Journal of Transport Geography 24 (2012) 370–382

Contents lists available at SciVerse ScienceDirect

Journal of Transport Geography

journal homepage: www.elsevier.com/locate/jtrangeo

Modelling transport and real-estate values interactions in urban systems

Ángel Ibeas [a], Ruben Cordera [a], Luigi dell'Olio [a,*], Pierluigi Coppola [b], Alberto Dominguez [a]

[a] Department of Transportation, University of Cantabria, Av. de Los Castros s/n, 39005 Santander, Cantabria, Spain
[b] Department of Civil Engineering, "Tor Vergata" University of Rome, Via del Politecnico 1, 00133 Roma, Italy

ARTICLE INFO

Keywords:
Hedonic pricing models
Spatial econometric models
Transport and land use interaction
Accessibility

ABSTRACT

This article presents hedonic Multiple Linear Regression models (MLR), spatial autoregressive hedonic models (SAR), Spatial autoregressive hedonic in the Error term Models (SEMs) and spatial Durbin hedonic models (SDMs) to estimate house price variations in metropolitan areas as a result of changing environmental and accessibility conditions. The goodness of fit of the different models has been compared along with a series of hypotheses about the performance of the specifications considering spatial relationships between observations. The case study for such analysis was the urban area of Santander (Spain). The models which considered spatial dependence between observations offered a greater degree of fit in a scenario showing strong spatial correlation in MLR residuals. The SEM model combined with a Queen-Contiguity matrix provided a good fit to the data and at the same time presented significant parameters with theoretically coherent signs. This model estimated increases of 1.8% for each additional transit line present in the areas of housing, as well as a reduction of 1.1% in their prices for each additional minute in travelling time to the Central Business District. Closeness to the train stations, however, implied reductions in house prices.

1. Introduction and objectives

Classic urban economic theory proposed by Alonso (1964), Muth (1969) and others is based on the trade-off between accessibility and space. Locations with better access to the Central Business District (CBD) have higher land values per unit area, because certain agents are more willing to pay higher prices for them. This fact implies that investment in transport can improve accessibility to certain locations and have repercussions on property values.

Hedonic studies stand out in the empirical literature which tries to verify hypotheses on urban economic theory. This technique has become the standard econometric tool for estimating the determinant factors on the prices of heterogeneous goods such as property values (Malpezzi, 2008). The development of hedonic studies had its roots in both early empirical studies (Court, 1939) and the reformulation of consumer theory carried out by Lancaster (1966). However, it was Rosen (1974) who finally formalised the theory of how markets worked for heterogeneous goods. According to this theory, real estate can be seen as goods priced as a function of the group of their characteristics. These characteristics may not only refer to the structural aspects of the properties but also to the characteristics of the surrounding area and their access to different land uses.

Hedonic models can help in estimating the increase in real estate prices derived from environmental and other local improvements, making them a potentially useful tool to support investment in transport projects through value capture policies. Furthermore, when integrated in Land Use/Transport Interaction models (LUTI models) hedonic models can help simulate the complex interactions caused in an urban system where location choices for housing or companies depend very strongly on the real estate market (Löchl and Axhausen, 2010; Waddell et al., 2007).

This article presents four hedonic regression estimators to verify, in accordance with the accepted economic theory, the hypothesis that the dwellings with better accessibility do capitalise, to a certain extent, these benefits; in other words, to verify to what extent a relationship between the accessibility conditions and the dwelling market values does exist. Accessibility is here measured by three types of indicators: an accumulated opportunities indicator, gravity-based indicators and the journey time to CBD taking into account congestion.

The application of hedonic regression techniques in research carried out within the real estate market has had various methodological problems which will be tackled throughout this study. One of the more basic problems is the existence of strong spatial relationships between observations. These relationships can violate the basic hypothesis of multiple linear regression model residual independence (LeSage and Pace, 2010). Anselin (1988) differentiates two basic types of spatial relationships. Spatial dependence or autocorrelation which is defined as the existence of a functional

* Corresponding author. Tel.: +34 942202262; fax: +34 942201703.
E-mail address: delloliol@unican.es (L. dell'Olio).

0966-6923/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jtrangeo.2012.04.012

relationship between what occurs at a point in space and what occurs at nearby or neighbouring points, and spatial heterogeneity (or spatial non-stationarity), that is the lack of structural stability in the parameters or of spatial errors in a model. In the context of real estate markets both effects could be present due to various factors: lack of equilibrium between housing supply and demand in different sectors of an urban area (Bitter et al., 2007), diffusion effects of market prices for housing in nearby areas or, simply, the omission of relevant variables that were not included in the model because of the lack of or poor quality available data. Therefore, it would be necessary to use spatial econometric models in order to avoid biased or inefficient parameters in case studies in which these effects play a significant role (LeSage and Pace, 2009).

Throughout this study, after Section 2 presents the state of the art of the models developed to estimate the impacts of changing accessibility on real estate, hedonic regression models will be estimated with and without considering the existence of spatial dependence between observations. In Section 3 a Multiple Linear Regression (MLR) model is estimated using the traditional hedonic regression technique applied to a dataset from the urban area of Santander. The residuals of this model show a strong degree of spatial autocorrelation between observations. To overcome such autocorrelation, spatial autoregressive models (SAR), spatial autoregressive model in the error term (SEM) and spatial Durbin models (SDMs) have been estimated along with the following questions of methodological interest:

1. do the models considering spatial dependency between observations have a significantly better fit to the data?
2. which spatial regression model combines a better fit with higher parsimony in the hedonic function?
3. which type of spatial relationship between observations provides a better fit?

The specification of the SAR, SEM and SDM models are presented in Section 4 and the model estimates are analysed to answer the questions posed. Finally, in Section 5, some conclusions are drawn on the opportunity to take into account the existence of spatial effects when modelling the real estate market as well as on the opportunity to use a given modelling specification to capitalise the accessibility conditions onto the housing market prices.

2. Bibliographic review

Research about how transport conditions influence real estate prices can be classified into two main streams of thought. On the one hand there are the more theoretical studies initiated by von Thünen (1826) in his work on agricultural land rents. This pioneering work later served as a basis for creating a theory about the distribution of land use and rents in urban areas proposed largely by Alonso (1964), Muth (1969) and Mills (1972). The nucleus of the theory lies in the modelling of certain trade-offs in the choice of location, mainly between the transport costs of getting to the CBD and the cost of the space, which can be modelled using bid-rent functions. This tradition, which makes up the theoretical nucleus of the urban economy, has continued to grow through the use of ever more complex models. An excellent systematic review of the various research work carried out can be found in Fujita (1989).

On the other hand, the second main body of research, based around the relationship between transport and real estate values, is of a more empirical nature and has provided a growing number of case studies. These have been generally supported by the well-known hedonic regression technique formulised by Rosen (1974) to describe how markets function with heterogeneous goods. The hedonic studies relating real estate prices with transport conditions have therefore complemented the theories on urban economy and tested their hypotheses through multiple case studies. Most of these studies have concentrated on the relationship between real estate prices and access to rail transport with very varied results (Pagliara and Papa, 2011). Debrezion et al. (2007) carried out a meta-analysis with more than 50 hedonic studies to explain the variability in the results of the research. The authors controlled the influence of variables like the type of property being modelled, the type of station or the functional form chosen for the hedonic model. The results detected a significant influence of variables such as the type of property or station being studied on the variability of the relationship between accessibility to railway stations and real estate prices. Commercial properties generally showed higher price rises than residential properties, while suburban train stations also had a greater influence on local real estate prices than light rail or metro stations did. Nevertheless, the authors found overvaluations of positive impacts if the specification of the models omitted variables which considered the influence of other modes of transport on accessibility.

Senior (2009) found that Metrolink had no effect on house prices in Greater Manchester (Forrest et al., 1996), however, Ovenell (2007) was later able to identify a positive effect on the prices of properties located between 0.5 and 1 km from the Metrolink stations. Andersson et al. (2010) studied the effects of accessibility to high speed rail in Taiwan and discovered that overall it had a very minor effect on property prices. This clearly contrasts with the results of Debrezion et al. (2006) in Holland, where the effects of proximity to a railways station were more than double those found in Taiwan. Banister and Thurstain-Goodwin (2011) found that, at the microlevel, non-transport benefits provided by investment in railways can be seen in the land and property markets. These effects are a reduction in land prices immediately around the railways stations due to increased noise levels and greater crime rates (Bowes and Ihlanfeldt, 2000), and these effects can radiate out up to a range of 1 km (Ovenell, 2007). It is important to point out that, although small investment may have local effects, it is the large investments that have really significant effects on the housing market (Colin Buchanan and Partners, 2003).

Another, less numerous, series of studies has concentrated on the impact caused by Bus Rapid Transit (BRT) systems on real estate prices. Rodríguez and Mojica (2009) studied the impact on property values caused by introducing a BRT system in the city of Bogotá in Bolivia and found price increases of between 13% and 14%. The influence of the same BRT system was again examined by Munoz-Raskin (2010) who found that the properties nearest the bus stops had a value 4.5% lower than the rest of the properties in Bogotá. However, they also found that properties located less than 5 min walk away from the stops were valued 8.7% higher than those located between 5 and 10 min walk away concluding that households were prepared to pay more to be located close to the BRT system. Cervero and Kang (2011) used multilevel hedonic regression to estimate the capitalisation of introducing a new BRT system in Seoul, South Korea and found increases in property values of up to 10% for residents less than 300 m from a stop on the network.

