Construction of Price Indexes, Discussion Paper 09-09. Department of Economics, University of British Columbia, Vancouver, Canada.

Ivancic, L., Erwin Diewert, W.E., Fox, K.J., 2011. Scanner Data, time aggregation and the construction of Price indexes. Journal of Econometrics 161, 24–35. https://doi.org/10.1016/j.jeconom.2010.09.003.

Khamis, S.H., 1970. Properties and conditions for the existence of a new type of index number. Sankhya B. 32, 81–98.

Kravis, I.B., Heston, A., Summers, R., 1982. World Product and Income: International Comparisons of Real Gross Product. Statistical Office of the United Nations and The World Bank, Baltimore. The Johns Hopkins University Press.

Krsinich, F., 2016. The FEWS index: Fixed effects with a window splice. Journal of Official Statistics 32, 375–404. https://doi.org/10.1515/jos-2016-0021.

Muellbauer, J., 1974. Household production theory, quality and the Hedonic technique. American Economic Review 64, 977–994.

Rao, D.S.P., 1995. On the Equivalence of the Generalized country-Product-Dummy (CPD) Method and the Rao-System for Multilateral Comparisons, Working Paper No. 5. Centre for International Comparisons, University of Pennsylvania, Philadelphia.

Rao, D.S.P., 2004. The country-Product-Dummy Method: A Stochastic Approach to the Computation of Purchasing Power Parities in the ICP, paper presented at the SSHRC Conference on Index Numbers and Productivity Measurement, June 30 – July 3, 2004, Vancouver, Canada.

Rao, D.S.P., 2005. On the equivalence of the weighted country product dummy (CPD) method and the Rao system for multilateral price comparisons, Review of Income and Wealth 51, 571–580. https://doi.org/10.1111/j.1475-4991.2005.00169.x.

Rao, D.S.P., Hajargasht, G., 2016. Stochastic approach to computation of purchasing power parities in the International Comparison Program (ICP). Journal of Econometrics 191, 414–425. https://doi.org/10.1016/j.jeconom.2015.12.012.

Rao, D.S.P., Timmer, M.P., 2003. Purchasing Power Parities for Industry Comparisons Using Weighted Elteto–Koves–Szulc (EKS) Methods. The Review of Income and Wealth 49, 491–511. https://doi.org/10.1111/j.0034-6586.2003.00101.x.

Rosen, S., 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy 82, 34–55. https://doi.org/10.1086/260169.

Sergeev, S., 2001. Measures of the similarity of the country's Price structures and their practical application, Conference on the European Comparison Program. United Nations Statistical Commission. Economic Commission for Europe, Geneva, November 12-14, 2001.

Sergeev, S., 2009. Aggregation methods based on structural international prices, in: Prasada Rao, D.S. (Ed.). Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications. Edward Elgar Publishing, Cheltenham UK, pp. 274–297.

Summers, R., 1973. International comparisons with incomplete Data. Rev. Income Wealth. 29, 1–16.

Szulc, B.J., 1983. Linking price index numbers, in: Diewert, W.E., Montmarquette, C. (Eds.). Price Level Measurement. Statistics Canada, Ottawa, pp. 537–566.

Triplett, J., 1987. Hedonic functions and hedonic indexes, in: Eatwell, J., Milgate, M., Newman, P. (Eds.). The New Palgrave: A Dictionary of Economics, Vol. 2. Stockton Press, New York, pp. 630–634.

Triplett, J., 2004. Handbook on hedonic indexes and quality adjustments in Price indexes, Directorate for Science. Technology Publishing and D./S./S./S./S./S.C., Industry, 9. OECD, Paris.

Triplett, J.E., McDonald, R.J., 1977. Assessing the quality error in output measures: The case of refrigerators. The Review of Income and Wealth 23, 137–156. https://doi.org/10.1111/j.1475-4991.1977.tb00008.x.

White, K.J., 2004. Shazam: User's Reference Manual*, version 10*. Northwest Econometrics Ltd, Vancouver, Canada.

Online Appendices : Data Construction and Supplementary Tables.

In this Appendix, we explain how our dummy variables for the various characteristics were constructed and we list various Tables for indexes which are described in the main text. The first eight subsections of this Appendix describe panel time dummy characteristics hedonic regressions as we introduce an additional explanatory characteristic one at a time. These regressions are unweighted Time Dummy Characteristics regressions of the type defined by (41) in the main text.

A.1 A Characteristics Hedonic Regression with Clock Speed as the Only Characteristic.

The price indexes $P_A{}^t$ and $P_{UV}{}^t$ make no adjustments for changes in the average quality of laptops over time. Let lnP denote the vector of dimension 2639 that consists of the logarithms of the monthly unit value prices of the subset of the 366 products that were sold in each month. We start our analysis by regressing the price vector lnP on the time dummy variables D_2, …, D_{24} and dummy variables for the clock speed of each laptop that was sold during the sample period.

The clock speeds range from 1.0 to 3.4 in increments of 0.1. Thus there are 25 possible clock speeds. Vectors of dummy variables of dimension 2639, D_{C1}, D_{C2}, …, D_{C25}, were generated using IF statements applied to the CLOCK variable.[52] The number of observations in each cell of clock speeds were as follows: 53, 280, 69, 18, 85, 51, 225, 0, 486, 104, 165, 201, 63, 186, 151, 31, 305, 12, 124, 10, 2, 10, 0, 4, 4. Thus D_{C8} and D_{C23} were vectors of zeros and there were no products that have clock speeds equal to 1.7 or 3.2. Also, several cells had very few members. Thus we reduced the number of cell speed categories from 25 to 7. We attempted to get approximately the same number of observations in each category except the highest cell speed category. New Groups 1 to 7 aggregated old groups 1-3, 4-8, 8-9, 10-12, 13-15, 16-18 and 19-25 respectively. Thus the new dummy variable vector D_{C1} equals the sum of the old vectors $D_{C1} + D_{C2} + D_{C3}$, the new D_{C2} equals the sum of the old vectors $D_{C4} + D_{C5} + D_{C6} + D_{C7} + D_{C8}$ and so on.

