Estimation Biases in Quality-Adjusted Hedonic Price Indices

Jan van Dalen [1]
Ben Bode

Rotterdam, October 1, 2004

[1]Rotterdam School of Management, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR, Rotterdam; e-mail: jdalen@fbk.eur.nl. The authors gratefully acknowledge useful comments by Mick Silver. The usual disclaimers apply.

Abstract

Log-linear hedonic models are widely used to construct quality-corrected price indices. It is well-known that least-squares estimation of these models yields biased estimates of the model parameters and of the expected value of the dependent price variable; see Goldberger (1968) and Teekens and Koerts (1972). The problem is cited by Diewert (2002), Triplett (2002) and Silver (2002) as an issue in price-index construction with reference to the amount of bias. However, many applied studies ignore the bias, either because it is considered non-substantial or because one is not aware of it.

This paper summarizes some of the consequences of the estimation bias associated with the use of log-linear models for the hedonic and matched-model price indices. After reviewing the background of the estimation bias, the impact of the estimation bias on various types of hedonic price indices is explained. Also, attention is paid to biases related with the hedonic imputation of missing prices. The consequences of ignoring the estimation bias are illustrated with an example of car price developments.

The matter is of particular interest for modern research on measuring quality-adjusted price developments, which often adopts log-linear hedonic specifications to analyze biases related with substitution, entry and exit of product varieties and aggregation of price information, but which tends to ignore biases related with the very instruments to examine these issues.

JEL classification: C23,C43, C51, C81, D12, E31, L15,L62, O47

1 Introduction

Least squares estimation of log-transformed multiplicative models leads to biased estimates of both the model parameters and the expected value of the dependent. This problem has been known at least since Goldberger (1968) and Teekens and Koerts (1972).[1] However, many applied studies ignore the bias, either because it is considered non-substantial or because one is not aware of it. The matter is of particular interest for current research on measuring quality-adjusted price developments, frequently adopting semi-log or double-log model specifications, which pays much attention to biases related with substitution, entry and exit of varieties, and aggregation of price information, but which tends to surpass biases related with the very instruments to examine these issues, namely model assumptions and related estimation problems.

This paper summarizes some of the bias problems related with the hedonic approach and the matched-model procedures. The biases are examined both at the level of model assumptions and on the level of estimated index values. Also, attention is paid to biases related with the imputation of missing prices. The biases tend to be particularly present in annually-estimated hedonic models, while price change estimates in pooled models seem to be limitedly biased. However, the estimation of these pooled models will be seen to be based on unsatisfactory assumptions, from an econometric point of view. The consequences of ignoring these biases are illustrated with an example of car price developments.

The paper is structured as follows. Section 2 explains the background of the problem introducing various technical concepts used in the sequel. Section 3 analyzes the biases associated with least-squares and maximum-likelihood estimation of two Laspeyres- and Paasche-like fully hedonic price indices. Specific attention is paid to the consequences of weighting. Also, an empirical example of the extent of the biases is provided. Section 4 examines some of the bias problems for matched model price indices, while section 5 comments on the bias for hedonic indices based on pooled estimation of the hedonic model. Section 6 reviews the main findings emphasizing that the bias associated with least-squares estimation of log-linear hedonic relationships is not innocent in general and that applied work in the area should at least indicate the magnitude of the biases resulting from this type of modelling.

[1] Related work can be found in, e.g. Aigner (1974), Evans and Shaban (1976), Kennedy (1981), Rukhin (1986), and Zellner (1971)

2 Background of the estimation problem

The estimation bias discussed stems from the statistical fact that the parameters (mean and variance) of a normally distributed random variable are not simply the natural log's of the corresponding parameters of a log-normally distributed random variable. As Berndt (1991, p.127) notes, an analysis of log prices is just not the same as an analysis of prices. The assumption of normality is not really necessary, as Goldberger (1968) notes, but greatly simplifies many derivations. This section briefly restates the known facts about the biased estimation of the parameters and expected value of log-transformed multiplicative models, and defines a general formulation to evaluate biases in specific situations examined in the next section.

Many price studies based on Lancaster (1966a,b)'s household production theory, adopt a linear relation between log prices on the one hand and product characteristics on the other hand in order to estimate quality corrected prices of products. The relationship is a consequence of the first order conditions associated with a two-stage optimization procedure, in which the utility derived from product services is maximized, and where the product services are the maximum attainable at given product characteristics.

Stochastic properties of the log-linear model

The log-relationship is either imposed a priori or selected on the basis of specification tests (Van Dalen and Bode, 2004). Let Y denote the price of a product variety, and $y = \ln Y$ its log-transformation. Moreover, let $\mathbf{x}$ be a $(K+1)$-vector with relevant product characteristics, which are non-stochastic and which may or may not have been suitably transformed by themselves. The $\mathbf{x}'$-vector has a 1 on the first position to cope with inhomogeneities in the price-characteristics specification or to reflect deviations from some reference product variety as in the pooled hedonic model. The commonly adopted log-linear price relationship may then be represented as:

$$y = \mathbf{x}'\boldsymbol{\beta} + \varepsilon \quad (1)$$

for arbitrary product varieties. The parameters to be estimated are summarized by the parameter vector $\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots, \beta_K)$, where β_0 represents the intercept and the β_k reflect the proportional price changes resulting from unit changes in the corresponding characteristics x_k, $\beta_k = (\partial Y/Y)/\partial x_k$ for $k = 1, \dots, K$.[2] The additive disturbance ε is assumed to be identically, independently normally distributed with mean

[2]It is tempting to interpret the β_k as the implicit prices of the characteristics of a product variety. This would indeed be the case in a linear price-characteristics relationship. The interpretation is not valid, however, in the log-linear model. If some measure of implicit prices is still desired, then $\partial Y/\partial x_k = \beta_k Y$ seems a suitable candidate.

μ and non-zero variance σ^2, $\varepsilon \sim n(\mu, \sigma^2)$. The model implies that the expected log-price for a given product variety is: $E(y|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}+\mu$, which is equal to the systematic part of (1) when the common assumption μ=0 is made.
The implications of the log-linear specification for the relation between the non-transformed, original price of a product variety Y and the associated product characteristics $\mathbf{x}$ can be analyzed by taking exponents of the log-price in (1):

$$Y = e^y = e^{\mathbf{x}'\boldsymbol{\beta}+\varepsilon} = e^{\mathbf{x}'\boldsymbol{\beta}}\nu \tag{2}$$

with $\nu = e^{\varepsilon}$, $\nu > 0$, the multiplicative disturbance term. Since ε is normally distributed with parameters μ and σ^2, ν is log-normally distributed with the same parameters μ and σ^2. Following these properties, the expected value and variance of the disturbance ν are equal to $E(\nu) = \exp\{\mu+\frac{1}{2}\sigma^2\}$ and $V(\nu) = \exp\{2\mu+\sigma^2\}(\exp\{\sigma^2\}-1)$, respectively. The expected price of a variety with characteristics $\mathbf{x}$ is read from (2) as: $E(Y|\mathbf{x}) = \exp\{\mathbf{x}'\boldsymbol{\beta}\}E(\nu)$, which is equal to the systematic part of (2) when $E(\nu)$ is assumed to be equal to 1, $E(\nu)$=1.
The stochastic properties of (1) and (2) make that the expected value of the price Y and that of the log-price y cannot at the same time be equal to the corresponding systematic parts of their model specifications. In other words, assuming that $E(\nu)$=1 in (2) is not consistent with assuming μ=0 in (1) except in the trivial situation when σ^2=0 (which is excluded by assumption). Suppose that the focus of analysis is on the non-transformed price Y described by (2). In this case, the requirement that the expected price of a product variety conforms with the systematic part of (2) implies that the expected value of the disturbance is equal to 1, $E(\nu) = \exp\{\mu + \frac{1}{2}\sigma^2\} = 1$, which in turn implies that the expected value of ε is equal to:

$$E(\varepsilon) = \mu = -\frac{1}{2}\sigma^2 \tag{3}$$

which is strictly negative for positive σ^2. The variance of ν becomes: $V(\nu) = \exp\{\sigma^2\}-1$. Alternatively, setting the expected value of the disturbance ε in (1) equal to $\mu = 0$, implies that the expected value of ν is larger than 1: $E(\nu) = \exp\{0 + \frac{1}{2}\sigma^2\} = \exp\{\frac{1}{2}\sigma^2\} > 1$. The inconsistency of the two requirements, $E(\varepsilon)$=0 and $E(\nu)$=1, together with the implication of $E(\nu)$=1 that $E(\varepsilon) = \mu = -\frac{1}{2}\sigma^2$ are at the core of the estimation bias elaborated below.[3]
Before discussing the estimation issue, it is convenient to quote the moment-generating function of the normally distributed disturbance term ε as: $M_{\varepsilon}(t) = E(e^{t\varepsilon}) =$

[3]Note that the discussion concentrates on the expected values of the disturbances in (1) and (2) and not on their variances. The reason for this is that no specific values for $V(\varepsilon) = \sigma^2$ and $V(\nu)$ have been imposed except that they are positive. The latter is not a restrictive assumption: if $V(\varepsilon) = \sigma^2 > 0$, then $V(\nu) = \exp\{\sigma^2\} - 1 > 0$, and the other way around.

$\exp\{t\mu + \frac{1}{2}t^2\sigma^2\}$, which in the case of $\mu = -\frac{1}{2}\sigma^2$ boils down to $M_\varepsilon(t) = \exp\{-\frac{1}{2}t\sigma^2 + \frac{1}{2}t^2\sigma^2\} = \exp\{-\frac{1}{2}\sigma^2 t(1-t)\}$ for real numbers t. The moment-generating function can be used, among other things, to easily calculate the mean value of ν as $E(\nu) = E(e^\varepsilon) = M_\varepsilon(1) = 1$. In the case of an arbitrary multivariate normal random variable $\mathbf{z}$, $\mathbf{z} \sim n(\boldsymbol{\delta}, \boldsymbol{\Delta})$, the moment-generating function of $\mathbf{z}$ is given as:

$$M_{\mathbf{z}}(\mathbf{t}) = E(e^{\mathbf{t}'\mathbf{z}}) = e^{\mathbf{t}'\boldsymbol{\delta} + \frac{1}{2}\mathbf{t}'\boldsymbol{\Delta}\mathbf{t}} \tag{4}$$

for real-valued vectors $\mathbf{t}$; all quantities appropriately dimensioned. The moment-generating function is frequently applied below to simplify calculations. For example, the variance of $e^{\mathbf{t}'\mathbf{z}}$, defined as $V(e^{\mathbf{t}'\mathbf{z}}) = E(e^{2\mathbf{t}'\mathbf{z}}) - E^2(e^{\mathbf{t}'\mathbf{z}})$, can be determined with the aid of (4) as:

$$\begin{aligned} V(e^{\mathbf{t}'\mathbf{z}}) &= M_{\mathbf{z}}(2\mathbf{t}) - M_{\mathbf{z}}^2(\mathbf{t}) \\ &= e^{2\mathbf{t}'\boldsymbol{\delta} + 2\mathbf{t}'\boldsymbol{\Delta}\mathbf{t}} - e^{2\mathbf{t}'\boldsymbol{\delta} + \mathbf{t}'\boldsymbol{\Delta}\mathbf{t}} \\ &= e^{2\mathbf{t}'\boldsymbol{\delta} + \mathbf{t}'\boldsymbol{\Delta}\mathbf{t}}(e^{\mathbf{t}'\boldsymbol{\Delta}\mathbf{t}} - 1) \end{aligned} \tag{5}$$

Properties of the least squares estimator of β

Much applied research in price index theory, determines the unknown parameters $\boldsymbol{\beta}$ in (2) by least squares estimation of (1) (incorrectly assuming $\mu = 0$) using a random sample of prices Y_i and characteristics $\mathbf{x}_i'$ of N product varieties $i = 1, \ldots, N$. The sample information is compactly summarized with a N-vector of prices $\mathbf{Y}$, a N-vector of log-prices $\mathbf{y}$ and a $N \times (K+1)$-matrix $\mathbf{X}$ of product characteristics with a N-vector $\boldsymbol{\iota}$ of 1's in the first column. The question for now is how the least squares estimator is affected by the stochastic properties of models (1) and (2).

The least squares estimator $\hat{\boldsymbol{\beta}}_{LS}$ of $\boldsymbol{\beta}$ in (1) is equal to:

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{LS} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \end{aligned} \tag{6}$$

where $\boldsymbol{\varepsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ according to (1). The expected value of the least squares estimator $\hat{\boldsymbol{\beta}}_{LS}$ is equal to $E(\hat{\boldsymbol{\beta}}_{LS}) = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon})$. Since the expected value of $\boldsymbol{\varepsilon}$ is equal to $E(\boldsymbol{\varepsilon}) = -\frac{1}{2}\sigma^2\boldsymbol{\iota}$ according to (3), the expected value of the least squares estimator $\hat{\boldsymbol{\beta}}_{LS}$ can be simplified to:

$$\begin{aligned} E(\hat{\boldsymbol{\beta}}_{LS}) &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(-\frac{1}{2}\sigma^2\boldsymbol{\iota}) \\ &= \boldsymbol{\beta} - \frac{1}{2}\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\iota} \\ &= \boldsymbol{\beta} - \frac{1}{2}\sigma^2\mathbf{i}_0 \end{aligned} \tag{7}$$

Here, $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\iota} = \mathbf{i}_0$, the first column of the $(K+1)$- identity matrix $\mathbf{I}_{K+1}$, as a result of the fact that $\boldsymbol{\iota}$ matches the first column of $\mathbf{X}$. This outcome shows that the least squares estimator of the intercept β_0 in (1) is downward biased by an amount $\frac{1}{2}\sigma^2$, while the least squares estimators of the other parameters β_k $(k = 1, \ldots, K)$ in (1) are unbiased. Moreover, the variance-covariance matrix of $\hat{\boldsymbol{\beta}}_{LS}$ is simply equal to:

$$\mathbf{V}(\hat{\boldsymbol{\beta}}_{LS}) = E(\hat{\boldsymbol{\beta}}_{LS} - E(\hat{\boldsymbol{\beta}}_{LS}))(\hat{\boldsymbol{\beta}}_{LS} - E(\hat{\boldsymbol{\beta}}_{LS}))' = \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \tag{8}$$

The unknown σ^2 can be estimated unbiasedly by the familiar mean square error: $\hat{\sigma}^2_{LS}$ = $\sum_i(y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{LS})^2/(N-K-1)$. The estimated standard errors of the least squares parameter estimates are therefore the same as in common applications of the log-linear regression model with $E(\varepsilon) = \mu = 0$.

It is important to note that the bias of the estimator of the intercept β_0 in (1) is not the same as the bias of its exponent in (2), and that the absence of biases in the least squares estimators of the other parameters β_k $(k = 1, \ldots, K)$ does not extend to the exponent of these parameters in (2). This is not merely an academic observation, but has practical relevance for inferences about the time dummy-effects in pooled hedonic regressions. This is easily demonstrated by means of the moment-generating function of $\hat{\boldsymbol{\beta}}_{LS}$. Since $\hat{\boldsymbol{\beta}} \sim n(\boldsymbol{\beta} - \frac{1}{2}\sigma^2\mathbf{i}_0, \sigma^2(\mathbf{X}'\mathbf{X})^{-1})$, applying (4) leads to:

$$\begin{aligned} M_{\hat{\boldsymbol{\beta}}}(\mathbf{t}) &= e^{\mathbf{t}'(\boldsymbol{\beta} - \frac{1}{2}\sigma^2\mathbf{i}_0) + \frac{1}{2}\mathbf{t}'\sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}} \\ &= e^{\mathbf{t}'\boldsymbol{\beta} - \frac{1}{2}\sigma^2 t_0 + \frac{1}{2}\sigma^2\mathbf{t}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{t}} \end{aligned} \tag{9}$$

with $\mathbf{t}$ an arbitrary real valued $(K+1)$-vector having t_0 as its first element. Using (9), the expected value of the estimator of the exponent of the intercept $e^{\hat{\beta}_0}$ is found by setting $\mathbf{t}'$ = $(1, 0, \ldots, 0)$: $E(e^{\hat{\beta}_0})$ = $\exp\{\beta_0 - \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2_{00}\}$, with σ^2_{00} element (1,1) of the variance-covariance matrix (8). The expected value of the exponent of the least squares estimator of an arbitrary parameter β_k is likewise obtained as $E(e^{\hat{\beta}_k})$ = $\exp\{\beta_k + \frac{1}{2}\sigma^2_{kk}\}$, with σ^2_{kk} element (k+1,k+1) of (8), $k = 1, \ldots, K$. So, even when the least squares estimators of the β_k are unbiased, their exponents are not.

