SFB
823

Separate estimation of spatial dependence parameters and variance parameters in a spatial model

Matthias Arnold, Dominik Wied

Discussion Paper

Nr. 8/2010

SFB
823

Separate estimation of spatial dependence parameters and variance parameters in a spatial model

Matthias Arnold* and Dominik Wied

Institut für Wirtschafts- und Sozialstatistik, TU Dortmund, Dortmund, Germany

March 31, 2010

Abstract

This paper suggests a two step estimation procedure for a spatial model with different kinds of spatial dependence and heteroscedastic innovations. Since maximum likelihood estimation is cumbersome due to the large number of parameters, we use a generalized method of moments approach to estimate the parameters of spatial correlation which does not need the large number of variance parameters to be known. For illustration purposes, we apply our estimation procedure to daily stock returns of the Euro Stoxx 50 members.

Keywords. spatial dependence, heteroscedasticity, GMM estimation, stock returns

JEL subject classifications: C13, C51, G12.

*Corresponding author: Matthias Arnold, Institut für Wirtschafts- und Sozialstatistik, Fakultät Statistik, TU Dortmund, 44221 Dortmund, Germany, email: arnold@statistik.tu-dortmund.de, Phone: +49/231/755 5203, Fax: +49/231/755 5284.

I. Introduction and Summary

Spatial modeling of dependence structures has become very popular over the last years. In many applications, it seems natural that observations at one location depend on neighboring observations at nearby locations. This phenomenon appears in different contexts like mining, agriculture, ecology or epidemiology, see e.g. Anselin (1988), Cressie (1991) and LeSage and Pace (2009) and the references therein. In some situations, there may be different kinds of spatial dependence in the data. A typical example are origin-destination flow models. LeSage and Pace (2008) distinguish between origin, destination and origin-to-destination dependence. Models of this kind can be used to analyze e.g. German journey-to-work data (Griffith (2009)), inter-provincial migration in Poland (Sarra and Signore (2010)) or Dutch museum visitor behavior (de Graaff et al. (2009)).

In this paper, we consider a model, where three different kinds of dependence may arise. In addition, we allow for heteroscedasticity. Thus, the number of unknown parameters to be estimated is large so that maximum likelihood estimation may be challenging. We suggest a two stage estimation procedure which can be easily implemented. First, we estimate the three parameters of spatial dependence by GMM in a way similar to Kelejian and Prucha (1999) and Kapoor et al. (2007). Here, we circumvent the large number of variance parameters by constructing the GMM estimator in such a way that the typically unknown variance parameters are not needed. In a second step, given the GMM estimates of the spatial dependence parameters, estimation of the variance parameters is straightforward.

As an illustration we apply our estimation procedure to daily stock returns of the 50 members of Dow Jones Euro Stoxx 50 for the period 2003-2009. So far, spatial modeling is not very popular in financial applications like stock returns. The reason for this might be that spatial models require some kind of distance measure between different locations in order to determine which locations should be considered as neighbors. Physical distance between the head offices does not seem to be a reasonable choice: Why should the stocks of Siemens and Allianz, both located in Munich, perform more homogenous than the stocks of Bayer and BASF, where the physical distance between the head quarters is larger? Consequently, up to now the literature concentrates on information spillovers where proximity to innovation clusters or patent activity plays an important role, see e.g. Boasson and MacPherson (2001) or Boasson et al. (2005). Stock performance is then used as a measure for economic success.

We suggest a more general form of spatial dependence for stock returns where we distinguish between three different kinds of spatial dependence. The first one is a general dependence which affects all stocks in the same way. The second one is global in nature and applies to firms that belong to the same branch: Since global input factors like commodity prices should have a similar effect on firms belonging to the same branch, the corresponding stock returns should display a similar behavior. The third one is a local form of dependence: Firms that are located in the same country should display similar behavior because they are exposed to the same surrounding conditions like regulatory frameworks or the business cycle in that country. In this model, we can compare the different spatial dependencies to each other. The innovation terms are allowed to be heteroscedastic. We restrict ourselves to the Euro Stoxx 50 in order to avoid the effect of exchange rate fluctuations.

