Hedonic functions, hedonic methods, estimation methods and Dutot and Jevons house price indexes: are there any links?*

Esmeralda A. Ramalho and Joaquim J.S. Ramalho

Department of Economics and *CEFAGE-UE, Universidade de Évora*

September 2011

Abstract

Hedonic methods are a prominent approach in the construction of house price indexes. This paper investigates in a comprehensive way whether or not there exists any kind of link between the type of price index to be computed (Dutot or Jevons) and the form of hedonic functions, hedonic methods and estimation methods, with a link being defined as a specific combination of price indexes, functions and methods that simplifies substantially the calculations required to compute hedonic price indexes. It is found that: (i) there is a link between Dutot indexes, exponential hedonic functions and the Poisson pseudo maximum likelihood estimator, on the one hand, and Jevons indexes, log-linear hedonic functions and ordinary least squares, on the other hand; and (ii) unlike implicitly assumed in the hedonic literature, there is no link between Jevons indexes and the time dummy variable method, since in this context quality-adjusted Dutot price indexes may also be simply computed as the exponential transformation of a time dummy variable coefficient, provided that an exponential hedonic function is used. A Monte Carlo simulation study illustrates both the convenience of the links identified and the biases that result from overlooking them or implementing bias corrections based on invalid assumptions.

Keywords: house prices, hedonic price indexes, quality adjustment, exponential regression model, log-linear regression model, retransformation.

JEL Classification: C43, C51, E31, R31.

*Financial support from Fundação para a Ciência e a Tecnologia, program FEDER/POCI 2010, is gratefully acknowledged. We are also indebted for helpful comments to Carlos Brás, Daniel Santos, Erwin Diewert, João Santos Silva, Rui Evangelista, Vanda Guerreiro and participants at the 26th Annual Congress of the European Economic Association, Oslo, and the 58th World Statistics Congress of the International Statistical Institute (ISI), Dublin. Address for correspondence: Joaquim J.S. Ramalho, Department of Economics, Universidade de Évora, Largo dos Colegiais, 7000-803 ÉVORA, Portugal (e-mail: jsr@uevora.pt).

1 Introduction

The construction of housing price indexes raises many conceptual and practical problems, because each house is typically a unique combination of many characteristics and in each year a very small percentage of the housing stock changes hand, implying that house prices are rarely observed. Therefore, house price indexes cannot be constructed simply by comparing the average price of houses sold in each time period, since the result would be dependent on the particular mix of dwellings that happened to be sold in that period. Instead, the heterogeneity of dwellings has somehow to be taken into account in order to separate the influences of quality changes from pure price movements. One way to do this is to use hedonic pricing methodologies, which over the past four decades have become the most relevant technique for dealing with housing heterogeneity. In fact, the first application of hedonic methods to construct house price indexes seems to have been made by the US Census Bureau and to date back to 1968 (Triplett, 2006); in the UK, two hedonic house price indexes (the Halifax and the Nationwide house price indexes) are produced since 1983;[1] and in France, quarterly hedonic housing price indexes have been computed since 1998 (Gourieroux and Laferrere, 2009). See Hill (2011) for other examples of countries where hedonic house price indexes dominate.

In the housing framework, hedonic pricing techniques build upon the idea that different characteristics of a dwelling impact differently on its evaluation by consumers. To measure those impacts, it is necessary to specify the so-called hedonic price function, which relates transaction prices to the relevant dwelling characteristics. Using regression techniques, it is then possible to estimate the implicit marginal prices of each dwelling characteristic. Finally, based on the estimated marginal prices, and using an appropriate (hedonic) method, housing prices can be straightforwardly adjusted in order to remove the effect of quality changes. Along this process, among other aspects, four important choices have to be made by empirical researchers: (i) the type of price index to compute (e.g. geometric, arithmetic); (ii) the form of the dependent variable in the hedonic price function (e.g. prices, logged prices); (iii) the hedonic method used for calculating the quality-adjusted price index, which reflects the assumptions made on the evolution of the marginal prices of the dwelling characteristics (e.g. imputation price method - prices are allowed to change every period; time dummy variable method - prices are assumed to be constant over time); and (iv) the method used to estimate the parameters of the hedonic function (e.g. ordinary or weighted least squares).

The choice of the form under which the price should be included in the hedonic function, and

[1] For information on the mentioned indexes see, respectively, http://www.lloydsbankinggroup.com/media1/economic_insight/halifax_house_price_index_page.asp and http://www.nationwide.co.uk/hpi.

its relationship with the choice of the price index, is one of the key issues in the general literature on constructing hedonic quality-adjusted price indexes, being the first in a list of unresolved issues discussed by Diewert (2003). In the context of the imputation price method, there are presently two very distinct approaches on this subject. Most authors (e.g. Triplett, 2006, p. 64) argue that the choice of an index number formula has to be an entirely separate matter from the choice of the form of the hedonic function. Otherwise, would the former require a specific form for the latter, researchers could be forced to use a functional form that is inconsistent with the data and might create an error in the quality adjustment procedure. According to this view, as the form of the hedonic function should depend only on the empirical relation between the prices of dwellings and their characteristics, its choice should be based exclusively on the use of statistical tools (e.g. goodness-of-fit criteria, specification tests). In contrast, other authors (e.g. Reis and Santos Silva, 2006) claim that the form under which the dependent variable appears in the hedonic function should correspond to the aggregator for the index. Therefore, Dutot (arithmetic) price indexes should be computed using estimates from hedonic functions with untransformed dwelling prices, while Jevons (geometric) price indexes should be based on hedonic functions using logged prices as the dependent variable. This second approach does not exclude the use of statistical tests to find the hedonic function that best fits the data but, in cases where the type of price index is defined *a priori*, restricts their application to the evaluation of the specification adopted for the right-hand side of the hedonic function.

While there is a clear divergence on the existence, or not, of links between price indexes and the form of the hedonic function in the context of the imputation price method, in the case of the time dummy variable method there is an apparent consensus in the hedonic literature that, in fact, there is a link between the Jevons price index and the log-linear hedonic function.[2] This is because the main attractiveness of the time dummy variable method, which requires heavier assumptions than the imputation price method, is the possibility of obtaining very simple expressions for quality-adjusted price indexes. As all authors seem to think that such simple expressions can only be obtained using the specific combination of price index and hedonic function referred to, the time dummy variable method has been considered in the hedonic literature, to the best of our knowledge, only in association with the Jevons price index and log-linear hedonic functions and never to compute Dutot quality-adjusted price indexes or in conjugation with other hedonic functions.

[2] In this paper, for simplicity, we use broadly the term 'log-linear' to denote any regression model that considers logged prices as the dependent variable (e.g. log-log, semi-log and translog models, index models with quadratic and/or interaction terms, etc.), since all the econometric analysis undertaken in the paper applies irrespective of the exact form under which the explanatory variables appear in the hedonic function.

Irrespective of the hedonic method employed, Reis and Santos Silva (2006) claim the existence of another link, this time involving the method used for estimating the hedonic function. They show that, in the context of weighted indexes (they were interested in price indexes for new passenger cars based on samples requiring the use of weights), any hedonic model linear in the parameters must be estimated by weighted least squares using as weights the same market shares employed to compute the indexes. Reis and Santos Silva (2006) proposed also a similar link for the case of a nonlinear regression function. In both cases, the extension to the case of non-weighted indexes is immediate.

The main aim of this paper is to investigate in a comprehensive way whether or not there exists any kind of link between the type of price index to be computed (Dutot or Jevons)[3] and the form of hedonic functions, hedonic methods and estimation methods. We consider that there is a link whenever a specific combination of price indexes, functions and methods simplifies substantially the calculations required to compute hedonic price indexes, while other combinations, although possible, require either additional assumptions and, in general, the use of bias corrections, or the estimation of hedonic equations for all time periods in the case of the imputation price method. We analyze two particular types of hedonic functions, one using logged prices as the dependent variable and the other the prices themselves. For the latter case, we adopt an exponential specification, which, to the best of our knowledge, has never been used in this framework but proves to be much more useful to deal with quality-adjusted price indexes than the more traditional linear regression model.

In contrast to previous papers, we use Monte Carlo methods to compare estimators of housing price indexes based on choices that do and do not respect the detected links. For the latter type of estimators, whenever they require additional assumptions and price index formula corrections, we also evaluate the biases that result from either the invalidity of those assumptions or the non-application of the corrections required. In order to obtain a realistic scenario for our experiments, we use the dataset of Anglin and Gençay (1996) as basis and simulate several patterns of evolution for dwelling prices and characteristics. Using controlled experiences instead of real data allow us to evaluate in a more precise way the consequences of employing different types of hedonic functions, estimation methods and hedonic methods.

This paper is organized as follows. Section 2 introduces some notation and reviews briefly the construction of hedonic quality-adjusted price indexes. Section 3 investigates whether there exists or not any link between price index formulas and the specification of hedonic functions. Section 4 examines the previous issue in the context of the time dummy variable method.

[3]For a comprehensive text on index number theory, see Balk (2008).

Section 5 analyzes the possible existence of an additional link involving the method chosen for estimating the hedonic function. Section 6 is dedicated to the Monte Carlo simulation study. Finally, Section 7 concludes.

2 The construction of hedonic house price indexes: a brief overview

Throughout this paper, p_{it} denotes the price p of dwelling i at period t, where, typically, the subscript i indexes different dwellings in each time period. We assume that either $t = 0$ (base period) or $t = s$ (current period). Let N_t be the number of dwellings observed at each time period. Let $X_{it,j}$ be the characteristic j of dwelling i at period t, $j = 1, ..., k$, and let x_{it} be the $1 \times (k+1)$ vector with elements $X_{it,j}$, $j = 0, ..., k$, where variable $X_{it,0} = 1$ denotes the constant term of the hedonic regression. Next, we provide a brief overview of the construction of hedonic quality-adjusted price indexes.

2.1 Dutot and Jevons price indexes

The main alternative elementary formulas for computing dwelling price indexes in the hedonic framework are the ratio of (unweighted) arithmetic means of prices (the Dutot price index) and the ratio of (unweighted) geometric means of prices (the Jevons price index). Let I^D and I^J be, respectively, the *population* Dutot and Jevons price indexes and let $\bar{I}^D$ and $\bar{I}^J$ be the corresponding *sample* estimators. At moment s, the sample Dutot price index is given by

$$\bar{I}_s^D = \frac{\frac{1}{N_s}\sum_{i=1}^{N_s} p_{is}}{\frac{1}{N_0}\sum_{i=1}^{N_0} p_{i0}}, \tag{1}$$

while the sample Jevons price index may be written as

$$\bar{I}_s^J = \frac{\prod_{i=1}^{N_s} p_{is}^{\frac{1}{N_s}}}{\prod_{i=1}^{N_0} p_{i0}^{\frac{1}{N_0}}} = \frac{\exp\left[\frac{1}{N_s}\sum_{i=1}^{N_s} \ln\left(p_{is}\right)\right]}{\exp\left[\frac{1}{N_0}\sum_{i=1}^{N_0} \ln\left(p_{i0}\right)\right]}. \tag{2}$$

It is straightforward to show that $\bar{I}_s^D$ is a consistent estimator of the population Dutot index

$$I_s^D = \frac{E\left(p_s\right)}{E\left(p_0\right)}, \tag{3}$$

while $\bar{I}_s^J$ is a consistent estimator of the population Jevons index

$$I_s^J = \frac{\exp\left\{E\left[\ln\left(p_s\right)\right]\right\}}{\exp\left\{E\left[\ln\left(p_0\right)\right]\right\}}. \tag{4}$$

As the exponential function cannot be taken through expected values, in general $I_s^D \neq I_s^J$; see Silver and Heravi (2007b) for a detailed study on the relationship between the population Dutot and Jevons price indexes.

The Dutot and Jevons price indexes just described measure the overall dwelling price change between period 0 and period s. That change may be due to the different characteristics of the dwellings sold in each period or may be the result of a pure price movement. Assuming that each characteristic of each dwelling may be evaluated and that dwellings may be interpreted as aggregations of characteristics, both Dutot and Jevons indexes may be decomposed into two components: a quality index ($I_s^{D_q}$ or $I_s^{J_q}$), which assumes that the implicit prices of the dwelling characteristics did not change over time and, therefore, measures the price change that is explained by changes in the dwelling characteristics; and a quality-adjusted price index ($I_s^{D_p}$ or $I_s^{J_p}$), which assumes that the characteristics of the dwellings are constant across time and measures the price change that is due to changes in the prices of the dwelling characteristics. Thus, we may write the population Dutot price index as

$$I_s^D = I_s^{D_q} \cdot I_s^{D_p} \tag{5}$$

and the population Jevons price index as

$$I_s^J = I_s^{J_q} \cdot I_s^{J_p}, \tag{6}$$

where

$$I_s^{D_q} = \frac{E\left(p_b \middle| x_{is}\right)}{E\left(p_b \middle| x_{i0}\right)}, \; I_s^{D_p} = \frac{E\left(p_s \middle| x_{ia}\right)}{E\left(p_0 \middle| x_{ia}\right)}, \tag{7}$$

$$I_s^{J_q} = \frac{\exp\left\{E\left[\ln\left(p_b \middle| x_{is}\right)\right]\right\}}{\exp\left\{E\left[\ln\left(p_b \middle| x_{i0}\right)\right]\right\}}, \; I_s^{J_p} = \frac{\exp\left\{E\left[\ln\left(p_s \middle| x_{ia}\right)\right]\right\}}{\exp\left\{E\left[\ln\left(p_0 \middle| x_{ia}\right)\right]\right\}} \tag{8}$$

and $(a,b) = (0,s)$ or $(s,0)$. Note that when $(a,b) = (0,s)$, $I_s^{D_p}$ and $I_s^{J_p}$ are Laspeyres-type quality-adjusted price indexes, since the comparison is based on the dwellings existing at the base period; and when $(a,b) = (s,0)$, $I_s^{D_p}$ and $I_s^{J_p}$ are Paasche-type quality-adjusted price indexes, since the comparison is based on the dwellings existing at the current period.

The prices of the dwelling characteristics are not observable, so the sample estimators $\bar{I}_s^D$ and $\bar{I}_s^J$ cannot be directly decomposed into quality and quality-adjusted price indexes. However, if a sample of the dwelling characteristics is available for each period, it is possible to estimate their implicit prices, and their evolution, using the so-called hedonic regression, which relates (dwelling) prices to (dwelling) characteristics. Based on this regression, alternative estimators for the unadjusted Dutot and Jevons price indexes may be constructed, being given by, respectively,

$$\hat{I}_s^D = \frac{\frac{1}{N_s}\sum_{i=1}^{N_s}\hat{p}_{is}}{\frac{1}{N_0}\sum_{i=1}^{N_0}\hat{p}_{i0}} \tag{9}$$

and

$$\hat{I}_s^J = \frac{\exp\left[\frac{1}{N_s}\sum_{i=1}^{N_s}\widehat{\ln(p_{is})}\right]}{\exp\left[\frac{1}{N_0}\sum_{i=1}^{N_0}\widehat{\ln(p_{i0})}\right]}, \tag{10}$$

which are consistent estimators of I_s^D and I_s^J, respectively, provided that the predictors $\hat{p}_{it}$ and $\widehat{\ln(p_{it})}$ are consistent estimators for $E(p_{it})$ and $E[\ln(p_{it})]$, respectively. As shown later in the paper, the *hedonic* estimators $\hat{I}_s^D$ and $\hat{I}_s^J$ may be straightforwardly decomposed into quality and quality-adjusted price indexes, which, under suitable assumptions, are consistent estimators of the corresponding population indexes defined in (7) and (8).

