Economic Measurement Group Workshop 2023

Scanner Data, Product Churn and Quality Adjustment

October 25, 2023

Erwin Diewert
(University of British Columbia)
and
Chihiro Shimizu
(Hitotsubashi University)

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 Section 4 are Time Dummy Characteristics type regressions.**
- **This section draws on the theory explained in Section 3 and runs panel data regressions. (Weighted by economic importance and unweighted).**

- **Section 5 is similar to section 4 except we run a Time Dummy Characteristics regression using the data of two consecutive periods and then the results of these separate regressions are chained together to generate the final index, which is called an Adjacent Period Time Dummy Characteristics index.**
- **Thus the indexes that are estimated in this section are real time "practical" indexes that statistical offices could produce.**

- **Section 6 draws on the theory explained in section 2; i.e., we consider weighted and unweighted Time Product Dummy hedonic regressions in this section.**
- **The models in this section use only a single product characteristic: the Japanese product code for each laptop sale.**
- **We consider a single panel regression versus a sequence of bilateral regressions that utilize the price and quantity data for two consecutive periods.**
- **The latter type of model can be implemented in real time and is called an Adjacent Period Time Product Dummy hedonic regression model.**

- **Section 7 considers alternatives to hedonic regression models based on standard index number theory; i.e., maximum overlap chained Laspeyres, Paasche and Fisher indexes are computed in this section.**
- **We also compute the Predicted Share Similarity linked price indexes which have only been developed recently.**
- **The indexes calculated in this section are also "practical" indexes.**

- **Section 8 introduces an additional characteristic into the hedonic Time Product Dummy regressions explained in section 6: the "newness" of the laptop model; i.e., the number of months that the model has been available in the marketplace.**
- **It could be the case that laptops are a "fashion" product where purchasers value a product just because it is new.**

- **Section 9 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 a product dummy variable; i.e., each product has its own unique bundle of price determining characteristics which can be represented by a unique product dummy variable.**
- **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.**
-

Section 2: Hedonic Regression Theory

- **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 $t = 1,...,T$.**
- **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 α_n are positive parameters, which can be interpreted as *quality adjustment factors.***
- **α_n is the marginal utility that an extra unit of product n gives to a purchaser of product n.**
- **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, it can be shown that the aggregate period t vector of purchases q^t is a solution to the aggregate period t utility maximization problem, $\max_q \{\alpha \cdot q : p^t \cdot q = e^t ; q \geq 0_N\}$ where e^t is equal to aggregate period t expenditure on the N products.**

- **The first order conditions for an interior solution, q^t, λ_t to the period t aggregate utility maximization problem are the following N+1 equations, where λ_t is a Lagrange multiplier:**

(2) $\alpha = \lambda_t p^t$;

(3) $p^t \cdot q^t = e^t$.

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

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

- **Define π_t as follows:**

(5) $\pi_t \equiv 1/\lambda_t = e^t/\alpha \cdot q^t = p^t \cdot q^t/\alpha \cdot q^t$

= a period t quality adjusted unit value price level.

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

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

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

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

(7) $Q^t \equiv \alpha \cdot q^t$; (so Q^t is essentially aggregate period t utility)

(8) $P^t \equiv e^t/Q^t = \pi_t$ (and P^t is a quality adjusted period t unit value)

- **where the second equation in (8) follows using (4) and (5).**
- **Thus equations (6) have the following interpretation: the period t price of product n, p_{tn}, is equal to the period t price level π_t times a quality adjustment parameter for product n, α_n.**
- **Empirically, equations (6) are unlikely to hold exactly and so we add error terms e_{tn} to the right hand sides of equations (6).**
- **The following slide shows the resulting (nonlinear in the parameters) least squares regression problem:**

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

- **However, Diewert (2023) showed that the estimated price levels π_t^* that solve the minimization problem (9) had unsatisfactory axiomatic properties.**
- **Thus we follow Court (1939) 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*:**

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

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

(11) $\rho_t \equiv \ln \pi_t$; $\quad t = 1,\ldots,T$;

(12) $\beta_n \equiv \ln \alpha_n$; $\quad n = 1,\ldots,N$.

- **The model defined by (10) is an adaptation of Summer's (1973) Country Product Dummy model to the time series context. Aizcorbe, Corrado and Doms (2000) had an early application of this model in the time series context.**

- **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 following Rao (1995) (2004) (2005; 574), we consider the following *weighted least squares minimization problem*:**

(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 $\rho^* \equiv [\rho_1^*,\ldots,\rho_T^*]$ and $\beta^* \equiv [\beta_1^*,\ldots,\beta_N^*]$ to solve (13) simplify to the following T equations (14) and N equations (15):**

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

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

- **Viewing the minimization problem (13), it can be seen that the ρ_t and β_n are not completely identified: we can add a constant to each ρ_t and subtract the same constant from each β_n and the predicted value for $\ln p_{tn}$ remains unchanged.**

- **Thus we impose the normalization $\rho_1 = 0$ (which corresponds to $\pi_1 = 1$) in order to identify all of the unknown parameters in (13).**

- **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.**

- **Alternatively, we 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 $t = 1,\ldots,T$ and $n = 1,\ldots,N$ where we set $\rho_1 = 0$ to avoid exact multicollinearity. The parameters estimated by this weighted linear regression will solve the weighted least squares minimization problem defined by (13) (along with the normalization $\rho_1 = 0$).**

- **Once the solution to (13) has been obtained, define $\pi_1^* = 1$ and the other estimated *price levels* π_t^* and the *quality adjustment factors* α_n^* as follows:**

(16) $\pi_t^* \equiv \exp[\rho_t^*]$; $t = 2,3,\ldots,T$; $\qquad \alpha_n^* \equiv \exp[\beta_n^*]$; $n = 1,\ldots,N$.

- **The price levels π_t^* defined by (16) are called the *Weighted Time Product Dummy price levels*.**

- **Note that the resulting *price index* between periods t and τ is defined as the ratio of the period t price level to the period τ price level and is equal to the following expression:**

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

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

The Two Sets of Estimates Problem

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

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

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

- Option (18) will tend to give us smoother price levels, P^{t*}, while option (19) will tend to give us smoother quantity levels.
- **If the fit of the weighted least squares regression problem (13) is perfect, then the two options will coïncide as will the weighted and unweighted indexes.**
- Slides 7-13 developed the theory behind unweighted and weighted Time Product Dummy hedonic régressions over a window of T periods in the case where **there were no missing observations**.

Time Product Dummy Regressions with Missing Observations

- **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 $t = 1,2,\ldots,T$. 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 $S^*(n) \equiv \{t: p_{tn} > 0\}$ for $n = 1,\ldots,N$. We 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*:**

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

Note that there are two equivalent ways of writing the least squares minimization problem.

- **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.**

- **The first order necessary conditions for $\rho_1,\ldots,\rho_T$ and $\beta_1,\ldots,\beta_N$ to solve (20) are the following counterparts to (14) and (15):**

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

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

- **In order to obtain a unique solution to (21) and (22), we set $\rho_1 = 0$ and we drop the first equation in (21) and use linear algebra to find a unique solution for the resulting equations. In practice, we ran an appropriate weighted least squares regression.**
- **Define the estimated *price levels* π_t^{*} and *quality adjustment factors* α_n^{*} by definitions (11) and (12). Substitute these definitions into equations (21) and (22).**
- **After some rearrangement, equations (21) and (22) become the following equations:**

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

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

- **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.**
- **The resulting system of T – 1 + N equations needs to be of full rank in order to obtain a unique solution. As noted above, the solution can also be obtained by running a linear regression.**

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

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

$$(26)\ Q^{t*} \equiv \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/P^{t*} \qquad t = 1,\ldots,T.$$

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

- **The counterparts to definitions (19) are the following definitions:**

$$(27)\ Q^{t**} \equiv \Sigma_{n\in S(t)}\, \alpha_n^* q_{tn}\ ; \qquad t = 1,\ldots,T;\ \text{Method 2}$$

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

- **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 $t = 1,\ldots,T$, $n \in S(t)$ 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.**
- **One perhaps unsatisfactory property of the WTPD price levels π_t^{*} is the following one: a product that is available in only one period out of the T periods has no influence on the aggregate price levels π_t^{*}. This means that the price of a new product that appears in period T has no influence on the price levels.**
- **The Time Dummy Characteristics 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.**

