Working Paper 3.5

Regime Switching Models: An application to the Real Estate Market in Ile-de-France

Charles Maurin
André de Palma
Nathalie Picard

Revision: 1

FP7-244557 12/04/2012

SustainCity

EUROPEAN COMMISSION
European Research Area
SEVENTH FRAMEWORK PROGRAMME
Funded under Socio-economic Sciences & Humanities

Contents

Regime Switching Models: An application to the Real Estate Market in Ile-de-France

Charles Maurin
ENS Cachan (ENSAE)
61, Av. du Président Wilson,
94235 Cachan Cedex, France

Teleph.: + 33 6 14 78 82 62
Telefax:
charles.maurin@ensae.fr

André de Palma
ENS Cachan (CES)
61, Av. du Président Wilson,
94235 Cachan Cedex, France

Teleph.: + 33 6 88 71 43 44
Telefax:
andre.depalma@ens-cachan.fr

Nathalie Picard
UCP (THEMA)
33, Bd du Port, 95011 Cergy-Pontoise Cedex, France

Teleph.: +33 1 34 25 61 78
Telefax: +33 1 34 25 62 33
nathalie.picard@-cergy.fr

12/04/2012

Abstract

In the context of the SustainCity project (www.sustaincity.org), we apply regime switching models on the real estate market. This study uses transactions recorded in Ile-de-France. The objective is to determine the different regimes characterizing the behavior of the real estate market.

We use regime switching models along with hedonic models, at the county level. Different specifications are tested, where the regime is exogenous or endogenous, and compared. Out-of-sample performances are calculated, and predictions are made. The hedonic approach allows predicting prices for different types of dwellings.

Real estate markets exhibit regime switching, when they are associated with hedonic regressions. Moreover, they allow for better prediction in future prices.

Two different regimes are found, for periods of expansion or recession, coming from shift in the trend parameter.

Keywords

Hedonic Price Models; Regime Switching Models; Predictions; Forecasts; Factors; Index

Preferred citation style

Maurin, C., A. de Palma and N. Picard (2012) Regime Switching Models: An application to the Real Estate Market in Ile de France, *SustainCity Working Paper*, **3.5**.

Executive Summary

In the context of SustainCity project, we apply regime switching models on the real estate market. In this study, we use transactions which occurred in Ile-de-France, the Greater Paris region, between 1991 and 2006. The objective is to estimate and predict the distribution of prices.

We use regime switching models and find that they fit the data better than usual auto-regressive hedonic regression models. We identify two regimes in the estimation period. We focus on the temporal dimension of real market prices in this region.

Data

The data comprise several hundreds of thousands of transactions, and come from the *Notaires* database. The notaries of the region are encouraged to report transactions in this centralized database. They include a variety of variables, which can be used in hedonic regressions. In particular, the number of rooms, the number of bathrooms, the floor level, the floor area, or the presence of a garden, a balcony, a swimming pool or of a garage is often reported in the database.

Hedonic Regressions

Hedonic regressions models consist in evaluating the value of a good from its components. In the special case of real estate market, one can measure how the price of dwellings varies when characteristics change. This consists in a simple linear regression of the price (usually expressed in logarithm) on a series of dwelling and/or environment characteristics.

INSEE Indices

INSEE, the French institute of statistics, developed a hedonic regression model using the Notaries database. The objective is to provide a quarterly real estate index for the region, and different indices for the different départements, or districts of the region. They model the logarithm of the price per square meter, and use various dwelling characteristics, but no environmental characteristics.

Their model can be decomposed in a few steps. First they identify sub-regions, which consist in areas with similar real estate dynamics. Then, they choose a “reference bundle” and an “estimation bundle” of dwellings. The reference bundle consists in the dwellings for which we want a price index. The “estimation bundle” consists in the transactions used, at the different periods, to estimate the coefficients of the hedonic regressions. This allows to evaluate the “reference bundle” at different periods, and to measure the evolution of the real estate market. An aggregation is then made across *départements* or the whole region to compute the price indices.

We use these indices to fit regime switching models. In this type of models, we suppose that there exists different states of the world, called regimes, and that the data have a different behavior in each of them. Parameters specific to these unobservable regimes characterize the behavior of the time series. Here, we are particularly interested in the evolution of the price indices. We can expect to see a regime characterizing upward trends, and another one characterizing periods of crisis.

Regime switching models

Regime switching models, and more particularly the methods of estimations, are presented in the study. The Baum-Welch estimation procedure, a special case of the expectation-maximization algorithm, is described, and different caveats that occur are discussed. In particular, we discuss how to deal with local maxima, and how to choose the optimal number of states.

Application of regime switching models to INSEE Index

This type of models is tested in a first approach with the INSEE indices. These indices are computed since 1980 for Paris, and since 1991 for most of the other *départements*. The data used end at the third quarter of 2010. The data comprise, for most of the *départements*, two periods of decline in real estate prices, in the early 1990s, and in the late 2000s. The estimation of a Hidden Markov Model, where each regime is characterized by a Gaussian distribution, finds two different states. In a first state, the mean is positive, and the transition to the other state is quite low. The other state, on the contrary, has a negative mean, and is less persistent. In both states, the variances have similar orders of magnitude. These two different states fit the data well, and detect rapidly the change in trend of the real estate market. Out of sample estimations also show that the new state is detected rapidly, when the real estate market starts declining.

These estimations allow making predictions of the real estate market, based on the inference of the market's state at the last period.

Another model is estimated, where borrowing rate is added. It is found that a one period lagged borrowing rate has the higher correlation with real estate indices, and is therefore used in the estimation. The states inferred have the same signification and we obtain a negative correlation between the prices of dwellings, and the interest rate, which is consistent with expectations.

Application of regime switching model to a hedonic regression model

We develop a new model, where we first estimate a hedonic model on the transactions of the database, focusing on apartments, and estimate a regime switching model on the time dummies intercept of these regressions. These models are estimated for each *départements* with dummy variables for the different communes. This allows, in practice, to make forecasts of prices in each city, for each type of apartments. We suppose in this approach that the coefficients of the dwelling characteristics do not evolve with time. We model the logarithm of the price per square meter of the dwellings, and we use a number of explanatory variables: the number of rooms, floor level, presence of a garage, a terrace, and of a swimming pool.

As expected, the presence of a swimming pool or of a garage has a positive effect on the price of the dwelling. The effects of the number of rooms vary across *départements*. In particular, in Paris, the higher the number of rooms for a given floor area, the higher the price of the dwelling. The reverse holds for *départements* further away from the capital.

The regime switching model on the different time dummy variables give similar results to those found with the INSEE indices. In particular, we most often find a regime with a positive mean, and another one with a negative mean, the former one corresponding to period of expansion, the latter one to the burst of a real estate bubble.

One important caveat occurs with these estimations. For some *départements*, data are only available between 1996 and 2006. During this time interval, no substantive shift in the real estate market occurred. That is why, for those time series, the two states are much less different, in particular, we do not find a state with a negative mean but only a state with a zero mean. This outlines the learning aspect of regime switching models. If data after 2006 were used, comprising the subprime crisis which did have a small impact on the real estate market in the region, then it is highly plausible that this “bad state” would have been estimated with a negative mean.

Conclusion

The models developed focused on the temporal dimension of real estate market, and the spatial correlation could be taken into account in a more specified way. Nevertheless, regime switching models manage to fit the data correctly, identifying periods of expansion and recession. In particular, these estimations can be used to forecast prices in different *départements* or cities, using the information at the last period and the inferred state of the market.

1 Introduction

The Real Estate Market is characterized by a strong heterogeneity in the goods exchanged. This makes it more difficult to estimate directly econometric models, as it is usually done for other financial assets. Nevertheless, several tools can be used to model the Real Estate Market.

The first category of models consists of hedonic regression models. The Hedonic approach is used to evaluate the economic value of local amenities, such as water, noise or air pollution. The assumption is that the price of a good depends on the price of its different attributes. For Real Estate goods, the attributes include the number of rooms, the number of bathrooms, the location and the proximity to an urban center, or the presence of transportation nearby. In the hedonic approach, we evaluate the implicit prices of these different characteristics, from the observed market price of the whole good. Other variables, which are exogenous to the building, can also explain part of the price, in particular the borrowing rate.

