# Construction of a Spatial Housing Price Index by Estimating an Almost Ideal Demand System

Víctor Iturra
Dusan Paredes

REAL 11-T-07 May, 2011

# Construction of a Spatial Housing Price Index by Estimating an Almost Ideal Demand System

Víctor Iturra
Universidad Católica del Norte, Chile
Dusan Paredes
Universidad Católica del Norte, Chile

**Abstract:** The aim of this article is to build a spatial housing price index for the Chilean communities (the *commune* political body similar to a municipality or county). The first step is to recover hedonic prices of three housing attributes by estimating a mixed index model (Bowden, 1992) using a generalized method of moments procedure. Secondly, a censored almost ideal demand system (Deaton and Muellbauer, 1980) is estimated to build expenditures for every community and to compare them among different spatial units, while maintaining a constant utility level. Using micro data from the 2009 CASEN survey, we show that there are important differences in housing prices among Chilean communities and the axiomatic approach tends to understate spatial index compared to economic approach.

Classification JEL: R21, C34, C36
Key words: spatial housing price index, mixed index, demand system with censored data, economic approach

## 1. Introduction

A cost of living index (CLI) measures the monetary cost differential for a representative consumer in the context of two distinct prices while maintaining a constant level of welfare (Konus, 1924). Both the literature and statistical offices relate the CLI to a temporal variation, or inflation. While this measurement is relevant for economic policies, such as monetary or wage adjustment policies, time is not the only dimension of analysis. Space is excluded from the theoretical and empirical discussion, namely, that two individuals in tow different locations face different expenditures to reach the same level of welfare. This implies that the CLI also measures the spatial differences of cost of living index (SCLI), correcting the comparison of monetary units such as the GDP or regional poverty rates. Moreover, while price differentials are marginal over time, they can be greater across space, reflecting variations in transport costs due to market clearing conditions. Thus, an SCLI could present greater variability than inflation, making its estimation necessary for regional economic analysis. Given that official statistics do not generate this measure for Chile (nor for the majority of countries), this article contributes to the estimation of a proxy of the SCLI, namely a spatial housing price index (SHPI) for 2009 at the smallest available scale, the community level.

The SCLI must be computed with cost information for each of the consumer goods and services in the two regions. However, these data are not readily accessible in official

statistics[1], making it necessary to work with limited information. In this sense, the literature recognizes the item "housing" as an adequate proxy for the SCLI because it is non-traded spatially, so that its differentials represent the price pressures associated with each spatial unit (Timmins, 2006). In the case of Chile, the National Socioeconomic Characterization Survey (CASEN) contains the final prices and housing characteristics for the major communities. Considering both, the theoretical argument of Timmins (2006) and the limited information available, this paper approximates the SCLI by estimating the SHPI, a task that has been developed in Chile in the earlier papers of Paredes and Aroca (2008) and Paredes (2009).

Housing is an heterogeneous good. Consequently, housing market prices should not be compared without correcting for their characteristics. In this respect Paredes and Aroca (2008) proposed estimating the hedonic prices of "similar" housing across space. Subsequently, they used average regional hedonic prices and a fixed consumer housing basket to estimate regional costs, a technique known as the axiomatic approach (Pollak, 1980). Paredes (2009) followed the axiomatic approach, but evaluating the robustness of diverse statistical techniques used to select housing units with "similar" characteristics.

Notwithstanding the earlier contributions of Paredes and Aroca (2008) and Paredes (2009), at least three weaknesses can be identified. First, the approach assumes that individuals necessarily acquire different housing across space. In contrast to Paredes and Aroca (2008) and Paredes (2009), it is not required that the SHPI represent the cost differential for a fixed consumer basket, but rather it is assumed that different geographic or urban characteristics imply the potential for a spatial substitution of the consumed characteristics[2].

Secondly, the methodology modifies the standard regression of hedonic prices in assuming that the socioeconomic characteristics of individuals affect their consumption of housing. Unlike Paredes and Aroca (2008) and Paredes (2009), the mixed index (Bowden, 1992) is used to estimate hedonic prices, thus reducing potential problems of consistency in the estimation and improving the development of the SHPI.

Finally, Paredes and Aroca (2008) and Paredes (2009) estimate the SHPI at the regional level, assuming that this measurement correctly represents intra-regional spatial heterogeneity. The present paper provides the first estimation of a community-based SHPI for Chile, providing price information at the lowest available spatial scale.

These contributions have two main elements. The first is related to the methodology to estimate an SHPI, in which space imposes critical restrictions in its construction. The second element is related to the use of the SHPI to adjust monetary units at the community level. This information will contribute to understanding diverse economic phenomenon, such as inter-community commuting or the development of urban zoning policies, among others.

In this context, the objective of this paper is to construct a community-based SHPI for Chile. The hypothesis of this work is that the purchasing power of consumers is not homogenous across different communes, that is, that consumers face substantial differences in prices to access the same level of utility between two distinct communities.

---

[1] While the 6th family budget survey undertaken by the National Statistics Institute (INE) gathered information about consumer expenditures in the regional capitals, the information is not applicable for regional comparison because it does not provide a representative sample.

[2] For example, housing in southern Chile has specific characteristics owing to the climate, such as the quality of roofs and walls to confront low temperatures and heavy rains, in contrast to housing in the north where the climate does not making such housing characteristics necessary.

This paper is organized as following. Section 2 presents a review of literature in Chile about hedonic prices and housing price index. The following section presents the theoretical framework underlying the contributions of this article. Section 4 discusses the methodology used and Section 5 presents the data to make estimations. The results are then presented and discussed, followed by presentation of the main conclusions.

## 2. Literature Review

There is scarce literature about hedonic prices and the construction of an SHPI. The only papers that have discussed the theme as it applies to Chile are those of Paredes and Aroca (2008) and Paredes (2009). This section reviews both papers, highlighting their contributions and limitations. Having identified the two elements, the main proposals found in the literature to deal with these problems and improve existing estimations will be discussed.

### 2.1 Literature in Chile

The estimation of a price index for homogeneous goods uses prices, quantities and a defined cost function (Pollak, 1980). However, heterogeneity in housing imposes special conditions given that only the final price is observed and not the prices of housing characteristics. This implies that two steps are needed to construct the SHPI. First, the shadow prices of each of these characteristics are obtained with hedonic regressions. With the prices and the housing characteristics (quantity), it is possible to reconstruct the cost and estimate the SHPI.