Nevertheless, in spite of the usefulness of hedonic studies they are not exempt from technical problems. Armstrong and Rodríguez (2006) point out three of them: the problem of omitting variables, the problem of choosing the functional form and the problem of spatial autocorrelation in sample observations. The omission of theoretically relevant variables may, as is well known, bias the estimated parameters (Gujarati and Porter, 2009) and the problems detected by Debrezion et al. (2007) in a great many models of omitting the accessibility provided by other modes of transport is a clear example of this. The problem of specifying the functional form is common to all hedonic studies. There is currently no theoretical basis which recommends using one particular functional specification rather than another, even if Cropper et al. (1988) showed that the linear form produced lower errors in cases where the model presented omitted variables. Malpezzi (2008) recommends the use of the log-linear form because it allows estimated parameters to be interpreted as semi-elasticities and has the capacity to reduce the problems derived from heterocedasticity. Finally, the problem of spatial autocorrelation in the sample may lead to the estimation of parameters which are inefficient or even biased, requiring the use of spatial econometric models (LeSage and Pace, 2009).

Within the range of available models in the field of spatial econometrics, the SAR and SEM models have enjoyed the greatest number of practical applications. These types of models were systematised for the first time in the early contribution made by Paelinck and Klaassen (1979) and received later contributions by various authors (Anselin, 2010). These types of models have received increasing attention in the field of hedonic studies applied to the relationship between transport and real estate prices. Armstrong and Rodríguez (2006) used a spatial autoregressive model to examine rising real estate values following the opening of a suburban railway in Eastern, MA, USA. The estimated model could capture the existence of spatial dependence between observations and price rises of up to 10% in the properties close to the stations. The properties located close to the lines also showed significant, but negative, changes in value. Martínez and Viegas (2009) examined the relationship between the availability of transport infrastructure and property values in Lisbon, Portugal in order to establish value capture schemes to introduce new public transport services. The authors used a MLR model and a SAR model to demonstrate the existence of spatial autocorrelation between observations. However, the MLR model showed similar parameters to those estimated by the SAR model and a lower Akaike information criteria (AIC) leading the authors to conclude that the MLR model was preferable because it offered sufficiently well-fitting predictions with greater parsimony. Löchl and Axhausen (2010) compared hedonic type MLR, SEM, SDM and based on geographically weighted regression (GWR) models. This comparison was made to establish the best specification of a hedonic regression to be introduced into a land use-transport interaction model (UrbanSim). The authors chose the SEM model as the most appropriate because the SDM model showed a large number of variables not significant and because the GWR showed a strong correlation between the estimated parameters, a phenomenon also found in other research (Ibeas et al., 2011; Wheeler and Tiefelsdorf, 2005).

3. Multiple Linear Regression (MLR) models

The data set used to estimate the hedonic models will be presented in this section. A second step will introduce the various specifications proposed for the MLR models along with a discussion about the parameters obtained. Finally, the presence of spatial autocorrelation in the residuals of the models will be evaluated, something that would violate one of the fundamental assumptions of the MLR models.

3.1. The data set

The hedonic regression models used in the present application will be estimated with data from the metropolitan area of Santander. Santander is a medium sized city, capital of the region of Cantabria in the North of Spain. The city currently has 182,700 inhabitants in its urban nucleus but the population rises to around 280,000 if the surrounding metropolitan area is taken into account. Apart from the capital, other important urban centres within the metropolitan area are Astillero (10,020 pop), Muriedas (11,279 pop) and Maliaño (5272 pop). The city of Santander is located to the North of the study area (see Figs. 1 and 2) and is connected to the other urban centres by transport networks and services. These networks are mainly made up of urban and interurban road systems, the urban and interurban public transport services and the interurban railway network connecting the most important nuclei in the study area.

The household sample comes from a cross sectional data base obtained from various on-line real estate platforms. The data was collected in June 2009 and contains information about asking prices and other structural characteristics for 1562 properties located in the metropolitan area. The availability of the address of each of the sample observations meant that they could be coded with a geographical information system (GIS).

The spatial distribution of the aggregated asking prices over large administrative areas shows how the highest average prices (around 500,000 Euros) are concentrated in the city of Santander and more specifically in the residential area El Sardinero, located to the east of the city. This neighbourhood is characterised by a range of environmental attractions such as its status as a garden city, the prestige associated with an address there, its closeness to beaches and parks. Other areas with high average prices are located along the central axis of the city of Santander (central zone) as well as in more recently developed neighbourhoods close to the el Sardinero area, which largely share the same environmental attractions (e.g. the Valdenoja neighbourhood). Two high price areas were detected away from the urban nucleus. The first of these was found to the west of Santander with a suburban style of development made up mainly of individual family houses and also close to the coast and beach areas (San Román, Liencres and Bezana). The second area is found to the east of Santander where several coastal settlements are made up mainly of second homes (Somo and Galizano, among others). The areas with the lowest average asking prices are located in a range of residential neighbourhoods found around the urban centre where the households with the lowest incomes live, as well as along the south eastern link around the Bay of Santander, where the area has been strongly influenced by negative environmental spillovers from industrial development and port activity. An overall north–south spatial gradient can be detected in the housing prices which is a function of proximity to the coast and the beaches. However, the existence of a pricing pattern depending on distance to the town centre of Santander is not so easy to identify using purely cartographic representation.

Putting all the data into the GIS meant a diverse range of variables relating to the environment and the location of the properties could be obtained when the data was crossed with the exact location of each household with various other socio-demographic characteristics present in the population and household census zoning data. The following variables are contained in the data base (see Table 1):

- LN(P) is the natural logarithm of the property asking price.
- IMPROV is a dummy variable taking a value of 1 if the property requires major improvement.
- DETAC is a dummy variable taking a value of 1 if the property is detached.
- ROOMS is the number of bedrooms at the property.
- BATH is the number of bathrooms at the property.
- FLOOR is the floor where the property is located in the building.
- LIFT is a dummy variable taking a value of 1 if the building where the property is located has a lift (elevator).
- TER is a dummy variable taking a value of 1 if the property has a terrace.

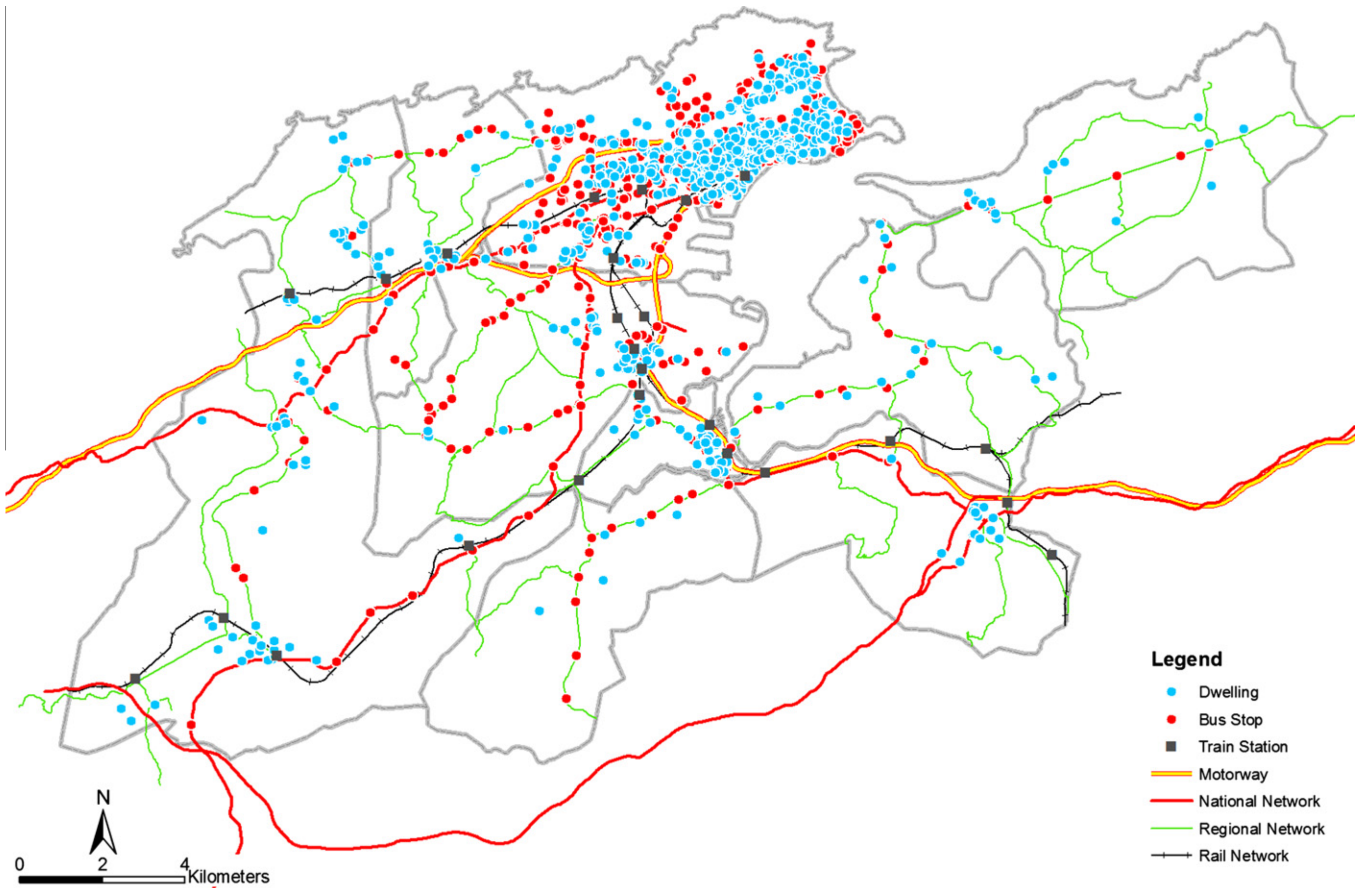

Fig. 1. Location of sampled dwellings, bus stops and transport networks in the study area.

