**Meeting of the Group of Experts on Consumer Price Indices, June 7-9, 2023**

# Scanner Data, Product Churn and Quality Adjustment

W. Erwin Diewert, University of British Columbia and University of New South Wales;
Email: erwin.diewert@ubc.ca

Chihiro Shimizu, Hitotsubashi University; Email: c.shimizu@r.hit-u.ac.jp

June 7, 2023

**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 or the product itself (Time Product Dummy regressions). The paper also considers regressions where the economic importance of products is taken into account (weighted versus unweighted regressions). Finally, traditional index numbers are calculated that do not make any special adjustments for quality change. The various approaches are implemented using Japanese price and quantity data on laptop sales in Japan for the 24 months in the years 2020-2021. Somewhat surprisingly, the "best" hedonic regression price index was virtually identical to the "best" traditional index.

**Key Words**

Quality adjustment, hedonic regressions, Predicted Share similarity linking, economic approach to index number theory.

**Journal of Economic Literature Classification Codes**

C32, C43, D20, D57, E31.

**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 either on 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 regressions while the second type of model is called a Time Product 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 the 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 Product Characteristics type regressions. We used characteristics data on 6 separate laptop characteristics in this section. We will 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. We will also consider the use of a hedonic regression that uses all of the data in a panel of data and the use of repeated hedonic regressions that use only the data of two consecutive periods and the results of these separate regressions are chained together to generate the final index, which is called an Adjacent Period Time Dummy Characteristics index.

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. We also consider panel regressions 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. This new methodology will be explained in section 6.

Section 7 lists some tentative conclusions that we draw from this study.

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

where the $\alpha_n$ 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, $q^t$, $\lambda_t$ to the period t aggregate utility maximization problem are the following N+1 equations, where $\lambda_t$ 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 $\lambda_t$. Using equation (3), we obtain the following expression for $\lambda_t$:

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

Define $\pi_t$ as follows:

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

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

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

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

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) $P^t \equiv e^t/Q^t$
$= \pi_t$

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 $\pi_t$ times a quality adjustment parameter for product n, $\alpha_n$.[5]

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 $\pi_t^*$ that solve the minimization problem (9) had unsatisfactory axiomatic properties. 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) $\min_{\rho, \beta} \Sigma_{n=1}^{N} \Sigma_{t=1}^{T} [\ln p_{tn} - \rho_t - \beta_n]^2$

[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) (2023a) 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 $\alpha_n$ is the marginal utility to a purchaser of a unit of product n for $n = 1,\ldots,N$. It can be shown that the period t price index $\pi_t$ 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).
[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.

where the new parameters $\rho_t$ and $\beta_n$ are defined as the logarithms of the $\pi_t$ and $\alpha_n$; 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 consider the following *weighted least squares minimization problem*:[7]

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

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

Solutions to (34) and (35) 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 $\lambda$. 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* $\pi_t^*$ and *quality adjustment factors* $\alpha_n^*$ 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 $\pi_t^*$ defined by (16) are called the *Weighted Time Product Dummy price levels*. Note that the resulting *price index* between periods t and $\tau$ is defined as the ratio of the period t price level to the period $\tau$ price level and is equal to the following expression:

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

If $s_{tn} = s_{\tau n}$ for n = 1,...,N, then $\pi_t^*/\pi_\tau^*$ 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 $\pi_t^*/\pi_\tau^*$ will not depend on the $\alpha_n^*$ in this case.
Once the estimates for the $\pi_t$ and $\alpha_n$ 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 $\pi_t^*$ estimates can be used to form the aggregates

---

[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).
[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 $\rho_1 = 0$.

using equations (18) or the $\alpha_n^*$ estimates can be used to form the aggregate period t price and quantity levels using equations (19):[10]

$$(18)\ P^{t*} \equiv \pi_t^* ;\quad Q^{t*} \equiv p^t\cdot q^t/\pi_t^* ; \qquad t = 1,...,T;$$
$$(19)\ Q^{t**} \equiv \alpha^*\cdot q^t ;\ P^{t**} \equiv p^t\cdot q^t/\alpha^*\cdot q^t ; \qquad t = 1,...,T.$$

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 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 $\lambda$. 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]

---

[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 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 observed 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.

Define the estimated *price levels* $\pi_t^*$ and *quality adjustment factors* $\alpha_n^*$ 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) $\pi_t^* = \exp[\Sigma_{n\in S(t)} s_{tn}\ln(p_{tn}/\alpha_n^*)]$ ; $\quad t = 1,...,T;$
(24) $\alpha_n^* = \exp[\Sigma_{t\in S^*(n)} s_{tn}\ln(p_{tn}/\pi_t^*)/\Sigma_{t\in S^*(n)} s_{tn}]$ ; $\quad n = 1,...,N.$

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

Thus $P^{t*}$ is a weighted geometric mean of the quality adjusted prices $p_{tn}/\alpha_n^*$ that are present in period t where the weight for $p_{tn}/\alpha_n^*$ 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) $Q^{t**} \equiv \Sigma_{n\in S(t)} \alpha_n^* q_{tn}$ ; $\quad t = 1,...,T;$
(28) $P^{t**} \equiv \Sigma_{n\in S(t)} p_{tn}q_{tn}/Q^{t**}$ $\quad t = 1,...,T;$

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

where the inequality follows from Schlömilch's inequality[15]; i.e., a weighted harmonic mean of the quality adjusted prices $p_{tn}/\alpha_n^*$ 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 models such as those to be considered in subsequent sections.

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

---

[15] See Hardy, Littlewood and Pólya (1934; 26).

models to generate imputed prices for missing prices and then use these imputed prices in their chosen index number formula.

One perhaps unsatisfactory property of the WTPD price levels $\pi_t^*$ 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 $\pi_t^*$.[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.**

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 $\alpha_n$ 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 $\alpha_n$ 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

[16] Diewert (2004) established this property.
[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.

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:

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

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.

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:[18]

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

Substitute equations (34) into equations and we obtain the following system of possible estimating equations where the $\pi_t$ and $a_{1j}$, $a_{2j}$, …, $a_{Kj}$ are unknown parameters:

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

We take logarithms of both sides of equations (35) in order to obtain the following system of estimating equations:[19]

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

Define the following parameters :

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

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

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

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

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

where $\rho$ 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….,K. Solutions to the least squares minimization problem will exist but a solution will not be unique.[20] Using equations (35), it can be seen that components of the vectors $\pi$ and $a(k) \equiv [a_{k1}, a_{k2}, \ldots, a_{kN(k)}]$ for k = 1,2,…,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 $b_{k1} = 0$ for k = 2,…,K (so $b_{11}$ is not normalized).

Once the normalizations suggested above have been imposed, the linear regression defined by (38) can be run and estimates for the unknown parameters $[\rho_1^*, \rho_2^*, \ldots, \rho_T^*]$ and $[b_{k1}^*, b_{k2}^*, \ldots, b_{kN(k)}^*]$ for k = 1,2….,K will be available. Use these estimates to define the logarithms of the quality adjustment factors $\alpha_n$ for all products n that were purchased in period t:[21]

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

The corresponding estimated product n quality adjustment factors $\alpha_{tn}^*$ are obtained by exponentiating the $\beta_{tn}^*$:

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

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

(42) $Q^{t**} \equiv \Sigma_{n\in S(t)} \alpha_{tn}^* q_{tn}$ ; 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:

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

In order to obtain a useful expression for the direct estimate for the period t price level, $\pi_t$, look at the first order conditions for minimizing (39) with respect to $\rho_t$:

(44) $0 = \Sigma_{n\in S(t)} \{\ln p_{tn} - \rho_t^* - \Sigma_{j=1}^{N(1)} b_{1j}^* D_{tn,1,j} - \Sigma_{j=1}^{N(2)} b_{2j}^* D_{tn,2,j} - \ldots - \Sigma_{j=1}^{N(K)} b_{Kj}^* D_{tn}\}$ t = 2,…,T
$= \Sigma_{n\in S(t)} \{\ln p_{tn} - \rho_t^* - \beta_n^*\}$

[20] Thus the X matrix that corresponds to the linear regression model defined by equations (36) will not have full column rank.
[21] If product n is available in multiple periods, the quality adjustment factors remain the same across periods.

where we used definitions (40) 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 (41), equations (44) imply that the direct estimate of the period t price level $\pi_t^*$ is equal to:

(45) $\pi_t^* = \Pi_{n\in S(t)}(p_{tn}/\alpha_{tn}^*)^{1/(N(t)} \equiv P^{t*}$ ; $\qquad$ t = 2,…,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 $(p_{tn}/\alpha_{tn}^*)$ 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 (39) 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:

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

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 by (39).

The new weighted counterpart to the linear regression model that was defined by equations (38) is given below:

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

In order to identify all of the parameters, make the same normalizations as were made above; i.e., set $\rho_1 = 0$ and $b_{k1} = 0$ for, k = 2,…,K. Use definitions (40), (41), (42) and (43) to define new $\beta_{tn}^*$, $\alpha_{tn}^*$, $Q^{t**}$ and $P^{t**}$. We rewrite $P^{t**}$ in a somewhat different manner as follows:

(48) $P^{t**} = \Sigma_{n\in S(t)} p_{tn}q_{tn}/\Sigma_{n\in S(t)} \alpha_{tn}^*q_{tn}$ $\qquad$ t = 1,…,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, $\pi_t$, look at the first order conditions for minimizing (46) with respect to $\rho_t$:

(49) $0 = \Sigma_{n\in S(t)} s_{tn}\{\ln p_{tn} - \rho_t^* - \Sigma_{j=1}^{N(1)} b_{1j}^* D_{tn,1,j} - \Sigma_{j=1}^{N(2)} b_{2j}^* D_{tn,2,j} - \ldots - \Sigma_{j=1}^{N(K)} b_{Kj}^* D_{tn}\}$ $\qquad$ t = 2,…,T
$= \Sigma_{n\in S(t)} s_{tn} \{\ln p_{tn} - \rho_t^* - \beta_n^*\}$

where we used definitions (40) to derive the second equality. Note that $\Sigma_{n\in S(t)} s_{tn} = 1$. Using definitions (37) and (41), equations (49) imply that the direct estimate of the period t price level $\pi_t^*$ is equal to:[22]

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

[22] Our normalizations imply $\pi_1^* = 1$.

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

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

Note that the direct period t price level defined by (50), $P^{t*}$, is a period t share weighted geometric mean of the period t quality adjusted prices $p_{tn}/\alpha_{tn}^{*}$ while the indirect period t price level $P^{t**}$ defined by (48) is a period t share weighted harmonic mean of the period t quality adjusted prices and thus we have the de Haan inequalities:

(52) $P^{t**} \leq P^{t*}$ and $Q^{t**} \geq Q^{t*}$ ; 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 Sample Wide Hedonic Regressions Using Characteristics.

### 4.1 The Laptop Data and Some Preliminary Price Indexes.

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. We 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.

The positive prices $p_{tn}$ and quantities $q_{tn}$ are listed in Table 1 in the Appendix 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 6 additional characteristics of the laptop product, which are discussed below.

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.

MEM is the memory capacity for the laptop. The mean memory size was 8188.9. There were only 4 clock speeds listed in our sample: 4096, 8192 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.

A priori, we expect 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.

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.

In summary, Table A1 in the Appendix lists the following 11 variables in vectors of dimension 2639: OBS (runs from 1 to 2639), TD, JAN, CLOCK, MEM, SIZE, PIX, HDMI, BRAND, Q and P.

The information in the column vectors TD and JAN were used to generate 24 time dummy variables, $D_1$, $D_2$, …, $D_{24}$[23] and 366 product dummy variable vectors, $D_{J1}$, $D_{J2}$,…,$D_{J366}$.[24]

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**

| Name | No. of Obs. | Mean | Std. Dev | Variance | Minimum | Maximum |
|---|---|---|---|---|---|---|
| JAN | 2639 | 195.75 | 103.94 | 10803 | 1 | 366 |
| CLOCK | 2639 | 1.9397 | 0.51807 | 0.2684 | 1 | 3.4 |
| MEM | 2639 | 8.1889 | 3.4357 | 11.804 | 4.096 | 16.384 |
| SIZE | 2639 | 1.4493 | 0.13807 | 0.0191 | 1.16 | 1.73 |
| PIX | 2639 | 2.482 | 1.2891 | 1.6617 | 1.049 | 8.294 |
| HDMI | 2639 | 0.75332 | 0.43116 | 0.1859 | 0 | 1 |
| BRAND | 2639 | 9.1527 | 2.2091 | 4.88 | 1 | 12 |
| Q | 2639 | 594.69 | 735.68 | 541230 | 100 | 5367 |
| P | 2639 | 1.1764 | 0.49155 | 0.24162 | 0.17381 | 2.8729 |

[23] Use IF statements to construct these dummy variables. Using the econometric package SHAZAM, the first two and last time dummy variable vectors of dimension 2639 were constructed using the following statements: GENR D1=(TD.EQ.20201); GENR D2=(TD.EQ.20202); …; GENR D24=(TD.EQ.202112). See White (2004) for information on Shazam.

