Estimation and Inference in Spatial Models with Dominant Units*

M. Hashem Pesaran

University of Southern California, USA, and Trinity College, Cambridge, UK

Cynthia Fan Yang

Florida State University

March 10, 2019

Abstract

Estimation and inference in the spatial econometrics literature are carried out assuming that the matrix of spatial or network connections has uniformly bounded absolute column sums in the number of cross-section units, n. In this paper, we consider spatial models where this restriction is relaxed. The linear-quadratic central limit theorem of Kelejian and Prucha (2001) is generalized and then used to establish the asymptotic properties of the GMM estimator due to Lee (2007) in the presence of dominant units. A new Bias-Corrected Method of Moments estimator is also proposed that avoids the problem of weak instruments by self-instrumenting the spatially lagged dependent variable. Both estimators are shown to be consistent and asymptotically normal, depending on the rate at which the maximum column sum of the weights matrix rises with n. The small sample properties of the estimators are investigated by Monte Carlo experiments and shown to be satisfactory. An empirical application to sectoral price changes in the US over the pre- and post-2008 financial crisis is also provided. It is shown that the share of capital can be estimated reasonably well from the degree of sectoral interdependence using the input-output tables, despite the evidence of dominant sectors being present in the US economy.

Keywords: spatial autoregressive models, central limit theorems for linear-quadratic forms, dominant units, GMM, bias-corrected method of moments (BMM), US input-output analysis, capital share.

JEL Classifications: C13, C21, C23, R15

---

*The authors would like to acknowledge helpful comments from Natalia Bailey, Alex Chudik, Simon Reese and Ron Smith. Part of this paper was written when Yang was a doctoral student at the University of Southern California. Corresponding author at: Department of Economics, Florida State University, 281 Bellamy Building, Tallahassee, FL 32306, USA. E-mail addresses: pesaran@usc.edu (Pesaran), cynthia.yang@fsu.edu (Yang).

1 Introduction

In spatial econometrics, the interdependence among cross-sectional units is captured via a spatial weights matrix, $\mathbf{W} = (w_{ij})$, which is usually constructed based on some measures of geographical, economic or social distance. A critical assumption that has been adopted in the existing literature is that the maximum absolute row and column sum norms of $\mathbf{W}$ are uniformly bounded in the number of cross section units, n. This assumption, which dates back to the seminal contributions of Kelejian and Prucha (1998, 1999), essentially imposes a strong restriction on the degree of cross-sectional dependence amongst the units in the spatial model or network. For example, the assumption will be satisfied if $\mathbf{W}$ is sparse in the sense that each unit has only a finite number of "neighbors", or if the strength of their connections decays sufficiently fast with their distance from one another. However, such sparsity conditions rule out the possibility that some units could be dominant or influential, in the sense that they might impact a large number of other units in the network. This could arise, for example, in the case of production or financial networks where a large number of firms or households could depend on one or more banks or sectors in the economy, as documented in the recent contributions by Acemoglu, Carvalho, Ozdaglar, and Tahbaz-Salehi (2012), Dungey and Volkov (2018) and Pesaran and Yang (2019). In such cases the standard proofs used to justify the consistency and asymptotic normality of the proposed estimators are no longer applicable.

In this paper we consider estimation and inference in spatial autoregressive (SAR) models where the maximum column sum norm of the weights matrix, denoted by $\|\mathbf{W}\|_1$, is allowed to rise with the dimension of the network, n. Specifically, we consider situations where $\|\mathbf{W}\|_1 = O\left(n^{\delta}\right)$, with $\delta \in [0, 1]$. The exponent δ measures the degree to which the most influential unit in the network impacts all other units. The condition imposed on $\mathbf{W}$ in the literature corresponds to assuming $\delta = 0$. But, as noted above, in many applications it is likely that $\delta > 0$, and it is therefore desirable to provide conditions under which standard estimators of SAR models continue to apply in such cases.

The exponent δ also relates to measures of network centrality. In the case of spatial models with row normalized weights matrices, the degree of centrality of unit j is typically measured by its (weighted) outdegree, defined by $d_j = \sum_{i=1}^{n} w_{ij}$. The degree of dominance of unit j can now be measured by the exponent δ_j, defined by $d_j = O\left(n^{\delta_j}\right)$, where $\delta_j \in [0, 1]$. Unit j is said to be strongly dominant if $\delta_j = 1$, weakly dominant if $\delta_j > 0$, and non-dominant if $\delta_j = 0$.[1] To simplify the exposition we refer to unit j as being dominant if $\delta_j > 0$, unless it is important to distinguish between cases of strong and weak dominance. Accordingly, the overall degree of network centrality is also given by $\delta = \max(\delta_1, \delta_2, \ldots, \delta_n)$.[2] From this

[1] For further details see Definition 1 in Pesaran and Yang (2019).
[2] Note that $\|\mathbf{W}\|_1 = \sup_j(d_j)$.

perspective, the assumption that $\mathbf{W}$ has bounded column sum norm requires that $\delta_j = 0$, for all j. The present paper relaxes this assumption and develops new estimation and inference theory allowing for the existence of dominant units ($\delta > 0$) in the network.[3]

We begin by generalizing the central limit theorem for linear-quadratic forms due to Kelejian and Prucha (2001), which requires $\delta = 0$. For our analysis we need to relax this restriction and allow the matrix in the quadratic form of their theorem to have column sums that are unbounded in n (namely allow for $\delta > 0$). The generalized central limit theorem is then used to establish the asymptotic properties of the estimators of the SAR model.

There are two main approaches to the estimation of spatial models, namely the maximum likelihood (ML) method developed by Cliff and Ord (1973, 1981), Upton and Fingleton (1985), and developed further by Anselin (1988), Lee (2004), and Lee and Yu (2010), amongst others. The second approach is the generalized method of moments (GMM) pioneered by Kelejian and Prucha (1998, 1999), and extended and further studied by Lee (2007), Kapoor et al. (2007), Lin and Lee (2010), and Lee and Yu (2014), amongst others. In this paper we consider the GMM approach developed by Lee (2007), which is generally applicable even if the SAR model does not contain any regressors, and establish conditions under which Lee's GMM estimator is consistent and asymptotically normal even if $\delta > 0$.

We also propose a new bias-corrected method of moments (BMM), which is also applicable generally and is simple to implement. The BMM approach was first introduced in a recent paper by Chudik and Pesaran (2017) for the estimation of dynamic panel data models with short time-dimension. In the context of the SAR model, the spatial lag variable is endogenous. Instead of looking for valid instruments, the BMM approach uses the spatial lag variable as an "instrument" for itself, but corrects the bias due to the non-zero correlation between the spatial lag variable and the error term. This method has the advantage of avoiding the weak instrument problem by construction. We show that both GMM and BMM estimators are consistent if $0 \leq \delta < 1$, and establish their asymptotic normality for values of δ in the range $0 \leq \delta < 1/2$.

An extensive set of Monte Carlo experiments lend support to the theoretical results and document that both estimators have satisfactory small sample properties, with the BMM estimator outperforming the GMM estimator when n is relatively small and δ is close to unity. The estimation techniques are shown to be robust to different degrees of spatial dependence, various specifications of the spatial weights matrix, and non-Gaussian errors.

[3] It is worth noting that in the current paper we assume $\mathbf{W}$ is known and focus on estimating the spatial parameters. In cases where information on direct connections of the network is unavailable, there exists a related literature that uses large panel data sets (with both n and T large) to detect which unit has the largest δ (when δ equals or is close to unity) from the pattern of correlation in the data without needing to know $\mathbf{W}$. See, for example, Bai and Ng (2006), Parker and Sul (2016), Brownlees and Mesters (2018), and Kapetanios et al. (2018). In a related literature, Bailey et al. (2016) also consider estimating δ using large panel data sets when $\mathbf{W}$ is not known.

As an empirical application we consider the sectoral price changes in the US over the pre- and post-2008 financial crisis, using 300×300 input-output tables as spatial weights. We show that the share of capital can be estimated from the degree of sectoral interdependence. We first investigate the presence of dominant sectors in the US economy by computing the extremum estimator of δ (the degree of network centrality) proposed in Pesaran and Yang (2019), and obtain estimates lying between 0.71 and 0.85, suggesting the existence of at least one dominant sector in the US economy. We then estimate a SAR model in the rate of sectoral price changes and provide estimates of the share of capital of around 0.4 during the pre-crisis period (1998–2006), and 0.3 over the post-crisis period (2007–2015). These estimates compare reasonably well with the share of capital calibrated in the literature.

The remainder of the paper is organized as follows: Section 2 describes the model and sets out its assumptions. Section 3 provides a generalization of Kelejian and Prucha's central limit theorem. The GMM and BMM estimation methods and their asymptotic properties are detailed in Sections 4 and 5, respectively. Section 6 presents the finite sample properties of the GMM and BMM estimators using Monte Carlo techniques. Section 7 contains an empirical application to the US sectoral prices, and Section 8 gives some concluding remarks. Proofs of theorems and propositions, together with statements and proofs of necessary lemmas, are provided in an online mathematical appendix. Additional empirical and Monte Carlo results are summarized in a supplement, which is available upon request.

Notations: Generic positive finite constants are denoted by K when they are large, and by ϵ when small. They can take different values at different instances. Let $\{f_n\}_{n=1}^{\infty}$ be a real sequence and $\{g_n\}_{n=1}^{\infty}$ be a real positive sequence. We write $f_n = O(g_n)$ if there exists a positive finite constant K_0 such that $|f_n|/g_n \leq K_0$ for all n; we write $f_n = o(g_n)$ if $f_n/g_n \to 0$ as $n \to \infty$. The symbols $\to_p$ and $\to_d$ indicate convergence in probability and in distribution as $n \to \infty$, respectively. Let $\{x_n\}$ be a sequence of random variables. We write $x_n = o_p(1)$ if $x_n \to_p 0$ as $n \to \infty$. $E_0(\cdot)$ denotes expectations taken under the true probability measure. For an $n \times n$ matrix $\mathbf{A} = (a_{ij})$, $\|\mathbf{A}\|_\infty = \sup_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|$ denotes the maximum absolute row sum norm (or row norm, for short) of $\mathbf{A}$; $\|\mathbf{A}\|_1 = \sup_{1\leq j\leq n} \sum_{i=1}^{n} |a_{ij}|$ denotes the maximum absolute column sum norm (or column norm, for short); and $\lambda_{\max}(\mathbf{A})$ denotes the largest eigenvalue of $\mathbf{A}$. The symbol $diag(\mathbf{A})$ represents a *vector* consisting of the diagonal elements of $\mathbf{A}$, namely, $diag(\mathbf{A}) = (a_{11}, a_{22}, \ldots, a_{nn})'$; whereas $Diag(\mathbf{A})$ represents a *diagonal matrix* formed by the diagonal entries of $\mathbf{A}$. $\mathbf{1}_n$ is an $n \times 1$ vector of ones, i.e., $\mathbf{1}_n = (1, 1, \ldots, 1)'$.

2 The model and its assumptions

We consider the following standard SAR model:

$$y_i = \rho y_i^* + \boldsymbol{\beta}'\mathbf{x}_i + \varepsilon_i, \text{ for } i = 1, 2, \ldots, n, \tag{1}$$

where y_i is the outcome variable on unit i, ρ is a fixed spatial coefficient, $\mathbf{x}_i$ is a $k \times 1$ vector of regressors on unit i with the associated vector of fixed coefficients $\boldsymbol{\beta}$, ε_i is a random error, y_i^* is the spatial variable, defined by

$$y_i^* = \sum_{j=1}^{n} w_{ij} y_j = \mathbf{w}_{i.}' \mathbf{y}, \tag{2}$$

$\mathbf{y} = (y_1, y_2, \dots, y_n)'$, $\mathbf{w}_{i.} = (w_{i1}, w_{i2}, \dots, w_{in})'$ is a vector of known constant weights and $w_{ij} \geq 0$ for all i and j. Let $\mathbf{y}^* = (y_1^*, y_2^*, \dots, y_n^*)'$. Then (2) implies that $\mathbf{y}^* = \mathbf{W}\mathbf{y}$, where $\mathbf{W} = (w_{ij}) = (\mathbf{w}_{1.}, \mathbf{w}_{2.}, \dots, \mathbf{w}_{n.})'$ is an $n \times n$ known matrix of spatial weights (or network connections).

We suppose $\mathbf{W}$ is row-standardized and assume that the j^{th} column sum of $\mathbf{W}$, $d_j = \sum_{i=1}^{n} w_{ij}$, is of order n^{δ_j} such that

$$d_j = \kappa_j n^{\delta_j}, \text{ for } j = 1, 2, \dots, n, \tag{3}$$

where δ_j is a fixed constant in the range $0 \leq \delta_j \leq 1$, and κ_j is a strictly positive random variable defined on $0 < \underline{\kappa} \leq \kappa_j \leq \bar{\kappa} < K$, where $\underline{\kappa}$ and $\bar{\kappa}$ are fixed constants. We also set

$$\delta = \max_{j=1,2,\dots,n} (\delta_j), \quad 0 \leq \delta \leq 1, \tag{4}$$

and note that $\max_j (d_j) = ||\mathbf{W}||_1 = O(n^{\delta})$. We further assume that the number of dominant units (with $\delta_j \neq 0$) is a finite number denoted by m. Without loss of generality, we presume that the first m units, $j = 1, 2, \dots, m$ (m is fixed) are δ_j-dominant (with $\delta_j \neq 0$), and the rest of the units, $j = m+1, m+2, \dots, n$, are non-dominant (with $\delta_j = 0$). In particular, the spatial weights matrix for the non-dominant units is denoted by $\mathbf{W}_{22}$, which is the $(n-m)$-dimensional square submatrix of $\mathbf{W}$ that captures the connections among the non-dominant units.

In matrix notation, model (1) can be rewritten as

$$\mathbf{y} = \rho \mathbf{y}^* + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ is an $n \times k$ matrix of observations on exogenous regressors, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$. The reduced-form representation of (5) is given by

$$\mathbf{y} = \mathbf{S}^{-1}(\rho)(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}), \tag{6}$$

where $\mathbf{S}(\rho) = \mathbf{I}_n - \rho \mathbf{W}$. The existence of $\mathbf{S}^{-1}(\rho)$ is ensured under the assumptions to be discussed below. It immediately follows from (6) that

$$\mathbf{y}^* = \mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{S}^{-1}(\rho)(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) = \mathbf{G}(\rho)(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}), \tag{7}$$

where $\mathbf{G}(\rho) = \mathbf{W}\mathbf{S}^{-1}(\rho)$. Note that the variables and spatial weights may depend on the sample size and form triangular arrays, although we suppress subscript n for notational simplicity.

The parameters of interest are ρ and $\boldsymbol{\beta}$, and their true values are denoted by ρ_0 and $\boldsymbol{\beta}_0$, respectively. For ease of exposition, we use $\mathbf{S}_0$ to denote the matrix $\mathbf{S}(\rho)$ evaluated at the true parameter value ρ_0, namely, $\mathbf{S}_0 = \mathbf{S}(\rho_0) = \mathbf{I}_n - \rho_0\mathbf{W}$. Similarly, we set

$$\mathbf{G}_0 = \mathbf{G}(\rho_0) = \mathbf{W}\left(\mathbf{I}_n - \rho_0\mathbf{W}\right)^{-1} = \mathbf{W}\mathbf{S}_0^{-1}, \tag{8}$$

and

$$\boldsymbol{\eta}_0 = \mathbf{G}_0\mathbf{X}\boldsymbol{\beta}_0. \tag{9}$$

The following assumptions are made to carry out the asymptotic analysis.

Assumption 1 *The idiosyncratic errors, ε_i, for $i = 1, 2, \ldots, n$,* in the SAR model given by (1) are independently and identically distributed (IID) with zero means and a constant variance, σ^2, *such that* $0 < \sigma^2 < K$, *and* $\sup_i E\left|\varepsilon_i\right|^{4+\epsilon} < K$.

Assumption 2 *The $(k+2)$-dimensional vector of parameters of model (1), $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}', \sigma^2)' \in \boldsymbol{\Theta} = \boldsymbol{\Theta}_\rho \times \boldsymbol{\Theta}_\beta \times \boldsymbol{\Theta}_{\sigma^2}$, where $\boldsymbol{\Theta}_\rho$, $\boldsymbol{\Theta}_\beta$, and $\boldsymbol{\Theta}_{\sigma^2}$ are compact subsets of $(-1, 1)$, $\mathbb{R}^k$ and $(0, \infty)$, respectively; the true value of $\boldsymbol{\theta}$, denoted by $\boldsymbol{\theta}_0 = (\rho_0, \boldsymbol{\beta}_0', \sigma_0^2)'$, lies in the interior of the parameter space, $\boldsymbol{\Theta}$.*

Assumption 3 *Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)'$ be the $n \times k$ matrix of observations on the regressors in model (1), where $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ik})'$. (a) $\mathbf{x}_i$, for $i = 1, 2, \ldots, n$, are distributed independently of the errors, ε_j, for all i and j, and $\sup_{i,s} E\left(|x_{is}|^{2+\epsilon}\right) < K$, (b) $n^{-1}\mathbf{X}'\mathbf{X} \to_p \boldsymbol{\Sigma}_{xx}$ is positive definite, and (c) $n^{-1}\mathbf{X}'\mathbf{G}_0\mathbf{X} \to_p \boldsymbol{\Sigma}_{xgx}$ and $n^{-1}\mathbf{X}'\mathbf{G}_0'\mathbf{G}_0\mathbf{X} \to_p \boldsymbol{\Sigma}_{xggx}$, where $\mathbf{G}_0$ is defined by (8).*

Assumption 4 *The spatial weights matrix, $\mathbf{W} = (w_{ij})$,* in the SAR model given by (1) *is non-negative, namely, $w_{ij} \geq 0$ for all i and j; it is row-standardized such that $\mathbf{W}\mathbf{1}_n = \mathbf{1}_n$.*

Assumption 5 *The column sums of the spatial weights matrix $\mathbf{W} = (\mathbf{w}_{ij})$, denoted by $d_j = \sum_{i=1}^n w_{ij}$, $j = 1, 2, \ldots, n$, are non-zero and follow the specification given by (3), where $\delta_j \neq 0$ for $j = 1, 2, \ldots, m$, and $\delta_j = 0$ for $j = m+1, m+2, \ldots, n$, with m being a fixed number. Also, $|\rho| \left\|\mathbf{W}_{22}\right\|_1 < 1$, where $\mathbf{W}_{22}$ is the $(n-m)$-dimensional square submatrix of $\mathbf{W}$ that represents the connections among the non-dominant units.*

Assumption 6 *There exists n_0 such that for all $n \geq n_0$ (including $n \to \infty$), either*

(a) $n^{-1}\mathbf{Q}_0'\mathbf{Q}_0$ is positive definite, where $\mathbf{Q}_0 = (\boldsymbol{\eta}_0, \mathbf{X})$, $\boldsymbol{\eta}_0$ is defined by (9),

and/or

(b) $h_n > \epsilon > 0$, where

$$h_n = n^{-1}Tr\left(\mathbf{G}_0'\mathbf{G}_0 + \mathbf{G}_0^2\right) - 2n^{-2}\left[Tr\left(\mathbf{G}_0\right)\right]^2, \tag{10}$$

and $\mathbf{G}_0$ is given by (8).

Remark 1 *It is worth noting that under Assumption 4, the matrix* $\mathbf{S}(\rho)$ *is invertible for all* ρ *satisfying* $|\rho| < 1$*, irrespective of whether the column sums of* $\mathbf{W}$ *are bounded.*[4]

Remark 2 *Under Assumption 5, the first* m *column sums of* $\mathbf{W}$ *are unbounded and rise with* n*, while the remaining* $(n - m)$ *column sums are bounded in* n*. Although* m *is assumed to be fixed and does not rise with* n*, it can be shown that this must be true if* δ_j *satisfies the summability condition:* $\sum_{j=1}^{n} \delta_j < K$*.*[5] *Also, the ordering of the dominant units does not affect the analysis. The current paper is not concerned with the identities of the dominant units, but rather it focuses on the estimation of* ρ *and* $\boldsymbol{\beta}$ *when there are dominant units, such that* $\delta > 0$*.*

Remark 3 *The non-negativity assumption,* $w_{ij} \geq 0$*, is imposed only for ease of exposition and is not restrictive. When it fails to hold, one can decompose* $\mathbf{W}$ *into two weights matrices with non-negative elements, namely,* $\mathbf{W} = \mathbf{W}^{+} - \mathbf{W}^{-} = (w_{ij}^{+}) - (w_{ij}^{-})$*, with* w_{ij}^{+} *and* $w_{ij}^{-} \geq 0$*. Then model (1) can be written as* $\mathbf{y} = \rho_1 \mathbf{W}^{+}\mathbf{y} + \rho_2 \mathbf{W}^{-}\mathbf{y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$*. See Bailey et al. (2016) for an empirical application employing this strategy.*

3 A generalization of the central limit theorem for linear-quadratic forms of Kelejian and Prucha (2001)

To allow for the presence of dominant units in the SAR model, we need to generalize the central limit theorem established in Theorem 1 of Kelejian and Prucha (2001) for linear-quadratic forms. First we consider the quadratic term which helps clarify the role played by the rate at which the column sum norm of the $n \times n$ weights matrix, $\mathbf{W}$, varies with n. We then consider the extension of this theorem to linear-quadratic forms needed for the analysis of SAR models with exogenous regressors. In what follows we state the theorems and relegate their proofs to Section A.2 of the online mathematical appendix.[6]

Theorem 1 *Let* $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ *denote the* $n \times 1$ *vector of* $IID\,(0, \sigma^2)$ random variables, where $0 < \sigma^2 < K$*. Suppose that* $\sup_i E\,|\varepsilon_i|^{4+\epsilon} < K$*, and denote the excess kurtosis of* $\{\varepsilon_i\}$ *by* $k_e = (\mu_4/\sigma^4) - 3$*, where* $\mu_4 = E\left(\varepsilon_i^4\right)$*. Let* $\mathbf{P} = (p_{ij})$ *be an array of* $n \times n$ *constant matrices that satisfy the following conditions*

$$\|\mathbf{P}\|_{\infty} = \sup_i \sum_{j=1}^{n} |p_{ij}| < K, \tag{11}$$

[4] See Lemma A.1 in Appendix A of Pesaran and Yang (2019).

[5] See Proposition 2 of Pesaran and Yang (2019).

[6] Note that the elements of the weights matrix, $\mathbf{W}$, and the error vector, $\boldsymbol{\varepsilon}$, typically depend on n, the sample size. But, unless required for clarity, we suppress subscript n to simplify the notations.

$$\|\mathbf{P}\|_1 = \sup_j \sum_{i=1}^{n} |p_{ij}| = O\left(n^{\delta}\right), \quad 0 \leq \delta < 1, \tag{12}$$

and $\mathbf{P}$ *has a finite number of unbounded columns. Define* $\mathbf{A} = (a_{ij}) = (\mathbf{P} + \mathbf{P}')/2$*. Also* $n^{-1}Tr(\mathbf{A}^2)$ *is such that*

$$n^{-1}Tr\left(\mathbf{A}^2\right) + \frac{1}{2}k_e\left(n^{-1}\sum_{i=1}^{n} a_{ii}^2\right) > \epsilon > 0, \text{ for all } n \text{ (including } n \to \infty). \tag{13}$$

Then if δ *lies in the range* $0 \leq \delta < 1/2$*, we have*

$$Q = \frac{\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon} - \sigma^2 Tr(\mathbf{A})}{\sqrt{n}\omega_n} \to_d N(0,1), \text{ as } n \to \infty, \tag{14}$$

where

$$\omega_n^2 = 2\sigma^4 n^{-1}Tr\left(\mathbf{A}^2\right) + k_e\sigma^4\left(n^{-1}\sum_{i=1}^{n} a_{ii}^2\right). \tag{15}$$

In application of the above theorem to GMM and BMM estimators of ρ, the column norm properties of the weights matrix, $\mathbf{W}$, carry over to matrix $\mathbf{P}$ in the above theorem, and allow us to establish asymptotic normality of the estimators even if $\mathbf{W}$ has unbounded column norms. It is also worth noting that matrix $\mathbf{P}$ in the above theorem need not be row-standardized, and our results hold as long as $\mathbf{P}$ is uniformly bounded in row norms, as stated in (11).

Remark 4 *It is straightforward to see that (13) implies* $\omega_n^2 > \epsilon > 0$*, for all* n *(including* $n \to \infty$*). If the errors are normally distributed, then* $k_e = 0$*, and condition (13) reduces to* $n^{-1}Tr(\mathbf{A}^2) > \epsilon > 0$*, which always holds true for finite* n *(except for the trivial case of* $\mathbf{A} = \mathbf{0}$*). Therefore in the case of* $k_e = 0$*, to ensure* $\omega_n^2 > \epsilon > 0$*, it is sufficient to assume that* $n^{-1}Tr(\mathbf{A}^2)$ *tends to a strictly positive limit as* $n \to \infty$*.*

The next theorem extends Theorem 1 to the linear-quadratic forms, which is required for establishing the asymptotic properties of the GMM and BMM estimators of SAR models with exogenous regressors.

Theorem 2 *Let* $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ *denote the* $n \times 1$ *vector of* $IID(0, \sigma^2)$ random variables, where $0 < \sigma^2 < K$*. Suppose that* $\sup_i E|\varepsilon_i|^{4+\epsilon} < K$*, and denote the excess kurtosis of* $\{\varepsilon_i\}$ *by* $k_e = (\mu_4/\sigma^4) - 3$*, where* $\mu_4 = E(\varepsilon_i^4)$*. Let* $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)'$ *be a vector of random variables with means* $\mu_{\eta,i}$ *and variances* $\mu_{2\eta,i}$*, distributed independently of* ε_j*, for all* i *and* j*, where* $\mu_{2\eta,i} > 0$*, for all* i*, and* $\sup_i E\left(|\eta_i|^{2+\epsilon}\right) < K$*. Let* $\mathbf{P} = (p_{ij})$ *be an array of* $n \times n$ *constant matrices that satisfy conditions (11) and (12), and* $\mathbf{P}$ *has a finite number of unbounded columns, with* $\delta \geq 0$*. Define* $\mathbf{A} = (a_{ij}) = (\mathbf{P} + \mathbf{P}')/2$*. Suppose* $n^{-1}Tr(\mathbf{A}^2)$ *is such that*

$$n^{-1}Tr\left(\mathbf{A}^2\right) + \frac{1}{2}k_e\left(n^{-1}\sum_{i=1}^{n} a_{ii}^2\right) + \frac{1}{2\sigma^2}\left(n^{-1}\sum_{i=1}^{n} \mu_{2\eta,i}\right) + \frac{\mu_3}{\sigma^4}\left(n^{-1}\sum_{i=1}^{n} a_{ii}\mu_{\eta,i}\right) > \epsilon > 0 \tag{16}$$

for all n (including $n \to \infty$). Then if δ lies in the range $0 \leq \delta < 1/2$, we have[7]

$$\tilde{Q} = \frac{\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\boldsymbol{\eta} - \sigma^2 Tr\left(\mathbf{A}\right)}{\sqrt{n}\tilde{\omega}_n} \to_d N(0,1), \text{ as } n \to \infty, \tag{17}$$

where

$$\tilde{\omega}_n^2 = 2\sigma^4 n^{-1} Tr\left(\mathbf{A}^2\right) + k_e \sigma^4 \left(n^{-1}\sum_{i=1}^n a_{ii}^2\right) + \sigma^2 \left(n^{-1}\sum_{i=1}^n \mu_{2\eta,i}\right) + 2\mu_3 \left(n^{-1}\sum_{i=1}^n a_{ii}\mu_{\eta,i}\right). \tag{18}$$

Remark 5 *Condition (16) ensures that $\tilde{\omega}_n^2 > \epsilon > 0$, for all n (including $n \to \infty$). If the errors are symmetrically distributed, then $\mu_3 = 0$. Since $\mu_{2\eta,i} > 0$ for all i, condition (16) in this case would reduce to (13) in Theorem 1.*

Remark 6 *Following a similar line of argument as in Kelejian and Prucha (2001), it can be shown that Theorems 1 and 2 can be extended to allow for heteroskedastic errors without affecting the conclusions.*

4 GMM estimation

We begin by extending the GMM method proposed by Lee (2007) for standard SAR models to the case where the column sums of the spatial weights matrix are not necessarily bounded in n. Lee (2007) suggests using both linear moment conditions formed with instruments and additional quadratic moments that are based on the properties of the idiosyncratic errors. Specifically, consider model (1) and let $\boldsymbol{\psi}$ denote the $(k+1)$-dimensional vector of parameters, $\boldsymbol{\psi} = (\rho, \boldsymbol{\beta}')' \in \boldsymbol{\Psi} = \boldsymbol{\Theta}_\rho \times \boldsymbol{\Theta}_\beta$. The true value of $\boldsymbol{\psi}$ is denoted by $\boldsymbol{\psi}_0 = (\rho_0, \boldsymbol{\beta}_0')'$. Suppose that $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n)'$ is an $n \times r$ ($r \geq k+1$) matrix of instruments for the regressors $(\mathbf{y}^*, \mathbf{X})$. Formally, $\mathbf{Z}$ satisfies the following assumption.

Assumption 7 *Let $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n)'$ be the $n \times r$ matrix of observations on the r instrumental variables, $\mathbf{z}_i = (z_{i1}, z_{i2}, \ldots, z_{ir})'$. (a) $\mathbf{z}_i$ is distributed independently of the errors, ε_j, for all i and $j = 1, 2, \ldots, n$, and $\sup_{i,s} E\left(|z_{is}|^{2+\epsilon}\right) < K$, (b) $n^{-1}\mathbf{Z}'\mathbf{Z} \to_p \boldsymbol{\Sigma}_{zz}$, a positive definite matrix, and (c) $n^{-1}\mathbf{Z}'\mathbf{Q}_0 \to_p \boldsymbol{\Sigma}_{zq}$ is a full column rank matrix, where $\mathbf{Q}_0 = (\boldsymbol{\eta}_0, \mathbf{X})$ and $\boldsymbol{\eta}_0$ is defined by (9).*

The r linear moment conditions are given by:

$$E_0\left[\mathbf{Z}'\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right)\right] = \mathbf{0}, \tag{19}$$

where

$$\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right) = \mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}. \tag{20}$$

[7] Recall that δ is defined by (12).