Our first hedonic regression regresses lnP on the time dummy variable vectors $D_2, D_3,\ldots,D_{24}$ and the 7 clock speed dummy variable vectors D_{C1}, D_{C2}, …, D_{C7}. The number of products that are in each of the 7 new clock speed cells are 402, 379, 486, 470, 400, 348 and 154. Since we have only one characteristic, our initial linear regression is the following one:

(A1) $\ln P = \Sigma_{t=2}^{24} \rho_t D_t + \Sigma_{j=1}^{7} b_{Cj} D_{Cj} + e$

where e is an error vector of dimension 2639.

We estimated the unknown parameters, $\rho_2^*, \rho_3^*, \ldots, \rho_{24}^*, b_{C1}^*, \ldots, b_{C7}^*$ in the above linear regression model using ordinary least squares. The log of the likelihood function was − 1401.58 and the R^2 between the observed price vector and the predicted price vector was only 0.2984. If increased clock speed is valuable to purchasers, we would expect the estimated b_{Cj}^* coefficients to increase as j increases. For this regression, the estimates for $b_{C1}^*, \ldots, b_{C7}^*$ were −0.4213, 0.0669, 0.1498, −0.0050, 0.2606, 0.3253 and 0.4535. These coefficients increase monotonically except for b_{C4}^*, so overall, it seems that purchasers do value increased clock speed.[53]

[52] Using SHAZAM, D_{C1}, D_{C2}, …, D_{C25} can be generated using the commands GENR DCL1=(CLOCK.EQ.1.0), GENR DCL2=(CLOCK.EQ.1.1), …, GENR DCL25=(CLOCK.EQ.3.4).
[53] Of course, these coefficients will change as we add other characteristics to the regression.

The estimated ρ_t^* are the logarithms of the price levels P^{t*} for t = 2,3,…,24 but we will not list the estimated price levels until we have entered all 8 of our characteristics listed in the data Appendix B into the regression.

Once the estimates for the b_{Cj} are available, we can calculate the logarithms of the appropriate quality adjustment factor α_{tn}^* that can be used to determine the quality of product n in month t. Denote the logarithm of α_{tn}^* by β_{tn}^* for t = 1,…,24 and $n \in S(t)$. Denote the vector of estimated quality adjustment factors (of dimension 2639) by β^*. It turns out that β^* can be calculated as follows:

(A2) $\beta^* = \Sigma_{j=1}^{7} b_{Cj}^* D_{Cj}$.

It is convenient to have a constant term in a linear regression: if this is the case, then the error terms must sum to zero across all observations. We can introduce a constant term into our regression model defined by (A1) as follows. First define ONE as a vector of ones of dimension 2639. Consider the following linear regression model:

(A3) $\ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + e$

where e is an error vector of dimension 2639. Thus we have added a vector of ones as an independent variable in the new regression defined by (A3) and dropped the first clock speed dummy variable vector D_{C1} as an explanatory variable. Denote the ordinary least squares estimates for the parameters in (A3) by $\rho_2^{**}, \rho_3^{**}, \ldots, \rho_{24}^{**}, b_0^{**}, b_{C2}^{**}, \ldots, b_{C7}^{**}$. It turns out that $\rho_t^{**} = \rho_t^*$ for t = 2,3,…,24 and the following vector equation also holds:

(A4) $b_0^* ONE + \Sigma_{j=2}^{7} b_{Cj}^* D_{Cj} = \Sigma_{j=1}^{7} b_{Cj}^* D_{Cj}$.

Thus the vector of log quality adjustment factors for the positive observed prices in the sample, β^* defined by (A2), is also equal to the following expression:

(A5) $\beta^* = b_0^* ONE + \Sigma_{j=2}^{7} b_{Cj}^* D_{Cj}$.

In the models which follow, we will add additional characteristics to the hedonic regression model defined by (A3) rather than adding addition explanatory variables to the model defined by (A1).

A.2 A Hedonic Regression that Added Memory Capacity as an Additional Characteristic.

We add memory capacity as another price determining characteristic of a laptop. There were only 3 sizes of memory capacity (the variable MEM in the Data Appendix B): 4096, 8192 and 16384. Construct dummy variable vectors of dimension 2639 for each value of MEM.[54] Denote these vectors as D_{M1}, D_{M2} and D_{M3}. The new log price time dummy characteristic hedonic regression is the following counterpart to (A3):

(A6) $\ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + e$.

---

[54] Using SHAZAM, the commands to create these dummy variable vectors D are: GENR DM1=(MEM.EQ.4096) ; GENR DM2=(MEM.EQ.8192) and GENR DM3=(MEM.EQ.16384). The number of products in each of these 3 cells are 620, 1710 and 309.

The log of the likelihood function was – 648.937, a gain of 752.64 log likelihood points for adding 2 new memory size parameters. The R^2 between the observed price vector and the predicted price vector was 0.6034. If increased memory capacity is valuable to purchasers, we would expect the estimated b_{Mj}^* coefficients to increase as j increases. For this regression, the estimates for b_{M2}^* and b_{M3}^* were .5493 and 0.9789. This regression indicates that purchasers do value increased memory capacity and are willing to pay a higher price for a laptop with greater memory capacity, other characteristics being held constant.