- GAR is a dummy variable taking a value of 1 if the property has a garage.
- SQM is the surface area of the property in square metres.
- LINES is a dummy variable taking a value of 1 if the property has a bus stop less than 400 m away interacting with the number of lines servicing that bus stop.
- CBD is the time in minutes which it takes at morning rush hour to reach the city's CBD from the property using the road network, considering congestion.
- TRAIN is a dummy variable taking a value of 1 if the property is less than 500 m from a suburban train station.
- ACC is a Hansen type measure (gravitational) of employment accessibility.
- CEN is a dummy variable taking a value of 1 if the property is located in the city centre.
- BCH is a dummy variable taking a value of 1 if the property is located at a beach.
- DEN is a measure of the zone's population density. Calculated as inhabitants per area.
- JOBS is the number of employments present in the area where the property is located.
- EXT is the proportion of the population from overseas living in the area where the property is located.

There are some problems associated with the characteristics of the data source used. The main restriction is that the property prices are not market values, they are asking prices. Nevertheless, previous research has shown that asking prices have a high correlation to selling prices and generally represent 90% of the equilibrium price (Hometrack, 2005). Therefore, changes in the dependent variable are not expected to cause significantly different parameter estimations. One of the aspects that could condition the use of this is the fact that it corresponds to a period which included the start of the housing crisis in Spain, narrowly connected to the international financial crisis. However, the price variations have been limited, especially in the urban centre of Santander, with a fall of between 10% and 15% compared with the current situation. Furthermore, changes in real estate prices are not thought to have altered the trade-offs between the characteristics of the properties being considered, meaning that the estimated parameters should not change. Additionally, the number of variables in the characteristics of the diverse properties included in the data base is limited. Unfortunately, there is no official data base currently available for public viewing in Spain showing the characteristics and final selling prices of real estate.

The dependent variable, the property asking price, has been specified in logarithmic form following the recommendations of Malpezzi (2008), meaning that the estimated parameters can be interpreted as semi-elasticities. The most relevant variables, in accordance with the aim of this study, are those which refer to transport conditions: LINES, CBD, TRAIN and ACC. The LINES variable, as described earlier, represents the interaction between the presence of a bus stop at least 400 m away and the number of lines which service that bus stop. This variable is therefore an indicator of the supply of bus services available to each of the properties. The possibility of measuring access to bus services by only using the 400 m zone was initially tested but it was discarded because of the lack of variability in the resulting dummy variable as 85% of properties had access to urban or interurban bus stops. The CBD variable represents access time in minutes to the urban centre of

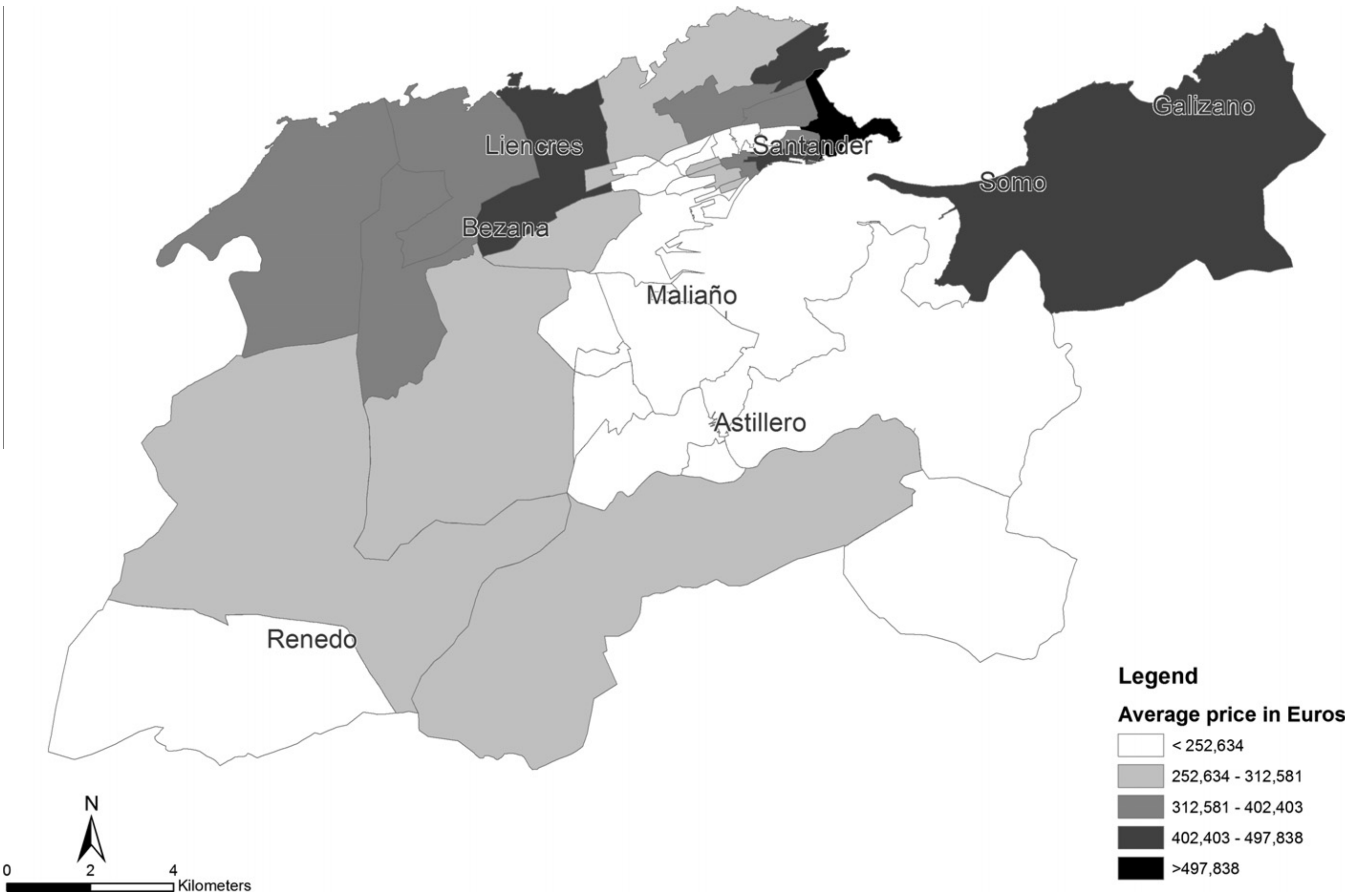

Fig. 2. Spatial distribution of average asking prices aggregated by administrative zones in the study area.

Table 1
Descriptive statistics of the variables contained in the residential property data base (N = 1562).

Variable	Minimum	Maximum	Mean	Std. deviation	Measurement unit
LN(P)	11	14.91	12.49	.55	Ln (Price €)
IMPROV	0	1	.07	.26	–
DETAC	0	1	.23	.41	–
ROOMS	0	12	2.97	1.15	No. of bedrooms
BATH	0	4	1.86	.83	No. of bathrooms
FLOOR	0	12	2.38	1.99	Floor number
LIFT	0	1	.52	.5	–
TER	0	1	.26	.43	–
GAR	0	1	.56	.49	–
SQM	20	850	124.41	79.73	m^2
LINES	0	15	4.26	4.62	No. of lines
CBD	.1	28.73	8.22	6.42	Minutes
TRAIN	0	1	.26	.43	–
ACC	5.01	48.05	18.56	12.12	–
CEN	0	1	.07	.26	–
BCH	0	1	.17	.37	–
DEN	.006	9.41	1.31	1.94	Pop/Area
JOBS	.006	4.60	.54	.63	No. Jobs in 1000s
EXT	.008	.322	.08	.05	Proportion of foreigners

Santander calculated with a transport model which uses real morning rush hour traffic flows. The times input into the data base assume the existence of congestion and represent a road accessibility indicator in accordance with the theoretical mono-centric model proposed by Alonso (1964). The TRAIN variable represents an accumulated opportunities measure of accessibility to metropolitan railway services (Handy and Niemeier, 1997). Therefore, it has been assumed that all the properties located at less than 500 m from one of the 25 railway stations in the study area (see Fig. 1) have access to the train mode. Finally, ACC is a Hansen (1959) type indicator of gravitational accessibility to employment which therefore, considers the possible multi-centric nature of the urban area. This indicator has been calculated from the employment figures available in census data. The expression used was as follows (Coppola and Nuzzolo, 2011):

$$Acc(o) = \sum_i [\exp(\alpha_2 \cdot Cost(o, d_i)) \cdot jobs(d_i)^{\alpha_1}] \qquad (1)$$

Table 2
Estimated parameters of the MLR models.