Properties of the least squares predictor $\hat{Y}_{LS,r}$

In addition, it should be noted that the main interest in hedonic analyses is in understanding variations in the price of product varieties Y. The question arises, therefore, how inferences about Y in (2) are influenced by the properties of the least squares estimator of the β's in (1). Below we illustrate the consequences of the stochastic properties of models (1) and (2) for such inferences by comparing the expected values of the price Y_r and of its least squares predictor $\hat{Y}_{LS,r} = \exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}\}$ of an arbitrary reference variety r with characteristics $\mathbf{x}_r$.

Starting from (2), the expected price $E(Y_r|\mathbf{x}_r)$ of the reference product is equal to $E(Y_r|\mathbf{x}_r) = \exp\{\mathbf{x}_r'\boldsymbol{\beta}\}E(\nu) = \exp\{\mathbf{x}_r'\boldsymbol{\beta}\}$. The variance of the price Y_r is obtained as: $V(Y_r|\mathbf{x}_r) = \exp\{2\mathbf{x}_r'\boldsymbol{\beta}\}V(\nu) = \exp\{2\mathbf{x}_r'\boldsymbol{\beta}\}(\exp\{\sigma^2\} - 1)$. Note that the variance of the price Y_r depends on the size of its systematic part $\exp\{\mathbf{x}_r'\boldsymbol{\beta}\}$, unlike the variance of the log-price y_r which is equal to σ^2 and therefore constant.

The least squares predictor $\hat{Y}_{LS,r}$ of the price Y_r is obtained by inserting the least squares estimator (6) into the systematic part of (2): $\hat{Y}_{LS,r} = \exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}\}$. The expected value of this predictor is equal to: $E(\hat{Y}_{LS,r}|\mathbf{x}_r) = E(\exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}\}) = \exp\{\mathbf{x}_r'\boldsymbol{\beta}\}\times E(\exp\{\mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\})$. The latter part of this expression can be determined with the moment-generating function (4), which for the disturbances $\boldsymbol{\varepsilon} \sim n(-\frac{1}{2}\sigma^2\boldsymbol{\iota}, \sigma^2\mathbf{I})$, is equal to $M_{\boldsymbol{\varepsilon}}(\mathbf{t}) = \exp\{-\frac{1}{2}\sigma^2\mathbf{t}'\boldsymbol{\iota} + \frac{1}{2}\sigma^2\mathbf{t}'\mathbf{t}\}$. Letting $\mathbf{t}' = \mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, the expected value of the least squares predictor is found as:

$$\begin{aligned}
E(\hat{Y}_{LS,r}|\mathbf{x}_r) &= E(e^{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}}) \\
&= e^{\mathbf{x}_r'\boldsymbol{\beta}}E(e^{\mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}}) \\
&= e^{\mathbf{x}_r'\boldsymbol{\beta}}M_{\boldsymbol{\varepsilon}}(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_r) \\
&= e^{\mathbf{x}_r'\boldsymbol{\beta}}e^{-\frac{1}{2}\sigma^2\mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\iota}+\frac{1}{2}\sigma^2\mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_r} \\
&= e^{\mathbf{x}_r'\boldsymbol{\beta}}e^{-\frac{1}{2}\sigma^2\mathbf{x}_r'\mathbf{i}_0+\frac{1}{2}\sigma^2\mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_r} \\
&= e^{\mathbf{x}_r'\boldsymbol{\beta}-\frac{1}{2}\sigma^2(1-h_r)} \qquad (10)
\end{aligned}$$

The quantity $h_r = \mathbf{x}_r'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_r$ is the leverage value for product variety r with characteristics $\mathbf{x}_r$. It represents the standardized distance of variety r's product characteristics $\mathbf{x}_r$ from the center of the sample space spanned by $\mathbf{X}$. The leverage value h_r is always in between $1/N$ and 1 for reference products that are part of the sample. It is equal to $1/N$, if $\mathbf{x}_r'$ represents a vector of (unweighted) average characteristics, $\mathbf{x}_r' = \boldsymbol{\iota}'\mathbf{X}/N$.

In like fashion, the variance of the price predictor $\hat{Y}_{LS,r}$ can be derived using (5) as: $V(\hat{Y}_{LS,r}|\mathbf{x}_r) = \exp\{2\mathbf{x}_r'\boldsymbol{\beta} - \sigma^2(1-h_r)\}(\exp\{\sigma^2 h_r\} - 1)$.

It is interesting to note that the least squares price predictor $\hat{Y}_{LS,r}$ is a biased estimator of $E(Y_r|\mathbf{x}_r) = \exp\{\mathbf{x}_r'\boldsymbol{\beta}\}$. The extent of the bias may be conveniently represented by the ratio of the expected values of the predictor and the predictand:

$$\begin{aligned}
B_{LS,r} &= E(\hat{Y}_{LS,r}|\mathbf{x}_r)/E(Y_r|\mathbf{x}_r) \\
&= e^{\mathbf{x}_r'\boldsymbol{\beta}-\frac{1}{2}\sigma^2(1-h_r)}/e^{\mathbf{x}_r'\boldsymbol{\beta}} \\
&= e^{-\frac{1}{2}\sigma^2(1-h_r)} \qquad (11)
\end{aligned}$$

The expression shows that price predictions made for inside the sample products (in terms of their characteristics $\mathbf{x}_r$), which have leverage values h_r in between $1/N$ and

1, are always downward biased: $0 < B_{LS,r} < 1$. The predicted prices are systematically below the expected price determined by (2). If price predictions are determined for new products, in the sense that the associated characteristics $\mathbf{x}_r$ have no antecedents in the sample but are instead far removed from the center of gravity, it may occur that $h_r > 1$, in which case the predictions are positively biased, $B_{LS,r} > 1$.

Properties of the maximum likelihood estimators of β and $E(Y_r|\mathbf{x}_r)$

Sofar, the discussion has focussed on biases associated with least squares estimation of the unknown $\boldsymbol{\beta}$ via (1). An alternative way to estimate the unknown $\boldsymbol{\beta}$, is to apply the maximum likelihood method directly to (2). The question is how inferences about $\boldsymbol{\beta}$ but more importantly about the price Y_r of some reference variety are affected by this alternative.

In the case of maximum likelihood applied to (2), it can be shown (see Teekens and Koerts, 1972) that the estimator $\hat{\boldsymbol{\beta}}_{ML}$ of $\boldsymbol{\beta}$ for known σ^2 follows as:

$$\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{LS} + \frac{1}{2}\sigma^2\mathbf{i}_0 \tag{12}$$

Unlike the least squares estimator, the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_{ML}$ of $\boldsymbol{\beta}$ is therefore seen not to underestimate the intercept by an amount $\frac{1}{2}\sigma^2$. In fact, it is an unbiased estimator of $\boldsymbol{\beta}$ in view of (7): $E(\hat{\boldsymbol{\beta}}_{ML}) = \boldsymbol{\beta}$. The maximum likelihood estimator of σ^2 is equal to $\hat{\sigma}^2_{ML} = \frac{N-K-1}{N}\hat{\sigma}^2_{LS}$, which is slightly biased but consistent. The variance may also be estimated by $\hat{\sigma}^2_{LS}$, which is unbiased but for which the likelihood function is no longer at its maximum.

Interestingly, the unbiased estimation of $\boldsymbol{\beta}$ by maximum likelihood does not expel the bias from the price predictor $\hat{Y}_{ML,r} = \exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{ML}\}$ as an estimator of the expected price of a reference variety $E(Y_r|\mathbf{x}_r)$. The expected value of the maximum likelihood predictor $\hat{Y}_{ML,r}$ is equal to: $E(\hat{Y}_{ML,r}|\mathbf{x}_r) = E(\exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{ML}\}) = E(\exp\{\mathbf{x}_r'(\hat{\boldsymbol{\beta}}_{LS} + \frac{1}{2}\sigma^2\mathbf{i}_0)\}) = E(\exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}\})\exp\{\frac{1}{2}\sigma^2\}$. Using the result for the expected value of the least squares predictor $\exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}\}$ in (10), the expected value of the maximum likelihood predictor becomes:

$$E(\hat{Y}_{ML,r}|\mathbf{x}_r) = E(e^{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}})e^{\frac{1}{2}\sigma^2} = e^{\mathbf{x}_r'\boldsymbol{\beta} + \frac{1}{2}\sigma^2 h_r} \tag{13}$$

The variance of $\hat{Y}_{ML,r}$ can be obtained as: $V(\hat{Y}_{ML,r}|\mathbf{x}_r) = \exp\{\sigma^2\}V(\exp\{\mathbf{x}_r'\hat{\boldsymbol{\beta}}_{LS}\}) = \exp\{2\mathbf{x}_r'\boldsymbol{\beta} + \sigma^2 h_r\}(\exp\{\sigma^2 h_r\} - 1)$.

The extent of the bias of $\hat{Y}_{ML,r}$ may again be expressed as a multiplicative factor $B_{ML} = E(\hat{Y}_{ML,r}|\mathbf{x}_r)/E(Y_r|\mathbf{x}_r)$, which is equal to:

$$\begin{aligned} B_{ML,r} &= E(\hat{Y}_{ML,r}|\mathbf{x}_r)/E(Y_r|\mathbf{x}_r) \\ &= e^{\mathbf{x}_r'\boldsymbol{\beta} + \frac{1}{2}\sigma^2 h_r}/e^{\mathbf{x}_r'\boldsymbol{\beta}} \\ &= e^{\frac{1}{2}\sigma^2 h_r} \end{aligned} \tag{14}$$

Since the bias factor B_{ML} is always larger than 1, the maximum likelihood predictor $\hat{Y}_{ML,r}$ is seen to systematically overstate the expected price of a reference variety.

Intermediate result
Summing up, the least squares estimator $\hat{\boldsymbol{\beta}}_{LS}$ is always a biased estimator of $\boldsymbol{\beta}$ in (2), while the maximum likelihood estimator $\hat{\boldsymbol{\beta}}_{ML}$ is practically unbiased (except for the slightly biased maximum likelihood estimate of σ^2). The least squares predictor $\hat{Y}_{LS,r}$ is a (usually) downward biased estimator of the expected price $E(Y_r|\mathbf{x}_r)$, while the maximum likelihood predictor $\hat{Y}_{ML,r}$ is upward biased. The size of the bias depends on two factors: the variance σ^2 of the disturbance ε in (1); and, in the case of the predicted prices, the leverage values h_r of the reference variety with product characteristics $\mathbf{x}_r$. The biases are small, when both factors are close to zero. The next session discusses the implications of these biases for the measurement of several hedonic price indices proposed in the literature.

3 Hedonic price indices based on annually estimated hedonic regressions

Several hedonic price indices can be calculated based on the log-linear models (1) and (2). One way is to define Laspeyres- and Paasche-like price indices as quality-adjusted price ratios of a typical product variety. Here, 'typical' is operationalized as a product variety with characteristics equal to the average product characteristics in either the reference period (Laspeyres) or the current period (Paasche). Another way is to adopt a matched-model approach, in which various price indices are constructed as (arithmetic or geometric) averages of price ratios of matching product varieties. The hedonic price-characteristics relationship is used here to impute the missing prices of entering or exiting product varieties. Yet another way is to estimate the hedonic model on a pooled sample of two or more periods, while using time-dummies to catch up with systematic price changes with respect to some base-period price level.
Starting point for all three methods of calculating quality-adjusted price indices is the assumption that the price-characteristics relationship (1) holds in both the current period (=1) and the reference period (=0):

$$\begin{aligned} y_0 &= \mathbf{x}_0'\boldsymbol{\beta}_0 + \varepsilon_0, & \varepsilon_0 &\sim n(-\frac{1}{2}\sigma_0^2, \sigma_0^2) \\ y_1 &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \varepsilon_1, & \varepsilon_1 &\sim n(-\frac{1}{2}\sigma_1^2, \sigma_1^2) \end{aligned} \quad (15)$$

for all product varieties offered for sale in the two periods of observation.
Although the difference between these approaches is not always strict, it perfectly serves to clarify the various ways in which the hedonics-based indices may suffer estimation biases. Below, we illustrate the consequences of least-squares and maximum-likelihood estimation of (1) for these calculation methods in the aforementioned order.

3.1 Definition of the quality-adjusted price indices, P_r^A and P_r^B

In a cost of living framework, price indices are defined as the ratio of the minimum expenditures necessary to attain some reference utility (or living standard) at current and reference period prices. One way to measure this ratio is to compare the characteristics of a typical product variety at current and base period implicit prices. This section discusses how this price ratio is defined, how it is estimated and how the estimates are biased due to the improper use of the log-linear model.

A price index defined as the expected value of a price ratio, P_r^A
Quality-adjusted price changes based on (15) may be defined as: (*i*) the expected value of the ratio of the price levels of some reference product (P_r^A); or (*ii*) the ratio of the expected price levels of a typical variety (P_r^B). The difference between the two indices concerns their treatment of non-systematic price variation in the model. The first index P_r^A is based on the price ratio $P(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0, \mathbf{x}_r) = Y_{1r}/Y_{0r}$ for a representative product variety with characteristics $\mathbf{x}_r$, which can be written by using (15) and (2) as:

$$\begin{aligned} P(\boldsymbol{\beta}_1, \boldsymbol{\beta}_0, \mathbf{x}_r) &= \frac{Y_{1r}}{Y_{0r}} \\ &= \frac{\exp\{\mathbf{x}_r'\boldsymbol{\beta}_1 + \varepsilon_{1r}\}}{\exp\{\mathbf{x}_r'\boldsymbol{\beta}_0 + \varepsilon_{0r}\}} \\ &= \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{\varepsilon_{1r} - \varepsilon_{0r}\} \end{aligned} \tag{16}$$

The reference variety r may be a single product variety satisfying the model properties (15) or a hypothetical product variety with $\mathbf{x}_r$ equal to the average characteristics $\bar{\mathbf{x}}_r$ of a reference set. In the latter instance, Y_{1r} and Y_{0r} can be interpreted as (weighted) geometric averages of prices Y_{1rj} and Y_{0rj} of varieties j in the reference set, which satisfy model (15); ε_{1r} and ε_{0r} are (weighted) averages of the corresponding disturbances ε_{1rj} and ε_{0rj}. Suppose that the reference set consists of n_r product varieties having weights w_{rj}, $0 \leq w_{rj} \leq 1$, which are conveniently summarized by a n_r-vector $\mathbf{w}_r' = (w_{r1}, \ldots, w_{r,n_r})$ satisfying $\mathbf{w}_r'\boldsymbol{\iota}_r = 1$. The reference prices are then obtained as $Y_{1r} = \prod_{j=1}^{n_r} Y_{1rj}^{w_{rj}} = \prod_{j=1}^{n_r} \exp\{\mathbf{x}_{rj}'\boldsymbol{\beta}_1 + \varepsilon_{1rj}\}^{w_{rj}} =$

$\exp\{\sum_{j=1}^{n_r} w_{rj}\mathbf{x}_{rj}'\boldsymbol{\beta}_1 + \sum_{j=1}^{n_r} w_{rj}\varepsilon_{1rj}\} = \exp\{\bar{\mathbf{x}}_r'\boldsymbol{\beta}_1 + \varepsilon_{1r}\}$ with $\mathbf{x}_r = \bar{\mathbf{x}}_r = \sum_{j=1}^{n_r} w_{rj}\mathbf{x}_{rj}$ and $\varepsilon_{1r} = \sum_{j=1}^{n_r} w_{rj}\varepsilon_{1rj} = \mathbf{w}_r'\boldsymbol{\varepsilon}_{1r}$. Similarly, $Y_{0r} = \exp\{\bar{\mathbf{x}}_r'\boldsymbol{\beta}_0 + \varepsilon_{0r}\}$ with $\varepsilon_{0r} = \mathbf{w}_r'\boldsymbol{\varepsilon}_{0r}$ and $\mathbf{w}_r$ and $\bar{\mathbf{x}}_r$ defined as before.

Clearly, the price ratio (16) is a stochastic quantity which partly consists of the systematic parts of the price-characteristics relationships (15) and partly of the random deviations from these systematic parts. It is itself not an attractive index, as the composing price levels Y_{1r} and Y_{0r} are hypothetical, not necessarily observable constructs and the index is subject to random variation. It seems more obvious, therefore to take the expected value of (16) as the relevant quality-adjusted period-to period price change, yielding $P_r^A = E(Y_{1r}/Y_{0r})$:

$$\begin{aligned} P_r^A &= E\left(\frac{Y_{1r}}{Y_{0r}}\right) \\ &= \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}E\left(\exp\{\varepsilon_{1r} - \varepsilon_{0r}\}\right) \\ &= \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{\sigma_r^2\} \end{aligned} \tag{17}$$

where $\sigma_r^2 = -\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}\sigma_1^2\mathbf{w}_r'\mathbf{w}_r + \frac{1}{2}\sigma_0^2\mathbf{w}_r'\mathbf{w}_r$. The expectation term in (17) is obtained by elaborating the moment-generating function (4) for a random variable $\mathbf{z}' = (\boldsymbol{\varepsilon}_{1r}', \boldsymbol{\varepsilon}_{0r}')$ with $\boldsymbol{\delta}' = E(\mathbf{z})' = (-\frac{1}{2}\sigma_1^2\boldsymbol{\iota}_r', -\frac{1}{2}\sigma_0^2\boldsymbol{\iota}_r')$, $\boldsymbol{\Delta}$ a block-diagonal matrix with $\sigma_1^2\mathbf{I}_r$ and $\sigma_0^2\mathbf{I}_r$ on the main diagonal, and $\mathbf{t}' = (\mathbf{w}_r', -\mathbf{w}_r')$.

