Introduction to Session VI

Marc Prud'Homme

Although hedonics is not always a perfect or desired methodology for producing a quality-adjusted price series, it is nevertheless the only practical solution under certain conditions. Products for which quality-adjustments should be applied to price series are numerous and are growing in number as the outputs of the economy becomes more complex. One only has to look at computers and automobiles, where rapid technological change has meant that the conventional way of measuring "pure" price change (matched-model or -sample) is not the best option and when applied results in biased numbers.

Housing and real estate are another good example of a product for which, although the dynamics may be different from, for example, computers, the underlying problem is the same: from one period to the next, an identical version is impossible to find and therefore matched sampling cannot be applied. Given the relative importance of housing in household budgets and its importance in most people's wealth portfolios, the use of hedonics, which is more resource intensive than some other approaches, can easily be justified. Compared to its alternatives, my own research shows that long-term price trends and short-term cyclical behaviour of prices are best measured using hedonics. Another benefit of using hedonics for measuring real estate prices is that you could actually attack two birds with one stone, meaning that you could use the same hedonic function and data with only the slightest of tweaking in order to do spatial (or geographical) price comparisons, which have an analytical value with regard to real estate.

I would like to mention a few words about what is going on at Statistics Canada with regard to real estate prices. We at Statistics Canada are very interested in this conference and encouraged by the papers being presented, and thank the organisers for inviting us. At present, if I have to describe our current situation with regard to real estate prices, I would have to say that we are simply house-poor. For residential houses, we have the New House Price Index (NHPI) produced by Statistics Canada, which is a monthly series that measures changes over time of the contractor's selling prices of new residential houses, where detailed specifications pertaining to each house remain the same between two consecutive periods.

The survey also collects contractors' estimates of the current value of the land from which a land price index is obtained. The resulting series is used for calculating some of the components of the consumer price index for shelter, depreciation to be exact, and for deflating the value of the national housing stock in the national accounts.

My own hedonic studies, or I should say my own exploratory hedonic studies, using resale house prices drawn from Multiple Listing Services (MLS) data, which is available in Canada, seem to confirm that trends in the NHPI appear to track pretty well what is going on in the resale housing market.

Our general situation with regard to real estate prices is likely to change in the near future, however, and we might be expanding our series on house price indexes or our real estate price indexes for a number of reasons. The first reason being our reaction to the comments by the Governor of the Bank of Canada delivered at the Conference of European Statisticians about the need for better real estate price indexes. The second reason is the federal government's need for better house price indicators, which are used in our provincial equalisation payment program. And the third reason is because right now at Statistics Canada there seems to be a growing need for a more comprehensive inter-area or inter-city price comparison program, for which we do not have accurate measures of house prices that could be used for this purpose at the moment. In fact, it is Bohdan Schultz, our now retired index number guru, who said that a spatial price index without good shelter prices is like a meal without its main course.

I have said enough about our situation at Statistics Canada. Turning the discussion on hedonics, I thought I would probably provide some background information on hedonics by presenting some of the milestones in the development of hedonics. This can be helpful given that a lot was said yesterday about hedonics but I suspect not many people in the room are necessarily familiar with the subject.

A person named Haas conducted one of the first studies using hedonics in 1922, and the interesting fact – or tidbit – about this price index was that it was actually a real estate price index for an area around Minnesota.

Andrew Court in 1939, in one of the most cited works using hedonics, produced a hedonic price index for automobiles. It is in this study that the term “hedonics” applied in the context of a quality-adjusted price index using regression analysis was first coined.

Bailey, Muth, and Norris, in 1963, produced a regression method for real estate. In 1968, the US Bureau of Census started to produce a price index for single-family homes under construction, and it was certainly the first hedonic index produced within the context of a regular statistical program. It is still produced today. In 1991, the American Real Estate and Urban Economics Association devoted a whole number of their journal to real estate price indexes.

Hedonics has become quite popular. Yesterday on Google, if you typed in the search word “hedonic”, you would get 71,000 results. In 2000, according to Grimm and Landefeld, just about 17 percent of GDP final expenditures are deflated using hedonic price indexes. And today, we have two more papers to add to the growing collection of papers on hedonics.

Housing stock in Brazil: estimation based on a hedonic price model

Luís Otávio Reiff[1] and Ana Luiza Barbosa[2]

1. Introduction

As an important aspect for the level of wealth, social welfare and economic growth, the housing stock, and more specifically, home ownership are significant structural mechanisms in the reduction of poverty and social inequalities.[3] Nevertheless, the difficulties in measuring the prices of real capital such as housing are very well known. These difficulties stem, in part, from the heterogeneity of these non-standard assets and the infrequency of observed transactions with individual properties. Therefore, accurate measurement of real estate prices is important for both practical and theoretical purposes. In particular, we believe that our empirical findings regarding this subject will make a valuable contribution to the housing policy debate in Brazil, where the housing sector must perform efficiently to provide affordable homes, especially to lower income people.

Statistics on the Brazilian economy show that in 1999 rent expenditures reached 14% of GDP and 17% of household consumption (National Accounts, IBGE (2000)). Ipea's estimates indicate that in this same year housing investment corresponded to 28% of total investment and 5% of gross domestic product (Morandi (2002)). According to the National Household Survey (*Pesquisa Nacional de Amostragem de Domicílios* - PNAD) conducted by the Brazilian Institute of Geography and Statistics (*Instituto Brasileiro de Geografia e Estatística* - IBGE), 74% of households inhabited their own property in 1999. These figures show how meaningful housing investment is in the capital accumulation process of the Brazilian economy. Moreover, it suggests that housing is the main component of wealth for most Brazilian households.

This study is part of a broader project called "Estimates of the Stock of Capital and Wealth of Brazil (1970-99) and applications in the analysis of public and regional policies". The main objective of the latter is to estimate the value of the capital stock and wealth in Brazil according to: (i) categories (residential, non-residential, machinery and equipment, domestic and imported); (ii) productive sectors (industry, agriculture and infrastructure); (iii) property (government, government-owned enterprises and private companies and families); and (iv) location (states and municipalities).

The purpose of this study is to estimate the housing stock in Brazil from 1970 to 1999. Estimations are based on the value of the rent payments of rented residences and the imputed rents of owned proprieties. Based on the monthly rent paid and the physical and locational characteristics of the property, a hedonic function is estimated to serve as a base to impute the rent for all residences. The conversion of the rent into the price of the property is done indirectly using an average discount rate of 0.75% per month (9.38% per year), found in empirical works on the subject. The source of data for the census years (1970, 1980, and 1991) was the Demographic Census, and for other years the National Household Survey (PNADs).

This study is organised as follows: in Sections 2 and 3 we present the hedonic price model and the model to evaluate the housing capital stock, respectively. Section 4 describes the data used for the estimation and Section 5 presents the estimation. The results are discussed in Sections 6 and we sum up with concluding remarks in Section 7.

---

[1] From Universidade Federal do Rio de Janeiro - UFRJ, Department of Economics, Rio de Janeiro, Brasil.

[2] From Instituto de Pesquisa Econômica Aplicada - IPEA, Rio de Janeiro, Brasil.

[3] See Wolff (2000), Spilerman (2000), Neri et al (1999), among others. Other relevant aspects for the analysis are: taxation on capital gains or taxation on imputed rents of owned residences (Poterba (1992), Hendershott and White (2000)); the construction of price indices (Zabel (1999)); and the demographic determinants of housing prices (Green and Hendershott (1993)).

2. Hedonic price model

Hedonic price models are used to identify factors or influences on the price of goods based on the idea that price is based on both intrinsic characteristics and external factors. These models are most commonly applied to housing markets in which the price of housing is based on the physical characteristics of the house (size, appearance, features) and the surrounding neighbourhood (accessibility to schools and shopping, quality of other houses, availability of public services). Estimating hedonic prices makes it possible to identify the extent to which specific factors affect the housing price.

In our model, we assume that the value of the monthly rent (R) of a property has a stable relation with its price (P), $R = \alpha P$, where α is the opportunity cost of renting housing.[4] Therefore, hedonic pricing relates the housing price to its attributes as follows:

$$R = X\beta + Z\delta + \varepsilon, \tag{1}$$

where R is a vector of the rents of the good (housing, in our case); the explanatory variables correspond to the physical attributes (X) of the housing unit and the location attributes (Z), which is related to amenities, public and service infrastructure, construction quality, among others; the coefficients β and δ correspond to the implicit price of housing attributes, and ε is the stochastic residual.

The vector of physical attributes aggregates 19 variables, where 17 are dichotomic variables and the other two are polythomic variables. The vector of location attributes consists of three variables, where one is a continuous variable and the remainders (two) are dichotomic variables.

Regarding location variables, one method of modelling it is to change each location's intercept. In other words, when we include dummy variables, we handle all specifications of each location *i* related to a reference community. The other alternative to attain the location effect in the model is to apply a proxy variable. As Zabel (1999), in our model we also use a proxy variable, the median income of the location, along with two other location dummies.

We assume that rents (R) have a distribution close to log-normal. The functional form that best fits this hypothesis is the log-linear form. The original equation which relates the rent to the other variables is:

$$R = \exp(X\beta + Z\delta + \varepsilon). \tag{2}$$

3. Method of hedonic income

In this section we briefly discuss four classical methods of measuring the capital stock. They are: (i) physical inventory stock; (ii) perpetual inventory; (iii) accounting value; and (iv) present value of hedonic prices. There are two other methods found in Brazilian literature: (v) data from income tax returns; and (vi) to use credit balance of the *Caixa Econômica Federal* (the federal savings and loan bank) as a proxy for the housing capital stock (Rebelo (1998)).

These methods are presented in the table below.

In this work we use an adaptation of Method IV - Present value of prices. The reason is the impossibility of using all the others. In Brazil, there is no information on residential housing prices or the flow of investments in residences. This makes it impossible to use Methods I and II. Method III also is not suitable because we do not have a financial census of households. There is only one for companies, which was conducted in 1985, and was subsequently substituted by sampling surveys. The information required by Method V is confidential and difficult to obtain. Method VI has a selection bias because only properties financed by the *Caixa Econômica Federal* (federal savings and loan) are covered, and these data are subject to great fluctuations that are more a reflection of macroeconomic financial conditions than real ones.

[4] We discuss more about this relation between rent and price of housing in the next section.

Table 1

Methods	Description
i. Physical inventory stock	It is a direct account of each capital unit through its average value. Example: estimation of the automobile stock in the United States;
ii. Perpetual inventory	The capital value is estimated through the accumulation of the investment flows over time. The net capital stock and depreciation rate are indirectly calculated. This is the method that has been most commonly used in the literature for its simplicity and adequacy to the available data;
iii. Accounting value	It uses the carrying value of asset on firms' balance sheets, collected by economic censuses;
iv. Present value of prices	It consists in discounting the estimated future income flows (wages and rents, respectively) to present values;
v. Data from income tax returns	Property statements are in income tax returns;
vi. Credit from *Caixa Econômica Federal* (the federal savings and loan bank)	Data from the credit balances of the *Caixa Econômica Federal* (the federal savings and loan bank) could be used as a proxy for the housing capital stock (Rebelo (1998)).

The method we use in this study is the present value of hedonic prices. We call here the "method of hedonic income". It differs from the "classical" present value of prices method in two ways. First, the data of rents used are not observed, but imputed rent for both rented and non-rented residences (owned, granted, others) through the hedonic price method. Second, we assume a perpetual capitalisation. Therefore, if we assume that residential property is a perpetual asset, its price P can be obtained by dividing the rent R by a parameter (α), which is the opportunity cost of renting a property:

$$P = R/\alpha. \tag{3}$$

Another way to shed light on the relationship between housing price and rent is to assume housing as a financial asset. In this case, "the families arbitrate between the gains in the rent market and those in the financial assets markets" (Rebelo (1998, p 25)). Parameter α is a function of the market interest rate (r) and the expected appreciation of the property (a), $\alpha = f(a, r)$. At equilibrium, $R = rP - aP$, which implies that $\alpha = r - a$.

