2014

Heteroskedasticity of Unknown Form in Spatial Autoregressive Models with Moving Average Disturbance Term

Osman Dogan
CUNY Graduate Center

CUNY GRADUATE CENTER PH.D PROGRAM IN ECONOMICS
WORKING PAPER SERIES

Heteroskedasticity of Unknown Form in Spatial Autoregressive Models with Moving Average Disturbance Term

Osman Dogan

Working Paper 2, Revised

Ph.D. Program in Economics
CUNY Graduate Center
365 Fifth Avenue
New York, NY 10016

December 2013
Revised December 2014

Heteroskedasticity of Unknown Form in Spatial Autoregressive Models with Moving Average Disturbance Term

Osman Dogan

CUNY Graduate Center Working Paper No. 2
December 2013
JEL No. C13, C21, C31

ABSTRACT

In this study, I investigate the necessary condition for consistency of the maximum likelihood estimator (MLE) of spatial models with a spatial moving average process in the disturbance term. I show that the MLE of spatial autoregressive and spatial moving average parameters is generally inconsistent when heteroskedasticity is not considered in the estimation. I also show that the MLE of parameters of exogenous variables is inconsistent and determine its asymptotic bias. I provide simulation results to evaluate the performance of the MLE. The simulation results indicate that the MLE imposes a substantial amount of bias on both autoregressive and moving average parameters.

Osman Dogan
Ph.D. Program in Economics
City University of New York
365 Fifth Avenue
New York, NY 10016
ODogan@gc.cuny.edu

1 Introduction

The spatial dependence among the disturbance terms of a spatial model is generally assumed to take the form of a spatial autoregressive process. The spatial model that has a spatial lag in the dependent variable and an autoregressive process in the disturbance term is known as the SARAR model. The main characteristic of an autoregressive process is that the effect of a location-specific shock transmits to all other locations with its effects gradually fading away for the higher order neighbors. The spatial autoregressive process may not be appropriate if there is strong evidence of localized transmission of shocks. That is, the autoregressive process is not the correct specification when the effects of shocks are contained within a small region and are not transmitted to other regions. An alternative to an autoregressive process is a spatial moving average process, where the effect of shocks are more localized. Haining (1978), Anselin (1988) and more recently Hepple (2003) and Fingleton (2008a,b) consider a spatial moving average process for the disturbance terms. The spatial model that contains a spatial lag of the dependent variable and a spatial moving average process for the disturbance term is known as the SARMA model.

In the literature, various estimation methods have been proposed (Das, Kelejian, and Prucha, 2003; Hepple, 1995a; Kelejian and Prucha, 1998, 1999, 2010; Lee, 2004, 2007a; Lee and Liu, 2010; LeSage and Pace, 2009; Lesage, 1997; Liu, Lee, and Bollinger, 2010). The ML method is the best known and most common estimator used in the literature for both SARAR and SARMA specifications. Lee (2004) shows the first order asymptotic properties of the MLE for the case of SARAR(1,0). The generalized method of moment (GMM) estimators are also considered for the estimation of the spatial models. Kelejian and Prucha (1998, 1999) suggest a two step GMM estimator for the SARAR(1,1) specification. One disadvantage of the two-step GMME is that it is usually inefficient relative to the MLE (Lee, 2007c; Liu, Lee, and Bollinger, 2010; Prucha, 2014).

To increase efficiency, Lee (2007a), Liu, Lee, and Bollinger (2010) and Lee and Liu (2010) formulate one step GMMEs based on a set of moment functions involving linear and quadratic moment functions. In this approach, the reduced form of spatial models motivates the formulation of moment functions. The reduced equations indicate that the endogenous variable, i.e., the spatial lag term, is a function of a stochastic and a non-stochastic term. The linear moment functions are based on the orthogonality condition between the non-stochastic term and the disturbance term, while the quadratic moment functions are formulated for the stochastic term. Then, the parameter vector is estimated simultaneously with a one-step GMME. Lee (2007a) shows that the one-step GMME can be asymptotically equivalent to the MLE when disturbance terms are i.i.d. normal. In the case where disturbances are simply i.i.d., Liu, Lee, and Bollinger (2010) and Lee and Liu (2010) suggest a one-step GMME that can be more efficient than the (quasi) MLE.

Fingleton (2008a,b) extends the two-step GMME suggested by Kelejian and Prucha (1998, 1999) for spatial models that have a moving average process in the disturbance term, i.e., SARMA(1,1). Baltagi and Liu (2011) modify the moment functions considered in Fingleton (2008a) in the manner of Arnold and Wied (2010), and suggest a GMME for the case of SARMA(0,1). The spatial moving average parameter in both Fingleton (2008a) and Baltagi and Liu (2011) is estimated by a nonlinear least squares estimator (NLSE). The asymptotic distribution for the NLSE of the spatial moving average parameter is not provided in either Fingleton (2008a) or Baltagi and Liu (2011). Recently, Kelejian and Prucha (2010) and Drukker, Egger, and Prucha (2013) provide a basic theorem regarding the asymptotic distribution of their estimator under fairly general conditions. The estimation approach suggested in Kelejian and Prucha (2010) and Drukker, Egger, and Prucha (2013) can easily be adapted for the estimation of the SARMA(1,1) and SARMA(0,1) models. Finally, although the Kelejian and Prucha approach in Fingleton (2008a) and Baltagi and Liu (2011) has computational advantages, it may be inefficient relative to the ML method.[1]

In the presence of an unknown form of heteroskedasticity, Lin and Lee (2010) show that the MLE for the cases of SARAR(1,0) may not be consistent as the log-likelihood function is not maximized at the true parameter vector. They suggest a robust GMME for the SARAR(1,0) specification by modifying the moment functions considered in Lee (2007a). Likewise, Kelejian and Prucha (2010) modify the moment functions of their previous two-step GMME to allow for an unknown form of heteroskedasticity.

The spatial moving average model introduces a different interaction structure. Therefore, it is of interest to investigate implications of a moving average process for estimation and testing issues. In this paper, I investigate the effect of heteroskedasticy on the MLE for the case of SARMA(1,1) and SARMA(0,1) along the lines of Lin and Lee (2010). The analytical results show that when heteroskedasticity is not considered in the estimation, the necessary condition for the consistency of the MLE is generally not satisfied for both SARMA(1,1) and SARMA(0,1) models. For the SARMA(1,1) specification, I also show that the MLE of other parameters is also inconsistent, and I determine its asymptotic bias. My simulation results indicate that the MLE imposes a substantial amount of bias on spatial autoregressive and moving average parameters. However, the simulation results also show that the MLE of other parameters reports a negligible amount of bias in large samples.

The rest of this paper is organized as follows. In Section 2, I specify the SARMA(1,1) model in more detail and list assumptions that are required for the asymptotic analysis. In Section 3, I briefly discuss implications of spatial processes proposed for the disturbance term in the literature. Section 4 investigates the necessary condition for the consistency of the MLE of autoregressive and moving average parameters. Section 5 provides expressions for the asymptotic bias of the MLE of parameters of the exogenous variables. Section 6 contains a small Monte Carlo simulation. Section 7 closes with concluding remarks.

2 Model Specification and Assumptions

In this study, the following first order SARMA(1,1) specification is considered:

$$Y_n = \lambda_0 W_n Y_n + X_n \beta_0 + u_n, \quad u_n = \varepsilon_n - \rho_0 M_n \varepsilon_n \tag{2.1}$$

[1]Fingleton (2008a) and Baltagi and Liu (2011) do not compare the finite sample efficiency of their estimators with the MLE.

where Y_n is an $n \times 1$ vector of observations of the dependent variable, X_n is an $n \times k$ matrix of non-stochastic exogenous variables, with an associated $k \times 1$ vector of population coefficients β_0, W_n and M_n are $n \times n$ spatial weight matrices of known constants with zero diagonal elements, and ε_n is an $n \times 1$ vector of disturbances. The variables $W_n Y_n$ and $M_n \varepsilon_n$ are known as the spatial lag of the dependent variable and the disturbance term, respectively. The spatial effect parameters λ_0 and ρ_0 are known as the spatial autoregressive and moving average parameters, respectively. As the spatial data is characterized with triangular arrays, the variables in (2.1) have subscript n.[2] The model specifications with $\lambda_0 \neq 0$, $\rho_0 \neq 0$, and $\lambda_0 = 0$, $\rho \neq 0$ are known, respectively, as SARMA(1,1) and SARMA(0,1) in the literature. Let Θ be the parameter space of the model. In order to distinguish the true parameter vector from other possible values in Θ, the model is stated with the true parameter vector $\theta_0 = \left(\beta_0', \delta_0'\right)'$ with $\delta_0 = (\lambda_0, \rho_0)'$.

For notational simplicity, we denote $S_n(\lambda) = (I_n - \lambda W_n)$, $R_n(\rho) = (I_n - \rho M_n)$, $G_n(\lambda) = W_n S_n^{-1}(\lambda)$, $H_n(\rho) = M_n R_n^{-1}(\rho)$, $\bar{X}_n(\rho) = R_n^{-1}(\rho) X_n$, and $\bar{G}_n(\delta) = R_n^{-1}(\rho) G_n(\lambda) R_n(\rho)$. Also, at the true parameter values (ρ_0, λ_0), we denote $S_n(\lambda_0) = S_n$, $R_n(\rho_0) = R_n$, $G_n(\lambda_0) = G_n$, $H_n(\rho_0) = H_n$, $\bar{X}_n(\rho_0) = \bar{X}_n$, and $\bar{G}_n(\delta_0) = \bar{G}_n$.

The model in (2.1) is considered under the following assumptions.

Assumption 1. *The elements ε_{ni} of the disturbance term ε_n are distributed independently with mean zero and variance σ_{ni}^2, and $\mathrm{E}\left|\varepsilon_{ni}\right|^{\nu} < \infty$ for some $\nu > 4$ for all n and i.*

The elements of the disturbance term have moments higher than the fourth moment. The existence moments condition is required for the application of the central limit theorem for the quadratic form given in Kelejian and Prucha (2010). In addition, the variance of a quadratic form in ε_n exists and is finite when the first four moments are finite. Finally, Liapunov's inequality guarantees that the moments less than ν are also uniformly bounded for all n and i.

Assumption 2. *The spatial weight matrices M_n and W_n are uniformly bounded in absolute value in row and column sums. Moreover, S_n^{-1}, $S_n^{-1}(\lambda)$, R_n^{-1} and $R_n^{-1}(\rho)$ exist and are uniformly bounded in absolute value in row and column sums for all values of ρ and λ in a compact parameter space.*

The uniform boundedness of terms in Assumption 2 is motivated to control spatial autocorrelations in the model at a tractable level (Kelejian and Prucha, 1998).[3] Assumption 2 also implies that the model in (2.1) represents an equilibrium relation for the dependent variable. By this assumption, the reduced form of the model becomes feasible as $Y_n = S_n^{-1} X_n \beta_0 + S_n^{-1} R_n \varepsilon_n$. The uniform boundedness of $S_n^{-1}(\lambda)$ and $R_n^{-1}(\rho)$ in Assumption 2 is only required for the MLE, not for the GMME (Liu, Lee, and Bollinger, 2010). When W_n is row normalized, a closed subset of interval $(1/\lambda_{min}, 1)$, where λ_{min} is the smallest eigenvalue of W_n, can be considered as the parameter space for λ_0. Analogously, a closed subset of $(1/\rho_{min}, 1)$, where ρ_{min} is the smallest eigenvalue of M_n, can be the parameter space of ρ_0 (LeSage and Pace, 2009, p.128).[4]

The next assumption states the regularity conditions for the exogenous variables.

[2]See Kelejian and Prucha (2010).

[3]For a definition and some properties of uniform boundedness, see Kelejian and Prucha (2010).

[4]There are some other formulations for the parameter spaces in the literature. For details see Kelejian and Prucha

Assumption 3. *The matrix X_n is an $n \times k$ matrix consisting of constant elements that are uniformly bounded. It has full column rank k. Moreover, $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ and $\lim_{n\to\infty} \frac{1}{n} \bar{X}_n'(\rho)\bar{X}_n(\rho)$ exist and are nonsingular for all values of ρ in a compact parameter space.*

3 Spatial Processes for the Disturbance Term

In the literature, there are three main parametric processes to model spatial autocorrelation among disturbance terms: (i) spatial autoregressive process (SAR), (ii) spatial moving average process (SMA), and (iii) spatial error components model (SEC). The implied covariance structure is different under each specification. In this section, I describe the transmission and the effect of shocks under each specification. The SAR process is specified as

$$u_n = \rho_0 M_n u_n + \varepsilon_n, \tag{3.1}$$

where u_n is an $n \times 1$ vector of regression disturbances, and ε_n is an $n \times 1$ vector of i.i.d. innovations with variance σ_0^2. Under the assumption of an equilibrium, i.e., R_n is invertible, the reduced from of (3.1) is $u_n = R_n^{-1}\varepsilon_n$ with the covariance matrix of $\mathrm{E}\left(u_n u_n'\right) = \Omega_n = \sigma_0^2 R_n^{-1} R_n^{-1'}$. Note that even if the innovations are homoskedastic, the diagonal elements of Ω_n are not equal suggesting heteroskedasticity for the regression disturbances. An expansion of $(I_n - \rho_0 M_n)^{-1}$ for $|\rho_0| < 1$ yields $(I_n - \rho_0 M_n)^{-1} = \sum_{j=0}^{\infty} \rho_0^j M_n^j = I_n + \rho_0 M_n + \rho_0^2 M_n^2 + \cdots$. Hence, the SAR specification of the disturbance term implies that a shock at location i is transmitted to all other locations. The first term I_n implies that the shock at location i directly affects location i, and through other terms denoted by the powers of M_n affects higher order neighbors. Eventually, the shock feeds back to location i through the interconnectedness of neighbors. Note that $|\rho_0| < 1$ ensures that the magnitude of the transmitted shock decreases for the higher orders of neighbors. As a result, the SAR specification allows researchers to model global transmission of shocks where the full effect of a shock to location i is the sum of initial shock and the feedback from other locations.

