SERC DISCUSSION PAPER **166**

Asymptotic Properties on Imputed Hedonic Price Indices

Olivier Schöni (SERC and University of Fribourg, Switzerland)

October 2014

This work is part of the research programme of the independent UK Spatial Economics Research Centre funded by a grant from the Economic and Social Research Council (ESRC), Department for Business, Innovation & Skills (BIS) and the Welsh Government. The support of the funders is acknowledged. The views expressed are those of the authors and do not represent the views of the funders.

Asymptotic Properties of Imputed hedonic Price Indices

Olivier Schöni

October 2014

* SERC and University of Fribourg, Switzerland

The present article is based on the early work contained in Schöni (2013). Therefore, I would like to sincerely thank Prof. Laurent Donzé for his exemplary guidance and advice.

Abstract
Hedonic price indices are currently considered to be the state-of-the-art approach to computing constant-quality price indices. In particular, hedonic price indices based on imputed prices have become popular both among practitioners and researchers to analyze price changes at an aggregate level. Although widely employed, little research has been conducted to investigate their asymptotic properties and the influence of the econometric model on the parameters estimated by these price indices. The present paper therefore tries to fill the actual knowledge gap by analyzing the asymptotic properties of the most commonly used imputed hedonic price indices in the case of linear and linearizable models. The obtained results are used to gauge the impact of bias adjusted predictions on hedonic imputed indices in the case of log-linear hedonic functions with normal distributed errors.

Keywords: Price indices, hedonic regression, imputation, asymptotic theory
JEL Classifications: C21; C43; C53; C58

1. Introduction

Price indices have become widespread in economic analyses to measure price changes of goods on an aggregate level with respect to a base period. They are defined according to standard formulae and do not involve any econometric model in their computation. However, a major drawback related to classic price indices is their inadequacy to take into account possible quality changes that may influence price changes at an aggregate level. This drawback has prompted a great amount of research, leading Rosen (1974) to define the economic framework necessary to the definition of the so-called hedonic price indices.

In the last thirty years, three main approaches based on imputed prices have established themselves as the most appropriate method for computing quality-adjusted hedonic price indices: Single, double, and characteristic imputation. In contrast to their unadjusted counterpart, hedonic indices do assume an underlying econometric model in their computation, thus worsening the price index problem[1], as illustrated by Hill and Melser (2008). Since economic theory does not provide indications on the choice of the imputation approach, researchers face the uncomfortable situation to choose among different imputation methods, wondering how econometric models estimated at the micro level affect affect the resulting price indices at the macro level.

A major domain of application of hedonic indices is represented by housing markets, where sound indicators of the general price level are of primary importance. Mark and Goldberg (1984) were among the first to compare hedonic imputed indices for housing goods to other classic price indices. Meese and Wallace (1991) and Wallace (1996) computed characteristic Fisher hedonic price indices based on non-parametric estimates of the hedonic regression function. Wallace and Meese (1997) compared characteristic hedonic indices with repeat-sales and hybrid-approaches. Kagie and Wezel (2007) used a boosting algorithm to improve the prediction performance of decision tree based hedonic models and computed single-imputed Fisher hedonic indices. Hill and Melser (2008) investigated the price index problem for single-imputed, double imputed, and characteristic hedonic price indices. von de Haan (2010) compared imputed Fisher and Törnquist

[1]The price index problem is defined as the non-equivalence of price index formulae: The computation of different price indices leads, in general, to different results.

hedonic indices to the time dummy method. Dorsey et al. (2010) used a double imputed Laspeyres index to measure the boom-bust housing cycle in the Los Angeles and San Diego metropolitan areas from 2000 to 2008. Diewert (2011) proposed, among other approaches, Laspeyres, Paashe, and Fisher imputed hedonic price indices to measure house price inflation.

As the above literature shows, research has focused on defining, computing, and finally comparing hedonic price indices. The computed hedonic indices aim to draw conclusions on the evolution of the market prices. However, these conclusions are usually based on informal considerations. In fact, hedonic price indices have been rarely employed as variable of interest in macro econometric analyses. This is mainly motivated by two factors. First, the unavailability of such indices: Often hedonic indices are computed by private real estate agencies that either do not publish them, or do not divulge the employed methodology. This situation, as suggested by the publication of the manuals Hill (2011), Eurostat (2012), OECD and Eurostat (2013), is probably going to change in the near future, since official statistic agencies are progressively adopting hedonic price indices. The second, and more important, reason is represented by the lack of theory surrounding these indices. Up to the present, hedonic indices have been used merely as descriptive statistical measures, neglecting the economic parameter they are actually estimating and the role played by the underlying econometric model.

The aim of the present paper is to fill the actual knowledge gap, by determining the asymptotic properties of Laspeyres, Paasche, and Fisher price indices for the single imputed, double imputed, and characteristic imputation methods in the case of linear and linearizable hedonic regression functions. This approach should provide a better understanding of the theoretical parameter a hedonic index estimates and the influence of the underlying econometric model on this parameter. In particular, a simulation study is performed to evaluate the impact of bias adjusted predictions on hedonic indices.

The present paper is structured as follows. Section 2 reviews the imputation approaches for the Laspeyres, Paasche, and Fisher price indices. The convergence in probability of hedonic price indices in the case of linear and linearizable hedonic functions is then established in Section 3. A simulation study to illustrate the obtained results is then performed in Section 4. Section 5 concludes the paper.

2. Hedonic imputed price indices: A review

The three main imputation approaches adopted in the literature are considered in the present paper to compute quality-adjusted price indices: Single, double, and characteristics imputation methods. Hedonic price indices based on time dummy variables are not analyzed in the present paper since they do not rely on traditional price index formulae. Moreover, only the hedonic counterpart of the classical Laspeyres, Paasche, and Fisher price indices are considered, although in section 3.3 we generalize the obtained results to composite price indices relying on these hedonic indices. The adopted terminology and the following definitions are based on Hill (2013).

Let $\mathbf{P}^t := (P_1^t, ..., P_{n_t}^t)' \in \mathbb{R}^{n_t}$ and $\mathbf{X}^t := (\mathbf{x}_1^t, ..., \mathbf{x}_{n_t}^t)' \in \mathbb{R}^{n_t \times K}$ denote a vector of independent random prices and a matrix of random characteristics in period t, respectively. The hedonic hypothesis states that in each time period the price of a good depends on its characteristics:

$$P_i^t = f^t(\mathbf{x}_i^t) + \epsilon_i^t = f^t(x_{i1}^t, ..., x_{iK}^t) + \epsilon_i^t, \ i = 1, ..., n_t, \tag{2.1}$$

where x_{ij}^t is the j-th characteristic of good i in period t, and the function f^t describes how the characteristics interact to build the price. The function f^t is usually called the hedonic regression function, or simply the hedonic function. We denote the hedonic function estimated in period t by $\hat{f}^t$. The set of the K observed characteristics is assumed constant through time, i.e., no new characteristic affecting the price appears in any time period $t = 1, ..., T$. The number of goods observed in period t is denoted by n_t, and it is assumed that $n_t \geq K$. The variable ϵ_i^t represents a stochastic error term with $\mathbb{E}(\epsilon_i^t) = 0 \ \forall i$.

According to (2.1), even if we identify a class of good with an appropriate set of characteristics, the price of the individual goods randomly varies. Therefore, assuming that characteristics and/or error terms are continuously distributed ,the quantity of a good purchased at a given price must be set equal to 1 in classic price index formulae. Let $\widehat{L}_{0,t}$, $\widehat{P}_{0,t}$, and $\widehat{F}_{0,t}$ denote the estimated Laspeyres, Paasche, and Fisher price index. Let us assume that the same set of n goods is observed in the base period 0 and the current period t. Under mild conditions on price distributions in the two time periods, the weak law of large numbers implies that

$$\widehat{L}_{0,t} = \widehat{P}_{0,t} = \widehat{F}_{0,t} = \frac{\sum_{i=1}^n P_i^t(\mathbf{x}_i^t)}{\sum_{i=1}^n P_i^0(\mathbf{x}_i^0)} \xrightarrow{P} \frac{\mu_{P^t}}{\mu_{P^0}} \text{ as } n \longrightarrow +\infty,$$

	Laspeyres	Paasche	Fisher
Single imputed	$\frac{\sum_{i=1}^{n_0} \hat{f}^t(\mathbf{x}_i^0)}{\sum_{i=1}^{n_0} P_i^0}$	$\frac{\sum_{i=1}^{n_t} P_i^t}{\sum_{i=1}^{n_t} \hat{f}^0(\mathbf{x}_i^t)}$	$\sqrt{\left(\frac{\sum_{i=1}^{n_0} \hat{f}^t(\mathbf{x}_i^0)}{\sum_{i=1}^{n_0} P_i^0}\right)\left(\frac{\sum_{i=1}^{n_t} P_i^t}{\sum_{i=1}^{n_t} \hat{f}^0(\mathbf{x}_i^t)}\right)}$
Double imputed	$\frac{\sum_{i=1}^{n_0} \hat{f}^t(\mathbf{x}_i^0)}{\sum_{i=1}^{n_0} \hat{f}^0(\mathbf{x}_i^0)}$	$\frac{\sum_{i=1}^{n_t} \hat{f}^t(\mathbf{x}_i^t)}{\sum_{i=1}^{n_t} \hat{f}^0(\mathbf{x}_i^t)}$	$\sqrt{\left(\frac{\sum_{i=1}^{n_0} \hat{f}^t(\mathbf{x}_i^0)}{\sum_{i=1}^{n_0} \hat{f}^0(\mathbf{x}_i^0)}\right)\left(\frac{\sum_{i=1}^{n_t} \hat{f}^t(\mathbf{x}_i^t)}{\sum_{i=1}^{n_t} \hat{f}^0(\mathbf{x}_i^t)}\right)}$
Characteristic	$\frac{\hat{f}^t(\overline{\mathbf{x}}^0)}{\hat{f}^0(\overline{\mathbf{x}}^0)}$	$\frac{\hat{f}^t(\overline{\mathbf{x}}^t)}{\hat{f}^0(\overline{\mathbf{x}}^t)}$	$\sqrt{\left(\frac{\hat{f}^t(\overline{\mathbf{x}}^0)}{\hat{f}^0(\overline{\mathbf{x}}^0)}\right)\left(\frac{\hat{f}^t(\overline{\mathbf{x}}^t)}{\hat{f}^0(\overline{\mathbf{x}}^t)}\right)}$

Table 1: Hedonic price indices with linear or linearizable hedonic functions

where μ_{P^0} and μ_{P^t} denote the mean price in the base and current period, respectively. In this case, classic Laspeyres, Paasche, and Fisher price indices thus converge toward the same unknown parameter given by the ratio of the mean prices in the two time periods. Although the same set of n goods has been observed in the two time periods, the estimated price ratio may be influenced by temporal changes in the underlying characteristics. Moreover, very often we don't observe exactly the same n goods in two time periods. This occurs, in particular, in the case of infrequently sold goods, e.g. houses, or goods possessing a rapidly changing technology, e.g. mobile phones, PC, etc. In this case, randomly sampling the same type of goods (e.g. single-family houses) in the two time periods and computing the ratio of prices $\sum_{i=1}^{n_t} P_i^t(\mathbf{x}_i^t)/\sum_{i=1}^{n_0} P_i^0(\mathbf{x}_i^0)$ leads to even greater quality variations, since the characteristics' distribution might have changed. To address this problem, several imputation approaches have been proposed. They are reviewed in the next section.

