TECHNISCHE UNIVERSITÄT WIEN

VIENNA UNIVERSITY OF TECHNOLOGY

DIPLOMARBEIT

Real Estate Price transmission in the City of Paris :

A comparison of spatial econometric methods

Ausgeführt am Institut für
Wirtschaftsmathematik
der Technischen Universität Wien

unter der Leitung von
ao. Univ.Prof. i.R Edwin Deutsch

durch

Alexandre Quettier
Matrikulnummer : 1328253
Wien, im September 2015

Eidesstaatliche Erklärung

Hiermit versichere ich, die vorliegende Arbeit selbstständig und unter ausschließlicher Verwendung der angegebenen Literatur und Hilfsmittel erstellt zu haben.
Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Ort, Datum Unterschrift

Abstract

This is an empirical study about the formation of homeownership prices in Paris. The main goal of the study is to determine to which extent the formation of prices is guided by selected city districts that influence the prices in other parts. For that purpose, we apply spatial econometric models at different scales to determine influent districts and justify the reasons for such influence. Using real estate data provided by the French statistical bureau INSEE, the study disposes of nearly 30.000 transactions in 2007.

Acknowledgments

First of all, I would like to thank my supervisor Professor Edwin Deutsch for guiding me and inspiring me all the way around during this thesis project period. I could not be more grateful for his detailed and elaborate feedbacks which help me self-improving and growing with this work.

Secondly I would like to express my sincere gratitude to Professor Gernot Tragler for all the courses that I attended during this two years in Vienna.

In addition, I must show my special appreciation to Professor Arnaud Simon from University Dauphine in Paris who showed continuous interest in this thesis work. He has always been willingly to share his wide knowledge and has contributed a lot to this work by having regular discussions with me and providing me with his innovative ideas.

I also want to thank Lisa Charlemagne, Alice Schoonbroodt, Jean-Charles Chapuis and all the people from Vienna who have helped me out during this journey in Austria.

Finally, special thanks to my parents, who always support me and encourage me.

Table of Contents

List of Tables

List of Figures

Introduction

The study investigates the ownership price formation process in the city of Paris. The topic is important because it touches not only the realm of marketing but even more the welfare conditions found in a metropolis. The realised prices of real estate sales depend on the characteristics of the objects under study, on locational characteristics and on personal attributes of the purchasers. The study disposes of a representative survey of nearly 30.000 transactions observed in Paris in 2007 that permits to analyse the price formation process in detail.

Simple OLS (Ordinary Least Squares) of prices against characteristics cannot provide satisfactory answers, because the sales prices do also follow a diffusion process such that the realised prices influence each other mutually. Prices formed in one part of the City may pull or push the prices in others parts, albeit the influence may decline with the distance.

The thesis aims at investigating that price diffusion by spatial econometric methods that are designed to uncover such mutual influences : Spatial Durbin Model (SDM) and his particular case, the Spatial Autoregressive Model (SAR).

The process called price diffusion can be evaluated by such spatial econometric models and the decisive parameter to be estimated is the spatial autocorrelation ρ, which captures the strength of that diffusion process. Much of the effort in this study was devoted to the appropriate design of a topographical grid of Paris that permits to apply the spatial analysis. Indeed, it was not possible to apply these models to nearly 30 000 individual observations. Another important part of the study was the choice of the significant characteristics in the formation of prices.

The study uses survey data provided by the French Notary Bureau BIEN and population data from the Statistical Institute INSEE. The Paris survey contains 28828 observations, with a degree of representation of 80% that is quite reasonable.

With the creation of an appropriate topographic grid, it was possible to evaluate the SAR-model to obtain an estimate ρ that characterises the diffusion. The innovative feature of the study is to evaluate the partial contributions of the districts to the overall parameter ρ. By ranking these contributions it is shown that several districts exert the strongest diffusion.

After working at the mesoscopic scale, the question of using individual data was a central issue in order to validate the results obtained with the grid. Several solutions are considered to use the information contained in individual data.

Structure

This thesis is structured as follows:

The first chapter gives a description of the data and an OLS-estimation by means of the survey, which reveals the properties of the selected data and calls for a spatial analysis.

The second chapter introduces the theoretical spatial model and describes the design of the grid, which permits highly instructive statistics of the distribution of prices and characteristics all over the city. The estimators and test statistics of the SAR and SDM methods are discussed and the ranking of districts according to their power of price diffusion is presented with some interpretations.

The third chapter provides methods for using the individual observations and allows to validate the results obtained in the second chapter.

Closing remarks and conclusions are summarized in the final chapter of this thesis.

CHAPTER I : Context and Data

The starting point of our study is the sample provided by the French Notary Bureau BIEN. With this representative survey of 28828 transactions observed in Paris during the year 2007, a hedonic approach was considered. Before looking at the sample, we will focus on the principles of this hedonic method.

The hedonic method was developed in the 1950s in the United States and is used to estimate economic values. It is most commonly applied to variations in housing prices that reflect the value of local environmental attributes. The basic premise of the hedonic pricing method is that the price of a marketed good is related to its characteristics.

This method is effective in the field of real estate for several reasons that we will explain.

Housing is a heterogeneous good. Each housing consists of a series of internal (size, number of rooms, type of heating, etc..) and external characteristics (accessibility, neighborhood, etc..). This why it is impossible to find strictly two identical housing and housing is considered as heterogeneous good.

A second reason is in the buyer's motivations in the acquisition of real estate. During the search of real estate, the buyer is looking for characteristics (internal and external) but also an investment that could bring him money during a possible resale. His choice will therefore be on maximizing preferences. The fact is that most buyers have similar preferences, and thus the housing having these

characteristics are the most popular and a priori the highest price. This partly explains the formation of prices by characteristic.
It is interesting to evaluate the importance and influence of each characteristic in the pricing of real estate. This helps to better understand consumer demand and adapt the construction of property to buyer preferences.

The first step of the method is to collect data on residential property sales in the region for a specific time period (usually one year). The required data include:

- selling prices and locations of residential properties
- property characteristics that affect selling prices, such as lot size, number and size of rooms, and number of bathrooms
- neighborhood characteristics that affect selling prices, such as property taxes, crime rates, and quality of schools
- accessibility characteristics that affect prices, such as distances to work and shopping centers, and availability of public transportation
- environmental characteristics that affect prices

The data are analyzed using regression analysis , which relates the price of the property to its characteristics and the environmental characteristic(s) of interest. Thus, the effects of different characteristics on price can be estimated. The regression results indicate how much property values will change for a small change in each characteristic, holding all other characteristics constant.

The analysis may be complicated by a number of factors. For example, the relationship between price and characteristics of the property may not be linear – prices may increase at an increasing or decreasing rate when characteristics change. In addition, many of the variables are likely to be correlated, so that their values change in similar ways. This can lead to understating the significance of some variables in the analysis. Thus, different functional forms and model specifications for the analysis must be considered.

In our case the number of available features is very high and the selection of significant characteristics is very important.

Sample

Our study disposes of a representative survey of 28828 transactions observed in Paris during the year 2007. The sample was supplied by University Dauphine which has collected the data from the French Notary Bureau BIEN and the statistical bureau INSEE.

For each observation we have information about :

- value of the transaction (total price and price per square meter)
- geographic situation of the real property (district, latitude and longitude)
- characteristics of the transactions and the estates

Those characteristics are classified into object characteristics, locational characteristics and personal characteristics of the buyers.

Objects characteristics :

- Surface
- Type of apartment
- Number of rooms
- Number of bathrooms

- Floor
- Period of construction
- Parking
- Elevator
- Basement
- Terrace
- Garden
- Extra Room

Locational and demographical characteristics (drawn from data offered by the Institute INSEE) :

- Log of the population per hectare (LNDENS) in 2008
- Annual growth rate of the population from 2008 to 2009 (POPGROW)

Characteristics of the buyers:

- Age
- Gender
- Profession
- Marital status

These data are available for each individual transaction. It is important to note that locational characteristics stated above are identical in each arrondissement, while the survey was undertaken in 2007, the population data referring to districts are available from the statistical bureau INSEE in the years 2008 and 2009.

The particular features of the Parisian market and the year 2007

The Parisian real estate market is a specific market and it is important to focus on its main features. It is so specific that its evolution is sometimes very different from what happens in the rest of France and even in the rest of the region "Ile-de-France". This is due to different characteristics :

- A very small area of just over 100 km² which is older than 150 years
- A very dense construction, which leaves little room to build new housing
- Housings mostly very old : 62% of them date from before 1949, 20% were built between 1949 and 1974. Thus 82% of the housing stock in Paris for over 40 years
- They are mainly very small units : studios and 2 rooms represent more than two thirds of 1.3 million apartments
- The average household size in Paris inhabitants is 1.9 inhabitants, it is much lower than the average in the "Ile-de-France" (2.33) or the national average (2.28).

The year 2007 is also specific since it marks the end of a period of rising prices, as shown in the following graphic (Figure 1). The following years (2008 and 2009) will be characterised by a decline in prices and transactions fall.

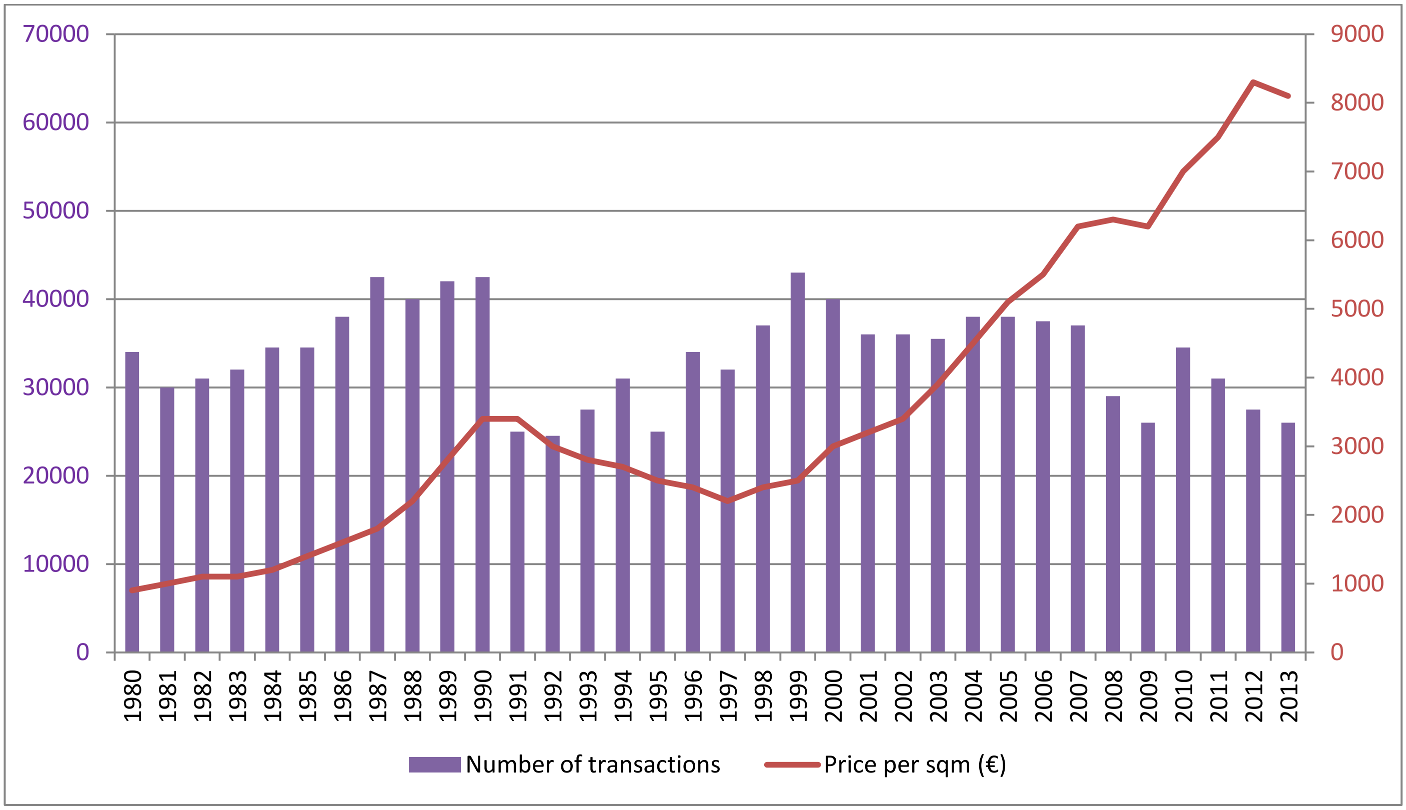


Figure 1 : Evolution of prices and transaction size in Paris

Descriptive statistics about the sample

The city of Paris is divided into twenty administrative districts, called "arrondissements" in French. We will use this term in the rest of the study.

The twenty arrondissements are arranged in the form of a clockwise spiral (often likened to a snail shell), starting from the middle of the city, with the first on the Right Bank (north bank) of the Seine. Paris is also composed of two wood : Bois de Vincennes in the East, Bois de Boulogne in the West.

