Residential property price index: preliminary results for Chile[1]

Camilo Vio[2]

1. Introduction

Monitoring the dynamic of residential prices, even though is not trivial, has a relevant importance especially in turbulent periods. Recent financial crises, triggered in the US due to a plunge on house prices, have highlighted the importance of monitoring house prices dynamics. In emerging economies, there is a lack of official house price indexes developed by public institutions. Moreover, in Latin America house prices data are only available for some countries, and even when available, time series are usually of short span, with coverage often limited to large metropolitan areas (Cubeddu et al, 2012). Notwithstanding, there are some studies developed by private consulting, whose purpose depends on their customers' needs, and due to the high costs of information gathering, are usually biased to a particular sample of properties.

Particularly, in Chile some studies displayed to estimate house prices were developed by Morandé (1992) using hedonic prices for Ñuñoa district,[3] between 1975 and 1989, Bergoeing et. al. (2002) who expand the prior mentioned work until 1998, which finally Desormeaux and Piguillem (2003) continued until 2003. Later, using two methodologies (repeat sales and hedonic prices), Parrado, Cox and Fuenzalida (2009) estimated house price index between 2001 and 2007, based on information from the Property Register Office (Conservador de Bienes Raíces). Moreover, using hedonic models, both Figueroa and Lever (1992) and Sagner (2009) estimate what factors determine house prices for Santiago.

This document presents alternative and preliminary estimations of the Chilean Residential Property Price Index (RPPI) employing a novel dataset that comes from the Internal Revenue Service, and includes all recorded transactions made within years 2001 and 2011. This dataset, considering its national coverage, allows exploring behaviors of property prices within cities, in order to make a more comprehensive monitoring of the real estate market and establishing connections with past events, due to the length of the data. These estimations include methodologies based on mix adjustment stratification, hedonic model and repeat sales.[4,5] The results presented in this paper are still preliminary and do not constitute an official statistic of the Central Bank of Chile.

This document goes through the following structure. Section 2 raises some international evidence on publication of house prices, section 3 describes the data used for the estimations. Section 4 makes a methodology description, highlighting the advantages and disadvantages of

---

1 All figures presented below, do not represent official statistics from the Central Bank of Chile.

2 Central Bank of Chile. cvio@bcentral.cl. I acknowledge the comments and suggestions of Luis Opazo, Érika Barrera, Romina Villarroel and Carolina Del Campo. The views and conclusions presented are exclusively those of the author and do not necessarily reflect the position of the Central Bank of Chile or of the Board members.

3 Ñuñoa is located in the west area of Santiago city.

4 Some countries that use mix adjustment methodology are, among others, Australia, Greece and Ireland, while the repeat sales method is used in US.

5 Other methodologies, used to compute Real Estate Index are weighted average, and sale-price appraisal ratio (SPAR).

each method used. Section 5 presents the main results of both methodologies and finally, section 6 exhibits an implementation of a micro analysis across districts.

2. International evidence

The process of constructing a RPPI is not simple. The methodology used can be constraint in large part by the nature of the data available, and on a regular and timely basis (Eurostat, 2011). Moreover, due to the cost of producing the data required, some indexes use part of the transactions – i.e constraint to the main cities – as information for its compute. Furthermore, this index construction is based on the assumption that the information will remain constant in the future. But this cannot be ensured, because changes in the information due to implementation or costs might be run by the sources. This last issue is especially important, because some methodologies demand very detailed data.

Among countries, there is a wide spectrum of methodologies underlying RPPIs. These differ in scopes such as quality and level of detail of the data. Also, the data sources and methodologies are not always well documented, which affect the possibility to make international comparison analyzes. Indeed, Eurostat (2011) states that "[c]omparability between indices can be very limited as a result of the different data sources (…) – mortgage versus cash purchases; urban versus rural prices; the prices of old properties versus new properties; valuation versus advertised price versus initial offer price versus final transaction price. The net result is that the published indices can, in practice, measure very different aspects of the price development in the housing markets."

Most of the selected countries use a hedonic approach to calculate a RPPI. A smaller group uses the most simple methodologies – average and common types –, three of them use a mix adjustment method, and only United States use the repeat sales method (table 1). As was mentioned above, the selection of each country is based on the data available to compute the index and also on the idiosyncrasy of each market.

