

# Quasi Maximum Likelihood Estimation of Spatial Models with Heterogeneous Coefficients\*

Michele Aquaro  
University of Warwick

Natalia Bailey  
Queen Mary, University of London

M. Hashem Pesaran  
Department of Economics & USC Dornsife INET, University of Southern California,  
and Trinity College, Cambridge

June 19, 2015

## Abstract

This paper considers spatial autoregressive panel data models and extends their analysis to the case where the spatial coefficients differ across the spatial units. It derives conditions under which the spatial coefficients are identified and develops a quasi maximum likelihood (QML) estimation procedure. Under certain regularity conditions, it is shown that the QML estimators of individual spatial coefficients are consistent and asymptotically normally distributed when both the time and cross section dimensions of the panel are large. It derives the asymptotic covariance matrix of the QML estimators allowing for the possibility of non-Gaussian error processes. Small sample properties of the proposed estimators are investigated by Monte Carlo simulations for Gaussian and non-Gaussian errors, and with spatial weight matrices of differing degree of sparseness. The simulation results are in line with the paper's key theoretical findings and show that the QML estimators have satisfactory small sample properties for panels with moderate time dimensions and irrespective of the number of cross section units in the panel, under certain sparsity conditions on the spatial weight matrix.

**Keywords:** Spatial panel data models, heterogeneous spatial lag coefficients, identification, quasi maximum likelihood (QML) estimators, non-Gaussian errors.

**JEL Codes:** C21, C23

---

\*The authors would like to acknowledge helpful comments from Bernard Fingleton, Harry Kelejian, Ron Smith and Cynthia Fan Yang. Financial support under ESRC Grant ES/I031626/1 is also gratefully acknowledged.

# 1 Introduction

Following the pioneering contributions of Whittle (1954) and Cliff and Ord (1973), important advances have been made in the analysis of spatial models. The original maximum likelihood approach of Cliff and Ord developed for a large number of spatial units ( $N$ ) observed at a point in time (or over a given time interval) has been extended to cover panel data models with fixed effects and dynamics. Other estimation and testing techniques, such as the generalised method of moments (GMM), have also been proposed. Some of the key references to the literature include Upton and Fingleton (1985), Anselin (1988), Cressie (1993), Kelejian and Robinson (1993), Ord and Getis (1995), Anselin and Bera (1998), and more recently, Haining (2003), Lee (2004), Kelejian and Prucha (1999), Kelejian and Prucha (2010), Lin and Lee (2010), Lee and Yu (2010), LeSage and Pace (2010), Arbia (2010), Cressie and Wikle (2011), and Elhorst (2014). Extensions to dynamic panels are provided by Anselin (2001), Baltagi et al. (2003), Kapoor et al. (2007), Baltagi et al. (2007), and Yu et al. (2008).

One important feature of the above contributions is the fact that they all assume that except for unit-specific effects, all other parameters, including the spatial lag coefficients, are homogeneous. This assumption might be needed in the case of pure spatial models or spatial panel data models with a short time dimension ( $T$ ), but with increasing availability of large panel data sets where  $N$  and  $T$  are both reasonably large it seems desirable to allow the spatial lag coefficients to vary across the spatial units. Examples of such data sets include large panels that cover regions, counties, states, or countries in the analysis of economic variables such as house prices, real wages, employment and income. For instance, in the empirical applications by Baltagi and Levin (1986) on demand for tobacco consumption, and by Holly et al. (2010) on house price diffusion across States in the US, the maintained assumption that spillover effects from neighbouring States are the same across all the 48 mainland States seems unduly restricting, particularly considering the large size of the US and the uneven distribution of economic activity across it.

This paper considers spatial autoregressive panel data models and extends their analysis to the case where the spatial coefficients differ across the spatial units. It derives conditions under which the spatial coefficients are globally and locally identified and proposes a quasi maximum likelihood (QML) estimation procedure for estimation and inference. It shows that the QML estimators are consistent and asymptotically normal under general regularity conditions when both the time and cross section dimensions of the panel are large. Asymptotic covariance matrices of the QML estimators are derived under Gaussian errors as well as when the errors are non-Gaussian, and consistent estimators of these covariance matrices are proposed for inference. The pure spatial model is further extended to include exogenous regressors, allowing the slope coefficients of the regressors to vary over the cross section units. The model allows for spatial dependence directly through contemporaneous dependence of individual units on their neighbours, and indirectly through possible cross-sectional dependence in the regressors.

The small sample performance of QML estimators are investigated by Monte Carlo simulations for different choices of the spatial weight matrices. The simulation results are in line with the paper's key theoretical findings, and show that the proposed estimators have good small sample properties for panels with moderate time dimensions and irrespective of the number of cross section units in the panel, although under non-Gaussian errors, tests based on QML estimators of the spatial parameters can be slightly distorted when the time dimension is relatively

small.

The rest of the paper is organised as follows. Section 2 sets up the first order spatial autoregressive model with heterogeneous coefficients (HSAR), formulates the assumptions, and derives the log-likelihood function. The identification problem is discussed in Section 3, followed by an account of consistency and asymptotic normality of the QML estimator in Sections 4 and 5, respectively. Section 6 considers the inclusion of heteroskedastic error variances and exogenous regressors in the HSAR model. Section 7 outlines the Monte Carlo design and reports small sample results (bias, root mean square errors, size and power) for different parameter values and sample size combinations. Some concluding remarks are provided in Section 8.

Notation: We denote the largest and the smallest eigenvalues of the  $N \times N$  matrix  $\mathbf{A} = (a_{ij})$  by  $\lambda_{\max}(\mathbf{A})$  and  $\lambda_{\min}(\mathbf{A})$ , respectively, its trace by  $\text{tr}(\mathbf{A}) = \sum_{i=1}^N a_{ii}$ , its maximum absolute column sum norm by  $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq N} \left( \sum_{i=1}^N |a_{ij}| \right)$ , and its maximum absolute row sum norm by  $\|\mathbf{A}\|_\infty = \max_{1 \leq i \leq N} \left( \sum_{j=1}^N |a_{ij}| \right)$ .  $\odot$  stands for Hadamard product or element-wise matrix product operator,  $\rightarrow_p$  denotes convergence in probability, and  $\rightarrow_d$  convergence in distribution. All asymptotics are carried out for a given  $N$  and as  $T \rightarrow \infty$ .  $K$  and  $\epsilon$  will be used to denote finite large and non-zero small positive numbers, respectively.

## 2 A heterogeneous spatial autoregressive model (HSAR)

The standard first-order spatial *autoregressive* panel data model is given by (see, for example, Anselin (1988))

$$y_{it} = \psi \sum_{j=1}^N w_{ij} y_{jt} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (1)$$

where scalar  $\psi$  is the spatial autoregressive parameter, assumed to be the same over all cross section units. Further,  $\mathbf{w}'_i \mathbf{y}_t = \sum_{j=1}^N w_{ij} y_{jt}$ , where  $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{iN})'$  with  $w_{ii} = 0$ , and  $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$ . Here  $\mathbf{w}_i$  denotes an  $N \times 1$  non-stochastic vector which is determined *a priori*. In its non-normalised form, it comprises of binary elements taking values of one or zero depending on whether unit  $i$  is ‘connected’ with unit  $j$  under a suitably defined distance metric, for  $i \neq j$ ,  $i, j = 1, 2, \dots, N$ . Stacking the observations by  $N$  individual units, (1) becomes:

$$\mathbf{y}_t = \psi \mathbf{W} \mathbf{y}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (2)$$

where  $\mathbf{W} = (w_{ij})$ ,  $i, j = 1, 2, \dots, N$ , is an  $N \times N$  ‘network’ or ‘spatial’ weight matrix that characterises all the connections, and  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ .

A heterogeneous version of (1) can be written as

$$y_{it} = \psi_i \sum_{j=1}^N w_{ij} y_{jt} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (3)$$

Initially, we assume that error variances are homoskedastic and set  $\sigma_i^2 = \sigma^2$  for all  $i$ , so that we focus on the heterogeneity of  $\psi_i$ . Extensions of the model to the case where the errors are heteroscedastic and (3) includes exogenous regressors are considered in Section 6.

Again, stacking the observations on individual units for each time period,  $t$ , we have

$$(\mathbf{I}_N - \mathbf{\Psi}\mathbf{W})\mathbf{y}_t = \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (4)$$

where  $\mathbf{\Psi} = \text{diag}(\boldsymbol{\psi})$ ,  $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_N)'$ , and  $\mathbf{I}_N$  is an  $N \times N$  identity matrix. For each  $i = 1, 2, \dots, N$ , the true value of  $\psi_i$  will be denoted by  $\psi_{i0}$  and accordingly  $\mathbf{\Psi}_0$  denotes the value of  $\mathbf{\Psi}$  evaluated at the true values,  $\boldsymbol{\psi}_0 = (\psi_{10}, \psi_{20}, \dots, \psi_{N0})'$ . In what follows, we also make use of the notations,  $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \mathbf{\Psi}\mathbf{W}$ , and  $\mathbf{S}_0 = \mathbf{I}_N - \mathbf{\Psi}_0\mathbf{W}$ .

For the analysis of identification and estimation of the heterogeneous spatial autoregressive (HSAR) model, we adopt the following assumptions:

**Assumption 1** *The  $N \times N$  spatial weight matrix,  $\mathbf{W} = (w_{ij})$ , is exactly sparse such that*

$$h_N = \max_{i \leq N} \sum_{j \leq N} I(w_{ij} \neq 0),$$

*is bounded in  $N$ , where  $I(A)$  denotes the indicator function that takes the value of 1 if  $A$  holds and 0 otherwise, and all diagonal elements of matrix  $\mathbf{W}$  are zero, i.e.  $w_{ii} = 0$ , for all  $i = 1, 2, \dots, N$ .*

**Remark 1** *This assumption ensures that the maximum number of non-zero elements in each row is bounded in  $N$ , and the matrix norms  $\|\mathbf{W}\|_1$  and  $\|\mathbf{W}\|_\infty$  are bounded in  $N$ . Note that we do not need the weight matrix to be exactly sparse, and Assumption 1 can be replaced by an approximate sparsity condition.*

**Assumption 2** *The error terms,  $\varepsilon_{it}$ ,  $i = 1, 2, \dots, N$ ,  $N \geq 1$ ,  $t = 1, 2, \dots, T$  are independently distributed over  $i$  and  $t$ , have zero means, and constant variances,  $E(\varepsilon_{it}^2) = \sigma^2$ , where  $0 < \kappa_1 \leq \sigma^2 \leq \kappa_2 < \infty$ , and  $\kappa_1, \kappa_2$  are finite generic constants independent of  $N$ .*

**Assumption 3** *The  $(N + 1) \times 1$  parameter vector,  $\boldsymbol{\theta} = (\boldsymbol{\psi}', \sigma^2)' \in \Theta$ , is a sub-set of the  $N + 1$  dimensional Euclidean space,  $\mathbb{R}^{N+1}$ .  $\Theta$  is a closed and bounded (compact) set and includes the true value of  $\boldsymbol{\theta}$ , denoted by  $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}'_0, \sigma_0^2)'$ , as an interior point.*

**Assumption 4**  $\lambda_{\min}[\mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})] > 0$ , for all values of  $\boldsymbol{\psi} \in \Theta$  and  $N$ .

**Assumption 5** Let  $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$ , and consider the sample covariance matrix of  $\mathbf{y}_t$ ,

$$\hat{\boldsymbol{\Sigma}}_T = \sum_{t=1}^T \mathbf{y}_t \mathbf{y}_t' / T. \quad (5)$$

For a given  $N$  we have (as  $T \rightarrow \infty$ )

$$\hat{\boldsymbol{\Sigma}}_T \rightarrow_p \boldsymbol{\Sigma}_0, \quad (6)$$

uniformly in  $\boldsymbol{\theta} = (\boldsymbol{\psi}', \sigma^2)'$ , where  $\boldsymbol{\Sigma}_0 = \sigma_0^2 (\mathbf{I}_N - \mathbf{\Psi}_0\mathbf{W})^{-1} (\mathbf{I}_N - \mathbf{\Psi}_0\mathbf{W})'^{-1} = \sigma_0^2 (\mathbf{S}'_0 \mathbf{S}_0)^{-1}$ .

**Remark 2** *Assumptions 1 and 2 are standard in the spatial econometrics literature. Also, it is easily seen that under Assumption 1,  $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \mathbf{\Psi}\mathbf{W}$  will be globally invertible if  $\sup_i |\psi_i| \|\mathbf{W}\|_\infty < 1$ . This follows since condition  $\sup_i |\psi_i| \|\mathbf{W}\|_\infty < 1$  ensures that matrix  $\mathbf{I}_N -$*

$\Psi\mathbf{W}$  is strictly diagonally dominant, and hence invertible.<sup>1</sup> But for identification, estimation, and inference when  $N$  is large, we also need a similar condition on the columns of  $\mathbf{W}$ , namely  $\sup_i |\psi_i| \|\mathbf{W}\|_1 < 1$ . Combining the two conditions we have

$$\sup_i |\psi_i| < \max \{1/\|\mathbf{W}\|_1, 1/\|\mathbf{W}\|_\infty\}. \quad (7)$$

See Lemma 1 in Appendix A. This result reduces to the condition obtained in Lemma 2 of Kelejian and Prucha (2010) for the homogeneous case where  $\psi_i = \psi$  for all  $i$ .

**Remark 3** Note also that under Assumption 1 and condition (7), for all values of  $\boldsymbol{\psi}$  and  $N$  we have

$$\begin{aligned} \lambda_{\max} [\mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})] &\leq \|\mathbf{S}(\boldsymbol{\psi})\|_1 \|\mathbf{S}(\boldsymbol{\psi})\|_\infty \\ &\leq \left(1 + \sup_i |\psi_i| \|\mathbf{W}\|_1\right) \left(1 + \sup_i |\psi_i| \|\mathbf{W}\|_\infty\right) < K. \end{aligned}$$

This result, together with Assumptions 2 and 4, ensures that for all  $\boldsymbol{\theta} = (\boldsymbol{\psi}', \sigma^2)' \in \Theta$ , and  $N$

$$\lambda_{\min} [\boldsymbol{\Sigma}(\boldsymbol{\theta})] > 0 \text{ and } \lambda_{\max} [\boldsymbol{\Sigma}(\boldsymbol{\theta})] < K < \infty, \quad (8)$$

and

$$\lambda_{\min} [\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] > 0 \text{ and } \lambda_{\max} [\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta})] < K < \infty, \quad (9)$$

where

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \sigma^2 (\mathbf{I}_N - \Psi\mathbf{W})^{-1} (\mathbf{I}_N - \Psi\mathbf{W})'^{-1} = \sigma^2 [\mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})]^{-1}.$$

## 2.1 The log-likelihood function

Under Assumption 1 and assuming condition (7) is met, then  $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \Psi\mathbf{W}$  is non-singular and (4) can be expressed as

$$\mathbf{y}_t = \mathbf{S}^{-1}(\boldsymbol{\psi})\boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T. \quad (10)$$

Under Assumption 2 and assuming that the errors,  $\boldsymbol{\varepsilon}_{it}$ , are normally distributed, then the joint density function of  $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T$  is given by

$$\prod_{t=1}^T \frac{|\mathbf{S}(\boldsymbol{\psi})|}{|\sigma^2 \mathbf{I}_N|^{1/2} (2\pi)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \mathbf{y}_t' \mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})\mathbf{y}_t\right),$$

and the (quasi) log-likelihood function can be written as

$$\ell(\boldsymbol{\theta}) = -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln \sigma^2 + T \ln |\mathbf{S}(\boldsymbol{\psi})| - \frac{1}{2\sigma^2} \sum_{t=1}^T \mathbf{y}_t' \mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})\mathbf{y}_t, \quad (11)$$

---

<sup>1</sup>The  $N \times N$  matrix  $\mathbf{A} = (a_{ij})$  is said to be strictly diagonally dominant if  $|a_{ii}| > \sum_{j \neq i}^N |a_{ij}|$ , for all  $i = 1, 2, \dots, N$ . Then by the Levy-Desplanques theorem it follows that  $\mathbf{A}$  is non-singular. See <http://planetmath.org/levydesplanquestheorem>.

where as before  $\boldsymbol{\theta} = (\boldsymbol{\psi}', \sigma^2)'$ . The last term of the log-likelihood function can be written more conveniently as

$$\begin{aligned} \sum_{t=1}^T \mathbf{y}_t' \mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi}) \mathbf{y}_t &= \sum_{t=1}^T \text{tr} [\mathbf{y}_t' \mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi}) \mathbf{y}_t] \\ &= T \text{tr} \left[ \mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi}) \frac{\sum_{t=1}^T \mathbf{y}_t \mathbf{y}_t'}{T} \right] \\ &= T \text{tr} \left[ \mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi}) \hat{\boldsymbol{\Sigma}}_T \right] \end{aligned}$$

where  $\hat{\boldsymbol{\Sigma}}_T$  is defined by (5). Hence,

$$\ell(\boldsymbol{\theta}) = -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln \sigma^2 + T \ln |\mathbf{S}(\boldsymbol{\psi})| - \frac{T}{2\sigma^2} \text{tr} \left[ \mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi}) \hat{\boldsymbol{\Sigma}}_T \right]. \quad (12)$$

### 3 Identification

We now investigate the conditions under which  $\boldsymbol{\theta}$  is identified. We first note that

$$E_0 \left( \hat{\boldsymbol{\Sigma}}_T \right) = \sum_{t=1}^T E_0 (\mathbf{y}_t \mathbf{y}_t') / T,$$

and using (10),

$$E_0 \left( \hat{\boldsymbol{\Sigma}}_T \right) = \sigma_0^2 (\mathbf{S}'_0 \mathbf{S}_0)^{-1},$$

where  $E_0$  denotes expectation taken under the DGP characterised by the true parameter vector  $\boldsymbol{\theta}_0$  and, as before,  $\mathbf{S}_0 = \mathbf{S}(\boldsymbol{\psi}_0) = \mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W}$ . Now taking expectations of the log likelihood function, (12), at the true value of  $\boldsymbol{\theta}$ ,  $\boldsymbol{\theta}_0$ , and using the above result we have

$$E_0 [\ell(\boldsymbol{\theta}_0)] = -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln \sigma_0^2 + T \ln |\mathbf{S}_0| - \frac{NT}{2}. \quad (13)$$

Next, we take expectations of the log-likelihood at some other parameter value,  $\boldsymbol{\theta}$ :

$$\begin{aligned} E_0 [\ell(\boldsymbol{\theta})] &= -\frac{NT}{2} \ln(2\pi) - \frac{NT}{2} \ln \sigma^2 + T \ln |\mathbf{S}(\boldsymbol{\psi})| \\ &\quad - \frac{T\sigma_0^2}{2\sigma^2} \text{tr} \left[ \mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi}) (\mathbf{S}'_0 \mathbf{S}_0)^{-1} \right]. \end{aligned} \quad (14)$$

Let  $\mathbf{P}_0 = \mathbf{I}_N - \mathbf{D}\mathbf{G}_0$ ,

$$\mathbf{G}_0 \equiv \mathbf{G}(\boldsymbol{\psi}_0) = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1}, \quad (15)$$

and

$$\mathbf{D} = \boldsymbol{\Psi} - \boldsymbol{\Psi}_0. \quad (16)$$

Then,

$$\begin{aligned}
\mathbf{S}(\boldsymbol{\psi}) &= \mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W} - (\boldsymbol{\Psi} - \boldsymbol{\Psi}_0) \mathbf{W} \\
&= \left[ \mathbf{I}_N - (\boldsymbol{\Psi} - \boldsymbol{\Psi}_0) \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \right] (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W}) \\
&= (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0) \mathbf{S}_0.
\end{aligned} \tag{17}$$

Note that  $(\mathbf{I}_N - \mathbf{D}\mathbf{G}_0) = \mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}$  is invertible so long as condition (7) is met for all values of  $\psi_i$ , including the true values,  $\psi_{i0}$ .

Consider now the expectations of the difference of the log-likelihoods evaluated at  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta}$  given in (13) and (14). We have,

$$\begin{aligned}
\frac{E_0 [\ell(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\theta})]}{NT} &= -\frac{1}{2} \ln \left( \frac{\sigma_0^2}{\sigma^2} \right) - \frac{1}{N} \ln |\mathbf{I}_N - \mathbf{D}\mathbf{G}_0| - \frac{1}{2} \\
&\quad + \frac{1}{2N} \frac{\sigma_0^2}{\sigma^2} \text{tr} \left\{ (\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W})' (\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}) (\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})^{-1} [(\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})']^{-1} \right\} \geq 0.
\end{aligned}$$

But,

$$\begin{aligned}
(\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}) (\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})^{-1} &= (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0) (\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W}) (\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})^{-1} \\
&= \mathbf{I}_N - \mathbf{D}\mathbf{G}_0.
\end{aligned} \tag{18}$$

Therefore, after some algebra, and denoting  $(NT)^{-1}E_0 [\ell(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\theta})]$  by  $Q_N(\boldsymbol{\varphi})$ , we have

$$\begin{aligned}
Q_N(\boldsymbol{\varphi}) &= -\frac{1}{2} [\ln(1 - \delta) + \delta] - \frac{1}{N} \ln |\mathbf{I}_N - \mathbf{D}\mathbf{G}_0| - \frac{1}{N} (1 - \delta) \text{tr}(\mathbf{D}\mathbf{G}_0) \\
&\quad + \frac{1}{2N} (1 - \delta) \text{tr}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}^2) \geq 0,
\end{aligned} \tag{19}$$

where  $\boldsymbol{\varphi} = (\mathbf{d}', \delta)'$ ,  $\delta = (\sigma^2 - \sigma_0^2) / \sigma^2 < 1$ , and  $\mathbf{d} = (d_1, d_2, \dots, d_N)'$ ,  $d_i = \psi_i - \psi_{i0}$ , for  $i = 1, 2, \dots, N$ . Using the derivative results in Appendix B, we have

$$\frac{\partial Q_N(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} = \left\{ \begin{array}{l} \frac{1}{N} \text{diag} \left[ \mathbf{G}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} \right] \boldsymbol{\tau}_N - \frac{1}{N} (1 - \delta) \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N + \frac{1}{N} (1 - \delta) \text{diag}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}) \boldsymbol{\tau}_N \\ \frac{1}{2} \frac{\delta}{1 - \delta} + \frac{1}{N} [\text{tr}(\mathbf{D}\mathbf{G}_0) - \frac{1}{2} \text{tr}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}^2)] \end{array} \right\},$$

and  $\partial^2 Q_N(\boldsymbol{\varphi}) / \partial \boldsymbol{\varphi}' \partial \boldsymbol{\varphi} = N^{-1} \boldsymbol{\Lambda}_N(\boldsymbol{\varphi})$ , where

$$\boldsymbol{\Lambda}_N(\boldsymbol{\varphi}) = \left[ \begin{array}{cc} (\mathbf{A}_0 \odot \mathbf{A}'_0) + (1 - \delta) \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) & \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N - \text{diag}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}) \boldsymbol{\tau}_N \\ \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) - \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}) & \frac{N}{2(1 - \delta)^2} \end{array} \right], \tag{20}$$

$\boldsymbol{\tau}_N$  is an  $N \times 1$  vector of ones,  $\odot$  is the Hadamard product matrix operator, and  $\mathbf{A}_0 = \mathbf{G}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W})^{-1}$ .<sup>2</sup> Hence,  $\partial Q_N(\mathbf{0}) / \partial \boldsymbol{\varphi} = \mathbf{0}$ , and using mean-value

<sup>2</sup>Using (18) we have  $\mathbf{I}_N - \mathbf{D}\mathbf{G}_0 = (\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}) (\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})^{-1}$ , and hence

$$\mathbf{G}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})^{-1} (\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W}) (\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W})^{-1} = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W})^{-1},$$

as required.

extension of  $Q_N(\boldsymbol{\varphi})$  we have

$$Q_N(\boldsymbol{\varphi}) = \frac{1}{2N} \boldsymbol{\varphi}' \boldsymbol{\Lambda}_N(\bar{\boldsymbol{\varphi}}) \boldsymbol{\varphi} \geq 0, \quad (21)$$

where  $\bar{\boldsymbol{\varphi}} = (\bar{\boldsymbol{d}}, \bar{\boldsymbol{\delta}})' = [\bar{\psi}_1 - \psi_{10}, \bar{\psi}_2 - \psi_{20}, \dots, \bar{\psi}_N - \psi_{N0}, (\bar{\sigma}^2 - \sigma_0^2) / \bar{\sigma}^2]'$ ,  $\bar{\psi}_i$  lies on the line segment joining 0 and  $\psi_{i0}$ , for  $i = 1, 2, \dots, N$  and  $0 < \bar{\sigma}^2 \leq \sigma_0^2$ . Furthermore,

$$Q_N(\boldsymbol{\varphi}) \geq \frac{\boldsymbol{\varphi}' \boldsymbol{\varphi}}{2N} \lambda_{\min}[\boldsymbol{\Lambda}_N(\bar{\boldsymbol{\varphi}})] \geq 0,$$

and  $Q_N(\boldsymbol{\varphi}) = 0$  if and only if  $N^{-1} \boldsymbol{\varphi}' \boldsymbol{\varphi} = 0$ , so long as  $\lambda_{\min}[\boldsymbol{\Lambda}_N(\bar{\boldsymbol{\varphi}})] > 0$  for all values of  $\bar{\psi}_i$  lying on the line segment joining 0 and  $\psi_{i0}$  and  $0 < \bar{\sigma}^2 \leq \sigma_0^2$ . Hence, for a fixed  $N$ , the parameters  $\psi_{i0}$ , for  $i = 1, 2, \dots, N$  and  $\sigma_0^2$  are globally identified if  $\lambda_{\min}[\boldsymbol{\Lambda}_N(\bar{\boldsymbol{\varphi}})] > 0$ , for all values of  $\bar{\psi}_i$  lying between 0 and  $\psi_{i0}$ , and  $0 < \bar{\sigma}^2 \leq \sigma_0^2$ , and are locally identified if  $\lambda_{\min}[\boldsymbol{\Lambda}_N(\mathbf{0})] > 0$ . For future reference we note that

$$\boldsymbol{\Lambda}_N(\mathbf{0}) = \begin{bmatrix} (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) & \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \\ \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) & \frac{N}{2} \end{bmatrix}. \quad (22)$$

Also, using the inverse of partitioned matrices, a necessary and sufficient condition for the identification of  $\boldsymbol{\psi}_0$  is given by the following rank condition

$$\text{Rank} \left[ (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) - \frac{2}{N} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) \right] = N, \quad (23)$$

which reduces to the simple inequality condition under homogeneity of spatial coefficients,  $\psi_i = \psi$ , for all  $i$  (or  $\boldsymbol{\Psi}_0 = \psi \mathbf{I}_N$ )

$$\boldsymbol{\tau}'_N (\mathbf{G}_0 \odot \mathbf{G}'_0) \boldsymbol{\tau}_N + \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) \boldsymbol{\tau}_N - \frac{2}{N} [\boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N]^2 > 0.$$

This condition further simplifies to

$$\frac{1}{2} [N^{-1} \text{tr}(\mathbf{G}_0^2) + N^{-1} \text{tr}(\mathbf{G}_0 \mathbf{G}'_0)] > [N^{-1} \text{tr}(\mathbf{G}_0)]^2. \quad (24)$$

The two identification conditions (23) and (24) provide an interesting contrast between the spatial models with homogeneous and heterogeneous spatial effects. Under the homogeneity restriction, using (24) it is easily seen that  $\psi_0 = 0$  is identified so long as  $N^{-1} \text{tr}(\mathbf{W}' \mathbf{W}) > 0$  for all  $N$  (including as  $N \rightarrow \infty$ ). But the same is not true of the individual spatial coefficients,  $\psi_i$ , under heterogeneity. To see this, note that when  $\psi_{i0} = 0$ , the  $i^{\text{th}}$  row of  $\mathbf{W}$  no longer enters the log-likelihood function and the  $i^{\text{th}}$  unit becomes totally disconnected from the rest of the units. As an example consider the case where  $N = 2$  and note that in this case

$$\mathbf{W} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$



and hence the rank condition (23) reduces to

$$\text{Rank} \begin{pmatrix} 1 + \psi_{20}^2 & 1 - \psi_{10}\psi_{20} \\ 1 - \psi_{10}\psi_{20} & 1 + \psi_{10}^2 \end{pmatrix} = 2.$$

This condition is clearly satisfied unless  $\psi_{10} = \psi_{20} = 0$ .