The imputation is obtained as follows: the regression coefficients β and δ of equation (2) are estimated for 1999 and are applied to the values of X and Z obtained in each year of the period under analysis. The estimation is made in terms of September 1999 prices and is based on fixed weights.

An alternative process to the rent imputation is to construct a housing capital index based on annual weights. The advantage of this process is that it is able to capture changes in the relative prices of the attributes and the possible changes in variable definitions. The drawback is that it also captures changes in rent relative to prices and the other prices of the economy, including real estate prices.

In particular, the 1980s and early 1990s were a period of severe macroeconomic instability in Brazil. The economy had to cope with chronic and accelerating inflation. As rents have a large rigidity in relation to the general level of prices, the housing capital index would not reflect the physical behaviour of the housing stock, but only the changes in the relative prices caused by macroeconomic imbalances. The solution in this case could be to impute trend values to the coefficients. However, using a tendency can be as arbitrary as the choice of a particular year (1999, in our case) as a fixed base for the prices of the attributes. For example, during the high inflation period, particularly from 1986 to 1994, when the Real Plan currency reform finally stabilised the economy, rent increases generally lagged behind the inflation rate, since readjustments were legally limited to every six months. Then, in the first year after the Real Plan, there was an over-correction until market forces came more fully into play and rents stabilised.

In terms of national accounts, using a fixed base means, for instance, that the difference of the value added of building an apartment and that of building a house remains constant over the years. Or, it means that the value added of building a house with piped water, sewage, electricity and masonry walls (generally the most desirable construction in Brazil), among other attributes, is the same as the

value added of building a house without these attributes. As we have no information available from the building industry, using a fixed base for the prices of attributes seems to be a reasonable assumption.

The correct calculus of each attribute's price depends on the objective of estimating the housing capital stock. In this study we use the fixed base as reference. In order to compare the housing index GDP in absolute terms, we assume that the parameter α is constant with the time and equals 0.75%.[5]

4. Data

The data sources for this study are the demographic censuses of the IBGE for 1970, 1980 and 1991 and all National Household Surveys (PNADs) conducted from 1981 to 1999. PNAD is a survey that interviews from 65,000 to 115,000 households yearly. The ratio between the sample and universe is close to 1:400. The interviews occur in the last quarter of each year, with September the reference month. The demographic censuses data consist of a sample of 25% of the housing in 1970 and 1980 and 12.5% from the 1991 census. The reference month of the censuses is August. We used 12,000 observations (number of rented residences) to estimate the hedonic model.

In this study we follow the definition of a housing property made by the demographic census, which classifies it as "permanent private houses". A house is considered permanent and private if at the moment of the data collection there are at most five households living in a residence, which provided specifically residential use for the households (FIBGE (1996, p 15)). Therefore, residential properties such as those classified as "collective houses"[6] and those classified as "improvised private houses"[7] are excluded in this study.

The variables used in the rent imputation are presented on Chart 1. The dependent variable is the declared value of the monthly rent payment stated in the PNADs and in the censuses.

The median incomes are obtained in the censuses. We expect that the larger the median income, the better the location quality (infrastructure, amenities, etc). Nevertheless, due to the high inflation faced by the Brazilian economy prior to mid-1994, we had to control the income variable in order to measure the location effect appropriately.[8]

The effect of changes in the inflation rate on real income is excluded from the calculation of property appreciation in the following manner: income for all years is calculated as a deviation from the 1999 average income, as we use the latter as a standard level. The adjusted median income in the year t (MEDIADJ(T)) is described as:

$$\text{MEDIADJ}(T) = \text{MEDI} + \text{LOG}(\text{AVERAGE INCOME\_1999}/\text{AVERAGE INCOME\_}T).$$

Therefore, the average of each year is modified, but the original income variance is maintained. Any change in the variance is treated as if the quality of the location has actually changed so that the average of the variable MEDIADJ of each year t does not correspond to the average for 1999. The impact of a residence's redistribution that favours higher income locations increases the average income and vice versa. For instance, imagine that the only change in the stock of residences for a year is the destruction of an apartment building in a poor neighbourhood and the construction of a building with the same characteristics in a rich neighbourhood. This fact increases the quantum of residences because a residence in a rich neighbourhood is worth more than one in poor neighbourhood. Regarding national accounts, the construction value added of a residence in a wealthy neighbourhood is larger than the same in a poor neighbourhood. One explanation for this is

---

[5] This value corresponds to the average of the monthly gross returns of 0.5% and 1% found in the literature. See Malpezzi (1991) and Halfeld (2002).

[6] Examples of collective dwellings are hotels, boarding houses, inns, nursing homes, orphanages, convents, penitentiaries, soldier's barracks, military posts, ships, workers' housing, etc.

[7] Improvised dwellings include those located in industrial and commercial establishments, vessels, truck trailers (but not house trailers/campers/motor homes, which are rare in Brazil), rail cars, tents, rudimentary shanties, lean-tos, etc.

[8] Considera et al (1997).

that the fit and finish of a residence in a rich neighbourhood, which is not an observable variable, demands more qualified labour with higher wages, implying larger profits.

Chart 1

Variables

RENT	Monthly rent payment
APT	1 if apartment; 0 if house.
	Wall type
WALL_1	1 if type1 (masonry); 0 otherwise (reused wood, straw or other lower quality material).
WALL_2	1 if type 2 (standard cut lumber); 0 otherwise (reused wood, straw or other lower quality material).
WALL_3	1 type 3 (bare wattle and daub); 0 if otherwise (reused wood, straw or other lower quality material)
	Roof type
CEILING_1	1 if ceiling type 1 (reinforced concrete slab); 0 otherwise (standard cut timber, reused wood, thatch or other lower quality material).
CEILING_2	1 if ceiling type 2 (clay tiles); 0 otherwise (standard cut lumber, reused wood, thatch or other lower quality material).
CEILING_3	1 ceiling type 3 (zinc sheeting); 0 otherwise (standard cut lumber, reused wood, thatch or other lower quality material).
	Water supply
WATER_1	1 if water system of type 1 (public water system); 0 otherwise (tank truck, rainwater collection or other).
WATER_2	1 if water system of type 2 (well or spring); 0 otherwise (tank truck, rainwater collection or other).
	Sewage type
SEWA_1	1 if sewer system of type 1 (public sewer system); 0 otherwise (ditch, river, lake or sea, others).
SEWA_2	1 sewer system of type 2 (septic tank); 0 otherwise (ditch, river, lake or sea, others).
SEWA_3	1 if sewer system of type 3 (rudimentary septic pit); 0 otherwise (ditch, river, lake or sea, others).
	Garbage collection
GARB_1	1 if garbage collection of type 1 (if there is direct or indirect garbage collection); 0 otherwise (thrown in river, lake or sea).
GARB_2	1 if garbage of type 2 (burned or buried); 0 otherwise (if thrown in river, lake or sea).
GARB_3	1 if garbage of type 3 (if disposed of on vacant lot); 0 otherwise (if thrown at river, lake or sea).
ELET	Electricity 1 if electric lighting; 0 if pressurised bottled gas, oil, kerosene, others.

Chart 1 (cont)

Variables

Garbage collection (cont)	
ROOMM	Number of rooms besides sleeping quarters (varies from 0 to 29).
ROOMSL	Number of rooms serving as sleeping quarters (varies from 1 to 15).
BATHROOM	1 if housing has indoor bathroom; 0 if no indoor bathroom.
MEDI	Median income of census sector.
MEDIADJ	Adjusted LMEDREN to measure deviation from the 1999 average.
H_MA	1 if housing is located in a metropolitan area. 0 otherwise.
H_AUTO	1 if housing is located in a non-metropolitan area. 0 otherwise

Table 2 presents the average of the variables shown in Chart 1.

Table 2

Descriptive statistics: average

Year	Sample	Expanded sample (thousand)	APT	WALL_1	WALL_2	WALL_3	CEILING_1	CEILING_2
1970	4,410,847	17.643	…	…	…	…	…	…
1980	6,302,660	25.211	0.12	0.77	0.18	0.04	0.17	0.69
1981	103,075	26.029	0.07	0.73	0.17	0.08	0.14	0.79
1982	111,359	27.401	0.08	0.75	0.16	0.07	0.14	0.79
1983	113,463	28.185	0.08	0.75	0.16	0.07	0.15	0.79
1984	115,748	29.164	0.08	0.77	0.15	0.06	0.16	0.79
1985	119,055	30.585	0.09	0.78	0.15	0.06	0.16	0.79
1986	65,236	31.100	0.09	0.78	0.14	0.06	0.18	0.76
1987	68,449	32.136	0.09	0.79	0.14	0.05	0.17	0.78
1988	68,773	33.167	0.10	0.80	0.13	0.05	0.17	0.78
1989	70,586	34.339	0.10	0.81	0.13	0.04	0.16	0.79
1990	72,941	34.111	0.10	0.82	0.13	0.04	0.17	0.78
1991	4,342,929	34.743	0.14	0.85	0.13	0.02	0.26	0.56
1992	78,058	35.903	0.09	0.82	0.12	0.04	0.16	0.80
1993	79,948	36.819	0.09	0.83	0.12	0.04	0.18	0.78
1995	85,043	38.474	0.09	0.84	0.11	0.04	0.18	0.79
1996	84,749	39.682	0.09	0.85	0.11	0.03	0.20	0.76
1997	89,696	40.645	0.09	0.86	0.10	0.03	0.19	0.78
1998	90,714	41.840	0.09	0.86	0.10	0.03	0.20	0.77
1999	93,793	42.851	0.09	0.87	0.10	0.03	0.19	0.78

Table 2 (cont)

Descriptive statistics: average

Year	CEILING_3	WATER_1	WATER_2	SEWA_1	SEWA_2	SEWA_3	GARB_1	GARB_2
1970	…	0.26	0.11	0.08	0.12	0.19	…	…
1980	0.01	0.76	0.15	0.38	0.20	0.30	ND	ND
1981	0.02	0.60	0.25	0.30	0.14	0.34	0.49	0.15
1982	0.02	0.61	0.24	0.29	0.17	0.31	0.51	0.15
1983	0.02	0.65	0.22	0.30	0.16	0.31	0.54	0.14
1984	0.02	0.66	0.23	0.31	0.17	0.31	0.56	0.15
1985	0.02	0.68	0.22	0.34	0.16	0.30	0.58	0.15
1986	0.02	0.70	0.21	0.38	0.14	0.29	0.58	0.16
1987	0.02	0.70	0.20	0.34	0.19	0.27	0.60	0.16
1988	0.02	0.71	0.19	0.39	0.16	0.26	0.60	0.15
1989	0.02	0.73	0.19	0.40	0.15	0.28	0.63	0.14
1990	0.02	0.73	0.19	0.41	0.16	0.27	0.64	0.14
1991	0.01	0.88	0.07	0.44	0.10	0.24	0.76	0.07
1992	0.01	0.68	0.09	0.39	0.10	0.24	0.62	0.16
1993	0.01	0.70	0.09	0.39	0.11	0.23	0.65	0.15
1995	0.01	0.71	0.09	0.40	0.12	0.23	0.66	0.14
1996	0.01	0.74	0.09	0.40	0.12	0.21	0.66	0.13
1997	0.01	0.74	0.09	0.41	0.12	0.23	0.69	0.13
1998	0.01	0.75	0.09	0.42	0.11	0.22	0.70	0.12
1999	0.01	0.76	0.09	0.44	0.12	0.23	0.72	0.12

Year	GARB_3	ELET	ROOMM	BATH-ROOM	ROOMSL	MEDI	H_MA	H_AUTO
1970	…	0.36	1.18	…	0.92	6.73	…	…
1980	…	0.88	…	…	2.68	7.03	…	…
1981	0.29	0.75	3.17	0.80	2.05	6.30	0.33	0.18
1982	0.25	0.76	3.21	0.80	2.00	6.29	0.34	0.18
1983	0.24	0.78	3.26	0.82	2.01	6.29	0.34	0.18
1984	0.23	0.79	3.31	0.83	2.01	6.31	0.34	0.19
1985	0.18	0.81	3.34	0.84	2.00	6.29	0.35	0.18
1986	0.18	0.83	3.34	0.84	2.05	6.30	0.35	0.19
1987	0.16	0.84	3.37	0.85	2.03	6.26	0.35	0.19
1988	0.16	0.86	3.50	0.85	2.00	6.22	0.35	0.19
1989	0.18	0.87	3.48	0.86	1.99	6.20	0.35	0.19
1990	0.17	0.87	3.52	0.86	2.00	6.23	0.33	0.20
1991	0.10	0.97	2.15	0.86	1.72	6.30	0.33	ND
1992	0.15	0.89	3.48	0.86	1.98	6.31	0.32	0.21
1993	0.13	0.90	3.52	0.87	1.98	6.30	0.33	0.21
1995	0.12	0.92	3.62	0.89	1.98	6.27	0.33	0.21
1996	0.12	0.93	3.62	0.90	2.01	6.26	0.32	0.21
1997	0.10	0.93	3.65	0.90	1.98	6.26	0.32	0.21
1998	0.08	0.94	3.69	0.91	1.98	6.27	0.32	0.22
1999	0.07	0.95	3.70	0.92	1.97	6.28	0.32	0.22

5. Estimation

The residential real estate market is characterised by its desegregation in physical and location attributes. Pavlov (2000) shows that the coefficients of hedonic attributes of a residence vary substantially among Los Angeles neighbourhoods. The author used an econometric technique which allows smooth space variation of parameters.

