Measuring the Spatio-Temporal Diffusion of Housing Rental Prices in Barcelona*

Oriol González Casasús

Final Year Project†
2017/2018

Abstract

Changes in housing rental prices in Barcelona are increasingly a source of debate. In this paper, I use a Spatial Durbin Model (SDM) to investigate what are the determinants of price movements with two different specifications for the weights matrix. I provide the results by distinguishing between the short and the long-run, and the direct and indirect effects. The findings suggest that there is clearly a spatio-temporal diffusion of housing rental prices in Barcelona. However, the estimates show that specification of the weights matrix is decisive to assess the effects of the different determinants. I also provide graphs with the estimated spatial Impulse Response Functions (IRF) to get a more clear idea of how housing rental prices spread across space and time.

JEL Classification: C23, C53, R31, R32.

Keywords: Spatial econometrics; Spatial Dynamic Panel Data models; Maximum Likelihood Estimation; Fixed effects; IRF; Housing rental prices; Barcelona.

*I am thankful to Víctor Quintás Martínez and Þórunn Helgadóttir for setting up the idea with me and for providing me with great contributions. I also thank Majid Al-Sadoon for his tutorage throughout this project and for his helpful comments and suggestions. All errors are my own.

†Degree in Economics, Universitat Pompeu Fabra. Project code: EDE02.

Contents

1 Introduction

Barcelona is the city in Spain with the highest housing prices, both of renting and buying[1]. These have increasingly been the subject of considerable debate, with the focus on its determinants. Specially, following the emergence of Airbnb and the rising number of tourists visiting the city of Barcelona.

Since 2013, rental prices have increased by 29% while property prices by 24%[1], much more than what wages have increased, therefore decreasing the purchasing power of the families and hence concerning both citizens and politicians. This has led to complaints of different neighbours' associations, such as the *Federació d'Associacions de Veïnes i Veïns de Barcelona* (FAVB), and to the creation of the first "tenants' union" in the entire State[2].

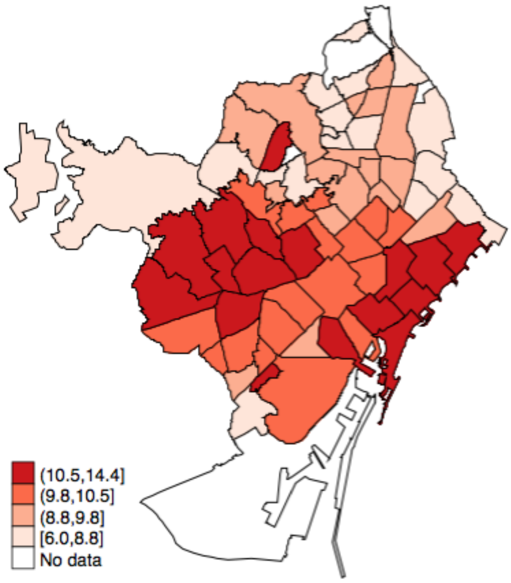


Figure 1: Distribution of housing monthly rents per m^2 by neighbourhood in Barcelona

This rise in prices has happened in an uneven way, yielding the location of a certain house to be a decisive factor for the resulting change in its price. In Figure 1 we can see how important this is; neighbourhoods with the highest rental prices are concentrated in two points, and prices decrease as soon as the neighbourhoods are further from them. Indeed, if we observe the

[1] Grau, X. (2018, April 8). *El flux i reflux metropolità que fa augmentar encara més el lloguer.* Retrieved April 20, 2018, from `https://www.ara.cat`.

[2] Vicens, L. (2017, May 9). *Neix a Barcelona el primer sindicat de llogaters per lluitar contra els preus "abusius" dels pisos.* Retrieved June 1, 2018, from `https://www.ara.cat`.

evolution of this plot from 2013 to the present period, we can see that the red has become darker the closer a certain neighbourhood is from these points, showing signs of spatial correlation[3].

Moreover, 38.2% of all homes in Barcelona are being rented, among which 72% are at short-term[4], what makes it difficult for families to settle down because of the housing uncertainty in the future. The larger percentage of houses being rented is located around El Barri Gòtic, where up to 80% of the houses are being rented, and it decreases as soon as we move away from it, the northern neighbourhoods being the ones with lower percentages[5]. These numbers show that renting is a major factor in the city of Barcelona, which motivates the research to find out what are the determinants for price changes.

When analysing the research conducted in the literature, in a number of studies spatial correlation has proven to be a significant factor in explaining housing prices across different regions. The degree to which a change in housing prices in a particular area spreads to another is therefore an important factor in determining the development of housing prices over time.

One of the main papers that show the potential of the spatio-temporal models is Brady (2011). He first estimates different spatial autoregressive models with a dynamic panel of average county housing prices from 31 California counties, monthly from 1995 to 2002. Then, he applies Jordà's (2005) local projection method in order to estimate impulse response functions. His results suggest that the diffusion of regional housing prices across space lasts up to two and a half years. Some years later, in Brady (2014), he extends his study to regional housing prices across U.S. states. There he finds that spatial diffusion is statistically significant for approximately three to four years following a housing price increase across regions. However, he only provides estimates constructing $\mathbf{W}$ by contiguity "for simplicity", and does not check different specifications to assure the robustness of his estimates, as suggested by different authors[6].

[3]See Appendix 1 for a Moran's I test of spatial correlation.

[4]Armengou, P.J. (2018, April 14). *Les llars que viuen de lloguer a Barcelona ja són prop del 40%.* Retrieved April 14, 2018, from `https://www.ara.cat`.

[5]See Appendix 2.

[6]See section 3.2 for more on that.

A similar result is found by Beenstock, M. and Felsenstein, D. (2007), who estimate a Spatial VAR model using annual data for Israel over the period 1987-2004 for nine regions and four variables: earnings, population, house prices and housing stock. The findings suggest that shocks die after about four years.

The majority of the literature on the diffusion of housing prices available is from the United States and I have not yet discovered a similar paper that covers the neighbourhoods of Barcelona. To my knowledge, the closest study to mine is McGreal, S. and Taltavull, P. (2013) who focus their analysis on Spanish provinces. They find that time and space changes in prices are mainly due to property size and economic and demographic characteristics. However, this study is too far to have external validity for the city of Barcelona. A specific analysis as the one I will do here is needed to answer whether there exists, and to what extent, spatio-temporal price diffusion in the city of Barcelona. Furthermore, the fact of checking different specifications for the weights matrix reinforce the robustness of the estimates presented.

