INSEE MÉTHODES

N° 128 – July 2014

THE NOTAIRES-INSEE HOUSING PRICES INDEXES

CONTENTS

Took part in this work

Philippe Clarenc
Jean-François Côte
Alain David
Jacques Friggit
Philippe Gallot
Stéphane Gregoir
Anne Laferrère
Adélia Nobre
Catherine Rougerie
Nelly Tauzin

Publication director
Jean-Luc Tavernier

ISSN : 1142-3080

Abstract

Measuring the changes in housing prices is a prerequisite to the proper functioning and transparency of these markets. In conjunction with the French civil-law notaries (*notaires*), INSEE has developed a method to produce regular and reliable price indices of used dwellings, the *Notaires-INSEE* housing price indices. It was first introduced in *INSEE Méthodes* no.98 (David *et al*., 2002), then in a revised version, "version 2", in *INSEE Méthodes* no. 111 (Beauvois *et al*., 2005). The latest *INSEE Méthodes* presents a new version, "version 3", of the *Notaires-INSEE* housing price indices.

Housing price indices are difficult to compute because the price of a given dwelling is seldom observed and because housing quality can change over time. The method used to overcome this difficulty involves econometric models (hedonic models), based on the estimation of relative prices of housing characteristics. In each primary area, a model is estimated from a housing stock, called the *estimation stock*; it is assumed that the model will apply for a few years. Thus in each current period, the value of a fixed housing stock, called the *reference stock*, can be estimated from all the transactions observed during that period, by averaging the reconstructed prices of a corresponding reference property. The index consists of the ratio of the estimated value of the reference stock in the current period to its value observed in the base period.

The data are supplied by notaries (*notaires*), who draw up contracts for sales of existing dwellings in France. These data currently cover about two thirds of all sales. Housing characteristics include floor space, number of rooms, number of bathrooms, floor in a building, whether there is a lift, a garage, etc.

As was originally intended, the initial models were revised after four years. This revision produced "version 2" of the indices, in force from July 2004 to October 2011. Again in accordance with the original commitment, a further revision was carried out (version 3), in force since November 2011. When new indices are introduced, this is accompanied by the more systematic publication of indices that have been adjusted for seasonal variations, in order to monitor market trends better.

This version comprises 293 basic hedonic models, estimated for geographic areas (strata), which are homogeneous in terms of prices. Separate models are used for houses and apartments: 15 models for apartments and 7 for houses in the Paris region (Île-de-France), 97 for apartments and 174 for houses outside the Paris region (hereafter the "Provinces").

The indices cover all of Metropolitan France and have been given the official "Notaires-INSEE" designation; production of the indices is monitored by a scientific board. A provisional version is published every quarter, two months after the end of the quarter in question, then a definitive version appears six months after the end of the quarter (five months for the Paris region indices). These quarterly indices of the prices of used dwellings in the Paris region have received official approval from the French Public Statistics Authority (ASP).

The present document describes the various improvements introduced into the calculation method and includes detailed appendices on the technical aspects of the new models.

Chapter 1: Introduction

Information and transparency are essential conditions for the smooth running of a competitive market. It is therefore crucial that economic actors have reliable information at their disposal on house prices and how these are changing, particularly as housing plays an essential part in the economy, in terms of household production, assets, budget and debt.

Housing for sale may be new (never been occupied) or used ("second-hand"). Since January 2013, a price index of used dwellings has been available on the INSEE website www.insee.fr as well as an index of all dwelling prices (new and used). *The present document covers only indices of used dwellings.*[1]

The price of a dwelling is seldom observed

The production of a housing price index raises the same problems as any other price index: how to separate "pure" price changes from changes in housing quality? But there are specific problems too.

Firstly, no two dwellings are ever exactly alike. A dwelling is a combination of characteristics (also referred to as "qualities") which only when taken together constitute the "housing good". A dwelling is not only a geographic location (and hence access to a neighbourhood defined by local public goods), but also a construction (defined by type, size and a multidimensional level of comfort) and often an emotional good, a place that holds family memories.

Secondly, a dwelling may change hands only infrequently. It is estimated that on average dwellings are sold once every 35 years[2] and undergo change, with or without financial consideration, every 25 years. Observation of the price of a given dwelling is therefore rare. This long time lag between purchase and resale of a given dwelling complicates the compiling of a price index. Some indices are based on the stated price or the asking price, others on estimates, but it is better to have access to market prices, i.e. prices of transactions actually completed.

How, then, should we measure the change in the price of a given dwelling, or more generally of a set of dwellings, even though only a few transactions are observed in each period?

Which prices are observed?

If we want to build an index of changes in the price of the entire housing stock, we cannot simply calculate the average current transaction price and compare it with that of the previous period. Such a comparison would mix price effects with changes in stock quality and in the non-representativeness of transactions. There are two parts to the problem: on the one hand, transactions observed in each period are not drawn at random from the housing stock; on the other, the housing stock changes and its quality does not remain constant from one period to the next.

Let us disregard the second of these problems and assume little or no change in the housing stock, i.e. that there is no new construction and that maintenance exactly offsets building wear and tear. If the transactions in a given period were in sufficient number and drawn at random from this fixed housing stock, then we could simply determine the average current transaction prices, compare them with that of the preceding period and so obtain an index of housing stock prices. If, however, as is often the case in practice, the transactions in a quarter are not a representative sample of the total housing stock, such a comparison would mix together changes in the quality of the sample with changes in prices.[3] To understand this, let us assume that the price of each dwelling remains absolutely constant over time; in a given quarter the majority of sales take place in the more prosperous

[1] In the years after the Second World War, new dwellings predominated in purchases of primary residences (and they were still in the majority at the beginning of the 1980s: 58% of transactions between 1980 and 1984, according to the 1984 Housing Survey). By 2006 they represented just under a third of purchases during the previous four years (source: Housing Survey 2006).

[2] This estimate disregards housing stock held by corporate bodies (mainly social housing bodies, but also institutional investors and real estate companies), which is usually sold in blocks. These average periods between two transactions are calculated on the basis of 800,000 transactions against payment and 330,000 transactions without financial consideration (where the reference person of the property-owning household changes) per year, reported in a stock of 28 million dwellings owned other than by a company (source: CGEDD).

[3] This problem concerning change in the quality of transactions observed from one period to the next is sometimes described as a problem of weighting. It has been well formulated by Triplett (1983) who makes a distinction between input-price and output-price indices.

neighbourhoods, then in the following quarter the majority of sales are at the lower end of the range. The average price will fall, but this will bear no relationship to the actual zero change, which is what a pure price index is designed to capture.

Changes in housing quality

Let us now turn to the second problem. Let us make the opposite assumption, namely that a sample of transactions representative of the housing stock is available at each date, but that this stock is not fixed. We cannot construct a price index in this case either, because such a comparison will combine a price effect and the effect of the change in quality of the stock as the older dwellings wear out or are destroyed and replaced by new ones.[4]

The problem posed by the change in quality is familiar to price-index producers. If the price of a light bulb goes up from 1 to 2 euros, but its life increases, its price appears to have doubled. However, if the consumer derives the same satisfaction from the long-life bulb as from two short-life ones, we will say that the quality has also doubled, and hence that the pure price has remained constant. For consumer goods, such changes in quality are common; sometimes comparing the goods themselves can be tricky when we are dealing with a new product.

When dealing with housing, the situation is both simpler but also more complicated. It is more complicated because, as noted earlier, quality cannot be measured according to a single dimension, like, for example, the lifetime of the light bulb. A large number of characteristics have to be taken into account.[5] This requires the use of "hedonic" econometric methods. However, the housing situation can also be simpler for the statistician to deal with because the number of technical revolutions is limited; over a given period of several years, changes in quality (defined as the emergence of a new dwelling characteristic) are slow to emerge. This might be a change in the number of bathrooms per dwelling or technical changes linked with compliance with new environmental standards.

We therefore make an initial assumption: the characteristics of dwellings that may influence their price, such as the quality of the neighbourhood, the number of rooms, whether there is a lift, which floor an apartment is on, etc. are finite in number and this number remains constant for the duration of the index-calculation period. If a new quality characteristic emerges to alter house prices, for instance the installation of air conditioners or mandatory asbestos testing, it will not at first be taken into account in the index estimation as a new quality, but as a price variation. In other words, we admit that consumer satisfaction is not yet truly influenced by air conditioning or asbestos testing, and that they consider this as a price increase. These changes in quality will be incorporated by the statisticians by adapting the specification of the models to regular intervals.

The hedonic method

Econometric methods are applied whenever a good is composite, whether this is in relation to the consumer price index (e.g. for computers, where their characteristics may change in value, but new characteristics do not appear as often[6]), the cost of a service (such as a bank,[7] or hospital), or labour market studies where workers' characteristics are heterogeneous and multidimensional.[8]

Hedonic methods were first developed in the United States between the two World Wars, in the midst of the Great Depression. According to official price indices estimated from average prices, automobile prices rose by 45% between 1925 and 1935 and strong pressure was being exerted on General Motors to reduce prices in order to maintain consumption and employment. So in 1939, Andrew Court, who worked for the Automobile Manufacturers Association and knew that the quality of the automobiles had also changed (safety glass, gear boxes, more powerful engines, etc.), developed a method to take into account the changes in characteristics of automobiles when calculating price indices.[9] Court defined a standard car, with given characteristics regarding speed, safety, windows, seat width, etc., which would serve as the benchmark for user comfort or *pleasure*.[10] He

[4] Curiously, this second problem receives little attention in the literature on housing prices.
[5] In fact, many consumer goods and durables also display this composite and multidimensional profile. We need only think of a computer, a plane trip or even a simple sugar cube whose packaging can differ.
[6] On the use of hedonic methods applied to computers, see Moreau (1996), Triplett (2004).
[7] In this case, it is rather the number of characteristics that changes.
[8] Zvi Griliches has done a considerable amount of work on this technique; see, for example, Griliches (1971).
[9] At the same time, the Bureau of Labor Statistics was trying to compare the qualities of tractors and trucks, in order to gauge what was quality change and what was price change.
[10] Hence the term "hedonic" to describe the method, which in reality is a perfectly commonplace econometric regression model.

then divided the average car price by this benchmark, i.e. its "hedonic content" and found that the price of the standard car had fallen by 55% in the period.[11]

Concerning housing prices, in addition to the first hypothesis (whereby each dwelling can be defined by the combination of a limited number of characteristics), we put forward a second hypothesis, whereby the relationship between the price per square metre of a dwelling and its characteristics remain fixed throughout the index-calculation period for a given type of dwelling and primary area. In a primary area, the housing price index is defined as the ratio of the value of a fixed reference housing stock during the current period to its value during the base period of the index. The value of each dwelling in this stock is estimated each quarter from transaction prices actually observed and econometric relations estimated over a specified time period.

Available data and implementation of the method

In France, real estate transactions are carried out before a *notaire* (hereafter: notary), a State-appointed public officer, who ensures that the deed of sale is recorded in the land register and collects certain taxes on behalf of the State. Although every deed of sale must include identification of the land base in the land tax register, this obligation does not extend to a description of the dwelling sold. Notaries are well aware of the importance of this information and for the last few years have made a concerted effort to collect this data.

In Paris, a price index for used apartments had been calculated since 1984 using a stratification method developed with INSEE. In 1997, the Higher Council of the Notariat (*Conseil supérieur du notariat*) decided to create a price index for dwellings in all areas outside Paris (hereafter: the Provinces). This was the perfect opportunity for INSEE to develop a rigorous method for calculating a price index for all used dwellings, following the path opened up by Court and Griliches.[12] This method was first introduced in 2002 in *INSEE Méthodes* no.98, *Les indices de prix des logements anciens.*[13] A second edition of this document was published in 2005.[14] The present volume (*version 3*) is therefore a second update.

Since the first version of the models was introduced, the collection of housing transaction data has improved: the number of transactions recorded has increased and so the indices can now be "pure" quarterly indices, unlike the original indices, which were calculated on an annual or semi-annual "period-over-period" basis. In addition, there are now more published indices which are made available more quickly: provisional indices are published about eight weeks after the end of the quarter, a similar delay to our counterparts in the United States and in most European countries. The definitive indices for the previous quarter are published at the same time[15]. Today, the interdepartmental chamber of Paris notaries (*Chambre interdépartementale des notaires de Paris - CINP*), through its own association, the *Paris notaires service* (PNS), calculates the indices for Paris every quarter from the gross data and then adjusted for seasonal variations. The company Min.not, a subsidiary of the ADSN group (*Association pour le développement du service notarial*), does the same for the Provinces and calculates the indices for the whole of France. Every year, INSEE calculates the seasonal adjustment coefficients and validates the quarterly indices before publication.

On 28 March 2011, the French National Assembly adopted law no. 2011-331 on the modernisation of the legal and judicial professions and some regulated professions. Articles 15 and 16 of this law establish that it is the responsibility of the notarial profession to carry out the public service consisting of collecting, transmitting, centralising and disseminating information on property transactions against payment (this includes not only sales of used dwellings, but also sales off-plan for new dwellings, sales of plots of land or sales of premises for non-residential use). The implementing decree was published on 3 September 2013.[16] The resulting improvements in coverage in the databases, should open the way to more comprehensive regional, departmental and local information, and in particular to the establishment of extra sub-national indices, modelled on those that are already published for the Rhône-Alpes and Provence-Alpes-Côte d'Azur regions and, more recently for Nord-Pas de Calais.

The indices in the new database (version 3), described in the present document, were first used towards the end of 2011. At the same time, seasonally adjusted indices were published more systematically, which gave a better view of market trends. In addition, the possibilities that have opened up with the development of electronic

[11] This story is told by Warsh (1999).
[12] The following worked on developing the method that was adopted: Alain David, François Dubujet, Christian Gouriéroux, Anne Laferrère and Claude Taffin.
[13] David *et al*. (2002).
[14] Beauvois *et al*. (2005).
[15] However, Corsica and the French overseas *départements* (DOM) are still excluded from the scope of the national indices.
[16] JORF no.0206 of 5 September 2013 page 14976, text no. 2, Decree no. 2013-803 of 3 September 2013 concerning notarial databases relating to real estate transfers against payment.

transmission of deeds by the notarial offices should reduce the time needed to process them. The quarterly indices of used dwelling prices in the Paris region were approved by the Public Statistics Authority (*Autorité de la statistique publique - ASP*) in June 2011.[17] This Authority checks that the principle of professional independence is complied with in the design, the production and the dissemination of public statistics and gives its quality certification to statistics produced for the general interest, by administrations, public and private bodies responsible for providing a public service. The indices covering the Provinces may also be checked by ASP subsequently.

Various extensions

The method used to calculate the housing price index can have various extensions.

One application is to create an expert system to assess housing prices. The aim is to see whether a property that is put up for sale is over- or under-priced compared with another property that is used as a reference. In addition, the tool may use the coefficients obtained when the hedonic equation was estimated to transform the price of the transaction into a reference-property equivalent price.

Another application could be to calculate price indices according to the modalities of a variable of interest. Initial investigations are underway and the possibility of defining indices according to the number of rooms in the dwelling is currently being examined.

Lastly, for a more reactive knowledge of the market, it has been decided that a new generation of indices will be created, so-called “advance” indices of the prices of used dwellings. In this case, instead of using transfers that have taken place in the course of a given quarter, it is the pre-contract sale agreements that will be considered. In fact, most pre-contract agreements become definitive transactions so these can give a good indication of market trends.

Content of this volume

After this introductory chapter, a second chapter sets out the theory of hedonic indices. Chapter 3 discusses the calculations applied and the new features in this second update. Chapter 4 describes the databases used. In Chapter 5, the method for estimating the hedonic equation is described, with some examples. Chapter 6 deals with seasonal adjustments to the indices, tools for quarterly tracking, and publications. Lastly, in Chapter 7, we compare the series of indices obtained in version 2 with those obtained in version 3.

[17] The Official Statistics Autthority (ASP) was created by Act no. 2008-776 of August 4, 2008. it ensures that oficial statistics are prepared and disseminated in all professional independence and according to the fundamental principles in the European Statistics Code of Pratice : impartiality, objectivity, relevance and data quality (art. 144). It notified the labeling of the quarterly statistics on housing prices in the Paris region (Île-de-France), jointly produced by the Paris Chamber of Notaries and INSEE, for a period of five years, in its opinion no. 2011-01 of June 21, 2011.

Chapter 2: Theory of hedonic indices

2.1 The rationale for hedonic approaches

The categories in a classification of goods and services, even if restricted, include some widely varying sets of characteristics. Dwellings, for instance (the goods being examined here) are defined by their location, their floor space, the number of rooms, whether occupied or vacant, age, and so on. Housing rentals (services) are differentiated further by lease durations, renewal terms, and early termination clauses.

This variety is reflected in the corresponding markets, leading to differences in housing turnover or occupation rates, and widely differing prices and this in turn can pose various problems for price analysis.

The main difficulty is due to the fact that the price of a dwelling can only be observed when the transactions take place, and these are infrequent (partial observability); similarly, a rent is observable only if the service is actually used. Outside of these situations, these values – prices and rents – have no existence in economic terms.

Traditionally, the way to get round this problem is to assume *implicit values*, also called, in our context, the "estimated price" of the dwelling and the "rental value". These implicit values are known only when they coincide with transaction prices or rents, and cannot therefore be reconstructed except from models describing the composition of the prices or rents, and any changes that occur.

Hedonic approaches rely on models such as these and explain how to use them in order to construct the non-observed values and define consistent sets of price indices.

2.1.1 Hedonic model with a priori *stratification*

We shall present the hedonic approach using a simplified formulation. Let us assume that the different goods can be aggregated into strata pre-defined in such a way that price changes are approximately parallel within the same stratum. Between strata, however, prices may differ considerably.

In our application, the strata will be geographic areas, the primary areas where dwellings are located. These strata are denoted *s, s=1,…,S.* Given a dwelling *i* in stratum *s*, with characteristics $z_{i,s,t}$ (floor space, number of rooms, etc.), we use a regression model to estimate its implicit value $p^*_{i,s,t}$ on date *t*.

Within the stratum *s*, the implicit values are assumed to be such that:

$$p^*_{i,s,t} \approx c(s, z_{i,s,t})\, p^*_{s,t} \qquad (2.1)$$

- where $p^*_{s,t}$ is an implicit reference value for stratum *s* on date *t*,
- $c(s,z_{i,s,t})$ is an adjustment coefficient that takes into account the characteristics of the good and which may be stratum-dependent,
- $\approx$ means "approximately equal to".

The *adjustment* and the *reference value* are defined to within one multiplicative scalar (which poses an identification problem). It is then possible to set the reference value as corresponding to a property of pre-specified quality (reference dwelling), z_0:

$$p^*_{0,s,t} \approx p^*_{s,t} \Leftrightarrow c(s, z_0) = 1$$

let us say, for example, a two-room, used apartment on the ground floor, etc.

$\dfrac{p^*_{i,s,t}}{c(s, z_{i,s,t})}$ will be called the "reference-property equivalent price" in terms of dwelling (i,s) on date t.

The approximation (2.1) can be made to resemble an econometric model more closely by introducing appropriate error terms and specifying a parameterized form for the adjustment coefficient. Such a specification might be, for example:

$$\log p^*_{i,s,t} = \sum_{k=1}^{K} \beta_{k,s} X_k\left(z_{i,s,t}\right) + \log p^*_{s,t} + \varepsilon^*_{i,s,t} \qquad (2.2)$$

where the error terms $\varepsilon^*_{i,s,t}$ are assumed to be independent, zero-mean and of variance $\eta^2_{s,t}$ in some cases stratum- and date-dependent. The $X_k\left(z_{i,s,t}\right)$ values are explanatory variables, *K* in number, which are functions of housing characteristics or of combinations of certain characteristics.

Formula (2.2) above has the advantage of expressing a linear model in the parameters $\hat{c}_0(s,z) = \exp\left[\sum_{k=1}^{K} \hat{\beta}_{k,s} X_k(z)\right]$, which will enable us to define the adjustment coefficient, and $\log p^*_{s,t}$, which will then give the stratum reference value.

Note that the model incorporates combined effects of the stratum and other housing characteristics, $\beta_{k,s}$ coefficients being stratum specific.

2.1.2 Problematic alternative approaches

Can one reconstruct price changes without using a hedonic model? Two alternative approaches are possible: one is the repeat-sales method, but this is difficult to implement rigorously, and the other is based on observation of average prices, but produces biased results.

Although a given property *i* is seldom traded on two given successive dates *t*, *t+1* (in which case we would speak of repeat data), we can nevertheless hope to measure its change in price by comparing prices of two similar properties. If the explanatory variables $X_k(z)$ are qualitative, for example, there may be several properties in the stratum with approximately the same price levels and not only parallel changes. These prices will be used to calculate price change. However, once the quality effects reach a certain intensity and number, the market may not be liquid enough for such a comparison to be possible. Furthermore, if we base our calculations solely on repeat sales prices, we neglect a large share of the information contained in transaction data.

A second approach often suggested is to compare the average price of properties in stratum *s* traded at *t+1*, with the average price of properties in the same stratum traded at *t*, in the hope of proxying change in the reference value. This approach is biased, however, as the quality structure of traded properties is not stable over time. To illustrate this difficulty, let us consider a case in which only a single property is traded on each date in the stratum, with the property noted as being of quality z_t at date *t*. The observed price ratio would be approximately:

$$\frac{c(s,z_{t+1})}{c(s,z_t)} \frac{p^*_{s,t+1}}{p^*_{s,t}}$$

and would differ from the change in reference value because of the change in the adjustment coefficient.

2.2 Using the hedonic model

The model is essentially used to compensate for the partial inobservability of the data and to adjust for quality effects. The process consists of several steps, as outlined below.

- Step 1: estimation of adjustment coefficients using a predefined set of transaction data, hereafter called *estimation stock*. Choice of a set of dwellings, called the *reference stock*, whose prices will be tracked.
- Step 2: at each date *t*, use of actual transaction data and adjustment coefficients estimated in step 1 to reconstruct the values of reference stock dwellings.
- Step 3: use of reference values and adjustment coefficients to construct an array of price indices and an expert valuation system.

2.2.1 Estimating adjustment coefficients

The model is mainly used to consider an estimation stock consisting of transactions occurring in a predefined period $t=1,\ldots,T_0$ (called *estimation period*). The transaction data provide price-quality pairs (price $p_{j,s,t}$, quality $z_{j,s,t}$), $j=1,\ldots, J_{s,t}$, $s=1,\ldots,S$, $t=1,\ldots,T_0$, where $J_{s,t}$ denotes the number of transactions in stratum s in period t.

We then estimate parameters $\beta_{k,s}, k=1,...,K$ in each stratum, using ordinary least squares, from equation (2.2). From this we deduce the estimated adjustment coefficient of the stratum:

$$\hat{c}_0(s,z) = \exp\left[\sum_{k=1}^{K} \hat{\beta}_{k,s} X_k(z)\right]$$

and also, if the precision of the estimated coefficients allows, a value range for each of these terms, :

$$\left[\hat{c}_1(s,z), \hat{c}_2(s,z)\right]$$

These adjustments will remain unchanged throughout a future period, whose length has to be specified.

2.2.2 Estimating reference value for date t

Let us now consider a date t, subsequent to the estimation period and transaction data: $p_{j,s,t}$, $z_{j,s,t}$, for each of the properties j in stratum s on date t. We have:

$$\log p_{j,s,t} = \log c\left(s, z_{j,s,t}\right) + \log p^*_{s,t} + \varepsilon_{j,s,t}$$

$$\cong \log \hat{c}_0\left(s, z_{j,s,t}\right) + \log p^*_{s,t} + \varepsilon_{j,s,t}$$

after replacing the adjustments by their proxies obtained in the first step.

We obtain $J_{s,t}$ estimations of the reference property price in stratum s on date t, $p^*_{s,t}$, and we deduce the approximation of $p^*_{s,t}$ using ordinary least squares:

$$\log \hat{p}^*_{s,t} = \frac{1}{J_{s,t}} \sum_{j=1}^{J_{s,t}} \left[\log p_{j,s,t} - \log \hat{c}_0\left(s, z_{j,s,t}\right)\right]$$

$$\Leftrightarrow \hat{p}^*_{s,t} = \prod_{j=1}^{J_{s,t}} \left[\frac{p_{j,s,t}}{\hat{c}_0\left(s, z_{j,s,t}\right)}\right]^{\frac{1}{J_{s,t}}}$$

$p^*_{s,t}$ is thus estimated from the geometric mean of the reference-property equivalent prices. By taking standard deviations into account, we can also propose a range for the reference value. More specifically, note that:

$\hat{\eta}^2_{s,t}$ is the empirical variance of the values $\log\left(p_{j,s,t} / \hat{c}_0\left(s, z_{j,s,t}\right)\right), j=1,...,J_{s,t}$.

This empirical variance proxies variance $\eta^2_{s,t}$ of the error term. The reference value admits a logarithm lying between $\log \hat{p}^*_{s,t} - 2\hat{\eta}_{s,t}$ and $\log \hat{p}^*_{s,t} + 2\hat{\eta}_{s,t}$, and the range is:

$$\left(\hat{p}^*_{1,s,t} = \exp\left(-2\hat{\eta}_{s,t}\right)\hat{p}^*_{s,t}\ , \hat{p}^*_{2,s,t} = \exp\left(2\hat{\eta}_{s,t}\right)\hat{p}^*_{s,t}\right)$$

2.2.3 Constructing a valuation system

We can now estimate the implicit values of any property at date *t*, of quality z, which is not necessarily traded.

The estimate is:

$$\hat{p}^{*}_{s,t}\hat{c}_0(s,z)$$

the range can be taken as equal to:

$$\left(\hat{p}^{*}_{1,s,t}\hat{c}_1(s,z),\ \hat{p}^{*}_{2,s,t}\hat{c}_2(s,z)\right)$$

2.2.4 Constructing an array of indices

The hedonic method has enabled us to define primary indices stratum by stratum:

$$I_{s,t} = \hat{p}^{*}_{s,t}, s = 1,...,S,\ t = 1,...,T$$

which can serve as a base for constructing composite indices. This construction uses traditional methods, such as a Laspeyres-index approach. For this, we define a basket of properties (housing stock), called the reference stock, from the initial period, specifying the qualities *z* and strata *s* concerned. *Z* denotes the different qualities introduced in the basket (reference stock) and $N_{z,s}$ the number of properties (dwellings) with characteristics *z* in stratum *s*.

the composite index is defined from the value of this basket:

$$I_t = \sum_{s=1}^{S}\sum_{z\in Z} N_{z,s}I_{s,t}$$

This basket can also serve to construct composite indices disaggregated in a coherent manner. For example, we can introduce a "two-room" index [French *deux pièces*] and consider the sub-stock composed only of two-room dwellings:

$$I_t(deux\,pièces) = \sum_{s=1}^{S}\ \sum_{z\in Z\ (deux\,pièces)}\ N_{z,s}\ I_{s,t}$$

or an index for a given region, calculated on a sub-set of strata:

$$I_t(région) = \sum_{s\in région}\ \sum_{z\in Z} N_{z,s}I_{s,t}$$

For a standardised approach, the convention is to take a base year, say *t = 0*. The standardisation should then be carried out index by index, giving indices with base *100* at *t = 0* written as:

$$I_{t/0} = 100\ I_t / I_0$$
$$I_{t/0}(deux\,pièces) = 100\ I_t(deux\,pièces) / I_0(deux\,pièces)$$

2.3 Making the hedonic approach more robust

The hedonic approach, which relies on estimates, can be sensitive to their precision or to parameter instability. It may be useful to aggregate some quality variables or strata in order to make the results more significant.

2.3.1 The search for underlying scores

Let us consider the adjustment coefficients. For each stratum, we have estimated a set of parameters $\hat{\beta}_{1,s},...\hat{\beta}_{K,s}$ defining the stratum *score* $s : \sum_{k=1}^{K} \hat{\beta}_{k,s} X_k$.

We can examine whether these *S* scores depend on a smaller number of underlying scores. This approach is as follows:

1. We define the matrix $\hat{B}$, of size *(K,S)*, whose columns are the vectors $\left(\hat{\beta}_{1,s},..., \hat{\beta}_{K,s}\right)$;
2. We determine the eigenvalues $\hat{\lambda}_1 \geq \hat{\lambda}_2 ... \geq \hat{\lambda}_S$ and the associated eigenvectors $\hat{\alpha}_1,...,\hat{\alpha}_S$ of matrix $\hat{B}'\hat{B}$, where $\hat{B}'$ denotes the transposition of $\hat{B}$.
3. The number of eigenvalues S_0 significantly different from zero yields the number of independent underlying scores which are given by:

$$Z_l = \hat{\gamma}_l' X = \hat{\alpha}_l' \hat{B}' X, l = 1,...,S_0$$

4. To make the model more robust, we constrain the adjustment coefficient to take the form:

$$c(s,z) = \exp\left[\sum_{l=1}^{S_0} \lambda_{l,s} Z_l\right]$$

As S_0 is less than both *K* and *S*, and often fairly small, there are far fewer parameters to estimate in this constrained form of adjustment coefficient.

Even if it requires a partial disaggregation of the scores Z_l deduced from the approach above, it is customary to select sub-scores that contain only variables with interpretations of the same type. For example, one sub-score will adjust for the physical characteristics of the dwelling, a second for its amenities, a third for the quality of the environment, a fourth for location, and so on. We thus obtain a hierarchical structure of the effects of the variables, making it easier to set up, interpret and update expert valuation systems.

2.3.2 Strata aggregation

Similarly, we can examine the strata via changes in the corresponding indices $I_{s,t}$. By analysing empirical correlations between these time series, we may be able to identify strata whose reference values are moving in parallel. If the result is interpretable, we can then aggregate those strata.

2.4 Monitoring the specification

Parameter values may change over time and adjustment coefficients determined in the estimation period may deteriorate. It is important to set up instruments to monitor the quality of the model so as to be able to identify the point at which we need to re-estimate it, and also to develop ideas about any changes that should be made. We can then go on to a suitable examination of estimation residuals.

At each date *t* the residuals are:

$$\hat{\varepsilon}_{j,s,t} = \log p_{j,s,t} - \log \hat{c}_0\left(s, z_{j,s,t}\right) - \log p^*_{s,t}, j = 1,..., J_{s,t}$$

They must be aggregated in order to eliminate quality effects and steer the monitoring process towards the parameters $\beta_{k,s}$ which are liable to be affected. To this end, we can consider various marginal characteristics,

for example: two rooms (d.p. [for French *deux pièces*]), used dwelling (a. [for French *ancien*]), etc., and compute the mean residuals for each.[18]

$\hat{\varepsilon}_{s,t}(d.p.) =$ mean of $\hat{\varepsilon}_{j,s,t}$ values for two-room dwellings of stratum *s* at date *t*,

$\hat{\varepsilon}_{s,t}(a.) =$ comparable mean for used dwellings, etc.

If the model is properly specified, such means should vary around zero. We shall therefore search for recurring divergences. For example, if we find $\hat{\varepsilon}_{s,t}(a.)$ values that are too often positive for stratum s_0, from a particular date t_0, we may need to adjust the value of a parameter for an explanatory variable X_{k_0} as a function of dwelling age. If a trend emerges in the series $t \to \hat{\varepsilon}_{S_0,t}(a.)$ this may mean that the proportionality hypothesis for price changes within stratum *s* is no longer fulfilled and that this stratum needs to be decomposed.

2.5 Extending the basic model

The hedonic model used by way of illustration until now has a disadvantage in that it pre-defines homogeneous strata for price changes. We can generalise this model as follows.

Let us begin by writing formula (2.2) in a condensed form:

$$\log\ p^*_{i,s,t} = \sum_{s_0=1}^{S}\sum_{k=1}^{K} \xi_{s_0}(i,s,t)\beta_{k,s_0} X_k(z_{i,s,t}) + \sum_{s_0=1}^{S} \log\ p^*_{s_0,t}\ \xi_{s_0}(i,s,t) + \varepsilon^*_{i,s,t}$$

where ξ_{s_0} denotes the dummy variable of stratum s_0, i.e. the variable that equals 1 if the observation is in stratum s_0, otherwise 0. In this form, the model applies to all data in all strata, and remains linear in the various parameters:

$$\beta_{k,s_0}, k = 1,...,K,\ s_0 = 1,...,S,\ \log p^*_{s_0,t},\ s_0 = 1,...,S,\ t = 1,...,T$$

Its limitations clearly appear in the expressions of the explanatory variables: the "constant term" part includes highly specific stratum × quality cross-effects, whereas the parts giving the dynamics of the reference values $p^*_{s_0,t}$ do not incorporate quality effects other than the stratum dummy.

An enlarged model may thus be written:

$$\log p^*_{i,t} = c_0(z_{i,t};\theta_0) + \sum_{l=1}^{L} c_l(z_{i,t};\theta_l) f_{l,t} + \varepsilon^*_{i,t} \qquad (2.3)$$

where $f_{1,t},...,f_{L,t}$ are dynamic factors to be determined, $c_0(z_{i,t};\theta_0),...,c_L(z_{i,t};\theta_L)$ are adjustment coefficients giving, for example, sensitivities to factors, with parameters $\theta_0,...,\theta_L$ to be estimated. There is no longer any need to distinguish stratum *s* which is reincorporated among the other characteristics of the dwelling.

The hedonic approach can be applied from such a model. For example, in the current situation at date *t*, factor values will be estimated using ordinary least squares on the proxied model:

$$\log p_{j,t} \cong \hat{c}_0(z_{j,t}) + \sum_{l=1}^{L} \hat{c}_l(z_{j,t}) f_{l,t}$$

where the $\hat{c}_l(z_{j,t})$ values are determined for the estimation period. We shall not go any further into a discussion on the corresponding estimation procedures. Let us simply note that model (2.3) enables us to determine the best factor forecasts, assuming all implicit prices are observable. These forecasts appear as linear combinations of price logarithms, with coefficients whose sum can be standardised to unity:

[18] If the number of transactions proves insufficient to compute such means, the period over which the mean is computed can always be lengthened by considering two or three consecutive dates.

$$f_{l,t}^{*} \cong \sum_{i=1}^{N} \pi_l(z_{it}) \log p_{i,t}^{*}$$

$$\Leftrightarrow \exp f_{l,t}^{*} = \prod_{i=1}^{N} \left[p_{i,t}^{*}\right]^{\prod_{l(z_{i,t})}}$$

Each factor is thus implicitly linked to a time-varying basket (stock) of composition $(\pi_l(z_{i,t}))$, with *i* varying), such that the change in f_l is close to that of the basket (stock) value. This is referred to as the *factor-mimicking basket (stock)* (Huberman, Kandel, Stambaugh, 1987). In other words, we can make a suitable choice of disaggregated indices whose changes will reproduce those of the underlying factors.

Chapter 3: Calculation procedures applied

The primary index is defined as the ratio of the current value of a reference housing stock to its base-period value. The method described in detail in this chapter is the same for all indices, with two reservations:

- the variables representing the quality of a property (characteristics) are not strictly the same for apartments and houses, or for the Paris Region and the Provinces;
- the procedure for stratifying the geographic area varies according to whether the observation concerns the Paris Region or the Provinces.

As we move from version 2 to version 3, the method used to calculate the indices has been changed. So that changes in the market can be better taken into account, the estimation stocks and reference stocks (cf. definitions below) will be updated more frequently (every two years). As a result, from now on the indices will be chain-linked. Three equations are used for chain-linking: the first concerns the primary indices and the other two the aggregated indices.

3.1 Reference stocks and estimation stocks

Definition 1: the housing stock and the reference period

The reference stock consists of the "basket of properties" or "portfolio" in which we will measure price variations. In this way we can be sure that the index effectively tracks price changes for the same dwellings and that it is not sensitive to variations in market structure.

We must therefore value the dwellings in the reference stock at the current date, despite the fact that they have not been sold, i.e. that we have been unable to observe their prices. To do this, we use an econometric model that links the price of a dwelling[19] to its physical characteristics and its location.

The *reference stock* consists of all the transactions in the period – called the *reference period* -, that are included within the scope of the index, except those situated at the extremes of the value distribution.[20]

In version 3 of the indices, the reference period consists of years *n-3* and *n-2* for the indices for years *n* and *n+1* (where *n* is an even number), since the stock has been updated every two years from 2008 (see below 3.5.1).

Definition 2: the housing stock and the estimation period

The estimation stock includes the transactions that are to be used to estimate the models. It consists of all transactions from the estimation period that fall within the scope of the index, except those for which the estimated value in the model diverges from the real value by more than two standard deviations. The reference stock and the estimation are divided into primary calculation areas or strata. There are 22 such areas in the Paris region and 271 in the Provinces (Table 3.2). [21]

In the Provinces, some strata consist of only one city, when there is a large enough real estate market to warrant this (in practice, over 110 transactions per quarter). In most strata, the models can bring out homogeneous price areas, using indicators. The other strata in the Provinces are defined by the set of towns obtained by combining the "region", "*département*", "urban unit" criteria, or other distinguishing geographic features (rural area, coastal resorts or Alpine region), or according to the average income per inhabitant (2006) for other municipalities.

The Paris region strata are defined according to geographic and price criteria.[22]

[19] Or more precisely the logarithm of the price.
[20] This is the price per square metre in the case of apartments and the total price for houses. We take off 1/20th at each end of the distribution.
[21] For the composition of the strata, see Appendix 2.
[22] For details of the calculation areas, see Appendix 2.

3.2 Model used for estimation

An econometric model is estimated for each area (stratum), hence for a market assumed to be homogeneous, based on estimation-stock transactions.

The model is written:[23]

$$\log p_i = \log p_0 + \sum_{a=1}^{2} \mu_a Y_{a,i} + \sum_{m=1}^{12} \theta_m M_{m,i} + \sum_{k=1}^{K} \beta_k X_{k,i} + \varepsilon_i \qquad (3.1)$$

with the following notation:

p_i : price per m^2 (for apartments) or total price (for houses) of property *i*

p_0 : price of the reference dwelling of the estimation stock

$Y_{a,i}$: dummy for year of sale of property *i*

$M_{m,i}$: dummy for month in which sale of property *I* occurred

$X_{k,i}$: physical characteristics of property *i*, indexed from *k=1* to *K*

$X_{k,i}$ are variables calculated from modalities of the initial available characteristics, Z_i , i.e. construction date, number of rooms, number of bathrooms, number of garages or parking spaces, condition of property; and in addition, for apartments only, which floor, whether there is a balcony, a cellar, a lift, the average floor space per room; and for houses only, whether there is a basement, the number of floors, size of plot and habitable space, etc.[24]

When necessary, the location of the dwelling within the stratum (neighbourhood), is regarded as a characteristic of the property, on a par with the number of rooms or floor space.

The $X_{k,i}$ variables may include cross-effects. They may be dummy variables, or continuous variables in the case of surface areas. Each of these characteristics has a price, assessed at the base period, relative to the characteristics of a reference property.

Definition 3: reference dwelling of estimation stock

The *reference dwelling* is the dwelling that has the characteristics used as references in the regressions (Table 3.1).

The non-observable or unobserved characteristics of dwellings are assumed to be uncorrelated with those observed quality characteristics. We assume separate markets for houses and apartments: for example, we assume that the value of an extra room may not be the same in an apartment and in a house. But in an apartment it is worth the same, irrespective of the floor it is located on. In the strata, the relative prices of dwelling characteristics are fixed. In other words, if a third-floor apartment is worth x% more than one on the ground floor, this ratio is the same for all apartments in the calculation stratum. On the other hand, the ratio may differ from one stratum to another.

[23] To simplify notations, we omit the stratum index *s*.

[24] These variables are not all present in both bases nor are they always used in the models. For details of the models used, see Chapters 3 and 4.

Table 3.1 - Characteristics of reference dwelling (estimation stock)[25]

Index	Reference dwelling in Paris Region	Reference dwelling in Provinces
Apartments	- 3 rooms, - ground floor, - average floor space per room for studios between 20 and 30 sq.m., for two rooms between 17 and 24 sq.m., for three rooms between 18 and 22 sq.m. and for four rooms and more between 17 and 21 sq.m., - no garage, - no cellar, - 1 bathroom, - Built between 1948 and 1969, - sold in the 12th month of the second year of the reference period,	
		- no terrace or balcony**, - in good condition**.
Houses	- 4 rooms, - 2 floors, - construction period not known, - 1 garage, - 1 bathroom, - 100 sq.m. living space and 610 sq.m. plot size, - sold in the 12th month of the second year of the reference period,	
	- 1 building*, - no cellar*.	- no basement**, - condition of property unknown**,

* Variable not present in database for Provinces
** Variable not present in database for Paris Region

Usually, the model introduces a dummy neighbourhood variable into the stratum. When this is the case, the reference dwelling then belongs to the neighbourhood omitted from the regression. The neighbourhood is introduced when the stratum concerns a agglomeration or a city, and when the volumes of transactions are sufficient (at least 110 transactions per year). Neighbourhoods do not necessarily correspond to administrative divisions.

Table 3.2 shows the number of strata in version 3: 22 in the Paris region, compared with 62 in version 2, 271 in the Provinces compared with 234 in version 2. The number of neighbourhoods considered has been reduced in the Paris region (from 230 to 102) and increased in the Provinces (from 1125 to 1509).

Table 3.2 – Number of primary strata in calculation and number of neighbourhoods

Index	Number of strata	Number of neighbourhoods
Paris Region	*22*	*102*
Apartments	15	66
Houses	7	36
Provinces	*271*	*1,509*
Apartments	97	516
Houses	174	993
Total	**293**	**1,611**

3.3 Current price of reference property

The same type of model can be used for the current period. The reference property has the characteristics described in the previous paragraph, but corresponds to the current sale period *t*. Its price is noted $p_{0,t}$. The change in the reference property price forms the basis for constructing the index. For this reason, it has to incorporate seasonal effects, trends and cycles. This is why seasonal and annual coefficients no longer appear in the current model.

The price per sq.m. of a property *j* sold during period *t* can be expressed using the model:

$$\log(p_{j,t}) = \log(p_{0,t}) + \sum_{k=1}^{K} \beta_{k,t} X_{k,j,t} + \varepsilon_{j,t}$$

[25] Variables are described in Chapter 4.

Note that the model (3.1) used in the estimation phase is compatible with the model above, used for current values. Year and month dummies need to be introduced because the estimation stock covers a period spanning several months. Similarly, the reference price for the estimation stock and the month *(a,m)* would be:

$$\log\left(p_{0,a,m}\right) = \log p_0 + \mu_a + \theta_m$$

We must now explain how to reconstruct the price of the reference property from current transaction data.

Let us first hypothesise that the $\beta_{k,t}$ coefficients are known and let us introduce $\tilde{p}_{j,t}$ the "reference-property equivalent price" of the property j,t :

$$\log\left(\tilde{p}_{j,t}\right) = \log\left(p_{j,t}\right) - \sum_{k=1}^{K} \beta_{k,t} X_{k,j,t}$$

The model can be rewritten:

$$\log\left(\tilde{p}_{j,t}\right) = \log\left(p_{0,t}\right) + \varepsilon_{j,t}$$

Thus, if we know the $\beta_{k,t}$ coefficients, the log of the reference-property price $\log\left(\mathrm{p}_{0,\mathrm{t}}\right)$ will be estimated using ordinary least squares, i.e. from the mean:

$$\log\left(\hat{p}_{0,t}\right) = \frac{1}{J_t} \sum_{j=1}^{J_t} \log\left(\tilde{p}_{j,t}\right)$$

where J_t is the number of transactions in period *t*.

If we assume that the model estimated for the reference period is stable over time (for the period when the index is calculated, which is now two years), we can replace the $\beta_{k,t}$ coefficients by the $\hat{\beta}_k$ values estimated for the reference period when calculating the reference-property equivalent prices:

$$\log\left(\tilde{p}_{j,t}\right) \cong \log\left(p_{j,t}\right) - \sum_{k=1}^{K} \hat{\beta}_k X_{k,j,t} = \log\left[\frac{p_{j,t}}{\exp\left(\sum_{k=1}^{K} \hat{\beta}_k X_{k,j,t}\right)}\right]$$

We then estimate $p_{0,t}$,the price per m^2 of the reference property in period *t*, using a geometric mean of the reference-property equivalent prices of the J_t properties sold in the course of period *t*:

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

i.e.:

$$\hat{p}_{0,t} = \left(\prod_{j=1}^{J_t} \tilde{p}_{j,t}\right)^{\frac{1}{J_t}}$$

In the following, we shall write $\hat{\alpha}_{0,t} = \log \hat{p}_{0,t}$.

3.4 Current value of reference stock

Having determined an approximate reference-property value (from estimating the reference-property equivalent price in period *t*), we can then reconstruct the approximate values of the properties in the reference stock and, by aggregation, the value of the stock itself. Calculations are performed stratum by stratum. For this reason we reintroduce the stratum index *s*.

We estimate the value of a property *i* in the reference stock of stratum *s* in the current period *t* from its characteristics $X_{k,i,s}$, which, it will be recalled, are independent of date *t*, by virtue of the very definition of the reference stock, of which the composition is stable over time.[26]

In the case of apartments, for example, the approximate value is:

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

where $A_{i,s}$ denotes the surface area of property *i,s*.

By summation of the estimated current value of the N_s properties of the reference stock of stratum *s,* we obtain the value of the reference stock on date *t*:

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

In the same way, we estimate the value of the reference stock in stratum *s*, in period *0*, called the index *base period*:

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

3.5 Calculation method

Three major changes have been introduced in relation to the previous version of the indices.

- The reference and estimation stocks are now updated every two years; the price vectors relating to the characteristics of the reference stocks, and hence their value at the base period, are also re-estimated every two years. The Scientific Board for the Notaries-INSEE indices[27] adopted this measure after observing that, over the planned new reference period (2002-2006), the hedonic models were not stable.[28] Since the coefficients and weightings are to be regularly updated, a chain-linked method was introduced. Specification of the models and definition of the strata should continue to be reviewed approximately every six years.

- The introduction of dummies into the equation (3.1) to represent months means that it is now possible to calculate monthly indices. These are calculated by PNS and Min.not but only PNS publishes them.

- In version 2, we used arithmetic means to aggregate the primary indices. In version 3, we use geometric means for infra-departmental and departmental levels. This method of calculation is used when constructing statistical indices if it is thought that there may be some switching between different goods of a similar nature. In our case, there may be the choice of buying properties in neighbouring geographic strata. We use arithmetic means for levels that are more aggregated than the departmental levels since there is less likelihood that a purchaser will switch between more distant geographic strata (region, province, whole of France) when moving house.

[26] Thus the quality of the dwellings does not vary.
[27] The composition of the Scientific Board for the Notaries-INSEE indices, responsible for supervising the indices, is described in agreements reached with the Notariat. Cf. Appendices 8 and 9.
[28] See Appendix 4, Stability test and duration of the reference period.

3.5.1 Renewal of reference stocks and estimation stocks every two years

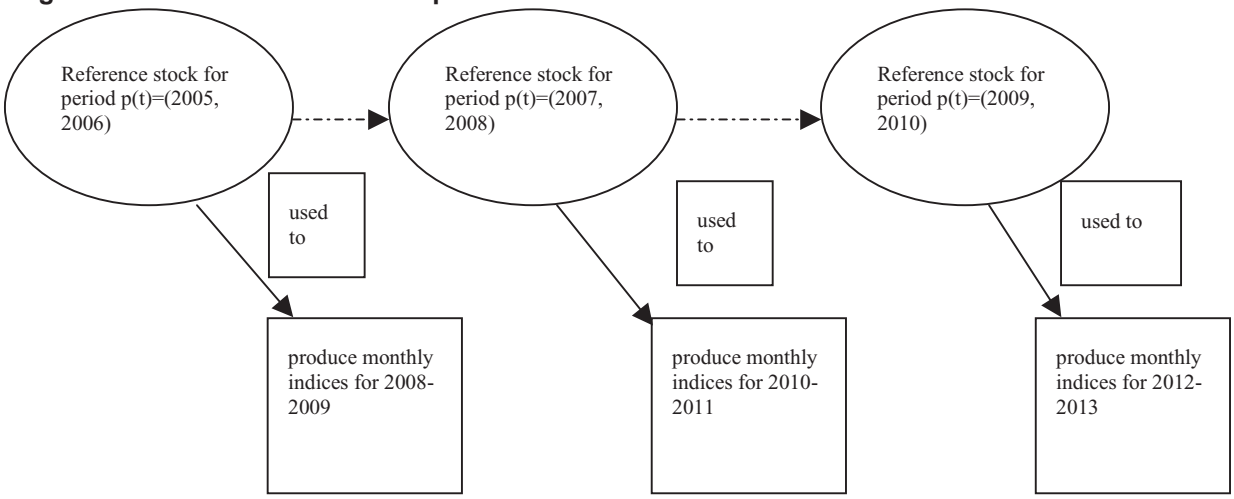
Reference stocks

In version 1 of the indices, the reference stock included transactions for between three and five years, depending on the indices (1994-96 for apartments in the Provinces, 1994-97 for houses and 1992-96 for apartments in Paris and the inner suburbs). In version 2, the reference stock included transactions from the four years 1998-2001.

In version 3, the reference stocks are updated every two years. The reference period *p(t)* is also set at two years of transactions, which is therefore shorter than before. It is defined in relation to the beginning of the current period, producing the indices in the following way: for current production of the index over the two years *(n, n+1)* where *n* is an even year, *p(t)* begins two years before, i.e. *p(t) = (n-3, n-2).*[29]

Figure 3.1 shows the succession of different reference stocks used to calculate monthly indices published from 2008-2013.

Figure 3.1 – reference stocks for period 2008-2013



Base period and publication base

The reference stocks are valued in the last quarter *q(t)* of the reference period, i.e. in the fourth quarter of the even year *n-2* for indices published over the period *(n, n+1).*[30] The weightings used to aggregate the primary indices are calculated on the basis of this housing stock valuation.

The quarter used to value the reference stock is called the *base period* for indices over the period *(n, n+1).* This should not be confused with the *base 100* of the published series, which was fixed in the first quarter of 2010.

Notations

The formula for calculating the value of the reference stock has to be generalised to take into account the renewal of the reference stock every two years by introducing a supplementary subscript, - written *p(t)*, the reference stock used -: the estimated value of the stock then becomes $\widehat{W}_{s,t}^{p(t)}$. In the same way, we introduce the value of the reference stock at the base period *q(t)*, which then chain-links the indices over the reference period *p(t)*. According to the different steps, it will be written $\hat{W}_{s,q(t),t}^{p(t)}$ or more simply, if possible in the context: $\hat{W}_{s,t}$ or $\hat{W}_{s,q(t),t}$ depending on the situation.

[29] For example, the first current period when indices were produced is years 2008 and 2009 thus *p(t)* = (2005, 2006). When the estimates are produced, *n-1* is not available.
[30] Thus, for example, the fourth quarter of 2006 will be the quarter when the reference stock for 2005-2006 is valued.