Variables	MLR1	MLR2	MLR3	MLR4	MLR5
(Constant)	11.367(.000)	11.393(.000)	11.227(.000)	11.402(.000)	11.240(.000)
IMPROV	−.085(.002)	−.080(.004)	−.087(.002)	−.079(.004)	−.086(.002)
DETAC	.038(.167)				
ROOMS	.069(.000)	.068(.000)	.068(.000)	.058(.000)	.058(.000)
BATH	.131(.000)	.134(.000)	.136(.000)	.123(.000)	.125(.000)
FLOOR	.011(.009)	.010(.012)	.012(.003)	.009(.019)	.011(.005)
LIFT	.244(.000)	.232(.000)	.253(.000)	.235(.000)	.256(.000)
TER	.046(.004)	.045(.005)	.039(.016)	.042(.007)	.036(.021)
GAR	.107(.000)	.112(.000)	.108(.000)	.109(.000)	.105(.000)
SQM	.003(.000)	.004(.000)	.004(.000)	.004(.000)	.004(.000)
LINES	.020(.000)	.021(.000)	.021(.000)	.021(.000)	.022(.000)
CBD	−.010(.000)	−.011(.000)		−.011(.000)	
TRAIN	−.060(.001)	−.062(.000)	−.076(.000)	−.065(.000)	−.075(.000)
ACC	.002(.269)		.005(.000)		.005(.000)
CEN	.123(.001)	.145(.000)	.088(.009)	.140(.000)	.084(.011)
BCH	.338(.000)	.336(.000)	.315(.000)	.326(.000)	.305(.000)
DEN	−.008(.130)				
JOBS	.002(.903)				
EXT	−.570(.001)	−.638(.000)	−.875(.000)	−.631(.000)	−.863(.000)
R^2	.767	.767	.763	.773	.770
R^2_{adj}	.765	.764	.761	.771	.768
F	282.67	362.92	355.54	375.13	367.30
p-Value F	.000	.000	.000	.000	.000
p-Value Moran's I	.000	.000	.000	.000	.000
AIC	334.75	331.56	356.10	252.48	277.75
Log-likelihood	−148.37	−150.78	−163.51	−111.24	−123.87
N	1562	1562	1562	1554	1554

where *Cost* is a measure of the journey cost between origin *o* and destination d_i calculated using a transport model considering congestion at morning rush hour. Jobs (d_i) are the current employments in the destination zone d_i and α_1, and α_2 are the parameters to be estimated. The parameters can be estimated by ordinary least squares, linearising (1) using logarithms at both sides of the expression. The transport flows generated by each zone have been selected as a proxy of the accessibility conditions (dependent variable). The parameter α_1 presented an estimated value of 0.26 while α_2 had a value of −0.12, showing a R^2 goodness of fit of 0.7.

3.2. MLR estimates

The first hedonic model (MLR1) was estimated with the 18 independent variables contained in the data base (see Table 2 showing the *p*-values of the parameters in brackets). All the parameters had theoretically correct signs, although four of them: DETAC, ACC, DEN and JOBS were not significant according to the *t*-test. The parameter of the DETAC variable indicated that being a detached property raises average house prices, although this effect was not high enough to be statistically significant, so it was eliminated in subsequent models. Furthermore, two of the variables which measure the influence of property accessibility, CBD and ACC had a high degree of colinearity with a correlation coefficient of −0.78 and a VIF value of 3.6 and 6.7 respectively. As the DETAC, DEN and JOBS variables were not significant they were removed from the MLR2 and MLR3 models and only one indicator of road accessibility was kept in each of them, CBD in MLR2 and ACC in MLR3. In both models, CBD and ACC were significant at a 99% confidence level. They also had the correct signs, although the MLR2 model had a slightly higher goodness of fit taking into account the R^2_{adj} as the Akaike information criteria (AIC). Both models showed moderate colinearity between independent variables and in no cases were Variance Inflation Factor (VIF) values over 3 nor condition indexes over 20 found. Finally, for the MLR4 and MLR5 models, respectively similar in their specification to MLR2 and MLR3, eight observations were removed as outliers because they showed studentized residuals[1] higher than three typical deviations. In five cases, this was due to negative residuals while in the remaining three cases the residuals were positive. Among the negative-residual observations, four of them concerned detached properties with residential surfaces and number of rooms and bathrooms high above the average. This causes the model to overestimate asking prices (although they were in themselves already high). The remaining observation with a negative residual corresponds to an apartment with an asking price of 75,000 Euros which is much lower than the average even when taking into account its characteristics; this appears to be due to an error in the digitisation of the data base. Finally, the three observations with positive residuals have very high prices (in all cases above 500,000 Euros) which the model underestimates. These high asking prices are because the properties are located in tourist zones, a factor which is partly captured by the BCH variable; however, in these three cases it has a much higher weight than normal.

The removal of the outliers slightly improved the fit of both models. The MLR4 and MLR5 models were therefore specified as (2) and (3) respectively:

$$Ln(\hat{P}_i) = \beta_0 + \beta_1 IMPROV_i + \beta_2 ROOMS + \beta_3 BATH_i + \beta_4 FLOOR_i + \beta_5 LIFT + \beta_6 TER_i + \beta_7 GAR + \beta_8 SQM_i + \beta_9 LINES_i + \beta_{10} CDB_i + \beta_{11} TRAIN_i + +\beta_{12} CEN_i + \beta_{13} BCH_i + \beta_{14} EXT_i + \varepsilon_i \quad (2)$$

[1] Given the residuals depend on the unit of measurement of the dependent variable, it would be better to standardise them by dividing them by the standard error of the regressions. This means the residuals can be compared with those of other regressions and facilitate the detection of outliers by distributing the residuals with an average of zero and a variance close to the unit. In this case it was chosen to use the studentized residuals, which are identical to the standardised residuals with the exception that the standard error of the *i*th residual is calculated by eliminating this observation. This guarantees that the variance of the distribution of the residuals is truly unitary (Gujarati and Porter, 2009).

$$Ln(\hat{P}_i) = \beta_0 + \beta_1 IMPROV_i + \beta_2 ROOMS + \beta_3 BATH_i + \beta_4 FLOOR_i + \beta_5 LIFT + \beta_6 TER_i + \beta_7 GAR + \beta_8 SQM_i + \beta_9 LINES_i + \beta_{10} ACC_i + \beta_{11} TRAIN_i + +\beta_{12} CEN_i + \beta_{13} BCH_i + \beta_{14} EXT_i + \varepsilon_i \quad (3)$$

Given the semi-logarithmic specification of the models, the parameters can be directly interpreted as semi-elasticities which allows the dependent variable's relative change caused by an absolute change in the value of an independent variable to be estimated ceteris paribus. This contrasts with log–log regressions where each parameter measures the independent variable elasticity, i.e. the dependent variable's relative change related to a relative change in the value of an independent variable (Gujarati and Porter, 2009). However, in the case of the dummy variables, the parameters cannot be directly interpreted in this way. These can be correctly calculated by applying the following expression (Halvorsen and Palmquist, 1980):

$$[\exp(\beta_n) - 1] * 100 \quad (4)$$

In the MLR4 and MLR5 models all the variables that were introduced were statistically significant and had theoretically believable signs (see Table 2). Among the variables related to the property's structural conditions, IMPROV was the only one that gave a negative sign which is interpreted as a reduction in property value of between 7.5% and 8.2% if the building had to be renovated. Noteworthy from among the variables which had a positive effect on property values, the presence of an additional bathroom, which increased prices by 12%, having a garage which increased prices by around 10% and, most strikingly, if the building was equipped with a lift it implied an increase in value of 20%. The important parameter related to this last variable may not only be due to the availability of a lift but also the age of the building because more modern and better quality construction normally includes a lift.

The parameters related to the environmental conditions of the properties showed positive signs if the buildings were located in the city centre or, even more so, in the beach areas, a variable which could also include the effects of better landscaping. A property's location in a beach area could imply an average increase of 38% in its value according to the parameter of the MLR4 model. The parameter of the EXT variable was negative and high which could also be due to the capture of other environmental effects by this variable such as population density (even though DEN did not turn out significant), worse urban services and a lower presence of public installations.