Our estimation provides the coefficients to evaluate residences. We divide the real estate market for residences into two sectors: the rural market and the urban market. The definition of rural and urban area follows census and PNAD criteria. According to these sources, an urban area is defined as all areas, whether urbanised or not, of cities and towns, besides isolated urban areas. Other areas are considered as rural: rural agglomerations of urban extensions; isolated rural agglomerations, villages, and other rural zones.

Table 3 presents the estimated coefficients. In the urban area regression, 18 of the 23 coefficients are statistically significant different than zero at the 5% level, and of these, 16 are statistically significant at 1%. The five that are not significant are: intercept, wall_2 (wood walls), wall_3 (wattle and daub walls), ceiling_2 (clay tile roofs) and ceiling_3 (zinc sheeting roofs). This is an expected result in the sense that the attributes do not differ significantly from the attributes used for the comparison. In particular, regarding the wall case, the comparison is made from "walls of reused wood, straw and other lower quality material", which was the case for close to 1% of all residences in 1999. The roof case is very similar. The "standard cut lumber, reused wood, thatch or other lower quality material" classification corresponded to less than 2% of the total residences in 1999.

Table 3

Estimated coefficients - 1999

Brazil: rural, urban and total

Observations Adjusted R2	Rural 346 0.57		Urban 12,390 0.67		Total 12,736 0.68		Difference
Variable	**Estimated**	**P**	**Estimated**	**P**	**Estimated**	**P**	**URB - RUR**
Intercept	–0.84	0.19	–0.04	0.79	–0.15	0.29	0.80
APT	0.05	0.85	0.12	0.00	0.12	0.00	0.07
WALL_1	0.49	0.09	0.24	0.00	0.24	0.00	–0.25
WALL_2	0.32	0.28	0.01	0.87	0.01	0.88	–0.31
WALL_3	0.05	0.89	–0.02	0.83	–0.02	0.83	–0.07
CEILING_1	0.23	0.44	0.19	0.00	0.17	0.01	–0.05
CEILING_2	0.01	0.97	0.10	0.13	0.07	0.24	0.09
CEILING_3	–0.45	0.50	0.14	0.08	0.10	0.17	0.59
WATER_1	0.11	0.27	0.28	0.00	0.27	0.00	0.17
WATER_2	0.19	0.10	0.28	0.00	0.26	0.00	0.08
SEWA_1	–0.26	0.08	0.08	0.00	0.08	0.00	0.34
SEWA_2	–0.11	0.44	–0.04	0.02	–0.04	0.02	0.07
SEWA_3	–0.14	0.29	–0.21	0.00	–0.20	0.00	–0.08
GARB_1	0.84	0.00	0.18	0.05	0.25	0.00	–0.66
GARB_2	0.35	0.22	–0.12	0.23	–0.08	0.40	–0.47
GARB_3	0.11	0.70	–0.12	0.23	–0.12	0.17	–0.24
ELET	0.28	0.12	0.24	0.00	0.24	0.00	–0.04
ROOMM	0.07	0.00	0.11	0.00	0.11	0.00	0.05
ROOMSL	0.01	0.91	0.17	0.00	0.16	0.00	0.16
BATHROOM	0.25	0.14	0.37	0.00	0.36	0.00	0.12
MEDI	0.52	0.00	0.42	0.00	0.42	0.00	–0.10
H_MA	0.52	0.00	0.43	0.00	0.44	0.00	–0.08
H_AUTO	0.22	0.03	0.22	0.00	0.23	0.00	0.00
Urban dummy					0.08	0.00	

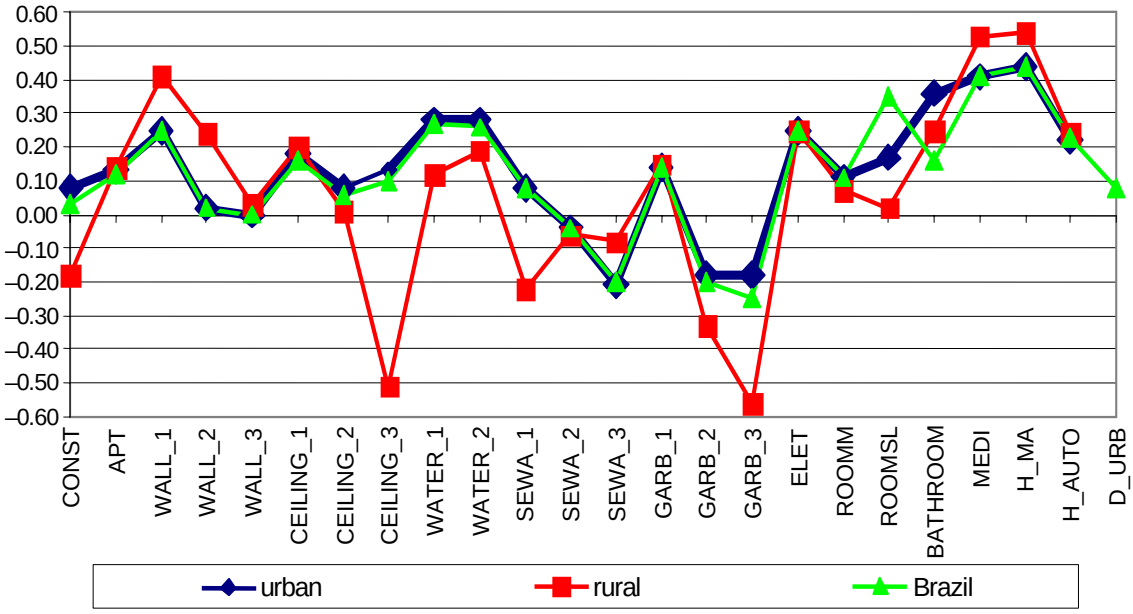
Source: PNAD (1999).

As we expected, two coefficients for the urban area are significantly negative. The sewa_2 (sewage through septic tank) and sewa_3 (sewage through rudimentary septic pit) variables are comparable to the variables sewage through ditch, river, lake and sea. Therefore, many of the residences are located close to beaches or lakes, which explains the higher values.

On the other hand, the regression for the rural area presents only six coefficients that are significantly different from zero. The lack of precision in the estimates is due to the reduced number of observations. However, this fact does not cause distortions in the value of the residence imputations. Figure 1 displays the differences between the urban and rural estimates. We note that the differences are not very high in absolute terms. For seven coefficients, the difference reaches 0.05 and for eight coefficients it attains no more than 0.20.

Figure 1

Estimated coefficients, Brazil: 1999



There are two additional problems regarding the parameter estimates. The first regards the lack of explanatory variables in the rent equation. The two main omitted variables are the state of conservation of the property (including the internal finish) and the existence of parking. These omissions may underestimate the residential values of large urban centres. The second problem is the sensitivity between stock and parameter estimates. It was beyond the scope of this work to conduct sensitivity tests of this nature.

6. Results

This section presents the residential capital and residential investment series estimated through the method of hedonic income (MHI).

Our methodology follows the estimation of the residential stock by the number of residential units. This stock aggregates its physical and location attributes and varies according to the variation of the units and the attributes. If there is an increase of more valued attributes, such as apartments, number of bedrooms, among others, the residential stock growth rate is higher than the residential units growth rate.

The problems mentioned in Section 4 indicate that the estimates should be taken with caution and could undergo future revisions.

Table 4 presents the results. The stock of residences (SR) is obtained from the rent series and it is transformed into monetary values based on the hypothesis that the median monthly rent/residential price relation is 0.75%. The stock of residences more than tripled in the last three decades, with annual growth of 4.2%. The stock of residences divided by the number of residential units corresponds to the average unitary value of the total of residences in the country (AUV). According to our estimates, the AUV increased from R$ 17,532 to R$ 23,755 (in R$ of 1999), corresponding to a growth of 35.5% or 1.05% a year.

Table 4

Housing capital value, number of residences, and GDP

Year	SR (R$ billion)	Residences (million)	AUV R$	GDP R$ billion	SR/GDP
1970	309.33	17.643	17,532	275.11	1.12
1980[1]	504.48	25.211	20,011	629.32	0.80
1981	520.86	26.029	20,011	634.18	0.82
1982	553.24	27.401	20,190	639.45	0.87
1983	584.88	28.185	20,751	620.71	0.94
1984	612.86	29.164	21,015	654.23	0.94
1985	652.29	30.585	21,327	705.59	0.92
1986	676.65	31.100	21,757	758.43	0.89
1987	696.44	32.136	21,672	785.21	0.89
1988	739.74	33.167	22,303	784.74	0.94
1989	761.27	34.339	22,169	809.53	0.94
1990	760.63	34.111	22,299	774.32	0.98
1991[1]	774.74	34.743	22,319	782.31	0.99
1992	790.21	35.903	22,010	778.06	1.02
1993	819.29	36.819	22,252	816.37	1.00
1994[1]	856.67	37.647	22,755	864.15	0.99
1995	894.05	38.474	23,238	900.65	0.99
1996	931.19	39.682	23,466	924.60	1.01
1997	951.97	40.645	23,422	954.85	1.00
1998	996.96	41.840	23,828	956.11	1.04
1999	1,017.94	42.851	23,755	963.87	1.06

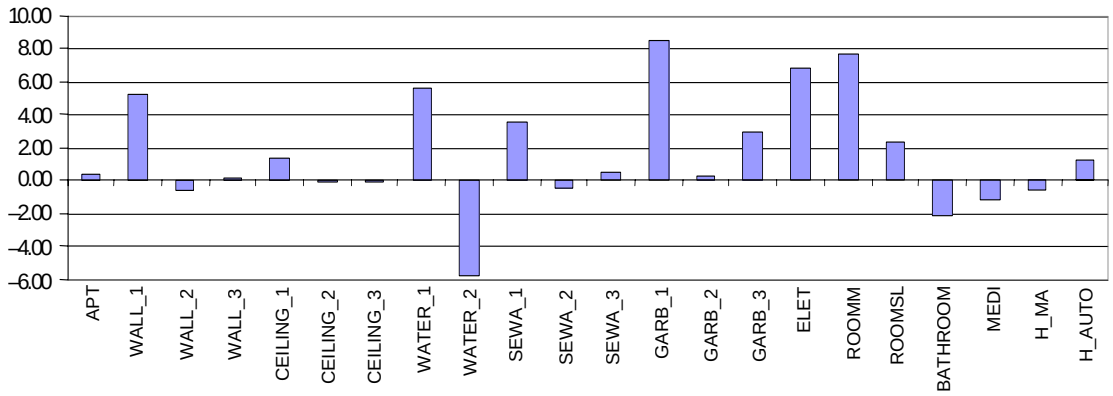
[1] Estimates obtained by interpolation.

Sources: IBGE; IPEA.

This variable measures the evolution of the quality of residences in Brazil. The residential quality growth is related to several factors. The three most important are: (i) the improvement of garbage collection; (ii) the addition of other rooms such as living rooms and kitchens in many Brazilian residences; and (iii) expanded electrification. These three characteristics contributed 22.9% out of the 35.5% total growth in quality, or 64.5% of that total (see Figure 2).

Figure 2

Growth decomposition of the residence attribute

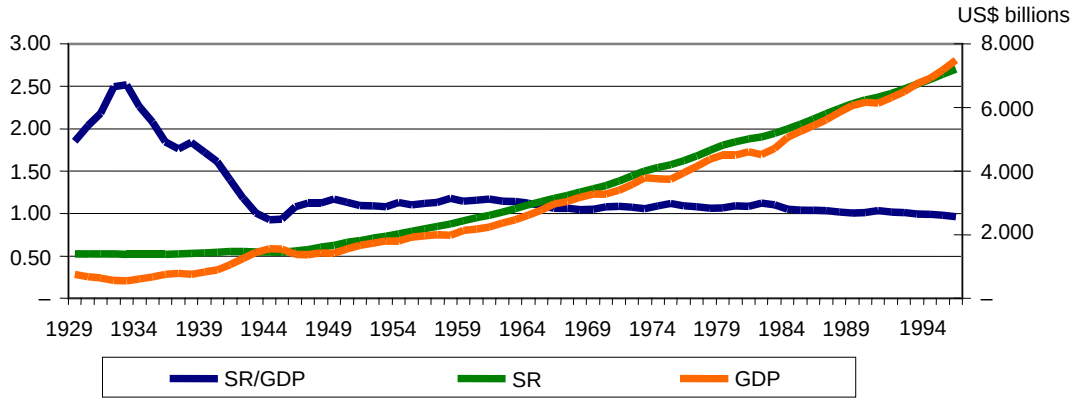


Regarding Brazil's gross domestic product (GDP), SR has shown a cyclical tendency over the last three decades. In the seventies, the SR/GDP ratio decreased from 1.12 to 0.85, which indicates that GDP increased 33% (1.12/0.85) above SR. In the two following decades, this ratio presented a constant growth trend. Therefore, we conclude that SR/GDP is strongly related to the cyclical movement of the product. See Table 4.

This cyclical movement is also observed in the North American economy. Figure 3 shows that the stock of residences behaves more smoothly than GDP, which indicates that the SR/GDP ratio changes according to GDP changes. It is worth noting the sudden fall of this ratio after the economic recovery from the economic crisis of the thirties.

Figure 3

Net stock of residences/GDP, US: 1929-97 (US$ 1992)



The SR/GDP ratio stabilised from 1946, varying only between 0.96 and 1.18. Figure 4 depicts this series and presents the economic growth cycles, such as the economic growth of the 1950s and 1960s, the two oil crises of the 1970s, and the vigorous growth of the 1990s.

Figure 5 presents the SR/GDP ratio for the period of the "Brazilian economic miracle" in the seventies and the stagnant economic growth in the eighties and nineties. The figures are similar to the ones for the North American economy, except for the tendency. While the American SR/GDP ratio presents a decreasing trend, there is an upward tendency for the Brazilian economy. There are two alternative views of these differences. One point of view is that the stock of residences in Brazil will tend to increase over GDP growth in upcoming years, and the SR/GDP ratio will tend to return to the pattern of the seventies (around 1.12). Another view is that this increase can indicate the end of a cycle of low growth of the product, which may point to economic recovery in the next few years.

Figure 4

Net stock of residences/GDP, US: 1946-97 (US$ 1992)

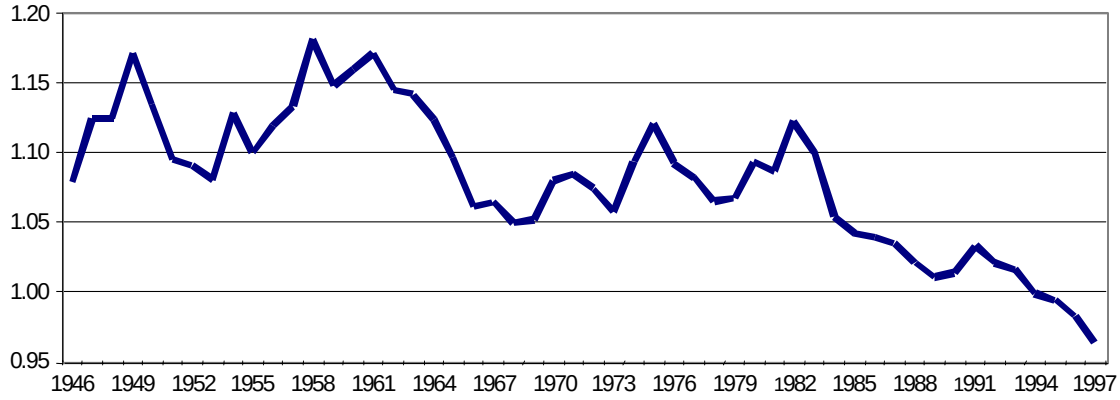
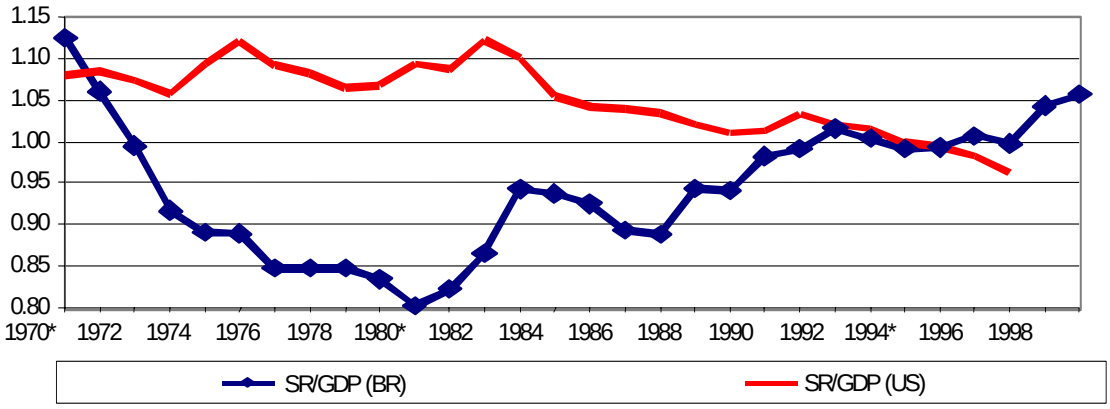


Figure 5

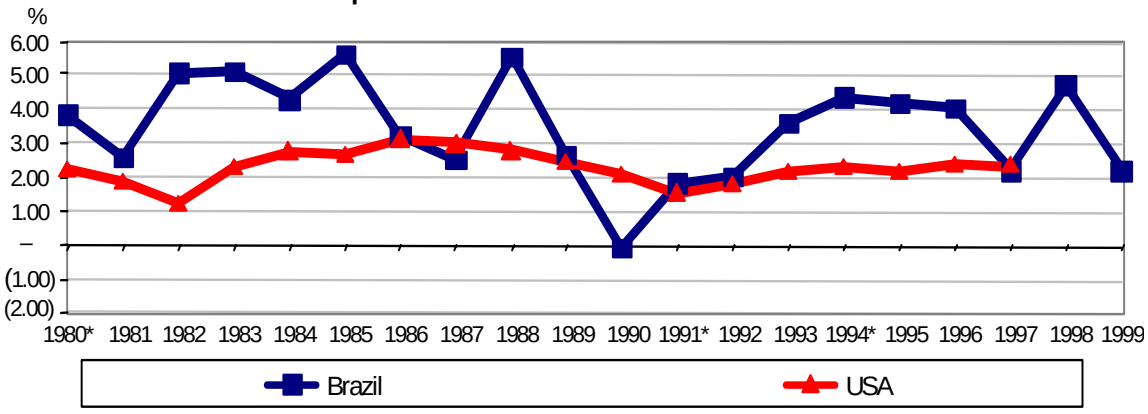
Stock of residential capital/GDP: Brazil and US: 1970-99



Finally, Figure 6 presents a comparison of the investment in residences in Brazil and in the United States from 1971 to 1999. Our estimates indicate that the residential investment in Brazil is more volatile than that in the United States of America. However, the tendency in both countries is the same. Brazilian housing investment attained around 4-5% of GDP in the seventies and decreased to 3% in the eighties and nineties. In the United States of America, the ratio of housing investment to GDP was 3% in the seventies and 2% in the nineties.

Figure 6

Net investment in residences/GDP: 1971-99
comparison between Brazil e United States



7. Concluding remarks

The work estimated the residential capital stock in Brazil for the period 1970-99. It is an exploratory work attempting to expand the range of statistics available on the residential property sector in Brazil. There are many applications of the data estimated here, such as use as an indicator of the level of real estate activity, an indicator of the level of welfare, disaggregation of investment in civil construction, imputation of rents in the national accounts, and compilation of consumer price indices.

The data available for Brazil do not allow us to use conventional methods to estimate the level of capital stock, namely the perpetual inventory method and the accounting value method. Hence, we used an innovative method called the "hedonic income method", by which we used the hedonic price model to impute the rent and then transform the rent (income) into a value, dividing by a discount rate, which is known as the "income method".

The results indicate that the magnitude of the residential capital (KR) is near the magnitude of the GDP. They also indicate that the residential capital series has lower variance than the GDP series, so that the fluctuation in the KR/GDP ratio is governed by the economic cycle. These results also are observed for the US economy, which we used as a benchmark in our analysis.

A secondary derivation of the residential capital series is the net residential investment series, which is not made available in the national accounts. Once again, using data from the United States as a benchmark, our estimates cannot be disregarded.

We can suggest three questions and extensions that can lead to future revisions of the residential capital series. The first refers to the starting data of the series data, 1970. The set of variables that describe residences is smaller than that from 1981 onward - there are nine variables available for 1970 against 22 for the remaining years. Besides this, one must consider that the source of data for 1970 is the Demographic Census, while for the period after 1981 the source is the PNAD, with the values for 1980 and 1991 estimated only using the number or residences obtained in the census. At first glance, this problem would lead us to conclude that the series is more consistent from 1981 on. But the fact that the data from 1970 capture the expanding cycle of the Brazilian economy in the 1970s is indicative that the estimate for that year is not totally absurd.

A possible extension of the work would be to transform the fixed-base series into a moving-base series. This would avoid the arbitrariness of choosing 1999 as the base year, besides capturing the price variables relative to the attributes of the properties occurring over the years.

A third question/extension, related with the second, is the estimation of the parameters. One could undertake some analysis of the consistency of the residential capital series by modifying or expanding the areas of the estimations of the parameters. For example, instead of estimating them by rural and urban area, they might be estimated for metropolitan and non-metropolitan region, or other divisions along these lines.

Finally, greater coverage of rural areas is needed.

References

Considera, C, M Neri and A Pinto (1997): *A evolução recente do bem-estar, da pobreza e da desigualdade*, IPEA, mimeo.

Green, R K and P H ndershott (1993): "Demographic factors and real house prices", *NBER Working Papers* 4332.

Halfeld, M (2002): "Seu imóvel", *Editora Fundamento*.

Haughwout, Andrew F (2000): *Public infrastructure investments, productivity and welfare in fixed geographic areas*, Federal Reserve Bank of New York, processed, May.

Hendershott, P and M White (2000): "The rise and fall of housing's favored investment status", *Journal of Housing Research*.

Instituto Brasileiro de Geografia e Estatística - IBGE (1970, 1980 and 1991): *Censos demográficos*, IBGE, Rio de Janeiro.

——— (1981 to 1999): *Pesquisa nacional por amostra de domicílios (PNAD)*, IBGE, Rio de Janeiro.

——— (2000): *Contas nacionais*, IBGE, Rio de Janeiro.

Kanemoto, Yoshitsugu (1997): "Housing question in Japan", forthcoming in *Regional Science and Urban Economics*.

