Spatial Autocorrelation and the Spatial Durbin Model

ROBUST Workshop

Essen/RWI 21 Feb 2019

Lea Eilers

Spatial Relations

Why is it important to consider spatial relations?

- „Everything is related to everything else, but near things are more related than distant things.“ (Tobler‘s first law of geography, (Tobler, 1979))

- It is important to consider two-way spatial relations when analyzing statistical data relating to regions
- In general: we can infer regional structural dependencies from observing a correlation between data points and if this correlation follows a certain spatial pattern (Anselin, LeGallo und Jayet, 2008).
- → Spatial correlation and spatial autocorrelation

Spatial correlation

Correlation: Relationship between two quantitative variables.

- Positive correlation: A high value of one variable tends to go with a high value of the other variable.
- Negative correlation: A high value of one variable tends to go with a low value of the other variable.
- Measurement: Correlation coefficient.

Spatial correlation:

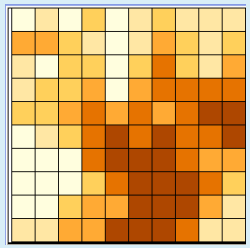
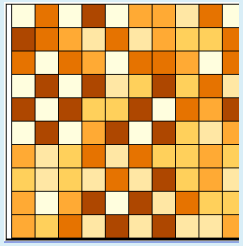
- Two features are spatially correlated if they are close to each other and are similar in their attribute values.
- Positive spatial correlation: A high value of the attribute of one feature tends to go with a high value of the other feature.
- Negative spatial correlation: A high value of the attribute of one feature tends to go with a low value of the other feature.
- Measurement: Semivariogram.

Spatial autocorrelation

- **Autocorrelation**: Observations of the same variable are correlated, i.e. observations are not independent but show a dependency.
- Measurement of time-series autocorrelation: Durbin Watson Test

- **Spatial autocorrelation:**
- Dependencies in all direction
  - i.e. the development of an object can be influenced by the events in all other objects of the same total space
- Neighboring observations of the same phenomenon are correlated.
- Measurement of spatial autocorrelation: Moran's *I* (spatial correlation coefficient) and LISA Cluster Map.

Spatial autocorrelation

Positive autocorrelation	Negative autocorrelation
	
• Development of clusters • Similar values of a variable are more concentrated than expected given a random distribution (Anselin, 1988).	• Dissimilar values of a variable are more often neighboring than in a random distribution. • Perfect negative autocorrelation: chequered pattern (Anselin und Bera; 1998).

Can we detect spatial clusters?

- Statistical concept for measuring spatial autocorrelation: **Moran's I** (1948).

- uses the cross products of geographical neighbors
- Null hypothesis: No spatial autocorrelation

$$I = \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{(\sum_i \sum_j w_{ij}) \ * \ \sum_i (x_i - \bar{x})^2}$$

- Positive autocorrelation: $0 < I \leq 1$
- Negative autocorrelation : $-1 \leq I < 0$
- No autocorrelation : $I = 0$

- The weighting matrix, and hence the computation of the autocorrelation, are influences by assumptions about the structure and intensity of spatial effects.

Spatial weight matrix

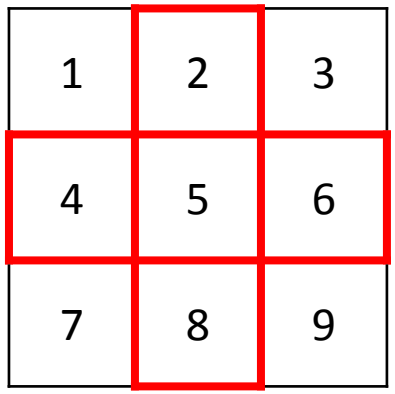
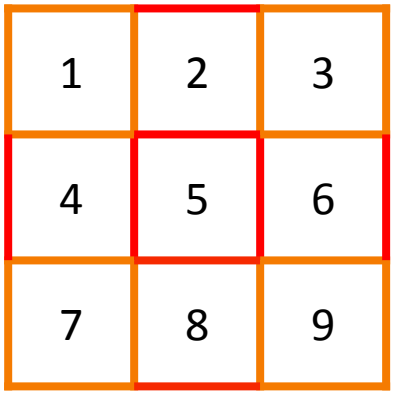
- Exogenous defined, based on prior knowledge
- Given as:

$$W^* = \begin{pmatrix} w_{11} & w_{12} & w_{13} & \cdots & w_{1n} \\ w_{21} & w_{22} & w_{23} & \cdots & w_{2n} \\ w_{31} & w_{32} & w_{33} & \cdots & w_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & w_{n3} & \cdots & w_{nn} \end{pmatrix}$$

- with w_{ij} being the spatial linkage between observation *i* and *j* with $w_{ii} = 0$ and $i = 1 \ldots n$ and $j = 1 \ldots n$.