A.3 A Hedonic Regression that Added Screen Size as an Additional Characteristic.

There were 10 different screen sizes (in units of 10 inches) in our sample of laptop observations. This variable is listed as SIZE in Appendix B. The 10 screen sizes in our sample were: 1.16, 1.2, 1.25, 1.33, 1.4, 1.54, 1.56, 1.6, 1.61 and 1.73. The usual commands were used to generate 10 dummy variables for this characteristic. However, for the screen sizes 1.2, 1.56 and 1.61, we had only 12, 14 and 35 observations in our sample for these three sizes. Thus we combined the dummy variable for size 1.2 with the dummy variable for 1.16,[55] combined the dummy variable for size 1.56 with size 1.54 and combined the dummy variables for sizes 1.6 and 1.61. Denote the resulting 7 dummy variables of dimension 2639 by D_{S1}, D_{S2}, …, D_{S7}. The number of observations in each of the 7 screen size cells was 98, 154, 810, 257, 1106, 114, 100.

The new log price Time Dummy Characteristics hedonic regression is the following counterpart to (A6):

$$\text{(A7)}\ \ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + e.$$

The log of the likelihood function was – 202.270, a gain of 446.667 log likelihood points for adding 6 new screen size parameters. The R^2 between the observed price vector and the predicted price vector was 0.7173. If increased screen size is valuable to purchasers, we would expect the estimated b_{Sj}^* coefficients to increase as j increases. For this regression, the estimates for b_{S2}^*-b_{S7}^* were 0.73371, 0.59447, 0.22923, 0.34524, 0.74190 and 0.68987. This regression indicates that purchasers prefer small and large screen sizes over intermediate screen sizes for laptops.

A.4 A Hedonic Regression that Added Pixels as an Additional Characteristic.

There were 10 different numbers of pixels in our sample of laptop observations. A larger number of pixels per unit of screen size will lead to clearer images on the screen and this may be utility increasing for purchasers. The pixel variable is listed as PIX in the Appendix B. There were 10 different PIX sizes in our sample. The 10 sizes (in transformed units of measurement) were: 1.049, 1.246, 1.296, 2.074, 3.318, 4.096, 5.184, 5.530, 5.898 and 8.294. The number of observations having these pixel sizes were as follows: 324, 4, 2, 1769, 5, 400, 14, 3, 79 and 39. The usual commands were used to generate the 10 pixel dummy variables, D_{P1}-D_{P10}. However, the number of observations in pixel groups 2, 3, 5, 7 and 8 were 14 or less so these groups of observations need to be combined with other categories. We ended up with 5 pixel groups: the new group 1 combined groups 1, 2 and 3; old group 4 became the new group 2, old groups 5 and 6 were combined to give us the new group 3, old groups 7, 8 and 9 were combined to be the new group 4 and the old group 10 became the new group 5. Denote the new pixel dummy variable vectors as D_{P1}-D_{P5}. The number of observations in each of these new pixel cells was 330, 1769, 405, 96, 39.

---

[55] GENR DS1=(SIZE.GE.1.16).AND.(SIZE.LE.1.20) is the SHAZAM command to construct the combined dummy variable.

The new log price time dummy characteristic hedonic regression is the following counterpart to (62):

(A8) $\ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + \Sigma_{j=2}^{5} b_{Pj} D_{Pj} + e.$

The log of the likelihood function for the hedonic regression defined by (A8) was − 71.1313, a gain of 131.139 log likelihood points for adding 4 new pixel number parameters. The R^2 between the observed price vector and the predicted price vector was 0.7440. If an increased number of pixels is valuable to purchasers, we would expect the estimated b_{Pj}^* coefficients to increase as j increases. For this regression, the estimates for b_{P2}^*-b_{P5}^* were 0.19750, 0.21889, 0.56884 and 0.69244. Thus the coefficients for the pixel dummy variables increase monotonically, indicating that purchasers are willing to pay more for an increase in screen clarity.

A.5 A Hedonic Regression that Added HDMI as an Additional Characteristic.

The dummy variable that indicates the presence of HDMI in the laptop has already been generated and is listed in Appendix B as the column vector HDMI. Denote this column vector as D_{H2} in the following hedonic regression which adds D_{H2} to the other regressor columns in (A8):

(A9) $\ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + \Sigma_{j=2}^{5} b_{Pj} D_{Pj} + b_{H2} D_{H2} + e.$

The log of the likelihood function for the hedonic regression defined by (A9) was 49.499, a gain of 120.631 log likelihood points for adding 1 new HDMI parameter. The R^2 between the observed price vector and the predicted price vector was 0.7764 which is a material increase over the R^2 of the previous model which was equal to 0.7440. If having HDMI capability in the laptop is valuable to purchasers, we would expect the estimated b_{H2}^* coefficient to be positive. Our estimated coefficient b_{H2}^* was equal to 0.36041 which is a positive number and hence, the presence of HDMI in the laptop increases utility.

A.6 A Hedonic Regression that Added Brand as an Additional Characteristic.

There are 11 brands in our sample. In Appendix B the variable BRAND takes on values from 1 to 12 but there are no brands that correspond to the number 2 in our sample for the 24 months in the years 2021 and 2022. Here are the numbers of observations in each of the 12 BRAND categories: 4, 0, 3,101, 6, 235, 107, 389, 489, 439, 327, 479. We calculated the sample wide average price for each brand and re-ordered the brands according to their average prices with the lowest average price brands listed first and the highest average brand listed last. After re-ordering (and dropping old brand 2), the new brand ordering from 1-11 consists of the following initial brands: 7, 6, 5, 9, 1, 12, 8, 4, 11, 10, 3. The number of observations in each new BRAND category are as follows: 107, 235, 66, 489, 4, 479, 389, 101, 327, 439, 3. Construct the 11 vectors of dummy variables for the 11 new brand categories and denote these vectors of dimension 2639 by D_{B1}-D_{B11}.

