DRAFT VERSION

Economic Impacts of Wetland Amenities: A Spatial Econometric Analysis of the Dutch Housing Market

by

René van der Kruk

presented at the 41st Congress of the European Regional Science Association, Zagreb, Croatia, from 29th August to 1st September 2001

Department of Spatial Economics
Free University Amsterdam - Tinbergen Institute
Keizersgracht 482
NL-1017 EG Amsterdam
the Netherlands
phone: +31(0)20 551 3525
fax: +31(0)20 551 3555
wetlands@renevdkruk.com
http://www.wetlands.renevdkruk.com

Abstract

The aim of this paper is to focus on the economic impacts of Dutch wetland amenities. In particular, a spatial statistical and econometric analysis of the housing market is performed in order to determine the relationship between the presence of wetland areas and the prices of nearby houses. For this purpose, a database with selling prices and characteristics of houses from the Dutch brokers association (NVM) is used. The approach followed here is closely related to the hedonic pricing method. This method determines the marginal value of various characteristics of a commodity.

In this paper a few novelties will be presented. The spatial cross-autocorrelation between housing prices and environmental (wetland) characteristics is inferred from local Moran's I. In addition a new spatial model called SARMA(**d**) is described together with the decomposition of the highest order spatial link matrix that is required for the estimation of this highly general model. A hybrid spatial link matrix is introduced that makes is possible to model relations between spatial units whose location can only be described by regions instead of (x,y) coordinates without losing information on the characteristics of individual observations.

1 Introduction

The Ramsar Convention (UNESCO, 1994) defines wetlands as: "areas of marsh, fen, peat land or water, whether natural or artificial, permanent or temporary, with water that is static or flowing, fresh, brackish or salt, including areas of marine water the depth of which at low tide does not exceed six meters". When countries join the Convention, they are enlisting in an international effort to ensure the conservation and wise use of wetlands. The Convention on Wetlands came into force for the Netherlands on 23 September 1980. As of February 2001 this country has 24 sites designated as "Wetlands of International Importance". In the near future a number of wetland sites will be added to this list. Figure 1 depicts both existing and future Dutch Ramsar wetland sites.

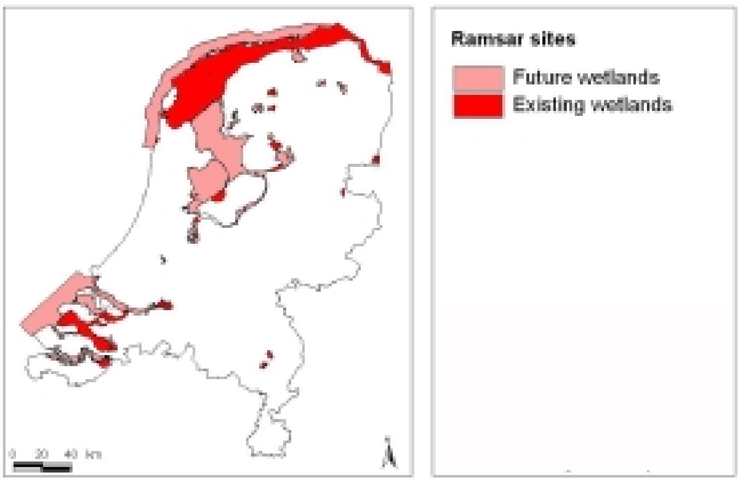


Figure 1: Existing and future Dutch Ramsar wetland sites

Wetlands are important not only because they are cradles of biological and genetic diversity. They also provide positive amenity values for nearby residents. These include open space, enhanced views, a buffer against noise and other forms of pollution. To date only limited research has been conducted which links wetland ecosystem characteristics and functions to the amenity values of wetlands. Mahan et al. (2000) estimate the value of wetland amenities in the Portland metropolitan area using the hedonic property price model. Their results indicate that wetlands influence the value of residential property: increasing the size of the nearest wetland to a residence by one acre increases the residence's value by $24. Similarly, reducing the distance to the nearest wetland by 1,000 feet increases the value by $436. In this paper the economic impact of Dutch

Ramsar wetland amenities on the housing market is analyzed using a spatial hedonic pricing analysis.

Hedonic pricing analysis is based on the hypothesis that differentiated products are valued for their utility-bearing attributes. The pioneering analysis dates back to an article by Court (1939) who used the term hedonic (in capitals) to describe the weighting of the relative importance of various automobile components in constructing price indices of "usefulness and desirability". The theoretical model of the market for a differentiated product developed by Rosen (1974) is still influential, although there have been significant modifications and improvements in the implementation of that model. Palmquist (1999) describes some recent developments in hedonic modeling. In environmental economics hedonic models have been used to estimate the willingness to pay for environmental improvements. Hedonic methods are revealed preference methods, and they represent one of the few instances where environmental quality is traded in actual markets. Housing markets are the most frequently used example of this. Hedonic models seek to extract information on the value of the environmental characteristics from the market for houses. Geoghegan et al (1997) estimate a hedonic model with spatial landscape indices to capture the amenity effects of surrounding land use patterns on the selling prices of houses. In this paper an "appropriate" area around each observation is chosen and measures of percent open space and diversity measured at different scales around that observation are used as indices. This approach goes beyond the usual approach in which spatial considerations are reduced to uni-dimensional measures. In a hedonic pricing framework the houses can also be considered as spatial units of observation. When Andrew Cliff and Ord (1973) published their book on spatial processes, the literature on spatial and space-time processes was "scant indeed". Anselin and Griffith (1988) and Can (1992) argue that methodological developments in spatial statistics and econometrics have shown that the straightforward use of traditional methods may not be adequate for the analysis and modeling of geographically referenced data due to spatial effects, namely spatial dependence and spatial heterogeneity. In the presence of spatial processes the assumption of the independence in the disturbances is violated. Another concept, which is important in spatial econometrics, is called spatial heterogeneity, i.e. functional form and parameters vary with location and are not homogeneous throughout the data set. The presence of spatial heterogeneity leads to a trade-off between locational specificity in the model and identifiability of the parameters and functional forms, within the constraints imposed by data availability. Anselin (1988) presents an introduction to spatial econometric issues. In real estate economics the importance of issues related to spatial processes is realized only recently. For example, Pace (1997) estimates a mixed regressive-spatially autoregressive hedonic model. However, environmental economists thus far did not enter the realm of space. This paper presents a spatial econometric framework in which the effect of environmental amenities can be analyzed. The following question will be answered:

Is it possible to detect and estimate the correlation between the presence of wetland areas and the prices of nearby houses in the Netherlands in 1996 using both spatial statistical and spatial econometric techniques?