  n: number of observation

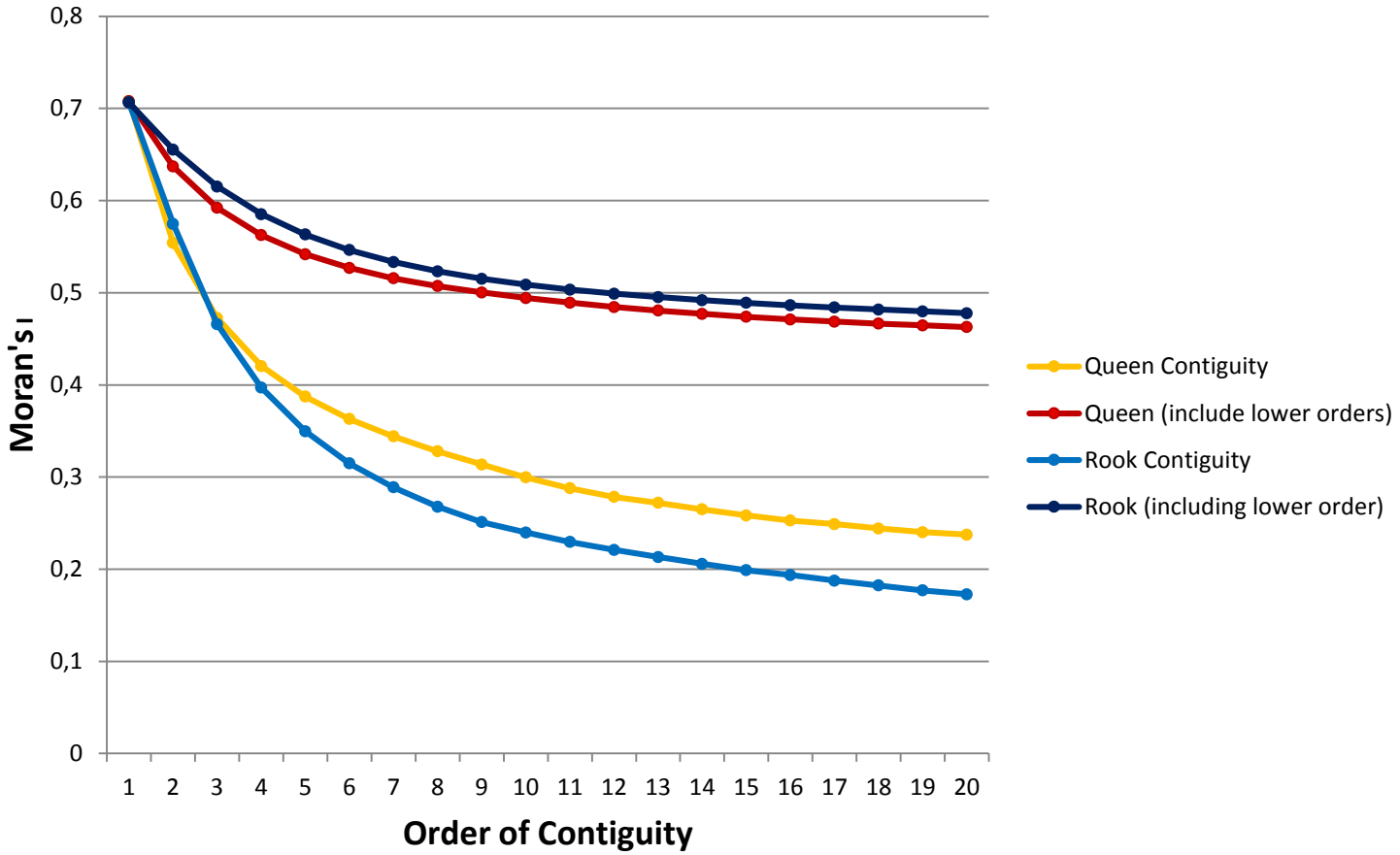
Definition

Rook Contiguity	Queen Contiguity
	
• Common border Neighbors 5: 2,4,6,8	• Common border and corner points Neighbors 5: 1,2,3,4,6,7,8,9

Definition

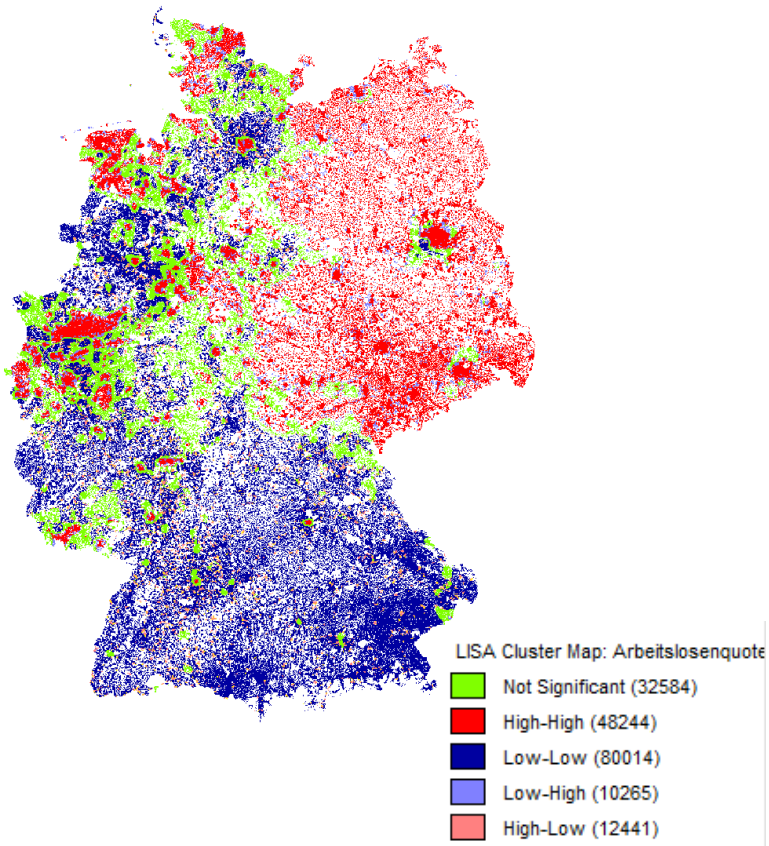
Rook Contiguity	Queen Contiguity
10 11 12 13 14 15 1 2 3 16 17 4 5 6 18 19 7 8 9 20 21 22 23 24 25	10 11 12 13 14 15 1 2 3 16 17 4 5 6 18 19 7 8 9 20 21 22 23 24 25
• Common border Neighbors 5: 2,4,6,8	• Common border and corner points Neighbors 5: 1,2,3,4,6,7,8,9