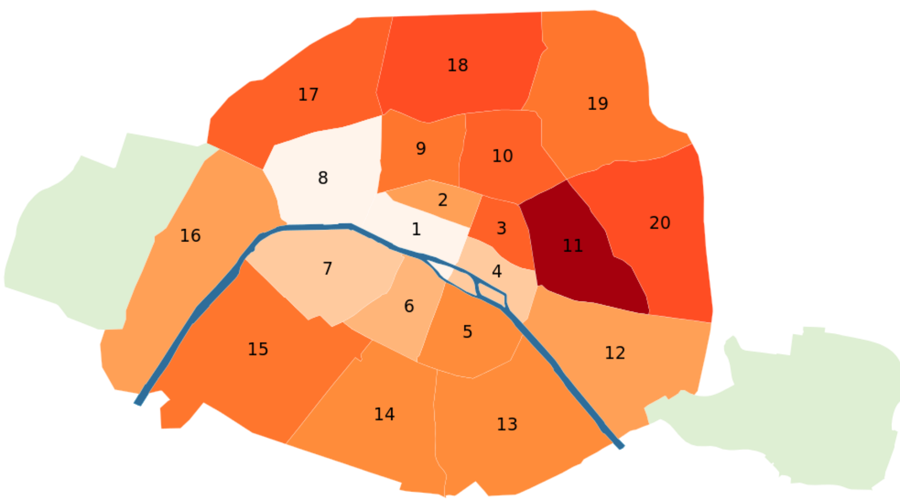


Figure 2 : Arrondissements of Paris

Arr.	Name	Surface (ha)	Population 2008	Density 2008
1	Louvre	183	17 584	9 609
2	Bourse	99	21 955	22 177
3	Temple	117	35 131	30 026
4	Hôtel-de-Ville	160	28 459	17 668
5	Panthéon	254	62 854	24 746
6	Luxembourg	215	45 147	20 999
7	Palais-Bourbon	409	57 895	14 155
8	Élysée	388	39 200	10 103
9	Opéra	218	59 840	27 450
10	Entrepôt	289	95 155	32 926
11	Popincourt	367	154 267	42 035
12	Reuilly	637	144 338	22 659
13	Gobelins	715	181 646	25 405
14	Observatoire	564	137 734	24 421
15	Vaugirard	848	236 490	27 888
16	Passy	791	167 384	21 161
17	Batignolles-Monceau	567	168 663	29 747
18	Buttes-Montmartre	601	197 173	32 807
19	Buttes-Chaumont	679	186 666	27 491
20	Ménilmontant	598	196 428	32 847
	Bois de Boulogne	846		
	Bois de Vincennes	995		
	Paris	**10 540**	**2 233 818**	**21194**

Table 1 : Population and density over Paris

Now we can look at some descriptive statistics about the sample. Because of the size difference between different districts, the number of observations in each district is quite variable. Figure 3 shows the most represented boroughs, who are naturally the largest districts and therefore located on the periphery.

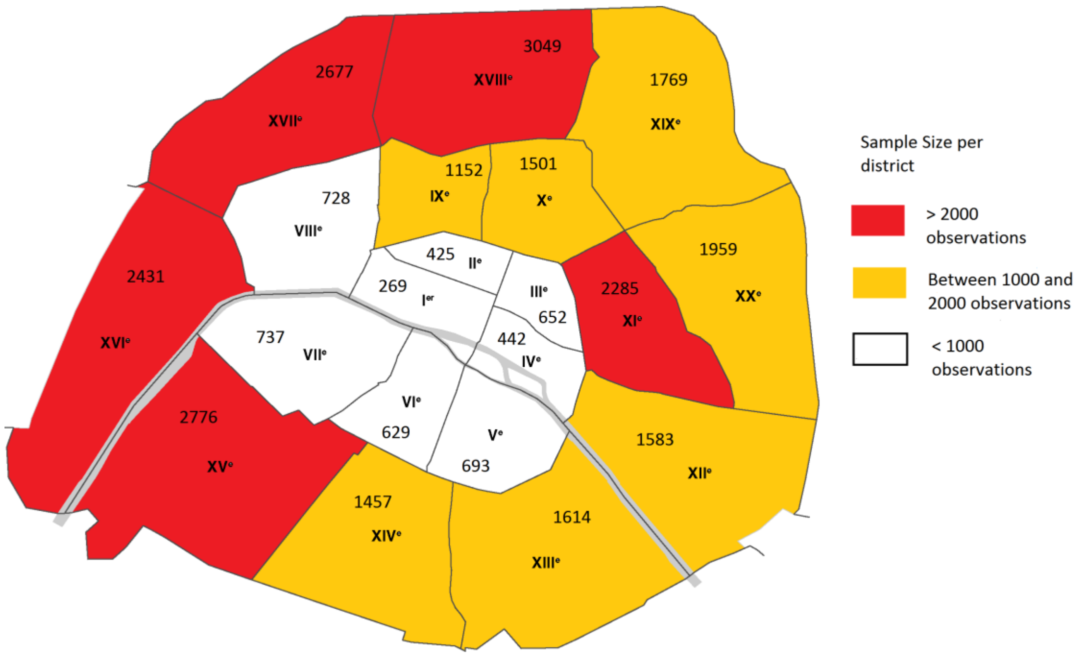


Figure 3 : Sample Size per district

Figure 4 provides the average price per square meter in each district. The districts located at the center / west side of the city are the most expensive and colored in red. Conversely, district northeast of the city are the cheapest and colored in blue. The districts with an average price between € 6,000 and € 7,000 per sqm are colored in white.

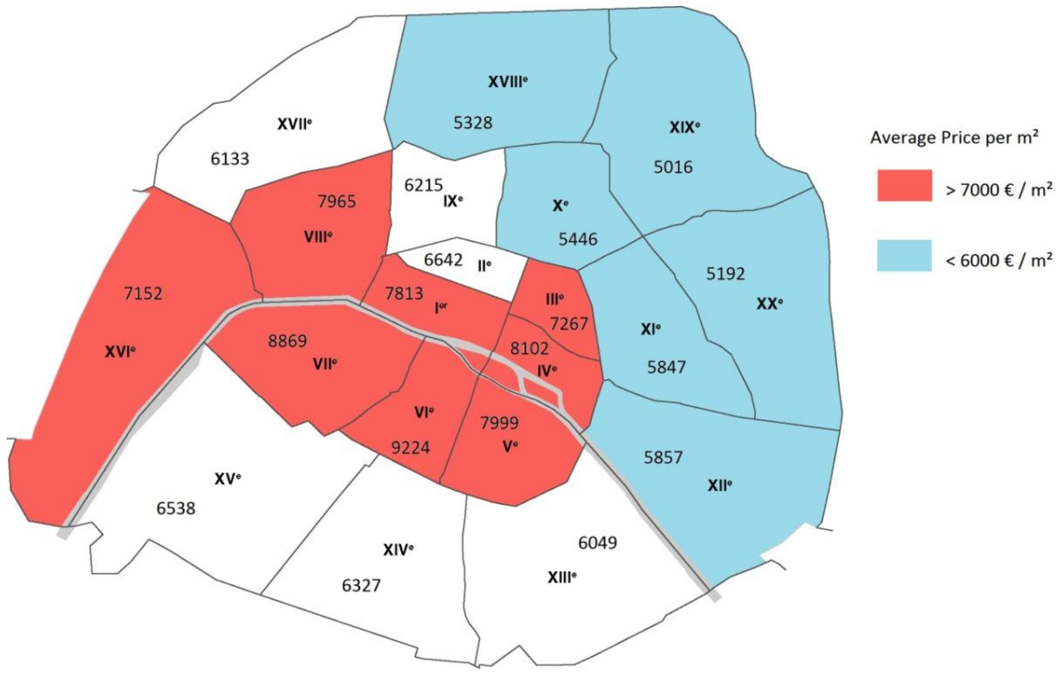


Figure 4 : Average Price per square meter

Selection of Criteria with OLS estimation

Because of the large number of characteristics in the sample, the first part of the study was to reduce the number of criteria to have a simpler model.

Initially we decided to perform an OLS estimation with the micro-data of the survey to determine the significant factors.

The prices and characteristics of objects and persons refer to the individual transactions. The characteristics of the estates and of the purchasers are 0/1 dummies. Instead the population data are the statistics observed in the 20 arrondissements.

After several tests, object characteristics selected for econometrics are:

- NEUF : age of the apartment if constructed after 1980
- SURF3 : the surface which was categorized initially into small (below 40 sqm), medium and large (above 100 sqm). Only the large one (SURF3) retained its significance also in the spatial setting (the medium size was excluded from estimation altogether).
- DUPLEX : an apartment being a duplex with two floors

Two personal characteristics of the buyers could be selected :

- OUVRIER : whether they are workers, craftsmen or retailers
- JEUNE : whether they are young people up to the age of 30

We also continue to consider the population data :

- LNDENS : Log of the population per hectare in 2008
- POPGROW : Annual growth rate of the population from 2008 to 2009

For descriptive purposes Table 2 contain some statistics about the sample. For each arrondissement and for entire Paris we have :

- Sample size
- Average Price per sqm
- Coefficient of variation : standard deviation / average in %
- Means of factors : the dummies were counted and divided by the sample sizes, to obtain the factor quotas observed in the respective arrondissements

	Sample	Average	Coeff. of	Means of factors by arrondissement				
Arronds	size	price m2	variation	NEUF	SURF3	DUPLEX	OUVRIER	JEUNE
1	269	7813,1	24,6	0,026	0,130	0,078	0,033	0,305
2	425	6642,0	26,8	0,024	0,064	0,068	0,073	0,362
3	652	7266,7	24,0	0,038	0,078	0,074	0,055	0,334
4	442	8101,7	25,8	0,025	0,093	0,102	0,061	0,290
5	693	7999,2	21,6	0,027	0,084	0,039	0,051	0,274
6	629	9224,5	26,9	0,027	0,151	0,064	0,029	0,316
7	737	8868,7	26,8	0,020	0,243	0,047	0,035	0,304
8	728	7964,7	29,7	0,054	0,383	0,044	0,034	0,396
9	1152	6211,9	22,3	0,043	0,117	0,030	0,081	0,333
10	1501	5446,1	21,3	0,033	0,075	0,035	0,079	0,358
11	2285	5847,2	19,2	0,053	0,048	0,030	0,088	0,323
12	1583	5857,0	20,2	0,066	0,044	0,016	0,106	0,289
13	1614	6048,9	23,8	0,126	0,035	0,022	0,094	0,297
14	1457	6327,2	19,9	0,076	0,048	0,025	0,078	0,278
15	2776	6538,3	19,0	0,062	0,057	0,026	0,087	0,251
16	2431	7151,9	23,5	0,042	0,298	0,038	0,049	0,292
17	2677	6133,1	22,4	0,033	0,118	0,028	0,089	0,347
18	3049	5328,4	24,3	0,036	0,021	0,017	0,120	0,321
19	1769	5016,0	20,7	0,133	0,028	0,020	0,146	0,292
20	1959	5192,2	19,2	0,084	0,026	0,031	0,121	0,281
Paris	28828	6282,6	28,1	0,057	0,093	0,032	0,087	0,308

Table 2 : Statistics of the sample

We can notice that the central districts have older housing that peripheral districts. This is not surprising since the construction of Paris occurred from the center to the periphery. Also, the share of workers is higher in the peripheral districts. This is consistent with the differences in prices between the center and peripheral districts.

The estimation shown in Tables 3 and 4 is performed by OLS, which does not take the spatial structure into account. Table 3 shows the results obtained for each arrondissement separately, where the population data had of course to be excluded. The last column lists the values of the coefficient of determination, denoted R2.

We use the following colors for the levels of significance :

- Green : 1%
- Yellow : 5%
- Blue : 10%
- White : Insignificant

Arronds	OLS estimators of factors by arrondissement						R2
	Constant	NEUF	SURF3	DUPLEX	OUVRIER	JEUNE	
1	7612,1	268,2	503,8	1089,2	349,1	104,3	0,038
2	6614,7	-695,1	233,4	721,1	-292,8	2,9	0,015
3	7210,5	224,4	688,1	1113,4	19,9	-267,4	0,051
4	8049,7	1245,7	648,5	536,9	-803,4	-154,5	0,036
5	7924,8	173,7	1182,6	896,8	-83,1	-219,2	0,055
6	9102,2	448,7	691,9	1094,3	-937,3	-117,4	0,029
7	8531,6	1271,2	953,4	1077,8	1,3	93,6	0,053
8	7533,8	1538,8	226,4	1116,3	-679,9	596,7	0,056
9	6218,9	299,4	44,5	791,6	-130,9	-116,1	0,014
10	5456,1	678,6	15,8	797,5	-269,8	-110,7	0,034
11	5838,9	572,0	147,7	697,1	-125,7	-120,6	0,034
12	5866,1	163,4	552,7	1278,8	-215,9	-144,3	0,040
13	5766,6	1643,9	628,3	665,6	-410,9	256,6	0,190
14	6339,5	512,1	303,7	616,7	-274,3	-216,9	0,035
15	6493,3	1085,1	570,1	566,3	-150,8	-224,6	0,084
16	6856,9	634,8	759,8	1093,3	-486,6	81,6	0,077
17	6079,5	312,4	847,6	844,0	-303,1	-152,4	0,069
18	5360,6	363,6	769,3	1073,0	-227,8	-163,7	0,032
19	5002,9	595,2	-5,2	422,1	-210,0	-149,9	0,053
20	5215,8	412,4	-355,9	1188,6	-87,3	-267,4	0,070
Paris	6184,7	420,9	1195,9	1043,0	-510,2	-85,3	0,070

Table 3 : OLS estimates based on survey sample without population data

In the table 4 follow the results for entire Paris, with and without the population data included.

	OLS estimators of individual price against factors									
	Const.	NEUF	SURF3	DUPLEX	LNDENS	POPGROW	OUVRIER	JEUNE	R2	AIC
Paris without	6184,7	420,9	1195,9	1043,0			-510,2	-85,3	0,070	14,88
Paris with population	19702,4	439,1	800,8	998,0	-2376,9	-1744,3	-399,1	-79,9	0,212	14,71

Table 4 : OLS estimates based on survey sample with and without population data

For entire Paris the OLS-estimators of the constant and factors are highly significant. Instead, the categorization into separate arrondissements yields mixed results, which nevertheless point to some issues that will be reconsidered in spatial estimation.