In the next part of this paper the data are described. In the third section spatial autocorrelation on the Dutch housing market is analyzed using Moran's *I*. In section 4 the estimates of two non-spatial multiple regression models are presented. The residuals of both models will be examined. In section 5 the spatial cross-autocorrelation between housing prices and environmental (wetland) characteristics is inferred from local Moran's *I*. In the next version of this paper (which will be downloadable on http://www.wetlands.renevdkruk.com) in section 6 a novel spatial model called SARIMA will be estimated using special higher order spatial link matrices. See the technical appendix for details. This paper ends with the main conclusions.

2 Data description

The data that will be used in this paper cover large parts of the Netherlands, in particular highly populated areas such as the *Randstad* and those regions that contain Ramsar wetland sites. Figure 2 depicts several land use categories within the study area in 1996.

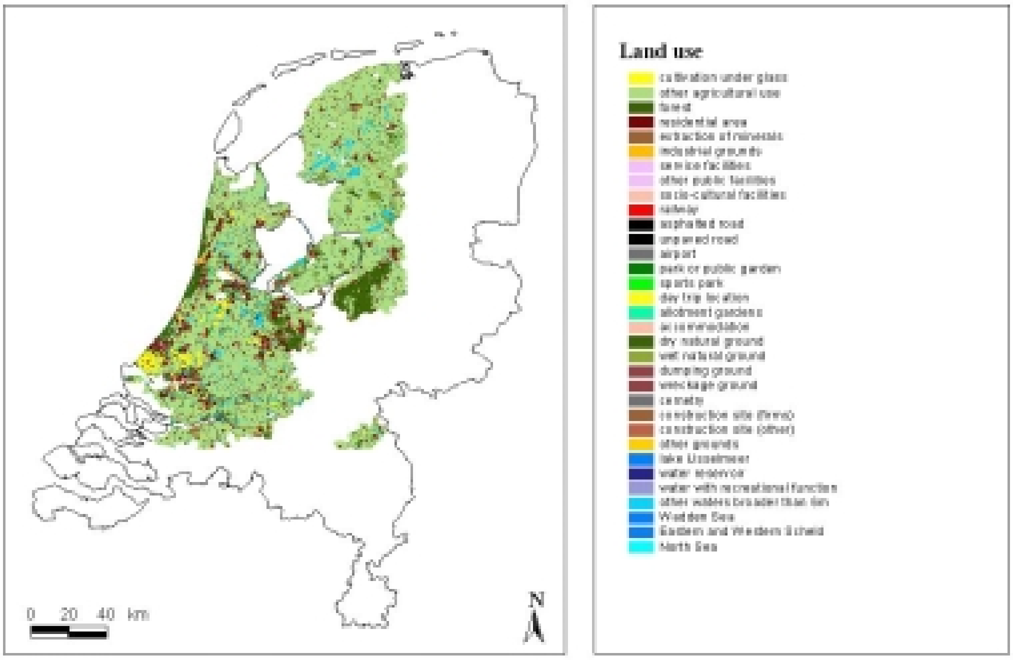


Figure 2: Land use within the study area in 1996

A database of the Dutch brokers association (NVM) will also be used. The total housing market share of the brokers that are member of the NVM is about 60%. This database contains data of 36,615 housing transactions in 1996. Only transactions on existing houses are studied, *id est* newly constructed houses are not considered. The following transaction data are available: transaction date; district number; transaction price; transaction costs; land ownership; capacity; parcel size; construction year; number of rooms; type of living room; type of garage; monument; inside maintenance; outside maintenance; length main garden; number of bathrooms; gas fire; fireplace. Note that the database does not contain grid coordinates of the house, which means that the precise location of a house is unknown. However, the district number gives some information on the location. Figure 3 depicts the total number of transactions in 1996 within each of the 309 districts that make up the study area.

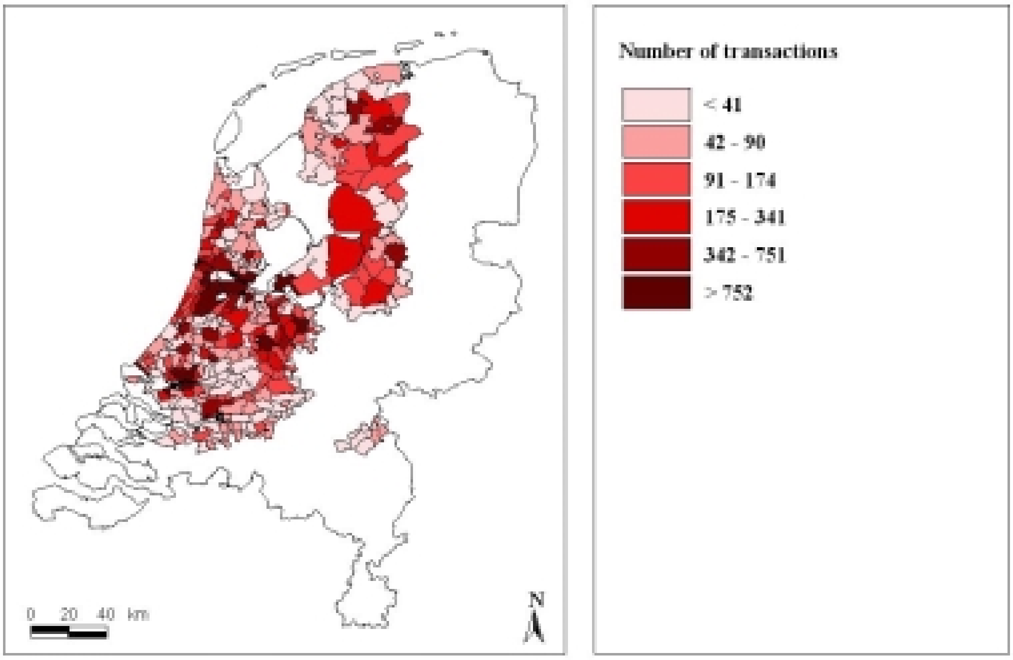


Figure 3: Number of housing transactions per district within the study area in 1996

It is relatively straightforward to show that spatial price patterns are present in the Dutch housing market. Figure 4 depicts the median transaction price for each district. It is clear from this picture that houses in contingent districts have similar transaction price levels.