II. Two Step Estimation Procedure

For $t = 1, \ldots, T$, let y_t be an n-dimensional random vector. We assume independence over time so that y_t is independent of y_s for $s \neq t$. In the cross-sectional dimension, the components of y_t are assumed to be spatially correlated where we allow for three different kinds of spatial dependence:

$$y_t = \rho_1 W_1 y_t + \rho_2 W_2 y_t + \rho_3 W_3 y_t + \varepsilon_t. \tag{1}$$

The spatial weight matrices W_1, W_2 and W_3 are known; the elements on the main diagonals are zero and the matrices are row-standardized. The elements of ε_t are not correlated, but they may be heteroscedastic, i.e., $\mathrm{Cov}(\varepsilon_t) = diag\{\sigma_1^2, \ldots, , \sigma_n^2\} =: \Sigma$. We assume that $\mathrm{E}(y_t) = 0$ which is suitable for our application; generalizations to cases where the expectation depends on explanatory variables are straightforward since the spatial correlation structure (1) could then be applied to the disturbances of the corresponding regression model. If the inverse of the matrix $(I_n - \rho_1 W_1 - \rho_2 W_2 - \rho_3 W_3)$ exists, our model leads to

$$\begin{aligned} \mathrm{Cov}(y_t) &= (I_n - \rho_1 W_1 - \rho_2 W_2 - \rho_3 W_3)^{-1} \Sigma \left(I_n - \rho_1 W_1^T - \rho_2 W_2^T - \rho_3 W_3^T\right)^{-1} \\ &=: V, \end{aligned}$$

where A^T denotes the transpose of a matrix A. Of course, the parameters could be estimated by way of maximum likelihood. Assuming normality and independence over time, the likelihood function would be

$$L(\rho_1, \rho_2, \rho_3, \Sigma) = (2\pi)^{-\frac{nT}{2}} (\det V)^{-\frac{T}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} y_t^T V^{-1} y_t\right).$$

Altogether, our model contains $n+3$ parameters, the three correlation parameters ρ_1, ρ_2 and ρ_3 and n parameters of variance, σ_i^2. Thus, the calculation of the maximum likelihood estimates can be computationally expensive, especially if n is large.

As an alternative, we suggest a two step estimation procedure which is easy to compute. First, we estimate the correlation parameters by generalized method of moments along the lines of Kelejian and Prucha (1999) or Kapoor et al. (2007). We will show that this step does not depend on the parameters of variance. Second, given the estimated correlation parameters it is straightforward to estimate the variance parameters.

The GMM estimator for the correlation parameters uses the following three moment conditions:

$$\begin{aligned} \mathrm{E}\left(\varepsilon_t^T W_1 \varepsilon_t\right) &= \mathrm{tr}(W_1\Sigma) = 0, \\ \mathrm{E}\left(\varepsilon_t^T W_2 \varepsilon_t\right) &= \mathrm{tr}(W_2\Sigma) = 0, \\ \mathrm{E}\left(\varepsilon_t^T W_3 \varepsilon_t\right) &= \mathrm{tr}(W_3\Sigma) = 0. \end{aligned}$$

Replacing ε_t by

$$\varepsilon_t = (I_n - \rho_1 W_1 - \rho_2 W_2 - \rho_3 W_3)\, y_t$$

and averaging over t gives the theoretical system of equations

$$\Gamma\theta + \gamma = 0,$$

where

$$\theta := \left(\rho_1, \rho_2, \rho_3, \rho_1^2, \rho_2^2, \rho_3^2, \rho_1\rho_2, \rho_1\rho_3, \rho_2\rho_3\right)^T$$

and for $i, j \in \{1, 2, 3\}$, the elements of $\Gamma \sim (3 \times 9)$ and $\gamma \sim (3 \times 1)$ are defined by

$$\begin{aligned}
\Gamma(ij) &= \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T} y_t^T \left(W_i + W_i^T\right) W_j y_t\right), \\
\Gamma(i, 3+j) &= \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T} y_t^T W_j^T W_i W_j y_t\right), \\
\Gamma(i, 7) &= \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T} y_t^T W_1^T \left(W_i + W_i^T\right) W_2 y_t\right), \\
\Gamma(i, 8) &= \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T} y_t^T W_1^T \left(W_i + W_i^T\right) W_3 y_t\right), \\
\Gamma(i, 9) &= \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T} y_t^T W_2^T \left(W_i + W_i^T\right) W_3 y_t\right), \\
\gamma_i &= \mathrm{E}\left(\frac{1}{T}\sum_{t=1}^{T} y_t^T W_i y_t\right).
\end{aligned}$$