In the estimates by "arrondissement" only the constant appears highly significant throughout. Among the factor quotas, only the large majority of the duplex is highly significant. Instead the significance of the other factor quotas depends on the position of the districts. In the inner districts the estimators are less significant or insignificant, while the significance in the outer districts appears satisfactory. One may conclude that the arrondissements in the center are characterized by an offer that satisfies a great diversity of preferences and wealth endowments, while the offer in the outer arrondissements is more addressed to specific tastes and social strata.

This assertion gets support by the fact that in average most square meter prices in the "arrondissements" 1 to 8 are above 7000 EUR, while the "arrondissements" 12 to 20 show prices mostly below 6000 EUR (see Figure 4). These thresholds will play an important role in spatial estimation. Moreover, the spread of prices can be captured by the coefficient of variation. In that respect the inner arrondissements show a larger spread than the outer ones. This may be one reason why the estimators in the outer districts are more significant.

Looking at the goodness of fit, the R^2-s appear rather poor. This is not astonishing because in large surveys the variables often exert a high volatility.
The critical point are the magnitudes of the estimators. The constant absorbs most of the explanatory power, by taking values close to the average prices. Instead the factor estimators are smaller in size, mostly positive for the object characteristics NEUF, SURF3 and DUPLEX, and negative for the personal characteristics OUVRIER and JEUNE.

This pattern does not only show up among the arrondissements, but also for entire Paris, see Table 4. In this Table, for entire Paris the population statistics of density and growth by arrondissement are included. Here we detect an extremely interesting fact.

By inclusion of population data all the estimators remain highly significant, and the $R^2 = 0.21$ shows a substantial increase in the goodness of fit. The improvement is apparently created by a trade-off between a much larger positive constant and equally large negative estimators of LNDENS and POPGROW. In contrast, most factor quotas are quite close to the case without population data. The only substantial deviation is observed for the large surface SURF3, whose estimator declines from 1195.9 to 800.9. The reason can be found in moderate multicollinearity.

Most partial correlations of the selected variables are significant but small in magnitude, see Table 5. Only the factor SURF3 is somewhat stronger correlated with LNDENS, with a correlation of -0.21.

Instead all other factors are less correlated with LNDENS, and the entirety of factors is weakly correlated with POPGROW. Hence one may conclude that the loss of magnitude in the estimator of SURF3 is generated by inclusion of LNDENS.

	PRIXM2	NEUF	SURF3	DUPLEX	LNDENS	POPGROW	OUVRIER	JEUNE
PRIXM2	1							
NEUF	0,0645	1						
SURF3	0,2202	0,0191	1					
DUPLEX	0,1376	0,0386	0,1509	1				
LNDENS	-0,3997	0,0120	-0,2119	-0,0396	1			
POPGROW	-0,0324	-0,0314	0,1246	0,0064	-0,1889	1		
OUVRIER	-0,0962	-0,0069	-0,0697	-0,0196	0,0619	-0,0141	1	
JEUNE	-0,0267	-0,0274	-0,0347	-0,0150	-0,0015	0,0252	-0,0684	1

Table 5 : Partial correlations between survey variables
Significance along Pearson Method : green 1% ; yellow 5%

There is a straightforward interpretation : apartments above 100 sqm are observed more often in arrondissements with lower population density, compare the means of factors in Table 2. Moreover, the partial correlation between LNDENS and POPGROW is negative. Thus the people who move to less densely populated arrondissements may increase the demand for larger apartments. By and large, new and large apartments realize selling prices above the average, while workers, craftsmen, small merchants and young people search apartments priced below the average.

To sum up the present section, the inclusion of population data does probably not invalidate the OLS-approach. But the size of the constant relative to the factor estimators gives rise to the hypothesis that there is a diffusion process that transmits the formation of prices from one arrondissement to the other ones. To show that, we have to use spatial econometric methods and to design a topographic grid covering Paris, which permits to study such a process in detail.

CHAPTER II : Comparison of spatial econometric methods

Some facts about spatial autocorrelation

Anselin and Bera (1998) propose an intuitive definition of spatial autocorrelation : "*Spatial autocorrelation can be loosely defined as the coincidence of value similarity with locational similarity*". In other words, the positive spatial autocorrelation is characterized by a tendency to concentrate in the area high or low values of a random variable.
In contrast, the negative spatial autocorrelation means that each location tends to be surrounded by neighboring locations which takes on very different values. Finally, the absence of spatial autocorrelation indicates that the spatial distribution of the variable values is random.

The presence of spatial autocorrelation for a variable means that there is a functional relationship between what happens at a point in space and what happens elsewhere.

Spatial autocorrelation differs from the temporal autocorrelation which is unidirectional as only the past influences the future. However, spatial autocorrelation is multidirectional since everything is connected to everything. This generalized interdependence has the effect of making more complex treatment methods of spatial autocorrelation. For example, some valid estimation methods for time series are not directly transferable to the space case.

To capture the interdependence between regions, we must consider their relative positions. For that, one must specify exogenously topology space system by constructing a matrix of weights W. This matrix is a square matrix having as many rows and columns as there are geographical areas (note N the number of regions), where each term **wij** of the matrix represents how region i and region j are connected spatially.

Theoretical models

The first model chosen to study the distribution and diffusion of prices within the city of Paris is a Spatial Autoregressive Model (SAR). This model is a special case of the Spatial Durbin Model (SDM) which will be discussed later.

This model says that levels of the dependent variable y depend on the levels of y in neighboring regions. It is thus a formulation of the idea of a spatial spillover, which characterize how the regional prices influence the prices in their neighborhood. The model maps a diffusion process where the price in one cell influences the prices in other cells.

The starting point is a classic linear regression model :

$$y = X\beta + u \quad , \qquad u \sim \mathrm{i.i.d.}\,(0, \sigma^2\, I_n)$$

where y is endogenous prices, X explanatory factors and u errors independent and identically distributed (i.i.d.)

One way to take into account spatial autocorrelation is to add an offset endogenous variable Wy.

Indeed, when the weight matrix W is a standardized matrix, the ith element of the spatial lag Wy contains the weighted average of observations from neighboring regions to region i.

The formal model is

$$\mathbf{y} = \boldsymbol{\rho}\mathbf{W}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \quad , \qquad \mathbf{u} \sim \mathbf{i.i.d.}\,(\mathbf{0}, \boldsymbol{\sigma}^{2}\ \mathbf{I_n})$$

where ρ is the spatial autocorrelation, y endogenous prices, X explanatory factors and W the weight matrix. This symmetric matrix W characterizes the type of neighborhood.

The parameter ρ is a coefficient on the spatially lagged dependent variable, Wy, and the parameters β reflect the influence of the explanatory variables on variation in the dependent variable y.

In full notation, the SAR Model is :

$$y = \begin{bmatrix} y_1 \\ \\ \\ y_N \end{bmatrix}, \quad W = \begin{bmatrix} w_{11} & \cdots & \cdots & w_{1N} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ w_{N1} & \cdots & \cdots & w_{NN} \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{1K} \\ \vdots & \vdots \\ \vdots & \vdots \\ x_{N1} & x_{NK} \end{bmatrix}, \quad u = \begin{bmatrix} u_1 \\ \\ \\ u_N \end{bmatrix}$$

A grid over Paris

For implementation reasons, it was impossible to directly apply this model with more than 28,000 individual observations. On the other hand, a grid by district is not satisfactory. That is why we decided to work at a meso scale and partition the city of Paris in a grid of 360 cells :

- 18 vertical index
- 20 horizontal index

The cells are nearly quadratic with 550m side length.

With this grid, the N equals 360 = 18 * 20, corresponding to 18 vertical grid cells times 20 horizontal grid cells. The matrix W has 129600 components accordingly. In estimation, cells without observation were sorted out.

Why this choice of 360 cells ?

The surface of Paris is an oval with 8.40 km vertical length (i.e. North-South) and 10.85 km horizontal length (i.e. West-East). Such a grid (20*18) allows both to have more detailed vision, but also a sufficient number of observations in each cell.
The cells are nearly quadratic with around 500m side length. It should be remarked that the choice of 20 horizontal cells has nothing to do with the 20 arrondissements of Paris.

We associate with each cell an average price and a district, which is the most represented in the cell (the exact location of each observation is given in the sample).

The topographical location of the observations is contained in the survey data. The relation between the grid and the arrondissements is shown in Table 6. The majority of cells belongs uniquely to some arrondissement. On the borders between two arrondissements the correspondence is not unique. For descriptive purposes in the Table 6, these cells are assigned to that arrondissement, which provides the majority of observations in the survey.

The river Seine is marked with light blue. The part of Paris located south of the Seine is called "rive gauche", the northern part including the light blue cells is the "rive droite".

NRV	NRH1	NRH2	NRH3	NRH4	NRH5	NRH6	NRH7	NRH8	NRH9	NRH10	NRH11	NRH12	NRH13	NRH14	NRH15	NRH16	NRH17	NRH18	NRH19	NRH20
18													18							
17								17	17	18	18	18	18	18	19	19	19			
16						17	17	17	17	18	18	18	18	18	19	19	19			
15				17	17	17	17	17	17	18	18	18	18	18	19	19	19	19		
14				17	17	17	17	17	17	9	18	9	10	10	19	19	19	19		
13				17	17	17	8	8	8	9	9	9	10	10	10	19	19	19	19	
12			16	16	16	8	8	8	8	9	9	9	10	10	10	19	20	20	20	20
11		16	16	16	16	16	8	8	8	1	2	2	3	3	11	11	20	20	20	20
10		16	16	16	16	7	7	7	7	1	1	1	3	3	11	11	11	20	20	20
9	16	16	16	16	16	7	7	7	7	7	6	1	4	4	11	11	11	20	20	20
8	16	16	16	16	15	15	7	7	7	6	6	5	4	4	11	11	11	11	20	20
7	16	16	16	15	15	15	15	7	6	6	5	5	5	4	12	12	12	12	12	20
6	16	16	15	15	15	15	15	15	14	6	5	5	5	5	12	12	12	12	12	12
5	16	16	15	15	15	15	15	15	14	14	14	13	13	13	13	12	12	12	12	12
4				15	15	15	15	14	14	14	14	13	13	13	13	13	12	12	12	
3						15	14	14	14	14	14	13	13	13	13	13	13			
2									14	14	13	13	13	13	13	13				
1											14	14		13						
NRV: Vertical grid index				NRH*: Horizontal grid index						Light blue : Border to river Seine						Dark blue: Île de la cité				

Table 6 : District of each cell

The sample sizes across the grid are shown in Table 7. The samples are unevenly distributed what follows from the magnitudes of the arrondissements and their populations. There are only a few cells with little observations. These cells often correspond to the boundaries of the city or the cells located "on the Seine".

NRV	NRH1	NRH2	NRH3	NRH4	NRH5	NRH6	NRH7	NRH8	NRH9	NRH10	NRH11	NRH12	NRH13	NRH14	NRH15	NRH16	NRH17	NRH18	NRH19	NRH20
18													2							
17								35	80	87	148	129	11	21	8	4	7			
16						3	9	83	465	284	320	424	69	184	116	248	38			
15				10	55	68	176	179	397	189	188	338	246	149	50	128	94	49		
14				69	185	161	96	211	187	261	255	253	37	198	125	145	119	37		
13				66	163	175	39	67	64	88	235	255	130	104	94	143	170	99	131	
12			17	99	99	51	95	75	62	33	87	147	272	204	201	167	141	216	127	20
11		5	106	123	104	91	68	3	15	51	69	219	195	208	257	295	103	185	197	21
10		20	147	114	24	44	55	10	10	22	84	110	213	178	198	229	84	44	166	17
9	2	89	139	165	23	77	182	35	45	58	69	50	133	129	190	269	193	102	112	51
8	40	135	132	53	157	89	34	55	35	96	143	87	47	92	117	181	220	172	173	24
7	82	104	109	96	215	170	103	89	119	65	56	46	70	16	106	191	127	136	109	45
6	150	133	22	167	180	189	164	97	79	161	58	95	108	39	3	15	103	90	93	44
5	105	33	6	110	191	179	158	47	101	34	39	140	170	32	19	1	64	187	189	22
4				22	139	92	142	144	168	136	89	141	76	93	55	40	6	60	47	
3						21	52	95	165	198	38	103	142	52	125	17	3			
2									65	44	64	116	86	179	60	5				
1											1	16		15						

Table 7 : Sample Size of each cell

The most interesting point are the related sqm prices shown in Table 8. They are calculated from the survey as cell averages of the prices.

The highest prices above 8500 EUR per sqm are located in the centre of Paris. They are contained in a strip along both sides of the Seine, reaching from the Champs Elysées in the West to the Quartier Marais in the East. High prices on the rive gauche are observed in the arrondissements 5, 6 and 7.

Contrary what one might believe, the 16th arrondissement in the West is characterised by prices that do not much exceed the average of Paris. In fact there are two parties in the 16th district. Only the cell near the Place de l'Étoile has a price of 9450 EUR per sqm, and therefore comes close to the highest price levels.