Rosen (1974) formalized these two stages. Firstly, the "hedonic prices" of each characteristic are estimated with a non-linear regression that allows price variations among consumers. The hedonic prices and the quantities consumed are used to construct costs. Papers dealing with this first step include Lira (1978) and Figueroa and Lever (1992). Lira (1978) used a sample of houses and apartments in fourteen communes in the Metropolitan Region (MR), and estimated that square meters and location positively affect rental price. Figueroa and Lever (1992) estimated that surface area and the socioeconomic level of the neighborhood positively affect the price of constructed square meter for the Metropolitan Region. While both papers represent pioneering studies in estimating hedonic prices for Chile, they focus on the Metropolitan Region and do not extend their results to other regions of the country or to the construction of an SHPI.

An advance emerged in the literature with Paredes and Aroca (2008), who constructed an SHPI in two steps. Firstly, they used a statistical technique to find similar housing units or "clones" across space and estimate the hedonic regression for these housing groups. They equalized housing units among regions to then estimate hedonic prices. The second step implies using a fixed consumer housing basket and the hedonic prices estimated to compute regional costs and obtain the SHPI. The results for 2003 indicate that Antofagasta has the highest housing prices in Chile.

Paredes (2009) evaluated the robustness of the methodology to find similar housing units among the regions. The author evaluated techniques that minimize the differences among housing characteristics in the base region and comparison regions. The results indicate that Antofagasta was again the most expensive region in Chile in 2006.

The literature in Chile has advanced, from estimating housing price determinants using hedonic prices, to estimating a regional SHPI for Chile. However, there are two problems that have not been addressed in the literature, which impedes considering the estimated indexes as

an adequate approximation of the SCLI. The first problem is bias in estimating hedonic prices and the second is the use of a fixed consumer housing basket across space that does not reflect the geographic and urban heterogeneity of housing characteristics. The following briefly reviews the literature that covers these problems and proposed solutions.

### 2.2 Hedonic prices for housing

Bowden (1992) suggested that the estimation of hedonic housing prices is theoretically biased when it does not consider the characteristics of the owner or renter. The author argued that prices have an allocation role, so that a correlation structure is created between the characteristics of the individual and those of the good or service in demand. The author called this the mixed index problem. In the case of cross-sectional studies, the mixed index problem is characterized because the housing characteristics and those of the individual contain two indices, one for the housing and the other for the individual, thus affecting the estimation of hedonic prices.

The characteristics of housing and of the individual are not independent, because of which the error term is related to the regressors. This situation does not allow for maintaining the assumption of strict exogeneity of the housing characteristics, producing bias in the estimation of hedonic prices and distorting the SHPI. Bowden (1992) proposed that the hedonic regressions should be estimated as a system of equations between the characteristics of housing and of the individual, such that the correlation of the elements is considered in the estimation of hedonic prices.

Brasington and Hite (2008) compared the performance of Bowden's mixed index to those of traditional hedonic price models. Their comparisons indicate that the mixed index provides better predictions than the traditional models, in particular for the variable quality of the housing environment. The comparison also suggested the estimation of hedonic prices is less skewed. The authors concluded that the mixed index improves the coefficients estimated from the hedonic regression.

The second step of this work implies estimating a cost function from a demand system to construct the SHPI for the communes of Chile, in which the consumer housing basket is not fixed *a priori*, but rather the level of utility of the consumer is fixed.

### 2.3 Demand system for housing

Obtaining the hedonic prices in the first step provides the prices for each characteristic per individual. The second step estimates a demand system to obtain costs in function of a constant level of welfare. Parsons (1986) estimated the demand functions for seven cities in the United States using an almost ideal demand system (AIDS) (Deaton and Muellbauer, 1980) for four housing attributes. The AIDS maintains the conventional properties of the consumer theory and recovers the cost function that depends on prices and utility from estimated parameters of the demand system. The comparison of costs among spatial units, while maintaining the level of consumer welfare constant, is known as the economic approach to construct SCLI. This paper follows this approach, that is, the consumer expenditure function is recovered by estimating a demand system.

The SHPI is estimated at the communal level, because of which it does not extrapolate regional information to the constituent communes of the respective regions. This avoids strong assumptions about the patterns of local supply of non-traded commodities (Timmins,

2006). As well, the high level of residential segregation in medium-sized and large cities implies that there is a high degree of heterogeneity in the SHPI at the intra-regional level. Consequently, using the commune as the spatial unit of analysis implies greater homogeneity in price information and characteristics. In this sense, the SHPI is constructed with greater precision, which allows for its use both for urban planning policies and for adjusting monetary unit for this spatial scale, as well as for policy evaluation (Parsons, 1986).

In summary, the estimation of the SHPI involves two steps, the first of which involves estimating hedonic prices, while the second estimates the demand system. Both steps recognize the contributions of the literature to overcome the limitations of the papers of Paredes and Aroca (2008) and Paredes (2009), that is, obtaining more precise hedonic prices by reducing bias through the use of the mixed index (Bowden, 1992) and assuming heterogeneity in housing characteristics, using the estimation of an almost ideal demand system (Deaton and Muellbauer, 1980).

## 3. Theoretical framework

This section reviews the theoretical elements in estimating hedonic prices using the mixed index. As well, two additional sources of bias in estimating the system of equations are explained. With the hedonic prices, the second step estimates a demand system. This section details the formulation of a demand system and the assumptions considered in estimating an AIDS, indicating its relationship to the construction of the CLI from this approach.

### 3.1 Hedonic Prices and the Mixed Index

Rosen (1974) defined hedonic prices as the implicit prices of attributes. Housing is defined as a good composed of attributes or characteristics summarized in a vector of $J$ elements $\boldsymbol{Z} = Z_1, \ldots., Z_j$. The housing price will depend on the set of its characteristics, consequently it is a function $P(\boldsymbol{Z})$, increasing in all its arguments, continuous and has a second defined derivative. The price of the housing can be expressed as:

$$P = P\left(Z_1, \ldots., Z_j\right) = P(\boldsymbol{Z}) \tag{1}$$

Rosen (1974) pointed out that to recover hedonic prices it is necessary to estimate a regression of observed prices of housing on all of their characteristics, using the best fitting functional form. Thus, the set of implicit marginal prices is computed as:

$$\frac{\partial P}{\partial Z_j} = p\left(Z_j\right) \tag{2}$$

Rosen (1974) did not indicate the appropriate functional form to specify the hedonic regression. However, the author indicated that a non-linear functional form should be used to ensure that it captured the variation of hedonic prices among the individuals of the selected sample.

The mixed index implies that the housing price depends on both the characteristics of housing and of the individuals, specifically their level of income, which are not independent. For example, better quality housing is associated with individuals with higher incomes. Consequently, there is no strict exogeneity between the regressors (such as quality) and the elements included in the error of estimation (such as certain unobserved determinants of income).