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

- **There are N products that are available over a window of T periods.**
- **The utility function for the N products is the linear function, $f(q) \equiv \alpha \cdot q = \Sigma_{n=1}^{N} \alpha_n q_n$ where q_n is the quantity of product n purchased or sold in the period under consideration and α_n is the quality adjustment factor for product n.**
- **What is new is the assumption that the quality adjustment factors are functions of a vector of K *characteristics* of the products: product n has the vector of characteristics $z^n \equiv [z_{n1},z_{n2},\ldots,z_{nK}]$ for $n = 1,\ldots,N$.**
- **The new assumption in this section is that the quality adjustment factors α_n are functions of the vector of characteristics z^n for each product and the same function, g(z) can be used to quality adjust each product:**

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

- **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. Of course, the fact that information on product characteristics must be collected is a disadvantage of the class of models studied in this section.**

- **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* $g(z) = g(z_1,...,z_K)$.**
- **Possible choices for $g(z_1,...,z_K)$ are:**

(i) Let $g(z_1,...,z_K)$ be a general nonparametric function of K variables;

(ii) $g(z_1,...,z_K) \equiv a_1 z_1 + a_2 z_2 + \ldots + a_K z_{1K}$ (linear function of the characteristics);

(iii) $g(z_1,...,z_K) \equiv g_1(z_1) + g_2(z_2) + \ldots + g_K(z_K)$ (additive separability) with the $g_k(z_k)$ being general nonparametric functions of one variable;

(iv) $g(z_1,...,z_K) \equiv g_1(z_1) g_2(z_2) x \ldots x g_K(z_K)$ (multiplicative separability).

- **In our paper, we assume that g(z) is defined as follows:**

(32) $g(z_1,...,z_K) \equiv g_1(z_1)g_2(z_2)\ldots g_K(z_K)$. **(We assume multiplicative separability).**

- **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. (Use linear splines instead of step functions?)**
- **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:**

(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 (31) and we obtain the following system of possible estimating equations where the π_t and a_{1j}, a_{2j}, …, a_{Kj} are unknown parameters:**

(35) $p_{tn} = \pi_t(\Sigma_{j=1}^{N(1)}\, a_{1j}D_{tn,1,j})(\Sigma_{j=1}^{N(2)}\, a_{2j}D_{tn,2,j}) \ldots (\Sigma_{j=1}^{N(K)}\, a_{Kj}D_{tn,K,j})\ ;$

$t = 1,\ldots,T;\ n \in S(t).$

- **We take logarithms of both sides of equations (35) in order to obtain the following system of estimating equations:**

(36) $\ln p_{tn} = \ln\pi_t + \Sigma_{j=1}^{N(1)}\, (\ln a_{1j})D_{tn,1,j} + \Sigma_{j=1}^{N(2)}\, (\ln a_{2j})D_{tn,2,j} + \ldots$

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

- **Note that all of the independent variables in the above linear regression are dummy variables. 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) $\ln p_{tn} = \rho_t + \Sigma_{j=1}^{N(1)} b_{1j}D_{tn,1,j} + \Sigma_{j=1}^{N(2)} b_{2j}D_{tn,2,j} + \ldots + \Sigma_{j=1}^{N(K)} b_{Kj}D_{tn,K,j} + e_{tn}$; $t = 1,\ldots,T;\ n \in S(t).$

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

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

- **where ρ is the vector $[\rho_1, \rho_2, \ldots, \rho_T]$ and $b(k)$ is the vector $[b_{k1}, b_{k2}, \ldots, b_{kN(k)}]$ for $k = 1,2\ldots,K$.**
- **Solutions to the least squares minimization problem will exist but a solution will not be unique. 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 $k = 2,\ldots,K$ (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 $k = 1,2\ldots,K$ will be available.**
- **Use these estimates to define the logarithms of the quality adjustment factors α_n for all products n that were purchased in period t:**

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

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

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

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

$$(42)\ Q^{t**} \equiv \Sigma_{n \in S(t)}\, \alpha_{tn}^* q_{tn}\ ; \qquad t = 1,\ldots,T.$$

- **If product n is available in multiple periods, the quality adjustment factors remain the same across periods.**

- **In order to obtain a useful expression for the direct estimate for the period t price level, π_t, look at the first order conditions for minimizing (39) with respect to ρ_t:**

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

$$= \Sigma_{n\in S(t)} \{\ln p_{tn} - \rho_t^* - \beta_n^*\}$$

- **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 π_t^* is equal to:**

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

- **Thus the direct estimate for the period t price level P^{t*} is equal to the geometric mean of the period t quality adjusted prices (p_{tn}/α_{tn}^*) for the products that were purchased in period t.**
- **Note that this price level can be calculated using price information alone whereas the indirect measure P^{t**} requires price and quantity information on the purchase of products during period t.**
- **If the unweighted linear regression defined by (38) fits the data exactly, there is no need to run the weighted counterpart regression to be defined shortly.**

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

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

- **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, $k = 2,\ldots,K$.**
- **Use definitions (40), (41), (42) and (43) to define new β_{tn}^{*}, α_{tn}^{*}, Q^{t**} and P^{t**}.**

- **We rewrite P^{t**} in a somewhat different manner as follows:**

$$(48)\ P^{t**} = \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_{tn}^{*}q_{tn} \qquad t = 1,\ldots,T$$
$$= \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, (\alpha_{tn}^{*}/p_{tn})p_{tn}q_{tn}$$
$$= [\Sigma_{n\in S(t)}\, s_{tn}(p_{tn}/\alpha_{tn}^{*})^{-1}]^{-1}.$$

- **In order to obtain a useful expression for the direct estimate for the period t price level, π_t, look at the first order conditions for minimizing (46) with respect to ρ_t:**

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

- **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 π_t^{*} is equal to:**

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

- **Our normalizations imply $\pi_1^{*} = 1$.**

- **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.**

Section 4: Laptop Data for Japan and Sample Wide Hedonic Regressions Using Characteristics.

*We obtained data from a private firm that collects price, quantity and characteristic information on the monthly sales of laptop computers across Japan.

*The data are thought to cover more than 60% of all laptop sales in Japan. We utilized the data for the 24 months in the years 2021 and 2022 for our regressions and index computations.

*There were 2639 monthly price and quantity observations on laptops sold in total over all months.

*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.

*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 was tremendous product churn in the sales of laptops in Japan, with individual products quickly entering and then exiting the market for laptops.

We have information on 9 characteristics of each laptop:

- **NEW is the number of months that the product has been available (or more precisely, sold) in Japan. NEW = 1 means the product was a new one. NEW ranges from 1 to 38 months in our sample.**
- **CLOCK is the clock speed of the laptop. The mean clock speed was 1.94 and the range of clock speeds was 1 to 3.4. The larger is the clock speed, the faster the computer can make computations. There were 23 distinct clock speeds for the laptops in our sample.**
- **MEM is the memory capacity for the laptop. The mean memory size was 8188.9. There were only 3 memory sizes listed in our sample: 4,096, 8,192 and 16,384.**
- **SIZE is the screen size of the laptop. The mean screen size (in inches) was 14.49. There were 10 distinct screen sizes in our sample: 11.6, 12, 12.5, 13.3, 14, 15.4, 15.6,16, 16.1 and 17.3.**
- **PIX is the number of pixels imbedded in the screen of the laptop. The mean number of pixels was 24.82. There were only 10 distinct number of pixels in our sample: 10.49, 12.46, 12.96, 20.74, 33.18, 40.96, 51.84, 55.30, 58.98 and 82.94.**

- **HDMI** is the presence (HDMI = 1) or absence (HDMI = 0) of a HDMI terminal in the laptop. If HDMI =1, then it is possible to display digitally recorded images without degradation.
- **WEIGHT** is the weight of the laptop in kilograms. Laptop weights ranged from 0.747 to 2.9 kilos.
- A priori, we expected that purchasers would value higher clock speed, memory capacity, screen size, the number of pixels and the availability of HDMI in a laptop, leading to increasing estimated coefficients for the dummy variables corresponding to higher values of the characteristic under consideration. We expected that purchasers would value a lighter laptop over a heavier one.
- **CPU** is the type of Central Processing Unit that the laptop used. There were 12 types of CPU in our sample.
- **BRAND** is the name of the manufacturer of the laptop. We have 11 brands in our sample. BRAND is frequently used as an explanatory variable in a hedonic regression as a proxy for company wide product characteristics that may be missing from the list of explicit product characteristics that are included in the regression.