NRV	NRH1	NRH2	NRH3	NRH4	NRH5	NRH6	NRH7	NRH8	NRH9	NRH10	NRH11	NRH12	NRH13	NRH14	NRH15	NRH16	NRH17	NRH18	NRH19	NRH20
18													4,8							
17								4,7	5,2	5,1	4,9	4,9	4,7	3,8	7,3	4,9	4,0			
16						5,0	5,6	4,9	5,3	5,7	5,7	5,3	4,4	4,3	4,5	4,9	4,8			
15				8,3	6,2	7,0	6,3	6,1	5,9	6,0	6,9	5,3	4,5	4,6	4,8	4,9	4,8	4,5		
14				6,0	6,7	6,8	6,9	6,5	6,3	6,1	6,2	5,7	4,9	5,1	5,1	5,8	5,2	4,8		
13				6,7	7,1	7,1	7,7	7,3	6,6	6,5	6,4	6,1	5,4	5,4	5,1	5,4	5,4	4,6	4,7	
12			8,3	7,6	7,7	9,2	8,1	7,7	8,0	6,5	6,1	6,0	5,4	5,7	5,2	5,0	5,2	5,1	5,1	4,6
11		6,9	7,6	7,6	7,6	8,1	10,7	9,8	9,1	8,8	8,1	6,5	6,1	6,4	5,7	5,4	5,0	5,3	5,3	4,8
10		7,2	8,0	7,7	7,4	9,2	9,0	9,6	9,7	7,8	8,1	6,9	7,4	7,3	6,1	5,8	5,7	5,6	5,4	4,9
9	9,5	7,2	7,4	7,4	8,9	8,4	8,2	9,2	10,5	9,6	9,2	7,2	8,1	8,4	6,5	6,0	5,7	5,3	5,0	4,5
8	7,0	6,9	6,9	7,0	7,3	7,7	8,3	8,9	10,5	10,2	9,9	9,2	10,4	7,7	6,8	6,2	5,9	5,6	5,4	5,2
7	6,8	7,1	6,4	6,2	6,8	6,7	7,6	8,2	8,4	9,4	8,6	8,5	8,1	7,1	6,2	5,9	6,1	5,9	5,8	5,5
6	6,3	6,5	6,3	6,4	6,4	6,5	6,5	7,1	6,8	8,0	7,9	8,2	7,9	7,2	6,9	6,2	5,5	5,9	6,0	5,4
5	5,7	5,8	6,3	6,1	6,4	6,5	6,2	7,0	6,8	7,6	7,7	7,2	6,9	6,1	6,1	7,2	5,7	5,8	5,9	4,9
4				5,7	6,3	6,1	6,1	6,0	6,4	6,6	6,4	7,0	6,2	6,4	5,6	8,2	6,4	5,0	5,5	
3						5,5	5,5	5,8	6,1	6,2	5,9	6,2	6,1	5,5	5,5	7,0	7,7			
2									5,7	6,2	6,7	5,9	5,8	4,9	4,6	5,4				
1											6,4	6,1		4,9						

Table 8 : Average Price of each cell (in 1000€/sqm)

The prices in the outer districts are much lower, in particular in the Eastern parts of Paris. The lowest price levels are observed in the arrondissements 11, 12, 13, 19 and 20. But interestingly enough, the spatial estimation will demonstrate that a few low price districts, the 12th arrondissement in particular, exert a strong price diffusion. Indeed, the results will reveal two types of strong diffusion: The one is created by "high price arrondissements" that raise the price level over all Paris, but there is a countervailing process generated by "low price arrondissements" that works in the opposite direction.

This process is controlled by the transaction conditions that prevail in the arrondissements. Although it is difficult to know the exact number of transaction, about 36000 transactions are estimated in Paris in 2007. The BIEN survey is therefore well stratified and reasonably representative at nearly 80%. Therefore we can draw useful information from relative data.

For that purpose we construct a variable that captures the relative magnitudes of transactions across the 20 districts. It is the ratio between the shares of the sample sizes of transactions (SHARESAMP), which measure the flows in 2007, and the shares of the district populations (SHAREPOP), which measure the stocks in 2008. The result is a variable called relative transaction size TRANSR, and shown in Table 9.

DISTRICT	DENSITY	POPULATION	WLNDENS	WPOPGROW	SHAREPOP	SHARESAMP	TRANSR	LRELPRIX
1	96,09	17584	4,735	1,20	0,787	0,933	1,185	0,2180
2	221,77	21955	5,275	2,54	0,983	1,474	1,500	0,0556
3	300,27	35131	5,698	2,96	1,573	2,262	1,438	0,1455
4	176,68	28268	5,233	0,95	1,265	1,533	1,212	0,2543
5	247,46	62854	5,499	-1,08	2,814	2,404	0,854	0,2416
6	209,99	45147	5,328	-2,01	2,021	2,182	1,080	0,3841
7	141,55	57895	4,990	0,62	2,592	2,557	0,986	0,3448
8	101,03	39200	4,644	4,07	1,755	2,525	1,439	0,2372
9	274,50	59840	5,619	2,08	2,679	3,996	1,492	-0,0113
10	329,26	95155	5,790	1,58	4,260	5,207	1,222	-0,1429
11	420,35	154267	5,976	0,22	6,906	7,926	1,148	-0,0718
12	226,59	144338	5,444	0,20	6,461	5,491	0,850	-0,0701
13	254,05	181646	5,537	1,19	8,132	5,599	0,689	-0,0379
14	244,21	137734	5,502	0,52	6,166	5,054	0,820	0,0071
15	278,88	236490	5,612	1,03	10,587	9,630	0,910	0,0399
16	211,61	167384	5,358	2,61	7,493	8,433	1,125	0,1296
17	297,47	168663	5,636	1,13	7,550	9,286	1,230	-0,0241
18	328,08	197173	5,781	2,31	8,827	10,577	1,198	-0,1647
19	274,91	186666	5,630	0,00	8,356	6,136	0,734	-0,2251
20	328,48	196428	5,784	1,10	8,793	6,795	0,773	-0,1906

Table 9 : Population and transaction intensity

The potential factors that influence the transaction size are the sqm price and the population variables. The sqm price is the log of the ratio between the district average divided by the total average, labelled as LRELPRIX. The population variables are the log density of population by hectare and the population growth. These data are available by districts in 2008. To moderate the lack of coincidence between the cells and districts, both variables were weighted with the relative sample sizes in the cells that belong to a given district. The result is the weighted log density WLNDENS and the weighted population growth WPOPGROW. The values by grid cells are shown in the Table 10.

WLNDENS

NRV	NRH1	NRH2	NRH3	NRH4	NRH5	NRH6	NRH7	NRH8	NRH9	NRH10	NRH11	NRH12	NRH13	NRH14	NRH15	NRH16	NRH17	NRH18	NRH19	NRH20
18													5,79							
17								5,70	5,70	5,79	5,79	5,79	5,79	5,79	5,62	5,62	5,62			
16						5,70	5,70	5,70	5,70	5,79	5,79	5,79	5,79	5,79	5,62	5,62	5,62			
15				5,70	5,70	5,70	5,70	5,70	5,70	5,79	5,79	5,79	5,79	5,79	5,62	5,62	5,62	5,62		
14				5,70	5,70	5,70	5,70	5,70	5,70	5,62	5,79	5,62	5,80	5,80	5,62	5,62	5,62	5,62		
13				5,70	5,70	5,70	4,62	4,62	4,62	5,62	5,62	5,62	5,80	5,80	5,80	5,62	5,62	5,62	5,62	
12			5,36	5,36	5,36	4,62	4,62	4,62	4,62	5,62	5,62	5,62	5,80	5,80	5,80	5,62	5,79	5,79	5,79	5,79
11		5,36	5,36	5,36	5,36	5,36	4,62	4,62	4,62	4,57	5,40	5,40	5,71	5,71	6,04	6,04	5,79	5,79	5,79	5,79
10		5,36	5,36	5,36	5,36	4,95	4,95	4,95	4,95	4,57	4,57	4,57	5,71	5,71	6,04	6,04	6,04	5,79	5,79	5,79
9	5,36	5,36	5,36	5,36	5,36	4,95	4,95	4,95	4,95	4,95	5,35	4,57	5,17	5,17	6,04	6,04	6,04	5,79	5,79	5,79
8	5,36	5,36	5,36	5,36	5,63	5,63	4,95	4,95	4,95	5,35	5,35	5,51	5,17	5,17	6,04	6,04	6,04	6,04	5,79	5,79
7	5,36	5,36	5,36	5,63	5,63	5,63	5,63	4,95	5,35	5,35	5,51	5,51	5,51	5,17	5,42	5,42	5,42	5,42	5,42	5,79
6	5,36	5,36	5,63	5,63	5,63	5,63	5,63	5,63	5,50	5,35	5,51	5,51	5,51	5,51	5,42	5,42	5,42	5,42	5,42	5,42
5	5,36	5,36	5,63	5,63	5,63	5,63	5,63	5,63	5,50	5,50	5,50	5,54	5,54	5,54	5,54	5,42	5,42	5,42	5,42	5,42
4				5,63	5,63	5,63	5,63	5,50	5,50	5,50	5,50	5,54	5,54	5,54	5,54	5,54	5,42	5,42	5,42	
3						5,63	5,50	5,50	5,50	5,50	5,50	5,54	5,54	5,54	5,54	5,54	5,54			
2									5,50	5,50	5,54	5,54	5,54	5,54	5,54	5,54				
1											5,50	5,50		5,54						

WPOPGROW

NRV	NRH1	NRH2	NRH3	NRH4	NRH5	NRH6	NRH7	NRH8	NRH9	NRH10	NRH11	NRH12	NRH13	NRH14	NRH15	NRH16	NRH17	NRH18	NRH19	NRH20
18													2,44							
17								0,93	0,92	2,07	2,44	2,44	2,44	2,44	-0,09	-0,09	-0,09			
16						0,92	0,92	0,92	0,94	2,34	2,44	2,44	2,44	2,44	-0,09	-0,09	-0,09			
15				0,92	0,92	0,92	0,92	0,92	1,04	2,44	2,44	2,44	2,44	2,20	-0,09	-0,09	-0,09	-0,09		
14				0,92	0,92	0,92	0,92	1,29	2,13	2,15	2,26	2,12	1,77	1,69	0,37	-0,09	-0,09	-0,09		
13				1,75	0,96	2,46	3,96	4,21	4,00	2,04	2,02	1,92	1,66	1,66	0,99	-0,09	-0,07	0,10	0,28	
12			2,69	2,69	2,65	4,12	4,21	4,21	4,03	2,14	2,32	1,95	1,66	1,66	1,51	0,53	1,00	1,15	1,15	1,15
11		1,98	2,69	2,69	2,69	3,17	4,21	4,21	3,58	1,60	2,27	2,81	2,86	2,32	0,20	0,25	1,13	1,15	1,15	1,15
10		2,69	2,69	2,69	2,69	1,20	0,61	0,72	0,72	1,04	1,04	1,43	3,27	3,39	0,11	0,09	0,36	1,15	1,15	1,15
9	2,69	2,69	2,69	2,69	1,68	0,72	0,72	0,72	0,72	-0,36	-1,10	0,87	0,81	2,01	0,59	0,09	0,17	1,00	1,15	1,15
8	2,69	2,69	2,69	1,87	1,02	0,99	0,72	0,72	0,54	-2,00	-2,58	-0,98	0,57	0,68	0,26	0,11	0,09	0,35	1,15	1,15
7	2,69	2,69	2,48	1,02	1,02	1,02	0,94	0,34	-2,11	-2,59	-1,47	-0,98	-0,98	0,68	0,18	0,18	0,17	0,15	0,51	0,68
6	2,69	2,68	1,55	1,02	1,02	1,02	1,02	1,02	-0,13	-1,70	-1,26	-0,98	-0,98	-0,98	0,18	0,18	0,18	0,18	0,18	0,18
5	2,67	2,69	1,02	1,02	1,04	1,02	1,02	0,79	0,53	0,53	0,09	0,25	0,23	1,10	1,31	1,02	1,02	0,79	0,53	0,18
4				1,02	1,02	1,02	0,98	0,53	0,53	0,53	0,64	1,31	1,31	1,31	1,31	1,28	0,18	0,18	0,18	
3						1,02	0,61	0,53	0,53	0,55	0,86	1,31	1,32	1,31	1,31	1,31	1,31			
2									0,53	0,53	1,07	1,29	1,31	1,31	1,31	1,31				
1											0,53	0,87		1,31						

Table 10 : WLNDENS and WPOPGROW for each cell

The highest densities are observed in the North-East, while the strongest growth occurs in the North-West, besides an isolated positive extreme in the 3d district. Instead the lowest densities and growth rates are observed in the center, with even negative growth rates in the 5th and 6th district. The OLS-estimates of the survey have already led us to assume that people may prefer less densely populated districts, thereby also searching large apartments that are more expensive. It is challenging to examine whether that result gets support at the district level.

To examine that question all variables were aggregated to the district level as shown in Table 9. The OLS-estimates with these data are demonstrated in Table 11.

Dependent					Mean
TRANSR					1,094
Explanatories		Coeff.	t-ratio	Prob	Means
CONSTANT		0,926	15,32	0	
WPOPGROW		0,122	3,76	0,002	1,161
LRELPRIX		0,489	1,96	0,067	0,056
	20	observations		R2 =	0,489

Table 11 : OLS-estimation of relative intensity

The OLS-estimators are highly significant for WPOPGROW and also significant for LRELPRIX, while WLNDENS did not turn out as significant. Hence a strong population growth is accompanied by an increase in transactions. Quite interestingly, higher prices exert also a stronger transaction effect, possibly because people agree to the offer of more expensive large and duplex apartments. Of course this must be stated with care, because the causal direction cannot unambiguously be derived from estimators in a static model. In particular, the year 2007 was characterized by an economic boost; thus the purchasers may have been guided by expectations of an everlasting economic progress and rising property prices. Purchasers endowed with substantial assets may have allocated their money in high price apartments.

Weight matrix

The content of the weight matrix depends on the neighborhood structure chosen. There are a lot of possibilities to construct such a matrix.

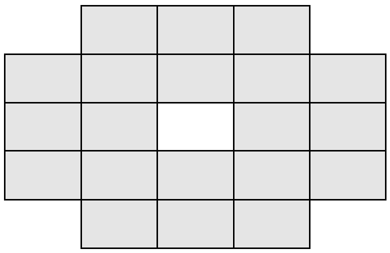
The most commonly used matrices are the contiguity matrices. The contiguity between two regions defined by the fact that they have a common border and each term of this matrix is equal to 1 if the regions are contiguous to order 1 and 0 otherwise (by convention, a region is not contiguous with itself: $w_{ii} = 0\ \forall i$).
These contiguity matrices are often used because of their simplicity but appear restrictive in terms of definition of the spatial connection between regions. Another possibility is the use of distance matrices. It is assumed in this case that the intensity of the interaction between two regions i and j depends on the distance between the centers of these regions.
Several indicators can be used as defined by the distance: bird's-distance, distance to roads or generalization transport time . Various specifications are available, the most widely used being the negative exponential function or a function of the inverse of the distance.