[24] Again use IF statements to construct the product dummy variables of dimension 2639. Using SHAZAM, these dummy variables $D_{J1}$-$D_{J366}$ were constructed using the following statements: DO #=1,366; GENR DJ#=(JAN.EQ.#) ; ENDO.

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:

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

where N(t) is the number of laptops sold in period t and S(t) is the set of products sold in period t.[25]

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:

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

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.[26]

We convert the average prices defined by (53) and (54) 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:

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

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 | 1.00000 | 1.00000 |
| 2 | 134 | 1.27876 | 1.28041 | 1.03525 | 0.99703 |
| 3 | 147 | 1.27849 | 1.29670 | 1.03503 | 1.00972 |
| 4 | 133 | 1.26150 | 1.28001 | 1.02127 | 0.99538 |
| 5 | 110 | 1.31278 | 1.30992 | 1.06279 | 1.02001 |
| 6 | 95 | 1.31639 | 1.28645 | 1.06571 | 1.00173 |
| 7 | 103 | 1.26883 | 1.26349 | 1.02721 | 0.98386 |
| 8 | 94 | 1.26053 | 1.25112 | 1.02049 | 0.97422 |
| 9 | 83 | 1.24859 | 1.22112 | 1.01082 | 0.95086 |
| 10 | 78 | 1.27961 | 1.27247 | 1.03594 | 0.99085 |
| 11 | 71 | 1.25161 | 1.21663 | 1.01327 | 0.94737 |
| 12 | 72 | 1.17273 | 1.12868 | 0.94941 | 0.87888 |
| 13 | 124 | 1.11517 | 1.08334 | 0.90281 | 0.84358 |
| 14 | 136 | 1.12928 | 1.08597 | 0.91423 | 0.84563 |

[25] Using SHAZAM and the time dummy variables D1-D24, the $PA^t$ can be generated by the following statements: first create a vector of ones of dimension 2639: SMPL 1 2639; GENR ONE=1. The N(t) for T = 1,…,24 can be generated as follows by taking the inner product of the time dummy variables D# with the vector of ones: DO #=1,24 ; MATRIX N#=ONE'D# ; ENDO . Now generate the sum of the prices in month # , SUM#, by taking the inner product of the complete price vector of dimension 2639 P with each time dummy variable, D#: DO #=1,24 ; MATRIX PA#=D#'P ; ENDO. Now form $PA^\#$ as SUM#/N# for #=1,…,24.

[26] Using SHAZAM, the $PUV^t$ can be generated as follows. First generate the vector of expenditures on purchased commodities E by the following statements: SMPL 1 2639 ; GENR E=P*Q. Now inner product the vectors E and Q with the time dummy variables to obtain the numerators and denominators for the $PUV^\#$: DO #=1,24 ; MATRIX NUM# = E'D# ; MATRIX DEN# = Q'D# ; ENDO. Then generate the $PUV^\#$ as NUM#/DEN# for # = 1,24.

| | | | | | |
|---|---|---|---|---|---|
| **15** | **150** | **1.11056** | **1.08594** | **0.89907** | **0.84560** |
| **16** | **135** | **1.15121** | **1.09629** | **0.93198** | **0.85366** |
| **17** | **105** | **1.10092** | **1.03040** | **0.89127** | **0.80235** |
| **18** | **109** | **1.06995** | **1.01540** | **0.86620** | **0.79067** |
| **19** | **107** | **1.05176** | **1.02634** | **0.85147** | **0.79919** |
| **20** | **101** | **1.02677** | **1.01863** | **0.83124** | **0.79319** |
| **21** | **100** | **1.04738** | **0.99001** | **0.84793** | **0.77090** |
| **22** | **91** | **1.11610** | **1.09602** | **0.90356** | **0.85345** |
| **23** | **96** | **1.06155** | **1.08657** | **0.85940** | **0.84609** |
| **24** | **119** | **1.10240** | **1.12772** | **0.89247** | **0.87814** |
| **Mean** | **109.96** | **1.17700** | **1.15970** | **0.95287** | **0.90302** |

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.

**4.2 A Hedonic Regression with Clock Speed as the Only Characteristic.**

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 (38) in the previous section. We start our analysis by regressing the price vector P on the time dummy variables $D_1$.,,,,$D_{24}$ 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.[27] 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 sets the dependent variable vector equal to the logarithms of the product price vector P (which we denote by lnP) and the vectors in the matrix of independent variables are the time dummy variable vectors $D_2$, $D_3$,…,$D_{24}$ and the new 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,

[27] 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).

348 and 154. Thus we have the following linear regression that is a special case of the class of models defined by (38) in the previous section:

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

where e is an error vector of dimension 2639.

We estimated the unknown parameters, $\rho_2^*$, $\rho_3^*$, …, $\rho_{24}^*$, $b_{C1}^*$, …, $b_{C7}^*$ in the linear regression model defined by (51) using ordinary least squares (the OLS command in Shazam). 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 $b_{C1}^*$, …, $b_{C7}^*$ 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.[28]

The estimated $\rho_t^*$ 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 6 of our characteristics listed in the data Appendix into the regression.

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

(57) $\beta^* = \Sigma_{j1}^{7} b_{Cj}^* D_{Cj}$.

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 (56) as follows. First define ONE as a vector of ones of dimension 2639. Consider the following linear regression model:

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

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 (58) 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 (58) by $\rho_2^{**}$, $\rho_3^{**}$, …, $\rho_{24}^{**}$, $b_0^{**}$, $b_{C2}^{**}$, …, $b_{C7}^{**}$. It turns out that $\rho_t^{**} = \rho_t^*$ for t = 2,3,…,24 and the following vector equation also holds:

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

Thus the vector of log quality adjustment factors for the positive observed prices in the sample, $\beta^*$ defined by (57), is also equal to the following expression:

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

---

[28] Of course, these coefficients will change as we add other characteristics to the regression.

In the models which follow, we will add additional characteristics to the hedonic regression model defined by (60) rather than adding addition explanatory variables to the model defined by (56).

**4.3 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): 4096, 8192 and 16384. Construct dummy variable vectors of dimension 2639 for each value of MEM.[29] 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 (58):

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

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.

**4.4 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 the Data Appendix. 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,[30] 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 $D_{S1}, D_{S2}, \ldots, D_{S7}$. 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 characteristic hedonic regression is the following counterpart to (61):

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

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,

---

[29] 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.

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

0.74190 and 0.68987. This regression indicates that purchasers prefer small and large screen sizes over intermediate screen sizes for laptops.

**4.5 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 Data Appendix. 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.[31] 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.

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

$$(63)\ \ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + \Sigma_{j=2}^{5} b_{Pj} D_{Pj} + e.$$

The log of the likelihood function for the hedonic regression defined by (63) 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.

**4.6 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 the Data Appendix 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 (63):

$$(64)\ \ln P = \Sigma_{t=2}^{24} \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7} b_{Cj} D_{Cj} + \Sigma_{j=2}^{3} b_{Mj} D_{Mj} + \Sigma_{j=2}^{7} b_{Sj} D_{Sj} + \Sigma_{j=2}^{5} b_{Pj} D_{Pj} + D_{H2} + e.$$

The log of the likelihood function for the hedonic regression defined by (64) 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

[31] The SHAZAM commands that generated the new groups were as follows: GENR DPI1=DPI1+DPI2+DPI3; GENR DPI2=DPI4; GENR DPI3=DPI5+DPI6; GENR DPI4=DPI7+DPI8+DPI9; GENR DPI5=DPI10.

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.

**4.7 A Hedonic Regression that Added Brand as an Additional Characteristic.**

As indicated above in section 4.1, there are 11 brands in our sample. In the Data Appendix 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:

$$(65)\ \ln P = \Sigma_{t=2}^{24}\, \rho_t D_t + b_0 ONE + \Sigma_{j=2}^{7}\, b_{Cj}\, D_{Cj} + \Sigma_{j=2}^{3}\, b_{Mj} D_{Mj} + \Sigma_{j=2}^{7}\, b_{Sj} D_{Sj} + \Sigma_{j=2}^{5}\, b_{Pj} D_{Pj} + b_{H2} D_{H2} + \Sigma_{j=2}^{11}\, b_{Bj} D_{Bj} + e.$$

The log of the likelihood function for the hedonic regression defined by (65) was 754.295, a 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.

The estimated coefficients on the time dummy variables in this regression are $\rho_2^*$, $\rho_3^*$, …, $\rho_{24}^*$. 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 4 in the following subsection.

The same definitions can be applied to the results of the hedonic regressions defined in sections 4.2-4.6; i.e., use the estimated $\rho_t^*$ generated by these 5 hedonic regressions to define the corresponding (incomplete) Time Dummy Characteristics Price Indexes, which we will denote by $P_C^t$, $P_{CM}^t$, $P_{CMS}^t$, $P_{CMSP}^t$ and $P_{CMSPH}^t$ for the hedonic regression models defined in sections 4.2, 4.3, 4.4, 4.5 and 4.6 respectively. These indexes are also listed in Table 4 below.

**4.8 The Weighted Time Dummy Characteristics Hedonic Regression Model.**

The price indexes defined in sections 4.2-4.7 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 can be constructed using the algebra around equations (46)-(52) in section 3 above.

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 $n\in S(t)$. To obtain the weighted counterpart to the hedonic regression model defined by (64) above, we just form a share vector of dimension 2639 that corresponds to the $\ln p_{tn}$ that appear in (64) 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 (64) by SS to obtain a new linear regression model which again provides estimates for the unknown parameters that appear in (64). The $R^2$ for this new weighted regression model turned out to be 0.8915 which is substantially higher than the $R^2$ for the counterpart unweighted model which was 0.8631.

The parameter estimates for this weighted hedonic regression model are listed in Table 3 below. This is our preferred regression from all of the regression models that have been presented thus far.