Given the simultaneousness of housing prices and the economic characteristics of the individual, the literature proposes a two stage economic model to reduce bias in the coefficients; the equation of hedonic housing prices, while the specification about the determinants of individual income comes in the second stage (Brasington and Hite, 2008):

$$\boldsymbol{P} = P\left(Z_1, \ldots., Z_j, \boldsymbol{Y}\right) \qquad (3)$$
$$\boldsymbol{Y} = Y(X_1, \ldots., X_l)$$

Where $\boldsymbol{P}$ and $\boldsymbol{Y}$ are the price vectors of housing and incomes of individuals, respectively, both of dimension $1xN$ and $\boldsymbol{X} = X_1, \ldots, X_l$ is the vector of $L$ determinants of the incomes of the individuals.

This model, in its estimable version, estimates hedonic housing prices and reduces bias in the coefficients. However, the present research addresses two additional sources of bias. The first source is those factors that affect the quality of housing, such as the physical attributes of the city or amenities (Glaeser, 2001), where variables related to the housing environment affect the quality of housing. However, information like this is rarely available at a small spatial scale (the block or neighborhood) to quantify this relationship, implying a potential source of endogeneity. As a result, the proposed approach will estimate the system of equations through the generalized method of moments to reduce bias in the estimation of the parameters related to the quality of housing.

Combes et al. (2008) proposed three sources to explain the spatial disparity of salaries. The first arises from the differences in the abilities of the labor force among different regions. The second comes directly from the initial endowment of some regions that gives rise to higher levels of productivity of the labor force. The third arises directly from the positive impact of the interaction among businesses on labor productivity.[3]

According to Combes et al. (2008), the Mincer equation used in the income regression in this research controls the first source of bias but does not consider the second and third. Thus, the impact of the initial endowments and agglomeration externalities are reflected in the coefficient of years of education of the individual. The use of instruments through the generalized method of moments filters the information from the agglomeration economies of the estimated coefficient of the years of education of the individual.

Once the hedonic prices of housing characteristics are obtained, the second step involves estimating a demand system to finally construct the SHPI. The following presents the theoretical basis for the construction of the SHPI, the approach previously used in the literature in Chile and the approach proposed by this research.

### 3.2 Cost of Living Index

Pollak (1980) pointed out that a CLI is the ratio of expenditures required to reach a particular base indifference curve given two distinct price regimes. In a spatial dimension, the CLI measures the minimum expenditure necessary to reach a level of utility in a region relative to the necessary expenditures in another reference spatial unit (Timmins, 2006).

[3] Marshall (1920) identified three sources of agglomeration economy, dense productive linkages, a dense labor market and technological externalities in the industrial localization.

Paredes and Aroca (2008) and Paredes (2009) approximated the SCLI using a fixed housing basket, that is, from an axiomatic approach. This research proposes approximating the SCLI from a heterogeneous consumer housing basket, that is, from an economic approach. The following reviews the theoretical basis for the construction of the CLI from this approach,[4] indicating the advantages of the economic approach proposed by this research and its relationship to the AIDS.

As Konus (1924) indicated, the calculation of the true cost of living index under the economic approach requires the combination of consumer goods that establishes a level of reference utility. The problem in constructing a true cost of living index is establishing a functional relationship between the quantity consumed and prices. Formally, the CLI is expressed as:

$$CLI = \frac{e(p^c, \bar{u})}{e(p^b, \bar{u})} \tag{4}$$

Where $p^c$ and $p^b$ represent the price vectors in the comparison (c) and base (b) spatial units, respectively, $e$ represents a homogeneous cost function of degree one that is nondecreasing, continuous and concave in prices and strictly increasing and continuous in utility, which provides the minimal cost to reach the level of utility $\bar{u}$ in the context of prices $p$. As can be appreciated in equation (4), the quantities consumed in the base and comparison regions are not fixed, given that they depend on the local price levels. This is the main difference from the axiomatic approach in which quantities and prices are independent (Breuer, 2007).

Paredes and Aroca (2008) and Paredes (2009) approximated the true CLI using the Fisher's ideal index, which uses a fixed consumer basket assuming an homogenous quadratic utility function. However, it is not known if this function represents average consumer preferences for every spatial unit. The economic approach, using the demand system estimation of Deaton and Muellbauer (1980), only maintains the utility level fixed among all spatial units, without assuming a functional form. The consumer housing basket in each community can vary according to the differences in the relative prices of the characteristics, which requires a functional relationship between the prices and characteristics of the housing consumed through a demand system.

Deaton and Muellbauer (1980) proposed using the AIDS to specify this functional form. The AIDS provides a first order approximation of any demand system, satisfying the axioms of choice and aggregation without requiring that the demand for goods is proportional to variations in income (Deaton and Muellbauer, 1980). The cost function for the model is defined as:

$$\ln e(p_k, \bar{u}) = \alpha_0 + \sum_i \alpha_i \ln p_{ik} + \frac{1}{2}\sum_i \sum_j \gamma_{ij} \ln p_{ik} \ln p_{jk} + \bar{u}\beta_0 \prod_i p_{ik}^{\beta_i} \tag{5}$$

Where $p_{ik}$ and $p_{jk}$ are the hedonic prices of the characteristics $i$ and $j$ of the community$k$; insofar as $\alpha_0, \alpha_1, \gamma_{ij}, \beta_0\ \beta_i$ are the parameters that define consumer preferences (Deaton and Muellbauer, 1980).

The derivation of the demand functions simplifies the estimation process, given that the number of parameters can be reduced. Using Shepard's lemma, the cost function is

[4] The stochastic approach is another method to approximate the CLI. For details on this methodology see Diewert (1995).

differentiated in respect to $\ln p_i$ to obtain the demand functions of the attributes as a proportion of the cost:

$$w_{ik} = \alpha_i + \sum_j \gamma_{ij} \ln p_{jk} + \beta_i \ln \left(\frac{x_k}{P_k}\right) \quad (6)$$

Where $w_{ik}$ is the budget share of i-ith characteristic of housing in region k; $x_k/P_k$ is the real expenditure of all housing attributes and the vector of $P_k$ is defined as:

$$\ln P_k = \alpha_0 + \sum_i \alpha_i \ln p_{ik} + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \ln p_{ik} \ln p_{jk} \quad (7)$$

Three properties of the consumer theory could be imposed in the parameters of the AIDS, namely: Slutsky symmetry, the homogeneity of degree zero in prices and expenditures and the adding-up condition:

$$\gamma_{ij} = \gamma_{ji} \quad (8)$$

$$\sum_j \gamma_{ij} = 0 \quad (9)$$

$$\sum_j \alpha_i = 1 \quad \sum_{i=1}^{n} \gamma_{ij} = 0 \quad \sum_{i=1}^{n} \beta_i = 0 \quad (10)$$