Since $\mathbf{X}$ is strictly exogenous under Assumption 3, a possible candidate for $\mathbf{Z}$ consists of linearly independent columns of $(\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \ldots)$. This choice of instruments was first proposed by Kelejian and Prucha (1998). To see why $\mathbf{Z}$ could take this form, note from (7) that $E(\mathbf{y}^*|\mathbf{X}) = \mathbf{G}(\rho)\mathbf{X}\boldsymbol{\beta}$. This term is clearly correlated with $\mathbf{y}^*$ but uncorrelated with $\boldsymbol{\varepsilon}$. Since $|\rho| \|\mathbf{W}\|_\infty < 1$ under Assumptions 2 and 4, $\mathbf{G}(\rho)$ can be expanded as

$$\mathbf{G}(\rho) = \mathbf{W}(\mathbf{I}_n - \rho\mathbf{W})^{-1} = \mathbf{W} + \rho\mathbf{W}^2 + \rho^2\mathbf{W}^3 + \ldots, \tag{21}$$

and then $\mathbf{G}(\rho)\mathbf{X}\boldsymbol{\beta} = \sum_{j=1}^{\infty} \rho^{j-1}\mathbf{W}^j\mathbf{X}\boldsymbol{\beta}$. This implies that the instruments for $\mathbf{y}^*$ can be chosen from the columns of $(\mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \ldots)$. Furthermore, Lee (2003) has shown that the asymptotically best IV matrix within the 2SLS framework is given by $\mathbf{Z}^* = \mathbf{Q}_0 = (\boldsymbol{\eta}_0, \mathbf{X})$. Since $\mathbf{Z}^*$ depends on the unknown parameters ρ_0 and $\boldsymbol{\beta}_0$, a feasible best IV can be constructed using some initial consistent estimates of the parameters.

Turning to the quadratic moment condition, we recall that the idiosyncratic errors are assumed to be cross-sectionally uncorrelated and homoskedastic. Using these properties, we have the following moment condition:

$$E_0\left[\boldsymbol{\varepsilon}'(\boldsymbol{\psi})\mathbf{C}\boldsymbol{\varepsilon}(\boldsymbol{\psi})\right] = 0, \tag{22}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{\psi})$ is defined by (20),

$$\mathbf{C} = (c_{ij}) = (\mathbf{B} + \mathbf{B}')/2, \tag{23}$$

and $\mathbf{B}$ is a matrix that satisfies the following assumption.

Assumption 8 *The matrix $\mathbf{B} = (b_{ij})$ is an $n \times n$ matrix of fixed constants such that (a) $Tr(\mathbf{B}) = 0$, (b) $||\mathbf{B}||_\infty < K$, (c) $||\mathbf{B}||_1 = O\left(n^{\delta_b}\right)$, where δ_b is a fixed constant in the range $0 \leq \delta_b < 1$, (d) $n^{-1}\mathbf{X}'\mathbf{C}\mathbf{X} \to_p \boldsymbol{\Sigma}_{xcx}$, $n^{-1}\boldsymbol{\eta}_0'\mathbf{C}\boldsymbol{\eta}_0 \to_p c_0$, and $n^{-1}\boldsymbol{\eta}_0'\mathbf{C}\mathbf{X} \to_p \mathbf{d}_0'$, where $\boldsymbol{\eta}_0$ is given by (9), $\mathbf{C} = (\mathbf{B} + \mathbf{B}')/2$, and $\mathbf{X}$ is the $n \times k$ matrix of observations on the regressors in model (1).*

Equation (22) is a valid moment condition since at the true value $\boldsymbol{\psi}_0$ we have

$$E_0\left[\boldsymbol{\varepsilon}'(\boldsymbol{\psi}_0)\mathbf{C}\boldsymbol{\varepsilon}(\boldsymbol{\psi}_0)\right] = E_0(\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}) = n\sigma_0^2 Tr(\mathbf{C}) = n\sigma_0^2 Tr(\mathbf{B}) = 0.$$

Here we consider a single quadratic moment for ease of exposition. In practice, one could use multiple quadratic moment conditions, namely $E_0(\boldsymbol{\varepsilon}'\mathbf{C}_\ell\boldsymbol{\varepsilon}) = 0$, for $\ell = 1, 2, \ldots, L$, where L is a finite number, $\mathbf{C}_\ell = (\mathbf{B}_\ell + \mathbf{B}_\ell')/2$, and $\mathbf{B}_\ell$ satisfies the conditions of Assumption 8. Lee (2007) assumes that $\mathbf{B}$ is uniformly bounded in both row and column sums in absolute value and suggests using $\mathbf{B}_\ell = \mathbf{W}^\ell - n^{-1}Tr(\mathbf{W}^\ell)\mathbf{I}_n$, for $\ell = 1, 2, \ldots, L$, in the quadratic moments, where $\mathbf{W}^\ell$ denotes the ℓ^{th} power of $\mathbf{W}$. However, in our set up where columns of $\mathbf{W}$ need not be bounded (see Assumptions 4 and 5), in part (a) of Assumption 8 we have relaxed Lee's boundedness condition on $\mathbf{B}$, and allow the column norm of $\mathbf{B}$ to rise with n at the rate of δ_b.

Remark 7 *If the idiosyncratic errors are heteroskedastic, condition (a) of Assumption 8, $Tr(\mathbf{B}) = 0$, needs to be replaced by the stronger condition: $b_{ii} = 0$, for $i = 1, 2, \ldots, n$. Practical choices of $\mathbf{B}$ in this case can be $\mathbf{B}_\ell = \mathbf{W}^\ell - Diag(\mathbf{W}^\ell)$, for $\ell = 1, 2, \ldots, L$.*

We are now ready to define the GMM estimator of $\boldsymbol{\psi}_0$ of model (1), denoted by $\tilde{\boldsymbol{\psi}} = (\tilde{\rho}, \tilde{\boldsymbol{\beta}}')'$, using both quadratic and linear moment conditions:

$$\tilde{\boldsymbol{\psi}} = \arg\min_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \boldsymbol{g}_n'(\boldsymbol{\psi}) \left(\mathbf{A}_n'\mathbf{A}_n\right) \boldsymbol{g}_n\left(\boldsymbol{\psi}\right), \tag{24}$$

where $\boldsymbol{g}_n(\boldsymbol{\psi})$ is a $(k+1) \times 1$ vector given by

$$\boldsymbol{g}_n(\boldsymbol{\psi}) = \begin{pmatrix} n^{-1}\boldsymbol{\varepsilon}'\left(\boldsymbol{\psi}\right)\mathbf{C}\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right) \\ n^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right) \end{pmatrix}, \tag{25}$$

and $\mathbf{A}_n$ is a fixed $(k+1) \times (r+1)$ matrix of full row rank, assumed to converge to a constant full row rank matrix $\mathbf{A}$.

Before proceeding to examine the asymptotic properties of $\tilde{\boldsymbol{\psi}}$, we first focus on the problem of identification in the case of pure SAR models without exogenous regressors. In this case (1) simplifies to,

$$\mathbf{y} = \rho\mathbf{y}^* + \boldsymbol{\varepsilon}, \tag{26}$$

with Assumption 2 replaced by

Assumption 9 *The parameter ρ of model (26) satisfies $\rho \in \boldsymbol{\Theta}_\rho$, where $\boldsymbol{\Theta}_\rho$ is a compact subset of $(-1, 1)$. The true value of ρ, denoted by ρ_0, lies in the interior of the parameter space, $\boldsymbol{\Theta}_\rho$.*

The GMM estimator of ρ_0 in model (26) can be obtained by

$$\tilde{\rho} = \arg\min_{\rho\in\boldsymbol{\Theta}_\rho} g_n^2\left(\rho\right), \tag{27}$$

where

$$g_n\left(\rho\right) = n^{-1}\boldsymbol{\varepsilon}'\left(\rho\right)\mathbf{C}\boldsymbol{\varepsilon}\left(\rho\right), \tag{28}$$

and

$$\boldsymbol{\varepsilon}\left(\rho\right) = \mathbf{y} - \rho\mathbf{y}^*. \tag{29}$$

Proposition 1 below shows that in order to uniquely identify ρ_0 in the pure SAR model (26), at least two moment conditions are required. Specifically, the GMM estimator of ρ_0 based on L quadratic moments (L is a finite number) is given by

$$\tilde{\rho} = \arg\min_{\rho\in\boldsymbol{\Theta}_\rho} \left[\mathbf{a}_n'\boldsymbol{g}_n\left(\rho\right)\right]^2, \tag{30}$$

where

$$\boldsymbol{g}_n\left(\rho\right) = \left[g_{1,n}\left(\rho\right), g_{2,n}\left(\rho\right), \ldots, g_{L,n}\left(\rho\right)\right]',$$

$$g_{\ell,n}(\rho) = n^{-1}\boldsymbol{\varepsilon}'(\rho)\,\mathbf{C}_{\ell}\boldsymbol{\varepsilon}(\rho)\,, \text{ for } \ell = 1,2,\ldots,L,$$

and $\mathbf{a}_n$ is a fixed $L \times 1$ non-zero non-negative vector.

Proposition 1 *Consider the SAR model given by (26), and suppose that Assumptions 1, 4, 5, 8(a)–(c), and 9 hold. Then to uniquely identify ρ_0 it is required that the GMM estimator, defined by (30), is based on at least two independent quadratic moment conditions, in the sense that the ratios $b_{\ell 0}/a_{\ell 0}$, are not all the same across $\ell = 1,2,\ldots,L \geq 2$, where $a_{\ell 0} = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}_{\ell}\mathbf{G}_0\right)$ and $b_{\ell 0} = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}_{\ell}\right)$.*

See Section A.2 of the online mathematical appendix for a proof.

Remark 8 *When the GMM estimator is based on a single quadratic moment condition, the parameter ρ_0 of model (26) in not uniquely identified and the GMM estimator of ρ computed by minimizing $g_n^2(\rho)$ defined by (28), converges in probability to ρ_0 or $\rho_0 + 2b_0/a_0$, where $a_0 = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}\mathbf{G}_0\right)$ and $b_0 = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}\right)$. In practice, we recommend using at least two quadratic moments if the SAR model does not contain exogenous regressors.*

Consider now the SAR model given by (1) that includes exogenous regressors. For ease of exposition, in what follows we set $\delta_b = \delta$, that is, $||\mathbf{B}||_1$ rises with n at the same rate as that of $||\mathbf{W}||_1$, since in practice $\mathbf{W}$ is commonly adopted as the $\mathbf{B}$ matrix. The following theorem shows that $\boldsymbol{\psi}_0 = (\rho_0, \boldsymbol{\beta}_0')$ of model (1) can be globally identified if we have enough instruments such that the rank condition in Assumption 7(c) holds. The theorem also establishes consistency and asymptotic normality of the GMM estimator defined by (24).

Theorem 3 *Consider the SAR model given by (1). Suppose that Assumptions 1–5, 7 and 8 hold, and $\delta_b = \delta$. Then*

(a) $\boldsymbol{\psi}_0 = (\rho_0, \boldsymbol{\beta}_0')$ is globally identified,

(b) the GMM estimator of $\boldsymbol{\psi}_0$, denoted by $\tilde{\boldsymbol{\psi}}$ and defined in (24), is consistent for $\boldsymbol{\psi}_0$ if δ (the degree of network centrality) defined by (4) lies in the range $0 \leq \delta < 1$,

(c) $\sqrt{n}\left(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right)$ is asymptotically normally distributed as $n \to \infty$, if δ lies in the range $0 \leq \delta < 1/2$, namely,

$$\sqrt{n}\left(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right) \to_d N\left[\mathbf{0}, \left(\mathbf{D}'\mathbf{A}'\mathbf{A}\mathbf{D}\right)^{-1}\mathbf{D}'\mathbf{A}'\mathbf{A}\mathbf{V}_g\mathbf{A}'\mathbf{A}\mathbf{D}\left(\mathbf{D}'\mathbf{A}'\mathbf{A}\mathbf{D}\right)^{-1}\right],$$

where

$$\mathbf{D} = \left[\left(2\sigma_0^2 b_0, \mathbf{0}_{1\times k}\right)', \boldsymbol{\Sigma}_{zq}'\right]', \qquad \mathbf{V}_g = \begin{pmatrix} v_1 & \mu_{30}\boldsymbol{\upsilon}' \\ \mu_{30}\boldsymbol{\upsilon} & \sigma_0^2\boldsymbol{\Sigma}_{zz} \end{pmatrix}, \tag{31}$$

$$v_1 = \lim_{n\to\infty}\left[\gamma_{20}n^{-1}\sum_{i=1}^{n} c_{ii}^2 + 2\sigma_0^4 Tr\left(n^{-1}\mathbf{C}^2\right)\right],$$

$b_0 = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}\right)$, $\mathbf{G}_0$ is defined by (8), $\mathbf{\Sigma}_{zq} = p\lim_{n\to\infty} n^{-1}\mathbf{Z}'\mathbf{Q}_0$, $\mathbf{\Sigma}_{zz} = p\lim_{n\to\infty} n^{-1}\mathbf{Z}'\mathbf{Z}$, $\boldsymbol{\upsilon} = p\lim_{n\to\infty} n^{-1}\mathbf{Z}'\left[diag\left(\mathbf{C}\right)\right]$, $\mu_{30} = E(\varepsilon_i^3)$, $\gamma_{20} = E(\varepsilon_i^4) - 3\sigma_0^4$, and c_{ii} is the i^{th} diagonal element of $\mathbf{C}$.

See Section A.2 of the online mathematical appendix for a proof.

Remark 9 *It is worth emphasizing that $\boldsymbol{\psi}_0$ is globally identified if we have enough instruments such that the rank condition in Assumption 7(c) is satisfied, irrespective of the number of quadratic moments included. Using quadratic moments in addition to linear moments improves efficiency.*

Remark 10 *Consistent estimators of μ_{30} and γ_{20} are given by $\tilde{\mu}_3 = n^{-1}\sum_{i=1}^n \tilde{\varepsilon}_i^3$, $\tilde{\gamma}_2 = n^{-1}\sum_{i=1}^n \tilde{\varepsilon}_i^4 - 3\left(\tilde{\sigma}^2\right)^2$, respectively, where $\tilde{\sigma}^2 = n^{-1}\sum_{j=1}^n \tilde{\varepsilon}_j^2$, and $\tilde{\varepsilon}_i = y_i - \tilde{\rho}y_i^* - \tilde{\boldsymbol{\beta}}'\mathbf{x}_i$.*

As is well known, the optimal moments weighting matrix is given by $\mathbf{V}_g^{-1}$, where $\mathbf{V}_g$ is defined in (31). A feasible optimal GMM (OGMM) estimator of $\boldsymbol{\psi}_0$, denoted by $\tilde{\boldsymbol{\psi}}_{opt}$, can be obtained by using a consistent estimator of $\mathbf{V}_g^{-1}$ for $\mathbf{A}_n'\mathbf{A}_n$, that is,

$$\tilde{\boldsymbol{\psi}}_{opt} = \arg\min_{\boldsymbol{\psi}\in\boldsymbol{\Psi}} \boldsymbol{g}_n'(\boldsymbol{\psi})\tilde{\mathbf{V}}_g^{-1}\boldsymbol{g}_n'(\boldsymbol{\psi}), \tag{32}$$

where $\boldsymbol{g}_n\left(\boldsymbol{\psi}\right)$ is given by (25) and $\tilde{\mathbf{V}}_g$ is a consistent estimator of $\mathbf{V}_g$. Then $\tilde{\boldsymbol{\psi}}_{opt}$ is consistent for $\boldsymbol{\psi}_0$ when δ is in the range $0 \leq \delta < 1$, and it has the following asymptotic distribution as $n \to \infty$ when δ is in the range $0 \leq \delta < 1/2$,

$$\sqrt{n}\left(\tilde{\boldsymbol{\psi}}_{opt} - \boldsymbol{\psi}_0\right) \to_d N\left[\mathbf{0}, \left(\mathbf{D}'\mathbf{V}_g^{-1}\mathbf{D}\right)^{-1}\right],$$

where $\mathbf{D}$ is given by (31).

The best choice of $\mathbf{B}$ exists under certain conditions. Lee (2007) shows that if the idiosyncratic errors are normally distributed, the OGMM estimator using $\mathbf{B}^* = \mathbf{G}_0 - n^{-1}Tr\left(\mathbf{G}_0\right)\mathbf{I}_n$ in the quadratic moment condition and $\mathbf{Z}^* = \mathbf{Q}_0 = \left(\boldsymbol{\eta}_0, \mathbf{X}\right)$ in the linear moment conditions, has the smallest asymptotic variance among the set of GMM estimators derived with quadratic matrices, $\mathbf{B}_\ell$, having zero trace. This estimator is referred as the best GMM estimator, and $\mathbf{B}^*$ is referred to as the best quadratic matrix.[8] By a similar argument and applying Lemma A.6 of the online mathematical appendix, it is straightforward to show that the asymptotic properties of the best GMM estimator can be extended to the case where the column sums of $\mathbf{W}$ rise with n, under the same conditions on δ as in Theorem 3. Since both $\mathbf{B}^*$ and $\mathbf{Z}^*$ depend on unknown parameters, a feasible best GMM estimator can be implemented in two steps: In the first step, we obtain a preliminary consistent estimate of $\boldsymbol{\psi}_0$. Then in the second

[8] Among the group of GMM estimators derived with the class of matrices having zero diagonal, the OGMM estimator using $\mathbf{B}^* = \mathbf{G}_0 - Diag\left(\mathbf{G}_0\right)$ and $\mathbf{Z}^* = \mathbf{Q}_0 = \left(\boldsymbol{\eta}_0, \mathbf{X}\right)$ in the moments has the smallest asymptotic variance. This result does not require the condition that the idiosyncratic errors are normally distributed. See Lee (2007) Proposition 3 for details.

step, we perform the optimal GMM estimation using the best IV and best quadratic matrices evaluated at the first-stage estimates. In the rest of this paper we focus on the feasible best GMM estimator and refer to it simply as the GMM estimator, for brevity.[9]

5 BMM estimation

In this section we develop the bias-corrected method of moments (BMM) estimator of $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}_{\mathbf{0}}',\sigma_0^2)' = (\rho_0, \boldsymbol{\beta}_0',\sigma_0^2)'$ for the SAR model given by (1). The BMM procedure uses least squares but corrects the bias due to the endogeneity of the spatial variable, $\mathbf{y}^*$. The application of BMM to the SAR model is straightforward. Using $\mathbf{y}^*$ and $\mathbf{X}$ as instruments, the bias-corrected population moments are given by

$$E\left[\mathbf{y}^{*\prime}\left(\mathbf{y}-\rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}\right)\right] = E\left(\mathbf{y}^{*\prime}\boldsymbol{\varepsilon}\right), \tag{33}$$

$$E\left[\mathbf{X}'\left(\mathbf{y}-\rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}\right)\right] = \mathbf{0}, \tag{34}$$

$$E\left[\left(\mathbf{y}-\rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}\right)'\left(\mathbf{y}-\rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}\right)\right] = n\sigma^2. \tag{35}$$

Using (7), we have

$$E\left(\mathbf{y}^{*\prime}\boldsymbol{\varepsilon}\right) = E\left[\left(\boldsymbol{\beta}'\mathbf{X}' + \boldsymbol{\varepsilon}'\right)\mathbf{G}'\left(\rho\right)\boldsymbol{\varepsilon}\right],$$

and under Assumptions 1 and 3, we obtain $E\left(\mathbf{y}^{*\prime}\boldsymbol{\varepsilon}\right) = \sigma^2 Tr\left[\mathbf{G}\left(\rho\right)\right]$. The sample version of the moment conditions (33)–(35) can now be written as

$$n^{-1}\mathbf{y}^{*\prime}\left(\mathbf{y}-\hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2 Tr\left[n^{-1}\mathbf{G}\left(\hat{\rho}\right)\right], \tag{36}$$

$$n^{-1}\mathbf{X}'\left(\mathbf{y}-\hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \mathbf{0}, \tag{37}$$

$$n^{-1}\left(\mathbf{y}-\hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y}-\hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \hat{\sigma}^2. \tag{38}$$

Let $\hat{\boldsymbol{\theta}} = (\hat{\rho}, \hat{\boldsymbol{\beta}}', \hat{\sigma}^2)'$ denote the BMM estimator of $\boldsymbol{\theta}_0$, which is the true value of $\boldsymbol{\theta} = (\rho, \boldsymbol{\beta}',\sigma^2)'$. The system of equations (36)–(38) can now be used to solve for $\hat{\boldsymbol{\theta}}$ as follows:

$$\hat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbf{m}_n'\left(\boldsymbol{\theta}\right)\mathbf{m}_n\left(\boldsymbol{\theta}\right), \tag{39}$$

where

$$\mathbf{m}_n\left(\boldsymbol{\theta}\right) = \left[m_{1,n}\left(\boldsymbol{\theta}\right), \mathbf{m}_{2,n}'\left(\boldsymbol{\theta}\right), m_{3,n}\left(\boldsymbol{\theta}\right)\right]',$$
$$m_{1,n}\left(\boldsymbol{\theta}\right) = n^{-1}\mathbf{y}^{*\prime}\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right) - \sigma^2 Tr\left[n^{-1}\mathbf{G}\left(\rho\right)\right],$$
$$\mathbf{m}_{2,n}\left(\boldsymbol{\theta}\right) = n^{-1}\mathbf{X}'\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right),\ m_{3,n}\left(\boldsymbol{\theta}\right) = n^{-1}\boldsymbol{\varepsilon}'\left(\boldsymbol{\psi}\right)\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right) - \sigma^2,$$

and $\boldsymbol{\varepsilon}\left(\boldsymbol{\psi}\right)$ is given by (20).

Unlike least squares, the BMM procedure is non-linear in $\hat{\rho}$, and its asymptotic properties

[9] We also examined the finite sample properties of other GMM estimators that do not use the best IV and best quadratic matrix. The results are available upon request.

critically depends on the assumptions regarding the rate at which the column sums of $\mathbf{W}$ rise with n. As we shall see, the BMM estimators are consistent and do not suffer from the weak instrument problem since $\mathbf{y}^*$ is instrumented with its own values. However, in small samples it might be beneficial to augment the system of estimating equations, (36)–(38), with additional moment conditions. See, for example, Lee (2007).

The following theorem summarizes the asymptotic distribution of the BMM estimator. Its proof is given in Section A.2 of the online mathematical appendix.

Theorem 4 *Consider the SAR model given by (1), and suppose that Assumptions 1–6 hold. Then*

(a) the bias-corrected method of moments (BMM) estimator of $\boldsymbol{\psi}_0 = (\rho_0, \boldsymbol{\beta}_0')'$, *denoted by* $\hat{\boldsymbol{\psi}} = (\hat{\rho}, \hat{\boldsymbol{\beta}}')'$ *and defined by (39), is consistent for* $\boldsymbol{\psi}_0$ *when* δ *is in the range* $0 \leq \delta < 1$, *where* δ *is a measure of network centrality, defined by (4).*

(b) $\sqrt{n}\left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right)$ *is asymptotically normally distributed as* $n \to \infty$ *when* δ *is in the range* $0 \leq \delta < 1/2$, *namely*

$$\sqrt{n}\left(\hat{\boldsymbol{\psi}} - \boldsymbol{\psi}_0\right) \to_d N\left[\mathbf{0}, \left(\mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1}\right)\right],$$

where

$$\mathbf{H} = \begin{pmatrix} \boldsymbol{\beta}_0'\boldsymbol{\Sigma}_{xggx}\boldsymbol{\beta}_0 + \sigma_0^2 h_0 & \boldsymbol{\beta}_0'\boldsymbol{\Sigma}_{xgx} \\ \boldsymbol{\Sigma}_{xgx}\boldsymbol{\beta}_0 & \boldsymbol{\Sigma}_{xx} \end{pmatrix}, \qquad \mathbf{V} = \begin{pmatrix} q^2 & \sigma_0^2\boldsymbol{\beta}_0'\boldsymbol{\Sigma}_{xgx} \\ \sigma_0^2\boldsymbol{\Sigma}_{xgx}\boldsymbol{\beta}_0 & \sigma_0^2\boldsymbol{\Sigma}_{xx} \end{pmatrix}, \tag{40}$$

$$q^2 = \sigma_0^2\boldsymbol{\beta}_0'\boldsymbol{\Sigma}_{xggx}\boldsymbol{\beta}_0 + \gamma_{20}\ p\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\pi_{ii,0}^2 + 2\mu_{30}\ p\lim_{n\to\infty} n^{-1}\left[diag\left(\boldsymbol{\Pi}_0\right)\right]'\boldsymbol{\eta}_0 \tag{41}$$

$$+ \sigma_0^4 p\lim_{n\to\infty}\left[Tr\left(n^{-1}\boldsymbol{\Pi}_0'\boldsymbol{\Pi}_0\right) + Tr\left(n^{-1}\boldsymbol{\Pi}_0^2\right)\right],$$

$$\boldsymbol{\Pi}_0 = \mathbf{G}_0 - \mathbf{M}_x Tr\left(n^{-1}\mathbf{G}_0\right), \ \ \mathbf{M}_x = \mathbf{I}_n - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}', \tag{42}$$

where $h_0 = \lim_{n\to\infty} h_n$, $\mathbf{G}_0$, $\boldsymbol{\eta}_0$ *and* h_n *are defined by (8), (9) and (10), respectively,* $\pi_{ii,0}$ *is the* i^{th} *diagonal element of* $\boldsymbol{\Pi}_0$, *and as before* $\mu_{30} = E(\varepsilon_i^3)$ *and* $\gamma_{20} = E(\varepsilon_i^4) - 3\sigma_0^4$.

Remark 11 *It can be seen from (41) that the variance formula will not involve the third and fourth moments of the error term if (i)* ε_i *is Gaussian, since under Gaussianity* $\gamma_{20} = 0$ *and* $\mu_{30} = 0$; *or (ii) the diagonal elements of* $\boldsymbol{\Pi}_0$ *are zero, which occurs if* $\mathbf{G}_0$ *has zero diagonal entries. Furthermore, the variances of both BMM and GMM estimators will not involve the third moment of the error term if the model does not contain* $\mathbf{X}$. *In general,* μ_{30} *can be estimated by* $\hat{\mu}_3 = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^3$, *where* $\hat{\varepsilon}_i = y_i - \hat{\rho}y_i^* - \hat{\boldsymbol{\beta}}'\mathbf{x}_i$, *and* γ_{20} *can be estimated by* $\hat{\gamma}_2 = n^{-1}\sum_{i=1}^{n}\hat{\varepsilon}_i^4 - 3\left(\hat{\sigma}^2\right)^2$.

Remark 12 *It is clear that* $\boldsymbol{\psi}_0$ *is identified if* $\mathbf{H}$, *defined in (40), is positive definite. Note that* $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2$, *where* $\mathbf{H}_1 = p\lim_{n\to\infty} n^{-1}\mathbf{Q}_0'\mathbf{Q}_0$, *and*

$$\mathbf{H}_2 = \begin{pmatrix} \sigma_0^2 h_0 & \mathbf{0}_{1\times k} \\ \mathbf{0}_{k\times 1} & \mathbf{0}_{k\times k} \end{pmatrix}.$$

Since $\mathbf{H}_1$ is positive semi-definite and $h_0 \geq 0$, it follows that $\mathbf{H}$ is positive definite if either $h_0 > 0$ and/or if $\mathbf{H}_1$ is positive definite. Therefore, Assumption 6 ensures that $\boldsymbol{\psi}_0$ is identified.

Remark 13 *It is meaningful to relate the identification condition given by Assumption 6(b) to the literature on social interactions. Let us first consider a simple example where there is only one social group in which everyone is connected with each other and self-influence is excluded. In this case, the matrix of the network is represented by*

$$\mathbf{W} = \frac{1}{n-1}\left(\mathbf{1}_n\mathbf{1}_n' - \mathbf{I}_n\right). \tag{43}$$

Yang (2018) has shown that a necessary condition for Assumption 6(b) is given by $n^{-1}Tr\left(\mathbf{W}'\mathbf{W}\right) > \epsilon > 0$ for all n (including $n \to \infty$). Given (43), it is easily verified that

$$\mathbf{W}^2 = \left(\frac{1}{n-1}\right)^2 \left[n\left(\mathbf{1}_n\mathbf{1}_n'\right) - 2\mathbf{1}_n\mathbf{1}_n' + \mathbf{I}_n\right],$$

and then $n^{-1}Tr\left(\mathbf{W}^2\right) = 1/\left(n-1\right)$, which tends to zero, as $n \to \infty$. Therefore, the identification condition is violated and we conclude that the endogenous social effect is unidentifiable without exogenous regressors. Now suppose that there are R groups and n_r units in the r^{th} group, for $r = 1, 2, \ldots, R$. Clearly, $\sum_{r=1}^{R} n_r = n$. The standard linear-in-means social interaction model assumes that individuals within a group have the same pairwise dependence, whereas individuals across different groups are not dependent. See Case (1991, 1992) for examples of empirical studies employing such a network structure. Then the matrix of group interactions, $\mathbf{W}$, can be represented by the following block diagonal matrix:

$$\mathbf{W} = Diag\left(\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_R\right), \quad \mathbf{W}_r = \frac{1}{n_r - 1}\left(\mathbf{1}_n\mathbf{1}_n' - \mathbf{I}_{n_r}\right), \quad r = 1, 2, \ldots, R.$$

Since we have shown that $Tr\left(\mathbf{W}_r^2\right) = n_r/\left(n_r - 1\right)$, it follows that

$$n^{-1}Tr\left(\mathbf{W}^2\right) = n^{-1}\sum_{r=1}^{R} Tr\left(\mathbf{W}_r^2\right) = \sum_{r=1}^{R}\left(\frac{1}{n_r - 1}\right)\pi_r,$$

where $\pi_r = n_r/n$ is the fraction of population in the r^{th} group. Suppose that n_r rises with n such that $\pi_r \geq 0$, as $n \to \infty$. If R is fixed, then $\lim_{n\to\infty} n^{-1}Tr\left(\mathbf{W}^2\right) = 0$ and the group interaction effect is unidentified in the absence of exogenous explanatory variables.

Interestingly, it turns out that the BMM estimator is related to the best GMM estimator under IID normal errors. The following proposition summarizes this relationship. Its proof is given in Section A.2 of the online mathematical appendix.