3 different options were considered in our study :

- Model EXTENDED : first and second order neighbors are considered (20 cells) and reciprocal distances are used
- Model LINEAR : 4 neighboring cells with reciprocal distances
- Model VOISINS : 4 neighboring cells with neighborhood marked by 1

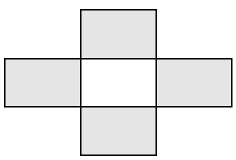
Here we can see a description of each model and the values of reciprocal distances :

Model EXTENDED : reciprocal distances



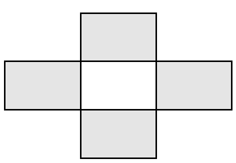
	0,926	1,070	0,926	
0,848	1,398	2,140	1,398	0,848
0,923	1,846	0	1,846	0,923
0,848	1,398	2,140	1,398	0,848
	0,926	1,070	0,926	

Model LINEAR : reciprocal distances



	2,140	
1,846	0	1,846
	2,140	

Model VOISINS : 1/0



	1	
1	0	1
	1	

After a transformation called “row-standardization” in which the rows of the neighbors matrix are made to sum to unity, we obtain the weight matrix W.

Calibration Design

This part is intended to assess the choice of the distribution of errors in the SAR model. The method used for calibration is as follows :

The spatial autoregressive (SAR) model over N cells is

$$y = \rho W y + X\beta + u$$

The problem is the shape of the distribution of errors u that determines the test statistics and the economic validity of the estimates. For that we get estimates $\hat{\rho}$ and $\hat{\beta}$.

With given non-stochastic explanatories this yields

$$\widehat{y_t} = \hat{\rho} W \widehat{y_t} + X_t\hat{\beta}$$

Hence, for the SAR the deterministic predictions are

$$\widehat{y_t} = (I - \hat{\rho} W)^{-1} X_t \hat{\beta}$$

Starting from these deterministic predictions as if they would reflect the true process, we calculate randomized endogenous variables :

$$\widetilde{y_t} = \widehat{y_t} + e_t$$

where the errors e_t are drawn from a random number generator. This is either a Gauss-Normal distribution or a skew distribution, in particular Gamma. To get asymptotic result statistics, the vectors $\tilde{y}$ have to be repeated at least 100 times.

We conduct the experiment with three different types of error :

- Normal distribution with expectation = 0 and standard deviation = 484
- Gamma distribution with parameters 4,27 and 234,36 (this provides expectation = 1000 and standard deviation = 484)
- Gamma distribution with parameters 1,07 and 468,72 (this provides expectation = 500 and standard deviation = 484)

This choice of parameters allows to have a standard deviation equal to the residual variance of the initial SAR model (484). Results of this initial SAR model are presented in figure 5. We test two different gamma distribution with expectation 500 and 1000, and we perform then a translation to position the random errors with expectation = 0

We perform then SAR model with the vector $\tilde{y}$. This operation has to be repeated 100 times for each type of error distribution.

We can then compare the results obtained for each type of errors (see table 11 and 12) :

- average and standard deviation of estimates
- average and standard deviation of standard errors

```
Call:lagsarlm(formula = prixm2n ~ neufqn + surf3qn + duplexn + wlndensn +
    wpgrown + ouvriern + jeunen, listw = mat2listw(as.matrix(wnorm),
    style = "W"), zero.policy = T, tol.solve = 1e-14)

Residuals:
      Min        1Q    Median        3Q       Max
-1564.363  -260.287   -14.146   216.782  2153.717

Type: lag
Coefficients: (asymptotic standard errors)
             Estimate Std. Error z value  Pr(>|z|)
(Intercept)  4269.754    938.201  4.5510 5.339e-06
neufqn        945.153    212.307  4.4518 8.515e-06
surf3qn      1417.149    281.555  5.0333 4.821e-07
duplexn      2793.411    779.710  3.5826 0.0003402
wlndensn     -507.197    141.501 -3.5844 0.0003379
wpgrown      -107.054     28.503 -3.7558 0.0001728
ouvriern    -1920.470    529.045 -3.6301 0.0002833
jeunen        659.726    272.801  2.4183 0.0155914

Rho: 0.73423, LR test value: 194.26, p-value: < 2.22e-16
Asymptotic standard error: 0.037444
    z-value: 19.609, p-value: < 2.22e-16
Wald statistic: 384.51, p-value: < 2.22e-16

Log likelihood: -2030.562 for lag model
ML residual variance (sigma squared): 234360, (sigma: 484.11)
Number of observations: 264
Number of parameters estimated: 10
AIC: 4081.1, (AIC for lm: 4273.4)
LM test for residual autocorrelation
test value: 5.4061, p-value: 0.020066
```

Figure 5 : Initial SAR results

Comparing standard deviations of estimates for each error pattern (Table 11 and 12), we see similar results for gamma and normal errors.

SAR	Normal (0 ; 484)			
	Estimates		Std Errors	
	Average	SD	Average	SD
Rho	**0,533**	0,043		
Constant	**8258,9**	877,5	**1184,9**	62,0
NEUF	**408,2**	197,3	**243,7**	13,5
SURF3	**1706,9**	277,6	**319,9**	17,0
DUPLEX	**3080,1**	766,5	**889,0**	49,3
WLNDENS	**-967,1**	116,5	**172,2**	9,2
WPOPGROW	**-163,1**	22,6	**33,3**	1,8
OUVRIER	**-3229,9**	532,7	**606,0**	32,2
JEUNE	**773,5**	286,5	**310,9**	17,0

Table 11 : Normal Error

SAR	Gamma (4,27 ; 234,36) + translation				Gamma Bis (1,07 ; 468,72) + translation			
	Estimates		Std Errors		Estimates		Std Errors	
	Average	SD	Average	SD	Average	SD	Average	SD
Rho	**0,538**	0,045			**0,537**	0,054		
Constant	**8175,5**	899,3	**1179,4**	64,9	**8187,9**	1101,0	**1176,2**	90,7
NEUF	**435,7**	238,7	**242,8**	14,2	**364,0**	184,4	**242,1**	20,3
SURF3	**1747,0**	282,8	**319,6**	18,1	**1661,9**	279,5	**317,6**	25,6
DUPLEX	**2895,0**	798,5	**886,2**	52,1	**3019,9**	744,6	**882,7**	74,4
WLNDENS	**-959,0**	116,4	**171,5**	9,6	**-959,0**	140,6	**170,9**	13,5
WPOPGROW	**-166,8**	22,1	**33,3**	1,9	**-161,4**	22,1	**33,0**	2,7
OUVRIER	**-3214,8**	630,7	**604,2**	34,4	**-3286,8**	603,6	**602,2**	49,2
JEUNE	**790,9**	219,9	**310,0**	18,0	**810,6**	299,5	**308,8**	25,7

Table 12 : Gamma errors

Results with SAR Method

The data used for evaluating the price diffusion process are based on the 264 observations that belong to the non-empty cells in the grid. The data form a meso-aggregate as the characteristics of objects and persons are averages within each cell. Therefore the statistics drawn from the grid slightly differ from the statistics drawn from the original survey.

Table 13 shows the results of the OLS (with and without population data) and SAR for different types of neighborhood.

Observed price per sqm:														
Average	6517,4													
STD	1386,9													
Item	Average		OLS		OLS		SAR-EXTENDED		SAR-LINEAR		SAR-VOISINS		AVPRIX	DIFFPRIX
Constant			6143,8	***	17953,7	***	2976,8	***	4269,6	***	4227,0	***		
QNEUF	0,0785		-389,0		-74,8		1139,5	***	945,1	***	965,3	***	6743,2	34,8
QSURFACE1	0,4185												6045,4	
QSURFACE2	0,4616												6274,3	
QSURFACE3	0,1199		4198,6	***	3205,1	***	1284,0	***	1417,2	***	1400,0	***	7494,9	53,5
QDUPLEX	0,0343		3136,8	*	2753,5	**	3061,0	***	2793,1	***	2790,4	***	7621,2	104,4
WLNDENS	5,5053				-2087,1	***	-416,5	***	-507,2	***	-503,2	***		-1908,0
WPOPGROW	1,1583				-310,4	***	-108,8	***	-107,1	***	-106,2	***		-403,6
QOUVRIER	0,0855		-7692,9	***	-5725,3	***	-1767,8	***	-1920,9	***	-1924,4	***	5733,4	-72,2
QJEUNE	0,3021		1490,6	***	1423,0	***	695,1	**	659,8	**	679,0	**	6212,1	24,9
rho							0,850	***	0,734	***	0,737	***		
pseudo-R2			0,511		0,694		0,856		0,854		0,855			
relstd(residuals)			0,1488		0,1177		0,0774		0,0744		0,0741			
logLikelihood			-2189,6		-2127,7		-2028,7		-2030,6		-2029,5			
AIC			4 393,2		4273,3		4077,3		4081,1		4079,1			
residual autocorr							0,630		0,020		0,025			
levels of significance		***	<1%	**	<5%	*	<10%		empty: insignificant					

Table 13 : SAR results under alternative specifications of neighborhoods

It can be seen that now the average price for total Paris is 6517.4, which exceeds the former price average of 6282.6 in Table 1. This results from the fact that we use the SAR-method without weights. Since the property prices in the inner districts are higher while the samples are smaller, the unweighted grid mean is necessarily higher.

The averages of the factor quotas shown in Table 13 are the means of the 264 meso-aggregates in the cells. Compared with the values in Tables 1 they differ in both directions. The differences between the survey and grid statistics could be reduced by making the grid finer.

The first two columns refer to the OLS-estimation of the prices against the characteristics. The results without and with population data can be compared. The fit given by the R2's satisfies. But in contrast to the OLS-results based on the survey, the differences between the factor estimators are larger now. Moreover, the estimator of QNEUF gets insignificant. Similar to the survey data, the magnitudes of the estimators Constant, log-density and population growth get large if population data are included. Note that the weighted log-densities WLNDENS and population growth rates WPOPGROW are used now. Again it is disappointing that the estimators Constant, population, and now even OUVRIERS, explain most of the endogenous price variable. However, by introduction of the spatial autocorrelation, this result improves substantially.

In the SAR model, the spatial autocorrelation parameter ρ measures the power of diffusion of the endogenous variable, here the average sqm-price in the grid cells PRIXM2. For that we have to design the matrix of weights W, which consists of the reciprocal of distances between the mid-points of the cells. Three types of distances are selected: the model EXTENDED that is based on the distances up to the over-next cell, the model LINEAR that is based on the distances to the straight neighboring cells, and the model of cell neighborhoods VOISINS that consists of 0/1 components with 1 for the straight neighbors.

Starting with the general statistics, no model appears uniformly better than other ones. The model EXTENDED is superior with regard to the log-Likelihood, the AIC criterion and the residual autocorrelation, which is the spatial autocorrelation left over after estimation (the null autocorrelation has a high probability). But the model EXTENDED provides the largest relative standard deviation of the residuals, in terms of the average price. The relative standard deviation assumes the lowest value in the model VOISINS. The pseudo-R2 along Nagelkerke is almost identical in all three models, and surprisingly satisfying with about 0.855. In all models, the spatial autocorrelation ρ is highly significant and rather substantial. This points to a strong diffusion process, which blurs the factor estimators, in particular in the model EXTENDED with the estimate $\rho=0.857$. Since the lowest value $\rho=0.742$ is assumed in the model LINEAR, we focus on the details of that model, for which the reduced form factor estimators DIFFPRIX are given, see below. There is no loss of information with regard to the model VOISINS, because the values of the coefficient estimators are quite close to the model LINEAR.

In all spatial models, the coefficient estimators are highly significant. The Constant and the estimators of WLNDENS and WPOPROW, which marked a trade-off in the OLS-model, are much lower in magnitude. In the LINEAR model, the Constant declines to 3906.6, while WLNDENS and WPOPGROW assume -451.0 and -104.0 resp. Recently built, large and duplex properties retain positive and relatively large estimators, with 944.1, 1435.6 and 2754.1 resp. The negative estimator -1918.0 of the OUVRIER's is substantial.

Instead, the young purchasers reveal a positive and moderately high estimator of 663.5. These coefficient values are in accordance with the average sqm-prices drawn from the survey, which are shown conditional upon object and personal characteristics in column AVPRIX, Table 13.

In contrast to the OLS-model the coefficient estimators in the SAR-model do not directly reveal the cet. par. price response to changing characteristics. The reason is the spatial autocorrelation, which transmits the cet. par. price effects over the whole grid. The reduced form price response can be obtained from the model predictors, when they get decomposed into predicted prices conditional on the model explanatories.

The price differences according to changing characteristics, labeled DIFFPRIX, is obtained by the decomposition of predicted prices $\hat{y}$ in k components $\widehat{y_{.k}}$ function of $X_{.k}$:

$$\hat{y} = (I_N - \hat{\rho}W)^{-1} X\hat{\beta} = \sum_{k=1}^{K} \widehat{y_{.k}} = \sum_{k=1}^{K} \widehat{\beta_k} (I_N - \hat{\rho}W)^{-1} X_{.k}$$

We can calculate from this decomposition the price difference related to the presence or not of a characteristic.

The factor quotas are analysed with regard to a 1%-point increase, that is by a difference of +0.01. Thus, if the quota of new apartments QNEUF increases by 1%-point, the average sqm-price in Paris rises by 36.6 EUR. If the quota of large apartments QSURF3 increases, the price response is 55.6 EUR, and for the quota QDUPLEX the price response is 106.7 EUR. Correspondingly, if there are 1%-point more QOUVRIER purchasers, it declines by 74.3 EUR, and if the quota of young purchasers QJEUNE increases by 1%-point, it rises by 25.7 EUR.