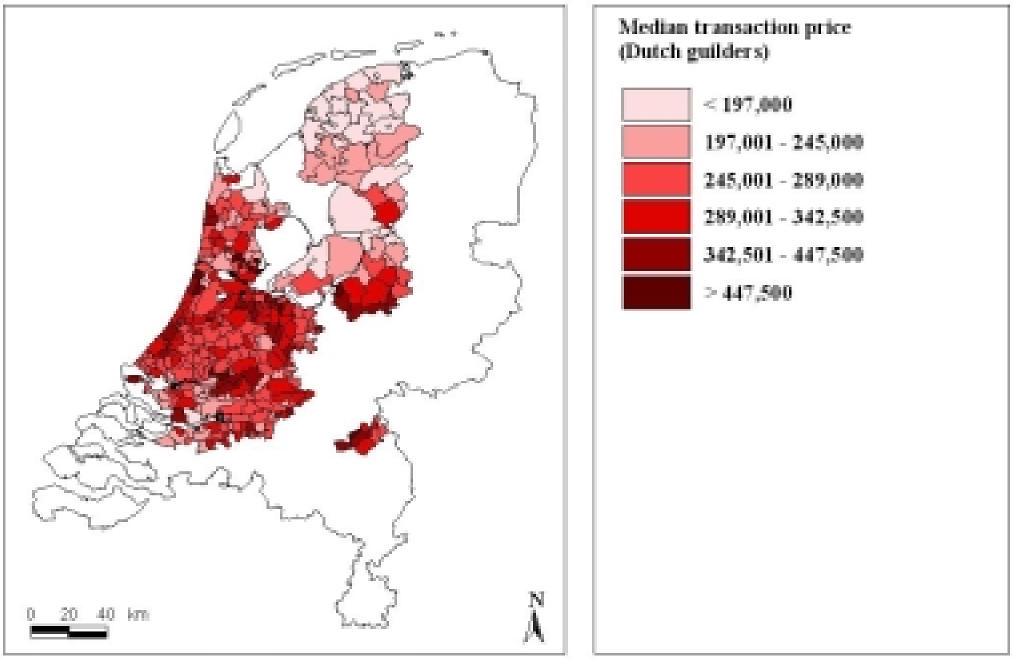


Figure 4: The median transaction price per district in 1996

This spatial autocorrelation is treated in a more formal way in the next section.

3 Spatial autocorrelation on the Dutch housing market

The original measures for spatial autocorrelation advanced by Moran (1948) were based on the notion of binary contiguity between spatial units. A more general concept is the $n \times n$ spatial link matrix $S = \{s_{ij}\}$ that represents the spatial relations between the housing prices in various districts, where $s_{ij} > 0$ if district i and district j are spatially tied together and $s_{ij} = 0$ otherwise ($s_{ii} = 0$ by convention) for $i,j = 1,..,n$. The spatial link matrix can also capture higher order dependence between spatial units that are not neighbors. See the technical appendix for details on a novel decomposition of the spatial link matrix.

One can test for the presence of different kinds of spatial autocorrelation in the housing market by using Moran's *I*. The observed value of this test statistic is defined as a ratio of quadratic forms in the regression residuals ***e***:

$$I_0 \equiv \frac{\mathbf{p}^T M \frac{1}{2}(S+S^T) M\mathbf{p}}{\mathbf{p}^T M\mathbf{p}} = \frac{\mathbf{e}^T \frac{1}{2}(S+S^T)\mathbf{e}}{\mathbf{e}^T\mathbf{e}}.$$

In this expression ***p*** is an n × 1 random vector of the dependent variable and the projection matrix $M \equiv I - X(X^T X)^{-1} X^T$. I is an $n \times n$ identity matrix and X is an $n \times k$ regression matrix of independent variables. S is an $n \times n$ spatial link matrix like the one introduced in the technical appendix. In the basic form of Moran's *I* the unity vector is the only independent variable (i.e. $k = 1$) and specifies the variation of ***y*** around its mean. Tiefelsdorf and Boots (1995; 1996) calculate the exact distribution of Moran's I assuming that the disturbances are normal distributed and that the spatial structure used to encode the underlying spatial relationship is well behaved. This warrants to approximate the significance of an observed value of Moran's I by the normal distribution. However, Tiefelsdorf (1999) notes that "for less well behaved spatial structures results from the normal approximation of the distribution of Moran's *I* can be misleading. Examples of less well-behaved structures are local spatial link matrices; spatial hierarchies or spatial link matrices associated with higher order spatial lags. Common among these spatial structures is that they lead to sparse spatial link matrices." In the same paper he introduces a saddle point approximation of the exact distribution of Moran's *I*. For more details see Tiefelsdorf (2000). If $\varepsilon \sim N(0,\sigma^2\Omega)$ the spectrum of eigenvalues of $D \equiv (\Omega^T)^{\frac{1}{2}} M(S - I_0 I)M\Omega^{\frac{1}{2}}$ determines the exact distribution of Moran's *I* and its saddlepoint approximation. Lieberman (1994) derives the saddlepoint approximation to the density and tail probability of a ratio of quadratic forms in normal variables where $\Omega = I$.