As already noted, estimating the SHPI with the economic approach requires a functional relationship between hedonic housing prices and the quantities consumed of each characteristic. The AIDS establishes this functional relationship without assuming a priori a functional form of the utility function and maintaining the conventional properties of consumer theory. Average consumer expenditures are recovered by estimating the parameters of the AIDS. Average consumer expenditure can vary over space owing to the differences in the relative prices of housing characteristics among communes. The utility level is the only fixed factor required to calculate the CLI from the economic approach. This condition can be clearly appreciated in expression (5):

$$\ln e(p_k, \bar{u}) = \alpha_0 + \sum_i \alpha_i \ln p_{ik} + \frac{1}{2} \sum_i \sum_j \gamma_{ij} \ln p_{ik} \ln p_{jk} + \bar{u} \beta_0 \prod_i p_{ik}^{\beta_i}$$

The construction of the SHPI under the economic approach compares costs between two communes, as expression (5) indicates, while maintaining the level of utility between these spatial units fixed. The prices facing consumers in the two communes are different. Consequently, the costs of housing attributes to reach the same level of utility are different.

## 4. Methodology

### 4.1 Hedonic prices regression

The hedonic prices of the housing characteristics are estimated using a non-linear functional form for both the hedonic and income regressions and a set of instruments for each regression. The following details the methodological considerations involved with reference to recovering hedonic prices.

As Halvorsen and Pollakowski (1981) indicated, the regression of hedonic prices, in their reduced form reflects the influences of supply and demand. Consequently, there is no

theoretically justified functional form for its estimation. Rosen (1974) indicated that the best fitting non-linear functional form should be used.

To carry out the estimation it is necessary to generate an estimable version of our economic model defined in (3). The housing price ($\boldsymbol{P}$) is assumed to be a random variable with $E[|\boldsymbol{P}|] < \infty$. Using the decomposition property, this random variable can be separated into a deterministic element, namely a conditional expectation and a random component (Angrist and Pischke, 2009; Wooldridge, 2002). Thus, the housing price paid by an individual $i$ is expressed as:

$$\boldsymbol{P}_i = E[P|Z] + \varepsilon_i \tag{11}$$

Where $\varepsilon_i$, by construction, is independent of $\boldsymbol{Z}$, that is $E[\varepsilon_i|\boldsymbol{Z}] = 0$. As well, the error term meets the condition $E[\varepsilon(\boldsymbol{Z})] = 0$ for any function defined for $\mathbf{Z}$. To obtain an estimable version, the researcher should assume an explicit functional form for the conditional expectation.

The literature in Chile in general has used a semi-logarithmic functional form (Paredes and Aroca, 2008; Paredes, 2009, Figueroa and Lever, 1992; Lira, 1978). This non-linear functional form allows recovered hedonic prices to vary for each observation of the sample. This variation at the individual level is fundamental to avoid the multico-linearity of the hedonic prices in the second step of the estimation. However, although the semi-logarithmic form is non-linear and recommended by the literature (Malpezzi, 2002), it presents a limitation for using recovered hedonic prices to estimate AIDS. For a semi-logarithmic form, the conditional expectation of the price paid by an individual $i$ for a housing unit with characteristics $j$ is expressed as:

$$E\left[\ln P_{ij}|Z\right] = \beta_0 + \sum_{j=1}^{J} \beta_{ij} Z_{ij} \tag{12}$$

Consequently, hedonic prices for housing characteristics are calculated with the following expression:

$$\hat{p}_i(Z_j) = e^{\widehat{\beta}_0 + \sum_{j=1}^{J} \widehat{\beta}_{ij} Z_{ij}} * \hat{\beta}_{ij} = \hat{P}_{ij} * \hat{\beta}_{ij} \tag{13}$$

Thus, the hedonic prices for the housing characteristics $j$ and $j+1$ are:

$$\hat{p}_i(Z_j) = \hat{P}_{ij} * \hat{\beta}_{ij} \tag{14}$$

$$\hat{p}_i(Z_{j+1}) = \hat{P}_{ij} * \hat{\beta}_{ij+1} \tag{15}$$

Expressions (14) and (15) correspond to two vectors of independent variables in the second step. Both indicate that for $n$ individuals in the sample, the difference between the hedonic prices of the characteristics $j$ and $j+1$ is $(\hat{\beta}_{ij} - \hat{\beta}_{ij+1})$, a constant difference for the $n$ individuals of the sample. This particularity of the semi-logarithmic functional form implies perfect multico-linearity among hedonic prices, that when used to estimate the parameters of the AIDS results in the matrix of $n$ individuals by $J$ housing characteristics being singular, it not being possible to estimate the demand system coefficients.

Following this outcome, the log-log functional form presents the non-linearity indicated by Rosen (1974), but unlike the semilogarithmic form, variability in hedonic prices is related to

the quantity of each attribute that is consumed by each individual in the sample (Pasha and Butt, 1996). The conditional expectation of this functional form is expressed as:

$$E\left[\ln P_{ij}|Z\right] = \beta_0 + \sum_{i=1}^{J} \beta_{ij} \ln Z_{ij} \tag{16}$$

The hedonic price for characteristic $j$ is calculated with the following expression:

$$\hat{p}_i(Z_j) = e^{\widehat{\beta}_0 + \sum_{i=1}^{J} \widehat{\beta}_{ij} Z_{ij}} * \frac{\widehat{\beta}_{ij}}{Z_{ij}} = \hat{P}_{ij} * \frac{\widehat{\beta}_{ij}}{Z_{ij}} \tag{17}$$

As can be appreciated in expression (17), the log-log functional form allows hedonic prices to vary among individuals as a function of the quantity of housing attributes they consume, without causing the perfect co-linearity of the semilogarithmic functional form. This functional form meets the necessary conditions to undertake the two steps proposed by Rosen (1974), that is, firstly the recovery of the hedonic prices of the characteristics and secondly the construction of consumer expenditure in each spatial unit.