My findings suggest that there exists a diffusion of housing rental prices across space and time. The main determinant of such movements appears to be the Euribor. Also, the unemployment rate shows an important direct effect in the short and in the long-run. When using a distance weights matrix specification in particular, other determinants such as population or the Index of Industrial Production (IPI) become of importance as well.

I therefore believe this paper will be a good addition to the research that has already been conducted on the housing prices in Barcelona and hopefully add something to the discussion that is already taking place. In this paper, I will explain what the spatio-temporal autoregressive (STAR) models are, their causal interpretation and how we can consistently identify the parameters of interest. Then, I will go over the data used and where it has been obtained. Finally, I will provide the results of the model estimations and the estimated spatial IRFs, to finish with some final conclusions and suggestions towards where future research should point.

2 Model

2.1 Specification

Anselin (2001)'s systematic description of the spatio-temporal autoregressive (STAR) models is frequently referenced in the literature[7]. The general form of the STAR model can be written as a dynamic panel equation:

$$y_{it} = \rho \sum_{j=1}^{N} w_{ij} y_{jt} + \alpha y_{i,t-1} + \gamma \sum_{j=1}^{N} w_{ij} y_{j,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$. We can express the model in vector form as:

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_t + \alpha \mathbf{y}_{t-1} + \gamma \mathbf{W} \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \qquad (1)$$

for $t = 1, 2, \ldots, T$. Here, $\mathbf{y}_t = (y_{1t},\, y_{2t},\, \ldots,\, y_{Nt})'$ is the $N \times 1$ vector of observations of the dependent variable for each individual at time t, $\mathbf{X}_t = (\mathbf{x}_{1t},\, \mathbf{x}_{2t},\, \ldots,\, \mathbf{x}_{Nt})'$ is an $N \times K$ matrix of (K) other explanatory variables for each individual at time t and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t},\, \varepsilon_{2t},\, \ldots,\, \varepsilon_{Nt})'$ is the $N \times 1$ vector of error terms for each individual at time t.

The $N \times N$ matrix $\mathbf{W} = (w_{ij})$ is called the *spatial weights matrix* of the STAR model, and each entry w_{ij} captures the extent to which the value of the dependent variable of individual j influences the value of the dependent variable of individual i. A possible specification for $\mathbf{W}$ is $w_{ij} = \mathbf{1}\{j \in S_i\}$, where the entry w_{ij} is 1 if individual j belongs to a neighbourhood S_i of individual i and 0 otherwise. When the spatial units are points, we may think of the neighbourhood as being determined by an upper threshold on the geographical distance between i and j. When the spatial units are regions, as it is our case, the neighbourhood of region i may be taken as the set of all its contiguous regions (the regions with which i shares a border).

[7]Also called Spatial Dynamic Panel Data (SDPD) models. See Anselin (2010) for a discussion on the evolution of Spatial Econometrics.

In both cases, according to Anselin (2001), it is common to normalize $\mathbf{W}$ so that $\sum_{j=1}^{N} w_{ij} = 1$ for each $i = 1, 2, \ldots, N$, i.e. rows add up to one. Then, we can think of $\mathbf{W}\mathbf{y}_t$ as a weighted average of the neighbours value of the dependent variable, with w_{ij} as weights.

Anselin (2001) classifies the class of models described by (1) in four categories:

1. If $\rho = \alpha = 0$ and $\gamma \neq 0$, the model is *pure space-recursive*: spatial dependence comes only from lagged values of neighbours.

2. If $\rho = 0$ and $\alpha, \gamma \neq 0$, the model is *time-space recursive*: spatial dependence comes from lagged values of neighbours and of the same individual.

3. If $\gamma = 0$ and $\rho, \alpha \neq 0$, the model is *time-space simultaneous*: spatial dependence comes from contemporaneous values of neighbours and lagged values of the same individual. In the literature, this model is also often referred to as Spatial Autoregressive (SAR) model. We can extend it to include the spatial lag of the independent variables as well, what yields the Spatial Durbin Model (SDM), of the form $\mathbf{y}_t = \rho\mathbf{W}\mathbf{y}_t + \alpha\mathbf{y}_{t-1} + \mathbf{W}\mathbf{X}_t\boldsymbol{\beta}_1 + \mathbf{X}_t\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_t$.

4. If $\rho, \alpha, \gamma \neq 0$, the model is *time-space dynamic*: spatial dependence comes from all of the sources above.

In his study for the Californian housing market, Brady (2011) assumes a time-space simultaneous (or SAR) model of the form

$$\mathbf{y}_t = \rho\mathbf{W}\mathbf{y}_t + \alpha\mathbf{y}_{t-1} + \mathbf{X}_t\boldsymbol{\beta} + \boldsymbol{\delta} + \boldsymbol{\varepsilon}_t, \tag{2}$$

to which he also incorporates an $N \times 1$ vector of individual fixed effects $\boldsymbol{\delta} = (\delta_1,\, \delta_2,\, \ldots,\, \delta_N)'$. He takes as the dependent variable in (2) the logarithm of average house sale prices in county i in month t. As other explanatory variables, he uses the unemployment rate, the logarithm of the number of new houses built per month and the logarithm of population at the county level, as well as the mortgage rate and the monthly industrial production index at the national level. I will follow a similar specification. I will assume that $\boldsymbol{\varepsilon}_t \sim i.i.d.N(0, \sigma_\varepsilon^2 \mathbf{I}_N)$.

2.2 Causal interpretation

Elhorst (2011) stresses the importance of not directly comparing different spatial models according to the coefficient estimates, variance-covariance matrix, standard errors and t-values because of the sensitivity to the weights matrix specification. LeSage and Pace (2009) suggest an approach that allows to do so which consists on a partial derivative method. What they show is that the effects of individual variables in a model are composed of a partial derivative of a combination of all model coefficients and the weights matrix, and that using the correct partial derivative interpretation of the parameters from various models results in less divergence in the inferences from different model specifications.