2.1. Imputation approaches in hedonic indices

Three main approaches have been proposed in the literature to cope with the quality variation problem: Single, double, and characteristic imputation. The index formulae of these approaches for the Laspeyres, Paasche, and Fisher indices are shown in Table 1 in the case of prices imputed in the original scale.

Let $\widehat{HIL}_{0,t}^{si}$, $\widehat{HIP}_{0,t}^{si}$, and $\widehat{HIF}_{0,t}^{si}$ denote the estimators of the Laspeyres, Paasche, and Fisher single imputed hedonic price indices, where 0 and t represent the base and current time periods, respectively. The base period is chosen among the time periods $t = 1, ..., T$. Single imputed hedonic price indices use the hedonic function to impute prices of each good in base/current period according to the hedonic function estimated in the other time period. Imputed prices in one period are then compared to observed prices in the other period.

In contrast to single imputed indices, double imputed hedonic price indices impute prices for both time periods. Once the hedonic functions for the two periods have been estimated, the set of characteristics in one period is evaluated according to the hedonic function estimated in the other period. By construction, this guarantees that the quality of the goods does not change between periods, and so prices can be directly compared. We denote the estimators of the Laspeyres, Paasche, and Fisher double imputed hedonic price indices with $\widehat{HIL}^{di}_{0,t}$, $\widehat{HIP}^{di}_{0,t}$ and $\widehat{HIF}^{di}_{0,t}$, where 0 and t represent the base and current time periods, respectively.

Instead of imputing prices for each good in a given period, characteristic hedonic price indices compute a representative good for one time period, and then impute the price of this characteristic good using the estimated hedonic functions in the two periods. The characteristic good is thought to appropriately represent the quality of the set of goods in one time period and is usually defined as being the mean vector of the characteristics. Also, in this case, since only the characteristic good is considered, quality does not change across periods, and prices are directly comparable. Let $\widehat{HIL}^{ch}_{0,t}$, $\widehat{HIP}^{ch}_{0,t}$, and $\widehat{HIF}^{ch}_{0,t}$ denote the estimators of the Laspeyres, Paasche, and Fisher characteristic hedonic price indices. The representative good in one time period is defined as the mean vector of the characteristics

$$\overline{\mathbf{x}}^t := (\overline{x}^t_1, ..., \overline{x}^t_K) = (\frac{1}{n_t}\sum_{i=1}^{n_t} x^t_{i1}, ..., \frac{1}{n_t}\sum_{i=1}^{n_t} x^t_{iK}).$$

3. Convergence in probability of imputed hedonic price indices

Although widely employed, the above defined indices have been used as empirical quantities, without knowing the parameter they are estimating. In particular, their asymptotic convergence has not been investigated, thus casting doubts on the use of such indices in official statistic and in macroeconomic analysis. To derive such properties, the hedonic function f^t considered in the hedonic hypothesis must be specified and an estimation technique accordingly adopted. We consider two case of figure, linear and linearizable hedonic functions.

In the rest of the paper we will refer to the *usual* hypothesis of the linear regression model. These hypothesis include a linear functional form of the data generating process, exogeneity of the regressors, linear independence of the regressors, and homoskedasticity of the error term (See Greene (2011), page 92 for an explicit statement of these hypotheses).

3.1. Imputed hedonic indices with linear or linearizable hedonic functions

The following linear hedonic model

$$P_i^t = f^t(x_{i1}^t, ..., x_{iK}^t) = \mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t = \beta_0^t + \beta_1^t x_{i1}^t + ... + \beta_K^t x_{iK}^t,\ i = 1, ..., n_t \tag{3.1}$$

is first assumed in each time period. Refer to Table 1 for the computation of imputed hedonic indices according to this model. The following proposition identifies the theoretic parameters toward which imputed hedonic price indices converge.

Theorem 3.1. *Let $(P_i^t, \boldsymbol{x}_i^t)$, $i = 1, ..., n_t$ be a random sample of n_t independent random variables belonging to period t $(t = 1, ..., T)$. We assume that the characteristics' vector $\boldsymbol{x}_i^t$ are i.i.d. with $\boldsymbol{\mu}_{\boldsymbol{x}^t} = \mathbb{E}(\boldsymbol{x}^t) < +\infty$ $\forall t$. If the usual hypotheses of the linear hedonic model in (3.1) hold in each time period, then*

i) $\widehat{HIL}_{0,t}^{si}$, $\widehat{HIL}_{0,t}^{di}$, *and* $\widehat{HIL}_{0,t}^{ch}$ *converge in probability toward* $\frac{\boldsymbol{\mu}_{\boldsymbol{x}^0}'\boldsymbol{\beta}^t}{\boldsymbol{\mu}_{\boldsymbol{x}^0}'\boldsymbol{\beta}^0}$.

ii) $\widehat{HIP}_{0,t}^{si}$, $\widehat{HIP}_{0,t}^{di}$, *and* $\widehat{HIP}_{0,t}^{ch}$ *converge in probability toward* $\frac{\boldsymbol{\mu}_{\boldsymbol{x}^t}'\boldsymbol{\beta}^t}{\boldsymbol{\mu}_{\boldsymbol{x}^t}'\boldsymbol{\beta}^0}$.

See the Appendix for a proof of the proposition. Since the three approaches are equivalent under the linear hedonic function assumption, their respective sample price indices converge toward the same parameter under very general assumptions. This is particularly important: In the case of linear hedonic functions the hedonic imputation approach does not worsen the price index problem. Moreover, a simple interpretation of the population parameter estimated by Laspeyres and Paasche indices is possible: By identifying the market quality with the mean vector of the populations' characteristics, price changes correspond to the re-pricing of the market quality in different time periods.

Although theoretically appealing, the asymptotic results demonstrated in Theorem 3.1 rest on the linearity of the hedonic function. In practice, however, linear hedonic functions are restricted to specific cases where an economic model implies the use of such functions. In most cases, the distribution of the error term in the regression model (2.1) is asymmetric and suffers of heteroskedasticity. To address these problems, a transformation of the dependent variable is usually performed, and a linear hedonic model assumed. In particular, the log-linear regression model represents the most used hedonic function in the literature, since it usually leads to normally distributed error terms. Therefore, we

	Laspeyres	Paasche	Fisher
Single	$\frac{\sum_{i=1}^{n_0} h(\hat{f}^t(\mathbf{x}_i^0))}{\sum_{i=1}^{n_0} P_i^0}$	$\frac{\sum_{i=1}^{n_t} P_i^t}{\sum_{i=1}^{n_t} h(\hat{f}^0(\mathbf{x}_i^t))}$	$\sqrt{\left(\frac{\sum_{i=1}^{n_0} h(\hat{f}^t(\mathbf{x}_i^0))}{\sum_{i=1}^{n_0} P_i^0}\right)\left(\frac{\sum_{i=1}^{n_t} P_i^t}{\sum_{i=1}^{n_t} h(\hat{f}^0(\mathbf{x}_i^t))}\right)}$
Double	$\frac{\sum_{i=1}^{n_0} h(\hat{f}^t(\mathbf{x}_i^0))}{\sum_{i=1}^{n_0} h(\hat{f}^0(\mathbf{x}_i^0))}$	$\frac{\sum_{i=1}^{n_t} h(\hat{f}^t(\mathbf{x}_i^t))}{\sum_{i=1}^{n_t} h(\hat{f}^0(\mathbf{x}_i^t))}$	$\sqrt{\left(\frac{\sum_{i=1}^{n_0} h(\hat{f}^t(\mathbf{x}_i^0))}{\sum_{i=1}^{n_0} h(\hat{f}^0(\mathbf{x}_i^0))}\right)\left(\frac{\sum_{i=1}^{n_t} h(\hat{f}^t(\mathbf{x}_i^t))}{\sum_{i=1}^{n_t} h(\hat{f}^0(\mathbf{x}_i^t))}\right)}$
Char	$\frac{h(\hat{f}^t(\overline{\mathbf{x}}^0))}{h(\hat{f}^0(\overline{\mathbf{x}}^0))}$	$\frac{h(\hat{f}^t(\overline{\mathbf{x}}^t))}{h(\hat{f}^0(\overline{\mathbf{x}}^t))}$	$\sqrt{\left(\frac{h(\hat{f}^t(\overline{\mathbf{x}}^0))}{h(\hat{f}^0(\overline{\mathbf{x}}^0))}\right)\left(\frac{h(\hat{f}^t(\overline{\mathbf{x}}^t))}{h(\hat{f}^0(\overline{\mathbf{x}}^t))}\right)}$

Table 2: Hedonic price indices with linearizable hedonic functions

further investigate the asymptotic convergence of imputed hedonic price indices in the case of linearizable hedonic functions. We consider the following hedonic model in each time period t:

$$g(P_i^t) = f^t(\mathbf{x}_i^s) + \epsilon_i^t = \mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t,\ \ i = 1, ..., n_t, \tag{3.2}$$

where g is a transformation of the dependent variable that is a priori known. For sake of simplicity we assume the same transformation g in all time periods. We assume that the transformed model satisfies the usual assumptions of the linear regression model. The error terms ϵ_i^t, in particular, are assumed to be homoskedastic.

Let $h = g^{-1}$ denote the inverse transformation. Table 2 shows Laspeyres, Paasche, and Fisher imputed hedonic indices in the case of linearizable hedonic functions. The formulas contained in Table 2 use biased predictions in the original scale. As we will see in Subsection 3.2, obtaining asymptotic results for the formulas of Table 2 when making unbiased predictions is straightforward in the standard case of log-linear models with normally distributed errors. However, without distributional assumptions on the error terms computations are considerably more complicate and are beyond the aim of the present paper. Moreover, many of the hedonic price indices published by statistics agencies use naive (i.e. biased) predictions when computing hedonic price indices in the original scale. The results of the present section may thus be useful to interpret hedonic price indices actually employed by statistics agencies and as a starting point for the asymptotic analysis with unbiased predictions.