Autocorrelation and Order of Contiguity

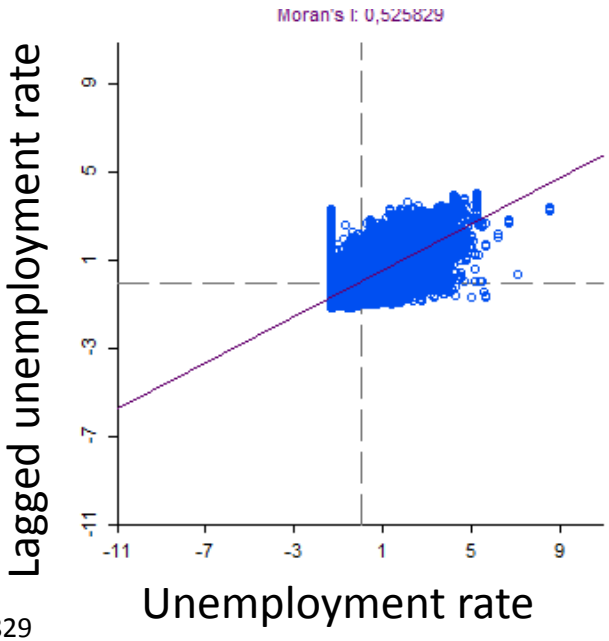


Spatial Autocorrelation: Unemployment rate

LISA- Cluster Map



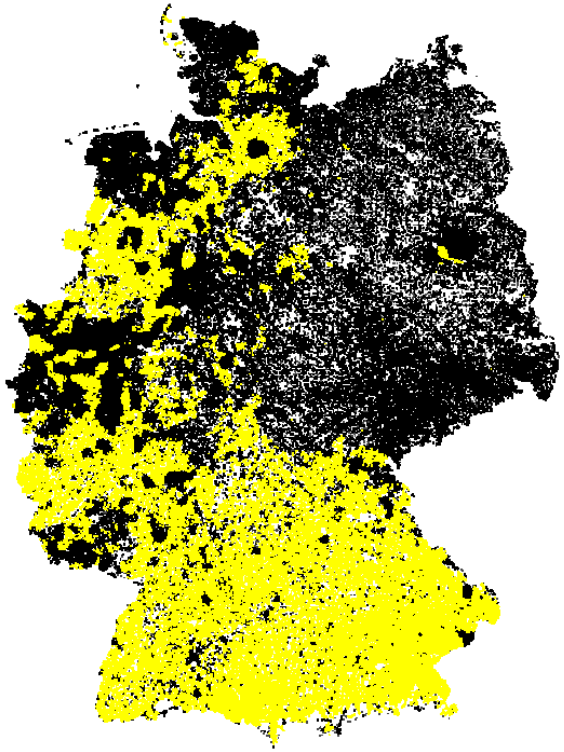
Moran's I



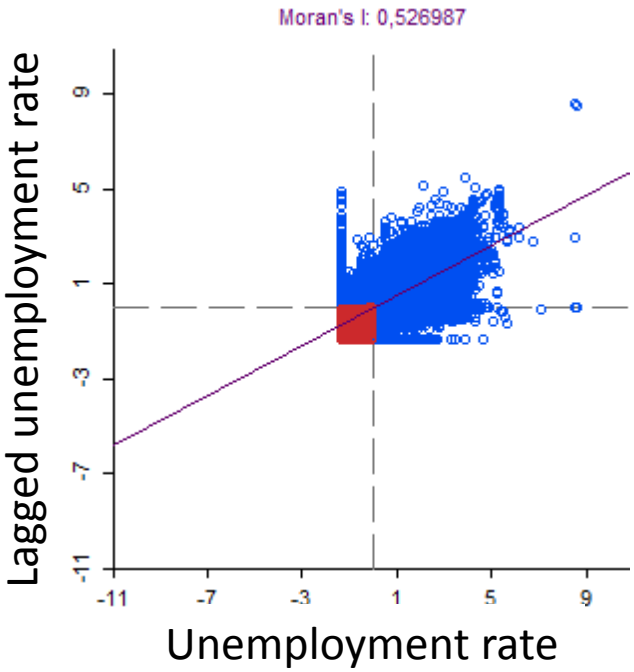
I= 0,525829
p= 0,001
sd= 0,0004
E[I]= 0

Spatial Autocorrelation: Unemployment rate

LISA- Cluster Map
Positive autocorrelation (Low-Low)