As time elapses, the implicit price (that we will simply call the price) of the different characteristics can vary. The price of an additional bedroom in the city of Paris may vary more than the price of new transportation.

The second approach is the Real Estate Indices. Different indices have been developed. In the Greater Paris, the INSEE has developed an Index, the *indice des Notaires-INSEE*, which is based on hedonic regressions, and evaluates the price evolution of the Real Estate Market in the Paris region. The index is based on a bundle of buildings, with specific characteristics, in order to account for the possible quality heterogeneity in the dwellings sold at different periods. We will present this index, and test econometric model to fit these data.

Another type of index consists of repeated sales transactions. It assumes that the different coefficients of the hedonic model do not vary over time. The main advantage is that it almost automatically takes into account the quality effect, as we work on the price evolution of the same buildings. They may be modified over time though, with extension for old buildings for example, or their quality may improve (or deteriorate). This is why, when sufficient data are available, it is preferable to use hedonic models.

The price indices give us the trend of the Real Estate markets. Nevertheless, contrary to stock indices, it is rather unusual in the literature to fit these indices with a dynamic model. We develop a regime switching model based on these indices, which takes into account exogenous macro-economic variables, such as the borrowing rate, to determine the different states of the Real Estate market. The different regimes are characterized by the trend and the volatility of the price indices. They allow us to determine the periods of crisis and bubbles in this market, and to calculate out-of-sample forecasts. The main objective is to detect the switch from the “good” regime to the “bad” regime.

In this paper, we use the Base des Notaires, a database of several hundreds of thousands transactions in the Paris area. Approximately 70% of all the transactions occurring in the region are reported in this database. We work on hedonic models to investigate the construction of a price index, we present the Notaire-INSEE index, and then we build a model on the indices to determine the different states of the market. We finally compare these results with an application of regime switching models to a dummy time variable in hedonic regressions.

2 Hedonic Regression Models

2.1 Overview

Hedonic comes from the Greek "hedone", which means enjoyment. The objective of the Hedonic approach is to evaluate the economic value of a good from its components. Some attributes do not have an individual price, because they cannot be sold separately. In the case of Real Estate for example, it is not possible to purchase separately one room or a preferred location. The Hedonic approach allows assessing the implicit prices of the different characteristics of a good based on its observed market value.

For example, the price of a dwelling will depend on its location (quality of neighborhood, proximity to the city center, or accessibility by public transportation) and its proper characteristics (floor area, number of bedrooms and bathrooms, presence of a garden, a garage, or a swimming pool).

The hedonic regression assumes a functional form between the whole price and the prices of the different characteristics. Different specifications are usually used, the level regression, where the coefficients of the regression are interpreted as the price of an additional unit of the good, and the log approach, where the coefficients are interpreted in terms of elasticity. This last approach is usually preferred, as it usually reduces heteroscedasticity (Malpezzi, 2003).

We present in the next section the model developed by INSEE, which uses a log specification, where all the discrete explanatory variables are transformed into dummy variables. All the explanatory variables are discrete; the dependent variable is the log of the price on the floor space. This type of model can be written as:

$$log\left(P_i^{surf}\right) = \sum_{k=1}^{K} \beta_k X_{k,i} + \epsilon_i, \qquad \forall i, \tag{1}$$

where P_i^{surf} is the price per square meter of dwelling, $X_{k,i}$ are the different characteristics, β_k the sensibilities to the different characteristics, and ϵ_i a white noise.

2.2 Base des Notaires

The Base des Notaires – BIEN – contains more than 70% of the transactions in the Paris region. The notaries of Ile-de-France voluntary report new transactions in the database. In 2009, 82% of the transactions were reported in the BIEN database. It appears that there is no selection bias between transactions reported in the database and those which are not. The French Institute of Statistics (INSEE) has constructed an Index from this database, called the Notaire-INSEE Index, which will be detailed in the next section.

The database contains more than 3 million transactions since 1990. The Paris City and the Petite-Couronne, which consists of the adjacent *départements* – Paris, Hauts-de-Seine, Seine-Saint-Denis and Val-de-Marne – have been reported in the database since 1991. The Grande-Couronne, which consists of the most remote *départements* of the Paris region – Yvelines, Essone, Val d'Oise and Seine-et-Marne – have been reported since 1996.

The database contains many variables. Some variables relate to the characteristic of the buildings, for example the floor area, the number of rooms and of bathrooms, the presence of a swimming pool, a garden, a lift or of a garage, the floor if it is a flat, the period of construction. The base also contains the location of the buildings, in terms of coordinates, and of address.

Other variables concern the previous owner. Most notably, the database contains his socioprofessional category, his gender, and his age. The time of the previous transaction for the building is provided, which makes it possible to measure how long the last owner stayed.

The database contains the same variables for the new owner.

Finally, the price of the transaction, and the amount of tax paid is almost always reported correctly in the database.

Unfortunately, there are a lot of missing values for some variables, which will require making a trade-off between the variables chosen in the different regressions, and the number of observations available. More specifically, almost 50% of the transactions did not report the floor surface of the dwellings, 51% did not report the date of construction, but less than 1% did not report the number of rooms.

Some variables have not been coded correctly, making it impossible to make the difference between non-answer and 0 answer for example. This is the case for the number of garages, the number of bathrooms, or again the floor level.

Some assumptions can be made, for example a non-answer for the number of bathrooms could mean that there is actually no bathroom, but this is a hypothesis hard to test. We present below the methodology developed by INSEE to build indices of the real estate market in *Ile de France.*

2.3 Notaires-INSEE indices

2.3.1 Approach

The Notaires-INSEE Index is constructed from this database. It is based on hedonic regression techniques. The index aims to eliminate the quality effect and the non-representativeness of transactions, in order to catch the evolution of the market.

The objective has been to build indices that catch the trend of the real estate market in the different *départements* of the Paris region. The main index concerns apartments, more specifically *anciens appartements*, which means that new constructions are eliminated of the analysis. In the Paris region, "old apartments" account for more than 80% of all transactions since the last decade, and this number has constantly increased since the Second World War, when a majority of transactions concerned new dwellings. Moreover, the transactions selected in the construction of the index only concern private contracts between private individuals.

The standard approach is to find different homogeneous areas in the Paris region characterized a similar price dynamics, or a similar market. These regions are determined *a priori*, and we will denote them by s. We denote the price of a dwelling (i, s) at time t by $p^*_{i,s,t}$, and the characteristics of dwelling i such as the floor area, the number of rooms …, by $z_{i,s,t}$.

$$p^*_{i,s,t} \approx c(s, z_{i,s,t}) p^*_{s,t}, \qquad (2)$$

where $p^*_{s,t}$ is the implicit reference price in region s, and $c(s, z_{i,s,t})$ is a corrective coefficient which depends on the quality of the dwelling i. The corrective term and the reference price are defined up to a multiplicative scalar, that is why, in order to identify the model, one can set the reference price, for a given set of characteristics z_0, as:

$$p^*_{0,s,t} \approx p^*_{s,t} \Leftrightarrow c(s, z_0) = 1. \qquad (3)$$

The econometric model used is then:

$$log(p^*_{i,s,t}) = \sum_{k=1}^{K} \beta_{k,s} X_k(z_{i,s,t}) + log(p^*_{s,t}) + \epsilon^*_{i,s,t}, \qquad (4)$$

where $\epsilon^*_{i,s,t}$ are independent noises, with zero-means and $\eta^2_{s,t}$ variances.

$X_k(z_{i,s,t})$ are the explanatory variables for dwelling i, which takes into account its quality.

$log(p^*_{s,t})$ is here the intercept since it corresponds to the reference dwelling, from which the different coefficients $\beta_{k,s}$ are constructed

The Notaires-INSEE approach is then based on the three following steps:

- Step 1: Estimation of the corrective coefficients calculated from a set of predefined transactions, an estimation bundle. Choice of the reference bundle, for which we want to get the price evolution.
- Step 2: For each date t, recovery of the prices of the dwellings of the reference bundle from the observed transaction data, and the corrective coefficients estimated in Step 1
- Step 3: Construction of difference price indices from the previous estimations.