Proposition 2 *Consider the SAR model given by (1), and assume that the errors are independently and normally distributed as $\varepsilon_i \sim IIDN\left(0, \sigma^2\right)$, for $i = 1, 2, \ldots, n$, and $0 < \sigma^2 < K$. Suppose that Assumptions 2–8 hold and the network centrality, δ, defined by (4), lies in the range $0 \leq \delta < 1/2$. Then the BMM estimator of $\boldsymbol{\psi}_0 = \left(\rho_0, \boldsymbol{\beta}_0'\right)'$, defined by (39), has* *the same asymptotic distribution as the best GMM estimator of* $\boldsymbol{\psi}_0$*, defined by (32) using* $\mathbf{B}^* = \mathbf{G}_0 - n^{-1}Tr\left(\mathbf{G}_0\right)\mathbf{I}_n$ *in the quadratic moment condition, and* $\mathbf{Z}^* = (\boldsymbol{\eta}_0, \mathbf{X})$ *in the linear moment conditions, where* $\mathbf{G}_0$ *and* $\boldsymbol{\eta}_0$ *are defined by (8) and (9), respectively.*

6 Monte Carlo experiments

We now examine the small sample properties of the GMM and BMM estimators for SAR models with dominant units using Monte Carlo techniques. The Data Generating Process (DGP) is specified as follows:

$$y_i = \alpha + \rho y_i^* + \beta x_i + \sigma_\varepsilon \varepsilon_i, \quad i = 1, 2, \ldots, n, \tag{44}$$

where $y_i^* = \mathbf{w}_{i.,y}^{\prime}\mathbf{y}$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)^{\prime}$, and $\mathbf{w}_{i.,y}^{\prime}$ is the i^{th} row of $\mathbf{W}_y$. The exogenous regressor, x_i, is generated to be spatially correlated as

$$x_i = \lambda x_i^* + \sigma_\nu \nu_i, \quad i = 1, 2, \ldots, n, \tag{45}$$

where $x_i^* = \mathbf{w}_{i.x}^{\prime}\mathbf{x}$, $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\prime}$, and $\mathbf{w}_{i.,x}^{\prime}$ is the i^{th} row of $\mathbf{W}_x$. Note that the spatial coefficients and weights matrices could be different for the $\mathbf{y}$ and $\mathbf{x}$ processes.

In matrix form, (44) can be rewritten as

$$\mathbf{y} = \mathbf{S}_y^{-1}\left(\rho\right)\left(\beta\mathbf{x}+\alpha\mathbf{1}_n\right) + \mathbf{u},$$

where $\mathbf{S}_y\left(\rho\right) = \mathbf{I}_n - \rho\mathbf{W}_y$, $\mathbf{u} = \sigma_\varepsilon\mathbf{S}_y^{-1}\left(\rho\right)\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\prime}$, and $\mathbf{u} = (u_1, u_2, \ldots, u_n)^{\prime}$. Similarly, (45) can be rewritten as $\mathbf{x} = \sigma_\nu\mathbf{S}_x^{-1}(\lambda)\boldsymbol{\nu}$, where $\mathbf{S}_x\left(\lambda\right) = \mathbf{I}_n - \lambda\mathbf{W}_x$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_n)^{\prime}$. For the idiosyncratic errors, we consider both Gaussian and non-Gaussian processes:

- Gaussian errors: $\varepsilon_i \sim IIDN\left(0, 1\right)$ and $\nu_i \sim IIDN\left(0, 1\right)$.
- Non-Gaussian errors: $\varepsilon_i \sim IID\left[\chi^2(2) - 2\right]/2$ and $\nu_i \sim IID\left[\chi^2(2) - 2\right]/2$, where $\chi^2(2)$ denotes a chi-square random variable with two degrees of freedom.

When $\beta = 0$, the fit of the SAR model is given by

$$R_0^2 = 1 - \frac{n\sigma_\varepsilon^2}{Tr\left[Var(\mathbf{y})\right]} = 1 - \frac{n}{Tr\left[\mathbf{S}_y^{-1}\left(\rho\right)\mathbf{S}_y^{\prime -1}\left(\rho\right)\right]}, \tag{46}$$

which does not depend on σ_ε^2, and is determined by the choice of ρ and $\mathbf{W}$. To control the fit of the SAR model when $\beta \neq 0$, we note that

$$R_\beta^2 = 1 - \frac{n\sigma_\varepsilon^2}{Tr\left[Var(\mathbf{y})\right]}, \tag{47}$$

where

$$Tr\left[Var(\mathbf{y})\right] = \beta^2\sigma_v^2 Tr\left[\mathbf{S}_y^{-1}\left(\rho\right)\mathbf{S}_x^{-1}\left(\lambda\right)\mathbf{S}_x^{\prime -1}\left(\lambda\right)\mathbf{S}_y^{\prime -1}\left(\rho\right)\right] + \sigma_\varepsilon^2 Tr\left[\mathbf{S}_y^{-1}\left(\rho\right)\mathbf{S}_y^{\prime -1}\left(\rho\right)\right].$$

It is also easily seen that

$$R_{\beta}^{2}-R_{0}^{2}=\frac{a_{n}s^{2}\left(1-R_{0}^{2}\right)}{1+a_{n}s^{2}}\geq 0,$$

where

$$a_{n}=\frac{Tr\left[\mathbf{S}_{y}^{-1}\left(\rho\right)\mathbf{S}_{x}^{-1}\left(\lambda\right)\mathbf{S}_{x}^{\prime -1}\left(\lambda\right)\mathbf{S}_{y}^{\prime -1}\left(\rho\right)\right]}{Tr\left[\mathbf{S}_{y}^{-1}\left(\rho\right)\mathbf{S}_{y}^{\prime -1}\left(\rho\right)\right]}>0,\ s^{2}=\frac{\beta^{2}\sigma_{v}^{2}}{\sigma_{\varepsilon}^{2}}\geq 0,$$

and note that s^2 is the signal-to-noise ratio. Since $a_n s^2 \geq 0$, we have $R_{\beta}^2 \geq R_0^2$, with equality holding if and only if $\beta = 0$. Therefore, given the values of $\mathbf{W}_y$ and ρ we can only control the value of $R_{\beta}^2 - R_0^2$. Since we are interested in the effects of changes in ρ and δ on the property of GMM and BMM estimators, without loss of generality we set the true parameter values to $\sigma_{\varepsilon,0}^2 = 1$, $\lambda_0 = 0.75$, $\alpha_0 = 1$ and $\beta_0 = 1$. The value of σ_v^2 is chosen to ensure that $R_{\beta}^2 = R_0^2 + 0.1$. This is achieved by setting σ_v^2 such that

$$\frac{\beta^{2}\sigma_{v}^{2}}{\sigma_{\varepsilon}^{2}}\frac{Tr\left[\mathbf{S}_{y}^{-1}\left(\rho\right)\mathbf{S}_{x}^{-1}\left(\lambda\right)\mathbf{S}_{x}^{\prime -1}\left(\lambda\right)\mathbf{S}_{y}^{\prime -1}\left(\rho\right)\right]}{Tr\left[\mathbf{S}_{y}^{-1}\left(\rho\right)\mathbf{S}_{y}^{\prime -1}\left(\rho\right)\right]}=\frac{0.1}{0.9-R_{0}^{2}},$$

or equivalently,

$$\sigma_{v}^{2}=\left(\frac{0.1}{0.9-R_{0}^{2}}\right)\frac{\sigma_{\varepsilon}^{2}}{\beta^{2}a_{n}}, \tag{48}$$

with the value of ρ chosen so that $R_0^2 < 0.9$.

Turning to the specifications of the spatial weights matrices, we consider the case where $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$ in the main text and relegate the results for $\mathbf{W}_x \neq \mathbf{W}_y$ to a Monte Carlo supplement which is available on request. The spatial weights matrix $\mathbf{W}$,

$$\mathbf{W}=(w_{ij})_{n\times n}=\begin{pmatrix}0 & \mathbf{w}_{12}^{\prime}\\ \mathbf{w}_{21} & \mathbf{W}_{22}\end{pmatrix},$$

is generated as follows: We assume, without loss of generality, that the first unit of the network is δ-dominant and the rest are non-dominant. Specifically, the first $\lfloor n^{\delta}\rfloor$ elements of the $(n-1)\times 1$ column vector $\mathbf{w}_{21}$ are drawn from $IIDU(0,1)$ and the rest are set to zero, where $\lfloor . \rfloor$ is the integer part operator. In this way, the sum of the first column of $\mathbf{W}$ expands with n at the rate of δ, i.e., $\sum_{i=1}^{n} w_{i1} = O\left(n^{\delta}\right)$. The first 8 elements of the $1\times(n-1)$ row vector $\mathbf{w}_{12}^{\prime}$ are set to one and the remaining elements to zero. $\mathbf{W}_{22}$ is a standard $(n-1)\times(n-1)$ spatial matrix with 8 connections (4-ahead-and-4-behind with equal weights), namely, $w_{i,j} = 0.125$ for $j = i-4, \ldots, i-1, i+1, \ldots, i+4$, and $w_{i,j} = 0$ otherwise. By construction, $\mathbf{W}_{22}$ is uniformly bounded in both row and column norms, namely, $\left\|\mathbf{W}_{22}\right\|_1 = O(1)$ and $\left\|\mathbf{W}_{22}\right\|_{\infty} = O(1)$. Finally, $\mathbf{W}$ is standardized so that each row sums to one.

We consider a number of different values of δ and ρ_0: $\delta = 0,\ 0.25,\ 0.50,\ 0.75,\ 0.95,\ 1$, and $\rho_0 = 0.2,\ 0.5,\ 0.75$;[10] and experiment with four sample sizes: $n = 100,\ 300,\ 500$, and

[10] The values of R_0^2 for different ρ and δ are reported in Table 46 in the Monte Carlo supplement (available

1,000. We include $\delta = 1$ in our experiments in order to see if the GMM and BMM estimators break down when $\delta = 1$, as predicted by the theory, and to see how the two estimators perform as δ approaches unity. The number of replications is set to 2,000, per experiment. We report results for both GMM and BMM estimates. The BMM estimator is computed by (39), and the GMM estimator is obtained with the optimal weight matrices for the linear and the quadratic moment conditions.[11] More specifically, the GMM estimator is computed in two steps: In the first step, the GMM estimates are computed with equal weights using $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr\left(\mathbf{W}^2\right)\mathbf{I}_n$, and $\mathbf{Z} = \left(\mathbf{1}_n, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x}\right)$. In the second step, we re-estimate with the optimal GMM weights using the best IV and best quadratic matrices evaluated at the first-step estimates, namely using $\tilde{\mathbf{Z}}^* = \left(\tilde{\mathbf{G}}\mathbf{x}\tilde{\alpha}, \tilde{\mathbf{G}}\mathbf{x}\tilde{\boldsymbol{\beta}}, \mathbf{1}_n, \mathbf{x}\right)$ and $\tilde{\mathbf{B}}^* = \tilde{\mathbf{G}} - n^{-1}Tr\left(\tilde{\mathbf{G}}\right)\mathbf{I}_n$, where $\tilde{\mathbf{G}} = \mathbf{G}\left(\tilde{\rho}\right)$, and $\tilde{\boldsymbol{\psi}} = \left(\tilde{\rho}, \tilde{\alpha}, \tilde{\beta}\right)'$ denote the first-step GMM estimates.

Tables 1a–2b summarize the results of the GMM and BMM estimators for the experiments with Gaussian errors, and Tables 3a–4b give the results for non-Gaussian errors. For each experiment, we report bias, root mean square error (RMSE), size, and power of both estimators for ρ and β. The estimates of the intercept term are omitted in order to save space. In addition, Figures 1a–2b plot the empirical power functions for ρ and β in the case of $\rho_0 = 0.5$ and $\beta_0 = 1$ for $\delta = 0$, 0.25, 0.75, 0.95, and $n = 100$ and 300, when the errors are non-Gaussian.[12]

Let us begin by examining the bias and RMSE results. We first observe that both GMM and BMM estimators display declining bias and RMSE as the sample size increases. On the whole, the bias and RMSE are very small even when $n = 100$, irrespective of the magnitude of the spatial autoregressive parameter, ρ. This result is in line with our theoretical finding that both estimators are consistent if $\delta < 1$. However, as the value of δ approaches one, we see a substantial increase in RMSE for both estimators. The two estimators perform similarly in terms of RMSE when $n > 300$, although the BMM estimator of β has smaller RMSE than the GMM estimator when $n = 100$, despite being more biased. The performance of the two estimators are even closer when we consider ρ, giving a very similar RMSEs for all sample sizes under consideration. Finally, the bias and RMSE of both methods are quite robust to non-Gaussian errors, as can be seen from Tables 3a to 4a.

We now turn to size and power properties of the BMM and GMM estimators. As can be seen from Table 1b, overall the tests of ρ have empirical size close to the nominal size of 5% when $\delta \leq 0.75$. This is true for both estimators. When the sample size is small ($n = 100$), the GMM estimator slightly over-rejects the null if the degree of spatial autocorrelation is high ($\rho_0 = 0.75$), and the size distortion becomes more severe as ρ_0 is increased towards unity. In

upon request). Note that $R_0^2 < 0.9$ holds when $\rho \leq 0.75$. When $\beta \neq 0$, we set $R_\beta^2 = R_0^2 + 0.1$. We have also examined the estimation of SAR models without exogenous regressors ($\beta = 0$). The results are also presented in the Monte Carlo supplement.

[11] We also consider GMM estimators using other instruments and quadratic matrices. The results are presented in the Monte Carlo supplement.

[12] More power function plots for other values of ρ_0, δ and n are documented in the Monte Carlo supplement.

comparison, the BMM estimator has the correct empirical size even when the sample size is small and ρ_0 is close to unity when $\delta \leq 0.75$. As the sample size becomes larger ($n \geq 300$), both estimators have the correct size and reasonable power for all values of ρ_0 if $\delta \leq 0.75$. These results suggest that the condition $\delta < 1/2$ assumed in this paper might be too conservative, and whilst sufficient it might not be necessary. Turning to size and power of the tests for β, summarized in Table 2b, we note that both estimators perform well, yielding the correct size and high power, and their performance is overall better as compared the results we obtain for ρ. Finally, these findings seem to be quite robust to non-Gaussian errors.

Figures 1a and 1b display the power functions for ρ when $\rho_0 = 0.5$ for $n = 100$ and 300, respectively. Overall, the tests of $\rho = \rho_0$ based on GMM and BMM estimators have similarly good power when $\delta \leq 0.5$. As δ moves towards one, the tests based on both estimators tend to over-reject the null. The over-rejection is more severe for the GMM estimator than the BMM estimator. For example, as shown in Figure 1a, when $\delta = 0.95$ and $n = 100$ the rejection frequency of the GMM estimator under the null is 25.8% as compared to 14.6% for the BMM estimator. A comparison of Figures 1a and 1b reveals that when $\delta \leq 0.75$ the size distortion is reduced as n expands from 100 to 300, but the over-rejection does not disappear with increasing sample size when $\delta = 0.95$. These findings are in line with our theoretical results.[13] We proceed with Figures 2a and 2b, which show the power functions for β when $\beta_0 = 1$ for $n = 100$ and 300, respectively. We see at once that the power curves for both estimators are very close. We also note that the over-rejection is less of a problem for the estimators of β than for ρ. The power is relatively low when $n = 100$ but rises notably as n increases to 300.

7 Empirical application to US sectoral prices

In earlier studies Acemoglu et al. (2012) and Pesaran and Yang (2019), using US input-output tables, find that δ for the US production network lies between 0.72 and 0.82, and accordingly the standard assumption in the spatial econometrics literature that presumes all units are non-dominant is violated. In what follows we first extend the closed economy multi-sectoral model in Pesaran and Yang (2019) to a small open economy in which production also requires imported intermediate inputs (raw materials). We then apply the GMM and BMM estimation techniques to investigate the degree of interdependence in sectoral price changes in the US economy.

For simplicity, we assume that there is only one type of imported intermediate good, whose quantity demanded for production by sector i at time t is denoted by m_{it}. Each sector i at

[13] Similar findings hold for different values of ρ_0 whether the errors are Gaussian or non-Gaussian, as can be seen from the power plots in the Monte Carlo supplement (available upon request).

time t produces output, q_{it}, by the following Cobb-Douglas production technology:

$$q_{it} = e^{\alpha u_{it}} l_{it}^{\alpha} m_{it}^{\vartheta} \prod_{j=1}^{n} q_{ij,t}^{(1-\alpha-\vartheta)w_{ij}}, \quad \text{for} \quad i = 1, 2, \ldots, n, \tag{49}$$

where l_{it} is the labor input, $q_{ij,t}$ is the amount of output of sector j used by sector i, u_{it} is the productivity shock that consists of two components: $u_{it} = \gamma_i f_t + v_{it}$, where v_{it} is a sector-specific shock, and f_t is a common factor with heterogeneous factor loadings, γ_i, for $i = 1, 2, \ldots, n$, The parameter α represents the share of labor, ϑ represents the share of imported intermediate goods, and w_{ij} is the share of sector j's output in the total domestic intermediate input use by sector i.

The representative household is assumed to have Cobb-Douglas preferences over n goods:

$$u\left(c_{1t}, c_{2t}, \ldots, c_{nt}\right) = A \prod_{i=1}^{n} c_{it}^{1/n},\ A > 0. \tag{50}$$

where c_{it} is the quantity consumed of good i. Furthermore, the household is endowed with l_t unit of labor, supplied inelastically at wage rate Wage_t. In equilibrium, the commodity markets clear,

$$c_{it} = q_{it} - \sum_{j=1}^{n} q_{ji,t} - q_{x,it}, \quad \text{for} \quad i = 1, 2, \ldots, n,$$

where $q_{x,it}$ is the quantity exported of good i; the labor market clears, $l_t = \sum_{i=1}^{n} l_{it}$; and trade is balanced, $P_{m,t} \sum_{i=1}^{n} m_{it} = \sum_{i=1}^{n} P_{it} q_{x,it}$, where P_{it} denotes the price of good i, and $P_{m,t}$ denotes the exogenous world price of the imported intermediate good.

Given prices $\{P_{1t}, P_{2t}, \ldots, P_{nt}, P_{m,t}, \text{Wage}_t\}$, the profit-maximization problem of sector i, for $i = 1, 2, \ldots, n$, is given by

$$\max_{q_{ij,t}, l_{it}, m_{it}} P_{it} e^{\alpha u_{it}} l_{it}^{\alpha} m_{it}^{\vartheta} \prod_{j=1}^{n} q_{ij,t}^{(1-\alpha-\vartheta)w_{ij}} - \text{Wage}_t \times l_{it} - P_{m,t} m_{it} - \sum_{j=1}^{n} P_{jt} q_{ij,t}.$$

The first-order conditions with respect to $q_{ij,t}$, l_{it}, and m_{it} imply that

$$q_{ij,t} = \frac{(1-\alpha-\vartheta)\, w_{ij} P_{it} q_{it}}{P_{jt}}, \quad l_{it} = \frac{\alpha P_{it} q_{it}}{\text{Wage}_t}, \quad m_{it} = \frac{\vartheta P_{it} q_{it}}{P_{m,t}}. \tag{51}$$

Substituting (51) into (49) and after some simplifications yields

$$p_{it} = \rho \sum_{j=1}^{n} w_{ij} p_{jt} + \alpha \omega_t + \vartheta p_{m,t} - b_i - \alpha\left(\gamma_i f_t + v_{it}\right), \text{ for } i = 1, 2, ..., n, \tag{52}$$

where $\rho = (1-\alpha-\vartheta)$, $p_{it} = \log\left(P_{it}\right)$, $\omega_t = \log\left(\text{Wage}_t\right)$, $p_{m,t} = \log\left(P_{m,t}\right)$, and $b_i = \alpha \log(\alpha) + \vartheta \log(\vartheta) + \rho \log(1-\alpha-\vartheta) + \rho \sum_{j=1}^{n} w_{ij} \log(w_{ij})$.

The system of price equations in (52) is in the form of a panel SAR model with fixed effects, observed (ω_t and $p_{m,t}$) and unobserved common factor (f_t). To transform these equations into a SAR model in observables we take first differences[14]

$$\Delta p_{it} = \rho \sum_{j=1}^{n} w_{ij} \Delta p_{jt} + \alpha \Delta \omega_t + \vartheta \Delta p_{m,t} - \alpha \left(\gamma_i \Delta f_t + \Delta v_{it} \right), \text{ for } t = 1, 2, \ldots, T, \tag{53}$$

and consider time averages computed over the sample period $t = 1, 2, \ldots, T$ to obtain

$$\overline{\Delta p}_i = \rho \sum_{j=1}^{n} w_{ij} \overline{\Delta p}_j + \alpha \overline{\Delta \omega} + \vartheta \overline{\Delta p}_m - \alpha \left(\gamma_i \overline{\Delta f} + \overline{\Delta v}_i \right), \tag{54}$$

where $\overline{\Delta p}_i = \frac{1}{T} \sum_{t=1}^{T} \Delta p_{it}$, $\overline{\Delta \omega} = \frac{1}{T} \sum_{t=1}^{T} \Delta \omega_t$, $\overline{\Delta p}_m = \frac{1}{T} \sum_{t=1}^{T} \Delta p_{m,t}$, $\overline{\Delta f} = \frac{1}{T} \sum_{t=1}^{T} \Delta f_t$, and $\overline{\Delta v}_i = \frac{1}{T} \sum_{t=1}^{T} \Delta v_{it}$. For a given sample period $\overline{\Delta \omega}$, $\overline{\Delta p}_m$, and $\overline{\Delta f}$ are fixed, and only cross section variations are relevant for estimation of ρ. We also assume that the factor loadings follow the random coefficient model $\gamma_i = \gamma_0 + \eta_i$, where $\eta_i \sim IID(0, \sigma_\eta^2)$, for $i = 1, 2, ..., n$. Using this result in (54), we now have

$$\overline{\Delta p}_i = a + \rho \sum_{j=1}^{n} w_{ij} \overline{\Delta p}_j + \varepsilon_i, \tag{55}$$

where

$$a = \alpha \overline{\Delta \omega} + \vartheta \overline{\Delta p}_m - \alpha \gamma_0 \overline{\Delta f}, \text{ and } \varepsilon_i = -\alpha \left(\overline{\Delta v}_i + \overline{\Delta f} \eta_i \right).$$

The SAR model in the rate of price changes, (55), can now be estimated by the methods of GMM and BMM. The parameter of interest is the spatial coefficient, ρ, which can be interpreted as capital's share of output. The $n \times n$ matrix $\mathbf{W} = (w_{ij})$ that summarizes the input-output relations corresponds to the spatial weights matrix.

The spatial weights matrix, $\mathbf{W}$, is constructed from the input-output tables at the most disaggregated level obtained from the website of the Bureau of Economic Analysis (BEA). These tables cover around 400 industries and are compiled by the BEA every five years. Specifically, $\mathbf{W}$ is a commodity-by-commodity direct requirements matrix, of which the $(i, j)^{th}$ entry represents the expense on commodity j per dollar of production of commodity i.[15] The commodity-by-commodity direct requirements (**DR**) tables are derived from the commodity-by-commodity total requirements (**TR**) tables by the following formula: $\mathbf{DR} = (\mathbf{TR} - \mathbf{I})(\mathbf{TR})^{-1}$, where $\mathbf{I}$ is the identity matrix of conformable dimension. The $\mathbf{W}$ matrix is taken as the transpose of **DR** and standardized so that the sum of intermediate input shares (the row sum of $\mathbf{W}$) equals unity for every sector. Since the vast majority of the elements in $\mathbf{W}$ are rather small numbers, in order to reduce noise in the system we construct a robust weights matrix by setting each element of $\mathbf{W}$ to one if it is greater than or equal to a given threshold value ϵ_w $(0 < \epsilon_w < 1)$, and to zero otherwise. Then the sectors with zero row sums are dropped and

[14] This paper focuses on cross section SAR models. The estimation of the panel data model given by (52) is beyond the scope of the current paper.

[15] The words commodity and sector are used interchangeably to convey the same meaning throughout this paper.

the matrix is row-standardized so that each row sums to one. The resulting matrix is denoted by $\tilde{\mathbf{W}}(\epsilon_w)$. The sector-specific price index at annual frequency are obtained from the BEA's gross domestic product by industry accounts. The annual rates of price changes are computed over the period 1998–2015, and they are matched to the sectors in the input-output tables using the BEA industry codes.

Given the time range of the price data, we consider two versions of $\tilde{\mathbf{W}}$ constructed from the input-output tables for the years 2002 and 2007, denoted by $\tilde{\mathbf{W}}_{2002}$ and $\tilde{\mathbf{W}}_{2007}$, respectively. In particular, we consider a cut-off value $\epsilon_w = 10\%$, which means that for any given sector only important suppliers that contribute at least 10% of the total input purchases are taken into account.[16]

We begin by examining the δ-dominance of the production networks for the years 2002 and 2007 by applying the extremum estimator developed by Pesaran and Yang (2019) to the outdegrees of the filtered input-output matrices $\tilde{\mathbf{W}}_{2002}(0.1)$ and $\tilde{\mathbf{W}}_{2007}(0.1)$. Table 5 reports the estimates of δ for the top five most important sectors for these weight matrices. The results show that the highest degree of dominance, $\widehat{\delta}_{(1)}$, lies between 0.71 and 0.85, and are not close to unity. Therefore, our proof of consistency of the GMM and BMM estimators of the spatial parameter applies to this empirical application. But for valid inference our proofs require $\delta < 1/2$, and special care must be exercised when carrying out inference on ρ in the present empirical application. Although, as noted above, our Monte Carlo experiments suggest that the degree of over-rejection of tests based on the BMM estimator of ρ is relatively low so long as δ is not too close to unity, and inference based on the BMM estimators seems to be acceptable for values of δ around 0.75.

Turning to the sectoral price changes, to allow for the possibility of structural breaks due to the 2007–2008 financial crisis, we consider two sub-samples: the pre-financial crises (1998–2006) and the post-financial crises (2007–2015) periods. The weights matrix $\tilde{\mathbf{W}}_{2002}(0.1)$ is used for the first sub-sample, while $\tilde{\mathbf{W}}_{2007}(0.1)$ is used for the second sub-sample. The BMM estimates are computed by (39). The GMM estimates are obtained in two steps: In the first step, we compute initial consistent estimate, $\tilde{\rho}$, by (30) using two equally weighted quadratic moments with $\mathbf{B}_1 = \tilde{\mathbf{W}}$ and $\mathbf{B}_2 = \tilde{\mathbf{W}}^2 - n^{-1}Tr\left(\tilde{\mathbf{W}}^2\right)\mathbf{I}_n$.[17] In the second step, we re-estimate the model using the best quadratic matrix, $\tilde{\mathbf{B}}^* = \tilde{\mathbf{G}} - n^{-1}Tr\left(\tilde{\mathbf{G}}\right)\mathbf{I}_n$, where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\rho})$ is evaluated at the first-step estimate. Table 6 presents the estimation results of model (55). It can be seen that the BMM and GMM estimates are very close and highly significant. The estimated share of capital is around 0.4 for the first sub-sample and 0.3 for the second sub-

[16] Our choice of the 10% threshold for non-zero elements of the weight matrix is in line with the US Regulation SFAS No. 131 that requires public firms to report customers representing more than 10% of their total yearly sales (see Cohen and Frazzini, 2008, p. 1978). The results for other cut-off values of $\epsilon_w = 5\%$ and 7.5%. are provided in the empirical supplement available upon request. Using lower threshold values tend to yield higher estimates of ρ.

[17] Here we denote $\tilde{\mathbf{W}}_{2002}(0.1)$ and $\tilde{\mathbf{W}}_{2007}(0.1)$ simply as $\tilde{\mathbf{W}}$ to simplify the notations.

Table 5: Estimates of the degree of dominance, δ, of the top five pervasive sectors using US input-output tables

	Input-output table for 2002		Input-output table for 2007	
	$\mathbf{W}_{2002}$	$\tilde{\mathbf{W}}_{2002}(0.1)$	$\mathbf{W}_{2007}$	$\tilde{\mathbf{W}}_{2007}(0.1)$
$\widehat{\delta}_{(1)}$	0.778	0.851	0.724	0.705
$\widehat{\delta}_{(2)}$	0.759	0.796	0.651	0.703
$\widehat{\delta}_{(3)}$	0.597	0.642	0.608	0.695
$\widehat{\delta}_{(4)}$	0.550	0.422	0.592	0.565
$\widehat{\delta}_{(5)}$	0.546	0.402	0.553	0.491
n	313 [301]	286 [114]	384 [364]	350 [140]
n^*	69,268 (70.70%)	581 (0.71%)	107,619 (72.98%)	616 (0.50%)

Notes: $\tilde{\mathbf{W}}(\epsilon_w = 0.1)$ denotes a filtered version of $\mathbf{W} = (w_{ij})$, defined by $\tilde{\mathbf{W}}(\epsilon_w) = (\tilde{w}_{ij}(\epsilon_w))$, where $\tilde{w}_{ij}(\epsilon_w)$ is a row-standardized version of $w_{ij}^*(\epsilon_w)$ defined by $w_{ij}^*(\epsilon_w) = w_{ij} I(w_{ij} \geq \epsilon_w)$, where $I(A)$ is an indicator variable which takes the value of unity if A holds and zero otherwise. We set $\epsilon_w = 10\%$, and report $\hat{\delta}_{(1)} > \hat{\delta}_{(2)} > \ldots > \hat{\delta}_{(5)}$; the five largest estimates of δ corresponding to the outdegrees of $\mathbf{W}$ and $\tilde{\mathbf{W}}(0.1)$, for the years 2002 and 2007. n is the total number of sectors with non-zero total demands (indegrees). The numbers in square brackets are the numbers of sectors with non-zero outdegrees. Note that a few sectors were dropped when constructing $\tilde{\mathbf{W}}$ from $\mathbf{W}$, since their total demands become zero. n^* is the number of non-zero elements. The percentages of non-zero elements are in parentheses.

sample. Although these estimates are not very precise, they match reasonably well with the commonly documented values of share of capital in the literature.[18]

8 Concluding remarks

A crucial assumption in the spatial econometrics literature requires that the weights (connections) matrix is uniformly bounded in both row and column sums. This assumption excludes the existence of dominant units in the network and is too restrictive for many applications. The current paper relaxes this assumption and allows the centrality of the connections matrix to rise at the rate of δ with n, as compared to the value of $\delta = 0$ assumed in the spatial literature. We also establish the asymptotic distribution of the GMM estimator due to Lee (2007) for this more general settings, and propose a new BMM estimator which is simple to compute and has better small sample properties as compared to the best GMM estimator when the degree of centrality of the weights matrix, δ, is relatively large. Both estimators are shown to be consistent and normally distributed if the maximum absolute column sum of

[18] The most commonly used value in calibration exercises is 0.36 (Hansen and Wright, 1998; Danthine et al., 2008). Other frequently used calibration values fall in the range 0.3–0.4. For example, Cooley and Prescott (1995) suggest 0.4; Gollin (2002) recommends a range of 0.23–0.34; Danthine et al. (2008) uses 0.3.