Table 1

International House Price Index (1)

Mix adjustment	Hedonic	Repeat Sales	SPAR	Average (2)	Common types (3)
Australia	Austria	United States	Netherlands	Belgium	Canada
Greece	Finland		New Zealand	Estonia	Czech Republic
Ireland	France			Russia	Denmark
	Norway			Slovak Republic	Sweden
	Poland			Switzerland	
	Slovenia				
	Spain				
	United Kingdom				

(1) Selected index for countries with more than one, are publish by government agencies. For more details, see Annex 1.

(2) Some countries use a weighted average.

(3) Follows the dynamic of a specific type dwelling's price.

Source: Author elaboration based on information from Silver (2012).

3. Data description

The data used to compute the RPPI is provided by the Internal Revenue Service (Servicio de Impuestos Internos). The information comes from two sources. First, the registry of trades, which has all the information of the properties transaction between two agents, excluded from inheritances and subsidies. Second, information used in Chile to characterize properties, with all their features, in order to charge taxes. These two sets of data are merged in one database, which contain the records of properties transactions – price, date of the trade, among others – and the characteristics of properties – geographic location, construction type, land and structure area, etc. The whole database contains approximately 1,683,000 trades accomplished between 2001 and 2011 whose primary use is residential.[6] After submitting the data to a cleaning process by removing outliers or trades with no price information, over 1,518,000 observations were obtained.

Data cleaning filters

In spite of the high quality of the information, both for its representativeness – it already contains all the trades made – and for its descriptive level, an important impediment is that part of it comes from an administrative form that is afterwards transferred into a digital platform. Given that, there are few chances of recognizing mistakes made during the transfer process.

Filters included in the final dataset, after identifying the transactions of residential properties are the following:

i) Duplicated observations: All duplicated observations are eliminated.

ii) Non valid date, or out of range: Observations whose date of trade does not correspond to the range of information (2001–2011) are eliminated.

iii) Building surface: Observations with a structure size less than 20 square meters are eliminated. This is below the first percentile of the distribution.

iv) Atypical sales price-appraisal ratio: All the observations whose sale price-appraisal ratio is atypical are eliminated. To define its boundaries, box-plot method was used.[7]

Considering the filters described above, the amount of deleted observations reaches a 10% of trades whose primary use is residential.

4. Methodology description

There is a wide spectrum of methodologies to compute a RPPI, which differ depending on the quality of information used to build it, and the main objectives desired to achieve. Furthermore, there are idiosyncratic factors that affect the relevant variables used to build an appropriate property price index. Given that, it is possible to find a wide amount of literature where some empirical comparison is developed for a specific country or city.[8] Considering

---

[6] Properties whose primary use is residential, represent around 75% of trades between 2001–2011.

[7] The Box-plot method considers as a lower bound the percentile 25 less 1.5 times the interquartile range, and as an upper bound the percentile 75 plus 1.5 times the interquartile range.

[8] Hansen (2006) performs a comparison among hedonic price measures and repeat sales for the Australian market, using three of the larger cities, Sydney, Melbourne and Brisbane. The same comparison is performed by Shimizu et al. (2010) applied to residential property prices in Tokio. Haan et al. (2007) generate a revision of the SPAR method, comparing it with the repeat sales for Netherlands.

this, the following document will focus on repeat sales, hedonic and mix-adjustment methodologies.

Repeat sales

Repeat sales method, proposed by Bailey et. al (1963), provides an estimation method based on changes of property prices sold more than once. It is usually assumed that the features of every dwelling does not change between sales, and hence, the dynamic of the price can be expressed as:

$$p_{it} - p_{it} = \sum_{t=1}^{T} G_{it} a_t + h_{it} , \quad (1)$$

where G_{it} is a temporal dummy that takes the value 1 when the resale was performed, –1 when the previous sale happens, and zero in any other case. h_{it} is the error term.

The main advantage of this method is its high accuracy in terms of controlling the quality of dwellings. Indeed, given that it uses the same property to compute the changes in price, qualitative characteristics that might not be observable or measured, would allow to control the bias associated with compositional changes. Another advantage is that it requires less information than other methods, given that it is possible to compute it using only price and time data.

However, one of its disadvantages is that, given the assumption that features of the dwelling are constant in the sample, any investment made in order to improve the quality of the dwelling, or depreciation due to neglect in the maintenance of the property, is not controlled. A good method to manage this, is to use sub-samples where the quality is assumed as relatively constant. The problem of this solution is that if the sub-sample is too small, changes in prices might not be indicative of the total sample of repeat sales.

