

Article

## Generalized Spatial Two Stage Least Squares Estimation of Spatial Autoregressive Models with Autoregressive Disturbances in the Presence of Endogenous Regressors and Many Instruments

Fei Jin <sup>1</sup> and Lung-fei Lee <sup>2,\*</sup>

<sup>1</sup> School of Economics, Shanghai University of Finance and Economics, Shanghai 200433, China; E-Mail: jinecon@gmail.com

<sup>2</sup> Department of Economics, The Ohio State University, Columbus, OH 43210, USA

\* Author to whom correspondence should be addressed; E-Mail: lee.1777@osu.edu; Tel.: +1-614-292-5508; Fax: +1-614-292-3906.

*Received: 25 March 2013; in revised form: 25 April 2013 / Accepted: 25 April 2013 /*

*Published: 27 May 2013*

---

**Abstract:** This paper studies the generalized spatial two stage least squares (GS2SLS) estimation of spatial autoregressive models with autoregressive disturbances when there are endogenous regressors with many valid instruments. Using many instruments may improve the efficiency of estimators asymptotically, but the bias might be large in finite samples, making the inference inaccurate. We consider the case that the number of instruments  $K$  increases with, but at a rate slower than, the sample size, and derive the approximate mean square errors (MSE) that account for the trade-offs between the bias and variance, for both the GS2SLS estimator and a bias-corrected GS2SLS estimator. A criterion function for the optimal  $K$  selection can be based on the approximate MSEs. Monte Carlo experiments are provided to show the performance of our procedure of choosing  $K$ .

**Keywords:** spatial autoregressive; spatial error; 2SLS; endogenous regressor; instrumental variable selection

---

## 1. Introduction

This paper considers the instrumental variable (IV) estimation of the spatial autoregressive (SAR) model with SAR disturbances (SARAR model) in the presence of endogenous regressors and many instruments. We study the case where the number of instruments increases with the sample size and derive asymptotic distributions of the generalized spatial two stage least squares (GS2SLS) estimator and a bias-corrected GS2SLS (CGS2SLS) estimator based on the leading-order many-instrument bias. Using many moments may improve the asymptotic efficiency but can make inference inaccurate in finite samples. [1] propose to minimize an approximate mean square error (MSE) as that of [2] for choosing the number of instruments in a cross section data model with endogenous regressors. The MSE takes into account an important bias term, so the method can avoid cases where asymptotic inferences are poor due to the bias being large relative to the standard deviation.

Ref [3] have derived the approximate MSEs of the two stage least squares (2SLS) and bias-corrected 2SLS (C2SLS) estimators for the SAR model with endogenous regressors and many instruments, but that SAR model has not included a SAR process in the disturbances. We extend the analysis to the SARAR model with endogenous regressors. The SARAR model combines spatial lag with spatial error dependence. The latter reflects spatial autocorrelation in measurement errors or in variables that are otherwise not crucial to the model [4,5]. It has a broader application than the simpler SAR model. It has been applied to empirical studies, e.g., Case's work [6–10]. Due to the presence of the spatial error dependence in addition to the spatial lag dependence, we consider the GS2SLS estimation of the model as in [11]. (Ref [12] have extended the estimation method in [11] to the SARAR model with endogenous regressors. Our focus here is on choosing the number of instruments by minimizing the approximated MSEs.) The estimation has taken into account the spatial error structure, based on a transformed equation. Because the transformation uses an initial consistent estimator of the spatial error dependence parameter, the impact from this initial estimator creates extra complexity that should be investigated. The analytical difficulty lies in determining the leading order terms depending on the number of instruments due to the presence of the spatial error process, whose orders cannot be expressed using terms appeared only in a SAR model without SAR disturbances. The approximated MSEs of the GS2SLS and CGS2SLS estimators turn out to be more complicated than those of the corresponding 2SLS and C2SLS estimators for the SAR model but are still tractable for empirical use. For the GS2SLS, the expression for the approximate MSE is similar to that for the 2SLS in [3], except for the presence of the filter for spatial error dependence in various matrices. If the formula for the approximate MSE in [3] is used for the SARAR models, then the derived number of instruments will not be asymptotically optimal. For the CGS2SLS estimator, however, except for the filter, the approximate MSE has additional terms compared with that for the C2SLS in [3], which are generated from the asymptotic distributions of the first two stage estimators.

We consider the following SARAR model:

$$y_n = \lambda W_n y_n + Z_{2n} \gamma + u_n, \quad u_n = \rho M_n u_n + \epsilon_n \quad (1)$$

where  $n$  is the number of spatial units,  $y_n$  is an  $n$ -dimensional vector of observations on the dependent variable, the  $n$ -dimensional vector of disturbances  $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nn})'$  has i.i.d. elements with mean zero and variance  $\sigma_\epsilon^2$ , and  $Z_{2n}$  is an  $n \times m$  matrix of variables that are possibly correlated with  $\epsilon_n$ ,  $W_n$

and  $M_n$  are  $n \times n$  spatial weights matrices that can be equal or different from each other, scalars  $\lambda$  and  $\rho$  are spatial autoregressive parameters, and  $\gamma$  is a parameter vector for  $Z_{2n}$ . Let  $Z_{2n} = \bar{Z}_{2n} + v_n$ , where  $\bar{Z}_{2n} = E(Z_{2n})$ . The  $\bar{Z}_{2n}$  is assumed to be an unknown function of  $X_n$ , which is an  $n \times k_x$  matrix of exogenous variables, and spatial lags of  $X_n$ :  $W_n X_n$ ,  $W_n^2 X_n$ , and so on. Model (1) can be an equation of the spatial simultaneous system as in [13]. In this case,  $y_n$  is a vector of observations on one of, say,  $k_y$  endogenous variables, and the equation for  $y_n$ , similar to those for other endogenous variables, is  $y_n = \lambda W_n y_n + X_{1n} \gamma_1 + Y_n \gamma_2 + u_n$ , where  $X_{1n}$  is the included exogenous variable matrix,  $Y_n$  is the endogenous variable matrix including all observations on the other  $(k_y - 1)$  endogenous variables and  $\gamma_1$  and  $\gamma_2$  are parameter vectors, then  $\bar{Z}_{2n} = \sum_{i=0}^{\infty} W_n^i X_n \Pi_i$ , where  $\Pi_i$ 's are matrices of parameters. Alternatively,  $Z_{2n}$  or some elements of  $Z_{2n}$  may be generated by an unknown nonlinear model [14], and thus we have an unknown nonlinear functional form for the conditional mean  $\bar{Z}_{2n}$  [1]. For  $v_n = (v'_{n1}, \dots, v'_{nn})'$ , we assume that  $v_{ni}$ 's are i.i.d. with mean zero and  $E(v'_{ni} v_{ni}) = \Sigma_v$ ,  $v_{ni}$  is independent of  $\epsilon_{nj}$  for  $j \neq i$ , but  $E(v_{ni} \epsilon_{ni}) = \sigma_{v\epsilon}$ . That is,  $v_{ni}$  and  $\epsilon_{ni}$  are correlated except with the exogenous explanatory variables. The  $i$ th variable in  $Z_{2n}$  is exogenous if the  $i$ th element of  $\sigma_{v\epsilon}$  is zero. Let  $Z_n = (W_n y_n, Z_{2n})$  and  $\delta = (\lambda, \gamma)'$ , then  $y_n = Z_n \delta + u_n$ .

We are interested in the parameter  $\delta$ . As in [11], the final generalized estimator for  $\delta$  is based on the Cochrane–Orcutt transformed equation:

$$R_n(\tilde{\rho}_n) y_n = \lambda R_n(\tilde{\rho}_n) Z_n \delta + R_n(\tilde{\rho}_n) u_n \quad (2)$$

where  $R_n(\tilde{\rho}_n) = I_n - \tilde{\rho}_n M_n$  with  $\tilde{\rho}_n$  being a consistent estimator of  $\rho$ . We consider the problem of choosing the number of instruments for  $R_n(\tilde{\rho}_n) Z_n$ , which can be many due to the unknown functional form of  $\bar{Z}_{2n}$  for its endogenous components. To derive  $\tilde{\rho}_n$ , we may first estimate the equation  $y_n = Z_n \delta + u_n$  by the 2SLS with a fixed number of instruments to obtain an initial estimator  $\check{\delta}_n$  of  $\delta$ , and then estimate  $\rho$  with a fixed number of quadratic moment equations that have the form  $\epsilon'_n(\rho, \check{\delta}_n) D_{nj} \epsilon_n(\rho, \check{\delta}_n) = 0$ , where the  $n \times n$  matrix  $D_{nj}$  has a zero trace, and  $\epsilon_n(\rho, \check{\delta}_n) = R_n(\rho)(y_n - Z_n \check{\delta}_n)$ . (The equation  $\epsilon'_n(\rho, \check{\delta}_n) D_{nj} \epsilon_n(\rho, \check{\delta}_n) = 0$  is a valid moment equation since  $E(\epsilon'_n D_{nj} \epsilon_n) = \sigma_\epsilon^2 \text{tr}(D_{nj}) = 0$  and  $\frac{1}{n}[\epsilon'_n(\rho_0, \check{\delta}_n) D_{nj} \epsilon_n(\rho_0, \check{\delta}_n) - \epsilon'_n D_{nj} \epsilon_n] = o_P(1)$  under regularity conditions.) The estimation thus involves three stages and the derivation of approximated MSEs is more complicated due to the presence of many terms with different orders. In [11], the asymptotic distribution of the third stage estimator  $\hat{\delta}_{2sls,n}$  is not affected by the estimators in the first two stages as long as  $\tilde{\rho}_n$  is a consistent estimator of  $\rho$ . For the approximate MSE of our GS2SLS estimator in the third stage, one may expect that it involves the asymptotic distributions of the first two stage estimators, since we use higher-order asymptotic theory for IV. However, it turns out that the variance of the dominant component related to the first two stage estimators in the expression for the GS2SLS estimator has a smaller order compared with other terms because of the i.i.d. property of  $\epsilon_{ni}$ 's. As a result, the leading order component of the MSE does not depend on the asymptotic distributions of the first two stage estimators and the expression for the approximate MSE is similar to that in [3] except for the filter for spatial error dependence. However, for the CGS2SLS estimator, the expression for the approximate MSE is more complicated than that in [3], because the term resulting from the estimation error of the leading order bias involves the asymptotic distributions of the first two stage estimators and an additional term appears due to the estimation of the spatial autoregressive parameter in the error process.

As  $\bar{Z}_{2n}$  is an unknown function of  $X_n, W_n X_n, W_n^2 X_n, \text{ etc.}$ , we may assume an infinite series approximation for  $\bar{Z}_{2n}$  and, in practice, use a known  $n \times q$  matrix  $\psi_{q,n}$  to approximate  $\bar{Z}_{2n}$ , where  $\psi_{q,n}$  depends on  $X_n, W_n X_n$  and so on. To closely approximate  $\bar{Z}_{2n}$  with a linear combination of  $\psi_{q,n}$ , we may need a large column number  $q$  as well as appropriate form of  $\psi_{q,n}$ . The instruments for  $W_n y_n$  can be based on  $\psi_{q,n}$ . Denote the true parameters for  $\delta$  and  $\rho$  by  $\delta_0$  and  $\rho_0$  respectively. As model (1) represents an equilibrium model,  $(I_n - \lambda_0 W_n)$  can be assumed to be invertible, where  $I_n$  is the  $n \times n$  identity matrix. (The SAR model is known as a simultaneous equation model in the spatial literature because the outcomes are determined by the interactions of spatial units. By assuming  $(I_n - \lambda_0 W_n)$  to be invertible, we have the equilibrium vector  $y_n$ .) Then, if  $\|\lambda_0 W_n\| < 1$  for some matrix norm  $\|\cdot\|$ , the equilibrium vector  $y_n = (I_n - \lambda_0 W_n)^{-1}(Z_{2n}\gamma_0 + u_n)$  can have an expansion  $\sum_{i=0}^{\infty} \lambda_0^i W_n^i (Z_{2n}\gamma_0 + u_n)$ . Therefore, the instruments for  $W_n y_n$  can be  $W_n \psi_{q,n}, W_n^2 \psi_{q,n}$  and so on, and the instruments for  $Z_n$  can be taken as the  $n \times K$  matrix

$$F_{K,n} = [\psi_{q,n}, W_n \psi_{q,n}, \dots, W_n^p \psi_{q,n}] \quad (3)$$

where  $K = (p + 1)q \geq m + 1$ . As an extension, we use the instrument matrix

$$Q_{K,n} = [F_{K,n}, M_n F_{K,n}] \quad (4)$$

for  $Z_n(\tilde{\rho}_n) = (I_n - \tilde{\rho}_n M_n)Z_n$ . (Due to technical difficulties in the presence of many IVs that involve estimated parameters in the literature, we do not use  $(I_n - \tilde{\rho}_n M_n)F_{K,n}$  as the instrument matrix for  $Z_n(\tilde{\rho}_n)$  (see [15]). If  $W_n = M_n$ , then  $M_n F_{K,n}$  generates some identical IVs as those in  $F_{K,n}$ . In this case, we can simply take  $Q_{K,n} = [F_{K,n}, W_n^{p+1} \psi_{q,n}]$ .) The asymptotic variance of the 2SLS estimator decreases when a linear combination of IVs approximates the conditional mean of the endogenous variables more closely. The efficiency (lower bound) of IV estimators is achieved when a linear combination of IVs equals the conditional mean [16]. Under regularity conditions, a linear combination of  $[I_n, W_n, W_n^2, \dots, W_n^p]$  can approximate  $(I_n - \rho W_n)^{-1}$  arbitrarily well as  $p \rightarrow \infty$ . Thus, if a linear combination of  $\psi_{q,n}$  can approximate  $\bar{Z}_{2n}$  well as  $n, q \rightarrow \infty$ , a linear combination of  $Q_{K,n}$  can approximate  $Z_n(\tilde{\rho}_n)$  arbitrarily well in probability as  $n, p, q \rightarrow \infty$ . On the other hand, if the number of instruments increases too fast relative to the sample size, they will lead to a bias of certain order for the corresponding IV estimators. The tradeoff between variance and bias can be summarized by the MSE of the estimator. So, minimizing the (approximated) MSE can reduce inaccurate inference due to the presence of many instruments. Following [1], we consider the case that the number of instruments  $K$  increases with, but at a rate slower than, the sample size  $n$ , which facilitates the investigation of the high order asymptotics of the MSEs.

The rest of the paper is organized as follows. Section 2 establishes asymptotic properties of the GS2SLS and CGS2SLS estimators. Section 3 derives the approximated MSEs for the estimators and gives a criterion function to choose the optimal number of IVs using the approximated MSEs. Section 4 presents some Monte Carlo results on the performance of the instrumental variable selection procedure in finite samples. Section 5 concludes. A list of notations, lemmas and proofs are collected in the appendices.

## 2. Properties of the GS2SLS and CGS2SLS Estimators

We establish the properties of the GS2SLS and CGS2SLS estimators in this section. Let  $R_n(\rho) = I_n - \rho M_n$ ,  $G_n(\lambda) = W_n(I_n - \lambda W_n)$ ,  $Z_n = \bar{Z}_n + \zeta_n$  with  $\bar{Z}_n = E(Z_n)$ , and  $\|A\| = \sqrt{\text{tr}(A'A)}$  be the Frobenius matrix norm for a matrix  $A$ . UB stands for boundedness of the sequences of both row and column sum matrix norms for a sequence of matrices. For simplicity, denote  $y_n(\rho) = R_n(\rho)y_n$ ,  $Z_n(\rho) = R_n(\rho)Z_n$ ,  $u_n(\rho) = R_n(\rho)u_n$ ,  $Z_{2n}(\rho) = R_n(\rho)Z_{2n}$ ,  $R_n = R_n(\rho_0)$ , and  $G_n = G_n(\lambda_0)$ . As  $y_n = (I_n - \lambda_0 W_n)^{-1}(Z_{2n}\gamma_0 + R_n^{-1}\epsilon_n)$ ,  $\bar{Z}_n = [G_n\bar{Z}_{2n}\gamma_0, \bar{Z}_{2n}]$  and  $\zeta_n = [G_n v_n \gamma_0 + G_n R_n^{-1}\epsilon_n, v_n]$ . The following are some basic regularity conditions.

**Assumption 1.**  $\{\epsilon_{ni}, v_{ni}\}$ 's,  $i = 1, \dots, n$ , are i.i.d. with mean zero,  $E(\epsilon_{ni}^2) = \sigma_\epsilon^2$ ,  $E(v_{ni}'v_{ni}) = \Sigma_v$  and  $E(v_{ni}\epsilon_{ni}) = \sigma_{v\epsilon}$ . The moments  $E|\epsilon_{ni}|^{4+\tau}$ ,  $E\|v_{ni}\|^4$  and  $E\|v_{ni}\epsilon_{ni}\|^2$  are finite, where  $\tau$  is some positive constant.

**Assumption 2.** (i) The sequences of matrices  $\{W_n\}$ ,  $\{M_n\}$ ,  $\{(I_n - \lambda_0 W_n)^{-1}\}$  and  $\{R_n^{-1}\}$  are UB; (ii)  $W_n$  and  $M_n$  have zero diagonals.

Since we use quadratic moments to estimate  $\rho$  in model (1), the existence of a moment of  $\epsilon_{ni}$  higher than the fourth order is required to properly apply the central limit theorem for linear-quadratic forms of disturbances in [17]. Some moment conditions are also imposed on  $v_{ni}$  and  $v_{ni}\epsilon_{ni}$  in Assumption 1. Assumption 2 (i), originated in [11,18], is a condition that bounds the degree of spatial dependence; Assumption 2 (ii) implies that no spatial unit is viewed as its own neighbor.

Let  $F_{0,n}$  be a full rank  $n \times k_f$  instrument matrix for  $Z_n$  in the first stage of the GS2SLS estimation. The number  $k_f$  of IVs is at least as large as the number  $(m + 1)$  of columns of  $Z_n$ , but is fixed for all  $n$ . Denote  $P_{F_n} = F_{0,n}(F_{0,n}'F_{0,n})^{-1}F_{0,n}'$ , where  $A^{-}$  is a generalized inverse for the matrix  $A$ . The first stage 2SLS estimator for  $\delta$  is  $\check{\delta}_n = (Z_n'P_{F_n}Z_n)^{-1}Z_n'P_{F_n}y_n$ . The following assumption about  $F_{0,n}$  is maintained.