Table 6: Estimation results of the cross-section model (55)

Year	Sub-sample 1998–2006		Sub-sample 2007–2015	
	BMM	GMM	BMM	GMM
$\hat{\rho}$ [Share of capital]	$0.397^{\dagger}$	$0.395^{\dagger}$	$0.287^{\dagger}$	$0.274^{\dagger}$
	(0.106)	(0.107)	(0.072)	(0.073)
$\hat{\sigma}_{\eta}^2$ [Error variance]	7.728	7.815	2.564	2.599
R^2	0.219	0.217	0.159	0.148
Weights matrix	$\tilde{\mathbf{W}}_{2002}\,(0.1)$		$\tilde{\mathbf{W}}_{2007}\,(0.1)$	
n [Number of sectors]	286		350	
T [Number of time periods]	9		9	

Notes: All estimations include an intercept (not shown here). The BMM estimates are computed by (39). The GMM estimates are computed by a two-step procedure following (30) using the best quadratic moment evaluated at the first-step estimate. R^2 is computed by (46). The spatial weights matrices are constructed with a threshold value of $\epsilon_w = 10\%$. $\tilde{\mathbf{W}}_{2002}\,(0.1)$ is used in the estimation over the period 1998–2006; $\tilde{\mathbf{W}}_{2007}\,(0.1)$ is used in the estimation over the period 2007–2015. Standard errors are in parentheses. † indicates significance at 1% level.

the interaction matrix does not increase too fast as n grows. For consistent estimation it is required that $\delta < 1$, and for the validity of the asymptotic distribution we need $\delta < 1/2$. But the extensive Monte Carlo experiments reported in the paper and in the supplement suggest that GMM and BMM estimators could perform reasonably well if $\delta \leq 0.75$. Thus, it might be conjectured that the sufficient condition of $\delta < 1/2$ might not be necessary for the validity of the asymptotic distribution of GMM and BMM estimators. Further analysis is required if $\delta > 1/2$. Such an analysis is beyond the scope of the present paper.

Table 1a: Bias and RMSE of the GMM and BMM estimators of ρ for the experiments with Gaussian errors

	GMM								BMM							
	Bias($\times$100)				RMSE($\times$100)				Bias($\times$100)				RMSE($\times$100)			
$\delta\backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	-6.76	-2.14	-1.37	-0.68	20.19	10.35	7.84	5.39	-9.06	-2.72	-1.67	-0.82	20.78	10.44	7.87	5.40
0.25	-6.73	-2.13	-1.36	-0.68	20.12	10.34	7.83	5.39	-8.99	-2.71	-1.67	-0.83	20.67	10.42	7.86	5.40
0.50	-6.86	-2.14	-1.35	-0.68	20.07	10.29	7.82	5.39	-9.06	-2.72	-1.66	-0.82	20.55	10.36	7.85	5.40
0.75	-6.58	-2.08	-1.27	-0.67	21.02	10.74	8.10	5.48	-9.45	-2.81	-1.66	-0.83	21.22	10.71	8.09	5.48
0.95	-5.63	-1.72	-1.30	-0.75	27.05	13.42	9.86	6.69	-11.57	-3.55	-2.19	-1.12	25.35	12.89	9.67	6.58
1.00	-4.85	-1.79	-1.27	-0.67	31.64	16.31	12.14	8.07	-13.10	-4.22	-2.65	-1.26	28.49	15.15	11.56	7.84
	$\rho_0 = 0.5$															
0.00	-5.77	-1.85	-1.13	-0.58	16.08	7.82	5.81	3.98	-8.06	-2.38	-1.41	-0.71	16.70	7.92	5.85	3.99
0.25	-5.76	-1.85	-1.13	-0.59	16.04	7.81	5.81	3.98	-8.02	-2.38	-1.41	-0.72	16.63	7.91	5.85	3.99
0.50	-5.84	-1.86	-1.13	-0.59	16.06	7.77	5.80	3.99	-8.06	-2.39	-1.42	-0.72	16.55	7.87	5.84	4.00
0.75	-5.74	-1.93	-1.12	-0.61	17.40	8.34	6.15	4.13	-8.77	-2.61	-1.49	-0.77	17.54	8.35	6.16	4.14
0.95	-5.41	-1.67	-1.18	-0.73	24.50	11.55	8.16	5.38	-12.33	-3.82	-2.24	-1.15	23.00	10.84	7.91	5.28
1.00	-4.20	-1.30	-0.85	-0.56	29.46	15.48	11.14	7.08	-14.77	-4.82	-2.92	-1.41	26.74	13.49	10.08	6.77
	$\rho_0 = 0.75$															
0.00	-4.02	-1.28	-0.76	-0.40	10.81	4.85	3.52	2.40	-5.98	-1.70	-0.98	-0.50	11.27	4.95	3.55	2.41
0.25	-4.02	-1.29	-0.76	-0.40	10.82	4.85	3.52	2.40	-5.97	-1.69	-0.98	-0.50	11.25	4.94	3.55	2.41
0.50	-4.02	-1.28	-0.76	-0.41	10.90	4.81	3.52	2.40	-5.96	-1.70	-0.99	-0.51	11.19	4.90	3.55	2.42
0.75	-4.08	-1.42	-0.80	-0.45	12.42	5.39	3.84	2.55	-6.82	-1.98	-1.10	-0.58	12.31	5.39	3.86	2.56
0.95	-5.30	-1.15	-0.77	-0.55	21.46	9.18	6.11	3.73	-12.10	-3.60	-2.01	-1.02	19.48	7.99	5.56	3.58
1.00	-6.07	-0.49	0.23	0.06	27.76	14.30	10.97	6.85	-16.37	-5.57	-3.35	-1.66	24.53	11.48	8.36	5.55

Notes: The data generating process (DGP) is given by (44) and (45), where the errors are generated as $IIDN\left(0,1\right)$. The true parameter values are $\alpha_0 = 1$, $\beta_0 = 1$, $\lambda_0 = 0.75$, $\sigma_{\varepsilon,0} = 1$, and $\sigma_{\nu,0}$ is computed by (48). $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. The first unit is δ-dominant, and the rest of the units are non-dominant. The number of replications is $2,000$. The BMM estimator is given by (39). The GMM estimator is computed in two steps: In the first step, we obtain preliminary GMM estimates, $\tilde{\boldsymbol{\psi}} = \left(\tilde{\rho}, \tilde{\alpha}, \tilde{\beta}\right)'$, following (24), where $\mathbf{Z} = \left(\mathbf{1}_n, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x}\right)$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr\left(\mathbf{W}^2\right)\mathbf{I}_n$, and $\mathbf{A} = \mathbf{I}_n$. In the second step, we set $\tilde{\mathbf{Z}}^* = \left(\tilde{\mathbf{G}}\mathbf{x}\tilde{\alpha}, \tilde{\mathbf{G}}\mathbf{x}\tilde{\beta}, \mathbf{1}_n, \mathbf{x}\right)$ and $\tilde{\mathbf{B}}^* = \tilde{\mathbf{G}} - n^{-1}Tr\left(\tilde{\mathbf{G}}\right)\mathbf{I}_n$, where $\tilde{\mathbf{G}} = \mathbf{W}\left(\mathbf{I}_n - \tilde{\rho}\mathbf{W}\right)^{-1}$, and compute the optimal GMM estimates by (32).

Table 1b: Size and power of the GMM and BMM estimators of ρ for the experiments with Gaussian errors

	GMM								BMM							
	Size(×100)				Power(×100)				Size(×100)				Power(×100)			
$\delta\backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	7.40	5.70	5.75	4.95	9.55	16.35	24.25	44.25	6.75	5.45	5.70	5.15	7.15	15.00	22.95	43.05
0.25	7.10	5.45	5.80	5.15	9.30	16.65	23.95	43.95	6.95	5.35	5.75	5.25	6.80	14.80	22.65	42.70
0.50	7.10	5.95	5.75	5.35	9.35	17.05	24.15	44.25	6.55	5.70	6.00	5.35	6.95	15.80	22.95	43.35
0.75	8.75	6.40	6.10	5.30	10.95	17.30	25.85	44.50	7.20	5.65	5.85	5.45	7.20	15.10	23.45	43.10
0.95	17.55	13.15	10.95	10.55	20.20	22.70	28.20	42.05	12.30	10.75	9.75	10.10	11.80	17.85	24.05	39.45
1.00	23.70	19.95	17.35	15.65	25.45	26.30	29.90	41.90	16.05	15.90	15.35	13.90	15.20	21.20	25.25	39.35
	$\rho_0 = 0.5$															
0.00	8.20	5.70	5.60	5.15	13.95	27.30	39.65	67.75	6.95	5.65	5.65	5.30	9.60	25.10	37.60	66.85
0.25	8.10	5.80	5.50	5.20	13.70	27.25	39.65	67.90	6.95	5.55	5.75	5.20	9.35	24.90	37.55	66.80
0.50	8.05	5.90	6.05	5.25	13.90	26.55	40.15	68.30	6.60	5.75	5.85	5.10	9.80	24.20	37.80	67.05
0.75	10.60	6.50	6.15	5.45	15.50	26.65	40.75	66.15	7.40	6.35	5.65	5.40	10.10	24.25	37.60	64.50
0.95	23.70	16.65	13.30	11.90	26.40	31.70	40.40	58.65	14.60	12.40	10.85	10.90	13.55	24.60	34.05	55.80
1.00	32.70	27.90	23.30	21.10	33.45	35.25	42.20	55.75	20.35	19.85	19.25	19.30	18.35	27.65	34.75	52.50
	$\rho_0 = 0.75$															
0.00	9.65	5.75	5.60	5.30	25.75	55.30	75.20	96.20	7.00	5.80	5.35	5.05	17.85	51.75	73.85	95.90
0.25	9.80	5.60	5.55	5.05	25.80	55.45	75.35	96.35	6.65	5.85	5.40	4.90	17.90	52.05	74.00	96.05
0.50	10.30	6.00	5.85	5.15	25.35	54.80	75.40	96.50	6.70	5.70	5.45	5.05	17.65	51.95	74.05	96.00
0.75	14.10	8.00	6.85	5.30	27.25	52.40	73.60	94.95	8.55	7.05	6.00	5.20	17.35	48.25	71.70	94.45
0.95	37.25	25.20	20.45	15.65	38.55	48.05	64.55	85.50	19.40	15.45	14.15	12.70	18.95	38.95	58.80	84.15
1.00	54.15	45.50	44.45	39.05	46.45	49.15	60.25	76.00	29.15	32.35	31.95	32.00	22.70	37.65	50.95	71.45

Notes: The power is calculated at $\rho_0 - 0.1$, where ρ_0 denotes the true value. See also the notes to Table 1a.

Table 2a: Bias and RMSE of the GMM and BMM estimators of β for the experiments with Gaussian errors

	GMM								BMM							
	Bias(×100)				RMSE(×100)				Bias(×100)				RMSE(×100)			
$\delta \backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	0.02	0.22	0.03	0.21	38.95	20.72	15.69	10.90	3.45	1.29	0.63	0.50	39.68	20.85	15.78	10.93
0.25	0.05	0.24	0.04	0.21	38.93	20.73	15.69	10.91	3.45	1.30	0.63	0.50	39.65	20.87	15.78	10.94
0.50	0.00	0.25	0.03	0.20	39.11	20.83	15.73	10.94	3.46	1.31	0.63	0.49	39.85	20.95	15.82	10.97
0.75	-0.38	0.08	-0.12	0.16	40.35	21.47	16.15	11.30	3.32	1.26	0.57	0.48	40.90	21.53	16.21	11.33
0.95	-1.67	-0.62	-0.56	0.09	45.88	26.05	20.20	15.40	2.06	0.91	0.34	0.51	46.17	25.94	20.17	15.37
1.00	-2.59	-1.16	-0.98	-0.14	50.19	30.56	25.28	20.64	1.06	0.18	-0.18	0.25	50.78	30.60	25.36	20.69
	$\rho_0 = 0.5$															
0.00	0.34	0.42	0.10	0.30	46.10	24.32	18.39	12.77	4.80	1.79	0.87	0.67	47.19	24.52	18.51	12.82
0.25	0.36	0.45	0.11	0.31	46.04	24.33	18.39	12.77	4.81	1.81	0.88	0.67	47.13	24.54	18.52	12.82
0.50	0.24	0.47	0.10	0.29	46.42	24.58	18.53	12.88	4.81	1.83	0.88	0.66	47.53	24.78	18.66	12.93
0.75	-0.11	0.30	-0.08	0.27	50.36	27.46	20.89	15.06	4.86	1.88	0.83	0.69	51.31	27.60	21.01	15.12
0.95	-1.79	-0.85	-0.88	0.22	69.42	47.37	39.81	34.35	3.40	1.61	0.43	0.90	70.40	47.31	39.82	34.38
1.00	-3.26	-2.18	-2.05	-0.19	82.96	63.84	58.79	54.16	1.85	0.22	-0.69	0.49	85.34	64.33	59.23	54.43
	$\rho_0 = 0.75$															
0.00	0.62	0.57	0.17	0.37	48.45	25.24	19.04	13.20	6.31	2.31	1.15	0.84	49.48	25.46	19.17	13.26
0.25	0.62	0.60	0.18	0.37	48.39	25.25	19.04	13.20	6.34	2.33	1.16	0.84	49.40	25.46	19.18	13.27
0.50	0.46	0.62	0.17	0.36	48.79	25.69	19.33	13.43	6.32	2.37	1.18	0.84	49.82	25.91	19.47	13.49
0.75	0.15	0.54	-0.04	0.38	57.15	32.27	24.97	18.64	6.72	2.63	1.17	0.95	58.05	32.45	25.15	18.75
0.95	-1.11	-1.26	-1.19	0.35	93.61	71.72	62.71	55.97	5.36	2.72	0.74	1.45	96.22	71.39	62.60	56.11
1.00	-3.03	-3.37	-3.61	-0.65	115.50	99.26	94.90	90.60	3.16	0.58	-1.04	0.94	120.65	101.25	96.41	91.57

Notes: The true parameter value is $\beta_0 = 1$. See also the notes to Table 1a.

Table 2b: Size and power of the GMM and BMM estimators of β for the experiments with Gaussian errors

	GMM								BMM							
	Size($\times$100)				Power($\times$100)				Size($\times$100)				Power($\times$100)			
$\delta\backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	7.20	5.55	4.45	4.25	10.05	16.95	24.75	44.35	7.85	5.45	4.50	4.30	11.55	18.70	26.20	45.20
0.25	7.30	5.60	4.45	4.35	9.90	17.00	25.00	44.10	7.85	5.50	4.50	4.35	11.25	18.80	26.15	44.95
0.50	7.45	5.40	4.35	4.40	9.40	16.80	24.35	44.35	7.95	5.35	4.40	4.50	11.20	18.55	25.85	45.10
0.75	7.15	5.50	4.60	4.40	9.50	17.05	23.15	42.65	7.50	5.25	4.60	4.55	11.40	18.50	24.65	43.50
0.95	7.35	5.00	4.70	4.60	8.10	12.85	16.25	25.65	7.20	5.30	4.70	4.55	9.50	14.20	16.85	26.60
1.00	6.55	5.05	4.70	4.20	7.80	9.75	11.85	15.25	7.00	5.10	4.90	4.15	8.90	10.65	12.55	15.50
	$\rho_0 = 0.5$															
0.00	7.40	5.65	4.45	4.25	9.35	13.50	19.00	34.95	7.80	5.70	4.65	4.30	10.85	14.70	20.95	36.05
0.25	7.35	5.70	4.55	4.30	9.25	13.40	18.85	35.05	7.75	5.70	4.65	4.35	10.65	14.80	20.95	36.05
0.50	7.40	5.70	4.55	4.35	8.85	13.60	19.05	34.75	7.90	5.60	4.60	4.50	10.25	14.60	20.90	35.90
0.75	7.30	5.70	4.55	4.45	8.40	11.85	15.40	27.00	7.65	5.30	4.70	4.50	10.05	13.15	16.80	27.90
0.95	7.30	5.05	5.30	4.65	7.15	7.35	8.25	8.25	7.40	5.40	5.20	4.75	8.40	8.15	8.70	8.65
1.00	6.25	4.90	4.85	4.20	6.65	5.95	5.80	5.75	6.90	5.25	5.30	4.25	7.70	6.65	6.00	6.00
	$\rho_0 = 0.75$															
0.00	7.65	5.90	4.75	4.20	9.10	12.90	18.45	33.20	8.15	5.85	4.80	4.40	10.50	14.10	20.35	34.65
0.25	7.50	5.75	4.80	4.25	9.15	13.10	18.35	33.50	8.20	5.80	4.80	4.40	10.55	14.45	20.30	34.85
0.50	7.70	5.85	4.75	4.25	8.95	13.10	17.80	32.80	8.20	5.65	4.75	4.45	10.45	14.05	19.75	34.00
0.75	7.85	5.85	5.05	4.45	8.50	10.40	12.40	18.50	8.55	5.55	5.05	4.55	10.10	11.70	13.55	19.45
0.95	7.30	5.35	5.50	4.80	7.05	6.55	6.15	6.10	7.80	5.40	5.60	4.55	8.35	7.10	6.70	6.40
1.00	6.20	5.05	4.80	4.20	6.30	4.85	4.80	4.60	7.25	5.35	5.35	4.35	7.80	5.55	5.10	4.75

Notes: The true parameter value is $\beta_0 = 1$ and power is calculated at 0.9. See also the notes to Table 1a.

Table 3a: Bias and RMSE of the GMM and BMM estimators of ρ for the experiments with non-Gaussian errors

	GMM								BMM							
	Bias($\times$100)				RMSE($\times$100)				Bias($\times$100)				RMSE($\times$100)			
$\delta\backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	-6.09	-1.87	-1.18	-0.59	19.86	9.83	7.55	5.28	-8.18	-2.39	-1.47	-0.73	20.28	9.90	7.59	5.30
0.25	-6.03	-1.85	-1.17	-0.59	19.83	9.82	7.55	5.28	-8.10	-2.37	-1.46	-0.73	20.24	9.88	7.59	5.30
0.50	-6.21	-1.89	-1.21	-0.60	19.84	9.94	7.58	5.29	-8.30	-2.42	-1.49	-0.74	20.26	9.99	7.62	5.31
0.75	-5.99	-1.84	-1.25	-0.60	20.96	10.35	7.92	5.49	-8.70	-2.52	-1.60	-0.76	21.18	10.38	7.94	5.50
0.95	-4.80	-1.60	-1.17	-0.59	27.26	13.04	9.70	6.68	-10.82	-3.13	-1.97	-0.96	25.51	12.62	9.52	6.58
1.00	-4.57	-1.37	-1.01	-0.55	31.80	15.63	11.78	8.07	-11.95	-3.50	-2.19	-1.10	28.55	14.55	11.28	7.84
	$\rho_0 = 0.5$															
0.00	-5.28	-1.62	-1.01	-0.50	15.88	7.40	5.61	3.90	-7.36	-2.11	-1.27	-0.63	16.39	7.49	5.66	3.92
0.25	-5.27	-1.61	-1.01	-0.51	15.89	7.39	5.61	3.90	-7.32	-2.10	-1.27	-0.63	16.38	7.48	5.65	3.92
0.50	-5.43	-1.65	-1.03	-0.52	15.96	7.48	5.63	3.91	-7.49	-2.14	-1.29	-0.64	16.43	7.57	5.68	3.94
0.75	-5.33	-1.70	-1.16	-0.56	17.34	7.99	6.06	4.14	-8.14	-2.34	-1.48	-0.71	17.60	8.07	6.10	4.16
0.95	-4.06	-1.55	-1.18	-0.59	24.92	11.14	8.14	5.39	-11.57	-3.45	-2.13	-1.03	22.97	10.65	7.88	5.32
1.00	-4.10	-0.81	-0.75	-0.46	29.49	14.60	10.70	7.07	-13.67	-4.10	-2.53	-1.25	26.63	12.80	9.80	6.73
	$\rho_0 = 0.75$															
0.00	-3.80	-1.17	-0.72	-0.36	10.68	4.61	3.41	2.35	-5.57	-1.55	-0.92	-0.45	11.18	4.69	3.46	2.37
0.25	-3.81	-1.17	-0.72	-0.36	10.71	4.60	3.41	2.35	-5.56	-1.55	-0.92	-0.46	11.19	4.69	3.46	2.37
0.50	-3.87	-1.19	-0.74	-0.37	10.83	4.64	3.42	2.37	-5.65	-1.57	-0.93	-0.46	11.25	4.72	3.47	2.39
0.75	-3.91	-1.30	-0.88	-0.42	12.49	5.15	3.84	2.57	-6.46	-1.82	-1.14	-0.54	12.59	5.24	3.88	2.59
0.95	-4.04	-0.94	-0.87	-0.46	22.04	9.04	6.15	3.80	-11.54	-3.38	-2.04	-0.97	19.40	7.96	5.65	3.67
1.00	-6.09	-0.10	0.22	0.13	27.25	14.06	10.46	6.97	-15.44	-4.97	-3.05	-1.51	24.28	10.80	8.10	5.49

Notes: The DGP is given by (44) and (45), where the errors are generated as $IID\left[\chi^2(2) - 2\right]/2$. See also the notes to Table 1a.

Table 3b: Size and Power of the GMM and BMM estimators of ρ for the experiments with non-Gaussian errors

	GMM								BMM							
	Size(×100)				Power(×100)				Size(×100)				Power(×100)			
$\delta\backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	7.45	4.30	5.30	5.20	9.85	16.30	24.25	43.15	6.60	3.95	5.20	5.20	7.80	14.85	22.70	42.30
0.25	7.45	4.30	5.35	5.25	9.65	16.15	24.20	43.20	6.90	4.00	5.25	5.25	7.65	14.80	22.75	42.35
0.50	7.40	4.45	5.20	5.15	9.85	16.60	24.50	42.90	6.95	4.35	5.25	5.10	7.45	15.05	22.95	42.30
0.75	9.05	5.45	6.15	6.65	11.70	18.45	24.95	42.75	8.15	5.20	6.15	6.40	8.35	16.35	23.60	41.95
0.95	18.70	12.55	12.00	10.00	21.75	23.25	26.85	41.60	13.30	9.85	10.35	9.45	12.95	18.65	23.75	40.00
1.00	24.70	17.55	17.45	16.05	25.45	26.20	29.80	41.35	16.85	13.75	15.35	13.80	16.45	21.25	25.85	38.45
	$\rho_0 = 0.5$															
0.00	7.70	4.40	5.25	5.65	14.00	25.80	39.40	68.05	6.60	4.35	4.90	5.50	8.95	23.80	37.50	67.10
0.25	7.95	4.50	5.25	5.55	14.00	25.95	39.25	68.15	6.85	4.35	4.90	5.50	9.05	23.65	37.50	67.25
0.50	8.00	4.75	5.20	5.80	14.40	25.95	38.95	67.90	6.75	4.45	5.30	5.50	9.15	24.20	37.15	66.85
0.75	11.00	5.75	6.95	6.40	16.95	26.65	38.20	65.85	7.95	5.60	6.45	6.10	9.85	23.75	36.25	64.10
0.95	25.80	16.45	14.50	12.15	28.65	32.15	37.75	57.60	14.55	11.85	12.25	10.45	14.45	23.95	32.65	54.85
1.00	32.90	25.30	23.55	22.15	33.05	34.80	39.80	55.95	20.35	18.45	19.60	19.60	19.05	27.10	33.85	51.80
	$\rho_0 = 0.75$															
0.00	9.55	4.75	4.95	5.75	25.80	56.40	77.25	96.55	6.65	4.30	4.95	5.85	18.40	52.65	75.45	96.20
0.25	9.85	4.75	4.85	5.50	25.85	56.50	77.45	96.45	6.60	4.35	4.75	5.55	18.65	52.65	75.55	96.10
0.50	10.35	4.95	5.25	5.50	26.05	56.50	77.30	96.60	6.95	4.65	4.80	5.65	18.85	53.00	75.60	96.35
0.75	14.85	7.00	7.70	6.80	29.30	54.80	73.50	94.45	8.55	6.40	7.15	6.75	18.55	49.45	70.80	94.25
0.95	40.75	26.00	23.05	17.20	42.00	51.15	62.70	86.15	17.75	14.90	15.50	13.25	19.15	39.20	56.10	84.30
1.00	52.40	43.00	42.95	38.70	45.85	49.25	59.60	75.85	27.40	30.75	31.40	32.10	23.35	37.00	49.45	72.10

Notes: The power is calculated at $\rho_0 - 0.1$, where ρ_0 denotes the true value. See also the notes to Table 3a.

Table 4a: Bias and RMSE of the GMM and BMM estimators of β for the experiments with non-Gaussian errors

	GMM								BMM							
	Bias(×100)				RMSE(×100)				Bias(×100)				RMSE(×100)			
$\delta \backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	1.30	0.43	0.74	0.24	39.71	21.06	15.91	11.32	4.68	1.43	1.30	0.53	40.54	21.26	16.04	11.36
0.25	1.27	0.40	0.73	0.24	39.67	21.05	15.89	11.31	4.63	1.40	1.29	0.53	40.51	21.25	16.02	11.35
0.50	1.29	0.33	0.71	0.22	39.77	21.11	15.96	11.33	4.67	1.36	1.27	0.50	40.60	21.32	16.08	11.38
0.75	0.80	0.37	0.74	0.29	41.20	21.77	16.50	11.79	4.41	1.52	1.35	0.60	41.90	21.95	16.62	11.82
0.95	0.45	-0.01	0.65	0.16	47.26	26.83	20.70	16.07	4.11	1.42	1.45	0.59	47.68	26.97	20.83	16.09
1.00	0.16	-0.08	0.68	0.08	51.74	31.10	25.58	21.66	3.52	1.16	1.42	0.44	52.29	31.33	25.75	21.70
	$\rho_0 = 0.5$															
0.00	1.83	0.65	0.95	0.33	47.05	24.73	18.63	13.23	6.26	1.91	1.66	0.69	48.26	25.02	18.81	13.30
0.25	1.81	0.62	0.94	0.33	47.00	24.72	18.61	13.22	6.22	1.89	1.65	0.69	48.20	25.00	18.79	13.28
0.50	1.84	0.54	0.92	0.30	47.25	24.92	18.77	13.33	6.29	1.85	1.63	0.66	48.48	25.24	18.97	13.39
0.75	1.38	0.66	1.07	0.45	51.53	27.83	21.36	15.75	6.23	2.16	1.87	0.85	52.61	28.17	21.57	15.80
0.95	1.55	0.43	1.56	0.55	71.50	48.66	40.72	35.96	6.55	2.52	2.70	1.19	72.66	49.25	41.18	36.07
1.00	1.48	0.42	2.02	0.48	85.79	64.95	59.45	56.96	5.97	2.34	3.15	1.03	87.77	65.91	60.20	57.22
	$\rho_0 = 0.75$															
0.00	2.33	0.84	1.09	0.40	49.36	25.64	19.22	13.65	7.89	2.43	1.99	0.85	50.77	25.98	19.45	13.72
0.25	2.33	0.83	1.09	0.40	49.30	25.62	19.20	13.63	7.87	2.41	1.98	0.85	50.70	25.95	19.42	13.70
0.50	2.31	0.73	1.07	0.38	49.61	26.00	19.50	13.87	7.99	2.38	1.97	0.82	51.06	26.37	19.74	13.94
0.75	2.19	0.96	1.38	0.62	58.15	32.64	25.54	19.54	8.35	2.91	2.45	1.15	59.66	33.12	25.85	19.61
0.95	3.44	0.91	2.63	0.97	96.55	73.12	63.65	58.72	9.66	4.05	4.26	1.97	99.21	74.27	64.60	58.90
1.00	3.63	0.78	3.13	0.65	120.03	101.13	95.82	95.44	8.90	3.87	5.28	1.87	124.04	103.80	98.10	96.30

Notes: The true parameter value is $\beta_0 = 1$. See also the notes to Table 3a.

Table 4b: Size and power of the GMM and BMM estimators of β for the experiments with non-Gaussian errors

	GMM								BMM							
	Size(×100)				Power(×100)				Size(×100)				Power(×100)			
$\delta\backslash n$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$															
0.00	6.70	5.05	5.60	5.95	10.45	17.05	25.15	44.40	6.85	5.45	5.85	6.15	11.90	18.85	26.65	45.60
0.25	6.70	5.00	5.50	5.90	10.30	16.95	25.15	44.35	6.90	5.35	5.70	6.15	11.80	18.60	26.85	45.60
0.50	6.45	5.55	5.45	5.65	9.65	16.85	26.15	44.25	6.60	5.75	5.60	5.75	11.20	18.35	27.15	45.05
0.75	6.80	5.00	5.05	6.00	10.15	16.00	24.40	41.75	6.80	5.20	5.20	6.00	11.60	17.55	25.80	42.60
0.95	6.70	6.05	5.95	5.70	9.30	14.00	18.10	26.55	6.65	5.90	5.90	5.85	10.50	15.15	18.75	27.35
1.00	6.20	5.95	5.30	5.30	8.35	10.50	13.00	16.55	6.35	6.00	5.50	5.60	9.35	11.55	13.95	17.05
	$\rho_0 = 0.5$															
0.00	6.75	5.15	5.65	5.80	9.55	14.25	20.15	34.70	6.80	5.35	5.85	6.05	11.00	15.25	21.25	35.75
0.25	6.75	5.20	5.60	5.85	9.55	14.15	20.35	34.75	6.80	5.30	5.75	5.95	11.00	15.15	21.30	35.50
0.50	6.60	5.35	5.35	5.75	8.95	13.85	19.55	34.15	6.60	5.65	5.65	5.75	10.55	14.85	20.70	35.30
0.75	7.00	5.05	5.45	6.05	8.65	12.65	17.20	26.60	7.05	5.25	5.55	6.10	10.00	13.45	18.45	27.60
0.95	6.70	6.00	5.85	5.70	7.80	8.25	9.10	9.25	7.05	5.85	5.95	5.90	8.85	9.20	9.55	9.45
1.00	6.05	5.80	5.35	5.75	6.90	6.30	6.20	6.75	6.35	6.15	5.60	5.90	7.90	6.95	6.60	7.00
	$\rho_0 = 0.75$															
0.00	7.05	5.00	5.40	5.70	9.45	13.95	19.30	32.85	7.35	5.40	5.65	5.95	10.80	15.25	20.55	33.95
0.25	7.00	5.05	5.45	5.70	9.60	14.05	19.20	33.10	7.40	5.45	5.60	5.75	11.00	15.35	20.65	34.15
0.50	7.45	5.30	5.30	5.70	9.25	13.00	19.00	31.60	7.35	5.60	5.70	5.80	10.40	14.65	20.05	32.70
0.75	7.40	5.25	5.45	6.25	8.95	10.75	14.00	19.60	7.35	5.75	5.65	6.25	10.10	11.95	14.75	20.40
0.95	6.90	5.85	5.75	6.10	6.90	6.85	6.95	7.50	7.60	5.95	6.15	6.10	8.60	7.65	7.40	7.60
1.00	6.10	5.55	5.20	5.60	6.75	5.70	5.35	5.95	6.95	6.40	5.60	6.10	7.65	6.15	6.15	6.20

Notes: The true parameter value is $\beta_0 = 1$ and power is calculated at 0.9. See also the notes to Table 3a.

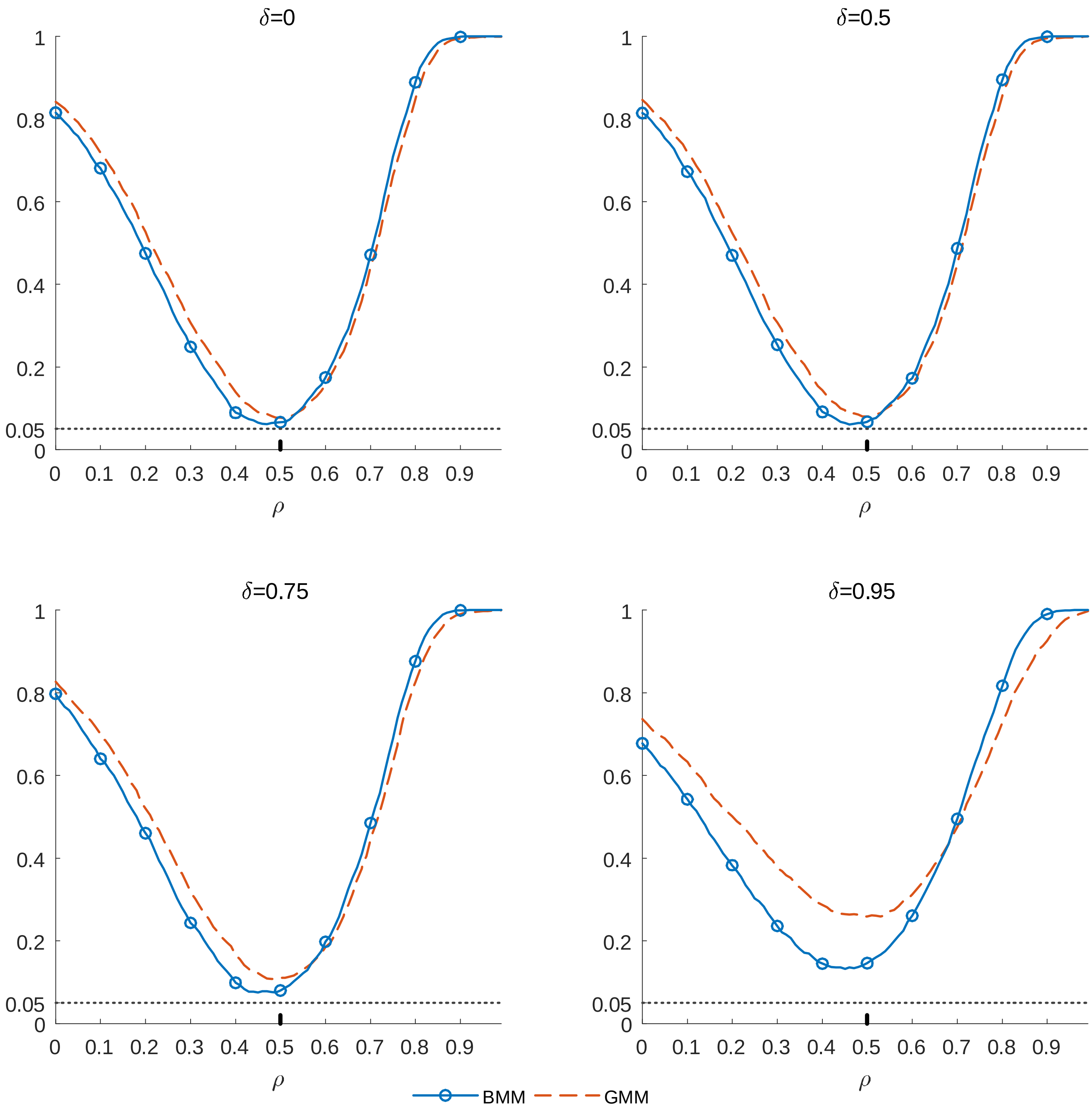

Figure 1a: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and non-Gaussian errors for different values of δ

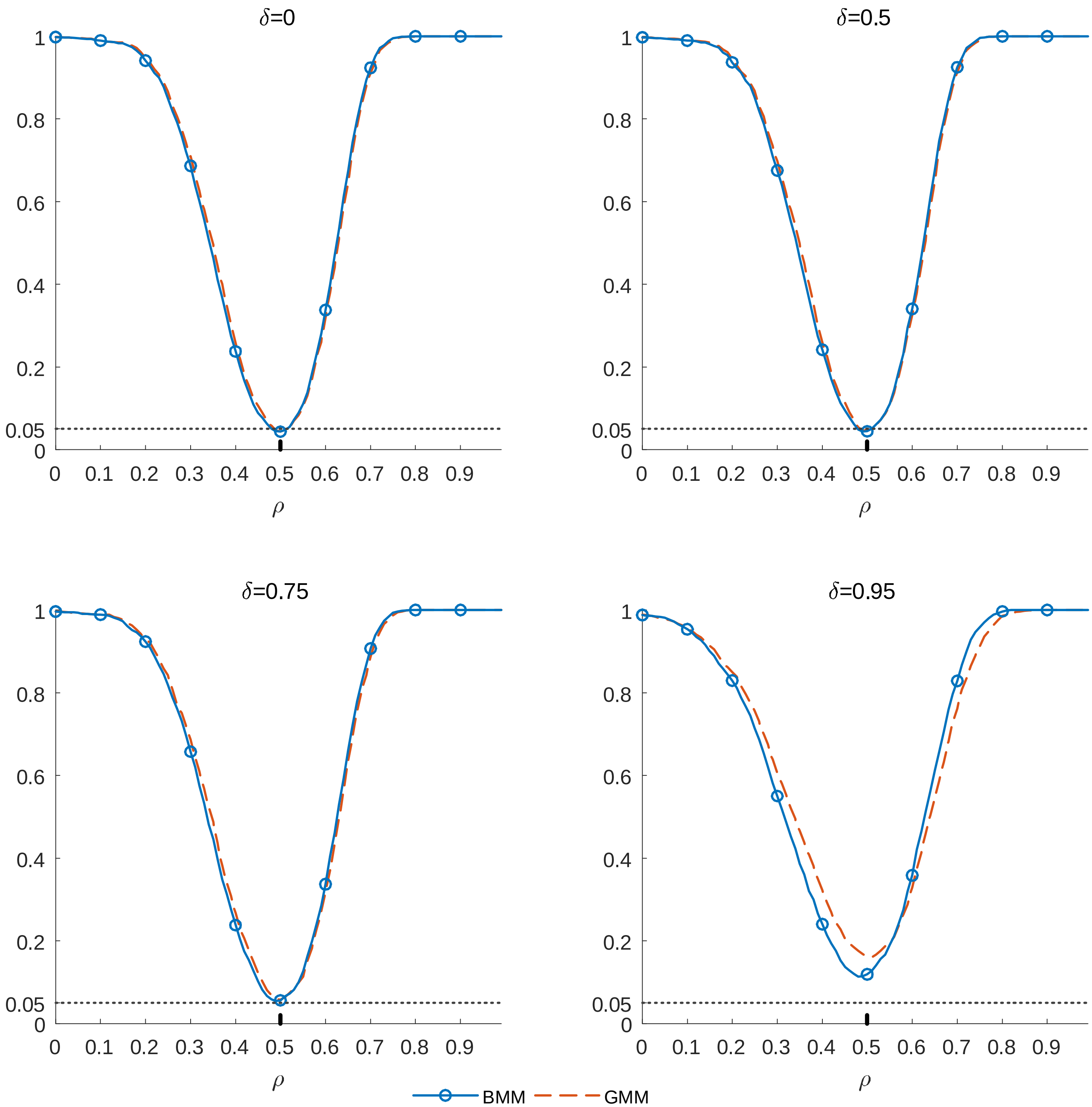

Figure 1b: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and non-Gaussian errors for different values of δ

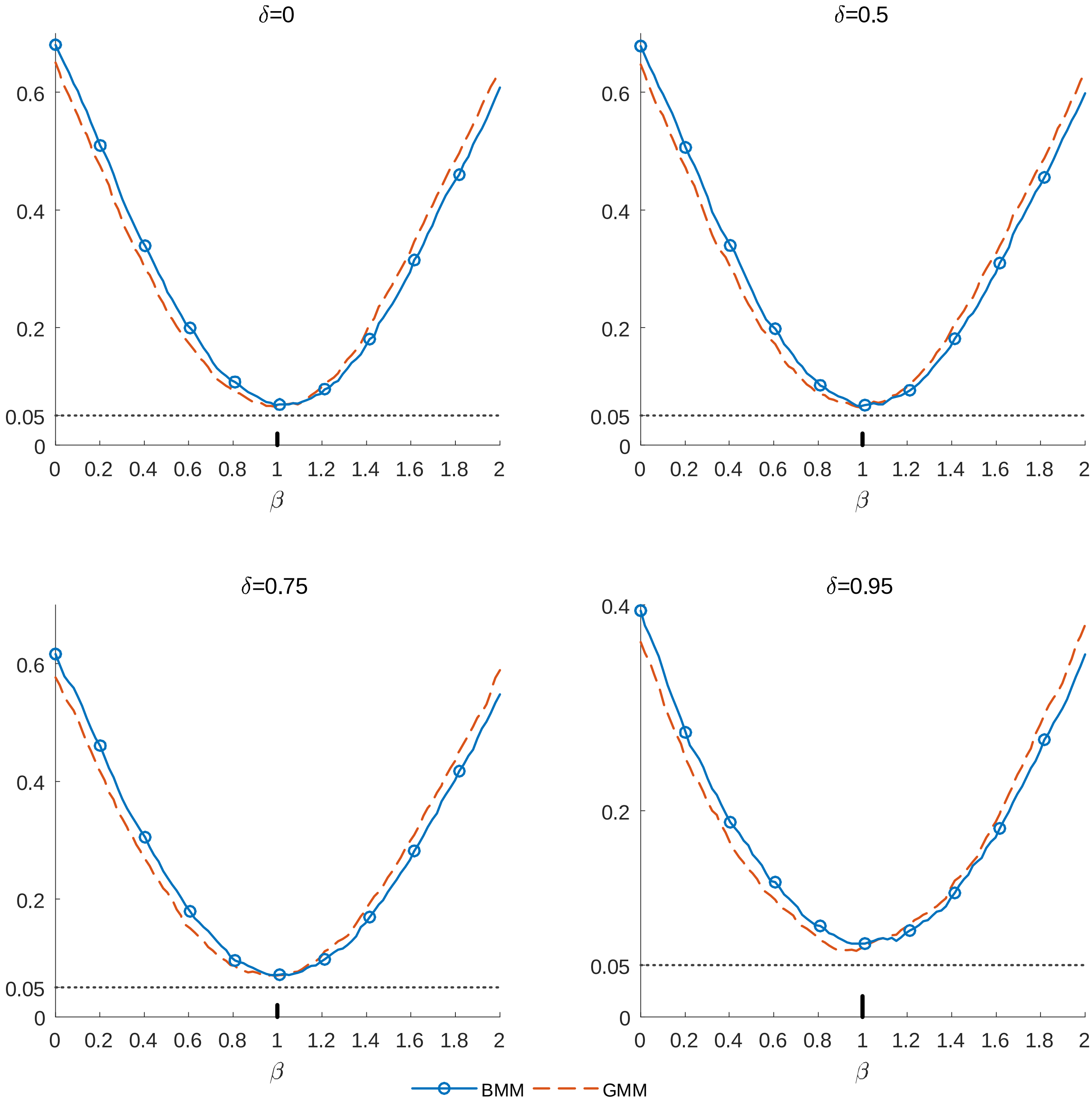

Figure 2a: Empirical power functions for β in the case of $\beta_0 = 1$, $n = 100$, and non-Gaussian errors for different values of δ

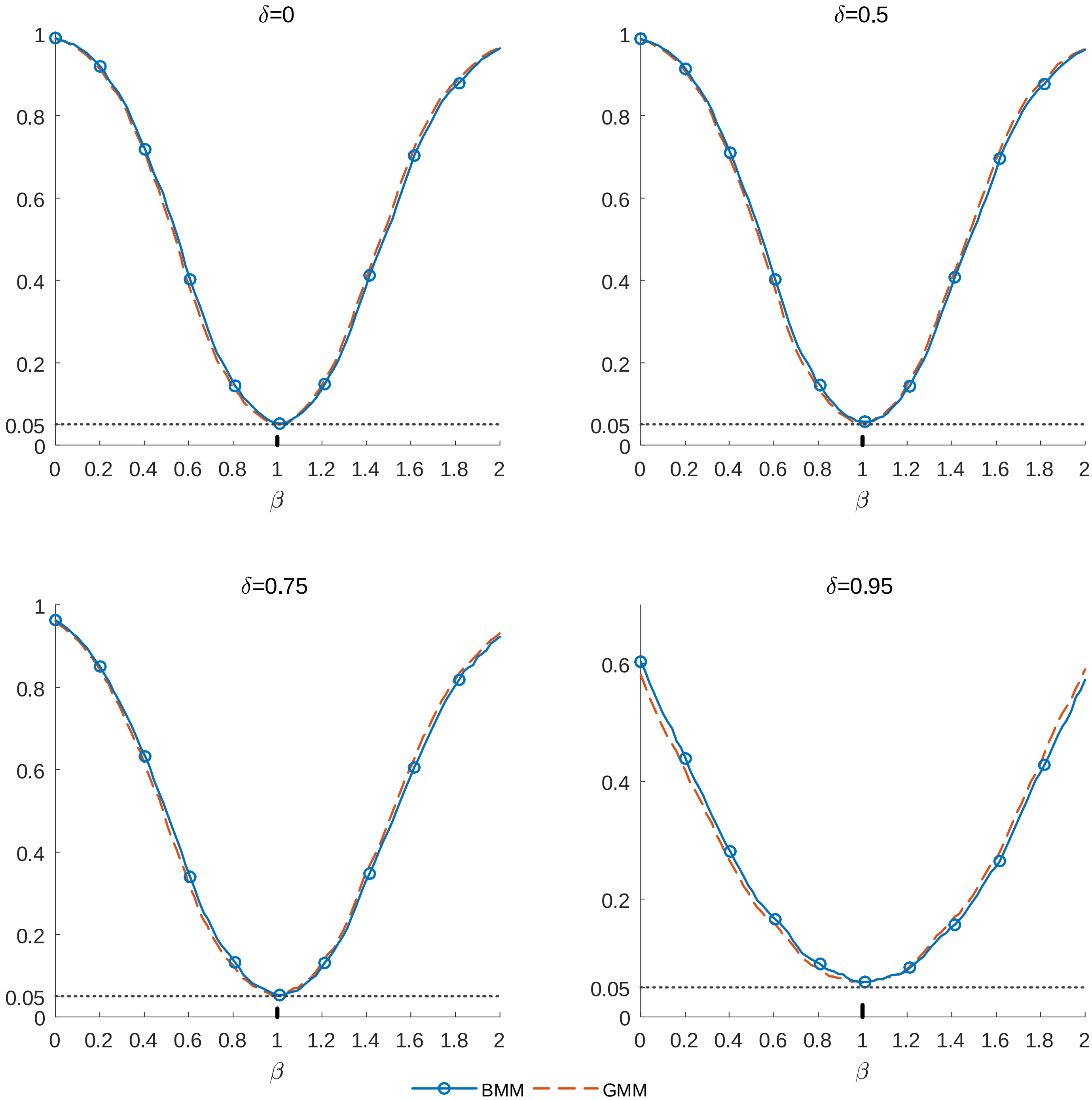

Figure 2b: Empirical power functions for β in the case of $\beta_0 = 1$, $n = 300$, and non-Gaussian errors for different values of δ

References

Acemoglu, D., V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi (2012). The network origins of aggregate fluctuations. *Econometrica 80*, 1977–2016.

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Volume 4. Springer Science & Business Media.

Bai, J. and S. Ng (2006). Evaluating latent and observed factors in macroeconomics and finance. *Journal of Econometrics 131*(1-2), 507–537.

Bailey, N., S. Holly, and M. H. Pesaran (2016). A two-stage approach to spatio-temporal analysis with strong and weak cross-sectional dependence. *Journal of Applied Econometrics 31*(1), 249–280.

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

Brownlees, C. T. and G. Mesters (2018). Detecting granular time series in large panels. Version dated to July 13, 2018. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.3040636`.

Case, A. (1992). Neighborhood influence and technological change. *Regional Science and Urban Economics 22*(3), 491–508.

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

Chudik, A. and M. H. Pesaran (2017). A bias-corrected method of moments approach for estimation of dynamic panels. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.3043407`.

Cliff, A. D. and J. K. Ord (1973). *Spatial Autocorrelation*. London: Pion.

Cliff, A. D. and J. K. Ord (1981). *Spatial Process: Models and Applications*. London: Pion.

Cohen, L. and A. Frazzini (2008). Economic links and predictable returns. *Journal of Finance 63*(4), 1977–2011.

Cooley, T. F. and E. C. Prescott (1995). Economic growth and business cycles. In T. F. Cooley (Ed.), *Frontiers of Business Cycle Research*, Chapter 1, pp. 1–38. Princeton, New Jersey: Princeton University Press.

Danthine, J.-P., J. B. Donaldson, and P. Siconolfi (2008). Distribution risk and equity returns. In R. Mehra (Ed.), *Handbook of the Equity Risk Premium*, Handbooks in Finance, Chapter 10, pp. 415–466. Elsevier Science.

Dungey, M. and V. Volkov (2018). R&D and wholesale trade are critical to the economy: Identifying dominant sectors from economic networks. *Economics Letters 162*, 81–85.

Gollin, D. (2002). Getting income shares right. *Journal of Political Economy 110*(2), 458–474.

Hansen, G. D. and R. Wright (1998). The labor market in real business cycle theory. In J. E. Harley, K. D. Hoover, and K. D. Salyer (Eds.), *Real Business Cycles: A Reader*, Chapter 9, pp. 168–178. London and New York: Routledge.

Kapetanios, G., M. H. Pesaran, and S. Reese (2018). A residual-based threshold method for detection of units that are too big to fail in large factor models. USC-INET Research Paper No. 18-23.

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 (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics 17*(1), 99–121.

Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*(2), 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*, 219–257.

Lee, L.-f. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews 22*, 307–335.

Lee, L.-f. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica 72*, 1899–1925.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics 137*, 489–514.

Lee, L.-f. and J. Yu (2010). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics 154*, 165–185.

Lee, L.-f. and J. Yu (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics 180*, 174–197.

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

Parker, J. and D. Sul (2016). Identification of unknown common factors: Leaders and followers. *Journal of Business & Economic Statistics 34*(2), 227–239.

Pesaran, M. H. and C. F. Yang (2019). Econometric analysis of production networks with dominant units. *Journal of Econometrics (forthcoming).*

Upton, G. and B. Fingleton (1985). *Spatial Data Analysis by Example. Volume 1: Point Pattern and Quantitative Data.* John Wiley & Sons Ltd.

Yang, C. F. (2018). Common factors and spatial dependence: An application to US house prices. Available at SSRN: `http://dx.doi.org/10.2139/ssrn.2856818`.

Online Mathematical Appendix to "Estimation and Inference in Spatial Models with Dominant Units"

M. Hashem Pesaran
University of Southern California, USA, and Trinity College, Cambridge, UK

Cynthia Fan Yang
Florida State University

March 10, 2019

This online mathematical appendix is organized into two sections. Section A.1 contains statements and proofs of necessary lemmas used in establishing the main theoretical results of the paper. Section A.2 provides proofs of the theorems and propositions in Sections 3–5 of the paper. Throughout this appendix, Assumptions 1–8 refer to the Assumptions made in the paper.

A.1 Lemmas

Lemma A.1 *Let* $\mathbf{A} = (a_{ij})$ *and* $\mathbf{B} = (b_{ij})$ *be* $n \times n$ *matrices, and suppose that* $\sup_{i,j} |a_{ij}| < K$.

(i) Let $\mathbf{C} = (c_{ij}) = \mathbf{AB}$. *If* $||\mathbf{B}||_1 < K$, *then* $\sup_{i,j} |c_{ij}| < K$ *and* $Tr\,(\mathbf{C}) = O\,(n)$.

(ii) Let $\mathbf{D} = (d_{ij}) = \mathbf{BA}$. *If* $||\mathbf{B}||_\infty < K$, *then* $\sup_{i,j} |d_{ij}| < K$ *and* $Tr\,(\mathbf{D}) = O\,(n)$.

Proof. This lemma is a special case of Lemma A.8 of Lee (2004). ■

Lemma A.2 *Suppose that* $\mathbf{A}$ *and* $\mathbf{B}$ *are* $n \times n$ *matrices that satisfy* $||\mathbf{A}||_\infty < K$ *and* $||\mathbf{B}||_\infty < K$, *then* $||\mathbf{AB}||_\infty < K$.

Proof. This result can be readily established by the submultiplicativity of the maximum row sum matrix norm, that is, $||\mathbf{AB}||_\infty \leq ||\mathbf{A}||_\infty ||\mathbf{B}||_\infty$. A proof can be found in, for example, Horn and Johnson (2012, Example 5.6.5). ■

Lemma A.3 *Let* $\mathbf{A}$ *be an* $n \times n$ *matrix and* $\mathbf{b}$ *be an* $n \times 1$ *vector.*

(i) If $||\mathbf{A}||_1 < K$, *and* $||\mathbf{b}||_1 = O(n^\delta)$, $0 \leq \delta \leq 1$, *then* $||\mathbf{Ab}||_1 = O(n^\delta)$.

(ii) If $||\mathbf{A}||_1 = O(n^\delta)$, $0 \leq \delta \leq 1$, *and* $||\mathbf{b}||_1 < K$, *then* $||\mathbf{Ab}||_1 = O(n^\delta)$.

Proof. (i) Let $\mathbf{c} = \mathbf{Ab}$ and its i^{th} element is denoted by c_i. Then

$$\sum_{i=1}^{n} |c_i| = \sum_{i=1}^{n} |\sum_{j=1}^{n} a_{ij} b_j| \leq \sum_{j=1}^{n} |b_j| \sum_{i=1}^{n} |a_{ij}| \leq \sum_{j=1}^{n} |b_j| \left(\sup_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{ij}| \right) = O(n^{\delta}),$$

The result in (ii) follows from similar reasoning. ■

Lemma A.4 *Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be $n \times n$ matrices such that $||\mathbf{A}||_{\infty} < K$, $||\mathbf{B}||_{\infty} < K$, and $||\mathbf{B}||_1 = O\left(n^{\delta}\right)$, where $0 \leq \delta \leq 1$. Then*

(i) $Tr\left(\mathbf{A}'\mathbf{B}\mathbf{B}'\mathbf{A}\right) = O\left(n^{\delta+1}\right)$,

(ii) $Tr\left[(\mathbf{A}'\mathbf{B})^2\right] = O\left(n^{\delta+1}\right)$,

(iii) $Tr\left(\mathbf{A}\mathbf{B}'\mathbf{C}\right) = O\left(n^{\delta+1}\right)$, where $\mathbf{C} = (c_{ij})$ is an $n \times n$ matrix such that $\sup_{i,j} |c_{ij}| < K$.

Proof. (i) From $||\mathbf{A}||_{\infty} < K$, it follows that $\sup_{i,j} |a_{ij}| < K$ and $\sum_{i=1}^{n} \sum_{j=1}^{n} |a_{ji}| < Kn$. Then

$$\begin{aligned} |Tr\left(\mathbf{A}'\mathbf{B}\mathbf{B}'\mathbf{A}\right)| &= |\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} a_{ji} b_{jk} b_{lk} a_{li}| \leq \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ji}| \sum_{l=1}^{n} |a_{li}| \sum_{k=1}^{n} |b_{jk}||b_{lk}| \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ji}| \left(\sup_{1 \leq i \leq n} \sum_{l=1}^{n} |a_{li}| \right) \sum_{k=1}^{n} |b_{jk}| \left(\sup_{1 \leq l,k \leq n} |b_{lk}| \right) \\ &\leq Kn^{\delta} \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ji}| \leq Kn^{\delta+1}, \end{aligned}$$

which establishes the claim.

(ii) Since $Tr\left[(\mathbf{A}'\mathbf{B})^2\right] \leq Tr\left(\mathbf{A}'\mathbf{B}\mathbf{B}'\mathbf{A}\right)$ by Schur's inequality, the result immediately follows from (i).

(iii) Note that

$$\begin{aligned} |Tr\left(\mathbf{A}\mathbf{B}'\mathbf{C}\right)| &= \left| \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij} b_{kj} c_{ki} \right| \leq \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}| \sum_{k=1}^{n} |b_{kj}| \, |c_{ki}| \\ &\leq \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}| \left(\sup_{1 \leq j \leq n} \sum_{k=1}^{n} |b_{kj}| \right) \left(\sup_{1 \leq i,k \leq n} |c_{ki}| \right) \leq Kn^{\delta+1}, \end{aligned}$$

and the result follows. ■

Lemma A.5 *Suppose that $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is a vector of random variables, where $\varepsilon_i \sim IID(0, \sigma^2)$ for all $i = 1, 2, \ldots, n$, and its fourth moment, $\mu_4 = E(\varepsilon_i^4)$, exists. Let $\gamma_2 = \mu_4 - 3\sigma^4$. Then for any $n \times n$ constant matrix (need not be symmetric) $\mathbf{A} = (a_{ij})$, we have*

(i) $E(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}) = \sigma^2 Tr(\mathbf{A})$,

(ii) $E(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon})^2 = \gamma_2 \sum_{i=1}^{n} a_{ii}^2 + \sigma^4 \left[Tr^2(\mathbf{A}) + Tr(\mathbf{A}\mathbf{A}') + Tr(\mathbf{A}^2)\right]$,

(iii) $Var(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}) = \gamma_2 \sum_{i=1}^{n} a_{ii}^2 + \sigma^4 \left[Tr(\mathbf{A}\mathbf{A}') + Tr(\mathbf{A}^2)\right] \leq K Tr\left(\mathbf{A}\mathbf{A}'\right)$.

Proof. See Lemma A.11 of Lee (2004). The inequality in (iii) follows from $\sum_{i=1}^{n} a_{ii}^2 \leq Tr(\mathbf{A}\mathbf{A}')$, $Tr\left(\mathbf{A}^2\right) \leq Tr(\mathbf{A}\mathbf{A}')$ (Schur's inequality), $\gamma_2 < K$ and $\sigma^2 < K$. ■

Lemma A.6 *Suppose that Assumptions 4 and 5 in the paper hold. Let* $\mathbf{S} = \mathbf{S}(\rho) = \mathbf{I}_n - \rho \mathbf{W}$, $\mathbf{G} = \mathbf{G}(\rho) = \mathbf{W}\mathbf{S}^{-1}(\rho) = \mathbf{W}(\mathbf{I}_n - \rho\mathbf{W})^{-1}$, *where* $|\rho| < 1$. *Then*

(i) $||\mathbf{S}^{-1}||_\infty < K$, *and* $||\mathbf{S}^{-1}||_1 = O(n^\delta)$.

(ii) $\|\mathbf{G}\|_\infty < K$, *and* $||\mathbf{G}||_1 = O(n^\delta)$.

Proof. (i) By assumption we have $||\rho\mathbf{W}||_\infty < 1$, and hence $\mathbf{S}^{-1} = \sum_{k=0}^{\infty} (\rho\mathbf{W})^k$.[A1] It follows that

$$||\mathbf{S}^{-1}||_\infty \leq 1 + |\rho|||\mathbf{W}||_\infty + |\rho|^2||\mathbf{W}||_\infty^2 + \ldots = \frac{1}{1 - |\rho|||\mathbf{W}||_\infty} < K.$$

We next prove that $||\mathbf{S}^{-1}||_1 = O(n^\delta)$. The matrix $\mathbf{W}$ can be partitioned as follows:

$$\underset{n\times n}{\mathbf{W}} = \begin{pmatrix} \underset{m\times m}{\mathbf{W}_{11}} & \underset{m\times(n-m)}{\mathbf{W}_{12}} \\ \underset{(n-m)\times m}{\mathbf{W}_{21}} & \underset{(n-m)\times(n-m)}{\mathbf{W}_{22}} \end{pmatrix}.$$

Applying the formula for the inverse of a partitioned matrix gives

$$\mathbf{S}^{-1} = \begin{pmatrix} \mathbf{\Phi}_1^{-1} & \rho\mathbf{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1} \\ \rho\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\mathbf{\Phi}_1^{-1} & \mathbf{S}_{22}^{-1} + \rho^2\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\mathbf{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1} \end{pmatrix},$$

where

$$\mathbf{\Phi}_1 = \mathbf{I}_m - \rho\mathbf{W}_{11} - \rho^2\mathbf{W}_{12}\mathbf{S}_{22}^{-1}\mathbf{W}_{21},$$

and $\mathbf{S}_{22} = \mathbf{I}_{n-m} - \rho\mathbf{W}_{22}$. Since under Assumptions 4 and 5 $|\rho \|\mathbf{W}_{22}\|_\infty < 1$, and $|\rho| \|\mathbf{W}_{22}\|_1 < 1$, then $\left\|\mathbf{S}_{22}^{-1}\right\|_\infty < K$ and $\left\|\mathbf{S}_{22}^{-1}\right\|_1 < K$. Also, since m is fixed and does not rise with n, it is sufficient to examine $||\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\mathbf{\Phi}_1^{-1}||_1$ and $||\mathbf{S}_{22}^{-1} + \rho^2\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\mathbf{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}||_1$. Let $\mathbf{w}_{\cdot j,21}$ denote the j^{th} column of $\mathbf{W}_{21}$. By Lemma A.3, $||\mathbf{S}_{22}^{-1}\mathbf{w}_{\cdot j,21}||_1 = O(n^{\delta_j})$, for $j = 1, 2, \ldots, m$, which yields $||\mathbf{S}_{22}^{-1}\mathbf{W}_{21}||_1 = O\left(n^\delta\right)$, where $\delta = \max_j(\delta_j)$. Therefore,

$$||\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\mathbf{\Phi}_1^{-1}||_1 \leq ||\mathbf{S}_{22}^{-1}\mathbf{W}_{21}||_1||\mathbf{\Phi}_1^{-1}||_1 = O(n^\delta), \tag{A.1}$$

noting that the norm of the $m \times m$ matrix $\mathbf{\Phi}_1^{-1}$ is bounded since m is fixed. Similarly, $||\mathbf{W}_{12}\mathbf{S}_{22}^{-1}||_1 \leq ||\mathbf{W}_{12}||_1||\mathbf{S}_{22}^{-1}||_1 < K$, and then

$$||\mathbf{S}_{22}^{-1}\mathbf{W}_{21}\mathbf{\Phi}_1^{-1}\mathbf{W}_{12}\mathbf{S}_{22}^{-1}||_1 \leq ||\mathbf{S}_{22}^{-1}\mathbf{W}_{21}||_1||\mathbf{\Phi}_1^{-1}||_1||\mathbf{W}_{12}\mathbf{S}_{22}^{-1}||_1 = O(n^\delta). \tag{A.2}$$

Combining (A.1) and (A.2), it follows that $||\mathbf{S}^{-1}||_1 = O(n^\delta)$.

(ii) The boundedness of the row norm of $\mathbf{G}$ is an immediate result of Assumption 4, Lemma A.2 and Lemma A.6(i). Let $\mathbf{G} = (g_{ij})_{n\times n}$ and $\mathbf{S}^{-1} = \left(s_{ij}^*\right)_{n\times n}$. For the j^{th} column of $\mathbf{G}$, $j = 1, 2, \ldots, n$, we have

$$\sum_{i=1}^{n} |g_{ij}| = \sum_{i=1}^{n} |\sum_{l=1}^{n} w_{il}s_{lj}^*| = \sum_{i=1}^{n} |\sum_{l=1}^{m} w_{il}s_{lj}^*| + \sum_{i=1}^{n} |\sum_{l=m+1}^{n} w_{il}s_{lj}^*|$$

[A1] See, for example, Horn and Johnson (2012, Corollary 5.6.16).

$$\leq \left(\sup_{1 \leq l \leq m} \sum_{i=1}^{n} |w_{il}| \right) \sum_{l=1}^{m} |s_{lj}^{*}| + \left(\sup_{m+1 \leq l \leq n} \sum_{i=1}^{n} |w_{il}| \right) \sum_{l=m+1}^{n} |s_{lj}^{*}|$$
$$\leq Kn^{\delta} m + Kn^{\delta}.$$

Since m is fixed, we obtain $\sum_{i=1}^{n} |g_{ij}| \leq Kn^{\delta}$, for all j, and this completes the proof. ■

Lemma A.7 *Suppose that Assumptions 1–5 and 8 in the paper hold, and* $\delta_b = \delta < 1$. *Let* $\mathbf{G} = \mathbf{G}(\rho) = \mathbf{W}(\mathbf{I}_n - \rho \mathbf{W})^{-1}$, and $\boldsymbol{\eta} = \mathbf{G}\mathbf{X}\boldsymbol{\beta}$. *Then*

(i) $n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon} = o_p(1)$,
(ii) $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\varepsilon} = \sigma^2 Tr(n^{-1}\mathbf{G}'\mathbf{C}) + o_p(1)$,
(iii) $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon} = \sigma^2 Tr(n^{-1}\mathbf{G}'\mathbf{C}\mathbf{G}) + o_p(1)$,
(iv) $n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon} = o_p(1)$,
(v) $n^{-1}\boldsymbol{\eta}'\mathbf{C}\boldsymbol{\varepsilon} = o_p(1)$,
(vi) $n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon} = o_p(1)$,
(vii) $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\eta} = o_p(1)$.

Proof. (i) Under Assumption 8, $\mathbf{C} = (\mathbf{B} + \mathbf{B}')/2$, and $Tr(\mathbf{B}) = 0$. Then $E(n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}) = n^{-1}\sigma^2 Tr(\mathbf{C}) = 0$, noting that $Tr(\mathbf{C}) = Tr(\mathbf{B}) = 0$. Also, by Lemma A.5,

$$Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}\right) = n^{-2}\gamma_2 \sum_{i=1}^{n} c_{ii}^2 + 2n^{-2}\sigma^4 Tr(\mathbf{C}^2) \leq Kn^{-2} Tr\left(\mathbf{C}^2\right),$$

where $\gamma_2 = E(\varepsilon_i^4) - 3\sigma^4$. Under Assumption 8, $\|\mathbf{B}\|_\infty < K$, which implies that $\sup_{i.j}|b_{ij}| < K$. By Lemma A.3 we obtain $Tr(\mathbf{B}^2) = O(n)$ and $Tr(\mathbf{B}\mathbf{B}') = O(n)$. Then $Tr(\mathbf{C}^2) = \frac{1}{2}\left[Tr(\mathbf{B}^2) + Tr(\mathbf{B}'\mathbf{B})\right] = O(n)$, and $lim_{n\to\infty} Var(n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}) = 0$. Hence, noting that $E(n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}) = 0$, then $n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\,\boldsymbol{\varepsilon} \to_p 0$, for all values of δ.

(ii) First note that $E(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\varepsilon}) = \sigma^2 Tr(n^{-1}\mathbf{C}\mathbf{G}')$. Since $\|\mathbf{G}'\|_1 < K$ by Lemma A.6, and $\sup_{i.j}|c_{ij}| \leq \sup_{i.j}|b_{ij}| < K$, applying Lemma A.1, then $Tr(n^{-1}\mathbf{C}\mathbf{G}') = O(1)$ which establishes that $E(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\varepsilon}) = O(1)$. Now, using Lemma A.5(iii) we have

$$\begin{aligned} Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\varepsilon}\right) &\leq Kn^{-2} Tr\left(\mathbf{G}'\mathbf{C}^2\mathbf{G}\right) \\ &\leq Kn^{-2}\left[2Tr\left(\mathbf{B}^2\mathbf{G}\mathbf{G}'\right) + Tr(\mathbf{G}'\mathbf{B}\mathbf{B}'\mathbf{G}) + Tr(\mathbf{B}\mathbf{G}\mathbf{G}'\mathbf{B}')\right]. \end{aligned}$$

Since $\|\mathbf{B}\|_\infty < K$ under Assumption 8, and $\|\mathbf{G}_0\|_\infty < K$ by Lemma A.6, applying Lemma A.2 yields $\|\mathbf{B}\mathbf{G}\|_\infty < K$ and $\|\mathbf{B}^2\mathbf{G}\|_\infty < K$. Then by Lemma A.1 we have $Tr\left[(\mathbf{B}^2\mathbf{G})\mathbf{G}'\right] = O(n)$ and $Tr\left[\mathbf{B}\mathbf{G}(\mathbf{B}\mathbf{G})'\right] = O(n)$. Since $\|\mathbf{B}\|_1 = O\left(n^{\delta}\right)$ under Assumption 8, by Lemma A.4 we obtain $Tr(\mathbf{G}'\mathbf{B}\mathbf{B}'\mathbf{G}) = O\left(n^{\delta+1}\right)$. Hence, $Var(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\varepsilon}) = O\left(n^{\delta-1}\right)$, and it follows that $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\,\boldsymbol{\varepsilon} \to_p \sigma^2 Tr(n^{-1}\mathbf{C}\mathbf{G}')$ if $\delta < 1$.

(iii) By Lemma A.5,

$$E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon}\right) = \sigma^2 Tr\left(n^{-1}\mathbf{G}'\mathbf{C}\mathbf{G}\right) = \sigma^2 Tr\left(n^{-1}\mathbf{G}'\mathbf{B}\mathbf{G}\right),$$

and

$$Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon}\right) \le Kn^{-2}Tr\left[\left(\mathbf{G}'\mathbf{C}\mathbf{G}\right)^2\right] = \frac{1}{2}Kn^{-2}\left\{Tr\left[\left(\mathbf{G}'\mathbf{B}\mathbf{G}\right)^2\right] + Tr\left(\mathbf{G}'\mathbf{B}\mathbf{G}\mathbf{G}'\mathbf{B}'\mathbf{G}\right)\right\}.$$

Since $\|\mathbf{B}\|_\infty < K$ under Assumption 8, and $\|\mathbf{G}_0\|_\infty < K$ by Lemma A.6, applying Lemma A.2 yields $\|\mathbf{B}\mathbf{G}\|_\infty < K$. Then by Lemma A.1, $Tr\left(\mathbf{G}'\mathbf{B}\mathbf{G}\right) = O(n)$ and hence $E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon}\right) = O(1)$. Since $\|\mathbf{G}\|_1 = O\left(n^\delta\right)$ by Lemma A.6, applying Lemma A.4 gives $Tr\left[\left(\mathbf{G}'\mathbf{B}\mathbf{G}\right)^2\right] = O\left(n^{\delta+1}\right)$ and $Tr\left(\mathbf{G}'\mathbf{B}\mathbf{G}\mathbf{G}'\mathbf{B}'\mathbf{G}\right) = O\left(n^{\delta+1}\right)$. Therefore, $Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon}\right) = O\left(n^{\delta-1}\right)$, and it follows that $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\mathbf{G}\,\boldsymbol{\varepsilon} \to_p \sigma^2 Tr\left(n^{-1}\mathbf{G}'\mathbf{B}\mathbf{G}\right)$, if $\delta < 1$.

(iv) $E\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}\right) = 0$ readily follows from the independence of $\mathbf{X}$ and $\boldsymbol{\varepsilon}$. Also,

$$Var\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}|\mathbf{X}\right) = n^{-2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}E\left(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\mathbf{C}'\mathbf{X}\boldsymbol{\beta} = n^{-2}\sigma^2\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}^2\mathbf{X}\boldsymbol{\beta},$$

and then

$$\begin{aligned} Var\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}\right) &= E\left[Var\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}|\mathbf{X}\right)\right] + Var\left[E\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}|\mathbf{X}\right)\right] \\ &= E\left[Var\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}|\mathbf{X}\right)\right] = n^{-2}\sigma^2 E\left(\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}^2\mathbf{X}\boldsymbol{\beta}\right) \\ &= n^{-2}\sigma^2 Tr\left(\mathbf{C}^2\mathbf{M}\right) = \frac{\sigma^2}{2n^2}\left[Tr\left(\mathbf{B}^2\mathbf{M}\right) + Tr\left(\mathbf{B}\mathbf{B}'\mathbf{M}\right)\right], \end{aligned}$$

where

$$\mathbf{M} = \left(m_{ij}\right) = E\left(\mathbf{X}\boldsymbol{\beta}\boldsymbol{\beta}'\mathbf{X}'\right). \tag{A.3}$$

Let β_l denote the l^{th} element of $\boldsymbol{\beta}$, for $l = 1, 2, \ldots, k$. Then, for any $1 \le i, j \le n$,

$$m_{ij} = E\left(\sum_{l=1}^{k}\sum_{l'=1}^{k} x_{il}x_{jl'}\beta_l\beta_{l'}\right) = \sum_{l=1}^{k}\sum_{l'=1}^{k}\beta_l\beta_{l'}E\left(x_{il}x_{jl'}\right).$$

Since

$$\left|E\left(x_{il}x_{jl'}\right)\right| \le E\left|x_{il}x_{jl'}\right| \le \left[E\left(x_{il}^2\right)E\left(x_{jl'}^2\right)\right]^{1/2} \le \sup_{1\le i\le n, 1\le l\le k} E\left(x_{il}^2\right) < K,$$

and $\sup_{1\le l\le k}|\beta_l| < K$, we have

$$\begin{aligned} \sup_{1\le i.j\le n}|m_{ij}| &\le \sup_{1\le i.j\le n}\left|\sum_{l=1}^{k}\sum_{l'=1}^{k}\beta_l\beta_{l'}E\left(x_{il}x_{jl'}\right)\right| \le \left(\sup_{1\le l\le k}|\beta_l|\right)^2 \sup_{1\le i.j\le n}\sum_{l=1}^{k}\sum_{l'=1}^{k}\left|E\left(x_{il}x_{jl'}\right)\right| \\ &\le k^2\left(\sup_{1\le l\le k}|\beta_l|\right)^2 \sup_{1\le i\le n, 1\le l\le k} E\left(x_{il}^2\right) < K. \end{aligned} \tag{A.4}$$

Also under Assumption 8 $\|\mathbf{B}\|_\infty < K$, and by Lemma A.2 $\|\mathbf{B}^2\|_\infty < K$. Then applying Lemma A.1(ii) we obtain $Tr\left(\mathbf{B}^2\mathbf{M}\right) = O(n)$. Moreover, as $\|\mathbf{B}\|_1 = O\left(n^\delta\right)$, applying Lemma A.4(iii) gives $Tr\left(\mathbf{B}\mathbf{B}'\mathbf{M}\right) = O\left(n^{\delta+1}\right)$. Therefore, $Tr\left(\mathbf{C}^2\mathbf{M}\right) = O\left(n^{\delta+1}\right)$ and $Var\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}\right) = O\left(n^{\delta-1}\right)$. It follows that $n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon} \to_p 0$, if $\delta < 1$.

(v) The proof is similar to that of (iv). The mean of $n^{-1}\boldsymbol{\eta}'\mathbf{C}\boldsymbol{\varepsilon}$ is zero and its variance is given by

$$Var\left(n^{-1}\boldsymbol{\eta}'\mathbf{C}\boldsymbol{\varepsilon}\right) = n^{-2}\sigma^2 E\left(\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{C}^2\mathbf{G}\mathbf{X}\boldsymbol{\beta}\right) = n^{-2}\sigma^2 Tr\left[\mathbf{C}^2\left(\mathbf{G}\mathbf{M}\mathbf{G}\right)\right],$$

where $\mathbf{M}$ is defined in (A.3). Let $\tilde{\mathbf{M}} = (\tilde{m}_{ij}) = \mathbf{G}\mathbf{M}\mathbf{G}'$. We have shown in (A.4) that $\sup_{1\leq i.j\leq n}|m_{ij}| < K$. Using $\|\mathbf{G}_0\|_\infty < K$ by Lemma A.6, and Lemma A.2(i) and (ii) yields $\sup_{1\leq i.j\leq n}|\tilde{m}_{ij}| < K$. Repeating the arguments for $Tr\left(\mathbf{C}^2\mathbf{M}\right)$ in (iv) leads to $Tr\left(\mathbf{C}^2\tilde{\mathbf{M}}\right) = O\left(n^{\delta+1}\right)$. Therefore, $Var\left(n^{-1}\boldsymbol{\eta}'\mathbf{C}\boldsymbol{\varepsilon}\right) = O\left(n^{\delta-1}\right)$ and it follows that $n^{-1}\boldsymbol{\eta}'\mathbf{C}\boldsymbol{\varepsilon} \to_p 0$ if $\delta < 1$.

(vi) Similar to proving the results in (iv) and (v), it can be shown that the mean of $n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon}$ is zero and its variance is

$$\begin{aligned} Var\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon}\right) &= n^{-2}\sigma^2 Tr\left(\mathbf{C}\mathbf{G}\mathbf{G}'\mathbf{C}\mathbf{M}\right) \\ &\leq Kn^{-2}\left[Tr\left(\mathbf{B}\mathbf{G}\mathbf{G}'\mathbf{B}\mathbf{M}\right) + Tr\left(\mathbf{B}'\mathbf{G}\mathbf{G}'\mathbf{B}\mathbf{M}\right)\right. \\ &\quad \left. + Tr\left(\mathbf{B}\mathbf{G}\mathbf{G}'\mathbf{B}'\mathbf{M}\right) + Tr\left(\mathbf{B}'\mathbf{G}\mathbf{G}'\mathbf{B}'\mathbf{M}\right)\right], \end{aligned} \tag{A.5}$$

where $\mathbf{M}$ is defined in (A.3). Let $\mathbf{P} = (p_{ij}) = \mathbf{B}\mathbf{M}$. Then $\sup_{1\leq i.j\leq n}|p_{ij}| < K$ follows from Lemma A.1, due to $\|\mathbf{B}\|_\infty < K$ by Assumption 8 and $\sup_{1\leq i.j\leq n}|m_{ij}| < K$, which is proved in (A.4). Since we have also shown in the proof of (iii) that $\|\mathbf{B}\mathbf{G}\|_\infty < K$, applying Lemma A.6(ii) and Lemma A.4(iii) leads to $Tr\left[\left(\mathbf{B}\mathbf{G}\right)\mathbf{G}'\mathbf{P}\right] = O\left(n^{\delta+1}\right)$. Similarly, the remaining three traces in (A.5) can be shown to be $O\left(n^{\delta+1}\right)$ by applying Lemmas A.1, A.2 and A.4. Consequently, $Var\left(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon}\right) = O\left(n^{\delta-1}\right)$, and we obtain $n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{C}\mathbf{G}\boldsymbol{\varepsilon} \to_p 0$, if $\delta < 1$.

(vii) It is easily seen that $E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\eta}\right) = 0$ and $Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\eta}\right) = n^{-2}\sigma^2 Tr\left(\mathbf{C}\mathbf{G}\mathbf{G}'\mathbf{C}\tilde{\mathbf{M}}\right)$, where as before $\tilde{\mathbf{M}} = (\tilde{m}_{ij}) = \mathbf{G}\mathbf{M}\mathbf{G}'$ and $\mathbf{M}$ is defined by (A.3). We have shown in the proof of (v) that $\sup_{1\leq i.j\leq n}|\tilde{m}_{ij}| < K$. Then by similar line of reasoning applied to (A.5), it follows that $Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\eta}\right) = O\left(n^{\delta-1}\right)$. Therefore, $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{C}\boldsymbol{\eta} \to_p 0$, if $\delta < 1$. ∎

Lemma A.8 *Suppose that Assumptions 4 and 5 in the paper hold. Consider* $\mathbf{G} = \mathbf{G}(\rho) = \mathbf{W}\left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}$, *where* $|\rho| < 1$. *Then*

$$\mathbf{1}_n'\mathbf{G}'\mathbf{1}_n = \mathbf{1}_n'\mathbf{G}\mathbf{1}_n = \frac{n}{1-\rho}, \tag{A.6}$$

$$\mathbf{1}_n'\mathbf{G}'\mathbf{G}\mathbf{1}_n = \frac{n}{(1-\rho)^2}, \tag{A.7}$$

$$Tr\left(n^{-1}\mathbf{G}^s\right) \leq K, \text{ for } s = 1, 2, \ldots, \tag{A.8}$$

$$Tr\left(n^{-1}\mathbf{G}'\mathbf{G}\right) \leq K. \tag{A.9}$$

Proof. First note that since $\mathbf{W}$ is a row-standardized stochastic matrix and $|\rho| < 1$, then

$$\mathbf{G}\mathbf{1}_n = \left(\mathbf{W} + \rho\mathbf{W}^2 + \rho^2\mathbf{W}^3 + \ldots\right)\mathbf{1}_n = \left(\frac{1}{1-\rho}\right)\mathbf{1}_n, \tag{A.10}$$

and (A.6) and (A.7) follow. Denote the diagonal elements of $\mathbf{G}^s$ by $g_{s,ii}$ and note that

$$Tr\left(n^{-1}\mathbf{G}^s\right) \leq \left|Tr\left(n^{-1}\mathbf{G}^s\right)\right| \leq n^{-1}\sum_{i=1}^{n}\left|g_{s,ii}\right|.$$

Also by Lemma A.6, $\|\mathbf{G}\|_\infty < K$, and since $\|\mathbf{G}^s\|_\infty \leq (\|\mathbf{G}\|_\infty)^s$, then $\|\mathbf{G}^s\|_\infty < K$. Hence, all elements of $\mathbf{G}^s$ must be bounded, specifically $|g_{s,ii}| < K$ and result (A.8) follows. Finally,

$$Tr\left(\mathbf{G}'\mathbf{G}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}^2 \leq \sum_{i=1}^{n}\left(\sum_{j=1}^{n}|g_{ij}|\right)^2,$$

but by Lemma A.6, $\sup_i \sum_{j=1}^{n}|g_{ij}| \leq K$, and hence $Tr\left(\mathbf{G}'\mathbf{G}\right) \leq nK$, and the result (A.9) follows. ■

Lemma A.9 *Suppose that Assumptions 1–5 in the paper hold. Let* $\mathbf{G} = \mathbf{G}(\rho) = \mathbf{W}\left(\mathbf{I}_n - \rho\mathbf{W}\right)^{-1}$, $\boldsymbol{\eta} = \mathbf{G}\mathbf{X}\boldsymbol{\beta}$, *and* $\mathbf{M}_x = \mathbf{I}_n - \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$. *Then*

$$E\left(n^{-1}\boldsymbol{\eta}'\mathbf{M}_x\boldsymbol{\eta}\right) = O(1), \tag{A.11}$$

$$E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) = n^{-1}\sigma^2 Tr\left(\mathbf{G}'\mathbf{M}_x\right) = O(1), \tag{A.12}$$

$$Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\eta}\right) = O\left(n^{-1}\right), \tag{A.13}$$

$$Var\left(n^{-1}\boldsymbol{\eta}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) = O\left(n^{-1}\right), \tag{A.14}$$

$$Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) = O\left(n^{-1}\right), \tag{A.15}$$

Proof. Let $\mathbf{G} = (g_{ij})$ and $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots, \xi_n)' = \mathbf{X}\boldsymbol{\beta}$. Note that $\boldsymbol{\eta} = \boldsymbol{G}\boldsymbol{\xi}$ and under the assumptions $|E(\xi_i\xi_j)| = |\sigma_\xi(i,j)| < K$, for all i and j. Since $\mathbf{M}_x$ is an idempotent matrix, we have

$$n^{-1}\boldsymbol{\eta}'\mathbf{M}_x\boldsymbol{\eta} \leq n^{-1}\boldsymbol{\xi}'\mathbf{G}'\mathbf{G}\boldsymbol{\xi} = n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j\sum_{s=1}^{n}g_{si}g_{sj} = n^{-1}\sum_{s=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\xi_i\xi_j g_{si}g_{sj},$$

and $n^{-1}E\left(\boldsymbol{\eta}'\mathbf{M}_x\boldsymbol{\eta}\right) \leq n^{-1}\sum_s\sum_i\sum_j g_{si}g_{sj}\sigma_\xi(i,j)$. Now noting that $|\sigma_\xi(i,j)| < K$, then $n^{-1}E\left(\boldsymbol{\eta}'\mathbf{M}_x\boldsymbol{\eta}\right) \leq Kn^{-1}\sum_s\left(\sum_i|g_{si}|\right)^2$, and noting from Lemma A.6 that $\sum_i|g_{si}| \leq K$, we obtain (A.11). To establish (A.12) note that

$$E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) = n^{-1}E\left[Tr\left(\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\right)\right] = n^{-1}\sigma^2 Tr\left(\mathbf{M}_x\mathbf{G}\right).$$

Applying Cauchy-Schwarz inequality to $Tr\left(\mathbf{M}_x\mathbf{G}\right)$ we have

$$E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) = n^{-1}\sigma^2 Tr\left(\mathbf{M}_x\mathbf{G}\right) \leq n^{-1}\sigma^2\sqrt{Tr\left(\mathbf{M}_x'\mathbf{M}_x\right)Tr\left(\mathbf{G}'\mathbf{G}\right)} = n^{-1}\sigma^2\sqrt{(n-k)Tr\left(\mathbf{G}'\mathbf{G}\right)},$$

and in view of (A.9) we have $E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) = n^{-1}\sigma^2 Tr\left(\mathbf{M}_x\mathbf{G}\right) < K$. Also since $E\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\eta}\right) = 0$, then

$$Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\eta}\right) = n^{-2}\sigma^2 E\left[\boldsymbol{\xi}'\left(\mathbf{G}'\mathbf{M}_x\mathbf{G}\right)^2\boldsymbol{\xi}\right] = O(n^{-1}),$$

which establishes (A.13), and similarly (A.14). Consider now $Var\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}\mathbf{M}_x\boldsymbol{\varepsilon}\right) = n^{-2}Var\left(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}\right)$, where $\mathbf{A} = \mathbf{G}\mathbf{M}_x$. Using Lemma A.5 we have

$$Var(\boldsymbol{\varepsilon}'\mathbf{A}\boldsymbol{\varepsilon}) = \gamma_2 \sum_{i=1}^{N} a_{ii}^2 + \sigma^4 \left[Tr(\mathbf{A}\mathbf{A}') + Tr(\mathbf{A}^2)\right] \leq KTr\left(\mathbf{A}\mathbf{A}'\right),$$

where $\gamma_2 = \mu_4 - 3\sigma^4$. But using result (12) in Lütkepohl (1996, p.44) yields

$$Tr(\mathbf{A}\mathbf{A}') = Tr\left(\mathbf{G}'\mathbf{M}_x\mathbf{G}\right) \leq \lambda_{\max}\left(\mathbf{M}_x\right) Tr\left(\mathbf{G}'\mathbf{G}\right) = Tr\left(\mathbf{G}'\mathbf{G}\right),$$

and using (A.9) we have $Tr(\mathbf{A}\mathbf{A}') \leq Tr\left(\mathbf{G}'\mathbf{G}\right) \leq nK$. Then the result (A.15) follows. ■

Lemma A.10 *Let* $\{X_{in},\ 1 \leq i \leq k_n,\ n \geq 1\}$ *be a martingale difference array with respect to the filtration* $\mathfrak{F}_{i-1,n}^x = \sigma\left[(X_{jn})_{j=1}^{i-1}\right]$. *Suppose that (a)* X_{in} *is square integrable, (b)* $\sum_{i=1}^{k_n} E\left|X_{in}\right|^{2+\epsilon} \to 0$, *and (c)* $\sum_{i=1}^{k_n} E(X_{in}^2|\mathfrak{F}_{i-1,n}^x) \to_p 1$. *Then* $\sum_{i=1}^{k_n} X_{in} \to_d N(0,1)$.

Proof. See Corollary 3.1 of Hall and Heyde (1980).[A2] ■

A.2 Proofs of theorems and propositions

The following proofs make use of the lemmas in Section A.1 of this online mathematical appendix. Note that the elements of the matrix and variables in the theorems and propositions may depend on sample size n and form triangular arrays, but we suppress subscript n in the proofs for notational simplicity.

Proof of Theorem 1. We first consider ω_n^2 given by (15) in the paper and show that ω_n^2 is bounded. Note that (11) in the paper implies that p_{ij} (or p_{ji}) must all be bounded in n. By definition, $a_{ij} = \left(p_{ij} + p_{ji}\right)/2$, and hence $\sup_{i,j}\left|a_{ij}\right| \leq \left(\sup_{i,j}\left|p_{ij}\right| + \sup_{i,j}\left|p_{ji}\right|\right)/2 < K$. Using (15) given in the paper we now have

$$\omega_n^2 \leq K \sup_n \left|\mu_4 - 3\sigma^4\right| + 2\left(\sup_n \sigma^4\right)\left[Tr\left(n^{-1}\mathbf{A}^2\right)\right].$$

Furthermore,

$$\begin{aligned} Tr\left(n^{-1}\mathbf{A}^2\right) &= \frac{1}{4n}\left[Tr\left(\mathbf{P}^2\right) + Tr\left(\mathbf{P}'^2\right) + 2Tr\left(\mathbf{P}'\mathbf{P}\right)\right] \\ &= \frac{1}{2}\left[Tr\left(n^{-1}\mathbf{P}^2\right) + Tr\left(n^{-1}\mathbf{P}'\mathbf{P}\right)\right], \end{aligned}$$

and

$$Tr\left(\mathbf{P}'\mathbf{P}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}^2 \leq \sum_{i=1}^{n}\left(\sum_{j=1}^{n}\left|p_{ij}\right|\right)^2 \leq \sum_{i=1}^{n}\left(\sup_i \sum_{j=1}^{n}\left|p_{ij}\right|\right)^2.$$

[A2] Condition (b) in Theorem A.10 is a sufficient condition for the conditional Lindeberg condition (3.7) in Corollary 3.1 (see Davidson, 1994, Theorem 23.11).

But under (11) of the paper, $\sup_i \sum_{j=1}^{n} |p_{ij}| < K$, and we have $Tr\left(n^{-1}\mathbf{P}'\mathbf{P}\right) \leq K$, which also implies that $Tr\left(n^{-1}\mathbf{P}^2\right) < K$. Hence, ω_n^2 is bounded in n for all values of $0 \leq \delta \leq 1$. Also note that condition (13) in the paper ensures that $\omega_n^2 > 0$, for all n (including $n \to \infty$).

Consider Q defined by (14) in the paper and following Kelejian and Prucha (2001) write it as $Q = \sum_{i=1}^{n} X_i$, where

$$X_i = \omega_n^{-1} n^{-1/2} a_{ii}\left(\varepsilon_i^2 - \sigma^2\right) + 2\omega_n^{-1} n^{-1/2} \varepsilon_i \zeta_{i-1}, \tag{A.16}$$

and

$$\zeta_{i-1} = \sum_{j=1}^{i-1} a_{ij}\varepsilon_j. \tag{A.17}$$

Clearly, $E\left(X_i\right) = 0$ and

$$\begin{aligned} E\left(X_i^2\right) &= \omega_n^{-2} n^{-1} E\left[a_{ii}\left(\varepsilon_i^2 - \sigma^2\right) + 2\varepsilon_i\zeta_{i-1}\right]^2 \\ &= \omega_n^{-2} n^{-1} E\left[a_{ii}^2\left(\varepsilon_i^4 + \sigma^4 - 2\varepsilon_i^2\sigma^2\right) + 4\varepsilon_i^2\zeta_{i-1}^2 + 4a_{ii}\left(\varepsilon_i^2 - \sigma^2\right)\varepsilon_i\zeta_{i-1}\right] \\ &= \omega_n^{-2} n^{-1}\left[a_{ii}^2\left(\mu_4 - \sigma^4\right) + 4\sigma^4 \sum_{j=1}^{i-1} a_{ij}^2\right]. \end{aligned} \tag{A.18}$$

Notice that (15) in the paper can be written equivalently as

$$\omega_n^2 = \left(\mu_4 - \sigma^4\right) n^{-1} \sum_{i=1}^{n} a_{ii}^2 + 4n^{-1}\sigma^4 \sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ij}^2 > 0 \tag{A.19}$$

Using (A.18) and (A.19) leads to $\sum_{i=1}^{n} E\left(X_i^2\right) = 1$. Note that $\{X_i, 1 \leq i \leq n\}$ forms a martingale difference array with respect to the filtration $\mathfrak{F}_{i-1}^{\varepsilon} = \sigma\left[\left(\varepsilon_j\right)_{j=1}^{i-1}\right]$ (with $\mathfrak{F}_0^{\varepsilon} = \{\emptyset, \Omega\}$). Since X_{i-1} depends on $\{\varepsilon_j\}_{j=1}^{i-1}$, it is readily seen that $\{X_i\}$ is a martingale difference array with respect to the filtration $\mathfrak{F}_{i-1}^{x} = \sigma\left[\left(X_j\right)_{j=1}^{i-1}\right]$. Hence, the central limit theorem given in Lemma A.10 is applicable to Q if the three conditions on $\left\{X_i, \mathfrak{F}_{i-1}^{x}\right\}$ can be established. Since we have shown that $\sum_{i=1}^{n} E\left(X_i^2\right) = 1$, and $E\left(X_i^2\right) \geq 0$ for all i, it follows that $E\left(X_i^2\right) \leq 1$, and hence X_i^2 is square integrable for all values of $0 \leq \delta \leq 1$. In what follows, we only need to show that conditions (b) and (c) of Lemma A.10 hold under $0 \leq \delta < 1/2$.

We now consider condition (b) of Lemma A.10. Let $q = 2 + \nu$, where $0 < \nu \leq \epsilon/2$. Then by Minkowski's inequality,

$$\begin{aligned} E|X_i|^q &= \omega_n^{-q} n^{-\frac{q}{2}} E\left|a_{ii}\left(\varepsilon_i^2 - \sigma^2\right) + 2\varepsilon_i\zeta_{i-1}\right|^q \\ &\leq \omega_n^{-q} n^{-\frac{q}{2}}\left[|a_{ii}|\left(E\left|\varepsilon_i^2 - \sigma^2\right|^q\right)^{1/q} + 2\left(E|\varepsilon_i|^q E\left|\zeta_{i-1}\right|^q\right)^{1/q}\right]^q \\ &\leq \omega_n^{-q} n^{-\frac{q}{2}}\left[|a_{ii}|\left(E\left|\varepsilon_i^2 - \sigma^2\right|^q\right)^{1/q} + 2\left(\sum_{j=1}^{i-1}|a_{ij}|\right)\left(E|\varepsilon_i|^q E\left|\varepsilon_j\right|^q\right)^{1/q}\right]^q. \end{aligned}$$

Since $\sup_i E\left|\varepsilon_i\right|^{4+\epsilon} < K$, we have $E\left|\varepsilon_i^2 - \sigma^2\right|^{2+\nu} \leq K$ and $E|\varepsilon_i|^{2+\nu} \leq K$ for all i, and it follows that

$$E|X_i|^{2+\nu} \leq \omega_n^{-(2+\nu)} n^{-\frac{2+\nu}{2}} K \left[|a_{ii}| + 2 \left(\sum_{j=1}^{i-1} |a_{ij}| \right) \right]^{2+\nu}$$
$$\leq \omega_n^{-(2+\nu)} n^{-\frac{2+\nu}{2}} K \left(\sum_{j=1}^{n} |a_{ij}| \right)^{2+\nu},$$

and

$$\sum_{i=1}^{n} E\,|X_i|^{2+\nu} \leq \omega_n^{-(2+\nu)} n^{-\frac{2+\nu}{2}} K \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}| \right)^{2+\nu}.$$

Using the definition,

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}| \right)^{2+\nu} = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} \frac{|p_{ij} + p_{ji}|}{2} \right)^{2+\nu}$$
$$\leq 2^{-(2+\nu)} \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |p_{ij}| + \sum_{j=1}^{n} |p_{ji}| \right)^{2+\nu},$$

and applying Loeve's c_r-inequality,[A3]

$$\left(\sum_{j=1}^{n} |p_{ij}| + \sum_{j=1}^{n} |p_{ji}| \right)^{2+\nu} \leq 2^{(2+\nu)-1} \left[\left(\sum_{j=1}^{n} |p_{ij}| \right)^{2+\nu} + \left(\sum_{j=1}^{n} |p_{ji}| \right)^{2+\nu} \right],$$

therefore we have

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}| \right)^{2+\nu} \leq \frac{1}{2} \left[\sum_{i=1}^{n} \left(\sum_{j=1}^{n} |p_{ij}| \right)^{2+\nu} + \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |p_{ji}| \right)^{2+\nu} \right].$$

But under assumption (11) in the paper, $\sum_{i=1}^{n} \left(\sum_{j=1}^{n} |p_{ij}| \right)^{2+\nu} = O(n)$. Also, letting m denote the number of unbounded columns of $\mathbf{P}_n$ and noting that m is finite by assumption, we obtain from (12) in the paper that

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{n} |p_{ji}| \right)^{2+\nu} \leq Kmn^{\delta(2+\nu)} + K(n-m) = O\left\{ n^{\max[\delta(2+\nu),1]} \right\}.$$

Hence, $\sum_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}| \right)^{2+\nu} = O\left\{ n^{\max\{[\delta(2+\nu),1]} \right\}$, and then

$$\sum_{i=1}^{n} E\,|X_i|^{2+\nu} \leq \omega_n^{-(2+\nu)} n^{-\frac{2+\nu}{2}} K \sum_{i=1}^{n} \left(\sum_{j=1}^{n} |a_{ij}| \right)^{2+\nu} = O\left\{ n^{-\frac{2+\nu}{2} + \max[\delta(2+\nu),1]} \right\},$$

[A3] See, for example, Davidson (1994), p. 140.

or equivalently,

$$\sum_{i=1}^{n} E\left|X_i\right|^{2+\nu} = \begin{cases} O\left(n^{-\frac{\nu}{2}}\right), & \text{if } \delta \le \frac{1}{2+\nu}, \\ O\left[n^{(\delta-\frac{1}{2})(2+\nu)}\right], & \text{if } \delta > \frac{1}{2+\nu}. \end{cases}$$

Therefore, $\sum_{i=1}^{n} E\left|X_i\right|^{2+\nu}$ converges to zero if $0 \le \delta < 1/2$, and this completes the proof of condition (b).

We now turn to establishing condition (c) of Lemma A.10. Note that

$$E\left(X_i^2|\mathfrak{F}_{i-1}^x\right) = \frac{a_{ii}^2\left(\mu_4 - \sigma^4\right)}{n\omega_n^2} + \frac{4\sigma^2\zeta_{i-1}^2}{n\omega_n^2} + \frac{4a_{ii}\mu_3\zeta_{i-1}}{n\omega_n^2},$$

and it follows that

$$\begin{aligned}
\sum_{i=1}^{n} E\left(X_i^2|\mathfrak{F}_{i-1}^x\right) - 1 &= \frac{\left(\sum_{i=1}^n a_{ii}^2\right)\left(\mu_4 - \sigma^4\right)}{n\omega_n^2} + \frac{4\sigma^2\sum_{i=1}^n \zeta_{i-1}^2}{n\omega_n^2} \\
&\quad + \frac{4\mu_3\sum_{i=1}^n a_{ii}\zeta_{i-1}}{n\omega_n^2} - \frac{\left(\mu_4 - \sigma^4\right)\sum_{i=1}^n a_{ii}^2 + 4\sigma_n^4\sum_{i=1}^n\sum_{j=1}^{i-1} a_{ij}^2}{n\omega_n^2} \\
&= \frac{4\sigma^2\left[\sum_{i=1}^n \zeta_{i-1}^2 - \sigma^2\sum_{i=1}^n\sum_{j=1}^{i-1} a_{ij}^2\right]}{n\omega_n^2} + \frac{4\mu_3\sum_{i=1}^n a_{ii}\zeta_{i-1}}{n\omega_n^2} \\
&= \omega_n^{-2}\left(8H_1 + 4H_2 + 4H_3\right),
\end{aligned}$$

where

$$H_1 = n^{-1}\sigma^2\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1} a_{ij}a_{ik}\varepsilon_j\varepsilon_k, \tag{A.20}$$

$$H_2 = n^{-1}\sigma^2\sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ij}^2\left(\varepsilon_j^2 - \sigma^2\right), \tag{A.21}$$

$$H_3 = n^{-1}\mu_3\sum_{i=1}^{n}\sum_{j=1}^{i-1} a_{ii}a_{ij}\varepsilon_j. \tag{A.22}$$

We need to show that H_s, for $s = 1, 2, 3$, tend to zero in probability as $n \to \infty$. For H_1, we have

$$H_1^2 = n^{-2}\sigma^4\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1}\sum_{l=1}^{n}\sum_{r=1}^{l-1}\sum_{s=1}^{r-1} a_{ij}a_{ik}a_{lr}a_{ls}\varepsilon_j\varepsilon_k\varepsilon_r\varepsilon_s.$$

Note that $E\left(\varepsilon_j\varepsilon_k\varepsilon_r\varepsilon_s\right) \ne 0$ only if $(j = r) \ne (k = s)$ or $(j = s) \ne (k = r)$, since $k \ne j$, $s \ne r$. Therefore,

$$\begin{aligned}
E\left(H_1^2\right) &= 2n^{-2}\sigma^8\sum_{l=1}^{n}\sum_{i=1}^{l}\sum_{j=1}^{i-1}\sum_{k=1}^{j-1} a_{ij}a_{ik}a_{lj}a_{lk} \\
&\le 2n^{-2}\sigma^8\sum_{l=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \left|a_{ij}\right|\left|a_{ik}\right|\left|a_{lj}\right|\left|a_{lk}\right|
\end{aligned}$$

$$
\begin{aligned}
&\leq n^{-2}\sigma^8 \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\sum_{l=1}^{n}|a_{lj}|\left(\sum_{k=1}^{n}|a_{ik}||p_{lk}|+\sum_{k=1}^{n}|a_{ik}||p_{kl}|\right)\\
&\leq n^{-2}\sigma^8 \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\sum_{l=1}^{n}|a_{lj}|\left(\sup_{1\leq l\leq n}\sum_{k=1}^{n}|p_{lk}|\right)\left(\sup_{1\leq i,k\leq n}|a_{ik}|\right)\\
&\quad + n^{-2}\sigma^8 \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\sum_{k=1}^{n}|a_{ik}|\left(\sup_{1\leq k\leq n}\sum_{l=1}^{n}|p_{kl}|\right)\left(\sup_{1\leq l,j\leq n}|a_{lj}|\right)\\
&\leq Kn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\left(\sup_{1\leq j\leq n}\sum_{l=1}^{n}|a_{lj}|+\sup_{1\leq i\leq n}\sum_{k=1}^{n}|a_{ik}|\right)\leq Kn^{\delta-1}.
\end{aligned}
$$

Noting also that $E(H_1)=0$, by Markov's inequality we conclude that $H_1=o_p(1)$ if $\delta<1$. Turning next to H_2. We have $E(H_2)=0$ and

$$
\begin{aligned}
H_2^2 &= n^{-2}\sigma^4\sum_{i=1}^{n}\sum_{j=1}^{i-1}a_{ij}^2\left(\varepsilon_j^2-\sigma^2\right)\sum_{k=1}^{n}\sum_{l=1}^{k-1}a_{kl}^2\left(\varepsilon_l^2-\sigma^2\right)\\
&= n^{-2}\sigma^4\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}\sum_{l=1}^{k-1}a_{ij}^2a_{kl}^2\varepsilon_j^2\varepsilon_l^2+n^{-2}\sigma^8\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}\sum_{l=1}^{k-1}a_{ij}^2a_{kl}^2\\
&\quad - n^{-2}\sigma^6\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}\sum_{l=1}^{k-1}a_{ij}^2a_{kl}^2\varepsilon_j^2-n^{-2}\sigma^6\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}\sum_{l=1}^{k-1}a_{ij}^2a_{kl}^2\varepsilon_l^2,
\end{aligned}
$$

which leads to

$$
\begin{aligned}
E\left(H_2^2\right) &= n^{-2}\sigma^4\left(\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}a_{ij}^2a_{kj}^2\mu_4+\sigma^4\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}\sum_{l=1,l\neq j}^{k-1}a_{ij}^2a_{kl}^2\right)\\
&\quad - n^{-2}\sigma^8\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}\sum_{l=1}^{k-1}a_{ij}^2a_{kl}^2\\
&= n^{-2}\sigma^4\left(\mu_4-\sigma^4\right)\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n}a_{ij}^2a_{kj}^2\\
&\leq n^{-2}\sigma^4\left(\mu_4-\sigma^4\right)\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2\left(\sup_{1\leq j\leq n}\sum_{k=1}^{n}|a_{kj}|\right)\left(\sup_{1\leq k,j\leq n}|a_{kj}|\right)\\
&\leq Kn^{\delta-1},
\end{aligned}
$$

where in the last line we used $n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}a_{ij}^2=Tr\left(n^{-1}\mathbf{A}'\mathbf{A}\right)<K$. Thus, we obtain that $H_2=o_p(1)$ if $\delta<1$. Lastly, $E(H_3)=0$, and

$$
H_3^2=n^{-2}\mu_3^2\sigma^2\sum_{i=1}^{n}\sum_{j=1}^{i-1}a_{ij}\varepsilon_j\sum_{k=1}^{n}\sum_{l=1}^{k-1}a_{kk}a_{kl}\varepsilon_l,
$$

and it follows that

$$E\left(H_3^2\right) = n^{-2}\mu_3^2\sigma^2 \sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{k=1}^{n} a_{ii}a_{ij}a_{kk}a_{kj}$$
$$\leq n^{-2}\mu_3^2\sigma^2 \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} |a_{ii}|\,|a_{ij}|\,|a_{kk}|\,|a_{kj}|$$
$$\leq Kn^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ii}|\,|a_{ij}|\left(\sup_{1\leq j\leq n}\sum_{k=1}^{n}|a_{kj}|\right)\left(\sup_{1\leq k\leq n}|a_{kk}|\right) \leq Kn^{\delta-1}.$$

Hence, $H_3 = o_p(1)$ if $\delta < 1$. Overall, we conclude that $\sum_{i=1}^{n} E(X_i^2|\mathfrak{F}_{i-1}^x) \to_p 1$ if $0 \leq \delta < 1$, which proves condition (c) of Lemma A.10. Combining our findings for (a)–(c) establishes the result in (14) of the paper under $0 \leq \delta < 1/2$. ■

Proof of Theorem 2. We begin by showing that $\tilde{\omega}_n^2$, which is defined in (18) of the paper, is bounded in n for all $0 \leq \delta \leq 1$. Note that

$$\tilde{\omega}_n^2 = \omega_n^2 + \sigma^2 n^{-1}\sum_{i=1}^{n}\mu_{2\eta,i} + 2\mu_3 n^{-1}\sum_{i=1}^{n} a_{ii}\mu_{\eta,i},$$

where ω_n^2 is defined by (15) in the paper. We have shown in the above proof of Theorem 1 that ω_n^2 is bounded in n for all $0 \leq \delta \leq 1$, and since $\sigma_{\eta,i}^2 < K$, $\mu_{\eta,i} < K$, $\mu_3 < K$, and $|a_{ii}| \leq |p_{ii}| < K$ for all i, it is immediate that $\tilde{\omega}_n^2$ is bounded in n for $0 \leq \delta \leq 1$. Also note that condition (16) in the paper implies that $\tilde{\omega}_n^2 > 0$, for all n (including $n \to \infty$).

Consider $\tilde{Q}$ defined by (17) in the paper and write it as $\tilde{Q} = \sum_{i=1}^{n} Y_i$, where

$$Y_i = \tilde{\omega}_n^{-1} n^{-1/2} a_{ii}\left(\varepsilon_i^2 - \sigma^2\right) + 2\tilde{\omega}_n^{-1} n^{-1/2}\varepsilon_i\zeta_{i-1} + \tilde{\omega}_n^{-1} n^{-1/2}\eta_i\varepsilon_i,$$

and ζ_{i-1} is defined in (A.17). It is easy to check that $\{Y_i, 1 \leq i \leq n\}$ forms a martingale difference array with respect to the filtration $\mathfrak{F}_{i-1}^{\eta,\varepsilon} = \sigma\left[(\eta_j)_{j=1}^{i-1}, (\varepsilon_j)_{j=1}^{i-1}\right]$ (with $\mathfrak{F}_0^{\eta,\varepsilon} = \{\emptyset, \Omega\}$), and therefore $\{Y_i\}$ is also a martingale difference array with respect to the filtration $\mathfrak{F}_{i-1}^y = \sigma\left[(Y_j)_{j=1}^{i-1}\right]$. To apply the central limit theorem given by Lemma A.10, we need to show in turn that the three conditions (a)–(c) are satisfied for $\left\{Y_i, \mathfrak{F}_{i-1}^y\right\}$.

First, we see that

$$E\left(Y_i^2\right) = \tilde{\omega}_n^{-2} n^{-1}\left[a_{ii}^2\left(\mu_4 - \sigma^4\right) + 4\sigma^4\sum_{j=1}^{i-1} a_{ij}^2 + \sigma^2\mu_{2\eta,i} + 2\mu_3 a_{ii}\mu_{\eta,i}\right].$$

Using (18) in the paper we obtain $\sum_{i=1}^{n} E\left(Y_i^2\right) = 1$. Since $E\left(Y_i^2\right) \geq 0$ for all i, we readily have $E\left(Y_i^2\right) \leq 1$ and hence Y_i is square integrable.

Turning to condition (b). Notice that Y_i can be rewritten as $Y_i = \tilde{\omega}_n^{-1} n^{-1/2}\left(Y_{1,i} + Y_{2,i}\right)$, where $Y_{1,i} = a_{ii}\left(\varepsilon_i^2 - \sigma^2\right) + 2\varepsilon_i\zeta_{i-1}$, and $Y_{2,i} = \eta_i\varepsilon_i$. Applying the c_r−inequality, we have

$$\sum_{i=1}^{n} E|Y_i|^{2+\epsilon} = \tilde{\omega}_n^{-(2+\epsilon)} n^{-\frac{2+\epsilon}{2}} \sum_{i=1}^{n} E\left|Y_{1,i} + Y_{2,i}\right|^{2+\epsilon}$$
$$\leq 2^{1+\epsilon}\tilde{\omega}_n^{-(2+\epsilon)} n^{-\frac{2+\epsilon}{2}} \sum_{i=1}^{n} \left(E\left|Y_{1,i}\right|^{2+\epsilon} + E\left|Y_{2,i}\right|^{2+\epsilon}\right).$$

Since

$$\sum_{i=1}^{n} E\left|Y_{2,i}\right|^{2+\epsilon} = \sum_{i=1}^{n} E\left|\eta_i\varepsilon_i\right|^{2+\epsilon} \leq n \sup_i E\left(|\varepsilon_i|^{2+\epsilon}\right) \sup_j E\left(|\eta_j|^{2+\epsilon}\right) \leq Kn,$$

it follows that $n^{-\frac{2+\epsilon}{2}} \sum_{i=1}^{n} E\left|Y_{2,i}\right|^{2+\epsilon} = O\left(n^{-\frac{\epsilon}{2}}\right)$, which converges to zero for all values of $0 \leq \delta \leq 1$. In addition, note that $Y_{1,i} = X_i n^{1/2}\omega_n$, where X_i is defined in (A.16). As we have shown in the proof of Theorem 1 that $\sum_{i=1}^{n} E\left|X_i\right|^{2+\epsilon} \to 0$ if $0 \leq \delta < 1/2$, we immediately obtain that $n^{-\frac{2+\epsilon}{2}} \sum_{i=1}^{n} E|Y_{1,i}|^{2+\epsilon} \to 0$ if $0 \leq \delta < 1/2$. Thus, overall we have $\sum_{i=1}^{n} E|Y_i|^{2+\epsilon} \to 0$ if $0 \leq \delta < 1/2$, and this completes the proof of condition (b).

Now it remains to establish condition (c): $\sum_{i=1}^{n} E\left(Y_i^2|\mathfrak{F}_{i-1}^y\right) \to_p 1$. Note that

$$\sum_{i=1}^{n} E\left(Y_i^2|\mathfrak{F}_{i-1}^y\right) - 1 = \tilde{\omega}_n^{-2}\left(8H_1 + 4H_2 + 4H_3 + 4H_4\right),$$

where H_s, $s = 1,2,3$, are given by (A.20)–(A.22), respectively, and $H_4 = n^{-1}\sigma^2 \sum_{i=1}^{n} \eta_i\zeta_{i-1}$. Since $E\left(H_4\right) = 0$ and

$$Var\left(H_4\right) = n^{-2}\sigma^4 \sum_{i=1}^{n} E\left(\eta_i^2\right) E\left(\zeta_{i-1}^2\right) \leq \sigma^4 \sup_i E\left(\eta_i^2\right)\left[n^{-2}\sum_{i=1}^{n} E\left(\zeta_{i-1}^2\right)\right] \leq Kn^{-1},$$

we have $H_4 \to_p 0$. As it has been shown in the proof of Theorem 1 that $H_s \to_p 0$, for $s = 1,2,3$, if $0 \leq \delta < 1$, overall we conclude that $\sum_{i=1}^{n} E\left(Y_i^2|\mathfrak{F}_{i-1}^y\right) \to_p 1$ if $0 \leq \delta < 1$. Combining conditions (a)–(c), Lemma A.10 is applicable and the result in (17) of the paper is established under $0 \leq \delta < 1/2$. ■

Proof of Proposition 1. Let us first consider the estimator defined by (27) in the paper using a single quadratic moment. We can rewrite $\boldsymbol{\varepsilon}\left(\rho\right)$, given by (29) in the paper, as

$$\boldsymbol{\varepsilon}\left(\rho\right) = \boldsymbol{\varepsilon} - \left(\rho - \rho_0\right)\mathbf{G}_0\boldsymbol{\varepsilon}. \tag{A.23}$$

Substituting (A.23) into $g_n\left(\rho\right)$, which is given by (28) in the paper, yields

$$g_n\left(\rho\right) = n^{-1}\left[\boldsymbol{\varepsilon} - \left(\rho - \rho_0\right)\mathbf{G}_0\boldsymbol{\varepsilon}\right]'\mathbf{C}\left[\boldsymbol{\varepsilon} - \left(\rho - \rho_0\right)\mathbf{G}_0\boldsymbol{\varepsilon}\right]$$
$$= \frac{\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}}{n} + \left(\rho - \rho_0\right)^2 \boldsymbol{\varepsilon}'\left(\frac{\mathbf{G}_0'\mathbf{C}\mathbf{G}_0}{n}\right)\boldsymbol{\varepsilon} - 2\left(\rho - \rho_0\right)\boldsymbol{\varepsilon}'\left(\frac{\mathbf{G}_0'\mathbf{C}}{n}\right)\boldsymbol{\varepsilon}. \tag{A.24}$$

Since $Tr\left(\mathbf{C}\right) = Tr\left(\mathbf{B}\right) = 0$ under Assumption 8, we have $E_0\left(\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}\right) = \sigma_0^2 Tr\left(\mathbf{C}\right) = 0$. Using the results in Lemma A.7(i)–(iii), we obtain

$$g_n(\rho) = (\rho - \rho_0)^2 \sigma_0^2 a_0 - 2(\rho - \rho_0)\sigma_0^2 b_0 + o_p(1),$$

where $a_0 = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}\mathbf{G}_0\right)$ and $b_0 = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}\right)$. Note that $g_n(\rho_0) = n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}$. Using (A.24), it follows that

$$g_n(\rho) - g_n(\rho_0) = (\rho - \rho_0)^2 \sigma_0^2 a_0 - 2(\rho - \rho_0)\sigma_0^2 b_0 + o_p(1).$$

Since $\tilde{\rho}$ is such that $g_n(\tilde{\rho}) \leq g_n(\rho_0)$, or equivalently $(\rho - \rho_0)^2 \sigma_0^2 a_0 - 2(\rho - \rho_0)\sigma_0^2 b_0 \leq 0$, then we will have global identification if $b_0 = 0$ and $a_0 \neq 0$. In this case, $(\tilde{\rho} - \rho_0)^2 a_0 \leq 0$, which is satisfied if and only if $\tilde{\rho} = \rho_0$. However, in general where $b_0 \neq 0$, and we must have either $\tilde{\rho} = \rho_0 + o_p(1)$, or $\tilde{\rho} = \rho_0 + 2b_0/a_0 + o_p(1)$. It is clear that ρ_0 is not globally identified if $b_0 \neq 0$.

Now suppose that we use at least two quadratic moments to obtain the GMM estimator. Formally, consider the estimator defined by (30) in the paper using L ($L \geq 2$) quadratic moments. The above arguments for a single quadratic moment readily extends to the case of multiple quadratic moments. Each (population) moment condition will have two solutions: $\tilde{\rho}_{1,\ell} = \rho_0$ and $\tilde{\rho}_{2,\ell} = \rho_0 + 2b_{\ell 0}/a_{\ell 0}$, for $\ell = 1, 2, \ldots, L$, where $a_{\ell 0} = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}_\ell\mathbf{G}_0\right)$ and $b_{\ell 0} = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}_\ell\right)$. Then it is clear that ρ_0 is uniquely identified as long as the ratios, $b_{\ell 0}/a_{\ell 0}$, are not all the same across $\ell = 1, 2, \ldots, L$. ■

Proof of Theorem 3. Consider $\boldsymbol{\varepsilon}(\boldsymbol{\psi})$ given by (20) in the paper. It can be rewritten as

$$\boldsymbol{\varepsilon}(\boldsymbol{\psi}) = \boldsymbol{\varepsilon} - (\rho - \rho_0)\mathbf{G}_0\boldsymbol{\varepsilon} - \mathbf{Q}_0(\boldsymbol{\psi} - \boldsymbol{\psi}_0), \tag{A.25}$$

where $\mathbf{Q}_0 = (\boldsymbol{\eta}_0, \mathbf{X})$ and $\boldsymbol{\eta}_0 = \mathbf{G}_0\mathbf{X}\boldsymbol{\beta}_0$, which is defined by (9) in the paper. Substituting (A.25) into the quadratic term in (25) of the paper and reorganizing yields

$$\begin{aligned} n^{-1}\boldsymbol{\varepsilon}'(\boldsymbol{\psi})\mathbf{C}\boldsymbol{\varepsilon}(\boldsymbol{\psi}) = {} & \frac{\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}}{n} - 2(\rho - \rho_0)\frac{\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{C}\boldsymbol{\varepsilon}}{n} - 2(\rho - \rho_0)\frac{\boldsymbol{\eta}_0'\mathbf{C}\boldsymbol{\varepsilon}}{n} - 2(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\frac{\mathbf{X}'\mathbf{C}\boldsymbol{\varepsilon}}{n} \\ & + (\rho - \rho_0)^2\frac{\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{C}\mathbf{G}_0\boldsymbol{\varepsilon}}{n} + (\rho - \rho_0)^2\frac{\boldsymbol{\eta}_0'\mathbf{C}\boldsymbol{\eta}_0}{n} + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\frac{\mathbf{X}'\mathbf{C}\mathbf{X}}{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ & + 2(\rho - \rho_0)^2\frac{\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{C}\boldsymbol{\eta}_0}{n} + 2(\rho - \rho_0)\frac{\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{C}\mathbf{X}}{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + 2(\rho - \rho_0)\frac{\boldsymbol{\eta}_0'\mathbf{C}\mathbf{X}}{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0). \end{aligned}$$

Using the results in Lemma A.7 and Assumption 7(d), the above equation becomes

$$\begin{aligned} n^{-1}\boldsymbol{\varepsilon}'(\boldsymbol{\psi})\mathbf{C}\boldsymbol{\varepsilon}(\boldsymbol{\psi}) = {} & (\rho - \rho_0)^2\left(\sigma_0^2 a_0 + c_0\right) - 2(\rho - \rho_0)\sigma_0^2 b_0 + 2(\rho - \rho_0)\mathbf{d}_0'(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ & + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\boldsymbol{\Sigma}_{xcx}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1), \end{aligned} \tag{A.26}$$

if $\delta < 1$, where $a_0 = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}\mathbf{G}_0\right)$, $b_0 = \lim_{n\to\infty} Tr\left(n^{-1}\mathbf{G}_0'\mathbf{C}\right)$, $c_0 = p\lim_{n\to\infty} n^{-1}\boldsymbol{\eta}_0'\mathbf{C}\boldsymbol{\eta}_0$, $\mathbf{d}_0' = p\lim_{n\to\infty} n^{-1}\boldsymbol{\eta}_0'\mathbf{C}\mathbf{X}$, and $\boldsymbol{\Sigma}_{xcx} = p\lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{C}\mathbf{X}$. Substituting (A.25) into the linear term in (25) in the paper yields

$$n^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\psi}) = n^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} - (\rho - \rho_0)n^{-1}\mathbf{Z}'\mathbf{G}_0\boldsymbol{\varepsilon} - n^{-1}\mathbf{Z}'\mathbf{Q}_0(\boldsymbol{\psi} - \boldsymbol{\psi}_0)$$

$$= \boldsymbol{\Sigma}_{zq} (\boldsymbol{\psi} - \boldsymbol{\psi}_0) + o_p (1) , \tag{A.27}$$

where $\boldsymbol{\Sigma}_{zq} = p \lim_{n\to\infty} n^{-1}\mathbf{Z}'\mathbf{Q}_0$ and $n^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} = o_p (1)$ readily follow Assumption 7. To see that $n^{-1}\mathbf{Z}'\mathbf{G}_0\boldsymbol{\varepsilon} = o_p (1)$, first note that its mean is zero due to independence of $\mathbf{Z}$ and $\boldsymbol{\varepsilon}$, and we only need to show that its variance is $o (1)$. Let $\mathbf{z}_{.l} = (z_{1l}, z_{2l,...,}z_{nl})'$ denote the l^{th} column of $\mathbf{Z}$, for $l = 1, 2, \ldots , r$. Then

$$Var \left(n^{-1}\mathbf{z}'_{.l}\mathbf{G}_0\boldsymbol{\varepsilon}\right) = E \left[Var \left(n^{-1}\mathbf{z}'_{.l}\mathbf{G}_0\boldsymbol{\varepsilon}|\mathbf{Z}\right)\right] = n^{-2}\sigma_0^2 Tr (\mathbf{G}_0\mathbf{G}'_0\mathbf{M}) ,$$

where $\mathbf{M} = (m_{ij}) = E (\mathbf{z}_{.l}\mathbf{z}'_{.l})$. Since $\sup_{i.j} |m_{ij}| = \sup_{i.j} |E (z_{il}z_{jl})| < K$ under Assumption 7, using Lemma A.4(iii) and Lemma A.6(ii) yields $Tr (\mathbf{G}_0\mathbf{G}'_0\mathbf{M}) = O \left(n^{\delta+1}\right)$ and then $Var (n^{-1}\mathbf{z}'_{.l}\mathbf{G}_0\boldsymbol{\varepsilon}) = O \left(n^{\delta-1}\right)$ for $l = 1, 2, \ldots , r$. Consequently, by Chebyshev's inequality $n^{-1}\mathbf{Z}'\mathbf{G}_0\boldsymbol{\varepsilon}$ converges in mean square and therefore also in probability to zero if $\delta < 1$.