If a more localized spatial dependence is conjectured for an economic model, then a spatial moving average process (SMA) specification is more suitable (Fingleton, 2008a,b; Haining, 1978; Hepple, 2003). The SMA process is specified as

$$u_n = \varepsilon_n - \rho_0 M_n \varepsilon_n, \tag{3.2}$$

where ρ_0 is the spatial moving average parameter. The reduced form does not involve an inverse of a square matrix. Hence, the transmission of a shock emanated from location i is limited to its immediate neighbors given by the nonzero elements in the ith row of M_n. Under this specification, the covariance matrix of u_n is $\Omega_n = \sigma_0^2 R_n R_n' = \sigma_0^2 \left(I_n - \rho_0 (M_n + M_n') + \rho_0^2 M_n M_n'\right)$. The spatial

(2010) and LeSage and Pace (2009). Note that the parameter spaces for β_0 and σ_0^2 are not required to be compact. As shown in (4.3a) and (4.3b), the MLE of these parameters is an OLS type estimator, hence boundedness is enough for the parameter spaces.

covariance is limited to nonzero elements of $\left(M_n + M_n'\right)$ and $M_n M_n'$. In comparison with the SAR specification, the range of covariance induced by the SMA model is much smaller.

Kelejian and Robinson (1993) suggest another specification which is called the spatial error components (SEC) model. This specification is similar to the SMA process in the sense that the implied covariance matrix does not involve a matrix inverse. Formally, the SEC model is given by $u_n = M_n \varepsilon_n + \epsilon_n$, where ε_n is an $n \times 1$ vector of regional innovations, whereas ϵ_n is an $n \times 1$ vector of locational innovations. Assuming that ε_n and ϵ_n are independent, the variance-covariance matrix becomes $\Omega_n = \sigma_\epsilon^2 I_n + \sigma_\varepsilon^2 M_n M_n'$, which indicates that the spatial correlation in a SEC specification is even more localized.

There have been some direct attempts to parametrize the covariance matrix of u_n, rather than defining a process for the disturbance term. For example, Besag (1974) considers a conditional first-order autoregressive model (CAR(1)) such that the covariance matrix of u_n takes the form of $\Omega_n = \sigma_0^2 (I_n - \rho_0 M_n)^{-1}$, where M_n is assumed to be a symmetric contiguity matrix. This covariance structure implies a process of $u_n = (I_n - \rho_0 M_n)^{-1/2} \varepsilon_n$. As in the case of the SAR process, a shock in a location is transmitted to all other locations, but now with a smaller amplitude. Another example is $\Omega_n = \sigma_0^2 (I_n + \rho_0 M_n)$, where M_n is assumed to be symmetric (Hepple, 1995b; Richardson, Guihenneuc, and Lasserre, 1992). In this case, the spatial correlation is restricted to first order neighbors, i.e., non-zero elements of M_n.

The elements of Ω_n can also be specified through a covariance generating function. For example, in Ripley (2005), the covariance generating function is defined in terms of distance between two locations in such a way that the resulting covariance is always non-negative definite. Let d_{ij} be the distance between location i and j, and $\Omega_{ij,n}$ be the covariance between these two locations. Then, the covariance generating function is defined by

$$\Omega_{ij,n} = \begin{cases} \sigma_0^2 \frac{2}{n} \left[\cos^{-1}\left(\frac{d_{ij}}{2\psi}\right) - \frac{d_{ij}}{2\psi}\left(1 - \frac{d_{ij}^2}{4\psi^2}\right)^{1/2} \right], & \text{if } d_{ij} \leq 2\psi \\ 0, & \text{otherwise.} \end{cases} \tag{3.3}$$

Intuitively, $\Omega_{ij,n}$ is proportional to the intersection area of two discs of common radius centered on locations i and j. The covariance generating function in (3.3) depends on the single parameter ψ, and has a fairly linear negative relationship with d_{ij} (Richardson, Guihenneuc, and Lasserre, 1992; Ripley, 2005). Another covariance generating function family, first introduced by Whittle in 1954, is a two parameter functions defined in terms of Gamma and Bessel functions. This family has the following specification:

$$\Omega_{ij,n} = \sigma_0^2 \left[2^{\nu-1}\Gamma(\nu)\right]^{-1} (\delta d_{ij})^{\nu} K_\nu(\delta d_{ij}), \tag{3.4}$$

where $K_\nu(\cdot)$ is the modified Bessel function, and $\Gamma(\cdot)$ is the standard Gamma function. The parameters $\nu > 0$ and $\delta > 0$ are respectively known as a shape parameter and a spatial parameter. The spatial parameter δ determines how far the spatial correlation will stretch. For the special case, where $\nu = \frac{1}{2}$, this covariance generating function gives an exponential decaying spatial correlations (Richardson, Guihenneuc, and Lasserre, 1992). There is also a more general exponential covariance generating function that depends on two parameters. This function is specified by $\Omega_{ij,n} = \sigma_0^2 \gamma \exp(\lambda d_{ij})$, where γ and λ are parameters need to be estimated. This function also exhibits exponential decay for the spatial correlations.

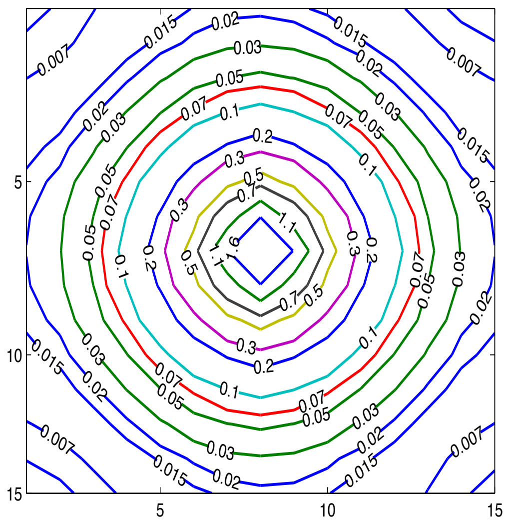
In the literature, there are some other covariance generating function families. However, the majority of these functions do not necessarily ensure that Ω_n is a positive definite matrix (Haining, 1987; Richardson, Guihenneuc, and Lasserre, 1992). The formal properties of the MLE for spatial models that have a covariance structure determined by a parametric function are investigated in an early study by Mardia and Marshall (1984). In this study, the authors state conditions under which the MLE is consistent and has asymptotic normal distribution.

In this study, the spatial model specified in (2.1) is considered. The interaction between the spatial autoregressive process and the moving average process for this model induces a complicated pattern for the transmission of a location specific shock. Under Assumption 2, the reduced form of the model is given by $Y_n = S_n^{-1} X_n \beta_0 + S_n^{-1} R_n \varepsilon_n$. The last term in the reduced form can be written as $S_n^{-1} R_n \varepsilon_n = \varepsilon_n - \rho_0 M_n \varepsilon_n + \sum_{l=1}^{\infty} \lambda_0^l W_n^l \varepsilon_n - \rho_0 M_n \sum_{l=1}^{\infty} \lambda_0^l W_n^l \varepsilon_n$. In this representation, the higher power of W_n does not have zero diagonal elements, which in turn implies that the total effect of a region specific shock also contains the feedback effects passed through other locations. The corresponding expression in the case of SARAR(1,1) specification is given by $S_n^{-1} R_n^{-1} \varepsilon_n = \sum_{l=0}^{\infty} \lambda_0^l W_n^l \sum_{k=0}^{\infty} \rho_0^k M_n^k \varepsilon_n$. Again, the induced pattern involves the interaction of two weight matrices and two parameters.

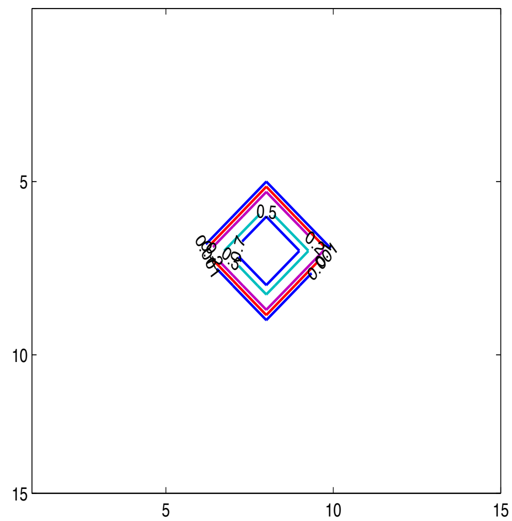
Following Fingleton (2008a), I illustrate the transmission pattern for a shock under each specification by using a rook weight matrix over a 15×15 lattice. Figure 1 shows the impact of a shock emanated from the unit located at the center of lattice.[5] In the case of SAR and SARAR(1,1), the effect of shock is more vigorous over the whole lattice. For the SMA specification, the shock is only transmitted to the immediate units as shown in Figure 1(b). In contrast, the effect of the shock gradually dies out under the SARMA(1,1) model.

[5]For easy comparison, we set $\lambda_0 = 0.9$ for SAR, $\rho_0 = -0.9$ for SMA, $(\lambda_0, \rho_0) = (0.5, 0.9)$ for SARAR(1,1) and $(\lambda_0, \rho_0) = (0.5, -0.9)$ for SARMA(1,1). The disturbance of the unit located at the center of the lattice is increased by 3.

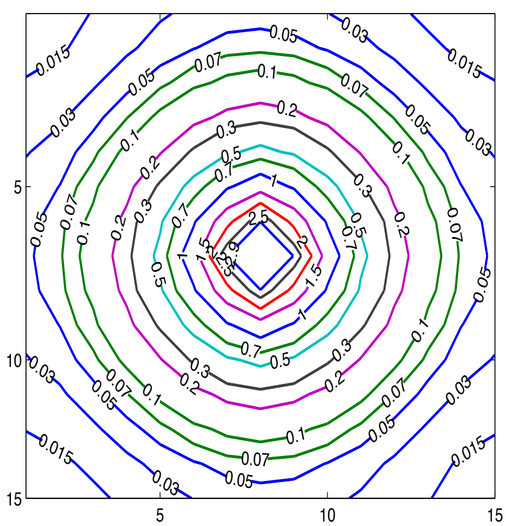
Figure 1: The Effect of a Shock



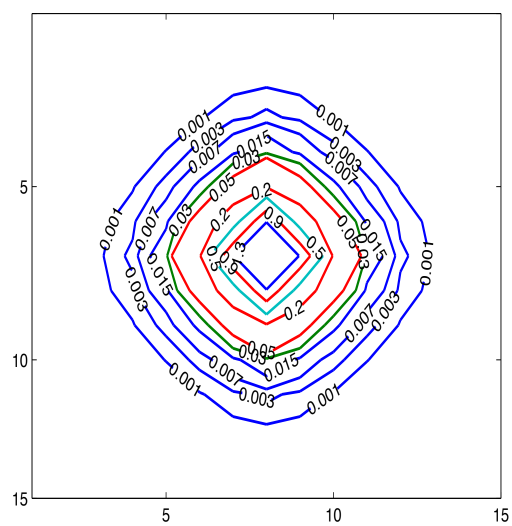
(a) The effect of a shock: SAR



(b) The effect of a shock: SMA



(c) The effect of a shock: SARAR(1,1)



(d) The effect of a shock: SARMA(1,1)

4 The MLE of λ_0 and ρ_0

The log-likelihood function for the model in (2.1) under the assumption that the disturbance terms of the model are i.i.d. normal with mean zero and variance σ_0^2 can be written as

$$\begin{aligned}\ln L_n(\zeta) = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) + \ln|S_n(\lambda)| - \ln|R_n(\rho)| \\ &- \frac{1}{2\sigma^2}\left(S_n(\lambda)Y_n - X_n\beta\right)' R_n'^{-1}(\rho)R_n^{-1}(\rho)\left(S_n(\lambda)Y_n - X_n\beta\right),\end{aligned} \tag{4.1}$$

where $\zeta = \left(\theta', \sigma^2\right)'$. The first order conditions with respect to β and σ^2 are respectively given by

$$\frac{\partial \ln L_n(\zeta)}{\partial \beta} = \frac{1}{\sigma^2}\bar{X}_n'(\rho)R_n^{-1}(\rho)\left(S_n(\lambda)Y_n - X_n\beta\right), \tag{4.2a}$$

$$\frac{\partial \ln L_n(\zeta)}{\partial \sigma^2} = \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\varepsilon_n'(\theta)\varepsilon_n(\theta), \tag{4.2b}$$

where $\varepsilon_n(\theta) = R_n^{-1}(\rho)\left(S_n(\lambda)Y_n - X_n\beta\right)$. The solutions of the first order conditions for a given δ yield the MLE of β_0 and σ_0^2:

$$\hat{\beta}_n(\delta) = \left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)R_n^{-1}(\rho)S_n(\lambda)Y_n, \tag{4.3a}$$

$$\hat{\sigma}_n^2(\theta) = \frac{1}{n}\varepsilon_n'(\theta)\varepsilon_n(\theta). \tag{4.3b}$$

Concentrating the log-likelihood function by eliminating σ^2 gives the following equation:

$$\ln L_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2} - \frac{n}{2}\ln\left(\frac{\varepsilon_n'(\theta)\varepsilon_n(\theta)}{|S_n(\lambda)|^{\frac{2}{n}}|R_n(\rho)|^{-\frac{2}{n}}}\right) \tag{4.4}$$

The above representation is useful for exploring the role of the Jacobian terms $|S_n(\lambda)|$ and $|R_n(\rho)|$ in the ML estimation. The MLE of θ is the extremum estimator obtained from the maximization of (4.4). In an equivalent way, the MLE of θ_0 can be defined by

$$\hat{\theta}_n = \operatorname{argmin}_{\theta\in\Theta}\left\{\frac{\varepsilon_n'(\theta)\varepsilon_n(\theta)}{|S_n(\lambda)|^{\frac{2}{n}}|R_n(\rho)|^{-\frac{2}{n}}}\right\} \tag{4.5}$$