Theorem 3.2. *Let $(P_i^t, \boldsymbol{x}_i^t)$, $i = 1, ..., n_t$ be a random sample of n_t independent random variables belonging to period t $(t = 1, ..., T)$. We assume that the characteristics' vector $\boldsymbol{x}_i^t$ are i.i.d. with $\boldsymbol{\mu}_{\boldsymbol{x}^t} = \mathbb{E}(\boldsymbol{x}^t) < +\infty\ \forall t$. If the usual hypotheses of the linear hedonic model (3.2) hold in each time period and $h \in C^\infty$, then* *i) If the l-th central moments $\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}$ and $\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^t}$ exist $\forall l$, then*

$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIL}^{si}_{0,t} = \frac{\sum_{l=0}^{\infty} \frac{h^{(l)}(\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^t)}{l!} \mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^t}}{\sum_{l=0}^{\infty} \frac{h^{(l)}(\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^0)}{l!} \mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}$$

ii) $$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIL}^{ch}_{0,t} = \frac{h(\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^t)}{h(\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^0)}.$$

iii) If the l-th central moments $\mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}$ and $\mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^0}$ exist $\forall l$, then

$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIP}^{si}_{0,t} = \frac{\sum_{l=0}^{\infty} \frac{h^{(l)}(\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^t)}{l!} \mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}}{\sum_{l=0}^{\infty} \frac{h^{(l)}(\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^0)}{l!} \mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^0}}$$

iv) $$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIP}^{ch}_{0,t} = \frac{h(\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^t)}{h(\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^0)},$$

where $h^{(l)}$ denotes the l-th derivative of h.

See the Appendix for a proof of the theorem for Laspeyres indices. The proof for Paashe indices is similar.

Theorems 3.2 is a generalization of theorem 3.1: When the identity transformation $g(x) = x$ if performed, we obtain the same population index formulas as in the linear case. When the transformation is different from the identity, however, the linearization of the hedonic model comes at a price: In contrast to the results obtained for linear hedonic functions, the different imputation approaches are neither equals nor asymptotically equivalents. Moreover, additional assumptions on the central moments of the transformed variable are needed to guarantee the convergence in probability of single imputed indices. Interestingly, characteristic hedonic price indices don't need additional hypothesis to converge in probability under model (3.2).

In the case of characteristic imputation, the quality is still identified with the mean vector of the characteristic in a given time period, and indices estimates the ratio of the repricing (in the original scale) of such a vector in another time period. On the other hand, single imputed hedonic indices also take into account higher moments to define the quality.

Attentive readers might have noted that Theorem 3.2 does not state any result concerning double imputed hedonic price indices. In fact, asymptotic convergence of double imputed indices could not have been established in the case of linearizable hedonic functions due to the predictions' dependence. Let us consider Laspeyres indices. Following an approach similar to the proof of Theorem 3.2, we have

$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIL}^{di}_{0,t} = \frac{\operatorname{plim}_{n_0\to+\infty} \frac{1}{n_0} \sum_{i=1}^{n_0} h(\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t)}{\operatorname{plim}_{n_0\to+\infty} \frac{1}{n_0} \sum_{i=1}^{n_0} h(\mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^0)}.$$

Unfortunately, whereas the convergence of the numerator can easily be assessed using the weak law of large numbers, the denominator's convergence is more difficult to determine. This is primarily due to the stochastic dependence of the $\mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^0$. Therefore, usual weak laws of large numbers cannot be used. Moreover, as demonstrated in Property 3 in the Appendix, the sufficient conditions for convergence in probability implied by Chebychev's inequality are not satisfied: The variance of the denominator will in general not tend to zero (results obtained in Proposition 3 thus show that even a second order approximation similar to the one performed in the proof of Theorem 3.2 is not possible). This seems to be of primary importance if double imputed hedonic indices are used as indicators of the general price level in econometric analyses of the market: Using a random variable that might not estimate a theoretic price change for a fixed quality may lead to erroneous conclusions.

3.2. Imputed hedonic indices and log-linear hedonic models

In the hedonic literature, the log-linear model has established itself as the reference model. We thus consider the regression model

$$\ln(P_i^t) = \mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t,\ \ i = 1, ..., n_t. \tag{3.3}$$

Since prices are usually skewed to the right, a log transformation usually leads to symmetrically distributed prices. The following Corollary shows the relation between single imputed and characteristic price indices when log-linear hedonic function are assumed:

Corollary 3.1. *Let $(P_i^t, \boldsymbol{x}_i^t),\ i = 1, ..., n_t$ be a random sample of n_t independent random variables belonging to period t $(t = 1, ..., T)$. We assume that the characteristics' vector $\boldsymbol{x}_i^t$ are i.i.d. with $\boldsymbol{\mu}_{\boldsymbol{x}^t} = \mathbb{E}(\boldsymbol{x}^t) < +\infty\ \forall t$. If the usual hypotheses of the linear hedonic model (3.3) hold in each time period, then*

i) $HIL_{0,t}^{si} = LR_{0,t} HIL_{0,t}^{ch}$ *with* $LR_{0,t} = \frac{\sum_{l=0}^{\infty} \frac{\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^t}}{l!}}{\sum_{l=0}^{\infty} \frac{\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!}}$.

ii) $HIP_{0,t}^{si} = PR_{0,t} HIP_{0,t}^{ch}$ *with* $PR_{0,t} = \frac{\sum_{l=0}^{\infty} \frac{\mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}}{l!}}{\sum_{l=0}^{\infty} \frac{\mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^0}}{l!}}$,

where $LR_{0,t}$ and $PR_{0,t}$ are defined as the Laspeyres and Paashe ratio, respectively.

Single imputed Laspeyres and Paasche hedonic indices thus converge toward a characteristic hedonic index of the same type times a factor under model (3.3). This factor may be interpreted as price variation measured by single imputed indices net of the price variation accounted by characteristic indices. To better understand its role, formulas of Corollary 3.1 with second order terms only (see Property 2 in the Appendix for the computation of the second central moment):

$$LR_{0,t} \approx \frac{2 + \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^0}}\boldsymbol{\beta}^t}{2 + \boldsymbol{\beta}^{0\prime}\Sigma_{\mathbf{x^0}}\boldsymbol{\beta}^0 + \sigma^2_{\epsilon^0}} \quad \text{and} \quad PR_{0,t} \approx \frac{2 + \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \sigma^2_{\epsilon^t}}{2 + \boldsymbol{\beta}^{0\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^0}.$$

The matrices $\Sigma_{\mathbf{x^0}}$ and $\Sigma_{\mathbf{x^t}}$ are positive-semidefinite, and the variances $\sigma^2_{\epsilon^0}$ and $\sigma^2_{\epsilon^t}$ are positive. Let us consider a simple case of figure in which only physical characteristic of the goods have been considered. In this case, characteristics' shadow prices $\boldsymbol{\beta}^0$ and $\boldsymbol{\beta}^t$ are expected to be positive in all time periods. Moreover, due both to production constraints and consumer preferences, we usually observe strongly positively correlated physical characteristics, such that the off-diagonal elements of the matrices $\Sigma_{\mathbf{x^0}}$ and $\Sigma_{\mathbf{x^t}}$ are all greater than zero. Therefore, factors $LR_{0,t}$ and $PR_{0,t}$ are expected to be positive in all time periods.

If shadow prices of one of more characteristics decrease $\boldsymbol{\beta}^t = \boldsymbol{\beta}^0 + \mathbf{c}$ with $\mathbf{c} = (c_1, ..., c_K),\ c_j \leq 0,\ j = 1, ..., K$, the multiplying factor for Laspeyres indices will be smaller than 1. In this case of figure, Laspeyres single imputed indices amplify the price drop as measured by Laspeyres characteristic indices ($LR_{0,t} < 1$). On the other hand, if shadow prices tend to be higher in the period under review $\boldsymbol{\beta}^t = \boldsymbol{\beta}^0 + \mathbf{c}$ with $\mathbf{c} = (c_1, ..., c_K),\ c_j \geq 0,\ j = 1, ..., K$, the Paashe factor amplifies the price increase as measured by Paashe characteristic indices ($PR_{0,t} > 1$). In general, if shadow prices $\boldsymbol{\beta}^t,\ t = 1, ..., T$ are approximatively constant through time, we expect to have $LR_{0,t} < 1$ and $PR_{0,t} > 1$ in every time period.

This systematic relationship is particularly relevant when cross-country comparisons of the general price level are effectuated using different quality-adjusted price indices (for example, Laspeyres single imputed and characteristic). For example, a country whose prices are measured with a single imputed approach might display a greater volatility with respect to the country whose price are measured with the characteristic approach (for example, see the difference already present in the same market for Laspeyres and Paashe single imputed and characteristic indices in the left side of Figure 2). This greater volatility, however, is only due to a different parameter being estimated, and not to any structural difference between the two markets.

The following Lemmas show that single imputed indices are finite when transformed prices and shadow prices follow a normal distribution.

Lemma 3.1. *Let us consider a log-linear hedonic function with inverse transformation $h(x) = e^x$. If both the transformed prices $\boldsymbol{x}_i^{0\prime}\boldsymbol{\beta}^t + \epsilon_i^0$ and the shadow prices $\boldsymbol{x}_i^{0\prime}\boldsymbol{\beta}^t$ (resp. $\boldsymbol{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t$ and $\boldsymbol{x}_i^{t\prime}\boldsymbol{\beta}^0$) are normally distributed, then Laspeyres (resp. Paashe) single imputed hedonic price index are finite.*

See the Appendix for a proof of the lemma in the case of Laspeyres indices. We now turn to the convergence of imputed price indices when unbiased prediction in the original scale are performed. Since the hedonic model is log-linear and we assume normal distributed errors, we simply add the usual bias correction term $\frac{1}{2}(\hat{\sigma}^t)^2$ to the predictions of the formulas contained in Table 2 (see Hill (2013)) before retransforming in the original scale. It is worth noting that to the author's knowledge no study has been conducted to evaluate the magnitude of this bias correction on imputed hedonic price indices.

Lemma 3.2. *Let $(P_i^t, \boldsymbol{x}_i^t)$, $i = 1, ..., n_t$ be a random sample of n_t independent random variables belonging to period t $(t = 1, ..., T)$. We assume that the characteristics' vector $\boldsymbol{x}_i^t$ are i.i.d. with $\boldsymbol{\mu}_{\boldsymbol{x}^t} = \mathbb{E}(\boldsymbol{x}^t) < +\infty$ $\forall t$. If the usual hypotheses of the linear hedonic model (3.3) hold in each time period and $\boldsymbol{\epsilon}^t \sim N(0, (\sigma^t)^2)$ $\forall t$, then*

i) If the l-th central moments $\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}$ and $\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^t}$ exist $\forall l$, then

$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIL}^{*si}_{0,t} = \frac{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^t+\frac{1}{2}(\sigma^t)^2}\sum_{l=0}^{\infty}\frac{\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^t}}{l!}}{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^0}\sum_{l=0}^{\infty}\frac{\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!}}.$$

ii)
$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIL}^{*ch}_{0,t} = \frac{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^t+\frac{1}{2}(\sigma^t)^2}}{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^0}\boldsymbol{\beta}^0+\frac{1}{2}(\sigma^0)^2}},$$

iii) If the l-th central moments $\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}$ and $\mu^l_{\boldsymbol{x}^{0\prime}\boldsymbol{\beta}^t}$ exist $\forall l$, then

$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIP}^{*si}_{0,t} = \frac{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^t}\sum_{l=0}^{\infty}\frac{\mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}}{l!}}{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^0+\frac{1}{2}(\sigma^0)^2}\sum_{l=0}^{\infty}\frac{\mu^l_{\boldsymbol{x}^{t\prime}\boldsymbol{\beta}^0}}{l!}}.$$

iv)
$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIP}^{*ch}_{0,t} = \frac{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^t+\frac{1}{2}(\sigma^t)^2}}{e^{\boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^0+\frac{1}{2}(\sigma^0)^2}},$$

where $\widehat{HIL}^{*si}_{0,t}$, $\widehat{HIL}^{*ch}_{0,t}$, $\widehat{HIP}^{*si}_{0,t}$ and $\widehat{HIP}^{*ch}_{0,t}$ denote the sample imputed hedonic indices based on unbiased prediction in the original scale.