**Table 3: Parameter Estimates for the Weighted Time Dummy Characteristics Hedonic Regression**

| Coef | Estimate | Std. Error | T Stat | Coef | Estimate | Std. Error | T Stat |
|---|---|---|---|---|---|---|---|
| $b_0^*$ | -1.1981 | 0.03714 | -32.260 | $b_{C5}^*$ | 0.2919 | 0.01477 | 19.760 |
| $\rho_2^*$ | 0.0156 | 0.01791 | 0.870 | $b_{C6}^*$ | 0.2495 | 0.01661 | 15.020 |
| $\rho_3^*$ | 0.0299 | 0.01797 | 1.662 | $b_{C7}^*$ | 0.3400 | 0.01798 | 18.910 |
| $\rho_4^*$ | 0.0321 | 0.01805 | 1.776 | $b_{M2}^*$ | 0.2393 | 0.01017 | 23.540 |
| $\rho_5^*$ | 0.0224 | 0.01803 | 1.245 | $b_{M3}^*$ | 0.5720 | 0.01687 | 33.900 |
| $\rho_6^*$ | 0.0079 | 0.01809 | 0.439 | $b_{S2}^*$ | 0.3568 | 0.03430 | 10.400 |
| $\rho_7^*$ | -0.0200 | 0.01813 | -1.104 | $b_{S3}^*$ | 0.4556 | 0.03246 | 14.040 |
| $\rho_8^*$ | -0.0235 | 0.01818 | -1.296 | $b_{S4}^*$ | 0.2590 | 0.03266 | 7.929 |
| $\rho_9^*$ | -0.0336 | 0.01823 | -1.841 | $b_{S5}^*$ | 0.3045 | 0.03150 | 9.665 |
| $\rho_{10}^*$ | -0.0260 | 0.01824 | -1.427 | $b_{S6}^*$ | 0.4730 | 0.04071 | 11.620 |
| $\rho_{11}^*$ | -0.0540 | 0.01827 | -2.958 | $b_{S7}^*$ | 0.5134 | 0.03508 | 14.640 |
| $\rho_{12}^*$ | -0.0884 | 0.01831 | -4.829 | $b_{P2}^*$ | 0.1488 | 0.01320 | 11.270 |
| $\rho_{13}^*$ | -0.0986 | 0.01833 | -5.383 | $b_{P3}^*$ | 0.4560 | 0.03566 | 12.790 |
| $\rho_{14}^*$ | -0.1042 | 0.01834 | -5.679 | $b_{P4}^*$ | 0.7055 | 0.04659 | 15.140 |
| $\rho_{15}^*$ | -0.0954 | 0.01845 | -5.167 | $b_{P5}^*$ | 0.5220 | 0.03061 | 17.050 |
| $\rho_{16}^*$ | -0.0765 | 0.01850 | -4.136 | $b_{H2}^*$ | 0.2996 | 0.02048 | 14.630 |
| $\rho_{17}^*$ | -0.0870 | 0.01859 | -4.680 | $b_{B2}^*$ | -0.2059 | 0.02512 | -8.197 |
| $\rho_{18}^*$ | -0.0974 | 0.01863 | -5.229 | $b_{B3}^*$ | 0.0021 | 0.03626 | 0.057 |
| $\rho_{19}^*$ | -0.0937 | 0.01873 | -5.003 | $b_{B4}^*$ | -0.0575 | 0.02363 | -2.435 |
| $\rho_{20}^*$ | -0.1110 | 0.01871 | -5.932 | $b_{B5}^*$ | -0.0618 | 0.14650 | -0.422 |
| $\rho_{21}^*$ | -0.1233 | 0.01870 | -6.593 | $b_{B6}^*$ | 0.3191 | 0.02316 | 13.780 |
| $\rho_{22}^*$ | -0.1174 | 0.01871 | -6.276 | $b_{B7}^*$ | 0.2144 | 0.02375 | 9.027 |
| $\rho_{23}^*$ | -0.1028 | 0.01877 | -5.477 | $b_{B8}^*$ | 0.0306 | 0.02984 | 1.025 |
| $\rho_{24}^*$ | -0.0823 | 0.01872 | -4.394 | $b_{B9}^*$ | 0.3261 | 0.02414 | 13.510 |
| $b_{C2}^*$ | 0.1565 | 0.01219 | 12.840 | $b_{B10}^*$ | 0.1684 | 0.03378 | 4.985 |
| $b_{C3}^*$ | 0.2821 | 0.01447 | 19.490 | $b_{B11}^*$ | 0.5110 | 0.15800 | 3.235 |
| $b_{C4}^*$ | 0.2301 | 0.01399 | 16.450 | | | | |

There are 53 parameters in this regression model that are estimated with 2586 degrees of freedom for the error terms. It can be seen that the clock speed parameters $b_{Cj}^*$ are only weakly increasing with respect to j; the memory capacity parameters $b_{M2}^*$ and $b_{M3}^*$ are monotonically increasing; the screen size parameters $b_{Sj}^*$ exhibit a U shaped pattern; the pixel parameters $b_{Pj}^t$ are monotonically increasing; 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 weakly preferred by purchasers.

The estimated coefficients on the time dummy variables in this regression are $\rho_2^*$, $\rho_3^*$, …, $\rho_{24}^*$. 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 4 (and plotted in Chart 1 below) 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 4 along with the unweighted Time Dummy Characteristics Indexes that are based on the regression models explained in sections 4.2-4.6. ($P_C^t$, $P_{CM}^t$, $P_{CMS}^t$, $P_{CMSP}^t$ and $P_{CMSPH}^t$). For comparison purposes, we also list the simple average laptop price indexes $P_A^t$ and $P_{UV}^t$ defined by definitions (55) in section 4.1.

**Table 4: Weighted and Unweighted Time Product Dummy Price Indexes**

| Month t | $P_{WTDC}^t$ | $P_{TDC}^t$ | $P_{CMSPH}^t$ | $P_{CMSP}^t$ | $P_{CMS}^t$ | $P_{CM}^t$ | $P_C^t$ | $P_A^t$ | $P_{UV}^t$ |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| 2 | 1.01571 | 1.03561 | 1.02620 | 1.02367 | 1.03230 | 1.01802 | 1.04123 | 1.03525 | 0.99703 |
| 3 | 1.03031 | 1.04665 | 1.03749 | 1.03260 | 1.03625 | 1.04575 | 1.09513 | 1.03503 | 1.00972 |
| 4 | 1.03257 | 1.03888 | 1.01851 | 1.01209 | 1.01869 | 1.03329 | 1.07238 | 1.02127 | 0.99538 |
| 5 | 1.02270 | 1.08280 | 1.08117 | 1.08253 | 1.08039 | 1.09031 | 1.15033 | 1.06279 | 1.02001 |
| 6 | 1.00797 | 1.07931 | 1.08333 | 1.08702 | 1.08707 | 1.10019 | 1.16008 | 1.06571 | 1.00173 |
| 7 | 0.98019 | 1.02240 | 1.02998 | 1.03049 | 1.03178 | 1.02851 | 1.09930 | 1.02721 | 0.98386 |
| 8 | 0.97673 | 1.02372 | 1.03536 | 1.03810 | 1.03602 | 1.03931 | 1.10055 | 1.02049 | 0.97422 |
| 9 | 0.96699 | 1.00763 | 1.01763 | 1.02219 | 1.02510 | 1.02037 | 1.08231 | 1.01082 | 0.95086 |
| 10 | 0.97431 | 1.02289 | 1.03329 | 1.03757 | 1.03760 | 1.03905 | 1.12498 | 1.03594 | 0.99085 |
| 11 | 0.94739 | 0.99707 | 1.00181 | 1.00575 | 1.00859 | 1.02131 | 1.11137 | 1.01327 | 0.94737 |
| 12 | 0.91540 | 0.94035 | 0.93111 | 0.93514 | 0.93850 | 0.94626 | 1.02127 | 0.94941 | 0.87888 |
| 13 | 0.90607 | 0.96932 | 0.91955 | 0.91411 | 0.91098 | 0.87076 | 0.95127 | 0.90281 | 0.84358 |
| 14 | 0.90108 | 0.95629 | 0.90833 | 0.90348 | 0.90146 | 0.86859 | 0.96108 | 0.91423 | 0.84563 |
| 15 | 0.90905 | 0.94247 | 0.89198 | 0.88531 | 0.88158 | 0.85448 | 0.93678 | 0.89907 | 0.84560 |
| 16 | 0.92634 | 0.95733 | 0.91131 | 0.89907 | 0.89222 | 0.86409 | 0.96173 | 0.93198 | 0.85366 |
| 17 | 0.91669 | 0.95014 | 0.89575 | 0.87694 | 0.87007 | 0.83104 | 0.90118 | 0.89127 | 0.80235 |
| 18 | 0.90717 | 0.94491 | 0.87540 | 0.85854 | 0.85243 | 0.80523 | 0.87761 | 0.86620 | 0.79067 |
| 19 | 0.91053 | 0.94595 | 0.86200 | 0.83793 | 0.82751 | 0.77520 | 0.82961 | 0.85147 | 0.79919 |
| 20 | 0.89493 | 0.92595 | 0.84228 | 0.82701 | 0.80855 | 0.75867 | 0.81446 | 0.83124 | 0.79319 |
| 21 | 0.88399 | 0.92104 | 0.84667 | 0.83211 | 0.81405 | 0.76625 | 0.82925 | 0.84793 | 0.77090 |
| 22 | 0.88920 | 0.92314 | 0.88356 | 0.86600 | 0.84461 | 0.80207 | 0.87828 | 0.90356 | 0.85345 |
| 23 | 0.90231 | 0.93081 | 0.88640 | 0.86528 | 0.84447 | 0.78950 | 0.83986 | 0.85940 | 0.84609 |
| 24 | 0.92102 | 0.91645 | 0.86613 | 0.85195 | 0.82916 | 0.77719 | 0.85181 | 0.89247 | 0.87814 |
| Mean | 0.94744 | 0.98255 | 0.95355 | 0.94687 | 0.94206 | 0.92273 | 0.98716 | 0.95287 | 0.90302 |

The results in Table 4 and Chart 1 are not very plausible. Our preferred hedonic index, $P_{WTDC}^t$, ends up at 0.92101 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 0.89247 and 0.87814). It seems unlikely that a quality adjusted price index for laptops could end up *higher* than a simple average price index for laptops. *The above results also show that missing characteristics can greatly affect the resulting hedonic price index*.

Although the weighted and unweighted time product characteristic indexes end up fairly close to each other in month 24 (0.92102 for the weighted index and 0.91645 for the unweighted hedonic index), there are substantial month to month differences between the two indexes. Moreover the mean of the weighted indexes $P_{WTPC}^t$ (0.94744) is substantially below the mean of the unweighted indexes $P_{TPC}^t$ (0.98255). 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.

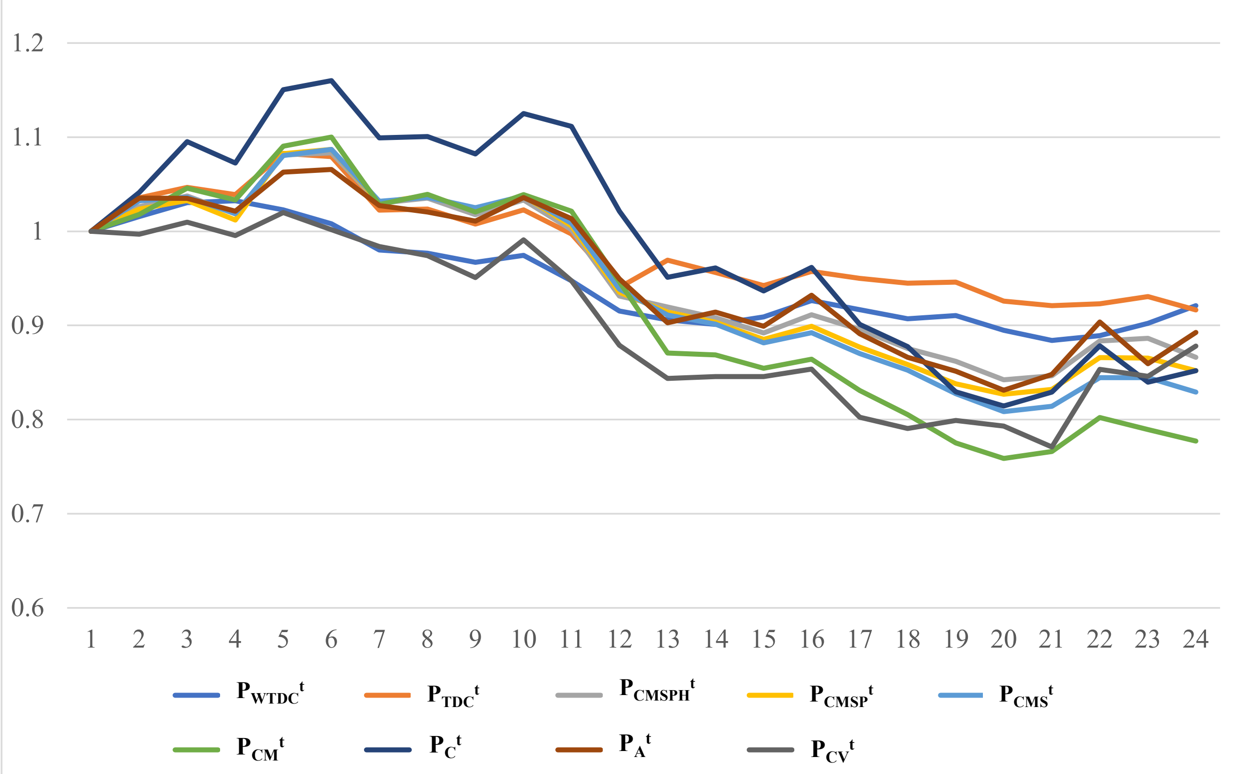



**Chart 1: Weighted and Unweighted Time Product Dummy Price Indexes**

**4.9 Direct and Indirect Weighted Time Dummy Characteristics Price Indexes.**

In this section, we will illustrate the relationship between *direct and indirect price levels* that can be derived from the hedonic regression described in section 4.8. We will use the results around equations (42)-(52) in section 3.