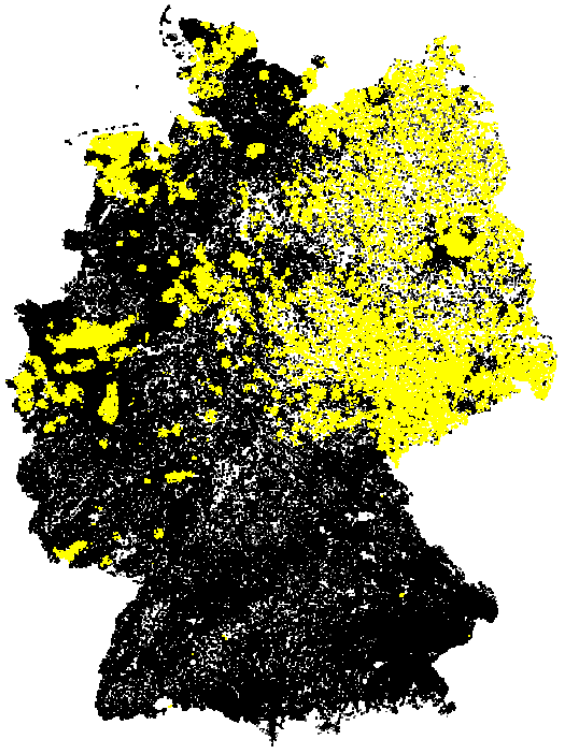


Moran‘s I
Positive autocorrelation (Low-Low)

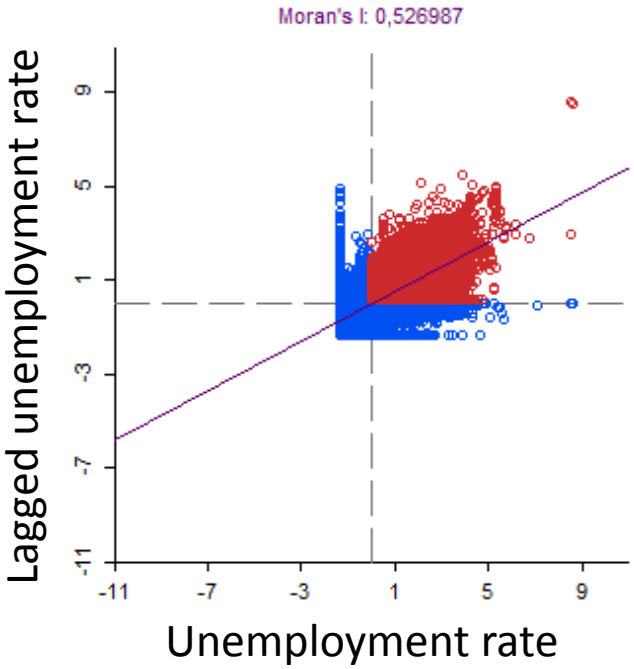


Spatial Autocorrelation: Unemployment rate

LISA- Cluster Map
Positive autocorrelation (High- High)



Moran‘s I
Positive autocorrelation (High- High)



Ignoring spatial (auto-)correlation

1. If spatial correlation is present, OLS can be biased and needs to be interpreted with caution.
   - Several OLS assumptions might not be satisfied.

2. Spatial dependencies in a dependent variable which is a weighted spatial average:
   - OLS estimates are biased and inconsistent

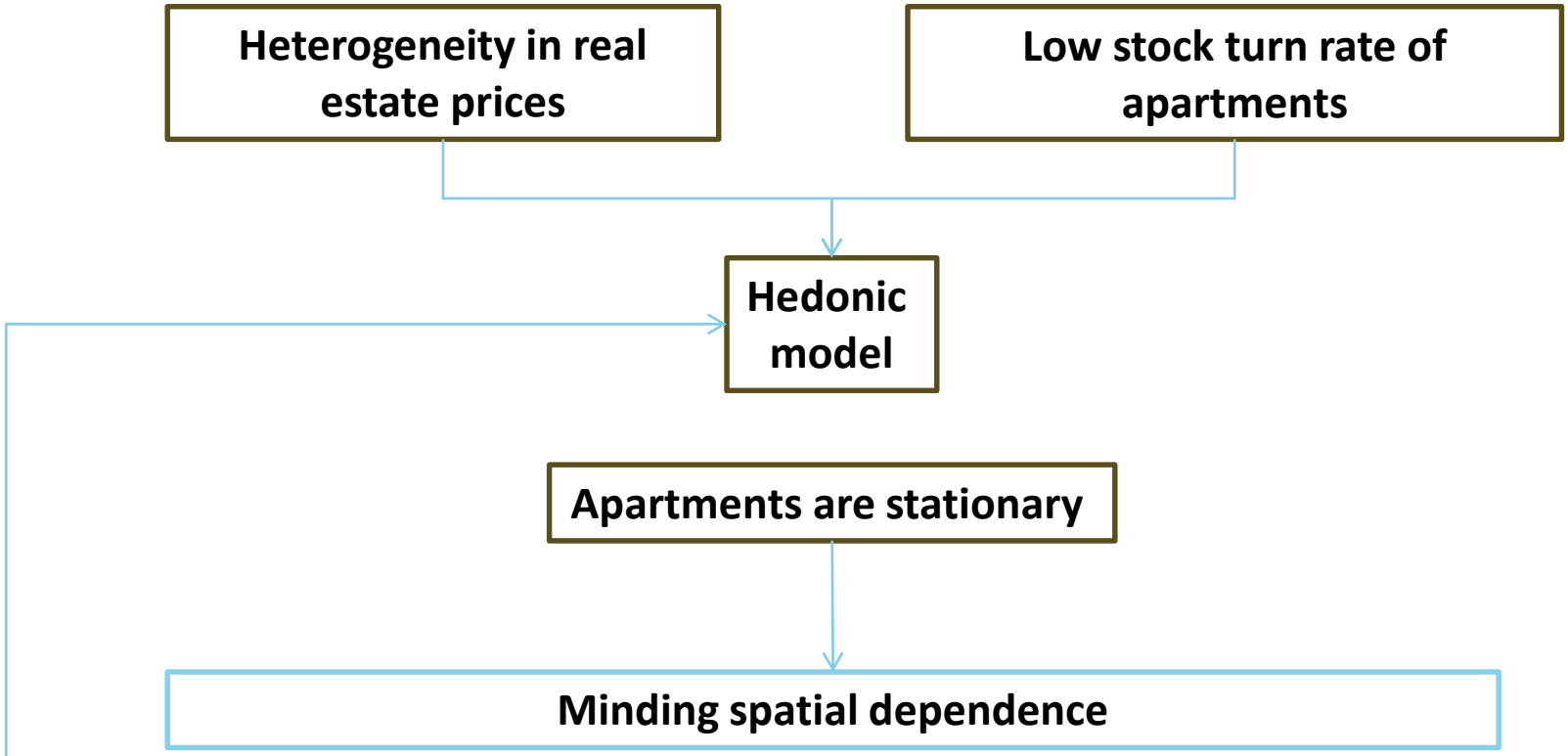
3. Spatial dependencies in the error term which is a weighted spatial average:
   - OLS estimates are unbiased for regression coefficients, however they are not efficient