See the Appendix for a proof of the lemma in the case of Laspeyres indices.
As shown in the above formula, we expect a much larger impact of bias correction on single imputed indices than on characteristic indices. In fact, if the volatility σ^t remains constant through time, bias correction does not affect population characteristic indices. We also have the counterpart of Corollary 3.1:

Lemma 3.3. *Let* $(P_i^t, \boldsymbol{x}_i^t)$, $i = 1, ..., n_t$ *be a random sample of* n_t *independent random variables belonging to period* t $(t = 1, ..., T)$. *We assume that the characteristics' vector* $\boldsymbol{x}_i^t$ *are i.i.d. with* $\boldsymbol{\mu}_{\boldsymbol{x}^t} = \mathbb{E}(\boldsymbol{x}^t) < +\infty \; \forall t$. *If the usual hypotheses of the linear hedonic model in (3.3) hold in each time period of the linear hedonic model in (3.3), then*

i) $HIL^{*si}_{0,t} = LR^*_{0,t} HIL^{*ch}_{0,t}$ *with* $LR^*_{0,t} = e^{\frac{1}{2}(\sigma^0)^2} \frac{\sum_{l=0}^{\infty} \frac{\mu^l_{(\boldsymbol{x}^0)'\boldsymbol{\beta}^t}}{l!}}{\sum_{l=0}^{\infty} \frac{\mu^l_{(\boldsymbol{x}^0)'\boldsymbol{\beta}^0+\epsilon^0}}{l!}}$.

ii) $HIP^{*si}_{0,t} = PR^*_{0,t} HIP^{*ch}_{0,t}$ *with* $PR^*_{0,t} = e^{-\frac{1}{2}(\sigma^t)^2} \frac{\sum_{l=0}^{\infty} \frac{\mu^l_{(\boldsymbol{x}^t)'\boldsymbol{\beta}^t+\epsilon^t}}{l!}}{\sum_{l=0}^{\infty} \frac{\mu^l_{(\boldsymbol{x}^t)'\boldsymbol{\beta}^0}}{l!}}$.

We close this section with two remarks. First, as sample indices' formulas Table 1 and 2 illustrate, there is not only a strong non-linearity in the sample indices used to estimate population indices, but also an apparent stochastic dependence between the numerator and the denominator in their formulas. For these reasons, it seems unrealistic to derive the asymptotic distribution of such indices with standard approaches even in the case of simple linear hedonic functions, therefore suggesting the use of resampling methods to determine their distribution. See Brachinger et al. (2012) for the construction of confidence intervals of elementary hedonic price indices.

Second, the convergence results obtained throughout this and the previous sections rely on the usual hypothesis of the linear regression model. However, as long as the estimation technique implies a convergence of the shadow prices $\hat{\beta}_0^t, ..., \hat{\beta}_K^t$ toward the $\beta_0^t, ..., \beta_K^t$ defining the theoretical linear data generating process, the convergence results remain valid. This is particularly importance, since it allows the use of other regression models/ regression techniques than the usual one. Ridge regression, for example, could be used to estimate penalized shadow prices in each time period, and subsequently compute sample hedonic indices that converge according to the formulae established in the previous sections.

3.3. Convergence in probability of composite indices

The convergence in probability of the Laspeyres and Paasche hedonic price indices can then be used to establish the convergence in probability of the Fisher indices.

Corollary 1. *The Fisher hedonic price indices* $\widehat{HIF}^{si}_{0,t}$, $\widehat{HIF}^{di}_{0,t}$, *and* $\widehat{HIF}^{ch}_{0,t}$ *converge in probability toward the geometric average of population Laspeyres and Paashe hedonic price indices.*

Based on the Fisher formulae contained in Table 1 and 2, the proof trivially follows from the continuous mapping theorem. Importantly, Corollary 1 can easily be generalized to show that convergence results of elementary (i.e. unweighted) price indices can easily be exploited to derive asymptotic results of composite ones. As long as the composite price index formula satisfies the smoothness condition, the continuous mapping theorem allows us to use the results obtained for elementary price indices to derive those of the composite ones.

An interesting case is represented by composite market indices that are given as a linear combination of elementary sub-indices. Let us consider, for example, a composite hedonic index $HC_{0,t}$ aiming to describe price changes of a market possessing S distinct segments. The usual approach is to compute the sample $\widehat{HC}_{0,t}$ as a convex combination of segment (elementary) price indices $\widehat{HI}_{0,t}$:

$$\widehat{HC}_{0,t} = c^1 \widehat{HI}_{0,t}{}^1 + ... + c^S \widehat{HI}_{0,t}{}^S, \; \sum_{i=1}^{S} c^i = 1, \; c^i \geq 0 \; \forall i,$$

where $\widehat{HI}_{0,t}$ denote either hedonic Laspeyres, Paashe, or Fisher price indices computed using a certain imputation method. Since we have demonstrated the convergence in probability of these indices, we can use the convergence preservation to obtain

$$\underset{n_0,n_t \to +\infty}{\text{plim}} \widehat{HC}_{0,t} = c^1 HI_{0,t}{}^1 + ... + c^S HI_{0,t}{}^S.$$

This results remains valid even if instead of predetermined weights $c^j, \; j = 1, ..., S$ we consider estimated relative expenditure weight $\hat{c}_j = \sum_{i=1}^{n_s} P_i^j / \sum_{i=1}^{n} P_i$, where P_i^s denotes prices observed in segment j. In this case, we simply have to consider the convergence in probability of weights in the computations.

4. Simulation study

In this section we perform a simulation study to further investigate the convergence properties established in Section 3. The simulated data will serve two purposes. The first purpose is the computation of Laspeyres, Paashe, and Fisher population hedonic indices for single imputed and characteristic approaches. The second purpose is to investigate the empirical convergence in probability of sample indices toward the population index they estimate. In particular, we analyse the impact of bias corrected predictions on indices ratios.

The simulated data are based on the hedonic housing data of the city of Ames (Iowa) presented by De Cock (2011), which are freely available on the data archives of the Journal of Statistics Education.

As explained in the data description file, the data set includes information from the Ames Assessor's Office used in computing assessed values for individual residential properties sold in Ames from 2006 to 2010 (monthly). For sake of simplicity, only 5 out of the 80 variables contained in the data set have been considered in the present simulation: Sale price (P, in USD), lot area (Lot, in square feet), total basement surface (Bas, in square feet), above ground living area (Liv, in square feet), size of garage (Gar, in square feet). Moreover, to base our simulation on more realistic estimates, the data has been grouped in 18 quarters, from the first quarter 2006 to the second quarter 2010.

4.1. Simulated data

In this section we describe the procedure employed to simulate the log-linear regression model presented in (3.3). The main objective is to simulate house prices and characteristics that are realistic according to the observed housing data of the city of Ames. The following procedure has thus been adopted in each quarter $t = 1, ..., 18$:

1. Estimation of the log-linear hedonic model using the observed data contained in the Ames data set:

$$\log(P_i^t) = \gamma_0^t + \gamma_1^t \text{Lot}_i^t + \gamma_2^t \text{Bas}_i^t + \gamma_3^t \text{Liv}_i^t + \gamma_4^t \text{Gar}_i^t + \epsilon_i^t,\ \ i = 1, ..., n_t.$$

   In particular, we estimate the variance $\sigma^2_{\epsilon^t}$ of the error term.

2. Simulation of 3000 vectors of characteristic $\boldsymbol{x}^{t*}$ using a truncated multivariate normal distribution: $\boldsymbol{x}^{t*} \sim N(\boldsymbol{\mu}_{\boldsymbol{x}^{t*}}, \Sigma_{\boldsymbol{x}^{t*}} \mid [\boldsymbol{x}^{t*}_{min}, \boldsymbol{x}^{t*}_{max}])$, where $\boldsymbol{\mu}_{\boldsymbol{x}^{t*}}$ and $\Sigma_{\boldsymbol{x}^{t*}}$ have been set equal to robust estimates of the characteristics' empirical mean and covariance matrix, respectively. The lower and upper bounds $\boldsymbol{x}^{t*}_{min}$ and $\boldsymbol{x}^{t*}_{max}$ of the density have been chosen equal to the minimum and maximum values observed in each time period.

3. Simulation of log-prices $\log(\boldsymbol{p}^t)^*$ are obtained by predicting the values of the simulated characteristics according to the hedonic models estimated at point 1 and adding a simulated random error:

$$\log(P_i^t)^* = \hat{\gamma}_0^t + \hat{\gamma}_1^t x_{i1}^{t*} + \hat{\gamma}_2^t x_{i2}^{t*} + \hat{\gamma}_3^t x_{i3}^{t*} + \hat{\gamma}_4^t x_{i4}^{t*} + \epsilon_i^{t*},\ \ i = 1, ..., 3000,$$

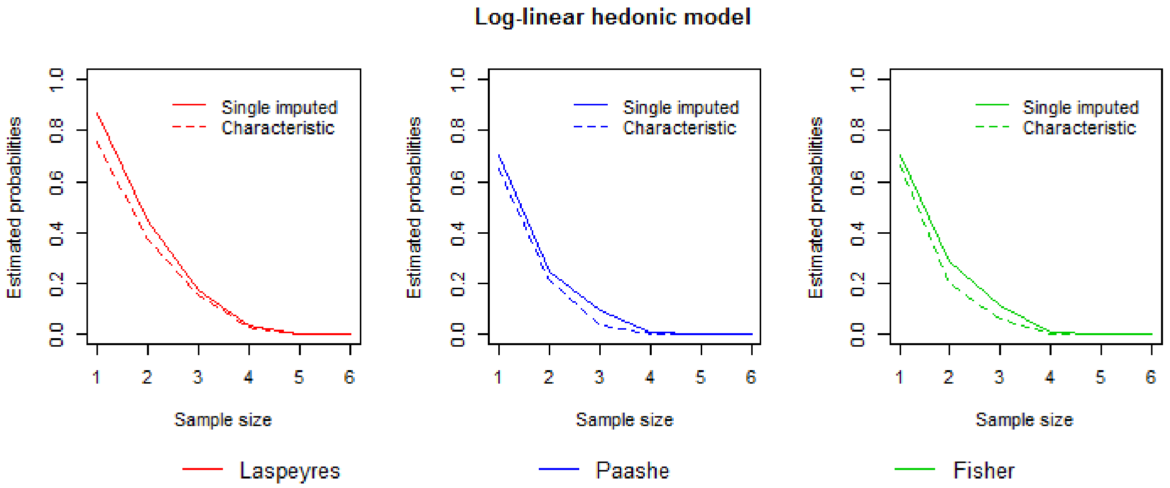

Figure 1: Convergence in probability of imputed indices

where ϵ_i^{t*} follows a normal distribution with mean equal to zero and variance $\sigma^2_{\epsilon^{t*}} = \hat{\sigma}^2_{\epsilon^t}$.

