NBER WORKING PAPER SERIES

HEDONIC PRICE INDEXES WITH UNOBSERVED PRODUCT CHARACTERISTICS, AND APPLICATION TO PC's

C. Lanier Benkard
Patrick Bajari

Working Paper 9980
http://www.nber.org/papers/w9980

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2003

We would like to thank Steven Berry, Gautam Gowrisankaran, Hidehiko Ichimura, Guido Imbens, Phillip Leslie, Ariel Pakes, Peter Reiss, Katja Seim, and three anonymous referees for many helpful comments. Both authors would also like to acknowledge the Hoover Institution and the National Science Foundation, SES-0112106, for support. Any remaining errors are our own. The views expressed herein are those of the authors and are not necessarily those of the National Bureau of Economic Research.

Hedonic Price Indexes with Unobserved Product Characteristics, and Application to PC's
C. Lanier Benkard and Patrick Bajari
NBER Working Paper No. 9980
September 2003
JEL No. E3, C1, D6

ABSTRACT

We show that hedonic price indexes may be biased when not all product characteristics are observed. We derive two primary sources of bias. The first is a classical selection problem that arises due to changes over time in the values of unobserved characteristics. The second comes from changes in the implicit prices of unobserved characteristics. Next, we show that the bias can be corrected for under fairly general assumptions using extensions of factor analysis methods. We test our methods empirically using a new comprehensive monthly data set for desktop personal computer systems. For this data we find that the standard hedonic index has a slight upward bias of approximately 1.4% per year. We also find that omitting an important characteristic (CPU benchmark) causes a large bias in the index with standard methods, but that this bias is essentially eliminated when the proposed correction is applied.

C. Lanier Benkard
Graduate School of Business
Stanford University
Stanford, CA 94305-5015
and NBER
lanierb@stanford.edu

Patrick Bajari
Department of Economics
Duke University
219B Social Science Building
Durham, NC 27708
and NBER
bajari@econ.duke.edu

1 Introduction

In recent years, U.S. statistical agencies have dramatically increased their use of hedonic methods in constructing official price indexes. While the first use of hedonic methods in the consumer price index did not occur until 1987[1], according to Landefeld and Grimm (2000), approximately 18 percent of U.S. GDP final expenditures are now deflated using indexes created using hedonic methods, and this number is rapidly growing.

Hedonic methods are being introduced into official indexes in order to correct for two well known problems with traditional matched model methods. First, in markets with rapid product turnover, the matched model index cannot be properly calculated because it is impossible to measure the prices of new products before they enter and old products after they exit. Pakes (2003) shows that if the matched model index is calculated only for those products that remain in the sample, then it is subject to a selection bias because the products that exit tend to be the ones that are less profitable. Second, the matched model index does not account for quality change. All price changes go into the index, even those associated with improvements in some product characteristics.

A long-standing problem with hedonic methods that has been widely recognized (Court (1939), Griliches (1961), Triplett (1969), Griliches and Ohta (1986)) but remains unresolved is that typically not all product characteristics are observable by researchers constructing price indexes. The importance of unobserved characteristics has been shown in recent work on demand estimation (e.g., Berry, Levinsohn, and Pakes (1995), Nevo (2001), Bajari and Benkard (2003), and others). Another indication that unobserved characteristics may be important is the fact that it is often the case that hedonic price regressions have a low goodness of fit as measured by the R^2. For example, Pakes (2003) reports R^2's for computers in the range of 0.26-0.52; Cockburn and Anis (1998) report

[1] See Moulton (2001). Other official indexes such as the Census Bureau's single family housing index and the BEA computer price index used hedonic methods prior to their adoption in the consumer price index.

R^2's for arthritis drugs in the range of 0.26-0.29.[2]

These observations motivate our three main research questions. First, what explains the errors made in the typical hedonic price regression? Candidate explanations include measurement error in prices, unobserved product characteristics, and approximation error due to functional form. The answer to this question is important because if price regression errors reflect, for example, only measurement error in prices, then all of the assumptions of standard hedonic methods are satisfied. Second, if the hedonic regression errors reflect unobserved product characteristics, to what extent is there a bias in the price index? Lastly, is it possible to construct hedonic price indexes that fully account for unobserved characteristics?

In section 2 we show that if some product characteristics are not observed then hedonic price indexes may be biased, and that this bias comes primarily from two sources.[3] The first is a classical selection problem that results when the average value of the unobserved characteristics for products in the market changes over time. In OLS estimates, the average value of the unobserved characteristics is absorbed into the period mean of the hedonic regression. This introduces a bias when the estimated hedonic surface from one period is used to predict the prices of products not observed in that period. For example, if the average value of unobserved characteristics is improving over time then, in later periods, hedonic methods would typically overpredict the prices of products that had dropped out of the sample in previous periods. In this example the price index would exhibit an upward bias.

The second source of bias is more subtle, and results from changes in the implicit prices of the unobserved characteristics over time. Consider the following simple example. Suppose that we wish to calculate a price index between two periods, t and $t+1$. For simplicity,

[2] Very low R^2's are not always the case. Berndt, Griliches, and Rappaport (1995) report R^2's of 0.77-0.83 for computers. Griliches (1961) reports R^2's in the range of 0.84-0.97 for automobiles.

[3] There is also a third source of bias, that we believe will be less important in practice, that occurs if the quantity weighted mean of the unobserved characteristics is substantially different from its unweighted mean.

assume that all products are observed in each period, so that both the matched model index and the hedonic index are defined and there is no selection problem. Assume that there are two observed characteristics, x_1 and x_2 (e.g., CPU speed and RAM), and one unobserved characteristic, ξ (e.g., quality). Suppose that the relationship between prices and product characteristics (both observed and unobserved) is linear so that in time period t:

$$p_{j,t} = \beta_{0,t} + \beta_{1,t} x_{1,j} + \beta_{2,t} x_{2,j} + \beta_{3,t} \xi_j.$$

Note that the price function is allowed to vary over time because the coefficients may change between periods.

Suppose that the econometrician is able to consistently estimate the intercept and the coefficients for the observed product characteristics, $\beta_{1,t}$ and $\beta_{2,t}$. Let $p_t(x_j)$ denote the predicted price of product j at time t using the hedonic surface:

$$p_t(x_j) = \beta_{0,t} + \beta_{1,t} x_{1,j} + \beta_{2,t} x_{2,j}.$$

In our example, it is easy to see that the matched model index and the hedonic price index differ due to changes in the valuation, $\beta_{3,t}$, of the unobserved characteristic. The matched model price adjustment between two periods t and $t+1$ for product j is

$$p_{j,t+1} - p_{j,t} = \beta_{0,t+1} - \beta_{0,t} + (\beta_{1,t+1} - \beta_{1,t}) x_{1,j} + (\beta_{2,t+1} - \beta_{2,t}) x_{2,j} + (\beta_{3,t+1} - \beta_{3,t}) \xi_j,$$

and the price adjustment using the hedonic surface is

$$p_{t+1}(x_j) - p_t(x_j) = \beta_{0,t+1} - \beta_{0,t} + (\beta_{1,t+1} - \beta_{1,t}) x_{1,j} + (\beta_{2,t+1} - \beta_{2,t}) x_{2,j}.$$

In this example, the matched model adjustment is the correct adjustment and the hedonic adjustment is incorrect. The hedonic adjustment leaves out the term that revalues the unobserved characteristic, $(\beta_{3,t+1} - \beta_{3,t})\xi_j$. Since the aggregate price index is a weighted average of the individual price adjustments, the aggregate price index typically would also be incorrect.[4]

[4] Note that in this simple example if $E[\xi_j | x_j] = 0$ then the hedonic adjustment is an unbiased estimate of the true adjustment. However, we show in section 2 that typically there would still be a statistical bias in the price index.

In section 2, we present our basic model and derive an analytical expression for the bias due to the two sources listed above. Table 1 provides an empirical example of the bias obtained in the price index when an important characteristic is known to be omitted. For our data on desktop personal computer systems (more details on the data provided in section 4.1), the table shows chained Fisher price indexes constructed using a standard hedonic approach (left) and then again using the same approach but with CPU benchmark omitted (right). As can be seen in the table, the indexes with CPU benchmark omitted exhibit substantial bias. Over just 29 months, the difference in overall inflation is approximately 9%, with the biased indexes showing less deflation. This variation is larger than any variation we were able to achieve through alternative methods of constructing the index or alternative functional forms. We therefore view it as potentially significant.

In section 3 we show how factor analysis methods can be extended to construct a fairly general statistical test for the presence, and even the dimension of the unobserved product characteristics. The intuition for this test is that, for products with similar values of the unobserved characteristics, the price regression errors should move similarly over time.

Next, we show how to use similar methods to construct hedonic indexes that account for the unobserved characteristics. If the unobserved characteristic is single dimensional, then it is possible to consistently estimate the hedonic surface (including recovering the unobserved characteristics) using a completely general functional form. If the dimension of the unobserved characteristics is two or greater, it is possible to consistently estimate the hedonic surface so long as there is a representation of the surface that is additively separable in the unobserved characteristics. Finally, we show that this methodology works, in general, if the unobserved characteristics are correlated with each other and, in certain cases, if they are correlated with the observed characteristics.

We apply our methods to a new data set for desktop personal computers. We find in this data that the dimension of the unobserved characteristics is likely to be either two and three. However, we also found that there was not enough data to get precise estimates of the price index when the unobserved characteristics were allowed to have dimension greater than one. Given the comprehensiveness of the data, this result sheds doubt on the practical ability to correct price indexes in the multi-dimensional case. However, these difficulties were exacerbated in our data by the extremely high rate of product turnover and the relatively high measurement error in the price data. Therefore, we believe that correcting for a multi-dimensional unobservable may be possible in other data sets with less rapid product turnover and better price measurement. Based upon the results of using a single dimensional unobserved characteristic, we find that the standard hedonic index is upwardly biased, by approximately 1.4% per year, and that this bias is primarily due to selection. Specifically, the unobserved characteristics for computers are improving over time, and this upwardly biases the standard hedonic price index.