Two aspects of this index are worth noting. First, the quality-adjusted price index P_r^A depends on the difference between the disturbance variances of both periods, $-\frac{1}{2}(\sigma_1^2 - \sigma_0^2)$, and on the average of the disturbance variances times the sum of squared weights, $\frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r$. If all varieties in the reference set are equally weighted, then $w_{rj} = 1/n_r$ for $j = 1, \ldots, n_r$ and $\sigma_r^2 = -\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}(\sigma_1^2 + \sigma_0^2)/n_r$. If the number of product varieties in the reference set n_r is relatively large, then the role of the latter part is comparatively small. If there is but a single reference variety, then $n_r = 1$ and σ_r^2 is seen to reflect uncertainty in the base-period hedonic relationship only, $\sigma_r^2 = \sigma_0^2$. Second, the disturbances ε_{1r} and ε_{0r} are supposed to be independently distributed, *cov*$(\varepsilon_{1r}, \varepsilon_{0r})$=0. If non-zero correlations ρ_{10} between the disturbances of the product varieties in the reference set are assumed, then the last part of (17) would become $\sigma_r^2 = -\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}(\sigma_1^2 + \sigma_0^2 - 2\rho_{01}\sigma_0\sigma_1)\mathbf{w}_r'\mathbf{w}_r$. This can be easily checked by applying (4) as before with $\boldsymbol{\Delta}$ having non-zero off-diagonal blocks $\sigma_{01}\mathbf{I}_r$, with the covariance $\sigma_{01} = \rho_{01}\sigma_0\sigma_1$. If the reference varieties are equally weighted, then $\mathbf{w}_r'\mathbf{w}_r = 1/n_r$ and the role of the correlation term is seen to diminish with n_r.

A price index defined as the ratio of expected prices, P_r^B

A second quality-adjusted price index P_r^B may be defined as the ratio of the systematic parts of the price-characteristics relationships:

$$P_r^B = \frac{\exp\{\mathbf{x}_r' \boldsymbol{\beta}_1\}}{\exp\{\mathbf{x}_r' \boldsymbol{\beta}_0\}} = \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\} \tag{18}$$

This approach links up with the theoretical outcome of Lancaster's household production theory: it directly relates the solutions of the first order conditions rather than their stochastic price-quality representations given by (15). In this respect, P_r^B resembles the pooled hedonic price index $P^P = \exp\{\pi\}$ in (48), which is similarly defined as a ratio of expected prices of a given product variety. Unlike P_r^A, the quality-adjusted price index P_r^B is not affected by the uncertainty in the price-characteristics equation reflected by σ_r^2.

The general definitions of the quality-adjusted price indices P_r^A and P_r^B can be further specified by explicating the characteristics $\mathbf{x}_r$ of the reference product variety r. Thus, Laspeyres-like indices P_L^A and P_L^B are obtained by choosing the reference product variety equal to the average amounts of product characteristics in the base period: $\mathbf{x}_r = \bar{\mathbf{x}}_0$. Paasche-like quality-adjusted price indices P_P^A and P_P^B are obtained by setting the reference variety characteristics equal to the average current period characteristics $\mathbf{x}_r = \bar{\mathbf{x}}_1$. Having defined the quality-adjusted price indices of interest, we now turn to their estimation.

3.2 Estimation of the quality-corrected price indices, P_r^A and P_r^B

The quality-adjusted price indices P_r^A and P_r^B are based on the unknown parameters of the hedonic models (15), which can be estimated by least-squares or by maximum-likelihood. This section discusses the consequences of these estimation methods starting with least-squares.

Least-squares estimation of the quality-corrected price indices

Estimates $\hat{P}_{LS,r}$ of the proposed quality-adjusted price indices are often constructed as the ratio of predicted price levels, $\hat{Y}_{1r}$ and $\hat{Y}_{0r}$, obtained by least squares estimation of the unknown parameters in (15) for the current and base period separately and inserting the relevant sample averages $\bar{\mathbf{x}}_1$ or $\bar{\mathbf{x}}_0$, leading to $\hat{P}_{LS,r}$:

$$\hat{P}_{LS,r} = \frac{\hat{Y}_{LS,1r}}{\hat{Y}_{LS,0r}} = \frac{\exp\{\mathbf{x}_r' \hat{\boldsymbol{\beta}}_{LS,1}\}}{\exp\{\mathbf{x}_r' \hat{\boldsymbol{\beta}}_{LS,0}\}} = \exp\{\mathbf{x}_r'(\hat{\boldsymbol{\beta}}_{LS,1} - \hat{\boldsymbol{\beta}}_{LS,0})\} \tag{19}$$

The question is whether, how and to what extent this procedure biases the price index measures, where for ease of exposition the population variances are treated as known.

Using (6), the expected value of $\hat{P}_{LS,r}$ can be obtained as:

$$\begin{aligned} E\left(\hat{P}_{LS,r}\right) &= E\left(\exp\{\mathbf{x}_r'(\hat{\boldsymbol{\beta}}_{LS,1} - \hat{\boldsymbol{\beta}}_{LS,0})\}\right) \\ &= \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\} \times \\ &\quad E\left(\exp\{\mathbf{x}_r'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon}_1 - \mathbf{x}_r'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\varepsilon}_0\}\right) \\ &= \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\} \end{aligned} \tag{20}$$

with $h_{0r} = \mathbf{x}_r'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{x}_r$ and $h_{1r} = \mathbf{x}_r'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{x}_r$. The expectation term in (20) follows the application of the moment-generating function (4) for a random variable $\mathbf{z}' = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_0') \sim n(\boldsymbol{\delta}, \boldsymbol{\Delta})$ with $\boldsymbol{\delta}' = (-\frac{1}{2}\sigma_1^2\boldsymbol{\iota}_1', -\frac{1}{2}\sigma_0^2\boldsymbol{\iota}_0')$, $\boldsymbol{\Delta}$ a block diagonal matrix with $\sigma_1^2\mathbf{I}_1$ and $\sigma_0^2\mathbf{I}_0$ on the main diagonal, and a vector $\mathbf{t}' = (\mathbf{x}_r'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1', -\mathbf{x}_r'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0')$. The variance of $\hat{P}_{LS,r}$ can be determined similarly as: $V(\hat{P}_{LS,r}) = \exp\{2\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) - \sigma_1^2(1 - h_{1r}) + \sigma_0^2(1 + h_{0r})\}(\exp\{\sigma_1^2 h_{1r} + \sigma_0^2 h_{0r}\} - 1)$.

Comparing the expected value of the quality-adjusted index (20) with the estimands in (17) or (18) reveals that the hedonic price indices based on log-linear regressions are clearly not unbiased. The extent of the bias may be represented by the ratio of expected price changes conform (11), which in the case of P_r^A leads to the bias factor $B_{LS,r}^A$:

$$\begin{aligned} B_{LS,r}^A &= \frac{E(\hat{P}_{LS,r})}{P_r^A} \\ &= \frac{\exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\}}{\exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r\}} \\ &= \exp\{\frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r} - \frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r\} \end{aligned} \tag{21}$$

In the case of P_r^B, the bias factor $B_{LS,r}^B$ for a product variety with quality characteristics $\mathbf{x}_r$ is equal to:

$$\begin{aligned} B_{LS,r}^B &= \frac{E(\hat{P}_{LS,r})}{P_r^B} \\ &= \frac{\exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\}}{\exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}} \\ &= \exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\} \end{aligned} \tag{22}$$

The bias factors can be further specified by evaluating (21) and (22) for a particular reference variety. In the case of the Laspeyres-like price-indices P_L^A and P_L^B, insertion of the base-period (weighted) average characteristics $\mathbf{x}_r = \bar{\mathbf{x}}_0$ leads to the bias factors $B^A_{LS,L}$ and $B^B_{LS,L}$, while in the case of the Paasche-like indices P^A_P and P^B_P, insertion of the current-period (weighted) averages $\mathbf{x}_r = \bar{\mathbf{x}}_1$ gives the bias factors $B^A_{LS,P}$ and $B^B_{LS,P}$. The practical implications of these biases are discussed below.

Maximum likelihood estimation of the quality-corrected price indices

In addition, quality-adjusted price indices can be based on maximum likelihood estimation of (15) leading to $\hat{P}_{ML,r}$. Analogously to (19) and recalling that the maximum likelihood estimator of the unknown $\boldsymbol{\beta}$ is equal to the least squares estimator plus half of the error variance (12), $\hat{P}_{ML,r}$ can be written as:

$$\begin{aligned}\hat{P}_{ML,r} &= \exp\{\mathbf{x}_r'(\hat{\boldsymbol{\beta}}_{ML,1} - \hat{\boldsymbol{\beta}}_{ML,0})\} \\ &= \exp\{\mathbf{x}_r'(\hat{\boldsymbol{\beta}}_{LS,1} - \hat{\boldsymbol{\beta}}_{LS,0}) + \frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \\ &= \hat{P}_{LS,r} \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \end{aligned} \tag{23}$$

The maximum-likelihood based estimator $\hat{P}_{ML,r}$ is therefore simply obtained by multiplying the least-squares estimated price-index $\hat{P}_{LS,r}$ with $\exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\}$. Using (20), the expected value of $\hat{P}_{ML,r}$ is found as:

$$\begin{aligned}E\left(\hat{P}_{ML,r}\right) &= E\left(\hat{P}_{LS,r}\right) \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \\ &= \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\} \times \\ &\quad \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \\ &= \exp\{\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\}\exp\{\frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\} \end{aligned} \tag{24}$$

where h_{0r} and h_{1r} are defined as before. Likewise, the variance of $\hat{P}_{ML,r}$ is obtained as $V(\hat{P}_{ML,r}) = V(\hat{P}_{LS,r})\exp\{\sigma_1^2 - \sigma_0^2\} = \exp\{2\mathbf{x}_r'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0) + \sigma_1^2 h_{1r} + \sigma_0^2 h_{0r}\}(\exp\{\sigma_1^2 h_{1r} + \sigma_0^2 h_{0r}\} - 1)$. Reference product varieties with base-period average characteristics $\mathbf{x}_r = \bar{\mathbf{x}}_0$ or current-period average characteristics $\mathbf{x}_r = \bar{\mathbf{x}}_1$ can be inserted into (24) to have the Laspeyres-like hedonic price index $\hat{P}_{ML,L}$ and the Paasche-like price index $\hat{P}_{ML,P}$.

The quality-adjusted, maximum-likelihood based price index $\hat{P}_{ML,r}$ is still not an unbiased estimator of the indices defined in (17) and (18). Combining the expected value of $\hat{P}_{ML,r}$ in (24) with P_r^A in (17) leads to the bias factor $B_{ML,r}^A$ equal to:

$$
\begin{aligned}
B_{ML,r}^A &= \frac{E(\hat{P}_{ML,r})}{P_r^A} \\
&= B_{LS,r}^A \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \\
&= \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r} - \frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r\} \qquad (25)
\end{aligned}
$$

Similarly, the bias factor $B_{ML,r}^B$ with respect to P_r^B is:

$$
\begin{aligned}
B_{ML,r}^B &= \frac{E(\hat{P}_{ML,r})}{P_r^B} \\
&= B_{LS,r}^B \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \\
&= \exp\{\frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\} \qquad (26)
\end{aligned}
$$

The bias factors, $B_{ML,L}^A$, $B_{ML,L}^B$, $B_{ML,P}^A$ and $B_{ML,P}^B$, for the Laspeyres- and Paasche-like price changes are again found by substituting $\mathbf{x}_r = \bar{\mathbf{x}}_0$ and $\mathbf{x}_r = \bar{\mathbf{x}}_1$ into (25) and (26), respectively.

3.3 Bias implications

A comparison of the bias factors (21), (22), (25) and (26) shows that the quality-adjusted hedonic price index is a biased estimator of both P_r^A and P_r^B, regardless of using the least-squares (19) or maximum likelihood (23) results. The differences between the bias factors are caused by three factors: (*i*) $\exp\{-\frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r\}$ due to the choice of estimand P_r^A in (17) or P_r^B in (18); (*ii*) $\exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\}$ due to the choice of estimation method, least squares (19) or maximum likelihood (23); and (*iii*) the shared component $\exp\{\frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\}$, which is due to the use of a non-linear, semi-log hedonic model (15) rather than a strictly linear specification. The first element $\exp\{-\frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r\}$ is always smaller than 1 causing a downward bias of the hedonic price indices. The third element, $\exp\{\frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\}$, is always larger than 1 causing an upward bias of the estimated hedonic price indices. The second element $\exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\}$ can be either larger or smaller than 1 depending on the difference between the variances.

The size of the biases therefore depends on: (*i*) the choice of the price index to be estimated P_r^A or P_r^B; (*ii*) the amount of unexplained variation in the hedonic relationships, σ_0^2 and σ_1^2; and (*iii*) the leverage values of the reference varieties, h_{0r} and h_{1r}. The first source of variation is not empirical, but a consequence of preferences for either P_r^A or P_r^B articulated on metaphysical grounds. The resulting bias can simply be resolved by multiplying the estimated hedonic indices (19) or (23) by $\exp\{\frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r\}$.

The second source of variation is more subtle. If the variances σ_0^2 and σ_1^2 are more or less equal-sized, such that the difference $\sigma_0^2 - \sigma_1^2$ is close to zero, then the bias due to choice of estimation method (least squares or maximum likelihood), is negligible. If the variances are small in relation to the leverage values, then also the bias due to the application of a non-linear hedonic model $\exp\{\frac{1}{2}(\sigma_1^2 h_{1r} + \sigma_0^2 h_{0r})\}$ will be small. The role of the variances therefore has partly to do with the stability of the hedonic relationship in time (equal σ_0^2 and σ_1^2) and partly with the absolute performance of the hedonic models (small σ_0^2 and σ_1^2).

The third source of variation concerns the leverage values h_{0r} and h_{1r}, which reflect the (standardized) distance between product varieties with characteristics $\mathbf{x}_r$ and the center of gravity of the explanatory information summarized by $\mathbf{X}_0$ or $\mathbf{X}_1$. In the case of typical product varieties with characteristics equal to the (unweighted) sample averages $\mathbf{x}_r = \bar{\mathbf{x}}_0$ or $\mathbf{x}_r = \bar{\mathbf{x}}_1$, the leverages h_{00} and h_{11} are equal to the reciprocal values of the corresponding sample sizes ($1/N_0$ and $1/N_1$). Moreover, if the characteristics space does not change too much from one period to another, such that $\bar{\mathbf{x}}_0$ and $\bar{\mathbf{x}}_1$ do not differ too much, then h_{01} and h_{10} will also be close to these reciprocal values. This implies that the bias due to the use of non-linear models, $\exp\{\frac{1}{2}(\sigma_1^2 h_{1r} + \sigma_0^2 h_{0r})\}$, decreases when the sample size increases, assuming that the variances σ_0^2 and σ_1^2 are fixed. Combining these insights, it follows that the maximum-likelihood based hedonic index usually is an adequate measure, especially in larger samples, while the adequacy of the least-squares based hedonic index depends on the size and temporal stability of the variances σ_0^2 and σ_1^2.

The expressions for the bias factors also show that a high explanatory power of the hedonic regressions is not a sufficient condition to prevent biases to occur, and not even a necessary condition to cause biases not to occur. The reason for this seemingly counterintuitive result is that the bias depends on the error variances which are not standardized measures, unlike the coefficient of determination R^2. For example, if the explanatory power of the estimated hedonic models is strong, such that the two variances σ_0^2 and σ_1^2 are small (that is close to zero, in absolute sense), then the resulting biases will indeed be small. However, if the explanatory power of the models is poor but the variances σ_0^2 and σ_1^2 remain stable over time, then the bias due to the choice of estimation method, least squares or maximum likelihood, will still be small. By contrast, if the difference between the two variances σ_0^2 and σ_1^2 is substantial despite a possibly high explanatory power, then also the bias may be quite large. The practical relevance of the biases will be further illustrated below using examples from the literature.