Add the column vectors D_{B2}-D_{B11} to the other regressor columns in (64) in order to obtain the following hedonic regression model:

(A10) $\ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + \Sigma_{j=2}^{5} b_{Pj} D_{Pj} + b_{H2} D_{H2}$
$+ \Sigma_{j=2}^{11} b_{Bj} D_{Bj} + e.$

The log of the likelihood function for the hedonic regression defined by (A10) was 754.295, a huge gain of 704.796 log likelihood points for adding 10 new brand parameters. The R^2 between the observed price vector and the predicted price vector was 0.8631 which is a very big increase over the R^2 of the previous model which was equal to 0.7764. The estimated brand coefficients b_{B2}^* - b_{B11}^* are as follows: – 0.1014, 0.1366, 0.0975, 0.1201, 0.5048, 0.4136, 0.1469, 0.4743, 0.2880, 0.6401. Thus there is a general tendency for the marginal utility of a more expensive brand to be higher than the marginal utility of a cheaper brand.

A.7 A Hedonic Regression that Added the Type of Central Processing Unit as an Additional Characteristic.

There are 10 types of Central Processing Units (CPUs) in our sample. Here are the numbers of observations in each of the 10 CPU categories: 245, 702, 766, 39, 66, 87, 255, 11, 462, 6. Construct the 10 vectors of dummy variables for the 10 CPU categories and denote these vectors of dimension 2639 by D_{U1}-D_{U10}.

Add the terms $\Sigma_{j=2}^{10} b_{Uj}D_{Uj}$ in order to obtain the new hedonic regression model. The log of the likelihood function for the new hedonic regression model was 1012.80, a large gain of 258.505 log likelihood points for adding 9 new CPU parameters. The R^2 between the observed price vector and the predicted price vector was 0.8874, which is an increase over the R^2 of the previous model which was equal to 0.8631.

A.8 A Hedonic Regression that Added Laptop Weight as an Additional Characteristic.

We defined 7 weight dummy variables, D_{W1}-D_{W7} by choosing the following break points for laptop weights: 1.0, 1.3, 1.6, 1.9, 2.1 and 2.3. The D_{W1} cell consisted of laptops that weighed less than 1 kilo, the D_{W2} cell consisted of laptops that were in the interval $1 \leq \text{WEIGHT} < 1.3$, ,,, , the D_{W6} cell consisted of laptops that were in the interval $2.1 \leq \text{WEIGHT} < 2.3$ and the D_{W7} cell consisted of laptops that satisfied the inequality $\text{WEIGHT} \geq 2.3$. The number of laptops in each of these cells was as follows: 417, 408, 477, 311, 297, 466, 263.

Add the column vectors D_{W2}-D_{W7} to the right hand side of the previous regression in order to obtain the following hedonic regression model:

$$\text{(A11)}\ \ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj}D_{Mj} + \Sigma_{j=2}^{7} b_{Sj}D_{Sj} + \Sigma_{j=2}^{5} b_{Pj}D_{Pj} + b_{H2}D_{H2} + \Sigma_{j=2}^{11} b_{Bj}D_{Bj} + \Sigma_{j=2}^{10} b_{Uj}D_{Uj} + \Sigma_{j=2}^{7} b_{Wj}D_{Wj} + e.$$

The log of the likelihood function for the hedonic regression defined by (66) was 1074.86, an increase of 62.06 over the previous log likelihood for adding 6 additional parameters. The R^2 between the observed price vector and the predicted price vector was 0.8926 which is a substantial increase over the R^2 of the previous model which was equal to 0.8631. The estimated weight coefficients b_{W2}^*- b_{W7}^* are as follows: 0.0765, 0.0018, -0.2094, -0.2447, -0.1852 and -0.2378. Thus a lighter laptop has on average a slightly positive price premium but the price premium becomes negative (and approximately constant) for laptops that weigh more that 1.6 kilos.

The estimated coefficients on the time dummy variables in this regression are ρ_2^*, ρ_3^*, …, ρ_{24}^*. Define $\rho_1^* \equiv 0$ and the estimated period t price levels $\pi_t^* \equiv \exp[\rho_t^*]$ for t = 1,2,…,24. Define the month t *Time Dummy Characteristics Price Index*, $P_{TDC}^t \equiv \pi_t^*$ for t = 1,…,24. This index is listed in Table A2 in the following subsection.

A.9 The Weighted Time Dummy Characteristics Hedonic Regression Model.

The price indexes defined in sections A4.1-A4.8 can be constructed by using information on product prices and the amounts of the various characteristics of each product. If in addition, information on quantities sold or purchased during each month in scope is available, then Weighted Time Dummy Characteristics price indexes defined by equations (49) in section 3 can be constructed.

Recall that the expenditure share that corresponds to purchased product n in month t is defined as $s_{tn} = p_{tn}q_{tn}/\Sigma_{j\in S(t)}\, p_{tj}q_{tj}$ for t = 1,…,24 and n∈S(t). To obtain the weighted counterpart to the hedonic regression model defined by (A11) above, we just form a share vector of dimension 2639 that corresponds to the $\ln p_{tn}$ that appear in (A11) and then form a new vector of dimension 2639 that consists of the positive square roots of each s_{tn}. Call this vector of square roots SS. Now multiply both sides of (A11) by SS to obtain a new linear regression model which again provides estimates for the unknown parameters that appear in (A11). The R^2 for this new weighted regression model turned out to be 0.9152 which is substantially higher than the R^2 for the counterpart unweighted model which was 0.8926.