**Assumption 3.** The instrument matrix  $F_{0,n}$  has full column rank  $k_f \geq m + 1$  for all  $n$ ,  $\lim_{n \rightarrow \infty} \frac{1}{n}F_{0,n}'F_{0,n}$  is finite and nonsingular, and  $\lim_{n \rightarrow \infty} \frac{1}{n}F_{0,n}'\bar{Z}_n$  is finite and has full column rank, where  $\bar{Z}_{2n}$  in  $\bar{Z}_n$  has uniformly bounded elements.

**Proposition 1.** Under Assumptions 1–3,  $\sqrt{n}(\check{\delta}_n - \delta_0) = (\frac{1}{n}\bar{Z}_n'P_{F_n}\bar{Z}_n)^{-1}\frac{1}{\sqrt{n}}\bar{Z}_n'P_{F_n}R_n^{-1}\epsilon_n + O_P(n^{-1/2}) \xrightarrow{d} N(0, \lim_{n \rightarrow \infty} (\frac{1}{n}\bar{Z}_n'P_{F_n}\bar{Z}_n)^{-1}\frac{\sigma_\epsilon^2}{n}\bar{Z}_n'P_{F_n}R_n^{-1}R_n^{-1}P_{F_n}\bar{Z}_n(\frac{1}{n}\bar{Z}_n'P_{F_n}\bar{Z}_n)^{-1})$ .

In the second stage of the GS2SLS estimation, we use a fixed number, say  $k_d$ , of quadratic moments to estimate  $\rho$  in model (1). Let  $g_n(\rho, \check{\delta}_n) = \frac{1}{n}[\epsilon_n'(\rho, \check{\delta}_n)D_{n1}\epsilon_n(\rho, \check{\delta}_n), \dots, \epsilon_n'(\rho, \check{\delta}_n)D_{n,k_d}\epsilon_n(\rho, \check{\delta}_n)]'$ , where  $\epsilon_n(\rho, \check{\delta}_n) = R_n(\rho)(y_n - Z_n\check{\delta}_n)$  and  $n \times n$  matrices  $D_{nj}$ 's have zero traces. The  $D_{nj}$ 's can be, e.g.,  $M_n$  and  $M_n^2 - I_n \text{tr}(M_n^2)/n$ . We maintain the following regularity condition on  $D_{nj}$ .

**Assumption 4.** The sequences of matrices  $\{D_{nj}\}$ ,  $j = 1, \dots, k_d$ , have zero traces and are UB.

Consider a generalized moments estimator  $\tilde{\rho}_n$  of  $\rho$ , which is

$$\tilde{\rho}_n = \arg \min_{\rho \in [-a,a]} g_n'(\rho, \check{\delta}_n)g_n(\rho, \check{\delta}_n) \tag{5}$$

for some  $a \geq 1$  so that  $[-a, a]$  contains  $\rho_0$ . It can be shown that  $g'_n(\rho, \check{\delta}_n)g_n(\rho, \check{\delta}_n) - E g'_n(\rho, \delta_0) E g_n(\rho, \delta_0)$  converges to zero in probability uniformly over  $[-a, a]$ . For the identification of  $\rho_0$ , it requires  $E g'_n(\rho, \delta_0) E g_n(\rho, \delta_0)$  to be zero uniquely at  $\rho_0$ . Let  $A^s = A + A'$  for any square matrix  $A$ . Note that  $E g_n(\rho, \delta_0) = \frac{\sigma_\epsilon^2}{2} \Xi_n [(\rho_0 - \rho), (\rho_0 - \rho)^2]'$ , where

$$\Xi_n = \frac{1}{n} \begin{pmatrix} \text{tr}[(M_n R_n^{-1})^s D_{n1}^s] & \text{tr}[(M_n R_n^{-1})' D_{n1}^s (M_n R_n^{-1})] \\ \vdots & \vdots \\ \text{tr}[(M_n R_n^{-1})^s D_{n,k_d}^s] & \text{tr}[(M_n R_n^{-1})' D_{n,k_d}^s (M_n R_n^{-1})] \end{pmatrix}$$

**Assumption 5.** The smallest eigenvalue of  $\Xi_n' \Xi_n$  is bounded away from zero.

Assumption 5 is satisfied if the limit of the  $2 \times 2$  matrix  $\Xi_n' \Xi_n$  exists and is nonsingular. With Assumption 5, there exists some  $\eta > 0$  such that  $E g'_n(\rho, \delta_0) E g_n(\rho, \delta_0) > \eta$  for any  $\rho \neq \rho_0$ . Thus for any  $\rho \neq \rho_0$ ,  $g'_n(\rho, \check{\delta}_n)g_n(\rho, \check{\delta}_n) > \eta/2$  with probability approaching 1 as  $n \rightarrow \infty$ .

**Proposition 2.** Under Assumptions 1–5,  $\tilde{\rho}_n$  is a consistent estimator of  $\rho_0$ , and

$$\sqrt{n}(\tilde{\rho}_n - \rho_0) = \frac{1}{\sqrt{n}}(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) + O_P(n^{-1/2})$$

is asymptotically normal with a finite variance, where

$$\mathbb{D}_n = \left( \frac{\sigma_\epsilon^2}{n^2} \sum_{j=1}^{k_d} \text{tr}^2(D_{nj}^s M_n R_n^{-1}) \right)^{-1} \sum_{j=1}^{k_d} \frac{1}{n} \text{tr}(D_{nj}^s M_n R_n^{-1}) D_{nj} \tag{6}$$

and

$$\mathbb{F}_n = - \left( \frac{\sigma_\epsilon^2}{n^2} \sum_{j=1}^{k_d} \text{tr}^2(D_{nj}^s M_n R_n^{-1}) \right)^{-1} \sum_{j=1}^{k_d} \frac{1}{n} \text{tr}(D_{nj}^s M_n R_n^{-1}) \frac{1}{n} E(\epsilon_n' D_{nj}^s R_n \zeta_n) \left( \frac{1}{n} \bar{Z}'_n P_{F_n} \bar{Z}_n \right)^{-1} \bar{Z}'_n P_{F_n} R_n^{-1} \tag{7}$$

with  $E(\epsilon_n' D_{nj}^s R_n \zeta_n) = [\text{tr}(D_{nj}^s R_n G_n) \sigma_{v\epsilon} \gamma_0 + \sigma_\epsilon^2 \text{tr}(D_{nj}^s R_n G_n R_n^{-1}), \text{tr}(D_{nj}^s R_n) \sigma_{v\epsilon}]$ .

In the expression for  $\sqrt{n}(\tilde{\rho}_n - \rho_0)$  above, the term  $\frac{1}{\sqrt{n}} \mathbb{F}_n \epsilon_n$  with the order  $O_P(1)$  is due to the usage of the first stage estimator  $\check{\delta}_n$ . That is to say that the asymptotic distribution of  $\check{\delta}_n$  has implication on the asymptotic distribution of  $\tilde{\rho}_n$ .

We now consider the GS2SLS estimator using the transformed Equation (2). With the instrument matrix  $Q_{K,n}$  in Equation (4), the GS2SLS estimator of  $\delta$  is

$$\hat{\delta}_{2sls,n} = [Z'_n(\tilde{\rho}_n) P_{K,n} Z_n(\tilde{\rho}_n)]^{-1} Z'_n(\tilde{\rho}_n) P_{K,n} y_n(\tilde{\rho}_n) \tag{8}$$

where  $P_{K,n} = Q_{K,n}(Q'_{K,n} Q_{K,n})^{-1} Q'_{K,n}$ .

**Assumption 6.** (i)  $\bar{H} = \lim_{n \rightarrow \infty} H_n$ , where  $H_n = \frac{1}{n} \bar{Z}'_n(\rho_0) \bar{Z}_n(\rho_0)$ , is a finite nonsingular  $(m+1) \times (m+1)$  matrix; (ii) for each  $Q_{K,n}$  in Equation (4), there exists  $\pi_{K,n}$  such that  $\frac{1}{n} \|\bar{Z}_n(\rho) - Q_{K,n} \pi_{K,n}\|^2 \rightarrow 0$  as  $n, K \rightarrow \infty$ .

**Assumption 6** (i) gives a sufficient condition for the identification of  $\delta_0$  in Equation (2); **Assumption 6** (ii) requires  $\bar{Z}_n(\rho)$  to be approximated arbitrarily well by a linear combination of  $Q_{K,n}$  for large enough  $K$  and  $n$ , which is implied by **Lemma 1** in **Section B** under some other basic assumptions. For analytical tractability, we maintain the following assumption.

**Assumption 7.** *The elements of  $Q_{K,n}$  in Equation (4) are uniformly bounded constants, and  $\lim_{n \rightarrow \infty} \frac{1}{n} Q'_{K,n} Q_{K,n}$  exists and is nonsingular for each  $K$ .*

The GS2SLS estimator  $\hat{\delta}_{2sls,n}$  is characterized by the first order condition  $\frac{1}{n} Z'_n(\tilde{\rho}_n) P_{K,n} [y_n(\tilde{\rho}_n) - Z_n(\tilde{\rho}_n) \hat{\delta}_{2sls,n}] = 0$ . By a Taylor expansion of this condition at  $\delta_0$ , the first term is  $\frac{1}{n} Z'_n(\tilde{\rho}_n) P_{K,n} u_n(\tilde{\rho}_n)$ , which has the dominant component  $\frac{1}{n} Z'_n(\rho_0) P_{K,n} \epsilon_n$  by **Lemma 8**. The expectation of this dominant component is  $\frac{1}{n} \Upsilon_n(K)$ , where

$$\Upsilon_n(K) = E(\zeta'_n R'_n P_{K,n} \epsilon_n) = [\text{tr}(\Gamma_{nK,2}) \sigma_{v\epsilon} \gamma_0 + \sigma_\epsilon^2 \text{tr}(\Gamma_{nK,3}), \text{tr}(\Gamma_{nK,1}) \sigma_{v\epsilon}]' = O(K) \tag{9}$$

with

$$\Gamma_{nK,1} = P_{K,n} R_n, \Gamma_{nK,2} = P_{K,n} R_n G_n, \text{ and } \Gamma_{nK,3} = P_{K,n} R_n G_n R_n^{-1} \tag{10}$$

Thus when  $K/n \rightarrow c \neq 0$ , the GS2SLS estimator  $\hat{\delta}_{2sls,n}$  is generally inconsistent. When  $K/n \rightarrow 0$ ,  $\hat{\delta}_{2sls,n}$  is consistent, but if the number of instruments  $K$  grows somehow fast relative to the sample size  $n$ , the asymptotic distribution may not center at the true  $\delta_0$ . The following proposition provides more information on this issue.

**Proposition 3.** *Under Assumptions 1–7,*

(i) *if  $K/n \rightarrow c \neq 0$ , then  $\hat{\delta}_{2sls,n} - \delta_0 \xrightarrow{P} \lim_{n \rightarrow \infty} \bar{b}_{nK,1}$ , where*

$$\bar{b}_{nK,1} = [\bar{Z}'_n(\rho_0) \bar{Z}_n(\rho_0) + \Omega_{n1}(K)]^{-1} \Upsilon_n(K) = O(K/n)$$

with

$$\begin{aligned} \Omega_{n1}(K) &= E(\zeta'_n R'_n P_{K,n} R_n \zeta_n) \\ &= \begin{pmatrix} \gamma'_0 \Sigma_v \gamma_0 \text{tr}(\Gamma'_{nK,2} \Gamma_{nK,2}) + \sigma_\epsilon^2 \text{tr}(\Gamma'_{nK,3} \Gamma_{nK,3}) + 2\sigma_{v\epsilon} \gamma_0 \text{tr}(\Gamma'_{nK,3} \Gamma_{nK,2}) & * \\ \Sigma_v \gamma_0 \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,2}) + \sigma'_{v\epsilon} \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,3}) & \Sigma_v \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,1}) \end{pmatrix} \end{aligned} \tag{11}$$

*might converge to a nonzero constant;*

(ii) *if  $K/n \rightarrow 0$ , then  $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0 - b_{n,K}) \xrightarrow{d} N(0, \sigma_\epsilon^2 \bar{H}^{-1})$ , where*

$$b_{n,K} = [Z'_n(\tilde{\rho}_n) P_{K,n} Z_n(\tilde{\rho}_n)]^{-1} \Upsilon_n(K) = \bar{b}_{nK,2} + o_P(K/n) \tag{12}$$

with  $\bar{b}_{nK,2} = [\bar{Z}'_n(\rho_0) \bar{Z}_n(\rho_0)]^{-1} \Upsilon_n(K) = O(K/n)$ .

From the above proposition, when  $K/n \rightarrow 0$ ,  $\hat{\delta}_{2sls,n}$  is consistent of  $\delta_0$ , but whether its asymptotic distribution is centered at  $\delta_0$  or not depends on the ratio  $K/\sqrt{n}$  as  $\sqrt{n} b_{n,K} = O_P(K/\sqrt{n})$ . The following corollary shows various scenarios.

**Corollary 1.** Under Assumptions 1–7,

- (i) if  $K^2/n \rightarrow 0$ ,  $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0) \xrightarrow{d} N(0, \sigma_\epsilon^2 \bar{H}^{-1})$ ;
- (ii) if  $K^2/n \rightarrow c < \infty$  and  $c \neq 0$ ,  $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0 - \bar{b}_{nK,2}) \xrightarrow{d} N(0, \sigma_\epsilon^2 \bar{H}^{-1})$ ;
- (iii) if  $K^2/n \rightarrow \infty$  but  $K^{1+\eta}/n \rightarrow 0$  for some  $0 < \eta < 1$ ,  $K^\eta(\hat{\delta}_{2sls,n} - \delta_0) \xrightarrow{p} 0$ .

When  $K^2/n \rightarrow 0$ , the number of instruments  $K$  increases slow relative to the sample size  $n$  and the asymptotic variance matrix  $\sigma_\epsilon^2 \bar{H}^{-1}$  achieves the efficiency lower bound for the class of IV estimators. When  $K^2/n$  goes to a non-zero limit as  $n$  goes to infinity,  $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0)$  is centered at  $\lim_{n \rightarrow \infty} \sqrt{n} \bar{b}_{nK,2}$ , which might be a non-zero finite constant and is a many instrument bias. Due to the spatial error dependence, the matrices  $\Gamma_{nK,1}$ ,  $\Gamma_{nK,2}$  and  $\Gamma_{nK,3}$  in Equation (10) of the bias component in Equation (9) play important roles. Without spatial error dependence, these matrices reduce to  $P_{K,n}$  and  $P_{K,n}G_n$ . Although the GS2SLS estimation is based on the spatial Cochrane–Orcutt transformed model (2), the asymptotic distribution of the estimator  $\tilde{\rho}_n$  in the transformation does not affect the asymptotic distribution of  $\hat{\delta}_{2sls,n}$ , as usual for the GS2SLS estimation.

To correct the many instrument bias, we consider a bias corrected estimator based on the estimation of the leading order bias  $b_{n,K}$  in Equation (12). Let  $Q_{0,n}$  be an instrument matrix with a fixed number of instruments and  $P_{0,n} = Q_{0,n}(Q'_{0,n}Q_{0,n})^{-1}Q'_{0,n}$ .

**Assumption 8.** The instrument matrix  $Q_{0,n}$  has full column rank  $k_q \geq m+1$  for all  $n$ ,  $\lim_{n \rightarrow \infty} \frac{1}{n}Q'_{0,n}Q_{0,n}$  is finite and nonsingular, and  $\lim_{n \rightarrow \infty} \frac{1}{n}Q'_{0,n}\bar{Z}_n(\rho_0)$  is finite and has full column rank.

The GS2SLS estimator

$$\tilde{\delta}_n = [Z'_n(\tilde{\rho}_n)P_{0,n}Z_n(\tilde{\rho}_n)]^{-1}Z'_n(\tilde{\rho}_n)P_{0,n}y_n(\tilde{\rho}_n) \tag{13}$$

and  $\tilde{\rho}_n$  together can be used to estimate  $b_{n,K}$ . Let  $\tilde{\Gamma}_{nK,1} = P_{K,n}R_n(\tilde{\rho}_n)$ ,  $\tilde{\Gamma}_{nK,2} = P_{K,n}R_n(\tilde{\rho}_n)G_n(\tilde{\lambda}_n)$ ,  $\tilde{\Gamma}_{nK,3} = P_{K,n}R_n(\tilde{\rho}_n)G_n(\tilde{\lambda}_n)R_n^{-1}(\tilde{\rho}_n)$ ,  $\tilde{\sigma}_\epsilon^2 = \frac{1}{n}(y_n - Z_n\tilde{\delta}_n)'R'_n(\tilde{\rho}_n)R_n(\tilde{\rho}_n)(y_n - Z_n\tilde{\delta}_n)$  and  $\tilde{\sigma}_{v\epsilon} = \frac{1}{n}(y_n - Z_n\tilde{\delta}_n)'R'_n(\tilde{\rho}_n)Z_{2n}$ . A bias-corrected GS2SLS (CGS2SLS) estimator is

$$\hat{\delta}_{c2sls,n} = \hat{\delta}_{2sls,n} - \tilde{b}_{n,K} \tag{14}$$

where  $\tilde{b}_{n,K} = [Z'_n(\tilde{\rho}_n)P_{K,n}Z_n(\tilde{\rho}_n)]^{-1}\tilde{\Upsilon}_n(K)$  with  $\tilde{\Upsilon}_n(K) = [\text{tr}(\tilde{\Gamma}_{nK,2})\tilde{\sigma}_{v\epsilon}\tilde{\gamma}_n + \tilde{\sigma}_\epsilon^2 \text{tr}(\tilde{\Gamma}_{nK,3}), \text{tr}(\tilde{\Gamma}_{nK,1})\tilde{\sigma}_{v\epsilon}]'$ .

**Proposition 4.** Under Assumptions 1–8, if  $K/n \rightarrow 0$ , then  $\sqrt{n}(\hat{\delta}_{c2sls,n} - \delta_0) \xrightarrow{d} N(0, \sigma_\epsilon^2 \bar{H}^{-1})$ .