Katz, Arnold J, Shelby W Herman and Michael Glenn (1999): *Fixed reproducible tangible wealth in the United States, 1929-94*, Bureau of Economic Analysis (BEA).

Lucena, José Mário P (1985): "O mercado habitacional no Brasil", FGV-RJ, *Série Teses* no 9, Rio de Janeiro.

Malpezzi, S (1991): "Rent control in developing countries", *BIRD Discussion Paper* 129, Washington.

Melazzo, E (1996): "Apontamentos para uma valoração do patrimônio Imobiliário urbano Brasileiro", *ANPUR*, Anais, IV Encontro, 504-8, Brasília.

Morandi, Lucilene (1995): *Estoque da riqueza interna tangível e reprodutível - Brasil, 1970/95*, Universidade Federal Fluminense (UFF), Niterói.

——— (1998): "Estoque de riqueza e a poupança do setor privado no Brasil", Instituto de Pesquisa Econômica Aplicada, Texto para Discussão, 572.

Morandi, Lucilene e Eustáquio J Reis (2003): "Estimativa do Estoque de Capital Fixo-Brasil, 1950-2000", working paper, IPEA-RJ, forthcoming.

Munnell, Alicia H (1990): "How does public infrastructure affect regional economic performance?", Chapter in Munnell (ed), *Is there a shortfall in public capital investment?*, Federal Reserve Bank of Boston, Boston, Massachussets.

Neri, M C et al (1999): "Assets, market and poverty in Brazil", Inter-American Development Bank, Research Network, *IRBD Working Papers Series* R-357.

O'Sullivan, Arthur (1996): *Urban Economics*, Homewood, Irwin.

Pavlov, Andrey D (2000): "Space-varying regression coefficients: a semi-parametric approach applied to real estate markets", *Real Estate Economics*, vol 28, pp 249-89.

Poterba, James M (1992): "Taxation and housing: old questions, new answers", *NBER Working Paper* 3963.

Rebelo, André M (1998): "O mercado habitacional brasileiro: aspectos teóricos e empíricos do investimento habitacional nas unidades da federação", *Estudos econômicos da construção*, vol 3, no 2(6), pp 21-39.

Santos, C H M e B O Cruz (2000): "A dinâmica dos mercados habitacionais metropolitanos: aspectos teóricos e uma aplicação para a Grande São Paulo", *IPEA Texto para discussão* 713, Brasília.

Sheppard, Stephen (1999): "Hedonic analysis of housing markets". Em Cheshire, Paul e Mills Edwin (eds), *Handbook of regional and urban economics*, vol 3, Elsevier Science BV, North-Holland.

Spilerman, S (2000): "Wealth and stratification processes", *Annual Review of Sociology*, no 26, 497-524.

Wolff, Edward N (2000): "Recent trends in wealth ownership", in T M Shapiro and E N Wolff (eds), *Benefits and mechanisms for spreading asset ownership in the United States*, Russell Sage Press, New York.

Zabel, Jeffrey E (1999): "Controlling for quality in house-price indices", *Journal of Real Estate Finance and Economics*, pp 223-41.

Hedonic housing price indexes: the French experience

Anne Laferrère

Summary

The whole of France is now covered by quarterly hedonic housing price indexes. They are the fruit of a unique collaboration between the notaries, who collect the data and do the computation, and INSEE, the French statistical agency, that provides the methodology and guarantees its proper use. The method relies on the econometric estimation of the value of fixed baskets of apartments or houses in some 300 geographic zones. The quarterly estimate is based on all sales recorded in a given period. Each zone has its own hedonic model of price. This allows to create indexes at various geographic levels. The regular publication of hedonic housing price indexes at a country level is rare, because it is costly. The French endeavour manages to be relatively cheap to the taxpayer because of a unique methodology and complementary interests between the involved parties.

Introduction

It is common knowledge that housing is for most households the main part of their wealth, and for all one of the main items of consumption. Many governments still regulate rents, grant rental allowances, allow tax deduction for mortgage payment, or subsidise the construction of public housing. Some enact laws making housing an "entitlement" for the poor. Most tax capital gains on houses, usually exempting a first home, many tax houses sales, regulate land use by zoning. They also decide on the spatial distribution of public goods such as schools, roads or public transportation, which have a direct effect on the price of land, hence of houses. Moreover all countries scrutinise inflation level. In view of this importance the absence of reliable data on the evolution of housing prices and the opacity of the housing market are often regretted.

The making of a price index requires reliable data, a strong methodology and a willingness for long term involvement. In France all housing sales have to take place in front of a notary who draws up the deed and is in charge of collecting stamp duty for the central government. In 1994 the notaries started to centralise those data. This had been done for the city of Paris since the 1980s and a housing price index was published under the name "Notaires-INSEE", because INSEE (the National Institute of Statistics and Economic Studies), the French statistical agency had helped with the method. In 1997 the notaries naturally turned to INSEE when they wanted to start computing indexes for the rest of France. INSEE agreed to collaborate, providing that a reliable econometric method would be applied. In a context of reduction of public expenses, the collaboration seemed a way of providing a public service at the lowest cost to the taxpayer. The existence of a research centre within INSEE (CREST, Centre de Recherche en Economie et Statistique) proved very important in spurring the methodological reflexion.[1] The new indexes started to be published quarterly in the summer of 2000, and the whole of France was covered at the end of 2002. In 2003 the first revision of the econometric specifications has been successfully done, and more detailed and reactive indexes will be published from 2004.

The methodological problem of a housing price index is no different from that of any price index: how is pure price evolution to be separated from changes in the quality of houses? But it also raises specific questions. First, two houses are never exactly the same, because they have many characteristics, the unique combination of which translates in a particular housing service. These

[1] Alain David (Perval), François Dubujet (INSEE), Christian Gouriéroux (CREST) and the author of this article, then head of INSEE Housing division, made up the team. Claude Taffin and Jacques Friggit also provided help.

"qualities" include construction type, number of rooms, plumbing, heating and other facilities, but also a specific location (hence access to local public goods) and neighbours, not to mention family ties to a place. Second, a house is a durable good: in France more than a third of first homes are more than 50 years old. As a result 75% of housing purchases are second hand, and, what is more important for a price index, homes rarely change hands. Price observation is rare. Those two features lead to the first problem: how to know the price evolution of a house or of a given group of dwellings, when very few prices are observed at each period, and, if observed, are those of dwellings different from those in the group?

The second problem is two-fold. On the one hand, not only the observed transactions are few, but they are not a random sample of the housing stock. On the other hand the housing stock itself is not fixed: it keeps changing through destruction, deterioration, improvement, new construction, extension, etc. Should the housing stock be perfectly fixed, transactions would have to be a large enough random sample of the housing stock to be validly used to compute price evolution. It never is the case, thus the comparison of average sale prices is a mixture of true price evolution and change in the quality of the sample of transactions drawn from the stock. For all these reasons econometric techniques cannot be avoided.

Section 1 dwells on the methodology of the indexes and their computation. In Section 2 the data collection and database are described, together with the process of index publication. Section 3 concludes.

1. Why a hedonic method and which?

As mentioned, dwellings have various idiosyncratic characteristics, including location, size, number of rooms, occupancy, age …. This heterogeneity translates into different sub-markets, various turnover rates, and prices. It leads to difficulties in analysing prices which are rarely observable. Indeed between two sales the value of a house, in the economic sense, cannot be given. It has to be estimated from a model of price.

We make two basic assumptions. The first is common to all hedonic indexes. Each dwelling is defined by the combination of a fixed number of characteristics. That is to say that if a new quality would appear, say air-conditioning or being asbestos free, it would not be considered a quality but an increase in price.[2] The idea is that consumer's utility is not yet changed by those types of quality changes, which are regarded as price changes. The assumption also stems from the fact that only a limited number of attributes are recorded in the data. The second assumption is less common: the relationship between the price of a house and its characteristics is fixed, in a given zone, during the index computation period. It greatly simplifies the quarterly computations as they imply no more econometrics.[3]

In a given zone, the price index is defined as the ratio of the estimated value of a reference stock of dwellings, a "basket of houses", to its value at the base period of the index. For each quarter, the value of each dwelling in the reference basket is estimated from the prices of all observed sales, thanks to the econometric models that have been estimated on the sales of the "estimation period".

The process has several steps, as follows:

1. define zones (strata), where price evolution is assumed to be homogeneous;
2. define a hedonic model of price, ie introduce correction coefficients for quality effects, for each zone;
3. estimate the correction coefficients from an estimation stock of dwellings in each zone;

---

[2] The "quality" may be a defect. For instance noise is not one of the characteristics usually observed or taken into account in hedonic price models. This means that an increase in neighbourhood noise level is implicitly interpreted as a price decline, if it is capitalised in prices.

[3] The validity of this assumption has been checked. See Appendix A1 for details.

4. compute the value of a reference stock at the base date for each zone;

5. compute the value of this reference stock, from data on all current period sales, by zone;

6. compute the price index as the evolution of the value of the reference stock between base and current date;

7. publish indexes and sub-indexes by aggregation of local zone indexes.

The first four steps are done once and for all (and only revised every five years), the three last ones are repeated every quarter.

1.1 The hedonic method

Among all hedonic housing price indexes that we are aware of, the French index has the unique features of combining a large number of geographic zones and the quarterly estimation of so-called "reference stocks" of dwellings in each zone. This section describes those features in some details.

Defining zones

Dwellings (houses and apartments are separated all along) are assumed to be stratified into groups where prices are homogeneous and price trends are roughly parallel. Ideally the elementary geographic zones could be very small sub-markets. Practically they were limited to a few hundreds to ensure a sufficient number of sales in each (over 400 per year). Typically, for large enough cities, above 10,000 inhabitants, a zone is a city centre or a city suburb; it is a group of rural areas or smaller towns in less densely populated regions; it is close to an arrondissement in Paris.[4]

Reference stock

The main idea of the hedonic method is to abstract from the variations of the structure of the market. We achieve this by estimating the value of a reference stock of dwellings at each date. Thus the index follows the price of the same dwellings. A reference stock is defined in each of the 315 elementary zones. It is made of all sales of a reference period of three to five years in that zone, excluding sales in the extreme quantiles of prices per square metre. The reference period is for instance 1994-96 for apartments in the Province;[5] the other reference periods are given in the last column of Table 1.[6] The reference stock in each zone is on average made of some 2,000 dwellings, and the reference stock for France is more than 630,000 dwellings (Table 1, col 4).

Hedonic models of price

To estimate the value of the reference stock econometric models are used, relating prices (the log of the price per square metre) to the characteristics of the dwellings. The characteristics Z_i include location (a neighbourhood within a zone), and quality of the dwelling itself (see Section 3.1.2 for details). Each model is estimated from a stock of dwellings called estimation stock. It includes all dwelling sales of the reference period, except those for which the number of rooms is not known, or the estimated price was found ex post to differ from the observed price by more than two standard errors. Thus is it close but not equal to the reference stock (see Table 1, col 5).[7] The econometric estimations are made separately for each elementary geographical zone *s*.

---

[4] See Appendix A2 for more details.

[5] By Province we mean all regions except Ile-de-France (Paris and the seven *départements* around it).

[6] The different reference periods stem from the different time implementation of the method, not from any technical reason.

[7] The first version (which is presented here) of the index was rather conservative and the reference stock excluded the bottom and top 10% of dwellings; hence the estimation stock could be larger than the reference stock. After the 2003 revision, only bottom and top 5% of dwellings are eliminated.

The model is the following:

$$\text{Log } p_i = \text{Log } p_0 + \sum_{a=1}^{3} \alpha_a Y_{a,i} + \sum_{t=1}^{4} \theta_t T_{t,i} + \sum_{k=1}^{K} \beta_k X_{k,i} + \varepsilon_i \quad (1)$$

where p_i is the price per m^2 of dwelling i, $Y_{a,i}$ is a dummy for year of sale of dwelling i, $T_{t,i}$ a dummy for quarter of sale of dwelling i and $X_{k,i}$ are the k characteristics of dwelling i.