Solution:

Extend the linear regression model by

- Spatially endogenous interactions: Spatial Lag Model
- Spatial interactions in the error: Spatial Error Model
- Both spatially endogenous interactions and spatial interactions in the error term as well as exogenous interactions: Spatial Durbin Model

Spatial dependence in apartment prices: Application of the spatial Durbin model



- Similar structural characteristics, such as building material, total living area, age of construction, garage and storage rooms.
- Households in the same neighbourhood share common social services (schools, health centres, libraries or malls…)
- Households in the same neighbourhood share the same distance to administrative and commercial agglomerations.

Contribution

i. We estimate a SDM model rather than a SAR or SEM model. The advantage of SDM is that its spillover effects are flexible, while those of SAR are not and those of SEM are even set to zero by construction.

ii. We control for neighborhood and time fixed effects. If these controls are not included, the spatial interaction effects and therefore the spillover effects may be biased, in most cases overestimated.

iii. We calculate direct and spillover effects, since the coefficient estimates do not reflect the impacts of the explanatory variables of one apartment on the price of another issues.

Ordinary-Least-Squares

- Hedonic apartment price equation:

$$P = \alpha \iota_n + X\beta + \varepsilon$$

- P: (n ×1) vector of the log of the apartment offering price
- n: number of observations included in the model
- ι_n: (n ×1) vector of ones associated with the constant term parameter α to be estimated
- β: ($k \times 1$) vector of unknown parameters associated with exogenous explanatory variables (apartment attributes), X, representing an ($n \times k$) matrix
- ε: (n x 1) vector whose elements follow $\varepsilon \overset{iid}{\sim} (0, \sigma^2 I_n)$.