Note that the asymptotic distribution of  $\hat{\delta}_{c2sls,n}$  in Equation (14) when  $K/n \rightarrow 0$  is the same as that of  $\hat{\delta}_{2sls,n}$  in Equation (8) when  $K^2/n \rightarrow 0$ . So the bias correction procedure has effectively relaxed some requirement on  $K$  in order for the corrected estimator to have a properly centered asymptotic distribution. The asymptotic distributions of the initial estimators  $\tilde{\delta}_n$  in Equation (13) and  $\tilde{\rho}_n$  in Proposition 2 used for the bias correction do not enter into the asymptotic distribution of  $\hat{\delta}_{c2sls,n}$ , when only the first order asymptotic expansion is considered. But when we investigate the approximated MSE of  $\hat{\delta}_{c2sls,n}$  later, as high order asymptotic expansions are considered, the asymptotic distributions of the estimators  $\tilde{\delta}_n$  and  $\tilde{\rho}_n$  used for the bias correction will generate additional terms for the approximated MSE.



### 3. Approximated MSE and Optimal $K$

For an estimator  $\hat{\delta}_n$  satisfying  $\sqrt{n}(\hat{\delta}_n - \delta_0) = \hat{H}_n^{-1}\hat{h}_n$ , [1] have derived a lemma that gives conditions on the decompositions of  $\hat{H}_n$  and  $\hat{h}_n$  such that the leading order term of the MSE depending on  $K$  is  $S_n(K)$ , in the sense that

$$n(\hat{\delta}_n - \delta_0)(\hat{\delta}_n - \delta_0)' = \hat{L}_n(K) + \hat{r}_n(K) \tag{15}$$

where  $E[\hat{L}_n(K)] = \sigma_\epsilon^2 H_n^{-1} + S_n(K) + T_n(K)$ , and  $T_n(K)$  and  $\hat{r}_n(K)$  are remainder terms that diminish faster than  $S_n(K)$ , such that  $[\hat{r}_n(K) + T_n(K)]/\text{tr}(S_n(K)) = o_P(1)$  as  $K, n \rightarrow \infty$ . A criterion function for the optimal  $K$  can be  $S_{n,\xi}(K) = \xi' S_n(K) \xi$ , the leading order MSE depending on  $K$  for a linear combination  $\xi' \hat{\delta}_n$ . In particular, one may use the unweighted version  $\text{tr}(S_n(K))$  as a practical criterion. Let  $\hat{S}_n(K)$  be an estimator of  $S_n(K)$ , then  $K$  can be chosen by minimizing the function  $\hat{S}_{n,\xi}(K) = \xi' \hat{S}_n(K) \xi$ .

In this section, we first derive the expression for  $S_n(K)$  for both the GS2SLS and CGS2SLS estimators and then show that the chosen  $K$  by minimizing  $\hat{S}_{n,\xi}(K)$  is asymptotically optimal in a sense in Equation (20) originated in [1]. Intuitively, this indicates that the error in the use of the feasible  $\hat{S}_{n,\xi}(K)$  criterion in place of the actual ideal  $S_{n,\xi}(K)$  is asymptotically negligible.

**Assumption 9.** (i)  $\frac{1}{K} \text{tr}(\Gamma_{nK,1}) \rightarrow c$ , where  $c \neq 0$ , as  $n, K \rightarrow \infty$ ;

(ii)  $\max_i |\Gamma_{nKj,ii}| \rightarrow 0$  for  $j = 1, 2, 3$ , as  $n, K \rightarrow \infty$ , where  $\Gamma_{nKj,ii}$  is the  $(i, i)$ th element of  $\Gamma_{nK,j}$ ;

(iii)  $\mu_3 = E(\epsilon_{ni}^3) = 0$  and  $E(\epsilon_{ni}^2 v_{ni}) = 0$ .

Assumption 9 (i) is for analytical tractability; Assumption 9 (ii) simplifies the expression for  $S_n(K)$  by imposing a restriction on the rate at which  $K$  increases with  $n$ ; Assumption 9 (iii) is also a condition that simplifies  $S_n(K)$ . These simplifications are adopted in [1,3]. (Without Assumption 9 (iii),  $S_n(K)$  for the GS2SLS will have an additional term  $\frac{1}{n} H_n^{-1} \{ \bar{Z}'_n(\rho_0) [E(\epsilon_{ni}^2 v_{ni}) \gamma_0 \text{vec}_D(\Gamma_{nK,2}) + \mu_3 \text{vec}_D(\Gamma_{nK,3}), E(\epsilon_{ni}^2 v_{ni}) \text{vec}_D(\Gamma_{nK,1})] \}^s H_n^{-1}$ , and  $S_n(K)$  for the CGS2SLS has an additional term that is much more complicated due to the estimator of  $\rho$  in the second stage of the GS2LS estimation and its use to correct the many instrument bias. Without Assumption 9 (ii),  $S_n(K)$  for the GS2SLS is not affected, but  $S_n(K)$  for the CGS2SLS has an additional term. Those additional terms can be estimated along with other terms, but they are not included here for simplicity.)

**Proposition 5.** Under Assumptions 1–9, if  $K^2/n \rightarrow 0$  and  $\sigma_{v\epsilon} \neq 0$ , then Equation (15) for the GS2SLS estimator  $\hat{\delta}_{2sls,n}$  is satisfied with

$$S_n(K) = \frac{1}{n} H_n^{-1} [\sigma_\epsilon^2 \bar{Z}'_n(\rho_0) (I_n - P_{K,n}) \bar{Z}_n(\rho_0) + \Omega_{n2}(K)] H_n^{-1} \tag{16}$$

where  $\Omega_{n2}(K) = \Upsilon_n(K) \Upsilon'_n(K)$ .

Note that  $S_n(K)$  above has a similar form as that in [3] except for the transformation  $R_n$  involved due to the spatial error dependence. The  $S_n(K)$  has a similar interpretation as that in [3]:  $\frac{\sigma_\epsilon^2}{n} H_n^{-1} \bar{Z}'_n(\rho_0) (I_n - P_{K,n}) \bar{Z}_n(\rho_0) H_n^{-1}$  is a variance term, which becomes smaller as a linear combination of  $Q_{K,n}$  approximates the mean of  $\bar{Z}_n(\rho_0)$  better;  $\frac{1}{n} H_n^{-1} \Omega_{n2}(K) H_n^{-1}$  is the leading order term in the MSE of  $\frac{1}{\sqrt{n}} H_n^{-1} \zeta'_n R'_n P_{K,n} \epsilon_n$  with the dominant component being from its expectation, which stands for the many instrument bias and increases as  $K$  increases. The minimization of a criterion function  $\xi' S_n(K) \xi$  thus takes into account the trade-off between the bias and variance.

**Proposition 6.** Under Assumptions 1–9, if  $K/n \rightarrow 0$  and  $\sigma_{ve} \neq 0$ , then Equation (15) for the CGS2SLS estimator  $\hat{\delta}_{c2sls,n}$  is satisfied with

$$S_n(K) = \frac{1}{n} H_n^{-1} [\sigma_\epsilon^2 \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) + \Pi_{n1}(K) + \Pi_{n2}(K) + \Pi_{n3}(K)] H_n^{-1} \quad (17)$$

where  $\Pi_{n1}(K)$ ,  $\Pi_{n2}(K)$  and  $\Pi_{n3}(K)$  are given in Equations (21), (22) and (25) respectively.

The first term in Equation (17) is the same as that in Equation (16). The second term  $\frac{1}{n} H_n^{-1} \Pi_{n1}(K) H_n^{-1}$  is the leading order term in the variance of  $\frac{1}{\sqrt{n}} H_n^{-1} [\zeta'_n R'_n P_{K,n} \epsilon_n - E(\zeta'_n R'_n P_{K,n} \epsilon_n)]$ . The third term  $\frac{1}{n} H_n^{-1} \Pi_{n2}(K) H_n^{-1}$  is due to the estimation error of the lead order bias of the GS2SLS estimator. This term becomes much more complicated than that for the SAR model because of the spatial error dependence. The last term  $\frac{1}{n} H_n^{-1} \Pi_{n3}(K) H_n^{-1}$  is an additional term compared with  $S_n(K)$  in [3], which is due to the estimation of  $\rho$ . (Thus, the  $\mathbb{F}_n$  in  $\Pi_{n2}(K)$  is from  $\tilde{\rho}_n$  used for the bias correction, and the  $\mathbb{F}_n$  in  $\Pi_{n3}(K)$  is from  $\tilde{\rho}_n$  in the spatial Cochrane–Orcutt transformation of the GS2SLS estimation.) The  $S_n(K)$  is a sum of different variance terms, because the bias terms have smaller orders compared with the variance terms.

We now consider the estimation of  $S_{n,\xi}(K) = \xi' S_n(K) \xi$ . Estimators for the parameters in  $S_{n,\xi}(K)$  can be constructed using a GS2SLS estimator. For the GS2SLS estimator, let the first stage IV matrix be  $F_{\bar{K},n}$  with  $\bar{K}$  instruments, the matrices for the quadratic moments in the second stage be  $D_{n1}, \dots, D_{n,\bar{k}_d}$ , and the last stage IV matrix be  $Q_{\bar{K},n} = [F_{\bar{K},n}, M_n F_{\bar{K},n}]$ . (The  $\bar{K}$  needs to increase with  $n$  so that the estimators for  $\sigma_{ve}$ ,  $\Sigma_v$  and  $\bar{H}$  defined below are consistent.) Then the first stage estimator for  $\delta$  is  $\hat{\delta} = (Z'_n P_{\bar{F}_n} Z_n)^{-1} Z'_n P_{\bar{F}_n} y_n$  with  $P_{\bar{F}_n} = F_{\bar{K},n} (F'_{\bar{K},n} F_{\bar{K},n})^{-1} F'_{\bar{K},n}$ , and the last stage estimator for  $\delta$  is  $\hat{\delta}_n = [Z'_n(\hat{\rho}_n) P_{\bar{K},n} Z_n(\hat{\rho}_n)]^{-1} Z'_n(\hat{\rho}_n) P_{\bar{K},n} y_n(\hat{\rho}_n)$  with  $P_{\bar{K},n} = Q_{\bar{K},n} (Q'_{\bar{K},n} Q_{\bar{K},n})^{-1} Q'_{\bar{K},n}$  and  $\hat{\rho}_n$  being the estimator for  $\rho$  in the second stage. Let the estimators for  $\sigma_\epsilon^2$ ,  $\sigma_{ve}$  and  $\Sigma_v$  be, respectively,  $\hat{\sigma}_\epsilon^2 = \frac{1}{n} \hat{\epsilon}'_n \hat{\epsilon}_n$ ,  $\hat{\sigma}_{ve} = \frac{1}{n} \hat{\epsilon}'_n \hat{v}_n$  and  $\hat{\Sigma}_v = \frac{1}{n} \hat{v}'_n \hat{v}_n$ , where  $\hat{\epsilon}_n = y_n(\hat{\rho}_n) - Z_n(\hat{\rho}_n) \hat{\delta}_n$  and  $\hat{v}_n = (I_n - P_{\bar{F}_n}) Z_{2n}$ . An estimator for  $\Omega_{n2}(K)$ ,  $\hat{\Omega}_{n2}(K)$ , can be derived by replacing the parameters with their respective estimators. An estimator for  $H_n$  is  $\hat{H}_n = \frac{1}{n} Z'_n(\hat{\rho}_n) P_{\bar{K},n} Z_n(\hat{\rho}_n)$ . For  $\frac{1}{n} \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)$ , note that

$$\begin{aligned} \frac{1}{n} E[Z'_n(\rho_0)(I_n - P_{K,n})Z_n(\rho_0)] &= \frac{1}{n} E\{[\bar{Z}_n(\rho_0) + R_n \zeta_n]'(I_n - P_{K,n})[\bar{Z}_n(\rho_0) + R_n \zeta_n]\} \\ &= \frac{1}{n} \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) + \frac{1}{n} E(\zeta'_n R'_n R_n \zeta_n) - \frac{1}{n} \Omega_{n1}(K) \end{aligned}$$

where  $\Omega_{n1}(K)$  is in Equation (11), thus  $\frac{1}{n} \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)$  can be estimated, up to an additive constant not depending on  $K$ , by  $\frac{1}{n} Z'_n(\hat{\rho}_n)(I_n - P_{K,n})Z_n(\hat{\rho}_n) + \frac{1}{n} \hat{\Omega}_{n1}(K)$ , where  $\hat{\Omega}_{n1}(K)$  is an estimator for  $\Omega_{n1}(K)$ , derived by replacing the parameters in  $\Omega_{n1}(K)$  by their estimators. Hence, for the GS2SLS,  $S_{n,\xi}(K) = \xi' S_n(K) \xi$  can be estimated, up to an additive constant not depending on  $K$ , by

$$\hat{S}_{n,\xi}(K) = \frac{1}{n} \xi' \hat{H}_n^{-1} [\hat{\sigma}_\epsilon^2 Z'_n(\hat{\rho}_n)(I_n - P_{K,n})Z_n(\hat{\rho}_n) + \hat{\sigma}_\epsilon^2 \hat{\Omega}_{n1}(K) + \hat{\Omega}_{n2}(K)] \hat{H}_n^{-1} \xi \quad (18)$$

Similarly, for the CGS2SLS,  $S_{n,\xi}(K)$  can be estimated, up to an additive constant not depending on  $K$ , by

$$\hat{S}_{n,\xi}(K) = \frac{1}{n} \xi' \hat{H}_n^{-1} [\hat{\sigma}_\epsilon^2 Z'_n(\hat{\rho}_n)(I_n - P_{K,n})Z_n(\hat{\rho}_n) + \hat{\sigma}_\epsilon^2 \hat{\Omega}_{n1}(K) + \hat{\Pi}_{n1}(K) + \hat{\Pi}_{n2}(K) + \hat{\Pi}_{n3}(K)] \hat{H}_n^{-1} \xi \quad (19)$$

where  $\hat{\Pi}_{n1}(K)$  is an estimator of  $\Pi_{n1}(K)$  derived by replacing the parameters in  $\Pi_{n1}(K)$  by their estimators,  $\hat{\Pi}_{n2}(K)$  is given in Equation (26) and  $\hat{\Pi}_{n3}(K)$  is given in Equation (27).

The optimal choice of  $K$  is the minimizer  $\hat{K}$  of  $\hat{S}_{n,\xi}(K)$ . The  $\hat{K}$  is optimal in the sense that  $S_{n,\xi}(\hat{K})$  is asymptotically as small as  $\min_K S_{n,\xi}(K)$ , i.e.,

$$\frac{S_{n,\xi}(\hat{K})}{\min_K S_{n,\xi}(K)} \xrightarrow{p} 1 \tag{20}$$

**Assumption 10.** (i)  $\sqrt{n}(\hat{\rho}_n - \rho_0) = O_P(1)$ ,  $\hat{\delta}_n \xrightarrow{p} \delta_0$ ,  $\hat{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2$ ,  $\hat{\sigma}_{v\epsilon} \xrightarrow{p} \sigma_{v\epsilon}$  and  $\hat{\Sigma}_v \xrightarrow{p} \Sigma_v$ ;  
 (ii) For the GS2SLS,  $|S_{n,\xi}(K)|/(K^2/n + \Delta_{nK,1}) > c$ , and for the CGS2SLS,  $|S_{n,\xi}(K)|/(K/n + \Delta_{nK,1}) > c$ , for some constant  $c > 0$ , where  $\Delta_{nK,1} = \frac{1}{n} \text{tr}[\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)]$ .

**Assumption 11.** For both the GS2SLS and CGS2SLS,  $\sum_K [nS_{n,\xi}(K)]^{-1} \rightarrow 0$ .

We assume the  $\sqrt{n}$ -consistency of  $\hat{\rho}_n$  and consistency of other preliminary estimators in Assumption 10 (i). Assumption 10 (ii) and Assumption 11 are similar to those in [3]. For the GS2SLS, from the proof of Proposition 5, the trace of the positive semi-definite matrix  $S_n(K)$  has exactly the same order as  $(K^2/n + \Delta_{nK,1})$ , then  $S_{n,\xi}(K)$  has the order  $O(K^2/n + \Delta_{nK,1})$ . Assumption 10 (ii) requires  $S_{n,\xi}(K)$  for the GS2SLS to have exactly the same order as  $(K^2/n + \Delta_{nK,1})$ . A similar condition on  $S_{n,\xi}(K)$  for the CGS2SLS is imposed. Assumption 11 imposes a restriction on the set of possible  $K$ .

**Proposition 7.** Under Assumptions 1–11, for  $\hat{K} = \arg \min_K \hat{S}_{n,\xi}(K)$ , Equation (20) is satisfied for both the GS2SLS and CGS2SLS.

#### 4. Monte Carlo Study

We demonstrate the finite sample performance of our instrument selection procedure with Monte Carlo experiments. Except for the additional spatial error dependence, most parts of the experimental design follow [3]. The model considered is

$$y_n = \lambda_0 W_n y_n + \gamma_0 Z_{2n} + u_n, \quad u_n = \rho_0 M_n u_n + \epsilon_n, \quad Z_{2n} = X_n \beta_0 + v_n$$

where  $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nn})'$ ,  $v_n = (v_{n1}, \dots, v_{nn})'$  and  $Z_{2n}$  is a vector. The  $(\epsilon_{ni}, v_{ni})$ 's are i.i.d. normal with mean zero,  $\epsilon_{ni}$  and  $v_{ni}$  both have unit variance, and the correlation coefficient between  $\epsilon_{ni}$  and  $v_{ni}$  is  $\sigma_{v\epsilon}$ , which will be varied by design. In the experiments,  $\gamma_0 = 1$ ,  $\lambda_0 = 0.6$ , and  $\rho_0 = 0.1$  or  $0.5$ . Elements of the  $n \times \bar{q}$  matrix  $X_n$  are random samples from the standard normal distribution. The specification implies a theoretical first stage coefficient of determination  $R_f^2 = \beta'_0 \beta_0 / (\beta'_0 \beta_0 + 1)$  (with the spatial dependence being ignored), according to [19]. The  $\bar{q}$  will be designed later on.

