Modeling Spatial Externalities: A Panel Data Approach

Christian Beer[*] and Aleksandra Riedl[†]

May 29, 2009

First draft, do not quote!

Abstract

In this paper we argue that the Spatial Durbin Model (SDM) is an appropriate framework to empirically quantify different kinds of externalities. Besides, it is also attractive from an econometric point of view as it nests several other models frequently employed. Up to now the SDM was applied in cross-sectional settings only, thereby ignoring individual heterogeneity. This paper extends the SDM to panel data allowing for non-spherical disturbances and proposes an estimator based on ML techniques. Results from a Monte Carlo study reveal that the estimator has satisfactory small sample properties and that neglecting the non-spherical nature of the errors leads to inflated standard errors. Moreover, we show that the incidence of type two errors in testing procedures for parameter significance of spatially lagged variables is the higher the denser the spatial weight matrix.

Keywords: Spatial panel data, Spatial Durbin Model, Maximum Likelihood, AR(1) and heteroskedastic errors, Monte Carlo simulation

JEL classification: C21, C23

1 Introduction

Externalities play a central role in various fields of economics. They often can be regarded as spatial phenomena as activities have to spread over space in order to affect the well-being of others. For example, homes that are surrounded by houses with beautiful gardens exhibit positive externalities. These externalities mainly affect direct neighbors and become less effective for distant houses. Further examples for spatial externalities include, amongst others, knowledge spillovers affecting economic growth, fiscal externalities relevant in the tax competition literature or pollution issues in environmental economics. Due to the increased amount of research in spatial econometrics more elaborate tools for analyzing such phenomena empirically have become available. So far,

[*]Institute for Economic Geography and GIScience, Vienna University of Economics and Business, E-Mail: christian.beer@wu-wien.ac.at

[†]Institute for Economic Geography and GIScience, Vienna University of Economics and Business, E-Mail: aleksandra.riedl@wu-wien.ac.at

the theoretical development of spatial econometric models[1] and their applications have mainly focused on modeling spatial dependence via two different sources: spatial correlation between non-observable explanatory variables (disturbance term), and spillover effects between the observations of the dependent variable. In contrast, spatial externalities working through the explanatory variables have received far less attention.

Only recently, the Spatial Durbin Model (SDM) (see Anselin, 1988) is gaining popularity in the economics literature as it includes a spatial lag on the dependent and independent variables and is thus suitable to capture externalities and spillovers arising from different sources. Applications in the field of regional science include Brasington and Hite (2005) who show that the SDM is the preferred model specification for explaining housing prices in a cross-sectional setting. Specifically, the spatial lag on the dependent variable controls for the fact that offer prices are often set with the knowledge of the selling prices of similar houses in the neighborhood, whereas the spatial lag on the explanatory variables, like school quality of neighboring entities, captures positive spillover effects arising through e.g. peer group effects. An interesting application of the SDM at the inter-country level can be found in Ertur and Koch (2007) who develop a spatially augmented Solow model assuming technological interdependence and physical capital externalities among economies. In their model, the SDM arises as the empirical representation of the derived steady state and convergence equations. Based on a sample of 91 countries they show that the nonlinear transformations of the SDM parameter estimates are in line with their assumptions concerning the existence of both, physical capital externalities arising through knowledge spillovers, and global technological interdependence among economies.

Besides its capability to reflect interdependencies originating from a variety of economic channels, the SDM also occupies an interesting position from an econometric point of view. Since it nests the models that include a spatial lag on the dependent variable or on the disturbances, it may be applied for model selection purposes. In this context, the common factor test (Burridge, 1981) is of particular importance as it permits discrimination between the SDM and the model with spatially lagged disturbances, and hence may indicate whether spatial externalities are substantive phenomena or rather random shocks diffusing through space.[2] Another important feature of the SDM that should be mentioned at this point is its performance in the case of spatially dependent omitted variables. As was shown in LeSage and Pace (2009, p.60), applying the SDM may mitigate the bias relative to OLS estimates when unobservable factors like location amenities

[1]Theoretical contributions include Baltagi et al. (2003), Conley (1999), Das et al. (2003), Driscoll and Kraay (1998), Kapoor et al. (2007), Kelejian and Prucha (1998), Kelejian and Prucha (1999), Lee (2004), Lee and Yu (2008) and LeSage and Pace (2009). For a detailed literature review on theoretical and empirical applications of spatial econometric models see Anselin et al. (2004).

[2]See also Mur and Angulo (2006) for alternative tests and a comparison based on Monte Carlo evidence.

or neighborhood prestige exert an influence on the dependent variable. This provides a strong econometric motivation for employing the SDM in applied work when modeling phenomena that are located in space.[3]

To our knowledge, the Spatial Durbin Model has been solely applied to cross-sectional data, thereby ignoring individual heterogeneity. The objective of this paper is to extend the SDM to panel data and to provide an adequate estimator within a fixed effects setting. To account for the non-spherical nature of disturbances often encountered when applying panel data models, we additionally allow for an AR(1) and heteroskedastic error structure. A distinctive feature of the SDM is that it cannot be estimated via Two-stage Least squares, as proposed by e.g. Kelejian and Prucha (1998), because the available instruments already enter as regressors in the model. We therefore propose an ML estimator, which involves a data transformation proposed by Lee and Yu (2008) to avoid the incidental parameter problem. Unlike often argued in the spatial econometrics literature, Maximum Likelihood estimation is also feasible for very large samples[4] due to many improvements in computational methods and computing power. Hence, the applied estimation routine for the proposed estimator remains feasible even when the number of time periods becomes large.[5]

To investigate the finite sample properties of the estimator we perform a Monte Carlo study. Results reveal that coefficient estimates are virtually unbiased for reasonable large samples and that ignoring serial correlation and heteroskedasticity inflates standard errors of the estimates. Moreover, we show by means of simulation that the occurrence of type 2 errors in the usual significance tests for parameter estimates of the spatially lagged variables crucially depends on the density of the spatial weight matrix. The more neighboring entities are exhibiting an influence on the dependent variable, the more often the null hypothesis of no influence of the parameter estimates is wrongly accepted. This points to the need to further investigate the properties of the power function for models with spatially lagged explanatory variables.

In the following section we extend the SDM to the panel case. The estimation approach is then presented in section 3. In section 4 we describe the Monte Carlo design, and present the results in section 5. Section 6 summarizes the main results and discusses several interesting and useful extensions for future research.

[3]An extensive analysis of the SDM for cross-section data and its properties are provided in LeSage and Pace (2009).

[4]Models involving samples of more than 60,000 cross-section observations can be estimated within a few seconds on desktop and laptop computers (see e.g. LeSage and Pace, 2009).

[5]The Matlab code is available from the authors upon request.