Other control methods are proposed by Diewert (2007), and Paredes and Aroca (2008). The first one incorporates the depreciation based on the length between the year of building and the date of sale, while the second one uses a matching method to pick commensurable dwellings according to the closest neighborhood's standard, in order to evaluate differences of prices among geographic regions. Even though this last method is used to compute an hedonic price index, it would be interesting to use it in the composition of repeat sales – i.e. using "similar" dwellings according to certain attributes in order to generate price differences. There is, furthermore, extensive literature on other variations of the estimators through repeat sales models, tending to reduce disadvantages and biases that might appear when using a general model, and to compare price dynamics against other methodologies.

Mix-adjustment

The mix-adjustment or stratification method is based on the measure of price variations of different types of dwellings, classifying the sample in groups of properties according to certain characteristics such as price, location, size, etc. Measures, such as mean or median, are calculated for every group before being combined in the index construction. The construction of its geometric adjustment, is based on the following expression:

$$MP_t = \prod_{i=1}^{n} P_{it}^{w_i} , \quad (2)$$

where MP_t is the adjusted price in *t*, w_i is the weight associated to the *i* group, (which can be computed through the participation of the group in total sales, *turnover* or participation of the group in the stock of dwellings), and P_{it} is the mean or median of group *I*'s dwelling price

at time *t*. Finally, *n* is the total number of groups. This method is simple and takes into account differences among property features.

The data specificity level of this method is not high – compared with hedonic measures, for example – only enough to identify and classify the properties into the different groups, and the effectiveness of its measure depends on the stratification used. Generally, this method only controls for compositional changes across the dimensions defined for each group, so one of the assumptions made is that quality changes occur for the whole group, and within each group they remain identical. In turn, this generates a restriction to the effect of controlling quality changes within each group.

Hedonic

Hedonic measures aim to control for compositional and quality changes through econometric techniques. This model, developed by Rosen (1974) is widely used by researchers, however it requires data with a high level of detail on specific characteristics of dwellings. Usually, a log-linear specification of the model is used, and the parameters of the hedonic model can be estimated by OLS regression in the sample data of each time period separately.[9] Thus, the estimating equation becomes as shown:

$$p_i^t = b_0 + \sum_{t=1}^{T} d^t D_i^t + \sum_{k=1}^{K} b_k z_{ik}^t + e_i^t, \qquad (3)$$

where p_i^t is the logarithmic of the price of dwelling i, sold at time t, D_i^t is a time dummy which has the value 1 if the observation comes from period τ, and 0 otherwise, z_{ik}^t represents the k characteristics of dwelling i, sold at time t. Vector b contains the implicit values of each characteristic.

After performing the estimation, the price change between the base period and each comparison period t, is represented by the exponential of each estimated dummy coefficient.

According to Hansen (2006), the main advantage of the hedonic model is that it provides direct estimators of pure prices, and can control for changes in composition and quality of dwellings. However, this measure has some limitations: (i) There could be some omitted variables – whether there is no information on them on the dataset, or they are unobservable – which could cause biased changes in pure price estimators. (ii) In its more general specification, they assume constant characteristic parameters through time.[10] (iii) For the purpose of publishing the level of prices, if the time series is extended to T+1, and new data is added, the characteristics of coefficients will change (Eurostat, 2011).

5. Results

The decision of what methodology is better to get a proper RPPI is not simple. There are some details that have to be considered, such as the availability of data, if the time series have to be corrected due to new past information, among others. In order to explore the advantages and disadvantages of the different choices, an index was computed based on

[9] For more details, see Eurostat (2011).

[10] However, there are some advanced estimations, which consider time-varying parameters. For example, Rambaldi and Prasada Rao (2010) use a time-varying hedonic model to estimate house price index for Brisbane.

three methodologies – repeat sales, mix adjustment and hedonic. On one side, the advantage of mix adjustment approach "is that it is easy to compute because it is based on simple medians from stratification and uses data that are readily available from most housing transactions database" (Prasad and Richards, 2008). On the other side, the disadvantage of repeat sales and hedonic method is that time series must be corrected in case new past information has to be considered or the value of each characteristic has changed.