Grosso modo, the method is the following. We have the dynamic equation

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_t + \alpha \mathbf{y}_{t-1} + \mathbf{X}_t \boldsymbol{\beta}_1 + \mathbf{W} \mathbf{X}_t \boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}_t$$

which can be rewritten as

$$\mathbf{y}_t = (\mathbf{I} - \rho \mathbf{W})^{-1} (\alpha \mathbf{I}) \mathbf{y}_{t-1} + (\mathbf{I} - \rho \mathbf{W})^{-1} (\mathbf{X}_t \boldsymbol{\beta}_1 + \mathbf{W} \mathbf{X}_t \boldsymbol{\beta}_2) + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}_t$$

Then, the matrix of partial derivatives of the expected value of $\mathbf{y}$ with respect to the k^{th} explanatory variable of $\mathbf{X}$ in unit 1 up to unit N at a particular point in time t, i.e. in the short-run, is

$$\left[\frac{\partial \mathbb{E}[\mathbf{y}]}{\partial x_{1k}} \cdots \frac{\partial \mathbb{E}[\mathbf{y}]}{\partial x_{Nk}} \right]_t = (\mathbf{I} - \rho \mathbf{W})^{-1} [\beta_{1k} \mathbf{I}_N + \beta_{2k} \mathbf{W}]$$

while, similarly, in the long-run is

$$\left[\frac{\partial \mathbb{E}[\mathbf{y}]}{\partial x_{1k}} \cdots \frac{\partial \mathbb{E}[\mathbf{y}]}{\partial x_{Nk}} \right] = ((1 - \alpha) \mathbf{I} - \rho \mathbf{W})^{-1} [\beta_{1k} \mathbf{I}_N + \beta_{2k} \mathbf{W}]$$

They also distinguish between the direct and indirect effects, the latter also called *spatial spillovers* for contemporaneous effects and *diffusion effects* when involving different time periods. The following list summarizes the results:

- Short-run Direct effect: $[(\mathbf{I} - \rho\mathbf{W})^{-1}(\beta_{1k}\mathbf{I}_N + \beta_{2k}\mathbf{W})]^{\bar{d}}$

- Short-run Indirect effect: $[(\mathbf{I} - \rho\mathbf{W})^{-1}(\beta_{1k}\mathbf{I}_N + \beta_{2k}\mathbf{W})]^{\overline{rsum}}$

- Long-run Direct effect: $[((1-\alpha)\mathbf{I} - \rho\mathbf{W})^{-1}(\beta_{1k}\mathbf{I}_N + \beta_{2k}\mathbf{W})]^{\bar{d}}$

- Long-run Indirect effect: $[((1-\alpha)\mathbf{I} - \rho\mathbf{W})^{-1}(\beta_{1k}\mathbf{I}_N + \beta_{2k}\mathbf{W})]^{\overline{rsum}}$

where the superscript $\bar{d}$ refers to the operator that calculates the mean diagonal element of a matrix and $\overline{rsum}$ denotes the operator that calculates the mean row sum of the non-diagonal elements[8]. Total effects are just the sum of the direct and indirect effects.

2.3 Estimation

As claimed by Arselin (2001), the standard panel regression models cannot be consistent because of the endogeneity of the spatial regressor $\mathbf{Wy}_t$ in equation (2). Different approaches have been proposed in order to consistently estimate the parameters of interest in a static spatial regression, such as maximum likelihood estimation (Ord, 1975), Bayesian Markov Chain Monte Carlo (LeSage, 1997), an instrumental variables generalized moments (IV/GM) approach suggested by (Kelejian and Prucha, 1998, 1999), spatial filtering (Griffith, 2003), generalized maximum entropy (Marsh and Mittelhammer, 2004), and use of matrix exponential transformations (LeSage and Pace, 2007). Among these, the bias-corrected maximum likelihood (ML) or quasi-maximum likelihood (QML) estimator, the instrumental variables or generalized method of moments (IV/GMM) method and the Bayesian Markov Chain Monte Carlo (MCMC) approach have been extended to dynamic spatial panels, solving for the endogeneity of the time

[8]For the complete derivation of this approach and a deeper explanation see LeSage and Pace (2009) and Elhorst (2011).

lag. However, two of these methods have acquired more use with respect to the rest, namely maximum likelihood and instrumental variables. For the former, a special (bias-corrected) likelihood function is needed for MLE to be consistent[9]. Lee, L. and Yu, J. (2010) prove that MLE appears to have excellent finite sample properties even when both N and T are quite small. Also, Elhorst (2005) shows that estimating a first-differenced model (to eliminate fixed effects) by ML yields a consistent estimator of the response parameters and the spatial autocorrelation coefficient when $N \to \infty$ regardless of the dimension of T. For the latter, it is common to use spatial lags of the explanatory variables (i.e. $\mathbf{WX}$) as instruments for $\mathbf{Wy}_t$. The estimation can be done by 2SLS or by GMM, but the moment conditions are very complicated for spatial models, so the first method is usually prioritized[10]. To address the dynamic dependence in equation (2), the standard techniques can be used. Anselin (2001) notes that the conditions for consistency in the model will be the same as in the standard dynamic panel model, so that, in the presence of fixed effects, the equation can be consistently estimated by taking first differences and then instrumenting the endogenous variable $\Delta\mathbf{y}_{t-1}$ by $\mathbf{y}_{t-2}$ (or further lags) à la Arellano and Bond (1991).

In the case of a short panel (T fixed and $N \longrightarrow \infty$), there is no consistent estimator for the individual effects $\boldsymbol{\delta}$ (because of the incidental parameter problem), so Anselin (2001) argues in favor of a specification based in random effects rather than in fixed effects. However, Elhorst (2003) states that, even for short panels, the inconsistency of the estimator for the individual effects $\boldsymbol{\delta}$ does not threaten the consistency of the parameters of interest in $\boldsymbol{\beta}$.

In this paper, I will carry out my estimations by the maximum likelihood approach because, as Elhorst (2010) notes, the number of studies regarding IV/GMM estimators of spatial panel data models is relatively scant.

[9]See Lee, L. and Yu, J. (2009) for derivations.

[10]Badinger et al. (2004) state that there is not a direct GMM estimator and that estimating a spatial dynamic model by GMM involves challenging complex moment conditions.