The use of a truncated normal distribution based on robust estimates of the mean vector and covariance matrix, avoids to create vectors with implausible characteristic values and/or characteristic combinations. Importantly, the estimates $\hat{\sigma}^2_{\epsilon^1}, \ldots, \hat{\sigma}^2_{\epsilon^{18}}$ allow us to simulate error terms that take into account volatility changes over time. Therefore, although simulated, the above procedure provides plausible data on which apply the results of the theoretical section.

4.2. Convergence in probability and population indices

Throughout this section, the base period has been set equal to the first quarter $t = 1$. To empirically analyse the convergence in probability of imputed hedonic indices, 300 samples have been drawn for a given sample size. The sample size n has then been progressively increased from 500 to 3000 observations by steps of 500 units. Using the same approach illustrated by Lafaye de Micheaux and Liquet (2009), we estimate the probabilities $\text{Pb}_n = P(\| \widehat{\boldsymbol{HI}}_n - \boldsymbol{HI} \| > c)$, $c > 0$, where $\widehat{\boldsymbol{HI}}_n = (\widehat{HI}^1_n, \ldots, \widehat{HI}^T_n)$ and $\boldsymbol{HI} = (HI^1, \ldots, HI^T)$ denote the vectors of sample and population hedonic indices for the $T = 18$ quarters, respectively.

Figure 1 shows estimated probabilities Pb_n $(n = 500k,\ k = 1, \ldots, 6)$ of Laspeyres, Paashe, and Fisher hedonic price indices for $c = 5\%$. The chosen threshold is extremely low, since on average a sample price index must not be distant more than $0.05/18 = 0.28\%$ in a given quarter. For each index, single imputed and characteristic approaches are considered. All

Quarter	$LR_{1,t}$	$LR^*_{1,t}$	$PR_{1,t}$	$PR^*_{1,t}$	$FR_{1,t}$	$FR^*_{1,t}$
1	0.97	1.00	1.03	1.01	1.00	1.00
2	0.96	0.99	1.01	0.99	0.99	0.99
3	0.96	0.99	1.01	0.99	0.99	0.99
4	0.96	0.99	1.01	0.98	0.99	0.98
5	0.96	0.99	1.00	0.98	0.98	0.99
6	0.95	0.98	1.00	0.99	0.98	0.98
7	0.96	0.99	1.02	0.99	0.99	0.99
8	0.96	0.98	1.02	0.99	0.99	0.99
9	0.99	1.01	1.05	1.01	1.02	1.01
10	0.96	0.98	1.01	0.99	0.98	0.99
11	0.97	0.99	1.01	1.00	0.99	1.00
12	0.96	0.98	1.01	0.99	0.98	0.99
13	0.98	1.01	1.02	1.01	1.00	1.01
14	0.96	0.99	1.01	0.98	0.98	0.99
15	0.96	0.98	1.01	0.99	0.99	0.99
16	0.97	1.00	1.02	0.99	1.00	1.00
17	0.96	0.99	1.01	0.99	0.99	0.99
18	0.97	1.00	1.03	1.00	1.00	1.00

Table 3: Hedonic imputed indices ratios

the indices display roughly similar convergence rates toward their population parameter, although Laspeyres indices seem to perform slightly worse than Paashe indices. Interestingly, Laspeyres indices showed a much slower convergence rate with respect to Paashe indices for some simulated data sets. This is probably due to a higher sensitivity of Laspeyres indices to extreme values. This issue, however, is not treated in the present paper.

Population indices of Laspeyres, Paashe, and Fisher hedonic price indices are shown in Figure 2. The left and right side of the Figure contain graphics of population indices resulting from biased and unbiased predicted prices in the original scale, respectively. Each graphic illustrates a specific population index for single imputation and characteristic approaches. Let us consider population indices based on biased predictions first.

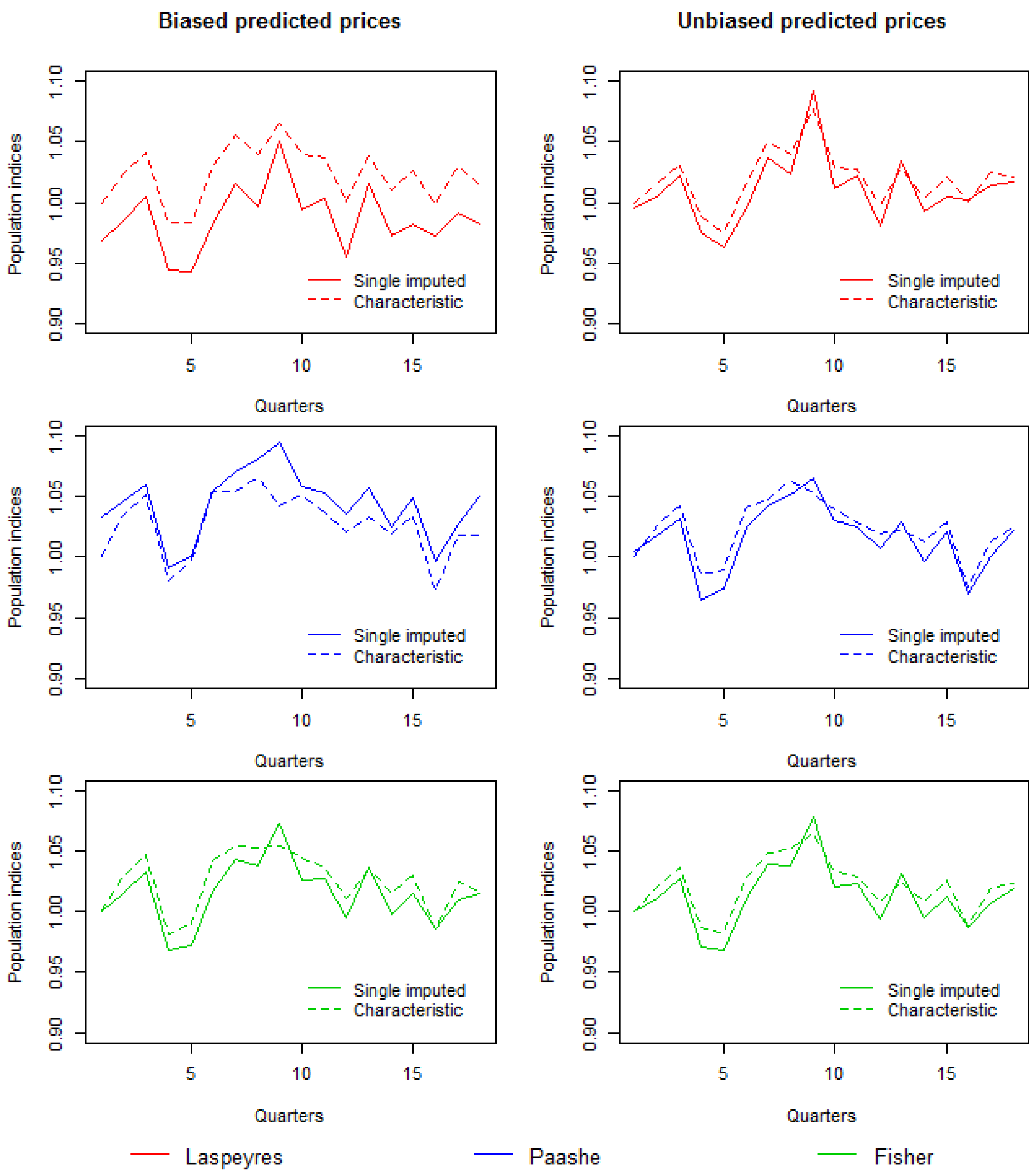

Figure 2: Population indices

As expected, Laspeyres ratios are all smaller than 1, whereas Paashe ratios are always greater than 1 (see Table 3). Fisher indices show approximatively the same value for both the single imputed and the characteristic approach. Nevertheless, characteristic indices are in general smoother than their single imputed counterpart, making the identification of potential cycles/trends difficult. This potential drawback, however, seems not to be present when unbiased predictions in the original scale are performed. In fact, in this case the bias adjustment lowers the distance between single imputed and characteristic indices, making the Laspeyres and Paashe ratios almost equal to 1 (see Table 3). Interestingly, the bias adjustment seems not to impact Fisher price indices, which posses ratios almost equal to 1 even before the bias adjustment. We thus reach the following important conclusion. Bias adjustment is of utmost important not only from a micro-econometric point of view, but also from a macro-econometric perspective: The price index problem caused by different imputation approaches seem to vanish when unbiased predicted prices are used.

5. Conclusions

Several important theoretical results have been obtained in the present paper. First, the asymptotic convergence of single imputed, double imputed, and characteristics hedonic price indices has been established in the case of goods possessing a linear hedonic function. In this case the price index problem is not worsened by imputation methods, alleviating an uncomfortable situation price statisticians have to face. Convergence results hold under mild assumptions, mainly dictated by the econometric models used to estimate the hedonic functions in different time periods.

Second, the asymptotic convergence of hedonic indices with linearizable hedonic functions has been established. The parameters estimated by single-imputed and characteristic hedonic price indices have been identified, and the double-imputed hedonic approach was found inappropriate to measure constant-quality price changes due to its possible lack of convergence in probability. The obtained results show how, in general, the functional form of the econometric model used to estimate the hedonic functions affects the parameter estimated by hedonic price indices. Depending on the hedonic approach, adopting a non-linear functional form modifies the type of quality adjustment: Single imputed hedonic indices also take the covariance structure of the regressors into account, whereas characteristic hedonic indices mainly identify the quality of the characteristic with their mean vector.

Third, under the assumption of log-linear hedonic functions, we established the analytical relationship between single-imputed and characteristic hedonic indices, highlighting the potential pitfalls this relationship might cause in econometric analyses. Finally, explicit formulas taking into account bias adjustment of predicted prices for single and characteristic imputed indices were given in the case of log-linear hedonic functions with normal distributed errors.

These theoretical findings are complemented with a simulation study. In particular, the convergence in probability of hedonic indices in the case of a log-linear hedonic model with normal distributed errors has been empirically analysed. Two main results are obtained with our simulation study. First, the convergence speed has been found similar for all the indices, although Laspeyres indices seem to be more affected by extreme values than Paashe indices. Second, adopting a bias correction for predicted prices reduces the distance between population indices estimated by single imputed and characteristic hedonic indices, virtually eliminating the price index problem.

References

Brachinger, H. W., Beer, M., Schöni, O., 2012. The Econometric Foundations of Hedonic Elementary Price Indices. Working paper, Department of Quantitative Economics, University of Fribourg.

DasGupta, A., 2011. Probability for Statistics and Machine Learning. Springer Texts in Statistics.