2 A Spatial Durbin Model for Panel Data

The Spatial Durbin Model (Anselin, 1988) is a modification of a model originally developed by Durbin (1960) in the context of time series analysis. In its spatial version, it is the unrestricted reduced form of a model with cross-sectional dependence in the errors and appears as the nesting model in a more general approach of model selection. Let w denote an $N \times N$ spatial weight matrix and y and ϵ vectors of dimension N including the dependent variable and the error term, respectively. Moreover, let X be an $N \times k$ matrix of independent variables. Then the cross-sectional SDM can be written in the following way:

$$y = \rho \mathrm{w} y + X\beta + \mathrm{w} X \gamma + \epsilon \quad (1)$$

$$\epsilon \sim N(0,1) \quad (2)$$

By imposing certain restrictions it turns out that several other spatial models can be regarded as special cases of the SDM. In particular, imposing the restriction $\gamma = 0$ leaves us with the spatial lag model, also referred to as the spatial autoregressive model (Cliff and Ord, 1981). If the restriction $\gamma = -\rho\beta$ is imposed the SDM reduces to the frequently applied spatial error model, where the spatial structure enters through the error term solely (Cliff and Ord, 1981). To discriminate between the SDM and the spatial error model the common factor test can be applied. For a discussion on testing procedures and their properties see e.g. Burridge (1981) and Mur and Angulo (2006). Lastly, if $\rho = 0$ an OLS type regression arises that includes a spatial lag on the regressors.

As in the cross-sectional setting, the extension of the SDM to the panel case may serve as a model selection framework for various spatial panel data models frequently applied in the literature. Additionally, the usual advantages of panel data can be exploited, including a higher sample variability, an increase in the degrees of freedom, more accurate inference and the possibility to control for the impact of time-constant omitted variables. To extend the model to the panel case we order the dependent variable as $y = (y_{11} \dots y_{1T}, \dots, y_{N1} \dots y_{NT})'$, where the slower index denotes the cross-sectional units $i = 1, \dots, N$ and the faster index refers to the time dimension $t = 1, \dots, T$. The model can then be written as follows:

$$
\begin{aligned}
y &= \rho W y + X\beta + WX\gamma + Z\mu_1 + WZ\mu_2 + \epsilon &(3)\\
\epsilon &\sim N(0, \sigma^2\Omega) &(4)\\
\Omega &= \Sigma_N \otimes \omega_T &(5)
\end{aligned}
$$

Assuming that the neighborhood relationship does not change over time, we can write the weight matrix as $W = \mathrm{w} \otimes I_T$[6]. To control for unobservable individual-specific effects each cross-section is assumed to have a time-constant term, modeled as a fixed parameter. These individual effects are collected in $Z = I_N \otimes \iota_T$, where ι_T is a vector of order T containing ones with μ_1 being the corresponding parameter vector. In contrast to random effects error components the fixed effects specification has the advantage of robustness as the fixed effects are allowed to be correlated with the regressors in the model. Since in the SDM each regressor enters in its spatial lag form, we also include WZ in our specification. As a generalization, the error term is assumed to follow an AR(1) process with ϕ being the autoregressive parameter, where serial correlation is accounted for in the matrix ω_T, where

$$
\omega_T = \frac{1}{1-\phi^2}\begin{pmatrix} 1 & \phi & \phi^2 & \dots & \phi^{T-1} \\ \phi & 1 & \phi & \dots & \phi^{T-2} \\ \phi^2 & \phi & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \phi \\ \phi^{T-1} & \phi^{T-2} & \dots & \phi & 1 \end{pmatrix} \quad (6)
$$

Moreover, the cross-sectional units do not necessarily exhibit the same variance. Specifically, we allow for heteroskedasticity by assuming that $\Sigma_N = \mathrm{diag}\left\langle \sigma_1^2, \dots, \sigma_N^2 \right\rangle$. Finally, it turns out that the variance-covariance matrix Ω is block-diagonal and has the form $\Omega = \mathrm{diag}\left\langle \sigma_1^2\omega_T \dots, \sigma_N^2\omega_T \right\rangle$.

3 Estimation Approach

In a frequentist framework, the SDM in its cross-sectional form has been estimated exclusively by implementing Maximum Likelihood techniques, because the Two-stage Least Squares estimator proposed by e.g. Kelejian and Prucha (1998) cannot be applied in this kind of setting. Specifically, the spatial lag on the dependent variable wy may no longer be instrumented by wX as the variables in wX themselves are part of the data generating process of y (see equation (1)). This distinctive

[6]In the following I_k denotes an identity matrix of dimension k.

feature of the SDM also applies to its panel version, so that we proceed to derive a ML estimator.

It is well known that in a fixed effects setting the number of parameters to be estimated increases with the number of cross-sectional units. This may lead to the incidental parameter problem and to an inconsistent estimator of the variance in a ML environment (see Neyman and Scott, 1948). In order to avoid these problems we apply a data transformation proposed by Lee and Yu (2008), which eliminates variables that are constant over time. Unlike the demeaning matrix proposed by e.g. Baltagi (2005) that subtracts the time mean of each observation, the transformation proposed by Lee and Yu (2008), which can be regarded as a generalization of the Helmert transformation, results in a variance-covariance matrix of full rank.

Before applying the above transformation to our model we first adopt a Cochrane-Orcutt (1949) procedure to deal with serial correlation in the disturbances. Contrary to the Prais-Winsten (1954) transformation, the Cochrane-Orcutt procedure leaves the fixed-effects constant over time, such that the Lee and Yu (2008) transformation can be effectively applied in a second step.[7] For this purpose, we construct the block diagonal matrix $P^{-1} = I_N \otimes H = \text{diag}\langle H, \dots H \rangle$, where

$$H_{((T-1)\times T)} = \begin{pmatrix} -\phi & 1 & 0 & \dots & \dots & 0 \\ 0 & -\phi & 1 & 0 & \dots & \vdots \\ \vdots & 0 & \ddots & \ddots & 0 & \vdots \\ \vdots & 0 & 0 & -\phi & 1 & 0 \\ 0 & \dots & \dots & 0 & -\phi & 1 \end{pmatrix} \tag{7}$$

Next, we follow Lee and Yu (2008) to construct the transformation matrix F. Define $FT_{T-1,T-2}$ as the matrix of eigenvectors corresponding to the eigenvalues of 1 of the matrix $I_{T-1} - \frac{1}{T-1} l_{T-1} l'_{T-1}$ where l_{T-1} is a $T-1$-dimensional vector containing the elements $\sqrt{T-1}$. The transformation matrix F is then given by $F = I_N \otimes FT'_{T-1,T-2}$ (see Lee and Yu, 2008, p.5). This transformation matrix deviates from Lee and Yu (2008) in two respects. First, it is rearranged according to our data organization. Second, due the application of the Cochrane-Orcutt procedure in a first step the number of available time periods is reduced by one, i.e. F is of dimension $N(T-2) \times N(T-1)$.

