

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

January 15, 2020

Abstract

In spatial econometrics literature estimation and inference are carried out assuming that the matrix of spatial or network connections has uniformly bounded absolute column sums in the number of units, n , in the network. This paper relaxes this restriction and allows for one or more units to have pervasive effects in the network. The linear-quadratic central limit theorem of Kelejian and Prucha (2001) is generalized to allow for such dominant units, and the asymptotic properties of the GMM estimators are established in this more general setting. A new bias-corrected method of moments (BMM) estimator is also proposed that avoids the problem of weak instruments by self-instrumenting the spatially lagged dependent variable. Both cases of homoskedastic and heteroskedastic errors are considered and the associated 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 GMM and BMM 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, heteroskedastic errors, GMM, bias-corrected method of moments (BMM), US input-output analysis, capital share.

JEL Classifications: C13, C21, C23, R15

*We acknowledge valuable comments from two anonymous reviewers and an associate editor. We also appreciate helpful comments from Natalia Bailey, Georges Bresson, Alex Chudik, Tadao Hoshino, Simon Reese, Ron Smith, and participants at the 2019 Asian Meeting of the Econometric Society, China Meeting of the Econometric Society, the 6th annual conference of the International Association for Applied Econometrics, the 2019 Midwest Econometrics Group Meeting, and the Spatial Econometrics Association XIII World Conference. 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. It will be satisfied, for example, 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). Acemoglu et al. (2012) show that in the US intersectoral network there exists a high degree of asymmetry in the roles that sectors play as suppliers to others, and such asymmetries in production networks could be an important source of aggregate fluctuations. Pesaran and Yang (2019) confirm the finding in Acemoglu et al. (2012) and further identify the wholesale trade sector as the dominant sector in the US over the period 1972–2007. Dungey and Volkov (2018) consider 49 OECD economies from 1996 to 2011 and identify wholesale trade and R&D as two dominant sectors to most economies. 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 suppose $\|\mathbf{W}\|_1 = \Theta(n^\delta)$, where $\Theta(n^\delta)$ denotes the expansion rate of $\|\mathbf{W}\|_1$ in terms of n , 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 $d_{jn} = \sum_{i=1}^n w_{ij}$. The degree of dominance of unit j can now be measured by the exponent δ_j , defined by $d_{jn} = \Theta(n^{\delta_j})$, 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 $\delta = \max(\delta_1, \delta_2, \dots, \delta_n)$.² From this 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 asymptotic properties of the GMM estimators of the SAR model under both homoskedastic and heteroskedastic errors, thus generalizing the results developed by Lee (2007) and Lin and Lee (2010) to the case of non-bounded spatial weights matrices. We establish conditions under which GMM estimators are consistent and asymptotically normal even if $\delta > 0$, under both homoskedastic and heteroskedastic errors.

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 design. We show that both GMM and BMM estimators are consistent

¹For further details see Definition 1 in Pesaran and Yang (2019).

²Note that when $w_{ij} \geq 0$, then $\|\mathbf{W}\|_1 = \sup_j(d_j)$.

³It 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, Parker and Sul (2016), Brownlees and Mesters (2018), and Kapetanios et al. (2019). In a related literature, Bailey et al. (2016) also consider estimating δ using large panel data sets when \mathbf{W} is not known.

if $0 \leq \delta < 1$, and establish their asymptotic normality for values of δ in the range $0 \leq \delta < 1/2$, irrespective of whether the errors are homoskedastic or not.

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. Both estimators also perform reasonably well under general error heteroskedasticity.

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 with homoskedastic errors 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). We obtained slightly larger estimates for the pre-crisis period when we allowed for heteroskedastic errors, which suggest error heteroskedasticity might have been serious in the pre-crisis period. But overall, the estimates of capital share obtained using the SAR model compare reasonably well with the estimates reported in the literature using very different calibration techniques.

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 the empirical application, and Section 8 gives some concluding remarks. An Online Supplement provides proofs of theorems and propositions, gives statements and proofs of necessary lemmas, and includes additional Monte Carlo and empirical results.

Notations: Generic positive finite constants are denoted by K when they are large, and by c 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 $f_n/g_n \rightarrow 0$ as $n \rightarrow \infty$. If $\{f_n\}_{n=1}^\infty$ and $\{g_n\}_{n=1}^\infty$ are both positive sequences of real numbers, then $f_n = \Theta(g_n)$ if there exists $N_0 \geq 1$ and positive finite constants K_0 and K_1 , such that $\inf_{n \geq N_0} (f_n/g_n) \geq K_0$, and $\sup_{n \geq N_0} (f_n/g_n) \leq K_1$. The symbols \rightarrow_p and \rightarrow_d indicate convergence in probability and in distribution as $n \rightarrow \infty$, respectively. Let $\{x_n\}$ be a sequence of random variables. We write

$x_n = o_p(1)$ if $x_n \rightarrow_p 0$ as $n \rightarrow \infty$. $E_0(\cdot)$ 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); and $\lambda_{\max}(\mathbf{A})$ ($\lambda_{\min}(\mathbf{A})$) denotes the largest (smallest) eigenvalue of \mathbf{A} . The symbol $\text{diag}(\mathbf{A})$ represents a vector consisting of the diagonal elements of \mathbf{A} , namely, $\text{diag}(\mathbf{A}) = (a_{11}, a_{22}, \dots, a_{nn})'$; whereas $\text{Diag}(\mathbf{A})$ 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., $\mathbf{1}_n = (1, 1, \dots, 1)'$.

2 The model and its assumptions

We consider the following standard SAR model:

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

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

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

$\mathbf{y} = (y_1, y_2, \dots, y_n)'$, $\mathbf{w}_{i.} = (w_{i1}, w_{i2}, \dots, w_{in})'$ 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 that the row sums of \mathbf{W} are uniformly bounded in n , but allow the column sums of \mathbf{W} to rise with n . Specifically, following Pesaran and Yang (2019), denote the j^{th} column sum of \mathbf{W} by $d_{jn} = \sum_{i=1}^n w_{ij}$, and assume that d_{jn} is of order n^{δ_j} such that

$$d_{jn} = \kappa_j n^{\delta_j}, \quad \text{for } j = 1, 2, \dots, n, \quad (3)$$

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

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

and note that $\max_j (d_{jn}) = \|\mathbf{W}\|_1 = \Theta(n^\delta)$. We further assume that the number of dominant units, m , is finite, and without loss of generality we suppose the first m units, $j = 1, 2, \dots, m$, are δ_j -dominant (with $\delta_j > 0$), 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 short, we consider weights matrices such that their first m column sums are unbounded in n , with the remaining $(n - m)$ column sums bounded. Although m is assumed to be fixed, it can be shown that this must be true if δ_j 's satisfy the summability condition, $\sum_{j=1}^n \delta_j < K$.⁴ Also, since our focus is on the estimation of ρ and β , the identity of the dominant units or the order by which there are included in \mathbf{W} do not affect the analysis.

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

$$\mathbf{y} = \rho \mathbf{y}^* + \mathbf{X}\beta + \varepsilon, \quad (5)$$

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 $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$. The reduced-form representation of (5) is given by

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

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

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

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 β , and their true values are denoted by ρ_0 and β_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

$$\mathbf{G}_0 = \mathbf{G}(\rho_0) = \mathbf{W}(\mathbf{I}_n - \rho_0 \mathbf{W})^{-1} = \mathbf{W}\mathbf{S}_0^{-1}, \text{ and } \eta_0 = \mathbf{G}_0 \mathbf{X}\beta_0. \quad (8)$$

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

Assumption 1 *The idiosyncratic errors, ε_i , for $i = 1, 2, \dots, n$, in the SAR model given by (1) are independently distributed over i with zero means and variances, σ_i^2 , such that $\inf_i(\sigma_i^2) > c > 0$, $\sup_i(\sigma_i^2) < K$, and $\sup_i E|\varepsilon_i|^{4+c} < K$, for some $c > 0$.*

Assumption 2 *The $(k+1)$ -dimensional parameter vector in (1), $\psi = (\rho, \beta')' \in \Psi = \Theta_\rho \times \Theta_\beta$, where Θ_ρ and Θ_β are compact subsets of \mathbb{R} and \mathbb{R}^k , respectively ($|\rho| < K$ and $\|\beta\|_1 < K$); the true value of ψ_0 , denoted by $\psi_0 = (\rho_0, \beta'_0)'$, lies in the interior of the parameter space, Ψ .*

Assumption 3 *Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ be the $n \times k$ matrix of observations on the regressors in (1), where $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})'$. (a) \mathbf{x}_i , for $i = 1, 2, \dots, n$, are distributed independently*

⁴See Proposition 2 of Pesaran and Yang (2019). The assumption of a fixed number of dominant units is analogous to the assumption of a fixed number of strong factors in factor models. As shown in Chudik et al. (2011), in order for the variances of the observables to be bounded, the number of strong factors must be fixed and cannot vary with n .

of the errors, ε_j , for all i and j , and $\sup_{i,s} E(|x_{is}|^{2+c}) < K$, (b) $n^{-1}\mathbf{X}'\mathbf{X} \rightarrow_p \Sigma_{xx}$ is positive definite, and (c) $n^{-1}\mathbf{X}'\mathbf{G}_0\mathbf{X} \rightarrow_p \Sigma_{xg_0x}$ and $n^{-1}\mathbf{X}'\mathbf{G}'_0\mathbf{G}_0\mathbf{X} \rightarrow_p \Sigma_{xg_0g_0x}$, where \mathbf{G}_0 is defined by (8), and Σ_{xg_0x} and $\Sigma_{xg_0g_0x}$ are finite $k \times k$ matrices.

Assumption 4 The spatial weights matrix, $\mathbf{W} = (w_{ij})$, in the SAR model given by (1) is such that (a) $w_{ij} \geq 0$ for all i and j , (b) $\|\mathbf{W}\|_\infty < K$, and $|\rho| \|\mathbf{W}\|_\infty < 1$, and (c) its column sums, denoted by $d_{jn} = \sum_{i=1}^n w_{ij}$, $j = 1, 2, \dots, n$, are non-zero and follow the specification given by (3), where $0 < \delta_j \leq 1$ for $j = 1, 2, \dots, m$, and $\delta_j = 0$ for $j = m+1, m+2, \dots, n$, with m being a fixed number. Also, $|\rho| \|\mathbf{W}_{22}\|_1 < 1$, where \mathbf{W}_{22} is the $(n-m)$ -dimensional square submatrix of \mathbf{W} that represents the connections among the non-dominant units.

Assumption 5 There exists n_0 such that for all $n \geq n_0$ (including $n \rightarrow \infty$), either

(a) $n^{-1}\mathbf{Q}'_0\mathbf{Q}_0$ is positive definite, where $\mathbf{Q}_0 = (\mathbf{G}_0\mathbf{X}\beta_0, \mathbf{X})$,

and/or

(b) $h_n > c > 0$, where

$$h_n = \text{Tr} [n^{-1} (\mathbf{G}_0^2 + \mathbf{G}'_0\mathbf{G}_0 - 2\check{\mathbf{G}}_0\mathbf{G}_0) \Sigma_0], \quad (9)$$

\mathbf{G}_0 is given by (8), $\check{\mathbf{G}}_0 = \text{Diag}(\mathbf{G}_0)$, and $\Sigma_0 = \text{Diag}(\sigma_{10}^2, \sigma_{20}^2, \dots, \sigma_{n0}^2)$, with σ_{i0}^2 denoting the true value of $\sigma_i^2 = \text{Var}(\varepsilon_i)$, for $i = 1, 2, \dots, n$.

Remark 1 It is worth noting that under Assumption 4, $|\lambda_{\max}(\rho\mathbf{W})| \leq |\rho| \|\mathbf{W}\|_\infty < 1$, and hence $\lambda_{\min}[\mathbf{S}(\rho)] = 1 - \lambda_{\max}(\rho\mathbf{W}) > c > 0$, and as a result $\mathbf{S}(\rho) = \mathbf{I}_n - \rho\mathbf{W}$ is invertible for all ρ satisfying $|\rho| \|\mathbf{W}\|_\infty < 1$, irrespective of whether the column sums of \mathbf{W} are bounded or not.⁵ Moreover, it can be seen from (6) that in order for the variance of y_i to be bounded, we need $\|\mathbf{S}^{-1}(\rho)\|_\infty < K$, which is ensured by the assumption $|\rho| \|\mathbf{W}\|_\infty < 1$.⁶ It is also clear that the condition $\|\mathbf{W}\|_\infty < K$ of Assumption 4 follows from $|\rho| \|\mathbf{W}\|_\infty < 1$ when $|\rho| > 0$, but it is required when $\rho = 0$. In the special case where \mathbf{W} is row-standardized such that each row sums up to one, the invertibility condition reduces to $|\rho| < 1$. Finally, dominant units with $\delta_j > 0$ can exist only in non-symmetric networks. This follows since when \mathbf{W} is symmetric the boundedness of $\text{Var}(y_i)$ excludes the possibility of dominant units.

Remark 2 The non-negativity condition, $w_{ij} \geq 0$, in Assumption 4(a) 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

⁵This can be seen by noting that $\mathbf{S}(\rho)$ is invertible if $|\lambda_{\max}(\rho\mathbf{W})| < 1$, and $|\lambda_{\max}(\rho\mathbf{W})| \leq |\rho| \min(\|\mathbf{W}\|_\infty, \|\mathbf{W}\|_1)$.

⁶Without loss of generality, let us abstract from the regressors and note that under the invertibility condition $\mathbf{y} = \mathbf{S}^{-1}(\rho)\varepsilon$. It then readily follows that $\inf_j (\sigma_j^2) \sum_{j=1}^n (s^{ij})^2 \leq \text{Var}(y_i) = \sum_{j=1}^n (s^{ij})^2 \sigma_j^2 \leq \sup_j (\sigma_j^2) \sum_{j=1}^n (s^{ij})^2$, where s^{ij} is the $(i,j)^{th}$ element of $\mathbf{S}^{-1}(\rho)$. Accordingly, for $\text{Var}(y_i)$ to be bounded in n and strictly positive, under Assumption 1 it is necessary and sufficient that the rows of $\mathbf{S}^{-1}(\rho)$ are square summable. This condition is in turn met when $|\rho| \|\mathbf{W}\|_\infty < 1$ even if $\|\mathbf{W}\|_1 = \Theta(n^\delta)$, with $\delta > 0$.

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.

Remark 3 In Assumption 3(c), we have assumed that $n^{-1} \mathbf{X}' \mathbf{G}_0' \mathbf{G}_0 \mathbf{X} \rightarrow_p \Sigma_{xg_0g_0x}$. As is proved in Lemma S.9 of the Online Supplement, $E(n^{-1} \mathbf{X}' \mathbf{G}_0' \mathbf{G}_0 \mathbf{X})$ exists and is bounded. Intuitively, since $\|\mathbf{G}_0\|_\infty < K$ by Lemma S.6, asymptotically $\mathbf{G}_0 \mathbf{X}$ behaves similarly to \mathbf{X} .

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. We first 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 S1.2 of the Online Supplement.⁷

Theorem 1 Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ denote the $n \times 1$ vector of random variables, where ε_i , for $i = 1, 2, \dots, n$, are independently distributed over i with zero means and variance σ_i^2 , where $\inf_i(\sigma_i^2) > c > 0$ and $\sup_i(\sigma_i^2) < K$. Suppose that $\sup_i E |\varepsilon_i|^{4+c} < K$, for some $c > 0$, and denote the excess kurtosis of $\{\varepsilon_i\}$ by $k_{e,i} = (\mu_{4i}/\sigma_i^4) - 3$, where $\mu_{4i} = E(\varepsilon_i^4)$. Let $\mathbf{P} = (p_{ij})$ be an array of $n \times n$ constant matrices that satisfy the following conditions

$$\|\mathbf{P}\|_\infty = \sup_i \sum_{j=1}^n |p_{ij}| < K, \quad (10)$$

$$\|\mathbf{P}\|_1 = \sup_j \sum_{i=1}^n |p_{ij}| = O(n^\alpha), \quad 0 \leq \alpha < 1, \quad (11)$$

where α is a fixed constant and \mathbf{P} has a finite number of unbounded columns. Define $\mathbf{A} = (a_{ij}) = (\mathbf{P} + \mathbf{P}')/2$. Suppose \mathbf{A} is such that

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \sigma_i^2 \sigma_j^2 + \frac{1}{2} n^{-1} \sum_{i=1}^n a_{ii}^2 \sigma_i^4 k_{e,i} > c > 0, \text{ for all } n \text{ (including } n \rightarrow \infty\text{)}. \quad (12)$$

Then if α lies in the range $0 \leq \alpha < 1/2$, we have

$$Q = \frac{\boldsymbol{\varepsilon}' \mathbf{A} \boldsymbol{\varepsilon} - \sum_{i=1}^n a_{ii} \sigma_i^2}{\sqrt{n} \varpi_n} \rightarrow_d N(0, 1), \text{ as } n \rightarrow \infty, \quad (13)$$

⁷Note 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.

where

$$\varpi_n^2 = 2n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \sigma_i^2 \sigma_j^2 + n^{-1} \sum_{i=1}^n a_{ii}^2 \sigma_i^4 k_{e,i}. \quad (14)$$

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 (10).

Remark 4 It is easily seen that condition (12) ensures $\varpi_n^2 > c > 0$, for all n (including $n \rightarrow \infty$). If the errors are normally distributed, then $k_{e,i} = 0$ for all i , and

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \sigma_i^2 \sigma_j^2 \geq \left(\inf_i \sigma_i^2 \right)^2 n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 = \left(\inf_i \sigma_i^2 \right)^2 n^{-1} \text{Tr}(\mathbf{A}\mathbf{A}') = \left(\inf_i \sigma_i^2 \right)^2 n^{-1} \text{Tr}(\mathbf{A}^2),$$

then (12) simplifies to $n^{-1} \text{Tr}(\mathbf{A}^2) > c > 0$, which always holds true for finite n (except for the trivial case of $\mathbf{A} = \mathbf{0}$). Therefore in the case of $k_{e,i} = 0$, to ensure $\varpi_n^2 > c > 0$, it is sufficient to assume that $n^{-1} \text{Tr}(\mathbf{A}^2)$ tends to a strictly positive limit as $n \rightarrow \infty$.

The next theorem extends Theorem 1 to linear-quadratic forms.

Theorem 2 Let $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ denote the $n \times 1$ vector of random variables, where ε_i , for $i = 1, 2, \dots, n$, are independently distributed over i with zero means and variance σ_i^2 , where $\inf_i(\sigma_i^2) > c > 0$ and $\sup_i(\sigma_i^2) < K$. Suppose that $\sup_i E |\varepsilon_i|^{4+c} < K$, for some $c > 0$, and denote the excess kurtosis of $\{\varepsilon_i\}$ by $k_{e,i} = (\mu_{4i}/\sigma_i^4) - 3$, where $\mu_{4i} = E(\varepsilon_i^4)$. Let $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)'$ be a vector of random variables with means $\mu_{\eta,i}$ and variances $\sigma_{\eta,i}^2$, distributed independently of ε_j , for all i and j , where $\sigma_{\eta,i}^2 > 0$, for all i , and $\sup_i E(|\eta_i|^{2+c}) < K$. Let $\mathbf{P} = (p_{ij})$ be an array of $n \times n$ constant matrices that satisfy conditions (10) and (11), and \mathbf{P} has a finite number of unbounded columns, with $\alpha \geq 0$, as defined by (11). Let $\mathbf{A} = (a_{ij}) = (\mathbf{P} + \mathbf{P}')/2$, and suppose that

$$n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \sigma_i^2 \sigma_j^2 + \frac{1}{2} n^{-1} \sum_{i=1}^n a_{ii}^2 \sigma_i^4 k_{e,i} + \frac{1}{2} n^{-1} \sum_{i=1}^n \sigma_{\eta,i}^2 \sigma_i^2 + n^{-1} \sum_{i=1}^n a_{ii} \mu_{\eta,i} \mu_{3i} > c > 0, \quad (15)$$

for all n (including $n \rightarrow \infty$), where $\mu_{3i} = E(\varepsilon_i^3)$. Then if α lies in the range $0 \leq \alpha < 1/2$, we have

$$\tilde{Q} = \frac{\boldsymbol{\varepsilon}' \mathbf{A} \boldsymbol{\varepsilon} + \boldsymbol{\varepsilon}' \boldsymbol{\eta} - \sum_{i=1}^n a_{ii} \sigma_i^2}{\sqrt{n} \tilde{\varpi}_n^2} \rightarrow_d N(0, 1), \text{ as } n \rightarrow \infty, \quad (16)$$

where

$$\tilde{\varpi}_n^2 = 2n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \sigma_i^2 \sigma_j^2 + n^{-1} \sum_{i=1}^n a_{ii}^2 \sigma_i^4 k_{e,i} + n^{-1} \sum_{i=1}^n \sigma_{\eta,i}^2 \sigma_i^2 + 2n^{-1} \sum_{i=1}^n a_{ii} \mu_{\eta,i} \mu_{3i}. \quad (17)$$

Remark 5 Condition (15) ensures that $\tilde{\omega}_n^2 > c > 0$, for all n (including $n \rightarrow \infty$). If the errors are symmetrically distributed, then $\mu_{3i} = 0$. Since $\sigma_{\eta,i}^2 > 0$ for all i , condition (15) in this case would reduce to (12) of Theorem 1.

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. Lee (2007) assumes that the errors are homoskedastic, and Lin and Lee (2010) further consider the GMM method in the presence of unknown heteroskedasticity. To make our analysis more general, our theoretical derivations allow for unknown heteroskedasticity and we will make comments on the special case of homoskedastic errors.

Specifically, consider model (1) and suppose that $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)'$ 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 6 Let $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n)'$ be the $n \times r$ matrix of observations on the r instrumental variables, $\mathbf{z}_i = (z_{i1}, z_{i2}, \dots, z_{ir})'$. (a) \mathbf{z}_i is distributed independently of the errors, ε_j , for all i and $j = 1, 2, \dots, n$, and $\sup_{i,s} E(|z_{is}|^{2+c}) < K$, (b) $n^{-1}\mathbf{Z}'\mathbf{Z} \rightarrow_p \boldsymbol{\Sigma}_{zz}$, a positive definite matrix, and (c) $n^{-1}\mathbf{Z}'\mathbf{Q}_0 \rightarrow_p \boldsymbol{\Sigma}_{zq,0}$ is a full column rank matrix, where $\mathbf{Q}_0 = (\mathbf{G}_0\mathbf{X}\boldsymbol{\beta}_0, \mathbf{X})$.

Recall that ψ denotes the $(k + 1)$ -dimensional vector of parameters, $\psi = (\rho, \boldsymbol{\beta}')'$ and its true value is denoted by $\psi_0 = (\rho_0, \boldsymbol{\beta}'_0)'$. The r linear moment conditions are given by:

$$E_0[\mathbf{Z}'\boldsymbol{\varepsilon}(\psi)] = \mathbf{0}, \quad (18)$$

where

$$\boldsymbol{\varepsilon}(\psi) = \mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta}. \quad (19)$$

Since \mathbf{X} is strictly exogenous under Assumption 3, a possible candidate for \mathbf{Z} consists of linearly independent columns of $(\mathbf{X}, \mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \dots)$. 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 Assumption 4(b), $\mathbf{G}(\rho)$ can be expanded as

$$\mathbf{G}(\rho) = \mathbf{W}(\mathbf{I}_n - \rho\mathbf{W})^{-1} = \mathbf{W} + \rho\mathbf{W}^2 + \rho^2\mathbf{W}^3 + \dots, \quad (20)$$

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 $(\mathbf{W}\mathbf{X}, \mathbf{W}^2\mathbf{X}, \dots)$. Furthermore, Lee (2003) has shown that the asymptotically best IV matrix within the 2SLS framework is given by $\mathbf{Q}_0 = (\mathbf{G}_0\mathbf{X}\boldsymbol{\beta}_0, \mathbf{X})$.

Since \mathbf{Q}_0 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. Using this property we have the following moment condition:

$$E_0 [\boldsymbol{\varepsilon}' (\boldsymbol{\psi}) \mathbf{C} \boldsymbol{\varepsilon} (\boldsymbol{\psi})] = 0, \quad (21)$$

where $\boldsymbol{\varepsilon} (\boldsymbol{\psi})$ is defined by (19),

$$\mathbf{C} = (c_{ij}) = (\mathbf{B} + \mathbf{B}') / 2, \quad (22)$$

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

Assumption 7 *The matrix $\mathbf{B} = (b_{ij})$ is an $n \times n$ matrix of fixed constants such that (a) $\text{diag}(\mathbf{B}) = \mathbf{0}$, (b) $\|\mathbf{B}\|_\infty < K$, (c) $\|\mathbf{B}\|_1 = O(n^{\delta_b})$, where δ_b is a fixed constant in the range $0 \leq \delta_b < 1$, (d) $n^{-1} \mathbf{X}' \mathbf{C} \mathbf{X} \rightarrow_p \boldsymbol{\Sigma}_{xcx}$, $n^{-1} \boldsymbol{\eta}'_0 \mathbf{C} \boldsymbol{\eta}_0 \rightarrow_p c_0$, and $n^{-1} \mathbf{X}' \mathbf{C} \boldsymbol{\eta}_0 \rightarrow_p \mathbf{d}_0$, where $\boldsymbol{\eta}_0$ is given by (8), $\mathbf{C} = (\mathbf{B} + \mathbf{B}') / 2$, and \mathbf{X} is the $n \times k$ matrix of observations on the regressors in model (1).*

Let $\boldsymbol{\Sigma} = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ and denote its true value by $\boldsymbol{\Sigma}_0 = \text{Diag}(\sigma_{10}^2, \sigma_{20}^2, \dots, \sigma_{n0}^2)$. Equation (21) is a valid moment condition under any unknown forms of heteroskedasticity since at the true value $\boldsymbol{\psi}_0$ we have

$$E_0 [\boldsymbol{\varepsilon}' (\boldsymbol{\psi}_0) \mathbf{C} \boldsymbol{\varepsilon} (\boldsymbol{\psi}_0)] = E_0 (\boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon}) = \text{Tr} (\mathbf{C} \boldsymbol{\Sigma}_0) = \text{Tr} (\mathbf{B} \boldsymbol{\Sigma}_0) = \sum_{i=1}^n b_{ii} \sigma_{i0}^2 = 0,$$

which does not require $\boldsymbol{\Sigma}_0$ to be known. 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 $\ell = 1, 2, \dots, L$, where L is a finite number, $\mathbf{C}_\ell = (\mathbf{B}_\ell + \mathbf{B}'_\ell) / 2$, and \mathbf{B}_ℓ satisfies the conditions of Assumption 7.

In the special case where the errors are homoskedastic, Assumption 7(a) can be replaced by the weaker requirement, namely $\text{Tr} (\mathbf{B}) = 0$. As pointed out by Lin and Lee (2010), using a \mathbf{B} matrix with zero diagonal elements will produce consistent estimates irrespective of whether the errors are homoskedastic or heteroskedastic, but it will be asymptotically less efficient if the errors happen to be homoskedastic. On the other hand, under heteroskedasticity using a \mathbf{B} matrix with zero trace but non-zero diagonal entries will yield inconsistent estimates.

Lee (2007) assumes that \mathbf{B} is uniformly bounded in both row and column sums in absolute value and suggests using $\mathbf{B}_\ell = \mathbf{W}^\ell - n^{-1} \text{Tr} (\mathbf{W}^\ell) \mathbf{I}_n$, for $\ell = 1, 2, \dots, L$, in the quadratic moments if the errors are homoskedastic, where \mathbf{W}^ℓ denotes the ℓ^{th} power of \mathbf{W} . If the errors are heteroskedastic, practical choices of \mathbf{B}_ℓ can be $\mathbf{B}_\ell = \mathbf{W}^\ell - \text{Diag}(\mathbf{W}^\ell)$. In contrast, in our set up where columns of \mathbf{W} need not be bounded (see Assumption 4), in part (c) of

Assumption 7 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 .

We are now ready to define the GMM estimator of ψ_0 of model (1), denoted by $\tilde{\psi} = (\tilde{\rho}, \tilde{\beta}')'$, using both quadratic and linear moment conditions:

$$\tilde{\psi} = \arg \min_{\psi \in \Psi} \mathbf{g}'_n(\psi) (\mathbf{A}'_n \mathbf{A}_n) \mathbf{g}_n(\psi), \quad (23)$$

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

$$\mathbf{g}_n(\psi) = \begin{pmatrix} n^{-1} \boldsymbol{\varepsilon}'(\psi) \mathbf{C} \boldsymbol{\varepsilon}(\psi) \\ n^{-1} \mathbf{Z}' \boldsymbol{\varepsilon}(\psi) \end{pmatrix}, \quad (24)$$

and \mathbf{A}_n is an $(k+1) \times (r+1)$ 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 $\tilde{\psi}$, we first focus on the problem of identification in the case of pure SAR models without exogenous regressors. In this case (1) simplifies to,

$$\mathbf{y} = \rho \mathbf{y}^* + \boldsymbol{\varepsilon}, \quad (25)$$

with Assumption 2 replaced by

Assumption 8 *The parameter ρ of model (25) satisfies $\rho \in \Theta_\rho$, where Θ_ρ is a compact subset of \mathbb{R} . The true value of ρ , denoted by ρ_0 , lies in the interior of the parameter space, Θ_ρ .*

The GMM estimator of ρ_0 in model (25) can be obtained by

$$\tilde{\rho} = \arg \min_{\rho \in \Theta_\rho} g_n^2(\rho), \quad (26)$$

where $g_n(\rho) = n^{-1} \boldsymbol{\varepsilon}'(\rho) \mathbf{C} \boldsymbol{\varepsilon}(\rho)$, and $\boldsymbol{\varepsilon}(\rho) = \mathbf{y} - \rho \mathbf{y}^*$. Proposition 1 below shows that in order to uniquely identify ρ_0 in the pure SAR model (25), 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

$$\tilde{\rho} = \arg \min_{\rho \in \Theta_\rho} [\mathbf{a}'_n \mathbf{g}_n(\rho)]^2, \quad (27)$$

where $\mathbf{g}_n(\rho) = [g_{1,n}(\rho), g_{2,n}(\rho), \dots, g_{L,n}(\rho)]'$, $g_{\ell,n}(\rho) = n^{-1} \boldsymbol{\varepsilon}'(\rho) \mathbf{C}_\ell \boldsymbol{\varepsilon}(\rho)$, for $\ell = 1, 2, \dots, L$, and \mathbf{a}_n is an $L \times 1$ non-zero non-negative vector.

Proposition 1 *Consider the SAR model given by (25), and suppose that Assumptions 1, 4, 7(a)–(c), and 8 hold. Then to uniquely identify ρ_0 it is required that the GMM estimator, defined by (27), 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 $\ell = 1, 2, \dots, L \geq 2$, where $a_{\ell 0} = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1} \mathbf{G}'_0 \mathbf{C}_\ell \mathbf{G}_0 \Sigma_0)$ and $b_{\ell 0} = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1} \mathbf{G}'_0 \mathbf{C}_\ell \Sigma_0)$.*

See Section S1.2 of the Online Supplement for a proof.

Remark 6 When the GMM estimator is based on a single quadratic moment condition, the parameter ρ_0 of model (25) is not uniquely identified and the GMM estimator of ρ computed by minimizing $g_n^2(\rho)$ defined by (26), converges in probability to ρ_0 or $\rho_0 + 2b_0/a_0$, where $a_0 = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0\mathbf{C}\mathbf{G}_0\Sigma_0)$ and $b_0 = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0\mathbf{C}\Sigma_0)$. In practice, it is advisable using at least two quadratic moments if the SAR model does not contain any 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 a rate equal to or slower than the rate of $\|\mathbf{W}\|_1$, since in practice \mathbf{W} is commonly adopted as the \mathbf{B} matrix. The following theorem shows that $\psi_0 = (\rho_0, \beta'_0)$ of model (1) can be globally identified if we have enough instruments such that the rank condition in Assumption 6(c) holds. The theorem also establishes consistency and asymptotic normality of the GMM estimator defined by (23).

Theorem 3 Consider the SAR model given by (1). Suppose that Assumptions 1–4, 6 and 7 hold, with $\delta_b = \delta$. Then

- (a) $\psi_0 = (\rho_0, \beta'_0)$ is globally identified,
- (b) the GMM estimator of ψ_0 , denoted by $\tilde{\psi}$ and defined in (23), is consistent for ψ_0 if δ (the degree of network centrality) defined by (4) lies in the range $0 \leq \delta < 1$,
- (c) $\sqrt{n}(\tilde{\psi} - \psi_0)$ is asymptotically normally distributed as $n \rightarrow \infty$, if δ lies in the range $0 \leq \delta < 1/2$, namely,

$$\sqrt{n}(\tilde{\psi} - \psi_0) \xrightarrow{d} N\left[\mathbf{0}, (\mathbf{D}'\mathcal{S}\mathbf{D})^{-1}(\mathbf{D}'\mathcal{S}\mathbf{V}_g\mathcal{S}\mathbf{D})(\mathbf{D}'\mathcal{S}\mathbf{D})^{-1}\right],$$

where $\mathcal{S} = \lim_{n \rightarrow \infty} \mathbf{A}'_n\mathbf{A}_n$, \mathbf{A}_n is the $(k+1) \times (r+1)$ matrix of full row rank used in the GMM minimand, (23),

$$\mathbf{D} = \left\{ \left[2 \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\Sigma_0\mathbf{C}\mathbf{G}_0), \mathbf{0}_{1 \times k} \right]', \Sigma'_{zq_0} \right\}', \quad (28)$$

$$\mathbf{V}_g = \begin{pmatrix} 2 \lim_{n \rightarrow \infty} \text{Tr}[n^{-1}(\Sigma_0\mathbf{C})^2] & \mathbf{0}_{1 \times r} \\ \mathbf{0}_{r \times 1} & p \lim_{n \rightarrow \infty} n^{-1}\mathbf{Z}'\Sigma_0\mathbf{Z} \end{pmatrix}, \quad (29)$$

\mathbf{G}_0 is defined by (8), $\Sigma_{zq_0} = p \lim_{n \rightarrow \infty} n^{-1}\mathbf{Z}'\mathbf{Q}_0$, and $\Sigma_0 = \text{Diag}(\sigma_{10}^2, \sigma_{20}^2, \dots, \sigma_{n0}^2)$.

The proof of Theorem 3 is given in Section S1.2 of the Online Supplement.

A few comments on Theorem 3 are in order. First, it is worth emphasizing that ψ_0 is globally identified if we have enough instruments such that the rank condition in Assumption 6(c) is satisfied, irrespective of the number of quadratic moments included. Using a finite number of quadratic moments in addition to linear moments improves efficiency.

Second, Theorem 3 is consistent with the result in Lin and Lee (2010), who extended the GMM estimator for SAR models proposed by Lee (2007) to the case of unknown heteroskedasticity assuming $\delta = 0$.

Third, if the errors are identically distributed with variances $\sigma_{i0}^2 = \sigma_0^2$, for all i , and Assumption 7(a) is replaced by $\text{Tr}(\mathbf{B}) = 0$, then the main conclusions in Theorem 3 are unaffected except that the expressions, (28) and (29), in the asymptotic variance need to be changed to:

$$\mathbf{D} = \left[(2\sigma_0^2 b_0, \mathbf{0}_{1 \times k})', \boldsymbol{\Sigma}'_{zq_0} \right]', \quad \mathbf{V}_g = \begin{pmatrix} v_{10} & \mu_{30} \mathbf{v}' \\ \mu_{30} \mathbf{v} & \sigma_0^2 \boldsymbol{\Sigma}_{zz} \end{pmatrix}, \quad (30)$$

$$v_{10} = \lim_{n \rightarrow \infty} \left[\gamma_{20} n^{-1} \sum_{i=1}^n c_{ii}^2 + 2\sigma_0^4 \text{Tr}(n^{-1} \mathbf{C}^2) \right],$$

where $b_0 = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1} \mathbf{G}'_0 \mathbf{C})$, $\boldsymbol{\Sigma}_{zz} = p \lim_{n \rightarrow \infty} n^{-1} \mathbf{Z}' \mathbf{Z}$, $\mathbf{v} = p \lim_{n \rightarrow \infty} n^{-1} \mathbf{Z}' [\text{diag}(\mathbf{C})]$, $\mu_{30} = E(\varepsilon_i^3)$, $\gamma_{20} = E(\varepsilon_i^4) - 3\sigma_0^4$, and c_{ii} is the i^{th} diagonal element of \mathbf{C} . Observe that the asymptotic distribution given by Theorem 3(c) does not depend on the third and fourth order moments of the errors since \mathbf{B} is restricted so that $\text{diag}(\mathbf{B}) = \mathbf{0}$. In contrast, (30) which assumes homoskedastic errors and only requires $\text{Tr}(\mathbf{B}) = 0$, does involve higher-order moments of the errors, ε_i .

Remark 7 To consistently estimate the variance of the GMM estimator, one can replace $\boldsymbol{\Sigma}$ with $\hat{\boldsymbol{\Sigma}} = \text{Diag}(\tilde{\varepsilon}_1^2, \tilde{\varepsilon}_2^2, \dots, \tilde{\varepsilon}_n^2)$, where $\tilde{\varepsilon}_i = y_i - \tilde{\rho}y_i^* - \tilde{\beta}' \mathbf{x}_i$. In the case of homoskedastic errors and if $\text{diag}(\mathbf{B}) \neq \mathbf{0}$, consistent estimators of μ_{30} and γ_{20} are given by $\tilde{\mu}_3 = n^{-1} \sum_{i=1}^n \tilde{\varepsilon}_i^3$ and $\tilde{\gamma}_2 = n^{-1} \sum_{i=1}^n \tilde{\varepsilon}_i^4 - 3(\tilde{\sigma}^2)^2$, respectively, where $\tilde{\sigma}^2 = n^{-1} \sum_{j=1}^n \tilde{\varepsilon}_j^2$.

As is well known in the GMM literature, the optimal moments weighting matrix is given by \mathbf{V}_g^{-1} . A feasible optimal GMM (OGMM) estimator of ψ_0 , denoted by $\tilde{\psi}_{opt}$, can be obtained by using a consistent estimator of \mathbf{V}_g^{-1} for $\mathbf{A}'_n \mathbf{A}_n$, that is,

$$\tilde{\psi}_{opt} = \arg \min_{\psi \in \Psi} \mathbf{g}'_n(\psi) \tilde{\mathbf{V}}_g^{-1} \mathbf{g}'_n(\psi), \quad (31)$$

where $\mathbf{g}_n(\psi)$ is given by (24) and $\tilde{\mathbf{V}}_g$ is a consistent estimator of \mathbf{V}_g . Then $\tilde{\psi}_{opt}$ is consistent for ψ_0 when δ is in the range $0 \leq \delta < 1$, and it has the following asymptotic distribution as $n \rightarrow \infty$ when δ lies in the range $0 \leq \delta < 1/2$,

$$\sqrt{n} (\tilde{\psi}_{opt} - \psi_0) \rightarrow_d N \left[\mathbf{0}, (\mathbf{D}' \mathbf{V}_g^{-1} \mathbf{D})^{-1} \right].$$

The best choice of \mathbf{B} exists under certain conditions. Lee (2007) shows that if the idiosyncratic errors are identically and normally distributed, the OGMM estimator using $\mathbf{G}_0 - n^{-1} \text{Tr}(\mathbf{G}_0) \mathbf{I}_n$ in the quadratic moment condition and $\mathbf{Q}_0 = (\mathbf{G}_0 \mathbf{X} \boldsymbol{\beta}_0, \mathbf{X})$ in the linear moment conditions, has the smallest asymptotic variance among the set of GMM estimators derived with the class of matrices, \mathbf{B}_ℓ , having zero trace. Among the group of GMM estimators derived with the class of matrices having zero diagonal, the OGMM estimator using

$\mathbf{G}_0 - Diag(\mathbf{G}_0)$ and \mathbf{Q}_0 in the moments has the smallest asymptotic variance.⁸ The OGMM estimator using $\mathbf{G}_0 - n^{-1}Tr(\mathbf{G}_0)\mathbf{I}_n$ (or $\mathbf{G}_0 - Diag(\mathbf{G}_0)$) and \mathbf{Q}_0 are referred to as the best GMM estimator. By a similar argument and applying Lemma S.6 of the Online Supplement, 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{G}_0 and \mathbf{Q}_0 depend on unknown parameters, a feasible best GMM estimator can be implemented in two steps: In the first step, a preliminary consistent estimate of ψ_0 is obtained, which is then used in the second step to compute the optimal GMM estimates using \mathbf{Q}_0 and $\mathbf{G}_0 - n^{-1}Tr(\mathbf{G}_0)\mathbf{I}_n$ if assuming homoskedasticity (or $\mathbf{G}_0 - Diag(\mathbf{G}_0)$ if assuming heteroskedasticity) 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 $\psi_0 = (\rho_0, \beta'_0)' = (\rho_0, \beta'_0, \sigma_0^2)'$ 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}^* . To clarify the idea, let us first consider the case of homoskedastic errors and then relax it to allow for unknown heteroskedasticity. Let $\theta_0 = (\psi'_0, \sigma_0^2)' = (\rho_0, \beta'_0, \sigma_0^2)'$ denote the vector of unknown parameters of the SAR model under homoskedasticity, where $\varepsilon_i \sim IID(0, \sigma_0^2)$ and $\sup_i E|\varepsilon_i|^{4+c} < K$, for some $c > 0$. 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

$$E[\mathbf{y}^{*'}(\mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\beta)] = E(\mathbf{y}^{*'}\varepsilon), \quad (32)$$

$$E[\mathbf{X}'(\mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\beta)] = \mathbf{0}, \quad (33)$$

$$E[(\mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\beta)'(\mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\beta)] = n\sigma^2. \quad (34)$$

Using (7), we have $E(\mathbf{y}^{*'}\varepsilon) = E[(\beta'\mathbf{X}' + \varepsilon')\mathbf{G}'(\rho)\varepsilon]$, and under Assumption 3(a) and the IID assumption of ε_i , we obtain $E(\mathbf{y}^{*'}\varepsilon) = \sigma^2 Tr[\mathbf{G}(\rho)]$. The sample version of the moment conditions (32)–(34) can now be written as

$$n^{-1}\mathbf{y}^{*'}(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\beta}) = \hat{\sigma}^2 Tr[n^{-1}\mathbf{G}(\hat{\rho})], \quad (35)$$

$$n^{-1}\mathbf{X}'(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\beta}) = \mathbf{0}, \quad (36)$$

⁸This result does not require the condition that the idiosyncratic errors are normally distributed. See Lee (2007) Proposition 3 for details.

⁹We also examined the finite sample properties of GMM estimators that use other instruments and \mathbf{B} matrices. The results are provided in the Online Supplement, and as can be seen overall they have less satisfactory small sample properties.

$$n^{-1} \left(\mathbf{y} - \hat{\rho} \mathbf{y}^* - \mathbf{X} \hat{\boldsymbol{\beta}} \right)' \left(\mathbf{y} - \hat{\rho} \mathbf{y}^* - \mathbf{X} \hat{\boldsymbol{\beta}} \right) = \hat{\sigma}^2. \quad (37)$$

Let $\hat{\boldsymbol{\theta}} = (\hat{\rho}, \hat{\boldsymbol{\beta}}', \hat{\sigma}^2)'$ denote the BMM estimator of $\boldsymbol{\theta}_0$. The system of equations (35)–(37) can now be used to solve for $\hat{\boldsymbol{\theta}}$ as follows:

$$\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta} \in \Theta} \mathbf{m}'_n(\boldsymbol{\theta}) \mathbf{m}_n(\boldsymbol{\theta}), \quad (38)$$

where $\mathbf{m}_n(\boldsymbol{\theta}) = [m_{1,n}(\boldsymbol{\theta}), \mathbf{m}'_{2,n}(\boldsymbol{\theta}), m_{3,n}(\boldsymbol{\theta})]'$ with $m_{1,n}(\boldsymbol{\theta}) = n^{-1} \mathbf{y}^{*\prime} \boldsymbol{\varepsilon}(\psi) - \sigma^2 \text{Tr}[n^{-1} \mathbf{G}(\rho)]$, $\mathbf{m}_{2,n}(\boldsymbol{\theta}) = n^{-1} \mathbf{X}' \boldsymbol{\varepsilon}(\psi)$, $m_{3,n}(\boldsymbol{\theta}) = n^{-1} \boldsymbol{\varepsilon}'(\psi) \boldsymbol{\varepsilon}(\psi) - \sigma^2$, $\boldsymbol{\varepsilon}(\psi)$ is given by (19), $\Theta = \Theta_\rho \times \Theta_\beta \times \Theta_{\sigma^2}$, and Θ_{σ^2} is a compact subspace of $(0, \infty)$ containing the true value σ_0^2 .

Unlike least squares, the BMM procedure is non-linear in $\hat{\rho}$, and its asymptotic properties 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, (35)–(37), with additional moment conditions. See, for example, Lee (2007).

We show in the Online Supplement that the BMM estimator under homoskedasticity, defined by (38), is a consistent estimator when δ is in the range $0 \leq \delta < 1$, where δ is a measure of network centrality, defined by (4). Moreover, when δ is in the range $0 \leq \delta < 1/2$, $\sqrt{n}(\hat{\psi} - \psi_0)$ is asymptotically normally distributed as:

$$\sqrt{n}(\hat{\psi} - \psi_0) \rightarrow_d N[\mathbf{0}, (\mathbf{H}^{-1} \mathbf{V} \mathbf{H}^{-1})], \quad (39)$$

as $n \rightarrow \infty$, where

$$\mathbf{H} = \begin{pmatrix} \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0g_0x} \boldsymbol{\beta}_0 + \sigma_0^2 h_0 & \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0x} \\ \boldsymbol{\Sigma}_{xg_0x} \boldsymbol{\beta}_0 & \boldsymbol{\Sigma}_{xx} \end{pmatrix}, \quad \mathbf{V} = \begin{pmatrix} q_0^2 & \sigma_0^2 \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0x} \\ \sigma_0^2 \boldsymbol{\Sigma}_{xg_0x} \boldsymbol{\beta}_0 & \sigma_0^2 \boldsymbol{\Sigma}_{xx} \end{pmatrix}, \quad (40)$$

$$q_0^2 = \sigma_0^2 \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0g_0x} \boldsymbol{\beta}_0 + \gamma_{20} p \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \pi_{ii,0}^2 + 2\mu_{30} p \lim_{n \rightarrow \infty} n^{-1} [\text{diag}(\boldsymbol{\Pi}_0)]' \mathbf{G}_0 \mathbf{X} \boldsymbol{\beta}_0 \quad (41)$$

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

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

$$h_0 = \lim_{n \rightarrow \infty} \{n^{-1} Tr(\mathbf{G}_0^2 + \mathbf{G}'_0 \mathbf{G}_0) - 2n^{-2} [Tr(\mathbf{G}_0)]^2\}, \quad (43)$$

where \mathbf{G}_0 is defined by (8), $\pi_{ii,0}$ is the i^{th} diagonal element of $\boldsymbol{\Pi}_0$, $\mu_{30} = E(\varepsilon_i^3)$ and $\gamma_{20} = E(\varepsilon_i^4) - 3\sigma_0^4$.

Remark 8 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 under homoskedasticity 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 $\hat{\varepsilon}_i = y_i - \hat{\rho}y_i^* - \hat{\beta}'\mathbf{x}_i$, and γ_{20} can be estimated by $\hat{\gamma}_2 = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i^4 - 3(\hat{\sigma}^2)^2$.

We now turn to considering the BMM estimation under general forms of heteroskedasticity where the errors satisfy Assumption 1. Recall that $\Sigma = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ and note that

$$E(\mathbf{y}^{*\prime}\boldsymbol{\varepsilon}) = E[(\boldsymbol{\beta}'\mathbf{X}' + \boldsymbol{\varepsilon}')\mathbf{G}'(\rho)\boldsymbol{\varepsilon}] = \text{Tr}[\mathbf{G}(\rho)\Sigma] = \sum_{i=1}^n g_{ii}(\rho) \sigma_i^2, \quad (44)$$

where $g_{ii}(\rho)$ is the i^{th} diagonal element of $\mathbf{G}(\rho)$. In view of (44), the moment condition (32) becomes $E[\mathbf{y}^{*\prime}(\mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})] = \sum_{i=1}^n g_{ii}(\rho) \sigma_i^2$, which can be written equivalently as

$$E[\mathbf{y}^{*\prime}(\mathbf{y} - \rho\mathbf{y}^* - \mathbf{X}\boldsymbol{\beta})] = E\left[\sum_{i=1}^n g_{ii}(\rho) \varepsilon_i^2\right] = E[\boldsymbol{\varepsilon}'\check{\mathbf{G}}(\rho)\boldsymbol{\varepsilon}], \quad (45)$$

where $\check{\mathbf{G}}(\rho) = \text{Diag}[\mathbf{G}(\rho)]$. In sum, the BMM population moment conditions under heteroskedasticity are given by (45) and (33), and the sample counterparts are

$$n^{-1}\mathbf{y}^{*\prime}(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}) = n^{-1}(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}})' \check{\mathbf{G}}(\hat{\rho})(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}), \quad (46)$$

$$n^{-1}\mathbf{X}'(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{0}. \quad (47)$$

Now let $\hat{\psi} = (\hat{\rho}, \hat{\boldsymbol{\beta}}')'$ denote the BMM estimator of ψ_0 assuming heteroskedastic errors. To avoid introducing further notations, we use the same notations as in the case of homoskedastic errors. The BMM estimator of ψ_0 , is obtained by solving the system of equations (46) and (47) for $\hat{\psi}$, which can be written equivalently as

$$\hat{\psi} = \arg \min_{\psi \in \Psi} \mathbf{m}_n'(\psi) \mathbf{m}_n(\psi), \quad (48)$$

where $\mathbf{m}_n(\psi) = [m_{1,n}(\psi), m_{2,n}(\psi)]'$ with $m_{1,n}(\psi) = n^{-1}\mathbf{y}^{*\prime}\boldsymbol{\varepsilon}(\psi) - n^{-1}\boldsymbol{\varepsilon}'(\psi)\check{\mathbf{G}}(\rho)\boldsymbol{\varepsilon}(\psi)$, $m_{2,n}(\psi) = n^{-1}\mathbf{X}'\boldsymbol{\varepsilon}(\psi)$, and as before $\boldsymbol{\varepsilon}(\psi)$ is given by (19).

The following theorem formally summarizes the asymptotic distribution of the BMM estimator assuming heteroskedasticity. Its proof is given in Section S1.2 of the Online Supplement.

Theorem 4 Consider the SAR model given by (1), and suppose that Assumptions 1–5 hold. Then

(a) the bias-corrected method of moments (BMM) estimator of $\psi_0 = (\rho_0, \boldsymbol{\beta}'_0)'$, denoted by $\hat{\psi} = (\hat{\rho}, \hat{\boldsymbol{\beta}}')'$ and defined by (48), is consistent for ψ_0 when δ , the centrality of the weights matrix \mathbf{W} defined by (4), lies in the range $0 \leq \delta < 1$.

(b) $\sqrt{n}(\hat{\psi} - \psi_0)$ is asymptotically normally distributed as $n \rightarrow \infty$, when δ is in the range $0 \leq \delta < 1/2$, namely

$$\sqrt{n}(\hat{\psi} - \psi_0) \rightarrow_d N[\mathbf{0}, (\mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1})], \quad (49)$$

where

$$\mathbf{H} = \begin{pmatrix} \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0g_0x} \boldsymbol{\beta}_0 + h_0 & \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0x} \\ \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xx} \end{pmatrix}, \quad \mathbf{V} = p \lim_{n \rightarrow \infty} \begin{pmatrix} q_0^2 & n^{-1} \boldsymbol{\beta}'_0 \mathbf{X}' \mathbf{G}'_0 \boldsymbol{\Sigma}_0 \mathbf{X} \\ n^{-1} \mathbf{X}' \boldsymbol{\Sigma}_0 \mathbf{G}_0 \mathbf{X} \boldsymbol{\beta}_0 & n^{-1} \mathbf{X}' \boldsymbol{\Sigma}_0 \mathbf{X} \end{pmatrix}, \quad (50)$$

$$q_0^2 = n^{-1} \boldsymbol{\beta}'_0 \mathbf{X}' \mathbf{G}'_0 \boldsymbol{\Sigma}_0 \mathbf{G}_0 \mathbf{X} \boldsymbol{\beta}_0 + \text{Tr} [n^{-1} \boldsymbol{\Sigma}_0 \boldsymbol{\Pi}_0 (\boldsymbol{\Sigma}_0 \boldsymbol{\Pi}_0 + \boldsymbol{\Sigma}_0 \boldsymbol{\Pi}'_0)], \quad \boldsymbol{\Pi}_0 = \mathbf{G}_0 - \check{\mathbf{G}}_0, \quad (51)$$

$$h_0 = \lim_{n \rightarrow \infty} \text{Tr} [n^{-1} (\mathbf{G}_0^2 + \mathbf{G}'_0 \mathbf{G}_0 - 2\check{\mathbf{G}}_0 \mathbf{G}_0) \boldsymbol{\Sigma}_0], \quad (52)$$

\mathbf{G}_0 is given by (8), $\check{\mathbf{G}}_0 = \text{Diag}(\mathbf{G}_0)$, and $\boldsymbol{\Sigma}_0 = \text{Diag}(\sigma_{10}^2, \sigma_{20}^2, \dots, \sigma_{n0}^2)$.

A number of remarks are in order. First, it should be noted that unlike linear regressions, but similar to the GMM estimator discussed above, the BMM estimator assuming homoskedasticity given by (48) will result in inconsistent estimates if the errors are in fact heteroskedastic. On the other hand, if the errors are homoskedastic, the BMM estimator assuming heteroskedasticity given by (38) will produce consistent but asymptotically less efficient estimates.

Second, it is clear from (49) that $\boldsymbol{\psi}_0$ is identified if \mathbf{H} , defined in (50), is positive definite. Notice that $\mathbf{H} = \mathbf{H}_1 + \mathbf{H}_2$, where $\mathbf{H}_1 = p \lim_{n \rightarrow \infty} n^{-1} \mathbf{Q}'_0 \mathbf{Q}_0$, and

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

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 5 ensures that $\boldsymbol{\psi}_0$ is identified. Moreover, if $\rho_0 \geq 0$, the identification condition given by Assumption 5(b) is ensured if

$$\text{Tr} [n^{-1} (\mathbf{G}_0^2 + \mathbf{G}'_0 \mathbf{G}_0 - 2\check{\mathbf{G}}_0 \mathbf{G}_0)] > c > 0, \quad (53)$$

for all n , a condition that does not depend on the unknown form of $\boldsymbol{\Sigma}_0$. That is, the degree of heteroskedasticity does not affect consistency of the estimator.¹⁰

Third, if the errors are homoskedastic, it is immediate from (39) that $\boldsymbol{\psi}_0$ is identified if $n^{-1} \mathbf{Q}'_0 \mathbf{Q}_0$ is positive definite, and/or

$$n^{-1} \text{Tr} (\mathbf{G}_0^2 + \mathbf{G}'_0 \mathbf{G}_0) - 2n^{-2} [\text{Tr} (\mathbf{G}_0)]^2 > c > 0, \quad (54)$$

for all n (including $n \rightarrow \infty$). Also notice that (54) is implied by (53) because

$$\text{Tr} (\check{\mathbf{G}}_0 \mathbf{G}_0) = \sum_{i=1}^n g_{ii,0}^2 \geq \frac{1}{n} \left(\sum_{i=1}^n g_{ii,0} \right)^2 = \frac{1}{n} [\text{Tr} (\mathbf{G}_0)]^2.$$

Therefore, if (53) holds and $\rho_0 \geq 0$, identification is achieved irrespective of whether the errors are homoskedastic or not.

¹⁰To see that (53) implies Assumption 5(b), let $\mathbf{A}_0 = (a_{ij,0}) = \mathbf{G}_0^2 + \mathbf{G}'_0 \mathbf{G}_0 - 2\check{\mathbf{G}}_0 \mathbf{G}_0$ and $\mathbf{G}_0 = (g_{ij,0})$. It is easy to verify that $a_{ii,0} = \sum_{j=1, j \neq i}^n (g_{ij,0} g_{ji,0} + g_{ji,0}^2)$. But under Assumption 4 and $\rho_0 \geq 0$, we have $g_{ij,0} \geq 0$, for all i and j , and then $\text{Tr} (n^{-1} \mathbf{A}_0 \boldsymbol{\Sigma}_0) = \sum_{i=1}^n \sigma_{i0}^2 a_{ii,0} \geq \inf_i (\sigma_{i0}^2) \text{Tr} (n^{-1} \mathbf{A}_0)$. Therefore, $\text{Tr} (n^{-1} \mathbf{A}_0) > c > 0$ implies $\text{Tr} (n^{-1} \mathbf{A}_0 \boldsymbol{\Sigma}_0) > c > 0$.

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 S1.2 of the Online Supplement.

Proposition 2 Consider the SAR model given by (1), and assume that the errors are independently and normally distributed as $\varepsilon_i \sim \text{IIDN}(0, \sigma^2)$, for $i = 1, 2, \dots, n$, and $0 < c < \sigma^2 < K$. Suppose that Assumptions 2–5 hold and the network centrality, δ , defined by (4), lies in the range $0 \leq \delta < 1/2$. Then the BMM estimator of $\psi_0 = (\rho_0, \beta'_0)'$, defined by (38), has the same asymptotic distribution as the best GMM estimator of ψ_0 , defined by (31) using $\mathbf{G}_0 - n^{-1} \text{Tr}(\mathbf{G}_0) \mathbf{I}_n$ in the quadratic moment condition, and $(\mathbf{G}_0 \mathbf{X} \beta_0, \mathbf{X})$ in the linear moment conditions, where \mathbf{G}_0 is defined by (8).

We complete the discussion of our theoretical results with the following two examples.

Example 1 Consider the following star network:

$$\mathbf{W} = \begin{pmatrix} 0 & w_{12} & 0 & 0 & \dots & 0 & 0 \\ w_{21} & 0 & w_{23} & 0 & \dots & 0 & 0 \\ w_{31} & 0 & 0 & w_{34} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ w_{n-1,1} & 0 & 0 & 0 & \dots & 0 & w_{n-1,n} \\ w_{n1} & 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}, \quad (55)$$

where, without loss of generality, the first unit is the star (dominant unit). As established in Lemma S.11 of the Online Supplement, a necessary condition for (54) is given by $n^{-1} \text{Tr}(\mathbf{W}' \mathbf{W}) > c > 0$, for all n (including $n \rightarrow \infty$). Note that this condition is necessary for (54), irrespective of whether there exist dominant units (with $\delta > 0$) or not,¹¹ and is also necessary for (53) under heteroskedasticity and $\rho_0 \geq 0$, as discussed above. For this example it is easily seen that

$$n^{-1} \text{Tr}(\mathbf{W}' \mathbf{W}) = n^{-1} \sum_{i=2}^n w_{i1}^2 + n^{-1} \sum_{i=1}^{n-1} w_{i,i+1}^2. \quad (56)$$

The first term refers to the strength of the dominant unit (the first unit), and the second term captures the strength of the connections of non-dominant units to the dominant unit. In the absence of regressors, identification of ρ_0 requires that at least one of the two terms of (56) to tend to a non-zero value as $n \rightarrow \infty$. In the case where there is no dominant unit or the strength of the dominant unit is weak then $n^{-1} \sum_{i=2}^n w_{i1}^2 \rightarrow 0$, as $n \rightarrow \infty$, and identification requires $\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^{n-1} w_{i,i+1}^2 > c > 0$. This condition is met if $\inf_i |w_{i,i+1}| > c$, namely if the dominant unit is impacted by almost all other units in the network. The \mathbf{W} matrix in our Monte Carlo experiments follows a similar structure as (55).

¹¹Yang (2018) discussed this identification condition for SAR models with no dominant units.

Example 2 Although the current paper considers asymmetric \mathbf{W} to allow for the presence of dominant units, the discussions on identification apply equally to symmetric \mathbf{W} . Let us first consider a simple example where there is only one social group in which everyone is connected to one another, and $w_{ii} = 0$. In this case, the matrix of connections is represented by $\mathbf{W} = (n - 1)^{-1} (\mathbf{1}_n \mathbf{1}'_n - \mathbf{I}_n)$. It is easily verified that $n^{-1} \text{Tr}(\mathbf{W}^2) = 1/(n - 1)$, which tends to zero, as $n \rightarrow \infty$. Therefore, the necessary condition for identification, namely, $n^{-1} \text{Tr}(\mathbf{W}'\mathbf{W}) > c > 0$ for all n (including $n \rightarrow \infty$), is violated and 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 $r = 1, 2, \dots, R$. 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:

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

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

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

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 $n \rightarrow \infty$. If R is fixed, then $\lim_{n \rightarrow \infty} n^{-1} \text{Tr}(\mathbf{W}^2) = 0$ and the group interaction effect is unidentified in the absence of exogenous explanatory variables.

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:

$$y_i = \alpha + \rho y_i^* + \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (57)$$

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

$$x_i = \lambda x_i^* + \nu_i, \quad i = 1, 2, \dots, n, \quad (58)$$

where $x_i^* = \mathbf{w}'_{i,x} \mathbf{x}$, $\mathbf{x} = (x_1, x_2, \dots, x_n)'$, and $\mathbf{w}'_{i,x}$ 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, (57) can be rewritten as

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

where $\mathbf{S}_y(\rho) = \mathbf{I}_n - \rho \mathbf{W}_y$, $\mathbf{u} = \mathbf{S}_y^{-1}(\rho) \boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$, and $\mathbf{u} = (u_1, u_2, \dots, u_n)'$. Similarly, (58) can be rewritten as $\mathbf{x} = \mathbf{S}_x^{-1}(\lambda) \boldsymbol{\nu}$, where $\mathbf{S}_x(\lambda) = \mathbf{I}_n - \lambda \mathbf{W}_x$ and $\boldsymbol{\nu} = (\nu_1, \nu_2, \dots, \nu_n)'$. For the idiosyncratic errors, we consider both Gaussian and non-Gaussian processes:

- Gaussian errors: $\varepsilon_i \sim IIDN(0, \sigma_i^2)$ and $\nu_i \sim IIDN(0, \sigma_{\nu,i}^2)$.
- Non-Gaussian errors: $\varepsilon_i/\sigma_i \sim IID[\chi^2(2) - 2]/2$ and $\nu_i/\sigma_{\nu,i} \sim IID[\chi^2(2) - 2]/2$, where $\chi^2(2)$ denotes a chi-square random variable with two degrees of freedom.

The error variances are generated as $\sigma_i^2 = \sigma_\varepsilon^2 \tau_{\varepsilon,i}^2$ and $\sigma_{\nu,i}^2 = \sigma_\nu^2 \tau_{\nu,i}^2$. Then $Var(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{D}_\varepsilon$, where $\mathbf{D}_\varepsilon = Diag(\tau_{\varepsilon,1}^2, \tau_{\varepsilon,2}^2, \dots, \tau_{\varepsilon,n}^2)$, and $Var(\boldsymbol{\nu}) = \sigma_\nu^2 \mathbf{D}_\nu$, where $\mathbf{D}_\nu = Diag(\tau_{\nu,1}^2, \tau_{\nu,2}^2, \dots, \tau_{\nu,n}^2)$. We consider both homoskedastic and heteroskedastic errors:

- Homoskedastic errors: $\tau_{\varepsilon,i}^2 = \tau_{\nu,i}^2 = 1$, for all i . Note that in this case $\sigma_i^2 = \sigma_\varepsilon^2$ and $\sigma_{\nu,i}^2 = \sigma_\nu^2$, for all i , and $\mathbf{D}_\varepsilon = \mathbf{D}_\nu = \mathbf{I}_n$.
- Heteroskedastic errors: $\tau_{\varepsilon,i}^2 \sim IIDU(0.5, 1.5)$ and $\tau_{\nu,i}^2 \sim IIDU(0.5, 1.5)$, where $U(a, b)$ denotes the uniform distribution on the interval (a, b) .

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

$$R_0^2 = 1 - \frac{Tr[Var(\boldsymbol{\varepsilon})]}{Tr[Var(\mathbf{y})]} = 1 - \frac{Tr(\mathbf{D}_\varepsilon)}{Tr[\mathbf{S}_y^{-1}(\rho) \mathbf{D}_\varepsilon \mathbf{S}_y'^{-1}(\rho)]}, \quad (59)$$

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

$$R_\beta^2 = 1 - \frac{Tr[Var(\boldsymbol{\varepsilon})]}{Tr[Var(\mathbf{y})]}, \quad (60)$$

where

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

It is also easily seen that

$$R_\beta^2 - R_0^2 = \frac{a_n s^2 (1 - R_0^2)}{1 + a_n s^2} \geq 0,$$

where

$$a_n = \frac{Tr[\mathbf{S}_y^{-1}(\rho) \mathbf{S}_x^{-1}(\lambda) \mathbf{D}_\nu \mathbf{S}_x'^{-1}(\lambda) \mathbf{S}_y'^{-1}(\rho)]}{Tr[\mathbf{S}_y^{-1}(\rho) \mathbf{D}_\varepsilon \mathbf{S}_y'^{-1}(\rho)]} > 0, \quad s^2 = \frac{\beta^2 \sigma_v^2}{\sigma_\varepsilon^2} \geq 0, \quad (61)$$

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 other 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

$$\frac{\beta^2 \sigma_v^2}{\sigma_\varepsilon^2} \times \frac{Tr [\mathbf{S}_y^{-1}(\rho) \mathbf{S}_x^{-1}(\lambda) \mathbf{D}_\nu \mathbf{S}_x'^{-1}(\lambda) \mathbf{S}_y'^{-1}(\rho)]}{Tr [\mathbf{S}_y^{-1}(\rho) \mathbf{D}_\varepsilon \mathbf{S}_y'^{-1}(\rho)]} = \frac{0.1}{0.9 - R_0^2},$$

or equivalently,

$$\sigma_v^2 = \left(\frac{0.1}{0.9 - R_0^2} \right) \frac{\sigma_\varepsilon^2}{\beta^2 a_n}, \quad (62)$$

where a_n is defined by (61). The value of ρ is now 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 report the results for the choices with $\mathbf{W}_x \neq \mathbf{W}_y$ in the Online Supplement. The spatial weights matrix \mathbf{W} ,

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

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 $\lfloor \cdot \rfloor$ 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., $\sum_{i=1}^n w_{i1} = \Theta(n^\delta)$. The first 8 elements of the $1 \times (n-1)$ row vector \mathbf{w}'_{12} 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 $j = i-4, \dots, i-1, i+1, \dots, i+4$, and $w_{i,j} = 0$ otherwise. By construction, \mathbf{W}_{22} is uniformly bounded in both row and column norms, namely, $\|\mathbf{W}_{22}\|_1 = O(1)$ and $\|\mathbf{W}_{22}\|_\infty = O(1)$. Finally, \mathbf{W} is standardized so that each row sums to one.¹²

We consider a number of different values of δ and ρ_0 : $\delta = 0, 0.25, 0.50, 0.75, 0.95, 1$, and $\rho_0 = 0.2, 0.5, 0.75$;¹³ and experiment with four sample sizes: $n = 100, 300, 500$, and $1,000$. 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 assuming homoskedastic errors is computed by (38), and by (48) under heteroskedastic errors. The GMM estimator reported here is the best GMM estimator.¹⁴ Specifically, the GMM

¹²Row standardization is unnecessary and only for convenience. Recall that our theory only requires the row sums of \mathbf{W} to be uniformly bounded.

¹³The values of R_0^2 for different ρ and δ are reported in Table 59 in the Online Supplement. 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 Online Supplement.

¹⁴We also consider GMM estimators using other IV matrix, \mathbf{Z} , and other \mathbf{B} matrices. The results are presented in the Monte Carlo supplement.

estimator assuming homoskedastic errors is computed in two steps: In the first step, the GMM estimates are computed with equal weights using $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr(\mathbf{W}^2) \mathbf{I}_n$, and $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x})$. In the second step, we re-estimate with the optimal GMM weights using the best IV matrix, $(\tilde{\mathbf{G}}\mathbf{x}\tilde{\alpha}, \tilde{\mathbf{G}}\mathbf{x}\tilde{\beta}, \mathbf{1}_n, \mathbf{x})$, and $\tilde{\mathbf{G}} - n^{-1}Tr(\tilde{\mathbf{G}}) \mathbf{I}_n$ in the quadratic moment, where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\rho})$, and $(\tilde{\rho}, \tilde{\alpha}, \tilde{\beta})'$ denote the first-step GMM estimates. Similarly, the GMM estimator assuming heteroskedastic errors is obtained in two steps, but it uses different matrices in the quadratic moments, namely, $\mathbf{B}_1 = \mathbf{W}$ and $\mathbf{B}_2 = \mathbf{W}^2 - Diag(\mathbf{W})$ in the first step, and $\tilde{\mathbf{G}} - Diag(\tilde{\mathbf{G}})$ in the second step.

Tables 1a–2b summarize the results of the GMM and BMM estimators for the experiments with homoskedastic Gaussian errors, Tables 3a–4b give the results for homoskedastic non-Gaussian errors, and Tables 5a–6b report the results for heteroskedastic non-Gaussian errors.¹⁵ For each experiment, we report bias, root mean square error (RMSE), size, and power of both estimators for ρ and β . The results for the intercept term are not reported 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 homoskedastic non-Gaussian.¹⁶

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. Both estimators exhibit similar bias and RMSE properties under heteroskedasticity, as shown in Tables 5a and 6a.

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 comparison, the BMM estimator has the correct empirical size even when the sample size is

¹⁵The Monte Carlo results for heteroskedastic Gaussian errors are provided in the Online Supplement to save space.

¹⁶Power function plots when errors are heteroskedastic and for other values of ρ_0 , δ and n are reported in the Online Supplement.

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 to the results we obtain for ρ . These findings seem to be quite robust to non-Gaussian errors. Finally, both estimators display larger size distortion under heteroskedasticity than under homoskedasticity when $n = 100$ and $\delta \leq 0.75$, but they become correctly sized when n increases to 300 and $\delta \leq 0.75$. As in the case of homoskedasticity, the GMM estimator is more severely over-sized than the BMM estimator under heteroskedasticity if δ is close to unity and the sample size is small. Overall, our extensive simulation results suggest that in practice one can perform sound inference with the BMM estimator so long as δ is not too close to unity.

Figures 1a and 1b display the power functions for ρ under homoskedasticity 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.¹⁷ 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. The power functions under heteroskedasticity resemble the plots under homoskedasticity with slightly larger size distortion when the sample size is small. See the Online Supplement for additional empirical power functions under heteroskedasticity.

7 Empirical application to US sectoral prices

In earlier studies using US input-output tables Acemoglu et al. (2012) and Pesaran and Yang (2019) find that δ , the degree of centrality of 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

¹⁷Similar 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 Online Supplement.

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

$$q_{it} = e^{\alpha u_{it}} l_{it}^\alpha m_{it}^\vartheta \prod_{j=1}^n q_{ij,t}^{(1-\alpha-\vartheta)w_{ij}}, \quad \text{for } i = 1, 2, \dots, n, \quad (63)$$

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 $i = 1, 2, \dots, n$. 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:

$$u(c_{1t}, c_{2t}, \dots, c_{nt}) = A \prod_{i=1}^n c_{it}^{1/n}, \quad A > 0. \quad (64)$$

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,

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

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 $\{P_{1t}, P_{2t}, \dots, P_{nt}, P_{m,t}, \text{Wage}_t\}$, the profit-maximization problem of sector i , for $i = 1, 2, \dots, n$, 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

$$q_{ij,t} = \frac{(1 - \alpha - \vartheta) w_{ij} P_{it} q_{it}}{P_{jt}}, \quad l_{it} = \frac{\alpha P_{it} q_{it}}{\text{Wage}_t}, \quad m_{it} = \frac{\vartheta P_{it} q_{it}}{P_{m,t}}. \quad (65)$$

Substituting (65) into (63) and after some simplifications yields

$$p_{it} = \rho \sum_{j=1}^n w_{ij} p_{jt} + \alpha \omega_t + \vartheta p_{m,t} - b_i - \alpha (\gamma_i f_t + v_{it}), \text{ for } i = 1, 2, \dots, n, \quad (66)$$

where $\rho = (1 - \alpha - \vartheta)$, $p_{it} = \log(P_{it})$, $\omega_t = \log(\text{Wage}_t)$, $p_{m,t} = \log(P_{m,t})$, 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 (66) 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 differences¹⁸

$$\Delta p_{it} = \rho \sum_{j=1}^n w_{ij} \Delta p_{jt} + \alpha \Delta \omega_t + \vartheta \Delta p_{m,t} - \alpha (\gamma_i \Delta f_t + \Delta v_{it}), \text{ for } t = 1, 2, \dots, T, \quad (67)$$

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

$$\overline{\Delta p}_i = \rho \sum_{j=1}^n w_{ij} \overline{\Delta p}_j + \alpha \overline{\Delta \omega} + \vartheta \overline{\Delta p}_m - \alpha (\gamma_i \overline{\Delta f} + \overline{\Delta v}_i), \quad (68)$$

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 the 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 $i = 1, 2, \dots, n$. Using this result in (68), we now have

$$\overline{\Delta p}_i = a + \rho \sum_{j=1}^n w_{ij} \overline{\Delta p}_j + \varepsilon_i, \quad (69)$$

where $a = \alpha \overline{\Delta \omega} + \vartheta \overline{\Delta p}_m - \alpha \gamma_0 \overline{\Delta f}$, and $\varepsilon_i = -\alpha (\overline{\Delta v}_i + \overline{\Delta f} \eta_i)$. The SAR model in the rate of price changes, (69), can now be estimated by 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 .¹⁹ The commodity-by-commodity direct requirements (**DR**) tables are derived from the commodity-

¹⁸This paper focuses on cross section SAR models. The estimation of the panel data model given by (66) is beyond the scope of the current paper.

¹⁹The words commodity and sector are used interchangeably to convey the same meaning throughout this paper.

by-commodity total requirements (\mathbf{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 \mathbf{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 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.²⁰

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 7 reports the estimates of δ for the top five most important sectors for these weights matrices. The results show that the highest degree of dominance, $\hat{\delta}_{(1)}$, 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 (38) and (48) if the errors are assumed to be homoskedastic and

²⁰Our choice of the 10% threshold for non-zero elements of the weights 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 Online Supplement. Using lower threshold values tend to yield higher estimates of ρ .

Table 7: 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}
$\hat{\delta}_{(1)}$	0.778	0.851	0.724
$\hat{\delta}_{(2)}$	0.759	0.796	0.651
$\hat{\delta}_{(3)}$	0.597	0.642	0.608
$\hat{\delta}_{(4)}$	0.550	0.422	0.592
$\hat{\delta}_{(5)}$	0.546	0.402	0.553
n	313 [301]	286 [114]	384 [364]
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 $\hat{\delta}_{(1)} > \hat{\delta}_{(2)} > \dots > \hat{\delta}_{(5)}$; 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.

heteroskedastic, respectively. The GMM estimates refer to the best GMM and are obtained in two steps: In the first step, we compute initial consistent estimate, $\tilde{\rho}$, by (27) using two equally weighted quadratic moments with $\mathbf{B}_\ell = \mathbf{W}^\ell - n^{-1}Tr(\mathbf{W}^\ell)\mathbf{I}_n$, $\ell = 1, 2$, if the errors are assumed to be homoskedastic (or $\mathbf{B}_\ell = \mathbf{W}^\ell - Diag(\mathbf{W}^\ell)$, $\ell = 1, 2$, if assuming heteroskedasticity).²¹ In the second step, we re-estimate the model using $\tilde{\mathbf{G}} - n^{-1}Tr(\tilde{\mathbf{G}})\mathbf{I}_n$ if assuming homoskedasticity (or $\tilde{\mathbf{G}} - Diag(\tilde{\mathbf{G}})$ if assuming heteroskedasticity), where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\rho})$ is evaluated at the first-step estimate.

Table 8 presents the estimation results of model (69). The top panel gives the estimates obtained under homoskedasticity and the bottom panel gives the estimates that allow for heteroskedasticity. As can be seen the BMM and GMM estimates are very close to one another, irrespective of whether the errors are assumed to be homoskedastic or not. Under homoskedastic errors, the estimated share of capital is around 0.40 for the first sub-sample and 0.29 for the second sub-sample. When we allow for heteroskedasticity we obtain a much larger estimate (0.54) for the first sub-sample, and a slightly smaller estimate (0.25) for the second sub-sample, which suggest heteroskedasticity is likely to be more serious in the first as compared to the second sub-sample. Despite the wide range of estimates of ρ obtained over

²¹Here 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 8: Estimation results of the cross-section model (69)

Year	Sub-sample		Sub-sample	
	1998–2006	2007–2015	BMM	GMM ^b
Assuming homoskedastic errors				
$\hat{\rho}$ [Share of capital]	0.397 [†] (0.106)	0.396 [†] (0.106)	0.287 [†] (0.072)	0.281 [†] (0.073)
$\hat{\sigma}_\eta^2$ [Error variance]	7.810	7.813	2.586	2.592
R^2	0.219	0.218	0.159	0.154
Assuming heteroskedastic errors				
$\hat{\rho}$ [Share of capital]	0.542 [†] (0.115)	0.542 [†] (0.114)	0.241 [†] (0.069)	0.246 [†] (0.069)
Weights matrix	$\tilde{\mathbf{W}}_{2002}(0.1)$		$\tilde{\mathbf{W}}_{2007}(0.1)$	
n [Number of sectors]	286		350	

Notes: All estimations include an intercept (not shown here). Standard errors are in parentheses. [†] indicates significance at 1% level. 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. R^2 is computed by (59) assuming homoskedasticity. The BMM estimates assuming homoskedastic errors are computed by (38), and computed by (48) if assuming heteroskedastic errors.

^b The GMM estimator refers to the best GMM estimator computed by a two-step procedure following (27) using the $\tilde{\mathbf{G}} - n^{-1} \text{Tr}(\tilde{\mathbf{G}}) \mathbf{I}_n$ if the errors are assumed to be homoskedastic, and $\tilde{\mathbf{G}} - \text{Diag}(\tilde{\mathbf{G}})$ if assuming heteroskedasticity, where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\rho})$ is evaluated at the first-step estimate, $\tilde{\rho}$.

the two sub-periods (0.28–0.54), they match reasonably well with the commonly documented values of share of capital in the literature. 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.

8 Concluding remarks

An important 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 economic applications. The current paper relaxes this assumption and allows the centrality of the connections to rise at the rate of δ with n , as compared to the value of $\delta = 0$ assumed in the literature. We also establish the asymptotic distribution of the GMM estimator due to

Lee (2007) and Lin and Lee (2010) under this more general setting, 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. Asymptotic properties of both estimators are investigated under homoskedastic as well as heteroskedastic errors, and shown to be consistent and normally distributed if the maximum absolute column sum of 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 homoskedastic Gaussian errors

$\delta \setminus n$	GMM ^b						BMM					
	Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)		
	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
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
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
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
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
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
$\rho_0 = 0.5$												
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
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
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
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
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
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
$\rho_0 = 0.75$												
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
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
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
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
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
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

Notes: The data generating process (DGP) is given by (57) and (58) with homoskedastic Gaussian errors. $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. 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 computed by (38).

^b The GMM estimator refers to the best GMM and is computed in two steps: In the first step, we obtain preliminary GMM estimates, $\tilde{\psi} = (\tilde{\rho}, \tilde{\alpha}, \tilde{\beta})'$, following (23), where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr(\mathbf{W}^2)\mathbf{I}_n$, and $\mathbf{A}_n = \mathbf{I}_n$. In the second step, we use $(\tilde{\mathbf{G}}\mathbf{x}\tilde{\alpha}, \tilde{\mathbf{G}}\mathbf{x}\tilde{\beta}, \mathbf{1}_n, \mathbf{x})$ and $\tilde{\mathbf{G}} - n^{-1}Tr(\tilde{\mathbf{G}})\mathbf{I}_n$, where $\tilde{\mathbf{G}} = \mathbf{W}(\mathbf{I}_n - \tilde{\rho}\mathbf{W})^{-1}$, in the linear and quadratic moments, respectively, and compute the optimal GMM estimates by (31).

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

$\delta \setminus n$	GMM ^b						BMM					
	Size($\times 100$)			Power($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	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
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
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
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
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
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
$\rho_0 = 0.5$												
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
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
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
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
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
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
$\rho_0 = 0.75$												
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
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
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
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
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
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

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 homoskedastic Gaussian errors

$\delta \setminus n$	GMM ^b						BMM					
	Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)		
	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
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
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
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
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
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
$\rho_0 = 0.5$												
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
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
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
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
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
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
$\rho_0 = 0.75$												
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
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
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
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
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
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

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 homoskedastic Gaussian errors

$\delta \setminus n$	GMM						BMM					
	Size($\times 100$)			Power($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	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
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
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
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
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
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
$\rho_0 = 0.5$												
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
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
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
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
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
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
$\rho_0 = 0.75$												
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
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
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
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
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
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

Notes: The true parameter value is $\beta_0 = 1$ and power is calculated at 0.8. See also the notes to Table 1a.

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

$\delta \setminus n$	GMM ^b						BMM					
	Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)		
	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
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
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
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
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
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
$\rho_0 = 0.5$												
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
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
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
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
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
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
$\rho_0 = 0.75$												
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
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
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
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
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
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

Notes: The DGP is given by (57) and (58) with homoskedastic non-Gaussian errors. See also the notes to Table 1a.

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

$\delta \setminus n$	GMM ^b						BMM					
	Size($\times 100$)			Power($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$	7.45	4.30	5.30	5.20	9.85	16.30	24.25	43.15	6.60	3.95	5.20	7.80
0.00	7.45	4.30	5.35	5.25	9.65	16.15	24.20	43.20	6.90	4.00	5.25	7.65
0.25	7.45	4.30	5.20	5.15	9.85	16.60	24.50	42.90	6.95	4.35	5.25	7.45
0.50	7.40	4.45	5.20	5.15	11.70	18.45	24.95	42.75	8.15	5.20	6.15	8.35
0.75	9.05	5.45	6.15	6.65	21.75	23.25	26.85	41.60	13.30	9.85	10.35	12.95
0.95	18.70	12.55	12.00	10.00	25.45	26.20	29.80	41.35	16.85	13.75	15.35	16.45
1.00	24.70	17.55	17.45	16.05	25.45	26.20	29.80	41.35	16.85	13.75	15.35	16.45
$\rho_0 = 0.5$	7.70	4.40	5.25	5.65	14.00	25.80	39.40	68.05	6.60	4.35	4.90	5.50
0.00	7.70	4.50	5.25	5.55	14.00	25.95	39.25	68.15	6.85	4.35	4.90	5.50
0.25	7.95	4.75	5.20	5.80	14.40	25.95	38.95	67.90	6.75	4.45	5.30	5.50
0.50	8.00	4.75	5.20	5.80	16.95	26.65	38.20	65.85	7.95	5.60	6.45	6.10
0.75	11.00	5.75	6.95	6.40	28.65	32.15	37.75	57.60	14.55	11.85	12.25	10.45
0.95	25.80	16.45	14.50	12.15	33.05	34.80	39.80	55.95	20.35	18.45	19.60	19.05
1.00	32.90	25.30	23.55	22.15	32.90	38.70	45.85	49.25	59.60	75.85	31.40	32.10
$\rho_0 = 0.75$	9.55	4.75	4.95	5.75	25.80	56.40	77.25	96.55	6.65	4.30	4.95	5.85
0.00	9.85	4.75	4.85	5.50	25.85	56.50	77.45	96.45	6.60	4.35	4.75	5.55
0.25	10.35	4.95	5.25	5.50	26.05	56.50	77.30	96.60	6.95	4.65	4.80	5.65
0.50	14.85	7.00	6.80	29.30	54.80	73.50	94.45	8.55	6.40	7.15	6.75	18.85
0.75	26.00	23.05	17.20	42.00	51.15	62.70	86.15	17.75	14.90	15.50	13.25	18.55
0.95	40.75	43.00	42.95	38.70	45.85	49.25	59.60	75.85	27.40	30.75	31.40	23.35
1.00	52.40	43.00	42.95	38.70	45.85	49.25	59.60	75.85	27.40	30.75	31.40	23.35

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 homoskedastic non-Gaussian errors

$\delta \setminus n$	GMM ^b						BMM					
	Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)		
	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
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
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
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
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
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
$\rho_0 = 0.5$												
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
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
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
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
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
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
$\rho_0 = 0.75$												
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
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
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
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
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
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

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 homoskedastic non-Gaussian errors

$\delta \setminus n$	GMM ^b						BMM					
	Size($\times 100$)			Power($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	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
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
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
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
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
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
$\rho_0 = 0.5$												
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
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
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
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
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
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
$\rho_0 = 0.75$												
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
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
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
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
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
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

Notes: The true parameter value is $\beta_0 = 1$ and power is calculated at 0.8. See also the notes to Table 3a.

Table 5a: Bias and RMSE of the GMM and BMM estimators of ρ for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	GMM ^b						BMM								
	Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)					
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000			
$\rho_0 = 0.2$	-6.10	-1.88	-1.22	-0.62	19.92	9.89	7.64	5.30	-8.06	-2.33	-1.45	-0.73			
	0.25	-6.05	-1.86	-1.21	-0.62	19.87	9.87	7.64	5.30	-7.98	-2.31	-1.43	-0.73		
	0.50	-6.22	-1.91	-1.24	-0.63	19.88	9.97	7.66	5.30	-8.16	-2.35	-1.46	-0.73		
	0.75	-6.24	-1.90	-1.31	-0.63	21.20	10.41	7.97	5.49	-8.62	-2.46	-1.56	-0.75		
	0.95	-5.97	-1.93	-1.34	-0.65	27.05	12.96	9.58	6.50	-10.54	-3.04	-1.86	-0.91		
	1.00	-6.64	-2.06	-1.42	-0.75	31.29	15.49	11.65	7.91	-11.81	-3.39	-2.10	-1.07		
	$\rho_0 = 0.5$	-5.32	-1.61	-1.02	-0.52	15.92	7.43	5.68	3.92	-7.28	-2.06	-1.26	-0.63		
		0.25	-5.30	-1.61	-1.02	-0.52	15.91	7.42	5.68	3.92	-7.24	-2.05	-1.25	-0.63	
		0.50	-5.40	-1.64	-1.04	-0.53	15.95	7.47	5.70	3.93	-7.35	-2.08	-1.27	-0.64	
		0.75	-5.52	-1.69	-1.15	-0.55	17.32	7.95	6.03	4.10	-7.97	-2.25	-1.41	-0.68	
		0.95	-5.44	-1.85	-1.24	-0.61	24.11	10.83	7.77	5.15	-10.90	-3.18	-1.90	-0.92	
		1.00	-6.41	-1.87	-1.30	-0.71	28.99	14.31	10.62	7.05	-13.09	-3.90	-2.38	-1.19	
		$\rho_0 = 0.75$	-3.86	-1.17	-0.73	-0.37	10.73	4.61	3.46	2.37	-5.51	-1.52	-0.91	-0.46	
			0.25	-3.85	-1.17	-0.73	-0.37	10.75	4.61	3.46	2.37	-5.50	-1.52	-0.91	-0.46
			0.50	-3.88	-1.18	-0.74	-0.38	10.79	4.61	3.47	2.38	-5.53	-1.52	-0.92	-0.46
			0.75	-3.95	-1.27	-0.85	-0.40	12.30	5.02	3.75	2.51	-6.22	-1.72	-1.06	-0.50
			0.95	-5.14	-1.37	-0.87	-0.46	21.12	8.29	5.68	3.55	-10.57	-2.99	-1.73	-0.82
			1.00	-8.48	-2.15	-1.20	-0.59	28.88	14.05	10.29	6.89	-14.50	-4.66	-2.84	-1.44

Notes: The data generating process (DGP) is given by (57) and (58) with heteroskedastic non-Gaussian errors. $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. 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 computed by (48).

^b The GMM estimator refers to the best GMM and is computed in two steps: In the first step, we obtain preliminary GMM estimates, $\tilde{\psi} = (\tilde{\rho}, \tilde{\alpha}, \tilde{\beta})'$, following (23), where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{Wx}, \mathbf{W}^2\mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - \text{Diag}(\mathbf{W}^2)$, and $\mathbf{A}_n = \mathbf{I}_n$. In the second step, we use $(\tilde{\mathbf{G}}_{\mathbf{x}\tilde{\alpha}}, \tilde{\mathbf{G}}_{\mathbf{x}\tilde{\beta}}, \mathbf{1}_n, \mathbf{x})'$ and $\tilde{\mathbf{G}} - \text{Diag}(\tilde{\mathbf{G}})$, where $\tilde{\mathbf{G}} = \mathbf{W}(\mathbf{I}_n - \tilde{\rho}\mathbf{W})^{-1}$, in the linear and quadratic moments, respectively, and compute the optimal GMM estimates by (31).

Table 5b: Size and Power of the GMM and BMM estimators of ρ for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	GMM ^b						BMM					
	Size($\times 100$)			Power($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	10.75	5.90	6.20	5.35	11.45	15.90	24.20	43.35	9.55	5.55	5.75	5.10
0.25	10.50	6.10	6.20	5.40	11.30	15.75	24.10	43.50	9.50	5.65	5.70	5.10
0.50	9.95	6.20	5.90	6.00	11.15	16.10	24.20	42.85	9.30	5.95	5.70	5.65
0.75	11.05	6.85	6.80	5.95	12.20	17.15	24.20	43.00	9.90	6.20	6.20	5.75
0.95	17.00	11.15	10.05	8.25	17.80	19.40	22.90	37.95	12.80	8.85	8.60	7.60
1.00	20.55	14.40	14.15	11.85	21.60	19.80	23.20	33.95	15.45	12.05	12.25	10.65
$\rho_0 = 0.5$												
0.00	10.40	6.10	6.15	5.95	15.10	25.80	40.80	68.65	9.30	5.70	5.60	5.55
0.25	10.50	6.10	6.20	5.80	15.30	25.50	41.00	68.55	9.10	5.60	5.55	5.55
0.50	11.05	6.05	6.05	5.75	14.95	26.65	40.20	68.80	9.50	6.00	6.00	5.75
0.75	12.30	6.90	7.30	6.15	16.55	26.10	39.05	66.15	10.10	6.80	6.85	6.35
0.95	21.85	13.85	11.90	9.30	24.45	26.75	35.45	56.15	14.35	10.10	9.30	8.20
1.00	27.85	20.00	18.95	15.85	27.40	27.45	31.50	45.45	18.75	15.75	15.75	13.50
$\rho_0 = 0.75$												
0.00	13.35	5.90	6.20	5.60	27.20	58.30	79.25	97.05	10.15	5.45	5.75	5.40
0.25	13.40	6.10	6.25	5.55	27.40	58.35	79.15	97.15	10.40	5.50	5.70	5.50
0.50	13.30	5.85	5.95	5.60	27.60	57.90	78.95	97.05	10.15	5.60	5.35	5.60
0.75	15.35	6.60	7.50	6.30	29.50	55.75	75.05	96.15	10.75	6.15	6.80	6.25
0.95	33.75	20.00	16.10	12.80	36.55	45.80	60.10	85.85	17.05	12.85	12.35	10.45
1.00	44.05	33.95	32.55	31.65	39.75	40.10	46.40	63.45	25.45	22.90	24.75	24.35

Notes: The power is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value. See also the notes to Table 5a.

Table 6a: Bias and RMSE of the GMM and BMM estimators of β for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	GMM ^b						BMM					
	Bias($\times 100$)			RMSE($\times 100$)			Bias($\times 100$)			RMSE($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	0.08	-0.31	0.12	-0.14	39.75	21.10	15.95	11.35	4.57	1.42	1.20	0.49
0.25	0.04	-0.35	0.10	-0.14	39.73	21.10	15.94	11.34	4.53	1.39	1.19	0.49
0.50	-0.08	-0.48	0.07	-0.17	39.83	21.20	15.98	11.37	4.58	1.34	1.16	0.46
0.75	-0.55	-0.37	0.09	-0.10	41.42	21.82	16.18	11.50	4.27	1.52	1.19	0.54
0.95	-0.40	-0.45	0.08	-0.19	46.08	25.03	16.88	12.10	3.92	1.39	1.11	0.45
1.00	-0.37	-0.31	0.09	-0.20	49.58	27.41	17.79	13.19	3.31	1.12	0.92	0.30
$\rho_0 = 0.5$												
0.00	0.63	-0.20	0.24	-0.11	47.05	24.76	18.68	13.27	6.11	1.91	1.54	0.64
0.25	0.61	-0.23	0.23	-0.11	47.03	24.76	18.66	13.27	6.08	1.88	1.53	0.64
0.50	0.47	-0.39	0.19	-0.14	47.25	24.97	18.74	13.32	6.14	1.83	1.49	0.61
0.75	0.06	-0.23	0.23	-0.05	50.30	26.20	19.10	13.58	5.92	2.10	1.58	0.72
0.95	0.29	-0.33	0.21	-0.19	58.05	31.50	20.25	14.66	5.54	2.05	1.51	0.64
1.00	-0.16	-0.36	0.13	-0.25	63.76	35.62	21.80	16.47	4.70	1.64	1.26	0.44
$\rho_0 = 0.75$												
0.00	1.56	0.10	0.45	-0.01	49.25	25.63	19.28	13.71	7.70	2.43	1.86	0.80
0.25	1.57	0.08	0.44	-0.01	49.22	25.63	19.26	13.69	7.68	2.40	1.85	0.80
0.50	1.47	-0.09	0.40	-0.04	49.37	25.90	19.33	13.77	7.77	2.35	1.82	0.77
0.75	1.25	0.09	0.46	0.05	54.05	27.80	19.92	14.22	7.75	2.72	1.96	0.90
0.95	1.74	-0.08	0.38	-0.15	64.67	34.92	21.81	15.97	7.40	2.88	1.99	0.88
1.00	0.74	-0.49	0.03	-0.36	71.34	40.14	23.79	18.39	6.18	2.32	1.69	0.65

Notes: The true parameter value is $\beta_0 = 1$. See also the notes to Table 5a.

Table 6b: Size and power of the GMM and BMM estimators of β for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	GMM ^b						BMM					
	Size($\times 100$)			Power($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$	9.80	6.65	6.35	6.20	10.50	16.35	23.85	44.15	7.80	5.85	5.80	5.95
0.00	9.80	6.65	6.35	6.20	10.60	16.20	23.80	44.30	7.75	5.90	5.85	5.95
0.25	9.75	6.70	6.20	6.20	10.50	15.40	24.05	44.15	7.80	6.00	5.65	6.05
0.50	10.05	6.70	6.40	6.20	10.45	15.75	24.05	43.90	8.55	5.80	5.25	6.10
0.75	10.75	6.90	5.95	6.25	10.45	15.75	24.05	43.90	8.40	6.30	6.10	10.35
0.95	10.75	7.10	6.90	6.05	10.30	13.65	23.60	39.60	8.75	6.45	5.45	5.85
1.00	10.10	6.85	6.25	6.35	9.05	11.40	20.70	34.20	8.70	6.30	8.70	9.80
$\rho_0 = 0.5$	9.80	6.90	6.35	6.45	10.30	13.20	19.05	33.95	7.75	5.95	5.90	6.15
0.00	9.80	6.95	6.30	6.40	10.20	13.20	19.15	33.90	7.80	6.10	6.00	6.15
0.25	9.60	6.95	6.55	6.40	9.90	12.80	18.95	33.85	8.00	6.10	5.90	6.10
0.50	10.00	7.10	6.55	6.40	10.00	12.70	19.00	33.15	8.80	5.95	5.50	6.30
0.75	10.65	7.05	5.95	6.25	9.70	11.00	17.50	29.05	8.70	6.45	6.10	9.45
0.95	11.15	7.05	7.05	6.20	9.75	8.75	14.50	23.80	8.80	6.50	5.70	6.35
1.00	10.10	7.10	6.50	6.35	8.65	7.90	12.50	19.85	8.80	6.70	5.75	8.15
$\rho_0 = 0.75$	9.80	6.90	6.50	6.50	10.20	12.55	18.30	32.30	8.10	6.00	5.90	6.10
0.00	9.80	6.90	6.45	6.35	10.30	12.60	18.45	32.35	8.00	5.85	5.85	6.05
0.25	9.75	6.90	6.35	6.35	10.15	12.45	18.50	31.55	8.00	6.20	5.80	6.05
0.50	9.95	7.25	6.35	6.15	6.40	10.05	11.90	18.10	30.80	9.05	5.85	5.70
0.75	11.10	6.95	6.15	6.40	9.60	10.30	15.85	25.70	9.10	6.80	6.40	6.05
0.95	10.95	7.35	7.00	6.30	8.65	7.90	12.50	19.85	8.80	6.70	5.75	6.45
1.00	9.80	7.40	6.70	6.60	8.65	7.90	12.50	19.85	8.80	6.70	5.75	8.50

Notes: The true parameter value is $\beta_0 = 1$ and power is calculated at 0.8. See also the notes to Table 5a.

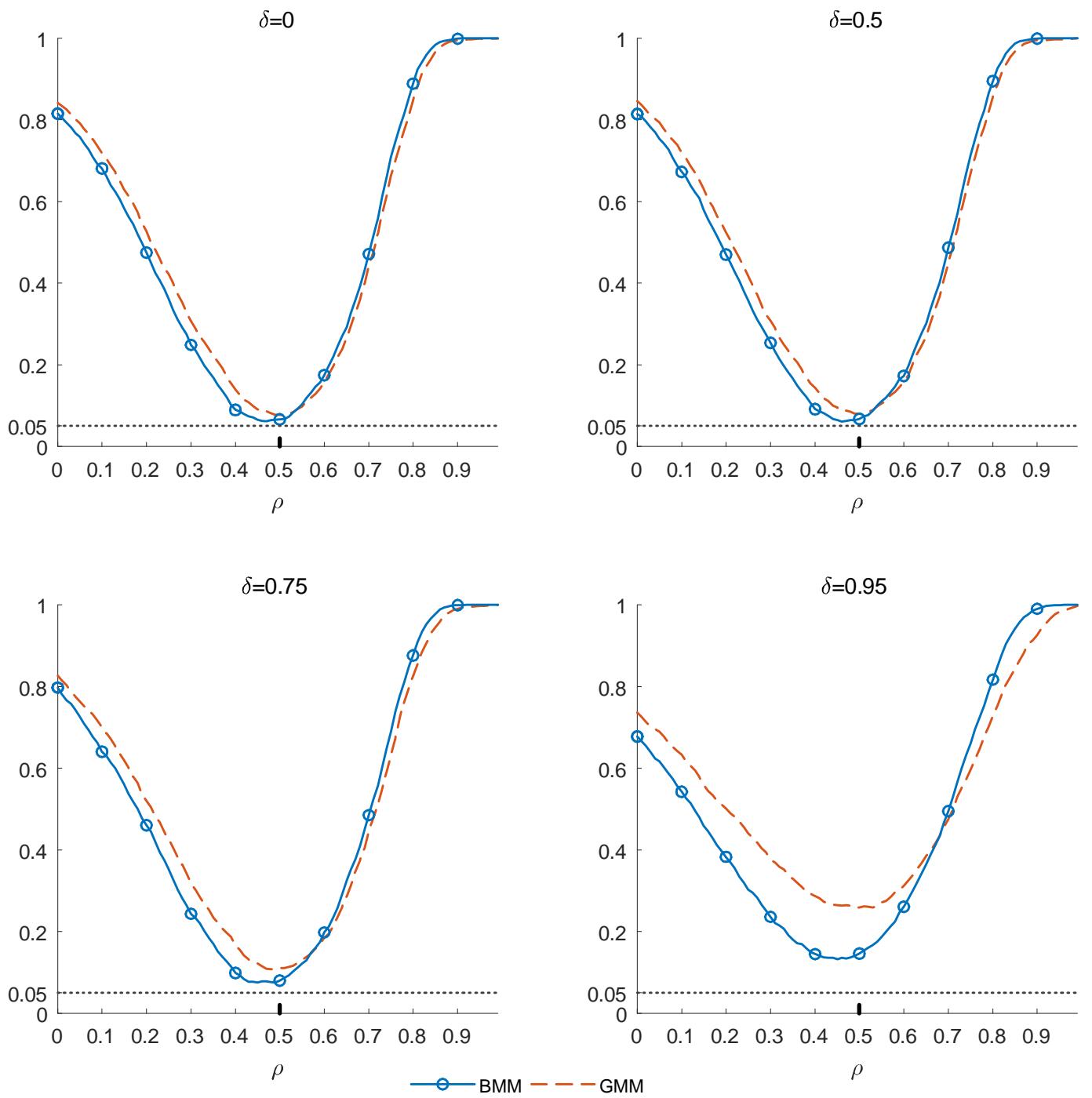


Figure 1a: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and homoskedastic non-Gaussian errors for different values of δ

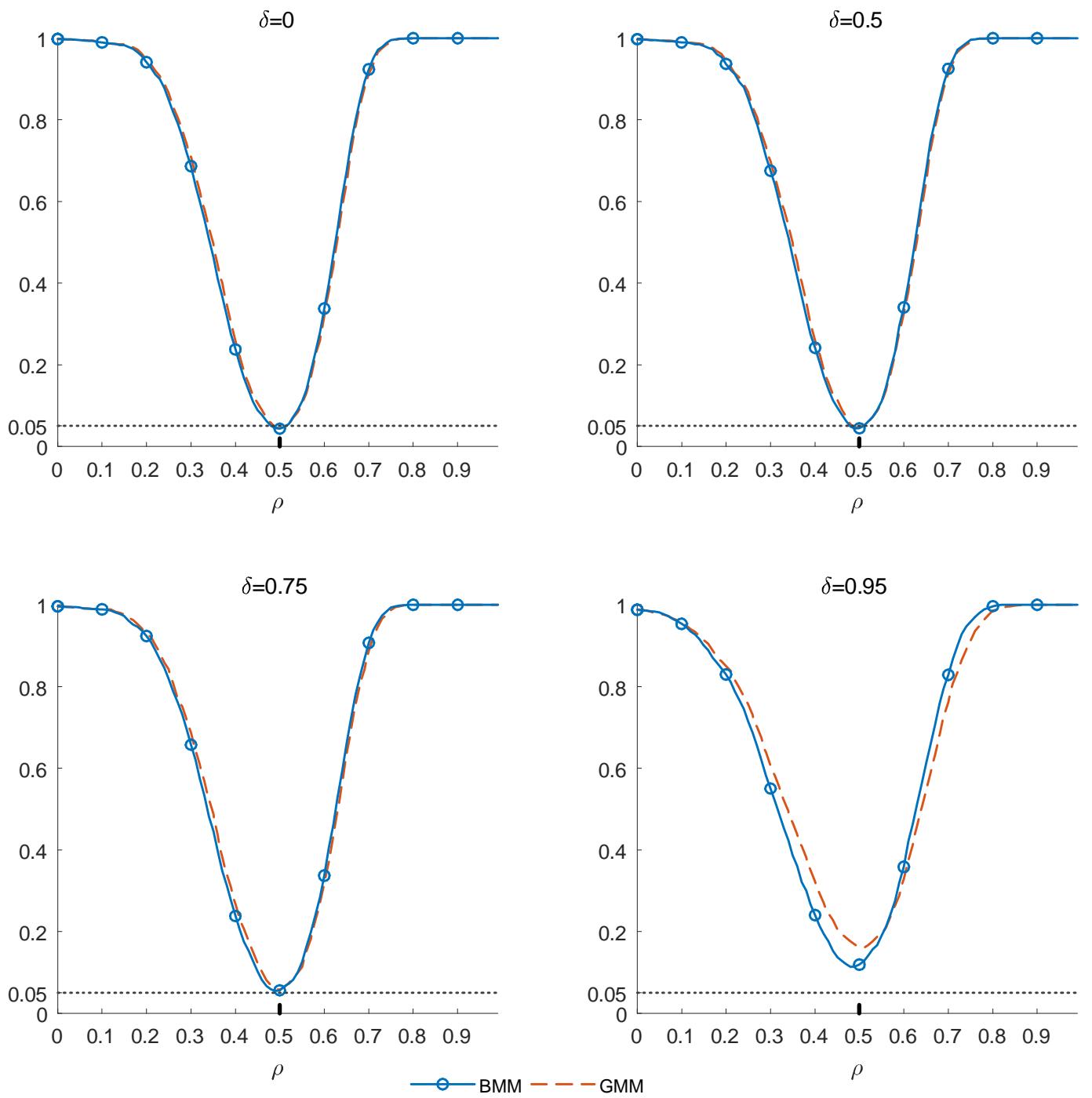


Figure 1b: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and homoskedastic non-Gaussian errors for different values of δ

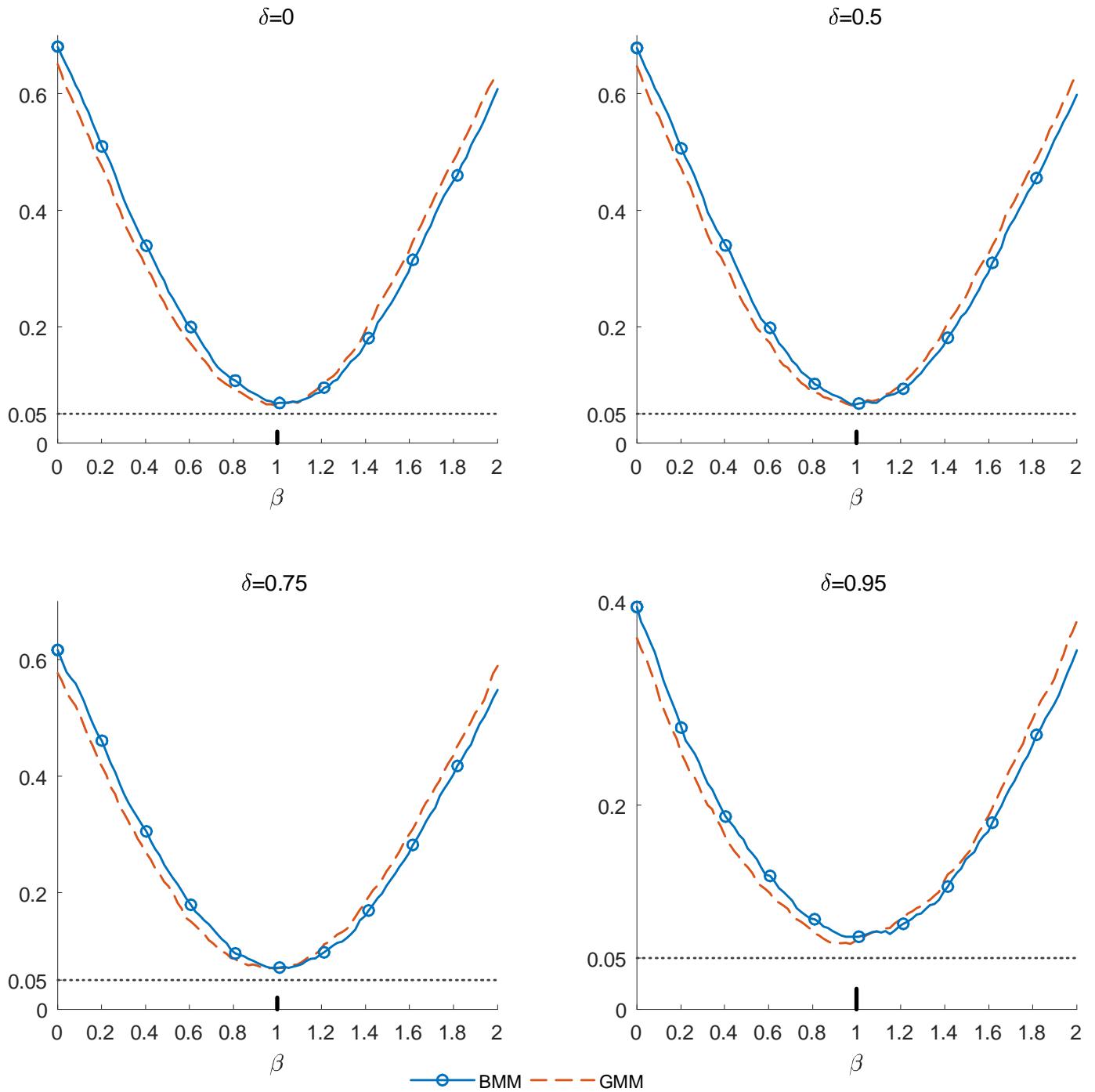


Figure 2a: Empirical power functions for β in the case of $\beta_0 = 1$, $n = 100$, and homoskedastic non-Gaussian errors for different values of δ

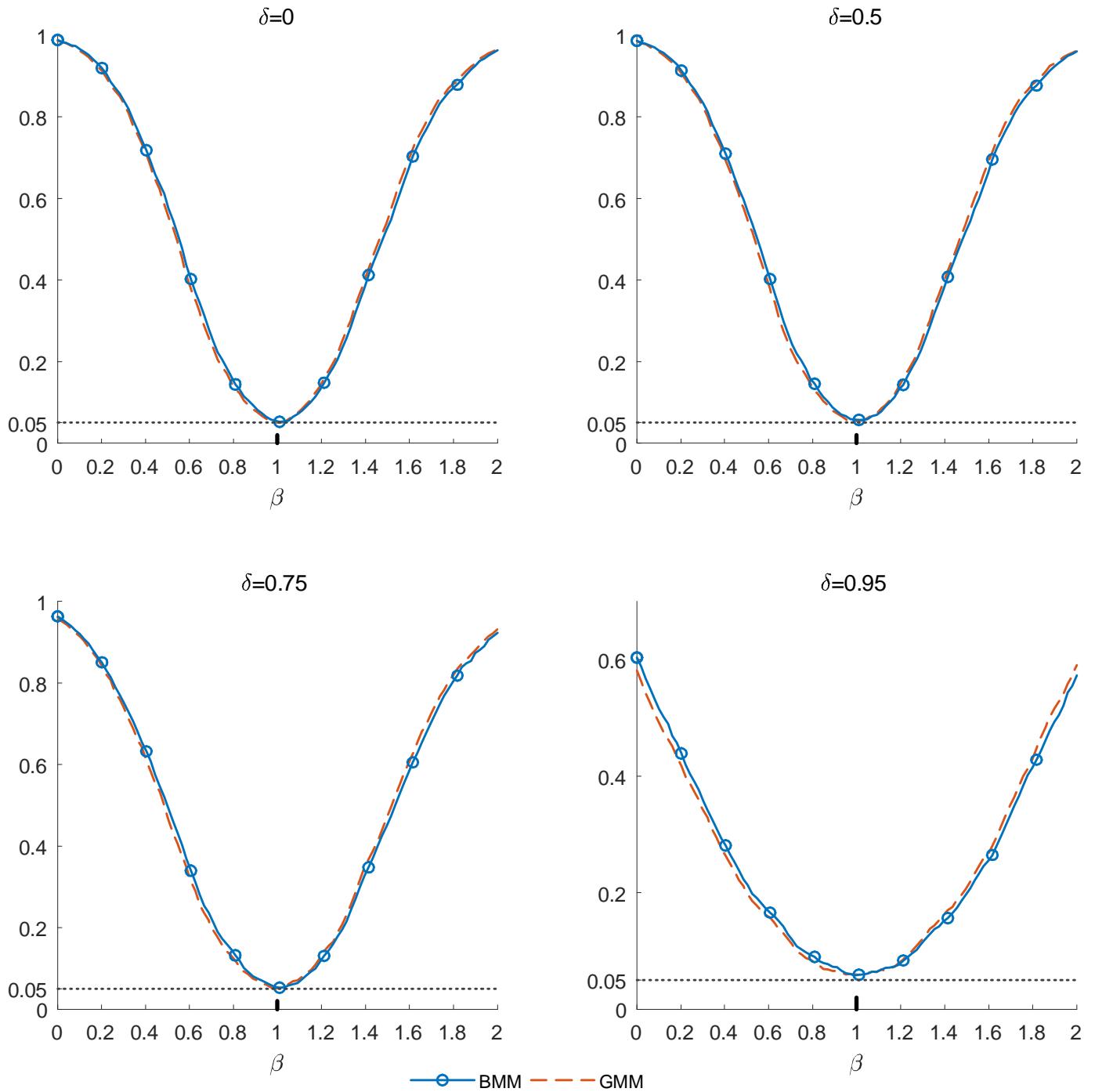


Figure 2b: Empirical power functions for β in the case of $\beta_0 = 1$, $n = 300$, and homoskedastic non-Gaussian errors for different values of δ

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.
- 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>.
- Chudik, A., M. H. Pesaran, and E. Tosetti (2011). Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal* 14, C45–C90.
- 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 (2019). Detection of units with pervasive effects in large panel data models. USC-INET Research Paper No. 19-09.
- Kapoор, 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 Supplement 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

January 15, 2020

This online supplement is organized into two sections. Section S1 provides statements and proofs of necessary lemmas, and gives proofs of the theorems and propositions in the paper. Section S2 presents additional Monte Carlo and empirical results.

S1 Theory supplement

This theory supplement begins by providing statements and proofs of necessary lemmas used in establishing the main theoretical results of the paper, and then provide proofs of the theorems and propositions set out in Sections 3–5 of the paper. Throughout this supplement, Assumptions 1–7 refer to the Assumptions made in the paper.

S1.1 Lemmas

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

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

Lemma S.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 S.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

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

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

Lemma S.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 $\|\mathbf{B}\|_1 = O(n^\delta)$, where $0 \leq \delta \leq 1$. Then

- (i) $\text{Tr}(\mathbf{A}'\mathbf{B}\mathbf{B}'\mathbf{A}) = O(n^{\delta+1})$,
- (ii) $\text{Tr}[(\mathbf{A}'\mathbf{B})^2] = O(n^{\delta+1})$,
- (iii) $\text{Tr}(\mathbf{AB}'\mathbf{C}) = O(n^{\delta+1})$, 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

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

which establishes the claim.

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

(iii) Note that

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

and the result follows. ■

Lemma S.5 Suppose that $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ is a vector of random variables, where ε_i , for $i = 1, 2, \dots, n$, are independently distributed over i with zero means and variances, σ_i^2 , such that $\inf_i (\sigma_i^2) > c > 0$ and $\sup_i (\sigma_i^2) < K$. Let $\boldsymbol{\Sigma} = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ and $\gamma_{2i} = \mu_{4i} - 3\sigma_i^4$,

where $\mu_{4i} = E(\varepsilon_i^4)$ and assume that $\sup_i |\mu_{4i}| < K$. Then for any $n \times n$ constant matrix (need not be symmetric) $\mathbf{A} = (a_{ij})$, we have

- (i) $E(\varepsilon' \mathbf{A} \varepsilon) = \sum_{i=1}^n a_{ii} \sigma_i^2 = Tr(\Sigma \mathbf{A})$,
- (ii) $E(\varepsilon' \mathbf{A} \varepsilon)^2 = \sum_{i=1}^n a_{ii}^2 \gamma_{2i} + Tr^2(\Sigma \mathbf{A}) + Tr[\Sigma \mathbf{A} (\Sigma \mathbf{A}' + \Sigma \mathbf{A})]$,
- (iii) $Var(\varepsilon' \mathbf{A} \varepsilon) = \sum_{i=1}^n a_{ii}^2 \gamma_{2i} + Tr[\Sigma \mathbf{A} (\Sigma \mathbf{A}' + \Sigma \mathbf{A})] \leq K Tr(\mathbf{A} \mathbf{A}')$.

Proof. See Lemma A.2 of Lin and Lee (2010). The inequality in (iii) follows from $\sup_i |\gamma_{2i}| < K$, $\sum_{i=1}^n a_{ii}^2 \leq Tr(\mathbf{A} \mathbf{A}')$, $Tr(\Sigma \mathbf{A} \Sigma \mathbf{A}') \leq (\sup_i \sigma_i^2)^2 Tr(\mathbf{A} \mathbf{A}')$ and $Tr(\Sigma \mathbf{A} \Sigma \mathbf{A}) \leq Tr(\Sigma \mathbf{A} \mathbf{A}' \Sigma) \leq (\sup_i \sigma_i^2)^2 Tr(\mathbf{A} \mathbf{A}')$ by Schur's inequality and $\sup_i (\sigma_i^2) < K$. ■

Remark S.1 In the special case where ε_i is homoskedastic with $\sigma_i^2 = \sigma^2$ and $\gamma_{2i} = \gamma_2$, for all i , then the results in Lemma S.5 reduce to

$$\begin{aligned} E(\varepsilon' \mathbf{A} \varepsilon) &= \sigma^2 Tr(\mathbf{A}), \\ E(\varepsilon' \mathbf{A} \varepsilon)^2 &= \gamma_2 \sum_{i=1}^n a_{ii}^2 + \sigma^4 [Tr^2(\mathbf{A}) + Tr(\mathbf{A} \mathbf{A}') + Tr(\mathbf{A}^2)], \\ Var(\varepsilon' \mathbf{A} \varepsilon) &= \gamma_2 \sum_{i=1}^n a_{ii}^2 + \sigma^4 [Tr(\mathbf{A} \mathbf{A}') + Tr(\mathbf{A}^2)] \leq K Tr(\mathbf{A} \mathbf{A}'). \end{aligned}$$

Lemma S.6 Suppose that ρ is a fixed constant and $\mathbf{W} = (w_{ij})$ is an $n \times n$ constant matrix such that (a) $w_{ij} \geq 0$ for all i and j , (b) $\|\mathbf{W}\|_\infty < K$, and $|\rho| \|\mathbf{W}\|_\infty < 1$, and (c) the column sums of \mathbf{W} , denoted by $d_{jn} = \sum_{i=1}^n w_{ij}$, $j = 1, 2, \dots, n$, are non-zero and follow the specification: $d_{jn} = \kappa_j n^{\delta_j}$, where κ_j is a strictly positive random variable defined on $0 < \underline{\kappa} \leq \kappa_j \leq \bar{\kappa} < K$, with $\underline{\kappa}$ and $\bar{\kappa}$ being fixed constants, δ_j is a fixed constant in the range $0 \leq \delta_j \leq 1$, with $\delta_j > 0$ for $j = 1, 2, \dots, m$, and $\delta_j = 0$ for $j = m+1, m+2, \dots, n$, where m is a fixed number. (d) $|\rho| \|\mathbf{W}_{22}\|_1 < 1$, where \mathbf{W}_{22} is the $(n-m)$ -dimensional square submatrix of \mathbf{W} that represents the connections among the non-dominant units.. Let $\mathbf{S} = \mathbf{S}(\rho) = \mathbf{I}_n - \rho \mathbf{W}$, $\mathbf{G} = \mathbf{G}(\rho) = \mathbf{W} \mathbf{S}^{-1}(\rho) = \mathbf{W} (\mathbf{I}_n - \rho \mathbf{W})^{-1}$, and $\delta = \max_{j=1,2,\dots,n} (\delta_j)$. 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) Since $\|\rho \mathbf{W}\|_\infty < 1$ by assumption, we have $\mathbf{S}^{-1} = \sum_{k=0}^{\infty} (\rho \mathbf{W})^k$ (see, for example, Horn and Johnson, 2012, Corollary 5.6.16). It follows that

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

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

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

Applying the formula for the inverse of a partitioned matrix gives

$$\mathbf{S}^{-1} = \begin{pmatrix} \Phi_1^{-1} & \rho \Phi_1^{-1} \mathbf{W}_{12} \mathbf{S}_{22}^{-1} \\ \rho \mathbf{S}_{22}^{-1} \mathbf{W}_{21} \Phi_1^{-1} & \mathbf{S}_{22}^{-1} + \rho^2 \mathbf{S}_{22}^{-1} \mathbf{W}_{21} \Phi_1^{-1} \mathbf{W}_{12} \mathbf{S}_{22}^{-1} \end{pmatrix},$$

where $\Phi_1 = \mathbf{I}_m - \rho \mathbf{W}_{11} - \rho^2 \mathbf{W}_{12} \mathbf{S}_{22}^{-1} \mathbf{W}_{21}$, and $\mathbf{S}_{22} = \mathbf{I}_{n-m} - \rho \mathbf{W}_{22}$. Since by assumption $|\rho| \|\mathbf{W}_{22}\|_\infty < 1$ and $|\rho| \|\mathbf{W}_{22}\|_1 < 1$, then $\|\mathbf{S}_{22}^{-1}\|_\infty < K$ and $\|\mathbf{S}_{22}^{-1}\|_1 < K$. Also, since m is fixed and does not rise with n , it is sufficient to examine $\|\mathbf{S}_{22}^{-1} \mathbf{W}_{21} \Phi_1^{-1}\|_1$ and $\|\mathbf{S}_{22}^{-1} + \rho^2 \mathbf{S}_{22}^{-1} \mathbf{W}_{21} \Phi_1^{-1} \mathbf{W}_{12} \mathbf{S}_{22}^{-1}\|_1$. Let $\mathbf{w}_{\cdot j, 21}$ denote the j^{th} column of \mathbf{W}_{21} . By Lemma S.3, $\|\mathbf{S}_{22}^{-1} \mathbf{w}_{\cdot j, 21}\|_1 = O(n^{\delta_j})$, for $j = 1, 2, \dots, m$, which yields $\|\mathbf{S}_{22}^{-1} \mathbf{W}_{21}\|_1 = O(n^\delta)$, where $\delta = \max_j(\delta_j)$. Therefore,

$$\|\mathbf{S}_{22}^{-1} \mathbf{W}_{21} \Phi_1^{-1}\|_1 \leq \|\mathbf{S}_{22}^{-1} \mathbf{W}_{21}\|_1 \|\Phi_1^{-1}\|_1 = O(n^\delta), \quad (\text{S.1})$$

noting that the norm of the $m \times m$ matrix Φ_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

$$\|\mathbf{S}_{22}^{-1} \mathbf{W}_{21} \Phi_1^{-1} \mathbf{W}_{12} \mathbf{S}_{22}^{-1}\|_1 \leq \|\mathbf{S}_{22}^{-1} \mathbf{W}_{21}\|_1 \|\Phi_1^{-1}\|_1 \|\mathbf{W}_{12} \mathbf{S}_{22}^{-1}\|_1 = O(n^\delta). \quad (\text{S.2})$$

Combining (S.1) and (S.2) leads to $\|\mathbf{S}^{-1}\|_1 = O(n^\delta)$.

(ii) Since $\|\mathbf{W}\|_\infty < K$ by assumption and $\|\mathbf{S}^{-1}\|_\infty < K$ by Lemma S.6(i), we immediately obtain $\|\mathbf{G}\|_\infty < K$ by Lemma S.2. Turning to $\|\mathbf{G}\|_1$. Let $\mathbf{G} = (g_{ij})_{n \times n}$ and $\mathbf{S}^{-1} = (s^{ij})_{n \times n}$. For the j^{th} column of \mathbf{G} , $j = 1, 2, \dots, n$, we have

$$\begin{aligned} \sum_{i=1}^n |g_{ij}| &= \sum_{i=1}^n \left| \sum_{l=1}^n w_{il} s^{ij} \right| = \sum_{i=1}^n \left| \sum_{l=1}^m w_{il} s^{ij} \right| + \sum_{i=1}^n \left| \sum_{l=m+1}^n w_{il} s^{ij} \right| \\ &\leq \left(\sup_{1 \leq l \leq m} \sum_{i=1}^n |w_{il}| \right) \sum_{l=1}^m |s^{ij}| + \left(\sup_{m+1 \leq l \leq n} \sum_{i=1}^n |w_{il}| \right) \sum_{l=m+1}^n |s^{ij}| \\ &\leq K n^\delta m + K n^\delta. \end{aligned}$$

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

Lemma S.7 Suppose that $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ is a vector of random variables, where ε_i , for $i = 1, 2, \dots, n$, are independently distributed over i with zero means and variances, σ_i^2 , such that $\inf_i(\sigma_i^2) > c > 0$ and $\sup_i(\sigma_i^2) < K$. Let $\boldsymbol{\Sigma} = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ and $\gamma_{2i} = \mu_{4i} - 3\sigma_i^4$, where $\mu_{4i} = E(\varepsilon_i^4)$ and assume that $\sup_i |\mu_{4i}| < K$. $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed coefficients such that $\|\boldsymbol{\beta}\|_1 < K$. $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ is an $n \times k$ matrix of random variables such that $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})'$ are distributed independently of the errors, ε_j , for all i and j , and $\sup_{i,s} E|x_{is}|^2 < K$. \mathbf{G} is an $n \times n$ matrix of fixed constants such that $\|\mathbf{G}\|_\infty < K$, and $\|\mathbf{G}\|_1 = O(n^\delta)$, where δ is a fixed constant in the range $0 \leq \delta < 1$. $\mathbf{C} = (c_{ij}) = (\mathbf{B} + \mathbf{B}')/2$, where \mathbf{B} is an $n \times n$ matrix of fixed constants such that $\|\mathbf{B}\|_\infty < K$ and $\|\mathbf{B}\|_1 = O(n^\delta)$. Then

- (i) $n^{-1} \boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon} = \text{Tr}(n^{-1} \mathbf{C} \boldsymbol{\Sigma}) + o_p(1)$,
- (ii) $n^{-1} \boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \boldsymbol{\varepsilon} = \text{Tr}(n^{-1} \mathbf{G}' \mathbf{C} \boldsymbol{\Sigma}) + o_p(1)$,
- (iii) $n^{-1} \boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \boldsymbol{\varepsilon} = \text{Tr}(n^{-1} \mathbf{G}' \mathbf{C} \mathbf{G} \boldsymbol{\Sigma}) + o_p(1)$,
- (iv) $n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{C} \boldsymbol{\varepsilon} = o_p(1)$,

- (v) $n^{-1}\beta' \mathbf{X}' \mathbf{G}' \mathbf{C} \boldsymbol{\varepsilon} = o_p(1)$,
- (vi) $n^{-1}\beta' \mathbf{X}' \mathbf{C} \mathbf{G} \boldsymbol{\varepsilon} = o_p(1)$,
- (vii) $n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \mathbf{X} \beta = o_p(1)$.

Proof. (i) First note that $E(n^{-1}\boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon}) = n^{-1}Tr(\mathbf{C} \Sigma) = n^{-1} \sum_{i=1}^n \sigma_i^2 c_{ii}$, and then

$$|E(n^{-1}\boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon})| \leq n^{-1} \sum_{i=1}^n |\sigma_i^2| |c_{ii}| \leq \sup_i (\sigma_i^2) \sup_i |c_{ii}|.$$

Since by assumption $\|\mathbf{B}\|_\infty < K$, which implies that $\sup_{i,j} |b_{ij}| < K$, and then $\sup_{i,j} |c_{ij}| \leq \sup_{i,j} |b_{ij}| < K$. Also as $\sup_i (\sigma_i^2) < K$, it follows that $E(n^{-1}\boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon}) = O(1)$. Next, consider the variance of $n^{-1}\boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon}$. By Lemma S.5(iii) and the definition of \mathbf{C} , we have

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

Since $\|\mathbf{B}\|_\infty < K$ and $\sup_{i,j} |b_{ij}| < K$, by Lemma S.3 we obtain $Tr(\mathbf{B}^2) = O(n)$ and $Tr(\mathbf{B}\mathbf{B}') = O(n)$. Therefore, $Var(n^{-1}\boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon}) = O(n^{-1})$, and then (i) immediately follows from the Chebyshev's inequality, for all values of δ .

(ii) First, $E(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \boldsymbol{\varepsilon}) = Tr(n^{-1}\mathbf{C} \Sigma \mathbf{G}')$. Note that $\sup_{i,j} |c_{ij}| < K$, $\sup_i (\sigma_i^2) < K$, and hence the elements of $\mathbf{C} \Sigma$ are uniformly bounded in n . Also, $\|\mathbf{G}'\|_1 = \|\mathbf{G}\|_\infty < K$ by assumption, and then by Lemma S.1 we have $Tr(n^{-1}\mathbf{C} \Sigma \mathbf{G}') = O(1)$, which establishes that $E(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \boldsymbol{\varepsilon}) = O(1)$. Second, by Lemma S.5(iii) we have

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

Since $\|\mathbf{B}\|_\infty < K$ and $\|\mathbf{G}\|_\infty < K$ by assumption, applying Lemma S.2 yields $\|\mathbf{B}\mathbf{G}\|_\infty < K$ and $\|\mathbf{B}^2\mathbf{G}\|_\infty < K$. Then by Lemma S.1 we have $Tr[(\mathbf{B}^2 \mathbf{G}) \mathbf{G}'] = O(n)$ and $Tr[\mathbf{B}\mathbf{G}(\mathbf{B}\mathbf{G})'] = O(n)$. Since $\|\mathbf{B}\|_1 = O(n^\delta)$ by assumption, by Lemma S.4 we obtain $Tr(\mathbf{G}' \mathbf{B} \mathbf{B}' \mathbf{G}) = O(n^{\delta+1})$. Hence, $Var(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \boldsymbol{\varepsilon}) = O(n^{\delta-1})$, and by the Chebyshev's inequality the result in (ii) follows if $\delta < 1$.

(iii) By Lemma S.5, $E(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \boldsymbol{\varepsilon}) = Tr(n^{-1}\mathbf{G}' \mathbf{C} \mathbf{G} \Sigma) = Tr(n^{-1}\mathbf{G}' \mathbf{B} \mathbf{G} \Sigma)$. Under the assumptions $\|\mathbf{B}\|_\infty < K$, $\|\Sigma\|_\infty < K$, and $\|\mathbf{G}\|_\infty < K$. Applying Lemma S.2 yields $\|\mathbf{B}\mathbf{G}\Sigma\|_\infty < K$. Then by Lemma S.1 $Tr(\mathbf{G}' \mathbf{B} \mathbf{G} \Sigma) = O(n)$ and hence $E(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \boldsymbol{\varepsilon}) = O(1)$. Next, by Lemma S.5(iii)

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

Since $\|\mathbf{G}\|_1 = O(n^\delta)$ by assumption, applying Lemma S.4 gives $Tr[(\mathbf{G}' \mathbf{B} \mathbf{G})^2] = O(n^{\delta+1})$ and $Tr(\mathbf{G}' \mathbf{B} \mathbf{G} \mathbf{G}' \mathbf{B}' \mathbf{G}) = O(n^{\delta+1})$. Therefore, $Var(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \boldsymbol{\varepsilon}) = O(n^{\delta-1})$, and the result in (iii) follows if $\delta < 1$.

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

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

and then

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

where

$$\mathbf{M} = (m_{ij}) = E(\mathbf{X} \beta' \mathbf{X}'). \quad (\text{S.3})$$

Let β_l denote the l^{th} element of β , for $l = 1, 2, \dots, k$. Then, for any $1 \leq i, j \leq n$,

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

Since

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

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

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

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

(v) The proof is similar to that of (iv). The mean of $n^{-1}\beta' \mathbf{X}' \mathbf{G}' \mathbf{C} \boldsymbol{\varepsilon}$ is zero and its variance is given by $Var(n^{-1}\beta' \mathbf{X}' \mathbf{G}' \mathbf{C} \boldsymbol{\varepsilon}) = n^{-2}E(\beta' \mathbf{X}' \mathbf{G}' \mathbf{C} \Sigma \mathbf{C} \mathbf{G} \mathbf{X} \beta) = n^{-2}Tr[\mathbf{C} \Sigma \mathbf{C} (\mathbf{G} \mathbf{M} \mathbf{G}')]$, where \mathbf{M} is defined in (S.3). Let $\tilde{\mathbf{M}} = (\tilde{m}_{ij}) = \mathbf{G} \mathbf{M} \mathbf{G}'$. We have shown in (S.4) that $\sup_{1 \leq i, j \leq n} |m_{ij}| < K$. Using $\|\mathbf{G}\|_\infty < K$, and Lemma S.1(i) and (ii) yields $\sup_{1 \leq i, j \leq n} |\tilde{m}_{ij}| < K$. Repeating the arguments for $Tr(\mathbf{C} \Sigma \mathbf{C} \mathbf{M})$ in (iv) leads to $Tr(\mathbf{C} \Sigma \mathbf{C} \tilde{\mathbf{M}}) = O(n^{\delta+1})$. Therefore, $Var(n^{-1}\eta' \mathbf{C} \boldsymbol{\varepsilon}) = O(n^{\delta-1})$ and it follows that $n^{-1}\eta' \mathbf{C} \boldsymbol{\varepsilon} \rightarrow_p 0$ if $\delta < 1$.

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

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

where \mathbf{M} is defined in (S.3). Let $\mathbf{P} = (p_{ij}) = \mathbf{B} \mathbf{M}$. Then $\sup_{1 \leq i,j \leq n} |p_{ij}| < K$ follows from Lemma S.1, due to $\|\mathbf{B}\|_\infty < K$ by assumption and $\sup_{1 \leq i,j \leq n} |m_{ij}| < K$, which is proved in (S.4). Since we have also shown in the proof of (iii) that $\|\mathbf{B} \mathbf{G} \boldsymbol{\Sigma}\|_\infty < K$ and by assumption $\|\mathbf{G}\|_1 = O(n^\delta)$, applying Lemma S.4(iii) leads to $Tr[(\mathbf{B} \mathbf{G} \boldsymbol{\Sigma}) \mathbf{G}' \mathbf{P}] = O(n^{\delta+1})$. Similarly, the remaining three traces in (S.5) can be shown to be $O(n^{\delta+1})$ by applying Lemmas S.1, S.2 and S.4. Consequently, $Var(n^{-1}\beta' \mathbf{X}' \mathbf{C} \mathbf{G} \boldsymbol{\varepsilon}) = O(n^{\delta-1})$, and we obtain $n^{-1}\beta' \mathbf{X}' \mathbf{C} \mathbf{G} \boldsymbol{\varepsilon} \rightarrow_p 0$, if $\delta < 1$.

(vii) It is easily seen that $n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \mathbf{X} \beta$ has mean zero and $Var(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \mathbf{X} \beta) = n^{-2} Tr(\mathbf{C} \mathbf{G} \boldsymbol{\Sigma} \mathbf{G}' \tilde{\mathbf{M}})$, where as before $\tilde{\mathbf{M}} = (\tilde{m}_{ij}) = \mathbf{G} \mathbf{M} \mathbf{G}'$ and \mathbf{M} is defined by (S.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 (S.5), it follows that $Var(n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \mathbf{X} \beta) = O(n^{\delta-1})$. Therefore, $n^{-1}\boldsymbol{\varepsilon}' \mathbf{G}' \mathbf{C} \mathbf{G} \mathbf{X} \beta \rightarrow_p 0$, if $\delta < 1$. ■

Lemma S.8 Suppose that \mathbf{G} is an $n \times n$ matrix of fixed constants such that $\|\mathbf{G}\|_\infty < K$ and $\|\mathbf{G}\|_1 = O(n^\delta)$, where δ is a fixed constant in the range $0 \leq \delta \leq 1$. Then

$$\mathbf{1}_n' \mathbf{G}' \mathbf{1}_n = \mathbf{1}_n' \mathbf{G} \mathbf{1}_n = O(n), \quad (\text{S.6})$$

$$\mathbf{1}_n' \mathbf{G}' \mathbf{G} \mathbf{1}_n = O(n), \quad (\text{S.7})$$

$$Tr(n^{-1}\mathbf{G}^s) = O(1), \text{ for } s = 1, 2, \dots, \quad (\text{S.8})$$

$$Tr(n^{-1}\mathbf{G}' \mathbf{G}) = O(1). \quad (\text{S.9})$$

Proof. First note that $\mathbf{G} \mathbf{1}_n \leq \|\mathbf{G}\|_\infty \mathbf{1}_n < K \mathbf{1}_n$, and then (S.6) and (S.7) readily follow. Denote the diagonal elements of \mathbf{G}^s by $g_{s,ii}$ and notice that

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

Since $\|\mathbf{G}^s\|_\infty \leq (\|\mathbf{G}\|_\infty)^s < K$, all elements of \mathbf{G}^s must be bounded, specifically $|g_{s,ii}| < K$ and result (S.8) follows. Finally, let $\mathbf{G} = (g_{ij})$ and then

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

but by assumption $\sup_i \sum_{j=1}^n |g_{ij}| = \|\mathbf{G}\|_\infty < K$, and hence $\text{Tr}(\mathbf{G}'\mathbf{G}) = O(n)$ and the result (S.9) is established. ■

Lemma S.9 Suppose that $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)'$ is a vector of random variables, where ε_i , for $i = 1, 2, \dots, n$, are independently distributed over i with zero means and variances, σ_i^2 , such that $\inf_i (\sigma_i^2) > c > 0$ and $\sup_i (\sigma_i^2) < K$. Let $\boldsymbol{\Sigma} = \text{Diag}(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$ and $\gamma_{2i} = \mu_{4i} - 3\sigma_i^4$, where $\mu_{4i} = E(\varepsilon_i^4)$ and assume that $\sup_i |\mu_{4i}| < K$. $\boldsymbol{\beta}$ is a $k \times 1$ vector of fixed coefficients such that $\|\boldsymbol{\beta}\|_1 < K$. $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)'$ is an $n \times k$ matrix of random variables such that $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ik})'$ are distributed independently of the errors, ε_j , for all i and j , $\sup_{i,s} E|x_{is}|^2 < K$, and $n^{-1}\mathbf{X}'\mathbf{X} \rightarrow_p \boldsymbol{\Sigma}_{xx}$ is positive definite. Let $\mathbf{M}_x = \mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. $\mathbf{G} = (g_{ij})$ is an $n \times n$ matrix of fixed constants such that $\|\mathbf{G}\|_\infty < K$, and $\|\mathbf{G}\|_1 = O(n^\delta)$, where δ is a fixed constant in the range $0 \leq \delta < 1$. Then

- (i) $E(n^{-1}\mathbf{X}'\mathbf{G}'\mathbf{G}\mathbf{X}) = O(1)$,
- (ii) $E(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\mathbf{G}\mathbf{X}\boldsymbol{\beta}) = O(1)$,
- (iii) $\text{Var}(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\mathbf{G}\mathbf{X}\boldsymbol{\beta}) = O(n^{\delta-1})$,
- (iv) $\text{Var}(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}) = O(n^{-1})$,
- (v) $n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\text{Diag}(\mathbf{G}^2)\mathbf{M}_x\boldsymbol{\varepsilon} = \text{Tr}[n^{-1}\text{Diag}(\mathbf{G}^2)\boldsymbol{\Sigma}] + O_p(n^{-1/2})$,
- (vi) $n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\text{Diag}(\mathbf{G})\mathbf{M}_x\boldsymbol{\varepsilon} = O_p(n^{-1/2})$,
- (vii) $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\text{Diag}(\mathbf{G})\mathbf{M}_x\boldsymbol{\varepsilon} = \text{Tr}[n^{-1}\mathbf{G}'\text{Diag}(\mathbf{G})\boldsymbol{\Sigma}] + O_p(n^{-1/2})$.

Proof. (i) Let $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$, where $\mathbf{x}_l = (x_{1l}, x_{2l}, \dots, x_{nl})'$, $l = 1, 2, \dots, k$. Then the $(l, l')^{th}$ element of $\mathbf{X}'\mathbf{G}'\mathbf{G}\mathbf{X}$ is given by $\mathbf{x}_l' \mathbf{G}' \mathbf{G} \mathbf{x}_{l'}$, and

$$E|\mathbf{x}_l' \mathbf{G}' \mathbf{G} \mathbf{x}_{l'}| \leq [E(\mathbf{x}_l' \mathbf{G}' \mathbf{G} \mathbf{x}_l)]^{1/2} [E(\mathbf{x}_{l'}' \mathbf{G}' \mathbf{G} \mathbf{x}_{l'})]^{1/2}.$$

Therefore, to examine the limiting properties of $\mathbf{X}'\mathbf{G}'\mathbf{G}\mathbf{X}$, it is sufficient to consider $E(\mathbf{x}_l' \mathbf{G}' \mathbf{G} \mathbf{x}_l)$. To this end, note that $\mathbf{x}_l' \mathbf{G}' \mathbf{G} \mathbf{x}_l = \sum_{i=1}^n \zeta_{il}^2$, where $\zeta_l = \mathbf{G} \mathbf{x}_l = (\zeta_{1l}, \zeta_{2l}, \dots, \zeta_{nl})'$. But $\zeta_{il} = \sum_{j=1}^n g_{ij} x_{jl}$, and by Minkowski's inequality

$$\|\zeta_l\|_2 \leq \sum_{j=1}^n |g_{ij}| \|x_{jl}\|_2 \leq \sup_j \|x_{jl}\|_2 \sum_{j=1}^n |g_{ij}|, \quad (\text{S.10})$$

where $\|x_{jl}\|_2 = [E(x_{jl}^2)]^{1/2}$. By assumption $\sup_i \sum_{j=1}^n |g_{ij}| = \|\mathbf{G}\|_\infty < K$, which leads to $\sup_{i,l} \|\zeta_{il}\|_2 \leq K \sup_{j,l} \|x_{jl}\|_2$. Since $\sup_{j,l} E|x_{js}|^2 < K$ by assumption, we obtain $\sup_{i,l} \|\zeta_{il}\|_2 < K$. Therefore, $E(\mathbf{x}_l' \mathbf{G}' \mathbf{G} \mathbf{x}_l) = \sum_{i=1}^n E(\zeta_{il}^2) \leq n \sup_i E(\zeta_{il}^2)$ and $\sup_l E(n^{-1}\mathbf{x}_l' \mathbf{G}' \mathbf{G} \mathbf{x}_l) = O(1)$, and hence $E(n^{-1}\mathbf{X}'\mathbf{G}'\mathbf{G}\mathbf{X}) = O(1)$.

(ii) Let $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)' = \mathbf{X}\boldsymbol{\beta}$. Since $\xi_i = \sum_{j=1}^k x_{ij} \beta_j$, for $i = 1, 2, \dots, n$, by Minkowski's inequality:

$$\sup_i \|\xi_i\|_2 \leq \sup_i \sum_{j=1}^k |\beta_j| \|x_{ij}\|_2 \leq \left(\sup_j |\beta_j| \right) \left(\sup_{i,j} \|x_{ij}\|_2 \right) < K. \quad (\text{S.11})$$

Since \mathbf{M}_x is an idempotent matrix, we have $n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\mathbf{G}\mathbf{X}\boldsymbol{\beta} \leq n^{-1}\boldsymbol{\xi}'\mathbf{G}'\mathbf{G}\boldsymbol{\xi} = n^{-1}\sum_{i=1}^n \eta_i^2$, where $\boldsymbol{\eta} = (\eta_1, \eta_2, \dots, \eta_n)' = \mathbf{G}\boldsymbol{\xi}$. Then $\eta_i = \sum_{j=1}^n g_{ij}\boldsymbol{\xi}_j$, and by the same argument as (S.10) we obtain

$$E(n^{-1}\boldsymbol{\xi}'\mathbf{G}'\mathbf{G}\boldsymbol{\xi}) = E\left(n^{-1}\sum_{i=1}^n \eta_i^2\right) \leq \sup_i E(\eta_i^2) = O(1), \quad (\text{S.12})$$

and the result in (ii) immediately follows.

(iii) Note that $E(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\mathbf{G}\mathbf{X}\boldsymbol{\beta}) = 0$ and

$$\begin{aligned} \text{Var}(n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'\mathbf{M}_x\mathbf{G}\mathbf{X}\boldsymbol{\beta}) &= E(n^{-2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\mathbf{G}\Sigma\mathbf{G}'\mathbf{M}_x\mathbf{G}\mathbf{X}\boldsymbol{\beta}) \\ &\leq \sup_i (\sigma_i^2) E\left[n^{-2}\boldsymbol{\beta}'\mathbf{X}'(\mathbf{G}'\mathbf{M}_x\mathbf{G})^2\mathbf{X}\boldsymbol{\beta}\right] \\ &\leq \sup_i (\sigma_i^2) E\left[n^{-2}\boldsymbol{\xi}'(\mathbf{G}'\mathbf{G})^2\boldsymbol{\xi}\right], \end{aligned} \quad (\text{S.13})$$

where $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_n)' = \mathbf{X}\boldsymbol{\beta}$. Let $\mathbf{d} = (d_1, d_2, \dots, d_n)' = \mathbf{G}'\mathbf{G}\boldsymbol{\xi}$. Then

$$E\left[n^{-2}\boldsymbol{\xi}'(\mathbf{G}'\mathbf{G})^2\boldsymbol{\xi}\right] = E\left(n^{-2}\sum_{i=1}^n d_i^2\right) \leq n^{-1} \sup_i E(d_i^2). \quad (\text{S.14})$$

Let a_{ij} denote the $(i, j)^{th}$ element of $\mathbf{G}'\mathbf{G}$. Then d_i can be written as $d_i = \sum_{j=1}^n a_{ij}\xi_j$. Using (S.11) and noting that $\sup_i \sum_{j=1}^n |a_{ij}| = O(n^\delta)$, we obtain

$$\sup_i \|d_i\|_2 \leq \sup_i \sum_{j=1}^n |a_{ij}| \|\xi_j\|_2 \leq \left(\sup_j \|\xi_j\|_2\right) \left(\sup_i \sum_{j=1}^n |a_{ij}|\right) = O(n^\delta).$$

Using this result in (S.14) yields $E[n^{-2}\boldsymbol{\xi}'(\mathbf{G}'\mathbf{G})^2\boldsymbol{\xi}] = O(n^{\delta-1})$, which together with (S.13) establishes the result in (iii).

(iv) Similarly to (iii), $E(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}) = 0$ and then

$$\begin{aligned} \text{Var}(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}) &= E(n^{-2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\Sigma\mathbf{M}_x\mathbf{G}\mathbf{X}\boldsymbol{\beta}) \\ &\leq \sup_i (\sigma_i^2) E(n^{-2}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{G}\mathbf{X}\boldsymbol{\beta}) \leq KE(n^{-2}\boldsymbol{\xi}'\mathbf{G}'\mathbf{G}\boldsymbol{\xi}). \end{aligned}$$

In view of (S.12), it readily follows that $\text{Var}(n^{-1}\boldsymbol{\beta}'\mathbf{X}'\mathbf{G}'\mathbf{M}_x\boldsymbol{\varepsilon}) = O(n^{-1})$.

(v) Let $\mathbf{P}_x = \mathbf{I}_n - \mathbf{M}_x = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ and $\mathbf{D} = \text{Diag}(\mathbf{G}^2)$. Then

$$n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\mathbf{D}\mathbf{M}_x\boldsymbol{\varepsilon} = n^{-1}\boldsymbol{\varepsilon}'\mathbf{D}\boldsymbol{\varepsilon} - 2n^{-1}\boldsymbol{\varepsilon}'\mathbf{P}_x\mathbf{D}\boldsymbol{\varepsilon} + n^{-1}\boldsymbol{\varepsilon}'\mathbf{P}_x\mathbf{D}\mathbf{P}_x\boldsymbol{\varepsilon}.$$

First note that $n^{-1}\boldsymbol{\varepsilon}'\mathbf{P}_x\mathbf{D}\boldsymbol{\varepsilon} = \frac{\boldsymbol{\varepsilon}'\mathbf{X}}{n} \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1} \frac{\mathbf{X}'\mathbf{D}\boldsymbol{\varepsilon}}{n}$, where it is immediate that $\frac{\boldsymbol{\varepsilon}'\mathbf{X}}{n} = O_p(n^{-1/2})$ and $\frac{\mathbf{X}'\mathbf{X}}{n} = O_p(1)$. To see $\frac{\mathbf{X}'\mathbf{D}\boldsymbol{\varepsilon}}{n} = O_p(1)$, notice that

$$\left\| \frac{\mathbf{X}'\mathbf{D}\boldsymbol{\varepsilon}}{n} \right\|_F \leq \frac{\|\mathbf{X}\|_F}{\sqrt{n}} \frac{\|\mathbf{D}\boldsymbol{\varepsilon}\|_F}{\sqrt{n}} = \left[\text{Tr}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right) \right]^{1/2} \left[\text{Tr}\left(\frac{\boldsymbol{\varepsilon}'\mathbf{D}^2\boldsymbol{\varepsilon}}{n}\right) \right]^{1/2},$$

which by the independence of \mathbf{X} and $\boldsymbol{\varepsilon}$ further implies

$$E \left\| \frac{\mathbf{X}' \mathbf{D} \boldsymbol{\varepsilon}}{n} \right\|_F^2 \leq E \left[Tr \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right) \right] E \left[Tr \left(\frac{\boldsymbol{\varepsilon}' \mathbf{D}^2 \boldsymbol{\varepsilon}}{n} \right) \right] = Tr \left[E \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right) \right] Tr \left(\frac{\mathbf{D}^2 \Sigma}{n} \right) < K.$$

Therefore, we have $n^{-1} \boldsymbol{\varepsilon}' \mathbf{P}_x \mathbf{D} \boldsymbol{\varepsilon} = O_p(n^{-1/2})$. Similarly,

$$\frac{1}{n} \boldsymbol{\varepsilon}' \mathbf{P}_x \mathbf{D} \mathbf{P}_x \boldsymbol{\varepsilon} = \frac{\boldsymbol{\varepsilon}' \mathbf{X}}{n} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}' \mathbf{D} \mathbf{X}}{n} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} = O_p(n^{-1/4}).$$

Hence, $n^{-1} \boldsymbol{\varepsilon}' \mathbf{M}_x \mathbf{D} \mathbf{M}_x \boldsymbol{\varepsilon} = n^{-1} \boldsymbol{\varepsilon}' \mathbf{D} \boldsymbol{\varepsilon} + O_p(n^{-1/2})$, and the result in (v) readily follows.

(vi) Let $\check{\mathbf{G}} = Diag(\mathbf{G})$. By expansion,

$$n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \mathbf{M}_x \check{\mathbf{G}} \mathbf{M}_x \boldsymbol{\varepsilon} = n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \mathbf{M}_x \check{\mathbf{G}} \boldsymbol{\varepsilon} - n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \check{\mathbf{G}} \mathbf{P}_x \boldsymbol{\varepsilon} + n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \mathbf{P}_x \check{\mathbf{G}} \mathbf{P}_x \boldsymbol{\varepsilon},$$

Following the same arguments as in the proof of (iii), it can be shown that $Var(n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \mathbf{M}_x \check{\mathbf{G}} \boldsymbol{\varepsilon}) = O(n^{-1})$. Since $E(n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \mathbf{M}_x \check{\mathbf{G}} \boldsymbol{\varepsilon}) = 0$, we then have $n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \mathbf{M}_x \check{\mathbf{G}} \boldsymbol{\varepsilon} = O_p(n^{-1/2})$.

Moreover, it is easy to see that

$$n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \check{\mathbf{G}} \mathbf{P}_x \boldsymbol{\varepsilon} = \frac{\boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \check{\mathbf{G}} \mathbf{X}}{n} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} = O_p(1) O_p(1) O_p(n^{-1/2}) = O_p(n^{-1/2}),$$

and

$$n^{-1} \boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \mathbf{P}_x \check{\mathbf{G}} \mathbf{P}_x \boldsymbol{\varepsilon} = \frac{\boldsymbol{\beta}' \mathbf{X}' \mathbf{G}' \check{\mathbf{G}} \mathbf{X}}{n} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}' \check{\mathbf{G}} \mathbf{X}}{n} \left(\frac{\mathbf{X}' \mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} = O_p(n^{-1/2}).$$

Then the desired result follows.

(vii) The proof is similar to that of (v). ■

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

This lemma is a variation of Corollary 3.1 in Hall and Heyde (1980) and a proof can be found therein. Note that condition (b) in Lemma S.10 is a sufficient condition for the conditional Lindeberg condition (3.7) stated in Corollary 3.1 (see, for example, Davidson, 1994, Theorem 23.11).

Lemma S.11 *Suppose that ρ is a fixed constant and $\mathbf{W} = (w_{ij})$ is an $n \times n$ constant matrix such that $\|\mathbf{W}\|_\infty < K$, $|\rho| \|\mathbf{W}\|_\infty < 1$, and $\|\mathbf{W}\|_1 = \Theta(n^\delta)$, where δ is a fixed constant in the range $0 \leq \delta \leq 1$. Let $\mathbf{G} = \mathbf{W} \mathbf{S}^{-1}(\rho)$, where $\mathbf{S}(\rho) = \mathbf{I}_n - \rho \mathbf{W}$. Let $h_n = n^{-1} Tr(\mathbf{G}^2 + \mathbf{G}' \mathbf{G}) - 2n^{-2} [Tr(\mathbf{G})]^2$, and suppose that Assumption 4 is met. Then $\lim_{n \rightarrow \infty} h_n = 0$, if $\lim_{n \rightarrow \infty} n^{-1} Tr(\mathbf{W}' \mathbf{W}) = 0$.*

Proof. Under Assumption 4, $\lambda_{\min}[\mathbf{S}(\rho)] > c > 0$, and $[\mathbf{S}'(\rho)\mathbf{S}(\rho)]^{-1}$ is a positive definite matrix. (See also Remark 1). Then by result (12) of Lütkepohl (1996, p.44) we obtain

$$\frac{1}{n} \text{Tr}(\mathbf{G}'\mathbf{G}) = \text{Tr}\left\{[\mathbf{S}'(\rho)\mathbf{S}(\rho)]^{-1} \frac{\mathbf{W}'\mathbf{W}}{n}\right\} \leq \lambda_{\max}\left\{[\mathbf{S}'(\rho)\mathbf{S}(\rho)]^{-1}\right\} \frac{\text{Tr}(\mathbf{W}'\mathbf{W})}{n}. \quad (\text{S.15})$$

Also

$$\lambda_{\max}[\mathbf{S}'(\rho)\mathbf{S}(\rho)] \leq \|\mathbf{S}(\rho)\|_{\infty} \|\mathbf{S}(\rho)\|_1 \leq (1 + |\rho| \|\mathbf{W}\|_{\infty}) (1 + |\rho| \|\mathbf{W}\|_1) = \Theta(n^{\delta}).$$

It follows that $c < \lambda_{\min}[\mathbf{S}'(\rho)\mathbf{S}(\rho)] \leq Kn^{\delta}$ and then

$$cn^{-\delta} \leq \lambda_{\max}\left\{[\mathbf{S}'(\rho)\mathbf{S}(\rho)]^{-1}\right\} = \{\lambda_{\min}[\mathbf{S}'(\rho)\mathbf{S}(\rho)]\}^{-1} < K. \quad (\text{S.16})$$

From (S.15) and (S.16), we see that if $\lim_{n \rightarrow \infty} n^{-1} \text{Tr}(\mathbf{W}'\mathbf{W}) = 0$ then we must have $\lim_{n \rightarrow \infty} n^{-1} \text{Tr}(\mathbf{G}'\mathbf{G}) = 0$. But by the Cauchy-Schwarz inequality, $n^{-1} \text{Tr}(\mathbf{G}'\mathbf{G}) \geq n^{-2} [\text{Tr}(\mathbf{G})]^2 \geq 0$, Therefore, if $\lim_{n \rightarrow \infty} n^{-1} \text{Tr}(\mathbf{G}'\mathbf{G}) = 0$, then $\lim_{n \rightarrow \infty} n^{-2} [\text{Tr}(\mathbf{G})]^2 = 0$. Finally, by Schur's inequality we have $n^{-1} \text{Tr}(\mathbf{G}^2) \leq n^{-1} \text{Tr}(\mathbf{G}'\mathbf{G})$, and then $0 \leq h_n \leq 2n^{-1} \text{Tr}(\mathbf{G}'\mathbf{G}) - 2n^{-2} [\text{Tr}(\mathbf{G})]^2$. Hence, if $\lim_{n \rightarrow \infty} n^{-1} \text{Tr}(\mathbf{G}'\mathbf{G}) = 0$, we must have $\lim_{n \rightarrow \infty} h_n = 0$, and this completes the proof. ■

Remark S.2 Lemma S.11 also holds for finite n , but this case is trivial since $n^{-1} \text{Tr}(\mathbf{W}'\mathbf{W}) = 0$ implies $\mathbf{W} = \mathbf{0}$.

S1.2 Proofs of theorems and propositions

The following proofs make use of the lemmas in Section S1.1 of this online supplement. 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 ϖ_n^2 given by (14) in the paper and show that ϖ_n^2 is bounded. Note that (10) in the paper implies that p_{ij} (or p_{ji}) must all be bounded in n . By definition, $a_{ij} = (p_{ij} + p_{ji})/2$, and hence $\sup_{i,j} |a_{ij}| \leq (\sup_{i,j} |p_{ij}| + \sup_{i,j} |p_{ji}|)/2 < K$. Using (14) given in the paper we now have

$$\varpi_n^2 \leq K \sup_{i,n} |\mu_{4i} - 3\sigma_i^4| + 2 \left(\sup_{i,n} \sigma_i^4 \right) [\text{Tr}(n^{-1}\mathbf{A}^2)].$$

Furthermore,

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

and

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

But under (10) of the paper, $\sup_i \sum_{j=1}^n |p_{ij}| < K$, and we have $Tr(n^{-1}\mathbf{P}'\mathbf{P}) \leq K$, which also implies that $Tr(n^{-1}\mathbf{P}^2) < K$. Hence, ϖ_n^2 is bounded in n for all values of $0 \leq \alpha \leq 1$. Also note that condition (12) in the paper ensures that $\varpi_n^2 > 0$, for all n (including $n \rightarrow \infty$).

Consider Q defined by (13) in the paper and following Kelejian and Prucha (2001) write it as $Q = \sum_{i=1}^n X_i$, where

$$X_i = \varpi_n^{-1} n^{-1/2} a_{ii} (\varepsilon_i^2 - \sigma_i^2) + 2\varpi_n^{-1} n^{-1/2} \varepsilon_i \zeta_{i-1}, \quad (\text{S.17})$$

and

$$\zeta_{i-1} = \sum_{j=1}^{i-1} a_{ij} \varepsilon_j. \quad (\text{S.18})$$

Clearly, $E(X_i) = 0$ and

$$\begin{aligned} E(X_i^2) &= \varpi_n^{-2} n^{-1} E[a_{ii} (\varepsilon_i^2 - \sigma_i^2) + 2\varepsilon_i \zeta_{i-1}]^2 \\ &= \varpi_n^{-2} n^{-1} E[a_{ii}^2 (\varepsilon_i^4 + \sigma_i^4 - 2\varepsilon_i^2 \sigma_i^2) + 4\varepsilon_i^2 \zeta_{i-1}^2 + 4a_{ii} (\varepsilon_i^2 - \sigma_i^2) \varepsilon_i \zeta_{i-1}] \\ &= \varpi_n^{-2} n^{-1} \left[a_{ii}^2 (\mu_{4i} - \sigma_i^4) + 4\sigma_i^2 \sum_{j=1}^{i-1} a_{ij}^2 \sigma_j^2 \right]. \end{aligned} \quad (\text{S.19})$$

Notice that (14) in the paper can be written equivalently as

$$\varpi_n^2 = n^{-1} \sum_{i=1}^n a_{ii}^2 (\mu_{4i} - \sigma_i^4) + 4n^{-1} \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^{i-1} a_{ij}^2 \sigma_j^2 > 0. \quad (\text{S.20})$$

Using (S.19) and (S.20) leads to $\sum_{i=1}^n E(X_i^2) = 1$. Note that $\{X_i, 1 \leq i \leq n\}$ forms a martingale difference array with respect to the filtration $\mathfrak{F}_{i-1}^\varepsilon = \sigma[(\varepsilon_j)_{j=1}^{i-1}]$ (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 $\mathfrak{F}_{i-1}^x = \sigma[(X_j)_{j=1}^{i-1}]$. Hence, the central limit theorem given in Lemma S.10 is applicable to Q if the three conditions on $\{X_i, \mathfrak{F}_{i-1}^x\}$ can be established. Since we have shown that $\sum_{i=1}^n E(X_i^2) = 1$, and $E(X_i^2) \geq 0$ for all i , it follows that $E(X_i^2) \leq 1$, and hence X_i^2 is square integrable for all values of $0 \leq \alpha \leq 1$. In what follows, we only need to show that conditions (b) and (c) of Lemma S.10 hold under $0 \leq \alpha < 1/2$.

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

$$E|X_i|^q = \varpi_n^{-q} n^{-\frac{q}{2}} E |a_{ii} (\varepsilon_i^2 - \sigma_i^2) + 2\varepsilon_i \zeta_{i-1}|^q$$

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

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

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

and

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

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 (see, for example, Davidson (1994), p.140),

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

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 (10) 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 (11) in the paper that

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

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

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

or equivalently,

$$\sum_{i=1}^n E |X_i|^{2+\nu} = \begin{cases} O\left(n^{-\frac{\nu}{2}}\right), & \text{if } \alpha \leq \frac{1}{2+\nu}, \\ O\left[n^{(\alpha-\frac{1}{2})(2+\nu)}\right], & \text{if } \alpha > \frac{1}{2+\nu}. \end{cases} \quad (\text{S.21})$$

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

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

$$E(X_i^2 | \mathfrak{F}_{i-1}^x) = \frac{a_{ii}^2(\mu_{4i} - \sigma_i^4)}{n\varpi_n^2} + \frac{4\sigma_i^2\zeta_{i-1}^2}{n\varpi_n^2} + \frac{4a_{ii}\mu_{3i}\zeta_{i-1}}{n\varpi_n^2},$$

and it follows that

$$\begin{aligned} \sum_{i=1}^n E(X_i^2 | \mathfrak{F}_{i-1}^x) - 1 &= \frac{\sum_{i=1}^n a_{ii}^2(\mu_{4i} - \sigma_i^4)}{n\varpi_n^2} + \frac{4\sum_{i=1}^n \sigma_i^2\zeta_{i-1}^2}{n\varpi_n^2} \\ &\quad + \frac{4\sum_{i=1}^n a_{ii}\mu_{3i}\zeta_{i-1}}{n\varpi_n^2} - \frac{\sum_{i=1}^n a_{ii}^2(\mu_{4i} - \sigma_i^4) + 4\sum_{i=1}^n \sigma_i^2 \sum_{j=1}^{i-1} a_{ij}^2 \sigma_j^2}{n\varpi_n^2} \\ &= \frac{4\left[\sum_{i=1}^n \sigma_i^2\zeta_{i-1}^2 - \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^{i-1} a_{ij}^2\right]}{n\varpi_n^2} + \frac{4\sum_{i=1}^n a_{ii}\mu_{3i}\zeta_{i-1}}{n\varpi_n^2} \\ &= \varpi_n^{-2}(8H_1 + 4H_2 + 4H_3), \end{aligned}$$

where

$$H_1 = n^{-1} \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} a_{ij} a_{ik} \varepsilon_j \varepsilon_k, \quad (\text{S.22})$$

$$H_2 = n^{-1} \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^{i-1} a_{ij}^2 (\varepsilon_j^2 - \sigma_j^2), \quad (\text{S.23})$$

$$H_3 = n^{-1} \sum_{i=1}^n a_{ii}\mu_{3i} \sum_{j=1}^{i-1} a_{ij} \varepsilon_j. \quad (\text{S.24})$$

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

$$H_1^2 = n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} \sum_{l=1}^n \sum_{r=1}^{l-1} \sum_{s=1}^{r-1} \sigma_i^2 \sigma_l^2 a_{ij} a_{ik} a_{lj} a_{lk} \varepsilon_j \varepsilon_k \varepsilon_r \varepsilon_s.$$

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

$$E(H_1^2) = 2n^{-2} \sum_{l=1}^n \sum_{i=1}^l \sum_{j=1}^{i-1} \sum_{k=1}^{j-1} a_{ij} a_{ik} a_{lj} a_{lk} \sigma_l^2 \sigma_i^2 \sigma_j^2 \sigma_k^2$$

$$\begin{aligned}
&\leq 2n^{-2} \left(\sup_i \sigma_i^2 \right)^4 \sum_{l=1}^n \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n |a_{ij}| |a_{ik}| |a_{lj}| |a_{lk}|, \\
&\leq Kn^{-2} \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \sum_{l=1}^n |a_{lj}| \left(\sum_{k=1}^n |a_{ik}| |p_{lk}| + \sum_{k=1}^n |a_{ik}| |p_{kl}| \right) \\
&\leq Kn^{-2} \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \sum_{l=1}^n |a_{lj}| \left(\sup_{1 \leq l \leq n} \sum_{k=1}^n |p_{lk}| \right) \left(\sup_{1 \leq i, k \leq n} |a_{ik}| \right) \\
&\quad + Kn^{-2} \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \sum_{k=1}^n |a_{ik}| \left(\sup_{1 \leq k \leq n} \sum_{l=1}^n |p_{kl}| \right) \left(\sup_{1 \leq l, j \leq n} |a_{lj}| \right) \\
&\leq Kn^{-2} \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| \left(\sup_{1 \leq j \leq n} \sum_{l=1}^n |a_{lj}| + \sup_{1 \leq i \leq n} \sum_{k=1}^n |a_{ik}| \right) \leq Kn^{\alpha-1}.
\end{aligned}$$

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

$$\begin{aligned}
H_2^2 &= n^{-2} \sum_{i=1}^n \sigma_i^2 \sum_{j=1}^{i-1} a_{ij}^2 (\varepsilon_j^2 - \sigma_j^2) \sum_{k=1}^n \sigma_k^2 \sum_{l=1}^{k-1} a_{kl}^2 (\varepsilon_l^2 - \sigma_l^2) \\
&= n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sum_{l=1}^{k-1} \sigma_i^2 \sigma_k^2 a_{ij}^2 a_{kl}^2 \varepsilon_j^2 \varepsilon_l^2 + n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sum_{l=1}^{k-1} \sigma_i^2 \sigma_j^2 \sigma_k^2 \sigma_l^2 a_{ij}^2 a_{kl}^2 \\
&\quad - n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sum_{l=1}^{k-1} \sigma_i^2 \sigma_k^2 \sigma_l^2 a_{ij}^2 a_{kl}^2 \varepsilon_j^2 - n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sum_{l=1}^{k-1} \sigma_i^2 \sigma_k^2 \sigma_j^2 a_{ij}^2 a_{kl}^2 \varepsilon_l^2,
\end{aligned}$$

which leads to

$$\begin{aligned}
E(H_2^2) &= n^{-2} \left(\sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sigma_i^2 \sigma_k^2 a_{ij}^2 a_{kj}^2 \mu_{4j} + \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sum_{l=1, l \neq j}^{k-1} \sigma_i^2 \sigma_k^2 \sigma_j^2 \sigma_l^2 a_{ij}^2 a_{kl}^2 \right) \\
&\quad - n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sum_{l=1}^{k-1} \sigma_i^2 \sigma_j^2 \sigma_k^2 \sigma_l^2 a_{ij}^2 a_{kl}^2 \\
&= n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \sigma_i^2 \sigma_k^2 a_{ij}^2 a_{kj}^2 (\mu_{4j} - \sigma_j^4) \\
&\leq n^{-2} \left(\sup_i \sigma_i^2 \right)^2 \sup_j |\mu_{4j} - \sigma_j^4| \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \left(\sup_{1 \leq j \leq n} \sum_{k=1}^n |a_{kj}| \right) \left(\sup_{1 \leq k, j \leq n} |a_{kj}| \right) \\
&\leq Kn^{\alpha-1},
\end{aligned}$$

where in the last line we used $n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 = Tr(n^{-1} \mathbf{A}' \mathbf{A}) < K$. Thus, we obtain that

$H_2 = o_p(1)$ if $\alpha < 1$. Lastly, $E(H_3) = 0$, and

$$H_3^2 = n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \mu_{3i} a_{ii} a_{ij} \varepsilon_j \sum_{k=1}^n \sum_{l=1}^{k-1} \mu_{3k} a_{kk} a_{kl} \varepsilon_l,$$

and it follows that

$$\begin{aligned} E(H_3^2) &= n^{-2} \sum_{i=1}^n \sum_{j=1}^{i-1} \sum_{k=1}^n \mu_{3i} \mu_{3k} \sigma_j^2 a_{ii} a_{ij} a_{kk} a_{kj} \\ &\leq n^{-2} \left(\sup_i |\mu_{3i}| \right)^2 \left(\sup_j \sigma_j^2 \right) \sum_{i=1}^n \sum_{j=1}^n |a_{ii}| |a_{ij}| \left(\sup_{1 \leq j \leq n} \sum_{k=1}^n |a_{kj}| \right) \left(\sup_{1 \leq k \leq n} |a_{kk}| \right) \\ &\leq K n^{\alpha-1}. \end{aligned}$$

Hence, $H_3 = o_p(1)$ if $\alpha < 1$. Overall, we conclude that $\sum_{i=1}^n E(X_i^2 | \mathfrak{F}_{i-1}^x) \rightarrow_p 1$ if $0 \leq \alpha < 1$, which proves condition (c) of Lemma S.10. Combining our findings for (a)–(c) establishes the result in (13) of the paper under $0 \leq \alpha < 1/2$. ■

Proof of Theorem 2. We begin by showing that $\tilde{\varpi}_n^2$, which is defined in (17) of the paper, is bounded in n for all $0 \leq \alpha \leq 1$. Note that

$$\tilde{\varpi}_n^2 = \varpi_n^2 + n^{-1} \sum_{i=1}^n \sigma_{\eta,i}^2 \sigma_i^2 + 2n^{-1} \sum_{i=1}^n a_{ii} \mu_{\eta,i} \mu_{3i},$$

where ϖ_n^2 is defined by (14) in the paper. We have shown in the above proof of Theorem 1 that ϖ_n^2 is bounded in n for all $0 \leq \alpha \leq 1$, and since $\sup_i \sigma_{\eta,i}^2 < K$, $\sup_i \mu_{\eta,i} < K$, $\sup_i \sigma_i^2 < K$, $\sup_i \mu_{3i} < K$ and $\sup_i |a_{ii}| < K$, it is immediate that $\tilde{\varpi}_n^2$ is bounded in n for $0 \leq \alpha \leq 1$. Also note that condition (15) in the paper implies that $\tilde{\varpi}_n^2 > 0$, for all n (including $n \rightarrow \infty$).

Consider \tilde{Q} defined by (16) in the paper and write it as $\tilde{Q} = \sum_{i=1}^n Y_i$, where

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

and ζ_{i-1} is defined in (S.18). 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 S.10, we need to show in turn that the three conditions (a)–(c) are satisfied for $\{Y_i, \mathfrak{F}_{i-1}^y\}$.

First, we see that

$$E(Y_i^2) = \tilde{\varpi}_n^{-2} n^{-1} \left[a_{ii}^2 (\mu_4 - \sigma_i^4) + 4\sigma_i^2 \sum_{j=1}^{i-1} a_{ij}^2 \sigma_j^2 + \sigma_i^2 \sigma_{\eta,i}^2 + 2\mu_{3i} a_{ii} \mu_{\eta,i} \right].$$

Using (17) in the paper we obtain $\sum_{i=1}^n E(Y_i^2) = 1$. Since $E(Y_i^2) \geq 0$ for all i , we readily have $E(Y_i^2) \leq 1$ and hence Y_i is square integrable.

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

$$\begin{aligned} \sum_{i=1}^n E|Y_i|^{2+c} &= \tilde{\omega}_n^{-(2+c)} n^{-\frac{2+c}{2}} \sum_{i=1}^n E|Y_{1,i} + Y_{2,i}|^{2+c} \\ &\leq 2^{1+c} \tilde{\omega}_n^{-(2+c)} n^{-\frac{2+c}{2}} \sum_{i=1}^n (E|Y_{1,i}|^{2+c} + E|Y_{2,i}|^{2+c}). \end{aligned}$$

Since

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

it follows that $n^{-\frac{2+c}{2}} \sum_{i=1}^n E|Y_{2,i}|^{2+c} = O(n^{-\frac{c}{2}})$, which converges to zero for all values of $0 \leq \alpha \leq 1$. In addition, note that $Y_{1,i} = n^{1/2} \varpi_n X_i$, where X_i is defined in (S.17). As we have shown in the proof of Theorem 1 that $\sum_{i=1}^n E|X_i|^{2+c} \rightarrow 0$ if $0 \leq \alpha < 1/2$, we immediately obtain that $n^{-\frac{2+c}{2}} \sum_{i=1}^n E|Y_{1,i}|^{2+c} \rightarrow 0$ if $0 \leq \alpha < 1/2$. Thus, overall we have $\sum_{i=1}^n E|Y_i|^{2+c} \rightarrow 0$ if $0 \leq \alpha < 1/2$, and this completes the proof of condition (b).

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

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

where H_s , $s = 1, 2, 3$, are given by (S.22)–(S.24), respectively, and $H_4 = n^{-1} \sum_{i=1}^n \eta_i \zeta_{i-1} \sigma_i^2$. Since $E(H_4) = 0$ and

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

we have $H_4 \rightarrow_p 0$. As it has been shown in the proof of Theorem 1 that $H_s \rightarrow_p 0$, for $s = 1, 2, 3$, if $0 \leq \alpha < 1$, overall we conclude that $\sum_{i=1}^n E(Y_i^2 | \mathfrak{F}_{i-1}^y) \rightarrow_p 1$ if $0 \leq \alpha < 1$. Combining conditions (a)–(c), Lemma S.10 is applicable and the result in (16) of the paper is established under $0 \leq \alpha < 1/2$. ■

Proof of Proposition 1. Let us first consider the estimator defined by (26) in the paper using a single quadratic moment. We can rewrite $\boldsymbol{\varepsilon}(\rho) = \mathbf{y} - \rho \mathbf{y}^*$ as

$$\boldsymbol{\varepsilon}(\rho) = \boldsymbol{\varepsilon} - (\rho - \rho_0) \mathbf{G}_0 \boldsymbol{\varepsilon}. \quad (\text{S.25})$$

Substituting (S.25) into $g_n(\rho) = n^{-1} \boldsymbol{\varepsilon}'(\rho) \mathbf{C} \boldsymbol{\varepsilon}(\rho)$ yields

$$\begin{aligned} g_n(\rho) &= n^{-1} [\boldsymbol{\varepsilon} - (\rho - \rho_0) \mathbf{G}_0 \boldsymbol{\varepsilon}]' \mathbf{C} [\boldsymbol{\varepsilon} - (\rho - \rho_0) \mathbf{G}_0 \boldsymbol{\varepsilon}] \\ &= \frac{\boldsymbol{\varepsilon}' \mathbf{C} \boldsymbol{\varepsilon}}{n} + (\rho - \rho_0)^2 \boldsymbol{\varepsilon}' \left(\frac{\mathbf{G}_0' \mathbf{C} \mathbf{G}_0}{n} \right) \boldsymbol{\varepsilon} - 2(\rho - \rho_0) \boldsymbol{\varepsilon}' \left(\frac{\mathbf{G}_0' \mathbf{C}}{n} \right) \boldsymbol{\varepsilon}. \end{aligned} \quad (\text{S.26})$$

Since $\text{diag}(\mathbf{C}) = \text{diag}(\mathbf{B}) = \mathbf{0}$ under Assumption 7, we have $E_0(\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}) = \text{Tr}(\mathbf{C}\boldsymbol{\Sigma}_0) = 0$. Using the results in Lemma S.7(i)–(iii), we obtain

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

where $a_0 = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0 \mathbf{C} \mathbf{G}_0 \boldsymbol{\Sigma}_0)$ and $b_0 = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0 \mathbf{C} \boldsymbol{\Sigma}_0)$. Note that $g_n(\rho_0) = n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\boldsymbol{\varepsilon}$. Using (S.26), it follows that

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

Since $\tilde{\rho}$ is such that $g_n(\tilde{\rho}) \leq g_n(\rho_0)$, or equivalently $(\rho - \rho_0)^2 a_0 - 2(\rho - \rho_0) b_0 \leq 0$, 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 (27) 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 $\ell = 1, 2, \dots, L$, where $a_{\ell 0} = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0 \mathbf{C}_\ell \mathbf{G}_0 \boldsymbol{\Sigma}_0)$ and $b_{\ell 0} = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0 \mathbf{C}_\ell \boldsymbol{\Sigma}_0)$. 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 $\ell = 1, 2, \dots, L$. ■

Proof of Theorem 3. Consider $\boldsymbol{\varepsilon}(\boldsymbol{\psi})$ given by (19) in the paper. It can be rewritten as

$$\boldsymbol{\varepsilon}(\boldsymbol{\psi}) = \boldsymbol{\varepsilon} - (\rho - \rho_0)\mathbf{G}_0\boldsymbol{\varepsilon} - \mathbf{Q}_0(\boldsymbol{\psi} - \boldsymbol{\psi}_0), \quad (\text{S.27})$$

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 (8) in the paper. Substituting (S.27) into the quadratic term in (24) of the paper and reorganizing yields

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

Using the results in Lemma S.7 and Assumption 6(d), the above equation becomes

$$\begin{aligned} n^{-1}\boldsymbol{\varepsilon}'(\boldsymbol{\psi})\mathbf{C}\boldsymbol{\varepsilon}(\boldsymbol{\psi}) &= (\rho - \rho_0)^2 (a_0 + c_0) - 2(\rho - \rho_0) b_0 + 2(\rho - \rho_0) \mathbf{d}'_0 (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ &\quad + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_{xcx} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) + o_p(1), \end{aligned} \quad (\text{S.28})$$

if $\delta < 1$, where $a_0 = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0 \mathbf{C} \mathbf{G}_0 \boldsymbol{\Sigma}_0)$, $b_0 = \lim_{n \rightarrow \infty} \text{Tr}(n^{-1}\mathbf{G}'_0 \mathbf{C} \boldsymbol{\Sigma}_0)$, $c_0 = p \lim_{n \rightarrow \infty} n^{-1}\boldsymbol{\eta}'_0\mathbf{C}\boldsymbol{\eta}_0$, $\mathbf{d}'_0 = p \lim_{n \rightarrow \infty} n^{-1}\boldsymbol{\eta}'_0\mathbf{C}\mathbf{X}$, and $\boldsymbol{\Sigma}_{xcx} = p \lim_{n \rightarrow \infty} n^{-1}\mathbf{X}'\mathbf{C}\mathbf{X}$. Substituting (S.27) into the linear term

in (24) in the paper yields

$$\begin{aligned} n^{-1}\mathbf{Z}'\boldsymbol{\varepsilon}(\boldsymbol{\psi}) &= n^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} - (\rho - \rho_0)n^{-1}\mathbf{Z}'\mathbf{G}_0\boldsymbol{\varepsilon} - n^{-1}\mathbf{Z}'\mathbf{Q}_0(\boldsymbol{\psi} - \boldsymbol{\psi}_0) \\ &= \boldsymbol{\Sigma}_{zq,0}(\boldsymbol{\psi} - \boldsymbol{\psi}_0) + o_p(1), \end{aligned} \quad (\text{S.29})$$

where $\boldsymbol{\Sigma}_{zq,0} = p \lim_{n \rightarrow \infty} n^{-1}\mathbf{Z}'\mathbf{Q}_0$ and $n^{-1}\mathbf{Z}'\boldsymbol{\varepsilon} = o_p(1)$ readily follow Assumption 6. 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 $\mathbf{z}_{\cdot l} = (z_{1l}, z_{2l}, \dots, z_{nl})'$ denote the l^{th} column of \mathbf{Z} , for $l = 1, 2, \dots, r$. Then

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

where $\mathbf{M} = (m_{ij}) = E(\mathbf{z}_{\cdot l}\mathbf{z}_{\cdot l}')$. Since $\sup_{i,j}|m_{ij}| = \sup_{i,j}|E(z_{il}z_{jl})| < K$ under Assumption 6, using Lemma S.4(iii) and Lemma S.6(ii) yields $\text{Tr}[(\mathbf{G}_0\boldsymbol{\Sigma}_0)\mathbf{G}_0'\mathbf{M}] = O(n^{\delta+1})$ and then $\text{Var}(n^{-1}\mathbf{z}_{\cdot l}'\mathbf{G}_0\boldsymbol{\varepsilon}) = O(n^{\delta-1})$ for $l = 1, 2, \dots, r$. 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 (S.28) and (S.29), we obtain

$$\mathbf{A}_n\mathbf{g}_n(\boldsymbol{\psi}) = \mathbf{A} \left[\begin{array}{c} (\rho - \rho_0)^2(a_0 + c_0) - 2(\rho - \rho_0)b_0 + 2(\rho - \rho_0)\mathbf{d}_0'(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \boldsymbol{\Sigma}_{xcx}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ \boldsymbol{\Sigma}_{zq,0}(\boldsymbol{\psi} - \boldsymbol{\psi}_0) \end{array} \right] + o_p(1),$$

or alternatively, $\mathbf{A}_n\mathbf{g}_n(\boldsymbol{\psi}) - \mathbf{A}E_0[\mathbf{g}_n(\boldsymbol{\psi})] = o_p(1)$. Under Assumption 6, $\boldsymbol{\Sigma}_{zq,0}$ has full column rank, then $\boldsymbol{\Sigma}_{zq,0}(\boldsymbol{\psi} - \boldsymbol{\psi}_0) = 0$ if and only if $\boldsymbol{\psi} = \boldsymbol{\psi}_0$. Hence, global identification is ensured without the quadratic moment. Moreover, it is readily seen that $\mathbf{g}_n(\boldsymbol{\psi})$ converges in probability uniformly in $\boldsymbol{\psi} \in \boldsymbol{\Psi}$ since $\boldsymbol{\Psi}$ is compact and $\mathbf{g}_n(\boldsymbol{\psi})$ is a continuous function. Thus, consistency of $\tilde{\boldsymbol{\psi}}$ can be established.

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

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

where $\bar{\boldsymbol{\psi}}$ lies element by element between $\boldsymbol{\psi}_0$ and $\tilde{\boldsymbol{\psi}}$. Note that $\frac{\partial \mathbf{g}_n(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}'} = -n^{-1}[2\mathbf{C}\boldsymbol{\varepsilon}(\boldsymbol{\psi}), \mathbf{Z}]'(\mathbf{y}^*, \mathbf{X})$, 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 (S.27), Lemma S.7, and Assumption 6(d) yields

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

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

$$= n^{-1} \text{Tr}(\mathbf{C}\mathbf{G}_0\boldsymbol{\Sigma}_0) - n^{-1}(\rho - \rho_0) \text{Tr}(\mathbf{G}'_0\mathbf{C}\mathbf{G}_0\boldsymbol{\Sigma}_0) + o_p(1),$$

if $\delta < 1$, and consequently

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

uniformly in $\boldsymbol{\psi} \in \Psi$. At $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, we have $\boldsymbol{\varepsilon}(\boldsymbol{\psi}_0) = \boldsymbol{\varepsilon}$, and it follows that $n^{-1}\boldsymbol{\varepsilon}'\mathbf{C}\mathbf{y}^* = n^{-1}\text{Tr}(\mathbf{C}\mathbf{G}_0\boldsymbol{\Sigma}_0) + o_p(1)$, 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, $\partial\mathbf{g}_n(\bar{\boldsymbol{\psi}})/\partial\boldsymbol{\psi}' = -\mathbf{D} + o_p(1)$, where \mathbf{D} is given by (28) in the paper. Moreover, by Theorem 2 in the paper we have $\mathbf{V}_g^{-1/2}\sqrt{n}\mathbf{g}_n(\boldsymbol{\psi}_0) \rightarrow_d N(0, \mathbf{I}_{k+1})$ if $\delta < 1/2$, where \mathbf{V}_g is given by (28). Hence, the asymptotic distribution of $\sqrt{n}(\tilde{\boldsymbol{\psi}} - \boldsymbol{\psi}_0)$ is as stated in Theorem 3.

■

Proof of (39): The asymptotic property of the BMM estimator under homoskedastic errors. To establish consistency and asymptotic distribution of the BMM estimators assuming the errors are homoskedastic, 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

$$\mathbf{M}_x(\mathbf{y} - \hat{\rho}\mathbf{y}^*) = -(\hat{\rho} - \rho_0)\mathbf{M}_x\mathbf{y}^* + \mathbf{M}_x\boldsymbol{\varepsilon}, \quad (\text{S.30})$$

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

$$n^{-1}(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \hat{\rho}\mathbf{y}^* - \mathbf{X}\hat{\boldsymbol{\beta}}) = n^{-1}(\mathbf{y} - \hat{\rho}\mathbf{y}^*)'\mathbf{M}_x(\mathbf{y} - \hat{\rho}\mathbf{y}^*) = \hat{\sigma}^2. \quad (\text{S.31})$$

Using the above results, the estimating equations (35)–(37) in the paper can now be written as

$$(n^{-1}\mathbf{y}^{*\prime}\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (n^{-1}\mathbf{y}^{*\prime}\mathbf{y}^*)(\hat{\rho} - \rho_0) = n^{-1}\mathbf{y}^{*\prime}\boldsymbol{\varepsilon} - \hat{\sigma}^2\text{Tr}[n^{-1}\mathbf{G}(\hat{\rho})], \quad (\text{S.32})$$

$$(n^{-1}\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (n^{-1}\mathbf{X}'\mathbf{y}^*)(\hat{\rho} - \rho_0) = n^{-1}\mathbf{X}'\boldsymbol{\varepsilon}, \quad (\text{S.33})$$

and

$$\hat{\sigma}^2 - \sigma_0^2 = [(n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\boldsymbol{\varepsilon}) - \sigma_0^2] - 2(\hat{\rho} - \rho_0)(n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\boldsymbol{\varepsilon}) + (\hat{\rho} - \rho_0)^2(n^{-1}\mathbf{y}^{*\prime}\mathbf{M}_x\mathbf{y}^*). \quad (\text{S.34})$$

Noting that $\mathbf{y}^* = \boldsymbol{\eta}_0 + \mathbf{G}_0\boldsymbol{\varepsilon}$, where $\boldsymbol{\eta}_0$ is given by (8) in the paper, we obtain

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

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

$$\begin{aligned}\hat{\sigma}^2 \operatorname{Tr}(n^{-1} \hat{\mathbf{G}}) - \sigma_0^2 \operatorname{Tr}(n^{-1} \mathbf{G}_0) &= (\hat{\sigma}^2 - \sigma_0^2) \operatorname{Tr}(n^{-1} \mathbf{G}_0) + \sigma_0^2 \left[\operatorname{Tr}(n^{-1} \hat{\mathbf{G}}) - \operatorname{Tr}(n^{-1} \mathbf{G}_0) \right] \\ &\quad + (\hat{\sigma}^2 - \sigma_0^2) \left[\operatorname{Tr}(n^{-1} \hat{\mathbf{G}}) - \operatorname{Tr}(n^{-1} \mathbf{G}_0) \right].\end{aligned}\quad (\text{S.35})$$

But

$$\begin{aligned}\hat{\mathbf{G}} - \mathbf{G}_0 &= \mathbf{W} (\mathbf{I}_n - \hat{\rho} \mathbf{W})^{-1} - \mathbf{W} (\mathbf{I}_n - \rho_0 \mathbf{W})^{-1} \\ &= \mathbf{W} (\mathbf{I}_n - \hat{\rho} \mathbf{W})^{-1} [(\mathbf{I}_n - \rho_0 \mathbf{W}) - (\mathbf{I}_n - \hat{\rho} \mathbf{W})] (\mathbf{I}_n - \rho_0 \mathbf{W})^{-1} \\ &= (\hat{\rho} - \rho_0) \mathbf{W} (\mathbf{I}_n - \hat{\rho} \mathbf{W})^{-1} \mathbf{W} (\mathbf{I}_n - \rho_0 \mathbf{W})^{-1} = (\hat{\rho} - \rho_0) \hat{\mathbf{G}} \mathbf{G}_0.\end{aligned}\quad (\text{S.36})$$

Hence, $\hat{\mathbf{G}} = \mathbf{G}_0 + (\hat{\rho} - \rho_0) \hat{\mathbf{G}} \mathbf{G}_0$, and using this result back in (S.36) now yields

$$\hat{\mathbf{G}} - \mathbf{G}_0 = (\hat{\rho} - \rho_0) \left[\mathbf{G}_0 + (\hat{\rho} - \rho_0) \hat{\mathbf{G}} \mathbf{G}_0 \right] \mathbf{G}_0 = (\hat{\rho} - \rho_0) \mathbf{G}_0^2 + \mathbf{R}_n(\hat{\rho}, \rho_0), \quad (\text{S.37})$$

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 S.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 $E |n^{-1} \operatorname{Tr}[\mathbf{R}_n(\hat{\rho}, \rho_0)]| \leq KE |\hat{\rho} - \rho_0|^2$, which establishes that

$$n^{-1} \operatorname{Tr}(\hat{\mathbf{G}} - \mathbf{G}_0) = (\hat{\rho} - \rho_0) \operatorname{Tr}(n^{-1} \mathbf{G}_0^2) + O_p[(\hat{\rho} - \rho_0)^2]. \quad (\text{S.38})$$

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

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

and hence

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

where notice that $O_p(n^{(\delta-1)/2}) = o_p(1)$ if $\delta < 1$. Using these results in (S.34) now yields

$$\begin{aligned}\hat{\sigma}^2 - \sigma_0^2 &= [(n^{-1} \boldsymbol{\varepsilon}' \mathbf{M}_x \boldsymbol{\varepsilon}) - \sigma_0^2] - 2(\hat{\rho} - \rho_0) \sigma_0^2 \operatorname{Tr}(n^{-1} \mathbf{G}_0 \mathbf{M}_x) \\ &\quad + O_p[(\hat{\rho} - \rho_0) n^{-1/2}] + O_p[(\hat{\rho} - \rho_0)^2].\end{aligned}\quad (\text{S.39})$$

Substituting (S.38) and (S.39) in (S.35) we have (noting that $\operatorname{Tr}(n^{-1} \mathbf{G}_0) < K$)

$$\begin{aligned}\hat{\sigma}^2 \operatorname{Tr}(n^{-1} \hat{\mathbf{G}}) - \sigma_0^2 \operatorname{Tr}(n^{-1} \mathbf{G}_0) \\ = \operatorname{Tr}(n^{-1} \mathbf{G}_0) [(n^{-1} \boldsymbol{\varepsilon}' \mathbf{M}_x \boldsymbol{\varepsilon}) - \sigma_0^2] - 2\sigma_0^2 (\hat{\rho} - \rho_0) \operatorname{Tr}(n^{-1} \mathbf{G}_0 \mathbf{M}_x) \operatorname{Tr}(n^{-1} \mathbf{G}_0)\end{aligned}$$

$$+ \sigma_0^2 (\hat{\rho} - \rho_0) \operatorname{Tr} (n^{-1} \mathbf{G}_0^2) + O_p [(\hat{\rho} - \rho_0)^2] + O_p [(\hat{\rho} - \rho_0) n^{-1/2}] . \quad (\text{S.40})$$

Using (S.40) in (S.32) and rearranging gives

$$(n^{-1} \mathbf{y}^{*\prime} \mathbf{X}) (\hat{\beta} - \boldsymbol{\beta}_0) + h_{n,\rho\rho} (\hat{\rho} - \rho_0) = h_{n,\rho\varepsilon} + O_p [(\hat{\rho} - \rho_0)^2] + O_p [(\hat{\rho} - \rho_0) n^{-1/2}] , \quad (\text{S.41})$$

where

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

Combining (S.41) and (S.33) we have

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

It is also easily seen that

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

Notice that

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

Under Assumption 3, we have

$$p \lim_{n \rightarrow \infty} n^{-1} \operatorname{Tr} (\mathbf{G}_0 \mathbf{M}_x) = \lim_{n \rightarrow \infty} n^{-1} \operatorname{Tr} (\mathbf{G}_0) - \lim_{n \rightarrow \infty} n^{-1} \operatorname{Tr} (\boldsymbol{\Sigma}_{xx} \boldsymbol{\Sigma}_{xg_0x}) = \lim_{n \rightarrow \infty} n^{-1} \operatorname{Tr} (\mathbf{G}_0) .$$

Hence, using results in Lemmas S.8 and S.9 we have $p \lim_{n \rightarrow \infty} h_{n,\rho\rho} = \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0g_0x} \boldsymbol{\beta}_0 + \sigma_0^2 h_0$, where h_0 is given by (43) in the paper; $p \lim_{n \rightarrow \infty} h_{n,\rho\varepsilon} = 0$; $p \lim_{n \rightarrow \infty} \frac{\mathbf{X}' \boldsymbol{\varepsilon}}{n} = \mathbf{0}$; $p \lim_{n \rightarrow \infty} \frac{\mathbf{y}^{*\prime} \mathbf{X}}{n} = \boldsymbol{\beta}'_0 \boldsymbol{\Sigma}_{xg_0x}$; and $p \lim_{n \rightarrow \infty} \frac{\mathbf{X}' \mathbf{X}}{n} = \boldsymbol{\Sigma}_{xx}$. 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

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

and

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

Consider now

$$\sqrt{n}h_{n,\rho\varepsilon} = \frac{\boldsymbol{\eta}'_0\varepsilon}{\sqrt{n}} + \frac{\varepsilon'\boldsymbol{\Pi}\varepsilon}{\sqrt{n}}, \quad (\text{S.42})$$

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 (S.42) conditional on \mathbf{X} . By Lemma S.6(ii), \mathbf{G}_0 satisfies the conditions in (10) and (11) in the paper. Since \mathbf{M}_x is an idempotent matrix, $\boldsymbol{\Pi}$ also satisfies (10) and (11). Therefore, applying Theorem 2 in the paper leads to

$$\frac{1}{\sqrt{n}} [\boldsymbol{\eta}'_0\varepsilon + \varepsilon'\boldsymbol{\Pi}\varepsilon - \sigma_0^2 \text{Tr}(\boldsymbol{\Pi})] \rightarrow_d N(0, q_0^2),$$

where q_0^2 is given by (41) in the paper. Notice that

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

and it follows that $p \lim_{n \rightarrow \infty} [n^{-1/2} \sigma_0^2 \text{Tr}(\boldsymbol{\Pi})] = 0$. Hence, by the Slutsky's theorem we obtain $n^{-1/2}(\boldsymbol{\eta}'_0\varepsilon + \varepsilon'\boldsymbol{\Pi}\varepsilon) \rightarrow_d N(0, q_0^2)$. In addition, it is readily seen that $n^{-1/2}\mathbf{X}'\varepsilon \rightarrow_d N(0, \sigma_0^2 \boldsymbol{\Sigma}_{xx})$. Thus, the asymptotic distribution given by (39) in the paper is established. ■

Proof of Theorem 4. When the errors are heteroskedastic, (S.31) need to be updated to

$$n^{-1} (\mathbf{y} - \hat{\mathbf{y}}^*)' \check{\mathbf{G}}(\rho) (\mathbf{y} - \hat{\mathbf{y}}^*) = n^{-1} (\mathbf{y} - \hat{\mathbf{y}}^*)' \mathbf{M}_x \check{\mathbf{G}}(\rho) \mathbf{M}_x (\mathbf{y} - \hat{\mathbf{y}}^*).$$

Then the estimating equations (46)–(47) in the main paper can be written as

$$(n^{-1}\mathbf{y}^{*'}\mathbf{X}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (n^{-1}\mathbf{y}^{*'}\mathbf{y}^*) (\hat{\rho} - \rho_0) = n^{-1}\mathbf{y}^{*'}\varepsilon - n^{-1}(\mathbf{y} - \hat{\mathbf{y}}^*)' \mathbf{M}_x \check{\mathbf{G}}(\rho) \mathbf{M}_x (\mathbf{y} - \hat{\mathbf{y}}^*), \quad (\text{S.43})$$

$$(n^{-1}\mathbf{X}'\mathbf{X}) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) + (n^{-1}\mathbf{X}'\mathbf{y}^*) (\hat{\rho} - \rho_0) = n^{-1}\mathbf{X}'\varepsilon, \quad (\text{S.44})$$

Using (S.30) we obtain

$$\begin{aligned} n^{-1} (\mathbf{y} - \hat{\mathbf{y}}^*)' \mathbf{M}_x \check{\mathbf{G}}(\rho) \mathbf{M}_x (\mathbf{y} - \hat{\mathbf{y}}^*) &= n^{-1} \varepsilon' \mathbf{M}_x \check{\mathbf{G}}(\rho) \mathbf{M}_x \varepsilon - 2(\hat{\rho} - \rho_0) [n^{-1} \mathbf{y}^{*'} \mathbf{M}_x \check{\mathbf{G}}(\rho) \mathbf{M}_x \varepsilon] \\ &\quad + (\hat{\rho} - \rho_0)^2 [n^{-1} \mathbf{y}^{*'} \mathbf{M}_x \check{\mathbf{G}}(\rho) \mathbf{M}_x \mathbf{y}^*] \\ &\equiv h_1 - 2(\hat{\rho} - \rho_0)h_2 + (\hat{\rho} - \rho_0)^2 h_3, \end{aligned} \quad (\text{S.45})$$

Also from (S.37) we have

$$\check{\mathbf{G}}(\rho) - \check{\mathbf{G}}_0 = (\hat{\rho} - \rho_0) \text{Diag}(\mathbf{G}_0^2) + O_p[(\hat{\rho} - \rho_0)^2]. \quad (\text{S.46})$$

We will consider h_1 , h_2 and h_3 in turn. Using (S.46) we obtain

$$\begin{aligned} h_1 &\equiv n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\check{\mathbf{G}}(\rho)\mathbf{M}_x\boldsymbol{\varepsilon} \\ &= n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} + (\hat{\rho} - \rho_0)n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\text{Diag}(\mathbf{G}_0^2)\mathbf{M}_x\boldsymbol{\varepsilon} \\ &\quad + (n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\boldsymbol{\varepsilon})O_p[(\hat{\rho} - \rho_0)^2]. \end{aligned}$$

Since $n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\boldsymbol{\varepsilon} \leq n^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = O_p(n^{-1/2})$, and $n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\text{Diag}(\mathbf{G}_0^2)\mathbf{M}_x\boldsymbol{\varepsilon} = \text{Tr}[n^{-1}\text{Diag}(\mathbf{G}_0^2)\boldsymbol{\Sigma}_0] + O_p(n^{-1/2})$ by Lemma S.9, we then have

$$h_1 = n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} + (\hat{\rho} - \rho_0)\text{Tr}[n^{-1}\text{Diag}(\mathbf{G}_0^2)\boldsymbol{\Sigma}_0] + O_p[(\hat{\rho} - \rho_0)n^{-1/2}]. \quad (\text{S.47})$$

Now consider h_2 . Using (S.46) and noting that $\mathbf{y}^* = \boldsymbol{\eta}_0 + \mathbf{G}_0\boldsymbol{\varepsilon}$, where $\boldsymbol{\eta}_0 = \mathbf{G}_0\mathbf{X}\boldsymbol{\beta}_0$, we have

$$\begin{aligned} h_2 &\equiv n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\check{\mathbf{G}}(\rho)\mathbf{M}_x\boldsymbol{\varepsilon} \\ &= n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} + 2(\hat{\rho} - \rho_0)[n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\text{Diag}(\mathbf{G}_0^2)\mathbf{M}_x\boldsymbol{\varepsilon}] \\ &\quad + (n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\boldsymbol{\varepsilon})O_p[(\hat{\rho} - \rho_0)^2] \\ &= n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} + O_p(\hat{\rho} - \rho_0) \\ &= n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'_0\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} + n^{-1}\boldsymbol{\eta}'_0\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} + O_p(\hat{\rho} - \rho_0). \end{aligned}$$

By Lemma S.9, $n^{-1}\boldsymbol{\eta}'_0\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} = O_p(n^{-1/2})$ and $n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'_0\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} = \text{Tr}[n^{-1}\mathbf{G}'_0\check{\mathbf{G}}_0\boldsymbol{\Sigma}_0] + O_p(n^{-1/2})$, we then obtain

$$-2(\hat{\rho} - \rho_0)h_2 = -2(\hat{\rho} - \rho_0)\text{Tr}[n^{-1}\mathbf{G}'_0\check{\mathbf{G}}_0\boldsymbol{\Sigma}_0] + O_p[(\hat{\rho} - \rho_0)n^{-1/2}] + O_p[(\hat{\rho} - \rho_0)^2]. \quad (\text{S.48})$$

Lastly,

$$\begin{aligned} (\hat{\rho} - \rho_0)^2 h_3 &\equiv (\hat{\rho} - \rho_0)^2 n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\check{\mathbf{G}}(\rho)\mathbf{M}_x\mathbf{y}^* \\ &= (\hat{\rho} - \rho_0)^2 n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\mathbf{y}^* + o_p[(\hat{\rho} - \rho_0)^2], \end{aligned}$$

where

$$n^{-1}\mathbf{y}^{*'}\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\mathbf{y}^* = n^{-1}\boldsymbol{\eta}'_0\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\eta}_0 + n^{-1}\boldsymbol{\varepsilon}'\mathbf{G}'_0\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\mathbf{G}_0\boldsymbol{\varepsilon} + 2n^{-1}\boldsymbol{\eta}'_0\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\mathbf{G}_0\boldsymbol{\varepsilon} = O_p(1).$$

It follows that

$$(\hat{\rho} - \rho_0)^2 h_3 = O_p[(\hat{\rho} - \rho_0)^2]. \quad (\text{S.49})$$

Then using (S.47), (S.48) and (S.49) in (S.45), we obtain

$$\begin{aligned} &n^{-1}(\mathbf{y} - \hat{\rho}\mathbf{y}^*)'\mathbf{M}_x\check{\mathbf{G}}(\rho)\mathbf{M}_x(\mathbf{y} - \hat{\rho}\mathbf{y}^*) \\ &= n^{-1}\boldsymbol{\varepsilon}'\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\boldsymbol{\varepsilon} + (\hat{\rho} - \rho_0)\{\text{Tr}[n^{-1}\text{Diag}(\mathbf{G}_0^2)\boldsymbol{\Sigma}_0] - 2\text{Tr}[n^{-1}\mathbf{G}'_0\check{\mathbf{G}}_0\boldsymbol{\Sigma}_0]\} \\ &\quad + O_p[(\hat{\rho} - \rho_0)^2] + O_p[(\hat{\rho} - \rho_0)n^{-1/2}] \end{aligned} \quad (\text{S.50})$$

Then substituting (S.50) in (S.43) yields

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

where

$$\begin{aligned} h_{n,\rho\varepsilon} &= n^{-1}\boldsymbol{\eta}'_0\varepsilon + n^{-1}\varepsilon'\left[\mathbf{G}'_0 - \mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\right]\varepsilon, \\ h_{n,\rho\rho} &= n^{-1}\mathbf{y}^{*\prime}\mathbf{y}^* + Tr[n^{-1}Diag(\mathbf{G}_0^2)\Sigma_0] - 2Tr[n^{-1}\mathbf{G}'_0\check{\mathbf{G}}_0\Sigma_0] \\ &= n^{-1}\boldsymbol{\eta}'_0\boldsymbol{\eta}_0 + n^{-1}\varepsilon'\mathbf{G}'_0\mathbf{G}_0\varepsilon + 2n^{-1}\varepsilon'\mathbf{G}'_0\boldsymbol{\eta}_0 + Tr[n^{-1}Diag(\mathbf{G}_0^2)\Sigma_0] - 2Tr[n^{-1}\mathbf{G}'_0\check{\mathbf{G}}_0\Sigma_0] \\ &= n^{-1}\boldsymbol{\eta}'_0\boldsymbol{\eta}_0 + Tr[\mathbf{G}'_0\mathbf{G}_0\Sigma_0] + Tr[n^{-1}Diag(\mathbf{G}_0^2)\Sigma_0] - 2Tr[n^{-1}\mathbf{G}'_0\check{\mathbf{G}}_0\Sigma_0] + O_p(n^{-1/2}). \end{aligned}$$

It follows that $p\lim_{n\rightarrow\infty}h_{n,\rho\rho} = \boldsymbol{\beta}'_0\Sigma_{xg_0g_0x}\boldsymbol{\beta}_0 + h_0$, where

$$h_0 = \lim_{n\rightarrow\infty}Tr[n^{-1}(\mathbf{G}_0^2 + \mathbf{G}'_0\mathbf{G}_0 - 2\check{\mathbf{G}}_0\mathbf{G}_0)\Sigma_0], \quad (\text{S.51})$$

The rest of the consistency proof runs as before under homoskedastic errors. Also note that

$$n^{-1}\varepsilon'\mathbf{M}_x\check{\mathbf{G}}_0\mathbf{M}_x\varepsilon = n^{-1}\varepsilon'\check{\mathbf{G}}_0\varepsilon - 2n^{-1}\varepsilon'\mathbf{P}_x\check{\mathbf{G}}_0\varepsilon + n^{-1}\varepsilon'\mathbf{P}_x\check{\mathbf{G}}_0\mathbf{P}_x\varepsilon,$$

where $\mathbf{P}_x = \mathbf{I}_n - \mathbf{M}_x = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Since

$$n^{-1}\varepsilon'\mathbf{P}_x\check{\mathbf{G}}_0\varepsilon = \frac{\varepsilon'\mathbf{X}}{n}\left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\frac{\mathbf{X}'\check{\mathbf{G}}_0\varepsilon}{n} = O_p(n^{-1/2})O_p(1)O_p(n^{-1/2}) = o_p(n^{-1/2}),$$

and similarly $n^{-1}\varepsilon'\mathbf{P}_x\check{\mathbf{G}}_0\mathbf{P}_x\varepsilon = o_p(n^{-1/2})$, we have

$$\sqrt{n}h_{n,\rho\varepsilon} = \frac{\boldsymbol{\eta}'_0\varepsilon}{\sqrt{n}} + \frac{\varepsilon'(\mathbf{G}'_0 - \check{\mathbf{G}}_0)\varepsilon}{\sqrt{n}} + o_p(1).$$

Applying Theorem 2 to the above linear-quadratic form and noticing $\frac{\mathbf{X}'\varepsilon}{\sqrt{n}} \rightarrow_d N(0, p\lim_{n\rightarrow\infty}\frac{1}{n}\mathbf{X}'\Sigma_0\mathbf{X})$ establishes the asymptotic distribution of the BMM estimator. ■

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

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

where $h_n = n^{-1} \text{Tr}(\mathbf{G}_0^2 + \mathbf{G}'_0 \mathbf{G}_0) - 2n^{-2} [\text{Tr}(\mathbf{G}_0)]^2$. Since $\Sigma_{xx} = p \lim_{n \rightarrow \infty} n^{-1} \mathbf{X}' \mathbf{X}$ and $\Sigma_{xg_0x} = p \lim_{n \rightarrow \infty} n^{-1} \mathbf{X}' \mathbf{G}_0 \mathbf{X}$ exist and they are k -dimensional square matrices (k is finite), it follows that

$$\begin{aligned} & p \lim_{n \rightarrow \infty} [\text{Tr}(n^{-1} \boldsymbol{\Pi}' \boldsymbol{\Pi}) + \text{Tr}(n^{-1} \boldsymbol{\Pi}^2)] \\ &= h_0 - 4 \lim_{n \rightarrow \infty} n^{-1} \text{Tr}(\Sigma_{xx} \Sigma_{xg_0x}) + 2 \lim_{n \rightarrow \infty} n^{-1} k [\text{Tr}(n^{-1} \mathbf{G}_0)]^2 = h_0, \end{aligned} \quad (\text{S.52})$$

where $h_0 = \lim_{n \rightarrow \infty} h_n$. 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 (S.52) leads to $\mathbf{V} = \sigma_0^2 \mathbf{H}$ and hence $\Omega_b = \sigma_0^2 \mathbf{H}^{-1}$, which is identical to the asymptotic variance of the best GMM estimator given by (4.5) of Lee (2007). ■

S2 Monte Carlo and empirical supplement

In what follows we provide additional Monte Carlo (MC) and empirical results in Sections S2.1 and S2.2, respectively. Section S2.1 aims to consider a wider class of data generating processes in the MC experiments. In particular, we consider spatial autoregressive (SAR) models without exogenous regressors, allowing for more than one dominant units, and including exponentially decaying degrees of dominance. The additional MC results also consider different spatial autoregressive processes for the regressors and the dependent variable. Finally, we provide MC results on the small sample performance of the maximum likelihood (ML) estimator of the SAR model with dominant units. Section S2.2 aims to provide evidence on the sensitivity of the empirical findings reported in the paper to the choice of the cut-off value, ϵ_w , in construction of the \mathbf{W} matrix.

S2.1 Additional Monte Carlo experiments

This Monte Carlo supplement provides additional simulation results. First we use the same MC designs as in the paper and examine the properties of additional GMM as well as ML estimators. We then consider other designs, including models without exogenous regressors, and models with more than one dominant units, and other choices of spatial weights matrices.

S2.1.1 Monte Carlo designs

Recall that the basic Data Generating Process (DGP) is given by (57) and (58) in the paper, which we reproduce here for convenience:

$$y_i = \alpha + \rho y_i^* + \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (\text{S.53})$$

$$x_i = \lambda x_i^* + \nu_i, \quad (\text{S.54})$$

where $y_i^* = \mathbf{w}'_{i,y} \mathbf{y}$, $\mathbf{y} = (y_1, y_2, \dots, y_n)'$, $x_i^* = \mathbf{w}'_{i,x} \mathbf{x}$, $\mathbf{x} = (x_1, x_2, \dots, x_n)'$, $\mathbf{w}'_{i,y}$ and $\mathbf{w}'_{i,x}$ are the i^{th} row of \mathbf{W}_y and \mathbf{W}_x , respectively. We consider both Gaussian and non-Gaussian error processes:

- Gaussian errors: $\varepsilon_i \sim IIDN(0, \sigma_i^2)$ and $\nu_i \sim IIDN(0, \sigma_{\nu,i}^2)$
- Non-Gaussian errors: $\varepsilon_i/\sigma_i \sim IID[\chi^2(2) - 2]/2$ and $\nu_i/\sigma_{\nu,i} \sim IID[\chi^2(2) - 2]/2$

The error variances are generated as $\sigma_i^2 = \sigma_\varepsilon^2 \tau_{\varepsilon,i}^2$ and $\sigma_{\nu,i}^2 = \sigma_v^2 \tau_{\nu,i}^2$. Then $Var(\boldsymbol{\varepsilon}) = \sigma_\varepsilon^2 \mathbf{D}_\varepsilon$, where $\mathbf{D}_\varepsilon = Diag(\tau_{\varepsilon,1}^2, \tau_{\varepsilon,2}^2, \dots, \tau_{\varepsilon,n}^2)$, and $Var(\boldsymbol{\nu}) = \sigma_v^2 \mathbf{D}_\nu$, where $\mathbf{D}_\nu = Diag(\tau_{\nu,1}^2, \tau_{\nu,2}^2, \dots, \tau_{\nu,n}^2)$. We consider both homoskedastic and heteroskedastic errors:

- Homoskedastic errors: $\tau_{\varepsilon,i}^2 = \tau_{\nu,i}^2 = 1$, for all i . Note that in this case $\sigma_i^2 = \sigma_\varepsilon^2$ and $\sigma_{\nu,i}^2 = \sigma_v^2$, for all i , and $\mathbf{D}_\varepsilon = \mathbf{D}_\nu = \mathbf{I}_n$.
- Heteroskedastic errors: $\tau_{\varepsilon,i}^2 \sim IIDU(0.5, 1.5)$ and $\tau_{\nu,i}^2 \sim IIDU(0.5, 1.5)$

The true parameter values are $\sigma_{\varepsilon,0}^2 = 1$, $\lambda_0 = 0.75$, $\alpha_0 = 1$, and $\sigma_{v,0}^2$ is set such that (62) in the paper holds. The number of replications is set to 2,000 for each experiment.

First, we assume that $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$, where \mathbf{W} is generated in the same way as in the main paper. We set $\beta_0 = 1$ and consider $\delta = 0, 0.25, 0.50, 0.75, 0.95, 1$, and $\rho_0 = 0.2, 0.5, 0.75$, for the sample sizes are $n = 100, 300, 500$, and $1,000$. The small sample properties of the BMM and the best GMM (BGMM) estimators for this design are already reported in the paper, except those for the experiments with heteroskedastic Gaussian errors. Here we report the results of the BMM and BGMM estimators under heteroskedastic Gaussian errors, and investigate the small sample performance of the ML and a few other GMM estimators. The standard ML estimator for SAR models without dominant units is described in detail in Anselin (1988, Chapter 6). Specifically, assume that the errors are homoskedastic and let $\boldsymbol{\psi} = (\rho, \alpha, \beta)'$ and $\boldsymbol{\theta} = (\boldsymbol{\psi}', \sigma_\varepsilon^2)'$ denote the parameters of model (S.53). Let $\boldsymbol{\theta}_0 = (\boldsymbol{\psi}'_0, \sigma_{\varepsilon,0}^2)' = (\rho_0, \alpha_0, \beta_0, \sigma_{\varepsilon,0}^2)'$ denote the true parameter values. The ML estimator of $\boldsymbol{\theta}_0$, denoted by $\hat{\boldsymbol{\theta}}$, is defined by

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} l_n(\boldsymbol{\theta}), \quad (S.55)$$

where the log-likelihood function, $l_n(\boldsymbol{\theta})$, is given by

$$l_n(\boldsymbol{\theta}) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma_\varepsilon^2 + \ln |\mathbf{I}_n - \rho \mathbf{W}_y| - \frac{1}{2\sigma_\varepsilon^2} \boldsymbol{\varepsilon}'(\boldsymbol{\psi}) \boldsymbol{\varepsilon}(\boldsymbol{\psi}), \quad (S.56)$$

$\boldsymbol{\varepsilon}(\boldsymbol{\psi}) = \mathbf{y} - \rho \mathbf{y}^* - \alpha \mathbf{1}_n - \beta \mathbf{x}$, $\mathbf{y}^* = \mathbf{W}_y \mathbf{y}$, and $\mathbf{1}_n$ is an $n \times 1$ vector of ones. Although, we have not investigated the properties of the ML estimator for SAR models with dominant units, we thought it is worthwhile to see how it performs in such cases.

We also consider two additional GMM estimators: a simple one that combines the moments using the identity matrix, and the optimal GMM (OGMM) estimator that uses the inverse of the covariance matrix to combine the moments. The moment conditions are given by (24) in the paper, where we set $\mathbf{B}_1 = \mathbf{W}$ and $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr(\mathbf{W}^2)\mathbf{I}_n$ under homoskedastic errors (or $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - Diag(\mathbf{W}^2)$ under heteroskedastic errors) in quadratic moments, and $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x})$ in linear moments. As explained in the paper, the BGMM estimator utilizes the (feasible) best \mathbf{B} matrix in the quadratic moment and the best instruments, $(\tilde{\mathbf{G}}\mathbf{x}\tilde{\alpha}, \tilde{\mathbf{G}}\mathbf{x}\tilde{\beta}, \mathbf{1}_n, \mathbf{x})$, in the linear moments, where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\rho})$ and $(\tilde{\rho}, \tilde{\alpha}, \tilde{\beta})'$ denote the first-step estimates. The BGMM estimator also uses the inverse of the covariance matrix to combine the moments. Recall that the (feasible) best \mathbf{B} matrix is given by $\tilde{\mathbf{G}} - n^{-1}Tr(\tilde{\mathbf{G}})\mathbf{I}_n$ if the errors are assumed to be homoskedastic, and $\tilde{\mathbf{G}} - Diag(\tilde{\mathbf{G}})$ if the errors are assumed to be homoskedastic. The results of the ML, GMM, and OGMM estimators of ρ_0 and β_0 are presented in Tables S.1–S.6 for homoskedastic Gaussian errors, and in Tables S.7–S.12 for homoskedastic non-Gaussian errors. The results of the BMM, GMM, OGMM, and BGMM estimators of ρ_0 and β_0 are summarized in Tables S.13–S.20 for heteroskedastic Gaussian errors, and the results of the GMM and OGMM estimators are given in Tables S.21–S.24 for heteroskedastic non-Gaussian errors (the results of the BMM and BGMM estimators in this case are presented in the main paper). In addition, Table S.25 reports the values of R_0^2 , defined by (59) in the paper assuming homoskedasticity, for various combinations of ρ_0 , δ , and n . Recall that the value of R_0^2 is used in (62) (given in the paper) in order to compute σ_v^2 .

Second, we consider the properties of the estimators of ρ_0 when $\beta_0 = 0$, namely a SAR model without any exogenous regressors, under homoskedastic errors. In this setup, we are able to consider an extremely high level of spatial autocorrelation: $\rho_0 = 0.95$. The results of the BMM, ML, GMM, OGMM, and BGMM estimators of ρ_0 for experiments with homoskedastic Gaussian errors are presented in Tables S.26–S.30, and for homoskedastic non-Gaussian errors are summarized in Tables S.31–S.35. Note that when $\beta_0 = 0$, the BMM estimator is computed using only the moment conditions (35) and (37) in the paper. The GMM and OGMM estimators are computed using two quadratic moment conditions with $\mathbf{B}_1 = \mathbf{W}$ and $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr(\mathbf{W}^2)\mathbf{I}_n$. The BGMM estimator is computed using $\tilde{\mathbf{G}} - n^{-1}Tr(\tilde{\mathbf{G}})\mathbf{I}_n$, where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\rho})$ and $\tilde{\rho}$ denotes the first-step GMM estimate.

Third, we consider SAR models with exogenous regressors (with $\beta_0 = 1$) for \mathbf{W} matrices with two dominant units under homoskedastic errors. Without loss of generality, we generate \mathbf{W} such that the sum of the first two columns rise with n at the same rate of δ . The non-zero elements of the first two columns are drawn from $IIDU(0, 1)$. The rest of the \mathbf{W} matrix is generated in the same way as in the paper. \mathbf{W} is row-standardized so that each row sums to unity. We focus on the BMM estimator of ρ_0 and β_0 in this case. The results are reported in Tables S.36 and S.37 for Gaussian errors, and Tables S.38 and S.39 for non-Gaussian errors.

Fourth, we examine \mathbf{W} matrices with exponentially decaying degrees of dominance. Specifically, we consider the basic design in the paper with homoskedastic errors except that the sum of the j^{th} column of \mathbf{W} rises with n at the rate of δ_j , where $\delta_j = 0.9^j$, for $j = 1, 2, \dots, n$. The non-zero elements of each column of \mathbf{W} are drawn from $IIDU(0, 1)$. Note that the order of the columns can be reshuffled and does not affect the results. According to our theory, only δ_{\max} , which equals 0.9 under this design, plays a role in the limiting distribution of the BMM and GMM estimators. We again focus on the BMM estimator of ρ and β . The estimation results are summarized in Table S.40.

Finally, we allow the spatial weights matrices in the y and x processes to be different. We assume that \mathbf{W}_x and \mathbf{W}_y follow the same structure but consider different values for their centrality, namely we set $\delta_x = 1$ and $\delta_y = 0, 0.25, 0.5, 0.75, 0.95, 1$. The results for the BMM estimators are displayed in Tables S.41 and S.42 for homoskedastic Gaussian errors, and Tables S.43 and S.44 for homoskedastic non-Gaussian errors.

S2.1.2 Summary of results

We first note that the performance of the ML estimator is similar to that of the BMM, and to a lesser extent, to that of the GMM estimators. Its bias and root mean square error (RMSE) falls with the sample size, for all values of ρ_0 and $\delta \leq 0.75$. It also has the correct size and reasonable power so long as $\delta \leq 0.75$.

Turning to the GMM and OGMM estimators, we observe that their small sample performance resembles that of the best GMM estimator. Perhaps not surprisingly, the GMM estimator has slightly larger RMSE than the OGMM estimator when n is small. The RMSE of the OGMM estimator is very close to that of the best GMM estimator, suggesting that the instruments and matrices of the quadratic moments used in the computation of the OGMM estimator approximate those of the best (infeasible) GMM estimator reasonably well.

These findings hold generally for the pure SAR model without any exogenous regressors. Again we obtain correct size and reasonable power so long as $\delta \leq 0.75$, irrespective of the values of ρ_0 (including $\rho_0 = 0.95$). All estimators tends to be quite robust to non-Gaussian errors.

When the weights matrix, \mathbf{W} , contains two dominant units or units whose degrees of dominance decay exponentially, the findings of the BMM estimator remain unaltered. These MC results are consistent with our theory, which shows that the limiting distribution of the BMM estimator is governed by the highest degree of dominance in the network. Furthermore, when the spatial weights matrices in the y and x processes are allowed to differ, the BMM estimator delivers robust performance and gives correct size and reasonable power if $\delta \leq 0.75$.

Lastly, the figures at the end of this supplement depict a set of empirical power functions for the BMM, best GMM, and ML estimators. In all figures we label the best GMM estimator

as GMM for convenience. Specifically, we plotted the power functions for ρ in the case of $\rho_0 = 0.2, 0.5, 0.75$ and homoskedastic Gaussian errors, when $\delta = 0, 0.25, 0.50, 0.75, 0.95, 1$, and $n = 100, 300, 500, 1,000$ under the same MC design as in the paper. In addition, to complement the power function plots in the main paper, we also plotted the empirical power functions for the BMM and best GMM estimators in the case of $\rho_0 = 0.5$ and heteroskedastic errors, when $\delta = 0, 0.25, 0.75, 0.95$, and $n = 100$ and 300 . It is readily seen that the power functions for all the three estimators are very close when $\delta \leq 0.75$, and the differences become hardly discernible as n rises above 500 . When $n = 100$ and $\delta \leq 0.75$, the tests based on the best GMM estimator slightly over-reject the null hypothesis, whereas tests based on the BMM and ML estimators have empirical sizes close to the 5% nominal level. When $\delta = 0.95$ or 1 , it is evident that all estimators fail to produce the desirable power functions. There is evidence of substantial size distortions, and the shape of the power functions strongly suggest that the asymptotic normal theory is no longer applicable when $\delta \geq 0.95$. Among the three estimators considered, the GMM estimator displays the highest degree of size distortions when δ is close to unity. These findings hold similarly for both homoskedastic and heteroskedastic errors.

Table S.1: Small sample properties of the ML estimator of ρ for the experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)					
	100	300	500	1,000	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$															
0.00	-9.14	-2.76	-1.71	-0.86	20.81	10.43	7.86	5.39	6.85	5.45	5.60	5.20	7.05	14.60	22.05
0.25	-9.05	-2.76	-1.71	-0.85	20.71	10.42	7.85	5.39	7.15	5.20	5.65	5.35	7.00	15.00	22.05
0.50	-9.13	-2.75	-1.69	-0.85	20.58	10.35	7.84	5.40	6.60	5.55	5.95	5.30	7.00	15.15	22.35
0.75	-9.51	-2.85	-1.71	-0.85	21.25	10.71	8.08	5.49	7.50	5.70	5.50	7.10	14.70	21.95	42.90
0.95	-11.65	-3.60	-2.22	-1.16	25.40	12.90	9.68	6.59	12.45	10.70	9.60	10.05	11.70	17.60	24.05
1.00	-13.18	-4.30	-2.72	-1.33	28.55	15.19	11.56	7.85	16.25	15.95	15.00	13.65	15.10	20.75	25.90
$\rho_0 = 0.5$															
0.00	-8.16	-2.45	-1.47	-0.70	16.74	7.95	5.86	3.97	6.90	5.75	5.85	5.10	9.30	21.10	38.10
0.25	-8.11	-2.45	-1.45	-0.72	16.67	7.94	5.85	3.99	6.85	5.80	6.10	5.20	9.25	21.55	38.40
0.50	-8.14	-2.44	-1.44	-0.71	16.59	7.89	5.85	4.00	6.65	5.50	5.95	5.35	9.45	21.70	38.75
0.75	-8.85	-2.67	-1.53	-0.77	17.57	8.37	6.15	4.13	7.40	6.35	5.70	5.50	9.25	21.35	38.25
0.95	-12.41	-3.86	-2.24	-1.13	23.04	10.86	7.91	5.28	14.60	12.25	11.10	11.15	13.50	23.20	35.90
1.00	-14.86	-4.82	-2.88	-1.32	26.78	13.53	10.11	6.78	20.15	20.00	20.40	19.40	17.95	27.10	37.20
$\rho_0 = 0.75$															
0.00	-6.02	-1.75	-1.07	-0.64	11.31	4.98	3.58	2.46	7.00	6.65	5.65	5.55	17.90	50.60	73.15
0.25	-6.03	-1.77	-1.07	-0.64	11.33	4.98	3.60	2.47	6.80	6.15	6.05	5.95	17.90	50.05	73.75
0.50	-6.03	-1.76	-1.07	-0.63	11.27	4.95	3.59	2.48	6.90	7.00	6.05	5.90	17.80	50.35	73.30
0.75	-6.90	-2.03	-1.18	-0.69	12.38	5.43	3.86	2.58	8.20	7.80	6.25	5.90	17.35	47.60	71.45
0.95	-12.21	-3.74	-2.10	-1.10	19.51	8.07	5.61	3.57	19.05	16.05	14.85	11.50	18.15	38.60	58.90
1.00	-16.44	-5.59	-3.35	-1.64	24.50	11.48	8.39	5.63	28.75	32.45	32.90	33.50	22.55	38.85	50.35

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic Gaussian errors. $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. The first unit is δ -dominant, and the rest of the units are non-dominant. The maximum likelihood (ML) estimator is given by (S.55) and was computed using the spatial econometrics functions contained in the Econometrics Toolbox Version 7 by James P. LeSage (www.spatial-econometrics.com). The power is calculated at $\rho_0 - 0.1$, where ρ_0 denotes the true value.

Table S.2: Small sample properties of the ML estimator of β for the experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	300	500	100	300
$\rho_0 = 0.2$												
0.00	3.49	1.32	0.65	0.52	39.70	20.86	15.78	10.93	7.85	5.40	4.50	4.35
0.25	3.48	1.33	0.66	0.52	39.66	20.86	15.78	10.93	7.85	5.60	4.50	4.30
0.50	3.50	1.33	0.65	0.50	39.87	20.95	15.80	10.97	7.90	5.40	4.40	4.45
0.75	3.34	1.29	0.60	0.49	40.91	21.53	16.21	11.33	7.45	5.20	4.45	4.70
0.95	2.08	0.93	0.35	0.53	46.17	25.94	20.16	15.38	7.20	5.25	4.70	4.55
1.00	1.08	0.20	-0.17	0.27	50.79	30.60	25.36	20.69	7.00	5.10	4.90	4.10
$\rho_0 = 0.5$												
0.00	4.88	1.86	0.91	0.65	47.23	24.55	18.50	12.79	7.90	5.65	4.60	4.30
0.25	4.87	1.88	0.92	0.68	47.16	24.56	18.51	12.80	7.90	5.65	4.75	4.40
0.50	4.87	1.88	0.91	0.65	47.55	24.79	18.65	12.90	7.95	5.45	4.40	4.55
0.75	4.92	1.94	0.87	0.69	51.36	27.62	21.01	15.10	7.75	5.45	4.70	4.50
0.95	3.43	1.63	0.43	0.88	70.42	47.31	39.80	34.39	7.35	5.30	5.15	4.55
1.00	1.86	0.21	-0.72	0.45	85.35	64.33	59.21	54.42	6.90	5.25	5.30	4.25
$\rho_0 = 0.75$												
0.00	6.37	2.39	1.30	1.07	49.51	25.50	19.21	13.31	8.20	5.75	4.90	4.60
0.25	6.43	2.45	1.31	1.08	49.49	25.52	19.21	13.32	8.20	5.80	4.70	4.60
0.50	6.41	2.48	1.34	1.05	49.88	25.98	19.54	13.55	8.30	5.85	4.70	4.40
0.75	6.82	2.72	1.30	1.15	58.10	32.49	25.18	18.80	8.55	5.60	5.10	4.45
0.95	5.44	2.83	0.79	1.55	96.26	71.47	62.66	56.15	7.85	5.45	5.55	4.55
1.00	3.18	0.54	-1.11	0.80	120.66	101.22	96.37	91.46	7.25	5.40	5.35	4.30

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.1.

Table S.3: Small sample properties of the GMM estimator of ρ for the experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-11.99	-3.64	-2.14	-1.08	24.34	11.50	8.45	5.76	8.70	6.65	6.05	5.60
0.25	-11.91	-3.64	-2.14	-1.08	24.19	11.48	8.44	5.76	8.45	6.55	6.00	5.55
0.50	-12.00	-3.66	-2.14	-1.08	24.09	11.43	8.44	5.77	8.15	6.10	6.30	5.35
0.75	-12.51	-3.81	-2.17	-1.10	25.08	11.88	8.73	5.85	9.55	7.05	6.50	7.40
0.95	-13.99	-4.37	-2.62	-1.34	29.34	14.00	10.28	6.85	15.20	12.05	9.65	9.00
1.00	-14.94	-4.59	-2.79	-1.34	32.43	16.09	12.01	8.05	20.15	16.80	15.75	13.35
$\rho_0 = 0.5$												
0.00	-9.37	-2.78	-1.62	-0.81	18.77	8.52	6.21	4.21	8.60	6.65	6.05	5.55
0.25	-9.32	-2.78	-1.62	-0.81	18.66	8.51	6.20	4.21	8.25	6.60	6.05	5.55
0.50	-9.37	-2.79	-1.62	-0.81	18.55	8.48	6.20	4.22	8.40	6.15	6.00	5.30
0.75	-10.00	-2.98	-1.68	-0.85	19.76	8.95	6.50	4.32	9.80	7.20	6.40	6.25
0.95	-12.26	-3.74	-2.20	-1.10	25.21	11.41	8.21	5.38	17.95	13.70	11.40	10.85
1.00	-13.61	-4.14	-2.49	-1.19	28.80	13.92	10.30	6.86	25.25	22.25	20.55	19.75
$\rho_0 = 0.75$												
0.00	-6.21	-1.77	-1.02	-0.51	12.36	5.34	3.84	2.59	8.45	6.60	6.05	5.50
0.25	-6.18	-1.77	-1.02	-0.51	12.29	5.33	3.84	2.59	8.45	6.50	5.95	5.55
0.50	-6.19	-1.78	-1.02	-0.51	12.17	5.30	3.84	2.60	8.65	6.45	6.15	5.30
0.75	-6.83	-1.95	-1.09	-0.54	13.43	5.73	4.10	2.70	10.40	7.50	6.55	6.45
0.95	-9.75	-2.74	-1.56	-0.76	20.10	8.34	5.85	3.73	24.75	17.65	15.00	13.60
1.00	-11.92	-3.51	-2.07	-0.95	24.59	11.58	8.55	5.66	37.60	35.60	33.25	31.35

Notes: The GMM estimator is given by (23) in the paper, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr(\mathbf{W}^2)\mathbf{I}_n$, and $\mathbf{A}_n = \mathbf{I}_n$. See also the notes to Table S.1.

Table S.4: Small sample properties of the GMM estimator of β for the experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)			
	100	300	500	1,000	300	500	1,000	300	500	1,000	300	500	1,000
$\rho_0 = 0.2$													
0.00	7.26	2.80	1.45	0.92	44.18	22.68	17.12	11.88	8.45	6.05	5.45	5.50	12.85
0.25	7.25	2.83	1.46	0.92	44.10	22.70	17.12	11.89	8.20	5.90	5.75	5.50	12.75
0.50	7.27	2.85	1.46	0.89	44.42	22.80	17.19	11.94	8.90	5.70	5.55	5.30	12.40
0.75	6.98	2.82	1.42	0.91	45.31	23.49	17.72	12.38	8.00	5.70	5.10	5.05	12.75
0.95	3.81	1.97	0.99	0.89	48.26	26.98	21.17	16.04	7.80	5.35	5.05	4.95	10.40
1.00	2.04	0.56	0.05	0.39	51.82	30.82	25.58	20.87	7.10	5.45	5.10	4.05	9.35
$\rho_0 = 0.5$													
0.00	8.63	3.30	1.70	1.08	52.61	26.61	20.02	13.87	8.70	5.85	5.55	5.75	11.85
0.25	8.64	3.33	1.71	1.08	52.50	26.63	20.02	13.88	8.55	5.75	5.55	5.75	11.65
0.50	8.66	3.37	1.72	1.05	53.05	26.90	20.20	14.01	8.80	5.70	5.50	5.30	12.15
0.75	8.57	3.45	1.69	1.12	56.94	29.96	22.79	16.38	7.95	5.65	5.15	4.95	11.20
0.95	4.84	2.63	1.22	1.34	74.21	49.64	42.17	36.60	8.35	5.65	5.60	5.30	9.35
1.00	2.41	0.35	-0.53	0.80	88.00	66.47	62.71	59.94	7.55	5.85	6.55	6.30	8.60
$\rho_0 = 0.75$													
0.00	9.97	3.76	1.97	1.23	55.58	27.83	20.85	14.44	8.80	6.05	5.55	5.45	11.30
0.25	9.99	3.80	1.97	1.23	55.43	27.85	20.84	14.45	8.60	6.05	5.60	5.60	11.55
0.50	10.04	3.86	1.99	1.20	56.05	28.35	21.19	14.71	8.80	5.65	5.40	5.25	11.75
0.75	10.35	4.15	2.02	1.35	64.99	35.33	27.27	20.29	8.30	5.80	5.20	5.20	11.40
0.95	5.77	3.72	1.91	1.78	103.45	79.53	72.31	67.41	9.40	7.15	8.05	8.65	9.75
1.00	2.82	0.08	-1.45	1.69	127.48	114.85	121.56	136.01	8.25	8.20	9.30	13.20	8.85

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.3.

Table S.5: Small sample properties of the OGMM estimator of ρ for the experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)							
				100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300
	$\rho_0 = 0.2$	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
0.00	-3.63	-1.41	-0.97	-0.49	20.66	10.25	7.80	5.38	9.80	5.85	5.90	4.80	13.75	18.90	26.35	45.95	
0.25	-3.63	-1.41	-0.96	-0.49	20.55	10.23	7.79	5.38	9.75	5.70	5.70	5.05	13.95	18.95	26.25	45.85	
0.50	-3.78	-1.40	-0.94	-0.48	20.54	10.19	7.78	5.40	9.95	5.85	5.20	13.55	19.15	25.85	45.85		
0.75	-2.22	-1.04	-0.74	-0.41	21.71	10.74	8.13	5.52	11.35	7.35	6.90	5.30	17.25	20.85	28.05	45.80	
0.95	1.66	0.79	0.11	-0.11	27.65	14.22	10.29	6.90	23.05	16.75	12.85	12.10	27.80	30.35	33.60	45.60	
1.00	2.29	0.76	0.10	0.10	32.74	17.77	13.19	8.75	30.05	23.95	21.95	19.55	33.80	33.60	35.40	46.25	
<hr/>																	
0.00	-3.22	-1.29	-0.83	-0.44	16.21	7.71	5.78	3.98	10.45	6.00	5.65	5.20	19.00	29.60	42.25	68.75	
0.25	-3.24	-1.29	-0.83	-0.44	16.15	7.69	5.77	3.98	10.35	6.15	5.70	5.35	19.30	29.25	41.90	68.75	
0.50	-3.31	-1.29	-0.82	-0.44	16.22	7.66	5.76	3.99	10.60	6.15	6.15	5.30	19.10	29.40	42.60	68.95	
0.75	-1.98	-1.14	-0.72	-0.42	17.75	8.26	6.15	4.16	13.85	7.70	6.80	5.60	23.00	31.00	43.15	66.75	
0.95	1.73	0.68	0.10	-0.15	24.02	12.03	8.43	5.51	29.35	20.15	14.75	13.55	36.75	38.95	45.10	62.00	
1.00	2.13	0.82	0.21	0.11	28.89	16.01	11.85	7.74	39.00	31.80	28.95	25.30	43.30	42.80	45.75	58.30	
<hr/>																	
0.00	-2.31	-0.91	-0.55	-0.30	10.64	4.79	3.52	2.41	12.00	6.65	5.75	5.55	31.20	57.45	77.50	96.40	
0.25	-2.35	-0.91	-0.55	-0.30	10.61	4.78	3.51	2.41	11.95	6.65	5.60	5.25	30.35	57.20	77.35	96.35	
0.50	-2.37	-0.91	-0.55	-0.31	10.68	4.76	3.51	2.42	12.30	6.80	5.90	5.00	31.05	57.15	77.25	96.55	
0.75	-1.60	-0.92	-0.54	-0.32	11.92	5.30	3.85	2.59	17.30	8.45	7.05	6.15	34.85	55.65	75.10	94.80	
0.95	0.69	0.54	0.07	-0.15	17.71	8.84	5.99	3.75	38.25	26.30	19.80	16.60	48.85	55.15	67.35	86.85	
1.00	-0.09	0.46	0.17	0.09	22.22	12.94	9.83	6.48	53.10	47.00	43.90	40.20	54.70	56.10	63.00	76.10	

Notes: The OGMM estimator is given by (31) in the paper, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{Wx}, \mathbf{W}^2\mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - n^{-1}Tr(\mathbf{W}^2)\mathbf{I}_n$, and the weighting matrix \mathbf{A}_n is the inverse of the estimated covariance of moments. See also the notes to Table S.1.

Table S.6: Small sample properties of the OGMM estimator of β for the experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)			
	100	300	500	1,000	300	500	1,000	300	500	1,000	300	500	1,000
$\rho_0 = 0.2$													
0.00	-1.12	-0.22	-0.20	0.11	39.48	20.87	15.75	10.91	7.55	5.70	4.60	4.20	9.85
0.25	-1.10	-0.19	-0.20	0.11	39.50	20.86	15.74	10.92	7.45	5.95	4.65	4.40	9.80
0.50	-1.18	-0.23	-0.24	0.08	39.70	20.95	15.77	10.94	7.20	5.70	4.90	4.35	9.60
0.75	-1.81	-0.46	-0.40	0.02	40.63	21.55	16.21	11.33	7.25	5.65	4.50	4.70	9.35
0.95	-2.68	-1.34	-0.92	-0.10	45.78	26.04	20.19	15.43	6.95	5.35	4.95	4.70	8.10
1.00	-2.75	-1.36	-1.06	-0.20	50.03	30.51	25.27	20.66	6.35	5.20	4.70	4.35	7.55
$\rho_0 = 0.5$													
0.00	-1.16	-0.09	-0.17	0.19	46.40	24.43	18.43	12.78	7.35	5.90	4.70	4.05	8.80
0.25	-1.11	-0.05	-0.16	0.19	46.38	24.41	18.43	12.78	7.30	5.95	4.80	4.20	8.95
0.50	-1.24	-0.08	-0.19	0.16	46.78	24.65	18.56	12.88	7.15	5.70	4.85	4.15	8.75
0.75	-1.99	-0.32	-0.39	0.12	50.34	27.50	20.92	15.09	7.15	5.65	4.65	4.50	8.45
0.95	-3.40	-1.84	-1.34	-0.03	68.63	47.04	39.67	34.40	6.80	5.20	5.00	4.75	7.35
1.00	-3.47	-2.19	-1.97	-0.17	82.62	63.65	58.73	54.18	6.25	5.10	4.65	4.15	6.95
$\rho_0 = 0.75$													
0.00	-1.25	-0.02	-0.15	0.23	48.28	25.27	19.06	13.22	7.30	6.05	4.70	4.15	8.85
0.25	-1.17	0.03	-0.14	0.23	48.21	25.26	19.06	13.22	7.35	5.95	4.80	4.10	8.95
0.50	-1.32	0.01	-0.16	0.22	48.60	25.70	19.34	13.44	7.50	6.15	4.95	4.05	8.45
0.75	-2.08	-0.15	-0.39	0.22	56.29	32.22	24.95	18.68	7.45	5.60	4.85	4.25	7.95
0.95	-3.47	-2.31	-1.60	0.04	91.90	70.68	62.23	56.01	6.60	5.30	5.70	4.40	7.10
1.00	-3.35	-2.90	-2.95	-0.17	114.70	99.32	95.07	90.86	6.00	5.15	4.85	4.05	6.35

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.5.

Table S.7: Small sample properties of the ML estimator of ρ for the experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)						
	100	300	500	1,000	300	500	1,000	100	300	500	1,000	300	500	1,000		
$\rho_0 = 0.2$																
0.00	-8.25	-2.44	-1.51	-0.75	20.32	9.90	7.58	5.30	6.60	4.10	5.20	5.25	7.65	14.50	21.90	42.65
0.25	-8.17	-2.41	-1.50	-0.75	20.26	9.88	7.58	5.30	6.85	4.05	5.15	4.85	7.60	14.55	21.95	42.80
0.50	-8.37	-2.47	-1.52	-0.77	20.29	9.99	7.62	5.31	6.85	4.45	5.05	5.20	7.30	15.10	22.60	42.00
0.75	-8.76	-2.56	-1.65	-0.77	21.21	10.38	7.94	5.51	8.05	5.20	6.05	6.45	8.05	16.40	22.35	42.15
0.95	-10.91	-3.19	-2.02	-1.01	25.55	12.63	9.51	6.57	13.30	9.55	10.10	9.45	12.95	18.15	23.35	39.70
1.00	-12.03	-3.58	-2.27	-1.18	28.62	14.58	11.30	7.83	17.05	13.90	14.95	13.80	16.60	20.65	26.00	38.05
$\rho_0 = 0.5$																
0.00	-7.44	-2.17	-1.30	-0.59	16.43	7.50	5.66	3.89	6.75	4.40	5.45	5.05	9.05	20.90	38.90	66.95
0.25	-7.41	-2.17	-1.32	-0.64	16.43	7.52	5.67	3.92	6.75	4.65	5.50	5.60	9.10	21.05	38.75	66.90
0.50	-7.57	-2.20	-1.32	-0.63	16.47	7.58	5.68	3.92	6.95	4.55	5.35	5.20	9.15	21.80	37.85	66.70
0.75	-8.22	-2.40	-1.49	-0.69	17.65	8.08	6.09	4.14	8.25	6.20	6.85	5.85	9.95	22.15	37.05	63.65
0.95	-11.65	-3.48	-2.14	-1.01	23.01	10.65	7.88	5.30	15.00	12.05	12.10	10.55	14.25	22.80	34.55	55.00
1.00	-13.76	-4.09	-2.48	-1.16	26.69	12.82	9.83	6.74	20.35	18.55	20.15	20.05	18.70	26.60	36.15	52.40
$\rho_0 = 0.75$																
0.00	-5.63	-1.60	-1.00	-0.57	11.23	4.71	3.52	2.42	7.05	5.15	5.45	6.15	19.15	51.65	75.20	95.85
0.25	-5.63	-1.62	-0.99	-0.58	11.25	4.74	3.48	2.42	7.35	5.15	5.45	6.00	19.40	51.40	75.55	95.50
0.50	-5.71	-1.66	-1.01	-0.57	11.31	4.76	3.52	2.45	7.45	5.10	5.80	6.40	19.60	50.85	75.25	96.05
0.75	-6.54	-1.87	-1.23	-0.66	12.66	5.27	3.91	2.64	8.90	7.85	8.00	7.25	19.25	48.80	71.00	93.35
0.95	-11.65	-3.52	-2.14	-1.03	19.44	8.05	5.70	3.66	18.50	16.25	16.50	12.95	18.65	39.55	55.85	84.85
1.00	-15.54	-5.00	-3.02	-1.51	24.29	10.83	8.12	5.57	27.80	30.75	32.60	33.80	23.15	39.30	48.75	70.30

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic non-Gaussian errors. See also the notes to Table S.1.

Table S.8: Small sample properties of the ML estimator of β for the experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	300	500	1,000	300	500	1,000	300	500
$\rho_0 = 0.2$												
0.00	4.71	1.46	1.33	0.54	40.55	21.26	16.04	11.36	6.85	5.45	5.80	6.20
0.25	4.67	1.43	1.32	0.54	40.53	21.26	16.02	11.36	6.90	5.35	5.75	6.10
0.50	4.70	1.39	1.29	0.52	40.61	21.33	16.09	11.39	6.65	5.80	5.55	5.75
0.75	4.44	1.54	1.38	0.60	41.90	21.95	16.62	11.82	6.80	5.25	5.25	6.05
0.95	4.14	1.44	1.48	0.61	47.69	26.98	20.84	16.09	6.65	5.90	5.75	10.50
1.00	3.53	1.18	1.43	0.46	52.29	31.34	25.76	21.71	6.35	5.95	5.60	5.50
$\rho_0 = 0.5$												
0.00	6.33	1.98	1.69	0.65	48.29	25.05	18.81	13.29	6.85	5.45	5.85	5.95
0.25	6.29	1.94	1.70	0.69	48.24	25.03	18.79	13.28	6.85	5.50	5.90	5.95
0.50	6.36	1.90	1.66	0.65	48.51	25.23	18.96	13.38	6.70	5.75	5.60	5.75
0.75	6.30	2.22	1.88	0.84	52.65	28.19	21.56	15.78	7.15	5.30	5.40	6.30
0.95	6.58	2.53	2.70	1.19	72.68	49.26	41.17	36.07	7.10	5.75	6.00	5.90
1.00	5.99	2.33	3.13	1.00	87.79	65.90	60.19	57.21	6.35	6.10	5.60	5.85
$\rho_0 = 0.75$												
0.00	7.97	2.54	2.13	1.05	50.83	26.06	19.54	13.77	7.40	5.50	5.70	6.05
0.25	7.96	2.52	2.10	1.06	50.76	25.99	19.45	13.76	7.60	5.55	5.75	5.95
0.50	8.07	2.55	2.09	1.00	51.14	26.43	19.82	14.00	7.55	5.55	5.70	5.95
0.75	8.43	3.01	2.60	1.35	59.71	33.18	25.92	19.65	7.35	5.65	5.60	6.30
0.95	9.73	4.14	4.34	2.02	99.23	74.32	64.65	58.92	7.65	5.90	6.15	5.95
1.00	8.92	3.86	5.19	1.75	124.01	103.76	98.00	96.18	6.90	6.35	5.60	6.10

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.7.

Table S.9: Small sample properties of the GMM estimator of ρ for the experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-10.66	-3.04	-1.85	-0.93	23.64	10.72	8.14	5.62	8.75	4.65	4.90	5.60
0.25	-10.58	-3.02	-1.84	-0.93	23.64	10.71	8.14	5.62	8.80	4.65	5.00	5.65
0.50	-10.86	-3.09	-1.87	-0.93	23.75	10.84	8.15	5.62	9.10	4.65	4.70	5.30
0.75	-11.28	-3.28	-2.05	-1.01	24.72	11.36	8.57	5.84	10.25	5.80	6.45	5.80
0.95	-12.90	-3.80	-2.41	-1.16	29.27	13.69	10.21	6.87	15.90	11.55	9.45	15.15
1.00	-13.60	-3.78	-2.38	-1.16	32.57	15.43	11.72	8.02	20.20	14.90	15.95	14.80
$\rho_0 = 0.5$												
0.00	-8.37	-2.32	-1.41	-0.71	18.32	7.94	5.98	4.11	8.55	4.50	4.80	5.80
0.25	-8.32	-2.31	-1.40	-0.71	18.33	7.93	5.97	4.10	8.80	4.45	4.90	5.70
0.50	-8.53	-2.36	-1.42	-0.71	18.42	8.03	5.99	4.11	9.35	4.70	4.95	5.50
0.75	-9.09	-2.57	-1.59	-0.78	19.71	8.57	6.39	4.31	10.15	6.30	6.50	5.70
0.95	-11.33	-3.26	-2.03	-0.96	25.18	11.15	8.16	5.40	18.80	14.05	12.75	10.75
1.00	-12.43	-3.40	-2.12	-1.02	28.94	13.24	10.00	6.81	25.15	20.50	20.80	19.50
$\rho_0 = 0.75$												
0.00	-5.58	-1.48	-0.89	-0.45	12.14	4.98	3.71	2.53	9.35	4.95	4.90	5.95
0.25	-5.55	-1.47	-0.88	-0.45	12.17	4.98	3.70	2.53	9.25	4.90	5.10	5.75
0.50	-5.68	-1.50	-0.90	-0.45	12.22	5.03	3.71	2.53	9.35	4.90	4.70	5.85
0.75	-6.25	-1.69	-1.04	-0.50	13.57	5.51	4.04	2.70	11.00	6.85	6.90	6.10
0.95	-9.06	-2.40	-1.46	-0.66	20.06	8.21	5.79	3.75	25.75	17.75	16.25	13.50
1.00	-10.98	-2.87	-1.76	-0.81	24.65	10.99	8.24	5.59	36.80	33.50	32.50	31.45

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic non-Gaussian errors. See also the notes to Table S.3.

Table S.10: Small sample properties of the GMM estimator of β for the experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	300	500	1,000	
$\rho_0 = 0.2$												
0.00	7.52	2.32	1.86	0.87	44.34	23.07	17.42	12.11	7.85	5.70	5.70	12.80
0.25	7.45	2.27	1.84	0.86	44.31	23.05	17.39	12.09	7.65	5.90	6.05	5.75
0.50	7.52	2.23	1.80	0.81	44.44	23.24	17.48	12.11	7.00	6.05	5.85	5.80
0.75	7.07	2.54	1.95	1.01	45.65	23.97	18.09	12.64	7.25	5.40	5.45	6.35
0.95	5.78	2.20	1.99	0.90	49.94	28.21	21.96	16.68	7.75	5.70	6.25	12.50
1.00	4.44	1.41	1.58	0.55	53.08	31.67	25.99	21.85	7.05	5.95	5.45	5.70
$\rho_0 = 0.5$												
0.00	9.06	2.74	2.19	1.02	52.88	27.10	20.39	14.15	7.85	5.85	6.00	5.90
0.25	8.97	2.68	2.17	1.01	52.79	27.07	20.35	14.13	7.70	6.00	6.00	5.80
0.50	9.06	2.64	2.11	0.95	53.15	27.46	20.57	14.23	7.25	5.85	5.85	5.95
0.75	8.81	3.16	2.40	1.26	57.55	30.67	23.34	16.78	7.70	5.80	5.55	6.40
0.95	8.44	3.67	3.44	1.79	78.04	51.99	43.87	38.29	8.55	6.10	6.75	6.20
1.00	6.94	2.65	3.38	1.98	89.68	67.69	61.98	61.28	7.15	6.50	5.90	6.90
$\rho_0 = 0.75$												
0.00	10.44	3.12	2.44	1.14	56.07	28.28	21.21	14.69	7.90	5.75	5.80	5.75
0.25	10.36	3.06	2.42	1.13	55.94	28.25	21.17	14.67	7.65	5.95	5.85	5.70
0.50	10.53	3.05	2.37	1.07	56.41	28.86	21.54	14.90	7.00	5.85	5.90	5.85
0.75	10.66	3.82	2.86	1.51	65.80	36.20	27.98	20.81	8.05	5.90	5.80	6.50
0.95	11.24	6.00	5.51	3.82	110.68	82.03	73.54	72.35	9.50	7.75	8.55	10.15
1.00	9.82	4.70	5.64	4.06	129.11	114.65	113.50	136.22	8.10	7.85	8.10	12.70

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.9.

Table S.11: Small sample properties of the OGMM estimator of ρ for the experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-2.90	-1.16	-0.80	-0.41	19.88	9.68	7.49	5.25	9.40	4.30	5.20	5.25
0.25	-2.85	-1.14	-0.79	-0.41	19.91	9.68	7.49	5.25	9.50	4.25	5.25	5.15
0.50	-3.20	-1.18	-0.82	-0.41	20.04	9.80	7.51	5.26	8.70	4.90	4.95	13.40
0.75	-1.96	-0.85	-0.69	-0.34	21.09	10.30	7.85	5.47	11.60	6.40	5.90	6.70
0.95	1.65	0.75	0.18	0.10	27.81	13.46	9.90	6.79	22.25	15.50	12.65	10.85
1.00	2.24	0.95	0.32	0.08	32.85	16.57	12.52	8.57	29.25	21.55	20.20	18.40
$\rho_0 = 0.5$												
0.00	-2.62	-1.06	-0.70	-0.36	15.67	7.28	5.56	3.88	10.15	4.55	5.05	5.80
0.25	-2.61	-1.05	-0.70	-0.36	15.69	7.28	5.56	3.88	10.20	4.55	5.15	5.80
0.50	-2.84	-1.08	-0.72	-0.37	15.85	7.36	5.57	3.89	10.15	4.75	5.05	5.85
0.75	-1.75	-0.87	-0.69	-0.34	17.13	7.91	5.96	4.11	13.70	6.80	6.45	6.40
0.95	2.07	0.71	0.11	0.04	24.07	11.26	8.21	5.44	30.60	19.30	16.35	12.05
1.00	2.12	1.12	0.34	0.07	28.58	14.75	11.05	7.46	38.10	28.15	27.05	24.45
$\rho_0 = 0.75$												
0.00	-1.96	-0.76	-0.49	-0.25	10.40	4.56	3.41	2.36	12.50	5.45	5.20	5.60
0.25	-1.99	-0.77	-0.49	-0.25	10.39	4.56	3.41	2.35	12.20	5.60	5.10	5.55
0.50	-2.11	-0.78	-0.50	-0.26	10.50	4.60	3.42	2.37	11.85	5.30	5.50	5.55
0.75	-1.37	-0.71	-0.54	-0.27	11.85	5.10	3.78	2.57	17.00	7.70	7.85	6.45
0.95	1.26	0.60	0.08	-0.01	17.94	8.38	5.93	3.74	40.05	25.45	22.30	16.30
1.00	0.04	0.84	0.32	0.05	22.01	11.97	9.20	6.15	52.10	42.10	40.45	37.75

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic non-Gaussian errors. See also the notes to Table S.5.

Table S.12: Small sample properties of the OGMM estimator of β for the experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	0.32	0.17	0.55	0.14	39.89	21.06	15.90	11.33	6.45	4.95	5.55	6.25
0.25	0.31	0.14	0.54	0.14	39.84	21.07	15.88	11.33	6.55	5.10	5.55	6.15
0.50	0.29	0.07	0.51	0.12	40.00	21.11	15.93	11.34	6.60	5.20	5.45	6.00
0.75	-0.35	0.05	0.52	0.17	41.26	21.78	16.49	11.77	6.70	5.20	5.10	5.85
0.95	-0.50	-0.46	0.33	-0.02	47.19	26.85	20.66	16.05	6.60	6.30	6.00	5.90
1.00	-0.07	-0.22	0.59	0.06	51.44	31.11	25.58	21.67	6.00	6.15	5.55	5.45
$\rho_0 = 0.5$												
0.00	0.61	0.32	0.72	0.21	47.02	24.71	18.60	13.25	6.60	5.15	5.60	6.10
0.25	0.61	0.29	0.71	0.21	46.93	24.70	18.58	13.24	6.75	5.25	5.50	6.00
0.50	0.56	0.21	0.68	0.18	47.26	24.89	18.74	13.33	6.50	5.25	5.50	6.00
0.75	-0.14	0.25	0.79	0.31	51.27	27.80	21.31	15.71	6.65	5.10	5.15	5.65
0.95	0.15	-0.16	1.16	0.34	71.09	48.49	40.57	35.86	6.80	5.90	5.60	5.75
1.00	1.28	0.50	2.04	0.58	85.14	64.93	59.46	56.97	5.85	5.80	5.25	5.50
$\rho_0 = 0.75$												
0.00	0.72	0.37	0.77	0.24	49.01	25.58	19.19	13.66	6.80	5.35	5.50	5.80
0.25	0.74	0.36	0.77	0.24	48.88	25.57	19.17	13.64	6.70	5.45	5.40	5.90
0.50	0.71	0.28	0.74	0.21	49.26	25.95	19.47	13.86	6.70	5.30	5.40	6.05
0.75	0.20	0.43	1.00	0.43	57.41	32.54	25.43	19.47	7.05	5.20	5.35	6.00
0.95	1.48	0.26	2.10	0.73	95.33	72.39	63.17	58.39	6.95	5.70	5.55	5.95
1.00	3.34	1.44	3.67	1.18	118.41	101.35	96.22	95.52	6.15	5.80	5.30	5.55

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.11.

Table S.13: Small sample properties of the BMM estimator of ρ for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-9.02	-2.71	-1.63	-0.82	20.65	10.43	7.85	5.43	7.30	5.85	6.05	5.00
0.25	-8.94	-2.70	-1.63	-0.83	20.50	10.41	7.84	5.42	7.50	5.85	5.90	4.95
0.50	-8.97	-2.69	-1.62	-0.82	20.37	10.33	7.82	5.44	7.05	6.05	6.20	5.30
0.75	-9.38	-2.79	-1.61	-0.83	21.25	10.72	8.07	5.50	7.45	6.00	6.30	5.70
0.95	-11.45	-3.50	-2.05	-1.06	25.30	12.82	9.50	6.46	10.70	9.30	9.05	8.00
1.00	-13.07	-4.17	-2.52	-1.21	28.68	15.09	11.42	7.71	14.20	14.00	13.45	12.90
$\rho_0 = 0.5$												
0.00	-8.02	-2.37	-1.39	-0.71	16.60	7.91	5.83	4.01	7.35	6.00	6.00	5.20
0.25	-7.97	-2.37	-1.38	-0.71	16.50	7.90	5.83	4.01	7.10	6.15	5.85	5.20
0.50	-7.97	-2.36	-1.38	-0.72	16.41	7.83	5.81	4.02	7.20	5.80	6.05	5.15
0.75	-8.62	-2.55	-1.42	-0.74	17.50	8.29	6.08	4.10	7.80	6.20	6.00	5.40
0.95	-11.82	-3.57	-2.00	-1.05	22.54	10.58	7.60	5.11	12.85	10.30	10.15	8.90
1.00	-14.37	-4.64	-2.75	-1.33	26.58	13.33	9.90	6.64	17.90	17.05	16.65	16.55
$\rho_0 = 0.75$												
0.00	-5.95	-1.69	-0.96	-0.50	11.23	4.93	3.54	2.42	7.10	6.00	5.65	5.15
0.25	-5.93	-1.69	-0.96	-0.50	11.18	4.93	3.54	2.42	6.95	5.85	5.55	5.30
0.50	-5.90	-1.67	-0.96	-0.50	11.11	4.87	3.52	2.43	7.80	5.75	5.55	5.10
0.75	-6.62	-1.89	-1.02	-0.53	12.20	5.29	3.75	2.50	8.50	6.95	5.55	5.45
0.95	-11.35	-3.25	-1.75	-0.90	18.78	7.66	5.24	3.42	16.10	12.05	12.65	10.95
1.00	-15.74	-5.32	-3.17	-1.58	24.27	11.33	8.23	5.46	23.95	26.35	26.30	26.50

Notes: The DGP is given by (S.53) and (S.54) with heteroskedastic Gaussian errors. $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. The first unit is δ -dominant, and the rest of the units are non-dominant. The BMM estimator is computed by (48) in the paper. The power is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value.

Table S.14: Small sample properties of the BMM estimator of β for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	3.41	1.20	0.50	0.50	39.62	20.81	15.78	10.99	8.50	5.55	4.75	4.60
0.25	3.42	1.22	0.50	0.50	39.59	20.82	15.78	10.99	8.60	5.70	4.85	4.65
0.50	3.47	1.23	0.49	0.48	39.86	20.94	15.81	11.01	8.75	5.45	5.15	4.55
0.75	3.38	1.15	0.39	0.46	41.10	21.40	15.83	11.06	8.80	5.40	4.60	4.60
0.95	2.09	0.75	0.20	0.46	45.30	24.06	16.49	11.67	8.25	5.80	4.65	4.75
1.00	1.04	0.09	-0.12	0.26	49.14	26.87	17.67	12.66	8.10	5.80	4.70	4.35
$\rho_0 = 0.5$												
0.00	4.75	1.68	0.72	0.67	47.09	24.47	18.50	12.87	8.70	5.70	4.80	4.65
0.25	4.77	1.71	0.72	0.67	47.04	24.49	18.50	12.88	8.90	5.80	4.80	4.75
0.50	4.82	1.73	0.71	0.65	47.48	24.72	18.56	12.92	8.80	5.60	5.20	4.70
0.75	4.86	1.71	0.62	0.64	50.04	25.77	18.68	13.07	8.90	5.55	4.90	4.65
0.95	3.34	1.30	0.42	0.67	57.24	30.34	19.83	14.15	8.35	5.70	4.95	4.75
1.00	1.84	0.32	-0.01	0.39	63.73	34.97	21.66	15.81	8.35	5.90	4.75	4.40
$\rho_0 = 0.75$												
0.00	6.27	2.19	1.00	0.83	49.36	25.39	19.15	13.31	9.15	5.75	4.90	4.75
0.25	6.29	2.22	1.00	0.83	49.29	25.41	19.15	13.32	9.15	5.85	4.90	4.75
0.50	6.33	2.25	1.00	0.82	49.63	25.72	19.22	13.39	9.15	6.00	5.40	4.95
0.75	6.60	2.36	0.93	0.83	53.73	27.45	19.53	13.68	9.25	5.55	5.25	4.65
0.95	5.13	2.16	0.79	0.93	63.81	33.65	21.41	15.41	8.85	6.00	5.05	4.85
1.00	3.16	0.89	0.32	0.61	71.05	39.41	23.56	17.58	8.55	5.85	4.90	4.50

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.13.

Table S.15: Small sample properties of the GMM estimator of ρ for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-11.89	-3.62	-2.08	-1.06	24.19	11.47	8.43	5.76	9.25	6.85	5.90	5.20
0.25	-11.81	-3.62	-2.08	-1.06	24.00	11.46	8.42	5.75	8.85	6.90	5.95	5.30
0.50	-11.86	-3.62	-2.07	-1.06	23.90	11.40	8.41	5.77	8.70	6.60	6.10	5.30
0.75	-12.39	-3.76	-2.08	-1.07	25.07	11.85	8.70	5.83	9.55	6.45	6.45	5.75
0.95	-13.69	-4.27	-2.41	-1.27	29.03	13.94	10.12	6.77	13.35	9.90	9.30	8.15
1.00	-14.63	-4.44	-2.61	-1.25	32.31	16.04	11.97	8.00	16.70	15.20	13.70	12.60
$\rho_0 = 0.5$												
0.00	-9.30	-2.76	-1.58	-0.80	18.67	8.50	6.19	4.20	8.75	6.70	6.00	5.35
0.25	-9.24	-2.76	-1.57	-0.80	18.53	8.49	6.18	4.20	8.55	6.70	6.05	5.35
0.50	-9.26	-2.76	-1.57	-0.80	18.43	8.45	6.18	4.22	8.75	6.25	5.85	5.25
0.75	-9.88	-2.93	-1.60	-0.82	19.74	8.93	6.47	4.30	10.20	6.50	6.25	5.85
0.95	-11.90	-3.60	-1.99	-1.04	24.79	11.31	8.06	5.32	15.05	11.40	10.95	8.75
1.00	-13.34	-4.01	-2.35	-1.12	28.64	13.87	10.28	6.83	21.60	18.60	17.95	16.75
$\rho_0 = 0.75$												
0.00	-6.17	-1.76	-0.99	-0.50	12.31	5.32	3.83	2.59	8.15	6.70	6.00	5.50
0.25	-6.13	-1.76	-0.99	-0.50	12.23	5.32	3.83	2.59	8.40	6.50	5.95	5.50
0.50	-6.13	-1.76	-0.99	-0.50	12.12	5.28	3.82	2.60	8.60	6.40	6.00	5.55
0.75	-6.71	-1.91	-1.03	-0.52	13.40	5.71	4.07	2.69	10.85	7.05	6.30	5.85
0.95	-9.44	-2.63	-1.41	-0.73	19.64	8.23	5.69	3.70	20.80	14.35	13.45	10.75
1.00	-11.77	-3.46	-2.01	-0.96	24.44	11.55	8.50	5.61	31.60	27.60	25.70	21.25

Notes: The DGP is given by (S.53) and (S.54) with heteroskedastic Gaussian errors. $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. The first unit is δ -dominant, and the rest of the units are non-dominant. The GMM estimator is given by (23) in the paper, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - Diag(\mathbf{W}^2)$, and $\mathbf{A}_n = \mathbf{I}_n$. The power is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value.

Table S.16: Small sample properties of the GMM estimator of β for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	7.15	2.69	1.27	0.88	44.06	22.60	17.06	11.92	9.10	5.75	5.40	5.80
0.25	7.15	2.73	1.28	0.88	43.99	22.63	17.06	11.93	9.20	5.70	5.45	5.70
0.50	7.27	2.76	1.26	0.84	44.44	22.80	17.12	11.96	10.10	5.90	5.70	5.80
0.75	7.04	2.68	1.14	0.84	45.57	23.41	17.20	12.05	8.95	5.75	5.65	5.95
0.95	3.73	1.77	0.70	0.76	47.30	25.15	17.26	12.19	8.75	5.65	5.65	4.80
1.00	1.92	0.42	0.07	0.36	50.18	27.08	17.80	12.75	8.40	5.85	5.10	4.45
$\rho_0 = 0.5$												
0.00	8.50	3.17	1.49	1.03	52.45	26.51	19.94	13.91	9.35	5.55	5.30	5.90
0.25	8.51	3.21	1.50	1.04	52.35	26.54	19.94	13.92	9.30	5.65	5.40	5.85
0.50	8.67	3.26	1.48	0.99	53.01	26.84	20.03	13.98	10.10	5.85	5.65	5.65
0.75	8.60	3.24	1.37	1.01	55.66	28.15	20.26	14.21	9.15	5.85	5.85	5.70
0.95	4.71	2.32	0.92	0.95	59.97	31.91	20.84	14.81	8.75	5.80	5.75	5.10
1.00	2.32	0.50	0.08	0.45	65.17	35.30	21.82	15.92	8.65	6.15	5.15	4.45
$\rho_0 = 0.75$												
0.00	9.83	3.63	1.75	1.18	55.37	27.72	20.77	14.48	9.35	5.70	5.30	5.85
0.25	9.86	3.67	1.75	1.18	55.24	27.76	20.76	14.49	9.35	5.85	5.40	5.90
0.50	10.06	3.74	1.74	1.14	55.87	28.15	20.88	14.59	10.25	5.75	5.70	5.80
0.75	10.28	3.86	1.66	1.18	60.44	30.28	21.35	15.00	9.35	6.25	5.85	5.80
0.95	5.69	2.96	1.19	1.15	66.92	35.88	22.77	16.33	9.35	6.10	6.20	5.40
1.00	2.71	0.63	0.14	0.51	72.26	39.76	23.72	17.70	8.70	6.10	5.20	4.60

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.15.

Table S.17: Small sample properties of the OGMM estimator of ρ for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	1,000	100	300	500
$\rho_0 = 0.2$												
0.00	-3.58	-1.41	-0.95	-0.50	20.51	10.26	7.79	5.41	10.25	6.55	6.05	5.20
0.25	-3.49	-1.41	-0.94	-0.50	20.42	10.24	7.78	5.41	10.65	6.55	6.05	5.20
0.50	-3.59	-1.40	-0.91	-0.49	20.33	10.19	7.77	5.43	10.40	6.25	6.20	5.30
0.75	-2.53	-1.13	-0.71	-0.42	21.33	10.75	8.12	5.54	12.05	7.35	7.15	5.80
0.95	-0.03	0.22	-0.03	-0.16	27.53	14.05	10.04	6.74	21.50	15.65	12.15	10.25
1.00	0.14	0.15	-0.13	-0.02	32.92	17.68	12.84	8.55	26.35	21.40	19.05	16.50
$\rho_0 = 0.5$												
0.00	-3.14	-1.29	-0.81	-0.44	16.19	7.71	5.77	3.99	11.35	6.70	6.10	5.30
0.25	-3.10	-1.29	-0.80	-0.44	16.13	7.69	5.76	3.99	10.75	6.65	5.90	5.50
0.50	-3.14	-1.28	-0.78	-0.44	16.19	7.66	5.75	4.01	10.95	6.70	6.10	5.35
0.75	-2.26	-1.15	-0.67	-0.40	17.40	8.23	6.09	4.13	13.60	7.80	7.25	5.75
0.95	-0.14	0.17	-0.01	-0.18	23.31	11.73	8.15	5.35	26.05	17.95	14.00	11.65
1.00	-0.11	0.19	-0.02	0.02	28.70	15.79	11.54	7.60	33.90	26.95	24.40	21.65
$\rho_0 = 0.75$												
0.00	-2.30	-0.90	-0.54	-0.30	10.60	4.79	3.51	2.42	12.40	7.05	6.10	5.30
0.25	-2.32	-0.91	-0.54	-0.30	10.53	4.78	3.50	2.42	12.15	7.20	6.15	5.35
0.50	-2.29	-0.89	-0.53	-0.31	10.56	4.75	3.50	2.43	12.85	7.15	6.20	5.20
0.75	-1.66	-0.89	-0.49	-0.29	11.73	5.23	3.76	2.53	16.90	8.40	6.95	6.00
0.95	-0.79	0.14	0.03	-0.13	17.15	8.47	5.73	3.66	33.20	21.60	17.55	14.00
1.00	-1.80	-0.06	-0.06	0.03	22.33	12.72	9.49	6.38	45.30	38.70	34.60	31.80

Notes: The DGP is given by (S.53) and (S.54) with heteroskedastic Gaussian errors. The OGMM estimator is given by (31) in the paper, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{Wx}, \mathbf{W}^2\mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - Diag(\mathbf{W}^2)$, and the weighting matrix \mathbf{A}_n is the inverse of the estimated covariance of moments. See also the notes to Table S.15.

Table S.18: Small sample properties of the OGMM estimator of β for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-0.82	-0.22	-0.27	0.10	39.93	20.92	15.75	10.99	9.15	6.05	4.90	4.80
0.25	-0.85	-0.18	-0.27	0.10	39.90	20.91	15.75	10.99	9.15	5.90	4.80	4.75
0.50	-0.82	-0.21	-0.31	0.06	40.02	20.98	15.77	11.01	8.90	5.60	4.90	4.70
0.75	-1.20	-0.43	-0.48	0.04	41.18	21.60	15.84	11.09	9.35	6.35	4.80	5.20
0.95	-1.81	-1.10	-0.84	-0.05	45.27	24.27	16.56	11.72	9.00	6.65	5.00	4.90
1.00	-2.30	-1.24	-0.78	-0.13	49.02	26.95	17.69	12.70	8.10	5.85	5.05	4.90
$\rho_0 = 0.5$												
0.00	-0.80	-0.09	-0.25	0.17	46.84	24.48	18.43	12.87	9.00	6.20	5.15	4.95
0.25	-0.80	-0.04	-0.24	0.17	46.75	24.48	18.43	12.87	9.15	6.10	5.15	4.90
0.50	-0.80	-0.07	-0.28	0.14	47.04	24.65	18.47	12.91	8.95	5.65	5.55	4.65
0.75	-1.21	-0.29	-0.47	0.12	49.56	25.93	18.65	13.09	9.15	6.40	5.25	5.20
0.95	-1.90	-1.25	-0.99	-0.04	56.46	30.48	19.88	14.21	8.90	6.55	5.25	4.85
1.00	-2.66	-1.55	-0.95	-0.15	62.81	34.94	21.64	15.85	8.10	5.95	5.00	4.80
$\rho_0 = 0.75$												
0.00	-0.86	-0.01	-0.23	0.21	48.59	25.34	19.06	13.31	9.10	6.25	5.60	4.80
0.25	-0.81	0.05	-0.22	0.22	48.41	25.34	19.05	13.31	9.00	6.30	5.50	4.90
0.50	-0.82	0.03	-0.25	0.19	48.61	25.58	19.11	13.38	8.95	6.05	5.85	4.75
0.75	-1.21	-0.13	-0.44	0.18	52.53	27.56	19.47	13.70	9.15	6.65	5.70	5.20
0.95	-1.58	-1.35	-1.14	-0.06	62.05	33.71	21.46	15.50	8.90	6.60	5.40	5.05
1.00	-2.15	-1.66	-1.05	-0.19	68.86	39.12	23.48	17.60	8.10	6.00	5.20	4.85

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.17.

Table S.19: Small sample properties of the BGMM estimator of ρ for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-6.79	-2.13	-1.33	-0.69	20.13	10.35	7.82	5.42	7.85	6.25	6.05	5.10
0.25	-6.76	-2.12	-1.33	-0.69	20.01	10.34	7.82	5.42	7.70	6.00	6.00	5.10
0.50	-6.85	-2.12	-1.30	-0.68	19.95	10.27	7.80	5.43	7.55	6.00	6.40	5.45
0.75	-6.74	-2.11	-1.23	-0.67	20.99	10.74	8.07	5.50	8.90	6.85	6.70	5.60
0.95	-6.27	-1.95	-1.24	-0.73	27.00	13.28	9.62	6.53	15.85	12.25	10.05	8.80
1.00	-6.21	-2.19	-1.33	-0.67	31.47	16.19	12.05	8.10	20.50	17.90	16.05	15.05
$\rho_0 = 0.5$												
0.00	-5.81	-1.84	-1.10	-0.58	16.03	7.81	5.80	4.00	8.05	6.10	6.00	5.05
0.25	-5.81	-1.84	-1.10	-0.59	15.96	7.80	5.79	4.00	8.15	6.40	5.80	5.20
0.50	-5.85	-1.83	-1.09	-0.59	15.96	7.75	5.78	4.01	8.65	6.25	6.00	5.45
0.75	-5.84	-1.91	-1.06	-0.59	17.29	8.29	6.07	4.10	10.45	6.95	6.15	5.45
0.95	-6.03	-1.78	-1.07	-0.67	23.99	11.26	7.81	5.19	21.20	14.10	12.45	9.70
1.00	-5.86	-1.76	-1.09	-0.57	29.11	15.36	11.18	7.38	27.45	23.70	22.75	20.60
$\rho_0 = 0.75$												
0.00	-4.06	-1.28	-0.74	-0.40	10.77	4.85	3.51	2.40	9.70	6.35	5.75	4.95
0.25	-4.08	-1.28	-0.74	-0.40	10.74	4.84	3.50	2.40	9.60	6.20	5.65	4.95
0.50	-4.03	-1.26	-0.73	-0.41	10.82	4.80	3.49	2.42	10.50	6.10	5.60	5.20
0.75	-4.04	-1.37	-0.74	-0.42	12.06	5.30	3.74	2.49	14.35	8.35	6.50	5.80
0.95	-5.74	-1.26	-0.65	-0.49	20.29	8.79	5.87	3.52	32.35	20.45	17.10	14.10
1.00	-7.93	-1.74	-0.93	-0.35	28.02	14.79	10.82	7.39	43.35	38.85	38.45	36.20

Notes: The DGP is given by (S.53) and (S.54) with heteroskedastic Gaussian errors. The BGMM estimator is computed in two steps: In the first step, we obtain preliminary GMM estimates, $\tilde{\psi} = (\tilde{\rho}, \tilde{\alpha}, \tilde{\beta})'$, following (23) in the paper, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{Wx}, \mathbf{W}^2 \mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - Diag(\mathbf{W}^2)$, and $\mathbf{A}_n = \mathbf{I}_n$. In the second step, we use $(\tilde{\mathbf{G}}\tilde{\mathbf{x}}\tilde{\mathbf{G}}\tilde{\mathbf{x}}\tilde{\beta}, \mathbf{1}_n, \mathbf{x})$ and $\tilde{\mathbf{G}} - Diag(\tilde{\mathbf{G}})$, where $\tilde{\mathbf{G}} = \mathbf{W}(\mathbf{I}_n - \hat{\rho}\mathbf{W})^{-1}$, in the linear and quadratic moments, respectively, and compute the optimal GMM estimates by (31) in the paper. See also the notes to Table S.15.

Table S.20: Small sample properties of the BGMM estimator of β for the experiments with heteroskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	0.32	0.21	-0.05	0.21	38.94	20.75	15.70	10.98	8.50	5.75	4.75	11.50
0.25	0.35	0.23	-0.05	0.21	38.91	20.76	15.70	10.99	8.55	5.80	4.70	4.90
0.50	0.37	0.26	-0.07	0.18	39.18	20.89	15.73	11.01	8.85	5.70	5.10	4.75
0.75	0.05	0.07	-0.24	0.16	40.64	21.44	15.77	11.06	9.05	5.80	4.75	4.80
0.95	-1.09	-0.54	-0.56	0.11	45.03	24.20	16.52	11.71	8.50	6.50	4.75	4.70
1.00	-2.03	-0.97	-0.74	-0.07	48.67	26.94	17.70	12.68	8.15	6.00	4.80	4.40
$\rho_0 = 0.5$												
0.00	0.66	0.41	0.01	0.30	46.06	24.36	18.38	12.85	8.45	5.80	5.00	4.65
0.25	0.70	0.44	0.02	0.30	46.01	24.37	18.38	12.86	8.40	5.80	4.95	4.80
0.50	0.71	0.48	-0.01	0.27	46.43	24.60	18.44	12.90	8.60	5.80	5.20	4.75
0.75	0.40	0.29	-0.21	0.25	49.20	25.78	18.59	13.05	8.90	6.10	4.95	5.00
0.95	-0.94	-0.56	-0.64	0.17	56.69	30.53	19.86	14.19	8.45	6.55	4.90	4.70
1.00	-2.34	-1.36	-0.98	-0.12	62.43	35.05	21.74	15.83	7.65	6.05	5.00	4.30
$\rho_0 = 0.75$												
0.00	0.88	0.56	0.07	0.36	48.35	25.27	19.02	13.27	8.80	6.10	5.00	4.55
0.25	0.93	0.59	0.08	0.37	48.27	25.28	19.01	13.27	8.80	5.90	5.10	4.60
0.50	0.93	0.63	0.06	0.35	48.56	25.57	19.08	13.34	8.80	6.05	5.20	4.85
0.75	0.55	0.49	-0.16	0.33	52.67	27.46	19.43	13.64	9.30	6.15	5.35	4.65
0.95	-0.01	-0.57	-0.75	0.19	62.98	34.07	21.52	15.47	8.45	6.85	5.35	4.80
1.00	-1.57	-1.63	-1.25	-0.33	68.63	39.34	23.80	17.70	8.00	6.15	5.20	4.95

S50

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.19.

Table S.21: Small sample properties of the GMM estimator of ρ for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-10.50	-2.97	-1.85	-0.93	23.48	10.70	8.21	5.64	11.90	6.10	6.25	5.65
0.25	-10.42	-2.94	-1.83	-0.93	23.44	10.68	8.21	5.63	12.20	6.10	6.35	5.75
0.50	-10.63	-2.99	-1.85	-0.92	23.54	10.78	8.21	5.63	11.90	6.15	5.60	5.55
0.75	-11.02	-3.18	-1.99	-0.98	24.50	11.29	8.58	5.82	11.70	6.75	6.95	6.15
0.95	-12.34	-3.64	-2.24	-1.07	28.74	13.57	10.10	6.75	15.40	9.80	8.75	7.40
1.00	-13.10	-3.55	-2.24	-1.08	32.32	15.42	11.76	7.99	18.70	12.40	12.95	11.00
$\rho_0 = 0.5$												
0.00	-8.25	-2.27	-1.40	-0.70	18.17	7.92	6.04	4.12	11.85	6.20	6.05	5.75
0.25	-8.19	-2.25	-1.39	-0.70	18.15	7.91	6.03	4.12	12.10	6.20	6.25	5.80
0.50	-8.34	-2.28	-1.40	-0.70	18.21	7.98	6.03	4.12	12.05	6.30	5.85	5.35
0.75	-8.83	-2.47	-1.54	-0.75	19.35	8.50	6.39	4.29	12.20	6.85	7.05	6.10
0.95	-10.75	-3.09	-1.87	-0.89	24.51	11.01	8.04	5.30	17.35	11.25	10.40	8.80
1.00	-11.93	-3.21	-2.01	-0.97	28.61	13.24	10.06	6.79	22.70	16.30	16.25	14.45
$\rho_0 = 0.75$												
0.00	-5.49	-1.45	-0.89	-0.44	12.02	4.96	3.74	2.54	12.15	6.55	6.30	5.70
0.25	-5.45	-1.43	-0.88	-0.44	12.01	4.95	3.74	2.53	12.35	6.45	6.25	5.60
0.50	-5.54	-1.45	-0.89	-0.44	12.04	4.99	3.74	2.54	12.90	6.55	5.85	5.70
0.75	-6.02	-1.61	-1.00	-0.48	13.16	5.45	4.04	2.68	12.85	7.40	7.10	6.15
0.95	-8.55	-2.27	-1.34	-0.63	19.36	8.08	5.69	3.67	23.05	14.80	13.10	10.35
1.00	-10.56	-2.77	-1.71	-0.82	24.33	10.99	8.29	5.56	31.45	25.05	24.90	23.10

S51

Notes: The DGP is given by (S.53) and (S.54) with heteroskedastic non-Gaussian errors. $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. The first unit is δ -dominant, and the rest of the units are non-dominant. The GMM estimator is given by (23) in the paper, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{W}_{\mathbf{x}}, \mathbf{W}_{\mathbf{x}}^2)$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - Diag(\mathbf{W}^2)$, and $\mathbf{A}_n = \mathbf{I}_n$. The power is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value.

Table S.22: Small sample properties of the GMM estimator of β for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	7.31	2.28	1.79	0.80	44.29	23.15	17.51	12.21	8.35	5.90	5.75	6.25
0.25	7.26	2.23	1.77	0.79	44.27	23.14	17.48	12.20	8.25	6.15	5.80	6.20
0.50	7.34	2.17	1.71	0.74	44.43	23.33	17.54	12.21	8.35	6.00	6.00	6.15
0.75	6.81	2.50	1.78	0.91	45.70	23.96	17.70	12.35	8.75	5.95	5.60	6.35
0.95	5.37	2.12	1.55	0.69	48.72	26.30	17.81	12.66	9.25	6.20	5.90	5.35
1.00	4.09	1.32	1.07	0.36	51.48	27.90	18.10	13.33	8.85	6.40	5.50	6.45
$\rho_0 = 0.5$												
0.00	8.77	2.71	2.11	0.94	52.80	27.19	20.50	14.27	8.55	6.00	5.70	6.15
0.25	8.72	2.65	2.08	0.93	52.75	27.17	20.47	14.26	8.50	6.15	5.70	6.15
0.50	8.82	2.57	2.02	0.87	53.10	27.51	20.57	14.29	8.50	6.10	6.05	6.15
0.75	8.40	3.05	2.13	1.08	56.02	28.84	20.90	14.56	8.85	5.90	5.70	6.45
0.95	6.98	2.80	1.92	0.87	62.00	33.30	21.49	15.39	9.40	6.40	6.15	5.35
1.00	5.26	1.71	1.32	0.45	66.76	36.35	22.19	16.64	8.90	6.50	5.60	6.45
$\rho_0 = 0.75$												
0.00	10.11	3.09	2.36	1.06	55.95	28.38	21.34	14.83	8.80	6.00	5.70	6.15
0.25	10.06	3.02	2.34	1.06	55.85	28.36	21.30	14.81	8.80	6.10	5.75	6.15
0.50	10.23	2.97	2.27	0.99	56.17	28.79	21.42	14.87	8.50	6.10	6.20	6.15
0.75	10.04	3.58	2.42	1.22	60.98	30.96	22.02	15.33	9.25	5.85	5.80	6.50
0.95	8.12	3.45	2.21	1.01	69.25	37.35	23.41	16.94	9.65	6.40	5.75	10.05
1.00	5.99	1.98	1.48	0.52	73.94	40.90	24.11	18.48	9.10	6.75	5.65	6.30

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.21.

Table S.23: Small sample properties of the OGMM estimator of ρ for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-2.80	-1.12	-0.77	-0.41	19.96	9.70	7.53	5.26	9.50	5.00	5.50	5.10
0.25	-2.75	-1.11	-0.76	-0.40	19.91	9.69	7.52	5.25	9.20	5.00	5.50	5.10
0.50	-2.97	-1.12	-0.78	-0.40	19.92	9.78	7.54	5.25	9.65	4.80	5.35	5.25
0.75	-1.76	-0.83	-0.69	-0.33	21.20	10.33	7.85	5.45	12.10	6.80	6.15	6.05
0.95	1.53	0.55	0.13	0.07	27.69	13.49	9.77	6.66	21.95	16.45	12.85	11.35
1.00	1.75	0.80	0.30	0.03	33.20	16.73	12.42	8.40	30.10	21.15	19.85	17.90
$\rho_0 = 0.5$												
0.00	-2.52	-1.26	-0.87	-0.47	16.28	7.42	5.67	3.92	14.05	6.40	6.25	6.00
0.25	-2.56	-1.26	-0.87	-0.47	16.22	7.41	5.67	3.92	13.80	6.40	6.25	6.05
0.50	-2.73	-1.28	-0.88	-0.48	16.21	7.47	5.69	3.93	13.90	6.75	6.00	5.95
0.75	-2.18	-1.18	-0.89	-0.45	17.26	7.97	6.01	4.10	15.45	7.45	7.20	6.35
0.95	-0.54	-0.29	-0.36	-0.21	22.99	11.03	7.92	5.20	24.70	16.25	13.65	10.75
1.00	-1.45	-0.24	-0.36	-0.26	28.35	14.33	10.79	7.19	31.95	22.50	20.85	18.20
$\rho_0 = 0.75$												
0.00	-1.94	-0.74	-0.47	-0.25	10.33	4.53	3.43	2.36	12.05	5.40	4.85	5.30
0.25	-1.95	-0.74	-0.47	-0.25	10.32	4.53	3.43	2.36	11.95	5.40	4.80	5.40
0.50	-2.01	-0.74	-0.48	-0.25	10.38	4.54	3.43	2.37	11.95	5.65	4.90	5.70
0.75	-1.23	-0.66	-0.50	-0.24	11.73	5.00	3.70	2.49	17.40	7.85	7.75	6.10
0.95	1.05	0.40	0.07	-0.01	17.49	8.07	5.55	3.53	40.25	26.15	20.65	16.40
1.00	-0.37	0.63	0.25	0.01	22.24	12.03	9.00	6.03	52.50	42.55	40.10	37.50

Notes: The OGMM estimator is given by (31) in the paper, where $\mathbf{Z} = (\mathbf{1}_n, \mathbf{x}, \mathbf{Wx}, \mathbf{W}^2\mathbf{x})$, $\mathbf{B}_1 = \mathbf{W}$, $\mathbf{B}_2 = \mathbf{W}^2 - \text{Diag}(\mathbf{W}^2)$, and the weighting matrix \mathbf{A}_n is the inverse of the estimated covariance of moments. See also the notes to Table S.21.

Table S.24: Small sample properties of the OGMM estimator of β for the experiments with heteroskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	0.20	0.16	0.44	0.10	39.98	21.12	15.92	11.38	6.20	5.70	5.15	6.20
0.25	0.19	0.13	0.43	0.10	39.94	21.12	15.90	11.37	6.25	5.75	5.00	6.15
0.50	0.13	0.03	0.39	0.08	40.09	21.20	15.94	11.38	6.45	5.70	4.95	6.10
0.75	-0.46	0.06	0.39	0.13	41.46	21.79	16.09	11.51	6.55	5.60	5.20	6.25
0.95	-0.49	-0.40	0.08	-0.10	46.15	25.01	16.82	12.20	6.90	6.40	5.60	6.00
1.00	-0.15	-0.24	0.14	-0.07	49.72	27.42	17.81	13.25	6.20	6.10	5.15	5.60
$\rho_0 = 0.5$												
0.00	-0.85	-0.53	0.00	-0.24	46.00	24.69	18.60	13.26	10.25	7.05	6.50	6.60
0.25	-0.79	-0.56	-0.01	-0.24	45.92	24.72	18.59	13.26	10.25	7.15	6.30	6.60
0.50	-0.97	-0.73	-0.11	-0.29	46.17	24.94	18.64	13.28	10.35	7.25	6.30	6.65
0.75	-1.86	-0.61	-0.11	-0.22	49.07	26.20	19.03	13.58	11.40	7.55	6.25	6.60
0.95	-1.40	-0.94	-0.31	-0.45	56.37	31.72	20.19	14.74	11.35	8.20	6.90	6.15
1.00	-0.31	-0.64	-0.11	-0.34	62.65	35.56	21.72	16.42	10.30	7.35	6.45	5.90
$\rho_0 = 0.75$												
0.00	0.56	0.37	0.64	0.19	48.97	25.61	19.23	13.72	6.60	5.85	5.20	6.25
0.25	0.57	0.34	0.63	0.19	48.86	25.60	19.21	13.70	6.55	5.90	5.10	6.30
0.50	0.46	0.23	0.59	0.16	49.10	25.86	19.27	13.76	6.70	5.85	4.90	6.20
0.75	-0.15	0.30	0.60	0.22	53.51	27.71	19.74	14.17	7.30	5.55	5.35	6.40
0.95	0.14	-0.47	0.05	-0.17	63.57	34.65	21.69	16.05	6.85	6.60	5.90	6.05
1.00	0.91	-0.12	0.20	-0.07	69.84	39.74	23.60	18.35	6.45	6.25	5.20	5.50
$\rho_0 = 0.95$												
0.00	0.20	0.16	0.44	0.10	39.98	21.12	15.92	11.38	6.20	5.70	5.15	6.20
0.25	0.19	0.13	0.43	0.10	39.94	21.12	15.90	11.37	6.25	5.75	5.00	6.15
0.50	0.13	0.03	0.39	0.08	40.09	21.20	15.94	11.38	6.45	5.70	4.95	6.10
0.75	-0.46	0.06	0.39	0.13	41.46	21.79	16.09	11.51	6.55	5.60	5.20	6.25
0.95	-0.49	-0.40	0.08	-0.10	46.15	25.01	16.82	12.20	6.90	6.40	5.60	6.00
1.00	-0.15	-0.24	0.14	-0.07	49.72	27.42	17.81	13.25	6.20	6.10	5.15	5.60

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.23.

Table S.25: Goodness-of-fit, R_0^2 , of the SAR model without exogenous regressors

$\delta \setminus n$	100	300	500	1,000
$\rho_0 = 0.2$				
0.00	0.018	0.018	0.018	0.018
0.25	0.018	0.018	0.018	0.018
0.50	0.018	0.018	0.018	0.018
0.75	0.017	0.017	0.017	0.017
0.95	0.015	0.015	0.015	0.015
1.00	0.015	0.014	0.014	0.014
$\rho_0 = 0.5$				
0.00	0.149	0.150	0.150	0.150
0.25	0.149	0.150	0.150	0.150
0.50	0.150	0.150	0.150	0.150
0.75	0.146	0.146	0.146	0.146
0.95	0.134	0.133	0.133	0.133
1.00	0.133	0.128	0.127	0.126
$\rho_0 = 0.75$				
0.00	0.463	0.464	0.464	0.464
0.25	0.463	0.464	0.464	0.465
0.50	0.467	0.466	0.465	0.465
0.75	0.467	0.464	0.463	0.463
0.95	0.463	0.458	0.454	0.453
1.00	0.483	0.471	0.469	0.467
$\rho_0 = 0.95$				
0.00	0.921	0.921	0.921	0.921
0.25	0.921	0.921	0.921	0.921
0.50	0.924	0.924	0.923	0.922
0.75	0.934	0.932	0.930	0.929
0.95	0.952	0.946	0.945	0.944
1.00	0.969	0.968	0.969	0.969

Notes: The DGP is given by (S.53). The first unit is δ -dominant, and the rest of the units are non-dominant. R_0^2 is computed by (59) in the paper assuming homoskedasticity. Note that $R_\beta^2 = R_0^2 + 0.1$, where R_β^2 is defined by (60) in the paper, and used in (62) in the paper to set the value of σ_v^2 .

Table S.26: Small sample properties of the BMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	500	100	300	500
$\rho_0 = 0.2$												
0.00	-8.07	-2.56	-1.54	-0.76	21.05	10.91	8.23	5.65	6.40	5.85	5.35	5.20
0.25	-8.00	-2.56	-1.54	-0.76	20.92	10.90	8.22	5.65	5.90	5.75	5.45	5.40
0.50	-8.03	-2.56	-1.53	-0.76	20.82	10.86	8.21	5.66	6.20	5.45	5.80	4.90
0.75	-8.46	-2.60	-1.51	-0.76	21.66	11.16	8.44	5.71	7.15	6.00	5.90	5.85
0.95	-10.86	-3.35	-2.00	-1.01	26.04	13.22	9.88	6.66	12.60	10.70	9.35	8.75
1.00	-12.40	-4.05	-2.47	-1.17	28.58	15.35	11.65	7.90	16.75	15.65	14.85	14.75
$\rho_0 = 0.5$												
0.00	-7.20	-2.23	-1.30	-0.66	16.68	8.20	6.07	4.15	5.90	6.00	5.65	5.10
0.25	-7.16	-2.23	-1.30	-0.66	16.56	8.19	6.06	4.14	5.80	6.05	5.60	5.05
0.50	-7.17	-2.25	-1.30	-0.66	16.48	8.17	6.06	4.16	6.15	5.80	5.50	5.05
0.75	-7.86	-2.39	-1.33	-0.69	17.57	8.53	6.31	4.23	7.00	6.20	6.05	5.45
S56	-11.44	-3.48	-1.98	-1.00	23.04	10.78	7.85	5.22	14.05	12.15	10.35	9.80
0.95	-13.91	-4.56	-2.72	-1.32	26.34	13.43	10.03	6.75	20.25	19.95	19.45	18.95
$\rho_0 = 0.75$												
0.00	-5.51	-1.64	-0.92	-0.48	11.33	5.15	3.71	2.51	6.00	6.05	5.25	4.95
0.25	-5.49	-1.64	-0.92	-0.48	11.23	5.15	3.71	2.51	5.80	6.05	5.30	4.90
0.50	-5.47	-1.65	-0.93	-0.48	11.14	5.12	3.71	2.52	5.90	6.00	5.30	5.20
0.75	-6.25	-1.84	-0.99	-0.52	12.31	5.46	3.92	2.60	7.45	6.30	5.80	5.45
0.95	-11.31	-3.29	-1.80	-0.90	19.20	7.83	5.45	3.51	17.55	14.70	13.70	11.95
1.00	-15.59	-5.34	-3.19	-1.57	23.97	11.36	8.29	5.52	27.75	31.60	31.90	31.95
$\rho_0 = 0.95$												
0.00	-2.95	-0.74	-0.40	-0.21	5.15	1.74	1.17	0.76	7.65	6.20	4.95	4.75
0.25	-2.93	-0.74	-0.40	-0.21	5.05	1.74	1.17	0.77	7.45	6.10	5.10	4.90
0.50	-2.87	-0.73	-0.40	-0.21	4.96	1.73	1.17	0.77	7.70	6.05	5.25	4.90
0.75	-3.46	-0.85	-0.44	-0.23	5.97	1.92	1.26	0.80	10.65	8.05	7.10	6.25
0.95	-14.96	-2.67	-1.23	-0.56	22.12	4.95	2.57	1.38	47.45	29.50	23.05	21.45
1.00	-33.90	-21.47	-17.70	-14.24	38.90	25.98	21.76	17.37	89.05	92.85	94.95	95.90

Notes: The DGP is given by (S.53) with homoskedastic Gaussian errors. The first column of \mathbf{W} is δ -dominant, and the rest of the columns are non-dominant. The BMM estimator is given by (38) in the paper. The power is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value.

Table S.27: Small sample properties of the ML estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-8.13	-2.63	-1.58	-0.80	21.07	10.90	8.22	5.66	6.70	5.65	5.30	5.15
0.25	-8.10	-2.61	-1.59	-0.79	20.95	10.90	8.21	5.65	6.35	5.55	5.30	5.20
0.50	-8.10	-2.61	-1.58	-0.79	20.85	10.85	8.20	5.66	6.30	5.40	5.70	5.00
0.75	-8.54	-2.66	-1.56	-0.79	21.70	11.15	8.43	5.70	7.30	6.05	5.90	5.60
0.95	-10.93	-3.40	-2.04	-1.04	26.09	13.23	9.89	6.65	12.90	10.75	9.30	8.65
1.00	-12.48	-4.13	-2.53	-1.25	28.61	15.39	11.67	7.89	16.95	15.80	14.95	14.75
$\rho_0 = 0.5$												
0.00	-7.29	-2.29	-1.34	-0.69	16.72	8.23	6.06	4.13	6.10	6.00	5.35	5.10
0.25	-7.24	-2.32	-1.34	-0.65	16.61	8.22	6.06	4.13	5.65	6.15	5.50	5.25
0.50	-7.27	-2.31	-1.36	-0.67	16.52	8.21	6.07	4.13	6.15	5.80	5.65	4.70
0.75	-7.93	-2.47	-1.38	-0.68	17.60	8.54	6.31	4.20	6.90	6.10	6.05	5.30
0.95	-11.54	-3.53	-2.00	-0.99	23.08	10.79	7.86	5.21	13.90	11.70	10.45	10.05
1.00	-13.99	-4.54	-2.69	-1.24	26.40	13.46	10.07	6.75	20.20	20.10	20.00	19.15
$\rho_0 = 0.75$												
0.00	-5.58	-1.71	-1.03	-0.62	11.41	5.19	3.76	2.57	5.85	6.70	6.10	6.00
0.25	-5.57	-1.69	-1.02	-0.62	11.31	5.19	3.75	2.57	5.70	6.70	5.75	5.60
0.50	-5.55	-1.71	-1.04	-0.61	11.22	5.18	3.76	2.58	5.55	7.05	6.20	5.70
0.75	-6.32	-1.89	-1.08	-0.65	12.37	5.51	3.96	2.62	6.90	7.50	6.15	5.55
0.95	-11.44	-3.42	-1.89	-0.97	19.26	7.90	5.48	3.53	17.45	15.35	13.95	11.25
1.00	-15.69	-5.38	-3.17	-1.56	23.98	11.35	8.30	5.59	27.45	31.50	33.00	33.75
$\rho_0 = 0.95$												
0.00	-2.78	-0.66	-0.34	-0.18	5.06	1.77	1.19	0.80	10.00	7.75	6.90	5.80
0.25	-2.77	-0.66	-0.34	-0.18	5.00	1.77	1.18	0.80	10.15	7.45	6.20	6.25
0.50	-2.72	-0.66	-0.35	-0.18	4.89	1.75	1.18	0.80	9.85	7.70	6.00	5.70
0.75	-3.26	-0.77	-0.39	-0.21	5.77	1.93	1.27	0.81	12.90	9.25	8.85	6.85
0.95	-10.49	-2.21	-1.04	-0.43	14.80	4.14	2.39	1.31	40.70	29.50	26.40	23.30
1.00	-18.39	-7.70	-5.11	-2.76	22.71	10.15	7.12	4.49	75.50	79.15	80.70	78.70

Notes: The DGP is given by (S.53) with homoskedastic Gaussian errors. See also the notes to Tables S.1 and S.26.

Table S.28: Small sample properties of the GMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)			
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	
$\rho_0 = 0.2$													
0.00	-8.96	-2.80	-1.63	-0.83	22.42	11.20	8.33	5.72	6.75	6.40	5.65	5.20	
0.25	-8.90	-2.80	-1.63	-0.83	22.25	11.19	8.33	5.72	6.70	6.35	5.65	5.20	
0.50	-8.94	-2.81	-1.62	-0.82	22.16	11.16	8.32	5.73	6.75	6.15	6.00	5.05	
0.75	-9.54	-2.91	-1.63	-0.84	23.40	11.53	8.59	5.79	8.15	6.50	6.25	6.05	
0.95	-12.14	-3.64	-2.10	-1.07	28.67	13.68	10.08	6.75	14.95	11.40	9.80	9.05	
1.00	-13.44	-4.19	-2.50	-1.20	31.47	15.91	11.89	8.00	19.60	16.45	15.30	14.85	
$\rho_0 = 0.5$													
0.00	-6.99	-2.14	-1.24	-0.62	17.10	8.25	6.09	4.16	6.45	6.25	5.70	5.35	
0.25	-6.94	-2.14	-1.23	-0.63	16.93	8.24	6.08	4.16	6.45	6.30	5.55	5.25	
0.50	-6.96	-2.15	-1.23	-0.62	16.83	8.21	6.08	4.17	6.85	6.05	5.75	5.20	
0.75	-7.57	-2.26	-1.25	-0.64	18.10	8.59	6.34	4.25	8.15	6.30	6.30	6.00	
S58	0.95	-10.43	-3.02	-1.73	-0.87	24.18	10.94	7.95	5.27	17.05	12.75	10.85	10.25
1.00	-12.17	-3.75	-2.23	-1.07	27.78	13.70	10.18	6.81	24.05	21.85	20.05	19.85	
$\rho_0 = 0.75$													
0.00	-4.67	-1.38	-0.79	-0.40	11.24	5.16	3.77	2.56	7.45	6.05	5.80	5.05	
0.25	-4.64	-1.38	-0.79	-0.40	11.11	5.15	3.77	2.56	7.35	6.20	5.60	5.05	
0.50	-4.64	-1.39	-0.79	-0.40	11.00	5.13	3.76	2.57	7.30	6.05	5.90	5.30	
0.75	-5.16	-1.48	-0.81	-0.41	12.21	5.46	3.98	2.65	9.00	6.90	6.70	6.10	
0.95	-8.29	-2.21	-1.24	-0.61	19.06	7.91	5.60	3.64	23.20	15.85	13.95	13.10	
1.00	-10.73	-3.23	-1.90	-0.91	23.65	11.39	8.40	5.59	35.80	34.60	33.40	32.15	
$\rho_0 = 0.95$													
0.00	-1.95	-0.49	-0.27	-0.13	4.71	1.85	1.31	0.88	11.85	7.45	6.35	5.50	
0.25	-1.93	-0.49	-0.27	-0.14	4.64	1.85	1.31	0.88	11.55	7.65	6.40	5.50	
0.50	-1.91	-0.49	-0.27	-0.14	4.55	1.83	1.31	0.88	11.55	7.85	6.30	5.75	
0.75	-2.26	-0.53	-0.28	-0.14	5.36	2.03	1.42	0.93	15.85	9.85	8.40	7.00	
0.95	-6.29	-1.13	-0.52	-0.23	12.76	3.95	2.64	1.63	44.20	30.80	25.10	21.55	
1.00	-10.59	-3.44	-2.01	-0.90	18.96	8.45	6.18	4.18	68.00	71.45	71.95	70.90	

Notes: The DGP is given by (S.53) with homoskedastic Gaussian errors. See also the notes to Tables S.3 and S.26.

Table S29: Small sample properties of the OGMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)			
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000	
$\rho_0 = 0.2$													
0.00	-5.72	-2.07	-1.29	-0.64	21.14	10.80	8.19	5.66	7.60	5.70	5.65	5.15	
0.25	-5.66	-2.07	-1.29	-0.64	21.12	10.79	8.18	5.65	7.20	5.50	5.80	5.20	
0.50	-5.77	-2.07	-1.27	-0.63	21.05	10.75	8.17	5.67	7.80	5.55	6.00	4.75	
0.75	-4.73	-1.86	-1.16	-0.58	22.50	11.16	8.44	5.73	10.05	6.40	6.10	5.75	
0.95	-3.28	-1.05	-0.83	-0.46	27.86	14.07	10.33	6.84	19.15	14.05	11.50	9.85	
1.00	-3.20	-1.21	-0.90	-0.40	30.96	16.58	12.62	8.47	23.85	20.40	18.85	18.15	
$\rho_0 = 0.5$													
0.00	-5.35	-1.90	-1.12	-0.58	16.64	8.12	6.04	4.15	7.35	6.00	5.70	5.05	
0.25	-5.34	-1.90	-1.12	-0.58	16.53	8.11	6.04	4.15	7.10	5.85	5.70	5.15	
0.50	-5.42	-1.91	-1.12	-0.57	16.55	8.08	6.03	4.16	8.00	5.80	5.75	4.90	
0.75	-4.91	-1.89	-1.10	-0.57	18.10	8.46	6.30	4.25	10.10	6.55	6.35	5.85	
S59	0.95	-3.84	-1.23	-0.87	-0.50	24.02	11.73	8.35	5.42	22.35	16.05	13.10	11.25
1.00	-4.21	-1.38	-0.99	-0.46	27.27	14.76	11.20	7.45	31.25	27.25	24.90	23.65	
$\rho_0 = 0.75$													
0.00	-4.21	-1.38	-0.78	-0.41	11.06	5.12	3.72	2.54	7.95	6.25	5.60	5.10	
0.25	-4.20	-1.38	-0.78	-0.41	10.94	5.11	3.71	2.54	7.60	6.20	5.70	5.10	
0.50	-4.21	-1.39	-0.78	-0.41	10.98	5.09	3.71	2.55	7.75	6.60	5.45	5.25	
0.75	-4.22	-1.48	-0.81	-0.43	12.30	5.43	3.93	2.63	10.75	7.45	5.95	6.05	
0.95	-4.23	-1.12	-0.74	-0.46	18.65	8.73	5.99	3.74	31.40	20.65	16.70	14.35	
1.00	-5.57	-1.57	-1.04	-0.49	22.44	12.28	9.32	6.25	42.60	41.40	40.20	37.80	
$\rho_0 = 0.95$													
0.00	-1.99	-0.53	-0.28	-0.15	4.65	1.80	1.24	0.83	11.60	7.30	5.95	5.50	
0.25	-1.98	-0.53	-0.28	-0.15	4.58	1.81	1.24	0.83	11.75	7.40	6.05	5.45	
0.50	-1.97	-0.53	-0.28	-0.15	4.50	1.79	1.24	0.84	11.45	7.70	6.05	5.40	
0.75	-2.27	-0.59	-0.30	-0.17	5.28	1.97	1.33	0.88	16.00	10.05	7.50	6.60	
0.95	-5.07	-1.05	-0.47	-0.23	11.93	3.90	2.58	1.56	53.35	33.70	26.55	22.75	
1.00	-8.35	-3.00	-1.82	-0.82	17.33	8.49	6.43	4.40	75.45	77.30	76.95	76.75	

S59

Notes: The DGP is given by (S.53) with homoskedastic Gaussian errors. See also the notes to Tables S.5 and S.26.

Table S.30: Small sample properties of the BGMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-7.84	-2.51	-1.52	-0.75	20.78	10.87	8.21	5.65	6.15	5.75	5.35	5.20
0.25	-7.78	-2.51	-1.51	-0.75	20.66	10.86	8.20	5.65	5.80	5.60	5.40	7.45
0.50	-7.82	-2.51	-1.51	-0.75	20.59	10.82	8.20	5.66	6.10	5.30	5.85	7.55
0.75	-8.20	-2.55	-1.48	-0.75	21.34	11.12	8.42	5.70	6.75	5.95	5.90	8.10
0.95	-10.57	-3.29	-1.97	-0.99	25.54	13.15	9.85	6.65	11.90	10.60	9.30	8.70
1.00	-12.23	-4.02	-2.46	-1.17	28.06	15.28	11.63	7.89	15.55	15.50	14.80	14.70
$\rho_0 = 0.5$												
0.00	-7.23	-2.25	-1.31	-0.67	16.58	8.19	6.06	4.14	5.75	6.00	5.65	5.10
0.25	-7.20	-2.25	-1.30	-0.67	16.48	8.19	6.06	4.14	5.60	6.05	5.60	5.00
0.50	-7.21	-2.26	-1.31	-0.67	16.41	8.16	6.05	4.16	6.05	5.80	5.50	5.05
0.75	-7.94	-2.41	-1.34	-0.70	17.47	8.52	6.31	4.23	6.85	6.10	6.05	5.45
0.95	-11.84	-3.56	-2.02	-1.02	23.10	10.75	7.84	5.22	13.30	11.75	10.30	9.80
1.00	-14.58	-4.67	-2.77	-1.34	26.11	13.41	10.02	6.75	18.00	19.60	19.35	18.95
$\rho_0 = 0.75$												
0.00	-5.71	-1.69	-0.95	-0.49	11.40	5.16	3.71	2.51	5.95	6.00	5.35	5.05
0.25	-5.70	-1.69	-0.95	-0.49	11.30	5.16	3.71	2.51	5.65	5.95	5.30	4.85
0.50	-5.67	-1.70	-0.96	-0.50	11.22	5.14	3.71	2.52	5.65	5.90	5.35	5.10
0.75	-6.62	-1.91	-1.03	-0.54	12.58	5.49	3.93	2.60	7.45	6.20	5.80	5.45
0.95	-13.41	-3.60	-1.93	-0.96	21.67	7.97	5.47	3.52	18.45	14.80	13.55	11.75
1.00	-20.19	-5.83	-3.37	-1.61	28.84	11.31	8.25	5.51	30.45	29.25	31.30	31.90
$\rho_0 = 0.95$												
0.00	-3.61	-0.80	-0.43	-0.22	6.06	1.77	1.18	0.77	9.55	6.00	4.80	4.80
0.25	-3.56	-0.80	-0.43	-0.22	6.02	1.77	1.18	0.77	9.45	5.95	4.90	4.80
0.50	-3.54	-0.80	-0.43	-0.22	6.73	1.76	1.18	0.77	9.60	5.95	5.20	4.85
0.75	-4.70	-0.93	-0.48	-0.24	10.83	1.98	1.28	0.82	15.15	8.40	7.25	6.50
0.95	-21.06	-2.80	-1.19	-0.56	38.29	7.23	2.79	1.43	65.85	38.20	27.60	22.15
1.00	-40.78	-26.42	-19.28	-7.64	57.55	51.18	44.91	19.50	99.75	99.85	99.00	95.70

Notes: The DGP is given by (S.53) with homoskedastic Gaussian errors. The BGMM estimator is computed using $\tilde{\mathbf{G}} - n^{-1}Tr(\tilde{\mathbf{G}})\mathbf{I}_n$ in the quadratic moment, where $\tilde{\mathbf{G}} = \mathbf{W}(\mathbf{I}_n - \hat{\rho}\mathbf{W})^{-1}$ and $\hat{\rho}$ denotes the first-step estimate. See also the note to Table S.26.

Table S.31: Small sample properties of the BMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-7.02	-1.91	-1.21	-0.63	20.40	10.09	7.89	5.52	6.30	4.60	4.85	5.35
0.25	-6.94	-1.89	-1.20	-0.63	20.41	10.08	7.88	5.52	6.40	4.65	4.95	5.25
0.50	-7.15	-1.92	-1.21	-0.62	20.44	10.17	7.88	5.51	6.45	4.75	4.85	5.20
0.75	-7.45	-2.08	-1.33	-0.66	21.24	10.61	8.23	5.71	7.55	5.15	5.25	5.55
0.95	-9.87	-2.66	-1.69	-0.85	25.67	12.82	9.84	6.70	11.45	9.95	10.55	8.60
1.00	-11.14	-3.12	-1.99	-0.98	28.54	14.60	11.38	7.90	15.85	13.90	15.35	14.25
$\rho_0 = 0.5$												
0.00	-6.36	-1.72	-1.06	-0.55	16.22	7.53	5.81	4.06	6.40	4.55	4.50	5.50
0.25	-6.33	-1.71	-1.06	-0.55	16.23	7.52	5.81	4.05	6.35	4.50	4.60	5.55
0.50	-6.50	-1.74	-1.07	-0.55	16.29	7.60	5.81	4.05	6.15	4.70	4.75	5.45
0.75	-7.03	-1.95	-1.24	-0.62	17.25	8.08	6.20	4.25	7.05	5.65	6.00	11.00
0.95	-10.50	-2.96	-1.81	-0.89	22.49	10.50	7.89	5.27	13.30	11.90	11.40	10.55
1.00	-12.75	-3.72	-2.30	-1.13	26.09	12.64	9.74	6.70	19.35	19.15	19.05	19.50
$\rho_0 = 0.75$												
0.00	-4.97	-1.32	-0.80	-0.41	11.06	4.70	3.56	2.47	6.55	4.35	4.95	5.00
0.25	-4.96	-1.31	-0.80	-0.41	11.08	4.70	3.56	2.47	6.90	4.30	4.95	4.90
0.50	-5.07	-1.33	-0.80	-0.41	11.16	4.74	3.56	2.47	6.65	4.45	4.90	5.15
0.75	-5.73	-1.56	-0.98	-0.48	12.19	5.20	3.90	2.63	7.60	5.75	6.60	6.45
0.95	-10.60	-3.01	-1.78	-0.86	18.70	7.76	5.57	3.60	17.40	15.25	14.85	12.95
1.00	-14.63	-4.65	-2.85	-1.41	23.57	10.59	8.00	5.45	26.45	29.90	31.55	31.70
$\rho_0 = 0.95$												
0.00	-2.75	-0.65	-0.39	-0.20	4.98	1.60	1.14	0.76	7.60	4.50	5.15	5.40
0.25	-2.75	-0.65	-0.39	-0.20	4.97	1.60	1.14	0.76	7.45	4.50	5.15	5.40
0.50	-2.75	-0.65	-0.39	-0.20	4.98	1.59	1.14	0.76	7.85	4.80	4.75	5.60
0.75	-3.39	-0.78	-0.47	-0.23	6.32	1.83	1.28	0.82	10.05	6.60	7.30	7.05
0.95	-14.38	-2.64	-1.32	-0.58	21.40	5.23	2.75	1.47	45.70	29.15	26.25	22.80
1.00	-33.52	-20.37	-17.33	-14.06	39.18	24.97	21.40	17.41	87.20	92.45	95.10	95.95

S61

Notes: The DGP is given by (S.53) with homoskedastic non-Gaussian errors. See also the notes to Table S.26.

Table S.32: Small sample properties of the ML estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-7.10	-1.97	-1.26	-0.65	20.41	10.07	7.87	5.53	6.50	4.20	4.80	5.15
0.25	-7.02	-1.95	-1.24	-0.66	20.44	10.07	7.87	5.53	6.60	4.10	4.75	5.10
0.50	-7.22	-1.98	-1.26	-0.66	20.45	10.16	7.87	5.52	6.75	4.65	4.75	5.00
0.75	-7.51	-2.13	-1.37	-0.70	21.25	10.59	8.21	5.72	7.85	5.10	5.05	5.40
0.95	-9.95	-2.71	-1.74	-0.90	25.70	12.83	9.84	6.69	12.05	9.60	10.35	8.75
1.00	-11.22	-3.19	-2.07	-1.06	28.58	14.61	11.39	7.90	16.35	13.60	14.95	14.20
$\rho_0 = 0.5$												
0.00	-6.47	-1.80	-1.10	-0.52	16.28	7.56	5.81	4.05	6.45	4.55	4.90	5.00
0.25	-6.42	-1.75	-1.11	-0.54	16.25	7.52	5.81	4.03	6.15	4.50	5.00	4.75
0.50	-6.59	-1.82	-1.13	-0.53	16.34	7.61	5.82	4.05	6.30	4.65	4.95	4.85
0.75	-7.10	-2.01	-1.26	-0.62	17.28	8.07	6.19	4.22	7.35	5.55	6.05	5.35
0.95	-10.57	-3.02	-1.83	-0.83	22.53	10.50	7.89	5.25	13.45	11.75	12.15	10.50
1.00	-12.84	-3.71	-2.27	-1.04	26.11	12.68	9.78	6.70	19.20	18.85	19.85	19.45
$\rho_0 = 0.75$												
0.00	-5.01	-1.36	-0.88	-0.53	11.11	4.72	3.61	2.54	6.20	4.95	5.50	6.15
0.25	-5.01	-1.37	-0.87	-0.55	11.12	4.72	3.58	2.53	6.25	5.05	5.45	5.60
0.50	-5.12	-1.39	-0.89	-0.54	11.22	4.77	3.63	2.54	6.30	5.55	5.55	6.25
0.75	-5.81	-1.63	-1.04	-0.61	12.26	5.22	3.91	2.68	7.45	6.70	7.10	6.90
0.95	-10.71	-3.13	-1.89	-0.92	18.75	7.83	5.62	3.60	17.50	16.15	15.95	12.40
1.00	-14.75	-4.70	-2.84	-1.40	23.60	10.61	8.03	5.51	26.45	29.90	32.65	33.05
$\rho_0 = 0.95$												
0.00	-2.57	-0.58	-0.34	-0.17	4.93	1.62	1.15	0.80	10.25	6.10	5.85	6.80
0.25	-2.58	-0.57	-0.34	-0.17	4.94	1.62	1.14	0.80	9.90	5.75	5.60	7.05
0.50	-2.58	-0.57	-0.33	-0.17	4.98	1.61	1.14	0.79	10.80	6.35	5.80	5.60
0.75	-3.14	-0.69	-0.42	-0.21	5.87	1.87	1.28	0.82	12.65	8.50	8.90	8.30
0.95	-10.22	-2.16	-1.14	-0.44	14.59	4.27	2.57	1.41	39.15	30.65	31.75	27.20
1.00	-17.73	-7.22	-4.90	-2.67	22.34	9.53	6.85	4.42	72.70	78.75	81.20	80.40

S62

Notes: The DGP is given by (S.53) with homoskedastic non-Gaussian errors. See also the notes to Table S.27.

Table S.33: Small sample properties of the GMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-7.84	-2.09	-1.30	-0.67	21.78	10.33	8.00	5.58	7.15	4.75	4.90	5.45
0.25	-7.77	-2.07	-1.29	-0.67	21.79	10.33	7.99	5.58	7.30	4.70	5.05	5.30
0.50	-7.98	-2.10	-1.30	-0.67	21.90	10.42	7.99	5.57	7.25	5.05	4.70	5.45
0.75	-8.33	-2.27	-1.43	-0.72	22.85	10.91	8.36	5.78	8.15	5.60	5.65	5.85
0.95	-10.76	-2.84	-1.78	-0.88	27.95	13.25	10.03	6.74	14.15	11.05	10.80	8.70
1.00	-11.90	-3.15	-1.99	-0.97	31.37	15.07	11.57	7.95	18.95	15.05	16.05	14.55
$\rho_0 = 0.5$												
0.00	-6.15	-1.61	-0.99	-0.51	16.63	7.60	5.84	4.06	6.85	4.80	4.95	5.45
0.25	-6.10	-1.59	-0.99	-0.51	16.65	7.60	5.84	4.06	7.20	4.70	4.90	5.45
0.50	-6.26	-1.62	-0.99	-0.51	16.71	7.67	5.84	4.06	7.00	4.90	4.75	5.40
0.75	-6.63	-1.77	-1.11	-0.55	17.74	8.13	6.18	4.24	8.10	5.90	5.65	5.80
0.95	-9.26	-2.40	-1.48	-0.72	23.52	10.64	7.91	5.25	16.75	13.30	12.25	10.10
1.00	-10.77	-2.83	-1.77	-0.87	27.60	12.90	9.85	6.74	24.75	20.15	20.30	19.45
$\rho_0 = 0.75$												
0.00	-4.13	-1.04	-0.64	-0.33	10.97	4.76	3.62	2.50	7.55	5.25	5.15	5.75
0.25	-4.11	-1.03	-0.63	-0.33	10.99	4.75	3.61	2.50	7.80	5.15	5.20	5.85
0.50	-4.20	-1.04	-0.64	-0.33	11.03	4.79	3.61	2.50	7.75	5.30	4.85	5.55
0.75	-4.55	-1.16	-0.72	-0.36	12.01	5.19	3.89	2.64	9.65	6.55	6.15	6.15
0.95	-7.41	-1.79	-1.08	-0.52	18.52	7.77	5.59	3.63	22.15	17.75	15.10	12.75
1.00	-9.56	-2.44	-1.51	-0.74	23.41	10.69	8.11	5.51	36.10	33.00	31.70	31.15
$\rho_0 = 0.95$												
0.00	-1.73	-0.36	-0.22	-0.11	4.59	1.73	1.27	0.86	12.65	7.60	6.75	6.10
0.25	-1.72	-0.36	-0.22	-0.11	4.60	1.73	1.27	0.86	13.00	7.50	6.70	6.20
0.50	-1.75	-0.36	-0.22	-0.11	4.60	1.73	1.27	0.86	12.80	7.75	6.95	6.00
0.75	-2.02	-0.40	-0.25	-0.12	5.33	1.96	1.41	0.93	16.35	10.40	8.65	7.70
0.95	-5.85	-0.96	-0.48	-0.18	12.42	3.96	2.67	1.66	43.20	31.40	26.60	22.55
1.00	-9.80	-2.86	-1.74	-0.78	18.65	7.72	5.84	4.06	67.60	70.25	70.85	71.20

S63

Notes: The DGP is given by (S.53) with homoskedastic non-Gaussian errors. See the notes to Table S.28.

Table S34: Small sample properties of the OGMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-4.94	-1.47	-0.98	-0.51	20.04	9.99	7.84	5.50	7.10	4.40	4.75	5.15
0.25	-4.84	-1.45	-0.96	-0.51	20.13	9.99	7.83	5.49	7.40	4.40	4.90	5.15
0.50	-5.15	-1.48	-0.97	-0.50	20.10	10.06	7.83	5.49	7.00	4.55	4.80	5.30
0.75	-4.43	-1.37	-0.96	-0.49	21.53	10.57	8.18	5.69	9.25	5.65	5.50	5.40
0.95	-3.59	-0.72	-0.65	-0.31	26.98	13.24	9.99	6.78	18.15	12.30	11.35	9.65
1.00	-3.41	-0.81	-0.73	-0.37	30.11	15.38	11.91	8.25	22.00	18.05	18.70	16.15
$\rho_0 = 0.5$												
0.00	-4.78	-1.39	-0.88	-0.46	15.86	7.47	5.79	4.04	6.90	4.65	5.00	5.45
0.25	-4.70	-1.38	-0.88	-0.46	15.92	7.47	5.79	4.04	7.00	4.60	5.00	5.35
0.50	-4.90	-1.41	-0.88	-0.45	15.95	7.54	5.79	4.04	6.85	4.70	4.85	5.60
0.75	-4.42	-1.39	-0.95	-0.48	17.26	8.05	6.14	4.23	9.60	6.25	5.80	5.85
0.95	-3.92	-1.00	-0.77	-0.38	23.16	10.88	8.04	5.34	20.85	15.55	13.40	10.50
1.00	-4.29	-1.02	-0.83	-0.44	26.59	13.55	10.39	7.12	28.90	23.70	23.90	22.10
$\rho_0 = 0.75$												
0.00	-3.73	-1.03	-0.63	-0.33	10.77	4.71	3.58	2.48	7.90	5.45	5.15	5.75
0.25	-3.69	-1.03	-0.63	-0.33	10.79	4.71	3.58	2.48	8.15	5.40	5.20	5.70
0.50	-3.82	-1.05	-0.64	-0.33	10.81	4.75	3.58	2.48	7.80	5.20	5.15	5.30
0.75	-3.61	-1.11	-0.74	-0.37	11.94	5.20	3.88	2.64	10.70	6.75	6.70	6.20
0.95	-4.26	-0.98	-0.74	-0.37	17.97	8.20	5.75	3.69	27.40	20.70	17.55	13.85
1.00	-5.33	-1.23	-0.84	-0.45	22.07	11.23	8.68	5.90	41.45	38.35	36.35	35.70
$\rho_0 = 0.95$												
0.00	-1.75	-0.40	-0.23	-0.12	4.52	1.68	1.22	0.83	12.75	7.10	6.85	6.20
0.25	-1.76	-0.40	-0.23	-0.12	4.53	1.67	1.22	0.82	12.95	7.10	6.90	6.00
0.50	-1.80	-0.40	-0.23	-0.12	4.55	1.68	1.22	0.83	12.85	7.15	7.10	6.00
0.75	-2.02	-0.46	-0.29	-0.14	5.24	1.92	1.37	0.90	17.30	10.15	9.15	7.60
0.95	-4.81	-0.87	-0.48	-0.20	11.68	3.94	2.64	1.62	50.50	33.95	30.40	24.05
1.00	-7.86	-2.50	-1.56	-0.70	17.11	7.71	6.00	4.22	73.35	74.80	76.45	75.80

S64

Notes: The DGP is given by (S.53) with homoskedastic non-Gaussian errors. See the notes to Table S.28.

Table S35: Small sample properties of the BGMM estimator of ρ from the SAR model without exogenous regressors for experiments with homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-6.82	-1.87	-1.19	-0.62	20.17	10.06	7.87	5.52	6.15	4.50	4.80	5.35
0.25	-6.75	-1.85	-1.18	-0.62	20.18	10.05	7.87	5.51	6.35	4.65	4.85	5.25
0.50	-6.94	-1.88	-1.19	-0.61	20.19	10.14	7.87	5.51	6.30	4.75	4.85	5.15
0.75	-7.18	-2.02	-1.30	-0.65	20.88	10.56	8.21	5.71	7.35	5.15	5.30	5.55
0.95	-9.58	-2.60	-1.65	-0.83	25.17	12.75	9.81	6.69	11.00	9.80	10.30	8.60
1.00	-10.92	-3.08	-1.97	-0.98	27.96	14.54	11.36	7.89	15.05	13.80	15.30	14.20
$\rho_0 = 0.5$												
0.00	-6.40	-1.73	-1.07	-0.55	16.15	7.52	5.81	4.06	6.15	4.45	4.50	5.45
0.25	-6.37	-1.72	-1.06	-0.55	16.16	7.51	5.81	4.05	6.10	4.30	4.60	5.55
0.50	-6.53	-1.75	-1.08	-0.55	16.22	7.59	5.81	4.05	6.00	4.65	4.75	5.45
0.75	-7.10	-1.98	-1.25	-0.62	17.13	8.07	6.19	4.25	6.80	5.70	6.10	6.00
0.95	-10.88	-3.04	-1.86	-0.91	22.19	10.46	7.88	5.27	12.25	11.75	11.50	10.55
1.00	-13.65	-3.82	-2.35	-1.15	25.95	12.60	9.73	6.70	17.60	18.90	18.80	19.40
$\rho_0 = 0.75$												
0.00	-5.18	-1.37	-0.82	-0.42	11.12	4.71	3.57	2.47	6.10	4.20	4.95	4.95
0.25	-5.18	-1.36	-0.82	-0.42	11.15	4.71	3.56	2.47	6.45	4.20	5.00	4.90
0.50	-5.28	-1.38	-0.83	-0.43	11.23	4.75	3.56	2.47	6.35	4.40	4.85	5.10
0.75	-6.17	-1.64	-1.02	-0.50	12.69	5.22	3.91	2.63	7.70	5.70	6.90	6.50
0.95	-12.88	-3.33	-1.93	-0.92	21.38	7.84	5.60	3.62	18.60	15.25	15.40	13.30
1.00	-20.05	-5.35	-3.05	-1.46	30.21	11.06	7.94	5.44	28.80	26.90	31.05	31.75
$\rho_0 = 0.95$												
0.00	-3.37	-0.72	-0.42	-0.21	6.03	1.62	1.15	0.77	9.50	4.55	5.15	5.30
0.25	-3.41	-0.72	-0.42	-0.21	6.42	1.62	1.15	0.77	9.70	4.45	5.05	5.30
0.50	-3.50	-0.71	-0.42	-0.21	8.57	1.62	1.15	0.77	10.35	4.70	4.70	5.50
0.75	-4.45	-0.86	-0.51	-0.25	9.59	1.94	1.30	0.82	15.50	7.40	7.70	7.50
0.95	-20.69	-2.96	-1.29	-0.57	37.62	8.04	3.05	1.55	65.05	38.50	30.85	24.55
1.00	-41.32	-29.87	-23.46	-8.92	58.28	56.70	53.57	22.70	99.75	99.75	95.05	58.55

Notes: The DGP is given by (S.53) with homoskedastic non-Gaussian errors. See the notes to Table S.28.

Table S.36: Small sample properties of the BMM estimator of ρ for experiments with two dominant units and homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-9.06	-2.72	-1.66	-0.82	20.84	10.43	7.87	5.40	7.40	5.30	7.45	15.05
0.25	-8.96	-2.72	-1.66	-0.83	20.67	10.42	7.86	5.40	7.00	5.40	7.00	14.85
0.50	-8.83	-2.73	-1.65	-0.83	20.39	10.40	7.85	5.41	7.00	5.85	5.95	5.25
0.75	-9.88	-2.98	-1.70	-0.86	21.92	10.94	8.15	5.51	7.90	6.30	5.95	5.25
0.95	-13.01	-4.06	-2.50	-1.23	27.80	14.14	10.60	7.13	13.90	11.40	11.95	10.20
1.00	-16.78	-5.65	-3.44	-1.61	34.29	18.74	14.43	9.82	22.45	21.80	22.25	19.75
$\rho_0 = 0.5$												
0.00	-8.08	-2.38	-1.41	-0.71	16.75	7.91	5.85	3.99	7.35	5.75	5.95	5.20
0.25	-8.02	-2.38	-1.40	-0.71	16.65	7.92	5.85	3.99	6.90	5.80	5.65	5.35
0.50	-7.96	-2.41	-1.41	-0.73	16.48	7.92	5.84	4.00	7.15	5.85	6.05	4.95
0.75	-9.33	-2.77	-1.54	-0.79	18.38	8.56	6.21	4.16	8.35	6.80	5.55	5.15
0.95	-14.86	-4.63	-2.73	-1.33	26.51	12.19	8.86	5.78	17.20	13.55	14.05	11.55
1.00	-20.82	-7.15	-4.34	-2.12	34.14	17.63	13.35	9.05	28.05	29.15	30.05	28.10
$\rho_0 = 0.75$												
0.00	-6.01	-1.70	-0.98	-0.50	11.32	4.94	3.56	2.41	7.25	5.80	5.50	5.00
0.25	-5.99	-1.70	-0.98	-0.50	11.30	4.95	3.55	2.41	6.70	5.80	5.55	4.95
0.50	-5.92	-1.72	-0.99	-0.51	11.16	4.94	3.55	2.42	6.95	5.80	5.55	5.00
0.75	-7.29	-2.08	-1.13	-0.59	13.06	5.52	3.86	2.56	9.15	6.80	6.25	5.45
0.95	-15.18	-4.39	-2.47	-1.18	23.73	9.12	6.26	3.89	22.75	17.80	17.65	14.80
1.00	-25.41	-9.58	-6.08	-3.17	34.45	16.48	12.20	8.18	40.50	44.65	46.05	46.75

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic Gaussian errors. $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. Both the first and second columns of \mathbf{W} are δ -dominant, and the rest of the columns are non-dominant. The BMM estimator is given by (38) in the paper. The power is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value.

Table S.37: Small sample properties of the BMM estimator of β for experiments with two dominant units and homoskedastic Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)			
				100	300	500	1,000	300	500	1,000	100	300	500
	$\rho_0 = 0.2$	100	300	500	1,000	300	500	1,000	100	300	500	100	300
0.00	3.40	1.27	0.63	0.51	39.68	20.86	15.79	10.93	7.30	5.20	4.50	4.35	11.70
0.25	3.40	1.30	0.64	0.51	39.64	20.88	15.79	10.94	7.55	5.30	4.50	4.40	11.45
0.50	3.33	1.27	0.58	0.46	39.97	20.97	15.86	11.00	7.50	5.10	4.45	4.70	11.05
0.75	3.14	1.17	0.57	0.52	41.46	21.57	16.28	11.34	7.75	5.50	4.45	4.75	11.65
0.95	1.32	0.95	0.44	0.52	47.91	26.50	20.91	16.61	6.75	4.95	4.35	4.55	8.95
1.00	0.49	-0.17	-0.35	0.19	53.00	31.93	27.18	22.29	6.55	5.10	4.85	4.60	8.10
<hr/>													
0.00	4.76	1.77	0.87	0.68	47.26	24.55	18.53	12.81	7.40	5.35	4.60	4.40	10.80
0.25	4.77	1.80	0.88	0.68	47.15	24.55	18.52	12.82	7.55	5.70	4.65	4.40	10.85
0.50	4.68	1.78	0.83	0.63	47.83	24.83	18.74	12.98	7.35	5.65	4.25	4.50	10.65
0.75	4.68	1.77	0.84	0.76	52.73	27.91	21.47	15.27	7.85	5.50	4.80	4.50	10.55
0.95	2.30	1.77	0.69	0.99	77.06	50.22	43.57	39.89	7.15	5.00	4.60	4.70	8.10
1.00	0.95	-0.46	-0.99	0.44	93.69	69.98	66.69	60.58	6.65	5.30	4.95	4.55	7.15
<hr/>													
0.00	6.31	2.28	1.15	0.84	49.68	25.49	19.19	13.26	7.70	5.60	4.85	4.35	11.40
0.25	6.32	2.32	1.17	0.84	49.57	25.50	19.19	13.27	7.85	5.75	4.85	4.45	10.90
0.50	6.19	2.34	1.14	0.81	50.31	26.03	19.62	13.57	7.85	5.90	4.50	4.30	10.75
0.75	6.52	2.49	1.18	1.03	60.78	33.36	26.34	19.25	8.30	5.50	5.10	4.45	10.20
0.95	3.66	3.01	1.11	1.66	109.27	78.51	70.75	66.65	7.80	5.35	5.35	4.80	8.25
1.00	1.81	-0.45	-1.33	1.07	134.65	113.43	110.76	103.88	6.75	5.50	5.20	4.60	6.65
<hr/>													

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.36.

Table S.38: Small sample properties of the BMM estimator of ρ for experiments with two dominant units and homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-8.26	-2.40	-1.45	-0.72	20.46	9.91	7.58	5.30	6.80	4.45	5.10	5.45
0.25	-8.17	-2.38	-1.45	-0.73	20.40	9.89	7.58	5.30	6.55	4.35	5.20	5.25
0.50	-8.28	-2.48	-1.52	-0.77	20.13	10.01	7.63	5.35	6.60	4.65	5.20	5.50
0.75	-9.13	-2.70	-1.75	-0.79	21.57	10.51	8.08	5.54	8.00	5.45	6.45	6.35
0.95	-12.82	-3.89	-2.26	-1.16	28.61	13.83	10.46	7.19	15.90	12.15	11.85	10.60
1.00	-15.72	-4.84	-2.89	-1.42	34.50	17.98	13.87	9.98	23.70	19.40	20.15	21.30
$\rho_0 = 0.5$												
0.00	-7.44	-2.11	-1.26	-0.63	16.52	7.50	5.65	3.92	6.35	4.20	4.75	5.50
0.25	-7.40	-2.11	-1.26	-0.63	16.50	7.49	5.65	3.92	6.55	4.30	5.00	5.65
0.50	-7.50	-2.18	-1.32	-0.67	16.35	7.59	5.69	3.96	6.80	4.40	5.25	5.95
0.75	-8.62	-2.52	-1.62	-0.73	18.01	8.22	6.24	4.19	8.25	5.40	7.15	6.05
0.95	-14.78	-4.43	-2.65	-1.30	27.11	11.99	8.88	5.86	17.90	13.80	13.60	12.45
1.00	-19.91	-6.34	-3.86	-1.90	34.22	16.70	12.78	9.08	28.20	26.75	28.05	28.30
$\rho_0 = 0.75$												
0.00	-5.63	-1.55	-0.91	-0.45	11.27	4.70	3.46	2.37	6.80	4.30	4.75	5.70
0.25	-5.63	-1.55	-0.92	-0.46	11.27	4.69	3.46	2.37	6.65	4.50	4.70	5.60
0.50	-5.64	-1.59	-0.95	-0.48	11.15	4.73	3.47	2.40	6.60	4.55	5.35	5.80
0.75	-6.78	-1.92	-1.22	-0.56	12.84	5.34	3.96	2.60	8.80	6.20	6.90	6.10
0.95	-15.34	-4.28	-2.53	-1.19	24.28	9.16	6.41	4.01	23.55	17.75	19.15	16.35
1.00	-24.72	-8.84	-5.64	-2.95	34.40	15.53	11.66	8.11	38.45	42.45	45.10	46.50

S68

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic non-Gaussian errors. See also the notes to Table S.36.

Table S.39: Small sample properties of the BMM estimator of β for experiments with two dominant units and homoskedastic non-Gaussian errors

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)						
	100			300			1,000			300						
	500	1,000	300	500	1,000	300	100	300	500	1,000	100	300	500	1,000		
$\rho_0 = 0.2$																
0.00	4.77	1.47	1.31	0.53	40.69	21.23	16.04	11.37	6.75	5.30	5.80	6.15	11.70	18.65	26.55	45.60
0.25	4.66	1.43	1.29	0.53	40.63	21.22	16.02	11.35	6.85	5.35	5.65	6.10	11.60	18.45	26.70	45.65
0.50	4.78	1.44	1.29	0.53	40.82	21.31	16.13	11.39	6.75	5.65	5.55	5.85	11.45	18.10	26.80	45.80
0.75	5.04	1.68	1.42	0.62	42.38	21.97	16.72	11.81	7.00	5.35	5.50	6.20	11.65	17.65	25.75	42.80
0.95	4.43	1.60	1.56	0.59	49.84	27.69	21.68	17.51	7.05	5.95	5.65	5.75	10.60	15.10	18.85	24.55
1.00	3.50	1.28	1.58	0.52	54.95	32.59	27.53	23.48	6.45	5.75	5.15	5.55	9.00	11.25	12.85	15.55
$\rho_0 = 0.5$																
0.00	6.38	1.96	1.67	0.70	48.51	25.00	18.82	13.31	7.10	5.40	5.80	6.20	10.95	15.40	21.35	35.50
0.25	6.27	1.91	1.65	0.69	48.39	24.98	18.79	13.29	7.10	5.25	5.80	6.10	10.90	14.95	21.15	35.55
0.50	6.50	1.94	1.66	0.70	48.92	25.26	19.06	13.42	6.85	5.65	5.75	5.95	10.55	14.80	21.10	35.15
0.75	7.14	2.38	1.97	0.88	53.94	28.44	22.09	15.92	7.35	5.40	5.15	6.20	10.05	13.65	17.85	26.30
0.95	7.47	2.92	3.07	1.26	80.06	52.48	45.11	42.03	7.45	5.65	5.70	6.15	9.00	9.05	8.80	8.80
1.00	6.33	2.76	3.82	1.39	97.12	71.42	67.53	63.91	6.65	5.90	5.35	5.60	7.35	7.10	6.05	6.85
$\rho_0 = 0.75$																
0.00	8.02	2.48	2.00	0.86	51.15	25.97	19.46	13.73	7.40	5.40	5.75	6.05	11.00	15.20	20.20	34.15
0.25	7.93	2.44	1.98	0.85	51.05	25.96	19.44	13.71	7.60	5.45	5.70	5.80	10.95	15.05	20.50	34.05
0.50	8.25	2.47	2.01	0.86	51.70	26.47	19.93	14.01	7.10	5.45	5.60	5.80	10.45	14.65	20.35	32.95
0.75	9.43	3.16	2.59	1.18	62.33	34.03	27.12	20.09	7.65	5.50	5.35	6.35	10.20	11.95	13.90	19.70
0.95	11.39	4.81	4.95	2.19	113.46	81.88	72.92	70.17	7.80	5.95	5.75	6.20	9.40	7.70	7.15	7.10
1.00	9.47	4.67	6.52	2.71	139.53	115.78	112.12	109.57	6.65	5.95	5.45	5.70	7.05	6.25	5.40	6.15

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.38.

Table S.40: Small sample properties of the BMM estimators of ρ and β for the experiments with exponentially decaying $\delta_{(i)}$, where $\delta_{(j)} = 0.9^j$

$\rho_0 \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
Experiments with homoskedastic Gaussian errors												
$\rho_0 = 0.2$	-3.95	-1.31	-0.76	-0.38	13.36	7.19	5.39	3.68	8.10	6.95	6.50	6.60
$\rho_0 = 0.5$	-5.53	-1.73	-0.99	-0.52	13.14	6.69	4.99	3.29	10.30	8.70	8.00	7.05
$\rho_0 = 0.75$	-7.07	-2.04	-1.14	-0.61	12.54	5.59	4.04	2.60	15.95	10.85	10.60	8.15
β												
$\rho_0 = 0.2$	1.07	0.73	-0.01	-0.02	45.15	25.09	20.01	15.30	6.45	6.15	5.20	4.55
$\rho_0 = 0.5$	2.52	1.32	-0.04	0.01	66.73	41.94	36.76	31.29	7.10	6.20	5.20	4.40
$\rho_0 = 0.75$	4.94	2.21	0.23	0.26	80.54	53.02	47.56	40.94	7.90	6.80	5.55	4.45
Experiments with homoskedastic non-Gaussian errors												
$\rho_0 = 0.2$	-3.04	-1.04	-0.67	-0.38	12.75	7.02	5.45	3.70	7.10	6.45	7.65	6.55
$\rho_0 = 0.5$	-4.58	-1.37	-0.93	-0.51	12.40	6.38	4.93	3.32	10.20	7.60	8.75	7.35
$\rho_0 = 0.75$	-6.17	-1.74	-1.12	-0.57	11.92	5.26	3.98	2.62	15.60	9.75	10.65	9.55
β												
$\rho_0 = 0.2$	4.55	1.57	1.50	0.48	48.16	25.53	20.28	15.32	7.65	5.75	5.35	5.30
$\rho_0 = 0.5$	7.40	2.68	2.77	0.94	70.90	42.67	37.19	31.47	7.75	5.90	5.40	5.40
$\rho_0 = 0.75$	10.57	3.93	3.93	1.28	85.18	54.00	48.15	41.38	8.90	6.50	5.55	5.55

S70

Notes: The DGP is given by (S.53) and (S.54). $\mathbf{W}_x = \mathbf{W}_y = \mathbf{W}$. The sum of the j^{th} column of \mathbf{W} rises with n at the rate of δ_j , where $\delta_j = 0.9^j$, for $j = 1, 2, \dots, n$. The BMM estimator is given by (38) in the paper. The power of $\hat{\rho}$ is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value. The power of $\hat{\beta}$ is calculated at 0.8.

Table S.41: Small sample properties of the BMM estimator of ρ for the experiments with homoskedastic Gaussian errors and $\mathbf{W}_x \neq \mathbf{W}_y$

$\delta_y \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
	$\rho_0 = 0.2$											
0.00	-8.56	-2.60	-1.60	-0.79	20.84	10.61	8.03	5.49	6.80	5.65	5.60	7.30
0.25	-8.49	-2.59	-1.59	-0.79	20.72	10.59	8.02	5.49	6.85	5.50	5.35	7.35
0.50	-8.53	-2.60	-1.58	-0.78	20.62	10.56	8.01	5.50	6.55	5.70	5.75	4.80
0.75	-8.93	-2.65	-1.57	-0.78	21.45	10.87	8.26	5.57	7.00	5.85	5.60	5.65
0.95	-11.36	-3.41	-2.10	-1.05	25.74	12.99	9.77	6.58	12.15	10.55	9.65	8.95
1.00	-13.04	-4.20	-2.63	-1.25	28.45	15.14	11.55	7.83	15.85	15.80	14.95	13.90
<hr/>												
0.00	-7.63	-2.28	-1.35	-0.68	16.67	8.04	5.96	4.06	6.40	5.60	5.60	5.25
0.25	-7.59	-2.28	-1.35	-0.68	16.57	8.03	5.96	4.06	6.40	5.60	5.60	5.25
0.50	-7.60	-2.29	-1.35	-0.69	16.47	8.00	5.95	4.07	6.60	5.80	5.80	4.90
0.75	-8.28	-2.46	-1.40	-0.72	17.58	8.40	6.24	4.17	7.00	6.40	6.00	5.15
0.95	-12.01	-3.62	-2.10	-1.06	23.14	10.79	7.88	5.23	13.90	12.10	10.65	10.30
1.00	-14.71	-4.80	-2.91	-1.40	26.69	13.48	10.07	6.77	20.60	20.05	19.50	19.45
<hr/>												
	$\rho_0 = 0.75$											
0.00	-5.77	-1.66	-0.96	-0.49	11.34	5.06	3.65	2.47	6.00	5.70	5.65	4.80
0.25	-5.75	-1.66	-0.96	-0.49	11.27	5.05	3.65	2.47	5.90	5.90	5.50	4.80
0.50	-5.72	-1.67	-0.96	-0.50	11.12	5.03	3.64	2.48	6.15	6.00	5.45	4.70
0.75	-6.52	-1.88	-1.03	-0.54	12.34	5.42	3.90	2.57	7.60	6.25	5.85	4.95
0.95	-11.77	-3.40	-1.88	-0.94	19.45	7.89	5.50	3.52	18.65	15.25	13.80	12.10
1.00	-16.31	-5.55	-3.34	-1.65	24.48	11.47	8.36	5.55	28.85	32.70	31.90	32.30
<hr/>												

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic Gaussian errors. \mathbf{W}_x and \mathbf{W}_y are generated following the same structure with the first column being δ_x — and δ_y —dominant, respectively, and the remaining columns are non-dominant. In all experiments $\delta_x = 1$. The BMM estimator is given by (38) in the paper. The power is calculated at $\rho_0 = 0.1$, where ρ_0 denotes the true value.

Table S.42: Small sample properties of the BMM estimator of β for the experiments with homoskedastic Gaussian errors and $\mathbf{W}_x \neq \mathbf{W}_y$

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	0.78	0.15	-0.08	0.18	47.04	25.19	18.94	13.12	6.50	5.25	4.60	4.40
0.25	0.75	0.15	-0.09	0.18	47.06	25.20	18.94	13.12	6.45	5.25	4.65	4.35
0.50	0.77	0.13	-0.09	0.18	47.17	25.21	18.95	13.14	6.55	5.30	4.70	4.40
0.75	0.65	0.15	-0.10	0.17	47.59	25.61	19.33	13.51	6.35	4.95	4.60	4.40
0.95	0.71	0.09	-0.16	0.20	49.89	28.71	22.61	17.30	6.50	5.40	4.95	4.50
1.00	0.76	0.11	-0.18	0.20	51.37	30.73	25.26	20.58	6.65	5.45	4.80	4.65
$\rho_0 = 0.5$												
0.00	1.42	0.36	-0.01	0.27	57.61	30.48	22.87	15.83	6.75	5.25	4.50	4.40
0.25	1.38	0.36	-0.02	0.27	57.62	30.48	22.87	15.83	6.65	5.15	4.50	4.40
0.50	1.41	0.33	-0.02	0.27	57.93	30.67	22.99	15.93	6.75	5.15	4.70	4.30
0.75	1.27	0.36	-0.08	0.27	61.18	33.75	25.90	18.65	6.75	5.10	4.75	4.45
0.95	1.34	0.16	-0.46	0.36	77.47	53.36	45.77	39.72	6.65	5.35	5.20	4.45
1.00	1.43	0.11	-0.64	0.39	86.41	64.61	59.03	54.50	6.90	5.35	5.00	4.50
$\rho_0 = 0.75$												
0.00	2.20	0.62	0.11	0.36	64.42	33.59	25.13	17.37	6.80	5.35	4.55	4.55
0.25	2.18	0.62	0.11	0.36	64.38	33.58	25.12	17.37	6.85	5.35	4.50	4.50
0.50	2.18	0.59	0.10	0.37	64.71	34.01	25.44	17.63	6.90	5.15	4.65	4.55
0.75	2.04	0.62	-0.03	0.38	72.56	41.50	32.56	24.16	7.20	5.15	4.90	4.50
0.95	2.14	0.31	-0.78	0.54	108.21	81.72	73.04	65.75	6.95	5.25	5.20	4.35
1.00	2.59	0.41	-0.95	0.84	122.97	102.62	97.01	92.87	7.25	5.35	5.15	4.55

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.41.

Table S43: Small sample properties of the BMM estimator of ρ for the experiments with homoskedastic non-Gaussian errors and $\mathbf{W}_x \neq \mathbf{W}_y$

$\delta_y \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	-7.57	-2.15	-1.33	-0.68	20.30	9.98	7.71	5.39	6.65	4.05	5.00	5.35
0.25	-7.49	-2.12	-1.31	-0.68	20.28	9.96	7.71	5.39	6.60	4.05	5.05	5.25
0.50	-7.69	-2.17	-1.34	-0.68	20.29	10.08	7.74	5.40	6.50	4.30	5.10	5.40
0.75	-8.01	-2.30	-1.45	-0.71	21.19	10.50	8.07	5.60	8.00	5.40	5.65	5.80
0.95	-10.42	-2.96	-1.83	-0.92	25.54	12.75	9.67	6.64	12.70	9.95	10.00	9.20
1.00	-11.87	-3.48	-2.18	-1.09	28.48	14.52	11.25	7.83	16.70	13.70	15.25	13.50
$\rho_0 = 0.5$												
0.00	-6.85	-1.91	-1.16	-0.59	16.33	7.52	5.74	3.99	6.30	4.40	4.65	5.55
0.25	-6.81	-1.90	-1.16	-0.59	16.33	7.51	5.73	3.98	6.40	4.40	4.65	5.50
0.50	-6.97	-1.94	-1.17	-0.59	16.37	7.60	5.75	4.00	6.45	4.65	4.95	5.50
0.75	-7.52	-2.14	-1.35	-0.66	17.44	8.09	6.15	4.21	7.35	5.70	6.30	6.05
0.95	-11.09	-3.23	-1.97	-0.96	22.78	10.61	7.89	5.29	14.30	12.00	11.15	10.75
1.00	-13.59	-4.08	-2.52	-1.24	26.58	12.77	9.79	6.72	20.35	18.60	19.30	19.70
$\rho_0 = 0.75$												
0.00	-5.30	-1.44	-0.86	-0.44	11.23	4.72	3.53	2.43	5.90	4.35	4.65	5.25
0.25	-5.29	-1.44	-0.86	-0.44	11.24	4.72	3.53	2.43	6.15	4.20	4.65	5.25
0.50	-5.38	-1.45	-0.87	-0.44	11.30	4.76	3.54	2.44	6.35	4.25	4.90	5.20
0.75	-6.06	-1.68	-1.06	-0.51	12.46	5.23	3.91	2.62	7.15	5.55	6.75	6.20
0.95	-11.08	-3.18	-1.89	-0.91	19.10	7.88	5.61	3.63	17.60	15.20	15.30	12.75
1.00	-15.39	-4.94	-3.04	-1.50	24.26	10.78	8.09	5.48	27.35	30.65	31.60	32.20

Notes: The DGP is given by (S.53) and (S.54) with homoskedastic non-Gaussian errors. See also the notes to Table S.41.

Table S.44: Small sample properties of the BMM estimator of β for the experiments with homoskedastic non-Gaussian errors and $\mathbf{W}_x \neq \mathbf{W}_y$

$\delta \setminus n$	Bias($\times 100$)			RMSE($\times 100$)			Size($\times 100$)			Power($\times 100$)		
	100	300	500	1,000	100	300	500	1,000	100	300	500	1,000
$\rho_0 = 0.2$												
0.00	3.06	1.01	1.10	0.37	48.45	25.52	19.21	13.76	6.10	6.05	5.40	5.55
0.25	3.05	1.00	1.09	0.36	48.44	25.52	19.21	13.76	6.15	6.05	5.40	5.55
0.50	3.07	1.00	1.09	0.36	48.47	25.55	19.24	13.78	6.10	5.95	5.45	5.50
0.75	3.06	1.02	1.13	0.37	49.01	25.98	19.65	14.16	6.25	5.65	5.30	5.75
0.95	3.18	1.16	1.27	0.45	51.48	29.15	22.92	18.18	6.40	5.90	5.45	5.95
1.00	3.32	1.22	1.44	0.55	52.90	31.19	25.65	21.66	6.10	6.00	5.55	5.80
$\rho_0 = 0.5$												
0.00	4.20	1.38	1.41	0.48	59.35	30.92	23.20	16.59	6.10	6.00	5.40	5.55
0.25	4.19	1.37	1.41	0.48	59.31	30.91	23.19	16.59	6.15	6.00	5.40	5.55
0.50	4.22	1.36	1.41	0.48	59.55	31.12	23.35	16.70	6.20	5.90	5.55	5.55
0.75	4.34	1.48	1.59	0.54	63.00	34.26	26.35	19.56	6.45	5.70	5.30	5.75
0.95	5.15	2.14	2.50	0.95	79.90	54.23	46.44	41.84	6.55	5.70	5.40	5.85
1.00	5.69	2.51	3.25	1.34	88.93	65.63	60.03	57.48	6.25	5.75	5.45	5.90
$\rho_0 = 0.75$												
0.00	5.30	1.72	1.68	0.60	66.38	34.12	25.49	18.19	6.35	5.85	5.45	5.60
0.25	5.29	1.71	1.67	0.59	66.31	34.10	25.48	18.19	6.35	5.85	5.50	5.60
0.50	5.34	1.71	1.68	0.60	66.57	34.55	25.82	18.47	6.40	5.85	5.55	5.60
0.75	5.67	1.95	2.07	0.74	74.74	42.15	33.15	25.36	6.40	5.70	5.45	5.80
0.95	7.51	3.33	3.94	1.56	111.51	83.07	74.14	69.35	6.70	5.70	5.45	5.85
1.00	8.55	4.15	5.49	2.36	126.49	104.27	98.79	97.91	6.35	5.70	5.60	6.05

Notes: The true value is $\beta_0 = 1$ and power is calculated at 0.8. See the notes to Table S.43.

Plots of empirical power functions

List of Figures

S.1	Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 100$, and homoskedastic Gaussian errors for different values of δ	S77
S.2	Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 300$, and homoskedastic Gaussian errors for different values of δ	S80
S.3	Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 500$, and homoskedastic Gaussian errors for different values of δ	S83
S.4	Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 1,000$, and homoskedastic Gaussian errors for different values of δ	S86
S.5	Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and homoskedastic Gaussian errors for different values of δ	S89
S.6	Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and homoskedastic Gaussian errors for different values of δ	S92
S.7	Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 500$, and homoskedastic Gaussian errors for different values of δ	S95
S.8	Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 1,000$, and homoskedastic Gaussian errors for different values of δ	S98
S.9	Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 100$, and homoskedastic Gaussian errors for different values of δ	S101
S.10	Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 300$, and homoskedastic Gaussian errors for different values of δ	S104
S.11	Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 500$, and homoskedastic Gaussian errors for different values of δ	S107
S.12	Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 1,000$, and homoskedastic Gaussian errors for different values of δ	S110
S.13	Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic Gaussian errors for different values of δ	S113
S.14	Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic Gaussian errors for different values of δ	S114
S.15	Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic Gaussian errors for different values of δ	S115
S.16	Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic Gaussian errors for different values of δ	S116
S.17	Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic non-Gaussian errors for different values of δ	S117

S.18 Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic non-Gaussian errors for different values of δ	S118
S.19 Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic non-Gaussian errors for different values of δ	S119
S.20 Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic non-Gaussian errors for different values of δ	S120

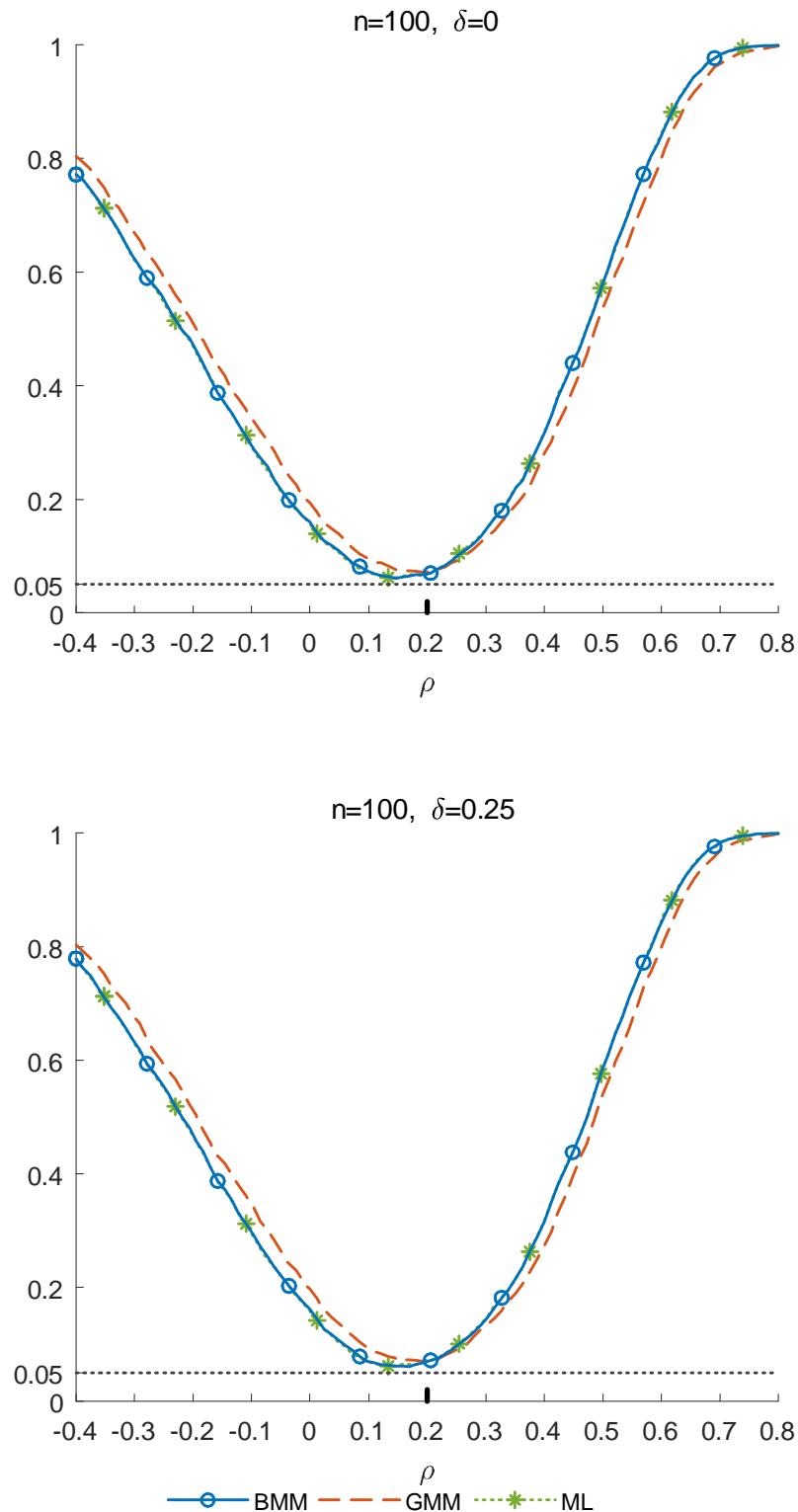


Figure S.1: Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 100$, and homoskedastic Gaussian errors for different values of δ

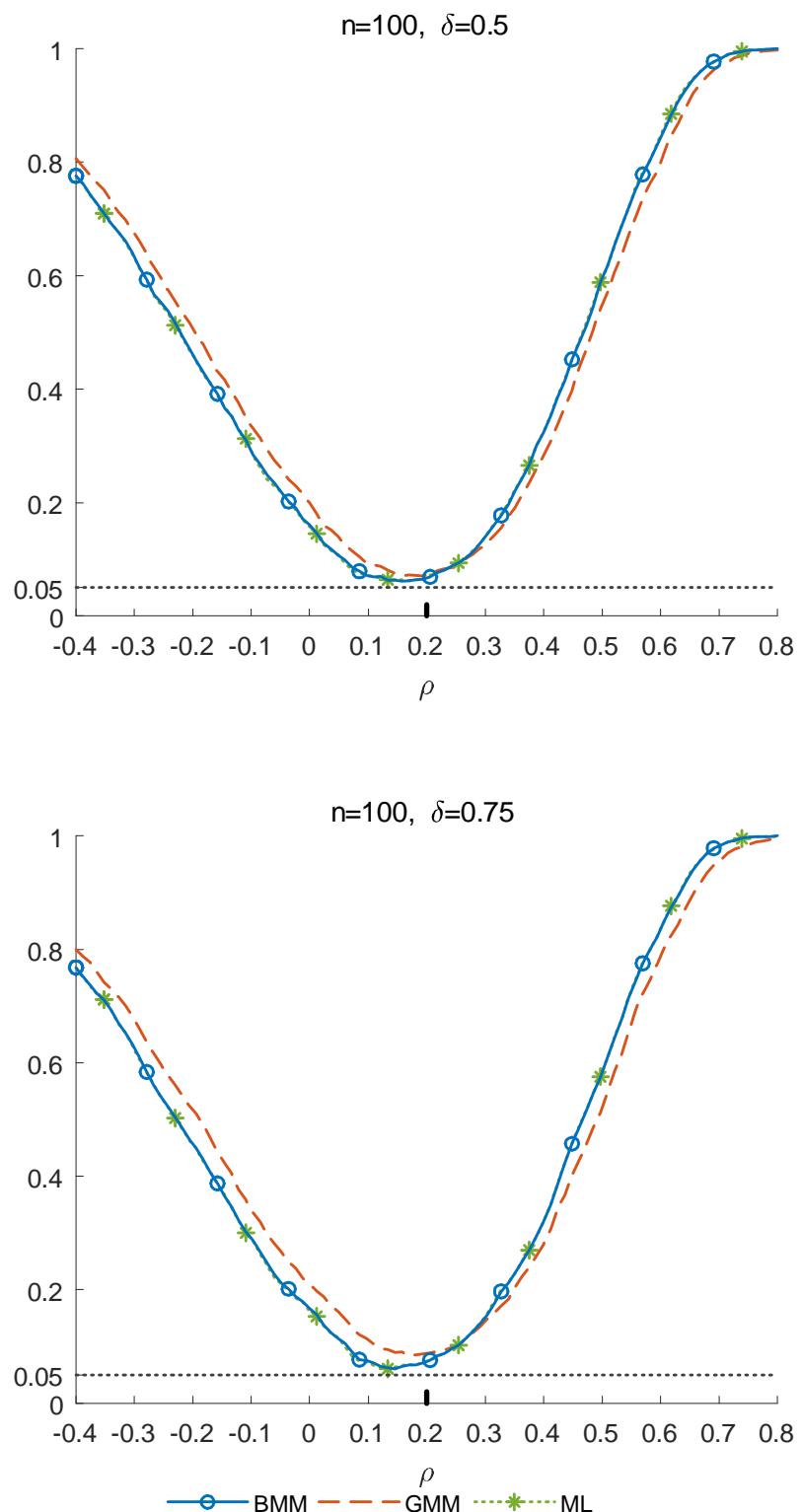


Figure S.1: (Continued)

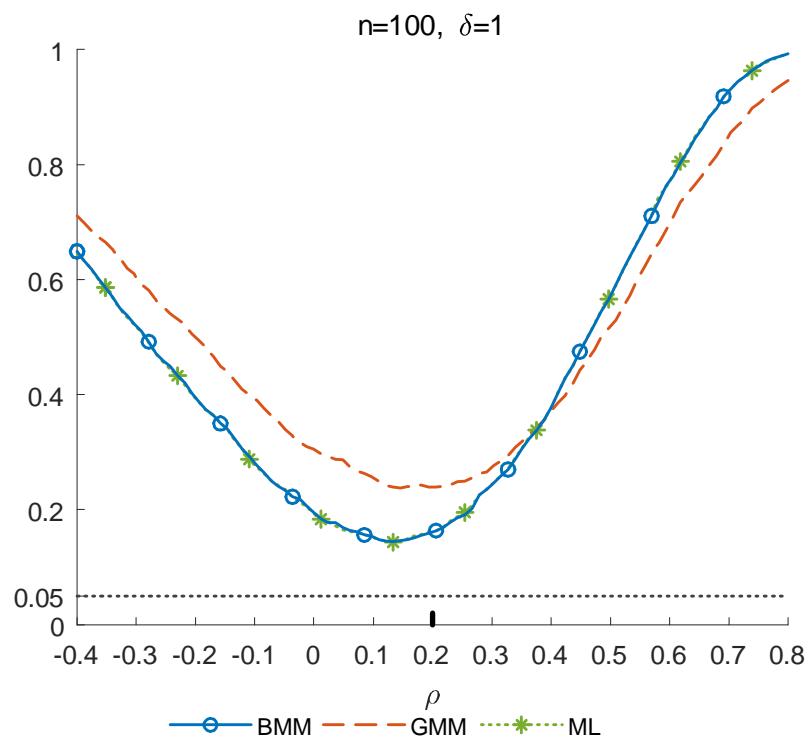
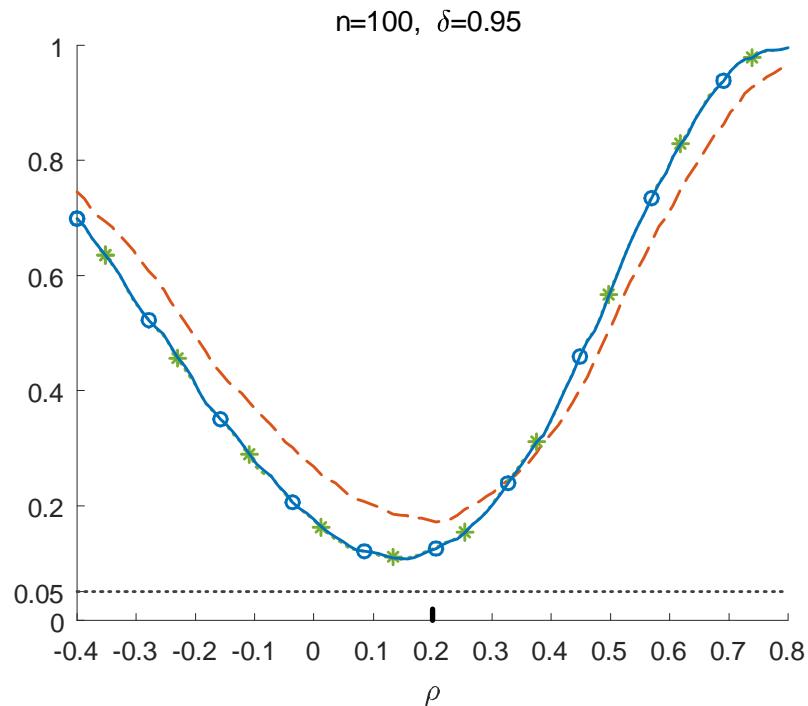


Figure S.1: (Continued)

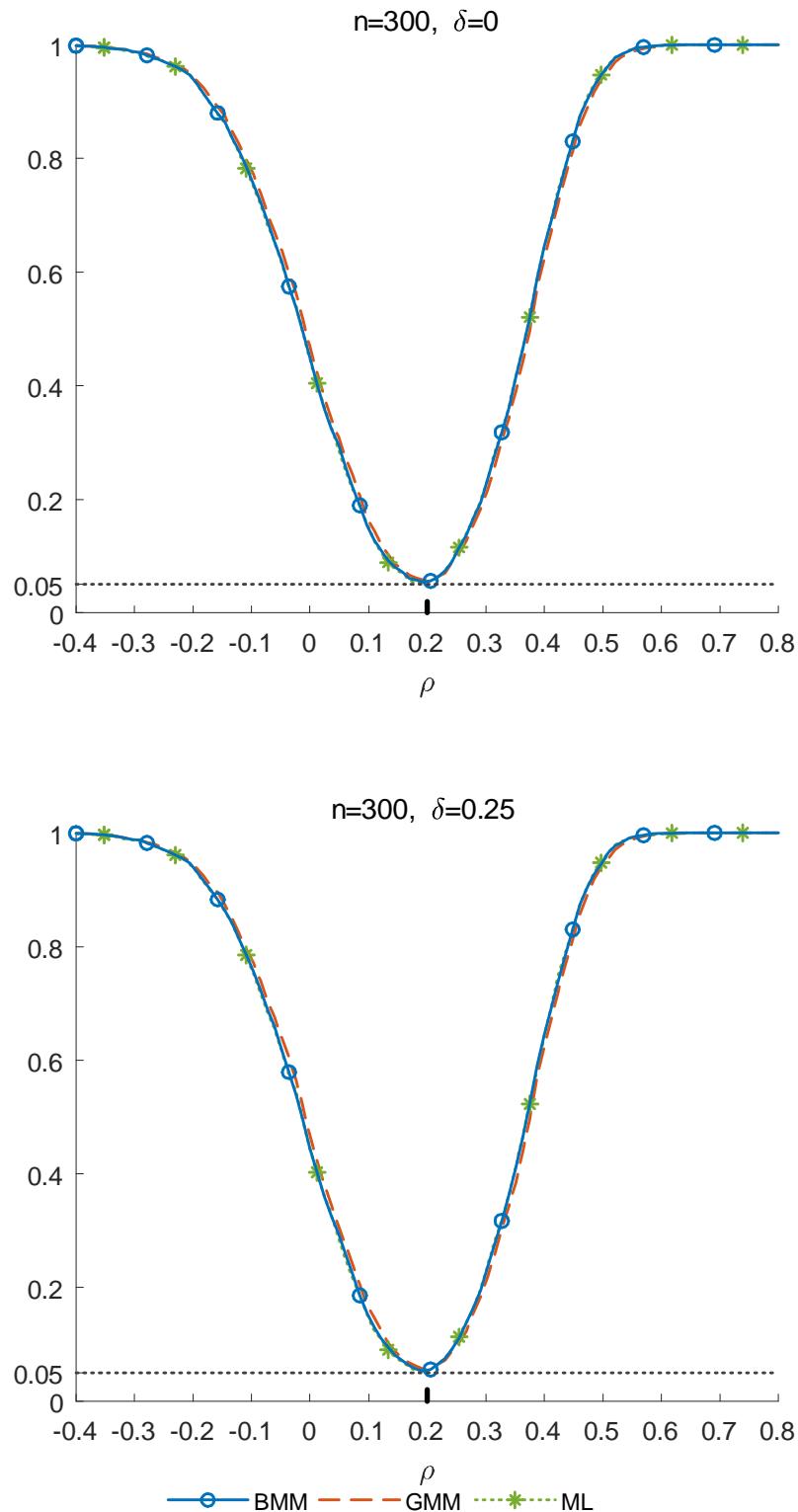


Figure S.2: Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 300$, and homoskedastic Gaussian errors for different values of δ

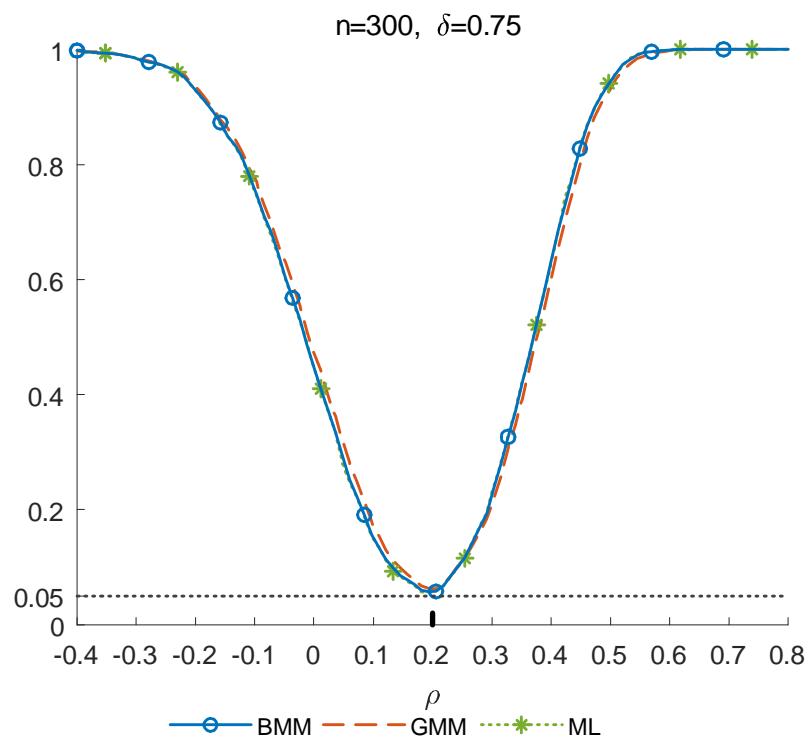
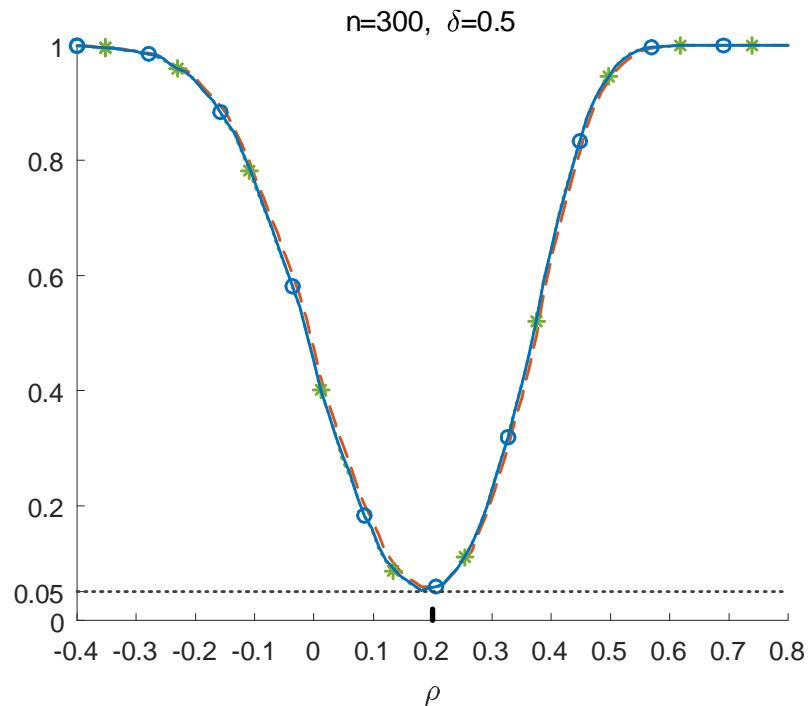


Figure S.2: (Continued)

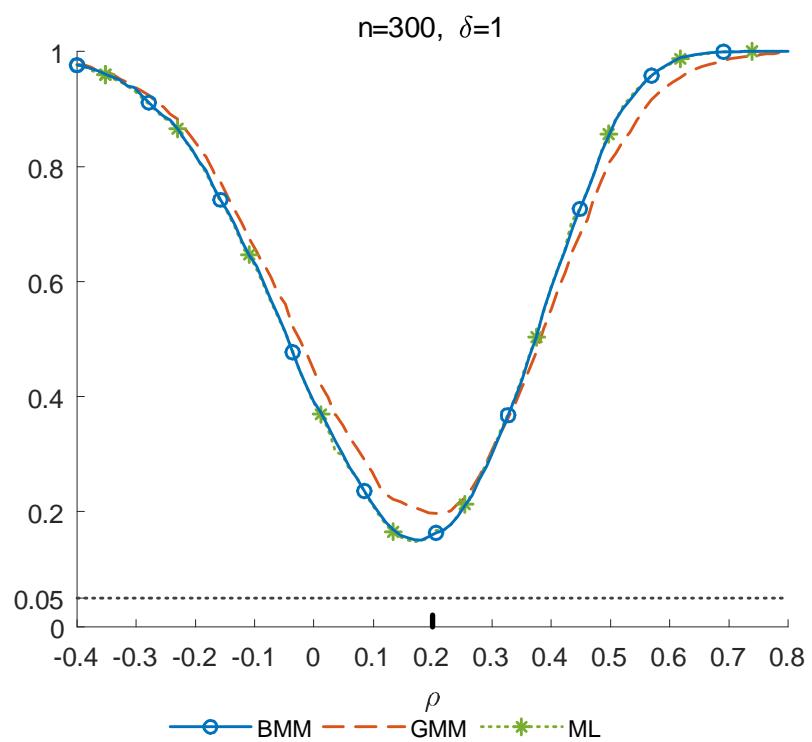
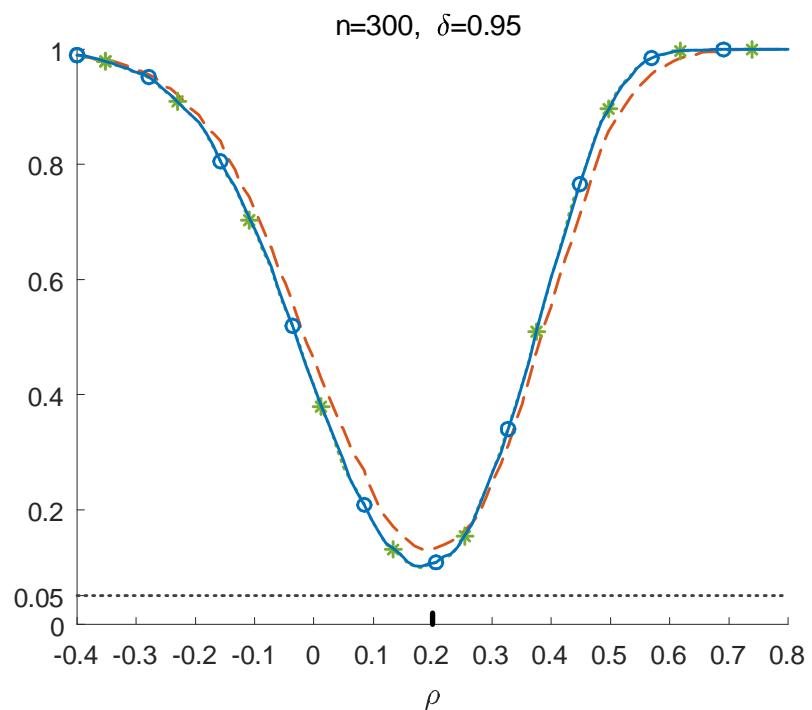


Figure S.2: (Continued)

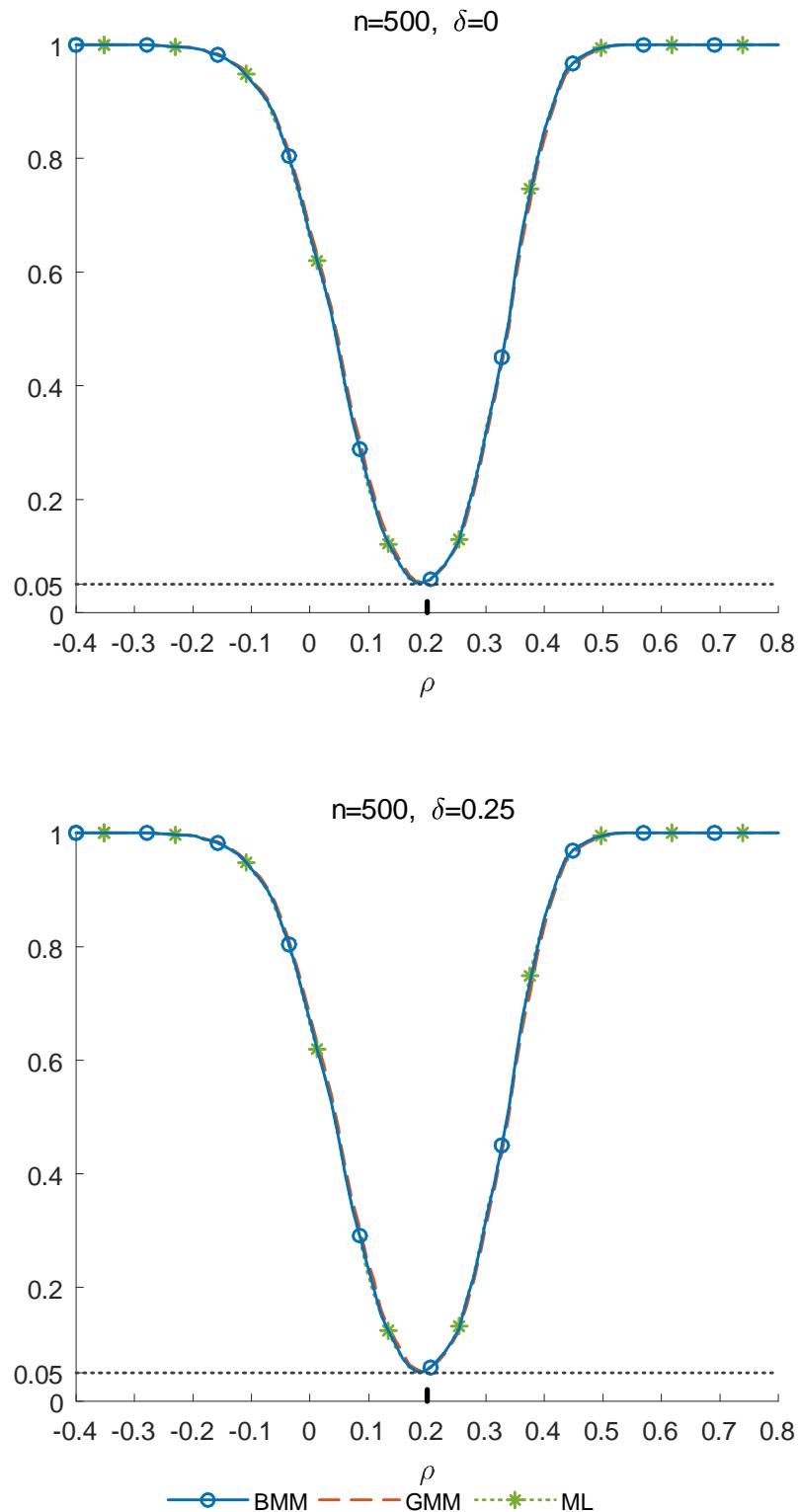


Figure S.3: Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 500$, and homoskedastic Gaussian errors for different values of δ

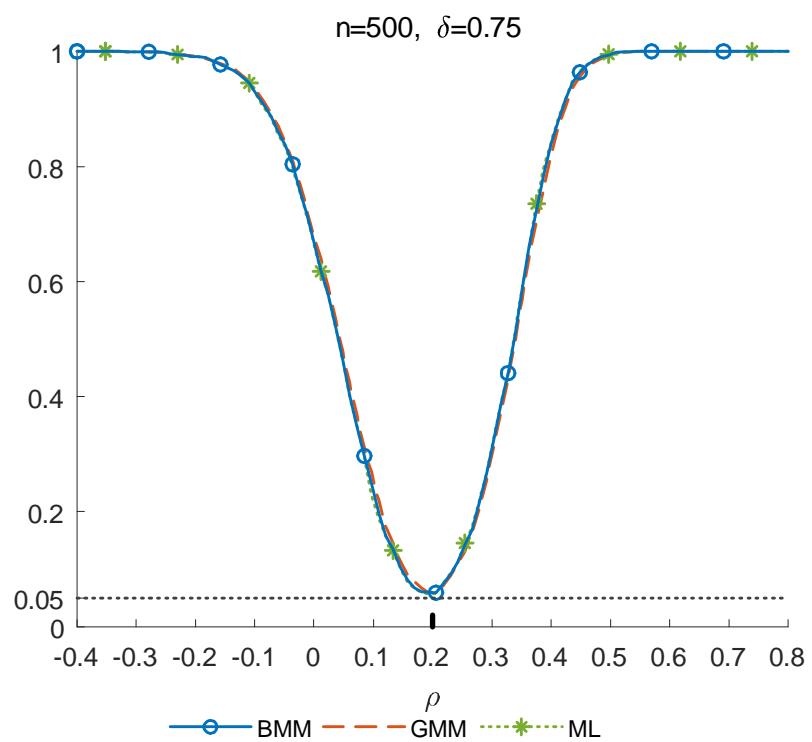
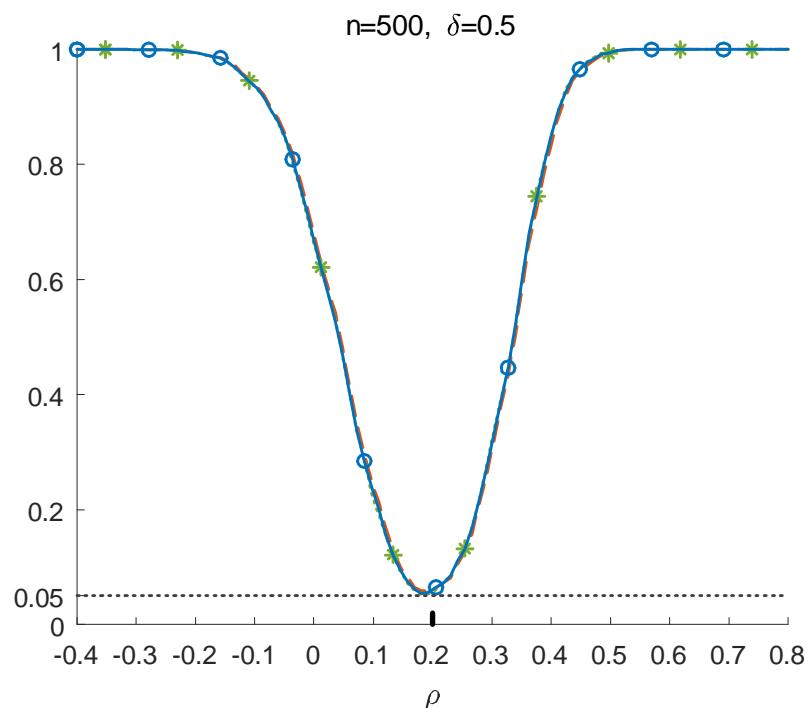


Figure S.3: (Continued)

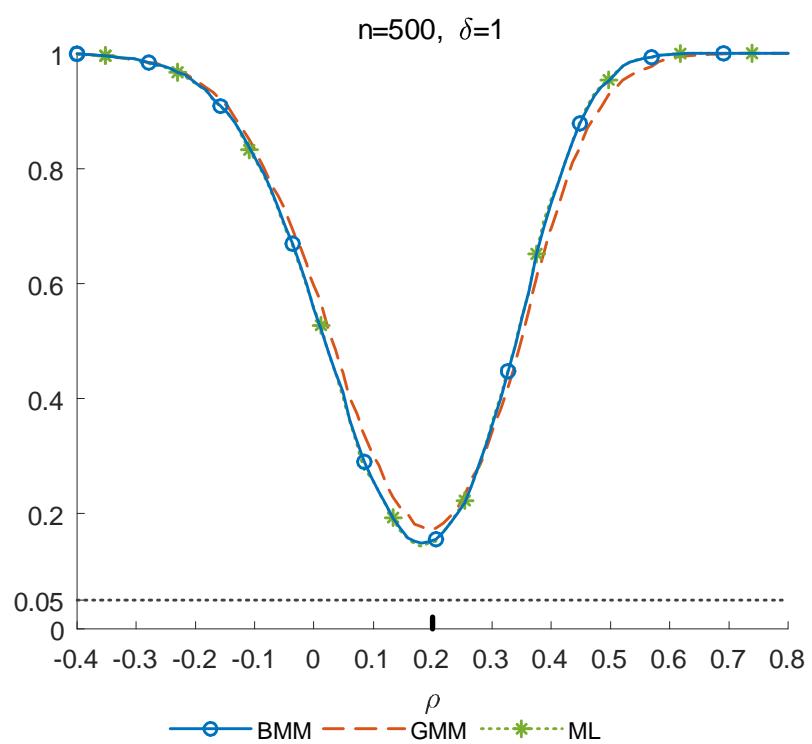
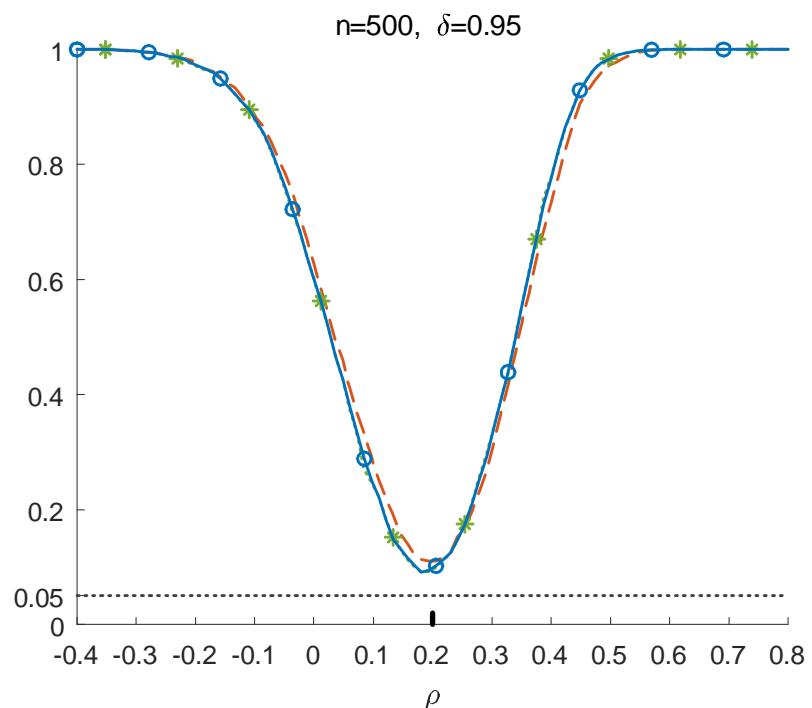


Figure S.3: (Continued)

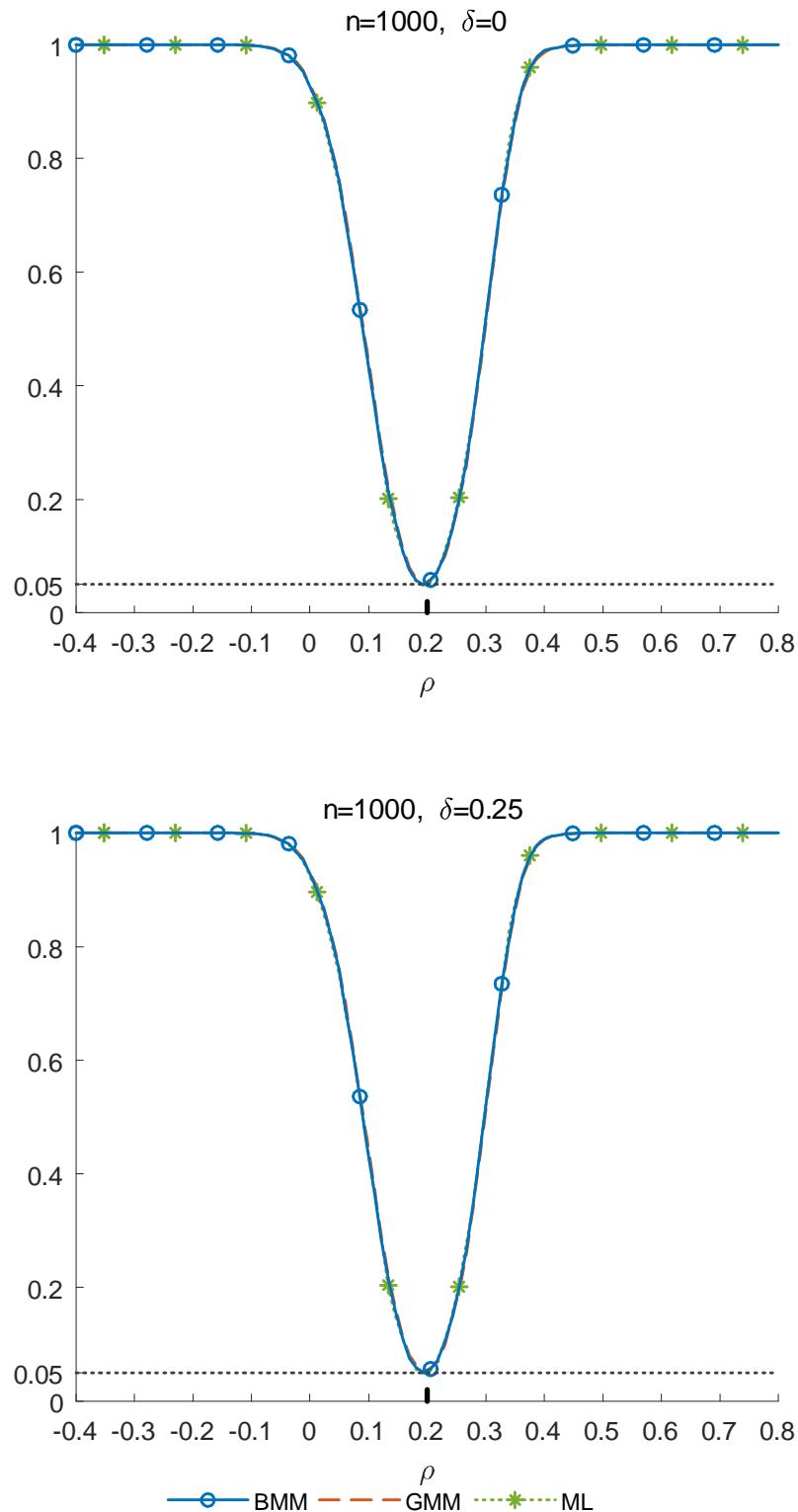


Figure S.4: Empirical power functions for ρ in the case of $\rho_0 = 0.2$, $n = 1,000$, and homoskedastic Gaussian errors for different values of δ

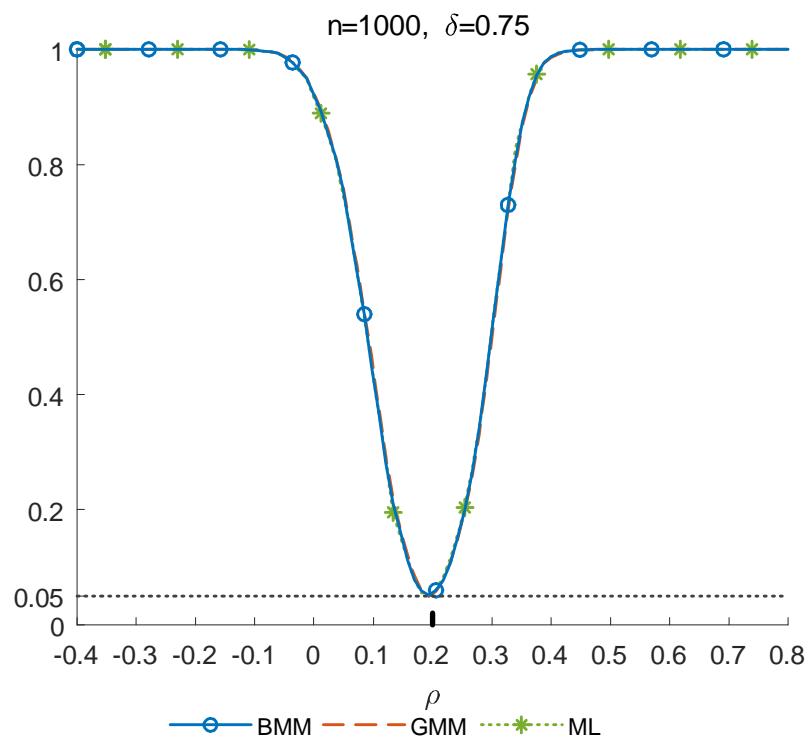
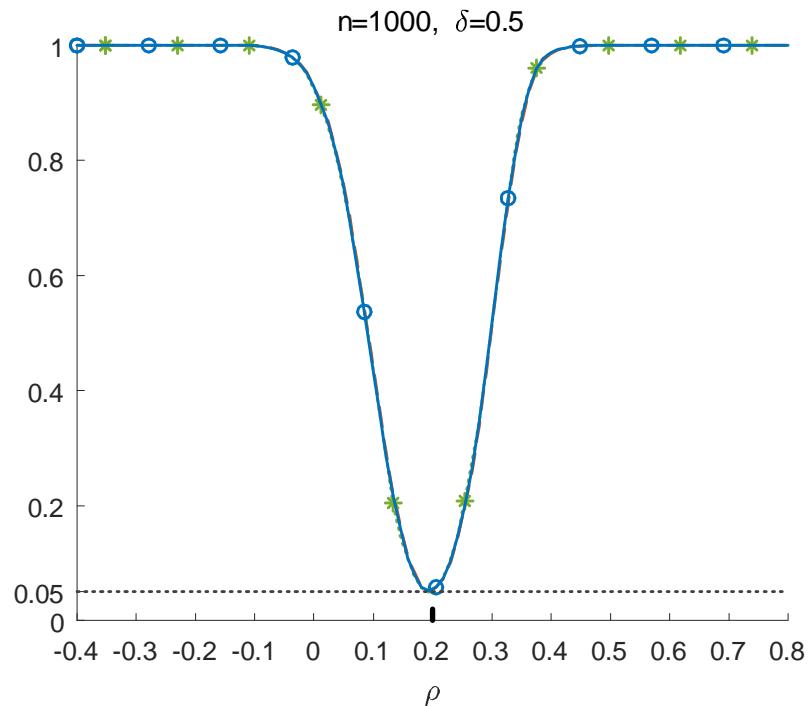


Figure S.4: (Continued)

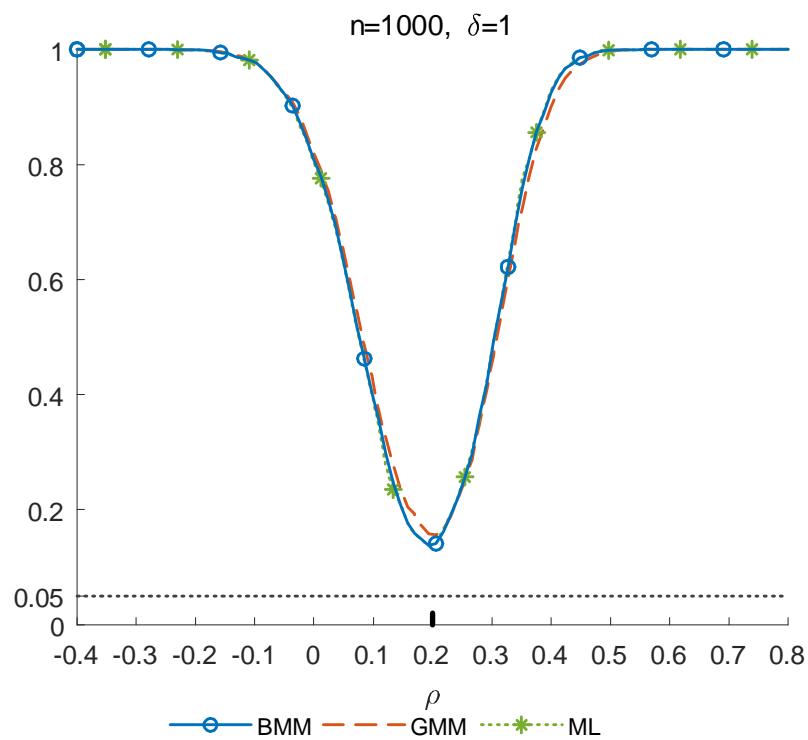
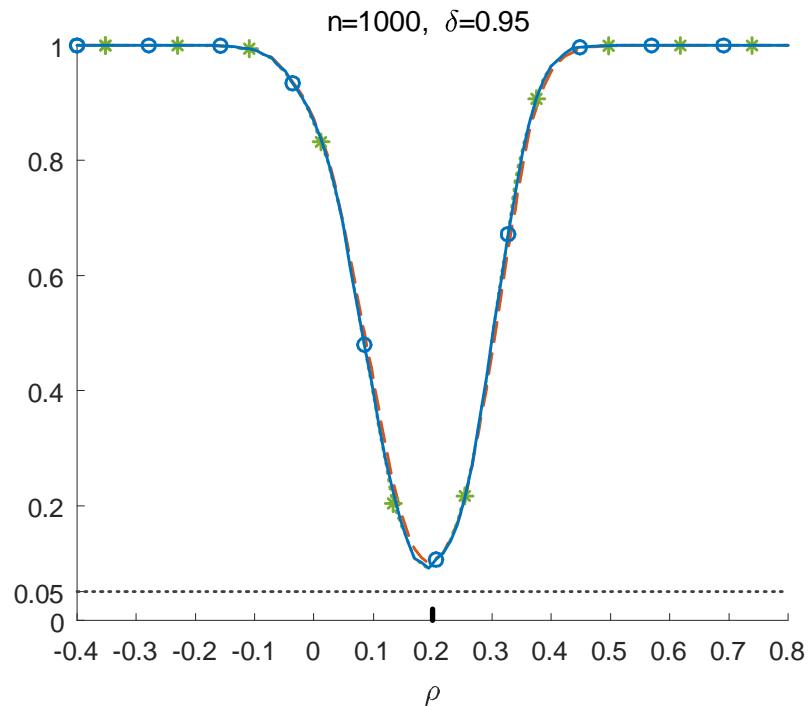


Figure S.4: (Continued)

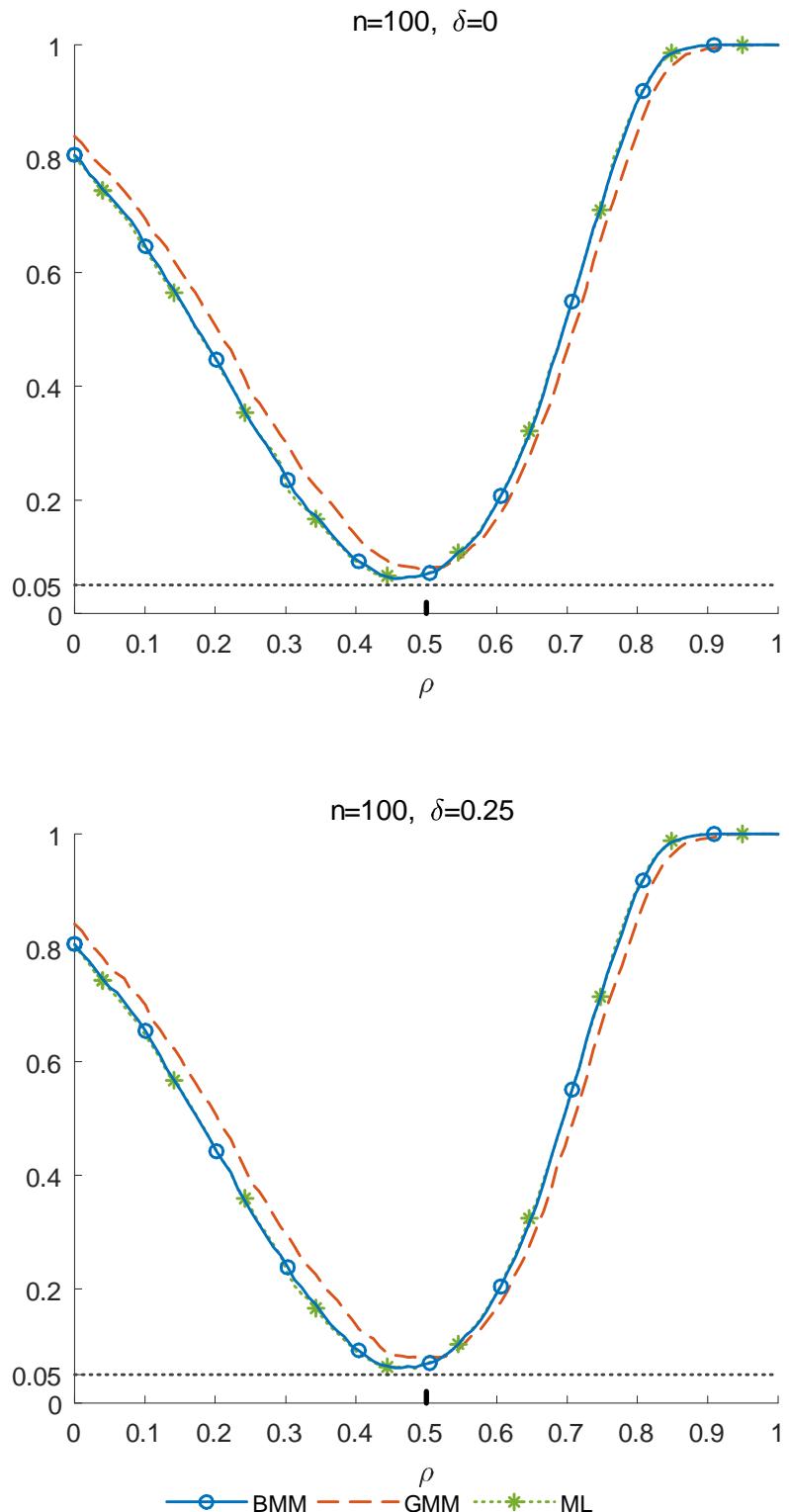


Figure S.5: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and homoskedastic Gaussian errors for different values of δ

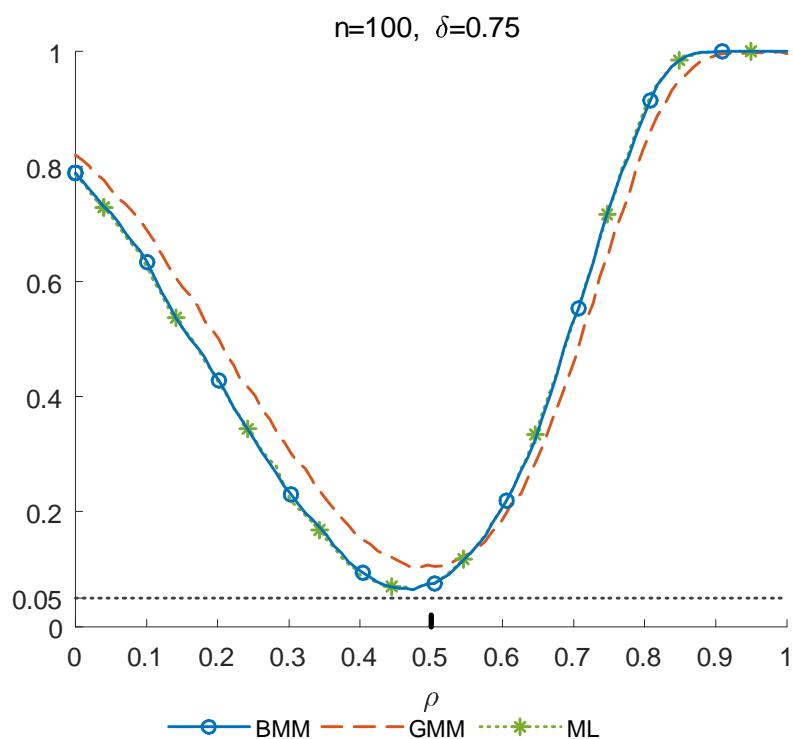
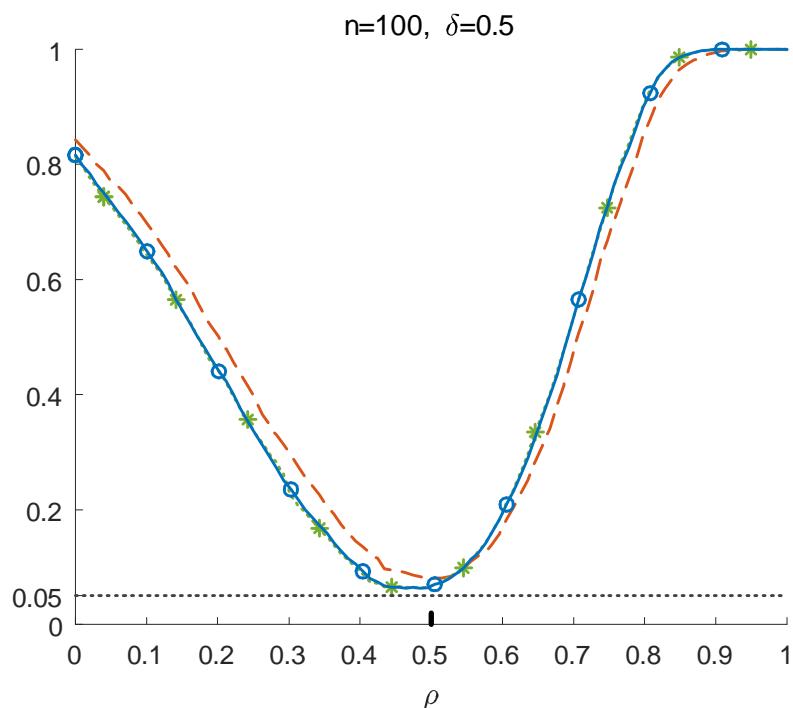


Figure S.5: (Continued)

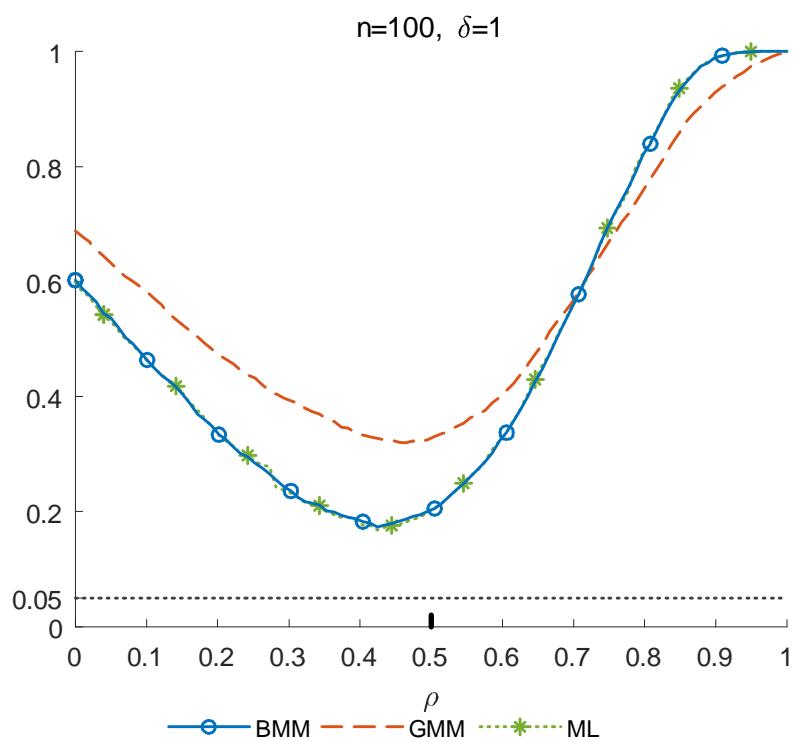
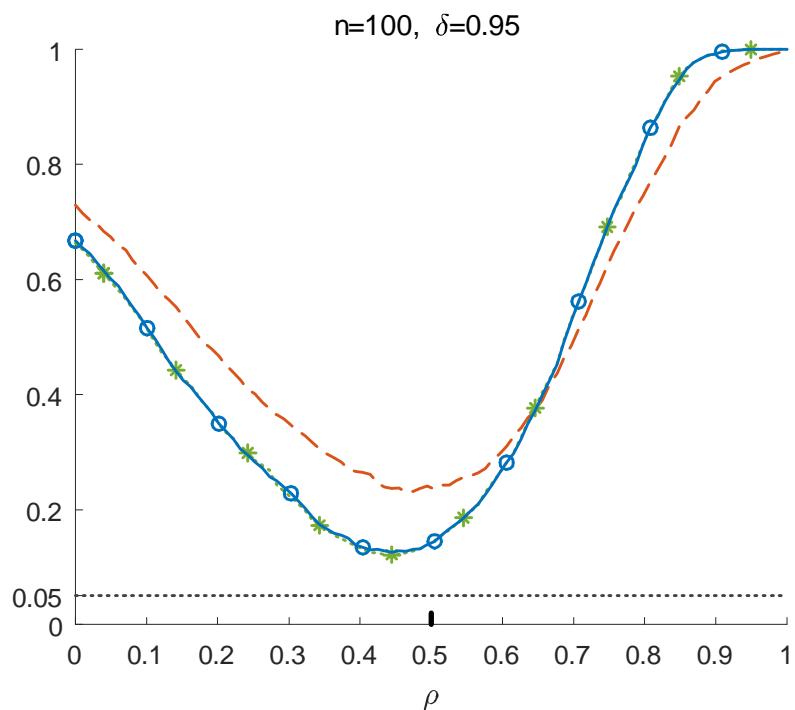


Figure S.5: (Continued)

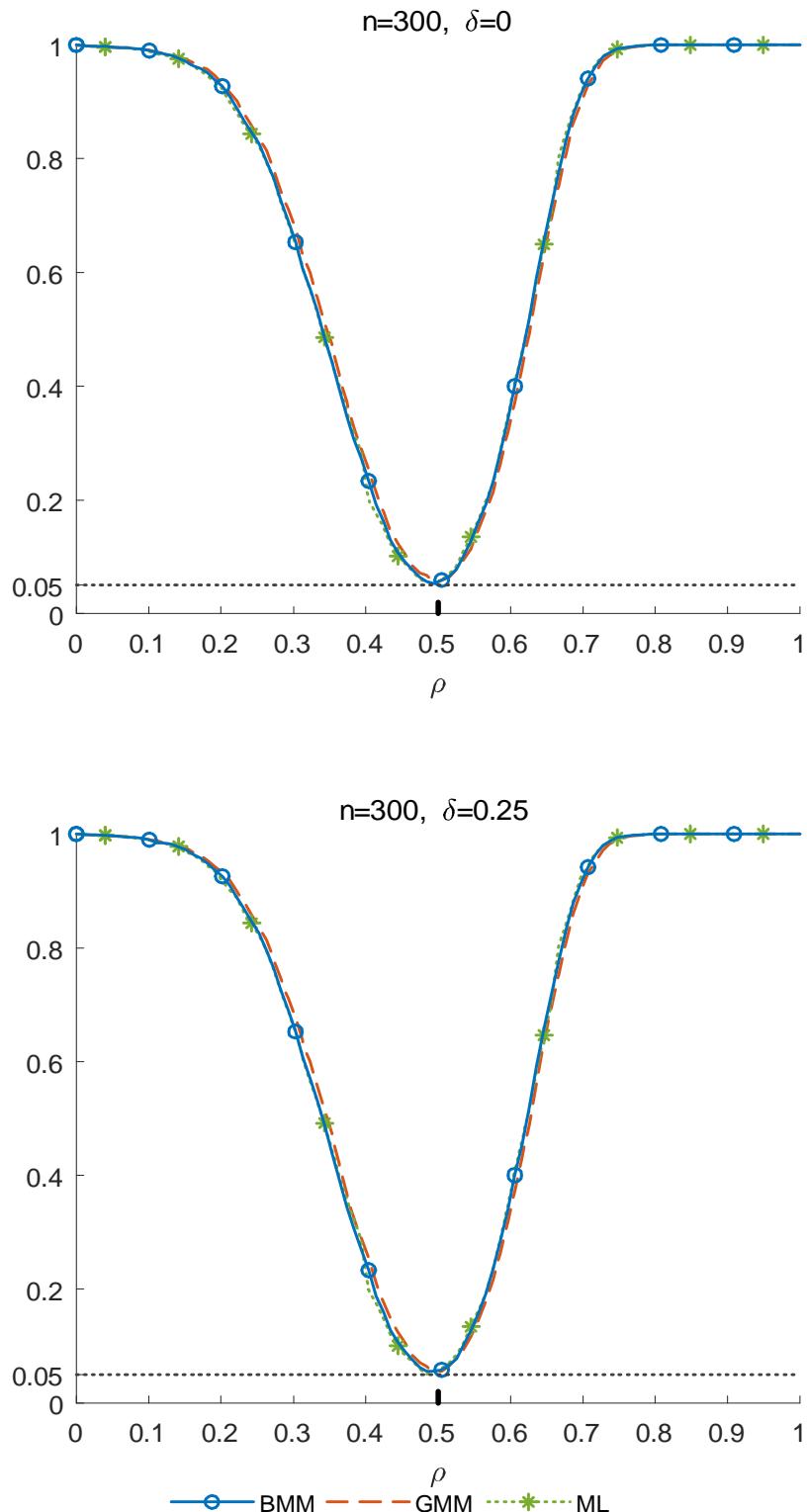


Figure S.6: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and homoskedastic Gaussian errors for different values of δ

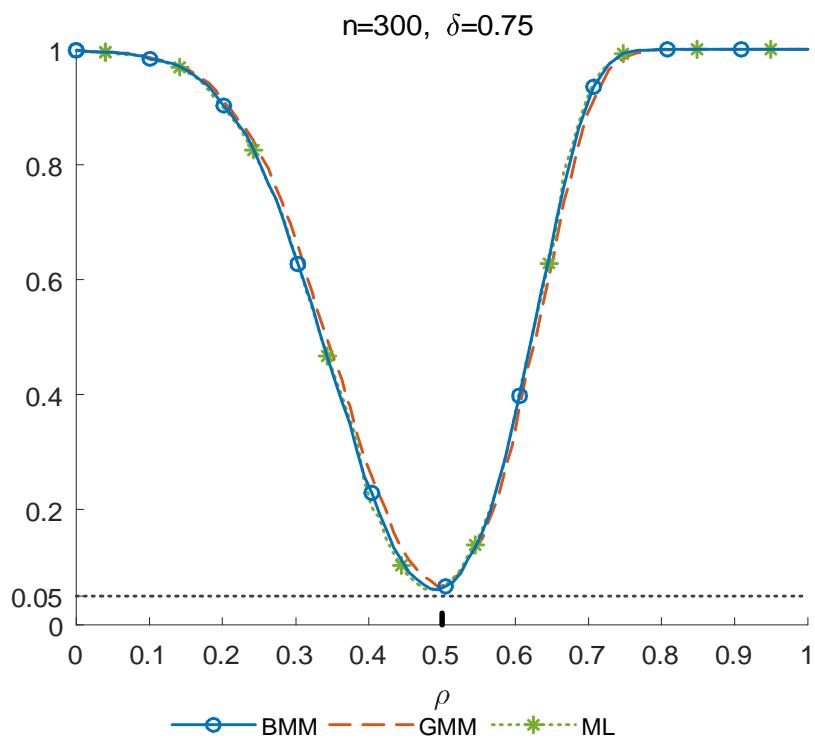
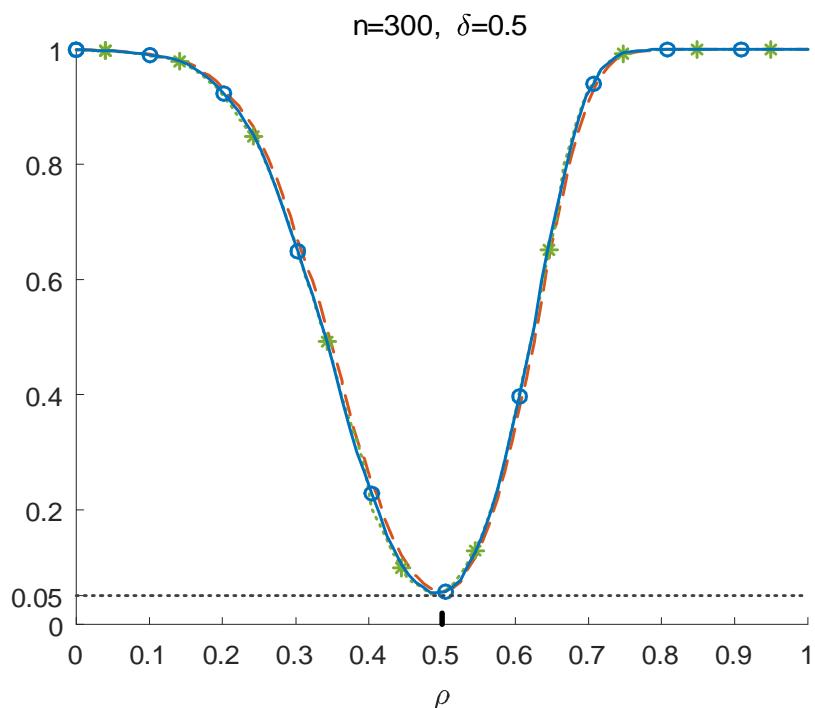


Figure S.6: (Continued)

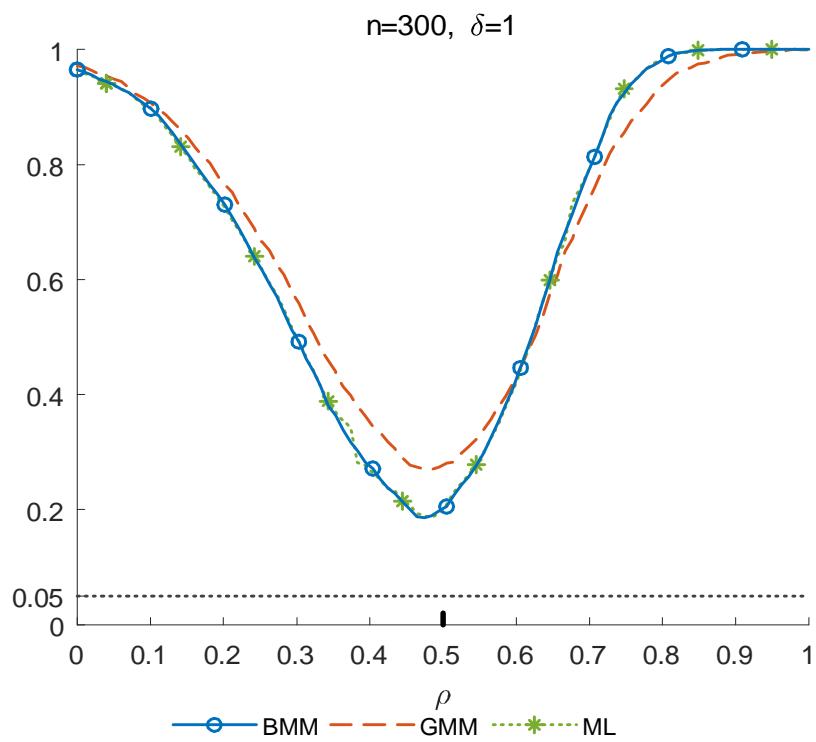
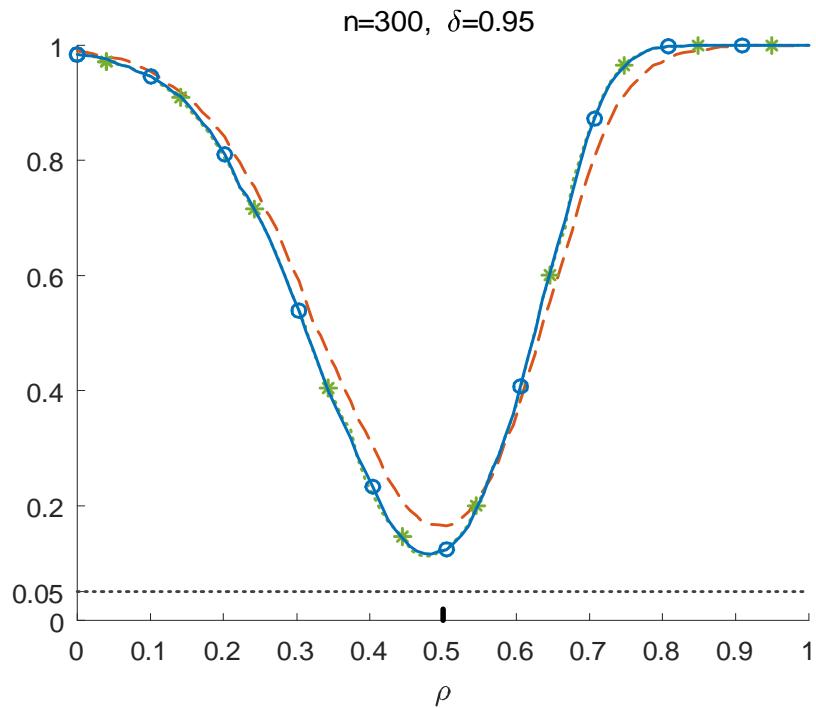


Figure S.6: (Continued)

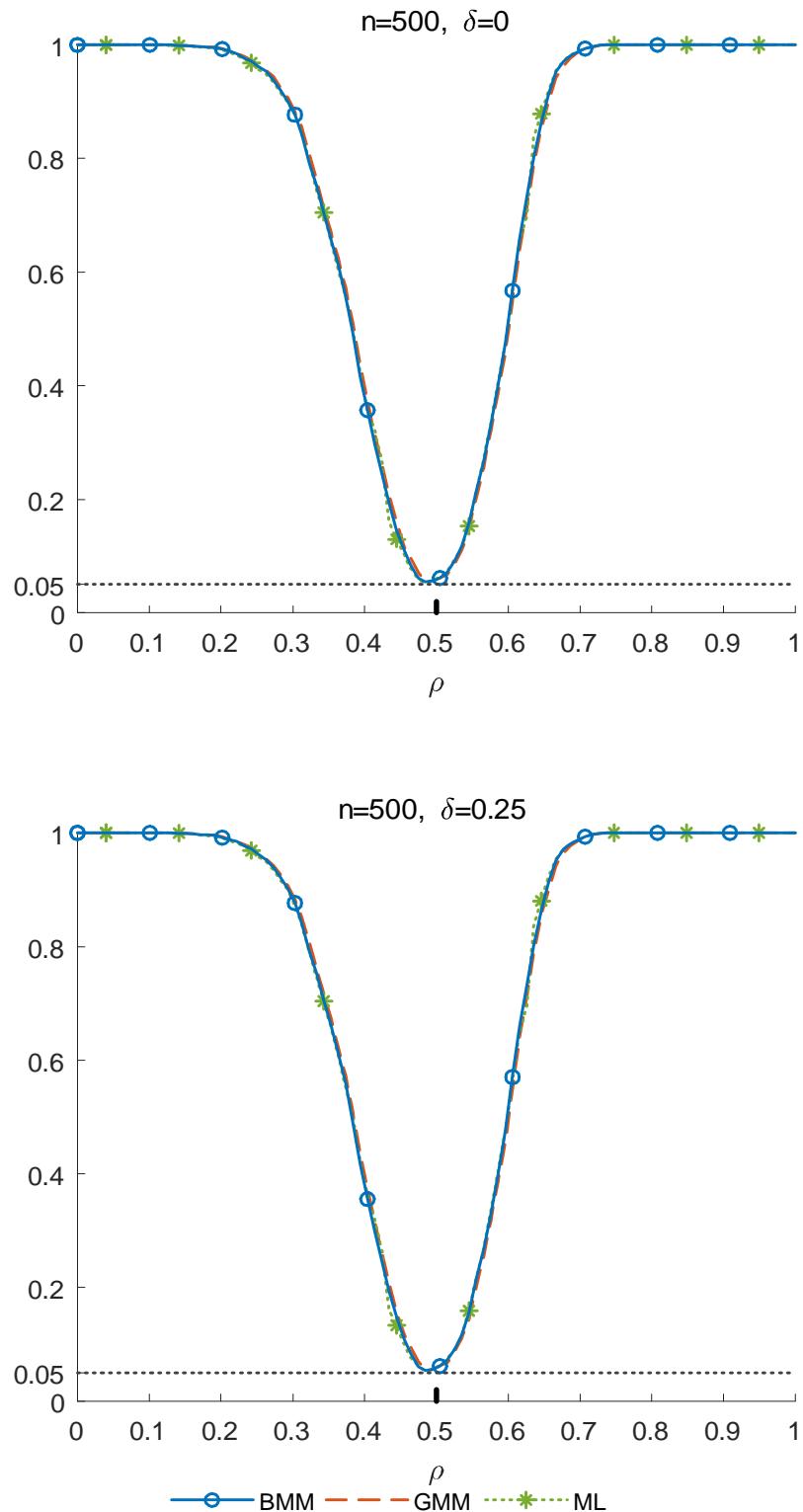


Figure S.7: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 500$, and homoskedastic Gaussian errors for different values of δ

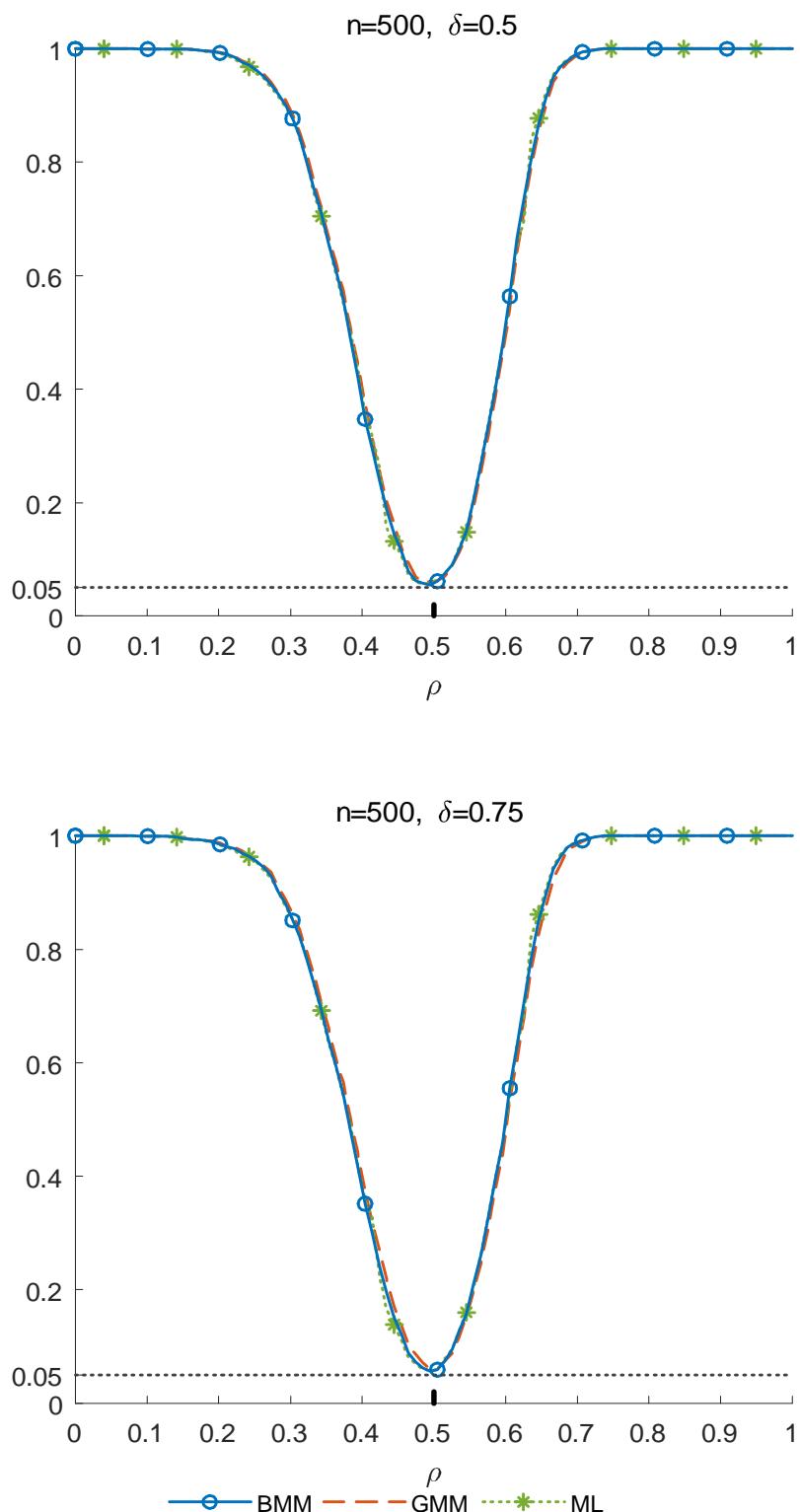


Figure S.7: (Continued)

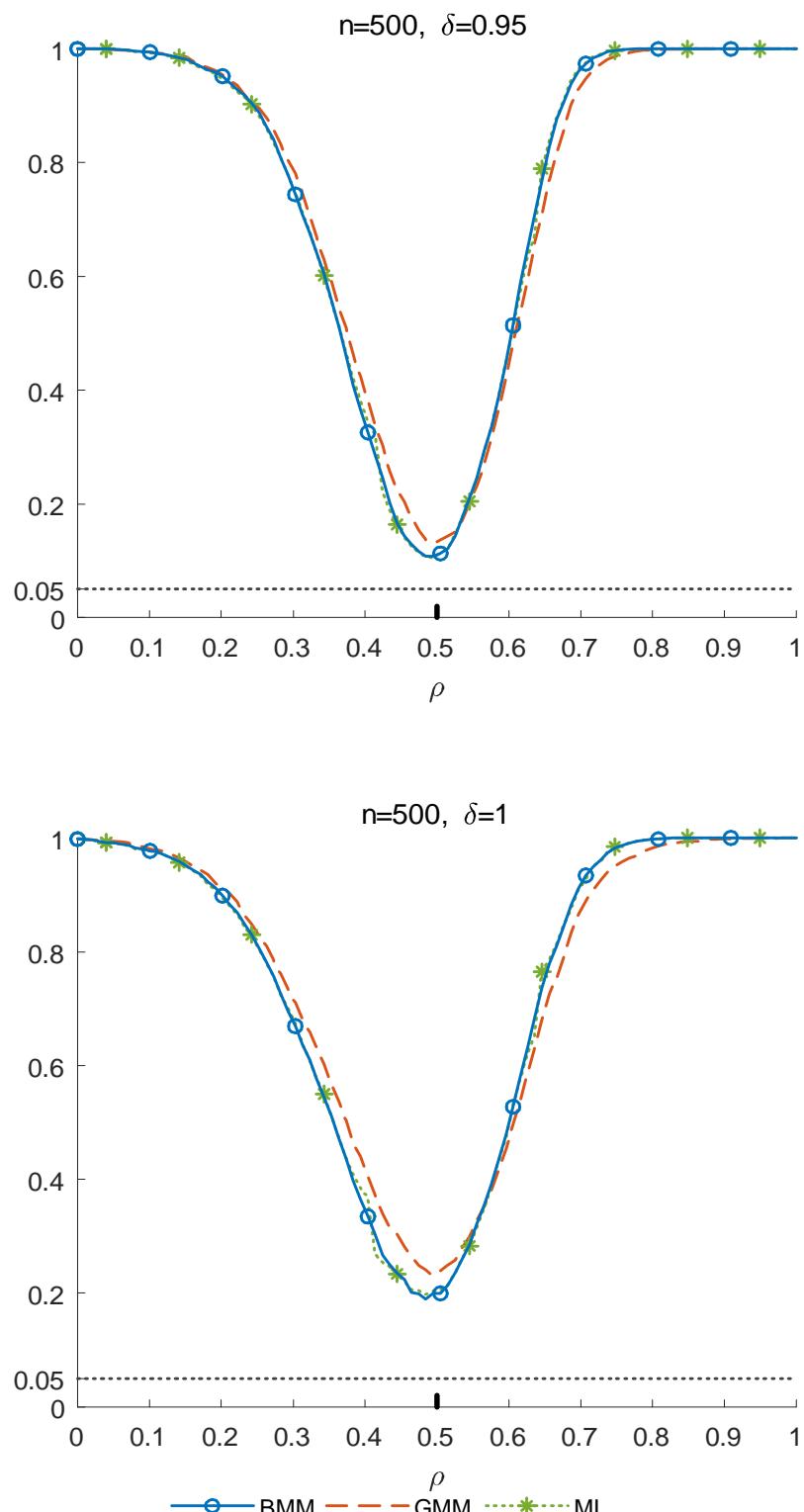


Figure S.7: (Continued)

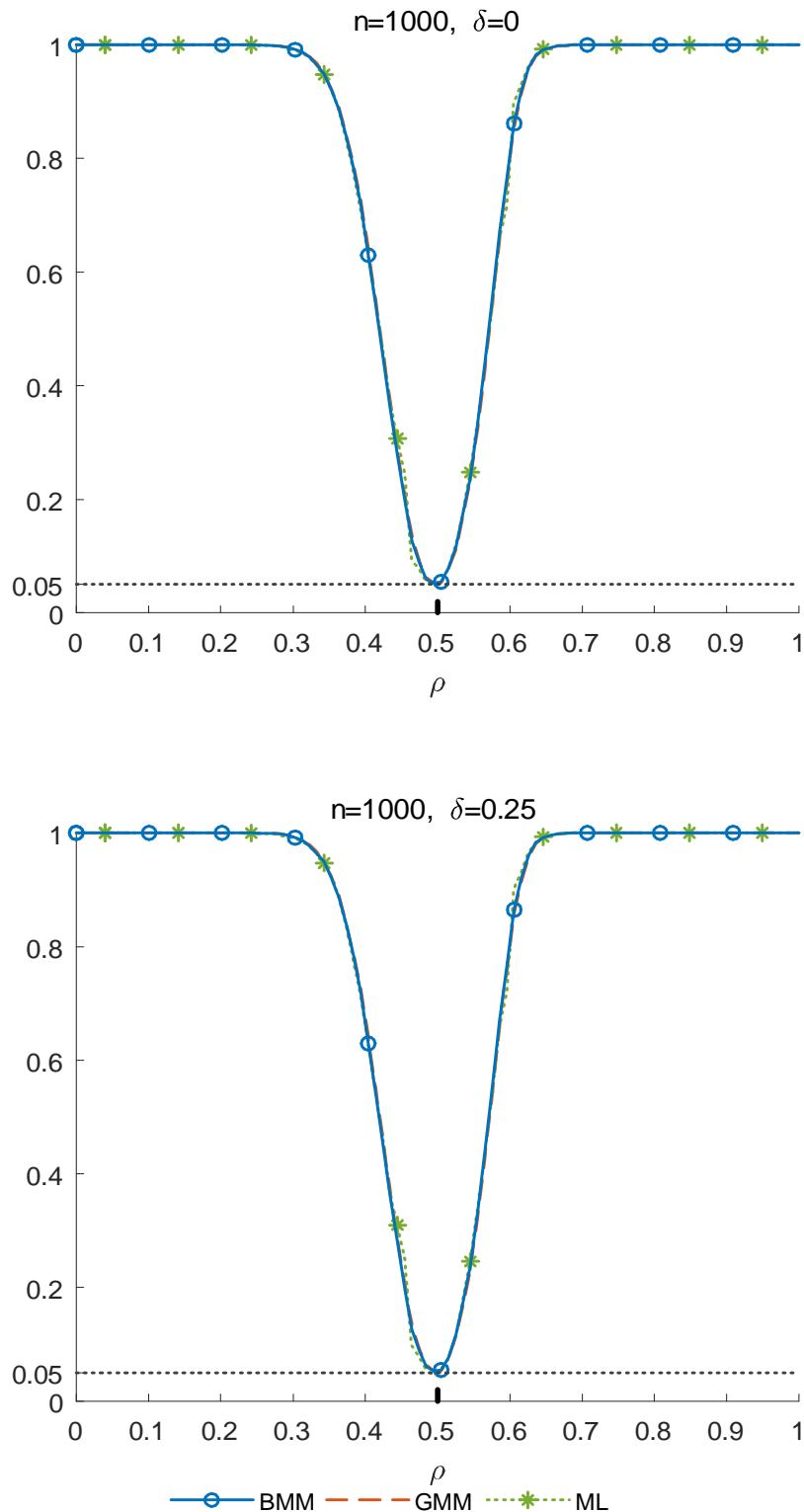


Figure S.8: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 1,000$, and homoskedastic Gaussian errors for different values of δ

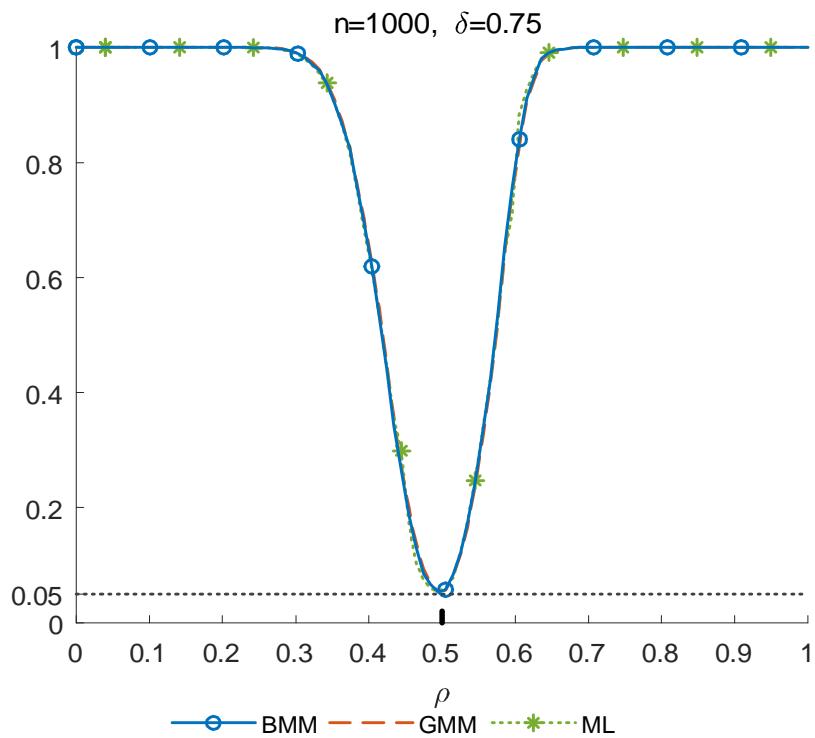
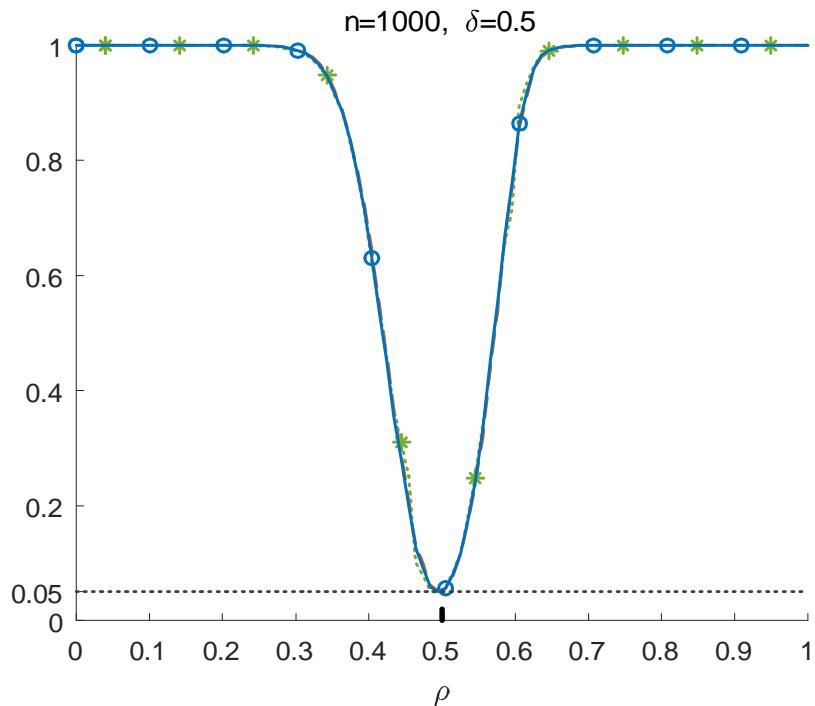


Figure S.8: (Continued)

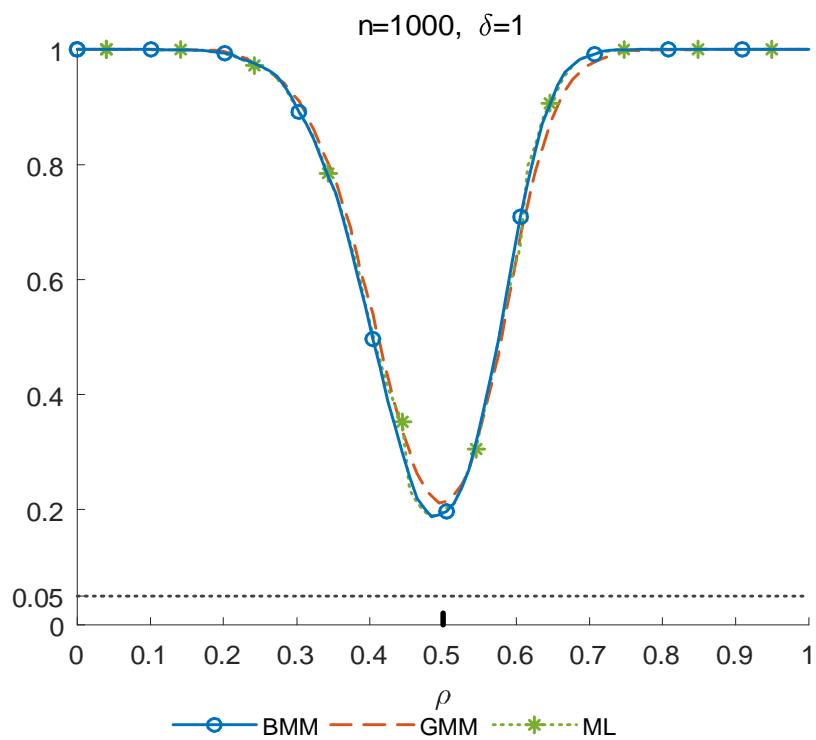
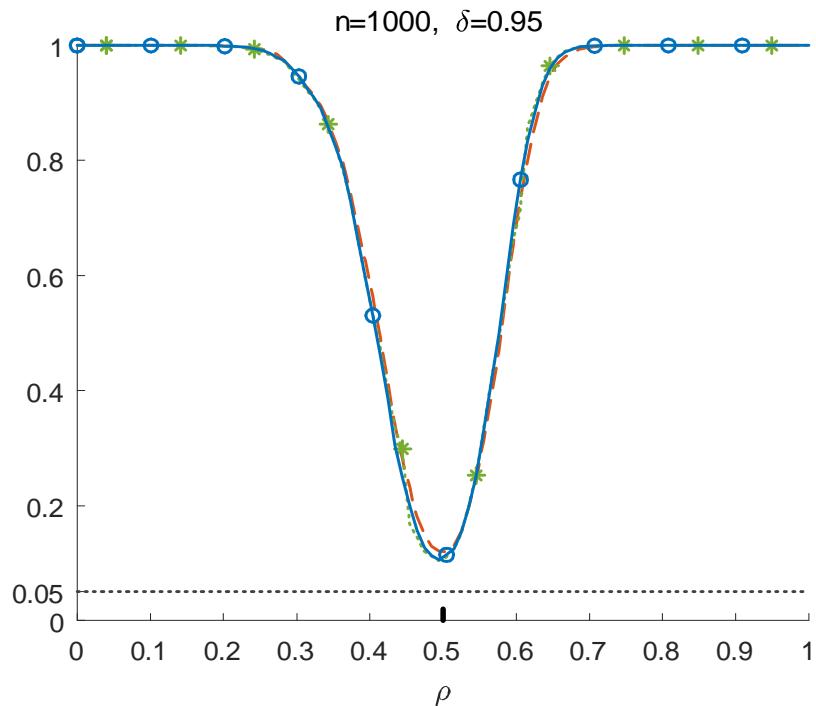


Figure S.8: (Continued)

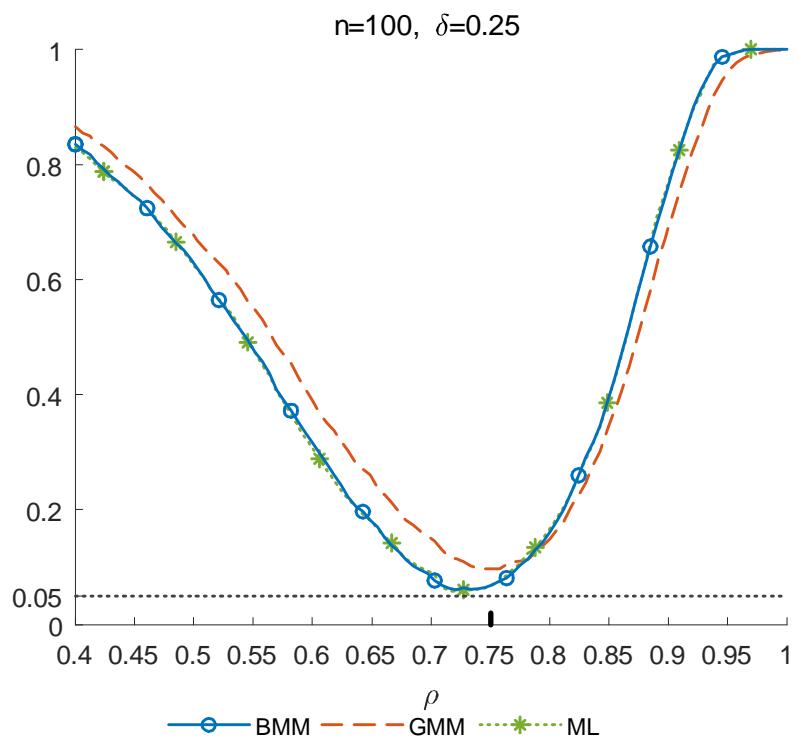
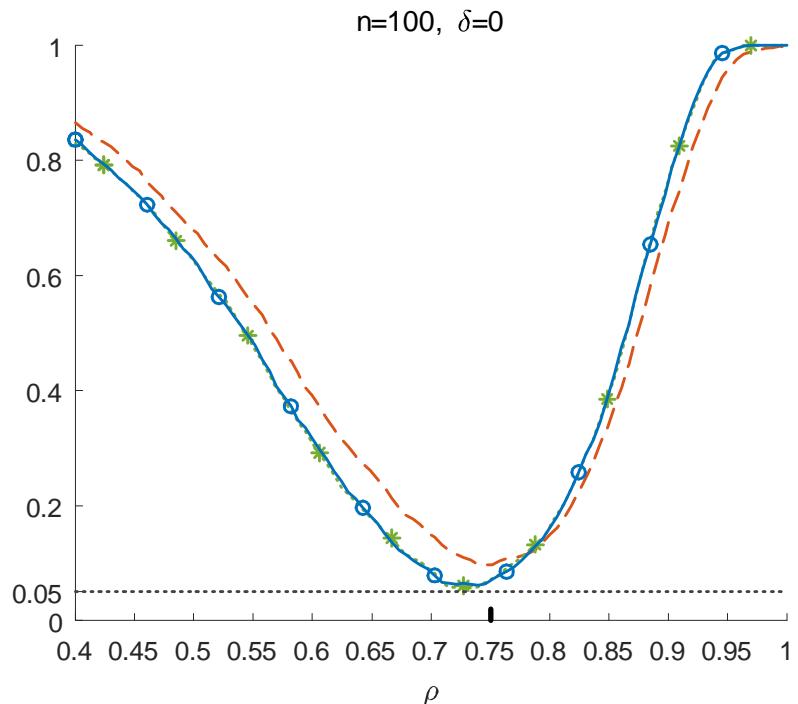


Figure S.9: Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 100$, and homoskedastic Gaussian errors for different values of δ

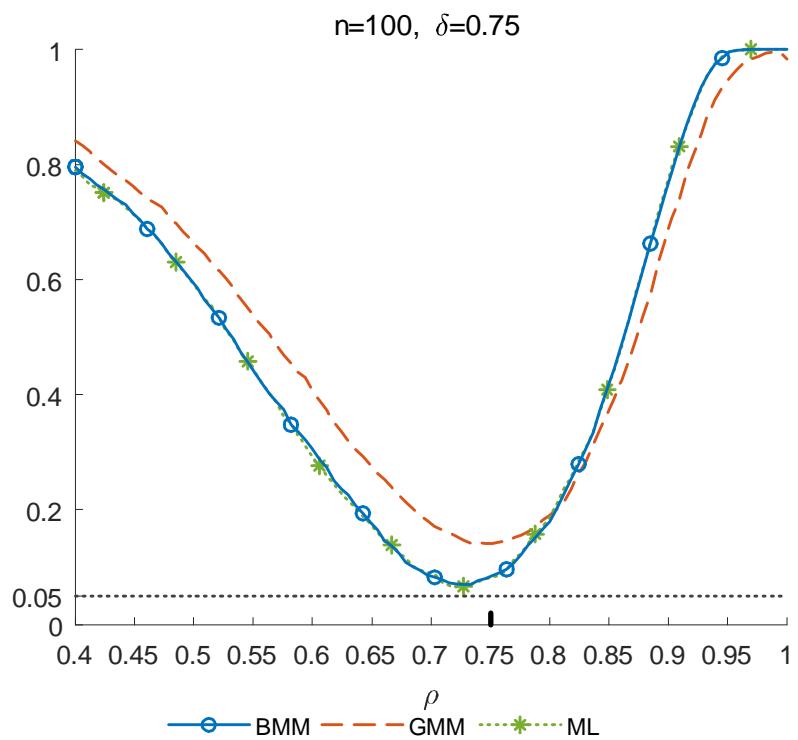
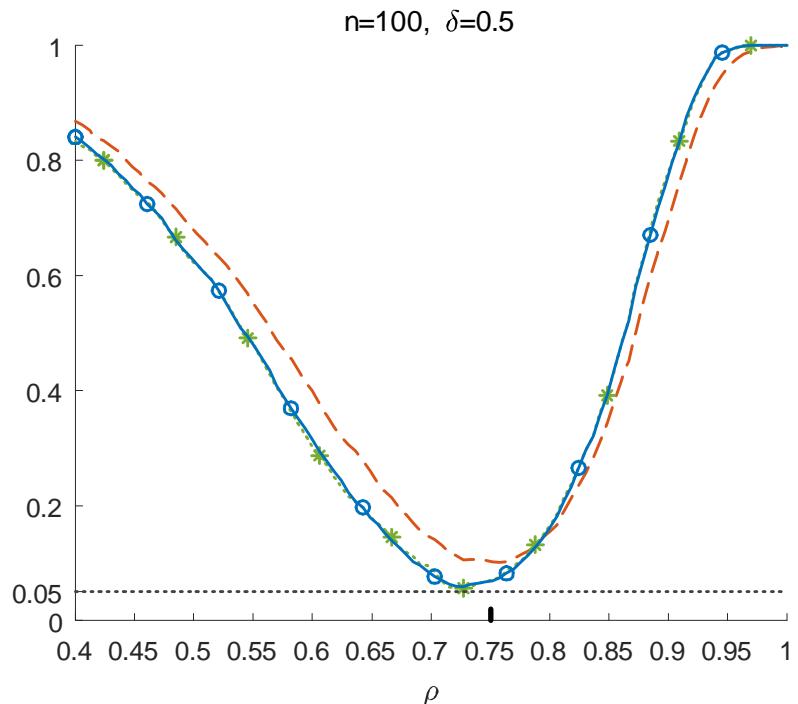


Figure S.9: (Continued)

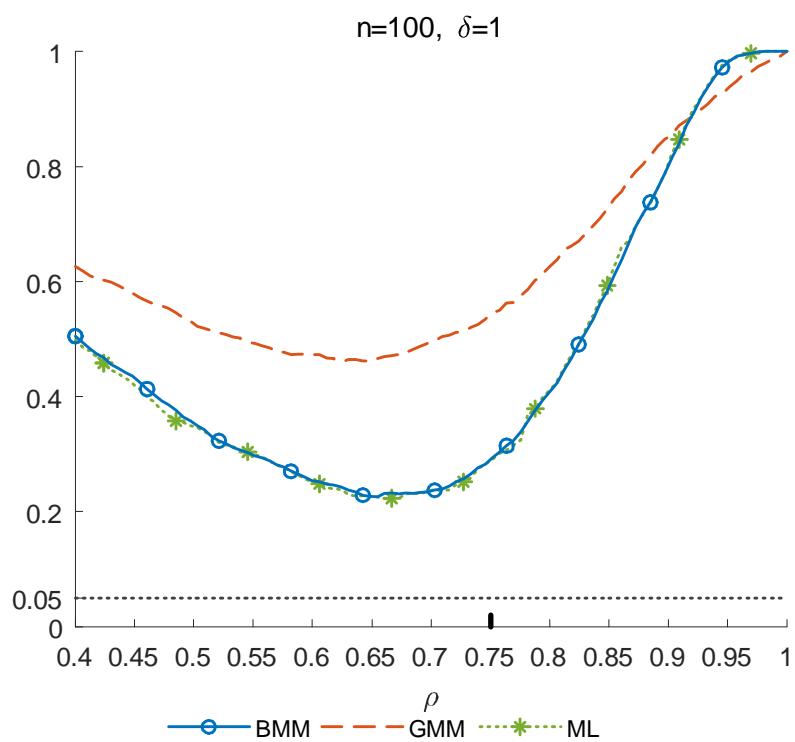
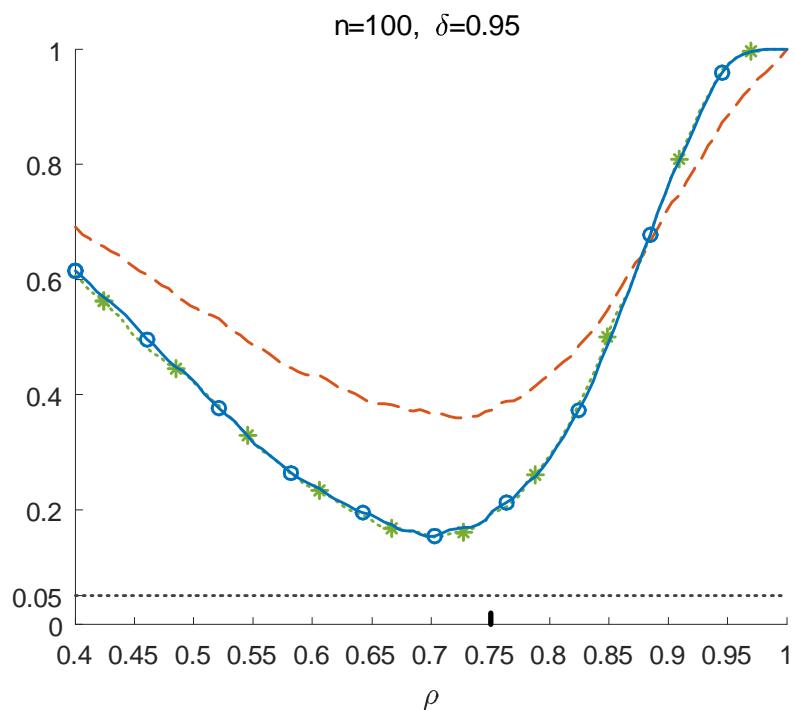


Figure S.9: (Continued)

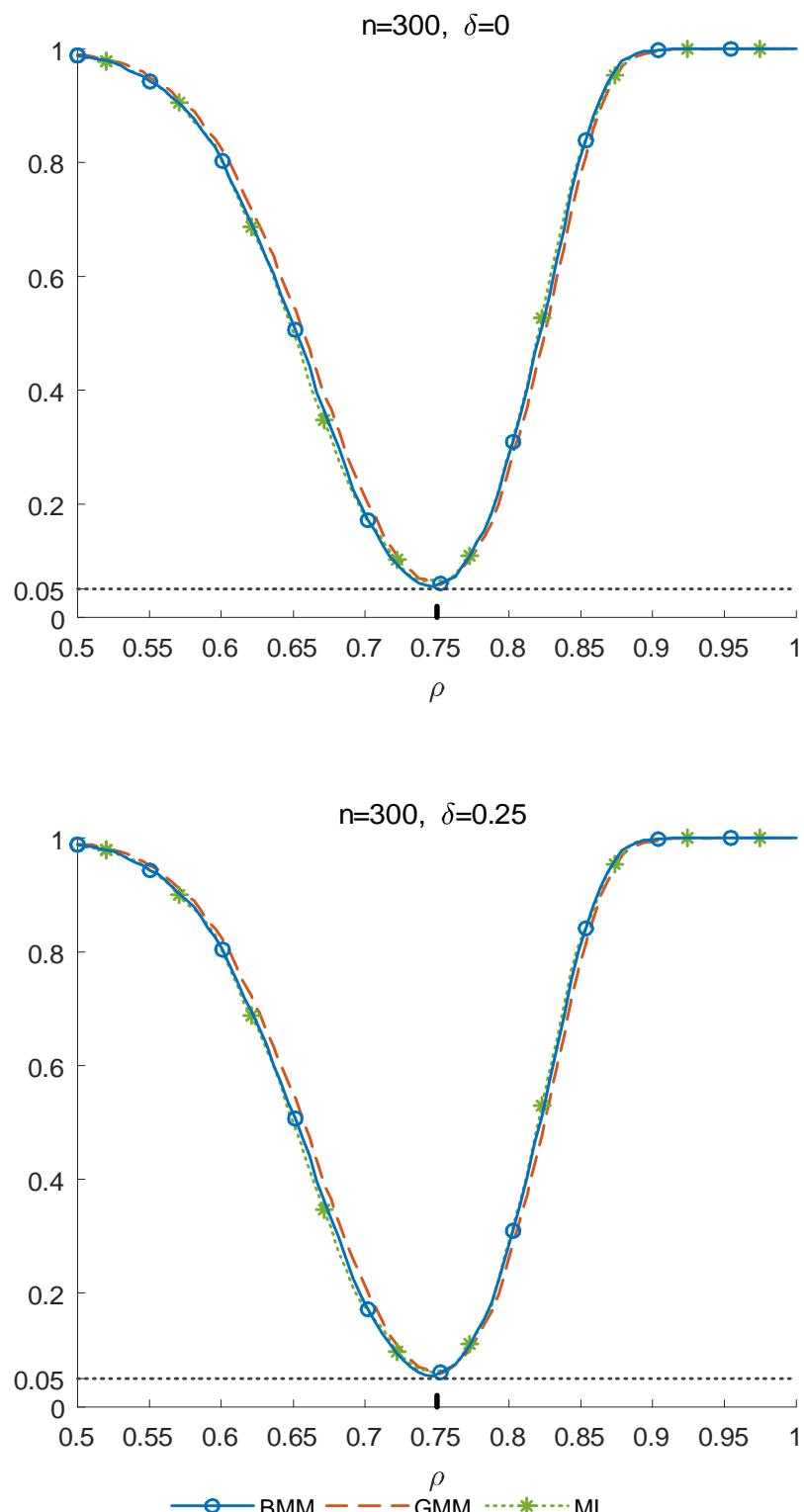


Figure S.10: Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 300$, and homoskedastic Gaussian errors for different values of δ

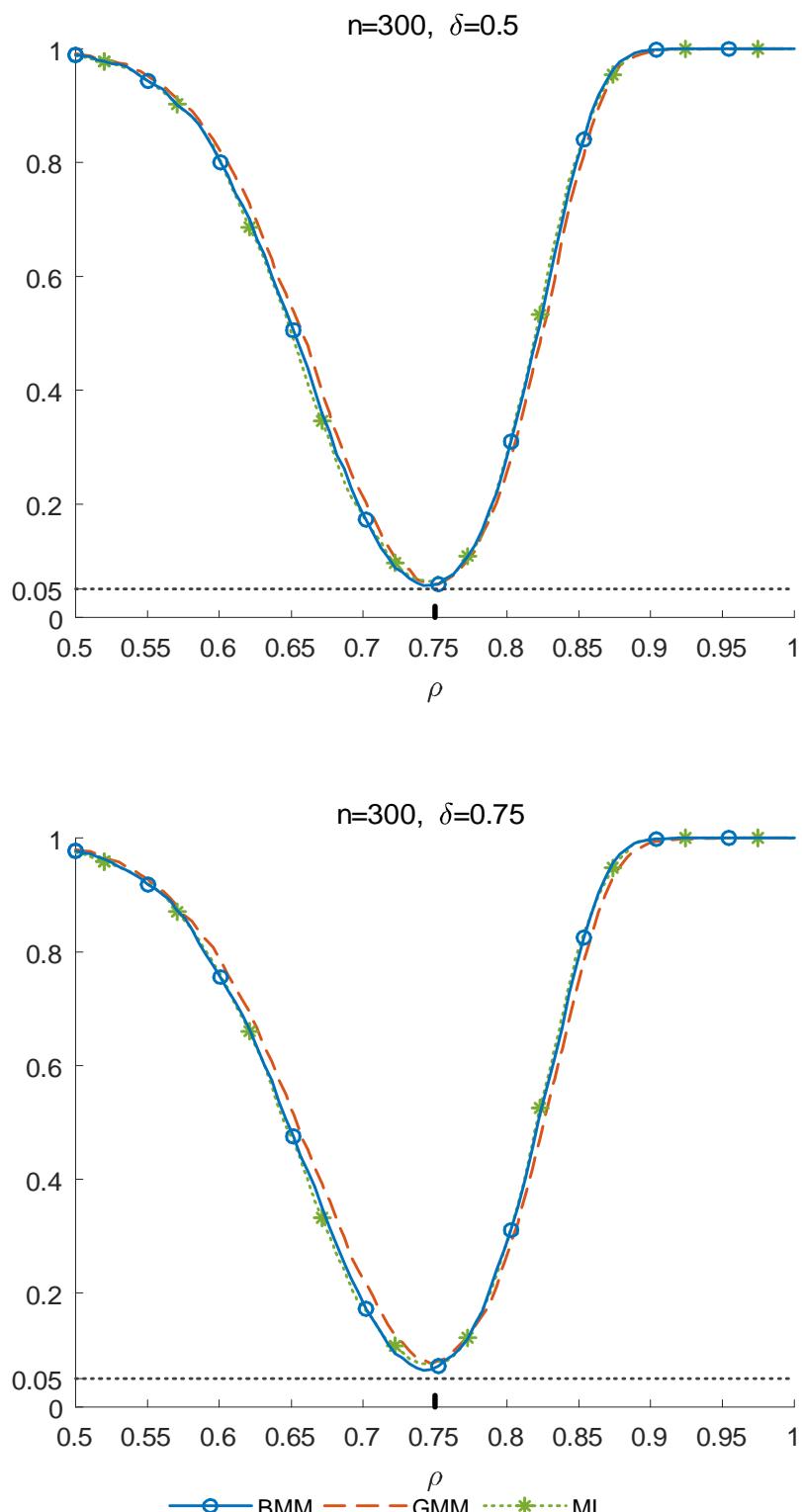


Figure S.10: (Continued)

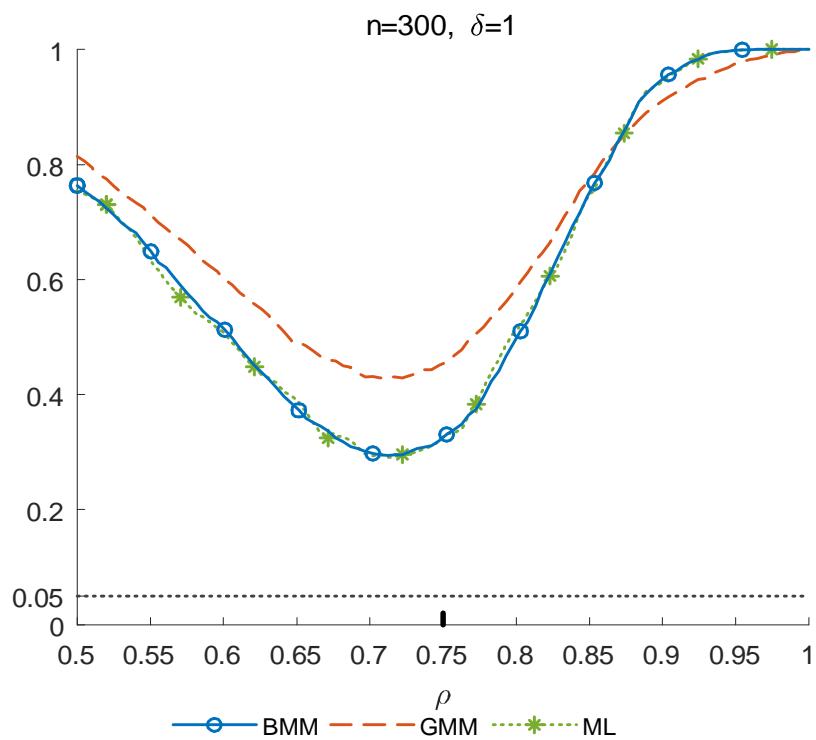
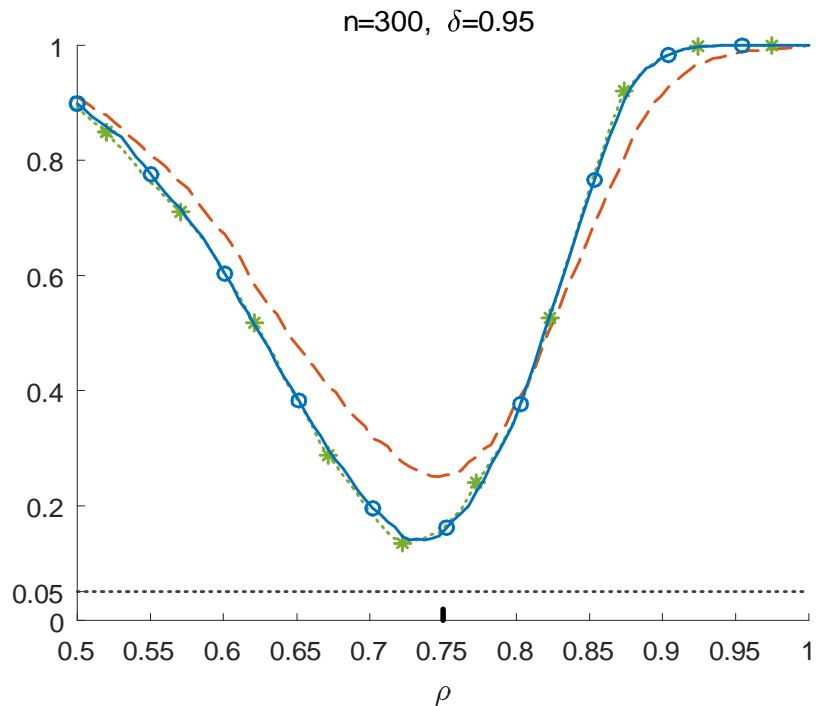


Figure S.10: (Continued)

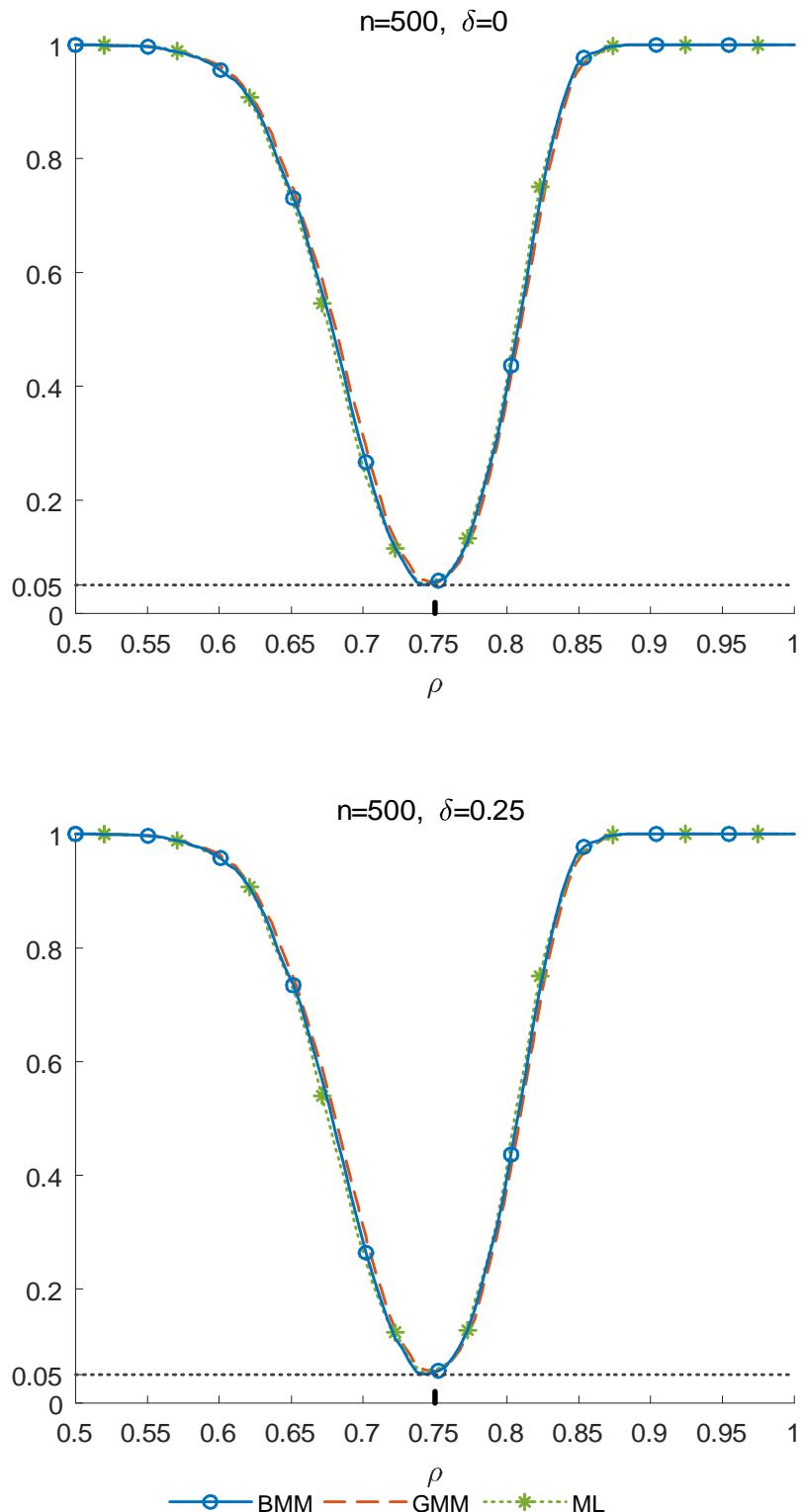


Figure S.11: Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 500$, and homoskedastic Gaussian errors for different values of δ

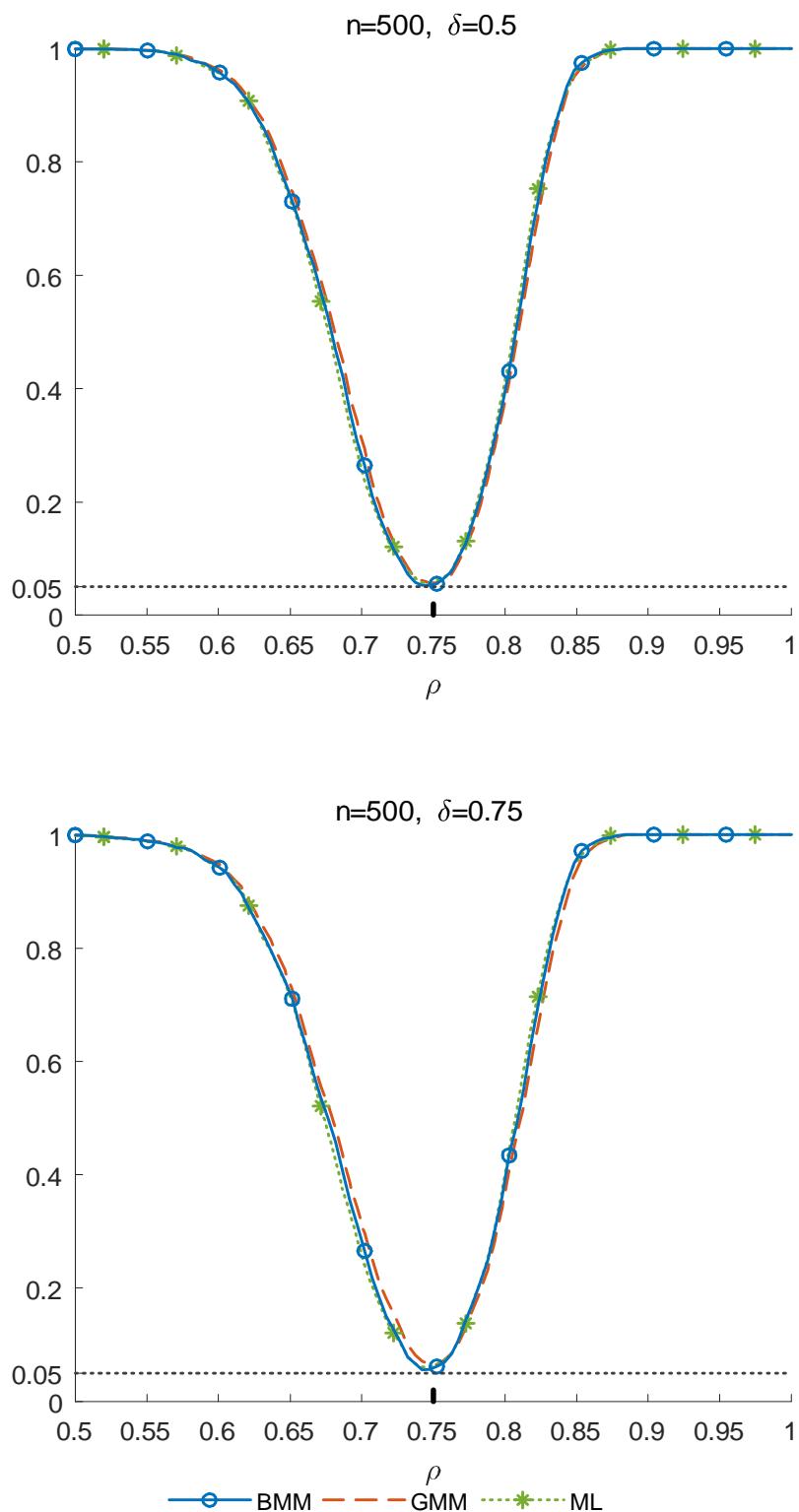


Figure S.11: (Continued)

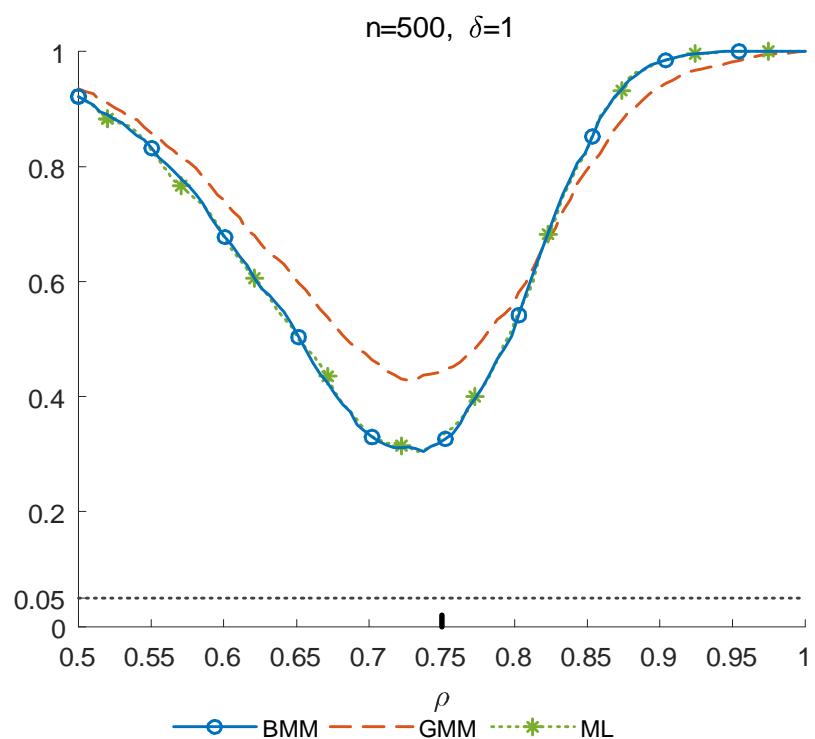
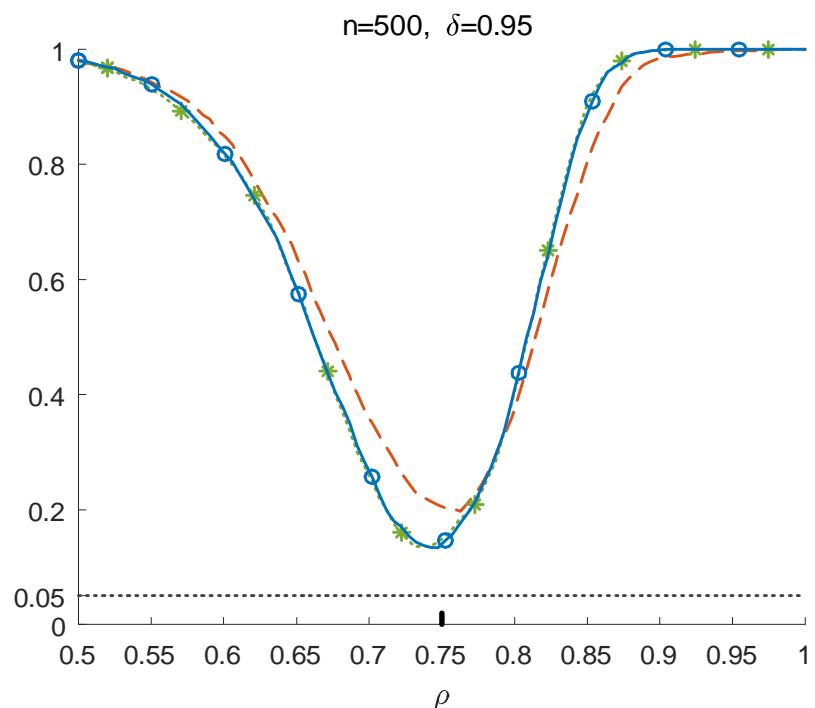


Figure S.11: (Continued)

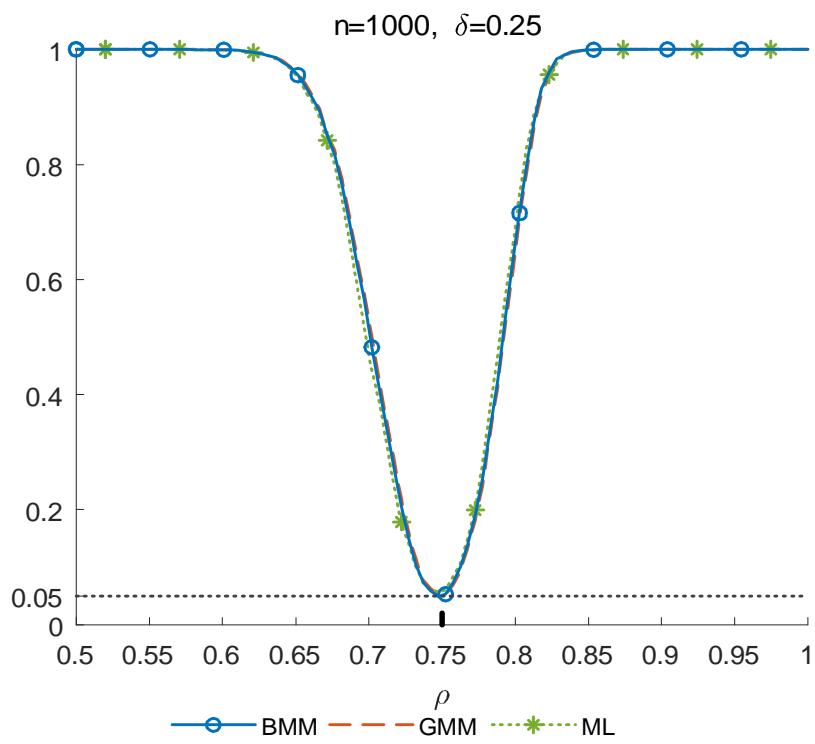
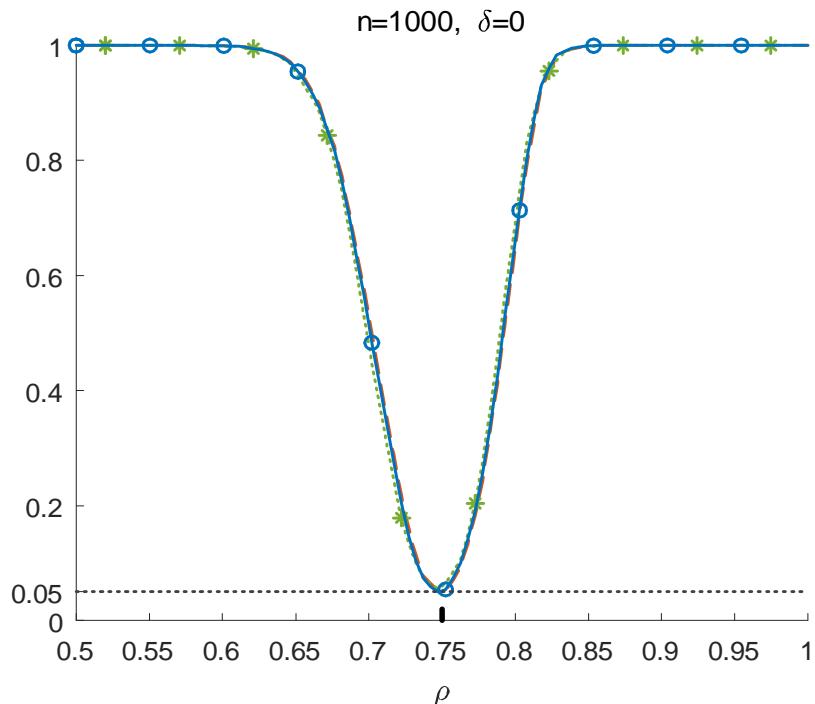


Figure S.12: Empirical power functions for ρ in the case of $\rho_0 = 0.75$, $n = 1,000$, and homoskedastic Gaussian errors for different values of δ

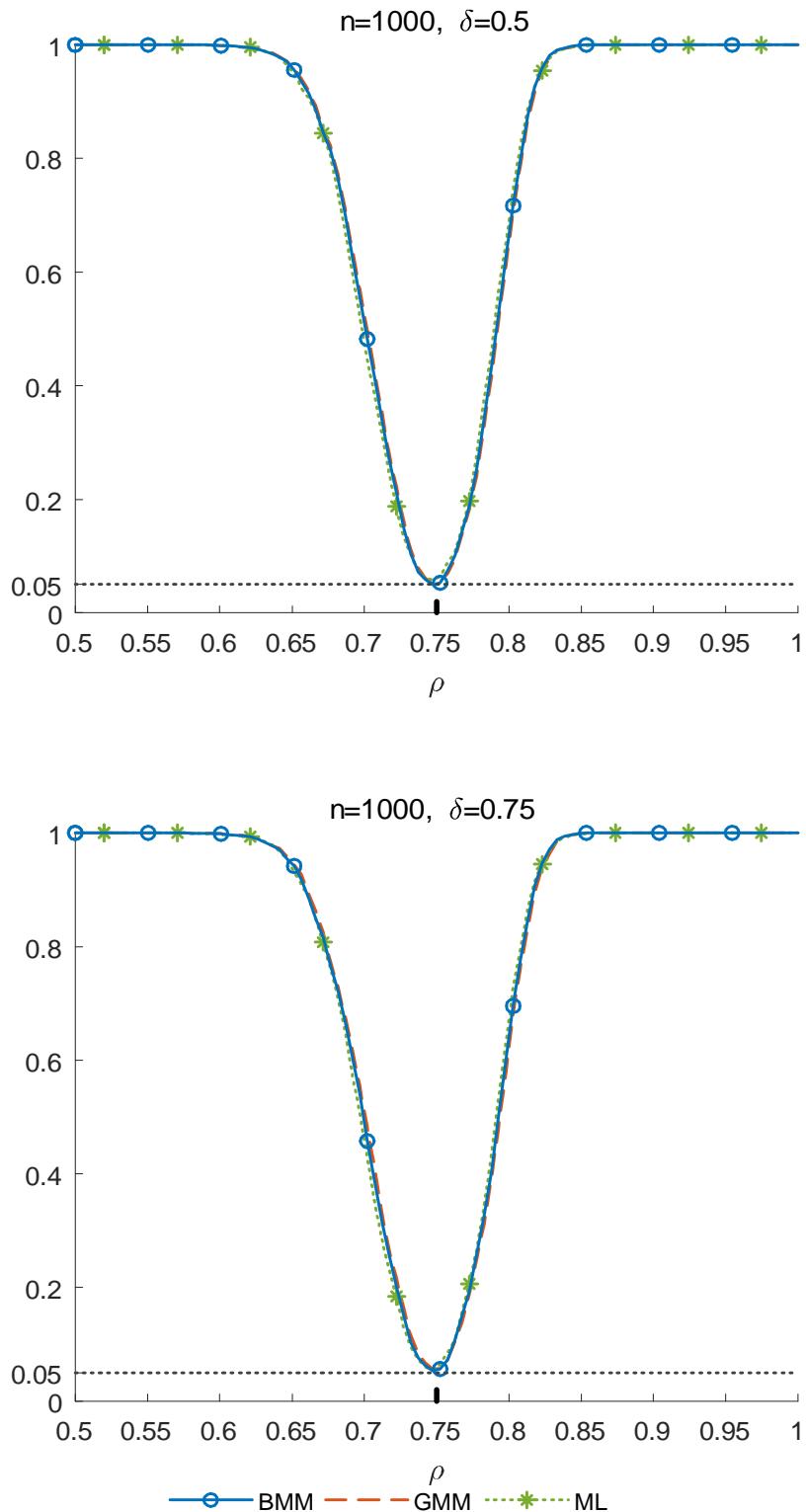


Figure S.12: (Continued)

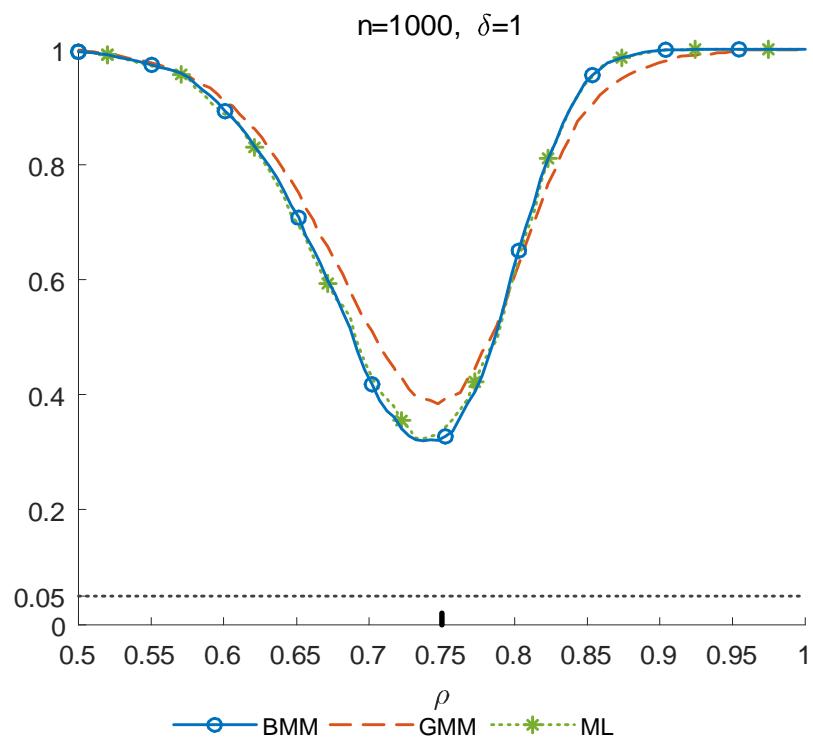
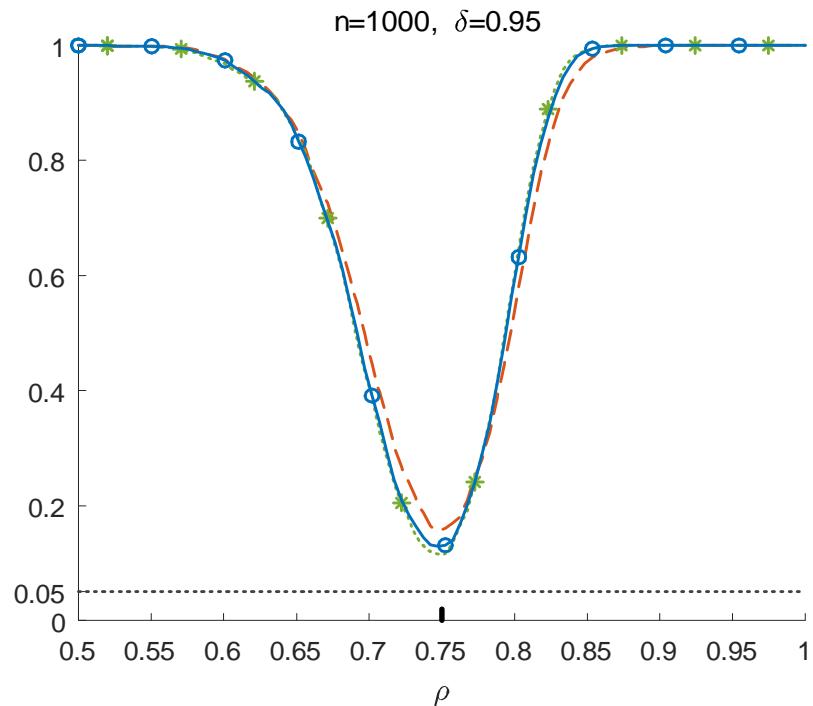


Figure S.12: (Continued)

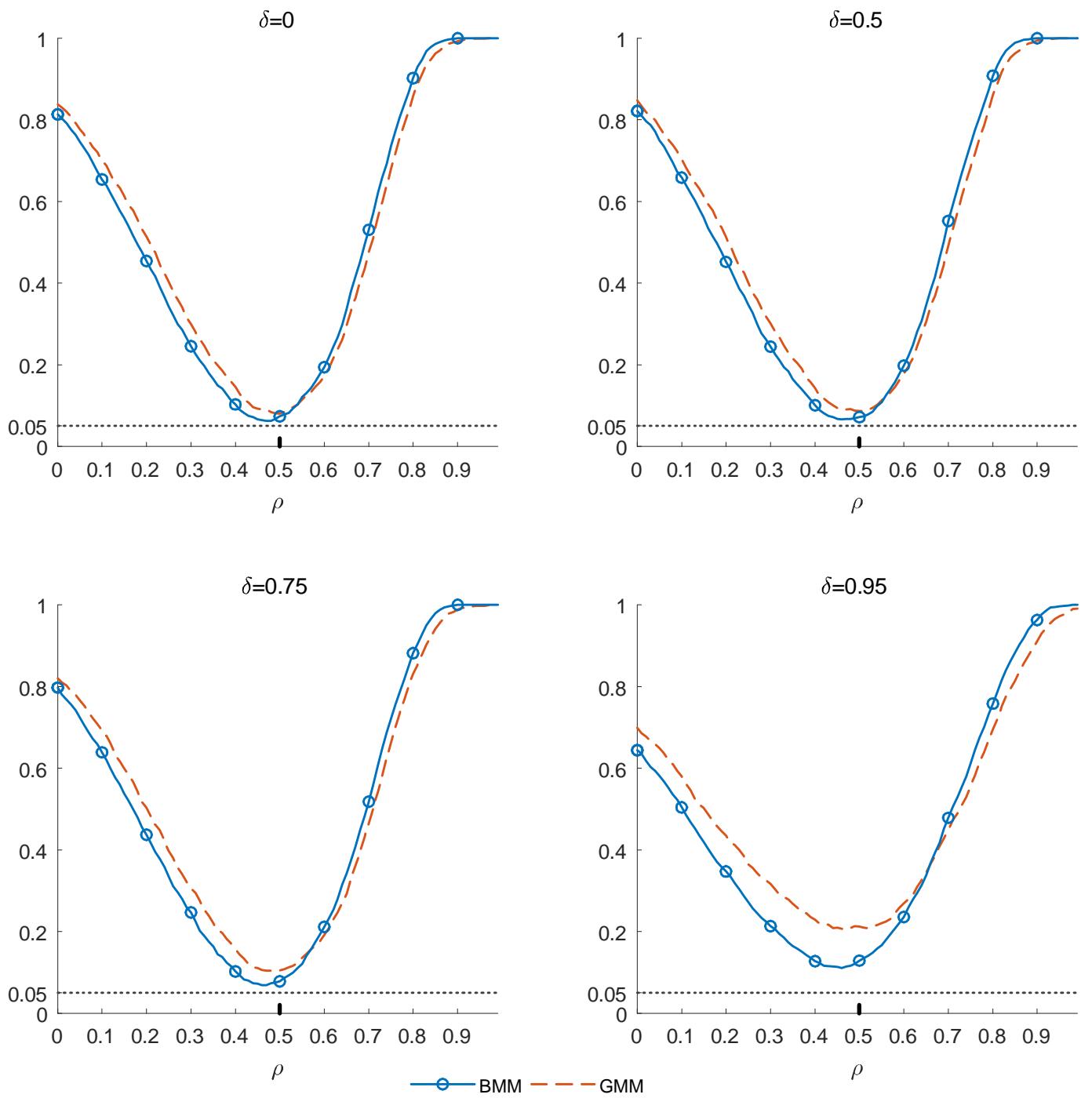


Figure S.13: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic Gaussian errors for different values of δ

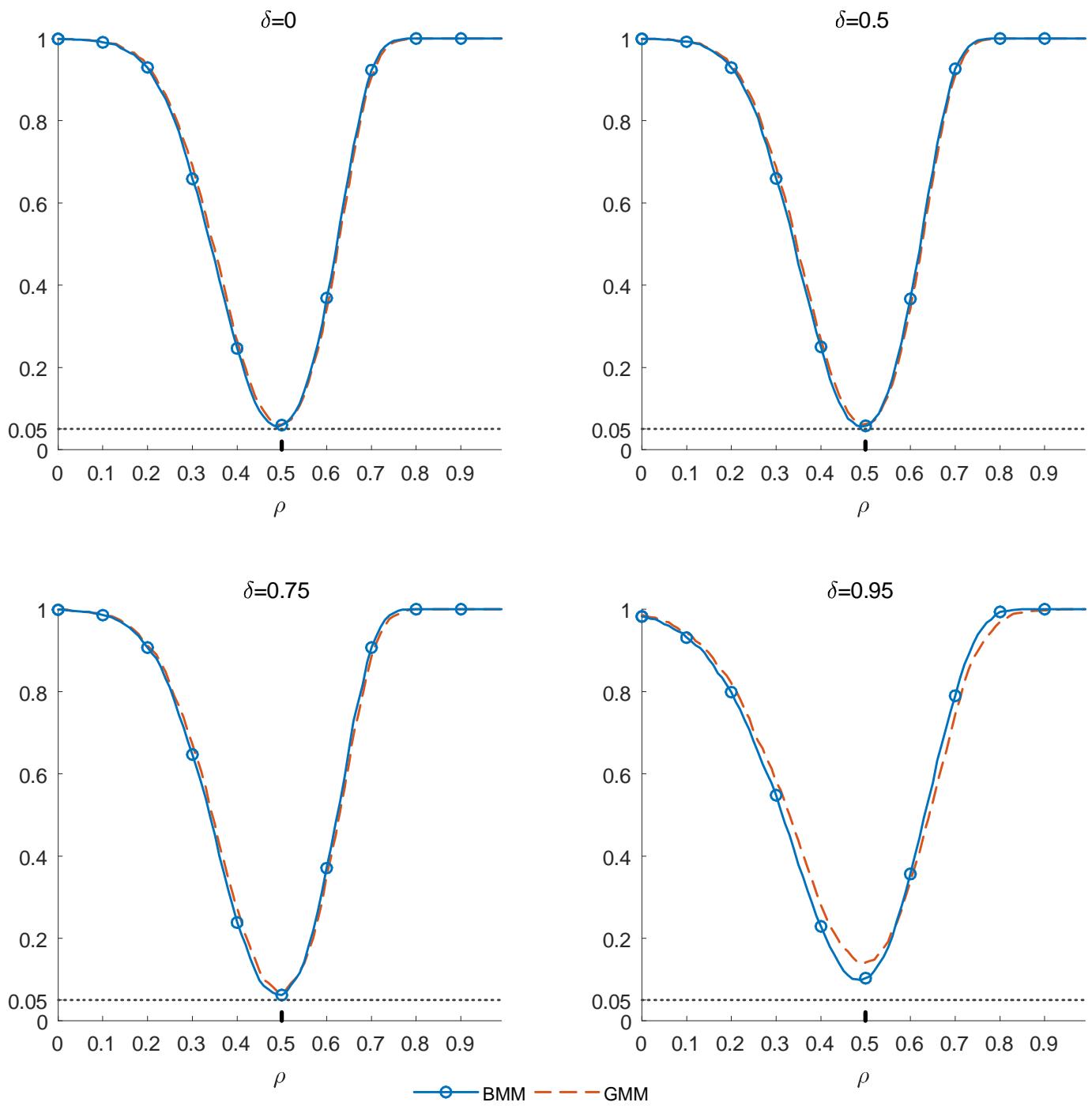


Figure S.14: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic Gaussian errors for different values of δ

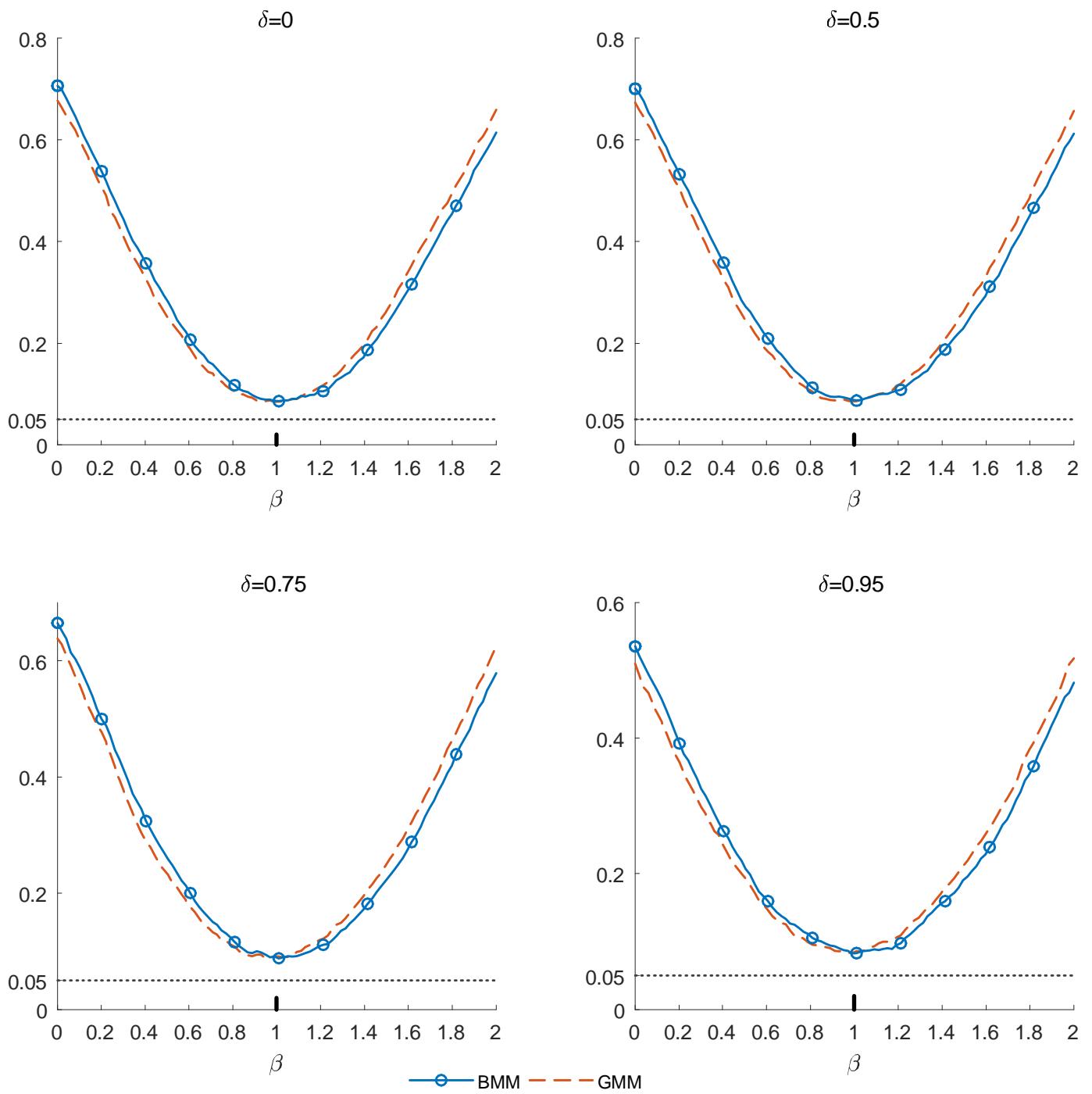


Figure S.15: Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic Gaussian errors for different values of δ

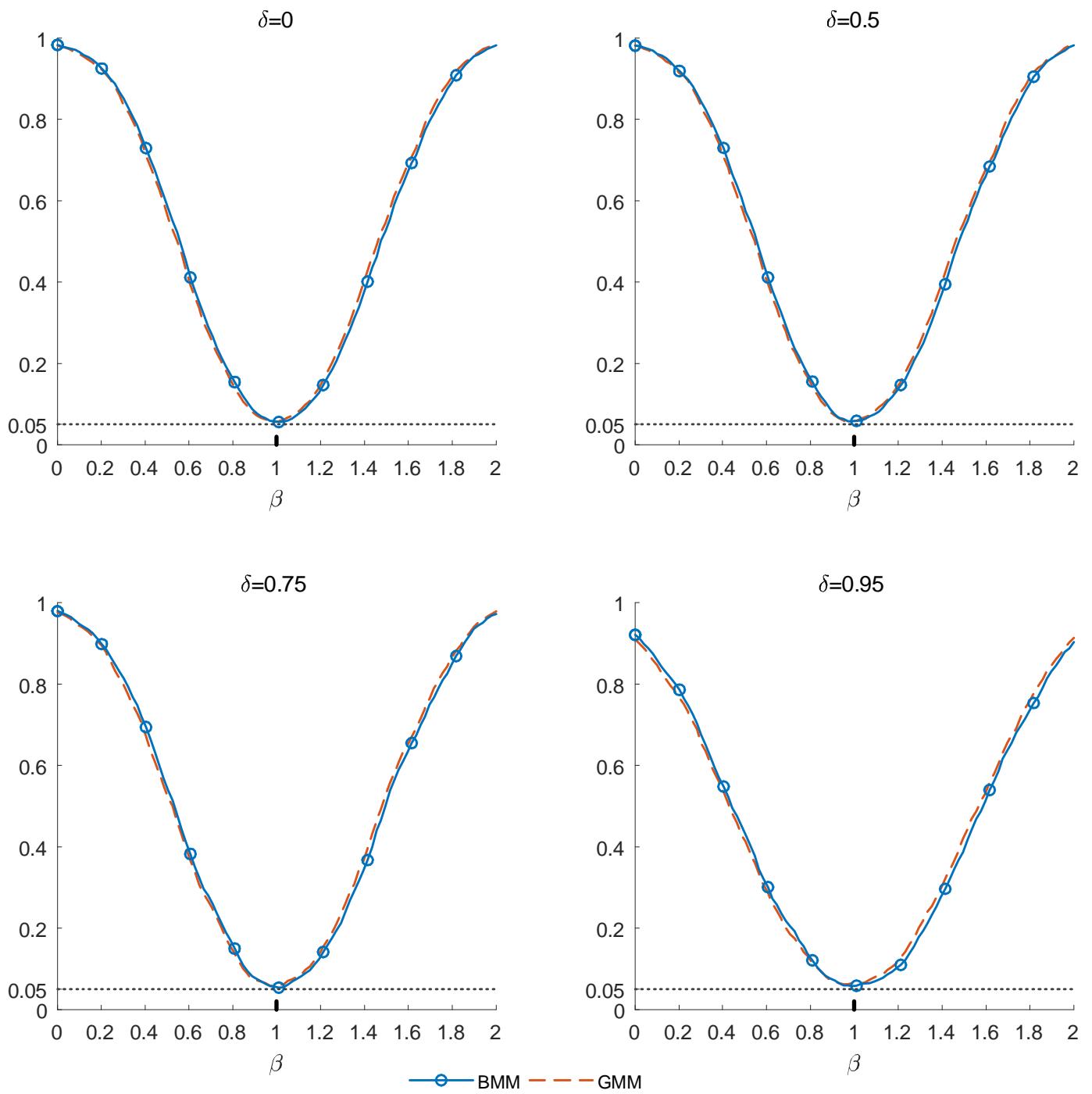


Figure S.16: Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic Gaussian errors for different values of δ

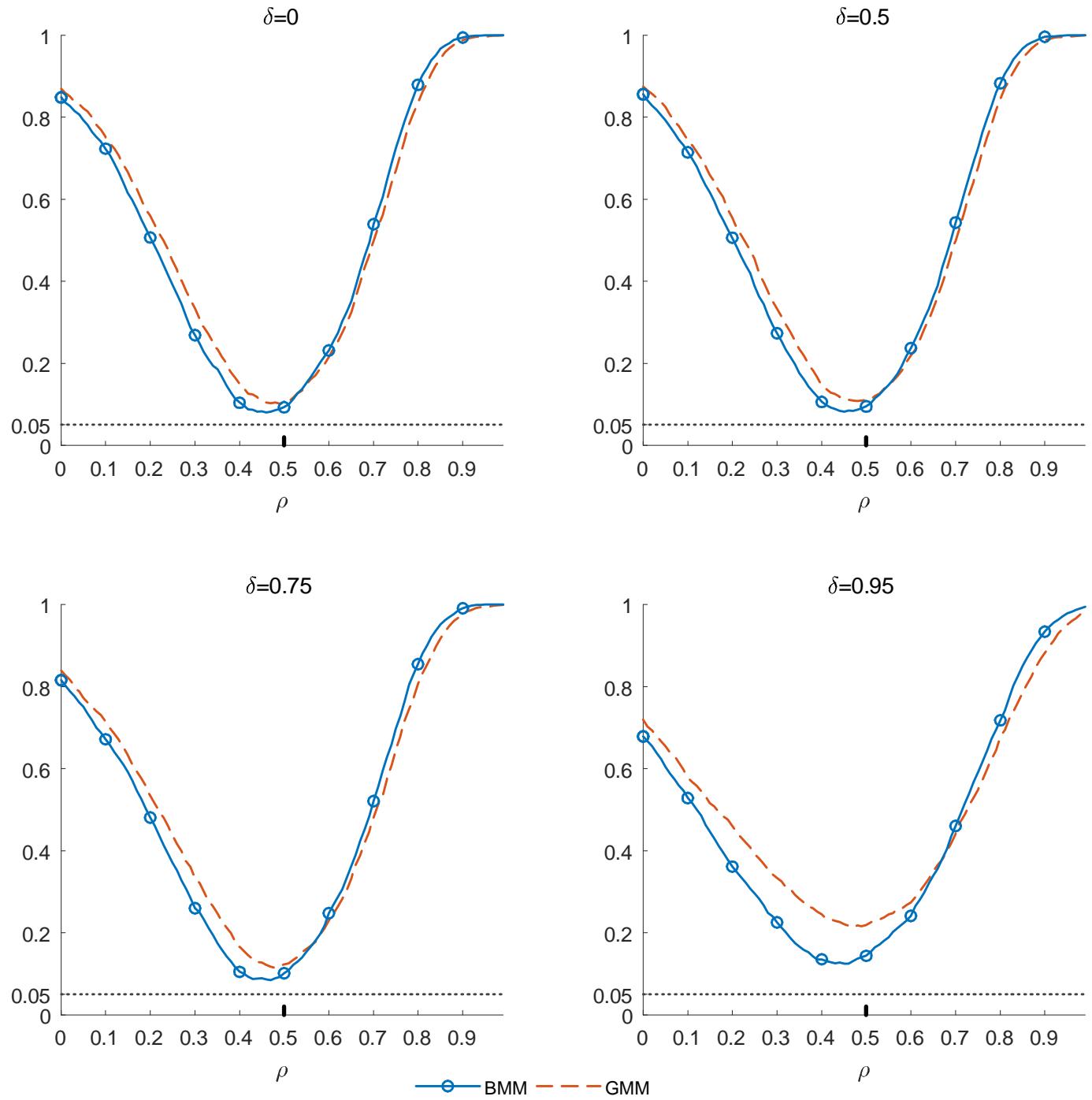


Figure S.17: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic non-Gaussian errors for different values of δ

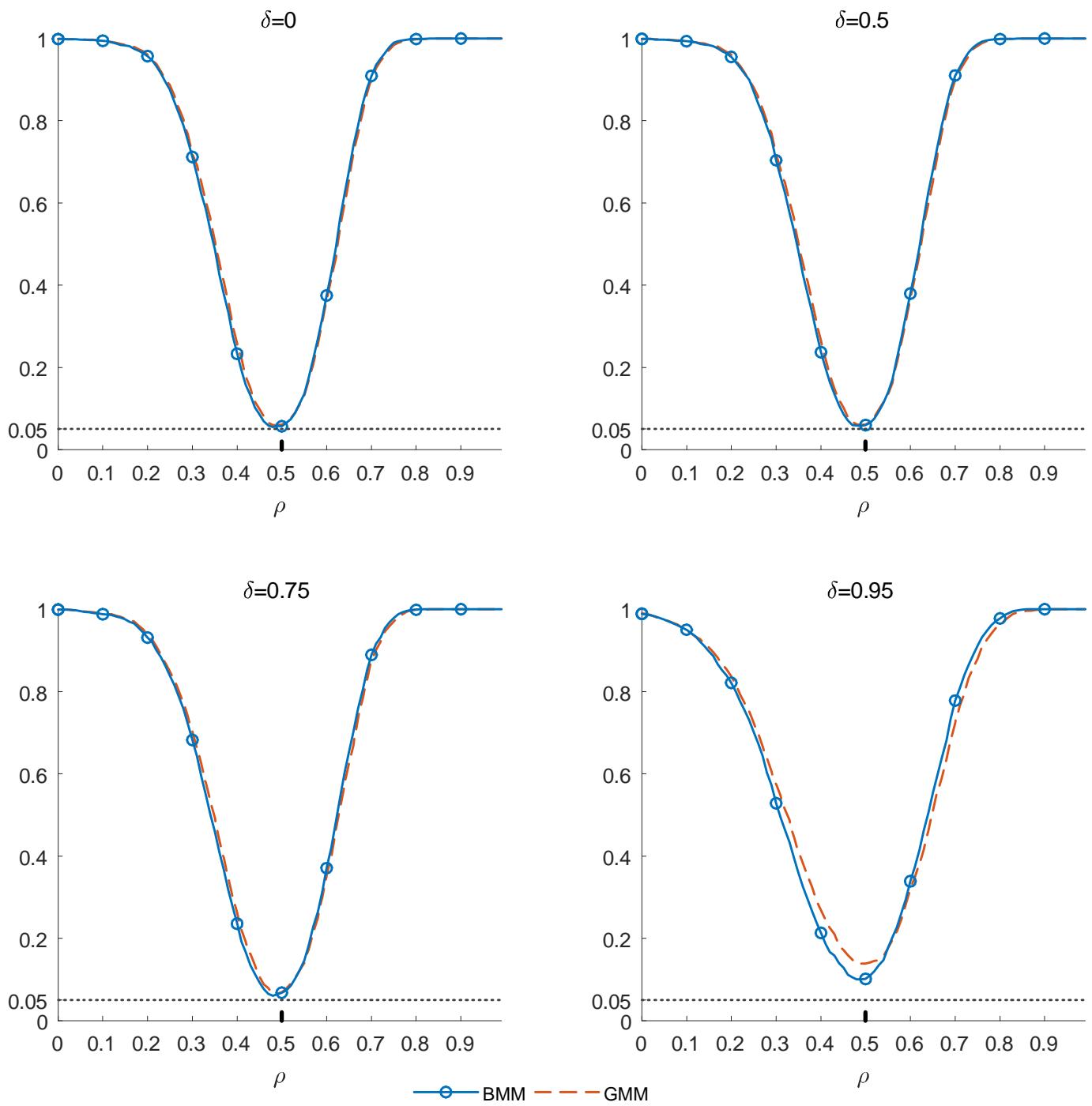


Figure S.18: Empirical power functions for ρ in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic non-Gaussian errors for different values of δ

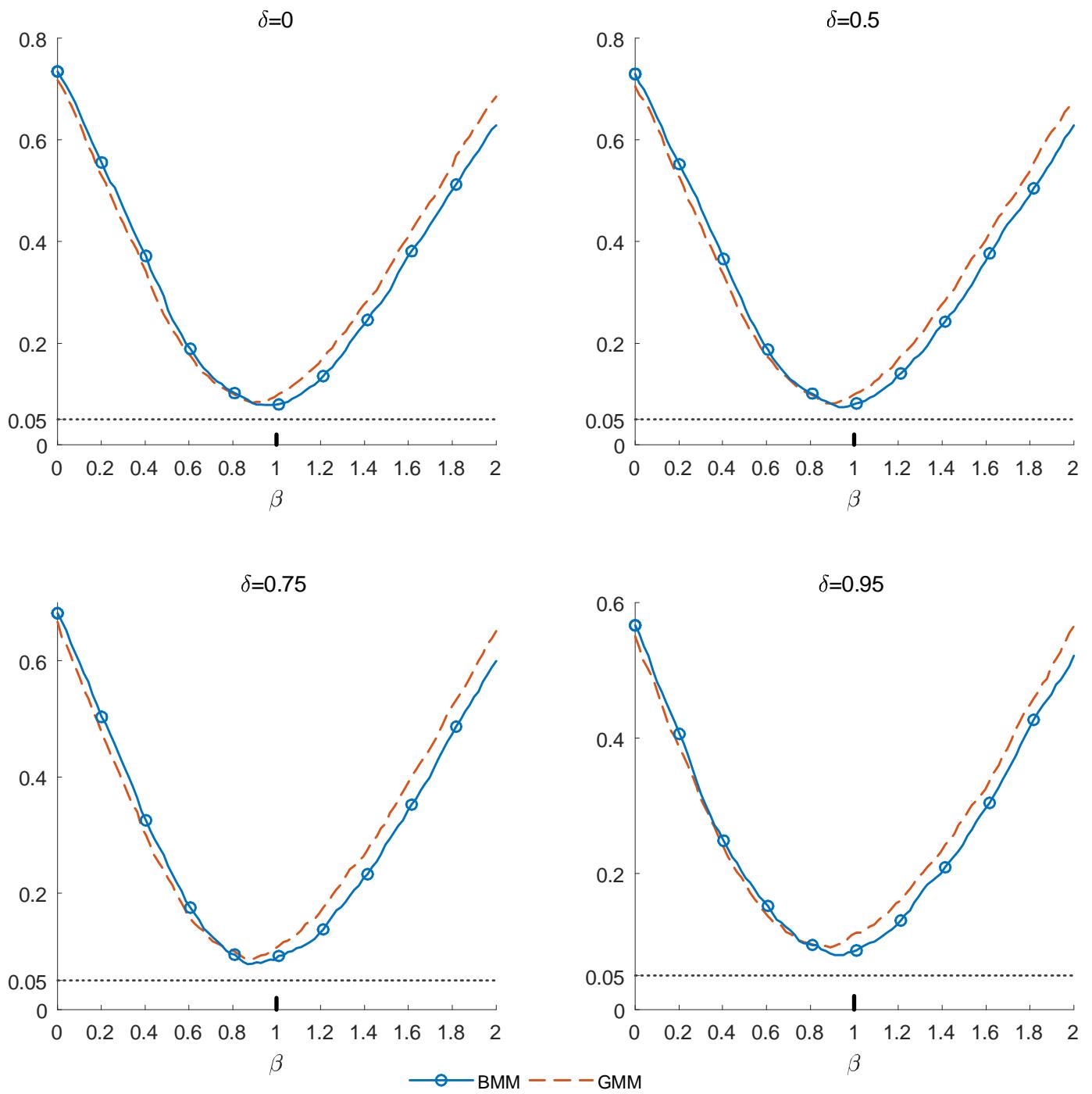


Figure S.19: Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 100$, and heteroskedastic non-Gaussian errors for different values of δ

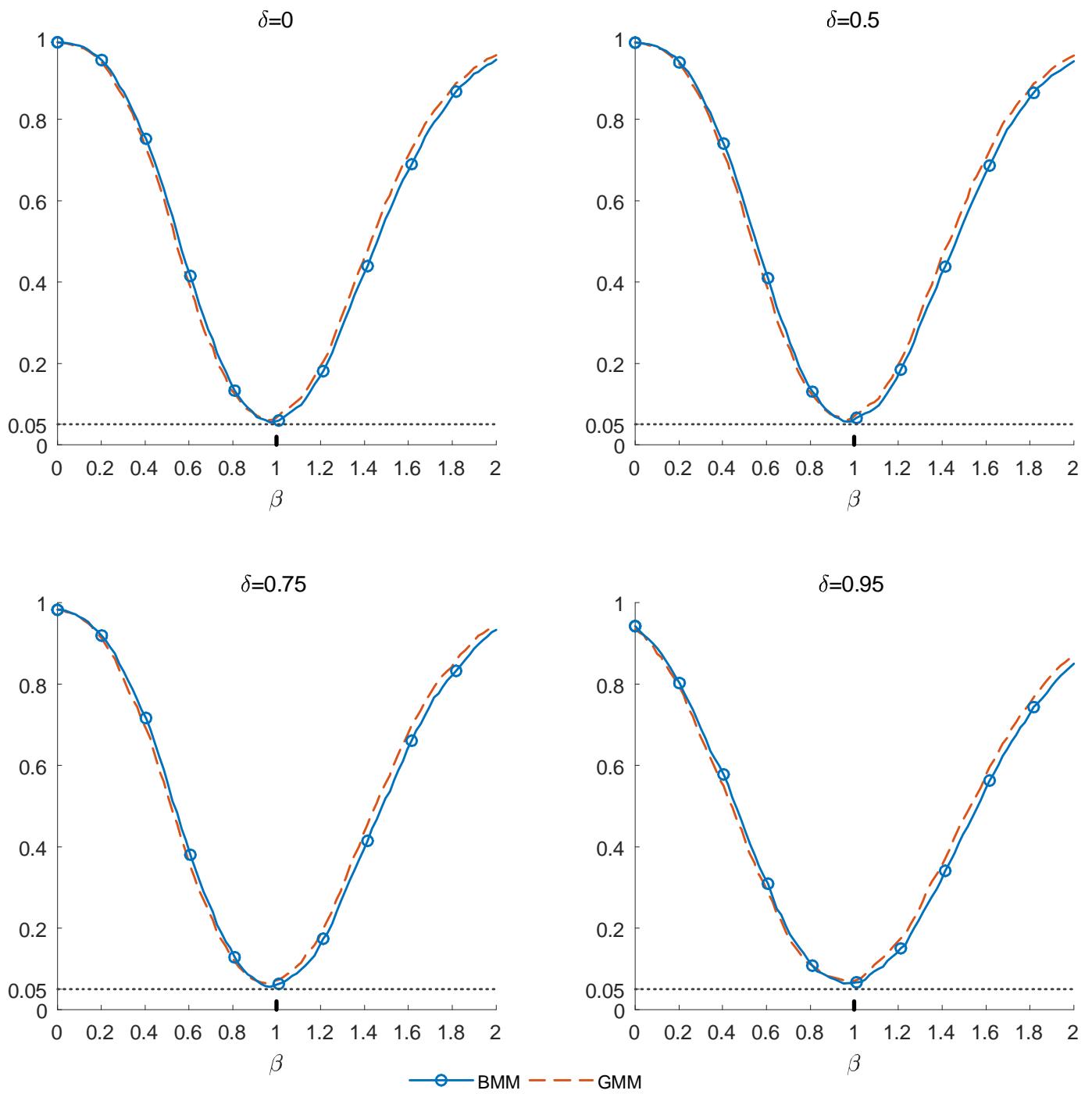


Figure S.20: Empirical power functions for β in the case of $\rho_0 = 0.5$, $n = 300$, and heteroskedastic non-Gaussian errors for different values of δ

S2.2 Additional empirical results

In the main paper we have presented results based on spatial weights matrices that are constructed using the threshold value of $\epsilon_w = 10\%$, with $\tilde{\mathbf{W}}(\epsilon_w)$ denoting a filtered version of $\mathbf{W} = (w_{ij})$. Specifically, $\tilde{\mathbf{W}}(\epsilon_w) = (\tilde{w}_{ij}(\epsilon_w))$, where $\tilde{w}_{ij}(\epsilon_w)$ is a row-standardized version of $w_{ij}^*(\epsilon_w)$ defined by $w_{ij}^*(\epsilon_w) = w_{ij}I(w_{ij} \geq \epsilon_w)$, and $I(A)$ is an indicator variable which takes the value of unity if A holds and zero otherwise. In this section we examine the robustness of our findings to different threshold values, including $\epsilon_w = 5\%$ and 7.5% . We begin by inspecting the δ -dominance of the top five pervasive sectors based on the three filtered weights, $\tilde{\mathbf{W}}(0.05) = (\tilde{w}_{ij}(0.05))$, $\tilde{\mathbf{W}}(0.075) = (\tilde{w}_{ij}(0.075))$ and $\tilde{\mathbf{W}}(0.1) = (\tilde{w}_{ij}(0.1))$. The results are summarized in Table S.45. The highest degree of dominance, $\hat{\delta}_{(1)}$, (which also measures the degree of network centrality) lies between 0.83 and 0.87, using the filtered weights matrices $\tilde{\mathbf{W}}(0.05)$ and $\tilde{\mathbf{W}}(0.075)$, respectively. In comparison, as documented in the paper $\hat{\delta}_{(1)}$ ranges between 0.71 to 0.85, if using the filtered weight matrix $\tilde{\mathbf{W}}(0.1)$. On average a lower cut-off value slightly increases the degree of network centrality.

Table S.45: 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		
	$\tilde{\mathbf{W}}_{2002}(0.05)$	$\tilde{\mathbf{W}}_{2002}(0.075)$	$\tilde{\mathbf{W}}_{2002}(0.1)$	$\tilde{\mathbf{W}}_{2007}(0.05)$	$\tilde{\mathbf{W}}_{2007}(0.075)$	$\tilde{\mathbf{W}}_{2007}(0.1)$
$\hat{\delta}_{(1)}$	0.830	0.870	0.851	0.843	0.837	0.705
$\hat{\delta}_{(2)}$	0.823	0.861	0.796	0.705	0.729	0.703
$\hat{\delta}_{(3)}$	0.583	0.594	0.642	0.679	0.650	0.695
$\hat{\delta}_{(4)}$	0.517	0.532	0.422	0.596	0.633	0.565
$\hat{\delta}_{(5)}$	0.476	0.513	0.402	0.539	0.558	0.491
n	309 [194]	306 [163]	286 [114]	384 [240]	380 [200]	350 [140]
n^*	1,396 (1.46%)	882 (0.94%)	581 (0.71%)	1,576 (1.07%)	968 (0.67%)	616 (0.50%)

Notes: $\hat{\delta}_{(1)} \geq \hat{\delta}_{(2)} \geq \dots \geq \hat{\delta}_{(5)}$ are the five largest estimates of the degree of dominance. $\tilde{\mathbf{W}}(\epsilon_w)$ denotes the robust \mathbf{W} matrix constructed with a threshold value of ϵ_w . The estimates for $\tilde{\mathbf{W}}_{2002}(0.1)$ and $\tilde{\mathbf{W}}_{2007}(0.1)$ reproduce the results in Table 7 of the paper, for ease of comparison. 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 in the related weight matrix. The percentages of non-zero elements are in parentheses.

Tables S.46 and S.47 report the estimation results of the SAR model given by (69) in the paper using threshold values $\epsilon_w = 5\%$ and 7.5% , respectively. The results suggest that a 5% cut-off value seem too low to yield a reasonable estimate of the share of capital, especially for the first sub-sample covering the years 1998–2006. This may be due to the inclusion of too many close to zero values in the network when a low cut-off value is selected. Compared with

Table S.46: Estimation results of the cross-section model (69) under threshold $\epsilon_w = 5\%$

Year	Sub-sample		Sub-sample	
	1998–2006	GMM ^b	2007–2015	GMM ^b
Assuming homoskedastic errors				
$\hat{\rho}$ [Share of capital]	0.772 [†] (0.063)	0.765 [†] (0.063)	0.487 [†] (0.081)	0.465 [†] (0.083)
$\hat{\sigma}_\eta^2$ [Error variance]	5.760	5.771	2.376	2.389
R^2	0.489	0.475	0.194	0.179
Assuming heteroskedastic errors				
$\hat{\rho}$ [Share of capital]	0.820 [†] (0.079)	0.812 [†] (0.077)	0.424 [†] (0.097)	0.411 [†] (0.098)
Weights matrix	$\tilde{\mathbf{W}}_{2002} (0.05)$		$\tilde{\mathbf{W}}_{2007} (0.05)$	
n [Number of sectors]	286		350	

Notes: The model is given by (69) in the paper, and includes an intercept (not reported). Standard errors are in parentheses. [†] indicates significance at the 1% level. The spatial weight matrices are constructed with the threshold value of $\epsilon_w = 5\%$. $\tilde{\mathbf{W}}_{2002} (0.05)$ is used when estimating the SAR model over the 1998–2006 period, and $\tilde{\mathbf{W}}_{2007} (0.05)$ is used when estimating the model over the 2007–2015 period. R^2 is computed by (59) assuming homoskedasticity in the paper. The BMM estimates assuming homoskedastic errors are computed by (38) in the paper, and computed by (48) in the paper if assuming heteroskedastic errors.

^b The GMM estimator refers to the best GMM estimator computed by a two-step procedure following (27) using the $\tilde{\mathbf{G}} - n^{-1} \text{Tr}(\tilde{\mathbf{G}}) \mathbf{I}_n$ if the errors are assumed to be homoskedastic, and $\tilde{\mathbf{G}} - \text{Diag}(\tilde{\mathbf{G}})$ if assuming heteroskedasticity, where $\tilde{\mathbf{G}} = \mathbf{G}(\tilde{\rho})$ is evaluated at the first-step estimate, $\tilde{\rho}$.

the results in Table 8 of the paper, we see that the estimates display less sensitivity to threshold values for the second sub-sample covering the years 2007–2015 as compared to the first sub-sample. Overall, it seems that the 10% cut-off value is a reasonable choice. Comparing the estimates obtained assuming homoskedastic errors with those assuming heteroskedastic errors, we consistently obtain slightly larger estimates for the pre-crisis period when allowing for heteroskedastic errors. As discussed in the main paper, this may suggest a high degree of heteroskedasticity in the pre-crisis period. Nonetheless, the estimates are overall close to each other within sampling errors.

Table S.47: Estimation results of the cross-section model (69) under threshold $\epsilon_w = 7.5\%$

Year	Sub-sample		Sub-sample	
	1998–2006	2007–2015	BMM	GMM ^b
Assuming homoskedastic errors				
$\hat{\rho}$ [Share of capital]	0.653 [†] (0.080)	0.633 [†] (0.082)	0.351 [†] (0.076)	0.332 [†] (0.078)
$\hat{\sigma}_\eta^2$ [Error variance]	6.275	6.316	2.469	2.484
R^2	0.424	0.398	0.160	0.147
Assuming heteroskedastic errors				
$\hat{\rho}$ [Share of capital]	0.736 [†] (0.103)	0.716 [†] (0.100)	0.309 [†] (0.076)	0.304 [†] (0.077)
Weights matrix	$\tilde{\mathbf{W}}_{2002} (0.075)$		$\tilde{\mathbf{W}}_{2007} (0.075)$	
n [Number of sectors]	306		380	

Notes: See the notes to Table S.46. The spatial weights matrices used to estimate ρ are constructed with the cut-off value of $\epsilon_w = 7.5\%$.

References

- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Volume 4. Springer Science & Business Media.
- 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.
- Lin, X. and Lee, L.F. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics*, 157(1), pp.34–52.
- Lütkepohl, H. (1996). *Handbook of Matrices*. John Wiley & Sons.