The parameter estimates for this weighted hedonic regression model are listed in Table A1. This is our preferred regression from all of the regression models that have been presented thus far.

Table A1: Parameter Estimates for the Weighted Time Dummy Characteristics Hedonic Regression

Coef	Estimate	Std. Error	t-stat	Coef	Estimate	Std. Error	t-stat
b_0^*	-1.146	0.038	-29.950	b_{S4}^*	0.374	0.031	12.000
ρ_2^*	-0.001	0.016	-0.053	b_{S5}^*	0.547	0.044	12.490
ρ_3^*	0.018	0.016	1.154	b_{S6}^*	0.727	0.050	14.630
ρ_4^*	0.024	0.016	1.517	b_{S7}^*	0.733	0.047	15.650
ρ_5^*	0.019	0.016	1.191	b_{P2}^*	0.001	0.018	0.046
ρ_6^*	0.004	0.016	0.255	b_{P3}^*	0.215	0.043	4.944
ρ_7^*	-0.026	0.016	-1.607	b_{P4}^*	0.362	0.055	6.600
ρ_8^*	-0.028	0.016	-1.713	b_{P5}^*	0.266	0.033	8.143
ρ_9^*	-0.038	0.016	-2.323	b_{H2}^*	0.296	0.019	15.460
ρ_{10}^*	-0.032	0.016	-1.972	b_{B2}^*	-0.191	0.025	-7.688
ρ_{11}^*	-0.060	0.016	-3.650	b_{B3}^*	0.083	0.033	2.485
ρ_{12}^*	-0.089	0.016	-5.433	b_{B4}^*	-0.032	0.022	-1.429
ρ_{13}^*	-0.101	0.016	-6.165	b_{B5}^*	0.007	0.133	0.050
ρ_{14}^*	-0.109	0.017	-6.592	b_{B6}^*	0.319	0.022	14.320
ρ_{15}^*	-0.103	0.017	-6.178	b_{B7}^*	0.240	0.023	10.350
ρ_{16}^*	-0.083	0.017	-4.942	b_{B8}^*	0.010	0.029	0.356
ρ_{17}^*	-0.092	0.017	-5.438	b_{B9}^*	0.310	0.024	12.710
ρ_{18}^*	-0.098	0.017	-5.778	b_{B10}^*	0.190	0.032	5.940
ρ_{19}^*	-0.096	0.017	-5.630	b_{B11}^*	0.838	0.143	5.857

ρ_{20}^*	-0.115	0.017	-6.809	b_{U2}^*	0.202	0.012	16.470
ρ_{21}^*	-0.122	0.017	-7.208	b_{U3}^*	0.149	0.015	9.831
ρ_{22}^*	-0.116	0.017	-6.825	b_{U4}^*	0.445	0.030	14.710
ρ_{23}^*	-0.098	0.017	-5.752	b_{U5}^*	-0.184	0.024	-7.660
ρ_{24}^*	-0.080	0.017	-4.681	b_{U6}^*	0.043	0.026	1.622
b_{C2}^*	0.076	0.015	5.039	b_{U7}^*	0.213	0.015	14.500
b_{C3}^*	0.239	0.017	14.230	b_{U8}^*	0.212	0.074	2.854
b_{C4}^*	0.177	0.017	10.670	b_{U9}^*	-0.113	0.019	-5.923
b_{C5}^*	0.206	0.017	11.950	b_{U10}^*	-0.160	0.127	-1.258
b_{C6}^*	0.341	0.021	16.410	b_{W2}^*	0.083	0.019	4.417
b_{C7}^*	0.293	0.021	14.090	b_{W3}^*	0.063	0.021	3.073
b_{M2}^*	0.093	0.014	6.509	b_{W4}^*	-0.137	0.039	-3.536
b_{M3}^*	0.399	0.019	20.810	b_{W5}^*	-0.102	0.039	-2.603
b_{S2}^*	0.476	0.034	14.110	b_{W6}^*	-0.142	0.040	-3.577
b_{S3}^*	0.598	0.032	18.880	b_{W7}^*	-0.154	0.039	-3.944

There are 68 parameters in this regression model with 2571 degrees of freedom for the error terms. It can be seen that the clock speed parameters b_{Cj}^* increase with respect to j up to a point and then basically level off; the memory capacity parameters b_{M2}^* and b_{M3}^* are monotonically increasing; the screen size parameters b_{Sj}^* are roughly increasing; the pixel parameters b_{Pj}^* are monotonically increasing (except for b_{P5}^*); the HDMI parameter b_{H2}^* is positive which indicates that the availability of HDMI is valued by purchasers and the brand parameters b_{Bj}^* are weakly increasing so that the higher price brands are mostly preferred by purchasers. With respect to weight, it appears that lighter models are preferred up to a point and then weight does not seem to matter much.

The estimated coefficients on the time dummy variables in this regression are ρ_2^*, ρ_3^*, …, ρ_{24}^*. Define $\rho_1^* \equiv 0$ and the estimated period t price levels $\pi_t^* \equiv \exp[\rho_t^*]$ for t = 1,2,…,24. Define the month t *Weighted Time Dummy Characteristics Price Index*, $P_{WTDC}^t \equiv \pi_t^*$ for t = 1,…,24. This index is listed in Table A2 (and plotted in Chart 1 in the main text) and it is our a priori preferred index thus far. The corresponding unweighted (or equally weighted) Time Dummy Characteristics Price Index P_{TDC}^t is also listed in Table 5 along with the simple average laptop price indexes P_A^t and P_{UV}^t defined in section 4 of the main text.