Applying the above transformation matrices F and P^{-1} to equation (3) and reducing the weight matrix by two time dimensions, i.e. $W = \mathrm{w} \otimes I_{T-2}$, the model can be written as

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

[7] Note that the Prais-Winsten (1954) transformation changes the first entry of the fixed-effects vector differently from the others. Hence, the transformation by Lee and Yu (2008) would not eliminate the respective vectors.

with $\tilde{y} = FP^{-1}y$, $\tilde{X} = FP^{-1}X$ and $\tilde{\epsilon} = FP^{-1}\epsilon$. Finally, the variance-covariance matrix $\sigma^2\tilde{\Omega}$ is equal to $\mathrm{E}[\tilde{\epsilon}\tilde{\epsilon}'] = \sigma^2(\Sigma_N \otimes I_{T-2})$ (see Appendix). Defining $A = (I - \rho W)$, $\tilde{Z} = [\tilde{X}, W\tilde{X}]$, and $\delta = (\beta', \gamma')'$ we can write the log-likelihood function $\mathcal{L}$ as

$$\mathcal{L} \propto -\frac{N(T-2)}{2}\log\sigma^2 - \frac{(T-2)}{2}\sum_{i=1}^{N}\log h_i(\alpha) - \frac{1}{2\sigma^2}(A\tilde{y} - \tilde{Z}\delta)'\tilde{\Omega}^{-1}(A\tilde{y} - \tilde{Z}\delta) + \log|A| \qquad (9)$$

To keep the number of parameters to be estimated small, we impose a functional form for the individual variances. This is particularly important in panel data settings with large N and small T. The individual variances σ_i^2 contained in Σ_N are represented by $h_i(\alpha)$ such that $\tilde{\Omega} = \mathrm{diag}\langle h_1(\alpha)I_{T-2} \dots h_N(\alpha)I_{T-2}\rangle$. Specifically, the function $h_i(\alpha) = h(\Lambda_i\alpha)$ models the individual variances as being dependent on the matrix Λ_i containing potential explanatory variables (see e.g. Griffiths, 2003). Note that the variables in Λ_i may be equal to those in $\tilde{Z}$. As the variances are assumed to be equal over time periods, Λ_i may be specified in terms of time averages of the explanatory variables.

For δ and σ^2 we can maximize the log-likelihood function analytically to obtain the following estimators:

$$\hat{\delta} = (\tilde{Z}'\tilde{\Omega}^{-1}\tilde{Z})^{-1}(\tilde{Z}'\tilde{\Omega}^{-1}A\tilde{y}) \qquad (10)$$

$$\hat{\sigma}^2 = \frac{1}{N(T-2)}(A\tilde{y} - \tilde{Z}\delta)'\tilde{\Omega}^{-1}(A\tilde{y} - \tilde{Z}\delta) \qquad (11)$$

For the other parameters the concentrated log-likelihood function,

$$\mathcal{L} \propto -\frac{N(T-2)}{2}\log\hat{\sigma}^2 - \frac{T-2}{2}\sum_{i=1}^{N}\log h_i(\alpha) + \log|A|, \qquad (12)$$

has to be maximized numerically. This is accomplished by applying a gradient-based method, implemented in the Optimization Toolbox of Matlab, that attempts to find a minimum of the negative multivariable log-likelihood function under the constraint that ϕ and ρ are smaller than one in absolute values. To derive standard errors for the Maximum Likelihood estimates we compute the second derivatives of the log-likelihood function, where the Hessian matrix is reported in the Appendix.

4 Design of the Monte Carlo Study

The purpose of our Monte Carlo experiments is twofold. First, we investigate the finite sample properties of the estimator outlined in the previous section (henceforth ML_1) and compare it to an estimator that does not take into account serial correlation and heteroskedasticity (henceforth ML_2). Second, we want to call attention to the potentially misleading conclusions drawn from applying standard $t-$tests in a Spatial Durbin Model for Panel data. Here we focus on the impact of the density of the weight matrix on the performance of the tests.

In both experiments the data generating process is assumed to be of the form specified in equations (3)-(5), where we additionally allow for spatial dependence in the explanatory variables X, i.e. :

$$
\begin{aligned}
y &= \rho W y + X\beta + WX\gamma + Z\mu_1 + WZ\mu_2 + \epsilon \qquad (13) \\
X &= \vartheta WX + \zeta \qquad (14) \\
\epsilon &\sim N(0, \sigma^2\Omega) \qquad (15)
\end{aligned}
$$

The explanatory variables X and the error terms ζ are drawn from a standard-normal distribution, where the observations in X are assumed to be weakly spatially dependent ($\vartheta = 0.3$). Regarding the matrix Ω, ω_T is constructed as in equation (6). The elements of Σ_N are defined by $\sigma_i = h_i(\alpha) = \exp(\alpha\overline{\Lambda}_i)$, where the $\overline{\Lambda}_i$ are drawn from a uniform-distribution with support $[0, 1]$. As discussed in section (3), $\overline{\Lambda}_i$ can be regarded as the average over time of a variable capable to explain the individual variances. The exponential specification was chosen to ensure that variances are nonnegative. Alternatively, this could also be achieved by using the squared sum (cf. e.g. Grffiths, 2003, p.87).

4.1 First experiment

In the first experiment we run $M = 2,000$ Monte Carlo trials and investigate the finite sample properties of the estimator by means of the average bias, the empirical standard deviation derived from the Monte Carlo trials (E-SD), and the root mean squared error (RMSE).[8] Additionally, we judge the quality of the variance estimates derived from the Hessian (T-SD) by comparing them to the empirical standard deviation.

We repeat this experiment with ML_2 to assess the impact of neglecting the non-spherical nature of the error term. Hence, instead of maximizing the log-likelihood function outlined in equation

[8] The average bias is defined as Bias= $\sum_{i=1}^{M} \frac{\hat{\theta}_i - \theta}{M}$, where θ refers to the value of the coefficient in the data generating process and $\hat{\theta}_i$ to the estimated value. Moreover, E-SD= $\sqrt{\frac{1}{M-1}\sum_{i=1}^{M}(\hat{\theta}_i - \text{mean}(\hat{\theta}_i))^2}$ and RMSE $= \sqrt{\text{E-SD}(\hat{\theta})^2 + \text{Bias}(\hat{\theta})^2}$.

(9), the ML_2 estimator maximizes

$$\mathcal{L} \propto -\frac{N(T-1)}{2}\log\sigma^2 - \frac{1}{2\sigma^2}(A\bar{y} - \bar{Z}\delta)'(A\bar{y} - \bar{Z}\delta) + \log|A| \tag{16}$$

where $\overline{y}$ and $\overline{Z}$ arrive by transforming equation (3) only by the matrix proposed by Lee and Yu (2008) to get rid of the fixed effects. Note that there is no need to apply the Cochrane-Orcutt-transformation as Ω in equation (4) reduces to the identity matrix. Consequently, $T-1$ instead of $T-2$ time periods are available in the estimation.[9]

The parameter vectors used in the first experiment are shown in table 1. We use θ_0 as basic specification and subsequently vary the values of γ, ρ, ϕ, and α. The first three modifications (θ_1 to θ_3) aim to investigate the behavior of the estimators under different specifications of spatial dependence. In the first modification (θ_1) the spatially lagged dependent variable has a negative impact on y and in the second modification (θ_2) it is not present in the data generating process. Modification (θ_3) evaluates the estimators in a scenario in which the spatially lagged independent variable does not show up in the data generating process (i.e. $\gamma = 0$).

By means of the final four modifications (θ_4 to θ_7) we want to assess the reliability of the estimators for different degrees of serial correlation and heteroskedasticity. First, we apply the estimators to a setting with relatively high serial correlation but homoskedasticity (θ_4), whereas in θ_5 the errors do not display serial correlation but pronounced heteroskedasticity. After investigating the impact of serial correlation and heteroskedasticiy separately, we apply the parameter vector θ_6 in which both, serial correlation and heteroskedasticity are high. To receive an impression regarding the effect of changing α on the extent of heteroskedasticity, observe that in the basic specification with $\alpha = 1$ the individual variance parameters σ_i range from 1 to $e \approx 2.27$ with a mean of $e^{\frac{1}{2}} \approx 1.65$, whereas with $\alpha = 2.5$, σ_i ranges from 0 to $e^{\frac{5}{2}} \approx 12.18$ with an expected value of $e^{\frac{5}{4}} \approx 3.49$. The aim of the final modification (θ_7) is to evaluate the estimators when neither heteroskedasticity nor serial correlation are present in the data generating process.

Table 1 about here

The spatial weight matrix w is constructed by applying first-order rook contiguity (see e.g. Anselin, 1988, p.17) to an r-board where $r = 4, \ldots, 11$. Hence, the number of cross sectional units N ranges from 16 to 121.

[9]To estimate this model we use the SDM function available in James LeSage's Econometric Toolbox (`http://www.spatial-econometrics.com`).

4.2 Second experiment

In the second experiment we are interested in the performance of simple hypothesis tests for the null-hypotheses that each, β, γ and ρ are equal to zero when their values in the data generating process are actually different. Specifically, we the set the values of the parameters of interest in the data generating process to 0.2, and use the values of the basic specification in the first Monte Carlo experiment for the other parameters. This results in the parameter vector $(\beta, \gamma, \rho, \sigma^2, \phi, \alpha) = (0.2, 0.2, 0.2, 1, 0.5, 1)$.

The elements of the weight matrix w are drawn from a Bernoulli distribution with success probability d. We allow for different degrees of sparseness of the weight matrix by varying the success probability of the Bernoulli distribution from 0.1 to 0.9 in steps of 0.1. We make sure that each observation has at least one neighbor by randomly assigning a neighbor in those cases where the procedure outlined above does not assign a neighbor. For both, a time dimension of $T = 10$ and $T = 20$ we consider $16, 36, 49$, and 64 cross-sectional units. In this experiment the number of trials equals 10,000. The number has to be considerably higher than in the first experiment because our interest focuses on the tails of the distribution (see e.g. Mooney, 1997, p.58). Apart from these modifications the setup of the second Monte Carlo experiment is the same as in the first experiment.

To gain insights in the performance of simple hypothesis tests, we calculate the fraction of Monte Carlo trials in which a Type 2 error occurs. This fraction is henceforth called acceptance rate and is calculated as follows: after estimating the coefficients, we obtain the standard errors of the estimates from the diagonal elements of the inverted Hessian and perform a t-test for the null-hypotheses that each, β, γ, and ρ is equal to zero. The significance level is set to 10%. We then calculate the acceptance rates i.e. the proportion of Monte Carlo trials in which the null hypotheses is wrongly not rejected. Hence, by subtracting the acceptance rates from 1, one obtains an estimate for the value of the power function at the paramter value set in the data generating process (i.e., 0.2).

5 Results

5.1 First experiment

The results of the first Monte Carlo experiment are presented in tables 2 and 3. Table 2 shows the results of different scenarios of spatial dependence (parameter specifications θ_0 to θ_3). For θ_0 we display a relatively broad selection of (N, T) pairs but refrain from reporting results for $N > 49$

as the properties of the estimator are already quite satisfactory for $N = 49$. For brevity, results corresponding to the remaining parameter specifications (θ_1 to θ_7) are shown for a small and a large sample only, i.e. we choose $(N, T) = (16, 5)$ and $(N, T) = (49, 20)$, respectively.[10]

Concerning the bias of β, γ and ρ the Monte Carlo experiments reveal that both estimators have satisfactory properties quite irrespective of the sample size and the specification used. The underestimation of ρ in specification θ_0 and the overestimation in specification θ_1 are in line with the results of Bao and Ullah (2007) who investigate the small sample properties of a spatial autoregressive model in the cross-section. The bias of α is quite small except in some settings with small N. The estimates for ϕ suffer from the scarcity of observations when T is small. However, the higher bias of ϕ and α is of less importance, if these parameters are regarded as nuisance parameters as their inclusion leads to an improvement in the estimation of the variance. This can be seen from the much more pronounced bias of σ^2 from the ML_2 estimator at least in settings with $T = 20$, even though serial correlation and heteroskedasticity are only of moderate size.

Table 2 about here

The impact of serial correlation and heteroskedasticity on the estimates can be analyzed in more detail by means of table 3, which shows the results of the first Monte Carlo experiment for different degrees of heteroskedasticity and serial correlation (i.e. for the parameter vectors θ_4 to θ_7). Concerning the results for $T = 5$ the assessment of the relative performance of the two estimators is somewhat mixed. The loss of an additional time period due to the Cochrane-Orcutt procedure has a negative impact on the ML_1 results. But, as can be seen from the results for θ_0 in table 2 a sufficiently large number of cross-sectional elements can outweigh the effects of the smaller number of time periods available for estimation. For $T = 20$ the results are more clear cut and show the advantages of taking into account serial correlation and heteroskedasticity in the estimator. A high degree of serial correlation in a setting with homoskedastic errors (θ_4) leads to much higher standard errors of the coefficient estimates of the ML_2 estimator and a considerable overestimation of the variance. Strong heteroskedasticity without serial correlation (θ_5) leads to even more biased estimates for the variance and consequently the inflation of standard errors is less pronounced. The combined effects of serial correlation and heteroskedasticity can be seen from the results for the parameter vector θ_6 for which both the variance is strongly upwardly biased and the standard deviation is much higher than for the ML_1 estimator. The results for specification θ_7 reveal that the ML_1 estimator leads to relative precise results even if neither serial correlation nor

[10] The full table of results is available from the authors upon request.

heteroskedasticity are present in the data generating process. The comparison between the two estimators suggests that in settings with few time periods and little heteroskedasticity and serial correlation the ML_2 estimator may be preferred. In all other settings the ML_1 estimator proves to be the preferred choice.

Table 3 about here

Regarding the ML_1 estimator in most cases the difference between the empirical standard deviation and the theoretical standard deviation is fairly small. Consequently, the Hessian of ML_1 provides good estimates for the variance of the estimator.

5.2 Second experiment

Figures 1 and 2 show the proportion of Monte Carlo trials in which the null-hypotheses $\beta = 0$, $\gamma = 0$, and $\rho = 0$ are wrongly accepted as a function of the density of the spatial weight matrix for different values of N and for $T = 10$ and $T = 20$. The figures show that the acceptance rate of β decreases with the sample size but does not depend on the density of the spatial weight matrix. In contrast, acceptance rates for γ and ρ increase with density. The number of cross-sectional units has hardly any influence but the acceptance rates are lower for $T = 20$ than for $T = 10$. This effect is most pronounced if d is low. The results for γ and ρ may be explained as follows: With a row standardized spatial weight matrix, γ and ρ can be regarded as the coefficients on the averages over the neighbors' explanatory variables and the neighbors' dependent variable, respectively. The averaging leads to a lower variance of the spatially lagged variables compared to the non-lagged variables. The variance is the lower the higher N because the number of non-zero elements of the spatial weight matrix is proportional to N. Given N, the number of neighbors increases with d. On the other hand the number of non-zero elements in the weight matrix is independent of T. For β an increase in sample size leads to lower acceptance rates, irrespective whether the increase is due to higher N or higher T. In contrast, for γ and ρ the positive effects of an increase in N on acceptance rates are outweighed by the negative effects of the higher number of neighbors.[11]

Figures 1 and 2 about here

[11] This is a special feature of the weight matrix used. If instead a weight matrix corresponding to first order rook contiguity is applied, acceptance rates for ρ and γ drop in N. (Results not reported here.)

These results point out an additional advantage of panel data models because in cross-sectional settings it is not possible to cope with this problem. Our results thus show that it is not unlikely to wrongly conclude that spatial lags of the dependent or the explanatory variable should be excluded from the econometric specification used in the estimation. As this could lead to an omitted variable bias we advise special caution when applying standard $t-$tests in a Spatial Durbin Model for Panel data.

6 Conclusion

In this paper we argue that the Spatial Durbin Model is a suitable framework for modeling spillovers and externalities of different sorts and that it may be used for model selection purposes as it nests several models frequently applied in empirical work. Motivated by its theoretical and econometric practicality, we extend the SDM to panel data allowing for non-spherical disturbances. We propose an estimator based on Maximum Likelihood techniques and show by means of a Monte Carlo analysis that it has fairly good small sample properties as measured by the bias, the empirical and theoretical standard deviations and the root mean squared error. A comparison with an estimator that ignores heteroskedasticity or serial correlation among disturbances reveals that standard errors of parameter estimates may be largely inflated when the non-spherical error structure is not considered. Whereas employing the proposed estimator in the case of spherical errors does hardly harm estimation results.

Additionally, the estimator is applicable in spatial autoregressive models with non-spherical disturbances. This is of particular relevance when the impact of the non-spatial regressors on the dependent variable is low, as IV estimates can be biased due to the weak instrument problem.

Moreover, we analyze the likelihood of committing a type 2 error in testing procedures for parameter significance of spatially lagged variables and show that it is the higher the higher the connectivity of regional entities. Applying standard testing procedures may therefore lead to wrong conclusions concerning the influence of spillovers and to incorrect model specifications. Therefore, we advise special caution when applying standard $t-$tests in spatial models. Further investigation of the power of hypothesis tests in such settings may thus be beneficial.

A final remark needs to be made regarding the interpretation of parameter estimates in the SDM. Unlike in OLS type regressions where independence of observations is assumed, parameter estimates in models containing spatial lags of the dependent variable have not a straightforward interpretation because of the embedded feedback effects among units. LeSage and Pace (2009) therefore develop summary measures for cross-sectional data reflecting the impact of a change in a single unit and of changes on a single unit. In order to meaningfully interpret parameter estimates in the context of the proposed SDM, the development of summary measures seems to be a promising field for future research.

References

Anselin, L. 1988 *Spatial econometrics: Methods and Models.* Kluwer Acad. Publ.: Kluwer Academic. Publishers.

Anselin, L., Florax, RJGM., Rey, SJ. 2004 *Advances in Spatial Econometrics: Methodology, Tools and Applications.* Springer.

Baltagi, BH. 2005 *Econometric analysis of panel data.* John Wiley & Sons.

Baltagi, BH., Song, SH., Koh, W. 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* **117**(1): 123–150.

Bao, Y., Ullah, A. 2007. Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* **137**(2): 396–413.

Brasington, DM., Hite, D. 2005. Demand for environmental quality: a spatial hedonic analysis. *Regional Science and Urban Economics* **35**(1): 57–82.

Burridge, P. 1981. Testing for a common factor in a spatial autoregression model. *Environment and Planning A* **13**(7): 795–800.

Cliff, AD., Ord, JK. 1981 *Spatial processes: models and applications.* Pion, London.

Cochrane, D., Orcutt, G. 1949. Application of Least Squares Regression to Relationships Containing Autocorelated Error Terms. *Journal of the American Statistical Association* **44**(2): 32–61.

Conley, TG. 1999. GMM estimation with cross sectional dependence. *Journal of Econometrics* **92**(1): 1–45.

Das, D., Kelejian, HH., Prucha, IR. 2003. Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* **82**(1): 1–26.

Driscoll, JC., Kraay, AC. 1998. Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *The Review of Economics and Statistics* **80**(4): 549–560.

Durbin, J. 1960. The Fitting of Time-Series Models. *Review of the International Statistical Institute* **28**(3): 233–244.

Ertur, C., Koch, W. 2007. Growth, technological interdependence and spatial externalities: theory and evidence. *Journal of Applied Econometrics* **22**(6): 1033–1062.

Griffiths, WE. 2003. Heteroskedasticity. In *A Companion to Theoretical Econometrics,* ed. BH. Baltagi. Blackwell Publishing Ltd. chapter 3, pp. 82–100.

Kapoor, M., Kelejian, HH., Prucha, IR. 2007. Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics* **140**(1): 97–130.

Kelejian, HH., Prucha, IR. 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics* **17**(4): 99–121.

_ 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* **40**(2): 509–533.

Lee, LF. 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* **72**(6): 1899–1925.

Lee, LF., Yu, J. 2008. Estimation of spatial autoregressive panel data models with fixed effects. Working paper downloaded on April, 17 2009 from `http://economics.sbs.ohio-state.edu/lee/` .

LeSage, JP., Pace, RK. 2009 *Introduction to Spatial Econometrics.* Chapman & Hall/CRC.

Mooney, CZ. 1997 *Monte Carlo Simulation* Quantitative Applications in the Social Sciences. SAGE Publications.

Mur, J., Angulo, A. 2006. The Spatial Durbin Model and the Common Factor Test. *Spatial Economic Analysis* **1**(2): 207–226.

Neyman, J., Scott, EL. 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica* **16**(1): 1–32.

Prais, S., Winsten, C. 1954. Trend estimation and serial correlation. Cowles Commission Discussion Paper 383, Cowles Commission.

A Appendix

A.1 Hessian

The second order derivatives read

$$
\begin{aligned}
\mathcal{L}_{\delta\delta} &= -\frac{1}{\sigma^2}\tilde{Z}'\tilde{\Omega}^{-1}\tilde{Z} && (17)\\
\mathcal{L}_{\delta\rho} &= -\frac{1}{\sigma^2}(W\tilde{y})'\tilde{\Omega}^{-1}\tilde{Z} && (18)\\
\mathcal{L}_{\delta\sigma^2} &= -\frac{1}{\sigma^4}\nu'\tilde{\Omega}^{-1}\tilde{Z} && (19)\\
\mathcal{L}_{\delta\phi} &= \frac{1}{\sigma^2}\left(\nu'\tilde{\Omega}^{-1}F\Xi Z + (AF\Xi y - F\Xi Z\delta)'\tilde{\Omega}^{-1}\tilde{Z}\right) && (20)\\
\mathcal{L}_{\delta\alpha} &= \frac{1}{\sigma^2}\nu'\frac{\partial\tilde{\Omega}^{-1}}{\partial\alpha}\tilde{Z} && (21)\\
& && (22)\\
\mathcal{L}_{\rho\rho} &= -\frac{1}{\sigma^2}\left(W\tilde{y}\right)'\tilde{\Omega}^{-1}W\tilde{y} - \mathrm{tr}(A^{-1}WA^{-1}W) && (23)\\
\mathcal{L}_{\rho\sigma^2} &= -\frac{1}{\sigma^4}\nu'\tilde{\Omega}^{-1}W\tilde{y} && (24)\\
\mathcal{L}_{\rho\phi} &= \frac{1}{\sigma^2}\left(\left(AF\Xi\tilde{y} - F\Xi Z\delta\right)'\tilde{\Omega}^{-1}W\tilde{y} + \nu'\tilde{\Omega}^{-1}WF\Xi y\right) && (25)\\
\mathcal{L}_{\rho\alpha} &= \frac{1}{\sigma^2}\nu'\frac{\partial\tilde{\Omega}^{-1}}{\partial\alpha}W\tilde{y} && (26)\\
& && (27)\\
\mathcal{L}_{\phi\sigma^2} &= \frac{1}{\sigma^4}\nu'\tilde{\Omega}^{-1}\left(AF\Xi y - F\Xi Z\delta\right) && (28)\\
\mathcal{L}_{\phi\phi} &= -\frac{1}{\sigma^2}\left(AF\Xi y - F\Xi Z\delta\right)'\tilde{\Omega}^{-1}\left(AF\Xi y - F\Xi Z\delta\right) && (29)\\
\mathcal{L}_{\phi\alpha} &= -\frac{1}{\sigma^2}\nu'\frac{\partial\tilde{\Omega}^{-1}}{\partial\alpha}\left(AF\Xi y - F\Xi Z\delta\right) && (30)\\
& && (31)\\
\mathcal{L}_{\alpha\sigma^2} &= \frac{1}{2\sigma^4}\nu'\frac{\partial\tilde{\Omega}^{-1}}{\partial\alpha}\nu && (32)\\
\mathcal{L}_{\alpha\alpha} &= -\frac{T-2}{2}\sum_{i=1}^{N}\frac{h_i(\alpha)h_i(\alpha)'' - (h_i(\alpha)')^2}{(h_i(\alpha))^2} - \frac{1}{2\sigma^2}\nu'\frac{\partial\partial\tilde{\Omega}^{-1}}{\partial\alpha\partial\alpha}\nu && (33)\\
& && (34)\\
\mathcal{L}_{\sigma^2\sigma^2} &= \frac{N\left(T-2\right)}{2\sigma^4} - \frac{1}{\sigma^6}\nu'\tilde{\Omega}^{-1}\nu && (35)
\end{aligned}
$$

where,

$$
\begin{aligned}
\nu &= A(\rho)FP^{-1}(\phi)y - FP^{-1}(\phi)Z\delta && (36)\\
\Xi &= \frac{\partial P^{-1}}{\partial\phi} = I_N \otimes \frac{\partial H}{\partial\phi} && (37)
\end{aligned}
$$

$$\frac{\partial H}{\partial \phi} = \begin{pmatrix} -1 & 0 & & \dots & & 0 \\ 0 & -1 & \ddots & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & -1 & 0 & 0 \\ 0 & \dots & \dots & 0 & -1 & 0 \end{pmatrix} \tag{38}$$

$$\frac{\partial \tilde{\Omega}^{-1}}{\partial \alpha} = \begin{pmatrix} -\frac{h_1'(\alpha)}{h_1(\alpha)^2} I_{T-2} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\frac{h_N'(\alpha)}{h_N(\alpha)^2} I_{T-2} \end{pmatrix} \tag{39}$$

$$\frac{\partial^2 \tilde{\Omega}^{-1}}{\partial^2 \alpha} = \begin{pmatrix} -\frac{h_1''(\alpha)h_1(\alpha)-2(h_1'(\alpha))^2}{h_1(\alpha)^3} I_{T-2} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\frac{h_N''(\alpha)h_N(\alpha)-2(h_N'(\alpha))^2}{h_N(\alpha)^3} I_{T-2} \end{pmatrix} \tag{40}$$

A.2 Transformed errors

To see that $\mathrm{E}[\tilde{\epsilon}\tilde{\epsilon}^{'}] = \Sigma_N \otimes I_{T-2}$, note that

$$\begin{aligned} \mathrm{E}[\tilde{\epsilon}\tilde{\epsilon}'] &= \mathrm{E}[FP^{-1}(\Sigma_N \otimes \omega_T)P^{-1'}F'] \\ &= F\mathrm{E}[P^{-1}(\Sigma_N \otimes \omega_T)P^{-1}]F' \\ &= F\mathrm{E}[(I_N \otimes H)(\Sigma_N \otimes \omega_T)(I_N \otimes H)]F' \\ &= F\mathrm{E}[(\Sigma_N \otimes H\omega_T)(I_N \otimes H)]F' \\ &= F\mathrm{E}[\Sigma_N \otimes H\omega_T H] \end{aligned} \tag{41}$$

Sine $H\omega_T H = I_{T-1}$, by denoting the expectations of the individual variance elements as σ_i we get $F\mathrm{E}[\Sigma_N \otimes H\omega_T H] =$

$$\begin{pmatrix} \sigma_1^2 FT' I_{T-1} FT & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_N^2 FT' I_{T-1} FT \end{pmatrix} = \begin{pmatrix} \sigma_1^2 I_{T-2} & 0 & \dots & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sigma_N^2 I_{T-2} \end{pmatrix} \tag{42}$$

where the second equality follows from $FT'FT = I_{T-2}$.