In the special case, where $|S_n(\lambda)| = |R_n(\rho)| = 1$, the MLE is the NLSE obtained from the minimization of $\varepsilon_n'(\theta)\varepsilon_n(\theta)$, i.e., $\hat{\theta}_{NLSE,n} = \operatorname{argmin}_{\theta\in\Theta}\varepsilon_n'(\theta)\varepsilon_n(\theta)$. It is clear that the Jacobian terms $|S_n(\lambda)|$ and $|R_n(\rho)|$ play a role of a weight (or a penalty) on $\varepsilon_n'(\theta)\varepsilon_n(\theta)$. The penalty is a function of the autoregressive parameters and the spatial weight matrices, which can be defined as $f\left(\lambda, \rho, W_n, M_n\right) = |S_n(\lambda)|^{\frac{2}{n}}|R_n(\rho)|^{-\frac{2}{n}}$. For the SARAR(1,1) specification, the last term in (4.4) is given by $-\frac{n}{2}\ln\left(\frac{\varepsilon_n'(\theta)\varepsilon_n(\theta)}{|S_n(\lambda)|^{\frac{2}{n}}|R_n(\rho)|^{\frac{2}{n}}}\right)$, where $\varepsilon_n(\theta) = R_n(\rho)\left(S_n(\lambda)Y_n - X_n\beta\right)$. Therefore, in the case of SARAR(1,1), the MLE of θ_0 is given by

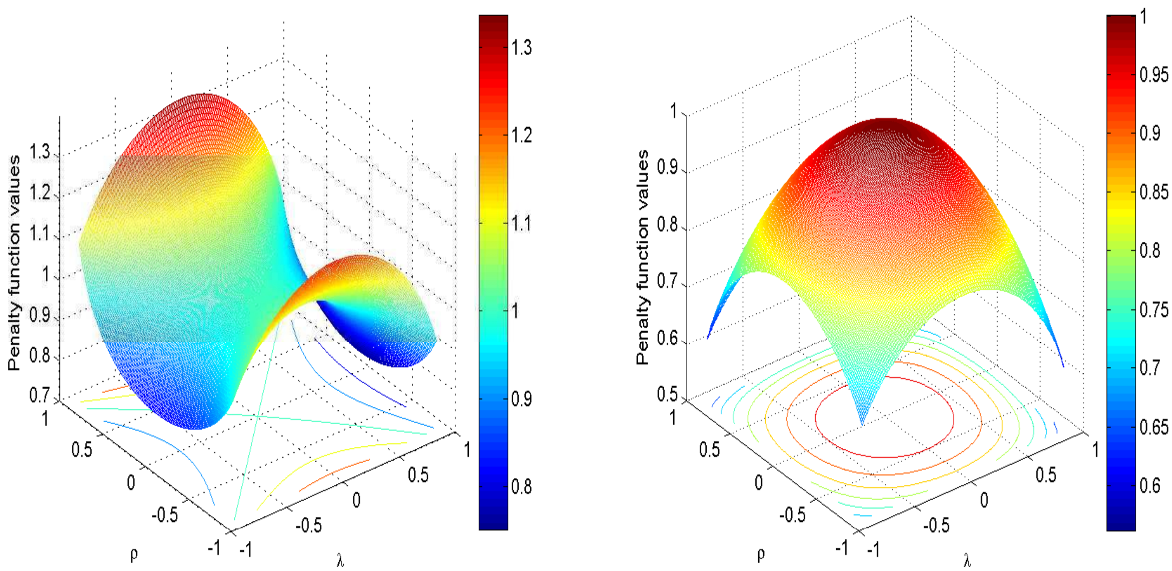
$$\hat{\theta}_n = \operatorname{argmin}_{\theta\in\Theta}\left\{\frac{\varepsilon_n'(\theta)\varepsilon_n(\theta)}{|S_n(\lambda)|^{\frac{2}{n}}|R_n(\rho)|^{\frac{2}{n}}}\right\} \tag{4.6}$$

It is hard to make any general statement about the effects and magnitudes of the penalty functions in both cases. Hepple (1976) illustrates that the Jacobian term imposes a substantial penalty for the SARAR(0,1) specification. To illustrate the effect of penalty functions for the case of SARMA(1,1) and SARAR(1,1), I use a distance based weight matrix for a sample of 91 countries such that each country is connected to every other country. The elements of the weight matrices are specified by

$$w_{ij} = m_{ij} = \begin{cases} 0 \quad \text{if} \quad i = j, \\ \frac{d_{ij}^{-2}}{\sum_{j=1}^{91} d_{ij}^{-2}} \quad \text{if} \quad i \neq j, \end{cases} \tag{4.7}$$

where d_{ij} between countries i and j is measured by the great-circle distance between country capitals.[6] Figure 2 shows the surface plots of penalty functions over a grid of spatial parameters.

Figure 2: The penalty functions for the dense weight matrix



(a) The penalty function for SARMA(1,1)

(b) The penalty function for SARAR(1,1)

For the SARAR(1,1) specification, the value of the penalty function decreases as the parameter combination (λ, ρ) moves away from $(0, 0)$ in any direction as shown in Figure 2(b).[7] On the

[6] $d_{ij} = R_0 \times \arccos\left(\cos\left(|longitude_i - longitude_j|\right)\cos(latitude_i)\cos(latitude_j) + \sin(latitude_i)\sin(latitude_j)\right)$, where R_0 is the Earth's radius.

[7] For SARAR(1,1), the penalty function is $f(\lambda, \rho, W_n, M_n) = |S_n(\lambda)|^{\frac{2}{n}}|R_n(\rho)|^{\frac{2}{n}}$.

other hand, there is no such monotonic decrease in the penalty function under the SARMA(1,1) specification as illustrated in Figure 2(a). The penalty function of SARMA(1,1) obtains relatively larger values when there is strong spatial dependence in the disturbance term, i.e., when ρ is near 1 or -1. In contrast, the penalty function has smaller values when there is strong spatial dependence in the dependent variable. This pattern indicates that the sum $\varepsilon_n^{'}(\theta)\varepsilon_n(\theta)$ is penalized as ρ moves toward to either 1 or -1. In the case of SARAR(1,1), this sum gets larger as (λ, ρ) moves toward $(\pm 1, \pm 1)$ in any direction, suggesting that the solution of the minimization problem is restricted to the region $(-1,-1) \times (+1,+1)$. Finally, in a small neighborhood of $(0,0)$, the surface plots in Figure 2 indicate that the penalty functions take values around 1, suggesting that the parameter estimates from the MLE can be similar to those from the NLSE under both specifications.

Next, I investigate the effect of heteroskedasticity on the MLE for the case of SARMA(1,1). I assume that the true data generating process is characterized by Assumption 1. More explicitly, the MLE $\hat{\sigma}_n^2(\delta)$ can be written as

$$\hat{\sigma}_n^2(\delta) = \frac{1}{n} Y_n^{'} S_n^{'}(\lambda) R_n^{'-1}(\rho) \bar{\mathbb{M}}_n(\rho) R_n^{-1}(\rho) S_n(\lambda) Y_n, \tag{4.8}$$

where $\bar{\mathbb{M}}_n(\rho) = (I_n - \mathbb{P}_n(\rho))$ is a projection type matrix with $\mathbb{P}_n(\rho) = \bar{X}_n(\rho)\left(\bar{X}_n^{'}(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n^{'}(\rho)$. Substituting $R_n^{-1}(\rho)S_n(\lambda)Y_n = R_n^{-1}(\rho)X_n\beta + \varepsilon_n$ into $\hat{\sigma}_n^2(\delta)$ and using the fact that $\bar{X}_n^{'}(\rho)\bar{\mathbb{M}}_n(\rho) = 0_{k\times n}$ and $\bar{\mathbb{M}}_n(\rho)\bar{X}_n(\rho) = 0_{n\times k}$, the MLE $\hat{\sigma}_n^2(\delta)$ can be written as

$$\hat{\sigma}_n^2(\delta) = \frac{1}{n}\varepsilon_n^{'}\bar{\mathbb{M}}_n(\rho)\varepsilon_n. \tag{4.9}$$

At δ_0, the probability limit of $\hat{\sigma}_n^2(\delta_0)$ is

$$\operatorname*{plim}_{n\to\infty} \hat{\sigma}_n^2(\delta_0) = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\varepsilon_n^{'}\varepsilon_n - \operatorname*{plim}_{n\to\infty} \frac{1}{n^2}\varepsilon_n \bar{X}_n \left(\frac{1}{n}\bar{X}_n^{'}\bar{X}_n\right)^{-1}\bar{X}_n^{'}\varepsilon_n. \tag{4.10}$$

For the first term on the right hand side, we have $\frac{1}{n}\varepsilon_n^{'}\varepsilon_n = \frac{1}{n}\sum_{i=1}^{n}\sigma_{ni}^2 + o_p(1)$ by Chebyshev's Weak Law of Large Numbers. The second term vanishes by virtue of Lemma 1(4) in Appendix 8.1, and Assumption 3. Therefore, we have

$$\hat{\sigma}_n^2(\delta_0) = \frac{1}{n}\sum_{i=1}^{n}\sigma_{ni}^2 + o_p(1). \tag{4.11}$$

The result in (4.11) indicates that the average of the individual variances is asymptotically equivalent to $\hat{\sigma}_n^2(\delta_0)$.

Concentrating out β and σ^2 from the log-likelihood function in (4.1) yields

$$\ln L_n(\delta) = -\frac{n}{2}(\ln(2\pi)+1) - \frac{n}{2}\ln\hat{\sigma}_n^2(\delta) + \ln|S_n(\lambda)| - \ln|R_n(\rho)|. \tag{4.12}$$

The MLE $\hat{\lambda}_n$ and $\hat{\rho}_n$ are extremum estimators obtained from the maximization of (4.12). The first order conditions of (4.12) with respect to ρ and λ are[8]

$$\frac{\partial \ln L_n(\delta)}{\partial \lambda} = -\frac{n}{2\hat{\sigma}_n^2(\delta)} \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \lambda} - \operatorname{tr}(G_n(\lambda)) \tag{4.13a}$$

$$\frac{\partial \ln L_n(\delta)}{\partial \rho} = -\frac{n}{2\hat{\sigma}_n^2(\delta)} \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \rho} + \operatorname{tr}(H_n(\rho)), \tag{4.13b}$$

where $G_n(\lambda) = W_n S_n^{-1}(\lambda)$ and $H_n(\rho) = M_n R_n^{-1}(\rho)$. For the consistency of $\hat{\lambda}_n$ and $\hat{\rho}_n$, the necessary condition is $\operatorname{plim}_{n\to\infty} \frac{1}{n}\frac{\partial \ln L_n(\delta_0)}{\partial \delta} = 0$. Below, I investigate the probability limit of the following expression:

$$\frac{1}{n}\frac{\partial \ln L_n(\delta_0)}{\partial \delta} = \begin{pmatrix} \frac{1}{n}\left(-\frac{n}{\frac{2}{n}\varepsilon_n' \overline{\mathbb{M}}_n \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \lambda}\right) - \frac{1}{n}\operatorname{tr}(G_n) \\ \frac{1}{n}\left(-\frac{n}{\frac{2}{n}\varepsilon_n' \overline{\mathbb{M}}_n \varepsilon_n} \frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \rho}\right) + \frac{1}{n}\operatorname{tr}(H_n) \end{pmatrix}. \tag{4.14}$$

Under Assumption 2, both H_n and G_n are uniformly bounded in absolute value in row and column sums. Therefore, $\frac{1}{n}\operatorname{tr}(H_n)$ and $\frac{1}{n}\operatorname{tr}(G_n)$ in (4.14) are of order $O(1)$. With these results for $\frac{1}{n}\operatorname{tr}(H_n)$ and $\frac{1}{n}\operatorname{tr}(G_n)$, a convenient result for the probability limit of (4.14) can be obtained, which is stated in the following proposition.

Proposition 1. *Under Assumptions 1 through 3, we have*

$$\frac{1}{n}\frac{\partial \ln L_n(\delta_0)}{\partial \delta} = \begin{pmatrix} \frac{\operatorname{Cov}\left(\overline{G}_{n,ii},\, \sigma_{ni}^2\right)}{\overline{\sigma}^2} + o_p(1) \\ -\frac{\operatorname{Cov}\left(H_{n,ii},\, \sigma_{ni}^2\right)}{\overline{\sigma}^2} + o_p(1) \end{pmatrix}, \tag{4.15}$$

where $\operatorname{Cov}\left(\overline{G}_{n,ii},\, \sigma_{ni}^2\right)$ *is the covariance between the diagonal elements of* $\overline{G}_n$, $\{\overline{G}_{n,11}, \overline{G}_{n,22}, \ldots, \overline{G}_{n,nn}\}$, *and the individual variances* $\{\sigma_{n1}^2, \sigma_{n2}^2, \ldots, \sigma_{nn}^2\}$. *Similarly,* $\operatorname{Cov}\left(H_{n,ii},\, \sigma_{ni}^2\right)$ *denotes the covariance between diagonal elements of* H_n, $\{H_{n,11}, H_{n,22}, \ldots, H_{n,nn}\}$, *and the individual variances* $\{\sigma_{n1}^2, \sigma_{n2}^2, \ldots, \sigma_{nn}^2\}$.

Proof. See Appendix 8.2. □

The above proposition indicates that the MLE of the spatial autoregressive and moving average parameters is not consistent as long as the covariance terms in (4.15) are not zero. Notice that, when the disturbance terms are homoskedastic, the covariance terms in (8.11) are zero. In the special case where $W_n = M_n$ and $\lambda_0 = \rho_0$, we have $S_n = R_n$ and $G_n = H_n$ so that $\overline{G}_n = R_n^{-1} G_n R_n =$

[8] For these results, we use the derivative rule given by $\frac{\partial \ln |R_n(\rho)|}{\partial \rho} = \operatorname{tr}\left(R_n^{-1}(\rho) \times \frac{\partial R_n(\rho)}{\partial \rho}\right)$. For a proof, see Abadir and Magnus (2005, p.372). Also note the commutative property of $R_n^{-1}(\rho) M_n = M_n R_n^{-1}(\rho) = H_n(\rho)$.

$R_n^{-1} H_n R_n = R_n^{-1} M_n R_n^{-1} R_n = H_n$. Hence, the necessary condition for the consistency of $\hat{\lambda}_n$ is identical to the one for $\hat{\rho}_n$.