De Cock, D., 2011. Ames, Iowa: Alternative to the boston housing data as an end of semester regression projec. Journal of Statistics Education 19 (3).

Diewert, W. E., 2011. Alternative approaches to measuring house price inflation. Economics working papers, Vancouver School of Economics.

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

Eurostat, 2012. Detailed technical manual on owner-occupied housing. Tech. rep., Eurostat.

Greene, W. H., 2011. Econometric Analysis - 7 Edition. Pearson Education.

Hill, R., 2011. Hedonic Price Indexes for Housing. Oecd statistics working papers, OECD Publishing.

Hill, R. J., 2013. Hedonic price indexes for residential housing: A survey, evaluation and taxonomy. Journal of Economic Surveys 27 (5), 879–914.

Hill, R. J., Melser, D., 2008. Hedonic Imputation and the Price Index Problem: an Application To Housing. Economic Inquiry 46 (4), 593–609.

Kagie, M., Wezel, M. V., 2007. Hedonic price models and indices based on boosting applied to the dutch housing market. Intelligent Systems in Accounting, Finance & Management 15 (3-4), 85–106.

Lafaye de Micheaux, P., Liquet, B., 2009. Understanding convergence concepts: A visual-minded and graphical simulation-based approach. The American Statistician 63 (2), 173–178.

Mark, J. H., Goldberg, M. A., 1984. Alternative housing price indices: An evaluation. Real Estate Economics 12 (1), 30–49.

Meese, R., Wallace, N., 1991. Nonparametric Estimation of Dynamic Hedonic Price Models and the Construction of Residential Housing Price Indices. Real Estate Economics 19 (3), 308–332.

OECD, Eurostat, 2013. Handbook on Residential Property Price Indices. OECD Library. OECD, Eurostat, International Labour Office and International Monetary Fund, and The United Nations Economic Commission for Europe and The World Bank.

Rosen, S., 1974. Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. Journal of Political Economy 82 (1), 34–55.

Schöni, O., 2013. Four essays on statistical problems of hedonic methods: An application to single- family houses. Ph.D. thesis, Faculty of Economics and Social Sciences, University of Fribourg.

von de Haan, J., 2010. Hedonic Price Indexes: A Comparison of Imputation, Time Dummy and Re-Pricing Methods. Journal of Economics and Statistics (Jahrbuecher fuer Nationaloekonomie und Statistik) 230 (6), 772–791.

Wallace, N., Meese, R., 1997. The construction of residential housing price indices: A comparison of repeat-sales, hedonic-regression, and hybrid approaches. The Journal of Real Estate Finance and Economics 14 (1-2), 51–73.

Wallace, N. E., 1996. Hedonic-based price indexes for housing: theory, estimation, and index construction. Economic Review, 34–48.

Appendix A. Asymptotic properties of hedonic models

The following properties hold under the classical linear model hypothesis and the hypothesis assumed in Proposition 3.1 and 3.2. The employed terminology is borrowed from DasGupta (2011).

Proof of Theorem 3.1. We start by demonstrating that the three approaches are identical when linear hedonic functions are considered, i.e. $\widehat{HIL}^{si}_{0,t} = \widehat{HIL}^{di}_{0,t} = \widehat{HIL}^{ch}_{0,t}$, and $\widehat{HIP}^{si}_{0,t} = \widehat{HIP}^{di}_{0,t} = \widehat{HIP}^{ch}_{0,t}$. This follows trivially from the fact that the average vector of characteristics always belong to the regression line:

$$\widehat{HIL}^{si}_{0,t} = \frac{\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t}{\sum_{i=1}^{n_0} P_i^0} = \frac{\frac{1}{n_0}\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t}{\frac{1}{n_0}\sum_{i=1}^{n_0} P_i^0} = \frac{\overline{\mathbf{x}}^{0\prime}\hat{\boldsymbol{\beta}}^t}{\overline{\mathbf{x}}^{0\prime}\hat{\boldsymbol{\beta}}^0} = \widehat{HIL}^{ch}_{0,t} = \frac{\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t}{\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^0} = \widehat{HIL}^{di}_{0,t}.$$

and

$$\widehat{HIP}^{si}_{0,t} = \frac{\sum_{i=1}^{n_t} P_i^t}{\sum_{i=1}^{n_t} \mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^0} = \frac{\frac{1}{n_t}\sum_{i=1}^{n_t} P_i^t}{\frac{1}{n_t}\sum_{i=1}^{n_0} \mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^0} = \frac{\overline{\mathbf{x}}^{t\prime}\hat{\boldsymbol{\beta}}^t}{\overline{\mathbf{x}}^{t\prime}\hat{\boldsymbol{\beta}}^0} = \widehat{HIP}^{ch}_{0,t} = \frac{\sum_{i=1}^{n_t} \mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t}{\sum_{i=1}^{n_t} \mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^0} = \widehat{HIP}^{di}_{0,t}.$$

Therefore, we demonstrate the convergence in probability for only one type of imputation method. Due to its simplicity, we restrict ourself to the characteristic approach. For Laspeyres indices we simply have

$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIL}^{ch}_{0,t} = \frac{(\text{plim}_{n_0\to+\infty}\overline{\mathbf{x}}^0)'(\text{plim}_{n_t\to+\infty}\hat{\boldsymbol{\beta}}^t)}{(\text{plim}_{n_0\to+\infty}\overline{\mathbf{x}}^0)'(\text{plim}_{n_0\to+\infty}\hat{\boldsymbol{\beta}}^0)} = \frac{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t}{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0}.$$

For Paasche indices we have:

$$\operatorname*{plim}_{n_0,n_t\to+\infty} \widehat{HIP}^{ch}_{0,t} = \frac{(\text{plim}_{n_t\to+\infty}\overline{\mathbf{x}}^t)'(\text{plim}_{n_t\to+\infty}\hat{\boldsymbol{\beta}}^t)}{(\text{plim}_{n_t\to+\infty}\overline{\mathbf{x}}^t)'(\text{plim}_{n_0\to+\infty}\hat{\boldsymbol{\beta}}^0)} = \frac{\boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t}{\boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^0}.$$

□

Property 1. *Let h denote a smooth function. In any time period t and base period 0, we have*

$$\operatorname*{plim}_{n_t \to +\infty} \frac{\sum_{i=1}^{n_0} h(\boldsymbol{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t)}{\sum_{i=1}^{n_0} P_i^0} = \frac{\sum_{i=1}^{n_0} h(\boldsymbol{x}_i^{0\prime}\boldsymbol{\beta}^t)}{\sum_{i=1}^{n_0} P_i^0}.$$

Proof. We first demonstrate that

$$\operatorname*{plim}_{n_t \to +\infty} \frac{\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t}{\sum_{i=1}^{n_0} P_i^0} = \frac{\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t}{\sum_{i=1}^{n_0} P_i^0}.$$

We consider first the convergence in probability of a single term $\mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t$ as $n_t \to +\infty$. The probability distribution of the K-dimensional random variable $(\mathbf{x}_i^0)$ does not depend on n_t. It can therefore be considered as converging in probability toward itself as $n_t \to +\infty$: $\text{plim}_{n_t \to +\infty} \mathbf{x}_i^0 = \mathbf{x}_i^0$. Under the classical hypothesis of the linear regression model estimated in period t, we have that $\text{plim}_{n_t \to +\infty} \hat{\boldsymbol{\beta}}^t = \boldsymbol{\beta}^t$. The multi-dimensional convergence preservation implies that $\text{plim}_{n_t \to +\infty} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t = \mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t$. Using again the convergence preservation, we obtain (since the sum does not depend on n_t)

$$\text{plim}_{n_t \to +\infty} \sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t = \sum_{i=1}^{n_0} \text{plim}_{n_t \to +\infty} \left(\mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t\right) = \sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t.$$

Since the denominator $\sum_{i=1}^{n_0} P_i^0$ does not depend on n_t, we also have that $\text{plim}_{n_t \to +\infty} \sum_{i=1}^{n_0} P_i^0 = \sum_{i=1}^{n_0} P_i^0$. Thus implying

$$\operatorname*{plim}_{n_t \to +\infty} \frac{\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t}{\sum_{i=1}^{n_0} P_i^0} = \frac{\text{plim}_{n_t \to +\infty} \sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t}{\text{plim}_{n_t \to +\infty} \sum_{i=1}^{n_0} P_i^0} = \frac{\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t}{\sum_{i=1}^{n_0} P_i^0}.$$

The continuous mapping theorem allow us to obtain the wanted result. Note: The proposition remains valid if the denominator is replaced with $\sum_{i=1}^{n_0} \mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^0$, since it represents a random variable not depending on n_t. □

Proof of Theorem 3.2. The convergence in probability of $\widehat{HIL}_{0,t}^{si}$ is first established. Using Property 1 we have

$$\operatorname*{plim}_{n_0, n_t \to +\infty} \widehat{HIL}_{0,t}^{si} = \operatorname*{plim}_{n_0, n_t \to +\infty} \frac{\sum_{i=1}^{n_0} h(\mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t)}{\sum_{i=1}^{n_0} P_i^0} = \operatorname*{plim}_{n_0 \to +\infty} \frac{\frac{1}{n_0}\sum_{i=1}^{n_0} h(\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t)}{\frac{1}{n_0}\sum_{i=1}^{n_0} P_i^0} =$$
$$= \frac{\text{plim}_{n_0 \to +\infty} \frac{1}{n_0}\sum_{i=1}^{n_0} h(\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t)}{\text{plim}_{n_0 \to +\infty} \frac{1}{n_0}\sum_{i=1}^{n_0} P_i^0}.$$

Let us first consider the convergence of the numerator. We perform a Taylor series expansion in $\boldsymbol{\mu}_{\mathbf{x}^0}'\boldsymbol{\beta}^t$:

$$\frac{1}{n_0}\sum_{i=1}^{n_0} h(\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t) = \frac{1}{n_0}\sum_{i=1}^{n_0}\sum_{l=0}^{\infty} \frac{h^{(l)}(\boldsymbol{\mu}_{\mathbf{x}^0}'\boldsymbol{\beta}^t)}{l!}(\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t - \boldsymbol{\mu}_{\mathbf{x}^0}'\boldsymbol{\beta}^t)^l$$

Since the $\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t - \boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t$ are independent and identically distributed random variables and we assumed that their central moments of order l existed $\forall l$, we have that (Khinchine's weak law of large numbers)

$$\operatorname*{plim}_{n_0\to+\infty}\frac{1}{n_0}\sum_{i=1}^{n_0} h(\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t) = \sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t)}{l!}\operatorname*{plim}_{n_0\to+\infty}\frac{1}{n_0}\sum_{i=1}^{n_0}(\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t - \boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t)^l =$$
$$= \sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t)}{l!}\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}.$$

For the denominator we perform a Taylor series expansion in $\boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t$:

$$\frac{1}{n_0}\sum_{i=1}^{n_0} P_i^0 = \frac{1}{n_0}\sum_{i=1}^{n_0} h(\mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t) = \frac{1}{n_0}\sum_{i=1}^{n_0}\sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t)}{l!}(\mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t - \boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t)^l$$

Since the $\mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t - \boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t$ are independent and identically distributed random variables and we assumed that their central moments of order l existed $\forall l$, we have that (Khinchine's weak law of large numbers)

$$\operatorname*{plim}_{n_0\to+\infty}\frac{1}{n_0}\sum_{i=1}^{n_0} P_i^0 = \sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t)}{l!}\operatorname*{plim}_{n_t\to+\infty}\frac{1}{n_t}\sum_{i=1}^{n_t}((\mathbf{x}_i^t)'\boldsymbol{\beta}^t + \epsilon_i^t - \boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t)^l =$$
$$= \sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t)}{l!}\mu^l_{\mathbf{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}.$$

Therefore,

$$\operatorname*{plim}_{n_0,n_t\to+\infty}\widehat{HIL}^{si}_{0,t} = \frac{\sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t)}{l!}\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{\sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0)}{l!}\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}.$$

Note: Although the above Taylor series expansion are not strictly necessary to establish a convergence result, they are useful to investigate the the relationship between single imputed and characteristic population indices.