For example, in 2003, the estimation bundle consisted of the transactions which occurred between 1998 and 2001 in region s. Some transactions, for which there are missing values, are deleted from the estimation procedure. These goods will determine the corrective coefficients. The reference is a subset of the estimation bundle; the dwellings with extreme prices have been deleted.

To estimate the corrective coefficients, in this example, we consider Equation (4) for the transactions that occurred between 1998 and 2001 in region s, for which we have the couples: price $p_{j,s,t}$, quality $z_{j,s,t}$ for $j = 1, , \ldots J_{s,t}$, $s = 1, \ldots, S$, $t = 1998, \ldots, 2001$. The OLS method gives the parameters $\beta_{k,s}$, $k = 1, \ldots, K$, and $\log(p^*_{s,t})$.

$$\hat{c}_0(s,z) = \exp\left(\sum_{k=1}^{K} \beta_{k,s} X_k(z)\right). \tag{5}$$

If we now consider transactions at a different period, with new prices and qualities $p_{j,s,t}$ and $z_{j,s,t}$, we have:

$$\log(p_{j,s,t}) = \log\left(c(s, z_{j,s,t})\right) + \log(p^*_{s,t}) + \epsilon_{j,s,t}, \qquad j = 1, \ldots, J_{s,t}$$

$$\cong \log\left(\hat{c}_0(s, z_{j,s,t})\right) + \log(p^*_{s,t}) + \epsilon_{j,s,t}. \tag{6}$$

This last approximation would be equality, if the corrective coefficients did not vary over time.

The OLS approximation of $p^*_{s,t}$ is then:

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

$$\Leftrightarrow \hat{p}^*_{s,t} = \prod_{j=1}^{J_{s,t}} \left[\frac{p_{j,s,t}}{\hat{c}_0(s, z_{j,s,t})}\right]^{\frac{1}{J_{s,t}}}, \tag{7}$$

which is the geometric mean of the reference bundle expressed in prices of the reference quality dwelling.

2.3.2 Model developed

The actual model for the estimation bundle, which consists of all transactions between 1998-2001 in a given region s is given by:

$$\log(p_i) = \log(p_0) + \sum_{a=1}^{4} \alpha_a Y_{a,i} + \sum_{t=1}^{4} \theta_t T_{t,i} + \sum_{k=1}^{K} \beta_k X_{k,i} + \epsilon_i, \tag{8}$$

where p_i is the price per square meter of dwelling i

$Y_{a,i}$ and $T_{t,i}$ are dummies variables that take into account seasonal price variations

$X_{k,i}$ contains characteristics of good i

The characteristics are dummies variables, for example there will be one dummy variable for dwellings with 2 rooms, one for dwellings with 3 rooms …

The reference dwelling has been chosen to be a 3 rooms, 1st floor, one bathroom, sold in T4 1998 apartment. Its price is denoted by p_0. To determine its price for other periods between 1998 and 2001, one has to add the time dummies. Its price at time τ, $p_{0,\tau}$ will be the core of the price Index. The objective is then to determine its evolution. We can say that the transactions sold in period τ constitute the reference bundle. The first objective is to determine the price of the reference dwelling at the current period $\tau, p_{0,\tau}$. We assume that the β coefficients estimated in the estimation period with the estimation bundle remains constant in the period τ.

The subscript j referred to a dwelling exchanged at period τ, whereas i referred to a dwelling of the estimation bundle. An asset j sold in period τ satisfies:

$$\log(p_{j,\tau}) = \log(p_{0,\tau}) + \sum_{k=1}^{K} \beta_k X_{k,j,\tau} + \epsilon_{j,\tau}. \tag{9}$$

Contrary to equation (8), the seasonal terms have been dropped since we want the index to take into account these effects. The index constructed are quarterly index, therefore τ represents quarters here.

We introduce the price of the dwelling at period τ expressed in terms of the reference dwelling:

$$\log(\tilde{p}_{j,\tau}) = \log(p_{j,\tau}) - \sum_{k=1}^{K} \beta_k X_{k,j,\tau} \tag{10}$$

(when $\beta_k = 0\ \forall k$, we get the reference dwelling).

Then:

$$\log(\tilde{p}_{j,\tau}) = \log(p_{0,\tau}) + \epsilon_{j,\tau}. \tag{11}$$

The OLS method gives an estimation of the reference dwelling's price in period τ:

$$\log(\hat{p}_{0,\tau}) = \frac{1}{J_\tau} \sum_{j=1}^{J_\tau} \log(\tilde{p}_{j,\tau}). \tag{12}$$

This supposed that we know what the values of the β coefficients are. By assuming that there are stable over time, we can replace β_k by $\hat{\beta}_k$ that have been computed in the estimation period with the estimation bundle. This hypothesis has been tested, and it seemed to hold. The estimation period consists of a four years interval, and the current period follows the estimation period. The estimation period is updated at least once every five years, so that the stability of the coefficients is verified.

Introducing the $\hat{\beta}$ coefficients, we now have:

$$\log(\tilde{p}_{j,\tau}) \cong \log(p_{j,\tau}) - \sum_{k=1}^{K} \hat{\beta}_k X_{k,j,\tau} = \log\left[\frac{p_{j,\tau}}{\exp(\sum_{k=1}^{K} \hat{\beta}_k X_{k,j,\tau})}\right]. \tag{13}$$

From this equation, one can get the price at the current period τ of the reference dwelling:

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

$$\hat{p}_{0,\tau} = \left(\prod_{j=1}^{J_\tau} \tilde{p}_{j,\tau} \right)^{\frac{1}{J_\tau}}. \tag{14}$$

Notation: $\hat{\alpha}_{0,\tau} = \hat{p}_{0,\tau}$

This allows us to calculate the total value of the reference bundle at the current period τ. Reintroducing subscript s, for a good i of the reference bundle, its price at period τ can be estimated by:

$$\hat{p}^*_{i,s,\tau} = \exp\left(\hat{\alpha}_{0,s,\tau} + \sum_{k=1}^{K} \hat{\beta}_k X_{k,i,s} \right) \times A_{i,s}. \tag{15}$$

Where $A_{i,s}$ is the floor area of the dwelling.

The total value of the reference bundle or region s at period τ is obtained by summing the prices the N_s goods of the reference bundle calculated at period τ.

$$\widehat{W}_{s,\tau} = \sum_{i=1}^{N_s} \hat{p}^*_{i,s,\tau}. \tag{16}$$

The price Index of region s at period τ is then defined by:

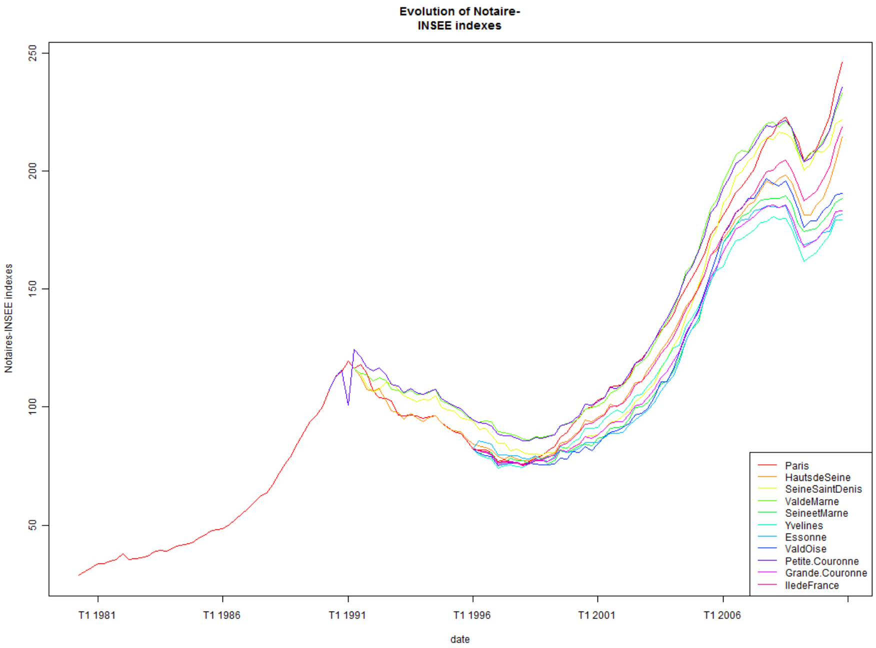
$$I_{t/_0}(s) = \frac{\widehat{W}_{s,\tau}}{\widehat{W}_{s,0}}. \tag{17}$$

For the Ile-de-France region, 62 regions have been determined, within which the elasticity of the different variables remains constant. Within the different regions, a number of sub-regions has been introduced to allow for different intercepts. The methodology used in the determination of the different regions is not detailed precisely, but they are supposed to represent similar real estate markets dynamics.