For the Residential Property Price Index computed for Chile, based in repeat sales methodology, properties with more than one transaction were used in the sample. The number of dwellings that have more than one transaction between 2001 and 2011 is near 340,000. However, the compute of the index was developed since 2004, in order to avoid undesirable effects of a lower population in the initial period of the sample. Furthermore, in order to exclude outliers, observations resold too rapidly – those whose distance between two sales is less than 180 days – were eliminated.[11]

In the Residential Property Price Index computed through the mix-adjustment methodology, all properties were included, both new and second-hand sales. Eurostat (2011) suggests that the stratification has to be made in order to control for compositional changes of the properties. For purposes of this study, the appraisal value of the property was used to stratify properties into five groups, given by the five quintiles of the distribution of appraisal values for the whole sample.[12] Then, the real price over the constructed area measured in squared meters was calculated for each transaction, and the median value of this index within each strata is obtained. After that, the index is collapsed using the sum of squared meters of constructed area traded on each quarter, as weight for each quintile. Finally, in order to chain the index, the Fisher method was used.[13]

In the computation of the hedonic model – described in (3) –, some categorical variables were created to control for differences in quality. For each category a dummy variable was created, which takes the value 1 if the property belongs to that category for that variable, or zero if otherwise. Considered variables were: Age of the building, separated on nine categories ranged between zero and at least 40 years; and quality of the building, separated in five categories given by the appraisal questionnaire. Due to the lack of information about number of bedrooms and bathrooms, the quality variable can be viewed as a proxy of those, as it includes some information about the material of construction, location, etc. Other quality control has to do with the constructed living space. This variable was included in two terms, one linear and the other quadratic, controlling for nonlinear effects.

The results for the three methodologies are shown in the figure 1, panel A. Also, annual growth is shown in panel B. Correlation between mix adjustment annual growth and hedonic annual growth is 0.73, while between mix adjustment annual growth and repeat sales annual growth is 0.26 and between repeat sales and hedonic, is 0.27. The last two correlations increase to 0.33 and 0.45 respectively, starting in 2007.

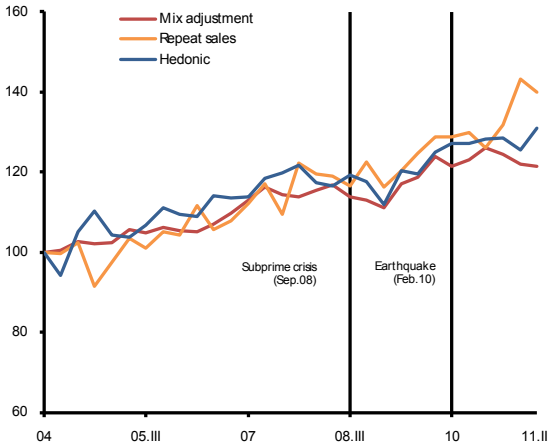
[11] In the compute of Case-Shiller index, they use the same criteria to filter for unusual transactions. S&P (2009).

[12] Appraisal of every property is developed by the Internal Revenue Service, filling a complete questionnaire where detailed characteristics are reported. Given that a unique document is completed for each property, and the same parameters are considered to value it, it could be considered as an objective measure.

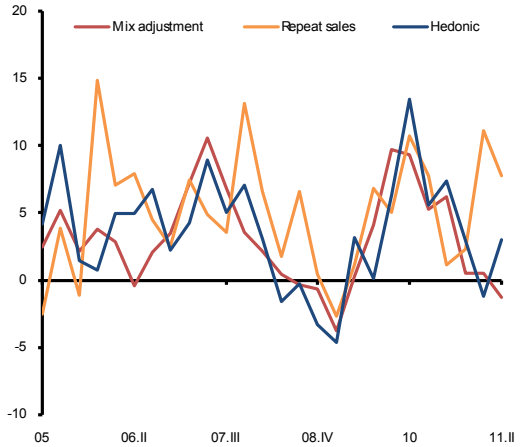
[13] Results does not change much when using the sum of squared meters sold the previous year as weights, and also when using Laspeyres method to chain the index.

Figure 1

Panel A: Residential property price index
(index, base mar.04=100)



Panel B: Annual growth
(percentage)



Source: Author's calculation.

6. Monitoring recent developments in residential market

Based on the three methodologies described above, persistent and generalized growth in house prices across the country has not been observed.[14] However, in order to get a more general view, some particular effects still have to be analyzed. First, due to the geographical distribution and economic activities of the principal cities in Chile, markets have some characteristics that throughout time remain heterogeneous.[15] Second, on February 2010 the occurrence of an earthquake affected significantly three regions of the country, while other areas were softly or not affected at all.

Furthermore, evidence on US housing shows that dynamics of house prices among MSAs are heterogeneous, not only for annual growth but also for the period where a structural change on growth rates is found (Ferreira and Gyourko, 2011). Using the same methodology as Ferreira and Gyourko (2011), a micro analysis of residential property prices dynamics among districts[16] was developed. A hedonic index for each city was computed, using some characteristics of the dwelling – size of the structure, quality and age of the building.[17] The implementation was made for 47 districts in Chile – those with more than 9.000 sales in the sample – from a universe of 346, which represents around 70% of sales between 2001 and 2011 (more than a million of observations).