Table 1

Number of local zones, number of neighbourhoods, size of reference and estimation stocks of the Notaires-INSEE indexes

Index	Zones	Neighbourhoods	Reference stock	Estimation stock	Reference period
Total Ile-de-France	**84**	**333**	**247,673**	**284,222**	
Apartments					
Paris	18	80	60,152	70,341	1992-96
Petite Couronne	24	92	53,014	62,098	1992-96
Grande Couronne	18	100	55,191	68,874	1998-2001
Houses	24	61	79,316	82,909	1997-2001
Total Province	**231**	**576**	**390,388**	**399,124**	
Apartments					
UU > 10,000 inhab.	113	302	123,739	123,296	1994-96
centre	84	236	90,571	90,228	1994-96
suburb	29	66	33,168	33,068	1994-96
UU < 10,000, rural	7	35	13,526	24,704	1994-96
Houses	111	239	253,123	251,124	1995-97
Total	**308**	**874**	**638,061**	**683,346**	

Note: Petite Couronne: the first circle of outskirts of Paris (Haut-de-Seine, Seine-Saint-Denis, Val-de-Marne). Grande Couronne: the rest of Ile-de-France, further from Paris (Essonne, Seine et Marne, Yvelines, Val d'Oise). Province: all other *départements* of metropolitan France, except Corsica.

The variables $X_{k,i}$ are continuous or dummy variables computed from the modalities of the variables Z_i and sometimes including interaction effects.[8] The coefficients of the model are prices defined relatively to reference characteristics, which together define a reference dwelling, the price of which is p_0.[9] The reference apartment has the following characteristics: three rooms, ground floor, rooms of average size, no service room, no parking, no terrace nor balcony, one bathroom, built between 1948 and 1980. The reference house has four rooms, two levels, one garage, one bathroom, a basement, no garden, its construction period is unknown.

[8] For instance the presence of an elevator is interacted with floor level. Note that, as there are as many models as zones, all variables are de facto interacted with the geographical zone.

[9] The reference dwelling is one of a precise quarter and year of sale. The value of a dwelling with the same characteristics, but sold at a different time would be computed from p_0 by adding the corresponding quarter and year parameters θ_t and α_t.

Current value of the reference dwelling

The same type of model could be used at the current period τ, with the same reference dwelling of price $p_{0,\tau}$. The price per m^2 of dwelling *j* sold in period τ is written as[10]

$$Log(p_{j,\tau}) = Log(p_{0,\tau}) + \sum_{K=1}^{K} \beta_K X_{K,j,\tau} + \varepsilon_{j,\tau} \ .$$

The period τ is chosen according to the type of index. More precisely, the index for a quarter *t* is computed over all arm-length transactions of a period τ ending with quarter *t*. Up to the end of 2003, the Parisian index was computed on a six-month basis, hence $\tau = [t-1;\ t]$; indexes for the Province were annual, $\tau = [t-3;\ t]$. From 2004 on they are all pure quarterly indexes, $\tau = t$, which makes them more reactive and also allows to study seasonal price variations. However quarterly indexes at a more local level remain semestral or annual to ensure a sufficient number of transactions in the zone.

Let me now explain how the price of the reference dwelling is computed from data on current sales. Assume that the $\beta_{k,\tau}$ coefficients are known, and call $\tilde{p}_{j,\tau}$ the price that would fetch dwelling *j* if it had the characteristics of the reference dwelling, then:

$$Log(\tilde{p}_{j,\tau)} = Log(p_{j,\tau}) - \sum_{K=1}^{K} \beta_K X_{K,j,\tau}$$

It is the "reference dwelling equivalent price" of dwelling *j*, τ. The model can be rewritten:

$$Log(\tilde{p}_{j,\tau}) = Log(p_{0,\tau}) + \varepsilon_{j,\tau}$$

Hence, if the β_k coefficients are known, the logarithm of the price of the reference dwelling Log ($p_{0,\tau}$) can be estimated as the means of all estimated prices:

$$Log(\hat{p}_{0,\tau}) = \frac{1}{J\tau} \sum_{j=1}^{J\tau} Log(\tilde{p}_{j,\tau}) \,,$$

where J_τ is the number of transactions of period τ.

Using our second assumption (the price model is stable over time), allows to replace the β_k coefficients by the $\hat{\beta}_k$ estimated over the reference period. As mentioned above this greatly simplifies the computation as there is no econometrics to be done:

$$\text{Log}(\tilde{\ }_{,\tau}) \text{ s Log}(\ \ _{,\tau}) - \sum_{=1} \hat{\beta} \qquad _{,\ ,\tau} = \text{Log}\left[\frac{_{,\tau}}{\exp\left(\sum\ _{=1} \hat{\beta}\ \ _{,\ ,\tau}\right)} \right]$$

Then the log of $p_{0,\tau}$, the price per m^2 of the reference dwelling in period τ, is estimated by a geometric mean of the "reference dwelling equivalent prices" of the J_τ dwellings sold in period τ:

$$\log \hat{p}_{0,\tau} = \frac{1}{J_\tau} \sum_{j=1}^{J_\tau} Log\tilde{p}_{j,\tau} = \frac{1}{J_\tau} Log\left(\prod_{j=1}^{J_\tau} \tilde{p}_{j,\tau} \right)$$

or:

$$\hat{p}_{0,\tau} = \left(\prod_{J}^{1} \tilde{p}_{j,\tau} \right)^{\frac{1}{J_\tau}} .$$

---

[10] The evolution of the price of the reference dwelling is the core of the index construction. For this reason it must include seasonal and cycle effects. This is why the quarter and year parameters are not in the current period model, while they were introduced in the first model because the estimation was made over more than one quarter. The price for a dwelling of quarter (*a*, *t*) would be: Log ($p_{0,a,t}$) = Log $p_0 + \alpha_a + \theta_t$. See examples of hedonic models in Appendix A3.

Current value of the reference stock

Knowing an estimated value of the reference dwelling, the estimated value of each dwelling of the reference stock can be computed, and, by aggregation, the value of the stock itself. Computations are made zone by zone. For this reason we re-introduce the index *s* of the zone. The value of dwelling *i* of the reference stock of zone *s* in the current period τ is estimated from its characteristics $X_{k,i,s}$, which are time invariant, by definition of the reference stock. The approached value is:

$$\hat{p}^*_{i,s,\tau} = \exp\left(Log\tilde{p}_{0,s,\tau} + \sum_{k=1}^{K} \hat{\beta}_{k,s} X_{k,i,s} \right) A_{i,s},$$

where $A_{i,s}$ is the surface of dwelling *i*, *s*.

The estimated current value of the N_s dwellings of the reference stock of zone *s* is obtained by a sum:

$$\hat{W}_{s,\tau} = \sum_{i=1}^{N_s} \hat{p}^*_{i,s,\tau}$$

In the same way, and once and for all, the value of the reference stock is estimated at period $t = 0$, the base period of the index. One notes:

$$\hat{W}_{s,0} = \sum_{i=1}^{N_s} \exp\left(Log\hat{p}_{0,s,0} + \sum_{k=1}^{K} \hat{\beta}_{k,s} X_{k,i,s} \right) A_{i,s}$$

the value of the transactions in zone *s* at date 0, as estimated by the model.

Quarterly computation of the index

The elementary index of a zone *s* is by definition the evolution of the value of the reference stock of that zone *s*. It is:

$$I_{t/0}(s) = \frac{\hat{W}_{w,\tau}}{\hat{W}_{s,0}} = \frac{\sum_{i=1}^{N_s} \exp\left(Log\hat{p}_{0,s,\tau} + \sum_{k=1}^{K} \hat{\beta}_{k,s} X_{k,i,s}\right) A_{i,s}}{\sum_{i=1}^{N_s} \exp\left(Log\hat{p}_{0,s,0} + \sum_{K=1}^{K} \hat{\beta}_{k,s} X_{k,i,s}\right) A_{i,s}}$$

Note that the index of a zone *s* can be written:

$$I_{t/0}(s) = \exp\left(Log\hat{p}_{0,s,\tau} - Log\hat{p}_{0,s}\right)$$

Practically, the computation of the index at date *t* does not need the computation of the implicit value of each dwelling of the reference stock, the coefficients Log $\hat{p}_{0,s,\tau}$ are obtained by:

$$\text{Log } \hat{}_{0,\tau} = \frac{1}{J_\tau \sum_{=1}^{\tau} \text{Log}(\quad_{,\,,\tau}) - \sum_{=1} \hat{\beta}_{\,,\,} \overline{\cdot}_{\,,\,,\tau}}$$

where $\overline{X}_{k,s,\tau}$ is the mean of the $X_{k,j,\tau}$ variables for the J_τ transactions of the current period in zone *s*.

Aggregated indexes

Most elementary indexes by zone are not published. They are aggregated at higher geographical levels. For instance the index for the "Province" measures the evolution of the value of the whole reference stock of Province.

$$I_{t/0} = \frac{\hat{W}_\tau}{\hat{W}_0} = \frac{\sum_s \hat{W}_{s,\tau}}{\sum_s \hat{W}_{s,0}},$$

It can be interpreted as the mean of the elementary indexes of zones, weighted by the sales value in the zone in the reference stock.

$$I_{t/0} = \sum_s \left(\frac{\hat{W}_{s,0}}{\sum_s \hat{W}_{s,0}} \right) I_{t/0}(s).$$

Practically, the weights of some indexes are redressed by a parameter δ_s for zones where the notary database is deemed to be non exhaustive.[11]

1.2 A numerical example for year 1999

This section briefly describes, on a concrete example, the steps of computation of the index of the fourth quarter of 1999 for a zone of the Province.

Step 1: extract data from the database

The index of 1999 fourth quarter is computed from all sales of the last quarters, that is from the whole of year 1999.[12] Four hundred and seven sales were registered in the zone for that period. For one of them, neither the number of rooms, nor the surface are known, it is eliminated, and 406 sales remain. The price per m^2 range from 2,825 to 8,065 Francs.

Step 2: estimate the price of the reference dwelling

The first of the 406 sales is a four-room apartment of 71,25 m^2, with one bathroom and a car-park, without service room, on the third floor, without lift, in a building dating from 1963. The price was 374,000 Francs, or 5,249 F/m^2. The computations have been done with decimal logarithms and $\log_{10}(5249) = 3.720$. Using the previously estimated model allows to know the price it would fetch if it had the characteristics of the reference dwelling.

$$\begin{aligned}\log(\tilde{\ }_{,\tau}) &= 3.720 \\ &\quad -0.0858 \ \text{(parking)} \\ &\quad -0.0231 \ \text{(third floor)} \\ &\quad +0.0184 \ \text{(four rooms)} \\ &\quad -0.0191 \ \text{(surface per room less than } 20\,\text{m}^2\text{)} \\ &= 3.6105.\end{aligned}$$

The "equivalent reference dwelling price" $\tilde{p}_{j,\tau}$ for this first sale is obtained as $10^{3.6105}$, that is 4,078 F/m^2.

The same operation is done for the 405 other sales. Sales for which the estimated equivalent reference price is deemed extreme (lower than one sixth (3,459 F/m^2) or higher than five sixths (5,336 F/m^2) of the price distribution) are excluded. Thus only 270 dwelling prices will be used in the index computation.

The estimated logarithm of the price of the reference dwelling, is the arithmetic mean of the logarithms of the 270 equivalent reference dwelling prices. It is 3.6334, which gives a price of 4,302 F/m^2.

Step 3: compute the value of the reference stock at current prices

Now the current value of each dwelling of the reference stock of the zone has to be estimated, from the same model. It can be interpreted as the price that the dwellings would fetch in 1999. Taking for example a one-room apartment of 28 m^2, with one bath, on the ground floor of a 1980 building, without a car-park.

$$\begin{aligned}\text{Log}(\hat{p}_i^*) &= 3.6334 \\ &\quad +0.0571\,\text{(studio)} \\ &\quad +0.0329\ \text{(built in 1980)} \\ &= 3.7476,\end{aligned}$$

or 5,592 F/m^2 for a total price of 156,588 Francs.

---

[11] Weights are estimated from stamp duty returns and correspond to the value of the stock in each zones.