In the current period, the value of the reference stock becomes:

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

with the same notation as in paragraph 3.4:

$\hat{W}_{s,t}$: current value of the reference stock in stratum s,

$\hat{\alpha}_{0,s,t}$: estimation of the log of the price/sq.m. of the reference property in stratum *s* on date *t*; the coefficients are estimated over the period $p(t)$,

N_s : number of transactions in stratum *s* during the reference period,

$\sum_{k=1}^{K} \hat{\beta}_{k,s} X_{k,i,s}$: correction to be applied to the estimation of the log of the reference property price in stratum *s* on date *t,* to estimate for this same date the log of the price of property *i* in stratum *s* purchased during the reference period,

$A_{i,s}$: floor space of property *i* in stratum *s* exchanged during the reference period.

3.5.2 Calculating primary indices in the current period

Calculating price changes

"Reference-property equivalent" prices are calculated using the price vector of the property characteristics, as described in paragraph 3.3. Using these estimated prices, we value the reference stock in quarter *t*.

By introducing dummy variables into the models for the month (3.1) instead of the quarter, we can now publish monthly indices. However, we do not have enough observations to calculate "pure" monthly indices. The transactions contributing to this calculation therefore still relate to a quarter. The monthly index is calculated using data from the last available quarter including this month: this is a monthly index, quarter-on-quarter.

Chain-linking

Price changes from quarter *t-1* to quarter *t* are obtained by comparing the value of the reference stock for quarter *t* with the value of the reference stock for quarter *t-1* observed three months earlier; *t* includes the months *m-2*, *m-1* and *m* and *t-1* includes months *m-5, m-4* and *m-3*. The formula for calculating price changes between quarters *t-1* and *t* is therefore:

$$I_{t/0}(s) = \frac{\hat{W}_{s,t}^{p(t)}}{\hat{W}_{s,t-1}^{p(t)}} \times I_{t-1/0}(s)$$

where *0* is the reference quarter and *t* is the corresponding previous quarter.

The price changes between quarters 0 and *t* are obtained by chain linking changes in the reference stock values between 0 and *t,* using the formula:

$$I_{t/0}(s) = \left[\prod_{u=1}^{t} \frac{\hat{W}_{s,u}^{p(u)}}{\hat{W}_{s,u-1}^{p(u)}} \right]$$

It has become necessary to introduce this chain-linking because the reference stock is now updated every two years, something which is new since the last version of the indices. In version 3, indices before 2008 have been calculated by applying index changes from version 2.

Putting the chain-linking in place

As the monthly indices are calculated on a quarter-on-quarter basis, a base is needed for the start of the chain-linking process for the three series of primary indices corresponding to the first, second and third months of the calendar quarters. The series of indices starts in January 2008. The base for the start of the chain-linking is set at October, November and December 2007.

Thus we calculate three start values in order to calculate the first three indices (for January, February and March 2008). We estimate the value of the reference stock *p(t)=(2007,2008)* in October 2007 (with the corresponding previous quarter consisting of sales from August, September and October 2007), in November and December 2007. The October 2007 index is then set at 100. The November index is obtained by comparing the November value for the reference stock *p(t)=(2007,2008)* with its October value. The December index is obtained by comparing the December value of the reference stock *p(t)=(2007,2008*) with the October value (*Table 3.3*).

For example, indices for January 2008, December 2009 and January 2010 for stratum *s* will be calculated as follows:

$$I_{janvier2008/0}(s) = \frac{\hat{W}_{s,janvier2008}^{parc2005-2006}}{\hat{W}_{s,octobre2007}^{parc2005-2006}} * I_{octobre2007/0}(s)$$

$$I_{décembre2009/0}(s) = \frac{\hat{W}_{s,décembre2009}^{parc2005-2006}}{\hat{W}_{s,septembre2009}^{parc2005-2006}} * I_{septembre2009/0}(s)$$

$$I_{janvier2010/0}(s) = \frac{\hat{W}_{s,janvier2010}^{parc2007-2008}}{\hat{W}_{s,octobre2009}^{parc2007-2008}} * I_{octobre2009/0}(s)$$

Table 3.3 – Calculation of price indices for last three months of 2007

Month of 4th quarter 2007	Estimated value of reference stock 2007 - 2008 * In month of publication	Formula	Value of index
1st month: October	500	100.0	100.0
2nd month: November	510	510/500 X 100	102.0
3rd month: December	513	513/500 X 100	102.6

* in € million

We can then start to chain-link the primary indices for the months of publication in the first quarter of 2008 using the following formula:

$$I_{t/0}(s) = \frac{\hat{W}_{s,t}}{\hat{W}_{s,t-1}} \times I_{t-1/0}(s)$$

where *t* is the corresponding quarter and *p(t)*=(2005,2006) as defined in 3.5.1.

The resulting primary indices can now be aggregated to obtain the indices for higher levels, such as the *département*, city, region, etc. For publication, the base 100 for the indices was set at the first quarter of 2010.

3.5.3 Calculating aggregate indices in the current period

To correct the non-exhaustiveness of the notarial database, we introduce an adjustment coefficient δ. This coefficient is estimated for each *département* and for each year of change of ownership, from tax data (amounts

of transfer taxes compiled by the Directorate-General for Taxation - DGFiP). This is obtained by dividing the sum of transactions in *département* d for year a, estimated from tax data, by this same amount recorded in the notarial database.

δ does not differentiate between apartments and houses. It is applied to all properties defined by the cross-tabulation (*département* x year of change of ownership).[31] It therefore modifies the term that weights the primary stratum indices, i.e. their value share of the reference stock.

First level of aggregation: calculating infra-departmental and departmental indices

For infra-departmental and departmental indices, the formula used is a geometric mean:

$$\frac{I_{t/0}}{I_{t-1/0}} = \prod_{s=1}^{n} \left(\frac{I_{t/0}(s)}{I_{t-1/0}(s)} \right) \left(\frac{\delta_s \hat{W}_{s,q(t)}}{\sum_{u=1}^{n} \delta_u \hat{W}_{u,q(t)}} \right)$$

where:

t, quarter

s, u, stratum

$I_{t/0}$, index of aggregate in t relative to reference quarter 0,

$I_{t/0}(s)$, index of stratum s in t relative to reference quarter 0,

$q(t)$, last quarter of the even year n-2 (base period),

$\hat{W}_{s,q(t)}$, value of properties in stratum s which changed ownership during period $p(t)$, estimated in quarter $q(t)$,

δ_s , correction coefficient for non-exhaustiveness of the notarial database.

Second level of aggregation: calculating supra-departmental indices

For supra-departmental indices with aggregate A, we use the arithmetic mean:

$$\frac{I_{t/0}(A)}{I_{t_0/0}(A)} = \sum_{d \in A} \left(\frac{\delta_d \hat{W}_{d,q(t)}}{\sum_{e \in A} \hat{W}_{e,q(t)}} \right) \left(\frac{I_{t/0}(d)}{I_{t_0/0}(d)} \right)$$

with the following additional notations:

t_0, last quarter of year n-1,

d, e, *département* or part of *département*

$I_{t/0}(A)$, index of aggregate A for quarter t relative to reference quarter 0,

$I_{t/0}(d)$, index of *département* d for quarter t relative to reference quarter 0,

$\hat{W}_{d,q(t)}$, value of properties in the *département* which changed ownership during period $p(t)$ estimated in quarter $q(t)$.

[31] This coefficient is the inverse of the coverage ratio in the notarial databases (described in Chapter 4, paragraph 4.3). However, when the coverage ratio is under 20%, the coefficient used is capped and it is therefore never greater than 5. As in version 2 of the indices, we use a value ratio rather than the volume ratio (number of transactions) that was used in version 1.

3.6 Numerical example for November 2012

There follows an example to describe the various steps in calculating a primary price index for November 2012. This example concerns apartments in stratum 2 of the Paris Region. The coefficients used for the period 2012-2013 are estimated from transactions in 2009-2010.

Step 1: extracting the base

The November 2012 index is calculated from all changes of ownership in September, October and November 2012. For the stratum and the quarter being considered, 222 changes of ownership were recorded in the base, falling within the scope of the index and providing the information necessary to calculate the indices. Prices per square metre range from €1,731 to €5,225.

Step 2: calculating the reference-property equivalent price

The first of the 222 transactions concerns a three-room apartment of 64 sq.m., with a bathroom, two garages and a cellar, located on the ground floor of a building of which the date of construction is not known; the dwelling is in neighbourhood no. 2. The selling price was €190,000, or €2,969/sq.m.

The coefficients of the characteristics that are specific to the property have to be removed to bring it down to the "reference-property equivalent" price.[32] We therefore have:

$\text{Log}(\widetilde{p}_{j,\tau})$ = Log (2,969), or 7.9959

+0.0598 (neighbourhood 2 → reference neighbourhood)

-0.0227 (1 or more cellars → no cellar)

-0.0670 (2 or more garages → no garage)

-0.1573 (date of construction unknown → 1948-1969)

= 7.8088

For the "reference-property equivalent" price $\widetilde{p}_{j,\tau}$ of this first transaction, we obtain $e^{7,8088}$, or €2,462/sq.m.

We repeat the operation for the other 221 transactions.

In each neighbourhood, transactions where the reference-property equivalent price is below the 2nd percentile or above the 98th percentile of the distribution are removed. In our example, the neighbourhood had only 36 transactions for the quarter. In this case we removed the most expensive and the least expensive properties. Across all of the five neighbourhoods in stratum 2, we withdrew 14 references where the reference-property equivalent prices were extreme values (6 in neighbourhood 1 and 2 in each of the other 4).

The price of the reference property, equal to the exponential of the arithmetic mean of the logs of the 208 reference-property equivalent prices of the remaining transactions, is Log (7.8314), or €2,518/sq.m.

Step 3: calculating the monthly stratum index for November 2012

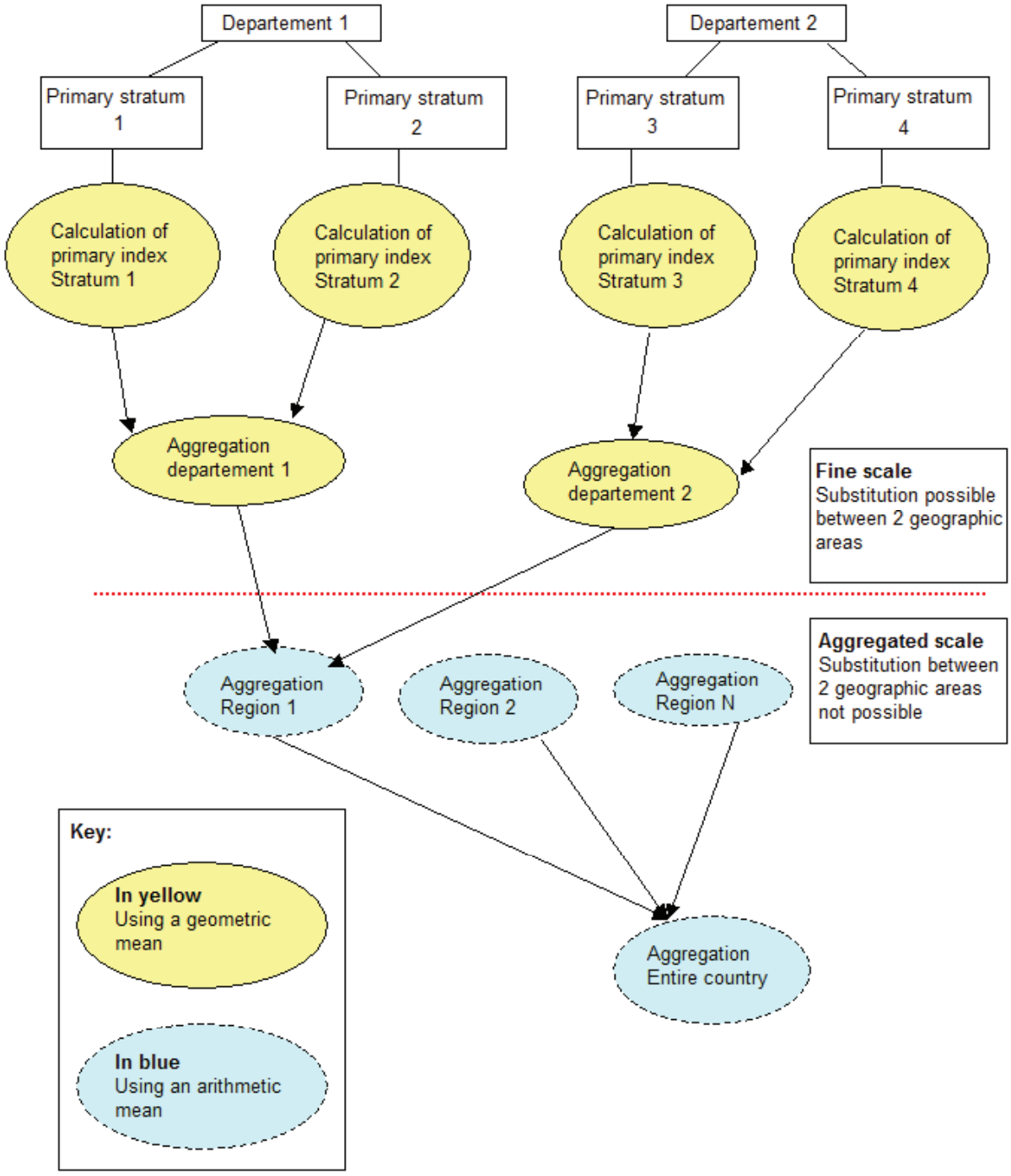
To determine the provisional monthly index[33] for a stratum in November 2012, we calculate the definitive price per sq.m. for the stratum in August 2012 (from transactions for June, July and August 2012), as the geometric mean of prices per sq.m. obtained in step 2 for each property in the stratum, or €2,498/sq.m. This price is compared with the provisional price three months later (€2,518/sq.m.) to determine the quarterly change in the stratum. It is this change in price per sq.m. in relation to the definitive price per sq.m. estimated for the stratum in November 2011 (reference period for price changes in 2012 and 2013) that will be aggregated to departmental and regional level to determine price changes by *département* and region.

[32] Cf. Table 3.1 on page 19.
[33] The definitive index is published three months later.

From primary indices to the national index: aggregating primary indices

Figure 3.2 below shows the different steps in the transition from primary index to the national index. *Départements* 1 and 2 belong to region 1. The geometric mean is applied for infra-departmental and departmental levels. The arithmetic mean is used for levels that are more aggregated.[34]

Figure 3.2 - Aggregation of primary indices



[34] Cf. page 21.

Chapter 4: Notarial databases

The Notaires-INSEE indices are calculated from transaction prices recorded in the notarial databases.[35]

4.1 Description of databases

There are two notarial databases:

- The "BIEN" database, managed by the PNS (Paris Notaires Services), covers the Paris Region. It was started in 1989 for Paris, 1991 for the inner suburbs (*Petite couronne*) and 1996 for the outer suburbs (*Grande couronne*); it has been used as a source for the indices since 1991.
- The "Perval" database, managed by ADSN (*Association pour le développement du service notarial*, a subsidiary of the Notariat), covers the Provinces and French overseas *départements*. It was set up in 1994.

The bases record transactions concerning all types of property, not only apartments and houses, but also buildings, business premises, land, garages, vineyards, and other agricultural properties. In August 2013, they contained 14.2 million records for transactions between 1990 and 2012 (Table 4.1). Every year, about 500,000 new transactions are added for used apartments and houses (25% in the Paris Region and 75% in the Provinces), which are suitable for inclusion in the calculation of the Notaires-INSEE indices. The use of the notarial databases and the calculation of the indices are governed by agreements between the CNIP (*Chambre interdépartementale des notaires de Paris*) or the CSN (*Conseil supérieur du notariat*) and INSEE.[36]

Articles 15 and 16 of law no.2011-331 of 28 March 2011 require a new public service obligation on the part of the legal profession. This law amends the legal provisions of the two laws that define the status and organisation of the profession of notary in France. Until 2011, the public service mission of the notariat was defined by article 1 of order no. 45-2590 of 2 November 1945 which defined their status: "Notaries are public officers, authorised to record any instrument or contract the parties which are obliged, or may wish, to invest with the authenticity associated with public authority instruments, and to guarantee their date, keep them safe, and issue principal and additional copies". From now on, the transmission of information about property transactions against payment to the CSN becomes a public service duty for notaries. In addition, this same law defines the activity of centralising and disseminating data collected by the CSN as a public service duty.

Until 2009, the transmission of data from notarial offices was paper-based. In 2010 the switch was made to electronic methods, which will ultimately enable the system to be much more flexible. In 2013, almost half of all records received by the databases arrived via electronic means. The introduction of electronic transmission has meant that waiting periods for sending in deeds have been reduced. There is at present one major drawback, however: data coding is now done by the notarial offices, rather than by the database administrators as was the case previously, and many more corrections are required now than with the paper deeds.

4.2 Delay in incorporating transactions

The time between the signing of a deed and the transaction being incorporated into the notarial databases obviously determines the responsiveness of the indices.

The target in 2013 in the Paris Region is a waiting period of 30 days to process data and add them to the database, after reception of the deeds. In 2012, the average total waiting period was 82 days.

In the Provinces, in 2010, the waiting time for processing was longer than in 2009 (127 days compared with 112), as the number of transactions had increased more rapidly than the facilities available for processing them. However, the backlog in processing the deeds was cleared by the end of 2011 (89 days in the second half-year of 2011). In 2012, the average was 76 days (54 to receive the deeds, 22 to input the data and incorporate them into the database.

Notaries have set themselves minimum stocks of non-coded deeds in hand, to ensure that the data coders have a steady stream of work. The minimum stock is around 30,000 deeds in the Provinces and should drop to 20,000

[35] For details on the transmission of data from deeds, cf. Appendix 1.
[36] The agreements are reproduced in Appendices 7 and 8.

by the end of 2013; in the Paris Region, the volume of non-coded deeds should be between 3,000 and 5,000 paper deeds and between 2,000 and 5,000 electronic deeds when the operation gets up to speed.

Table 4.1 - Records in notarial databases, by year of transaction

Metropolitan France					
Year of transaction	Number of records	Incl. apartments and houses	Incl. used apartments and houses	Used apartments	Used houses
1990	105,186	82,475	54,227	41,027	13,200
1991	154,485	114,236	90,180	52,737	37,443
1992	279,572	195,304	165,453	85,085	80,368
1993	327,631	227,823	195,929	95,225	100,704
1994	394,505	281,141	236,020	114,531	121,489
1995	402,567	280,963	242,414	109,178	133,236
1996	557,948	412,273	365,524	159,694	205,830
1997	555,070	397,629	345,942	150,894	195,048
1998	645,241	467,347	409,326	183,282	226,044
1999	787,764	571,815	493,771	227,086	266,685
2000	778,810	560,500	510,719	236,243	274,476
2001	780,481	571,173	515,391	242,854	272,537
2002	754,277	556,753	496,082	238,902	257,180
2003	802,454	588,543	515,064	250,533	264,531
2004	823,556	598,580	517,971	254,352	263,619
2005	851,706	623,377	533,046	264,876	268,170
2006	853,088	621,774	528,903	265,762	263,141
2007	855,971	621,867	528,480	264,472	264,008
2008	740,125	521,952	441,499	218,401	223,098
2009	627,281	454,096	370,111	180,816	189,295
2010	778,846	591,965	489,837	237,719	252,118
2011	719,914	534,713	463,821	224,390	239,431
2012	626,746	458,612	406,714	190,773	215,941
Total	14,203,224	10,334,911	8,916,424	4,288,832	4,627,592

Source: BIEN and PERVAL databases - August 2013

Table 4.2 – Records in PERVAL database, by year of transaction

Provinces					
Year of transaction	Number of records	Incl. apartments and houses	Incl. used apartments and houses	Used apartments	Used houses
1990	41,476	27,500	23,619	11,341	12,278
1991	97,448	67,126	56,323	24,275	32,048
1992	220,889	147,032	123,075	51,224	71,851
1993	265,316	176,123	150,620	58,891	91,729
1994	314,822	214,935	179,791	69,771	110,020
1995	330,619	222,636	192,180	70,194	121,986
1996	436,646	310,901	276,769	97,984	178,785
1997	433,432	296,838	259,422	91,109	168,313
1998	503,638	349,206	307,319	112,883	194,436
1999	606,950	421,348	364,425	138,231	226,194
2000	605,940	416,751	381,083	146,593	234,490
2001	607,127	426,318	385,757	151,799	233,958
2002	576,602	406,180	361,322	143,985	217,337
2003	615,992	429,736	375,242	151,370	223,872
2004	622,696	428,530	368,987	147,973	221,014
2005	643,823	447,509	380,739	155,639	225,100
2006	653,110	453,308	381,721	160,336	221,385
2007	661,782	458,936	385,306	161,772	223,534
2008	581,206	389,982	326,399	135,229	191,170
2009	491,910	340,026	273,998	112,013	161,985
2010	596,982	434,569	359,958	146,287	213,671
2011	564,361	402,584	351,476	145,198	206,278
2012	493,752	346,968	310,787	122,682	188,105
Total	10,966,519	7,615,042	6,576,318	2,606,779	3,969,539

Source: PERVAL database - August 2013

Table 4.3 – Records in BIEN database, by year of transaction

Paris Region					
Year of transaction	Number of records	Incl. apartments and houses	Incl. used apartments and houses	Used apartments	Used houses
1990	63,710	54,975	30,608	29,686	922
1991	57,037	47,110	33,857	28,462	5,395
1992	58,683	48,272	42,378	33,861	8,517
1993	62,315	51,700	45,309	36,334	8,975
1994	79,683	66,206	56,229	44,760	11,469
1995	71,948	58,327	50,234	38,984	11,250
1996	121,302	101,372	88,755	61,710	27,045
1997	121,638	100,791	86,520	59,785	26,735
1998	141,603	118,141	102,007	70,399	31,608
1999	180,814	150,467	129,346	88,855	40,491
2000	172,870	143,749	129,636	89,650	39,986
2001	173,354	144,855	129,634	91,055	38,579
2002	177,675	150,573	134,760	94,917	39,843
2003	186,462	158,807	139,822	99,163	40,659
2004	200,860	170,050	148,984	106,379	42,605
2005	207,883	175,868	152,307	109,237	43,070
2006	199,978	168,466	147,182	105,426	41,756
2007	194,189	162,931	143,174	102,700	40,474
2008	158,919	131,970	115,100	83,172	31,928
2009	135,371	114,070	96,113	68,803	27,310
2010	181,864	157,396	129,879	91,432	38,447
2011	155,553	132,129	112,345	79,192	33,153
2012	132,994	111,644	95,927	68,091	27,836
Total	3,236,705	2,719,869	2,340,106	1,682,053	658,053

Source: BIEN database - August 2013

4.3 Coverage rate

The notarial databases have until now been supplied with data on a voluntary basis, and they are not exhaustive; the coverage rate, defined here as the ratio of known amounts in the databases to true transaction amounts, is therefore not 100%.

Knowing the coverage rate is important, for two main reasons. First, if a certain type of transaction were under-recorded and changes in these prices showed a specific profile, the index would be biased.[37] This would happen if the notaries' behaviour in forwarding information to the databases varied with the characteristics of property-price changes. Hence the importance of knowing the coverage rate by property type and by geographic area. Second, it is interesting *per se* to be aware of the number of real-estate transactions in order to monitor their trends. This cannot be done using only the transactions recorded in the notarial databases unless the coverage rate is constant. At present it does fluctuate and is likely to increase further mainly because notaries are now legally obliged to supply information to the databases.

In the absence of exhaustive national data on sales of used housing by dwelling type, we must take an indirect approach. There are three possibilities: using an *ad hoc* survey, using a breakdown of notarial deeds or using tax data.

- *Ad hoc* survey

Existan, a survey of local tax offices carried out every year in the 1990s by the Ministry for Infrastructure, has been an invaluable tool. Sampling was based on paper documents held by the tax offices. The survey continued until 2000, and covered all of France; since then, it has been abandoned when the tax departments have moved on to computerisation (and the paper documents have disappeared), for cost reasons and pending the extraction of data from the asset database compiled by the tax offices. This survey categorised different types of property

[37] We can minimise this potential bias by choosing a reference stock whose structure is not too different from the stock that can be determined from censuses. In fact, two issues are interrelated. First, the reference stock must give an unbiased picture of all transactions; the issue here is the representativeness of the notarial databases, of which the coverage rate is only one factor. Second, the reference stock must also be an acceptable proxy for the housing stocks, as the purpose of the index is to measure the change in (theoretical) value of the housing stocks (see Chapter 1) and not only the value of flows of traded goods. This second issue justifies the comparison (by structure) of the reference stock with the census. See tables in paragraph 4.4.5.

(apartments, houses, entire buildings) as well as the characteristics of the buyers and sellers. It allowed comparisons, by *département*, of the number of transactions in notarial databases with those recorded by the survey, and hence gave an estimate of coverage rates.

- Breakdown of notarial deeds

For the Paris Region, the BIEN database is linked to a survey on notarial office activity which gives a monthly breakdown, by *département*, of the number of notarial deeds, and hence of transactions. We estimate the coverage rate of the database from the ratio of the number of deeds recorded in the database to the total number of deeds of sale signed. Thus, in 2010, the overall coverage rate was 82% in the Paris Region (86% for Paris, 82% for the inner suburbs and 79% for the outer suburbs). However, the breakdown of the deeds does not differentiate between transactions concerning new properties and those concerning used properties, nor does it distinguish dwellings in different types of property; in addition, the location taken into account is that of the notarial office and not that of the property.

For the Provinces, there are no figures for the number of deeds; we therefore have to use a different method to estimate the coverage rate.

- Tax data

Since Existan was discontinued, we have used tax sources to estimate coverage rates.[38] The basic information consists of the amount of transfer-tax revenue (*droits de mutation*) collected in each *département* and according to departmental taxation rates (currently 3.8% or 0.6%). The 0.6% tax assessment basis includes properties transferred free of charge (donations) and can therefore not be used. The 3.8% bracket, however, (calculated by dividing the revenue from taxation by 3.8%) can be compared with the same tax assessment base in the notarial databases. The result[39] is the coverage rate (in monetary value terms) of the notarial databases for all properties taxed at 3.8%: used dwellings, used business real estate and land not subject to VAT, with a few exceptions. By applying certain approximations, we can determine a coverage rate for used housing alone.[40]

Using the above procedure, the coverage rate of the notarial databases for existing dwellings in 2010 was estimated at 63% for all of France (Corsica and overseas *départements* included), 80% for the Paris Region and 56% for the Provinces. It varies greatly from one *département* to another: in 2010, excluding Corsica and overseas *départements* and territories, it was under 30% in 7 *départements*, between 30 and 50% in 34 *départements*, between 50% and 70% in 40 *départements* and more than 70% in 13 *départements* (*Figure 4.1*).

A low coverage rate does not in itself pose a statistical problem, provided that the number of records exceeds a given threshold; what is more problematic is the fact of not knowing whether the deeds recorded in the database are representative of all transactions or whether they are biased. This would happen, for instance, if some notarial offices never sent in deeds relating to particular markets or if some notaries did not send in the deeds concerning a certain category of client.[41] Partial comparisons made so far give no reason to think that there is any significant bias, except in the few *départements* with the lowest coverage rate.

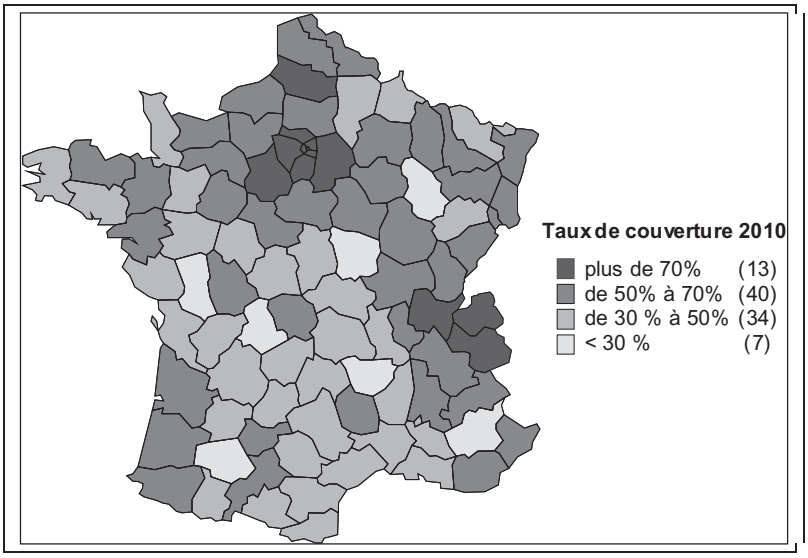
[38] Results from the two sources were compared for 1997. Transaction numbers and values were identical to within 1% for all of France, with some wider gaps observed in certain *départements*.

[39] Deeds are signed on average 1.7 months before the transfer taxes are recorded in the mortgage register. We compare the value of the deeds that actually appear in the notarial databases for month n with the value of taxes recorded by the Directorate General for Public Finances (DGFiP) for month n+2.

[40] Details of calculations and the results are given in Appendix 1.

[41] In the Provinces, according to the database administrators, returns for houses and apartments are usually of good quality. The same is not always true for other types of property and some offices never or seldom send in information on the deeds for rural properties (agricultural land and vineyards), however, these properties are not included when constructing the indices. In the Paris Region, it is possible, again according to the database administrators, that exceptional properties (e.g. very expensive) are transmitted less frequently than the more common transactions (because of the purchaser, or even the notary).

Figure 4.1 - Estimated coverage rate of notarial databases, for existing dwellings, by department, 2010



Lecture : Les taux de couverture sont calculés en montants de transactions.
Source : Bases Perval et BIEN

4.4 Scope of coverage of indices, data quality

4.4.1 Defining the scope

The scope of the Notaires-INSEE indices is confined to existing dwellings as defined for tax purposes (apartments or houses subject to registration duties and not to VAT). New dwellings are therefore not included.

We also exclude dwellings that are considered to be non-standard such as rooms, attics, lofts, workshops, *concierges'* lodgings, manors, large properties, town houses.

The dwelling must be unoccupied at the time of sale, intended for residential purposes only and acquired with full property rights by a private individual or a real-estate company.[42] We therefore remove dwellings when the period of occupancy by a third party or by the seller exceeds six months, as we consider that, given the regulations on rental leases, these dwellings usually sold below the going price.

Lastly, we consider only private sales, i.e. excluding auction sales.

4.4.2 Transactions used for the calculation

The transactions selected must also have the following characteristics:

- for apartments:
  - the number of rooms must be fewer than 9,
  - the habitable space is between 10 sq. m. and 200 sq. m.,
  - the sale price is between €1,500 and €5,000,000,
  - the price per sq. m. is less than €25,000.
- for houses:
  - the number of room is fewer than 13,
  - the habitable space is between 20 sq. m. and 300 sq. m.,
  - the plot size (including the floor space of the house) is more than 9 sq. m.,

[42] Numbers of dwellings sold by type of purchaser are given in Appendix 9, Tables A9.1 and A9.2.

- the sale price is more than €1,500,
- (only in the Provinces) the sale price is less than €15,000,000[43].

Table 4.4 – Number of observations excluded at each stage (2009-2010 reference stock)

	Paris Region		Provinces	
	Apartments	**Houses**	**Apartments**	**Houses**
Number of transactions entered	**116,544**	**40,857**	**205,791**	**224,696**
Dwellings occupied (partially or totally)	10,454 (6.6%)		18,822 (4.4%)	
Use other than residential	2,413 (1.5%)		10,416 (2.4%)	
Other than full property rights	1,321 (0.8%)		230 (0.1%)	
Type of transaction other than private sale	1,081 (0.7%)		174 (0.0%)	
Purchaser out of scope (=other than individuals, real estate companies, SCI and unknown)	2,733 (1.7%)		7,529 (1.7%)	
Property not fully second-hand	27,590 (17.5%)		72,498 (16.8%)	
Apartments excl. studios, duplex, triplex, standard apartments with no information	1,007 (1.2%)		913 (0.7%)	
Apartments where number of rooms and floor space not provided	none[44]		4,921 (3.8%)	
Apartments with 9 rooms and more	30 (0.0%)		73 (0.1%)	
Apartments of 1 to 9 sq.m.	23 (0.0%)		145 (0,1%)	
Apartments of at least 201 sq.m.	308 (0.4%)		290 (0.2%)	
Apartments priced at more than €25,000/sq.m.	9 (0.0%)		32 (0.0%)	
Apartments priced at €1,500 or less	none		299 (0.2%)	
Apartments priced at €5,000,000 or more	14 (0.0%)		3 (0.0%)	
Houses out of scope (=other than town houses, detached houses, villas, farms and unspecified)		412 (0.5%)		8 182 (4.0%)
House where number of rooms and floor space not provided		none		19,749 (9.8%)
Houses with 13 rooms and over		13 (0.0%)		346 (0.2%)
Houses of 1 to 19 sq.m.		7 (0.0%)		95 (0.0%)
Houses of at least 301 sq.m.		72 (0.2%)		665 (0.3%)
House with plot size 0 to 9 sq.m.		8 (0.0%)		1,885 (0.9%)
Houses priced at €1,500 or under		none		516 (0.3%)
Houses priced at €15,000,000 or over				2 (0.0%)
Number of valid transactions for purposes of calculating indices	**79,264 (68.0%)**	**32,057 (78.5%)**	**178,657 (86.8%)**	**180,568 (80.3%)**

[43] In the Paris Region the selection is made beforehand, when codification is carried out. All sale deeds for houses with a price of more than 3.5 times the median current price (obtained by multiplying the median price in Q4 2000, base period of the indices V2 Notaries - Insee, by the evolution of the index) are checked by a coder; all houses with a price of over 7 million euros are also checked by a coder.

[44] In the Paris Region, the number of rooms is estimated when this information is not provided (cf. Table 4.7)

4.4.3 Available variables

More and more deeds are codified in the notarial offices and transmitted electronically. The others are still transmitted in paper format : photocopies of extracts from deeds or standardised mortgage documents [*documents hypothécaires normalisés*: DHN), accompanied by further information not given in the documents sent. Deeds are then codified by the database administrators. However, variables are sometimes not properly filled in and are not used in the hedonic models. The explanatory variables used in the models are described in Table 4.5.

Table 4.5 - Variables used to define dwelling quality

Variable	Number of modalities	Definition
- number of rooms (apartments)	5	"1 room" to "5 and+"
- number of rooms (houses)	5	"less than 3 rooms" to "7 rooms and+"
- floor of building and presence of lift (apartments only)	6	Ground floor, 1st floor, 2nd floor, 3rd floor, 4th floor or higher without lift, 4th floor or higher with lift
- number of floors (houses only)	3	1, 2, 3 or more
- average living space per room by property size (apartments only)		For studios: 20 to 30 sq.m., two-room apartments: 17 to 24 sq.m., three-room apartments: 18 to 22 sq.m., four-room apartments and more: 17 to 21 m²
- living space (houses only),		in sq.m.
- plot size (houses only)		in sq.m.
- number of garages or parking spaces	3	0, 1, 2 and +
- number of bathrooms	4	0, 1, 2 and +, not determined
- construction period	10	before 1850, 1850-1913, 1914-1947, 1948-1969, 1970-1980, 1981-1991, 1992-2000, 2001-2010, 2011-2020[45], not determined
- date of transaction		
- presence of a balcony or terrace (apartments in the Provinces only)	2	yes/no
- number of buildings (houses in Paris Region only)	2	yes/no
- presence of a cellar (apartments)	3	yes/no, not determined
- presence of a cellar (houses in Paris Region only)	3	0, 1, 2 and +
- presence of a basement (houses in the Provinces only)	2	Yes, no
- condition of property (Provinces only)	3	good condition, needs work, needs renovation

Lastly, some variables, which are recent or perhaps where information is inaccurate, are not used for the moment but could be interesting to exploit in future:

- type of heating (whether there is central heating),
- energy rating (A to G according to the level of primary energy consumption),
- climate label (ratings A to G according to the level of greenhouse gas emissions),
- status of the buyer (first-time buyer or not),
- type of negotiation (with a notary, with an agency, private sale),
- type of renewable energy,
- materials used for construction (concrete, cut stone, metal materials, stone, brick, wood, earth, other),
- fact that the building has a building quality label (BBC, HQE, etc.),
- first sale of the property since its completion (yes/no),
- property occupied as a primary residence (yes/no), etc.

[45] This modality was created at the beginning of the 2010s and will be included in the hedonic models when there are sufficient observations.

4.4.4 Dealing with missing information

In order to be used in the calculations, a transaction must include at least the price of the property, the place where the transaction took place, the date, and also the type of property.

Information on at least one of the variables for surface area or number of rooms must be provided.

In other cases, variables may be imputed (by estimation or recoding, depending on the rules given in Table 4.7), or the missing value is processed as just that (Tables 4.6 and 4.7).

Table 4.6 – Share of observations where imputation applied

Property type and area	Living space	Number of rooms	Period of construction	Number of garages or parking spaces	Number of bathrooms	Floor or number of levels	Lift
Collective							
Paris	9.2	2.5	11.4	0.6	11.6	2.1	44.7
Inner suburbs	9.1	1.3	17.6	0.3	7.8	2.9	46.8
Outer suburbs	9.4	1.1	18.4	0.2	4.8	4.0	47.3
Provinces	6.9	3.3	44.6	26.8	4.9	5.5	62.4
Individual							
Paris, inner suburbs	59.4	2.0	69.8	5.4	6.6	0.7	N/A
Outer suburbs	51.7	1.1	52.3	4.0	3.7	0.6	N/A
Provinces	41.9	*	36.9	22.2	8.8	17	N/A

Scope: all valid transactions for purposes of calculating index, as in table 4.4, including observations not retained in estimation and reference stocks; 2007-2008
"*": no imputation
" N/A": not applicable

Table 4.7 - Non-response by variable

Type of non-response	Geographic area	Modalities required	Action	Value if recoding/comment
Price	Paris Region, Provinces	Non-zero numeric value	Rejected	
Nature and use of property	Paris Region, Provinces	Apartment or house	Rejected	
Place of transaction	Paris Region, Provinces		Rejected	
Plot size (houses)	Paris Region, Provinces	Non-zero numeric value	Rejected	
Living space and number of rooms	Paris Region, Provinces		Rejected	Rejected if both variables missing
Living space (apartments)	Paris Region	Non-zero numeric value	Imputed	Econometric estimation according to number of bathrooms, number of rooms, period of construction, number of floors (or floor of building) and property type
Living space (apartments)	Provinces	Non-zero numeric value	Imputed	Observation rejected for constructing estimation or reference stocks; estimation in current year according to number of bathrooms, number of rooms and period of construction.
Living space (houses)	Paris Region	Non-zero numeric value	Imputed	Econometric estimation according to number of bathrooms, number of rooms, period of construction, number of floors (or floor of building), property type and plot size
Living space (houses)	Provinces	Non-zero numeric value	Imputed	Econometric estimation according to number of bathrooms, number of rooms, period of construction and number of floors
Number of buildings (houses)	Paris Region	Non-zero numeric value	Recoded	1 building
Number of bathrooms	Paris Region, Provinces	Non-zero numeric value	Recoded	Number of bathrooms unknown
Number of rooms	Paris Region	Non-zero numeric value	Imputed	Econometric estimation according to surface area and stratum and property type

Type of non-response	Geographic area	Modalities required	Action	Value if recoding/comment
Number of rooms	Provinces	Non-zero numeric value	Rejected	
Period	Paris Region, Provinces	10 date ranges	Recoded	1848-1969 for apartments; Period unknown for houses
Presence of lift (apartments)	Paris Region, Provinces	Yes/no	Recoded	Yes
Building floor (apartments)	Paris Region, Provinces	Non-zero numeric value	Recoded	Ground floor
Number of floors (houses)	Paris Region, Provinces	Non-zero numeric value	Recoded	2 floors
Number of cellars (apartments)	Paris Region	Numeric value	Recoded	0 (no cellar)
Number of cellars (apartments)	Provinces	Numeric value	Recoded	Number of cellars unknown
Number of cellars (houses)	Paris Region	Numeric value	Recoded	0 (no cellar)
Presence of basement (houses)	Provinces	Yes/no	Recoded	No
Number of garages (apartments)	Paris Region, Province	Numeric value	Recoded	0 (No garage)
Number of garages (houses)	Paris Region, Provinces	Numeric value	Recoded	1 garage for Paris Region 0 (No garage) for Provinces
Presence of balcony (apartments)	Provinces	Yes/no	Recoded	No
Presence of terrace (apartments)	Provinces	Yes/no	Recoded	No
Condition of property	Provinces	3 modalities	Recoded	Condition of property unknown

4.4.3 Structure of estimation stocks

In the following tables, we see the structure of the estimation stocks, according to the main variables used in the regressions. For comparison, percentages from the 2008 population census are given whenever possible.

Table 4.8 - Structure of estimation stock 2007-2008 and comparison with population census; Paris Region (apartments)

	Estimation stock 2007-2008		Stock in 2008 census (%)	
	Number	%		
Total	**146,089**			
Number of rooms				
1	22,329	15%	17%	
2	43,849	30%	29%	
3	42,627	29%	28%	
4	26,601	18%	17%	
5 or more	10,683	7%	9%	
Floor space				
<40 sq. m.	42,421	29%	31%	
40 to <70 sq. m.	64,673	44%	40%	
70 to <100 sq. m.	30,553	21%	22%	
100 <150 sq. m.	7,329	5%	6%	
150 sq. m. or more	1,113	1%	1%	
Construction period				**Census periods**
Not given	22,039	15%		
Before 1914	23,091	16%		
1914-1947	19,341	13%	38%	Before 1949
1947-1969	33,055	23%	32%	1949-1974
1970-1980	25,830	18%	11%	1975-1981
1981-1991	10,123	7%	5%	1982-1989
After 1991	12,610	9%	14%	1990 -
Bathrooms				
Not given	11,005	8%		
0	2,432	2%		
1	123,703	85%		
2 or more	8,949	6%		
Garage, parking space				
Not given	463	0%		
0	83,600	57%	52%	no
1	54,726	37%	48%	yes
2 or more	7,300	5%		

Table 4.9 - Structure of estimation stock and comparison with population census; Paris Region (houses)

	Estimation stock 2007-2008		Stock in 2008 census (%)	
	Number	%		
Total	**55,792**			
Number of rooms				
1 - 3	9,299	17%	17%	
4	13,732	25%	26%	
5	15,691	28%	28%	
6 or more	17,070	31%	29%	
Total				
Floor space				
<40 sq. m.	215	0%	2%	
40 to <70 sq. m.	7,109	13%	13%	
70 to <100 sq. m.	20,842	37%	37%	
100 to <150 sq. m.	22,470	40%	35%	
150 sq. m. or more	5,156	9%	13%	
Construction period				**Census periods**
Not given	31,411	56%		
Before 1914	2,524	5%		
1914-1947	4,818	9%	30%	before 1949
1947-1969	4,114	7%	27%	1949-1974
1970-1980	5,423	10%	13%	1975-1981
1981-1991	4,364	8%	13%	1982-1989
After 1991	3,138	6%	17%	1990-
Bathrooms				
Not given	1,756	3%		
0	464	1%		
1	35,468	64%		
2 or more	18,104	32%		
Garage, parking space				
Not given	2,213	4%		
0	15,425	28%	19%	no
1	33,545	60%	81%	yes
2 or more	4,609	8%		

Table 4.10 - Structure of estimation stock 2007-2008 and comparison with population census; Provinces (apartments)

	Estimation stock 2007-2008		Stock in 2008 census (%)	
	Number	%		
Total	**222,026**			
Number of rooms				
1	37,103	17%	13%	
2	59,507	27%	26%	
3	63,289	29%	32%	
4	44,909	20%	20%	
5 or more	17,218	8%	9%	
Floor space				
<40 sq. m.	57,044	26%	23%	
40 to <70 sq. m.	92,236	42%	42%	
70 to <100 sq. m.	57,146	26%	28%	
100 to <150 sq. m.	13,926	6%	6%	
150 sq. m. or more	1,674	1%	1%	
Construction period				**Census periods**
Not given	84,066	38%		
Before 1914	10,008	5%		
1914-1947	13,600	6%	30%	before 1949
1948-1969	38,941	18%	32%	1949-1974
1970-1980	34,231	15%	11%	1975-1981
1981-1991	22,384	10%	7%	1982-1989
after 1991	18,796	8%	20%	1990-
Garage, parking space				
0 or not given	118,872	54%	51%	no
1	91,965	41%	49%	yes
2 or more	11,189	5%		
Bathrooms				
0 or not given	9,860	4 %		
1	202,873	91 %		
2 or more	9,293	4 %		
Floor				
0 or not given	50,308	23%		
1	52,787	24%		
2	45,397	20%		
3	32,600	15%		
4 or more	40,934	18%		
Lift (floor > 3)				
No	3,043	7%		
Yes or not given	37,891	93%		
Cellar				
0 or not given	99,225	45%		
1 or more	122,801	55%		
Terrace or balcony				
No or not given	147,416	66%		
Yes	74,610	34%		
Condition of property				
Not given	113,718	51%		
Good	90,128	41%		
Needs work	14,973	7%		
Needs renovation	3,207	1%		

Table 4.11 - Structure of estimation stock 2007-2008 and comparison with population census; Provinces (houses)

	Estimation stock 2007-2008		Stock in 2008 census (%)	
	Number	%		
Total	**328,847**			
Number of rooms				
1 to 3	65,115	20%	15%	
4	83,402	25%	28%	
5	87,586	27%	30%	
6 or more	92,744	28%	27%	
Total				
Floor space				
<40 sq. m.	2,703	1%	2%	
40 to <70 sq. m.	34,764	11%	11%	
70 to <100 sq. m.	115,433	35%	40%	
100 to <150 sq. m.	136,532	42%	35%	
150 sq. m. or more	39,415	12%	12%	
Construction period				**Census periods**
Not given	105,280	32%		
Before 1914	41,975	13%		
1914-1947	45,173	14%	35%	before 1949
1948-1969	38,267	12%	21%	1949-1974
1970-1980	42,099	13%	13%	1975-1981
1981-1991	28,212	9%	11%	1982-1989
after 1991	27,841	8%	20%	1990-
Garage, parking space				
0 or not given	124,530	38%	21%	no
1	170,351	52%	79%	yes
2 ou plus	33,966	10%		
Bathrooms				
0 or not given	21,389	7 %		
1	235,118	71 %		
2 or more	72,340	22 %		
Floors				
0 or not given	29,536	9%		
1	79,051	24%		
2	183,452	56%		
3 or more	36,808	11%		
Plot size				
<500 sq. m.	135,690	41%		
500-1,000 sq. m.	89,983	27%		
1,000-1,500 sq. m.	39,356	12%		
1,500-2,500 sq. m.	30,269	9%		
>2,500 sq. m.	33,549	10%		
Condition of property				
Not given	144,590	44%		
Good	122,674	37%		
Needs work	42,631	13%		
Needs renovation	18,952	6%		
Basement				
No or not given	251,234	76%		
Yes	77,613	24%		

Chapter 5: Estimating the hedonic equation, practical application

5.1 Estimating the hedonic equation in two steps

The vector of relative prices of the characteristics is obtained after two estimation steps, where the second step reduces the influence of extreme values. The aim is to achieve a more robust estimation of the coefficients of relative prices.

See Appendix 3 for new additions in version 3 on model specification.

Step 1: determining the estimation stock

An estimation of the basic model is carried out, using the reference stock for each stratum, as mentioned in Chapter 3.

After this first estimation, extreme observations are eliminated, associated with a large residual in absolute value which has little compatibility with the postulated model. The aim is to produce a robust estimation of the coefficients of the model, which is less sensitive to the specific characteristics of the data sample used. In this way, the empirical distribution of the residuals is closer to a normal law (0, σ^2), for which the frequency of very large observations in proportion to the standard deviation (more than 3 standard deviations) and in absolute value is negligible.

There are several methods that can be used to detect these points. Here, we use the standardised residuals method. We standardise estimated residual $\hat{\varepsilon}_i$ using the standard deviation of estimated variance $\hat{\varepsilon}_i$. Using this method, it is possible to detect non-standard observations in cases where the estimated residual $\hat{r}_i$ from observation i is too large.

The formula for calculation for observation i is therefore:

$$\hat{r}_i = \hat{\varepsilon}_i \Big/ \hat{s}\sqrt{1-h_{ii}} \qquad (5.1)$$

where $\hat{s}$: the square root of estimated variance of $\hat{\varepsilon}_i$ equal to $\hat{s}^2 = \dfrac{\sum_{i=1}^{n} \hat{\varepsilon}_i^2}{n-(k+1)}$

n : number of observations

k+1 : number of parameters of the econometric equation

$\hat{\varepsilon}_i$, : estimated residuals of the regression

$h_{ii} = x_i'(X'X)^{-1}x_i$: lever effect of observation i

The amount (5.1) follows a Beta law if the numerator and the denominator are independent. This amount shows whether the standardised residual $\hat{r}_i$ is atypical or not and acts as a test statistic for the hypotheses:

H0: observation i is not atypical;

H1: observation i is atypical.

After carrying out the regression, we remove from the estimation stock all observations where the standardised residual $\hat{r}_i$ is not included in the interval $-2 \prec \hat{r}_i \prec 2$, for a significance level of α=0.05. At the end of this step, we have a new estimation stock which can give a more robust estimation of the hedonic model.

Step 2: refining specification to make it parsimonious

A new estimation of the econometric equation was carried out on the estimation stock from step 1. The aim of step 2 was to reduce the number of explanatory variables and thus increase the validity and the robustness of the model. Thus at the same time, we reduce the risk of colinearity between variables, a risk that increases with a greater number of variables and which increases the dependency of the estimated values to the available sample. The logic behind the selection of variables is to bring into the equation only variables that are most correlated with the explanatory variable. We are therefore looking for a combination of variables which, when associated with the explanatory variable, give a high R^2 when the model is estimated. There are various methods for selecting variables; we used the ascending method, known as the Forward method.

The Forward method is a step-by-step method, where variables are added one after the other. At each step we use an entry test for the variable in the equation (Fisher's test). The process starts with a model with no variables at step 0. At step 1, we choose the variable that most accounts for the increase in the explained variation. The process stops when none of the remaining variables satisfies the entry conditions (entry threshold fixed at 0.5) or, in other words, when no variable has a test value above this threshold. The entry threshold [French *seuil*] for the j^{th} variable corresponds to a Fisher's test: we compare the sum of the squares (SSE_{M1}) of the complete model *(M1)* made up of the first *j* variables, with the sum of the squares (SSE_{M2}) of model *(M2)* limited to the first *j-1* variables, i.e.:

$$Seuil_{Variable_j} = \frac{\left(SSE_{M1} - SSE_{M2}\right)}{SSE_{M2} \big/ (n-k-1)}$$

where n is the number of observations and k is the number of variables in the model.

5.2. Numerical examples

We present below six examples of models, successively for apartments and houses in three different strata each time. In each model and for each descriptive variable, a reference property was chosen, which defines the reference modality. In the second "robust" step, the estimation procedure as described in the previous paragraph is based on a Forward approach. Variables are introduced progressively in the model, according to their explanatory power (that is to say, their contribution to the reduction of the sample variance of the error terms of the regression).