We further test our estimation approach by leaving out an important characteristic (CPU benchmark) and reestimating the price index. While the standard hedonic index is severely biased in this case, our approach essentially removes the bias even if only a single dimensional unobserved product characteristic is used (see also Table 9). Thus, although the results above suggest that correcting the index for a multi-dimensional unobservable may be difficult, these results show that corrections based on a single dimensional unobservable provide a good approximation to the multi-dimensional case.

Our results suggest that there is a tradeoff between the hedonic and matched model approaches. Hedonic methods are better at capturing quality change and also can solve the product entry and exit problem of the matched model approach. However, hedonic methods may be biased due to unobserved characteristics. Our approach of including unobserved product characteristics in the hedonic index can be viewed as achieving a middle ground between the two standard approaches. Our approach also lies between the two standard approaches in terms of data requirements. A limitation of our approach is that products must be observed in several time periods, or several spatially separated markets, in order to estimate the vector of unobserved product characteristics. The number of periods required depends on the dimension of the unobserved characteristics.

This falls considerably short of that required to construct the matched model index, for which every product must be observed in every period.

2 An Expression for the Bias Due to Unobserved Characteristics in Hedonic Price Functions

2.1 Model and Notation

We assume that the econometrician has data for $t = 0..T$ periods and, without loss of generality, we assume that the base period of the price index is period $t = 0$.[5] We assume that each commodity, j, can be represented as a finite dimensional vector of attributes. In most applications, the economist does not observe all of the product attributes relevant to the consumer. Therefore, in the model, the economist perfectly observes the first K attributes, which we denote by the vector $x_j = (x_{j1}, ..., x_{jK})$, but does not observe an L-vector of attributes $\xi_j = (\xi_{j1}, ..., \xi_{jL})$.

Let $\mathcal{C}_t$ be the set of products in market t and denote the set of products that are available in both markets s and t as $\mathcal{C}_{s,t} = \mathcal{C}_s \cap \mathcal{C}_t$. Let F_t be the joint distribution of (x, ξ) in market t, with support $\mathcal{X}_t \subset \mathbb{R}^{K+L}$, where $\mathcal{X}_t$ is assumed to be compact.

Implicit in the notation above is the assumption that products are readily identifiable in the sense that it is possible to identify the same product across different time periods t. Under this assumption, the entire vector of product characteristics, (x_j, ξ_j), is fixed across markets for each product. If a product's characteristics change between two periods, then we define the two products to be different products.

The assumption that a product's characteristics stay fixed over time may be unrealistic

[5] It would not change anything in our analysis to consider instead either T spatially separated markets, or a total of T observations for a set of spatially separated markets over time.

in some industries. For example, if one characteristic of a product is the manufacturer's reputation for providing good service, then that could change over time even if the physical aspects of the product do not. Examples of this might include the average hold time on the company's customer service hotline.

2.2 Price Index Formulas

In this paper, we concentrate on what we believe are the most commonly used forms of the price index: plutocratic weighted average indexes with base period (Laspeyres') or reference period (Paasche) weights. We define the standard matched model indexes as follows:

$$\mathrm{M}_t^L = \frac{\sum_{j \in \mathcal{C}_0} p_{jt} q_{j0}}{\sum_{j \in \mathcal{C}_0} p_{j0} q_{j0}}, \tag{1}$$

$$\mathrm{M}_t^P = \frac{\sum_{j \in \mathcal{C}_t} p_{jt} q_{jt}}{\sum_{j \in \mathcal{C}_t} p_{j0} q_{jt}}. \tag{2}$$

Standard results show that M_t^L is an upper bound and M_t^P is a lower bound to the exact price index.[6] In the empirical section of the paper we find that, due to a high rate of product turnover, we must instead apply the "chained" versions of these indexes, in which the weights are constantly updated from one period to the next.[7] Therefore, we also calculate chained Fisher indexes since several papers (e.g., Aizcorbe, Corrado, and Doms (2003)) have argued that the chained Fisher index provides a better approximation to the true index in markets with high product turnover.

Hedonic methods substitute prices predicted from the estimated hedonic surface, $p(x)$, into (1) and (2) in the place of actual prices, not all of which are observed. The primary differences in hedonic methods arise in the details of how prices are predicted and whether

[6] These correspond to the classical bounds of Konus (1924). See also Pakes (2003) for ways of deriving these bounds more generally.

[7] Note that the chained forms of these indexes no longer represent proper bounds and are therefore simply approximations to the true index.

the predicted prices should always be used, or whether they should only be used where actual prices are unavailable, or some combination of these options. In this paper, we compare the matched model indexes (M_t^L and M_t^P) with hedonic indexes (H_t^L and H_t^P) in which all of the prices are replaced with prices predicted by the hedonic index:

$$\mathrm{H}_t^L = \frac{\sum_{j \in \mathcal{C}_0} p_t(x_j) q_{j0}}{\sum_{j \in \mathcal{C}_0} p_0(x_j) q_{j0}} \tag{3}$$

$$\mathrm{H}_t^P = \frac{\sum_{j \in \mathcal{C}_t} p_t(x_j) q_{jt}}{\sum_{j \in \mathcal{C}_t} p_0(x_j) q_{jt}} \tag{4}$$

Our approach differs slightly from the methods proposed by Pakes (2003), which substitutes all prices in the numerator with prices predicted using the hedonic surface, but uses actual prices in the denominator. It also differs from the method used by the BLS, which uses a hybrid of the hedonic and matched model methods that substitutes predicted prices only in cases where products drop out of the sample. However, our techniques can just as easily be applied to hedonic indexes of those forms.[8]

2.3 An Analytical Expression of the Unobserved Characteristics Bias for the Linear Case

In order to better understand how unobserved characteristics lead to bias in the aggregate price index, in this section we derive analytical expressions for the bias in the index. Because this is difficult to do in general, we concentrate on the simple case in which the price function is linear. We thus write the price function as follows,

$$p_{jt} = \beta_{0,t} + x_j' \beta_{x,t} + \xi_j' \beta_{\xi,t}, \tag{5}$$

where both x_j and ξ_j are vectors.

[8] The alternative methods suggested in Pakes (2003) and used by the BLS are in part designed to allow construction of the index under time constraints. We ignore these practical issues in this paper.

In characterizing the bias in H^L, it is helpful to rewrite the index as follows:

$$\mathrm{H}_t^L = \frac{\sum_{j\in\mathcal{C}_0} p_t(x_j)q_{j0}}{\sum_{j\in\mathcal{C}_0} p_0(x_j)q_{j0}} = 1 + \frac{\sum_{j\in\mathcal{C}_0}(p_t(x_j) - p_0(x_j))q_{j0}}{\sum_{j\in\mathcal{C}_0} p_0(x_j)q_{j0}} \tag{6}$$

where the functions $p_t(x_j)$ are the hedonic surface in period t which are only a function of the observed characteristics as is common in practice.

Consider what happens if we estimate (5) using standard techniques. Suppose, for the sake of simplicity, that we estimate (5) under the assumption that ξ and x are mean independent, $E_t[\xi|x] = E_t[\xi] = \mu_t$. If the mean independence assumption holds and there are a large number of observations in each period, then the parameter estimates obtained from the T regressions are:

$$\tilde{\beta}_{0,t} \approx \beta_{0,t} + \mu_t'\beta_{\xi,t} \tag{7}$$

$$\tilde{\beta}_{x,t} \approx \beta_{x,t} \tag{8}$$

Note that the intercept captures the average change over time in both the price of ξ and the mean of ξ.

We can now use (7) and (8), in conjunction with (6), to characterize the bias in H. The bias in the numerator of H^L is:

$$\mathrm{Bias}(\mathrm{Num}(\mathrm{H}^L))_t = (\mu_t - \mu_0)'\beta_{\xi,t}Q_0 + \sum_{j\in\mathcal{C}_0}(\mu_0 - \xi_j)'(\beta_{\xi,t} - \beta_{\xi,0})q_{j0} \tag{9}$$

where Q_0 is total sales of the good in the base period. Similarly, the bias in the numerator of H^P is,

$$\mathrm{Bias}(\mathrm{Num}(\mathrm{H}^P))_t = (\mu_t - \mu_0)'\beta_{\xi,0}Q_t + \sum_{j\in\mathcal{C}_t}(\mu_t - \xi_j)'(\beta_{\xi,t} - \beta_{\xi,0})q_{jt} \tag{10}$$

where Q_t is total sales of the good in the reference period.

The expressions for the bias in the numerator involve two main terms. The first term depends on how much the mean of ξ changes over time and therefore reflects selection bias. If there is no selection, such that mean of ξ is constant over time, then the first term is zero. The second term reflects the extent to which the unobserved characteristics are revalued over time, $(\beta_{\xi,t} - \beta_{\xi,0})$. If the value of the unobserved characteristics is constant over time then the second term is zero. These are the two sources of bias mentioned in the introduction.[9]

The expressions for the bias in the denominator involve similar terms. The bias in the denominator of H^L is

$$\text{Bias}(\text{Den}(\mathrm{H}^L))_t = \beta_{\xi,0} \sum_{j \in \mathcal{C}_0} (\mu_0 - \xi_j)' q_{j0}, \tag{11}$$

while the bias in the denominator of H^P is

$$\text{Bias}(\text{Den}(\mathrm{H}^P))_t = \beta_{\xi,t} \sum_{j \in \mathcal{C}_t} (\mu_t - \xi_j)' q_{jt}. \tag{12}$$

The bias in the denominator of the index reflects the extent to which the quantity weighted mean of the unobserved characteristic differs from its unweighted mean. It is difficult to sign this bias in general since the quantity weights depend on consumer tastes. Assuming that the unobserved characteristics carry positive prices, if demand is higher for goods with higher values of the unobserved characteristics, then the denominator is downwardly biased, leading to an upward bias in the index. Note also that the bias is constant over time for the Laspeyre's index, and likely to be fairly constant for the Paasche index. This means that if the denominator is biased downward, then price changes in the index for all periods will be biased upward. Based on our experiences, our prior is that this source of bias is likely to be less important than the previous two.