Finally, most of the parameters related to the transport conditions had signs which agreed with the hypothesis that improved transport conditions resulted in increased real estate values. The exception was the TRAIN variable which had a negative sign implying that properties with a train station less than 500 m away had a value which was between 6% and 7% lower. This result which goes against established theory has also been found in other previous studies (Forrest et al., 1996). But, unlike in the case of Metrolink, where the stations and most of the nearby housing were built in the 19th and early 20th centuries (making them less desirable and with lower prices), the reason is unclear and may be because railway installations and infrastructure carry with them a series of negative spillovers associated with noise and a generally lower landscaping and environmental quality (Armstrong and Rodríguez, 2006). We must also consider the fact that the railway constitutes a minimum percentage of the modal split (around 1%) which helps to explain why the presence of a railway station does not have a positive impact on property prices.

The parameter of the bus accessibility indicator LINES indicates that each additional nearby public transport line can imply an additional increase of up to 2.2% in property values. However, this result should be interpreted with care because the city centre areas of Santander are the ones that are best served by public transport. In this sense, the hypothesis that a better supply of public transport services implies increased property values could be reversed and be due simply to the fact that the more central areas have more public transport services due to the demand for travelling to that area from all the other parts of the city. This phenomenon is very typical of European cities where long standing historic urban centres have always had a high residential and commercial presence and high real estate values (Felsenstein et al., 2010).

The road network accessibility indicators produced believable signs, negative in the case of CBD in the MLR4 model and positive in the case of ACC in the MLR5 model. The CBD parameter indicates that an additional minute in travelling time to the urban centre of Santander could imply slightly over a 1% reduction in property values which confirms the existence of the price gradient assumed by urban economic theory. The parameter of the ACC variable indicates that an increase of one unit in the employment accessibility index implies an increase of 0.5% in real estate values.

For both the MLR4 and the MLR5 models the parameters of the CBD and ACC variables had the correct signs and were significant, meaning that both these indicators of accessibility to employment opportunities and commercial activities can be considered valid. However, a variable like CBD can be adapted to most traditional mono-centric urban areas while the ACC variable can be considered as a more valid indicator for polycentric zones (Ottensmann et al., 2008). In the case of this research, the nuclei of the alternative urban centres to Santander within the metropolitan area do not as yet show enough development in the number of job opportunities available to be able to compete with the city centre of Santander which has more than 20% of the total jobs and around 65% if the city as a whole is considered. It can therefore be stated that the metropolitan area still presents a strong mono-centric character meaning that the CBD variable can be considered as a strong indicator for measuring accessibility. The MLR4 model also had a slightly better fit in both its R^2_{adj} and AIC and will therefore be chosen as the reference model for specifying the models which consider the existence of spatial dependence.

3.3. Autocorrelation analysis

The well-known Moran's I index (Griffith, 1987) was used to test for the presence of residual autocorrelation. Before the index was calculated, the geographical point information on property location was transformed into zonal information using Thiessen polygons (Maguire et al., 2005). The index was initially applied with Queen type spatial contiguity (Griffith, 1987) and later at different fixed distances and clearly significant values appeared in all cases (see Table 1).

A Getis-Ord Gi* statistic was also calculated (see Fig. 3 for the residuals of the MLR4 model). This index is useful for detecting clusters of autocorrelated residuals which may not become evident using only a global autocorrelation index (Ord and Getis, 1995). The Gi index showed significant values at a 95% confidence level for the presence in various zones of both positive and negative residual clusters. The zones showing significant correlation in the positive residuals are located in the northern part of the study area, specifically in the zones close to the Bay of Santander where there is a strong demand for housing due to the attractive characteristics of the area. Of particular relevance was the case of the El Sardinero sector where properties had very high prices which were almost certainly due to the lack of supply and the high demand, apart from the factors mentioned above. The negative residuals were spatially autocorrelated in certain nuclei to the south of the Bay and to the west of the study area, zones which currently have a high number of commuting trips to the city centre.

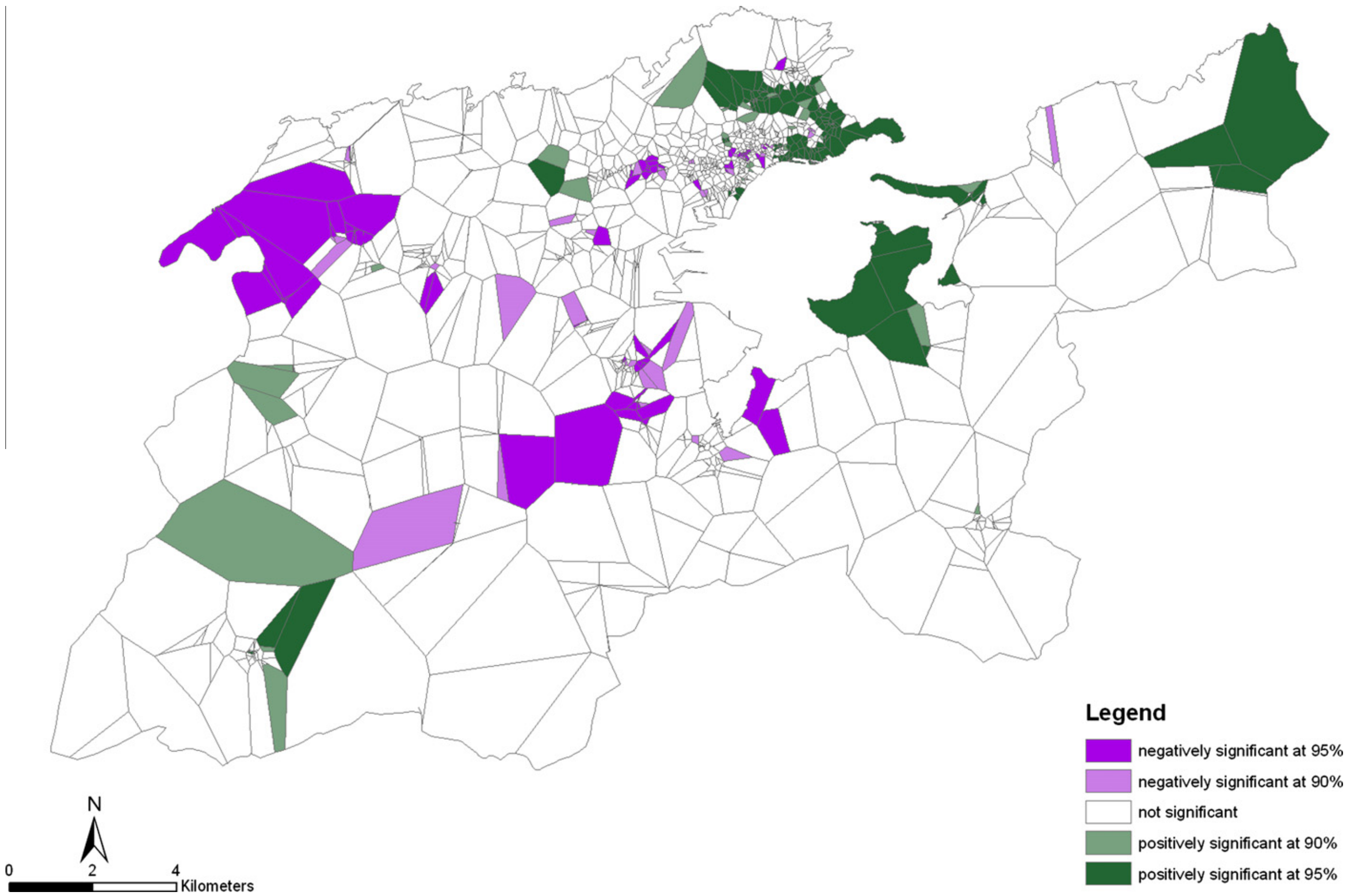

Fig. 3. Significance of the Getis-Ord Gi* statistic values on the residuals of the MLR4 model.

Anselin (1988) recommends using the Lagrange Multiplier test (LM) to detect specification errors due to not considering spatial dependence in MLR models. This test can detect specification errors caused by not including the autoregressive parameter in the dependent variable (LM-Lag) or in the error term (LM-Error). These tests can also provide robust versions if both specification errors are significant. In this case study they prove significant values both in their LM-Lag and LM-Error versions. The robust tests were also significant except in the case of the spatial dependence of the dependent variable using a nearest neighbour distance matrix (see the following section).

4. Spatial econometric models: SAR, SEM and SDM

To capture the effect of the strong spatial autocorrelation present in the residuals of the MLR models, a further series of models were estimated considering the existence of spatial dependence between observations. Both of the functional forms more commonly found in the literature, SAR and SEM, and the spatial Durbin model (SDM) were applied. The models considering spatial relationships were estimated using the specification of the MLR4 model as this gave a better fit and the measure of road accessibility from the journey time to the CBD was considered to be a correct hypothesis.

4.1. SAR, SEM and SDM specifications

LeSage and Pace (2009) provided an extensive introduction to the spatial econometric models developed in the literature. The most well-known spatial model is the simultaneous autoregressive (SAR) model which assumes the existence of a diffusion process in the dependent variable and can be specified as follows:

$$y = \rho W y + X\beta + \varepsilon \tag{5}$$

where y is a vector of observations for the dependent variable, ρ is the parameter of spatial autocorrelation, W is a matrix $N \times N$ of spatial weightings where N *is* the number of observations, β is a vector of estimated parameters, X is a matrix with observations for the independent variables and ε is a vector of independent and identically distributed (IID) error terms.