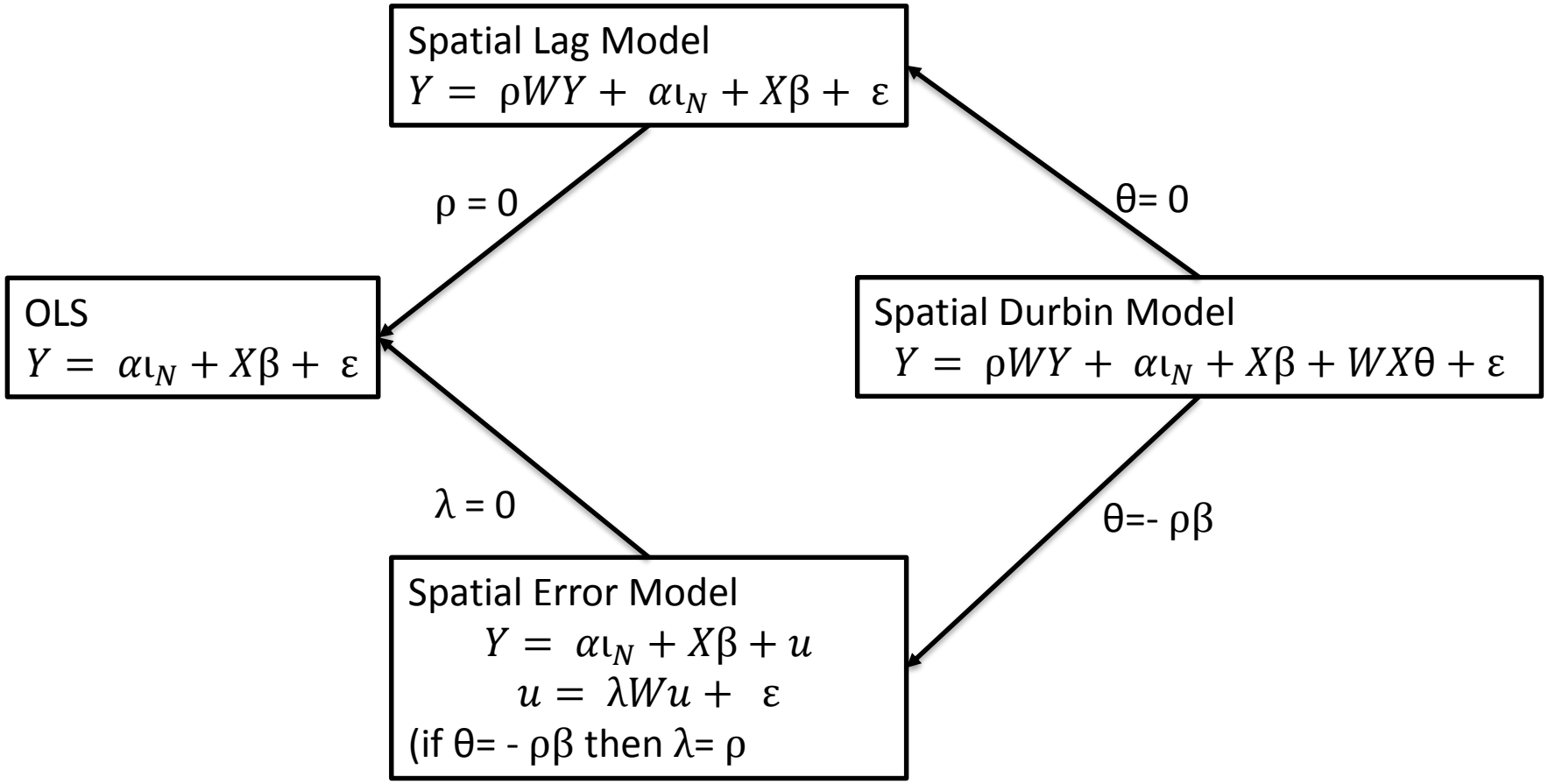
Spatial Durbin Model

- Includes spatially lagged dependent variable and spatially lagged explanatory variables:

$$P = \rho WP + \alpha \iota_n + X\beta + WX\theta + \varepsilon$$

- P, X and β are defined as above
- ρ: spatial autoregressive parameter
- WP: is the spatially lagged offering price accounting for various spatial dependencies with W defined as (*n x n*) spatial weight matrix
- ρWP: Endogenous interaction effect
- θ: (*k x 1*) vector of unknown parameters
- $\theta\ WX$: Exogenous interaction effect

The relationship between different spatial dependence models for cross-section data



Spatial Weight Matrix

- W based on prior knowledge (exogenously determined)
- *k*-nearest neighbours' (based on actual distances) W defined as in Baumont et al. (2004):

$$\begin{cases} w_{ij}(k) = 0 \text{ if } i = j, \forall k \\ w_{ij} = 1 \text{ if } d_{ij} \leq d_i(k) \text{ and } w_{ij}(k) = \dfrac{w_{ij}}{\sum_j w_{ij}(k)} \\ w_{ij}(k) = 0 \text{ if } d_{ij} > d_i(k) \end{cases}$$

- W row normalized: each row sums up to one and leads to asymmetry in the case of actual distances

Direct and Indirect Effects

- **Idea**: point estimates for one or more spatial regression specifications to test the hypothesis as to whether or not spatial spillover exist leads to erroneous conclusions (LeSage & Pace 2009).

- **Solution:** Partial derivatives interpretation of the impact from changes to the variable of different models specifications represent a more valid basis for testing the hypothesis.

  - **Direct effect:** change of a particular apartment characteristic in a particular apartment changes the dependent variable.
  - **Indirect effect:** measures the impact on the price of a particular apartment from changing an exogenous variable in another apartment.

Direct, Indirect and Total Effects

- Partial derivatives of the expected values of P with respect to the explanatory variables x_{nk}:

$$\begin{bmatrix} \frac{\partial E(P_1)}{\partial x_{1k}} & \frac{\partial E(P_1)}{\partial x_{2k}} & \cdots & \frac{\partial E(P_1)}{\partial x_{nk}} \\ \frac{\partial E(P_2)}{\partial x_{1k}} & \frac{\partial E(P_2)}{\partial x_{2k}} & \cdots & \frac{\partial E(P_2)}{\partial x_{nk}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E(P_N)}{\partial x_{1k}} & \frac{\partial E(P_N)}{\partial x_{2k}} & \cdots & \frac{\partial E(P_N)}{\partial x_{nk}} \end{bmatrix} = ((I_N - \rho W)^{-1}) \begin{bmatrix} \beta_k & w_{12}\theta_k & \cdots & w_{1N}\theta_k \\ w_{21}\theta_k & \beta_k & \cdots & w_{2N}\theta_k \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1}\theta_k & w_{N2}\theta_k & \cdots & \beta_k \end{bmatrix}$$

- No prior restrictions are imposed on the magnitude of direct and indirect effects
- Ratio between the indirect effects and the direct effect may be different for different explanatory variables

Data

Data Source:

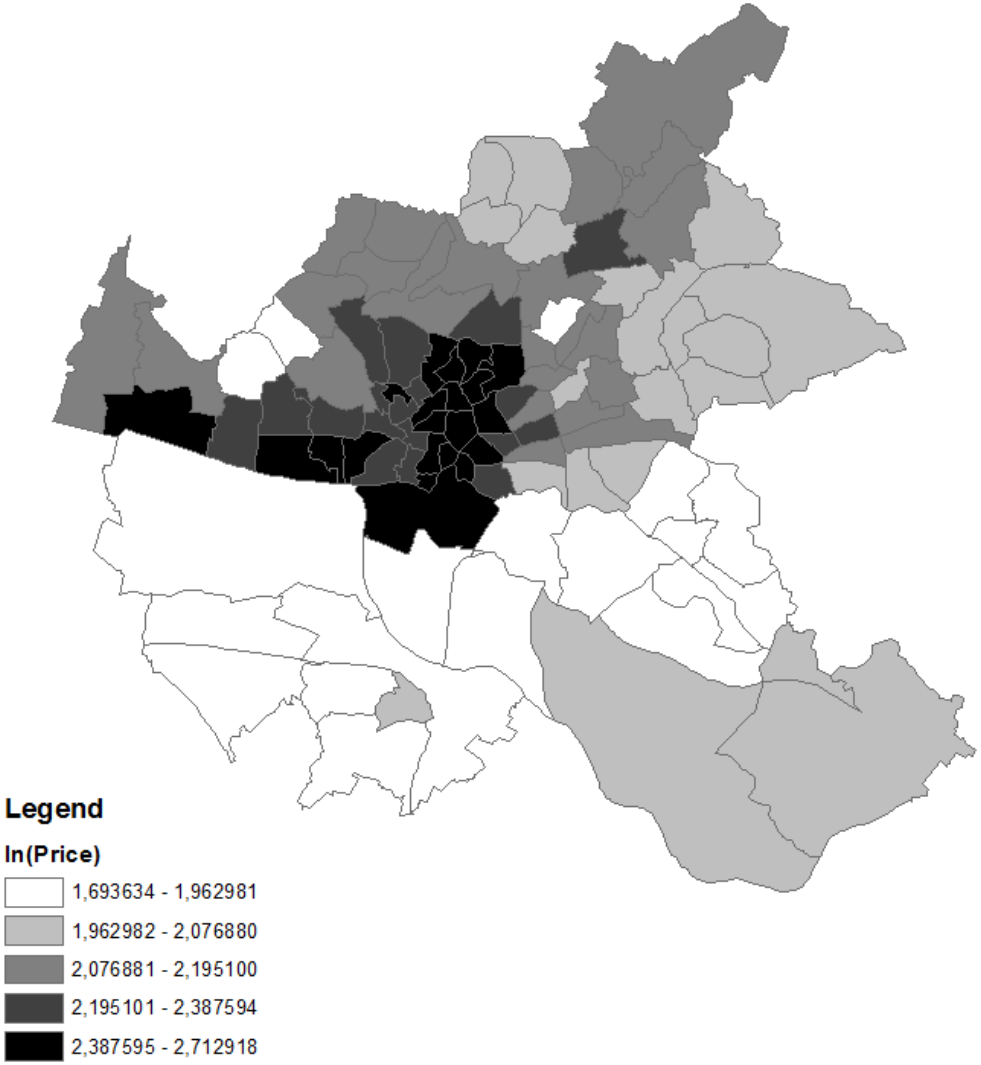
- IDN ImmoDaten GmbH provided by vdpResearch

Scope of the data

- Apartment offering prices between 2008 and 2010 measured in €/m²
- Adjusted for doublings, incomplete observations are removed
- Cross section with 4,029 observations

Group	Price determining variables
Basic attributes	no. of rooms, size in m², age, age², rented
Equipment characteristics	Elevator, Balcony, Fitted Kitchen, Garage, Fireplace, Terrace, Winter Garden, Central Heating
Quality indicator	Attic Flat, Fist Occupancy, Premium, Newly build, Smooth, Refurbished
Location	Postal codes
Year of offer	2008, 2009, 2010

Study Region – Average Price per sqm on post-code level



Diagnostic Dependency Test

Test	H_o	H_1	Procedures	Results
LM-error test	No spatial autocorrelation (λ=0) given the assumption that ρ=0.	Spatial autocorrelation (λ≠0).	If H_o is rejected, estimate a spatial error model.	LM_λ= 1850.14 LM_λ^r= 908.81
LM-lag test	No spatial autocorrelation (ρ=0) given the assumption that λ=0.	Spatial autocorrelation (ρ≠0).	If H_o is rejected, estimate a spatial lag model.	LM_ρ= 1096.15 LM_ρ^r= 154.82
If both null hypotheses of the LM-tests are rejected, perform the robust tests!				
LR-spatial error	Θ+ρβ=0	Θ+ρβ≠0	If H_o is rejected, estimate a spatial Durbin model.	$LR_{\theta+\rho\beta=0}$= 94.84
LR-spatial lag	Θ=0	Θ≠0	If H_o is rejected, estimate a spatial Durbin model.	$LR_{\theta=0}$= 235.21

λ: parameter Spatial Error Model

Determinant	(1)	(2)				
		(a)		(b)	(c)	(d)
	OLS	Spatial Durbin Model				
	Coeff	**Coeff** X	**Coeff** W*X	**Dir effects**	**Indir effects**	**Total effects**
2009	0.0248	0.0330	-0.0864	0.0293	-0.1513	-0.1219
	2.23	3.41	-2.95	3.02	-2.62	-2.03
2010	0.0993	0.1068	-0.1079	0.1041	-0.1062	-0.0011
	7.82	9.41	-3.34	9.09	-1.73	-0.03
No. of rooms	0.0137	0.0126	-0.0182	0.0121	-0.0239	-0.0117
	3.23	3.39	-1.39	3.27	-0.92	-0.43
Total living	0.0743	0.058	0.086	0.063	0.2658	0.3295
Area* 100	7.78	6.98	-1.39	7.44	4.11	4.91
Age	-0.0049	-0.005	0.0034	-0.0051	0.0012	-0.0038
	-15.13	-19.11	6.37	-17.34	0.83	-2.57
Age	0.025	0.023	-0.013	0.023	0.002	0.025
squ*1000	12.20	14.65	-4.05	12.38	0.165	2.43
Rented	-0.0025	-0.0078	0.0045	-0.0076	-0.0005	0.0081
	-0.26	-0.93	0.16	-0.87	-0.009	-0.15
Elevator	0.0173	0.0047	0.0431	0.0071	0.1043	0.1115
	1.41	0.42	1.56	0.64	2.16	2.25
Balcony	0.0470	0.0311	0.030	0.0338	0.1097	0.1435
	4.56	3.41	1.03	3.65	1.86	2.34
Attic	0.006	0.0101	-0.032	0.0087	-0.0609	-0.0522
	0.06	1.07	-0.95	0.92	-0.91	-0.76
Fitted	0.0072	0.0157	0.0046	0.0161	0.0308	0.0468
Kitchen	0.63	1.54	0.14	1.58	0.49	0.71
First	0.1370	0.1199	-0.1024	0.1187	-0.0805	0.0382
Occupancy	11.31	11.05	-3.10	10.96	-1.27	0.58