[12] Remember that in 1999 the index was quarterly published but on an annual basis. From 2004 on it will be a pure quarterly index.

Adding all the estimated prices gives the current value of the reference stock of the zone, 139.214 million Francs.

Step 4: compute the index in the zone

The value of the reference stock at the base period was 127.902 million Francs. The index 1999 fourth quarter, base 100 in 1994 fourth quarter, is:

$$I_{1999Q4/0} = \frac{139.214}{127.902} \times 100 = 108.84.$$

This computation, done in April 2000, gives a provisional index for 1999 fourth quarter because all transactions for the last quarter of 1999 are not yet registered at that date. About 85% of the definitive total is present. A special weighting system accounts for the discrepancy. The computation is done again in July 2000 and yields a definitive index.[13]

2. Database and publication policy

All real estate transactions in France have to be registered in front of a notary. There are 4,600 notary practices over the country. The role of a notary is to verify and certify the legal sale contract and deed, send the records to the Conservation des Hypothèques (Mortgage Register), and collect the stamp duty for the government. A notary is at the same time an *officier ministériel*, a public official, and a private entrepreneur.[14] Thanks to this feature of the French legislation, the data collected by the notaries are well suited to computing quarterly housing price indexes.[15] Compared to other well-known indexes such as Freddie Mac in the United States of America or Halifax in Great Britain, the data include purchases made without a mortgage.

2.1 Database

A "Notaires-INSEE" index was created in 1983 for second-hand apartments in Paris.[16] In 1997, the Conseil Supérieur du Notariat (CSN) wanted to create a price index for dwellings outside the Paris region, in the Province. The separation of the Paris region and the Province, each with its own database, is a complication that is due to a history of high centralisation in France with the Paris region concentrating a large part of the wealth.[17] INSEE agreed to provide a methodology, because a public service of reliable housing price indexes was lacking in France. To ensure long-term involvement of both parties, formal agreements were signed between the CSN and INSEE in 1998 for indexes for apartments in the centre large urban units of the province, then in 1999 for houses, and apartments in the outskirts of cities and rural areas, and later between the CINP and INSEE. A renovated Parisian index was published in the spring of 2000, indexes for apartments of Petite Couronne were made available in 2002, and indexes for the Grande Couronne followed in 2003.

Each notary is asked to send for key-boarding an extract or a photocopy of the sale deed, plus some extra notes on the dwelling characteristics. This is done on a voluntary basis. In the near future the sale contracts should be normalised and computerised, thus the process could be automatic and use

---

[13] If sales are registered later than July 2000, the index does take them into account.

[14] Part of his fees are regulated by the government, with a fixed part and one roughly proportional to the sale value.

[15] FNAIM (an association of real estate brokers) also collects prices from "commitments to sell", that is earlier in the sale process, and calculates price evolution, but it covers only 13% of transactions.

[16] It was called so because INSEE had helped by defining cells and providing weights from the Census.

[17] The oldest database is the one collected by the CINP (Chambre Interdépartementale des Notaires de Paris, de la Seine-Saint-Denis et du Val-de-Marne) from 1979 for Paris and Petite Couronne, from 1995 for the Grande Couronne. The database of the CSN is more recent: created in 1990, and really in operation in 1993, this base is managed by Perval, a company owned by the CSN, who is in charge of getting the data from the notaries of province. The making of the indexes brought the Parisian, and Province databases closer. For instance a sale of a Parisian dwelling made by a notary of province is now included in the Parisian database and vice-versa.

electronic mail. This is not now the case. The data on a particular sale are integrated in the database within three to four months. The index is restricted to arm-length transactions of second-hand dwellings.[18] To enter the computation a dwelling has to be free for occupation (not rented), only used for habitation (no professional use), and has to be acquired in full property, by a private individual or by a SCI (Société civile immobilière). Exceptional homes such as single rooms, attics, studios, castles, are excluded. The database encompasses all real estate transactions, including for instance parking lots, new buildings, or land. The housing database has some five million references in 2003 (25% are in the Ile-de-France database) and roughly half of them are apartments, half are houses. Some 510,000 housing sales have been registered in 2002.

2.1.1 Coverage rate

The rate of coverage of the notary data compared to the total housing transactions is not perfectly known because there are no official statistics on housing transactions.[19] However test estimates of this coverage have shown it to be around 80-90% for Paris; for the outskirts of Paris it has increased from 60% in 1991 to 80% in 1999. It is around 64% for the rest of France and varies from one district to the other.[20] Overall the coverage has been estimated around 71% of the value of transactions in 2001. As already mentioned, to compensate for the spatial variation in coverage rates, the reference stock for the province and Grande Couronne has been re-weighted (see footnote 11).

Actually a 100% coverage rate is not necessary to compute an index, if the sales are a randomly selected and large enough sample. The way each notary chooses to send the data is not well known. But there is no reason to think that it generates a significant bias on the index at an aggregated level. Moreover the choice of a fixed basket of dwellings over a large enough period to compute the index minimises the adverse effect of potential selection bias.[21]

2.1.2 Characteristics of dwellings, treatment of non-responses

The database is anonymous, to comply with the French law. The precise address of the dwelling is included but is not made public, and is not used in the index computation. Besides the date of the sale, the dwelling characteristics are the following: surface (in m^2), location (census track), time of construction (eight categories: $<$1850, 1850-1913, 1914-47, 1948-69, 1970-80, 1981-91, 1992-2000, $>$2000), number of rooms (from one, to five and more), mean surface per room, number of bathrooms (zero, one, or two and more), number of garages or car parks (zero, one, or two and more) and for apartments, floor level and existence of a lift (ground floor, first, second, third, fourth or more no lift, fourth or more with lift),[22] existence of a service room (zero, one or more). For houses, the number of levels (one, two, three or more), the presence of a basement and the surface of the plot are also known. The rate of non-response varies among variables (Table 2). In case of non-response, either the sale does not enter the computation, or the characteristic is imputed (Table 3).

---

[18] New dwellings are submitted to VAT, which is lower than stamp duty. The first sale of a new building if it takes place after five years is no more under the VAT regime, thus enters the scope of the indexes.

[19] Only the amount of stamp duties collected in each of the 95 French *départements* is known and since the tax rate no more distinguishes dwellings from other real estate, a coverage rate can only be estimated. The notaries themselves are in the process of collecting statistics on their activity, but it is for the moment done only in the Paris area, and is partial.

[20] It is lower than 30% in six *départements*, over 50% in 69 *départements* and over 70% in 36 *départements*. There are some 87 *départements* outside Ile-de-France. Corsica is not covered.

[21] The fact that the sales are a non random sample of the stock of dwellings, and that registration in the database is also potentially non random.

[22] Data analysis have been used to group or interact some characteristics.

Table 2

Non-response rates in %

Zone of the index	Surface	Number of rooms	Age	Garages car park	Bathrooms	Floor or number of levels	Lift
Province							
House	40.2	6.3	27.4	40.6	11.5	8.6	–
Apartment	9.0	1.8	25.1	56.2	7.0	4.6	58.9
Paris apartment	25.9 (7.7)	0	82.0 (8.1)	0.4	0	0.9	48.8
Petite Couronne apartment	46.1	0	23.5	0	0	3.4	59.3
Grande Couronne apartment	14.4	0.7	19.4	0	0	1.9	47.7
Ile-de-France house	47.0	0.2	51.5	0	0	0.3	–

Note: Computed on the reference stocks.

In some cases the missing information can be recovered from sales of other dwellings in the same building at another date. Then the non-response rates above are upper limits of the final rates which appear in parenthesis. In Ile-de-France, for garages and bathrooms the rate is low, but there is a confusion between “no bathroom” (or “no garage”) and non-response.

Table 3

Treatment of non-responses

Non-response	Action	Method, if imputation
Price	Rejected	
Surface and number of rooms	Rejected	
Surface	Imputed	Reference period mean of dwellings of same number of rooms in the commune or, if unknown, in the region; econometric imputation in Petite and Grande Couronne
Number of rooms	Imputed	From the surface at national level.
Type of dwelling	Rejected	
Lift	Imputed	Non-response = yes
Floor	Imputed	Ground floor
Bathroom	Imputed	No bathroom
Garage, parking	Imputed	No garage, parking
Date of construction		“Non-response” isolated
Nature of buyer	Imputed	Private individual or SCI
Nature of seller		Non-used
Destination of dwelling	Imputed	Habitation, full property
Surface of plot (for houses)	Rejected	

2.2 A quarterly INSEE supervision

The notaries collect the data and compute the indexes at their own cost. The question of the cost is not dwelt upon here. I only mention that by-products of the index computation are or could be sold. They go from part of the database, statistics on buyers and sellers, to a complete valuation system of dwellings and expertise on real estate prices.[23] The indexes themselves are publicly available and free.

INSEE is answerable for the index method, but does not compute the indexes. Hence a procedure of quarterly quality verification of the main indexes has been established. It relies on information on the gathering of the data (time of integration in the databases, quality controls) and on the comparison of the evolution of means prices and indexes, to detect structural modification. Volumes of sales, their structure by dwelling type (typically the number of rooms) are followed and compared to the reference stock. Zones with extreme variations of price or volume compared to the preceding quarter or to other zones are also detected and checked for potential errors.

2.3 The currently published indexes

Nine indexes are currently published at the national level. Seven are indexes for apartments: Paris, Petite Couronne, urban units of more than 10,000 inhabitants (city centres and suburbs), and small urban units and rural areas. One is a house index for the Province. The indexes for apartments in the Parisian Grande Couronne and houses in Ile-de-France will be added in Spring 2004. In some urban units or districts with enough sales, local indexes are also computed, but not published by INSEE. They might be in the near future.[24] Not waiting for the entire coverage of the national territory, INSEE estimates, a quarterly national index for "France". It will be replaced by one estimated on all sub-indexes "Notaires-INSEE" as soon as all are available and published, in the summer of 2004. All indexes can be found in a paper INSEE publication Bulletin Mensuel de Statistique (BMS), in January, April, July, and October, as well as on INSEE website.

Each published index is identified in the BMS by a code. They can also be found at http://www.indices.insee.fr (Indices et séries statistiques, Construction Logement, Indices trimestriels des prix des logements anciens). For each quarter t two new indexes are produced: a provisional index for $t - 1$ and a revised final index for $t - 2$.[25] The main indexes are presented in Figure 1.

---

[23] Quarterly or annual press conferences held by the notaries of the Paris area are available at http://www.paris.notaires.fr. For the rest of France see, http://www.immoprix.com/

[24] See Appendix A2 for details.

[25] The base 100 of the indexes was fixed at the second quarter of 1994 for Paris, at the fourth quarter 1994 for the Province.