Table A2: Weighted and Unweighted Time Dummy Characteristics Price Indexes

Month t	P_{WTDC}^t	P_{TDC}^t	P_A^t	P_{UV}^t
1	1.000	1.000	1.000	1.000
2	0.999	1.034	1.035	0.997
3	1.019	1.041	1.035	1.010
4	1.025	1.037	1.021	0.995
5	1.019	1.080	1.063	1.020
6	1.004	1.070	1.066	1.002
7	0.974	1.018	1.027	0.984
8	0.972	1.014	1.020	0.974
9	0.963	1.002	1.011	0.951

10	0.968	1.020	1.036	0.991
11	0.942	0.994	1.013	0.947
12	0.914	0.947	0.949	0.879
13	0.904	0.970	0.903	0.844
14	0.897	0.957	0.914	0.846
15	0.902	0.943	0.899	0.846
16	0.921	0.955	0.932	0.854
17	0.912	0.950	0.891	0.802
18	0.907	0.948	0.866	0.791
19	0.909	0.951	0.851	0.799
20	0.891	0.932	0.831	0.793
21	0.885	0.931	0.848	0.771
22	0.891	0.938	0.904	0.853
23	0.906	0.952	0.859	0.846
24	0.923	0.935	0.892	0.878
Mean	0.944	0.984	0.953	0.903

A.10 Supplementary Tables for Sections 4-7

Table A3: Sample Wide and Adjacent Period Weighted and Unweighted Characteristics Price Indexes.

Month t	P_{WATDC}^t	P_{ATDC}^t	P_{WTDC}^t	P_{TDC}^t	P_A^t	P_{UV}^t
1	1.000	1.000	1.000	1.000	1.000	1.000
2	0.999	1.026	0.999	1.034	1.035	0.997
3	1.007	1.031	1.019	1.041	1.035	1.010
4	1.010	1.021	1.025	1.037	1.021	0.995
5	1.007	1.039	1.019	1.080	1.063	1.020
6	0.995	1.030	1.004	1.070	1.066	1.002
7	0.972	0.989	0.974	1.018	1.027	0.984
8	0.960	0.981	0.972	1.014	1.020	0.974
9	0.953	0.965	0.963	1.002	1.011	0.951
10	0.949	0.972	0.968	1.020	1.036	0.991
11	0.929	0.945	0.942	0.994	1.013	0.947
12	0.899	0.913	0.914	0.947	0.949	0.879
13	0.875	0.911	0.904	0.970	0.903	0.844
14	0.863	0.897	0.897	0.957	0.914	0.846
15	0.851	0.884	0.902	0.943	0.899	0.846
16	0.862	0.891	0.921	0.955	0.932	0.854
17	0.857	0.884	0.912	0.950	0.891	0.802
18	0.846	0.873	0.907	0.948	0.866	0.791
19	0.841	0.869	0.909	0.951	0.851	0.799
20	0.825	0.849	0.891	0.932	0.831	0.793
21	0.818	0.847	0.885	0.931	0.848	0.771
22	0.807	0.845	0.891	0.938	0.904	0.853
23	0.795	0.843	0.906	0.952	0.859	0.846
24	0.788	0.831	0.923	0.935	0.892	0.878
Mean	0.905	0.931	0.944	0.984	0.953	0.903

Table A4: Sample Wide and Adjacent Period Weighted and Unweighted Time Product Dummy Price Indexes

Month t	P_{WATPD}^t	P_{ATPD}^t	P_{WTPD}^t	P_{TPD}^t	P_A^t	P_{UV}^t
1	1.000	1.000	1.000	1.000	1.000	1.000
2	0.994	0.988	0.988	0.983	1.035	0.997
3	0.985	0.981	0.982	0.978	1.035	1.010
4	0.985	0.967	0.980	0.965	1.021	0.995
5	0.975	0.949	0.969	0.953	1.063	1.020
6	0.964	0.931	0.951	0.937	1.066	1.002
7	0.944	0.907	0.923	0.906	1.027	0.984
8	0.930	0.886	0.918	0.889	1.020	0.974
9	0.920	0.869	0.910	0.877	1.011	0.951
10	0.916	0.863	0.903	0.874	1.036	0.991
11	0.891	0.835	0.879	0.853	1.013	0.947
12	0.859	0.801	0.851	0.825	0.949	0.879
13	0.826	0.776	0.833	0.808	0.903	0.844
14	0.815	0.764	0.826	0.795	0.914	0.846
15	0.796	0.749	0.814	0.779	0.899	0.846
16	0.795	0.747	0.823	0.779	0.932	0.854
17	0.787	0.734	0.820	0.771	0.891	0.802
18	0.778	0.723	0.810	0.759	0.866	0.791
19	0.767	0.708	0.809	0.754	0.851	0.799
20	0.752	0.694	0.798	0.745	0.831	0.793
21	0.743	0.685	0.788	0.737	0.848	0.771
22	0.734	0.675	0.785	0.733	0.904	0.853
23	0.715	0.661	0.768	0.724	0.859	0.846
24	0.694	0.646	0.745	0.707	0.892	0.878
Mean	0.857	0.814	0.870	0.839	0.953	0.903