Consider now the case at the other extreme where  $N$  is allowed to rise without bounds. Then, for identification, it is also required that  $Q_N(\varphi)$  tends to a finite limit as  $N \rightarrow \infty$ . To simplify the exposition, we focus on local identification. Using (22) we note that

$$Q_N(\varphi) = \frac{1}{2N} \mathbf{d}' [(\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0)] \mathbf{d} + \delta \left[ \frac{\mathbf{d}' \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N}{N} \right] + \frac{\delta^2}{4},$$

where  $\mathbf{d}' \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N = \sum_{i=1}^N d_i g_{0,ii}$ , where  $g_{0,ii}$  is the  $i^{\text{th}}$  diagonal element of  $\mathbf{G}_0$ . Also, since  $Q_N(\varphi) \geq 0$ , then

$$\begin{aligned} Q_N(\varphi) = |Q_N(\varphi)| &\leq \frac{1}{2} \left( \frac{\mathbf{d}' \mathbf{d}}{N} \right) \lambda_{\max} [(\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0)] \\ &+ |\delta| \left( \frac{\sum_{i=1}^N g_{0,ii}^2}{N} \right)^{1/2} \left( \frac{\mathbf{d}' \mathbf{d}}{N} \right)^{1/2} + \frac{\delta^2}{4}. \end{aligned}$$

Furthermore,

$$\lambda_{\max} [(\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0)] \leq \|(\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0)\|_1,$$

and by Lemma 2 of Appendix A, and noting that

$$\|\mathbf{G}_0 \odot \mathbf{G}'_0\|_1 \leq \|\mathbf{G}_0\|_1 \|\mathbf{G}'_0\|_{\infty},$$

then

$$Q_N(\varphi) \leq \frac{K}{2} \left( \frac{\mathbf{d}' \mathbf{d}}{N} \right) + \frac{\delta^2}{4},$$

where  $K$  is a fixed constant which is bounded in  $N$ . As a result,  $Q_N(\varphi)$  is also bounded in  $N$ , noting that  $\psi_i$  and  $\sigma^2$  are assumed be bounded. Therefore, assuming that  $\lambda_{\min}[\boldsymbol{\Lambda}_N(\mathbf{0})] > 0$  for all values of  $N$  (including when  $N \rightarrow \infty$ ), we have

$$\frac{K}{2} \left( \frac{\mathbf{d}' \mathbf{d}}{N} \right) + \frac{\delta^2}{4} \geq Q_N(\varphi) \geq \left( \frac{\boldsymbol{\varphi}' \boldsymbol{\varphi}}{2N} \right) \lambda_{\min}[\boldsymbol{\Lambda}_N(\mathbf{0})] > 0,$$

and the  $N$  spatial parameters,  $\psi_i$  for  $i = 1, 2, \dots, N$ , are identified by the condition

$$\lim_{N \rightarrow \infty} \left( \frac{\mathbf{d}' \mathbf{d}}{N} \right) = \lim_{N \rightarrow \infty} \left[ \frac{\sum_{i=1}^N (\psi_i - \psi_{i0})^2}{N} \right] = 0.$$

Clearly, as we have shown above, all the spatial parameters are identified if  $\lambda_{\min}[\boldsymbol{\Lambda}_N(\bar{\boldsymbol{\varphi}})] > 0$ . But if  $N$  is sufficiently large, some of the parameters might be unidentified. This is because the

condition  $\lim_{N \rightarrow \infty} \sum_{i=1}^N (\psi_i - \psi_{i0})^2 / N = 0$ , does not necessarily imply  $\psi_i = \psi_{i0}$  for all  $i$ .

The main identification result is summarised in the following proposition.

**Proposition 1** *Consider the heterogeneous spatial autoregressive (HSAR) model given by (3) and suppose that Assumptions 1 to 5 hold and the invertibility condition (7) is met. Then the true parameter values,  $\sigma_0^2$  and  $\psi_{i0}$ , for  $i = 1, 2, \dots, N$ , are identified if  $\lambda_{\min}[\mathbf{\Lambda}_N(\bar{\varphi})] > 0$ , where  $\mathbf{\Lambda}_N(\varphi)$  is defined by (20). In the case where  $N$  is large and rising without bounds,  $\lambda_{\min}[\mathbf{\Lambda}_N(\bar{\varphi})] > 0$  is necessary for identification but need not be sufficient if the aim is to identify all the spatial coefficients,  $\psi_i$  for  $i = 1, 2, \dots, N$ , as  $N \rightarrow \infty$ . For local identification the condition simplifies to*

$$\lambda_{\min}(\mathbf{H}_0) > \epsilon > 0, \text{ for all } N,$$

where

$$\mathbf{H}_0 = \mathbf{G}_0 \odot \mathbf{G}'_0 + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) - \frac{2}{N} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0), \quad (25)$$

$\mathbf{G}_0 \equiv \mathbf{G}(\psi_0) = \mathbf{W}(\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1}$ ,  $\boldsymbol{\Psi}_0 = \text{diag}(\psi_0)$ ,  $\psi_0 = (\psi_{10}, \psi_{20}, \dots, \psi_{N0})'$ , the  $i^{\text{th}}$  element of  $\text{diag}(\mathbf{G}_0 \mathbf{G}'_0)$  is given by  $\mathbf{g}'_{0i} \mathbf{g}_{0i}$ , and  $\mathbf{W}$  is the spatial weight matrix defined under Assumption 1. For large  $N$  it is also required that  $\lambda_{\max}(\mathbf{H}_0) < K$  for all  $N$ .

## 4 Consistency of the QML estimator

The quasi maximum likelihood estimator,  $\hat{\boldsymbol{\theta}}_T = (\hat{\boldsymbol{\psi}}'_T, \hat{\sigma}_T^2)'$ , is defined by

$$\hat{\boldsymbol{\theta}}_T = \arg \max_{\boldsymbol{\theta} \in \Theta} \ell(\boldsymbol{\theta}). \quad (26)$$

We now show that, under Assumptions 1 to 5 and assuming that the invertibility condition (7) holds, we have  $p \lim_{T \rightarrow \infty} (\hat{\boldsymbol{\theta}}_T) = \boldsymbol{\theta}_0 = (\boldsymbol{\psi}'_0, \sigma_0^2)'$ , namely,  $\hat{\boldsymbol{\theta}}_T$  is a consistent estimator of  $\boldsymbol{\theta}$ . First, for the true parameter values  $\boldsymbol{\theta}_0 \in \Theta$  we have

$$\boldsymbol{\theta}_0 = \arg \max_{\boldsymbol{\theta} \in \Theta} E_0 \left[ \frac{\ell(\boldsymbol{\theta})}{T} \right],$$

which is unique over  $\Theta$  (see Section 3). Therefore,

$$\lim_{T \rightarrow \infty} E_0 \left[ \frac{\ell(\boldsymbol{\theta}_0)}{T} \right] \geq \lim_{T \rightarrow \infty} E_0 \left[ \frac{\ell(\boldsymbol{\theta})}{T} \right]. \quad (27)$$

Also,

$$p \lim_{T \rightarrow \infty} \frac{\ell(\hat{\boldsymbol{\theta}}_T)}{T} \geq p \lim_{T \rightarrow \infty} \frac{\ell(\boldsymbol{\theta})}{T}.$$

But under Assumption 4 we have

$$p \lim_{T \rightarrow \infty} \left[ \frac{\ell(\boldsymbol{\theta})}{T} \right] = \lim_{T \rightarrow \infty} E_0 \left[ \frac{\ell(\boldsymbol{\theta})}{T} \right],$$

and given (27) we have that

$$p \lim_{T \rightarrow \infty} \frac{\ell(\hat{\boldsymbol{\theta}}_T)}{T} \geq \lim_{T \rightarrow \infty} E_0 \left[ \frac{\ell(\boldsymbol{\theta})}{T} \right] \leq \lim_{T \rightarrow \infty} E_0 \left[ \frac{\ell(\boldsymbol{\theta}_0)}{T} \right].$$

Therefore,

$$p \lim_{T \rightarrow \infty} \frac{\ell(\hat{\boldsymbol{\theta}}_T)}{T} = \lim_{T \rightarrow \infty} E_0 \left[ \frac{\ell(\boldsymbol{\theta}_0)}{T} \right],$$

and since  $\boldsymbol{\theta}_0$  presents the unique maximum of  $E_0 [T^{-1}\ell(\boldsymbol{\theta})]$ , then

$$p \lim_{T \rightarrow \infty} (\hat{\boldsymbol{\theta}}_T) = \boldsymbol{\theta}_0. \quad (28)$$

## 5 Asymptotic normality

The QML estimator,  $\hat{\boldsymbol{\theta}}_T = (\hat{\boldsymbol{\psi}}_T', \hat{\sigma}_T^2)'$ , solves the score function,  $\mathbf{U}_T(\hat{\boldsymbol{\theta}}_T) = \partial \ell(\hat{\boldsymbol{\theta}}_T) / \partial \boldsymbol{\theta} = \mathbf{0}$ , associated with the log-likelihood function given by (11). For a given  $N$ , consider now the mean expansion of  $\mathbf{U}(\hat{\boldsymbol{\theta}}_T)$  around  $\boldsymbol{\theta}_0$

$$\mathbf{0} = \frac{1}{\sqrt{T}} \mathbf{U}_T(\hat{\boldsymbol{\theta}}_T) = \frac{1}{\sqrt{T}} \mathbf{U}_T(\boldsymbol{\theta}_0) - \mathbf{H}_T(\bar{\boldsymbol{\theta}}_T) \sqrt{T} (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0),$$

where  $\mathbf{H}_T(\bar{\boldsymbol{\theta}}_T) = -\frac{1}{T} \frac{\partial^2 \ell(\bar{\boldsymbol{\theta}}_T)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$ , and  $\bar{\boldsymbol{\theta}}_T$  lies on the line segments joining  $\hat{\boldsymbol{\theta}}_T$  and  $\boldsymbol{\theta}_0$ .<sup>3</sup> In view of the consistency result, (28), we also have  $p \lim_{T \rightarrow \infty} (\bar{\boldsymbol{\theta}}_T) = \boldsymbol{\theta}_0$ , and given that under our assumptions  $\mathbf{H}_T(\bar{\boldsymbol{\theta}}_T)$  is a smooth function of  $\bar{\boldsymbol{\theta}}_T$ , then

$$\mathbf{H}_T(\bar{\boldsymbol{\theta}}_T) \rightarrow_p E_0 \left( -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = \mathbf{H}(\boldsymbol{\theta}_0),$$

which is a positive definite matrix under the identification condition. Thus, the limiting distribution of  $\sqrt{T}(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0)$  is given by the limiting distribution of  $T^{-1/2} \mathbf{U}_T(\boldsymbol{\theta}_0)$ . But  $\mathbf{U}_T(\boldsymbol{\theta}_0) = \left( \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}'}, \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2} \right)'$ ,

$$\frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}} = -T \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N + \frac{1}{\sigma_0^2} \sum_{t=1}^T \mathbf{y}_t^* \odot (\mathbf{y}_t - \boldsymbol{\Psi}_0 \mathbf{y}_t^*),$$

---

<sup>3</sup>More specifically, denoting typical elements of  $\boldsymbol{\theta}_0$ ,  $\bar{\boldsymbol{\theta}}_T$ , and  $\hat{\boldsymbol{\theta}}_T$  by  $\theta_{i0}$ ,  $\bar{\theta}_{iT}$ , and  $\hat{\theta}_{iT}$ , then the  $(i, j)$  element of  $\mathbf{H}(\bar{\boldsymbol{\theta}}_T)$  is evaluated at  $(\bar{\theta}_{iT}, \bar{\theta}_{jT})$ , where  $\bar{\theta}_{iT}$  is a convex combination of  $\hat{\theta}_{iT}$  and  $\theta_{i0}$ .

and

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2} &= -\frac{NT}{2\sigma_0^2} + \frac{1}{2\sigma_0^4} \sum_{t=1}^T \mathbf{y}'_t (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})' (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W}) \mathbf{y}_t \\ &= \frac{NT}{2\sigma_0^2} \left[ \frac{1}{\sigma_0^2} \left( \frac{1}{NT} \sum_{t=1}^T \boldsymbol{\varepsilon}'_t \boldsymbol{\varepsilon}_t \right) - 1 \right],\end{aligned}$$

where  $\mathbf{G}_0 = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} = (g_{0,ij})$ ,  $\mathbf{y}_t^* = (y_{1t}^*, y_{2t}^*, \dots, y_{Nt}^*)' = \mathbf{W} \mathbf{y}_t$ , and  $\boldsymbol{\tau}_N$  is an  $N \times 1$  vector of ones.

Consider first the  $i^{\text{th}}$  component of  $\partial \ell(\boldsymbol{\theta}_0) / \partial \psi$ , and note that it can be written as

$$\frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i} = -T g_{0,ii} + \frac{1}{\sigma_0^2} \sum_{t=1}^T y_{it}^* \varepsilon_{it}.$$

Also  $y_{it}^* = \mathbf{e}'_{i,N} \mathbf{G}_0 \boldsymbol{\varepsilon}_t$ , where  $\mathbf{e}_{i,N}$  is an  $N$  dimensional vector with its  $i^{\text{th}}$  element unity and zeros elsewhere. Then

$$\frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i} = \frac{1}{\sqrt{T}} \sum_{t=1}^T \eta_{it}, \quad (29)$$

where

$$\eta_{it} = \mathbf{e}'_{i,N} \mathbf{G}_0 \boldsymbol{\zeta}_t \zeta_{it} - g_{0,ii},$$

$\boldsymbol{\zeta}_t = \boldsymbol{\varepsilon}_t / \sigma_0 = (\zeta_{1t}, \zeta_{2t}, \dots, \zeta_{Nt})'$ ,  $E(\boldsymbol{\zeta}_t) = \mathbf{0}$ ,  $E(\boldsymbol{\zeta}_t \zeta_{it}) = \mathbf{e}_{i,N}$ , and  $\text{Var}(\boldsymbol{\zeta}_t) = \mathbf{I}_N$ . Also,  $E(\eta_{it}) = 0$ , and

$$E(\eta_{it}^2) = \mathbf{e}'_{i,N} \mathbf{G}_0 E(\boldsymbol{\zeta}_t \boldsymbol{\zeta}'_t \zeta_{it}^2) \mathbf{G}'_0 \mathbf{e}_{i,N} - g_{0,ii}^2,$$

which yields

$$\text{Var}(\eta_{it}) = g_{0,ii}^2 [E(\zeta_{it}^4) - 2] + \sum_{j=1}^N g_{0,ij}^2. \quad (30)$$

Hence,  $E_0 \left[ \frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i} \right] = 0$ , and since, under Assumption 2,  $\eta_{it}$ 's are distributed independently over  $t$ , then so long as  $E(|\varepsilon_{it}|^{4+\epsilon}) < K$  for some small positive  $\epsilon > 0$ ,  $\frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i}$  will be distributed as  $N(0, \omega_{ii})$ , where

$$\omega_{ii} = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \text{Var}(\eta_{it}) = g_{0,ii}^2 \left[ \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(\zeta_{it}^4) - 2 \right] + \sum_{j=1}^N g_{0,ij}^2.$$

The above result holds even if  $N \rightarrow \infty$ , so long as  $\sum_{j=1}^N g_{0,ij}^2$  is bounded in  $N$ . This latter condition is met if the maximum row sum norm of matrix  $\mathbf{G}_0$  is bounded, which is ensured under Assumption 1 and assuming condition (7) holds. See Lemma 2 of Appendix A.

Similarly,  $T^{-1/2} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2}$  can be written as

$$\frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^2} \left[ \frac{1}{\sqrt{T}} \sum_{t=1}^T \left( \frac{\boldsymbol{\varepsilon}'_t \boldsymbol{\varepsilon}_t}{\sigma_0^2} - N \right) \right].$$

or

$$\frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^2 \sqrt{T}} \sum_{t=1}^T \xi_{tN}, \quad (31)$$

where  $\xi_{tN} = \sum_{i=1}^N (\zeta_{it}^2 - 1)$ . Hence, for a fixed  $N$  and under Assumption 2,  $\xi_{tN}$ , for  $t = 1, 2, \dots, T$ , have zero means and are serially independent with a finite variance, if  $E(|\varepsilon_{it}|^{4+\epsilon}) < K$ , for some small positive  $\epsilon$ . Hence, under these conditions,  $T^{-1/2} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2}$  also tends to  $N(0, \omega_0^2)$ , where

$$\omega_0^2 = \left( \frac{1}{4\sigma_0^4} \right) \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \sum_{i=1}^N \text{Var}(\zeta_{it}^2).$$

For large  $N$ , one needs to consider the limiting distribution of  $(NT)^{-1/2} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2}$ , which is also asymptotically normally distributed with variance now given by

$$\left( \frac{1}{4\sigma_0^4} \right) \lim_{(N,T) \rightarrow \infty} (NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N \text{Var}(\zeta_{it}^2).$$

Finally, using (29) and (31) we have (for  $i \neq j$ )

$$\begin{aligned} E_0 \left[ \frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i} \frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_j} \right] &= \frac{1}{T} \sum_{t=1}^T \sum_{t'=1}^T E(\eta_{it} \eta_{jt'}) \\ &= \frac{1}{T} \sum_{t=1}^T \sum_{t'=1}^T E[(\mathbf{e}'_{i,N} \mathbf{G}_0 \boldsymbol{\zeta}_t \zeta_{it} - g_{0,ii}) (\mathbf{e}'_{j,N} \mathbf{G}_0 \boldsymbol{\zeta}_{t'} \zeta_{jt'} - g_{0,jj})] \\ &= \frac{1}{T} \sum_{t=1}^T E[(\mathbf{e}'_{i,N} \mathbf{G}_0 \boldsymbol{\zeta}_t \zeta_{it} - g_{0,ii}) (\mathbf{e}'_{j,N} \mathbf{G}_0 \boldsymbol{\zeta}_t \zeta_{jt} - g_{0,jj})] \\ &= \mathbf{e}'_{i,N} \mathbf{G}_0 E(\zeta_{it} \zeta_{jt} \boldsymbol{\zeta}_t \boldsymbol{\zeta}'_t) \mathbf{G}'_0 \mathbf{e}_{j,N} - g_{0,ii} g_{0,jj} \\ &= g_{0,ij} g_{0,ji}, \text{ for } i \neq j. \end{aligned}$$

Recall from (30) that

$$E_0 \left[ \frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i} \frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i} \right] = E(\eta_{it}^2) = g_{0,ii}^2 [E(\zeta_{it}^4) - 2] + \sum_{j=1}^N g_{0,ij}^2.$$

Finally,

$$\begin{aligned}
E_0 \left[ \frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \sigma^2} \frac{1}{\sqrt{T}} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \psi_i} \right] &= \frac{1}{2\sigma_0^2 T} \sum_{t=1}^T \sum_{t'=1}^T E(\xi_{tN} \eta_{it'}) \\
&= \frac{1}{2\sigma_0^2 T} \sum_{t=1}^T \sum_{t'=1}^T \sum_{i=1}^N E[(\zeta_{it}^2 - 1)(\mathbf{e}'_{i,N} \mathbf{G}_0 \boldsymbol{\zeta}_{t'} \zeta_{it'} - g_{0,ii})] \\
&= \frac{1}{2\sigma_0^2 T} \sum_{t=1}^T \sum_{t'=1}^T \sum_{i=1}^N E[\zeta_{it}^2 (\mathbf{e}'_{i,N} \mathbf{G}_0 \boldsymbol{\zeta}_{t'} \zeta_{it'} - g_{0,ii})] \\
&= \frac{g_{0,ii} \left[ \frac{1}{T} \sum_{t=1}^T E(\zeta_{it}^4) - 1 \right]}{2\sigma_0^2}.
\end{aligned}$$

Let

$$\gamma = \left[ \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T E(\zeta_{it}^4) - 1 \right] = \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \text{Var}(\zeta_{it}^2). \quad (32)$$

We collect the various terms of

$$\mathbf{J}(\boldsymbol{\theta}_0, \gamma) = \lim_{T \rightarrow \infty} E_0 \left[ \frac{1}{T} \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right)' \right] = (J_{0,ij}),$$

where

$$\begin{aligned}
J_{0,ii} &= g_{0,ii}^2 (\gamma - 1) + \sum_{j=1}^N g_{0,ij}^2, \text{ for } i = 1, 2, \dots, N \\
J_{0,ij} &= g_{0,ij} g_{0,ji}, \text{ for } i \neq j = 1, 2, \dots, N \\
J_{0,iN+1} &= \frac{\gamma g_{0,ii}}{2\sigma_0^2}, \text{ for } i = 1, 2, \dots, N, \\
J_{0,ii} &= \frac{N\gamma}{4\sigma_0^4}, \text{ for } i = N + 1.
\end{aligned}$$

Alternatively, in matrix notation, we have

$$\mathbf{J}(\boldsymbol{\theta}_0, \gamma) = \begin{pmatrix} (\mathbf{G}_0 \odot \mathbf{G}'_0) \odot [\boldsymbol{\tau}_N \boldsymbol{\tau}'_N + (\gamma - 2)\mathbf{I}_N] + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) & \frac{\gamma}{2\sigma_0^2} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \\ \frac{\gamma}{2\sigma_0^2} \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) & \frac{N\gamma}{4\sigma_0^4} \end{pmatrix}. \quad (33)$$

Having established that the score vector is asymptotically normally distributed, it is now easily seen that, for a fixed  $N$  and as  $T \rightarrow \infty$ ,

$$\sqrt{T} (\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta}_0) \rightarrow_d N \left[ \mathbf{0}, \text{AsyVar}(\hat{\boldsymbol{\theta}}_T) \right], \quad (34)$$

where  $\text{AsyVar}(\hat{\boldsymbol{\theta}}_T) = \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0, \gamma) \mathbf{H}^{-1}(\boldsymbol{\theta}_0)$  and  $\mathbf{H}(\boldsymbol{\theta}_0) = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]$ . In the case where the errors,  $\varepsilon_{it}$ , are Gaussian,  $\gamma = 2$  and, as to be expected,  $\mathbf{H}(\boldsymbol{\theta}_0) = \mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ . This is

easily verified noting that

$$\mathbf{H}(\boldsymbol{\theta}_0) = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} \\ \mathbf{H}_{21} & \mathbf{H}_{22} \end{pmatrix}_{(N+1) \times (N+1)}, \quad (35)$$

where

$$\begin{aligned} \mathbf{H}_{11} &= E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right] = (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} \left[ \frac{1}{\sigma_0^2} \frac{\sum_{t=1}^T E_0 (y_{it}^* \odot y_{it}^*)}{T}, i = 1, \dots, N \right] \\ \mathbf{H}_{12} &= E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi} \partial \sigma^2} \right] = \frac{1}{\sigma_0^4} \frac{\sum_{t=1}^T E_0 [\mathbf{y}_t^* \odot (\mathbf{y}_t - \boldsymbol{\Psi}_0 \mathbf{y}_t^*)]}{T} \\ \mathbf{H}_{22} &= E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial (\sigma^2)^2} \right] = -\frac{N}{2\sigma_0^4} + \frac{1}{\sigma_0^6} \sum_{t=1}^T \frac{E_0 [\mathbf{y}_t' (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})' (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W}) \mathbf{y}_t]}{T} \\ \mathbf{H}_{21} &= \mathbf{H}'_{12}. \end{aligned}$$

Since  $\mathbf{y}_t^* = (y_{1t}^*, y_{2t}^*, \dots, y_{Nt}^*)' = \mathbf{W} \mathbf{y}_t$  then, for each  $i = 1, 2, \dots, N$ , we have

$$\frac{\sum_{t=1}^T y_{it}^* \odot y_{it}^*}{T} = \frac{\sum_{t=1}^T (y_{it}^*)^2}{T} = \frac{\sum_{t=1}^T (\mathbf{w}'_i \mathbf{y}_t)^2}{T} = \mathbf{w}'_i \left( \frac{1}{T} \sum_{t=1}^T \mathbf{y}_t \mathbf{y}_t' \right) \mathbf{w}_i.$$

Then,

$$\begin{aligned} E_0 \left[ \frac{\sum_{t=1}^T (y_{it}^*)^2}{T \sigma_0^2} \right] &= \frac{1}{\sigma_0^2} \mathbf{w}'_i \left[ \frac{1}{T} \sum_{t=1}^T E_0 (\mathbf{y}_t \mathbf{y}_t') \right] \mathbf{w}_i \\ &= \frac{1}{\sigma_0^2} \mathbf{w}'_i (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \left[ \frac{1}{T} \sum_{t=1}^T E_0 (\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t') \right] (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi}_0)^{-1} \mathbf{w}_i \\ &= \frac{1}{\sigma_0^2} \mathbf{w}'_i (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} (\sigma_0^2 \mathbf{I}_N) (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi}_0)^{-1} \mathbf{w}_i \\ &= \mathbf{g}'_{0i} \mathbf{g}_{0i}, \end{aligned}$$

where  $\mathbf{g}'_{0i}$  is the  $i^{\text{th}}$  row of  $\mathbf{G}_0$ . Stacking over the  $N$  units and collecting the different components we obtain

$$\mathbf{H}_{11} = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right] = (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} (\mathbf{G}_0 \mathbf{G}'_0).$$

Next, for each  $i = 1, 2, \dots, N$ , we have (recall that  $\varepsilon_{it} = y_{it} - \psi_{i0} y_{it}^*$ ),

$$E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \psi_i \partial \sigma^2} \right] = \frac{1}{\sigma_0^4} E_0 \left( \frac{\sum_{t=1}^T y_{it}^* \varepsilon_{it}}{T} \right) = \frac{1}{\sigma_0^2} g_{0,ii},$$

where the last equality follows from

$$\begin{aligned}
E_0 \left( \frac{\sum_{t=1}^T y_{it}^* \varepsilon_{it}}{T} \right) &= E_0 \left( \frac{\sum_{t=1}^T \mathbf{w}'_i \mathbf{y}_t \varepsilon_{it}}{T} \right) \\
&= E_0 \left[ \frac{\sum_{t=1}^T \mathbf{w}'_i (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \boldsymbol{\varepsilon}_t \varepsilon_{it}}{T} \right] = \sigma_0^2 \mathbf{w}'_i (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \mathbf{e}_{i,N} \\
&= \sigma_0^2 g_{0,ii},
\end{aligned}$$

and, as before,  $\mathbf{e}_{i,N}$  is an  $N$  dimensional vector with its  $i^{\text{th}}$  element unity and zeros elsewhere. Stacking over the  $N$  units we obtain

$$\mathbf{H}_{12} = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \psi \partial \sigma^2} \right] = \frac{1}{\sigma_0^2} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N,$$

where as before  $\boldsymbol{\tau}_N$  is an  $N \times 1$  vector of ones. Finally,

$$\mathbf{H}_{22} = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial (\sigma^2)^2} \right] = E_0 \left( -\frac{N}{2\sigma_0^4} + \frac{1}{\sigma_0^6} \sum_{t=1}^T \frac{\boldsymbol{\varepsilon}'_t \boldsymbol{\varepsilon}_t}{T} \right) = \frac{N}{2\sigma_0^4}.$$

Using the above results we now have

$$\mathbf{H}(\boldsymbol{\theta}_0) = \begin{pmatrix} (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) & \frac{1}{\sigma_0^2} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \\ \frac{1}{\sigma_0^2} \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) & \frac{N}{2\sigma_0^4} \end{pmatrix}, \quad (36)$$

which is equal to  $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$  defined by (33) for  $\gamma = 2$ , as required.

To obtain the inverse of  $\mathbf{H}(\boldsymbol{\theta}_0)$ , let  $\mathbf{H}_{11.2} = \mathbf{H}_{11} - \mathbf{H}_{12} \mathbf{H}_{22}^{-1} \mathbf{H}_{21}$  and note that

$$\mathbf{H}^{-1}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{H}_{11.2}^{-1} & -\mathbf{H}_{11.2}^{-1} \mathbf{H}_{12} \mathbf{H}_{22}^{-1} \\ -\mathbf{H}_{22}^{-1} \mathbf{H}_{21} \mathbf{H}_{11.2}^{-1} & \mathbf{H}_{22}^{-1} + \mathbf{H}_{22}^{-1} \mathbf{H}_{21} \mathbf{H}_{11.2}^{-1} \mathbf{H}_{12} \mathbf{H}_{22}^{-1} \end{pmatrix}.$$

The asymptotic covariance matrix of  $\sqrt{T} \hat{\boldsymbol{\psi}}_T$  is given by  $\mathbf{H}_{11.2}^{-1}$ , where

$$\begin{aligned}
\mathbf{H}_{11.2} &= \mathbf{H}_{11} - \mathbf{H}_{12} \mathbf{H}_{22}^{-1} \mathbf{H}_{21} \\
&= (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) - \frac{2}{N} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0).
\end{aligned}$$

Under the identification conditions established in Proposition 1,  $\mathbf{H}_{11.2}$  is full rank. Further,  $\mathbf{G}_0$  has bounded maximum absolute row and column sum norms by Lemma 2 of Appendix A, thus it also follows that  $(\mathbf{G}_0 \odot \mathbf{G}'_0)$  has a bounded maximum absolute row (column) sum norm, and hence  $\mathbf{H}_{11.2}^{-1}$  has a bounded maximum absolute row (column) sum norm.

The main result of this section can now be summarised in the following proposition.

**Proposition 2** *Consider the heterogeneous spatial autoregressive (HSAR) model given by (3) and suppose that: (a) Assumptions 1 to 5 hold, (b) the invertibility condition (7) is met, (c) the*



$N \times N$  information matrix

$$\mathbf{H}_{11.2} = (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) - \frac{2}{N} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0),$$

is full rank, where  $\mathbf{G}_0 = \mathbf{W}(\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1}$ ,  $\boldsymbol{\Psi}_0 = \text{diag}(\boldsymbol{\psi}_0)$ ,  $\boldsymbol{\psi}_0 = (\psi_{10}, \psi_{20}, \dots, \psi_{N0})'$ , the  $i^{\text{th}}$  element of  $\text{diag}(\mathbf{G}_0 \mathbf{G}'_0)$  is given by  $\mathbf{g}'_{0i} \mathbf{g}_{0i}$ , and  $\mathbf{W}$  is the spatial weight matrix, and (d)  $\varepsilon_{it} \sim \text{IIDN}(0, \sigma_0^2)$ . Then the maximum likelihood estimator of  $\boldsymbol{\psi}_0$ , denoted by  $\hat{\boldsymbol{\psi}}_T$  and computed by (26), has the following asymptotic distribution as  $T \rightarrow \infty$ ,

$$\sqrt{T}(\hat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}_0) \rightarrow_d N\left[0, \text{AsyVar}(\hat{\boldsymbol{\psi}}_T)\right],$$

where

$$\text{AsyVar}(\sqrt{T}\hat{\boldsymbol{\psi}}_T) = \left[ (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) - \frac{2}{N} \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) \right]^{-1}, \quad (37)$$

which does not depend on  $\sigma_0^2$ .

**Remark 4** In the case where  $\varepsilon_{it}$  are non-Gaussian but  $E(|\varepsilon_{it}|^{4+\epsilon}) < K$  holds for some small positive  $\epsilon$ , the quasi maximum likelihood estimator,  $\hat{\boldsymbol{\psi}}_T$ , continues to be normally distributed but its asymptotic covariance matrix is given by the upper  $N \times N$  partition of  $\mathbf{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0, \gamma) \mathbf{H}^{-1}(\boldsymbol{\theta}_0)$  where  $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$  and  $\mathbf{H}(\boldsymbol{\theta}_0)$  are defined by (33) and (36), respectively. Note that  $\gamma$  is defined by (32), and as noted earlier under Gaussian errors it takes the value of  $\gamma = 2$ , and we have  $\mathbf{J}(\boldsymbol{\theta}_0, 2) = \mathbf{H}(\boldsymbol{\theta}_0)$ .

### 5.1 Consistent estimation of $\text{AsyVar}(\hat{\boldsymbol{\theta}}_T)$

The asymptotic covariance matrix of  $\hat{\boldsymbol{\theta}}_T$  can be constructed directly by using the expressions derived for (33) and (36), yielding the standard formula

$$\text{AsyVar}(\hat{\boldsymbol{\theta}}_T) = \mathbf{H}^{-1}(\boldsymbol{\theta}_0), \quad (38)$$

when the information matrix equality holds in the case of  $\varepsilon_{it} \sim \text{IIDN}(0, \sigma_0^2)$  and  $\gamma = 2$ , and the sandwich formula

$$\text{AsyVar}(\hat{\boldsymbol{\theta}}_T) = \mathbf{H}^{-1}(\boldsymbol{\theta}_0) \mathbf{J}(\boldsymbol{\theta}_0, \gamma) \mathbf{H}^{-1}(\boldsymbol{\theta}_0), \quad (39)$$

otherwise. Consistent estimators of  $\mathbf{H}(\boldsymbol{\theta}_0)$  and  $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$  can be obtained by replacing  $\boldsymbol{\theta}_0$  with its QML estimator,  $\hat{\boldsymbol{\theta}}_T$ , and estimating  $\gamma$  by

$$\hat{\gamma} = (NT)^{-1} \sum_{t=1}^T \sum_{i=1}^N \left( \frac{\hat{\varepsilon}_{it}}{\hat{\sigma}} \right)^4 - 1,$$

where  $\hat{\varepsilon}_{it} = y_{it} - \hat{\psi}_i \sum_{j=1}^N w_{ij} y_{jt}$ , with  $\hat{\sigma}$  and  $\hat{\psi}_i$  being the QML estimators of  $\sigma_0$  and  $\psi_{i0}$ , respectively.

Alternatively, one can use the sample counterparts of  $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$  and  $\mathbf{H}(\boldsymbol{\theta}_0)$  and estimate the

covariance matrix of the QML estimators by

$$\widehat{Var}(\hat{\boldsymbol{\theta}}_T) = \mathbf{H}_T^{-1}(\hat{\boldsymbol{\theta}}_T), \quad (40)$$

and

$$\widehat{Var}(\hat{\boldsymbol{\theta}}_T) = \mathbf{H}_T^{-1}(\hat{\boldsymbol{\theta}}_T) \mathbf{J}_T(\hat{\boldsymbol{\theta}}_T, \hat{\gamma}) \mathbf{H}_T^{-1}(\hat{\boldsymbol{\theta}}_T), \quad (41)$$

where  $\mathbf{H}_T(\boldsymbol{\theta}) = -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$  and  $\mathbf{J}_T(\boldsymbol{\theta}) = \frac{1}{T} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)'$  with  $\ell(\boldsymbol{\theta}) = \sum_{t=1}^T \ell_t(\boldsymbol{\theta})$ . The first and second derivatives are provided in Appendix C for the general case discussed in the next section.

## 6 The HSAR model with heteroskedastic error variances and exogenous regressors

The heterogeneous spatial autoregressive model (3) can be extended to include exogenous regressors as well as heteroskedastic errors. In this case the spatial model can be written as

$$y_{it} = \psi_i \sum_{j=1}^N w_{ij} y_{jt} + \boldsymbol{\beta}'_i \mathbf{x}_{it} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T, \quad (42)$$

where, as before,  $\sum_{j=1}^N w_{ij} y_{jt} = \mathbf{w}'_i \mathbf{y}_t$ ,  $\mathbf{y}_t = (y_{1t}, y_{2t}, \dots, y_{Nt})'$  and  $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{iN})'$  with  $w_{ii} = 0$ . Now, we also introduce a  $k \times 1$  vector of exogenous regressors  $\mathbf{x}_{it} = (x_{i1,t}, x_{i2,t}, \dots, x_{ik,t})'$  with parameters  $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik})'$ . The above specification is sufficiently general and allows for the inclusion of fixed effects by setting one of the regressors, say  $x_{i1,t}$ , equal to unity. We also allow the errors,  $\varepsilon_{it}$ , to be cross-sectionally heteroskedastic, namely  $Var(\varepsilon_{it}) = \sigma_i^2$  for  $i = 1, 2, \dots, N$ . Stacking by individual units for each time period  $t$ , (42) becomes

$$\mathbf{y}_t = \boldsymbol{\Psi} \mathbf{W} \mathbf{y}_t + \mathbf{B} \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (43)$$

where  $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\psi})$  and  $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_N)'$ ,  $\mathbf{W} = (w_{ij})$ ,  $i, j = 1, 2, \dots, N$ ,  $\mathbf{B} = \text{diag}(\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_N)$ ,  $\mathbf{x}_t = (\mathbf{x}'_{1t}, \mathbf{x}'_{2t}, \dots, \mathbf{x}'_{Nt})'$ , and  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ . Under Assumption 1 and assuming condition (7) holds, then (43) can be written as

$$\mathbf{y}_t = (\mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W})^{-1} (\mathbf{B} \mathbf{x}_t + \boldsymbol{\varepsilon}_t), \quad t = 1, 2, \dots, T. \quad (44)$$

The (quasi) log-likelihood function can then be written as (assuming that the errors are Gaussian)

$$\begin{aligned} \ell(\boldsymbol{\theta}) = & -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^N \ln \sigma_i^2 + T \ln |\mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W}| \\ & - \frac{1}{2} \sum_{t=1}^T [(\mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W}) \mathbf{y}_t - \mathbf{B} \mathbf{x}_t]' \boldsymbol{\Sigma}_\varepsilon [(\mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W}) \mathbf{y}_t - \mathbf{B} \mathbf{x}_t], \end{aligned} \quad (45)$$

where  $\boldsymbol{\Sigma}_\varepsilon = \text{diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$ . Alternatively, it is often more convenient to write the above log-likelihood function as

$$\ell(\boldsymbol{\theta}) = -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^N \ln \sigma_i^2 + T \ln |\mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W}| - \frac{1}{2} \sum_{i=1}^N \frac{(\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{\sigma_i^2}, \quad (46)$$

where  $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_N, \sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$  with  $\boldsymbol{\beta}_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik})'$ ,  $\mathbf{X}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})'$  is the  $T \times k$  matrix of regressors on the  $i^{\text{th}}$  cross section unit with  $\mathbf{x}_{it} = (x_{i1,t}, x_{i2,t}, \dots, x_{ik,t})'$ ,  $\mathbf{y}_i = (y_{i1}, y_{i2}, \dots, y_{iT})'$ , and  $\mathbf{y}_i^* = (y_{i1}^*, y_{i2}^*, \dots, y_{iT}^*)'$ , with the elements  $y_{it}^* = \sum_{j=1}^N w_{ij} y_{jt} = \mathbf{w}'_i \mathbf{y}_t$ .

The analysis of this model can proceed as in the case of the HSAR model discussed earlier without further conceptual complications. We make the following assumptions:

**Assumption 6** The  $N(k+2) \times 1$  parameter vector,  $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_N, \sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$   $\in \Theta$ , is a sub-set of the  $N(k+2)$  dimensional Euclidean space,  $\mathbb{R}^{N(k+2)}$ .  $\Theta$  is a closed and bounded (compact) set and includes the true value of  $\boldsymbol{\theta}$ , denoted by  $\boldsymbol{\theta}_0$ , as an interior point, and  $\sup_i \|\boldsymbol{\beta}_i\|_1 < K$ .

**Assumption 7** The error terms,  $\varepsilon_{it}$ ,  $i = 1, 2, \dots, N$ ,  $N \geq 1$ ,  $t = 1, 2, \dots, T$  are independently distributed over  $i$  and  $t$ , have zero means, and constant variances,  $E(\varepsilon_{it}^2) = \sigma_i^2$ , where  $0 < \kappa_1 \leq \sigma_i^2 \leq \kappa_2 < \infty$ , and  $\kappa_1, \kappa_2$  are finite generic constants independent of  $N$ .

**Assumption 8** The regressors,  $\mathbf{x}_{it}$ , for  $i = 1, 2, \dots, N$ , are exogenous such that  $E(\mathbf{x}_{it} \varepsilon_{jt}) = 0$  for all  $i$  and  $j$ ,

$$T^{-1} \sum_{t=1}^T \mathbf{x}_{it} \varepsilon_{jt} \rightarrow_p \mathbf{0}, \text{ uniformly in } i \text{ and } j = 1, 2, \dots, N.$$

The covariance matrices  $E(\mathbf{x}_{it} \mathbf{x}'_{jt}) = \boldsymbol{\Sigma}_{ij}$ , for all  $i$  and  $j$ , are time-invariant and finite,  $\boldsymbol{\Sigma}_{ii}$  is non-singular,  $T^{-1} \mathbf{X}'_i \mathbf{X}_j \rightarrow_p \boldsymbol{\Sigma}_{ij}$ ,  $\sup_i [\lambda_{\max}(T^{-1} \mathbf{X}'_i \mathbf{X}_i)] < K$ , and  $\inf_i [\lambda_{\min}(T^{-1} \mathbf{X}'_i \mathbf{X}_i)] > 0$ .

**Remark 5** Assumption 8 is standard and allows for the regressors to be cross-sectionally correlated. This is sufficiently general and applies both when  $N$  is finite and when it rises without bounds.

The analysis of identification can now proceed as before and will be based on  $\mathbf{H}(\boldsymbol{\theta}_0) = E_0 \left[ -T^{-1} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right]$ . The relevant partial derivatives of the associated Hessian matrix are given in Appendix C. Setting  $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_N)'$ , and  $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$ , we write

$$\mathbf{H}(\boldsymbol{\theta}_0) = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right] = \begin{pmatrix} \mathbf{H}_{11} & \mathbf{H}_{12} & \mathbf{H}_{13} \\ \cdot & \mathbf{H}_{22} & \mathbf{H}_{23} \\ \cdot & \cdot & \mathbf{H}_{33} \end{pmatrix}_{N(k+2) \times N(k+2)}, \quad (47)$$

where

$$\begin{aligned} \mathbf{H}_{11} &= E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right], \quad \mathbf{H}_{12} = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\beta}'} \right], \quad \mathbf{H}_{13} = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi} \partial \boldsymbol{\sigma}^{2'}} \right], \\ \mathbf{H}_{22} &= E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \right], \quad \mathbf{H}_{23} = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\sigma}^{2'}} \right], \quad \mathbf{H}_{33} = E_0 \left[ -\frac{1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta}_0)}{\partial (\boldsymbol{\sigma}^2) \partial (\boldsymbol{\sigma}^{2'})} \right]. \end{aligned}$$

$\mathbf{H}_{11}$  is given by the  $N \times N$  matrix

$$\mathbf{H}_{11} = (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} \left[ \frac{1}{\sigma_i^2} \frac{1}{T} \sum_{t=1}^T E_0(y_{it}^{*2}), i = 1, 2, \dots, N \right],$$

where, as before,  $\mathbf{G}_0 = \mathbf{W}(\mathbf{I}_N - \Psi_0 \mathbf{W})^{-1}$  with its  $i^{\text{th}}$  row denoted by  $\mathbf{g}'_{0i}$ , and

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T E_0(y_{it}^{*2}) &= \mathbf{w}'_i (\mathbf{I}_N - \Psi_0 \mathbf{W})^{-1} [\mathbf{B}E(\mathbf{x}_t \mathbf{x}'_t) \mathbf{B}' + \Sigma_\varepsilon] (\mathbf{I}_N - \mathbf{W}' \Psi_0)^{-1} \mathbf{w}_i \\ &= \mathbf{g}'_{0i} [\mathbf{B}E(\mathbf{x}_t \mathbf{x}'_t) \mathbf{B}' + \Sigma_\varepsilon] \mathbf{g}_{0i}, \end{aligned}$$

where  $\mathbf{B}E(\mathbf{x}_t \mathbf{x}'_t) \mathbf{B}'$  is an  $N \times N$  matrix with its  $(r, s)$  element given by  $\beta'_r \Sigma_{rs} \beta_s$ .  $\mathbf{H}_{12}$  is an  $N \times kN$  matrix with its  $i^{\text{th}}$  row given by a  $1 \times kN$  vector of zeros except for its  $i^{\text{th}}$  block which is given by the  $1 \times k$  vector  $\sigma_i^{-2} E_0(T^{-1} \mathbf{y}_i^{*'} \mathbf{X}_i)$ , namely

$$\mathbf{H}_{12} = \begin{pmatrix} \sigma_1^{-2} E_0(T^{-1} \mathbf{y}_1^{*'} \mathbf{X}_1) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \sigma_2^{-2} E_0(T^{-1} \mathbf{y}_2^{*'} \mathbf{X}_2) & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \sigma_N^{-2} E_0(T^{-1} \mathbf{y}_N^{*'} \mathbf{X}_N) \end{pmatrix},$$

where

$$\begin{aligned} E_0(T^{-1} \mathbf{y}_i^{*'} \mathbf{X}_i) &= E_0 \left( T^{-1} \sum_{t=1}^T y_{it}^* \mathbf{x}'_{it} \right) = E_0 \left( T^{-1} \sum_{t=1}^T \mathbf{w}'_i \mathbf{y}_t \mathbf{x}'_{it} \right) \\ &= \mathbf{w}'_i (\mathbf{I}_N - \Psi_0 \mathbf{W})^{-1} \mathbf{B}E(\mathbf{x}_t \mathbf{x}'_{it}) \\ &= \mathbf{g}'_{0i} (\Sigma_{i1} \beta_1, \Sigma_{i2} \beta_2, \dots, \Sigma_{iN} \beta_N)'. \end{aligned}$$

$\mathbf{H}_{13}$  is an  $N \times N$  diagonal matrix with its  $i^{\text{th}}$  element given by  $\sigma_i^{-2} \mathbf{w}'_i (\mathbf{I}_N - \Psi_0 \mathbf{W})^{-1} \mathbf{e}_{i,N}$  =  $\sigma_i^{-2} g_{0,ii}$ ,  $\mathbf{H}_{22}$  is an  $Nk \times Nk$  block diagonal matrix with its  $i^{\text{th}}$  block given by  $\sigma_i^{-2} \Sigma_{ii}$ ,  $\mathbf{H}_{23} = \mathbf{0}$ , and  $\mathbf{H}_{33} = \text{diag}(1/2\sigma_1^4, 1/2\sigma_2^4, \dots, 1/2\sigma_N^4)$ .

For a finite  $N$  and as  $T \rightarrow \infty$  all the parameters are identified under Assumptions 1, 4, 5, 6, 7, and 8, and assuming that  $\mathbf{H}(\boldsymbol{\theta}_0)$ , defined by (47), is a positive definite matrix. Since under these Assumptions,  $\sigma_i^2 > 0$ ,  $\Sigma_{ii} > 0$ , for all  $i$ , then once  $\psi'_i$ 's are identified, then  $\beta'_i$ 's and  $\sigma_i^2$ 's are also identified conditional on  $\psi'_i$ 's. To derive the necessary and sufficient condition for identification of spatial parameters we partition  $\mathbf{H}(\boldsymbol{\theta}_0)$  as follows

$$\mathbf{H}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathbf{H}_{11} & \tilde{\mathbf{H}}_{12} \\ \cdot & \tilde{\mathbf{H}}_{22} \end{pmatrix},$$

where  $\tilde{\mathbf{H}}_{12} = (\mathbf{H}_{12}, \mathbf{H}_{13})$  is an  $N \times (Nk + N)$  matrix, and since  $\mathbf{H}_{23} = \mathbf{H}_{32} = \mathbf{0}$ , then  $\tilde{\mathbf{H}}_{22} = \text{diag}(\mathbf{H}_{22}, \mathbf{H}_{33})$ , which is an  $(Nk + N) \times (Nk + N)$  matrix. The asymptotic covariance matrix

of  $\sqrt{T}\hat{\boldsymbol{\psi}}_T$  is given by  $\tilde{\mathbf{H}}_{11.2}^{-1}$ , where

$$\tilde{\mathbf{H}}_{11.2} = \mathbf{H}_{11} - \tilde{\mathbf{H}}_{12}\tilde{\mathbf{H}}_{22}^{-1}\tilde{\mathbf{H}}_{21} = \mathbf{H}_{11} - \mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21} - \mathbf{H}_{13}\mathbf{H}_{33}^{-1}\mathbf{H}_{31}. \quad (48)$$

But

$$\mathbf{H}_{11} = (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} [\sigma_i^{-2} \mathbf{g}'_{0i} (\mathbf{B}\boldsymbol{\Sigma}_{xx}\mathbf{B}' + \boldsymbol{\Sigma}_\varepsilon) \mathbf{g}_{0i}, i = 1, 2, \dots, N],$$

where the  $(r, s)$  element of  $\mathbf{B}\boldsymbol{\Sigma}_{xx}\mathbf{B}'$  is given by  $\boldsymbol{\beta}'_r \boldsymbol{\Sigma}_{rs} \boldsymbol{\beta}_s$  as before, and  $\mathbf{g}'_{0i} \boldsymbol{\Sigma}_\varepsilon \mathbf{g}_{0i} = \sum_{s=1}^N \sigma_s^2 g_{0,is}^2$ . Hence

$$\begin{aligned} \mathbf{H}_{11} &= (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} \left[ \sum_{s=1}^N (\sigma_s^2 / \sigma_i^2) g_{0,is}^2, i = 1, 2, \dots, N \right] \\ &+ \text{diag} \left[ \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \boldsymbol{\beta}'_r \boldsymbol{\Sigma}_{rs} \boldsymbol{\beta}_s, i = 1, 2, \dots, N \right]. \end{aligned}$$

Similarly,

$$\begin{aligned} &\mathbf{H}_{12}\mathbf{H}_{22}^{-1}\mathbf{H}_{21} \\ &= \text{diag} \left[ \sigma_i^{-2} \mathbf{w}'_i (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} \mathbf{B} E(\mathbf{x}_i \mathbf{x}'_i) \boldsymbol{\Sigma}_{ii}^{-1} E(\mathbf{x}_i \mathbf{x}'_i) \mathbf{B}' (\mathbf{I}_N - \mathbf{W}' \boldsymbol{\Psi}_0)^{-1} \mathbf{w}_i, i = 1, 2, \dots, N \right] \\ &= \text{diag} \left[ \sigma_i^{-2} \mathbf{g}'_{0i} (\boldsymbol{\Sigma}_{i1} \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_{i2} \boldsymbol{\beta}_2, \dots, \boldsymbol{\Sigma}_{iN} \boldsymbol{\beta}_N)' \boldsymbol{\Sigma}_{ii}^{-1} (\boldsymbol{\Sigma}_{i1} \boldsymbol{\beta}_1, \boldsymbol{\Sigma}_{i2} \boldsymbol{\beta}_2, \dots, \boldsymbol{\Sigma}_{iN} \boldsymbol{\beta}_N) \mathbf{g}_{0i}, i = 1, 2, \dots, N \right] \\ &= \text{diag} \left[ \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \boldsymbol{\beta}'_r \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is} \boldsymbol{\beta}_s, i = 1, 2, \dots, N \right], \end{aligned}$$

and

$$\mathbf{H}_{13}\mathbf{H}_{33}^{-1}\mathbf{H}_{31} = 2 \text{diag} [g_{0,ii}^2, i = 1, 2, \dots, N].$$

Using the above results in (48), now yields,

$$\begin{aligned} \tilde{\mathbf{H}}_{11.2} &= (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} \left[ -g_{0,ii}^2 + \sum_{s=1, s \neq i}^N (\sigma_s^2 / \sigma_i^2) g_{0,is}^2, i = 1, 2, \dots, N \right] \\ &+ \text{diag} \left[ \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \boldsymbol{\beta}'_r (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \boldsymbol{\beta}_s, i = 1, 2, \dots, N \right]. \end{aligned} \quad (49)$$

The necessary and sufficient condition for identification of  $\psi'_i$ s in the general model, (42), is now give by  $\text{Rank}(\tilde{\mathbf{H}}_{11.2}) = N$ .

In the special case where the regressors are cross-sectionally uncorrelated, namely when  $\boldsymbol{\Sigma}_{rs} = \mathbf{0}$ , if  $r \neq s$ , the third term in the above result vanishes and we have

$$\tilde{\mathbf{H}}_{11.2} = (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} \left[ -g_{0,ii}^2 + \sum_{s=1, s \neq i}^N (\sigma_s^2 / \sigma_i^2) g_{0,is}^2, i = 1, 2, \dots, N \right],$$

which does not depend on  $\boldsymbol{\beta}'_i$ s or the exogenous regressors.

In the case where  $N$  is large, for identification of  $\psi$  we require the following two conditions

$$\begin{aligned}\lambda_{\min}(\tilde{\mathbf{H}}_{11.2}) &> \epsilon > 0, \text{ for all } N, \\ \lambda_{\max}(\tilde{\mathbf{H}}_{11.2}) &< K < \infty, \text{ for all } N.\end{aligned}$$

The first condition is the usual rank condition when  $N$  is finite. Regarding the second condition we first note that since under the identification condition  $\tilde{\mathbf{H}}_{11.2}$  is a symmetric positive definite matrix then  $\lambda_{\max}(\tilde{\mathbf{H}}_{11.2}) = \left| \lambda_{\max}(\tilde{\mathbf{H}}_{11.2}) \right| \leq \left\| \tilde{\mathbf{H}}_{11.2} \right\|_{\infty}$ . Also it is easily seen that

$$\begin{aligned}\left\| \tilde{\mathbf{H}}_{11.2} \right\|_{\infty} &\leq \left\| (\mathbf{G}_0 \odot \mathbf{G}'_0) \right\|_{\infty} + \sup_i \left\| -g_{0,ii}^2 + \sum_{s=1, s \neq i}^N (\sigma_s^2 / \sigma_i^2) g_{0,is}^2 \right\|_{\infty} \\ &+ \sup_i \left\| \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \beta'_r (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \beta_s \right\|_{\infty}.\end{aligned}\quad (50)$$

Recall that  $\|\mathbf{G}_0\|_1 < K$  and  $\|\mathbf{G}_0\|_{\infty} < K$ , are bounded in  $N$  when Assumption 1 and (7) hold. See Lemma 2 in Appendix A, from which it also follows that  $\|\mathbf{G}_0 \odot \mathbf{G}'_0\|_{\infty} < K$ . Furthermore,

$$\left\| -g_{0,ii}^2 + \sum_{s=1, s \neq i}^N (\sigma_s^2 / \sigma_i^2) g_{0,is}^2 \right\|_{\infty} \leq \frac{\sup_i (\sigma_i^2)}{\inf_i (\sigma_i^2)} \sum_{s=1}^N g_{0,is}^2 < K,$$

since the maximum absolute row (column) norm of matrix  $\mathbf{G}_0$  is bounded, and  $0 < \kappa_1 \leq \sigma_i^2 \leq \kappa_2 < \infty$  by assumption. Consider now the last term of (50) and note that

$$\begin{aligned}&\left\| \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \beta'_r (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \beta_s \right\|_{\infty} \\ &\leq \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N |g_{0,is}| |g_{0,ir}| \left\| \beta'_r (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \beta_s \right\|_{\infty} \\ &\leq \sigma_i^{-2} \sup_{r,s} \left\| \beta'_r (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \beta_s \right\|_{\infty} \sum_{r=1}^N \sum_{s=1}^N |g_{0,is}| |g_{0,ir}| \\ &\leq \sigma_i^{-2} \sup_r \|\beta_r\|_1 \sup_s \|\beta_s\|_{\infty} \sup_{r,s} \left\| (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \right\|_{\infty} \left( \sum_{s=1}^N |g_{0,is}| \right)^2.\end{aligned}$$

Under Assumptions 6, 7 and 8,  $\inf \sigma_i^2 > 0$ ,  $\sup_r \|\beta_r\|_1$ ,  $\sup_s \|\beta_s\|_{\infty}$ ,  $\boldsymbol{\Sigma}_{rs}$  and  $\boldsymbol{\Sigma}_{ii}^{-1}$  exist and are finite. Also  $\sup_i \sum_{s=1}^N |g_{0,is}| = \|\mathbf{G}_0\|_{\infty}$ , which is bounded under our assumptions. Hence it follows that

$$\sup_i \left\| \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \beta'_r (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \beta_s \right\|_{\infty} < K.$$

Therefore, overall we have  $\left| \lambda_{\max}(\tilde{\mathbf{H}}_{11.2}) \right| = \lambda_{\max}(\tilde{\mathbf{H}}_{11.2}) \leq \left\| \tilde{\mathbf{H}}_{11.2} \right\|_{\infty} < K$ , for all  $N$ . Com-

binning this result together with the identification condition we have

$$\epsilon < \lambda_{\min} \left( \tilde{\mathbf{H}}_{11 \cdot 2} \right) < \lambda_{\max} \left( \tilde{\mathbf{H}}_{11 \cdot 2} \right) < K,$$

for some small positive,  $\epsilon$ , and a finite (possibly large) positive constant,  $K$ . Inverting the above inequality also yields<sup>4</sup>

$$\frac{1}{K} \leq \lambda_{\min} \left( \tilde{\mathbf{H}}_{11 \cdot 2}^{-1} \right) \leq \lambda_{\max} \left( \tilde{\mathbf{H}}_{11 \cdot 2}^{-1} \right) \leq \frac{1}{\epsilon},$$

which establishes that  $\tilde{\mathbf{H}}_{11 \cdot 2}^{-1}$  exists and is bounded even for  $N$  large.

The main result of this section can now be summarised in the following proposition.

**Proposition 3** *Consider the heterogeneous spatial autoregressive (HSAR) model given by (42) and suppose that: (a) Assumptions 1, 4, 5, and 6, 7, and 8 hold, (b) the invertibility condition (7) is met, (c)  $\lambda_{\min} \left( \tilde{\mathbf{H}}_{11 \cdot 2} \right) > \epsilon > 0$ , for all  $N$ , where  $\tilde{\mathbf{H}}_{11 \cdot 2}$  is the  $N \times N$  matrix*

$$\begin{aligned} \tilde{\mathbf{H}}_{11 \cdot 2} = & (\mathbf{G}_0 \odot \mathbf{G}'_0) + \text{diag} \left[ -g_{0,ii}^2 + \sum_{s=1, s \neq i}^N (\sigma_s^2 / \sigma_i^2) g_{0,is}^2, i = 1, 2, \dots, N \right] \\ & + \text{diag} \left[ \sigma_i^{-2} \sum_{r=1}^N \sum_{s=1}^N g_{0,is} g_{0,ir} \beta'_r (\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}) \boldsymbol{\beta}_s, i = 1, 2, \dots, N \right], \end{aligned}$$

$\mathbf{G}_0 = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}_0 \mathbf{W})^{-1} = (g_{0,ij})$ ,  $\boldsymbol{\Psi}_0 = \text{diag}(\boldsymbol{\psi}_0)$ ,  $\boldsymbol{\psi}_0 = (\psi_{10}, \psi_{20}, \dots, \psi_{N0})'$ , and  $\mathbf{W}$  is the spatial weight matrix, and (d)  $\varepsilon_{it} \sim \text{IIDN}(0, \sigma_{0i}^2)$ . Then the maximum likelihood estimator of  $\boldsymbol{\psi}_0$ , denoted by  $\hat{\boldsymbol{\psi}}_T$  and computed by (26), has the following asymptotic distribution as  $T \rightarrow \infty$ ,

$$\sqrt{T} \left( \hat{\boldsymbol{\psi}}_T - \boldsymbol{\psi}_0 \right) \rightarrow_d N \left[ 0, \text{AsyVar}(\hat{\boldsymbol{\psi}}_T) \right],$$

where

$$\text{AsyVar}(\sqrt{T} \hat{\boldsymbol{\psi}}_T) = \left[ \tilde{\mathbf{H}}_{11 \cdot 2} \right]^{-1}.$$

The asymptotic results hold for any  $N$ .

**Remark 6** *The results of the above theorem readily extend to spatial models with non-Gaussian errors, but as in the case of pure spatial models the estimates of the variance matrix of the QML estimators must be based on the sandwich formula*

$$\widehat{\text{Var}} \left( \hat{\boldsymbol{\theta}}_T \right) = \mathbf{H}_T^{-1} \left( \hat{\boldsymbol{\theta}}_T \right) \mathbf{J}_T \left( \hat{\boldsymbol{\theta}}_T \right) \mathbf{H}_T^{-1} \left( \hat{\boldsymbol{\theta}}_T \right),$$

where  $\mathbf{H}_T \left( \hat{\boldsymbol{\theta}}_T \right)$  and  $\mathbf{J}_T \left( \hat{\boldsymbol{\theta}}_T \right)$  are given in Appendix C.

---

<sup>4</sup>Recall that for any positive definite matrix  $\mathbf{A}$ , we have  $\lambda_{\max}(\mathbf{A}) = \mathbf{1} / \lambda_{\min}(\mathbf{A}^{-1})$ .

## 7 Monte Carlo study

We investigate the small sample properties of the proposed QML estimator through a Monte Carlo simulation study. We consider the following data generating process

$$y_{it} = a_i + \psi_i \sum_{j=1}^N w_{ij} y_{jt} + \beta_i x_{it} + \varepsilon_{it}, \quad i = 1, 2, \dots, N; \quad t = 1, 2, \dots, T. \quad (51)$$

We include one exogenous regressor,  $x_{it}$ , with coefficient  $\beta_i$  in each regression.

Stacking by individual units, (51) becomes

$$\mathbf{y}_t = \mathbf{a} + \mathbf{\Psi} \mathbf{W} \mathbf{y}_t + \mathbf{B} \mathbf{x}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, T, \quad (52)$$

where  $\mathbf{a} = (a_1, a_2, \dots, a_N)'$ ,  $\mathbf{\Psi} = \text{diag}(\boldsymbol{\psi})$  and  $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_N)'$ ,  $\mathbf{W} = (w_{ij})$ ,  $i, j = 1, 2, \dots, N$ ,  $\mathbf{B} = \text{diag}(\boldsymbol{\beta})$ , where  $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_N)'$ ,  $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{Nt})'$ , and  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ . The unknown parameters are summarised in the vector  $\boldsymbol{\theta} = (\mathbf{a}', \boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^2)'$ ,  $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$ . In total there are  $4N$  unknown parameters.

We generate  $x_{it}$  as

$$x_{it} = \phi_i \mathbf{w}'_i \mathbf{x}_t + v_{it}, \quad (53)$$

or in matrix form

$$\mathbf{x}_t = (\mathbf{I}_N - \mathbf{\Phi} \mathbf{W})^{-1} \mathbf{v}_t,$$

where  $\mathbf{x}_t = (x_{1t}, x_{2t}, \dots, x_{Nt})'$ ,  $\mathbf{\Phi} = \text{diag}(\phi_1, \phi_2, \dots, \phi_N)$ , and  $\mathbf{v}_t = (v_{1t}, v_{2t}, \dots, v_{Nt})'$ , with  $v_{it} \sim \text{IIDN}(0, \sigma_v^2)$ . We set  $\phi_i = 0.5$  (representing a moderate degree of spatial dependence), and

$$\sigma_v^2 = \frac{N}{\text{tr} \left[ (\mathbf{I}_N - \mathbf{\Phi} \mathbf{W})^{-1} (\mathbf{I}_N - \mathbf{\Phi} \mathbf{W})'^{-1} \right]}.$$

This ensures that

$$N^{-1} \sum_{i=1}^N \text{Var}(x_{it}) = 1.$$

For  $\mathbf{W} = (w_{ij})$ ,  $i, j = 1, 2, \dots, N$ , we use the 4 connection spatial matrix below.

We consider both Gaussian and non-Gaussian errors. Specifically we consider the following two error generating processes

$$\varepsilon_{it}/\sigma_{i0} \sim \text{IIDN}(0, 1),$$

and

$$\varepsilon_{it}/\sigma_{i0} \sim \text{IID} [\chi^2(2) - 2] / 2,$$

for  $i = 1, 2, \dots, N$ , and  $t = 1, 2, \dots, T$ , where  $\chi^2(2)$  is a chi-square variate with 2 degrees of freedom.  $\sigma_{i0}^2$  are generated as independent draws from  $\chi^2(2)/4 + 0.50$ , for  $i = 1, 2, \dots, N$ , and kept fixed across the replications.

For the weight matrix,  $\mathbf{W} = (w_{ij})$ , we first use contiguity criteria to generate the non-normalized weights,  $w_{ij}^o$ , then row normalise the resultant weight matrices to obtain  $w_{ij}$ . More specifically, we consider  $\mathbf{W}$  matrices with 2, 4 and 10 connections and generate  $w_{ij}^o$ , for  $i, j = 1, 2, \dots, N$ , as



2 connections:  $w_{i,j-1}^o = w_{i,j+1}^o = 1$ , and zero otherwise,

4 connections:  $w_{i,j-1}^o = w_{i,j-2}^o = w_{i,j+1}^o = w_{i,j+2}^o = 1$ , and zero otherwise,

10 connections:  $w_{i,j-1}^o = \dots = w_{i,j-5}^o = w_{i,j+1}^o = \dots = w_{i,j+5}^o = 1$ , and zero otherwise.

Since by construction  $\|\mathbf{W}\|_\infty = 1$ , then condition (7) is satisfied if  $\sup_i |\psi_i| < 1$ , and ensures that  $\mathbf{I}_N - \Psi\mathbf{W}$  is invertible. Then,

$$\mathbf{y}_t = (\mathbf{I}_N - \Psi\mathbf{W})^{-1} (\mathbf{a} + \mathbf{B}\mathbf{x}_t + \boldsymbol{\varepsilon}_t); \quad t = 1, 2, \dots, T.$$

We consider two main experiments: **Experiment A**, where  $a_{i0} \sim IIDN(1, 1)$ ,  $\beta_{i0}$  are set to zero and  $\psi_{i0} \sim IIDU(0, 0.8)$ , for  $i = 1, 2, \dots, N$ , and **Experiment B**, where  $a_{i0} \sim IIDN(1, 1)$ ,  $\beta_{i0} \sim IIDU(0, 1)$  and  $\psi_{i0} \sim IIDU(0, 0.8)$ , for  $i = 1, 2, \dots, N$ .

We consider the following  $N, T$  combinations:  $N = 25, 50, 75, 100$  and  $T = 25, 50, 100, 200$ , and use  $R = 2000$  replications for each experiment. Across the replications,  $\boldsymbol{\theta}_0$ , and the weight matrix,  $\mathbf{W}$ , are kept fixed, whilst the errors and the regressors,  $\varepsilon_{it}$  and  $x_{it}$  (and hence  $y_{it}$ ), are re-generated randomly in each replication. Note that, as  $N$  increases, supplementary units are added to the original vector  $\boldsymbol{\theta}_0$  generated initially for  $N = 25$ . Due to the problem of simultaneity, the degree of time variation in  $y_{it}^*$  for each unit  $i$  depends on the choice of  $\mathbf{W}$  as well as the number of cross section units,  $N$ . Naturally, this is reflected in the performance of the  $\psi_i$  estimators and the tests based on them.

We compute both bias and RMSE of the QML estimators for individual cross section units, as well as their averages across all  $N$  cross section units. In addition, we report empirical sizes based on the individual spatial autoregressive parameter estimates and power functions for three units whose distinct true spatial autoregressive parameters,  $\psi_{i0}$ , are selected to be low, medium and large in magnitude. The experiments are carried out for spatial weight matrices,  $\mathbf{W}$ , with two, four and ten connections and the spatial parameter estimates are denoted by  $\hat{\psi}_{i(R)} = R^{-1} \sum_{r=1}^R \hat{\psi}_{i,r}$ , where  $\hat{\psi}_{i,r}$  refers to the QML estimator of  $\psi_i$  in the  $r^{th}$  replication.

## 7.1 Bias and RMSE results

The QML estimators are computed using the likelihood function (46). We start with the basic model in Experiment A that assumes a first order heterogeneous spatial autoregressive model without any exogenous variables and with normally distributed errors. Table A1 reports the average bias and RMSE of the mean  $\hat{\psi}_{i(R)}$  estimates, averaged across replications and over all cross section units. For each spatial weight matrix used ( $\mathbf{W}$  with two, four or ten connections) the average biases lie in the vicinity of zero while their corresponding RMSEs decline with  $T$  for a given  $N$ , as to be expected. Table A2 displays bias and RMSE results for the spatial autoregressive parameters,  $\hat{\psi}_{i(R)}$ , of individual units arranged with respect to their true values in an ascending order, from lowest to the highest, for clarity of exposition. These are shown as  $N$  increases from 25 to 100. Due to the large number of spatial parameter estimates for the cross section sizes  $N$  considered, some of these have been excluded from Table A2 and are available upon request from the authors. Also to save space, we are only reporting results based on the spatial weight matrix,  $\mathbf{W}$ , with four connections. Results for other choices of spatial weight matrices are available upon request.

Overall, it is evident that the biases of the individual spatial autoregressive parameters are close to zero even for small values of  $T$ , and irrespective of the magnitude of the spatial

parameter. As the theory suggests, the quality of the individual spatial parameter estimates are not affected by the size of  $N$ , but improve with the time dimension. The reported RMSEs decline with  $T$  for all values of  $N$ .

Similar results can be obtained even if the errors are non-Gaussian. Tables A4 and A5 give bias and RMSE of the QML estimators of  $\psi_i$  when the errors are generated as iid  $\chi^2(2)$  random variables. These tables follow the same lay out as those of Tables A1 and A2, and differ from them only in the way the errors are generated. A comparison of the results across these two sets of tables show that the QML estimator is reasonably robust to non-normal errors, and the rate at which RMSEs decline with  $T$  when the errors are generated as iid  $\chi^2(2)$  tends to be similar to when the errors are generated as iid  $N(0, 1)$ .

The bias and RMSE results for Experiment B with normally distributed errors are summarised in Table B1 for the average estimates across the spatial units, and in Table B2 for the individual spatial estimates (as for Experiment A). It is clear that adding exogenous regressors to the spatial model does not alter the main conclusions and, if anything, their inclusion can marginally improve the precision with which individual spatial parameters are estimated. To save space results for the non-Gaussian errors scenario for Experiment B are not shown but conclusions are qualitatively analogous to those for Experiment A. These results are available upon request from the authors.

## 7.2 Size and power results

The empirical size of the tests based on the individual QML estimates of  $\psi_i$  for Experiment A, in the case where the errors are iid Normal are summarized in Table A3. We present results based on the sample standard and sandwich covariance matrix formulae given by (40) and (41), respectively. As can be seen, in general the tests are correctly sized at 5 per cent for  $T$  relatively large, although for small values of  $T$  there are size distortions when both the standard and sandwich formulae are used. The results based on the standard and sandwich formulae both converge to 5 per cent as  $T$  increases, irrespective of the value of  $N$ .

The empirical power function of the tests are displayed in Figures A1-A3 for three cross section units with a low, medium and high spatial parameter. More precisely, Figure A1 shows a number of power functions under different alternative hypotheses  $\psi_i = \psi_{i0} + \delta$ , with  $\psi_{i0} = 0.3374$  and  $\delta = -0.8, -0.791, \dots, 0.791, 0.8$ , or until the parameter space boundaries of  $-1$  and  $1$  are reached for  $i = 1, 2, \dots, N$ . Similarly, Figures A2 and A3 depict the same functions when  $\psi_{i0}$  takes a medium value ( $\psi_{i0} = 0.5059$ ) and a high value ( $\psi_{i0} = 0.7676$ ), respectively. Note that in order to save space, the depicted power functions are based on tests where the sandwich formula is used.<sup>5</sup> Overall, for a specific cross section unit,  $i$ , the empirical power functions are similar for all  $N$  but improve with  $T$ . Furthermore, perhaps not surprisingly, empirical power functions become more and more asymmetrical as  $\psi_i$ 's move closer and closer to the boundary value of 1.

Empirical size estimates for Experiment A when errors are non-normal are summarized in Table A6. There are some size distortions when  $T$  is small irrespective of whether the standard or sandwich formulae are used. But as  $T$  increases the size distortion of the tests based on both formulae tend to zero. See, for example, the size estimates for  $N = 50$  and  $T = 200$ . The estimates of the power functions, computed using the sandwich formula, are provided in Figures A4-A6, and are comparable to those shown above for the Gaussian case.

<sup>5</sup>Equivalent power functions based on the standard covariance matrix estimator are available upon request.

Size and power results for Experiment B with Gaussian errors are summarised in Tables B3-B4, and in Figures B1-B6. The results are comparable to those obtained for Experiment A, and there are some size distortions when  $T$  is small, but the size estimates converge to their nominal value of 5 per cent as  $T$  increases. For completeness, the associated power functions are displayed in Figures B1-B3, which tend to be sharper than those shown in Figures A1-A3, for the experiments without exogenous regressors. Adding an exogenous regressor to the spatial model seems to result in more precise estimates of the spatial coefficients.

Turning to size and power of tests based on the regression coefficients,  $\beta_i$ , the size results are summarized in Table B4. As can be seen, empirical sizes using the standard variance formula are very close to the nominal value of 5 per cent for all  $N$  and  $T$  combinations, although there are some size distortions when  $T$  is small. On the other hand, using the sandwich formula produces larger size distortions when  $T$  is small. The associated power functions computed using the sandwich formula are displayed in Figures B4-B6 for a low value of  $\beta_i$  ( $\beta_{i0} = 0.0344$ ), a medium ( $\beta_{i0} = 0.4898$ ), and for a high value ( $\beta_{i0} = 0.9649$ ), respectively. Again the empirical power functions are similar across  $N$  and improve with  $T$ .

## 8 Conclusion

Standard spatial econometric models assume a single parameter to characterise the intensity or strength of spatial dependence. In the case of pure cross section models or panel data models with a short  $T$ , such restrictive parameter specification might be inevitable. However, in a data rich environment where both the time ( $T$ ) and cross section ( $N$ ) dimensions are large, this assumption can be relaxed. This paper investigates a spatial autoregressive panel data model with fully heterogeneous spatial parameters. The asymptotic properties of the quasi maximum likelihood estimator are analysed assuming a sparse spatial structure with each individual unit having at least one connection. Conditions under which the QML estimator of spatial parameters,  $\psi_i$ , are consistent and asymptotically normal are derived. It is also shown that under certain bound conditions on  $\mathbf{W}(\mathbf{I}_N - \Psi\mathbf{W})^{-1}$  the asymptotic properties of the individual estimates are not affected by the size of  $N$ . Monte Carlo simulation results provided are supportive of the theoretical findings. Extensions of this model specification that incorporate richer temporal and spatial dynamics and that accommodate negative as well as positive connections are interesting avenues for future research. The methods developed in the paper can also be applied to hierarchical panel data models where spatial parameters are assumed to be the same within regions (groups) but allowed to differ across regions or groups.

## Appendix A Technical lemmas

**Lemma 1** *Let  $\mathbf{W}$  comply to Assumption 1. Then matrix  $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}$  is non-singular if*

$$\sup_i |\psi_i| < \max \{1/\|\mathbf{W}\|_1, 1/\|\mathbf{W}\|_\infty\}, \quad i = 1, 2, \dots, N.$$

**Proof.** Let  $\varrho(\boldsymbol{\Psi}\mathbf{W})$  be the spectral radius of matrix  $\boldsymbol{\Psi}\mathbf{W}$ . Non-singularity of  $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}$  is ensured if

$$\varrho(\boldsymbol{\Psi}\mathbf{W}) < 1. \quad (\text{A.1})$$

However, since for any matrix norm  $\|\mathbf{A}\|$ ,  $\varrho(\mathbf{A}) \leq \|\mathbf{A}\|$ , then using the maximum column sum matrix norm we have

$$\varrho(\boldsymbol{\Psi}\mathbf{W}) \leq \|\boldsymbol{\Psi}\mathbf{W}\|_1 \leq \|\boldsymbol{\Psi}\|_1 \|\mathbf{W}\|_1 = \sup_i |\psi_i| \|\mathbf{W}\|_1, \quad (\text{A.2})$$

and from (A.1) we have

$$\sup_i |\psi_i| \|\mathbf{W}\|_1 < 1.$$

Similarly, using a maximum row sum matrix norm we have

$$\sup_i |\psi_i| \|\mathbf{W}\|_\infty < 1,$$

where we have used the result  $\|\boldsymbol{\Psi}\|_1 = \|\boldsymbol{\Psi}\|_\infty = \sup_i |\psi_i|$ . Therefore, matrix  $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}$  is invertible if

$$\sup_i |\psi_i| < \max \{1/\|\mathbf{W}\|_1, 1/\|\mathbf{W}\|_\infty\}, \quad i = 1, 2, \dots, N. \quad (\text{A.3})$$

■

**Lemma 2** *Let  $\mathbf{G}(\boldsymbol{\psi}) = \mathbf{W}(\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W})^{-1}$ , and suppose that Assumption 1 holds. Then*

$$\|\mathbf{G}(\boldsymbol{\psi})\|_1 < K \text{ and } \|\mathbf{G}(\boldsymbol{\psi})\|_\infty < K, \quad (\text{A.4})$$

*for all values of  $\boldsymbol{\psi} = (\psi_1, \psi_2, \dots, \psi_N)'$  that satisfy condition (7).*

**Proof.** Under condition (7), we have

$$\mathbf{G}(\boldsymbol{\psi}) = \mathbf{W} + \mathbf{W}\boldsymbol{\Psi}\mathbf{W} + \mathbf{W}\boldsymbol{\Psi}\mathbf{W}\boldsymbol{\Psi}\mathbf{W} + \dots,$$

and

$$\|\mathbf{G}(\boldsymbol{\psi})\|_1 \leq \|\mathbf{W}\|_1 + \|\mathbf{W}\|_1^2 \|\boldsymbol{\Psi}\|_1 + \|\mathbf{W}\|_1^3 \|\boldsymbol{\Psi}\|_1^2 + \dots$$

But  $\|\boldsymbol{\Psi}\|_1^s = [\sup_i |\psi_i|]^s$ , and under condition (7) we have  $\sup_i |\psi_i| \|\mathbf{W}\|_1 < 1$ . Hence

$$\|\mathbf{G}(\boldsymbol{\psi})\|_1 \leq \|\mathbf{W}\|_1 \left( \frac{1}{1 - \sup_i |\psi_i| \|\mathbf{W}\|_1} \right).$$

Similarly,

$$\|\mathbf{G}(\boldsymbol{\psi})\|_\infty = \|\mathbf{W}\|_\infty \left( \frac{1}{1 - \sup_i |\psi_i| \|\mathbf{W}\|_\infty} \right).$$

The boundedness of column and row matrix norms of  $\mathbf{G}(\boldsymbol{\psi})$  now follow since, under Assumption 1,  $\|\mathbf{W}\|_1$  and  $\|\mathbf{W}\|_\infty$  are bounded,  $1 - \sup_i |\psi_i| \|\mathbf{W}\|_1 > 0$ , and  $1 - \sup_i |\psi_i| \|\mathbf{W}\|_\infty > 0$ . ■

## Appendix B Derivatives of $Q_N(\boldsymbol{\varphi})$

We consider the first and second derivatives of function  $Q_N(\boldsymbol{\varphi})$ , as specified in (19) and repeated below for convenience:

$$\begin{aligned} Q_N(\boldsymbol{\varphi}) &= -\frac{1}{2} [\ln(1 - \delta) + \delta] - \frac{1}{N} \ln |\mathbf{I}_N - \mathbf{D}\mathbf{G}_0| - \frac{1}{N} (1 - \delta) \text{tr}(\mathbf{D}\mathbf{G}_0) \\ &\quad + \frac{1}{2N} (1 - \delta) \text{tr}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}^2) \geq 0 \end{aligned}$$

where  $\boldsymbol{\varphi} = (\mathbf{d}', \delta)'$ ,  $\delta = (\sigma^2 - \sigma_0^2) / \sigma^2 < 1$ , and  $\mathbf{d} = (d_1, d_2, \dots, d_N)'$ ,  $d_i = \psi_i - \psi_{i0}$ , for  $i = 1, 2, \dots, N$ .

### First derivatives

For the first derivatives, we have:

$$\frac{\partial Q_N(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} = \begin{pmatrix} \frac{\partial Q_N(\boldsymbol{\varphi})}{\partial \mathbf{d}} \\ \frac{\partial Q_N(\boldsymbol{\varphi})}{\partial \delta} \end{pmatrix}_{(N+1) \times 1}.$$

$$\begin{aligned} \frac{\partial Q_N(\boldsymbol{\varphi})}{\partial \mathbf{d}} &= \begin{pmatrix} \frac{\partial Q_N(\boldsymbol{\varphi})}{\partial d_i} \end{pmatrix} \\ &= \left( \frac{1}{N} \mathbf{g}'_{0i} (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} \mathbf{e}_i - \frac{1}{N} (1 - \delta) g_{0,ii} + \frac{1}{N} (1 - \delta) \mathbf{g}'_{0i} \mathbf{g}_{0i} d_i \right), \end{aligned}$$

where  $\mathbf{e}_i$  is the  $i^{\text{th}}$  column of  $\mathbf{E}_{ij}$  matrix of zeros bar element  $(i, j)$  which is equal to unity, and  $\mathbf{g}'_{0i}$  is the  $i^{\text{th}}$  row of  $\mathbf{G}_0$ .

$$\frac{\partial Q_N(\boldsymbol{\varphi})}{\partial \delta} = \frac{1}{2} \frac{\delta}{1 - \delta} + \frac{1}{N} \left[ \text{tr}(\mathbf{D}\mathbf{G}_0) - \frac{1}{2} \text{tr}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}^2) \right].$$

Overall, stacking over cross section units we get:

$$\frac{\partial Q_N(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}} = \begin{pmatrix} \frac{1}{N} \text{diag} \left[ \mathbf{G}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} \right] \boldsymbol{\tau}_N - \frac{1}{N} (1 - \delta) \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N + \frac{1}{N} (1 - \delta) \text{diag}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}) \boldsymbol{\tau}_N \\ \frac{1}{2} \frac{\delta}{1 - \delta} + \frac{1}{N} \left[ \text{tr}(\mathbf{D}\mathbf{G}_0) - \frac{1}{2} \text{tr}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}^2) \right] \end{pmatrix},$$

where  $\boldsymbol{\tau}_N$  is a  $N \times 1$  vector of ones.

## Second derivatives

For the second derivatives, we have:

$$\frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \boldsymbol{\varphi}' \partial \boldsymbol{\varphi}} = \begin{pmatrix} \frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \mathbf{d} \partial \mathbf{d}'} & \frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \mathbf{d} \partial \delta} \\ \frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \delta \partial \mathbf{d}'} & \frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \delta^2} \end{pmatrix}_{(N+1) \times (N+1)}.$$

First,

$$\frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \mathbf{d} \partial \mathbf{d}'} = \begin{pmatrix} \frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial d_i \partial d_j} \end{pmatrix},$$

where

$$\frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial d_i \partial d_j} = \begin{cases} \frac{1}{N} \mathbf{g}'_{0i} (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} \mathbf{E}_{ii} \mathbf{G}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} \mathbf{e}_i + \frac{1}{N} (1 - \delta) \mathbf{g}'_{0i} \mathbf{g}_{0i}, & \text{if } i = j \\ \frac{1}{N} \mathbf{g}'_{0i} (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} \mathbf{E}_{jj} \mathbf{G}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} \mathbf{e}_i, & \text{if } i \neq j \end{cases}.$$

Further,

$$\frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \mathbf{d} \partial \delta} = \begin{pmatrix} \frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial d_i \partial \delta} \end{pmatrix} = \begin{pmatrix} \frac{1}{N} g_{0,ii} - \frac{1}{N} \mathbf{g}'_{0i} \mathbf{g}_{0i} d_i \end{pmatrix},$$

and finally

$$\frac{\partial^2 Q_N(\boldsymbol{\varphi})}{\partial \delta^2} = \frac{1}{2} \frac{1}{(1 - \delta)^2}.$$

Overall, stacking over cross section units we get  $\partial^2 Q_N(\boldsymbol{\varphi}) / \partial \boldsymbol{\varphi}' \partial \boldsymbol{\varphi} = N^{-1} \boldsymbol{\Lambda}_N(\boldsymbol{\varphi})$ , where:

$$\boldsymbol{\Lambda}_N(\boldsymbol{\varphi}) = \begin{bmatrix} (\mathbf{A}_0 \odot \mathbf{A}'_0) + (1 - \delta) \text{diag}(\mathbf{G}_0 \mathbf{G}'_0) & \text{diag}(\mathbf{G}_0) \boldsymbol{\tau}_N - \text{diag}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}) \boldsymbol{\tau}_N \\ \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0) - \boldsymbol{\tau}'_N \text{diag}(\mathbf{G}_0 \mathbf{G}'_0 \mathbf{D}) & \frac{N}{2(1 - \delta)^2} \end{bmatrix},$$

where  $\boldsymbol{\tau}_N$  is an  $N \times 1$  vector of ones,  $\odot$  is the Hadamard product matrix operator, and  $\mathbf{A}_0 = \mathbf{G}_0 (\mathbf{I}_N - \mathbf{D}\mathbf{G}_0)^{-1} = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W})^{-1}$ .

## Appendix C Estimator of $AsyVar(\hat{\boldsymbol{\theta}}_T)$

### Derivatives of the log-likelihood function

The vector of maximum likelihood estimates,  $\hat{\boldsymbol{\theta}}_T$ , in Section 6 is obtained by maximising the log-likelihood function (46) which we reproduce for convenience here

$$\ell(\boldsymbol{\theta}) = -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^N \ln \sigma_i^2 + T \ln |\mathbf{I}_N - \boldsymbol{\Psi}\mathbf{W}| - \frac{1}{2} \sum_{i=1}^N \frac{(\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{\sigma_i^2}, \quad (\text{C.5})$$

where  $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_N, \sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)' = (\boldsymbol{\psi}', \boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \dots, \boldsymbol{\beta}'_N, \boldsymbol{\sigma}^2)'$ , with  $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)'$ .

## First derivatives

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta})}{\partial \psi_i} &= -T \operatorname{tr}[(\mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W})^{-1} \mathbf{E}_{ii} \mathbf{W}] + \frac{\mathbf{y}_i^{*'} (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{\sigma_i^2}, \text{ for } i = 1, 2, \dots, N, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i} &= \frac{\mathbf{X}_i' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{\sigma_i^2}, \text{ for } i = 1, 2, \dots, N, \\ \frac{\partial \ell(\boldsymbol{\theta})}{\partial \sigma_i^2} &= -\frac{T}{2\sigma_i^2} + \frac{1}{2\sigma_i^4} (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i), i = 1, 2, \dots, N.\end{aligned}$$

## Second derivatives

$$\begin{aligned}\mathbf{H}_T(\boldsymbol{\theta}) &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{pmatrix} \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \psi'} & \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \boldsymbol{\beta}'} & \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \sigma^{2'}} \\ \cdot & \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} & \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma^{2'}} \\ \cdot & \cdot & \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial (\sigma^2) \partial (\sigma^{2'})} \end{pmatrix}, \\ &= \begin{pmatrix} \mathbf{H}_{T,11} & \mathbf{H}_{T,12} & \mathbf{H}_{T,13} \\ \cdot & \mathbf{H}_{T,22} & \mathbf{H}_{T,23} \\ \cdot & \cdot & \mathbf{H}_{T,33} \end{pmatrix},\end{aligned}$$

such that

$$\mathbf{H}_{T,11} = \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \psi'} = \left[ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \psi_j} \right],$$

$$\frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \psi_j} = \begin{cases} \operatorname{tr}[(\mathbf{I} - \boldsymbol{\Psi} \mathbf{W})^{-1} \mathbf{E}_{jj} \mathbf{W} (\mathbf{I} - \boldsymbol{\Psi} \mathbf{W})^{-1} \mathbf{E}_{ii} \mathbf{W}] & \text{if } i \neq j \\ \operatorname{tr}[(\mathbf{I} - \boldsymbol{\Psi} \mathbf{W})^{-1} \mathbf{E}_{ii} \mathbf{W} (\mathbf{I} - \boldsymbol{\Psi} \mathbf{W})^{-1} \mathbf{E}_{ii} \mathbf{W}] + \frac{1}{\sigma_i^2} \frac{\mathbf{y}_i^{*'} \mathbf{y}_i^*}{T} & \text{if } i = j \end{cases},$$

$$\frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \psi_j} = \begin{cases} g_{ij} g_{ji} & \text{if } i \neq j \\ g_{ii}^2 + \frac{1}{\sigma_i^2} \frac{\mathbf{y}_i^{*'} \mathbf{y}_i^*}{T} & \text{if } i = j \end{cases},$$

where  $\mathbf{E}_{ij}$  is the  $N \times N$  matrix whose  $(i, j)$  element is 1 and zero elsewhere and  $\mathbf{G} = (g_{ij}) = \mathbf{W} (\mathbf{I}_N - \boldsymbol{\Psi} \mathbf{W})^{-1}$ . Further, with the  $(i, j)$  or  $i^{\text{th}}$  element of associated matrix or vector given in  $\{\}$ ,

$$\begin{aligned}
\mathbf{H}_{T,12} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \boldsymbol{\beta}'} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \boldsymbol{\beta}'_j} \right\} = \left\{ \frac{1}{\sigma_i^2} \frac{\mathbf{y}_i^{*'} \mathbf{X}_i}{T}, \text{ if } i = j, \text{ and } \mathbf{0}, \text{ if } i \neq j \right\}, \\
\mathbf{H}_{T,13} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \sigma^{2'}} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \sigma_j^2} \right\} = \left\{ \frac{1}{\sigma_i^4} \frac{\mathbf{y}_i^{*'} (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{T}, \text{ if } i = j, \text{ and } \mathbf{0}, \text{ if } i \neq j \right\}, \\
\mathbf{H}_{T,22} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}'_j} \right\} = \left\{ \frac{1}{\sigma_i^2} \frac{\mathbf{X}_i' \mathbf{X}_i}{T}, \text{ if } i = j, \text{ and } \mathbf{0}, \text{ if } i \neq j \right\}, \\
\mathbf{H}_{T,23} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma^{2'}} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \sigma_j^2} \right\} = \left\{ \frac{1}{\sigma_i^4} \frac{\mathbf{X}_i' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{T}, \text{ if } i = j, \text{ and } \mathbf{0}, \text{ if } i \neq j \right\} \\
\mathbf{H}_{T,33} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial (\sigma^2) \partial (\sigma^{2'})} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \sigma_i^2 \partial \sigma_j^2} \right\} \\
&= \left\{ -\frac{1}{2\sigma_i^4} + \frac{1}{\sigma_i^6} \frac{1}{T} (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i), \text{ if } i = j, \text{ and } \mathbf{0}, \text{ if } i \neq j \right\}
\end{aligned}$$

In the homoskedastic case,  $\sigma_i^2 = \sigma^2$ , for  $i = 1, 2, \dots, N$ , we have

$$\begin{aligned}
\mathbf{H}_{T,12} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \boldsymbol{\beta}'} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \boldsymbol{\beta}'_j} \right\} = \left\{ \frac{1}{\sigma^2} \frac{\mathbf{y}_i^{*'} \mathbf{X}_i}{T}, \text{ if } i = j, \text{ and } \mathbf{0}, \text{ if } i \neq j \right\}, \\
\mathbf{H}_{T,13} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi \partial \sigma^2} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \psi_i \partial \sigma^2} \right\} = \left\{ \frac{1}{\sigma^4} \frac{\mathbf{y}_i^{*'} (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{T} \right\}, \\
\mathbf{H}_{T,22} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}'_j} \right\} = \left\{ \frac{1}{\sigma^2} \frac{\mathbf{X}_i' \mathbf{X}_i}{T}, \text{ if } i = j, \text{ and } \mathbf{0}, \text{ if } i \neq j \right\}, \\
\mathbf{H}_{T,23} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \sigma^2} = \left\{ \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i \partial \sigma^2} \right\} = \left\{ \frac{1}{\sigma^4} \frac{\mathbf{X}_i' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)}{T} \right\} \\
\mathbf{H}_{T,33} &= \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial (\sigma^2)^2} = -\frac{N}{2\sigma^4} + \frac{1}{\sigma^6} \frac{1}{T} \sum_{i=1}^N (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i)' (\mathbf{y}_i - \psi_i \mathbf{y}_i^* - \mathbf{X}_i \boldsymbol{\beta}_i),
\end{aligned}$$

Finally, from the above results we obtain:

$$\mathbf{J}_T(\boldsymbol{\theta}) = \frac{1}{T} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \left( \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \text{ and } \mathbf{H}_T(\boldsymbol{\theta}) = \frac{-1}{T} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'},$$

from which the standard and sandwich covariance matrix estimators (40) and (41), are given by

$$\widehat{Var}(\hat{\boldsymbol{\theta}}_T) = \mathbf{H}_T^{-1}(\hat{\boldsymbol{\theta}}_T),$$

and

$$\widehat{Var}(\hat{\boldsymbol{\theta}}_T) = \mathbf{H}_T^{-1}(\hat{\boldsymbol{\theta}}_T) \mathbf{J}_T(\hat{\boldsymbol{\theta}}_T) \mathbf{H}_T^{-1}(\hat{\boldsymbol{\theta}}_T),$$

respectively.



## References

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Kluwer Academic: The Netherlands.
- Anselin, L. (2001). Spatial econometrics. In B. H. Baltagi (Ed.), *A Companion to Theoretical Econometrics*. Blackwell: Massachusetts.
- Anselin, L. and A. K. Bera (1998). Spatial dependence in linear regression models with an introduction to spatial econometrics. In A. Ullah and D. E. A. Giles (Eds.), *Handbook of Applied Economic Statistics*. Marcel Dekker: New York.
- Arbia, G. (2010). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer: Berlin Heidelberg New York.
- Baltagi, B. and D. Levin (1986). Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics* 68, 148–155.
- Baltagi, B., S. H. Song, B. C. Jung, and W. Kon (2007). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* 140, 5–51.
- Baltagi, B., S. H. Song, and W. Kon (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics* 117, 123–150.
- Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation*. Pion Ltd.: London.
- Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons: New York.
- Cressie, N. and C. K. Wikle (2011). *Statistics for Spatio-Temporal Data*. Wiley-Blackwell: New York.
- Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer: Berlin Heidelberg.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge University Press.
- Holly, S., M. H. Pesaran, and T. Yamagata (2010). A spatio-temporal model of house prices in the USA. *Journal of Econometrics* 158, 160–173.
- Kapoor, M., H. H. Kelejian, and I. R. Prucha (2007). Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97–130.
- Kelejian, H. H. and I. R. Prucha (1999). A generalised moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509–533.
- Kelejian, H. H. and I. R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53–67.
- Kelejian, H. 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. *Papers in Regional Science* 72, 297–312.

- Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899–1925.
- Lee, L.-F. and J. Yu (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154, 165–185.
- LeSage, J. and K. Pace (2010). Spatial econometrics. *Web Book of Regional Science Regional Research Institute* 50, 1014–1015.
- Lin, X. and L.-F. Lee (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157, 34–52.
- Ord, J. K. and A. Getis (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis* 27, 286–306.
- Upton, G. J. G. and B. Fingleton (1985). *Spatial Data Analysis by Example, Volume 1*. Wiley: Chichester.
- Whittle, P. (1954). On stationary processes on the plane. *Biometrika* 41, 434–449.
- Yu, J., R. M. de Jong, and L.-F. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both  $n$  and  $T$  are large. *Journal of Econometrics* 146, 118–134.

Table A1: Average bias and RMSE of spatial parameter estimates from HSAR(1) model with Gaussian errors over all  $N$  cross section units for different spatial weight matrices ( $\mathbf{W}$ ), and  $N$  and  $T$  combinations

$N$	25			50			75			100		
	$T$	25	50	200	25	50	100	200	25	50	100	200
$\mathbf{W}$ with 2 connections												
Average Bias												
$\bar{a}_R$	0.0658	0.0363	0.0185	0.0086	0.0422	0.0215	0.0116	0.0048	0.0400	0.0180	0.0096	0.0037
$\bar{\psi}_R$	-0.0298	-0.0157	-0.0080	-0.0037	-0.0231	-0.0126	-0.0060	-0.0028	-0.0227	-0.0100	-0.0053	-0.0021
Average RMSE												
$\bar{a}_R$	0.8414	0.6227	0.4494	0.3171	0.7629	0.5712	0.4119	0.2898	0.7518	0.5522	0.3964	0.2786
$\bar{\psi}_R$	0.3729	0.2765	0.2001	0.1413	0.3862	0.2943	0.2151	0.1537	0.3774	0.2814	0.2035	0.1440
$\mathbf{W}$ with 4 connections												
Average Bias												
$\bar{a}_R$	0.0132	0.0049	0.0024	-0.0002	0.0120	0.0040	0.0021	-0.0005	0.0123	0.0016	0.0006	-0.0007
$\bar{\psi}_R$	-0.0042	-0.0009	-0.0007	0.0002	-0.0033	-0.0007	-0.0002	0.0004	-0.0041	0.0003	0.0002	0.0004
Average RMSE												
$\bar{a}_R$	0.7398	0.5334	0.3785	0.2665	0.6996	0.5001	0.3560	0.2495	0.6897	0.4903	0.3468	0.2464
$\bar{\psi}_R$	0.3194	0.2308	0.1637	0.1149	0.3443	0.2471	0.1757	0.1232	0.3356	0.2393	0.1692	0.1199
$\mathbf{W}$ with 10 connections												
Average Bias												
$\bar{a}_R$	0.0155	0.0049	0.0013	-0.0009	0.0215	0.0061	0.0016	-0.0011	0.0241	0.0044	0.0003	-0.0011
$\bar{\psi}_R$	-0.0079	-0.0020	-0.0006	0.0003	-0.0096	-0.0026	-0.0005	0.0006	-0.0121	-0.0022	0.0000	0.0005
Average RMSE												
$\bar{a}_R$	0.8414	0.6212	0.4530	0.3223	0.8223	0.6081	0.4422	0.3139	0.8201	0.6018	0.4385	0.3142
$\bar{\psi}_R$	0.3747	0.2775	0.2030	0.1443	0.4130	0.3060	0.2226	0.1583	0.4078	0.3002	0.2189	0.1566

Notes: True parameter values are generated as  $\psi_{i0} \sim IIDU(0, 0.8)$  and  $a_{i0} \sim IIDN(1, 1)$ , for  $i = 1, 2, \dots, N$ . Gaussian errors are generated as  $\varepsilon_{it}/\sigma_{i0} \sim IIDN(0, 1)$ , with  $\sigma_{i0}^2 \sim IID[\chi^2(2)/4 + 0.5]$ , for  $i = 1, 2, \dots, N$ . Average Bias and Average RMSE for  $\bar{\psi}_R$  are calculated as  $(NR)^{-1} \sum_{i=1}^N \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})$  and  $N^{-1} \sum_{i=1}^N \left[ \sqrt{R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})^2} \right]$ , respectively. Average Bias and Average RMSE for  $\bar{a}_R$  are computed similarly. The number of replications is set to  $R = 2000$ .

Table A2: Bias and RMSE for individual spatial parameter estimates from the HSAR(1) model with Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections for different  $N$  and  $T$  combinations

$T$	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$N = 25$								
0.0255	-0.0145	0.6162	-0.0019	0.4841	-0.0173	0.3509	-0.0025	0.2413
0.0777	0.0173	0.4396	0.0074	0.3259	0.0018	0.2199	0.0027	0.1525
0.1109	0.0316	0.3023	0.0108	0.2072	0.0037	0.1435	0.0011	0.1017
0.1261	0.0070	0.2829	-0.0017	0.1979	0.0027	0.1362	0.0024	0.0909
0.1948	0.0003	0.2264	-0.0055	0.1497	-0.0028	0.1022	0.0029	0.0732
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3883	-0.0104	0.4211	0.0025	0.3019	0.0053	0.2117	-0.0008	0.1564
0.3987	0.0130	0.2435	0.0036	0.1684	0.0073	0.1198	-0.0002	0.0818
0.4375	-0.0006	0.3704	0.0049	0.2676	-0.0020	0.1799	0.0013	0.1279
0.4682	0.0197	0.3650	0.0144	0.2622	0.0074	0.1879	0.0015	0.1317
0.5059	0.0075	0.2297	0.0095	0.1624	-0.0029	0.1092	0.0025	0.0769
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.6726	0.0116	0.2279	0.0126	0.1669	0.0045	0.1162	0.0034	0.0848
0.6793	-0.0061	0.2387	0.0030	0.1794	-0.0010	0.1300	-0.0018	0.0912
0.7127	-0.1494	0.5234	-0.0648	0.3715	-0.0342	0.2745	-0.0134	0.2009
0.7246	-0.0265	0.2694	-0.0143	0.2043	-0.0002	0.1461	-0.0034	0.1076
0.7676	-0.0107	0.2025	-0.0008	0.1497	-0.0012	0.1139	0.0021	0.0804
$N = 50$								
0.0255	-0.0081	0.6162	-0.0210	0.4810	-0.0157	0.3587	0.0004	0.2395
0.0432	0.0269	0.3346	0.0153	0.2334	0.0036	0.1601	0.0074	0.1117
0.0608	-0.0081	0.6428	0.0093	0.4947	0.0012	0.3598	0.0020	0.2594
0.0625	0.0182	0.2793	0.0133	0.1925	0.0017	0.1329	0.0011	0.0945
0.0777	0.0094	0.4382	0.0163	0.3168	0.0126	0.2219	0.0032	0.1493
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3451	0.0037	0.4101	-0.0014	0.2856	0.0052	0.1990	-0.0002	0.1428
0.3510	-0.0003	0.3537	0.0010	0.2423	-0.0067	0.1676	0.0012	0.1203
0.3565	0.0041	0.2387	0.0087	0.1664	0.0027	0.1132	0.0025	0.0789
0.3755	0.0278	0.3557	0.0083	0.2505	0.0044	0.1734	-0.0019	0.1203
0.3786	-0.0122	0.4528	0.0021	0.3305	-0.0036	0.2302	-0.0001	0.1622
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7127	-0.1012	0.4838	-0.0641	0.3625	-0.0398	0.2752	-0.0129	0.1997
0.7246	-0.0366	0.2797	-0.0031	0.1958	-0.0065	0.1486	-0.0024	0.1069
0.7472	-0.1198	0.4484	-0.0631	0.3248	-0.0229	0.2367	-0.0090	0.1797
0.7676	-0.0166	0.2133	-0.0038	0.1553	0.0004	0.1148	0.0011	0.0809
0.7695	-0.0083	0.1885	-0.0001	0.1377	-0.0003	0.0995	0.0006	0.0711

Continued...

Notes: True parameter values are generated as  $\psi_{i0} \sim IIDU(0, 0.8)$  and  $a_{i0} \sim IIDN(1, 1)$ , for  $i = 1, 2, \dots, N$ . Gaussian errors are generated as  $\varepsilon_{it}/\sigma_{i0} \sim IIDN(0, 1)$ , with  $\sigma_{i0}^2 \sim IID[\chi^2(2)/4 + 0.5]$ , for  $i = 1, 2, \dots, N$ . Biases and RMSEs are computed as  $R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})$  and  $\sqrt{R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})^2}$ , for  $i = 1, 2, \dots, N$ . The spatial weight matrix  $\mathbf{W} = (w_{ij})$  has four connections so that  $w_{ij} = 1$  if  $j$  is equal to:  $i - 2, i - 1, i + 1, i + 2$ , and zero otherwise, for  $i, j = 1, 2, \dots, N$ . The number of replications is set to  $R = 2000$ . Estimates are sorted in ascending order according to their true values.

Table A2: ...Continued

$T$ $\psi_{i0}$	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$N = 75$								
0.0255	0.0267	0.6202	-0.0245	0.4836	0.0008	0.3483	-0.0117	0.2483
0.0397	0.0132	0.4191	0.0069	0.2911	0.0070	0.1953	0.0027	0.1411
0.0432	0.0206	0.3407	0.0045	0.2252	0.0118	0.1557	0.0010	0.1122
0.0478	0.0313	0.5108	0.0020	0.3792	-0.0022	0.2660	0.0034	0.1895
0.0608	0.0014	0.6206	0.0026	0.4802	-0.0032	0.3481	0.0011	0.2441
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3755	0.0320	0.3433	0.0193	0.2416	0.0105	0.1736	0.0040	0.1170
0.3786	-0.0032	0.4520	0.0111	0.3272	0.0087	0.2322	-0.0047	0.1655
0.3883	-0.0070	0.4295	-0.0062	0.3084	-0.0014	0.2189	0.0020	0.1469
0.3927	-0.0027	0.3001	-0.0082	0.2137	-0.0019	0.1508	0.0006	0.1035
0.3987	0.0068	0.2483	0.0008	0.1691	-0.0006	0.1154	0.0015	0.0840
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7472	-0.1273	0.4612	-0.0642	0.3257	-0.0277	0.2372	-0.0123	0.1796
0.7512	-0.0179	0.2441	-0.0098	0.1829	0.0023	0.1321	0.0006	0.0983
0.7536	-0.0333	0.2799	-0.0118	0.2091	0.0011	0.1569	0.0014	0.1157
0.7676	-0.0089	0.2052	-0.0015	0.1570	-0.0026	0.1120	0.0013	0.0795
0.7695	-0.0030	0.1801	0.0026	0.1372	-0.0006	0.1011	0.0012	0.0700
$N = 100$								
0.0244	0.0180	0.4109	0.0058	0.2973	-0.0004	0.2011	-0.0035	0.1410
0.0255	0.0034	0.6187	-0.0171	0.4866	-0.0220	0.3665	0.0036	0.2484
0.0397	0.0256	0.4217	0.0190	0.2996	0.0084	0.2004	0.0091	0.1397
0.0432	0.0298	0.3409	0.0137	0.2200	0.0108	0.1600	0.0015	0.1101
0.0478	0.0006	0.5147	0.0018	0.3793	-0.0068	0.2718	0.0018	0.1841
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3786	0.0054	0.4539	-0.0004	0.3203	-0.0041	0.2257	0.0018	0.1637
0.3839	0.0161	0.2757	0.0071	0.1969	-0.0041	0.1401	0.0062	0.0961
0.3883	-0.0086	0.4249	0.0075	0.3071	0.0051	0.2132	0.0032	0.1524
0.3927	-0.0025	0.3062	0.0011	0.2126	0.0014	0.1487	-0.0031	0.1060
0.3987	0.0065	0.2460	0.0047	0.1713	0.0031	0.1198	-0.0014	0.0847
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7536	-0.0410	0.2897	-0.0172	0.2097	-0.0016	0.1572	0.0009	0.1142
0.7676	-0.0099	0.2027	-0.0011	0.1550	0.0031	0.1093	0.0000	0.0778
0.7695	-0.0084	0.1883	0.0044	0.1332	0.0028	0.1032	0.0019	0.0704
0.7705	-0.0518	0.3042	-0.0215	0.2182	-0.0061	0.1645	-0.0021	0.1223
0.7904	-0.0152	0.1978	-0.0067	0.1482	0.0026	0.1090	0.0035	0.0803

Table A3: Empirical sizes of tests for individual spatial parameters from HSAR(1) model with Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections for different  $N$  and  $T$  combinations, using standard and sandwich formulae for the variances

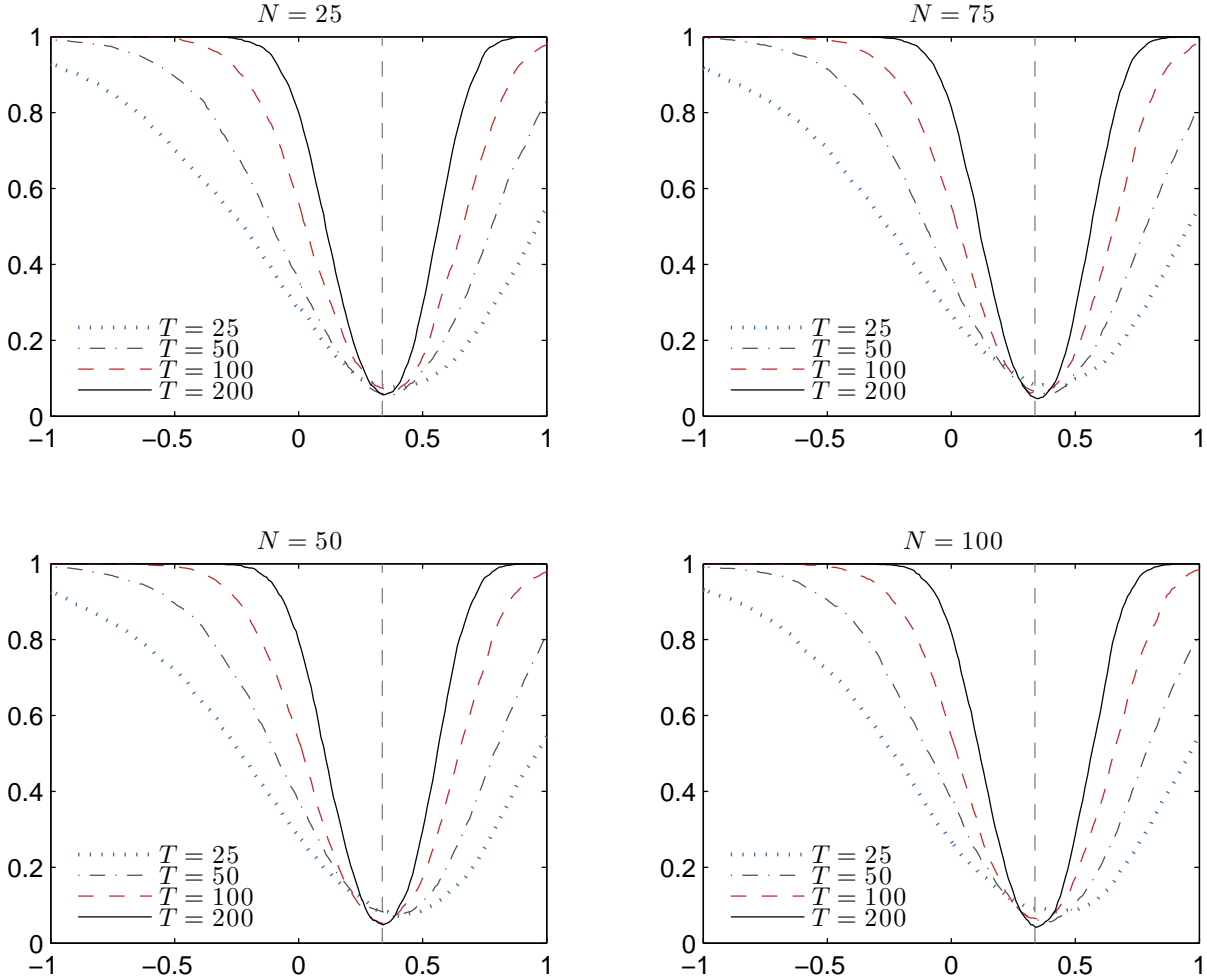
$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\psi_{i0}$	$N = 25$							
0.0255	0.0258	0.0400	0.0470	0.0460	0.0580	0.0635	0.0495	0.0510
0.0777	0.0546	0.0635	0.0540	0.0495	0.0960	0.0715	0.0575	0.0505
0.1109	0.0616	0.0610	0.0470	0.0505	0.0950	0.0725	0.0535	0.0495
0.1261	0.0581	0.0600	0.0545	0.0415	0.0860	0.0760	0.0610	0.0455
0.1948	0.0743	0.0605	0.0510	0.0505	0.0965	0.0715	0.0520	0.0510
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.3883	0.0263	0.0480	0.0585	0.0570	0.0550	0.0675	0.0640	0.0595
0.3987	0.0601	0.0640	0.0540	0.0505	0.0930	0.0780	0.0655	0.0530
0.4375	0.0323	0.0565	0.0465	0.0550	0.0625	0.0720	0.0565	0.0605
0.4682	0.0232	0.0435	0.0480	0.0480	0.0460	0.0585	0.0570	0.0470
0.5059	0.0677	0.0630	0.0490	0.0525	0.0920	0.0805	0.0525	0.0505
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.6726	0.0222	0.0375	0.0435	0.0490	0.0410	0.0585	0.0545	0.0530
0.6793	0.0227	0.0400	0.0570	0.0565	0.0435	0.0590	0.0670	0.0585
0.7127	0.0051	0.0130	0.0135	0.0185	0.0210	0.0175	0.0175	0.0225
0.7246	0.0081	0.0205	0.0350	0.0570	0.0260	0.0335	0.0445	0.0610
0.7676	0.0197	0.0170	0.0500	0.0465	0.0305	0.0265	0.0585	0.0540
	$N = 50$							
0.0255	0.0265	0.0456	0.0520	0.0460	0.0555	0.0590	0.0605	0.0500
0.0432	0.0672	0.0676	0.0580	0.0480	0.0835	0.0735	0.0650	0.0510
0.0608	0.0137	0.0320	0.0490	0.0530	0.0450	0.0570	0.0535	0.0585
0.0625	0.0631	0.0636	0.0495	0.0545	0.0920	0.0720	0.0595	0.0605
0.0777	0.0483	0.0581	0.0510	0.0430	0.0785	0.0635	0.0645	0.0425
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.3451	0.0356	0.0511	0.0545	0.0585	0.0635	0.0645	0.0660	0.0560
0.3510	0.0560	0.0641	0.0540	0.0510	0.0810	0.0800	0.0615	0.0605
0.3565	0.0656	0.0611	0.0535	0.0475	0.0890	0.0720	0.0585	0.0555
0.3755	0.0458	0.0581	0.0540	0.0505	0.0795	0.0710	0.0605	0.0545
0.3786	0.0254	0.0391	0.0455	0.0510	0.0495	0.0620	0.0615	0.0575
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.7127	0.0081	0.0095	0.0185	0.0225	0.0210	0.0180	0.0200	0.0215
0.7246	0.0153	0.0095	0.0380	0.0510	0.0275	0.0170	0.0490	0.0540
0.7472	0.0209	0.0160	0.0175	0.0150	0.0340	0.0255	0.0210	0.0195
0.7676	0.0224	0.0245	0.0525	0.0525	0.0420	0.0340	0.0600	0.0540
0.7695	0.0224	0.0290	0.0445	0.0500	0.0415	0.0460	0.0490	0.0555
	Continued...							

Notes: Nominal size is set to 5%. The standard and sandwich formulae are given by (40) and (41), respectively. See the notes to Table A2 for details of the data generating process. Size is computed under  $H_{i0}$ :  $\psi_i = \psi_{i0}$ , using a two-sided alternative where  $\psi_{i0}$  takes values in the range  $[0.0, 0.8]$  for  $i = 1, 2, \dots, N$ . The number of replications is set to  $R = 2000$ . Estimates are sorted in ascending order according to the true values of the spatial autoregressive parameters.

Table A3: ...Continued

$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\psi_{i0}$	$N = 75$							
0.0255	0.0240	0.0405	0.0490	0.0535	0.0545	0.0565	0.0535	0.0570
0.0397	0.0602	0.0565	0.0495	0.0510	0.0815	0.0720	0.0550	0.0530
0.0432	0.0724	0.0470	0.0480	0.0535	0.0975	0.0595	0.0545	0.0560
0.0478	0.0541	0.0555	0.0500	0.0565	0.0755	0.0705	0.0600	0.0605
0.0608	0.0112	0.0285	0.0505	0.0525	0.0365	0.0490	0.0580	0.0595
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3755	0.0500	0.0585	0.0580	0.0455	0.0760	0.0700	0.0645	0.0485
0.3786	0.0250	0.0420	0.0520	0.0575	0.0470	0.0625	0.0605	0.0650
0.3883	0.0270	0.0505	0.0635	0.0400	0.0580	0.0720	0.0705	0.0480
0.3927	0.0541	0.0550	0.0560	0.0520	0.0890	0.0645	0.0660	0.0550
0.3987	0.0704	0.0650	0.0540	0.0550	0.0925	0.0740	0.0595	0.0590
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7472	0.0173	0.0210	0.0220	0.0215	0.0270	0.0235	0.0250	0.0215
0.7512	0.0148	0.0160	0.0375	0.0590	0.0275	0.0250	0.0400	0.0650
0.7536	0.0168	0.0170	0.0215	0.0525	0.0290	0.0230	0.0305	0.0515
0.7676	0.0184	0.0240	0.0440	0.0435	0.0315	0.0365	0.0545	0.0495
0.7695	0.0209	0.0245	0.0560	0.0490	0.0370	0.0520	0.0615	0.0480
	$N = 100$							
0.0244	0.0554	0.0620	0.0500	0.0525	0.0765	0.0660	0.0570	0.0520
0.0255	0.0277	0.0420	0.0570	0.0590	0.0520	0.0635	0.0690	0.0600
0.0397	0.0620	0.0660	0.0490	0.0485	0.0850	0.0790	0.0530	0.0560
0.0432	0.0682	0.0480	0.0595	0.0505	0.0895	0.0570	0.0695	0.0545
0.0478	0.0538	0.0515	0.0530	0.0490	0.0815	0.0715	0.0595	0.0500
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3786	0.0210	0.0345	0.0505	0.0545	0.0460	0.0495	0.0515	0.0580
0.3839	0.0610	0.0580	0.0530	0.0525	0.0895	0.0665	0.0625	0.0600
0.3883	0.0354	0.0565	0.0575	0.0575	0.0645	0.0655	0.0670	0.0600
0.3927	0.0482	0.0570	0.0560	0.0570	0.0725	0.0750	0.0570	0.0580
0.3987	0.0687	0.0685	0.0570	0.0600	0.0990	0.0845	0.0625	0.0675
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7536	0.0179	0.0185	0.0230	0.0470	0.0285	0.0270	0.0255	0.0510
0.7676	0.0174	0.0265	0.0425	0.0425	0.0300	0.0380	0.0505	0.0465
0.7695	0.0297	0.0215	0.0595	0.0510	0.0420	0.0350	0.0665	0.0555
0.7705	0.0113	0.0150	0.0215	0.0315	0.0200	0.0210	0.0260	0.0460
0.7904	0.0287	0.0185	0.0370	0.0555	0.0430	0.0325	0.0475	0.0570

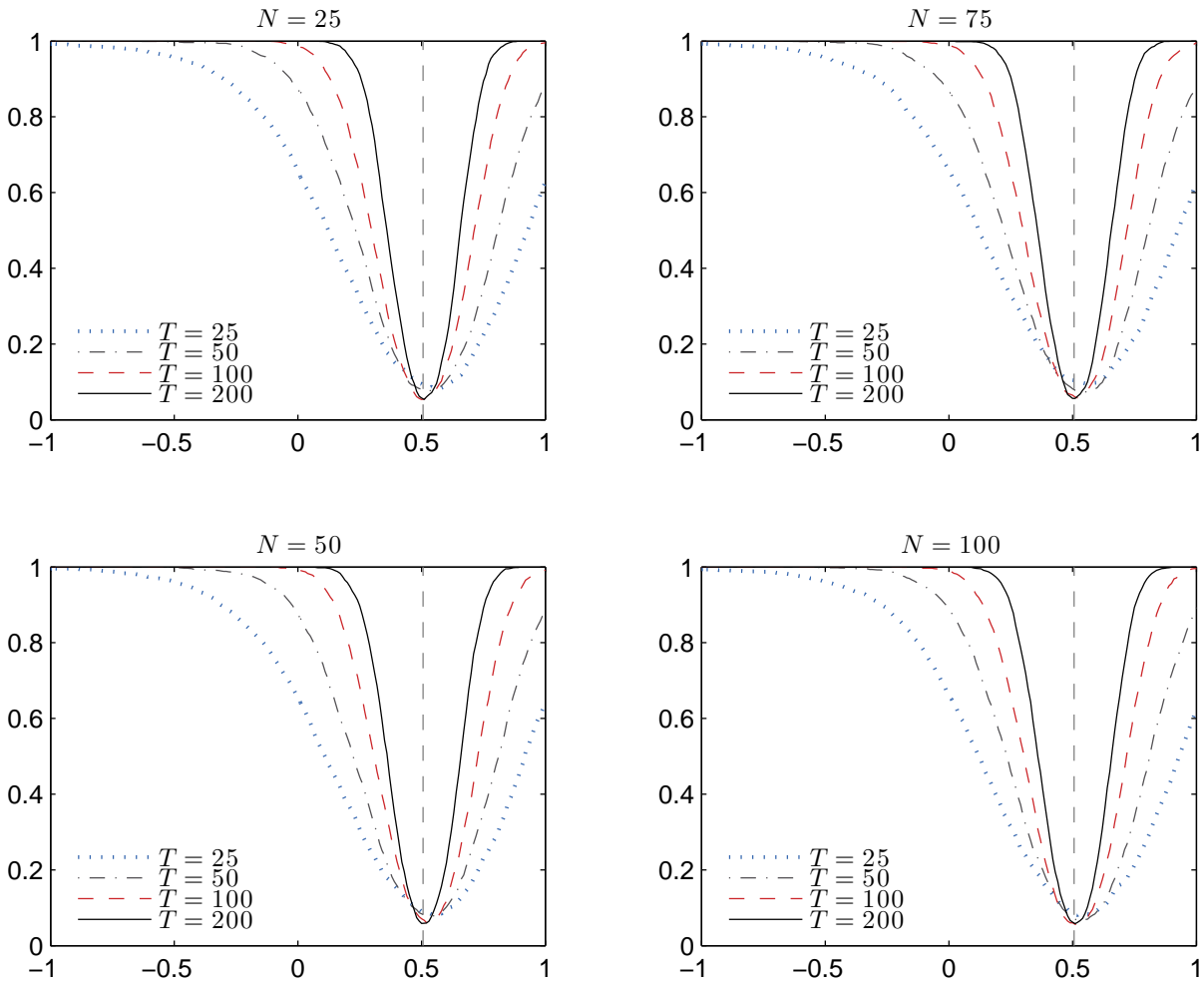
Figure A1: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.3374$  from HSAR(1) model with Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: The power functions are based on the sandwich formula given by (41). See the notes to Table A2 for details of the data generating process. Power is computed under  $\psi_i = \psi_{i0} + \delta$ , where  $\delta = -0.8, -0.791, \dots, 0.791, 0.8$  or until the parameter space boundaries of -1 and 1 are reached. The number of replications is set to  $R = 2000$ .

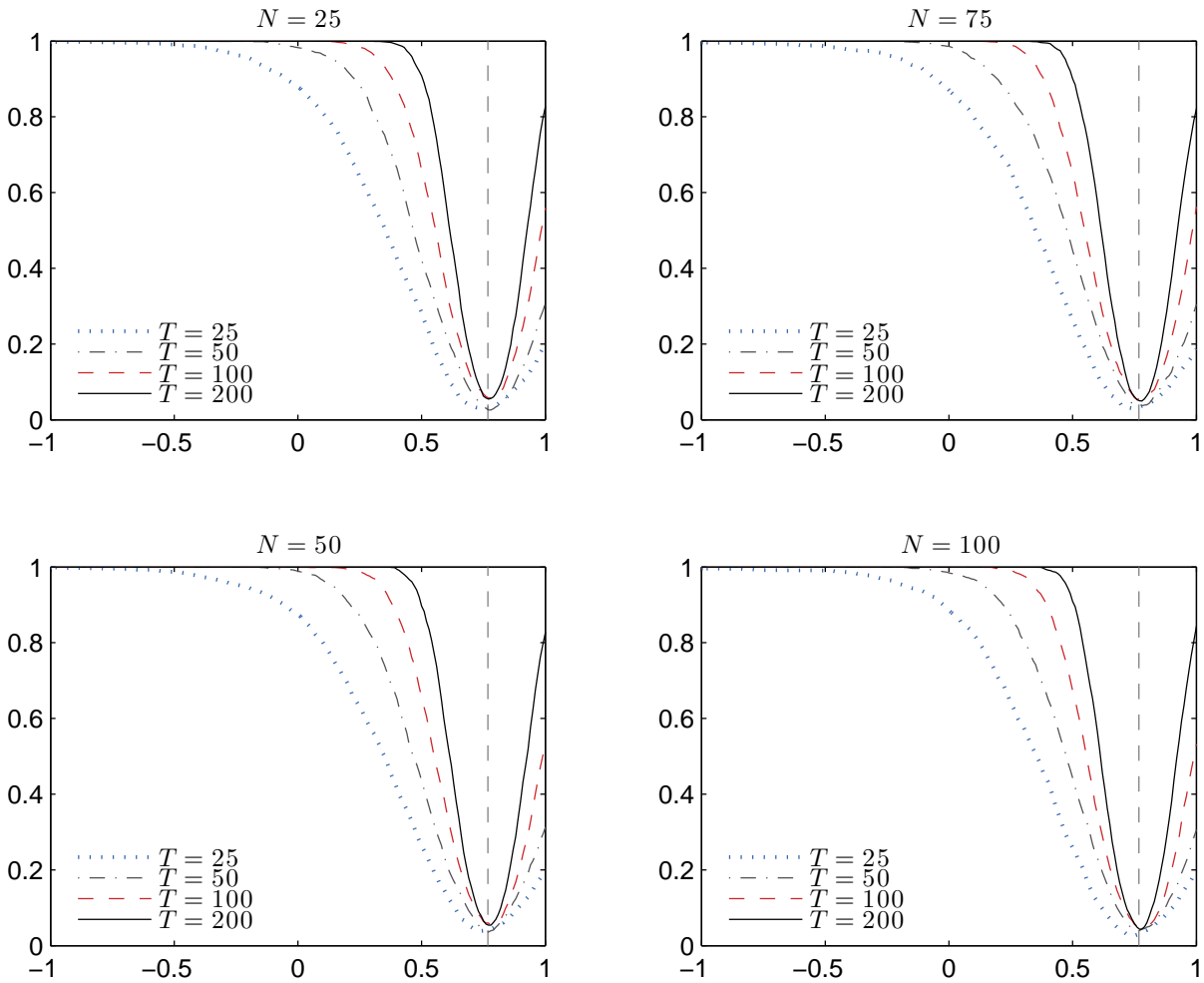


Figure A2: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.5059$  from HSAR(1) model with Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure A1.

Figure A3: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.7676$  from HSAR(1) model with Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure A1.

Table A4: Average bias and RMSE of spatial parameter estimates from HSAR(1) model with non-Gaussian errors over all  $N$  cross section units for different spatial weight matrices ( $\mathbf{W}$ ), and  $N$  and  $T$  combinations

$N$	25				50				75				100			
	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
$\mathbf{W}$ with 2 connections																
Average Bias																
$\tilde{a}_R$	0.0768	0.0424	0.0181	0.0078	0.0498	0.0250	0.0117	0.0054	0.0468	0.0233	0.0105	0.0037	0.0479	0.0239	0.0099	0.0046
$\tilde{\psi}_R$	-0.0353	-0.0190	-0.0078	-0.0034	-0.0270	-0.0140	-0.0066	-0.0028	-0.0265	-0.0130	-0.0057	-0.0022	-0.0253	-0.0124	-0.0055	-0.0023
Average RMSE																
$\tilde{a}_R$	0.8747	0.6456	0.4520	0.3226	0.7829	0.5836	0.4216	0.2934	0.7762	0.5662	0.4035	0.2798	0.7913	0.5813	0.4098	0.2840
$\tilde{\psi}_R$	0.3856	0.2853	0.2008	0.1435	0.3970	0.3006	0.2199	0.1555	0.3886	0.2877	0.2080	0.1448	0.3858	0.2879	0.2049	0.1425
$\mathbf{W}$ with 4 connections																
Average Bias																
$\tilde{a}_R$	0.0273	0.0099	0.0003	-0.0019	0.0206	0.0060	0.0002	-0.0006	0.0192	0.0053	0.0011	-0.0010	0.0180	0.0047	0.0002	-0.0007
$\tilde{\psi}_R$	-0.0104	-0.0037	0.0002	0.0009	-0.0078	-0.0016	0.0000	0.0005	-0.0080	-0.0020	-0.0000	0.0004	-0.0076	-0.0014	0.0001	0.0005
Average RMSE																
$\tilde{a}_R$	0.7593	0.5448	0.3814	0.2687	0.7109	0.5098	0.3593	0.2518	0.7025	0.4998	0.3525	0.2470	0.7148	0.5102	0.3582	0.2512
$\tilde{\psi}_R$	0.3269	0.2349	0.1653	0.1160	0.3501	0.2520	0.1777	0.1247	0.3417	0.2445	0.1722	0.1207	0.3416	0.2451	0.1724	0.1206
$\mathbf{W}$ with 10 connections																
Average Bias																
$\tilde{a}_R$	0.0282	0.0103	0.0010	-0.0027	0.0264	0.0069	0.0006	-0.0005	0.0282	0.0073	0.0011	-0.0018	0.0279	0.0075	0.0013	-0.0011
$\tilde{\psi}_R$	-0.0135	-0.0050	-0.0005	0.0014	-0.0125	-0.0031	-0.0002	0.0003	-0.0139	-0.0036	-0.0004	0.0007	-0.0138	-0.0038	-0.0006	0.0006
Average RMSE																
$\tilde{a}_R$	0.8362	0.6245	0.4536	0.3229	0.8127	0.6037	0.4397	0.3142	0.8078	0.6011	0.4394	0.3137	0.8182	0.6108	0.4441	0.3179
$\tilde{\psi}_R$	0.3711	0.2784	0.2032	0.1448	0.4068	0.3038	0.2216	0.1583	0.4001	0.2992	0.2192	0.1566	0.4023	0.3022	0.2203	0.1576

Notes: True parameter values are generated as  $\psi_{i0} \sim IIDU(0,0.8)$  and  $a_{i0} \sim IIDN(1,1)$ , for  $i = 1, 2, \dots, N$ . Non-Gaussian errors are generated as  $\varepsilon_{it}/\sigma_{i0} \sim IID\chi^2(2)-2$ , with  $\sigma_{i0}^2 \sim IID\chi^2(2)/4 + 0.5$ , for  $i = 1, 2, \dots, N$ . Average Bias and Average RMSE for  $\psi_R$  are calculated as  $(NR)^{-1} \sum_{i=1}^N \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})$  and  $N^{-1} \sum_{i=1}^N \left[ \sqrt{R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})^2} \right]$ , respectively. Average Bias and Average RMSE for  $\tilde{a}_R$  are computed similarly. The number of replications is set to  $R = 2000$ .

Table A5: Bias and RMSE for individual spatial parameter estimates from the HSAR(1) model with non-Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections for different  $N$  and  $T$  combinations

$T$	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$N = 25$								
0.0255	0.0209	0.6032	-0.0216	0.4843	0.0020	0.3612	-0.0052	0.2451
0.0777	-0.0162	0.4369	0.0127	0.3172	0.0026	0.2254	0.0006	0.1535
0.1109	0.0172	0.3166	0.0128	0.2034	0.0077	0.1474	0.0062	0.1026
0.1261	0.0042	0.2960	0.0064	0.2026	-0.0007	0.1371	0.0045	0.0945
0.1948	-0.0024	0.2250	-0.0043	0.1504	0.0026	0.1065	0.0020	0.0733
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3883	-0.0218	0.4323	-0.0047	0.3154	0.0016	0.2197	0.0013	0.1541
0.3987	0.0173	0.2490	0.0014	0.1818	0.0015	0.1210	0.0054	0.0855
0.4375	-0.0027	0.3798	-0.0092	0.2726	-0.0070	0.1890	-0.0019	0.1310
0.4682	0.0271	0.3660	0.0090	0.2687	0.0020	0.1878	0.0007	0.1377
0.5059	-0.0045	0.2411	0.0053	0.1660	0.0052	0.1133	0.0026	0.0787
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.6726	0.0042	0.2392	0.0109	0.1676	-0.0000	0.1191	0.0019	0.0873
0.6793	-0.0068	0.2444	-0.0088	0.1791	-0.0049	0.1303	0.0034	0.0919
0.7127	-0.1359	0.5146	-0.0809	0.3791	-0.0289	0.2629	-0.0139	0.1960
0.7246	-0.0413	0.2965	-0.0176	0.2102	0.0047	0.1477	0.0029	0.1048
0.7676	-0.0172	0.2093	-0.0084	0.1540	0.0045	0.1090	0.0034	0.0805
$N = 50$								
0.0255	0.0108	0.6082	-0.0040	0.4821	-0.0053	0.3524	-0.0160	0.2469
0.0432	0.0140	0.3467	0.0010	0.2366	0.0080	0.1634	0.0040	0.1151
0.0608	-0.0006	0.6348	-0.0052	0.4967	-0.0075	0.3684	0.0042	0.2562
0.0625	0.0128	0.2799	0.0136	0.1937	0.0021	0.1318	-0.0022	0.0963
0.0777	-0.0021	0.4554	0.0061	0.3173	-0.0001	0.2229	0.0055	0.1578
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3451	-0.0070	0.4296	0.0106	0.3033	0.0017	0.2103	0.0019	0.1445
0.3510	-0.0024	0.3472	0.0056	0.2495	-0.0006	0.1736	0.0008	0.1200
0.3565	0.0123	0.2566	0.0136	0.1691	0.0001	0.1161	0.0010	0.0840
0.3755	0.0202	0.3517	0.0058	0.2472	0.0086	0.1745	0.0046	0.1198
0.3786	-0.0009	0.4447	0.0021	0.3439	-0.0026	0.2350	0.0017	0.1664
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7127	-0.1152	0.4772	-0.0560	0.3609	-0.0336	0.2715	-0.0026	0.1936
0.7246	-0.0272	0.2825	-0.0144	0.2056	-0.0027	0.1518	-0.0029	0.1094
0.7472	-0.1391	0.4807	-0.0687	0.3342	-0.0273	0.2451	-0.0126	0.1840
0.7676	-0.0260	0.2154	0.0009	0.1523	-0.0039	0.1134	0.0017	0.0825
0.7695	-0.0030	0.1810	0.0005	0.1397	0.0018	0.1023	0.0025	0.0719

Continued...

Notes: True parameter values are generated as  $\psi_{i0} \sim IIDU(0, 0.8)$  and  $a_{i0} \sim IIDN(1, 1)$ , for  $i = 1, 2, \dots, N$ . Non-Gaussian errors are generated as  $\varepsilon_{it}/\sigma_{i0} \sim IID\chi^2_{(2)} - 2$ , with  $\sigma_{i0}^2 \sim IID[\chi^2(2)/4 + 0.5]$ , for  $i = 1, 2, \dots, N$ . Biases and RMSEs are computed as  $R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})$  and  $\sqrt{R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})^2}$ , for  $i = 1, 2, \dots, N$ . The spatial weight matrix  $\mathbf{W} = (w_{ij})$  has four connections so that  $w_{ij} = 1$  if  $j$  is equal to:  $i - 2, i - 1, i + 1, i + 2$ , and zero otherwise, for  $i, j = 1, 2, \dots, N$ . The number of replications is set to  $R = 2000$ . Estimates are sorted in ascending order according to their true values.

Table A5: ...Continued

$T$ $\psi_{i0}$	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$N = 75$								
0.0255	0.0117	0.6045	-0.0186	0.4884	-0.0045	0.3536	0.0006	0.2509
0.0397	0.0032	0.4148	0.0169	0.3004	-0.0007	0.1998	0.0015	0.1445
0.0432	0.0255	0.3465	0.0205	0.2336	0.0110	0.1589	0.0082	0.1138
0.0478	-0.0129	0.5195	-0.0032	0.3722	0.0036	0.2658	-0.0003	0.1891
0.0608	0.0050	0.6098	-0.0248	0.5085	-0.0086	0.3599	-0.0070	0.2434
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3755	0.0130	0.3460	0.0088	0.2563	0.0085	0.1752	-0.0009	0.1229
0.3786	-0.0263	0.4596	-0.0030	0.3395	0.0063	0.2390	0.0014	0.1685
0.3883	-0.0066	0.4182	0.0015	0.3066	0.0006	0.2131	0.0030	0.1467
0.3927	0.0094	0.3107	0.0076	0.2184	-0.0072	0.1534	0.0001	0.1049
0.3987	0.0031	0.2573	0.0029	0.1690	0.0057	0.1198	-0.0014	0.0826
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7472	-0.1263	0.4652	-0.0637	0.3292	-0.0346	0.2463	-0.0149	0.1779
0.7512	-0.0263	0.2590	-0.0135	0.1854	-0.0024	0.1346	-0.0033	0.0955
0.7536	-0.0428	0.3074	-0.0220	0.2118	-0.0084	0.1569	0.0012	0.1133
0.7676	-0.0214	0.2239	-0.0035	0.1530	-0.0007	0.1123	0.0000	0.0784
0.7695	-0.0102	0.1797	-0.0043	0.1322	-0.0018	0.1009	0.0000	0.0706
$N = 100$								
0.0244	-0.0003	0.4104	-0.0062	0.3049	0.0078	0.2058	-0.0014	0.1431
0.0255	-0.0041	0.6076	0.0040	0.4716	-0.0182	0.3635	-0.0104	0.2431
0.0397	0.0150	0.4155	0.0063	0.3024	-0.0049	0.2144	0.0016	0.1430
0.0432	0.0182	0.3408	0.0099	0.2339	0.0063	0.1627	0.0033	0.1108
0.0478	0.0306	0.5120	0.0106	0.3849	0.0035	0.2614	-0.0038	0.1864
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3786	-0.0173	0.4676	0.0157	0.3286	0.0082	0.2337	-0.0013	0.1665
0.3839	0.0041	0.2833	0.0100	0.2016	0.0012	0.1384	0.0012	0.0952
0.3883	-0.0137	0.4248	-0.0020	0.3067	-0.0020	0.2216	0.0003	0.1511
0.3927	0.0039	0.3064	0.0060	0.2231	0.0049	0.1517	0.0017	0.1037
0.3987	-0.0036	0.2557	0.0058	0.1784	0.0035	0.1205	0.0010	0.0815
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7536	-0.0480	0.3011	-0.0151	0.2115	-0.0024	0.1510	0.0053	0.1156
0.7676	-0.0122	0.2089	-0.0054	0.1494	0.0018	0.1125	0.0077	0.0814
0.7695	-0.0084	0.1895	0.0007	0.1347	0.0060	0.0999	-0.0010	0.0715
0.7705	-0.0844	0.3640	-0.0224	0.2283	-0.0083	0.1664	-0.0016	0.1236
0.7904	-0.0207	0.1999	-0.0124	0.1508	0.0017	0.1100	0.0012	0.0792

Table A6: Empirical sizes of tests for individual spatial parameters from HSAR(1) model with non-Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections for different  $N$  and  $T$  combinations, using standard and sandwich formulae for the variances

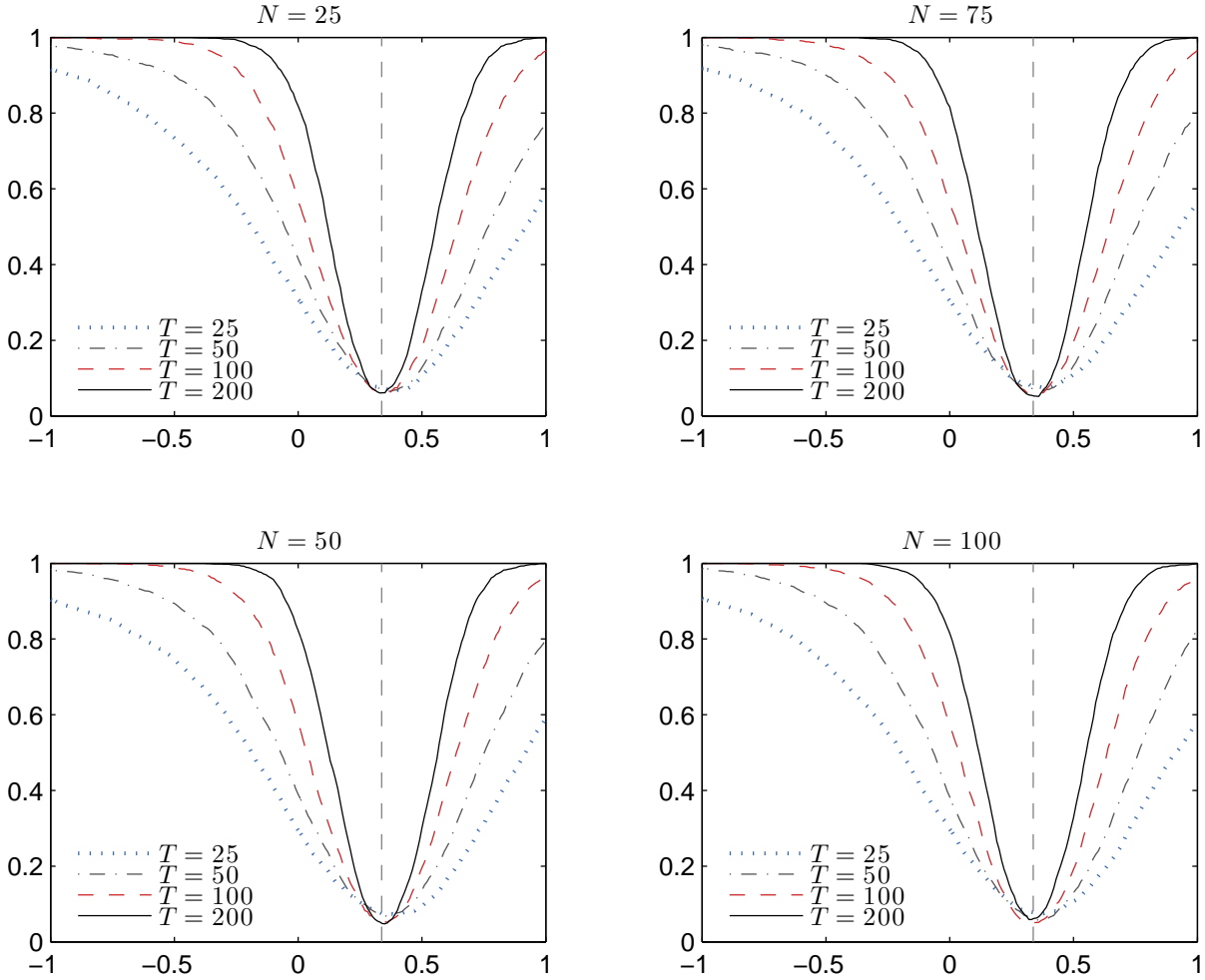
$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\psi_{i0}$	$N = 25$							
0.0255	0.0329	0.0435	0.0555	0.0500	0.0590	0.0625	0.0625	0.0510
0.0777	0.0463	0.0525	0.0550	0.0535	0.0680	0.0600	0.0540	0.0470
0.1109	0.0602	0.0575	0.0475	0.0500	0.0885	0.0630	0.0525	0.0490
0.1261	0.0633	0.0695	0.0505	0.0530	0.0815	0.0775	0.0565	0.0595
0.1948	0.0643	0.0530	0.0560	0.0440	0.0860	0.0750	0.0680	0.0545
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3883	0.0237	0.0475	0.0535	0.0535	0.0515	0.0695	0.0615	0.0555
0.3987	0.0607	0.0715	0.0530	0.0605	0.0805	0.0865	0.0620	0.0610
0.4375	0.0412	0.0540	0.0475	0.0520	0.0685	0.0620	0.0535	0.0540
0.4682	0.0273	0.0430	0.0515	0.0530	0.0475	0.0480	0.0545	0.0540
0.5059	0.0535	0.0515	0.0565	0.0595	0.0735	0.0655	0.0625	0.0620
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.6726	0.0304	0.0350	0.0415	0.0480	0.0565	0.0495	0.0525	0.0570
0.6793	0.0268	0.0330	0.0510	0.0520	0.0485	0.0490	0.0620	0.0540
0.7127	0.0026	0.0075	0.0065	0.0165	0.0175	0.0165	0.0145	0.0210
0.7246	0.0149	0.0160	0.0405	0.0505	0.0375	0.0320	0.0505	0.0550
0.7676	0.0180	0.0245	0.0420	0.0540	0.0470	0.0415	0.0470	0.0515
	$N = 50$							
0.0255	0.0265	0.0400	0.0530	0.0480	0.0560	0.0580	0.0550	0.0540
0.0432	0.0690	0.0561	0.0530	0.0450	0.0885	0.0675	0.0635	0.0530
0.0608	0.0254	0.0355	0.0430	0.0525	0.0615	0.0605	0.0585	0.0590
0.0625	0.0628	0.0576	0.0505	0.0525	0.0705	0.0625	0.0575	0.0590
0.0777	0.0513	0.0495	0.0575	0.0555	0.0835	0.0640	0.0575	0.0540
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3451	0.0420	0.0591	0.0575	0.0565	0.0725	0.0675	0.0570	0.0535
0.3510	0.0482	0.0541	0.0605	0.0555	0.0725	0.0650	0.0650	0.0510
0.3565	0.0731	0.0671	0.0515	0.0635	0.0840	0.0680	0.0650	0.0690
0.3755	0.0425	0.0526	0.0555	0.0530	0.0675	0.0640	0.0565	0.0515
0.3786	0.0249	0.0395	0.0460	0.0535	0.0495	0.0490	0.0515	0.0585
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7127	0.0057	0.0060	0.0120	0.0130	0.0165	0.0125	0.0210	0.0200
0.7246	0.0104	0.0140	0.0385	0.0475	0.0320	0.0285	0.0480	0.0530
0.7472	0.0182	0.0110	0.0185	0.0225	0.0295	0.0210	0.0295	0.0335
0.7676	0.0156	0.0285	0.0395	0.0625	0.0450	0.0475	0.0485	0.0615
0.7695	0.0228	0.0305	0.0485	0.0540	0.0445	0.0450	0.0565	0.0595
	Continued...							

Notes: Nominal size is set to 5%. The standard and sandwich formulae are given by (40) and (41), respectively. See the notes to Table A5 for details of the data generating process. Size is computed under  $H_{i0}$ :  $\psi_i = \psi_{i0}$ , using a two-sided alternative where  $\psi_{i0}$  takes values in the range  $[0.0, 0.8]$  for  $i = 1, 2, \dots, N$ . The number of replications is set to  $R = 2000$ . Estimates are sorted in ascending order according to the true values of the spatial autoregressive parameters.

Table A6: ...Continued

$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\psi_{i0}$	$N = 75$							
0.0255	0.0283	0.0445	0.0545	0.0590	0.0590	0.0600	0.0630	0.0605
0.0397	0.0561	0.0560	0.0485	0.0565	0.0760	0.0690	0.0540	0.0540
0.0432	0.0614	0.0575	0.0565	0.0535	0.0790	0.0640	0.0615	0.0520
0.0478	0.0519	0.0470	0.0480	0.0575	0.0720	0.0615	0.0525	0.0555
0.0608	0.0226	0.0360	0.0560	0.0490	0.0535	0.0515	0.0515	0.0480
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3755	0.0456	0.0620	0.0575	0.0495	0.0715	0.0715	0.0595	0.0585
0.3786	0.0262	0.0460	0.0550	0.0640	0.0525	0.0570	0.0620	0.0645
0.3883	0.0325	0.0480	0.0515	0.0470	0.0535	0.0635	0.0550	0.0465
0.3927	0.0614	0.0590	0.0565	0.0520	0.0815	0.0605	0.0685	0.0595
0.3987	0.0672	0.0565	0.0595	0.0575	0.0875	0.0745	0.0630	0.0630
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7472	0.0115	0.0145	0.0185	0.0235	0.0260	0.0285	0.0255	0.0320
0.7512	0.0110	0.0180	0.0325	0.0460	0.0350	0.0405	0.0410	0.0495
0.7536	0.0121	0.0125	0.0255	0.0425	0.0205	0.0300	0.0335	0.0490
0.7676	0.0241	0.0220	0.0465	0.0455	0.0470	0.0415	0.0620	0.0505
0.7695	0.0199	0.0255	0.0545	0.0505	0.0505	0.0485	0.0560	0.0600
	$N = 100$							
0.0244	0.0535	0.0541	0.0550	0.0585	0.0745	0.0630	0.0625	0.0620
0.0255	0.0270	0.0455	0.0520	0.0495	0.0520	0.0590	0.0575	0.0520
0.0397	0.0577	0.0591	0.0570	0.0530	0.0765	0.0615	0.0615	0.0490
0.0432	0.0572	0.0531	0.0505	0.0405	0.0755	0.0580	0.0515	0.0440
0.0478	0.0556	0.0616	0.0510	0.0440	0.0760	0.0680	0.0610	0.0415
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3786	0.0355	0.0410	0.0550	0.0550	0.0605	0.0580	0.0580	0.0545
0.3839	0.0572	0.0571	0.0565	0.0465	0.0735	0.0690	0.0685	0.0495
0.3883	0.0328	0.0460	0.0530	0.0460	0.0475	0.0515	0.0570	0.0470
0.3927	0.0588	0.0596	0.0630	0.0515	0.0740	0.0750	0.0670	0.0515
0.3987	0.0678	0.0581	0.0510	0.0535	0.0855	0.0655	0.0605	0.0520
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7536	0.0127	0.0175	0.0220	0.0470	0.0325	0.0270	0.0310	0.0450
0.7676	0.0143	0.0205	0.0445	0.0515	0.0335	0.0425	0.0565	0.0530
0.7695	0.0217	0.0295	0.0570	0.0510	0.0480	0.0495	0.0645	0.0575
0.7705	0.0111	0.0110	0.0185	0.0330	0.0275	0.0230	0.0305	0.0360
0.7904	0.0169	0.0285	0.0380	0.0535	0.0415	0.0485	0.0420	0.0525

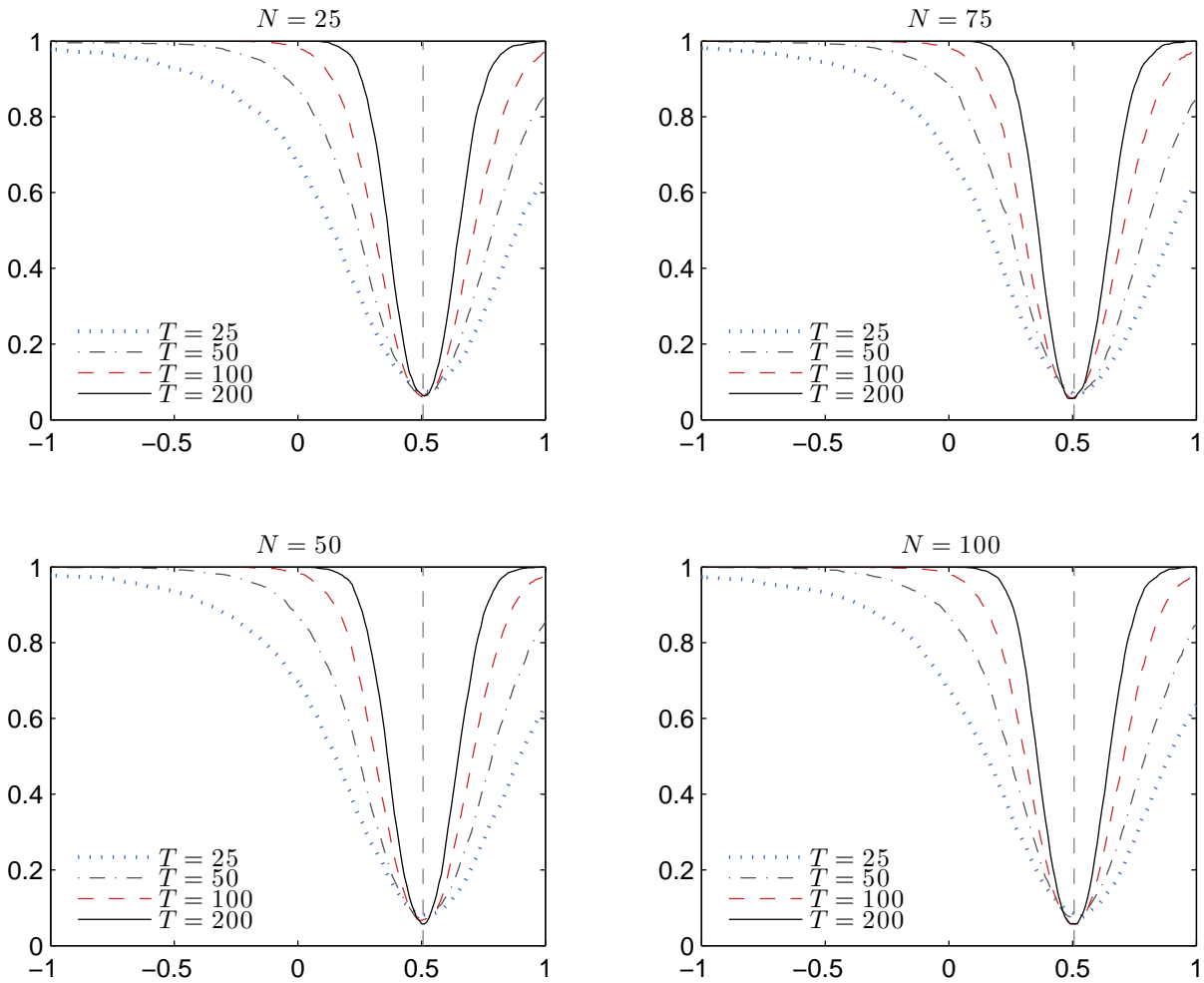
Figure A4: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.3374$  from HSAR(1) model with non-Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: The power functions are based on the sandwich formula given by (41). See the notes to Table A5 for details of the data generating process. Power is computed under  $\psi_i = \psi_{i0} + \delta$ , where  $\delta = -0.8, -0.791, \dots, 0.791, 0.8$  or until the parameter space boundaries of -1 and 1 are reached. The number of replications is set to  $R = 2000$ .