Table 1: Parameter vectors

	β	γ	ρ	σ^2	ϕ	α
θ_0	1	1	0.5	1	0.5	1
θ_1	1	1	-0.5	1	0.5	1
θ_2	1	1	0	1	0.5	1
θ_3	1	0	0.5	1	0.5	1
θ_4	1	1	0.5	1	0.9	0
θ_5	1	1	0.5	1	0	2.5
θ_6	1	1	0.5	1	0.9	2.5
θ_7	1	1	0.5	1	0	0

Table 2: Monte Carlo Results for θ_0-θ_3

θ	T	N		β		γ		ρ		σ^2		ϕ	α
				ML1	ML2	ML1	ML2	ML1	ML2	ML1	ML2	ML1	ML1
θ_0	5	16	Bias	0.011	0.012	0.040	0.039	-0.021	-0.017	-0.182	0.518	-0.434	0.101
			E-SD	0.193	0.173	0.395	0.346	0.123	0.109	0.447	0.345	0.151	0.974
			T-SD	0.180	0.170	0.366	0.334	0.116	0.102	0.384	0.275	0.149	0.831
			RMSE	0.194	0.174	0.397	0.348	0.125	0.110	0.482	0.623	0.460	0.979
	5	49	Bias	0.008	0.009	0.016	0.015	-0.009	-0.008	-0.164	0.570	-0.417	0.029
			E-SD	0.103	0.096	0.225	0.212	0.076	0.070	0.217	0.197	0.081	0.458
			T-SD	0.102	0.097	0.219	0.204	0.071	0.064	0.204	0.162	0.083	0.424
			RMSE	0.103	0.097	0.226	0.213	0.077	0.070	0.272	0.603	0.425	0.459
	10	16	Bias	0.004	0.002	0.015	0.022	-0.008	-0.011	-0.055	0.845	-0.190	0.030
			E-SD	0.110	0.127	0.220	0.252	0.072	0.082	0.279	0.319	0.089	0.503
			T-SD	0.110	0.123	0.220	0.235	0.070	0.069	0.257	0.223	0.085	0.475
			RMSE	0.110	0.127	0.220	0.253	0.073	0.083	0.284	0.903	0.210	0.504
	10	49	Bias	0.004	0.004	0.003	0.004	-0.002	-0.003	-0.037	0.878	-0.183	0.004
			E-SD	0.062	0.072	0.131	0.155	0.043	0.050	0.147	0.175	0.048	0.261
			T-SD	0.063	0.070	0.134	0.144	0.043	0.043	0.142	0.130	0.048	0.255
			RMSE	0.062	0.072	0.131	0.155	0.043	0.050	0.152	0.895	0.189	0.261
	20	16	Bias	0.005	0.007	0.004	0.011	-0.004	-0.008	-0.019	1.058	-0.086	0.015
			E-SD	0.072	0.089	0.143	0.179	0.046	0.056	0.176	0.269	0.056	0.317
			T-SD	0.073	0.089	0.145	0.168	0.046	0.048	0.174	0.172	0.054	0.310
			RMSE	0.072	0.089	0.143	0.179	0.046	0.057	0.177	1.092	0.102	0.317
	20	49	Bias	0.001	0.001	0.000	0.002	0.000	-0.001	-0.011	1.067	-0.083	0.001
			E-SD	0.040	0.050	0.087	0.107	0.029	0.035	0.098	0.154	0.031	0.170
			T-SD	0.041	0.050	0.089	0.102	0.029	0.030	0.096	0.098	0.031	0.168
			RMSE	0.040	0.050	0.087	0.107	0.029	0.035	0.099	1.078	0.089	0.170
θ_1	5	16	Bias	-0.003	-0.002	-0.018	-0.010	0.025	0.023	-0.184	0.519	-0.435	0.100
			E-SD	0.184	0.166	0.338	0.295	0.141	0.125	0.435	0.346	0.151	0.964
			T-SD	0.170	0.161	0.306	0.283	0.135	0.116	0.383	0.277	0.149	0.829
			RMSE	0.184	0.166	0.338	0.296	0.143	0.127	0.472	0.624	0.461	0.969
	20	49	Bias	0.000	-0.001	-0.002	-0.003	0.003	0.004	-0.011	1.069	-0.084	0.001
			E-SD	0.038	0.048	0.074	0.091	0.035	0.042	0.098	0.155	0.031	0.170
			T-SD	0.039	0.048	0.074	0.089	0.033	0.033	0.097	0.099	0.031	0.168
			RMSE	0.038	0.048	0.074	0.091	0.035	0.042	0.099	1.080	0.089	0.170
θ_2	5	16	Bias	0.002	0.003	0.003	0.008	-0.002	-0.001	-0.191	0.506	-0.435	0.101
			E-SD	0.188	0.169	0.367	0.321	0.164	0.147	0.433	0.336	0.151	0.969
			T-SD	0.175	0.165	0.339	0.311	0.160	0.139	0.377	0.267	0.149	0.830
			RMSE	0.188	0.169	0.367	0.321	0.164	0.147	0.473	0.607	0.460	0.974
	20	49	Bias	0.000	0.000	-0.002	-0.001	0.002	0.002	-0.012	1.066	-0.084	0.001
			E-SD	0.039	0.049	0.083	0.100	0.041	0.049	0.097	0.152	0.031	0.170
			T-SD	0.040	0.049	0.083	0.096	0.040	0.040	0.096	0.096	0.031	0.168
			RMSE	0.039	0.049	0.083	0.100	0.041	0.049	0.098	1.077	0.089	0.170
θ_3	5	16	Bias	0.005	0.006	0.044	0.043	-0.029	-0.025	-0.183	0.519	-0.436	0.102
			E-SD	0.183	0.165	0.361	0.316	0.143	0.126	0.440	0.346	0.150	0.970
			T-SD	0.170	0.161	0.334	0.305	0.135	0.116	0.383	0.277	0.149	0.829
			RMSE	0.183	0.165	0.364	0.319	0.146	0.128	0.477	0.624	0.461	0.975
	20	49	Bias	0.000	0.000	0.006	0.005	-0.004	-0.004	-0.011	1.069	-0.083	0.004
			E-SD	0.038	0.049	0.081	0.100	0.034	0.041	0.098	0.153	0.031	0.168
			T-SD	0.039	0.048	0.081	0.094	0.033	0.033	0.097	0.099	0.031	0.169
			RMSE	0.038	0.049	0.081	0.100	0.034	0.041	0.098	1.080	0.089	0.168

Notes: The weight matrix is derived using first order rook contiguity. E-SD denotes the empirical standard deviation of the Monte Carlo results, RMSE the root mean squared error and T-SD the theoretical standard deviation derived from the inverse of the Hessian.