---

[14] Cubeddu et al. (2012) make an analysis of vulnerabilities for Latin America. They found few signs of misalignment in the mortgage and the real estate sector, and house prices in most markets are near equilibrium levels.

[15] For example, the north area of the country is very intensive on mining industry, and the south area has more activity on agriculture. Therefore, during a boom on a specific industry, impact on the region which is more dependent on that industry would increase.

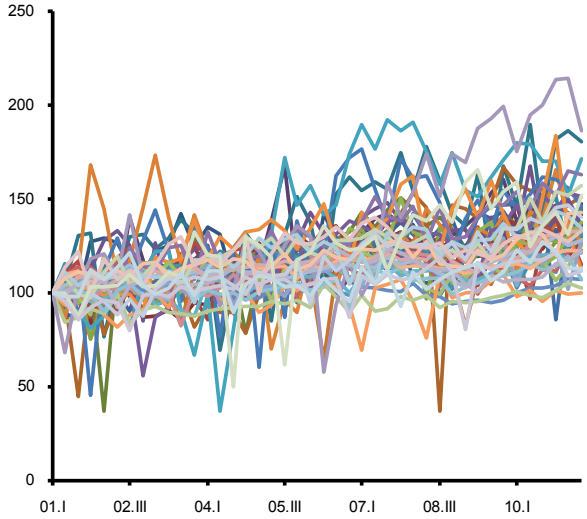
[16] A district is the basic and minor politic-administrative division in Chile. Might be understanding as a group of neighborhoods, and generally, the biggest cities have more than one district (i.e., in Santiago, there are more than 30 districts).

[17] The reason to compute a hedonic index is that within each district there are a short number of transactions, therefore it is not possible to compute an index based on repeat sales or mix adjustment method.

Results show that even though there is some heterogeneity in price dynamics across cities, persistent and generalized growth is absent (figure 2, panel A). This differs on what Ferreira and Gyourko (2011) found for metropolitan areas in US before subprime crisis, where many of them presented persistent growth beginning at different time periods.

Figure 2

Residential property price index across location
(index, base mar.04=100)



Source: Author calculations based on Ferreira and Gyourko (2011).

Furthermore, an analysis in order to obtain the timing of the start of housing boom was also developed. This analysis is based on searches of structural changes on annual growth index prices. Therefore, for each one of the 47 districts, a dummy regression was estimated, considering a dummy variable to account for the magnitude of the change in the growth, based on the following equation:[18]

$$PG_{d,t} = a_d + D_d 1[q_{d,t} \; ^3 \; q^*_{d,t}] + e_{d,t}, \qquad (4)$$

where $PG_{d,t}$ is the annual housing price growth in each district, $D_d 1[q_{d,t} \; ^3 \; q^*_{d,t}]$ is a set of relative quarter dummies, which take the value 1 for $t>t^*$, a_d is a district fixed effect, and $e_{d,t}$ is the error term. The analysis was developed between the first quarter of 2006 and the third quarter of 2009. In order to obtain the structural change, for each district if the dummy coefficient of the highest R-squared of the regression is significant, then there is a structural change on growth of that district house prices.

Results show that there are five districts with structural changes, which occurred on the third quarter of 2009 (table 2). Therefore, while there are no signs of a generalized boom, it is needed to continue with the micro analysis in order to account for further booms, and to understand which are the factors influencing those growths.

[18] This is the same equation used by Ferreira and Gyourko (2011).

Table 2

Real House Price Growth (*)

(percent)

	02.I–09.II	**09.III–11.II**
Santiago	–0.6	5.1
Ñuñoa	0.8	4.0
Las Condes	0.5	8.5
La Reina	2.8	10.2
La Florida	1.6	6.2
Vitacura	1.3	7.4

(*) Average annual growth.

Source: Author calculation based on Ferreira and Gyourko (2011).