If we only wish to specify the existence of spatial dependence in the error term of the observations, then an autoregressive simultaneous spatial error model (SEM) can be used. This type of model can be written as:

$$y = X\beta + u \tag{6}$$

$$u = \lambda W u + \varepsilon \tag{7}$$

where W is a $N \times N$ matrix of spatial weightings, λ is a spatial autocorrelation parameter of errors u and ε is a vector of independent and identically distributed errors. So in this model, the dependent variable of any location is a function not only of the independent variables but also of the errors u of the neighbouring locations.

Finally, although applied in fewer research works, a third type of model found in the literature is the spatial Durbin model (SDM):

$$y = \rho W y + X\beta + \gamma W X + \varepsilon \tag{8}$$

where W is a $N \times N$ matrix of spatial weightings, ρ and γ are spatial autocorrelation parameters of the dependent and independent variables, respectively and ε is a vector of independent and identically distributed errors. The application of this model is especially recommended by LeSage and Pace (2009) because of two of its properties. In the first place its greater robustness against the omission of relevant independent variables. Secondly, given its more general model characteristics than those of SAR and SEM which may be considered as nested versions of SDM. So the addition of the constraint $\gamma = 0$ to the SDM model leads to the SAR model, while the constraint $\gamma = -\rho\beta$ leads to a SEM model and, finally, the constraints $\rho = 0$ and $\gamma = 0$ produce a standard MLR model. Furthermore, the SDM model is the only one that produces unbiased parameters even when the true data generation processes are provided by SAR, SEM, SDM models or even by a general spatial model which includes dependence in both the dependent variable and the error terms.

The W spatial weighting matrices present in these models define the connectivity between the units of analysis. It is important to correctly specify the matrix elements, w_{ij}, to ensure that spatial econometric models are correctly applied. The four most common ways of defining connectivity are: Queen type contiguity, Rook type contiguity, fixed number of closest neighbours and neighbours located at a maximum determined distance. Queen type and Rook type are two different contiguity definitions coming from the game of chess. Thus, Queen type contiguity considers all adjacent locations sharing an edge or a vertex with a given location as neighbours, while Rook type contiguity considers neighbours to be only those adjacent locations sharing an edge with a given location (Anselin, 1988).

LeSage and Pace (2009) argued that the estimated parameters should not be excessively variable under changes in the definition of neighbourhood matrices. This aspect will be verified in the following application. Finally, spatially dependent model parameters should be estimated using maximum likelihood because estimation using ordinary least squares could lead both to inconsistent parameters and standard errors.

Although in spatial econometric models the estimated parameters are directly interpreted using SEM type models, the same does not occur in the cases of SAR and SDM models which consider lags in the dependent variable. Simultaneous feedback exists with these specifications because a change in the dependent variable of an observation causes changes in the neighbouring observations which in turn have repercussions on the first observation. Therefore, in the cases of the SAR and SDM specifications, the estimated parameters should be seen as the representation of a state of equilibrium in the modelling process which includes the effects of spatial diffusion (Ward and Gleditsch, 2008). Given that the effects provided by each variable take the form of a matrix in this situation, LeSage and Pace (2009) recommend the use of a series of scaling indicators to correctly interpret the functional relationships:

a. *Average direct effect:* calculated as the mean of the elements of the main diagonal of the parameter matrix. It can be interpreted as the effects caused by the group of observations of an independent variable on the dependent variable.
b. *Average indirect effect:* calculated as the mean of the elements outside the main diagonal of the parameter matrix. It can be interpreted as the diffusion effects between observations caused by changes in an independent variable.
c. *Average total effect:* calculated as the mean of the elements of the parameter matrix of observations. It can be interpreted as the total effect, direct and indirect, received by the dependent variable.

These measures can be calculated separately and require the use of simulation techniques if inferences are to be made of their significance.

4.2. SAR, SEM and SDM estimates

The models were estimated with three types of neighbourhood matrices once again starting from the transformation of the point information to zonal information using Thiessen polygons. An initial matrix assumes first order Queen type spatial relationship between observations, that is considered that there is contiguity between observations only when they have common boundaries. The second of the proposed matrices (referred as "K10", in Tables 3 and 4) considers the 10 closest observations to each property. The third and final matrix (referred as "D1750", in Tables 3 and 4) starts from a distance of 1750 m and considers all the observations found within this radius to be neighbours of the property. This distance was chosen because it was the minimum in which all the observations had at least one neighbouring observation. The three matrices gave a progressively increasing number of average links: 5.7 for the Queen type contiguity matrix, 10 for the matrix with k close neighbours and 219 using the distance matrix. The three matrices progressively consider wider neighbourhood relationships, which should prove useful when checking the hypothesis that the different neighbourhood matrices do not affect the estimated parameters to any large degree.

A total of nine models were estimated considering spatial dependency between observations, combining the different functional specifications SAR, SEM and SDM with three types of neighbourhood matrices. Although the interpretation of the estimated parameters is direct in the case of the SEM type models, the same does not occur with the SAR and SDM models because these consider lags in the dependent variable. In this latter case the scaling measures proposed by LeSage and Pace (2009) should be used to obtain the authentic magnitudes of the direct and indirect effects of the independent variables.

The parameters estimated using the SAR and SEM models in all cases showed identical signs to those obtained using ordinary least squares in the MLR4 model. Only the variables TRAIN and FLOOR, the latter only in the SEM with the D1750 neighbourhood specification, were not significant at a 95% confidence level.

Conversely, in the SDM models there was a greater presence of not significant parameters, especially for the theoretically interesting variables CBD and TRAIN in the SDM-QUEEN and SDM-D1750 models. With the SDM-K10 model all the regressors were significant and showed identical signs to those present in the SAR, SEM and MLR4 models. The spatial lag of the independent variables turned out to be not significant in many cases and in others had signs which were theoretically difficult to interpret such as the positive signs for IMPROV or the negative signs for GAR and SQM using SDM-QUEEN.

The parameters estimated for the different neighbourhood matrices (see Table 3) did not show any great changes with respect to the MRL models and in no case did they show sign changes except in the CBD and TRAIN variables between, on the one hand, the SDM-D1750 model and the SDM-QUEEN and SDM-K10 models. Nevertheless, these latter parameters showed little significance in the SDM models, meaning that their change of sign did not imply that the parameter estimations were clearly different. Moreover, in the three types of models the fit provided by the specifications considering Queen-Contiguity were superior to those provided by the 10 closest neighbours (K10) and maximum distance matrices (D1750). This statement is valid using both the log-likelihood and the AIC indexes.

The Morańs I index was used to test for the presence of residual autocorrelation in the spatial models. Only four models did not show any autocorrelation that was not significant: SEM-QUEEN, SEM-D1750, SDM-QUEEN and SDM-D1750. In all the other cases the spatial residual autocorrelation was significant even though

Table 3
Estimated parameters for the SAR, SEM and SDM models.

Variables	SAR			SEM			SDM		
	QUEEN	K10	D1750	QUEEN	K10	D1750	QUEEN	K10	D1750
(Constant)	8.507	8.520	6.313	11.443	11.430	11.469	6.607	7.643	4.812
IMPROV	−.098	−.095	−.082	−.093	−.090	−.085	−.092	−.096	−.089
ROOMS	.061	.054	.059	.060	.054	.057	.059	.051	.058
BATH	.109	.120	.117	.118	.126	.123	.114	.126	.118
FLOOR	.011	.009	.008	.009	.010	.006*	.010	.009	.006*
LIFT	.212	.228	.242	.196	.225	.232	.193	.221	.229
TER	.035	.038	.034	.034	.040	.036	.033	.037	.032
GAR	.104	.110	.113	.114	.108	.127	.112	.108	.135
SQM	.003	.003	.003	.003	.003	.003	.003	.003	.003
LINES	.017	.020	.022	.018	.020	.015	.004*	.018	.014
CBD	−.009	−.007	−.007	−.011	−.009	−.010	−.008*	−.005	.001*
TRAIN	−.052	−.052	−.017*	−.051	−.047	−.006*	−.000*	−.028	.013*
CEN	.094	.140	.104	.129	.133	.076	.065*	.131	.086
BCH	.223	.245	.205	.295	.264	.227	.084*	.195	.269
EXT	−.288	−.609	−.547	−.505	−.710	−1.00	−.213*	−.683	−1.12
Lag.IMPROV							.118	.144*	.134*
Lag.ROOMS							−.047	.004*	.061*
Lag.BATH							−.031*	−.072*	.102*
Lag.FLOOR							−.006*	−.000*	−.031*
Lag.LIFT							.037*	−.015*	.049*
Lag.TER							.034*	−.020*	−.051*
Lag.GAR							−.077	.054*	−.031*
Lag.SQM							−.001	−.000*	−.001*
Lag.LINES							.009*	−.000*	.015*
Lag.CBD							.005*	−.083*	−.001*
Lag.TRAIN							−.057*	−.083	−.065*
Lag.CEN							.005*	.014*	.890
Lag.BCH							.099*	.090*	−.163
Lag.EXT							−.224*	.935	.963
ρ	.235	.232	.406				.415	.311	.568
λ				.447	.429	.756			
p-Value ρ/λ	.000	.000	.000	.000	.000	.000	.000	.000	.000
p-Value Morańs I	.000	.000	.000	.060	.000	.791	.068	.000	.631
Log-likelihood	−49.66	−63.91	−75.01	−22.51	−74.33	−34.80	7.99	−42.42	−2.06
AIC	133.33	161.82	184.04	79.02	182.67	103.62	46.02	146.85	66.13

* Not significant at 0.05.