In section 4.8, we defined the estimated direct monthly price levels, $\pi_t^*$, by exponentiating the estimated coefficients $\rho_t^*$. Define the month t *direct price level* $P^{t*}$ as follows:

(66) $P^{t*} \equiv \pi_t^* = P_{WTDC}^t$ ; t = 1,…,24.

Because $\pi_1^* = 1$, the directly estimated monthly price levels $P^{t*}$ also equal the corresponding Weighted Time Dummy Characteristics price indexes, $P_{WTDC}^t$, which are listed in Table 4 above.

Define month t *total expenditures* (or sales) of laptops in our sample, $e^t$, as follows:

(67) $e^t \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}$ ; t = 1,..,24.

The (indirectly) estimated *aggregate quantity level* for month t, $Q^{t*}$, is defined by deflating month t expenditures $e^t$ by $P^{t*}$:

(68) $Q^{t*} \equiv e^t/P^{t*}$ ; t = 1,…,24.

$P^{t*}$, $e^t$ and $Q^{t*}$ are listed in Table 5 below.

We now show how the parameter estimates listed in Table 4 above can be used to form monthly direct aggregate quantity indexes $Q^{t**}$ for each month t. First, form the vector of dimension 2639 of *logarithms of the product quality adjustment parameters* $\beta^*$ as follows:

(69) $\beta^* \equiv b_0^*\mathrm{ONE} + \Sigma_{j=2}^{7}\, b_{Cj}^* D_{Cj} + \Sigma_{j=2}^{3}\, b_{Mj}^* D_{Mj} + \Sigma_{j=2}^{7}\, b_{Sj}^* D_{Sj} + \Sigma_{j=2}^{5}\, b_{Pj}^* D_{Pj} + b_{H2}^* D_{H2} + \Sigma_{j=2}^{11}\, b_{Bj}^* D_{Bj}$.

Denote the component of $\beta^*$ that corresponds to product n sold in month t by $\beta_{tn}^*$ for t = 1,…,24 and $n \in S(t)$. Define the quality adjustment parameter for purchased product n in period t, $\alpha_{tn}^*$, by exponentiating $\beta_{tn}^*$:

(70) $\alpha_{tn}^* \equiv \exp[\beta_{tn}^*]$ ; $\qquad$ t = 1,…,24; $n \in S(t)$.

Using the above quality adjustment parameters $\alpha_{tn}^*$, we can form a month t *direct estimate for the aggregate quantity* or *utility* obtained by purchasers during period t:

(71) $Q^{t**} \equiv \Sigma_{n\in S(t)}\, \alpha_{tn}^* q_{tn}$ ; $\qquad$ t = 1,...,24.

The corresponding month t *indirect price level*, $P^{t**}$, is defined by deflating month t expenditure $e^t$ by the month t aggregate quantity $Q^{t**}$:

(72) $P^{t**} \equiv e^t/Q^{t**} = \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_{tn}^* q_{tn}$ ; $\qquad$ t = 1,…,24.

The price and quantity level series, $P^{t**}$ and $Q^{t**}$, are listed in Table 5 below. It can be seen $P^{t*}$, $P^{t**}$, $Q^{t*}$ and $Q^{t**}$ satisfy the de Haan inequalities (52); i.e., these series satisfy the following inequalities:

(73) $P^{t**} \le P^{t*}$ and $Q^{t**} \ge Q^{t*}$ ; $\qquad$ t = 1,…,24.

If the $R^2$ for the weighted hedonic regression defined in section 4.8 were equal to 1, then the direct and indirectly defined monthly price and quantity levels would coincide; i.e., we would have $P^{t**} = P^{t*}$ and $Q^{t**} = Q^{t*}$ for t = 1,…,24.

The indirectly defined price *level* series, $P^{t**}$, can be turned into the *Weighted Time Dummy Characteristics Price Index* series, $P_{WTPC}^t$, by dividing the $P^{t**}$ by $P^{1**}$:

(74) $P_{IWTPC}^t \equiv P^{t**}/P^{1**}$ ; $\qquad$ t = 1,…,24.

The series $P_{WTPC}^t$ is also listed in Table 5.

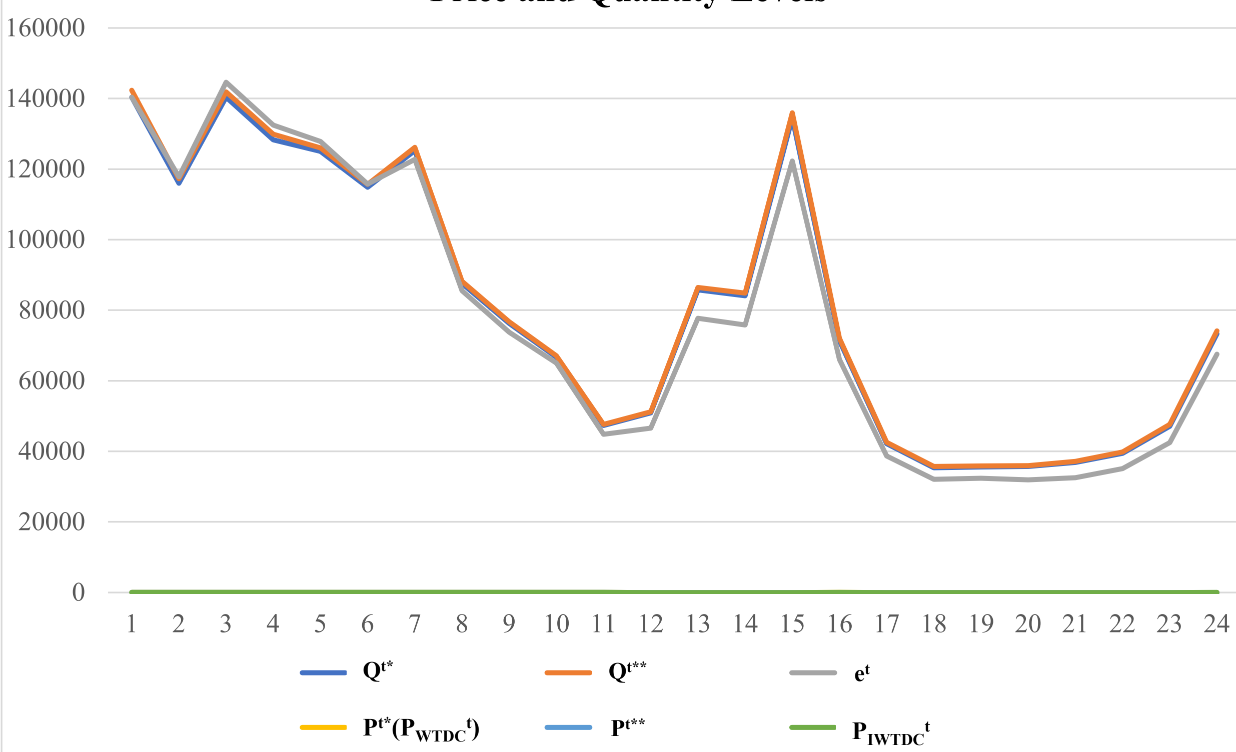
**Table 5: Direct and Indirect Weighted Time Dummy Characteristics Price and Quantity Levels**

| Month t | $Q^{t*}$ | $Q^{t**}$ | $e^t$ | $P^{t*}$ ($P_{WTDC}^t$) | $P^{t**}$ | $P_{IWTDC}^t$ |
|---|---|---|---|---|---|---|
| 1 | 140388 | 142306 | 140388 | 1.00000 | 0.98653 | 1.00000 |
| 2 | 115958 | 117271 | 117780 | 1.01571 | 1.00434 | 1.01806 |
| 3 | 140351 | 141842 | 144604 | 1.03031 | 1.01948 | 1.03340 |
| 4 | 128314 | 129847 | 132494 | 1.03257 | 1.02039 | 1.03433 |

| | | | | | | |
|---|---|---|---|---|---|---|
| **5** | **125022** | **126026** | **127860** | **1.02270** | **1.01455** | **1.02841** |
| **6** | **114803** | **115637** | **115717** | **1.00797** | **1.00069** | **1.01436** |
| **7** | **125235** | **126134** | **122755** | **0.98019** | **0.97321** | **0.98650** |
| **8** | **87567** | **88148** | **85529** | **0.97673** | **0.97028** | **0.98354** |
| **9** | **76291** | **76718** | **73773** | **0.96699** | **0.96161** | **0.97474** |
| **10** | **66703** | **67084** | **64990** | **0.97431** | **0.96879** | **0.98202** |
| **11** | **47313** | **47594** | **44824** | **0.94739** | **0.94181** | **0.95468** |
| **12** | **50869** | **51213** | **46566** | **0.91540** | **0.90925** | **0.92167** |
| **13** | **85751** | **86402** | **77696** | **0.90607** | **0.89924** | **0.91152** |
| **14** | **84089** | **84823** | **75771** | **0.90108** | **0.89329** | **0.90549** |
| **15** | **134545** | **135966** | **122309** | **0.90905** | **0.89955** | **0.91184** |
| **16** | **71296** | **72011** | **66044** | **0.92634** | **0.91713** | **0.92966** |
| **17** | **42172** | **42550** | **38659** | **0.91669** | **0.90855** | **0.92096** |
| **18** | **35359** | **35711** | **32077** | **0.90717** | **0.89822** | **0.91048** |
| **19** | **35549** | **35853** | **32369** | **0.91053** | **0.90282** | **0.91515** |
| **20** | **35699** | **35957** | **31948** | **0.89493** | **0.88851** | **0.90065** |
| **21** | **36822** | **37186** | **32550** | **0.88399** | **0.87535** | **0.88730** |
| **22** | **39437** | **39776** | **35067** | **0.88920** | **0.88161** | **0.89366** |
| **23** | **47104** | **47636** | **42502** | **0.90231** | **0.89222** | **0.90441** |
| **24** | **73319** | **74114** | **67528** | **0.92102** | **0.91114** | **0.92358** |
| **Mean** | **80832** | **81575** | **77992** | **0.94744** | **0.93911** | **0.95193** |

It can be seen that the direct and indirectly defined Weighted Time Dummy Characteristic Price Indexes, $P_{WTPC}^{t}$ and $P_{IWTPC}^{t}$, are fairly close to each other but both indexes seem to end up implausibly high; see Chart 2. Thus in the following section, we will implement adjacent period hedonic regressions using the same 6 characteristics.

**Chart 2: Direct and Indirect Weighted Time Dummy Characteristics Price and Quantity Levels**



**5. Adjacent Period Characteristics Hedonic Regression Models.**

There are two problems with our "best" directly defined weighted hedonic price index using characteristics, $P_{WTPC}{}^t$, which was defined in the previous section:

- It is not a real time index; i.e., it is a retrospective index that is calculated using the data covering two years;[32]
- It does not allow for gradual taste change on the part of purchasers.

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.[33] 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 6 characteristics of the products that were used in section 4.7 above. The counterpart regression to the unweighted time dummy characteristic hedonic regression defined by (65) in section 4.7 becomes the following regression model:

$$(75)\ \ln P = \rho_2 D_2 + b_0 ONE + \Sigma_{j=2}{}^7\, b_{Cj}\, D_{Cj} + \Sigma_{j=2}{}^3\, b_{Mj} D_{Mj} + \Sigma_{j=2}{}^7\, b_{Sj} D_{Sj} + \Sigma_{j=2}{}^5\, b_{Pj} D_{Pj} + b_{H2} D_{H2} + \Sigma_{j=2}{}^{11}\, b_{Bj} D_{Bj} + e$$

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 (75) 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 75 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 (75) 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.[34]

---

[32] This difficulty can be overcome by using rolling window hedonic regressions. See Diewert and Fox (2022; 360-361) for a discussion of the issues surrounding linking the results from a new panel of data with the results from a previous panel.

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

[34] 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 (75) 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 (75) 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 (75) were replaced by the terms $\Sigma_{j=2}{}^9\, b_{Bj} D_{Bj}$. 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.