The estimation of the hedonic regression coefficients, based on Bowden (1992), considers the application of a system of equations in which for a housing unit with characteristics $j$ demanded by individual $i$ this system is expressed as:

$$\begin{aligned} P_{ij} &= E\left[P|Z_{ij}, Y_i\right] + \varepsilon_{ij} \\ Y_i &= E[Y_i|X_i] + \eta_i \end{aligned} \tag{18}$$

Thus, it holds that:

$$E\left[Z_{ij} * Y_i\right] \neq 0 \tag{19}$$

$$E\left[Z_{ij} * \varepsilon_{ij}\right] \neq 0 \tag{20}$$

$$E[X_i * \eta_i] \neq 0 \tag{21}$$

The first condition reflects the dependence between the housing characteristics and those of the individuals, for which this system of equations is estimated to reduce bias in the estimation of the coefficients of the hedonic regression. According to our theoretical framework, the second and third conditions indicate that there is no strict exogeneity of the regressors with respect to the error term. To deal with these sources of endogeneity, two matrices of instruments were used, $\boldsymbol{K}$ and $\boldsymbol{T}$, of equal dimension to the number of communes multiplied by the number of instruments used in each matrix, which are applied to the hedonic regression and the income regression, respectively.

With regard to the income regression, this article considers the semilogarithmic functional form proposed by Mincer (1974) in which the variable to be explained is the hourly income logarithm of the individuals ($\ln Y_i$), as a function of their number of years of educational ($Ed_i$), years of work experience ($Exp_i$) and this term squared ($Exp_i^2$). Dummies were also used for sex ($S_i$) (1 man and 0 women) and economic field of work of the individuals ($A_i$), which considered nine groups, namely, hunting and forestry, mining, manufacturing, utilities (power, water and gas), construction, commerce, transportation, financial establishments, communal and social services, as well as unspecified activities:

$$E[\ln Y_i|X_i] = \delta_0 + \delta_1 Ed_i + \delta_2 Exp_i + \delta_2 Exp_i^2 + \delta_3 S_i + \sum_{m=1}^{9} \delta_{im} A_{im} \tag{22}$$

In summary, to obtain the hedonic prices, the system of equations (18) is estimated to reduce three sources of endogeneity. Given that different instruments are used for each equation whose number is greater than the regressors, the generalized method of moments (GMM) was used. The use of different instruments for each equation implies that the system of equations should be estimated with a full-information instrumental variables efficient estimator (FIVE) under conditional homocedasticity among the equations of the system (Hayashi, 2000).

### 4.2 Estimating the almost ideal demand system with censored data

Once obtained, the hedonic prices of housing characteristics are used to estimate the demand system. However, there are individuals within the sample that do not demand certain housing attributes, imposing a restriction in two areas. First, it is not possible to recover hedonic prices from expression (17) for these individuals given that the denominator is zero. Secondly, the sample is censored and does not consider individuals with zero consumption, which implies a source of selection bias. That is to say, the non-consumption of certain characteristics can be associated with latent variables that determine the decision to consume the housing characteristics or not (Bakhshoodeh, 2009).

To deal with this problem, Helen and Wessells (1990) used the two-step Heckman technique (1979). The first step estimates a probit regression $d_j = l'_j\lambda_j + e_j$ for every housing characteristic to determine the probability that a consumer demands this characteristic. The dependent variable $d_j$ of the regression is a binary variable that has the value of 1 if the consumer demands the characteristic and zero if he/she does not. The explanatory variables are latent variables that induce the consumer's decision to demand the characteristic or not. With this information the density function $\phi$ and the accumulated probability function $\Phi$ are calculated for consumers of each of the housing attributes. With this information, in turn, the inverse Mills ratio is determined as the density function ratio over 1 minus the accumulated probability function. The inverse Mills ratio for every attribute is used as an instrument to estimate the AIDS (Bakhshoodeh, 2009).

Shonkwiller and Yen (1999) suggested that using the two-step technique of Helen and Wessells (1990) produces the estimation of inconsistent coefficients of the demand system. The authors proposed incorporating the information from the first step by multiplying the explanatory variables of the demand system by $\Phi(l'_j\lambda_j)$, and density $\phi(l'_j\lambda_j)$ appears as an additional explanatory variable for each equation of the demand system, which is the solution we adopted for the second step in our research.

The estimation of the AIDS for the housing characteristics takes the following form:

$$w_i = \Phi(l'_j\lambda_j)\left\{\alpha_i + \sum_j \gamma_{ij} \ln p_j + \beta_i \ln\left(\frac{x}{P}\right)\right\} + \varphi_i \phi(l'_j\lambda_j) + \omega_i \qquad (23)$$

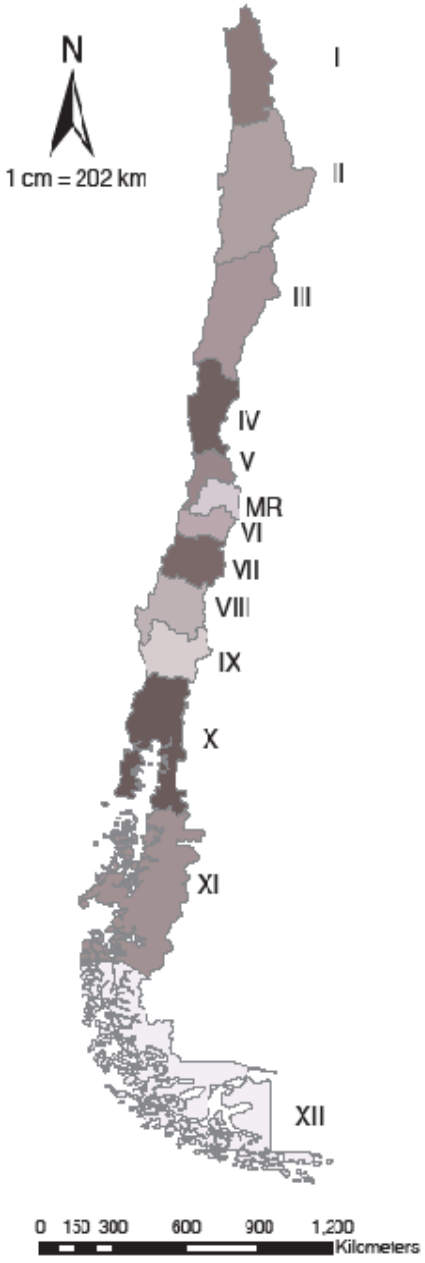
Where:

$w_i = \frac{p_j z_j}{x}$ is the proportion of expenditure on the characteristic $j$ over the total expenditure on housing for individual $i$.
$p_j$ is the hedonic price of housing characteristic $j$.
$x$ is the vector of the total expenditure on housing characteristics.
$\beta_i, \alpha_i, \gamma_{ij}, \varphi_i$ are the vectors of parameters to be estimated.
$\omega_i$ is the vector of errors of dimension n. It is assumed that $\omega_i \sim N(0, \sigma^2 I)$ for each characteristic $j$ and $E[\omega_i \omega_j] \neq 0$, with $i \neq j$.