Before Moran's I can be applied to the Dutch housing market there is only one issue that must be resolved: in the previous section it is noted that the database used here does not contain grid coordinates of the houses. Only the district number provides information on the location of the house. One solution would be to make additional assumptions about the spatial contiguity of houses both within a single district and of houses in different districts. However, this approach would come at the cost of maximum number of observations per district that can be used and very restrictive and arbitrary assumptions. One way out would be to consider ZIP code areas instead of district numbers. In any other case even sparse matrix algorithms will not be sufficient to prevent the occurrence of computer memory problems. See the technical appendix for more on this issue.

These remarks provide good reasons to "scale up" the analysis to a higher spatial level: the spatial units of analysis are districts instead of houses. The total number of observations is thus reduced from 36,615 to only 309. A drawback of this approach is that much information on the characteristics of each individual house is lost: restrictive assumptions must be made regarding the homogeneity of the houses within a certain area. Since Moran's I will be used in this section to detect large-scale spatial autocorrelation patterns this approach will be followed here. For each district the median transaction price of the houses, which have been sold in 1996 is determined. In terms of Moran's I the dependent variable $\boldsymbol{p}$ contains these median prices. The 309×309 spatial link matrix S expresses the spatial relations between districts. We first consider the variation of the median prices around the overall mean. Table 1 contains values of the I_0, the sample version of Moran's I. See the technical appendix for more information on coding schemes.

Table 1: Moran's I_0 and p-values for 8 cases

	No coding	*C*-coding	*S*-coding	*W*-coding
I_0 using S_1	1.22	0.25	0.26	0.28
p-value	0.00	0.00	0.00	0.00
I_0 using S	3.60	0.09	0.12	0.15
p-value	0.00	0.00	0.00	0.00

It is clear that there is a strong positive autocorrelation between the median prices of the houses within the study area.

A spatial link matrix (of any order) can also be decomposed into local spatial link matrices as defined by Tiefelsdorf and Boots (1997). These matrices can be applied to analyze the heterogeneity of spatial autocorrelations using local Moran's *I*. In section 5 this test statistic will be used to detect spatial (cross-)autocorrelation in error terms of ordinary regression models.

4 A non-spatial multiple regression model of the Dutch housing market

As noted in the introduction the assumption of the independence in the disturbances is violated in the presence of spatial autocorrelation. In the previous section it is shown that there is a strong (positive) autocorrelation in the Dutch housing market. If one would nevertheless estimate ordinary regression models the results in table 2 are obtained.

Table 2: OLS regression results of both a model with and a model without land use variables. The dependent variable is the transaction price. The independent variables are the attributes of the house. Share of agricultural use other than cultivation under glass is the default land use category.

Variable	Model 1	Model 2	Variable	Model 1	Model 2	Variable	Model 1	Model 2
Constant	216294	-99102	dummy det. stone garage	40749	40206	railway	-	468098
dummy January	-25683	-27184	dummy wooden garage	-1313	14254	asphalted road	-	433124
dummy February	-23662	-24474	dummy built-in garage	45200	48902	unpaved road	-	-2696181
dummy March	-14167	-15515	dummy monument	56068	58936	airport	-	-1055183
dummy April	-13102	-14079	In maint. (1:good – bad:5)	-5920	-8403	parc or public garden	-	-261035
dummy May	-11420	-13824	out maint. (1:good – bad:5)	-25877	-21652	sports park	-	-61280
dummy June	-9382	-9154	dummy 5-10 meter garden	-14092	-702	day trip location	-	471440
dummy July	-7879	-9064	dummy 10-15 meter garden	-10482	7727	allotment gardens	-	706170
dummy August	-6456	-6136	dummy 15-20 meter garden	23462	41560	dry natural ground	-	260587
dummy September	-4224	-4663	dummy 20-50 meter garden	60317	77554	**wet natural ground (wetland)**	-	-271430
dummy October	-1264	-1772	dummy > 50 meter garden	70764	94196	dumping ground	-	3173462
dummy November	690	95	number of bathrooms	32506	31129	wreckage ground	-	-3598969
dummy not KK	31518	33839	dummy gas	-21648	-22971	cemetery	-	1416348
dummy fixed lease	10459	-14144	dummy fireplace	42169	35309	construction site (firms)	-	-416745
dummy variable lease	-5503	-24116	cultivation under glass	-	152531	construction site (other)	-	-178819
capacity	438	427	forest	-	158266	other grounds	-	197297
parcel size	14	15	residential area	-	138804	IJssel Lake	-	-262215
construction year	-52	69	extraction of minerals	-	435679	water reservoir	-	-289675
number of rooms	7504	7404	industrial ground	-	-433093	water with recreational function	-	1085231
Dummy through room	-8200	-11681	service facilities	-	645680	other waters broader than 6 m	-	223281
dummy room and suite	41796	25509	other public facilities	-	29193	Wadden Sea	-	-2960544
dummy undet. stone garage	29965	41661	socio-cultural facilities	-	81831	North Sea	-	-206190

If one would be unaware of the presence of spatial autocorrelation a first glance at the table on the previous page would point towards a negative impact of wetlands on the housing prices. However, the other important spatial concept (heterogeneity) will prove to be important as will become clear later on.

The justification of adding neighborhood characteristics to the first model above is given by figure 6. This map illustrates the spatial distribution of the median value of the errors of model 1 for each district.

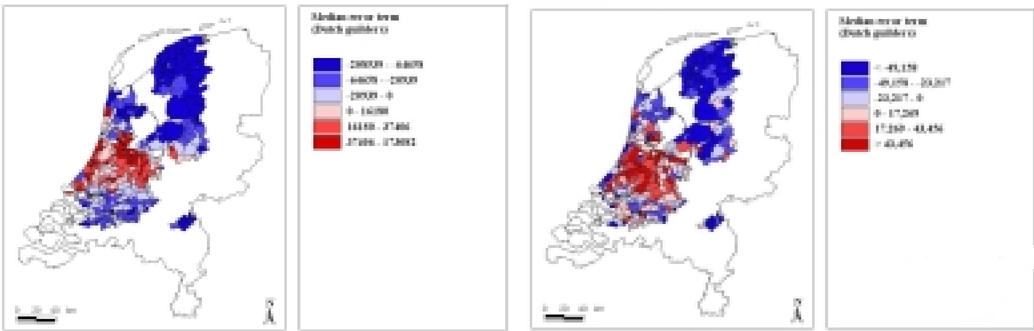


Figure 6: The median error of model 1…. …. and model 2 per district

From figure 6 (and the Moran's I test) it is clear that after correcting for housing characteristics there is still some spatial autocorrelation present in the (median) error terms, i.e. the part of the housing price that cannot be explained by the regressors used either in model 1 or in model 2. The median errors in the *Randstad* districts are positive, while in other parts of the Netherlands the errors are negative. Apparently, people want to live in this part of the country. Note the error differences (from red to blue) in the districts with large shares of dry natural ground such as the dunes near the coast and parts of the *Veluwe*. What about the wetlands?

Lieberman (1994) observes that a ratio of a bilinear form to a quadratic form can be easily transformed into a ratio of quadratic forms. The cross-autocorrelation coefficient

$$I_0 \equiv \frac{\mathbf{y}^T M \frac{1}{2}(H+H^T)M\mathbf{z}}{\mathbf{y}^T M\mathbf{z}}$$

can be used as a test statistic for spatial cross-autocorrelation between the variables $\boldsymbol{y}$ and $\boldsymbol{z}$. In the next section the communality in the spatial patterns of housing prices and wetland characteristics is investigated in more detail using local Moran's I where M is made up of the regressors used in model 1.

5 Spatial cross-autocorrelation between wetlands and housing prices

In the introduction op this paper the question is raised whether it would be possible to detect and estimate the correlation between the presence of wetland areas and the prices of nearby houses in the Netherlands in 1996 using both spatial statistical and spatial econometric techniques. In this section the local cross-autocorrelation test statistic will be used to answer the first part of this question. The second part will be answered in the next section. Figure 7 shows the approximated value of the CDF of local Moran's *I*.

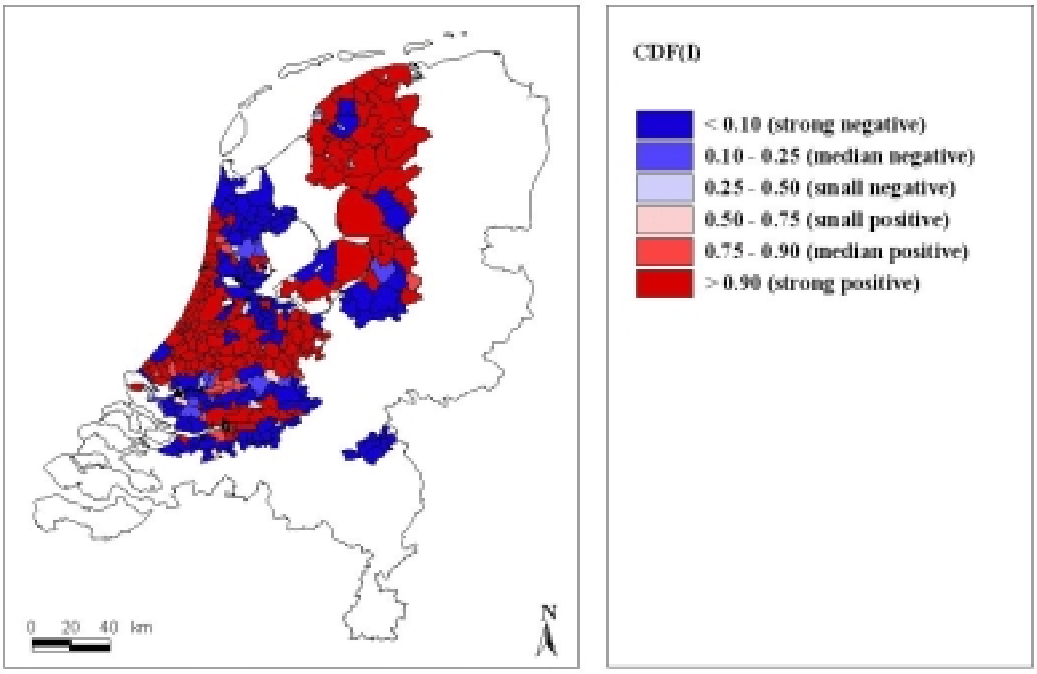


Figure **7:** the approximated value of the CDF of local Moran's *I* per district

From figure 7 and from the value of the global Moran's *I* the conclusion can be drawn that there is an overall strong positive autocorrelation between the median transaction price and the presence of wetlands. There are, however also districts with strong negative autocorrelation.

6 A SARMA model of the Dutch housing market

When applied to spatially distributed observations, ignoring spatial autocorrelation may lead to a serious violation of the assumptions underlying ordinary least squares regression which can result in erroneous statistical inference. The previous sections have shown that some strong (cross-)autocorrelation in the Dutch housing market exists. Fortunately, a variety of spatial models can adjust for this problem. In this section the estimates of some spatial econometric models of the housing market in the Netherlands will be presented. The starting point from the analysis is the following (very general) theoretical model:

$$p_{(x,y),t} \equiv p_{(x,y),t}(\mathbf{z}_{(x,y),t} \mid \boldsymbol{\beta}_{(x,y),t}, P, Z) \quad \text{where } P = \{p_{(\tilde{x},\tilde{y}),\tilde{t}}\}; Z = \{\mathbf{z}_{(\tilde{x},\tilde{y}),\tilde{t}}\}; (\tilde{x},\tilde{y}) \in (X,Y); \tilde{t} \in T$$

In this expression $p_{(x,y),t}$ denotes the revealed (transaction) price of a house at a certain moment in time *t*. The house is situated at a certain point in space, which is defined by the grid coordinates *(x,y)*. The vector variable $\mathbf{z}_{(x,y),t}$ contains the attributes of the house. Given these characteristics of the house, the transaction price is determined by the vector of parameters $\boldsymbol{\beta}_{(x,y),t}$. Note that this formulation allows for a change in both the characteristics of the house and the implicit price of the attributes. Moreover, the functional form of the relationship between the price and the attributes is indefinite. *P* and *Z* denote the sets of the transaction prices and attributes both of houses at other locations, as defined by elements of the set *(X,Y)*, and of the same house at another moment in time, which is defined by elements of the set *T*. Although this highly general model incorporates important features such as space-time autocorrelation and heterogeneity and allows for general function forms such as the quadratic Box-Cox specification, it is obvious that it is impossible to identify the parameters and functional forms, within the constraints imposed by data availability.