2.3.3 INSEE Indices

Price indices for *Ile de France* real estate market has been developed by INSEE, the French national institute of statistics. These indices are published quarterly, and are built on a hedonic approach. We discuss below the evolution of these indices. An index characterizes the market in the whole region, and different sub-indices are constructed for the different *départements*. They are calculating by aggregating the different index of the sub regions defined by (17).

Figure 1 Evolution of the Notaires-INSEE indices



We see that there is a strong correlation between the price indices of the different *départements*. The price indices have been normalized so that they start at the same level as the Paris Index, which can be seen as the reference index, because it is the oldest one. In T1 2010, we see that the Paris index is the highest of all indices, which confirms that Paris city prices have resisted quite strongly to the last crisis.

This graphic also shows that there has been a long upward trend in the Real Estate Prices in the Paris region, except for the crisis between 1991 and 1995, and the recent subprime crisis, whose effects was very short on the Real Estate Market in Ile-de-France.

The following table confirms the strong relationship between the evolutions of price indices in Ile-de-France. We calculated the correlations between the quarter to quarter evolution of the different indices, defined by, for each index m and each quarter q_t, q_{t+1}:

$$R_{t+1}{}^{m} = \log\left(\frac{I_{q_{t+1}}{}^{m}}{I_{q_t}{}^{m}}\right). \tag{18}$$

Table 1 Correlation between returns of price indices in IDF

	P	HdS	SSD	VdM	SeM	Y	E	VdO	PC	GC	IdF
Paris	1	0.79	0.72	0.78	0.72	0.78	0.7	0.77	0.82	0.82	0.97
HautsdeSeine	0.79	1	0.73	0.86	0.78	0.81	0.73	0.76	0.97	0.85	0.94
SeineSaintDenis	0.72	0.73	1	0.83	0.74	0.74	0.78	0.8	0.85	0.83	0.84
ValdeMarne	0.78	0.86	0.83	1	0.83	0.82	0.85	0.8	0.94	0.9	0.91
SeineetMarne	0.72	0.78	0.74	0.83	1	0.77	0.8	0.84	0.83	0.89	0.83
Yvelines	0.78	0.81	0.74	0.82	0.77	1	0.71	0.79	0.85	0.94	0.87
Essonne	0.7	0.73	0.78	0.85	0.8	0.71	1	0.8	0.81	0.88	0.8
ValdOise	0.77	0.76	0.8	0.8	0.84	0.79	0.8	1	0.83	0.92	0.85
Petite.Couronne	0.82	0.97	0.85	0.94	0.83	0.85	0.81	0.83	1	0.91	0.97
Grande.Couronne	0.82	0.85	0.83	0.9	0.89	0.94	0.88	0.92	0.91	1	0.92
IledeFrance	0.97	0.94	0.84	0.91	0.83	0.87	0.8	0.85	0.97	0.92	1

This table shows that the correlation between the term to term evolutions of the price indices is very high between the different *départements*, suggesting that the real estate market may be driven by a common factor influencing the different *départements.*

We will use these price indices with the model exposed in the next section to try to make forecasts of the real estate market.

3 Regime Switching Model

3.1 Presentation

A Hidden Markov Model is a doubly embedded stochastic process. The first process consists of the hidden sequence of states, which is not observed, but can be inferred from the other process which produces the sequence of observations. We denote the N states by $1, \dots, n, \dots, N$, and the state at date t by S_t. In the specific case of financial data, the state can be interpreted as "the state of the world" and essentially characterizes periods of high uncertainty or bubbles and periods of calm, or period of expansion and recession.

The *Markov* term in the name of the model refers to the definition of the process generating the hidden sequence of states. In this paper, we assume that this process is driven by a first-order Markov chain. A transition matrix drives the process, and the transition's rule between two states satisfies the Markov property:

$$P(S_{t+1} = n_1 | S_t = n_2,\ S_{t-1} = n_3 \dots) = P(S_{t+1} = n_1 | S_t = n_2) = Tr(n_2, n_1)$$

Where Tr denotes the transition matrix.

This transition matrix is a stochastic matrix:

$$\forall n_1 \in [\![1, N]\!], \sum_{n=1}^{N} Tr(n_1, n) = 1,\ \ \forall n_1, n_2 \in [\![1, N]\!], 0 \le Tr(n_1, n_2) \le 1.$$

We will discuss the case of Hidden Markov Models which are parametric random processes. In these models, the number of parameters is finite. Some of them will characterize the distribution of the process conditionally on the state, and the others will determine how the process switches from one state to another. The model depends on the number of states N, on the parameters of the conditional distribution, and on the initial probability. The parameters are represented by:

$$\theta = (Tr, B, \pi).$$

Tr is the transition matrix, B contains the parameters of the conditional density, and π represents the probability of the initial state.

We will denote the sequence of observations by $R = (R_1, \dots, R_T)$, where T is the length of the sequence. R will represent the returns of an asset, or the term to term evolution of the price index.

The density of the returns, conditionally on the state is supposed to be Gaussian here, and the density will be denoted by f_n. In this case, and when we consider one index, B contains the N means and the N variances parameters of the time series, in the different states.

The density at time t, once we have determined the probability of being in the different states, and for a set of parameters θ, is given by:

$$f\left(R_t|\overline{R}_{t-1},\theta\right) = \sum_{n=1}^{N} P\left(S_t = n\middle|\overline{R}_{t-1},\theta\right) * f(R_t\,|S_t = n,\theta,\overline{R}_{t-1}), \quad (19)$$

where f_n simply is the density of the Gaussian distribution in state n.

$\mathcal{R}_T$ is the notation employed to say that we use all the sequence data.

$\overline{R}_t$ will be used when we use all the data up to date t.

The log-likelihood of the model, which will be maximized to get the maximum likelihood parameters, is given by:

$$L(R|\theta) = \sum_{t=1}^{T} \log f\left(R_t|\overline{R}_{t-1},\theta\right). \quad (20)$$

The total number of parameters, when we work with N states and with M assets, is given by: $N \times M$ parameters for the different means, $(N-1) \times N$ parameters for the transition matrix (we assume that the states are the same date in each date, and their dynamic is therefore driven by the same transition matrix), $N \times M \times (M+1)/2$ parameters for the N variance-covariance matrices, and finally N parameters for the $N-1$ parameters for the probability of the different states in the first date. For instance, if $N = 2$ and $M = 1$, we have to estimate 7 parameters, when $N = 3$ and $M = 1$, 14 parameters, when $N = 2$ and $M = 2$, 11 parameters, and when $N = 3$ and $M = 2$, 20 parameters.

We see that, because of the quadratic term in M in the number of parameters for the variance-covariance matrices, the number of parameters grow fast with M. This is one inconvenient of this model, and it has to be dealt with when we want to model different assets jointly.

In order to estimate this model, we will use the Baum-Welch algorithm, which is a part of the Expectation-Maximization class of algorithms.

This main problem with this algorithm is that it can converge towards a local maximum. In order to find the global maximum of the likelihood function, we usually estimate the models with different initial conditions, and choose in the end the solution with the higher value.

The initial conditions can be drawn randomly, respecting the constraints for the different parameters, the initial transition matrix has to be a stochastic matrix for instance, and the variance-covariance matrices have to be definite positive. We can also set the different parameters with specific values. For example, for the means and the variances parameters, we can estimate the Gaussian Mixture Model on the observations in order to have a first guess of these parameters.