Table A5: Predicted Share Measures of Relative Price Similarity for 24 Months

r	Δ(r,1)	Δ(r,2)	Δ(r,3)	Δ(r,4)	Δ(r,5)	Δ(r,6)	Δ(r,7)	Δ(r,8)	Δ(r,9)	Δ(r,10)	Δ(r,11)	Δ(r,12)
1	0.000	0.010	0.009	0.017	0.031	0.049	0.051	0.051	0.072	0.064	0.088	0.101
2	0.010	0.000	0.001	0.009	0.015	0.026	0.027	0.033	0.041	0.045	0.055	0.055
3	0.009	0.001	0.000	0.005	0.006	0.012	0.016	0.017	0.023	0.024	0.032	0.034
4	0.017	0.009	0.005	0.000	0.012	0.015	0.021	0.020	0.027	0.027	0.041	0.046
5	0.031	0.015	0.006	0.012	0.000	0.001	0.008	0.003	0.007	0.007	0.017	0.022
6	0.049	0.026	0.012	0.015	0.001	0.000	0.008	0.003	0.007	0.006	0.016	0.021
7	0.051	0.027	0.016	0.021	0.008	0.008	0.000	0.005	0.004	0.006	0.007	0.008
8	0.051	0.033	0.017	0.020	0.003	0.003	0.005	0.000	0.000	0.001	0.001	0.001
9	0.072	0.041	0.023	0.027	0.007	0.007	0.004	0.000	0.000	0.001	0.000	0.001
10	0.064	0.045	0.024	0.027	0.007	0.006	0.006	0.001	0.001	0.000	0.001	0.004
11	0.088	0.055	0.032	0.041	0.017	0.016	0.007	0.001	0.000	0.001	0.000	0.000
12	0.101	0.055	0.034	0.046	0.022	0.021	0.008	0.001	0.001	0.004	0.000	0.000
13	0.140	0.083	0.050	0.050	0.029	0.028	0.024	0.016	0.014	0.017	0.013	0.013
14	0.141	0.094	0.057	0.055	0.035	0.034	0.032	0.022	0.024	0.023	0.019	0.018
15	0.149	0.101	0.062	0.057	0.041	0.040	0.037	0.027	0.030	0.029	0.024	0.024

16	0.178	0.116	0.080	0.077	0.051	0.048	0.046	0.035	0.037	0.037	0.032	0.034
17	0.300	0.236	0.148	0.129	0.093	0.087	0.093	0.076	0.076	0.078	0.074	0.086
18	0.380	0.299	0.172	0.144	0.085	0.077	0.083	0.067	0.069	0.067	0.068	0.082
19	0.394	0.343	0.284	0.255	0.155	0.155	0.158	0.138	0.139	0.141	0.134	0.143
20	0.608	0.507	0.326	0.253	0.173	0.166	0.172	0.153	0.153	0.154	0.157	0.185
21	0.589	0.501	0.284	0.223	0.155	0.147	0.185	0.166	0.166	0.168	0.171	0.196
22	0.850	0.671	0.445	0.380	0.232	0.222	0.246	0.247	0.244	0.246	0.246	0.290
23	0.865	0.657	0.491	0.457	0.363	0.373	0.427	0.406	0.406	0.410	0.417	0.463
24	1.013	0.856	0.613	0.459	0.318	0.307	0.354	0.261	0.263	0.261	0.282	0.325

r	**Δ(r,13)**	**Δ(r,14)**	**Δ(r,15)**	**Δ(r,16)**	**Δ(r,17)**	**Δ(r,18)**	**Δ(r,19)**	**Δ(r,20)**	**Δ(r,21)**	**Δ(r,22)**	**Δ(r,23)**	**Δ(r,24)**
1	0.140	0.141	0.149	0.178	0.300	0.380	0.394	0.608	0.589	0.850	0.865	1.013
2	0.083	0.094	0.101	0.116	0.236	0.299	0.343	0.507	0.501	0.671	0.657	0.856
3	0.050	0.057	0.062	0.080	0.148	0.172	0.284	0.326	0.284	0.445	0.491	0.613
4	0.050	0.055	0.057	0.077	0.129	0.144	0.255	0.253	0.223	0.380	0.457	0.459
5	0.029	0.035	0.041	0.051	0.093	0.085	0.155	0.173	0.155	0.232	0.363	0.318
6	0.028	0.034	0.040	0.048	0.087	0.077	0.155	0.166	0.147	0.222	0.373	0.307
7	0.024	0.032	0.037	0.046	0.093	0.083	0.158	0.172	0.185	0.246	0.427	0.354
8	0.016	0.022	0.027	0.035	0.076	0.067	0.138	0.153	0.166	0.247	0.406	0.261
9	0.014	0.024	0.030	0.037	0.076	0.069	0.139	0.153	0.166	0.244	0.406	0.263
10	0.017	0.023	0.029	0.037	0.078	0.067	0.141	0.154	0.168	0.246	0.410	0.261
11	0.013	0.019	0.024	0.032	0.074	0.068	0.134	0.157	0.171	0.246	0.417	0.282
12	0.013	0.018	0.024	0.034	0.086	0.082	0.143	0.185	0.196	0.290	0.463	0.325
13	0.000	0.004	0.003	0.006	0.018	0.023	0.036	0.038	0.044	0.084	0.102	0.094
14	0.004	0.000	0.001	0.003	0.011	0.017	0.025	0.025	0.030	0.066	0.077	0.076
15	0.003	0.001	0.000	0.000	0.004	0.007	0.011	0.010	0.015	0.049	0.055	0.057
16	0.006	0.003	0.000	0.000	0.001	0.004	0.004	0.005	0.006	0.041	0.043	0.046
17	0.018	0.011	0.004	0.001	0.000	0.002	0.003	0.003	0.004	0.039	0.041	0.044
18	0.023	0.017	0.007	0.004	0.002	0.000	0.001	0.003	0.002	0.036	0.036	0.040
19	0.036	0.025	0.011	0.004	0.003	0.001	0.000	0.001	0.001	0.035	0.033	0.037
20	0.038	0.025	0.010	0.005	0.003	0.003	0.001	0.000	0.001	0.034	0.034	0.037
21	0.044	0.030	0.015	0.006	0.004	0.002	0.001	0.001	0.000	0.033	0.031	0.036
22	0.084	0.066	0.049	0.041	0.039	0.036	0.035	0.034	0.033	0.000	0.001	0.004
23	0.102	0.077	0.055	0.043	0.041	0.036	0.033	0.034	0.031	0.001	0.000	0.001
24	0.094	0.076	0.057	0.046	0.044	0.040	0.037	0.037	0.036	0.004	0.001	0.000