The result in Proposition 1 indicates that the consistency of the MLE depends on the specification of weight matrices. It is of interest to investigate specifications that yield zero covariances. An obvious case is when there is no variation in the diagonal elements of $\bar{G}_n$ and H_n. Then, the necessary condition for the consistency of $\hat{\lambda}_n$ and $\hat{\rho}_n$ is not violated, even if the disturbances are heteroskedastic. For example, there is no variations in the diagonal elements of $\bar{G}_n$ and H_n when W_n and M_n are block-diagonal matrices with an identical sub-matrix in the diagonal blocks and zeros elsewhere. This type of block diagonal weight matrix can be seen in social interaction scenarios where a block represents a group in which each individual is equally affected by the members of the group (Lee, 2007b; Lee, Liu, and Lin, 2010). Suppose that there are R groups each of which has m members so that $n = mR$. If we assign equal weight to each member of a group, then $W_n = M_n = I_R \otimes B_m$, where $B_m = \frac{1}{m-1}\left(l_m l_m' - I_m\right)$, and l_m is an m-dimensional column vector of ones. For this set up, there is no variation in the diagonal elements of $\bar{G}_n$ and H_n, therefore $\text{Cov}\left(\bar{G}_{n,ii},\ \sigma_{ni}^2\right) = \text{Cov}\left(H_{n,ii},\ \sigma_{ni}^2\right) = 0$.

There is also no variation in the diagonal elements of $\bar{G}_n$ and H_n when the circular world weight matrices considered in Kelejian and Prucha (1999) are employed. In these weight matrices, the order of observations is important since the observations are related to some units in front and to some in back. As an example consider a "1 ahead and 1 behind" weight matrix where each observation is related to the one immediately after and immediately before it. For this scenario, we also have $\text{Cov}\left(\bar{G}_{n,ii},\ \sigma_{ni}^2\right) = \text{Cov}\left(H_{n,ii},\ \sigma_{ni}^2\right) = 0$. The circular world weight matrices can be adjusted to create some variation in the diagonal elements of $\bar{G}_n$ and H_n. For example, Kelejian and Prucha (2007) construct a different version in which the first and the last one third of the sample observations have 5 neighbors in front and 5 in back, while the middle third only has 1 neighbor in front and 1 in back. Under this scenario, the Monte Carlo results in Kelejian and Prucha (2007) show that the MLE is significantly biased for the case of SARAR(1,1).

5 The MLE of β_0

In the previous section, I showed that the consistency of the MLE of the spatial autoregressive and moving average parameters is not ensured. In this section, I investigate the consistency of the MLE of β_0. The result in (4.3a) indicates that the MLE $\hat{\beta}_n(\hat{\delta}_n)$ is also inconsistent, since it is based on the inconsistent estimators $\hat{\lambda}_n$ and $\hat{\rho}_n$. The asymptotic bias of $\hat{\beta}_n(\hat{\delta}_n)$ can be determined from (4.3a). By using $S_n(\lambda) = S_n + (\lambda_0 - \lambda)\, W_n$, the MLE $\hat{\beta}_n(\delta)$ can be written as

$$\begin{aligned}
\hat{\beta}_n(\delta) =& \beta_0 + \left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1} \bar{X}_n'(\rho) R_n^{-1}(\rho) R_n \varepsilon_n \\
&+ (\lambda_0 - \lambda)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1} \bar{X}_n'(\rho) R_n^{-1}(\rho) G_n X_n \beta_0 \\
&+ (\lambda_0 - \lambda)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1} \bar{X}_n'(\rho) R_n^{-1}(\rho) G_n R_n \varepsilon_n,
\end{aligned} \tag{5.1}$$

Under Assumption 3, the term $\left(\frac{1}{n}\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}$ is uniformly bounded in absolute value in row and column sums. By Lemma 1(5) of Appendix 8.1, terms involving ε_n in (5.1) vanish in probability. Thus,

$$\hat{\beta}_n(\delta) = \beta_0 + (\lambda_0 - \lambda)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)R_n^{-1}(\rho)G_nX_n\beta_0 + o_p(1) \tag{5.2}$$

The asymptotic bias of $\hat{\beta}_n(\hat{\delta}_n)$ follows from (5.2), which is given by $\left(\lambda_0 - \hat{\lambda}_n\right)\left(\bar{X}_n'(\hat{\rho}_n)\bar{X}_n(\hat{\rho}_n)\right)^{-1}\bar{X}_n'(\hat{\rho}_n)R_n^{-1}(\hat{\rho}_n)G_nX_n\beta_0$. This result shows that the asymptotic bias of $\hat{\beta}_n(\hat{\delta}_n)$ depends on weight matrices and the regressors matrix, and is not zero unless spatial parameters are consistent. Note that the bias is the OLS estimator obtained from the artificial regression of $R_n^{-1}(\hat{\rho}_n)G_nX_n\beta_0$ on $\bar{X}_n'(\hat{\rho}_n)$. For the special case of $\hat{\lambda}_n = \lambda_0 + o_p(1)$, we have $\hat{\beta}_n(\delta) = \beta_0 + o_p(1)$. In this case, there is no asymptotic bias and the inconsistency of $\hat{\rho}_n$ has no effect on $\hat{\beta}_n(\hat{\delta}_n)$.

The specification with $\lambda_0 = 0$ in (2.1) is called the spatial moving average model (SARMA(0,1) or SMA). For the SARMA(0,1) specification, the log-likelihood function simplifies to

$$\ln L_n(\zeta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \ln|R_n(\rho)| - \frac{1}{2\sigma^2}\left(Y_n - X_n\beta\right)' R_n'^{-1}(\rho)R_n^{-1}(\rho)\left(Y_n - X_n\beta\right), \tag{5.3}$$

where $\zeta = \left(\theta', \sigma^2\right)'$ with $\theta = \left(\rho, \beta'\right)'$. For a given value of ρ, the first order conditions yield

$$\hat{\beta}_n(\rho) = \left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)R_n^{-1}(\rho)Y_n$$
$$\hat{\sigma}_n^2(\rho) = \frac{1}{n}\varepsilon_n'(\theta)\varepsilon_n(\theta),$$

where $\varepsilon_n(\theta) = R_n^{-1}Y_n - \bar{X}_n\beta$. The necessary condition for the consistency of the MLE $\hat{\rho}_n$ can be obtained from (4.15). From the second row of (4.15), we have $\frac{1}{n}\frac{\partial \ln L_n(\rho_0)}{\partial \rho} = -\frac{\mathrm{Cov}\left(H_{n,ii},\, \sigma_{ni}^2\right)}{\overline{\sigma}^2} + o_p(1)$, which implies that the MLE $\hat{\rho}_n$ is inconsistent. Substitution of $Y_n = X_n\beta_0 + R_n\varepsilon_n$ into $\hat{\beta}_n(\rho)$ yields

$$\hat{\beta}_n(\rho) = \beta_0 + \left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)R_n^{-1}(\rho)\varepsilon_n. \tag{5.4}$$

The variance of $\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)D_n^{-1}(\rho)\varepsilon_n$ in (5.4) has an order of $O(\frac{1}{n})$ by Lemma 1(5) of Appendix 8.1. Then, Chebyshev's inequality implies that $\hat{\beta}_n(\rho) = \beta_0 + o_p(1)$. Hence, $\hat{\beta}_n(\hat{\rho}_n)$ has no asymptotic bias even though $\hat{\rho}_n$ is inconsistent.

For the spatial autoregressive model, where $\rho_0 = 0$ in (2.1), the result in (4.15) simplifies to $\frac{1}{n}\frac{\partial \ln L_n(\lambda_0)}{\partial \lambda} = \frac{\mathrm{Cov}\left(G_{n,ii},\, \sigma_{ni}^2\right)}{\overline{\sigma}^2} + o_p(1)$. The term $\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)R_n^{-1}(\rho)G_nX_n\beta_0$ in (5.2) simplifies to $\left(X_n^{'}X_n\right)^{-1} X^{'}G_nX_n\beta_0$ so that

$$\hat{\beta}_n(\lambda) = \beta_0 + (\lambda_0 - \lambda)\left(X_n^{'}X_n\right)^{-1} X^{'}G_nX_n\beta_0 + o_p(1), \tag{5.5}$$

The result in (5.5) is the exact result stated in Lin and Lee (2010) for the case of SARAR(1,0).

I collect the above results for the MLE of β_0 in the following proposition.

Proposition 2. *Consider the model in* (2.1) *under Assumptions 1 through 3, then*

1. *For the SARMA(1,1) model, we have*

$$\hat{\beta}_n(\delta) = \beta_0 + (\lambda_0 - \lambda)\left(\bar{X}_n^{'}(\rho)\bar{X}_n(\rho)\right)^{-1} \bar{X}_n^{'}(\rho)R_n^{-1}(\rho)G_nX_n\beta_0 + o_p(1). \tag{5.6}$$

2. *For the SARMA(0,1) model, where* $\lambda_0 = 0$, *we have* $\hat{\beta}_n(\rho) = \beta_0 + o_p(1)$.

3. *For the SARMA(1,0) model, where* $\rho_0 = 0$, *we have*

$$\hat{\beta}_n(\lambda) = \beta_0 + (\lambda_0 - \lambda)\left(X_n^{'}X_n\right)^{-1} X^{'}G_nX_n\beta_0 + o_p(1). \tag{5.7}$$

In Section 4 and 5, I showed that the MLE of δ_0 and β_0 is generally inconsistent when heteroskedasticity is present in the model. Besides its computational burden, the consistency of MLE is not ensured. In the next section, I confirm these large sample results through a Monte Carlo simulation.

6 Monte Carlo Simulation

In this section, the finite sample properties of the MLE are investigated through a Monte Carlo experiment for the cases of (i) SARMA(0,1), and (ii) SARMA(1,1). For both models, we assume heteroskedastic innovations in the data generating processes.

6.1 Design

There are two regressors and no intercept term such that $X_n = [x_{n,1}, x_{n,2}]$ and $\beta_0 = (\beta_{10}, \beta_{20})'$, where $x_{n,1}$ and $x_{n,2}$ are $n \times 1$ independent random vectors that are generated from a Normal(0,1). We consider $n = 100, 500, 1000$ and let $W_n = M_n$ and set $\beta_0 = (1,\ 1)^{'}$ for all experiments. For the spatial autoregressive parameters $(\lambda_0,\ \rho_0)$, we employ combinations from the set $\mathscr{B} = (-0.6,\ -0.3,\ 0,\ 0.3,\ 0.6)$ to allow for weak and strong spatial interactions.

The row normalized spatial weight matrix is based on the small group interaction scenario described in Lin and Lee (2010). In this scenario, the weight matrix is a block diagonal matrix where each block represents a group interaction. The size of each block is determined by the group size, which are determined by a random draw from Uniform(15,50). Let $\{g_1, \ldots, g_G\}$ be the set of groups, where G is the total number of groups. Denote the size of each group by m_i for $i = 1, \ldots, G$. Then, the block for group i is given by $B_i = \frac{1}{m_i - 1}\left(l_{m_i} l'_{m_i} - I_{m_i}\right)$, where l_{m_i} is the $m_i \times 1$ vector of ones. Then, $W_n = M_n = \text{Diag}\left(B_1, \ldots, B_G\right)$.[9]

The observations in a group have the same variance, and I use the group size to create heteroskedasticity. If the group size is greater than 35, we set the variance of that group equal to its size raised to 0.4 power; otherwise we let the variance be the square of the inverse of the group size. Then, the i-th element of the innovation vector ε_n is generated according to $\varepsilon_{ni} = \sigma_{ni}\xi_{ni}$, where σ_{ni} is the standard error for the i-th observation and ξ_{ni}'s are i.i.d. Normal(0,1).

I use the following expressions to measure the level of signal-to-noise in our step up (Pace, LeSage, and Zhu, 2012):

$$R^2_{SARMA(1,1)} = 1 - \frac{\text{tr}\left(R'_n S_n^{-1\prime} S_n^{-1} R_n \Sigma_n\right)}{\beta'_0 X'_n S_n^{-1\prime} S_n^{-1} X_n \beta_0 + \text{tr}\left(R'_n S_n^{-1\prime} S_n^{-1} R_n \Sigma_n\right)}, \tag{6.1}$$

$$R^2_{SARMA(0,1)} = 1 - \frac{\text{tr}\left(R'_n R_n \Sigma_n\right)}{\beta'_0 X'_n X_n \beta_0 + \text{tr}\left(R'_n R_n \Sigma_n\right)}, \tag{6.2}$$

where Σ_n is the diagonal $n \times n$ covariance matrix of the disturbance terms. This set-up yields an R^2 value close to 0.55. For each specification, the Monte Carlo experiment is based on 1000 repetitions.

6.2 Simulation Results

The simulation results are presented in Appendix 8.3 and 8.4. In each table, the empirical mean (Mean), the bias (Bias), the empirical standard error (Std.), and the root mean square error (RMSE) of parameter estimates are presented next to each other.

First, we consider the simulation results for the SARMA (0,1) model. The simulation results are presented in Table 1 of Appendix 8.3. The MLE imposes almost no bias on β_{10} and β_{20} in all cases. The moving average parameter ρ_0 has substantial amount of bias when $n = 100$, but the amount of bias decreases as the sample size increases. Despite this, the MLE imposes significant amount of bias on ρ_0 when $n = 500$ and $n = 1000$ in cases where the true value of ρ_0 is nonzero. Overall, the simulation results are consistent with our large sample results. That is, the MLE of β_{10} and β_{20} is consistent, while the MLE of ρ_0 is inconsistent in the presence of heteroskedasticity.

Now, we turn to the simulation results for the case of SARMA(1,1). First, we consider the simulation results for λ_0 and ρ_0. Table 2 shows the estimation results for $n = 100$. The MLE imposes substantial amount of bias on both parameters in all cases. The amount of bias for λ_0 is relatively larger when there exists a strong negative spatial dependence in the dependent variable. There is a similar pattern for ρ_0, where the amount of bias and RMSE is, in general, larger for the

[9] Here, $\text{Diag}\left(B_1, \ldots, B_G\right)$ denotes the block diagonal matrix in which the diagonal blocks are $m_i \times m_i$ matrices B_is.

cases of high negative spatial dependence in both dependent variable and disturbance term. The pattern that we see for λ_0 and ρ_0 shows itself for the estimation results of β_{10} and β_{20}. That is, the reported biases and RMSEs are relatively larger for β_{10} and β_{20}, when there are strong spatial dependence in the model.