It can be noted that the Gaussian Mixture Model, where we assume the observations come from a mixture of Gaussian distributions, is a special case of the Hidden Markov Models. In fact, if we suppose that there are two states, and that the observations come from a mixture of Gaussians with weights ρ and $1-\rho$, the equivalent Hidden Markov Model is defined by a two states model, with the same Gaussian distributions in the two states, and by setting the initial probability equaled to the ergodic probability. If we denote the different transition coefficients by p_{11}, p_{22}, then the ergodic probability is given by:

$$\Pi = \begin{bmatrix} \frac{1-p_{22}}{2-p_{11}-p_{22}} \\ \frac{1-p_{11}}{2-p_{11}-p_{22}} \end{bmatrix}.$$

Starting with this initial probability we assure that the probability of being in the different states at each date will remain the same. Then, to obtain the same model as the Gaussian Mixture Model, we just have to choose the transition matrix coefficients such as $\rho = \frac{1-p_{22}}{2-p_{11}-p_{22}}$. Therefore, Hidden Markov Model is richer than a standard Gaussian Mixture Model. In particular, when predictions are made, the weights of the different Gaussian components vary over time.

3.2 Estimation of the model: the Expectation-Maximization algorithm

The Baum-Welch algorithm is iterative, and at each iteration, it gives parameters' updates. The algorithm is based on the inference that can be made at each period about the state of the process. We introduce the probability of being in the different states, knowing only the past, or the whole sequence. In the latter case, we say that they are *smoothed probabilities.*

We need a initial probability, $\hat{\pi}$, the probability of being in the different states at $t = 1$.

Then, a recursive algorithm gives us the probability sequence of being in the different states. This is the expectation step of the algorithm, which is based on Bayesian updating. It relies on the recursive calculation of $P\left(S_t = n \middle| \overline{R}_t, \theta\right)$ and $P\left(S_{t+1} = n \middle| \overline{R}_t, \theta\right)$.

$$P\left(S_t = n \middle| \overline{R}_t, \theta\right) = \frac{P\left(S_t = n \middle| \overline{R}_{t-1}, \theta\right) * f(R_t \mid S_t = n, \theta, \overline{R}_{t-1})}{\sum_{n=1}^{N} P\left(S_t = n \middle| \overline{R}_{t-1}, \theta\right) * f(R_t \mid S_t = n, \theta, \overline{R}_{t-1})}, \qquad \forall\, n \in [\![1, N]\!] \qquad (21)$$

where we can see that:

$$\left(S_t = n \middle| \overline{R}_{t-1}, \theta\right) * f(R_t \mid S_t = n, \theta, \overline{R}_{t-1}) = P(R_t, S_t = n \mid \overline{R}_{t-1}, \theta) \qquad (22)$$

and:

$$\left[P\left(S_{t+1}=n\middle|\overline{R}_t,\theta\right)\right]_{n=1\ldots N}=Tr'\times\left[P\left(S_t=n\middle|\overline{R}_t,\theta\right)\right]_{n=1\ldots N}. \quad (23)$$

Moreover, once these probability sequences have been determined, it is important to calculate the backward probabilities, also called the smoothed probabilities. The method has been developed by Kim (1993).

We have the recursive formula, that has to be applied for $t=T-1,\ldots,1$:

$$P(S_t=j|\mathcal{R}_T,\theta)=\sum_{i=1}^{N}P(S_t=j,S_{t+1}=i\ |\mathcal{R}_T,\theta), \quad (24)$$

where:

$$P(S_t=j,S_{t+1}=i\ |\mathcal{R}_T,\theta)=P(S_{t+1}=i|\mathcal{R}_T,\theta)\frac{Tr_{ji}*P(S_t=j\ |\ \overline{R}_{t-1},\theta)}{P(S_{t+1}=i\ |\ \overline{R}_t,\theta)}. \quad (25)$$

Then, conditionally on the probability of being in the different states, we can update the means and the variances parameters, along with the transition matrix. This is done in the maximization step. In this simple case, where we work with Gaussian distributions, the good thing is that we have analytical formulae to update these parameters. It has been showed that the likelihood obtained with this method does converge towards the likelihood of the model.

The update formulae for the means and the variances are given by:

$$\hat{\mu}(S=n|\mathcal{R}_T)=\frac{\sum_{t=1}^{T}P\left(S_t=n\middle|\mathcal{R}_T,\hat{\theta}\right)\times R_t}{\sum_{t=1}^{T}P\left(S_t=n\middle|\mathcal{R}_T,\hat{\theta}\right)}, \quad (26)$$

$$\hat{\sigma}^2(S=n|\mathcal{R}_T)=\frac{\sum_{t=1}^{T}\left[P\left(S_t=n\middle|\mathcal{R}_T,\hat{\theta}\right)\times R_t-\hat{\mu}(S=n\ |\mathcal{R}_T)\right]^2}{\sum_{t=1}^{T}P\left(S_t=n\middle|\mathcal{R}_T,\hat{\theta}\right)}, \quad (27)$$

$$\widehat{Tr}_{ij}=\frac{\sum_{t=2}^{T}P\left(S_t=j,S_{t-1}=i\ \middle|\mathcal{R}_T,\hat{\theta}\right)}{\sum_{t=2}^{T}P\left(\ S_{t-1}=i\ \middle|\mathcal{R}_T,\hat{\theta}\right)}, \quad (28)$$

$$\hat{\pi}=P\left(S_1=n\middle|\mathcal{R}_T,\hat{\theta}\right). \quad (29)$$

For the means and the variances parameters, the updating formulae are like the standard maximum likelihood formulae, except that we weight the observations by the probability that they are drawn in state n. The probability of the different states at time 1 is then the smoothed probability when $t=1$. The transition matrix parameters is calculated from the number of times it appeared the process has switched from state one state to the other, divided by the number of times the process has been in this state.

Dempster, Laird and Rubin have developed the Expectation-Maximization algorithm in 1977. At each iteration, the likelihood function increases. This assures the convergence of the algorithm.

3.3 Simulation of the model

Another interesting aspect of the model is that it is quite straightforward to draw Monte-Carlo simulations. To draw samples from a Hidden Markov Model, we need to have the different parameters, the transition matrix, the different means and variances, and finally the initial probability of the process.

Then, if we want to get Msimulations, we just have to follow this procedure, where H denotes the horizon of the simulation:

- If $h = 1$:
  - Draw the state according to the initial probability
  - Draw an observation with the parameters of the state
  - $h = h + 1$

- If $h > 1$ and $h \leq H$
  - Draw the state from the transition matrix: $P(S_h = n_1) = P(S_{h-1} = n_2, S_h = n_1)$ (it corresponds to the n_2 line of the transition matrix)
  - Draw an observation with the parameters of the state
  - $h = h + 1$

This is an intensive task which has to be repeated several thousand times in order to obtain good estimates of cumulative distributions for example. Nevertheless, the procedure is easy and allows making forecasts.

In the case of log-returns, there is an analytical formula for the cumulative returns, but in practice, the computational time is longer if we want to obtain cumulative returns' distributions for each date before the horizon H.

We first apply these models to the price index developed by INSEE, before presenting an example with a hedonic regression where we apply regime switching models to the intercept that have been estimated.

3.4 Application of Regime Switching Model to INSEE indices

The objective is to find a model that can fit the Notaire-INSEE indices for each *département*. For Paris, the length of the observation sequence is 123 trimesters, ranging from T2 1980 to T4 2010, whereas for the Petite-Couronne, it is only 79 trimesters, from T1 1991 to T4 2010, and 60 for the Grande-Couronne, from T1 1996 to T4 2010. The time series are not very long, but one could hope that they restitute the trend of the different markets faithfully, as they aim to eliminate most of the noise, most notably the quality heterogeneity in the data.

Different models have been tested. In the first case, we decided to assume that the dynamics of the different indices where independent, and depended only on its past evolution. We decided to model the term to term evolution of the different indices (returns of the indices).

One key aspect of Regime Switching Models is to determine the number of states. As the number of states grow, the number of parameters increases linearly. With this type of data, the problem is that the number of observations is quite low, so the addition of a new state will have a negative impact on the estimation of the different parameters. In order to select the best model, it can be interesting to look at Information Criterion, and more specifically the Bayesian Information Criterion.

The BIC for the Paris index with two states is much lower than the BIC when we estimated the model with three states. That indicates that the better fit given by the introduction of a third state is not sufficient to choose the three states model over the two states model. The same conclusion is found for the other indices.

We have summarized the estimations of the different models in the following table, for the different indices. We have to remember that, given the low number of observations available, the precision of the estimators is not very good, but the main objective here is to try to detect different states of the real estate market.