Table 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

- 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 is

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

- where N(t) is the number of laptops sold in period t and S(t) is the set of products sold in period t.
- The average period t price of a laptop, PA^t, weights each period t laptop price equally and thus does not take the economic importance of each type of laptop into account. A more representative measure of average laptop price in period t is the period t unit value price, PUV^t, defined as follows:

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

- where $e_{tn} \equiv p_{tn}q_{tn}$ is expenditure or sales of product n in period t for t = 1,…,24 and n = 1,…,366.
- We convert the average prices defined by (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) $P_A{}^t \equiv PA^t/PA^1$; $P_{UV}{}^t \equiv PUV^t/PUV^1$; $\quad t = 1,\ldots,24.$

- **We will list our average price index, $P_A{}^t$, and the period t unit value price index $P_{UV}{}^t$ shortly for the 24 months in our sample.**
- **But in this section, we will focus on time dummy regressions that use the characteristics information on the products. (Section 3 regressions with T = 24).**
- **We will run sample wide regressions (24 months) and introduce the 9 characteristics one by one and see if the fit of the regression improves significantly.**
- **Our first such regression is:**

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

- **We grouped the 25 Clock speeds into 7 groups and ran the following regression:**

(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, $\rho_2{}^*, \rho_3{}^*, \ldots, \rho_{24}{}^*, b_{C1}{}^*, \ldots, b_{C7}{}^*$ 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.**

4.2 A Hedonic Regression with Clock Speed as the Only Characteristic (cont)

- **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.**
- **The estimated ρ_t^* are the logarithms of the price levels P^{t*} for t = 2,3,…,24 but we will not list the estimated price levels until we have entered all 8 of our characteristics into the regression.**
- **Note that these coefficients will change as we add other characteristics to the regression.**

4.3 Adding Memory Capacity as an Additional Characteristic.

- **There were only 3 sizes of memory capacity (the variable MEM in the Data Appendix): 4096, 8192 and 16384.**
- **The new model is:**

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

- **We had 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.**

4.3 Adding Memory Capacity as an Additional Characteristic. (cont)

- 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 Adding Screen Size as an Additional Characteristic.

- There were 10 different screen sizes but some screen sizes had very few observations so we combined these small cells with a neighbouring cell to end up with 7 screen size dummy variables.
- 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.

4.4 Adding Screen Size as an Additional Characteristic (continued)

- **If increased screen size is valuable to purchasers, we would expect the estimated b_{Sj}^{*} coefficients to increase as j increases.**
- **For this regression, the estimates for b_{S2}^{*}-b_{S7}^{*} were 0.73371, 0.59447, 0.22923, 0.34524, 0.74190 and 0.68987. This regression indicates that purchasers prefer small and large screen sizes over intermediate screen sizes for laptops.**

4.5 Adding 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.**
- **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 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.**

4.5 Adding Pixels as an Additional Characteristic (cont)

- 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 Adding HDMI as an Additional Characteristic

- The dummy variable that indicates the presence of HDMI in the laptop is the column vector D_{H2} in the following hedonic regression:

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

- 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 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 Adding Brand as an Additional Characteristic.

- **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}.**
- **The brands were reordered according to their average prices with the lowest average price brands listed first and the highest average brand listed last.**
- **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).**
- **The log of the likelihood function for the hedonic regression defined by (65) was 754.295, a huge gain of 704.796 log likelihood points for adding 10 new brand parameters.**
- **The R^2 between the observed price vector and the predicted price vector was 0.8631 which is a very big increase over the R^2 of the previous model which was equal to 0.7764.**
- **The estimated brand coefficients b_{B2}^*- b_{B11}^* are: – 0.1014, 0.1366, 0.0975, 0.1201, 0.5048, 0.4136, 0.1469, 0.4743, 0.2880, 0.6401. Thus there is a weak general tendency for the marginal utility of a more expensive brand to be higher than the marginal utility of a cheaper brand.**

4.9 Adding Laptop Weight as an Additional Characteristic.

- We defined 7 weight dummy variables, D_{W1}-D_{W7} by choosing the following break points for laptop weights: 1.0, 1.3, 1.6, 1.9, 2.1 and 2.3.
- The D_{W1} cell consisted of laptops that weighed less than 1 kilo, the D_{W2} cell consisted of laptops that were in the interval $1 \leq \text{WEIGHT} < 1.3$, ,,, , the D_{W6} cell consisted of laptops that were in the interval $2.1 \leq \text{WEIGHT} < 2.3$ and the D_{W7} cell consisted of laptops that satisfied the inequality $\text{WEIGHT} \geq 2.3$.
- The number of laptops in each of these cells was as follows: 417, 408, 477, 311, 297, 466, 263. Add the column vectors D_{W2}-D_{W7} to the right hand side of the previous regression to get the new regression.
- The log of the likelihood function for the hedonic regression defined by (66) was 1074.86, an increase of 62.06 over the previous log likelihood for adding 6 additional parameters.
- The R^2 between the observed price vector and the predicted price vector was 0.8926 which is a substantial increase over the R^2 of the previous model which was equal to 0.8631.
- The estimated weight coefficients b_{W2}^*- b_{W7}^* are as follows: 0.0765, 0.0018, -0.2094, -0.2447, -0.1852 and -0.2378. Thus a lighter laptop has on average a slightly positive price premium but the price premium becomes negative (and approximately constant) for laptops that weigh more that 1.6 kilos.

4.9 Adding Laptop Weight as an Additional Characteristic. (cont)

- The estimated coefficients on the time dummy variables in this regression are ρ_2^*, ρ_3^*, …, ρ_{24}^*. Define $\rho_1^* \equiv 0$ and the estimated period t price levels $\pi_t^* \equiv \exp[\rho_t^*]$ for t = 1,2,…,24.
- Define the month t Time Dummy Characteristics Price Index, $P_{TDC}^t \equiv \pi_t^*$ for t = 1,…,24. This index is listed in Table 4 in the following subsection, which simply adds the share weights to the least squares minimization problem considered in this subsection.

4.10 The Weighted Time Dummy Characteristics Hedonic Regression Model.

- To obtain the weighted counterpart to the hedonic regression model defined by in section 4.9, form a share vector of dimension 2639 that corresponds to the $\ln p_{tn}$ that appear in (66) 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 (66) by SS to obtain a new linear regression model which again provides estimates for the unknown parameters that appear in (66).
- The R^2 for this new weighted regression model turned out to be 0.9152 which is substantially higher than the R^2 for the counterpart unweighted model which was 0.8926.

- **4.10 The Weighted Time Dummy Characteristics Hedonic Regression Model**
- **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 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.**

$1 = P_{WTDC}{}^{t}$; $2 = P_{TDC}{}^{t}$; $3 = P_{CMSPH}{}^{t}$; $4 = P_{CMSP}{}^{t}$; $5 = P_{CMS}{}^{t}$;

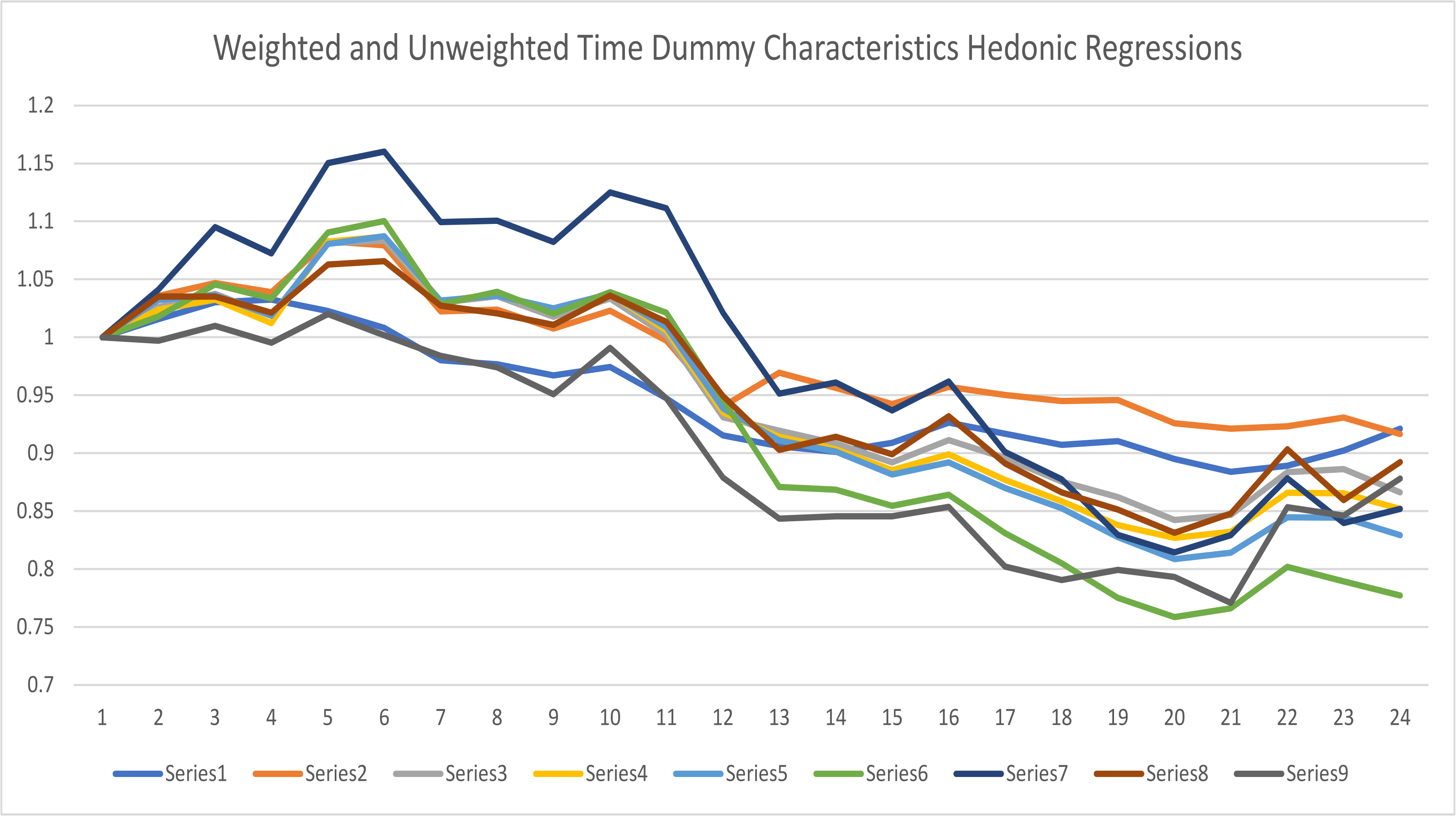
$6 = P_{CM}{}^{t}$; $7 = P_{C}{}^{t}$; $8 = P_{A}{}^{t}$; $9 = P_{UV}{}^{t}$.

Table 4: Weighted and Unweighted Time Dummy Characteristics Price Indexes

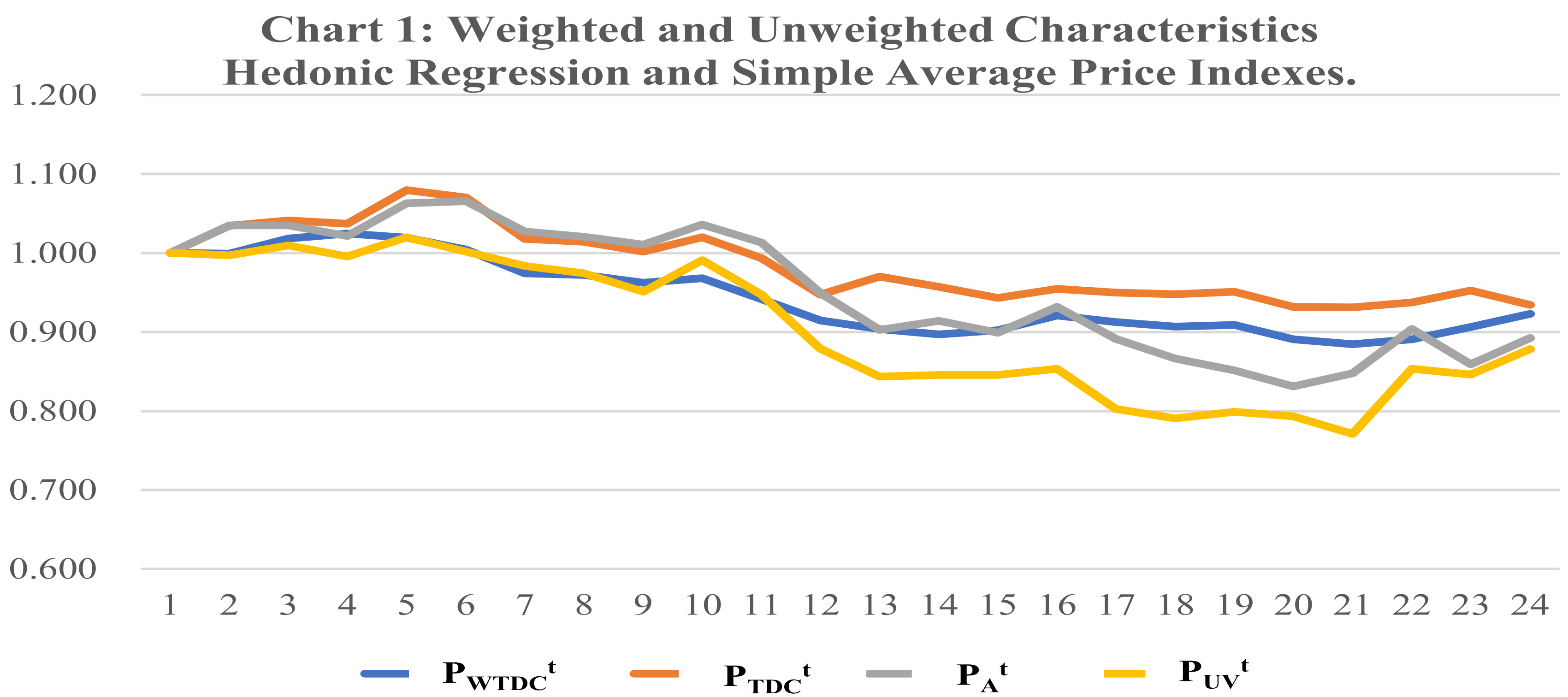
Month t	P_{WTDC}^{t}	P_{TDC}^{t}	P_{CMSPH}^{t}	P_{CMSP}^{t}	P_{CMS}^{t}	P_{CM}^{t}	P_{C}^{t}	P_{A}^{t}	P_{UV}^{t}
1	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000
2	1.01571	1.03561	1.02620	1.02367	1.03230	1.01802	1.04123	1.03525	0.99703
3	1.03031	1.04665	1.03749	1.03260	1.03625	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

$1 = P_{WTDC}{}^t$; $2 = P_{TDC}{}^t$; $3 = P_{CMSPH}{}^t$; $4 = P_{CMSP}{}^t$; $5 = P_{CMS}{}^t$;
$6 = P_{CM}{}^t$; $7 = P_C{}^t$; $8 = P_A{}^t$; $9 = P_{UV}{}^t$.

Adding characteristics changes the resulting index substantially!



A Simplified Chart 1
The Chart below shows the Weighted and Unweighted Time Dummy Characteristics Indexes along with the Average Laptop Price Index and the Unit Value Price Index.



- **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*: as we added characteristics to the regression, the resulting indexes changed significantly.**
- **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 $P_{WTDC}{}^{t}$ (0.94744) is substantially below the mean of the unweighted indexes $P_{TDC}{}^{t}$ (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.**

4.11 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 4.10. We will use the results around equations (42)-(52) in section 3.**
- **In section 4.10, we defined the estimated direct monthly price levels, π_t^*, by exponentiating the estimated coefficients ρ_t^*. Define the month t *direct price level* P^{t*} as follows:**

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

- **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:**

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

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

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

- **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* β^* as follows:**

$$(70)\ \beta^* \equiv b_0 ONE + \Sigma_{j=2}^{7}\, b_{Cj}\, D_{Cj} + \Sigma_{j=2}^{3}\, b_{Mj} D_{Mj} + \Sigma_{j=2}^{7}\, b_{Sj} D_{Sj} + \Sigma_{j=2}^{5}\, b_{Pj} D_{Pj} + b_{H2} D_{H2} + \Sigma_{j=2}^{11}\, b_{Bj} D_{Bj} + \Sigma_{j=2}^{10}\, b_{Uj} D_{Uj} + \Sigma_{j=2}^{7}\, b_{Wj} D_{Wj}.$$

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

$$(71)\ \alpha_{tn}^* \equiv \exp[\beta_{tn}^*]\ ; \qquad t = 1,\ldots,24;\ n \in S(t).$$

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

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

- **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**}:**

(73) $P^{t**} \equiv e^t/Q^{t**} = \Sigma_{n\in S(t)}\, p_{tn}q_{tn}/\Sigma_{n\in S(t)}\, \alpha_{tn}{}^{*}q_{tn}$; 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:**

(74) $P^{t**} \leq P^{t*}$ and $Q^{t**} \geq Q^{t*}$; t = 1,…,24.

- **If the R^2 for the weighted hedonic regression defined in section 4.10 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, $P_{IWTDC}{}^{t}$, by dividing the P^{t**} by P^{1**}:**

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

- **The series $P_{IWTDC}{}^{t}$ is also listed in Table 5.**

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

- **The point here is that the direct and indirect methods generate much the same indexes because the R square for the Weighted Time Dummy Characteristics regression was high.**

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

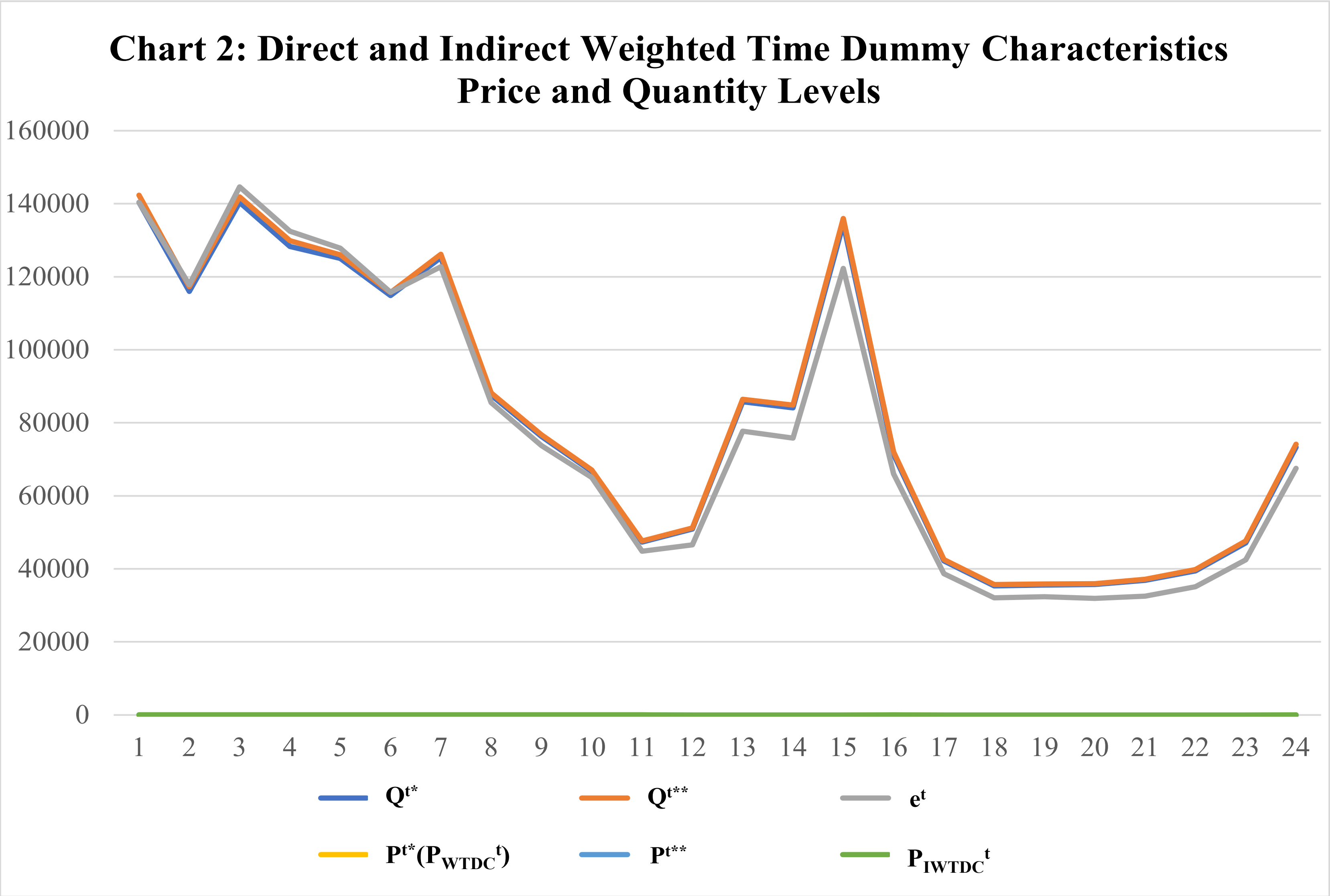


Chart 2: Direct and Indirect Weighted Time Dummy Characteristics
Price and Quantity Levels
160000
140000
120000
100000
80000
60000
40000
20000
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Q^{t*}
Q^{t**}
e^{t}
$P^{t*}(P_{WTDC}^{t})$
P^{t**}
P_{IWTDC}^{t}

5. Adjacent Period Time Dummy Characteristics Hedonic Regression Models.

There are two problems with our "best" directly defined weighted hedonic price index using characteristics, $P_{WTDC}{}^{t}$, 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;**
- **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. Thus we can use the analytical framework presented in section 3 and simply apply it to the case where T = 2.**
- **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.**

- **As in the previous section, weighted and unweighted (equally weighted) versions of the adjacent period time dummy characteristics regressions can be constructed.**
- **We constructed both weighted and unweighted adjacent period time dummy characteristics regressions.**
- **P_{WATDC}^t is the Weighted Adjacent Period Time Dummy Characteristics Price Index for month t.**
- **Its unweighted (or equally weighted) counterpart index is P_{ATDC}^t.**
- **These indexes are listed in Table 6 below.**
- **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.**
- **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
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