In the next version of this paper a class of spatial autoregressive models that is outlined in the technical appendix will also be applied in order to analyze cross-sectional spatial data on the Dutch housing market. Spatial-temporal models like the one introduced by Pace et al. (2000) will not be considered.

6 Conclusions

This paper addresses the following question: Is it possible to detect and estimate the correlation between the presence of wetland areas and the prices of nearby houses in the Netherlands in 1996 using both spatial statistical and spatial econometric techniques? Before answering this question this paper first gives an impression of the data and the spatial autocorrelation in the Dutch housing market. In the third section spatial autocorrelation is detected using Moran's *I*. In section 4 the estimates of two non-spatial multiple regression models are presented. The residuals of both models still contain strong spatial autocorrelation. This warrants the use of spatial models. In section 5 the spatial cross-autocorrelation between housing prices and wetland characteristics is inferred from local Moran's *I*. It is not possible to give a straight answer as to whether wetlands have a positive influence on housing prices. There is a lot of spatial heterogeneity in this relation. There is however a remarkable spatial pattern in the cross-autocorrelation that cannot be explained by the data. In the next version of this paper (which will be downloadable on http://www.wetlands.renevdkruk.com) in section 6 a novel spatial model called SARIMA will be estimated using special higher order spatial link matrices. The technical appendix of this paper already presents a theoretical framework.

Literature

Anselin, Luc. 1988. Spatial econometrics: methods and models. Kluwer Academic Publishers, Dordrecht.

Anselin, Luc and Daniel A. Griffith. 1988. "Do spatial effects really matter in regression analysis?" Papers of the Regional Science Association 65:11-34.

Anselin, Luc and Oleg Smirnov. 1996. "Efficient algorithms for constructing proper higher order spatial lag operators." Journal of Regional Science 36:67-89

Can, Ayse. 1992. "Specification and estimation of hedonic housing price models." Regional Science and Urban Economics 22:453-474.

Cliff, Andrew and J. Keith Ord. 1973. Spatial autocorrelation. Pion, London.

Court, Andrew T. 1939. "Hedonic price indexes with automotive examples." The Dynamics of Automobile Demand." General Motors, New York, pp. 98-119.

Geoghegan, Jacqueline, Lisa A. Wainger, and Nancy E. Bockstael. 1997. "Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS." Ecological Economics 23:251-264.

Lieberman, Offer. 1994. "Saddlepoint approximation for the distribution of a ratio of quadratic forms in normal variables." Journal of the American Statistical Association 89:924-928.

Mahan, Brent L., Stephen Polasky, and Richard M. Adams. 2000. "Valuing urban wetlands: A property price approach." Land Economics 76:100-113.

Moran, P. A. P. 1948. "The interpretation of statistical maps." Journal of the Royal Statistical Society B 10:243-251.

Pace, R. Kelley. 1997. "Performing large spatial regressions and autoregressions." Economics Letters 54:283-291.

Pace, R. Kelley, Ronald Barry, Otis W. Gilley, and C.F. Sirmans. 2000. "A method for spatial-temporal forecasting with an application to real estate prices." International Journal of Forecasting 16:229-246.

Palmquist, Raymond B. 1999. "Hedonic models." in Jeroen C. J. M.van den Bergh (ed.) The Handbook on Environmental and Resource Economics. Edward Elgar Publishing, pp. 765-776.

Rosen, Sherwin. 1974. "Hedonic prices and implicit markets: Product differentiation in pure competition." Journal of Political Economy 82:34-55.

Tiefelsdorf, Michael. 1999. "Approximations of the exact distribution of global and local Moran's I." Paper presented at the 46th North American Meetings of the Regional Science Association International in Montreal, November 11-14, 1999.

Tiefelsdorf, Michael. 2000. Modeling spatial processes: the identification and analysis of spatial relationships in regression residuals by means of Moran’s I. Springer-Verlag Berlin Heidelberg.

Tiefelsdorf, Michael, and Barry Boots. 1995. "The exact distribution of Moran's I." Environment and Planning A 27:985-999.

Tiefelsdorf, Michael, and Barry Boots. 1996. "Letters to the editor. The exact distribution of Moran's I." Environment and Planning A 28:1900.

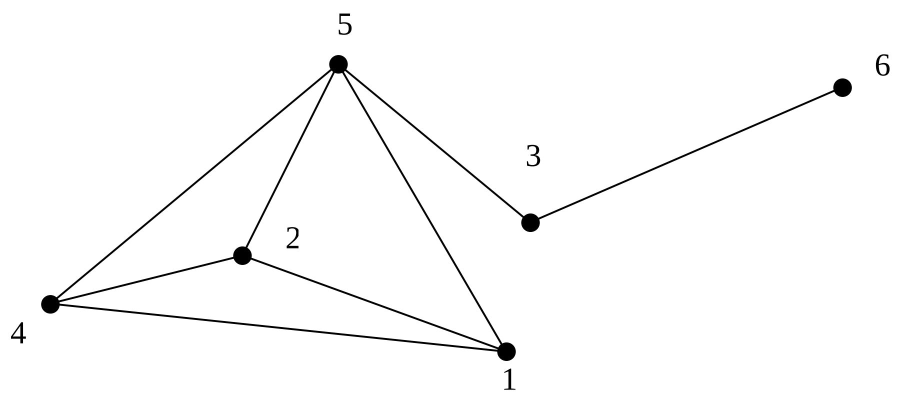
Tiefelsdorf, Michael, and Barry Boots. 1997. "A note on the extremeties of local Moran's I's and their impact on global Moran's I." Geographical Analysis 29:248-257

Tiefelsdorf, Michael, Daniel A. Griffith, and Barry Boots. 1999. "A variance-stabilizing coding scheme for spatial link matrices." Environment and Planning A 31:165-180.