Table 3 contains the simulation results when $n = 500$. The same pattern that I described for λ_0 and ρ_0 is also prevalent in Table 3. The MLE still imposes substantial amount of bias on λ_0 and ρ_0. The noticeable improvement in the estimation results for β_{10} and β_{20} suggests that these parameters are less affected by the inconsistency of the MLE of λ_0 and ρ_0, when the sample size is moderately large. The estimation results in Table 4 for β_{10} and β_{20} are also consistent with this claim. That is, when the sample size is large , i.e., $n = 1000$, the MLE imposes trivial bias on β_{10} and β_{20} in most cases. On the other hand, the estimation results in Table 4 show that the MLE imposes significant bias on λ_0 and ρ_0, which in turn implies the inconsistency of the MLE for these parameters.

I now evaluate the finite sample efficiency measured by RMSE of the MLE through the surface plots given in Appendix 8.5. Figure 3 shows the surface plots of RMSEs for β_{10} and β_{20}. It is clear from the surface plots that the MLE has higher RMSEs when strong spatial dependence exists in the model. The surface plots in Figure 4 are for λ_0 and ρ_0. These surface plots indicate that the MLE of these parameters has higher RMSEs when there exists strong negative spatial dependence in both the dependent variable and disturbance term.

7 Conclusion

In this study, I show that the MLE of the spatial autoregressive and moving average parameters for the SARMA(1,1) specification is generally inconsistent in the presence of heteroskedastic disturbances. The analytical results indicate that the concentrated log-likelihood function is not maximized at the true parameter values when heteroskedasticity is not considered in the estimation. The necessary condition for the consistency of the MLE depends on the specification of spatial weight matrices. We also show that the MLE of the parameters of the exogenous variables is inconsistent, and we state the expression for the corresponding asymptotic bias.

The Monte Carlo results show that the MLE imposes substantial amount of bias on the spatial autoregressive and moving average parameters in all cases for all sample sizes when the spatial weight matrix has non-identical blocks on the diagonals. The simulation results also show that the inconsistency of the spatial autoregressive and moving average parameters has almost no effect on the estimates of parameters of the exogenous variables for cases where the sample size is large.

8 Appendix

8.1 Some Useful Lemmas

Lemma 1. *Let A_n, B_n and C_n be $n \times n$ matrices with (i,j)th elements respectively denoted by $a_{n,ij}$, $b_{n,ij}$ and $c_{n,ij}$. Assume that A_n and B_n have zero diagonal elements, and C_n has uniformly bounded row and column sums in absolute value. Let q_n be $n \times 1$ vector with uniformly bounded elements in absolute value. Assume that ε_n satisfies Assumption 1 with covariance matrix denoted by Σ_n=*Diag$\{\sigma^2_{n1}, \ldots, \sigma^2_{nn}\}$. *Then,*

(1) $$\mathrm{E}\left(\varepsilon_n^{'} A_n \varepsilon_n \cdot \varepsilon_n^{'} B_n \varepsilon_n\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{n,ij}\left(b_{n,ij} + b_{n,ji}\right)\sigma^2_{ni}\sigma^2_{nj} = \mathrm{tr}\left(\Sigma_n A_n \left(B_n^{'}\Sigma_n + \Sigma_n B_n\right)\right)$$

(2) $$\begin{aligned}\mathrm{E}\left(\varepsilon_n C_n \varepsilon_n\right)^2 &= \sum_{i=1}^{n} c^2_{n,ii}\left[\mathrm{E}\left(\varepsilon^4_{ni}\right) - 3\sigma^4_{ni}\right] + \left(\sum_{i=1}^{n} c_{n,ii}\sigma^2_{ni}\right)^2 \\ &+ \sum_{i=1}^{n}\sum_{j=1}^{n} c_{n,ij}\left(c_{n,ij} + c_{n,ji}\right)\sigma^2_{ni}\sigma^2_{nj} \\ &= \sum_{i=1}^{n} c^2_{n,ii}\left[\mathrm{E}\left(\varepsilon^4_{ni}\right) - 3\sigma^4_{ni}\right] + \mathrm{tr}^2\left(\Sigma_n C_n\right) + \mathrm{tr}\left(\Sigma_n C_n C_n^{'}\Sigma_n + \Sigma_n C_n \Sigma_n C_n\right),\end{aligned}$$

(3) $$\begin{aligned}\mathrm{Var}\left(\varepsilon_n C_n \varepsilon_n\right) &= \sum_{i=1}^{n} c^2_{n,ii}\left[\mathrm{E}(\varepsilon^4_{ni}) - 3\sigma^4_{ni}\right] + \sum_{i=1}^{n}\sum_{j=1}^{n} c_{n,ij}(c_{n,ij} + c_{n,ji})\sigma^2_{ni}\sigma^2_{nj} \\ &= \sum_{i=1}^{n} c^2_{n,ii}\left[\mathrm{E}(\varepsilon^4_{ni}) - 3\sigma^4_{ni}\right] + \mathrm{tr}\left(\Sigma_n C_n C_n^{'}\Sigma_n + \Sigma_n C_n \Sigma_n C_n\right).\end{aligned}$$

(4) $$\mathrm{E}\left(\varepsilon_n^{'} C_n \varepsilon_n\right) = O(n),\ \mathrm{Var}\left(\varepsilon_n^{'} C_n \varepsilon_n\right) = O(n),\ \varepsilon_n^{'} C_n \varepsilon_n = O_p(n).$$

(5) $$\mathrm{E}\left(C_n \varepsilon_n\right) = 0,\ \mathrm{Var}\left(C_n \varepsilon_n\right) = O(n),\ C_n\varepsilon_n = O_p(n),\ \mathrm{Var}\left(q_n^{'} C_n \varepsilon_n\right) = O(n),\ q_n^{'} C_n \varepsilon_n = O_p(n).$$

Proof. For (1), (2), (3), (4) and (5) see *Lemmas A.1* through *A.4* in Lin and Lee (2010) and *Lemma 2* in Dogan and Suleyman (2013). □

Lemma 2. *Consider $\bar{\mathbb{M}}_n = (I_n - \mathbb{P}_n)$, where $\mathbb{P}_n = \bar{X}_n(\bar{X}_n^{'}\bar{X}_n)^{-1}\bar{X}_n^{'}$ under Assumption 3. Assume that ε_n satisfies Assumption 1 with covariance matrix denoted by Σ_n=*Diag$\{\sigma^2_{n1}, \ldots, \sigma^2_{nn}\}$. *Then,*

(1) *$\bar{\mathbb{M}}_n$ and $\mathbb{P}_n$ are uniformly bounded in absolute value in both row and column sums.*

(2) $$\mathrm{Var}\left(\mathbb{P}_n\varepsilon_n\right) = O\left(\frac{1}{n}\right),\ \mathbb{P}_n\varepsilon_n = o_p(1),\ \mathrm{Var}\left(\varepsilon_n\mathbb{P}_n\varepsilon_n\right) = O\left(\frac{1}{n}\right),\ \varepsilon_n\mathbb{P}_n\varepsilon_n = O_p(1).$$

(3) *Elements of $\mathbb{P}_n$ are $O\left(\frac{1}{n}\right)$.*

Proof. The proof is similar to the proof of *Lemma 3* in Dogan and Suleyman (2013). Hence, it is omitted. $\square$

8.2 Proof of Proposition 1

For the probability limit of terms in (4.14), the partial derivatives $\frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \rho}$, $\frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \lambda}$ and $\frac{\partial \bar{\mathbb{M}}_n(\rho)}{\partial \rho}$ are required, which are given by

$$
\begin{aligned}
(1)\quad \frac{\partial \bar{\mathbb{M}}_n(\rho)}{\partial \rho} &= -\left[R_n^{-1}(\rho) M_n \bar{X}_n(\rho)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)\right] - \left[\bar{X}_n(\rho)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\right. \\
&\quad \left.\times \bar{X}_n'(\rho) M_n' R_n'^{-1}(\rho)\right] + \left[\bar{X}_n(\rho)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho) H_n'(\rho)\bar{X}_n(\rho)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)\right] \\
&\quad + \left[\bar{X}_n(\rho)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho) H_n(\rho)\bar{X}_n(\rho)\left(\bar{X}_n'(\rho)\bar{X}_n(\rho)\right)^{-1}\bar{X}_n'(\rho)\right]
\end{aligned}
$$

$$
\begin{aligned}
(2)\quad \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \rho} &= \left[\frac{2}{n} Y_n' S_n'(\lambda) H_n'(\rho) R_n'^{-1}(\rho)\bar{\mathbb{M}}_n(\rho) R_n^{-1}(\rho) S_n(\lambda) Y_n\right] \\
&\quad - \left[\frac{2}{n} Y_n' S_n'(\lambda) R_n'^{-1}(\rho)\mathbb{P}_n(\rho) H_n' \bar{\mathbb{M}}_n(\rho) R_n^{-1}(\rho) S_n(\lambda) Y_n\right].
\end{aligned}
$$

$$
(3)\quad \frac{\partial \hat{\sigma}_n^2(\delta)}{\partial \lambda} = -\left[\frac{2}{n} Y_n' S_n'(\lambda) R_n'^{-1}(\rho)\bar{\mathbb{M}}_n(\rho) R_n^{-1}(\rho) W_n Y_n\right].
$$

First, the probability limit of the first row in (4.14) is investigated:

$$
\operatorname*{plim}_{n\to\infty} \frac{1}{n}\left(-\frac{n}{\frac{2}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}\frac{\partial \hat{\sigma}_n^2(\delta_0)}{\partial \lambda}\right) = \operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\bar{G}_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} + \operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\bar{G}_n\bar{X}_n\beta_0}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}, \tag{8.1}
$$

where we use $\bar{X}_n'\bar{\mathbb{M}}_n = 0_{k\times n}$. For the second term on the r.h.s. of (8.1), we have

$$
\operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n R_n^{-1} G_n X_n\beta_0}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} = 0, \tag{8.2}
$$

since the numerator converges in probability to zero by Lemma 1(5) and Lemma 2(1), and for the term in the denominator we have $\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n = \frac{1}{n}\sum_{i=1}^n \sigma_{ni}^2 + o_p(1)$ as shown in (4.11). The overall result is zero since $\frac{1}{n}\sum_{i=1}^n \sigma_{ni}^2$ is uniformly bounded for all n by Assumption 1. As for the first term on the r.h.s of (8.1), we have

$$
\operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\bar{G}_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} = \operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\varepsilon_n\bar{G}_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} - \operatorname*{plim}_{n\to\infty} \frac{\frac{1}{n}\varepsilon_n\bar{X}_n\left(\bar{X}_n'\bar{X}_n\right)^{-1}\bar{X}_n'\bar{G}_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}. \tag{8.3}
$$

We first evaluate the last term in (8.3). The numerator of this term tends to zero in probability as n goes to infinity by Lemma 1(4) and Assumption 3. Hence, the last term in (8.3) vanishes.

Now, we return to the first term in the r.h.s. of (8.3). By Lemma 1(4), $\mathrm{Var}\left(\frac{1}{n}\varepsilon_n'\bar{G}_n\varepsilon_n\right) = O\left(\frac{1}{n}\right) = o(1)$. Then, the Chebyshev inequality implies that $\mathrm{plim}_{n\to\infty}\left(\frac{1}{n}\varepsilon_n'\bar{G}_n\varepsilon_n - \mathrm{E}\left(\frac{1}{n}\varepsilon_n'\bar{G}_n\varepsilon_n\right)\right) = \mathrm{plim}_{n\to\infty}\left(\frac{1}{n}\varepsilon_n'\bar{G}_n\varepsilon_n - \frac{1}{n}\sum_{i=1}^n \bar{G}_{n.ii}\sigma_{ni}^2\right) = 0$. Hence,

$$\frac{\frac{1}{n}\varepsilon_n\bar{G}_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} = \frac{\frac{1}{n}\sum_{i=1}^n \bar{G}_{n,ii}\sigma_{ni}^2}{\frac{1}{n}\sum_{i=1}^n \sigma_{ni}^2} + o_p(1). \tag{8.4}$$

These results imply the following one:

$$\frac{1}{n}\left(-\frac{n}{\frac{2}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}\frac{\partial\hat{\sigma}_n^2(\delta_0)}{\partial\lambda}\right) = \frac{\frac{1}{n}\sum_{i=1}^n \bar{G}_{n.ii}\sigma_{ni}^2}{\frac{1}{n}\sum_{i=1}^n \sigma_{ni}^2} + o_p(1). \tag{8.5}$$

Now, we return to the first term in the second row of (4.14):

$$\begin{aligned}\underset{n\to\infty}{\mathrm{plim}}\frac{1}{n}\left(-\frac{n}{\frac{2}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}\frac{\partial\hat{\sigma}_n^2(\delta_0)}{\partial\rho}\right) &= -\underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}Y_n'S_n'H_n'R_n'^{-1}\bar{\mathbb{M}}_nR_n^{-1}S_nY_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}\\ &\quad+\underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}Y_n'S_n'R_n'^{-1}\mathbb{P}_nH_n'\bar{\mathbb{M}}_nR_n^{-1}S_nY_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}.\end{aligned} \tag{8.6}$$

Each term is handled separately below by using $R_n^{-1}S_nY_n = \bar{X}_n\beta_0 + \varepsilon_n$, $S_nY_n = X_n\beta_0 + R_n\varepsilon_n$, $\bar{X}_n'\bar{\mathbb{M}}_n = 0_{k\times n}$ and $\bar{\mathbb{M}}_n\bar{X}_n = 0_{n\times k}$. Note that $\frac{1}{n}Y_n'S_n'R_n'^{-1}\mathbb{P}_nH_n'\bar{\mathbb{M}}_nR_n^{-1}S_nY_n = \frac{1}{n}\beta_0'\bar{X}_n'\mathbb{P}_nH_n'\bar{M}_n\varepsilon_n + \frac{1}{n}\varepsilon_n\mathbb{P}_nH_n'\bar{M}_n\varepsilon_n$. By Lemma 1(5) and Lemma 2(1), $\frac{1}{n}\beta_0'\bar{X}_n'H_n'\bar{M}_n\varepsilon_n = o_p(1)$. By Lemma 1(4) and Assumption 3, we have $\frac{1}{n}\varepsilon_n'\mathbb{P}_nH_n\varepsilon_n = o_p(1)$. For the remaining term, by Lemma 2, we have $\frac{1}{n}\varepsilon_n\mathbb{P}_nH_n'\bar{M}_n\varepsilon_n = o_p(1)$. Hence, the second term on the r.h.s. of (8.6) vanishes.