Let G and g be the empirical counterparts of Γ and γ, i.e., for $i \in \{1, 2, 3\}$, $j \in \{1, \ldots, 9\}$, $G(ij)$ and g_i are given by $\Gamma(ij)$ and γ_i with the expectation operator left out, respectively. The GMM estimator for ρ_1, ρ_2 and ρ_3 is defined as

$$(\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3)^T_{GMM} := \arg \min_{\rho_1 \in [-1,1], \rho_2 \in [-1,1], \rho_3 \in [-1,1]} ||G\theta + g||.$$

The theoretical term $\Gamma\theta + \gamma$ is equal to zero for the true parameter values. Our GMM estimator is calculated by finding the values for ρ_1, ρ_2 and ρ_3 for which the corresponding empirical system $G\theta + g$ is closest to zero. Compared to the ML estimator of the model parameters, this GMM estimator is easy to calculate: We just have to minimize $||G\theta + g||$ with respect to ρ_1, ρ_2 and ρ_3. Even for large n, this is easy to handle. In particular, the parameters of variance σ_i^2 are not needed to calculate the GMM estimators for the correlation parameters. This GMM estimator is consistent for $T \to \infty$ as long as the theoretical system of equations has a unique solution in ρ_1, ρ_2, and ρ_3. The proof is a straightforward extension of Kelejian and Prucha's proof.

Given the estimates for the correlation parameters, estimation of the parameters of variance in the second step is straightforward: We just take the averages over the estimated $\hat{\varepsilon}_{i,t}^2$:

$$\hat{\sigma}_i^2 := \frac{1}{T}\sum_{t=1}^{T} \hat{\varepsilon}_{i,t}^2 := \frac{1}{T}\sum_{t=1}^{T} \left[\left(I_n - \hat{\rho}_1 W_1 - \hat{\rho}_2 W_2 - \hat{\rho}_3 W_3 \right) y_t \right]_i^2 .$$

For $T \to \infty$ and $i = 1, \ldots, n$, the estimator $\hat{\sigma}_i^2$ is consistent for σ_i^2 by the law of large numbers as long as the y_t have finite absolute $(4 + \delta) - th$ moments for some $\delta > 0$.

III. Monte Carlo Simulation

We investigate the finite sample properties of our estimation procedure by Monte Carlo simulation studies. We choose $n = 50$ and construct the adjacency matrices in the following way. For W_1, all off-diagonal elements are equal to 1/49 so that each observation is affected by every other observation. The second matrix W_2 reflects a dependence between blocks of five observations each so that e.g. observation 17 depends on observations 16, 18, 19 and 20. The corresponding non-zero elements of W_2 are all equal to 1/4. Finally, for W_3 the first 25 observations depend on each other as well as the last 25 observations so that the non-zero elements of W_3 are equal to 1/24. We consider three different sample sizes T = 100, 500, 2000 and three different settings for the spatial correlations parameters. The first one reflects small correlations ($\rho_1 = \rho_2 = \rho_3 = 0.1$), the second one large correlations ($\rho_1 = \rho_2 = \rho_3 = 0.3$) and the third one different amounts of correlation ($\rho_1 = 0.1$, $\rho_2 = 0.3$, $\rho_3 = 0.5$). For the innovation terms ε_i we consider homoscedasticity ($\sigma_i^2 = 1$) as well as heteroscedasticity ($\sigma_i^2 = i$). For each combination of T, correlation structure and variance structure, we generate 10000 replications of the data and compute simulated biases and MSEs of the correlation parameter estimators as well as the sum of the biases and the sum of the relative MSEs of the $n = 50$ variance parameters, e.g. the MSEs are divided by the true variance parameters before summation.