As in [3], we consider two models with different specifications of  $\beta_0$ . In Model 1, the coefficients are decreasing, i.e., the  $j$ th element of  $\beta_0$  is

$$\beta_{0j} = c(\bar{q}) \left(1 - \frac{j}{\bar{q} + 1}\right)^4, \text{ for } j = 1, \dots, \bar{q}$$

where  $c(\bar{q})$  is chosen such that  $R_f^2$  is equal to some specified value in the experiments; in Model 2, the coefficients are all equal, i.e.,

$$\beta_{0j} = \sqrt{\frac{R_f^2}{\bar{q}(1 - R_f^2)}}, \text{ for } j = 1, \dots, \bar{q}$$

These two specifications stand for, respectively, the case that some instruments are more important than others and the other case that no instrument should be preferred over others [1]. In the experiments,  $R_f^2$  is equal to 0.02 or 0.1,  $\sigma_{v\epsilon}$  is equal to 0.1, 0.5 or 0.9, and  $n = 98$  or 490. The  $W_n$  is a block diagonal matrix with each block in the diagonal being the row normalized matrix used for the study of crimes across 49 districts in Columbus, OH in [20]. The spatial weights matrix  $M_n$  in the error process is set to be the same as the spatial weights matrix  $W_n$ . The number of Monte Carlo repetitions is 2000.

Let  $X_n^q$  be a matrix consisting of the first  $q$  columns of  $X_n$ , and  $Q_{p,q} = [X_n^q, W_n X_n^q, \dots, W_n^p X_n^q]$ , for  $p = 1, 2, \dots, \bar{p}$  and  $q = 1, 2, \dots, \bar{q}$ . For  $n = 98$ , we set  $\bar{p} = 4$  and  $\bar{q} = 5$ ; for  $n = 490$ , we set  $\bar{p} = 10$  and  $\bar{q} = 10$ . The following estimators are considered:

- (i) GS2SLS-min: the GS2SLS with  $Q_{1,1}$  (as the instrument matrix in the third stage);
- (ii) GS2SLS-max: the GS2SLS with  $Q_{\bar{p},\bar{q}}$ ;
- (iii) GS2SLS-op: the GS2SLS with  $Q_{p,q}$ , where  $(p, q)$  minimizes  $\hat{S}_{n,\xi}(K)$  in Equation (18) with  $\xi = (1, 1)'$ ;
- (iv) CGS2SLS-max: the CGS2SLS with  $Q_{\bar{p},\bar{q}}$ ;
- (v) CGS2SLS-op: the CGS2SLS with  $Q_{p,q}$ , where  $(p, q)$  minimizes  $\hat{S}_{n,\xi}(K)$  in Equation (19) with  $\xi = (1, 1)'$ .

The leading order bias for the CGS2SLS and the approximated MSEs are estimated using the GS2SLS with  $Q_{2,\bar{q}}$  as the instrument matrix in the third stage. For all the GS2SLS and CGS2SLS estimators considered, the instrument matrix used in the first stage is  $Q_{2,\bar{q}}$ , and the matrices used for the quadratic moments in the second stage are  $W_n$  and  $W_n^2 - I_n \text{tr}(W_n^2)/n$ . (As  $\bar{q}$  is relatively large compared with the sample size, for the first stage estimator of the GS2SLS estimation and the estimator for the bias correction, we use  $p = 2$  as suggested by [11].)

For each estimator, the following robust measures of central tendency and dispersion are reported: (There are some outliers in the GS2SLS and CGS2SLS estimates, thus the mean and variance of the estimators are not reported.) the median bias (MB), the median of the absolute deviations (MAD), the difference between the 0.1 and 0.9 quantiles (DQ) in the empirical distribution, and the coverage rate (CR) of a nominal 95% confidence interval.

The summary statistics of the estimators for Model 1 are reported in Tables 1–4. We first compare GS2SLS-min, GS2SLS-max and GS2SLS-op. The GS2SLS-max has the largest median bias in most cases, and the GS2SLS-op has the smallest median bias for half of the cases when  $n = 98$  but it has the intermediate medium bias when  $n = 490$ . The GS2SLS-max has the smallest MAD and DQ in all cases, the GS2SLS-op of  $\lambda_0$  has the intermediate MAD and DQ, and GS2SLS-op of  $\gamma_0$  has the intermediate MAD and DQ when  $R_f^2 = 0.02$  but largest MAD and DQ when  $R_f^2 = 0.1$ . The CR of GS2SLS-op is closest to the nominal level in most cases, while the CR of GS2SLS-max is significantly lower than the nominal level in many cases. The CGS2SLS-max generally reduces the bias of GS2SLS-max significantly, has similar magnitudes of MAD and DQ to those of GS2SLS-max, and has a CR closer to the nominal level compared with GS2SLS-max but still significantly lower than the nominal level in many cases. Compared with the GS2SLS-op, in most cases, the CGS2SLS-op has much larger MAD and DQ, similar CR, and has smaller median bias for  $\lambda_0$  but larger median bias for  $\gamma_0$ .

**Table 1.** Estimation of Model 1 with  $R_f^2 = 0.02$  and  $n = 98$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.174	0.375	2.327	1.000	-0.011	0.612	3.618	1.000	
	GS2SLS-max	0.242	0.083	0.323	0.810	-0.065	0.175	0.654	0.992	
	GS2SLS-op	0.171	0.297	1.702	0.999	0.015	0.468	2.307	1.000	
	CGS2SLS-max	-0.046	0.125	0.667	0.870	0.098	0.295	1.248	0.974	
	CGS2SLS-op	-0.375	0.581	10.489	0.991	0.332	0.719	8.277	1.000	
	0.5	GS2SLS-min	0.157	0.428	2.917	1.000	0.188	0.582	3.415	1.000
	GS2SLS-max	0.156	0.071	0.279	0.921	0.347	0.154	0.609	0.932	
	GS2SLS-op	0.129	0.235	1.357	1.000	0.333	0.382	1.883	0.999	
	CGS2SLS-max	-0.008	0.081	0.374	0.958	0.407	0.246	0.983	0.824	
	CGS2SLS-op	-0.190	0.383	5.713	0.999	0.501	0.531	4.274	1.000	
	0.9	GS2SLS-min	0.148	0.295	2.039	1.000	0.293	0.456	3.633	0.982
	GS2SLS-max	0.064	0.031	0.120	0.968	0.791	0.081	0.306	0.033	
	GS2SLS-op	0.074	0.129	0.814	1.000	0.700	0.291	1.492	0.782	
	CGS2SLS-max	0.032	0.034	0.136	0.997	0.723	0.152	0.608	0.189	
CGS2SLS-op	0.011	0.160	1.544	1.000	0.628	0.355	2.091	0.820		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.349	0.342	2.413	0.992	0.026	0.572	3.234	1.000	
	GS2SLS-max	0.344	0.059	0.241	0.414	-0.070	0.176	0.696	0.992	
	GS2SLS-op	0.310	0.272	1.755	0.984	0.031	0.428	2.375	1.000	
	CGS2SLS-max	0.057	0.160	1.432	0.720	0.049	0.330	1.544	0.967	
	CGS2SLS-op	-0.137	0.730	10.221	0.972	0.203	0.810	7.607	1.000	
	0.5	GS2SLS-min	0.262	0.347	2.225	1.000	0.195	0.556	3.487	1.000
	GS2SLS-max	0.261	0.053	0.208	0.503	0.335	0.155	0.578	0.934	
	GS2SLS-op	0.227	0.208	1.342	0.991	0.350	0.403	2.108	1.000	
	CGS2SLS-max	0.092	0.077	0.401	0.855	0.400	0.254	1.022	0.848	
	CGS2SLS-op	-0.085	0.368	6.524	0.996	0.464	0.565	4.658	1.000	
	0.9	GS2SLS-min	0.228	0.219	1.672	0.992	0.339	0.460	3.461	0.973
	GS2SLS-max	0.181	0.027	0.103	0.224	0.775	0.066	0.264	0.023	
	GS2SLS-op	0.191	0.114	0.659	0.952	0.690	0.262	1.390	0.768	
	CGS2SLS-max	0.140	0.029	0.124	0.546	0.705	0.127	0.542	0.165	
CGS2SLS-op	0.115	0.133	1.600	0.966	0.639	0.310	2.055	0.809		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

**Table 2.** Estimation of Model 1 with  $R_f^2 = 0.1$  and  $n = 98$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.081	0.283	1.528	1.000	-0.024	0.240	1.026	1.000	
	GS2SLS-max	0.225	0.076	0.302	0.835	-0.062	0.135	0.548	0.995	
	GS2SLS-op	0.116	0.254	1.437	1.000	0.001	0.288	1.365	1.000	
	CGS2SLS-max	-0.033	0.109	0.558	0.887	0.039	0.197	0.829	0.979	
	CGS2SLS-op	-0.212	0.379	5.158	0.997	0.211	0.461	3.997	1.000	
	0.5	GS2SLS-min	0.082	0.231	1.368	1.000	-0.031	0.255	1.127	1.000
	GS2SLS-max	0.149	0.066	0.263	0.903	0.206	0.136	0.534	0.965	
	GS2SLS-op	0.078	0.213	1.203	0.998	0.147	0.312	1.452	1.000	
	CGS2SLS-max	-0.004	0.080	0.361	0.961	0.174	0.174	0.693	0.949	
	CGS2SLS-op	-0.155	0.283	3.963	0.999	0.312	0.381	2.418	1.000	
	0.9	GS2SLS-min	0.103	0.270	1.683	1.000	0.027	0.298	1.763	0.995
	GS2SLS-max	0.075	0.044	0.171	0.914	0.595	0.095	0.368	0.207	
	GS2SLS-op	0.071	0.182	1.185	0.998	0.284	0.357	1.671	0.928	
	CGS2SLS-max	0.022	0.049	0.210	0.985	0.407	0.153	0.605	0.598	
CGS2SLS-op	-0.034	0.190	2.795	1.000	0.374	0.394	2.175	0.913		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.253	0.313	1.991	0.996	0.021	0.273	1.243	1.000	
	GS2SLS-max	0.327	0.059	0.237	0.412	-0.058	0.146	0.568	0.995	
	GS2SLS-op	0.257	0.253	1.611	0.983	0.019	0.290	1.374	1.000	
	CGS2SLS-max	0.055	0.127	1.086	0.766	0.020	0.229	1.014	0.981	
	CGS2SLS-op	-0.159	0.472	6.950	0.972	0.144	0.557	5.132	1.000	
	0.5	GS2SLS-min	0.197	0.280	1.646	0.997	0.002	0.278	1.253	1.000
	GS2SLS-max	0.268	0.055	0.213	0.444	0.214	0.138	0.527	0.959	
	GS2SLS-op	0.217	0.232	1.415	0.991	0.166	0.316	1.515	1.000	
	CGS2SLS-max	0.087	0.083	0.421	0.826	0.197	0.192	0.802	0.941	
	CGS2SLS-op	-0.047	0.309	4.120	0.987	0.282	0.400	2.706	1.000	
	0.9	GS2SLS-min	0.222	0.262	1.671	0.994	0.013	0.239	1.165	0.995
	GS2SLS-max	0.217	0.030	0.118	0.156	0.488	0.080	0.322	0.334	
	GS2SLS-op	0.216	0.190	1.288	0.958	0.148	0.246	1.216	0.968	
	CGS2SLS-max	0.140	0.043	0.185	0.669	0.310	0.129	0.527	0.753	
CGS2SLS-op	0.077	0.185	2.591	0.966	0.249	0.327	1.756	0.947		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

**Table 3.** Estimation of Model 1 with  $R_f^2 = 0.02$  and  $n = 490$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.124	0.348	2.079	1.000	0.000	0.383	1.736	1.000	
	GS2SLS-max	0.245	0.040	0.147	0.168	-0.116	0.094	0.360	0.958	
	GS2SLS-op	0.158	0.210	1.203	0.995	-0.009	0.307	1.431	1.000	
	CGS2SLS-max	-0.023	0.056	0.310	0.866	0.046	0.149	0.603	0.904	
	CGS2SLS-op	-0.326	0.433	7.033	0.993	0.265	0.425	4.096	1.000	
	0.5	GS2SLS-min	0.145	0.322	1.972	1.000	0.024	0.364	1.766	1.000
	GS2SLS-max	0.156	0.031	0.117	0.367	0.306	0.080	0.302	0.587	
	GS2SLS-op	0.117	0.166	0.984	1.000	0.301	0.276	1.301	0.998	
	CGS2SLS-max	0.014	0.035	0.138	0.978	0.299	0.135	0.502	0.587	
	CGS2SLS-op	-0.128	0.269	4.360	1.000	0.364	0.360	1.961	0.999	
	0.9	GS2SLS-min	0.143	0.271	1.772	0.999	0.016	0.295	1.569	0.995
	GS2SLS-max	0.067	0.016	0.061	0.514	0.757	0.041	0.155	0.000	
	GS2SLS-op	0.089	0.183	1.014	0.998	0.348	0.274	1.361	0.898	
	CGS2SLS-max	0.038	0.019	0.076	0.934	0.558	0.088	0.342	0.043	
CGS2SLS-op	-0.011	0.163	1.762	1.000	0.423	0.284	1.513	0.850		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.241	0.333	2.121	0.996	0.009	0.382	1.682	1.000	
	GS2SLS-max	0.338	0.029	0.111	0.001	-0.111	0.098	0.370	0.948	
	GS2SLS-op	0.248	0.220	1.452	0.978	0.015	0.331	1.472	1.000	
	CGS2SLS-max	0.057	0.079	0.723	0.634	0.015	0.188	0.860	0.855	
	CGS2SLS-op	-0.241	0.530	9.160	0.936	0.218	0.572	5.418	1.000	
	0.5	GS2SLS-min	0.241	0.266	1.641	0.996	0.030	0.315	1.491	1.000
	GS2SLS-max	0.265	0.025	0.094	0.002	0.308	0.079	0.311	0.552	
	GS2SLS-op	0.230	0.163	0.956	0.971	0.274	0.284	1.231	1.000	
	CGS2SLS-max	0.106	0.038	0.179	0.575	0.302	0.140	0.551	0.572	
	CGS2SLS-op	-0.077	0.292	4.467	0.984	0.344	0.358	2.332	0.999	
	0.9	GS2SLS-min	0.218	0.263	1.765	0.994	0.075	0.294	1.820	0.995
	GS2SLS-max	0.184	0.012	0.046	0.000	0.754	0.037	0.138	0.000	
	GS2SLS-op	0.204	0.161	0.961	0.963	0.377	0.256	1.220	0.887	
	CGS2SLS-max	0.142	0.015	0.058	0.032	0.580	0.084	0.319	0.031	
CGS2SLS-op	0.111	0.151	2.019	0.950	0.421	0.287	1.530	0.836		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

**Table 4.** Estimation of Model 1 with  $R_f^2 = 0.1$  and  $n = 490$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.032	0.154	0.801	0.999	-0.016	0.131	0.526	1.000	
	GS2SLS-max	0.214	0.037	0.144	0.257	-0.068	0.076	0.274	0.984	
	GS2SLS-op	0.126	0.211	1.258	0.999	-0.004	0.296	1.475	1.000	
	CGS2SLS-max	0.007	0.044	0.176	0.970	0.009	0.093	0.362	0.979	
	CGS2SLS-op	-0.172	0.301	3.807	0.999	0.209	0.388	2.928	1.000	
	0.5	GS2SLS-min	0.045	0.150	0.834	0.999	-0.015	0.136	0.553	1.000
	GS2SLS-max	0.165	0.031	0.121	0.290	0.199	0.067	0.260	0.792	
	GS2SLS-op	0.097	0.221	1.402	1.000	0.110	0.301	1.433	1.000	
	CGS2SLS-max	0.029	0.035	0.139	0.967	0.112	0.086	0.328	0.922	
	CGS2SLS-op	-0.113	0.258	3.874	1.000	0.248	0.338	2.203	1.000	
	0.9	GS2SLS-min	0.053	0.147	0.975	1.000	-0.003	0.136	0.574	0.998
	GS2SLS-max	0.114	0.019	0.073	0.144	0.503	0.044	0.167	0.003	
	GS2SLS-op	0.107	0.182	1.080	0.996	0.106	0.220	0.980	0.986	
	CGS2SLS-max	0.060	0.026	0.103	0.861	0.217	0.075	0.273	0.643	
CGS2SLS-op	-0.046	0.216	2.924	1.000	0.280	0.364	2.083	0.957		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.072	0.189	1.255	0.996	0.003	0.131	0.525	1.000	
	GS2SLS-max	0.316	0.030	0.115	0.006	-0.054	0.073	0.287	0.983	
	GS2SLS-op	0.211	0.238	1.563	0.986	0.020	0.276	1.241	1.000	
	CGS2SLS-max	0.079	0.054	0.277	0.718	0.014	0.108	0.431	0.957	
	CGS2SLS-op	-0.137	0.382	5.654	0.967	0.205	0.453	3.517	1.000	
	0.5	GS2SLS-min	0.097	0.173	1.275	0.993	-0.006	0.140	0.595	1.000
	GS2SLS-max	0.264	0.025	0.101	0.006	0.200	0.068	0.263	0.776	
	GS2SLS-op	0.191	0.219	1.377	0.991	0.116	0.258	1.184	0.999	
	CGS2SLS-max	0.110	0.034	0.150	0.570	0.108	0.092	0.361	0.910	
	CGS2SLS-op	-0.028	0.291	4.901	0.985	0.206	0.330	2.157	1.000	
	0.9	GS2SLS-min	0.098	0.156	1.341	0.989	-0.005	0.148	0.638	0.999
	GS2SLS-max	0.210	0.017	0.064	0.000	0.482	0.044	0.167	0.004	
	GS2SLS-op	0.150	0.180	1.114	0.977	0.120	0.191	0.833	0.996	
	CGS2SLS-max	0.138	0.022	0.088	0.183	0.195	0.078	0.300	0.702	
CGS2SLS-op	0.039	0.213	3.970	0.974	0.205	0.307	1.865	0.969		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

Tables 5–8 report the summary statistics of the estimators for Model 2. Among GS2SLS-min, GS2SLS-max and GS2SLS-op, in most cases, the GS2SLS-max has the largest median bias, the GS2SLS-op of  $\lambda_0$  has the smallest median bias, and the GS2SLS-op of  $\gamma_0$  has the intermediate median bias. The GS2SLS-max has the smallest MAD and DQ, and the GS2SLS-op has the intermediate MAD and DQ. The CR of GS2SLS-op is closest to the nominal level, while the CR of GS2SLS-max is significantly lower than the nominal level in many cases. The performance of CGS2SLS-max for Model 2 is similar to that for Model 1. Compared with the GS2SLS-op, the CGS2SLS-op has much larger MAD and DQ in most cases, similar CR, and has smaller median bias in more than half of the cases when  $\rho_0 = 0.5$  but larger median bias in most cases when  $\rho_0 = 0.1$ .