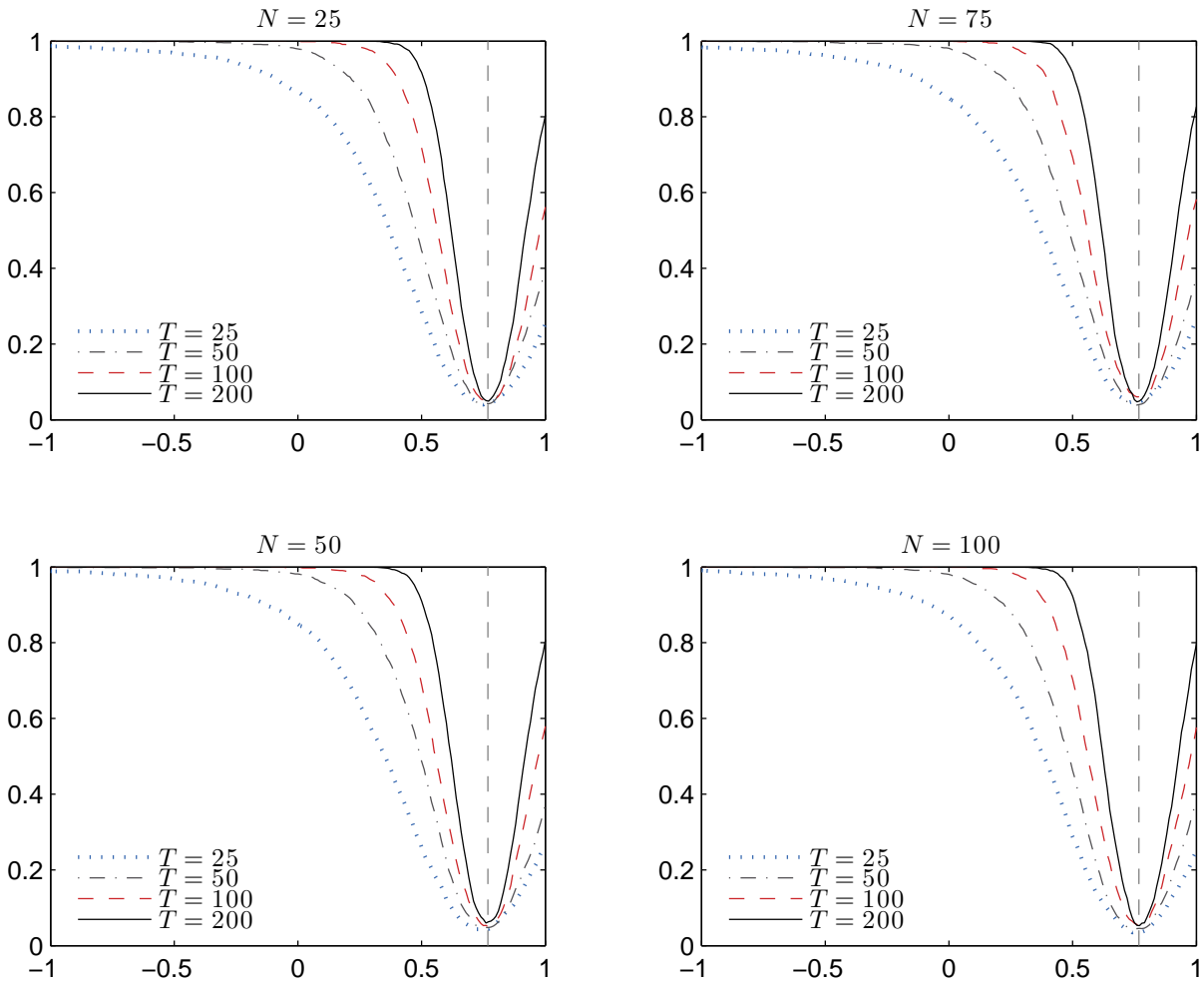


Figure A5: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.5059$  from HSAR(1) model with non-Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure A4.

Figure A6: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.7676$  from HSAR(1) model with non-Gaussian errors and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure A4.

Table B1: Average bias and RMSE of spatial parameter estimates from HSAR(1) model with Gaussian errors and one exogenous regressor over all  $N$  cross section units for different spatial weight matrices ( $\mathbf{W}$ ), and  $N$  and  $T$  combinations

$N$	25				50				75				100			
	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
$\mathbf{W}$ with 2 connections																
Average Bias																
$\bar{\alpha}_R$	0.0399	0.0167	0.0071	0.0025	0.0324	0.0131	0.0055	0.0022	0.0310	0.0118	0.0058	0.0021	0.0322	0.0135	0.0048	0.0026
$\bar{\psi}_R$	-0.0178	-0.0069	-0.0031	-0.0011	-0.0175	-0.0073	-0.0032	-0.0014	-0.0168	-0.0063	-0.0030	-0.0012	-0.0161	-0.0070	-0.0028	-0.0012
$\bar{\beta}_R$	0.0166	0.0070	0.0036	0.0019	0.0139	0.0067	0.0034	0.0017	0.0124	0.0062	0.0029	0.0012	0.0133	0.0063	0.0035	0.0013
Average RMSE																
$\bar{\alpha}_R$	0.6369	0.4294	0.2958	0.2040	0.6060	0.4039	0.2790	0.1932	0.6134	0.4130	0.2821	0.1968	0.6316	0.4228	0.2899	0.2009
$\bar{\psi}_R$	0.2729	0.1838	0.1263	0.0873	0.2983	0.2006	0.1388	0.0957	0.3004	0.2026	0.1381	0.0962	0.3004	0.2017	0.1375	0.0953
$\bar{\beta}_R$	0.2518	0.1655	0.1136	0.0793	0.2530	0.1664	0.1144	0.0793	0.2484	0.1637	0.1119	0.0774	0.2456	0.1629	0.1109	0.0774
$\mathbf{W}$ with 4 connections																
Average Bias																
$\bar{\alpha}_R$	0.0161	0.0049	0.0028	0.0003	0.0150	0.0059	0.0029	-0.0001	0.0135	0.0027	0.0016	0.0001	0.0128	0.0041	0.0001	0.0004
$\bar{\psi}_R$	-0.0064	-0.0015	-0.0011	-0.0001	-0.0056	-0.0021	-0.0010	-0.0000	-0.0054	-0.0009	-0.0004	-0.0000	-0.0051	-0.0015	-0.0001	-0.0001
$\bar{\beta}_R$	0.0068	0.0023	0.0015	0.0008	0.0051	0.0025	0.0014	0.0008	0.0042	0.0021	0.0009	0.0002	0.0051	0.0022	0.0015	0.0004
Average RMSE																
$\bar{\alpha}_R$	0.6125	0.4216	0.2953	0.2061	0.5910	0.4062	0.2845	0.1991	0.5945	0.4100	0.2866	0.2018	0.6072	0.4171	0.2925	0.2047
$\bar{\psi}_R$	0.2582	0.1778	0.1243	0.0868	0.2836	0.1949	0.1363	0.0951	0.2839	0.1958	0.1367	0.0961	0.2848	0.1963	0.1371	0.0960
$\bar{\beta}_R$	0.2255	0.1511	0.1049	0.0736	0.2283	0.1533	0.1063	0.0742	0.2255	0.1507	0.1044	0.0727	0.2227	0.1504	0.1036	0.0728
$\mathbf{W}$ with 10 connections																
Average Bias																
$\bar{\alpha}_R$	0.0161	0.0048	0.0016	-0.0000	0.0207	0.0068	0.0030	-0.0009	0.0198	0.0040	0.0006	-0.0003	0.0211	0.0067	0.0010	0.0001
$\bar{\psi}_R$	-0.0077	-0.0017	-0.0008	-0.0001	-0.0093	-0.0031	-0.0012	0.0004	-0.0099	-0.0020	-0.0002	0.0001	-0.0100	-0.0031	-0.0005	-0.0000
$\bar{\beta}_R$	0.0044	0.0009	0.0009	0.0006	0.0027	0.0015	0.0009	0.0006	0.0021	0.0011	0.0003	-0.0000	0.0033	0.0013	0.0010	0.0002
Average RMSE																
$\bar{\alpha}_R$	0.6863	0.4848	0.3438	0.2419	0.6893	0.4868	0.3449	0.2425	0.6990	0.4962	0.3538	0.2499	0.7104	0.5038	0.3579	0.2532
$\bar{\psi}_R$	0.3005	0.2131	0.1512	0.1065	0.3417	0.2416	0.1712	0.1204	0.3440	0.2444	0.1744	0.1230	0.3469	0.2469	0.1752	0.1241
$\bar{\beta}_R$	0.2124	0.1435	0.0997	0.0700	0.2162	0.1462	0.1016	0.0712	0.2139	0.1444	0.1005	0.0700	0.2117	0.1439	0.0995	0.0701

Notes: True parameter values are generated as  $\psi_{i0} \sim IIDU(0, 0.8)$ ,  $\alpha_{i0} \sim IIDN(1, 1)$  and  $\beta_{i0} \sim IIDU(0, 1)$ , for  $i = 1, 2, \dots, N$ . Gaussian errors are generated as  $\varepsilon_{it}/\sigma_{i0} \sim IIDN(0, 1)$ , with  $\sigma_{i0}^2 \sim IID[\chi^2(2)/4 + 0.5]$ , for  $i = 1, 2, \dots, N$ . Exogenous regressors are spatially correlated across  $i$  and generated by (53), with  $\phi_i = 0.5$ . Average Bias and Average RMSE for  $\bar{\psi}_R$  are calculated as  $(NR)^{-1} \sum_{i=1}^N \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})$  and  $N^{-1} \sum_{i=1}^N \left[ \sqrt{R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})^2} \right]$ , respectively. Average Bias and Average RMSE

for  $\bar{\alpha}_R$  and  $\bar{\beta}_R$  are computed similarly. The number of replications is set to  $R = 2000$ .

Table B2: Bias and RMSE for individual spatial parameter estimates from the HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections for different  $N$  and  $T$  combinations

$T$	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$N = 25$								
0.0255	-0.0337	0.5203	0.0012	0.3751	-0.0087	0.2565	-0.0009	0.1787
0.0777	0.0124	0.3699	0.0000	0.2522	-0.0017	0.1694	0.0024	0.1169
0.1109	0.0146	0.2377	0.0036	0.1594	-0.0002	0.1109	-0.0012	0.0772
0.1261	0.0006	0.2136	0.0025	0.1465	-0.0010	0.1021	0.0009	0.0692
0.1948	-0.0017	0.1981	-0.0081	0.1289	-0.0022	0.0866	0.0039	0.0610
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3883	-0.0028	0.3281	-0.0023	0.2182	0.0033	0.1526	-0.0034	0.1080
0.3987	0.0012	0.1973	0.0010	0.1333	0.0044	0.0938	-0.0015	0.0627
0.4375	-0.0003	0.2200	-0.0011	0.1568	-0.0014	0.1022	-0.0003	0.0733
0.4682	0.0125	0.3026	0.0090	0.2060	0.0055	0.1453	-0.0007	0.0981
0.5059	-0.0050	0.1782	0.0018	0.1242	-0.0031	0.0815	0.0014	0.0566
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.6726	-0.0010	0.1922	0.0069	0.1336	0.0018	0.0908	0.0005	0.0674
0.6793	-0.0006	0.1962	0.0038	0.1372	-0.0009	0.0963	-0.0007	0.0673
0.7127	-0.0983	0.4073	-0.0381	0.2762	-0.0161	0.2096	-0.0025	0.1491
0.7246	-0.0122	0.2029	-0.0045	0.1464	0.0000	0.1080	-0.0035	0.0738
0.7676	-0.0115	0.1776	-0.0018	0.1324	-0.0020	0.0927	0.0004	0.0666
$N = 50$								
0.0255	-0.0138	0.5170	-0.0099	0.3696	-0.0081	0.2658	-0.0035	0.1775
0.0432	0.0022	0.2436	0.0008	0.1570	0.0020	0.1132	0.0033	0.0759
0.0608	-0.0314	0.5547	0.0148	0.3788	0.0016	0.2727	-0.0020	0.1844
0.0625	0.0096	0.2626	0.0101	0.1757	0.0001	0.1202	-0.0005	0.0856
0.0777	0.0053	0.3656	0.0049	0.2467	0.0058	0.1719	0.0042	0.1149
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3451	-0.0011	0.3629	-0.0006	0.2523	-0.0013	0.1690	-0.0003	0.1211
0.3510	0.0026	0.2719	0.0031	0.1813	-0.0042	0.1264	0.0002	0.0881
0.3565	-0.0012	0.1894	0.0002	0.1303	-0.0001	0.0884	0.0021	0.0603
0.3755	0.0197	0.3169	0.0041	0.2158	0.0018	0.1471	-0.0042	0.1031
0.3786	-0.0018	0.3209	-0.0040	0.2211	-0.0020	0.1541	-0.0011	0.1065
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7127	-0.0728	0.3885	-0.0340	0.2853	-0.0126	0.2119	-0.0002	0.1505
0.7246	-0.0226	0.2105	-0.0029	0.1474	-0.0032	0.1046	-0.0009	0.0726
0.7472	-0.0833	0.3773	-0.0428	0.2690	-0.0150	0.1940	0.0015	0.1473
0.7676	-0.0111	0.1844	-0.0035	0.1310	-0.0009	0.0948	-0.0010	0.0652
0.7695	-0.0037	0.1640	-0.0013	0.1181	-0.0019	0.0812	-0.0002	0.0594

Continued...

Notes: True parameter values are generated as  $\psi_{i0} \sim IIDU(0, 0.8)$ ,  $a_{i0} \sim IIDN(1, 1)$  and  $\beta_{i0} \sim IIDU(0, 1)$ , for  $i = 1, 2, \dots, N$ . Gaussian errors are generated as  $\varepsilon_{it}/\sigma_{i0} \sim IIDN(0, 1)$ , with  $\sigma_{i0}^2 \sim IID[\chi^2(2)/4 + 0.5]$ , for  $i = 1, 2, \dots, N$ . Exogenous regressors are spatially correlated across  $i$  and generated by (53), with  $\phi_i = 0.5$ . Biases and RMSEs are computed as  $R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})$  and  $\sqrt{R^{-1} \sum_{r=1}^R (\hat{\psi}_{i,r} - \psi_{i0})^2}$ , for  $i = 1, 2, \dots, N$ . The spatial weight matrix  $\mathbf{W} = (w_{ij})$  has four connections so that  $w_{ij} = 1$  if  $j$  is equal to  $i - 2, i - 1, i + 1, i + 2$ , and zero otherwise, for  $i, j = 1, 2, \dots, N$ . The number of replications is set to  $R = 2000$ . Estimates are sorted in ascending order according to their true values.

Table B2: ...Continued

$T$ $\psi_{i0}$	25		50		100		200	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
$N = 75$								
0.0255	0.0135	0.5149	-0.0100	0.3766	0.0029	0.2486	0.0011	0.1788
0.0397	0.0091	0.3646	0.0004	0.2362	-0.0001	0.1584	0.0013	0.1176
0.0432	0.0054	0.2436	0.0025	0.1554	0.0038	0.1094	0.0016	0.0801
0.0478	0.0251	0.4871	0.0005	0.3508	-0.0027	0.2424	0.0022	0.1711
0.0608	0.0113	0.5441	0.0007	0.3930	0.0039	0.2774	-0.0000	0.1987
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3755	0.0142	0.3024	0.0124	0.2095	0.0074	0.1535	0.0009	0.1010
0.3786	0.0005	0.3202	0.0052	0.2235	0.0037	0.1527	-0.0032	0.1058
0.3883	-0.0115	0.3280	-0.0037	0.2187	-0.0042	0.1534	0.0002	0.1065
0.3927	-0.0082	0.2937	-0.0073	0.1972	-0.0023	0.1370	0.0009	0.0936
0.3987	-0.0001	0.1977	0.0003	0.1347	-0.0013	0.0921	0.0003	0.0645
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7472	-0.0783	0.3641	-0.0400	0.2682	-0.0129	0.1974	-0.0087	0.1462
0.7512	-0.0107	0.2058	-0.0037	0.1512	0.0002	0.1088	-0.0001	0.0783
0.7536	-0.0336	0.2663	-0.0091	0.1952	0.0010	0.1434	0.0004	0.1049
0.7676	-0.0079	0.1785	0.0032	0.1335	-0.0011	0.0959	0.0006	0.0628
0.7695	-0.0026	0.1579	0.0009	0.1157	-0.0013	0.0854	-0.0002	0.0584
$N = 100$								
0.0244	-0.0007	0.3198	0.0015	0.2107	-0.0000	0.1435	-0.0015	0.0994
0.0255	0.0075	0.5309	-0.0159	0.3772	-0.0105	0.2603	0.0006	0.1814
0.0397	-0.0004	0.3504	0.0137	0.2389	0.0031	0.1612	0.0028	0.1133
0.0432	0.0036	0.2389	0.0126	0.1572	0.0030	0.1105	-0.0021	0.0769
0.0478	-0.0126	0.4955	-0.0040	0.3562	-0.0109	0.2451	0.0042	0.1668
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3786	0.0055	0.3169	-0.0016	0.2188	-0.0040	0.1516	-0.0038	0.1056
0.3839	0.0041	0.2386	-0.0033	0.1621	-0.0035	0.1118	0.0040	0.0759
0.3883	-0.0062	0.3252	0.0013	0.2185	0.0002	0.1482	0.0007	0.1052
0.3927	-0.0043	0.2917	-0.0002	0.1984	0.0001	0.1338	-0.0027	0.0981
0.3987	-0.0004	0.1940	0.0029	0.1335	0.0003	0.0922	-0.0013	0.0653
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7536	-0.0338	0.2762	-0.0125	0.1900	-0.0006	0.1440	-0.0003	0.1036
0.7676	-0.0071	0.1768	-0.0015	0.1310	0.0009	0.0917	0.0006	0.0641
0.7695	-0.0058	0.1627	0.0029	0.1153	0.0003	0.0867	0.0016	0.0569
0.7705	-0.0476	0.2718	-0.0119	0.1912	-0.0037	0.1433	-0.0051	0.1035
0.7904	-0.0125	0.1693	0.0032	0.1174	0.0011	0.0884	0.0008	0.0618

Table B3: Empirical sizes of tests for individual spatial parameters from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections for different  $N$  and  $T$  combinations, using standard and sandwich formulae for the variances

$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\psi_{i0}$	$N = 25$							
0.0255	0.0545	0.0635	0.0530	0.0455	0.0755	0.0730	0.0535	0.0520
0.0777	0.0835	0.0660	0.0550	0.0465	0.1065	0.0765	0.0600	0.0500
0.1109	0.0885	0.0660	0.0610	0.0510	0.1095	0.0750	0.0580	0.0540
0.1261	0.0730	0.0580	0.0675	0.0440	0.1005	0.0790	0.0780	0.0495
0.1948	0.0910	0.0660	0.0535	0.0565	0.1085	0.0790	0.0575	0.0595
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.3883	0.0755	0.0610	0.0585	0.0580	0.1075	0.0845	0.0710	0.0635
0.3987	0.0750	0.0600	0.0570	0.0480	0.1035	0.0715	0.0635	0.0535
0.4375	0.0705	0.0615	0.0485	0.0570	0.0910	0.0710	0.0570	0.0610
0.4682	0.0700	0.0685	0.0570	0.0480	0.0960	0.0770	0.0625	0.0525
0.5059	0.0805	0.0740	0.0460	0.0460	0.1075	0.0910	0.0505	0.0520
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.6726	0.0545	0.0645	0.0490	0.0640	0.0825	0.0750	0.0585	0.0665
0.6793	0.0605	0.0700	0.0595	0.0525	0.0935	0.0810	0.0665	0.0515
0.7127	0.0205	0.0190	0.0280	0.0350	0.0310	0.0265	0.0305	0.0380
0.7246	0.0275	0.0470	0.0560	0.0515	0.0540	0.0610	0.0690	0.0545
0.7676	0.0290	0.0410	0.0550	0.0595	0.0515	0.0610	0.0655	0.0665
	$N = 50$							
0.0255	0.0547	0.0580	0.0630	0.0530	0.0785	0.0750	0.0720	0.0555
0.0432	0.0818	0.0590	0.0610	0.0525	0.1075	0.0680	0.0670	0.0545
0.0608	0.0512	0.0595	0.0645	0.0565	0.0785	0.0725	0.0745	0.0550
0.0625	0.0783	0.0675	0.0520	0.0495	0.0950	0.0790	0.0620	0.0565
0.0777	0.0783	0.0645	0.0565	0.0435	0.0990	0.0790	0.0650	0.0435
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.3451	0.0587	0.0600	0.0550	0.0545	0.0845	0.0740	0.0610	0.0590
0.3510	0.0858	0.0640	0.0535	0.0560	0.1175	0.0740	0.0595	0.0580
0.3565	0.0778	0.0680	0.0550	0.0535	0.1090	0.0790	0.0605	0.0575
0.3755	0.0748	0.0680	0.0600	0.0540	0.1040	0.0825	0.0690	0.0610
0.3786	0.0723	0.0655	0.0625	0.0525	0.0985	0.0760	0.0735	0.0550
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.7127	0.0231	0.0215	0.0220	0.0260	0.0345	0.0265	0.0240	0.0340
0.7246	0.0351	0.0465	0.0485	0.0435	0.0630	0.0575	0.0600	0.0520
0.7472	0.0296	0.0225	0.0245	0.0185	0.0400	0.0305	0.0290	0.0230
0.7676	0.0296	0.0445	0.0615	0.0485	0.0580	0.0670	0.0745	0.0485
0.7695	0.0411	0.0490	0.0485	0.0590	0.0575	0.0665	0.0510	0.0620
	Continued...							

Notes: Nominal size is set to 5%. The standard and sandwich formulae are given by (40) and (41), respectively. See the notes to Table B2 for details of the data generating process. Size is computed under  $H_0: \psi_i = \psi_{i0}$ , using a two-sided alternative where  $\psi_{i0}$  takes values in the range  $[0.0, 0.8]$  for  $i = 1, 2, \dots, N$ . The number of replications is set to  $R = 2000$ . Estimates are sorted in ascending order according to the true values of the spatial autoregressive parameters.

Table B3: ...Continued

$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\psi_{i0}$	$N = 75$							
0.0255	0.0642	0.0555	0.0535	0.0475	0.0855	0.0685	0.0550	0.0505
0.0397	0.0843	0.0620	0.0490	0.0595	0.1065	0.0755	0.0545	0.0645
0.0432	0.0777	0.0510	0.0530	0.0610	0.1170	0.0680	0.0595	0.0690
0.0478	0.0667	0.0595	0.0565	0.0575	0.0890	0.0750	0.0615	0.0595
0.0608	0.0386	0.0580	0.0550	0.0545	0.0645	0.0690	0.0590	0.0585
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3755	0.0617	0.0615	0.0680	0.0490	0.0950	0.0830	0.0775	0.0510
0.3786	0.0757	0.0670	0.0545	0.0505	0.0945	0.0840	0.0600	0.0570
0.3883	0.0787	0.0655	0.0635	0.0495	0.1090	0.0770	0.0740	0.0550
0.3927	0.0657	0.0550	0.0580	0.0470	0.1045	0.0675	0.0585	0.0485
0.3987	0.0777	0.0690	0.0540	0.0560	0.0990	0.0895	0.0595	0.0580
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7472	0.0266	0.0280	0.0245	0.0220	0.0370	0.0335	0.0275	0.0250
0.7512	0.0176	0.0250	0.0555	0.0490	0.0430	0.0360	0.0585	0.0545
0.7536	0.0226	0.0250	0.0235	0.0490	0.0310	0.0290	0.0335	0.0530
0.7676	0.0271	0.0430	0.0635	0.0460	0.0565	0.0720	0.0705	0.0495
0.7695	0.0351	0.0520	0.0610	0.0535	0.0575	0.0625	0.0640	0.0570
	$N = 100$							
0.0244	0.0884	0.0585	0.0505	0.0500	0.1040	0.0730	0.0560	0.0535
0.0255	0.0588	0.0570	0.0535	0.0550	0.0935	0.0705	0.0635	0.0585
0.0397	0.0773	0.0675	0.0550	0.0550	0.0985	0.0805	0.0590	0.0600
0.0432	0.0773	0.0605	0.0580	0.0550	0.1000	0.0720	0.0680	0.0560
0.0478	0.0613	0.0665	0.0525	0.0505	0.0820	0.0815	0.0520	0.0530
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.3786	0.0668	0.0685	0.0565	0.0525	0.0860	0.0800	0.0620	0.0565
0.3839	0.0784	0.0630	0.0555	0.0480	0.1075	0.0780	0.0565	0.0515
0.3883	0.0713	0.0620	0.0550	0.0485	0.1045	0.0790	0.0620	0.0505
0.3927	0.0658	0.0625	0.0490	0.0565	0.0865	0.0700	0.0600	0.0590
0.3987	0.0733	0.0680	0.0525	0.0545	0.1020	0.0840	0.0590	0.0565
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.7536	0.0266	0.0215	0.0295	0.0530	0.0330	0.0295	0.0365	0.0540
0.7676	0.0256	0.0430	0.0565	0.0490	0.0415	0.0600	0.0630	0.0515
0.7695	0.0417	0.0445	0.0665	0.0505	0.0675	0.0650	0.0745	0.0540
0.7705	0.0206	0.0205	0.0260	0.0590	0.0380	0.0310	0.0325	0.0585
0.7904	0.0286	0.0310	0.0550	0.0455	0.0525	0.0485	0.0615	0.0475

Table B4: Empirical sizes of tests for individual slope parameters from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections for different  $N$  and  $T$  combinations, using standard and sandwich formulae for the variances

$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\beta_{i0}$	$N = 25$							
0.0344	0.0700	0.0550	0.0525	0.0510	0.1040	0.0730	0.0615	0.0595
0.0357	0.0775	0.0635	0.0540	0.0485	0.1050	0.0800	0.0590	0.0550
0.0462	0.0820	0.0615	0.0590	0.0565	0.1120	0.0805	0.0685	0.0605
0.1190	0.0780	0.0615	0.0570	0.0580	0.1080	0.0775	0.0655	0.0595
0.1419	0.0710	0.0585	0.0560	0.0535	0.0965	0.0785	0.0645	0.0565
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.5472	0.0890	0.0635	0.0515	0.0530	0.1175	0.0750	0.0620	0.0560
0.6787	0.0730	0.0535	0.0520	0.0550	0.0935	0.0685	0.0595	0.0545
0.6797	0.0770	0.0690	0.0560	0.0575	0.0975	0.0790	0.0620	0.0565
0.6948	0.0750	0.0765	0.0605	0.0460	0.1035	0.0925	0.0685	0.0520
0.6991	0.0630	0.0520	0.0510	0.0540	0.0895	0.0700	0.0585	0.0620
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.8143	0.0700	0.0745	0.0540	0.0640	0.1005	0.0915	0.0615	0.0670
0.8147	0.0675	0.0545	0.0575	0.0540	0.0895	0.0690	0.0690	0.0550
0.9134	0.0780	0.0640	0.0535	0.0500	0.1055	0.0755	0.0565	0.0570
0.9572	0.0655	0.0535	0.0580	0.0520	0.0900	0.0650	0.0605	0.0540
0.9649	0.0870	0.0650	0.0545	0.0540	0.1110	0.0815	0.0620	0.0610
	$N = 50$							
0.0344	0.0787	0.0695	0.0515	0.0450	0.1015	0.0855	0.0560	0.0535
0.0357	0.0757	0.0635	0.0485	0.0540	0.1025	0.0785	0.0575	0.0585
0.0462	0.0822	0.0665	0.0515	0.0555	0.1100	0.0795	0.0580	0.0590
0.0759	0.0898	0.0560	0.0620	0.0475	0.1140	0.0705	0.0670	0.0520
0.0838	0.0712	0.0620	0.0515	0.0660	0.0970	0.0760	0.0595	0.0730
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.5472	0.0827	0.0580	0.0580	0.0510	0.1035	0.0750	0.0640	0.0560
0.5499	0.0762	0.0605	0.0560	0.0550	0.0985	0.0760	0.0645	0.0580
0.5688	0.0777	0.0610	0.0560	0.0585	0.1025	0.0790	0.0620	0.0620
0.6020	0.0732	0.0585	0.0560	0.0555	0.1115	0.0770	0.0665	0.0635
0.6160	0.0772	0.0605	0.0555	0.0525	0.1095	0.0710	0.0620	0.0575
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
0.9134	0.0712	0.0605	0.0560	0.0490	0.0915	0.0750	0.0630	0.0500
0.9172	0.0757	0.0600	0.0580	0.0460	0.1110	0.0755	0.0600	0.0505
0.9572	0.0827	0.0620	0.0490	0.0525	0.1105	0.0715	0.0580	0.0565
0.9649	0.0827	0.0600	0.0515	0.0560	0.1120	0.0715	0.0580	0.0575
0.9961	0.0732	0.0675	0.0515	0.0550	0.0980	0.0805	0.0555	0.0620
	Continued...							