Table 2 Correlation between returns of price indices in IDF

	Paris	HdS	SSD	VdM	SeM	Y	E	VdO	PC	GC	IdF
Mean											
State 1	-1.59	-1.37	-1.00	-0.87	0.27	-1.78	0.28	0.17	-0.97	-0.61	-1.03
State2	3.13	2.50	2.75	2.60	3.22	2.06	3.97	3.40	2.57	2.40	2.50
Variances											
State 1	4.96	3.93	3.47	2.76	4.11	3.92	3.47	4.34	2.82	3.44	2.92
State 2	2.89	2.64	3.45	2.33	3.78	3.06	2.63	3.42	3.21	2.57	1.64
Transition											
State 1	0.91	0.94	0.95	0.95	0.96	0.80	0.97	0.97	0.95	0.90	0.85
State 2	0.96	0.97	0.97	0.95	0.97	0.97	0.88	0.95	0.95	0.97	0.98
Nb. Obs.											
	123	79	79	79	60	60	60	60	79	60	60

Three of the models estimated do not have a state with a negative mean. The three *départements* are part of the six for which only 60 observations are available. These time series do not begin before the

1991-1995 crisis, and therefore only catch the last financial crisis. The state with a mean around 0 has a higher variance than the other state for these three *départements*. Plotting the sequence of probabilities of being in the different states show that this state occurred in 1996 and 2007-2009, so the switch occurred when the prices started to stop increasing, just before the subprime crisis.

We can compare the results obtained above with those of a joint regime switching regime. For sake of simplicity, we only consider 5 *départements* from the *Petite Couronne*, those for which at least 79 observations are available. In each state, we assume that the returns of the indices follow a multivariate Gaussian distribution. Moreover, in each state, we allow for different correlations between the different indices. The results confirm the one obtained above, and show that it is a good approximation to consider that the states are common to the different *départements.* The following table reports the results of this model.

Table 3 Results of the switching regime model for the different *départements*

	Paris	HdS	SSD	VdM
Mean				
State 1	-1.65	-1.62	-1.64	-1.21
State2	2.64	2.33	2.29	2.24
Variances				
State 1	4.21	3.77	3.43	2.64
State 2	1.51	2.78	3.83	2.72
Transition				
State 1			0.93	
State 2			0.98	

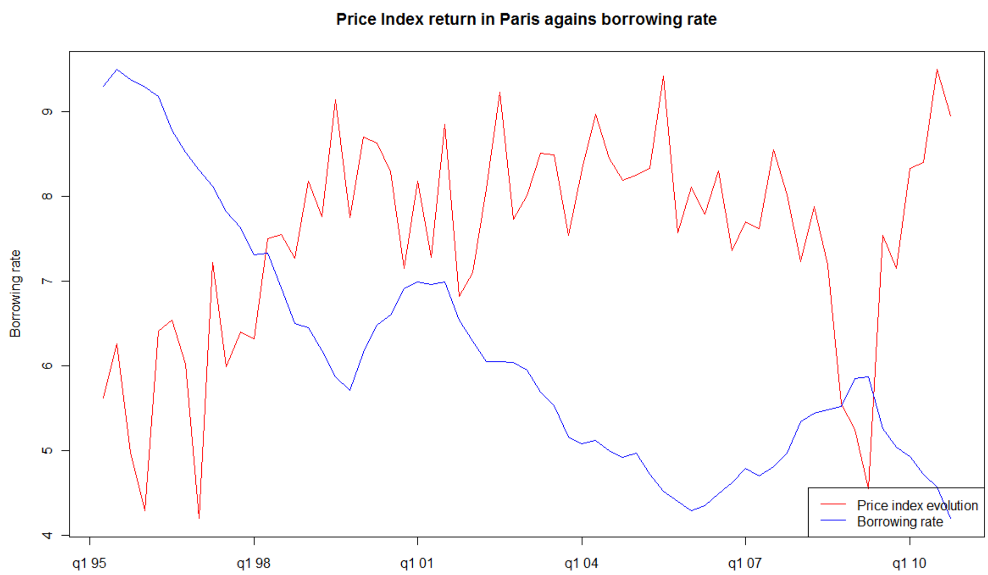
Correl S1	Paris	HdS	SSD	VdM
Paris	1	0.27	0.35	0.42
HdS	0.27	1	0.22	0.64
SSD	0.35	0.22	1	0.52
VdM	0.41	0.64	0.52	1

Correl S2	Paris	HdS	SSD	VdM
Paris	1	0.80	0.48	0.62
HdS	0.80	1	0.64	0.74
SSD	0.48	0.64	1	0.76
VdM	0.62	0.74	0.76	1

The different parameters estimated here are quite close to those found in the univariate case. There is however one more information conveyed by these estimations, namely the correlations between the different indices in each state. It appears that these correlations are much higher in state 2, the "good" state, than in the "bad" state. This would imply that periods of recession may exhibit less similitude among different *départements*.

A second model was developed, and uses an explicative variable: the borrowing rate. Other explicative variables have been tried. The correlation between the Notaires-INSEE index and the CAC 40, the main French stock index was tested, but no correlation was found. The correlation between the borrowing rate, available each quarter since 1995 and the real estate index is significant. A linear regression of the term to term evolution of the Paris index against the one quarter lagged value of the borrowing rate gives a R^2 of 0.4. One problem is that the average duration of the loans is not available, and it should be used along with the borrowing rate to have a better idea of the price really paid by buyers.

Figure 2 Price index return in Paris, and the borrowing rate



The model can be written as:

$$R_{t,m} = \alpha_m(S_t) + \beta_m(S_t)r_{t-2} + \epsilon_t, \quad (30)$$

where $R_{t,m}$ represents the "return" of the Index in region m, $\alpha_m(S_t)$ the mean in state S_t, $\beta_m(S_t)$ measures the impact of the borrowing rate in state S_t and ϵ the residuals, has the same variance σ_m^2 in the different states.

The regime switching model fitted with the Paris Index found two states.

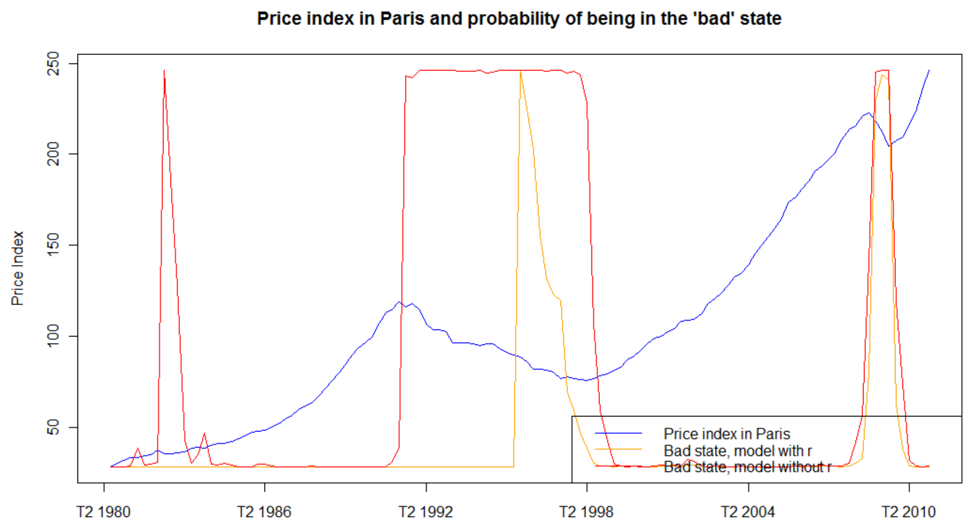
Table 4 Result of the Regime Switching Model with the borrowing rate

	Mean	Variances	Transitions	Probaf	Beta
State 1	-2.45	2.07	0.748	0.022	0.04
State 2	7.36	2.07	0.978	0.978	-0.867

The first state is the state of crisis, with a negative mean, and is less persistent than the second test. The effect of the borrowing rate in this state is 0, whereas, in the "good state", a decrease of the borrowing rate leads to an increase of the price index.

The states found in this model can be compared to those found in the previous model. The following graphic plots the probability of being in the different states over time:

Figure 3 Probability of being in the "bad" state



This illustrates that the states found in the two models are almost the same (the orange series only start in 1995). The main difference is that the model with borrowing rate switches to the "good" state before the other one in 1997.