**Table 5.** Estimation of Model 2 with  $R_f^2 = 0.02$  and  $n = 98$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.246	0.611	3.434	1.000	0.046	0.774	4.234	1.000	
	GS2SLS-max	0.250	0.082	0.317	0.797	-0.079	0.177	0.661	0.994	
	GS2SLS-op	0.198	0.326	2.034	0.999	0.059	0.472	2.364	1.000	
	CGS2SLS-max	-0.055	0.132	0.730	0.849	0.107	0.299	1.340	0.967	
	CGS2SLS-op	-0.391	0.602	10.313	0.992	0.360	0.766	7.293	1.000	
	0.5	GS2SLS-min	0.160	0.354	2.317	1.000	0.408	0.796	4.989	1.000
	GS2SLS-max	0.155	0.065	0.250	0.910	0.338	0.146	0.576	0.943	
	GS2SLS-op	0.128	0.216	1.228	1.000	0.399	0.393	1.964	1.000	
	CGS2SLS-max	-0.008	0.078	0.354	0.960	0.418	0.233	0.949	0.845	
	CGS2SLS-op	-0.204	0.361	6.722	0.998	0.572	0.531	4.621	1.000	
	0.9	GS2SLS-min	0.050	0.210	1.433	1.000	0.741	0.523	3.243	0.963
	GS2SLS-max	0.063	0.032	0.133	0.968	0.793	0.080	0.316	0.038	
	GS2SLS-op	0.051	0.121	0.699	1.000	0.763	0.238	1.278	0.775	
	CGS2SLS-max	0.030	0.036	0.155	0.993	0.721	0.147	0.609	0.193	
CGS2SLS-op	-0.006	0.148	1.445	1.000	0.714	0.286	2.544	0.800		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.289	0.363	2.264	0.994	0.059	0.829	5.260	1.000	
	GS2SLS-max	0.342	0.061	0.238	0.367	-0.091	0.180	0.712	0.991	
	GS2SLS-op	0.267	0.274	1.645	0.985	0.071	0.523	3.235	1.000	
	CGS2SLS-max	0.063	0.160	1.484	0.698	0.023	0.356	1.665	0.958	
	CGS2SLS-op	-0.167	0.675	9.228	0.966	0.254	0.807	7.433	1.000	
	0.5	GS2SLS-min	0.277	0.342	2.408	0.997	0.303	0.694	4.551	1.000
	GS2SLS-max	0.264	0.052	0.203	0.449	0.330	0.151	0.585	0.934	
	GS2SLS-op	0.226	0.196	1.324	0.986	0.356	0.394	2.001	0.999	
	CGS2SLS-max	0.100	0.073	0.372	0.844	0.356	0.242	1.027	0.853	
	CGS2SLS-op	-0.098	0.362	7.297	0.989	0.475	0.584	5.336	1.000	
	0.9	GS2SLS-min	0.182	0.181	1.172	0.992	0.689	0.470	2.978	0.962
	GS2SLS-max	0.184	0.027	0.105	0.240	0.777	0.073	0.285	0.024	
	GS2SLS-op	0.179	0.109	0.682	0.969	0.762	0.220	1.184	0.779	
	CGS2SLS-max	0.144	0.030	0.130	0.568	0.710	0.137	0.559	0.183	
CGS2SLS-op	0.099	0.146	1.737	0.972	0.700	0.299	2.223	0.812		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

**Table 6.** Estimation of Model 2 with  $R_f^2 = 0.1$  and  $n = 98$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.199	0.439	2.673	1.000	-0.001	0.482	2.470	1.000	
	GS2SLS-max	0.230	0.076	0.295	0.804	-0.064	0.151	0.573	0.996	
	GS2SLS-op	0.190	0.290	1.703	0.999	0.017	0.364	1.702	1.000	
	CGS2SLS-max	-0.039	0.115	0.562	0.892	0.069	0.209	0.839	0.983	
	CGS2SLS-op	-0.285	0.461	6.720	0.994	0.206	0.479	3.688	1.000	
	0.5	GS2SLS-min	0.002	0.385	2.408	1.000	0.198	0.669	4.025	1.000
	GS2SLS-max	0.137	0.068	0.266	0.907	0.217	0.135	0.531	0.963	
	GS2SLS-op	0.058	0.217	1.323	1.000	0.198	0.337	1.710	1.000	
	CGS2SLS-max	-0.003	0.076	0.335	0.964	0.177	0.178	0.709	0.942	
	CGS2SLS-op	-0.173	0.302	4.685	0.999	0.263	0.364	2.475	0.999	
	0.9	GS2SLS-min	0.058	0.364	2.209	0.999	0.260	0.504	3.843	0.992
	GS2SLS-max	0.102	0.042	0.170	0.887	0.522	0.085	0.333	0.311	
	GS2SLS-op	0.103	0.231	1.521	0.999	0.369	0.282	2.034	0.958	
	CGS2SLS-max	0.039	0.053	0.220	0.982	0.331	0.134	0.528	0.728	
CGS2SLS-op	-0.023	0.197	2.793	1.000	0.339	0.268	1.738	0.955		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.290	0.364	2.446	0.998	0.053	0.601	3.690	1.000	
	GS2SLS-max	0.319	0.068	0.265	0.454	-0.064	0.162	0.632	0.989	
	GS2SLS-op	0.252	0.278	1.901	0.991	0.056	0.443	2.357	1.000	
	CGS2SLS-max	0.063	0.120	1.088	0.777	0.016	0.251	1.169	0.966	
	CGS2SLS-op	-0.149	0.495	7.195	0.970	0.220	0.642	5.969	1.000	
	0.5	GS2SLS-min	0.244	0.309	1.949	0.997	0.329	0.728	4.353	1.000
	GS2SLS-max	0.268	0.051	0.203	0.440	0.233	0.129	0.507	0.961	
	GS2SLS-op	0.243	0.214	1.317	0.986	0.222	0.366	1.924	1.000	
	CGS2SLS-max	0.091	0.082	0.445	0.825	0.213	0.182	0.780	0.944	
	CGS2SLS-op	-0.052	0.321	5.613	0.988	0.259	0.395	2.986	1.000	
	0.9	GS2SLS-min	0.163	0.261	1.781	0.984	0.088	0.387	2.616	0.991
	GS2SLS-max	0.196	0.038	0.150	0.307	0.487	0.086	0.330	0.371	
	GS2SLS-op	0.149	0.184	1.207	0.970	0.290	0.247	1.503	0.965	
	CGS2SLS-max	0.117	0.050	0.220	0.774	0.291	0.141	0.556	0.787	
CGS2SLS-op	0.049	0.195	3.284	0.978	0.279	0.270	1.742	0.970		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

**Table 7.** Estimation of Model 2 with  $R_f^2 = 0.02$  and  $n = 490$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.193	0.416	2.823	1.000	0.070	0.704	4.351	1.000	
	GS2SLS-max	0.242	0.037	0.145	0.160	-0.100	0.093	0.351	0.967	
	GS2SLS-op	0.169	0.218	1.220	0.998	0.014	0.347	1.673	1.000	
	CGS2SLS-max	-0.017	0.056	0.273	0.893	0.040	0.144	0.580	0.919	
	CGS2SLS-op	-0.348	0.468	11.234	0.992	0.321	0.517	5.647	1.000	
	0.5	GS2SLS-min	0.130	0.354	2.272	1.000	0.259	0.652	3.957	1.000
	GS2SLS-max	0.154	0.031	0.120	0.389	0.316	0.078	0.297	0.557	
	GS2SLS-op	0.104	0.171	0.980	1.000	0.349	0.288	1.408	0.999	
	CGS2SLS-max	0.015	0.033	0.136	0.977	0.303	0.132	0.502	0.581	
	CGS2SLS-op	-0.153	0.293	3.525	1.000	0.405	0.384	2.560	0.999	
	0.9	GS2SLS-min	0.100	0.263	1.769	1.000	0.412	0.541	4.162	0.986
	GS2SLS-max	0.070	0.015	0.059	0.472	0.748	0.041	0.159	0.000	
	GS2SLS-op	0.086	0.155	0.995	1.000	0.546	0.263	1.447	0.860	
	CGS2SLS-max	0.041	0.020	0.074	0.925	0.538	0.083	0.338	0.051	
CGS2SLS-op	-0.008	0.169	2.335	1.000	0.490	0.247	1.964	0.863		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.322	0.398	2.574	0.997	-0.005	0.723	3.984	1.000	
	GS2SLS-max	0.338	0.029	0.110	0.002	-0.115	0.099	0.382	0.940	
	GS2SLS-op	0.271	0.243	1.508	0.976	-0.008	0.407	2.041	1.000	
	CGS2SLS-max	0.060	0.082	0.657	0.634	0.014	0.189	0.862	0.855	
	CGS2SLS-op	-0.300	0.587	11.233	0.939	0.252	0.675	6.906	1.000	
	0.5	GS2SLS-min	0.251	0.281	1.692	0.997	0.291	0.661	3.651	1.000
	GS2SLS-max	0.263	0.025	0.096	0.004	0.306	0.082	0.307	0.553	
	GS2SLS-op	0.239	0.172	1.055	0.971	0.337	0.295	1.385	0.998	
	CGS2SLS-max	0.104	0.038	0.181	0.576	0.302	0.140	0.554	0.580	
	CGS2SLS-op	-0.086	0.316	6.578	0.984	0.375	0.373	2.944	0.999	
	0.9	GS2SLS-min	0.252	0.236	1.595	0.991	0.240	0.400	3.186	0.988
	GS2SLS-max	0.184	0.012	0.046	0.000	0.754	0.037	0.142	0.000	
	GS2SLS-op	0.212	0.134	0.943	0.961	0.534	0.256	1.511	0.831	
	CGS2SLS-max	0.142	0.015	0.059	0.035	0.584	0.082	0.320	0.026	
CGS2SLS-op	0.098	0.156	2.048	0.958	0.503	0.263	1.808	0.823		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

**Table 8.** Estimation of Model 2 with  $R_f^2 = 0.1$  and  $n = 490$ .

		$\lambda_0 = 0.6$				$\gamma_0 = 1.0$				
		MB	MAD	DQ	CR	MB	MAD	DQ	CR	
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.1$									
	GS2SLS-min	0.138	0.318	1.886	1.000	0.019	0.342	1.584	1.000	
	GS2SLS-max	0.215	0.038	0.147	0.246	-0.073	0.075	0.282	0.983	
	GS2SLS-op	0.123	0.241	1.339	1.000	0.000	0.295	1.242	1.000	
	CGS2SLS-max	0.008	0.044	0.182	0.956	0.002	0.099	0.368	0.982	
	CGS2SLS-op	-0.261	0.416	7.521	0.998	0.128	0.393	2.719	1.000	
	0.5	GS2SLS-min	0.143	0.279	1.654	1.000	0.037	0.322	1.661	1.000
	GS2SLS-max	0.164	0.032	0.121	0.286	0.201	0.072	0.269	0.784	
	GS2SLS-op	0.094	0.223	1.218	0.999	0.120	0.273	1.162	1.000	
	CGS2SLS-max	0.028	0.035	0.138	0.970	0.108	0.091	0.343	0.912	
	CGS2SLS-op	-0.181	0.357	7.336	1.000	0.129	0.384	2.341	1.000	
	0.9	GS2SLS-min	0.193	0.206	1.298	1.000	0.056	0.316	1.821	0.997
	GS2SLS-max	0.118	0.019	0.075	0.117	0.476	0.045	0.182	0.005	
	GS2SLS-op	0.117	0.185	1.127	0.997	0.236	0.196	1.018	0.981	
	CGS2SLS-max	0.059	0.027	0.106	0.854	0.200	0.071	0.285	0.703	
CGS2SLS-op	-0.073	0.284	5.465	0.998	0.069	0.275	1.452	0.994		
$\sigma_{v\epsilon} = 0.1$	$\rho_0 = 0.5$									
	GS2SLS-min	0.236	0.269	1.706	0.992	0.026	0.275	1.236	1.000	
	GS2SLS-max	0.319	0.030	0.113	0.008	-0.057	0.071	0.275	0.982	
	GS2SLS-op	0.224	0.220	1.365	0.988	0.030	0.237	1.036	1.000	
	CGS2SLS-max	0.078	0.054	0.306	0.724	0.009	0.102	0.402	0.967	
	CGS2SLS-op	-0.160	0.389	6.310	0.962	0.065	0.298	1.734	1.000	
	0.5	GS2SLS-min	0.255	0.245	1.559	0.993	0.082	0.374	1.865	1.000
	GS2SLS-max	0.267	0.026	0.098	0.005	0.215	0.073	0.267	0.740	
	GS2SLS-op	0.202	0.216	1.233	0.987	0.139	0.253	1.200	1.000	
	CGS2SLS-max	0.109	0.037	0.162	0.566	0.136	0.097	0.385	0.882	
	CGS2SLS-op	-0.112	0.372	7.572	0.988	0.091	0.406	2.593	1.000	
	0.9	GS2SLS-min	0.250	0.200	1.235	0.986	0.060	0.271	1.456	0.996
	GS2SLS-max	0.211	0.015	0.059	0.000	0.492	0.042	0.158	0.001	
	GS2SLS-op	0.186	0.160	0.987	0.973	0.247	0.172	0.857	0.978	
	CGS2SLS-max	0.142	0.022	0.089	0.164	0.211	0.076	0.299	0.667	
CGS2SLS-op	0.022	0.249	4.416	0.985	0.078	0.266	1.447	0.993		

MB: median bias; MAD: median of the absolute deviations; DQ: difference between the 0.1 and 0.9 quantiles; CR: coverage rate of a nominal 95% confidence interval.

From the Monte Carlo results of both models, we can see that the proposed CGS2SLS estimator can effectively reduce the many instrument bias, and the estimators derived by choosing the number of instruments to minimize their respective approximated MSEs, GS2SLS-op and CGS2SLS-op, have coverage rates closer to the nominal level than the estimators using very few or many instruments, *i.e.*, GS2SLS-op and CGS2SLS-op can make inference more reliable. Between GS2SLS-op and CGS2SLS-op, no one is always better than the other in terms of central tendency or coverage rate, but the GS2SLS-op has much smaller dispersion in most cases.

The summary statistics of the estimated  $p$  and  $q$  are presented in Tables 9 and 10. Consistent with [3], in most cases for both models, only the first spatial lag ( $p = 1$ ) is used. For Model 1, in most cases,  $\hat{q}$  is 1 or 2 with  $n = 98$ , and is larger with  $n = 490$  but is smaller than the maximum number of instruments

$\bar{q} = 10$ . For Model 2,  $\hat{q}$  tends to be larger, which might be due to the fact that the variables in  $X_n$  of Model 2 are equally important but the importance of the variables in  $X_n$  of Model 1 is in decreasing order. For both models,  $\hat{q}$  tends to be larger with a larger  $R_f^2$ .

**Table 9.** The Distributions of  $\hat{p}$  and  $\hat{q}$  in Model 1.

	GS2SLS								CGS2SLS								
	$\hat{p}$				$\hat{q}$				$\hat{p}$				$\hat{q}$				
	MO	LQ	ME	UQ	MO	LQ	ME	UQ	MO	LQ	ME	UQ	MO	LQ	ME	UQ	
$n = 98$																	
$R_f^2 = 0.02, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	4	1	1	1	5	1	1	2	4	1	1	2	5
	0.5	1	1	1	4	1	1	2	5	1	1	1	4	1	1	2	5
	0.9	1	1	1	4	1	1	1	5	1	1	1	4	1	1	1	5
$R_f^2 = 0.02, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	4	1	1	1	4	4	1	3	4	5	1	3	4
	0.5	1	1	1	4	1	1	1	5	1	1	1	4	1	1	2	5
	0.9	1	1	1	4	1	1	1	5	1	1	1	4	1	1	1	5
$R_f^2 = 0.1, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	4	1	1	1	5	1	1	2	4	1	1	3	5
	0.5	1	1	1	4	1	1	2	4	1	1	2	4	2	1	2	4
	0.9	1	1	1	3	1	1	1	3	1	1	1	3	1	1	2	3
$R_f^2 = 0.1, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	4	1	1	1	4	1	1	2	4	5	1	4	4
	0.5	1	1	1	3	1	1	2	4	1	1	2	3	1	1	2	4
	0.9	1	1	1	2	1	1	1	2	1	1	1	2	1	1	2	2
$n = 490$																	
$R_f^2 = 0.02, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	2	9	1	1	3	9	1	1	2	9	1	1	3	9
	0.5	1	1	2	9	2	1	3	9	1	1	1	9	2	1	3	9
	0.9	1	1	1	3	1	1	2	3	1	1	1	3	1	1	2	3
$R_f^2 = 0.02, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	6	1	1	2	7	1	1	2	6	1	1	3	7
	0.5	1	1	1	7	1	1	2	8	1	1	1	7	1	1	2	8
	0.9	1	1	1	3	1	1	2	3	1	1	1	3	1	1	2	3
$R_f^2 = 0.1, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	7	3	2	4	8	1	1	2	7	3	2	4	8
	0.5	1	1	1	2	3	2	3	5	1	1	1	2	3	2	4	5
	0.9	1	1	1	1	2	2	2	3	1	1	1	1	3	2	4	3
$R_f^2 = 0.1, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	3	3	1	3	5	1	1	1	3	3	2	4	5
	0.5	1	1	1	2	3	2	3	4	1	1	1	2	3	2	4	4
	0.9	1	1	1	1	2	2	2	3	1	1	1	1	3	2	3	3

MO: mode; LQ: 0.1 quantile; ME: median; UQ: 0.9 quantile.