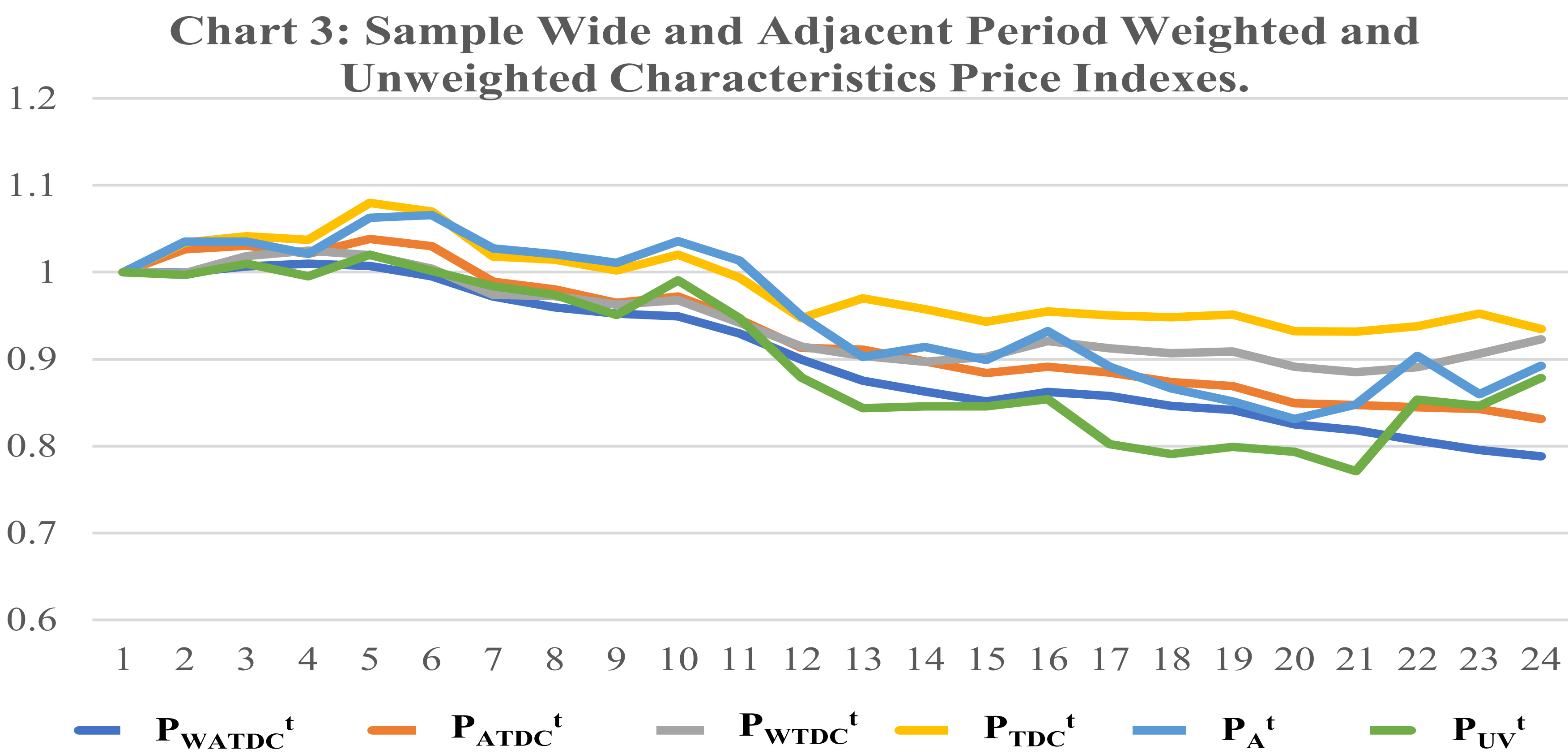


Chart 3: Sample Wide and Adjacent Period Weighted and Unweighted Characteristics Price Indexes.
1.2
1.1
1
0.9
0.8
0.7
0.6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
$P_{WATDC}{}^t$
$P_{ATDC}{}^t$
$P_{WTDC}{}^t$
$P_{TDC}{}^t$
$P_A{}^t$
$P_{UV}{}^t$

- **Our new adjacent period characteristics price indexes, P_{WATDC}^{24} and P_{ATDC}^{24}, finish well below their single regression counterpart indexes, P_{WTDC}^{24} and P_{TDC}^{24}.**
- **More importantly, the new indexes finish below the Average Price index P_{A}^{24} and the Unit Value index P_{UV}^{24}, so that there was some positive quality improvement in laptops over our sample period.**
- **Thus the new adjacent period indexes are more plausible than the corresponding single regression based indexes.**
- **Looking at the effects of weighting, it can be seen that the adjacent period equally weighted characteristics index P_{ATDC}^{t} finishes 4.3 percentage points above its weighted counterpart P_{WATDC}^{t} for t = 24 and on average, P_{ATDC}^{t} is 2.6 percentage points above the average for P_{WATDC}^{t}.**
- **Since this equally weighted index gives too much weight to unrepresentative products, we prefer the Weighted Adjacent Period Time Dummy Characteristics Index P_{WATDC}^{t}.**

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

- **The adjacent period indexes fit the data much better since each bilateral regression estimates a new set of quality adjustment parameters whereas the panel regression approach fixes the quality adjustment parameters over the entire window of observations.**
- **The adjacent period methodology that allows the quality adjustment parameters to change every month means that purchasers may not have stable consistent preferences over time and some economists may object to the resulting inconsistency of these indexes.**
- **There may be external environmental factors (that change over time) which affect the utility to purchasers of the products in scope. We are assuming that purchasers of the products in scope have preferences that are separable from other products which can only be a rough approximation to reality. Also, the "newness" or "oldness" of a product may affect purchaser utility. We will add "newness" as a price determining characteristic in section 8 below.**

6. Time Product Dummy Regression Models.

- **We have seen that missing characteristics can have a material effect on the price index.**
- **A model that includes all possible product characteristics is the Time Product Dummy model presented in section 2.**
- **Thus in this section, we will consider weighted and unweighted time product dummy hedonic regression models.**
- **In order to set up the unweighted regression problem for our particular application, we make use of the vectors of time dummy variables, $D_1, \ldots, D_{24}$, 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 (sample wide) 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 $\beta_1, \beta_2, \ldots, \beta_{366}$:**

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

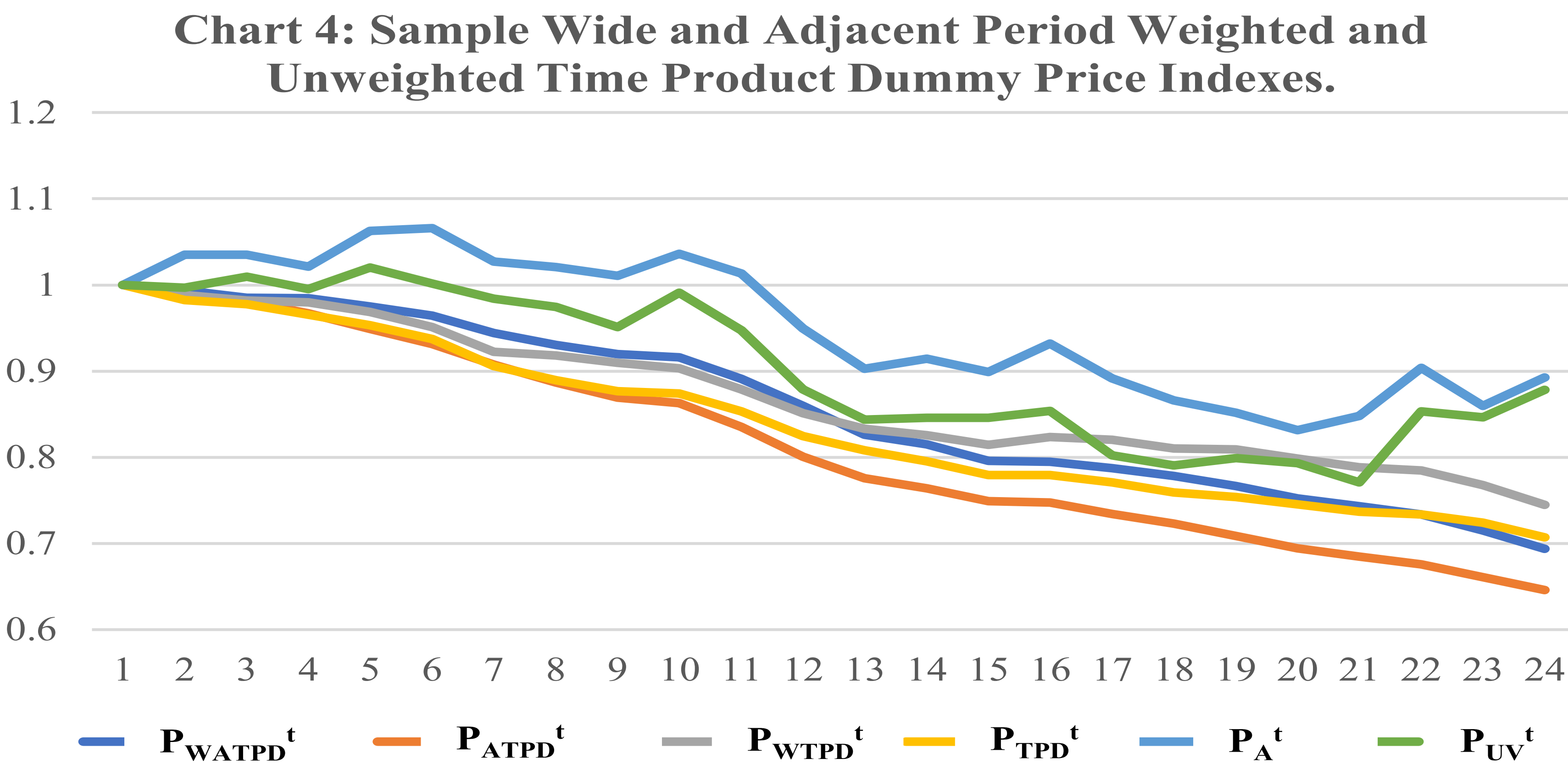
- **The R^2 for the above regression turned out to be 0.9836. We set ρ_t^* equal to one. The estimated ρ_t^* were exponentiated and the sequence of the $\pi_t^* \equiv \exp[\rho_t^*]$ are the *Time Product Dummy Price Indexes P_{TPD}^t* which are listed in Table 7 below.**
- **To obtain the Weighted Time Product Dummy Price Indexes, multiply the vectors on both sides of (78) (excluding the error vector e) by the vector of positive square roots of the month by month expenditure shares s_{tn} on the products which were purchased in each period.**
- **The resulting linear regression in the same parameters $\rho_2, \rho_3, \ldots, \rho_{24}$ and $\beta_1, \beta_2, \ldots, \beta_{366}$ was run and the R^2 for this weighted time product dummy regression turned to be 0.9840.**
- **Again, set ρ_1^* equal to one. The estimated ρ_t^* were exponentiated and the new sequence of the $\pi_t^* \equiv \exp[\rho_t^*]$ are the Weighted Time Product Dummy Price Indexes P_{WTPD}^t which are listed in Table 7 below.**
- **As in the previous section, we can calculate Adjacent Period Time Product Dummy regressions.**

- **As in the previous section, to obtain Weighted Adjacent Period Time Product Dummy Price Indexes, P_{WAPD}^{t} we took the 23 bilateral regressions that were used to form the unweighted indexes and multiplied the dependent and independent variables in each of these regressions by the square root of the appropriate expenditure share.**
- **Table 7 lists the Adjacent Period Weighted and Unweighted Time Product Dummy price indexes, P_{WATPD}^{t} and P_{ATPD}^{t}, as well as the simple average and unit value price indexes, P_{A}^{t} and P_{UV}^{t}.**
- **Chart 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
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

- **We prefer the weighted indexes over their unweighted counterparts.**
- **$P_{WATPD}{}^{t}$ (the red line) finished well below the single regression Time Product Dummy Price Indexes $P_{TPD}{}^{t}$ (the grey line). It may be that the Adjacent Period index $P_{WATPD}{}^{t}$ is subject to some chain drift ($P_{TPD}{}^{t}$ is not subject to chain drift).**



7. Similarity Linked Price Indexes for Laptops.

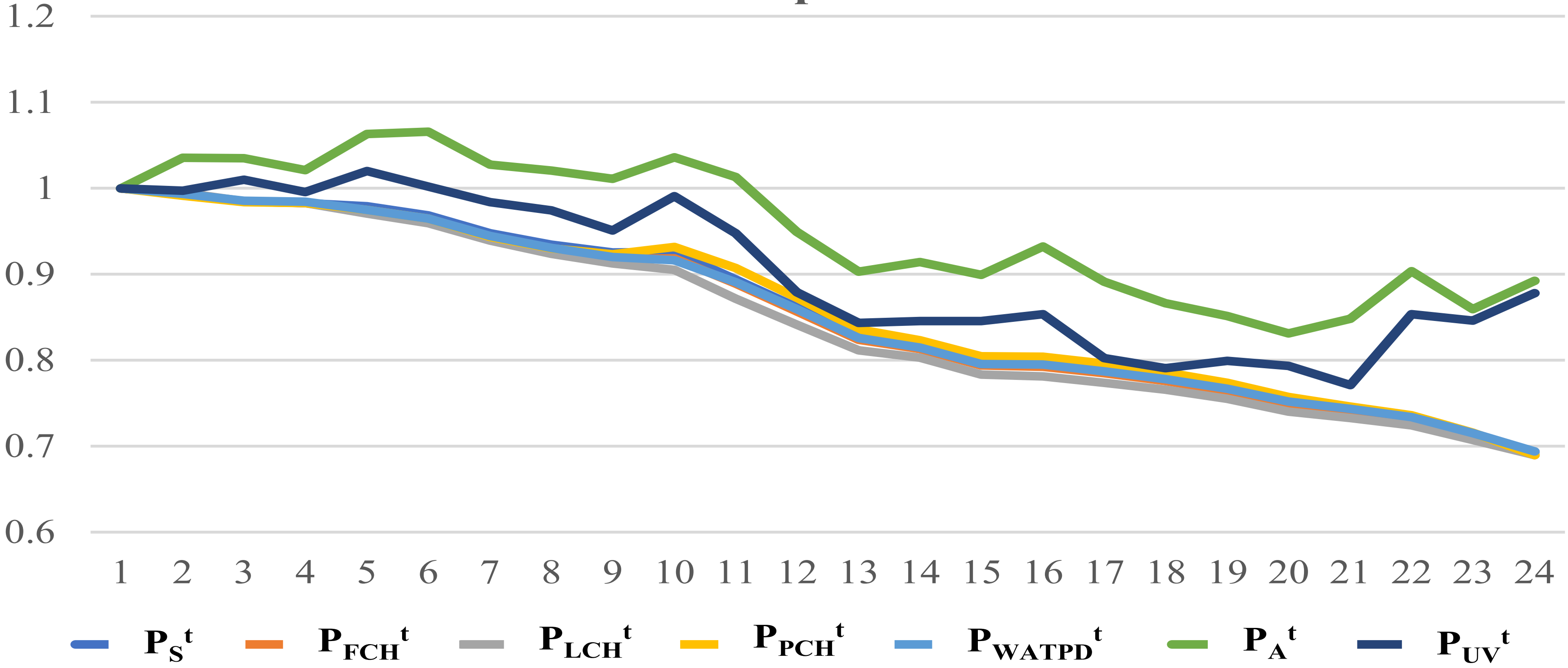
- **The indexes defined in the previous sections that made use of 23 adjacent period regressions were *chained* indexes; i.e., the index constructed for month t compared the prices for month t with the prices for month t – 1.**
- **However, it is not the case that all bilateral comparisons of prices 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) 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.**

- **The *Predicted Share* method of linking months with the most similar structure of relative prices is explained and implemented in the paper.**
- **However, the Predicted Share method generated indexes that are almost identical to the chained maximum overlap Fisher index.**
- **Moreover, the Predicted Share and Chained Fisher indexes were also very close to the Weighted Adjacent Period Time Product Dummy Price Indexes, P_{WAPD}^t, that were constructed in the previous section.**
- **This close correspondence is explained by the fact that the bilateral Weighted Time Product Dummy index formula is approximated to the second order around an equal (or proportional) price and quantity point by the Fisher index if there are no missing products.**
- **Thus we will not present the algebra for the Predicted Share method for forming price indexes in this presentation since it did not lead to a new index.**
- **The 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.Table 9 also 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

Chart 5: The Predicted Share Similarity Linked Price Index and Other Comparison Price Indexes.



8. Newness as a Characteristic

- **Consumers are sometimes willing to pay a premium for a product that has just appeared in the marketplace; i.e., some consumers are willing to pay a higher price for this type of product simply due to its "newness".**
- **Possible examples of this type of *fashion product* are certain types of clothing, cell phones and cars. In this section, we attempt to determine whether laptops are a fashion product.**
- **In principle, hedonic regression models can be used to determine whether a product is a fashion product.**
- **A special case of the characteristics model explained in section 3 can be used where there are only two characteristics: the product itself and the number of months that the product has been on the marketplace.**
- **Thus if p_{tn} is the price of product n in month t, the basic model is the following one:**

(85) $p_{tn} \approx \pi_t \alpha_n \delta_a$

- **where π_t is the month t price level, α_n is a product specific quality adjustment parameter for product n and δ_a is an additional quality adjustment parameter that adjusts the price according to the age "a" of product n sold in month t.**
- **If we assume that purchaser preferences are constant over the entire sample period, then t equals 1 to 24, n equals 1 to 366 and "a" equals 1 to 38 months.**
- **Define the vector of dimension 2639 of the logarithms of the observed laptop prices as lnP as was done in previous sections. Define the vectors of time dummy variables, D_1, …, D_{24}, as in section 4.1 above. This section also defined the 366 product dummy variable vectors of dimension 2639, D_{J1}, …, D_{J366}.**
- **The approximate model of price behavior defined by (85) was modified so that there were 7 age or newness cells instead of 38. Upon taking logarithms of both sides of the modified equations (85), we obtained the following approximate model:**