3.4 Weighted least squares and maximum likelihood

Much applied work in price index measurement is based on weighted least squares of the hedonic relations (15); examples are Cole et al. (1986), Aizcorbe, Corrado and Doms (2000), and Silver and Heravi (2001). The weighting variable is often taken to be some measure of sales volume in line with previous suggestions by Griliches (1971). A consequence of weighting is that the previously developed formulae no longer apply and that published information about mean square errors and sample sizes is insufficient to gauge on the extent of the biases. The question addressed in this section is how the previous results are affected when applying weighted rather than unweighted estimation of the unknown parameters in the hedonic model (15).
Suppose that the hedonic model is estimated by weighting each product variety with an amount q_{it}, $i = 1, \ldots, N_t$, t = 0 or 1. The weights are positive, $q_{it} > 0$, but do not necessarily add up to 1, which conforms with the commonly used sales volume measures. The weights are arranged in diagonal N_t-matrices $\mathbf{Q}_t$ = diag$(q_{1t}, \ldots, q_{N_t,t})$ or N_t-vectors $\mathbf{q}'_t = (q_{1t}, \ldots, q_{N_t,t})$, where $\mathbf{q}_t = \mathbf{Q}_t \boldsymbol{\iota}_t$ and t = 0 or 1, to facilitate the estimation procedure. After specification of the weight variable, most software packages produce outcomes of the weighted least squares estimator $\hat{\boldsymbol{\beta}}_{WS,t}$ defined as (cf. (6)):

$$\hat{\boldsymbol{\beta}}_{WS,t} = (\mathbf{X}'_t \mathbf{Q}_t \mathbf{X}_t)^{-1} \mathbf{X}'_t \mathbf{Q}_t \mathbf{y}_t = \boldsymbol{\beta}_t + (\mathbf{X}'_t \mathbf{Q}_t \mathbf{X}_t)^{-1} \mathbf{X}'_t \mathbf{Q}_t \boldsymbol{\varepsilon}_t \tag{27}$$

which is a solution to the minimization of the weighted sum of squares $(\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)' \times \mathbf{Q}_t (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}_t)$ over $\boldsymbol{\beta}_t$. The approach is consistent with assuming that the disturbance vectors are distributed as $\boldsymbol{\varepsilon}_t \sim n(\boldsymbol{\mu}_t, \sigma_t^2 \mathbf{Q}_t^{-1})$ for both periods, where the $\boldsymbol{\mu}_t$ are yet to be determined.
The impact of weighting on the biases depends on the objective of the weighting scheme. First, weighting may be considered a simple way to take heteroscedasticity into account. In this view, the variances of the product varieties are allowed to differ, $V(\varepsilon_{it}) = \sigma_{it}^2$ with $\sigma_{it}^2 = \sigma_t^2 / q_{it}$. Applying the condition $E(\nu_{it}) = 1$, conform (1) and (3), gives $E(\varepsilon_{it}) = -\frac{1}{2}\sigma_{it}^2 = -\frac{1}{2}\sigma_t^2/q_{it}$. Hence, the disturbance vectors are distributed as $\boldsymbol{\varepsilon}_t \sim n(-\frac{1}{2}\sigma_t^2 \mathbf{Q}_t^{-1} \boldsymbol{\iota}_t, \sigma_t^2 \mathbf{Q}_t^{-1})$. Second, weighting may be considered an attempt to analyze the actually sold items of a product variety realizing that some product varieties are more popular than others. In this view, the observed prices Y_{it} in the sample may be interpreted as geometric averages of the unobserved prices $Y_{it,j}$ of q_{it} sold items of the product variety, each item satisfying the log-linear model (15) as before: $Y_{it} = \prod_{j=1}^{q_{it}} Y_{it,j}^{1/q_{it}} = \prod_{j=1}^{q_{it}} \exp\{\mathbf{x}'_{it}\boldsymbol{\beta}_t + \varepsilon_{it,j}\}^{1/q_{it}} = \exp\{\frac{1}{q_{it}} \sum_{j=1}^{q_{it}} \mathbf{x}'_{it}\boldsymbol{\beta}_t + \frac{1}{q_{it}} \sum_{j=1}^{q_{it}} \varepsilon_{it,j}\}$ $= \exp\{\mathbf{x}'_{it}\boldsymbol{\beta}_t + \varepsilon_{it}\}$ with $\varepsilon_{it} = \frac{1}{q_{it}} \sum_{j=1}^{q_{it}} \varepsilon_{it,j}$. Assuming the $\varepsilon_{it,j}$ to be identically and independently normally distributed satisfying $E(\nu_{it,j}) = 1$ in accordance with (15), $\varepsilon_{it,j} \sim n(-\frac{1}{2}\sigma_t^2, \sigma_t^2)$, implies that the ε_{it} are also normally distributed with expectation $-\frac{1}{2}\sigma_t^2$, but with variance σ_t^2/q_{it}, $\varepsilon_{it} \sim n(-\frac{1}{2}\sigma_t^2, \sigma_t^2/q_{it})$. The disturbance vector

is therefore distributed as $\boldsymbol{\varepsilon}_t \sim n(-\frac{1}{2}\sigma_t^2\boldsymbol{\iota}_t, \sigma_t^2\mathbf{Q}_t^{-1})$. So, changing the order of applying the weights and imposing the condition $E(\nu) = 1$ affects the expected value of the disturbances $\boldsymbol{\varepsilon}_t$, while the variance-covariance matrix of $\boldsymbol{\varepsilon}_t$ remains unchanged. Both approaches are consistent with the weighted least squares procedure yielding (27). But obviously, the expected value of $\hat{\boldsymbol{\beta}}_{WS,t}$ depends on whether one takes $\boldsymbol{\mu}_t = -\frac{1}{2}\sigma_t^2\mathbf{Q}_t^{-1}\boldsymbol{\iota}_t$ or $\boldsymbol{\mu}_t = -\frac{1}{2}\sigma_t^2\boldsymbol{\iota}_t$ as the expected value of $\boldsymbol{\varepsilon}_t$.
Here, we adopt the second motivation for weighting, assuming that weighting serves to take varying sales volumes into account rather than to cope with varying price heterogeneity between product varieties. The expected value of $\hat{\boldsymbol{\beta}}_{WS,t}$ in (27) is consequently found as:

$$E\left(\hat{\boldsymbol{\beta}}_{WS,t}\right) = \boldsymbol{\beta}_t - \frac{1}{2}\sigma_t^2\mathbf{i}_0 \tag{28}$$

which is similar to (7) but with σ_t^2 the multiplicative constant in $\sigma_t^2\mathbf{Q}_t^{-1}$ instead of in $\sigma_t^2\mathbf{I}_t$. The weighted maximum likelihood estimator $\hat{\boldsymbol{\beta}}_{WL,t}$ can be found as $\hat{\boldsymbol{\beta}}_{WL,t} = \hat{\boldsymbol{\beta}}_{WS,t} + \frac{1}{2}\sigma_t^2\mathbf{i}_0$ in accordance with (12). In addition, all previous derivations remain the same except that all cross-products $\mathbf{X}_t'\mathbf{X}_t$ become weighted cross-products $\mathbf{X}_t'\mathbf{Q}_t\mathbf{X}_t$. Moreover, σ_t^2 is estimated by the weighted least squares estimator: $\hat{\sigma}_{WS,t}^2 = (\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_{WS,t})'\mathbf{Q}_t(\mathbf{y}_t - \mathbf{X}_t\hat{\boldsymbol{\beta}}_{WS,t})/(N_t - K_t - 1)$. The impact of weighting on the theoretical results therefore seems rather limited. But note that this would definitely not be the case when adopting the heteroscedasticity consideration with $E(\boldsymbol{\varepsilon}_t) = -\frac{1}{2}\sigma_t^2\mathbf{Q}_t^{-1}\boldsymbol{\iota}_t$, which leads to an expected value of $\hat{\boldsymbol{\beta}}_{WS,t}$ equal to $E(\hat{\boldsymbol{\beta}}_{WS,t}) = \boldsymbol{\beta}_t - \frac{1}{2}\sigma_t^2(\mathbf{X}_t'\mathbf{Q}_t\mathbf{X}_t)^{-1}\mathbf{X}_t'\boldsymbol{\iota}_t$.

3.5 Some empirical results

The consequences of using log-linear models for price index construction are illustrated with an example of new passenger cars sold in the Dutch market during the period 1990-1999. The data about prices, quality characteristics and sales have been described in Van Dalen and Bode (2004). Their table 3 presents various hedonic price indices of interest, and their table C.1 gives the regression results of the hedonic regressions used. Some relevant features of these results, such as sample sizes, R^2's, mean square errors and brand-weighted, least-squares based Laspeyres and Paasche-like price indices, have been put in tables 1 and 2 below. The results indicate considerable fluctuation in the number of product varieties offered for sale during the observation period, starting from 1816 product varieties in 1990 and growing to 2937 varieties in 1999. The explanatory power of the hedonic regressions varies from 93.1% to 95.5%, while the mean square error lies between 0.524 in 1997 and 2.03 in 1990. The quality-corrected chained Laspeyres-like price index is seen to increase to

1.048 in 1993, after which it declines to 0.963 in 1999. Likewise the chained Paasche-like price index increases to 1.069 in 1993 and then decreases to 0.990 in 1999. The question is how these least-squares based results are affected by the biases associated by the use of log-linear hedonic models.

Table 1: Summary of hedonic indices and bias factors (Laspeyres)

Year	N_t	R_t^2	$\hat{\sigma}_t^2$	$I\hat{P}_{LS,L}$	$I\hat{B}_{LS,L}^A$	$I\hat{B}_{LS,L}^B$	$I\hat{P}_{ML,L}$	$I\hat{B}_{ML,L}^A$	$I\hat{B}_{ML,L}^B$	$I\hat{P}_L^A$	$I\hat{P}_L^B$
1990	1816	0.931	2.028	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1991	2270	0.935	1.844	1.020	0.998	1.096	0.930	0.911	1.000	1.021	0.930
1992	2400	0.938	1.507	1.042	0.997	1.298	0.803	0.769	1.000	1.045	0.803
1993	2751	0.946	0.925	1.048	0.997	1.736	0.604	0.574	1.000	1.051	0.604
1994	3095	0.946	0.856	1.045	0.996	1.797	0.581	0.554	1.000	1.048	0.581
1995	3282	0.946	0.781	1.023	0.996	1.866	0.549	0.534	1.000	1.027	0.548
1996	3696	0.944	0.690	0.993	0.996	1.952	0.509	0.510	1.000	0.998	0.509
1997	3885	0.951	0.524	0.984	0.996	2.121	0.464	0.469	1.000	0.989	0.464
1998	3697	0.952	0.598	0.972	0.995	2.045	0.475	0.487	1.000	0.977	0.475
1999	2937	0.955	0.752	0.963	0.995	1.893	0.509	0.526	1.000	0.968	0.509

Table 2: Summary of hedonic indices and bias factors (Paasche)

Year	N_t	R_t^2	$\hat{\sigma}_t^2$	$I\hat{P}_{LS,P}$	$I\hat{B}_{LS,P}^A$	$I\hat{B}_{LS,P}^B$	$I\hat{P}_{ML,P}$	$I\hat{B}_{ML,P}^A$	$I\hat{B}_{ML,P}^B$	$I\hat{P}_P^A$	$I\hat{P}_P^B$
1990	1816	0.931	2.028	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1991	2270	0.935	1.844	1.027	0.999	1.096	0.936	0.911	1.000	1.028	0.936
1992	2400	0.938	1.507	1.060	0.998	1.298	0.817	0.769	1.000	1.062	0.817
1993	2751	0.946	0.925	1.069	0.997	1.736	0.616	0.575	1.000	1.072	0.616
1994	3095	0.946	0.856	1.062	0.997	1.797	0.591	0.555	1.000	1.066	0.591
1995	3282	0.946	0.781	1.045	0.997	1.866	0.560	0.534	1.000	1.048	0.560
1996	3696	0.944	0.690	1.018	0.996	1.952	0.521	0.510	1.000	1.021	0.521
1997	3885	0.951	0.524	1.006	0.996	2.121	0.474	0.470	1.000	1.010	0.474
1998	3697	0.952	0.598	0.996	0.996	2.045	0.487	0.487	1.000	1.000	0.487
1999	2937	0.955	0.752	0.990	0.996	1.893	0.523	0.526	1.000	0.994	0.523

Straightforward correction of the year-to-year quality-adjusted price changes $\hat{P}_{WS,L}$ and $\hat{P}_{WS,P}$ in (19) by multiplying with $\exp\{\frac{1}{2}(\hat{\sigma}_1^2 - \hat{\sigma}_0^2)\}$ gives the maximum-likelihood based indices $\hat{P}_{WL,L}$ and $\hat{P}_{WL,P}$ defined in (23). The impact of the correction is huge: the maximum-likelihood based, quality-corrected price indices, $\hat{P}_{WL,L}$ and $\hat{P}_{WL,P}$, are below the least-squares based price indices, $\hat{P}_{WS,L}$ and $\hat{P}_{WS,P}$, in all years, pointing at a quality-corrected price level of passenger cars in 1999 of about half the price-level of new cars in 1990. There is little difference between the Laspeyres and Paasche results in this respect.

The extent of the biases depends on the intended use of the quality-corrected indices as estimators of either P_r^A in (17) or P_r^B in (18). The estimated expected values of both price indices $\hat{P}_r^A$ and $\hat{P}_r^B$ obtained after dividing either $\hat{P}_{WS,r}^A$ or $\hat{P}_{WL,r}^B$ by their respective bias factors, are shown in the last two columns of tables 1 and 2, while chained indices based on the year-to-year bias factors $\hat{B}_{WS,r}^A$, $\hat{B}_{WS,r}^B$, $\hat{B}_{WL,r}^A$ and $\hat{B}_{WL,r}^B$ are presented next to the corresponding price indices. The results show that the (weighted) least-squares price indices $\hat{P}_{WS,L}$ and $\hat{P}_{WS,P}$ are practically unbiased estimators of P_L^A and P_P^A, but are extremely upward biased estimators of P_L^B and P_P^B. In the former instance, the bias factor $\hat{B}_{WS,L}^A$ in 1999 is only 99.5 of the 1990 level, while in the latter instance the bias factor $\hat{B}_{WS,L}^B$ is 1.893, almost double the 1990 level. The outcomes for the maximum-likelihood based estimators $\hat{P}_{WL,L}$ and $\hat{P}_{WL,P}$ are almost reversed: these indices appear to be almost unbiased when estimating P_L^B and P_P^B, but are substantially biased when estimating P_L^A and P_P^A.

The main source of the biases is the factor $\exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\}$, whose estimates vary between 0.747 in 1993 and 1.080 in 1999. The other two factors are relatively small. The factor $\exp\{-\frac{1}{2}(\sigma_1^2 + \sigma_0^2)\mathbf{w}_r'\mathbf{w}_r\}$ is estimated between 0.998 and 1.000 for the Laspeyres-like indices and between 0.999 and 1.000 for the Paasche-like indices. The estimated factor $\exp\{\frac{1}{2}\sigma_1^2 h_{1r} + \frac{1}{2}\sigma_0^2 h_{0r}\}$ does not differ from 1 in the third decimal in any situation. The biases are therefore largely due to the dynamic performance of the hedonic regressions: the mean squared errors $\hat{\sigma}_t^2$ reveal considerable variation in time causing the least-squares and maximum likelihood based price indices to diverge. In this particular sample, the $\hat{\sigma}_t^2$ are seen to decrease in time until 1997, which makes that $\exp\{\frac{1}{2}(\hat{\sigma}_1^2 - \hat{\sigma}_0^2)\}$ is systematically below 1 (and $\hat{P}_{WL,r}$ below $\hat{P}_{WS,r}$) over the period.

In all, the results demonstrate the importance of explicating the object of estimation P_r^A or P_r^B as well as of properly estimating the hedonic relationships.

4 Matched-model and composite price indices

An often-used method to correct for quality differences between periods is to calculate matched-model price indices, which basically are weighted aggregates of price relatives of the relevant product varieties. The approach has intuitive appeal as it compares like with like (Silver and Heravi, 2001) and thereby effectively controls for quality variations, at least as far as the matched set is concerned. The hedonic methodology is sometimes used in conjunction with the matched-model approach to impute the missing prices (or price relatives) of entering and exiting product varieties. The combination of both approaches leads to so-called composite price indices that are based partly on observed price relatives and partly on hedonically imputed prices. The question is how the matched-model and composite price indices are affected by the use of the log-linear hedonic model (15).

4.1 Definition of the price indices

Different composite and matched-model price indices exist, which vary with respect to the use of weights, the period from which the weights are taken (base-period Laspeyres or current-period Paasche), and the type of price averages (arithmetic or geometric) applied. Here, we consider sales-weighted, geometrically-averaged Törnqvist-like indices, in which the price relatives are either partially imputed when missing (leading to the composite price index P^C) or wholly represented by their predicted values based on hedonic models (leading to a matched-model hedonic price index P^M). The weights are treated as exogenous to the price index construction: they are considered non-random and have no impact on the interpretation of the included prices. The first assumption serves to simplify the exposition for the common situation that weights are based on sales values which themselves contain a price component. The second assumption is necessary to avoid inconsistencies with the interpretation of the used prices implied by weighted regression of the hedonic model as explained in section 3.4.