Define $P^{1*} \equiv 1$ as the month 1 index level. Define $\rho_2^*$ as the estimated month 2 time dummy coefficient for the bilateral regression defined by (75)[35] and define $\pi_2^*$ as the exponential of $\rho_2^*$; 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 (75) except the time dummy term $\rho_2 D_2$ on the right hand side of (75) was replaced with the term $\rho_3 D_3$ where $D_3$ takes on the value 1 for the products sold in month 3 and the value 0 for the products sold in month 2. Once $\rho_3^*$ 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 $\rho_{24}^*$. Once $\rho_{24}^*$ 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 Product* (Unweighted) *Characteristics Price Index* for month t, $P_{ATPC}^t$, was defined as follows:

$$(76)\quad P_{ATPC}^t \equiv P^{t*}/P^{1*}\,; \qquad t = 1,\ldots,24.$$

The price index defined by (76) 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 (75) 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 (75). 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 $\pi_t^*$ for t = 2,3,…,24. Now follow the same steps as were made in the paragraphs above definitions (76) in order to define the *Weighted Adjacent Period Time Product Characteristics Price Index* for month t, $P_{WATPC}^t$, for t = 1,2,…,24. This index along with its unweighted (or equally weighted) counterpart index, $P_{ATPC}^t$, are listed in Table 6 below. For comparison purposes, Table 6 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 $P_{UV}^t$.[36] See Chart 3 for plots of the indexes listed in Table 6.

**Table 6: Sample Wide and Adjacent Period Weighted and Unweighted Characteristics Price Indexes**

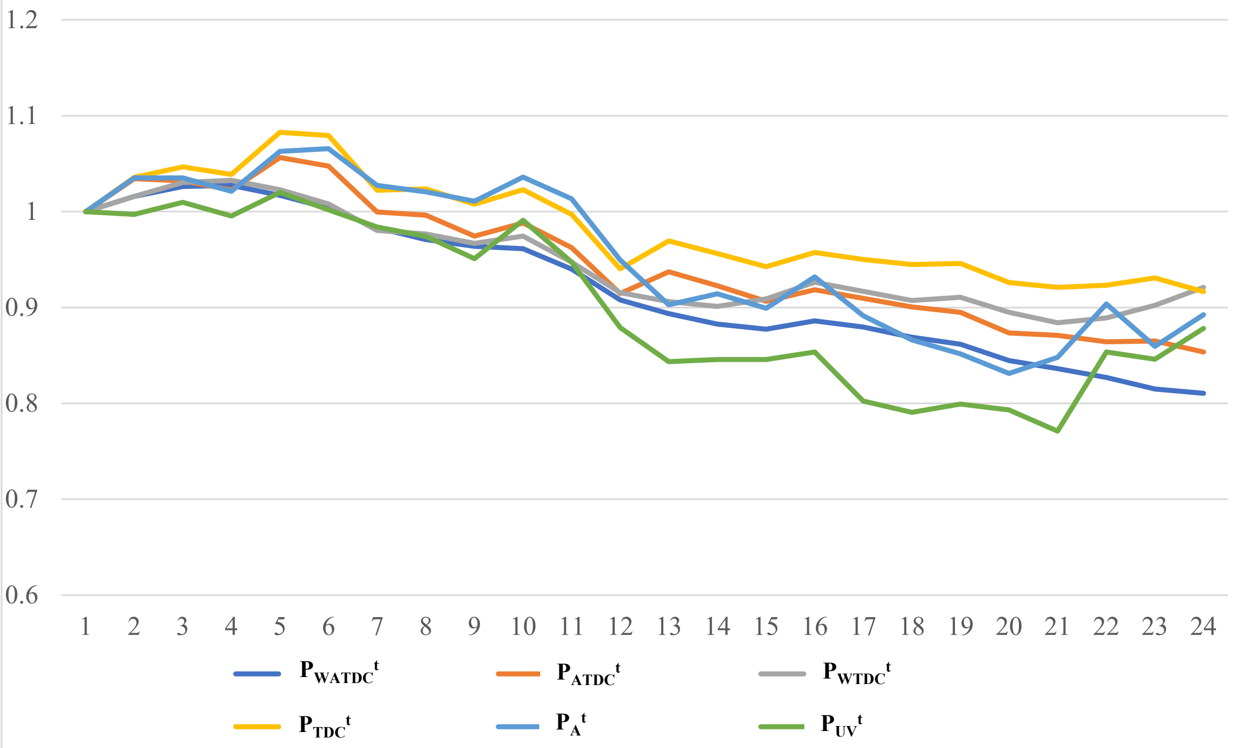
| Month t | $P_{WATDC}^t$ | $P_{ATDC}^t$ | $P_{WTDC}^t$ | $P_{TDC}^t$ | $P_A^t$ | $P_{UV}^t$ |
|---|---|---|---|---|---|---|
| 1 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| 2 | 1.01597 | 1.03434 | 1.01571 | 1.03561 | 1.03525 | 0.99703 |
| 3 | 1.02612 | 1.03214 | 1.03031 | 1.04665 | 1.03503 | 1.00972 |
| 4 | 1.02732 | 1.02268 | 1.03257 | 1.03888 | 1.02127 | 0.99538 |
| 5 | 1.01684 | 1.05650 | 1.02270 | 1.08280 | 1.06279 | 1.02001 |
| 6 | 1.00363 | 1.04757 | 1.00797 | 1.07931 | 1.06571 | 1.00173 |

[35] Taking into account the reduction in the number of cells for the various characteristics, (75) became: $\ln P = \rho_2 D_2 + b_0 ONE + \Sigma_{j=2}^{6}\, b_{Cj}\, D_{Cj} + \Sigma_{j=2}^{3}\, b_{Mj} D_{Mj} + \Sigma_{j=2}^{4}\, b_{Sj} D_{Sj} + \Sigma_{j=2}^{4}\, b_{Pj} D_{Pj} + b_{H2} D_{H2} + \Sigma_{j=2}^{9}\, b_{Bj} D_{Bj} + e$.

[36] These latter four indexes were listed in Table 4.

| | | | | | | |
|---|---|---|---|---|---|---|
| **7** | **0.98301** | **0.99975** | **0.98019** | **1.02240** | **1.02721** | **0.98386** |
| **8** | **0.97090** | **0.99619** | **0.97673** | **1.02372** | **1.02049** | **0.97422** |
| **9** | **0.96368** | **0.97454** | **0.96699** | **1.00763** | **1.01082** | **0.95086** |
| **10** | **0.96133** | **0.98820** | **0.97431** | **1.02289** | **1.03594** | **0.99085** |
| **11** | **0.94000** | **0.96227** | **0.94739** | **0.99707** | **1.01327** | **0.94737** |
| **12** | **0.90779** | **0.91460** | **0.91540** | **0.94035** | **0.94941** | **0.87888** |
| **13** | **0.89365** | **0.93709** | **0.90607** | **0.96932** | **0.90281** | **0.84358** |
| **14** | **0.88269** | **0.92254** | **0.90108** | **0.95629** | **0.91423** | **0.84563** |
| **15** | **0.87733** | **0.90649** | **0.90905** | **0.94247** | **0.89907** | **0.84560** |
| **16** | **0.88593** | **0.91854** | **0.92634** | **0.95733** | **0.93198** | **0.85366** |
| **17** | **0.87962** | **0.90962** | **0.91669** | **0.95014** | **0.89127** | **0.80235** |
| **18** | **0.86894** | **0.90062** | **0.90717** | **0.94491** | **0.86620** | **0.79067** |
| **19** | **0.86163** | **0.89505** | **0.91053** | **0.94595** | **0.85147** | **0.79919** |
| **20** | **0.84450** | **0.87334** | **0.89493** | **0.92595** | **0.83124** | **0.79319** |
| **21** | **0.83613** | **0.87088** | **0.88399** | **0.92104** | **0.84793** | **0.77090** |
| **22** | **0.82692** | **0.86431** | **0.88920** | **0.92314** | **0.90356** | **0.85345** |
| **23** | **0.81487** | **0.86516** | **0.90231** | **0.93081** | **0.85940** | **0.84609** |
| **24** | **0.81055** | **0.85353** | **0.92102** | **0.91645** | **0.89247** | **0.87814** |
| **Mean** | **0.92081** | **0.94775** | **0.94744** | **0.98255** | **0.95287** | **0.90302** |

**Chart 3: Sample Wide and Adjacent Period Weighted and Unweighted Characteristics Price Indexes**




It can be seen that the adjacent period equally weighted characteristics index $P_{ATDC}^t$ finishes above its weighted counterpart $P_{WATDC}^t$ for t = 24 and on average, $P_{ATDC}^t$ is 2.7 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 $P_{WATDC}^t$. Although $P_{WATDC}^t$. index finishes substantially below the month 24 Unit Value Price Index $P_{UV}^{24}$, we note that the average of the

$P_{WATDC}^t$ is 0.92081, which is substantially higher than the average of the Unit Value Price Index $P_{UV}^t$. Thus it seems that the quality adjustment provided by the quality adjusted indexes exhibited thus far is incomplete.

Here are some of the advantages and disadvantages of the Weighted Adjacent Period Time Dummy Characteristics indexes $P_{WATDC}^t$ over the Weighted Time Dummy Characteristics indexes $P_{WTPC}^t$:

- 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.
- If the number of characteristics is large relative to the number of observations in a bilateral regression, the estimates for the quality adjustment parameters could be unreliable which could lead to unreliable estimates for the price levels.
- 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 this fact.

The results presented in sections 4 and 5 of this paper indicate that missing characteristics can have a material effect on the price index. A model that includes all possible product characteristics[37] 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.

**6. Time Product Dummy Variable 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 $\rho_t = 0$. The corresponding equally weighted Unweighted Time Product Dummy least squares minimization problem is defined by (20) with all expenditure shares $s_{tn}$ set equal to 1.

In order to set up the unweighted regression problem for our particular application, we make us of the vectors of time dummy variables, $D_1, \ldots, D_{24}$, which were defined in section 4.1 above. This section also defined 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 previous sections. Then the unweighted Time Product Dummy regression model can be expressed as the following estimating equation for the log price levels $\rho_2, \rho_3, \ldots, \rho_{24}$ and the 366 product log quality adjustment factors $\beta_1, \beta_2, \ldots, \beta_{366}$:

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

The $R^2$ for the above regression turned out to be 0.9836. We set $\rho_t^*$ equal to one. The estimated $\rho_t^*$ 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 7 below.

[37] There may be external environmental factors (that change over time) which affect the utility to purchasers of the products in scope. Also, the "newness" or "oldness" of a product may affect purchaser utility. A fashion product is a product whose utility falls due to the length of time the product has been available in the marketplace. A proven or reliable product is a product whose utility rises as the length of time it has been available for purchase. It is possible to measure this effect using a hedonic regression approach but we do not include the "newness" of a product as a price determining characteristic in this study.

To obtain the Weighted Time Product Dummy Price Indexes, multiply the vectors on both sides of (77) (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 $\rho_t^*$ equal to one. The estimated $\rho_t^*$ 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 7 below.

As in the previous section, 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 (77) that links the prices of months 1 and 2 is the following regression model:

$$(78)\ \ln P^* = \rho_2 D_2{}^* + \Sigma_{k=1}{}^{366}\, \beta_k D_{Jk}{}^* + e^t$$
$$= \rho_2 D_2{}^* + \Sigma_{k\in S(1,2)}\, \beta_k D_{Jk}{}^* + e^t$$

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 only defined for products n that were actually sold in periods 1 and 2. The first vector equation in (78) 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 (78) sums over the nonzero product dummy vectors.[38]

In any case, 23 unweighted bilateral time product dummy variable regressions were run and the estimated $\rho_t^*$ were converted into $\pi_t^*$ and the $\pi_t^*$ 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 7 below.[39]

As usual, 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.

Table 7 lists the Adjacent Period Weighted and Unweighted Time Product Dummy price indexes, $P_{WATPD}{}^t$ and $P_{ATPC}{}^t$, as well as the simple average and unit value price indexes, $P_A{}^t$ and $P_{UV}{}^t$.[40] Chart 4 plots the indexes listed in Table 7.