(86) $\ln p_{tn} \approx \rho_t + \beta_n + \gamma_a$

- **where $\rho_t \equiv \ln\pi_t$ for t = 1,…,24; $\beta_n \equiv \ln\alpha_n$ for $n \in S(t)$ and $\gamma_a \equiv \ln\delta_a$ where "a" indicates the appropriate age cell for observation n in month t. Recall that S(t) is the set of products that were sold in month t.**
- **Note that equations (86) form the basis for a linear regression but it can be seen that not all parameters on the right hand side of equations (86) can be identified. We require at least two normalizations (such as $\rho_1 = 0$ and $\gamma_1 = 0$) in order to uniquely determine the remaining parameters.**
- **Define $S(1,2) \equiv S(1) \cup S(2)$ as the set of products that were purchased in months 1 and 2.**
- **The new hedonic regression model based on equations (86) that used only the prices of months 1 and 2 is the following regression model:**

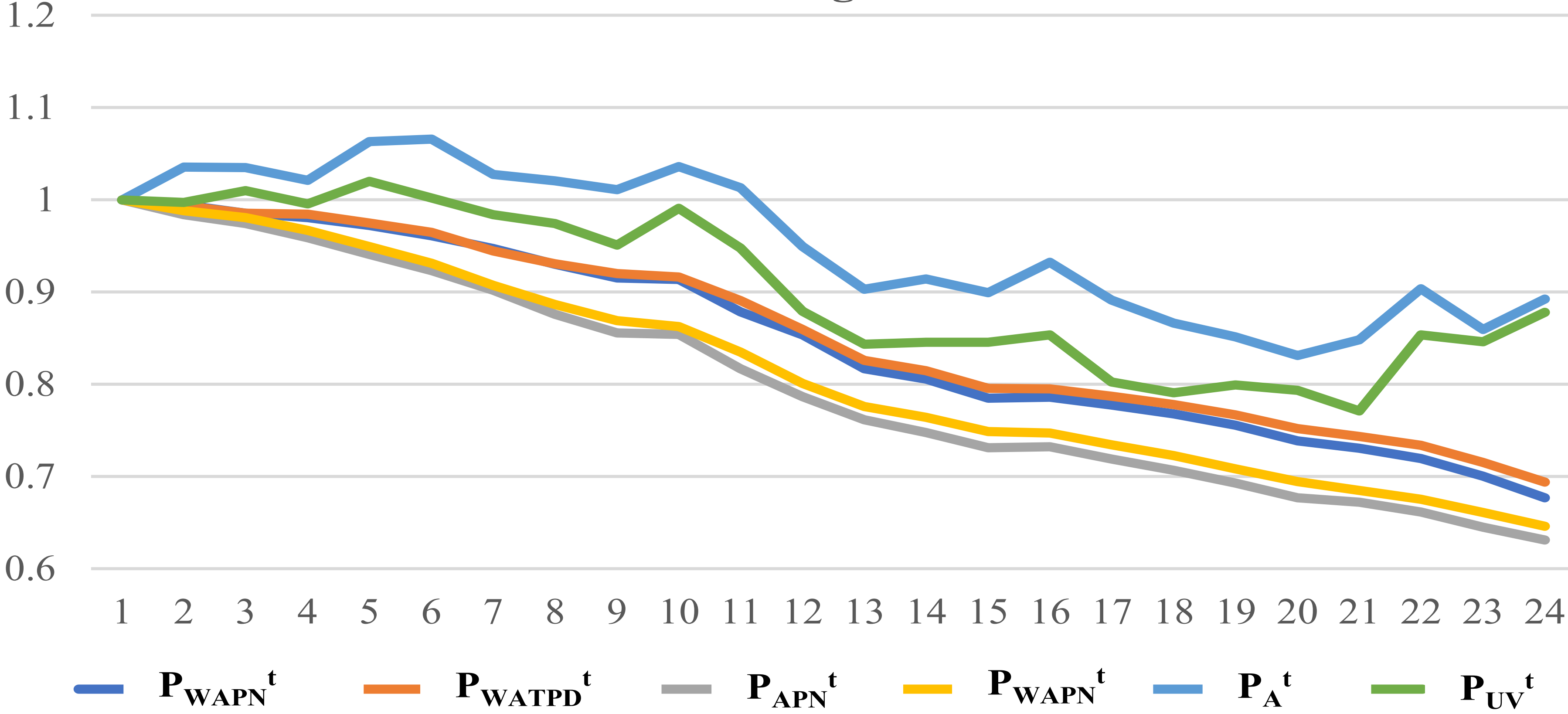
(87) $\ln P^* = \rho_2 D_2^* + \Sigma_{k \in S(1,2)}\, \beta_k D_{Jk}^* + \Sigma_{i=1}^{7}\, \gamma_i D_{Ai}^* + e^{t*}$

- **where the log price vector $\ln P^*$, the month 2 time dummy vector D_2^*, the product dummy variable vectors $D_{J1}^*, \ldots, D_{J366}^*$ and the newness dummy variable vectors D_{Ai}^* are restricted to products n that were actually sold in months 1 and 2.**

- **The results for the bilateral regression (87) can be compared to the corresponding Adjacent Period Time Product Dummy section 6 regression that excluded the age dummy variables.**
- **The log of the likelihood function increased by 47.38 log likelihood points for adding 5 new age parameters γ_1^*-γ_5^*. The R^2 for the new regression between the observed price vector and the predicted price vector was 0.9990 while the R^2 for the corresponding section 6 regression was 0.9989.**
- **The estimates for γ_1^*-γ_5^* were –0.0675, –0.0508, –0.0215, –0.0158 and –0.0296. These parameter estimates indicate that laptop purchasers were willing to pay a price premium for very old laptops that were available on the marketplace and thus laptops were not a fashion product.**
- **Instead of an age of product discount, there was a product premium for the oldest models.**
- **In general, the estimated gamma parameters did not show much consistency as we moved from one bilateral regression to the next one.**
- **Define the month t (Unweighted) Adjacent Period Newness Price Index that uses the product code and the "newness" of the product as price determining characteristics as P_{APN}^t for t = 1,…,24. The corresponding Weighted index is P_{WAPN}^t for t = 1,…,24.**

- **Newness did not add much to the TPD adjacent period regressions that did not use "newness". Below is a Chart that compares the newness series, P_{APN}^{t} and P_{WAPN}^{t} , to their counterparts without newness.**

Chart 6: Time Product Dummy Indexes with and without Newness and Average Price Indexes.



- **We prefer the weighted adjacent period index that used only the product code as a characteristic $P_{WATPD}{}^{t}$ over the index $P_{WAPN}{}^{t}$ that used product code and age of the product as characteristics for a number of reasons:**
- **A simpler model is in general preferred to a more complex model. Since adding age as a characteristic does not significantly change the index, why add age as a characteristic? The signs of the age parameters are not stable as we move from one bilateral regression to the next. This casts some doubt on the validity of the more general hedonic model that has age as a price determining characteristic.**
- **Adding age as an explanatory variable is not completely straightforward. In order to avoid multicollinearity problems, we had to modify our original definitions of the age cells. Thus the resulting index is not completely *reproducible*: different econometricians may define the age cells differently, leading to different indexes. On the other hand, treating the product code as a characteristic is (almost) completely reproducible.**
- **It is not completely reproducible because producers of some products relabel their products as being "new" when in fact, the product has not fundamentally changed.**

- **Treating age of the product on the marketplace as a fundamental utility enhancing (or reducing) characteristic introduces a psychological element into utility that seems somewhat different from other product characteristics that determine the usefulness of a product.**
- **In order to avoid possible arbitrary adjustments to the utility of products, the newness of a product should be allowed as a characteristic only if there is clear evidence that the age of a product influences its price.**
- **In retrospect, we should have run a sample wide regression with the newness variable to see if a clear difference emerged.**
- **There is also the suspicion that the use of adjacent period time product dummy regressions are subject to possible chain drift.**
- **We need to explore Rolling Window time product dummy regressions to check for chain drift.**

9. Conclusion.

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

- **If quantity or expenditure weights for products are available in addition to price information, then it is important to use these weights in the calculation of representative 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.**
- **The Adjacent Period Weighted Time Product Dummy index can be 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.**
- **However, it is difficult to rule out possible chain drift when using Adjacent Period Time Product Dummy regressions. It may be "safer" to use Rolling Window Time Product Dummy regressions. Determining how long the window should be is an open question.**
- **Newness should be introduced as a product characteristic with some degree of caution.**

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