The first term on r.h.s. of (8.6) can be written as

$$-\underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}Y_n'S_n'R_n'^{-1}\bar{\mathbb{M}}_nH_nR_n^{-1}S_nY_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} = -\underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_nH_n\bar{X}_n\beta_0}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} - \underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_nH_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}. \tag{8.7}$$

Substituting $\bar{\mathbb{M}}_n = \mathbb{I}_n - \bar{X}_n\left(\bar{X}_n'\bar{X}_n\right)^{-1}\bar{X}_n'$ into (8.7) yields

$$\begin{aligned}-\underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}Y_n'S_n'R_n'^{-1}\bar{\mathbb{M}}_nH_nR_n^{-1}S_nY_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} &= -\underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}\varepsilon_n'H_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n} - \underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_nH_n\bar{X}_n\beta_0}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}\\ &\quad+\underset{n\to\infty}{\mathrm{plim}}\frac{\frac{1}{n^2}\varepsilon_n'\bar{X}_n\left(\frac{1}{n}\bar{X}_n'\bar{X}_n\right)^{-1}\bar{X}_n'H_n\varepsilon_n}{\frac{1}{n}\varepsilon_n'\bar{\mathbb{M}}_n\varepsilon_n}.\end{aligned} \tag{8.8}$$

By Lemma 1(5) and (4.11), the second term on the r.h.s of (8.8) vanishes. The third term vanishes by Lemma 1(4) and (4.11). The probability limit of the remaining term can be found by the Chebyshev inequality. By Lemma 1(4), we have $\mathrm{Var}\left(\frac{1}{n}\varepsilon_n'H_n\varepsilon_n\right) = O(\frac{1}{n}) = o(1)$. Hence, $\mathrm{plim}_{n\to\infty}\left(\frac{1}{n}\varepsilon_n'H_n\varepsilon_n - E(\frac{1}{n}\varepsilon_n'H_n\varepsilon_n)\right) = \mathrm{plim}_{n\to\infty}\left(\frac{1}{n}\varepsilon_n'H_n\varepsilon_n - \frac{1}{n}\sum_{i=1}^n H_{n,ii}\sigma_{ni}^2\right) = 0$. Combining these results, we get the following result for the first term in the first row of (4.14):

$$\frac{1}{n}\left(-\frac{n}{\frac{2}{n}\varepsilon_n'\mathbb{M}_n\varepsilon_n}\frac{\partial\hat{\sigma}_n^2(\delta_0)}{\partial\rho}\right)=-\frac{\frac{1}{n}\sum_{i=1}^n H_{n,ii}\sigma_{ni}^2}{\frac{1}{n}\sum_{i=1}^n\sigma_{ni}^2}+o_p(1). \tag{8.9}$$

By combining the results in (8.5) and (8.9), we obtain:

$$\frac{1}{n}\frac{\partial\ln L_n(\delta_0)}{\partial\delta}=\begin{pmatrix}\frac{\frac{1}{n}\sum_{i=1}^n\overline{G}_{n.ii}\sigma_{ni}^2}{\frac{1}{n}\sum_{i=1}^n\sigma_{ni}^2}-\frac{1}{n}tr(G_n)+o_p(1)\\ -\left(\frac{\frac{1}{n}\sum_{i=1}^n H_{n,ii}\sigma_{ni}^2}{\frac{1}{n}\sum_{i=1}^n\sigma_{ni}^2}-\frac{1}{n}tr(H_n)\right)+o_p(1)\end{pmatrix}. \tag{8.10}$$

For the notational simplification, denote $H_n^*=\frac{1}{n}tr(H_n)=\frac{1}{n}\sum_{i=1}^n H_{n,ii}$, $\overline{G}_n^*=\frac{1}{n}tr(\overline{G}_n)=\frac{1}{n}\sum_{i=1}^n\overline{G}_{n,ii}$, and $\overline{\sigma}^2=\frac{1}{n}\sum_{i=1}^n\sigma_{ni}^2$. Then, (8.10) can be written in a more convenient form as[10]

$$\begin{aligned}\frac{1}{n}\frac{\partial\ln L_n(\delta_0)}{\partial\delta}&=\begin{pmatrix}\frac{\frac{1}{n}\sum_{i=1}^n\left(\overline{G}_{n.ii}-\overline{G}^*\right)\left(\sigma_{ni}^2-\overline{\sigma}^2\right)}{\overline{\sigma}^2}-\frac{1}{n}tr(\overline{G}_n-G_n)+o_p(1)\\ -\frac{\frac{1}{n}\sum_{i=1}^n\left(H_{n,ii}-H_n^*\right)\left(\sigma_{ni}^2-\overline{\sigma}^2\right)}{\overline{\sigma}^2}+o_p(1)\end{pmatrix}\\ &=\begin{pmatrix}\frac{\mathrm{cov}\left(\overline{G}_{n,ii},\,\sigma_{ni}^2\right)}{\overline{\sigma}^2}+o_p(1)\\ -\frac{\mathrm{cov}\left(H_{n,ii},\,\sigma_{ni}^2\right)}{\overline{\sigma}^2}+o_p(1)\end{pmatrix}.\end{aligned} \tag{8.11}$$

[10]Note that $\frac{1}{n}tr(\overline{G}_n-G_n)=0$.

8.3 Simulation Results for SARMA(0,1)

Table 1: Simulation Results for SARMA(0,1)

$n = 100$	β_1	β_2	ρ
ρ	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]
-0.6	(0.987)[-0.013](0.209)[0.209]	(1.009)[0.009](0.218)[0.218]	(-0.405)[0.195](0.618)[0.648]
-0.3	(1.000)[-0.000](0.211)[0.211]	(0.998)[-0.002](0.203)[0.203]	(-0.205)[0.095](0.630)[0.637]
0.0	(1.001)[0.001](0.214)[0.214]	(1.008)[0.008](0.229)[0.229]	(0.101)[0.101](0.565)[0.574]
0.3	(1.000)[0.000](0.217)[0.217]	(0.993)[-0.007](0.222)[0.222]	(0.434)[0.134](0.386)[0.409]
0.6	(0.996)[-0.004](0.212)[0.212]	(0.998)[-0.002](0.210)[0.210]	(0.710)[0.110](0.204)[0.232]
$n = 500$			
-0.6	(1.006)[0.006](0.083)[0.083]	(0.995)[-0.005](0.082)[0.082]	(-0.652)[-0.052](0.377)[0.380]
-0.3	(1.001)[0.001](0.083)[0.083]	(1.002)[0.002](0.084)[0.084]	(-0.354)[-0.054](0.388)[0.392]
0.0	(0.998)[-0.002](0.084)[0.084]	(1.002)[0.002](0.082)[0.082]	(0.007)[0.007](0.293)[0.293]
0.3	(1.005)[0.005](0.085)[0.085]	(0.998)[-0.002](0.080)[0.080]	(0.346)[0.046](0.189)[0.194]
0.6	(0.997)[-0.003](0.078)[0.078]	(1.002)[0.002](0.081)[0.081]	(0.652)[0.052](0.095)[0.108]
$n = 1000$			
-0.6	(1.000)[-0.000](0.058)[0.058]	(1.000)[0.000](0.059)[0.059]	(-0.682)[-0.082](0.284)[0.296]
-0.3	(0.999)[-0.001](0.059)[0.059]	(0.998)[-0.002](0.058)[0.058]	(-0.342)[-0.042](0.274)[0.277]
0.0	(1.000)[-0.000](0.057)[0.057]	(1.004)[0.004](0.059)[0.059]	(0.010)[0.010](0.191)[0.191]
0.3	(0.998)[-0.002](0.058)[0.058]	(1.000)[0.000](0.058)[0.058]	(0.330)[0.030](0.125)[0.128]
0.6	(1.002)[0.002](0.057)[0.057]	(0.999)[-0.001](0.057)[0.057]	(0.630)[0.030](0.072)[0.078]

8.4 Simulation Results for SARMA(1,1)

Table 2: Simulation Results for SARMA(1,1): $n = 100$

		λ	β_1	β_2	ρ
λ	ρ	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]
-0.6	-0.6	(-1.583)[-0.983](4.262)[4.374]	(0.874)[-0.126](0.342)[0.364]	(0.898)[-0.102](0.350)[0.365]	(-0.273)[0.327](0.981)[1.034]
-0.6	-0.3	(-1.790)[-1.190](4.346)[4.506]	(0.848)[-0.152](0.371)[0.401]	(0.847)[-0.153](0.361)[0.392]	(-0.178)[0.122](0.997)[1.004]
-0.6	0.0	(-1.794)[-1.194](4.355)[4.516]	(0.867)[-0.133](0.357)[0.381]	(0.865)[-0.135](0.353)[0.378]	(0.021)[0.021](0.934)[0.934]
-0.6	0.3	(-1.404)[-0.804](3.687)[3.773]	(0.839)[-0.161](0.379)[0.412]	(0.851)[-0.149](0.382)[0.410]	(0.264)[-0.036](0.709)[0.710]
-0.6	0.6	(-0.591)[0.009](1.108)[1.108]	(0.760)[-0.240](0.455)[0.515]	(0.760)[-0.240](0.455)[0.515]	(0.470)[-0.130](0.342)[0.366]
-0.3	-0.6	(-0.907)[-0.607](3.275)[3.331]	(0.912)[-0.088](0.325)[0.337]	(0.907)[-0.093](0.324)[0.337]	(-0.259)[0.341](0.822)[0.890]
-0.3	-0.3	(-1.132)[-0.832](3.497)[3.594]	(0.882)[-0.118](0.351)[0.370]	(0.881)[-0.119](0.362)[0.381]	(-0.136)[0.164](0.906)[0.920]
-0.3	0.0	(-1.335)[-1.035](3.840)[3.977]	(0.857)[-0.143](0.361)[0.388]	(0.861)[-0.139](0.367)[0.393]	(-0.005)[-0.005](0.861)[0.861]
-0.3	0.3	(-1.045)[-0.745](3.364)[3.445]	(0.840)[-0.160](0.399)[0.430]	(0.835)[-0.165](0.400)[0.433]	(0.220)[-0.080](0.709)[0.714]
-0.3	0.6	(-0.574)[-0.274](1.873)[1.893]	(0.768)[-0.232](0.466)[0.521]	(0.758)[-0.242](0.459)[0.519]	(0.436)[-0.164](0.390)[0.423]
0.0	-0.6	(-0.452)[-0.452](2.570)[2.609]	(0.904)[-0.096](0.354)[0.367]	(0.898)[-0.102](0.350)[0.365]	(-0.292)[0.308](0.721)[0.784]
0.0	-0.3	(-0.690)[-0.690](3.123)[3.199]	(0.903)[-0.097](0.337)[0.350]	(0.889)[-0.111](0.340)[0.358]	(-0.208)[0.092](0.772)[0.778]
0.0	0.0	(-0.834)[-0.834](3.174)[3.282]	(0.841)[-0.159](0.383)[0.415]	(0.857)[-0.143](0.391)[0.416]	(-0.079)[-0.079](0.804)[0.808]
0.0	0.3	(-0.450)[-0.450](2.131)[2.178]	(0.839)[-0.161](0.407)[0.438]	(0.838)[-0.162](0.412)[0.442]	(0.238)[-0.062](0.590)[0.593]
0.0	0.6	(-0.278)[-0.278](1.068)[1.104]	(0.768)[-0.232](0.469)[0.523]	(0.763)[-0.237](0.463)[0.521]	(0.411)[-0.189](0.349)[0.397]
0.3	-0.6	(0.068)[-0.232](1.429)[1.448]	(0.938)[-0.062](0.311)[0.317]	(0.951)[-0.049](0.307)[0.311]	(-0.384)[0.216](0.543)[0.585]
0.3	-0.3	(-0.157)[-0.457](2.174)[2.221]	(0.903)[-0.097](0.344)[0.358]	(0.902)[-0.098](0.345)[0.359]	(-0.279)[0.021](0.623)[0.623]
0.3	0.0	(-0.211)[-0.511](2.030)[2.094]	(0.867)[-0.133](0.376)[0.399]	(0.864)[-0.136](0.381)[0.404]	(-0.161)[-0.161](0.660)[0.679]
0.3	0.3	(-0.203)[-0.503](2.007)[2.069]	(0.819)[-0.181](0.437)[0.473]	(0.813)[-0.187](0.432)[0.471]	(0.095)[-0.205](0.621)[0.654]
0.3	0.6	(-0.022)[-0.322](0.735)[0.802]	(0.659)[-0.341](0.508)[0.612]	(0.657)[-0.343](0.503)[0.609]	(0.329)[-0.271](0.381)[0.468]
0.6	-0.6	(0.422)[-0.178](0.712)[0.733]	(0.981)[-0.019](0.231)[0.232]	(0.981)[-0.019](0.230)[0.231]	(-0.584)[0.016](0.346)[0.346]
0.6	-0.3	(0.376)[-0.224](0.580)[0.621]	(0.976)[-0.024](0.253)[0.254]	(0.965)[-0.035](0.255)[0.257]	(-0.511)[-0.211](0.329)[0.391]
0.6	0.0	(0.270)[-0.330](0.842)[0.905]	(0.961)[-0.039](0.294)[0.296]	(0.945)[-0.055](0.292)[0.297]	(-0.412)[-0.412](0.386)[0.564]
0.6	0.3	(0.152)[-0.448](1.345)[1.418]	(0.921)[-0.079](0.326)[0.335]	(0.920)[-0.080](0.335)[0.344]	(-0.286)[-0.586](0.415)[0.719]
0.6	0.6	(0.159)[-0.441](0.767)[0.884]	(0.802)[-0.198](0.436)[0.479]	(0.800)[-0.200](0.432)[0.476]	(-0.059)[-0.659](0.414)[0.779]