It should be remarked that the price responses hold for a price structure that remains otherwise invariant. In simple words, the changes of the average sqm-price in the city of Paris are generated by a reallocation of the purchasing population. This also demonstrates the limits of the cet. par. analysis. For instance, if the quota of large apartments QSURF3 rises, the quotas of smaller apartments QSURF1 and QSURF2 must decline to some unknown extent. Fortunately, these quotas turned out to be insignificant in SAR-estimation and do not appear among the DIFFPRIX-components.

But with regard to the population data a most severe problem arises. If the log-density rises by 1%, the price response is -642.6 EUR, and if the population growth rises by 1%-point, it is -402.6 EUR. The negative responses are due to the fact that people choose properties all over Paris, where the "arrondissements" with lower prices are more populated. However, a substantial population increase is unlikely to leave the sqm-prices of properties unchanged. The dynamics resulting from such a process are not captured by the static SAR-model.

Hence the analysis of the price differences should be confined to the object and personal factors where the information about the price response is most useful.

Price diffusion and Interpretation

The final step is the ranking of "arrondissements" according to the power of price diffusion. The idea is simple: Over the grid we calculate the partial contribution of the "arrondissements" to the size of ρ.

The "arrondissements" with strong diffusion exert a decline in the estimate ρ if we dispense of that arrondissement in estimating the price for entire Paris. Conversely, those with weak diffusion raise the estimate ρ. By approximation, in the selection of districts we had to assign the cells to the districts that create the majority of observations there. In addition, the "arrondissements" with strong diffusion can be categorized into high price and low price ones Both generate a diffusion process albeit in opposite directions.

The results for the three types of neighborhood are listed in Table 14. The high price "arrondissements" are marked with light pink, while those with low prices are marked with aquamarine.

	> 7000	high price		
	< 6000	low price		
ARROND	PRIXM2	EXTENDED	LINEAR	VOISINS
1	7769,2	0,856	0,717	0,719
2	7288,3	0,847	0,735	0,737
3	6789,9	0,831	0,716	0,719
4	8333,6	0,852	0,734	0,736
5	8218,9	0,830	0,712	0,713
6	9181,6	0,846	0,725	0,728
7	9184,0	0,844	0,727	0,731
8	8421,3	0,860	0,750	0,752
9	6176,8	0,845	0,739	0,740
10	5286,2	0,839	0,729	0,732
11	5930,8	0,841	0,727	0,728
12	5910,8	0,809	0,702	0,704
13	6168,9	0,848	0,735	0,737
14	6365,3	0,840	0,723	0,727
15	6513,8	0,800	0,700	0,703
16	7304,5	0,882	0,766	0,770
17	6293,4	0,853	0,750	0,751
18	5099,8	0,860	0,746	0,749
19	5022,7	0,850	0,735	0,738
20	5139,1	0,851	0,730	0,732
Paris	6517,4	0,857	0,742	0,744

Table 14 : Models SAR with alternative neighbourhoods
Rho by selected exclusion of arrondissements

Ranking by rho are shown in the table 14 for EXTENDED and LINEAR model (Sorting rho in VOISINS yields order identical to rho in LINEAR). The thresholds of strong and weak power of diffusion are of course somewhat arbitrary. We suggest to set the borders as they are marked with bold lines in the Table 15.

Sorted by rho in EXTENDED		
ARROND	PRIXM2	EXTENDED
15	6513,8	0,800
12	5910,8	0,809
5	8218,9	0,830
3	6789,9	0,831
10	5286,2	0,839
14	6365,3	0,840
11	5930,8	0,841
7	9184,0	0,844
9	6176,8	0,845
6	9181,6	0,846
2	7288,3	0,847
13	6168,9	0,848
19	5022,7	0,850
20	5139,1	0,851
4	8333,6	0,852
17	6293,4	0,853
1	7769,2	0,856
8	8421,3	0,860
18	5099,8	0,860
16	7304,5	0,882

Sorted by rho in LINEAR		
ARROND	PRIXM2	LINEAR
15	6513,8	0,700
12	5910,8	0,702
5	8218,9	0,712
3	6789,9	0,716
1	7769,2	0,717
14	6365,3	0,723
6	9181,6	0,725
11	5930,8	0,727
7	9184,0	0,727
10	5286,2	0,729
20	5139,1	0,730
4	8333,6	0,734
13	6168,9	0,735
2	7288,3	0,735
19	5022,7	0,735
9	6176,8	0,739
18	5099,8	0,746
8	8421,3	0,750
17	6293,4	0,750
16	7304,5	0,766

Table 15 : Sorted by rho in EXTENDED and LINEAR

Then the arrondissement with strong diffusion are 3, 5, 12 and 15; in the models LINEAR and VOISINS they are supplemented by the arrondissements 1 and 17. The arrondissements with weak diffusion are 8, 16 and 18 in all cases. Looking at the range of the resulting ρ's, we should better say "strong" versus "not so strong", but we stay with "strong" versus "weak" for convenience.

Now, by combining high prices with strong diffusion, the price leader in that sense (the quartiers primes) is the rive gauche with the 5th district (the Quartier Latin), while the 1st district (with the Île St. Louis) is small and rather isolated. Medium price districts with strong diffusion are the arrondissement 3 in the Center and the large arrondissements 15 and 17 in the West. Quite remarkably, there is finally the low price arrondissement 12 in the East that exerts a countervailing force in price formation.

It is also interesting to mention the districts with weak power of diffusion. These are the high price arrondissements 8 and 16, and the low price arrondissement 18. The arrondissement 16 is astonishing because it exerts the weakest diffusion by far.

From a statistical point of view one may ask why the price diffusion is strong or weak. We raise the hypothesis that the spatial autocorrelation is affected by the pattern of factor quotas within the districts. That pattern can be revealed by the cell means belonging to an arrondissement. If these means deviate much from the city averages, the contribution to ρ should decline by weak diffusion. The opposite case of strong diffusion should be created by a small deviation, thus the pattern conforms to the city average.

To justify this interpretation we calculate a variable called PATTERN that quantifies the similarity of the characteristics of a district with the average of Paris (see Table 16).

Indeed, for the weak diffusion such a pattern was detected in the heterogeneous arrondissements 8 and 16. The object characteristics in these districts deviate from the average quotas by far, in particular by the absence of new and small apartments, and by the prevalence of large properties. For the social strata expressed by the quotas of workers QOUVRIER and young purchasers QJEUNE the largest deviation obtains in the arrondissement 8 where business dominates, while 16 is not so far from the city average. Taken together, the factor quotas with the largest deviation are again in 8 and 16, see column PATTERN in Table 16. Quite interestingly, an almost identical pattern obtains when drawing from the survey data.

Sorted by rho in LINEAR			
ARROND	PRIXM2	LINEAR	PATTERN
15	6513,8	0,700	0,030
12	5910,8	0,702	0,035
5	8218,9	0,712	0,055
3	6789,9	0,716	0,059
1	7769,2	0,717	0,043
14	6365,3	0,723	0,041
6	9181,6	0,725	0,038
11	5930,8	0,727	0,057
7	9184,0	0,727	0,090
10	5286,2	0,729	0,038
20	5139,1	0,730	0,048
4	8333,6	0,734	0,037
13	6168,9	0,735	0,050
2	7288,3	0,735	0,070
19	5022,7	0,735	0,055
9	6176,8	0,739	0,025
18	5099,8	0,746	0,079
8	8421,3	0,750	0,106
17	6293,4	0,750	0,038
16	7304,5	0,766	0,117

Table 16 : Ranking by rho and pattern

Concerning the opposite case of strong diffusion the patterns deviating least from the city factor averages are derived for the arrondissements 12 and 15. In the leading arrondissement 17, the social strata almost coincide with the city averages, but the object characteristics are farer away. Instead, from the survey data the closest pattern of all factors are derived just in that district. Whether the outcome of the SAR-model presented here can be replicated with the original survey is a topic for future investigation.

In the 5th district the pattern of factor quotas is a bit more heterogenous. But the price leadership obtained for that district can be interpreted from an urban perspective. The arrondissement 5 contains the famous Quartier Latin around the University Sorbonne and many other research and cultural institutes. Although the population in that district is stagnant with moderate fluctuation (see TRANSR in 9), the social climate may signal the conditions of property sales to the other parts of Paris. Admittedly, this interpretation cannot fully satisfy, because it does not rely on statistical arguments.

Since the rho values found with the SAR model are rather high, it means that the model describes a homogeneous diffusion and can't strictly explain the influence of variables. It is useful to obtain lower rho values to answer the following problematic : How the distribution of characteristics in the districts affects prices ? This why we decide to use a more general model : SDM (Spatial Durbin Model).

Comparative analysis with Spatial Durbin Model

Presentation of the Spatial Durbin Model

Spatial Durbin Model is a more general model that takes into account the spatial distribution of characteristics.

The formal model is

$$y = \rho W y + X\beta + WX\theta + u \quad , \qquad u \sim i.i.d.(0, \sigma^2 I_n)$$

Our model specification will allow characteristics (variables contained in the matrix X) from neighboring regions to exert an influence. This is accomplished by entering an average of the explanatory variables from neighboring regions, created using the matrix product W X.
The only difference with the SAR model is this term "WXθ", which takes into account the lag values of characteristics.

Now neighborhood characteristics have influence to determine the price. For example a high rate of workers in a neighborhood will influence the price in the area and tend to lower the price. Conversely, the presence of luxury buildings in a neighborhood will tend to increase the price in the area.

Table 17 shows the results of the OLS (with population data), SAR and SDM for LINEAR neighborhood. We can observe that the value of rho falls to a more reasonable level (0.668) with a SDM. We can look at the influence of the variables and see how they are distributed over the territory. Especially we can see that the lag QNEUF has a significant role.

Observed price per sqm:							
Average	6517,4						
STD	1386,9						
Item	Average	OLS		SDM-LINEAR		SAR-LINEAR	
Constant		17953,67	***	6663,2	***	4269,6	***
QNEUF	0,0785	-74,81		1226,0	***	945,1	***
lag QNEUF				-1655,0	***		
QSURFACE3	0,1199	3205,05	***	1566,1	***	1417,2	***
lag QSURFACE3				-706,3			
QDUPLEX	0,0343	2753,5	**	2380,5	***	1417,2	***
lag QDUPLEX				1312,7			
WLNDENS	5,5053	-2087,14	***	158,1		2793,1	***
lag WLNDENS				-950,6	*		
WPOPGROW	1,1583	-310,42	***	51,2		-507,2	***
lag WPOPGROW				-192,6	*		
QOUVRIER	0,0855	-5725,31	***	-1996,2	***	-107,1	***
lag QOUVRIER				-1836,9	*		
QJEUNE	0,3021	1422,96	***	646,9	**	-1920,9	***
lag QJEUNE				-139,4			
rho				**0,668**		**0,734**	
pseudo-R2		0,694		0,880		0,854	
logLikelihood		-2127,7		-2012,9		-2030,6	
AIC		4273,3		4059,7		4081,1	
residual autocorr				11,9460	***	0,0201	

Table 17 : Comparison OLS, SAR and SDM

Table 18 provides the new classification with the SDM. The results are similar with SAR model for most districts. However, we note the appearance of the 18th district among the districts with a strong spatial distribution. The districts 16 and 17 are still those with the lowest spatial diffusion. The SDM has therefore confirmed largely the results obtained with the SAR. These findings are illustrated in the Figure 6.

Sorted by rho in LINEAR with SDM		
ARROND	PRIXM2	LINEAR
15	6513,8	0,623
5	8218,9	0,639
14	6365,3	0,644
12	5910,8	0,649
18	5099,8	0,655
10	5286,2	0,659
8	8421,3	0,660
11	5930,8	0,660
6	9181,6	0,664
3	6789,9	0,664
20	5139,1	0,665
19	5022,7	0,666
1	7769,2	0,667
4	8333,6	0,672
2	7288,3	0,675
9	6176,8	0,675
7	9184,0	0,677
13	6168,9	0,689
16	7304,5	0,689
17	6293,4	0,716

Sorted by rho in LINEAR with SAR		
ARROND	PRIXM2	LINEAR
15	6513,76	0,700
12	5910,79	0,702
5	8218,9	0,712
3	6789,93	0,716
1	7769,23	0,717
14	6365,34	0,723
6	9181,58	0,725
11	5930,75	0,727
7	9184,03	0,727
10	5286,22	0,729
20	5139,08	0,730
4	8333,61	0,734
13	6168,93	0,735
2	7288,27	0,735
19	5022,67	0,735
9	6176,75	0,739
18	5099,78	0,746
8	8421,25	0,750
17	6293,39	0,750
16	7304,5	0,766

Table 18 : Ranking by rho (SDM) and comparison with SAR

The districts 12 and 18 are "low price arrondissement" and have a high diffusing power. They tend to reduce prices of neighboring districts.

The opposite phenomenon occurs with the 5th district, which is a "high price arrondissement". Note that this phenomenon is even more important than the price difference between two neighboring districts is significant.