Table A5 can be used to construct the relative price similarity linked Predicted Share Price index, P_S^t, for t = 1,…,24. We set $P_S^1 = 1$. When comparing the prices of month 2 to the prices of previous months, there is only one possible comparison in our window of data so that we must compare p^2 to p^1. We use the matched model Fisher index $P_F(1,2)$ defined by (64) to define the similarity linked month 2 index. Thus $P_S^2 \equiv P_F(1,2)$. Now look at the column in Table A5 that has the heading Δ(r,3). Look at the first 2 entries in this column. We have $\Delta(1,3) = 0.0088$ and $\Delta(2,3) = 0.0007$. Since $\Delta(2,3)$ is smaller than $\Delta(1,3)$, we link month 3 to month 2 using the matched model Fisher index $P_F(2,3)$. Thus $P_S^3 \equiv P_S^2\, P_F(2,3)$. Now look at the column in Table A5 that has the heading Δ(r,4). Look at the first 3 entries in this column. We have $\Delta(1,4) = 0.0170$, $\Delta(2,4) = 0.0092$ and $\Delta(3,4) = 0.0046$. Since $\Delta(3,4)$ is the smallest of these 3 measures, we link month 4 to month 3 using the matched model Fisher index $P_F(3,4)$. Thus $P_S^4 \equiv P_S^3\, P_F(3,4)$. This procedure can be continued until we look down the column that has the heading Δ(r,24). The smallest measure of relative price similarity in the first 23 rows of this column is the entry for row 23 which has measure 0.0013. Thus we link month 24 to month 23 using the matched model Fisher index $P_F(23,24)$ which leads to the following definition for $P_S^{24} \equiv P_S^{23} P_F(23,24)$.[56]

The relative price Predicted Share Similarity Linked indexes P_S^t are listed in Table A6 below. We also list the chained maximum overlap Laspeyres, Paasche and Fisher indexes, P_{LCH}^t, P_{PCH}^t and P_{FCH}^t in Table A6. Finally, for comparison purposes, Table A6 lists our "best" hedonic price index from the previous sections, the Weighted Adjacent Period Time Product Dummy Index, P_{WATPD}^t, as well as the average laptop price index P_A^t and the Unit Value price index P_{UV}^t. See Chart 3 for plots of the indexes listed in Table A6.

Table A6: The Predicted Share Similarity Linked Price Index and Other Comparison Price Indexes

Month t	P_S^t	P_{FCH}^t	P_{LCH}^t	P_{PCH}^t	P_{WATPD}^t	P_A^t	P_{UV}^t
1	1.000	1.000	1.000	1.000	1.000	1.000	1.000
2	0.993	0.993	0.995	0.991	0.994	1.035	0.997
3	0.985	0.985	0.985	0.984	0.985	1.035	1.010
4	0.983	0.983	0.983	0.983	0.985	1.021	0.995
5	0.979	0.972	0.970	0.975	0.975	1.063	1.020
6	0.968	0.962	0.959	0.965	0.964	1.066	1.002
7	0.948	0.941	0.939	0.944	0.944	1.027	0.984
8	0.935	0.927	0.924	0.930	0.930	1.020	0.974
9	0.925	0.918	0.912	0.923	0.920	1.011	0.951
10	0.926	0.918	0.905	0.932	0.916	1.036	0.991
11	0.894	0.889	0.872	0.907	0.891	1.013	0.947
12	0.862	0.857	0.841	0.873	0.859	0.949	0.879
13	0.828	0.824	0.811	0.836	0.826	0.903	0.844
14	0.817	0.813	0.803	0.823	0.815	0.914	0.846
15	0.798	0.794	0.784	0.805	0.796	0.899	0.846
16	0.797	0.792	0.781	0.804	0.795	0.932	0.854
17	0.789	0.785	0.773	0.796	0.787	0.891	0.802
18	0.780	0.776	0.765	0.786	0.778	0.866	0.791
19	0.768	0.764	0.755	0.773	0.767	0.851	0.799
20	0.753	0.749	0.740	0.757	0.752	0.831	0.793
21	0.743	0.739	0.733	0.746	0.743	0.848	0.771
22	0.734	0.730	0.724	0.736	0.734	0.904	0.853
23	0.715	0.711	0.707	0.716	0.715	0.859	0.846
24	0.693	0.690	0.689	0.690	0.694	0.892	0.878
Mean	0.859	0.855	0.848	0.861	0.857	0.953	0.903

[56] The entire set of bilateral matched model Fisher links is as follows: 2-1; 3-2; 4-3; 5-3*; 6-5; 7-6; 8-6*; 9-8; 10-9; 11-9*; 12-11; 13-12; 14-13; 15-14; 16-15; 17-16; 18-17; 19-18; 20-19; 21-20; 22-21; 23-22; 24-23. Note that there are only 3 bilateral links that are not chain links. Thus the similarity linked indexes for our data are likely to be close to the corresponding chained maximum overlap Fisher index.