Table 3: Simulation Results for SARMA(1,1): $n = 500$

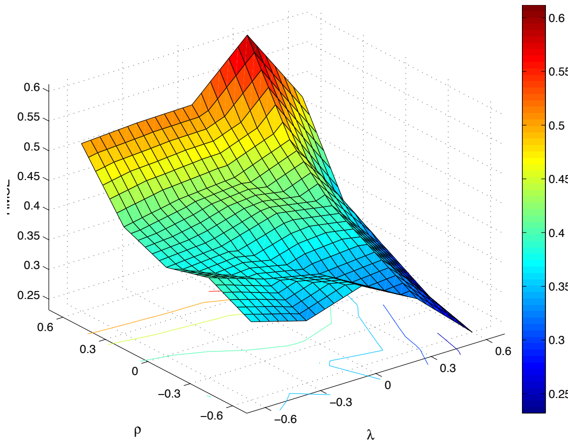
		λ	β_1	β_2	ρ
λ	ρ	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]
-0.6	-0.6	(-3.051)[-2.451](7.286)[7.687]	(0.914)[-0.086](0.257)[0.271]	(0.911)[-0.089](0.256)[0.271]	(-1.040)[-0.440](1.921)[1.970]
-0.6	-0.3	(-2.905)[-2.305](7.213)[7.572]	(0.916)[-0.084](0.253)[0.267]	(0.918)[-0.082](0.254)[0.267]	(-0.725)[-0.425](1.913)[1.960]
-0.6	0.0	(-1.771)[-1.171](5.677)[5.797]	(0.953)[-0.047](0.203)[0.208]	(0.949)[-0.051](0.204)[0.210]	(-0.123)[-0.123](1.427)[1.432]
-0.6	0.3	(-0.977)[-0.377](3.577)[3.597]	(0.985)[-0.015](0.142)[0.143]	(0.982)[-0.018](0.140)[0.141]	(0.303)[0.003](0.814)[0.814]
-0.6	0.6	(-0.667)[-0.067](0.139)[0.154]	(1.003)[0.003](0.088)[0.088]	(1.006)[0.006](0.085)[0.085]	(0.609)[0.009](0.087)[0.087]
-0.3	-0.6	(-0.985)[-0.685](4.189)[4.244]	(0.979)[-0.021](0.163)[0.165]	(0.975)[-0.025](0.162)[0.164]	(-0.608)[-0.008](1.201)[1.201]
-0.3	-0.3	(-1.513)[-1.213](5.164)[5.304]	(0.953)[-0.047](0.187)[0.193]	(0.960)[-0.040](0.189)[0.193]	(-0.577)[-0.277](1.472)[1.498]
-0.3	0.0	(-1.196)[-0.896](4.602)[4.689]	(0.972)[-0.028](0.174)[0.177]	(0.968)[-0.032](0.171)[0.174]	(-0.155)[-0.155](1.284)[1.293]
-0.3	0.3	(-0.457)[-0.157](1.797)[1.804]	(0.996)[-0.004](0.103)[0.103]	(0.994)[-0.006](0.102)[0.102]	(0.312)[0.012](0.449)[0.449]
-0.3	0.6	(-0.460)[-0.160](0.212)[0.266]	(1.000)[-0.000](0.082)[0.082]	(1.006)[0.006](0.082)[0.082]	(0.557)[-0.043](0.132)[0.139]
0.0	-0.6	(0.040)[0.040](1.086)[1.087]	(0.998)[-0.002](0.090)[0.090]	(0.998)[-0.002](0.086)[0.086]	(-0.371)[0.229](0.468)[0.521]
0.0	-0.3	(-0.220)[-0.220](2.089)[2.100]	(0.994)[-0.006](0.103)[0.103]	(0.994)[-0.006](0.109)[0.109]	(-0.333)[-0.033](0.703)[0.704]
0.0	0.0	(-0.205)[-0.205](1.807)[1.819]	(0.996)[-0.004](0.101)[0.101]	(0.995)[-0.005](0.101)[0.101]	(-0.075)[-0.075](0.681)[0.685]
0.0	0.3	(-0.077)[-0.077](0.731)[0.735]	(0.996)[-0.004](0.085)[0.085]	(0.998)[-0.002](0.087)[0.087]	(0.298)[-0.002](0.328)[0.328]
0.0	0.6	(-0.153)[-0.153](0.253)[0.296]	(0.987)[-0.013](0.136)[0.137]	(0.989)[-0.011](0.135)[0.136]	(0.521)[-0.079](0.197)[0.213]
0.3	-0.6	(0.317)[0.017](0.140)[0.141]	(1.003)[0.003](0.084)[0.084]	(1.000)[0.000](0.082)[0.082]	(-0.430)[0.170](0.201)[0.263]
0.3	-0.3	(0.228)[-0.072](0.173)[0.188]	(1.003)[0.003](0.086)[0.086]	(0.998)[-0.002](0.083)[0.083]	(-0.323)[-0.023](0.272)[0.273]
0.3	0.0	(0.137)[-0.163](0.715)[0.734]	(0.998)[-0.002](0.086)[0.087]	(0.997)[-0.003](0.086)[0.086]	(-0.174)[-0.174](0.408)[0.444]
0.3	0.3	(0.199)[-0.101](0.211)[0.234]	(0.996)[-0.004](0.100)[0.100]	(0.996)[-0.004](0.100)[0.100]	(0.216)[-0.084](0.362)[0.372]
0.3	0.6	(0.245)[-0.055](0.194)[0.202]	(0.961)[-0.039](0.211)[0.214]	(0.958)[-0.042](0.209)[0.213]	(0.587)[-0.013](0.205)[0.205]
0.6	-0.6	(0.545)[-0.055](0.086)[0.102]	(0.998)[-0.002](0.082)[0.082]	(1.000)[-0.000](0.084)[0.084]	(-0.652)[-0.052](0.102)[0.115]
0.6	-0.3	(0.486)[-0.114](0.082)[0.141]	(0.998)[-0.002](0.082)[0.082]	(1.001)[0.001](0.081)[0.081]	(-0.583)[-0.283](0.103)[0.301]
0.6	0.0	(0.411)[-0.189](0.091)[0.209]	(1.000)[0.000](0.083)[0.083]	(0.997)[-0.003](0.081)[0.081]	(-0.490)[-0.490](0.124)[0.505]
0.6	0.3	(0.324)[-0.276](0.088)[0.290]	(1.007)[0.007](0.089)[0.090]	(1.003)[0.003](0.092)[0.092]	(-0.344)[-0.644](0.200)[0.674]
0.6	0.6	(0.288)[-0.312](0.159)[0.350]	(0.943)[-0.057](0.253)[0.259]	(0.941)[-0.059](0.253)[0.260]	(-0.070)[-0.670](0.387)[0.774]

Table 4: Simulation Results for SARMA(1,1): $n = 1000$

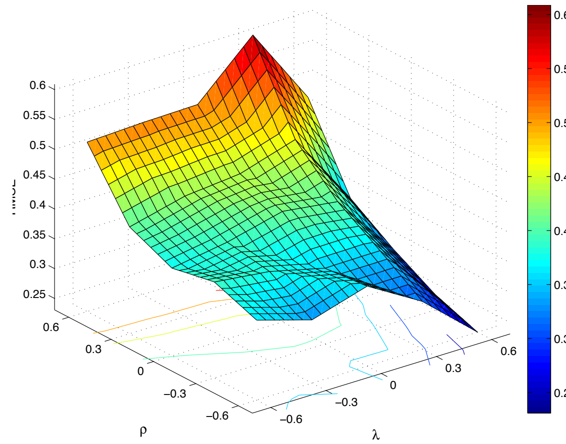
		λ	β_1	β_2	ρ
λ	ρ	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]	(Mean)[Bias](Std.)[RMSE]
-0.6	-0.6	(-3.449)[-2.849](8.618)[9.077]	(0.907)[-0.093](0.283)[0.298]	(0.906)[-0.094](0.282)[0.297]	(-1.323)[-0.723](2.487)[2.590]
-0.6	-0.3	(-4.151)[-3.551](9.648)[10.280]	(0.877)[-0.123](0.311)[0.334]	(0.880)[-0.120](0.312)[0.335]	(-1.135)[-0.835](2.798)[2.920]
-0.6	0.0	(-1.675)[-1.075](6.110)[6.204]	(0.957)[-0.043](0.201)[0.205]	(0.958)[-0.042](0.199)[0.204]	(-0.148)[-0.148](1.666)[1.672]
-0.6	0.3	(-0.650)[-0.050](2.400)[2.401]	(0.991)[-0.009](0.092)[0.093]	(0.991)[-0.009](0.093)[0.093]	(0.352)[0.052](0.568)[0.570]
-0.6	0.6	(-0.682)[-0.082](0.095)[0.126]	(1.007)[0.007](0.059)[0.060]	(1.007)[0.007](0.057)[0.058]	(0.595)[-0.005](0.054)[0.055]
-0.3	-0.6	(-0.698)[-0.398](3.631)[3.653]	(0.983)[-0.017](0.128)[0.129]	(0.985)[-0.015](0.129)[0.129]	(-0.624)[-0.024](1.152)[1.153]
-0.3	-0.3	(-1.691)[-1.391](6.083)[6.240]	(0.952)[-0.048](0.204)[0.210]	(0.954)[-0.046](0.204)[0.209]	(-0.704)[-0.404](1.839)[1.883]
-0.3	0.0	(-0.829)[-0.529](4.086)[4.120]	(0.981)[-0.019](0.141)[0.143]	(0.982)[-0.018](0.142)[0.143]	(-0.103)[-0.103](1.241)[1.245]
-0.3	0.3	(-0.385)[-0.085](1.415)[1.418]	(1.000)[-0.000](0.073)[0.073]	(0.999)[-0.001](0.074)[0.074]	(0.300)[-0.000](0.335)[0.335]
-0.3	0.6	(-0.476)[-0.176](0.169)[0.244]	(1.005)[0.005](0.058)[0.058]	(1.007)[0.007](0.058)[0.058]	(0.524)[-0.076](0.105)[0.130]
0.0	-0.6	(0.090)[0.090](0.866)[0.870]	(0.998)[-0.002](0.064)[0.064]	(1.000)[-0.000](0.067)[0.067]	(-0.361)[0.239](0.373)[0.443]
0.0	-0.3	(-0.096)[-0.096](1.508)[1.511]	(0.996)[-0.004](0.083)[0.083]	(0.995)[-0.005](0.078)[0.078]	(-0.323)[-0.023](0.575)[0.575]
0.0	0.0	(-0.111)[-0.111](1.287)[1.291]	(0.997)[-0.003](0.071)[0.071]	(0.994)[-0.006](0.070)[0.070]	(-0.063)[-0.063](0.525)[0.529]
0.0	0.3	(-0.074)[-0.074](0.181)[0.195]	(0.997)[-0.003](0.060)[0.060]	(0.999)[-0.001](0.060)[0.060]	(0.260)[-0.040](0.169)[0.174]
0.0	-0.6	(0.068)[-0.232](1.429)[1.448]	(0.938)[-0.062](0.311)[0.317]	(0.951)[-0.049](0.307)[0.311]	(-0.384)[0.216](0.543)[0.585]
0.3	-0.6	(0.342)[0.042](0.090)[0.099]	(1.000)[0.000](0.060)[0.060]	(1.001)[0.001](0.059)[0.059]	(-0.429)[0.171](0.118)[0.208]
0.3	-0.3	(0.251)[-0.049](0.111)[0.121]	(1.000)[-0.000](0.063)[0.063]	(1.004)[0.004](0.060)[0.061]	(-0.338)[-0.038](0.152)[0.157]
0.3	0.0	(0.174)[-0.126](0.125)[0.178]	(1.001)[0.001](0.058)[0.058]	(0.998)[-0.002](0.059)[0.059]	(-0.188)[-0.188](0.229)[0.296]
0.3	0.3	(0.225)[-0.075](0.145)[0.163]	(0.999)[-0.001](0.061)[0.061]	(0.999)[-0.001](0.058)[0.058]	(0.223)[-0.077](0.271)[0.281]
0.3	0.6	(0.274)[-0.026](0.147)[0.149]	(0.996)[-0.004](0.097)[0.097]	(0.992)[-0.008](0.097)[0.097]	(0.609)[0.009](0.148)[0.148]
0.6	-0.6	(0.562)[-0.038](0.055)[0.067]	(1.002)[0.002](0.059)[0.059]	(1.000)[0.000](0.059)[0.059]	(-0.668)[-0.068](0.066)[0.095]
0.6	-0.3	(0.496)[-0.104](0.058)[0.119]	(1.002)[0.002](0.059)[0.059]	(1.000)[0.000](0.058)[0.058]	(-0.590)[-0.290](0.071)[0.299]
0.6	0.0	(0.417)[-0.183](0.058)[0.192]	(0.998)[-0.002](0.062)[0.062]	(1.000)[0.000](0.061)[0.061]	(-0.495)[-0.495](0.072)[0.501]
0.6	0.3	(0.324)[-0.276](0.061)[0.282]	(1.004)[0.004](0.060)[0.060]	(1.002)[0.002](0.060)[0.060]	(-0.372)[-0.672](0.114)[0.682]
0.6	0.6	(0.320)[-0.280](0.176)[0.331]	(0.977)[-0.023](0.175)[0.177]	(0.975)[-0.025](0.176)[0.177]	(0.007)[-0.593](0.416)[0.725]