**Table 7: Sample Wide and Adjacent Period Weighted and Unweighted Time Product Dummy Price Indexes**

| Month t | $P_{WATPD}{}^t$ | $P_{ATPD}{}^t$ | $P_{WTPD}{}^t$ | $P_{TPD}{}^t$ | $P_A{}^t$ | $P_{UV}{}^t$ |
|---|---|---|---|---|---|---|

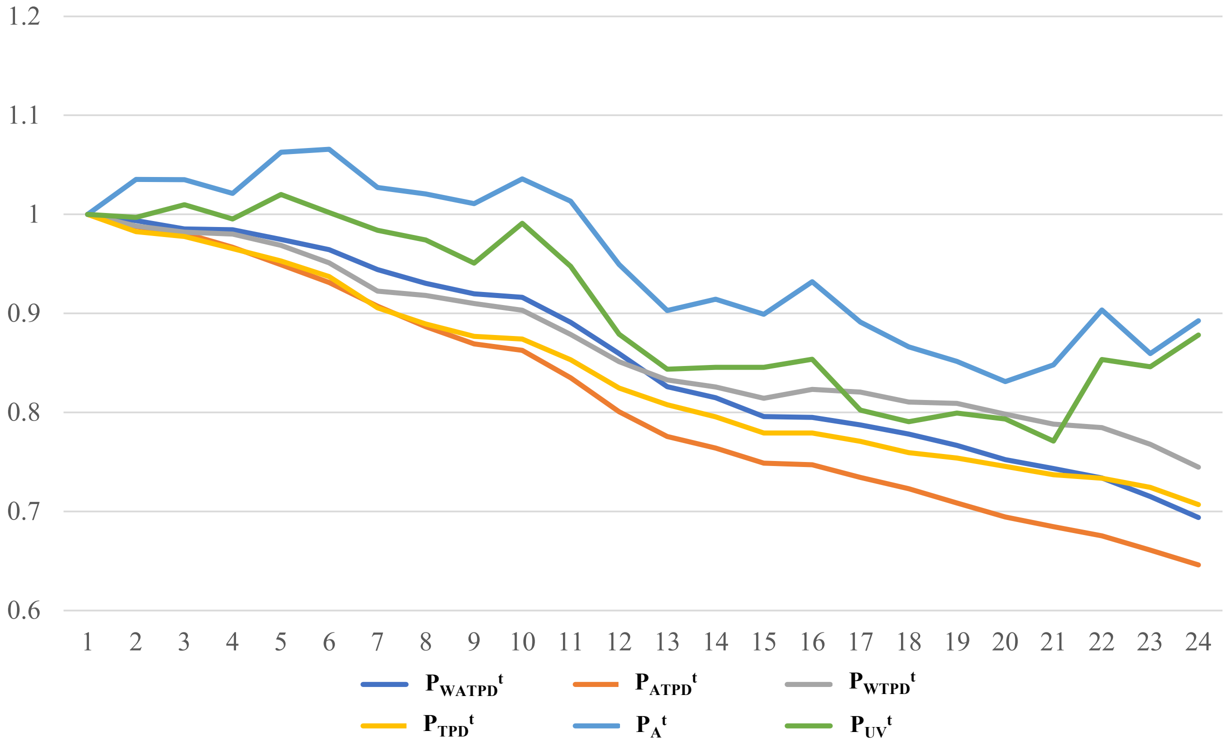
[38] It turned out to be somewhat difficult to go from line 1 in (78) to line 2 in (78). However, it is possible to construct programs that overcome these difficulties.

[39] The $R^2$ for the 23 bilateral 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.

[40] These latter four indexes were also listed in Table 4.

| | | | | | | |
|---|---|---|---|---|---|---|
| **1** | **1.00000** | **1.00000** | **1.00000** | **1.00000** | **1.00000** | **1.00000** |
| **2** | **0.99358** | **0.98781** | **0.98828** | **0.98257** | **1.03525** | **0.99703** |
| **3** | **0.98526** | **0.98084** | **0.98205** | **0.97768** | **1.03503** | **1.00972** |
| **4** | **0.98456** | **0.96681** | **0.98006** | **0.96541** | **1.02127** | **0.99538** |
| **5** | **0.97476** | **0.94903** | **0.96878** | **0.95302** | **1.06279** | **1.02001** |
| **6** | **0.96444** | **0.93115** | **0.95087** | **0.93711** | **1.06571** | **1.00173** |
| **7** | **0.94422** | **0.90729** | **0.92250** | **0.90572** | **1.02721** | **0.98386** |
| **8** | **0.93034** | **0.88649** | **0.91801** | **0.88931** | **1.02049** | **0.97422** |
| **9** | **0.91971** | **0.86908** | **0.90983** | **0.87676** | **1.01082** | **0.95086** |
| **10** | **0.91611** | **0.86254** | **0.90323** | **0.87407** | **1.03594** | **0.99085** |
| **11** | **0.89088** | **0.83488** | **0.87881** | **0.85326** | **1.01327** | **0.94737** |
| **12** | **0.85948** | **0.80071** | **0.85129** | **0.82468** | **0.94941** | **0.87888** |
| **13** | **0.82589** | **0.77569** | **0.83276** | **0.80777** | **0.90281** | **0.84358** |
| **14** | **0.81473** | **0.76387** | **0.82554** | **0.79541** | **0.91423** | **0.84563** |
| **15** | **0.79577** | **0.74871** | **0.81431** | **0.77924** | **0.89907** | **0.84560** |
| **16** | **0.79492** | **0.74716** | **0.82328** | **0.77927** | **0.93198** | **0.85366** |
| **17** | **0.78726** | **0.73419** | **0.82048** | **0.77078** | **0.89127** | **0.80235** |
| **18** | **0.77805** | **0.72286** | **0.81037** | **0.75921** | **0.86620** | **0.79067** |
| **19** | **0.76665** | **0.70844** | **0.80906** | **0.75392** | **0.85147** | **0.79919** |
| **20** | **0.75214** | **0.69445** | **0.79830** | **0.74549** | **0.83124** | **0.79319** |
| **21** | **0.74318** | **0.68464** | **0.78818** | **0.73698** | **0.84793** | **0.77090** |
| **22** | **0.73369** | **0.67542** | **0.78460** | **0.73339** | **0.90356** | **0.85345** |
| **23** | **0.71498** | **0.66085** | **0.76781** | **0.72413** | **0.85940** | **0.84609** |
| **24** | **0.69385** | **0.64587** | **0.74478** | **0.70698** | **0.89247** | **0.87814** |
| **Mean** | **0.85685** | **0.81411** | **0.86972** | **0.83884** | **0.95287** | **0.90302** |

**Chart 4: Sample Wide and Adjacent Period Weighted and Unweighted Time Product Dummy Price Indexes**



As usual, 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. Finally, we prefer the Adjacent Period Weighted Time Product Dummy Indexes $P_{WATPD}{}^t$ over their single regression counterpart indexes, the Weighted Time Product Dummy Indexes $P_{WTPF}{}^t$ for two reasons: (i) the regressions which generate the $P_{WATPD}{}^t$ fit the data much better than the single regression which generated the $P_{WTPD}{}^t$ and (ii) the $P_{WATPD}{}^t$ appear to be s biy smoother than the $P_{WTPD}{}^t$. $P_{WATPD}{}^t$ is our preferred index thus far.

Our preferred index, the Adjacent Period Weighted Time Product Dummy Index $P_{WATPD}{}^t$, is a chained index and thus, it is subject to possible chain drift.[41] In order to reduce or eliminate possible chain drift, in the following section we will calculate Predicted Share Price Similarity linked indexes as well as some traditional indexes.

**7. 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[42]) 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.[43] 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.[44]

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:

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

[41] Chain drift typically results from prices and quantities that exhibit large temporary fluctuations; see Szulc (1983) and Hill (1988). But the laptop price data seem to move quite smoothly so a priori, we did not think that chain drift would be a problem for this data set.
[42] See Diewert (1976) for the relationship of the Fisher index to the economic approach to index number theory.
[43] 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) (2018) and Hill, Rao, Shankar and Hajargasht (2017). Hill (2001) (2004) also pursued this similarity of relative prices approach in the time series context.
[44] This method is explained more fully in Diewert (2023).

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

Note that the prices of the matched models for month t are in the numerators of definitions (78) and (79) and the corresponding prices of the matched models for month r in the denominators of definitions (78) and (79). 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 $P_P(r,t)$:[45]

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

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:

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

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:[46]

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

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 (83) 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 $\lambda$, then the matched model Fisher index $P_F(r,t)$ defined by (81) will equal $\lambda$. Another very important property of the measure of relative price similarity defined by (83) 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 8.

**Table 8: 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) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|

[45] Note that there are 576 = 24x24 matched model bilateral Fisher (1922) indexes $P_F(r,t)$.
[46] See Diewert (2023) for the axiomatic properties of this measure.

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **1** | **0.0000** | **0.0103** | **0.0088** | **0.0170** | **0.0312** | **0.0492** | **0.0514** | **0.0506** | **0.0719** | **0.0643** | **0.0876** | **0.1009** |
| **2** | **0.0103** | **0.0000** | **0.0007** | **0.0092** | **0.0146** | **0.0257** | **0.0268** | **0.0325** | **0.0410** | **0.0448** | **0.0546** | **0.0554** |
| **3** | **0.0088** | **0.0007** | **0.0000** | **0.0046** | **0.0057** | **0.0119** | **0.0163** | **0.0168** | **0.0229** | **0.0236** | **0.0319** | **0.0340** |
| **4** | **0.0170** | **0.0092** | **0.0046** | **0.0000** | **0.0116** | **0.0149** | **0.0210** | **0.0196** | **0.0267** | **0.0268** | **0.0414** | **0.0459** |
| **5** | **0.0312** | **0.0146** | **0.0057** | **0.0116** | **0.0000** | **0.0005** | **0.0079** | **0.0030** | **0.0074** | **0.0071** | **0.0173** | **0.0215** |
| **6** | **0.0492** | **0.0257** | **0.0119** | **0.0149** | **0.0005** | **0.0000** | **0.0075** | **0.0027** | **0.0066** | **0.0059** | **0.0164** | **0.0207** |
| **7** | **0.0514** | **0.0268** | **0.0163** | **0.0210** | **0.0079** | **0.0075** | **0.0000** | **0.0045** | **0.0044** | **0.0057** | **0.0067** | **0.0075** |
| **8** | **0.0506** | **0.0325** | **0.0168** | **0.0196** | **0.0030** | **0.0027** | **0.0045** | **0.0000** | **0.0002** | **0.0013** | **0.0007** | **0.0012** |
| **9** | **0.0719** | **0.0410** | **0.0229** | **0.0267** | **0.0074** | **0.0066** | **0.0044** | **0.0002** | **0.0000** | **0.0009** | **0.0002** | **0.0005** |
| **10** | **0.0643** | **0.0448** | **0.0236** | **0.0268** | **0.0071** | **0.0059** | **0.0057** | **0.0013** | **0.0009** | **0.0000** | **0.0007** | **0.0039** |
| **11** | **0.0876** | **0.0546** | **0.0319** | **0.0414** | **0.0173** | **0.0164** | **0.0067** | **0.0007** | **0.0002** | **0.0007** | **0.0000** | **0.0002** |
| **12** | **0.1009** | **0.0554** | **0.0340** | **0.0459** | **0.0215** | **0.0207** | **0.0075** | **0.0012** | **0.0005** | **0.0039** | **0.0002** | **0.0000** |
| **13** | **0.1396** | **0.0832** | **0.0497** | **0.0500** | **0.0285** | **0.0276** | **0.0240** | **0.0160** | **0.0144** | **0.0174** | **0.0133** | **0.0132** |
| **14** | **0.1412** | **0.0935** | **0.0568** | **0.0545** | **0.0347** | **0.0335** | **0.0320** | **0.0220** | **0.0240** | **0.0230** | **0.0185** | **0.0181** |
| **15** | **0.1487** | **0.1013** | **0.0620** | **0.0566** | **0.0405** | **0.0397** | **0.0368** | **0.0266** | **0.0295** | **0.0289** | **0.0239** | **0.0237** |
| **16** | **0.1784** | **0.1158** | **0.0799** | **0.0767** | **0.0511** | **0.0483** | **0.0457** | **0.0345** | **0.0374** | **0.0367** | **0.0320** | **0.0342** |
| **17** | **0.2995** | **0.2356** | **0.1480** | **0.1292** | **0.0929** | **0.0865** | **0.0926** | **0.0758** | **0.0763** | **0.0775** | **0.0744** | **0.0860** |
| **18** | **0.3798** | **0.2993** | **0.1719** | **0.1442** | **0.0852** | **0.0768** | **0.0829** | **0.0667** | **0.0687** | **0.0665** | **0.0682** | **0.0821** |
| **19** | **0.3937** | **0.3428** | **0.2843** | **0.2545** | **0.1547** | **0.1549** | **0.1583** | **0.1381** | **0.1392** | **0.1409** | **0.1344** | **0.1429** |
| **20** | **0.6077** | **0.5073** | **0.3255** | **0.2534** | **0.1732** | **0.1664** | **0.1724** | **0.1525** | **0.1532** | **0.1543** | **0.1571** | **0.1850** |
| **21** | **0.5892** | **0.5008** | **0.2837** | **0.2233** | **0.1554** | **0.1473** | **0.1849** | **0.1659** | **0.1657** | **0.1677** | **0.1711** | **0.1964** |
| **22** | **0.8498** | **0.6705** | **0.4450** | **0.3799** | **0.2317** | **0.2216** | **0.2461** | **0.2465** | **0.2442** | **0.2463** | **0.2457** | **0.2896** |
| **23** | **0.8646** | **0.6571** | **0.4914** | **0.4568** | **0.3629** | **0.3730** | **0.4268** | **0.4061** | **0.4061** | **0.4102** | **0.4165** | **0.4628** |
| **24** | **1.0132** | **0.8555** | **0.6126** | **0.4593** | **0.3182** | **0.3071** | **0.3539** | **0.2608** | **0.2626** | **0.2612** | **0.2816** | **0.3249** |