Table 1 gives the simulated biases. In general, the biases are small. For settings with different amounts of dependence ($\rho_1 = 0.1$, $\rho_2 = 0.3$, $\rho_3 = 0.5$), $\hat{\rho}_1$ seems to be

Table 1: Simulated biases of the estimators for $n = 50$, 10000 repetitions each

T	ρ_1	ρ_2	ρ_3	σ_i^2	$\hat{\rho_1}$	$\hat{\rho_2}$	$\hat{\rho_3}$	$\hat{\sigma}_i^2$
100	0.1	0.1	0.1	1	-0.00026	-0.00030	-0.00470	-0.03266
100	0.3	0.3	0.3	1	0.00228	-0.00048	-0.00273	-0.01119
100	0.1	0.3	0.5	1	0.01520	-0.00047	-0.00052	0.00189
500	0.1	0.1	0.1	1	0.00042	-0.00018	-0.00010	-0.00954
500	0.3	0.3	0.3	1	0.00030	-0.00014	-0.00029	-0.00576
500	0.1	0.3	0.5	1	0.00610	0.00001	-0.00019	0.01114
2000	0.1	0.1	0.1	1	-0.00013	0.00001	-0.00022	-0.00176
2000	0.3	0.3	0.3	1	0.00026	0.00004	-0.00031	-0.00056
2000	0.1	0.3	0.5	1	0.00050	0.00003	-0.00011	-0.00008
100	0.1	0.1	0.1	i	0.00004	-0.00059	-0.00566	-1.04455
100	0.3	0.3	0.3	i	0.00211	-0.00055	-0.00229	-0.06369
100	0.1	0.3	0.5	i	0.01436	-0.00041	-0.00064	0.25778
500	0.1	0.1	0.1	i	0.00031	-0.00014	-0.00117	-0.17334
500	0.3	0.3	0.3	i	0.00037	-0.00004	-0.00056	0.05469
500	0.1	0.3	0.5	i	0.00670	-0.00000	-0.00019	0.41660
2000	0.1	0.1	0.1	i	0.00001	0.00006	-0.00022	-0.06643
2000	0.3	0.3	0.3	i	0.00021	0.00002	-0.00025	-0.05600
2000	0.1	0.3	0.5	i	0.00066	0.00004	-0.00008	-0.04628

Table 2: Simulated MSE of the estimators for $n = 50$, 10000 repetitions each

T	ρ_1	ρ_2	ρ_3	σ_i^2	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	$\hat{\sigma}_i^2$
100	0.1	0.1	0.1	1	0.00575	0.00041	0.00362	1.00283
100	0.3	0.3	0.3	1	0.00095	0.00028	0.00116	1.00195
100	0.1	0.3	0.5	1	0.00298	0.00028	0.00054	1.00296
500	0.1	0.1	0.1	1	0.00108	0.00008	0.00069	0.20079
500	0.3	0.3	0.3	1	0.00017	0.00005	0.00021	0.19985
500	0.1	0.3	0.5	1	0.00125	0.00006	0.00010	0.20085
2000	0.1	0.1	0.1	1	0.00028	0.00002	0.00017	0.04997
2000	0.3	0.3	0.3	1	0.00042	0.00001	0.00005	0.04995
2000	0.1	0.3	0.5	1	0.00009	0.00001	0.00002	0.05000
100	0.1	0.1	0.1	i	0.00439	0.00054	0.00366	25.48836
100	0.3	0.3	0.3	i	0.00096	0.00036	0.00130	25.51620
100	0.1	0.3	0.5	i	0.00275	0.00036	0.00056	26.32263
500	0.1	0.1	0.1	i	0.00083	0.00011	0.00071	5.10533
500	0.3	0.3	0.3	i	0.00015	0.00007	0.00022	5.11178
500	0.1	0.3	0.5	i	0.00137	0.00007	0.00011	5.45464
2000	0.1	0.1	0.1	i	0.00021	0.00003	0.00018	1.27290
2000	0.3	0.3	0.3	i	0.00004	0.00002	0.00005	1.27235
2000	0.1	0.3	0.5	i	0.00013	0.00002	0.00003	1.30723

slightly upward biased. For the other settings, the simulated biases do not show a clear pattern. There are positive as well as negative signs so that the true biases seem to be close to zero. The variance structure of the innovations does not seem to influence our results. This fits to the fact that the variance parameters are not needed to calculate the spatial dependence parameters.