Table 4
Average direct, indirect and total impacts estimated for the SAR-QUEEN and SDM-K10 models.

Variables	SAR-QUEEN			SDM-K10		
	Average direct impact	Average indirect impact	Total average impact	Average direct impact	Average indirect impact	Total average impact
IMPROV	−.099	−.029	−.128	−.088	.158*	.070*
ROOMS	.062	.018	.080	.053	.029*	.082*
BATH	.110	.032	.143	.123	−.046*	.077*
FLOOR	.011	.003	.015	.009	.003*	.012*
LIFT	.214	.063	.277	.225	.074*	.300
TER	.036	.010	.046	.036	−.012*	.024*
GAR	.105	.030	.136	.113	.122*	.236
SQM	.003	.001	.004	.003	.000*	.004
LINES	.017	.005	.022	.018	.003*	.022
CBD	−.009	−.002	−.012	−.005	−.001*	−.007*
TRAIN	−.052	−.015	−.068	−.033	−.127	−.161
CEN	.095	.028	.124	.135	.076*	.212
BCH	.225	.066	.292	.204	.210	.414
EXT	−.291	−.085	−.376	−.637	1.00	.365*

* Not significant at 0.05.

the index showed values that were clearly inferior to those presented by the MLR models.

Taking into account these factors, the models considered as being the best for each type were chosen. Three selection criteria were used: Firstly, the goodness of fit of the models with the sample data; secondly, considering the significance and coherence of the parameters present in the models with the initial hypotheses derived from geographic and economic theory, if and when the unexpected signs were not clearly significant; finally, it was considered important that the residuals derived from the models should not show any significant degree of spatial autocorrelation in order to maintain parameter unbiassedness and efficiency.

From among the SAR and SEM models, the best were thought to be those estimated using QUEEN type neighbourhood matrices because of their better fit. However, from among the SDM models, the SDM-K10 model was thought to be better than the SDM-QUEEN

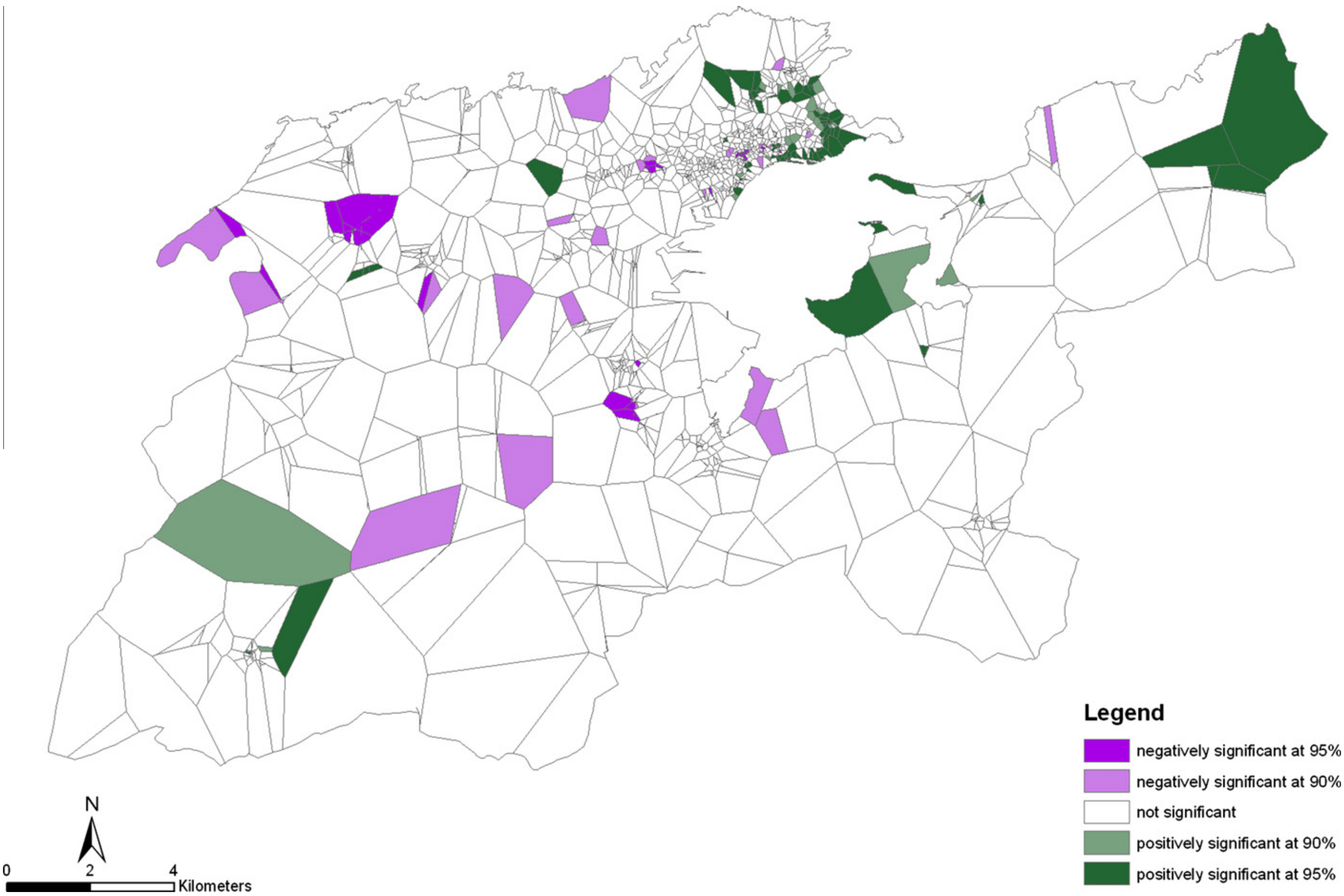

Fig. 4. Significance of the Getis-Ord Gi* statistic values on the residuals of the SEM-QUEEN model.

and SDM-D1750 in spite of its inferior fit, because it was the only model in which the most important variables in this study: LINES, CBD and TRAIN were significant in all cases.

Table 4 shows the average and total direct and indirect impacts of the autoregressive models on the dependent variable. Simulation based on 100 samples was used to estimate the inference measures. The number of samples obtained had to be reduced to 100 because of the amount of calculation involved in simulating models with a great number of parameters. A comparison between Tables 3 and 4 show that the estimated parameters experienced slight variations. Furthermore, in cases like that of the parameter of the BCH variable, situations could occur where the direct or indirect effects of a variable can change its significance. The SAR-QUEEN model also showed indirect impacts in all the significant cases with theoretically believable signs.

Finally, the best model was chosen from the three models selected above. The SEM-QUEEN was chosen because it provided a better fit, theoretically coherent signs (except in the case of the TRAIN variable, mentioned above) and no significant spatial autocorrelation in the residuals at a 95% confidence level. The Getis-Ord Gi* statistic was calculated for the residuals of this model (see Fig. 4). A visual comparison between Figs. 3 and 4 show that the SEM-QUEEN model notably reduced residual autocorrelation in various zones even though it continued to be high in the South East zone of Santander.

5. Conclusions

The work described in this article specified four types of hedonic models: multiple linear regression, spatial autoregressive, spatial autoregressive model in the error term and spatial autoregressive Durbin in order to determine the influence of transport conditions on the prices of a series of real estate properties. These models were estimated using data collected in the metropolitan area of the city of Santander and were compared to determine which of them presented the best fit and how the estimated parameters were affected by setting different neighbourhood relationships between the observations.

The estimated models were useful in calculating how different transport characteristics affected the prices of real estate properties. The LINES variable, as a measure of accessibility to bus transport, had a positive sign in all cases with estimations of its influence on property prices of between 1.4% and 2.2% depending on the model and 1.8% according to SEM-QUEEN. However, this result should be interpreted with care because the causality direction, increase in supply of bus services – higher property prices is arguable for the study area used in this research. The city centre of Santander has a large supply of available public transport which is mainly geared to providing access to the city centre from the surrounding areas. The CBD variable was used as an indicator of journey time by car to the city centre and in all the models the parameter had a negative sign with estimations of property price reductions of between -0.5% and -1.1% per minute of journey time. The specifications that measured accessibility through the CBD variable were chosen rather than through a gravity type indicator like ACC because the models gave a better fit and the study area could be characterised as very mono-centric. The TRAIN variable indicated accessibility using suburban railway transport and presented a negative parameter, which, in agreement with previous research (Armstrong and Rodríguez, 2006), was almost certainly due to noise and other spillovers associated with this type of infrastructure. The impact of the TRAIN parameter on real estate values was −4.9% in the SAR-QUEEN model and oscillated between −2.7% and −6% depending on the specification. These results were not far from those obtained by Forrest et al. (1996), who got an estimation of −4.5% in the average property prices placed less than 1 km from a railway station in Greater Manchester. In addition, the models confirmed the existence of a price gradient as a function of accessibility to the CBD. However, the effect that accessibility to public transport has on property prices was more debatable in spite of the fact that the zones with the most bus lines, especially the centre of Santander, had higher average house prices.