	OLS	X	W*X	Direct	Indirect	Total
Garage	0.0467	0.0400	-0.0164	0.0399	0.0119	0.0518
	4.61	4.45	-0.62	4.39	0.25	1.07
Garden	0.0063	0.009	0.0513	0.0119	0.1273	0.1393
	0.66	1.13	1.79	1.41	2.42	2.53
Premium	0.1320	0.0858	-0.0075	0.0880	0.0918	0.1798
	11.39	8.13	-0.26	8.52	1.75	3.36
Fireplace	0.0698	0.0615	-0.0231	0.0632	0.0329	0.0962
	2.83	2.85	-0.31	2.91	0.21	0.60
New built	0.1115	0.0942	-0.0067	0.0965	0.1044	0.2011
	8.31	7.73	-0.22	8.07	1.87	3.48
Smooth	0.0009	0.016	-0.0999	0.0114	-0.2009	-0.1894
	0.09	1.84	-4.05	1.32	-4.16	-3.78
Refurbished	0.0842	0.0733	0.0683	0.0784	0.2472	0.3256
	5.93	5.88	1.63	6.33	2.92	3.72
Terrace	0.0306	0.0243	-0.0391	0.0226	-0.0561	-0.0335
	3.04	2.76	-1.28	2.62	-0.99	-0.57
Winter garden	0.0424	0.0342	0.0741	0.0397	0.2009	0.2406
	1.32	1.22	0.84	1.40	0.22	1.37
Central heating	-0.0117	-0.00151	0.0305	0.0019	0.1334	0.1353
	-1.29	-0.18	2.89	0.23	3.50	3.47
ρ		0.5639				
		28.32				
R^2	0.3853	0.6087	0.6427			
R^2_{adj}	0.3820	0.6005	0.6333			
Log-likelihood			316.69			

Spatial Lag, OLS model

LM_ρ	1096.15	p=0
LM_ρ^{robust}	154.82	p=0

Spatial error, OLS model

LM_λ	1850.14	p=0
LM_λ^{robust}	908.81	p=0

Spatial lag, spatial Durbin model

$LR_{\theta=0}$	235.21	p = 0

Spatial error, spatial Durbin model

$LR_{\theta+\rho\beta=0}$	94.84	p = 0

*t-values presented in the second line; test for significant level: LM statistics are based on OLS residuals. LR-statistics are based on log-likelihood values.

Results

OLS-Regression

- Apartment characteristics show the expected sign
- Including postal code dummies increases the explanatory power.

Spatial Durbin Model

Direct effects

- OLS and direct effects show similar significant effects, but the coefficients differ: OLS-regression is biased and over- or underestimate the coefficients of various degrees.
  - For instance, ‘premium’ is overestimated by 50%.

Indirect effects

- Given the estimation result, some apartment characteristics have effects on surrounding apartments
- Seven explanatory apartment characteristics are significant at the 5% significance level.

Three Main Points following the method suggested by LeSage and Pace (2009)

1. LeSage and Pace (2009) show similarity between the direct impact estimates and the response parameters.

   -> response estimates and direct effects only differ in the second or third decimal place.

2. Existence of large discrepancies between the indirect impact and the spatially lagged coefficients in the spatial Durbin model.

   -> for instance ‘garden’: statistically significant indirect effect but statistically insignificant for the spatially lagged variable.

3. Total effects differ from the sum of the response parameters and the spatially lagged coefficients.

   -> for instance ‘garden‘: total impact in the apartment price is positive (0.1393) while the total impact suggested by X and W*X would equal less than half of this magnitude (0.0603).

Conclusion

Hedonic apartment price regressions usually rely on individual characteristics which may exclude important spatially related neighbourhood variables.

Current study improves past hedonic modelling efforts by directly incorporating spatial effects into the apartment hedonic price model.

Main Results

- Direct and indirect parameters show significant results
- Direct effects and the response parameters are similar
- Differences between the spatially lagged and the indirect effects

Reference

Anselin, Luc. "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity." *Geographical analysis* 20.1 (1988): 1-17.

Anselin, Luc, and Anil K. Bera. "Introduction to spatial econometrics." *Handbook of applied economic statistics* 237 (1998).

Anselin, Luc, Julie Le Gallo, and Hubert Jayet. "Spatial panel econometrics." *The econometrics of panel data*. Springer, Berlin, Heidelberg, 2008. 625-660.

Baumont, Catherine, Cem Ertur, and Julie Le Gallo. "Spatial analysis of employment and population density: the case of the agglomeration of Dijon 1999." *Geographical analysis* 36.2 (2004): 146-176.

LeSage, James, and Robert Kelley Pace. *Introduction to spatial econometrics*. Chapman and Hall/CRC, 2009.

Moran, Patrick AP. "The interpretation of statistical maps." *Journal of the Royal Statistical Society. Series B (Methodological)* 10.2 (1948): 243-251.