Table 2 shows the results for the MSEs. In general, the MSEs of the correlation parameters are very small even for moderate T. This is true for homoscedastic as well as heteroscedastic innovations. In return, the sum of the MSEs of the variance parameters is clearly robust against different correlation settings. As T increases, all MSEs seem to decrease by order $1/T$. Comparing the different correlation parameters to each other we conclude that MSEs decrease when the true parameter increases. Furthermore, MSEs are smallest for $\hat{\rho}_2$ which captures a dependence of only 5 observations each.

IV. Application to stock returns

We analyze the spatial dependencies in the daily stock returns of the Euro Stoxx 50 members in the composition of January 2010 for the period from 2003 until 2009. The data we use are adjusted stock prices from Datastream which we transfer to log returns. Our basic model for the stock returns on day t, $t = 1, \ldots, T$, is

$$y_t = \rho_g W_g y_t + \rho_b W_b y_t + \rho_l W_l y_t + \varepsilon_t,$$

where y_t is the vector of stock returns on day t, the weight matrices W_g, W_b and W_l capture general dependencies, dependencies inside branches and local dependencies and the unknown parameters ρ_g, ρ_b and ρ_l represent the amount of the three kinds of dependencies, respectively. Our main interest is to distinguish between dependencies inside branches and local dependence. In addition, we introduce a third kind of dependence called general dependence to capture impacts which effect all stocks in a similar way like prior performances of stock markets in the USA or Asia. If we did not do that, the dependencies of main interest could be superposed by this general dependence.

Table 3 shows the partitioning of the Euro Stoxx 50 members into branches and countries. Nokia and CRH are the only representatives of their home countries, respectively, but in order to avoid singularities, groups must not consist of only one member. We consider two different groupings. In model 1, we impose a group called „others“ for Finland and Ireland, where only one company is part of the Euro Stoxx 50, respectively. In model 2, we put Nokia and CRH to the Benelux group which would then be labeled „small countries“. According to these groupings, the adjacency matrices are constructed in the following way. The off-diagonal elements of the general adjacency matrix W_g are $1/(n-1)$. In W_b and W_l, the element in the

Table 3: Partitioning of Euro Stoxx 50 members into branches and countries in model 1; groups „Benelux“ and „others“ are merged to the new group „small countries“ in model 2

Finance	Aegon, Allianz, AXA, Banco Bilbao, Banco Santander, BNP, Crédit Agricole, Deutsche Bank, Deutsche Börse, Generali, ING, Intesa, Münchener Rück, Société Générale, Unicredit
Automobil	Daimler, Renault, VW
Energy	Alstom, E.ON, ENEL, ENI, Iberdrola, Repsol, RWE, SUEZ, Total
Telecom and Media	Deutsche Telekom, France Telecom, Telecom Italia, Telefonica, Vivendi
Pharma and Chemicals	Air Liquide, BASF, Bayer, Sanofi
Construction	Vinci, Saint-Gobain
Consumer Electronics	Nokia, Philips, SAP, Siemens, Schneider
Consumer retail	Anheuser Busch, Carrefour, Danone, L'Oreal, LVMH, Unilever
Basic Industry	Arcelor Mittal, CRH, Saint Gobain, Vinci
Benelux	Aegon, Anheuser Busch, Arcelor, ING, Philips, Unilever
France	Air Liquide, Alstom, AXA, BNP, Carrefour, Crédit Agricole, France Telecom, Danone, L'Oreal, LVMH, Saint Gobain, Sanofi, Schneider, Société Générale, SUEZ, Total, Vinci, Vivendi
Germany	Allianz, BASF, Bayer, Daimler, Deutsche Bank, Deutsche Börse, Deutsche Telekom, E.ON, Münchner Rück, RWE, SAP, Siemens, VW
Italy	Generali, ENEL, ENI, Intesa, Telecom Italia, Unicredito
Spain	Banco Bilbao, Banco Santander, Iberdrola, Repsol, Telefonica
Others	CRH, Nokia