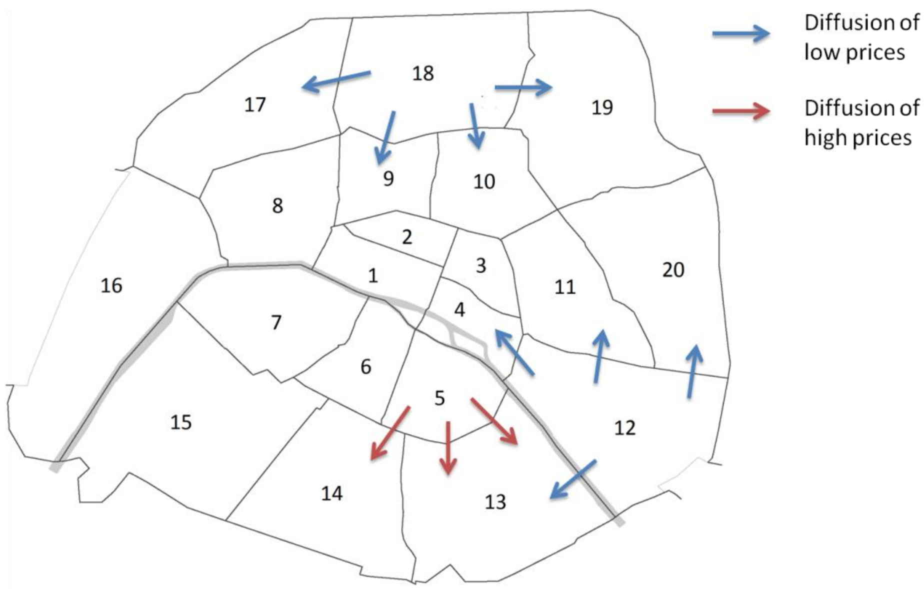


Figure 6 : Diffusion process in Paris

The districts 12 and 18 are "low price arrondissement" and have a high diffusing power. They tend to reduce prices of neighboring districts.

The opposite phenomenon occurs with the 5th district, which is a "high price arrondissement". Note that this phenomenon is even more important than the price difference between two neighboring districts is significant.

CHAPTER III : Approach on individual observations and results validation

The previous chapter with the use of the SAR and SDM methods allowed to obtain significant and interesting results regarding the spatial diffusion within the city of Paris. Nevertheless it was impossible to directly use individual data and this constitutes a loss of information. To remedy it, several solutions are possible.

One can for example consider a random sampling among the individual observations. We know that the R software is able to apply the SAR method to about 2000 observations. That is why we decided to perform tests with 2,000 observations drawn at random.

Another solution is to create a finer grid that would provide a more accurate view of the diffusion phenomenon. The use of individual data is still not direct but one approaches the microscopic level with this solution.

Nearest Neighbors method with R software

Since we are now working directly with individual observations, the latitude and longitude characteristics now have a direct impact. In the previous chapter these characteristics were only used to associate a grid cell to each observation. The weight matrix was then generated according to the distance between two cells of the grid and a choice of neighborhood. The purpose of this part is to use position data directly to create the weight matrix.

For each observation, we have a geographical location and can associate neighbors located within a perimeter. The choice of the perimeter, which corresponds to determining the maximum distance from a neighboring, highly influence results.

The software R allows to create such a weight matrix directly from the geographical coordinates of each observation.
In order to validate and to explain the process, we test the algorithm with the network of 360 cells. In other words, we consider that each cell of the grid corresponds to an observation and a geographical position. Geographical position of a cell is the average position of the observations in the cell (see Figure 7)

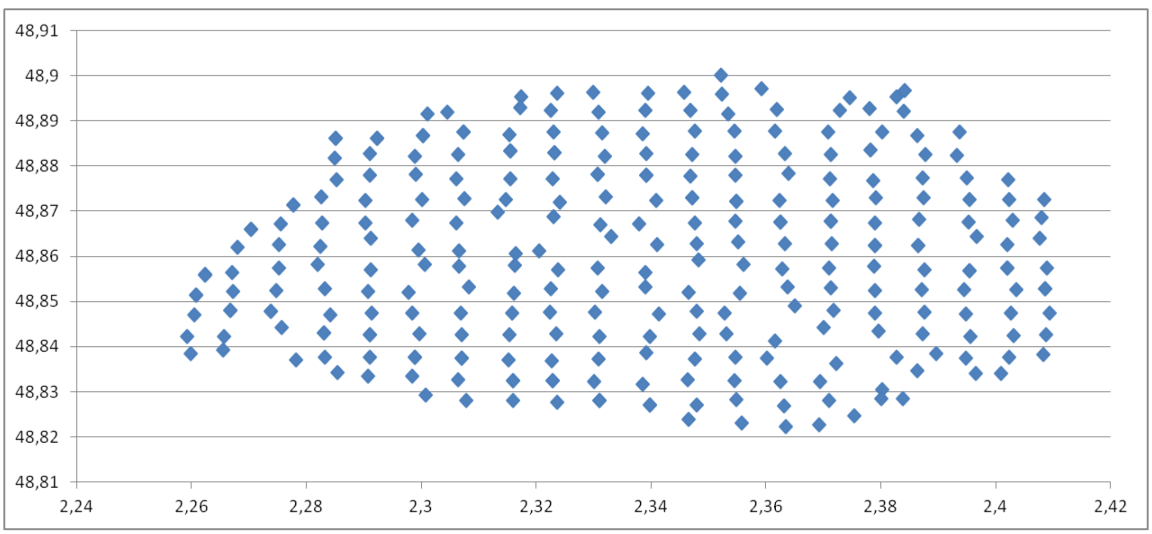


Figure 7 : Geographical positions

The "**dnearnneigh**" command of R software provides the list of neighbors of each cell, as shown in Figure 8. We must choose the maximum distance between two neighboring. Here the maximum distance is 1.5 km. After various tests, this value allows to have results close to those obtained in the previous chapter. This distance of 1,5 km corresponds approximately to the first and second order neighbors.

For example, neighboring cells of the seventh cell are : 1, 5, 6, 8, 18, 19, 20, 32, 33, 34. Obviously, the number of neighboring cells increases with increasing maximum distance. This also means that the implementation is longer with a higher maximum distance. However the interest in considering a very important number of neighbors is limited since the weight associated with these neighbors quickly becomes lower with distance.

1	c(5, 6, 7, 8, 18, 19, 20, 32, 33)	
2	c(3, 4, 13, 14, 15, 27, 28, 29, 42, 43, 58)	
3	c(2, 4, 14, 15, 16, 28, 29, 30, 44, 59)	
4	c(2, 3, 5, 14, 15, 16, 17, 29, 30, 31)	
5	c(1, 4, 6, 7, 16, 17, 18, 30, 31, 32, 46)	
6	c(1, 5, 7, 8, 17, 18, 19, 31, 32, 33)	
7	c(1, 5, 6, 8, 18, 19, 20, 32, 33, 34)	
8	c(1, 6, 7, 18, 19, 20, 33, 34)	
9	c(10, 11, 20, 21, 22, 23, 35, 36, 50, 51)	
10	c(9, 11, 21, 22, 23, 36, 37, 38, 51)	
11	c(9, 10, 21, 22, 23, 36, 37, 38)	
12	c(13, 25, 26, 27, 40, 41, 42, 55, 56, 57, 73)	
13	c(2, 12, 14, 25, 26, 27, 28, 40, 41, 42, 56, 57)	
14	c(2, 3, 4, 13, 15, 27, 28, 29, 42, 43, 44, 58, 59)	
15	c(2, 3, 4, 14, 16, 28, 29, 30, 43, 44, 58, 59)	

Figure 8 : List of neighbors

The " **nbdists**" command of R software provides now the list of distances for each cell, as shown in Figure 9. In other words, for each cell, we have the distance with each neighboring cell. To obtain the weight matrix, it remains to choose a mathematical function, here the inverse function. After a transformation called “row-standardization” in which the rows of the neighbors matrix are made to sum to unity, we obtain the weight matrix W.

```
distance                      List of 264
   : num [1:9] 1.473 0.833 0.477 0.842 1.061 ...
   : num [1:11] 0.697 1.385 1.477 0.278 0.662 ...
   : num [1:10] 0.697 0.689 0.789 0.436 0.909 ...
   : num [1:10] 1.385 0.689 1.055 1.448 0.935 ...
   : num [1:11] 1.473 1.055 0.695 1.412 1.059 ...
   : num [1:10] 0.833 0.695 0.718 1.489 0.865 ...
   : num [1:10] 0.477 1.412 0.718 0.782 0.724 ...
   : num [1:8] 0.842 1.489 0.782 1.476 0.895 ...
   : num [1:10] 0.887 1.062 1.442 0.364 0.466 ...
   : num [1:9] 0.887 0.226 1.129 0.585 0.38 ...
   : num [1:8] 1.062 0.226 1.333 0.805 0.512 ...
   : num [1:11] 0.387 1.133 0.526 0.823 1.133 ...
   : num [1:12] 1.477 0.387 1.416 1.496 0.736 ...
   : num [1:13] 0.278 0.789 1.448 1.416 0.583 ...
   : num [1:12] 0.662 0.436 0.935 0.583 0.921 ...
```

Figure 9 : List of distances

Test of the algorithm with the grid of 360 cells

Before applying the method to individual observations, we tested the weight matrix creation algorithm on the network to ensure that the results were consistent with results from previous chapters. As said previously, we associate with each cell of the grid a geographical position corresponding to the centroid of all observations present in the cell. With a maximum distance of 1.5 km we get the results shown in Table 19.

Observed price per sqm:										
Average	6517,4									
STD	1386,9									
Item	Average		SAR-EXTENDED		SAR-LINEAR		SAR-VOISINS		Dmax = 1,5	
Constant			2976,8	***	4269,6	***	4227,0	***	5004,4	***
QNEUF	0,0785		1139,5	***	945,1	***	965,3	***	911,1	***
QSURFACE1	0,4185									
QSURFACE2	0,4616									
QSURFACE3	0,1199		1284,0	***	1417,2	***	1400,0	***	1407,1	***
QDUPLEX	0,0343		3061,0	***	2793,1	***	2790,4	***	3473,8	***
WLNDENS	5,5053		-416,5	***	-507,2	***	-503,2	***	-694,9	***
WPOPGROW	1,1583		-108,8	***	-107,1	***	-106,2	***	-125,4	***
QOUVRIER	0,0855		-1767,8	***	-1920,9	***	-1924,4	***	-1906,3	***
QJEUNE	0,3021		695,1	**	659,8	**	679,0	**	744,8	**
rho			0,850	***	0,734	***	0,737	***	0,775	***
pseudo-R2			0,856		0,854		0,855		0,855	
relstd(residuals)			0,0774		0,0744		0,0741		0,0751	
logLikelihood			-2028,7		-2030,6		-2029,5		-2043,5	
AIC			4077,3		4081,1		4079,1		4107,1	
residual autocorr			0,630		0,020		0,025		0,025	
levels of significance		***	<10%		empty: insignificant					

Table 19 : SAR method with Dmax = 1.5 km

The results are very similar to those obtained with manual neighborhood matrices and encourage us to apply the method with individual observations.

Random sampling among the individual observations

We can now work with individual data sample. However, it is impossible to implement the algorithm with our 28828 observations. This corresponds to a weight matrix with 831053584 elements, which cannot be managed by the software R. The limit is around 2,000 observations. That is why we decided to make a random selection of 2,000 observations from the sample (28828 observations). We repeat the process 14 times to get a good sample coverage.

The maximum distance used for each test is still 1.5 km. This corresponds to a number of neighbors on the order of a hundred.

The results of the 14 experiments are presented in Table 20. The colors correspond to the significance of each estimator.

Mean and standard deviation of each estimator are summarized in Table 21.

We note that standard deviation of the estimators are relatively low, except for DUPLEX. This is logically due to the small number of duplex in the sample.

This low volatility of most of estimators allows us to use the results of the random sampling.

First of all we can notice that the average value of rho (0.822) is close to the value of EXTENDED model (0.85) and higher than in models LINEAR (0.734) and VOISINS (0.737). Note that this value of rho is in direct relation with the choice of the maximum distance of neighborhood.

The value of the constant (5428.9) is higher than with the grid but remains similar to models LINEAR (4269.6) and VOISINS (4227.0). This constant is still highly significant.

Regarding factors, some remain highly significant (NEUF, SURF3, DUPLEX, WLNDENS) while others lose part of their significance (WPOPGROW, OUVRIER, JEUNE).

Test	1	2	3	4	5	6	7
Rho	0,811	0,794	0,858	0,829	0,821	0,838	0,850
Constant	5263,3	5912,8	4816,8	5033,3	5579,1	4660,4	5227,1
NEUF	689,7	555,4	555,3	859,3	814,1	843,5	601,4
SURF3	395,9	549,6	365,2	339,1	436,6	446,1	368,0
DUPLEX	567,2	1050,2	1006,3	902,4	820,6	1087,8	1206,1
WLNDENS	-734,3	-831,2	-704,7	-712,4	-807,8	-659,2	-767,4
WPOPGROW	-11,5	-36,2	-9,7	-3,6	-50,8	5,0	-15,0
OUVRIER	-86,7	-308,2	-204,5	-341,7	-15,0	-242,4	-98,0
JEUNE	-85,9	-37,4	-74,3	-94,5	40,6	-165,2	-187,9

Test	8	9	10	11	12	13	14
Rho	0,867	0,783	0,806	0,859	0,804	0,797	0,791
Constant	4419,2	6033,2	5505,2	4862,8	5911,7	5792,1	6987,8
NEUF	529,9	575,1	659,7	588,9	678,7	648,7	498,6
SURF3	416,5	586,4	579,6	355,3	392,2	265,2	324,9
DUPLEX	766,4	740,8	494,8	928,8	734,5	1063,2	1042,8
WLNDENS	-647,0	-839,6	-763,6	-713,5	-834,7	-806,4	-1005,8
WPOPGROW	-22,2	-42,3	-51,4	-43,6	-31,5	-46,3	-29,1
OUVRIER	-47,9	-161,9	-234,7	-124,5	-278,7	-273,1	-346,8
JEUNE	-30,8	2,3	-124,3	-34,3	-109,7	-54,0	-209,4

< 1%

< 5 %

< 10%

Insignifiant

Table 20 : Random Sampling

For objects characteristics (NEUF, SURF3 and DUPLEX), magnitudes of the coefficients are smaller with the individual data than with the grid. This is particularly apparent with the DUPLEX (886.6 with individual data, 2793.1 with the LINEAR model over the grid).