| **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** | **0.1396** | **0.1412** | **0.1487** | **0.1784** | **0.2995** | **0.3798** | **0.3937** | **0.6077** | **0.5892** | **0.8498** | **0.8646** | **1.0132** |
| **2** | **0.0832** | **0.0935** | **0.1013** | **0.1158** | **0.2356** | **0.2993** | **0.3428** | **0.5073** | **0.5008** | **0.6705** | **0.6571** | **0.8555** |
| **3** | **0.0497** | **0.0568** | **0.0620** | **0.0799** | **0.1480** | **0.1719** | **0.2843** | **0.3255** | **0.2837** | **0.4450** | **0.4914** | **0.6126** |
| **4** | **0.0500** | **0.0545** | **0.0566** | **0.0767** | **0.1292** | **0.1442** | **0.2545** | **0.2534** | **0.2233** | **0.3799** | **0.4568** | **0.4593** |
| **5** | **0.0285** | **0.0347** | **0.0405** | **0.0511** | **0.0929** | **0.0852** | **0.1547** | **0.1732** | **0.1554** | **0.2317** | **0.3629** | **0.3182** |
| **6** | **0.0276** | **0.0335** | **0.0397** | **0.0483** | **0.0865** | **0.0768** | **0.1549** | **0.1664** | **0.1473** | **0.2216** | **0.3730** | **0.3071** |
| **7** | **0.0240** | **0.0320** | **0.0368** | **0.0457** | **0.0926** | **0.0829** | **0.1583** | **0.1724** | **0.1849** | **0.2461** | **0.4268** | **0.3539** |
| **8** | **0.0160** | **0.0220** | **0.0266** | **0.0345** | **0.0758** | **0.0667** | **0.1381** | **0.1525** | **0.1659** | **0.2465** | **0.4061** | **0.2608** |
| **9** | **0.0144** | **0.0240** | **0.0295** | **0.0374** | **0.0763** | **0.0687** | **0.1392** | **0.1532** | **0.1657** | **0.2442** | **0.4061** | **0.2626** |
| **10** | **0.0174** | **0.0230** | **0.0289** | **0.0367** | **0.0775** | **0.0665** | **0.1409** | **0.1543** | **0.1677** | **0.2463** | **0.4102** | **0.2612** |
| **11** | **0.0133** | **0.0185** | **0.0239** | **0.0320** | **0.0744** | **0.0682** | **0.1344** | **0.1571** | **0.1711** | **0.2457** | **0.4165** | **0.2816** |
| **12** | **0.0132** | **0.0181** | **0.0237** | **0.0342** | **0.0860** | **0.0821** | **0.1429** | **0.1850** | **0.1964** | **0.2896** | **0.4628** | **0.3249** |
| **13** | **0.0000** | **0.0035** | **0.0032** | **0.0057** | **0.0184** | **0.0230** | **0.0355** | **0.0380** | **0.0443** | **0.0842** | **0.1022** | **0.0937** |
| **14** | **0.0035** | **0.0000** | **0.0006** | **0.0031** | **0.0111** | **0.0170** | **0.0248** | **0.0254** | **0.0299** | **0.0656** | **0.0767** | **0.0762** |
| **15** | **0.0032** | **0.0006** | **0.0000** | **0.0003** | **0.0039** | **0.0072** | **0.0112** | **0.0101** | **0.0148** | **0.0486** | **0.0550** | **0.0567** |
| **16** | **0.0057** | **0.0031** | **0.0003** | **0.0000** | **0.0014** | **0.0035** | **0.0044** | **0.0045** | **0.0064** | **0.0407** | **0.0434** | **0.0458** |
| **17** | **0.0184** | **0.0111** | **0.0039** | **0.0014** | **0.0000** | **0.0020** | **0.0025** | **0.0025** | **0.0036** | **0.0391** | **0.0412** | **0.0438** |
| **18** | **0.0230** | **0.0170** | **0.0072** | **0.0035** | **0.0020** | **0.0000** | **0.0012** | **0.0031** | **0.0019** | **0.0359** | **0.0358** | **0.0396** |
| **19** | **0.0355** | **0.0248** | **0.0112** | **0.0044** | **0.0025** | **0.0012** | **0.0000** | **0.0006** | **0.0010** | **0.0349** | **0.0332** | **0.0367** |
| **20** | **0.0380** | **0.0254** | **0.0101** | **0.0045** | **0.0025** | **0.0031** | **0.0006** | **0.0000** | **0.0006** | **0.0341** | **0.0336** | **0.0370** |
| **21** | **0.0443** | **0.0299** | **0.0148** | **0.0064** | **0.0036** | **0.0019** | **0.0010** | **0.0006** | **0.0000** | **0.0330** | **0.0313** | **0.0356** |
| **22** | **0.0842** | **0.0656** | **0.0486** | **0.0407** | **0.0391** | **0.0359** | **0.0349** | **0.0341** | **0.0330** | **0.0000** | **0.0009** | **0.0043** |
| **23** | **0.1022** | **0.0767** | **0.0550** | **0.0434** | **0.0412** | **0.0358** | **0.0332** | **0.0336** | **0.0313** | **0.0009** | **0.0000** | **0.0013** |
| **24** | **0.0937** | **0.0762** | **0.0567** | **0.0458** | **0.0438** | **0.0396** | **0.0367** | **0.0370** | **0.0356** | **0.0043** | **0.0013** | **0.0000** |

Table 8 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 $P_F(p^1,p^2)$ defined by (83) to define the similarity linked month 2 index. Thus $P_S^2 \equiv P_F(p^1,p^2)$. Now look at the column in Table 8 that has the heading Δ(r,3). Look at the first 2 entries in this column. We have Δ(1,3) = 0.0088 and Δ(2,3) = 0.0007. Since Δ(2,3) is smaller than Δ(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 8 that has the heading $\Delta(r,4)$. 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 $\Delta(r,24)$. 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 $P_S^{24} \equiv P_S^{23}\, P_F(23,24)$.[47]

The relative price Predicted Share Similarity Linked indexes $P_S^t$ are listed in Table 9 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 9. Finally, for comparison purposes, Table 9 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 5 for plots of the indexes listed in Table 9.

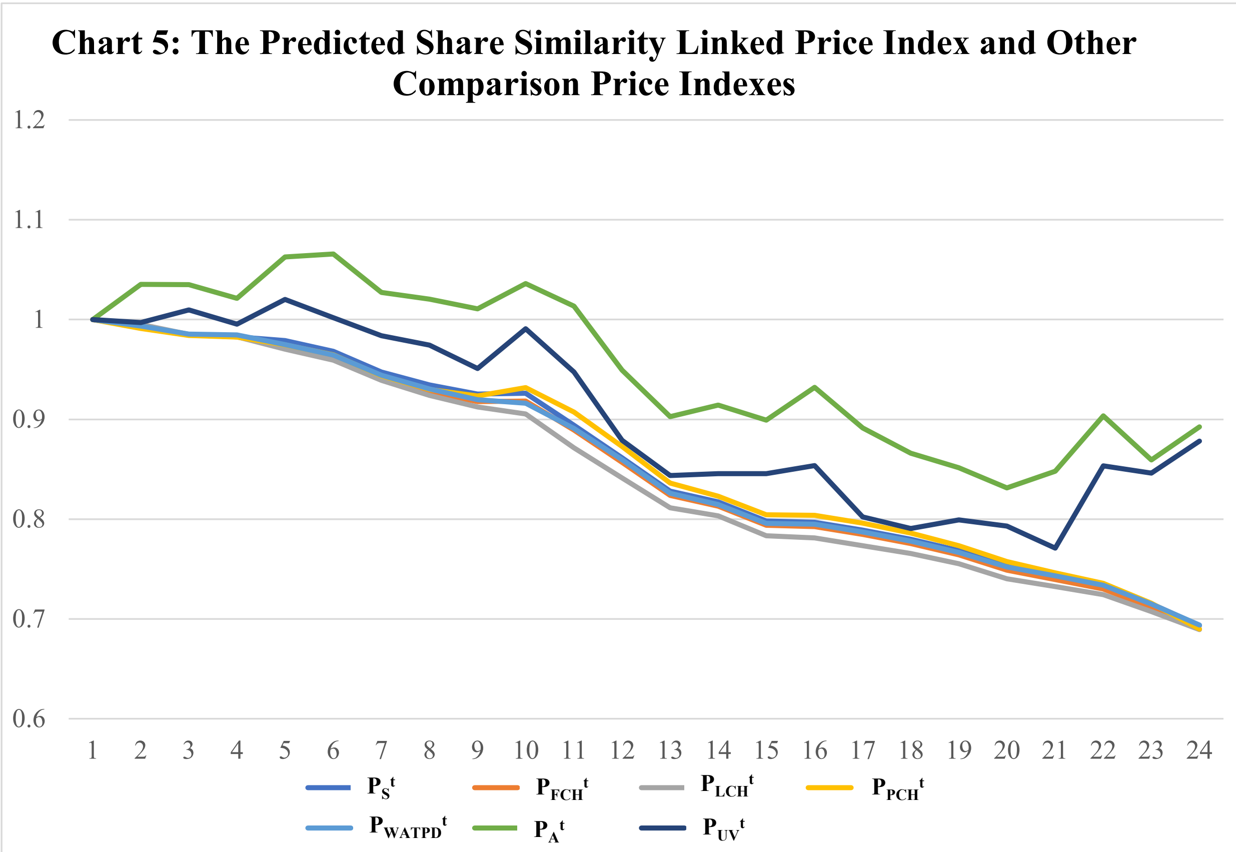
**Table 9: The Predicted Share Similarity Linked Price Index and Other Comparison Price Indexes**