Following this step, a number of variables have not been introduced in the equation due to their influence not significantly different from the associated reference modality. The lines marked "ns" in the tables below correspond to these unselected modalities. For example, in Table 5.1, the final model is obtained after 49 variables were added to the constant exploratory in 48 exploratory steps. This model is satisfactory insofar 73.1% of the variance in prices per square meter (in logarithm) are associated with these 49 variables, describing the properties and therefore taken into account by the model.

5.2.1. Examples of regressions (2009-2010 stock)

Apartments

Table 5.1 - Apartments in stratum 3 in Paris Region, excluding Paris

(R^2 = 0.731; s=0.190;[46] number of observations: 8,059 – 48 steps in FORWARD)

Variables	Coefficient	Standard deviation	P-value
(Intercept)	7.844	0.009	0.000
Year 2009	-0.045	0.004	0.000
Year 2010	Reference		

[46] s is the mean quadratic error (or root mean square deviation).

Variables	Coefficient	Standard deviation	P-value
January	-0.025	0.008	0.002
February	-0.033	0.007	0.000
March	-0.052	0.007	0.000
April	-0.042	0.007	0.000
May	-0.038	0.008	0.000
June	-0.018	0.007	0.006
July	-0.023	0.006	0.000
August	ns	-	-
September	ns	-	-
October	ns	-	-
November	ns	-	-
December	Reference		
Before 1850	ns	-	-
1850 – 1913	0.108	0.034	0.002
1914 – 1947	0.078	0.022	0.000
1948 – 1969	Reference		
1970 – 1980	-0.032	0.005	0.000
1981 – 1991	0.022	0.005	0.000
1992 – 2010	0.115	0.006	0.000
Construction period unknown	ns	-	-
0 bathrooms	ns	-	-
1 bathroom	Reference		
2+ bathrooms	0.018	0.008	0.032
0 cellar	Reference		
1+ cellar	0.041	0.004	0.000
0 garage	Reference		
1 garage	0.071	0.006	0.000
2+ garages	0.113	0.008	0.000
Ground floor	Reference		
1st floor	ns	-	-
2nd floor	ns	-	-
3rd floor	ns	-	-
4th floor with lift	-0.034	0.005	0.000
4th floor without lift	ns	-	-
Floor space per room of studios <20 sq.m.	-0.050	0.017	0.003
Floor space per room of studios 20-30 sq.m.	Reference		
Floor space per room of studios >30 sq.m.	-0.103	0.013	0.000
Floor space per room of 2-room apt. <17 sq.m.	0.157	0.018	0.000
Floor space per room of 2-room apt. 17-24 sq.m.	Reference		
Floor space per room of 2-room apt. >24 sq.m.	-0.124	0.008	0.000
Floor space per room of 3-room apt. <18 sq.m.	0.061	0.017	0.000
Floor space per room of 3-room apt. 18-22 sq.m.	Reference		
Floor space per room of 3-room apt.>22 sq.m.	-0.081	0.007	0.000
Floor space per room of 4-room apt. and more <17 sq.m.	ns	-	-
Floor space per room of 4-room apt. 17-21 sq.m.	Reference		
Floor space per room of 4-room apt. and more >21 sq.m.	-0.059	0.007	0.000
Studio in neighbourhood 1	0.293	0.014	0.000
2 rooms in neighbourhood 1	0.145	0.010	0.000
3 rooms in neighbourhood 1	Reference		
4 rooms in neighbourhood 1	-0.102	0.010	0.000
5+ rooms in neighbourhood 1	-0.253	0.016	0.000
Studio in neighbourhood 2	0.141	0.018	0.000
2 rooms in neighbourhood 2	-0.043	0.012	0.000
3 rooms in neighbourhood 2	-0.209	0.009	0.000
4 rooms in neighbourhood 2	-0.308	0.010	0.000
5+ rooms in neighbourhood 2	-0.419	0.015	0.000
Studio in neighbourhood 3	0.338	0.015	0.000
2 rooms in neighbourhood 3	0.181	0.011	0.000
3 rooms in neighbourhood 3	0.033	0.010	0.001
4 rooms in neighbourhood 3	-0.091	0.011	0.000
5+ rooms in neighbourhood 3	-0.193	0.022	0.000

Variables	Coefficient	Standard deviation	P-value
Studio in neighbourhood 4	0.346	0.017	0.000
2 rooms in neighbourhood 4	0.174	0.012	0.000
3 rooms in neighbourhood 4	0.019	0.009	0.039
4 rooms in neighbourhood 4	-0.132	0.010	0.000
5+ rooms in neighbourhood 4	-0.232	0.015	0.000
Studio in neighbourhood 5	0.410	0.030	0.000
2 rooms in neighbourhood 5	0.302	0.016	0.000
3 rooms in neighbourhood 5	0.214	0.011	0.000
4 rooms in neighbourhood 5	0.079	0.014	0.000
5+ rooms in neighbourhood 5	ns	-	-

Table 5.2 - Apartments in stratum 7502 in Paris

($R^2 = 0.280$; s=0.190; number of observations: 13,702 - 58 steps in FORWARD)

Variables	Coefficient	Standard deviation	P-value
(Intercept)	8.717	0.008	0.000
Year 2009	-0.118	0.003	0.000
Year 2010	Reference		
January	-0.101	0.007	0.000
February	-0.101	0.007	0.000
March	-0.115	0.007	0.000
April	-0.097	0.007	0.000
May	-0.098	0.007	0.000
June	-0.079	0.006	0.000
July	-0.056	0.006	0.000
August	-0.048	0.007	0.000
September	-0.036	0.006	0.000
October	-0.032	0.006	0.000
November	-0.019	0.007	0.006
December	Reference		
Before 1850	0.072	0.011	0.000
1850 – 1913	0.035	0.004	0.000
1914 – 1947	0.034	0.005	0.000
1948 – 1969	Reference		
1970 – 1980	0.020	0.006	0.001
1981 – 1991	0.054	0.010	0.000
1992 – 2010	0.070	0.010	0.000
Construction period unknown	0.037	0.005	0.000
0 bathrooms	-0.056	0.007	0.000
1 bathroom	Reference		
2+ bathrooms	0.045	0.009	0.000
Number of bathrooms unknown	-0.027	0.005	0.000
0 cellar	Reference		
1+ cellars	ns	-	-
0 garage	Reference		
1 garage	ns	-	-
2+ garages	0.062	0.019	0.001
Ground floor	Reference		
1st floor	0.013	0.005	0.017
2nd floor	0.037	0.005	0.000
3rd floor	0.041	0.005	0.000
4th floor with lift	0.056	0.005	0.000
4th floor without lift	0.045	0.009	0.000
Floor space per room of studios <20 sq.m.	0.036	0.007	0.000
Floor space per room of studios 20-30 sq.m.	Reference		
Floor space per room of studios >30 sq.m.	-0.025	0.009	0.007
Floor space per room of 2-room apt. <17 sq.m.	ns	-	-
Floor space per room of 2-room apt. 17-24 sq.m.	Reference		
Floor space per room of 2-room apt. >24 sq.m.	ns	-	-
Floor space per room of 3-room apt. <18 sq.m.	-0.024	0.006	0.000
Floor space per room of 3-room apt. 18-22 sq.m.	Reference		
Floor space per room of 3-room apt. >22 sq.m.	ns	-	-

Variables	Coefficient	Standard deviation	P-value
Floor space per room of 4-room apt. and more <17 sq.m.	ns	-	-
Floor space per room of 4-room apt. 17-21 sq.m.		Reference	
Floor space per room of 4-room apt. and more >21 sq.m.	0.023	0.008	0.003
Studio in neighbourhood 1	-0.010	0.007	0.167
2 rooms in neighbourhood 1	-0.025	0.006	0.000
3 rooms in neighbourhood 1		Reference	
4 rooms in neighbourhood 1	ns	-	-
5+ rooms in neighbourhood 1	ns	-	-
Studio in neighbourhood 2	0.237	0.014	0.000
2 rooms in neighbourhood 2	0.219	0.011	0.000
3 rooms in neighbourhood 2	0.223	0.014	0.000
4 rooms in neighbourhood 2	0.193	0.021	0.000
5+ rooms in neighbourhood 2	0.285	0.032	0.000
Studio in neighbourhood 3	0.073	0.008	0.000
2 rooms in neighbourhood 3	0.074	0.006	0.000
3 rooms in neighbourhood 3	0.092	0.007	0.000
4 rooms in neighbourhood 3	0.081	0.010	0.000
5+ rooms in neighbourhood 3	0.089	0.014	0.000
Studio in neighbourhood 4	0.173	0.014	0.000
2 rooms in neighbourhood 4	0.131	0.010	0.000
3 rooms in neighbourhood 4	0.129	0.013	0.000
4 rooms in neighbourhood 4	0.101	0.019	0.000
5+ rooms in neighbourhood 4	ns	-	-
Studio in neighbourhood 5	0.051	0.010	0.000
2 rooms in neighbourhood 5	0.036	0.007	0.000
3 rooms in neighbourhood 5	0.092	0.008	0.000
4 rooms in neighbourhood 5	0.098	0.011	0.000
5+ rooms in neighbourhood 5	0.110	0.017	0.000
Studio in neighbourhood 6	-0.022	0.011	0.035
2 rooms in neighbourhood 6	-0.037	0.008	0.000
3 rooms in neighbourhood 6	-0.035	0.010	0.001
4 rooms in neighbourhood 6	-0.066	0.015	0.000
5+ rooms in neighbourhood 6	-0.071	0.025	0.004

Table 5.3 - Apartments in Toulouse

($R^2 = 0.57$; s=0.17; number of observations: 5,538 - 70 steps in FORWARD)

Variables	Coefficient	Standard deviation	P-value
(Intercept)	7.505	0.013	0.000
Year 2009	-0.053	0.005	0.000
Year 2010		Reference	
January	-0.023	0.009	0.011
February	-0.051	0.010	0.000
March	-0.038	0.009	0.000
April	-0.048	0.010	0.000
May	-0.026	0.009	0.007
June	ns	-	-
July	ns	-	-
August	ns	-	-
September	ns	-	-
October	ns	-	-
November	ns	-	-
December		Reference	
Before 1850	0.103	0.024	0.000
1850 – 1913	0.144	0.018	0.000
1914 – 1947	0.130	0.017	0.000
1948 – 1969		Reference	
1970 – 1980	0.001	0.010	0.883
1981 – 1991	0.084	0.010	0.000
1992 – 2010	0.080	0.009	0.000
Construction period unknown	0.076	0.009	0.000
0 bathrooms	0.026	0.013	0.044
1 bathroom		Reference	
2+ bathrooms	0.084	0.015	0.000
Number of bathrooms unknown	ns	-	-

Variables	Coefficient	Standard deviation	P-value
0 cellar		Reference	
1+ cellars	-0.064	0.006	0.000
Number of cellars unknown	ns	-	-
0 garage		Reference	
1 garage	0.077	0.006	0.000
2+ garages	0.155	0.011	0.000
Ground floor		Reference	
1st floor	ns	-	-
2nd floor	ns	-	-
3rd floor	ns	-	-
4th floor with lift	-0.019	0.006	0.003
4th floor without lift	ns	-	-
Floor space per room of studios <20 sq.m.	0.144	0.014	0.000
Floor space per room of studios 20-30 sq.m.		Reference	
Floor space per room of studios >30 sq.m.	-0.103	0.013	0.000
Floor space per room of 2-room apt. <17 sq.m.	0.082	0.012	0.000
Floor space per room of 2-room apt. 17-24 sq.m.		Reference	
Floor space per room of 2-room apt. >24 sq.m.	-0.047	0.010	0.000
Floor space per room of 3-room apt. <18 sq.m.	ns	-	-
Floor space per room of 3-room apt. 18-22 sq.m.		Reference	
Floor space per room of 3-room apt.>22 sq.m.	0.023	0.009	0.016
Floor space per room of 4-room apt. and more <17 sq.m.	ns	-	-
Floor space per room of 4-room apt. 17-21 sq.m.		Reference	
Floor space per room of 4-room apt. and more >21 sq.m.	0.058	0.011	0.000
Good condition		Reference	
Needs work	-0.062	0.010	0.000
Needs renovation	-0.187	0.019	0.000
Condition unknown	-0.037	0.006	0.000
Without terrace or balcony		Reference	
With terrace(s) or balcony(ies)	0.049	0.005	0.000
Studio in neighbourhood 1	0.176	0.021	0.000
2 rooms in neighbourhood 1	0.010	0.019	0.597
3 rooms in neighbourhood 1		Reference	
4 rooms in neighbourhood 1	-0.249	0.019	0.000
5+ rooms in neighbourhood 1	-0.248	0.030	0.000
Studio in neighbourhood 2	0.303	0.018	0.000
2 rooms in neighbourhood 2	0.194	0.018	0.000
3 rooms in neighbourhood 2	0.096	0.021	0.000
4 rooms in neighbourhood 2	ns	-	-
5+ rooms in neighbourhood 2	ns	-	-
Studio in neighbourhood 3	0.534	0.021	0.000
2 rooms in neighbourhood 3	0.522	0.022	0.000
3 rooms in neighbourhood 3	0.520	0.022	0.000
4 rooms in neighbourhood 3	0.482	0.032	0.000
5+ rooms in neighbourhood 3	0.468	0.029	0.000
Studio in neighbourhood 4	0.277	0.027	0.000
2 rooms in neighbourhood 4	0.219	0.020	0.000
3 rooms in neighbourhood 4	0.114	0.020	0.000
4 rooms in neighbourhood 4	0.148	0.022	0.000
5+ rooms in neighbourhood 4	0.096	0.044	0.031
Studio in neighbourhood 5	0.233	0.029	0.000
2 rooms in neighbourhood 5	0.145	0.020	0.000
3 rooms in neighbourhood 5	0.056	0.017	0.001
4 rooms in neighbourhood 5	0.002	0.020	0.916
5+ rooms in neighbourhood 5	ns	-	-
Studio in neighbourhood 6	0.185	0.027	0.000
2 rooms in neighbourhood 6	0.063	0.015	0.000
3 rooms in neighbourhood 6	-0.017	0.015	0.254
4 rooms in neighbourhood 6	-0.097	0.019	0.000
5+ rooms in neighbourhood 6	-0.149	0.044	0.001

Variables	Coefficient	Standard deviation	P-value
Studio in neighbourhood 7	0.274	0.028	0.000
2 rooms in neighbourhood 7	0.188	0.018	0.000
3 rooms in neighbourhood 7	0.084	0.017	0.000
4 rooms in neighbourhood 7	-0.043	0.019	0.020
5+ rooms in neighbourhood 7	-0.126	0.032	0.000
Studio in neighbourhood 8	0.202	0.026	0.000
2 rooms in neighbourhood 8	0.104	0.020	0.000
3 rooms in neighbourhood 8	0.047	0.020	0.021
4 rooms in neighbourhood 8	-0.066	0.024	0.006
5+ rooms in neighbourhood 8	-0.131	0.040	0.001
Studio in neighbourhood 9	0.384	0.015	0.000
2 rooms in neighbourhood 9	0.313	0.014	0.000
3 rooms in neighbourhood 9	0.277	0.014	0.000
4 rooms in neighbourhood 9	0.252	0.016	0.000
5+ rooms in neighbourhood 9	0.322	0.021	0.000

Table 5.4 - Apartments in Lyon

($R^2 = 0.44$; s=0.17; number of observations: 7,168 – 69 steps in FORWARD)

Variables	Coefficient	Standard deviation	P-value
(Intercept)	7.947	0.010	0.000
Year 2009	-0.086	0.004	0.000
Year 2010	Reference		
January	-0.050	0.009	0.000
February	-0.054	0.009	0.000
March	-0.062	0.008	0.000
April	-0.047	0.008	0.000
May	-0.044	0.008	0.000
June	-0.025	0.007	0.000
July	ns	-	-
August	ns	-	-
September	ns	-	-
October	ns	-	-
November	ns	-	-
December	Reference		
Before 1850	0.102	0.014	0.000
1850 – 1913	0.079	0.009	0.000
1914 – 1947	0.041	0.009	0.000
1948 – 1969	Reference		
1970 – 1980	-0.031	0.008	0.000
1981 – 1991	0.061	0.008	0.000
1992 – 2010	0.132	0.008	0.000
Construction period unknown	0.050	0.006	0.000
0 bathroom	ns	-	-
1 bathroom	Reference		
2+ bathrooms	0.040	0.009	0.000
Number of bathrooms unknown	-0.041	0.009	0.000
0 cellar	Reference		
1+ cellar	ns	-	-
Number of cellars unknown	ns	-	-
0 garage	Reference		
1 garage	0.044	0.005	0.000
2+ garages	0.092	0.010	0.000
Ground floor	Reference		
1st floor	0.035	0.007	0.000
2nd floor	0.053	0.007	0.000
3rd floor	0.045	0.007	0.000
4th floor with lift	0.056	0.006	0.000
4th floor without lift	ns	-	-

Variables	Coefficient	Standard deviation	P-value
Floor space per room of studios <20 sq.m.	0.097	0.018	0.000
Floor space per room of studios 20-30 sq.m.		Reference	
Floor space per room of studios >30 sq.m.	-0.127	0.011	0.000
Floor space per room of 2-room apt. <17 sq.m.	0.038	0.015	0.011
Floor space per room of 2-room apt. 17-24 sq.m.		Reference	
Floor space per room of 2-room apt. >24 sq.m.	-0.057	0.008	0.000
Floor space per room of 3-room apt. <18 sq.m.	0.060	0.012	0.000
Floor space per room of 3-room apt. 18-22 sq.m.		Reference	
Floor space per room of 3-room apt.>22 sq.m.	ns	-	-
Floor space per room of 4-room apt. and more <17 sq.m.	ns	-	-
Floor space per room of 4-room apt. 17-21 sq.m.		Reference	
Floor space per room of 4-room apt. and more >21 sq.m.	0.032	0.007	0.000
Good condition		Reference	
Needs work	-0.104	0.007	0.000
Needs renovation	-0.199	0.014	0.000
Condition unknown	-0.020	0.005	0.000
Without terrace or balcony		Reference	
With terrace(s) or balcony(ies)	0.023	0.005	0.000
Studio in neighbourhood 1	0.134	0.021	0.000
2 rooms in neighbourhood 1	0.036	0.014	0.014
3 rooms in neighbourhood 1		Reference	
4 rooms in neighbourhood 1	ns	-	-
5+ rooms in neighbourhood 1	ns	-	-
Studio in neighbourhood 2	0.258	0.025	0.000
2 rooms in neighbourhood 2	0.145	0.020	0.000
3 rooms in neighbourhood 2	0.048	0.018	0.007
4 rooms in neighbourhood 2	0.064	0.021	0.002
5+ rooms in neighbourhood 2	0.073	0.029	0.010
Studio in neighbourhood 3	0.069	0.015	0.000
2 rooms in neighbourhood 3	-0.049	0.011	0.000
3 rooms in neighbourhood 3	-0.096	0.010	0.000
4 rooms in neighbourhood 3	-0.121	0.011	0.000
5+ rooms in neighbourhood 3	-0.134	0.015	0.000
Studio in neighbourhood 4	0.155	0.022	0.000
2 rooms in neighbourhood 4	0.059	0.015	0.000
3 rooms in neighbourhood 4	ns	-	-
4 rooms in neighbourhood 4	ns	-	-
5+ rooms in neighbourhood 4	ns	-	-
Studio in neighbourhood 5	-0.081	0.022	0.000
2 rooms in neighbourhood 5	-0.111	0.017	0.000
3 rooms in neighbourhood 5	-0.199	0.014	0.000
4 rooms in neighbourhood 5	-0.260	0.016	0.000
5+ rooms in neighbourhood 5	-0.286	0.022	0.000
Studio in neighbourhood 6	0.123	0.026	0.000
2 rooms in neighbourhood 6	ns	-	-
3 rooms in neighbourhood 6	-0.155	0.013	0.000
4 rooms in neighbourhood 6	-0.227	0.012	0.000
5+ rooms in neighbourhood 6	-0.199	0.016	0.000
Studio in neighbourhood 7	0.232	0.019	0.000
2 rooms in neighbourhood 7	0.139	0.014	0.000
3 rooms in neighbourhood 7	0.114	0.012	0.000
4 rooms in neighbourhood 7	0.085	0.014	0.000
5+ rooms in neighbourhood 7	0.100	0.015	0.000
Studio in neighbourhood 8	ns	-	-
2 rooms in neighbourhood 8	-0.072	0.012	0.000
3 rooms in neighbourhood 8	-0.134	0.011	0.000
4 rooms in neighbourhood 8	-0.199	0.014	0.000
5+ rooms in neighbourhood 8	-0.161	0.019	0.000
Studio in neighbourhood 9	0.057	0.020	0.004
2 rooms in neighbourhood 9	-0.096	0.013	0.000
3 rooms in neighbourhood 9	-0.201	0.012	0.000
4 rooms in neighbourhood 9	-0.273	0.013	0.000
5+ rooms in neighbourhood 9	-0.332	0.019	0.000

Houses

Table 5.5 – Houses in stratum 5 in Paris Region [47]

(R^2 = 0.678; s=0.190; number of observations: 7120 - 34 steps in FORWARD)

Variables	Coefficient	Standard deviation	P-value
(Intercept)	9.995	0.052	0.000
Living space (sq.m.)	0.353	0.012	0.000
Plot size (sq.m.)	0.148	0.003	0.000
Year 2009	-0.049	0.003	0.000
Year 2010	Reference		
January	ns	-	-
February	-0.019	0.007	0.005
March	-0.016	0.007	0.024
April	-0.019	0.007	0.005
May	-0.016	0.007	0.014
June	ns	-	-
July	ns	-	-
August	ns	-	-
September	ns	-	-
October	ns	-	-
November	ns	-	-
December	Reference		
Before 1913	ns	-	-
1914 – 1947	ns	-	-
1948 – 1969	-0.019	0.008	0.000
1970 – 1980	-0.018	0.005	0.000
After 1980	0.037	0.004	0.000
Period of construction unknown	Reference		
0 bathroom	-0.163	0.032	0.000
1 bathroom	Reference		
2 bathrooms	0.050	0.004	0.000
3+ bathrooms	-0.078	0.013	0.000
0 cellar	Reference		
1 cellar	0.026	0.005	0.000
2+ cellars	ns	-	-
0 garage	-0.036	0.005	0.000
1 garage	Reference		
2+ garages	0.039	0.006	0.000
1 floor	-0.041	0.005	0.000
2 floors	Reference		
3+ floors	ns	-	-
Up to 3 rooms in neighbourhood 1	-0.059	0.009	0.000
4 rooms in neighbourhood 1	Reference		
5 rooms in neighbourhood 1	ns	-	-
6 rooms in neighbourhood 1	0.024	0.007	0.000
7+ rooms in neighbourhood 1	0.042	0.009	0.000
Up to 3 rooms in neighbourhood 2	ns	-	-
4 rooms in neighbourhood 2	-0.033	0.007	0.000
5 rooms in neighbourhood 2	-0.037	0.006	0.000
6 rooms in neighbourhood 2	-0.028	0.008	0.000
7+ rooms in neighbourhood 2	-0.040	0.011	0.000
Up to 3 rooms in neighbourhood 3	0.120	0.021	0.000
4 rooms in neighbourhood 3	0.169	0.010	0.000
5 rooms in neighbourhood 3	0.158	0.008	0.000
6 rooms in neighbourhood 3	0.188	0.012	0.000
7+ rooms in neighbourhood 3	0.182	0.016	0.000
Up to 3 rooms in neighbourhood 4	0.139	0.017	0.000
4 rooms in neighbourhood 4	0.112	0.013	0.000
5 rooms in neighbourhood 4	0.129	0.012	0.000
6 rooms in neighbourhood 4	0.153	0.014	0.000
7+ rooms in neighbourhood 4	0.188	0.022	0.000

[47] In this example, the variable "number of buildings" does not appear. This variable is present in only one stratum.

Table 5.6 – Houses in suburbs of Lille

(R^2 = 0.85; s=0.17; number of observations: 6,702 - 75 steps in FORWARD)

Variables	Coefficient	Standard deviation	P-value
(Intercept)	8.689	0.047	0.000
Living space (sq.m.)	0.531	0.011	0.000
Plot size (sq.m.)	0.180	0.004	0.000
Year 2009	-0.053	0.004	0.000
Year 2010	Reference		
January	-0.040	0.009	0.000
February	ns	-	-
March	-0.021	0.009	0.018
April	-0.033	0.009	0.000
May	-0.026	0.009	0.004
June	-0.016	0.007	0.026
July	ns	-	-
August	ns	-	-
September	ns	-	-
October	ns	-	-
November	ns	-	-
December	Reference		
Before 1913	-0.064	0.011	0.000
1914 – 1947	-0.039	0.009	0.000
1948 – 1969	-0.008	0.009	0.421
1970 – 1980	0.041	0.010	0.000
After 1980	0.137	0.010	0.000
Period of construction unknown	Reference		
0 bathrooms	ns	-	-
1 bathrooms	Reference		
2+ bathrooms	0.123	0.007	0.000
Presence of basement	0.019	0.005	0.000
No basement	Reference		
0 garage	-0.055	0.005	0.000
1 garage	Reference		
2+ garages	0.026	0.008	0.002
1 floor	ns	-	-
2 floors	Reference		
3+ floors	-0.016	0.005	0.002
Good condition	0.029	0.006	0.000
Needs work	-0.138	0.007	0.000
Needs renovating	-0.357	0.012	0.000
Condition unknown	Reference		
Up to 3 rooms in neighbourhood 1	ns	-	-
4 rooms in neighbourhood 1	Reference		
5 rooms in neighbourhood 1	0.116	0.017	0.000
6 rooms in neighbourhood 1	0.170	0.023	0.000
7+ rooms in neighbourhood 1	0.252	0.034	0.000
Up to 3 rooms in neighbourhood 2	ns	-	-
4 rooms in neighbourhood 2	ns	-	-
5 rooms in neighbourhood 2	0.076	0.019	0.000
6 rooms in neighbourhood 2	ns	-	-
7+ rooms in neighbourhood 2	ns	-	-
Up to 3 rooms in neighbourhood 3	ns	-	-
4 rooms in neighbourhood 3	-0.039	0.017	0.021
5 rooms in neighbourhood 3	ns	-	-
6 rooms in neighbourhood 3	ns	-	-
7+ rooms in neighbourhood 3	-0.142	0.072	0.047
Up to 3 rooms in neighbourhood 4	-0.088	0.029	0.003
4 rooms in neighbourhood 4	ns	-	-
5 rooms in neighbourhood 4	-0.092	0.022	0.000
6 rooms in neighbourhood 4	-0.201	0.040	0.000
7+ rooms in neighbourhood 4	ns	-	-

Variables	Coefficient	Standard deviation	P-value
Up to 3 rooms in neighbourhood 5	0.096	0.038	0.013
4 rooms in neighbourhood 5	0.197	0.021	0.000
5 rooms in neighbourhood 5	0.190	0.018	0.000
6 rooms in neighbourhood 5	0.308	0.028	0.000
7+ rooms in neighbourhood 5	0.344	0.032	0.000
Up to 3 rooms in neighbourhood 6	ns	-	-
4 rooms in neighbourhood 6	ns	-	-
5 rooms in neighbourhood 6	-0.064	0.023	0.006
6 rooms in neighbourhood 6	ns	-	-
7+ rooms in neighbourhood 6	ns	-	-
Up to 3 rooms in neighbourhood 7	0.129	0.047	0.006
4 rooms in neighbourhood 7	0.170	0.027	0.000
5 rooms in neighbourhood 7	0.236	0.024	0.000
6 rooms in neighbourhood 7	0.335	0.032	0.000
7 rooms in neighbourhood 7	0.472	0.044	0.000
Up to 3 rooms in neighbourhood 8	0.164	0.032	0.000
4 rooms in neighbourhood 8	0.212	0.016	0.000
5 rooms in neighbourhood 8	0.242	0.016	0.000
6 rooms in neighbourhood 8	0.234	0.020	0.000
7+ rooms in neighbourhood 8	0.421	0.025	0.000
Up to 3 rooms in neighbourhood 9	0.150	0.035	0.000
4 rooms in neighbourhood 9	0.154	0.022	0.000
5 rooms in neighbourhood 9	0.219	0.024	0.000
6 rooms in neighbourhood 9	0.191	0.030	0.000
7+ rooms in neighbourhood 9	0.295	0.036	0.000
Up to 3 rooms in neighbourhood 10	ns	-	-
4 rooms in neighbourhood 10	0.045	0.021	0.032
5 rooms in neighbourhood 10	ns	-	-
6 rooms in neighbourhood 10	ns	-	-
7+ rooms in neighbourhood 10	0.230	0.078	0.003
Up to 3 rooms in neighbourhood 11	-0.295	0.023	0.000
4 rooms in neighbourhood 11	-0.200	0.013	0.000
5 rooms in neighbourhood 11	-0.187	0.013	0.000
6 rooms in neighbourhood 11	-0.158	0.020	0.000
7+ rooms in neighbourhood 11	-0.195	0.024	0.000
Up to 3 rooms in neighbourhood 12	0.074	0.032	0.020
4 rooms in neighbourhood 12	0.144	0.019	0.000
5 rooms in neighbourhood 12	0.084	0.020	0.000
6 rooms in neighbourhood 12	0.175	0.029	0.000
7+ rooms in neighbourhood 12	0.135	0.045	0.003
Up to 3 rooms in neighbourhood 13	-0.114	0.025	0.000
4 rooms in neighbourhood 13	-0.086	0.014	0.000
5 rooms in neighbourhood 13	-0.107	0.017	0.000
6 rooms in neighbourhood 13	-0.128	0.030	0.000
7+ rooms in neighbourhood 13	-0.144	0.058	0.014
Up to 3 rooms in neighbourhood 14	-0.063	0.023	0.006
4 rooms in neighbourhood 14	ns	-	-
5 rooms in neighbourhood 14	-0.030	0.012	0.015
6 rooms in neighbourhood 14	ns	-	-
7+ rooms in neighbourhood 14	ns	-	-
Up to 3 rooms in neighbourhood 15	ns	-	-
4 rooms in neighbourhood 15	ns	-	-
5 rooms in neighbourhood 15	ns	-	-
6 rooms in neighbourhood 15	ns	-	-
7+ rooms in neighbourhood 15	ns	-	-
Up to 3 rooms in neighbourhood 16	-0.062	0.019	0.001
4 rooms in neighbourhood 16	ns	-	-
5 rooms in neighbourhood 16	ns	-	-
6 rooms in neighbourhood 16	ns	-	-
7+ rooms in neighbourhood 16	ns	-	-
Up to 3 rooms in neighbourhood 17	ns	-	-
4 rooms in neighbourhood 17	0.059	0.013	0.000
5 rooms in neighbourhood 17	0.118	0.011	0.000
6 rooms in neighbourhood 17	0.132	0.015	0.000
7+ rooms in neighbourhood 17	0.153	0.021	0.000

Table 5.7 – Houses in suburbs of Bordeaux

(R^2 = 0.67; s=0.19; number of observations: 3 818 - 53 steps in FORWARD)

Variables	Coefficient	Standard deviation	P-value
(Intercept)	9.414	0.074	0.000
Living space (sq.m.)	0.534	0.017	0.000
Plot size (sq.m.)	0.074	0.005	0.000
Year 2009	-0.063	0.006	0.000
Year 2010	Reference		
January	-0.030	0.014	0.034
February	-0.054	0.014	0.000
March	-0.078	0.013	0.000
April	-0.044	0.013	0.001
May	-0.058	0.013	0.000
June	-0.026	0.011	0.015
July	ns	-	-
August	0.022	0.011	0.040
September	ns	-	-
October	ns	-	-
November	ns	-	-
December	Reference		
Before 1913	0.073	0.021	0.001
1914 – 1947	0.067	0.012	0.000
1948 – 1969	ns	-	-
1970 – 1980	ns	-	-
After 1980	0.066	0.008	0.000
Period of construction unknown	Reference		
0 bathrooms	ns	-	-
1 bathrooms	Reference		
2+ bathrooms	0.082	0.009	0.000
Presence of basement	0.036	0.015	0.013
No basement	Reference		
0 garage	-0.030	0.007	0.000
1 garage	Reference		
2+ garages	0.038	0.013	0.004
1 floor	0.017	0.007	0.013
2 floors	Reference		
3+ floors	ns	-	-
Good condition	0.014	0.007	0.041
Needs work	-0.102	0.013	0.000
Needs renovating	-0.324	0.020	0.000
Condition unknown	Reference		
Up to 3 rooms in neighbourhood 1	ns	-	-
4 rooms in neighbourhood 1	Reference		
5 rooms in neighbourhood 1	0.091	0.021	0.000
6 rooms in neighbourhood 1	0.117	0.027	0.000
7+ rooms in neighbourhood 1	0.151	0.034	0.000
Up to 3 rooms in neighbourhood 2	ns	-	-
4 rooms in neighbourhood 2	0.029	0.019	0.128
5 rooms in neighbourhood 2	ns	-	-
6 rooms in neighbourhood 2	ns	-	-
7+ rooms in neighbourhood 2	ns	-	-
Up to 3 rooms in neighbourhood 3	ns	-	-
4 rooms in neighbourhood 3	ns	-	-
5 rooms in neighbourhood 3	ns	-	-
6 rooms in neighbourhood 3	ns	-	-
7+ rooms in neighbourhood 3	-0.090	0.042	0.031
Up to 3 rooms in neighbourhood 4	0.068	0.026	0.010
4 rooms in neighbourhood 4	0.106	0.026	0.000
5 rooms in neighbourhood 4	0.086	0.030	0.004
6 rooms in neighbourhood 4	0.156	0.043	0.000
7+ rooms in neighbourhood 4	ns	-	-

Variables	Coefficient	Standard deviation	P-value
Up to 3 rooms in neighbourhood 5	-0.117	0.031	0.000
4 rooms in neighbourhood 5	-0.065	0.022	0.003
5 rooms in neighbourhood 5	-0.069	0.022	0.002
6 rooms in neighbourhood 5	-0.108	0.042	0.011
7+ rooms in neighbourhood 5	ns	-	-
Up to 3 rooms in neighbourhood 6	-0.103	0.019	0.000
4 rooms in neighbourhood 6	-0.161	0.015	0.000
5 rooms in neighbourhood 6	-0.145	0.018	0.000
6 rooms in neighbourhood 6	-0.126	0.028	0.000
7+ rooms in neighbourhood 6	-0.239	0.045	0.000
Up to 3 rooms in neighbourhood 7	-0.282	0.023	0.000
4 rooms in neighbourhood 7	-0.181	0.015	0.000
5 rooms in neighbourhood 7	-0.166	0.015	0.000
6 rooms in neighbourhood 7	-0.112	0.022	0.000
7+ rooms in neighbourhood 7	-0.180	0.035	0.000
Up to 3 rooms in neighbourhood 8	-0.062	0.022	0.004
4 rooms in neighbourhood 8	-0.064	0.015	0.000
5 rooms in neighbourhood 8	0.011	0.016	0.480
6 rooms in neighbourhood 8	ns	-	-
7+ rooms in neighbourhood 8	0.074	0.028	0.008
Up to 3 rooms in neighbourhood 9	-0.136	0.028	0.000
4 rooms in neighbourhood 9	-0.076	0.017	0.000
5 rooms in neighbourhood 9	-0.044	0.015	0.004
6 rooms in neighbourhood 9	-0.053	0.019	0.005
7+ rooms in neighbourhood 9	0.033	0.025	0.184

5.2.2. Quality of hedonic regressions

The quality of the hedonic regressions is usually measured by the determination coefficient R^2, which ranges from 0 to 1. The higher the coefficient, the greater the regression's predictive power. For cross-sectional individual data, good values for R^2 are roughly 0.25-0.4 for 1,000 to 3,000 observations and about twenty explanatory variables. This is the level of quality that we observe here for the various strata. Remember that the dependent variable is the Napierian logarithm of the price per sq. m. for apartments and of the property price for houses.

Apartments

Table 5.8 – Quality of regressions and number of observations from estimation stock by stratum: Paris Region, apartments

Stratum	Steps	R^2	Estimation stock
1	45	0.478	5,348
2	48	0.635	2,860
3	48	0.731	4,615
4	52	0.601	9,287
5	52	0.508	11,833
6	58	0.519	17,154
7	28	0.681	3,225
8	51	0.640	8,667
9	39	0.292	11,139
10	50	0.415	16,746
7501	47	0.501	8,463
7502	58	0.280	13,702
7503	56	0.397	9,895
7504	32	0.393	4,251
7505	38	0.338	5,626
Total			132,811

Table 5.9 – Quality of regressions and number of observations from estimation stock by stratum: Provinces, apartments

Stratum	Steps	R^2	Estimation stock
Agglomération de Genève - Annemasse (partie française)	42	0,47	1 876
Agglomération de Bayonne	53	0,54	2 253
Banlieue de Toulon	57	0,60	3 895
Banlieue de Lyon	64	0,54	4 993
Banlieue de Marseille-Aix-en-Provence	49	0,60	1 939
Antibes	32	0,36	1 480
Cannes	49	0,49	2 470
Cannet	23	0,39	1 121
Menton	16	0,27	761
Nice	69	0,43	8 251
Banlieue de Nice	71	0,46	4 671
Marseille	86	0,59	6 289
Caen	23	0,48	993
Dijon	58	0,54	2 434
Besançon	32	0,46	1 806
Brest	45	0,43	1 750
Toulouse	70	0,57	5 538
Bordeaux	55	0,37	2 537
Banlieue de Bordeaux	46	0,60	2 058
Agde	23	0,41	1 412
Montpellier	48	0,45	1 947
Rennes	47	0,59	2 507
Tours	30	0,54	1 181
Grenoble	60	0,56	2 351
Banlieue de Grenoble	42	0,53	2 163
Saint-Etienne	17	0,32	1 490
Nantes	56	0,45	3 472
Agglomération de Saint-Nazaire	26	0,74	865
Banlieue de Nantes	32	0,65	1 081
Orléans	39	0,52	1 259
Reims	41	0,57	1 558
Nancy	25	0,25	1 484
Banlieue de Nancy	33	0,43	1 181
Metz	31	0,40	1 308
Lille	47	0,59	2 366
Banlieue de Lille (partie française)	40	0,59	2 032
Clermont-Ferrand	26	0,42	1 284
Pau	19	0,31	1 375
Strasbourg	39	0,50	2 422
Banlieue de Strasbourg (partie française)	30	0,43	1 302
Agglomération de Mulhouse	18	0,42	944
Lyon	69	0,44	7 168
Villeurbanne	30	0,46	1 854
Agglomération de Chambéry	25	0,47	1 244
Annecy	30	0,52	1 186
Banlieue de Annecy	35	0,63	1 269
Le Havre	34	0,50	1 269
Rouen	40	0,44	1 356
Banlieue de Rouen	25	0,56	795
Toulon	41	0,43	1 960
Littoral Grand Nord (Régions 31,22)	45	0,74	2 042
Littoral de Normandie (Région 25, 23)	49	0,65	2 249
Littoral Bretagne Est (dép. 35,56)	50	0,66	1 974
Littoral Bretagne Ouest (dép 22,29)	41	0,65	1 735
Littoral Pays de la Loire (Région 52)	28	0,38	1 069
Littoral Sud-Ouest (Régions 54,72)	49	0,46	2 695
Littoral Languedoc-Roussillon Sud (dép 11,66 et 34)	38	0,44	2 252
Littoral Languedoc-Roussillon Sud (dép. 34, 30)	40	0,58	1 612
Littoral Provence-Alpes-Côte d'Azur Ouest (dép. 13, 83)	28	0,25	1 374
Littoral Provence-Alpes-Côte d'Azur Est (dép. 83,06)	43	0,62	2 324
Stations de ski de Tarentaise	34	0,66	1 887

Stratum	Steps	R²	Estimation stock
Stations de ski de Maurienne, Val d'Arly, Beaufortin, les Bauges, Chablais, Giffre, Les Bornes et Chartreuse.	37	0,45	1 809
Stations de ski de Mont-Blanc et Aravis.	40	0,65	1 417
Stations de ski de l'Isère	29	0,57	1 149
Autres communes du Nord sauf Picardie, à revenu faible	24	0,28	1 758
Autres communes du Nord sauf Picardie, à revenu moyen	49	0,47	1 880
Autres communes du Nord sauf Picardie, à revenu élevé	34	0,37	1 514
Autres communes de l'Ouest sauf Pays de la Loire, à revenu faible	44	0,51	2 563
Autres communes de l'Ouest sauf Pays de la Loire, à revenu élevé	23	0,38	1 747
Autres communes du Limousin et d'Auvergne, à revenu faible	33	0,52	1 605
Autres communes du Limousin et d'Auvergne, à revenu élevé	35	0,44	1 620
Autres communes de Picardie, à revenu faible	44	0,62	2 029
Autres communes de Picardie, à revenu élevé	24	0,38	1 222
Autres communes du Centre, à revenu faible	40	0,59	1 999
Autres communes du Centre, à revenu élevé	34	0,49	1 546
Autres communes de Bourgogne, à revenu faible	36	0,50	1 407
Autres communes de Bourgogne, à revenu élevé	27	0,42	1 336
Autres communes de Lorraine, à revenu faible	38	0,44	1 768
Autres communes de Lorraine, à revenu élevé	27	0,28	1 713
Autres communes d'Alsace, à revenu faible	32	0,48	1 861
Autres communes d'Alsace, à revenu élevé	23	0,36	1 316
Autres communes de Franche-Comté, à revenu faible	51	0,49	1 578
Autres communes de Franche-Comté, à revenu élevé	28	0,33	1 341
Autres communes du Pays de la Loire, à revenu faible	36	0,40	1 696
Autres communes du Pays de la Loire, à revenu élevé	35	0,62	1 518
Autres communes du Midi-Pyrénées, à revenu faible	51	0,70	2 414
Autres communes du Midi-Pyrénées, à revenu élevé	26	0,43	1 655
Autres communes Du Rhône-Alpes-Est, à revenu faible	27	0,37	2 068
Autres communes Du Rhône-Alpes-Est, à revenu moyen	32	0,52	1 852
Autres communes Du Rhône-Alpes-Est, à revenu élevé	36	0,38	2 456
Autres communes Du Rhône-Alpes-Ouest, à revenu faible	41	0,57	2 210
Autres communes Du Rhône-Alpes-Ouest, à revenu élevé	45	0,53	2 368
Autres communes du Sud-Est (Languedoc-Roussillon et PACA), à revenu faible	34	0,38	2 133
Autres communes du Sud-Est (Languedoc-Roussillon et PACA), à revenu élevé	39	0,38	3 250
Autres communes de Provence-Alpes-Côte d'Azur Nord, à revenu faible	44	0,51	1 738
Autres communes de Provence-Alpes-Côte d'Azur Nord, à revenu moyen	34	0,60	1 598
Autres communes de Provence-Alpes-Côte d'Azur Nord, à revenu élevé	35	0,50	1 448
TOTAL			197 995

Houses

Table 5.10 - Quality of regressions and number of observations from estimation stock by stratum: Paris Region, houses

Stratum	Steps	R^2	Estimation stock
1	53	0.795	5,977
2	24	0.693	1,652
3	41	0.654	11,314
4	37	0.681	5,423
5	34	0.678	7,120
6	37	0.644	9,974
7	39	0.622	11,115
Total			52,575

Table 5.11 - Quality of regressions and number of observations from estimation stock by stratum: Provinces, houses

Stratum	Steps	R²	Estimation stock
Unité Urbaine d'Avignon	19	0,73	953
Unité Urbaine de Béthune	29	0,72	2 005
Unité Urbaine de Metz	23	0,66	1 254
Unité Urbaine de Douai-Lens	32	0,72	3 436

Stratum	Steps	R^2	Estimation stock
Unité Urbaine de Toulon	33	0,69	2 136
Unité Urbaine de Lyon	26	0,68	4 003
Banlieue de Marseille-Aix-en-Provence	35	0,72	2 067
Unité Urbaine de Nice	36	0,67	2 895
Marseille	24	0,78	910
Unité Urbaine de Dijon	16	0,74	892
Unité Urbaine de Brest	17	0,64	1 001
Toulouse	14	0,67	860
Banlieue de Toulouse	22	0,73	2 221
Bordeaux	25	0,78	1 087
Banlieue de Bordeaux	53	0,67	3 818
Unité Urbaine de Montpellier	13	0,71	701
Unité Urbaine de Rennes	19	0,70	1 059
Unité Urbaine de Tours	24	0,75	1 424
Unité Urbaine de Grenoble	19	0,72	933
Nantes	17	0,74	1 107
Unité Urbaine de Saint-Nazaire	18	0,67	835
Banlieue de Nantes	35	0,78	2 514
Unité Urbaine de Orléans	26	0,76	1 685
Unité Urbaine de Reims	13	0,76	715
Unité Urbaine de Nancy	26	0,67	1 954
Lille	18	0,77	825
Unité Urbaine de Maubeuge (partie française)	18	0,72	1 049
Tourcoing	12	0,73	1 011
Unité Urbaine de Dunkerque	21	0,69	760
Unité Urbaine de Valenciennes (partie française)	31	0,70	2 357
Banlieue de Lille (partie française)	75	0,85	6 702
Unité Urbaine de Calais	14	0,73	1 040
Unité Urbaine de Clermont-Ferrand	17	0,75	963
Le Mans	24	0,77	1 258
Le Havre	24	0,71	864
Unité Urbaine de Rouen	36	0,81	2 332
Amiens	20	0,80	1 196
Littoral du Finistère et des Côtes d'Armor, à revenu très faible	15	0,60	1 266
Littoral du Finistère et des Côtes d'Armor, à revenu faible	17	0,59	1 062
Littoral du Finistère et des Côtes d'Armor, à revenu moyen	13	0,59	981
Littoral du Finistère et des Côtes d'Armor, à revenu élevé	15	0,62	1 087
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu très faible	24	0,61	1 084
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu faible	17	0,63	1 063
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu moyen	18	0,59	919
Littoral d'Ille-et-Vilaine, et du Morbihan, à revenu élevé	14	0,57	1 073
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu très faible	26	0,68	1 297
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu faible	16	0,58	1 315
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu moyen	18	0,58	1 366
Littoral du Nord-Pas-de-Calais et des deux Normandie, à revenu élevé	19	0,66	1 215
Littoral Méditerranéen, à revenu faible	24	0,72	1 457
Littoral Méditerranéen, à revenu moyen	18	0,75	1 054
Littoral Méditerranéen, à revenu élevé	30	0,70	1 106
Littoral de Loire-Atlantique, Vendée et Charente-Maritime, à revenu faible	35	0,65	2 211
Littoral de Loire-Atlantique, Vendée et Charente-Maritime, à revenu élevé	23	0,67	1 742
Littoral de Gironde, des Landes et des Pyrénées-Atlantiques, à revenu faible	28	0,63	1 558
Littoral de Gironde, des Landes et des Pyrénées-Atlantiques, à revenu élevé	24	0,66	1 111
Stations de Ski d'Isère, Savoie et Haute-Savoie	31	0,68	1 664
Autres communes d'Indre et d'Indre-et-Loire, à revenu faible	18	0,70	859
Autres communes d'Indre et d'Indre-et-Loire, à revenu moyen	24	0,72	1 165
Autres communes d'Indre et d'Indre-et-Loire, à revenu élevé	22	0,70	1 247
Autres communes d'Eure-et-Loir, à revenu faible	35	0,66	1 544
Autres communes d'Eure-et-Loir, à revenu moyen	25	0,67	1 557
Autres communes d'Eure-et-Loir, à revenu élevé	30	0,66	1 639
Autres communes du Loiret, à revenu faible	23	0,68	1 369
Autres communes du Loiret, à revenu moyen	19	0,68	1 278
Autres communes du Loiret, à revenu élevé	22	0,68	1 410
Autres communes du Loir et Cher, à revenu faible	26	0,73	1 532
Autres communes du Loir et Cher, à revenu élevé	26	0,70	1 518
Autres communes du Cher, à revenu faible	20	0,74	995
Autres communes du Cher, à revenu élevé	34	0,78	1 042

Stratum	Steps	R^2	Estimation stock
Autres communes de Savoie et de Haute-Savoie, à revenu faible	30	0,71	1 636
Autres communes de Savoie et de Haute-Savoie, à revenu moyen	29	0,64	1 594
Autres communes de Savoie et de Haute-Savoie, à revenu élevé	29	0,61	1 529
Autres communes de Rhône-Alpes Ouest, à revenu très faible	26	0,69	1 905
Autres communes de Rhône-Alpes Ouest, à revenu faible	35	0,69	2 079
Autres communes de Rhône-Alpes Ouest, à revenu moyen	25	0,67	1 936
Autres communes de Rhône-Alpes Ouest, à revenu élevé	24	0,61	1 998
Autres communes de l'Ain, à revenu faible	25	0,65	2 028
Autres communes de l'Ain, à revenu élevé	29	0,70	2 315
Autres communes de l'Isère, à revenu faible	22	0,61	1 910
Autres communes de l'Isère, à revenu élevé	35	0,64	2 006
Autres communes de Champagne-Ardenne Nord, à revenu faible	27	0,69	1 377
Autres communes de Champagne-Ardenne Nord, à revenu moyen	40	0,64	1 558
Autres communes de Champagne-Ardenne Nord, à revenu élevé	28	0,68	1 583
Autres communes de Champagne-Ardenne Sud, à revenu faible	31	0,68	1 505
Autres communes de Champagne-Ardenne Sud, à revenu élevé	28	0,68	1 533
Autres communes de la Somme, à revenu faible	25	0,67	1 225
Autres communes de la Somme, à revenu moyen	20	0,70	1 459
Autres communes de la Somme, à revenu élevé	28	0,74	1 547
Autres communes de l'Aisne, à revenu très faible	28	0,74	1 564
Autres communes de l'Aisne, à revenu faible	33	0,73	1 716
Autres communes de l'Aisne, à revenu moyen	26	0,68	1 678
Autres communes de l'Aisne, à revenu élevé	33	0,74	1 794
Autres communes de l'Oise, à revenu faible	26	0,71	1 270
Autres communes de l'Oise, à revenu moyen	21	0,64	1 357
Autres communes de l'Oise, à revenu élevé	16	0,68	1 376
Autres communes de Seine-Maritime, à revenu faible	25	0,73	1 544
Autres communes de Seine-Maritime, à revenu moyen	23	0,71	1 811
Autres communes de Seine-Maritime, à revenu élevé	29	0,70	1 600
Autres communes de l'Eure, à revenu faible	35	0,65	1 856
Autres communes de l'Eure, à revenu moyen	26	0,64	1 698
Autres communes de l'Eure, à revenu élevé	17	0,66	1 779
Autres communes de Basse-Normandie, à revenu très faible	31	0,68	1 862
Autres communes de Basse-Normandie, à revenu faible	31	0,67	2 082
Autres communes de Basse-Normandie, à revenu moyen	46	0,73	2 166
Autres communes de Basse-Normandie, à revenu élevé	30	0,68	2 251
Autres communes de Bourgogne Est (21,71), à revenu faible	25	0,67	2 043
Autres communes de Bourgogne Est (21,71), à revenu moyen	24	0,61	2 116
Autres communes de Bourgogne Est (21,71), à revenu élevé	30	0,71	2 237
Autres communes de Bourgogne Ouest (89,58), à revenu faible	21	0,69	1 058
Autres communes de Bourgogne Ouest (89,58), à revenu moyen	26	0,62	1 257
Autres communes de Bourgogne Ouest (89,58), à revenu élevé	21	0,62	1 564
Autres communes du Pas-de-Calais, à revenu faible	28	0,67	1 776
Autres communes du Pas-de-Calais, à revenu moyen	28	0,70	1 617
Autres communes du Pas-de-Calais, à revenu élevé	28	0,72	1 932
Autres communes du Nord, à revenu très faible	25	0,73	1 595
Autres communes du Nord, à revenu faible	27	0,72	1 624
Autres communes du Nord, à revenu moyen	33	0,71	1 705
Autres communes du Nord, à revenu élevé	28	0,78	1 706
Autres communes de Lorraine Ouest (55,88), à revenu faible	28	0,68	1 807
Autres communes de Lorraine Ouest (55,88), à revenu élevé	30	0,64	1 836
Autres communes de Lorraine Est (54,57), à revenu faible	31	0,56	1 813
Autres communes de Lorraine Est (54,57), à revenu moyen	37	0,56	1 913
Autres communes de Lorraine Est (54,57), à revenu élevé	38	0,62	2 098
Autres communes d'Alsace, à revenu faible	24	0,61	1 174
Autres communes d'Alsace, à revenu moyen	32	0,62	1 316
Autres communes d'Alsace, à revenu élevé	25	0,69	1 436
Autres communes de Franche-Comté, à revenu faible	25	0,64	2 097
Autres communes de Franche-Comté, à revenu moyen	27	0,64	2 181
Autres communes de Franche-Comté, à revenu élevé	29	0,64	2 386
Autres communes du Pays de la Loire Nord, à revenu très faible	35	0,72	2 241
Autres communes du Pays de la Loire Nord, à revenu faible	28	0,71	2 275
Autres communes du Pays de la Loire Nord, à revenu moyen	47	0,76	2 502
Autres communes du Pays de la Loire Nord, à revenu élevé	41	0,76	2 427
Autres communes de Loire-Atlantique, à revenu faible	27	0,66	2 202
Autres communes de Loire-Atlantique, à revenu élevé	33	0,75	2 163

Stratum	Steps	R^2	Estimation stock
Autres communes de Vendée, à revenu faible	24	0,70	1 413
Autres communes de Vendée, à revenu élevé	28	0,71	1 525
Autres communes de Bretagne du Nord, à revenu faible	31	0,68	2 195
Autres communes de Bretagne du Nord, à revenu moyen	43	0,70	2 342
Autres communes de Bretagne du Nord, à revenu élevé	32	0,73	2 243
Autres communes de Bretagne du Sud, à revenu faible	32	0,71	2 364
Autres communes de Bretagne du Sud, à revenu élevé	30	0,69	2 739
Autres communes de Poitou-Charentes Sud, à revenu faible	33	0,67	2 389
Autres communes de Poitou-Charentes Sud, à revenu élevé	41	0,70	2 921
Autres communes de Poitou-Charentes Nord, à revenu faible	20	0,69	1 505
Autres communes de Poitou-Charentes Nord, à revenu moyen	30	0,70	1 654
Autres communes de Poitou-Charentes Nord, à revenu élevé	32	0,71	1 682
Autres communes d'Aquitaine du Nord-est, à revenu faible	20	0,66	2 016
Autres communes d'Aquitaine du Nord-est, à revenu élevé	28	0,67	1 923
Autres communes d'Aquitaine du Sud-ouest, à revenu faible	44	0,60	2 276
Autres communes d'Aquitaine du Sud-ouest, à revenu élevé	31	0,64	2 448
Autres communes de Gironde, à revenu faible	33	0,69	1 671
Autres communes de Gironde, à revenu élevé	29	0,65	1 971
Autres communes de Midi-Pyrénées du Nord, à revenu faible	14	0,69	1 113
Autres communes de Midi-Pyrénées du Nord, à revenu moyen	16	0,71	1 462
Autres communes de Midi-Pyrénées du Nord, à revenu élevé	23	0,67	1 731
Autres communes de Midi-Pyrénées du Sud, à revenu faible	23	0,71	1 470
Autres communes de Midi-Pyrénées du Sud, à revenu moyen	24	0,66	1 553
Autres communes de Midi-Pyrénées du Sud, à revenu élevé	32	0,71	1 805
Autres communes du Limousin, à revenu faible	22	0,73	1 894
Autres communes du Limousin, à revenu élevé	34	0,71	1 998
Autres communes d'Auvergne, à revenu faible	18	0,65	1 408
Autres communes d'Auvergne, à revenu moyen	25	0,64	1 738
Autres communes d'Auvergne, à revenu élevé	28	0,71	2 067
Autres communes du Languedoc-Roussillon, à revenu très faible	22	0,72	1 724
Autres communes du Languedoc-Roussillon, à revenu faible	16	0,74	1 530
Autres communes du Languedoc-Roussillon, à revenu moyen	23	0,72	1 909
Autres communes du Languedoc-Roussillon, à revenu élevé	29	0,70	1 717
Autres communes de Provence-Alpes-Côte d'Azur du Nord, à revenu faible	19	0,68	1 517
Autres communes de Provence-Alpes-Côte d'Azur du Nord, à revenu élevé	21	0,72	1 621
Autres communes de Provence-Alpes-Côte d'Azur du Sud, à revenu faible	20	0,74	1 354
Autres communes de Provence-Alpes-Côte d'Azur du Sud, à revenu moyen	14	0,76	1 368
Autres communes de Provence-Alpes-Côte d'Azur du Sud, à revenu élevé	19	0,71	1 322
TOTAL			293 410

Chapter 6: Seasonal adjustments to indices, monitoring and publication

The main used-housing price indices have been analysed over the long term and their good quality has been certified by the official "Notaires-INSEE" designation. Prior to certification, they are examined by a Scientific Board, formed under the terms of an agreement between the French notariat and INSEE.[48] In addition, the French public statistics authority (ASP) has approved the indices for the Paris Region produced by Paris Notaires Services (notice no. 2011-01 by the *Autorité de la statistique publique* on 21 June 2011 on the certification of quarterly statistics on housing prices in the Paris Region). Each of these indices is produced as a gross index and a seasonally adjusted index.

Supplementary indices covering the regions, *départements*, major cities or urban units may be certified as "Notaires-INSEE" indices provided they appear to be stable and rely on a sufficiently large dataset, in accordance with the criteria set out by the Scientific Board of the Notaires-INSEE indices.

Since November 2011, the old series have been replaced by new ones (based on regular updates of hedonic models, with base 100 in the first quarter of 2010).

The indices are monitored regularly by the Scientific Board of the Notaires-INSEE indices, via dashboards.

6.1 Seasonal adjustments to indices

6.1.1 Method for calculating corrections for seasonal adjustments

In version 3, the seasonal adjustment method selected is CENSUS X12. The method used in version 2 was CENSUS X11. With X12, estimates for seasonal adjustment coefficients are improved. In particular, by calculating moving averages the length of series can be shortened. With X12, series can be extended using Arima models to estimate the missing quarters or months. Method X12 is available in the SAS and Démétra software packages. It was the Démétra software that was selected for the seasonal adjustments in the Notaires-INSEE indices.

6.1.2 Taking corrections into account for trading days

The number of trading days that make up the quarters has no notable impact on determining prices. The price indices are not processed in any way to correct for the effect of trading days.

6.1.3 Nature of seasonally adjusted data

The indices are based on one quarter and are published every three months, covering a calendar quarter, for example, the first quarter of 2013 (January 2013, February 2013 and March 2013). "Quarter-on-quarter" indices can also be calculated, based on a quarter, but published every month (covering, for example, the three-month period: February 2013, March 2013 and April 2013). They correspond for a given month to the average of the data for the reference month and the two preceding months.[49] These calculations are made by the notaries. Seasonal adjustment was designed with this feature in mind: it is applied to quarter-on-quarter indices.

6.1.4 Frequency of calculating seasonal adjustment coefficients

The seasonal coefficients are revised once a year. They are updated in August, once the definitive data from the 4th quarter of the previous year has become available. An annual update of the seasonal adjustment seems preferable to a monthly or quarterly update. In this way, the number of revisions to the coefficients is limited, which is important since they may prove to be somewhat fragile, being estimated from non-definitive data. At the time of the annual revision of the seasonally adjusted coefficients, the Arima models used to extend the raw series are also updated.

The seasonally adjusted coefficients used to calculate indices later than the period of seasonal adjustment are the result of projections. These projections are made over two years.

[48] See the composition of the Scientific Board for the Notaires-INSEE indices in the agreements concerning the partnership between INSEE and the French Notariat (Appendices 7 et 8).
[49] Cf. Chapter 3.

6.1.5 Direct or indirect method

The first method is what is called the direct method, where each series is seasonally adjusted, whatever its aggregation level. The indirect method, on the other hand, consists of making seasonal adjustments to the primary series first then aggregating the seasonally adjusted series.

In theory, neither method is better than the other. However, the indirect method does have the advantage, when adapted specifically to the Notaires-INSEE indices, of ensuring coherence between changes in different aggregation levels. The direct method, however, can give contradictory results: it is possible that all the seasonally adjusted components of an aggregate (e.g. the *départements*) change in one direction while the seasonal adjustment of the aggregate (e.g. the region) changes in the other. This would be difficult to express.

We therefore perform seasonal adjustments using the indirect method. For the aggregation, the primary seasonally adjusted indices are weighted by the values of the reference stock at the baseline period q(t) used to chain indices over the reference period p(t).[50] As a result, the weightings selected are the same as those for the raw indices.

6.1.6 Calculation

To calculate seasonally adjusted coefficients, we can choose an additive or a multiplicative model. With the additive model, we suppose that the components of the series are independent one from the other. Thus the level of seasonal variation is independent of the level of the series. The multiplicative model, on the other hand, supposes that the components of the series are dependent one on the other. For housing prices, it clearly has to be the multiplicative method that is used in all cases because the level of seasonal variation increases and decreases with the level of the series.

Using X12, it is also possible to define the length of the moving average. If seasonality changes rapidly, in particular, it may be preferable to reduce the length of the moving averages.

In concrete terms, two sets of strata are used, the first for apartments and the second for houses. For each type of dwelling, the seasonal adjustments are first made at primary level. The seasonally adjusted indices for the non-primary geographic areas and data sets grouping apartments and houses together are obtained by aggregation.

For apartments, the primary geographic areas are the following:

- *départements* in the Paris Region (8 series);
- all city centres in the Provinces with more than 10,000 inhabitants;
- all suburban towns in the Provinces with more than 10,000 inhabitants;
- all towns in the Provinces with fewer than 10,000 inhabitants and rural municipalities in the Provinces;
- regions of Rhône-Alpes, Provence-Alpes-Côte d'Azur and Nord-Pas-de-Calais.[51]

For houses, the primary geographic areas are the following:

- *départements* in the Paris Region, excluding Paris (7 series);
- all of the Provinces;
- regions of Rhône-Alpes, Provence-Alpes-Côte d'Azur and Nord-Pas-de-Calais.[52]

The number of series to be seasonally adjusted is 28: 15 for the Paris Region and 13 for the Provinces.

[50] Cf. Chapter 3.
[51] Regional seasonally adjusted indices are not used when calculating seasonally adjusted indices relating to metropolitan France. They are used, however, to calculate seasonally adjusted indices for regional prices of all dwellings.
[52] Cf. note 51.

Table 6.1 – Seasonally adjusted "Notaires-INSEE" series

Apartments in Paris
Apartments in Seine-et-Marne
Houses in Seine-et-Marne
Apartments in Yvelines
Houses in Yvelines
Apartments in Essonne
Houses in Essonne
Apartments in Hauts-de-Seine
Houses in Hauts-de-Seine
Apartments in Seine-Saint-Denis
Houses in Seine-Saint-Denis
Apartments in Val de Marne
Houses in Val de Marne
Apartments in Val d'Oise
Houses in Val d'Oise
Apartments in city-centres of agglomerations of more than 10,000 inhabitants
Apartments in suburbs of agglomerations of more than 10,000 inhabitants
Apartments in agglomerations of fewer than 10,000 and rural areas
Houses in the Provinces
Apartments in Provence-Alpes-Côte d'Azur
Houses in Provence-Alpes-Côte d'Azur
Apartments in Marseille
Apartments in Rhône-Alpes
Houses in Rhône-Alpes
Apartments in Lyon
Apartments in Nord-Pas-de-Calais
Houses in Nord-Pas-de-Calais
Houses in Lille

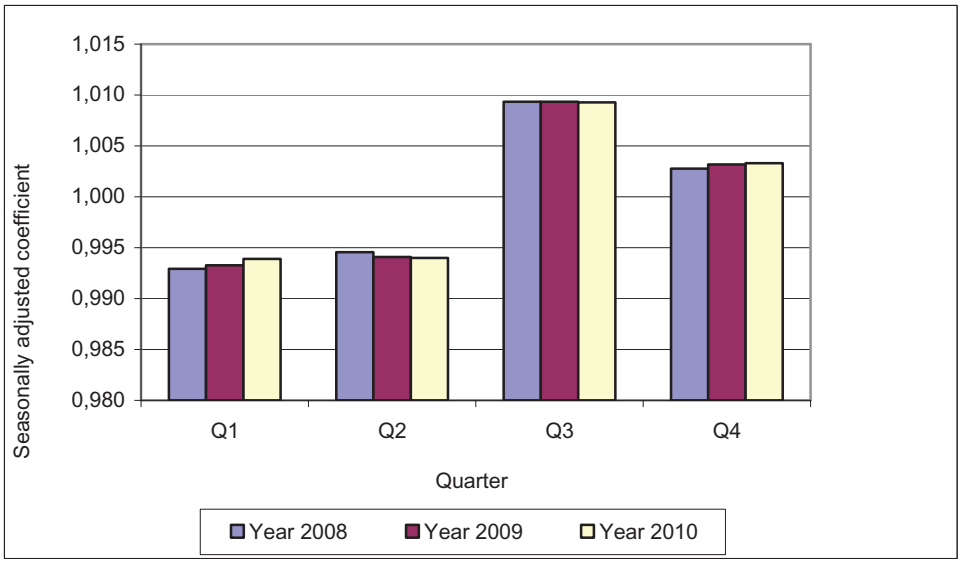
Table 6.2 - Correspondence between primary series and aggregates

Aggregated series	Corresponding primary series
Provinces	
Apartments in agglomerations of more than 10,000 inhabitants	- Apartments in city centres - Apartments in suburbs
Apartments	- Apartments in city centres - Apartments in suburbs - Apartments in rural areas
Apartments and houses	- Apartments in city centres - Apartments in suburbs - Apartments in rural areas - Houses
Apartments and houses: Rhône-Alpes	- Apartments Rhône-Alpes - Houses Rhône-Alpes
Apartments and houses: Provence-Alpes-Côte d'Azur	- Apartments Provence-Alpes-Côte d'Azur - Houses Provence-Alpes-Côte d'Azur
Apartments and houses: Nord-Pas-de-Calais	- Apartments Nord-Pas-de-Calais - Houses Nord-Pas-de-Calais
Paris Region	
Apartments	- Apartments by *département*
Apartments Paris Region excluding Paris	- Apartments by *département*
Apartments outer suburbs	- Apartments by *département* (nos. 77, 78, 91 and 95)
Apartments inner suburbs	- Apartments by *département* (nos. 92, 93 and 94)
Houses	- Houses by *département*
Houses outer suburbs	- Houses by *département* (nos. 77, 78, 91 and 95)
Houses inner suburbs	- Houses by *département* (nos. 92, 93 and 94)
Apartments and houses	- Apartments by *département* - Houses by *département*
Metropolitan France	
Apartments	- Apartments in city centres - Apartments in suburbs - Apartments in rural areas - Apartments in Paris Region by *département*
Houses	- Houses in Provinces - Houses in Paris Region by *département*
Apartments and houses	- Apartments in city centres - Apartments in suburbs - Apartments in rural areas - Houses in Provinces - Apartments in Paris Region by *département* - Houses in Paris Region by *département*

6.1.7 Results

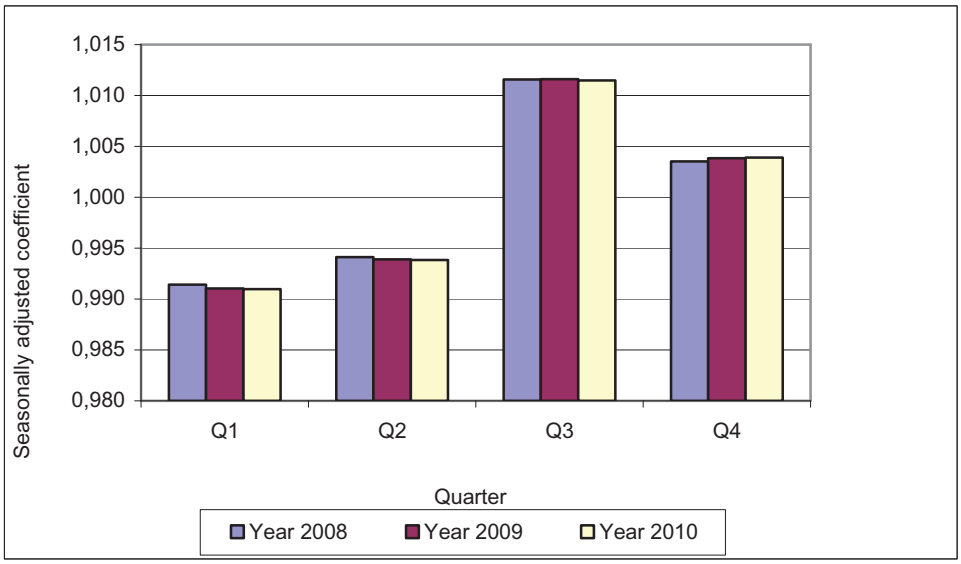
The prices of apartments, just like those of houses, experience seasonal fluctuations. All other things being equal, prices are slightly higher in the third quarter and this seasonal effect is more pronounced for houses than for apartments (Figures 6.1 and 6.2).

Figure 6.1: Seasonally adjusted coefficients of indices for apartments in all of metropolitan France



How to read this chart: apartment prices are highest in the third quarter. Seasonal adjustment changes little over time.

Figure 6.2: Seasonally adjusted coefficients of indices for houses in all of metropolitan France



How to read this chart: as for apartments, house prices are highest in the third quarter. Seasonal adjustment coefficients for houses are also fairly stable over time.

6.2 Dashboard

Each of the quarterly indices is available in the form of raw data and seasonally adjusted data, advance provisional, provisional, semi-final and final data.[53]

6.2.1 Objectives and composition of the dashboard

The dashboard is designed to comment on the indices, to uncover any possible anomalies and ultimately to validate publication. Because speed is of the essence, INSEE only validates the indices certified with the "Notaires-INSEE" designation.

The principle behind validation is based on the monitoring of the databases (volumes, coverage rates), analysing revisions and changes, and comparing changes in average prices and indices, to detect any structural effects and to find out their cause.

We analyse advance provisional indices and the provisional indices for the current quarter, semi-definitive and definitive indices from the previous quarter, also the corresponding volumes of transactions. The main figures to be checked are those published by INSEE in the *Informations rapides* collection and the macro-economic data-bank (BDM).

6.2.2 Examples of tables

To make it easier to identify suspect data, the dashboard brings together information on the levels of the indices, changes, revisions, contributions to overall changes, contributions to overall revisions, transaction volumes and coverage rates in the form of composite tables.

The contribution of a given area *d* to quarterly changes in supra-departmental aggregates is given by the following formula:

$$C_{d,t} = \frac{I_{A,t0/0}}{I_{A,t-1/0}} * \left(\frac{\delta_d \hat{W}_{d,p(t),q(t)}}{\sum_{e \in A} \delta_e \hat{W}_{e,p(t),q(t)}} \right) * \frac{I_{t/0}(d) - I_{t-1/0}(d)}{I_{t_0/0}(d)}$$

where:

t represents a monthly date after December 2007 which corresponds to year *a* within the set {*N, N+1*} where *N* is even,

t_0 the corresponding date in the same month (1st, 2nd or 3rd month of the quarter) of the last quarter of the year *N-1* where *N* is even,

$C_{d,t}$: contribution of area *d* to the change in aggregate *A* between *t-1* and *t*,

$I_{A,t/0}$: index of aggregate *A* on date *t* compared with reference quarter *0*,

$I_{A,t0/0}$: index of aggregate *A* on date t_0 compared with reference quarter *0*,

$I_{t/0}(d)$: index of area *d* (or part of area *d* if the aggregate does not include the entire area) on date *t* compared with reference quarter *0*,

$I_{t0/0}(d)$: index of area *d* (or part of area *d* if the aggregate does not include the entire area) on date t_0 compared with reference quarter *0*,

$\hat{W}_{d,p(t),q(t)}$: value of properties in relation to area *d* (or part of area *d* if the aggregate does not include the entire area) traded during the period *p(t)* estimated on date *q(t)*, *p(t)* corresponding to years *N-3* and *N-2* where *N* is even and *q(t)* is the last quarter of year *N-2* where *N* is even,

[53] Cf. Appendix 11.

δ_d , δ_e coefficients applied to take into account the non-exhaustivity of the notaries' database in the course of year *N-2* where *N* is even.

$\hat{W}_{e,p(t),q(t)}$: value of properties in area *e* (or part of area *e* if the aggregate does not include the entire area) purchased during the period *p(t)* estimated on date *q(t)*, *p(t)* corresponding to years *N-3* and *N-2* where *N* is even and *q(t)* is the last quarter of year *N-2* where *N* is even.

Sets of tables are produced to monitor changes and to monitor revisions, at an aggregate level and at a fine level. From the different contributions made to changes and revisions we can quickly identify the detailed series that are giving rise to marked changes or revisions in certain aggregates (*Table 6.3*).

The summary tables also include the following information:

- quarterly changes;
- contributions to quarterly changes overall;
- revisions;
- contributions to revisions overall.

Table 6.3 – Summary of changes relating to apartments in the Provinces (example shows 3rd quarter 2012)

Area	Weighting (%)	Seasonally adjusted indices		Quarterly variations in seasonal adjustment		Year-on-year changes in seasonal adjustment	Volumes		Change in volumes		Coverage rate	
	Q	Q-1	Q	Q-1/Q-2	Q/Q-1	Q/Q-4	Q-1	Q	Q/Q-1	Q/Q-4	Q-1	Q
France		114.0	113.8	-0.4%	-0.2%	-0.7%					50.3%	42.0%
Provinces	100.0%	107.2	106.6	-0.5%	-0.6%	-1.3%	22,936	15,480	-32.5%	-50.2%	44.5%	34.8%
Rural	9.7%	104.1	104.3	-1.6%	0.2%	-2.3%	2,810	1,871	-33.4%	-46.1%		
City centre	59.7%	107.7	107.2	-0.5%	-0.4%	-1.2%	13,547	9,233	-31.8%	-51.1%		
Suburbs	30.6%	107.6	106.3	-0.2%	-1.3%	-1.3%	6,579	4,376	-33.5%	-50.1%		
31 Nord-Pas de Calais	3.3%	111.4	112.0	1.2%	0.5%	2.5%	734	534	-27.2%	-50.6%	45.3%	41.1%
82 Rhône-Alpes	22.4%	110.5	110.0	0.3%	-0.5%	-0.2%	5,253	3,850	-26.7%	-45.3%	51.3%	40.8%
69123 Lyon	5.1%	118.9	118.1	0.6%	-0.6%	1.0%	769	715	-7.0%	-44.2%		
93 PACA	27.4%	105.6	105.1	-1.8%	-0.5%	-2.5%	5,247	3,020	-42.4%	-55.8%	51.2%	34.0%
13055 Marseille	4.3%	105.5	105.4	-1.2%	-0.1%	-3.7%	798	577	-27.7%	-48.7%		

How to read this chart: The provisional price index for apartments in all of metropolitan France is 113.8 in the 3rd quarter 2012 (seasonally adjusted data). This index is lower in the Provinces (106.6 points). Between the 2nd and 3rd quarters 2012, there was a downturn in prices. The decline was a little more pronounced in the Provinces that nationally. Prices also fell over the year (-0.7% for all of metropolitan France).
PACA= Provence-Alpes-Côte d'Azur

Coverage rates are estimated for France and the *départements* (cf. Chapter 4).

Table 6.4 – Greatest provisional changes in absolute value (at stratum level)

stratum	Name of stratum	Weighting (%)	Index Q-1	Index Q	quarterly variation at date T *	Volumes Q	Contribution to quarterly variation in Provinces at date T
54701	Suburb of Nancy	0.4%	101.5	112.7	11.1%	76	0.000404015
97009	Coastal towns PACA West (*dép.* 13, 83)	1.0%	102.3	113.2	10.7%	53	0.001073172
97004	Coastal towns Bretagne West (*dép.* 22, 29)	0.7%	102.8	113.2	10.2%	131	0.000672242
99521	Other municipalities Pays de la Loire, low income	0.7%	97.4	105.3	8.2%	110	0.00056283
99222	Other municipalities Picardie, high income	0.4%	96.8	104.2	7.6%	74	0.000326289
99082	Other municipalities Limousin and Auvergne, high income	0.6%	101.1	108.2	7.1%	78	0.000395908
99431	Other municipalities Franche-Comté, low income	0.5%	102.5	109.4	6.7%	91	0.000342599
74601	Suburb of Annecy	0.7%	113.6	120.4	6.0%	94	0.000423566
97010	Coastal towns PACA East (*dép* 83, 06)	1.8%	102.5	108.7	6.0%	123	0.001052136
651	Genève-Annemasse agglomeration (French part)	1.0%	120.2	126.6	5.3%	141	0.000530451
67701	Suburb of Strasbourg (French part)	0.7%	108.1	113.2	4.7%	69	0.000305825
33063	BORDEAUX	1.4%	118.0	123.2	4.4%	237	0.000589698
59350	LILLE	1.2%	116.5	121.5	4.3%	154	0.000508159
99911	Other municipalities South-East (Languedoc-Roussillon and PACA), low income	1.3%	105.3	109.8	4.3%	230	0.000531111
29019	BREST	0.7%	104.4	108.5	3.9%	138	0.000262074
97008	Coastal towns Languedoc-Roussillon South (*dép.* 34, 30)	0.9%	108.8	113.0	3.9%	189	0.000349547
99933	Other municipalities PACA North, high income	0.7%	104.4	108.5	3.9%	54	0.000251746
99412	Other municipalities Lorraine, high income	0.5%	99.5	103.4	3.9%	125	0.000195157
64445	PAU	0.5%	95.0	98.7	3.8%	67	0.00017471
35238	RENNES	1.2%	112.4	116.7	3.8%	224	0.000461684
…	…	…	…	…	…	…	…

How to read this chart: The greatest variation between quarter Q-1 and quarter Q is found in the Nancy suburbs (+11.1%). The change in this series makes only a small contribution to the change in the index overall (0.04 percentage points).
PACA= Provence-Alpes-Côte d'Azur

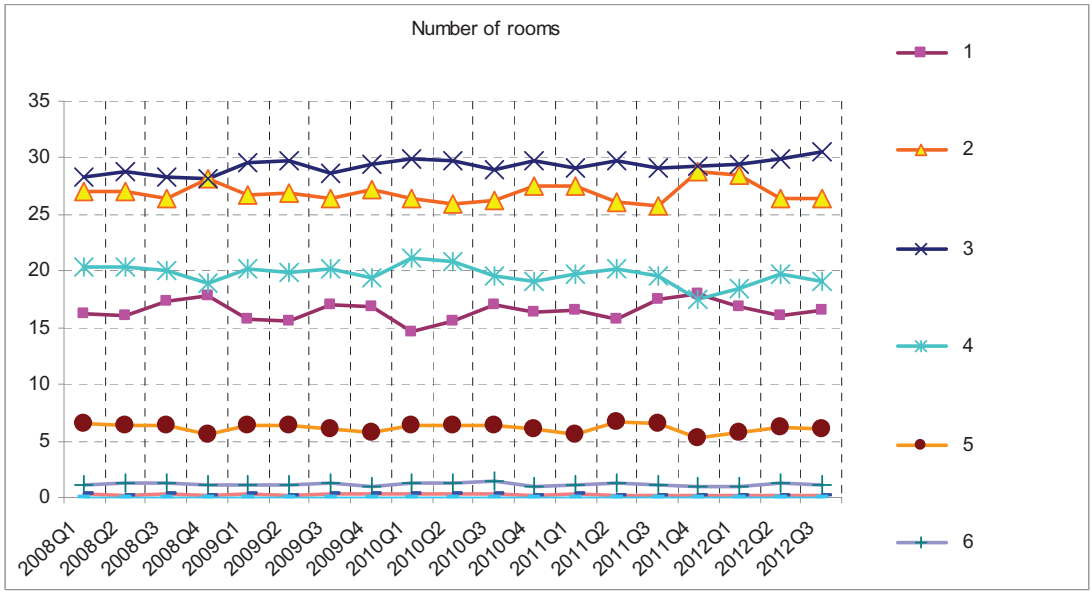
These data are classified in decreasing order of absolute value, to make it easier to see any errors. For example, in the suburbs of Nancy, prices increase by 11.1% compared with the previous quarter. Because of the size of this increase, a check was carried out based on the early provisional indices and the corresponding volumes.

In addition to items for comparison, the dashboard also contains information on the indices by theme, relating to changes, revisions, average prices, volumes, coverage rates and weightings.

The dashboard also contains a page of graphs in order to visualise the Notaires-INSEE indices, showing average price indices, volumes of transactions, coverage rates, etc.

The dashboard can also show changes in the distribution of modalities of the variables used for price estimation. The aim is to detect any possible changes or systematic errors in coding these modalities (see for example Figure 6.3, showing the variable “number of rooms”).

Figure 6.3 – Change in structure of transactions by number of rooms for apartments in the Provinces



How to read this chart: the structure of transactions by number of rooms appears to be relatively stable over time.

6.3 Publication

Since July 2004, the Notaires-INSEE indices have been published four times a year by INSEE in *Informations Rapides*. They can also be found in the INSEE macro-economic databank (BDM) and they are posted on the INSEE website, http://www.insee.fr.

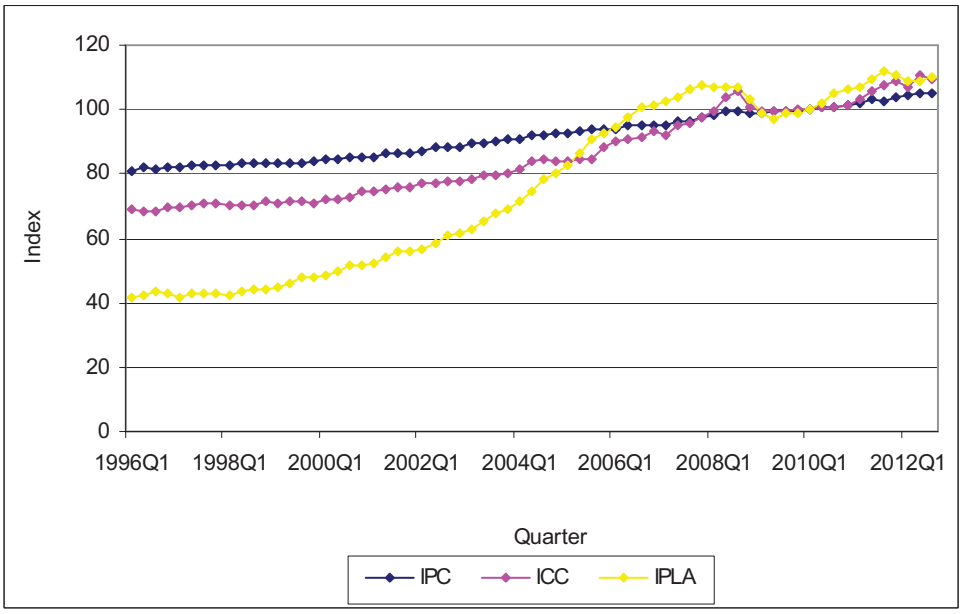
The list of published indices grew in 2011 when the Nord-Pas de Calais indices were published, after tests were carried out and they proved to show good coverage of the region in the notarial databases in previous years.

The indices are also disseminated by notaries on their own websites, http://www.paris.notaires.fr for the Paris Region and http://www.immoprix.com for the Provinces.

Typically, the indices for a given quarter are published about two months after the end of the quarter. Details of the dissemination criteria and the type of indices published are given in Appendix 12.

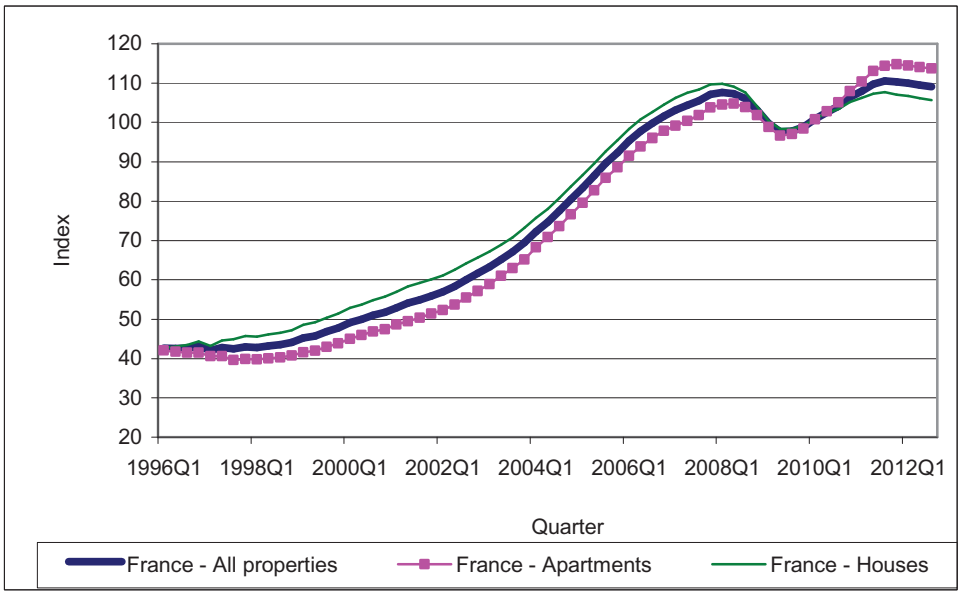
The charts below show changes in the main Notaires-INSEE indices since 1996 (when series were first published for metropolitan France).

Figure 6.4 - Used housing price index, consumer price index and cost-of-construction index (France, all property types)



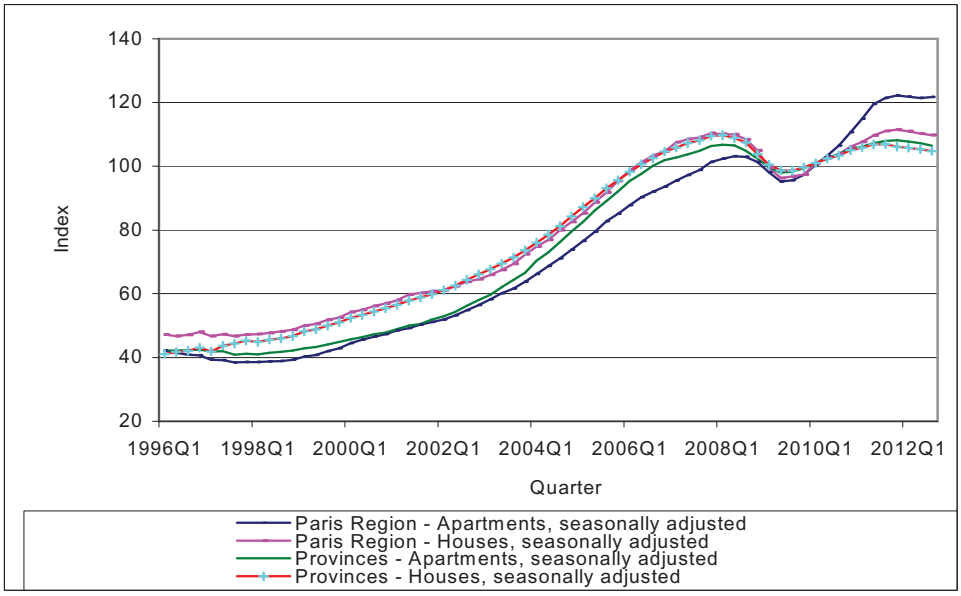
How to read this chart: Between 1998 and 2011, used-housing prices increased much more rapidly than construction costs, which had in turn increased much more markedly than consumer prices.
IPLA: used-housing price index
IPC: consumer price index
ICC: cost-of-construction index
Base 100=2010 Q1

Figure 6.5 - Used-housing price indices by property type (apartments or houses) adjusted for seasonal variations



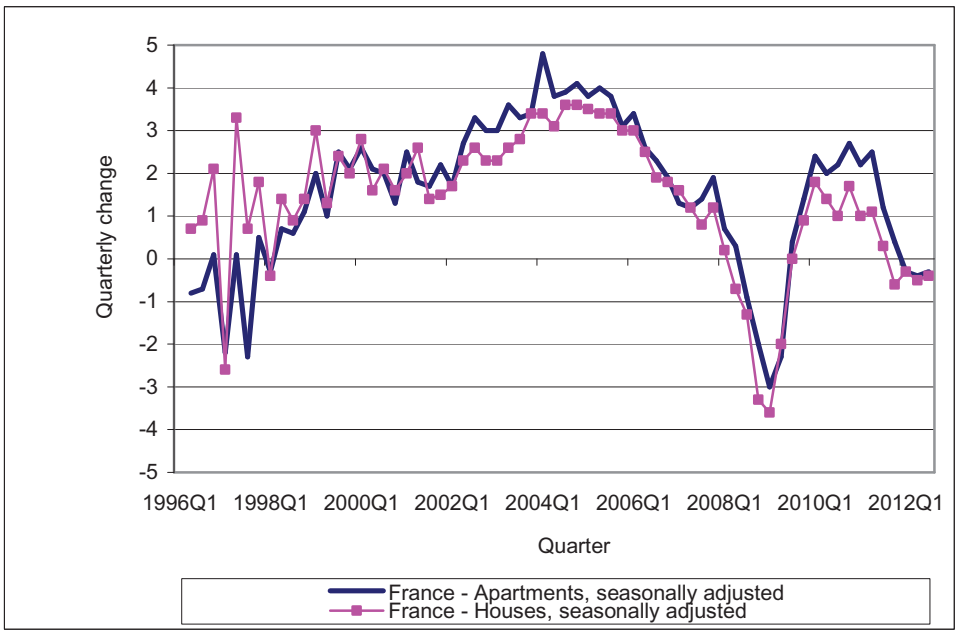
How to read this chart: between 1998 and 2011, apartment prices increased more rapidly than house prices.
Base 100=2010 Q1

Figure 6.6 - Used-housing price indices in Paris Region and the Provinces by property type (apartments or houses) adjusted for seasonal variations



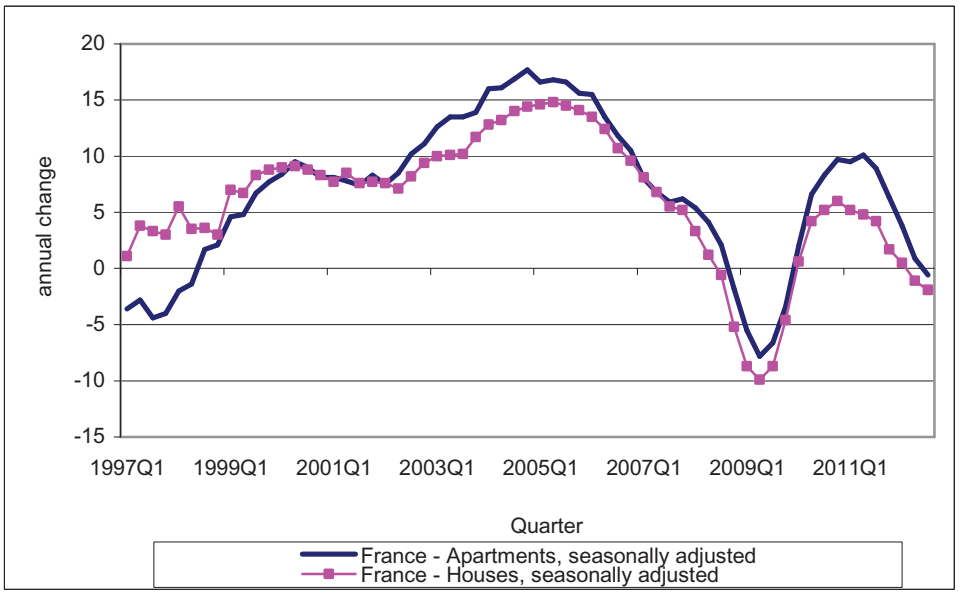
How to read this chart: since 1998, prices of apartments in the Paris Region have increased most.
Base 100=2010 Q1

Figure 6.7 - Quarterly change in used-housing price indices by property type (apartments or houses) adjusted for seasonal variations



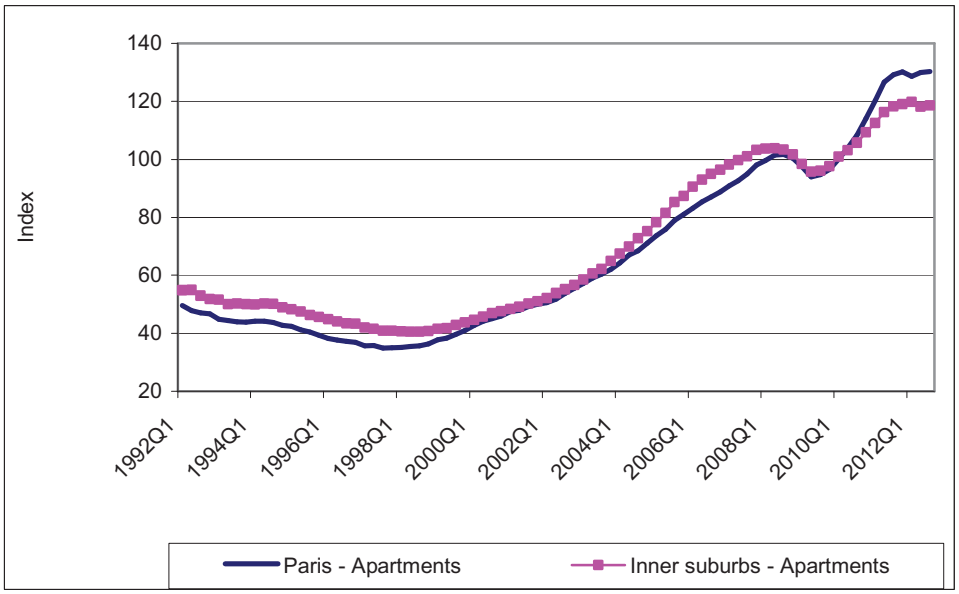
How to read this chart: since 2001, in most cases, quarterly changes in apartment prices have exceeded those of house prices.
Base 100=2010 Q1

Figure 6.8 - Annual change in used-housing price indices by property type (apartments or houses)



How to read this chart: since 2001, annual changes in apartment prices have been greater than changes in house prices.
Base 100=2010 Q1

Figure 6.9 - Used-housing price indices for apartments in Paris and the inner suburbs, adjusted for seasonal variations



How to read this chart: between 1998 and 2012, the prices of apartments in Paris increased more quickly than those of apartments in the inner suburbs. Between 2009 and 2011 in particular, the rebound was much more pronounced in Paris than in the inner suburbs.
Base 100=2010 Q1

Chapter 7: Comparison of V2 and V3

The methods used to calculate the indices in version 2 are described in Beauvois *et al*. (2005). In this chapter, we compare the indices calculated in version 2 with those in version 3. To do this, we make a breakdown of the different changes that have happened to assess their respective impacts. These changes concern, first, the introduction of new strata (zoning patterns), and second, a number of changes that will be grouped together under the heading of a new method of calculation. These are: new specifications for the models, more regular updating of corrective coefficients, chain-linking indices and moving to aggregation by geometric mean.

Indices for apartments are not very much affected by the new zoning and the change of method. In contrast, the introduction of newly defined areas in the Provinces and of chaining have a significant impact on price indices for houses.

To recalculate indices later than 2007 and revise indices for houses built prior to 2008, we used more comprehensive data than that used originally. This had the effect of accentuating changes in the index in periods of rising prices.

An error was corrected in the coding of the number of garages in houses in the Provinces. This had a significant effect, but was fortunately concentrated in 2010 and 2011.

7.1 Different sources of divergence

For apartments, the new method was applied from 2008 onwards, and indices prior to 2008 were linked to the new indices. For houses, indices were also totally recalculated, as there were new specifications relating to living space and surface areas of plots (cf. Appendix 3). Seasonal adjustments were recalculated, with method X11 being replaced by method X12-Arima.

.

There are four sources of divergence from the indices in version 2:

- "super-definitive" data is taken into account when recalculating the indices: deeds continue to be recorded after the indices are published; "super-definitive" data are data that are available today and which therefore include deeds that did not contribute to the calculation of the index when it was published;
- using a new zoning system: in order to take into account the distortions over time of relative prices of dwelling characteristics, the geographic boundaries of the strata were reviewed. The use of hedonic methods is based on the hypothesis that price changes are homogeneous within the strata;
- applying a new method of calculation: the specification for the econometric models was refined. In addition, in order to better take account of market changes, it was decided to update the coefficients used to calculate the indices more often. This change involved introducing an index chaining procedure. The method used to aggregate the primary indices was also reviewed. A geometric method was selected for use at *département* level. A new method of seasonal adjustment was also introduced.
- correction of an error: the switch to version 3 of the Notaires-INSEE indices coincided with the correction of an error in indices from 2009 to 2011. This error was due to a break in the coding of data from the Provinces concerning one variable, an incident that happened after the introduction of new coding software.

We first look at the overall divergence between versions 2 and 3 (7.2), then concentrate on the effect of each source of divergence.

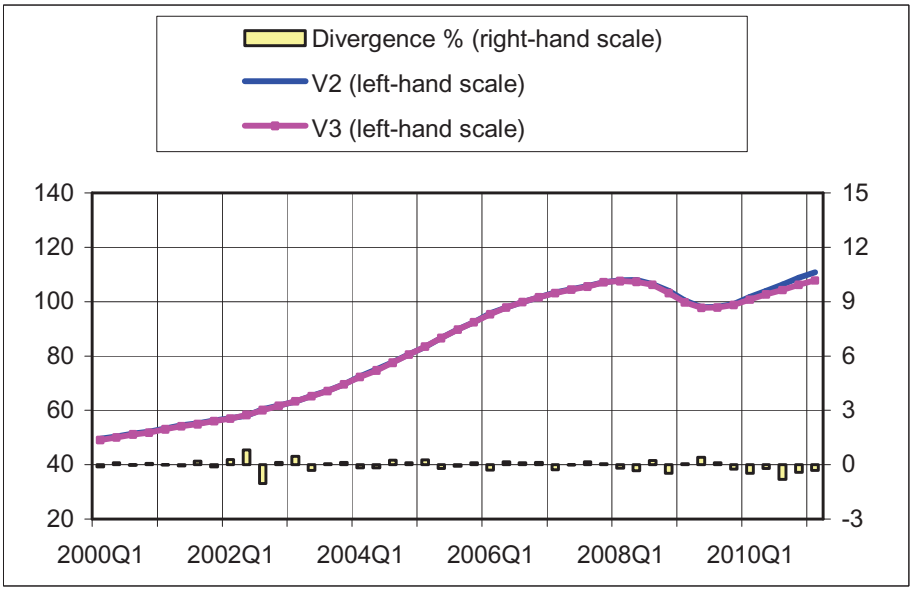
7.2 Overall divergence between versions 2 and 3

The switch from version 2 to version 3 had a significant impact on quarterly changes in prices relating to the years 2010 and 2011. Price changes between the fourth quarter of 2009 and the fourth quarter of 2010 were

therefore revised downwards by 2.0 points. Conversely, price changes prior to 2008 were revised only slightly (Figure 7.1),[54] if we exclude the second and third quarters of 2002.

Warning : until the end of this chapter, the differences are calculated as the difference (in% points) between the changes compared to the previous quarter, obtained with each version.

Figure 7.1 - Seasonally adjusted used-housing price indices for metropolitan France in versions 2 and 3, metropolitan France, base 100=2010Q1



Concerning apartments, only very minor revisions were necessary over the entire period (Figure 7.2). Revisions to pre-2008 prices are simply due to the application of a new method of applying seasonal adjustments (using X12-Arima instead of X11). From 2008 onwards, apart from the new method of making seasonal adjustments, revisions are associated with the use of super-definitive data, the introduction of a new zoning system and the introduction of a new method of calculation (new model, regular updating of coefficients, geometric mean and the use of chain-linking).

Revisions concerning houses were minor for years prior to 2010, if we exclude the second and third quarters of 2002. However, considerable revisions were necessary for 2010 and 2011 (Figure 7.3). For these two years, apart from the use of the new method for calculating the indices, a new way of adjusting for seasonal variations and the introduction of super-definitive data, revisions were linked with correcting the error mentioned above. Changes in house prices between the end of 2009 and the end of 2010 were therefore revised downwards by 3.7 points.

[54] In order to facilitate comparisons between the version 2 and version 3 indices, Figures 7.1 to 7.9 show "recalculated" indices from version 2. A coefficient multiplier was applied to each series of indices in version 2 so that indices for the 4th quarter of 2007 would be equal in both versions. The 4th quarter of 2007 was chosen to give a better understanding of the differences due to recalculating the version 2 indices and those due to the transition to version 3.

Figure 7.2 - Seasonally adjusted used-housing price indices for apartments in metropolitan France in versions 2 and 3, base 100=2010Q1

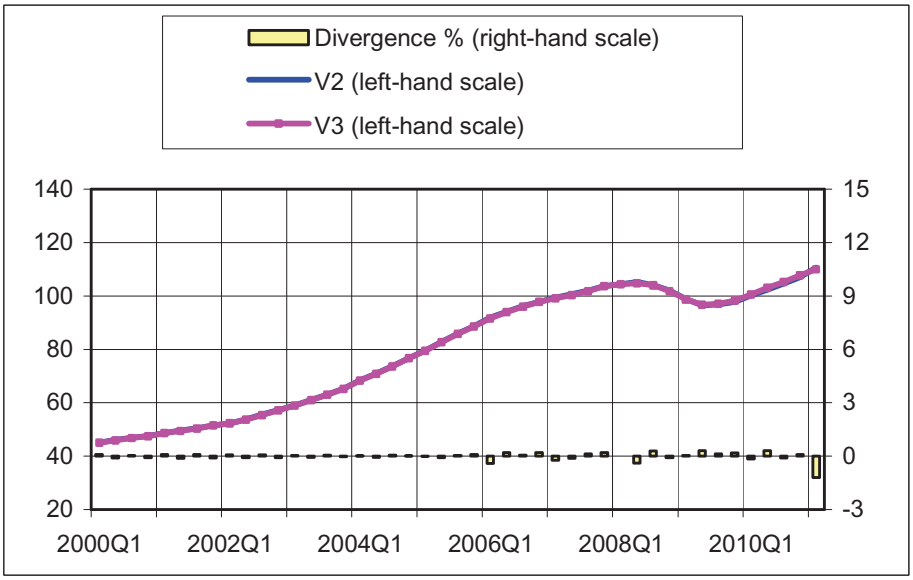
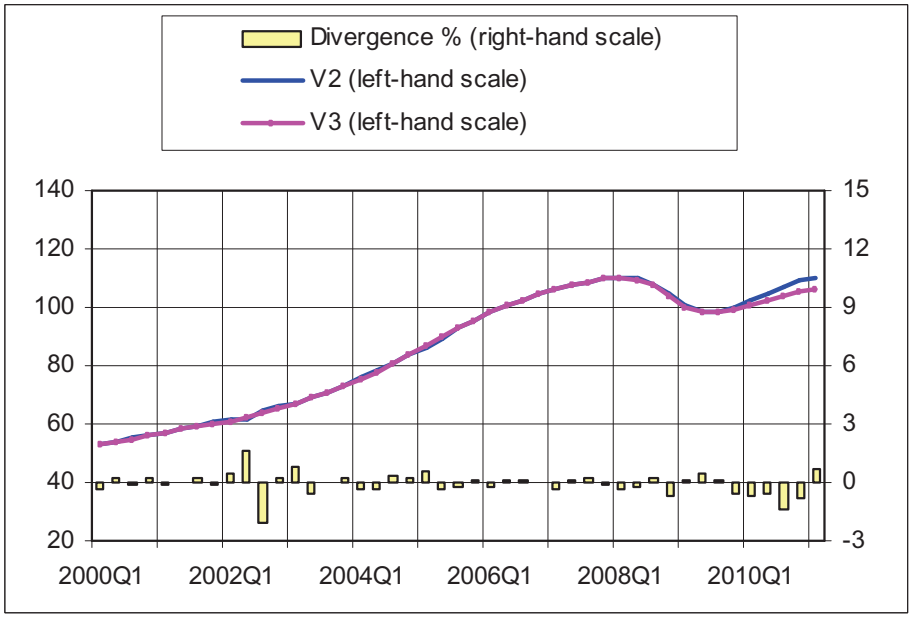


Figure 7.3 - Seasonally adjusted used-housing price indices for houses in metropolitan France in versions 2 and 3, base 100=2010Q1



Changes concerning the Paris Region required little revision (Figure 7.4). Revisions relating to apartments were also small (Figure 7.5). Those concerning houses were a little more pronounced at the beginning of the 2000s, due mainly to a change in the specification for the models. Living space and plot size have now been better taken into account (Figure 7.6).

Figure 7.4 - Seasonally adjusted used-housing price indices for the Paris Region in versions 2 and 3, base 100=2010Q1

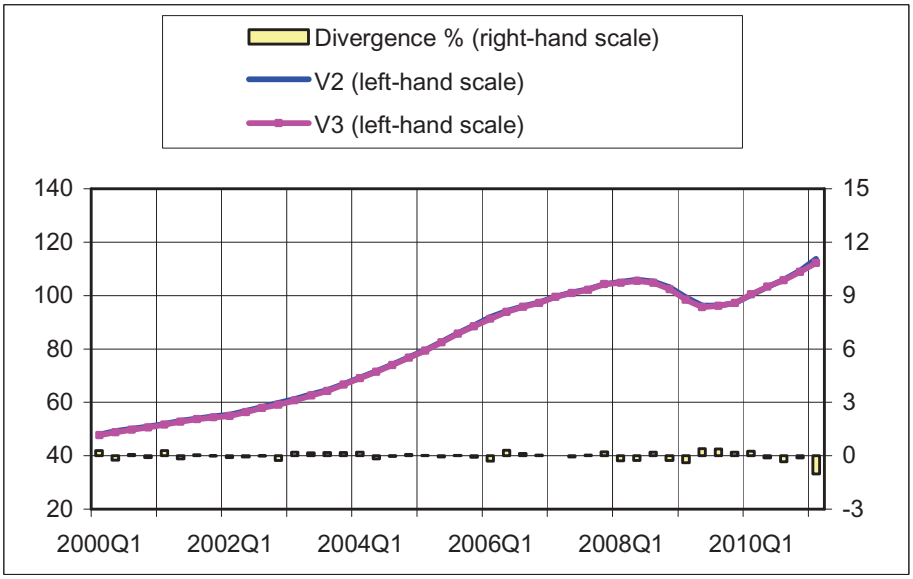


Figure 7.5 - Seasonally adjusted used-housing price indices for apartments in the Paris Region in versions 2 and 3, base 100=2010Q1

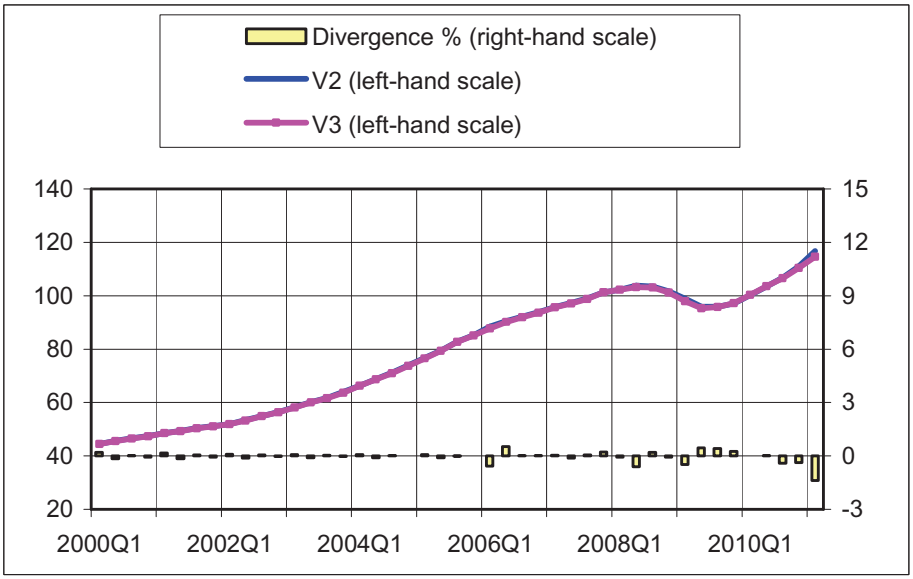
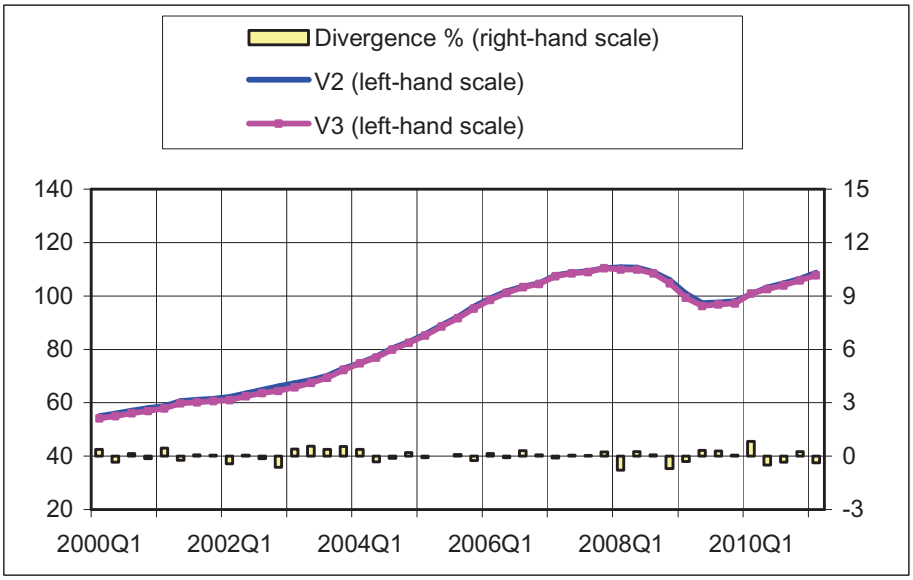


Figure 7.6 - Seasonally adjusted used-housing price indices for houses in the Paris Region in versions 2 and 3, base 100=2010Q1



Price changes in the Provinces were revised considerably for 2010 and 2011 (Figure 7.7), as a result of correcting the error in the houses index (Figure 7.9). However, only moderate revisions were necessary in the index for apartments in the provinces (Figure 7.8).

Figure 7.7 - Seasonally adjusted used-housing price indices for the Provinces in versions 2 and 3, base 100=2010Q1

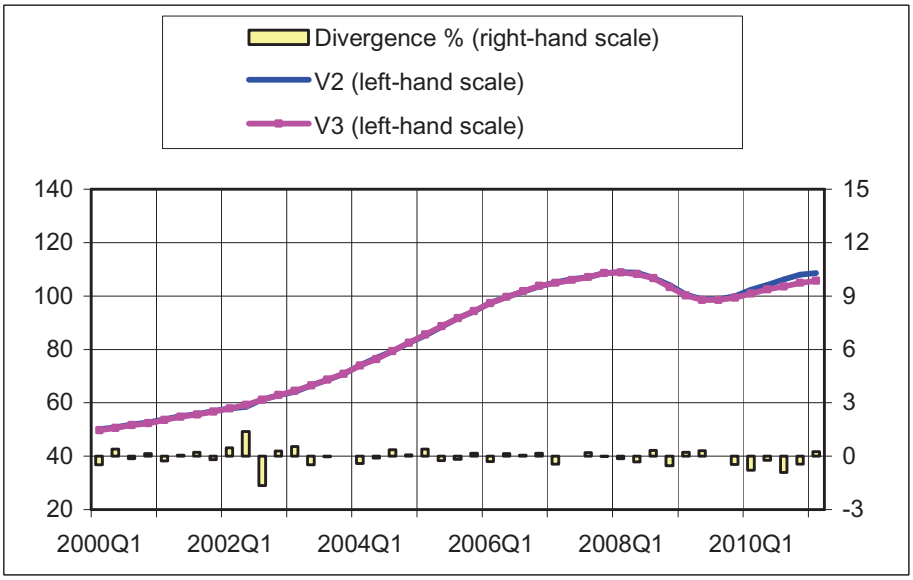


Figure 7.8 - Seasonally adjusted used-housing price indices for apartments in the Provinces in versions 2 and 3, base 100=2010Q1

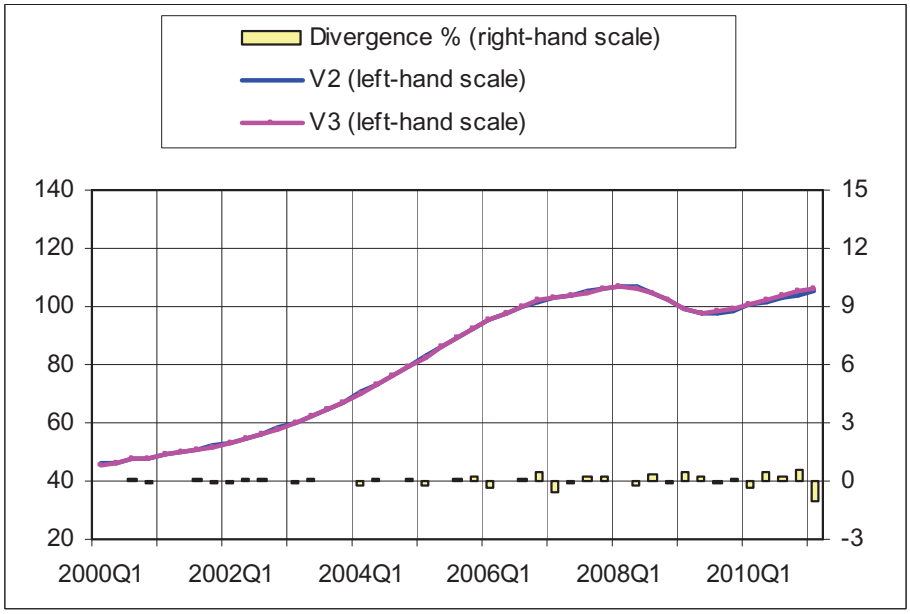
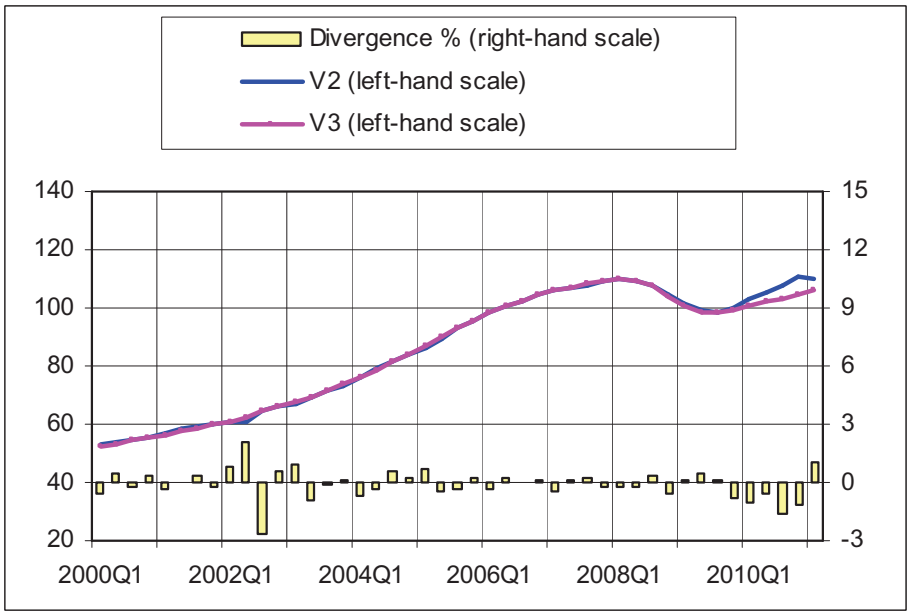


Figure 7.9 - Seasonally adjusted used-housing price indices for houses in the Provinces in versions 2 and 3, base 100=2010Q1



7.3 Analysis of divergence between versions 2 and 3

7.3.1 Divergence due to using super-definitive data

For apartments, divergence between definitive and super-definitive indices impacted on V3 only from 2008 onwards. In fact, V3 indices prior to 2008 are based on definitive V2 data (the only data available when the indices were calculated). For houses, on the other hand, V3 indices prior to 2008, like the V3 indices later than 2007, were recalculated from super-definitive data.

For the Provinces, divergence between definitive and super-definitive indices appears small (Figures 7.10 and 7.11). The average deviation from 2002 to 2009 is 0.1 of a point for apartments and 0.6 for houses. The reason

for this upward impact after the move to super-definitive data lies in the fact that prices rose. As any new deeds being considered were mainly signed towards the end of the quarter, they were at much higher prices than the average quarterly prices calculated from definitive data. The considerable difference between variations from the second to third quarters of 2002 in house prices in the Provinces was due to a great extent to the fact that new data were included in the second quarter 2002 (as a result of problems with data collection at the time: under-estimation in one quarter, catch-up in the next).

Figure 7.10 - Definitive and super-definitive raw price indices for used apartments in the Provinces in version 2, base 100=2000Q4

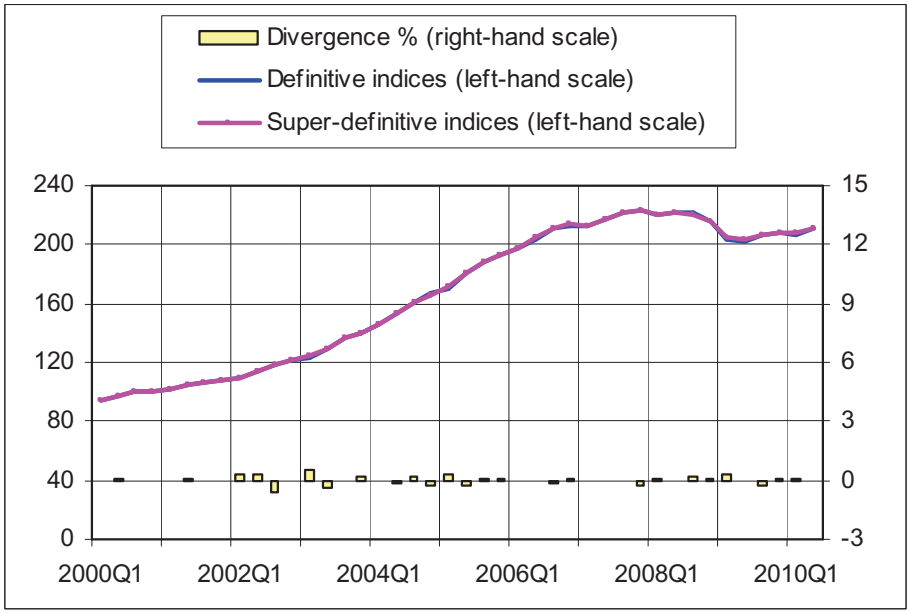
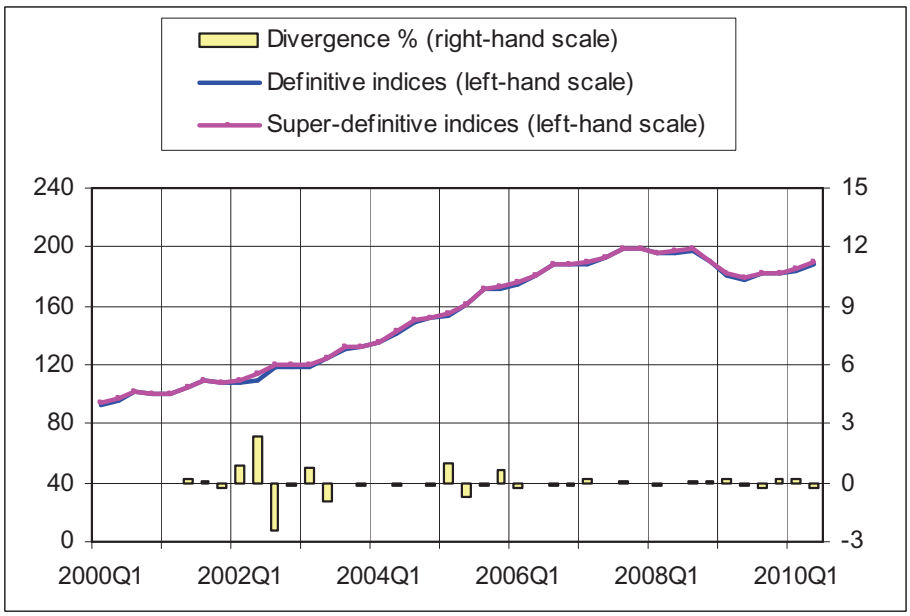


Figure 7.11 - Definitive and super-definitive raw price indices for used houses in the provinces in version 2, base 100=2000Q4



7.3.2 Impact of updating the zoning system

For both apartments and houses, updating the zoning had an impact on V3 indices from 2008 onwards. V3 indices prior to 2008 were based on the zoning in place in V2.

For the Provinces, the impact of updating the zoning system is slightly more marked than the effect of changing from definitive to super-definitive data. Between 2002 and 2009, the average divergence is 0.3 points for apartments and 1.4 for houses (Figures 7.12 and 7.13). In the case of houses in the Provinces in particular, the number of strata increased in version 3.

Figure 7.12 - Super-definitive raw price indices for used apartments in the Provinces in version 2 with version 2 and version 3 zoning, base 100=2000Q4

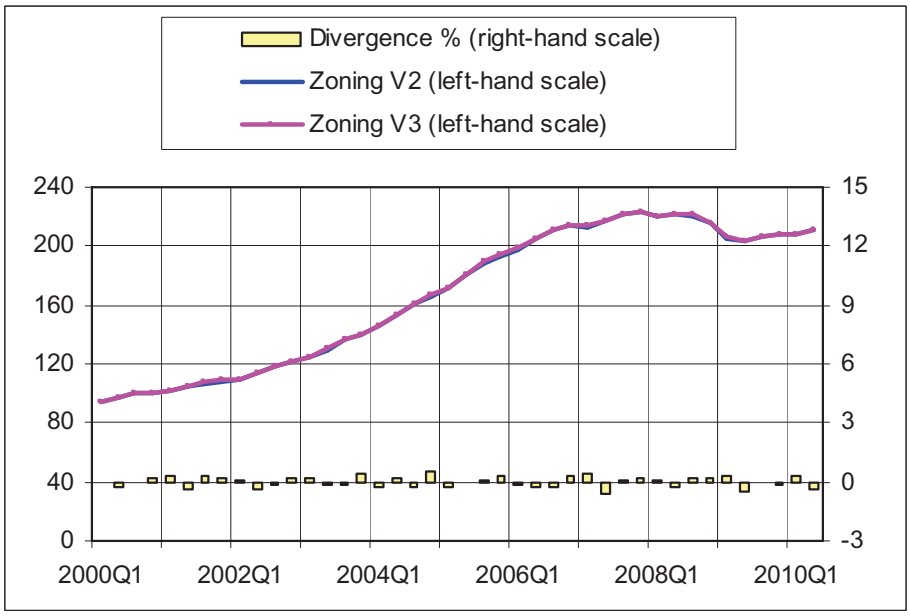
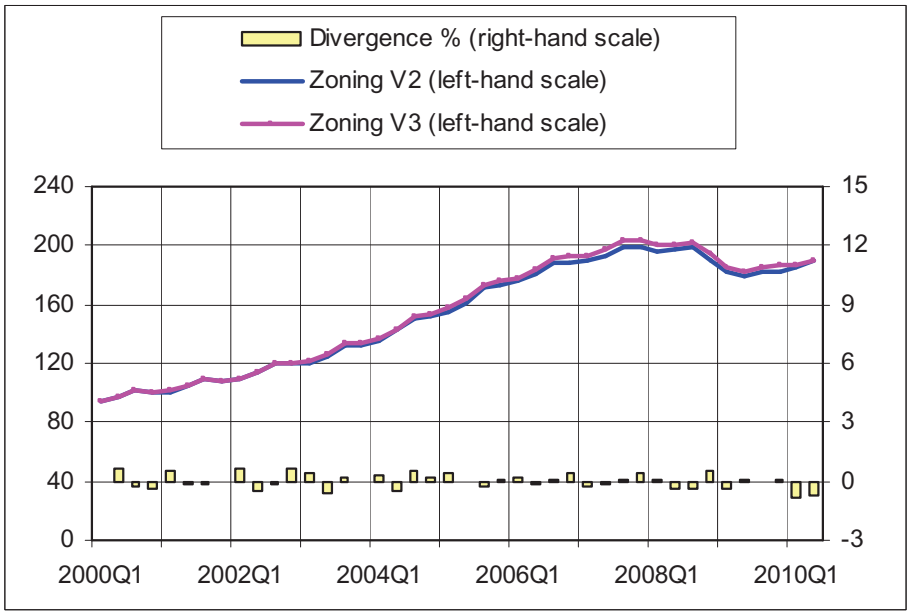


Figure 7.13 - Super-definitive raw price indices for used houses in the Provinces in version 2 with version 2 and version 3 zoning an, base 100=2000Q4



7.3.3 Impact of the new calculation method

With the new calculation method, revisions were needed in the V3 indices, mainly from 2008 onwards. For indices prior to 2008, only the specifications regarding living space and plot size for houses were changed.

Changing from arithmetic mean to geometric mean had little impact on the results, either for apartments or houses in the Provinces (Figures 7.14 and 7.15). For apartments, all the aggregations were done arithmetically. For houses, data at *département* level were aggregated geometrically while supra-departmental data were aggregated arithmetically.

Figure 7.14 - Super-definitive raw price indices for used apartments in the Provinces in version 2 with arithmetic and geometric aggregation , base 100=2000Q4

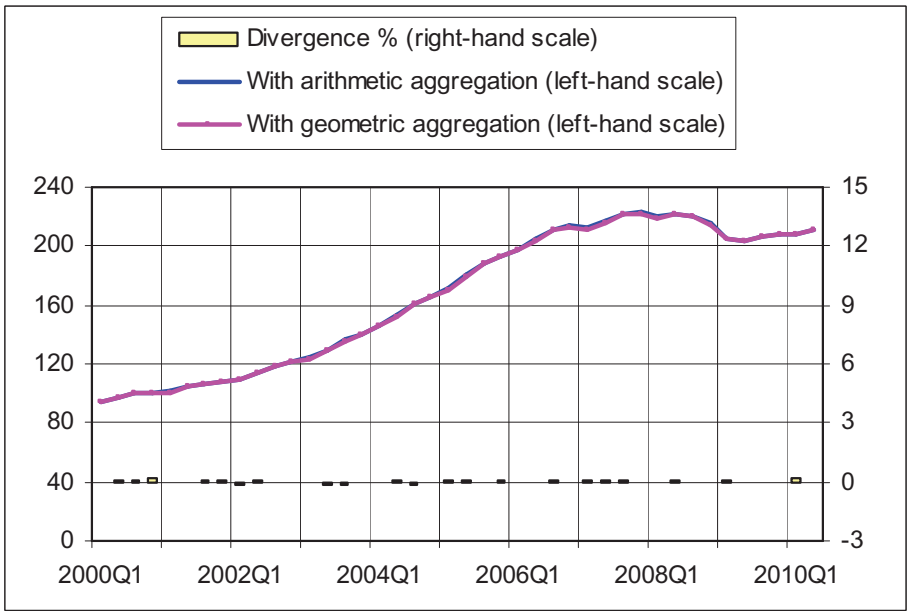
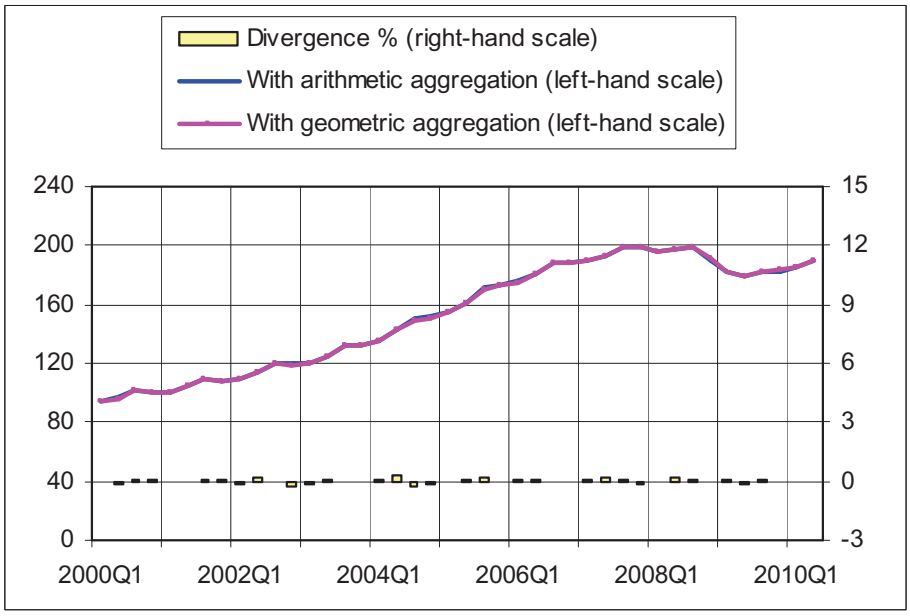


Figure 7.15 - Super-definitive raw price indices for used houses in the Provinces in version 2 with arithmetic and geometric aggregation, base 100=2000Q4



The introduction of chain-linking, on the other hand, which became necessary when correction coefficients were updated more frequently, had a significant effect on price indices for houses (Figures 7.16 and 7.17). The levels of the mean deviation between the chained and non-chained indices from 2002 to 2009 was -0.2 percentage points for apartments and 1.5 points for houses. To calculate the chained indices, the reference and estimation stocks are updated every two years. The indices are aggregated arithmetically. The zoning system used is that of V2.

Figure 7.16 - Super-definitive raw price indices for used apartments in the provinces in version 2 with and without chaining, base 100=2000Q4

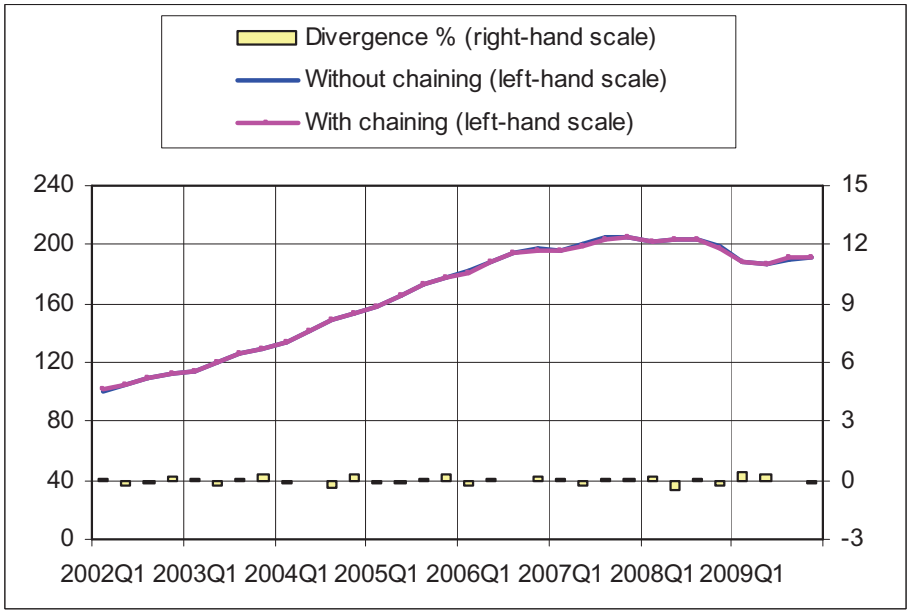
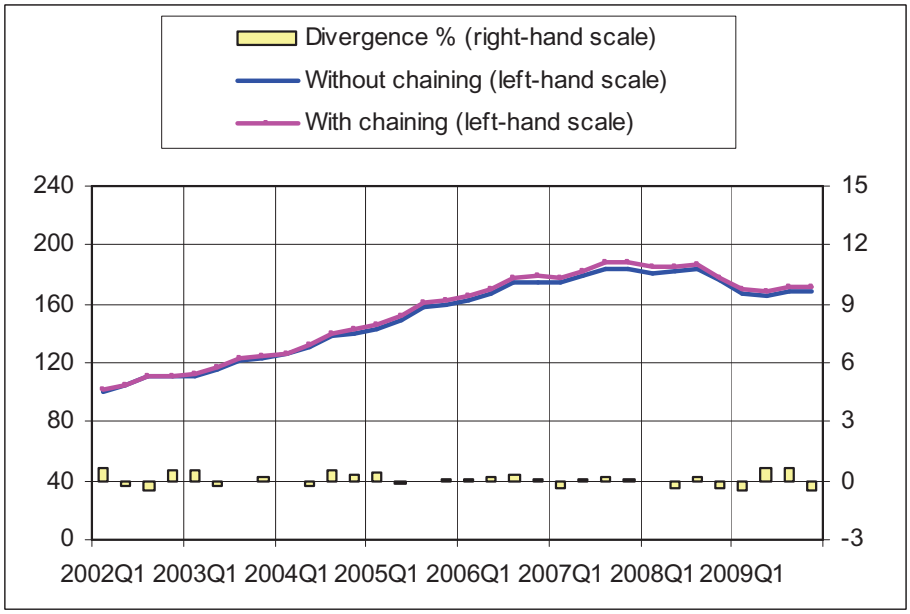


Figure 7.17 - Super-definitive raw price indices for used houses in the Provinces in version 2 with and without chaining, base 100=2000Q4

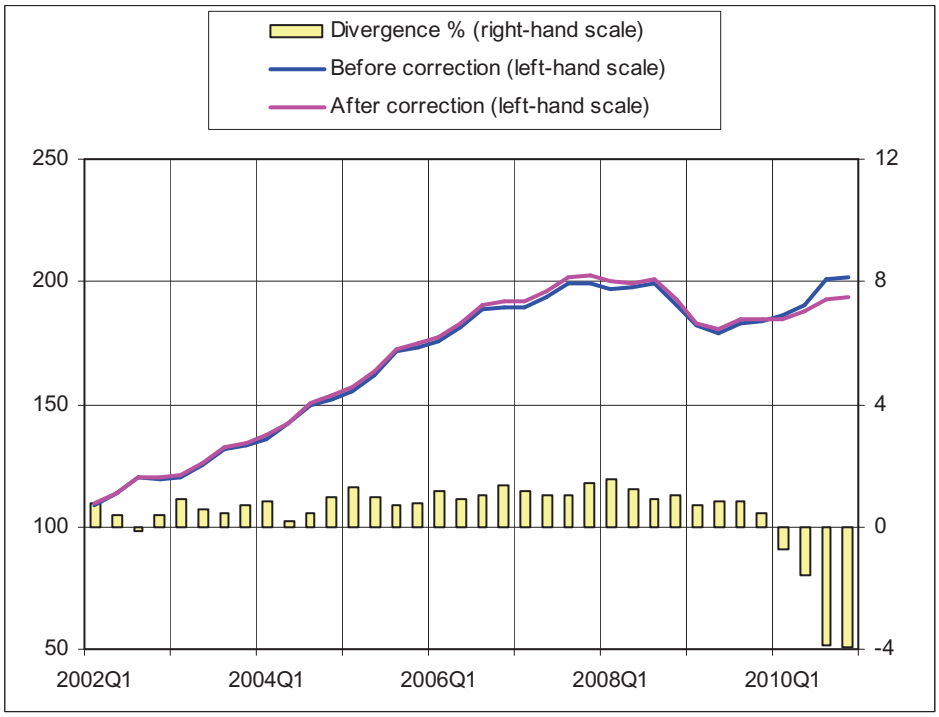


7.3.4 Impact of correcting an error

The housing price indices for the provinces published in version 2 were flawed due to the introduction of an error from mid-2009, when there was a break in coding for “number of garages”. The missing data relating to

this variable up until mid-2009 were not recoded, which resulted implicitly in these cases being assigned to the reference modality, i.e. "1 garage". When new software was introduced, the missing data relating to this variable were replaced using the modality "no garage". The share of this modality then increased steadily, from less than 5% in mid-2009 to more than 35% by the end of 2010. After the error had been corrected, the change in prices between the fourth quarter of 2009 and the fourth quarter of 2010 was then revised downwards by 4% (Figure 7.18). In fact, the error in applying the imputation led to an under-estimation of the change in house quality.

Figure 7.18 - Super-definitive price indices for used houses in the Provinces in version 2 before and after error correction, base 100=2000Q4

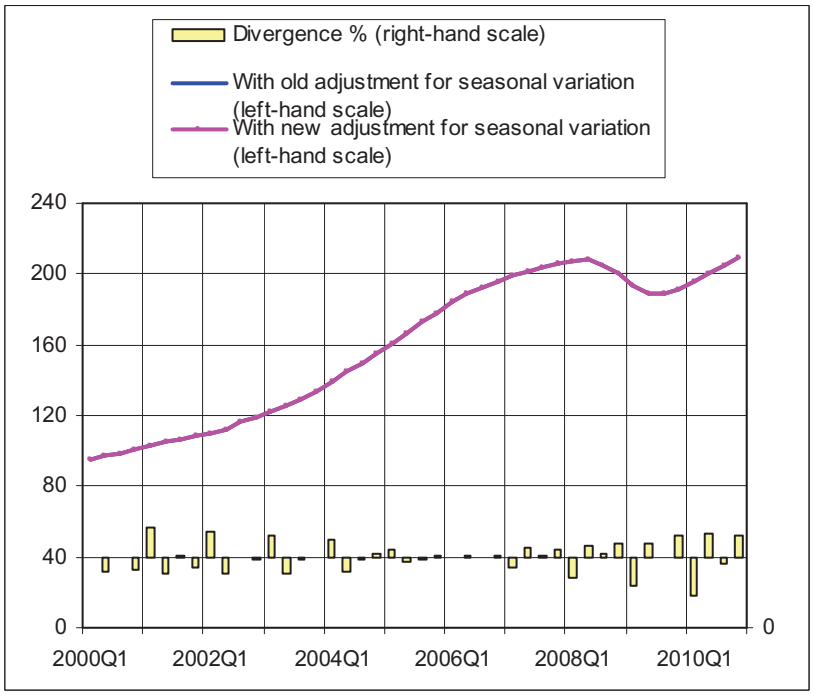


7.3.5 Impact of the new method of seasonal variation adjustment

When version 3 of the Notaires-INSEE indices was adopted, the method used to adjust for seasonal variation was improved. We switched from X11 to X12-Arima. A particular feature is that now, before seasonal adjustment, raw data series are extended by Arima models, which means improved estimates of seasonal coefficients.

The move to the new method of seasonal adjustment had very little effect on the indices, with divergences of no more than 0.1 percentage points for all of metropolitan France (*Figure 7.19*). The differences are highest at the beginning and end of the period, which is consistent with the continuation of the series by the X12-Arima method.

Figure 7.19 - Housing price indices for metropolitan France in version 2 after old and new adjustment for seasonal variation, base 100=2000Q4[55]



[55] N. B.: the right-hand scale in this figure is not the same as in the preceding graphs.

Appendix 1: Transmission of deeds of sale to the notarial databases and estimating coverage rates

Virtually all housing transactions[56] generate a deed (*acte*) certified by a notary. This document is passed on to the tax authorities (first the Mortgage Registry [*Conservation des hypothèques*], then the Land Registry [*Services du Cadastre et des Domaines*]), for registration and payment of transfer tax in most cases, and VAT if the sale concerns a new dwelling. An exhaustive source does therefore exist. However, this is a paper document and does not resemble a standard administrative form as it consists of a written text. Consequently, if this is to become a usable computer file, i.e. a database, the information has to be captured in a coded file. This is a cumbersome and costly operation given the size of the documents (usually several pages) and the volume they represent (approximately 800,000 transactions for used dwellings per year across the whole of France).

The transmission of information from the deeds and the data capture was undertaken by the notaries themselves, first by the Chambre interdépartementale des Notaires de Paris (CINP) in the late 1970s (and especially since 1990), then by the Conseil Supérieur du Notariat (CSN), which set up a limited company, Perval, specifically for this purpose in 1993. These two bodies receive information from the notarial offices and feed it into one of the databases; the first covers the Paris Region, the second covers the Provinces and the French overseas *départements*. It was this arrangement that made it possible – again on the notaries' initiative and with the collaboration of INSEE – to create the price indices described here. Note, however, that there was an index already in existence: it was launched in 1983 and tracked sales of used apartments in Paris. This index was calculated from an average of prices weighted into 72 strata according to the housing stock at the last population census. It was revamped and now applies the same methodology as the other Notaires-INSEE indices.

Until 2010 the information was transferred on paper.[57] Since this date, some of the deeds are now captured in digital format. Teletransmission is still in the process of being introduced in the notarial offices, for feeding the mortgage registry (Projet Télé@ctes) as well as the notarial databases (project to introduce teletransmission to the notarial databases). These long-term undertakings have involved modifying all the software for drawing up deeds (LRA) and changing the way the notarial offices were organised, to ensure that all the necessary information was entered in the LRA. By the end of 2012, about 50% of the deeds had been teletransmitted for the purposes of the indices.

Unfortunately, the notarial offices do not pass on all their deeds, and not all offices transmit anything at all. The coverage rate of the notarial databases varies from one *département* to another. It is important to be aware of this since it does affect the calculation applied when weighting the indices.

Coverage rate

The coverage rate of the notarial databases is obtained from the ratio of the total value of transactions recorded in the databases to the total value of all transactions.[58] The rate is estimated by comparing the value of transactions in the notarial databases with the assessment base for transfer taxes collected by the Directorate General for Public Finances (DGFiP). As the DGFiP supplies the amounts of transfer taxes collected in each *département*, the coverage rate can be calculated by *département*.

However, it is not possible to compute the rate separately for used housing, as there is no difference in the amount of transfer tax to which they are subject. Before 1999, there was a specific transfer-tax rate for used housing and the DGFiP accounts recorded the corresponding revenues. Then all that had to be done was to divide this amount by the tax rate to obtain the assessment base. Since the reforms of 1998 and 1999, used housing transactions are included in the standard regime assessment base. Today there are three tax regimes that cover property transfer tax:

- The "ordinary law" (*droit commun*) regime

[56] With the exception of certain transactions carried out by general government, or property transfers arising from corporate acquisitions.

[57] The notarial offices photocopied the deed (or more often an extract from the deed) and to this they attached a slip of paper which gave information not necessarily provided on the deed extract.

[58] Rates calculated from values are higher than those calculated from numbers of transactions, because the *départements* where the coverage rate is highest are usually the *départements* where the average transaction amount is highest, especially in the Paris Region.

This applies exclusively to transactions involving payment for used property and real estate not subject to value-added tax. Since January 1, 2011, it includes a levy of 3.8%[59] that goes to the *département* and 1.2% to the municipality, 5.0% in all. It used to be 3.6% to the *département*, 1.2% to the municipality and in the last years of the period considered, 0.2% to the State, but this is no longer collected.

- The "special rules" regime.

  This regime applies:

  - to sales of new and assimilated properties purchased with a view to a quick resale ("property developers") and real estate subject to value-added tax,
  - to certain property transactions where there is no charge (donations); these transactions represent about 40% of the assessment base for this tax regime on average in the whole of France.

  Since January 1, 2011, the special rules regime includes a levy for the *département* of 0.7% on transactions with a financial consideration and 0.6% on donations. Previously, this regime included 0.6% for the *département*, and in recent years during the period considered and for transactions with a financial consideration, a levy of 0.1% to the State, which is no longer collected.

- Transfer tax exemption

  This regime applies mainly to purchases by the State, local government and some public establishments.

Since 1999, as used housing transactions are no longer counted separately in the DGFiP accounts, we need to estimate their total value. At the end of 1999, used housing transactions represented about 80% of the ordinary law assessment base. These have been checked for coherence since then and this high proportion has proved to have varied little.[60]

If the coverage rate of the notarial databases were the same for used housing and for total property sales covered by the ordinary law regime, the problem would be solved: the proportion of used housing covered by the entire ordinary law regime would be the same whether all transactions were considered together or only sales recorded in the notarial databases. However, for years prior to 2000, when different types of properties could be differentiated within the transfer tax receipts, the coverage rate of the notarial databases was slightly higher (by about 2%) for used housing than for all transactions covered by the current ordinary law regime. This difference varied considerably from one *département* to another.

We hypothesise that the used housing coverage rate and the rate for all transactions under the ordinary law regime are similar. This may result in a large error in the coverage rate for a given *département*, but this has no effect on the value of the Notaires-INSEE index for this *département*. Any possible error in the value of very aggregated indices (France, Paris Region and Provinces) is also low, as errors in weighting for the entire country are mutually compensated. It is at supra-departmental level, where data are not very aggregated (e.g. a region made up of two *départements*) that the effect on the value of the indices is potentially highest, however, tests have been carried out which have shown that it still remains very low.

[59] At the time of writing, a project was announced to raise this rate to 4.5% (a 0.7% increase) from March 1, 2014 in those *départements* who so wish.

[60] These are checks made by the Scientific Board of the Notaires-INSEE indices when calculating the number of used housing transactions.

Table A1.1 – Coverage rate of notarial databases (transfer tax under ordinary law regime)

Year	Paris Region	Provinces	France
2000	74.0%	58.2%	62.7%
2001	78.4%	59.0%	64.5%
2002	77.4%	53.6%	60.7%
2003	80.3%	55.3%	62.8%
2004	81.7%	54.1%	62.3%
2005	81.5%	55.6%	63.3%
2006	80.3%	56.5%	63.6%
2007	83.0%	57.6%	64.8%
2008	80.4%	58.6%	64.8%
2009	76.6%	56.0%	62.2%
2010	80.0%	56.5%	63.1%
2011	70.1%	51.0%	56.8%
2012	73.4%	51.9%	58.4%

Source: DGFiP and Notaires of France - PERVAL and BIEN databases
How to read this table: coverage rates are relatively stable over time, however, they did fall considerably in 2011. This can be explained by the trend in the number of transactions, which picked up after decreasing substantially in 2008-2009. The increase was not matched by the coding measures when they were put in place.

Appendix 2: Geographic divisions

The primary used-housing price indices are calculated at a very precise geographic level, called the stratum, then aggregated in the form of totals weighted by the structure of the reference stock.[61] Thus the country is divided up into units that are as small as possible to give a network on which to base calculations. A stratum must respect two conditions for it to be of an adequate standard:

- It must include a minimum number of transactions, 110 in each quarter for the period being studied.
- Within the stratum, prices must be homogeneous.

Stratification was carried out using different methods for the Provinces and the Paris Region, given the different contexts of these real estate markets.

- In the Provinces, local markets may be very compartmentalised (with coastal resorts, ski resorts etc.). The process is based on a geographic proximity criterion with the municipality as the primary unit.
- In the Paris Region, we first define price areas: two for the apartment market (where the city of Paris makes up one area and the rest of the Paris Region the other), one for houses (because there are too few houses in Paris). Next, the municipalities are grouped together according to criteria of price level homogeneity. Municipalities assigned to the same stratum are not always contiguous.

In both cases, the aim is to have homogeneous prices within a stratum. This homogeneity is gauged according to price: sale price for houses, sale price per sq.m. for apartments. We also check afterwards that the strata differ one from another.

In the Provinces, studies have determined a higher number of strata in the new version of the indices:

- 97 strata for apartments, compared with 88 in 2002,
- 174 strata for houses in 2008 compared with 146 in 2002.

The increased number of strata stems from the improvements made in feeding the database as well as from an increase in the total number of transactions.

The exhaustive list of municipalities that make up each stratum is given in the attached Excel file.

In the Paris Region, on the other hand, studies have determined that the number of strata should be reduced:

- 5 strata for apartments in Paris, compared with 18 in 2002,
- 10 strata for apartments in the Paris Region, excluding Paris, compared with 37 in 2002,
- 7 strata for houses in the Paris Region, as in 2002.

The exhaustive list of municipalities that make up each stratum is given in the attached Excel file.

Stratification in the Provinces

The transactions analysed cover the period from the 1st quarter of 1998 to the 3rd quarter of 2007, or 39 quarters in all. Division of the territory begins at a fine level of geographic stratification, i.e. the municipality. Any municipality that exceeds the minimum threshold of 110 transactions per quarter is designated a stratum. In practice this only happens for the largest municipalities. When this is not the case, municipalities are aggregated into the same stratum if their price levels are consistent. There are also two types of municipality that are differentiated as being specific cases: ski resorts and coastal resorts.

Thus the Provinces are divided into strata by geographic level and according to different types of property market. The levels of division may be:

- the municipality: this is the smallest unit.[62] This division is the first step in the process. The number of quarterly sales is calculated for all the municipalities in France; we retain those that exceed the minimum threshold of 110 transactions per quarter.
- the urban unit: strata are combined into urban units and we remove those municipalities that have already been designated as strata. Thus there may be strata that consist of only one suburb or others

[61] See definitions in Chapter 3.
[62] There is no stratum for *arrondissements*. The *arrondissements* of Lyon and Marseille are dealt with using dummies in the regressions.

that consist of the entire agglomeration when a large city (or municipality) does not have sufficient transactions to form a stratum by itself.

- Alpine winter sports resorts: these are grouped into strata by mountain range or by *département* when this is possible. The *départements* concerned are Savoie, Haute-Savoie and Isère, given the size of their ski areas. Three strata of municipalities are created for apartments, by mountain range:
    - Stratum 01: Tarentaise
    - Stratum 02: Maurienne, Val d'Arly, Beaufortin, Les Bauges, Chablais, Giffre, Les Bornes, and Chartreuse,
    - Stratum 03: Mont-Blanc and Aravis.

For Isère, we create a specific class for municipalities that include a ski resort (i.e. stratum 04: Isère). For the other mountain ranges (southern Alps, Pyrénées, Vosges and Massif Central), it is not possible to create specific strata because the volume of sales is too small.

- coastal municipalities: the strata are obtained by grouping together all municipalities in the same *département* or region on a geographic basis. They are differentiated from nearby municipalities because their average house price is higher than in the surrounding municipalities (the divergence is set at 15 or 20% depending on the specific case).

Municipalities are grouped together on the basis of geographic proximity, and according to national territorial divisions: the municipality, the urban unit, group of coastal municipalities, group of ski resorts, up to a *département* or even a region.

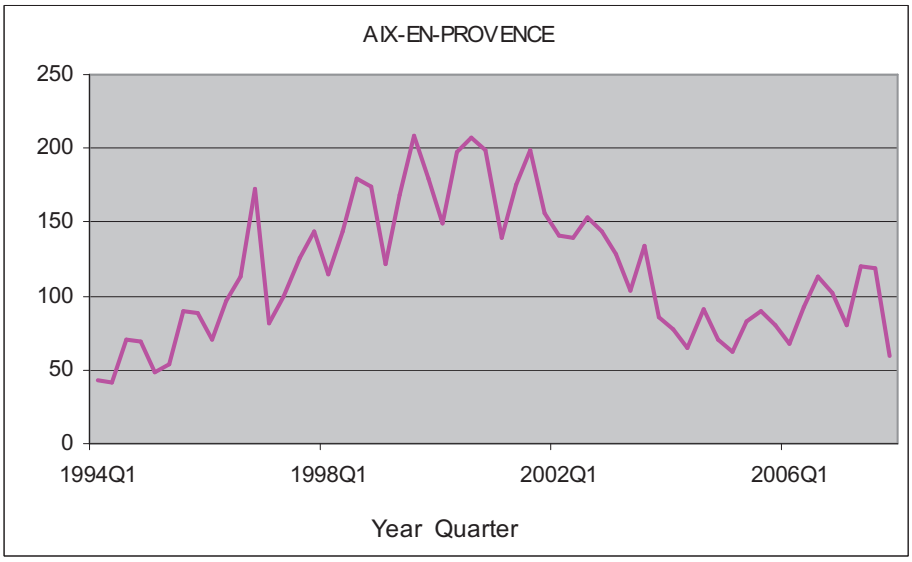
A few strata that come just below the 110 transaction cut-off for just a few quarters could be selected if the development in the number of transactions seemed favourable. Thus, for apartments, the Aix-en-Provence stratum was removed because the number of transactions fell considerably. It regularly dropped below the threshold from 2003 onwards. In Caen, on the other hand, the number of transactions increased over the period and remained above the threshold from 1998. This stratum was therefore created. (Figures A2.1 and A2.2).

Next we deal with urban units, coastal municipalities and municipalities within the mountain ranges most popular with tourists, and rural municipalities. Rural municipalities do not have enough sales per quarter to be selected, nor do they have any particular feature that could bring them up. We therefore group them into classes according to the quantiles for average income per inhabitant for 2006*,* with each class constituting a stratum. This is done using a descending process, looking first at national economic and development areas ("Zones for Study and Development" - ZEAT). Each stratum has to fulfil the following two criteria:

- a volume of transactions greater than 8,000 or 9,000 transactions between 1998 and 2007,
- a minimum of 110 transactions per quarter between 1998 and 2007.

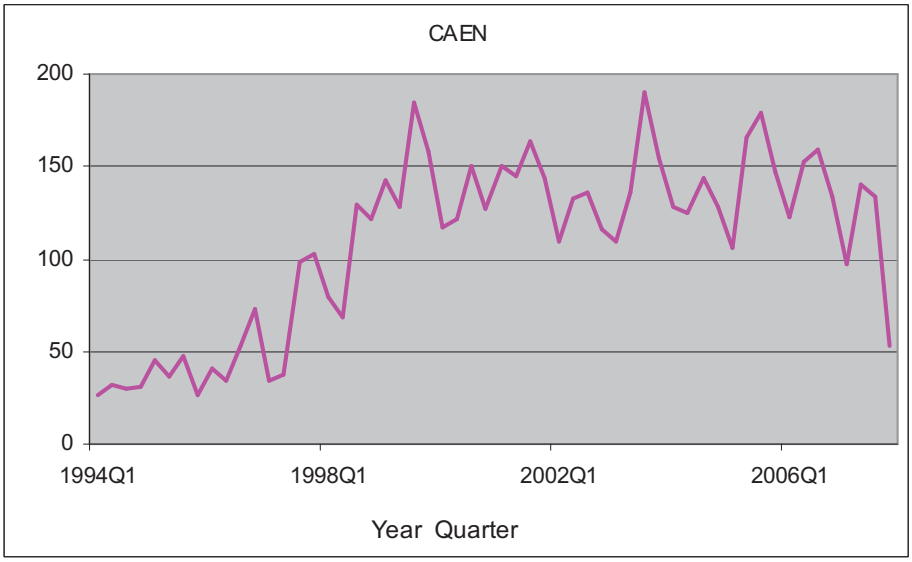
If the thresholds set for a ZEAT are widely exceeded, then we move down to regional level. The same is then done for a given region, and sometimes we have to move down to an even smaller level, the *département*. In other instances, we group together two *départements* from the same region to form one stratum (cf. Table A2.1 showing the grouping for the Ardennes and Marne *départements,* for example).

Figure A2.1 - Number of quarterly transactions in Aix-en-Provence (apartments)



Source: Perval database

Figure A2.2 - Number of quarterly transactions in Caen (apartments)



Source: Perval database

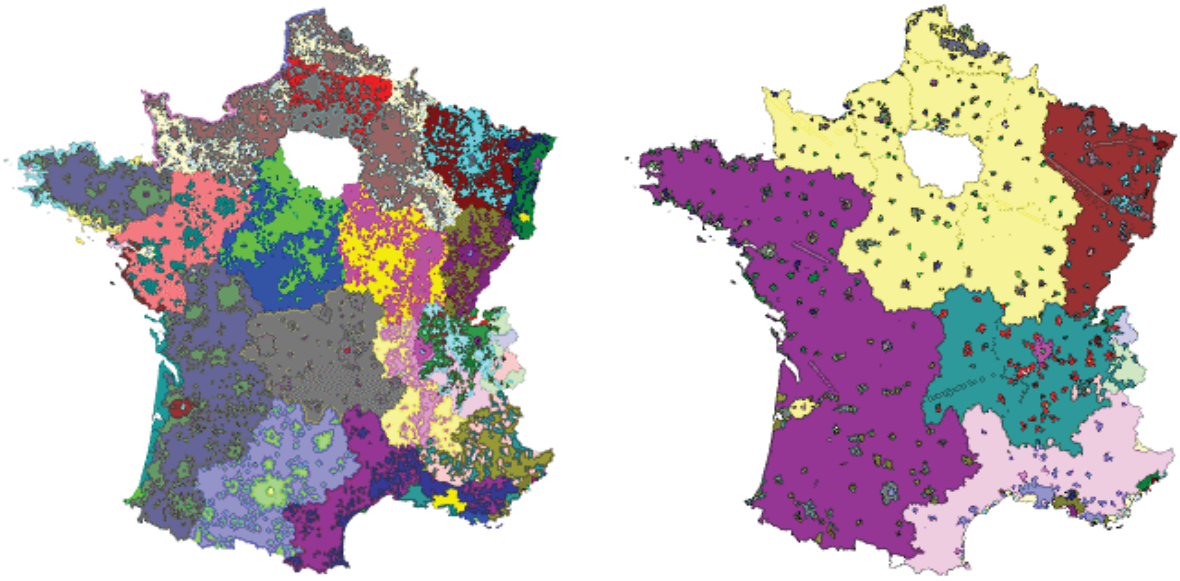
Table A2.1 - Example of breakdown of geographic areas by "income" variable (houses)

Stratum code	*Département*	Region	Lower bound of income for the stratum (€)	Upper bound of income for the stratum (€)	Total number of quarters when 110 dwellings threshold exceeded	Total sales from 1998 to 2007
99210	Grouping of Ardennes and Marne *départements*	Champagne Ardennes	7,824	14,480	39	9,564
99211			14,480	18,821	39	11,535
99212			18,821	55,946	39	10,678
99213	Grouping of Aube and Haute-Marne *départements*		8 ,321	15,106	39	8,921
99214			15,106	55,835	39	8,640
99220	Somme	Picardie	9,342	13,543	39	8,848
99221			13,543	15,638	39	9,373
99222			15,638	40,192	39	9,159
99223	Aisne		9,998	14,889	39	11,616
99224			14,889	17,456	39	11,741
99225			17,456	20,190	39	11,635
99226			20,190	72,021	39	12,147
99227	Oise		8,299	13,431	39	8,717
99228			13,431	15,505	39	8,868
99229			15,505	44,817	39	9,118

Source: PERVAL database

Figure A2.3 – Strata for apartments in the Provinces

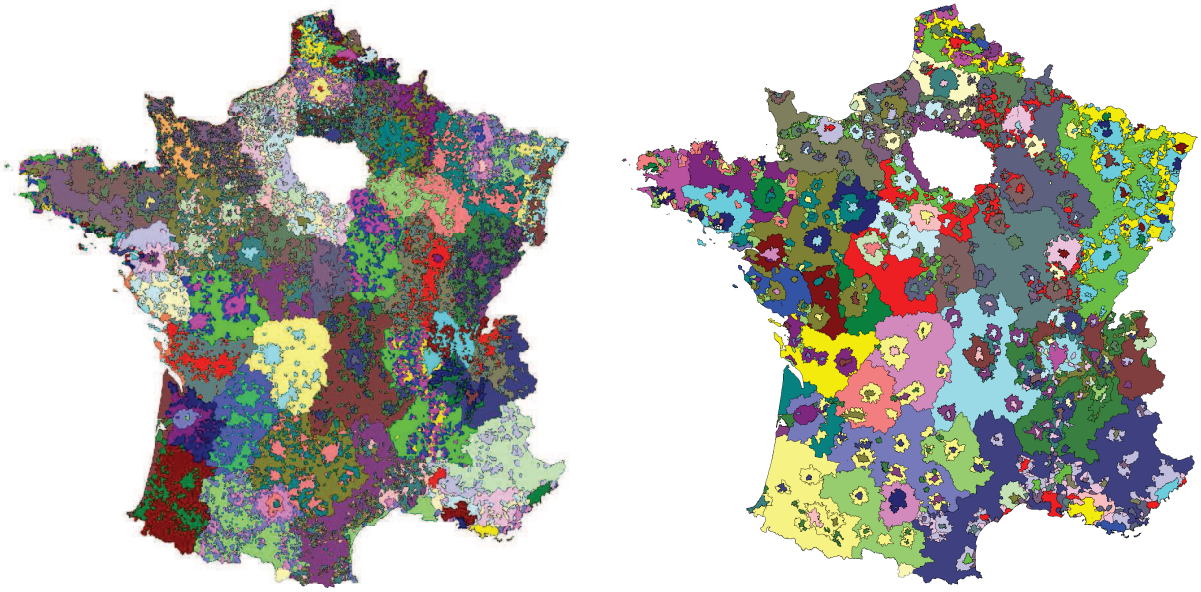
New database (indices version 3) *Old database (indices version 2)*



Source: PERVAL database
How to read this chart: strata are distinguished by their colour. For example, in version 2, the large yellow expanse around the Paris Region corresponds to the strata of rural municipalities in North-West France. As there are a large number of strata, some are shown on the maps in the same colour. They can be differentiated one from another by the fact that they are not contiguous.

Figure A2.4 – Strata for houses in the Provinces

New database (indices version 3) *Old database (indices version 2)*



Source: PERVAL database
How to read this chart: see above.

Stratification in the Paris Region

In the Paris Region, the transactions analysed are those from 2003 to 2007, or 20 quarters in all.

The first step consists of combining the information using factorial correspondence analysis (CA). The stratification per se is carried out in a second step using a hierarchical ascendant classification (HAC).

- First step: combine the information

As the HAC can accept only qualitative variables, the continuous variables are discretised. We look for the best possible summary of the information by reducing the number of axes for analysis. Each variable modality can be seen as an axis for analysis. We check that the contribution of each variable is not too dominant compared with the average contribution on the first axes. To establish the cut-off points, the method relies on statistical indicators, such as centiles, mean and standard deviations. Next, the little used modalities are combined. If the number of instances for a modality represents less than 5% of the total, it is combined with another.

- Second step: compile the strata

Before starting the stratification, we compile statistics at neighbourhood level (for Paris) or canton level (for the rest of the Paris Region), to cross the following variables: living space in the dwelling (bracketed values), plot size for houses (also by bracketed values), number of bathrooms, number of rooms, construction period, presence of lift for apartments, floor of the building, presence of cellars, number of garages and presence of balconies or terraces. Paris, where average prices are higher, is dealt with separately. The price per sq.m. for apartments and sale price for houses are incorporated as supplementary variables in our analysis. They can also be used to define the resulting classes.

The classification is done step by step, aggregating observations that are “most similar”. To do this, we use the distance between two observations or two classes to group together those that are closest. As the aggregation progresses, the classes include more and more observations.

- Example: the classification for apartments in Paris

First the number of classes is determined. Using the *Tree analysis* (*Figure A2.5*) we can determine visually the optimal number of classes. This graph shows the order in which classes are formed and also measures the value of aggregation distance between classes. A large “jump” on the tree corresponds to a large increase in the value for this distance. In practice, we select a cut-off in the upper part of the diagram, which is where we have classes that are still constructed with fairly short aggregation distances (hence classes that are fairly homogeneous), and just before a large jump (making sure not to go too

high up in the tree where the classes are too varied). In our case, the large jump corresponds to a cut-off below the split into 5 classes.

A second diagram (Figure A2.6) shows for each class, the modalities of variables and the individual variables (neighbourhoods grouped by *arrondissement*) that contribute most to defining the class. The diagram shows, in terms of weight in the total for transactions in Paris, the five classes of modalities of variables that are well-represented in the class and the geographic distribution of dwellings. Crossing these three criteria gives the profile for each class.

Profile of class 1

This class represents 20% of transactions in Paris. The best represented modalities of variables are:

- high price per sq.m.,
- older construction period,
- apartment without cellar,
- apartment without lift,
- apartment without garage.

The geographic distribution of dwellings in this class is as follows:

- 100% of dwellings in the 1st, 2nd,3rd, 4th and 6th *arrondissements*,
- 100% of dwellings in the 9th *arrondissement*,
- 72% of dwellings in the 10th *arrondissement*,
- 70% of dwellings in the 5th *arrondissement*,
- 50% of dwellings in the 7th *arrondissement*.
- 20% of dwellings in the 12th *arrondissement.*

Profile of class 2

This class represents 34% of transactions in Paris. The best represented modalities of variables are:

- medium price per sq.m.,
- older construction period,
- small living space between 20 and 40 sq.m.,
- apartment without lift,
- apartment without bathroom,
- apartment without garage.

The geographic distribution of dwellings in this class is as follows:

- 100% of dwellings in the 11th and 18th *arrondissements*,
- 63% of dwellings in the 17th *arrondissement*,
- 50% of dwellings in the 20th *arrondissement*,
- 39% of dwellings in the 14th *arrondissement*
- 30% of dwellings in the 5th *arrondissement*,
- 28% of dwellings in the 10th and 13th *arrondissements*.

Profile of class 3

This class covers 24% of transactions in Paris. The best represented modalities of variables in this class are:

- fairly high price per sq. m.,
- fairly recent construction period,
- average living space,
- apartment with garage,

- apartment with cellar
- apartment with lift.

The geographic distribution of dwellings in this class is as follows:

- 100% of dwellings in the 15th *arrondissement*,
- 80% of dwellings in the 12th *arrondissement*,
- 47% of dwellings in the 14th *arrondissement*,
- 30% of dwellings in the 19th *arrondissement*,
- 27% of dwellings in the 20th *arrondissement*.
- 12% of dwellings in the 17th *arrondissement*.

Profile of class 4

This class represents 10% of transactions in Paris. The best represented modalities of variables are:

- low price per sq. m.,
- recent construction period,
- apartment with garage,
- apartment with cellar.

The geographic distribution of dwellings in this class is as follows:

- 72% of dwellings in the 13th *arrondissement*,
- 70% of dwellings in the 19th *arrondissement*,
- 23% of dwellings in the 20th *arrondissement*,
- 15% of dwellings in the 14th *arrondissement*.

Profile of class 5

This class represents 12% of transactions in Paris. The best represented modalities of variables are:

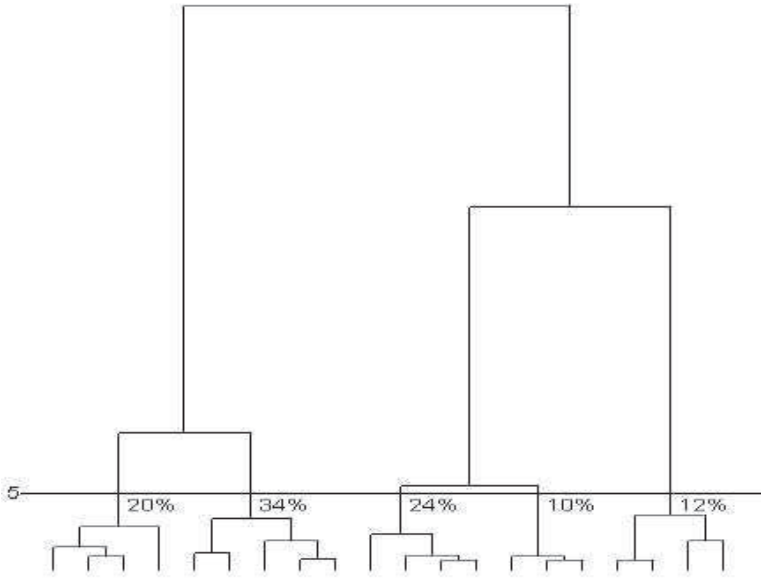
- high price per sq. m.,
- construction period between 1950 and 1970,
- apartment with large living space,
- apartment with 2 or more bathrooms,
- apartment with lift.

The geographic distribution of dwellings in this class is as follows:

- 100% of dwellings in the 8th and 16th *arrondissements*,
- 50% of dwellings in the 7th *arrondissement*,
- 15% of dwellings in the 17th *arrondissement*.

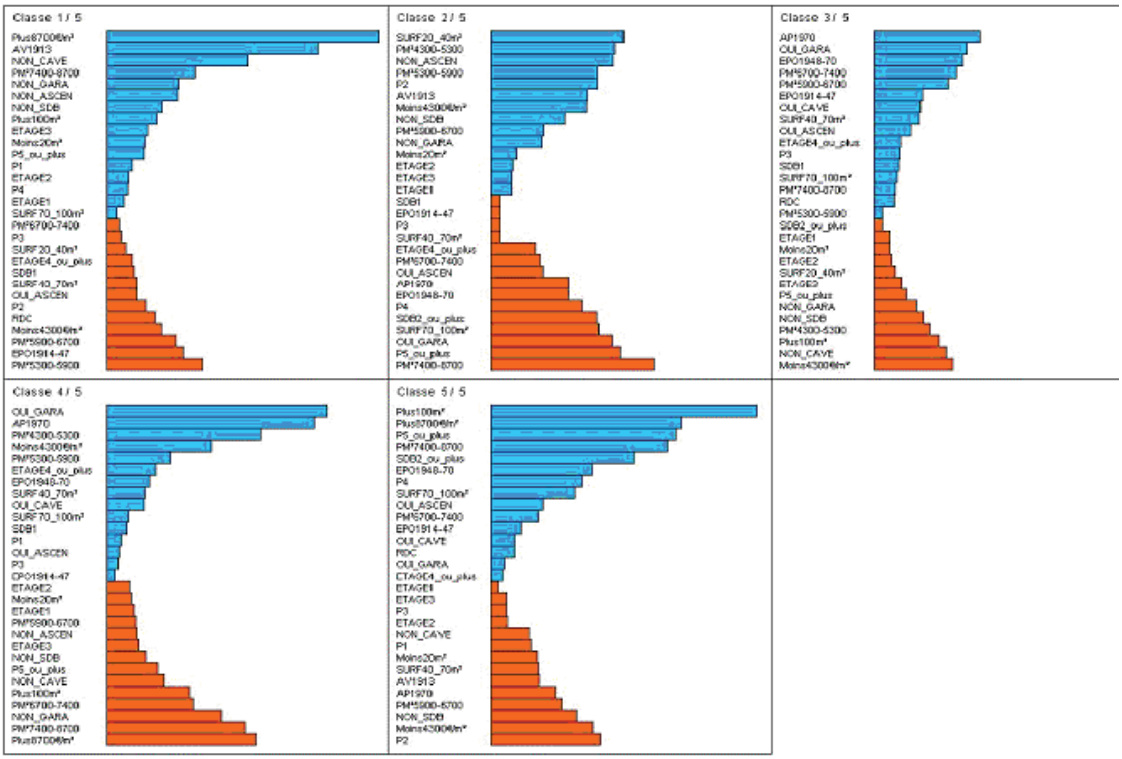
Figure A2.7 gives a visualisation of the spatial distribution of the five classes for apartments in Paris.

Figure A2.5 – Tree analysis for apartments in Paris



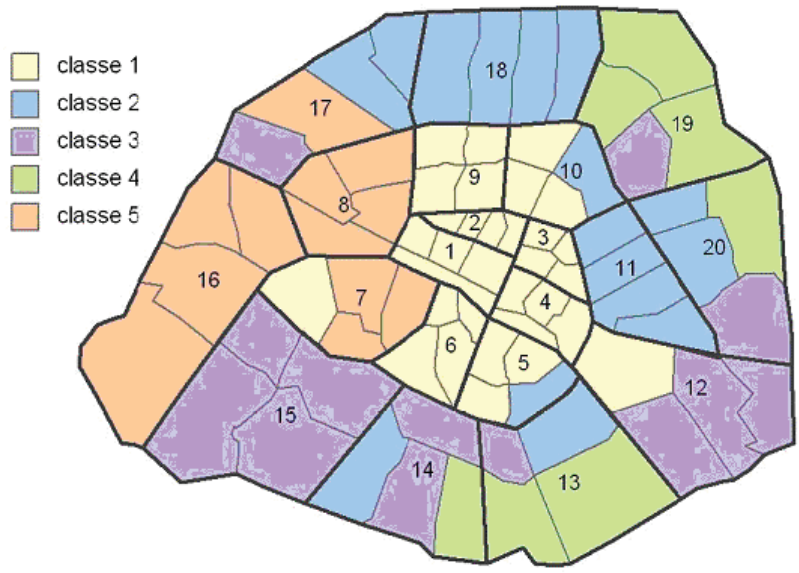
Source: BIEN database
How to read this chart: the straight line defining the division into 5 classes corresponds to an optimal choice in terms of strategy.

Figure A2.6 – Looking for over-represented variables for each class



Source: BIEN database
How to read this chart: The blue bars correspond to over-represented variables in the class.

Figure A2.7 - Division into classes for apartments in Paris



Source: BIEN database

Testing for strata homogeneity

Once the division into strata has been completed, we check the homogeneity of prices within a stratum, and also heterogeneity between strata. We propose to take as an example the studies carried out by the PNS for apartments in Paris. The studies by Min.not are similar.

Table A2.2 provides a first visual examination of average prices by strata for apartments in Paris. From 2003 - 2007, these averages varied from €4,514 to €6,549 per sq.m. Confidence intervals are very often unconnected, which seems to indicate that prices are different from one stratum to another.

Table A2.2 – Statistics for prices per sq.m. in apartments in Paris strata

Stratum	Number of transactions	Average	Standard deviation	Standard error	95% confidence interval for mean		Minimum	Maximum
					Lower bound	Upper bound		
7501	24,724	6,190.61	2,037.04	12.96	6,165.21	6,216.00	322.58	2,4851.37
7502	40,714	4,919.52	1,323.92	6.56	4,906.66	4,932.38	148.24	1,8152.17
7503	28,771	5,453.07	1,361.13	8.02	5,437.34	5,468.80	435.15	1,9402.71
7504	12,488	4,514.04	1,208.99	10.82	4,492.83	4,535.24	431.97	1,7784.15
7505	16,339	6,549.16	1,874.81	14.67	6,520.41	6,577.90	877.37	2,2000.00
Total	123,036	5,474.97	1,707.11	4.87	5,465.43	5,484.51	148.24	2,4851.37

Source: PNS

The Levene method is used to ensure that n samples extracted from the same population are of equal variance (Table A2.3). Finally, we carry out a Fisher's test (Table A2.4). One-factor analysis of variance also enables us to see whether the geographic area (factor) has an influence on the price of the property. The homogeneity hypothesis in terms of prices of strata can be rejected when significance level $\alpha = 0.05$. There are therefore different variances and different mean effects. But we do not know if the five strata are different or if it is only some of the strata that are not homogeneous one with another. To solve this problem, a post-hoc test is subsequently carried out (cf. Table A2.5) to see if all the strata are different when taken two by two. In our example, this test compares the five strata in Paris. For each stratum, the significance level is below the threshold $\alpha = 0.05$. Price differences are therefore observed between the two strata for apartments in Paris. Similar results are found for the other strata in the Paris Region and in the Provinces.

Table A2.3 - Test for equality of variances

Levene's test statistic W	Degrees of freedom 1	Degrees of freedom 2	P value of test
1,276.53	4	123,031	0.000

Source: BIEN database
H0: no difference in variance between strata
H1: at least one different variance in one stratum
If the hypothesis H0 is not confirmed, this indicates that the result does not derive from a simple sampling from one and the same population. In this case, variances of the groups are no longer equal amongst themselves at significance level α.
When significance level α is equal to 0.05, hypothesis H0 is rejected; the variances are not equal.

Table A2.4 – Analysis of variance for strata of apartments in Paris

	Sum of squares	Degrees of freedom	Mean square	Fisher's statistic	Significance level of Fisher's test F (Pvalue)
Inter-stratum variance	55,621,427,170.7	4	13,905,356,792.5	5,647.5	>0.000
Intra-stratum variance	302,929,525,100.8	123,031	2,462,221.1		
Total	358,550,952,270.9	123,035			

Source: BIEN database
H0: there is no difference in price between means for the strata
H1: there is a difference in price between means for the strata

Table A2.5 - *Post-hoc* test for strata of apartments in Paris

(I) Reference stratum	(J) Compared with strata	Difference in means (I-J)	Standard error	Significance	95% confidence interval	
					Upper bound	Lower bound
7501	7502	1,271.08	14.52	0.03	1,227.42	1,314.75
	7503	737.54	15.24	0.03	690.34 1	784.74 1
	7504	1,676.57	16.88	0.02	626.24	726.90
	7505	-358.55	19.57	0.02	-417.15	-299.95
7502	7501	-1,271.08	14.52	0.03	-1,314.75	-1,227.42
	7503	-533.55	10.37	0.03	-566.18	-500.92
	7504	405.49	12.65	0.00	370.30	440.67
	7505	-1,629.63	16.07	0.00	-1,674.40	-1,584.86
7503	7501	-737.54	15.24	0.03	-784.74	-690.34
	7502	533.55	10.37	0.03	500.92 901.05	566.18
	7504	939.04	13.47	0.00	-1,143.17	977.02
	7505	-1,096.08	16.72	0.00		-1 049.00
7504	7501	-1,676.57	16.88	0.02	-1,726.90	-1,626.24
	7502	-405.49	12.65	0.00	-440.67	-370.30
	7503	-939.04	13.47	0.00	-977.02	-901.05
	7505	-2,035.12	18.23	0.01	-2,086.88	-1,983.36
7505	7501	358.55	19.57	0.02	299.95	4,17.15
	7502	1,629.63	16.07	0.00	1,584.86	1,674.40
	7503	1,096.08	16.72	0.00	1,049.00	1,143.17
	7504	2,035.12	18.23	0.01	1,983.36	2,086.88

Source: BIEN database

Description of strata

Strata are geographic areas within which prices are homogeneous (cf. Chapter 3). Details of their breakdown into municipalities are given in an Excel file that can be obtained on request from notaries.

Strata in the Provinces

For the Provinces, the construction of the strata is based on a geographic proximity criterion, starting with the municipality. Table A2.6 shows the distribution of transactions in the reference stock of 2007-2008 in the Provinces, by part of urban unit 1999.

Table A2.6 - Provinces: transactions (apartments and houses) by part of urban unit and municipality grouping (%)

Part of urban unit 1999	Rural	UU < 5,000 inhab.*	5,000-10,000 inhab.	10,000-20,000 inhab.	20,000-50,000 inhab.	50,000-100,000 inhab.	100,000-200,000 inhab.	Over 200,000 inhab.	Total
Apartments									
Rural	6%	0%	0%	0%	0%	0%	0%	0%	6%
Central city	0%	2%	3%	5%	6%	9%	8%	33%	67%
Suburb	0%	0%	0%	1%	1%	2%	3%	20%	27%
Total	6%	2%	3%	6%	8%	11%	11%	53%	100%
Houses									
Rural	39%	0%	0%	0%	0%	0%	0%	0%	39%
Central city	0%	8%	6%	5%	5%	4%	2%	5%	35%
Suburb	0%	1%	1%	2%	2%	3%	3%	14%	26%
Total	39%	8%	7%	7%	7%	8%	5%	20%	100%

* Inhab. = inhabitants
2007-2008 Reference stock

Strata for apartment indices in the Provinces

The entire country was divided into 97 strata for used apartments: 33 strata consist of a single city (Table A2.7), thus an index is calculated for each of these 33 cities. For the other strata (Tables A2.8 et seq.), 12 are in the suburbs of these cities, 5 correspond to an entire urban unit, 14 consist of atypical municipalities (coastal towns and winter sports resorts); the remaining 33 are groupings of towns made according to the region and the average income per inhabitant (see definitions below).

A certain number of series of indices disseminated by the notaries do not have the Notaires-INSEE designation. These are mainly regional and departmental series and indices for the 33 municipalities listed in Table A2.7.

These indices are disseminated on the websites immoprix.com and perval.fr, also via press conferences organised by the Chambers of Notaries on property prices and via survey reports produced by the notaries.

The list of indices that may be disseminated is checked approximately every two years. An index series can be disseminated when the number of transactions per quarter is greater than or equal to 110 for several consecutive quarters.

Table A2.7 – List of 33 cities in the Provinces for which an apartment index is calculated

Municipality	Population*	Number of dwellings*	Number of transactions in reference stock**
Antibes	76,994	27,135	1,871
Agde	22,487	3,862	1,719
Annecy	50,115	20,372	1,020
Besançon	117,599	36,396	1,974
Bordeaux	235,891	78,724	2,539
Brest	142,097	40,430	2,411
Caen	109,899	33,206	942
Cannes	72,939	26,181	2,836
Le Cannet	40,940	15,169	1,205
Clermont-Ferrand	139,006	42,322	1,349
Dijon	151,576	52,136	2,628
Grenoble	156,659	61,561	2,212
Le Havre	178,769	32,486	1,304
Lille	225,784	61,103	2,328
Lyon	474,946	188,776	6,808
Marseille	851,420	244,867	6,160
Menton	28,833	9,775	1,133
Metz	122,838	35,076	1,566
Montpellier	252,998	84,784	2,399
Nancy	106,361	40,915	1,756
Nantes	283,288	82,438	4,123
Nice	439,553	132,175	8,541
Orléans	113,257	28,652	1,593
Pau	84,036	27,637	1,209
Reims	181,468	33,262	1,913
Rennes	206,655	65,019	2,972
Rouen	109,425	38,595	1,448
Saint-Étienne	172,696	53,944	1,743
Strasbourg	272,116	84,070	2,929
Toulon	166,733	50,808	1,906
Toulouse	439,553	152,830	4,674
Tours	135,480	37,007	1,380
Villeurbanne	141,106	47,334	1,963

* Source: population census 2008

** 2007-2008 stock

With the exception of Lyon and Marseille, the associated indices do not have the Notaires-INSEE designation.

Table A2.8 - List of 17 strata (suburb or urban unit in the Provinces) for which an apartment index is calculated

Stratum	Number of transactions in reference stock*
Suburb of Lille (French part)	1,903
Suburb of Annecy	1,111
Suburb of Bordeaux	1,861
Suburb of Grenoble	2,340
Suburb of Nancy	1,458
Suburb of Nantes	1,099
Suburb of Rouen	936
Suburb of Strasbourg (French part)	1,547
Suburb of Nice	4,941
Suburb of Toulon	3,807
Suburb of Lyon	4,689
Suburb of Marseille-Aix-en-Provence	2,108
Urban unit of Chambéry	1,148
Urban unit of Mulhouse	1,602
Urban unit of Saint-Nazaire	848
Urban unit of Genève-Annemasse (French part)	1,944
Urban unit of Bayonne	1,953

* 2007-2008 stock

Twelve strata are composed of the suburbs of an urban unit and five are composed of the entire urban unit. The associated indices do not have the Notaires-INSEE designation and are not published.

Table A2.9 – List of 14 strata of winter sports and coastal resorts for which an apartment index is calculated

Stratum	Number of transactions in reference stock*
Coastal towns in 'Grand Nord', Nord - Pas-de-Calais and Picardie regions	2,013
Coastal towns in Haute and Basse Normandie regions	2,228
Coastal towns in Bretagne *départements* Ille-et-Vilaine and Morbihan	2,062
Coastal towns in Bretagne *départements* Finistère and Côte d'Armor	2,262
Coastal towns in Pays de la Loire region	1,140
Coastal towns in South-West Poitou-Charentes and Aquitaine regions	2,584
Coastal towns in Languedoc-Roussillon South	2,457
Coastal towns in Languedoc-Roussillon North	2,038
Coastal towns in PACA West	1,551
Coastal towns in PACA East	3,058
Ski resorts in Tarentaise	1,871
Ski resorts in Maurienne, Val d'Arly, Beaufortin, les Bauges, Chablais, Giffre, Les Bornes and Chartreuse	1,775
Ski resorts in Mont-Blanc and Aravis	1,136
Ski resorts in Isère	1,198

* 2007-2008 stock

Fourteen strata are composed of "atypical" municipalities (e.g. coastal municipalities, winter sports resorts), which had to be isolated in order to obtain coherent strata in the regressions. The associated indices do not have the Notaires-INSEE designation and are not published.

Table A2.10 - Other strata for which apartment indices are calculated, by average income 2006

Stratum	Number of transactions in reference stock by income bracket*		
	Low income	Medium income	High income
Other municipalities in Alsace	2,052		1,594
Other municipalities in Bourgogne	1,703		1,444
Other municipalities in Franche-Comté	1,657		1,512
Other municipalities in Lorraine	2,669		2,505
Other municipalities in the West except Pays de la Loire	2,580		1,812
Other municipalities in PACA North	1,775	1,667	1,454
Other municipalities in Picardie	1,867		1,419
Other municipalities in the Centre	2,526		1,956
Other municipalities in Limousin and Auvergne	1,763		1,503
Other municipalities in Midi-Pyrénées	2,076		1,596
Other municipalities in the North except Picardie	1,908	2,118	1,518
Other municipalities in Pays de la Loire	1,936		1,689
Other municipalities in Rhône-Alpes-East	2,051	1,900	2,498
Other municipalities in Rhône-Alpes-West	2,404		2,198
Other municipalities in the South-East (Languedoc-Roussillon and PACA)	2,016		3,236

* 2007-2008 stock

Municipalities in the 33 strata listed above have insufficient sales per quarter, and do not have any particular features to distinguish them. They have therefore been classified according to the average income per inhabitant for 2006.
The associated indices do not have the Notaires-INSEE designation and are not published.

Strata for house indices in the Provinces

There are 37 strata that concern towns, suburbs or an entire urban unit (Table A2.11); the other 137 strata are groupings of towns by their region and their average income per inhabitant in 2006 (Table A2.12).

Table A2.11 – List of 37 towns, suburbs and urban units for which a house index is calculated

Municipality	Number of transactions in reference stock**
Amiens	970
Bordeaux	1,192
Le Havre	868
Le Mans	1,394
Lille	931
Marseille	826
Nantes	1,146
Toulouse	744
Tourcoing	1,060
Suburb of Bordeaux	3,631
Suburb of Lille (French part)	6,772
Suburb of Marseille-Aix-en-Provence	1,990
Suburb of Nantes	2,640
Suburb of Toulouse	2,222
Urban unit of Avignon	949
Urban unit of Béthune	1,417
Urban unit of Brest	1,227
Urban unit of Calais	1,129
Urban unit of Clermont-Ferrand	929
Urban unit of Dijon	907
Urban unit of Douai-Lens	3,298
Urban unit of Dunkerque	970
Urban unit of Grenoble	800
Urban unit of Lyon	3,811
Urban unit of Maubeuge (French part)	903
Urban unit of Metz	1,547
Urban unit of Montpellier	815
Urban unit of Nancy	2,258
Urban unit of Nice	3,241
Urban unit of Orléans	1,966
Urban unit of Reims	906
Urban unit of Rennes	1,047
Urban unit of Rouen	2,708
Urban unit of Saint-Nazaire	770
Urban unit of Toulon	2,089
Urban unit of Tours	1,709
Urban unit of Valenciennes (French part)	1,991

* 2007-2008 stock

With the exception of the Lille index, these indices do not have the Notaires-INSEE designation and are not published.

Table A2.12 - Other strata for houses in the Provinces

Stratum	Number of transactions from the reference stock**			
	Income			
	Very low	Low	Medium	High
Ski resorts in Isère, Savoie and Haute Savoie				1,597
Coastal towns in Gironde, Landes and Pyrénées-Atlantiques		1,573		1,127
Coastal towns in Loire-Atlantique, Vendée and Charente-Maritime		2,186		1,762
Coastal towns in Ille-et-Vilaine, and Morbihan	1,128	985	958	1,044
Coastal towns in Finistère and Côte d'Armor	1,434	1,263	1,139	1,186
Coastal towns in Nord - Pas-de-Calais and the two Normandies	1,244	1,385	1,359	1,148
Coastal towns in the Mediterranean		1,789	1,490	1,583
Other municipalities in Alsace		1,567	1,685	1,686
Other municipalities in Aquitaine North-East		2,020		1,852
Other municipalities in Aquitaine South-West		2,179		2,301
Other municipalities in Auvergne		1,409	1,779	1,985
Other municipalities in Basse-Normandie	2,320	2,220	2,219	2,134
Other municipalities in Bourgogne East (21,71)		2,287	2,254	2,184
Other municipalities in Bourgogne West (89,58)		1,453	1,620	1,967
Other municipalities in Bretagne North		2,640	2,455	2,309
Other municipalities in Bretagne South		2,371		2,451
Other municipalities in Champagne-Ardenne North		1,951	1,996	1,940
Other municipalities in Champagne-Ardenne South		1,836		1,703
Other municipalities in Franche-Comté		2,308	2,148	2,134
Other municipalities in Gironde		1,908		2,003
Other municipalities in Somme		1,611	1,633	1,511
Other municipalities in Ain		2,117		2,236
Other municipalities in Aisne	1,760	1,976	1,723	1,934
Other municipalities in Eure		1,999	1,853	1,870
Other municipalities in Isère		1,889		1,973
Other municipalities in Loire-Atlantique		2,354		2,141
Other municipalities in Oise		1,468	1,455	1,450
Other municipalities in Lorraine East (54,57)		2,429	2,408	2,548
Other municipalities in Lorraine West (55,88)		2,279		2,126
Other municipalities in Midi-Pyrénées North		1,272	1,144	1,537
Other municipalities in Midi-Pyrénées South		1,500	1,479	1,624
Other municipalities in PACA North		1,708		1,717
Other municipalities in PACA South		1,398	1,570	1,187
Other municipalities in Poitou-Charentes North		1,703	1,792	1,621
Other municipalities in Poitou-Charentes South		3,159		3,103
Other municipalities in Savoie and Haute-Savoie		1,532	1,525	1,425
Other municipalities in Seine-Maritime		1,670	1,917	1,844
Other municipalities in Vendée		1,703		1,609
Other municipalities in Eure-et-Loir		1,733	1,713	1,870
Other municipalities in Indre and Indre-et-Loire		1,314	1,442	1,536
Other municipalities in Languedoc-Roussillon	1,569	1,611	1,917	1,763
Other municipalities in Limousin		1,953		2,034
Other municipalities in Loir-et-Cher		1,879		1,785
Other municipalities in Loiret		1,493	1,403	1,574
Other municipalities in Cher		1,362		1,283
Other municipalities in Nord	1,608	1,604	1,580	1,542
Other municipalities in Pas-de-Calais		1,625	1,660	1,768
Other municipalities in Pays de la Loire North	2,904	2,731	3,012	2,942
Other municipalities in Rhône-Alpes West	1,883	2,031	1,936	2,017

* 2007-2008 stock

The associated indices do not have the Notaires-INSEE certification and are not published.

Strata in the Paris Region

In the Paris Region, strata were determined by principal component analysis of the average property price by municipality. The result was that the strata were not constructed according to the geographical proximity criterion and could not be identified by name as was the case in the Provinces.

In the table below, we show the distribution by *département* of the number of transactions by stratum (transactions in reference stock 2007 -2008) and by dwelling type.

Table A2.13 – Paris Region: number of strata by *département* (excl. Paris) and corresponding numbers in the reference stock*; apartments

Apartments*								
Département	77	78	91	92	93	94	95	Total
Stratum								
1	2,089	434	2,086	-	-	126	986	5,721
2	2,229	158	73	-	-	162	314	2,936
3	1,217	475	1,119	-	1,204	-	1,546	5,561
4	523	1,532	1,728	-	2,348	2,863	1,895	10,889
5	1,612	3,628	3,927	-	846	831	2,021	12,865
6	-	1,596	-	5,308	4,319	6,102	2,082	19,407
7	-	2,172	-	1,527	-	-	-	3,699
8	271	3,636	1,781	3,147	-	31	326	9,192
9	-	-	-	8,528	-	3,238	-	11,766
10	-	-	149	7,294	6,150	4,649	-	18,242
Total	7,941	13,631	10,863	25,804	14,867	18,002	9,170	100,278

* 2007-2008 stock

Table A2.14 – Paris Region: number of strata by *département* (incl. Paris) and corresponding numbers in the reference stock*; houses

Houses*									
Stratum	75	77	78	91	92	93	94	95	Total
1	133	234	1,321	259	2,248	-	1,163	1,081	6,439
2	-	-	-	-	690	709	379	-	1,778
3	-	-	984	1,082	109	4,988	2,040	2,623	11,826
4	-	1,603	466	1,959	-	770	829	-	5,627
5	-	2,304	956	828	-	-	484	3,193	7,765
6	-	2,867	3,918	3,009	49	-	332	154	10,329
7	-	6,801	2,110	1,817	-	-	-	1,300	12,028
Total	133	13,809	9,755	8,954	3,096	6,467	5,227	8,351	55,792

How to read this chart: For houses, stratum 1 includes Paris.
* 2007-2008 stock

Strata for apartments in the Paris Region

Table A2.15 - List of 5 strata for apartment indices in Paris

Stratum	Neighbourhood	Number of transactions in reference stock*
7501	1 to 17, 19 to 24, 28, 33 to 39 and 48	9,346
7502	18, 40 to 44, 49, 56, 67 to 72, 77, 79	15,139
7503	45 to 47, 52, 53, 55, 57 to 60, 65, 76, 80	10,689
7504	50, 51, 54, 73 to 75, 78	4,635
7505	25 to 27, 29 to 32, 61 to 64, 66	6,002
Total	80 neighbourhoods	45,811 transactions

* 2007-2008 reference stock
Five strata were constructed from the 80 neighbourhoods in Paris.

Table A2.16 – List of 10 strata for apartment indices in Paris Region, excl. Paris

Stratum	Number of municipalities	Number of transactions in reference stock*
1	790	5,721
2	87	2,936
3	23	5,561
4	41	10,889
5	184	12,865
6	51	19,407
7	3	3,699
8	61	9,192
9	13	11,766
10	27	18,242
Total	1,280 municipalities	100,278 transactions

* 2007-2008 reference stock

Ten strata were constructed for apartments from the 1,280 municipalities in the Paris Region, excl. Paris. These indices do not have the Notaires-INSEE designation and are not published.

Strata for houses in the Paris Region

Table A2.17 – List of 7 strata for house indices in Paris Region

Stratum	Number of municipalities	Number of transactions in reference stock*
1	64	6,439
2	23	1,778
3	81	11,826
4	56	5,627
5	98	7,765
6	286	10,329
7	673	12,028
Total	1,281 municipalities	55,792 transactions

* 2007-2008 reference stock

Seven strata were constructed for houses in the Paris Region, from the 1,281 municipalities. Paris is included in stratum 1 because the volume of house sales is low. These indices do not have the Notaires-INSEE designation and are not published.

Appendix 3: Updating the model specifications

The specifications for the models are reviewed each time the database is revised. Most of the changes in version 3 compared with version 2 relate to modalities for the plot size and living space variables in houses; we also wanted to harmonise as completely as possible the models for the Paris Region and the Provinces.

Changes to processing the surface area in houses

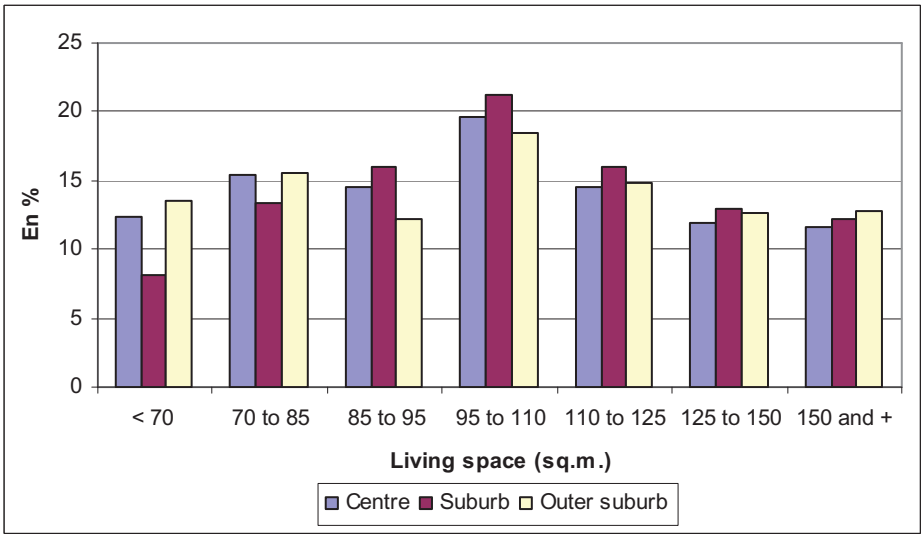
Analysis of surface area variables

The prices that we observe are used to estimate the value of a reference property, which in turn is used to estimate the value of the reference stock at time *t*. In the earlier version of the indices, the reference value for continuous variables was not fixed, and thus had a zero default setting. The estimation of relative prices has not changed. However, it is more correct to choose as a reference property one with a plot size and living space that are not zero.

The question then had to be asked as to whether different reference vales should be used for central cities, suburbs and rural municipalities. We therefore compared the respective distributions according to municipality type. In the end, we used the same values for the Paris Region and the Provinces.

For the Provinces, the modal class of living space in a dwelling is the same (Figure A3.1) whatever the municipality type [class 95 to 110 sq.m.]. There is not a great deal of difference in the three series, although we note that municipalities in the suburbs are less well represented in the classes for small living space (less than 85 sq.m.).

Figure A3.1 - Distribution of transactions by living space, houses in the Provinces

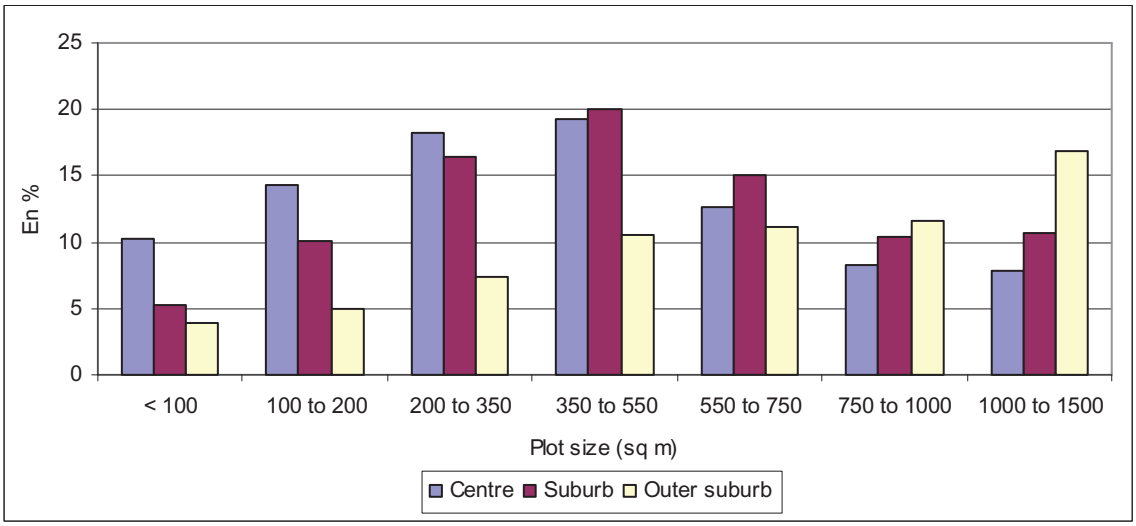


Source: PERVAL database, transactions observed between 2004 and 2008
How to read this chart: Houses with 95 to 110 sq.m. represent more than 20% of all houses located in suburbs.

Differences are more pronounced when considering plot size (Figure A3.2). In outer-suburban municipalities, plots are larger: 50% of houses in outer-suburban municipalities have plots of over 1,000 sq.m. compared with only 23% of suburban municipalities and 17% of central city municipalities. This is reflected in the value of the modal classes for the series: for outer-suburban municipalities the interval is 1000 to 1500 sq.m. compared with a smaller interval class of 350 to 550 sq.m. for the other two types of municipality.

However, it is not possible in all cases to take different reference modalities according to municipality type, as some strata contain both central city municipalities and outer-suburban municipalities.

Figure A3.2 - Distribution of transactions by plot size, houses in the Provinces

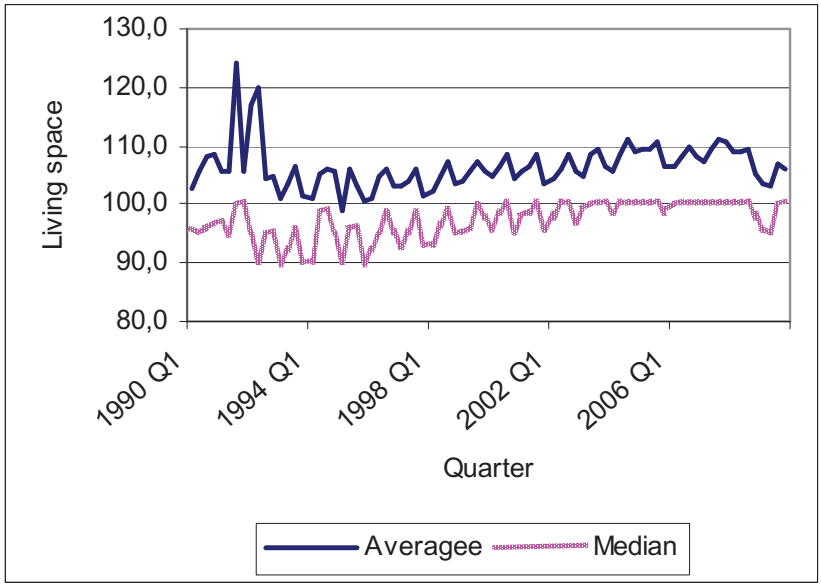


Source: PERVAL database, transactions observed between 2004 and 2008
How to read this chart: Houses with a plot size of 350 to 550 sq.m. represent almost 20% of all houses located in suburbs.

With regards to changes over time, if we take the example of the Paris region, the quarterly living space average has an irregular profile in the period from 1990 Q1 to 2009 Q4. The median value, which is less sensitive to extreme values, is more stable, especially from 2003 onwards. Considerable seasonal variation can be seen (Figure A3.3).

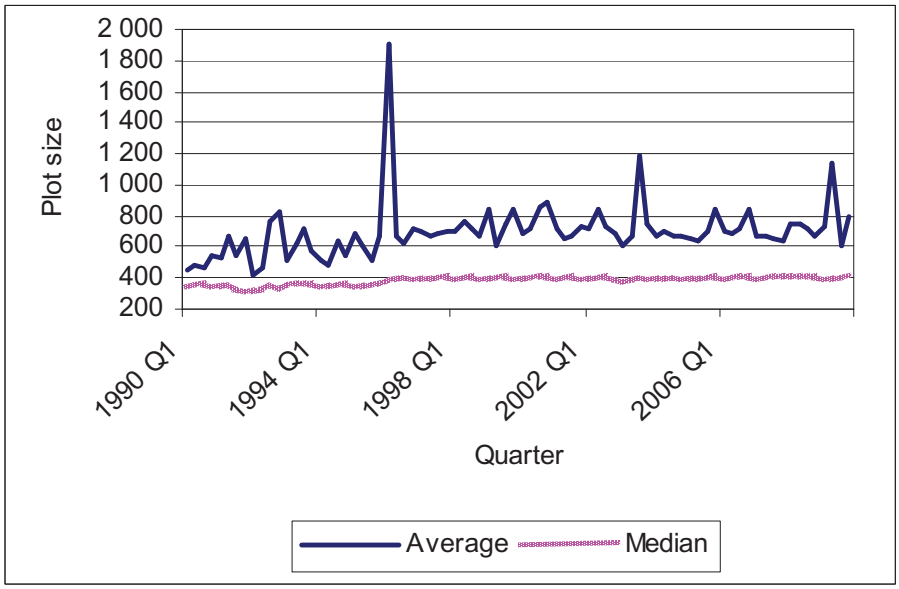
With regards to plot size, the median fluctuates little, while the average is much more volatile (Figure A3.4).

Figure A3.3 - Living space, median and average values, houses in Paris Region



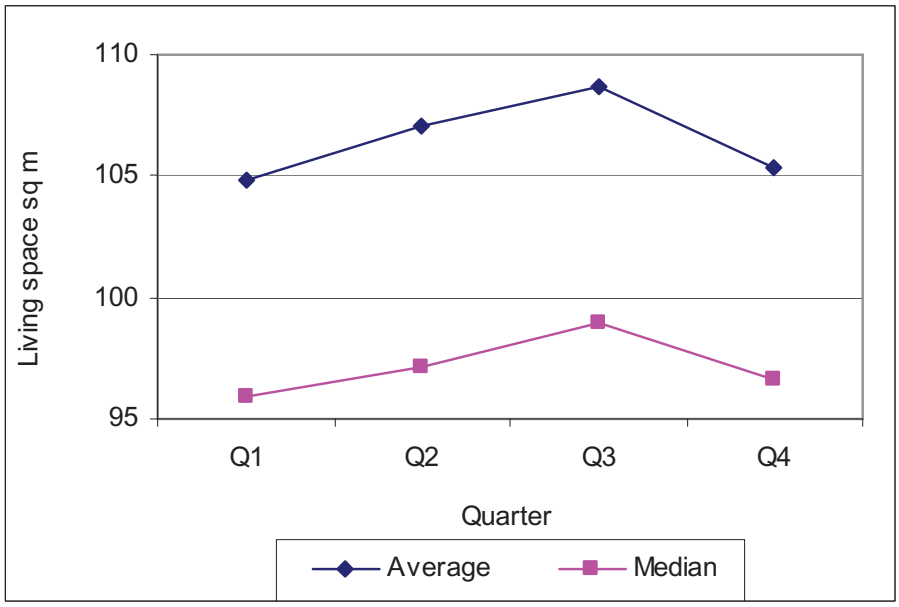
Source: BIEN database, transactions observed between 1990 Q1 and 2009 Q4

Figure A3.4 - Plot size, median and average values, houses in Paris Region



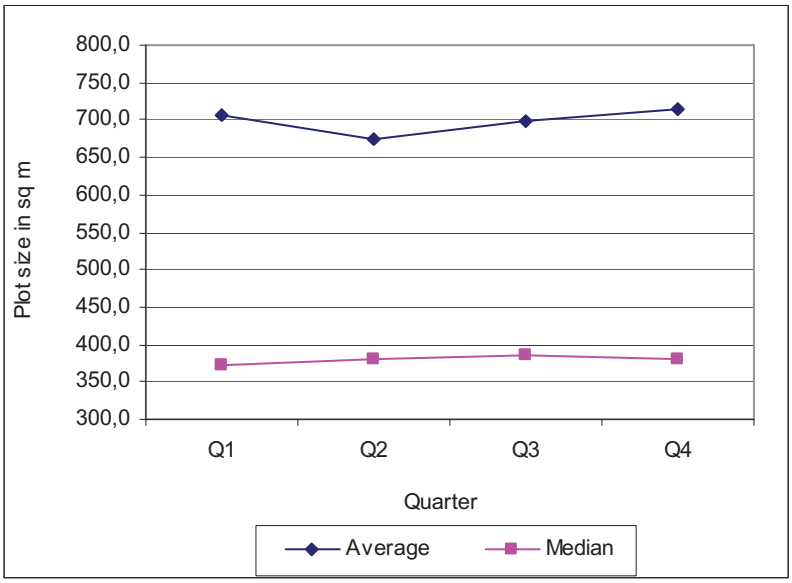
Source: BIEN database, transactions observed between 1990 Q1 and 2009 Q4
Comment: the change in the series is affected by the addition of the area of the outer Paris suburbs from 1996

Figure A3.5 - Living space, median and average values, houses, Paris Region by quarter



Source: BIEN database, transactions observed by quarter (between 1990 Q1 and 2009 Q4)

Figure A3.6 - Plot size, median and average values, houses in Paris Region by quarter



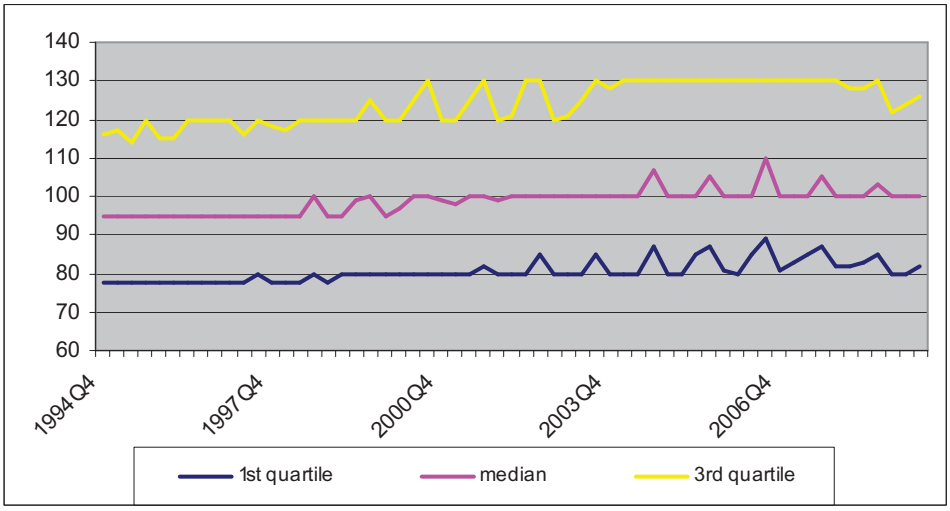
Source: BIEN database, transactions observed by quarter (between 1990 Q1 and 2009 Q4)

Choice of reference values

The volatility of plot size is less than that of living space. There is little seasonal variation in the median value.

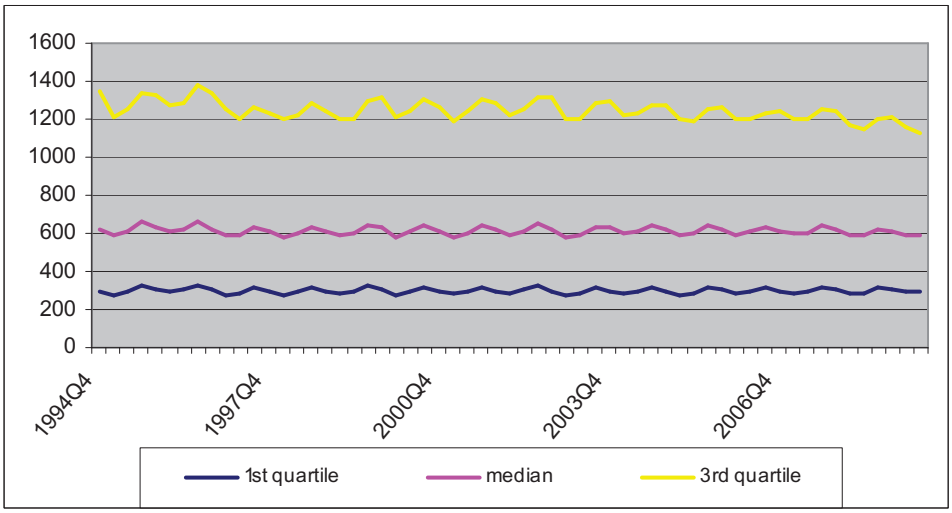
Figures A3.7 and A3.8 show the change in the median and the two extreme deciles for the two variables for the Provinces.

Figure A3.7 - Living space in houses in the Provinces for three quantiles



Source: PERVAL database, transactions observed between 1994 Q4 and 2009 Q2
How to read this chart: the median of living space in houses was relatively stable over the last ten years.

Figure A3.8 - Plot size in houses in the Provinces for three quantiles



Source: PERVAL database, transactions observed between 1994 Q4 and 2009 Q2

Taking living space into account in the models

The reference house selected has living space measuring 100 sq.m., and plot size 610 sq.m. These values are the same for the Paris Region and the Provinces.

We also need to know if we must introduce a transformation of the living space variables, for example by using their logarithm or breaking them down into segments. The choice depends on our idea of the link between the price of the property and living space (or plot size). If we think that the link is linear, then we opt for logarithms. The hypothesis that we then make is that price-elasticity is constant. A log-log model has an advantage in that we are able to read elasticities directly. If, on the other hand, we think that the link between price and plot size or living space is not linear, then a solution is to divide the variables into classes. In addition, this discretisation makes it easier to deal with discontinuity in the living space. Living space in houses is often rounded up or down (to the nearest 5 or 10 square metres).[63]

In the previous version of the indices, plot size and living space were expressed as raw data and not as logarithms (we hypothesised that house prices were an exponential function of $\beta_{k,\tau}$ coefficients of the hedonic model, instead of being a linear function). When determining house prices, we therefore tended to assign rather too little influence to these two variables compared with the other variables. This resulted in an adjustment coefficient close to zero, which in turn increased the seasonal variation in the series of price indices artificially.

In order to simplify this revision, the choice was made to use logarithms. Bracketing categories requires many steps, which can be carried out during the next revision: finding the optimal number of categories (5 to 7 normally[64]), determining their values; differentiating them, where necessary, according to strata; testing the hypothesis of price elasticity stability for each stratum in relation to living space.

Impact of change in the method on the series of indices

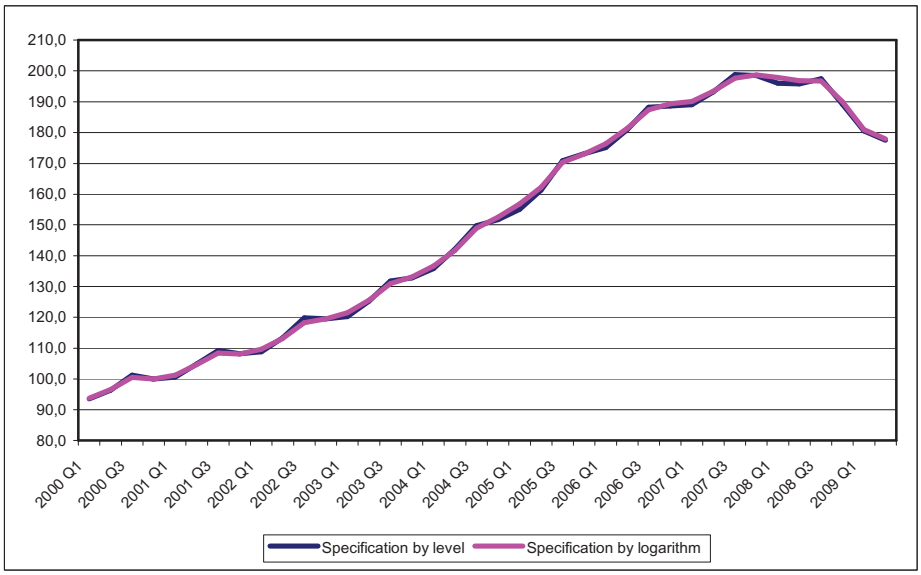
We also studied the impact of the new method on determining coefficients of the characteristics vector and on the price index series. Here, we present results for the Provinces.

The old method gives a more volatile series of results than the new (Figure A3.9). Seasonal variation in the indices is reduced using the new method. Divergences between the indices are fairly low, less than about 2% in absolute value.

[63] Plot sizes are measured more carefully, however, because they correspond to the cadastral surface areas provided in the notarial deeds.

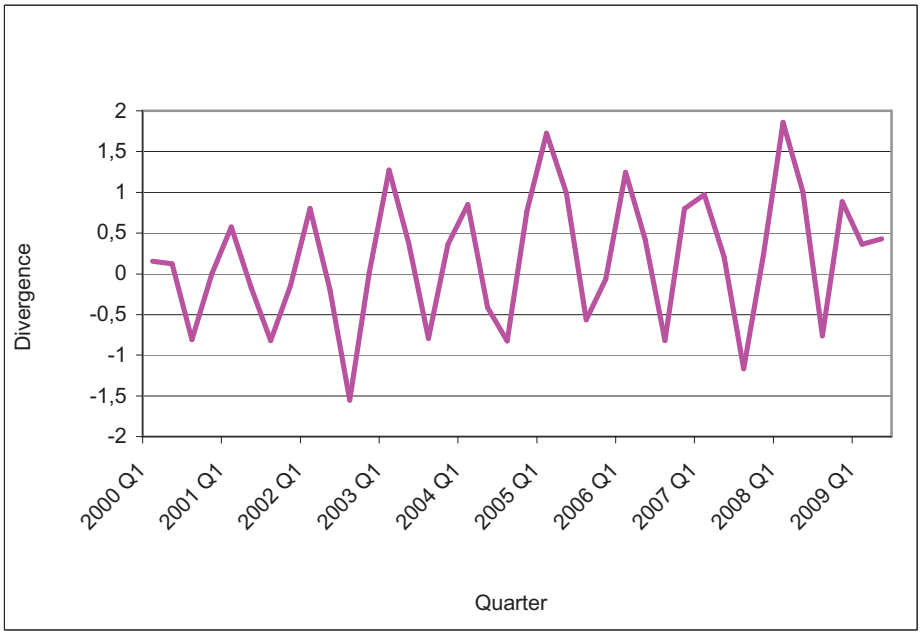
[64] There should not be too many groupings, otherwise we tend to revert to a single continuous case. This also increases the number of coefficients to be estimated.

Figure A3.9 - House price indexes for the Provinces



Source: PERVAL database, transactions observed between 2000 Q1 and 2009 Q3
How to read this chart: the series with surface area variables given by level shows a more pronounced seasonal variation in its profile than the series with the surface area expressed as logarithm.

Figure A3.10 - Divergences in indexes between the old method and the new method



Source: PERVAL database, transactions observed between 2000 Q1 and 2009 Q3
How to read this chart: the divergence corresponds to the "old method – surface area variables by level" index minus the "new method - surface area variables as logarithm" index.

Specification of the new models

The dependent variables have not been modified: for houses, the total price is used whereas for apartments, we take the price per sq.m. For the Provinces, we keep the same explanatory variables as in version 2. There is one change: in version 2, the number of rooms and the neighbourhood[65] were introduced separately, whereas in version 3, they have instead been cross-tabulated.

[65] In the general sense of an area included in a stratum: it can be the neighborhoods in the cities but also larger groupings outside the cities.

The models for the Paris Region and the Provinces cannot be exactly alike, because some explanatory variables are not found in both notarial databases (e.g. the variable to codify the general condition of the property is only present in the PERVAL database).

See the tables below for the list of variables introduced into the models. All qualitative variables have been discretised.

Table A3.1 – List of explanatory variables in models for apartments

Explanatory variables	Present in Paris Region model	Present in Provinces model
Year of transaction	Yes	Yes
Month of transaction	Yes	Yes
Construction period	Yes	Yes
Number of bathrooms	Yes	Yes
Cellar	Yes	Yes
Number of garages	Yes	Yes
Floor*	Yes	Yes
Lift*	Yes	Yes
Cross-tabulation lift and number of floors*	Yes	Yes
Cross-tabulation number of rooms and neighbourhood	Yes	Yes
Floor space per room**	Yes	Yes
Condition of property	No	Yes
Terrace or balcony	No	Yes

* For apartments located on upper floors (4th floor and higher), we crossed the variables "number of floors" and "presence of lift". Below this, we believe that the variable "lift" has no effect on prices and we used only the variable "floor".
** Logarithm

Table A3.2 – List of explanatory variables in models for houses

Explanatory variables	Present in Paris Region model	Present in Provinces model
Living space*	Yes	Yes
Plot size*	Yes	Yes
Year of transaction	Yes	Yes
Month of transaction	Yes	Yes
Construction period	Yes	Yes
Number of bathrooms	Yes	Yes
Basement	No	Yes
Number of garages	Yes	Yes
Number of storeys	Yes	Yes
Condition of property	No	Yes
Cellar	Yes	No
Cross-tabulation number of rooms and neighbourhood	Yes	Yes

* Logarithm

Appendix 4: Stability tests and duration of reference period

The relative prices of property characteristics, estimated from the hedonic models, will be used to calculate price indices with constant characteristics. Hedonic models are estimated from transactions over a given period, called the "estimation period".[66] As transactions are relatively infrequent but relative prices change over time, the estimation period selected must be long enough to provide sufficient observations but not so long that the model coefficients can be considered as constants. Stability test are carried out to determine the optimal duration that should be selected (in number of years). We must ensure that in statistical terms the $\hat{\beta}_k$ coefficients of the characteristics can be considered as constant over the estimation period. The time effect is captured by the time dummies in the econometric equation.

Stability tests in version 2

In version 2 the estimation period was sixteen quarters (1998 to 2001). We tested the stability of the $\hat{\beta}_k$ coefficients across the estimation period by verifying that the gap between the value of properties with the characteristic X_k estimated by the model, and their actual selling price (residual ε_i) meets the model's stochastic hypotheses, and in particular that it does not contain any overlooked deterministic trend.

Let $\overline{\varepsilon}_{t,1}$ be the mean of the residuals measured for all dwellings in quarter t that have characteristic X_1 . We build the sequence of mean residuals $\overline{\varepsilon}_{1,1}$, $\overline{\varepsilon}_{2,1}$,...., $\overline{\varepsilon}_{16,1}$ for the sixteen quarters in the estimation period. Bearing in mind that the residuals have a zero mathematical mean, and diverge from 0, we were able to conclude that overall the coefficients were stable across the period 1998 to 2001.

Stability tests in version 3

The sharp rise in prices observed in the 2000s which was then interrupted in 2008 mean that it is particularly crucial to consider this question of model stability. A stability test in the form of a Chow test was applied to each econometric equation. In addition, to ensure that conditions were valid for the Chow test, in a second step we estimated the models using the quasi-generalised least squares estimator, to guard against any distribution problems that might arise due to the existence of heteroscedasticity.

The Chow test is carried out on the sum of the squared residuals from the second step in the estimation. This is a traditional analysis of variance test, in the form of a Fisher's test. The scope of the analysis is observations from the estimation stock. Observations are divided into two groups of years and two sub-models: the years 2004-2005, the "years at the beginning of the period" group and years 2006-2007, the "years at the end of the period" group. Tests showed that the coefficients estimated across the two periods were significantly different. An estimation was applied with variable selection but did not improve the results.

Table A4.1 shows an example of the analysis results for houses and apartments in the Provinces. It gives the number and the weight of the strata where the coefficient stability hypothesis was rejected. We could have gone on to differentiate the duration of the reference stock by strata according to the test results: for example, four years for the models where the stability hypothesis was accepted and two years for the rest. However, for practical reasons –in particular, different methods were needed for chain-linking the price indices– we decided to retain a single estimation duration in the model for all strata. The proportion of strata where the zero hypothesis was rejected increased naturally with the type 1 error value. For the most frequently uses error level (α=0.05), this proportion was high for houses (46.0%). It was even more so for apartments (for α=0.01, this share reached 80.4% and rose to 85.6% for α=0.05). Table A4.2 shows that correcting the heteroscedasticity does not alter the results: the model stability hypothesis cannot be sustained, either for houses or apartments, across the period 2004-2007. The relative prices of the characteristics are not constant over time. The hypothesis of hedonic model stability over a four-year period cannot be sustained. The same result was obtained for the Paris Region.

[66] See definitions in Chapter 3.

Table A4.1 – Number of strata where stability hypothesis is rejected at threshold α,[67] OLS estimator

Significance level	Houses		Apartments	
Level of risk	Number of strata	%	Number of strata	%
α=0.001	27	15.5	60	61.9
α=0.01	52	29.9	78	80.4
α=0.05	80	46.0	83	85.6

Source: PERVAL database
Number of strata for houses, 174; for apartments, 97
H0: the model is stable over time
H1: the model is not stable (at least one coefficient is significantly modified)

Table A4.2 – Number of strata where stability hypothesis is rejected at threshold α; QGLS estimator

	Houses		Apartments	
Significance level	Number of strata	%	Number of strata	%
α=0.001	20	11.5	58	59.8
α=0.01	44	25.3	72	74.2
α=0.05	81	46.6	86	88.7

Source: PERVAL database
Number of strata for houses, 174; for apartments, 97
With the OLS estimator a majority of hedonic models do not satisfy the stability hypothesis. However, the reliability of the tests may need to be treated with caution when heteroscedasticity is present, in which case the OLS estimator is not effective and the usual tests no longer work. Graphs were produced linking the squared residuals with the predicted values for the most influential variables but were not able to determine which variables were associated with the heteroscedasticity. We therefore calculated the quasi-generalised least squares (QGLS) estimator, by hypothesising a different variance across each period. This method was used when the shape of the variance-covariance matrix was unknown. We applied this adjustment when estimating the hedonic function in the second step, i.e. after removing the atypical points discovered during the first step. The aim was to find a good estimator of the variances-covariances of the risks.

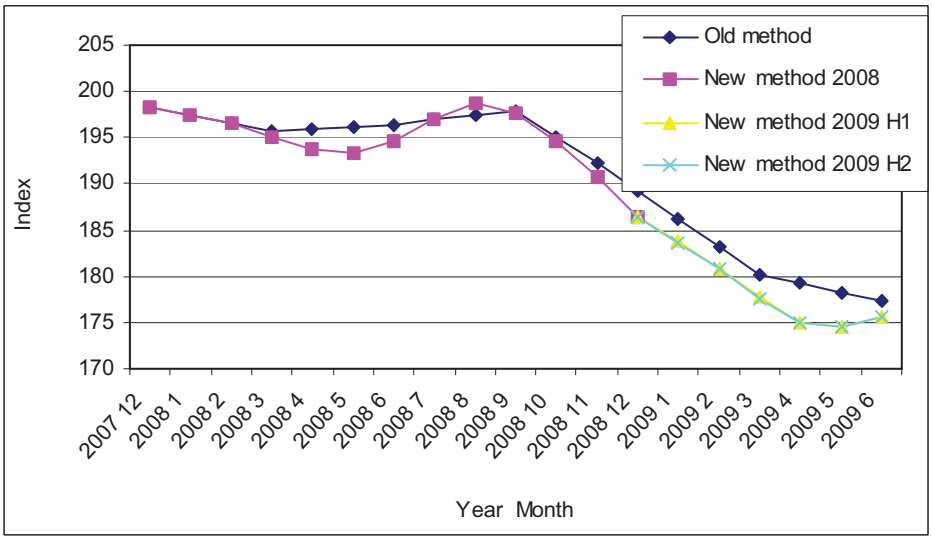
Choice of reference periods and updating the vector of relative prices of dwelling characteristics

Thus in version 3, the estimation period was reduced to two years, whereas it had been four years in version 2. This reduction was also applied to the reference stock. The definition of the strata and the specification for the econometric equations will remain the same for the lifetime of version 3 but the β_k coefficients will be updated every two years. Note that we could have decided to update them every year. These decisions were based on a comparison for 2008 and 2009 of quarterly price indices calculated using different methods relating to the length of the estimation period and the frequency with which coefficients are updated (Figure A4.1).

All three graphs have almost the same pattern of change. However, the shape of the series calculated using a four-year estimation stock ("old method") is less uneven than the other two series, which in fact are combined for 2008 as they were calculated from the same estimation stock (that of 2006-2007). Divergences between the series calculated with the four-year estimation stock and those based on the two-year stock are relatively large, ranging from 1.5 points to -4 points. The graph showing the "old method" series is always higher than the other two except for the 2nd and 3rd quarters of 2008. For the series calculated using the two-year estimation stock, the calculation method made little difference: divergences were zero in 2008 and ranged from -0.0012 to 0.1317 percentage points in 2009.

[67] A type 1 error (error of the first kind) is the probability that the hypothesis of coefficient stability be incorrectly rejected.

Figure A4.1 - Index of used houses for the Provinces

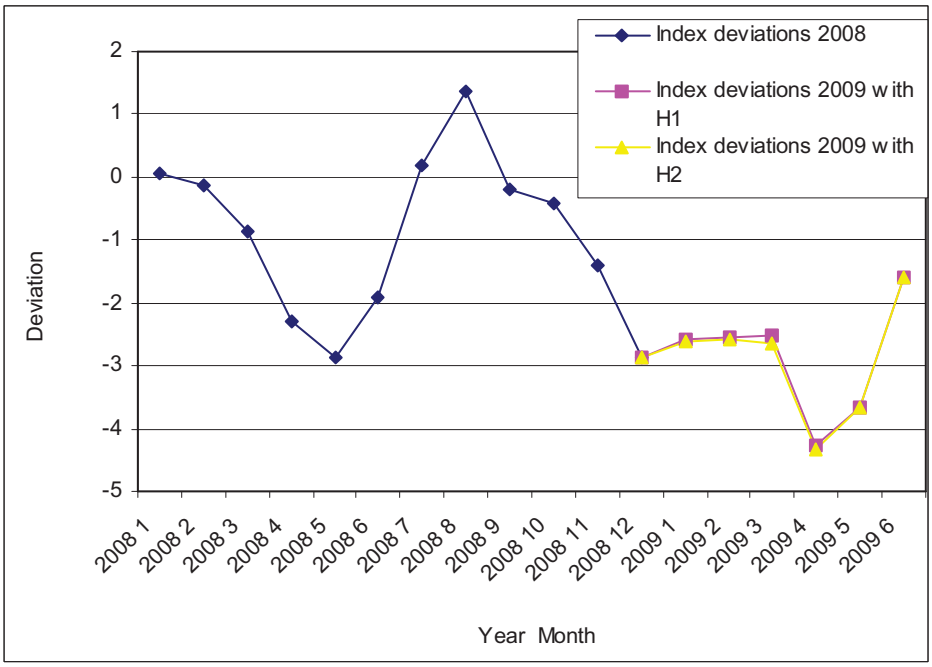


Source: PERVAL database

How to read this chart:

- old method = 4-year estimation stock;
- new method H1: 2-year estimation stock without updating β_k across the period); the β_k coefficients for 2008 and 2009 are estimated from 2006-2007;
- new method H2: 2-year estimation stock, with annual update of β_k.; the β_k coefficients for 2008 are estimated from 2006-2007 and those for 2009, from 2007-2008;

Figure A4.2 - Index of used houses for the Provinces (index deviations)



Source: PERVAL database

How to read this chart: the "new methods" series are identical in 2008 (we use the same set of coefficients) and different in 2009 (we have two different sets).

Table A4.3 - Examples of Chow test for strata for houses in the Provinces

	Total model		Model 1: 2004 and 2005		Model 2: 2006 and 2007			
Stratum code	degree of freedom	sum of squared residuals	degree of freedom	sum of squared residuals	degree of freedom	sum of squared residuals	Chow test statistic	Pvalue
751	1,915	114.8	874	54.2	1,007	56.9	1.82	0.0028
752	2,690	142.4	1,325	76.3	1,326	61.7	2.23	0.0000
754	3,136	211.2	1,477	105.3	1,620	102.3	1.38	0.0581
755	7,182	412.1	3,456	204.1	3,677	202.2	2.09	0.0000
756	4,656	288.1	2,289	142.6	2,313	140.0	1.66	0.0018
757	8,938	428.0	4,574	225.8	4,320	198.4	1.80	0.0009
758	4,055	211.7	1,820	100.4	2,186	108.6	1.07	0.3401
6701	7,477	906.8	3,601	427.4	3,817	464.5	2.11	0.0000
13055	1,676	123.9	824	66.0	818	53.7	1.67	0.0092
21701	2,418	75.3	1,262	41.7	1,122	32.4	1.06	0.3742
29701	2,825	122.2	1,430	63.2	1,351	56.4	1.36	0.0573
31555	1,556	83.5	774	42.8	743	38.3	1.18	0.2137
31701	5,032	150.5	2,614	82.4	2,379	66.4	1.50	0.0245
33063	2,742	174.6	1,346	80.3	1,352	87.4	2.53	0.0000
33701	8,568	382.9	4,379	205.9	4,125	172.8	1.48	0.0080
34701	1,734	81.3	780	39.4	920	39.8	1.32	0.1061
35701	2,257	93.6	1,154	46.7	1,069	43.7	2.32	0.0000
37701	3,801	146.3	1,756	70.9	2,001	73.5	1.12	0.2656
38701	1,920	94.0	968	49.8	918	42.0	1.33	0.0977
44109	2,617	111.9	1,245	54.9	1,333	55.7	0.76	0.8619
44601	2,324	211.3	1,264	120.0	1,026	85.7	1.85	0.0021
44701	5,774	178.3	2,772	89.9	2,943	86.2	1.25	0.0982
45701	4,853	134.6	2,501	72.2	2,303	60.8	1.17	0.1984
51701	2,245	80.7	1,115	41.8	1,096	36.0	2.34	0.0000
54701	4,770	229.5	2,190	105.6	2,536	121.4	1.19	0.1790
59350	1,874	95.1	941	48.9	889	42.1	1.90	0.0004
59502	2,049	184.8	1,020	90.1	995	91.1	1.17	0.2332
59599	2,198	91.0	1,083	47.9	1,091	42.0	1.07	0.3675
59601	2,133	90.0	1,115	49.6	979	38.5	1.16	0.2291
59701	3,622	321.8	1,894	176.3	1,684	140.9	1.17	0.2019
59702	15,671	652.9	7,985	348.2	7,582	294.6	2.35	0.0000
62601	2,537	130.2	1,317	72.0	1,191	55.6	1.76	0.0072
63701	2,105	126.3	1,070	66.5	1,001	55.9	1.91	0.0013

How to read this chart: In the Pvalue column, shaded boxes indicate strata where the stability hypothesis H0: $\beta^1 = \beta^2 = \beta$, is rejected for a significance level $\alpha=0.05$.
Source: PERVAL database

Table A4.4 - Examples of Chow test for strata after adjusting for heteroscedasticity for houses in the Provinces

	Total model		Model 1: 2004 and 2005		Model 2: 2006 and 2007			
Stratum code	degree of freedom	sum of squared residuals	degree of freedom	sum of squared residuals	degree of freedom	sum of squared residuals	Chow test statistic	Pvalue
751	1,921	462.5	865	209.0	1,014	238.3	1.52	0.0186
752	2,687	608.3	1,316	310.9	1,324	278.4	1.81	0.0007
754	3,150	812.7	1,471	393.5	1,632	402.7	1.37	0.0478
755	7,180	1,699.6	3,452	826.1	3,671	846.4	2.03	0.0000
756	4,653	1,143.1	2,281	560.9	2,310	557.9	1.61	0.0018
757	9,047	1,944.6	4,567	997.2	4,428	926.5	1.88	0.0001
758	4,064	919.9	1,814	415.3	2,193	490.9	1.06	0.3476
6701	7,474	2,607.6	3,588	1 235.4	3,819	1,340.0	1.38	0.0216
13055	1,700	465.3	817	229.1	841	218.7	1.55	0.0145
21701	2,408	422.3	1,252	227.8	1,114	185.3	1.25	0.1307
29701	2,882	593.9	1,423	293.7	1,407	284.9	1.45	0.0205
31555	1,634	419.7	810	212.2	792	198.1	1.15	0.2649
31701	5,028	865.1	2,607	461.2	2,374	392.3	1.44	0.0269
33063	2,734	693.5	1,339	332.5	1,343	335.6	1.95	0.0001
33701	8,561	1,782.7	4,367	926.2	4,122	834.3	1.49	0.0049
34701	1,734	384.0	774	176.4	918	197.7	1.07	0.3569
35701	2,270	462.2	1,151	231.6	1,077	215.3	1.81	0.0012
37701	3,798	737.0	1,750	347.8	1,996	379.3	0.97	0.5372
38701	1,914	431.9	960	222.5	912	198.4	1.16	0.2183
44109	2,611	538.1	1,239	262.0	1,325	267.4	0.90	0.6630
44601	2,326	698.5	1,257	383.6	1,027	294.5	1.64	0.0061
44701	5,775	1,006.5	2,771	498.0	2,937	495.4	1.12	0.2280
45701	4,850	800.1	2,497	422.9	2,296	366.6	1.13	0.2318
51701	2,238	418.5	1,106	213.5	1,090	192.0	1.68	0.0042
54701	4,789	1,044.2	2,185	477.8	2,552	553.2	1.17	0.1855
59350	1,868	417.4	930	208.7	886	189.9	1.65	0.0027
59502	2,039	610.0	1,014	302.4	983	293.6	1.11	0.2883
59599	2,190	458.6	1,074	228.2	1,084	223.7	1.00	0.4673
59601	2,126	437.4	1,107	232.5	972	193.7	1.16	0.2139
59701	3,616	1,074.0	1,885	571.3	1,679	487.5	0.98	0.5089
59702	15,683	3,148.4	7,976	1 634.3	7,595	1,467.6	2.08	0.0000
62601	2,541	576.5	1,311	304.4	1,193	259.6	1.51	0.0259
63701	2,097	505.9	1,061	260.8	994	230.9	1.41	0.0429

How to read this chart: In the Pvalue column, the shaded boxes indicate strata where the stability hypothesis is rejected H0: $\beta^1 = \beta^2 = \beta$, for a significance level $\alpha=0.05$.
Source: PERVAL database

Appendix 5: Alternative calculation methods

We have described in this volume the method for constructing a hedonic price index and implementing it to produce the Notaires-INSEE indices. There are several steps in this method:

- definition of strata, where price changes are assumed to be homogeneous;
- introduction of adjustments for quality effects, ar stratum by stratum;
- estimation of adjustment effects from an estimation stock;
- calculation of price changes by stratum from total transactions;
- calculation of the index by observing the change in value of a reference stock;
- regular publication of indices and sub-indices.

A procedure like this must be systematic and integrated. In this chapter, we examine alternative methods used in France and in other countries too, to determine whether they are theoretically better, or if they are preferred for practical reasons (budgetary constraints, data availability, different publication requirements), or because of differences in household behaviour (greater mobility leading to repeat sales). In the last section, we describe procedures used by other bodies in France, and in several other countries, with an emphasis on data-collection issues and the bodies responsible for calculating the indices.

If we consider the range of methods used internationally to compile housing price indices, we find some that are relatively simple and others that are more sophisticated. For practical reasons related to cost or for legal reasons, many housing price statistics use average or median prices to track changes in the housing market. We have explained why these are biased (Chapter 2, paragraph 2.1.2).

The most sophisticated methods usually use the hedonic approach or the repeat-sales approach. The former is an econometric approach; it uses regressions incorporating quality effects and a time effect, and assimilates this time effect to a pure constant-quality price effect. The second approach aims to eliminate quality effects by retaining only data on successive sales, or repeat-sales. We shall now look briefly at their underlying principles and hypotheses and discuss some of their variants.

Econometric (hedonic) approach: interpreting time coefficients in a regression as a pure constant-quality price effect

Let us assume that we have several instantaneous cross-sections of samples of real-estate transactions, which give us a measurement for the price and dwelling/building/neighbourhood characteristics and which we can use to estimate a hedonic model. The model defines the price level as a function of the dwelling characteristics. The coefficients of these characteristics are assumed to be stable over time, and a time effect v_t is introduced into the constant term.

The model is written:

$$\log p_{i,t} = a + \sum_{k=1}^{n} \beta_k X_{k,i,t} + v_t + e_{i,t}$$

where the errors $e_{i,t}$ are assumed to be independent, identically distributed and zero-mean. Note that in order to simplify the discussion, the stratum effect is not introduced. To make the time effect identifiable, we assume $v_{t_0} = 0$ for a given date, which then becomes the base period. The change in the index between t_0 and t is equated with the value v_t and the change in the index between *t-1* and *t* is equated with $v_t - v_{t-1}$.

The hypotheses underlying this specification are as follows:

- The variables selected to characterise a dwelling are incorporated in additive form (recall that this is after a transformation of initial characteristics, if needed);
- The relative prices β_k of the characteristics X_k are time-independent, which means that these variables are assumed to have no cross-effects over time;

- After adjustment for effects of characteristics, price variabilities are constant (absence of heteroscedasticity);
- Transactions introduced as observations are representative of the total set of properties for which we want to calculate the index, after quality-effect adjustment.

Some of these aspects, such as the time-dependence of the parameters β_k , are classic subjects for study in the literature. While the long-term constancy of these coefficients is usually rejected, which explains the need for regularly revising the index, it is generally accepted for shorter periods of around 4 to 5 years. There is a simple way of verifying this stability over time. If, for example, the coefficient of X_1 depends on date *t*, the initial model will be seen as missppecified, as the cross variables $1_\tau(t)X_{1,i,t}$, with τ varying, have been omitted. These variables are orthogonal to one another. We need only record the empirical correlation between the omitted variable $1_\tau(t)X_{1,i,t}$ and the estimation residual $\hat{e}_{i,t}$ as a function of date τ. If these correlations are close to zero, the β_1 coefficient is regarded as stable over time. Otherwise, the form of the change in this correlation as a function of τ gives information on the type of change in the coefficient.

One example of this method is the method over "adjacent periods". We consider a model with time dummies based on two consecutive periods. At each period *t* the model is reestimated. The time dummy measures price change with constant characteristics between dates *t-1* and *t*. the index between two dates is obtained by chaining one to another the indices of the periods between these two dates.

The representativeness of the transaction samples at each date *t* relative to what is theoretically desirable for an index is an issue that is less often discussed. For example, if the sample at date *t* includes transactions of which the prices are systematically higher than the theoretical price for the overall population (i.e. the price of a fixed reference stock and not of transactions that vary from one date to another), the time coefficient $\hat{v}_t$ estimated from data will reflect both the theoretical price level and the bias due to the non-representativeness of the sample, although it will not be possible to distinguish these effects easily (problem of identifiability). However, there are two approaches that we can use to detect any non-representativeness.

- The first was suggested by Griliches (1971, p.7-8) in a different setting (price index for cars). It consists of distinguishing repeat sales in samples for dates *t-1* and *t* for example. We can then calculate the mean residuals for these repeat sales on date *t-1* and date *t* and see if they are close to zero. However, the aim of this approach is to verify the representativeness of the repeat-sales sub-sample rather than that of the complete samples. It is also difficult to implement in our context, where the number of repeat-sales is small.
- Another approach relies on the assumed change in the index. Given a set of observations $\hat{v}_1, \ldots, \hat{v}_T$ of the index between 1 and *T*, we can construct a dynamic model which will supply a forecasting interval:

$$\left[\tilde{v}_{T+1}, \tilde{\tilde{v}}_{T+1}\right]$$

for the future value v_{T+1} . If the estimate for date $T+1$ does not lie within this interval, we can either assume that the sample for date $T+1$ is unrepresentative, or we can look for a structural cause for this sudden price change.

Repeat-sales method

In 1943, Gaston Duon, who worked for the French National Statistics Service, the forerunner of INSEE, applied what was called the repeat-sales method to compile price indices for dwellings in Paris between 1790 and 1944 (Duon, 1943 and 1946). In 1956, Léo Grebler, who was apparently unaware of the work being done by Gaston Duon, then applied this method to compile a housing price index in 22 cities of the United States.[68] Both Duon and Grebler were well aware that the quality of a given property changes over time as a result of the effects of two conflicting phenomena, on the one hand the general obsolescence of buildings and on the other, the improvements made over the years. They both compiled two series of indices, before and after adjustment for

[68] Historically, housing price indices usually use the repeat-sales method. This was the method used by Piet Eicholtz to construct the Herengracht index. The exceptions are the indices by d'Avenel covering the period 1200-1800 and which are thus, apparently, the property indices that go back farthest in time, but which may contain some uncertain data.

these effects. The deviation in growth for these two series was around 1% per year both for Duon and for Grebler.[69]

The repeat-sales method was then applied, with more powerful computing resources, by Bailey, Muth and Nourse (1963). The authors do not directly pose the question of the fixed basket of goods, but they do make two comments. First, "because of the variation in the quality of sold dwellings from one period to another, the mean price varies more than the price of each property taken individually"; second, a gradual change in the quality of traded dwellings over time biases the change in average prices.

It was difficult to specify a hedonic model with a time dummy, given the frequent lack of data on the characteristics of dwellings. To overcome this difficulty, they suggested using the fact that certain dwellings are sold several times in succession. These repeat-sales data would obviate the need for detailed information on the characteristics of the properties.[70]

The lack of data on dwelling characteristics and the technical difficulty of applying the econometric approach are still today the main reasons given in favour of a repeat-sales approach. This approach is easy to explain when repeat-sales occur on dates *t* -1 and *t*. In this case, we equate the price change to the mean change observed for these repeat sales, to give a model of this type:

$$\log p_{i,t} = \log p_{i,t-1} + b_t + u_{i,t-1,t}, i \in I_{t-1,t}$$

at a given date *t*. $I_{t-1,t}$ denotes total dwellings traded both at *t* -1 and *t*, and b_t is the change that we are looking for.

However, even with high mobility, few repeat sales are so very close together in time. The approach is therefore extended to take account of repeat sales taking place at two dates t_1 and t_2 , $t_1 < t_2$, which can be further apart. The underlying model now becomes:

$$\log p_{i,t_2} = \log p_{i,t_1} + \sum_{t=t_1+1}^{t_2} b_t + u_{i,t_1,t_2}, i \in I_{t_1,t_2}$$

with t_1 and t_2 varying.

It can be rewritten by including the explanatory variables for dates $Z_{i,t} = 1$, if date *t* lies in the period between the trading dates, if not it is $Z_{i,t} = 0$. The model now becomes:

$$\log \frac{p_{i,t_2}}{p_{i,t_1}} = \sum_{t=T_1}^{T_2} b_t Z_{i,t} + u_{i,t_1,t_2}, i \in I_{t_1,t_2}$$

where $[T_1, T_2]$ gives the time interval that is the union of all the intervals $[t_1, t_2]$. In this form, it is a linear model in the parameters of interest b_t, with *t* varying, which gives price changes at the various dates. This model is generally estimated using ordinary least squares.

However, if the time interval between repeat sales covers more than two periods, then overlaps will occur for periods concerning the various dwellings, and hence there may be correlations between error terms, which should be taken into account in the estimation method. To illustrate this potential problem and understand why the quality effects have disappeared, it is worth returning to the hedonic model described in the previous section:

$$\log p_{i,t} = a + \sum \beta_k X_{k,i,t} + v_t + e_{i,t}$$

with $e_{i,t} = u_{i,t} + \eta_i$,

[69] On secular housing price indices, and in particular a comparison of the work of Duon and Grebler, cf. Comparing Four Secular Home Price Indices, J. Friggit, June 2008, http://www.cgedd.developpement-durable.gouv.fr/IMG/doc/house-price-index-Paris-and-others-secular_cle7fed11.doc.

where the error terms $u_{i,t}$, η_i are assumed to be independent of each other, and zero-mean. If all repeat sales taking place in t_1, t_2 are representative of sales in the total set of dwellings (after adjustment for the characteristic effect), we have:

$$\log \frac{p_{i,t_2}}{p_{i,t_1}} = v_{t_2} - v_{t_1} + u_{i,t_2} - u_{i,t_1}, i \in I_{t_1,t_2}$$

We confirm that the repeat sales approach with a least-squares estimation is consistent with $b_t = v_t - v_{t-1}$ and $\sum_{t_1+1}^{t_2} b_t = v_{t_2} - v_{t_1}$, since the error terms $u_{i,t_1,t_2} = u_{i,t_2} - u_{i,t_1}$ are indeed independent, with identical distribution.

Finally, we can make the model more complex and refine it by incorporating the natural depreciation of dwellings (net of improvements made), or other quality variables that may change between two sales (e.g. a comfort feature added to the dwelling).

To sum up, the repeat-sales method is based on the following assumptions:

- relative prices of characteristics are constant over time. This is the same hypothesis as in the standard applications of the hedonic method. But it is undoubtedly less likely to be fulfilled over the relatively long period between two successive sales of the same dwelling;
- there is no selection bias. However, dwellings sold frequently are probably not representative either of total transactions, or of the housing stock. These may be, for example, small dwellings (first homes for young couples), whose prices may move differently from those of larger dwellings; or they may be dwellings that are resold very quickly and which may have unobservable characteristics, related to the vendor, which may account for large capital gains. Clapp *et al*. (1991) find a difference in the short-term change in repeat-sales indices relative to hedonic indices, but this difference disappears in the long term (3 years), which they see as logical: if the market is working, there can be no long-term disequilibrium in relative prices.[71] Given the distortions for periods of less than three years, they recommend using hedonic methods. Case *et a*l. (1997) suggest adjusting for this effect by incorporating information on the link between a dwelling's appreciation rate and transaction frequency;
- the dwelling is effectively the same. Of course renovations or even more substantial alterations (extensions, improvements) are common, with the result that the dwelling is no longer the same. These last two problems can be addressed by combining the hedonic model with repeat-sales data; this will also correct a flaw in the repeat-sales method: the fact that it uses few observations relative to total transactions. The reason is that dwellings change hands fairly infrequently, at least this is the case in France;
- the error terms $u_{i,t}$ are indeed independent, zero-mean and with identical variance. However, this hypothesis is presumably not satisfied. These errors concern prices, which, on average, increase over time, hence generating heteroscedasticity. We also expect a stronger correlation between prices for dates that are close than for dates that are further apart. It therefore seems important to introduce a time correlation between errors $e_{i,t}$, which modifies the method for estimating rates of change b_t.

Note that for France, results from the repeat-sales method,[72] for overlapping periods and areas, do not differ very much from those of the Notaires-INSEE indices.[73]

[71] Mark and Goldberg (1984) for their part, find a persistent long-term difference.

[72] Applied to notarial data over a long period (price comparison, without econometric approach). See Friggit J., "Comparing Four Secular Home Price Indices", CGEDD, June 2008, http://www.cgedd.developpement-durable.gouv.fr/prix-immobilier-evolution-1200-a1048.html.

[73] The repeat-sales method gives poor results for recent years, because of the small number of observations and the selection bias mentioned above for short periods of ownership (capital gains on properties resold quickly exceed the capital gains on the index). The comparison was therefore made after incorporating an adjustment coefficient benchmarked to Paris.

Laspeyres index or chain index

A large amount of classical index theory is devoted to the choice of weightings. Three options are available in general: they can be set once and for all, defined as equal to a set of quantities traded at an initial date (Laspeyres index), or at a terminal date (Paasche index), or else they can be modified at each index calculation date (chain index).

The quantities exchanged are related to the composition of the housing stock by stratum. The repeat-sales method leads naturally to the perspective of a chain index, where the weights assigned to the thinnest stratum level vary over time, depending on the degree of aggregation used for the calculation. Different degrees of detail may become necessary in the aggregation, according to the number of useable sales there are in the calculations. Since the structure of repeat sales shifts over time, the way the chaining procedure is applied cannot be specified *a priori*. This problem is not addressed by the users or the advocates of the method.

By contrast, the hedonic approach can be used to calculate indices of the Laspeyres, Paasche or chain type, since it comes to the prix of a reference property. However, in order to facilitate comparison with other types of investments, eg financial ones, it seems preferable not to choose chaining from one period to another, but to maintain a reference stock over a certain period of time. Indeed, for financial investment the usual practice is to track a fixed ("crystallised") portfolio so as not to mix up price effects with the effects of updating the portfolio (investment strategy).[74]

This seems all the more justified for housing price indices since housing is a durable asset that carries high transaction costs and for which investor preferences between different types of goods evolve slowly. As a result, the index of housing prices is different from that of consumer prices. The argument for chaining consumer price indices (or those for the annual weighting revision) is based on the fact that we want to track changes in consumer preference, that are revealed by the share of the budget allocated to each major category of assets. In the housing sphere, adaptations are slow in all likelihood. At the detailed level, the indices are aggregated multiplicatively to take into account a possible substitutability of properties. At the higher level of aggregation, they are aggregated additively.

[74] The housing price indices are chained Laspeyres indices. Compared to version 2, the new calculation method corresponds to a succession of calculations of Laspeyres closer in time– every two years instead of every five years (cf. Chapter 3).

Appendix 6: Advance price indices

Usually, before a property is sold, a pre-contract is signed.[75] There are normally about three months between the signing of the pre-contract and the signing of the actual sale contract. The introduction of advanced price indices calculated from the pre-contracts will therefore enable us to produce indices that are more reactive to the state of the market.

Collecting pre-contracts

Notaries are legally obliged to pass on information about pre-sale contracts to the housing databases, according to the Law of March 2011. However, the notariat has collected pre-contract data from the notarial offices since mid-2009. There have been two different systems. First, a portal was created for the notarial offices in the provinces to enter basic pre-contract information. With this system a large number of pre-contracts were collected (Table A6.1), however, it did not contain all the variables needed for calculating the indices.

Table A6.1: Number of sales and pre-contract sales of used dwellings collected in 2012

	Sales	Pre-sales
Paris Region	95,927	15,526
Provinces	308,650	111,216

The decision was therefore taken to develop a specific module for the teletransmission of pre-contracts, incorporated into the software used in the notarial offices to draw up deeds of sale. This was along the same lines as the teletransmission module for sales, which gradually replaced the need for the notaries to send paper copies of deeds.

The pre-contract teletransmission tool was introduced gradually into the notarial offices from 2010 (using IT systems and the deed of sale software). It has grown since then depending on the demands of database managers and developments in notarial IT services.

Based on the new version of the software, a procedure to pair up the pre-contract and the sale was put in place during 2013. It is now possible to carry out a number of calculations that had not been done previously due to a lack of information: in particular, it is now possible to study the time lag between pre-contract and sale, or the proportion of pre-contracts that do not ultimately become a sale.

Methodology and dissemination

Given this background and the volume of pre-contracts transmitted, it is not yet possible to produce a price index based solely on this data, however, this could be an ultimate goal. A provisional calculation method has been developed based on current indices.

The working hypothesis is that the econometric models developed for sales should also be valid for pre-sale contracts. Thus by applying these models to the pre-contract data it should be possible to calculate standardised prices ("reference-property equivalent" prices[76]) for the pre-contracts and compare them with the standardised prices for sales for the preceding quarters. We will therefore be able to calculate advance indicators, provided that the volume of pre-contracts received is sufficient.

Advance indicators for price changes in Paris and in the Hauts-de-Seine *département* are disseminated by the Chamber of Notaries of Paris through their economic outlook reports and press conferences. In the Provinces, a general indicator is given in the quarterly economic outlook reports and indications about price changes in some large provincial cities are also given at the annual national conference of the *Conseil Supérieur du Notariat*.

[75] All pre-contracts or "promises of sale" (*compromis* or *promesse de vente*) are collected by the notarial databases.
[76] Cf. Chapter 3.

Appendix 7: Agreement with PNS

Agreement regarding the continuation and expansion of the partnership between INSEE and PNS in the field of housing price indices in the Paris Region.

between:

- **CHAMBRE INTERDEPARTEMENTALE DES NOTAIRES DE PARIS**, represented by Mr. Benasse, notary, its chairman,

- CHAMBRE DES NOTAIRES DE SEINE ET MARNE, represented by Mr. Hautebas, notary, its chairman,

- CHAMBRE INTERDEPARTEMENTALE DES NOTAIRES DE VERSAILLES, represented by Mr. Savoure, notary, its chairman,

- **CHAMBRE DES NOTAIRES DE L'ESSONNE,** represented by Mr. Lemoine, notary, its chairman,

- CHAMBRE DES NOTAIRES DES HAUTS DE SEINE, represented by Mr. Herrnberger, notary, its chairman,

- **ASSOCIATION DES NOTAIRES DU CHATELET** (**P**ARIS **N**OTAIRES **S**ERVICES), a not-for-profit organisation administered under the 1901 Act, represented by Mr. Cauro, notary, its chairman;

Acting jointly and hereafter referred to as PNS,

the first party,

and:

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES, hereafter referred to as INSEE and represented by its Director-General Mr. Jean-Luc Tavernier,

the second party,

it has been agreed and decided as follows:

Preamble

PNS and INSEE signed an initial agreement on December 6, 1990, establishing a partnership for the calculation and publication of the price index for vacant used apartments sold in Paris.

A second agreement, dated May 16, 2000 (agreement 2000 00094), and an addendum dated November 26, 2002 have revised the method for calculating this index and extended the scope of application of this partnership to other price indices for vacant used housing sold in all or part of the Paris Region.

A third agreement, dated December 8, 2005 (agreement 2005 00353), and an addendum dated November 26, 2009 widened the scope of application of the indices even further and put in place combined dissemination and publication of the said indices.

The agreements signed in 1990, 2000 and 2005 have now reached their expiry dates. They are replaced by the present agreement, the purpose of which is to continue and strengthen the partnership between PNS and INSEE in the field of housing price indices. The present agreement takes into account recent progress in this partnership.

Article 1: Purpose of agreement

The present agreement defines the conditions for PNS and INSEE to participate in the calculation, validation, publication and dissemination of price indices for vacant used housing. These are calculated from property transaction data in the BIEN database, belonging to and managed by PNS, and fed with data by the notaries of the Paris Region.

PNS informs INSEE that the computerised system for processing information on named individuals in the BIEN database has been declared to the National Commission on Information Technology and Civil Liberties (*Commission Nationale de l'Iinformatique et des Libertés – CNIL*), in accordance with the provisions of Article 16 of the Act of January 6, 1978 Act on Information Technology, Data Files and Civil Liberties, amended by law 2004-801 of 6 August 2004.

The indices covered by the present agreement are the following, it being understood that the quarters referred to are calendar quarters:

- quarterly housing price indices (apartments and houses) for the entire Paris Region,
- quarterly housing price indices (apartments and houses) for each of the seven *départements* in the Paris Region, excluding Paris, for the three *départements* in the inner suburbs taken together, for the four *départements* in the outer suburbs taken together and for all of the Paris Region, excluding Paris,
- quarterly price indices for apartments for the entire Paris region,
- quarterly price indices for apartments in each of the eight *départements* in the Paris Region, for the three *départements* in the inner suburbs taken together, for the four *départements* in the outer suburbs taken together and for all of the Paris Region, excluding Paris,
- quarterly price indices for houses for the entire Paris Region,
- quarterly price indices for houses in each of the seven *départements* in the Paris Region excluding Paris, for the three *départements* in the inner suburbs taken together, for the four *départements* in the outer suburbs taken together and for all of the Paris Region, excluding Paris.

Each index is to be produced as a provisional index, calculated about one and a half months after the end of the quarter to which it relates, and a definitive index, calculated about four and a half months after the end of the quarter to which it relates.

In addition, each of these indices will be produced without adjustment for seasonal variations and also with adjustment for seasonal variations.

Hereafter, the indices covered by the present agreement will be referred to generically as “the indices”.

PNS informs INSEE that monthly used housing price indices are also calculated quarter-on-quarter by PNS for quarters that do not correspond to calendar quarters. These monthly quarter-on-quarter indices are calculated using the same method as the indices covered by the present agreement and cover the same time period.

Article 2: The “Indice Notaires-INSEE” designation and approval by the *Autorité de la statistique publique* (National public statistics authority)

Subject to the provisions of Article 5 of the present agreement, the indices will be allowed to use the designation “Indice Notaires-INSEE”. In this respect, INSEE authorises PNS to use this designation in external communications and in relations with third parties.

PNS is informed that similar indices calculated for the Provinces by the company Min.not and the Conseil Supérieur du Notariat in partnership with INSEE also use the “Indice Notaires-INSEE” designation, as do the indices for metropolitan France calculated by Min.not from the “Paris Region” index by PNS and the “Province” index by Min.not.

In addition, the Paris Region indices have received approval from the *Autorité de la statistique publique* (National public statistics authority) to use the designation. The Notaires-INSEE indices for the Paris Region have also received this approval.

Article 3: Scientific Board

The role of the Scientific Board for the Notaires-INSEE indices will be one of deliberation and advice for the parties of the present agreement. Its work will concern mainly (but not exclusively) the way in which the indices are calculated and the way the notarial databases are fed with data.

For example, it defines coverage rates below which the indices are not validated.

The Scientific Board is composed of:

i) a chairperson, chosen jointly by INSEE, PNS, Min.not and the Conseil Supérieur du Notariat;

ii) two representatives from INSEE;

iii) two representatives from PNS;

iv) one representative from Min.not;

v) one representative from Conseil Supérieur du Notariat.

The Scientific Board may call on the assistance of any qualified persons that it may judge necessary for carrying out its duties, for the duration of the present agreement or only occasionally.

The Board shall meet, at the request of the chairperson, at least once per quarter. Meetings shall be organised by INSEE and written minutes shall be kept of the proceedings. Each party may put before the chairperson a request for a meeting, as required.

The Scientific Board is the same for both agreements concerning the calculation, validation and dissemination of indices with the "Indice Notaires-INSEE" designation:

- the agreement linking PNS and INSEE;
- the agreement linking Min.not, the Conseil Supérieur du Notariat and INSEE.

Article 4: Amendments to the present agreement

Any change made to the present agreement shall be set out in an addendum.

Article 5: INSEE commitments

INSEE shall:

(i) provide PNS with its statistical expertise and advice for calculating the indices and shall supply data in its possession that can ensure the quality of the indices;

(ii) provide PNS in August every year with the correction coefficients for seasonal adjustments needed to calculate the seasonally adjusted indices in their index series;

(iii) include the indices covered by the present agreement in its own publications where dissemination seems justified;

INSEE shall put in place a method for monitoring the quality of the indices. At least one week before publication, the indices shall be submitted to INSEE for approval. INSEE shall respond within no more than two working days. If there is no response after two working days, this shall be the equivalent of approval.

Achieving minimal coverage rates is now a necessary condition for validation of the indices. The methodology for the indices is described in detail in a special issue of the *INSEE-Méthodes* series which is regularly updated by INSEE.

Article 6: PNS commitments

PNS shall:

(i) comply with the index calculation procedures recommended by the Notaires-INSEE indices Scientific Board,

(ii) allocate the human and physical resources (especially IT resources) needed to maintain and update these procedures and in particular shall adapt these resources to the increase in workload that will result when the obligation for notaries to supply data to the databases used to calculate the indices comes into force; PNS shall put in place a system to assess these human and financial needs on a regular basis and will evaluate the resources actually used over the last few quarters and anticipated use in the next quarters. These assessments will be put before the Scientific Board of the Notaires-INSEE indices;

(iii) preserve and improve the current system for collecting information on real estate sales in the Paris Region: in particular the collection rate and data incorporation time shall be monitored via dashboards (number of notarised transactions and number of transactions added per month, stock awaiting codification); these will be submitted to INSEE every quarter. PNS agrees to inform INSEE before any change is made to the process of creating the notarial databases (definition of fields, changes that could impact the collection and incorporation rates, data-completion rates for some fields, index values);

(iv) provide INSEE with the information defined as agreed with them so that they can approve the indices. In particular this will include detailed indices, price levels and the transaction amounts and volumes in the course of the quarter;

(v) publish the indices quarterly;

(vi) in accordance with the aim of the agreement, provide INSEE with all information extracted from PNS databases that could be of use for INSEE's internal statistical studies, and any price index series calculated by PNS for its own requirements and which would be useful to the Institute for study purposes;

(vii) announce in its publications any change made to the method used to calculate the monthly quarter-on-quarter indices not covered by the present agreement.

Article 7: Joint publication and dissemination of the indices

Publication and dissemination of the indices, by INSEE and PNS, are subject to embargo.

The date and time when the embargo is lifted, i.e. the date and time when INSEE and PNS are authorised to publish and disseminate the indices, are determined as follows:

- The dates for publication by INSEE and for the Chambre interdépartementale des notaires de Paris to hold its press conference are proposed by the Notaires-INSEE indices Scientific Board and approved by INSEE at least one quarter in advance. Basically, the principle behind the schedule is to have the dates in the last week of February, May, August and November. However, the date for the end of August will systematically be moved back by about two weeks.
- The time that the embargo is lifted is 8.45am.

Information about the indices is disseminated and published as follows:

i) for the attention of the media, by PNS:

By means of a press conference, which will take place systematically on the morning of the day the embargo is lifted, at 8.45am, barring exceptions. Once the press conference is over, information can be disseminated to all the media by any appropriate means;

In accordance with the recommendations of the Autorité de la Statistique Publique, the press file handed to journalists and the documents presented at the conference must clearly differentiate those indices which are designated "Indice Notaires-INSEE" from any other information.

ii) for the attention of the media, by INSEE:

Information is relayed to the press agencies at 8.30am and to the rest of the media at 8.45am, on the morning the embargo is lifted;

iii) for the attention of the general public, by PNS and / or by INSEE:

Information is communicated from 8.45am onwards, by any appropriate means.

The cut-off dates for the BIEN database when calculating the indices are determined by PNS in such a way as to ensure that the indices can be disseminated on the date the embargo is lifted, taking into account the time needed to carry out the necessary calculations, to check them and validate them, and also to prepare publications.

Coefficients and details of the zoning used in the regressions may not be made public except with the express agreement of INSEE and PNS.

Indices that are not approved by INSEE may not be published with the "Indice Notaires INSEE" designation.

Article 8: Financial conditions

In consideration of the fact that the preparation, monitoring and regular publishing of the indices is a task that falls within the respective remits of all partners, and that the cooperation required under the terms of the present agreement entails an evenly balanced workload for each party, the present agreement shall be entered into without financial considerations.

Article 9: Duration of contract

The present agreement will come into force as soon as it is signed by all the partners. It is concluded for a period of five years.

Article 10: Termination of the agreement

The present agreement may be cancelled by any one of the parties by registered letter with acknowledgement of receipt. Cancellation shall take effect six months after receipt of this letter.

In the event of cancellation, PNS and INSEE waive the right to future use of the "Indice Notaires-INSEE" designation.

However, the parties agree in advance to consult with each other in order to allow the continuation of financial operations undertaken by financial institutions or others, under any licensing agreements entered into by PNS.

Article 11: INSEE and PNS representatives

The following persons shall execute the terms of the present agreement:

i) on behalf of INSEE, the head of the Housing division in the Department of Consumer Prices, Household Income and Living Conditions;

ii) on behalf of PNS, the current chairman of the ASSOCIATION DES NOTAIRES DU CHATELET (PARIS NOTAIRES SERVICES).

Article 12: Enforcement clauses

The present agreement is exempt from stamp duty and formal registration requirements.

The implementation of the present agreement shall have as an immediate consequence the abrogation of agreement 2005 00353 and its addendum dated 2009.

Signed in Paris in seven original copies, 13 November 2013,

For PNS and the CHAMBRES DES NOTAIRES D'ILE-DE-FRANCE

Signed: Maître Bénasse

Signed: Maître Hautebas

Signed: Maître Savoure

Signed: Maître Lemoine

Signed: Maître Herrnberger

Signed: Maître Cauro

For INSEE, the Director General

Signed: Monsieur Jean-Luc Tavernier

Appendix 8: Agreement with the Conseil supérieur du Notariat *and Min.not*

Agreement regarding the continuation and expansion of the partnership between INSEE, the *Conseil supérieur du Notariat* and Min.not in the field of housing price indices in the provinces.

between:

- The **CONSEIL SUPERIEUR DU NOTARIAT,** represented by Me Jean Tarrade, notary, its chairman,

- The Min.not Company - **Real Estate Market Notaries**, SASU capital of € 2,450,000, registered with the RCS Aix en Provence under number 381 000 611, represented by M. Daniel BOUCHON, its chairman,

the first party,

and:

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES, hereafter referred to as INSEE and represented by its Director-General Mr. Jean-Luc Tavernier,

the second party,

it has been agreed and decided as follows:

Preamble

The CNS, Perval and INSEE signed an initial agreement on June 15, 1998, establishing a partnership for the calculation and publication of the price index for vacant used apartments sold in the central cities of agglomerations of 10,000 or more inhabitants, located in France outside Ile-de-France.

A second agreement, dated October 7, 1999 has extended the scope of application of this partnership to the calculation and dissemination of a price index of used houses in all municipalities of the Provinces, as well as an index of prices of used apartments in all municipalities of province towns of over 10,000 inhabitants (instead of only central cities).

The CNS, Perval and INSEE signed an second agreement in 2005 to continue and strengthen the partnership between the CSN, Perval and INSEE in calculating indices in housing prices, by extending the scope to indices calculated in the administrative regions and large cities, since statistical quality of such evidence would be considered satisfactory by the parties. This second agreement was the subject of an amendment in 2010 on the one hand to describe the nature of the information transmitted by the Scientific Board for the Notaries-INSEE indices, for the quarterly validation of indices, and on the other hand to specify the conditions for the dissemination of data, including advancing the dates of publication of indices to reflect the improvement of the information collected by notarial databases.

The agreements signed in 1998 and 2005 have now reached their expiry dates. They are replaced by the present agreement, the purpose of which is to continue and strengthen the partnership between the CNS, Perval and INSEE in the field of housing price indices. The present agreement takes into account recent progress in this partnership.

Article 1: Purpose of agreement

The present agreement defines the conditions for the CNS, Perval and INSEE to participate in the calculation, validation, publication and dissemination of price indices for vacant used housing. These are calculated from property transaction data fed by the notaries in the Provinces on the one hand and by the price indices of used housing in Paris région calculated par Paris Notaires Service (PNS) on the other hand.

Min.not informs INSEE that the computerised system for processing information on named individuals from the housing transaction by the notaries in the Provinces and managed by ADSN is referenced in the list maintained by the Correspondent Infromatique et liberté, designated by Min.not in accordance with Act no. 78-17 "Informatique et Libertés" of January 6, 1978.

Each index covered by the present agreement is to be produced as an advanced provisional index calculated about one and a half months after the end of the quarter to which it relates, a provisional index, calculated about three months after the end of the quarter to which it relates, a semi-definitive index calculated about four and a half months after the end of the quarter to which it relates, and a definitive index, calculated about six months after the end of the quarter to which it relates.

In addition, each of these indices will be produced without adjustment for seasonal variations and also with adjustment for seasonal variations.

The indices covered by the present agreement are quarterly used housing price indices, without adjustment for seasonal variations and also with adjustment for seasonal variations, for:

- apartments and houses taken together,
- apartments,
- houses;

in the following geographic fields:

- metropolitan France,
- Provinces,
- Nord-Pas-de-Calais Region (provisional, semi-final and final indices only),
- Provence-Alpes-Côte-d'Azur Region (provisional, semi-final and final indices only),
- Rhône-Alpes Region (provisional, semi-final and final indices only) ;

In addition, the following indices will be produced without adjustment for seasonal variations and also with adjustment for seasonal variations:

- houses in the agglomeration of Lille (provisional, semi-final and final indices only),
- apartments:
    - in the agglomerations of 10,000 or more inhabitants in the Provinces taken together (provisional, semi-final and final indices only),
    - in the central towns of the agglomerations of 10,000 or more inhabitants in the Provinces taken together (provisional, semi-final and final indices only),
    - in the suburbs of the agglomerations of 10,000 or more inhabitants in the Provinces taken together (provisional, semi-final and final indices only),
    - in the rural municipalities and municipalities of the agglomerations with less than 10,000 inhabitants in the Provinces taken together (provisional, semi-final and final indices only),
    - in the municipality of Lyon (provisional, semi-final and final indices only),
    - in the municipality of Marseille (provisional, semi-final and final indices only).

Min.not informs INSEE that monthly used housing price indices are also calculated quarter-on-quarter by Min.not for quarters that do not correspond to calendar quarters. These monthly quarter-on-quarter indices are calculated using the same method as the indices covered by the present agreement and cover the same time period.

Hereafter, the indices covered by the present agreement will be referred to generically as "the indices".

Article 2: The "Indice Notaires-INSEE" designation

Subject to the provisions of Article 5 of the present agreement, the indices will be allowed to use the designation "Indice Notaires-INSEE" and optionally a custom region as "Index Notaires-INSEE Rhône-Alpes" appellation. In this respect, each party, the CSN, Min.not and INSEE, is authorized to use this designation in external communications and in relations with third parties.

The CSN and Min.not are informed that similar indices calculated for The Paris region by Paris Notaires Services 'PNS) in partnership with INSEE also use the "Indice Notaires-INSEE" designation.

Article 3: Scientific Board

The role of the Scientific Board for the Notaires-INSEE indices will be one of deliberation and advice for the parties of the present agreement. Its work will concern mainly (but not exclusively) the way in which the indices are calculated and the way the notarial databases are fed with data.

For example, it defines coverage rates below which the indices are not validated.

The Scientific Board is composed of:

i) a chairperson, chosen jointly by INSEE, PNS, Min.not and the Conseil Supérieur du Notariat;

ii) two representatives from INSEE;

iii) two representatives from PNS;

iv) one representative from Min.not;

v) one representative from Conseil Supérieur du Notariat.

The Scientific Board may call on the assistance of any qualified persons that it may judge necessary for carrying out its duties, for the duration of the present agreement or only occasionally.

The Board shall meet, at the request of the chairperson, at least once per quarter. Each party may put before the chairperson a request for a meeting, as required. Meetings shall be organised by INSEE and written minutes shall be kept of the proceedings.

The Scientific Board is the same for both agreements concerning the calculation, validation and dissemination of indices with the "Indice Notaires-INSEE" designation:

- the agreement linking PNS and INSEE;
- the agreement linking Min.not, the Conseil Supérieur du Notariat and INSEE.

Article 4: Amendments to the present agreement

Any change made to the present agreement shall be set out in an addendum.

Article 5: INSEE commitments

INSEE shall:

(i) provide Min.not with its statistical expertise and advice for calculating the indices and shall supply data in its possession that can ensure the quality of the indices;

(ii) provide Min.not in August every year with the correction coefficients for seasonal adjustments needed to calculate the seasonally adjusted indices in their index series;

(iii) include the indices covered by the present agreement in its own publications, where dissemination meets dissemination criteria defined by the Scientific Board.

INSEE shall put in place a method for monitoring the quality of the indices. At least one week before publication, the indices shall be submitted to INSEE for approval. INSEE shall respond within no more than two working days. If there is no response after two working days, this shall be the equivalent of approval.

Achieving minimal coverage rates is now a necessary condition for validation of the indices. The methodology for the indices is described in detail in a special issue of the *INSEE-Méthodes* series which is regularly updated by INSEE.

Article 6: CSN and Min.not commitments

Min.not shall:

(i) comply with the index calculation procedures recommended by the Notaires-INSEE indices Scientific Board,

(ii) allocate the human and physical resources (especially IT resources) needed to maintain and update these procedures and in particular shall adapt these resources to the increase in workload that will result when the obligation for notaries to supply data to the databases used to calculate the indices comes into force;

(iii) monitor the collection rate and data incorporation time via dashboards (number of notarised transactions and number of transactions added per month, stock awaiting codification); these will be submitted to INSEE every quarter. Min.not agrees to inform INSEE before any change is made to the process of creating the notarial databases (definition of fields, changes that could impact the collection and incorporation rates, data-completion rates for some fields, index values);

(iv) provide INSEE with the information defined as agreed with them so that they can approve the indices. In particular this will include detailed indices, price levels and the transaction amounts and volumes in the course of the quarter;

(v) publish the indices quarterly;

(vi) in accordance with the aim of the agreement, provide INSEE with any price index series calculated by Min.not for its own requirements and which would be useful to the Institute for study purposes;

The CSN shall :

(i) preserve and improve the current system for collecting information on real estate sales in the provinces: in particular the collection rate and data incorporation time.

Article 6 bis: Protection of personal data

The CSN, Min.not and INSEE, each commit to complying with the regulations on the protection of personal data, according to Act no. 78-17 "Informatique et Libertés" of January 6, 1978 and its decrees.

Thus, they shall take all necessary precautions to maintain the security, confidentiality of personal data which they are aware in the framework of this agreement and in particular to prevent it from being distorted, damaged or communicated to persons not expressly authorized.

CSN, Min.not and INSEE commit to making any administrative process under their responsbility according to regulations related to the protection of personal data.

In this respect, it is reported that the CSN and Min.not, have each appointed a correspondant « Informatique et Libertés » responsible for maintaining and updating the register of treatment and makes it available to the Commission Nationale Informatique et Libertés.

The CSN Min.not and INSEE vouch, as defined in Article 1120 of the Civil Code, for the respect by their employees and/or possible subcontractors duly authorized, of this section. It was agreed that the present Agreement may be the object of a communication to the Commission Nationale Informatique et Libertés.

Article 6 ter: – confidentiality

The CSN Min.not and INSEE are conventionally required to absolute secrecy about all information from databases, documents or any element, in particular technical, financial or organizational they would access, under this Agreement and / or previous conventions. Each party undertakes to respect this confidentiality obligation by all of its staff, any officer, any contractor and any subcontractor.

Breach of that duty of confidentiality commits the responsibility of the CSN, or Min.Notaires-Insee or INSEE and constitutes a case of early termination which is described in section 10 "Terms of denunciation of the agreement."

Article 7: Joint publication and dissemination of the indices

Publication and dissemination of the indices, by INSEE, the CSN and Min.not, are subject to embargo.

The date and time when the embargo is lifted, i.e. the date and time when INSEE, The CSN and Min.not are authorised to publish and disseminate the indices, are determined as follows:

Dates INSEE publications are proposed by the Scientific Board for the Notaires-INSEE and validated by the INSEE at least one quarter in advance.

The time that the embargo is lifted is 8.45am.

Information about the indices is disseminated and published as follows:

i) for the attention of the media, by the CSN and Min.not:

By means of a press conference, which will take place systematically on the morning of the day the embargo is lifted, at 8.45am, barring exceptions. Once the press conference is over, information can be disseminated to all the media by any appropriate means;

In accordance with the recommendations of the Autorité de la Statistique Publique, the press file handed to journalists and the documents presented at the conference must clearly differentiate those indices which are designated "Indice Notaires-INSEE" from any other information.

ii) for the attention of the media, by INSEE:

Information is relayed to the press agencies at 8.30am and to the rest of the media at 8.45am, on the morning the embargo is lifted;

iii) for the attention of the general public, the CSN, Min.not and / or by INSEE:

Information is communicated from 8.45am onwards, by any appropriate means.

The cut-off dates for the BIEN database when calculating the indices are determined by Min.not in such a way as to ensure that the indices can be disseminated on the date the embargo is lifted, taking into account the time needed to carry out the necessary calculations, to check them and validate them, and also to prepare publications.

Coefficients and details of the zoning used in the regressions may not be made public except with the express agreement of INSEE, the CSN and Min.not.

Indices that are not approved by INSEE may not be published with the "Indice Notaires INSEE" designation.

Article 8: Financial conditions

In consideration of the fact that the preparation, monitoring and regular publishing of the indices is a task that falls within the respective remits of all partners, and that the cooperation required under the terms of the present agreement entails an evenly balanced workload for each party, the present agreement shall be entered into without financial considerations.

Article 9: Duration of contract

The present agreement will come into force as soon as it is signed by all the partners. It is concluded for a period of five years.

Article 10: Termination of the agreement

The present agreement may be cancelled by any one of the parties by registered letter with acknowledgement of receipt. Cancellation shall take effect six months after receipt of this letter.

In the event of cancellation, The CSN, Min.not and INSEE waive the right to future use of the "Indice Notaires-INSEE" designation. However, the parties agree in advance to consult with each other in order to allow the continuation of financial operations undertaken by financial institutions or others, under any licensing agreements entered into by Min.not and the CSN.

Article 11: INSEE, CSN and Min.not representatives

The following persons shall execute the terms of the present agreement:

i) on behalf of INSEE, the head of the Housing division in the Department of Consumer Prices, Household Income and Living Conditions;

ii) on behalf of the CSN, its chairman,

ii) on behalf of Min.not, its chairman.

Article 12: Enforcement clauses

The present agreement is exempt from stamp duty and formal registration requirements.

The implementation of the present agreement shall have as an immediate consequence the abrogation of agreement 2005 00329 and its addendum dated 2010.

Signed in Paris in three original copies, 25 june 2014

A Paris, le 21 mai 2014	A Paris, le 21 mai 2014	A Paris, le 11 juin 2014	A Venelles, le 20 juin 2014
Pour le ministre des Finances et des Comptes publics et par délégation le directeur général de l'Insee	Pour le ministre de l'Économie, du Redressement productif et du Numérique et par délégation le directeur général de l'Insee	Pour le CSN	Pour Min.not
Signé : Jean-Luc TAVERNIER	Signé : Jean-Luc TAVERNIER	Signé : Me Jean TARRADE	Signé : Christine REY DU BOISSIEU

Appendix 9: Definitions

Used housing

We use a tax definition for used (second-hand) housing. A transaction is deemed to concern a used dwelling if it is the first sale more than five years after the construction completion date, or, if it is a second sale, then irrespective of the construction completion date. Thus the transaction may consist of the first sale of a property that is already used (more than five years old), or the second sale of a nearly-new property. The tax distinction is reflected in different tax rates: 0.6% for new dwellings, 4.8% for existing dwellings.

Suburb

Suburbs are suburban municipalities in urban units of 10,000+ inhabitants: all units not classified as "central city" are suburbs.

Reference property

Property whose price is tracked to calculate the indices. The characteristics of this property are the reference modalities for the explanatory variables of the transaction price model (e.g. 4-room house, on two floors, with a garage and bathroom). The reference property is described in Chapter 3 (Table 3.1).

Apartment buildings

Properties in apartment buildings are categorised as apartments (studio, apartment, duplex, triplex). We exclude rooms, attics, lofts, workshops and superintendents' lodgings.

Occupancy status and purpose

The dwellings tracked in the indices are vacant at the time of sale; they are intended for residential use only, and are acquired with full property rights. We do not exclude apartments leased by the purchaser before the sale. We do remove apartments occupied by a third party or by the seller when loss of use exceeds six months and when there is a right of use and residence or a right of usufruct.

Single-family dwellings

Properties are called single-family dwellings if they are houses, whether detached or not. They have a separate, direct private entrance from outside. To maintain consistency, we exclude property types such as large estates, manors, luxury townhouses, towers, and converted watermills/windmills. We therefore include farms, townhouses in cities and villages, detached houses and villas. We also include properties of which the type is not specified.

Transaction type

The only transactions considered are private sales conducted directly between sellers and buyers or through a real estate professional. Sales by voluntary auction on the Notaires' Real Estate Market (*Marché Immobilier des Notaires*) are therefore excluded.

Seller and buyer categories

The seller may be an individual, a real estate professional, or a company. Only dwellings purchased by individuals or by real estate companies (*Société Civile Immobilière (SCI)*) are tracked to calculate the indices. Dwellings purchased by real estate professionals are therefore outside the scope of the indices. Non-responses, which are much more numerous, are included in the scope of the indices.

Table A9.1 – Number of dwellings sold by buyer type: Provinces

	Apartments						Houses	
	Central cities of UU >10,000 inhab.		Suburbs of UU >10,000 inhab.		Rural and UU<10,000 inhab.			
Buyer type	Number	%	Number	%	Number	%	Number	%
Individuals	78,763	89%	34,883	90%	14,789	88%	193,190	91%
Real-estate companies (SCI)	4,602	5%	1,714	4%	808	5%	8,377	4%
Real-estate dealers	546	1%	210	1%	53	0%	1,288	1%
Other (non-professionals)*	949	1%	480	1%	210	1%	2,912	1%
Non-responses	3,830	4%	1,632	4%	959	6%	7,310	3%
Total	88,690	100%	38,919	100%	16,819	100%	213,077	100%

* Administrations, businesses, etc.

Scope: used housing transactions in notarial databases, 2010.

Table A9.2 - Number of dwellings sold by buyer type: Paris Region (apartments)

Buyer type	Apartments							
	Paris		Inner suburbs		Outer suburbs		Total	
	Number	%	Number	%	Number	%	Number	%
Individuals	25,376	85%	33,125	92%	24,092	95%	82,593	90%
Real-estate companies (SCI)	2,140	7%	1,471	4%	661	3%	4,272	5%
Real-estate dealers	231	1%	92	0%	28	0%	351	0%
Other	691	2%	599	2%	187	1%	1,477	2%
Non- responses	1,435	5%	859	2%	449	2%	2,743	3%
Total	29,873	100%	36,146	100%	25,417	100%	91,436	100%

Scope: used housing transactions in notarial databases, 2010.

Table A9.3 - Number of dwellings sold by buyer type: Paris Region (houses)

Buyer type	Houses					
	Paris and inner suburbs		Outer suburbs		Total	
	Number	%	Number	%	Number	%
Individuals	9,303	90%	26,476	94%	35,779	93%
Real-estate companies (SCI)	468	5%	779	3%	1,247	3%
Real-estate dealers	82	1%	147	1%	229	1%
Other	229	2%	268	1%	497	1%
Non- responses	257	2%	438	2%	695	2%
Total	10,339	100%	28,108	100%	38,447	100%

Scope: used housing transactions in notarial databases, 2010.

Reference stock

Stock of which the change in value constitutes the index (cf. Chapter 3). For a given neighbourhood or city, it consists of all the transactions in the reference period that fall within the scope of the index, except sales where the price is judged to be an outlier ($1/20^{th}$ at each end of distribution).[77] The reference stock forms the basket or portfolio of properties for the index.

Estimation stock

Housing stock of which the values are used to estimate the relative prices of property characteristics. It consists of all dwellings sold in the reference period that fall within the scope of the index. We exclude transactions considered to be aberrant, i.e. those whose residuals are greater than two standard deviations, and which therefore lie outside the interval $\left[\overline{x} - 2\sigma; \overline{x} + 2\sigma\right]$. The estimation stock and the reference stock are therefore sub-sets of the set of transactions for the estimation period that fall within the scope of the index (*Table A9.4*). Non-responses and the processing of missing observations are dealt with in Chapter 4 (*Table 4.7*).

[77] In version 1 of the model, one sixth of transactions with extreme values were removed as a precaution and because the quality of these data was not known (David *et al*., 2002). Later tests showed that the indices were in fact robust and could withstand a more parsimonious elimination of outliers.

For example: for the "Lille suburbs, used houses" stratum, we have 10,337 transactions in 2007 and 2008. Of these, 2,831 had some variables missing or were not properly documented, and in 340 cases the price was considered to be higher or lower than the observed price by more than two standard deviations. This left 7,166 transactions in the estimation stock. The reference stock comprises 6,772 transactions: the original 7,506 transactions of the estimation stock (7,166 + 340), minus the outliers.

Table A9.4 – Size of reference stock and estimation stock of Notaires-INSEE indices

	Number of dwellings in reference stock*	Number of transactions In estimation stock*
Metropolitan France	723,691	764,084
Paris Region		
Apartments	146,089	154,318
Paris	45,811	48,269
Inner suburbs	58,673	62,126
Outer suburbs	41,605	43,923
Houses (total)	55,792	58,893
Inner suburbs (+Paris)	14,923	15,733
Outer suburbs	40,869	43,160
Provinces		
Apartments	209,824	222,026
UUs > 10,000 centre	128,992	137,162
UUs > 10,000 suburbs	56,502	59,827
Rural and UUs < 10,000	24,330	25,037
Houses	311,986	328,847

* Stock in 2007-2008

Table A9.5 - Paris: comparison of structures in total stock, reference stock and annual transactions, %

	Number of dwellings*	Number of reference stock transactions**	Number of transactions in 2010	Number of transactions in 2011
Stratum				
1	20.0	20.4	19.7	20.2
2	30.0	33.0	32.4	31.8
3	26.0	23.3	23.6	24.1
4	108	10.1	10.5	10.2
5	13.2	13.1	13.8	13.6
Total	100.0	100.0	100.0	100.0
Construction period				
Before 1992	95.0	86.8	79.8	75.6
Since 1992	5.0	2.2	2.6	2.5
Unknown	0.0	11.0	17.6	21.7
Total	100.0	100	100.0	100.0
Size				
1 or 2 rooms	55.8	57.0	55.7	57.9
3+ rooms	44.2	41.2	42.5	40.6
Unknown	0.0	1.8	1.8	1.5
Total	100.0	100.0	100.0	100.0

* Source: 2008 Population census
** Stock 2007-2008

Table A9.5 compares the structure of the index reference stock (2007 - 2008) with that of the housing stock in the 2008 population census, and transactions from 2010 and 2011 for Paris. For a considerable proportion of the variables, information was not supplied for the databases and in addition, the number of transactions does not include new housing stock. Comparisons are therefore approximate. However, overall, these structures show that transactions give a good representation of the total housing stock.

A detailed breakdown of the estimation stocks for apartments and houses in the Paris Region and the Provinces is given at the end of the appendices.

Base period, index base quarter

1st quarter 2010.

Estimation period or reference period

Period for which we estimate the basic models, i.e. the relative prices of property characteristics. In version 3 of the indices, this period is two years.

Property price

Seller's net price (after deduction of agency commission if this is recorded in the deed of sale), therefore excluding taxes and legal fees.

Stratum, neighbourhood

Strata are geographic areas within which prices are homogeneous (cf. Chapter 3). For the method used to determine the strata, cf. Appendix 2. For the stratification of the municipalities, cf. Excel file, obtained on request from notaries.

Thus a stratum corresponds to the scope of application of a distinct hedonic model. A neighbourhood is a finer division, whose influence is demonstrated by suitable dummies within a given hedonic model. Neighbourhoods do not necessarily correspond to administrative boundaries.

Urban unit

An urban unit is defined as a set of dwellings arranged so that no two units are more than 200 metres apart, and containing at least 2,000 inhabitants. Municipalities (*communes*) that meet these criteria make up urban units (UU), the others are said to be "rural". The concept of the urban unit, based on the continuity of the built-up area and the threshold of 2,000 "agglomerated" inhabitants, is thus a mainly visual notion, based on demographics and settlement patterns.

Central city

A municipality (*commune*) is defined as a central city if it contains more than half of the population of the agglomeration. Otherwise, all municipalities with a population exceeding half the population of the largest municipality in the agglomeration are classified as central cities, as is the largest municipality.

If the annual number of transactions is less than 110, the central city does not have its own specific index.

Appendix 10: Bibliography

For a more comprehensive bibliography, see Beauvois et al. (2005), Insee Méthodes no. 111, available online on this website.

Bailey, M., Muth, R., Nourse H.: "A Regression Method for Real Estate Price Index Construction", *Journal of the American Statistical Association*, 58, 933-942, 1963.

Beauvois, M., David A., Dubujet F., Friggit, J., Gouriéroux C., Laferrère A., Massonnet, S., Vrancken, E. : "Les indices Notaires-Insee de prix des logements anciens, Version 2 des modèles hédoniques", Insee Méthodes, no. 111, 2005.

Case, B., Pollakowski, H., Wachter S.: "Frequency of Transaction and House Price Modelling", *JREFE*, 14, 173-188, 1997.

Clapp, J., Giacotto, C.: "House Price Indices Based on All Transactions Compared to Repeat Subsamples", *AREUEA*, 19, 270-285, 1991.

Court A.: "Hedonic Price Indices with Automotive Examples", p. 99-117, in The Dynamics of Automobile Demand, New-York, General Motors Company, 1939.

David A., Dubujet F., Gouriéroux C. and Laferrère A.: "Les indices de prix des logements anciens", Insee méthodes, no. 98, 2002.

Duon, G.: "Documents sur le problème du logement", Ministère de l'Économie, Service National de la Statistique, *Études Économiques, n°1,* 1946.

Eichholtz Piet M.A., "A Long Run House Price Index : the Herengracht Index, 1629-1973", REE, 25 (2),175-192, Amsterdam, August 1996.

Grebler, Blank and Winnick, Annexe C of "Capital Formation in Residential Real Estate: Trends and Prospects", *National Bureau of Economic Research,* Princeton University Press, Princeton, 1956.

Griliches Z.: "Price Indices and Quality Change", Cambridge, Harvard University Press, 1971.

Mark, J., Goldberg M.: "Alternative Housing Price Indices: An Evaluation", *AREUEA*, 12, 30-49, 1984.

Moreau, A.: "Methodology of the Price Index for Microcomputers and Printers in France", in *Industry Productivity: International Comparison and Measurement Issues*, OECD Proceedings, 99-118, 1996.

Triplett, Jack E.: "Concept of Quality in Input and Output Price Measures: A Resolution of the User-Value Resource Cost Debate", in *The US Income and Product Accounts: Selected Topics*, ed M.F. Foss, NBER Studies in Income and Wealth, University of Chicago Press, Chicago, 1983.

Triplett Jack E. : "Handbook on Hedonic Indexes and Quality Adjustments in Prices Indexes: Special Application to Information Technology Products", *STI Working papers 2004/9, OECD*, 2004.

Warsh, D.: "Economics principals, Roots and Wings", Globe Columnist, 14 March 1999.

Appendix 11: Current developments concerning the indices

Updated March 2014

List of indices approved by the French public statistics authority (*Autorité de la statistique publique*)

The *Autorité de la Statistique Publique (ASP)* has approved the indices for the Paris Region, produced by Paris Notaires Services (Notice no. 2011-01 by the *Autorité de la statistique publique* on 21 June 2011 concerning the approval of quarterly statistics on housing prices in the Paris Region).

Each of the quarterly indices is available as raw data and seasonally adjusted data, as follows:

- provisional index *(calculated about one and a half months after the end of the quarter to which it relates),*
- definitive index (*calculated about four and a half months after the end of the quarter to which it relates).*

Indices with the Notaires-INSEE designation:

- Apartments and houses:
  - all Paris Region,
  - all Paris Region, excl. Paris,
  - all three inner suburb *départements*,
  - all four outer suburb *départements*.
  - each of the seven *départements* in Paris Region, excl. Paris,
- Apartments:
  - all Paris Region,
  - all Paris Region, excl. Paris,
  - all three inner suburb *départements*,
  - all four outer suburb *départements*.
  - each of the eight *départements* in Paris Region,
- Houses:
  - all Paris Region,
  - all Paris Region, excl. Paris,
  - all three inner suburb *départements*,
  - all four outer suburb *départements*.
  - each of the seven *départements* in Paris Region, excl. Paris,

List of Notaires-INSEE indices

There are two separate bodies who disseminate indices with the “Notaires-INSEE” designation:

- Paris Region: Paris Notaires Services (PNS),
- Provinces and metropolitan France: Min.not.

Paris Region

Indices disseminated by PNS are covered by an agreement between PNS and INSEE.[78] They are available as raw data and seasonally adjusted data, as follows:

- provisional index *(calculated about one and a half months after the end of the quarter to which it relates),*
- definitive index *(calculated about four and a half months after the end of the quarter to which it relates).*

Indices with the Notaires-INSEE designation for the Paris Region:

- Apartments and houses:
  - all Paris Region,
  - all Paris Region, excl. Paris,
  - all three inner suburb *départements*,
  - all four outer suburb *départements*.
  - each of the seven *départements* in Paris Region, excl. Paris,
- Apartments:
  - all Paris Region,
  - all Paris Region, excl. Paris,
  - all three inner suburb *départements*,
  - all four outer suburb *départements*.
  - each of the eight *départements* in Paris Region,
- Houses:
  - all Paris Region,
  - all Paris Region, excl. Paris,
  - all three inner suburb *départements*,
  - all four outer suburb *départements*.
  - each of the seven *départements* in Paris Region, excl. Paris,

Provinces and metropolitan France

Min.not disseminates the indices for the Provinces and the national aggregate data. These indices are covered by an agreement between the *Conseil Supérieur du Notariat*, Min.not and INSEE.[79] They are available as raw data and as seasonally adjusted data, as follows:

- advance provisional index *(calculated about one and a half months after the end of the quarter to which it relates[80]),*
- provisional index *(calculated about three months after the end of the quarter to which it relates),*
- semi-definitive index *(calculated about four and a half months after the end of the quarter to which it relates),*
- definitive index *(calculated about six months after the end of the quarter to which it relates).*

Indices with the Notaires-INSEE designation for the Provinces and metropolitan France:

[78] Cf. Appendix 7.
[79] Cf. Appendix 8.
[80] This index is disseminated only for all of the Provinces and all of metropolitan France.

- Apartments and houses:
    - entire country,
    - Provinces,
    - Nord-Pas-de-Calais region,
    - Provence-Alpes-Côte- D'Azur region,
    - Rhône-Alpes region.
- Apartments:
    - entire country,
    - Provinces,
    - agglomerations with more than 10,000 inhabitants in the Provinces,
    - central cities in agglomerations with more than 10,000 inhabitants in the Provinces,
    - agglomerations with more than 10,000 inhabitants in the Provinces,
    - rural municipalities and municipalities in agglomerations with less than 10,000 inhabitants in the provinces,
    - Nord-Pas-de-Calais region,[81]
    - Provence-Alpes-Côte-D'Azur region,[82]
    - Rhône-Alpes region,[83]
    - municipality of Lyon,[84]
    - municipality of Marseille.[85]
- Houses:
    - entire country,
    - Provinces,
    - Nord-Pas-de-Calais region,
    - Provence-Alpes-Côte-D'Azur region,
    - Rhône-Alpes region.
    - Lille agglomeration (only semi-definitive and definitive indices).

[81] Only provisional, semi-definitive and definitive indices.
[82] Idem.
[83] Idem.
[84] Idem.
[85] Idem.

Table A11.1 - Series identifiers in the Macro-economic Database (BDM) on the insee.fr website

	Seasonally adjusted identifier	Raw data identifier
Municipality of Marseille - Apartments	1587626	1587625
Agglomeration of Lille - Houses	1587618	1587617
Municipality of Lyon - Apartments	1587634	1587633
Paris - Apartments	1587636	1587635
Seine-et-Marne - Apartments	1587638	1587637
Seine-et-Marne - Houses	1587640	1587639
Yvelines - Apartments	1587642	1587641
Yvelines - Houses	1587644	1587643
Essonne - Apartments	1587646	1587645
Essonne - Houses	1587648	1587647
Hauts-de-Seine - Apartments	1587650	1587649
Hauts-de-Seine - Houses	1587652	1587651
Seine Saint Denis - Apartments	1587654	1587653
Seine Saint Denis - Houses	1587656	1587655
Val de Marne - Apartments	1587658	1587657
Val de Marne - Houses	1587660	1587659
Val d'Oise - Apartments	1587662	1587661
Val d'Oise - Houses	1587664	1587663
Metropolitan France - Apartments	1587576	1587575
Metropolitan France - Total	1587580	1587579
Metropolitan France - Houses	1587578	1587577
Provinces - Agglomerations with >10,000 inhabitants - Apartments	1587588	1587587
Provinces - Agglomerations with >10,000 inhabitants - Suburbs	1587592	1587591
Provinces - Agglomerations with >10,000 inhabitants – Central cities - Apartments	1587590	1587589
Provinces - Agglomerations with >10,000 inhabitants and rural - Apartments	1587594	1587593
Provinces - Apartments	1587582	1587581
Provinces - Total	1587586	1587585
Provinces - Houses	1587584	1587583
Paris Region – Outer suburbs - Apartments	1587608	1587607
Paris Region – Outer suburbs - Houses	1587610	1587609
Paris Region excl. Paris - Apartments	1587596	1587602
Paris Region – Inner suburbs - Apartments	1587604	1587603
Paris Region - Inner suburbs - Houses	1587606	1587605
Paris Region - Apartments	1587597	1587595
Paris Region - Total	1587601	1587600
Paris Region - Houses	1587599	1587598
Nord Pas de Calais - Apartments	1587612	1587611
Nord Pas de Calais - Total	1587616	1587615
Nord Pas de Calais - Houses	1587614	1587613
Rhône Alpes - Apartments	1587628	1587627
Rhône Alpes - Total	1587632	1587631
Rhône Alpes - Houses	1587630	1587629
Provence Alpes Cote d'Azur - Apartments	1587620	1587619
Provence Alpes Cote d'Azur - Total	1587624	1587623
Provence Alpes Cote d'Azur - Houses	1587622	1587621
Paris Region excl. Paris - Total	1634730	1634729
Paris Region – Inner suburbs - Total	1634732	1634731
Paris Region – Outer suburbs - Total	1634734	1634733
Paris Region - Seine et Marne - Total	1634736	1634735
Paris Region - Yvelines - Total	1634738	1634737
Paris Region - Essonne - Total	1634740	1634739
Paris Region - Hauts de Seine - Total	1634742	1634741
Paris Region - Seine Saint Denis - Total	1634744	1634743
Paris Region - Val de Marne - Total	1634746	1634745
Paris Region - Val d'Oise - Total	1634748	1634747

Conditions for disseminating the quarterly Notaires-INSEE indices

In order to be able to use the "Notaires-INSEE" designation, an index must adhere to a certain number of criteria that are strictly monitored by the Scientific Board of the Notaires-INSEE indices: coverage rates, scale of revisions, numbers of deeds taken into account, time taken to incorporate deeds into the databases.

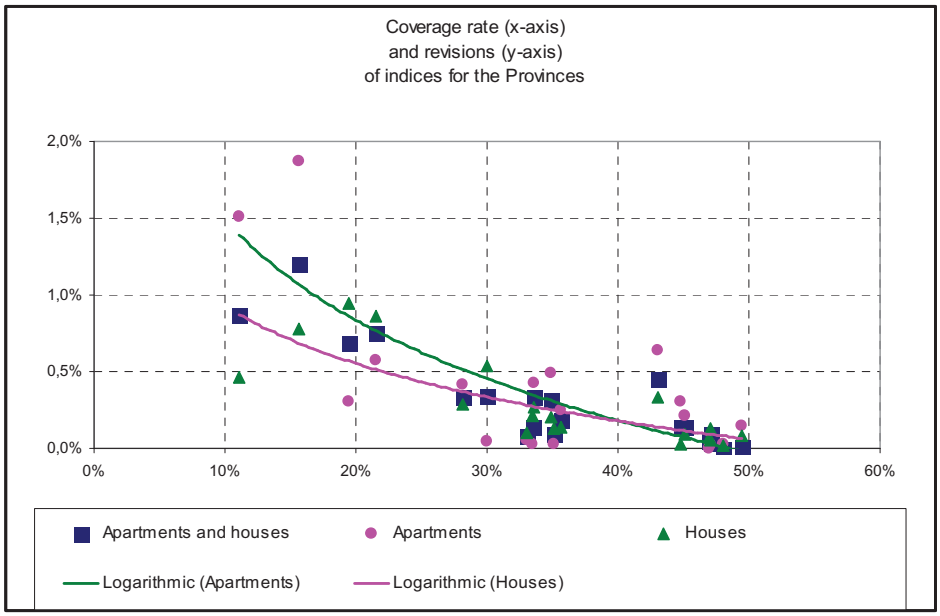
Minimum coverage rates have been set, by index type (*Table A11.2*). This is to ensure that few revisions to the index are required between dissemination of the first version ("advance provisional" index) and the last ("definitive" index). Thresholds were determined based on an analysis of the links between coverage rates and the scale of revisions required in indices covering the provinces (*Fig. A11.1*). If these thresholds are not respected, then the dissemination of the index is interrupted.

Table A11.2 – Minimum coverage rate for regions and entire country

By geographic level	Advance Provisional	Provisional	Semi-definitive	Definitive
Entire country	20%	20%	20%	20%
Province	30%	30%	30%	30%
Other supra-regional indices	Not disseminated	50%	50%	50%
Region	Not disseminated	50%	50%	50%

N. B.: transaction volumes and coverage rates in the Paris Region are always considerably greater than the minimum thresholds set for the provinces.

Figure A11.1: Link between coverage rate and index revisions for the Provinces



Reference period for reference and estimation stocks

Reference and estimation stocks have the same period of reference in all versions for a primary index.

Reference period before version V3

The reference period for reference and estimation stocks remains the same for the lifetime of the database.

Table A11.3 – Reference period for reference and estimation stocks in V1 and V2

Version	Geographic area	Dwelling type	Reference period
Version 1	Provinces	Apartment	1994 - 1996
	Provinces	House	1994 - 1997
	Paris	Apartment	1992 - 1996
	Inner suburbs (Paris Region)	Apartment	1992 - 1996
Version 2	Provinces	Apartment	1998 - 2001
	Provinces	House	1998 - 2001
	Paris Region	Apartment	1998 - 2001
	Paris Region	House	1998 - 2001

Reference period in V3

In version 3, the reference period is not the same as the database lifetime. Reference and estimation stocks are renewed every two years. We take two years, N-3 and N-2, to produce the price indices for years N and N+1. The reference periods are the same for the Provinces and the Paris Region and are independent of dwelling type.

Table A11.4 – Reference period and production period in V3

Reference period	Period when indices produced
2005 - 2006	2008 - 2009
2007 - 2008	2010 - 2011
2009 - 2010	2012 - 2013
2011 - 2012	2014 - 2015
2013 - 2014	2016 - 2017
2015 - 2016	2018 - 2019

Useful links

INSEE

- *Bulletin statistique* : http://www.insee.fr/fr/bases-de-donnees/bsweb/

- Macro-economic database (BDM): http://www.bdm.insee.fr/bdm2/index.action

- Quick quarterly information on used housing prices:

http://www.insee.fr/fr/themes/indicateur.asp?id=96

- Used housing price indices, version 2 of hedonic models, *Insee Méthodes* no.111, December 2005.

http://www.insee.fr/fr/publications-et-services/sommaire.asp?codesage=IMET111&nivgeo=0

http://www.insee.fr/en/publications-et-services/sommaire.asp?codesage=IMET111&nivgeo=0

- Used housing price indices, *Insee Méthodes* no. 98, September 2002
http://www.insee.fr/fr/publications-et-services/sommaire.asp?codesage=IMET098&nivgeo=0

Chambre interdépartementale des notaires de Paris (CINP) : http://www.paris.notaires.fr/

- Past press conferences by Paris Region Notaries:

http://www.paris.notaires.fr/presse/conferences-presse-immobilieres

- Latest market trends presented by notaries:

http://www.paris.notaires.fr/outil/immobilier/prix-et-nombre-de-ventes-paris-idf

- Observations on real estate trends by French notaries:

http://www.notaires.fr/notaires/communiques-de-presse

Council for the Environment and Sustainable Development (*Conseil général de l'Environnement et du Développement durable - CGEDD*)

- Historical statistics and analyses of long-term changes in the residential property market:

http ://www.cgedd.developpement-durable.gouv.fr/rubrique.php3?id_rubrique=138

- Numbers and amounts of property sales, estimate:

http://www.cgedd.developpement-durable.gouv.fr/nombre-et-montant-des-ventes-a1

- Secular indices of housing prices:

http://www.cgedd.developpement-durable.gouv.fr/IMG/doc/house-price-index-Paris-and-others-secular_cle7fed11.doc

Eurostat

Handbook on Residential Property Price Indices (RPPIs):

http://epp.eurostat.ec.europa.eu/portal/page/portal/hicp/methodology/hps/rppi_handbook

OECD

Triplett, Jack E. (2004): "Handbook on Hedonic Indices and Quality Adjustments in Price Indices: Special Application to Information Technology Products", STI Working papers 2004/9, OECD:

http://www.oecd.org/dataoecd/37/31/33789552.pdf