**Table 10.** The Distributions of  $\hat{p}$  and  $\hat{q}$  in Model 2.

	GS2SLS								CGS2SLS								
	$\hat{p}$				$\hat{q}$				$\hat{p}$				$\hat{q}$				
	MO	LQ	ME	UQ	MO	LQ	ME	UQ	MO	LQ	ME	UQ	MO	LQ	ME	UQ	
<i>n</i> = 98																	
$R_f^2 = 0.02, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	4	1	1	1	5	1	1	2	4	5	1	3	5
	0.5	1	1	1	4	1	1	2	5	1	1	1	4	1	1	2	5
	0.9	1	1	1	4	1	1	2	5	1	1	1	4	1	1	1	5
$R_f^2 = 0.02, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	4	1	1	1	5	4	1	2	4	5	1	3	5
	0.5	1	1	1	4	1	1	2	5	1	1	1	4	1	1	2	5
	0.9	1	1	1	4	1	1	1	5	1	1	1	4	1	1	1	5
$R_f^2 = 0.1, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	4	1	1	2	5	1	1	2	4	5	1	4	5
	0.5	1	1	1	3	1	1	3	5	1	1	1	3	5	1	4	5
	0.9	1	1	1	2	1	1	1	3	1	1	1	2	5	1	4	3
$R_f^2 = 0.1, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	3	1	1	2	5	1	1	2	3	5	2	5	5
	0.5	1	1	1	3	1	1	2	5	1	1	1	3	5	1	4	5
	0.9	1	1	1	2	1	1	1	3	1	1	1	2	5	1	4	3
<i>n</i> = 490																	
$R_f^2 = 0.02, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	2	9	1	1	4	10	1	1	2	9	1	1	4	10
	0.5	1	1	2	9	1	1	4	10	1	1	1	9	1	1	4	10
	0.9	1	1	1	3	1	1	2	4	1	1	1	3	1	1	4	4
$R_f^2 = 0.02, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	6	1	1	2	9	1	1	2	6	10	1	5	9
	0.5	1	1	1	8	1	1	3	10	1	1	1	8	1	1	4	10
	0.9	1	1	1	3.5	1	1	1	4	1	1	1	3.5	1	1	2	4
$R_f^2 = 0.1, \rho_0 = 0.1, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	5	10	3	9	10	1	1	1	5	10	8	10	10
	0.5	1	1	1	1	10	2	6	10	1	1	1	1	10	8	10	10
	0.9	1	1	1	1	3	1	2	5	1	1	1	1	10	8	10	5
$R_f^2 = 0.1, \rho_0 = 0.5, \sigma_{v\epsilon} = 0.1$	0.1	1	1	1	1	10	1	7	10	1	1	1	1	10	8	10	10
	0.5	1	1	1	1	4	2	4	10	1	1	1	1	10	8	10	10
	0.9	1	1	1	1	3	1	2	4	1	1	1	1	10	8	10	4

MO: mode; LQ: 0.1 quantile; ME: median; UQ: 0.9 quantile.

### 5. Conclusions

In this paper, we derive an approximated MSE of the GS2SLS estimator and a bias corrected GS2SLS (CGS2SLS) estimator for the SARAR model in the presence of endogenous variables and many instruments. We propose a instrument selection procedure by minimizing the approximated MSEs. Our Monte Carlo experiments show that the CGS2SLS can effectively correct the many instrument bias and the instrument selection procedure generally makes inference in finite samples more accurate.

### Acknowledgements

We are grateful to the editor and one anonymous referee for helpful comments that have improved the presentation of the paper.

**Appendix**

**A. Notations**

$A^s = A + A'$  for a square matrix  $A$ .

$\|A\| = \sqrt{\text{tr}(A'A)}$  is the Frobenius matrix norm for a matrix  $A$ .

$\text{vec}_D(A)$  is a column vector whose elements are the diagonal elements of a square matrix  $A$ .

$R_n = R_n(\rho_0)$  and  $G_n = G_n(\lambda_0)$ , where  $R_n(\rho) = I_n - \rho M_n$  and  $G_n(\lambda) = W_n(I_n - \lambda W_n)^{-1}$ .

$y_n(\rho) = R_n(\rho)y_n$ ,  $Z_{2n}(\rho) = R_n(\rho)Z_{2n}$ ,  $Z_n(\rho) = R_n(\rho)Z_n$  and  $u_n(\rho) = R_n(\rho)u_n$ .

$Z_{2n} = \bar{Z}_{2n} + v_n$ , where  $\bar{Z}_{2n} = E(Z_{2n})$ .

$Z_n = [W_n y_n, Z_{2n}] = \bar{Z}_n + \zeta_n$ , where  $\bar{Z}_n = E(Z_n) = [G_n \bar{Z}_{2n} \gamma_0, \bar{Z}_{2n}]$  and  $\zeta_n = [G_n v_n \gamma_0 + G_n R_n^{-1} \epsilon_n, v_n]$ .

$P_{K,n} = Q_{K,n}(Q'_{K,n}Q_{K,n})^{-}Q'_{K,n}$ , where  $(Q'_{K,n}Q_{K,n})^{-}$  is a generalized inverse of  $Q'_{K,n}Q_{K,n}$ .

$\Gamma_{nK,1} = P_{K,n}R_n$ ,  $\Gamma_{nK,2} = P_{K,n}R_nG_n$  and  $\Gamma_{nK,3} = P_{K,n}R_nG_nR_n^{-1}$ .

$\Delta_{nK,1} = \frac{1}{n} \text{tr}[\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)]$  and  $\Delta_{nK,2} = \frac{1}{n} \text{tr}[\bar{Z}'_n M'_n(I_n - P_{K,n})M_n \bar{Z}_n]$ .

$h_n = \frac{1}{\sqrt{n}} \bar{Z}'_n(\rho_0)\epsilon_n$  and  $H_n = \frac{1}{n} \bar{Z}'_n(\rho_0)\bar{Z}_n(\rho_0)$ .

For the GS2SLS,

$$S_n(K) = \frac{1}{n} H_n^{-1} [\sigma_\epsilon^2 \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) + \Omega_{n2}(K)] H_n^{-1}$$

where

$$\Omega_{n2}(K) = \Upsilon_n(K)\Upsilon'_n(K)$$

with

$$\Upsilon_n(K) = E(\zeta'_n R'_n P_{K,n} \epsilon_n) = [\text{tr}(\Gamma_{nK,2})\sigma_{v\epsilon}\gamma_0 + \sigma_\epsilon^2 \text{tr}(\Gamma_{nK,3}), \text{tr}(\Gamma_{nK,1})\sigma_{v\epsilon}]'$$

For the CGS2SLS,

$$S_n(K) = \frac{1}{n} H_n^{-1} [\sigma_\epsilon^2 \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) + \Pi_{n1}(K) + \Pi_{n2}(K) + \Pi_{n3}(K)] H_n^{-1}$$

where  $\Pi_{n1}(K)$  is a symmetric matrix equal to

$$\Pi_{n1}(K) = \begin{pmatrix} \gamma'_0 \sigma'_{v\epsilon} \sigma_{v\epsilon} \gamma_0 \text{tr}(\Gamma_{nK,2}^2) + \sigma_\epsilon^2 \gamma'_0 \Sigma_v \gamma_0 \text{tr}(\Gamma_{nK,2} \Gamma'_{nK,2}) + \sigma_\epsilon^4 \text{tr}(\Gamma_{nK,3} \Gamma_{nK,3}^s) + 2\sigma_\epsilon^2 \sigma_{v\epsilon} \gamma_0 \text{tr}(\Gamma_{nK,2} \Gamma_{nK,3}^s) & * \\ \sigma'_{v\epsilon} \sigma_{v\epsilon} \gamma_0 \text{tr}(\Gamma_{nK,1} \Gamma_{nK,2}) + \sigma_\epsilon^2 \Sigma_v \gamma_0 \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,2}) + \sigma_\epsilon^2 \sigma'_{v\epsilon} \text{tr}(\Gamma_{nK,3} \Gamma_{nK,1}^s) & * \end{pmatrix} \tag{21}$$

with the (2, 2)th block being  $\sigma'_{v\epsilon} \sigma_{v\epsilon} \text{tr}(\Gamma_{nK,1}^2) + \sigma_\epsilon^2 \Sigma_v \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,1})$ ,

$$\Pi_{n2}(K) = -[\Pi_{n2,1}(K), \Pi_{n2,2}(K)]^s - 2\sigma_\epsilon^2 \Omega_{n1}(K) \tag{22}$$

where

$$\begin{aligned} \Omega_{n1}(K) &= E(\zeta'_n R'_n P_{K,n} R_n \zeta_n) \\ &= \begin{pmatrix} \gamma'_0 \Sigma_v \gamma_0 \text{tr}(\Gamma'_{nK,2} \Gamma_{nK,2}) + \sigma_\epsilon^2 \text{tr}(\Gamma'_{nK,3} \Gamma_{nK,3}) + 2\sigma_{v\epsilon} \gamma_0 \text{tr}(\Gamma'_{nK,3} \Gamma_{nK,2}) & * \\ \Sigma_v \gamma_0 \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,2}) + \sigma'_{v\epsilon} \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,3}) & \Sigma_v \text{tr}(\Gamma'_{nK,1} \Gamma_{nK,1}) \end{pmatrix} \\ \Pi_{n2,1}(K) &= V_{1n}[\sigma_\epsilon^2 \text{tr}(\Gamma_{nK,3} M_n R_n^{-1}) - \sigma_\epsilon^2 \text{tr}(P_{K,n} M_n G_n R_n^{-1}) - \text{tr}(P_{K,n} M_n G_n) \sigma_{v\epsilon} \gamma_0] \\ &\quad + V_{2n}[\text{tr}(\Gamma_{nK,2} G_n) \sigma_{v\epsilon} \gamma_0 + \sigma_\epsilon^2 \text{tr}(\Gamma_{nK,2} G_n R_n^{-1}), \sigma_{v\epsilon} \text{tr}(\Gamma_{nK,2})]' \\ &\quad + V_{3n} \gamma_0 \text{tr}(\Gamma_{nK,2}) + V_{4n} \text{tr}(\Gamma_{nK,3}) \end{aligned} \tag{23}$$

and

$$\Pi_{n2,2}(K) = -V_{1n} \sigma_{v\epsilon} \text{tr}(P_{K,n} M_n) + V_{3n} \text{tr}(\Gamma_{nK,1}) \tag{24}$$

with  $V_{1n} = \frac{\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0) \mathbb{F}'_n$   $V_{2n} = \sigma_\epsilon^2 I_{m+1}$

$$V_{3n} = \frac{\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0) \bar{Z}_{2n} - \frac{\sigma_\epsilon^2}{n^2} \bar{Z}'_n(\rho_0) \mathbb{F}'_n \sigma_{v\epsilon} \text{tr}(M_n R_n^{-1}) - \frac{\sigma_\epsilon^2}{n} (\bar{Z}'_n R'_n \bar{Z}_{2n} + [\Sigma_v \gamma_0 \text{tr}(R_n G_n) + \sigma'_{v\epsilon} \text{tr}(G_n), \Sigma_v \text{tr}(R_n)]')$$

and

$$V_{4n} = -\frac{2\sigma_\epsilon^4}{n^2} \bar{Z}'_n(\rho_0) \mathbb{F}'_n \text{tr}(M_n R_n^{-1}) - \frac{2\sigma_\epsilon^2}{n} [\text{tr}(R_n G_n) \sigma_{v\epsilon} \gamma_0 + \sigma_\epsilon^2 \text{tr}(G_n), \sigma_{v\epsilon} \text{tr}(R_n)]'$$

and

$$\Pi_{n3}(K) = -\frac{\sigma_\epsilon^2}{n^2} \{ \bar{Z}'_n(\rho_0) \mathbb{F}'_n [E(u'_n M'_n P_{K,n} R_n \zeta_n) + E(\epsilon'_n P_{K,n} M_n \zeta_n)] \}^s \tag{25}$$

with

$$E(u'_n M'_n P_{K,n} R_n \zeta_n) = [\sigma_{v\epsilon} \gamma_0 \text{tr}(R_n^{-1} M'_n \Gamma_{nK,2}) + \sigma_\epsilon^2 \text{tr}(R_n^{-1} M'_n \Gamma_{nK,3}), \sigma_{v\epsilon} \text{tr}(R_n^{-1} M'_n \Gamma_{nK,1})]$$

and

$$E(\epsilon'_n P_{K,n} M_n \zeta_n) = [\sigma_{v\epsilon} \gamma_0 \text{tr}(P_{K,n} M_n G_n) + \sigma_\epsilon^2 \text{tr}(P_{K,n} M_n G_n R_n^{-1}), \sigma_{v\epsilon} \text{tr}(P_{K,n} M_n)]$$

Let  $\hat{V}_{1n} = \frac{\hat{\sigma}_\epsilon^2}{n} Z'_n(\hat{\rho}_n) \hat{\mathbb{F}}'_n$ ,  $\hat{V}_{2n} = \hat{\sigma}_\epsilon^2 I_{m+1}$ ,

$$\hat{V}_{3n} = \frac{\hat{\sigma}_\epsilon^2}{n} [Z'_n(\hat{\rho}_n) Z_{2n} - \hat{E}(\zeta'_n R'_n v_n)] - \frac{\hat{\sigma}_\epsilon^2}{n^2} Z'_n(\hat{\rho}_n) \hat{\mathbb{F}}'_n \hat{\sigma}_{v\epsilon} \text{tr}(M_n \hat{R}_n^{-1}) - \frac{\hat{\sigma}_\epsilon^2}{n} Z'_n \hat{R}'_n Z_{2n}$$

and

$$\hat{V}_{4n} = -\frac{2\hat{\sigma}_\epsilon^4}{n^2} Z'_n(\hat{\rho}_n) \hat{\mathbb{F}}'_n \text{tr}(M_n \hat{R}_n^{-1}) - \frac{2\hat{\sigma}_\epsilon^2}{n} [\text{tr}(\hat{R}_n \hat{G}_n) \hat{\sigma}_{v\epsilon} \hat{\gamma}_n + \hat{\sigma}_\epsilon^2 \text{tr}(\hat{G}_n), \hat{\sigma}_{v\epsilon} \text{tr}(\hat{R}_n)]'$$

where  $\hat{R}_n = R_n(\hat{\rho}_n)$ ,  $\hat{G}_n = G_n(\hat{\rho}_n)$ ,  $\hat{E}(\zeta'_n R'_n v_n) = [\hat{\Sigma}_v \hat{\gamma}_n \text{tr}(\hat{R}_n \hat{G}_n) + \hat{\sigma}'_{v\epsilon} \text{tr}(\hat{G}_n), \hat{\Sigma}_v \text{tr}(\hat{R}_n)]'$  and  $\hat{\mathbb{F}}_n$  is an estimator of  $\mathbb{F}_n$  in (7) derived by replacing  $\bar{Z}_n$  by  $Z_n$  and true parameters by their estimators. An estimation for  $\Pi_{n2}(K)$  is

$$\hat{\Pi}_{n2}(K) = -[\hat{\Pi}_{n2,1}(K), \hat{\Pi}_{n2,2}(K)]^s - 2\hat{\sigma}_\epsilon^2 \hat{\Omega}_{n1}(K) \tag{26}$$

where  $\hat{\Pi}_{n2,1}(K)$ ,  $\hat{\Pi}_{n2,2}(K)$  and  $\hat{\Omega}_{n1}(K)$  are derived respectively from  $\Pi_{n2,1}(K)$ ,  $\Pi_{n2,2}(K)$  and  $\Omega_{n1}(K)$  by replacing  $V_{jn}$ 's by  $\hat{V}_{jn}$ 's and the rest of involved parameters by their respective estimators.



An estimator for  $\Pi_{n3}(K)$  is

$$\hat{\Pi}_{n3}(K) = -\frac{\hat{\sigma}_\epsilon^2}{n^2} \{Z'_n(\hat{\rho}_n)\hat{F}'_n[\hat{E}(u'_n M'_n P_{K,n} R_n \zeta_n) + \hat{E}(\epsilon'_n P_{K,n} M_n \zeta_n)]\}^s \tag{27}$$

where  $\hat{E}(\zeta'_n R'_n P_{K,n} M_n u_n)$  and  $\hat{E}(\zeta'_n M'_n P_{K,n} \epsilon_n)$  are derived by replacing the parameters in, respectively,  $E(\zeta'_n R'_n P_{K,n} M_n u_n)$  and  $E(\zeta'_n M'_n P_{K,n} \epsilon_n)$  by their estimators.

**B. Lemmas**

The following lemma gives sufficient conditions under which  $\bar{Z}_n(\rho)$  can be approximated arbitrarily well by a linear combination of  $Q_{K,n}$  as  $n, K \rightarrow \infty$ . When the approximation of  $\bar{Z}_n(\rho)$  becomes better as the number of instruments  $K$  increases, the variance part of the MSE becomes smaller.

**Lemma 1.** *Suppose that  $\sup_n \|\lambda_0 W_n\|_\infty < 1$ , elements of  $\psi_{q,n}$  are uniformly bounded constants, and there exists  $\pi_q^0$  such that  $\|\bar{Z}_{2n} - \psi_{q,n} \pi_q^0\|_\infty \rightarrow 0$  as  $n, q \rightarrow \infty$ . Then, for  $F_{K,n} = [\psi_{q,n}, W_n \psi_{q,n}, \dots, W_n^p \psi_{q,n}]$ , where  $p, q \rightarrow \infty$  as  $n \rightarrow \infty$ ,*

- (i) *there exists  $\pi_{K,n}$  such that  $\frac{1}{n} \|\bar{Z}_n - F_{K,n} \pi_{K,n}\|^2 \rightarrow 0$  as  $n, K \rightarrow \infty$ ,*
- (ii)  *$\frac{1}{n} \|\bar{Z}_n(\rho) - Q_{K,n} \pi_{K,n,1}\|^2 \leq \frac{c}{n} \|\bar{Z}_n - F_{K,n} \pi_{K,n}\|^2$  for some  $c > 0$ , where  $\pi_{K,n,1} = [\pi'_{K,n}, -\rho \pi'_{K,n}]'$ , which implies that  $\frac{1}{n} \|\bar{Z}_n(\rho) - Q_{K,n} \pi_{K,n,1}\|^2 \rightarrow 0$  as  $n, K \rightarrow \infty$ .*

*Proof.* (i) is Lemma 2.1 in [3]. The argument is as follows. Let

$$\pi_{K,n} = \begin{pmatrix} 0 & (\pi_q^0 \gamma_0)' & (\lambda_0 \pi_q^0 \gamma_0)' & \dots & (\lambda_0^{p-1} \pi_q^0 \gamma_0)' \\ \pi_q^0 & 0 & 0 & \dots & 0 \end{pmatrix}'$$

Then,  $F_{K,n} \pi_{K,n} = [W_n \sum_{j=0}^{p-1} \lambda_0^j W_n^j \psi_{q,n} \pi_q^0 \gamma_0, \psi_{q,n} \pi_q^0] = [(I_n - \lambda_0^p W_n^p) G_n \psi_{q,n} \pi_q^0 \gamma_0, \psi_{q,n} \pi_q^0]$  and

$$\bar{Z}_n - F_{K,n} \pi_{K,n} = [\lambda_0^p W_n^p G_n \bar{Z}_{2n} \gamma_0 + (I_n - \lambda_0^p W_n^p) G_n (\bar{Z}_{2n} - \psi_{q,n} \pi_q^0) \gamma_0, \bar{Z}_{2n} - \psi_{q,n} \pi_q^0]$$

Thus

$$\begin{aligned} & \|\bar{Z}_n - F_{K,n} \pi_{K,n}\|_\infty \\ & \leq \|\lambda_0 W_n\|_\infty^p \|G_n\|_\infty \|\bar{Z}_{2n} \gamma_0\|_\infty + (1 + \|\lambda_0 W_n\|_\infty^p) \|G_n\|_\infty \|\bar{Z}_{2n} - \psi_{q,n} \pi_q^0\|_\infty \|\gamma_0\|_\infty + \|\bar{Z}_{2n} - \psi_{q,n} \pi_q^0\|_\infty \\ & \rightarrow 0, \end{aligned}$$

as  $n, p, q \rightarrow \infty$ . Since  $\frac{1}{n} \|\bar{Z}_{2n} - \psi_{q,n} \pi_q^0\|^2 \leq (\|\bar{Z}_{2n} - \psi_{q,n} \pi_q^0\|_\infty)^2$ , the result follows.

(ii) Let  $R'_n(\rho) R_n(\rho) = R'_{1n}(\rho) R_{2n}(\rho) R_{1n}(\rho)$  be an eigenvalue-eigenvector decomposition, where  $R_{2n}(\rho)$  is a diagonal matrix whose diagonal elements are the eigenvalues of  $R'_n(\rho) R_n(\rho)$  and  $R_{1n}(\rho)$  is an orthonormal matrix whose columns are eigenvectors of  $R'_n(\rho) R_n(\rho)$ . Then,

$$\begin{aligned} \frac{1}{n} \|\bar{Z}_n(\rho) - Q_{K,n} \pi_{K,n,1}\|^2 &= \frac{1}{n} \|R_n(\rho) (\bar{Z}_n - F_{K,n} \pi_{K,n})\|^2 \\ &= \frac{1}{n} \text{tr}[(\bar{Z}_n - F_{K,n} \pi_{K,n})' R'_n(\rho) R_n(\rho) (\bar{Z}_n - F_{K,n} \pi_{K,n})] \\ &= \frac{1}{n} \text{tr}[(\bar{Z}_n - F_{K,n} \pi_{K,n})' R'_{1n}(\rho) R_{2n}(\rho) R_{1n}(\rho) (\bar{Z}_n - F_{K,n} \pi_{K,n})] \\ &\leq \frac{1}{n} r_{n,\max} \text{tr}[(\bar{Z}_n - F_{K,n} \pi_{K,n})' (\bar{Z}_n - F_{K,n} \pi_{K,n})] \end{aligned}$$

where  $r_{n,\max}$  is the largest eigenvalue of  $R'_n(\rho)R_n(\rho)$ . By the spectral radius theorem,

$$r_{n,\max} \leq \|R'_n(\rho)R_n(\rho)\|_\infty \leq \|R'_n(\rho)\|_\infty \|R_n(\rho)\|_\infty \leq c$$

for some  $c > 0$  and all  $n$ . Thus (ii) holds. □

The following lemma, Lemma A.1 in [1], gives conditions on the decomposition of an estimator, such that the dominant component of the MSE depending on the number of instruments can be derived.

**Lemma 2.** For an estimator given by  $\sqrt{n}(\hat{\delta}_n - \delta_0) = \hat{H}_n^{-1}\hat{h}_n$ , suppose that there is a decomposition,  $\hat{h}_n = h_n + T_n^h + Z_n^h$ ,  $\hat{H}_n = H_n + T_n^H + Z_n^H$ ,

$$(h_n + T_n^h)(h_n + T_n^h)' - h_n h_n' H_n^{-1} T_n^{H'} - T_n^H H_n^{-1} h_n h_n' = \hat{A}_n(K) + Z_n^A(K)$$

such that

- (i)  $T_n^h = o_P(1)$ ,  $h_n = O_P(1)$ ,  $H_n = O(1)$ ,
- (ii) the determinant of  $H_n$  is bounded away from zero,
- (iii)  $\rho_{K,n} = \text{tr}(S_n(K)) = o(1)$ ,
- (iv)  $\|T_n^H\|^2 = o_P(\rho_{K,n})$ ,  $\|T_n^h\| \|T_n^H\| = o_P(\rho_{K,n})$ ,  $\|Z_n^h\| = o_P(\rho_{K,n})$ ,  $\|Z_n^H\| = o_P(\rho_{K,n})$ ,  $Z_n^A(K) = o_P(\rho_{K,n})$ ,
- (v)  $E[\hat{A}_n(K)] = \sigma_\epsilon^2 H_n + H_n S_n(K) H_n + o(\rho_{K,n})$ .

Then (15) is satisfied.

**Lemma 3.** Let  $A_n = [a_{n,ij}]$  and  $B_n = [b_{n,ij}]$  be  $n \times n$  matrices, then

- (i)  $E(\epsilon'_n A_n v_n) = \sigma_{v\epsilon} \text{tr}(A_n)$ ,
- (ii)  $E(\epsilon'_n A_n \epsilon_n) = \sigma_\epsilon^2 \text{tr}(A_n)$ ,
- (iii)  $E(v'_n A_n v_n) = \Sigma_v \text{tr}(A_n)$ ,
- (iv)  $E(v'_n A_n \epsilon_n \epsilon'_n B_n v_n) = [E(\epsilon_{ni}^2 v'_{ni} v_{ni}) - 2\sigma'_{v\epsilon} \sigma_{v\epsilon} - \sigma_\epsilon^2 \Sigma_v] \text{vec}_D'(A_n) \text{vec}_D(B_n) + \sigma'_{v\epsilon} \sigma_{v\epsilon} [\text{tr}(A_n) \text{tr}(B_n) + \text{tr}(A_n B'_n)] + \sigma_\epsilon^2 \Sigma_v \text{tr}(A_n B_n)$ ,
- (v)  $E(\epsilon'_n A_n \epsilon_n \epsilon'_n B_n v_n) = [E(\epsilon_{ni}^3 v_{ni}) - 3\sigma_\epsilon^2 \sigma_{v\epsilon}] \text{vec}_D'(A_n) \text{vec}_D(B_n) + \sigma_\epsilon^2 \sigma_{v\epsilon} [\text{tr}(A_n) \text{tr}(B_n) + \text{tr}(A_n B_n^s)]$ ,
- (vi)  $E(\epsilon'_n A_n \epsilon_n \epsilon'_n B_n \epsilon_n) = (\mu_4 - 3\sigma_\epsilon^4) \text{vec}_D'(A_n) \text{vec}_D(B_n) + \sigma_\epsilon^4 [\text{tr}(A_n) \text{tr}(B_n) + \text{tr}(A_n B_n^s)]$ ,
- (vii)  $E(v'_n A_n v_n v'_n B_n v_n) = [E(v'_{ni} v_{ni})^2 - \Sigma_v^2 - E(v'_{ni} v_{nj} v'_{nj} v_{ni}) - E(v'_{ni} v_{nj})^2] \text{vec}_D'(A_n) \text{vec}_D(B_n) + \Sigma_v^2 \text{tr}(A_n) \text{tr}(B_n) + E(v'_{ni} v_{nj} v'_{nj} v_{ni}) \text{tr}(A_n B_n) + E(v'_{ni} v_{nj})^2 \text{tr}(A_n B'_n)$ .

*Proof.* For (i)–(iii), we only prove (i), as the other two follow similarly; for (iv)–(vii), we only prove (iv) for the same reason.

For (i),  $\epsilon'_n A_n v_n = \sum_{i=1}^n a_{n,ii} \epsilon_{ni} v_{ni} + \sum_{i=1}^n \sum_{j \neq i} a_{n,ij} \epsilon_{ni} v_{nj}$ . As  $E(\epsilon_{ni} v_{ni}) = \sigma_{v\epsilon}$  and  $E(\epsilon_{ni} v_{nj}) = 0$  for  $i \neq j$ , the result follows. For (iv),

$$E(v'_n A_n \epsilon_n \epsilon'_n B_n v_n) = \sum_{i=1}^n \sum_{j=1}^n \sum_{r=1}^n \sum_{s=1}^n a_{n,ij} b_{n,rs} E(v'_{ni} v_{ns} \epsilon_{nj} \epsilon_{nr})$$

where  $E(v'_{ni}v_{ns}\epsilon_{nj}\epsilon_{nr}) \neq 0$  in one of the following situations:  $i = j = r = s$ ;  $i = j$  and  $r = s$ , but  $i \neq r$ ;  $i = r$  and  $j = s$ , but  $i \neq j$ ;  $i = s$  and  $j = r$ , but  $i \neq j$ . Then

$$\begin{aligned} & E(v'_n A_n \epsilon_n \epsilon'_n B_n v_n) \\ &= \sum_{i=1}^n a_{n,ii} b_{n,ii} E(v'_{ni} v_{ni} \epsilon_{ni}^2) + \sum_{i=1}^n \sum_{j \neq i} E[(a_{n,ii} b_{n,jj} + a_{n,ij} b_{n,ij}) v'_{ni} \epsilon_{ni} v_{nj} \epsilon_{nj} + a_{n,ij} b_{n,ji} v'_{ni} v_{ni} \epsilon_{nj}^2] \\ &= [E(\epsilon_{ni}^2 v'_{ni} v_{ni}) - 2\sigma'_{v\epsilon} \sigma_{v\epsilon} - \sigma_\epsilon^2 \Sigma_v] \text{vec}_D'(A_n) \text{vec}_D(B_n) + \sigma'_{v\epsilon} \sigma_{v\epsilon} [\text{tr}(A_n) \text{tr}(B_n) + \text{tr}(A_n B'_n)] \\ &\quad + \sigma_\epsilon^2 \Sigma_v \text{tr}(A_n B_n) \end{aligned} \quad \square$$

**Lemma 4.** Suppose that  $n \times n$  matrices  $\{A_n\}$  and  $\{B_n\}$  are UB,  $C_n = P_{K,n} A_n = [c_{n,ij}]$  and  $D_n = P_{K,n} B_n = [d_{n,ij}]$ . Then

- (i)  $\text{tr}(P_{K,n}) = K$ ,
- (ii)  $|\text{tr}(C_n)| = O(K)$ ,  $|\text{tr}(C_n^2)| = O(K)$  and  $\sum_{i=1}^n c_{n,ii}^2 = O(K)$ ,
- (iii)  $|\text{tr}(C_n D_n)| = O(K)$  and  $\sum_{i=1}^n c_{n,ii} d_{n,ii} = O(K)$ .

*Proof.* (i) and (ii) are Lemma B.2 in [21]; (iii) By the Cauchy–Schwarz inequality,

$$\text{tr}^2(C_n D_n) \leq \text{tr}(C_n C'_n) \text{tr}(D_n D'_n)$$

and

$$\left(\sum_{i=1}^n c_{n,ii} d_{n,ii}\right)^2 \leq \sum_{i=1}^n c_{n,ii}^2 \sum_{i=1}^n d_{n,ii}^2$$

where  $\text{tr}(C_n C'_n) = \text{tr}(P_{K,n} A_n A'_n)$  and  $\text{tr}(D_n D'_n) = \text{tr}(P_{K,n} B_n B'_n)$ , thus the results follow by (ii).  $\square$

**Lemma 5.** Suppose that  $\{A_n\}$  and  $\{B_n\}$  are  $n \times n$  matrices that are UB and  $C_n = A_n P_{K,n} B_n$ , then

- (i)  $\frac{1}{n} \epsilon'_n A_n \epsilon_n = O_P(1)$ ,  $\frac{1}{n} \epsilon'_n A_n v_n = O_P(1)$ , and  $\frac{1}{n} v'_n A_n v_n = O_P(1)$ ;
- (ii)  $\frac{1}{\sqrt{n}} [\epsilon'_n A_n \epsilon_n - E(\epsilon'_n A_n \epsilon_n)] = O_P(1)$ ,  $\frac{1}{\sqrt{n}} [\epsilon'_n A_n v_n - E(\epsilon'_n A_n v_n)] = O_P(1)$ , and  $\frac{1}{\sqrt{n}} [v'_n A_n v_n - E(v'_n A_n v_n)] = O_P(1)$ ;
- (iii)  $\frac{1}{\sqrt{n}} [\epsilon'_n C_n \epsilon_n - E(\epsilon'_n C_n \epsilon_n)] = O_P(\sqrt{K/n})$ ,  $\frac{1}{\sqrt{n}} [\epsilon'_n C_n v_n - E(\epsilon'_n C_n v_n)] = O_P(\sqrt{K/n})$ , and  $\frac{1}{\sqrt{n}} [v'_n C_n v_n - E(v'_n C_n v_n)] = O_P(\sqrt{K/n})$ .

*Proof.* All the results follow by Chebyshev’s inequality and Lemmas 3–4. We only prove the last result in (iii). Let  $e_i$  be the  $i$ th column of the  $m \times m$  identity matrix. Then the variance of the  $(i, j)$ th element of  $[v'_n C_n v_n - E(v'_n C_n v_n)]$  is  $\frac{1}{n} E\{e'_i [v'_n C_n v_n - E(v'_n C_n v_n)] e_j e'_j [v'_n C_n v_n - E(v'_n C_n v_n)]' e_i\}$ , which is smaller than or equal to

$$\begin{aligned} & \frac{1}{n} E\{e'_i [v'_n C_n v_n - E(v'_n C_n v_n)] [v'_n C_n v_n - E(v'_n C_n v_n)]' e_i\} \\ &= \frac{1}{n} e'_i [E(v'_n C_n v_n v'_n C'_n v_n) - E(v'_n C_n v_n) E(v'_n C'_n v_n)] e_i \\ &= O(K/n) \end{aligned}$$

by Lemmas 3–4. Thus, the  $(i, j)$ th element of  $\frac{1}{\sqrt{n}} [\epsilon'_n C_n v_n - E(\epsilon'_n C_n v_n)]$  is  $O_P(\sqrt{K/n})$  by Chebyshev’s inequality. The result follows as  $i$  and  $j$  are arbitrary.  $\square$

**Lemma 6.** Suppose that  $\{A_n\}$  is a sequence of  $n \times n$  matrices that are bounded in the column sum matrix norm, the elements of the  $n \times k$  matrix  $C_n$  are uniformly bounded, and  $\epsilon_{ni}$ 's in  $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nn})'$  are i.i.d. with zero mean and finite variance  $\sigma_\epsilon^2$ . Then  $\frac{1}{\sqrt{n}}C_n'A_n\epsilon_n = O_P(1)$ .

Furthermore, if the limit of  $\frac{1}{n}C_n'A_nA_n'C_n$  exists and is positive definite, then

$$\frac{1}{\sqrt{n}}C_n'A_n\epsilon_n \xrightarrow{d} N(0, \lim_{n \rightarrow \infty} \frac{\sigma_\epsilon^2}{n}C_n'A_nA_n'C_n)$$

*Proof.* See [22]. □

The following two lemmas show the orders of relevant terms in deriving the decompositions for the GS2SLS and CGS2SLS estimators.

**Lemma 7.** Suppose that  $n \times n$  matrices  $\{A_n\}$  are UB, then

- (i)  $\Delta_{nK,1} = \frac{1}{n} \text{tr}[\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)] = o(1)$ ,
- (ii)  $\Delta_{nK,2} = \frac{1}{n} \text{tr}[\bar{Z}'_nM'_n(I_n - P_{K,n})M_n\bar{Z}_n] = o(1)$ ,
- (iii)  $\frac{1}{n}||\bar{Z}'_nM'_n(I_n - P_{K,n})\bar{Z}_n(\rho_0)|| = O(\sqrt{\Delta_{nK,1}\Delta_{nK,2}})$  and  $\frac{1}{n}\bar{Z}'_nM'_n(I_n - P_{K,n})A_n\bar{Z}_n(\rho_0) = O(\sqrt{\Delta_{nK,2}})$ ,
- (iv)  $\frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_n\epsilon_n = O_P(\Delta_{nK,1}^{1/2})$ ,  $\frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_nv_n = O_P(\Delta_{nK,1}^{1/2})$ ,  $\frac{1}{\sqrt{n}}\bar{Z}'_nM'_n(I_n - P_{K,n})A_n\epsilon_n = O_P(\Delta_{nK,2}^{1/2})$  and  $\frac{1}{\sqrt{n}}\bar{Z}'_nM'_n(I_n - P_{K,n})A_nv_n = O_P(\Delta_{nK,2}^{1/2})$ ,
- (v)  $\frac{1}{n} \text{tr}[\bar{Z}'_nM'_n(I_n - P_{K,n})A_nA'_n(I_n - P_{K,n})M_n\bar{Z}_n] = O(\Delta_{nK,2})$ ,
- (vi)  $\sqrt{\Delta_{nK,1}/n} = o(K/n + \Delta_{nK,1})$ ,  $\sqrt{\Delta_{nK,2}/n} = o(K/n + \Delta_{nK,2})$ ,  $\sqrt{K\Delta_{nK,1}/n} = o(K^2/n + \Delta_{nK,1})$  and  $\sqrt{K\Delta_{nK,2}/n} = o(K^2/n + \Delta_{nK,2})$ .