The AIDS is estimated by a nonlinear seemingly unrelated regression using generalized non-linear least squares. This methodology is recommended when it is necessary to impose restrictions on the parameters and there are correlation errors among the equations of the system, as is the case of the demand regression of the AIDS.

Finally, with the estimated AIDS parameters, the expenditure function for the average individual in every commune on housing characteristics was recovered with the estimated AIDS parameters in order to construct the SHPI.

Figure 1: Map of Chile



## 5. Data

The data were obtained from the 2009 National Socioeconomic Characterization Survey (CASEN) undertaken by the Ministry of Planning (MIDEPLAN). CASEN is a representative household survey at the national, regional, communal, urban and rural levels. CASEN 2009 surveyed 246,924 persons in 71,460 households in 334 communities.

Several filters were applied to the original survey for the purpose of this research. The first of these filters was used to separate heads of households who report the rental price of their housing. The other filters were applied at the level of housing and homeowners. The final sample for the estimation of the system of equations (18) corresponds to 5,630 observations. The information for the instruments was obtained from the 2009 Municipal Information System.

The variables used for the regression of the hedonic prices are the housing rental price (dependent variable), number of bathrooms, number of bedrooms, the global housing quality index (GHQI)[5], dummies variables for the square meters of the housing units and dummies variables for each region in Chile. The set of instruments for this regression are population density and crowding index.

The variables used for the income regression are the hourly income of the worker (dependent variable), years of education, years of work experience and years of experience squared, dummies by sex and economic area of the work of individuals. The set of instruments for this regression are distance to the national capital from each commune, distance to the regional capital from each commune, communal population density, proportion of communal population economically active in the primary sector, proportion of the communal population economically active in the secondary sector and proportion of professionals in total communal employment.

Table 1: Description of the main variables to estimate the System of Equations

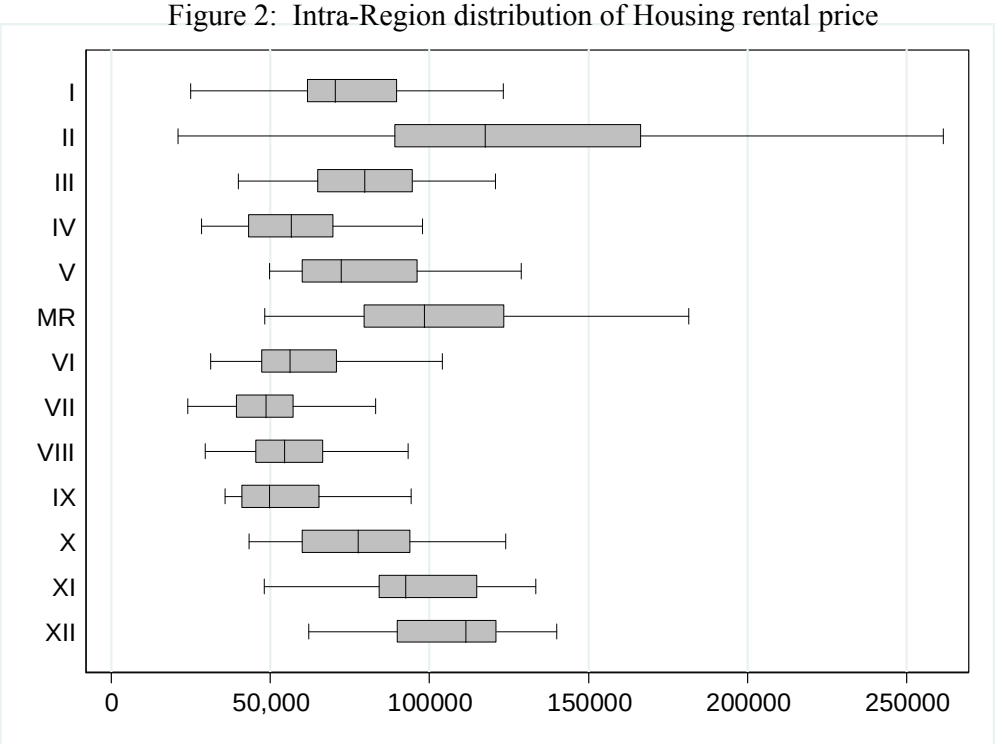
| Variables[a] | Description | Mean | SD | Min. | Max. |
|---|---|---|---|---|---|
| Housing rental price | Number (CLP) | 90,161 | 71,625 | 5,000 | 1,500,000 |
| Bedrooms | Number | 2.41 | 0.92 | 1 | 12 |
| Bathrooms | Number | 1.12 | 0.39 | 1 | 5 |
| GHQI | Between 0 and 1 | 0.88 | 0.1 | 0.29 | 0.99 |
| Square meters | Dummy 1= more than 100; 0 = 100 or less | 0.07 | 0.25 | 0 | 1 |
| Education | Years | 11.33 | 3.67 | 0 | 20 |
| Experience | Years | 28.75 | 12.52 | 4 | 81 |
| Sex | Dummy 1= man, 0 = women | 0.76 | 0.43 | 0 | 1 |

[a] Descriptive statistics for 5,630 observations.

As can be appreciated in Figure 2, there is a great dispersion of housing rental price both intra and inter regions, highlighting Second and Metropolitan regions, which present the largest rental price variation, showing a great heterogeneity among regions.

[5] Index constructed according to the methodology of the Latin American and Caribbean Demography Center (Centro Latinoamericano and Caribeño de Demografía - CELADE).

Figure 2: Intra-Region distribution of Housing rental price




Source: Authors

## 6. Results

### 6.1 Estimation of hedonic prices

To recover the hedonic prices of the housing characteristics requires estimating the system of equations (18), which incorporates both the instruments for the regression of hedonic prices and individual income equation.

Table 2: Coefficients estimated for the hedonic regression through the system of equations

| Variables | Coefficients | Variables | Coefficients |
|---|---|---|---|
| Ln (Bedrooms) | 0.17 | Ln (Income) | 0.21 |
| | (3.86) | | (6.01) |
| Ln (Bathrooms) | 0.59 | Education | 0.28 |
| | (12.81) | | (19.79) |
| Ln (GHQI) | 1.38 | Experience | 0.008 |
| | (1.26) | | (1.09) |
| Square meters | 0.16 | Experience$^2$ | 0.0004 |
| | (4.03) | | (3.26) |
| Constant | 9.87 | Sex | 0.27 |
| | (26.73) | | (5.95) |

z-values in brackets, 5,630 observations; Method of estimation: GMM-FIVE. Robust standard errors.