UNESCO. 1994. "Convention on Wetlands of International Importance especially as Waterfowl Habitat". Paris.

Technical appendix

In this appendix a new spatial link matrix is introduced, which captures higher order spatial dependence between spatial units. This matrix can be obtained by a small change in the so-called higher order spatial lag operators. Anselin and Smirnov (1996) present efficient algorithms to compute higher order spatial lag operators without redundant and circular patterns. They use a simple example of spatial dependence, which can be represented by the following graph.



A novelty in this paper is the introduction and decomposition of a spatial link matrix which summarizes the spatial dependence:

$$S \equiv \sum_{d=1}^{D} S_d = S_1 + S_2 + S_3 =$$

$$= \begin{bmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 0 & 1 & \frac{1}{2} & 1 & 1 & \frac{1}{3} \\ 1 & 0 & \frac{1}{2} & 1 & 1 & \frac{1}{3} \\ \frac{1}{2} & \frac{1}{2} & 0 & \frac{1}{2} & 1 & 1 \\ 1 & 1 & \frac{1}{2} & 0 & 1 & \frac{1}{3} \\ 1 & 1 & 1 & 1 & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & 1 & \frac{1}{3} & \frac{1}{2} & 0 \end{bmatrix}$$

S_d is a d^{th} order spatial link matrix. The index d runs from 1 until D, which is the diameter of the graph. By definition the diameter is the highest order link in the system of spatial units. S_d represents the d^{th} order spatial link between the spatial units. In this matrix a d^{-1} appears when spatial units are linked together via d steps in the graph. The difference with the paper by Anselin and Smirnov (1996) is that *d* is inverted. This is done because it is easier to use S_d as a spatial weight matrix. The intuition is that the spatial link between units that are further away from each other is weighted less. S_1 is the first order spatial link matrix, i.e. the contiguity matrix defined on the first order spatial neighborhood relation between adjacent spatial objects. S_D is the highest order spatial link matrix. It represents the highest order spatial contiguity. The sum of all spatial link matrices yields the matrix S, which captures all the spatial link relations of the spatial units. For most spatial research the matrix S_1 is used. However, if one wants to detect higher order spatial dependence, it is more appropriate to use either one or a combination of higher order spatial link matrices.

Coding schemes

In order to cope with heterogeneity, which is induced by the different linkage degrees of the spatial objects a spatial link matrix is converted using coding schemes. The paper by Tiefelsdorf et al. (1999) describes the *C*-coding, *W*-coding, and *S*-coding schemes that can be used. The different results of applying the coding schemes to the first order spatial link matrix S_1 and the spatial link matrix S are presented below.

$$\text{C}-\text{coding of } S_1 = \begin{bmatrix} 0.00 & 0.38 & 0.00 & 0.38 & 0.38 & 0.00 \\ 0.38 & 0.00 & 0.00 & 0.38 & 0.38 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.38 & 0.38 \\ 0.38 & 0.38 & 0.00 & 0.00 & 0.38 & 0.00 \\ 0.38 & 0.38 & 0.38 & 0.38 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.38 & 0.00 & 0.00 & 0.00 \end{bmatrix} \quad \text{C}-\text{coding of } S = \begin{bmatrix} 0.00 & 0.27 & 0.14 & 0.27 & 0.27 & 0.09 \\ 0.27 & 0.00 & 0.14 & 0.27 & 0.27 & 0.09 \\ 0.14 & 0.14 & 0.00 & 0.14 & 0.27 & 0.27 \\ 0.27 & 0.27 & 0.14 & 0.00 & 0.27 & 0.09 \\ 0.27 & 0.27 & 0.27 & 0.27 & 0.00 & 0.14 \\ 0.09 & 0.09 & 0.27 & 0.09 & 0.14 & 0.00 \end{bmatrix}$$

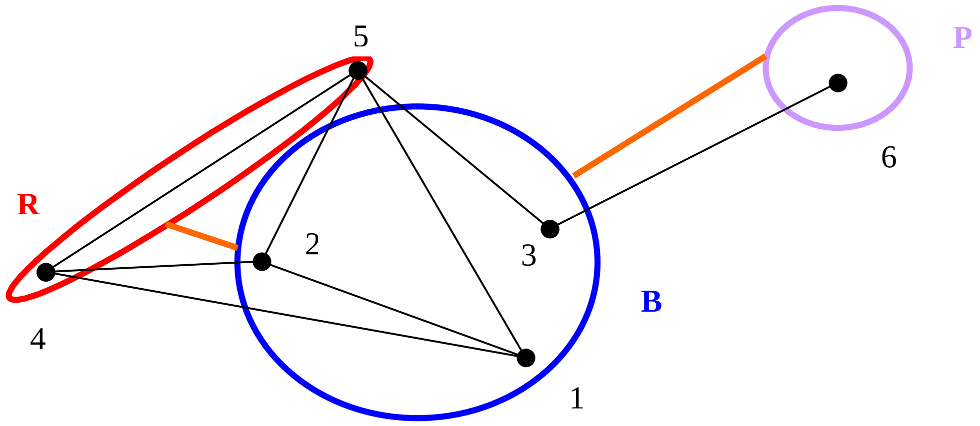
$$\text{S}-\text{coding of } S_1 = \begin{bmatrix} 0.00 & 0.36 & 0.00 & 0.36 & 0.36 & 0.00 \\ 0.36 & 0.00 & 0.00 & 0.36 & 0.36 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.44 & 0.44 \\ 0.36 & 0.36 & 0.00 & 0.00 & 0.36 & 0.00 \\ 0.31 & 0.31 & 0.31 & 0.31 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.62 & 0.00 & 0.00 & 0.00 \end{bmatrix} \quad \text{S}-\text{coding of } S = \begin{bmatrix} 0.00 & 0.27 & 0.13 & 0.27 & 0.27 & 0.09 \\ 0.27 & 0.00 & 0.13 & 0.27 & 0.27 & 0.09 \\ 0.14 & 0.14 & 0.00 & 0.14 & 0.28 & 0.28 \\ 0.27 & 0.27 & 0.13 & 0.00 & 0.27 & 0.09 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.00 & 0.12 \\ 0.11 & 0.11 & 0.33 & 0.11 & 0.17 & 0.00 \end{bmatrix}$$