Annex 1

Selected International House Price Index

Country	Institution	Periodicity (*)
Australia	Australian Bureau of Statistics	Quarterly
Austria	Oesterreichische Nationalbank	Quarterly
Belgium	Statistics Belgium & SPF Economie	Quarterly
Canada	Statistics Canada	Monthly
Czech Republic	Czech Statistical Office	Quarterly
Denmark	Statistics Denmark	Quarterly
Estonia	Statistics Estonia	Quarterly
Finland	Statistics Finland	Quarterly
France	National Institute of Statistics and Economic Research (INSEE)	Quarterly
Greece	Bank of Greece	Quarterly
Ireland	Department of Environment, Heritage, and Local Government. Central Statistics Office (CSO)	Quarterly
Netherlands	Central Bureau voor de Statistiek (CBS)	Monthly
New Zealand	Reserve Bank of New Zealand	Quarterly
Norway	Statistics Norway	Quarterly
Poland	National Bank of Poland	Quarterly
Russia	Federal State Statistics Service	Quarterly
Slovak Republic	National Bank of Slovakia	Quarterly
Slovenia	Statistical Office of the Republic of Slovenia	Quarterly
Spain	Instituto Nacional de Estadísticas (INE)	Quarterly
Sweden	Statistics Sweden	Quarterly
Switzerland	Swiss National Bank	Quarterly
United Kingdom	Nationwide	Monthly
United States	Federal Housing Finance Agency (FHFA)	Monthly

(*) The most frequently reported.

Source: Author own elaboration based on information from Silver (2012).

References

Bailey, MJ, RF. Ruth and HO. Nourse (1963). "A Regression Method for Real Estate Price Index Construction." Journal of the American Statistical Association, 58, pp 933–942.

Bergoeing, R., F. Morandé and R. Soto (2002). "Asset Prices in Chile: Facts and Fads." In *Banking, Financial Integration and International Crises*, edited by L. Hernández and K. Schmidt-Hebbel. Central Bank of Chile.

Cubeddu, L., C. Tobar and E. Tsounta (2012). "Latin America: Vulnerabilities Under Construction?" IMF Working Paper N°12/193.

Desormeaux, D. and F. Piguillem (2003). "Precios Hedónicos e Índices de Precios de Vivienda." Documento de trabajo N°12. Cámara Chilena de la Construcción.

Diewert, W. E. (2007). "The Paris OECD-IMF Workshop on Real Estate Indexes: Conclusions and Future Directions." The University of British Columbia, Department of Economics, Discusión Paper 07–01.

Eurostat (2011). Handbook on Residential Property Price Indices. European Comission. Version 3.0. January.

Ferreira, F. and J. Gyourko (2011). "Anatomy of the Beginning of the Housing Boom: U.S. Neighborhoods and Metropolitan." NBER Working Paper No. 17374. August.

Figueroa, E. and G. Lever (1992). "Determinantes del Precio de Vivienda en Santiago: Una Estimación Hedónica." Estudios de Economía 19(1): 67–84.

Haan, J de, E van der Wal, D ter Steege and P. de Vries (2007). "The Measurement of House Prices: A Review of the SPAR Method." Statistics Netherlands.

Hansen, J (2006). "Australian House Prices: A Comparison of Hedonic and Repeat-Sales Measures." Research Discussion Paper. Reserve Bank of Australia.

IMF (2004). World Economic Outlook. April.

Morandé, F. (1992). "The Dynamics of Real Asset Prices, The Real Exchange Rate, Trade Reforms, and Foreign Capital Inflows. Chile 1976–1989." *Journal of Development Economics* 39: 111–39.

Paredes, D. and P. Aroca (2008). "Metodología para Estimar un Índice Regional de Costo de Vivienda en Chile." Cuadernos de Economía 45, pp 129–143. May.

Parrado, E., P. Cox and M. Fuenzalida (2009). "Evolución de los Precios de Viviendas en Chile." *Revista de Economía Chilena* 12(1): 51–68.

Prasad, N. and A. Richards (2008). "Improving Median Housing Price Indexes through Stratification." Journal of Real Estate Research 30(1): 45–71.

Rambaldi, A. and D.S. Prasada Rao (2010). "The Effect of Hedonic Modelling and Index Weights on Hedonic Imputation Indices." Mimeo.

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

Sagner, A. (2009). "Determinantes del Precio de Viviendas en Chile." Documento de Trabajo N°549. Central Bank of Chile.

Shimizu, C., K Nishimura and T. Watanabe (2010). "House Prices in Tokio: A Comparison of Repeat-Sales and Hedonic Measures." Research Center for Price Dynamics Discussion Paper No.52.

Silver, M. (2012). "Why House Price Indexes Differ: Measurement and Analysis." IMF Working Paper N°12/125.

S&P (2009). "S&P Case-Shiller Home Price Indices: Index Methodology." November.