3 Data

3.1 Variables of interest

The data I will be using is from the Department of Statistics of *Ajuntament de Barcelona* and includes information on different demographic, social, labour and economic indicators. The main advantage of using this data is that it is organized by neighbourhoods (total of 73; N=73), which provides me with a large cross-section and hence allows for potential spatial interactions. I will be using quarterly data from the first quarter of 2014 (first time period where information is broken down by neighbourhood) until the last quarter of 2017 (last update), and so I will work with 16 time periods (T=16). The main variables of interest are the logarithm of the average price per squared meter in the neighbourhood, the number of contracts signed in the neighbourhood, the unemployment rate in the neighbourhood, the population in the neighbourhood, the Index of Industrial Production (IPI), and the Euribor, in order to control for the evolution of the mortgage rate.

	Ciutat Vella	Eixample	Sants-Montjuïc	Les Corts	Sarrià-St Gervasi
price	13.74	11.75	11.04	12.94	13.89
contracts	267.29	363.07	158.48	153.97	170.85
unemployment	10.38	7.22	9.24	6.53	4.67
population	25141.06	44076.83	25710.96	27200.42	24579.17

	Gràcia	Horta-Guinardó	Nou Barris	Sant Andreu	Sant Martí
price	11.84	10.34	9.15	9.66	11.38
contracts	210.62	90.94	76.42	110.56	116.18
unemployment	7.64	9.57	13.11	10.75	9.61
population	24146.95	16681.8	16876.53	24033.04	23393.22

Table 1: Means of the main variables by district for the period 2014-2017

Table 1 provides the average of the space-variant variables of interest for the entire period of study for each district. We can see that there is large variability in all of them. Although it is not at neighbourhood level[11], it still allows us to get an initial idea of the spatial differences.

3.2 Spatial information

For the spatial information, I obtained a shapefile (SHP) containing the information on the coordinates of the neighbourhoods in Barcelona from the Cartographic Institute of *Ajuntament de Barcelona*[12]. This allows to construct the spatial matrices to be used in the regressions.

There are different possible ways to construct the spatial weights matrix, such as contiguity, assigning a value of 1 if two neighbourhoods are adjoining and 0 otherwise (Brady, 2009); distance, giving a lower value the further away a neighbourhood is (Anselin, 1980); on the structure of a social network (Doreian, 1980); on economic distance (Case, Rosen and Hines, 1993); k-nearest neighbours (Pinkse and Slade, 1998); among others. As previously explained, the main two ways to construct it are by contiguity and by (inverse) distance. It is straightforward to see that these matrices will be symmetric. We can plot the information contained in these matrices to see how neighbourhoods are distributed in space[13]. The intensity plots in Figure 2 give a darker value the larger the value in the matrix. Hence, for the contiguity matrix it will be either black or white, and for the distance matrix it will have different shades of gray.

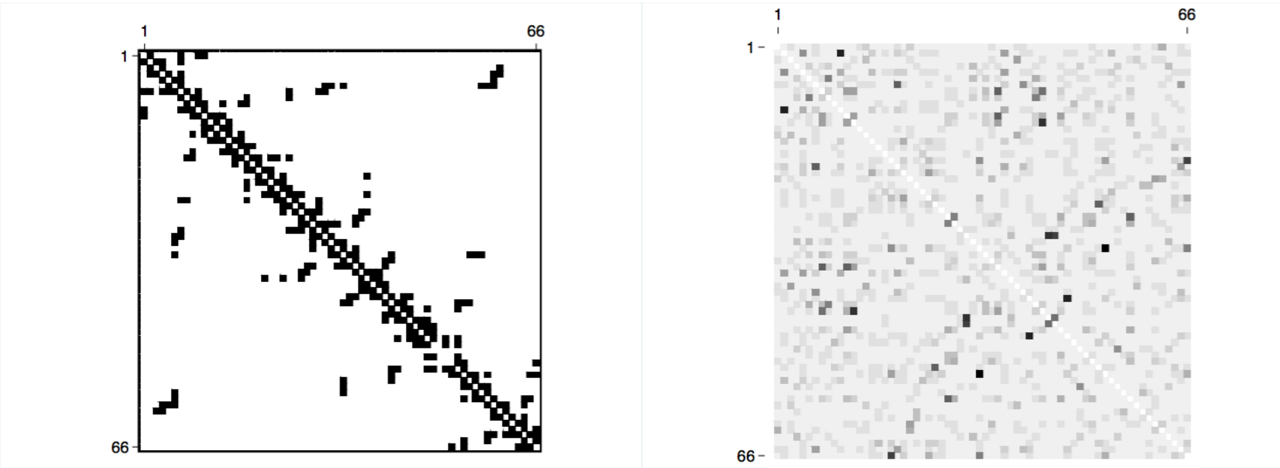
Different authors warn that the specification of the spatial weights matrix has a direct effect on the estimation results[14]. There are two main trends in literature, those who assume a certain construction of the spatial weights matrix and those who use different specifications in order to check the robustness of their estimations. Also recently, some studies such as Bhattacharjee and Jensen-Butler (2006) have started to propose methodologies to empirically estimate these

[11]For practical reasons, since there are 73 neighbourhoods while only 10 districts.

[12]The file is ETRS89. DIVISIONS ADMINISTRATIVES, retrieved from `http://w20.bcn.cat/cartobcn/`.

[13]There are 66 neighbourhoods instead of the original 73 because I had to get rid of those who did not have enough information on housing rental prices to be able to get consistent estimates.

[14]See, for example, Anselin (1988), Anselin (1999) and Tiefelsdorf, Griffith, and Boots (1999).



(a) Contiguity Matrix (b) Inverse-distance Matrix

Figure 2: Intensity plots of the Spatial Matrix, computed by the two different procedures

spatial weights matrices. I am not going to do so here because what I am mainly interested in is to consistently identify the coefficients of the spatial and time lags, and as Manski (1993) notes, in order to avoid identification problems, the weights should truly be exogenous to the model. I will provide the results using the two weighting matrices above in order to assure the robustness of my estimates. For the distance matrix, the distance function I use is the Euclidean distance.