The composite price index P^C

A Törnqvist-like composite price index P^C may be defined as:

$$P^C = \left(\prod_{i \in \mathcal{A}^1 \cap \mathcal{A}^0} \left(\frac{Y_{1i}}{Y_{0i}} \right)^{w_{0i}} \prod_{i \in \bar{\mathcal{A}}^1 \cap \mathcal{A}^0} \left(\frac{E(Y_{1i})}{Y_{0i}} \right)^{w_{0i}} \right)^{\frac{1}{2}} \times \left(\prod_{i \in \mathcal{A}^1 \cap \mathcal{A}^0} \left(\frac{Y_{1i}}{Y_{0i}} \right)^{w_{1i}} \prod_{i \in \mathcal{A}^1 \cap \bar{\mathcal{A}}^0} \left(\frac{Y_{1i}}{E(Y_{0i})} \right)^{w_{1i}} \right)^{\frac{1}{2}} \quad (29)$$

Here, $\mathcal{A}^1$ and $\mathcal{A}^0$ represent index sets of N_1 and N_0 product varieties offered for sale in the current (=1) and base (=0) periods. The union $\mathcal{A}^1 \cup \mathcal{A}^0$ of both index sets contains $N_0 + N_1 - N_b$ product varieties, of which N_b elements are in the set of resident varieties $\mathcal{A}^1 \cap \mathcal{A}^0$ that are sold in both periods, $N_b \leq N_0$ and $N_b \leq N_1$; $N_0 - N_b$ elements are in the set of exiting product varieties $\bar{\mathcal{A}}^1 \cap \mathcal{A}^0$; and $N_1 - N_b$ are in the set of entering product varieties $\mathcal{A}^1 \cap \bar{\mathcal{A}}^0$. The prices Y_{1i} and Y_{0i} are observable for all varieties i and periods except for the entering and exiting product varieties. In the latter case, the missing prices are represented by their expected values which will be based on the hedonic models formulated in (15). The weights w_{0i} and w_{1i} are sales value shares of product varieties in the base and current periods. These weights differ therefore from the volume weights used in (17) but have the same mathematical properties (that is, they are non-negative and add up to 1 over the base and current periods respectively). Accordingly, we use the same notation.
The expected value of P^C is determined by elaborating (29) in terms of the underlying model (15) and then applying the expectation operator:[4]

$$
\begin{aligned}
E(P^C) &= E\left(\exp\{\frac{1}{2}\sum_{i\in\mathcal{A}^0} w_{0i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+\frac{1}{2}\sum_{i\in\mathcal{A}^1\cap\mathcal{A}^0} w_{0i}\varepsilon_{1i}-\frac{1}{2}\sum_{i\in\mathcal{A}^0} w_{0i}\varepsilon_{0i}\}\times\right.\\
&\quad \left.\times\exp\{\frac{1}{2}\sum_{i\in\mathcal{A}^1} w_{1i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+\frac{1}{2}\sum_{i\in\mathcal{A}^1} w_{1i}\varepsilon_{1i}-\frac{1}{2}\sum_{i\in\mathcal{A}^1\cap\mathcal{A}^0} w_{1i}\varepsilon_{0i}\}\right)\\
&= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times\\
&\quad \times E\left(\exp\{\frac{1}{2}(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)'\dot{\boldsymbol{\varepsilon}}_1-\frac{1}{2}(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)'\dot{\boldsymbol{\varepsilon}}_0\}\right)\\
&= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times\exp\{-\frac{1}{2}(\sigma_1^2-\sigma_0^2)\}\times \qquad (30)\\
&\quad \times\exp\{\frac{1}{8}(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)'(\sigma_1^2\dot{\mathbf{Q}}_1^{-1}+\sigma_0^2\dot{\mathbf{Q}}_0^{-1})(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)\}
\end{aligned}
$$

The dotted symbols are $(N_0 + N_1 - N_b)$-vectors or matrices containing the original undotted information and zero's otherwise. Their elements correspond with the unique elements of the union of $\mathcal{A}^0$ and $\mathcal{A}^1$. Thus, $\dot{\mathbf{w}}_0' = (\mathbf{w}_0', \mathbf{0}')$ is a vector with base-period weights and zero's for the entering product varieties; $\dot{\mathbf{w}}_1 = (\mathbf{0}', \dot{\mathbf{w}}_1')$ is a vector with current-period weights and zero's for the exiting product varieties. Likewise, $\dot{\boldsymbol{\varepsilon}}_0' = (\boldsymbol{\varepsilon}_0', \mathbf{0}')$ and $\dot{\boldsymbol{\varepsilon}}_1' = (\mathbf{0}', \boldsymbol{\varepsilon}_1')$ represent inflated disturbance vectors, and $\dot{\mathbf{Q}}_1$ and $\dot{\mathbf{Q}}_0$ are inflated versions of the weight matrices $\mathbf{Q}_1$ and $\mathbf{Q}_0$ used in the hedonic model.

[4]It is not feasible in general to determine the expected values of the price indices as the product of the expected price relatives, because the price relatives are not independently distributed.

The expectation part of (30) has been calculated with the use of (4) taking $\mathbf{z}' = (\dot{\boldsymbol{\varepsilon}}_1', \dot{\boldsymbol{\varepsilon}}_0') \sim n(\boldsymbol{\delta}, \boldsymbol{\Delta})$ with $\boldsymbol{\delta}' = (-\frac{1}{2}\sigma_1^2\boldsymbol{\iota}_{01}', -\frac{1}{2}\sigma_0^2\boldsymbol{\iota}_{01}')$, $\boldsymbol{\Delta}$ a block-diagonal matrix with $\sigma_1^2\dot{\mathbf{Q}}_1^{-1}$ and $\sigma_0^2\dot{\mathbf{Q}}_0^{-1}$ on the main diagonal, and letting $\mathbf{t}' = (\frac{1}{2}(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)', -\frac{1}{2}(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)')$. Here, $\boldsymbol{\iota}_{01}$ is a $(N_0+N_1-N_b)$-vector. Note that $\dot{\mathbf{w}}_0'\boldsymbol{\iota}_{01} = \mathbf{w}_0'\boldsymbol{\iota}_0 = 1$ and $\dot{\mathbf{w}}_1'\boldsymbol{\iota}_{01} = \mathbf{w}_1'\boldsymbol{\iota}_1 = 1$. The variance of P^C can be obtained as: $V(P^C) = \exp\{(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)-(\sigma_1^2-\sigma_0^2)+\frac{1}{4}(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)'(\sigma_1^2\dot{\mathbf{Q}}_1^{-1}+\sigma_0^2\dot{\mathbf{Q}}_0^{-1})(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)\}(\exp\{\frac{1}{4}(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)'(\sigma_1^2\dot{\mathbf{Q}}_1^{-1}+\sigma_0^2\dot{\mathbf{Q}}_0^{-1})(\dot{\mathbf{w}}_1+\dot{\mathbf{w}}_0)\}-1)$, where use has been made of (5).

If all product varieties in the index are equally weighted with $w_{0i} = 1/N_0$ for $i \in \mathcal{A}^0$ and $w_{1i} = 1/N_1$ for $i \in \mathcal{A}^1$ and the product varieties in (15) are also not weighted, then the expected value of P^C in (30) becomes equal to $E(P^C) = \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times\exp\{-\frac{1}{2}(\sigma_1^2-\sigma_0^2)+\frac{1}{8}\sigma_1^2(\frac{1}{N_1}+2\frac{N_b}{N_0N_1}+\frac{N_b}{N_0^2})+\frac{1}{8}\sigma_0^2(\frac{1}{N_0}+2\frac{N_b}{N_0N_1}+\frac{N_b}{N_1^2})\}$. Moreover, if markets are perfectly stable, such that no exit of obsolete models nor entry of new product varieties occurs, $N_1 = N_0 = N_b$, then the expected value of the composite index P^C is equal to the Fisher-like $E(P^C) = P_L^{A\,\frac{1}{2}}P_P^{A\,\frac{1}{2}}$ based on (17).

The matched-model price index P^M

Analogous to (18), a matched-model hedonic price index P^M may also be defined as:

$$\begin{aligned} P^M &= \left(\prod_{i\in\mathcal{A}^0}\left(\frac{E(Y_{1i})}{E(Y_{0i})}\right)^{w_{0i}}\right)^{\frac{1}{2}}\left(\prod_{i\in\mathcal{A}^1}\left(\frac{E(Y_{1i})}{E(Y_{0i})}\right)^{w_{1i}}\right)^{\frac{1}{2}} \\ &= \exp\{\frac{1}{2}\sum_{i\in\mathcal{A}^0}w_{0i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+\frac{1}{2}\sum_{i\in\mathcal{A}^1}w_{1i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\} \\ &= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\} \end{aligned} \tag{31}$$

Note that $P^M = E(P^M)$ as P^M is a parameter and not a random variable. Moreover, the matched-model price index P^M in (31) is seen to be equal to the geometric average of the Laspeyres and Paasche-like hedonic indices P_L^B and P_P^B in (18): $P^M = P_L^{B\,\frac{1}{2}}P_P^{B\,\frac{1}{2}}$, and therefore again a Fisher-like price index. This correspondence is not affected by the entry or exit of product varieties.

4.2 Estimation of P^C and P^M

Estimates of the composite and matched-model price indices P^C and P^M can be obtained by substituting the least-squares or maximum-likelihood based predictions of the product variety price for the unobserved quantities in (29) and (31). The question is how these estimators (or predictors) of the price indices behave statistically. This section begins with the behavior of least-squares based predictors and continues with the maximum-likelihood based estimators.

Least-squares based estimators of P^C and P^M

Least-squares estimation of the hedonic equations in (15) for each period leads to estimates $\hat{Y}_{LS,1i}$ and $\hat{Y}_{LS,0i}$ of the expected prices of (missing) product varieties defined as $\hat{Y}_{LS,1i} = \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{LS,1} = \mathbf{x}_i'\boldsymbol{\beta}_1 + \mathbf{x}_i'(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\boldsymbol{\varepsilon}_1$ and $\hat{Y}_{LS,0i} = \mathbf{x}_i'\hat{\boldsymbol{\beta}}_{LS,0} = \mathbf{x}_i'\boldsymbol{\beta}_0 + \mathbf{x}_i'(\mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_0'\boldsymbol{\varepsilon}_0$. Inserting these predicted values into the composite price index P^C (29) gives the predicted quality-adjusted price index $\hat{P}^C_{LS}$ as:

$$\hat{P}^C_{LS} = \left(\prod_{i\in\mathcal{A}^1\cap\mathcal{A}^0} \left(\frac{Y_{1i}}{Y_{0i}}\right)^{w_{0i}} \prod_{i\in\bar{\mathcal{A}}^1\cap\mathcal{A}^0} \left(\frac{\hat{Y}_{1i}}{Y_{0i}}\right)^{w_{0i}} \right)^{\frac{1}{2}} \times \left(\prod_{i\in\mathcal{A}^1\cap\mathcal{A}^0} \left(\frac{Y_{1i}}{Y_{0i}}\right)^{w_{1i}} \prod_{i\in\mathcal{A}^1\cap\bar{\mathcal{A}}^0} \left(\frac{Y_{1i}}{\hat{Y}_{0i}}\right)^{w_{1i}} \right)^{\frac{1}{2}} \tag{32}$$

In order to determine the expected value of $\hat{P}^C_{LS}$, it is helpful to introduce the following notation. A $(N_0 + N_1 - N_b) \times (K + 1)$-matrix $\mathbf{X}$ with the characteristics of all product varieties ($i \in \mathcal{A}^1 \cup \mathcal{A}^0$) is defined such that the first $N_0 - N_b$ rows contain information about the exiting product varieties ($i \in \bar{\mathcal{A}}^1 \cap \mathcal{A}^0$), the next N_b rows contain information about the resident product varieties ($i \in \mathcal{A}^1 \cap \mathcal{A}^0$), and the last $N_1 - N_b$ rows that of the entering product varieties ($i \in \mathcal{A}^1 \cap \bar{\mathcal{A}}^0$). Three diagonal $(N_0 + N_1 - N_b)$-matrices $\mathbf{D}_b$, $\mathbf{D}_1$ and $\mathbf{D}_0$ are defined with one's on the main diagonal for the elements corresponding with product varieties available in both periods ($\mathbf{D}_b$), the base-period ($\mathbf{D}_0$) and the current period ($\mathbf{D}_1$), and with zero's elsewhere. These matrices serve to select the relevant information from the $(N_0 + N_1 - N_b)$ characteristics vectors of $\mathbf{X}$ and other vectors. It is easily verified that $\mathbf{D}_b = \mathbf{D}_0 + \mathbf{D}_1 - \mathbf{I}_{01}$ and that $\mathbf{D}_b\mathbf{D}_0 = \mathbf{D}_b\mathbf{D}_1 = \mathbf{D}_0\mathbf{D}_1 = \mathbf{D}_b$. Associated with these matrices are $(N_0 + N_1 - N_b)$-vectors $\mathbf{d}_b$, $\mathbf{d}_1$ and $\mathbf{d}_0$ defined as: $\mathbf{d}_b = \mathbf{D}_b\boldsymbol{\iota}_{01}$, $\mathbf{d}_1 = \mathbf{D}_1\boldsymbol{\iota}_{01}$ and $\mathbf{d}_0 = \mathbf{D}_0\boldsymbol{\iota}_{01}$. Note that $\mathbf{X}'\mathbf{D}_1\boldsymbol{\iota}_{01} = \mathbf{X}'\mathbf{d}_1 = \mathbf{X}_1'\boldsymbol{\iota}_1$ and $\mathbf{X}'\mathbf{D}_0\boldsymbol{\iota}_{01} = \mathbf{X}'\mathbf{d}_0 = \mathbf{X}_0'\boldsymbol{\iota}_0$, that $\mathbf{X}'\mathbf{D}_1\dot{\boldsymbol{\varepsilon}}_1 = \mathbf{X}_1'\boldsymbol{\varepsilon}_1$ and $\mathbf{X}'\mathbf{D}_0\dot{\boldsymbol{\varepsilon}}_0 = \mathbf{X}_0'\boldsymbol{\varepsilon}_0$, and that $\mathbf{X}'\mathbf{D}_1\dot{\mathbf{w}}_1 = \mathbf{X}_1'\mathbf{w}_1$ and $\mathbf{X}'\mathbf{D}_0\dot{\mathbf{w}}_0 = \mathbf{X}_0'\mathbf{w}_0$.

Using these conventions, the expected value of the least-squares based $\hat{P}^C_{LS}$ (32) can be determined as:

$$\begin{aligned}
E(\hat{P}^C_{LS}) &= E\Bigg(\exp\{\frac{1}{2}\sum_{\mathcal{A}^1\cap\mathcal{A}^0} w_{0i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+w_{0i}\varepsilon_{1i}-w_{0i}\varepsilon_{0i}\}\times \\
&\quad\times\exp\{\frac{1}{2}\sum_{\bar{\mathcal{A}}^1\cap\mathcal{A}^0} w_{0i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+w_{0i}\mathbf{x}_i'(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Q}_1\boldsymbol{\varepsilon}_1-w_{0i}\varepsilon_{0i}\}\times \\
&\quad\times\exp\{\frac{1}{2}\sum_{\mathcal{A}^1\cap\mathcal{A}^0} w_{1i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+w_{1i}\varepsilon_{1i}-w_{1i}\varepsilon_{0i}\}\times \\
&\quad\times\exp\{\frac{1}{2}\sum_{\mathcal{A}^1\cap\bar{\mathcal{A}}^0} w_{1i}\mathbf{x}_i'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)+w_{1i}\varepsilon_{1i}-w_{1i}\mathbf{x}_i'(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Q}_0\boldsymbol{\varepsilon}_0\}\Bigg) \\
&= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times \\
&\quad\times E\Bigg(\exp\{\frac{1}{2}(\dot{\mathbf{w}}_1'+\dot{\mathbf{w}}_0'\mathbf{D}_b+\dot{\mathbf{w}}_0'(\mathbf{D}_0-\mathbf{D}_b)\mathbf{X}(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}\mathbf{X}'\dot{\mathbf{Q}}_1)\dot{\varepsilon}_1\}\times \\
&\quad\times\exp\{-\frac{1}{2}(\dot{\mathbf{w}}_0'+\dot{\mathbf{w}}_1'\mathbf{D}_b+\dot{\mathbf{w}}_1'(\mathbf{D}_1-\mathbf{D}_b)\mathbf{X}(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}\mathbf{X}'\dot{\mathbf{Q}}_0)\dot{\varepsilon}_0\}\Bigg) \\
&= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times\exp\{-\frac{1}{2}(\sigma_1^2-\sigma_0^2)\}\times \qquad (33)\\
&\quad\times\exp\{\frac{1}{8}\sigma_1^2((\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)'\dot{\mathbf{Q}}_1^{-1}(\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)+g_{10})\} \\
&\quad\times\exp\{\frac{1}{8}\sigma_0^2((\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)'\dot{\mathbf{Q}}_0^{-1}(\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)+g_{01})\}
\end{aligned}$$

Use has been made of (4) with $\mathbf{z}=(\dot{\varepsilon}_1',\dot{\varepsilon}_0')$ as before and $\mathbf{t}'=(\frac{1}{2}(\dot{\mathbf{w}}_1'+\dot{\mathbf{w}}_0'\mathbf{D}_b+\dot{\mathbf{w}}_0'(\mathbf{D}_0-\mathbf{D}_b)\mathbf{X}(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}\mathbf{X}'\dot{\mathbf{Q}}_1),-\frac{1}{2}(\dot{\mathbf{w}}_0'+\dot{\mathbf{w}}_1'\mathbf{D}_b+\dot{\mathbf{w}}_1'(\mathbf{D}_1-\mathbf{D}_b)\mathbf{X}(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}\mathbf{X}'\dot{\mathbf{Q}}_0))$. The terms $(\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)'\dot{\mathbf{Q}}_1^{-1}(\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)=\dot{\mathbf{w}}_0'\mathbf{D}_b\dot{\mathbf{Q}}_1^{-1}\mathbf{D}_b\dot{\mathbf{w}}_0+2\dot{\mathbf{w}}_0'\mathbf{D}_b\dot{\mathbf{Q}}_1^{-1}\dot{\mathbf{w}}_1+\mathbf{w}_1'\mathbf{Q}_1^{-1}\mathbf{w}_1$ and $(\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)'\dot{\mathbf{Q}}_0^{-1}(\dot{\mathbf{w}}_0+\dot{\mathbf{w}}_1)=\dot{\mathbf{w}}_1'\mathbf{D}_b\dot{\mathbf{Q}}_0^{-1}\mathbf{D}_b\dot{\mathbf{w}}_1+2\dot{\mathbf{w}}_0'\dot{\mathbf{Q}}_0^{-1}\mathbf{D}_b\dot{\mathbf{w}}_1+\mathbf{w}_0'\mathbf{Q}_0^{-1}\mathbf{w}_0$ reflect sum of squared index weights variably summated over the common or the whole set of product varieties in the current and base periods.