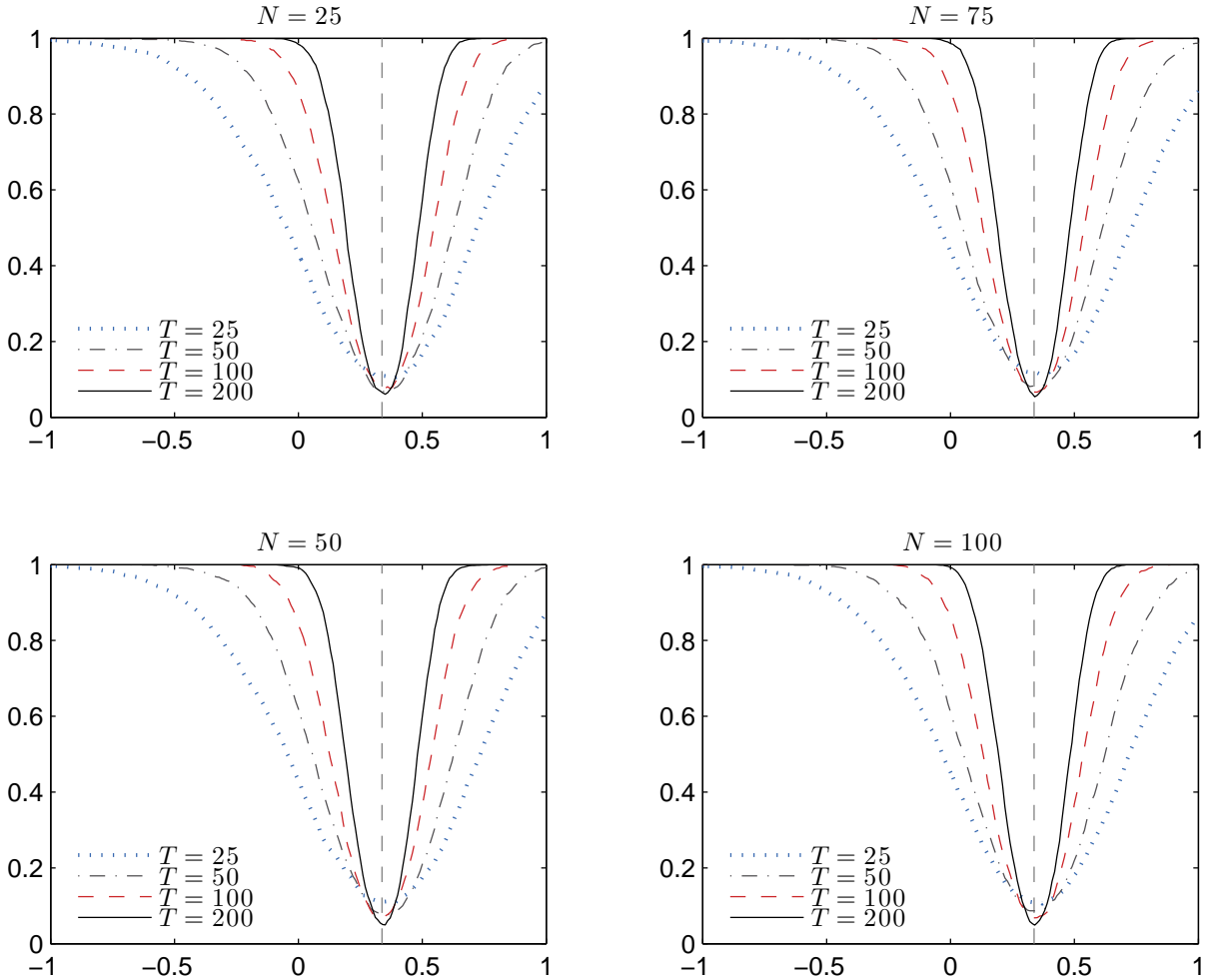
Notes: Nominal size is set to 5%. The standard and sandwich formulae are given by (40) and (41), respectively. See the notes to Table B2 for details of the data generating process. Size is computed under  $H_0: \beta_i = \beta_{i0}$ , using a two-sided alternative where  $\beta_{i0}$  takes values in the range  $[0.0, 1.0]$  for  $i = 1, 2, \dots, N$ . The number of replications is set to  $R = 2000$ . Estimates are sorted in ascending order according to the true values of the spatial autoregressive parameters.



Table B4: ...Continued

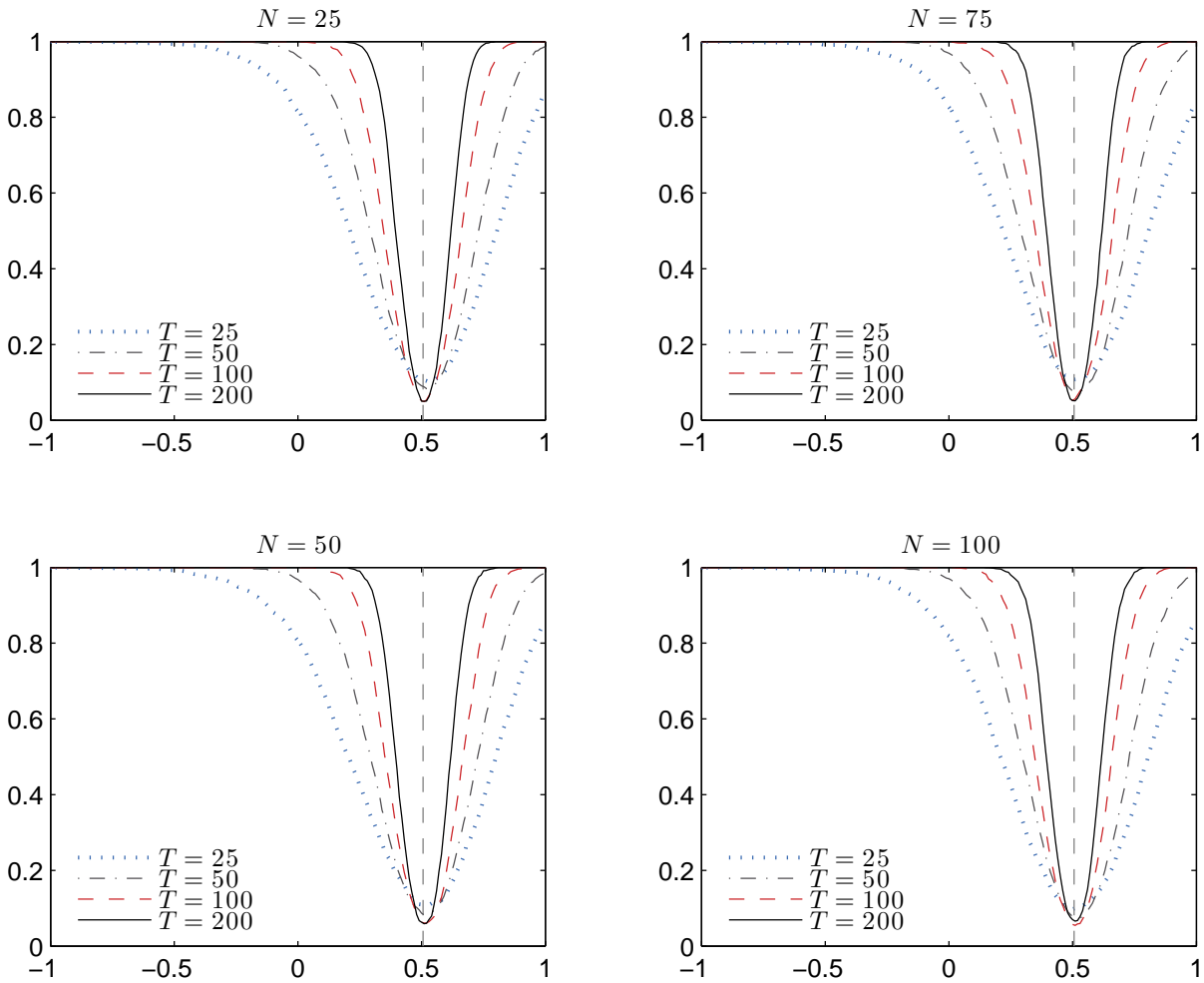
$T$	Empirical sizes: Standard				Empirical sizes: Sandwich			
	25	50	100	200	25	50	100	200
$\beta_{i0}$	$N = 75$							
0.0344	0.0902	0.0620	0.0565	0.0520	0.1255	0.0785	0.0670	0.0555
0.0357	0.0842	0.0675	0.0595	0.0490	0.1110	0.0750	0.0730	0.0495
0.0430	0.0822	0.0570	0.0550	0.0570	0.1050	0.0665	0.0630	0.0650
0.0462	0.0746	0.0630	0.0545	0.0420	0.1015	0.0715	0.0620	0.0455
0.0759	0.0812	0.0680	0.0530	0.0560	0.1050	0.0775	0.0615	0.0620
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.4018	0.0797	0.0655	0.0545	0.0565	0.1020	0.0810	0.0630	0.0630
0.4173	0.0822	0.0615	0.0555	0.0570	0.1050	0.0735	0.0635	0.0645
0.4302	0.0847	0.0680	0.0520	0.0545	0.1105	0.0805	0.0585	0.0530
0.4357	0.0887	0.0660	0.0505	0.0540	0.1110	0.0810	0.0555	0.0600
0.4868	0.0746	0.0665	0.0590	0.0525	0.1065	0.0810	0.0690	0.0505
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9172	0.0767	0.0605	0.0630	0.0525	0.1105	0.0710	0.0760	0.0585
0.9448	0.0762	0.0555	0.0585	0.0480	0.0980	0.0705	0.0650	0.0545
0.9572	0.0767	0.0595	0.0570	0.0505	0.1035	0.0680	0.0670	0.0535
0.9649	0.0822	0.0580	0.0525	0.0520	0.1055	0.0830	0.0630	0.0590
0.9961	0.0792	0.0615	0.0585	0.0460	0.0995	0.0765	0.0595	0.0485
	$N = 100$							
0.0292	0.0732	0.0655	0.0575	0.0585	0.0940	0.0775	0.0650	0.0620
0.0344	0.0812	0.0670	0.0490	0.0480	0.1140	0.0810	0.0520	0.0515
0.0357	0.0757	0.0560	0.0640	0.0525	0.1040	0.0720	0.0740	0.0575
0.0430	0.0817	0.0630	0.0515	0.0520	0.1000	0.0745	0.0560	0.0580
0.0462	0.0747	0.0695	0.0565	0.0525	0.1100	0.0870	0.0620	0.0615
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.4173	0.0752	0.0610	0.0585	0.0570	0.1025	0.0730	0.0695	0.0615
0.4302	0.0838	0.0630	0.0515	0.0580	0.1055	0.0770	0.0605	0.0610
0.4357	0.0747	0.0590	0.0475	0.0565	0.1010	0.0765	0.0540	0.0600
0.4588	0.0647	0.0630	0.0560	0.0510	0.0890	0.0825	0.0595	0.0560
0.4868	0.0742	0.0700	0.0525	0.0515	0.1145	0.0885	0.0590	0.0555
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.9448	0.0742	0.0610	0.0490	0.0600	0.1055	0.0685	0.0545	0.0635
0.9572	0.0787	0.0595	0.0550	0.0515	0.1135	0.0755	0.0600	0.0525
0.9649	0.0737	0.0615	0.0575	0.0575	0.1075	0.0740	0.0655	0.0640
0.9797	0.0757	0.0575	0.0470	0.0535	0.1025	0.0790	0.0525	0.0585
0.9961	0.0672	0.0620	0.0605	0.0510	0.0970	0.0740	0.0640	0.0580

Figure B1: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.3374$  from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



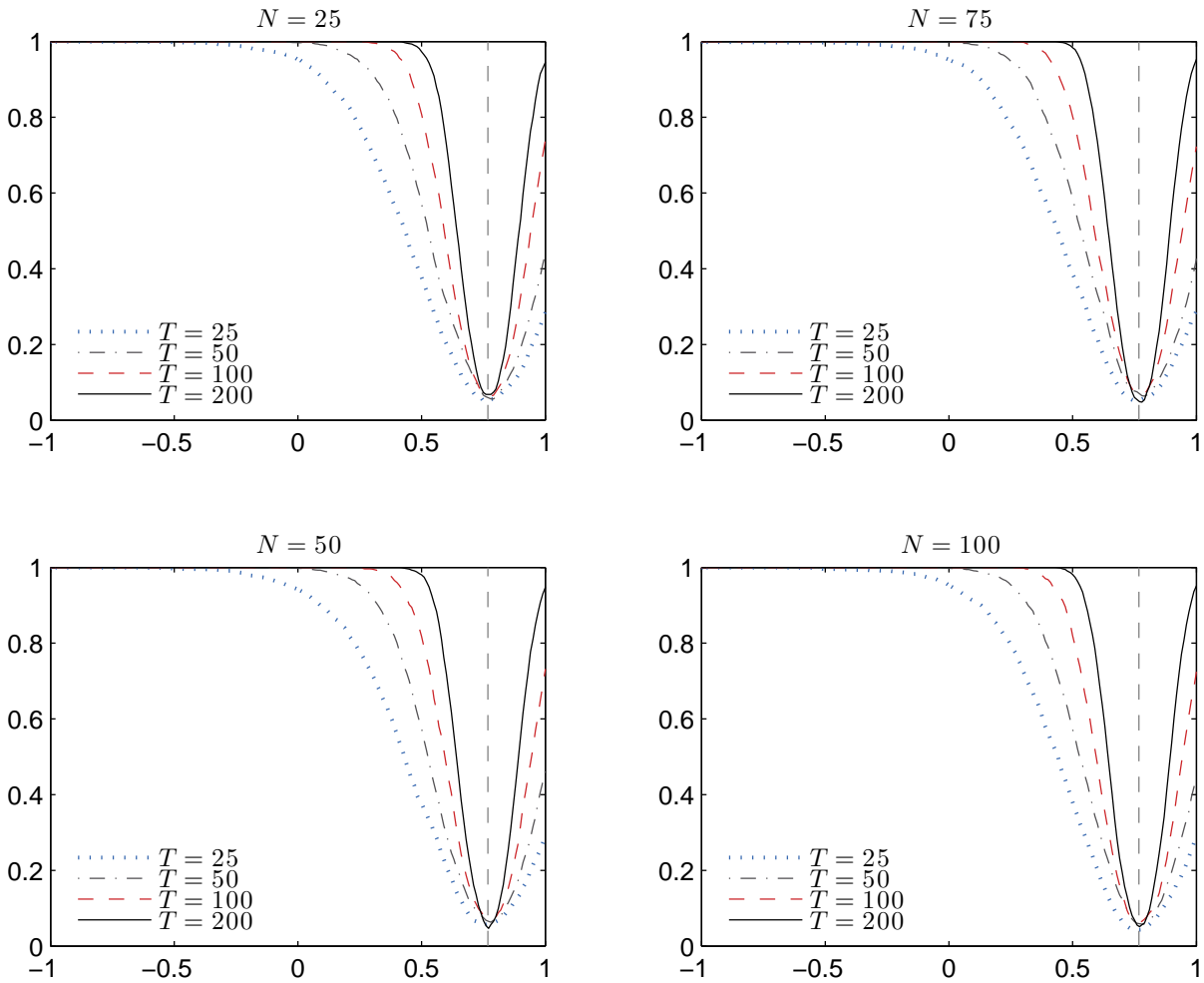
Notes: The power functions are based on the sandwich formula given by (41). See the notes to Table B2 for details of the data generating process. Power is computed under  $\psi_i = \psi_{i0} + \delta$ , where  $\delta = -0.8, -0.791, \dots, 0.791, 0.8$  or until the parameter space boundaries of -1 and 1 are reached. The number of replications is set to  $R = 2000$ .

Figure B2: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.5059$  from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



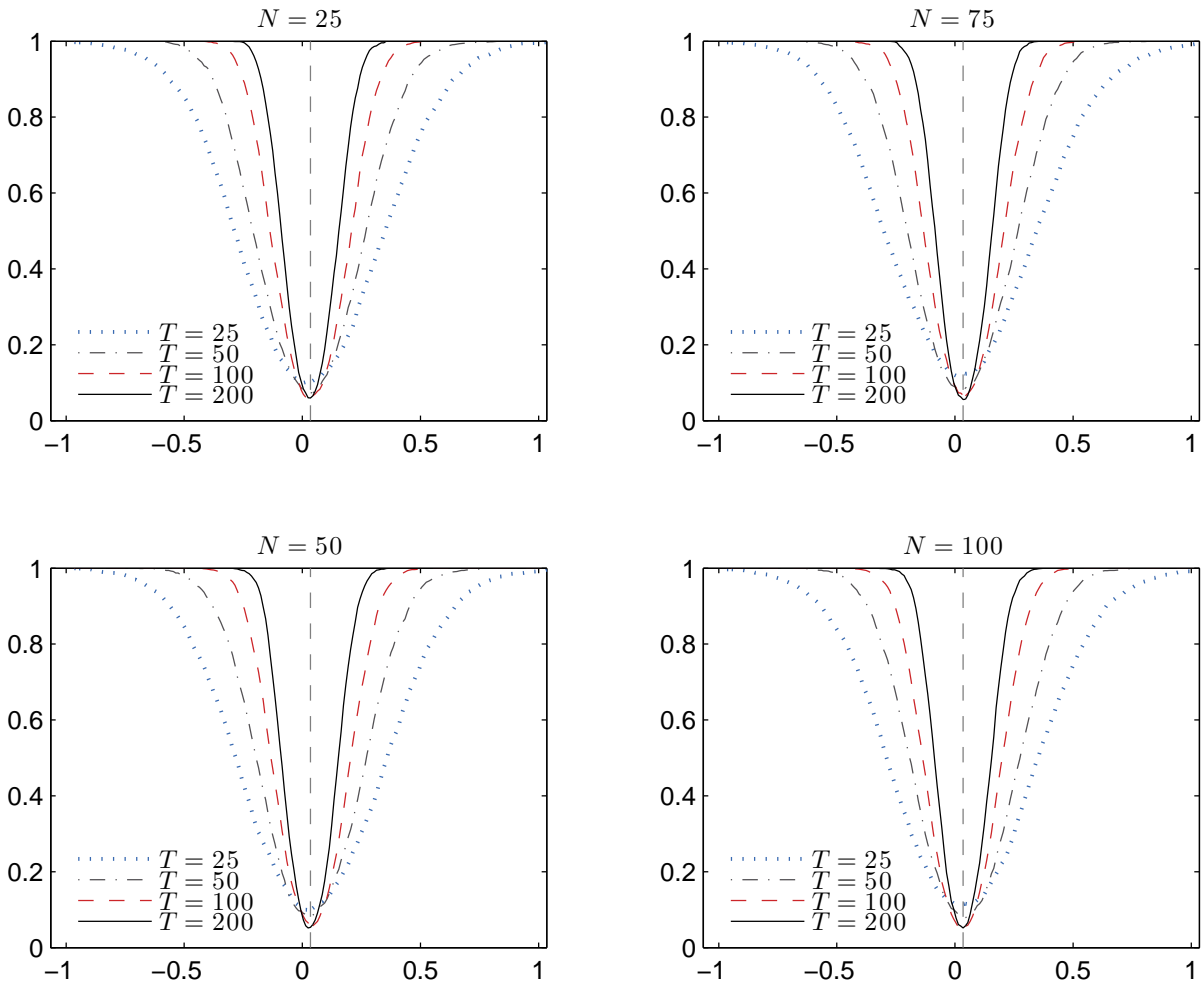
Notes: See the notes to Figure B1.

Figure B3: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the spatial parameter value  $\psi_{i0} = 0.7676$  from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



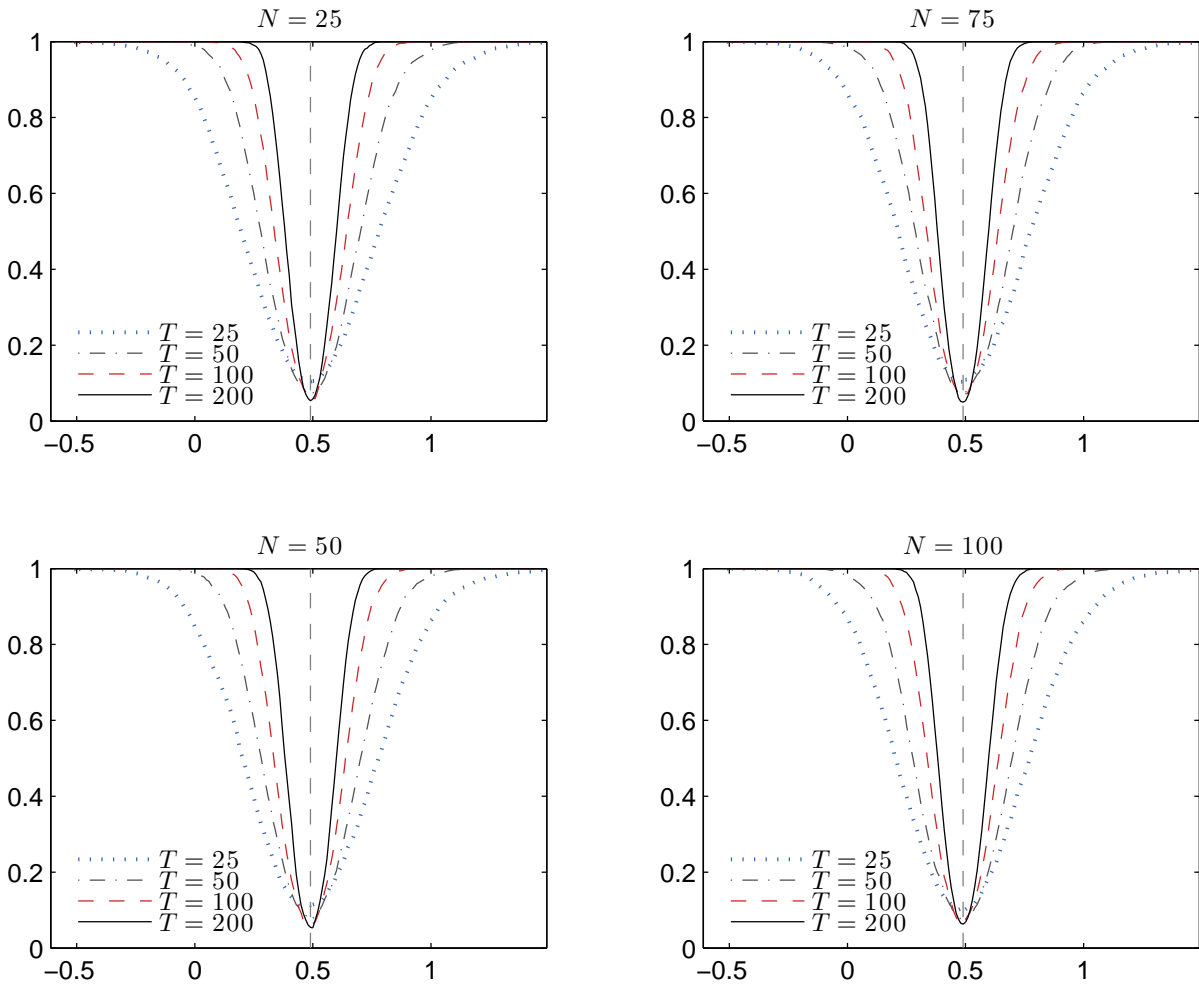
Notes: See the notes to Figure B1.

Figure B4: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the slope parameter value  $\beta_{i0} = 0.0344$  from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



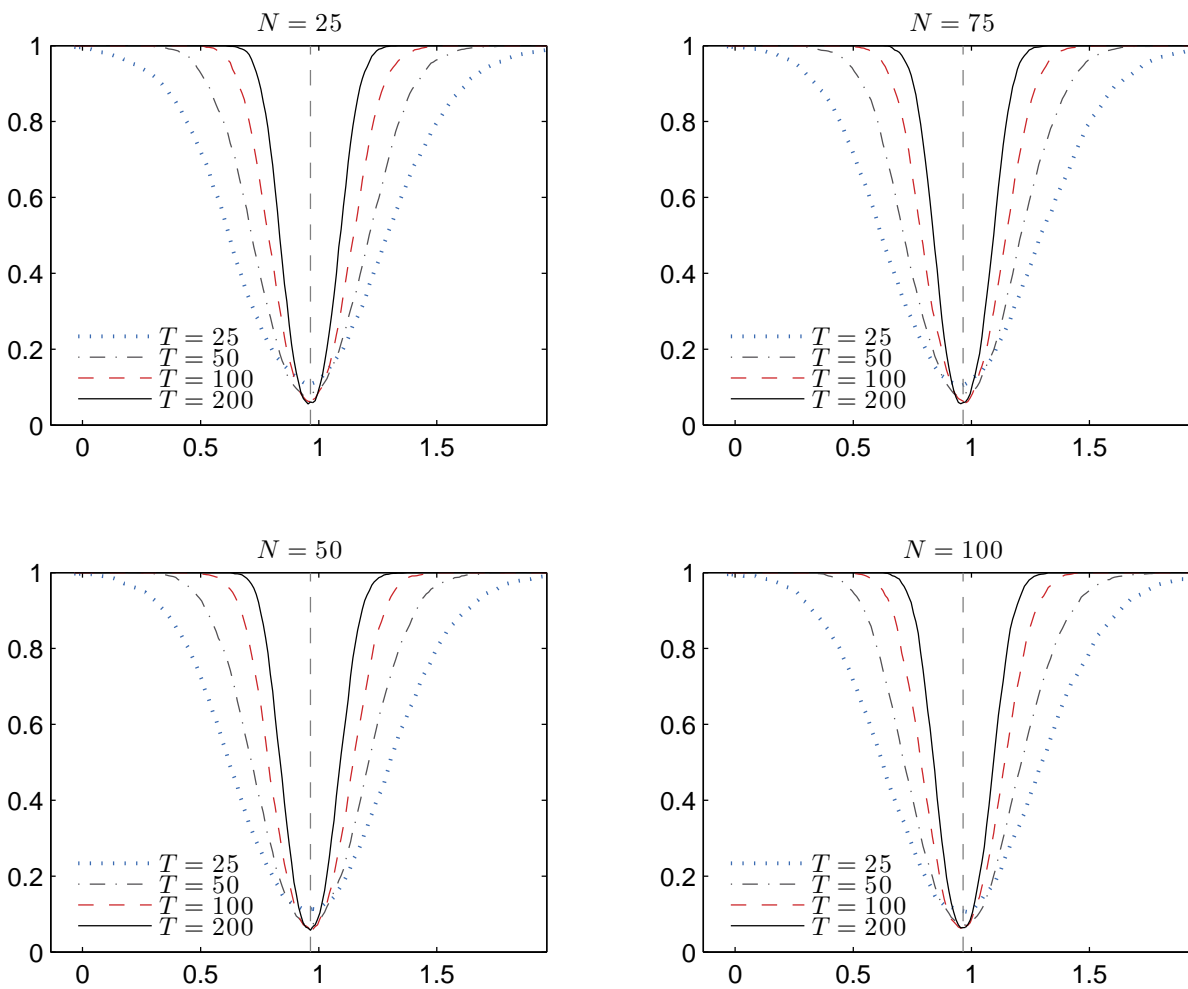
Notes: The power functions are based on the sandwich formula given by (41). See the notes to Table B2 for details of the data generating process. Power is computed under  $\beta_i = \beta_{i0} + \delta$ , where  $\delta = -1.0, -0.991, \dots, 0.991, 1.0$ . The number of replications is set to  $R = 2000$ .

Figure B5: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the slope parameter value  $\beta_{i0} = 0.4898$  from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure B4.

Figure B6: Empirical power functions for different  $N$  and  $T$  combinations, associated with testing the slope parameter value  $\beta_{i0} = 0.9649$  from HSAR(1) model with Gaussian errors, one exogenous regressor and spatial weight matrix  $\mathbf{W}$  having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure B4.