2.2 Specification of hedonic functions

On the basis of economic theory, very few restrictions are placed on the form of the hedonic price equation; see e.g. Cropper, Deck and McConnell (1988). As practically there is no *a priori* structural restriction on its form, several alternative specifications have been adopted for the hedonic function in empirical studies. Most of those specifications differ essentially in the form under which the explanatory variables appear in the hedonic equation, with the dependent variable appearing either in levels or in logarithms. In this paper we focus on the latter choice because, for the purposes of this paper, the exact specification of the explanatory variables is irrelevant in the sense that any function of the dwelling characteristics (e.g. logs, squares, interaction terms) is easily accommodated by the procedures proposed in the next sections to compute Jevons and Dutot hedonic price indexes. Therefore, although, for simplicity, all hedonic functions considered in this paper are based on index models linear in the parameters, all results remain valid if more complex, nonlinear index models are used.

Given that prices are strictly positive, the most plausible specifications for hedonic functions are probably the log-linear model

$$\ln p_{it} = x_{it}\beta_t + u_{it} \tag{11}$$

and the exponential regression model

$$p_{it} = \exp\left(x_{it}\beta_t^* + u_{it}^*\right), \tag{12}$$

where u_{it} (u_{it}^*) is the error term, standing for the non-explained part of the price, e.g. unregistered attributes of the dwelling, and β_t (β_t^*) is the $(k+1) \times 1$ vector of parameters, with elements $\beta_{t,j}$ ($\beta_{t,j}^*$), $j = 0, ..., k$, to be estimated. The parameter $\beta_{t,j}$ ($\beta_{t,j}^*$) is often interpreted as the implicit marginal price for (some function of) characteristic $X_{t,j}$ and is allowed to change over time.

In a nonstochastic form (i.e. without an error term), models (11) and (12) would represent exactly the same relationship between p_{it} and x_{it}. In that case, $\beta_t = \beta_t^*$ and the same theoretical arguments used for justifying specification (11) can also be applied to justify model (12). However, due to the presence of the stochastic error terms u_{it} and u_{it}^*, the two models are not equivalent, since the former requires the assumption $E\left(u_{it} | x_{it}\right) = 0$, while the latter assumes $E\left[\exp\left(u_{it}^*\right) | x_{it}\right] = 1$. As it is well known, neither of those assumptions imply the other, i.e. $E\left[\exp\left(u_{it}\right) | x_{it}\right] \neq 1$ and $E\left(u_{it}^* | x_{it}\right) \neq 0$. In fact, as demonstrated by Santos Silva and Tenreyo (2006), only under very specific conditions on the error term would the two models describe simultaneously the same data generating process.

In empirical work, due to the fact of being linear in the parameters and hence easily estimable, the log-linear model (11) has been widely applied in the construction of hedonic price indexes. In contrast, the exponential regression model (12), to the best of our knowledge, has not been ever considered in the applied hedonic literature. Nevertheless, in this paper we focus on both specifications, because, as it will become clear soon, a crucial issue in the construction of Jevons and Dutot quality-adjusted price indexes is whether the dependent variable of the hedonic function should be the price itself or its logarithm.[4] Moreover, as shown in Section 4, the computation of Dutot price indexes in the time dummy variable method framework may be substantially simplified if an exponential hedonic function is used.

3 Links between price indexes and hedonic functions

As briefly discussed in the previous section, to construct hedonic price indexes it is necessary first to use the hedonic regression to obtain consistent estimators of unadjusted price indexes and then to decompose the unadjusted index into quality and quality-adjusted price components. This section considers the exponential and log-linear hedonic functions described above and examines how unadjusted and quality-adjusted Dutot and Jevons price indexes may be consistently estimated in each case. First, we consider the case of Dutot price indexes and then the Jevons case.

[4] In this sense, instead of the exponential regression model, we could have considered the much more popular linear hedonic function $p_{it} = x_{it}\beta_t^* + u_{it}^*$. However, as the linear model does not take into account the positiveness of p_{it}, it may generate negative price estimates. This problem was noted *inter alia* by Hill and Melser (2008), which had to drop dwellings with negative price predictions before computing price indexes.

3.1 Links in the Dutot framework

The analysis that follows is made first under the assumption that the true and specified hedonic functions coincide and then considering the opposite case. Given the procedures and assumptions underlying the use of each type of hedonic function, we conclude whether quality-adjusted Dutot price indexes may be indifferently estimated using exponential or log-linear hedonic functions or, instead, it is clearly preferable to use only one of those specifications.

3.1.1 True and assumed data generating process: exponential hedonic function

Assume that the true generating process of dwelling prices is appropriately described by the exponential hedonic function (12), with $E\left[\exp\left(u_{it}^{*}\right)\middle| x_{it}\right] = 1$. Assume also that the researcher specifies and estimates that same hedonic function. In this framework, a consistent predictor of dwelling prices is simply given by $\hat{p}_{it} = \exp\left(x_{it}\hat{\beta}_t^{*}\right)$. Therefore, it follows immediately that a consistent estimator of I_s^D of (3) is given by the hedonic estimator $\hat{I}_s^D$ of (9), with $\hat{p}_{it}$ replaced by $\exp\left(x_{it}\hat{\beta}_t^{*}\right)$:

$$\hat{I}_s^D = \frac{\frac{1}{N_s}\sum_{i=1}^{N_s}\exp\left(x_{is}\hat{\beta}_s^{*}\right)}{\frac{1}{N_0}\sum_{i=1}^{N_0}\exp\left(x_{i0}\hat{\beta}_0^{*}\right)}. \tag{13}$$

Moreover, $\hat{I}_s^D$ can be straightforwardly decomposed into a quality index and a quality-adjusted price index:

$$\hat{I}_s^D = \underbrace{\frac{\frac{1}{N_s}\sum_{i=1}^{N_s}\exp\left(x_{is}\hat{\beta}_b^{*}\right)}{\frac{1}{N_0}\sum_{i=1}^{N_0}\exp\left(x_{i0}\hat{\beta}_b^{*}\right)}}_{\hat{I}_s^{D_q}} \underbrace{\frac{\frac{1}{N_a}\sum_{i=1}^{N_a}\exp\left(x_{ia}\hat{\beta}_s^{*}\right)}{\frac{1}{N_a}\sum_{i=1}^{N_a}\exp\left(x_{ia}\hat{\beta}_0^{*}\right)}}_{\hat{I}_s^{D_p}}. \tag{14}$$

where $(a, b) = (0, s)$ or $(s, 0)$ and $\hat{I}_s^{D_q}$ and $\hat{I}_s^{D_p}$ are consistent estimators for, respectively, $I_s^{D_q}$ and $I_s^{D_p}$ of (7). Clearly, $\hat{I}_s^{D_p}$ may be interpreted as a quality-adjusted price index because, on the one hand, it compares the values of the characteristics of the dwellings observed in period a using the implicit prices of the characteristics estimated for periods 0 and s and, on the other hand, the only other source of price variation in the index assumes that the implicit prices of the dwelling characteristics do not change over time.

3.1.2 True and assumed data generating process: log-linear hedonic function

While the construction of quality-adjusted Dutot price indexes is very simple when both the true and specified hedonic functions have an exponential form, the same does not happen when those functions are both log-linear. The problem is that the estimation of a log-linear hedonic function yields directly consistent estimates for the logarithm of the dwelling price, $\widehat{\ln\left(p_{it}\right)} = x_{it}\hat{\beta}_t$ (see equation 11), not for the price itself, but Dutot price indexes require consistent estimates of prices, not logged prices. Moreover, due to the stochastic nature of hedonic functions, the antilog of $\widehat{\ln(p_{it})}$, $\exp\left[\widehat{\ln(p_{it})}\right] = \exp\left(x_{it}\hat{\beta}_t\right)$, is not in general a consistent estimator of $E(p_t|x_{it})$. Indeed, the log-linear hedonic function (11) implicitly assumes that $p_{it} = \exp(x_{it}\beta_t + u_{it})$, i.e.

$$E(p_{it}|x_{it}) = \exp(x_{it}\beta_t) E[\exp(u_{it})|x_{it}], \tag{15}$$

where, in general, $E[\exp(u_{it})|x_{it}] \neq 1$; see Section 2.2. Therefore, in the log-linear context, consistent estimates of dwelling prices require inevitably the previous estimation of $E[\exp(u_{it})|x_{it}]$.

Let $\mu_{it} \equiv E[\exp(u_{it})|x_{it}]$ and assume that

$$\mu_{it} = g(x_{it}^{*}\alpha_t), \tag{16}$$

where $g(\cdot)$ may be a nonlinear function, x_{it}^{*} is some function of x_{it} and α_t is the associated $(k_\alpha + 1)$-vector of parameters. For the moment, assume that $g(\cdot)$ is a known function and that a consistent estimator for μ_{it}, $\hat{\mu}_{it} = g(x_{it}^{*}\hat{\alpha}_t)$, is available. Then, a consistent estimator of I_s^D is given by

$$\hat{I}_s^D = \frac{\frac{1}{N_s}\sum_{i=1}^{N_s}\exp\left(x_{is}\hat{\beta}_s\right)g(x_{is}^{*}\hat{\alpha}_s)}{\frac{1}{N_0}\sum_{i=1}^{N_0}\exp\left(x_{i0}\hat{\beta}_0\right)g(x_{i0}^{*}\hat{\alpha}_0)}. \tag{17}$$

Therefore, in general, consistent estimation of unadjusted Dutot price indexes requires the availability of a consistent estimator for both μ_{is} and μ_{i0}. The only case where the naive estimator $\exp\left(x_{it}\hat{\beta}_t\right)$ for the price can be used for consistent estimation of I_s^D occurs when $\mu_{is} = \mu_{i0} = \mu$, i.e. μ_{it} is constant across dwellings and over time.

Although more complex due to the presence of the adjustment term, expression (17) can still be straightforwardly decomposed into quality and quality-adjusted price components. Indeed, by analogy with the decomposition for the exponential model in (14), we obtain the decomposition

$$\hat{I}_s^D = \underbrace{\frac{\frac{1}{N_s}\sum_{i=1}^{N_s}\exp\left(x_{is}\hat{\beta}_b\right)g(x_{is}^{*}\hat{\alpha}_b)}{\frac{1}{N_0}\sum_{i=1}^{N_0}\exp\left(x_{i0}\hat{\beta}_b\right)g(x_{i0}^{*}\hat{\alpha}_b)}}_{\hat{I}_s^{Dq}}\underbrace{\frac{\frac{1}{N_a}\sum_{i=1}^{N_a}\exp\left(x_{ia}\hat{\beta}_s\right)g(x_{ia}^{*}\hat{\alpha}_s)}{\frac{1}{N_a}\sum_{i=1}^{N_a}\exp\left(x_{ia}\hat{\beta}_0\right)g(x_{ia}^{*}\hat{\alpha}_0)}}_{\hat{I}_s^{Dp}}, \tag{18}$$

in which $\hat{I}_s^{Dq}$ and $\hat{I}_s^{Dp}$ contain similar corrections to that of the unadjusted index $\hat{I}_s^D$ in (17). From (18), it is clear that in the scale that is of interest for the construction of Dutot price indexes, the implicit price of each characteristic is now a function of both $\hat{\beta}_t$ and $\hat{\alpha}_t$. Therefore, both types of parameters must be kept fixed over time when calculating quality indexes and both must be evaluated at the base and current periods in the computation of quality-adjusted

Dutot price indexes. Hence, the parameter constancy of the parameters β_t that appear in the log-linear hedonic function is by no means a sufficient condition for the constancy of quality-adjusted price indexes. An important consequence of this finding is that if one is interested in testing whether prices have changed significantly between two periods, the traditional practice of applying a Chow test for assessing the null hypothesis of equal β_t coefficients in the two periods may lead to wrong conclusions: the constancy of the parameters α_t must be tested too.

Thus, the estimation of hedonic Dutot price indexes based on log-linear hedonic functions requires more calculations but, nevertheless, it is still very simple when consistent estimates of α_0 and α_s are available. However, such estimates are not easy to obtain, since they require exact knowledge on how $E[\exp(u_{it})|x_{it}]$ is related to the dwelling characteristics, i.e. they require the specification of the $g(\cdot)$ function in (16). We may either make a direct functional form assumption for $g(\cdot)$, which has been relatively uncommon in applied work, or make further assumptions on the distribution of the error term u_{it} of the log-linear hedonic function that will imply a specific form for $g(\cdot)$, which has been the most common approach in the econometrics literature on retransformation issues.[5]

There are two popular sets of assumptions that are typically made on the error term distribution. The first consists of assuming that u_{it} is homoskedastic. As shown by Duan (1983), this assumption implies that $E[\exp(u_{it})|x_{it}]$ does not depend on the individual characteristics x_{it}. Duan (1983) considered a single cross-section. Applying his assumption to our framework, we may either assume that the variance of the error term is identical over time ($\mu_{it} = \mu$) or allow it to change over time ($\mu_{it} = \mu_t$). In the former case there is no need to estimate μ, as discussed before. In the latter case, a consistent estimator of μ_t is given by Duan's (1983) smearing estimator, which consists of estimating the unknown error distribution by the empirical distribution function of the ordinary least squares (OLS) residuals of the log-linear model and then taking expectations with respect to that distribution:

$$\hat{\mu}_t = \frac{1}{N_t}\sum_{i=1}^{N_t} \exp(\hat{u}_{it}). \tag{19}$$

Alternatively to the assumptions underlying the application of the smearing estimator, it is also usual in the retransformation literature to assume that u_{it} has a normal distribution with a variance of a known form, $u_{it} \sim \mathcal{N}(0, x_{it}^{*}\alpha_t)$; see *inter alia* Meulenberg (1965) and Ai and Norton (2000). As it is well known, this implies that $\exp(u_{it})$ has a log-normal distribution,

[5] The standard problem of going from log predictions to level predictions has been commonly referred to in the econometrics literature as the 'retransformation problem'. This issue has been studied particularly in the area of health economics; see *inter alia* Mullahy (1998) and Manning and Mullahy (2001).

with mean given by:

$$\mu_{it} = \exp\left(0.5 x_{it}^{*} \alpha_t\right). \tag{20}$$

In this case, an estimate of α_t can be obtained by regressing the squared OLS residuals of the log-linear model on x_{it}^{*}. See van Dalen and Bode (2004) for a comprehensive analysis of the biases that may arise in the construction of Dutot price indexes based on log-linear hedonic functions under the assumption that u_{it} has a normal distribution. From now on, we use the term 'normal-smearing estimator' to denote the estimator computed according to (20).

Many authors in the hedonic housing literature are aware of the need for applying bias corrections when computing quality-adjusted Dutot price indexes from log-linear hedonic functions. Clearly, most authors prefer to apply the normal-smearing estimator (e.g. Malpezzi, Chun and Green, 1998; Triplett, 2006; Dorsey, Hu, Mayer and Wang, 2010; Coulson, 2011; and Hill, 2011) rather than the smearing correction (to the best of our knowledge, García and Hernández (2007) are the only authors to use this estimator). Moreover, all authors assume homoskedasticity, allowing for error variance changes only over time. However, typically, the assumptions underlying the application of the chosen bias correction are not discussed, much less are they tested. Furthermore, some authors still do not apply any bias correction in this context, either because it is considered negligenciable or because practitioners are simply not aware of it.[6] In the Monte Carlo study in Section 6, we show that very important biases may arise if a consistent estimate of $E\left[\exp\left(u_{it}\right)|x_{it}\right]$ is not used in the computation of hedonic Dutot price indexes.[7] Next, we discuss an alternative method to compute these indexes which never requires the application of bias corrections, irrespective of the form of the true hedonic function.