$$\text{W}-\text{coding of } S_1 = \begin{bmatrix} 0.00 & 0.33 & 0.00 & 0.33 & 0.33 & 0.00 \\ 0.33 & 0.00 & 0.00 & 0.33 & 0.33 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 0.50 & 0.50 \\ 0.33 & 0.33 & 0.00 & 0.00 & 0.33 & 0.00 \\ 0.25 & 0.25 & 0.25 & 0.25 & 0.00 & 0.00 \\ 0.00 & 0.00 & 1.00 & 0.00 & 0.00 & 0.00 \end{bmatrix} \quad \text{W}-\text{coding of } S = \begin{bmatrix} 0.00 & 0.26 & 0.13 & 0.26 & 0.26 & 0.09 \\ 0.26 & 0.00 & 0.13 & 0.26 & 0.26 & 0.09 \\ 0.14 & 0.14 & 0.00 & 0.14 & 0.29 & 0.29 \\ 0.26 & 0.26 & 0.13 & 0.00 & 0.26 & 0.09 \\ 0.22 & 0.22 & 0.22 & 0.22 & 0.00 & 0.11 \\ 0.13 & 0.13 & 0.40 & 0.13 & 0.20 & 0.00 \end{bmatrix}$$

Note that the numbers are rounded off at two digits. By construction each of the entries of the *S*-coding schemes are in between the corresponding elements of the *C*-coding and *W*-coding schemes. The variation of the entries among the coding schemes using the matrix S is much smaller than the variation of the elements of the matrices corresponding to the first order spatial link matrix. The explanation for this observation is the higher order spatial character of the matrix S that has an extra stabilizing effect. These findings indicate that the effect of the chosen coding scheme is more important if one only considers first order spatial links. If higher order spatial links are also important, the choice of the coding scheme is less relevant.

A solution to the location issue

If there is more than one observation (house) in at least one spatial unit (district) the following method can be used to construct an artificial spatial link matrix. Consider the example used in this technical appendix.



The graph above is made up of the spatial units in the graph used earlier. However, There is also a higher spatial scale. The spatial observations 1, 2, and 3 lie within the district marked by the blew oval. The spatial unit 6 is a singleton observation set within the purple district. If the spatial links between houses are not known, but if each spatial connection between districts is known the following matrix can be constructed.

$$S^{nw} \equiv \begin{bmatrix} 0 & 2\cdot\frac{1}{3} & 2\cdot\frac{1}{3} & 1\cdot\frac{1}{2} & 1\cdot\frac{1}{2} & 1\cdot\frac{1}{1} \\ 2\cdot\frac{1}{3} & 0 & 2\cdot\frac{1}{3} & 1\cdot\frac{1}{2} & 1\cdot\frac{1}{2} & 1\cdot\frac{1}{1} \\ 2\cdot\frac{1}{3} & 2\cdot\frac{1}{3} & 0 & 1\cdot\frac{1}{2} & 1\cdot\frac{1}{2} & 1\cdot\frac{1}{1} \\ 1\cdot\frac{1}{3} & 1\cdot\frac{1}{3} & 1\cdot\frac{1}{3} & 0 & 2\cdot\frac{1}{2} & \frac{1}{2}\cdot\frac{1}{1} \\ 1\cdot\frac{1}{3} & 1\cdot\frac{1}{3} & 1\cdot\frac{1}{3} & 2\cdot\frac{1}{2} & 0 & \frac{1}{2}\cdot\frac{1}{1} \\ 1\cdot\frac{1}{3} & 1\cdot\frac{1}{3} & 1\cdot\frac{1}{3} & \frac{1}{2}\cdot\frac{1}{2} & \frac{1}{2}\cdot\frac{1}{2} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{2}{3} & \frac{2}{3} & \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{2}{3} & 0 & \frac{2}{3} & \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{2}{3} & \frac{2}{3} & 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 1 & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{4} & \frac{1}{4} & 0 \end{bmatrix}$$

The first factor of the matrix element equals d and the second factor is the inverse of the total number of observations within the other district. Note that the element that represents the spatial link between houses within a district is set equal to 2.

SARMA(d) model

The results in the previous sections of this technical appendix can be used to construct a new spatial ARMA($\boldsymbol{d}$) or SARMA($\boldsymbol{d}$) model where $d = [d_i]_{i=1,2,3}$.

$$\boldsymbol{y} = \boldsymbol{\alpha W}(d_1)\boldsymbol{y} + \boldsymbol{\beta W}(d_2)\boldsymbol{X} + \boldsymbol{\gamma}Y + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} = \boldsymbol{\delta W}(d_3)\boldsymbol{\varepsilon} + \boldsymbol{\chi}$$
$$\boldsymbol{\chi} \sim \mathrm{N}(0, \sigma^2 I_n)$$

In the SARMA model $\boldsymbol{W}(d_i) = [S_k]_{k=1,..,d[i]}$ for $0 \leq d_i \leq D$. Note that in literature $\boldsymbol{W}(d_1) = \boldsymbol{W}(d_3) = S_1$ while $\boldsymbol{W}(d_2) = I_n$. The advantage of this general model is that higher order spatial links can also be taken into account. This would solve the issue raised by Dubin (1992) who argues, "even if a set of variables could be agreed upon, a severe measurement problem exists. Neighborhood measures are necessarily geographic in nature. Therefore, in order to measure some aspect of the neighborhood, one must first know what its boundaries are."