The scalar $g_{10}=[\dot{\mathbf{w}}_0'(\mathbf{D}_b+\mathbf{D}_0)\mathbf{X}+2\mathbf{w}_1'\mathbf{X}_1](\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}\mathbf{X}'(\mathbf{D}_0-\mathbf{D}_b)\dot{\mathbf{w}}_0$ may be interpreted as a standardized distance between the base-period average characteristics of the exiting product varieties ($\mathbf{X}'(\mathbf{D}_0-\mathbf{D}_b)\dot{\mathbf{w}}_0$) and the weighted sum of the base-period average characteristics of the resident varieties ($\dot{\mathbf{w}}_0'\mathbf{D}_b\mathbf{X}$), the base-period average of all base-period varieties ($\dot{\mathbf{w}}_0'\mathbf{D}_0\mathbf{X}=\mathbf{w}_0'\mathbf{X}_0$) and twice the current-period average of all current-period varieties ($2\mathbf{w}_1'\mathbf{X}_1$) in the space defined by the current period characteristics. A similar but reverse interpretation holds for the scalar $g_{01}=(\dot{\mathbf{w}}_1'(\mathbf{D}_b+\mathbf{D}_1)\mathbf{X}+2\mathbf{w}_0'\mathbf{X}_0)(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}\mathbf{X}'(\mathbf{D}_1-\mathbf{D}_b)\dot{\mathbf{w}}_1$. The scalars g_{10} and $g_{01}$are both zero when no entry or exit of product varieties occurs. In this specific case, $E(\hat{P}^C_{LS})$ is equal to the expected value of P^C, $E(P^C_{LS})$ in (30) and to the hedonic index P^A_r in (17) with $\mathbf{x}_r=\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)$. Calculation of the variance of $\hat{P}^C_{LS}$ is suspended.

A least-squares based estimate $\hat{P}^M_{LS}$ of the matched-model price index P^M is obtained by inserting the predicted hedonic prices $\hat{Y}_{LS,1i}$ and $\hat{Y}_{LS,0i}$ for all the prices in (31):

$$\begin{aligned}\hat{P}^M_{LS} &= \left(\prod_{i\in\mathcal{A}^0}\left(\frac{\hat{Y}_{1i}}{\hat{Y}_{0i}}\right)^{w_{0i}}\right)^{\frac{1}{2}}\left(\prod_{i\in\mathcal{A}^1}\left(\frac{\hat{Y}_{1i}}{\hat{Y}_{0i}}\right)^{w_{1i}}\right)^{\frac{1}{2}} \\ &= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\hat{\boldsymbol{\beta}}_{LS,1}-\hat{\boldsymbol{\beta}}_{LS,0})\}\end{aligned} \tag{34}$$

The expected value of $\hat{P}^M_{LS}$ can be determined as:

$$\begin{aligned}E(\hat{P}^M_{LS}) &= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times \\ &\times E\Bigg(\exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Q}_1\boldsymbol{\varepsilon}_1 \\ &-\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Q}_0\boldsymbol{\varepsilon}_0\}\Bigg) \\ &= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times\exp\{-\frac{1}{2}(\sigma_1^2-\sigma_0^2)\}\times \\ &\times\exp\{\frac{1}{8}\sigma_1^2(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)\} \\ &\times\exp\{\frac{1}{8}\sigma_0^2(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)\}\end{aligned} \tag{35}$$

The moment-generating function (4) has been applied to calculate the expectation term with $\mathbf{z} = (\boldsymbol{\varepsilon}_1', \boldsymbol{\varepsilon}_0')$ and the vector $\mathbf{t}' = (\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Q}_1, -\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}\mathbf{X}_0'\mathbf{Q}_0)$.
The variance of $\hat{P}^M_{LS}$ can be determined as: $V(\hat{P}^M_{LS}) = \exp\{(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\}\times \exp\{-\sigma_1^2+\sigma_0^2\}\times \exp\{\frac{1}{4}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\sigma_1^2(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}+\sigma_0^2(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1})(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)\}\times (\exp\{\frac{1}{4}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\sigma_1^2(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}+\sigma_0^2(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1})(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)\}-1)$.

Maximum-likelihood based estimators of P^C and P^M

In addition, maximum-likelihood can be used to estimate (15) and to find the (expected) prices in (29) and (31) as $\hat{Y}_{ML,1i}$ and $\hat{Y}_{ML,0i}$. In light of $\hat{\boldsymbol{\beta}}_{ML} = \hat{\boldsymbol{\beta}}_{LS} + \sigma^2\mathbf{i}_0$ (12), these maximum-likelihood based price predictions can be written in terms of the least-squares predicted prices as: $\hat{Y}_{ML,1i} = \exp\{\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{ML,1}\} = \exp\{\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{LS,1}+\frac{1}{2}\sigma_1^2\} = \hat{Y}_{LS,1i}\exp\{\frac{1}{2}\sigma_1^2\}$ and $\hat{Y}_{ML,0i} = \hat{Y}_{LS,0i}\exp\{\frac{1}{2}\sigma_0^2\}$. Using these expressions, the maximum-likelihood based price index $\hat{P}^C_{ML}$ can be reformulated in terms of the least-squares based estimate $\hat{P}^C_{LS}$ as:

$$\hat{P}^C_{ML} = \hat{P}^C_{LS}\times\exp\{\frac{1}{4}\sigma_1^2\dot{\mathbf{w}}_0'(\mathbf{d}_0-\mathbf{d}_b)-\frac{1}{4}\sigma_0^2\dot{\mathbf{w}}_1'(\mathbf{d}_1-\mathbf{d}_b)\} \tag{36}$$

If the price relatives are equally weighted with $w_{0i} = 1/N_0$ and $w_{1i} = 1/N_1$, then $\hat{P}^C_{ML} = \hat{P}^C_{LS} \times \exp\{\frac{1}{4}\sigma_1^2(1-\frac{N_b}{N_0}) - \frac{1}{4}\sigma_0^2(1-\frac{N_b}{N_1})\}$. The two estimated price indices $\hat{P}^C_{ML}$ and $\hat{P}^C_{LS}$ are equal, when there is no entry or exit of product varieties, which makes intuitively sense because in this situation no imputation is performed.
The expected value of $\hat{P}^C_{ML}$ is determined with the help of (36) and (33) as:

$$E(\hat{P}^C_{ML}) = E(\hat{P}^C_{LS})\exp\{\frac{1}{4}\sigma_1^2\dot{\mathbf{w}}_0'(\mathbf{d}_0-\mathbf{d}_b) - \frac{1}{4}\sigma_0^2\dot{\mathbf{w}}_1'(\mathbf{d}_1-\mathbf{d}_b)\} \tag{37}$$

The variance of $\hat{P}^C_{ML}$ is related to the variance of $\hat{P}^C_{LS}$ as: $V(\hat{P}^C_{ML}) = V(\hat{P}^C_{LS}) \times \exp\{\frac{1}{2}\sigma_1^2\dot{\mathbf{w}}_0'(\mathbf{d}_0-\mathbf{d}_b) - \frac{1}{2}\sigma_0^2\dot{\mathbf{w}}_1'(\mathbf{d}_1-\mathbf{d}_b)\}$.
Furthermore, the unobserved expected prices of the matched-model price index P^M in (31) can be estimated by the maximum-likelihood predicted prices giving $\hat{P}^M_{ML}$. As $\hat{Y}_{ML,1i} = \hat{Y}_{LS,1i}\exp\{\frac{1}{2}\sigma_1^2\}$ and $\hat{Y}_{ML,0i} = \hat{Y}_{LS,0i}\exp\{\frac{1}{2}\sigma_0^2\}$ following (12), $\hat{P}^M_{ML}$ can be written in terms of the least-squares based matched-model index $\hat{P}^M_{LS}$ (34) as:

$$\hat{P}^M_{ML} = \hat{P}^M_{LS} \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \tag{38}$$

The expected value of $\hat{P}^M_{ML}$ is obtained via (35) and (38) as:

$$\begin{aligned} E(\hat{P}^M_{ML}) &= E(\hat{P}^M_{LS})\exp\{\frac{1}{2}(\sigma_1^2-\sigma_0^2)\} \\ &= \exp\{\frac{1}{2}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\boldsymbol{\beta}_1-\boldsymbol{\beta}_0)\} \\ &\quad \times\exp\{\frac{1}{8}\sigma_1^2(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)\} \\ &\quad \times\exp\{\frac{1}{8}\sigma_0^2(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)'(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1}(\bar{\mathbf{x}}_1+\bar{\mathbf{x}}_0)\} \end{aligned} \tag{39}$$

The variance of $\hat{P}^M_{ML}$ is equal to $V(\hat{P}^M_{ML}) = V(\hat{P}^M_{LS}) \times \exp\{\sigma_1^2 - \sigma_0^2\}$.

4.3 Bias factors associated with the estimation of P^C and P^M

The extent of the biases associated with the estimated composite and matched-model prices indices, $\hat{P}^C_{LS}$, $\hat{P}^C_{ML}$, $\hat{P}^M_{LS}$ and $\hat{P}^M_{ML}$, depends on the specific price index selected but also on the entity that is to be estimated P^C or P^M. The combination of four predictors and two predictands leads to 8 multiplicative bias factors defined as in (21), (22), (25) and (26).
The bias factor $B^C_{LS,C}$ associated with the estimation of the composite price index P^C with the least-squares composite index $\hat{P}^C_{LS}$ is defined as the ratio of the expected values of the two quantities involved, (33) and (30):

$$\begin{aligned} B^C_{LS,C} &= \frac{E(\hat{P}^C_{LS})}{E(P^C)} \\ &= \exp\{\frac{1}{8}\sigma_1^2 g_{10} + \frac{1}{8}\sigma_0^2 g_{01}\} \end{aligned} \tag{40}$$

If there is no entry or exit of product varieties and g_{10}= g_{01} = 0, then $B^C_{LS,C}$=1 and $\hat{P}^C_{LS}$ is an unbiased estimator of $E(P^C)$. This is in fact a trivial case, in which imputation is not an issue: $\hat{P}^C_{LS}$ = P^C, even without taking expectations. The more relevant question is of course whether $E(P^C)$ in (30) really is the quantity to be estimated.
The corresponding bias factor associated with the use of the maximum-likelihood based composite price index, $B^C_{ML,C}$, is found by relating the expected value of $\hat{P}^C_{ML}$ in (37) and $E(P^C)$ in (30):

$$\begin{aligned} B^C_{ML,C} &= \frac{E(\hat{P}^C_{ML})}{E(P^C)} \qquad (41) \\ &= B^C_{LS,C} \times \exp\{\frac{1}{4}\sigma_1^2\dot{\mathbf{w}}_0'(\mathbf{d}_0 - \mathbf{d}_b) - \frac{1}{4}\sigma_0^2\dot{\mathbf{w}}_1'(\mathbf{d}_1 - \mathbf{d}_b)\} \end{aligned}$$

In the absence of market dynamics, $\mathbf{d}_0 = \mathbf{d}_1 = \mathbf{d}_b$ and $B^C_{ML,C}$ is equal to 1 implying that $\hat{P}^C_{ML}$ like $\hat{P}^C_{LS}$ is an unbiased estimator of $E(P^C)$.
In addition, the composite price index P^C can be estimated with the matched-model price indices $\hat{P}^M_{LS}$ and $\hat{P}^M_{ML}$, see (35) and (39), which involves bias factors $B^C_{LS,M}$ and $B^C_{ML,M}$:

$$\begin{aligned} B^C_{LS,M} &= \frac{E(\hat{P}^M_{LS})}{E(P^C)} \qquad (42) \\ &= \exp\{\frac{1}{8}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)'(\sigma_0^2(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1} + \sigma_1^2(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1})(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)\} \\ &\quad \times \exp\{-\frac{1}{8}(\dot{\mathbf{w}}_1 + \dot{\mathbf{w}}_0)'(\sigma_1^2\dot{\mathbf{Q}}_1^{-1} + \sigma_0^2\dot{\mathbf{Q}}_0^{-1})(\dot{\mathbf{w}}_1 + \dot{\mathbf{w}}_0)\} \\ B^C_{ML,M} &= \frac{E(\hat{P}^M_{ML})}{E(P^C)} = B^C_{LS,M} \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \qquad (43) \end{aligned}$$

Calculation of the corresponding mean square errors of these estimators provides insight into the relative efficiency of these estimators, but is left as a future exercise.
Moreover, the biases involved with estimation of the matched-model price index P^M with the least-squares based $\hat{P}^M_{LS}$ and the maximum-likelihood based $\hat{P}^M_{ML}$ can be expressed as multiplicative factors $B^M_{LS,M}$ and $B^M_{ML,M}$ obtained by dividing the

expected values of $\hat{P}^M_{LS}$ in (35) and $\hat{P}^M_{ML}$ in (39) by that of P^M in (31):

$$\begin{aligned}
B^M_{LS,M} &= \frac{E(\hat{P}^M_{LS})}{P^M} \\
&= \exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \times \qquad (44)\\
&\quad \times \exp\{\frac{1}{8}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)'(\sigma_0^2(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1} + \sigma_1^2(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1})(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)\} \\
B^M_{ML,M} &= \frac{E(\hat{P}^M_{ML})}{P^M} \\
&= B^M_{LS,M} \times \exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\} \qquad (45)\\
&= \exp\{\frac{1}{8}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)'(\sigma_0^2(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1} + \sigma_1^2(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1})(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)\}
\end{aligned}$$

Finally, the matched-model index P^M can be estimated with the least-squares based composite index $\hat{P}^C_{LS}$ and the maximum-likelihood based composite index $\hat{P}^C_{ML}$ which gives rise to bias factors $B^M_{LS,C}$ and $B^M_{ML,C}$ determined in line with (33) and (37) and (31) as:

$$\begin{aligned}
B^M_{LS,C} &= \frac{E(\hat{P}^C_{LS})}{P^M} \\
&= \exp\{-\frac{1}{2}(\sigma_1^2 - \sigma_0^2) + \frac{1}{8}\sigma_1^2 g_{10} + \frac{1}{8}\sigma_0^2 g_{01}\} \times \qquad (46)\\
&\quad \times \exp\{+\frac{1}{8}(\dot{\mathbf{w}}_1 + \dot{\mathbf{w}}_0)'(\sigma_1^2\dot{\mathbf{Q}}_1^{-1} + \sigma_0^2\dot{\mathbf{Q}}_0^{-1})(\dot{\mathbf{w}}_1 + \dot{\mathbf{w}}_0)\} \\
B^M_{ML,C} &= \frac{E(\hat{P}^C_{ML})}{P^M} \\
&= B^M_{LS,C} \times \exp\{\frac{1}{4}\sigma_1^2\dot{\mathbf{w}}_0'(\mathbf{d}_0 - \mathbf{d}_b) - \frac{1}{4}\sigma_0^2\dot{\mathbf{w}}_1'(\mathbf{d}_1 - \mathbf{d}_b)\}
\end{aligned}$$

The next section briefly reflects on the various sources of the bias factors.