Figure 1

Main notaires INSEE housing price indexes

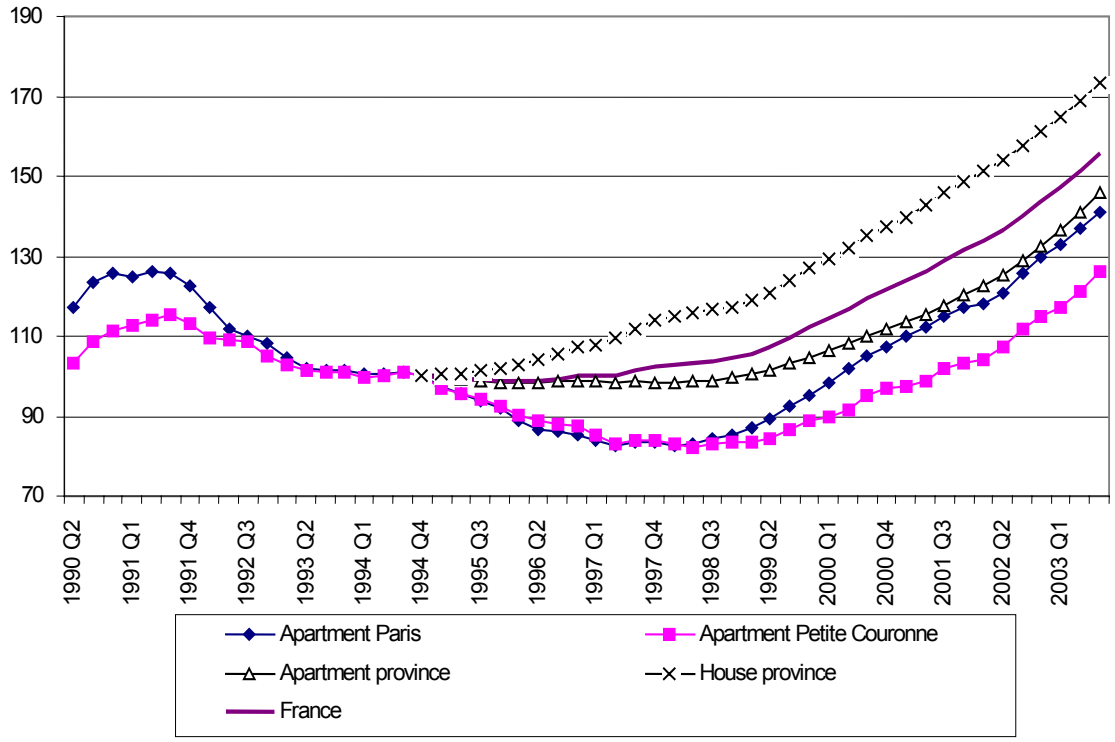


Table 4

Aggregate indexes

Code			Type of index
Website	**Paper**	**BMS**	
081767865	00000	00	France
			Ile-de-France
067517858	11000	10	Paris - apartments
080557385	12000	10	Petite Couronne - apartments
			Grande Couronne - apartments (in March 2004)
			Ile-de-France - houses (in March 2004)
			Province
080557486	20000	00	Province
080557587	20000	10	Province - apartments
067517959	20000	20	Province - houses
067517959	21000	10	Urban units above 10,000 inhabitants - apartments
067518060	21100	10	Urban units above 10,000 inhabitants - city centre - apartments
067518262	21200	10	Urban units above 10,000 inhabitants - suburbs - apartments
080557688	22000	10	Urban units less than 10,000 inhabitants and rural - apartments

3. Conclusion

Thanks to the propitious conjunction of sales data, goodwill and accurate methodology, reliable housing price indexes now exist for France. All three elements are necessary and it is important that they persist over the long run. The data should go on being collected, that is the notaries have to settle on a durable way of funding them. The tax authorities are unifying and computerising the real estate sale documents. A side effect is likely to be a reduced cost for data gathering and better quality. But the information necessary for the hedonic models, and not requested by tax authorities, has to be provided for the index and the hundreds of small notary practice have to be motivated. This leads to the second element, goodwill. It is fuelled by information about the use of the indexes. To the notaries, they should become a trademark, and the valuation system that goes with the indexes should prove useful, and a means to make the enterprise profitable. On the INSEE and academic side, and for the general public, the mere existence of reliable indexes and of all the information that goes with them, has begun to fuel new types of studies. As prices can be compared both in space and over time, microeconomic models of agents' decisions can incorporate them, and provide more reliable guidelines to public and individual choices. As housing and more generally real estate prices and consumer prices evolution can differ widely, it is of the utmost importance for economic policy to make use of both. As for methodology, its unique feature is the use of the valuation of reference parks at a detailed geographic level. Besides, the assumption of time stability of the models implies that there is no more econometric in the current period computation of the index, which saves time and cost. The first revision of period base, reference stock and model specification just took place. It proved not too difficult, and to have no major effect on the index profile, comforting the long term resilience of our hedonic methodology.

Appendix

A1. The time stability of the models

The models assume that the temporal effect is captured by the term $\sum_{a=1}^{3} \alpha_a Y_{a,i} + \sum_{t=1}^{4} \theta_t T_{t,i}$ and that the coefficients $\hat{\beta}_k$ are time invariant during the years following the estimation period, as the same coefficients are applied at each quarter to compute the index. The time invariance assumption of the $\hat{\beta}_k$ was checked. It was verified that the difference between the estimated value of dwellings with characteristic X_k and their actual sale price, the residual u_i, satisfied the stochastic assumption of the model, and did not include an unobserved deterministic component. The time evolution of the mean of the residuals in some zones was computed for each of the coefficients $\hat{\beta}_k$, $k = 2$ to K. They were found stable over time.

After a maximum of five years, they are checked and changed if necessary. This has just been done, and found to have no major effect on the index profile.

A2. Choice of zones/strata and sub-indexes

It is important to estimate the hedonic models on homogeneous price zones, that is zones where prices are not too different, and move in the same way over time. The strata of the publication of an index are not necessarily homogeneous. This is why they have been cut or grouped. The neighbourhoods have been defined locally by interviewing real estate experts. Then tree analysis allowed to group similar neighbourhoods. Ideally a model could be estimated in each neighbourhood, but as estimations are costly, there is only one model by zone, but dummy variables for particular neighbourhoods were added when necessary.

Indexes at a detailed geographical level are not yet published by INSEE but they are available from the notaries. The zones are the strata, that is the geographical level at which the hedonic models are estimated.

In Paris a zone can be one arrondissement, but also a group of arrondissements or a part of a larger one. In the rest of Ile-de-France, the strata are subdivisions of the five *départements* of the Petite and Grande Couronne (listed in the footnote of Table 1).

In the rest of France, a city centre has no proper apartment index if the annual number of sales is less than 150. Fifty-three zones are made of a single city. They range from Lyon, Marseille and Toulouse to Bayonne, Vichy and Annemasse. For the 33 largest a provisional index can also be published. Outside those large cities, 31 other zones are made of cities grouped by region interacted with four urban unit size (10,000-20,000, 20,000-50,000, 50,000-100,000 and over 100,000 inhabitants). There are 20 administrative regions grouped here into ten larger units. For the “suburbs” of cities, 21 zones are made of the suburb of a single urban unit (from Lyon, Toulon and Grenoble, to Dijon, Toulouse and Tours), and eight zones are groups of suburbs in groups of regions. The zoning for the indexes of apartment prices in rural areas is more heterogeneous. Seven main zones are defined (“Rural Grand Nord”, “Rural Est”, “Rural and Littoral, Grand Ouest”, “Rhône-Alpes, outside ski resorts”, “Rhône-Alpes, ski resorts”, “Languedoc Roussillon”, “Provence-Alpes-Côte d’Azur”), each comprising from two to 10 smaller zones defined by interacting regions and size of the urban unit. Note that for the Alpine regions, the ski resorts are isolated, because their housing market is very different from the surrounding areas.

For houses price indexes, the zones are different because the cities with an index for apartments do not all have enough house sales to compute their own index. Houses are traditionally numerous in cities of the North of France, hence Calais, Douai, Lens and Valenciennes have a house index but not an apartment index, while Chambéry, Annecy or Blois have not. All in all, only 47 cities have a house price index. Outside those cities, zones for houses price indexes are defined by interacting regions and size of urban units (rural villages, suburban zones, urban unit under or over 20,000 inhabitants).

A3. The hedonic models

There are as many hedonic models as elementary price zones. Two examples are presented here for apartments in zone 13 of Paris, zone two of Hauts-de-Seine in Petite Couronne. The dependent variable is the logarithm (base 10) of the price per square metre in Francs. The estimated coefficients, while being close from one zone to the other, differ to some extent. For instance a garage brings more added value in Paris than on the outskirts of the city. This change in the relative price of characteristics justifies the use of the hedonic method at a decentralised geographical level. It is equivalent to relate all variables with a zone dummy.

The hedonic regressions quality as measured by R^2, varies between 0.152 and 0.789. For individual cross-section data values of R^2 in the range of 0.25-0.40 for 1,000-3,000 observations and around 20 variables are considered good. This is what is obtained in most zones.

Table A1

Examples of hedonic regressions: apartments, Paris, zone 13

$R^2 = 0.42$, nb obs: 5520

Variables	Coefficients	Standard error	p-value
Constant	4,120	0,007	0,000
Year 1992	0,119	0,004	0,000
Year 1993	0,076	0,004	0,000
Year 1994	0,078	0,004	0,000
Year 1995	0,045	0,004	0,000
Year 1996	Reference		
Quarter one	0,019	0,004	0,000
Quarter two	0,016	0,004	0,000
Quarter three	0,013	0,003	0,000
Quarter four	Reference		
Neighbourhood one	Reference		
Neighbourhood two	–0,015	0,003	0,000
Neighbourhood three	0,027	0,003	0,000
Before 1850	–0,042	0,016	0,006
1850-1913	Reference		
1914-47	–0,011	0,004	0,002
1948-69	0,009	0,004	0,017
1970-80	0,053	0,005	0,000
After 1981	0,094	0,011	0,000
Unknown	0,015	0,005	0,003
No bathroom	–0,075	0,003	0,000
One bathroom	Reference		
Two bathrooms or +	0,027	0,006	0,000
No garage	Reference		
One garage	0,072	0,004	0,0000
Two garages or more	0,098	0,015	0,000
Service room	0,066	0,006	0,000

Table A1 (cont)

Examples of hedonic regressions: apartments, Paris, zone 13

$R^2 = 0.42$, nb obs: 5520

Variables	Coefficients	Standard error	p-value
Ground floor	Reference		
Floor one	0,045	0,006	0,000
Floor Two	0,076	0,006	0,000
Floor three	0,076	0,006	0,000
Four or more with lift	0,091	0,005	0,000
Four or more no lift	0,037	0,008	0,000
One room	0,016	0,004	0,000
Two rooms	Reference		
Three rooms	0,008	0,004	0,018
Four rooms	0,020	0,004	0,000
Five rooms or more	0,020	0,006	0,001
Small surface/room	0,007	0,003	0,030
Average surface/room	reference		
Large surface/room	–0,007	0,003	0,040

Table A2

Apartments, Haut-de-Seine, zone 2

$R^2 = 0.48$, nb obs: 2476

Variables	Coefficients	Standard error	p-value
Constant	4,0349	0,0109	0,0000
Year 1992	0,1118	0,0069	0,0000
Year 1993	0,0736	0,0065	0,0000
Year 1994	0,0748	0,0063	0,0000
Year 1995	0,0307	0,0065	0,0000
Year 1996	Reference		
Quarter one	0,0194	0,0059	0,0010
Quarter two	0,0182	0,0059	0,0020
Quarter three	0,0047	0,0060	0,4282
Quarter four	Reference		
Neighbourhood one	Reference		
Neighbourhood two	–0,1275	0,0045	0,0000
Before 1850	0,1013	0,0233	0,0000
1850-1913	Reference		

Table A2 (cont)

Apartments, Haut-de-Seine, zone 2

$R^2 = 0.48$, nb obs: 2476

Variables	Coefficients	Standard error	p-value
1914-47	0,0105	0,0059	0,0736
1948-69	0,0100	0,0074	0,1772
1970-80	0,0241	0,0100	0,0158
After 1981	0,0904	0,0123	0,0000
Date unknown	0,0111	0,0061	0,0695
No bathroom	–0,0540	0,0047	0,0000
One bathroom	Reference		
Two bathrooms or more	–0,0245	0,0140	0,0795
No garage	Reference		
One garage	0,0565	0,0083	0,0000
Two garages	0,0559	0,0180	0,0019
Service room	–0,0012	0,0178	0,9463
Ground floor	Reference		
First floor	0,0402	0,0086	0,0000
Floor two	0,0518	0,0084	0,0000
Floor three	0,0475	0,0086	0,0000
Four or more with lift	0,0628	0,0079	0,0000
Four or more no lift	0,0185	0,0107	0,0838
One room	0,0028	0,0069	0,6805
Two rooms	Reference		
Three rooms	0,0265	0,0054	0,0000
Four rooms	0,0397	0,0084	0,0000
Five rooms or more	0,0482	0,0113	0,0000
Small surface/room	0,0066	0,0059	0,2624
Average surface/room	Reference		
Large surface/room	0,0305	0,0084	0,0003

Reference

David, Alain, François Dubujet, Christian Gouriéroux and Anne Laferrère (2002): “Les indices de prix des logements anciens”, *INSEE Méthode*, 98, p 119.