3.1.3 The link between Dutot price indexes and exponential hedonic functions

The previous analysis shows unequivocally that, when the hedonic function has a log-linear form, the estimation of quality-adjusted Dutot price indexes is somewhat complex, since, in general, it will be necessary to get an estimate of the retransformation bias (μ_{it}) for each time period. Moreover, the estimation of μ_{it} by simple methods requires some stringent assumptions

[6] Actually, many authors seem to confuse the bias corrections analyzed in this section, which are necessary for obtaining *consistent* predictors for dwelling prices, with those discussed later on at the end of Section 4, which aim only at reducing the *finite sample bias* of those predictors and, thus, are not important asymptotically. van Dalen and Bode (2004) consider both types of bias corrections.

[7] For comparisons of corrected (based on the normal-smearing estimator assuming homoskedasticity) and uncorrected price-adjusted Dutot price indexes using actual data, see Malpezzi, Chun and Green (1998), Pakes (2003) and van Dalen and Bode (2004), which carried out applications involving, respectively, the housing sector, personal computers and new passenger cars. Substantial differences between the two types of indexes were found in all cases.

on the distribution of the error term, which, *a priori*, there is no reason to believe that will hold with actual data. For example, when working with log-linear hedonic functions, Goodman and Thibodeu (1995), Fletcher, Gallimore and Mangan (2000) and Stevenson (2004) found dwelling age-induced heteroskedasticity, which prevents application of the simple smearing estimator. Furthermore, empirical researchers working with hedonic functions but not interested in making predictions typically do not assume normality and/or specify the heteroskedastic pattern that characterizes their data, much less are used to specify $E\left[\exp\left(u_{it}\right)|x_{it}\right]$.

Given that it is much more simple to calculate Dutot price indexes using exponential hedonic functions, it is important to examine the effects of estimating an exponential regression model in cases where the true data generating process has a log-linear representation. Consider first the augmented log-linear model that assumes also that

$$\mu_{it} = \exp\left(x_{it}\alpha_t\right), \tag{21}$$

which, among other alternatives, is indeed a plausible assumption for $E\left[\exp\left(u_{it}\right)|x_{it}\right]$. Then, from (15), it follows that

$$\begin{aligned} E\left(p_{it}|x_{it}\right) &= \exp\left(x_{it}\beta_t\right)\exp\left(x_{it}\alpha_t\right) \\ &= \exp\left[x_{it}\left(\beta_t+\alpha_t\right)\right] \\ &= \exp\left(x_{it}\beta_t^*\right), \end{aligned} \tag{22}$$

where $\beta_t^* \equiv \beta_t + \alpha_t$. Clearly, for our purposes, the addition of assumption (21) to the log-linear model is equivalent to assume from the start that the generating process of dwelling prices is appropriately described by the exponential hedonic function (12). In fact, although β_t cannot be identified when the exponential model (22) is estimated, consistent and asymptotically equivalent estimators for $E\left(p_{it}|x_{it}\right)$ are produced by both the augmented log-linear and the exponential models. This implies that, instead of the multi-step procedures described in the previous section, quality-adjusted Dutot price indexes can be consistently estimated using the standard procedures described in Section 3.1.1 for the exponential regression model, *even when the true hedonic function has a log-linear form*, provided that assumption (21) holds in the data. This is because the retransformation bias is automatically captured by the parameters β_t^*.

Assumption (21) is decisive for the previous analysis. However, note that this assumption is by no means heavier than those made above to ignore or to simplify the estimation of the μ_{it}. Let $\alpha_{t,0}$ be the intercept, let $\alpha_{t,+}$ be the remaining component of α_t and let $\alpha_{\cdot,0}$, $\alpha_{\cdot,+}$ and α denote, respectively, the previous parameters when they are assumed to be constant over time. The bias correction may be ignored only if $\mu_{it} = \mu = \exp(\alpha_{\cdot,0})$, where $\alpha_{\cdot,0} = \ln(\mu)$, which, relative to (21), imposes two additional constraints: $\alpha_t = \alpha$ and $\alpha_{\cdot,+} = 0$. The smearing estimator, by assuming $\mu_{it} = \mu_t = \exp(\alpha_{t,0})$, where $\alpha_{t,0} = \ln(\mu_t)$, is also more restrictive than the estimator resultant from the augmented log-linear model, because it requires further that $\alpha_{t,+} = 0$. If true, the restrictions imposed in either case will be automatically accommodated by a standard estimation of the exponential regression model, as shown in equation (22). Therefore, it makes more sense to estimate an exponential hedonic function than assuming *a priori* the restrictions needed to use a log-linear model without any bias corrections or plus a smearing estimator.

Relative to the normal-smearing estimator, the augmented log-linear formulation does not require normality of u_{it} but adds the assumption $x_{it}^* = x_{it}$. However, functions of x_{it} can be straightforwardly added to the index function in (21). For example, assume that the true hedonic function is log-linear and that $\mu_{it} = \exp\left(0.5x_{it}^2\alpha_t\right)$. Then:

$$\begin{aligned} E(p_{it}|x_{it}) &= \exp(x_{it}\beta_t)\exp\left(0.5x_{it}^2\alpha_t\right) \\ &= \exp(z_{it}\delta_t^*), \end{aligned} \tag{23}$$

where z_{it} is a vector containing the distinct elements of both x_{it} and x_{it}^2 and δ_t^* is the associated vector of parameters. Therefore, assumption (20) is also easy to accommodate in a standard exponential regression model. Moreover, because we only need to worry about the specification of the hedonic function, also in this case it is preferable to estimate quality-adjusted Dutot price indexes using exponential hedonic functions. This way, we may focus on the issue of choosing the (functions of) dwelling characteristics that should appear in the hedonic function and we may use standard functional form tests (e.g. the RESET test) to assess whether (22) or (23) are in fact appropriate specifications for $E(p_{it}|x_{it})$. In contrast, if we decide to deal directly with a log-linear hedonic specification, not only there is one additional function to be dealt with (the error variance function) but also it is typically less clear how to specify and test it.

Thus, the same three sets of assumptions that simplify the calculation of quality-adjusted Dutot price indexes when the hedonic function is log-linear, also ensure that the exponential regression model yields consistent estimators for dwelling prices. Given that no bias correction is needed in the latter case and that simple, standard specification tests may be used to select the explanatory variables and to assess the model functional form, it seems to be strongly recommended to use exponential hedonic functions whenever Dutot price indexes are to be computed. Of course, there may be instances where the true hedonic function is log-linear but the assumption of an exponential function for μ_{it} as in (22) or (23) is not valid at all. However, even in that case, typically it will be much easier to specify, estimate and test an exponential hedonic function augmented by (nonlinear) functions of the dwelling characteristics (e.g. polynomials and interaction terms) in order to approximate the true data generating process than to insist on the use of a log-linear model and try to find an appropriate specification for the uncommon $g(\cdot)$ function. In this sense, what is effectively relevant for a simple computation of Dutot price indexes is that the dependent variable of the hedonic function is the dwelling price itself and not some transformation of it. Indeed, this is all that is necessary to ensure that no bias correction needs to be applied to produce Dutot price indexes. Therefore, we conclude that *there exists a clear link between the computation of quality-adjusted Dutot price indexes and the exponential hedonic function, in particular, and hedonic functions which consider the unstransformed dwelling price as dependent variable, in general.*

3.2 Links in the Jevons framework

This section demonstrates that, similarly to the Dutot case, there is a link between Jevons price indexes and a particular hedonic function. We adopt the same structure as that of Section 3.1 but, given the similarity of the arguments put forward, the present section is substantially abbreviated relative to the previous one.

3.2.1 True and assumed data generating process: log-linear hedonic function

In the case of Jevons price indexes, it is the use of a log-linear hedonic function that simplifies considerably the construction of quality-adjusted indexes. Indeed, when the true data generating process is suitably described by a log-linear model, a consistent predictor of the logged price is given by $\widehat{\ln(p_{it})} = x_{it}\hat{\beta}_t$ and, therefore, a consistent estimator for I_s^J of (4) is the hedonic estimator

$$\hat{I}_s^J = \frac{\exp\left(\frac{1}{N_s}\sum_{i=1}^{N_s} x_{is}\hat{\beta}_s\right)}{\exp\left(\frac{1}{N_0}\sum_{i=1}^{N_0} x_{i0}\hat{\beta}_0\right)}. \tag{24}$$

This estimator can be decomposed into a quality ($\hat{I}_s^{J_q}$) and a quality-adjusted ($\hat{I}_s^{J_p}$) price index:

$$\hat{I}_s^J = \underbrace{\frac{\exp\left(\frac{1}{N_s}\sum_{i=1}^{N_s} x_{is}\hat{\beta}_b\right)}{\exp\left(\frac{1}{N_0}\sum_{i=1}^{N_0} x_{i0}\hat{\beta}_b\right)}}_{\hat{I}_s^{J_q}} \underbrace{\frac{\exp\left(\frac{1}{N_a}\sum_{i=1}^{N_a} x_{ia}\hat{\beta}_s\right)}{\exp\left(\frac{1}{N_a}\sum_{i=1}^{N_a} x_{ia}\hat{\beta}_0\right)}}_{\hat{I}_s^{J_p}}, \tag{25}$$

where $(a, b) = (0, s)$ or $(s, 0)$ and $\hat{I}_s^{J_q}$ and $\hat{I}_s^{J_p}$ are consistent estimators for, respectively, $I_s^{J_q}$ and $I_s^{J_p}$ of (8).

3.2.2 True and assumed data generating process: exponential hedonic function

On the other hand, when the true hedonic function has an exponential form, specification and estimation of an exponential regression model yields directly consistent estimates for the dwelling price, $\hat{p}_{it} = \exp\left(x_{it}\hat{\beta}_t^*\right)$, not for the logged prices that appear in the Jevons price index formula. Moreover, the naive estimator given by the logarithm of $\hat{p}_{it}$, $\ln\left(\hat{p}_{it}\right) = x_{it}\hat{\beta}_t^*$, is not in general a consistent estimator for $E\left[\ln\left(p_t\right)|x_{it}\right]$. Indeed, from (12), it follows that

$$E\left[\ln\left(p_{it}\right)|x_{it}\right] = x_{it}\beta_t^* + E\left(u_{it}^*|x_{it}\right), \tag{26}$$

where $E\left(u_{it}^*|x_{it}\right) \neq 0$; see Section 2.2.

To the best of our knowledge, the problem of going from level predictions to log predictions has never been analyzed in the econometrics literature. However, this is a very similar issue to that created by the assumption of a log-linear hedonic function in the Dutot framework, being necessary to estimate a bias correction. In the Jevons case, consistent estimation of logged dwelling prices requires the previous estimation of $E\left(u_{it}^*|x_{it}\right)$. Let

$$E\left(u_{it}^*|x_{it}\right) \equiv \mu_{it}^* = h\left(x_{it}^*\alpha_t^*\right), \tag{27}$$

where $h\left(\cdot\right)$ is a known function and α_t^* is a vector of parameters for which a consistent estimator, $\hat{\alpha}_t^*$, is available. Then, a consistent estimator for $\ln\left(p_{it}\right)$ is given by $\widehat{\ln\left(p_{it}\right)} = x_{it}\hat{\beta}_t^* + \hat{\mu}_{it}^*$, which yields the following estimator for the unadjusted Jevons price index:

$$\hat{I}_s^J = \frac{\exp\left\{\frac{1}{N_s}\sum_{i=1}^{N_s}\left[x_{is}\hat{\beta}_s^* + h\left(x_{is}^*\hat{\alpha}_s^*\right)\right]\right\}}{\exp\left\{\frac{1}{N_0}\sum_{i=1}^{N_0}\left[x_{i0}\hat{\beta}_0^* + h\left(x_{i0}^*\hat{\alpha}_0^*\right)\right]\right\}}. \tag{28}$$

This estimator may be decomposed as follows:

$$\hat{I}_s^J = \underbrace{\frac{\exp\left\{\frac{1}{N_s}\sum_{i=1}^{N_s}\left[x_{is}\hat{\beta}_b^* + h\left(x_{is}^*\hat{\alpha}_b^*\right)\right]\right\}}{\exp\left\{\frac{1}{N_0}\sum_{i=1}^{N_0}\left[x_{i0}\hat{\beta}_b^* + h\left(x_{i0}^*\hat{\alpha}_b^*\right)\right]\right\}}}_{\hat{I}_s^{Jq}} \underbrace{\frac{\exp\left\{\frac{1}{N_a}\sum_{i=1}^{N_a}\left[x_{ia}\hat{\beta}_s^* + h\left(x_{ia}^*\hat{\alpha}_s^*\right)\right]\right\}}{\exp\left\{\frac{1}{N_a}\sum_{i=1}^{N_a}\left[x_{ia}\hat{\beta}_0^* + h\left(x_{ia}^*\hat{\alpha}_0^*\right)\right]\right\}}}_{\hat{I}_s^{Jp}}. \tag{29}$$

Therefore, unless $\mu_{it}^* = \mu^*$, the estimation of quality-adjusted Jevons price indexes based on exponential hedonic functions requires the previous estimation of $\hat{\alpha}_s^*$ and $\hat{\alpha}_0^*$.

Similarly to the Dutot case, some assumptions may be made in order to simplify the estimation of μ_{it}^*. In particular, we may apply the same principle underlying the smearing estimator and estimate

$$\hat{\mu}_t^* = \sum_{i=1}^{N_t} \hat{u}_{it}^*, \tag{30}$$

provided that $E\left(u_{it}^*|x_{it}\right)$ does not depend on x_{it}; or we may assume that $\exp\left(u_{it}^*|x_{it}\right)$, in addition to unity mean, has a lognormal distribution with variance $\left[\exp\left(x_{it}^*\alpha_t^*\right)-1\right]$ such that u_{it}^* has a normal (conditional) distribution with mean given by

$$\mu_{it}^* = -0.5x_{it}^*\hat{\alpha}_t^*, \tag{31}$$

where $\hat{\alpha}_t^*$ results from regressing the squared residuals of the exponential hedonic function (plus one) on $\exp\left(x_{it}^*\alpha_t^*\right)$.

3.2.3 The link between Jevons price indexes and log-linear hedonic functions

Consider now the estimation of a log-linear model when the true hedonic function has an exponential form and it is further assumed that $h(\cdot)$ is a linear function:

$$\mu_{it}^* = x_{it}^*\alpha_t^*. \tag{32}$$

Then,

$$\begin{aligned} E\left[\ln\left(p_{it}\right)|x_{it}\right] &= x_{it}\beta_t^* + x_{it}^*\alpha_t^* \\ &= z_{it}\delta_t, \end{aligned} \tag{33}$$

which shows that assumptions for μ_{it}^* of the type made in (32), such as those underlying the smearing and the normal-smearing estimators, are easily accommodated by the log-linear model.

Irrespective of assumption (32) being true or not, and irrespective of the true generating process of dwelling prices being appropriately represented by an exponential function or not, using a log-linear hedonic function is the only form of ensuring that no bias corrections are necessary for computing quality-adjusted Jevons price indexes. Such indexes when based on an hedonic function that uses untransformed dwelling prices as dependent variable, even when it corresponds to the true specification, will require in general the specification and estimation of $E\left(u_{it}^*|x_{it}\right)$. Hence, *there exists a clear link between the computation of quality-adjusted Jevons price indexes and the log-linear hedonic function, in particular, and hedonic functions which consider logged dwelling prices as dependent variable, in general.*

4 Links in the context of the time dummy variable hedonic method

In Section 3, we assumed that the implicit prices of the dwelling characteristics change from one period to the other, which implies that separate hedonic regressions have to be estimated using the N_t observations available for each period and that unadjusted Dutot and Jevons price indexes may be decomposed as in (14) and (25), respectively. This decomposition method, known as the imputation price index method, is the most general technique for computing hedonic price indexes. However, there exist other hedonic methods, such as the also popular time dummy variable method; see *inter alia* Hill (2011) and Triplett (2006) for other alternatives.

The time dummy variable method assumes that the implicit prices of the dwelling characteristics are constant across a certain number of time periods. Let T denote that number of periods, let T_i be a vector of $T-1$ dummy variables whose elements T_{it} $(t = 1, ..., T-1)$ take the value unity if dwelling i was sold at period t (and zero otherwise), and let λ (λ^*) be the associated vector of coefficients with elements λ_t (λ_t^*). Let also r_{it} be a vector containing all dwelling characteristics other than the period of sale and θ (θ^*) be the associated vector of parameters. Under the assumption of that θ and θ^* are constant, only one hedonic function needs to be estimated for the whole period, using a sample that comprises observations from all the T periods. In the log-linear case, the hedonic function may be written as

$$\ln(p_{it}) = r_{it}\theta + T_{it}\lambda_t + u_{it}, \tag{34}$$

while in the exponential case it is given by

$$p_{it} = \exp\left(r_{it}\theta^* + T_{it}\lambda_t^* + u_{it}^*\right). \tag{35}$$

Under suitable assumptions, consistent predictors for logged dwelling prices in periods 0 and s are given by, respectively, $\widehat{\ln(p_{i0})} = r_{i0}\hat{\theta}$ and $\widehat{\ln(p_{is})} = r_{is}\hat{\theta} + \hat{\lambda}_s$ and consistent predictors for dwelling prices are given by, respectively, $\hat{p}_{i0} = \exp\left(r_{i0}\hat{\theta}^*\right)$ and $\hat{p}_{is} = \exp\left(r_{is}\hat{\theta}^* + \hat{\lambda}_s^*\right)$.

From (25), it follows that the quality-adjusted Jevons price index based on the log-linear hedonic function (34) simplifies to

$$\hat{I}_s^{J_p} = \frac{\exp\left(\frac{1}{N_a}\sum_{i=1}^{N_a} r_{ia}\hat{\theta} + \hat{\lambda}_s\right)}{\exp\left(\frac{1}{N_a}\sum_{i=1}^{N_a} r_{ia}\hat{\theta}\right)} = \exp\left(\hat{\lambda}_s\right), \tag{36}$$

which is a well known result in the hedonic literature and, in fact, the main attractiveness of using the time dummy variable method. For this reason, and because most authors seem to think that a similar result is not possible in the Dutot framework, there is an apparent consensus in the hedonic literature that there is a link between the time dummy variable method and the Jevons price index in the sense that only with this specific combination of hedonic methods and price indexes is the calculation of quality-adjusted price indexes substantially simplified. See, for example, Silver and Heravi (2007a), Diewert, Heravi and Silver (2009), Haan (2010) and Hill (2011), which, in their sections dedicated to the time dummy variable method, restrict their attention to Jevons indexes calculated from hedonic functions based on the logged price,[8] and Triplett (2006) and Diewert (2011), which consider a linear regression model and conclude that no expression similar to (36) is available in the Dutot framework. However, as shown next, a similar simplification applies to quality-adjusted Dutot price indexes when used in association with the exponential hedonic function (35). Indeed, from (14), it follows that:

$$\hat{I}_s^{D_p} = \frac{\frac{1}{N_a}\sum_{i=1}^{N_a} \exp\left(r_{ia}\hat{\theta}^* + \hat{\lambda}_s^*\right)}{\frac{1}{N_a}\sum_{i=1}^{N_a} \exp\left(r_{ia}\hat{\theta}^*\right)} = \exp\left(\hat{\lambda}_s^*\right), \tag{37}$$

which implies that the calculation of quality-adjusted price indexes using the time dummy variable method is as simple for Dutot indexes as for Jevons indexes. Therefore, unlike claimed by many, *there is no link between the time dummy variable method and the Jevons price index.*

Naturally, as for the imputation price method, the simple expressions (36) and (37) are valid only if the links detected in the previous section are respected or, alternatively, if the parameters that appear in the bias functions (16) and (27) are constant over time. Indeed, only under the latter assumption is the quality-adjusted Jevons price index based on an exponential hedonic function given by

$$\hat{I}_s^{J_p} = \frac{\exp\left\{\frac{1}{N_a}\sum_{i=1}^{N_a}\left[r_{ia}\hat{\theta}^* + \hat{\lambda}_s^* + h\left(r_{ia}^*\hat{\alpha}^*\right)\right]\right\}}{\exp\left\{\frac{1}{N_a}\sum_{i=1}^{N_a}\left[r_{ia}\hat{\theta}^* + h\left(r_{ia}^*\hat{\alpha}^*\right)\right]\right\}} = \exp\left(\hat{\lambda}_s^*\right) \tag{38}$$

and the quality-adjusted Dutot price index based on a log-linear hedonic function given by:

$$\hat{I}_s^{D_p} = \frac{\frac{1}{N_a}\sum_{i=1}^{N_a} \exp\left[r_{ia}\hat{\theta} + \hat{\lambda}_s + g\left(r_{ia}^*\hat{\alpha}\right)\right]}{\frac{1}{N_a}\sum_{i=1}^{N_a} \exp\left[r_{ia}\hat{\theta} + g\left(r_{ia}^*\hat{\alpha}\right)\right]} = \exp\left(\hat{\lambda}_s\right). \tag{39}$$

Although, given the parameter constancy assumed for the hedonic function, the assumption of

[8] For instance, Hill (2011, p. 40) writes: 'The index could be constructed using the time-dummy, imputation or characteristics methods. For the former, the next task is to choose a functional form for the hedonic model. For the latter two methods, it is necessary to choose both a price index formula and a functional form.'

constant α_t and α_t^* is probably more plausible in this context than in the case of the imputation price method, we should not take for granted the validity of that assumption and be aware that, in case of invalidity, the same bias corrections derived in the previous section also apply to the time dummy variable method.

Many authors in the hedonic literature have pointed out the fact that, although $\hat{\lambda}_s$ is an unbiased estimator of λ_s, $\exp\left(\hat{\lambda}_s\right)$ is a consistent but not unbiased estimator of $\exp\left(\lambda_s\right)$.[9] As suggested by Goldberger (1968), under the assumption of a normal and homoskedastic error term, a less biased (but still not unbiased) estimator is given by $\exp\left(\hat{\lambda}_s - 0.5\hat{\sigma}^2_{\hat{\alpha}_s}\right)$, where $\hat{\sigma}^2_{\hat{\alpha}}$ is the OLS estimator of the variance of $\hat{\alpha}$. Typically, the effect of this bias correction in the computation of hedonic indexes is quite small, which is not surprising since it vanishes asymptotically; see *inter alia* Berndt (1991, p. 144), van Dalen and Bode (2004), Triplett (2006) and Syed, Hill and Melser (2008). The resemblance of the expressions defining this bias correction and one of those analyzed in the previous section (see equation 20), which seems to have confused some authors, and the minimal practical utility of the former correction, are possibly the main reasons why some practitioners still do not apply in empirical work the latter correction or respect the links identified in Section 3. In fact, it is quite puzzling the large attention that Goldberger's (1968) correction has received in the hedonic literature on the time dummy variable method, in contrast to the null discussion of the much more relevant bias corrections discussed in the previous section, which, unlike the former, may be essential for obtaining consistent estimators in cases where the links identified in Section 3 are not respected.

5 Links between estimation methods and hedonic functions

A final issue that is worth to investigate is the relation between the two estimators of unadjusted Dutot and Jevons price indexes that were introduced in Section 2.1: the sample estimators $\bar{I}_s^D$ of (1) and $\bar{I}_s^J$ of (2) and the hedonic estimators $\hat{I}_t^D$ of (9) and $\hat{I}_t^J$ of (10). Taking into account that the former are the most natural and simple estimators for the population indexes I_s^D and I_s^J, it is specially interesting to use hedonic estimators that, besides being decomposable into a quality and a price component, are equal to the corresponding sample estimators. Next, we discuss, first for Jevons indexes and then for the Dutot case, under which circumstances sample and hedonic estimators produce identical estimates of unadjusted price variation.

Comparing expressions (2) and (10), it follows that a sufficient condition for ensuring that

[9] The same applies to λ_s^* and $\exp\left(\hat{\lambda}_s^*\right)$.

$\bar{I}_s^J = \hat{I}_s^J$ is given by:

$$N_t^{-1} \sum_{i=1}^{N_t} \ln p_{it} = N_t^{-1} \sum_{i=1}^{N_t} \widehat{\ln (p_{it})}. \tag{40}$$

In general, this equality does not hold. However, as noted by Reis and Silva (2006), there is a very simple, common case in which equation (40) is satisfied. When the hedonic function has a log-linear specification and the parameters of the model are estimated by OLS, the estimator $\hat{\beta}_t$ for β_t satisfies the following set of orthogonality conditions between the residuals $\hat{u}_{it}$ and the explanatory variables:

$$\sum_{i=1}^{N_t} x_{it}^{'} \hat{u}_{it} = \sum_{i=1}^{N_t} x_{it}^{'} \left(\ln p_{it} - \widehat{\ln p_{it}} \right) = 0. \tag{41}$$

Typically, as we are also assuming throughout this paper, x_{it} includes an intercept, implying that $\sum_{i=1}^{N_t} \hat{u}_{it} = 0$ and, hence, the averages of both the observed and OLS predicted logged prices are identical, as in (40).

Similarly, equality $\bar{I}_t^D = \hat{I}_t^D$ is only satisfied when the averages of observed and predicted prices are equal,

$$\frac{1}{N_t} \sum_{i=1}^{N_t} p_{it} = \frac{1}{N_t} \sum_{i=1}^{N_t} \hat{p}_{it}; \tag{42}$$

see expressions (1) and (9). Using the exponential regression model (12) as hedonic function, equation (42) is only satisfied when the parameters of interest β_t^* are estimated by the so-called Poisson pseudo maximum likelihood (PPML) method, where $\hat{\beta}_t^*$ is defined by the set of first-order conditions

$$\sum_{i=1}^{N_t} x_{it}^{\prime} \hat{u}_{it}^* = \sum_{i=1}^{N_t} x_{it}^{\prime} \left(p_{it} - \hat{p}_{it} \right) = 0. \tag{43}$$

No other alternative estimation method for exponential regression models, such as nonlinear least squares or other pseudo maximum likelihood methods, produces estimators that satisfy equation (42). The PPML method may be straightforwardly implemented in most standard econometric packages, including the recommended Eicker-White robust version; see Santos Silva and Tenreyo (2006) for details.

A very useful implication of equations (42) and (43) is that the process of producing Paasche-type quality-adjusted price indexes in the context of the imputation price method is substantially simplified. Consider $\hat{I}_s^{D_p}$ of (14), with $a = s$. Denote by $\bar{p}_s$ the arithmetic mean of the actual dwelling prices in period s. Using the PPML method to estimate $\hat{\beta}_t^*$, it follows from (43) that $\hat{I}_s^{D_p}$ reduces to:

$$\hat{I}_s^{D_p} = \frac{\bar{p}_s}{\frac{1}{N_s} \sum_{i=1}^{N_s} \exp \left(x_{is} \hat{\beta}_0 \right)}. \tag{44}$$

Similarly, $\hat{I}_s^{J_p}$ of (25), with $a = s$ and $\hat{\beta}_t$ estimated by OLS, may be simplified to

$$\hat{I}_s^{J_p} = \frac{\bar{p}_s}{\exp\left[\frac{1}{N_s}\sum_{i=1}^{N_s} x_{is}\hat{\beta}_0\right]}, \tag{45}$$

where $\bar{p}_s$ now denotes a geometric mean. These simplified Paasche-type price indexes are very attractive for statistical agencies because allow them to compute hedonic housing price indexes in a more timely and simple manner: the hedonic function needs to be estimated only at the base period.

Thus, *there is a link between Dutot price indexes, exponential hedonic functions and the PPML estimation method, on the one hand, and a link between Jevons price indexes, log-linear hedonic functions and the OLS method, on the other hand.* If these links are not respected, hedonic equations have to be estimated for each time period in the case of the imputation price method and sample and hedonic estimates of unadjusted price variation will be in general different, which is an uncomfortable result given that the main aim of applying the hedonic methodology is the decomposition into quality and price components of the observed (*sampled*) dwelling price variation. Nevertheless, note that, in terms of the latter feature, the consequences of not respecting these new links are negligenciable in asymptotic terms, since any alternative estimation method appropriate for the type of model specified produces consistent estimators for I_s^D or I_s^J.

6 Monte Carlo simulation study

This section investigates the finite sample properties of the estimators proposed in the previous sections for Paasche-type Dutot price indexes and the main consequences of using estimators that do not respect the identified links. Given the similarity of the conclusions achieved in some preliminary experiments involving Laspeyres-type and Jevons indexes, such indexes are not considered in this Monte Carlo study.

6.1 Experimental design

This paper claims the existence of a link between the exponential regression model and Dutot price indexes. Therefore, we are particularly interested in investigating the performance of that model when the true hedonic function has a log-linear specification in order to show that even in such a situation it is in general preferable to compute a Dutot price index using an exponential hedonic function than a log-linear one. To define the parameters and the variables of this hedonic function and in order to obtain a realistic scenario for this Monte Carlo study, we consider as a starting point for our experiments a dataset for the Canadian housing market, namely for the city of Windsor, which was first analyzed by Anglin and Gençay (1996). This dataset consists of 546 observations for the year of 1987, each with one continuous regressor, four count variables and six binary regressors. To simplify our investigation, without loss of generality, we consider only one of each type of explanatory variable in our Monte Carlo study, namely the natural logarithm of the lot size of the property in square feet (LOT), the number of bedrooms ($BDMS$) and a dummy variable which equals one if the dwelling is located in a preferred neighborhood of the city (REG). In all regressions the dependent variable is the sale price in Canadian dollars, divided by 100000, or its logarithm.

Regressing the logarithm of the price on a constant term and the three mentioned explanatory variables produces the following results:

$$\widehat{\ln (p_i)} = -4.809 + 0.460 LOT_i + 0.141 BDMS_i + 0.184 REG_i, \tag{46}$$

with $R^2 = 0.460$ and $\hat{\sigma}^2 = 0.075$, where σ^2 is the error term variance under the assumption of homoskedasticity. In order to establish a possible heteroskedasticity pattern, we regressed also the squared residuals, $\hat{u}_i^2$, of (46) on LOT_i and its square:

$$\hat{u}_i^2 = -0.007 LOT_i + 0.002 LOT_i^2 + error. \tag{47}$$

Therefore, in most experiments, we use the following log-linear model to generate the dwelling prices in each of the $t = 0, ..., 20$ periods that this study comprises:

$$\ln (p_{it}) = \beta_{t,0} + \beta_{t,1} LOT_{it} + \beta_{t,2} BDMS_{it} + \beta_{t,3} REG_{it} + u_{it}, \tag{48}$$

where $\beta_0' = [-4.809, 0.460, 0.141, 0.184]$ and the error term may be homoskedastic or heteroskedastic. In particular, we consider the following expressions to generate distinct patterns for the error term variance: (i) $\sigma_{it}^2 = 0.075$ (homoskedasticity); (ii) $\sigma_{it}^2 = \sigma_t^2 \in [0.075, 0.375]$ (time-varying error variance), either because $\sigma_t^2 = 0.075 + 0.015t$ or is randomly drawn from a Uniform distribution on that interval; and (iii) $\sigma_{it}^2 = -0.007 LOT_{it} + c_t LOT_{it}^2$, where $c_t \in [0.002, 0.010]$ (heteroskedasticity), with $c_t = 0.002 + 0.0004t$ or drawn from a Uniform distribution on the mentioned interval.

We draw u_{it} using three alternative distributions: a normal distribution $\mathcal{N}\left(0, \sigma_{it}^2\right)$, a displaced Gamma distribution $Gamma\left(\gamma^2 \sigma_{it}^2, \gamma\right) - \gamma \sigma_{it}^2$, where γ is a fixed parameter, and an Extreme Value or Gumbel distribution $Gumbel\left(-0.577216 \eta_{it}, \eta_{it}\right)$, where $\eta_{it} = \sqrt{6\sigma_{it}^2 / \pi^2}$. The reparametrization used in the last two cases ensures that, similarly to the normal distribution, the error term variance is simply σ_{it}^2, while $E\left[\exp\left(u_{it}\right)\right]$ is given by $\left[\gamma/\left(\gamma-1\right)\right]^{\gamma^2\sigma_{it}^2}\exp\left(-\gamma\sigma_{it}^2\right)$ in the Gamma case and $\exp\left(-0.577216\eta_{it}\right)\Gamma\left(1-\eta_{it}\right)$ in the Gumbel case; see *inter alia* Mood, Graybill and Boes (1974, pp. 540-543). As mentioned earlier, most empirical studies in this area typically assume a normal disturbance but the consideration of other distributions for the error term is important, for several reasons. First, it allows us to examine the robustness of the widely used normal-smearing estimator to distributional assumptions. Second, the Gamma distribution has two interesting features: it tends to the normal distribution as $\gamma^2\sigma_{it}^2$ gets larger; and by varying the parameter γ, it is possible to keep the error term variance σ_{it}^2 fixed while changing the value of $E\left[\exp\left(u_{it}\right)|x_{it}\right]$, which is not possible in the normal and Gumbel cases. This allows us to investigate whether larger deviations from the normal distribution and larger values of $E\left[\exp\left(u_{it}\right)|x_{it}\right]$ affect the ability of the different estimators to estimate $I_s^{D_p}$ in finite samples. Finally, when u_{it} has a Gumbel distribution, $E\left(p_{it}|\,x_{it}\right)$ may be written as:

$$\begin{aligned} E\left(p_{it}|\,x_{it}\right) &= \exp\left(x_{it}\beta_t\right)E\left[\exp\left(u_{it}\right)|\,x_{it}\right] \\ &= \exp\left(x_{it}\beta_t\right)\exp\left(-0.577216\eta_{it}\right)\Gamma\left(1-\eta_{it}\right) \\ &= \exp\left\{x_{it}\beta_t-0.577216\eta_{it}+\ln\left[\Gamma\left(1-\eta_{it}\right)\right]\right\} \\ &= \exp\left\{x_{it}\beta_t-\frac{\sqrt{6}}{\pi}\sigma_{it}+\ln\left[\Gamma\left(1-\frac{\sqrt{6}}{\pi}\sigma_{it}\right)\right]\right\}. \end{aligned} \tag{49}$$

While in case of homoskedasticity expression (49) reduces to the standard exponential regression model $\exp\left(x_{it}\beta_t^*\right)$ as in (22), the same does not happen under heteroskedasticity, since, unlike the normal and Gamma cases, (49) cannot be written as $\exp\left(z_{it}\delta_t^*\right)$ of (23). Thus, the consideration of a Gumbel distribution for the error term allows us to assess the robustness of the exponential regression model in cases where the true hedonic function cannot be expressed as a typical exponential regression model, i.e. with an index function linear in the parameters. Note that when u_{it} has a Gamma distribution, $E\left(p_{it}|\,x_{it}\right)$ is given by:

$$\begin{aligned} E\left(p_{it}|\,x_{it}\right) &= \exp\left(x_{it}\beta_t\right)\left[\gamma/\left(\gamma-1\right)\right]^{\gamma^2\sigma_{it}^2}\exp\left(-\gamma\sigma_{it}^2\right) \\ &= \exp\left(x_{it}\beta_t\right)\exp\left\langle\left\{\ln\left[\gamma^2/\left(\gamma-1\right)\right]-\gamma\right\}\sigma_{it}^2\right\rangle \\ &= \exp\left(z_{it}\delta_t^*\right), \end{aligned} \tag{50}$$

where $z_{it}=x_{it}$ (if σ_{it}^2 is constant across dwellings) or $z_{it}=\left(x_{it},LOT_{it}^2\right)$ (heteroskedastic case).

Regarding the dwelling characteristics, for period 0 we considered as base sample the original dataset of dwelling characteristics. For the remaining 20 periods, we constructed base samples as follows. First, we sorted the dwellings in the original sample according to the actual sale price of each dwelling. Then, we constructed four strata, where the first stratum contains the 25% cheapest dwellings, the second comprises the next 25% and so on. Let f_t be a four-element vector of probabilities assigned to each stratum. We next drew f_t from a Dirichlet distribution with parameter $\varsigma_t = \phi f_t^B$, where $\phi = 5$ is a precision parameter, $f_t^B = f_{t-1}^B + \Delta f_t^B$ is the expected value of f_t, $\Delta f_t^B = [-0.01, 0, 0.005, 0.005] * t$ and $f_0^B = [0.25, 0.25, 0.25, 0.25]$.[10] Then, for each time period, we generated a base sample of 546 observations, drawing with replacement from the original dataset a stratified sample based on f_t. Finally, for each one of the 21 time periods, we drew from the base samples, with replacement, 5000 random samples of N_t observations, where N_t was either set at the original sample size (experiments involving price prediction) or previously drawn from an Uniform distribution with limit points 250 and 500 in order to mimic the fact that with actual data the sample size typically differs across periods (all the remaining experiments). Experiments involving tenfold samples were also performed, in which case the same procedures where applied to generate the Monte Carlo sample but only after replicating the original sample ten times.

For the parameters of the hedonic function, we considered two alternative experimental designs. In the first (Design A), we consider $\beta_t = \beta_{t-1}(1 + \Delta\beta_t)$, $t \geq 1$, where β_0 was defined above and the four elements of $\Delta\beta_t$ are drawn independently from a Normal distribution with mean zero and variance 0.0001/50. In the second (Design B), we considered a similar setting but the variance of $\Delta\beta_t$ is multiplied by 50. Thus, while in Design A the parameters β_t are relatively stable over time, in Design B they display much more variability.

To illustrate the main practical characteristics of the experimental designs simulated, Figure 1 displays unadjusted and quality-adjusted population Dutot price indexes, as well as the associated quality index, for both Designs A and B when the error term has a normal distribution and its variance is defined according to the homoskedasticity (across dwellings) patterns (i) and (ii) defined above. These are fixed base indexes and represent 'population' indexes, since they were calculated using the base samples, the true β_t parameters and the known bias correction $E[\exp(u_{it})|x_{it}]$.

Figure 1 about here

As Figure 1 shows, although the simulated quality changes are identical across experiments, the pure price evolution is quite distinct in Designs A and B, being much more irregular and displaying much larger absolute variations in the latter case. As a consequence, in Design A I_s^D and $I_s^{D_q}$ display typically a similar evolving pattern, while in Design B the evolving of I_s^D

[10] See Kotz, Balakrishnan and Johnson (2000, ch. 49) for a general discussion of the Dirichlet distribution and Murteira, Ramalho and Ramalho (2011) for details on the reparameterized version considered in this paper.

is dictated mainly by the evolution of $I_s^{D_p}$. Note also that, although the parameters of the hedonic function are kept fixed across homoskedasticity patterns in each experimental design, both the unadjusted and quality-adjusted price indexes vary from experiment to experiment due to different assumptions on the error term variance. For example, consider the three graphs of the first row of Figure 1. In the first graph $I_s^{D_p}$ changes very little over time. In the second graph, as the result of an increasing variance of the error term, at a constant rate, $I_s^{D_p}$ also increases at a relatively constant rate. Finally, in the third graph, due to the random nature of σ_{it}^2, the time trajectory of $I_s^{D_p}$ is much less regular. This illustrates clearly the need for implementing bias corrections in cases where the hedonic function is specified in a scale that is not the one of interest for calculating the index: when the hedonic function is log-linear, while $\ln(p_{it})$, and Jevons price indexes, do not change as a result of a variation of σ_{it}^2, p_{it} and, hence, Dutot price indexes do change.

In all experiments that follow, two alternative hedonic functions are estimated: a log-linear function, which is estimated by OLS; and an exponential regression model, which is estimated by PPML. In both cases, we assume that all relevant regressors are known, which implies that in the latter case some additional regressors may be added to those that appear in equation (48) (i.e. when σ_{it}^2 depends on LOT_{it}^2, this variable is added to the original set of regressors). The only exception occurs when u_{it} has a Gumbel distribution, where in the computation of the PPML estimator we simply add LOT_{it}^2 to the original set of regressors and keep a linear index function, in spite of in this case the true hedonic function in the exponential scale, see equation (49), having other (functions of the included) regressors and being a nonlinear index model. In the log-linear case, three alternative predictors of dwelling prices are computed: (i) the naive OLS estimator, which first estimates $\ln(p_{it})$ and then uses its antilog as predictor of p_{it}; (ii) the normal-smearing OLS estimator (OLSn), which applies the bias correction (20) to the OLS estimator and assumes a normal-distributed error term and knowledge on the error term variance function; and (iii) the smearing OLS estimator (OLSs), which applies the bias correction (19) to the OLS estimator and assumes a homoskedastic error term. In all experiments we implement the OLSn estimator using the correct variance function for the error term.

6.2 The ability of alternative estimators for predicting dwelling prices

Before focussing on the main aim of this Monte Carlo study, the comparison of alternative hedonic estimators of quality-adjusted Dutot price indexes, we run a first set of Monte Carlo experiments to illustrate the effects of different assumptions on the error term over the prediction of dwelling prices by OLS, OLSn, OLSn and PPML. We simulate 5000 random samples of size

546 drawn with replacement from the actual sample of regressors. We used model (48) to generate dwelling prices for a single period (say, $t = 0$), considering the vector β_0 of implicit prices defined above and three distributions and two distinct variance functions for u_{i0}. For each sample, we first estimate the log-linear and exponential models and then, using the parameter estimates, compute the four alternative predictors of dwelling prices mentioned above. Figure 2 displays the average predictions across replications yielded by each estimator for dwellings which have three bedrooms and are not located in a preferred neighborhood of the city and whose lot size ranges, in steps of 50, from 1650 to 15600 square feet. The first two values represent the mode for the variables $BDMS$ and REG, while the boundary values chosen for the lot size are its minimum and maximum values in the sub-sample of dwellings for which $(BDMS, REG) = (3, 0)$. Figure 2 displays also the expected value of dwelling prices (denoted as $True$), i.e. the values obtained from (15) considering the true vector β_0, the original data for the regressors and the correct form for $E\left[\exp\left(u_{it}\right)|x_{it}\right]$.

Figure 2 about here

The first two graphs of the first row of Figure 2 consider the case of a normal, homoskedastic error term. In the first graph $\sigma_{i0}^2 = 0.075$ and in the second $\sigma_{i0}^2 = 0.375$. Clearly, the only estimator that yields predictions that may deviate substantially from the actual dwelling prices is the naive OLS estimator. As follows from equations (15) and (20), these deviations become more important as the variance of the error term increases.

The remaining graphs of the first row of Figure 2 consider non-normal distributions for u_{i0} but keep $\sigma_{i0}^2 = 0.375$. In all cases, both the naive and the normal-smearing estimator yield inconsistent predictions of dwelling prices, which shows that the latter estimator is not robust to deviations from the normal assumption made for the error term, unlike the PPML and OLSs estimators. When u_{i0} has a Gamma distribution, the deviations are more substantial for lower γ, which is a direct consequence of the higher value implied for $E\left[\exp\left(u_{i0}\right)|\, x_{i0}\right]$: for $\gamma = 1.5$, $E\left[\exp\left(u_{i0}\right)|\, x_{i0}\right] = 1.440$; for $\gamma = 3$, $E\left[\exp\left(u_{i0}\right)|\, x_{i0}\right] = 1.276$. The bias of the OLS estimator also increases with $E\left[\exp\left(u_{i0}\right)|\, x_{i0}\right]$, which equals 1.206 when u_{i0} has a normal distribution and 1.289 in the Gumbel case.

In the last row of Figure 2, we define the variance of u_{i0} as a function of both LOT_{i0} and LOT_{i0}^2, with the parameter associated to the latter variable being 0.002 or 0.01. Unlike the homoskedastic case, the smearing estimator clearly fails in consistently predicting dwelling prices, particularly when σ_{i0}^2 is more sensitive to LOT_{i0}^2, overpredicting the price for low values of the lot size and underpredicting it for high values. The non-robustness of the OLSn estimator to departures from a normal error term distribution is again apparent. In contrast, the PPML estimator performs well even in the case of a Gumbel-distributed error term, where the exponential regression model estimated is not well specified.

Overall, the PPML is the only estimator that performs well in all experiments, which confirms the ability of the exponential regression model to generate consistent predictions of dwelling prices even in cases where the true model has a log-linear functional form. Note that the small bias displayed in some cases by the PPML estimator for larger values of LOT is a small sample issue, disappearing asymptotically, as can be confirmed in Figure 3, based on 5000 samples of 5460 dwellings drawn with replacement from the original sample. In contrast, all the other biases illustrated in Figure 2 do not vanish as the sample size gets larger.

Figure 3 about here

6.3 Results for Paasche-type Dutot price indexes

In this section we present the main Monte Carlo results obtained for Paasche-type Dutot price indexes. We first report and analyze the results obtained in the framework of the imputation price method using the experimental designs previously described. Then, we make some minor adaptations to those designs and present results for the links concerning the use of the time dummy variable method and the estimation of the exponential regression model using different PPML methods.

6.3.1 Imputation price method

Figure 4 considers the case of a normal-distributed error term and reports four alternative quality-adjusted Dutot price indexes. When the error term variance is constant over time, the four estimators yield consistent (and indistinguishable) estimators for $I_s^{D_p}$. Note that in spite of OLS producing inconsistent estimates of dwelling prices, as shown in Figure 2, it estimates consistently $I_s^{D_p}$. This is because when σ_{it}^2 is constant across dwellings and over time, $E\left[\exp\left(u_{it}\right)|x_{it}\right]$ is also constant and, therefore, the correction factors considered in equation (18) cancel out. When σ_{it}^2 changes over time, then any estimates of $I_s^{D_p}$ based on the naive OLS estimator become inconsistent, while, as expected, all the other methods still give rise to consistent estimates. In fact, note that the estimates of $I_s^{D_p}$ produced by OLS are independent of the value of σ_{it}^2, while all the other estimators, automatically (PPML) or through a bias correction (OLSn and OLSs), have into account the effect of a varying error term variance in the unstransformed price scale, irrespective of the variation being constant or random.

Figure 4 about here

Figure 5 considers a similar experimental design (we omit the results for σ_{it}^2 constant over time) but the error term now has a displaced Gamma distribution, $G\left(\gamma^2\sigma_{it}^2, \gamma\right) - \gamma\sigma_{it}^2$, where $\gamma = 1.5$ or 3. Clearly, large deviations from the normal assumption required by the OLSn estimator, even when the error variance function is of a known form, may induce very large biases in the estimation of quality-adjusted Dutot price indexes. The naive estimator OLS displays even larger biases, while both the PPML and OLSs estimators are clearly robust to deviations from the normality assumption. Given that many empirical applications of hedonic methods to housing markets employ the OLSn estimator without testing the implied normality assumption (see e.g. the references given in the last paragraph of Section 3.1.2), we hope that in face of these results that practice is abandoned.

Figure 5 about here

Finally, Figure 6 considers heteroskedasticity across dwellings. Unlike predicted by the theory, the smearing estimator produces consistent estimates of $I_s^{D_p}$, in spite of dwelling prices not being consistently estimated for large values of the coefficient associated to LOT_{it}^2, as shown previously in Figure 2. We simulated many other experiments with different heteroskedasticity patterns across dwellings and in all cases we failed to find an example in which the OLSs-based hedonic indexes would deviate in a sizable manner from the true value of the quality-adjusted Dutot price indexes. Notice also the good performance of the PPML estimator, even in the Gumbel case.

Figure 6 about here

Overall, the results found in this section confirm the existence of a very convenient link between exponential hedonic functions and Dutot price indexes, since, despite the log-linear form of the hedonic function used to generate the prices, in all cases consistent estimates of $I_s^{D_p}$ were produced without necessity of applying any type of bias correction. Somewhat surprisingly, the smearing estimator seems to be an attractive alternative for computing $I_s^{D_p}$, given its apparent robustness to heteroskedasticity. Nevertheless, note that this estimator has the undesirable feature of yielding inconsistent estimates of dwelling prices under heteroskedasticity. Regarding the two most popular predictors of $I_s^{D_p}$ in applied work, the naive and the normal-smearing OLS estimators, this Monte Carlo investigation has shown clearly that, in general, they should be used only in very specific situations.

In order to examine whether the smearing estimator is really an attractive alternative to the PPML estimator for calculating quality-adjusted Dutot price indexes, we computed their root mean squared errors (RMSE) in the previous experiments. Figure 7 reports the results obtained for the Gamma ($\gamma = 1.5$) and Gumbel cases, for both $N_t \in \{250, 500\}$ and $N_t \in \{2500, 5000\}$. Clearly, another disadvantage of the smearing estimator relative to the PPML estimator is its much larger variability, especially for small samples. Interestingly, the OLSn estimator displays the lowest RMSE when the sample size is small. However, this can be hardly seen as a positive feature of this estimator: it just means that OLSn estimates are concentrated far away from the true price indexes. When the sample size increases, the PPML estimator is the best RMSE performer in most cases.

Figure 7 about here

6.3.2 Time dummy variable method

In contrast to the previous experiments, in this section we consider that the effect of the variables LOT, $BDMS$ and REG on dwelling prices is constant over time, so that the performance of the alternative estimators for Dutot price indexes can be evaluated in the context of the time dummy variable method. Therefore, the dwelling prices are now generated using the log-linear model that corresponds to equation (34):

$$\ln(p_{it}) = -4.809 + 0.460LOT_{it} + 0.141BDMS_{it} + 0.184REG_{it} + \lambda_1 T_{i1} + ... + \lambda_{T-1} T_{iT-1} + u_{it}, \quad (51)$$

where the parameters λ_t are drawn independently from a Normal distribution with mean zero and variance 0.015. In order to avoid very time-consuming experiments as consequence of pooling all observations to estimate equation (51), we set $T = 11$.[11] The quality-adjusted Dutot price indexes based on an exponential hedonic function was estimated using the simplified formula (37). The remaining characteristics of the Monte Carlo study remain unchanged.

Figure 8 exhibits the results obtained for some of the experimental designs considered before. In all cases, conclusions similar to those of the previous section are achieved, which demonstrates that the link between Dutot price indexes and exponential hedonic functions is valid irrespective of the hedonic method applied. In particular, note again the deleterious effect of heteroskedasticity on the bias of the OLS and (for non-normal error terms) OLSn estimators. These results confirm also that, similarly to the Jevons / log-linear hedonic function case, quality-adjusted Dutot price indexes based on exponential hedonic functions may also be computed simply as $\exp(\lambda_s)$, as argued in Section 4.

Figure 8 about here

[11] Actually, in applied work, the number of time periods pooled together is typically very small.

6.3.3 Alternative estimation methods for exponential hedonic functions

Finally, in this section we examine one of the links established in Section 5, namely that relative to the connection between Dutot price indexes, exponential hedonic functions and the PPML method. To perform this investigation, we consider now two alternative estimation methods to PPML: nonlinear least squares (NLS) and the Gamma pseudo maximum likelihood (GPML) method. As discussed in Santos Silva and Tenreyo (2006), the main difference between these methods is the functional form assumed for the conditional variance of $\exp(u_{it}^*)$ in (12):

$$V\left[\exp\left(u_{it}^*\right)\middle|\, x_{it}\right] = \tau \exp\left(x_{it}\beta_t^*\right)^{-\rho}, \tag{52}$$

where $\rho = 0$ (GPML), $\rho = 1$ (PPML) or $\rho = 2$ (NLS) and τ is a constant term. From (12) and (52), it follows that

$$V\left(p_{it}\middle|\, x_{it}\right) = \tau \exp\left(x_{it}\beta_t^*\right)^{2-\rho}, \tag{53}$$

that is, the NLS estimator assumes a constant conditional variance for dwelling prices, $V\left(p_{it}\middle|\, x_{it}\right) = \tau$, the PPML estimator assumes that the conditional variance is proportional to the conditional mean, $V\left(p_{it}\middle|\, x_{it}\right) = \tau E\left(p_{it}\middle|\, x_{it}\right)$, and the GPML estimator assumes that the conditional variance is a quadratic function of the conditional mean, $V\left(p_{it}\middle|\, x_{it}\right) = \tau E\left(p_{it}\middle|\, x_{it}\right)^2$. These different assumptions on $V\left(u_{it}^*\middle|\, x_{it}\right)$ imply also that the set of first-order conditions defining each estimator is given by:

$$\sum_{i=1}^{N_t} x_{it}'\left[p_{it} - \exp\left(x_{it}\beta_t^*\right)\right]\exp\left(x_{it}\beta_t^*\right)^{\rho-1} = 0 \tag{54}$$

See Santos Silva and Tenreyo (2006) for a comprehensive analysis of the three estimators.

Clearly, only when $\rho = 1$ (PPML estimator) are the averages of observed and predicted dwelling prices identical and, hence, as noted in Section 5: the sample and hedonic estimators of unadjusted price variation are also identical; and the Paasche-type quality-adjusted Dutot price index based on the imputation price method can be computed estimating the hedonic function only at the base period. In spite of these advantages of the PPML estimator, as the other alternative methods also produce consistent estimators for $I_s^{D_p}$, there may be circumstances where it may be preferable to use the NLS or the GPML estimator. For example, if the true error term variance is given by (52) but $\rho \neq 1$, then more precise estimators are potentially obtained if the GPML (ρ close to 0) or NLS (ρ close to 2) estimators are employed. Next, we investigate this issue using Monte Carlo methods.

In contrast to the previous experiments, now we use the following exponential hedonic function to generate dwelling prices:

$$p_{it} = \exp\left(\beta^*_{t,0} + \beta^*_{t,1} LOT_{it} + \beta^*_{t,2} BDMS_{it} + \beta^*_{t,3} REG_{it} + u^*_{it}\right), \qquad (55)$$

where $\beta_0^{*\prime} = [-4.770, 0.458, 0.147, 0.168]$, which is the set of coefficients that results from estimating equation (55) using the original data set, and $\exp(u^*_{it})$ is a lognormal random variable with mean one and variance as in (52), with $\tau = 1$ and $\rho = -1, 0, 1, 2$. The dwelling characteristics are generated as in the other experiments, while for β^*_t we consider a similar experimental design to that defining Design B in the previous analysis.

Figure 9 displays 95% and 99% confidence intervals and RMSE for quality-adjusted Dutot price indexes. While all methods produce very similar 95% confidence intervals, the other statistics analyzed in Figure 9 show clearly that NLS estimators are often much less precise than their competitors, which is a consequence of the extreme values that NLS occasionally yields. These results mimic the erratic behavior of NLS in the estimation of regression coefficients already detected in studies by Manning and Mullahy (2001) and Santos Silva and Tenreyo (2006). Therefore, there is strong evidence that NLS should not be used for estimating exponential hedonic functions. Regarding PPML and GPML, no substantial efficiency gains arise from using one or the other estimator, so, given the attractive features of the former estimator discussed before, in general there will be no reasons for using other estimator than PPML in this context.

Figure 9 about here

7 Concluding remarks

Quality-adjusted house price indexes are typically computed using hedonic pricing methodologies. In practice, the various choices underlying the estimation of hedonic indexes (price index formula, hedonic function, hedonic method, estimation method) are usually made independently from one another. In this paper, we have shown that there is a strong link between Dutot price indexes, exponential hedonic functions and the PPML method, on the one hand, and Jevons price indexes, log-linear hedonic functions and OLS estimation, on the other hand. In fact, when these links are ignored, the implicit price of each dwelling characteristic is a function not only of the parameters that appear in the hedonic function but also of the nuisance parameters that characterize the error term distribution, which leads to the counterintuitive result that the constancy of all regression coefficients of the hedonic function does not always imply null price inflation. To illustrate the importance of respecting the identified links, we provided a comprehensive Monte Carlo analysis of the substantial biases that may arise in the construction of

Dutot (Jevons) price indexes when a log-linear (exponential) function is used to relate dwelling prices and characteristics and wrong assumptions are made on the error term distribution.

The use of exponential regression seems to have never been previously considered in the hedonic literature but proves clearly to be more useful to deal with hedonic price indexes than the more popular linear regression model. Besides the obvious advantage of taking into account the positiveness of dwelling prices, the use of an exponential hedonic function in the context of the time dummy variable method allows the computation of quality-adjusted Dutot price indexes simply as the exponential transformation of a time dummy variable coefficient. So far, such simplification was thought to be valid only for computing Jevons indexes based on log-linear hedonic functions.

In this paper we focussed on the construction of Dutot and Jevons hedonic indexes. However, there are alternative price index formulas that are commonly used in the computation of housing quality-adjusted price indexes, such as Fisher and Torqvist indexes. Actually, based on the so-called economic (Diewert, 1976) and axiomatic (Balk, 1995) approaches, see Hill (2011), many authors recommend the use of Fisher and Torqvist indexes over the elementary indexes discussed in this paper. The Fisher quality-adjusted price index (I_s^F) is given by the geometric mean of Laspeyres and Paasche quality-adjusted Dutot indexes,

$$I_s^F = \sqrt{\frac{E\left(p_s|\, x_{i0}\right)}{E\left(p_0|\, x_{i0}\right)}\frac{E\left(p_s|\, x_{is}\right)}{E\left(p_0|\, x_{is}\right)}}, \tag{56}$$

while the Tornqvist quality-adjusted price index (I_s^T) is given by the geometric mean of Laspeyres and Paasche quality-adjusted Jevons indexes,

$$I_s^T = \sqrt{\frac{\exp\left\{E\left[\ln\left(p_s|\, x_{i0}\right)\right]\right\}}{\exp\left\{E\left[\ln\left(p_0|\, x_{i0}\right)\right]\right\}}\frac{\exp\left\{E\left[\ln\left(p_s|\, x_{is}\right)\right]\right\}}{\exp\left\{E\left[\ln\left(p_0|\, x_{is}\right)\right]\right\}}}. \tag{57}$$

Clearly, given that they are a function of two versions of either I_s^D or I_s^J, all the links identified is this paper are also (and probably even more) relevant for the computation of Fisher and Torqvist quality-adjusted price indexes.

Finally, note that all conclusions of the paper apply also to other markets with similar characteristics to the housing sector (e.g. the art market, where each art work is unique and its price is rarely observed even over periods spanning decades; see Collins, Scorcu and Zanola, 2009). With a few adaptations, namely the use of weighting schemes along the lines of Reis and Santos Silva (2006), the analysis of this paper will apply also to heterogeneous goods frequently transacted.

References

Ai, C. and Norton, E.C. (2000), "Standard errors for the retransformation problem with heteroscedasticity", *Journal of Health Economics*, 19, 697-718.

Anglin, P.M. and Gençay, R. (1996), "Semiparametric estimation of an hedonic price function", *Journal of Applied Econometrics*, 11, 633-648.

Balk, B.M. (1995), "Axiomatic price index theory: a survey", *International Statistical Review*, 63(1), 69-93.

Balk, B.M. (2008), *Price and Quantity Index Numbers: Models for Measuring Aggregate Change and Difference*, Cambridge University Press.

Berndt, E.R. (1991), *The Practice of Econometrics. Classic and Contemporary*, Addison Wesley.

Collins, A., Scorcu, A., and Zanola, R. (2009), "Reconsidering hedonic art price indexes", *Economics Letters*, 104(2), 57-60.

Coulson, N.E. (2011), "House price index methodologies", in S. J. Smith, M. Elsinga, L. Fox-O'Mahony, S. E. Ong and S. Wachter (eds.), *International Encyclopedia of Housing and Home*, Elsevier.

Cropper, M.L., Deck, L.B., and Mcconnell, K.E. (1988), "On the choice of functional form for hedonic price functions", *Review of Economics and Statistics*, 70(4), 668-675.

Dalen, J. van and Bode, B. (2004), "Estimation biases in quality-adjusted hedonic price indices", Paper presented at the *SSHRC International Conference on Index Number Theory and the Measurement of Prices and Productivity*, Vancouver.

Diewert, W.E. (1976), "Exact and superlative index numbers", *Journal of Econometrics*, 4, 115-145.

Diewert, W.E. (2003), "Hedonic regressions: a review of some unresolved issues", *Proceedings of the 7th Meeting of the Ottawa Group*, Paris.

Diewert, W.E. (2011), "Alternative approaches to measuring house price inflation", mimeo.

Diewert, W.E., Heravi, S. and Silver, M. (2009), "Hedonic imputation versus time dummy indexes", in W.E. Diewert, J. Greenlees and C.R. Hulten (eds.), *Price Index Concepts and Measurement*, NBER, Studies in Income and Wealth, Vol. 70, Chicago: University of Chicago Press, 161-202.

Dorsey, R.E., Hu, H., Mayer, W.J., and Wang, H.C. (2010), "Hedonic versus repeat-sales housing price indexes for measuring the recent boom-bust cycle", *Journal of Housing Economics*, 19(2), 75-93.

Duan, N. (1983), "Smearing estimate: a nonparametric retransformation method", *Journal of the American Statistical Association*, 78(383), 605-610.

Fletcher, M., Gallimore, P., and Mangan, J. (2000), "Heteroskedasticity in hedonic house price models", *Journal of Property Research*, 17(2), 93-108.

García, J.A.B. and Hernández, J.E.R. (2007), "Housing and urban location decisions in Spain: an econometric analysis with unobserved heterogeneity", *Urban Studies*, 44(9), 1657-1676.

Goldberger, A.S. (1968), "The interpretation and estimation of Cobb-Douglas functions", *Econometrica*, 35(3/4), 464-472.

Goodman, A.C. and Thibodeau, T.G. (1995), "Age-related heteroskedasticity in hedonic house price equations", *Journal of Housing Research*, 6(1), 25-42.

Gouriéroux, C. and Laferrère, A. (2009), "Managing hedonic housing price indexes: the French experience", *Journal of Housing Economics*, 18(3), 94-108.

Haan, J. de. (2010), "Hedonic price indexes: a comparison of imputation, time dummy and 're-pricing' methods", *Jahrbucher fur National Okonomie und Statistik*, 230(6), 772-791.

Hill, R. (2011), "Hedonic price indexes for housing", OECD Statistics Working Papers 2011/1, OECD Publishing.

Hill, R.J. and Melser, D. (2008), "Hedonic imputation and the price index problem: an application to housing", *Economic Inquiry*, 46(4), 593-609.

Kotz, S., Balakrishnan, N. and Johnson, N.L. (2000), *Continuous Multivariate Distributions, Volume 1: Methods and Applications*, John Wiley & Sons.

Malpezzi, S., Chun, G.H., and Green, R.K. (1998), "New place-to-place housing price indexes for U.S. metropolitan areas, and their determinants", *Real Estate Economics*, 26(2), 235-274.

Manning, W.G. and Mullahy, J. (2001), "Estimating log models: to transform or not transform?", *Journal of Health Economics*, 20, 461-494.

Meulenberg, M.T.G. (1965), On the estimation of an exponential function", *Econometrica*, 33(4), 863-868.

Mullahy, J. (1998), "Much ado about two: reconsidering retransformation and the two-part model in health econometrics", *Journal of Health Economics*, 17, 247-281.

Murteira, J., Ramalho, E.A. and Ramalho, J.J.S., "Regression analysis of multivariate fractional data", mimeo.

Pakes, A. (2003), "A reconsideration of hedonic price indexes with an application to PC's", *American Economic Review*, 93(5), 1578-1596.

Reis, H. and Santos Silva, J.M.C. (2006), "Hedonic indexes for new passenger cars in Portugal (1997-2001)", *Economic Modelling*, 23(6), 890-908.

Santos Silva, J.M.C. and Tenreyo, S. (2006), "The log of gravity", *Review of Economics and Statistics*, 88(4), 641-658.

Silver, M. and Heravi, S. (2007a), "The difference between hedonic imputation indexes and time dummy hedonic indexes", *Journal of Business & Economic Statistics*, 25(2), 239-246.

Silver, M. and Heravi, S. (2007b), "Why elementary price index number formulas differ: evidence on price dispersion", *Journal of Econometrics*, 140, 874-883.

Syed, I., Hill, R.J., and Melser, D. (2008), "Flexible spatial and temporal hedonic price indexes for housing in the presence of missing data", Discussion Paper 2008-14, School of Economics, The University of New South Wales.

Stevenson, S. (2004), "New empirical evidence on heteroscedasticity in hedonic housing models", *Journal of Housing Economics*, 13(2), 136-153.

Triplett, J. (2006), *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes: Special Application to Information Technology Products*, OECD.

Figure 1: Population Dutot price indexes – $u_{it} \sim N(0, \sigma_t^2)$

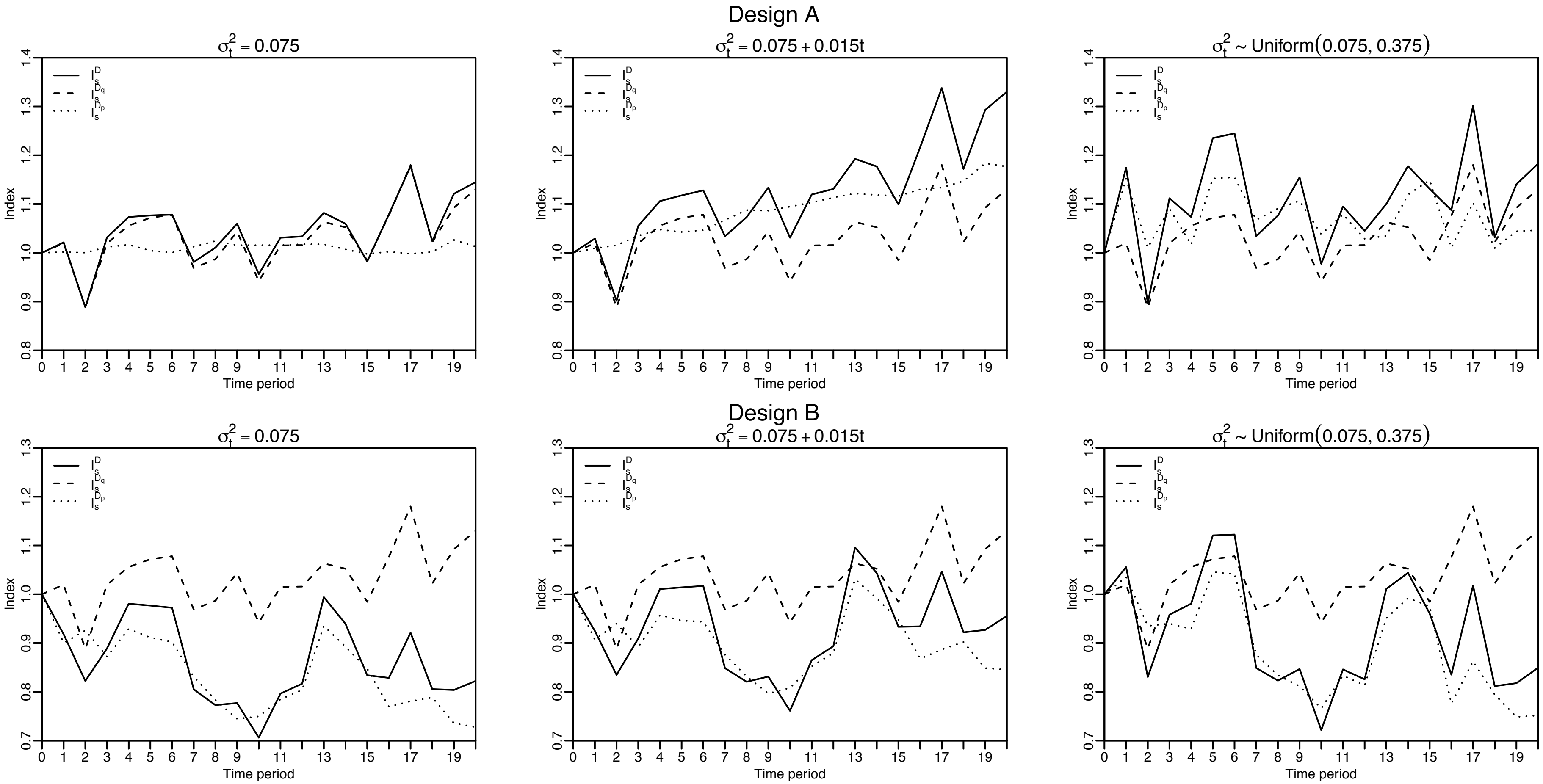

Figure 2: Alternative methods for predicting dwelling prices $\left(N_0 = 546\right)$

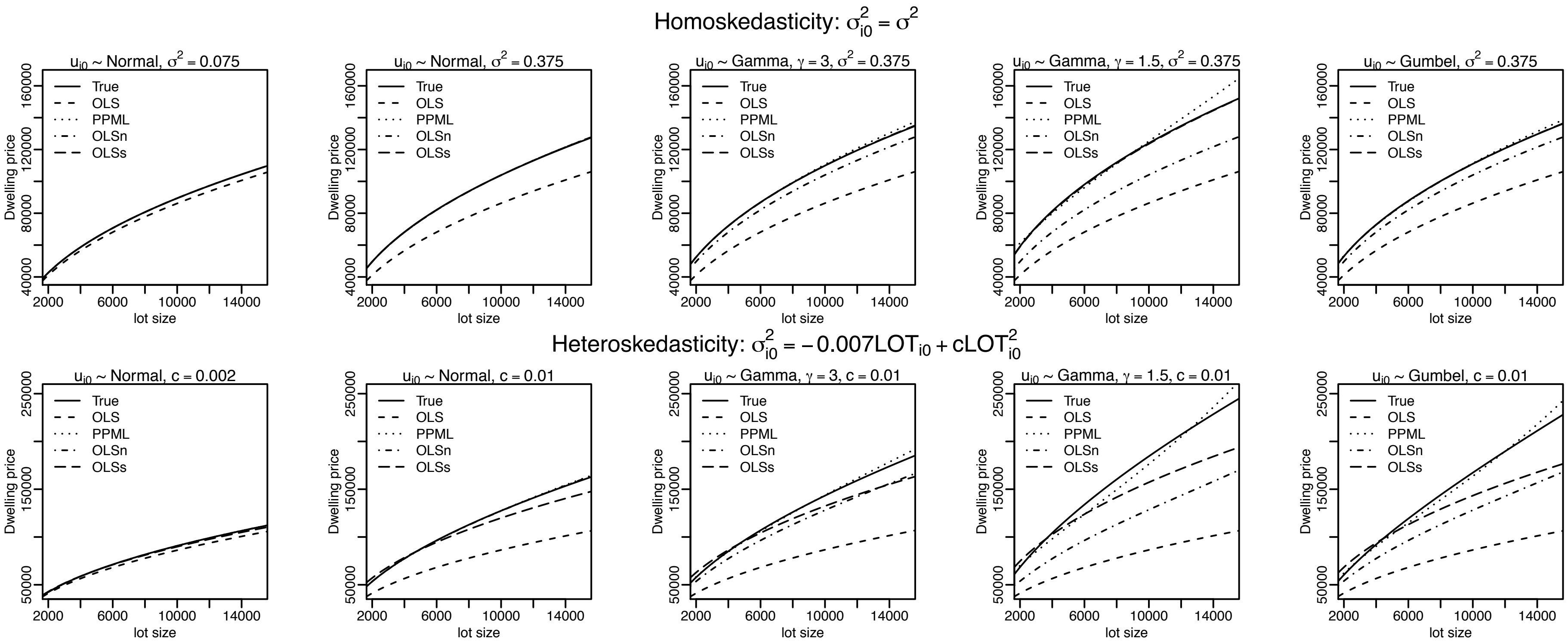

Figure 3: Alternative methods for predicting dwelling prices $(N_0 = 5460)$

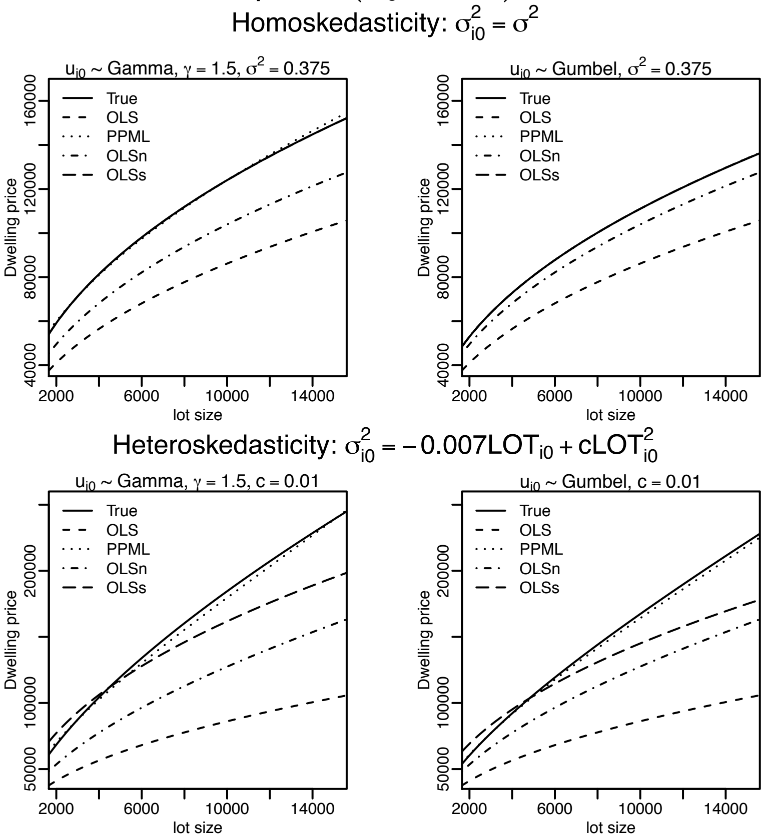

Figure 4: Quality−adjusted Dutot price indexes − homoskedasticity and time−varying error variance cases; $u_{it} \sim N(0, \sigma_t^2)$

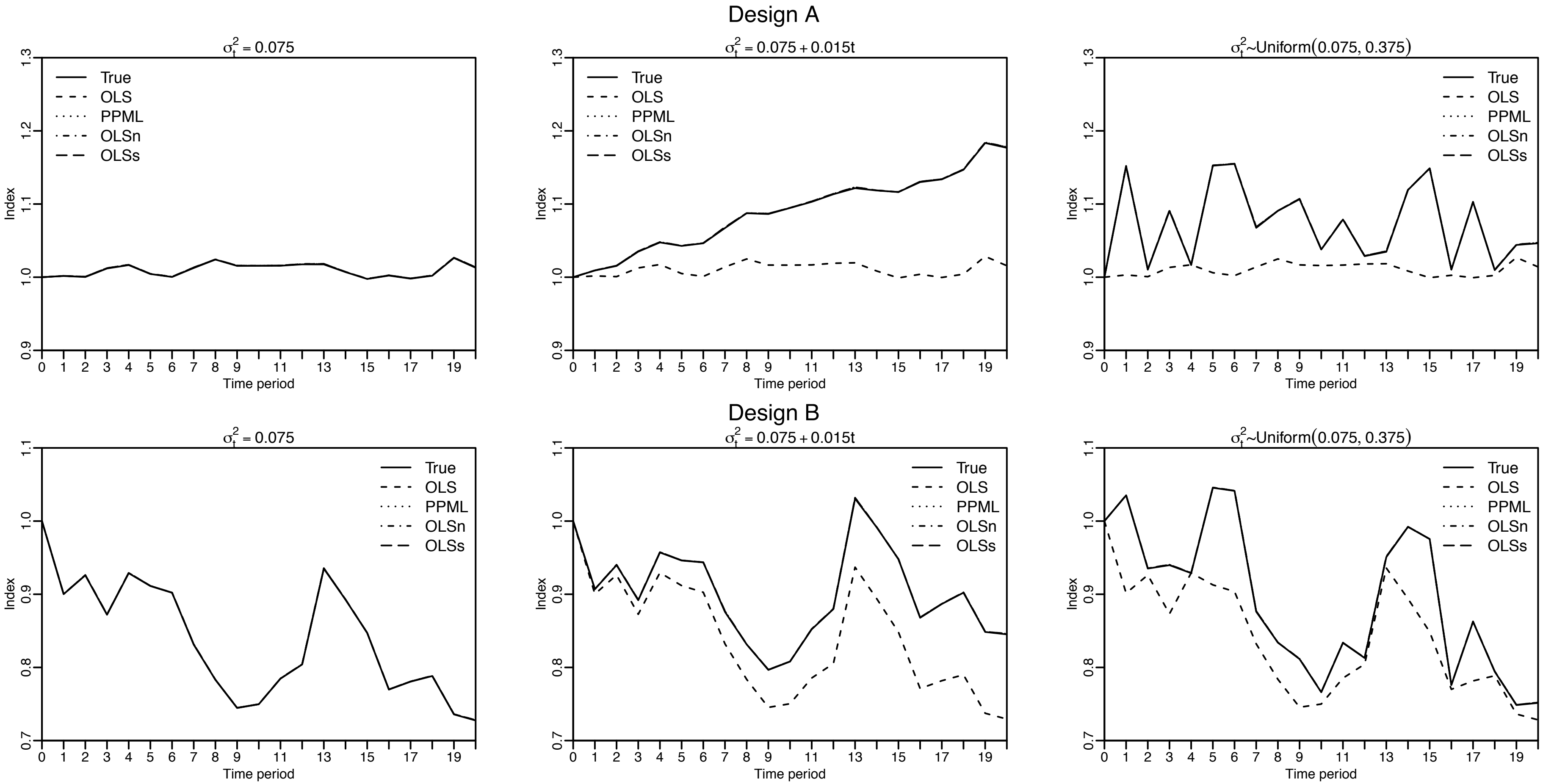

Figure 5: Quality–adjusted Dutot price indexes – time–varying error variance; $u_{it} = v_{it} - \gamma\sigma_t^2$, $v_{it} \sim \text{Gamma}\left(\gamma^2\sigma_t^2, \gamma\right)$