Table 3: Monte Carlo Results for θ_4-θ_7

θ	T	N		β		γ		ρ		σ^2		ϕ	α
				ML1	ML2	ML1	ML2	ML1	ML2	ML1	ML2	ML1	ML1
θ_4	5	16	Bias	0.009	0.009	0.029	0.036	-0.016	-0.017	-0.272	-0.078	-0.585	0.038
			E-SD	0.133	0.141	0.298	0.311	0.108	0.113	0.422	0.219	0.157	0.987
			T-SD	0.126	0.135	0.277	0.280	0.101	0.094	0.344	0.167	0.140	0.835
			RMSE	0.133	0.141	0.300	0.313	0.109	0.115	0.502	0.232	0.606	0.988
	20	49	Bias	0.001	0.002	0.005	0.013	-0.002	-0.007	-0.040	1.455	-0.128	0.002
			E-SD	0.028	0.063	0.067	0.152	0.025	0.061	0.095	0.254	0.024	0.170
			T-SD	0.029	0.054	0.067	0.109	0.025	0.031	0.093	0.117	0.021	0.169
			RMSE	0.028	0.063	0.067	0.152	0.025	0.062	0.103	1.477	0.130	0.170
θ_5	5	16	Bias	0.013	0.022	0.046	0.045	-0.022	-0.021	-0.148	3.311	-0.264	0.207
			E-SD	0.286	0.287	0.525	0.507	0.120	0.111	0.459	1.222	0.134	0.958
			T-SD	0.255	0.275	0.475	0.496	0.115	0.111	0.398	0.785	0.144	0.827
			RMSE	0.287	0.288	0.527	0.508	0.122	0.113	0.482	3.529	0.297	0.980
	20	49	Bias	0.002	0.000	0.006	0.006	-0.003	-0.003	-0.007	3.460	-0.052	0.008
			E-SD	0.059	0.073	0.117	0.140	0.028	0.032	0.098	0.502	0.033	0.167
			T-SD	0.058	0.072	0.113	0.137	0.028	0.032	0.097	0.213	0.034	0.169
			RMSE	0.059	0.073	0.117	0.140	0.029	0.032	0.098	3.496	0.062	0.167
θ_6	5	16	Bias	0.010	0.012	0.036	0.063	-0.019	-0.025	-0.330	3.138	-0.588	0.196
			E-SD	0.210	0.281	0.414	0.523	0.118	0.134	0.366	1.410	0.156	0.966
			T-SD	0.198	0.269	0.385	0.486	0.110	0.111	0.314	0.753	0.140	0.830
			RMSE	0.210	0.281	0.415	0.527	0.120	0.136	0.493	3.440	0.608	0.986
	20	49	Bias	0.001	0.002	0.000	0.012	0.000	-0.006	-0.043	9.988	-0.128	0.005
			E-SD	0.044	0.114	0.090	0.233	0.027	0.068	0.097	1.744	0.025	0.174
			T-SD	0.045	0.111	0.092	0.202	0.027	0.033	0.093	0.525	0.021	0.168
			RMSE	0.044	0.114	0.090	0.233	0.027	0.069	0.106	10.139	0.130	0.174
θ_7	5	16	Bias	0.013	0.012	0.040	0.028	-0.020	-0.015	-0.073	-0.044	-0.263	0.037
			E-SD	0.178	0.140	0.364	0.290	0.119	0.098	0.506	0.176	0.136	0.976
			T-SD	0.160	0.138	0.336	0.284	0.113	0.095	0.435	0.173	0.144	0.832
			RMSE	0.178	0.141	0.366	0.291	0.120	0.099	0.512	0.181	0.296	0.977
	20	49	Bias	0.001	0.001	0.006	0.005	-0.003	-0.002	-0.004	-0.002	-0.052	0.001
			E-SD	0.038	0.037	0.084	0.082	0.028	0.028	0.098	0.047	0.033	0.167
			T-SD	0.037	0.036	0.082	0.079	0.028	0.027	0.097	0.047	0.034	0.169
			RMSE	0.038	0.037	0.084	0.082	0.029	0.028	0.098	0.047	0.062	0.167

Notes: The weight matrix is derived using first order rook contiguity. E-SD denotes the empirical standard deviation of the Monte Carlo results, RMSE the root mean squared error and T-SD the theoretical standard deviation derived from the inverse of the Hessian.

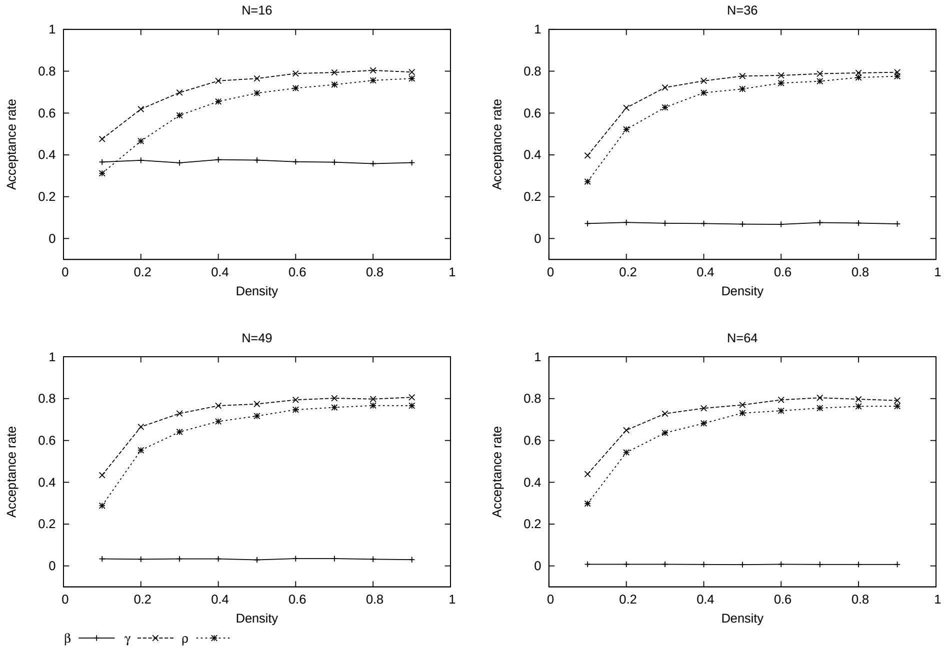

Figure 1: Acceptance rates: $T = 10$

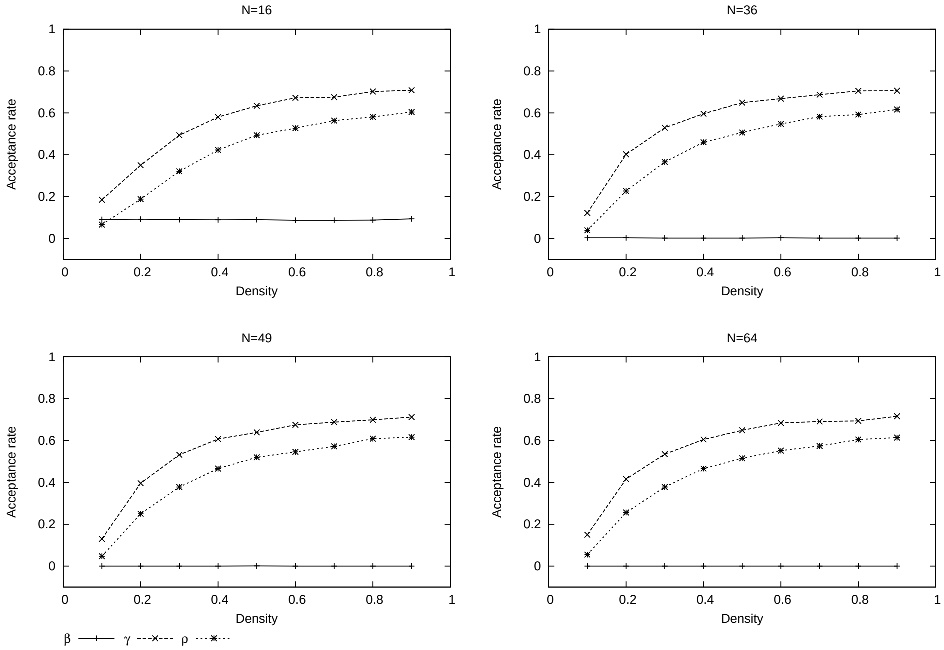

Figure 2: Acceptance rates: $T = 20$