8.5 Surface Plots of RMSEs for SARMA(1,1)

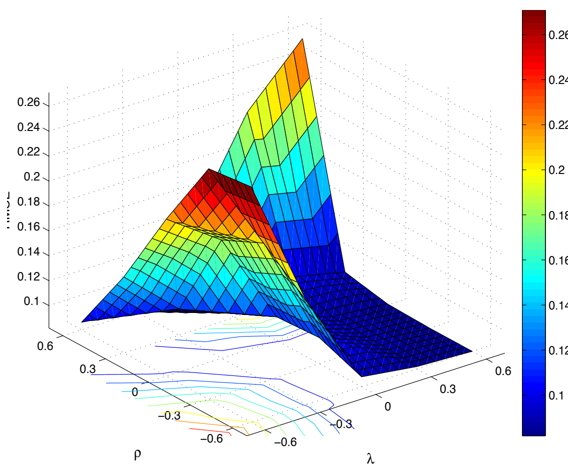
Figure 3: RMSE of β_1 and β_2



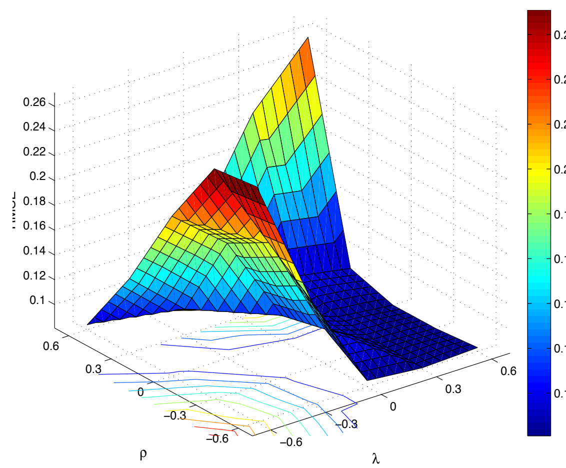
(a) $\beta_1, n = 100$



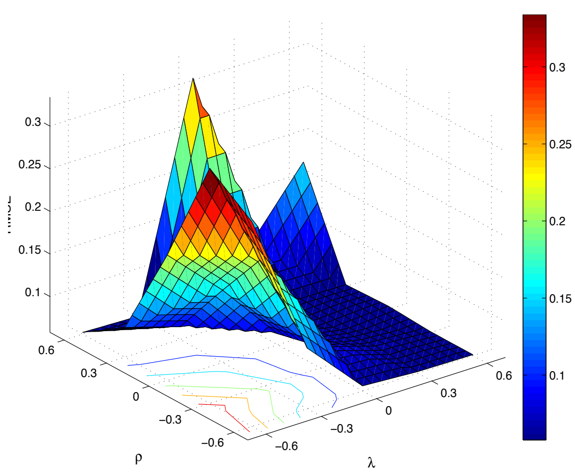
(b) $\beta_2, n = 100$



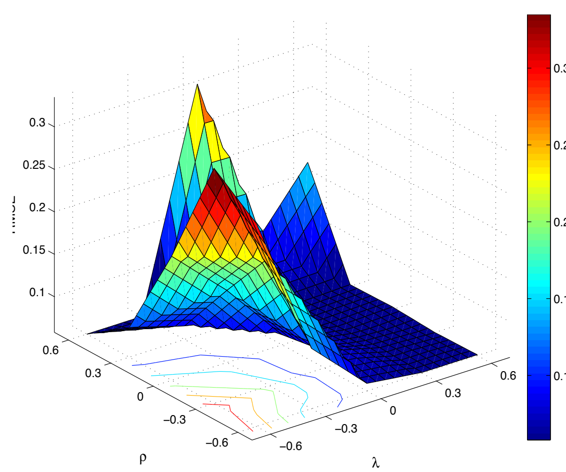
(c) $\beta_1, n = 500$



(d) $\beta_2, n = 500$

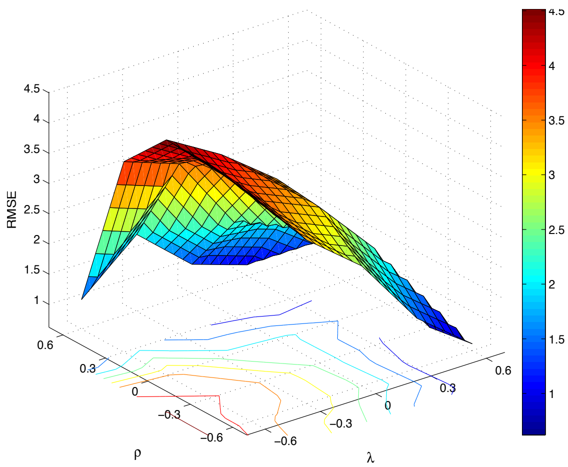


(e) $\beta_1, n = 1000$

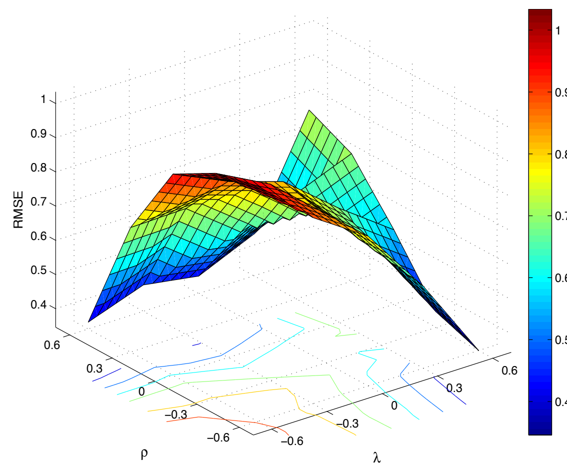


(f) $\beta_2, n = 1000$

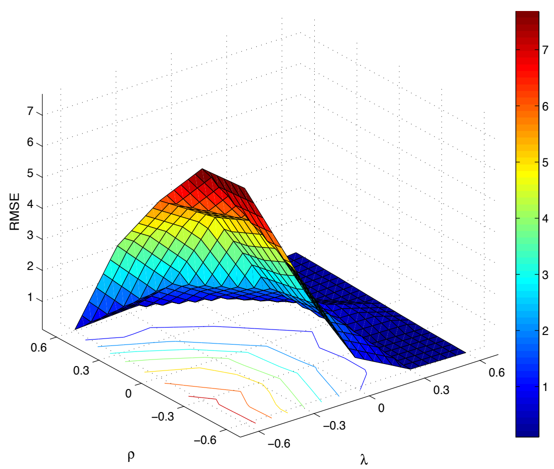
Figure 4: RMSE of λ and ρ



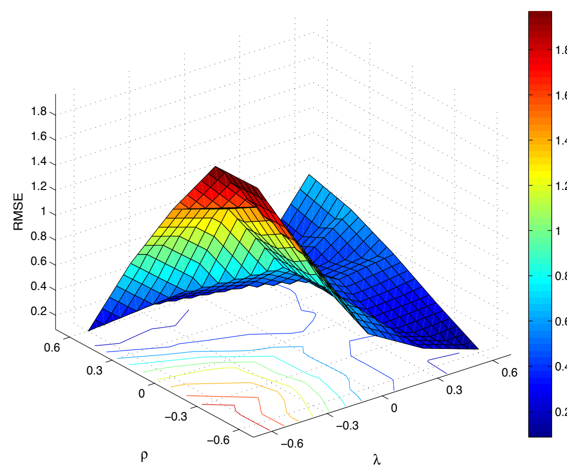
(a) $\lambda, n = 100$



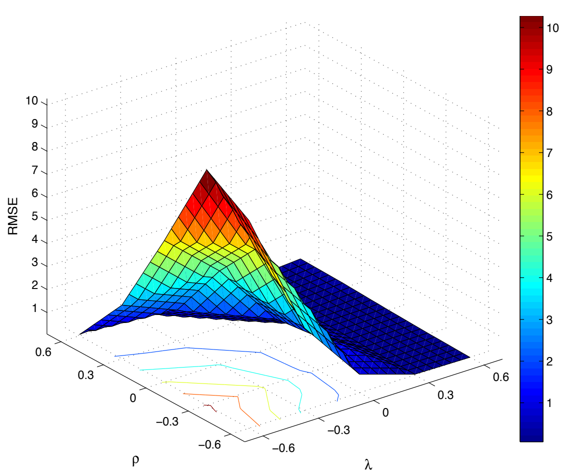
(b) $\rho, n = 100$



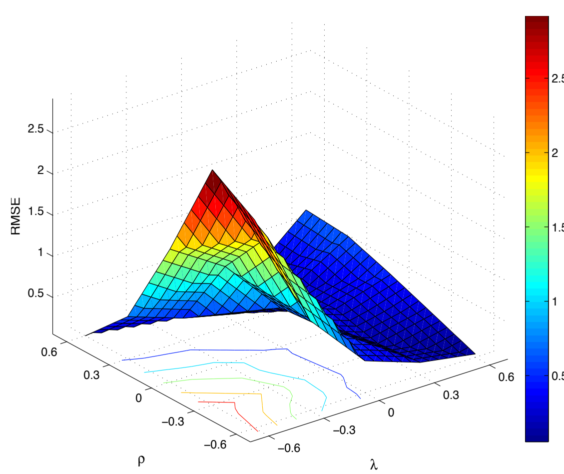
(c) $\lambda, n = 500$



(d) $\rho, n = 500$



(e) $\lambda, n = 1000$



(f) $\rho, n = 1000$

References

Abadir, Karim M. and Jan R. Magnus (2005). *Matrix Algebra*. Econometric Exercises. New York: Cambridge University Press.

Anselin, Luc (1988). *Spatial econometrics: Methods and Models*. New York: Springer.

Arnold, Matthias and Dominik Wied (2010). “Improved GMM estimation of the spatial autoregressive error model”. In: *Economics Letters* 108.1, pp. 65 –68.

Baltagi, Badi H. and Long Liu (2011). “An improved generalized moments estimator for a spatial moving average error model”. In: *Economics Letters* 113.3, pp. 282 –284.

Besag, Julian (1974). “Spatial Interaction and the Statistical Analysis of Lattice Systems”. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 36.2, pp. 192–236.

Das, Debabrata, Harry H. Kelejian, and Ingmar R. Prucha (2003). “Small Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances”. In: *Papers in Regional Science* 82, pp. 1–26.

Dogan, Osman and Taspinar Suleyman (2013). *GMM Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances*. Working Papers 001. City University of New York Graduate Center, Ph.D. Program in Economics.

Drukker, David M., Peter Egger, and Ingmar R. Prucha (2013). “On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors”. In: *Econometric Reviews* 32.5-6, pp. 686–733.

Fingleton, Bernard (2008a). “A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices”. In: *Empirical Economics* 34 (1), pp. 35–37.

— (2008b). “A Generalized Method of Moments Estimator for a Spatial Panel Model with an Endogenous Spatial Lag and Spatial Moving Average Errors”. In: *Spatial Economic Analysis* 3.1, pp. 27–44.

Haining, R. P. (1978). “The Moving Average Model for Spatial Interaction”. In: *Transactions of the Institute of British Geographers*. New Series 3.2, pp. 202–225.

Haining, Robert (1987). “Trend-Surface Models with Regional and Local Scales of Variation with an Application to Aerial Survey Data”. In: *Technometrics* 29.4, pp. 461–469.

Hepple, Leslie W. (1976). “A Maximum Likelihood Model for Econometric Estimation with Spatial Series”. In: *Theory and practice in regional science*. Ed. by Ian Masser. London papers in regional science 6. London: Pion Limited.

— (1995a). “Bayesian techniques in spatial and network econometrics: 1. Model comparison and posterior odds”. In: *Environment and Planning* 27.3, pp. 247–469.

— (1995b). “Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms”. In: *Environment and Planning* 27.4, pp. 615–644.

— (2003). *Bayesian and maximum likelihood estimation of the linear model with spatial moving average disturbances*. School of Geographical Sciences, University of Bristol, Working Papers Series.

Kelejian, Harry H. and Ingmar R. Prucha (1998). “A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances”. In: *Journal of Real Estate Finance and Economics* 17.1, pp. 1899–1926.

— (1999). “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model”. In: *International Economic Review* 40.2, pp. 509–533.

— (2007). “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, Department of Economics, University of Maryland”.

— (2010). “Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances”. In: *Journal of Econometrics* 157, pp. 53–67.

Kelejian, Harry H. and D Robinson (1993). “A Suggested Method of Estimation for Spatial Interdependent Models with Autocorrelated Errors, and an Application to a County Expenditure Model”. In: *Papers in Regional Science* 72, pp. 297–312.

Lee, Lung-fei (2004). “Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models”. In: *Econometrica* 72.6, pp. 1899–1925.

— (2007a). “GMM and 2SLS estimation of mixed regressive, spatial autoregressive models”. In: *Journal of Econometrics* 137.2, pp. 489–514.

— (2007b). “Identification and estimation of econometric models with group interactions, contextual factors and fixed effects”. In: *Journal of Econometrics* 140, pp. 333–374.

— (2007c). “The Method of Elimination and Substitution in the GMM Estimation of Mixed Regressive, Spatial Autoregressive Models”. In: *Journal of Econometrics* 140, pp. 155–189.

Lee, Lung-fei and Xiaodong Liu (2010). “Efficient GMM Estimation Of High Order Spatial Autoregressive Models With Autoregressive Disturbances”. In: *Econometric Theory* 26.01, pp. 187–230.

Lee, Lung-fei, Xiaodong Liu, and Xu Lin (2010). “Specification and estimation of social interaction models with network structures”. In: *The Econometrics Journal* 13, pp. 145–176.

LeSage, James and Robert K. Pace (2009). *Introduction to Spatial Econometrics (Statistics: A Series of Textbooks and Monographs*. London: Chapman and Hall/CRC.

Lesage, James P. (1997). “Bayesian Estimation of Spatial Autoregressive Models”. In: *International Regional Science Review* 20.1-2, pp. 113–129.

Lin, Xu and Lung-fei Lee (2010). “GMM estimation of spatial autoregressive models with unknown heteroskedasticity”. In: *Journal of Econometrics* 157.1, pp. 34–52.

Liu, Xiaodong, Lung-fei Lee, and Christopher R. Bollinger (2010). “An efficient GMM estimator of spatial autoregressive models”. In: *Journal of Econometrics* 159.2, pp. 303–319.

Mardia, K. V. and R. J. Marshall (1984). “Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression”. In: *Biometrika* 71.1, pp. 135–146.

Pace, Robert K., James P. LeSage, and Shuang Zhu (2012). “Spatial Dependence in Regressors and its Effect on Performance of Likelihood-Based and Instrumental Variable Estimators”. In: ed. by Daniel Millimet Dek Terrell. 30th Anniversary Edition (Advances in Econometrics, Volume 30). Emerald Group Publishing Limited, pp. 257–295.

Prucha, Ingmar R. (2014). “Instrumental Variables/Method of Moments Estimation”. In: *Handbook of Regional Science*. Ed. by Manfred M. Fischer and Peter Nijkamp. Springer Berlin Heidelberg, pp. 1597–1617.

Richardson, S., C. Guihenneuc, and V. Lasserre (1992). “Spatial Linear Models with Autocorrelated Error Structure”. In: *Journal of the Royal Statistical Society. Series D (The Statistician)* 41.5, pp. 539–557.

Ripley, Brian D. (2005). *Spatial Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons.