CESifo WORKING PAPERS

7542
2019

March 2019

Estimation and inference for spatial models with heterogeneous coefficients: an application to U.S. house prices

Michele Aquaro, Natalia Bailey, M. Hashem Pesaran

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

CESifo *Working Paper No. 7542*
Category 12: Empirical and Theoretical Methods

Estimation and inference for spatial models with heterogeneous coefficients: an application to U.S. house prices

Abstract

This paper considers the problem of identification, estimation and inference in the case of spatial panel data models with heterogeneous spatial lag coefficients, with and without (weakly) exogenous regressors, and subject to heteroskedastic errors. A quasi maximum likelihood (QML) estimation procedure is developed and the conditions for identification of spatial coefficients are derived. Regularity conditions are established for the QML estimators of individual spatial coefficients, as well as their means (the mean group estimators), to be consistent and asymptotically normal. 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 degrees of sparsity. The simulation results are in line with the paper's key theoretical findings even for panels with moderate time dimensions, irrespective of the number of cross section units. An empirical application to U.S. house price changes during the 1975-2014 period shows a significant degree of heterogeneity in spill-over effects over the 338 Metropolitan Statistical Areas considered.

JEL-Codes: C210, C230.

Keywords: spatial panel data models, heterogeneous spatial lag coefficients, identification, quasi maximum likelihood (QML) estimators, non-Gaussian errors, house price changes, Metropolitan Statistical Areas.

Michele Aquaro
Joint Research Centre
European Commission
M.Aquaro@warwick.ac.uk

Natalia Bailey
Department of Econometrics & Business Statistics,
Monash University
Clayton / Victoria / Australia
natalia.bailey@monash.edu

M. Hashem Pesaran
Department of Economics
University of Southern California / Los Angeles / CA / USA
pesaran@usc.edu

February 23, 2019
The authors would like to acknowledge helpful comments from Alex Chudik, Bernard Fingleton, Harry Kelejian, Ron Smith and Cynthia Fan Yang. Financial support under ESRC Grant ES/I031626/1 is also gratefully acknowledged. All errors and omissions are our own. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of this paper. *A preliminary version of this paper appeared under the title "Quasi-maximum likelihood estimation of spatial models with heterogeneous coefficient", CESifo Working Paper No.5428. This paper is a substantially revised version and supersedes Aquaro et al. (2015).*

1 Introduction

This paper considers a heterogeneous version of the standard spatial autoregressive (SAR) panel data model whereby the spatial lag coefficients are allowed to differ over the cross section units. We refer to this generalized specification as the heterogeneous SAR (or HSAR) model. The model also features weakly exogenous regressors, possible fixed effects and heteroskedastic error variances, and provides a reasonably general framework for the analysis of heterogeneous interactions, where it is important to distinguish between the average intensity of spillover effects as characterized by standard spatial models, and the heterogeneity of such effects over different geographical units such as counties, regions or countries. Importantly, the framework studied in this paper allows for spatial dependence directly through contemporaneous dependence of individual units on their connections, and indirectly through possible cross-sectional dependence in the regressors. The econometric analysis of HSAR models presents new technical difficulties, both for identification and estimation of a large set of spatial lag coefficients that must be simultaneously estimated.

Our analysis builds on the existing literature on SAR models, pioneered by Whittle (1954) and Cliff and Ord (1973), and further advanced in a number of important directions. The maximum likelihood approach of Cliff and Ord, which was developed for a pure spatial model, has been extended to cover panel data models with fixed effects and dynamics. Other estimation and testing techniques, such as the generalized method of moments (GMM), also have been proposed. Some of the key references to this 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), Yu et al. (2008) and Yu et al. (2012). Spatial techniques also have proved useful when analyzing network effects as can be seen in the pioneering work of Case (1991) and Manski (1993).

Almost all these contributions (whether in the context of spatial or network models) assume that, apart from possibly fixed effects, spatial spill-over or network effects are homogeneous. However, even if all units in a network have the exact same number of connections, it can be the case that not all units are equally important or influential. Therefore, the assumption of a homogeneous spatial coefficient, though needed when analyzing pure spatial models or spatial panel data models with a short time dimension (T), is likely to be restrictive, and its validity should at least be tested. Also when T is large, the HSAR model can be estimated for any N and it is not required that $N \to \infty$, which is clearly needed when T is small. Examples of such data sets include large panels that cover counties, regions, 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 U.S., it is interesting to investigate whether the maintained assumption that spillover effects from neighboring states are the same across all the 48 mainland states in fact holds, particularly considering the large size of the U.S. and the uneven distribution of economic activity across it.

Whilst estimation of HSAR panel data models can be carried out using MLE and GMM approaches, in this paper we focus on the former and discuss identification, estimation and inference using the quasi maximum likelihood (QML) method. We derive conditions under which the QML estimators of the *individual* parameters are locally identified, and establish consistency and asymptotic normality of the estimators under certain regularity conditions. Asymptotic covariance matrices of the QML estimators are derived under both Gaussian and non-Gaussian errors, and consistent estimators of these covariances are proposed. Alternative estimation methods based on our HSAR model include the Bayesian Markov Chain Monte Carlo approach of LeSage and Chih (2018) and the generalized Yule-Walker estimation method of Dou et al. (2016).

We also propose an estimator of the cross section mean of the individual parameters (also known in the literature as the Mean Group, MG, estimator) assuming a random coefficient model. It is shown that MG estimators are consistent and asymptotically normal for N and $T \to \infty$, jointly, so long as $\sqrt{N}/T \to 0$, and the spatial dependence is sufficiently weak. Such estimators are helpful in two respects. They provide an overall average estimator of the spatial effects that could be compared to corresponding estimates obtained using standard homogeneous SAR models. They can also be used to obtain average estimators across sub-spatial groupings such as states or regions, or sub-groups within a production or financial network, such as industry types. Following Pesaran et al. (1996), the deviations of the MG estimators from their homogenous counterparts can be used to develop simple Hausman type tests of the homogeneity of the underlying individual parameters.

The small sample performance of the QML estimator is investigated by Monte Carlo simulations for different values of N and T and alternative 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. We also investigate the small sample performance of the MG estimator and find its performance to be satisfactory with biases that are universally negligible, and RMSEs that decline with T and quite rapidly with N. Regarding size and power, tests based on the MG estimator exhibit some downward size distortions when T is small, but such distortions disappear as T rises for all values of N. The small sample bias of the MG estimator can be reduced using half-Jackknife procedure as discussed in Chudik and Pesaran (2019).

As an empirical application, we fit HSAR models to U.S. quarterly house price changes over the period 1975Q1-2014Q4 observed at Metropolitan Statistical Areas (MSAs). Not surprisingly we find a considerable degree of heterogeneity across the MSA specific estimates. As to be expected, with a few exceptions, the estimates of spatial coefficients are positive and statistically significant, suggesting a high degree of spill-over effects of house price changes to neighboring MSAs. There were 11 MSA (out of 338 considered in our analysis) with statistically significant negative spatial effects, and included Cheyenne (Wyoming), Coeur d'Alene (Idaho) or Hot Springs (Arkansas). These MSAs tend to be relatively remote with outward migratory flows to neighboring regions.

We also consider MG estimators obtained for six U.S. regions based on the individual MSA level estimates. Spatial parameter estimates are positive and statistically significant for all regions. The average estimate of the spatial lag obtained for the U.S. (around 0.51) is lower than the estimate of around 0.65 reported by Yang (2018) who considers a homogeneous SAR specification estimated on a similar data set. The differences between the two estimates could be due to the considerable degree of heterogeneity that we observe across the regions in the U.S., which is being neglected under the homogeneity assumption. We also find positive and statistically significant effects of population and income growth on house price changes, again with a high degree of heterogeneity across the regions.

The rest of the paper is organized as follows: Section 2 introduces the first order spatial autoregressive model with heterogeneous coefficients and some useful generalizations, and derives its log-likelihood function. Section 3 sets out the assumptions of the model, derives the identification conditions and proves consistency and asymptotic normality of the QML estimator when the time dimension is large. Section 4 outlines the Mean Group estimator derived from the heterogeneous spatial coefficients of the HSAR model. Section 5 presents the Monte Carlo design and reports small sample results (bias, root mean square errors, size and power) of the QML and MG estimators for different parameter values and sample size combinations. Section 6 reports the results of our empirical application to the U.S. house price changes across MSAs. Some concluding remarks are provided in Section 7. Mathematical proofs, data sources and additional Monte Carlo results are provided in an online supplement.

Notations: We denote the largest and the smallest eigenvalues of the $N \times N$ matrix $\boldsymbol{A} = (a_{ij})$ by $\lambda_{\max}(\boldsymbol{A})$ and $\lambda_{\min}(\boldsymbol{A})$, respectively, its trace by $\mathrm{tr}(\boldsymbol{A}) = \sum_{i=1}^{N} a_{ii}$, its spectral radius by $\rho(\boldsymbol{A}) = |\lambda_{\max}(\boldsymbol{A})|$, its spectral norm by $\|\boldsymbol{A}\| = \lambda_{\max}^{1/2}(\boldsymbol{A}'\boldsymbol{A})$, its maximum absolute column sum norm by $\|\boldsymbol{A}\|_1 = \max_{1\leq j\leq N}\left(\sum_{i=1}^{N}|a_{ij}|\right)$, and its maximum absolute row sum norm by $\|\boldsymbol{A}\|_\infty = \max_{1\leq i\leq N}\left(\sum_{j=1}^{N}|a_{ij}|\right)$. $\mathrm{Diag}(\boldsymbol{A}) = \mathrm{Diag}(a_{11}, a_{22}, \ldots, a_{NN})$ represents an $N \times N$ diagonal matrix formed by the diagonal elements of $\boldsymbol{A}$, while $\mathrm{diag}(\boldsymbol{A}) = (a_{11}, a_{22}, \ldots, a_{NN})'$ denotes an $N \times 1$ vector. We denote the ℓ_p-norm of the random variable x by $\|x\|_p = E(|x|^p)^{1/p}$ for $p \geq 1$, assuming that $E(|x|^p) < K$. $\odot$ stands for Hadamard product or element-wise matrix product operator, $\rightarrow_p$ denotes convergence in probability, $\overset{a.s.}{\rightarrow}$ almost sure convergence, $\rightarrow_d$ convergence in distribution, and $\overset{a}{\sim}$ asymptotic equivalence in distribution. Asymptotics for estimation of individual parameters are carried out for N finite and as $T \to \infty$. K and c will be used to denote finite large and non-zero small positive numbers, respectively, that do not depend on N and T.

2 A heterogeneous spatial autoregressive model (HSAR)

2.1 Model specification

We consider the following SAR model with heterogeneous slopes:

$$y_{it} = \psi_{i0}\left(\sum_{j=1}^{N} w_{ij} y_{jt}\right) + \boldsymbol{\beta}_{i0}'\boldsymbol{x}_{it} + \varepsilon_{it}, \text{ for } i = 1, 2, \ldots, N;\ t = 1, 2, \ldots, T, \tag{1}$$

where y_{it} is the dependent variable for unit i observed at time t, $\boldsymbol{x}_{it} = (x_{i1,t}, x_{i2,t}, \ldots, x_{ik,t})'$ is a $k \times 1$ vector of exogenous regressors, with the associated $k \times 1$ vector of slope parameters, $\boldsymbol{\beta}_{i0} = (\beta_{i1,0}, \beta_{i2,0}, \ldots, \beta_{ik,0})'$. ε_{it} is the unexplained component of y_{it}, which we refer to as the error of the i^{th} cross section unit, or the 'error' for short. Finally, $y_{it}^* = \sum_{j=1}^{N} w_{ij} y_{jt} = \mathbf{w}_i'\mathbf{y}_t$ is the average effect of other units on unit i at time t, where $\mathbf{y}_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$ and $\mathbf{w}_i'$ is the

i^{th} row of the $N \times N$ spatial weight matrix, $\mathbf{W} = (w_{ij})$, with w_{ij}, for $i, j = 1, 2, \ldots, N$ being the spatial weights. Without loss of generality we set $w_{ii} = 0$, for all i, assume that $w_{ij} \geq 0$, and normalize the spatial weights so that $\sum_{j=1}^{N} w_{ij} = 1$.[1] When the weights are not normalized, (1) continues to hold with ψ_{i0} re-defined as ψ_{i0}/v_i, where $\sum_{j=1}^{N} w_{ij} = v_i$. Consequently, in the heterogeneous case the normalization of the weights is innocuous, and can be viewed as an identifying restriction, so that ψ_{i0} can be distinguished from v_i, which is achieved by setting $v_i = 1$. The same is not true in the homogeneous case where $\psi_{i0} = \psi_0$ for all i, and the use of non-normalized weights is equivalent to setting $\psi_{i0} = \psi_0/v_i$, which is not an innocuous restriction. The HSAR model (1) can also be viewed as a generalization of the random coefficient panel data model reviewed, for example, by Hsiao and Pesaran (2008). However, this is not a straightforward generalization due to the endogeneity of $y_{it}^{*} = \mathbf{w}_i' \mathbf{y}_t$ in (1).

The assumption of non-negative weights ($w_{ij} \geq 0$) can be relaxed by replacing $\mathbf{W}$ with two weight matrices: one for positive weights, $\mathbf{W}^{+} = (w_{ij}^{+})$, where $w_{ij}^{+} = w_{ij}$ if $w_{ij} > 0$ and zero otherwise, and one for negative weights, $\mathbf{W}^{-} = (w_{ij}^{-})$, where $w_{ij}^{-} = -w_{ij}$ if $w_{ij} < 0$ and zero otherwise. Then (1) can be written more generally as

$$y_{it} = \psi_{i0}^{+} \left(\sum_{j=1}^{N} w_{ij}^{+} y_{jt} \right) + \psi_{i0}^{-} \left(\sum_{j=1}^{N} w_{ij}^{-} y_{jt} \right) + \boldsymbol{\beta}_{i0}' \boldsymbol{x}_{it} + \varepsilon_{it}, \tag{2}$$

where ψ_{i0}^{+} and ψ_{i0}^{-} measure the effects of positively and negatively connected units on y_{it}, and allowed to vary across units. For an empirical application of such a setting see Bailey et al. (2016).[2] The HSAR model can also be generalized further by estimating the weights, w_{ij}, so long as each unit has a finite number of known neighbors. In such a setting the HSAR model can be written as

$$y_{it} = \sum_{j=1}^{N} \psi_{ij0} I\left(w_{ij}\right) y_{jt} + \boldsymbol{\beta}_{i0}' \boldsymbol{x}_{it} + \varepsilon_{it}, \tag{3}$$

where $I\left(w_{ij}\right) = 1$ if $w_{ij} \neq 0$ and 0 otherwise, and $\sup_i \sum_{j=1}^{N} \left|\psi_{ij0}\right| I\left(w_{ij}\right) < K$. This specification only exploits the qualitative information contained in $I\left(w_{ij}\right)$ and represents another important generalization of the homogeneous spatial model. In what follows we focus on the basic HSAR specification given by (1) and note that estimation and inference for models (2) and (3) can be conducted along the lines set out in this paper.

Stacking the observations by the N individual units for each time period t, (1) can be written more compactly as

$$\left(\boldsymbol{I}_N - \boldsymbol{\Psi}_0 \boldsymbol{W}\right) \boldsymbol{y}_{\circ t} = \boldsymbol{B}_0 \boldsymbol{x}_{\circ t} + \boldsymbol{\varepsilon}_{\circ t},\ t = 1, 2, \ldots, T, \tag{4}$$

where $\boldsymbol{y}_{\circ t} = (y_{1t}, y_{2t}, \ldots, y_{Nt})'$, $\boldsymbol{I}_N$ is an $N \times N$ identity matrix, $\boldsymbol{\Psi}_0 = \mathrm{Diag}\left(\boldsymbol{\psi}_0\right)$ with $\boldsymbol{\psi}_0 =$

[1] Strictly speaking, the weights, w_{ij}, are N-dependent and should be denoted as $w_{ij,N}$. The same also applies to y_{it}, β_{i0}, and ε_{it}. But we abstract from including the subscript N when denoting w_{ij}, y_{it} and ε_{it}, to keep the notations simple and manageable.

[2] It is also possible to allow for spatial effects in the errors and the regressors. For example, ε_{it} can be replaced by $\varepsilon_{it} = \varphi_{i0}\left(\sum_{j=1}^{N} w_{\varepsilon,ij} \varepsilon_{jt}\right) + \nu_{it}$, and the regressors augmented with spatial effects such as $x_{i\ell,t}^{*} = \sum_{j=1}^{N} w_{\ell,ij} x_{j\ell,t}$, for $\ell = 1, 2, \ldots, k$, where $w_{\varepsilon,ij}$ and $w_{\ell,ij}$ are the spatial weights. To simplify the exposition in this paper we abstract from spatial error and regressor processes and focus on the contemporaneous spatial effects in the dependent variable, y_{it}.

$(\psi_{10}, \psi_{20}, \ldots, \psi_{N0})'$, and $\boldsymbol{B}_0$ is the $N \times kN$ block diagonal matrix

$$\boldsymbol{B}_0 = \begin{pmatrix} \boldsymbol{\beta}_{10}' & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\beta}_{20}' & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\beta}_{N-1,0}' & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{\beta}_{N,0}' \end{pmatrix}, \tag{5}$$

and $\boldsymbol{x}_{\circ t} = (\boldsymbol{x}_{1t}', \boldsymbol{x}_{2t}', \ldots, \boldsymbol{x}_{Nt}')'$ is the $kN \times 1$ vector of observations on the exogenous regressors. Finally, $Var\,(\boldsymbol{\varepsilon}_{\circ t}) = \boldsymbol{\Sigma}_0 = \text{Diag}\left(\boldsymbol{\sigma}_0^2\right)$, with $\boldsymbol{\sigma}_0^2 = (\sigma_{10}^2, \sigma_{20}^2, \ldots, \sigma_{N0}^2)'$. We set $\mathbf{S}_0 \equiv \mathbf{S}(\boldsymbol{\psi}_0) = \mathbf{I}_N - \boldsymbol{\Psi}_0 \boldsymbol{W}$, and assume that $\mathbf{S}_0$ is invertible.[3] Then, the reduced form of (4) can be expressed as

$$\boldsymbol{y}_{\circ t} = \mathbf{S}^{-1}(\boldsymbol{\psi}_0)\left[\boldsymbol{B}_0 \boldsymbol{x}_{\circ t} + \boldsymbol{\varepsilon}_{\circ t}\right], \; t = 1, 2, \ldots, T. \tag{6}$$

Remark 1 *If $\psi_{10} = \psi_{20} = \cdots = \psi_{N0} = \psi_0$, $\boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20} = \cdots = \boldsymbol{\beta}_{N0} = \boldsymbol{\beta}_0$ and $\sigma_{10}^2 = \sigma_{20}^2 = \cdots = \sigma_{N0}^2 = \sigma_0^2$, then (1) collapses to the standard first order SAR model. Alternatively, if $\boldsymbol{\beta}_{10} = \boldsymbol{\beta}_{20} = \cdots = \boldsymbol{\beta}_{N0} = \mathbf{0}$ and $\sigma_{10}^2 = \sigma_{20}^2 = \cdots = \sigma_{N0}^2 = \sigma_0^2$, then (1) reverts to a first order HSAR model with no exogenous regressors and with homoskedastic errors.*

2.2 The log-likelihood function

To estimate the unit-specific coefficients we collect all the parameters of the N units in the $N(k+2) \times 1$ vector $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})'$ where $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_N)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_N')'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)'$, and denote the associated vector of true values by $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}_0', \boldsymbol{\beta}_0', \boldsymbol{\sigma}_0^{2\prime})'$. The log-likelihood function of (6) can be written as

$$\begin{aligned} \ell_T(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) &= -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + \frac{T}{2}\ln\left|\mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})\right| \\ &\quad - \frac{1}{2}\sum_{t=1}^{T}\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right]' \boldsymbol{\Sigma}^{-1}\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right], \end{aligned} \tag{7}$$

where $\boldsymbol{\Sigma} = \text{Diag}\,(\boldsymbol{\sigma}^2)$, $\boldsymbol{\Psi} = \text{Diag}\,(\boldsymbol{\psi})$, and $\mathbf{S} \equiv \mathbf{S}(\boldsymbol{\psi}) = \boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}$.

The quasi maximum likelihood estimators (QMLE), $\hat{\boldsymbol{\theta}}$, are the extreme value estimators obtained by maximization of (7). When the error terms, $\boldsymbol{\varepsilon}_{\circ t}(\boldsymbol{\theta}_0) = \mathbf{S}(\boldsymbol{\psi}_0)\boldsymbol{y}_{\circ t} - \boldsymbol{B}_0\boldsymbol{x}_{\circ t}$, are normally distributed, then vector $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}$, while under non-Gaussian errors, $\hat{\boldsymbol{\theta}}$ is the QMLE of $\boldsymbol{\theta}$.

3 Asymptotic properties of QML estimators

3.1 Assumptions

In order to investigate the conditions under which $\boldsymbol{\theta}_0$ is identified, and to establish consistency and the asymptotic normality of $\hat{\boldsymbol{\theta}}$, we make the following assumptions, using the filtration $\mathcal{F}_t = (\boldsymbol{x}_{\circ t}, \boldsymbol{x}_{\circ t-1}, \boldsymbol{x}_{\circ t-2}, \ldots)$, where $\boldsymbol{x}_{\circ t} = (\boldsymbol{x}_{1t}', \boldsymbol{x}_{2t}', \ldots, \boldsymbol{x}_{Nt}')'$:

[3] Conditions under which $\mathbf{S}_0$ is invertible are discussed in Section 3.1.

Assumption 1 *The error terms* $\{\varepsilon_{it},\ i=1,2,\ldots,N; t=1,2,\ldots,T\}$ *are independently distributed over* i *and* t*;* $E(\varepsilon_{it}|\mathcal{F}_t)=0$*,* $E(\varepsilon_{it}^2|\mathcal{F}_t)=\sigma_{i0}^2$*, for* $i=1,2,\ldots,N$*, where* $\inf_i\left(\sigma_{i0}^2\right)>c>0$*,* $\sup_i\left(\sigma_{i0}^2\right)<K<\infty$*, and* $E\left(|\varepsilon_{it}|^p|\mathcal{F}_t\right)=E\left(|\varepsilon_{it}|^p\right)=\varpi_{ip}<K$*, for all* i *and* t*, where* ϖ_{ip} *is a time-invariant constant,* $1\leq p\leq 4+\epsilon$*, for some* $\epsilon>0$*.* [4]

Assumption 2 *(a)* $\boldsymbol{x}_{\circ t}$ *are stationary processes with mean zero, and satisfy the moment condition* $\sup_{i,\ell,t}E\left(|x_{i\ell,t}|^{2+c}\right)<K$*, for some* $c>0$*,* $i=1,2,\ldots,N$*,* $\ell=1,2,\ldots,k$*, and* $t=1,2,\ldots,T$*. (b)* $E\left(\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}'\right)=\boldsymbol{\Sigma}_{\boldsymbol{xx}}=\left(\boldsymbol{\Sigma}_{ij}\right)$*, where* $\boldsymbol{\Sigma}_{ij}=E(\boldsymbol{x}_{it}\boldsymbol{x}_{jt}')$ *exists for all* i *and* j*, such that* $\sup_{i,j}\|\boldsymbol{\Sigma}_{ij}\|<K$*, and* $\boldsymbol{\Sigma}_{ii}$ *is a* $k\times k$ *non-singular matrix with* $\inf_i\left[\lambda_{\min}\left(\boldsymbol{\Sigma}_{ii}\right)\right]>c>0$*, and* $\sup_i\left[\lambda_{\max}\left(\boldsymbol{\Sigma}_{ii}\right)\right]<K$*; (c)* $T^{-1}\sum_{t=1}^T\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}'\overset{a.s.}{\to}\boldsymbol{\Sigma}_{\boldsymbol{xx}}$*, as* $T\to\infty$*.*

Assumption 3 *The* $N(k+2)\times 1$ *parameter vector* $\boldsymbol{\theta}=(\boldsymbol{\psi}',\boldsymbol{\beta}',\boldsymbol{\sigma}^{2\prime})'$ *belongs to* $\boldsymbol{\Theta}=\boldsymbol{\Theta}_\psi\times\boldsymbol{\Theta}_\beta\times\boldsymbol{\Theta}_\sigma\subset\mathbb{R}^N\times\mathbb{R}^{Nk}\times\mathbb{R}^N$*, a sub-set of the* $N(k+2)$ *dimensional Euclidean space,* $\mathbb{R}^{N(k+2)}$*.* $\boldsymbol{\Theta}$ *is a closed and bounded (compact) set and includes the true value of* $\boldsymbol{\theta}$*, denoted by* $\boldsymbol{\theta}_0$*, which is an interior point of* $\boldsymbol{\Theta}$*, and* $\sup_i\|\boldsymbol{\beta}_i\|_1<K$*.*

Assumption 4 $\boldsymbol{W}=(w_{ij})$ *is a constant known weights matrix which is uniformly bounded in row and column sums in absolute value, i.e.* $\|\boldsymbol{W}\|_\infty<K<\infty$ *and* $\|\boldsymbol{W}\|_1<K<\infty$*, and its diagonal elements are zero, that is* $w_{ii}=0$*, for* $i=1,2,\ldots,N$*. In addition, the spatial autoregressive parameters reside in the range set by*

$$\sup_i|\psi_i|<\max\left\{1/\|\boldsymbol{W}\|_1,1/\|\boldsymbol{W}\|_\infty\right\},\tag{8}$$

for all values of ψ_i*,* $i=1,2,\ldots,N$*, in* $\boldsymbol{\Theta}_\psi$*.*

Remark 2 *Assumption 1 implies that* $E(\varepsilon_{it})=0$*,* $E(\varepsilon_{it}^2)=\sigma_{i0}^2$*, for* $i=1,2,\ldots,N$*, and does not allow for conditional heteroskedasticity. But it is possible to allow for time variations in* $E\left(|\varepsilon_{it}|^{4+\epsilon}|\mathcal{F}_t\right)$ *by relaxing the moment conditions on* ε_{it} *and* $x_{i\ell,t}$*.*

Remark 3 *Assumption 2 is standard and allows for the regressors to be cross-sectionally correlated, and hence* $\boldsymbol{x}_{\circ t}$ *can also include observable common factors. This is sufficiently general and applies for all* N*. Further, it allows the regressors to be weakly exogenous, thus allowing the spatial model to include lagged values of the dependent variable. Finally, the theoretical model (1) can be modified to include an intercept (fixed effects) by setting one of the elements of* $\boldsymbol{x}_{it}$ *to unity, at the expense of complicating the algebra. Such a setting is analyzed in the Monte Carlo simulation study of Section 5.*

Remark 4 *Assumption 4 is sufficiently general and allows the spatial weights to take negative values. But, as noted above, in empirical applications one might wish to distinguish between positive and negative connections as they might have differential effects on the outcomes. This assumption does not require the weights to be normalized either, so long as condition (8) is met.*

Remark 5 *Condition (8) is sufficient for global invertibility of matrix* $\mathbf{S}(\boldsymbol{\psi})=\boldsymbol{I}_N-\boldsymbol{\Psi}\boldsymbol{W}$ *on* $\boldsymbol{\Theta}_\psi$*. See Lemma 1 in the online 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*.*

[4]Clearly, $\varpi_{i2}=\sigma_{i0}^2$.

Remark 6 *Let* $\mathbf{V}(\boldsymbol{\psi}) = \mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})$*, where* $\mathbf{S}(\boldsymbol{\psi}) = \boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}$*. Then under Assumption 4, for all values of* $\boldsymbol{\psi} \in \boldsymbol{\Theta}_\psi$*, and for all* N *we have*

$$\lambda_{\min}\left[\mathbf{V}(\boldsymbol{\psi})\right] > c > 0, \tag{9}$$

and

$$\begin{aligned}\lambda_{\max}\left[\mathbf{V}(\boldsymbol{\psi})\right] &\leq \|\mathbf{S}(\boldsymbol{\psi})\|_1 \|\mathbf{S}(\boldsymbol{\psi})\|_\infty \\ &\leq \left(1 + \sup_i |\psi_i| \, \|\boldsymbol{W}\|_1\right)\left(1 + \sup_i |\psi_i| \, \|\boldsymbol{W}\|_\infty\right) < K < \infty.\end{aligned} \tag{10}$$

These results, together with Assumptions 1, 2(b) and 3, ensure that for all values of N *and* $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})' \in \boldsymbol{\Theta}$,

$$\lambda_{\min}\left[\boldsymbol{\Sigma}_y(\boldsymbol{\theta})\right] > c > 0, \text{ and } \lambda_{\max}\left[\boldsymbol{\Sigma}_y(\boldsymbol{\theta})\right] < K < \infty, \tag{11}$$

and

$$\lambda_{\min}\left[\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\theta})\right] > c > 0, \text{ and } \lambda_{\max}\left[\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\theta})\right] < K < \infty, \tag{12}$$

where $\boldsymbol{\Sigma}_y(\boldsymbol{\theta}) = \mathbf{S}(\boldsymbol{\psi})^{-1}\left[\boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\boldsymbol{B}' + \boldsymbol{\Sigma}\right]\mathbf{S}(\boldsymbol{\psi})'^{-1}$*.*

3.2 Identification

Here we focus on the problem of identification of the individual parameters in $N(k+2) \times 1$ vector $\boldsymbol{\theta}_0$ for a given N, and as $T \to \infty$. To highlight the main issues involved in the identification of spatial parameters under the heterogeneous setting, first we consider the HSAR model (4) without the exogenous regressors, namely

$$y_{it} = \psi_{i0}\sum_{j=1}^{N} w_{ij}y_{jt} + \varepsilon_{it}, \; i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T, \tag{13}$$

where $\varepsilon_{it} \sim IIDN\left(0, \sigma_{i0}^2\right)$ for $i = 1, 2, \ldots, N$. Under Assumption 4, (6) is then simplified to

$$\boldsymbol{y}_{\circ t} = \mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_{\circ t}, \; t = 1, 2, \ldots, T. \tag{14}$$

With a slight abuse of notation let $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\sigma}^{2\prime})'$, and note that in this case the log-likelihood function is given by

$$\ell_T(\boldsymbol{\theta}) = -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + \frac{T}{2}\ln|\mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})| - \frac{1}{2}\sum_{t=1}^{T}\boldsymbol{y}_{\circ t}'\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t}. \tag{15}$$

It is also helpful to write the associated average log-likelihood function as

$$\bar{\ell}_T(\boldsymbol{\theta}) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + \frac{1}{2}\ln|\mathbf{V}(\boldsymbol{\psi})| - \frac{1}{2}\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{y}_{\circ t}'\mathbf{P}(\boldsymbol{\theta})\boldsymbol{y}_{\circ t}\right), \tag{16}$$

where

$$\mathbf{V}(\boldsymbol{\psi}) = \mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi}),\ \ \mathbf{P}(\boldsymbol{\theta}) = \mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi}), \text{ and } \mathbf{S}(\boldsymbol{\psi}) = \boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}. \tag{17}$$

Let $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})$, in which $\bar{\ell}_T(\boldsymbol{\theta}_0)$ is $\bar{\ell}_T(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. Then, for a given N and as $T \to \infty$, we have (see Lemma 3 of the online appendix A when setting $\boldsymbol{B} = \boldsymbol{O}$ in (A.5)) $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) - E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] \overset{a.s.}{\to} 0$, where

$$\begin{aligned} E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = E_0\left[\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})\right] &= -\frac{1}{2}\sum_{i=1}^{N} \ln\left(\sigma_{i0}^2/\sigma_i^2\right) - \frac{N}{2} \\ &+ \frac{1}{2}\left[\ln\left(\frac{|\mathbf{V}(\boldsymbol{\psi}_0)|}{|\mathbf{V}(\boldsymbol{\psi})|}\right)\right] + \frac{1}{2}\operatorname{tr}\left[\mathbf{P}(\boldsymbol{\theta})\mathbf{P}^{-1}(\boldsymbol{\theta}_0)\right]. \end{aligned} \tag{18}$$

Alternatively, we can express (18) in terms of the eigenvalues of $\mathbf{V}(\boldsymbol{\psi})$ and $\mathbf{V}(\boldsymbol{\psi}_0)$ which we denote by λ_i^2 and λ_{i0}^2, respectively. Recall from Remark 6 that $0 < \lambda_i^2, \lambda_{i0}^2 < K$. Therefore, (18) reduces to

$$E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = \frac{1}{2}\sum_{i=1}^{N}\left[\left(\tilde{\lambda}_i^2/\tilde{\lambda}_{i0}^2\right) - \ln\left(\tilde{\lambda}_i^2/\tilde{\lambda}_{i0}^2\right) - 1\right], \tag{19}$$

where $\tilde{\lambda}_i^2 = \lambda_i^2/\sigma_i^2$, $\tilde{\lambda}_{i0}^2 = \lambda_{i0}^2/\sigma_{i0}^2$, and $0 < \sigma_i^2, \sigma_{i0}^2 < K$, for $i = 1, 2, \ldots, N$. The above results imply that

$$\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \overset{a.s.}{\to} E_0\left[\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})\right] = E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] \geq 0, \tag{20}$$

and in view of (19) we have $\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \overset{a.s.}{\to} 0$ if

$$\sum_{i=1}^{N}\left[\left(\tilde{\lambda}_i^2/\tilde{\lambda}_{i0}^2\right) - \ln\left(\tilde{\lambda}_i^2/\tilde{\lambda}_{i0}^2\right) - 1\right] = 0. \tag{21}$$

Hence, for a given N, it readily follows that $\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \overset{a.s.}{\to} 0$, as $T \to \infty$, if and only if $\lambda_i^2/\sigma_i^2 = \lambda_{i0}^2/\sigma_{i0}^2$ for all i.[5] Therefore, the ratio $\lambda_{i0}^2/\sigma_{i0}^2$ is globally identified, although without further restrictions on $\mathbf{P}(\boldsymbol{\theta})$ and $\mathbf{S}(\boldsymbol{\psi})$, it will not be possible to separately identify λ_{i0}^2 and σ_{i0}^2.

Consider now the problem of identification of $\boldsymbol{\psi}_0$, which is the parameter vector of interest, and note that

$$\begin{aligned} \frac{|\mathbf{V}(\boldsymbol{\psi}_0)|}{|\mathbf{V}(\boldsymbol{\psi})|} &= \frac{|\mathbf{S}(\boldsymbol{\psi}_0)|^2}{|\mathbf{S}(\boldsymbol{\psi})|^2} = \left|\mathbf{S}(\boldsymbol{\psi}_0)\mathbf{S}^{-1}(\boldsymbol{\psi})\right|^2 = \left|\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\right|^{-2}, \\ \operatorname{tr}\left[\mathbf{P}(\boldsymbol{\theta})\mathbf{P}^{-1}(\boldsymbol{\theta}_0)\right] &= \operatorname{tr}\left[\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\Sigma}_0\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\right] \\ &= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1/2}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\Sigma}_0\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1/2}\right], \end{aligned}$$

and rewrite (18) as

$$\begin{aligned} E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = -\frac{N}{2} - \frac{1}{2}\sum_{i=1}^{N}\ln\left(\sigma_{i0}^2/\sigma_i^2\right) - \left[\ln\left(\left|\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\right|\right)\right] \\ + \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1/2}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\Sigma}_0\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1/2}\right]. \end{aligned}$$

[5] Note that $a - \ln(a) - 1 \geq 0$, for any $a > 0$, with the equality holding if and only if $a = 1$.

Further, we note that $\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0) = \boldsymbol{I}_N - \boldsymbol{D}\boldsymbol{G}_0$, where $\boldsymbol{G}_0 = \boldsymbol{W}\left(\boldsymbol{I}_N - \boldsymbol{\Psi}_0\boldsymbol{W}\right)^{-1}$, and $\boldsymbol{D} = \boldsymbol{\Psi} - \boldsymbol{\Psi}_0$, is a diagonal matrix with elements $d_i = \psi_i - \psi_{i0}$. Using these results, the above expression for $E_0\left[Q_T\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right)\right]$ can be written equivalently as

$$
\begin{aligned}
E_0\left[Q_T\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right)\right] &= -\frac{N}{2} - \frac{1}{2}\sum_{i=1}^{N}\ln\left(\sigma_{i0}^2/\sigma_i^2\right) - \ln\left|\boldsymbol{I}_N - \boldsymbol{D}\boldsymbol{G}_0\right| \\
&+ \frac{1}{2}\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1/2}\left(\boldsymbol{I}_N - \boldsymbol{D}\boldsymbol{G}_0\right)'\boldsymbol{\Sigma}_0\left(\boldsymbol{I}_N - \boldsymbol{D}\boldsymbol{G}_0\right)\boldsymbol{\Sigma}^{-1/2}\right] \\
&= A_N + B_N,
\end{aligned}
\tag{22}
$$

where

$$
A_N = \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\sigma_{i0}^2}{\sigma_i^2} - \ln\left(\sigma_{i0}^2/\sigma_i^2\right) - 1\right] - \ln\left|\boldsymbol{I}_N - \boldsymbol{D}\boldsymbol{G}_0\right| - \operatorname{tr}\left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}_0\boldsymbol{D}\boldsymbol{G}_0\right), \tag{23}
$$

$$
B_N = \frac{1}{2}\operatorname{tr}\left(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{G}_0'\boldsymbol{D}\boldsymbol{\Sigma}_0\boldsymbol{D}\boldsymbol{G}_0\boldsymbol{\Sigma}^{-1/2}\right). \tag{24}
$$

We first note that $B_N \geq 0$, since we can write $B_N = (1/2)\operatorname{tr}\left(\mathbf{A}_0'\mathbf{A}_0\right)$, with $\mathbf{A}_0 = \boldsymbol{\Sigma}_0^{1/2}\boldsymbol{D}\boldsymbol{G}_0\boldsymbol{\Sigma}^{-1/2}$. Consider now A_N, denote the i^{th} eigenvalue of $\boldsymbol{D}\boldsymbol{G}_0$ by μ_i, and note that since $\boldsymbol{I}_N - \boldsymbol{D}\boldsymbol{G}_0 = \mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)$, then the eigenvalues of $\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)$ are also given by $1 - \mu_i$, for $i = 1, 2, \ldots, N$. Further, by Lemma 1 of the online appendix A, $\lambda_{\min}\left[\mathbf{S}(\boldsymbol{\psi})\right] > 0$ for all ψ_i that satisfy condition (8). Hence, we must also have $1 - \mu_i > 0$, for all i. Using these results, A_N can now be written as

$$
A_N = \frac{1}{2}\sum_{i=1}^{N}\left[\frac{\sigma_{i0}^2}{\sigma_i^2} - \ln\left(\sigma_{i0}^2/\sigma_i^2\right) - 1\right] - \sum_{i=1}^{N}\ln(1-\mu_i) - \sum_{i=1}^{N}\left(\frac{\sigma_{i0}^2}{\sigma_i^2}\right)\mu_i.
$$

Let $\delta_{\sigma i} = \sigma_{i0}^2/\sigma_i^2$ and $\delta_{\psi i} = (1 - \mu_i) > 0$, for all i. Then, write $E_0\left[Q_T\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right)\right]$ as

$$
\begin{aligned}
E_0\left[Q_T\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right)\right] &= A_N + B_N \\
&= \frac{1}{2}\sum_{i=1}^{N}\left[\delta_{\sigma i} - \ln\left(\delta_{\sigma i}\right) - 1\right] - \sum_{i=1}^{N}\ln\delta_{\psi i} - \sum_{i=1}^{N}\delta_{\sigma i}(1 - \delta_{\psi i}) + B_N \\
&= \frac{1}{2}\sum_{i=1}^{N}\left[\delta_{\sigma i} - \ln\left(\delta_{\sigma i}\right) - 1\right] + \left[\sum_{i=1}^{N}\delta_{\sigma i}\left(\delta_{\psi i} - \ln\delta_{\psi i} - 1\right)\right] \\
&+ \left[\sum_{i=1}^{N}\left(\delta_{\sigma i} - 1\right)\ln\delta_{\psi i}\right] + B_N \\
&= A_{1,N} + A_{2,N} + \left(A_{3,N} + B_N\right).
\end{aligned}
$$

Since $\delta_{\sigma i} > 0$, and $\delta_{\psi i} > 0$ for all i, then $\delta_{\sigma i} - \ln\left(\delta_{\sigma i}\right) - 1 \geq 0$, and $\delta_{\psi i} - \ln\delta_{\psi i} - 1 \geq 0$ for all i, with equalities holding if and only if $\delta_{\sigma i} = 1$ and $\delta_{\psi i} = 1$ for all i. Hence, $A_{1,N} \geq 0$, and $A_{2,N} \geq 0$ for all values of N, and global identification of σ_{i0}^2 will be possible only if we are able to show that $A_{3,N} + B_N$ is non-negative. But it is easily seen that the non-negativity of $A_{3,N} + B_N$

can not be guaranteed without further restrictions. This follows since

$$A_{3,N} = \sum_{i=1}^{N} (\delta_{\sigma i} - 1) \ln \delta_{\psi i},$$

and there are values of $\delta_{\sigma i}$ and $\delta_{\psi i}$ in $\boldsymbol{\Theta} = \boldsymbol{\Theta}_\psi \times \boldsymbol{\Theta}_\sigma$ for which $A_{3,N} < 0$. Considering $(A_{3,N} + B_N)$ somewhat weakens the requirement since $B_N \geq 0$, but still does not guarantee that $(A_{3,N} + B_N) \geq 0$, for all values of $\delta_{\sigma i} > 0$ and $\delta_{\psi i} > 0$. Therefore, global identification of $\boldsymbol{\psi}_0$ can not be guaranteed. To investigate the possibility of local identification we introduce the following definition:

Definition 1 *Consider the set $\mathcal{N}_c(\boldsymbol{\sigma}_0^2)$ in the closed neighborhood of $\boldsymbol{\sigma}_0^2$ defined by*

$$\mathcal{N}_c(\boldsymbol{\sigma}_0^2) = \left\{ \boldsymbol{\sigma}_0^2 \in \boldsymbol{\Theta}_\sigma,\ \left| \sigma_{i0}^2 / \sigma_i^2 - 1 \right| < c_i, \text{ for } i = 1, 2, \ldots, N \right\},$$

for some $c_i > 0$, $i = 1, 2, \ldots, N$, where $\boldsymbol{\Theta}_\sigma$ is a compact subset of $\mathbb{R}^N$.

We now show that $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}_0', \boldsymbol{\sigma}_0^{2\prime})'$ is identified on $\boldsymbol{\Theta}_c = \boldsymbol{\Theta}_\psi \times \mathcal{N}_c(\boldsymbol{\sigma}_0^2)$. Consider values of $\delta_{\sigma i}$ within the local neighborhood of $\delta_{\sigma i} = 1$ for all i. Recall that $A_{1,N} + A_{2,N} \geq 0$, and the boundary values $A_{1,N} = 0$ or $A_{2,N} = 0$ can occur if and only if $\delta_{\sigma i} = 1$ and $\delta_{\psi i} = 1$ for all i, respectively. Therefore, $A_N \geq 0$ if $\delta_{\sigma i} = 1$, otherwise $A_{1,N} > 0$. Similarly, $A_N \geq 0$ if $\delta_{\psi i} = 1$, otherwise $A_{2,N} > 0$. Therefore, there must exist $\mathbf{c} = (c_1, c_2, \ldots, c_N) > \mathbf{0}$, such that $A_N = 0$ on $\boldsymbol{\Theta}_c$ if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, which in turn establishes that $\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \overset{a.s.}{\to} 0$, as $T \to \infty$, on the set $\boldsymbol{\Theta}_c$ *if and only if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Next, consider the HSAR model (4) with exogenous regressors. The average log-likelihood in this case is given by (see (7))

$$\begin{aligned} \bar{\ell}_T(\boldsymbol{\theta}) = {} & -\frac{N}{2} \ln(2\pi) - \frac{1}{2} \sum_{i=1}^{N} \ln \sigma_i^2 + \frac{1}{2} \ln |\mathbf{V}(\boldsymbol{\psi})| \\ & - \frac{1}{2} \left(\frac{1}{T} \sum_{t=1}^{T} \left[\mathbf{S}(\boldsymbol{\psi}) \boldsymbol{y}_{\circ t} - \boldsymbol{B} \boldsymbol{x}_{\circ t} \right]' \boldsymbol{\Sigma}^{-1} \left[\mathbf{S}(\boldsymbol{\psi}) \boldsymbol{y}_{\circ t} - \boldsymbol{B} \boldsymbol{x}_{\circ t} \right] \right), \end{aligned} \tag{25}$$

where $\boldsymbol{\theta}$ is now defined by $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})'$ and $\boldsymbol{B}$ has the same form as in (5). Following a similar line of reasoning as in the case without exogenous regressors (see Lemma 3 of the online appendix A), we have that $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})$, where $\bar{\ell}_T(\boldsymbol{\theta})$ is now given by (25), and $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) - E_0 \left[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \right] \overset{a.s.}{\to} 0$, (as $T \to \infty$) where

$$E_0 \left[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \right] = A_N + B_N + C_N. \tag{26}$$

A_N and B_N are defined as before by (23) and (24), and C_N is given by

$$\begin{aligned} C_N &= \frac{1}{2}\sum_{i=1}^{N}\frac{(\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i0})' \boldsymbol{\Sigma}_{ii} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i0})}{\sigma_i^2} + \mathrm{tr}\left[\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{B} - \boldsymbol{B}_0)\boldsymbol{\Sigma}_{xx}\boldsymbol{\Xi}_0'\right] \\ &+ \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_{xx}\boldsymbol{\Xi}_0'\boldsymbol{\Xi}_0\right) \\ &= C_{1,N} + C_{2,N} + C_{3,N}, \end{aligned} \tag{27}$$

where $\boldsymbol{\Xi}_0 = \boldsymbol{\Sigma}^{-1/2}\boldsymbol{D}\boldsymbol{G}_0\boldsymbol{B}_0$, and as before $\mathbf{D} = \mathrm{Diag}(\boldsymbol{\psi} - \boldsymbol{\psi}_0)$. Consider now $C_{3,N}$ and note that since $\boldsymbol{\Sigma}_{xx} = E(\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}')$ and $\boldsymbol{\Xi}_0'\boldsymbol{\Xi}_0$ are positive semi-definite matrices, then using result (9) on p. 44 of Lütkepohl (1996),

$$\mathrm{tr}\left(\boldsymbol{\Sigma}_{xx}\boldsymbol{\Xi}_0'\boldsymbol{\Xi}_0\right) \geq N\left[\det\left(\boldsymbol{\Sigma}_{xx}\right)\right]^{1/N}\left[\det\left(\boldsymbol{\Xi}_0'\boldsymbol{\Xi}_0\right)\right]^{1/N} \geq 0,$$

and hence $C_{3,N} \geq 0$. Also, as shown above, on the subset $\boldsymbol{\Theta}_c = \boldsymbol{\Theta}_\psi \times \boldsymbol{\Theta}_\beta \times \mathcal{N}_c(\boldsymbol{\sigma}_0^2)$, $A_N + B_N = 0$ if and only if $\mathbf{D} = \mathrm{Diag}(\boldsymbol{\psi} - \boldsymbol{\psi}_0) = \mathbf{0}$, and hence it must also follow that $C_{2,N} = 0$ on $\boldsymbol{\Theta}_c$. Thus, overall $\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \overset{a.s.}{\to} 0$ on $\boldsymbol{\Theta}_c$ if and only if

$$\sum_{i=1}^{N}\frac{(\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i0})' \boldsymbol{\Sigma}_{ii} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i0})}{\sigma_i^2} = 0. \tag{28}$$

This equality holds for all N if and only if $(\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i0})' \boldsymbol{\Sigma}_{ii} (\boldsymbol{\beta}_i - \boldsymbol{\beta}_{i0}) = 0$, for all i, and since under Assumption 2(b) $\boldsymbol{\Sigma}_{ii}$ is a positive definite matrix this can occur if and only if $\boldsymbol{\beta}_i = \boldsymbol{\beta}_{i0}$ for all i.

Before we state the identification result for the general model (4), we require the following modification of Assumption 3:

Assumption 5 *The $N(k+2) \times 1$ parameter vector $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})'$ belongs to $\boldsymbol{\Theta}_c = \boldsymbol{\Theta}_\psi \times \boldsymbol{\Theta}_\beta \times \mathcal{N}_c(\boldsymbol{\sigma}_0^2)$, where $\boldsymbol{\Theta}_\psi$ and $\boldsymbol{\Theta}_\beta$ are compact subsets of $\mathbb{R}^N$ and $\mathbb{R}^{Nk}$, respectively, and $\mathcal{N}_c(\boldsymbol{\sigma}_0^2)$ is given in Definition 1, and $\boldsymbol{\Theta}_c$ is a sub-set of the $N(k+2)$ dimensional Euclidean space, $\mathbb{R}^{N(k+2)}$.*

The main identification result of the paper is summarized in the following proposition:

Proposition 1 *Consider the heterogeneous spatial autoregressive (HSAR) model given by (4) with the associated log-likelihood function given by (7). Suppose that Assumptions 1, 2, 4 and 5 hold. Then, the $N(k+2)$ dimensional true parameter vector $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}_0', \boldsymbol{\beta}_0', \boldsymbol{\sigma}_0^{2\prime})'$ is almost surely locally identified on $\boldsymbol{\Theta}_c$.*

3.3 Consistency and asymptotic normality

We are now in a position to consider consistency and asymptotic normality of the QML estimator of $\boldsymbol{\theta}$, given by $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \bar{\ell}_T(\boldsymbol{\theta})$, where $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\psi}}', \hat{\boldsymbol{\beta}}', \hat{\boldsymbol{\sigma}}^{2\prime})'$. We establish the results for a given N, and as $T \to \infty$. First, we focus on the proof of consistency. Under Assumptions 1, 2, 4 and 5, we have: (i) $\boldsymbol{\Theta}_c$, being a subset of $\boldsymbol{\Theta}$, is compact, (ii) $\boldsymbol{\theta}_0$ is an interior point of $\boldsymbol{\Theta}_c$, (iii) $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) \overset{a.s.}{\to} E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})]$, with $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})$ and $E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = A_N + B_N + C_N$,where A_N, B_N and C_N are given by (23), (24) and (27), respectively, and (iv)

$\boldsymbol{\theta}_0$ is a unique maximum of $E_0\left[Q_T\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right)\right]$ on $\boldsymbol{\Theta}_c$. The last result follows from the identification analysis of Section 3.2. It is clear that all conditions of Theorem 9.3.1 of Davidson (2000) are satisfied, therefore almost sure local consistency of $\hat{\boldsymbol{\theta}}$ is ensured, with $\hat{\boldsymbol{\theta}} \stackrel{a.s.}{\rightarrow} \boldsymbol{\theta}_0$ on $\boldsymbol{\Theta}_c$, as $T \rightarrow \infty$. To establish asymptotic normality of $\hat{\boldsymbol{\theta}}$, we apply the mean value theorem to $\bar{\ell}_T(\boldsymbol{\theta})$ such that

$$\bar{\ell}_T(\boldsymbol{\theta}) - \bar{\ell}_T(\boldsymbol{\theta}_0) = \left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' \bar{\boldsymbol{s}}_T\left(\boldsymbol{\theta}_0\right) - \frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' \bar{\mathbf{H}}_T(\bar{\boldsymbol{\theta}})\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right), \tag{29}$$

where $\bar{\boldsymbol{s}}_T\left(\boldsymbol{\theta}\right) = \partial\bar{\ell}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\bar{\mathbf{H}}_T(\boldsymbol{\theta}) = -\partial^2\bar{\ell}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$, and $\bar{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. By Lemma 5 of the online appendix A we have $\bar{\boldsymbol{s}}_T\left(\boldsymbol{\theta}_0\right) \stackrel{a.s.}{\rightarrow} \mathbf{0}$, and by the results of Section 3.2 we also have $\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \stackrel{a.s.}{\rightarrow} E_0\left[\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})\right] \geq 0$. Hence, in view of (29) it must also hold that (as $T \rightarrow \infty$)

$$\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right)' \bar{\mathbf{H}}_T(\bar{\boldsymbol{\theta}})\left(\boldsymbol{\theta} - \boldsymbol{\theta}_0\right) \stackrel{a.s.}{\rightarrow} E_0\left[Q_T\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right)\right],$$

where $E_0\left[Q_T\left(\boldsymbol{\theta}_0, \boldsymbol{\theta}\right)\right]$ is given by (26). But we have already established that on $\boldsymbol{\Theta}_c$, the right hand side of the above expression can be equal to zero if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, and hence it must be that $\bar{\mathbf{H}}_T(\bar{\boldsymbol{\theta}}) \stackrel{a.s.}{\rightarrow} \bar{\mathbf{H}}(\boldsymbol{\theta}_0)$, where $\bar{\mathbf{H}}(\boldsymbol{\theta}_0)$ *must* be a positive definite matrix given by

$$\bar{\mathbf{H}}(\boldsymbol{\theta}_0) = \lim_{T \rightarrow \infty} E_0\left(-\partial^2\bar{\ell}_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right).$$

Next, for a given N we apply the mean value theorem to $\bar{\boldsymbol{s}}_T\left(\boldsymbol{\theta}\right)$ so that

$$\mathbf{0} = \sqrt{T}\bar{\boldsymbol{s}}_T\left(\hat{\boldsymbol{\theta}}\right) = \sqrt{T}\bar{\boldsymbol{s}}_T\left(\boldsymbol{\theta}_0\right) - \bar{\mathbf{H}}_T(\breve{\boldsymbol{\theta}})\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),$$

or equivalently

$$\mathbf{0} = \frac{1}{\sqrt{T}}\boldsymbol{s}_T\left(\hat{\boldsymbol{\theta}}\right) = \frac{1}{\sqrt{T}}\boldsymbol{s}_T\left(\boldsymbol{\theta}_0\right) - \mathbf{H}_T(\breve{\boldsymbol{\theta}})\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right),$$

where $\boldsymbol{s}_T\left(\boldsymbol{\theta}\right) = \partial\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$, $\mathbf{H}_T(\boldsymbol{\theta}) = -\frac{1}{T}\partial^2\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$, and $\breve{\boldsymbol{\theta}}$ lies between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. Therefore,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) = \mathbf{H}_T^{-1}(\breve{\boldsymbol{\theta}})\left[\sqrt{T}\boldsymbol{s}_T\left(\boldsymbol{\theta}_0\right)\right],$$

and since $\hat{\boldsymbol{\theta}}$ is consistent on $\boldsymbol{\Theta}_c$, then

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \stackrel{a}{\sim} \mathbf{H}^{-1}(\boldsymbol{\theta}_0)\left[\sqrt{T}\boldsymbol{s}_T\left(\boldsymbol{\theta}_0\right)\right],$$

where $\boldsymbol{H}\left(\boldsymbol{\theta}_0\right) = \lim_{T \rightarrow \infty} E_0\left[-\frac{1}{T}\partial^2\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right]$, with

$$E_0\left[-\frac{1}{T}\partial^2\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'\right] = \begin{pmatrix} \boldsymbol{H}_{11} & \boldsymbol{H}_{12} & \boldsymbol{H}_{13} \\ \boldsymbol{H}_{12}' & \boldsymbol{H}_{22} & \boldsymbol{H}_{23} \\ \boldsymbol{H}_{13}' & \boldsymbol{H}_{23}' & \boldsymbol{H}_{33} \end{pmatrix}_{N(k+2)\times N(k+2)}.$$

The expressions for $\mathbf{H}_{ij}$ can be obtained using the partial derivative $\partial^2 \ell_T(\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'$ given in the online appendix B. Specifically we have

$$\boldsymbol{H}\left(\boldsymbol{\theta}_0\right) = \begin{pmatrix} \left(\mathbf{G}_0 \odot \mathbf{G}_0'\right) + \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}\left(\boldsymbol{G}_0\boldsymbol{\Sigma}_0\boldsymbol{G}_0'\right) + \boldsymbol{\Delta}_{\boldsymbol{\beta}_0} & \mathbb{E}_{\boldsymbol{\beta}_0} & \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}\left(\mathbf{G}_0\right) \\ \mathbb{E}_{\boldsymbol{\beta}_0} & \boldsymbol{Z}_0 & \mathbf{0} \\ \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}\left(\mathbf{G}_0\right) & \mathbf{0}' & \frac{1}{2}\boldsymbol{\Sigma}_0^{-2} \end{pmatrix}, \tag{30}$$

where $\boldsymbol{\Delta}_{\boldsymbol{\beta}_0}$, $\mathbb{E}_{\boldsymbol{\beta}_0}$, and $\boldsymbol{Z}_0$ are diagonal matrices given by

$$\boldsymbol{\Delta}_{\boldsymbol{\beta}_0} = \mathrm{Diag}\left[\sigma_{i0}^{-2}\sum_{r=1}^{N}\sum_{s=1}^{N} g_{0,is}g_{0,ir}\boldsymbol{\beta}_{r0}'\boldsymbol{\Sigma}_{rs}\boldsymbol{\beta}_{s0},\ i=1,2,\ldots,N\right], \tag{31}$$

$$\mathbb{E}_{\boldsymbol{\beta}_0} = \mathrm{Diag}\left[\sigma_{i0}^{-2}\sum_{s=1}^{N} g_{0,is}\boldsymbol{\beta}_{s0}'\boldsymbol{\Sigma}_{is},\ i=1,2,\ldots,N\right], \tag{32}$$

and

$$\boldsymbol{Z}_0 = \mathrm{Diag}\left[\sigma_{i0}^{-2}\boldsymbol{\Sigma}_{ii},\ i=1,2,\ldots,N\right]. \tag{33}$$

Again by Lemma 5 of the online appendix A, we have that

$$\left[\frac{1}{\sqrt{T}}\boldsymbol{s}_T\left(\boldsymbol{\theta}_0\right)\right] \rightarrow_d N\left[\mathbf{0}, \mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\right]$$

where

$$\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right) = \lim_{T\to\infty}\begin{pmatrix} \begin{array}{c}\left(\mathbf{G}_0 \odot \mathbf{G}_0'\right) + \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}\left(\boldsymbol{G}_0\boldsymbol{\Sigma}_0\boldsymbol{G}_0'\right) + \boldsymbol{\Delta}_{\boldsymbol{\beta}_0} \\ +\left(\gamma-2\right)\mathrm{Diag}\left(\mathbf{G}_0\odot\mathbf{G}_0'\right)\end{array} & \mathbb{E}_{\boldsymbol{\beta}_0} & \frac{\gamma}{2}\boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}\left(\mathbf{G}_0\right) \\ \mathbb{E}_{\boldsymbol{\beta}_0} & \boldsymbol{Z}_0 & \mathbf{0} \\ \frac{\gamma}{2}\boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}\left(\mathbf{G}_0\right) & \mathbf{0}' & \frac{\gamma}{4}\boldsymbol{\Sigma}_0^{-2} \end{pmatrix} \tag{34}$$

and

$$\gamma = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} Var(\zeta_{it}^2) = \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\left[E\left(\zeta_{it}^4\right)-1\right], \tag{35}$$

with $\zeta_{it} = \varepsilon_{it}/\sigma_{i0}$, for $i=1,2,\ldots,N$. Hence,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \rightarrow_d N\left(\mathbf{0}, \mathbf{V}_{\boldsymbol{\theta}}\right), \tag{36}$$

where $\mathbf{V}_{\boldsymbol{\theta}}$ has the usual sandwich formula

$$\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}_0)\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\mathbf{H}^{-1}(\boldsymbol{\theta}_0). \tag{37}$$

In the case where the errors, ε_{it}, are Gaussian, $\gamma = 2$ and, as to be expected, $\boldsymbol{H}\left(\boldsymbol{\theta}_0\right) = \mathbf{J}\left(\boldsymbol{\theta}_0, 2\right)$. This is easily verified by referring back to (30) which is equal to $\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)$ defined by (34) for $\gamma = 2$, as required.

Remark 7 *When no exogenous regressors are included in the HSAR specification (1), then the* *asymptotic variance,* $\mathbf{V}_{\boldsymbol{\theta}} = \boldsymbol{H}^{-\mathbf{1}}(\boldsymbol{\theta}_0)\,\mathbf{J}(\boldsymbol{\theta}_0,\gamma)\,\boldsymbol{H}^{-\mathbf{1}}(\boldsymbol{\theta}_0)$, *simplifies so that:*

$$\boldsymbol{H}(\boldsymbol{\theta}_0) = \begin{pmatrix} (\boldsymbol{G}_0 \odot \boldsymbol{G}_0') + \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}(\boldsymbol{G}_0\boldsymbol{\Sigma}_0\boldsymbol{G}_0') & \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}(\mathbf{G}_0) \\ \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}(\mathbf{G}_0) & \frac{1}{2}\boldsymbol{\Sigma}_0^{-2} \end{pmatrix}_{2N\times 2N}, \tag{38}$$

and

$$\mathbf{J}(\boldsymbol{\theta}_0,\gamma) = \begin{pmatrix} \begin{matrix} (\boldsymbol{G}_0 \odot \boldsymbol{G}_0') + \boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}(\boldsymbol{G}_0\boldsymbol{\Sigma}_0\boldsymbol{G}_0') \\ +(\gamma-2)\,\mathrm{Diag}(\boldsymbol{G}_0 \odot \boldsymbol{G}_0') \end{matrix} & \frac{\gamma}{2}\boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}(\mathbf{G}_0) \\ \frac{\gamma}{2}\boldsymbol{\Sigma}_0^{-1}\,\mathrm{Diag}(\mathbf{G}_0) & \frac{\gamma}{4}\boldsymbol{\Sigma}_0^{-2} \end{pmatrix}. \tag{39}$$

Again, under Gaussian errors we have $\mathbf{J}(\boldsymbol{\theta}_0,2) = \boldsymbol{H}(\boldsymbol{\theta}_0)$.

The main result of this section is summarized in the following proposition:

Proposition 2 *Consider the heterogeneous spatial autoregressive (HSAR) model given by (1). Suppose that Assumptions 1, 2, 4, 5, and conditions (21) and (28) hold. Denote the* $N(k+2)$ *dimensional (quasi-) maximum likelihood estimator of* $\boldsymbol{\theta}_0$ *by* $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \bar{\ell}_T(\boldsymbol{\theta})$, *where* $\bar{\ell}_T(\boldsymbol{\theta})$ *is given by (25). Then,* $\hat{\boldsymbol{\theta}}$ *is almost surely locally consistent for* $\boldsymbol{\theta}_0$ *on* $\boldsymbol{\Theta}_c$, *and has the following asymptotic distribution*

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \to_d N(\mathbf{0}, \mathbf{V}_{\theta}), \tag{40}$$

where $\mathbf{V}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}(\boldsymbol{\theta}_0)\mathbf{J}(\boldsymbol{\theta}_0,\gamma)\,\mathbf{H}^{-1}(\boldsymbol{\theta}_0)$, *and* $\boldsymbol{H}(\boldsymbol{\theta}_0)$ *and* $\mathbf{J}(\boldsymbol{\theta}_0,\gamma)$ *are defined by (30) and (34), respectively.*

Proof. See the online appendix B. ■

Focusing on the inverse of the information matrix, we partition $\boldsymbol{H}(\boldsymbol{\theta}_0)$ as follows

$$\boldsymbol{H}(\boldsymbol{\theta}_0) = \begin{pmatrix} \boldsymbol{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{12}' & \mathcal{H}_{22} \end{pmatrix}, \tag{41}$$

where $\mathcal{H}_{12} = (\boldsymbol{H}_{12}, \boldsymbol{H}_{13})$ is an $N \times (Nk+N)$ matrix, and since $\boldsymbol{H}_{23} = \boldsymbol{H}_{32} = \mathbf{0}$, then $\mathcal{H}_{22} = \mathrm{Diag}(\boldsymbol{H}_{22}, \boldsymbol{H}_{33})$, which is an $(Nk+N)\times(Nk+N)$ matrix. Then,

$$\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0) = \begin{pmatrix} \mathcal{H}_{11\cdot 2}^{-1} & -\mathcal{H}_{11\cdot 2}^{-1}\mathcal{H}_{12}\mathcal{H}_{22}^{-1} \\ -\mathcal{H}_{22}^{-1}\mathcal{H}_{21}\mathcal{H}_{11\cdot 2}^{-1} & \mathcal{H}_{22}^{-1} + \mathcal{H}_{22}^{-1}\mathcal{H}_{21}\mathcal{H}_{11\cdot 2}^{-1}\mathcal{H}_{12}\mathcal{H}_{22}^{-1} \end{pmatrix}.$$

Of interest is matrix $\mathcal{H}_{11\cdot 2}$ given by

$$\begin{aligned} \mathcal{H}_{11\cdot 2} &= \boldsymbol{H}_{11} - \mathcal{H}_{12}\mathcal{H}_{22}^{-1}\mathcal{H}_{21} = \boldsymbol{H}_{11} - \boldsymbol{H}_{12}\boldsymbol{H}_{22}^{-1}\boldsymbol{H}_{21} - \boldsymbol{H}_{13}\boldsymbol{H}_{33}^{-1}\boldsymbol{H}_{31} \\ &= \left(\mathbf{G}_0 \odot \mathbf{G}_0'\right) + \mathrm{Diag}\left[-g_{0,ii}^2 + \sum_{s=1,s\neq i}^{N}\left(\sigma_{s0}^2/\sigma_{i0}^2\right) g_{0,is}^2,\ i = 1,2,\ldots,N\right] \\ &\quad + \mathrm{Diag}\left[\sigma_{i0}^{-2}\sum_{r=1}^{N}\sum_{s=1}^{N} g_{0,is}g_{0,ir}\boldsymbol{\beta}_{r0}'\left(\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri}\boldsymbol{\Sigma}_{ii}^{-1}\boldsymbol{\Sigma}_{is}\right)\boldsymbol{\beta}_{s0},\ i = 1,2,\ldots,N\right], \end{aligned} \tag{42}$$

since its inverse, $\mathcal{H}_{11\cdot 2}^{-1}$, represents the asymptotic covariance matrix of $\sqrt{T}\hat{\boldsymbol{\psi}}$ under normality of the error term. This last result can be summarized in the following corollary:

Corollary 1 *Consider the heterogeneous spatial autoregressive (HSAR) model given by (1). Suppose that Assumptions 1, 2, 4, 5, and conditions (21) and (28) hold. Then, the $N \times N$ information matrix*

$$\mathcal{H}_{11\cdot 2} = \left(\mathbf{G}_0 \odot \mathbf{G}_0'\right) + \mathrm{Diag}\left[-g_{0,ii}^2 + \sum_{s=1, s\neq i}^{N} \left(\sigma_{s0}^2/\sigma_{i0}^2\right) g_{0,is}^2,\ i = 1, 2, \ldots, N\right] \tag{43}$$
$$+ \mathrm{Diag}\left[\sigma_{i0}^{-2} \sum_{r=1}^{N} \sum_{s=1}^{N} g_{0,is} g_{0,ir} \boldsymbol{\beta}_{r0}' \left(\boldsymbol{\Sigma}_{rs} - \boldsymbol{\Sigma}_{ri} \boldsymbol{\Sigma}_{ii}^{-1} \boldsymbol{\Sigma}_{is}\right) \boldsymbol{\beta}_{s0},\ i = 1, 2, \ldots, N\right],$$

is full rank, where $\mathbf{G}_0 = \boldsymbol{W}\left(\boldsymbol{I}_N - \boldsymbol{\Psi}_0 \boldsymbol{W}\right)^{-1} = (g_{0,ij})$, $\boldsymbol{\Psi}_0 = \mathrm{Diag}(\boldsymbol{\psi}_0)$, $\boldsymbol{\psi}_0 = (\psi_{10}, \psi_{20}, \ldots, \psi_{N0})'$, and $\boldsymbol{W}$ is the spatial weight matrix, and $\varepsilon_{it} \sim IIDN(0, \sigma_{i0}^2)$. Then the maximum likelihood estimator of $\boldsymbol{\psi}_0$, denoted by $\hat{\boldsymbol{\psi}}$ and computed by maximizing (A.21), has the following asymptotic distribution,

$$\sqrt{T}\left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right) \rightarrow_d N\left(0, \mathbf{V}_{\boldsymbol{\psi}}\right), \tag{44}$$

where

$$\mathbf{V}_{\boldsymbol{\psi}} = \left[\mathcal{H}_{11\cdot 2}\right]^{-1}. \tag{45}$$

Remark 8 *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 (43) vanishes and we have*

$$\mathcal{H}_{11\cdot 2} = \left(\mathbf{G}_0 \odot \mathbf{G}_0'\right) + \mathrm{Diag}\left[-g_{0,ii}^2 + \sum_{s=1, s\neq i}^{N} \left(\sigma_{s0}^2/\sigma_{i0}^2\right) g_{0,is}^2,\ i = 1, 2, \ldots, N\right], \tag{46}$$

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

Remark 9 *In the case where ε_{it} are non-Gaussian but $E(|\varepsilon_{it}|^{4+\epsilon}) < K$ holds for some $\epsilon > 0$, the quasi maximum likelihood estimator, $\hat{\boldsymbol{\psi}}$, continues to be normally distributed but its asymptotic covariance matrix is given by the upper $N \times N$ partition of $\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\, \mathbf{J}(\boldsymbol{\theta}_0, \gamma)\, \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)$, where $\boldsymbol{H}(\boldsymbol{\theta}_0)$ and $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ are defined by (30) and (34), respectively. Recall that γ is defined by (35), and under Gaussian errors it takes the value of $\gamma = 2$, so that we have $\mathbf{J}(\boldsymbol{\theta}_0, 2) = \boldsymbol{H}(\boldsymbol{\theta}_0)$.*

3.3.1 Consistent estimation of $\boldsymbol{V}_\theta$

The asymptotic covariance matrix of $\hat{\boldsymbol{\theta}}$ can be consistently estimated using the expressions given by (30) and (34), yielding the standard formula

$$\boldsymbol{V}_\theta = \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0), \tag{47}$$

when the information matrix equality holds in the case of $\varepsilon_{it} \sim IIDN(0, \sigma_{i0}^2)$ and $\gamma = 2$, and the sandwich formula

$$\boldsymbol{V}_\theta = \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\, \mathbf{J}(\boldsymbol{\theta}_0, \gamma)\, \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0), \tag{48}$$

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

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

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

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

$$\hat{\boldsymbol{V}}_{\hat{\theta}} = \hat{\boldsymbol{H}}_T^{-1} \left(\hat{\boldsymbol{\theta}} \right), \tag{49}$$

and

$$\hat{\boldsymbol{V}}_{\hat{\theta}} = \hat{\boldsymbol{H}}_T^{-1} \left(\hat{\boldsymbol{\theta}} \right) \hat{\mathbf{J}}_T \left(\hat{\boldsymbol{\theta}}, \hat{\gamma} \right) \hat{\boldsymbol{H}}_T^{-1} \left(\hat{\boldsymbol{\theta}} \right), \tag{50}$$

where $\hat{\mathbf{J}}_T(\boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)'$, $\ell_t(\boldsymbol{\theta})$ is defined by (A.20) and $\hat{\boldsymbol{H}}_T(\boldsymbol{\theta}) = -\frac{1}{T} \frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$. Consistency of $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ for $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ follows from consistency of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, of $\hat{\gamma}$ for γ and the independence of $\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ over t, as shown in Lemma 5 of the online appendix A. The first and second derivatives are provided in the online appendix C.

4 Mean group estimators

So far we have focussed on estimation of the unit-specific parameters and have derived the asymptotic results for a given N and as $T \to \infty$. But in practice it is often of interest to obtain average estimates across all the units or a sub-group of the units in the panel, assuming that the individual coefficients follow a random coefficient model. In the context of the HSAR model, (1), suppose that $\{\psi_{i0}, \boldsymbol{\beta}_{i0}, i = 1, 2, \ldots, N\}$ are randomly distributed around the common means, ψ_0 and $\boldsymbol{\beta}_0$, such that

$$\psi_{i0} = \psi_0 + \eta_{i\psi}, \text{ and } \boldsymbol{\beta}_{i0} = \boldsymbol{\beta}_0 + \boldsymbol{\eta}_{i\beta} \text{ for } i = 1, 2, \ldots, N, \tag{51}$$

where $\boldsymbol{\eta}_i = \left(\eta_{i\psi}, \boldsymbol{\eta}_{i\beta}' \right) \sim IID(\mathbf{0}, \boldsymbol{\Omega}_\eta)$, $\boldsymbol{\Omega}_\eta > 0$ is a positive definite matrix, and it is assumed that $E \|\boldsymbol{\eta}_i\|^{2+c} < K$, for some $c > 0$. The parameters of interest are ψ_0 and $\boldsymbol{\beta}_0$ which are the population means of spatial lags and slope parameters of the underlying HSAR model. For consistent estimation of ψ_0 and $\boldsymbol{\beta}_0$ we now need N and T sufficiently large. Large T is required to consistently estimate the unit-specific coefficients, and large N is required for estimation of the common means, ψ_0 and $\boldsymbol{\beta}_0$. It is also possible to apply this procedure to subsets of the units, so long as the number of units in each set is reasonably large.

Consistent estimators of ψ_0 and $\boldsymbol{\beta}_0$ are given by the mean group (MG) estimators,

$$\hat{\psi}_{MG} = N^{-1} \sum_{i=1}^{N} \hat{\psi}_i, \text{ and } \hat{\boldsymbol{\beta}}_{MG} = N^{-1} \sum_{i=1}^{N} \hat{\boldsymbol{\beta}}_i, \tag{52}$$

where $\hat{\psi}_i$ and $\hat{\boldsymbol{\beta}}_i$ are the underlying unit-specific estimators. The MG estimator was originally developed by Pesaran and Smith (1995) who show that in the standard case where $\hat{\psi}_i$ and $\hat{\boldsymbol{\beta}}_i$ are independently distributed, then $\hat{\psi}_{MG}$ and $\hat{\boldsymbol{\beta}}_{MG}$ will be consistent and asymptotically normal. Recently, Chudik and Pesaran (2019) extend this analysis and consider MG estimators based on possibly cross correlated estimators and show that the standard MG estimation continues to apply so long as the underlying unit-specific estimators are weakly cross correlated.

In the case of the present application the asymptotic validity of the MG estimator can be established by first noting that

$$\sqrt{N}\left(\hat{\psi}_{MG}-\psi_0\right)=N^{-1/2}\sum_{i=1}^{N}\left(\hat{\psi}_i-\psi_{i0}\right)+N^{-1/2}\sum_{i=1}^{N}\left(\psi_{i0}-\psi_0\right),$$

where upon using (51) can also be written as

$$\begin{aligned}\sqrt{N}\left(\hat{\psi}_{MG}-\psi_0\right)&=\frac{\sqrt{N}}{T}\left[N^{-1}\sum_{i=1}^{N}T\left(\hat{\psi}_i-\psi_{i0}\right)\right]+N^{-1/2}\sum_{i=1}^{N}\eta_{i\psi}\\&=q_{NT}+\xi_{NT}.\end{aligned}\tag{53}$$

Consider now q_{NT}, the first term of the above expression, and recall that under the regularity conditions set out in the previous sections, $\sqrt{T}\left(\hat{\psi}_i-\psi_{i0}\right)\overset{a}{\sim}N(0,\omega^2_{\psi_i})$, where $\sup_i \omega^2_{\psi_i}<K$, and $E\left(\hat{\psi}_i-\psi_{i0}\right)=O\left(T^{-1}\right)$, for all i. Hence

$$E\left(q_{NT}\right)=O\left(\frac{\sqrt{N}}{T}\right).\tag{54}$$

Furthermore we note that q_{NT} can also be written as

$$q_{NT}=N^{-1/2}\sum_{i=1}^{N}\left(\hat{\psi}_i-\psi_{i0}\right)=T^{-1/2}N^{-1/2}\boldsymbol{\tau}'_N\left[\sqrt{T}\left(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}_0\right)\right],$$

where $\hat{\boldsymbol{\psi}}=(\hat{\psi}_1,\hat{\psi}_2,\ldots,\hat{\psi}_N)'$, and $\boldsymbol{\tau}_N$ is an $N\times 1$ vector of ones. Denote the $N\times N$ asymptotic covariance matrix of $\hat{\boldsymbol{\psi}}$ by $\mathbf{V}_{\boldsymbol{\psi}}=AsyVar\left[\sqrt{T}\left(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}_0\right)\right]$, then

$$\begin{aligned}\lim_{N,T\to\infty}Var\left(q_{NT}\right)&=\lim_{N,T\to\infty}Var\left[T^{-1/2}N^{-1/2}\boldsymbol{\tau}'_N\sqrt{T}\left(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}_0\right)\right]\\&=\lim_{N,T\to\infty}\left(\frac{\boldsymbol{\tau}'_N\mathbf{V}_{\boldsymbol{\psi}}\boldsymbol{\tau}_N}{NT}\right)\leq\lim_{N,T\to\infty}\left[\frac{\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\psi}}\right)}{T}\right].\end{aligned}\tag{55}$$

Suppose now that $\sqrt{N}/T\to 0$, and $\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\psi}}\right)<K$ as N and $T\to\infty$. Then using (54) and (55) it readily follows that

$$\lim_{N,T\to\infty}E\left(q_{NT}\right)=0,\text{ and }\lim_{N,T\to\infty}Var\left(q_{NT}\right)=0,$$

and hence as $\sqrt{N}/T\to 0$, $q_{NT}=o_p(1)$, and in view of (53) $\sqrt{N}\left(\hat{\psi}_{MG}-\psi_0\right)\overset{a}{\sim}\xi_{NT}=$

$N^{-1/2}\sum_{i=1}^{N}\eta_{i\psi}$. Finally, under the random coefficient model where $\{\eta_{i\psi}$, for $i=1,2,\ldots,N\}$ are assumed to be independently distributed with zero means and finite variances, $\xi_{NT}\overset{a}{\sim} N\left[0,Var\left(\eta_{i\psi}\right)\right]$, and therefore under the additional conditions, $\sqrt{N}/T\rightarrow 0$ and $\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\psi}}\right)<K$, we have (as N,$T\rightarrow\infty$, jointly):

$$\sqrt{N}\left(\hat{\psi}_{MG}-\psi_0\right)\overset{a}{\sim}N\left(0,\omega_{\psi}^2\right), \tag{56}$$

where $\omega_{\psi}^2=Var\left(\eta_{i\psi}\right)$. It is also easily seen that ω_{ψ}^2 can be consistently estimated by

$$\hat{\omega}_{\psi}^2=\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\psi}_i-\hat{\psi}_{MG}\right)^2. \tag{57}$$

Similarly,

$$\sqrt{N}\left(\hat{\boldsymbol{\beta}}_{MG}-\boldsymbol{\beta}_0\right)\overset{a}{\sim}N\left(\mathbf{0},\boldsymbol{\Omega}_{\beta}\right), \tag{58}$$

so long as $\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\beta}}\right)<K$, where $\mathbf{V}_{\boldsymbol{\beta}}$ is the $kN\times kN$ asymptotic covariance matrix of $\sqrt{T}\,\hat{\boldsymbol{\beta}}=(\sqrt{T}\hat{\boldsymbol{\beta}}_1^{\prime},\sqrt{T}\hat{\boldsymbol{\beta}}_2^{\prime},\ldots,\sqrt{T}\hat{\boldsymbol{\beta}}_N^{\prime})^{\prime}$. A consistent estimator of $\boldsymbol{\Omega}_{\beta}$ is given by

$$\hat{\boldsymbol{\Omega}}_{\beta}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\hat{\boldsymbol{\beta}}_i-\hat{\boldsymbol{\beta}}_{MG}\right)\left(\hat{\boldsymbol{\beta}}_i-\hat{\boldsymbol{\beta}}_{MG}\right)^{\prime}. \tag{59}$$

It now remains to establish conditions under which $\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\psi}}\right)<K$ and $\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\beta}}\right)<K$ hold. We first note that $\mathbf{V}_{\boldsymbol{\psi}}$ and $\mathbf{V}_{\boldsymbol{\beta}}$ are sub-matrices of $\mathbf{V}_{\boldsymbol{\theta}}$ defined by (37) which we reproduce here for convenience:

$$\mathbf{V}_{\boldsymbol{\theta}}=\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\mathbf{H}^{-1}(\boldsymbol{\theta}_0),$$

where $\boldsymbol{H}\left(\boldsymbol{\theta}_0\right)$ and $\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)$ are given by (30) and (34), respectively. Hence, it is sufficient to show that $\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\theta}}\right)$ is bounded in N. To this end we first note that

$$\left\|\mathbf{V}_{\boldsymbol{\theta}}\right\|\leq\left\|\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right\|^2\left\|\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\right\|, \tag{60}$$

where $\left\|\mathbf{A}\right\|=\lambda_{\max}^{1/2}\left(\mathbf{A}^{\prime}\mathbf{A}\right)$ is the spectral norm of $\mathbf{A}$. However, since $\mathbf{V}_{\boldsymbol{\theta}}$, $\mathbf{H}^{-1}(\boldsymbol{\theta}_0)$ and $\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)$ are symmetric matrices, then $\left\|\mathbf{V}_{\boldsymbol{\theta}}\right\|=\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\theta}}\right)$, $\left\|\mathbf{H}^{-1}(\boldsymbol{\theta}_0)\right\|^2=\lambda_{\max}^2\left[\boldsymbol{H}^{-1}\left(\boldsymbol{\theta}_0\right)\right]$, and $\left\|\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\right\|=\lambda_{\max}\left[\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\right]$, and (60) can also be written as

$$\lambda_{\max}\left(\mathbf{V}_{\boldsymbol{\theta}}\right)\leq\lambda_{\max}^2\left[\boldsymbol{H}^{-1}\left(\boldsymbol{\theta}_0\right)\right]\lambda_{\max}\left[\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\right]. \tag{61}$$

But $\lambda_{\max}\left[\boldsymbol{H}^{-1}\left(\boldsymbol{\theta}_0\right)\right]=1/\lambda_{\min}\left[\boldsymbol{H}\left(\boldsymbol{\theta}_0\right)\right]$, and under the identification conditions established in Section 3.2, we have $\lambda_{\min}\left[\boldsymbol{H}\left(\boldsymbol{\theta}_0\right)\right]>0$, which ensures that $\lambda_{\max}\left[\boldsymbol{H}^{-1}\left(\boldsymbol{\theta}_0\right)\right]<K$ is bounded in N. Finally, we note that by Theorem 5.6.9 of Horn and Johnson (1985),

$$\lambda_{\max}\left[\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\right]\leq\left\|\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)\right\|_{\infty}, \tag{62}$$

and using (34) it is easily seen that the column (row) norm of $\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)$ is dominated by matrices $\left(\boldsymbol{G}_0\odot\boldsymbol{G}_0^{\prime}\right)$, $\boldsymbol{\Delta}_{\boldsymbol{\beta}_0}$ and $\mathbb{E}_{\boldsymbol{\beta}_0}$, where the latter two matrices are diagonal. The other matrices in $\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)$, namely $\boldsymbol{\Sigma}_0$ and $\mathbf{Z}_0$, are also diagonal matrices whose elements do not vary with N.

Consider $\mathbf{\Delta}_{\boldsymbol{\beta}_0}$ defined by (31) and note that

$$\sup_i \left(\sum_{s=1}^{N} g_{0,is}^2 \sigma_{s0}^2 \right) \leq \sup_s \left(\sigma_{s0}^2 \right) \sup_i \left(\sum_{s=1}^{N} |g_{0,is}| \right) = \sup_s \left(\sigma_{s0}^2 \right) \|\boldsymbol{G}_0\|_\infty .$$

Similarly,

$$\begin{aligned} \sup_i \left| \sum_{r=1}^{N} \sum_{s=1}^{N} g_{0,is} g_{0,ir} \boldsymbol{\beta}_{r0}' \mathbf{\Sigma}_{rs} \boldsymbol{\beta}_{s0} \right| &\leq \sup_i \sum_{r=1}^{N} \sum_{s=1}^{N} |g_{0,is}| \, |g_{0,ir}| \left\| \boldsymbol{\beta}_{r0}' \mathbf{\Sigma}_{rs} \boldsymbol{\beta}_{s0} \right\| \\ &\leq \sup_{r,s} \left\| \boldsymbol{\beta}_{r0}' \mathbf{\Sigma}_{rs} \boldsymbol{\beta}_{s0} \right\| \sup_i \sum_{r=1}^{N} \sum_{s=1}^{N} |g_{0,is}| \, |g_{0,ir}| \\ &= \sup_s \|\boldsymbol{\beta}_{s0}\| \sup_r \|\boldsymbol{\beta}_{r0}\| \sup_{r,s} \|\mathbf{\Sigma}_{rs}\| \, \|\boldsymbol{G}_0\|_\infty^2 . \end{aligned}$$

However, under Assumptions 2(b) and 3 we have $\sup_s \|\boldsymbol{\beta}_{s0}\| < K$ and $\sup_{r,s} \|\mathbf{\Sigma}_{rs}\| < K$, and under Assumption 4 it follows that $\|\boldsymbol{G}_0\|_\infty < K$ (see Lemma 2 of the online appendix A). Hence, $\left\|\mathbf{\Delta}_{\boldsymbol{\beta}_0}\right\| < K$. Similarly, it is also easily established that all elements of $\mathbb{E}_{\boldsymbol{\beta}_0}$ are bounded in N. Finally, again under Assumption 4 and as shown in Lemma 2 of the online appendix A, $\|\boldsymbol{G}_0 \odot \boldsymbol{G}_0'\|_\infty < K$. Consequently, $\|\mathbf{J}(\boldsymbol{\theta}_0, \gamma)\|_\infty < K$, and in view of (62) it follows that $\lambda_{\max}[\mathbf{J}(\boldsymbol{\theta}_0, \gamma)] < K$. Using this result in (61) and recalling that $\lambda_{\max}\left[\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\right] < K$, then overall we have $\lambda_{\max}(\mathbf{V}_{\boldsymbol{\theta}}) < K$, as required. Note that this result does not need the exogenous regressors to be weakly cross-correlated; it is sufficient that $\sup_{r,s} \|\mathbf{\Sigma}_{rs}\| < K$.

The main result of this section is summarized in the following proposition:

Proposition 3 *Consider the heterogeneous spatial autoregressive (HSAR) model given by (1) where the coefficients* $\{\psi_{i0}, \boldsymbol{\beta}_{i0}, i = 1, 2, \ldots, N\}$ *are distributed randomly around the common means* ψ_0 *and* $\boldsymbol{\beta}_0$ *following (51). Suppose that Assumptions 1, 2, 4, 5, and conditions (21) and (28) hold. Then, as* $N, T \to \infty$*, jointly such that* $\sqrt{N}/T \to 0$*, the mean group estimators,* $\hat{\psi}_{MG}$ *and* $\hat{\beta}_{MG}$*, defined by (52) have the following asymptotic distributions*

$$\sqrt{N}\left(\hat{\psi}_{MG} - \psi_0\right) \overset{a}{\sim} N\left(0, \omega_\psi^2\right) \ \text{ and } \sqrt{N}\left(\hat{\boldsymbol{\beta}}_{MG} - \boldsymbol{\beta}_0\right) \overset{a}{\sim} N\left(\mathbf{0}, \mathbf{\Omega}_\beta\right),$$

with consistent estimators of ω_ψ^2 *and* $\mathbf{\Omega}_\beta$ *provided in (57) and (59), respectively.*

5 Small sample properties of the QMLE

We investigate the small sample properties of the proposed QML estimator and the associated MG estimator using Monte Carlo simulations. We consider the following data generating process (DGP)

$$y_{it} = a_i + \psi_i \sum_{j=1}^{N} w_{ij} y_{jt} + \beta_i x_{it} + \varepsilon_{it}, \ i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T. \tag{63}$$

We include one exogenous regressor, x_{it}, with coefficient β_i as well as fixed effects, a_i, in unit-specific regressions. Stacking these regressions we have

$$\boldsymbol{y}_{\circ t} = \boldsymbol{a} + \boldsymbol{\Psi}\boldsymbol{W}\boldsymbol{y}_{\circ t} + \boldsymbol{B}\boldsymbol{x}_{\circ t} + \boldsymbol{\varepsilon}_{\circ t},\ t = 1, 2, \ldots, T, \tag{64}$$

where $\boldsymbol{a} = (a_1, a_2, \ldots, a_N)'$, $\boldsymbol{\Psi} = \text{Diag}(\boldsymbol{\psi})$ and $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_N)'$, $\boldsymbol{W} = (w_{ij})$, $i, j = 1, 2, \ldots, N$, $\boldsymbol{B} = \text{Diag}(\boldsymbol{\beta})$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)'$, $\boldsymbol{x}_{\circ t} = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$, and $\boldsymbol{\varepsilon}_{\circ t} = (\varepsilon_{1t}, \varepsilon_{2t}, \ldots, \varepsilon_{Nt})'$. Note that since we explicitly account for fixed effects which we separate out from the remaining regressors included in $\boldsymbol{x}_{\circ t}$, the unknown parameters are summarized in vector $\boldsymbol{\theta}$, as follows: $\boldsymbol{\theta} = \left(\boldsymbol{a}', \boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime}\right)'$, $\boldsymbol{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2)'$. In total there are $4N$ unknown parameters.

We allow for spatial dependence in the regressors, x_{it}, and generate them as

$$x_{it} = \phi_i \boldsymbol{w}_{x,i}' \boldsymbol{x}_{\circ t} + v_{it}, \tag{65}$$

or in matrix form

$$\boldsymbol{x}_{\circ t} = \left(\mathbf{I}_N - \boldsymbol{\Phi}\boldsymbol{W}_x\right)^{-1} \boldsymbol{v}_{\circ t},$$

where $\boldsymbol{\Phi} = \text{Diag}(\phi_1, \phi_2, \ldots, \phi_N)$, and $\boldsymbol{v}_{\circ t} = (v_{1t}, v_{2t}, \ldots, v_{Nt})'$, with $v_{it} \sim IIDN(0, \sigma_v^2)$. We set $\phi_i = 0.5$ (representing a moderate degree of spatial dependence), and set

$$\sigma_v^2 = \frac{N}{\text{tr}\left[\left(\mathbf{I}_N - \boldsymbol{\Phi}\boldsymbol{W}_x\right)^{-1}\left(\mathbf{I}_N - \boldsymbol{\Phi}\boldsymbol{W}_x\right)'^{-1}\right]}, \tag{66}$$

which ensures that $N^{-1}\sum_{i=1}^{N} Var(x_{it}) = 1$. We set $\boldsymbol{W}_x = \boldsymbol{W} = (w_{ij})$, $i, j = 1, 2, \ldots, N$, and use the 4-connection spatial matrix described below.

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

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

and

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

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

For the weight matrix, $\boldsymbol{W} = (w_{ij})$, first we use contiguity criteria to generate the non-normalized weights, w_{ij}^o, then row normalize the resultant weight matrices to obtain w_{ij}. More specifically, we consider $\boldsymbol{W}$ matrices with 2, 4 and 10 connections and generate w_{ij}^o, for $i = 1, 2, \ldots, N$, as[6]

- 2 connections: $w_{i,j}^o = 1$ if $j = i-1, i+1$, and zero otherwise,
- 4 connections: $w_{i,j}^o = 1$ if $j = i-2, i-1, i+1, i+2$, and zero otherwise,
- 10 connections: $w_{i,j}^o = 1$ if $j = i-5, i-4, \ldots, i-1, i+1, i+2, \ldots, i+5$, and zero otherwise.

[6]By construction, the first and the last units have fewer neighbors as compared to the other units.

Since by construction $\|\boldsymbol{W}\|_{\infty} = 1$, then condition (8) is satisfied if $\sup_i |\psi_i| < 1$, and ensures that $\mathbf{I}_N - \boldsymbol{\Psi W}$ is invertible. We generate the unit-specific coefficients of the HSAR model as $a_{i0} \sim IIDN(1,1)$, $\beta_{i0} \sim IIDU\,(0,1)$, and $\psi_{i0} \sim IIDU(0,0.8)$, for $i = 1,2,\ldots,N$.[7] Given the DGP in (63), values of y_{it} are now generated as

$$\boldsymbol{y}_{\circ t} = \left(\mathbf{I}_N - \boldsymbol{\Psi W}\right)^{-1} \left(\boldsymbol{a} + \boldsymbol{B x}_{\circ t} + \boldsymbol{\varepsilon}_{\circ t}\right), t = 1,2,\ldots,T.$$

Initially, to illustrate that our proposed estimator applies to both cases where N is small and large, we considered the two polar cases of $N = 5$ and $N = 100$, and set $T = 25, 50, 100, 200$. We then considered a more comprehensive set of N values, namely $N = 25, 50, 75, 100$. For each experiment we used $R = 2,000$ replications. Across the replications, $\boldsymbol{\theta}_0$, and the weight matrix, $\boldsymbol{W}$, are kept fixed, whilst the errors and the regressors, ε_{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 $N = 5$. Due to the problem of simultaneity, the degree of time variations in y^*_{it} for each unit i depends on the choice of $\boldsymbol{W}$ and the number of cross section units, N. Naturally, this is reflected in the performance of the estimators and the power properties of the tests based on them.

We report bias and RMSE of the QML estimators for individual cross section units, as well as their corresponding empirical sizes. In addition, we report power functions for three units with true spatial autoregressive parameters, ψ_{i0}, selected to be low, medium and large in magnitude. The experiments are carried out for spatial weight matrices, $\boldsymbol{W}$, with two, four and ten connections. The mean of simulated parameter estimates are computed as

$$\hat{\psi}_{i(R)} = R^{-1} \sum_{r=1}^{R} \hat{\psi}_{i,r}, \text{ and } \hat{\beta}_{i(R)} = R^{-1} \sum_{r=1}^{R} \hat{\beta}_{i,r},$$

where $\hat{\psi}_{i,r}$ and $\hat{\beta}_{i,r}$ refer to the QML estimates of ψ_i and β_i in the r^{th} replication. The QML estimators are computed using the log-likelihood function (7). We also report small sample results for the MG estimators of ψ_0 and β_0, defined by (52), using the experiment described in Section 5.2 below.

5.1 Results for individual estimates

Since the results based on the Gaussian and non-Gaussian errors are very close, in what follows we only report the results for the non-Gaussian case where the errors are generated as iid $\chi^2(2)$ random variables, and use the sandwich formula (50) to compute standard errors. Also to save space, we focus on results based on the spatial weight matrix, $\boldsymbol{W}$, with four connections.[8] Initially, to highlight the applicability of the proposed estimators to small as well as large dimensional HSAR panels, we provide detailed results for the experiments with $N = 5$ and $N = 100$.

[7] We also carried out experiments without exogenous regressors with $\beta_{i0} = 0$, for all i, corresponding to model (14) in Section 3.2. The results of these experiments are available upon request.

[8] Results for Gaussian errors and other choices of spatial weight matrices are available upon request.

5.1.1 Two polar cases: $N = 5$ and $N = 100$

Table 1 reports the bias, RMSE, empirical size and power of the individual parameters, ψ_{i0} and β_{i0}, $i = 1, 2, \ldots, N$, for the experiments with $N = 5$. The bias of estimating ψ_{i0} tends to be small but negative when $T = 25$, whilst estimates of β_{i0} show an upward bias when T is small ($T = 25$). But the biases of both estimators fall quite rapidly with T, for all i. A similar pattern can be seen in the RMSEs, again declining with T reasonably fast. Turning to size and power of the tests based on the QML estimates, there is some evidence of over-rejection when T is small ($T = 25$). But the size distortion gets eliminated as T is increased, with the tests having the correct size for values of $T \geq 50$. This pattern is shared by both ψ_{i0} and β_{i0}. Similarly, power is low when $T = 25$ but improves markedly for all 5 units as T is increased.[9] Overall the small sample results are in line with our theoretical findings, and give satisfactory results for values of $T \geq 50$; a property which is repeated for other experiments considered in this paper.

For $N = 100$ we report the results only for a selected number of units, namely units with the three smallest and largest population values for ψ_{i0} and a few in between, and the associated β_{i0} values. The small sample results for these experiments are summarized in Table 2, and are qualitatively similar to those reported in Table 1 for $N = 5$, indicating that the theoretical framework of Section 3 can be applied equally to data sets with small and large numbers of cross section units.

5.1.2 RMSE, size and power for all N and T combinations

We now turn to the rest of the results and consider all the combinations of $N \in \{25, 50, 75, 100\}$ and $T \in \{25, 50, 100, 200\}$. To save space we use boxplots to summarize the results for RMSE and size, and use empirical rejection frequency plots for power.[10] All results are shown in the online appendix E. The RMSE boxplots for ψ_{i0} and β_{i0} are given in Figures A1 and A2, respectively.[11] Overall, the RMSE values are small for both parameters and fall with T but are not affected by changes in the cross section dimension, N, which is in line with the theory developed in Section 3.

The boxplots for the size of the tests based on the QML estimates of ψ_{i0} and β_{i0} are given in Figures A3 and A7, respectively. These results are based on the sandwich covariance matrix formula given by (50). 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 some size distortions. Once again the size estimates are not affected by N, and tend to 5 per cent as T increases, irrespective of the value of N.

To save space we only report the empirical power functions of the tests for three cross section units with low, medium and high parameter values. The power plots are computed for different values of ψ_i and β_i defined by $\psi_i = \psi_{i0} + \delta$, and $\beta_i = \beta_{i0} + \delta$, for $i = 1, 2, \ldots, N$, where

[9] Clearly, improvements in power can be achieved by reducing the error variances, σ_{i0}^2. Some supporting evidence is provided in Tables S1 and S2 in the online appendix E.

[10] The boxplots for bias of the estimators are similar to those of RMSE and are available upon request. The corresponding tables that show bias and RMSE results for the individuals estimates ($\hat{\psi}_{i(R)}$, and $\hat{\beta}_{i(R)}$, $i = 1, 2, \ldots, N$) are also available upon request.

[11] In each boxplot, the central mark indicates the median, while the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers. Finally, outliers are defined as values greater than $q_3 + 1.5\,(q_3 - q_1)$ or smaller than $q_1 - 1.5\,(q_3 - q_1)$, where q_1 and q_3 are the 25th and 75th percentiles, respectively.

Table 1: Bias, RMSE, size and power for parameters of individual units in the HSAR(1) model with one exogenous regressor and non-Gaussian errors for $N = 5$ and $T \in \{25, 50, 100, 200\}$.

T	25		50		100		200	
Parameter	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ψ_{i0}								
$\psi_{1,0} = 0.1261$	-0.0056	0.1891	0.0005	0.1230	-0.0023	0.0851	0.0010	0.0592
$\psi_{2,0} = 0.3883$	-0.0051	0.2495	-0.0058	0.1687	-0.0006	0.1148	-0.0003	0.0803
$\psi_{3,0} = 0.4375$	-0.0115	0.2436	-0.0022	0.1499	0.0034	0.1041	-0.0003	0.0743
$\psi_{4,0} = 0.5059$	0.0050	0.1769	-0.0040	0.1221	-0.0028	0.0820	-0.0010	0.0571
$\psi_{5,0} = 0.7246$	-0.0109	0.2089	-0.0031	0.1502	-0.0009	0.1071	0.0006	0.0721
β_{i0}								
$\beta_{1,0} = 0.9649$	0.0125	0.2236	0.0069	0.1472	0.0024	0.1008	-0.0020	0.0717
$\beta_{2,0} = 0.9572$	0.0100	0.2674	0.0068	0.1833	-0.0022	0.1272	-0.0025	0.0892
$\beta_{3,0} = 0.2785$	0.0078	0.2908	-0.0012	0.1806	-0.0026	0.1252	0.0022	0.0907
$\beta_{4,0} = 0.9134$	-0.0020	0.2195	0.0072	0.1461	0.0012	0.1000	0.0000	0.0684
$\beta_{5,0} = 0.8147$	0.0104	0.2842	0.0108	0.1950	0.0081	0.1341	0.0003	0.0911
T	25	50	100	200	25	50	100	200
Parameter	Size				Power			
ψ_{i0}								
$\psi_{1,0} = 0.1261$	0.1040	0.0675	0.0535	0.0515	0.3410	0.4665	0.7060	0.9065
$\psi_{2,0} = 0.3883$	0.0950	0.0690	0.0560	0.0580	0.2515	0.3525	0.4900	0.7315
$\psi_{3,0} = 0.4375$	0.0935	0.0620	0.0560	0.0510	0.2245	0.3355	0.5115	0.7975
$\psi_{4,0} = 0.5059$	0.0835	0.0740	0.0660	0.0485	0.3430	0.5025	0.7355	0.9345
$\psi_{5,0} = 0.7246$	0.0660	0.0670	0.0645	0.0530	0.2450	0.3610	0.5410	0.7975
β_{i0}								
$\beta_{1,0} = 0.9649$	0.0900	0.0645	0.0525	0.0530	0.2845	0.3825	0.5360	0.8075
$\beta_{2,0} = 0.9572$	0.0930	0.0725	0.0625	0.0570	0.2165	0.2885	0.4380	0.6535
$\beta_{3,0} = 0.2785$	0.0960	0.0710	0.0515	0.0585	0.2585	0.3000	0.4565	0.6375
$\beta_{4,0} = 0.9134$	0.0865	0.0630	0.0565	0.0485	0.3055	0.3845	0.5715	0.8245
$\beta_{5,0} = 0.8147$	0.0890	0.0705	0.0555	0.0510	0.2005	0.2570	0.3700	0.6080

Notes: True parameter values are generated as $\psi_{i0} \sim \mathit{IIDU}(0, 0.8)$, $\alpha_{i0} \sim \mathit{IIDN}(1, 1)$, and $\beta_{i0} \sim \mathit{IIDU}(0, 1)$ for $i = 1, 2, \dots, N$. Non-Gaussian errors are generated as $\varepsilon_{i0}/\sigma_{i0} \sim \mathit{IID}[\chi^2(2) - 2]/2$, with $\sigma^2_{i0} \sim \mathit{IIDU}[\chi^2(2)/4 + 0.5]$ for $i = 1, 2, \dots, N$. The spatial weight matrix $\boldsymbol{W} = (w_{ij})$ 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 $i = 1, 2, \dots, N$. 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 $i = 1, 2, \dots, N$. Empirical size and empirical power are based on the sandwich formula given by (50). The nominal size is set to 5%. Size is computed under $H_{i0}\colon \psi_i = \psi_{i0}$, using a two-sided alternative, for $i = 1, 2, \dots, N$. Power is computed under $\psi_i = \psi_{i0} + 0.2$, for $i = 1, 2, \dots, N$. The number of replications is set to $R = 2,000$. Estimates are sorted in ascending order according to the true values of the spatial autoregressive parameters. Biases, RMSEs, sizes and powers for β_i, $i = 1, 2, \dots, N$, are computed similarly, with power computed under $\beta_i = \beta_{i0} + 0.2$.

Table 2: Bias, RMSE, size and power for parameters of individual units in the HSAR(1) model with one exogenous regressor and non-Gaussian errors for $N = 100$ and $T \in \{25, 50, 100, 200\}$.

T	25		50		100		200	
Parameter	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ψ_{i0}								
$\psi_{1,0} = 0.0244$	-0.0005	0.3152	-0.0049	0.2138	0.0021	0.1415	-0.0001	0.1010
$\psi_{2,0} = 0.0255$	-0.0330	0.5189	0.0034	0.3674	-0.0137	0.2641	-0.0033	0.1794
$\psi_{3,0} = 0.0397$	0.0129	0.3509	-0.0017	0.2448	-0.0014	0.1681	0.0013	0.1183
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{51,0} = 0.3927$	-0.0027	0.2912	0.0038	0.2056	0.0009	0.1395	0.0005	0.0960
$\psi_{52,0} = 0.3987$	0.0001	0.1994	-0.0031	0.1381	0.0029	0.0921	0.0005	0.0638
$\psi_{53,0} = 0.4004$	-0.0112	0.3063	0.0075	0.2049	0.0033	0.1392	-0.0015	0.0991
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{98,0} = 0.7695$	-0.0031	0.1621	0.0018	0.1149	0.0055	0.0824	-0.0003	0.0586
$\psi_{99,0} = 0.7705$	-0.0530	0.2903	-0.0126	0.1895	0.0002	0.1401	0.0003	0.1041
$\psi_{100,0} = 0.7904$	-0.0125	0.1716	-0.0094	0.1275	0.0011	0.0897	0.0008	0.0613
β_{i0}								
$\beta_{1,0} = 0.1978$	0.0089	0.2782	0.0017	0.1771	0.0007	0.1192	-0.0073	0.0824
$\beta_{2,0} = 0.7060$	0.0252	0.3699	0.0016	0.2359	-0.0005	0.1608	0.0049	0.1144
$\beta_{3,0} = 0.4173$	0.0107	0.2541	0.0034	0.1733	0.0000	0.1157	0.0028	0.0821
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{51,0} = 0.9448$	0.0060	0.1924	-0.0024	0.1294	0.0018	0.0896	0.0009	0.0634
$\beta_{52,0} = 0.1190$	0.0046	0.1824	0.0026	0.1259	-0.0005	0.0853	0.0021	0.0578
$\beta_{53,0} = 0.7127$	0.0026	0.2630	-0.0050	0.1654	0.0038	0.1201	0.0012	0.0831
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{98,0} = 0.1067$	0.0041	0.1688	-0.0031	0.1115	0.0010	0.0762	-0.0002	0.0550
$\beta_{99,0} = 0.4588$	0.0207	0.2643	0.0039	0.1788	0.0033	0.1232	0.0027	0.0888
$\beta_{100,0} = 0.3674$	0.0056	0.1691	0.0032	0.1179	0.0009	0.0830	0.0004	0.0560

T	25	50	100	200	25	50	100	200
Parameter	Size				Power			
ψ_{i0}								
$\psi_{1,0} = 0.0244$	0.0890	0.0810	0.0520	0.0590	0.1820	0.2200	0.3290	0.5480
$\psi_{2,0} = 0.0255$	0.0705	0.0595	0.0555	0.0490	0.0945	0.0895	0.1495	0.2140
$\psi_{3,0} = 0.0397$	0.0905	0.0745	0.0585	0.0575	0.1555	0.1895	0.2805	0.4450
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{51,0} = 0.3927$	0.0950	0.0645	0.0590	0.0535	0.1785	0.2625	0.3590	0.5810
$\psi_{52,0} = 0.3987$	0.0850	0.0620	0.0625	0.0505	0.3050	0.4390	0.6285	0.8660
$\psi_{53,0} = 0.4004$	0.0885	0.0785	0.0570	0.0585	0.1995	0.2490	0.3745	0.5800
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{98,0} = 0.7695$	0.0635	0.0630	0.0660	0.0610	0.3340	0.4755	0.6935	0.9145
$\psi_{99,0} = 0.7705$	0.0300	0.0285	0.0370	0.0495	0.1455	0.2045	0.3095	0.5205
$\psi_{100,0} = 0.7904$	0.0545	0.0570	0.0625	0.0505	0.3120	0.4665	0.6455	0.8845
β_{i0}								
$\beta_{1,0} = 0.1978$	0.1160	0.0700	0.0610	0.0505	0.2405	0.3040	0.4380	0.7115
$\beta_{2,0} = 0.7060$	0.1025	0.0580	0.0505	0.0545	0.1725	0.2095	0.2710	0.4510
$\beta_{3,0} = 0.4173$	0.0950	0.0780	0.0550	0.0595	0.2450	0.3160	0.4655	0.7080
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{51,0} = 0.9448$	0.0910	0.0685	0.0590	0.0570	0.3260	0.4665	0.6500	0.8880
$\beta_{52,0} = 0.1190$	0.0970	0.0800	0.0505	0.0440	0.3500	0.4840	0.7030	0.9150
$\beta_{53,0} = 0.7127$	0.1075	0.0660	0.0665	0.0515	0.2420	0.3150	0.4410	0.6810
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{98,0} = 0.1067$	0.0960	0.0660	0.0530	0.0545	0.3950	0.5500	0.7605	0.9475
$\beta_{99,0} = 0.4588$	0.0725	0.0615	0.0545	0.0595	0.2015	0.2775	0.4225	0.6415
$\beta_{100,0} = 0.3674$	0.0935	0.0660	0.0695	0.0540	0.3605	0.5025	0.7255	0.9370

Notes: See notes to Table 1.

$\delta = -0.800, -0.791, \ldots, 0.791, 0.800$. We only consider values of ψ_i that satisfy the condition $|\psi_i| < 1$.[12]

The power results for the spatial parameters, ψ_{i0}, are displayed in Figures A4-A6, that correspond to the low value ($\psi_{i0} = 0.3374$), the medium value ($\psi_{i0} = 0.5059$) and the high value ($\psi_{i0} = 0.7676$), respectively. As to be expected the power depends on the choice of ψ_{i0} and rises with T, but does not seem to be affected by N. Furthermore, perhaps not surprisingly, empirical power functions for ψ_{i0} become more and more asymmetrical as ψ_{i0}'s move closer and closer to the boundary value of 1. The power functions for the three associated values of β_{i0} are shown in Figures A8-A10 for the low value of β_{i0} ($\beta_{i0} = 0.0344$), the medium value ($\beta_{i0} = 0.4898$), and the high value ($\beta_{i0} = 0.9649$), respectively. Again the empirical power functions are similar across N and improve with T.

5.2 Small sample properties of the MG estimators

We employ the same data generating process, defined by (63), and set $a_{i0} = a_0 + \epsilon_{1i}$, with $a_0 = 1$ and $\epsilon_{1i} \sim IIDN(0,1)$, $\psi_{i0} = \psi_0 + \epsilon_{2i}$, with $\psi_0 = 0.4$ and $\epsilon_{2i} \sim IIDU(-0.4, 0.4)$ and $\beta_{i0} = \beta_0 + \epsilon_{3i}$, with $\beta_0 = 0.5$ and $\epsilon_{3i} \sim IIDU\,(-0.5, 0.5)$. Parameters a_0, ψ_0 and β_0 are fixed while parameters a_{i0}, ψ_{i0} and β_{i0} vary across replications, for $i = 1, 2, \ldots, N$, in accordance to the random coefficients model. The MG estimators and their standard errors are computed using (52), (57) and (59), and the number of replications is set to $R = 2,000$. The small sample properties of the mean group estimators of ψ_0 and β_0 are summarized in Table B of the online appendix E. The top panel gives the results for Gaussian errors, and the bottom panel for non-Gaussian errors. As to be expected the bias and RMSE of the MG estimators decline steadily with both N and T, and it does not matter whether the errors are Gaussian or not. There are some small size distortions when $N = T = 25$, but the size rapidly converges to the nominal value of 5 percent as N and T are increased. For example for $T = 25$ the size is always within the simulation standard errors when $N \geq 50$.

6 Heterogeneous spatial spill-over effects in U.S. housing market

As an empirical application we estimate HSAR models for real house price changes in the United States at Metropolitan Statistical Areas (MSAs) over the period 1975Q1-2014Q4. Accurately modelling and forecasting the housing market cycle is of paramount importance for prospective owners, investors, and real estate market participants such as insurers and mortgage lenders (Agnello et al., 2015). Determinants of US house price variations are numerous and well-documented in the literature, two prominent fundamentals being real per capita disposable income and population - see for example Malpezzi (1999) and Gallin (2006) among others. An important aspect of the modelling strategy is to account for the existence of co-movements in house prices within and across MSAs. Recently, Bailey et al. (2016) (hereafter BHP) highlight the importance of distinguishing between types of cross-sectional dependence in the analysis of US house price changes, which if ignored can lead to biased parameter estimates. See, for example, the studies by Swoboda et al. (2015) and Munro (2018). BHP distinguish between spatial

[12]The empirical power functions are computed using the sandwich formula for the covariance matrix of the underlying estimators.

dependence that originates from economy-wide common shocks such as changes in interest rates, oil prices and technology, and the dependence across MSAs due to local spill-over effects arising from differences in house prices, incomes and demographics across MSAs.[13] Here, we use an extended version of the panel dataset employed by BHP and further augmented with population and per capita real income data by Yang (2018) to estimate HSAR models, after filtering out the effects of common factors on house price changes.[14] We provide MSA specific estimates of spill-over effects, as well as population and income elasticities of house prices as compared to the homogeneous spatial parameter estimates obtained in Yang (2018). To compare our individual estimates with those of Yang, we also report MG estimates both at the national and regional levels. As we shall see, we find considerable heterogeneity across MSAs and regions.

6.1 Data description and transformations

The U.S. Office of Management and Budget (OMB) delineates metropolitan statistical areas (MSAs) according to published standards that are applied to Census Bureau data. These are revised periodically. A total of 381 MSAs fall under the February 2013 definition.[15] We consider 377 of these from the contiguous United States.[16] Accordingly we compile quarterly nominal house prices (HP) for 377 MSAs over the period 1975Q1-2014Q4. In addition, we obtained nominal income per capita (INC) and population (POP) at the MSA level over the same period. Both real house prices and real per capita income for all MSAs are then computed by deflating their nominal values by State level Consumer Price Index data (CPI) which are matched to the corresponding MSAs.[17] Further details on data sources can be found in the online appendix D.

We denote the variables that are included in our model by: Π_{it} for percent quarterly rate of change of real house prices of MSA i in quarter t (dependent variable), $GPOP_{it}$ for percent quarterly rate of change of population (regressor), and $GINC_{it}$ for percent quarterly rate of change in real per capita income (regressor). Specifically,

$$\begin{aligned}
\Pi_{it} &= 100 \times \left[\ln\left(\frac{HP_{it}}{CPI_{it}}\right) - \ln\left(\frac{HP_{it-1}}{CPI_{it-1}}\right)\right], \\
GPOP_{it} &= 100 \times \left[\ln\left(POP_{it}\right) - \ln\left(POP_{it-1}\right)\right], \text{ and} \\
GINC_{it} &= 100 \times \left[\ln\left(\frac{INC_{it}}{CPI_{it}}\right) - \ln\left(\frac{INC_{it-1}}{CPI_{it-1}}\right)\right],
\end{aligned}$$

[13]For a theoretical analysis of the interactions between regional house prices, migration flows and income shocks see Cun and Pesaran (2018).

[14]The authors would like to thank Cynthia Yang for providing them with the updated dataset originally used in Yang (2018).

[15]The February 2013 delineation states that 'metropolitan statistical areas have at least one urbanised area of 50,000 or more population, plus adjacent territory that has a high degree of social and economic integration with the core as measured by commuting ties'. For further details see:
https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/bulletins/2013/b13-01.pdf

[16]This excludes the non-contiguous states of Alaska (2 MSAs) and Hawaii (2MSAs) and all other off-shore insular areas.

[17]The quarterly figures for nominal house prices (HP) are arithmetic averages of monthly observations of HP. Further, per capita income (INC), population (POP) and consumer price index (CPI) are annual data which are converted into quarterly observations by following the interpolation method provided in the GVAR Toolbox User Guide which can be found at: https://sites.google.com/site/gvarmodelling/gvar-toolbox.

for $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$ ($1975Q1 - 2014Q4$), where $N = 377$ MSAs, and $T = 160$ quarters.

Implementation of our approach requires the panel of variable Π_{it} to be weakly cross-sectionally dependent by Assumption 4. Hence, we apply the CD test developed in Pesaran (2004, 2015) to Π_{it} in order to assess the strength of cross-sectional dependence (CSD) in real house price changes. The CD statistic turns out to be 1621.22 which is substantially higher than the 1.96 critical value at 5 per cent level. With the null hypothesis of weak CSD soundly rejected, we then estimated the exponent of cross-sectional dependence, α, due to Bailey et al. (2016) which measures the degree of cross sectional dependence of house price changes. Values of α close to unity are indicative of strong cross-sectional dependence. We obtained $\hat{\alpha} = 1.001(0.03)$, where the standard error of the estimate is given in brackets. It is clear that real house prices changes, Π_{it}, are strongly correlated across MSAs, and before estimating local spill-over effects using the HSAR model, we must first purge the house price inflation series of the common sources of their dependence, as suggested in BHP.

Accordingly, we de-seasonalize and de-factor the three variables that we use to estimate the HSAR specifications, and use residuals from OLS regressions of Π_{it}, $GPOP_{it}$ and $GINC_{it}$ on: (i) an intercept, (ii) 3 quarterly dummies and (iii) national and regional cross-sectional averages of Π_{it}, $GPOP_{it}$ and $GINC_{it}$ respectively.[18] We denote these de-seasonalized and de-factored variables by π_{it}, $gpop_{it}$ and $ginc_{it}$, respectively.[19] The CD statistic for the filtered series π_{it} now stands at -0.200, which is not statistically significant and π_{it} satisfy the condition of weak cross-section dependence required when estimating HSAR models.

6.2 Modelling de-factored house price changes

We now consider the following HSAR specification for π_{it}:

$$\pi_{it} = a_i + \psi_i \sum_{j=1}^{N} w_{ij}\pi_{jt} + \beta_{1i}^{pop}\ gpop_{it} + \beta_{2i}^{pop}\ gpop_{i,t-1} + \beta_{1i}^{inc} ginc_{it} + \beta_{2i}^{inc} ginc_{i,t-1} + \varepsilon_{it}, \quad (67)$$

which allows for fixed effects and full heterogeneity in both the spatial coefficients of real house price changes (ψ_i), and the slope coefficients for the two regressors and their lagged values ($\beta_{1i}^{pop}, \beta_{2i}^{pop}, \beta_{1i}^{inc}, \beta_{2i}^{inc}$). Innovations are assumed to be distributed as $\varepsilon_{it} \sim IID\left(0, \sigma_{i\varepsilon}^2\right)$.[20] (67) is in accordance with the theoretical model (1) analyzed in Sections 2 and 3.[21] With regard to the construction of the weights matrix $\boldsymbol{W} = (w_{ij})$, we consider a distance based weighting

[18] We partition the MSAs into $R = 8$ regions, in line with the Bureau of Economic Analysis classification, each region $r = 1, 2, \ldots, R$, containing a total of N_r MSAs. The eight regions are: New England (15 MSAs), Mid East (41 MSAs), South East (120 MSAs), Great Lakes (59 MSAs), Plains (33 MSAs), South West (39 MSAs), Rocky Mountains (22 MSAs) and Far West (48 MSAs).

[19] This transformation of the data follows Yang (2018). She also includes local cross-sectional averages of house price changes in her defactoring procedure. Given their limited explanatory power we abstract from incorporating local averages when defactoring the series.

[20] In performing the data transformations of Section 6.1, we abstract from the sampling uncertainty related to using defactored series when estimating HSAR models. In principle, one could estimate the common and local effects simultaneously, instead of the two-stage procedure being followed. However, such an endeavour is beyond the scope of the present paper.

[21] We have considered alternative models to (67): one assuming no time lags in the exogenous variables and another that allows for lagged dependent variables as well as lagged regressors. Overall, the results convey the same message as that from running regression (67). For brevity of exposition, these results are not included in the paper, but are available upon request.

scheme implemented in Yang (2018), which is common in the spatial econometrics literature. More precisely, the calculation of the geodesic distance between each pair of latitude/longitude coordinates for the MSAs included in our sample uses the Haversine formula. Then, we determine a specific radius threshold, d (miles), within which MSAs are considered to be neighbors. In this case, the relevant entries in the un-normalized weights matrix $\boldsymbol{W}^0$ are set to unity. The MSAs that fall outside this radius are labelled non-neighbors and their corresponding entries in $\boldsymbol{W}^0$ are set to zero. Finally, we row-normalize $\boldsymbol{W}^0$ and obtain $\boldsymbol{W}$ which is used in (67).

We consider three versions of $\boldsymbol{W}$ constructed with the radius threshold values of $d = 75$, 100 and 125, miles. We name the adjacency matrices $\boldsymbol{W}_{75}$, $\boldsymbol{W}_{100}$ and $\boldsymbol{W}_{125}$, respectively. For brevity of exposition, in what follows we focus on the version of (67) that uses $\boldsymbol{W}_{75}$ which gives a reasonably sparse weight matrix with 0.88% non-zero elements. Other types of weighting schemes can also be entertained. For example, BHP consider two separate adjacency matrices determined by the statistically positive and negative pairwise correlations of de-factored real house price changes. Another scheme is proposed by Zhou et al. (2017) who use a sample-based adjacency matrix to approximate the true network structure by focusing on an estimation framework that incorporates just the degree (number of connections) of each unit in the network. Since our primary focus is on the estimation of heterogeneous spatial coefficients, we do not consider such alternative weigh matrices, which can be easily pursued if needed.

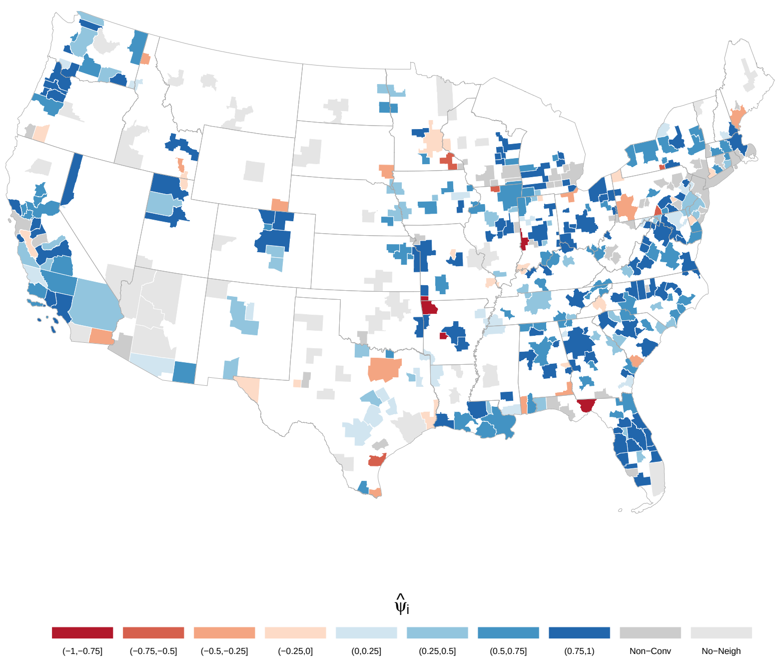
6.3 Estimation results

First we present the estimates of individual spatial effects by MSA. Note that when using adjacency matrix $\boldsymbol{W}_{75}$ in (67) there are 39 out of the total 377 MSAs that are completely isolated (have no neighbors) and are thus excluded from the analysis. This leaves us with a reduced sample of $N = 338$ MSAs. For ease of exposition the individual spatial lag coefficient estimates for these 338 MSAs are displayed in Figure 1. Each estimate, $\hat{\psi}_i$, is matched to its corresponding MSA on the map of the U.S.. MSAs colored in blue depict positive spatial lag coefficients, with different shades of blue corresponding to differing ranges within which each $\hat{\psi}_i$ falls: lighter shades refer to ranges closer to zero while darker shades relate to spatial lag coefficient estimates closer to the boundary value of unity. Similarly, red areas are associated with negative spatial lag coefficient estimates, with the lighter shade of red indicating $\hat{\psi}_i$ falling in ranges closer to zero while darker red areas refer to more sizable spatial coefficient estimates in absolute terms.[22]

It is evident from Figure 1 that spatial coefficients are estimated to be predominantly positive and in general relatively sizeable. Indeed, 255 MSAs have positive spatial lag coefficients of which 226 are statistically significant. This points to the existence of important spill-over effects in the U.S. housing market even when the influence of national (common) factors are filtered out. It is easy to show spill-over effects in house price changes across MSAs without de-factoring, but such evidence suffers from the conjunctions of national and local influences, and can be misleading. The spatial display of the estimates in Figure 1 shows how the strength of local spill-over effects changes as we move from the sparsely populated areas in the middle of the US (Plains, Rocky Mountains and South West), and towards the two coastal areas (South East, Mid East and Far West) which have a much higher population density.

[22]The spatial lag coefficients of 44 MSAs hit the upper or lower bound of 0.994/-0.994 set in our optimization procedure. These are shown as a separate category in Figure 1. Of these 40 $\hat{\psi}_i$ are positive and 4 are negative. Widening the bounds to (-0.995,0.995) reduces the number of MSAs that fall outside the bounds to 30.

Figure 1: Spatial autoregressive parameter estimates ($\hat{\psi}_i$) for Metropolitan Statistical Areas in the Unites States



Notes: Each $\hat{\psi}_i$ is mapped to a Metropolitan Statistitial Area (MSA) in the U.S.. A total of 338 MSAs are included in model (67). MSAs coloured in blue correspond to positive spatial parameter estimates while MSAs coloured in red match to negative spatial parameter estimates. Darker shades of blue or red indicate more sizable $\hat{\psi}_i$ while lighter shades related to $\hat{\psi}_i$ closer to zero in absolute terms. Category 'Non-conv' includes MSAs whose $\hat{\psi}_i$ estimates hit the upper/lower bound in the optimisation procedure, while category 'No-Neigh' includes MSAs that have no neighbours when using $\boldsymbol{W}_{75}$.

Similar differences can also be seen in the estimates of the elasticities of house price changes to population and real per capita income changes, as shown in Figures 2(a) and 2(b). Focussing on contemporaneous effects, we observe that in 276 MSAs the population or income variables have a positive impact on house price changes, although the population effects tend to be more significant and sizeable. Of these, around two thirds tend to coincide with areas also reporting positive estimates for the spatial lag coefficients. Important examples of such MSAs include Seattle (Washington), San Francisco (California) or Boston (Massachusetts).[23] In contrast, the number of MSAs with negative estimates of the spatial lag coefficients is substantially lower, amounting to 39 of which only 11 are significantly different from zero. These are spread more evenly across the United States and correspond to economically less active areas in the U.S., such as Cheyenne (Wyoming), Coeur d'Alene (Idaho), Hot Springs (Arkansas), and Dothan (Alabama). It is also interesting that out of these 11 MSAs 7 have in fact experienced stagnant or declining population over our sample period, which could be the main reason behind the negative estimates of ψ_i obtained for these MSAs. One can extend the analysis further by computing marginal direct as well as spill-in and spill-out indirect effects of each explanatory variable on changes in real house prices, as discussed in LeSage and Chih (2016).[24]

[23]The estimates of the lagged population and real per capita income variables in (67) are generally small and less statistically significant as compared to their contemporaneous effects. These estimates are available upon request.

[24]All individual spatial and slope coefficient estimates with their standard errors from model (67) are available

The heterogeneity in spatial effects and the population and income elasticities across the U.S. can be seen even if we average the estimates across different regions in the U.S.. Table 3 reports the mean group estimates of the parameters grouped by six regions. We started with the standard eight regional classification, but combined New England and the Mid East, and South West and Rocky Mountains to ensure a reasonable number of MSAs (N_r in Table 3) per region. The MG estimates of the spatial lag coefficients are quite close for Great Lakes, South East and Far West (in the range 0.573 to 0.599), but differ markedly from the other three regions, namely New England & Mid East, Plains, and South West & Rocky Mountains, once again largely reflecting the different degrees of population density across the U.S.. We notice even larger differences in the MG estimates of population and real income variables across the regions, with much larger estimates for the effects of population changes on house prices as compared to income changes. For the U.S. as a whole, the MG estimate of the spatial effects amount to 0.509 (0.025) which points to the existence of non-negligible spatial dynamics in the U.S. even after conditioning on factors that generate strong spatial correlation between disaggregate house price fluctuations. This result is comparable though slightly lower than the homogeneous estimates of 0.643 (0.005) and 0.612 (0.003) obtained in Yang (2018) using the GMM and MLE approaches, respectively. Finally, the MG estimates of the contemporaneous effects of population and income variables for the U.S. as a whole are 0.446 (0.047) and 0.092 (0.009), respectively. The associated estimates for the lagged values of these variables are 0.155 (0.032) and 0.027 (0.007), all of which are statistically significant, and economically sizeable.

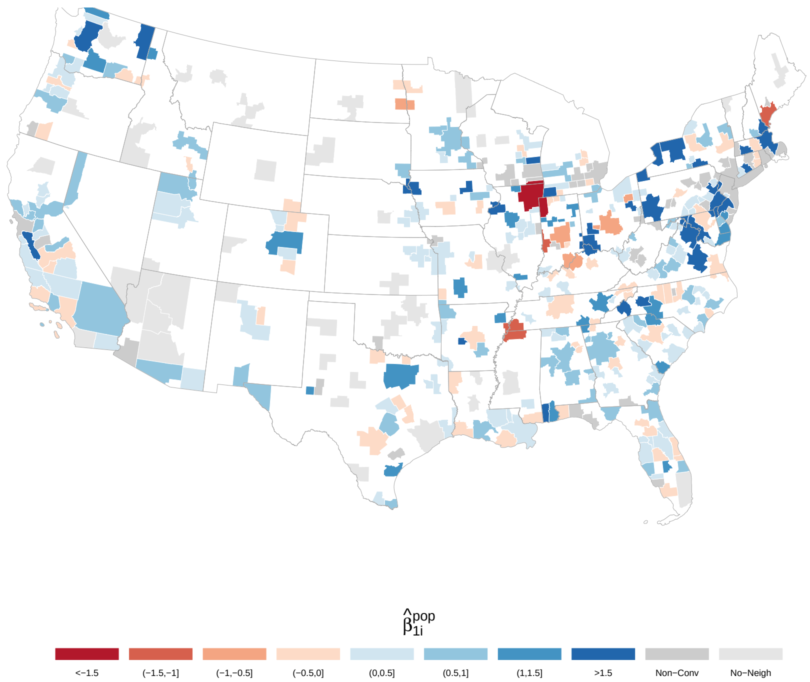
7 Conclusion

Standard spatial econometric models assume a single parameter to characterize the intensity or strength of spatial dependence across all units. In the case of pure cross section models or panel data models with a short time dimension, this assumption is inevitable. However, in a data rich environment where both the time (T) and cross section (N) dimensions are large, this can be relaxed. This paper investigates a spatial autoregressive panel data model with fully heterogeneous spatial parameters (HSAR) where the spatial dependence can arise directly through contemporaneous dependence of individual units on their neighbors, and indirectly through possible cross-sectional dependence in the regressors.

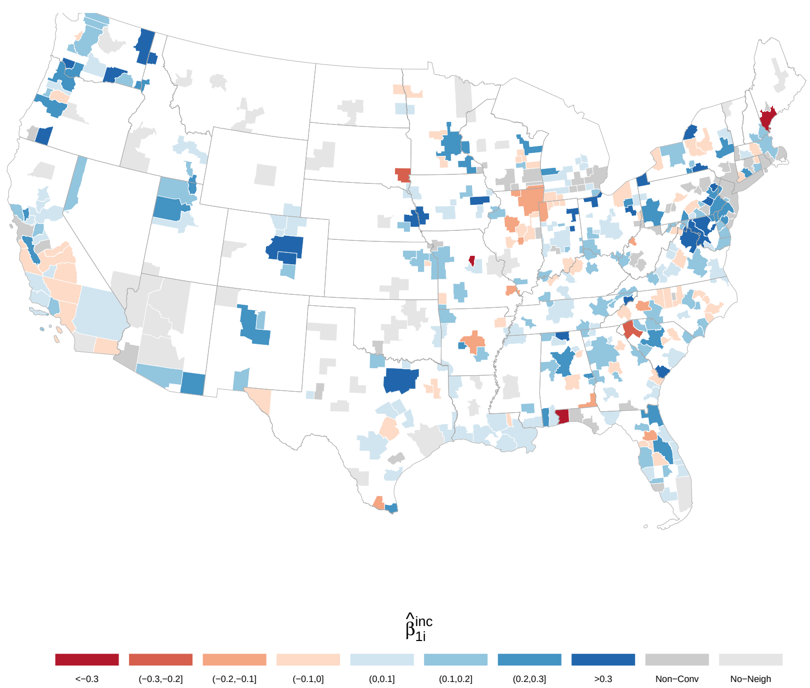
The asymptotic properties of the quasi maximum likelihood estimator are analyzed assuming a sparse spatial structure with each individual unit having at least one connection. Conditions under which the QML estimator of spatial parameters are consistent and asymptotically normal are derived. It is also shown that under certain conditions on spatial coefficients and the spatial weights, the asymptotic properties of the individual estimates are not affected by the size of cross section dimension N. An estimator of the cross section mean of the individual parameters (MG estimators) is also analyzed which can be used for comparisons with outcomes from standard homogeneous SAR models. It is shown that MG estimators are consistent and asymptotically normal as N and $T \to \infty$, jointly, so long as $\sqrt{N}/T \to 0$, and the spatial dependence is sufficiently weak. Monte Carlo simulation results provided are supportive of the theoretical findings. As an application of the HSAR model we investigate the potential heterogeneity in spatial spill-over effects in the U.S. housing market across the 338 MSAs included in our sample.

upon request.

Figure 2: Contemporaneous elasticities of house price changes to population growth ($\hat{\beta}_{1i}^{pop}$) and real income growth ($\hat{\beta}_{1i}^{inc}$) for Metropolitan Statistical Areas in the Unites States



(a) Population growth



(b) Real income growth

Notes: Each $\hat{\beta}_{1i}^{pop}$ and $\hat{\beta}_{1i}^{inc}$ is mapped to a Metropolitan Statistical Area (MSA) in the U.S.. A total of 338 MSAs are included in model (67). MSAs coloured in blue correspond to positive slope parameter estimates while MSAs coloured in red match to negative slope parameter estimates. Darker shades of blue or red indicate more sizable $\hat{\beta}_{1i}^{pop}$ and $\hat{\beta}_{1i}^{inc}$ while lighter shades related to $\hat{\beta}_{1i}^{pop}$ and $\hat{\beta}_{1i}^{inc}$ closer to zero in absolute terms. Category 'Non-conv' includes MSAs whose $\hat{\psi}_i$ estimates hit the upper/lower bound in the optimisation procedure, while category 'No-Neigh' includes MSAs that have no neighbours when using $\boldsymbol{W}_{75}$.

Table 3: Mean group estimates (MGE) of spatial coefficients and elasticities of house price changes to population and real income growth by six major U.S. regions, and the U.S. as a whole

r	Name	N_r	$\hat{\psi}_r$	$\hat{\beta}^{pop}_{1r}$	$\hat{\beta}^{pop}_{2r}$	$\hat{\beta}^{inc}_{1r}$	$\hat{\beta}^{inc}_{2r}$
1 &2	New England & Mid East	39	$0.398^{\ddagger}$	$0.959^{\ddagger}$	0.218	$0.165^{\ddagger}$	$-0.046^{\dagger}$
			(0.074)	(0.214)	(0.138)	(0.032)	(0.024)
3	Great Lakes	45	$0.599^{\ddagger}$	$0.367^{\ddagger}$	0.156	$0.071^{\ddagger}$	0.001
			(0.065)	(0.151)	(0.103)	(0.021)	(0.014)
4	Plains	26	$0.444^{\ddagger}$	$0.431^{\ddagger}$	0.191^{*}	0.064^{*}	0.011
			(0.083)	(0.128)	(0.100)	(0.035)	(0.016)
5	South East	107	$0.598^{\ddagger}$	$0.320^{\ddagger}$	$0.123^{\ddagger}$	$0.056^{\ddagger}$	$0.030^{\ddagger}$
			(0.039)	(0.054)	(0.043)	(0.011)	(0.012)
6 &7	South West & Rocky Mountains	37	$0.238^{\ddagger}$	$0.355^{\ddagger}$	$0.197^{\dagger}$	$0.136^{\ddagger}$	$0.086^{\ddagger}$
			(0.070)	(0.075)	(0.062)	(0.029)	(0.020)
8	Far West	40	$0.573^{\ddagger}$	$0.468^{\ddagger}$	0.119	$0.116^{\ddagger}$	$0.074^{\ddagger}$
			(0.051)	(0.095)	(0.076)	(0.025)	(0.023)
	U.S.	294	$0.509^{\ddagger}$	$0.446^{\ddagger}$	$0.155^{\ddagger}$	$0.092^{\ddagger}$	$0.027^{\ddagger}$
			(0.025)	(0.047)	(0.032)	(0.009)	(0.007)

Notes: * $p < 0.1$, † $p < 0.05$, ‡ $p < 0.01$. Non-parametric robust standard errors in parentheses (see below). For $r = 1, \ldots, 6$, $\hat{\psi}_{MG,r} = N_r^{-1} \sum_{i \in I_r} \hat{\psi}_i$, and s.e.$(\hat{\psi}_{MG,r}) = \sqrt{[N_r(N_r-1)]^{-1} \sum_{i \in I_r} (\hat{\psi}_i - \hat{\psi}_{MG,r})^2}$, where I_r is the set of units belonging to region r, $I_r = \{i\colon i \text{ is in region } r\}$, and N_r is the number of units per region, $N_r = \#(I_r)$. New England (8 MSAs) and Mid East (31 MSAs) as well as South West (22 MSAs) and Rocky Mountains (15 MSAs) have been merged in order to obtain a sufficiently large number of MSAs in the two broader regions. For the U.S. as a whole: $\hat{\psi}_{MG,US} = N^{-1} \sum_{i=1}^{N} \hat{\psi}_i$, and s.e.$(\hat{\psi}_{MG,US}) = \sqrt{[N(N-1)]^{-1} \sum_{i=1}^{N} (\hat{\psi}_i - \hat{\psi}_{MG,US})^2}$. The MGE of coefficient estimates of house price changes to population and real income changes ($\hat{\beta}^{pop}_{1i}$, $\hat{\beta}^{pop}_{2i}$, $\hat{\beta}^{inc}_{1i}$ and $\hat{\beta}^{inc}_{2i}$) are computed similarly. The computations of all MG estimates exclude the MSAs whose spatial lag coefficients hit the upper/lower bound in the optimisation procedure.

The methods developed in this paper can be extended to consider cases where the spatial parameter corresponding to each neighbor of unit i is estimated distinctly, as well as to the case of hierarchical panel data models where spatial parameters are assumed to be the same within regions (groups) but allowed to differ across regions or groups.

References

Agnello, L., V. Castro, and R. M. Sousa (2015). Booms, busts, and normal times in the housing market. *Journal of Business and Economic Statistics 33*(1), 25–45.

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.

Aquaro, M., N. Bailey, and M. H. Pesaran (2015). Quasi maximum likelihood estimation of spatial models with heterogeneous coefficients. USC-INET Research Paper No. 15-17.

Arbia, G. (2010). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Springer: Berlin Heidelberg New York.

Bailey, N., G. Kapetatios, and M. H. Pesaran (2016). Exponent of cross-sectional dependence: estimation and inference. *Journal of Applied Econometrics 31*(6), 929–960.

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.

Case, A. C. (1991). Spatial patterns in household demand. *Econometrica 59*, 953–965.

Chudik, A. and M. H. Pesaran (2019). Mean group estimator in presence of weakly cross-correlated estimators. *Economics Letters 175*, 101–105.

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.

Cun, W. and M. H. Pesaran (2018). Land use regulations, migration and rising house price dispersion in the U.S. Available at $https://papers.ssrn.com/sol3/papers.cfm?abstract_id = 3162399$.

Davidson, J. (2000). *Econometric Theory*. Blackwell, Oxford, UK.

Dou, B., M. L. Parrella, and Q. Yao (2016). Generalized Yule-Walker estimation for spatio-temporal models with unknown diagonal coefficients. *Journal of Econometrics 194*(2), 369–382.

Elhorst, J. P. (2014). *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Springer: Berlin Heidelberg.

Gallin, J. (2006). The long run relationship between housing prices and income: Evidence from local housing markets. *Real Estate Economics 34*(3), 417–438.

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.

Horn, R. A. and C. R. Johnson (1985). *Matrix analysis.* Cambridge University Press.

Hsiao, C. and M. H. Pesaran (2008). Random coefficient models. In L. Matyas and P. Sevestre (Eds.), *The Econometrics of Panel Data (Third Edition)*. Springer.

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 (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics 104*(2), 219–257.

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.

LeSage, J. P. and Y.-Y. Chih (2016). Interpreting heterogeneous coefficient spatial autoregressive panel models. *Economics Letters 142*, 1–5.

LeSage, J. P. and Y.-Y. Chih (2018). A bayesian spatial panel model with heterogeneous coefficients. *Regional Science and Urban Economics 72*, 58–73.

Lin, X. and L.-F. Lee (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics 157*, 34–52.

Lütkepohl, H. (1996). *Handbook of Matrices.* John Wiley and Sons.

Malpezzi, S. (1999). A simple error correction model of house prices. *Journal of Housing Economics 8*(1), 27–62.

Manski, C. F. (1993). Identification of endogenous social effects: the reflection problem. *The Review of Economic Studies 60*(3), 531–542.

Munro, A. (2018). Economic valuation of a rise in ambient radiation risks: hedonic pricing evidence from the accident in fukushima. *Journal of Regional Science 58*, 635–658.

Ord, J. K. and A. Getis (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis 27*, 286–306.

Pesaran, M. H. (2004). General diagnostic tests for cross section dependence in panels. CESifo working paper series no. 1229. Available at $http://ssrn.com/abstract = 572504$.

Pesaran, M. H. (2015). Testing weak cross-sectional dependence in large panels. *Econometrics Reviews 34*(6-10), 1089–1117.

Pesaran, M. H. and R. Smith (1995). Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics 68*, 79–113.

Pesaran, M. H., R. P. Smith, and K. S. Im (1996). Dynamic linear models for heterogenous panels. In L. Mátyás and P. S. (eds) (Eds.), *The Econometrics of Panel Data*. Dordrecht, Kluwer Academic Publishers, Ch.8, ISBN 0 7923 3787 5.

Swoboda, A., T. Nega, and M. Timm (2015). Hedonic analysis over time and space: the case of house price and traffic noise. *Journal of Regional Science 55*, 644–670.

Upton, G. J. G. and B. Fingleton (1985). *Spatial Data Analysis by Example, Volume 1*. Wiley: Chichester.

White, H. (1984). *Asymptotic Theory for Econometricians*. Academic Press, Inc.

Whittle, P. (1954). On stationary processes on the plane. *Biometrika 41*, 434–449.

Yang, C. F. (2018). Common factors and spatial dependence: An application to US house prices. Available at https://sites.google.com/site/cynthiafanyang/research.

Yu, J., R. de Jong, and L. F. Lee (2012). Estimation for spatial dynamic panel data with fixed effects: the case of spatial cointegration. *Journal of Econometrics 167*, 16–37.

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.

Zhou, J., Y. Tu, Y. Chen, and H. Wang (2017). Estimating spatial autocorrelation with sampled network data. *Journal of Business and Economic Statistics 35*(1), 130–138.

Online Supplement

for

Estimation and Inference for Spatial Models with Heterogeneous Coefficients: An Application to U.S. House Prices

by

Michele Aquaro
Joint Research Centre, European Commission

Natalia Bailey
Monash University

M. Hashem Pesaran
Department of Economics,
University of Southern California and Trinity College, Cambridge

February 23, 2019

Introduction

This online supplement is composed of Appendices A-E. Appendix A includes statements and proofs of lemmas used in the derivations of Sections 3.2, 3.3 and 4 of the paper. Appendix B provides proof of Proposition 2 in Section 3.3 of the paper, while Appendix C gives the first and second derivatives of the log likelihood function of the HSAR model with exogenous regressors. Appendix D describes the data sources used in Section 6, and Appendix E displays additional Monte Carlo results based on the designs set out in Section 5 of the paper.

Appendix A Technical lemmas

Lemma 1 *Consider the weight matrix* $\boldsymbol{W}$ *and suppose that Assumption 4 holds. Then matrix* $\mathbf{S}(\boldsymbol{\psi}) = \mathbf{I}_N - \boldsymbol{\Psi}\boldsymbol{W}$ *is non-singular with positive eigenvalues, namely* $\lambda_{\min}[\mathbf{S}(\boldsymbol{\psi})] > 0$.

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

$$\varrho(\boldsymbol{\Psi}\boldsymbol{W}) < 1. \tag{A.1}$$

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

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

and from (A.1) we have

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

Similarly, using a maximum row sum matrix norm we have

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

where we have used the result $\|\boldsymbol{\Psi}\|_1 = \|\boldsymbol{\Psi}\|_\infty = \sup_i |\psi_i|$. Therefore, matrix $\mathbf{S}(\boldsymbol{\psi}) = \boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}$ is invertible under condition (8) of Assumption 4. Also all eigenvalues of $\mathbf{S}(\boldsymbol{\psi})$ are necessarily positive, since $\lambda_{\min}[\mathbf{S}(\boldsymbol{\psi})] = 1 - \lambda_{\max}(\boldsymbol{\Psi}\boldsymbol{W}) \geq 1 - |\lambda_{\max}(\boldsymbol{\Psi}\boldsymbol{W})| = 1 - \varrho(\boldsymbol{\Psi}\boldsymbol{W}) > 0$. ∎

Lemma 2 *Let* $\mathbf{G}(\boldsymbol{\psi}) = \boldsymbol{W}(\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W})^{-1}$, *and suppose that Assumption 4 holds. Then*

$$\|\mathbf{G}(\boldsymbol{\psi})\|_1 < K \text{ and } \|\mathbf{G}(\boldsymbol{\psi})\|_\infty < K, \tag{A.3}$$

and

$$\left\|\mathbf{G}(\boldsymbol{\psi}) \odot \mathbf{G}'(\boldsymbol{\psi})\right\|_1 < K \text{ and } \left\|\mathbf{G}(\boldsymbol{\psi}) \odot \mathbf{G}'(\boldsymbol{\psi})\right\|_\infty < K, \tag{A.4}$$

for all values of $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_N)'$ *that satisfy condition (8).*

Proof. Under condition (8), we have

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

and

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

But $\|\boldsymbol{\Psi}\|_1^s = [\sup_i |\psi_i|]^s$, and under condition (8) we have $\sup\limits_i |\psi_i| \|\boldsymbol{W}\|_1 < 1$. Hence,

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

Similarly,

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

The boundedness of column and row matrix norms of $\mathbf{G}(\boldsymbol{\psi})$ now follow since, under Assumption 4, $\|\boldsymbol{W}\|_1$ and $\|\boldsymbol{W}\|_\infty$ are bounded, $1 - \sup\limits_i |\psi_i| \|\boldsymbol{W}\|_1 > 0$, and $1 - \sup_i |\psi_i| \|\boldsymbol{W}\|_\infty > 0$. Finally, (A.4) follows since

$$\left\|\mathbf{G}(\boldsymbol{\psi}) \odot \mathbf{G}'(\boldsymbol{\psi})\right\|_1 = \max_{1 \leq j \leq N} \left(\sum_{i=1}^N |g_{ij} g_{ji}| \right) \leq \max_{1 \leq j \leq N} \left(\sup_j |g_{ij}| \sum_{i=1}^N |g_{ji}| \right) < K,$$

and

$$\left\|\mathbf{G}(\boldsymbol{\psi}) \odot \mathbf{G}'(\boldsymbol{\psi})\right\|_\infty = \max_{1 \leq i \leq N} \left(\sum_{j=1}^N |g_{ij} g_{ji}| \right) \leq \max_{1 \leq i \leq N} \left(\sup_i |g_{ij}| \sum_{j=1}^N |g_{ji}| \right) < K,$$

from (A.3). ■

Lemma 3 *Consider the average log-likelihood function of (4):*

$$\bar{\ell}_T(\boldsymbol{\theta}) = T^{-1} \ell_T(\boldsymbol{\theta}) = -\frac{NT}{2} \ln(2\pi) - \frac{T}{2} \sum_{i=1}^N \ln \sigma_i^2 + \frac{T}{2} \ln |\mathbf{V}(\boldsymbol{\psi})| \tag{A.5}$$
$$- \frac{1}{2} \sum_{t=1}^T [\mathbf{S}(\boldsymbol{\psi}) \boldsymbol{y}_{\circ t} - \boldsymbol{B} \boldsymbol{x}_{\circ t}]' \boldsymbol{\Sigma}^{-1} [\mathbf{S}(\boldsymbol{\psi}) \boldsymbol{y}_{\circ t} - \boldsymbol{B} \boldsymbol{x}_{\circ t}],$$

where $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})'$, $\boldsymbol{B}$ is given in (5), $\bar{\ell}_T(\boldsymbol{\theta}) = T^{-1} \ell_T(\boldsymbol{\theta})$ and $\ell_T(\boldsymbol{\theta})$ is defined by (7). Also, $\mathbf{V}(\boldsymbol{\psi}) = \mathbf{S}'(\boldsymbol{\psi}) \mathbf{S}(\boldsymbol{\psi})$. Denote the true parameter vector of $\boldsymbol{\theta}$ by $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}_0', \boldsymbol{\beta}_0', \boldsymbol{\sigma}_0^{2\prime})'$ which lies in the interior of $\boldsymbol{\Theta} = \boldsymbol{\Theta}_\psi \times \boldsymbol{\Theta}_\beta \times \boldsymbol{\Theta}_\sigma \subset \mathbb{R}^N \times \mathbb{R}^{Nk} \times \mathbb{R}^N$. Then, under Assumptions 1 and 2 we have

$$\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \overset{a.s.}{\to} E_0 \left[\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta}) \right], \tag{A.6}$$

where E_0 represents expectations taken under $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof. Let $Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})$, and evaluating (A.5) at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, note that

$$
\begin{aligned}
Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = & -\frac{1}{2}\sum_{i=1}^{N} \ln\left(\sigma_{i0}^2/\sigma_i^2\right) + \frac{1}{2}\left[\ln\left(\frac{|\mathbf{V}(\boldsymbol{\psi}_0)|}{|\mathbf{V}(\boldsymbol{\psi})|}\right)\right] \\
& -\frac{1}{2}\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{S}(\boldsymbol{\psi}_0)\boldsymbol{y}_{\circ t} - \boldsymbol{B}_0\boldsymbol{x}_{\circ t}\right]' \boldsymbol{\Sigma}_0^{-1}\left[\mathbf{S}(\boldsymbol{\psi}_0)\boldsymbol{y}_{\circ t} - \boldsymbol{B}_0\boldsymbol{x}_{\circ t}\right]\right\} \\
& +\frac{1}{2}\left\{\frac{1}{T}\sum_{t=1}^{T}\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right]' \boldsymbol{\Sigma}^{-1}\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right]\right\},
\end{aligned}
\tag{A.7}
$$

where, by Remark 6, $\lambda_{\min}[\mathbf{V}(\boldsymbol{\psi})] > 0$ and $\lambda_{\max}[\mathbf{V}(\boldsymbol{\psi})] < K$, so that $|\mathbf{V}(\boldsymbol{\psi})| = |\mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})| = |\mathbf{S}(\boldsymbol{\psi})|^2$. Also, by first taking conditional expectations with respect to $\mathcal{F}_t$, and then taking expectations with respect to $\boldsymbol{x}_{\circ t}$, we have

$$
\begin{aligned}
\sum_{t=1}^{T} E_0\left\{\operatorname{tr}\left[\boldsymbol{y}_{\circ t}'\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t}\right]\right\} &= T\operatorname{tr}\left[\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\boldsymbol{\Sigma}_{y0}\right], \\
\sum_{t=1}^{T} E_0\left\{\operatorname{tr}\left[\boldsymbol{y}_{\circ t}'\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{x}_{\circ t}\right]\right\} &= T\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\boldsymbol{B}_0'\boldsymbol{S}_0'^{-1}\mathbf{S}'(\boldsymbol{\psi})\right], \\
\sum_{t=1}^{T} E_0\left\{\operatorname{tr}\left[\boldsymbol{x}_{\circ t}'\boldsymbol{B}'\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{x}_{\circ t}\right]\right\} &= T\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\boldsymbol{B}'\right],
\end{aligned}
$$

and hence

$$
\frac{1}{T}\sum_{t=1}^{T} E_0\left\{\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right]' \boldsymbol{\Sigma}^{-1}\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right]\right\} = \left\{\begin{array}{c} \operatorname{tr}\left[\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\boldsymbol{\Sigma}_{y0}\right] \\ -2\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1}\boldsymbol{B}\boldsymbol{\Sigma}_{\boldsymbol{xx}}\boldsymbol{B}_0'\mathbf{S}_0'^{-1}\mathbf{S}'(\boldsymbol{\psi})\right] \\ +\operatorname{tr}\left[\boldsymbol{\Sigma}^{-1/2}\boldsymbol{B}\boldsymbol{\Sigma}_{xx}\boldsymbol{B}'\boldsymbol{\Sigma}^{-1/2}\right] \end{array}\right\}.
$$

Using the above results in (A.7) we now obtain

$$
\begin{aligned}
E_0\left[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})\right] &= E_0\left[\bar{\ell}_T(\boldsymbol{\theta}_0) - \bar{\ell}_T(\boldsymbol{\theta})\right] \\
&= -\frac{1}{2}\sum_{i=1}^{N}\ln\left(\sigma_{i0}^2/\sigma_i^2\right) - \frac{N}{2} + \frac{1}{2}\left[\ln\left(\frac{|\mathbf{V}(\boldsymbol{\psi}_0)|}{|\mathbf{V}(\boldsymbol{\psi})|}\right)\right] \\
&+ \frac{1}{2}\operatorname{tr}\left[\mathbf{P}(\boldsymbol{\theta})\mathbf{P}^{-1}(\boldsymbol{\theta}_0)\right] \\
&+ \frac{1}{2}\operatorname{tr}\left\{\boldsymbol{\Sigma}^{-1/2}\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]\boldsymbol{\Sigma}_{\boldsymbol{xx}}\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]'\boldsymbol{\Sigma}^{-1/2}\right\},
\end{aligned}
$$

where

$$
\frac{|\mathbf{V}(\boldsymbol{\psi}_0)|}{|\mathbf{V}(\boldsymbol{\psi})|} = \frac{|\mathbf{S}(\boldsymbol{\psi}_0)|^2}{|\mathbf{S}(\boldsymbol{\psi})|^2} = \left|\mathbf{S}(\boldsymbol{\psi}_0)\mathbf{S}^{-1}(\boldsymbol{\psi})\right|^2 = \left|\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\right|^{-2},
$$

$$
\begin{aligned}
\operatorname{tr}\left[\mathbf{P}(\boldsymbol{\theta})\mathbf{P}^{-1}(\boldsymbol{\theta}_0)\right] &= \operatorname{tr}\left[\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\Sigma}_0\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\right] \\
&= \operatorname{tr}\left[\boldsymbol{\Sigma}^{-1/2}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\Sigma}_0\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1/2}\right],
\end{aligned}
$$

and $\mathbf{P}(\boldsymbol{\theta}) = \mathbf{S}'(\boldsymbol{\psi})\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})$. To establish (A.6) we show that

$$Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) - E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] \overset{a.s.}{\to} 0. \tag{A.8}$$

To this end we note that under (4),

$$\frac{1}{T}\sum_{t=1}^{T}[\mathbf{S}(\boldsymbol{\psi}_0)\boldsymbol{y}_{\circ t} - \boldsymbol{B}_0\boldsymbol{x}_{\circ t}]' \boldsymbol{\Sigma}_0^{-1}[\mathbf{S}(\boldsymbol{\psi}_0)\boldsymbol{y}_{\circ t} - \boldsymbol{B}_0\boldsymbol{x}_{\circ t}] = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\zeta}_{\circ t}'\boldsymbol{\zeta}_{\circ t},$$

where $\boldsymbol{\zeta}_{\circ t} = (\zeta_{1t}, \zeta_{2t}, \ldots, \zeta_{Nt}) \sim IID(\mathbf{0}, \mathbf{I}_N)$, $\zeta_{it} = \varepsilon_{it}/\sigma_{i0}$, for $i = 1, 2, \ldots, N$. Also,

$$\begin{aligned}
&\frac{1}{2}\left\{\frac{1}{T}\sum_{t=1}^{T}[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}]' \boldsymbol{\Sigma}^{-1}[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}]\right\} \\
&= \frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{\circ t}'\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\mathbf{P}(\boldsymbol{\theta})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_{\circ t}\right] \\
&+ \frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}\left\{\boldsymbol{\Sigma}^{-1/2}\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]\boldsymbol{x}_{\circ t}\right\}'\left\{\boldsymbol{\Sigma}^{-1/2}\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]\boldsymbol{x}_{\circ t}\right\}\right] \\
&+ \left[\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{\circ t}'\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]'\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{\varepsilon}_{\circ t}\right].
\end{aligned}$$

Using the above results in (A.7) and after some simplifications we have

$$\begin{aligned}
Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta}) - E_0[Q_T(\boldsymbol{\theta}_0, \boldsymbol{\theta})] = &-\frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}z_{1t,N}(\boldsymbol{\theta}_0)\right] + \frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}z_{2t,N}(\boldsymbol{\theta})\right] \\
&+ \frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}z_{3t,N}(\boldsymbol{\theta}_0, \boldsymbol{\theta})\right] + \frac{1}{2}\left[\frac{1}{T}\sum_{t=1}^{T}z_{4t,N}(\boldsymbol{\theta})\right],
\end{aligned} \tag{A.9}$$

where

$$z_{1t,N}(\boldsymbol{\theta}_0) = \boldsymbol{\zeta}_{\circ t}'\boldsymbol{\zeta}_{\circ t} - N = \sum_{i=1}^{N}\left(\zeta_{it}^2 - 1\right), \tag{A.10}$$

$$\begin{aligned}
z_{2t,N}(\boldsymbol{\theta}) &= \boldsymbol{\varepsilon}_{\circ t}'\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\mathbf{P}(\boldsymbol{\theta})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_{\circ t} - E_0\left[\boldsymbol{\varepsilon}_{\circ t}'\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\mathbf{P}(\boldsymbol{\theta})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\varepsilon}_{\circ t}\right] \\
&= \boldsymbol{\zeta}_{\circ t}'\mathcal{A}(\boldsymbol{\theta}_0, \boldsymbol{\theta})\boldsymbol{\zeta}_{\circ t} - \mathrm{tr}[\mathcal{A}(\boldsymbol{\theta}_0, \boldsymbol{\theta})],
\end{aligned} \tag{A.11}$$

in which $\mathcal{A}(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \boldsymbol{\Sigma}_0^{1/2}\mathbf{S}'^{-1}(\boldsymbol{\psi}_0)\mathbf{P}(\boldsymbol{\theta})\mathbf{S}^{-1}(\boldsymbol{\psi}_0)\boldsymbol{\Sigma}_0^{1/2}$,

$$z_{3t,N}(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \boldsymbol{x}_{\circ t}'\mathcal{B}(\boldsymbol{\theta}_0, \boldsymbol{\theta})\boldsymbol{x}_{\circ t} - \mathrm{tr}[\mathcal{B}(\boldsymbol{\theta}_0, \boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{xx}}], \tag{A.12}$$

in which $\mathcal{B}(\boldsymbol{\theta}_0, \boldsymbol{\theta}) = \left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]'\boldsymbol{\Sigma}^{-1}\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]$, and

$$z_{4t,N}(\boldsymbol{\theta}) = \boldsymbol{x}_{\circ t}'\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]'\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{\varepsilon}_{\circ t}. \tag{A.13}$$

We establish that each $z_{jt,N}$, $j = 1,2,3,4$ is a martingale difference process with finite second order moments. Starting with $z_{1t,N}(\boldsymbol{\theta}_0)$, we have that under Assumption 1, $\sup_{it} E\left|\zeta_{it}\right|^{4+\epsilon} < K$, for some $\epsilon > 0$. Then the elements in (A.10) are L_2 bounded, in the sense that $\sup_{it} E\left|\zeta_{it}^2 - 1\right|^2 < K$, and

$$\frac{1}{T}\sum_{t=1}^{T}\left(\zeta_{it}^2 - 1\right) \overset{a.s.}{\to} 0, \tag{A.14}$$

for a given N and as $T \to \infty$. Consider now (A.11), and note that $z_{2t,N}(\boldsymbol{\theta})$ is serially independent (over t), and has mean zero. Under our assumptions (see Assumption 1, Remarks 4 and 6)

$$\left\|\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\right\|_1 < K, \text{ and } \left\|\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\right\|_\infty < K. \tag{A.15}$$

Further, note that $z_{2t,N}$ is a de-meaned quadratic form in ζ_{it} and Theorem 1 of Kelejian and Prucha (2001) applies to $z_{2t,N}$. Denote the (i,j) element of $\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})$, by a_{ij}, and note that $a_{ij} = a_{ji}$. Then using (3.2) in Kelejian and Prucha (2001), we have (recall that $E\left(\zeta_{it}^2\right) = 1$)

$$\begin{aligned} Var\left[z_{2t,N}(\boldsymbol{\theta})\right] &= Var\left[\boldsymbol{\zeta}_{\circ t}'\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\boldsymbol{\zeta}_{\circ t}\right] \\ &= 4\sum_{i=1}^{N}\sum_{j=1}^{i-1} a_{ij}^2 + \sum_{i=1}^{N} a_{ii}^2\left[E\left(\zeta_{it}^4\right) - 1\right]. \end{aligned}$$

But

$$\sum_{i=1}^{N}\sum_{j=1}^{i-1} a_{ij}^2 \leq \sum_{i=1}^{N}\sum_{j=1}^{N} a_{ij}^2 = \operatorname{tr}\left[\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})'\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\right] = \left[\left\|\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\right\|_F\right]^2,$$

and using (A.15),

$$\left\|\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\right\|_F \leq \sqrt{\left\|\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\right\|_1\left\|\mathcal{A}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\right\|_\infty} < K.$$

Hence, $\sum_{i=1}^{N} a_{ii}^2 < K$ and $\sum_{i=1}^{N}\sum_{j=1}^{i-1} a_{ij}^2 < K$. Furthermore,

$$\left|\sum_{i=1}^{N} a_{ii}^2\left[E\left(\zeta_{it}^4\right) - 1\right]\right| < \sup_{it}\left|E\left(\zeta_{it}^4\right) - 1\right|\sum_{i=1}^{N} a_{ii}^2,$$

and under Assumption 1, $\sup_{it} E\left(\zeta_{it}^4\right) < K$, and hence $Var\left[z_{2t,N}(\boldsymbol{\theta})\right] < K$. Further, since $z_{2t,N}(\boldsymbol{\theta})$ are independently distributed over t, then we have (see, for example, White (1984))

$$\frac{1}{T}\sum_{t=1}^{T} z_{2t,N}(\boldsymbol{\theta}) \overset{a.s.}{\to} 0. \tag{A.16}$$

Next, using (A.12),

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} z_{3t,N}(\boldsymbol{\theta}) &= \operatorname{tr}\left[\mathcal{B}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}'\right)\right] - \operatorname{tr}\left[\mathcal{B}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\boldsymbol{\Sigma}_{\boldsymbol{xx}}\right] \\ &= \operatorname{tr}\left\{\mathcal{B}(\boldsymbol{\theta}_0,\boldsymbol{\theta})\left[T^{-1}\sum_{t=1}^{T}\left(\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}' - \boldsymbol{\Sigma}_{\boldsymbol{xx}}\right)\right]\right\}, \end{aligned}$$

But by Assumption 2(b) and (c) we have that $E\left(\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}' - \boldsymbol{\Sigma}_{\boldsymbol{xx}}|\mathcal{F}_t\right) = \mathbf{0}$, and $T^{-1}\sum_{t=1}^{T}\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}' \overset{a.s.}{\to} \boldsymbol{\Sigma}_{\boldsymbol{xx}}$, as $T\to\infty$, which establishes that $T^{-1}\sum_{t=1}^{T} z_{3t,N}\left(\boldsymbol{\theta}\right) \overset{a.s.}{\to} 0$, as required. Finally, using (A.13),

$$\frac{1}{T}\sum_{t=1}^{T} z_{4t,N}\left(\boldsymbol{\theta}\right) = \operatorname{tr}\left\{\left[\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}\boldsymbol{B}_0 - \boldsymbol{B}\right]'\boldsymbol{\Sigma}^{-1}\mathbf{S}(\boldsymbol{\psi})\mathbf{S}_0^{-1}T^{-1}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{\circ t}\boldsymbol{x}_{\circ t}'\right\}.$$

But by Assumption 2(a), we have that $E\left(\boldsymbol{\varepsilon}_{\circ t}\boldsymbol{x}_{\circ t}'|\mathcal{F}_t\right) = 0$ and $E\left(|\boldsymbol{\varepsilon}_{\circ t}\boldsymbol{x}_{\circ t}'|^p \mid \mathcal{F}_t\right) \le E\left(|\boldsymbol{\varepsilon}_{\circ t}|^p |\boldsymbol{x}_{\circ t}'|^p \mid \mathcal{F}_t\right)$, for $p=2$, which is bounded. Hence, $T^{-1}\sum_{t=1}^{T}\boldsymbol{\varepsilon}_{\circ t}\boldsymbol{x}_{\circ t}' \overset{a.s.}{\to} \mathbf{0}$, as $T\to\infty$ and $\frac{1}{T}\sum_{t=1}^{T} z_{4t,N}\left(\boldsymbol{\theta}\right) \overset{a.s.}{\to} 0$. (see, for example, White (1984)). Finally, (A.8) and (A.6) follow similarly. ■

Lemma 4 *Let*

$$\eta_{it} = \sigma_{i0}^{-1}\mathbf{e}_{i,N}'\boldsymbol{G}_0\mathbf{B}_0\boldsymbol{x}_{\circ t}\zeta_{it} + \sigma_{i0}^{-1}\mathbf{e}_{i,N}'\boldsymbol{G}_0\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\zeta}_{\circ t}\zeta_{it} - g_{0,ii}, \tag{A.17}$$

where $\boldsymbol{G}_0 = \boldsymbol{W}\left(\mathbf{I}_N - \boldsymbol{\Psi}_0\boldsymbol{W}\right)^{-1} = (g_{0,ij})$ *and* $\mathbf{e}_{i,N}$ *is an* N *dimensional vector with its* i^{th} *element unity and zeros elsewhere, and* $\boldsymbol{\zeta}_{\circ t} = (\zeta_{1t},\zeta_{2t},\ldots,\zeta_{Nt})' = (\varepsilon_{1t}/\sigma_{10},\varepsilon_{2t}/\sigma_{20},\ldots,\varepsilon_{Nt}/\sigma_{N0})'$. *Then under Assumptions 1 and 2,* η_{it} *is a martingale difference process with respect to the filtration,* $\mathcal{F}_t = (\boldsymbol{x}_{\circ t},\boldsymbol{x}_{\circ t-1},\boldsymbol{x}_{\circ t-2},\ldots)$, *namely* $E\left(\eta_{it}|\mathcal{F}_t\right) = 0$, *and*

$$\sup_{i,t} E\left|\eta_{it}\right|^p < K, \text{ for } 1\le p\le 2+c, \text{ and some } c>0. \tag{A.18}$$

Proof. We first recall that $E\left(\boldsymbol{\zeta}_t|\mathcal{F}_t\right) = \mathbf{0}$, and hence $E\left(\boldsymbol{\zeta}_t\right) = \mathbf{0}$. Also $E\left(\boldsymbol{\zeta}_{\circ t}\zeta_{it}\right) = \mathbf{e}_{i,N}$ and $Var(\boldsymbol{\zeta}_{\circ t}) = \mathbf{I}_N$. Now under Assumption 1 it follows that

$$\begin{aligned} E\left(\eta_{it}|\mathcal{F}_t\right) &= E\left(\sigma_{i0}^{-1}\mathbf{e}_{i,N}'\boldsymbol{G}_0\mathbf{B}_0\boldsymbol{x}_{\circ t}\zeta_{it}|\mathcal{F}_t\right) + E\left(\sigma_{i0}^{-1}\mathbf{e}_{i,N}'\boldsymbol{G}_0\boldsymbol{\Sigma}_0^{1/2}\boldsymbol{\zeta}_{\circ t}\zeta_{it}|\mathcal{F}_t\right) - g_{0,ii} \\ &= 0 + g_{0,ii} - g_{0,ii} = 0, \end{aligned}$$

and establishes that η_{it} is a martingale difference process with respect to $\mathcal{F}_t$, as required. To establish (A.18), since $\zeta_{it} = \varepsilon_{it}/\sigma_{i0}$ then by Minkowski's inequality for $p\ge 1$ we have:

$$\left\|\eta_{it}\right\|_p \le \sigma_{i0}^{-2}\left\|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\varepsilon_{it}\right\|_p + \sigma_{i0}^{-1}\left\|\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it}\right\|_p + \left|g_{0,ii}\right|, \tag{A.19}$$

where $\boldsymbol{\varphi}_i' = \mathbf{e}_{i,N}'\boldsymbol{G}_0\mathbf{B}_0$, $\boldsymbol{\vartheta}_i' = \mathbf{e}_{i,N}'\boldsymbol{G}_0\boldsymbol{\Sigma}_0^{1/2} = \left(g_{i1}\sigma_{10}, g_{i2}\sigma_{20},\ldots,g_{iN}\sigma_{N0}\right)$, and $|g_{0,ii}|<K$. Consider now the first term of (A.19), and note that since conditional on $\mathcal{F}_t$, $\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}$ is given, and noting that by Assumption 1 $E\left(|\varepsilon_{it}|^p|\mathcal{F}_t\right) = \varpi_{ip} < K$, then

$$\left\|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\varepsilon_{it}\right\|_p^p = E\left[E\left(\left|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\varepsilon_{it}\right|^p \mid \mathcal{F}_t\right)\right] \le E\left[\left|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\right|^p\left|E\left(|\varepsilon_{it}|^p \mid \mathcal{F}_t\right)\right] = E\left(\left|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\right|^p\right)\varpi_{ip},$$

and hence $\left\|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\varepsilon_{it}\right\|_p \le \varpi_{ip}^{1/p}\left\|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\right\|_p$. Also

$$\begin{aligned} \left\|\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\right\|_p &= \left\|\sum_{j=1}^{N} g_{ij,0}\boldsymbol{\beta}_{j0}'\boldsymbol{x}_{jt}\right\|_p \le \sum_{j=1}^{N}\left|g_{ij,0}\right|\left\|\boldsymbol{\beta}_{j0}'\boldsymbol{x}_{jt}\right\|_p \\ &\le \left(\sup_{j,t} E\left\|\boldsymbol{\beta}_{j0}'\boldsymbol{x}_{jt}\right\|_p\right)\sum_{j=1}^{N}\left|g_{ij,0}\right|. \end{aligned}$$

The first term on the right hand side is bounded by Assumption 2(a), for $p \leq 2+c$, and $\sup_i \sum_{j=1}^{N} |g_{ij,0}|$ is bounded by Lemma 2. Hence, $\sup_{i,t} \|\boldsymbol{\varphi}_i' \boldsymbol{x}_{\circ t}\|_p < K$, and overall we have $\sup_{i,t} \|\boldsymbol{\varphi}_i' \boldsymbol{x}_{\circ t}\varepsilon_{it}\|_p < K$. Consider now the second term of (A.19) and note that

$$\left\|\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it}\right\|_p \leq \sum_{j=1}^{N} \|g_{ij,0}\sigma_{j0}\zeta_{jt}\zeta_{it}\|_p \leq \frac{1}{\sigma_{i0}} \sum_{j=1}^{N} |g_{ij,0}| \left|\varepsilon_{jt}\varepsilon_{it}\right|_p = \frac{1}{\sigma_{i0}} \sum_{j=1}^{N} |g_{ij,0}| \left[E\left|\varepsilon_{jt}\varepsilon_{it}\right|^p\right]^{1/p}.$$

But $\sup_i (1/\sigma_{i0}) < K$ by Assumption 1, and using Cauchy–Schwarz inequality we obtain[25]

$$\begin{aligned}
\left\|\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it}\right\|_p &\leq K \sum_{j=1}^{N} |g_{ij,0}| \left[E\left(\varepsilon_{jt}^{2p}\right)\right]^{1/2p} \left[E\left(\varepsilon_{it}^{2p}\right)\right]^{1/2p} \\
&\leq K \left\{\sup_{i,t} \left[E\left(\varepsilon_{it}^{2p}\right)\right]^{1/p}\right\} \sum_{j=1}^{N} |g_{ij,0}|.
\end{aligned}$$

Again $\sup_i \sum_{j=1}^{N} |g_{ij,0}| < K$ under Lemma 2 , and $E\left(\varepsilon_{it}^{2p}\right) < K$ for $2p = 4+\epsilon$ under Assumption 1, and hence $\left\|\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it}\right\|_p < K$. Using this result together with $\sup_{i,t} \|\boldsymbol{\varphi}_i' \boldsymbol{x}_{\circ t}\varepsilon_{it}\|_p < K$ (established above) in (A.19) now yields (A.18) by setting $c = 2\epsilon$. ■

Lemma 5 *Let*

$$\begin{aligned}
\ell_t(\boldsymbol{\theta}) = &-\frac{N}{2}\ln(2\pi) - \frac{1}{2}\sum_{i=1}^{N} \ln\sigma_i^2 + \frac{1}{2}\ln|\mathbf{V}(\boldsymbol{\psi})| \\
&- \frac{1}{2}\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right]' \boldsymbol{\Sigma}^{-1} \left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right],
\end{aligned} \tag{A.20}$$

where $\mathbf{V}(\boldsymbol{\psi}) = \mathbf{S}'(\boldsymbol{\psi})\mathbf{S}(\boldsymbol{\psi})$, and note that the log-likelihood function is given by

$$\begin{aligned}
\ell_T(\boldsymbol{\theta}) = \sum_{t=1}^{T}\ell_t(\boldsymbol{\theta}) = &-\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\sum_{i=1}^{N} \ln\sigma_i^2 + \frac{T}{2}\ln|\mathbf{V}(\boldsymbol{\psi})| \\
&- \frac{1}{2}\sum_{t=1}^{T}\left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right]' \boldsymbol{\Sigma}^{-1} \left[\mathbf{S}(\boldsymbol{\psi})\boldsymbol{y}_{\circ t} - \boldsymbol{B}\boldsymbol{x}_{\circ t}\right],
\end{aligned}$$

(see also (7)) which can be written equivalently as

$$\begin{aligned}
\ell_T(\boldsymbol{\theta}) = &-\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\sum_{i=1}^{N} \ln\sigma_i^2 + \frac{T}{2}\ln|\mathbf{V}(\boldsymbol{\psi})| \\
&- \frac{1}{2}\left\{\sum_{i=1}^{N} \frac{\left(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i\right)'\left(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i\right)}{\sigma_i^2}\right\},
\end{aligned} \tag{A.21}$$

where $\boldsymbol{y}_{i\circ} = (y_{i1}, y_{i2}, \ldots, y_{iT})'$ and $\boldsymbol{y}_{i\circ}^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{iT}^*)'$ are $T\times 1$ vectors, and $\mathbf{X}_{i\circ} = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})'$ is the $T\times k$ matrix of observations on regressors specific to the i^{th} cross

[25] Note that since by Assumption 1 $E\left(|\varepsilon_{it}|^p |\mathcal{F}_t\right) = \varpi_{ip} < K$, then for a given i we also have $E\left(|\varepsilon_{it}|^p\right) = \varpi_{ip}$, unconditionally.

section unit. Suppose that Assumptions 1, 2, 4 and 5, and conditions (21) and (28) hold. Denote the score function by $\boldsymbol{s}_T(\boldsymbol{\theta}) = \partial\ell_T(\boldsymbol{\theta})/\partial\boldsymbol{\theta} = \sum_{t=1}^T \partial\ell_t(\boldsymbol{\theta})/\partial\boldsymbol{\theta}$. *Then*

$$T^{-1}\boldsymbol{s}_T(\boldsymbol{\theta}_0) \overset{a.s.}{\to} \mathbf{0}, \tag{A.22}$$

and

$$T^{-1/2}\boldsymbol{s}_T(\boldsymbol{\theta}_0) \to_d N\left[\mathbf{0}, \mathbf{J}(\boldsymbol{\theta}_0, \gamma)\right], \tag{A.23}$$

where

$$\mathbf{J}(\boldsymbol{\theta}_0, \gamma) = (J_{0,ij}) = \lim_{T\to\infty}\sum_{t=1}^T E_0\left[\frac{1}{T}\left(\frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)\left(\frac{\partial\ell_t(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right)'\right] \tag{A.24}$$

and

$$\gamma = \left[\lim_{T\to\infty} T^{-1}\sum_{t=1}^T E\left(\zeta_{it}^4\right) - 1\right] = \lim_{T\to\infty} T^{-1}\sum_{t=1}^T Var(\zeta_{it}^2),$$

with $\zeta_{it} \sim IID(0,1)$, $\zeta_{it} = \varepsilon_{it}/\sigma_{i0}$, *for* $i = 1, 2, \ldots, N$. *A consistent estimator of* $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ *is given by*

$$\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}, \hat{\gamma}) = \frac{1}{T}\left\{\left[\sum_{t=1}^T \frac{\partial\ell_t\left(\hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\theta}}\right]\left[\sum_{t=1}^T \frac{\partial\ell_t\left(\hat{\boldsymbol{\theta}}\right)}{\partial\boldsymbol{\theta}}\right]'\right\},$$

where $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \bar{\ell}_T(\boldsymbol{\theta})$ *and*

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

with $\hat{\varepsilon}_{it} = y_{it} - \hat{\psi}_i\sum_{j=1}^N w_{ij}y_{jt} - \hat{\boldsymbol{\beta}}_i'\boldsymbol{x}_{it}$. $\hat{\sigma}_i$, $\hat{\boldsymbol{\beta}}_i$ *and* $\hat{\psi}_i$ *are the QML estimators of* σ_{i0}, $\boldsymbol{\beta}_{i0}$ *and* ψ_{i0}, *respectively.*

Proof. For a given N, the $N(k+2)\times 1$ score vector $\boldsymbol{s}_T(\boldsymbol{\theta}_0) = \left(\frac{\partial\ell_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\psi}'}, \frac{\partial\ell_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}'}, \frac{\partial\ell_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\sigma}^{2\prime}}\right)'$, where

$$\begin{pmatrix} \frac{\partial\ell_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\psi}} \\ \frac{\partial\ell_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\beta}} \\ \frac{\partial\ell_T(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\sigma}^2} \end{pmatrix}_{N(k+2)\times 1} = \begin{pmatrix} \left[-T\,\mathrm{Diag}(\boldsymbol{G}_0) + \mathrm{Diag}\left(\frac{\boldsymbol{y}_{i\circ}^{*\prime}\boldsymbol{\varepsilon}_{i\circ}}{\sigma_{i0}^2}, i = 1, 2, \ldots, N\right)\right]\boldsymbol{\tau}_N \\ \mathrm{Diag}\left(\frac{\mathbf{X}_{i\circ}'\boldsymbol{\varepsilon}_{i\circ}}{\sigma_{i0}^2}, i = 1, 2, \ldots, N\right)\boldsymbol{\tau}_{Nk} \\ \mathrm{Diag}\left[-\frac{T}{2\sigma_{i0}^2} + \frac{1}{2\sigma_{i0}^4}\left(\boldsymbol{\varepsilon}_{i\circ}'\boldsymbol{\varepsilon}_{i\circ}\right), i = 1, 2, \ldots, N\right]\boldsymbol{\tau}_N \end{pmatrix}, \tag{A.25}$$

$\boldsymbol{y}_{i\circ}^* = (y_{i1}^*, y_{i2}^*, \ldots, y_{iT}^*)$, $\boldsymbol{\varepsilon}_{i\circ} = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{iT})$, $\boldsymbol{\varepsilon}_{i\circ} = \boldsymbol{y}_{i\circ} - \psi_{i0}\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_{i0}$, and $\boldsymbol{\tau}_\kappa$ is a $\kappa\times 1$ vector of ones. Consider first the i^{th} component of $\partial\ell_T(\boldsymbol{\theta}_0)/\partial\boldsymbol{\psi}$, and note that it can be written as

$$\frac{\partial\ell_T(\boldsymbol{\theta}_0)}{\partial\psi_i} = -Tg_{0,ii} + \frac{1}{\sigma_{i0}^2}\sum_{t=1}^T y_{it}^*\varepsilon_{it}.$$

Also $y_{it}^* = \mathbf{e}_{i,N}'\boldsymbol{G}_0(\mathbf{B}_0\boldsymbol{x}_{\circ t} + \boldsymbol{\varepsilon}_{\circ t})$, where $\boldsymbol{G}_0 = \boldsymbol{W}(\mathbf{I}_N - \boldsymbol{\Psi}_0\boldsymbol{W})^{-1}$ and $\boldsymbol{e}_{i,N}$ is an N dimensional vector with its i^{th} element unity and zeros elsewhere. Then

$$\frac{1}{T}\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \psi_i} = \frac{1}{T}\sum_{t=1}^{T}\eta_{it}, \tag{A.26}$$

where η_{it} is already defined by (A.19) which we write as

$$\eta_{it} = \sigma_{i0}^{-1}\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\zeta_{it} + \sigma_{i0}^{-1}\sigma_{i0}^{-1}\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it} - g_{0,ii}, \tag{A.27}$$

and as in proof of Lemma 4, $\boldsymbol{\varphi}_i' = \mathbf{e}_{i,N}'\boldsymbol{G}_0\mathbf{B}_0$, $\boldsymbol{\vartheta}_i' = \mathbf{e}_{i,N}'\boldsymbol{G}_0\boldsymbol{\Sigma}_0^{1/2} = (g_{i1}\sigma_{10}, g_{i2}\sigma_{20}, \ldots, g_{iN}\sigma_{N0})$, $\boldsymbol{\zeta}_{\circ t} = (\zeta_{1t}, \zeta_{2t}, \ldots, \zeta_{Nt})'$, and $\zeta_{it} = \varepsilon_{it}/\sigma_{i0}$. Also recall that by Lemma 4, $E(\eta_{it}|\mathcal{F}_t) = 0$, and $\sup_{i,t} E|\eta_{it}|^{2+c} < K$, for some $c > 0$. Therefore, using (A.26) by the strong law of large numbers for martingales we have (see, for example, White (1984))

$$\frac{1}{T}\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\psi}} \overset{a.s.}{\to} 0. \tag{A.28}$$

Further, since $E(\eta_{it}|\mathcal{F}_t) = 0$, then using (A.27)

$$\begin{aligned} Var(\eta_{it}) = E[Var(\eta_{it} \mid \mathcal{F}_t)] &= \sigma_{i0}^{-2}\mathbf{g}_{0,i}'\mathbf{B}_0\boldsymbol{\Sigma}_{xx}\mathbf{B}_0'\mathbf{g}_{0,i} + \sigma_{i0}^{-2}\sum_{j=1}^{N}\sigma_{j0}^2 g_{0,ij}^2 + g_{0,ii}^2\left[E\left(\zeta_{it}^4\right) - 2\right] \\ &= \sigma_{i0}^{-2}\sum_{r=1}^{N}\sum_{s=1}^{N} g_{0,is}g_{0,ir}\boldsymbol{\beta}_{r0}'\boldsymbol{\Sigma}_{rs}\boldsymbol{\beta}_{s0} + \sigma_{i0}^{-2}\sum_{j=1}^{N}\sigma_{j0}^2 g_{0,ij}^2 + g_{0,ii}^2\left[E\left(\zeta_{it}^4\right) - 2\right]. \end{aligned} \tag{A.29}$$

Consider now the limiting distribution of $\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \psi_i} = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_{it}$, and note that by Lemma 4, $\sup_{i,t} E|\eta_{it}|^{2+c} < K$ for some $c > 0$, and by Corollary 5.25 in White (1984) it follows that

$$\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \psi_i} \to_d N(0, \omega_{ii}), \text{ as } T \to \infty, \tag{A.30}$$

where (using (A.29))

$$\begin{aligned} \omega_{ii} &= \lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} Var(\eta_{it}) \\ &= g_{0,ii}^2\left[\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left(\zeta_{it}^4\right) - 2\right] + \sigma_{i0}^{-2}\mathbf{g}_{0,i}'\mathbf{B}_0\boldsymbol{\Sigma}_{xx}\mathbf{B}_0'\mathbf{g}_{0,i} + \sigma_{i0}^{-2}\sum_{j=1}^{N}\sigma_{j0}^2 g_{0,ij}^2, \end{aligned}$$

which exists and is finite under Assumptions 1 and 2(b).
Similarly, consider the i^{th} component of $\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}}$. Then, write

$$\frac{1}{T}\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_i} = \frac{1}{\sigma_{i0}^2}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\varepsilon_{it} = \frac{1}{\sigma_{i0}}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_{it}\zeta_{it}. \tag{A.31}$$

But by Assumption 2(a), $E(\boldsymbol{x}_{it}\zeta_{it} \mid \mathcal{F}_t) = (1/\sigma_{i0})\,\boldsymbol{x}_{it}E(\varepsilon_{it}|\mathcal{F}_t) = \mathbf{0}$, and $Var(\boldsymbol{x}_{it}\zeta_{it} \mid \mathcal{F}_t) =$

$\boldsymbol{x}_{it}\boldsymbol{x}_{it}'E\left(\zeta_{it}^2 \mid \mathcal{F}_t\right) = \boldsymbol{x}_{it}\boldsymbol{x}_{it}'$. Hence $E\left(\boldsymbol{x}_{it}\zeta_{it}\right) = \mathbf{0}$, and $Var\left(\boldsymbol{x}_{it}\zeta_{it}\right) = \boldsymbol{\Sigma}_{ii} < K$. Therefore, noting that $\boldsymbol{x}_{it}\zeta_{it}$ is a martingale difference process with finite second-order moments, it follows that

$$\frac{1}{T}\frac{\partial \ell_T\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\beta}} \overset{a.s.}{\to} \mathbf{0}, \text{ as } T \to \infty. \tag{A.32}$$

Denote the ℓ^{th} element of $\boldsymbol{x}_{it}\zeta_{it}$ by $z_{i\ell,t} = x_{i\ell,t}\zeta_{it}$ for $\ell = 1, 2, \dots, k$, and note that the ℓ^{th} element of $\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\beta}_i}$ is given $\frac{1}{\sigma_{i0}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{i\ell,t}$, where $z_{i\ell,t}$ is a martingale difference process with respect to $\mathcal{F}_t$. Also, by Assumptions 1 and 2(a),

$$\begin{aligned}
\sup_{i,\ell,t} E\left|z_{i\ell,t}\right|^p &= \sup_{i,\ell,t} E\left|x_{i\ell,t}\zeta_{it}\right|^p = \sup_{i,\ell,t} E\left[E\left(\left|x_{i\ell,t}\zeta_{it}\right|^p \left|\mathcal{F}_t\right.\right)\right] \leq \sup_{i,\ell,t} E\left[\left|x_{i\ell,t}\right|^p E\left(\left|\zeta_{it}\right|^p \left|\mathcal{F}_t\right.\right)\right] \\
&= \sup_{i,\ell,t}\left(E\left|x_{i\ell,t}\right|^p\right)\sigma_{i0}^{-p}\varpi_{ip} < K,
\end{aligned}$$

for $p = 2 + c$, $c > 0$. Hence, by Corollary 5.25 in White (1984) it follows that for each i and ℓ and as $T \to \infty$, $\frac{1}{\sigma_{i0}}\frac{1}{\sqrt{T}}\sum_{t=1}^{T} z_{i\ell,t}$ tends to a normal distribution and as whole we have

$$\frac{1}{\sqrt{T}}\frac{\partial \ell_T\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\beta}_i} \to_d N\left(\mathbf{0}, \boldsymbol{\Omega}_i\right), \tag{A.33}$$

where

$$\boldsymbol{\Omega}_i = \frac{1}{\sigma_{i0}^2}\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T} E\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right). \tag{A.34}$$

Finally, consider the i^{th} component of $\frac{\partial \ell_T(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\sigma}^2}$, and note that

$$\frac{1}{T}\frac{\partial \ell_T\left(\boldsymbol{\theta}_0\right)}{\partial \sigma_i^2} = \frac{1}{2\sigma_{i0}^2}\frac{1}{T}\sum_{t=1}^{T}\left(\frac{\varepsilon_{it}^2}{\sigma_{i0}^2} - 1\right) = \frac{1}{2\sigma_{i0}^2}\left[\frac{1}{T}\sum_{t=1}^{T}\left(\zeta_{it}^2 - 1\right)\right].$$

Let $\xi_{it} = \zeta_{it}^2 - 1$,where $\zeta_{it} = \varepsilon_{it}/\sigma_{i0}$. Then

$$\frac{1}{T}\frac{\partial \ell_T\left(\boldsymbol{\theta}_0\right)}{\partial \sigma_i^2} = \frac{1}{2\sigma_{i0}^2}\left[\frac{1}{T}\sum_{t=1}^{T}\xi_{it}\right]. \tag{A.35}$$

We have $E\left(\xi_{it} \mid \mathcal{F}_t\right) = E\left(\zeta_{it}^2 \mid \mathcal{F}_t\right) - 1 = 0$, and $E\left(\xi_{it}^2 \mid \mathcal{F}_t\right) = E\left(\zeta_{it}^4 \mid \mathcal{F}_t\right) - 1$, so that, since under Assumption 1 ξ_{it}'s are martingale difference processes and $E(\left|\varepsilon_{it}\right|^{4+\epsilon}\left|\mathcal{F}_t\right.) < K$, for some small positive ϵ, then $\sup_i E\left|\xi_{it}\right|^2 < K$ and by the strong law of large numbers for martingale processes we have

$$\frac{1}{T}\frac{\partial \ell_T\left(\boldsymbol{\theta}_0\right)}{\partial \boldsymbol{\sigma}^2} \overset{a.s.}{\to} 0, \text{ as } T \to \infty. \tag{A.36}$$

Similarly, since $\sup_i E\left|\xi_{it}\right|^{2+c} < K$ for some $c > 0$, then as before

$$T^{-1/2}\frac{\partial \ell_T\left(\boldsymbol{\theta}_0\right)}{\partial \sigma_i^2} \to_d N(0, v_{ii}), \tag{A.37}$$

where

$$v_{ii} = \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} \left[\frac{1}{4\sigma_{i0}^4} Var(\xi_{it}) \right] = \left(\frac{1}{4\sigma_{i0}^4} \right) \lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} Var(\zeta_{it}^2). \tag{A.38}$$

Now results (A.28), (A.32) and (A.36) establish (A.22), and results (A.30), (A.33) and (A.37) establish (A.23), as required, with $\mathbf{J}(\boldsymbol{\theta}_0, \gamma) = \lim_{T\to\infty} \sum_{t=1}^{T} E_0 \left[\frac{1}{T} \left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \left(\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)' \right]$. Consistency of $\hat{\mathbf{J}}(\hat{\boldsymbol{\theta}}, \hat{\gamma})$ for $\mathbf{J}(\boldsymbol{\theta}_0, \gamma)$ follows from consistency of $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}_0$, and $\hat{\gamma}$ for γ, and independence of $\frac{\partial \ell_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ over t. Further, since $\hat{\boldsymbol{\theta}} \overset{a.s.}{\to} \boldsymbol{\theta}_0$ on $\boldsymbol{\Theta}_c$, as $T \to \infty$, as shown in Section 3.3, we have

$$\hat{\varepsilon}_{it} = \varepsilon_{it} + O_p\left(\frac{1}{\sqrt{T}} \right), \text{ and } \hat{\sigma}_i^2 = \sigma_i^2 + O_p\left(\frac{1}{\sqrt{T}} \right),$$

which establishes that

$$\hat{\gamma} = (NT)^{-1} \sum_{t=1}^{T} \sum_{i=1}^{N} \left(\frac{\hat{\varepsilon}_{it}}{\hat{\sigma}_i} \right)^4 - 1 \to_p \gamma, \text{ as } T \to \infty, \text{ for any } N.$$

■

Appendix B Proof of Proposition 2

Proof of Proposition 2. First, we consider the information matrix $\boldsymbol{H}(\boldsymbol{\theta}_0)$ given by

$$\boldsymbol{H}(\boldsymbol{\theta}_0) = \lim_{T\to\infty} E_0 \left[-\frac{1}{T} \frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right], \tag{B.39}$$

where

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

We evaluate each partial derivative in (B.39):
$\boldsymbol{H}_{11} = E_0 \left[-\frac{1}{T} \frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\psi} \partial \boldsymbol{\psi}'} \right]$ is given by the $N \times N$ matrix

$$\boldsymbol{H}_{11} = \left(\mathbf{G}_0 \odot \mathbf{G}_0' \right) + \text{Diag} \left[\frac{1}{\sigma_{i0}^2} \frac{1}{T} \sum_{t=1}^{T} E_0 \left(y_{it}^{*2} \right), \, i = 1, 2, \ldots, N \right],$$

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

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^{T} E_0 \left(y_{it}^{*2} \right) &= \boldsymbol{w}_i' (\mathbf{I}_N - \boldsymbol{\Psi}_0 \boldsymbol{W})^{-1} \left[\boldsymbol{B}_0 E \left(\boldsymbol{x}_{\circ t} \boldsymbol{x}_{\circ t}' \right) \boldsymbol{B}_0' + \boldsymbol{\Sigma}_0 \right] (\mathbf{I}_N - \boldsymbol{W}' \boldsymbol{\Psi}_0)^{-1} \boldsymbol{w}_i \\ &= \boldsymbol{g}_{0i}' \left(\boldsymbol{B}_0 \boldsymbol{\Sigma}_{xx} \boldsymbol{B}_0' + \boldsymbol{\Sigma}_0 \right) \boldsymbol{g}_{0i} \\ &= \sum_{r=1}^{N} \sum_{s=1}^{N} g_{0,is} g_{0,ir} \boldsymbol{\beta}_{r0}' \boldsymbol{\Sigma}_{rs} \boldsymbol{\beta}_{s0} + \sum_{s=1}^{N} g_{0,is}^2 \sigma_{s0}^2. \end{aligned}$$

Note that as shown in Lemmas 2 and 5,

$$\|\mathbf{G}_0\|_\infty < K, \ \ \left\|\mathbf{G}_0 \odot \mathbf{G}_0'\right\|_\infty < K \text{ and } \left\|\sum_{r=1}^{N}\sum_{s=1}^{N} g_{0,is}g_{0,ir}\boldsymbol{\beta}_{r0}'\boldsymbol{\Sigma}_{rs}\boldsymbol{\beta}_{s0}\right\|_\infty < K.$$

$\boldsymbol{H}_{12} = E_0\left[-\frac{1}{T}\frac{\partial^2\ell_T(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\beta}'}\right]$ is an $N \times kN$ matrix with its i^{th} row given by a $1 \times kN$ vector of zeros except for its i^{th} block which is given by the $1 \times k$ vector $\sigma_{i0}^{-2}E_0(T^{-1}\boldsymbol{y}_i^{*\prime}\mathbf{X}_i)$, namely

$$\boldsymbol{H}_{12} = \begin{pmatrix} \sigma_{10}^{-2}E_0(T^{-1}\boldsymbol{y}_1^{*\prime}\mathbf{X}_1) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \sigma_{20}^{-2}E_0(T^{-1}\boldsymbol{y}_2^{*\prime}\mathbf{X}_2) & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \sigma_{N0}^{-2}E_0(T^{-1}\boldsymbol{y}_N^{*\prime}\mathbf{X}_N) \end{pmatrix},$$

where

$$\begin{aligned} E_0(T^{-1}\boldsymbol{y}_i^{*\prime}\mathbf{X}_i) &= E_0\left(T^{-1}\sum_{t=1}^{T} y_{it}^{*}\boldsymbol{x}_{it}'\right) = E_0\left(T^{-1}\sum_{t=1}^{T}\boldsymbol{w}_i'\boldsymbol{y}_{\circ t}\boldsymbol{x}_{it}'\right) \\ &= \boldsymbol{w}_i'(\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})^{-1}\mathbf{B}E\left(\boldsymbol{x}_{\circ t}\boldsymbol{x}_{it}'\right) \\ &= \boldsymbol{g}_{0i}'\left(\boldsymbol{\Sigma}_{i1}\boldsymbol{\beta}_{10}, \boldsymbol{\Sigma}_{i2}\boldsymbol{\beta}_{20}, \ldots, \boldsymbol{\Sigma}_{iN}\boldsymbol{\beta}_{N0}\right)' = \sum_{s=1}^{N} g_{0,is}\boldsymbol{\beta}_{s0}'\boldsymbol{\Sigma}_{is}. \end{aligned}$$

Again by Assumptions 2(b), 3 and 5, $\sup_s \|\boldsymbol{\beta}_{s0}\|_1$ and $\boldsymbol{\Sigma}_{is}$ exist and are finite. Also, $\max_i \sum_{s=1}^{N}|g_{0,is}| = \|\boldsymbol{G}_0\|_\infty$ which is bounded under our assumptions. $\mathbf{H}_{13} = E_0\left[-\frac{1}{T}\frac{\partial^2\ell_T(\boldsymbol{\theta})}{\partial\boldsymbol{\psi}\partial\boldsymbol{\sigma}^{2\prime}}\right]$ is an $N \times N$ diagonal matrix with its i^{th} element given by $\sigma_{i0}^{-2}\boldsymbol{w}_i'(\mathbf{I}_N - \boldsymbol{\Psi}_0\mathbf{W})^{-1}\mathbf{e}_{i,N} = \sigma_{i0}^{-2}g_{0,ii}$, where $\mathbf{e}_{i,N}$ is an N dimensional vector with its i^{th} element unity and zeros elsewhere. $\boldsymbol{H}_{22} = E_0\left[-\frac{1}{T}\frac{\partial^2\ell_T(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right]$ is an $Nk \times Nk$ block diagonal matrix with its i^{th} block given by $\sigma_{i0}^{-2}\boldsymbol{\Sigma}_{ii}$. $\boldsymbol{H}_{23} = \mathbf{0}$, and finally $\boldsymbol{H}_{33} = E_0\left[-\frac{1}{T}\frac{\partial^2\ell_T(\boldsymbol{\theta})}{\partial(\boldsymbol{\sigma}^2)\partial(\boldsymbol{\sigma}^{2\prime})}\right] = \text{Diag}(1/2\sigma_{10}^4, 1/2\sigma_{20}^4, \ldots, 1/2\sigma_{N0}^4)$. Collecting all terms, we obtain (B.39).

Next, recalling from Lemma 4 that

$$\eta_{it} = \sigma_{i0}^{-1}\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\zeta_{it} + \sigma_{i0}^{-1}\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it} - g_{0,ii},$$

where $\boldsymbol{\varphi}_i' = \mathbf{e}_{i,N}'\boldsymbol{G}_0\mathbf{B}_0$ and $\boldsymbol{\vartheta}_i' = \mathbf{e}_{i,N}'\boldsymbol{G}_0\boldsymbol{\Sigma}_0^{1/2}$, and using (A.26) and (A.29) of Lemma 5, we have the following cross-products (for $i \neq j$)

$$\begin{aligned}
E_0\left[\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \psi_i}\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \psi_j}\right] &= \frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E_0\left(\eta_{it}\eta_{jt'}\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left\{\begin{array}{c} \left[\sigma_{i0}^{-1}\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\zeta_{it} + \sigma_{i0}^{-1}\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it} - g_{0,ii}\right] \\ \left[\sigma_{j0}^{-1}\boldsymbol{\varphi}_j'\boldsymbol{x}_{\circ t'}\zeta_{jt'} + \sigma_{j0}^{-1}\boldsymbol{\vartheta}_j'\boldsymbol{\zeta}_{\circ t'}\zeta_{jt'} - g_{0,jj}\right] \end{array}\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T} E\left\{\begin{array}{c} \left[\sigma_{i0}^{-1}\boldsymbol{\varphi}_i'\boldsymbol{x}_{\circ t}\zeta_{it} + \sigma_{i0}^{-1}\boldsymbol{\vartheta}_i'\boldsymbol{\zeta}_{\circ t}\zeta_{it} - g_{0,ii}\right] \\ \left[\sigma_{j0}^{-1}\boldsymbol{\varphi}_j'\boldsymbol{x}_{\circ t}\zeta_{jt} + \sigma_{j0}^{-1}\boldsymbol{\vartheta}_j'\boldsymbol{\zeta}_{\circ t}\zeta_{jt} - g_{0,jj}\right] \end{array}\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\left\{\begin{array}{c} \sigma_{i0}^{-1}\sigma_{j0}^{-1}\boldsymbol{\varphi}_i'E\left(\boldsymbol{x}_{\circ t}\boldsymbol{x}_{\circ t}'\zeta_{it}\zeta_{jt}\right)\boldsymbol{\varphi}_j + \sigma_{i0}^{-1}\sigma_{j0}^{-1}\boldsymbol{\vartheta}_i'E\left(\zeta_{it}\zeta_{jt}\boldsymbol{\zeta}_{\circ t}\boldsymbol{\zeta}_{\circ t}'\right)\boldsymbol{\vartheta}_j - g_{0,ii}g_{0,jj} \\ +\sigma_{i0}^{-1}\sigma_{j0}^{-1}\boldsymbol{\varphi}_i'E_0\left(\boldsymbol{x}_{\circ t}\boldsymbol{\zeta}_{\circ t}'\boldsymbol{\vartheta}_j\zeta_{it}\zeta_{jt}\right) + \sigma_{j0}^{-1}\sigma_{i0}^{-1}\boldsymbol{\varphi}_j'E_0\left(\boldsymbol{x}_{\circ t}\boldsymbol{\zeta}_{\circ t}'\boldsymbol{\vartheta}_j\zeta_{jt}\zeta_{it}\right) \end{array}\right. \\
&= \left\{\begin{array}{ll} \left\{\begin{array}{c} g_{0,ii}^2\left[\frac{1}{T}\sum_{t=1}^{T}E\left(\zeta_{it}^4\right) - 2\right] + \sigma_{i0}^{-2}\mathbf{g}_{0,i}'\mathbf{B}_0\boldsymbol{\Sigma}_{xx}\mathbf{B}_0'\mathbf{g}_{0,i} \\ +\sigma_{i0}^{-2}\sum_{j=1}^{N}\sigma_{j0}^2 g_{0,ij}^2 \end{array}\right. & \text{, for } i=j \\ g_{0,ij}g_{0,ji}, & \text{,for } i \neq j \end{array}\right. .
\end{aligned}$$

Further, using (A.31) and (A.34) of Lemma 5 we have

$$\begin{aligned}
E_0\left[\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i}\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_j'}\right] &= \frac{1}{\sigma_{i0}\sigma_{j0}}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left(\boldsymbol{x}_{it}\zeta_{it}\zeta_{jt'}\boldsymbol{x}_{jt'}'\right) \\
&= \left\{\begin{array}{c} \frac{1}{\sigma_{i0}^2 T}\sum_{t=1}^{T}E\left(\boldsymbol{x}_{it}\boldsymbol{x}_{it}'\right), \text{ for } i=j \\ \mathbf{0}, \text{ for } i \neq j \end{array}\right. ,
\end{aligned}$$

and using (A.35) and (A.38) of Lemma 5 we have

$$\begin{aligned}
E_0\left[\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \sigma_i^2}\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \sigma_j^2}\right] &= \frac{1}{4\sigma_{i0}^2\sigma_{j0}^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left(\xi_{it}\xi_{jt'}\right) \\
&= \frac{1}{4\sigma_{i0}^2\sigma_{j0}^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left[\left(\zeta_{it}^2-1\right)\left(\zeta_{jt'}^2-1\right)\right] \\
&= \frac{1}{4\sigma_{i0}^2\sigma_{j0}^2}\frac{1}{T}\sum_{t=1}^{T} E\left(\zeta_{it}^2\zeta_{jt}^2 - \zeta_{it}^2 - \zeta_{jt}^2 + 1\right) \\
&= \left\{\begin{array}{c} \frac{1}{4\sigma_{i0}^4 T}\left[\sum_{t=1}^{T}E\left(\zeta_{it}^4\right) - 1\right], \text{ for } i=j \\ 0, \text{ for } i \neq j \end{array}\right. .
\end{aligned}$$

In addition,

$$
\begin{aligned}
E_0\left[\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i}\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \psi_j}\right] &= \frac{1}{\sigma_{i0}T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E_0\left[\left(\boldsymbol{x}_{it}\zeta_{it}\right)\eta_{jt'}\right] \\
&= \frac{1}{\sigma_{i0}T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E_0\left[\boldsymbol{x}_{it}\zeta_{it}\left(\sigma_{j0}^{-1}\boldsymbol{\varphi}_j'\boldsymbol{x}_{\circ t'}\zeta_{jt'}+\sigma_{j0}^{-1}\boldsymbol{\vartheta}_j'\boldsymbol{\zeta}_{\circ t'}\zeta_{jt'}-g_{0,jj}\right)\right] \\
&= \frac{1}{\sigma_{i0}T}\sum_{t=1}^{T}\left[\begin{array}{c}\sigma_{j0}^{-1}E_0\left(\boldsymbol{x}_{it}\boldsymbol{\varphi}_j'\boldsymbol{x}_{\circ t}\zeta_{it}\zeta_{jt}\right)+\sigma_{j0}^{-1}E_0\left(\boldsymbol{x}_{it}\zeta_{it}\boldsymbol{\vartheta}_j'\boldsymbol{\zeta}_{\circ t}\zeta_{jt}\right) \\ -g_{0,jj}E\left(\boldsymbol{x}_{it}\zeta_{it}\right)\end{array}\right] \\
&= \left\{\begin{array}{c}\sigma_{i0}^{-2}\boldsymbol{g}_{0i}'\left(\boldsymbol{\Sigma}_{i1}\boldsymbol{\beta}_{10},\boldsymbol{\Sigma}_{i2}\boldsymbol{\beta}_{20},\ldots,\boldsymbol{\Sigma}_{iN}\boldsymbol{\beta}_{N0}\right)',\ \text{for } i=j \\ \mathbf{0},\ \text{for } i\neq j\end{array}\right. .
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
E_0\left[\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \sigma_i^2}\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \psi_j}\right] &= \frac{1}{2\sigma_{i0}^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left(\xi_{it}\eta_{jt'}\right) \\
&= \frac{1}{2\sigma_{i0}^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left[\left(\zeta_{it}^2-1\right)\left(\sigma_{j0}^{-1}\boldsymbol{\varphi}_j'\boldsymbol{x}_{\circ t'}\zeta_{jt'}+\sigma_{j0}^{-1}\boldsymbol{\vartheta}_j'\boldsymbol{\zeta}_{\circ t'}\zeta_{jt'}-g_{0,jj}\right)\right] \\
&= \frac{1}{2\sigma_{i0}^2}\frac{1}{T}\sum_{t=1}^{T} E\left[\zeta_{it}^2\left(\sigma_{j0}^{-1}\boldsymbol{\varphi}_j'\boldsymbol{x}_{\circ t}\zeta_{jt}+\sigma_{j0}^{-1}\boldsymbol{\vartheta}_j'\boldsymbol{\zeta}_{\circ t}\zeta_{jt}-g_{0,jj}\right)\right] \\
&= \left\{\begin{array}{c}\frac{g_{0,ii}\left[\frac{1}{T}\sum_{t=1}^{T}E\left(\zeta_{it}^4\right)-1\right]}{2\sigma_{i0}^2},\ \text{for } i=j \\ 0,\ \text{for } i\neq j\end{array}\right. ,
\end{aligned}
$$

and finally,

$$
\begin{aligned}
E_0\left[\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}_i}\frac{1}{\sqrt{T}}\frac{\partial \ell_T(\boldsymbol{\theta})}{\partial \sigma_j^2}\right] &= \frac{1}{2\sigma_{i0}\sigma_{j0}^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left[\left(\boldsymbol{x}_{it}\zeta_{it}\right)\xi_{jt'}\right] \\
&= \frac{1}{2\sigma_{i0}\sigma_{j0}^2}\frac{1}{T}\sum_{t=1}^{T}\sum_{t'=1}^{T} E\left[\left(\boldsymbol{x}_{it}\zeta_{it}\right)\left(\zeta_{jt'}^2-1\right)\right] \\
&= 0,\ \text{for all } i,j=1,2,\ldots,N.
\end{aligned}
$$

Overall, let

$$
\gamma=\left[\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}E\left(\zeta_{it}^4\right)-1\right]=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}Var(\zeta_{it}^2). \tag{B.40}
$$

We can collect the various terms and construct matrix

$$
\mathbf{J}\left(\boldsymbol{\theta}_0,\gamma\right)=\left(J_{0,ij}\right)=\lim_{T\to\infty}E_0\left[\frac{1}{T}\left(\sum_{t=1}^{T}\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\sum_{t=1}^{T}\frac{\partial \ell_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)'\right]=\left(\begin{array}{ccc}\boldsymbol{J}_{11} & \boldsymbol{J}_{12} & \boldsymbol{J}_{13} \\ . & \boldsymbol{J}_{22} & \boldsymbol{J}_{23} \\ . & . & \boldsymbol{J}_{33}\end{array}\right)_{N(k+2)\times N(k+2)}, \tag{B.41}
$$

where $\ell_t(\boldsymbol{\theta})$ is defined in (A.20) and

$$\boldsymbol{J}_{11} = \begin{cases} (\boldsymbol{G}_0 \odot \boldsymbol{G}_0') + (\gamma - 2)\operatorname{Diag}(\boldsymbol{G}_0 \odot \boldsymbol{G}_0') \\ +\operatorname{Diag}\left[\sigma_{i0}^{-2}\boldsymbol{g}_{0i}'(\boldsymbol{B}_0\boldsymbol{\Sigma}_{xx}\boldsymbol{B}_0' + \boldsymbol{\Sigma}_0)\boldsymbol{g}_{0i},\ i = 1, 2, \ldots, N\right] \end{cases},$$

$$\boldsymbol{J}_{12} = \operatorname{Diag}\left[\sigma_{i0}^{-2}\sum_{s=1}^{N} g_{0,is}\boldsymbol{\beta}_{s0}'\boldsymbol{\Sigma}_{is},\ i = 1, 2, \ldots, N\right],$$

$$\boldsymbol{J}_{13} = \frac{\gamma}{2}\operatorname{Diag}\left(\sigma_{i0}^{-2}g_{0,ii},\ i = 1, 2, \ldots, N\right),\ \boldsymbol{J}_{22} = \operatorname{Diag}\left(\sigma_{i0}^{-2}\boldsymbol{\Sigma}_{ii},\ i = 1, 2, \ldots, N\right),$$

$$\boldsymbol{J}_{23} = \boldsymbol{0},\ \boldsymbol{J}_{33} = \frac{\gamma}{4}\operatorname{Diag}\left(1/\sigma_{i0}^{4},\ i = 1, 2, \ldots, N\right).$$

Having established that the score vector is asymptotically normally distributed, it is now easily seen that as $T \to \infty$,

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \to_d N\left(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\theta}}\right), \tag{B.42}$$

where $\boldsymbol{V}_{\boldsymbol{\theta}} = \boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)\,\mathbf{J}(\boldsymbol{\theta}_0, \gamma)\,\boldsymbol{H}^{-1}(\boldsymbol{\theta}_0)$. ■

Appendix C Estimator of $\boldsymbol{V}\left(\hat{\boldsymbol{\theta}}\right)$

Derivatives of the log-likelihood function

The vector of maximum likelihood estimates, $\hat{\boldsymbol{\theta}}_T$, in Section 2 is obtained by maximizing the log-likelihood function (A.21) which we reproduce here for convenience[26]

$$\ell_T(\boldsymbol{\theta}) = -\frac{NT}{2}\ln(2\pi) - \frac{T}{2}\sum_{i=1}^{N}\ln\sigma_i^2 + T\ln|\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}| - \frac{1}{2}\sum_{i=1}^{N}\frac{(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i)'(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i)}{\sigma_i^2}, \tag{C.43}$$

where $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\beta}', \boldsymbol{\sigma}^{2\prime})'$.

First derivatives

We have

$$\frac{\partial\ell_T(\boldsymbol{\theta})}{\partial\psi_i} = -T\operatorname{tr}[(\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W})^{-1}\boldsymbol{E}_{ii}\boldsymbol{W}] + \frac{\boldsymbol{y}_{i\circ}^{*\prime}(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i)}{\sigma_i^2}, \text{ for } i = 1, 2, \ldots, N,$$

$$\frac{\partial\ell_T(\boldsymbol{\theta})}{\partial\boldsymbol{\beta}_i} = \frac{\mathbf{X}_{i\circ}'(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i)}{\sigma_i^2}, \text{ for } i = 1, 2, \ldots, N,$$

$$\frac{\partial\ell_T(\boldsymbol{\theta})}{\partial\sigma_i^2} = -\frac{T}{2\sigma_i^2} + \frac{1}{2\sigma_i^4}(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i)'(\boldsymbol{y}_{i\circ} - \psi_i\boldsymbol{y}_{i\circ}^* - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i), \text{ for } i = 1, 2, \ldots, N,$$

where $\boldsymbol{E}_{ii}$ is the $N \times N$ matrix whose (i, i) element is 1 and zero elsewhere.

[26]Note that $\ln|\boldsymbol{S}'(\psi)\mathbf{S}(\psi)| = 2\ln|\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}|$.

Second derivatives

We have

$$\hat{\boldsymbol{H}}\left(\boldsymbol{\theta}\right) = -\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'} = \begin{pmatrix} \hat{\boldsymbol{H}}_{11} & \hat{\boldsymbol{H}}_{12} & \hat{\boldsymbol{H}}_{13} \\ . & \hat{\boldsymbol{H}}_{22} & \hat{\boldsymbol{H}}_{23} \\ . & . & \hat{\boldsymbol{H}}_{33} \end{pmatrix},$$

$$= \begin{pmatrix} -\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'} & -\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\beta}'} & -\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\psi}\partial \boldsymbol{\sigma}^{2\prime}} \\ . & -\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} & -\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\sigma}^{2\prime}} \\ . & . & -\frac{1}{T}\frac{\partial^2 \ell_T(\boldsymbol{\theta})}{\partial (\boldsymbol{\sigma}^2)\partial (\boldsymbol{\sigma}^{2\prime})} \end{pmatrix}.$$

With the (i,j) or i^{th} element of associated matrix or vector given in $\{\}$, we have

$$\hat{\boldsymbol{H}}_{11} = -\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\psi}'} = \left\{-\frac{1}{T}\frac{\partial^2 \ell_{NT}\left(\boldsymbol{\theta}\right)}{\partial \psi_i \partial \psi_j}\right\},$$

$$-\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \psi_i \partial \psi_j} = \begin{cases} \operatorname{tr}\left[(\boldsymbol{I}-\boldsymbol{\Psi}\boldsymbol{W})^{-1}\boldsymbol{E}_{ii}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{\Psi}\boldsymbol{W})^{-1}\boldsymbol{E}_{ii}\boldsymbol{W}\right] + \frac{1}{\sigma_i^2}\frac{\boldsymbol{y}_{i\circ}^{*\prime}\boldsymbol{y}_{i\circ}^{*}}{T} & \text{if } i=j \\ \operatorname{tr}\left[(\boldsymbol{I}-\boldsymbol{\Psi}\boldsymbol{W})^{-1}\boldsymbol{E}_{jj}\boldsymbol{W}(\boldsymbol{I}-\boldsymbol{\Psi}\boldsymbol{W})^{-1}\boldsymbol{E}_{ii}\boldsymbol{W}\right] & \text{if } i \neq j \end{cases},$$

$$-\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \psi_i \partial \psi_j} = \begin{cases} g_{ii}^2 + \frac{1}{\sigma_i^2}\frac{\boldsymbol{y}_{i\circ}^{*\prime}\boldsymbol{y}_{i\circ}^{*}}{T} & \text{if } i=j \\ g_{ij}g_{ji} & \text{if } i \neq j \end{cases},$$

where $\boldsymbol{E}_{ij}$ is the $N \times N$ matrix whose (i,j) element is 1 and zero elsewhere and $\mathbf{G} = (g_{ij}) = \boldsymbol{W}\left(\boldsymbol{I}_N - \boldsymbol{\Psi}\boldsymbol{W}\right)^{-1}$. Further,

$$\hat{\boldsymbol{H}}_{12} = -\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\beta}'} = \left\{-\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \psi_i \partial \boldsymbol{\beta}_j'}\right\} = \left\{\frac{1}{\sigma_i^2}\frac{\boldsymbol{y}_{i\circ}^{*\prime}\boldsymbol{X}_{i\circ}}{T}, \text{ if } i=j, \text{ and } \mathbf{0}, \text{ if } i \neq j\right\},$$

$$\hat{\boldsymbol{H}}_{13} = -\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\psi}\partial \boldsymbol{\sigma}^{2\prime}} = \left\{-\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \psi_i \partial \sigma_j^2}\right\} = \left\{\frac{1}{\sigma_i^4}\frac{\boldsymbol{y}_{i\circ}^{*\prime}\left(\boldsymbol{y}_{i\circ} - \psi_i \boldsymbol{y}_{i\circ}^{*} - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i\right)}{T}, \text{ if } i=j, \text{ and } 0, \text{ if } i \neq j\right\},$$

$$\hat{\boldsymbol{H}}_{22} = -\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = \left\{-\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\beta}_i \partial \boldsymbol{\beta}_j'}\right\} = \left\{\frac{1}{\sigma_i^2}\frac{\boldsymbol{X}_{i\circ}'\boldsymbol{X}_{i\circ}}{T}, \text{ if } i=j, \text{ and } \mathbf{0}, \text{ if } i \neq j\right\},$$

$$\hat{\boldsymbol{H}}_{23} = -\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\beta}\partial \boldsymbol{\sigma}^{2\prime}} = \left\{-\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\beta}_i \partial \sigma_j^2}\right\} = \left\{\frac{1}{\sigma_i^4}\frac{\mathbf{X}_{i\circ}'(\boldsymbol{y}_{i\circ} - \psi_i \boldsymbol{y}_{i\circ}^{*} - \boldsymbol{X}_{i\circ}\boldsymbol{\beta}_i)}{T}, \text{ if } i=j, \text{ and } \mathbf{0}, \text{ if } i \neq j\right\},$$

$$\hat{\boldsymbol{H}}_{33} = -\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial\left(\boldsymbol{\sigma}^2\right)\partial\left(\boldsymbol{\sigma}^{2\prime}\right)} = \left\{-\frac{1}{T}\frac{\partial^2 \ell_T\left(\boldsymbol{\theta}\right)}{\partial \sigma_i^2 \partial \sigma_j^2}\right\}$$

$$= \left\{-\frac{1}{2\sigma_i^4} + \frac{1}{\sigma_i^6}\frac{1}{T}(\boldsymbol{y}_{i\circ} - \psi_i \boldsymbol{y}_{i\circ}^{*} - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i)'(\boldsymbol{y}_{i\circ} - \psi_i \boldsymbol{y}_{i\circ}^{*} - \mathbf{X}_{i\circ}\boldsymbol{\beta}_i), \text{ if } i=j, \text{ and } 0, \text{ if } i \neq j\right\}.$$

Finally, from the above results we obtain:

$$\hat{\boldsymbol{J}}\left(\boldsymbol{\theta}\right)=\frac{1}{T}\left\{\left[\sum_{t=1}^{T}\frac{\partial\ell_{t}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right]\left[\sum_{t=1}^{T}\frac{\partial\ell_{t}\left(\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\right]'\right\} \text{ and } \hat{\boldsymbol{H}}\left(\boldsymbol{\theta}\right)=-\frac{1}{T}\frac{\partial^{2}\ell_{T}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'},$$

from which the standard and sandwich covariance matrix estimators (49) and (50), are given by

$$\hat{\boldsymbol{V}}_{\hat{\theta}}=\hat{\boldsymbol{H}}^{-1}\left(\hat{\boldsymbol{\theta}}\right) \text{ and } \hat{\boldsymbol{V}}_{\hat{\theta}}=\hat{\boldsymbol{H}}^{-1}\left(\hat{\boldsymbol{\theta}}\right)\hat{\boldsymbol{J}}\left(\hat{\boldsymbol{\theta}}\right)\hat{\boldsymbol{H}}^{-1}\left(\hat{\boldsymbol{\theta}}\right).$$

Appendix D Data sources

Monthly data for U.S. house prices over the period January 1975 to December 2014 are obtained from the Freddie Mac House Price Index (FMHPI). These data are available at: http://www.freddiemac.com/research/.

Annual data on nominal income per capita and population at MSA level are acquired from the Bureau of Economic Analysis website for the same period. These data are available at: https://www.bea.gov/data/.

Annual State level Consumer Price Index data are obtained from the Bureau of Labour Statistics: https://www.bls.gov/cpi/. These are matched to the corresponding MSAs. In some cases where area data are missing then the U.S. average CPI is used instead.

Appendix E Additional Monte Carlo results

The Monte Carlo results provided in the tables and plots below are based on the designs set out in Section 5 of the paper.

Table S1: Bias, RMSE, size and power for parameters of individual units in the HSAR(1) model with one exogenous regressor and non-Gaussian errors for $N = 5$ and $T \in \{25, 50, 100, 200\}$.

T	25		50		100		200	
Parameter	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ψ_{i0}								
$\psi_{1,0} = 0.1261$	-0.0063	0.1537	0.0001	0.1012	-0.0023	0.0707	0.0007	0.0490
$\psi_{2,0} = 0.3883$	-0.0035	0.2078	-0.0049	0.1392	-0.0006	0.0955	-0.0002	0.0666
$\psi_{3,0} = 0.4375$	-0.0096	0.1852	-0.0016	0.1155	0.0020	0.0807	0.0000	0.0578
$\psi_{4,0} = 0.5059$	0.0022	0.1478	-0.0039	0.1018	-0.0020	0.0686	-0.0008	0.0481
$\psi_{5,0} = 0.7246$	-0.0040	0.1747	-0.0016	0.1248	-0.0008	0.0880	0.0001	0.0597
β_{i0}								
$\beta_{1,0} = 0.9649$	0.0097	0.1663	0.0047	0.1102	0.0020	0.0758	-0.0015	0.0540
$\beta_{2,0} = 0.9572$	0.0065	0.1968	0.0048	0.1349	-0.0017	0.0938	-0.0018	0.0657
$\beta_{3,0} = 0.2785$	0.0054	0.2088	-0.0012	0.1316	-0.0017	0.0918	0.0013	0.0666
$\beta_{4,0} = 0.9134$	-0.0011	0.1643	0.0054	0.1098	0.0007	0.0749	0.0000	0.0515
$\beta_{5,0} = 0.8147$	0.0039	0.2079	0.0069	0.1440	0.0056	0.0989	0.0003	0.0674

T	25	50	100	200	25	50	100	200
Parameter	Size				Power			
ψ_{i0}								
$\psi_{1,0} = 0.1261$	0.1025	0.0665	0.0545	0.0505	0.4350	0.6010	0.8260	0.9725
$\psi_{2,0} = 0.3883$	0.1000	0.0690	0.0575	0.0555	0.3015	0.4520	0.6205	0.8535
$\psi_{3,0} = 0.4375$	0.0990	0.0630	0.0570	0.0520	0.3395	0.4910	0.7365	0.9400
$\psi_{4,0} = 0.5059$	0.0795	0.0670	0.0615	0.0460	0.4265	0.6165	0.8475	0.9780
$\psi_{5,0} = 0.7246$	0.0770	0.0720	0.0700	0.0560	0.3270	0.4680	0.6765	0.9125
β_{i0}								
$\beta_{1,0} = 0.9649$	0.0920	0.0645	0.0535	0.0610	0.3985	0.5485	0.7610	0.9500
$\beta_{2,0} = 0.9572$	0.0965	0.0725	0.0595	0.0545	0.3100	0.4265	0.6445	0.8690
$\beta_{3,0} = 0.2785$	0.0965	0.0695	0.0500	0.0625	0.3370	0.4380	0.6435	0.8525
$\beta_{4,0} = 0.9134$	0.0890	0.0685	0.0550	0.0470	0.4230	0.5305	0.7830	0.9545
$\beta_{5,0} = 0.8147$	0.0940	0.0730	0.0590	0.0525	0.2885	0.3880	0.5590	0.8350

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 $i = 1, 2, \ldots, N$. Non-Gaussian errors are generated as $\varepsilon_{i0}/\sigma_{i0} \sim IID[\chi^2(2) - 2]/2$, with $\sigma^2_{i0} \sim IIDU[\chi^2(2)/8 + 0.25]$ for $i = 1, 2, \ldots, N$. The spatial weight matrix $\boldsymbol{W} = (w_{ij})$ 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 $i = 1, 2, \ldots, N$. 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 $i = 1, 2, \ldots, N$. Empirical size and empirical power are based on the sandwich formula given by (47). The nominal size is set to 5%. Size is computed under $H_{i0}\colon \psi_i = \psi_{i0}$, using a two-sided alternative, for $i = 1, 2, \ldots, N$. Power is computed under $\psi_i = \psi_{i0} + 0.2$, for $i = 1, 2, \ldots, N$. The number of replications is set to $R = 2,000$. Estimates are sorted in ascending order according to the true values of the spatial autoregressive parameters. Biases, RMSEs, sizes and powers for β_i, $i = 1, 2, \ldots, N$, are computed similarly, with power computed under $\beta_i = \beta_{i0} + 0.2$.

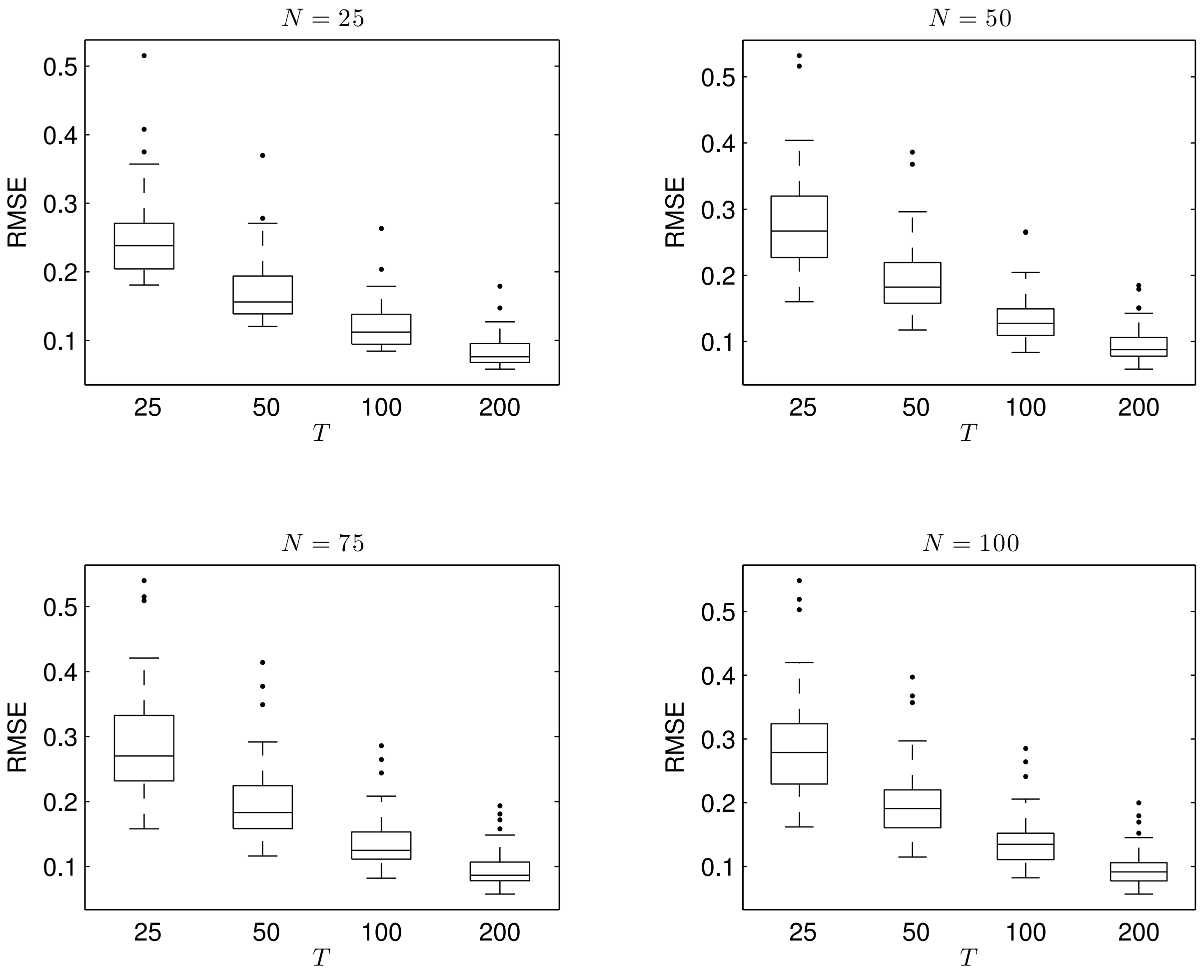
Table S2: Bias, RMSE, size and power for parameters of individual units in the HSAR(1) model with one exogenous regressor and non-Gaussian errors for $N = 100$ and $T \in \{25, 50, 100, 200\}$.

T	25		50		100		200	
Parameter	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
ψ_{i0}								
$\psi_{1,0} = 0.0244$	-0.0021	0.2572	-0.0036	0.1741	0.0009	0.1160	0.0000	0.0828
$\psi_{2,0} = 0.0255$	-0.0356	0.4532	0.0022	0.3105	-0.0114	0.2191	-0.0022	0.1499
$\psi_{3,0} = 0.0397$	0.0078	0.3029	-0.0032	0.2099	-0.0006	0.1438	0.0011	0.1025
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{51,0} = 0.3927$	-0.0022	0.2698	0.0020	0.1914	-0.0003	0.1305	0.0001	0.0896
$\psi_{52,0} = 0.3987$	0.0002	0.1694	-0.0039	0.1166	0.0021	0.0777	0.0004	0.0545
$\psi_{53,0} = 0.4004$	-0.0097	0.2691	0.0054	0.1770	0.0025	0.1205	-0.0017	0.0855
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{98,0} = 0.7695$	-0.0012	0.1433	0.0019	0.1013	0.0046	0.0724	-0.0001	0.0510
$\psi_{99,0} = 0.7705$	-0.0397	0.2546	-0.0088	0.1692	0.0020	0.1248	0.0007	0.0920
$\psi_{100,0} = 0.7904$	-0.0084	0.1514	-0.0070	0.1113	0.0008	0.0771	0.0006	0.0521
β_{i0}								
$\beta_{1,0} = 0.1978$	0.0066	0.2012	0.0010	0.1280	0.0006	0.0868	-0.0052	0.0598
$\beta_{2,0} = 0.7060$	0.0202	0.2711	0.0005	0.1720	-0.0001	0.1176	0.0033	0.0837
$\beta_{3,0} = 0.4173$	0.0077	0.1852	0.0027	0.1254	-0.0002	0.0842	0.0019	0.0597
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{51,0} = 0.9448$	0.0043	0.1415	-0.0015	0.0962	0.0015	0.0665	0.0007	0.0471
$\beta_{52,0} = 0.1190$	0.0030	0.1324	0.0023	0.0913	-0.0004	0.0619	0.0014	0.0421
$\beta_{53,0} = 0.7127$	0.0019	0.1941	-0.0036	0.1226	0.0025	0.0893	0.0010	0.0615
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{98,0} = 0.1067$	0.0024	0.1221	-0.0024	0.0807	0.0005	0.0553	-0.0002	0.0399
$\beta_{99,0} = 0.4588$	0.0147	0.1909	0.0026	0.1300	0.0017	0.0899	0.0017	0.0650
$\beta_{100,0} = 0.3674$	0.0035	0.1239	0.0022	0.0865	0.0006	0.0607	0.0002	0.0408

T	25	50	100	200	25	50	100	200
Parameter	Size				Power			
ψ_{i0}								
$\psi_{1,0} = 0.0244$	0.0915	0.0805	0.0560	0.0600	0.2255	0.3025	0.4560	0.7070
$\psi_{2,0} = 0.0255$	0.0830	0.0645	0.0580	0.0525	0.1225	0.1170	0.1940	0.2890
$\psi_{3,0} = 0.0397$	0.0905	0.0785	0.0605	0.0630	0.1815	0.2390	0.3450	0.5405
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{51,0} = 0.3927$	0.0995	0.0640	0.0595	0.0530	0.1975	0.2915	0.4020	0.6325
$\psi_{52,0} = 0.3987$	0.0865	0.0660	0.0620	0.0525	0.3785	0.5380	0.7440	0.9395
$\psi_{53,0} = 0.4004$	0.0960	0.0810	0.0520	0.0590	0.2400	0.3020	0.4540	0.6760
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\psi_{98,0} = 0.7695$	0.0710	0.0665	0.0675	0.0660	0.4015	0.5760	0.7930	0.9650
$\psi_{99,0} = 0.7705$	0.0390	0.0320	0.0405	0.0535	0.1750	0.2380	0.3820	0.6095
$\psi_{100,0} = 0.7904$	0.0690	0.0655	0.0575	0.0510	0.3845	0.5705	0.7605	0.9500
β_{i0}								
$\beta_{1,0} = 0.1978$	0.1085	0.0710	0.0570	0.0485	0.3320	0.4715	0.6715	0.9195
$\beta_{2,0} = 0.7060$	0.1055	0.0580	0.0495	0.0570	0.2315	0.3110	0.4530	0.6680
$\beta_{3,0} = 0.4173$	0.0935	0.0805	0.0520	0.0590	0.3585	0.4815	0.6985	0.9120
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{51,0} = 0.9448$	0.0940	0.0705	0.0535	0.0545	0.4710	0.6580	0.8445	0.9745
$\beta_{52,0} = 0.1190$	0.0950	0.0845	0.0525	0.0470	0.5250	0.6900	0.8940	0.9920
$\beta_{53,0} = 0.7127$	0.1055	0.0690	0.0685	0.0545	0.3615	0.4800	0.6485	0.8910
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\beta_{98,0} = 0.1067$	0.0945	0.0685	0.0520	0.0570	0.5775	0.7650	0.9320	0.9960
$\beta_{99,0} = 0.4588$	0.0745	0.0625	0.0505	0.0580	0.2845	0.4355	0.6460	0.8600
$\beta_{100,0} = 0.3674$	0.1000	0.0710	0.0680	0.0540	0.5480	0.7205	0.9070	0.9935

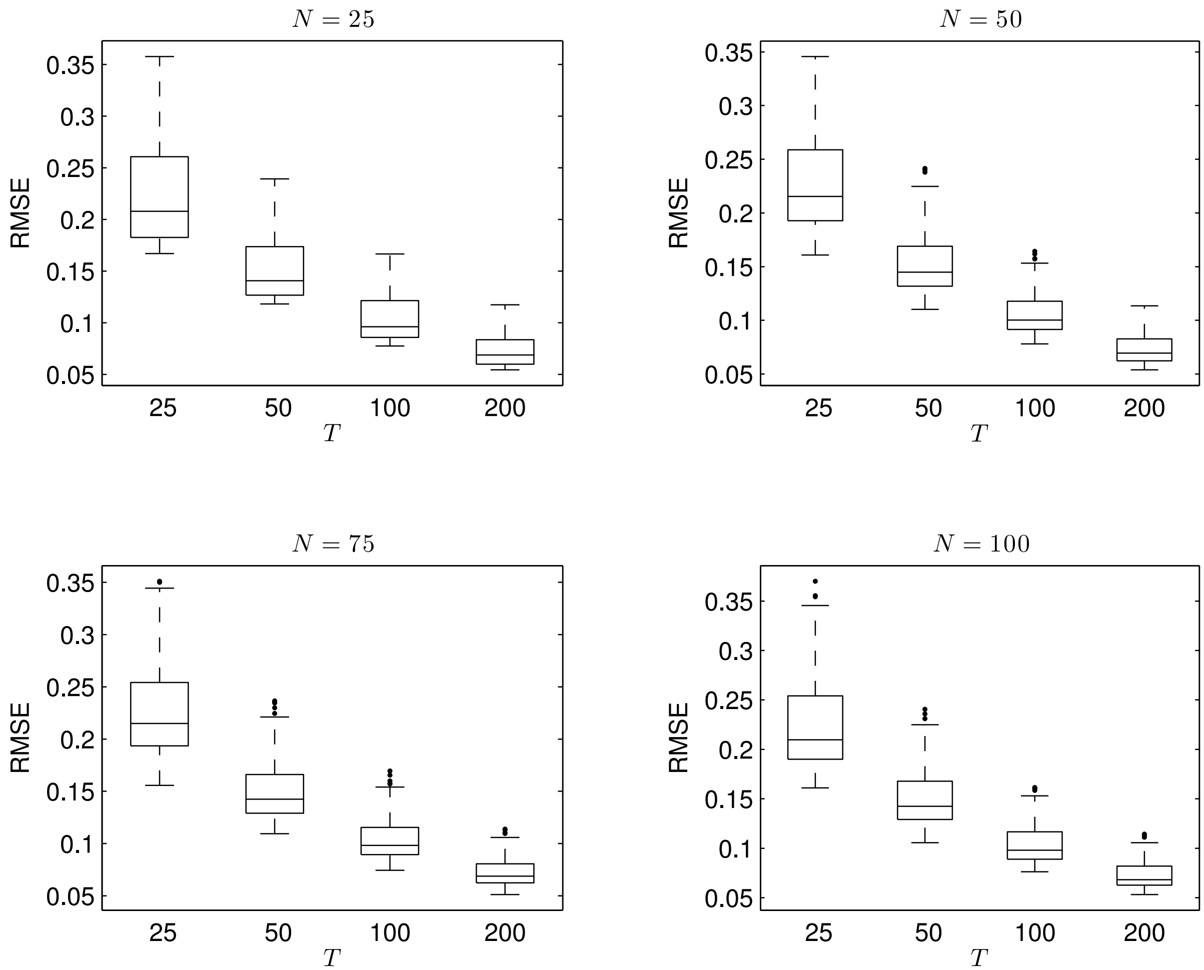
Notes: See notes to Table S1.

Figure A1: Boxplots of RMSEs for the individual autoregressive spatial parameter estimates from the HSAR(1) model with non-Gaussian errors, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections for different N and T combinations



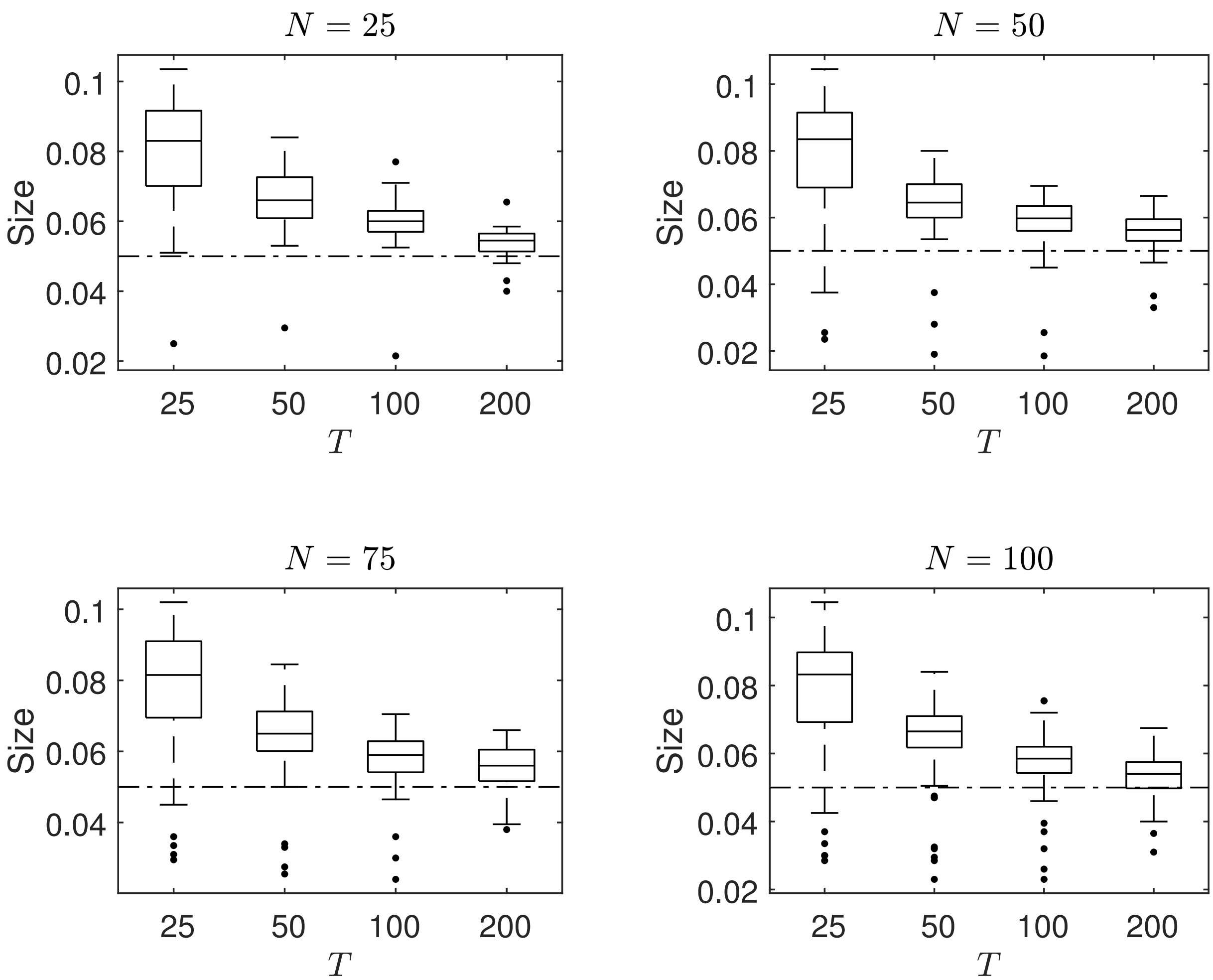
Notes: True parameter values are generated as $\psi_{i0} \sim IIDU(0, 0.8)$, $a_{i0} \sim IIDN(1, 1)$ and $\beta_{i0} \sim IIDU(0, 1)$, for $i = 1, 2, \ldots, N$. Non-Gaussian errors are generated as $\varepsilon_{it}/\sigma_{i0} \sim IID[\chi^2(2) - 2]/2$, with $\sigma_{i0}^2 \sim IID[\chi^2(2)/4 + 0.5]$, for $i = 1, 2, \ldots, N$. Exogenous regressors are spatially correlated across i and generated by (65), with $\phi_i = 0.5$. The spatial weight matrix $\boldsymbol{W} = (w_{ij})$ 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 $i = 1, 2, \ldots, N$. RMSEs are computed as $\sqrt{R^{-1}\sum_{r=1}^{R}(\hat{\psi}_{i,r} - \psi_{i0})^2}$ for $i = 1, 2, \cdots, N$. The number of replications is set to $R = 2,000$.