For $\widehat{HIL}^{ch}_{0,t}$, we simply have

$$\operatorname*{plim}_{n_0,n_t\to+\infty}\widehat{HIL}^{ch}_{0,t} = \frac{h\Big((\operatorname{plim}_{n_0\to+\infty}\overline{\mathbf{x}}^0)'(\operatorname{plim}_{n_t\to+\infty}\hat{\boldsymbol{\beta}}^t)\Big)}{h\Big((\operatorname{plim}_{n_0\to+\infty}\overline{\mathbf{x}}^0)'(\operatorname{plim}_{n_0\to+\infty}\hat{\boldsymbol{\beta}}^0)\Big)} = \frac{h(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t)}{h(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0)}.$$

□

Proof of Corollary 3.1. We demonstrate point i). The proof of point ii) is similar. Using the fact that $h^{(l)}(x) = e^x\ \forall l$, we simply have

$$HIL^{si}_{0,t} = \frac{\sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t)}{l!}\mu^l_{\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t}}{\sum_{l=0}^{\infty}\frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0)}{l!}\mu^l_{\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^0+\epsilon_i^0}}\frac{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t}}{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0}}\frac{\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t}}{l!}}{\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^0+\epsilon_i^0}}{l!}} = \frac{\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{l!}}{\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!}}HIL^{ch}_{0,t}.$$

□

Proof of Lemma 3.1. As demonstrated in Theorem 3.2, we have that

$$\underset{n_0,n_t\to+\infty}{\text{plim}} \widehat{HIL}^{si}_{0,t} = \frac{\sum_{l=0}^{\infty} \frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t)}{l!}\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{\sum_{l=0}^{\infty} \frac{h^{(l)}(\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0)}{l!}\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}} = \frac{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t}\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{l!}}{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0}\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!}}.$$

We demonstrate the convergence of the series in the denominator and numerator. Let $\sigma_{\mathbf{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}$ denote the standard deviation of the transformed prices. Since $\mu^l_{(\mathbf{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}$ corresponds to the l-th central moment of a normally distributed variable, we have that $\mu^l_{\mathbf{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t} = (l-1)!!\sigma^l_{\mathbf{x}^{t\prime}\boldsymbol{\beta}^t+\epsilon^t}$ if l is even and 0 otherwise, where $(l-1)!!$ denotes the double factorial of $l-1$. We thus have

$$\begin{aligned}
\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!} &= \sum_{l\ even}^{\infty}\frac{(l-1)!!\sigma^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!} = \sum_{k=1}^{\infty}\frac{(2k-1)!!\sigma^{2k}_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{(2k)!} = \\
&= \sum_{k=1}^{\infty}\frac{\frac{(2k)!\sigma^{2k}_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{2^k k!}}{(2k)!} = \sum_{k=1}^{\infty}\frac{\sigma^{2k}_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{k!2^k} = \\
&= \sum_{k=1}^{\infty}\frac{(\sigma^2_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0})^k}{k!}\frac{1}{2^k} < \sum_{k=1}^{\infty}\frac{(\sigma^2_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0})^k}{k!} = e^{\sigma^2_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}.
\end{aligned}$$

Since the series is monotonic and bounded, it is convergent.

For the series in the numerator we similarly have

$$\begin{aligned}
\sum_{l=0}^{\infty}\frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{l!} &= \sum_{l\ even}^{\infty}\frac{(l-1)!!\sigma^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{l!} = \sum_{k=1}^{\infty}\frac{(2k-1)!!\sigma^{2k}_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{(2k)!} = \\
&= \sum_{k=1}^{\infty}\frac{\frac{(2k)!\sigma^{2k}_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{2^k k!}}{(2k)!} = \sum_{k=1}^{\infty}\frac{\sigma^{2k}_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{k!2^k} = \sum_{k=1}^{\infty}\frac{(\sigma^2_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t})^k}{k!}\frac{1}{2^k} < \\
&< \sum_{k=1}^{\infty}\frac{(\sigma^2_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t})^k}{k!} = e^{\sigma^2_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}.
\end{aligned}$$

Since the series is monotonic and bounded, it is convergent. The ratio of two convergent series is convergent. □

Proof of Lemma 3.2. The proof is similar to the proof of theorem 3.2. The convergence in probability of $\widehat{HIL}^{si}_{0,t}$ is first established:

$$\begin{aligned}
\underset{n_0,n_t\to+\infty}{\text{plim}} \widehat{HIL}^{*si}_{0,t} &= \underset{n_0,n_t\to+\infty}{\text{plim}} \frac{\sum_{i=1}^{n_0} e^{\mathbf{x}_i^{0\prime}\hat{\boldsymbol{\beta}}^t+\frac{1}{2}(\hat{\sigma}^t)^2}}{\sum_{i=1}^{n_0} P_i^0} = \\
&= \underset{n_0\to+\infty}{\text{plim}} \frac{\sum_{i=1}^{n_0} e^{\mathbf{x}_i^{0\prime}\,\text{plim}_{n_t\to+\infty}\hat{\boldsymbol{\beta}}^t+\frac{1}{2}\,\text{plim}_{n_t\to+\infty}(\hat{\sigma}^t)^2}}{\sum_{i=1}^{n_0} P_i^0} = \\
&= \underset{n_0\to+\infty}{\text{plim}} \frac{\frac{1}{n_0}\sum_{i=1}^{n_0} e^{\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t+\frac{1}{2}(\sigma^t)^2}}{\frac{1}{n_0}\sum_{i=1}^{n_0} P_i^0} =
\end{aligned}$$

$$= e^{\frac{1}{2}(\sigma^t)^2} \frac{\text{plim}_{n_0\to+\infty} \frac{1}{n_0}\sum_{i=1}^{n_0} e^{\mathbf{x}_i^{0\prime}\boldsymbol{\beta}^t}}{\text{plim}_{n_0\to+\infty} \frac{1}{n_0}\sum_{i=1}^{n_0} P_i^0} =$$

$$= e^{\frac{1}{2}(\sigma^t)^2} \frac{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t} \sum_{l=0}^{\infty} \frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{l!}}{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0} \sum_{l=0}^{\infty} \frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!}}.$$

For $\widehat{HIL}^{*ch}_{0,t}$, we have

$$\underset{n_0,n_t\to+\infty}{\text{plim}} \widehat{HIL}^{*ch}_{0,t} = \underset{n_0,n_t\to+\infty}{\text{plim}} \frac{e^{\overline{\mathbf{x}}^{0\prime}\hat{\boldsymbol{\beta}}^t+\frac{1}{2}(\hat{\sigma}^t)^2}}{e^{\overline{\mathbf{x}}^{0\prime}\hat{\boldsymbol{\beta}}^0+\frac{1}{2}(\hat{\sigma}^0)^2}} =$$

$$= \underset{n_t\to+\infty}{\text{plim}} \frac{e^{\text{plim}_{n_0\to+\infty}\overline{\mathbf{x}}^{0\prime}\hat{\boldsymbol{\beta}}^t+\frac{1}{2}(\hat{\sigma}^t)^2}}{e^{\text{plim}_{n_0\to+\infty}\overline{\mathbf{x}}^{0\prime}\hat{\boldsymbol{\beta}}^0+\frac{1}{2}\text{plim}_{n_0\to+\infty}(\hat{\sigma}^0)^2}} =$$

$$= \underset{n_t\to+\infty}{\text{plim}} \frac{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\hat{\boldsymbol{\beta}}^t+\frac{1}{2}(\hat{\sigma}^t)^2}}{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0+\frac{1}{2}(\sigma^0)^2}} = \frac{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t+\frac{1}{2}(\sigma^t)^2}}{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0+\frac{1}{2}(\sigma^0)^2}}.$$

□

Proof of Lemma 3.3. We only demonstrate point i). The proof of point ii) is similar. Using the fact that $h^{(l)}(x) = e^x \ \forall l$, we simply have

$$HIL^{si}_{0,t} = e^{\frac{1}{2}(\sigma^t)^2} \frac{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^t} \sum_{l=0}^{\infty} \frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^t}}{l!}}{e^{\boldsymbol{\mu}'_{\mathbf{x}^0}\boldsymbol{\beta}^0} \sum_{l=0}^{\infty} \frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!}} = e^{\frac{1}{2}(\sigma^0)^2} \frac{\sum_{l=0}^{\infty} \frac{\mu^l_{(\mathbf{x}^0)'\boldsymbol{\beta}^t}}{l!}}{\sum_{l=0}^{\infty} \frac{\mu^l_{\mathbf{x}^{0\prime}\boldsymbol{\beta}^0+\epsilon^0}}{l!}} HIL^{*ch}_{0,t}.$$

□

Property 2. *If in time period t a linear hedonic function is assumed, then*

i) $\underset{n_t\to+\infty}{\text{plim}} \frac{1}{n_t}\sum_{i=1}^{n_t} P_i^t = \boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^t$

ii) $\underset{n_t\to+\infty}{\text{plim}} \frac{1}{n_t}\sum_{i=1}^{n_t} \boldsymbol{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t = \boldsymbol{\mu}'_{\boldsymbol{x}^t}\boldsymbol{\beta}^t$.