| Month t | $P_S^t$ | $P_{FCH}^t$ | $P_{LCH}^t$ | $P_{PCH}^t$ | $P_{WATPD}^t$ | $P_A^t$ | $P_{UV}^t$ |
|---|---|---|---|---|---|---|---|
| 1 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 | 1.00000 |
| 2 | 0.99299 | 0.99299 | 0.99499 | 0.99099 | 0.99358 | 1.03525 | 0.99703 |
| 3 | 0.98452 | 0.98452 | 0.98509 | 0.98395 | 0.98526 | 1.03503 | 1.00972 |
| 4 | 0.98264 | 0.98264 | 0.98278 | 0.98250 | 0.98456 | 1.02127 | 0.99538 |
| 5 | 0.97885 | 0.97249 | 0.97035 | 0.97463 | 0.97476 | 1.06279 | 1.02001 |
| 6 | 0.96824 | 0.96195 | 0.95918 | 0.96472 | 0.96444 | 1.06571 | 1.00173 |
| 7 | 0.94753 | 0.94137 | 0.93918 | 0.94357 | 0.94422 | 1.02721 | 0.98386 |
| 8 | 0.93457 | 0.92689 | 0.92393 | 0.92986 | 0.93034 | 1.02049 | 0.97422 |
| 9 | 0.92543 | 0.91782 | 0.91232 | 0.92335 | 0.91971 | 1.01082 | 0.95086 |
| 10 | 0.92600 | 0.91838 | 0.90527 | 0.93168 | 0.91611 | 1.03594 | 0.99085 |
| 11 | 0.89409 | 0.88924 | 0.87157 | 0.90727 | 0.89088 | 1.01327 | 0.94737 |
| 12 | 0.86152 | 0.85685 | 0.84120 | 0.87279 | 0.85948 | 0.94941 | 0.87888 |
| 13 | 0.82820 | 0.82371 | 0.81147 | 0.83614 | 0.82589 | 0.90281 | 0.84358 |
| 14 | 0.81744 | 0.81301 | 0.80318 | 0.82295 | 0.81473 | 0.91423 | 0.84563 |
| 15 | 0.79826 | 0.79394 | 0.78350 | 0.80451 | 0.79577 | 0.89907 | 0.84560 |
| 16 | 0.79677 | 0.79245 | 0.78126 | 0.80379 | 0.79492 | 0.93198 | 0.85366 |
| 17 | 0.78900 | 0.78472 | 0.77346 | 0.79615 | 0.78726 | 0.89127 | 0.80235 |
| 18 | 0.77988 | 0.77565 | 0.76547 | 0.78596 | 0.77805 | 0.86620 | 0.79067 |
| 19 | 0.76847 | 0.76431 | 0.75526 | 0.77346 | 0.76665 | 0.85147 | 0.79919 |
| 20 | 0.75289 | 0.74881 | 0.74032 | 0.75740 | 0.75214 | 0.83124 | 0.79319 |
| 21 | 0.74342 | 0.73939 | 0.73261 | 0.74623 | 0.74318 | 0.84793 | 0.77090 |
| 22 | 0.73398 | 0.73000 | 0.72431 | 0.73573 | 0.73369 | 0.90356 | 0.85345 |
| 23 | 0.71536 | 0.71148 | 0.70730 | 0.71569 | 0.71498 | 0.85940 | 0.84609 |
| 24 | 0.69347 | 0.68971 | 0.68948 | 0.68993 | 0.69385 | 0.89247 | 0.87814 |
| Mean | 0.85890 | 0.85468 | 0.84806 | 0.86139 | 0.85685 | 0.95287 | 0.90302 |

It can be seen 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. These three indexes seem to be "best" for our particular

[47] 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

application. It can also be seen that the chained Laspeyres and Paasche indexes, $P_{LCH}^t$ and $P_{PCH}^t$, are very close to our "best" 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 they are strongly seasonal products, chained Fisher indexes may be subject to some chain drift and so the use of the similarity linked indexes is recommended in this case. The disadvantages of the similarity linked indexes are that the programming required to produce these indexes is more complex and the indexes will be difficult to explain to the public.

The Adjacent Period Weighted Time Product Dummy indexes performed well in this application. But in other applications where the products are not close substitutes, this method can be biased because it basically assumes linear preferences for purchasers of the group of products in scope.[48] Also if there is price bouncing behavior, this method will be subject to possible chain drift.

## 8. Conclusion.

The following tentative conclusions emerge from our study of laptop prices in Japan:

[48] See Diewert and Fox (2022) on this point.

- If quantity or expenditure weights are available in addition to price information, then it is important to use these weights in the calculation of a weighted by economic importance price index.
- Hedonic regressions that use amounts of product characteristics as independent variables in the regressions are not recommended for two reasons: (i) it is expensive to collect information on characteristics and (ii) it is likely that some important price determining characteristics are not included in the list of characteristics.[49]
- The Adjacent Period Weighted Time Product Dummy index is a preferred index provided that: (i) prices and quantities do not fluctuate violently from period to period due to product sales or strong seasonality and (ii) the products in scope are thought to be close substitutes.
- The Predicted Share Similarity Linked index is also a preferred index that should be satisfactory even if there are product sales or strong seasonality or if the products in scope are not close substitutes. The disadvantages of this method are the complexity of the computations and the difficulty of explaining the method to the public.
- In our particular application, our two preferred indexes were virtually identical. The chained maximum overlap Fisher indexes were also extremely close to our two preferred indexes and the chained maximum overlap Laspeyres and Paasche indexes were very close to our preferred indexes. However, we do not expect these close approximations to occur in other applications.

**Data Appendix**

The data can be obtained on request by emailing one of the authors.

**References**


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

Aizcorbe, A., C. Corrado and M. Doms (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 DC.

Allen, R.C. and W.E. Diewert (1981), "Direct versus Implicit Superlative Index Number Formulae", *Review of Economics and Statistics* 63, 430-435

.

Aten, B. and A. Heston (2009), "Chaining Methods for International Real Product and Purchasing Power Comparisons: Issues and Alternatives", pp. 245-273 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.


[49] There is at least one important exception to this "rule". Property price indexes must use property characteristics (such as the land plot area, the floor space area of the structure on the property and the location of the property) in order to construct a property price index. The reason why the time product dummy approach and normal index number theory cannot be used in this context is the fact that a property with a structure *does not remain the same* from period to period due to structure ageing and renovation expenditures. For examples of the use of hedonic characteristics regressions in property markets, see Diewert and Shimizu (2015) (2016) (2022).

Court, A.T. (1939), “Hedonic Price Indexes with Automotive Examples”, pp. 99-117 in The *Dynamics of Automobile Demand*, New York: General Motors Corporation.

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 Re-pricing Methods”, Jahrbücher für Nationökonomie und Statistik 230, 772-791.

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

de Haan, J. and F. Krsinich (2018), “Time Dummy Hedonic and Quality-Adjusted Unit Value Indexes: Do They Really Differ?”, *Review of Income and Wealth* 64:4, 757-776.

Diewert, W.E. (1976), “Exact and Superlative Index Numbers”, *Journal of Econometrics* 4, 114-145.

Diewert, W.E. (2002), “Weighted Country Product Dummy Variable Regressions and Index Number Formulae”, Department of Economics, Discussion Paper 02-15, University of British Columbia, Vancouver, B.C., Canada, V6T 1Z1.

Diewert, W.E. (2003), “Hedonic Regressions: A Consumer Theory Approach”, in *Scanner Data and Price Indexes*, Studies in Income and Wealth (Vol. 61), eds. R.C. Feenstra and M.D. Shapiro, Chicago: University of Chicago Press, pp. 317-348.

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. (2005a), “Weighted Country Product Dummy Variable Regressions and Index Number Formulae”, *Review of Income and Wealth* 51, 561-570.

Diewert, W.E. (2005b), “Adjacent Period Dummy Variable Hedonic Regressions and Bilateral Index Number Theory”, *Annales D’Économie et de Statistique*, No. 79/80, 759-786.

Diewert, W.E. (2009), “Similarity Indexes and Criteria for Spatial Linking”, pp. 183-216 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham, UK: Edward Elgar.

Diewert, W.E. (2013), “Methods of Aggregation above the Basic Heading Level within Regions”, pp. 121-167 in *Measuring the Real Size of the World Economy: The Framework, Methodology and Results of the International Comparison Program—ICP*, Washington D.C.: The World Bank.

Diewert, W.E. (2022), "Quality Adjustment Methods", Chapter 8 in *Consumer Price Index Theory*, Washington D.C.: International Monetary Fund, published online at: https://www.imf.org/en/Data/Statistics/cpi-manual.

Diewert, W.E. (2023), "Scanner Data, Elementary Price Indexes and the Chain Drift Problem", pp. 445-606 in *Advances in Economic Measurement*, D. Chotikapanich, A.N. Rambaldi and N. Rhode (eds.), Singapore: Palgrave Macmillan.

Diewert, W.E. and K.J. Fox (2022), "Substitution Bias in Multilateral Methods for CPI Construction," *Journal of Business and Economic Statistics* 40:1, 355-369.

Diewert, W.E. and C. Shimizu (2015), "Residential Property Price Indices for Tokyo", *Macroeconomic Dynamics* 19, 1659-1714.

Diewert, W. E. and C. Shimizu (2016), "Hedonic Regression Models for Tokyo Condominium Sales," *Regional Science and Urban Economics* 60, 300-315.

Diewert, W.E. and C. Shimizu (2022), Residential Property Price Indexes: Spatial Coordinates versus Neighbourhood Dummy Variables", *Review of Income and Wealth* 68:3, 770-796.

Fisher, I. (1922), *The Making of Index Numbers*, Boston: Houghton-Mifflin.

Griliches, Z. (1971), "Introduction: Hedonic Price Indexes Revisited", pp. 3-15 in *Price Indexes and Quality Change*, Z. Griliches (ed.), Cambridge MA: Harvard University Press.

Hardy, G.H., J.E. Littlewood and G. Pólya (1934), *Inequalities*, Cambridge: Cambridge University Press.

Hill, R.J. (1997), "A Taxonomy of Multilateral Methods for Making International Comparisons of Prices and Quantities", *Review of Income and Wealth* 43(1), 49-69.

Hill, R.J. (1999a), "Comparing Price Levels across Countries Using Minimum Spanning Trees", *The Review of Economics and Statistics* 81, 135-142.

Hill, R.J. (1999b), "International Comparisons using Spanning Trees", pp. 109-120 in *International and Interarea Comparisons of Income, Output and Prices*, A. Heston and R.E. Lipsey (eds.), Studies in Income and Wealth Volume 61, NBER, Chicago: The University of Chicago Press.

Hill, R.J. (2001), "Measuring Inflation and Growth Using Spanning Trees", *International Economic Review* 42, 167-185.

Hill, R.J. (2004), "Constructing Price Indexes Across Space and Time: The Case of the European Union", *American Economic Review* 94, 1379-1410.

Hill, R.J. (2009), "Comparing Per Capita Income Levels Across Countries Using Spanning Trees: Robustness, Prior Restrictions, Hybrids and Hierarchies", pp. 217-244 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Hill, R.J., D.S. Prasada Rao, S. Shankar and R. Hajargasht (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. and M.P. Timmer (2006), "Standard Errors as Weights in Multilateral Price Indexes", *Journal of Business and Economic Statistics* 24:3, 366-377.

Hill, T.P. (1988), "Recent Developments in Index Number Theory and Practice", *OECD Economic Studies* 10, 123-148.

Kravis, I.B., A. Heston and R. Summers (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.

Muellbauer, J. (1974), "Household Production Theory, Quality and the .Hedonic Technique", *American Economic Review* 64:6, 977-994.

Rao, D.S. Prasada (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. Prasada (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. Prasada (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* 51:4, 571-580.

Rao, D.S. Prasada and G. Hajargasht (2016), "Stochastic Approach to Computation of Purchasing Power Parities in the International Comparison Program", *Journal of Econometrics* 191:2, 414-425.

Rao, D.S. Prasada and M.P Timmer (2003), "Purchasing Power Parities for Industry Comparisons Using Weighted Eltetö-Köves-Szulc (EKS) Methods", *Review of Income and Wealth* 49, 491-511.

Rosen, S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy* 82, 34–55.

Sergeev, S. (2001), "Measures of the Similarity of the Country's Price Structures and their Practical Application", Conference on the European Comparison Program, U. N. Statistical Commission. Economic Commission for Europe, Geneva, November 12-14, 2001.

Sergeev, S. (2009), "Aggregation Methods Based on Structural International Prices", pp. 274-297 in *Purchasing Power Parities of Currencies: Recent Advances in Methods and Applications*, D.S. Prasada Rao (ed.), Cheltenham UK: Edward Elgar.

Summers, R. (1973), "International Comparisons with Incomplete Data", *Review of Income and Wealth* 29:1, 1-16.

Szulc, B.J. (1983), "Linking Price Index Numbers," pp. 537-566 *in Price Level Measurement*, W.E. Diewert and C. Montmarquette (eds.), Ottawa: Statistics Canada.

Triplett, J. (2004), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, Directorate for Science, Technology and Industry, DSTI/DOC(2004)9, Paris: OECD.

Triplett, J. E. and R. J. McDonald (1977), "Assessing the Quality Error in Output Measures: The Case of Refrigerators", *The Review of Income and Wealth* 23:2, 137-156.

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