As previously explained, it is common to normalize these matrices because in applications it could happen that the un-normalized matrix $\mathbf{W}$ causes $(\mathbf{I} - \rho\mathbf{W})$ to be singular for some values of ρ. I apply a spectral normalization, which consists in dividing each element by the modulus of the largest eigenvalue of the matrix. As Kelejian and Prucha (2010) note, this avoids a possible misspecification of the model and the problematic of comparing rows arising from row-normalization.

4 Results

Despite having assumed that the true DGP in determining the level of housing rental prices in Barcelona follows a Spatial Autoregressive (SAR) model, I will estimate a Spatial Durbin Model (SDM), i.e. I will include as other explanatory variables the spatial lag of the current explanatory variables ($\mathbf{WX}_t$)[15]. The reason is that, as LeSage and Pace (2009) claim, including these variables when the true data generating process is a SAR model does not threaten the unbiasedness of the estimates for the explanatory variable parameters. However, estimating a SAR model when the true DGP is a SDM could lead to omitted variables problem, and consequently biased estimates. This result is also supported by Lee, L. and Yu, J. (2016), by means of a Monte Carlo experiment. This way I guard myself against having biased estimates. Then, it remains to test the joint significance of the estimates of the coefficients of $\mathbf{WX}_t$ with an F-test to check whether they are statistically different from zero and see therefore whether including them worsens the efficiency of the estimates.

As explained in section 2.2, the point estimates obtained from a spatial regression are not directly interpretable. Hence, I will use the method described in LeSage and Pace (2009) and directly provide the final interpretable results. Bear in mind that the explanatory variables are expressed in levels while the dependent variable is expressed in logs, so the interpretation of the coefficients must be that by changing the explanatory variable by 1 unit, our dependent variable will change on average by $100 \times \beta$ percent, *ceteris paribus.*

Table 2 gathers the estimates of the direct and indirect effects for both the short and the long-run, using the two different procedures to construct a matrix described above. The first thing that stands out is that both the spatial and the time lag are statistically significant at the 1% significance level for both types of matrix specifications. This makes it clear that there is indeed spatio-temporal diffusion. So, next paragraph investigates its determinants.

Overall, we can see that the variable Euribor is the only one which is significant across most of the cases. Its coefficients are relatively large and negative, so any increase in the Euribor is likely to lead to an important decrease in the rental prices, both in the short and in the long-run.

[15] I do not consider the spatial lag of the variables IPI and Euribor because they have the same values for all the neighbourhoods for each time period.

	Dependent variable: log rent			Dependent variable: log rent	
	Contiguity	Distance		Contiguity	Distance
Spatial lag	0.1786***	0.4426***	Time lag	0.2763***	0.2488***
	(0.0515)	(0.0929)		(0.0305)	(0.0309)
Short-run			**Long-run**		
Direct			Direct		
N. of contracts	-0.0001	-0.0000	N. of contracts	-0.0002	-0.0001
	(0.0001)	(0.0000)		(0.0002)	(0.0001)
Unemployment	-0.0165***	-0.0008	Unemployment	-0.0228***	-0.0011
	(0.0055)	(0.0062)		(0.0076)	(0.0082)
Population	0.0000	0.0000	Population	0.0000	0.0000
	(0.0000)	(0.0000)		(0.0000)	(0.0000)
IPI	0.0035	0.0104***	IPI	0.0048	0.0139***
	(0.0022)	(0.0038)		(0.0030)	(0.0051)
Euribor	-0.1249***	-0.0602*	Euribor	-0.1736***	-0.0810*
	(0.0268)	(0.0364)		(0.0371)	(0.0489)
Indirect			Indirect		
N. of contracts	0.0004**	0.0008**	N. of contracts	0.0005**	0.0015*
	(0.0001)	(0.0003)		(0.0002)	(0.0009)
Unemployment	0.0005	0.0025	Unemployment	-0.0010	0.0083
	(0.0057)	(0.0177)		(0.0082)	(0.0405)
Population	-0.0000	-0.0004**	Population	0.0000	-0.0007
	(0.0000)	(0.0002)		(0.0000)	(0.0005)
IPI	0.0006	0.0079*	IPI	0.0013	0.0212
	(0.0005)	(0.0043)		(0.0010)	(0.0182)
Euribor	-0.0227***	-0.0462	Euribor	-0.0472**	-0.1218
	(0.0086)	(0.0338)		(0.0190)	(0.1131)
N	66	66	Log-likelihood	1353.6032	1365.2885
T	15	15	AIC	-2693.231	-2716.424

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: MLE results for a SDM for both types of spatial matrices

Unemployment seems to have an important direct effect both in the short and in the long-run, when using a contiguity weights matrix. That means that it is specially important, in order to determine the rental price of a house, the unemployment rate of that same neighbourhood. The estimates suggest that if the unemployment rate in a specific neighbourhood increases by 1 percentage point, the rental prices in that same neighbourhood will decrease, on average, by 1.65% in the short-run and 2.28% in the long-run, *ceteris paribus*. The population seems to have very little effect, and only shows significance in the short-run in an indirect way, when using the distance matrix. The Index of Industrial Production (IPI) is specially significant when using the distance weights matrix. It influences the price in the short-run in a direct and indirect way, and in a direct way in the long-run.

Last, point out that, although there is a large variation in the statistical significance of the determinants depending on the type of weights matrix used, the sign of the coefficient remains the same. These changes in the significance of some coefficients could be due to the fact that, as later shown in section 4.1, the true DGP is a SAR model when using the contiguity matrix, while when using the distance matrix the estimates suggest it is a SDM.

4.1 Robustness checks

So far, the biggest assumption made regarding the model is that the true DGP follows a SAR model. In order to check what is the true DGP, we can simply test the joint significance of the coefficients of $\mathbf{WX}_t$. The F-test leads to reject the null hypothesis that the coefficients are jointly equal to zero at the 3% significance level when using the distance matrix. However, we can only reject the null at the 13.78% significance level when we estimate the model with the contiguity weights matrix. This indicates that the true DGP follows a Spatial Autoregressive (SAR) model if the specification of the weights matrix is by contiguity and a Spatial Durbin Model (SDM) if it is by distance, so we indeed needed to estimate a SDM in order to obtain unbiased estimates. Still, some may argue that the spatial dependence can also be on the error component, leading to a Spatial Error Model (SEM). If we test this, we obtain a result that suggests that it follows a SEM at the 14.61% significance level, when using a contiguity matrix, and at the 4.21% significance level, when using the distance matrix. I do not consider this result determinant enough to be concerned about my estimates in case the true DGP is a SEM. Moreover, as LeSage and Pace (2009) claim, "the cost of ignoring spatial dependence in the dependent variable is relatively high since biased estimates will result if this type of dependence is ignored. In addition, ignoring this type of dependence will also lead to an inappropriate interpretation of the explanatory variable coefficients as representing partial derivative impacts arising from changes in the explanatory variables. In contrast, ignoring spatial dependence in the disturbances will lead to a loss of efficiency in the estimates. As samples become large, efficiency becomes less of a problem relative to bias".