The comparison between the MLR model and the SAR, SEM and SDM models which considered spatial dependency between observations allowed to verify the hypothesis put forward at the start of the study. The spatial econometric models were specified using different neighbourhood matrices, making it possible to check if they had an important influence on the estimated parameters. In general, the estimated parameters were found not to show any important variations or sign changes. Only in the case of the SDM model some parameters, bordering on 0 and not significant, change their sign. These results show that in the case of this research, the different neighbourhood matrices did not have any notable influence on the results obtained.

The SEM-QUEEN specification was chosen as the best model because it gave a better fit and also had parameters that were clearly significant with theoretically consistent signs. The good performance of the SEM models indicates that the MRL models had specification problems by omitting variables related to the local environment resulting in their overvaluation of properties in certain zones and undervaluation in others. The models which considered spatial dependence between observations, particularly the SEM models, helped to reduce this effect with better fits than the MLR models using both the log-likelihood and the AIC. Furthermore, the spatial models reduced the presence of residual spatial autocorrelation according to the Moran's I index and in some cases this did not turn out to be significant. Nevertheless, in spite of their better fit, the SDM models which consider the existence of spatial lags in the independent variables, in some cases had theoretically inconsistent signs and non-significant parameters mainly in the lag variables but also in some key variables like CBD or TRAIN. For this reason they were considered to be less valid for the purposes of this research.

Acknowledgements

This study was supported by research funding form the Spanish Ministry of Science and Innovation through the Project E 21/08-INTERLAND. The authors would also like to thank the City Council of Santander which has allowed this research to take place.

References

Alonso, W., 1964. Location and Land Use: Toward a General Theory of Land Rent. Harvard University Press, Cambridge.
Andersson, D.E., Shyr, O.F., Fu, J., 2010. Does high-speed rail accessibility influence residential property prices? Hedonic estimates from southern Taiwan. Journal of Transport Geography 18, 166–174.
Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht, Boston.
Anselin, L., 2010. Thirty years of spatial econometrics. Papers in Regional Science 89, 3–25.
Armstrong, R., Rodríguez, D., 2006. An evaluation of the accessibility benefits of commuter rail in eastern Massachusetts using spatial hedonic price functions. Transportation 33, 21–43.
Banister, D., Thurstain-Goodwin, M., 2011. Quantification of the non-transport benefits resulting from rail investment. Journal of Transport Geography 19, 212–223.
Bitter, C., Mulligan, G., Dall'erba, S., 2007. Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. Journal of Geographical Systems 9, 7–27.
Bowes, D.R., Ihlanfeldt, K.R., 2000. Identifying the impacts of rail transit stations on residential property values. Journal of Urban Economics 50, 1–25.
Cervero, R., Kang, C.D., 2011. Bus rapid transit impacts on land uses and land values in Seoul, Korea. Transport Policy 18, 102–116.
Colin Buchanan and Partners, 2003. Economic and Regeneration Impact of Tramlink. Report for the South London Partnership. Funded by the London Development Agency. Transport for London and the London Boroughs of Croydon and Sutton, July.
Coppola, P., Nuzzolo, A., 2011. Changing accessibility, dwelling price and the spatial distribution of socio-economic activities. Research in Transportation Economics 31, 63–71.
Court, A.T., 1939. Hedonic Price Indexes with Automobile Examples, the Dynamics of Automobile Demand. General Motors, New York.
Cropper, M.L., Leland, B.D., McConnell, K.E., 1988. On the choice of functional form for hedonic price functions. The Review of Economics and Statistics 70, 668–675.
Debrezion, G., Pels, E., Rietveld, P., 2006. The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market. Tinbergen Institute Discussion Papers, 06-031/3.
Debrezion, G., Pels, E., Rietveld, P., 2007. The impact of railway stations on residential and commercial property value: a meta-analysis. The Journal of Real Estate Finance and Economics 35, 161–180.
Felsenstein, D., Axhausen, K.W., Waddell, P., 2010. Land use–transportation modeling with urbanism: experiences and progress. The Journal of Transport and Land Use 3 (2), 1–3.
Forrest, D., Glen, J., Ward, R., 1996. The impact of a light rail system on the structure of house prices: a hedonic longitudinal study. Journal of Transport Economics and Policy 30, 15–29.
Fujita, M., 1989. Urban Economic Theory: Land Use and City Size. Cambridge University Press, Cambridge, New York.
Griffith, D., 1987. Spatial Autocorrelation: A Primer. Association of American Geographers Resource Publication, Washington, DC.
Gujarati, D.N., Porter, D.C., 2009. Basic Econometrics. McGraw-Hill Irwin, Boston.
Halvorsen, R., Palmquist, R., 1980. The interpretation of dummy variables in semilogarithmic equations. American Economic Review 70, 474–475.
Handy, S.L., Niemeier, D.A., 1997. Measuring accessibility: an exploration of issues and alternatives. Environment and Planning A 29, 1175–1194.
Hansen, W.G., 1959. How accessibility shapes land use. Journal of American Institute of Planners 25 (1), 73–76.
Hometrack, 2005. Key Housing Market Statistics for England and Wales, The Hometrack National May 2004 Survey of the Housing Market. UK House Price Statistics.
Ibeas, A., Cordera, R., dell'Olio, L., Moura, J.L., 2011. Modelling demand in restricted parking zones. Transportation Research Part A: Policy and Practice 45, 485–498.
Lancaster, K.J., 1966. A new approach to consumer theory. Journal of Political Economy 74, 132–157.
LeSage, J.P., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, Boca Raton.
LeSage, J.P., Pace, R.K., 2010. Spatial econometric models. In: Fischer, M.M., Getis, A. (Eds.), Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer, Berlin, pp. 355–376.
Löchl, M., Axhausen, K.W., 2010. Modelling hedonic residential rents for land use and transport simulation while considering spatial effects. The Journal of Transport and Land Use 3, 39–63.
Maguire, D.J., Batty, M., Goodchild, M.F., 2005. GIS, Spatial Analysis, and Modeling. ESRI Press, Redlands.
Malpezzi, S., 2008. Hedonic pricing models: a selective and applied review. In: O'Sullivan, T., Gibb, K. (Eds.), Housing Economics and Public Policy. Blackwell Science, Oxford, pp. 67–89.
Martínez, L., Viegas, J., 2009. Effects of transportation accessibility on residential property values. Transportation Research Record: Journal of the Transportation Research Board 2115, 127–137.
Mills, E.S., 1972. Studies in the Structure of the Urban Economy. Johns Hopkins Press, Baltimore.
Munoz-Raskin, R., 2010. Walking accessibility to bus rapid transit: does it affect property values? The case of Bogotá, Colombia. Transport Policy 17, 72–84.
Muth, R.F., 1969. Cities and Housing: The Spatial Pattern of Urban Residential Land Use. University of Chicago Press, Chicago.
Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geographical Analysis 27, 286–306.
Ottensmann, J.R., Payton, S., Man, J., 2008. Urban location and housing prices within a hedonic model. Journal of Regional Analysis and Policy 38, 19–35.
Ovenell, N., 2007. A Second Hedonic Longitudinal Study on the Effect on House Prices of Proximity to the Metrolink Light Rail System in Greater Manchester. Unpublished MSc Transport Engineering and Planning Dissertation, University of Salford.
Paelinck, J.H.P., Klaassen, L.H., 1979. Spatial Econometrics. Saxon House, Farnborough.
Pagliara, F., Papa, E., 2011. Urban rail systems investments: an analysis of the impacts on property values and residents' location. Journal of Transport Geography 19, 200–211.
Rodríguez, D.A., Mojica, C.H., 2009. Capitalization of BRT network expansions effects into prices of non-expansion areas. Transportation Research Part A: Policy and Practice 43, 560–571.
Rosen, S., 1974. Hedonic prices and implicit markets: product differentiation in pure competition. Journal of Political Economy 82, 34–55.

Senior, M.L., 2009. Impacts on travel behaviour of greater Manchester's light rail investment (metrolink phase 1): evidence from household surveys and census data. Journal of Transport Geography 17, 187–197.
von Thünen, J.H., 1966. Der isolierte staat in beziehung auf landwirtschaft und nationalökonomie. Oxford University Press, Oxford (translated by Wartenburg, C.M., 1966).
Waddell, P., Ulfarsson, G.F., Franklin, J.P., Lobb, J., 2007. Incorporating land use in metropolitan transportation planning. Transportation Research Part A: Policy and Practice 41, 382–410.
Ward, M.D., Gleditsch, K.S., 2008. Spatial Regression Models. Sage Publications, Thousand Oaks.
Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. Journal of Geographical Systems 7, 161–187.