The bias in the index as a whole is easiest to evaluate asymptotically since, by the Slutzky theorem, the bias in the index can then be evaluated by considering the biases in the numerator and denominator separately. This leaves three overall sources of bias: two in

[9] Note that revaluation of the unobserved characteristics would not bias the index if there was no selection and the quantity weighted mean of ξ was the same as its unweighted mean, μ. Interestingly, this suggests that, if there were no selection problem, using quantity weights in the hedonic regression would eliminate the unobserved characteristics bias in the index.

the numerator (selection and repricing of the unobserved characteristic) and one in the denominator (the difference between the quantity weighted mean of ξ and its unweighted mean). The bias in the index will reflect the sum of these three sources.

In our opinion, there are many industries in which the mean of the unobserved characteristics and the price of the unobserved characteristics are likely to change over time, particularly high technology industries. In that case, it is likely that there would be unobserved characteristics bias in standard hedonic indexes.

3 Modeling Unobservables in the Hedonic Price Function

In this section, we outline an approach to estimating hedonic price functions in the presence of unobserved characteristics. Our approach is similar to the factor analysis literature, especially Lawley and Maxwell (1971), Goldberger (1974), and Cragg and Donald (1995, 1997), except that we have found it necessary to extend that literature in several ways, most notably to account for selection.

3.1 The Hedonic Price Function

Bajari and Benkard (2003) provide a set of primitive conditions under which there exists a price surface, denoted $p_t(x_j, \xi_j)$, in each market t. For the remainder of the paper we implicitly rely on the results of this theorem in the sense that we assume that there exists a function mapping product characteristics to prices.[10]

[10] In Bajari and Benkard (2003), ξ_j is single dimensional. However, extending the theorem to the case in which ξ is L-dimensional is straightforward. If the assumptions of the theorem were to not hold, then the hedonic approach could still be viewed as an approximation to the truth. However, we cannot say how good the approximation would be without making additional assumptions.

To simplify the analysis and estimation, we assume that the price function can be written as additively separable in the observed and unobserved product characteristics and linear in the unobserved characteristics,

$$p_{jt} = f_t(x_j) + \beta'_{\xi,t}\xi_j + \nu_{jt}, \tag{13}$$

where $f_t(\cdot)$ is a function, possibly of unknown parametric form, and ν_{jt} represents measurement error in the observed price.

Equation (13) places some restrictions on the functional form of the price function, but retains perhaps more generality than it first appears. All of the analysis that follows is general to nonlinear transformations of the right and left hand side variables. Additionally, $f_t(\cdot)$ can be a general nonparametric function within any one of those forms. Since we allow for the unobserved product characteristics to be correlated with each other, higher order terms in ξ may appear as additional dimensions.[11] Since our analysis is general to the case where ξ is correlated with x (if the relationship is stable over time — see section 3.3), interactions between ξ and x may also appear as additional dimensions. We allow for measurement error in prices because in our experience with price data in I.O. applications we have found this can happen for a variety of reasons, and furthermore, we believe it to be true in our data.

For ease of exposition, in this section we maintain several assumptions that we later relax. First, we assume that the unobserved product characteristics are mean independent of the observed product characteristics. This assumption is common in the hedonics literature implicitly, and also in the literature on demand estimation explicitly. It seems likely that it is violated to some extent in practice so we show that it is possible to substantially relax this assumption in section 3.3. Second, we assume that the measurement error is *iid* and independent of x and ξ. It is straightforward to generalize the specification of the measurement error in several ways, including AR(p), and heteroskedasticity of unknown

[11] For example, if the price surface was a function of ξ_{j1} and ξ_{j1}^2, then we could let $\xi_{j2} \equiv \xi_{j1}^2$ and the price surface would be of the form in (13). If the price surface was a function of ξ_{j1}, ξ_{j2}, and $\xi_{j1} * \xi_{j2}$, then we could let $\xi_{j3} \equiv \xi_{j1} * \xi_{j2}$ and the price surface would be of the form in (13).

form. We consider the latter case below. Finally, the analysis is substantially easier to follow if we assume that there is no selection in that data, i.e., we assume that the distribution of ξ_j is constant over time. We add selection to the model in section 3.2.

3.1.1 Estimating $f_t(\cdot)$

Let $\epsilon_{jt} \equiv \beta'_{\xi,t}\xi_j + \nu_{jt}$ represent the error terms in the period-by-period hedonic price regressions. Under the assumptions listed above, $E_t[\epsilon_{jt}|x] = \beta'_{\xi,t}\mu$, where $\mu = E[\xi_j|x_j]$ is a constant. Therefore, the functions $f_t(\cdot)$ can be estimated using standard techniques. For example, if the parametric form of the functions $f_t(\cdot)$ is known, then they could be estimated using least squares. Otherwise, kernel or series based nonparametric regression techniques could be used. Note that the $f_t(\cdot)$ functions absorb the mean of the unobservable, $\beta'_{\xi,t}\mu$, so at this point the functions can only be estimated up to an additive constant term.[12] Because these estimation approaches are standard, we omit a detailed discussion of them and proceed as if $f_t(\cdot)$ were known.

3.1.2 Estimating $\beta_{\xi,t}$

What makes it possible to identify and estimate the complete model (13) is the fact that this model places tight restrictions on the covariance matrix of the errors in the hedonic regressions, ϵ_{jt}. In order to derive those restrictions we need first to make some normalizations. The normalizations we use are standard to factor analysis and are without loss of generality.[13] We normalize the mean of ξ to be zero, $E[\xi_j] = 0$. We also normalize ξ to have covariance matrix I_L across all periods $0..T$. The reason that the normalizations are necessary is that ξ is not observable and thus has no inherent units. It is multiplied by a coefficient vector that is also unknown. Thus, neither the mean nor the variance of ξ

[12] If there is selection, then each function $f_t(\cdot)$ absorbs the period mean of the unobservable, $\beta'_{\xi,t}\mu_t$.

[13] See Lawley and Maxwell (1971) for a good discussion. We remind readers that at this point we are maintaining the assumption that there is no selection. In the event that there is selection, we have to be careful in applying the normalizations. See section 3.2 for details.

is identified separately from the coefficients β_ξ. Importantly, knowledge of the normalized coefficients is sufficient for construction of the price index.

Let ϵ_j be the T-vector of errors for product j. Then under the assumptions given above,

$$\Sigma \equiv E[\epsilon_j \epsilon_j'] = \beta_\xi \beta_\xi' + \sigma_\nu^2 I_T. \tag{14}$$

Without any restrictions, $E[\epsilon_j \epsilon_j']$ has $\frac{T(T+1)}{2}$ unique elements. However, our model contains only $T * L + 1$ parameters. Thus, for small values of L the model places significant restrictions on this matrix. In fact, it is possible to estimate the entire matrix of parameters β_ξ so long as $L \leq \frac{T}{2}$ (approximately). Since most price index applications have data for a large number of time periods or spatially separated markets, the model is typically overidentified for reasonable values of L.

Estimation of β_ξ can be achieved in several ways. The traditional approach of the factor analysis literature (e.g., Lawley and Maxwell (1971)) has been to assume normality for the unobserved product characteristics and then use maximum likelihood. However, such an approach would be inappropriate here because the model provides us with only first and second moment information and nothing more. If we were to assume normality of the unobserved product characteristics, then in conjunction with the normalization of their covariance matrix to the I_L matrix, we would be implicitly assuming full independence of the unobserved product characteristics. We do not want to assume full independence because we want to allow for functional form flexibility in (13). Thus, we instead proceed using GMM with the moment conditions provided by (14).[14]

Assuming that there is no selection, the model can be estimated as follows. Let

$$S = \frac{1}{J} \sum_{j=1}^{J} \epsilon_j \epsilon_j'. \tag{15}$$

[14] In a previous version of the paper we used likelihood methods and found that they led to an overestimate of the number of unobserved product characteristics, L. We also ran monte carlo experiments in which there were two unobserved characteristics: ξ_j, and ξ_j^2. The GMM approach below correctly identified that there were two unobserved characteristics. The MLE approach needed anywhere from 5-8 (independent) unobserved characteristics to match this data.

Since (15) is the empirical counterpart to (14), our model gives us $T(T+1)/2$ unique moment conditions,

$$E[S] = \Sigma. \tag{16}$$

The natural GMM estimator would minimize a quadratic form in these moment conditions,

$$\{\hat{\beta}_\xi, \hat{\sigma}_\nu\} = \arg\min(\text{vech}S - \text{vech}\Sigma)' A(\text{vech}S - \text{vech}\Sigma)$$

for some positive definite weight matrix A. Under standard conditions, $\hat{\beta}_\xi$ and $\hat{\sigma}_\nu$ are consistent and asymptotically normal for any positive definite weight matrix A. For example, the I matrix could be used. Also under standard conditions,[15]

$$\sqrt{J}(\text{vech}S - \text{vech}\Sigma) \rightarrow N(0, V),$$

and it is well known that the optimal weight matrix to use in the GMM objective function is $A = V^{-1}$.[16]

3.1.3 Hypothesis Tests for the Dimension L

The above estimation algorithm is conditional on knowing the dimension L. Cragg and Donald (1997) shows that if the optimal weight matrix is used, then the value of the objective function can also be used as a statistical test for the true dimension of the model. The difficulty of applying the approach of Cragg and Donald (1997) in our application comes in estimating V. Typically, a consistent estimator of V can be obtained using the sample moments of ϵ_j. For example, an estimator for the covariance between the (q, r) and (s, t) elements of S is given by,

$$\frac{1}{N}\sum_{j=1}^{N}(\epsilon_{j,q}\epsilon_{j,r}\epsilon_{j,s}\epsilon_{j,t} - S_{q,r}S_{s,t}).$$

[15] In our application the number of observations will typically vary by cell of S so the asymptotic approximations have to be corrected appropriately.

[16] See Hansen (1982).

However, in our application products tend not to last longer than about twelve months, so there are many combinations of (q, r, s, t) for which there are very few or even zero observations. Thus, while it is still possible to estimate V, it is not possible to estimate it very well, and in our experience not well enough to construct reliable hypothesis tests.[17] We solve this problem by using subsamples of data for which V can be estimated well.

A potential problem with these hypothesis tests is that the errors used to calculate the moment conditions (16) are estimated and are therefore not equal to the true error terms. While this does not affect consistency of the index, the additional noise might influence the hypothesis tests toward causing false rejections (i.e., toward supporting too many unobserved factors). The extent of the problem would likely depend on the number of first stage observations and the variance of the measurement error in price. One possible solution to this problem would be to estimate the first and second stages jointly using GMM.[18] That is, the first stage consists of a set of OLS moments (one set for each time period) given by

$$E[\epsilon_{jt}|x] = 0.$$

These moments could be combined with those in (16) in one large joint GMM estimation procedure. Hypothesis tests based on the joint GMM objective function would then account for first stage estimation error. The problem with the joint approach is that it has a massive data requirement. Each time period in the first stage estimation adds a set of K moment conditions, leading to a total of $\text{NMOM} = T * (K + (T + 1)/2)$ moments. In order to run hypothesis tests it is necessary to obtain a good estimate of the variance covariance matrix of the moment conditions, which has $\text{NMOM} * (\text{NMOM} + 1)/2$ unique elements. For many data sets, including the one used in this paper, this will not be possible. We discuss this issue further in the empirical section of the paper.

[17] One problem we had was that the differing number of observations in every cell led to an estimate of V which was not positive definite due to sampling error, and thus not invertible to obtain the weight matrix.

[18] We thank an anonymous referee for suggesting this solution.

3.1.4 Estimating ξ

The two-step approach above provides estimates of all of the parameters of the model. However, in order to construct price indexes, it is also necessary to estimate the vector of unobserved product characteristics for each product. The vector of errors for each product j can be written as

$$\epsilon_{jt} = \beta'_{\xi,t}\xi_j + \nu_{jt}. \tag{17}$$

At this point we assume that the parameters $\beta'_{\xi,t}$ are known because they have been previously estimated.

Since $\beta_{\xi,t}$ is known and the measurement error is *iid* and independent of everything, equation (17) becomes a standard linear regression model with $\beta_{\xi,t}$ as the observed covariates and ξ_j as the unknown parameter vector. Estimation of the equation is straightforward via OLS. A problem that is likely to be encountered is that (17) can only be estimated for those products whose prices are observed in $T_j \geq L$ periods and, depending on the variance of the measurement error, can only be estimated well if T_j is large. In that case, if $L > 1$ it is not in general possible to estimate ξ_j for all products, and it may be difficult to estimate ξ_j well unless T_j is large or the variance of the measurement error is small. This also introduces some selection into the index as some products would have to be dropped in calculating the index.

In application, $\beta_{\xi,t}$ is not known, but is instead replaced by a consistent estimator $\hat{\beta}_{\xi,t}$. This introduces finite sample bias into the estimates of ξ similar to that of measurement error in the standard regression model. Since $\hat{\beta}_{\xi,t}$ is consistent as the number of products goes to infinity, this bias goes to zero with the number of products. Provided that there is measurement error ($\sigma^2_\nu > 0$), consistency of $\hat{\xi}_j$ also requires the number of time periods (or spatially separated markets) to become large for each product. Consistency of $\hat{\xi}_j$ would thus be obtained as both the number of products and the number of time periods become large.

3.2 Selection

An important problem with proceeding using the GMM approach described above is that there is substantial selection in our data for PC's.[19] As technology improves, lower quality products exit while higher quality products enter. Thus, for example, it is unlikely that the products that we observe in period 1 are a random sample of products from the distribution of all products observed in all periods, as is required by the moment conditions (16). In this section we allow there to be selection on both observed product characteristics, x_j, and unobserved product characteristics, ξ_j. We continue to assume that the measurement error in price is *iid* and thus not subject to selection.

Selection introduces two main problems to the analysis. The first is that, even if the mean of the unobserved product characteristics is normalized to zero overall, the mean of the unobserved product characteristics is not necessarily zero among products observed in any given period: $\mu_t \equiv E_t[\xi_j|x_j] \neq 0$.[20] One way that this shows up is that the errors from the hedonic price regressions, $\epsilon_{j,t}$, will include a term in the period mean of the unobserved characteristics,

$$\epsilon_{j,t} = \beta'_{\xi,t}(\xi_j - \mu_t) + \nu_{jt}.$$

This extra term must be accounted for when estimating ξ_j. This can be done using multivariate and partitioned regression techniques or an equivalent iterative procedure.

The second problem caused by selection is that we only observe the covariance of the errors in the price regression between two periods for products that are observed in both periods. If selection influences these covariances, then it is impossible to calculate sample moments that correspond to the population moments given by (16). Formally, for any pair of periods $s, t \in 0..T$, the moments in (16) represent $E[\epsilon_{j,s}\epsilon_{j,t}]$. Instead, we observe the sample counterpart to the population moment, $E[\epsilon_{j,s}\epsilon_{j,t}|j \in \mathcal{C}_{s,t}]$, where $\mathcal{C}_{s,t}$ represents the set of products observed in periods s and t. Therefore, if we were to proceed as described

[19] We are grateful to Ariel Pakes for helping us to clarify our thoughts in this section.

[20] The same is true among products observed in any pair of periods s and t.

above and ignore the selection problem, we may bias the estimates of β_ξ as well as the statistical dimension tests.

Our approach to handling selection is two-fold. First, in running the hypothesis tests for the dimension L on subsamples as described above, instead of using all data points for all products observed at any point during the subsample, we reduce the data to a balanced panel. Formally, we choose a balanced panel, $\mathcal{C}_{s..t}$, representing all products observed in all periods, $s, ..., t$. In running the statistical dimension tests, we then use the moment conditions in (16) but only for products in the balanced panel,

$$E[S|j \in \mathcal{C}_{s..t}] = \Sigma.$$

Note that the fact that we are using a balanced panel does mean that we are using a selected group of products. For example, because these products were observed over the entire panel, they are likely to be better than products that dropped out at some point. The way that this selection would show up is that the unobserved characteristics in the balanced panel would have a different (perhaps higher) mean and different covariance matrix than an unselected sample would. However, the mean and covariance matrix are normalized away in the estimation so the fact that they are different than what would be obtained without selection does not matter. What matters for the estimation is that the mean and covariance matrix are held constant across the moment conditions (the entire matrix S). Holding the selection constant over the panel allows us to discern common movements in the price regression errors, which allows us to estimate the coefficients, β_ξ. As long as there are enough products in the balanced panel to identify the coefficients, the selection no longer matters. The downside of the balanced panel approach is that it forces us to throw out some of the information available in the data. The benefit is that it allows us to test for the dimension L while allowing for selection without restrictions.

The approach of using a balanced panel does not allow us to estimate the price index as a whole because very few, if any, products are observed in every period in the data. However, we can extend the intuition of the balanced panel forward in several ways. Perhaps the easiest approach would be to chain together balanced panels for several subsamples of the data in order to construct the whole index. This approach should in principle work, but at the expense of not using all of the available information in the data. Instead, we propose using the moment conditions from overlapping balanced panels in conjunction with each other in order to estimate the overall index. However, in order to do this, we have to explicitly account for the varying selection across different panels.

Suppose that we apply the factor analysis normalizations relative to the products in some group $\mathcal{C}_\eta$, such that the ξ's for those products have mean zero and covariance matrix I_L. For example, the group $\mathcal{C}_\eta$ could be the balanced panel of all products observed in all periods 1..10. Then, as above, these products provide us with a series of moment conditions,

$$E[S|\mathcal{C}_{1..10}] = \Sigma_\eta \equiv \beta_\xi \beta_\xi' + \sigma_\nu^2 I.$$

Now, consider a second group of products, $\mathcal{C}_\tau$. For example, $\mathcal{C}_\tau$ could be the balanced panel of all products observed in all periods 2..11. If we allow the selection process to be completely unrestricted, then we know nothing about the mean and variance of ξ among this second group relative to the normalization from the first group. However, we still have an equivalent set of moment conditions,

$$E[S|\mathcal{C}_{2..11}] = \Sigma_\tau \equiv \beta_\xi \Psi_\tau \beta_\xi' + \sigma_\nu^2 I,$$

where Ψ_τ is the covariance matrix of ξ_j among products in $\mathcal{C}_\tau$. Because of the presence of the new parameters, Ψ_τ, this second set of moment conditions does not provide as much information as the first. However, for small dimensions L, they should still provide a great deal of information. For example, if $L = 1$ then these moment conditions are simply shifted by a constant relative to the first group. In this manner we can use many moment conditions from successive overlapping balanced panels to estimate the parameters β_ξ while allowing for selection without restriction. Note that this procedure introduces new incidental parameters, Ψ_τ, for each set of moment conditions used, and therefore increases the computational burden of the estimation.

3.3 The Non-Independent Case

We have already shown that if the unobserved product characteristics are correlated with each other, but independent of the observed characteristics, then we can estimate the price index by normalizing them to be uncorrelated with each other. In this section, we consider the case where the unobserved product characteristics are also correlated with the observed product characteristics.

We consider the case where the functions $f_t(\cdot)$ are estimated using a nonparametric series estimator. This approach is suggested by Pakes (2003) and also used in the empirical section of this paper. It also nests many parametric approaches, including linear, semi-log, and log-log, and can be viewed as an approximation to other nonparametric approaches. In that case, the price equation can be written as,

$$p_{jt} = \beta_{0,t} + \beta'_{x,t}\phi(x_j) + \beta'_{\xi,t}\xi_j + \nu_{jt}, \tag{18}$$

where $\phi(x_j)$ is a $M \times 1$ vector of basis functions of x_j and $\beta_{x,t}$ is a $M \times 1$ vector of parameters.

Without loss of generality, ξ_j can be written as,

$$\xi_j = \mu_t + \gamma_t\phi(x_j) + \zeta_{jt}, \tag{19}$$

where $E_t[\zeta_{jt}|\phi(x_j)] = 0$ and γ_t is a $L \times M$ matrix of parameters. The expression (19) represents the period t projection of ξ_j on $\phi(x_j)$ with respect to the period t sampling distribution of (x_j, ξ_j).[21] In general, the relationship between ξ_j and $\phi(x_j)$ may vary over time, depending on such things as changes in production technology and selection. However, suppose the projection were stable over time, such that

$$\xi_j = \gamma\phi(x_j) + \zeta_j, \tag{20}$$

where $E_t[\zeta_j|\phi(x_j)] = \mu_t$ for all time periods, t. Then combining (18) and (20) gives

$$p_{jt} = \beta_{0,t} + (\beta'_{x,t} + \beta'_{\xi,t}\gamma)\phi(x_j) + \beta'_{\xi,t}\zeta_j + \nu_{jt}, \tag{21}$$

[21] For clarification, μ_t and γ_t are the coefficients from a regression of ξ_j on $\phi(x_j)$ for all products observed in period t.

This model is analogous to that estimated above for the independent case. Thus, using the approach outlined above, we can consistently estimate the quantities $\beta_{0,t}$, $(\beta'_{x,t} + \beta'_{\xi,t}\gamma)$, $\beta_{\xi,t}$, and ζ_j, under the correct assumption that $E_t[\zeta_j|\phi(x_j)] = \mu_t$. The quantities $\beta_{x,t}$, γ, and ξ_j are not separately identified/estimable using this approach. However, we do not require these quantities in order to evaluate $p_t(x_j, \xi_j)$. The estimable quantities are sufficient for evaluation of this function and thus sufficient for construction of the hedonic price index. Thus, so long as the relationship between ξ_j and x_j is stable over time, the estimation approach described in the previous section provides consistent estimates of the price index.

Though it is substantially more general than assuming that they are independent, the assumption that the relationship between ξ_j and x_j is stable over time is somewhat restrictive, particularly if we were considering long panels. However, it would also be possible to use balanced panels in the first stage hedonic regressions to solve the correlation problem more generally. The reason balanced panels would work is the same as before. They hold the set of products fixed over time, thus making the relationship between ξ_j and x_j fixed over time.

4 Empirical Results

4.1 Data

Our data comes from *PC Data Retail Hardware Monthly Report* and includes quantity sold, average sales price, and a long list of machine characteristics for desktop computers sold over a 29 month period from August 1997 to December 1999. The data set reportedly covers approximately 75% of U.S. retail computer sales. The price data is collected from cash register receipts and is constructed by taking total sales of each product over a month and dividing by quantity sold. It therefore represents the average retail sales price of the machine in that month.[22]

The characteristics data included 65 product characteristics, including 23 processor type dummies and 9 operating system type dummies. In order to reduce the dimension of the characteristics space, rather than use the 23 processor type dummies and the speed rating of the chip as separate characteristics, we instead obtained CPU benchmarks for each machine from *The CPU Scorecard* (www.cpuscorecard.com). Despite having considerable variation, a regression of the CPU benchmark variable on processor dummies interacted with speed of the chip had an R^2 of 0.995, justifying its use in their place.

Of the remaining 41 characteristic fields, we eliminated those fields that were either not reliable (not always recorded) or only applied to a handful of machines. However, despite the need to drop several of the characteristics fields, we are left with an extremely rich set of characteristics. The final characteristics set included nine operating system dummies (Win 3.11, Win 3.1, NT 3.51, NT3.2, NT 4.0, NT, Win 98, Win 95, Other) plus CPU benchmark, MMX, ram capacity, hard drive capacity, SCSI, CDROM, DVD, modem, modem speed, NIC, monitor dummy, monitor size, zip drive, desktop (versus tower), refurbished, dual hard drive, and dual processor, for a total of 26 characteristics.

Tables 2 and 3 contain summary statistics for the final data set. Table 2 shows that there are approximately 600 machines per month in the data, representing an average of approximately 300,000 units sold. The sales-weighted average price of machines drops by approximately 40% over the 29 month period. The unweighted average price is generally higher, but moves similarly. At the same time, Table 3 shows that sales-weighted average CPU benchmark, and sales-weighted average hard drive capacity all go up by approximately a factor of four, while sales-weighted average ram goes up by a approximately a

[22] In working with the raw data, we discovered two problems that we felt needed addressing. First, the data for machines with very few sales was highly variable from month to month. Second, sometimes machines are recorded as having been sold at very low prices (e.g., $0.01) when they were in fact taken off the books for other reasons, such as because the unit was stolen. Thus, in order to remove both of these problems, we dropped all price observations for units that sold fewer than 10 units in a given period. After dropping these observations, 3853 machines remained, from an original sample of approximately 8000.

factor of three. To summarize, prices for a constant quality machine are dropping rapidly but consumers are also rapidly substituting toward higher quality machines. The net result is that average purchase prices still drop by 50% over the 29 month period.

Table 3 shows that, despite the fact that our data only covers 29 months, there is considerable shift in the boundaries of the characteristics space over time. The shift in the minimum set of characteristics available is only slight. However, there is a considerable shift upward in the maximum characteristics available, particularly with respect to CPU benchmark and hard drive capacity. Table 3 also leaves out some shifts in the characteristics space with respect to the other product characteristics. For example, in our data, Windows NT 3.51 is unavailable after May 1998.

4.2 Price Index Calculations

4.2.1 Standard Indexes

Table 4 lists matched model price indexes calculated using the final data set. Even though our data is quite high frequency relative to that used by the BLS, the standard matched model indexes are quite unreliable here because there is so much attrition in the sample. The standard indexes suffer from both a selection bias, present even in the initial periods, and from considerable noise in later periods due to there being so few matched products (note the drop from the Nov 1999 to Dec 1999). In our opinion, table 4 shows how difficult it is to construct a matched model price index for a fast paced high technology industry like personal computers. Even in such a short span as two years and even using a very comprehensive data set covering nearly 4000 machines, it is nearly impossible to use the matched model method to construct a reliable price index. On the other hand, there are enough observations common to any two neighboring periods that the chained indexes do not suffer from the same sampling noise problem. However, with the chained indexes there is still a potential selection problem with respect to which products remain in the market from period to period.

Table 5 shows standard hedonic indexes (H) calculated using the same data set. In implementing the hedonic indexes it was necessary to choose a baseline functional form. In the spirit of nonparametric estimation, in choosing the baseline functional form our goal was to find the functional form that provided the best fit for the hedonic surface. We tried several functional forms, including linear, semi-log,[23] and log-log. An analysis of the residuals from these functional forms over several time periods revealed that log-log provided a very poor fit. The linear form fit the high end machines well but did not fit the low end machines well (it vastly underpredicted price). The semi-log fit the low end machines well, but showed some slight problems at the high end (slightly underpredicted price). We judged that the best of the three was the semi-log so we proceed using this as our baseline form.[24]

Coefficients in the hedonic regressions generally had the expected signs, the main exception being the "Modem" variable, which consistently was estimated to have a negative coefficient.[25] We speculate that this may be due to the fact that computers with modems are generally intended for home use and may be of lower average quality in other respects that are not observed.

The standard hedonic index and the matched model index are quite different over some ranges (e.g., from Aug 1997 - Sep 1997), which probably reflects the selection problem in the matched model indexes. However, their movement over the whole sample is surprisingly similar. This result is contrary to the results of Pakes (2003), in which it is found that the selection problem is so bad that the matched model indexes actually rise for some periods instead of falling. We do not know for sure why our results differ so much in this respect. However, we speculate that selection is not nearly as bad a problem in monthly

[23] In the semi-log form, only the left hand side (price) is in log form.

[24] Our judgement was based on a series of statistical tests as well as "eyeing" the fit via residual charts. These results are also consistent with the arguments of Diewert (2003), which argues that the left hand side variable in hedonic regressions should be in log form.

[25] Individual coefficients are not reported because there were 29 regressions with approximately 26 coefficients each, for a total of 754 coefficient estimates.

data as it is in yearly data. In our data it is typical for over 90% of the products in one month to be observed in the next, whereas typically fewer than 10% are observed 12 months later.[26] From a policy point of view, this evidence may suggest that it is worthwhile to use higher frequency data in industries that have a lot of product turnover. Note that our results also show slightly faster rates of decline than those of Aizcorbe, Corrado, and Doms (2003) for the period where the data overlaps.

We found that the standard (non-chained) hedonic indexes were subject to some variability with respect to changing the functional form of the hedonic price function. This variability arises because of the fact that the product space is changing over time. Because computers are improving over time, in calculating the price index for periods that are far apart in time it is typically necessary to extrapolate the hedonic price function outside the range of characteristics space on which it was estimated. We found that this introduced substantial variability into the index to the point where we are not confident in the results of the non-chained indexes even for the best of the functional forms. On the other hand, changing the functional form had very little effect on the chained indexes because very little extrapolation was needed between adjacent periods.

Despite the fact that our data contains many characteristics, we found that the R^2 statistics in the hedonic regressions ranged from 0.40 to 0.78. While these were lower than expected, they are in the same range as those found in Pakes (2003). They are lower than those of Holdway (2000). However, Holdway (2000) uses data obtained solely from large web-based retailers and is hence likely to be holding many unobserved factors constant. This result suggests that either there are still some important characteristics, such as sales outlet, or quality, that we do not observe, or that there is substantial measurement error in prices.

Table 6 shows standard hedonic indexes (H) similar to those above except that a polynomial series was used on the right hand side (as suggested by Pakes (2003)), but retaining

[26] However, the fact that 90% of the products are observed from one month to the next does not preclude selection being a bad problem.

the semi-log form. In general these indexes resulted in better in sample fit of the hedonic function, particularly for those periods in which fit was previously the poorest. A large number of the coefficients on the second order terms were also statistically significant. However, this improvement comes at some cost with respect to prediction near the boundaries of the sample in characteristics space. The result is that even with just a second order polynomial, there are some wild fluctuations in the standard price indexes (see the Paasche price index for July-December 1999). We found that going to higher order polynomials further improves the fit of the model, but makes the price index even wilder. Again, the problem was not as bad for the chained indexes, as can be seen in the table.

Because of the unreliability of the non-chained indexes here and above, below we only report results for the chained indexes. For similar reasons (sampling error), we use the second order polynomial indexes, rather than the third order ones, as our base case index.

4.2.2 The Multidimensional Case

Table 7 reports p-values for hypothesis tests based on the GMM objective function for various values of L. The first three columns are the baseline tests run for *iid* measurement error and three subsamples of approximately ten periods that roughly divide the data into three pieces. We chose to use ten period subsamples because ten periods provided enough degrees of freedom to run tests for values of L up to about five while maintaining sufficiently many observations in the balanced panel. Tests with smaller and larger subsamples generated similar results, as did tests using different subsamples of ten periods. Results from the three tests are slightly inconsistent, with the early periods requiring a higher dimensional unobservable, but generally imply that the true dimension L is greater than or equal to four.

Because the coverage of the data changes slightly over the panel and because the variance in computer prices generally falls over our sample, we were concerned that the measurement error variance might not be constant over time. Thus, we re-ran the hypothesis tests allowing for heteroskedasticity of unknown form in the measurement error. Relative to the baseline tests, in order to allow for heteroskedasticity of unknown form, we need only run the same estimation procedure but throwing out the moment conditions corresponding to the diagonal of the covariance matrix in (16). Allowing for heteroskedasticity of unknown form, all three tests suggest that we cannot reject that the true value is $L = 3$. If we allow L to be large, it is impossible to test whether the measurement error is heteroskedastic or not because the model can always match the data equally well by increasing L. However, in our opinion, given the comprehensiveness of the characteristics data, the result that $L = 3$ is more reasonable than those above. We also find further support for heteroskedastic measurement error below.

We also worried that the large sample tests may over-reject due to the fact that first stage estimates are used in place of the true error terms in the hypothesis tests. Since our data did not contain enough data points to allow joint estimation of the two stages, we instead simulated finite sample critical values using the coefficient estimates obtained below and the assumption that both the unobserved characteristics and the measurement error were normally distributed. The results of these tests showed that the finite sample critical values were indeed larger than the asymptotic critical values. However, as shown in Table 7, the hypothesis that $L = 1$ is still rejected in all cases. We therefore conclude that in this data set, $L \in \{2, 3\}$.

Table 8 shows chained price indexes constructed for the $L = 0$ and $L = 1$ cases.[27] Comparing the $L = 0$ and $L = 1$ cases, we find that correcting for unobserved characteristics substantially reduces the index, by 2.9% over the 29 month period for the Fisher index. The reason for this bias is primarily selection. We find that the unobserved characteristics

[27] The $L = 1$ case was estimated using the moments from all balanced panels of length 3, 4, 5, 9, and 10 periods. We found that the results were *extremely* stable over choices of which panel lengths to use (to within 0.1 in the overall index). The primary reason for choosing these period lengths was a trade off between efficiency and computational burden. The more period lengths we use, the more efficient the estimates, but at the expense of higher computational burden. We wanted to include several short panels because they generally have more data points, as well as several long panels so as to incorporate information from periods that are far apart in the data.

are substantially improving over time (their normalized mean moves steadily upward from approximately -0.3 to 0.2 over the period). The standard hedonic index ($L = 0$) absorbs the mean unobserved characteristic in each period into the intercept of the price function. Then, in predicting the prices of goods from previous periods that were not observed in later ones, it overpredicts their prices, raising the overall index. Most of this movement in the average unobserved characteristic takes place during the first 18 months in the data, which is also reflected in the estimated indexes.

We found that allowing for heteroskedastic measurement error did significantly affect the parameter estimates and, based on this, we feel that this is the correct specification. However, we found that it had very little impact on the overall price index. The primary effect was to increase the rate of price deflation in early periods, and slow the rate of price deflation midway through the sample.

Unfortunately, we found that we were unable to estimate the price index for $L > 1$ cases in our data because it was not possible to estimate a two-dimensional ξ very precisely for many products. We found that there was so much noise in the estimates of ξ that the price index calculations were also too noisy to be reliable. Thus, since we found that $L \in \{2, 3\}$ above, the $L = 1$ case has to be viewed as an approximation to the true index. In view of the comprehensive nature of our data, this finding leads us to be slightly pessimistic with respect to the ability to correct for multiple unobserved factors in other industries.

However, there are several factors that also make this data more difficult to work with than other industries. One is the high rate of product turnover, which leads to products often only being observed in a handful of periods. If products are observed more often, estimation of multiple factors becomes easier. Another is the fact that our data was not collected as carefully as that of the BLS. With less measurement error, estimation would also be easier.

4.2.3 Omitting An Important Characteristic

Table 9 tests our approach for the case when an important characteristic, the CPU benchmark, is known to be omitted. We compare the estimated price index when all of the observed product characteristics are included (as above), against those constructed using all of the observed characteristics except CPU benchmark. This experiment tests the ability of the approach to reduce the bias from unobserved product characteristics for a case in which the bias is quantifiable. Note that in our model leaving out CPU benchmark is equivalent to leaving out several characteristics since our base case model uses a second order polynomial in all of the continuous characteristics.

In the first two columns of Table 9 we compute standard chained Fisher indexes first including all characteristics, and then omitting CPU benchmark. As indicated above, the results show that a substantial bias occurs in this case. Over the entire period, the difference between the two indexes is approximately 9%.

In the third and fourth columns of the table, we report the same results after controlling for the unobserved characteristics. The two indexes do not agree entirely, but are much closer than the two standard indexes. Over the entire sample, the difference is now just 2.3%. In fact, the first, third, and fourth columns of the table are remarkably similar.

We also calculated the correlation between the estimated values of ξ and the left out CPU benchmark variable to see if the estimated ξ's reflected the left out characteristic. We found the correlation with CPU benchmark was 0.41. We take this as evidence that the procedure is working the way it is supposed to. Note that when CPU benchmark is omitted, the unobserved characteristics only pick up the residual correlation of CPU benchmark with prices once the effect of RAM and Hard Drive and the other characteristics is already accounted for. Thus, we view the 0.41 figure as being quite high.

In theory, with enough data, the procedure should provide the same results whether or not CPU is included, so what explains the fact that the results are not 100% consistent?

Part of the difference between the two is almost surely explainable by the fact that we were unable to estimate the index for higher values of L. The procedure is using a single dimensional unobservable to try to match the previous results ($L \in \{2, 3\}$), plus now CPU benchmark is left out, so we should expect that $L \geq 4$. Thus, the procedure is relying on an approximation. A second reason for the difference would be if the relationship between CPU benchmark and the other observed characteristics among the observed products is changing substantially over the sample period.

5 Conclusions

In conclusion, we have presented both theoretical and empirical evidence that omitted product characteristics can lead to a severe bias in hedonic price indexes. Moreover, we have shown that, at least for our data on desktop PC's, this bias is of practical relevance. In the case of PC's we have found evidence that there is a selection bias in the standard hedonic index that biases the index upward by about 1.4% per year in our sample.

Given the comprehensiveness of the characteristics data available for computers, we found it somewhat surprising that the unobserved characteristics bias was this large. In other industries where hedonic techniques are currently used by the BLS, such as housing and apparel, we might expect it to be more difficult to collect such comprehensive data and thus more likely that there are important unobserved characteristics. On the other hand, mitigating this effect is the fact that unobserved characteristics in these industries are likely to change less quickly over time, reducing the selection bias.

We have also presented an approach for constructing hedonic indexes that control for unobserved product characteristics under quite general assumptions. This approach can be viewed as a middle ground between the standard hedonic approach and the matched model approach. The drawback of our approach is that it requires more data than the standard hedonic approach because it requires data on the same products over several time periods. However, its data requirements fall far short of those of the matched model approach. Our methodology also shows how to do factor analysis more generally for unbalanced panels when selection is present.

6 References

Aizcorbe, A., Corrado, C. and M. Doms (2003), "Constructing Price and Quantity Indexes for High Technology Goods," Working Paper, Federal Reserve Board.

Bajari, P., and C. L. Benkard (2003), "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach", Working Paper, Stanford GSB.

Berndt, E. T., Griliches, Z. and N. R. Rappaport (1995), "Econometric Estimates of Price Indexes for Personal Computers in the 1990's," *Journal of Econometrics*, 68, 243-268.

Berry, S., J. Levinsohn, and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, **63:4**, 841-89.

Cockburn, I. and A. Anis (1998), "Hedonic Analysis of Arthritis Drugs," NBER Working Paper #W6574.

Court, A. T. (1939), "Hedonic Price Indexes with Automotive Examples," in General motors Corporation, *The Dynamics of Automobile Demand,* New York.

Cragg, J. G. and S. G. Donald (1995), "Factor Analysis Under More General Unknown Form," in G.S. Maddala, P.C.B. Phillips, and T.N. Srinivasan, eds., *Advances in Econometrics and Quantitative Economics*, Cambridge, Mass.: Basil Blackwell, 1995.

Cragg, J. G. and S. G. Donald (1997), "Inferring the Rank of a Matrix," *Journal of Econometrics*, Vol. 76, 223-251.

Diewert, E. (2003), "Hedonic Regressions: A Review of Some Unresolved Issues," presented at "International Conference in honor of Zvi Griliches: R&D, Education and Productivity", Paris, August 25-27, 2003.

Goldberger, A. S. (1974), "Unobservable Variables in Economics," in *Frontiers in Econometrics*, Zarembka ed., New York: Academic Press, 193-211.

Griliches, Z. (1961), "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," in *The Price Statistics of the Federal Government*, New York: National Bureau of Economic Research.

Griliches, Z. and M. Ohta (1986), "Automobile Prices and Quality: Did the Gasoline Price Increase Change Consumer Tastes in the U.S.?" *Journal of Business and Economic Statistics,* 4:2, l87-198.

Hansen, L. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, **50**, 1029-1054.

Holdway, M. (2001), "Quality-Adjusting Computer Prices in the Producer Price Index: An Overview," Bureau of Labor Statistics.

Imbens, G. and W. K. Newey (2001), "Identification and Estimation of Triangular Simultaneous Equations Models Without Additivity," Working Paper.

Konus, A. (1924), "The Problem of the True Cost of Living Index," translated in *Econometrica*, 7, (1939), 10-29.

Landefeld, J. S. and B. T. Grimm (2000), "A Note on the Impact of Hedonics and Computers on Real GDP," *Survey of Current Business*, 80, 17-22.

Lawley, D. N., and A. E. Maxwell (1971), *Factor Analysis as a Statistical Method,* New York: American Elsevier Publishing.

Matzkin, Rosa (2003), "Nonparametric Estimation of Nonadditive Random Functions," *Econometrica*, forthcoming.

Moulton, B. R. (2001), "The Expanding Role of Hedonic Methods in the Official Statistics of the United States," Mimeo, U.S. Bureau of Economic Analysis.

Nevo, A. (2001), "New Products, Quality Changes, and Welfare Measures Computed from Estimated Demand Systems," NBER Working Paper #W8425.

Pakes, A. (2003), "A Reconsideration of Hedonic Price Indices With an Application to PC's", *American Economic Review, forthcoming.*

Triplett, J. E. (1969), "Automobiles and Hedonic Quality Measurement," Journal of Political Economy, 77:3, 408-417.

7 Tables and Graphs

Table 1: Left Out Characteristics Bias

	Chained Fisher Indexes	
Period	All Chars. Included	CPU Benchmark Omitted
Aug '97	100.0	100.0
Sep '97	92.3	93.4
Oct '97	85.3	86.8
Nov '97	81.0	83.9
Dec '97	76.0	78.8
Jan '98	66.9	68.9
Feb '98	65.1	68.4
Mar '98	62.0	65.8
Apr '98	59.3	64.3
May '98	54.3	59.1
Jun '98	51.4	56.4
Jul '98	47.8	53.0
Aug '98	43.5	49.5
Sep '98	41.0	47.7
Oct '98	37.4	43.7
Nov '98	35.5	42.3
Dec '98	32.0	38.5
Jan '99	29.7	37.0
Feb '99	28.8	37.1
Mar '99	27.9	37.5
Apr '99	27.2	37.3
May '99	25.4	34.9
Jun '99	22.5	30.4
Jul '99	21.5	30.7
Aug '99	20.1	29.5
Sep '99	18.2	27.6
Oct '99	17.4	27.4
Nov '99	16.6	26.5
Dec '99	15.7	25.1

Table 2: Summary of Computer Data

Period	Unique Machines	Total Sales	Avg. Price (Unweighted)	Avg. Price (Sales-Weighted)
Aug '97	577	226029	1396	1422
Sep '97	556	239417	1408	1437
Oct '97	562	211610	1411	1423
Nov '97	517	265070	1358	1351
Dec '97	524	345153	1308	1321
Jan '98	572	328028	1224	1200
Feb '98	525	331262	1172	1217
Mar '98	614	371337	1187	1194
Apr '98	601	260173	1206	1179
May '98	547	210834	1182	1134
Jun '98	660	278002	1160	1111
Jul '98	563	250110	1156	1133
Aug '98	615	345183	1177	1092
Sep '98	649	393909	1131	1113
Oct '98	647	296737	1128	1032
Nov '98	563	428776	1046	1099
Dec '98	644	592138	1042	995
Jan '99	593	406644	981	1028
Feb '99	569	371586	998	1056
Mar '99	675	452156	1025	1046
Apr '99	635	313716	977	1061
May '99	608	285353	968	1033
Jun '99	692	378476	947	1002
Jul '99	614	330798	878	1020
Aug '99	616	478200	841	992
Sep '99	672	571820	848	953
Oct '99	710	379487	866	914
Nov '99	661	484269	861	925
Dec '99	747	664983	912	879

Table 3: Summary of Product Characteristics

	CPU Benchmark				RAM				Hard Drive Capacity			
Period	Avg*	S.D.*	Min	Max	Avg*	S.D.*	Min	Max	Avg*	S.D.*	Min	Max
Aug '97	333	178	17	781	24	12	4	128	2340	1147	420	7000
Sep '97	343	194	17	855	25	14	4	128	2509	1219	420	7000
Oct '97	383	203	17	855	27	14	4	128	2733	1292	420	7000
Nov '97	400	216	17	982	26	12	4	128	2817	1355	420	7000
Dec '97	422	213	17	982	27	12	8	128	2910	1382	420	8000
Jan '98	428	219	17	982	27	12	4	128	2998	1538	250	12000
Feb '98	472	222	17	982	30	14	4	128	3174	1587	420	12000
Mar '98	501	226	17	1130	31	15	8	128	3302	1718	420	12000
Apr '98	532	236	17	1131	32	15	4	128	3474	1826	80	12000
May '98	572	237	17	1131	33	15	8	128	3665	1832	500	12000
Jun '98	599	251	17	1131	36	18	8	128	3900	2004	420	12000
Jul '98	661	252	17	1131	39	19	8	128	4239	2065	800	12000
Aug '98	700	248	17	1344	41	21	8	128	4464	2148	500	12000
Sep '98	730	256	17	1240	44	23	8	128	4697	2364	420	16800
Oct '98	743	271	17	1240	45	24	8	128	4794	2407	540	16800
Nov '98	802	261	17	1240	49	27	8	256	5127	2481	850	16800
Dec '98	806	265	17	1270	51	30	8	256	5292	2644	250	18000
Jan '99	843	264	17	1468	53	30	8	256	5490	2696	250	19000
Feb '99	899	249	17	1651	57	32	8	256	5919	2982	250	19000
Mar '99	929	285	17	1651	57	30	8	256	6058	3105	250	20000
Apr '99	991	275	17	1651	60	29	8	256	6449	3307	250	32000
May '99	1049	276	17	1780	63	30	16	256	6925	3337	250	20400
Jun '99	1080	303	17	1814	64	31	4	256	7221	3651	340	32000
Jul '99	1151	287	17	1930	68	37	16	512	7608	3789	500	32000
Aug '99	1183	299	17	2254	69	34	16	256	7765	3829	500	27000
Sep '99	1237	328	17	2347	72	35	16	256	8202	4197	500	27000
Oct '99	1278	343	17	2399	72	33	16	256	8545	4362	64	27000
Nov '99	1329	343	17	2510	73	33	16	256	9027	4556	64	36500
Dec '99	1339	381	17	2544	73	35	8	256	9167	4905	64	40000

*Averages and standard deviations are sales weighted.

Table 4: Matched Model Indexes

Period	N(L)	Laspeyre's	Paasche	Fisher	N	Chained Laspeyre's	Chained Paasche	Chained Fisher
Aug '97	NA	100.0	100.0	100.0	NA	100.0	100.0	100.0
Sep '97	425	96.0	96.2	96.1	425	96.0	96.2	96.1
Oct '97	353	91.2	91.7	91.5	405	91.9	92.6	92.3
Nov '97	294	82.5	81.1	81.8	412	85.5	84.1	84.8
Dec '97	266	77.6	76.6	77.1	400	81.3	79.5	80.4
Jan '98	253	70.3	72.0	71.2	405	75.8	73.8	74.8
Feb '98	206	67.3	67.8	67.5	416	69.7	69.7	69.7
Mar '98	198	60.0	59.8	59.9	431	67.0	66.0	66.5
Apr '98	172	59.0	63.2	61.0	473	62.6	61.9	62.2
May '98	122	56.8	62.5	59.6	427	58.8	58.6	58.7
Jun '98	147	50.7	56.0	53.2	439	53.8	52.6	53.2
Jul '98	88	49.7	53.6	51.6	429	49.3	48.8	49.1
Aug '98	83	50.6	46.8	48.7	443	46.1	46.0	46.1
Sep '98	91	65.1	43.7	53.3	453	43.7	43.3	43.5
Oct '98	103	44.9	44.9	44.9	478	39.7	40.4	40.1
Nov '98	49	57.3	68.9	62.8	429	37.1	37.6	37.4
Dec '98	73	36.4	41.3	38.8	459	35.5	35.8	35.7
Jan '99	45	53.3	45.0	49.0	447	31.8	33.3	32.5
Feb '99	27	77.9	59.7	68.2	419	30.4	32.3	31.3
Mar '99	37	87.9	98.2	92.9	449	29.0	30.6	29.8
Apr '99	12	86.9	105.	95.7	471	27.7	29.1	28.4
May '99	4	16.3	33.5	23.4	426	26.6	28.0	27.3
Jun '99	13	14.4	47.7	26.2	459	24.9	26.2	25.6
Jul '99	1	65.5	65.5	65.5	468	23.0	24.6	23.8
Aug '99	2	23.9	47.9	33.8	459	21.8	22.9	22.4
Sep '99	1	65.4	65.4	65.4	473	20.8	21.6	21.2
Oct '99	2	62.3	61.2	61.7	499	19.2	20.6	19.9
Nov '99	2	58.3	58.9	58.6	509	18.1	19.8	18.9
Dec '99	5	13.1	20.1	16.3	532	17.7	19.5	18.5

N(L) is number of units used to construct Laspeyre's index.
N is number of units used to construct Chained indexes.

Table 5: Standard Hedonic Indexes (H)

Period	N(L)	Laspeyre's	Paasche	Fisher	N	Chained Laspeyre's	Chained Paasche	Chained Fisher
Aug '97	NA	100.0	100.0	100.0	NA	100.0	100.0	100.0
Sep '97	550	92.9	92.5	92.7	541	92.9	92.5	92.7
Oct '97	550	86.0	86.4	86.2	545	86.5	86.8	86.6
Nov '97	550	84.4	82.5	83.4	502	83.9	83.2	83.5
Dec '97	550	79.3	78.1	78.7	518	78.3	77.9	78.1
Jan '98	550	67.7	69.0	68.3	561	68.1	67.3	67.7
Feb '98	550	70.3	67.6	69.0	517	67.7	67.0	67.3
Mar '98	550	68.4	66.2	67.3	594	65.0	64.4	64.7
Apr '98	550	66.5	63.5	65.0	583	63.4	62.9	63.1
May '98	550	59.3	59.5	59.4	533	57.7	57.4	57.5
Jun '98	550	59.1	52.8	55.9	642	53.4	53.4	53.4
Jul '98	550	56.1	45.2	50.4	547	49.6	50.6	50.1
Aug '98	550	51.3	40.3	45.5	600	45.4	45.9	45.6
Sep '98	550	57.7	39.7	47.9	635	43.5	43.9	43.7
Oct '98	550	47.7	35.4	41.1	635	39.2	40.6	39.9
Nov '98	550	58.3	32.6	43.5	550	38.1	38.5	38.3
Dec '98	550	43.9	28.5	35.4	624	34.6	34.8	34.7
Jan '99	550	49.1	30.5	38.7	579	32.0	32.5	32.3
Feb '99	550	53.7	31.1	40.8	556	31.7	32.4	32.0
Mar '99	550	52.2	30.1	39.6	659	30.9	31.9	31.4
Apr '99	550	47.9	32.1	39.2	622	29.8	30.3	30.0
May '99	550	32.0	31.6	31.8	593	27.2	27.7	27.5
Jun '99	550	24.8	27.6	26.1	663	23.9	24.4	24.1
Jul '99	550	29.3	26.1	27.7	588	22.9	23.3	23.1
Aug '99	550	30.1	21.9	25.7	588	21.3	21.5	21.4
Sep '99	550	27.8	19.5	23.3	646	19.4	19.4	19.4
Oct '99	550	25.4	18.9	21.9	677	18.9	18.7	18.8
Nov '99	550	26.2	17.5	21.4	626	17.9	17.7	17.8
Dec '99	550	16.5	16.4	16.4	705	16.8	16.7	16.7

Functional form is semi-log. R^2 ranges from 0.40-0.78.
N(L) is number of units used to construct Laspeyre's index.
N(CL) is number of units used for Chained Laspeyre's index.