These results are encouraging; they show that the regime switching models are able to take into consideration the different trends of the real estate market. When forecasting, this leads to a bimodal distribution for the near future, since there is a strictly positive probability to go back from one state to another, and the means are very different.

4 Regime Switching and Hedonic Model

4.1 Presentation

In this section, we introduce a new model that aims to take directly into account the temporal dynamic of the real estate market in the hedonic model, using the *Notaires* database.

For each *département*, we fit a specific model. We allow for a dummy variable specific to the communes to take into account different levels of prices. The coefficients relating the prices of the different attributes to the price of the whole dwelling are supposed to be constant over time in each *département.*

In fact, the regime switching approach is implemented in the intercept specific to the *départements.* Therefore, the regimes are specific to the different *départements*, but they can still be easily compared across *départements.*

The specification we use is the following:

$$log\left(P_i^{surf}\right) = \alpha_t + \sum_{k=1}^{K} \beta_k X_{k,i} + \epsilon_i, \qquad \forall i \in [\![1, N]\!]. \tag{31}$$

β_k represents the different parameters, and is assumed to be constant over time, $X_{k,i}$ represents the characteristics of the dwelling used in the hedonic regression, and P_i^{surf} represents the price per square meter of the dwelling. We use the logarithm of the price per unit of surface as it has been shown that this transformation decreases heteroscedasticity.

We estimate two types of model, where t will either represent quarterly, or monthly period of times. Taking a monthly time step gives less robust estimation for the parameters of the hedonic model, but gives more observations to be analyzed by a regime switching model. Taking quarterly period gives more robust estimation, but the number of observations is then very low for the regime switching model.

We assume that the time dummy variable has the following specification:

$$\alpha_t = \alpha_{t-1} + \mu^{S_t} + \sigma^{S_t}\epsilon_t, \tag{32}$$

where $\epsilon_t \sim N(0,1)$.

This specification is justified in the next part. The objective is to measure the trend of the real estate market in this variable. We allow for two different regimes, since the number of observations, even when taking monthly periods, is not sufficient to estimate correctly more regimes.

We run these types of models for the different *départements*, and we add a dummy variable for the *communes* in the regressions. We use different characteristics of the dwellings in the hedonic regressions: the number of rooms, the presence of a garage, a swimming pool, or a terrace and the floor level.

We focus on the transactions that are made between two consenting individuals or *gré-à-gré*. We eliminate the transactions for which the price or the floor area is not available. We also eliminate the transactions for which the data entered are outliers, for instance when the price is below 10 000 €. Assumptions have been made, for example a missing value for swimming pool is interpreted as no swimming pool for the dwelling consider.

4.2 Estimations

The estimations of the hedonic models in the different *départements* give R^2 around 0.6 in average. There is a difference between Paris and the other *départements* concerning the coefficients of the number of rooms. As the number of rooms increases in Paris, the price per unit of surface decreases, whereas the converse is true in the other *départements*. This can be viewed as a preference to big rooms in the suburbs, whereas in Paris, more and smaller rooms are preferred to fewer and larger rooms.

There is no other major surprise, the price increases with the presence of a terrace, a garage, the presence of a swimming pool. Prices tend to be slightly lower when the floor level increases.

The introduction of a dummy variable for the different *communes*, increases a lot the R^2 of the estimations, this shows the disparity of the real estate market across *communes*, even in the same *départements*.

We report in Appendix the time dummies for the different *départements*, for the monthly and quarterly data.

These variables exhibit a non-stationary behavior; it seems that a period of expansion follows a period of recession, in the early 1990's. This is confirmed by the historical crisis of the real estate market at that time, followed by a dramatic increase ever since. A first-difference is required to make the process stationary. We use a regime switching approach to take into account the different trends that appear in these variables.

A first important observation is that we do not have the same number of data in the different *départements*. For Paris, and the surrounding *départements* – 92, 93, 94 – data are available between 1991 and 2005. For the other *départements*, farther away from Paris: 77, 78, 91, 95, data are only available between 1996 and 2005. Data between 1991 and 1996 did not specify the quarter or the month of the transactions. We used the frequency of transactions between 1996 and 2005 to affect the different transactions to different months and quarters.

It can be observed that the time dummies have a decreasing trend between 1991 and 1997, and an increasing trend from 1997 onwards. Unfortunately, for the four *départements* for which data are only available since 1996, the decreasing trend is very short and only slight. The regime switching approach may not be very appropriate for these time variables.

Moreover, contrary to the previous study on INSEE indices, the last financial crisis that did affect the real estate market in Ile de France is not comprised in the data used.

We report below the different estimations with a quarterly time step:

Table 5 Parameters of the regime switching model of the time dummy variable for the different *départements* with quarterly data

Départements	*Mean 1*	*Mean 2*	*V 1*	*V2*	*Tr1-2*	*Tr2-1*
75	*0.029*	*-0.013*	*0.0003*	*0.0005*	*0.00*	*0.04*
92	*0.030*	*-0.008*	*0.0005*	*0.0008*	*0.00*	*0.03*
93	*0.037*	*-0.002*	*0.0006*	*0.0008*	*0.00*	*0.03*
94	*0.032*	*-0.045*	*0.0008*	*0.0009*	*0.00*	*0.03*
77	*0.027*	*0.008*	*0.0007*	*0.0005*	*0.00*	*0.05*
78	*0.026*	*0.0001*	*0.0006*	*0.0004*	*0.00*	*0.10*
91	*0.041*	*0.008*	*0.0007*	*0.0005*	*0.00*	*0.039*
95	*0.027*	*0.0015*	*0.001*	*0.0001*	*0.00*	*0.087*

The first two columns contain the means estimated in the two states. Column V1 and V2 contain the variances of the regime switching model, and the last two columns contain the transition probabilities between the two states.

We can see that the first regime has always the highest mean, is more persistent, and has a lower variance. It corresponds mostly to the period from 1997 onwards. The second regime has a negative mean for the three *départements* for which we have data since 1991, and includes the downward period between 1991 and 1997. For the other *départements* the mean is very close to 0 in the second regime and the variance is still higher than in the first regime. This shows that this model was able to capture the time varying component of the hedonic model.

The most striking result is that state 1 seems to be an absorbing state in most of the cases; the transition from 1 to 2 is very low. The model would predict that it is almost not possible to go from the "good state" to the "bad state". This is obviously not a very satisfactory result, but can be explained by the data which were used in this model. The data do not include a second period of decline in the real estate market, which would have allowed a significant transition probability between the “good” and the “bad” state.

If data after 2005 were used, there would have included the subprime crisis. This would have probably allow to estimate different parameters for the transition matrix, increasing the probability of switching from state 1 to state 2, and improve the prediction power of the model. This also shows that this kind of model “learns” with time, and must be used with a sufficiently large time window.

The second table contains the results of the estimation with monthly data:

Table 6 Parameters of the regime switching model of the time dummy variable for the different *départements* with monthly data

Départements	*Mean 1*	*Mean 2*	*V 1*	*V2*	*Tr1-2*	*Tr2-1*
75	*0.010*	*-0.0047*	*0.0004*	*0.0008*	*0.00*	*0.012*
92	*0.0047*	*-0.013*	*0.001*	*0.007*	*0.02*	*0.11*
93	*0.0020*	*0.0044*	*0.001*	*0.007*	*0.11*	*0.32*
94	*0.0048*	*-0.0036*	*0.001*	*0.0029*	*0.00*	*0.02*
77	*0.0075*	*0.0001*	*0.0006*	*0.004*	*0.022*	*0.10*
78	*0.029*	*-0.010*	*0.0004*	*0.0008*	*0.001*	*0.62*
91	*0.051*	*0.0022*	*0.0009*	*0.0001*	*0.017*	*0.14*
95	*0.0049*	*0.0089*	*0.0008*	*0.003*	*0.073*	*0.40*

Here again, the mean is overall higher in state 1 than in state 2, and the variance is much lower. Once again, state 1 is much more persistent than state 2, and is absorbent for two of the three *départements* for which data are available since 1991. For the other 2 *département*, the process switches several times during the different states, most particularly between 1991 and 1997, which explains the non-zero transition probability.