Table 4: Spatial dependencies in Euro Stoxx 50 stock returns

	model 1			model 2		
period	$\hat{\rho}_g$	$\hat{\rho}_b$	$\hat{\rho}_l$	$\hat{\rho}_g$	$\hat{\rho}_b$	$\hat{\rho}_l$
2003-2009	0.544	0.192	0.101	0.504	0.190	0.143
2003	0.440	0.219	0.170	0.381	0.215	0.233
2004	0.579	0.148	0.058	0.509	0.144	0.132
2005	0.571	0.159	0.066	0.515	0.155	0.126
2006	0.538	0.201	0.092	0.522	0.198	0.111
2007	0.459	0.231	0.143	0.374	0.230	0.231
2008	0.637	0.129	0.081	0.609	0.128	0.112
2009	0.418	0.332	0.095	0.396	0.331	0.118

i^{th} row and j^{th} column is nonzero if the corresponding stocks belong to the same branch (W_b) or country (W_l). In each row, the nonzero entries are identical. The matrix W_g is the same in both models as well as W_b, whereas W_l is different between both models. We estimate the dependence parameters ρ_g, ρ_b and ρ_l on the whole data set as well as on subsamples which contain the daily returns of only one year, respectively.

Table 4 shows the results. For each year as well as for the whole data set, general dependence is the largest in both models. In model 1, where we have the local group „others", dependence inside branches is about twice as large as local dependence. In model 2, where all companies of small countries are put together in one group, local dependence increases by about 0.05, whereas general dependence decreases correspondingly. Dependence inside branches is practically the same for both models. We conclude that in model 2, local dependence captures general dependence to some extent because we invented a new group of small countries which are not really locally connected. Consequently, we prefer model 1 to model 2 because it seems to capture the different kinds of dependence more accurately. It is quite interesting to see that for model 1 in 2009, $\hat{\rho}_b$ rises to 0.332, whereas $\hat{\rho}_l$ is only 0.095. This could mean that recently, dependencies inside branches became more important than local dependencies.

Acknowledgement: The financial support by the Deutsche Forschungsgemeinschaft (SFB 823, Statistik nichtlinearer dynamischer Prozesse) is gratefully acknowledged.

References

L. Anselin. *Spatial Econometrics: Methods and Models.* Kluwer, Dordrecht, Boston and London, 1988.

V. Boasson, E. Boasson, and A. MacPherson. Firm value and geographic competitive advantage: Evidence from the US pharmaceutical industry. *Journal of Business*, 78(6):2465–2495, 2005.

V. Boasson and A. MacPherson. The role of geographic location in the financial and innovation performance of publicly traded pharmaceutical companies: empirical evidence from the United States. *Environment and Planning A*, 33(8):1431–1444, 2001.

N. A. C. Cressie. *Statistics for Spatial Data.* John Wiley & Sons, New York, 1991.

T. de Graaff, J. Boter, and J. Rouwendal. On spatial differences in the attractiveness of Dutch museums. *Environment and Planning A*, 41(11):2778–2797, 2009.

D. A. Griffith. Modeling spatial autocorrelation in spatial interaction data: empirical evidence from 2002 Germany journey-to-work flows. *Journal of Geographical Systems*, 11:117–140, 2009.

M. Kapoor, H. H. Kelejian, and I. R. Prucha. Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1):97–130, 2007.

H. H. Kelejian and I. R. Prucha. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, 40: 509–533, 1999.

J. LeSage and R. K. Pace. *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, London and New York, 2009.

J. P. LeSage and R. K. Pace. Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48(5):941–967, 2008.

A.L. Sarra and M. Del Signore. A Dynamic Origin-constrained Spatial Interaction Model Applied to Poland's Inter-provincial Migration. *Spatial Economic Analysis*, 5(1):29–41, 2010.