Figure A2: Boxplots of RMSEs for the individual slope parameter estimates from the HSAR(1) model with non-Gaussian errors, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections for different N and T combinations



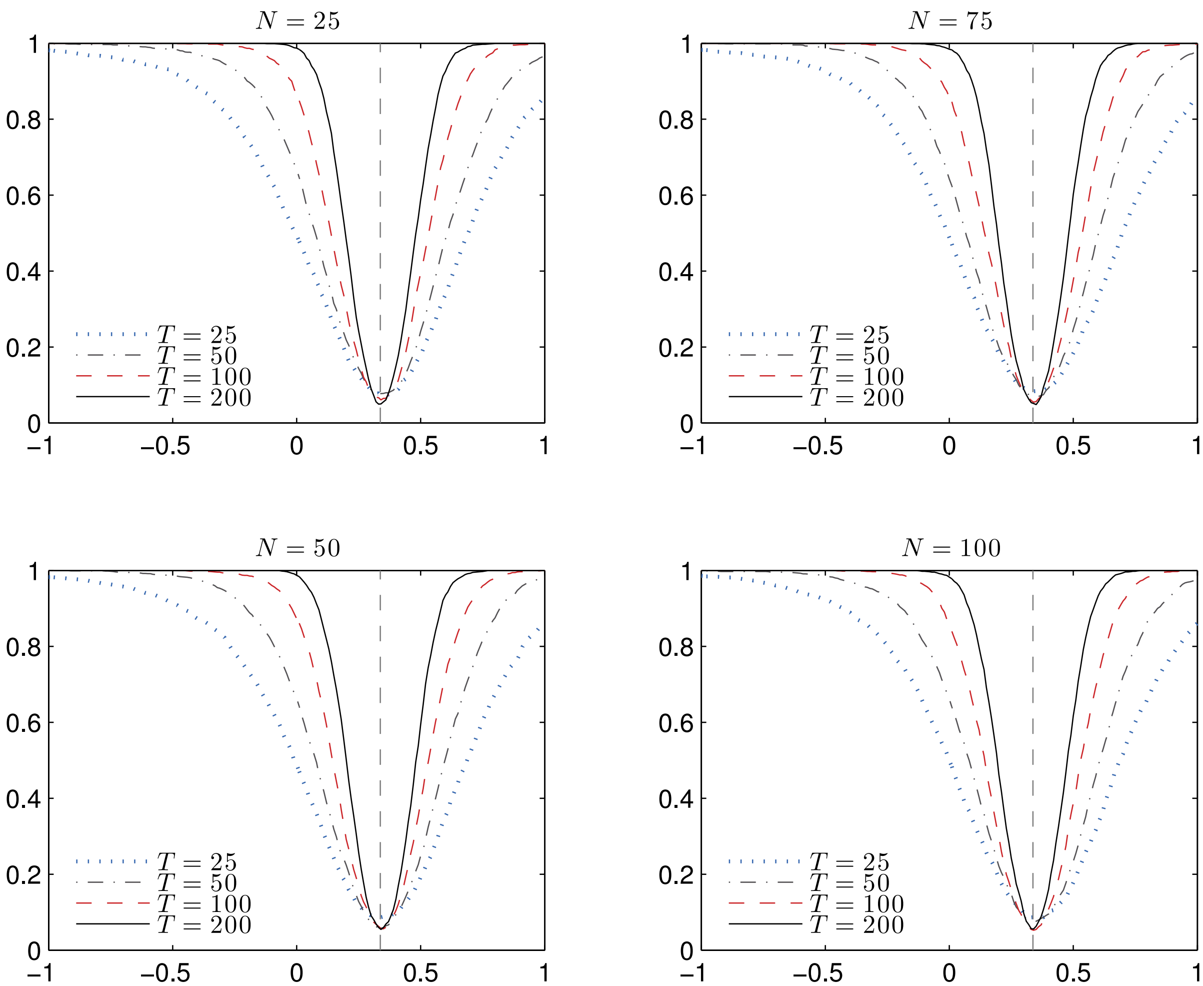
Notes: RMSEs are computed as $\sqrt{R^{-1}\sum_{r=1}^{R}(\hat{\beta}_{i,r} - \beta_{i0})^2}$ for $i = 1, 2, \cdots, N$. See the notes to Figure A1 for details of the data generating process. The number of replications is set to $R = 2,000$.

Figure A3: Boxplots of empirical sizes of tests for individual spatial parameters from HSAR(1) model with non-Gaussian errors, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections for different N and T combinations, using the sandwich formula for the variance



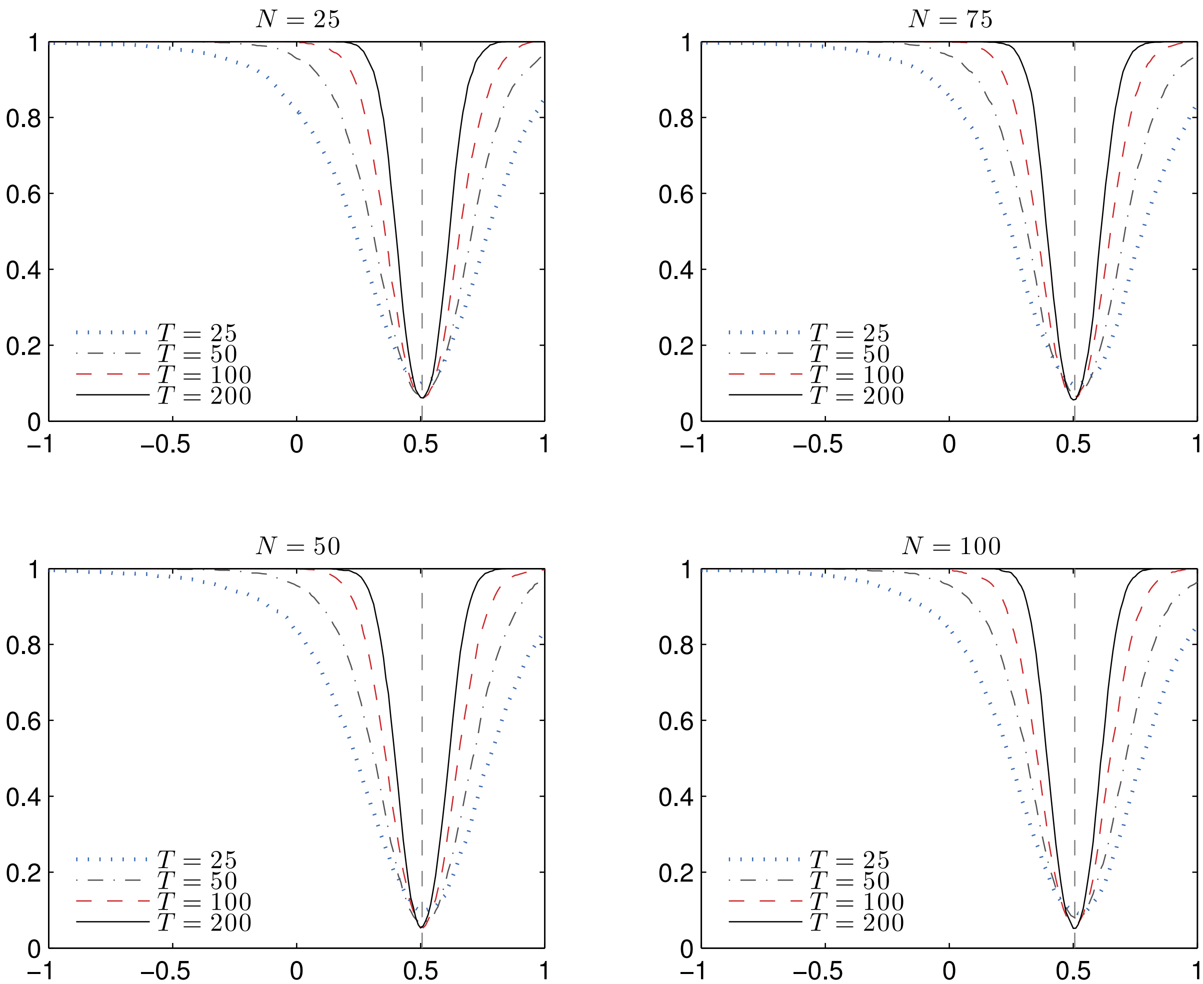
Notes: Nominal size is set to 5%. The sandwich formula is given by (50). See the notes to Figure A1 for details of the data generating process. Size is computed under H_0: $\psi_i = \psi_{i0}$, using a two-sided alternative where ψ_{i0} takes values in the range [0.0, 0.8] for $i = 1, 2, \ldots, N$. The number of replications is set to $R = 2,000$.

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, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections, using the sandwich formula for the variance



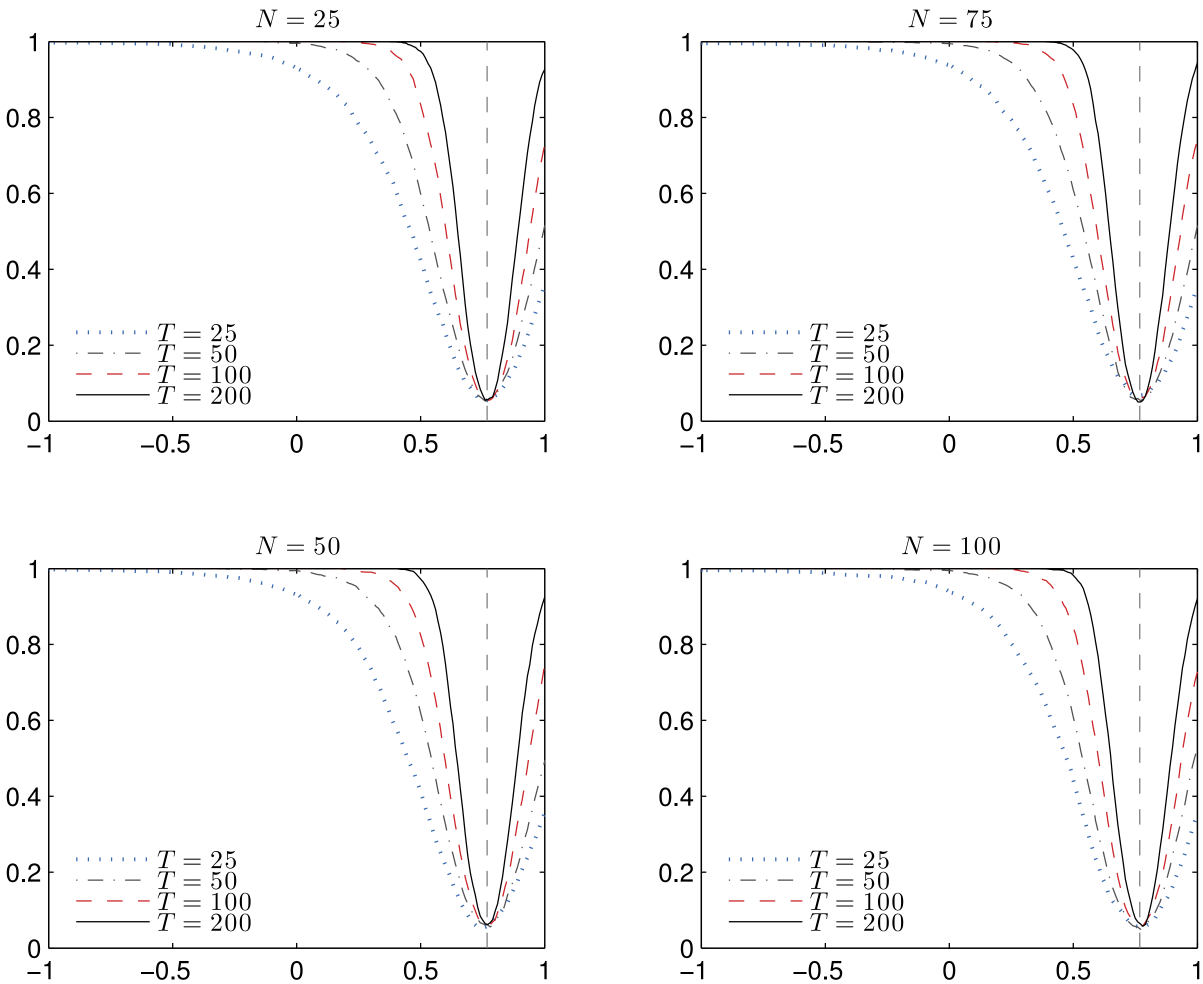
Notes: The power functions are based on the sandwich formula given by (50). See the notes to Figure A1 for details of the data generating process. Power is computed under $\psi_i = \psi_{i0} + \delta$, where $\delta = -0.8, -0.791, \ldots, 0.791, 0.8$ or until the parameter space boundaries of -1 and 1 are reached. The number of replications is set to $R = 2,000$.

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, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections, using the sandwich formula for the variance



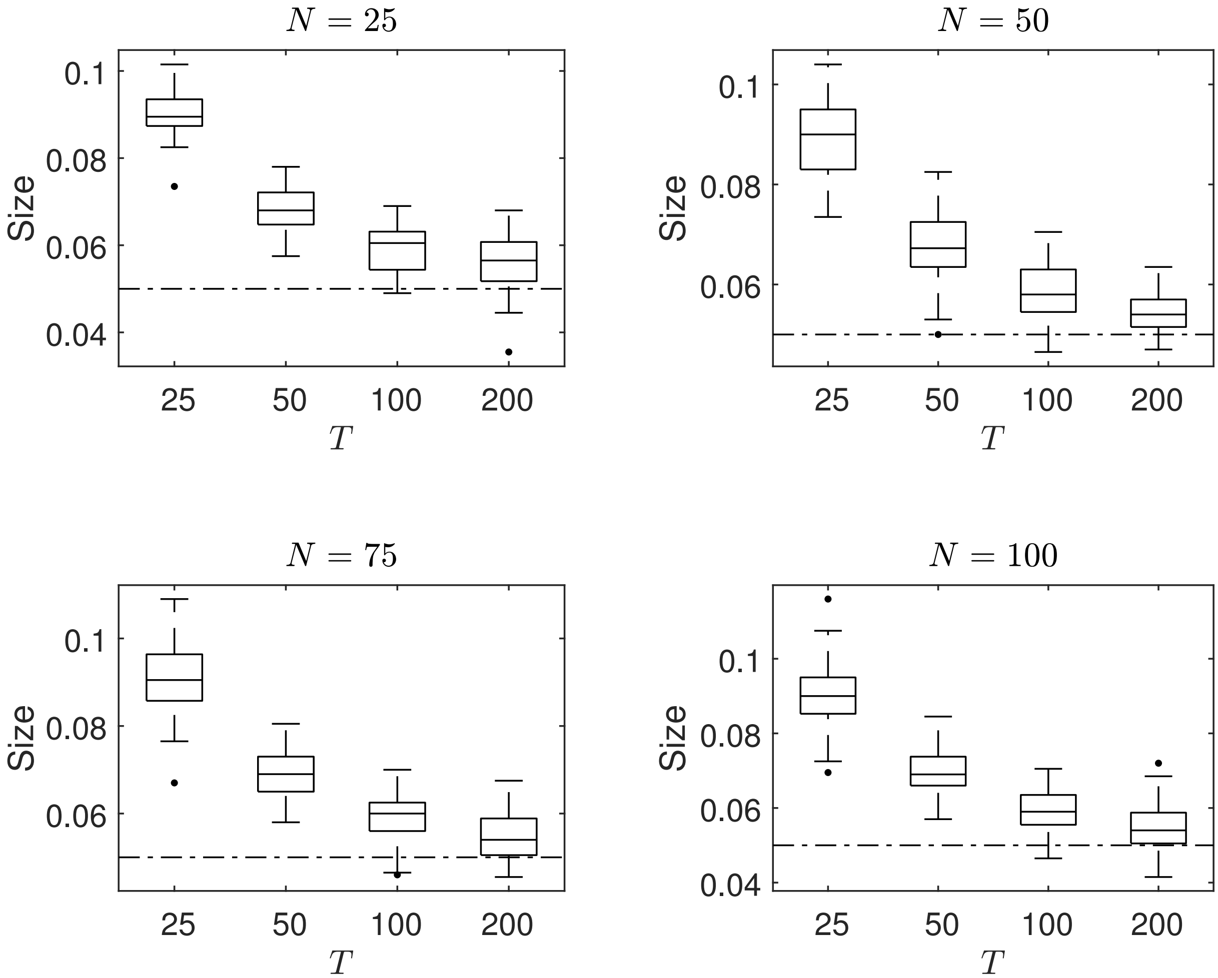
Notes: See the notes to Figure A4.

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, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections, using the sandwich formula for the variance



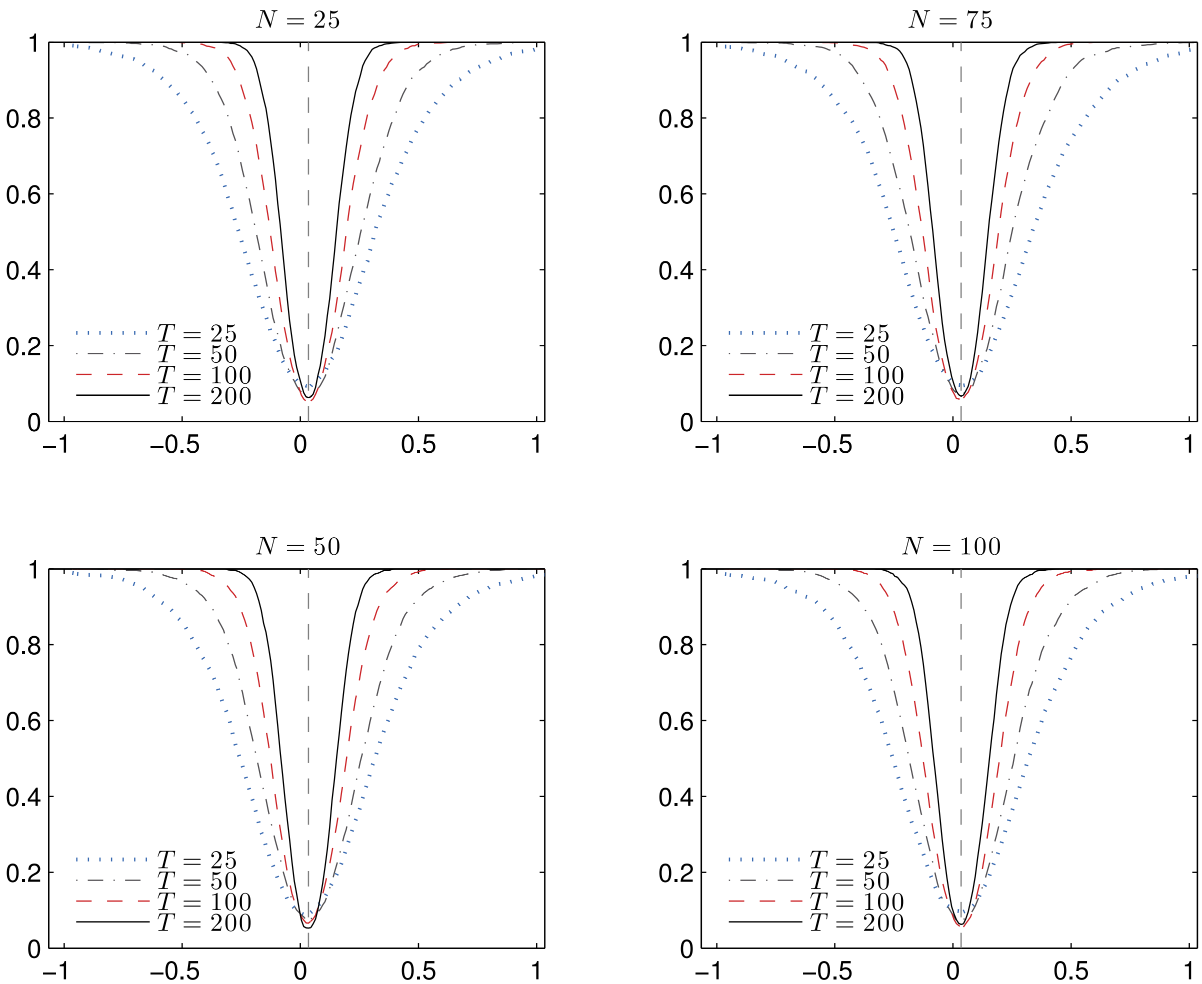
Notes: See the notes to Figure A4.

Figure A7: Boxplots of empirical sizes of tests for individual slope parameters from HSAR(1) model with non-Gaussian errors, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections for different N and T combinations, using the sandwich formula for the variance



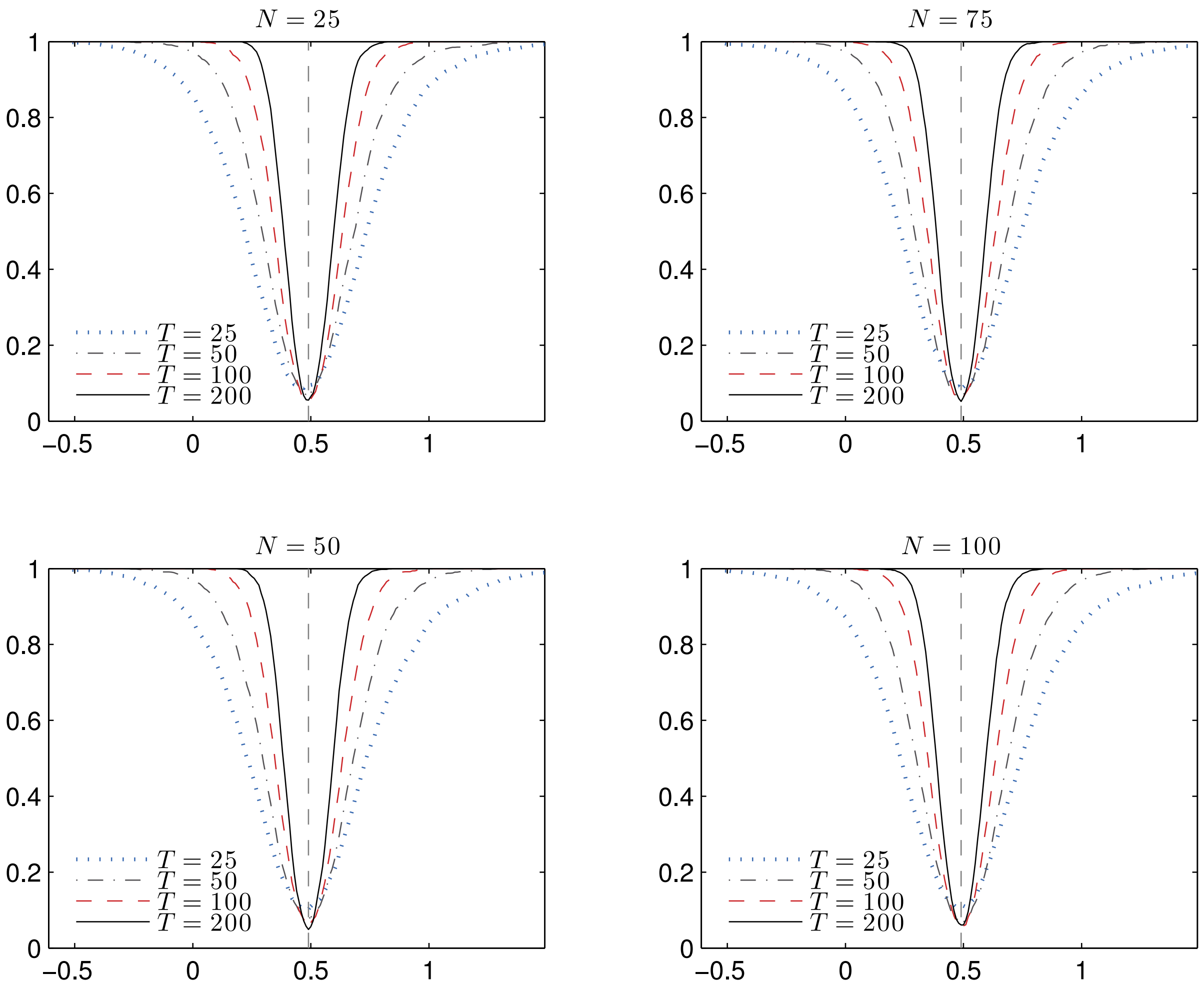
Notes: Nominal size is set to 5%. The sandwich formula is given by (50). See the notes to Figure A1 for details of the data generating process. Size is computed under H_0: $\beta_i=\beta_{i0}$, using a two-sided alternative where β_{i0} takes values in the range [0.0, 1.0] for $i = 1, 2, \ldots, N$. The number of replications is set to $R = 2,000$.

Figure A8: 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 non-Gaussian errors, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections, using the sandwich formula for the variance



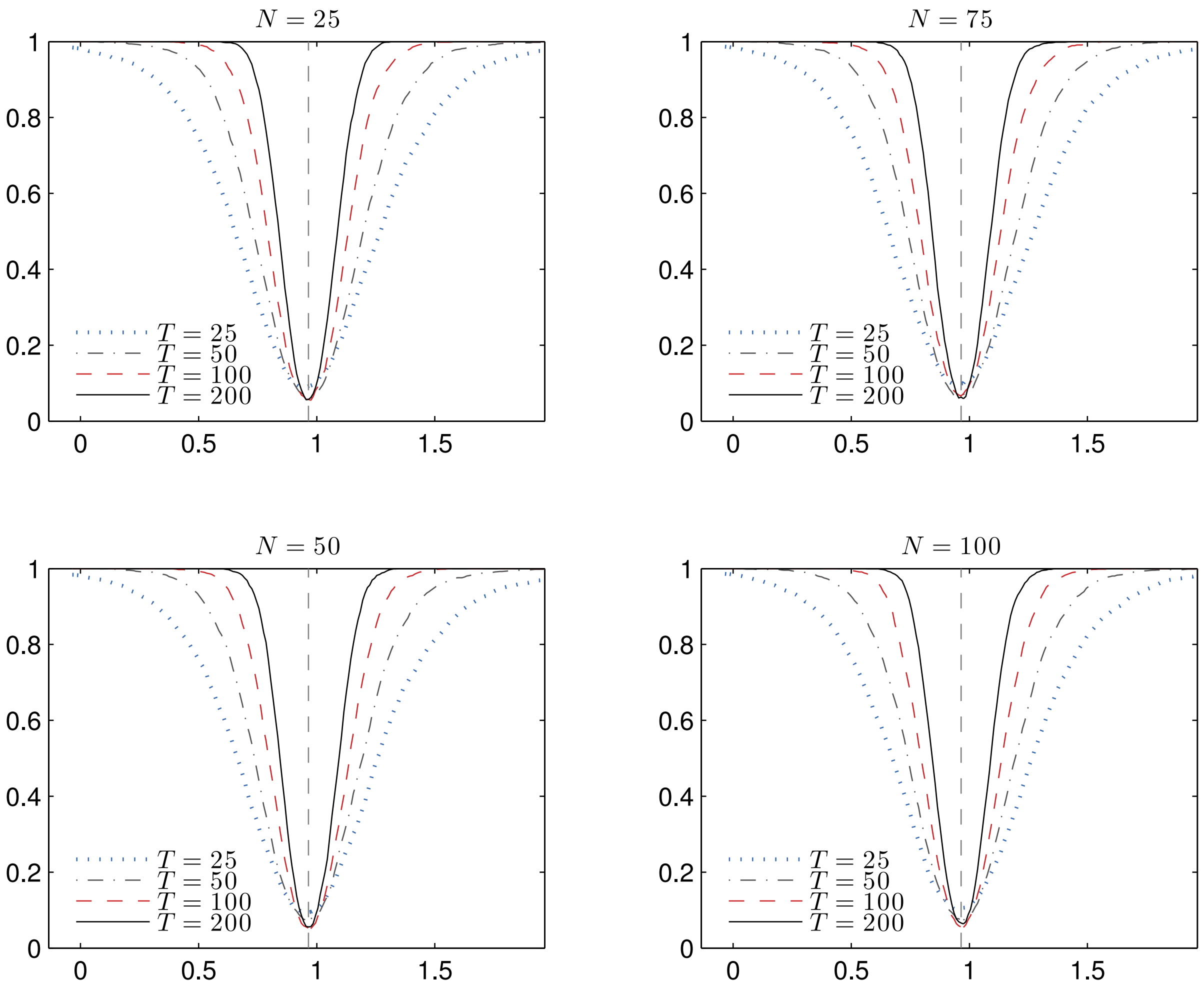
Notes: The power functions are based on the sandwich formula given by (50). See the notes to Figure A1 for details of the data generating process. Power is computed under $\beta_i{=}\beta_{i0} + \delta$, where $\delta = -1.0, -0.991, \ldots, 0.991, 1.0$. The number of replications is set to $R = 2,000$.

Figure A9: 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 non-Gaussian errors, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure A8.

Figure A10: 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 non-Gaussian errors, one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections, using the sandwich formula for the variance



Notes: See the notes to Figure A8.

Table B: Bias, RMSE and size for the Mean Group (MG) estimator from the HSAR(1) model with one exogenous regressor and spatial weight matrix $\boldsymbol{W}$ having 4 connections for different N and T combinations

N	25				50				75				100			
T	25	50	100	200	25	50	100	200	25	50	100	200	25	50	100	200
Gaussian errors																
Bias																
$\hat{\hat{\psi}}_{MG}$	-0.0060	-0.0005	-0.0013	-0.0007	-0.0053	-0.0013	-0.0008	0.0009	-0.0075	-0.0014	0.0006	-0.0005	-0.0070	-0.0012	-0.0007	-0.0001
$\hat{\hat{\beta}}_{MG}$	0.0061	0.0011	0.0023	-0.0008	0.0052	0.0005	0.0028	0.0012	0.0065	0.0024	0.0010	0.0014	0.0055	0.0033	0.0020	-0.0001
RMSE																
$\hat{\hat{\psi}}_{MG}$	0.0599	0.0542	0.0508	0.0482	0.0446	0.0381	0.0357	0.0339	0.0358	0.0311	0.0286	0.0276	0.0310	0.0270	0.0251	0.0237
$\hat{\hat{\beta}}_{MG}$	0.0752	0.0662	0.0602	0.0595	0.0522	0.0477	0.0433	0.0422	0.0435	0.0378	0.0360	0.0340	0.0371	0.0324	0.0311	0.0298
Size																
$\hat{\hat{\psi}}_{MG}$	0.0255	0.0415	0.0435	0.0585	0.0270	0.0335	0.0430	0.0475	0.0205	0.0285	0.0370	0.0445	0.0200	0.0325	0.0335	0.0350
$\hat{\hat{\beta}}_{MG}$	0.0595	0.0635	0.0530	0.0630	0.0510	0.0625	0.0625	0.0545	0.0560	0.0515	0.0525	0.0485	0.0485	0.0475	0.0580	0.0520
non-Gaussian errors																
Bias																
$\hat{\hat{\psi}}_{MG}$	-0.0102	-0.0018	-0.0010	-0.0003	-0.0088	-0.0018	-0.0007	0.0008	-0.0097	-0.0017	0.0001	-0.0007	-0.0079	-0.0010	-0.0010	-0.0005
$\hat{\hat{\beta}}_{MG}$	0.0087	0.0016	0.0025	-0.0011	0.0048	0.0013	0.0026	0.0010	0.0067	0.0023	0.0016	0.0011	0.0047	0.0028	0.0014	0.0001
RMSE																
$\hat{\hat{\psi}}_{MG}$	0.0638	0.0558	0.0508	0.0480	0.0451	0.0389	0.0361	0.0336	0.0369	0.0313	0.0285	0.0271	0.0319	0.0275	0.0250	0.0237
$\hat{\hat{\beta}}_{MG}$	0.0755	0.0648	0.0610	0.0593	0.0528	0.0461	0.0440	0.0420	0.0436	0.0372	0.0362	0.0338	0.0380	0.0325	0.0307	0.0300
Size																
$\hat{\hat{\psi}}_{MG}$	0.0330	0.0410	0.0520	0.0535	0.0280	0.0330	0.0425	0.0480	0.0260	0.0350	0.0360	0.0390	0.0225	0.0320	0.0350	0.0375
$\hat{\hat{\beta}}_{MG}$	0.0625	0.0620	0.0600	0.0600	0.0540	0.0505	0.0625	0.0560	0.0555	0.0515	0.0580	0.0510	0.0515	0.0495	0.0525	0.0540

Notes: True parameter values are generated as $a_{i0} = a_0 + \epsilon_{1i}$, with $a_0 = 1$ and $\epsilon_{1i} \sim IIDN(0,1)$, $\psi_{i0} = \psi_0 + \epsilon_{2i}$, with $\psi_0 = 0.4$ and $\epsilon_{2i} \sim IIDU(-0.4, 0.4)$ and $\beta_{i0} = \beta_0 + \epsilon_{3i}$, with $\beta_0 = 0.5$ and $\epsilon_{3i} \sim IIDU(-0.5, 0.5)$, where ϵ_{1i}, ϵ_{2i}, and ϵ_{3i} are drawn for each replication. Gaussian errors are generated as $\varepsilon_{it}/\sigma_{i0} \sim N(0,1)$, while non-Gaussian errors are generated as $\varepsilon_{it}/\sigma_{i0} \sim IID[\chi^2(2) - 2]/2$, where $\sigma_{i0}^2 \sim IID[\chi^2(2)/4 + 0.5]$, for $i = 1, 2, \ldots, N$ in both instances. Exogenous regressors are spatially correlated across i and generated by (65), with $\phi_i = 0.5$. Let $\hat{\psi}_r$ denote the MG estimator of ψ_0 for replication r, $\hat{\psi}_r = N^{-1} \sum_{i=1}^{N} \hat{\psi}_{i,r}$. $r = 1, \ldots, R$, Bias and RMSE are computed as $R^{-1} \sum_{r=1}^{R} (\hat{\psi}_r - \psi_0)$ and $\sqrt{R^{-1} \sum_{r=1}^{R} (\hat{\psi}_r - \psi_0)^2}$; bias and RMSE for $\hat{\beta}_r$ are computed accordingly. The spatial weight matrix $\boldsymbol{W} = (w_{ij})$ 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 $i = 1, 2, \ldots, N$. The number of replications is set to $R = 2,000$.