Table 6: Nonparametric Hedonic Indexes (H)

Order:		2	2		2	2	2	3	3	3
Period	N(L)	L	P	N(CL)	CL	CP	CF	CL	CP	CF
Aug '97	NA	100.0	100.0	NA	100.0	100.0	100.0	100.0	100.0	100.0
Sep '97	550	92.4	92.2	541	92.4	92.2	92.3	92.6	92.7	92.6
Oct '97	550	84.5	85.4	545	85.1	85.5	85.3	85.7	86.4	86.0
Nov '97	550	82.1	80.6	502	81.6	80.4	81.0	81.5	80.9	81.2
Dec '97	550	77.6	75.8	518	76.7	75.4	76.0	76.6	76.2	76.4
Jan '98	550	64.6	66.9	561	67.7	66.2	66.9	67.7	68.2	67.9
Feb '98	550	66.8	64.3	517	66.0	64.3	65.1	65.5	65.8	65.6
Mar '98	550	64.6	62.3	594	62.9	61.2	62.0	62.0	62.7	62.4
Apr '98	550	63.5	59.2	583	60.1	58.5	59.3	59.5	60.0	59.8
May '98	550	56.4	54.6	533	54.9	53.6	54.3	54.4	54.9	54.7
Jun '98	550	56.0	50.1	642	51.6	51.3	51.4	51.2	52.3	51.8
Jul '98	550	54.7	45.5	547	47.4	48.2	47.8	47.1	49.6	48.3
Aug '98	550	50.3	40.0	600	43.5	43.5	43.5	43.2	44.5	43.8
Sep '98	550	59.2	38.8	635	40.9	41.0	41.0	40.6	41.9	41.3
Oct '98	550	47.4	41.5	635	36.9	38.0	37.4	36.5	38.3	37.4
Nov '98	550	60.5	38.9	550	35.4	35.6	35.5	35.3	39.7	37.4
Dec '98	550	46.8	38.5	624	31.9	32.0	32.0	32.7	36.9	34.7
Jan '99	550	53.9	35.7	579	29.7	29.7	29.7	30.6	34.2	32.4
Feb '99	550	63.2	36.1	556	28.9	28.7	28.8	29.7	33.1	31.4
Mar '99	550	59.4	35.9	659	28.0	27.7	27.9	28.8	32.1	30.4
Apr '99	550	54.2	36.0	622	27.6	26.8	27.2	28.4	31.2	29.8
May '99	550	31.2	32.0	593	25.8	25.0	25.4	26.2	28.9	27.5
Jun '99	550	20.6	31.2	663	22.8	22.1	22.5	22.9	25.2	24.0
Jul '99	550	23.8	13612488	588	21.9	21.0	21.5	22.1	24.1	23.1
Aug '99	550	23.7	48.9	588	20.6	19.6	20.1	20.8	22.4	21.6
Sep '99	550	22.1	523.9	646	18.7	17.8	18.2	18.8	20.3	19.6
Oct '99	550	24.8	2153.1	677	18.0	16.9	17.4	18.0	19.4	18.7
Nov '99	550	25.0	2115.6	626	17.2	16.1	16.6	17.1	18.3	17.7
Dec '99	550	13.4	5073.8	705	16.2	15.3	15.7	16.2	17.6	16.8

Functional form is semi-log with polynomial series. R^2 ranges from 0.50-0.79.
N(L) is number of units used to construct Laspeyre's index.
N(CL) is number of units used for Chained Laspeyre's index.
Order is polynomial order.

Table 7: P-Values For Dimensionality Tests

	Homoskedastic M.E.			Heteroskedastic M.E.		
Subsample (Months):	1-10	11-20	21-29	1-10	11-20	21-29
Large Sample Tests:						
Dimension (L)						
0	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.000	0.000
2	0.000	0.000	0.000	0.000	0.000	0.046
3	0.000	0.001	0.006	0.147	0.913	0.958
4	0.001	0.244	0.749	0.757	0.963	—
5	0.025	0.248	—	—	—	—
Small Sample Tests:						
Dimension (L)						
0	0.000	0.000	0.000	0.000	0.000	0.000
1	0.000	0.000	0.000	0.000	0.003	0.000
Number of Obs:	93	58	137	93	58	137
	"—": too few degrees of freedom to calculate.					

Table 8: Chained Price Indexes for $L = 0$ and $L = 1$

				Homosk. M.E.			Heterosk. M.E.		
Dimension	$L = 0$			$L = 1$			$L = 1$		
Period	CL	CP	CF	CL	CP	CF	CL	CP	CF
Aug '97	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Sep '97	92.4	92.2	92.3	93.9	93.8	93.8	93.1	93.1	93.8
Oct '97	85.1	85.5	85.3	86.7	87.0	86.9	85.8	86.3	86.9
Nov '97	81.6	80.4	81.0	82.1	81.0	81.5	81.1	80.3	81.4
Dec '97	76.7	75.4	76.0	76.8	75.6	76.2	75.9	74.9	76.1
Jan '98	67.7	66.2	66.9	70.5	69.6	70.0	70.2	69.5	70.0
Feb '98	66.0	64.3	65.1	66.2	65.2	65.7	65.2	64.4	65.6
Mar '98	62.9	61.2	62.0	61.7	61.0	61.3	60.9	60.4	61.2
Apr '98	60.1	58.5	59.3	58.4	57.8	58.1	57.6	57.2	58.0
May '98	54.9	53.6	54.3	53.8	53.4	53.6	53.2	53.0	53.7
Jun '98	51.6	51.3	51.4	49.4	50.2	49.8	48.9	49.9	49.8
Jul '98	47.4	48.2	47.8	45.0	46.9	46.0	44.5	46.5	45.9
Aug '98	43.5	43.5	43.5	41.3	42.4	41.8	40.9	42.2	41.8
Sep '98	40.9	41.0	41.0	37.6	38.6	38.1	37.0	38.2	38.0
Oct '98	36.9	38.0	37.4	33.5	35.4	34.5	33.5	35.5	34.4
Nov '98	35.4	35.6	35.5	30.9	31.7	31.3	30.5	31.4	31.3
Dec '98	31.9	32.0	32.0	28.2	29.2	28.7	28.2	29.1	28.7
Jan '99	29.7	29.7	29.7	24.9	25.9	25.4	25.1	25.9	25.6
Feb '99	28.9	28.7	28.8	23.2	24.1	23.7	23.2	24.1	23.9
Mar '99	28.0	27.7	27.9	22.4	23.5	22.9	22.4	23.6	23.1
Apr '99	27.6	26.8	27.2	21.8	22.5	22.2	21.8	22.4	22.4
May '99	25.8	25.0	25.4	20.1	20.8	20.5	20.0	20.6	20.6
Jun '99	22.8	22.1	22.5	18.1	18.6	18.4	18.1	18.6	18.6
Jul '99	21.9	21.0	21.5	17.1	17.5	17.3	17.0	17.4	17.5
Aug '99	20.6	19.6	20.1	16.0	16.3	16.2	16.0	16.2	16.3
Sep '99	18.7	17.8	18.2	14.6	14.8	14.7	14.6	14.7	14.8
Oct '99	18.0	16.9	17.4	13.8	14.0	13.9	13.8	13.9	14.0
Nov '99	17.2	16.1	16.6	13.0	13.1	13.1	13.0	13.1	13.2
Dec '99	16.2	15.3	15.7	12.8	12.9	12.8	12.8	12.9	12.9

Table 9: Chained Price Indexes: Multidimensional Cases with and without CPU Benchmark

	Chained Fisher Indexes			
	Standard Indexes		Corrected Indexes	
Period	All Chars.	CPU Omit.	All Chars.	CPU Omit.
Aug '97	100.0	100.0	100.0	100.0
Sep '97	92.3	93.4	93.8	93.6
Oct '97	85.3	86.8	86.9	86.8
Nov '97	81.0	83.9	81.4	82.8
Dec '97	76.0	78.8	76.1	77.2
Jan '98	66.9	68.9	70.0	69.6
Feb '98	65.1	68.4	65.6	67.8
Mar '98	62.0	65.8	61.2	64.3
Apr '98	59.3	64.3	58.0	61.4
May '98	54.3	59.1	53.7	57.2
Jun '98	51.4	56.4	49.8	53.0
Jul '98	47.8	53.0	45.9	48.9
Aug '98	43.5	49.5	41.8	45.7
Sep '98	41.0	47.7	38.0	41.8
Oct '98	37.4	43.7	34.4	38.3
Nov '98	35.5	42.3	31.3	34.8
Dec '98	32.0	38.5	28.7	32.2
Jan '99	29.7	37.0	25.6	29.2
Feb '99	28.8	37.1	23.9	27.9
Mar '99	27.9	37.5	23.1	27.7
Apr '99	27.2	37.3	22.4	26.4
May '99	25.4	34.9	20.6	23.8
Jun '99	22.5	30.4	18.6	21.3
Jul '99	21.5	30.7	17.5	20.3
Aug '99	20.1	29.5	16.3	19.1
Sep '99	18.2	27.6	14.8	17.3
Oct '99	17.4	27.4	14.0	16.5
Nov '99	16.6	26.5	13.2	15.6
Dec '99	15.7	25.1	12.9	15.2
Average N	573	573	573	573
Corrected indexes robust to heteroskedastic measurement error.				