A final observation could be made about the missing data. We can see that for 7 neighbourhoods, out of the total 73, there is not (enough) information on housing rental prices, yielding a sample size of 66 individuals. This situation should not affect the estimates in a significant way, since they represent a small fraction of the entire population and they happen to be the neighbourhoods with the lowest renting activity. Henceforth, however, it would be interesting to be able to collect data from those neighbourhoods to make sure whether this statement indeed holds.

4.2 Spatial Impulse Response Functions

With the estimates just obtained, we can construct spatial Impulse Response Functions to better understand this spatio-temporal diffusion[16]. By expressing the spatio-temporal equation in reduced form, we have

$$\mathbf{y}_t = \rho \mathbf{W} \mathbf{y}_t + \alpha \mathbf{y}_{t-1} + \mathbf{\Gamma}_t + \boldsymbol{\varepsilon}_t$$

This equation can be rewritten as

$$\mathbf{y}_t = (\mathbf{I} - \rho \mathbf{W})^{-1} \alpha \mathbf{y}_{t-1} + (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{\Gamma}_t + (\mathbf{I} - \rho \mathbf{W})^{-1} \boldsymbol{\varepsilon}_t$$

where $\mathbf{\Gamma}_t$ is $\mathbf{X}_t \boldsymbol{\beta}_1 + \mathbf{W} \mathbf{X}_t \boldsymbol{\beta}_2$ in the case of a SDM and $\mathbf{X}_t \boldsymbol{\beta}$ when using a SAR model. The spatial Impulse Response Function (IRF) of a shock at period t on the housing rental prices can be expressed as[17]

$$\begin{aligned}
\text{At } t = 1: \quad & \frac{\partial \mathbf{y}_t}{\partial \boldsymbol{\varepsilon}_t} = (\mathbf{I} - \rho \mathbf{W})^{-1} \\
\text{At } t = 2: \quad & \frac{\partial \mathbf{y}_{t+1}}{\partial \boldsymbol{\varepsilon}_t} = \alpha (\mathbf{I} - \rho \mathbf{W})^{-2} \\
\vdots \quad & \\
\text{At } t = k: \quad & \frac{\partial \mathbf{y}_{t+k}}{\partial \boldsymbol{\varepsilon}_t} = \alpha^{k-1} (\mathbf{I} - \rho \mathbf{W})^{-k}
\end{aligned}$$

This allows us to see, given a shock to a specific neighbourhood, how does this effect spread across all the neighbourhoods and time periods through prices. To analyse this, we can plot

[16] The name "Spatial Impulse Response Functions" follows the notation in Brady(2009) and Brady(2014).
[17] Assuming $t = 1$ is the current period.

the results in a map graph in order to obtain an easier interpretation. I will assume there is a shock at present period ($t = 1$) in the neighbourhood La Dreta de l'Eixample, everything else constant. I will show what happens on the first four periods following the shock, i.e. on the following year, using both types of weights matrices. One reason to choose this neighbourhood responds to one of the motivations of this study; see how the recent tourism increase can affect the prices. La Dreta de l'Eixample is one of the neighbourhoods where there are the most houses dedicated to tourists, where more than 50% of the house rental supply is earmarked to tourism. This is something that really worries the neighbours, as the *Federació d'Associacions de Veïnes i Veïns de Barcelona* claims[18]. Moreover, l'Eixample is the district with the largest population, so it is important to analyse any price change there because it will affect a larger number of individuals. In any case, this analysis can easily be extended to other neighbourhoods.

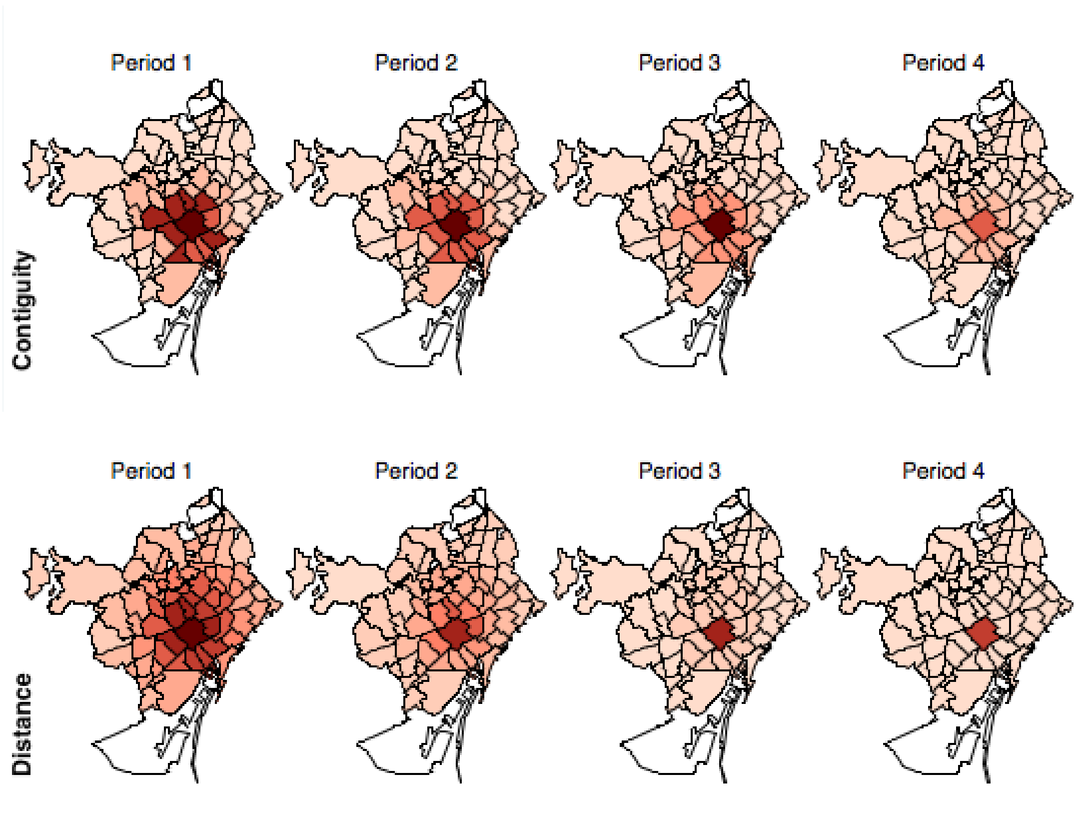


Figure 3: IRF estimates to a shock in La Dreta de l'Eixample using both weights matrices

[18]Mezquita, R. (2017, March 30). *Los pisos turísticos superan el 50% de los alquileres en Ciutat Vella y la Dreta de l'Eixample.* Retrieved June 2, 2018, from `https://cronicaglobal.elespanol.com`.

In Figure 3, we can see the estimates of these IRFs for the first four periods following the shock, computed for each weights matrix. The results obtained using the contiguity matrix are the ones shown at the top, whereas in the second row we can find the estimates obtained with the distance matrix. Higher values have a darker tone of red. As the shock fades away, the colour becomes more clear.[19]

5 Conclusions

The results of this paper provide an insight into the determinants of housing rental prices in the city of Barcelona. The estimations for the neighbourhoods of Barcelona are in concordance with the literature: house rental prices spread across space and time. Moreover, this paper shows the importance of analysing and differentiating the effects of price determinants in the short and in the long-run, and to distinguish between the effects of a variable in the same neighbourhood (direct) and the effects of that same variable in nearby neighbourhoods (indirect). The differences in the estimates depending on the weights matrix used show the great impact its specification has and gives support to check multiple specifications in order to assure robustness.

The fact that the coefficients on the space and time lags are statistically significant could suggest some sort of market inefficiency, as shown in other studies such as Tirtiroğlu (1992) or Case and Shiller (2003). These findings of spatio-temporal persistence can help explain housing bubbles formation and be of much use for housing market analysts as well as for policy makers and institutions. However, a deeper analysis has to be conducted in this direction to make stronger claims.

Also, the forecastability of price movements allows us to predict how the prices will move in the future due to external shocks. With the spatial IRF estimates obtained, we can get a

[19]Neighbourhoods with missing data are represented in white.

clearer idea of how housing rental prices spread across space and time.

As previously noted, it was not only a concern the unbiasedness of the estimates, but also its efficiency. It is important to include more time periods as soon as new information is available to check the robustness of the current estimates and the consistency of the results presented here. Because this is a seminal paper on spatio-temporal diffusion of housing rental prices in the neighbourhoods of Barcelona, it remains for a future extension of this paper to check whether this holds.

References

[1] Anselin, L. (1980). Estimation Methods for Spatial Autoregressive Structures. Regional Science Dissertation and Monograph Series 8. Field of Regional Science, Cornell University, Ithaca, N.Y.

[2] Anselin, L. (1998). GIS Research Infrastructure for Spatial Analysis of Real Estate Markets. *Journal of Housing Research*, 9(1), pp. 113-133.

[3] Anselin, L. (1999). Spatial econometrics. An updated review. Regional Economics Applications Laboratory (REAL), University of Illinois, Urbana-Champaign.

[4] Anselin, L. (2001). Spatial econometrics. In Baltagi, Badi, editor, *A Companion to Theoretical Econometrics*, pp. 310-330. Blackwell, Oxford.

[5] Anselin, L. (2010). Thirty years of Spatial Econometrics. *Papers in Regional Science*, 89(1), pp. 3-25.

[6] Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies*, 58(2), pp. 277-297.

[7] Badinger, H., Müller, W. and Tondl, G. (2004). Regional Convergence in the European Union (1985-1999): A Spatial Dynamic Panel Analysis. *Regional Studies*, 38(3), pp. 241-253.

[8] Beenstock, M. and Felsenstein, D. (2007). Spatial Vector Autoregressions. *Spatial Economic Analysis*, 2(2), pp. 167-196.

[9] Bhattacharjee, A. and Jensen-Butler, C. (2006). Estimation of Spatial Weights Matrix in a Spatial Error Model, with an Application to Diffusion in Housing Demand. *CRIEFF Discussion Paper*, number 0519, University of St. Andrews.

[10] Brady, R. (2011). Measuring the diffusion of housing prices across space and over time. *Journal of Applied Econometrics*, 26(2), pp. 213-231.

[11] Brady, R. (2014). The Spatial Diffusion of Regional Housing Prices Across U.S. States. *Regional Science and Urban Economics*, 46, pp. 150-166.

[12] Case, A.C., Rosen, H.S. and Hines, J.R. (1993). Budget spillovers and fiscal policy interdependence. Evidence from the states. *Journal of Public Economics*, 52(3), pp. 285-307.

[13] Case, K. and Shiller, R. (2003). Is There a Bubble in the Housing Market?. *Brookings Papers on Economic Activity*, 2003(2), pp. 299-342.

[14] Doreian, P. (1980). Linear Models with Spatially Distributed Data: Spatial Disturbances or Spatial Effects?. *Sociological Methods and Research*, 9(1), pp. 29-60.

[15] Elhorst, J.P. (2003). Specification and estimation of spatial panel data models. *International Regional Science Review*, 26(3), pp. 244-268.

[16] Elhorst, J.P. (2005). Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis*, 37(1), pp. 85-106.

[17] Elhorst, J.P. (2010). Spatial Panel Data Models. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications, pp. 377-407. Springer.

[18] Elhorst, J.P. (2011). Spatial Panel Models. Handbook of Regional Science. Springer.

[19] Griffith, D.A. (2003). Spatial Autocorrelation and Spatial Filtering. Gaining Understanding Through Theory and Scientific Visualization. Springer, Berlin Heidelberg.

[20] Kelejian, H.H. and Prucha, I.R. (1998). A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 17(1), pp. 99-121.