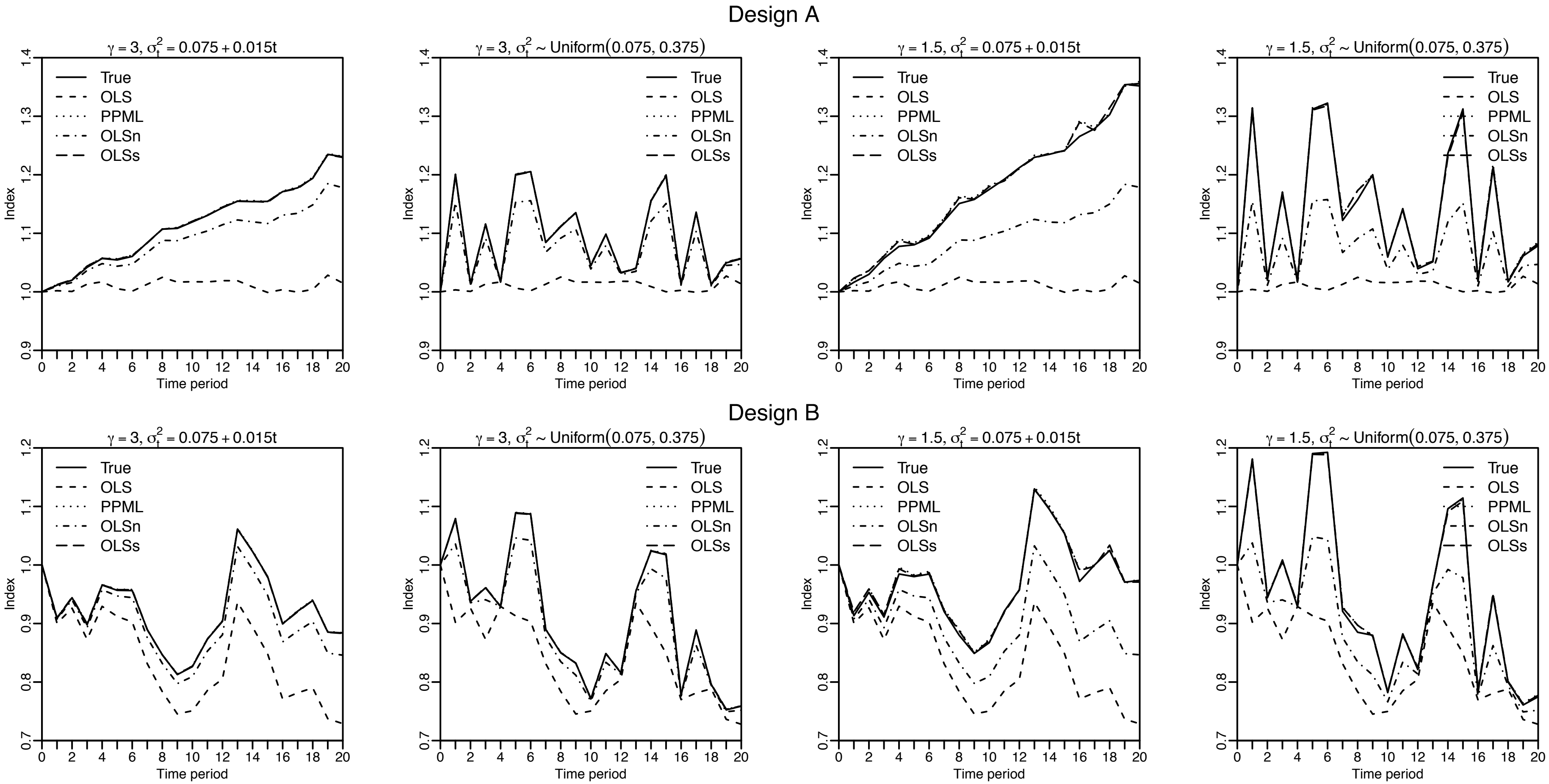

Figure 6: Quality−adjusted Dutot price indexes − heteroskedasticity: $\sigma^2_{it} = -0.007LOT_{it} + c_t LOT^2_{it}$

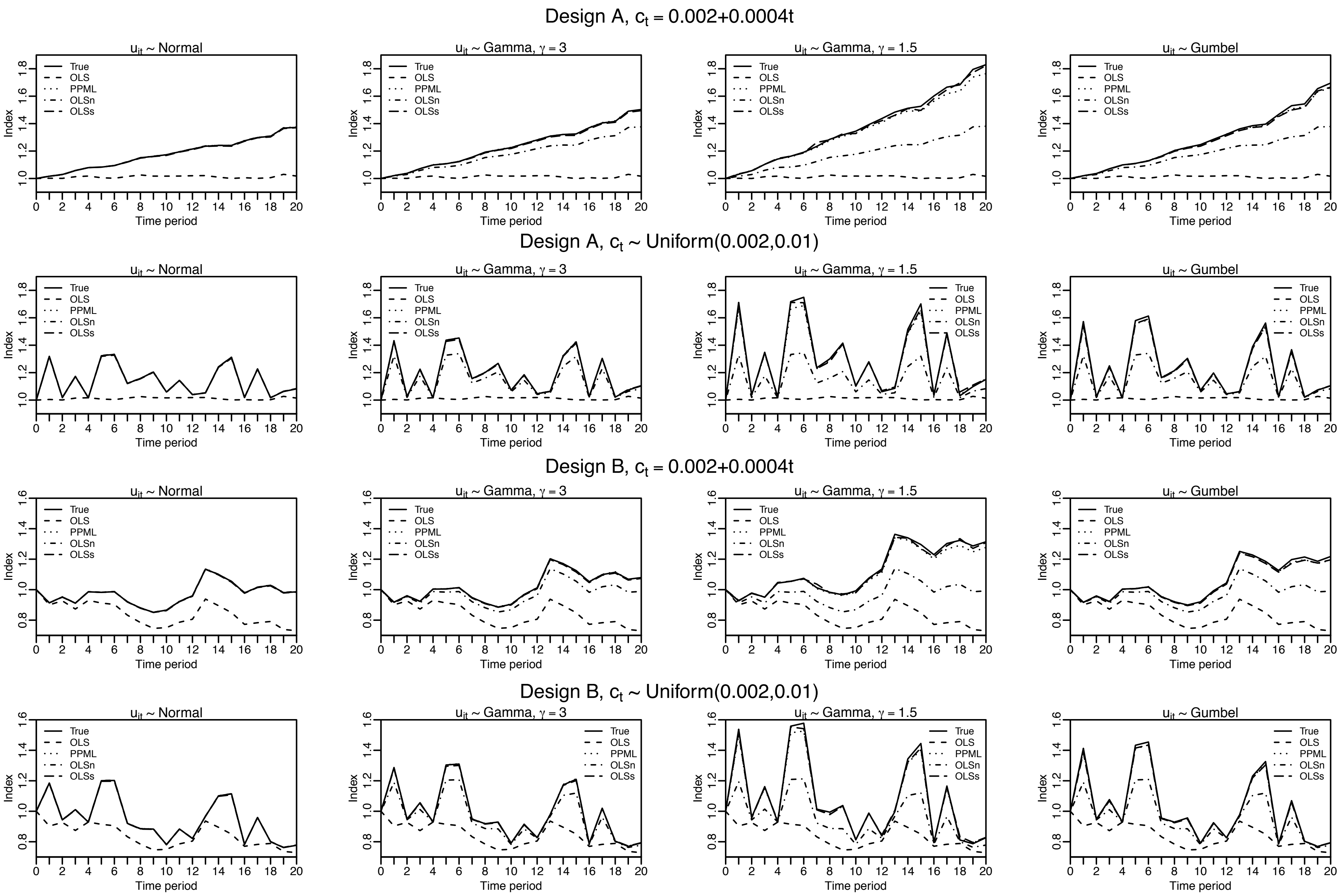

Figure 7: Root mean squared errors of quality-adjusted Dutot price indexes – heteroskedasticity: $\sigma^2_{it} = -0.007LOT_{it} + c_t LOT^2_{it}$

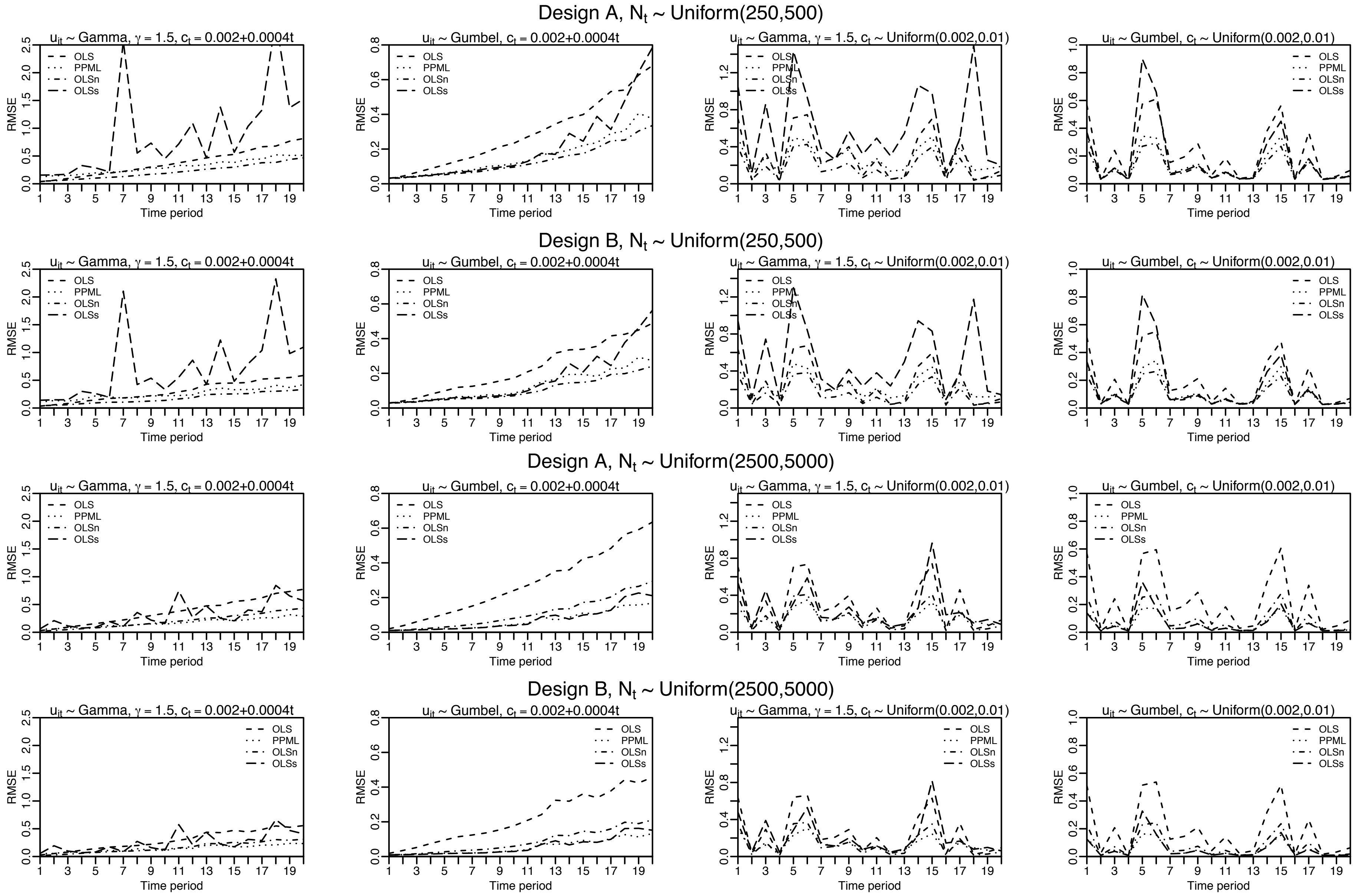

Figure 8: Quality−adjusted Dutot price indexes − time dummy variable method

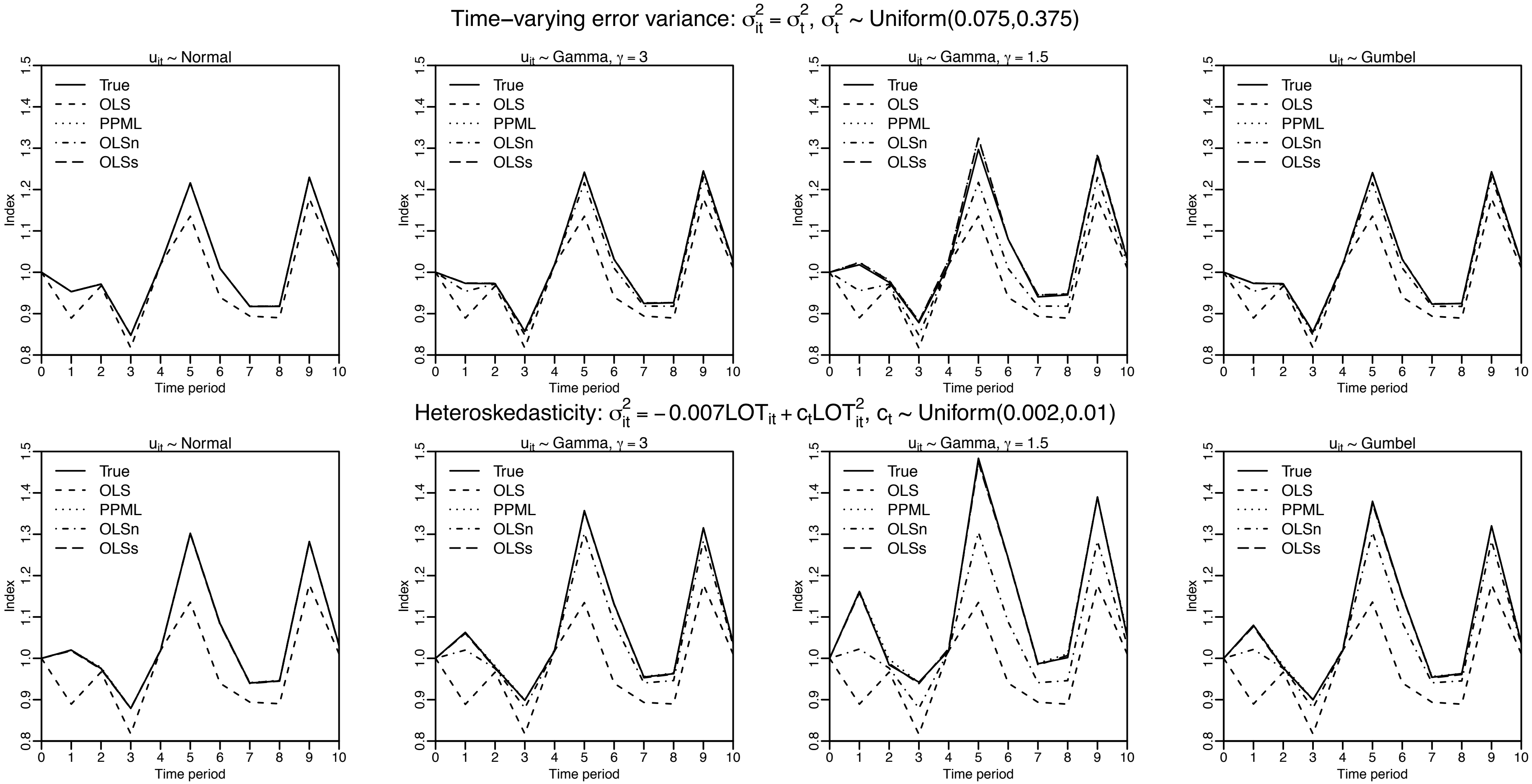

Figure 9: Alternative quality–adjusted Dutot price indexes based on exponential hedonic functions