



AUSTRALIAN
SCHOOL OF BUSINESS™

THE UNIVERSITY OF NEW SOUTH WALES

The University of New South Wales Australian School of Business

Australian School of Business Research Paper No. 2010 ECON 03

Consistency of Hedonic Price Indexes with Unobserved Characteristics

Iqbal Syed

This paper can be downloaded without charge from
The Social Science Research Network Electronic Paper Collection:
<http://ssrn.com/abstract=1550108>

Consistency of Hedonic Price Indexes with Unobserved Characteristics

Iqbal Syed

*School of Economics, University of New South Wales,
Sydney, NSW 2052, Australia.
(email: i.syed@unsw.edu.au)*

Abstract

Hedonic regressions are prone to omitted variable bias because many of the price determining characteristics are typically unobserved. The estimation of price relatives for new and disappearing goods using hedonic imputation methods involves taking ratios of two hedonic models corresponding to two consecutive periods. This may lead to a situation where the omitted variable bias in one of the hedonic regressions offsets the other. This study finds that the single imputation hedonic method estimates inconsistent price relatives, while the double imputation method may produce consistent price relatives depending on the behavior of unobserved characteristics in the comparison periods. The study outlines a methodology to estimate the magnitude of bias in single and double imputation price relatives. The results of this study have implications with regard to the construction of quality adjusted indexes.

JEL Classification Numbers: C43; C52; E31

Keywords: Hedonic imputation method; Model selection; New and disappearing goods; Omitted variable bias; Quality adjusted price indexes.

1. Introduction

An hedonic regression model specifies the price of a variety v at time t , $P_{v,t}$, as a function of its determining characteristics and a random error term. The most common use of hedonic regressions has been to disentangle the quality component of price change from the observed price change to achieve better measures of ‘pure’ inflation. Hedonic regressions have been used for quality adjustment in price changes since Waugh (1928) and Court (1939), with Griliches (1961) reviving interest in the methodology. Boskin et al. (1996) consider the hedonic regression to be the most promising approach to control for quality changes, whereas Schultze and Mackie (2002) recommend following a cautionary approach and emphasize the need for further research. Moulton (2001) reports that approximately 18% of the US GDP final expenditures are deflated using price indexes that use hedonic methods and this share is expected to rise.

Price comparison of ‘like with like’ is an essential requirement in the construction of price indexes. This requirement raises problems in many markets, including super-market product, electronic and housing markets. For example, electronic products, such as computers, are subject to rapid technological innovation, leading to quick product turnovers and short life cycles. $P_{v,t}$ is observed if the model v was sold in period t , but $P_{v,t+1}$ is unobserved if the model exited out of the market before the beginning of period $t + 1$. In this case, the price relative $P_{v,t+1}/P_{v,t}$ is unobserved. In housing markets, every house is somewhat different and the same house is unlikely to be sold in two adjacent periods. This implies that if v refers to a particular house, either $P_{v,t}$ or $P_{v,t+1}$ is unobserved. Hedonic regressions impute these unobserved prices, and let us estimate the price relatives of unmatched, i.e. new and disappearing, items.

Price indexes calculated from only the matched varieties, i.e. the varieties that are sold in both t and $t + 1$, and leaving out the new and disappearing varieties, suffer from sample selection bias (Boskin et al., 1996; Pakes, 2003; Benkard and Bajari, 2005; Silver and Heravi, 2005). For example, Silver and Heravi, using comprehensive scanner data of five products (washing machines, dishwashers, television sets, cameras and vacuum

cleaners), show that the sample degradation is substantial even in a short period of one year. The percentage of models of these five products that disappeared within one year ranged between 29% and 47%, and the decline in the sales value ranged between 5% and 18%. Pakes (2003) and Benkard and Bajari (2005) find the sample degradation to be more severe for personal computers—by 85% and 90%, respectively, in a year—leaving ample room for sample selection bias.

It is a common concern that hedonic regressions, largely because of the unavailability of data, omit relevant characteristics. Some of these characteristics may be correlated with the included characteristics, leading to biased and inconsistent estimates.¹ Over the years, many authors have explicitly or implicitly acknowledged the omitted variable bias problem in hedonic regressions, including Court (1939), Griliches (1961), Pakes (2003), Hulten (2003), Benkard and Bajari (2005), Triplett (2006), and Hill and Melser (2008). However, there has been little investigation of the omitted variable problem specific to hedonic regressions, and particularly to hedonic imputation methods, beyond a general understanding of the omitted variable problem in regression analysis.²

The study is based on the premise that omitted variable bias is a fact in hedonic regressions. The study focusses on hedonic imputation methods, where separate hedonic regressions are hypothesized for each period of price comparison. Typically, and in this study, prices are compared between two adjacent periods. Between two such periods, the behavior of some characteristics may remain stable, while the behavior of other characteristics may be unstable. These two sets of characteristics, stable and unstable, may be treated differently in the omitted variable analysis of hedonic imputation methods. The conjecture of this study is that it is the unstable characteristics that drive price change, and therefore they should be included in the model in order to attain consistency of the estimated price relatives. The stable characteristics, though

¹The minimal requirement of an estimator is ‘consistency’. For practical purposes, inconsistency can be viewed as being the same as bias.

²There may be other sources of bias including incorrect specification of the functional forms and measurement errors in prices. This paper addresses only the omitted variable bias problem.

important in each period, do not drive price change and therefore may be excluded, as far as the consistency of the estimated price relatives is concerned. That is, under certain stability conditions the omitted variable bias corresponding to each regression may be canceled out. This study provides detailed expression of such stability conditions for, and evaluates the performance of, both the level and log hedonic models in the single and double imputation methods.

This study analyzes the omitted variable bias generated by each hedonic regression and then extends the analysis to hedonic imputation methods. Let us suppose that the conjecture of this paper is correct, i.e. the relevant characteristics that are stable between the periods can be omitted without having any effect on the consistency of the estimated price relatives. This may have an important implication with regard to the application of hedonic methods in order to construct quality adjusted indexes. For example, in the used car market there are many characteristics that are relevant to consumers in each period, including the make and model, age of car, engine size, odometer reading, color and fuel efficiency. However, there may be only a few characteristics that drive price change across two consecutive periods (such as fuel efficiency during a period of volatile oil prices), while the implicit value of other characteristics remains stable. This implies that data compilers may focus on a few select characteristics and, as a result, reduce cost.³

The importance of obtaining consistent price relatives in order to calculate indexes should be emphasized. Price relatives, including estimated price relatives, are the building blocks for the construction of indexes. It can be shown that elementary indexes (such as the Carli, Dutot and Jevons indexes) constructed on consistent price relatives are themselves consistent. Let us suppose that \hat{p}_v are consistent price relatives of their corresponding parameters θ_v , $\forall v = 1, \dots, V$. The estimated Jevons index, comparing prices between the period t and $t + 1$, is the geometric mean of all the price relatives \hat{p}_v : $\widehat{P}_{t,t+1} = \prod_{v=1}^V (\hat{p}_v)^{1/V}$. Using a property of probability limits (*plim*)—that if $g(\cdot)$ is

³Benkard and Bajari (2005) study the omitted variable bias problem in hedonic imputation methods. The focus of their paper, different to this paper, is on outlining an estimation methodology using factor analysis.

a continuous function, then $plim[g(x)] = g[plim(x)]$ —the following is obtained:

$$plim\left(\widehat{P}_{t,t+1}\right) = plim\left(\prod_{v=1}^V (\hat{p}_v)^{1/V}\right) = \prod_{v=1}^V [plim(\hat{p}_v)]^{1/V} = \prod_{v=1}^V (\theta_v)^{1/V}$$

that is, the estimated Jevons index is a consistent estimate of the true Jevons index. However, if the price relatives are inconsistent, and if there is no justification that biases tend to cancel each other out, the elementary indexes, and any other price indexes, including superlative indexes such as the Fisher and Törnqvist indexes, which are weighted averages of individual price relatives, are also inconsistent.

This paper is organized as follows. The next section introduces hedonic imputation methods and presents some observations on recent literature that compares and contrasts different hedonic methods. In section 3, the hedonic models are specified for the log of prices and the coefficient stability condition required to attain consistency of estimated price relatives is derived (a concise derivation for hedonic models specified on price levels is provided in appendix A). Section 4 shows results of Monte Carlo simulations which are conducted in order to check whether the stability condition derived analytically for large samples also holds for small samples. Section 5 discusses the implications of the findings with regard to their application. Conclusions are drawn in section 6.

2. Time Dummy, Single and Double Imputation Hedonic Methods

The dominant hedonic regression methods are the time-dummy hedonic method and the hedonic imputation method. In the time-dummy method, prices are hypothesized as a function of time dummies and the characteristics variables. The characteristics variables play the role of ‘controlling variables’ in the regression, and the coefficients of time dummies reflect the average price change between the periods after holding

the characteristics constant. If a relevant characteristic is omitted that is correlated with any of the included characteristics or the time dummies, then the estimated time-dummy coefficients are biased and inconsistent.

In the hedonic imputation method, separate regressions are specified for each period. The method essentially involves predictions of the left hand side variable (price or log-price) of the hedonic regressions. The varieties of goods sold in, say, period 0 are predicted for period 1 by putting the estimated characteristics coefficients of period 1 into the value of the characteristics of period 0. Thus, if $P_{v,t+1}$ refers to a new variety v that appeared in the market at period $t + 1$, then the hedonic imputation method imputes $P_{v,t}$, i.e. the price of variety v before it appeared in the market. Let us call this imputed price $\hat{P}_{v,t}$. Similarly, $\hat{P}_{v,t+1}$ is estimated from a hedonic regression when variety v disappeared from the market at period t . Once all the varieties are ‘matched’ in this way, standard price index formulas can be used.

The hedonic imputation method offers two approaches—single imputation or double imputation approaches. In the single imputation approach, only the unobserved price is imputed. The single imputation price relative is $P_{v,t+1}/\hat{P}_{v,t}$ or $\hat{P}_{v,t+1}/P_{v,t}$, depending on whether item v is a new or disappearing item. In the double imputation approach, on the other hand, both the observed and unobserved prices are imputed, and the estimated price relative is $\hat{P}_{v,t+1}/\hat{P}_{v,t}$. The imputed prices such as $\hat{P}_{v,t}$ and $\hat{P}_{v,t+1}$ are biased and inconsistent due to omission of characteristics. But the question is whether the price relatives $P_{v,t+1}/\hat{P}_{v,t}$, $\hat{P}_{v,t+1}/P_{v,t}$ or $\hat{P}_{v,t+1}/\hat{P}_{v,t}$ are also biased and inconsistent or can they be unbiased and consistent under particular conditions?

In the existing literature, various authors have discussed which of the single or double imputation methods is more appropriate. Triplett (2006) favors the minimum use of imputation on the grounds of minimizing estimation variance unless there is reason to believe that omitted characteristics have not changed between the comparison periods. Hill and Melser (2008) suggest that the double imputation method is preferable because of the potential of omitted variable biases being canceled out, however, they do not provide any formal analysis in support of their argument. Pakes (2003) does

not find any difference between the calculated indexes for personal computers obtained from the single and double imputation methods and, therefore, chooses to report the indexes using the single imputation price relatives. On the other hand, van Mulligen (2003) finds a systematic difference in indexes for personal computers, notebooks and servers—the double imputation index lies in between the single imputation index and the matched model index.

The current study, conducting a formal analysis on the estimators of imputation methods, finds that the single imputation method, similar to the time-dummy method, produces inconsistent estimates, whereas the double imputation method may produce consistent estimates under some stability conditions related to omitted characteristics and depending on whether the hedonic models are estimated for the log or level of prices. Consistency may be achieved even when each of the hedonic regressions produces inconsistent estimates, and at the same time by being less demanding on data. Because of the high prevalence of matched items between two adjacent periods in most markets, the study shows that it is possible to evaluate the performance of alternative models in terms of the magnitude of omitted variable bias.

3. Hedonic Models of Log Prices

3.1 Estimation of single and double imputation price relatives from the correctly specified models

This section begins with setting the parameters of interest and estimating the single and double imputation price relatives from the correctly specified log hedonic models corresponding to periods 0 and 1. The following two equations specify the hedonic

regressions for periods 0 and 1, respectively:

$$\ln p_i^0 = \sum_{k=1}^K \beta_k z_{k,i}^0 + \epsilon_i^0 \quad \forall i = 1, \dots, I \quad (1)$$

$$\ln p_v^1 = \sum_{k=1}^K \delta_k z_{k,v}^1 + \epsilon_v^1 \quad \forall v = 1, \dots, V \quad (2)$$

In equation 1, $\ln p_i^0$ denotes the log of price of item i in period 0, $z_{k,i}^0$ refers to the value of characteristic k for item i in period 0, and ϵ_i^0 is the error term assumed to be *i.i.d.* with zero mean and constant variance. In equation 2, the notations have similar interpretations for item v in period 1. The first characteristics in both the equations, z_1^0 and z_1^1 , take the value of 1 for all observations, indicating that they refer to the intercept terms in the equations. This way of denoting the intercept terms eases the use of notations in later sections. The parameters of interest for item i in period 0 and item v in period 1 are the following, respectively (*plim* refers to probability limit):

$$plim (p_i^0 | z_1^0, z_2^0, \dots, z_k^0) = exp \left(\sum_{k=1}^K \beta_k z_{k,i}^0 \right) = exp (\theta_i^0) \quad (3)$$

$$plim (p_v^1 | z_1^1, z_2^1, \dots, z_k^1) = exp \left(\sum_{k=1}^K \delta_k z_{k,v}^1 \right) = exp (\theta_v^1) \quad (4)$$

Now let us suppose that item i and item v are the same item. Then, from equations 3 and 4, the following is obtained:

$$plim \left(\frac{p_i^1}{p_i^0} \right) = \frac{exp(\theta_i^1)}{exp(\theta_i^0)} = exp(\Theta_i) \quad (5)$$

This sets the parameter of interest for the price relative of item i to be $exp(\Theta_i)$. If any other alternative price relatives of item i converge to $exp(\Theta_i)$, then these price relatives are consistent; otherwise they are asymptotically biased.

Let us now suppose that equations 1 and 2 satisfy the classical linear regression model assumptions, including that each of z_k^t is uncorrelated with disturbance ϵ^t for

$t=0,1$. The ordinary least squares method (OLS) provides consistent estimates of the parameters in equations 1 and 2. That is, if $\widehat{\beta}_k$ are estimates of β_k , and $\widehat{\delta}_k$ are estimates of δ_k , $plim(\widehat{\beta}_k) = \beta_k$ and $plim(\widehat{\delta}_k) = \delta_k$, $\forall k = 1, \dots, K$. The predicted prices obtained from both equations are also consistent.⁴ That is, if \widehat{p}_i^0 refers to the predicted price of item i in period 0, then the probability limit of \widehat{p}_i^0 is:⁵

$$plim \left(\widehat{p}_i^0 | z_1^0, z_2^0, \dots, z_k^0 \right) = plim \left(\exp \left(\sum_{k=1}^K \widehat{\beta}_k z_{k,i}^0 \right) \right) = \exp(\theta_i^0) \quad (6)$$

Turning now to period 1, if \widehat{p}_v^1 is the predicted price of item v , then the probability limit of \widehat{p}_v^1 is:

$$plim \left(\widehat{p}_v^1 | z_1^1, z_2^1, \dots, z_k^1 \right) = plim \left(\exp \left(\sum_{k=1}^K \widehat{\delta}_k z_{k,v}^1 \right) \right) = \exp(\theta_v^1) \quad (7)$$

Now hedonic imputation methods are applied in order to obtain the price relative between periods 0 and 1 for item i sold in period 0 (hence, item i is a disappearing item). The single imputation price relative is as follows:

$$\frac{\widehat{p}_i^1}{\widehat{p}_i^0} = \frac{\exp \left(\sum_{k=1}^K \widehat{\delta}_k z_{k,i}^0 \right)}{\exp \left(\sum_{k=1}^K \beta_k z_{k,i}^0 + \epsilon_i^0 \right)} = \prod_{k=1}^K \exp \left(\widehat{\delta}_k - \beta_k \right) z_{k,i}^0 \times \exp(-\epsilon_i^0) \quad (8)$$

where \widehat{p}_i^1 in the numerator is obtained by multiplying the estimated coefficients of the characteristics of period 1 with the value of the corresponding characteristics in period 0. The regression function specified in equation 1 is substituted for \widehat{p}_i^0 in the denominator. In the double imputation method, \widehat{p}_i^0 is used instead of p_i^0 . The double imputation price relative for the same item i is obtained as follows:

⁴This can be shown by using the rules of probability limits. The rules are: (1) if $g(x_n)$ is a continuous function that is not a function of n , $plim(g(x_n)) = g(plim(x_n))$; and (2) if x_n and y_n are random variables with $plim(x_n) = c$ and $plim(y_n) = d$, then $plim(x_n \pm y_n) = c \pm d$, $plim(x_n \times y_n) = c \times d$ and $plim(x_n \div y_n) = c \div d$ (for $d \neq 0$). For other rules of probability limits, see, for example, appendix C, Wooldridge (2009).

⁵Goldberger (1968) shows that even if $\widehat{\beta}$ is an unbiased estimator of β , because of taking a non-linear transformation, $\exp(\widehat{\beta})$ is a biased estimator of $\exp(\beta)$. To correct for the bias, Kennedy (1981) suggests the use of $[\exp(\widehat{\beta} + 0.5\widehat{\sigma}^2)]$ for the estimator of $\exp(\beta)$, where $\widehat{\sigma}^2$ is an estimate of the variance of $\widehat{\beta}$. However, in this paper the correction factor for this bias is not incorporated in the derivation. This does not have any impact on the results of the paper.

$$\frac{\widehat{p}_i^1}{\widehat{p}_i^0} = \frac{\exp\left(\sum_{k=1}^K \widehat{\delta}_k z_{k,i}^0\right)}{\exp\left(\sum_{k=1}^K \widehat{\beta}_k z_{k,i}^0\right)} = \prod_{k=1}^K \exp\left(\widehat{\delta}_k - \widehat{\beta}_k\right) z_{k,i}^0 \quad (9)$$

By using the assumption of equation 1 that ϵ_i^0 is an *i.i.d* with a zero mean, $plim(-\epsilon_i^0) = 0$. Hence, the *plim* of the single imputation price relative:

$$plim\left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0}\right) = \prod_{k=1}^K \exp\left[plim\left(\widehat{\delta}_k\right) - plim\left(\beta_k\right)\right] z_{k,i}^0$$

Substituting equations 3 and 7, the following is obtained:

$$plim\left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0}\right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} = \exp(\Theta_i) \quad (10)$$

that is, the single imputation price relatives estimated for disappearing items using the imputed prices from equation 2 are consistent estimates of the true price relatives. Furthermore, it can be shown that for the double imputation price relatives:

$$plim\left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0}\right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} = \exp(\Theta_i) \quad (11)$$

The above derivation indicates that both \widehat{p}_i^1/p_i^0 and $\widehat{p}_i^1/\widehat{p}_i^0$ provide the correct measure of the price change of item i between periods 0 and 1 in the sense that they converge to the true price relative, $\exp(\Theta_i)$. This implies that if there is no potential for omission of characteristics, one can use either the single or the double imputation method. The next section shows that the prevalence of omitted characteristics shifts the preference towards the double imputation method.

3.2 Coefficient stability conditions required to attain consistency in the imputation methods

Let $\mathbf{Z} = [\mathbf{Z}_a \ \mathbf{Z}_b] = [\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_s; \ \mathbf{z}_{s+1} \ \dots \ \mathbf{z}_K]$, where \mathbf{Z} includes all the characteristics specified in equations 1 and 2. Let us now suppose that \mathbf{Z}_a contains the set

of observed characteristics and \mathbf{Z}_b contains the set of unobserved characteristics. Thus, the log of prices is now estimated on the characteristics z_1, z_2, \dots, z_s for both periods 0 and 1. A further assumption is made that each of the unobserved characteristics is correlated with the set of observed characteristics. This implies that the models estimated on the reduced set of characteristics suffer from the omitted variable problem, leading to biased and inconsistent estimates of the coefficients and predicted prices.⁶

Now, using the ordinary least squares (OLS) method, the following predicted log price for item i is obtained:

$$\widetilde{\ln p}_i^0 = \sum_{j=1}^s \tilde{\xi}_j z_{j,i}^0 \quad (12)$$

Let $\gamma_{j,k}^0 = E(z_k^0)$ for $j = 1$ (i.e. for the intercept term) and $\forall k = s + 1, \dots, K$, and $\gamma_{j,k}^0 = Cov(z_j^0, z_k^0)/Var(z_j^0)$, $\forall j = 2, \dots, s$ and $\forall k = s + 1, \dots, K$. The probability limit of $\widetilde{\ln p}_i^0$ in equation 12 is taken, and $E(u) = 0$ and $Cov(z_j^0, u) = 0$, $\forall j = 1, \dots, s$ are set to get the following:

$$plim(\widetilde{\ln p}_i^0) = \sum_{j=1}^s \beta_j z_{j,i}^0 + \sum_{j=1}^s \sum_{k=s+1}^K \gamma_{j,k}^0 \beta_k z_{j,i}^0 \quad (13)$$

Then, $\sum_{k=s+1}^K \beta_k z_{k,i}^0$ is added to the first term and subtracted from the second term of the right-hand side of the equation to get:

$$\begin{aligned} plim(\widetilde{\ln p}_i^0) &= \sum_{j=1}^K \beta_j z_{j,i}^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \\ &= \theta_i^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \end{aligned} \quad (14)$$

Consistency requires that the term $\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)$ equals zero. This term is zero if $\forall k = s + 1, \dots, K$ either $\beta_k = 0$ or $\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 = 0$.

⁶For a detailed derivation of omitted variable bias in a multi-variable context, see, for example, Johnston (1984).

But $\beta_k \neq 0$ since by assumption z_k^0 is a relevant characteristic. The other term $\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 = 0$, if for item i , z_k^0 is a perfect linear combination of the included characteristics, i.e. $z_{k,i}^0 = \gamma_{1,k} z_{1,i}^0 + \gamma_{2,k} z_{2,i}^0 + \dots + \gamma_{s,k} z_{s,i}^0$. This may happen for an item for a given z_k^0 , but is unlikely to happen for $\forall k = s+1, \dots, K$. Now let us consider all items, $i = 1, \dots, I$. For all i , $\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0$ implies that z_k^0 is a perfect linear combination of the included characteristics. This is a redundant condition because if a characteristic is a perfect linear combination of the other characteristics then it would not have entered in the model specified in equation 1. There is another possibility that since the term $\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)$ involves summation, and there are positive and negative numbers, these numbers somehow may cancel each other out. However, there is no justification for this to happen. Hence, it can be concluded that:

$$plim(\widetilde{lnp}_i^0) \neq \theta_i^0$$

that is, \widetilde{lnp}_i^0 is an inconsistent estimator of θ_i^0 .

For period 1, let us denote the OLS estimates of the coefficient of the z_j^1 characteristic by $\widetilde{\eta}_j$. The imputed price of item i with the values of the characteristics set $z_1^0, z_2^0, \dots, z_j^0$ is $\widetilde{lnp}_i^1 = \sum_{j=1}^s \widetilde{\eta}_j z_{j,i}^0$. Note that item i is sold in period 0, not in period 1. Let $\gamma_{j,k}^1 = E(z_k^1)$ for $j = 1$ and $\forall k = 1, \dots, K$, and $\gamma_{j,k}^1 = Cov(z_j^1, z_k^1) / Var(z_j^1)$, $\forall j = 2, \dots, s$ and $\forall k = 1, \dots, K$. Following the similar operations shown in equations 13 and 14, it can be shown that:

$$plim(\widetilde{lnp}_i^1) = \theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right), \quad (15)$$

where, following on from the explanation provided for equation 14, it can be shown that the term $\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \neq 0$. This is mainly because $\delta_k \neq 0$ and z_k^0 is not a perfect linear combination of the set of included characteristics, $\forall k = s+1, \dots, K$. Hence:

$$plim(\widetilde{lnp}_i^1) \neq \theta_i^1$$

that is, $\widetilde{\ln p}_i^1$ is an inconsistent estimator of θ_i^1 .

The single and double imputation price relatives estimated from models with reduced sets of characteristics are \widetilde{p}_i^1/p_i^0 and $\widetilde{p}_i^1/\widetilde{p}_i^0$, respectively. In order to derive the conditions required for the price relatives to be consistent with respect to $\exp(\Theta_i)$, their probability limits are taken. The following is the derivation for the single imputation method:

$$plim \left(\frac{\widetilde{p}_i^1}{p_i^0} \right) = plim \left[\frac{\exp(\widetilde{\ln p}_i^1)}{\exp(\ln p_i^0)} \right] = \frac{\exp \left[plim(\widetilde{\ln p}_i^1) \right]}{\exp \left[plim(\ln p_i^0) \right]}$$

By substituting equation 3 in the denominator and equation 15 in the numerator, and setting $\exp(\theta_i^1)/\exp(\theta_i^0) = \exp(\Theta_i)$ (see equation 5), the following is obtained:

$$\begin{aligned} plim \left(\frac{\widetilde{p}_i^1}{p_i^0} \right) &= \frac{\exp \left[\theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right]}{\exp \left[\theta_i^0 \right]} \\ &= \exp(\Theta_i) \times \exp \left(\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right) \end{aligned} \quad (16)$$

Now the term $\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \neq 0$. The explanation is the same as is provided in equations 14 and 15. Hence:

$$plim \left(\frac{\widetilde{p}_i^1}{p_i^0} \right) \neq \exp(\Theta_i)$$

i.e. the single imputation method estimates inconsistent price relatives.

The bias is given by $\exp \left(\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right)$. The direction and magnitude of the bias depend on a number of factors, including some unknown values (for example, $\gamma_{j,k}^1$). Hence, in practice it is impossible to find the direction and magnitude of bias corresponding to each of the price relatives. Moreover, Silver and Heravi (2005), Haan (2007), and Melser and Syed (2008) argue that the pricing patterns be-

tween new and disappearing items may vary systematically. This difference in pricing patterns may lead to a systematic difference in the estimated residuals obtained from hedonic regressions corresponding to new and disappearing items, implying that the biases do not offset each other. It should also be noted that the bias depends on the parameters of period 1 only, i.e. the period for which the price of item i is imputed, implying that the expressions for bias vary between the single and double imputation methods.

Now, in order to derive the conditions for consistency for the double imputation method, the probability limit of the double imputation price relative, $\tilde{p}_i^1/\tilde{p}_i^0$, is taken:

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = plim \left[\prod_{j=1}^s \exp \left(\tilde{\delta}_j - \tilde{\beta}_j \right) z_{j,i}^0 \right]$$

Using the rules of probability limits mentioned earlier, the following is obtained:

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = \frac{plim \left[\prod_{j=1}^s \exp \left(\tilde{\delta}_j z_{j,i}^0 \right) \right]}{plim \left[\prod_{j=1}^s \exp \left(\tilde{\beta}_j z_{j,i}^0 \right) \right]} = \frac{\exp \left[plim \left(\sum_{j=1}^s \tilde{\delta}_j z_{j,i}^0 \right) \right]}{\exp \left[plim \left(\sum_{j=1}^s \tilde{\beta}_j z_{j,i}^0 \right) \right]}$$

By substituting equations 14 and 15, it can be shown that:

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} \times \frac{\exp \left[\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right]}{\exp \left[\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \right]} \quad (17)$$

Hence, it can be seen from equation 17 that:

If

$$\frac{\exp \left[\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right]}{\exp \left[\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \right]} = 1 \quad (18)$$

or

$$\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) = \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \quad (19)$$

then

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} = \exp(\Theta_i)$$

The magnitude of the bias is given by the difference between between 1 and the left-hand side of equation 18. If the difference is positive, then the model overestimates P_i^1/P_i^0 , and if the difference is negative, the model underestimates P_i^1/P_i^0 . In the next section, a formula for the empirical estimates of the magnitude of bias is derived.

Equation 19 is referred to as the stability condition because consistency requires the stability of two sets of parameters. First, the regression coefficients of the unobserved characteristics in the hedonic equations are required to be the same across periods. If they are not the same across periods, then this may bias the hedonic price indexes. In appendix A, similar derivations are undertaken for the level hedonic models. It is shown that the exact expressions of the stability conditions or the sources of bias differ between the log and the level hedonic models (compare equations 19 and A.5).

The second set of parameters is related to the regression coefficients obtained when each of the unobserved characteristics is regressed on each of the observed characteristics. These regression coefficients are required to be the same across periods, i.e. $\gamma_{j,k}^0 = \gamma_{j,k}^1, \forall j = 1, \dots, s$ and $\forall k = s+1, \dots, K$. The second set of regression coefficients reflect how each of the excluded characteristics is configured with the set of included characteristics.⁷ In the next section, a technique is derived to empirically test whether equation 19 holds. This derivation will also provide a better intuitive understanding of what equation 19 implies.

⁷Diewert et al. (2008), while comparing the time-dummy and hedonic imputation methods, find that the change in the configuration of characteristics between the comparison periods is one of the factors that determine the difference in the indexes obtained from the time-dummy and the hedonic imputation methods.

3.3 Using matched items to estimate the magnitude of omitted variable bias

Let us now denote the items that were sold in both periods, the matched items, by $m = 1, \dots, M$. For these items, both the price relatives—ratios of observed prices and double imputation price relatives—are available. This enables the evaluation of the performance of the estimated hedonic models—the closer the estimated price relatives to the observed price relatives, the lower is the omitted variable bias contaminating the estimated price relatives.

It should be noted that the matched items dominate the total number of items sold between two adjacent periods.⁸ For the purpose of constructing price indexes, prices are usually compared between two adjacent months or between two adjacent quarters. Benkard and Bajari (2005) report that more than 90% of personal computer models observed in one month are also observed in the next month, though they argue at the same time that it is impossible to obtain a reliable measure of price comparison using the matched models because more than 90% of the models drop out within a period of one year. Similarly, from table 2 of Silver and Heravi (2005), it can be calculated that typically more than 95% of models are matched models between two adjacent months.⁹

Following from equation 12, it is possible to write the equation for period 0 as follows:

$$\ln p_m^0 = \widetilde{\ln p_m^0} + \widetilde{e}_m^0 \quad \forall m = 1, \dots, M \quad (20)$$

where \widetilde{e}_m^0 is the estimated error when $\ln p_m^0$ is estimated on the reduced set of characteristics. By rearranging terms in equation 20 and taking the probability limits of both sides of the equation, the following is obtained:

⁸An exception is the housing market.

⁹It should be emphasized that although the majority of items are matched between two consecutive months, this does not undermine the matching problem caused by new and disappearing items. This is because, as the evidence shows, the problem becomes severe—both in terms of the turnover rates and the resulting bias in indexes—because of the cumulative effect over a period of time.

$$plim \left(\widetilde{e}_m^0 \right) = plim \left(lnp_m^0 \right) - plim \left(\widetilde{lnp}_m^0 \right) \quad \forall m = 1, \dots, M$$

By substituting $plim \left(lnp_m^0 \right) = \theta_m^0$ (see equation 3) and equation 14 for $plim \left(\widetilde{lnp}_m^0 \right)$, the following is obtained:

$$plim \left(\widetilde{e}_m^0 \right) = - \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,m}^0 - z_{k,m}^0 \right) \quad \forall m = 1, \dots, M \quad (21)$$

Similarly, for period 1, the following is obtained:

$$plim \left(\widetilde{e}_m^1 \right) = - \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,m}^1 - z_{k,m}^1 \right) \quad \forall m = 1, \dots, M \quad (22)$$

where \widetilde{e}_m^1 is the estimated error when lnp_m^1 is estimated on the reduced set of characteristics. Equations 4 and 15 are used for the derivation of equation 22.

Note that equations 21 and 22 refer to matched item m , implying that each of the characteristics of item m for the two periods are the same, i.e. $z_{k,m}^0 = z_{k,m}^1 = z_{k,m}$, $\forall k = 1, \dots, K$. This implies that equations 21 and 22 are simply the negative of the left- and right-hand sides of equation 19, respectively, when equation 19 corresponds to item m , $\forall m = 1, \dots, M$. Thus, the bias in equation 18 can be estimated by:

$$\frac{exp \left(\widetilde{e}_m^0 \right)}{exp \left(\widetilde{e}_m^1 \right)} - 1 \quad (23)$$

If $\left[exp \left(\widetilde{e}_m^0 \right) / exp \left(\widetilde{e}_m^1 \right) - 1 \right]$ is greater than 0, then the double imputation overestimates the true price relative of item m ; if the difference is less than 0, the double imputation underestimates the true price relative. Among alternative models with different sets of characteristics, the model that minimizes $\left| exp \left(\widetilde{e}_m^0 \right) / exp \left(\widetilde{e}_m^1 \right) - 1 \right|$ estimates price relatives that are contaminated the least by the omission of relevant characteristics. In a similar way, in the case of the single imputation method the bias in equation 16 can be estimated by:

$$\frac{1}{\exp\left(\widetilde{e}_m^1\right)} - 1 \quad (24)$$

3.4 Plausibility of equal correlations between the characteristics across periods

Let us assume that the regression coefficients between the observed and unobserved characteristics are equal between the two periods, i.e. $\gamma_{j,k}^0 = \gamma_{j,k}^1, \forall j = 1, \dots, s$ and $k = s+1, \dots, K$. Hence, in equation 19, $\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 = \sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 = w_{k,i}^0$. Thus, the stability condition in equation 19 reduces to:

$$\sum_{k=s+1}^K \delta_k w_{k,i}^0 = \sum_{k=s+1}^K \beta_k w_{k,i}^0$$

or

$$\delta_k = \beta_k, \quad \forall k = s+1, \dots, K \quad (25)$$

This implies that if the hedonic models are estimated for the log of prices, the consistency of the estimated price relatives requires that the coefficients of the same unobserved characteristic be equal between the comparison periods, provided that there has not been any significant change in the configuration of the characteristics of items between the comparison periods.

The assumption of the equality of the regression coefficients may be more plausible in one market over another. Let us consider the housing market. For the housing market this assumption would mean, say with regard to the relationship between the number of bedrooms and lot size, that for a given number of bedrooms, the average lot size is the same across periods. The housing market may be characterized by a high degree of sluggishness from the supply side, where transactions of old houses by far dominate the market. The matching problem arises because every house is different, but it may be reasonable to assume that the configuration between the characteristics remains stable in the housing market between two adjacent periods (for example, between two quarters

of a year). However, this assumption is unlikely to hold true in markets characterized by a rapid rate of technological innovation, such as the personal computer market. This is because technological advancement occurs at different rates with regard to different features of personal computers. Thus, quality change is reflected in the change in the configuration of characteristics, such as between the characteristics ‘CPU speed’ and ‘hard-disk size’.

Perhaps the housing market and the personal computer market are two extreme examples in terms of the expected nature of qualitative change between two adjacent periods. There are many other products where the matching problem arises in the construction of indexes and where the changes in the configuration of characteristics are not as pronounced as in the personal computer market, yet not as rigid as in the housing market. This may include markets for many electronic products other than computers, as well as the used car market and the markets for supermarket products such as laundry products, beverages, etc. In these markets, qualitative changes may occur in a few characteristics, and the stability condition indicates that in order to attain consistency these characteristics should be included in the model. Testing $\gamma_{j,k}^0 = \gamma_{j,k}^1$ in different markets may be an interesting area of further research.

4. Monte Carlo Simulations and Unbiased Price Relatives

The analytical results in the previous section are obtained for large samples (or as asymptotic properties) in order to estimate consistent price relatives. Now the question is: do the results also hold for finite or small samples, and equivalently apply in order to estimate unbiased price relatives?¹⁰ In this section, Monte Carlo simulations are conducted: first, to check whether equation 18 or 19 which produces consistent price

¹⁰Note that it is more desirable to estimate unbiased and consistent estimators than to estimate consistent but biased estimators. The analytical results focus on consistency instead of unbiasedness because the derivations involving ratios and multiplications of random variables are relatively straightforward, with probability limits rather than expectational operators.

relatives in the case of the double imputation method, produces unbiased price relatives at the same time; and, second, to estimate the bias specified in equations 23 and 24 for both small and large samples. The simulations are conducted with different specifications of the true and estimated hedonic models.

In period 0, the true model is obtained from the process $\ln p_i^{(0)} = 0.9 + 0.9x_{1,i} + 0.9x_{2,i} + \varepsilon_i$ and, in period 1, the true model is obtained from the process $\ln p_i^{(1)} = 0.9 + \kappa_1 x_{1,i} + \kappa_2 x_{2,i} + \varepsilon_i$. ε_i is drawn from a normal distribution, $\varepsilon_i \sim N(0, 1)$. The simulation exercise includes changing κ_1 and κ_2 to produce different scenarios. The only two characteristics, x_1 and x_2 , are drawn from a multivariate normal distribution with means π_1 and π_2 , variances ω_1^2 and ω_2^2 , respectively, and covariance ρ_{12} . For the matched items, which account for 90% of the total items in periods 0 and 1, the means and the variances are $[\pi_1, \pi_2, \omega_1^2, \omega_2^2, \rho_{12}] = [3, 3, 1, 1, 0.5]$.

The disappearing and new items account for 10% of the items in periods 0 and 1, respectively. The technological innovation occurs through changes in one of the characteristics, x_1 or x_2 , which may be reflected in the simulations through changes in the means and marginal valuations of x_1 or x_2 . For example, if the innovation occurs through x_1 , then the mean of x_1 for the disappearing items (π_1^d) may be expected to be lower than the mean of x_1 for the matched items (π_1), which in turn may be expected to be lower than the mean of x_1 for the new items (π_1^n). Examples of x_1 , in the case of computers, are the RAM or hard disk sizes of computers. Another way to look at innovation is through cost savings. If the cost saving occurs through x_1 then this may be reflected in the simulation through a fall in κ_1 between period 0 and 1.

The models in period 0 and 1 are estimated on an intercept and x_1 , hence the models omit x_2 . Whether x_1 is the stable or the unstable characteristic depends on whether π_1^d or π_1^n is different from 3 or whether κ_1 is different from 0.9. Let $[\kappa_1, \kappa_2] = [0.1, 0.9]$, i.e. the marginal valuation of x_1 falls in period 1, while the marginal valuation of x_2 remains the same. Furthermore, let $[\pi_1^d, \pi_1^n] = [3, 3]$ and $[\pi_2^d, \pi_2^n] = [3, 3]$, i.e. the mean values of the characteristics remain the same between the disappearing, matched and new items. In this example, the unstable characteristic is included and the stable char-

acteristic is excluded. Hence, according to equations 18 or 19, the double imputation method is expected to produce consistent price relatives.

In order to check for unbiasedness, 8 Monte Carlo simulations corresponding to $10 \times 3^{k-1}$ iterations for $k = 1, \dots, 8$, and each having a sample of 1000 observations, are conducted. The means of the estimates of equation 23 converge towards 0, implying that the double imputation price relatives are unbiased (see top part of columns 2 and 3 of table 1). In order to check for consistency, the sample size is increased by $10 \times 3^{k-1}$ observations for $k = 1, \dots, 8$, while the number of iterations is kept fixed at 10. The results show that the mean and variance converge to zero rapidly with the increase in sample size (see the bottom part of columns 2 and 3 in table 1). Thus the simulation results show that the double imputation price relatives are unbiased and consistent.¹¹

Insert table 1 here.

The same estimates corresponding to equation 24, i.e. for the single imputation method, are shown in columns 4 and 5 of table 1. The results show that there is no tendency for the mean to converge to 0 with the increase in the number of iterations, and the mean and variance to converge to 0 with the increase in the sample size. This provides evidence that the single imputation price relatives are biased and inconsistent.

Now let $[\pi_1^d, \pi_1^n] = [1, 5]$, $[\pi_2^d, \pi_2^n] = [3, 3]$ and $[\kappa_1, \kappa_2] = [0.1, 0.9]$. In this case, the technological progress occurs through the provision of an increasing number of x_1 for a given x_2 , and through cost savings in x_1 . Because of differing means, the variance of x_1 differs between periods 0 and 1. This leads the correlation coefficient to change between the two periods (i.e. $\gamma_{j,k}^0 \neq \gamma_{j,k}^1$ in equation 19). The simulation results provided in columns 3 and 4 of table 2 show no tendency for the estimates of the means of equation 23 to converge to zero, implying that the double imputation method estimates biased price relatives.¹² The same conclusion is drawn with regard to equation 24 for the single

¹¹If a large difference is created at the initial stage by setting a large difference between κ_2 and 0.9, between π_1 and π_2 and by increasing the variance of ε_i , the results with regard to convergence are the same.

¹²If the change in mean value had happened with x_2 , then the double imputation price relatives would have been consistent.

imputation method (see columns 5 and 6 of table 2). Similarly, if $[\kappa_1, \kappa_2] = [0.9, 0.1]$, i.e. the omitted characteristic sees a drop in the marginal valuation, both the single and double imputation price relatives are biased and inconsistent (see bottom part of table 2). However, importantly, the biases are found to be higher for the single imputation method than for the double imputation method.¹³

Insert table 2 here.

With regard to the time-dummy method, it can be easily shown that the bias generated by the time-dummy method is substantial even in a trivial case where the double imputation method produces unbiased and consistent price indexes. Let us suppose that all the items are matched between periods 0 and 1. Furthermore, let us suppose that $[\kappa_1, \kappa_2] = [0.9, 0.9]$. As before, the models in both periods are estimated on intercept and x_1 . It can be shown that the Jevons index constructed from taking the geometric mean of the estimated double imputation price relatives is 1, i.e. the double imputation method provides the correct measure of price change. On the contrary, the time-dummy index—which is equivalent to the Jevons index—provides a biased estimate of price change, where the magnitude of bias depends on the correlation between x_2 and the time dummy. In the above example, even if the correlation is as low as 0.1, the time-dummy index overestimates the price change by 9%.

5. Implications for Applications

There are a number of important implications of the results with regard to the application of hedonic methods in the construction of quality adjusted indexes. First, if the omitted variable bias is one of the main problems inhibiting the application of hedonic methods, the single imputation method does not offer a better solution than the time-dummy method. Both methods are equally demanding of the characteristics information that is required to produce unbiased and consistent estimates of price change. The double imputation method may produce unbiased and consistent estimates of price

¹³The results with regard to consistency are similar to that of unbiasedness.

change even in the presence of a large number of unobserved characteristics. Also, the simulation results show that even if the bias exists, the magnitude of bias is in general lower in the double imputation than in the single imputation price relatives.

Second, the data requirement shifts from ‘all price determining characteristics’ to ‘price determining characteristics that drive the price changes between the comparison periods’. If the log model is the preferred model, the data compilers may focus on collection of information on the characteristics whose contribution, both in terms of their marginal values to price and how they are configured with the other characteristics, is expected to change between the periods. This may, on the one hand, reduce the data requirement by a large amount while, on the other hand, make it possible to obtain estimates that are ‘near consistent’, if not consistent.

The third implication is related to the second implication. The idea of what constitutes an important variable differs between the single regression equation (hedonic or otherwise) and the hedonic double imputation method. In the case of single regression equations, an important variable is one that has a relatively large explanatory power, whereas in the case of the hedonic imputation method an important variable is one that has a large influence on the price change between the periods. In both cases, the important variables may coincide in many markets but not necessarily in all markets.

Fourth, the stability conditions on the unobserved characteristics, conditions which are required for unbiasedness and consistency, differ between the log and level models (the derivation for the level hedonic models is provided in appendix A). In general, while log models may be preferable where price changes are driven by a selected number of characteristics, level models may be appropriate where a large number of characteristics coefficients are expected to change by the same proportion. For example, the log models may be appropriate for electronic markets because technological advancement may be driven by a selected number of features in a particular electronic product, and the level models may be more appropriate for the housing market because price is driven by expectations of the future flow of income, where the expectations formed in two adjacent periods may be similar with regard to the different physical features of a

house.

Finally, in recent years a number of studies comparing the time-dummy and hedonic imputation methods have argued in favor of using the hedonic imputation method (see, for example, Silver and Heravi, 2007; Diewert et al., 2008). This is because the hedonic imputation method allows the characteristics coefficients to vary, whereas the time-dummy method constrains them to remain fixed between the periods. These studies analyze the observed characteristics only and do not draw any distinction between the single and double imputation methods. In contrast, the current study focusses on the unobserved characteristics and conducts a separate analysis on the single and double imputation methods based on the omitted variable bias generated by the unobserved characteristics. Although the methodological approach is different, the conclusion of this study accords with the recent literature which argues in favor of using hedonic imputation methods. Moreover, this study finds justification in favor of the double imputation method. This is because the double imputation method has the potential to address the omitted bias problem in hedonic regressions with the careful choice of limited but selected characteristics.

6. Conclusion

In the case of single hedonic regressions, the omitted variable bias problem is similar to the problem with regressions in general, i.e. if the relevant variables that are correlated with the included variables are omitted, the estimates are biased and inconsistent. However, this is not the case for the price relatives estimated from the double imputation method, which involves the ratio of two hedonic regressions, because of the potential of biases being canceled out. This study has taken a systematic look at the omitted variable problem of hedonic imputation methods by analyzing the asymptotic bias generated by each of the hedonic regressions. The study analyzed the log and level hedonic models for both the single and double imputation methods.

The study found that the single imputation method does not have any advantage

over the time-dummy method in terms of dealing with the omitted variable problem. If the hedonic regressions exclude any relevant price determining characteristics, both methods produce biased and inconsistent estimates. The double imputation method has a clear advantage in this case. This study found that double imputation price relatives can be unbiased and consistent even if the important price determining characteristics corresponding to each of the hedonic regressions are omitted. The requirement for attaining unbiasedness and consistency is that the contributions of the unobserved characteristics in price determination are stable between the periods. Depending on the products, this has the potential of reducing the data requirement, in terms of characteristics information, by a large extent.

This study has outlined a method to estimate the magnitude of bias in the single and double imputation price relatives for matched items. This method may indicate how the hedonic imputation method performs in predicting the price relatives of unmatched, i.e. new and disappearing, items. This estimated magnitude of bias may provide a benchmark summary measure that can be used to compare the performance of different hedonic imputation models applied to different data sets.

Appendix A: Hedonic models of price levels

Let us specify the models for periods 0 and 1 on price levels as follows:

$$p_i^0 = \sum_{k=1}^K \beta_k z_{k,i}^0 + \epsilon_i^0 \quad \forall i = 1, \dots, I \quad (\text{A.1})$$

$$p_v^1 = \sum_{k=1}^K \delta_k z_{k,v}^1 + \epsilon_v^1 \quad \forall v = 1, \dots, V \quad (\text{A.2})$$

where, similar to equations 1 and 2, $z_{k,i}^0$ and $z_{k,i}^1$ refer to the value of the characteristics k for item i in periods 0 and 1, respectively. The first characteristics, z_1^0 and z_1^1 , are the intercept terms of the equations. The error terms ϵ_i^0 and ϵ_v^1 are *i.i.d.* with zero mean and constant variance. Furthermore, the classical linear regression model assumptions

hold for models in equations A.1 and A.2.

Following similar steps to those shown in equations 6 to 11 for the log models, it can be shown that the single and double imputation price relatives estimated from equations A.1 and A.2 are consistent. That is:

$$plim \left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0} \right) = plim \left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0} \right) = \frac{\theta_i^1}{\theta_i^0} = \Phi_i \quad (\text{A.3})$$

Let us now suppose that instead of estimating models on $z_1, \dots, z_s, z_{s+1}, \dots, z_K$ characteristics as specified in equations A.1 and A.2, the models are estimated on z_1, \dots, z_s . Let \widetilde{p}_i^0 and \widetilde{p}_v^1 be the predicted prices obtained from the estimated models for items i and v , respectively. Following steps similar to those in equations 13 and 14, it can be shown that $plim \left(\frac{\widetilde{p}_i^0}{\widetilde{p}_i^0} \right) = \theta_i^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \neq \theta_i^0$, and, following steps similar to those in equation 15, it can be shown for the imputed price of item i in period 1 that $plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right) = \theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \neq \theta_i^1$. Thus, the probability limit of the single imputation price relative for item i is:

$$plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right) \neq \frac{\theta_i^1}{\theta_i^0} = \Phi_i$$

that is, the single imputation price relative for item i is inconsistent.

Now to the double imputation price relatives. Substituting the expressions for $plim \left(\frac{\widetilde{p}_i^0}{\widetilde{p}_i^0} \right)$ and $plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right)$, the following is obtained:

$$plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right) = \frac{\theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right)}{\theta_i^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)} \quad (\text{A.4})$$

Equation A.4 implies that if

$$\frac{\theta_i^1}{\theta_i^0} = \frac{\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right)}{\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)} \quad (\text{A.5})$$

the double imputation price relatives estimated from level hedonic models are consistent.

Note that equation A.5 is the level hedonic model equivalent of equation 19 for the log

hedonic model.

Thus to evaluate the performance of alternative models, the following is estimated for the matched item m :

$$\left(\frac{p_m^1}{p_m^0}\right) - \left(\frac{\widetilde{e}_m^1}{\widetilde{e}_m^0}\right) \quad (\text{A.6})$$

The closer the value is towards zero, the lower is the omitted variable bias in the estimated price relative of item m .

If $\gamma_{j,k}^0 = \gamma_{j,k}^1, \forall j = 1, \dots, s$ and $k = s+1, \dots, K$ are set, then the stability condition in equation A.5 reduces to:

$$\frac{\sum_{j=1}^s \delta_j z_{j,i}}{\sum_{j=1}^s \beta_j z_{j,i}} = \frac{\sum_{k=s+1}^K \delta_k z_{k,i}}{\sum_{k=s+1}^K \beta_k z_{k,i}} \quad (\text{A.7})$$

Thus, the stability condition for the level models specified in equation A.7 differs from that for the log models specified in equation 25. This implies that hypotheses about the behavior of unobserved characteristics across periods have implications in the choice between the level and the log models in the double imputation hedonic method.

References

- Benkard, C.L., Bajari, P. (2005): Hedonic Price Indexes with Unobserved Product Characteristics, and Application to Personal Computers, *Journal of Business & Economic Statistics* 23(1), 61-75.
- Berndt, E.R., Rappaport, N.J. (2001): Price and Quality of Desktop and Mobile Personal Computers: A Quarter-Century Historical Overview, *American Economic Review* 91(2), 268-273.
- Boskin, M., Dullberger, E., Gordon, R., Griliches, Z., Jorgenson, D. (1996): Towards a More Accurate Measure of the Cost of Living, Advisory Commission to Study the Consumer Price Index, Final Report to the Senate Finance Committee, Washington, D.C.

- Court, A.T. (1939): Hedonic Price Indexes with Automotive Examples, in *The Dynamics of Automobile Demand*, General Motors Corporation, New York, 98-119.
- Diewert, W.E., Heravi, S., Silver, M. (2008): Hedonic Imputation versus Time Dummy Hedonic Indexes, NBER Working Paper 14018, National Bureau of Economic Research, Cambridge, <http://www.nber.org/papers/w14018>.
- Goldberger, A.S. (1968): The Interpretation and Estimation of Cobb-Douglas Functions, *Econometrica* 36(3/4), 464-472.
- Griliches, Z. (1961): Hedonic Price Indexes for Automobiles: An Econometric Analysis of Duality Change, in *The Price Statistics of the Federal Government*, General Series 73, Columbia University and National Bureau of Economic Research, 137-196.
- Haan, J.de (2007): Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches, Paper presented at the 10th Ottawa Group Meeting, October 9-12, 2007, Ottawa, http://www.ottawagroup2007.ca/r004/pdf/ogo04_033_e.pdf.
- Hill, R.H., Melser, D. (2008): Hedonic Imputation and the Price Index Problem: An Application to Housing, *Economics Inquiry* 46(4), 593-609.
- Hulten, C.R. (2003): Price Hedonics: A Critical Review, *FRBNY Economic Policy Review* 9, 5-15.
- Johnston, J. (1984): *Econometric Methods* (3rd ed.), New York: McGraw-Hill Book Company.
- Kennedy, P.E. (1981): Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations, *American Economic Review* 71(4), 801.
- Melser, D., Syed, I. (2008): Pricing over the Product Life Cycle: An Empirical Analysis, Research Paper No. 2008 ECON 25, Australian School of Business, University of New South Wales, Sydney.

- Moulton, B.R. (2001): The Expanding Role of Hedonic Methods in the Official Statistics of the United States, U.S. Bureau of Economic Analysis, Washington, D.C, <http://www.bea.gov/about/pdf/expand3.pdf>.
- Pakes, A. (2003): A Reconsideration of Hedonic Price Indexes with an Application to PCs, *American Economic Review* 93(5), 1578-1596.
- Schultze, C.L., Mackie, C. (2002): *At What Price? Conceptualizing and Measuring the Cost-of-Living and Price Indexes*, Washington, D.C.: National Academy Press.
- Silver, M., Heravi, S. (2005): A Failure in the Measurement of Inflation: Results From a Hedonic and Matched Experiment Using Scanner Data, *Journal of Business & Economic Statistics* 23(3), 269-281.
- (2007): The Difference Between Hedonic Imputation Indexes and Time Dummy Hedonic Indexes, *Journal of Business & Economic Statistics* 25(2), 239-246.
- Tripllett, J.E. (2006): *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD Directorate for Science, Technology and Industry, Paris.
- van Mulligen, P.H. (2003): *Quality Aspects in Price Indices and International Comparisons: Application of the Hedonic Method*, Unpublished Ph.D Dissertation, University of Groningen.
- Waugh, F.V. (1928): Quality Factors Influencing Vegetable Prices, *Journal of Farm Economics* 10, 186-196.
- Wooldridge, J.M. (2009): *Introductory Econometrics: A Modern Approach*. Ohio: Thomson South-Western.

Table 1: Simulation Results for Double and Single Imputation Price Relatives when the Stable Characteristic is Omitted

Simulations	Double Imputation		Single Imputation	
	$\kappa_1 = 0.1, \kappa_2 = 0.9$		$\kappa_1 = 0.1, \kappa_2 = 0.9$	
	Mean	Variance	Mean	Variance
No of iterations ¹				
10	0.0042	0.0244	0.0123	1.0064
30	0.0072	0.0313	0.8390	3.3538
90	-0.0006	0.0281	0.4580	1.5262
270	0.0016	0.0246	1.0442	3.5721
810	0.0006	0.0253	1.2725	4.1336
2460	0.0005	0.0247	1.2531	3.7287
7290	0.0004	0.0251	1.1870	3.7190
21870	0.0002	0.0255	1.2032	4.1710
Sample sizes ²				
100	0.0534	0.0863	0.7052	2.2129
300	-0.0204	0.0548	0.4410	0.9403
900	-0.0031	0.0192	0.4976	1.3767
2700	0.0017	0.0148	4.5119	12.9150
8100	0.0028	0.0104	1.1284	4.6546
24600	-0.0001	0.0063	2.1059	3.4185
72900	0.0012	0.0021	0.4586	2.3750
218700	-0.0002	0.0016	1.2400	2.5616

Notes: 1. Sample size=1000; 2. No. of iterations=10

Table 2: Simulation Results for Double and Single Imputation Price Relatives when the Configuration of Characteristics Changes and the Unstable Characteristics are Omitted

Simulations	No. of Iterations	Double Imputation		Single Imputation	
		Mean	Variance	Mean	Variance
$\pi_1^d = 1, \pi_1^n = 5$ $\kappa_1 = 0.1, \kappa_2 = 0.9$	10	-0.0417	0.1137	0.8475	1.8530
	30	0.0027	0.1593	0.4219	1.2269
	90	0.0139	0.1837	0.4857	1.7682
	270	0.0174	0.1386	1.5814	6.1601
	810	0.0193	0.1556	1.1893	4.0624
	2460	-0.0035	0.1499	1.0130	4.3506
	7290	0.0029	0.1522	1.1450	4.4066
	21870	0.0043	0.1490	1.1796	4.2468
$\pi_1^d = 3, \pi_1^n = 3$ $\kappa_1 = 0.9, \kappa_2 = 0.1$	10	0.0852	0.3870	1.1917	2.1996
	30	-0.1728	0.7752	1.2507	2.5072
	90	0.2815	0.8591	0.3956	1.8394
	270	0.2448	0.8736	0.8178	2.3800
	810	0.2609	1.1236	0.7106	2.4986
	2460	0.2692	0.9780	0.5711	1.8508
	7290	0.2649	0.9805	0.6481	2.1202
	21870	0.2751	0.9913	0.6563	2.1727

Note: Sample size for all simulations is 1000.

Consistency of Hedonic Price Indexes with Unobserved Characteristics

Iqbal Syed

*School of Economics, University of New South Wales,
Sydney, NSW 2052, Australia.
(email: i.syed@unsw.edu.au)*

Abstract

Hedonic regressions are prone to omitted variable bias because many of the price determining characteristics are typically unobserved. The estimation of price relatives for new and disappearing goods using hedonic imputation methods involves taking ratios of two hedonic models corresponding to two consecutive periods. This may lead to a situation where the omitted variable bias in one of the hedonic regressions offsets the other. This study finds that the single imputation hedonic method estimates inconsistent price relatives, while the double imputation method may produce consistent price relatives depending on the behavior of unobserved characteristics in the comparison periods. The study outlines a methodology to estimate the magnitude of bias in single and double imputation price relatives. The results of this study have implications with regard to the construction of quality adjusted indexes.

JEL Classification Numbers: C43; C52; E31

Keywords: Hedonic imputation method; Model selection; New and disappearing goods; Omitted variable bias; Quality adjusted price indexes.

1. Introduction

An hedonic regression model specifies the price of a variety v at time t , $P_{v,t}$, as a function of its determining characteristics and a random error term. The most common use of hedonic regressions has been to disentangle the quality component of price change from the observed price change to achieve better measures of ‘pure’ inflation. Hedonic regressions have been used for quality adjustment in price changes since Waugh (1928) and Court (1939), with Griliches (1961) reviving interest in the methodology. Boskin et al. (1996) consider the hedonic regression to be the most promising approach to control for quality changes, whereas Schultze and Mackie (2002) recommend following a cautionary approach and emphasize the need for further research. Moulton (2001) reports that approximately 18% of the US GDP final expenditures are deflated using price indexes that use hedonic methods and this share is expected to rise.

Price comparison of ‘like with like’ is an essential requirement in the construction of price indexes. This requirement raises problems in many markets, including super-market product, electronic and housing markets. For example, electronic products, such as computers, are subject to rapid technological innovation, leading to quick product turnovers and short life cycles. $P_{v,t}$ is observed if the model v was sold in period t , but $P_{v,t+1}$ is unobserved if the model exited out of the market before the beginning of period $t + 1$. In this case, the price relative $P_{v,t+1}/P_{v,t}$ is unobserved. In housing markets, every house is somewhat different and the same house is unlikely to be sold in two adjacent periods. This implies that if v refers to a particular house, either $P_{v,t}$ or $P_{v,t+1}$ is unobserved. Hedonic regressions impute these unobserved prices, and let us estimate the price relatives of unmatched, i.e. new and disappearing, items.

Price indexes calculated from only the matched varieties, i.e. the varieties that are sold in both t and $t + 1$, and leaving out the new and disappearing varieties, suffer from sample selection bias (Boskin et al., 1996; Pakes, 2003; Benkard and Bajari, 2005; Silver and Heravi, 2005). For example, Silver and Heravi, using comprehensive scanner data of five products (washing machines, dishwashers, television sets, cameras and vacuum

cleaners), show that the sample degradation is substantial even in a short period of one year. The percentage of models of these five products that disappeared within one year ranged between 29% and 47%, and the decline in the sales value ranged between 5% and 18%. Pakes (2003) and Benkard and Bajari (2005) find the sample degradation to be more severe for personal computers—by 85% and 90%, respectively, in a year—leaving ample room for sample selection bias.

It is a common concern that hedonic regressions, largely because of the unavailability of data, omit relevant characteristics. Some of these characteristics may be correlated with the included characteristics, leading to biased and inconsistent estimates.¹ Over the years, many authors have explicitly or implicitly acknowledged the omitted variable bias problem in hedonic regressions, including Court (1939), Griliches (1961), Pakes (2003), Hulten (2003), Benkard and Bajari (2005), Triplett (2006), and Hill and Melser (2008). However, there has been little investigation of the omitted variable problem specific to hedonic regressions, and particularly to hedonic imputation methods, beyond a general understanding of the omitted variable problem in regression analysis.²

The study is based on the premise that omitted variable bias is a fact in hedonic regressions. The study focusses on hedonic imputation methods, where separate hedonic regressions are hypothesized for each period of price comparison. Typically, and in this study, prices are compared between two adjacent periods. Between two such periods, the behavior of some characteristics may remain stable, while the behavior of other characteristics may be unstable. These two sets of characteristics, stable and unstable, may be treated differently in the omitted variable analysis of hedonic imputation methods. The conjecture of this study is that it is the unstable characteristics that drive price change, and therefore they should be included in the model in order to attain consistency of the estimated price relatives. The stable characteristics, though

¹The minimal requirement of an estimator is ‘consistency’. For practical purposes, inconsistency can be viewed as being the same as bias.

²There may be other sources of bias including incorrect specification of the functional forms and measurement errors in prices. This paper addresses only the omitted variable bias problem.

important in each period, do not drive price change and therefore may be excluded, as far as the consistency of the estimated price relatives is concerned. That is, under certain stability conditions the omitted variable bias corresponding to each regression may be canceled out. This study provides detailed expression of such stability conditions for, and evaluates the performance of, both the level and log hedonic models in the single and double imputation methods.

This study analyzes the omitted variable bias generated by each hedonic regression and then extends the analysis to hedonic imputation methods. Let us suppose that the conjecture of this paper is correct, i.e. the relevant characteristics that are stable between the periods can be omitted without having any effect on the consistency of the estimated price relatives. This may have an important implication with regard to the application of hedonic methods in order to construct quality adjusted indexes. For example, in the used car market there are many characteristics that are relevant to consumers in each period, including the make and model, age of car, engine size, odometer reading, color and fuel efficiency. However, there may be only a few characteristics that drive price change across two consecutive periods (such as fuel efficiency during a period of volatile oil prices), while the implicit value of other characteristics remains stable. This implies that data compilers may focus on a few select characteristics and, as a result, reduce cost.³

The importance of obtaining consistent price relatives in order to calculate indexes should be emphasized. Price relatives, including estimated price relatives, are the building blocks for the construction of indexes. It can be shown that elementary indexes (such as the Carli, Dutot and Jevons indexes) constructed on consistent price relatives are themselves consistent. Let us suppose that \hat{p}_v are consistent price relatives of their corresponding parameters θ_v , $\forall v = 1, \dots, V$. The estimated Jevons index, comparing prices between the period t and $t + 1$, is the geometric mean of all the price relatives \hat{p}_v : $\widehat{P}_{t,t+1} = \prod_{v=1}^V (\hat{p}_v)^{1/V}$. Using a property of probability limits (*plim*)—that if $g(\cdot)$ is

³Benkard and Bajari (2005) study the omitted variable bias problem in hedonic imputation methods. The focus of their paper, different to this paper, is on outlining an estimation methodology using factor analysis.

a continuous function, then $plim[g(x)] = g[plim(x)]$ —the following is obtained:

$$plim\left(\widehat{P}_{t,t+1}\right) = plim\left(\prod_{v=1}^V (\hat{p}_v)^{1/V}\right) = \prod_{v=1}^V [plim(\hat{p}_v)]^{1/V} = \prod_{v=1}^V (\theta_v)^{1/V}$$

that is, the estimated Jevons index is a consistent estimate of the true Jevons index. However, if the price relatives are inconsistent, and if there is no justification that biases tend to cancel each other out, the elementary indexes, and any other price indexes, including superlative indexes such as the Fisher and Törnqvist indexes, which are weighted averages of individual price relatives, are also inconsistent.

This paper is organized as follows. The next section introduces hedonic imputation methods and presents some observations on recent literature that compares and contrasts different hedonic methods. In section 3, the hedonic models are specified for the log of prices and the coefficient stability condition required to attain consistency of estimated price relatives is derived (a concise derivation for hedonic models specified on price levels is provided in appendix A). Section 4 shows results of Monte Carlo simulations which are conducted in order to check whether the stability condition derived analytically for large samples also holds for small samples. Section 5 discusses the implications of the findings with regard to their application. Conclusions are drawn in section 6.

2. Time Dummy, Single and Double Imputation Hedonic Methods

The dominant hedonic regression methods are the time-dummy hedonic method and the hedonic imputation method. In the time-dummy method, prices are hypothesized as a function of time dummies and the characteristics variables. The characteristics variables play the role of ‘controlling variables’ in the regression, and the coefficients of time dummies reflect the average price change between the periods after holding

the characteristics constant. If a relevant characteristic is omitted that is correlated with any of the included characteristics or the time dummies, then the estimated time-dummy coefficients are biased and inconsistent.

In the hedonic imputation method, separate regressions are specified for each period. The method essentially involves predictions of the left hand side variable (price or log-price) of the hedonic regressions. The varieties of goods sold in, say, period 0 are predicted for period 1 by putting the estimated characteristics coefficients of period 1 into the value of the characteristics of period 0. Thus, if $P_{v,t+1}$ refers to a new variety v that appeared in the market at period $t + 1$, then the hedonic imputation method imputes $P_{v,t}$, i.e. the price of variety v before it appeared in the market. Let us call this imputed price $\hat{P}_{v,t}$. Similarly, $\hat{P}_{v,t+1}$ is estimated from a hedonic regression when variety v disappeared from the market at period t . Once all the varieties are ‘matched’ in this way, standard price index formulas can be used.

The hedonic imputation method offers two approaches—single imputation or double imputation approaches. In the single imputation approach, only the unobserved price is imputed. The single imputation price relative is $P_{v,t+1}/\hat{P}_{v,t}$ or $\hat{P}_{v,t+1}/P_{v,t}$, depending on whether item v is a new or disappearing item. In the double imputation approach, on the other hand, both the observed and unobserved prices are imputed, and the estimated price relative is $\hat{P}_{v,t+1}/\hat{P}_{v,t}$. The imputed prices such as $\hat{P}_{v,t}$ and $\hat{P}_{v,t+1}$ are biased and inconsistent due to omission of characteristics. But the question is whether the price relatives $P_{v,t+1}/\hat{P}_{v,t}$, $\hat{P}_{v,t+1}/P_{v,t}$ or $\hat{P}_{v,t+1}/\hat{P}_{v,t}$ are also biased and inconsistent or can they be unbiased and consistent under particular conditions?

In the existing literature, various authors have discussed which of the single or double imputation methods is more appropriate. Triplett (2006) favors the minimum use of imputation on the grounds of minimizing estimation variance unless there is reason to believe that omitted characteristics have not changed between the comparison periods. Hill and Melser (2008) suggest that the double imputation method is preferable because of the potential of omitted variable biases being canceled out, however, they do not provide any formal analysis in support of their argument. Pakes (2003) does

not find any difference between the calculated indexes for personal computers obtained from the single and double imputation methods and, therefore, chooses to report the indexes using the single imputation price relatives. On the other hand, van Mulligen (2003) finds a systematic difference in indexes for personal computers, notebooks and servers—the double imputation index lies in between the single imputation index and the matched model index.

The current study, conducting a formal analysis on the estimators of imputation methods, finds that the single imputation method, similar to the time-dummy method, produces inconsistent estimates, whereas the double imputation method may produce consistent estimates under some stability conditions related to omitted characteristics and depending on whether the hedonic models are estimated for the log or level of prices. Consistency may be achieved even when each of the hedonic regressions produces inconsistent estimates, and at the same time by being less demanding on data. Because of the high prevalence of matched items between two adjacent periods in most markets, the study shows that it is possible to evaluate the performance of alternative models in terms of the magnitude of omitted variable bias.

3. Hedonic Models of Log Prices

3.1 Estimation of single and double imputation price relatives from the correctly specified models

This section begins with setting the parameters of interest and estimating the single and double imputation price relatives from the correctly specified log hedonic models corresponding to periods 0 and 1. The following two equations specify the hedonic

regressions for periods 0 and 1, respectively:

$$\ln p_i^0 = \sum_{k=1}^K \beta_k z_{k,i}^0 + \epsilon_i^0 \quad \forall i = 1, \dots, I \quad (1)$$

$$\ln p_v^1 = \sum_{k=1}^K \delta_k z_{k,v}^1 + \epsilon_v^1 \quad \forall v = 1, \dots, V \quad (2)$$

In equation 1, $\ln p_i^0$ denotes the log of price of item i in period 0, $z_{k,i}^0$ refers to the value of characteristic k for item i in period 0, and ϵ_i^0 is the error term assumed to be *i.i.d.* with zero mean and constant variance. In equation 2, the notations have similar interpretations for item v in period 1. The first characteristics in both the equations, z_1^0 and z_1^1 , take the value of 1 for all observations, indicating that they refer to the intercept terms in the equations. This way of denoting the intercept terms eases the use of notations in later sections. The parameters of interest for item i in period 0 and item v in period 1 are the following, respectively (*plim* refers to probability limit):

$$plim (p_i^0 | z_1^0, z_2^0, \dots, z_k^0) = exp \left(\sum_{k=1}^K \beta_k z_{k,i}^0 \right) = exp (\theta_i^0) \quad (3)$$

$$plim (p_v^1 | z_1^1, z_2^1, \dots, z_k^1) = exp \left(\sum_{k=1}^K \delta_k z_{k,v}^1 \right) = exp (\theta_v^1) \quad (4)$$

Now let us suppose that item i and item v are the same item. Then, from equations 3 and 4, the following is obtained:

$$plim \left(\frac{p_i^1}{p_i^0} \right) = \frac{exp(\theta_i^1)}{exp(\theta_i^0)} = exp(\Theta_i) \quad (5)$$

This sets the parameter of interest for the price relative of item i to be $exp(\Theta_i)$. If any other alternative price relatives of item i converge to $exp(\Theta_i)$, then these price relatives are consistent; otherwise they are asymptotically biased.

Let us now suppose that equations 1 and 2 satisfy the classical linear regression model assumptions, including that each of z_k^t is uncorrelated with disturbance ϵ^t for

$t=0,1$. The ordinary least squares method (OLS) provides consistent estimates of the parameters in equations 1 and 2. That is, if $\widehat{\beta}_k$ are estimates of β_k , and $\widehat{\delta}_k$ are estimates of δ_k , $plim(\widehat{\beta}_k) = \beta_k$ and $plim(\widehat{\delta}_k) = \delta_k$, $\forall k = 1, \dots, K$. The predicted prices obtained from both equations are also consistent.⁴ That is, if \widehat{p}_i^0 refers to the predicted price of item i in period 0, then the probability limit of \widehat{p}_i^0 is:⁵

$$plim \left(\widehat{p}_i^0 | z_1^0, z_2^0, \dots, z_k^0 \right) = plim \left(\exp \left(\sum_{k=1}^K \widehat{\beta}_k z_{k,i}^0 \right) \right) = \exp(\theta_i^0) \quad (6)$$

Turning now to period 1, if \widehat{p}_v^1 is the predicted price of item v , then the probability limit of \widehat{p}_v^1 is:

$$plim \left(\widehat{p}_v^1 | z_1^1, z_2^1, \dots, z_k^1 \right) = plim \left(\exp \left(\sum_{k=1}^K \widehat{\delta}_k z_{k,v}^1 \right) \right) = \exp(\theta_v^1) \quad (7)$$

Now hedonic imputation methods are applied in order to obtain the price relative between periods 0 and 1 for item i sold in period 0 (hence, item i is a disappearing item). The single imputation price relative is as follows:

$$\frac{\widehat{p}_i^1}{\widehat{p}_i^0} = \frac{\exp \left(\sum_{k=1}^K \widehat{\delta}_k z_{k,i}^0 \right)}{\exp \left(\sum_{k=1}^K \beta_k z_{k,i}^0 + \epsilon_i^0 \right)} = \prod_{k=1}^K \exp \left(\widehat{\delta}_k - \beta_k \right) z_{k,i}^0 \times \exp(-\epsilon_i^0) \quad (8)$$

where \widehat{p}_i^1 in the numerator is obtained by multiplying the estimated coefficients of the characteristics of period 1 with the value of the corresponding characteristics in period 0. The regression function specified in equation 1 is substituted for \widehat{p}_i^0 in the denominator. In the double imputation method, \widehat{p}_i^0 is used instead of p_i^0 . The double imputation price relative for the same item i is obtained as follows:

⁴This can be shown by using the rules of probability limits. The rules are: (1) if $g(x_n)$ is a continuous function that is not a function of n , $plim(g(x_n)) = g(plim(x_n))$; and (2) if x_n and y_n are random variables with $plim(x_n) = c$ and $plim(y_n) = d$, then $plim(x_n \pm y_n) = c \pm d$, $plim(x_n \times y_n) = c \times d$ and $plim(x_n \div y_n) = c \div d$ (for $d \neq 0$). For other rules of probability limits, see, for example, appendix C, Wooldridge (2009).

⁵Goldberger (1968) shows that even if $\widehat{\beta}$ is an unbiased estimator of β , because of taking a non-linear transformation, $\exp(\widehat{\beta})$ is a biased estimator of $\exp(\beta)$. To correct for the bias, Kennedy (1981) suggests the use of $[\exp(\widehat{\beta} + 0.5\widehat{\sigma}^2)]$ for the estimator of $\exp(\beta)$, where $\widehat{\sigma}^2$ is an estimate of the variance of $\widehat{\beta}$. However, in this paper the correction factor for this bias is not incorporated in the derivation. This does not have any impact on the results of the paper.

$$\frac{\widehat{p}_i^1}{\widehat{p}_i^0} = \frac{\exp\left(\sum_{k=1}^K \widehat{\delta}_k z_{k,i}^0\right)}{\exp\left(\sum_{k=1}^K \widehat{\beta}_k z_{k,i}^0\right)} = \prod_{k=1}^K \exp\left(\widehat{\delta}_k - \widehat{\beta}_k\right) z_{k,i}^0 \quad (9)$$

By using the assumption of equation 1 that ϵ_i^0 is an *i.i.d* with a zero mean, $plim(-\epsilon_i^0) = 0$. Hence, the *plim* of the single imputation price relative:

$$plim\left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0}\right) = \prod_{k=1}^K \exp\left[plim(\widehat{\delta}_k) - plim(\beta_k)\right] z_{k,i}^0$$

Substituting equations 3 and 7, the following is obtained:

$$plim\left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0}\right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} = \exp(\Theta_i) \quad (10)$$

that is, the single imputation price relatives estimated for disappearing items using the imputed prices from equation 2 are consistent estimates of the true price relatives. Furthermore, it can be shown that for the double imputation price relatives:

$$plim\left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0}\right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} = \exp(\Theta_i) \quad (11)$$

The above derivation indicates that both \widehat{p}_i^1/p_i^0 and $\widehat{p}_i^1/\widehat{p}_i^0$ provide the correct measure of the price change of item i between periods 0 and 1 in the sense that they converge to the true price relative, $\exp(\Theta_i)$. This implies that if there is no potential for omission of characteristics, one can use either the single or the double imputation method. The next section shows that the prevalence of omitted characteristics shifts the preference towards the double imputation method.

3.2 Coefficient stability conditions required to attain consistency in the imputation methods

Let $\mathbf{Z} = [\mathbf{Z}_a \ \mathbf{Z}_b] = [\mathbf{z}_1 \ \mathbf{z}_2 \ \dots \ \mathbf{z}_s; \ \mathbf{z}_{s+1} \ \dots \ \mathbf{z}_K]$, where \mathbf{Z} includes all the characteristics specified in equations 1 and 2. Let us now suppose that \mathbf{Z}_a contains the set

of observed characteristics and \mathbf{Z}_b contains the set of unobserved characteristics. Thus, the log of prices is now estimated on the characteristics z_1, z_2, \dots, z_s for both periods 0 and 1. A further assumption is made that each of the unobserved characteristics is correlated with the set of observed characteristics. This implies that the models estimated on the reduced set of characteristics suffer from the omitted variable problem, leading to biased and inconsistent estimates of the coefficients and predicted prices.⁶

Now, using the ordinary least squares (OLS) method, the following predicted log price for item i is obtained:

$$\widetilde{\ln p}_i^0 = \sum_{j=1}^s \tilde{\xi}_j z_{j,i}^0 \quad (12)$$

Let $\gamma_{j,k}^0 = E(z_k^0)$ for $j = 1$ (i.e. for the intercept term) and $\forall k = s + 1, \dots, K$, and $\gamma_{j,k}^0 = Cov(z_j^0, z_k^0)/Var(z_j^0)$, $\forall j = 2, \dots, s$ and $\forall k = s + 1, \dots, K$. The probability limit of $\widetilde{\ln p}_i^0$ in equation 12 is taken, and $E(u) = 0$ and $Cov(z_j^0, u) = 0$, $\forall j = 1, \dots, s$ are set to get the following:

$$plim(\widetilde{\ln p}_i^0) = \sum_{j=1}^s \beta_j z_{j,i}^0 + \sum_{j=1}^s \sum_{k=s+1}^K \gamma_{j,k}^0 \beta_k z_{j,i}^0 \quad (13)$$

Then, $\sum_{k=s+1}^K \beta_k z_{k,i}^0$ is added to the first term and subtracted from the second term of the right-hand side of the equation to get:

$$\begin{aligned} plim(\widetilde{\ln p}_i^0) &= \sum_{j=1}^K \beta_j z_{j,i}^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \\ &= \theta_i^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \end{aligned} \quad (14)$$

Consistency requires that the term $\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)$ equals zero. This term is zero if $\forall k = s + 1, \dots, K$ either $\beta_k = 0$ or $\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 = 0$.

⁶For a detailed derivation of omitted variable bias in a multi-variable context, see, for example, Johnston (1984).

But $\beta_k \neq 0$ since by assumption z_k^0 is a relevant characteristic. The other term $\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 = 0$, if for item i , z_k^0 is a perfect linear combination of the included characteristics, i.e. $z_{k,i}^0 = \gamma_{1,k} z_{1,i}^0 + \gamma_{2,k} z_{2,i}^0 + \dots + \gamma_{s,k} z_{s,i}^0$. This may happen for an item for a given z_k^0 , but is unlikely to happen for $\forall k = s+1, \dots, K$. Now let us consider all items, $i = 1, \dots, I$. For all i , $\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0$ implies that z_k^0 is a perfect linear combination of the included characteristics. This is a redundant condition because if a characteristic is a perfect linear combination of the other characteristics then it would not have entered in the model specified in equation 1. There is another possibility that since the term $\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)$ involves summation, and there are positive and negative numbers, these numbers somehow may cancel each other out. However, there is no justification for this to happen. Hence, it can be concluded that:

$$plim(\widetilde{lnp}_i^0) \neq \theta_i^0$$

that is, \widetilde{lnp}_i^0 is an inconsistent estimator of θ_i^0 .

For period 1, let us denote the OLS estimates of the coefficient of the z_j^1 characteristic by $\widetilde{\eta}_j$. The imputed price of item i with the values of the characteristics set $z_1^0, z_2^0, \dots, z_j^0$ is $\widetilde{lnp}_i^1 = \sum_{j=1}^s \widetilde{\eta}_j z_{j,i}^0$. Note that item i is sold in period 0, not in period 1. Let $\gamma_{j,k}^1 = E(z_k^1)$ for $j = 1$ and $\forall k = 1, \dots, K$, and $\gamma_{j,k}^1 = Cov(z_j^1, z_k^1) / Var(z_j^1)$, $\forall j = 2, \dots, s$ and $\forall k = 1, \dots, K$. Following the similar operations shown in equations 13 and 14, it can be shown that:

$$plim(\widetilde{lnp}_i^1) = \theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right), \quad (15)$$

where, following on from the explanation provided for equation 14, it can be shown that the term $\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \neq 0$. This is mainly because $\delta_k \neq 0$ and z_k^0 is not a perfect linear combination of the set of included characteristics, $\forall k = s+1, \dots, K$. Hence:

$$plim(\widetilde{lnp}_i^1) \neq \theta_i^1$$

that is, $\widetilde{\ln p}_i^1$ is an inconsistent estimator of θ_i^1 .

The single and double imputation price relatives estimated from models with reduced sets of characteristics are \widetilde{p}_i^1/p_i^0 and $\widetilde{p}_i^1/\widetilde{p}_i^0$, respectively. In order to derive the conditions required for the price relatives to be consistent with respect to $\exp(\Theta_i)$, their probability limits are taken. The following is the derivation for the single imputation method:

$$plim \left(\frac{\widetilde{p}_i^1}{p_i^0} \right) = plim \left[\frac{\exp(\widetilde{\ln p}_i^1)}{\exp(\ln p_i^0)} \right] = \frac{\exp \left[plim(\widetilde{\ln p}_i^1) \right]}{\exp \left[plim(\ln p_i^0) \right]}$$

By substituting equation 3 in the denominator and equation 15 in the numerator, and setting $\exp(\theta_i^1)/\exp(\theta_i^0) = \exp(\Theta_i)$ (see equation 5), the following is obtained:

$$\begin{aligned} plim \left(\frac{\widetilde{p}_i^1}{p_i^0} \right) &= \frac{\exp \left[\theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right]}{\exp \left[\theta_i^0 \right]} \\ &= \exp(\Theta_i) \times \exp \left(\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right) \end{aligned} \quad (16)$$

Now the term $\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \neq 0$. The explanation is the same as is provided in equations 14 and 15. Hence:

$$plim \left(\frac{\widetilde{p}_i^1}{p_i^0} \right) \neq \exp(\Theta_i)$$

i.e. the single imputation method estimates inconsistent price relatives.

The bias is given by $\exp \left(\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right)$. The direction and magnitude of the bias depend on a number of factors, including some unknown values (for example, $\gamma_{j,k}^1$). Hence, in practice it is impossible to find the direction and magnitude of bias corresponding to each of the price relatives. Moreover, Silver and Heravi (2005), Haan (2007), and Melser and Syed (2008) argue that the pricing patterns be-

tween new and disappearing items may vary systematically. This difference in pricing patterns may lead to a systematic difference in the estimated residuals obtained from hedonic regressions corresponding to new and disappearing items, implying that the biases do not offset each other. It should also be noted that the bias depends on the parameters of period 1 only, i.e. the period for which the price of item i is imputed, implying that the expressions for bias vary between the single and double imputation methods.

Now, in order to derive the conditions for consistency for the double imputation method, the probability limit of the double imputation price relative, $\tilde{p}_i^1/\tilde{p}_i^0$, is taken:

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = plim \left[\prod_{j=1}^s \exp \left(\tilde{\delta}_j - \tilde{\beta}_j \right) z_{j,i}^0 \right]$$

Using the rules of probability limits mentioned earlier, the following is obtained:

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = \frac{plim \left[\prod_{j=1}^s \exp \left(\tilde{\delta}_j z_{j,i}^0 \right) \right]}{plim \left[\prod_{j=1}^s \exp \left(\tilde{\beta}_j z_{j,i}^0 \right) \right]} = \frac{\exp \left[plim \left(\sum_{j=1}^s \tilde{\delta}_j z_{j,i}^0 \right) \right]}{\exp \left[plim \left(\sum_{j=1}^s \tilde{\beta}_j z_{j,i}^0 \right) \right]}$$

By substituting equations 14 and 15, it can be shown that:

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} \times \frac{\exp \left[\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right]}{\exp \left[\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \right]} \quad (17)$$

Hence, it can be seen from equation 17 that:

If

$$\frac{\exp \left[\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \right]}{\exp \left[\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \right]} = 1 \quad (18)$$

or

$$\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) = \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \quad (19)$$

then

$$plim \left(\frac{\tilde{p}_i^1}{\tilde{p}_i^0} \right) = \frac{\exp(\theta_i^1)}{\exp(\theta_i^0)} = \exp(\Theta_i)$$

The magnitude of the bias is given by the difference between between 1 and the left-hand side of equation 18. If the difference is positive, then the model overestimates P_i^1/P_i^0 , and if the difference is negative, the model underestimates P_i^1/P_i^0 . In the next section, a formula for the empirical estimates of the magnitude of bias is derived.

Equation 19 is referred to as the stability condition because consistency requires the stability of two sets of parameters. First, the regression coefficients of the unobserved characteristics in the hedonic equations are required to be the same across periods. If they are not the same across periods, then this may bias the hedonic price indexes. In appendix A, similar derivations are undertaken for the level hedonic models. It is shown that the exact expressions of the stability conditions or the sources of bias differ between the log and the level hedonic models (compare equations 19 and A.5).

The second set of parameters is related to the regression coefficients obtained when each of the unobserved characteristics is regressed on each of the observed characteristics. These regression coefficients are required to be the same across periods, i.e. $\gamma_{j,k}^0 = \gamma_{j,k}^1, \forall j = 1, \dots, s$ and $\forall k = s+1, \dots, K$. The second set of regression coefficients reflect how each of the excluded characteristics is configured with the set of included characteristics.⁷ In the next section, a technique is derived to empirically test whether equation 19 holds. This derivation will also provide a better intuitive understanding of what equation 19 implies.

⁷Diewert et al. (2008), while comparing the time-dummy and hedonic imputation methods, find that the change in the configuration of characteristics between the comparison periods is one of the factors that determine the difference in the indexes obtained from the time-dummy and the hedonic imputation methods.

3.3 Using matched items to estimate the magnitude of omitted variable bias

Let us now denote the items that were sold in both periods, the matched items, by $m = 1, \dots, M$. For these items, both the price relatives—ratios of observed prices and double imputation price relatives—are available. This enables the evaluation of the performance of the estimated hedonic models—the closer the estimated price relatives to the observed price relatives, the lower is the omitted variable bias contaminating the estimated price relatives.

It should be noted that the matched items dominate the total number of items sold between two adjacent periods.⁸ For the purpose of constructing price indexes, prices are usually compared between two adjacent months or between two adjacent quarters. Benkard and Bajari (2005) report that more than 90% of personal computer models observed in one month are also observed in the next month, though they argue at the same time that it is impossible to obtain a reliable measure of price comparison using the matched models because more than 90% of the models drop out within a period of one year. Similarly, from table 2 of Silver and Heravi (2005), it can be calculated that typically more than 95% of models are matched models between two adjacent months.⁹

Following from equation 12, it is possible to write the equation for period 0 as follows:

$$\ln p_m^0 = \widetilde{\ln p_m^0} + \widetilde{e_m^0} \quad \forall m = 1, \dots, M \quad (20)$$

where $\widetilde{e_m^0}$ is the estimated error when $\ln p_m^0$ is estimated on the reduced set of characteristics. By rearranging terms in equation 20 and taking the probability limits of both sides of the equation, the following is obtained:

⁸An exception is the housing market.

⁹It should be emphasized that although the majority of items are matched between two consecutive months, this does not undermine the matching problem caused by new and disappearing items. This is because, as the evidence shows, the problem becomes severe—both in terms of the turnover rates and the resulting bias in indexes—because of the cumulative effect over a period of time.

$$plim \left(\widetilde{e}_m^0 \right) = plim \left(lnp_m^0 \right) - plim \left(\widetilde{lnp}_m^0 \right) \quad \forall m = 1, \dots, M$$

By substituting $plim \left(lnp_m^0 \right) = \theta_m^0$ (see equation 3) and equation 14 for $plim \left(\widetilde{lnp}_m^0 \right)$, the following is obtained:

$$plim \left(\widetilde{e}_m^0 \right) = - \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,m}^0 - z_{k,m}^0 \right) \quad \forall m = 1, \dots, M \quad (21)$$

Similarly, for period 1, the following is obtained:

$$plim \left(\widetilde{e}_m^1 \right) = - \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,m}^1 - z_{k,m}^1 \right) \quad \forall m = 1, \dots, M \quad (22)$$

where \widetilde{e}_m^1 is the estimated error when lnp_m^1 is estimated on the reduced set of characteristics. Equations 4 and 15 are used for the derivation of equation 22.

Note that equations 21 and 22 refer to matched item m , implying that each of the characteristics of item m for the two periods are the same, i.e. $z_{k,m}^0 = z_{k,m}^1 = z_{k,m}$, $\forall k = 1, \dots, K$. This implies that equations 21 and 22 are simply the negative of the left- and right-hand sides of equation 19, respectively, when equation 19 corresponds to item m , $\forall m = 1, \dots, M$. Thus, the bias in equation 18 can be estimated by:

$$\frac{exp \left(\widetilde{e}_m^0 \right)}{exp \left(\widetilde{e}_m^1 \right)} - 1 \quad (23)$$

If $\left[exp \left(\widetilde{e}_m^0 \right) / exp \left(\widetilde{e}_m^1 \right) - 1 \right]$ is greater than 0, then the double imputation overestimates the true price relative of item m ; if the difference is less than 0, the double imputation underestimates the true price relative. Among alternative models with different sets of characteristics, the model that minimizes $\left| exp \left(\widetilde{e}_m^0 \right) / exp \left(\widetilde{e}_m^1 \right) - 1 \right|$ estimates price relatives that are contaminated the least by the omission of relevant characteristics. In a similar way, in the case of the single imputation method the bias in equation 16 can be estimated by:

$$\frac{1}{\exp\left(\widetilde{e}_m^1\right)} - 1 \quad (24)$$

3.4 Plausibility of equal correlations between the characteristics across periods

Let us assume that the regression coefficients between the observed and unobserved characteristics are equal between the two periods, i.e. $\gamma_{j,k}^0 = \gamma_{j,k}^1$, $\forall j = 1, \dots, s$ and $k = s+1, \dots, K$. Hence, in equation 19, $\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 = \sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 = w_{k,i}^0$. Thus, the stability condition in equation 19 reduces to:

$$\sum_{k=s+1}^K \delta_k w_{k,i}^0 = \sum_{k=s+1}^K \beta_k w_{k,i}^0$$

or

$$\delta_k = \beta_k, \quad \forall k = s+1, \dots, K \quad (25)$$

This implies that if the hedonic models are estimated for the log of prices, the consistency of the estimated price relatives requires that the coefficients of the same unobserved characteristic be equal between the comparison periods, provided that there has not been any significant change in the configuration of the characteristics of items between the comparison periods.

The assumption of the equality of the regression coefficients may be more plausible in one market over another. Let us consider the housing market. For the housing market this assumption would mean, say with regard to the relationship between the number of bedrooms and lot size, that for a given number of bedrooms, the average lot size is the same across periods. The housing market may be characterized by a high degree of sluggishness from the supply side, where transactions of old houses by far dominate the market. The matching problem arises because every house is different, but it may be reasonable to assume that the configuration between the characteristics remains stable in the housing market between two adjacent periods (for example, between two quarters

of a year). However, this assumption is unlikely to hold true in markets characterized by a rapid rate of technological innovation, such as the personal computer market. This is because technological advancement occurs at different rates with regard to different features of personal computers. Thus, quality change is reflected in the change in the configuration of characteristics, such as between the characteristics ‘CPU speed’ and ‘hard-disk size’.

Perhaps the housing market and the personal computer market are two extreme examples in terms of the expected nature of qualitative change between two adjacent periods. There are many other products where the matching problem arises in the construction of indexes and where the changes in the configuration of characteristics are not as pronounced as in the personal computer market, yet not as rigid as in the housing market. This may include markets for many electronic products other than computers, as well as the used car market and the markets for supermarket products such as laundry products, beverages, etc. In these markets, qualitative changes may occur in a few characteristics, and the stability condition indicates that in order to attain consistency these characteristics should be included in the model. Testing $\gamma_{j,k}^0 = \gamma_{j,k}^1$ in different markets may be an interesting area of further research.

4. Monte Carlo Simulations and Unbiased Price Relatives

The analytical results in the previous section are obtained for large samples (or as asymptotic properties) in order to estimate consistent price relatives. Now the question is: do the results also hold for finite or small samples, and equivalently apply in order to estimate unbiased price relatives?¹⁰ In this section, Monte Carlo simulations are conducted: first, to check whether equation 18 or 19 which produces consistent price

¹⁰Note that it is more desirable to estimate unbiased and consistent estimators than to estimate consistent but biased estimators. The analytical results focus on consistency instead of unbiasedness because the derivations involving ratios and multiplications of random variables are relatively straightforward, with probability limits rather than expectational operators.

relatives in the case of the double imputation method, produces unbiased price relatives at the same time; and, second, to estimate the bias specified in equations 23 and 24 for both small and large samples. The simulations are conducted with different specifications of the true and estimated hedonic models.

In period 0, the true model is obtained from the process $\ln p_i^{(0)} = 0.9 + 0.9x_{1,i} + 0.9x_{2,i} + \varepsilon_i$ and, in period 1, the true model is obtained from the process $\ln p_i^{(1)} = 0.9 + \kappa_1 x_{1,i} + \kappa_2 x_{2,i} + \varepsilon_i$. ε_i is drawn from a normal distribution, $\varepsilon_i \sim N(0, 1)$. The simulation exercise includes changing κ_1 and κ_2 to produce different scenarios. The only two characteristics, x_1 and x_2 , are drawn from a multivariate normal distribution with means π_1 and π_2 , variances ω_1^2 and ω_2^2 , respectively, and covariance ρ_{12} . For the matched items, which account for 90% of the total items in periods 0 and 1, the means and the variances are $[\pi_1, \pi_2, \omega_1^2, \omega_2^2, \rho_{12}] = [3, 3, 1, 1, 0.5]$.

The disappearing and new items account for 10% of the items in periods 0 and 1, respectively. The technological innovation occurs through changes in one of the characteristics, x_1 or x_2 , which may be reflected in the simulations through changes in the means and marginal valuations of x_1 or x_2 . For example, if the innovation occurs through x_1 , then the mean of x_1 for the disappearing items (π_1^d) may be expected to be lower than the mean of x_1 for the matched items (π_1), which in turn may be expected to be lower than the mean of x_1 for the new items (π_1^n). Examples of x_1 , in the case of computers, are the RAM or hard disk sizes of computers. Another way to look at innovation is through cost savings. If the cost saving occurs through x_1 then this may be reflected in the simulation through a fall in κ_1 between period 0 and 1.

The models in period 0 and 1 are estimated on an intercept and x_1 , hence the models omit x_2 . Whether x_1 is the stable or the unstable characteristic depends on whether π_1^d or π_1^n is different from 3 or whether κ_1 is different from 0.9. Let $[\kappa_1, \kappa_2] = [0.1, 0.9]$, i.e. the marginal valuation of x_1 falls in period 1, while the marginal valuation of x_2 remains the same. Furthermore, let $[\pi_1^d, \pi_1^n] = [3, 3]$ and $[\pi_2^d, \pi_2^n] = [3, 3]$, i.e. the mean values of the characteristics remain the same between the disappearing, matched and new items. In this example, the unstable characteristic is included and the stable char-

acteristic is excluded. Hence, according to equations 18 or 19, the double imputation method is expected to produce consistent price relatives.

In order to check for unbiasedness, 8 Monte Carlo simulations corresponding to $10 \times 3^{k-1}$ iterations for $k = 1, \dots, 8$, and each having a sample of 1000 observations, are conducted. The means of the estimates of equation 23 converge towards 0, implying that the double imputation price relatives are unbiased (see top part of columns 2 and 3 of table 1). In order to check for consistency, the sample size is increased by $10 \times 3^{k-1}$ observations for $k = 1, \dots, 8$, while the number of iterations is kept fixed at 10. The results show that the mean and variance converge to zero rapidly with the increase in sample size (see the bottom part of columns 2 and 3 in table 1). Thus the simulation results show that the double imputation price relatives are unbiased and consistent.¹¹

Insert table 1 here.

The same estimates corresponding to equation 24, i.e. for the single imputation method, are shown in columns 4 and 5 of table 1. The results show that there is no tendency for the mean to converge to 0 with the increase in the number of iterations, and the mean and variance to converge to 0 with the increase in the sample size. This provides evidence that the single imputation price relatives are biased and inconsistent.

Now let $[\pi_1^d, \pi_1^n] = [1, 5]$, $[\pi_2^d, \pi_2^n] = [3, 3]$ and $[\kappa_1, \kappa_2] = [0.1, 0.9]$. In this case, the technological progress occurs through the provision of an increasing number of x_1 for a given x_2 , and through cost savings in x_1 . Because of differing means, the variance of x_1 differs between periods 0 and 1. This leads the correlation coefficient to change between the two periods (i.e. $\gamma_{j,k}^0 \neq \gamma_{j,k}^1$ in equation 19). The simulation results provided in columns 3 and 4 of table 2 show no tendency for the estimates of the means of equation 23 to converge to zero, implying that the double imputation method estimates biased price relatives.¹² The same conclusion is drawn with regard to equation 24 for the single

¹¹If a large difference is created at the initial stage by setting a large difference between κ_2 and 0.9, between π_1 and π_2 and by increasing the variance of ε_i , the results with regard to convergence are the same.

¹²If the change in mean value had happened with x_2 , then the double imputation price relatives would have been consistent.

imputation method (see columns 5 and 6 of table 2). Similarly, if $[\kappa_1, \kappa_2] = [0.9, 0.1]$, i.e. the omitted characteristic sees a drop in the marginal valuation, both the single and double imputation price relatives are biased and inconsistent (see bottom part of table 2). However, importantly, the biases are found to be higher for the single imputation method than for the double imputation method.¹³

Insert table 2 here.

With regard to the time-dummy method, it can be easily shown that the bias generated by the time-dummy method is substantial even in a trivial case where the double imputation method produces unbiased and consistent price indexes. Let us suppose that all the items are matched between periods 0 and 1. Furthermore, let us suppose that $[\kappa_1, \kappa_2] = [0.9, 0.9]$. As before, the models in both periods are estimated on intercept and x_1 . It can be shown that the Jevons index constructed from taking the geometric mean of the estimated double imputation price relatives is 1, i.e. the double imputation method provides the correct measure of price change. On the contrary, the time-dummy index—which is equivalent to the Jevons index—provides a biased estimate of price change, where the magnitude of bias depends on the correlation between x_2 and the time dummy. In the above example, even if the correlation is as low as 0.1, the time-dummy index overestimates the price change by 9%.

5. Implications for Applications

There are a number of important implications of the results with regard to the application of hedonic methods in the construction of quality adjusted indexes. First, if the omitted variable bias is one of the main problems inhibiting the application of hedonic methods, the single imputation method does not offer a better solution than the time-dummy method. Both methods are equally demanding of the characteristics information that is required to produce unbiased and consistent estimates of price change. The double imputation method may produce unbiased and consistent estimates of price

¹³The results with regard to consistency are similar to that of unbiasedness.

change even in the presence of a large number of unobserved characteristics. Also, the simulation results show that even if the bias exists, the magnitude of bias is in general lower in the double imputation than in the single imputation price relatives.

Second, the data requirement shifts from ‘all price determining characteristics’ to ‘price determining characteristics that drive the price changes between the comparison periods’. If the log model is the preferred model, the data compilers may focus on collection of information on the characteristics whose contribution, both in terms of their marginal values to price and how they are configured with the other characteristics, is expected to change between the periods. This may, on the one hand, reduce the data requirement by a large amount while, on the other hand, make it possible to obtain estimates that are ‘near consistent’, if not consistent.

The third implication is related to the second implication. The idea of what constitutes an important variable differs between the single regression equation (hedonic or otherwise) and the hedonic double imputation method. In the case of single regression equations, an important variable is one that has a relatively large explanatory power, whereas in the case of the hedonic imputation method an important variable is one that has a large influence on the price change between the periods. In both cases, the important variables may coincide in many markets but not necessarily in all markets.

Fourth, the stability conditions on the unobserved characteristics, conditions which are required for unbiasedness and consistency, differ between the log and level models (the derivation for the level hedonic models is provided in appendix A). In general, while log models may be preferable where price changes are driven by a selected number of characteristics, level models may be appropriate where a large number of characteristics coefficients are expected to change by the same proportion. For example, the log models may be appropriate for electronic markets because technological advancement may be driven by a selected number of features in a particular electronic product, and the level models may be more appropriate for the housing market because price is driven by expectations of the future flow of income, where the expectations formed in two adjacent periods may be similar with regard to the different physical features of a

house.

Finally, in recent years a number of studies comparing the time-dummy and hedonic imputation methods have argued in favor of using the hedonic imputation method (see, for example, Silver and Heravi, 2007; Diewert et al., 2008). This is because the hedonic imputation method allows the characteristics coefficients to vary, whereas the time-dummy method constrains them to remain fixed between the periods. These studies analyze the observed characteristics only and do not draw any distinction between the single and double imputation methods. In contrast, the current study focusses on the unobserved characteristics and conducts a separate analysis on the single and double imputation methods based on the omitted variable bias generated by the unobserved characteristics. Although the methodological approach is different, the conclusion of this study accords with the recent literature which argues in favor of using hedonic imputation methods. Moreover, this study finds justification in favor of the double imputation method. This is because the double imputation method has the potential to address the omitted bias problem in hedonic regressions with the careful choice of limited but selected characteristics.

6. Conclusion

In the case of single hedonic regressions, the omitted variable bias problem is similar to the problem with regressions in general, i.e. if the relevant variables that are correlated with the included variables are omitted, the estimates are biased and inconsistent. However, this is not the case for the price relatives estimated from the double imputation method, which involves the ratio of two hedonic regressions, because of the potential of biases being canceled out. This study has taken a systematic look at the omitted variable problem of hedonic imputation methods by analyzing the asymptotic bias generated by each of the hedonic regressions. The study analyzed the log and level hedonic models for both the single and double imputation methods.

The study found that the single imputation method does not have any advantage

over the time-dummy method in terms of dealing with the omitted variable problem. If the hedonic regressions exclude any relevant price determining characteristics, both methods produce biased and inconsistent estimates. The double imputation method has a clear advantage in this case. This study found that double imputation price relatives can be unbiased and consistent even if the important price determining characteristics corresponding to each of the hedonic regressions are omitted. The requirement for attaining unbiasedness and consistency is that the contributions of the unobserved characteristics in price determination are stable between the periods. Depending on the products, this has the potential of reducing the data requirement, in terms of characteristics information, by a large extent.

This study has outlined a method to estimate the magnitude of bias in the single and double imputation price relatives for matched items. This method may indicate how the hedonic imputation method performs in predicting the price relatives of unmatched, i.e. new and disappearing, items. This estimated magnitude of bias may provide a benchmark summary measure that can be used to compare the performance of different hedonic imputation models applied to different data sets.

Appendix A: Hedonic models of price levels

Let us specify the models for periods 0 and 1 on price levels as follows:

$$p_i^0 = \sum_{k=1}^K \beta_k z_{k,i}^0 + \epsilon_i^0 \quad \forall i = 1, \dots, I \quad (\text{A.1})$$

$$p_v^1 = \sum_{k=1}^K \delta_k z_{k,v}^1 + \epsilon_v^1 \quad \forall v = 1, \dots, V \quad (\text{A.2})$$

where, similar to equations 1 and 2, $z_{k,i}^0$ and $z_{k,i}^1$ refer to the value of the characteristics k for item i in periods 0 and 1, respectively. The first characteristics, z_1^0 and z_1^1 , are the intercept terms of the equations. The error terms ϵ_i^0 and ϵ_v^1 are *i.i.d.* with zero mean and constant variance. Furthermore, the classical linear regression model assumptions

hold for models in equations A.1 and A.2.

Following similar steps to those shown in equations 6 to 11 for the log models, it can be shown that the single and double imputation price relatives estimated from equations A.1 and A.2 are consistent. That is:

$$plim \left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0} \right) = plim \left(\frac{\widehat{p}_i^1}{\widehat{p}_i^0} \right) = \frac{\theta_i^1}{\theta_i^0} = \Phi_i \quad (\text{A.3})$$

Let us now suppose that instead of estimating models on $z_1, \dots, z_s, z_{s+1}, \dots, z_K$ characteristics as specified in equations A.1 and A.2, the models are estimated on z_1, \dots, z_s . Let \widetilde{p}_i^0 and \widetilde{p}_v^1 be the predicted prices obtained from the estimated models for items i and v , respectively. Following steps similar to those in equations 13 and 14, it can be shown that $plim \left(\frac{\widetilde{p}_i^0}{\widetilde{p}_i^0} \right) = \theta_i^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right) \neq \theta_i^0$, and, following steps similar to those in equation 15, it can be shown for the imputed price of item i in period 1 that $plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right) = \theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right) \neq \theta_i^1$. Thus, the probability limit of the single imputation price relative for item i is:

$$plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right) \neq \frac{\theta_i^1}{\theta_i^0} = \Phi_i$$

that is, the single imputation price relative for item i is inconsistent.

Now to the double imputation price relatives. Substituting the expressions for $plim \left(\frac{\widetilde{p}_i^0}{\widetilde{p}_i^0} \right)$ and $plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right)$, the following is obtained:

$$plim \left(\frac{\widetilde{p}_i^1}{\widetilde{p}_i^0} \right) = \frac{\theta_i^1 + \sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right)}{\theta_i^0 + \sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)} \quad (\text{A.4})$$

Equation A.4 implies that if

$$\frac{\theta_i^1}{\theta_i^0} = \frac{\sum_{k=s+1}^K \delta_k \left(\sum_{j=1}^s \gamma_{j,k}^1 z_{j,i}^0 - z_{k,i}^0 \right)}{\sum_{k=s+1}^K \beta_k \left(\sum_{j=1}^s \gamma_{j,k}^0 z_{j,i}^0 - z_{k,i}^0 \right)} \quad (\text{A.5})$$

the double imputation price relatives estimated from level hedonic models are consistent.

Note that equation A.5 is the level hedonic model equivalent of equation 19 for the log

hedonic model.

Thus to evaluate the performance of alternative models, the following is estimated for the matched item m :

$$\left(\frac{p_m^1}{p_m^0}\right) - \left(\frac{\widetilde{e}_m^1}{\widetilde{e}_m^0}\right) \quad (\text{A.6})$$

The closer the value is towards zero, the lower is the omitted variable bias in the estimated price relative of item m .

If $\gamma_{j,k}^0 = \gamma_{j,k}^1, \forall j = 1, \dots, s$ and $k = s+1, \dots, K$ are set, then the stability condition in equation A.5 reduces to:

$$\frac{\sum_{j=1}^s \delta_j z_{j,i}}{\sum_{j=1}^s \beta_j z_{j,i}} = \frac{\sum_{k=s+1}^K \delta_k z_{k,i}}{\sum_{k=s+1}^K \beta_k z_{k,i}} \quad (\text{A.7})$$

Thus, the stability condition for the level models specified in equation A.7 differs from that for the log models specified in equation 25. This implies that hypotheses about the behavior of unobserved characteristics across periods have implications in the choice between the level and the log models in the double imputation hedonic method.

References

- Benkard, C.L., Bajari, P. (2005): Hedonic Price Indexes with Unobserved Product Characteristics, and Application to Personal Computers, *Journal of Business & Economic Statistics* 23(1), 61-75.
- Berndt, E.R., Rappaport, N.J. (2001): Price and Quality of Desktop and Mobile Personal Computers: A Quarter-Century Historical Overview, *American Economic Review* 91(2), 268-273.
- Boskin, M., Dullberger, E., Gordon, R., Griliches, Z., Jorgenson, D. (1996): Towards a More Accurate Measure of the Cost of Living, Advisory Commission to Study the Consumer Price Index, Final Report to the Senate Finance Committee, Washington, D.C.

- Court, A.T. (1939): Hedonic Price Indexes with Automotive Examples, in *The Dynamics of Automobile Demand*, General Motors Corporation, New York, 98-119.
- Diewert, W.E., Heravi, S., Silver, M. (2008): Hedonic Imputation versus Time Dummy Hedonic Indexes, NBER Working Paper 14018, National Bureau of Economic Research, Cambridge, <http://www.nber.org/papers/w14018>.
- Goldberger, A.S. (1968): The Interpretation and Estimation of Cobb-Douglas Functions, *Econometrica* 36(3/4), 464-472.
- Griliches, Z. (1961): Hedonic Price Indexes for Automobiles: An Econometric Analysis of Duality Change, in *The Price Statistics of the Federal Government*, General Series 73, Columbia University and National Bureau of Economic Research, 137-196.
- Haan, J.de (2007): Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Other Approaches, Paper presented at the 10th Ottawa Group Meeting, October 9-12, 2007, Ottawa, http://www.ottawagroup2007.ca/r004/pdf/ogo04_033_e.pdf.
- Hill, R.H., Melsner, D. (2008): Hedonic Imputation and the Price Index Problem: An Application to Housing, *Economics Inquiry* 46(4), 593-609.
- Hulten, C.R. (2003): Price Hedonics: A Critical Review, *FRBNY Economic Policy Review* 9, 5-15.
- Johnston, J. (1984): *Econometric Methods* (3rd ed.), New York: McGraw-Hill Book Company.
- Kennedy, P.E. (1981): Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations, *American Economic Review* 71(4), 801.
- Melsner, D., Syed, I. (2008): Pricing over the Product Life Cycle: An Empirical Analysis, Research Paper No. 2008 ECON 25, Australian School of Business, University of New South Wales, Sydney.

- Moulton, B.R. (2001): The Expanding Role of Hedonic Methods in the Official Statistics of the United States, U.S. Bureau of Economic Analysis, Washington, D.C, <http://www.bea.gov/about/pdf/expand3.pdf>.
- Pakes, A. (2003): A Reconsideration of Hedonic Price Indexes with an Application to PCs, *American Economic Review* 93(5), 1578-1596.
- Schultze, C.L., Mackie, C. (2002): *At What Price? Conceptualizing and Measuring the Cost-of-Living and Price Indexes*, Washington, D.C.: National Academy Press.
- Silver, M., Heravi, S. (2005): A Failure in the Measurement of Inflation: Results From a Hedonic and Matched Experiment Using Scanner Data, *Journal of Business & Economic Statistics* 23(3), 269-281.
- (2007): The Difference Between Hedonic Imputation Indexes and Time Dummy Hedonic Indexes, *Journal of Business & Economic Statistics* 25(2), 239-246.
- Tripllett, J.E. (2006): *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD Directorate for Science, Technology and Industry, Paris.
- van Mulligen, P.H. (2003): *Quality Aspects in Price Indices and International Comparisons: Application of the Hedonic Method*, Unpublished Ph.D Dissertation, University of Groningen.
- Waugh, F.V. (1928): Quality Factors Influencing Vegetable Prices, *Journal of Farm Economics* 10, 186-196.
- Wooldridge, J.M. (2009): *Introductory Econometrics: A Modern Approach*. Ohio: Thomson South-Western.

Table 1: Simulation Results for Double and Single Imputation Price Relatives when the Stable Characteristic is Omitted

Simulations	Double Imputation		Single Imputation	
	$\kappa_1 = 0.1, \kappa_2 = 0.9$		$\kappa_1 = 0.1, \kappa_2 = 0.9$	
	Mean	Variance	Mean	Variance
No of iterations ¹				
10	0.0042	0.0244	0.0123	1.0064
30	0.0072	0.0313	0.8390	3.3538
90	-0.0006	0.0281	0.4580	1.5262
270	0.0016	0.0246	1.0442	3.5721
810	0.0006	0.0253	1.2725	4.1336
2460	0.0005	0.0247	1.2531	3.7287
7290	0.0004	0.0251	1.1870	3.7190
21870	0.0002	0.0255	1.2032	4.1710
Sample sizes ²				
100	0.0534	0.0863	0.7052	2.2129
300	-0.0204	0.0548	0.4410	0.9403
900	-0.0031	0.0192	0.4976	1.3767
2700	0.0017	0.0148	4.5119	12.9150
8100	0.0028	0.0104	1.1284	4.6546
24600	-0.0001	0.0063	2.1059	3.4185
72900	0.0012	0.0021	0.4586	2.3750
218700	-0.0002	0.0016	1.2400	2.5616

Notes: 1. Sample size=1000; 2. No. of iterations=10

Table 2: Simulation Results for Double and Single Imputation Price Relatives when the Configuration of Characteristics Changes and the Unstable Characteristics are Omitted

Simulations	No. of Iterations	Double Imputation		Single Imputation	
		Mean	Variance	Mean	Variance
$\pi_1^d = 1, \pi_1^n = 5$ $\kappa_1 = 0.1, \kappa_2 = 0.9$	10	-0.0417	0.1137	0.8475	1.8530
	30	0.0027	0.1593	0.4219	1.2269
	90	0.0139	0.1837	0.4857	1.7682
	270	0.0174	0.1386	1.5814	6.1601
	810	0.0193	0.1556	1.1893	4.0624
	2460	-0.0035	0.1499	1.0130	4.3506
	7290	0.0029	0.1522	1.1450	4.4066
	21870	0.0043	0.1490	1.1796	4.2468
$\pi_1^d = 3, \pi_1^n = 3$ $\kappa_1 = 0.9, \kappa_2 = 0.1$	10	0.0852	0.3870	1.1917	2.1996
	30	-0.1728	0.7752	1.2507	2.5072
	90	0.2815	0.8591	0.3956	1.8394
	270	0.2448	0.8736	0.8178	2.3800
	810	0.2609	1.1236	0.7106	2.4986
	2460	0.2692	0.9780	0.5711	1.8508
	7290	0.2649	0.9805	0.6481	2.1202
	21870	0.2751	0.9913	0.6563	2.1727

Note: Sample size for all simulations is 1000.