[21] Kelejian, H.H. and Prucha, I.R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, 40(2), pp. 509-533.

[22] Kelejian, H.H. and Prucha, I.R. (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1), pp. 53-67.

[23] Lee, L. and Yu, J. (2009). Some recent developments in spatial panel data models. *Regional Science and Urban Economics*, 40(5), pp. 255-271.

[24] Lee, L. and Yu, J. (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154(2), pp. 165-185.

[25] Lee, L. and Yu, J. (2016). Identification of Spatial Durbin Panel Models. *Journal of Applied Econometrics*, 31(1), pp. 133-162.

[26] LeSage, J.P. (1997). Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review*, 20(1,2), pp. 113-129.

[27] LeSage, J.P. and Pace, R.K. (2007). A Matrix Exponential Spatial Specification. *Journal of Econometrics*, 140(1), pp. 190-214.

[28] LeSage, J.P. and Pace, R.K. (2009). Introduction to Spatial Econometrics. CRC Press, Taylor & Francis Group.

[29] Manski, C.F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3), pp. 531-542.

[30] Marsh, T.L. and Mittelhammer, R. (2004). Generalized maximum entropy estimation of a first order spatial autoregressive model. In *Spatial and Spatiotemporal Econometrics*, pp. 199-234. Emerald Group Publishing Limited.

[31] McGreal, S. and Taltavull, P. (2013). Implicit House Prices: Variation over Time and Space in Spain. *Urban Studies*, 50(10), pp. 2024-2043.

[32] Moran, P.A.P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society*, 10(2), pp. 243-251.

[33] Ord, K. (1975). Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70(349), pp. 120-126.

[34] Pinkse, J. and Slade, M. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, 85(1), pp. 125-154.

[35] Tiefelsdorf, M., Griffith, D.A. and Boots, B. (1999). A Variance-Stabilizing Coding Scheme for Spatial Link Matrices. *Environment and Planning A*, 31(1), pp. 165-180.

[36] Tirtiroğlu, D. (1992). Efficiency in housing markets: Temporal and spatial dimensions. *Journal of Housing Economics*, 2(3), pp. 276-292.

Appendices

Appendix 1

Moran's I is a global statistic used to examine spatial autocorrelation. It was first introduced in Moran (1948) and it is computed as

$$I = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2 \sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij}}$$

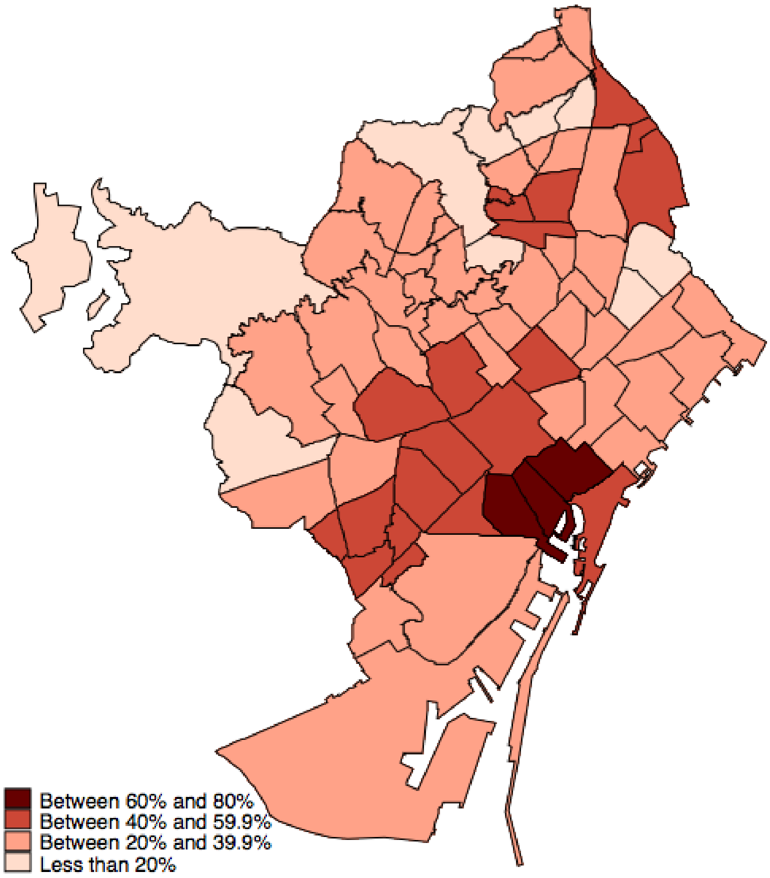
where y_i is the value taken by Y in region r_i, y_j is the value taken by Y in region r_j, and $\bar{y}$ is the average of Y. Under the null hypothesis of no global spatial autocorrelation, the expected value of I is $\mathbb{E}(I) = -\frac{1}{N-1}$. When $I > \mathbb{E}(I)$, this indicates a positive spatial autocorrelation, i.e. the values of Y at nearby regions are similar. When $I < \mathbb{E}(I)$, it is the other way around. I will just provide the results when using the contiguity matrix as an illustration.

	Variable: log rent		
Time	I	$\mathbb{E}(I)$	z
2014q1	0.431	-0.015	5.926
2014q2	0.322	-0.015	4.674
2014q3	0.410	-0.015	5.875
2014q4	0.407	-0.015	5.874
2015q1	0.360	-0.015	5.210
2015q2	0.440	-0.015	6.276
2015q3	0.425	-0.015	6.094
2015q4	0.392	-0.015	5.598
2016q1	0.419	-0.015	5.977
2016q2	0.477	-0.015	6.764
2016q3	0.445	-0.015	6.364
2016q4	0.450	-0.015	6.393
2017q1	0.371	-0.015	5.299
2017q2	0.469	-0.015	6.682
2017q3	0.457	-0.015	6.525
2017q4	0.414	-0.015	5.931

Moran's I statistic for the variable log rent

We can clearly reject the null hypothesis of no spatial autocorrelation for all time periods.

Appendix 2



Percentage of houses being rented in each neighbourhood in Barcelona. Source: Armengou, P.J. (2018, April 14). *Les llars que viuen de lloguer a Barcelona ja són prop del 40%.* Retrieved April 14, 2018, from `https://www.ara.cat`.