As can be appreciated in Table 2, except from GHQI, the coefficients of the hedonic regression are statistically significant and with the expected signs. They have a positive effect on the housing rental price. Housing with more square meters that also belong to individuals with higher incomes also presents higher rental prices. Recovery of the hedonic prices of housing prices were made using (17), which are used as regressors to estimate the AIDS. As

mentioned in the methodology section, some consumers do not demand certain housing attributes, because of which there is a potential selection bias which leads to estimating the AIDS as a censored model to obtain consistent coefficients.

### 6.2 Estimation of the almost ideal demand system

The estimation of the AIDS with censored data was carried out using expression (23). Table 3 shows the results of the estimation using the correction by bias with censored data and without this correction.

Table 3: Coefficients estimated for the almost ideal demand system

| Variables | Uncensored Coefficients | Censored Coefficients | Variables | Uncensored Coefficients | Censored Coefficients |
|---|---|---|---|---|---|
| $\alpha_1$ | 0.08 | 0.09 | $\gamma_{12}$ | -5.34e-11 | 0.0002 |
| | (1.4e+07) | (92.14) | | (-0.14) | (5.54) |
| $\alpha_2$ | 0.28 | 0.33 | $\gamma_{22}$ | -5.98e-10 | 0.002 |
| | (1.6e+07) | (53.11) | | (-0.45) | (4.13) |
| $\beta_1$ | -9.31e-09 | 0.001 | $\varphi_1$ | | 0.05 |
| | (-1.11) | (9.26) | | | (49.33) |
| $\beta_2$ | -2.87e-09 | 0.009 | $\varphi_2$ | | 0.19 |
| | (-1.07) | (8.93) | | | (67.7) |
| $\gamma_{11}$ | -1.58e-10 | 0.0001 | | | |
| | (-0.53) | (3.56) | | | |

z-values in brackets . Sample size: 5,630 observations distributed in 331 communes. Method of estimation: Seemingly unrelated regressions (SUR). Reference share $w_3$ (GHQI).

The results of estimating the AIDS under an uncensored model versus a censored model highlights the change in the statistical significance in the most of the estimated parameters. With respect to the estimation of the model with censored data, the positive signs of the coefficients $\beta_i$, indicate that housing characteristics are defined as luxury goods, that is, in the context of an increase in real expenditure on housing, the proportion spent on housing attributes increases. Parameters $\gamma_{ij}$ are all statistical significant, each $\gamma_{ij}$ represents the effect on budget share of an increase in housing attribute $j$ with $\frac{x}{P}$ held constant. The parameters that correct the selection bias of the sample $\varphi_i$ are all statistically significant.

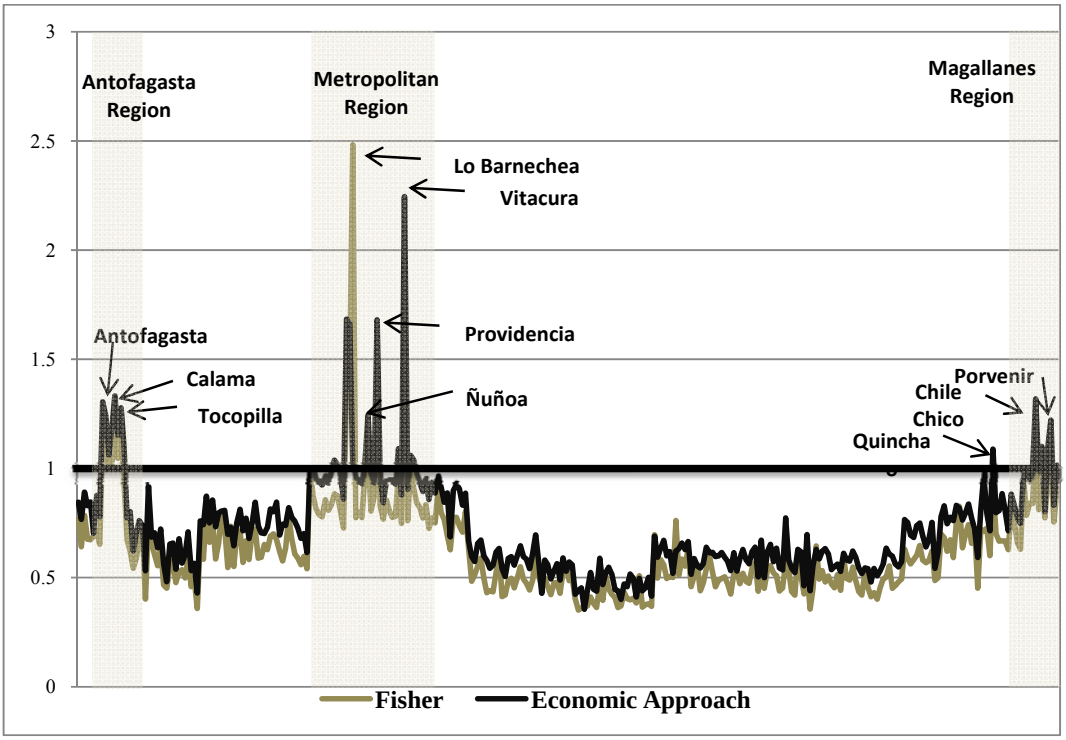
Once the AIDS coefficients have been estimated, the final step to construct the SHPI is to recover the expenditure function (5) to compute the housing expenditure of a consumer situated in the median[6] for each community, given a constant level of utility. The reference level of utility corresponds to the community of Santiago, the capital of the Metropolitan Region, that is, the SHPI indicates the compensation in the cost of moving from Santiago to any other community.

Figure 3 shows the SHPI constructed under the axiomatic approach using the Fisher index and the economic approach. The information is shown in accordance with the geographic configuration of Chile (see Figure 1), for convenience the communities are presented in a different way, that is to say, the communes from north to south are ordered along the left to right axis. The SHPI constructed under the economic approach are situated over the Fisher index for the majority of the Chilean communes. The communities with the highest housing

[6] The median is the most appropriate measurement of the central tendency for assymetric distributions, such as income given that there are few individuals that accumulate an important part of their income.

prices are in the Metropolitan Region, followed by the communities of Calama and Antofagasta in the Second Region. It is worth noting that the Metropolitan Region has the highest concentration of population in Chile, while the Second Region has the fifth largest population among the regions in Chile according to the 2002 Census.

Figure 3: SHPI for Chilean communes according to the axiomatic approach versus the economic approach.