Proof.

i) Due to the linear model hypotheses, we have that $\mathbb{E}(\epsilon_i^t|\mathbf{x}_i^t) = 0$. A stronger form of exogeneity is not necessary, since the characteristics vectors $\mathbf{x}_i^t$, $i = 1, ..., n_t$ are assumed to be independent in a given time period. Using the law of iterated expectations, we have

$$\mathbb{E}(P_i^t) = \mathbb{E}_{\mathbf{x}^t}(\mathbb{E}(P_i^t|\mathbf{x}_i^t)) = \mathbb{E}_{\mathbf{x}^t}(\mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t) = \boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t < +\infty \ \forall i.$$

According to Khinchine's weak law of large numbers, the proof is complete.

ii) Due to the fact that the estimated prices $\hat{P}_i^t = \mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t$ are not independent, we use the Chebyshev's inequality to demonstrate the convergence in probability. Let $\boldsymbol{\epsilon}^t := (\epsilon_1^t, ..., \epsilon_{n_t}^t)'$ denote the vector of the random errors. We start by computing the mean of the considered random variable. Using the exogeneity hypothesis, we have $\mathbb{E}(\boldsymbol{\epsilon}^t|\mathbf{X}^t) = \mathbf{0}$. We condition on the whole characteristics matrix $\mathbf{X}^t$ and use the law of iterated expectations.

$$\begin{aligned}
\mathbb{E}(\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t) &= \mathbb{E}_{\mathbf{X}^t}(\mathbb{E}((\mathbf{x}_i^t)'\hat{\boldsymbol{\beta}}^t)|\mathbf{X}^t) = \\
&= \mathbb{E}_{\mathbf{X}^t}\left(\mathbb{E}\left((\mathbf{x}_i^t)'(\boldsymbol{\beta}^t + ((\mathbf{X}^t)'(\mathbf{X}^t))^{-1}(\mathbf{X}^t)'\boldsymbol{\epsilon}^t)|\mathbf{X}^t\right)\right) = \\
&= \mathbb{E}_{\mathbf{X}^t}\left(\mathbb{E}\left((\mathbf{x}_i^t)'\boldsymbol{\beta}^t + (\mathbf{x}_i^t)'((\mathbf{X}^t)'(\mathbf{X}^t))^{-1}(\mathbf{X}^t)'\boldsymbol{\epsilon}^t|\mathbf{X}^t\right)\right) = \\
&= \mathbb{E}_{\mathbf{x}_i^t}((\mathbf{x}_i^t)'\boldsymbol{\beta}^t) = \boldsymbol{\mu}'_{\mathbf{x}^t}\boldsymbol{\beta}^t \ \forall i.
\end{aligned}$$

We still have to show that $\lim_{n_t \to +\infty} V(\frac{1}{n_t}\sum_{i=1}^{n_t}\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t) = 0$. We start by examining the variance matrix of the vector $\hat{\mathbf{P}}^t := (\hat{P}_1^t, ..., \hat{P}_{n_t}^t)$. Let $\mathbf{H}^t := \mathbf{X}^t(\mathbf{X}^{t\prime}\mathbf{X}^t)^{-1}\mathbf{X}^{t\prime}$ denote the hat matrix at time t. Using the variance decomposition we have

$$\begin{aligned}
V(\hat{\mathbf{P}}^t) &= \mathbb{E}(V(\hat{\mathbf{P}}^t|\mathbf{X}^t)) + V(\mathbb{E}(\hat{\mathbf{P}}^t|\mathbf{X}^t)) = \mathbb{E}(V(\mathbf{H}^t\mathbf{P}^t|\mathbf{X}^t)) + V(\mathbf{X}^t\boldsymbol{\beta}^t) = \\
&= \mathbb{E}(\mathbf{H}^t V(\mathbf{P}^t|\mathbf{X}^t)\mathbf{H}^{t\prime}) + V(\mathbf{X}^t\boldsymbol{\beta}^t) = \sigma^2\mathbb{E}(\mathbf{H}^t) + V(\mathbf{X}^t\boldsymbol{\beta}^t)
\end{aligned}$$

Since the random variables $\mathbf{x}_1^t, ..., \mathbf{x}_{n_t}^t$ are independent, the $\mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t, ..., \mathbf{x}_{n_t}^{t\prime}\boldsymbol{\beta}^t$ are also independent. The off-diagonal elements of $V(\mathbf{X}^t\boldsymbol{\beta}^t)$ are thus equal to zero. On the contrary, its diagonal elements are equal to $V(\mathbf{x_i^t}'\boldsymbol{\beta}^t) = \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t$. We thus have

$$V(\hat{P}_i^t) = \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \sigma^2\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_i),$$

where $\mathbf{e}_i = (0...010...0)'$ denotes a n_t-dimensional column vector with the i-th component equal to 1 and zero otherwise. The covariances are given by

$$Cov(\hat{P}_i^t, \hat{P}_j^t) = \sigma^2\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_j).$$

Let Tr denote the trace operator and $\mathbf{U}$ a $n_t \times n_t$ matrix with components equal to 1. We have

$$\begin{aligned}
V(\frac{1}{n_t}\sum_{i=1}^{n_t}\hat{P}_i^t) &= \frac{1}{n_t^2}\sum_{i=1}^{n_t}V(\hat{P}_i^t) + \frac{1}{n_t^2}\sum_{i\neq j}^{n_t}Cov(\hat{P}_i^t, \hat{P}_j^t) = \\
&= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{1}{n_t^2}\sum_{i=1}^{n_t}\sigma^2\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_i) + \frac{1}{n_t^2}\sum_{i\neq j}^{n_t}\sigma^2\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_j) =
\end{aligned}$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t^2}\sum_{i=1}^{n_t}\mathrm{Tr}(\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_i)) + \frac{\sigma^2}{n_t^2}\sum_{i\neq j}^{n_t}\mathrm{Tr}(\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_j)) =$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\mathrm{Tr}(\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_j)) =$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t^2}\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\mathbb{E}(\mathrm{Tr}(\mathbf{H}^t\mathbf{e}_j\mathbf{e}_i')) =$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t^2}\mathbb{E}(\mathrm{Tr}(\mathbf{H}^t\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}\mathbf{e}_j\mathbf{e}_i')) =$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t^2}\mathbb{E}(\mathrm{Tr}(\mathbf{H}^t\mathbf{U})) =$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t^2}\mathbb{E}(\mathrm{Tr}(\mathbf{U})) =$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t^2}\mathbb{E}(n_t) =$$

$$= \frac{1}{n_t}\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{n_t\sigma^2}{n_t^2} \longrightarrow 0 \text{ as } n_t \longrightarrow \infty,$$

where we have used the fact that, since the column vectors of $\mathbf{U}$ are equal to $(1, ..., 1)'$, their projection in the vector space generated by $\mathbf{X^t}$ correspond to the identity function (the linear model contains a constant term).

□

Property 3. *We assume a linearizable hedonic function in period t.*

i) $\underset{n_t\to+\infty}{\text{plim}} \frac{1}{n_t}\sum_{i=1}^{n_t}((\boldsymbol{x}_i^t)'\boldsymbol{\beta}^t + \epsilon_i^t - \boldsymbol{\mu}_{\boldsymbol{x}^t}'\boldsymbol{\beta}^t)^2 = \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \sigma^2$

ii) The mean $\frac{1}{n_t}\sum_{i=1}^{n_t}((\boldsymbol{x}_i^t)'\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\boldsymbol{x}^t}'\boldsymbol{\beta}^t)^2$ does not satisfy Chebychev's sufficient conditions for convergence in probability.

Proof.

i) The mean the considered sequence of random variables is equal to $\mathbb{E}((\mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2) = V(P_i^t) = \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \sigma^2 \forall\ i$ (see Property 2). Since the $(\mathbf{x}_i^{t\prime}\boldsymbol{\beta}^t + \epsilon_i^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2$, $i = 1, ..., n_t$ are independent and identically distributed, this completes the proof (Khinchine's weak law of large numbers).

ii) Since the $(\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2$ are neither identically distributed, nor independent, we use the Chebyshev's inequality to prove the convergence in probability. The mean is equal to

$$
\begin{aligned}
\mathbb{E}(\frac{1}{n_t}\sum_{i=1}^{n_t}(\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2) &= \frac{1}{n_t}\sum_{i=1}^{n_t}\mathbb{E}((\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2) = \\
&= \frac{1}{n_t}\sum_{i=1}^{n_t}V(\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t) = \\
&= \frac{1}{n_t}\sum_{i=1}^{n_t}\Big(\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \sigma^2\mathbb{E}(\mathbf{e}_i'\mathbf{H}^t\mathbf{e}_i)\Big) = \\
&= \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t}\mathbb{E}(\mathrm{Tr}(\mathbf{H}^t\sum_{i=1}^{n_t}\mathbf{e}_i\mathbf{e}_i')) = \\
&= \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{\sigma^2}{n_t}\mathbb{E}(\mathrm{Tr}(\mathbf{H}^t)) = \\
&= \boldsymbol{\beta}^{t}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{(K+1)\sigma^2}{n_t} \longrightarrow \boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t \text{ as } n_t \to \infty.
\end{aligned}
$$

The variance is given by

$$
\begin{aligned}
V(\frac{1}{n_t}\sum_{i=1}^{n_t}(\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2) &= \mathbb{E}((\frac{1}{n_t}\sum_{i=1}^{n_t}(\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2)^2) - \\
&(\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t + \frac{(K+1)\sigma^2}{n_t})^2 = \\
&= \frac{1}{n_t^2}\sum_{i=1}^{n_t}\mathbb{E}((\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^4) + \\
&\frac{1}{n_t^2}\sum_{i\neq j}^{n_t}\mathbb{E}((\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2(\mathbf{x}_j^{t\prime}\hat{\boldsymbol{\beta}}^t - \boldsymbol{\mu}_{\mathbf{x}^t}'\boldsymbol{\beta}^t)^2) \\
&- (\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t)^2 - 2(\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t\frac{(K+1)\sigma^2}{n_t}) + \\
&(\frac{(K+1)\sigma^2}{n_t})^2 = \\
&= \frac{1}{n_t^2}\sum_{i\neq j}^{n_t}\mu^4_{\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t} + \frac{1}{n_t^2}\sum_{i\neq j}^{n_t}\mu^{2,2}_{\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t} - (\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t)^2 \\
&- 2(\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t\frac{(K+1)\sigma^2}{n_t}) - (\frac{(K+1)\sigma^2}{n_t})^2,
\end{aligned}
$$

where $\mu^4_{\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t}$ and $\mu^{2,2}_{\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t}$ represent the fourth central moment and the higher order covariance of the variables of $\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t,\ i = 1, ..., n_t$. Clearly, the last two terms of the above expression tend to zero when n_t goes to infinity. Thus, in order to obtain a zero variance when n_t goes to infinity, we should have that

$$
\frac{1}{n_t^2}\sum_{i\neq j}^{n_t}\mu^4_{\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t} + \frac{1}{n_t^2}\sum_{i\neq j}^{n_t}\mu^{2,2}_{\mathbf{x}_i^{t\prime}\hat{\boldsymbol{\beta}}^t} \to (\boldsymbol{\beta}^{t\prime}\Sigma_{\mathbf{x^t}}\boldsymbol{\beta}^t)^2.
$$

Since in general this condition is not satisfied, the variance of the sample second central moment will not converge toward zero.

□

BIS Department for Business Innovation & Skills

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Llywodraeth Cymru
Welsh Government

Spatial Economics Research Centre (SERC)
London School of Economics
Houghton Street
London WC2A 2AE

Tel: 020 7852 3565
Fax: 020 7955 6848
Web: www.spatialeconomics.ac.uk

SERC is an independent research centre funded by the Economic and Social Research Council (ESRC), Department for Business Innovation and Skills (BIS) and the Welsh Government.