An interesting point is the magnitude of WLNDENS (-773.4 with individual data, -507.2 with the LINEAR model over the grid). This increase confirms the major role of this variable, which remains highly significant.

	Average	Standard Deviation
Rho	0,822	0,028
Constant	5428,917	676,686
NEUF	649,870	116,904
SURF3	415,752	96,664
DUPLEX	886,569	207,552
WLNDENS	-773,398	92,195
WPOPGROW	-27,720	18,451
OUVRIER	-197,436	109,401
JEUNE	-83,205	71,770

Table 21 : Average and standard deviation over the 14 experiences

To sum up this section we can say that the direct use of individual data largely validates the results of Chapter 2. This however limits importance and significance of the characteristics of the purchaser.

Let consider now the solution of a finer grid.

A finer grid over Paris

We worked in the second chapter with a grid of 360 cells. Now we consider a finer grid of 1140 cells :

- 36 vertical index
- 40 horizontal index

The cells are nearly quadratic with 275m side length.

In estimation, cells without observation were sorted out and we have at least 1013 cells with observations. Average number of observations per cell is 28,5 observations.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1																								2	4								1							
2																9	11	37	45	19	26	79	65	26			20						5							
3															9	40	80	76	38	39	65	62	95	79	40		20	42		10	52	34	29							
4												2	9			23	136	94	73	46	65	57	68	82	37		36	47		38	110	46	20							
5												1	34	2		24	95	118	74	62	75	65	96	109	60	26	81	36	11	13	39	13	39	2			4			
6									6	7	1	30	40	37	27	38	98	89	12	39	31	24	57	108	87	71	55	2	12	7	15	27	34	22	31	1	4			
7								10	14	54	20	27	20	54	71	72	58	97	60	36	90	63	19	69	20	43	16	22	29	10	6	30	64	16	27	6	2			
8								33	61	28	37	23	24	15	28	88	33	52	21	80	52	63	39	90	19		69	37	52	21	75	29	30	39	16	8				
9							14	28	32	36	45	43	28	15	19	26	20	44	32	70	77	49	56	64	20		9	45	31	28	24	10	2	34	22	6	4			
10								29	49	42	57	48	9	2	18	9	19	5	32	23	48	55	80	56	39	28	13	46	35	11	38	7	20	32	29	25	24	23		
11							4	29	37	31	21	41	12	6	8	23	13	2	17	11	50	50	49	53	29	27	9	13	11	32	32	42	33	54	35	11	39	55	11	
12						8	36	21	25	9	6	19	23	25	12	18	9	24	6	18	8	15	47	43	43	58	27	23	13	45	36	30	58	43	37	60	31	35	5	1
13						8	16	14	17	23	5	21	16	30	25	18	8	8	8	2		22	35	13	49	57	60	66	25	59	49	31	34	29	31	45	50	35	20	10
14					5	16	22	19	34	25	26	13	8	9	6	4	13	25	8	5	20	20	18	37	97	65	52	34	31	75	45	36	25	24	52	30	36	36	28	
15				3	41	25	38	17	24	3	37	16	19	18			1	5	1	4	33	4	32	41	43	37	70	20	57	65	79	42	58	36	34	41	62	64	31	5
16				9	29	29	25	27	36	11	6	5	16	7					14	25	23	15	29	69	54	43	43	29	44	39	71	65	31	6	11	23	56	16	23	
17			2	7	18	52	28	7	6	8	3		4	3	1					9	1	25	10	28	42	71	29	70	33	59	47	59	19	26	9	2	37	47	20	3
18			11	8	32	32	55	14			7	29	31	33	3	13	11					21	27	27	23	48	20	36	50	25	27	77	33	38			32	23	9	
19			12		29	38	57	26	9		16	33	36	38	7	8	9	5	28	6		5	9	18	22	48	27	19	42	24	65	47	38	23	40	1	23	35	13	10
20		6	41	31	32	38	28	17		12	1	27	65	19		11	8	13	8	17	26	32	2		2	29	49	33	40	32	52	52	58	37	55	39	43	39	9	23
21		5	17	29	33	34	23		28	19	19	1	7	16	1	5	8	11	19	12	41	48	23	2	2	2	8	39	49	40	57	36	49	68	39	44	30	25	43	10
22		37	40	23	30	8	1	22	54	26	39	6		9	19	15	9	5	22	25	26	15	37	30	4	31	12	21	12	7	34	27	45	16	74	14	37	42	42	
23	14	17	50	36	43	16	2	22	44	50	38	22	9	14	10	22	25	17	20	18	12	17	8	24	13	8	1	11	8	28	41	64	18	45	53	26	26	23	12	8
24	14	27	23	28	32	1	15	44	26	56	59	34	28	37	18	19	35	29	16	19		7	10	11	28	9		4	12	26	50	52	10	30	25	10	23	32	17	12
25	26	28	11	7	8	29	23	53	51	56	57	26	14	31	16	20	36	13	11	11		21	19	30	37	15				11	4	24	50	17	33	42	19	20	40	7
26	53	25	51	18		9	57	31	36	28	13	72	40	45	19	27	9	31	42	56	8	16	2	25	48	14	8	8			2	3	20	28	9	12	16	12	25	16
27	46	45	49	2			12	29	41	26	35	49	34	27	20	13	10	21	16	36	15	27	13	25	34	15	12	14			3			28	15	36	56	27	14	9
28	40	51	13			1	25	32	48	61	44	26	27	55	18	2	13	10		11	6	5	13	41	40	42	27						3	10	66	18	42	44	18	12
29		5				10	22	24	29	46	40	39	29	32	2	10	24	34	17	10		4	25	41	45	27	18		2	13	4			1	41	44	46	36	21	
30								31	72	28	39	25	35	34	4	57	39	31	46	10	8	6	34	25	17	26	23	8	22	3			1	6		30	50	23		
31									19	16	12	8	49	13	31	45	34	40	39	7	44	17	52	14	12	16	19		35	7	5	19					7			
32											7	23	9	18	14	44	52	43	52	68	12	22	16	25	49	37	10	29	9	45	11	21	10							
33													7	24	9	25	8	60	37	39	14	7	7	22	20	29	19	16	32	30	3	4	1							
34															2	4	37	50	33	19	13	26	58	47	51	17	26	30	9	16	13	1								
35																	4	3	7		3	14	21	23	5	19	49	37	43	1										
36																						4	20			6	11	54	28											

Table 22 : Number of observations per cell

As we can see from the table 22, the cells near the Seine or on the periphery often have a low number of observations.
This topographic choice provides however an interesting vision of the socio demographic characteristics : observations of the same cell are located very close geographically and correspond to a neighborhood. This thinner gate reduces smoothing phenomena within the same cell.

As in the first grid, we associate with each cell an average price and a district, which is the most represented in the cell (the exact location of each observation is given in the sample).

In order to have results comparable to those of the previous parts, we choose a maximum distance of 500m to calculate the weight matrix. This choice gives a fairly low value of rho (0.71) and then allows to have rather different value of rho by district exclusion.

Table 23 shows the results of the SAR for Paris with this grid of 1140 cells. We note that as with the random sampling method, the characteristics of the buyers (QOUVRIER, QJEUNE) are less significants.

SAR results				
Item	SAR Paris Grid of 1140 cells		SAR-LINEAR Grid of 360 cells	
Constant	5946,6	***	4269,6	***
QNEUF	486,2	***	945,1	***
QSURFACE3	1102,4	***	1417,2	***
QDUPLEX	1262,6	***	2793,1	***
WLNDENS	-756,8	***	-507,2	***
WPOPGROW	-91,7	***	-107,1	***
QOUVRIER	-609,7	**	-1920,9	***
QJEUNE	255,8	*	659,8	**
rho	0,709	***	0,734	***
logLikelihood	-8053,4		-2030,6	
AIC	16127,0		4081,1	
residual autocorr	0,003		0,020	
levels of significance	***	<1%		
	**	<5%		
	*	<10%		

Table 23 : SAR Results for Paris with grid of 1140 cells and comparison with results of previous chapter

The final step is the ranking of "arrondissements" according to the power of price diffusion. As before we use a maximum distance of 500m.

The results of the SAR by exclusion are shown in the following table 24.

ARROND	PRIXM2	SAR
1	7769,2	0,719
2	7288,3	0,708
3	6789,9	0,701
4	8333,6	0,710
5	8218,9	0,696
6	9181,6	0,703
7	9184,0	0,700
8	8421,3	0,709
9	6176,8	0,708
10	5286,2	0,708
11	5930,8	0,680
12	5910,8	0,701
13	6168,9	0,712
14	6365,3	0,704
15	6513,8	0,653
16	7304,5	0,721
17	6293,4	0,730
18	5099,8	0,704
19	5022,7	0,670
20	5139,1	0,706
Paris	6517,4	0,709

Sorted by rho

ARROND	PRIXM2	SAR	PATTERN
15	6513,8	0,653	0,030
19	5022,7	0,670	0,055
11	5930,8	0,680	0,057
5	8218,9	0,696	0,055
7	9184,0	0,700	0,090
12	5910,8	0,701	0,035
3	6789,9	0,701	0,059
6	9181,6	0,703	0,038
14	6365,3	0,704	0,041
18	5099,8	0,704	0,079
20	5139,1	0,706	0,048
2	7288,3	0,708	0,070
10	5286,2	0,708	0,038
9	6176,8	0,708	0,025
8	8421,3	0,709	0,106
4	8333,6	0,710	0,037
13	6168,9	0,712	0,050
1	7769,2	0,719	0,043
16	7304,5	0,721	0,117
17	6293,4	0,730	0,038

Dmax = 500m

Table 24 : SAR results by exclusion

The 15th district keeps the place of the district with the highest diffusion. The 5th arrondissement remains influential in this new ranking. It is interesting to note the appearance of the 19th and 11th.

In the bottom of the rankings, we still find the districts 16 and 17.

To sum up, this new method confirms the general trend of the previous chapter.

Regarding the link between PATTERN and diffusion power, we want to check if the new method confirms our hypothesis.

As remind, we raise the hypothesis that the diffusion power is affected by the pattern of factor quotas within the districts. That pattern can be revealed by the cell means belonging to an arrondissement. If these means deviate much from the city averages, the contribution to ρ should decline by weak diffusion. The opposite case of strong diffusion should be created by a small deviation, thus the pattern conforms to the city average.

The new method confirms this hypothesis for most of the districts but we have to be careful with some exception districts. Some rather heterogeneous districts have a fairly strong diffusion, the 7th arrondissement for example.

CONCLUSIONS

The spatial econometric model of price formation in the city of Paris brought about a number of unexpected and new results. To make the SAR model viable we designed first a grid over Paris, which provided the data aggregated to the meso-scale of grid cells used in estimation. Central for our analysis is the categorization into high price inner and lower price outer districts, as well as the chacteristics of the sold properties, which are classified according to their deviation from the city average.

From simple OLS estimation of the survey we obtained the first result that the estimators in the outer districts are more significant than in the inner ones, most likely because the inner districts are marked by preferences and wealth endowments that create a high variability of prices. But, since the price is mainly explained by the constant, the OLS-estimators cannot satisfy. This improves indeed in the spatial setting.

In SAR-estimation we tested alternative structures of the cell neighborhoods, with different maximum distances to the neighboring cells. In all cases, the price response is positive for new, large and duplex apartments. It is negative when the purchasers are workers, craftsmen and small merchants, and it is slightly positive among young purchasers up to 30 years. For the evaluation of the price diffusion the model LINEAR with the smallest maximum distance and the smallest spatial autocorrelation was chosen. For this model we calculated the response of prices to changing object and personal characteristics, or factors for short.

The ranking of districts according to their contribution to rho corroborated the hypothesis that districts creating a strong diffusion, hence containing quartiers primes, have factor quotas close to the city average, while districts creating weak diffusion have factor quotas that are distinct. The hypothesis could be confirmed for the strong diffusion districts 12 and 15, and for the weak diffusion districts 8 and 16. The price leadership of the strong diffusion district 5 can be interpreted from an urban perspective.

A important question arises from the limits of the grid and the possibility of using individual data. In the last chapter different approaches were examined that permit to estimate the original survey data, and to validate main findings.

Random sampling among the individual observations has confirmed the significance of objects characteristics and WLNDENS.

Finally, the basic hypothesis about the causes of diffusion deserves further elaboration. If it is true, the implication for urban policies might be important. On the lines of research and expertises that recommend mixed social structures, if the various parts of city come close to a city average of mixed structures and moderate average city prices, such a structure would create a diffusion process that moderates the price all over the community.

BIBLIOGRAPHY

1. **Silverman, Bernard. W.** *Density Estimation for Statistics and Data Analysis.* 1986.

2. **James LeSage, Robert Kelley Pace.** *Introduction to Spatial Econometrics.* 2009.

3. **Baltagi and Bresson.** Maximum likelihood estimation and Lagrange multiplier tests for panel seemingly unrelated regressions with spatial lag and spatial errors: an application to hedonic housing prices in Paris. *Journal of Urban Economics.* 2011.

4. **Bhattacharjee, Arnab, Castro, Eduardo and Marques, Joao.** Spatial Interactions in Hedonic Pricing Models : The Urban Housing Market of Aveiro, Portugal. *Spatial Economic Analysis.* 2012, Vol. 7, 1.

5. **Crawley, Michael J.** *The R Book.* 2012.

6. **Bhattacharjee, Arnab and Holly.** Interaction in social networks and committees. *Spatial Economic Analysis.* 2013, Vol. 8, 1.

7. **Anselin, L., Florax, R.J., Rey, S.J.** *Advances in spatial econometrics. Methodology, tools and applications.* s.l. : Springer, 2004.