Now combining (A.26) and (A.27), we obtain

$$\mathbf{A}_n\boldsymbol{g}_n (\boldsymbol{\psi}) = \mathbf{A} \begin{bmatrix} (\rho - \rho_0)^2 (\sigma_0^2 a_0 + c_0) - 2 (\rho - \rho_0) \sigma_0^2 b_0 + 2 (\rho - \rho_0) \mathbf{d}'_0 (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_{xcx} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ \boldsymbol{\Sigma}_{zq} (\boldsymbol{\psi} - \boldsymbol{\psi}_0) \end{bmatrix} + o_p (1) ,$$

or alternatively,

$$\mathbf{A}_n\boldsymbol{g}_n (\boldsymbol{\psi}) - \mathbf{A}E_0 [\boldsymbol{g}_n (\boldsymbol{\psi})] = o_p (1) .$$

Under Assumption 7, $\boldsymbol{\Sigma}_{zq}$ has full column rank, then $\boldsymbol{\Sigma}_{zq} (\boldsymbol{\psi} - \boldsymbol{\psi}_0) = 0$ if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Hence, global identification is ensured without the quadratic moment. Moreover, it is readily seen that $\boldsymbol{g}_n (\boldsymbol{\psi})$ converges in probability uniformly in $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ since $\boldsymbol{\Psi}$ is compact and $\boldsymbol{g}_n (\boldsymbol{\psi})$ is a continuous function. Thus, consistency of $\tilde{\boldsymbol{\psi}}$ can be established.

Consider now to the asymptotic distribution of $\tilde{\boldsymbol{\psi}}$. By a mean-value expansion of $\frac{\partial \boldsymbol{g}'_n(\tilde{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}}\mathbf{A}'_n\mathbf{A}_n\boldsymbol{g}_n(\tilde{\boldsymbol{\psi}}) = 0$ around $\boldsymbol{\psi}_0$, we obtain

$$\sqrt{n}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0) = -\left(\frac{\partial \boldsymbol{g}'_n(\tilde{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}}\mathbf{A}'_n\mathbf{A}_n\frac{\partial \mathbf{g}(\bar{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}'}\right)^{-1} \frac{\partial \boldsymbol{g}'_n(\tilde{\boldsymbol{\psi}})}{\partial \boldsymbol{\psi}}\mathbf{A}'_n\sqrt{n}\mathbf{A}_n\boldsymbol{g}_n(\boldsymbol{\psi}_0),$$

where $\bar{\boldsymbol{\psi}}$ lies element by element between $\boldsymbol{\psi}_0$ and $\hat{\boldsymbol{\psi}}$. Note that

$$\frac{\partial \boldsymbol{g}_n(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'} = -n^{-1} [2\mathbf{C}\boldsymbol{\varepsilon} (\boldsymbol{\psi}) , \mathbf{Z}]' (\mathbf{y}^*, \mathbf{X}) ,$$

and $\mathbf{y}^* = \boldsymbol{\eta}_0 + \mathbf{G}_0\boldsymbol{\varepsilon}$, we have

$$n^{-1}\boldsymbol{\varepsilon}' (\boldsymbol{\psi}) \mathbf{C}\mathbf{y}^* = n^{-1}\boldsymbol{\varepsilon}' (\boldsymbol{\psi}) \mathbf{C}\boldsymbol{\eta}_0 + n^{-1}\boldsymbol{\varepsilon}' (\boldsymbol{\psi}) \mathbf{C}\mathbf{G}_0\boldsymbol{\varepsilon}.$$

Using (A.25), Lemma A.7, and Assumption 7(d) yields

$$\begin{aligned} n^{-1}\boldsymbol{\varepsilon}' (\boldsymbol{\psi}) \mathbf{C}\boldsymbol{\eta}_0 &= n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\eta}_0 - n^{-1}(\rho - \rho_0)\boldsymbol{\varepsilon}'\mathbf{G}'_0\mathbf{C}\boldsymbol{\eta}_0 - n^{-1} (\boldsymbol{\psi} - \boldsymbol{\psi}_0)' \mathbf{Q}'_0\mathbf{C}\boldsymbol{\eta}_0 \\ &= -n^{-1} (\boldsymbol{\psi} - \boldsymbol{\psi}_0)' \mathbf{Q}'_0\mathbf{C}\boldsymbol{\eta}_0 + o_p (1) , \end{aligned}$$

$$n^{-1}\boldsymbol{\varepsilon}'(\boldsymbol{\psi})\,\mathbf{C}\mathbf{G}_0\boldsymbol{\varepsilon} = n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\mathbf{G}_0\boldsymbol{\varepsilon} - n^{-1}(\rho-\rho_0)\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{C}\mathbf{G}_0\boldsymbol{\varepsilon} - n^{-1}(\boldsymbol{\psi}-\boldsymbol{\psi}_0)'\,\mathbf{Q}_0'\mathbf{C}\mathbf{G}_0\boldsymbol{\varepsilon}$$
$$= n^{-1}\sigma_0^2 Tr\,(\mathbf{C}\mathbf{G}_0) - n^{-1}\sigma_0^2(\rho-\rho_0)Tr\,(\mathbf{G}_0'\mathbf{C}\mathbf{G}_0) + o_p(1),$$

if $\delta < 1$, and consequently

$$n^{-1}\boldsymbol{\varepsilon}'(\boldsymbol{\psi})\,\mathbf{C}\mathbf{y}^* = -n^{-1}(\boldsymbol{\psi}-\boldsymbol{\psi}_0)'\,\mathbf{Q}_0'\mathbf{C}\boldsymbol{\eta}_0 + n^{-1}\sigma_0^2 Tr\,(\mathbf{C}\mathbf{G}_0) - n^{-1}\sigma_0^2(\rho-\rho_0)Tr\,(\mathbf{G}_0'\mathbf{C}\mathbf{G}_0) + o_p(1),$$

uniformly in $\boldsymbol{\psi} \in \boldsymbol{\Psi}$. At $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, we have $\boldsymbol{\varepsilon}(\boldsymbol{\psi}_0) = \boldsymbol{\varepsilon}$, and it follows that $n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\mathbf{y}^* = n^{-1}\sigma_0^2 Tr\,(\mathbf{C}\mathbf{G}_0) + o_p(1)$, and

$$n^{-1}\mathbf{Z}'\mathbf{y}^* = n^{-1}\mathbf{Z}'\boldsymbol{\eta}_0 + n^{-1}\mathbf{Z}'\mathbf{G}_0\boldsymbol{\varepsilon} = n^{-1}\mathbf{Z}'\boldsymbol{\eta}_0 + o_p(1),$$

if $\delta < 1$. Thus, $\partial \boldsymbol{g}_n(\bar{\boldsymbol{\psi}})/\partial\boldsymbol{\psi}' = -\mathbf{D} + o_p(1)$, where $\mathbf{D}$ is given by (31) in the paper. Moreover, by Theorem 2 in the paper we have $\mathbf{V}_g^{-1/2}\sqrt{n}\boldsymbol{g}_n(\boldsymbol{\psi}_0) \to_d N(0, \mathbf{I}_{k+1})$ if $\delta < 1/2$, where $\mathbf{V}_g$ is given by (31). Hence, the asymptotic distribution of $\sqrt{n}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ is as stated in Theorem 3. ■

Proof of Theorem 4. To establish consistency and asymptotic distribution of the BMM estimators, we first note that under model (5) in the paper with $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ we have

$$\mathbf{y} - \hat{\rho}\mathbf{y}^* = -(\hat{\rho} - \rho_0)\,\mathbf{y}^* + \mathbf{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon},$$

and hence

$$\mathbf{M}_x(\mathbf{y} - \hat{\rho}\mathbf{y}^*) = -(\hat{\rho} - \rho_0)\,\mathbf{M}_x\mathbf{y}^* + \mathbf{M}_x\boldsymbol{\varepsilon},$$

where $\mathbf{M}_x$ is given by (42) in the paper. Also note that

$$n^{-1}\left(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = n^{-1}(\mathbf{y} - \hat{\rho}\mathbf{y}^*)'\,\mathbf{M}_x(\mathbf{y} - \hat{\rho}\mathbf{y}^*) = \hat{\sigma}^2.$$

Using the above results, the estimating equations (36)–(38) in the paper can now be written as

$$\left(n^{-1}\mathbf{y}^{*\prime}\mathbf{X}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \left(n^{-1}\mathbf{y}^{*\prime}\mathbf{y}^*\right)(\hat{\rho} - \rho_0) = n^{-1}\mathbf{y}^{*\prime}\boldsymbol{\varepsilon} - \hat{\sigma}^2 Tr\left[n^{-1}\mathbf{G}(\hat{\rho})\right], \tag{A.28}$$
$$\left(n^{-1}\mathbf{X}'\mathbf{X}\right)\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\right) + \left(n^{-1}\mathbf{X}'\mathbf{y}^*\right)(\hat{\rho} - \rho_0) = n^{-1}\mathbf{X}'\boldsymbol{\varepsilon}, \tag{A.29}$$

and

$$\hat{\sigma}^2 - \sigma_0^2 = \left[\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) - \sigma_0^2\right] - 2(\hat{\rho} - \rho_0)\left(n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\boldsymbol{\varepsilon}\right) + (\hat{\rho} - \rho_0)^2\left(n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\mathbf{y}^*\right). \tag{A.30}$$

Noting that $\mathbf{y}^* = \boldsymbol{\eta}_0 + \mathbf{G}_0\boldsymbol{\varepsilon}$, where $\boldsymbol{\eta}_0$ is given by (9) in the paper, we obtain

$$n^{-1}\mathbf{y}^{*\prime}\mathbf{X} = n^{-1}\boldsymbol{\eta}_0'\mathbf{X} + n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{X}, \quad n^{-1}\mathbf{y}^{*\prime}\boldsymbol{\varepsilon} = n^{-1}\boldsymbol{\eta}_0'\boldsymbol{\varepsilon} + n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\boldsymbol{\varepsilon},$$
$$n^{-1}\mathbf{y}^{*\prime}\mathbf{y}^* = n^{-1}\boldsymbol{\eta}_0'\boldsymbol{\eta}_0 + n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{G}_0\boldsymbol{\varepsilon} + 2n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\boldsymbol{\eta}_0,$$
$$n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\mathbf{y}^* = n^{-1}\boldsymbol{\eta}_0'\mathbf{M}_x\boldsymbol{\eta}_0 + n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{M}_x\mathbf{G}_0\boldsymbol{\varepsilon} + 2n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{M}_x\boldsymbol{\eta}_0,$$
$$n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\boldsymbol{\varepsilon} = n^{-1}\boldsymbol{\eta}_0'\mathbf{M}_x\boldsymbol{\varepsilon} + n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{M}_x\boldsymbol{\varepsilon}.$$

Also, denoting $\mathbf{G}(\hat{\rho})$ by $\hat{\mathbf{G}}$, we have

$$\hat{\sigma}^2 Tr\left(n^{-1}\hat{\mathbf{G}}\right) - \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\right) = \left(\hat{\sigma}^2 - \sigma_0^2\right) Tr\left(n^{-1}\mathbf{G}_0\right) + \sigma_0^2\left[Tr\left(n^{-1}\hat{\mathbf{G}}\right) - Tr\left(n^{-1}\mathbf{G}_0\right)\right] + \left(\hat{\sigma}^2 - \sigma_0^2\right)\left[Tr\left(n^{-1}\hat{\mathbf{G}}\right) - Tr\left(n^{-1}\mathbf{G}_0\right)\right]. \tag{A.31}$$

But

$$\begin{aligned}\hat{\mathbf{G}} - \mathbf{G}_0 &= \mathbf{W}\left(\mathbf{I}_n - \hat{\rho}\mathbf{W}\right)^{-1} - \mathbf{W}\left(\mathbf{I}_n - \rho_0\mathbf{W}\right)^{-1} \\ &= \mathbf{W}\left(\mathbf{I}_n - \hat{\rho}\mathbf{W}\right)^{-1}\left[\left(\mathbf{I}_n - \rho_0\mathbf{W}\right) - \left(\mathbf{I}_n - \hat{\rho}\mathbf{W}\right)\right]\left(\mathbf{I}_n - \rho_0\mathbf{W}\right)^{-1} \\ &= \left(\hat{\rho} - \rho_0\right)\mathbf{W}\left(\mathbf{I}_n - \hat{\rho}\mathbf{W}\right)^{-1}\mathbf{W}\left(\mathbf{I}_n - \rho_0\mathbf{W}\right)^{-1} = \left(\hat{\rho} - \rho_0\right)\hat{\mathbf{G}}\mathbf{G}_0. \end{aligned} \tag{A.32}$$

Hence, $\hat{\mathbf{G}} = \mathbf{G}_0 + (\hat{\rho} - \rho_0)\hat{\mathbf{G}}\mathbf{G}_0$, and using this result back in (A.32) now yields

$$\hat{\mathbf{G}} - \mathbf{G}_0 = \left(\hat{\rho} - \rho_0\right)\left[\mathbf{G}_0 + \left(\hat{\rho} - \rho_0\right)\hat{\mathbf{G}}\mathbf{G}_0\right]\mathbf{G}_0 = \left(\hat{\rho} - \rho_0\right)\mathbf{G}_0^2 + \mathbf{R}_n\left(\hat{\rho}, \rho_0\right),$$

where $\mathbf{R}_n(\hat{\rho}, \rho_0) = (\hat{\rho} - \rho_0)^2\,\mathbf{G}(\hat{\rho})\,\mathbf{G}^2(\rho_0)$. But by Lemma A.6, $\|\mathbf{G}(\rho)\|_\infty < K$, and only considering estimates of ρ that satisfy the condition $|\hat{\rho}| < 1$, we have $\|\mathbf{R}_n(\hat{\rho}, \rho_0)\|_\infty \leq K|\hat{\rho} - \rho_0|^2$, and hence $E\left|n^{-1}Tr\left[\mathbf{R}_n(\hat{\rho}, \rho_0)\right]\right| \leq KE|\hat{\rho} - \rho_0|^2$, which establishes that

$$n^{-1}Tr\left(\hat{\mathbf{G}} - \mathbf{G}_0\right) = \left(\hat{\rho} - \rho_0\right) Tr\left(n^{-1}\mathbf{G}_0^2\right) + O_p\left[\left(\hat{\rho} - \rho_0\right)^2\right]. \tag{A.33}$$

Using results in Lemmas A.8 and A.9, it is now readily established that

$$\begin{gathered} n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{X} = O_p\left(n^{-1/2}\right), \quad n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\boldsymbol{\eta}_0 = O_p\left(n^{-1/2}\right), \quad n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{M}_x\boldsymbol{\eta}_0 = O_p\left(n^{-1/2}\right), \\ n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\boldsymbol{\varepsilon} = \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\right) + O_p\left(n^{-1/2}\right), \quad n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{M}_x\mathbf{G}_0\boldsymbol{\varepsilon} = \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0'\mathbf{M}_x\mathbf{G}_0\right) + O_p\left(n^{-1/2}\right), \\ n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{G}_0\boldsymbol{\varepsilon} = \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0'\mathbf{G}_0\right) + O_p\left(n^{-1/2}\right), \quad n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{M}_x\boldsymbol{\varepsilon} = \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right) + O_p\left(n^{-1/2}\right), \end{gathered}$$

and hence

$$\begin{aligned} n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\boldsymbol{\varepsilon} &= \sigma_0^2 + O_p\left(n^{-1/2}\right), \quad n^{-1}\mathbf{y}^{*\prime}\boldsymbol{\varepsilon} = \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\right) + O_p\left(n^{-1/2}\right), \\ n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\boldsymbol{\varepsilon} &= \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right) + O_p\left(n^{-1/2}\right), \\ n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\mathbf{y}^* &= n^{-1}\boldsymbol{\eta}_0'\mathbf{M}_x\boldsymbol{\eta}_0 + \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0'\mathbf{M}_x\mathbf{G}_0\right) + O_p\left(n^{-1/2}\right). \end{aligned}$$

Using these results in (A.30) now yields

$$\hat{\sigma}^2 - \sigma_0^2 = \left[\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) - \sigma_0^2\right] - 2\left(\hat{\rho} - \rho_0\right)\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right) + O_p\left[\left(\hat{\rho} - \rho_0\right)n^{-1/2}\right] + O_p\left[\left(\hat{\rho} - \rho_0\right)^2\right]. \tag{A.34}$$

Substituting (A.33) and (A.34) in (A.31) we have (noting that $Tr\left(n^{-1}\mathbf{G}_0\right) < K$)

$$\begin{aligned} &\hat{\sigma}^2 Tr\left(n^{-1}\hat{\mathbf{G}}\right) - \sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\right) \\ =& Tr\left(n^{-1}\mathbf{G}_0\right)\left[\left(n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\boldsymbol{\varepsilon}\right) - \sigma_0^2\right] - 2\sigma_0^2\left(\hat{\rho} - \rho_0\right) Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right) Tr\left(n^{-1}\mathbf{G}_0\right) \\ &+ \sigma_0^2\left(\hat{\rho} - \rho_0\right) Tr\left(n^{-1}\mathbf{G}_0^2\right) + O_p\left[\left(\hat{\rho} - \rho_0\right)^2\right] + O_p\left[\left(\hat{\rho} - \rho_0\right)n^{-1/2}\right]. \end{aligned} \tag{A.35}$$

Using (A.35) in (A.28) and rearranging gives

$$\left(n^{-1}\mathbf{y}^{*\prime}\mathbf{X}\right)\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)+h_{n,\rho\rho}\left(\hat{\rho}-\rho_0\right)=h_{n,\rho\varepsilon}+O_p\left[\left(\hat{\rho}-\rho_0\right)^2\right]+O_p\left[\left(\hat{\rho}-\rho_0\right)n^{-1/2}\right], \quad \text{(A.36)}$$

where

$$h_{n,\rho\varepsilon}=n^{-1}\boldsymbol{\eta}_0'\boldsymbol{\varepsilon}+n^{-1}\boldsymbol{\varepsilon}'\left[\mathbf{G}_0'-\mathbf{M}_x Tr\left(n^{-1}\mathbf{G}_0\right)\right]\boldsymbol{\varepsilon},$$
$$h_{n,\rho\rho}=n^{-1}\mathbf{y}^{*\prime}\mathbf{y}^*+\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0^2\right)-2\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right)Tr\left(n^{-1}\mathbf{G}_0\right).$$

Combining (A.36) and (A.29) we have

$$\begin{pmatrix} h_{n,\rho\rho} & \frac{\mathbf{y}^{*\prime}\mathbf{X}}{n} \\ \frac{\mathbf{X}'\mathbf{y}^*}{n} & \frac{\mathbf{X}'\mathbf{X}}{n}\end{pmatrix}\begin{pmatrix}\hat{\rho}-\rho_0 \\ \hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\end{pmatrix}=\begin{pmatrix} h_{n,\rho\varepsilon} \\ \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n}\end{pmatrix}+\begin{pmatrix} O_p\left[\left(\hat{\rho}-\rho_0\right)^2\right]+O_p\left[\left(\hat{\rho}-\rho_0\right)n^{-1/2}\right] \\ \mathbf{0}\end{pmatrix}.$$

It is also easily seen that

$$\begin{aligned} h_{n,\rho\rho} &= n^{-1}\boldsymbol{\eta}_0'\boldsymbol{\eta}_0+n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\mathbf{G}_0\boldsymbol{\varepsilon}+2n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}_0'\boldsymbol{\eta}_0 \\ &\quad +\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0^2\right)-2\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right)Tr\left(n^{-1}\mathbf{G}_0\right) \\ &= n^{-1}\boldsymbol{\eta}_0'\boldsymbol{\eta}_0+\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0'\mathbf{G}_0\right)+\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0^2\right) \\ &\quad -2\sigma_0^2 Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right)Tr\left(n^{-1}\mathbf{G}_0\right)+O_p\left(n^{-1/2}\right).\end{aligned}$$

Notice that

$$\begin{aligned} Tr\left(n^{-1}\mathbf{G}_0\mathbf{M}_x\right) &= n^{-1}Tr\left(\mathbf{G}_0\right)-n^{-1}Tr\left[\mathbf{G}_0\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\right] \\ &= n^{-1}Tr\left(\mathbf{G}_0\right)-n^{-1}Tr\left[\left(n^{-1}\mathbf{X}'\mathbf{X}\right)^{-1}\left(n^{-1}\mathbf{X}'\mathbf{G}_0\mathbf{X}\right)\right].\end{aligned}$$

Under Assumption 3, we have

$$p\lim_{n\to\infty} n^{-1}Tr\left(\mathbf{G}_0\mathbf{M}_x\right)=\lim_{n\to\infty} n^{-1}Tr\left(\mathbf{G}_0\right)-\lim_{n\to\infty} n^{-1}Tr\left(\boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{xgx}\right)=\lim_{n\to\infty} n^{-1}Tr\left(\mathbf{G}_0\right).$$

Hence, using results in Lemmas A.8 and A.9 we have

$$p\lim_{n\to\infty} h_{n,\rho\rho}=\boldsymbol{\beta}_0'\boldsymbol{\Sigma}_{xggx}\boldsymbol{\beta}_0+\sigma_0^2 h_0,$$

where $h_0=\lim_{n\to\infty}h_n$, and h_n is given by (10); $p\lim_{n\to\infty}h_{n,\rho\varepsilon}=0$; $p\lim_{n\to\infty}\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n}=\mathbf{0}$; $p\lim_{n\to\infty}\frac{\mathbf{y}^{*\prime}\mathbf{X}}{n}=\boldsymbol{\beta}_0'\boldsymbol{\Sigma}_{xgx}$; and $p\lim_{n\to\infty}\frac{\mathbf{X}'\mathbf{X}}{n}=\boldsymbol{\Sigma}_{xx}$. Therefore, the BMM estimators are consistent if $\mathbf{H}$, defined in (40) in the paper, is a non-singular matrix. In particular, under this condition $\hat{\rho}-\rho_0=O_p(n^{-1/2})$.

To derive the asymptotic distribution of the BMM estimators, we first note that

$$\begin{pmatrix} h_{n,\rho\rho} & \frac{\mathbf{y}^{*\prime}\mathbf{X}}{n} \\ \frac{\mathbf{X}'\mathbf{y}^*}{n} & \frac{\mathbf{X}'\mathbf{X}}{n}\end{pmatrix}\begin{pmatrix}\sqrt{n}\left(\hat{\rho}-\rho_0\right) \\ \sqrt{n}\left(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\right)\end{pmatrix}=\begin{pmatrix}\sqrt{n}h_{n,\rho\varepsilon} \\ \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sqrt{n}}\end{pmatrix}+\begin{pmatrix} O_p\left[\sqrt{n}\left(\hat{\rho}-\rho_0\right)^2\right]+O_p\left[\left(\hat{\rho}-\rho_0\right)\right] \\ \mathbf{0}\end{pmatrix},$$

and

$$\mathbf{H}\sqrt{n}\left(\hat{\boldsymbol{\psi}}-\boldsymbol{\psi}_0\right)=\begin{pmatrix}\sqrt{n}h_{n,\rho\varepsilon}+O_p\left(n^{-1/2}\right) \\ \frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sqrt{n}}\end{pmatrix}.$$

Consider now

$$\sqrt{n}h_{n,\rho\varepsilon} = \frac{\boldsymbol{\eta}_0'\boldsymbol{\varepsilon}}{\sqrt{n}} + \frac{\boldsymbol{\varepsilon}'\boldsymbol{\Pi}\boldsymbol{\varepsilon}}{\sqrt{n}}, \tag{A.37}$$

where $\boldsymbol{\Pi}$ is given by (42) in the paper. Since $\mathbf{X}$ is strictly exogenous under Assumption 3, we carry on the analysis of (A.37) conditional on $\mathbf{X}$. By Lemma A.6(ii), $\mathbf{G}_0$ satisfies the conditions in (11) and (12) in the paper. Since $\mathbf{M}_x$ is an idempotent matrix, $\boldsymbol{\Pi}$ also satisfies (11) and (12). Therefore, applying Theorem 2 in the paper leads to

$$\frac{1}{\sqrt{n}}\left[\boldsymbol{\eta}_0'\boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}'\boldsymbol{\Pi}\boldsymbol{\varepsilon} - \sigma_0^2 Tr\left(\boldsymbol{\Pi}\right)\right] \rightarrow_d N(0,\omega_n^2),$$

where ω_n^2 is given by (41) in the paper. Notice that

$$Tr\left(\boldsymbol{\Pi}\right) = Tr\left(\mathbf{G}_0\right) - Tr\left(\mathbf{M}_x\right)Tr\left(n^{-1}\mathbf{G}_0\right) = Tr\left(\mathbf{G}_0\right) - \frac{n-k}{n}Tr\left(\mathbf{G}_0\right) = \frac{k}{n}Tr\left(\mathbf{G}_0\right) < K,$$

and it follows that $p\lim_{n\rightarrow\infty}\left[n^{-1/2}\sigma_0^2 Tr\left(\boldsymbol{\Pi}\right)\right] = 0$. Hence, by Slutsky's theorem we obtain

$$\frac{\boldsymbol{\eta}_0'\boldsymbol{\varepsilon}}{\sqrt{n}} + \frac{\boldsymbol{\varepsilon}'\boldsymbol{\Pi}\boldsymbol{\varepsilon}}{\sqrt{n}} \rightarrow_d N(0,\omega_n^2).$$

In addition, it is readily seen that $\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{\sqrt{n}} \rightarrow_d N\left(0,\sigma_0^2\boldsymbol{\Sigma}_{xx}\right)$. Thus, the asymptotic distribution of $\hat{\boldsymbol{\psi}}$ as stated in Theorem 4 in the paper is established. ■

Proof of Proposition 2. We will show that under the stated conditions the limiting distribution of the BMM estimator given by Theorem 4 in the paper is equivalent to the distribution of the best GMM estimator given by (4.5) of Proposition 3 in Lee (2007). Note that the last term of (41) in the paper can be rewritten as

$$\begin{aligned}
& Tr\left(n^{-1}\boldsymbol{\Pi}'\boldsymbol{\Pi}\right) + Tr\left(n^{-1}\boldsymbol{\Pi}^2\right) \\
&= Tr\left(n^{-1}\mathbf{G}_0'\mathbf{G}_0 + n^{-1}\mathbf{G}_0^2\right) - 4Tr\left(n^{-1}\mathbf{M}_x\mathbf{G}_0\right)Tr\left(n^{-1}\mathbf{G}_0\right) \\
&\quad + 2Tr\left(n^{-1}\mathbf{M}_x\right)\left[Tr\left(n^{-1}\mathbf{G}_0\right)\right]^2 \\
&= n^{-1}Tr\left(\mathbf{G}_0'\mathbf{G}_0 + \mathbf{G}_0^2\right) - 2\left[Tr\left(n^{-1}\mathbf{G}_0\right)\right]^2 \\
&\quad - 4Tr\left[n^{-1}\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{G}_0\right]Tr\left(n^{-1}\mathbf{G}_0\right) + 2n^{-1}k\left[Tr\left(n^{-1}\mathbf{G}_0\right)\right]^2 \\
&= h_n - 4n^{-1}Tr\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{G}_0\mathbf{X}\right)\right]Tr\left(n^{-1}\mathbf{G}_0\right) + 2n^{-1}k\left[Tr\left(n^{-1}\mathbf{G}_0\right)\right]^2,
\end{aligned}$$

where h_n is given by (10) in the paper. Since $\boldsymbol{\Sigma}_{xx} = p\lim_{n\rightarrow\infty} n^{-1}\mathbf{X}'\mathbf{X}$ and $\boldsymbol{\Sigma}_{xgx} = p\lim_{n\rightarrow\infty} n^{-1}\mathbf{X}'\mathbf{G}_0\mathbf{X}$ exist and they are k-dimensional square matrices (k is finite), it follows that

$$\begin{aligned}
& p\lim_{n\rightarrow\infty}\left[Tr\left(n^{-1}\boldsymbol{\Pi}'\boldsymbol{\Pi}\right) + Tr\left(n^{-1}\boldsymbol{\Pi}^2\right)\right] \\
&= h - 4\lim_{n\rightarrow\infty} n^{-1}Tr\left(\boldsymbol{\Sigma}_{xx}\boldsymbol{\Sigma}_{xgx}\right) + 2\lim_{n\rightarrow\infty} n^{-1}k\left[Tr\left(n^{-1}\mathbf{G}_0\right)\right]^2 = h,
\end{aligned} \tag{A.38}$$

where $h = \lim_{n\rightarrow\infty} h_n$. In addition, the assumption of normally distributed errors imply that $\gamma_2 = 0$ and $\mu_3 = 0$. Finally, combining (40) and (41) in the paper with (A.38) leads to

$\mathbf{V} = \sigma_0^2 \mathbf{H}$ and hence $\mathbf{\Omega}_b = \sigma_0^2 \mathbf{H}^{-1}$, which is identical to the asymptotic variance of the best GMM estimator given by (4.5) of Lee (2007). ■

References

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians.* Oxford: Oxford University Press.

Hall, P. and C. C. Heyde (1980). *Martingale Limit Theory and Its Application.* London: Academic Press.

Horn, R. A. and C. R. Johnson (2012). *Matrix Analysis.* Cambridge: Cambridge University Press.

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran *I* test statistic with applications. *Journal of Econometrics* 104, 219–257.

Lee, L.-f. (2004). A supplement to "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models". Unpublished manuscript. Department of Economics, The Ohio State University. Available at: `http://www.econ.ohio-state.edu/lee/wp/sar-qml-r-appen-04feb.pdf`.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489–514.

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