4.4 Bias implications

The various bias factors depend on five typical elements that each have a specific influence on the size of the biases. These elements are:

- $\exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\}$, which distinguishes the estimated $\hat{P}^M_{ML}$ in (38) from $\hat{P}^M_{LS}$ in (34). It is part of the bias factors $B^C_{ML,M}$ in (43), $B^M_{LS,M}$ in (44), $B^M_{LS,C}$ in (46) and $B^M_{ML,C}$ in (47). The nature of the bias, above or below 1, is not certain in advance as it depends on the temporal difference between the error variances, $\sigma_1^2 - \sigma_0^2$.
- $\exp\{\frac{1}{4}\sigma_1^2\dot{\mathbf{w}}_0'(\mathbf{d}_0 - \mathbf{d}_b) - \frac{1}{4}\sigma_0^2\dot{\mathbf{w}}_1'(\mathbf{d}_1 - \mathbf{d}_b)\}$, which distinguishes the estimated $\hat{P}^C_{ML}$ in (36) from $\hat{P}^C_{LS}$ in (32). It is part of the bias factors $B^C_{ML,C}$ in (41) and $B^M_{ML,C}$ in

(47). It depends on the temporal difference between the error variances, σ_1^2 and σ_0^2, corrected for the sum of base-period index weights of the exiting product varieties and the sum of current-period index weights of the entering product varieties.

- $\exp\{\frac{1}{8}(\sigma_0^2 g_{01} + \sigma_1^2 g_{10})\}$ defined below (33), which is related with the imputation of missing prices in the composite price index. It is therefore present in all bias factors involving the estimated composite price index: $B^C_{LS,C}$ in (40), $B^C_{ML,C}$ in (41), $B^M_{LS,C}$ in (46) and $B^M_{ML,C}$ in (47). As g_{10} and g_{01} can be interpreted as standardized distances, which are positive by definition, this particular element of the bias factors is always larger than 1.

- $\exp\{\frac{1}{8}(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)'(\sigma_0^2(\mathbf{X}_0'\mathbf{Q}_0\mathbf{X}_0)^{-1} + \sigma_1^2(\mathbf{X}_1'\mathbf{Q}_1\mathbf{X}_1)^{-1})(\bar{\mathbf{x}}_1 + \bar{\mathbf{x}}_0)\}$, which is associated with (expected values of) the estimated matched-model price indices, $E(\hat{P}^M_{LS})$ in (35) and $E(\hat{P}^M_{ML})$ in (39), and therefore part of the bias factors: $B^C_{LS,M}$ in (42), $B^C_{ML,M}$ in (43), $B^M_{LS,M}$ in (44) and $B^M_{ML,M}$ in (45). The argument of the exponent can be interpreted as a sum of two squared standardized distances and is accordingly always positive. The exponent causes an upward bias on the corresponding estimated price indices.

- $\exp\{\frac{1}{8}(\dot{\mathbf{w}}_1 + \dot{\mathbf{w}}_0)'(\sigma_1^2\dot{\mathbf{Q}}_1^{-1} + \sigma_0^2\dot{\mathbf{Q}}_0^{-1})(\dot{\mathbf{w}}_1 + \dot{\mathbf{w}}_0)\}$ is the exponent of the sum of sums of squared weights weighted by the variances of the disturbances of the hedonic models. It is part of the composite indices and absent for the matched-model indices. It therefore affects all situations in which composite price indices are applied to estimate P^M or matched-model indices are used to estimate $E(P^C)$: $B^C_{LS,M}$ in (42), $B^C_{ML,M}$ in 43), $B^M_{LS,C}$ in (46) and $B^M_{ML,C}$ in (47). Being a sum of sum squares, this element has an upward effect on the bias factors (depending on a possible different plus or minus sign associated with it).

The second and third bias parts vary with the extent of entry and exit. If markets are perfectly stable with no exiting or entering product varieties, then these bias elements would be equal to 1 (that is, they have no effect on the bias). In this unique instance, the estimated composite price indices $P_{LS,C}$ and $P_{ML,C}$ are equal to P^C, and accordingly unbiased. The third, fourth and fifth bias parts are exponents of sums of squares, which all have a positive influence on the total amount of bias (unless accompanied with a minus sign). The size of these three bias elements cannot be indicated in advance. The first bias part, is a familiar quantity, which is also part of the biases associated with the estimation of P^A and P^B. As before, the extent and direction of this particular bias element depends on the performance of the hedonic models in time.

4.5 Some empirical results of the biases

The size of the various biases is again illustrated with an example of the price developments of new passenger cars in the Netherlands, 1990-1999. Referring to table 5 in Van Dalen and Bode (2004), table 3 below shows the least-squares based chained composite price index $I\hat{P}^C_{LS}$ (32) and table 4 the least-squares based matched-model index $I\hat{P}^M_{LS}$ (34). The indices have been obtained by using sales-volume weights in the regression analyses and sales-value weights in the calculation of the indices. Conform the results in tables 1 and 2, the estimated quality-adjusted price indices are seen to increase until 1993 or 1994 and to decrease thereafter. The matched-model index is below the composite index in all years and ends at 1.03 in 1999, about 5 percentage points below the latter.

The maximum-likelihood based indices strongly deviate from the least-squares based estimates. The chained maximum-likelihood based, matched-model index $I\hat{P}^M_{ML}$ reveals a 1999 quality-corrected price level that is about half the price level of new passenger cars in 1990. This is in line with the results of $I\hat{P}^A_{ML}$ and $I\hat{P}^B_{ML}$ in tables 1 and 2, but has rather dramatic implications. Even in the case of the maximum-likelihood based composite price index $I\hat{P}^C_{ML}$, the estimated price index values are below the least-squares based results $I\hat{P}^C_{LS}$ and the quality-adjusted price level in 1999 is substantially (about 21%) below the 1990 price level.

The pattern of the bias factors observed in tables 3 and 4 is roughly the same is that observed in tables 1 and 2. The bias factors associated with the least-squares based estimate of the composite price index $I\hat{B}^C_{LS,C}$ in table 3 and those with the maximum-likelihood based matched model price index $I\hat{B}^M_{ML,M}$ in table 4 appear to be negligible. Conversely, the bias factors associated with the least-squares based estimated of the matched-model index $I\hat{B}^M_{LS,M}$ in table 4 and with the maximum-likelihood based estimate of the composite index $I\hat{B}^C_{ML,C}$ in table 3 are substantial. The largest bias factors for both indices are observed in 1997, where $I\hat{B}^M_{LS,M}$ is equal to 2.126 and $I\hat{B}^C_{ML,C}$ is equal to 0.766.

In addition, tables 3 and 4 present the bias factors associated with 'wrongly' applying composite price indices to estimate the matched-model index P^M in (31), $I\hat{B}^M_{LS,C}$ and $I\hat{B}^M_{ML,C}$, and with applying matched-model price indices to estimate the composite price index P^C in (29), $I\hat{B}^C_{LS,M}$ and $I\hat{B}^C_{ML,M}$. The outcomes suggest that the bias involved with using the least-squares based matched-model index to estimate P^C, $I\hat{B}^C_{LS,M}$ is negligible, while the biases associated with the other three combinations are substantial. Using the composite price indices $I\hat{P}^C_{LS}$ and $I\hat{P}^C_{ML}$ to estimate

Table 3: Törnqvist-like composite price indices and bias factors

Year	$I\hat{P}^C_{LS}$	$IB^C_{LS,C}$	$I\hat{P}^C_{ML}$	$IB^C_{ML,C}$	$I\hat{P}^C$	$IB^C_{LS,M}$	$IB^C_{ML,M}$
1990	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1991	1.021	1.000	0.943	0.923	1.021	1.000	0.909
1992	1.065	1.000	0.954	0.896	1.065	1.000	0.774
1993	1.088	1.000	0.910	0.836	1.088	1.000	0.579
1994	1.093	1.000	0.890	0.814	1.093	1.000	0.563
1995	1.089	1.000	0.866	0.795	1.089	1.000	0.536
1996	1.080	1.000	0.846	0.783	1.080	1.000	0.514
1997	1.079	1.000	0.827	0.766	1.079	1.000	0.470
1998	1.079	1.000	0.830	0.769	1.079	1.000	0.487
1999	1.082	1.000	0.854	0.789	1.082	1.000	0.531

Table 4: Törnqvist-like matched-model price indices and bias factors

Year	$I\hat{P}^M_{LS}$	$IB^M_{LS,M}$	$I\hat{P}^M_{ML}$	$IB^M_{ML,M}$	$I\hat{P}^M$	$IB^M_{LS,C}$	$IB^M_{ML,C}$
1990	1.000	1.000	1.000	1.000	1.000	1.000	1.000
1991	1.025	1.100	0.932	1.000	0.932	1.100	1.016
1992	1.063	1.293	0.822	1.000	0.822	1.293	1.158
1993	1.079	1.726	0.625	1.000	0.625	1.726	1.443
1994	1.079	1.778	0.607	1.000	0.607	1.778	1.447
1995	1.063	1.866	0.569	1.000	0.569	1.867	1.484
1996	1.041	1.946	0.535	1.000	0.535	1.946	1.524
1997	1.038	2.126	0.488	1.000	0.488	2.126	1.630
1998	1.032	2.054	0.502	1.000	0.502	2.054	1.579
1999	1.030	1.885	0.546	1.000	0.546	1.885	1.487

P^M is seen to lead to systematic and substantial over-estimation, while the use of the maximum-likelihood based matched model index $I\hat{P}^M_{ML}$ to estimate P^C leads to under-estimation.

5 Pooled hedonic indices

A related approach to the estimation of quality-adjusted price indices is based on the repackaging theory of Fisher and Shell (1971). It amounts to estimating a hedonic relationship using pooled information about prices and product characteristics of two (or more) adjacent periods. The quality-adjusted price index estimate is obtained as the (exponent of the) estimate of a dummy variable that identifies the constituting periods. More accurately, let $\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}_1$, $\boldsymbol{\varepsilon}_1 \sim n(-\frac{1}{2}\sigma_1^2\boldsymbol{\iota}_1, \sigma_1^2\mathbf{I}_1)$, and $\mathbf{y}_0 = \mathbf{X}_0\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_0$, $\boldsymbol{\varepsilon}_0 \sim n(-\frac{1}{2}\sigma_0^2\boldsymbol{\iota}_0, \sigma_0^2\mathbf{I}_0)$, the matrix representations of the hedonic equations (15) with $\mathbf{X}_1 = (\boldsymbol{\iota}_1, \mathbf{X}_{r1})$ and $\mathbf{X}_0 = (\boldsymbol{\iota}_0, \mathbf{X}_{r0})$ - the subscript r stands for 'remaining characteristics'. The pooled hedonic model can then be written as:

$$\mathbf{y} = \mathbf{d}_1\pi + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{W}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim n(-\frac{1}{2}\sigma^2\boldsymbol{\iota}, \sigma^2\mathbf{I}) \tag{47}$$

where the $(N_1 + N_0)$-vector $\mathbf{y}' = (\mathbf{y}_1', \mathbf{y}_0')$ contains the log-price information, $\mathbf{X}$ is the stacked $(N_1 + N_0) \times (K + 1)$-matrix with product characteristics $\mathbf{X}' = (\mathbf{X}_1', \mathbf{X}_0')$ (including the vector of 1's in the first column), the $(N_1 + N_0)$-vector $\mathbf{d}_1' = (\boldsymbol{\iota}_1', \mathbf{0}')$ is a dummy vector with 1's for product varieties in the current period and 0's for product varieties in the base period. The $(N_1 + N_0) \times (K + 2)$-matrix $\mathbf{W} = (\mathbf{d}_1, \mathbf{X})$ and the $(K + 2)$-vector $\boldsymbol{\eta}' = (\pi, \boldsymbol{\beta}')$ have been introduced to simplify further analysis. In the pooled specification (47), π is the parameter of interest since $\exp\{\pi\}$ is the ratio of the expected price levels in the two periods for a given set of product characteristics:

$$\begin{aligned} P^P &= \exp\{E\,(y|d_1 = 1, \mathbf{x} = \mathbf{x}^*) - E\,(y|d_1 = 0, \mathbf{x} = \mathbf{x}^*)\} \\ &= \exp\{\pi + \mathbf{x}^{*\prime}\boldsymbol{\beta} - \mathbf{x}^{*\prime}\boldsymbol{\beta}\} \\ &= \exp\{\pi\} \end{aligned} \tag{48}$$

Comparing the pooled model (47) with the hedonic relationships specified in (15) shows two important restrictions: $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$ and $\sigma^2 = \sigma_1^2 = \sigma_0^2$. The first restriction, $\boldsymbol{\beta} = \boldsymbol{\beta}_1 = \boldsymbol{\beta}_0$, states that in the pooled model the effects of quality on prices are assumed to be the same in both periods. The assumption is usually evaluated with a multiple F-test; see Hall (1971) for an early example, and Berndt, Griliches and Rappaport (1995) for recent applications. A practical implication of the restriction is that a main source of bias in the various hedonic price indices, the part related with the differences between $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$, is non-existent by assumption. The second restriction, $\sigma^2 = \sigma_1^2 = \sigma_0^2$, means that the amount of unexplained price variation among product varieties within each period is the same. The assumption is rarely explicitly tested.[5] The implication of the restriction again is that an important source of biases in the hedonic price indices, related with the difference between σ_1^2 and σ_0^2, is set aside by assumption. In light of these assumptions, it is obvious to expect that the amount of bias observed for the pooled hedonic price index is smaller than the biases observed for the various hedonic indices.

The parameter π can be estimated unbiasedly by the first element $\hat{\pi}$ of the least squares estimator $\hat{\boldsymbol{\eta}} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{y} = \boldsymbol{\eta} + (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\varepsilon$. The estimator of the quality-adjusted price is $\exp\{\hat{\pi}\}$, which has expected value $\exp\{\pi + \frac{1}{2}\sigma_{11}^2\}$, where σ_{11}^2 is the first element of the estimated covariance matrix $\sigma^2(\mathbf{W}'\mathbf{W})^{-1}$ as explained below (9). The multiplicative bias factor B^P associated with the pooled estimation of the quality-adjusted price change, $\exp\{\pi\}$, is therefore equal to:

$$B^P = \exp\{\frac{1}{2}\sigma_{11}^2\} \tag{49}$$

The size of this bias can be assessed by elaborating the variance-covariance matrix of $\hat{\boldsymbol{\eta}}$, which, following the partitioning of $\mathbf{W} = (\mathbf{d}_1, \mathbf{X})$, reads as:

$$\sigma^2(\mathbf{W}'\mathbf{W})^{-1} = \begin{pmatrix} \mathbf{d}_1'\mathbf{d}_1 & \mathbf{d}_1'\mathbf{X} \\ \mathbf{X}'\mathbf{d}_1 & \mathbf{X}'\mathbf{X} \end{pmatrix}^{-1} = \begin{pmatrix} N_1 & \boldsymbol{\iota}_1'\mathbf{X}_1 \\ \mathbf{X}_1'\boldsymbol{\iota}_1 & \mathbf{X}_1'\mathbf{X}_1 + \mathbf{X}_0'\mathbf{X}_0 \end{pmatrix}^{-1} \tag{50}$$

The first diagonal element of this matrix can be determined with the rules for inverting partitioned matrices as:

$$\begin{aligned} \sigma_{11}^2 &= \sigma^2\frac{1}{N_1} - \sigma^2\frac{1}{N_1^2}\boldsymbol{\iota}_1'\mathbf{X}_1(\frac{1}{N_1}\mathbf{X}_1'\boldsymbol{\iota}_1\boldsymbol{\iota}_1'\mathbf{X}_1 - \mathbf{X}_1'\mathbf{X}_1 - \mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_1'\boldsymbol{\iota}_1 \\ &= \sigma^2\frac{1}{N_1} + \sigma^2\frac{1}{N_1^2}\boldsymbol{\iota}_1'\mathbf{X}_1(\mathbf{X}_1'\mathbf{M}_1\mathbf{X}_1 + \mathbf{X}_0'\mathbf{X}_0)^{-1}\mathbf{X}_1'\boldsymbol{\iota}_1 \end{aligned} \tag{51}$$

with $\mathbf{M}_1 = \mathbf{I}_1 - \boldsymbol{\iota}_1\boldsymbol{\iota}_1'/N_1$ an idempotent $(N_1 \times N_1)$-matrix. Since $\mathbf{X}_1 = (\boldsymbol{\iota}_1, \mathbf{X}_{r1})$ has a vector of 1's in the first column, the $(K+1)$-matrix $\mathbf{X}_1'\mathbf{M}_1\mathbf{X}_1$ is block diagonal with a K-matrix $\mathbf{X}_{r1}'\mathbf{M}_1\mathbf{X}_{r1}$ in the right corner and 0's otherwise. The $(K+1)$-matrix $(\mathbf{X}_1'\mathbf{M}_1\mathbf{X}_1 + \mathbf{X}_0'\mathbf{X}_0)^{-1}$ can therefore be written as:

$$(\mathbf{X}_1'\mathbf{M}_1\mathbf{X}_1 + \mathbf{X}_0'\mathbf{X}_0)^{-1} = \begin{pmatrix} N_0 & \boldsymbol{\iota}_0'\mathbf{X}_{r0} \\ \mathbf{X}_{r0}'\boldsymbol{\iota}_0 & \mathbf{X}_{r1}'\mathbf{M}_1\mathbf{X}_{r1} + \mathbf{X}_{r0}'\mathbf{X}_{r0} \end{pmatrix}^{-1} \tag{52}$$

[5] An extension to the model (48) might be to relax the assumption that $\sigma^2 = \sigma_1^2 = \sigma_0^2$ somewhat by letting $\sigma_i^2 = \sigma_0^2 exp\{c \times d_{1,i}\}$ for all i in the sample. The multiplicative constant c can be estimated by $\hat{c} = (\mathbf{d}_1'\mathbf{d}_1)^{-1}\mathbf{d}_1\mathbf{v}$, where $\mathbf{v}$ is the vector with the natural logarithms of the squared residuals from the least squares regression of (47).

Again applying the rules of matrix inversion to (52) and substituting the result into (51) gives the desired expression of σ_{11}^2:

$$\sigma_{11}^2 = \sigma^2 \frac{1}{N_1} + \sigma^2 \frac{1}{N_0} + (\bar{\mathbf{x}}_{r1} - \bar{\mathbf{x}}_{r0})'(\mathbf{X}_{r1}'\mathbf{M}_1\mathbf{X}_{r1} + \mathbf{X}_{r0}'\mathbf{M}_0\mathbf{X}_{r0})^{-1}(\bar{\mathbf{x}}_{r1} - \bar{\mathbf{x}}_{r0}) \quad (53)$$

with $\mathbf{M}_0 = \mathbf{I}_0 - \boldsymbol{\iota}_0\boldsymbol{\iota}_0'/N_0$ an idempotent $(N_0 \times N_0)$-matrix. Accordingly, the bias factor $B^P = \exp\{\frac{1}{2}\sigma_{11}^2\}$ in (49) associated with the pooled approach is always larger than 1 (σ_{11}^2 is always positive) and depends on two factors: the amount of within-period price variation and the amount of technological progress. The impact of the unexplained price variation σ^2 diminishes with the sample size: in larger samples the bias resulting from imperfectly explaining hedonic model (47) is less than that in smaller samples (all else equal). The impact of technological progress depends on the situation at hand. If there is no progress at all, such that the average product characteristics in the two periods are close ($\bar{\mathbf{x}}_{r1} = \bar{\mathbf{x}}_{r0}$), then the bias factor B^P is relatively close to 1 (σ_{11}^2 close to 0) and the bias is small. However, if technological changes in the industry are relatively large, then the averages of the product characteristics will differ from period to period and the bias will be comparatively large ($B^P > 1$). Thus, there exists intuitive correspondence between the sources of bias variations between the pooled and other hedonic estimators of quality-adjusted price changes, despite the substantial technical differences between the various approaches.

Table 5: Summary of pooled hedonic index and bias factors

Year	N_t	$\hat{\sigma}_t^2$	R^2	$I\hat{P}_{LS}^P$	$\hat{P}_{LS}^P$	$\hat{\sigma}_{22}$	$I\hat{B}_{LS}^P$
1990	.	.	.	1.000	.	.	1.000
1991	4086	1.939	0.932	1.023	1.023	0.004	1.000
1992	4670	1.714	0.935	1.052	1.028	0.004	1.000
1993	5151	1.221	0.940	1.061	1.008	0.003	1.000
1994	5846	0.926	0.943	1.057	0.996	0.003	1.000
1995	6377	0.824	0.945	1.037	0.981	0.003	1.000
1996	6978	0.744	0.944	1.009	0.973	0.003	1.000
1997	7581	0.611	0.947	0.998	0.989	0.002	1.000
1998	7582	0.566	0.951	0.989	0.991	0.002	1.000
1999	6634	0.684	0.952	0.982	0.993	0.002	1.000

A practical example of the bias associated with the pooled hedonic index is given in table 5, which is again based on estimation results presented in Van Dalen and Bode (2004). The first three columns summarize the performance of the bi-annually pooled regression models, while $I\hat{P}_{LS}^P$ represents the pooled hedonic index determined by chaining the successively estimated $\hat{P}^P_{LS} = exp\{\hat{\pi}\}$. The estimated index closely resembles the results for the least-squares based Laspeyres and Paasche-like hedonic indices, $I\hat{P}_{LS,L}$ and $I\hat{P}_{LS,P}$ discussed before: quality-adjusted prices increase until about 1993 and decrease afterward; they end at a quality-adjusted price level slightly below that in 1990. More interesting, is the estimated bias factor $I\hat{B}_{LS,P}$ obtained by chaining the exponent of the squared standard error of the time dummy $\exp\{\hat{\sigma}^2_{22}\}$. The influence of the bias factor appears to be negligible differing from 1 only in the fifth decimal. This may seem a highly attractive result, but it should be born in mind that it is obtained only after imposing the relatively restrictive conditions explained before. Moreover, there is no guarantee that the result carries over to other hedonic analyses as well.

6 Summary, results and concluding remarks

This study has reviewed the consequences of the estimation bias associated with the estimation of log-linear hedonic models for the construction of quality-adjusted hedonic price indices. Basically, the estimation bias is due to a mismatch between a legitimate desire to unbiasedly predict the (non-transformed) prices of product varieties and the commonly applied zero-mean assumption in the hedonic regression model explaining log-linearly transformed prices in terms of the quality characteristics of product varieties. Least-squares based price predictions are biased by a multiplicative factor $\exp\{-\frac{1}{2}\sigma^2\}$ associated with the expectation of the disturbance term of the hedonic equation and a factor $\exp\{\frac{1}{2}\sigma^2 h_r\}$ associated with the presence of a variance term in the expected value of the predicted (non-transformed) prices. Maximum-likelihood based price predictions are only affected by the latter bias factor, which is usually quite small.

The impact of the estimation bias has been analyzed for various hedonic price indices, which vary with respect to: (*i*) the kind of imputation employed (complete imputation or imputation of missing prices of entering and exiting product varieties only); (*ii*) the definition of price indices as an expected value of price relatives or as a relative of expected prices in successive periods; (*iii*) the construction of price indices in either a Laspeyres- or Paasche-like fashion or in a superlative Törnqvist-like manner; (*iv*) and the use of annual or pooled estimation results. Also, attention has been paid to the role of weighting, both weighting as part of the estimation procedures employed and weighting as part of the price index construction. The practical relevance of the theoretical implications have been illustrated with an example of car price developments of new passenger cars sold in the Dutch market during the period 1990-1999; see Van Dalen and Bode (2004).

Results
Various results have been obtained, some of which are highlighted below:

- Hedonically estimated quality-adjusted price indices. Hedonic price indices have been defined as: (*i*) the expected value of the ratio of the price levels of some reference product (P_r^A); and (*ii*) the ratio of the systematic parts of the price-characteristics relationships of a typical variety (P_r^B). Both the least-squares or maximum likelihood estimation lead to biased estimators of both P_r^A and P_r^B. The derived bias factors show that the maximum-likelihood based hedonic index usually is an adequate measure, especially in larger samples, while the adequacy of the least-squares based hedonic index depends on the size and temporal stability of the variances (σ_0^2 and σ_1^2). Interestingly, a high explanatory power of the hedonic regressions is not a sufficient condition to prevent biases to occur, and not even a necessary condition to cause biases not to occur.

  The empirical example demonstrates that the (weighted) least-squares (Laspeyres and Paasche) price indices $\hat{P}_{WS}$ are practically unbiased estimators of P^A, but are extremely upward biased estimators of P^B. In the latter instance the (Laspeyres) bias factor $\hat{B}^B_{WS,L}$ is 1.893 in 1999 indicating that the estimated price change is about twice as much as should be. The outcomes for the maximum-likelihood based (Laspeyres and Paasche) estimators $\hat{P}_{WL}$ are almost reversed: these indices appear to be almost unbiased when estimating P^B, but are substantially biased when estimating P^A. The main source of the biases appeared to be the factor $\exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\}$, whose estimates vary between 0.747 in 1993 and 1.080 in 1999. The biases were therefore largely due to the dynamic performance of the hedonic regressions: the mean squared errors $\hat{\sigma}_t^2$ revealed considerable variation in time causing the least-squares and maximum likelihood based price indices to diverge. In all, the results demonstrated the importance of explicating the object of estimation (P_r^A or P_r^B) as well as of properly estimating the hedonic relationships.

- Composite and matched-model quality adjusted price indices. In this case, sales-weighted, geometrically-averaged Törnqvist-like indices have been used, in which the price relatives are either partially imputed when missing (leading to the composite price index P^C) or wholly represented by their predicted values based on hedonic models (leading to a matched-model hedonic price index P^M). The extent of the biases associated with the estimated composite and matched-model prices indices depends on the specific price index selected as well as on the entity that is to be estimated P^C or P^M.

The various bias factors appeared to depend on five typical elements that were discussed in section 4.4. One of these bias elements ($\exp\{\frac{1}{2}(\sigma_1^2 - \sigma_0^2)\}$) is also part of the biases associated with the estimation of P^A and P^B. As before, the extent and direction of this particular bias element depends on the performance of the hedonic models in time.

The size of the various biases was again illustrated using the car price developments example. Least-squares based matched-model and chained composite price indices were obtained using sales-volume weights in the regression analyses and sales-value weights in the calculation of the indices. The matched-model index was below the composite index in all years and ends at 1.03 in 1999, about 5 percentage points below the latter. The maximum-likelihood based indices again strongly deviated from the least-squares based estimates. Even in the case of the maximum-likelihood based composite price index $I\hat{P}^C_{ML}$, the estimated price index values were below the least-squares based results $I\hat{P}^C_{LS}$ and the quality-adjusted price level in 1999 was substantially (about 21%) below the 1990 price level. The pattern of the bias factors is roughly the same as observed for the hedonic indices.

- Pooled hedonic price indices. Pooled estimation of the hedonic relationship utilizing information about prices and product characteristics from two (or more) adjacent periods yield quality-adjusted price index estimates as the (exponent of the) estimate of a time-dummy variable. This approach implicitly assumes that the effects of quality on prices are the same in both periods, and that the amount of unexplained price variation among product varieties within each period is the same. The restrictions imply that two important sources of biases in the hedonic price indices, viz. the parts related with the differences between $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_0$ and the difference between σ_1^2 and σ_0^2 in (15), are set aside by assumption. In light of these restrictive assumptions, it is obvious to expect that the amount of bias observed for the pooled hedonic price index is smaller than the biases observed for the various hedonic indices. This was confirmed by the car price development example.

- Weighting. Weighting plays a role both in estimating the parameters of the hedonic model and in constructing the price indices. The impact of weighting as part of the regression procedure depends on the objective of the weighting scheme. In our study the purpose has been to take varying sales volumes of product varieties into account. This has obvious consequences for the variance of the disturbance term, but has no effects its expected value. This would definitely not be the case when weighting is applied to cope with varying price heterogeneity (heteroscedasticity) between product varieties.

In the case of the composite and matched-model price indices, the weights used in the calculation of the price indices have been considered to be non-random and to have no impact on the interpretation of the prices included. The first assumption reduces the complexity of the analysis (sales-weights involve a price component that is itself subject of analysis) and the second assumption is required to avoid inconsistencies between the interpretation of the prices implied by weighted regression (section 3.4) and that of the prices used in the price index.

Limitations

Our analysis has made various assumptions that proved to be convenient, but which have not been evaluated and consequently are subject of future exploration. The following assumptions are mentioned explicitly:

- Throughout it is has been assumed that log-transformed prices are normally distributed. The assumption is necessary to arrive at analytical expressions for the size and scope of the biases. Changing the assumed distribution of the disturbances will also affect these bias expressions. However, the main source of the estimation bias is the inconsistency between the conditions that both log-prices and non-transformed prices should be unbiased, which is a somewhat different issue than that covered by the normality assumption.

- Random deviations from the hedonic models in two successive periods have been assumed to be independently distributed. This assumption is widely imposed in applied work, but it may not be considered plausible. We did not test the independency, and have not presented empirical evidence of the consequences of allowing for non-zero correlations. At the same time, our framework is general enough to allow such testing if considered appropriate.

- Expected prices of entering or exiting product varieties are assumed to satisfy the price-characteristics structure implied by the hedonic models. This is of course a quite common assumption, but alternative views such as a reservation price approach would definitely give different results.

Concluding remarks

The role of the bias associated with the use of log-linear hedonic models may be large or small depending on the situation at hand, but the general impression gained from the analysis is that it can not be considered negligible a priori. It is recommended, therefore, that applications of the log-linear model for the construction of quality adjusted price indices are accompanied with estimates of the bias.

Moreover, the fact that different estimation methods affect different types of price indices in different ways, stresses the importance of transparent communication about which price index estimate has been obtained with which method to estimate the hedonic model, and ideally also of presenting the results of some sensitivity analysis involving different combinations of estimation method and type of price index selected.

References

Aigner, D.J., 1974. Asymptotic Minimum-MSE Prediction in the Cobb-Douglas Model with a Multiplicative Disturbance Term, *Econometrica* **42**(4), 737-748.

Aizcorbe, A.M., C. Corrado, and M. Doms, 2000. Constructing Price and Quantity Indexes for High Technology Goods, Paper presented at NBER Summer Institute about Price, Output, and Productivity Measurement, NBER: Cambridge, MA.

Berndt, E.R., 1991. *The Practice of Econometrics: Classic and Contempory*, Reading, MA: Addison-Wesley.

Berndt, E.R., Z. Griliches en N. Rappaport, 1995. Econometric Estimates of Price Indexes for Personal Computers in the 1990's, *Journal of Econometrics* **68**(1), 243-268.

Cole, R., Y.C. Chen, J.A. Barquin-Stolleman, E. Dulberger, N. Helvacian and J.H. Hodge, 1986. Quality-Adjusted Price Indexes for Computer Processors and Selected Peripheral Equipment, *Survey of Current Business* **66**, 41-50.

Diewert, W.E., 2002. Hedonic Regressions: A Consumer Theory Approach. In: Matthew Shapiro and Robbert Feenstra (eds), *Scanner Data and Price Indexes*, NBER, Studies in Income and wealth **61**, Chicago: University of Chicago Press.

Evans, I. G., and S. A. Shaban, 1976. Point Estimation in Multiplicative Models, *Econometrica* **44**(3), 467-473.

Goldberger, A.S., 1968. The Interpretation and Estimation of Cobb-Douglas Functions, *Econometrica* **36**(3/4), 464-472.

Griliches, Z. (ed.), 1971, *Price Indexes and Quality Change*, Cambridge, Massachusetts: Harvard University Press.

Hall, R.E., 1971, The Measurement of Quality Change from Vintage Price Data, in: Z. Griliches (ed.), *Price Indexes and Quality Change*, Cambridge, Massachusetts: Harvard University Press, 240-271.

Kennedy, P.E., 1981. Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations, *The American Economic Review* **71**(4), 801.

Lancaster, K.J., 1966a. A new Approach to Consumer Theory, *Journal of Political Economy* **74**, 132-157.

Lancaster, K.J., 1966b, Change and Innovation in the Technology of Consumption, *American Economic Review* **56**, 14-23.

Rukhin, A.L., 1986. Improved Estimation in Lognormal Models, *Journal of the American Statistical Association* **81**(396), 1046-1049.

Silver, M., and S. Heravi, 2001. The Measurement of Quality Adjusted Price Changes, Paper presented at NBER Conference on Research in Income and Wealth, Washingtion DC: september 2000.

Silver, M., 2002. Quality change and hedonics In: Revision of the ILO manual on CPI, Chapter 21. http://www.ilo.org/public/english/bureau/-stat/guides/cpi/index.htm.

Teekens, R., and J. Koerts, 1972. Some Statistical Implications of the Log Transformation of Multiplicative Models, *Econometrica* **40**(5), 793-819.

Triplett, J.E., 2002. *Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products*, OECD, Directorate for Science, Technology and Industry, Draft, OECD, Paris.

Van Dalen, J., and B. Bode, 2004. Quality-Corrected Price Indices: the case of the Dutch new passenger car market, 1990-1999, *Applied Economics* **36**(11), 1169-1197.

Zellner, A., 1971. Bayesian and Non-Bayesian Analysis of the Log-Normal Distribution and Log-Normal Regression, *Journal of the American Statistical Association* **66**(334), 327-330.