Source: Authors

To determine if there are significant differences between the SHPI calculated using the economic and the axiomatic approaches, a test was performed on the mean differences between the two indices. The results reject the null hypothesis of equality of means between the two indices. Although the index constructed under the economic approach was situated very close to the axiomatic approach, the differences were statistically significant. The axiomatic approach tends to underestimate the SHPI and there are substantial differences in housing prices among communes within the same region, revealing heterogeneity in the configuration of housing markets. Finally, given that the SHPI constructed under the economic approach is based on economic theory, it has advantages over the axiomatic approach whenever the AIDS information allows evaluating consumer welfare in the context of changes in housing policies or urban planning.

## 7. Conclusions

This paper proposes a spatial index of housing prices for the communes of Chile using the economic approach. Bias in estimating hedonic prices was reduced by using a system of equations and instruments for the regression of hedonic prices and the incomes of individuals. Expenditure was constructed for the consumer situated in the median of the 325 communities, incorporating into the analysis the substitution made among housing attributes. The estimation

of the almost ideal demand system considered correction by selection bias, that is, a censored system, given that not all the individuals in the sample demand the housing characteristics.

Three relevant results emerge from this work: first, there are important differences in the expenditure of a consumer in Santiago to access the same level of welfare in another community in the country. This expenditure is lower in most of the communities, but in the Second and the Metropolitan Regions a positive compensation in spending is required to reach the same level of welfare. Secondly, the spatial heterogeneity of the housing market is reflected. The fact that there are differences in housing prices within a region indicates that the consumer must modify costs to access the same level of welfare in communities within the same region of residence. This result reinforces the idea that aggregating the communities to obtain a regional measurement of price indices is not adequate, given the high degree of spatial heterogeneity. Thirdly, the axiomatic approach tends to underestimate the SHPI for a major proportion of the communities in the sample. This index falls below the index constructed under the economic approach, showing the substitution bias of the axiomatic approach for not considering the substitution among attributes by the consumer given the different price vectors in each spatial unit.

Finally, the utility of having an economically-based instrument to construct the SHPI provides coherence in the analysis of consumer behavior by incorporating the ability to substitute goods when there are differences in their relative prices. The SHPI allows evaluating the impact of policies in terms of changes in the welfare of the consumer in the context of differences in the quality of housing or housing attributes, such as the change in housing policy since 2007 oriented to the quality instead of the quantity of housing units. This economic measurement can be incorporated in the analysis of commuting among communes or regions, adding the dimension of the cost of living to the variables that induce the decision of workers to work in communes or regions different from they live. Future research could be oriented to analyze the spatial distribution of the SHPI to contrast the hypothesis that there is a certain spatial segregation of high-priced housing, and at the same time to understand the variables that affect the configuration of the housing market.

**References**


Angrist, J. & Pischke, J., (2009). Mostly Harmless Econometrics. Princeton University Press.

Bowden, R., (1992). Competitive selection and market data: The mixed index problem. *Review of Economic Studies*. 59,625-633.

Bakhshoodeh, M., (2009). Impacts of world prices to domestic rice markets in rural Iran. *Food Policy*.35.12-19.

Bransington, D., & Hite, D., (2008). A mixed index approach to identifying hedonic price models. *Regional Science and Urban Economics*. 38, 271-284.

Breuer, C., (2007), Cost-of-living indexes for Germany. Working Papers.

Combes, P., Duranton, G., & Gobillon, L., (2008). Spatial wage disparities: Sorting matters¡. *Journal of Urban Economics*. 63.723-742.

Deaton, A., & Muelbauer, J., (1980). An Almost Ideal Demand System. *American Economic Review.* 312-326

Diewert, WE. (1978). Superlative Index Numbers and Consistency in Aggregation. *Econometrica.* 46,4, 883-900.

Diewert, WE. (1995). Axiomatic and Economic Approaches to Elementary Price Index.NBER Working Papers 5104. National Bureau of Economic Research, Inc.

Figueroa, E., & G. Lever (1992), Determinantes del precio de la vivienda en Santiago: una estimación hedónica. *Estudios de Economía*. 19(1), 5-31.

Glaeser, E., Kolko, J., & Saiz, A., (2001).Consumer city. *Journal of Economic Geography*.1, 27-50.
Halvorsen, R., & Pollakowski, H., (1981). Choice of Functional Form for Hedonic Price Equations. *Journal of Urban Economics*, 10, 37-49.
Hayashi, F. (2000). Econometrics. Princeton University Press.
Heckman, J., (1979), Sample selection bias as a specification error. *Econometrica.* 47, 153-162.
Helen, D., & Wessells, C., (1990). Demand System Estimation with Microdata: A Censored Regression Approach. *Journal of Business & Economic Statistcs.*Vol 8, 3, pp.365-371.
Koo, J., Phillips, K., & Sigalla.F (2000). Measuring Regional Cost of Living. *Journal of Business & Economic Statistics*. 18, N°1, 127-136.
Konus, A., (1924). The Problem of the True Index of The Cost of Living. *Econométrica,* 7. N°1, 10-29.
Lira, R., (1978), Precios implícitos de características de vivienda en Santiago. *Cuadernos de Economía*.15(44),67-90.
Malpezzi, S., (2002), Hedonic Prices Models: A Selective and Applied Review. *Working Paper.*
Marshall, A. (1920). Principles of Economics. Macmillan, London.
Mincer, J., (1974). Schooling, Experience and Earnings, Columbia Press University.
Paredes, D., (2009), A methodology to compute regional housing price index using matching estimator methods. *The Annals of Regional Science*.doi:10.1007/s00168-009-0346-z.
Paredes, D., & Aroca, P., (2008), Metodología para estimar un índice de regional de costo de vivienda en Chile. *Cuadernos de Economía*, 45, 129-143.
Parson, G., (1986), An Almost Ideal Demand System for Housing Attributes. *Southern Economic Journal*, 53, N°2, 347-363.
Pasha, H., & Butt, M., (1996). Demand Attributes in Developing Countries: A Study of Pakistan. *Urban Studies,* 33, N° 7, 1141-1154.
Pollak., R., (1980), Group-Cost-of-Living Indexes. *The American Economic Review.*70, N°2, 273-278.
Rosen, S., (1974). Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition. *The Journal of Political Economy*. 82. N°1, 34-55.
Shonkwiller, J., & Yen, S., (1990). Two-Step Estimation of a Censored System of Equation. *American Journal of Agricultural Economics*.81,N°4, pp.972-982.
Timmins, C.,(2006). Estimating spatial differences in the Brazilian cost of living with household location choices. *Journal of Development Economics*. 80, 59-83.
Wooldridge, J., (2002). Econometric Analysis of Cross Section and Panel Data. The MIT Press.