We can compare these results with those obtained if we simply added a time trend to the Hedonic model, instead of modeling a time component by a regime switching model. In this case, we obtain the following results:

Table 7 Trend coefficients o hedonic regression models for the different départements

	75	92	93	94	77	78	91	95
Quarterly	0.014	0.013	0.0092	0.011	0.018	0.020	0.016	0.018
Monthly	0.0048	0.0042	0.0031	0.0037	0.0059	0.0066	0.0054	0.0061

We can see that the results obtained by a simple linear trend are between the means in the two states obtained with the regime switching model. This shows that this kind of model take into account the different behaviors of the real estate market, and improve a simple hedonic model with linear trend.

As mentioned before, the most important caveat in the model developed here is the transition coefficients between the different states. We need to use more data in order to be able to measure the probability of switching from the “good” state to the “bad” state. The data used here did not contain such a period, therefore the prediction made in this framework will probably be too optimistic.

Nevertheless, if this type of model is estimated on data containing such periods, the price predictions are likely to be better than the ones given by a hedonic model with a time-trend.

5 Conclusion and Extensions

We have used hedonic regression models for the Paris region, and will continue to work on the development of a more sophisticated model that takes into account spatial correlation. Two options are available: either define regions with similar market dynamics, as done by INSEE, or take a specification that accounts for spatial correlation.

For example, Löchl and Axhausen (2009) applied the hedonic approach to residential rents in Zurich. We will try the different specifications on dwelling prices. They used three models: ordinary least square (OLS), spatial autoregressive (SAR) and geographically weighted regression (GWR). The OLS model lacks the ability to consider spatial dependency and spatial heterogeneity, consequently leading to biased and inefficient estimations.

SAR is a popular approach to incorporate spatial effects in a regression model. It assumes that the response variable at each location is a function not only of the explanatory variable at that location, but of the response at neighboring locations as well.

The GWR method, developed by Brunsdon, Fotheringham and Charlton (1998), attempts to incorporate geographical information into a regression model using a series of distance related weights. Essentially, it consists of a series of locally linear regressions that utilize distance-weighted overlapping samples of the data.

These developments could be then used with the regime switching approach presented here, and allow to make predictions in time and space for any type of dwellings.

Regime switching models have proved to be able to detect structural changes in the real estate market. Used with INSEE indices, they were able to detect rapidly switch in regimes. Moreover, using an explicative variable such as the borrowing rate confirmed the results found.
In the hedonic model developed, the time dummy variable exhibited behavior of a non-stationary variable which is suitable for fitting with a regime switching model. A larger window of observations would have allowed us to eliminate the absorbing states which were sometimes estimated. This type of models allow however to detect the different behaviors of the real estate markets comprised in the data.

6 References

Baroni, M., F.Barthélémy and M.Mokrane (2005) A PCA Factor Repeat Sales Index (1973-2001) To Forecast Apartment Prices in Paris (France), *ESSEC Working Papers* DR 05002, ESSEC Research Center, ESSEC Business School.

Brunsdon, C., A.S. Fotheringham and M. Charlton (1998) Geographically weighted regression – modelling spatial non-stationarity, *The Statistician*, **47** (3) 431-443.

David A., C. Gouriéroux and A. Laferrère (2002) Les indices de prix des logements anciens, *INSEE méthodes*, **98**.

Dempster A. P., N. M. Laird and D. B. Rubin (1977) *Journal of the Royal Statistical Society. Series B (Methodological)*, **39** (1) 1-38.

Hamilton, J.D. (1989) A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle, *Econometrica* **57**, 357-84.

Hamilton, J.D. (1994) *Time Series Analysis*, Princeton Univ. Press.

Kim, C-J (1993) Dynamic Linear Models with Markov-Switching, *Journal of Econometrics, Elsevier*, **60** (1-2) 1-22.

Löchl, M. and K.W. Axhausen (2009) Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, *Arbeitsberichte Verkehrs- und Raumplanung, 5XX*, Institut für Verkehrsplanung und Transportsysteme (IVT), ETH Zürich, Zürich.

Malpezzi, S. (2003) Hedonic pricing models: A selective and applied review, *Housing Economics and Public Policy*, 67-74, T. O'Sullivan and K. Gibb, Blackwell Science, Oxford.

Picard, N., C. Antoniou and A. de Palma (2010) Econometric Models, *SustainCity Working Paper*, **2.4**, THEMA, Université de Cergy-Pontoise.

Vuong, Q. (1989) Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses, *Econometrica*, **57**, 307-333.

7 Appendix

We represent below, for the different départements, the time dummy coefficients estimated in the hedonic regression models with quarterly periods. They exhibit non stationary behavior, which is why we take first order differences before estimating a regime switching model.

Figure 4 Time dummy coefficients in the hedonic regression models with quarterly periods

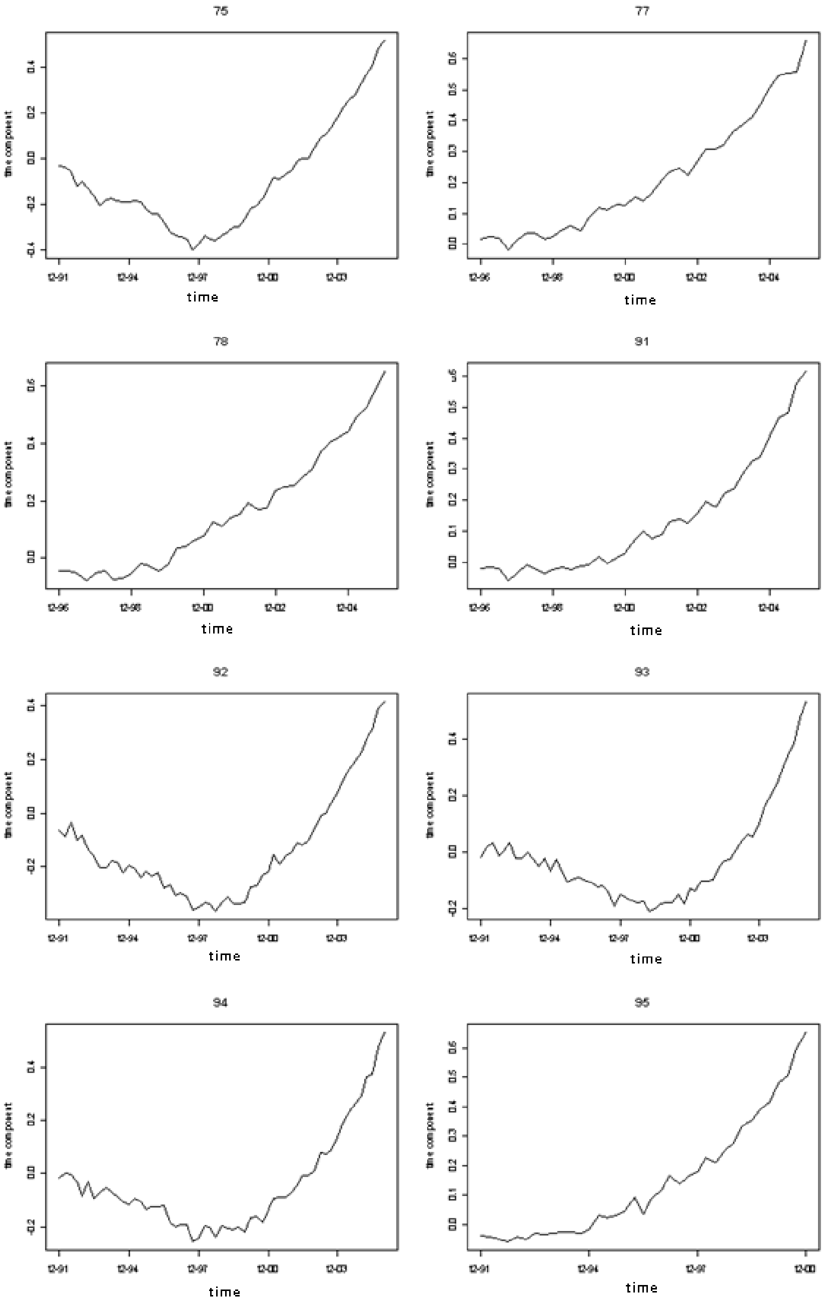


Figure 5 Time dummy coefficients in the hedonic regression models with monthly periods