*Proof.* (i) By Assumption 6, there exists  $\pi_{K,n}$  such that  $\frac{1}{n}||\bar{Z}_n(\rho_0) - Q_{K,n}\pi_{K,n}||^2 \rightarrow 0$  as  $n, K \rightarrow \infty$ . Then

$$\begin{aligned} \Delta_{nK,1} &= \frac{1}{n} \text{tr}[(\bar{Z}_n(\rho_0) - Q_{K,n}\pi_{K,n})'(I_n - P_{K,n})(\bar{Z}_n(\rho_0) - Q_{K,n}\pi_{K,n})] \\ &\leq \frac{1}{n} \text{tr}[(\bar{Z}_n(\rho_0) - Q_{K,n}\pi_{K,n})'(\bar{Z}_n(\rho_0) - Q_{K,n}\pi_{K,n})] \\ &= \frac{1}{n} ||\bar{Z}_n(\rho_0) - Q_{K,n}\pi_{K,n}||^2 \\ &= o(1) \end{aligned}$$

(ii) As  $\rho M_n\bar{Z}_n = \bar{Z}_n(0) - \bar{Z}_n(\rho)$ , there exist  $\pi_{nK,1}$  such that  $\frac{1}{n}||M_n\bar{Z}_n - Q_{K,n}\pi_{nK,1}||^2 \rightarrow 0$  as  $n, K \rightarrow \infty$ . Then (ii) holds by an argument similar to that for (i).

(iii) By the Cauchy–Schwarz inequality,

$$\begin{aligned} |\frac{1}{n}e'_i\bar{Z}'_nM'_n(I_n - P_{K,n})\bar{Z}_n(\rho_0)e_j|^2 &\leq \frac{1}{n}e'_i\bar{Z}'_nM'_n(I_n - P_{K,n})M_n\bar{Z}_ne_i \cdot \frac{1}{n}e'_j\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)e_j \\ &\leq \Delta_{nK,1}\Delta_{nK,2} \end{aligned}$$

where  $e_i$  denotes the  $i$ th column of the  $(m + 1) \times (m + 1)$  identity matrix. Thus the first result follows. The second result in (iii) follows by

$$\begin{aligned} |\frac{1}{n}e'_i\bar{Z}'_nM'_n(I_n - P_{K,n})A_n\bar{Z}_n(\rho_0)e_j|^2 &\leq \frac{1}{n}e'_i\bar{Z}'_nM'_n(I_n - P_{K,n})M_n\bar{Z}_ne_i \cdot \frac{1}{n}e'_j\bar{Z}'_n(\rho_0)A'_nA_n\bar{Z}_n(\rho_0)e_j \\ &= O(\Delta_{nK,2}) \end{aligned}$$

(iv) By Chebyshev’s inequality,

$$P\left(\left|\frac{1}{\sqrt{n}}e_i'\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_n\epsilon_n\right| > \eta\right) \leq \frac{\sigma_\epsilon^2}{n\eta^2}e_i'\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_nA_n'(I_n - P_{K,n})\bar{Z}_n(\rho_0)e_i$$

for some  $\eta > 0$ . Let  $A_nA_n' = A_{1n}A_{2n}A_{1n}'$ , where  $A_{1n}$  is an orthonormal matrix whose columns are  $A_nA_n'$ ’s eigenvectors and  $A_{2n}$  is a diagonal matrix with the diagonal elements being  $A_nA_n'$ ’s eigenvalues. Then

$$\begin{aligned} & \frac{1}{n}e_i'\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_nA_n'(I_n - P_{K,n})\bar{Z}_n(\rho_0)e_i \\ &= \frac{1}{n}e_i'\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_{1n}A_{2n}A_{1n}'(I_n - P_{K,n})\bar{Z}_n(\rho_0)e_i \\ &\leq \frac{1}{n}\iota_n e_i'\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)e_i \\ &\leq \frac{1}{n}\|A_nA_n'\|_\infty e_i'\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0)e_i \\ &= O(\Delta_{nK,1}) \end{aligned}$$

where  $\iota_n$  is the largest eigenvalue of  $A_nA_n'$  and the last inequality follows by the spectral radius theorem. Thus  $\frac{1}{\sqrt{n}}e_i'\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_n\epsilon_n = O_P(\Delta_{nK,1}^{1/2})$  and  $\frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)(I_n - P_{K,n})A_n\epsilon_n = O_P(\Delta_{nK,1}^{1/2})$ . Other results follow similarly.

(v) Use the expression  $A_nA_n' = A_{1n}A_{2n}A_{1n}'$  as in the proof of (iv), then

$$\begin{aligned} & \frac{1}{n}\text{tr}[\bar{Z}'_nM_n'(I_n - P_{K,n})A_nA_n'(I_n - P_{K,n})M_n\bar{Z}_n] \\ &\leq \frac{1}{n}\iota_n \text{tr}[\bar{Z}'_nM_n'(I_n - P_{K,n})A_{1n}A_{1n}'(I_n - P_{K,n})M_n\bar{Z}_n] \\ &= \frac{1}{n}\iota_n \text{tr}[\bar{Z}'_nM_n'(I_n - P_{K,n})M_n\bar{Z}_n] \\ &= O(\Delta_{nK,2}) \end{aligned}$$

(vi) The first two results are Lemma A.3 (vi) in [1]. For the third result, either  $\Delta_{nK,1} = 0$ , in which case  $\sqrt{K\Delta_{nK,1}/n}/(K^2/n + \Delta_{nK,1}) = 0$ , or  $\sqrt{K\Delta_{nK,1}/n}/(K^2/n + \Delta_{nK,1}) = \frac{1}{K^2/\sqrt{nK\Delta_{nK,1}} + \sqrt{n\Delta_{nK,1}/K}} \leq \frac{1}{2\sqrt{K}} \rightarrow 0$ , by the inequality of arithmetic and geometric means. Thus the result follows. The last result follows similarly.  $\square$

**Lemma 8.** With  $\Delta_{nK,1}$  and  $\Delta_{nK,2}$  defined, respectively, in Lemma 7 (i) and (ii),

(i)  $\frac{1}{n}Z'_n(\rho_0)P_{K,n}Z_n(\rho_0) = H_n + T_{1n}^H + T_{2n}^H + T_{3n}^H + T_{4n}^H$ ,

where

- (a)  $H_n = \frac{1}{n}\bar{Z}'_n(\rho_0)\bar{Z}_n(\rho_0) = O(1)$ ,
- (b)  $T_{1n}^H = -\frac{1}{n}\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) = O(\Delta_{nK,1})$ ,
- (c)  $T_{2n}^H = \frac{1}{n}[\bar{Z}'_n(\rho_0)R_n\zeta_n + \zeta_n'R_n\bar{Z}_n(\rho_0)] = O_P(n^{-1/2})$ ,
- (d)  $T_{3n}^H = \frac{1}{n}\zeta_n'R_nP_{K,n}R_n\zeta_n = O_P(K/n)$  and
- (e)  $T_{4n}^H = -\frac{1}{n}[\bar{Z}'_n(\rho_0)(I_n - P_{K,n})R_n\zeta_n + \zeta_n'R_n(I_n - P_{K,n})\bar{Z}_n(\rho_0)] = O_P(\sqrt{\Delta_{nK,1}/n}) = O_P(K/n + \Delta_{nK,1})$ .

- (ii)  $\frac{1}{n} Z_n'(\rho_0) P_{K,n} M_n Z_n = \frac{1}{n} \bar{Z}_n'(\rho_0) M_n \bar{Z}_n + O(\sqrt{\Delta_{nK,1} \Delta_{nK,2}}) + O_P(n^{-1/2}) + O_P(K/n) + o_P(K/n + \Delta_{nK,1}) + O_P(\sqrt{\Delta_{nK,2}/n})$ .
- (iii)  $\frac{1}{n} Z_n' M_n' P_{K,n} M_n Z_n = \frac{1}{n} \bar{Z}_n' M_n' M_n \bar{Z}_n + O(\Delta_{nK,2}) + O_P(n^{-1/2}) + O_P(K/n) + O_P(\sqrt{\Delta_{nK,2}/n})$ .
- (iv)  $\frac{1}{\sqrt{n}} Z_n'(\rho_0) P_{K,n} \epsilon_n = h_n + T_{1n}^h + T_{2n}^h$ ,  
 where  $h_n = \frac{1}{\sqrt{n}} \bar{Z}_n'(\rho_0) \epsilon_n = O_P(1)$ ,  $T_{1n}^h = -\frac{1}{\sqrt{n}} \bar{Z}_n'(\rho_0) (I_n - P_{K,n}) \epsilon_n = O_P(\Delta_{nK,1}^{1/2})$  and  $T_{2n}^h = \frac{1}{\sqrt{n}} \zeta_n' R_n' P_{K,n} \epsilon_n = O_P(K/\sqrt{n})$ .
- (v)  $\frac{1}{\sqrt{n}} Z_n'(\rho_0) P_{K,n} M_n u_n = \frac{1}{\sqrt{n}} \bar{Z}_n'(\rho_0) M_n u_n + O_P(\Delta_{nK,1}^{1/2}) + \frac{1}{\sqrt{n}} \zeta_n' R_n' P_{K,n} M_n u_n$ ,  
 where  $\frac{1}{\sqrt{n}} \zeta_n' R_n' P_{K,n} M_n u_n = O_P(K/\sqrt{n})$ .
- (vi)  $\frac{1}{\sqrt{n}} Z_n' M_n' P_{K,n} \epsilon_n = \frac{1}{\sqrt{n}} \bar{Z}_n' M_n' \epsilon_n - \frac{1}{\sqrt{n}} \bar{Z}_n' M_n' (I_n - P_{K,n}) \epsilon_n + \frac{1}{\sqrt{n}} \zeta_n' M_n' P_{K,n} \epsilon_n$ ,  
 where  $\frac{1}{\sqrt{n}} \bar{Z}_n' M_n' (I_n - P_{K,n}) \epsilon_n = O_P(\Delta_{nK,2}^{1/2})$  and  $\frac{1}{\sqrt{n}} \zeta_n' M_n' P_{K,n} \epsilon_n = O_P(K/\sqrt{n})$ .
- (vii)  $\frac{1}{\sqrt{n}} Z_n' M_n' P_{K,n} M_n u_n = \frac{1}{\sqrt{n}} \bar{Z}_n' M_n' M_n u_n + O_P(\Delta_{nK,2}^{1/2}) + O_P(K/\sqrt{n})$ .
- (viii)  $\frac{1}{\sqrt{n}} [\zeta_n' R_n' P_{K,n} R_n \zeta_n - E(\zeta_n' R_n' P_{K,n} R_n \zeta_n)] = O_P(\sqrt{K/n})$  and  $\frac{1}{\sqrt{n}} [\zeta_n' A_n \epsilon_n - E(\zeta_n' A_n \epsilon_n)] = O_P(\sqrt{K/n})$ ,  
 where  $A_n = M_n' P_{K,n}$ ,  $R_n' P_{K,n}$  or  $R_n' P_{K,n} M_n R_n^{-1}$ .

*Proof.* (i) Because  $Z_n = \bar{Z}_n + \zeta_n$ , we have the decomposition that  $\frac{1}{n} Z_n'(\rho_0) P_{K,n} Z_n(\rho_0) = H_n + T_{1n}^H + T_{2n}^H + T_{3n}^H + T_{4n}^H$ . Since elements of  $\bar{Z}_n$  are uniformly bounded,  $H_n = O(1)$ . By Lemma 7 (i),  $T_{1n}^H = O(\Delta_{nK,1})$ . By Lemma 6,  $T_{2n}^H = O_P(n^{-1/2})$ . By Lemmas 3 and 4,  $E(T_{3n}^H) = O(K/n)$ , and, hence,  $T_{3n}^H = O_P(K/n)$  by Markov's inequality. By Lemma 7 (iv) and (vi),  $T_{4n}^H = O_P(\sqrt{\Delta_{nK,1}/n}) = o_P(K/n + \Delta_{nK,1})$ .

(ii) and (iii) follow similarly to (i).

(iv) Because  $Z_n = \bar{Z}_n + \zeta_n$ , we have  $\frac{1}{\sqrt{n}} Z_n'(\rho_0) P_{K,n} u_n = h_n + T_{1n}^h + T_{2n}^h$ . By Lemma 6,  $h_n = O_P(1)$ . By Lemma 7 (iv),  $T_{1n}^h = O_P(\Delta_{nK,1}^{1/2})$ . As  $\zeta_n = [G_n v_n \gamma_0 + G_n R_n^{-1} \epsilon_n, v_n]$ , by Lemmas 3 and 4,  $E(T_{2n}^h T_{2n}^{h'}) = O(K^2/n)$ , and, hence,  $T_{2n}^h = O_P(K/\sqrt{n})$  by Chebyshev's inequality.

(v), (vi) and (vii) follow similarly to (iv). (viii) follows directly by Lemma 5. □

The following lemma shows the orders of some expectation terms, which are helpful in determining the approximated MSEs of the GS2SLS and CGS2SLS estimators.

- Lemma 9.** (i)  $E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n']$  is UB, where  $\mathbb{D}_n$  and  $\mathbb{F}_n$  are given in Proposition 2.
- (ii) Let the elements of  $n \times n$  matrices  $\{A_n\}$  be uniformly bounded, then elements of  $E(\epsilon_n \epsilon_n' A_n v_n)$  and  $E(\epsilon_n \epsilon_n' A_n \epsilon_n)$  are uniformly bounded.
- (iii) Let the elements of  $n \times m$  matrices  $\{B_n\}$  be uniformly bounded, where  $m$  is a finite fixed number, then  $E(\epsilon_n \epsilon_n' B_n v_n')$  is UB.
- (iv) Let the elements of  $n$ -dimensional vectors  $\{C_n\}$  be uniformly bounded, then  $E(\epsilon_n \epsilon_n' C_n \epsilon_n')$  is UB.

*Proof.* (i) Let  $\mathbb{D}_n = [d_{n,ij}]$  and  $\mathbb{F}_n = [f_{n1}, \dots, f_{nn}]$ . The  $(i, i)$ th element of  $E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n']$  is

$$E(\epsilon_{ni}^4) d_{n,ii} + \mu_3 f_{ni} + \sigma_\epsilon^4 [\text{tr}(\mathbb{D}_n) - d_{n,ii}] = [E(\epsilon_{ni}^4) - \sigma_\epsilon^4] d_{n,ii} + \mu_3 f_{ni}$$

and the  $(i, j)$ th element for  $i \neq j$  is  $\sigma_\epsilon^4 (d_{n,ij} + d_{n,ji})$ . Since  $\mathbb{D}_n$  is UB and elements of  $\mathbb{F}_n$  are uniformly bounded,  $E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n']$  is UB.

(ii) Let  $A_n = [a_{n,ij}]$ , then the  $i$ th row of  $E(\epsilon_n \epsilon_n' A_n v_n)$  is  $\sum_{r=1}^n \sum_{s=1}^n a_{n,rs} E(\epsilon_{ni} \epsilon_{nr} v_{ns}) = a_{n,ii} E(\epsilon_{ni}^2 v_{ni})$ . Thus elements of  $E(\epsilon_n \epsilon_n' A_n v_n)$  are uniformly bounded. Similarly, elements of  $E(\epsilon_n \epsilon_n' A_n \epsilon_n)$  are uniformly bounded.

(iii) Let  $B_n = [b'_{n1}, \dots, b'_{nn}]'$ , then the  $(i, j)$ th element of  $E(\epsilon_n \epsilon_n' B_n v_n')$  is  $\sum_{l=1}^n b_{nl} E(\epsilon_{ni} \epsilon_{nj} v_{lj}')$ , which is equal to  $b_{ni} E(\epsilon_{ni}^2 v_{ni}')$  when  $i = j$  and 0 otherwise. Thus  $E(\epsilon_n \epsilon_n' B_n v_n')$  is UB.

(iv) follows similarly to (iii). □

**Lemma 10.** *The sequence of matrices  $\{(I_n - \lambda W_n)^{-1}\}$  is UB in a neighborhood of  $\lambda_0$  and  $\{R_n^{-1}(\rho)\}$  is UB in a neighborhood of  $\rho_0$ .*

*Proof.* See [22]. □

The following lemma shows the dominant components of estimation errors for parameters of the model or for  $\Upsilon_n(K)$  in (9), which help to derive the approximated MSE of the CGS2SLS estimator.

**Lemma 11.** *Let  $L_{1n} = \frac{1}{\sqrt{n}}(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n)$  as in Proposition 2, then*

(i) *the GS2SLS estimator  $\tilde{\delta}_n = [Z_n'(\tilde{\rho}_n) P_{0,n} Z_n(\tilde{\rho}_n)]^{-1} Z_n'(\tilde{\rho}_n) P_{0,n} y_n(\tilde{\rho}_n)$  satisfies  $\sqrt{n}(\tilde{\delta}_n - \delta_0) = L_{2n} + o_P(1)$ , where  $L_{2n} = [\frac{1}{n} \bar{Z}_n'(\rho_0) P_{0,n} \bar{Z}_n(\rho_0)]^{-1} \frac{1}{\sqrt{n}} \bar{Z}_n'(\rho_0) P_{0,n} \epsilon_n$ ,*

(ii)  *$\sqrt{n}(\tilde{\sigma}_{v\epsilon} - \sigma_{v\epsilon})' = L_{3n} + o_P(1)$ , where*

$$L_{3n} = \sqrt{n} \left( \frac{1}{n} v_n' \epsilon_n - \sigma_{v\epsilon}' \right) + \frac{1}{\sqrt{n}} \bar{Z}_n' \epsilon_n - \frac{\sigma_{v\epsilon}'}{n} \text{tr}(M_n R_n^{-1}) L_{1n} - \frac{1}{n} [\bar{Z}_n' R_n \bar{Z}_n + E(v_n' R_n \zeta_n)] L_{2n}$$

with  $E(v_n' R_n \zeta_n) = [\text{tr}(R_n G_n) \Sigma_v \gamma_0 + \text{tr}(G_n) \sigma_{v\epsilon}', \Sigma_v \text{tr}(R_n)]$ ,

(iii)  *$\sqrt{n}(\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2) = L_{4n} + o_P(1)$ , where*

$$L_{4n} = \sqrt{n} \left( \frac{1}{n} \epsilon_n' \epsilon_n - \sigma_{v\epsilon}^2 \right) - \frac{2\sigma_\epsilon^2}{n} \text{tr}(M_n R_n^{-1}) L_{1n} - \frac{2}{n} E(\epsilon_n' R_n \zeta_n) L_{2n}$$

with  $E(\epsilon_n' R_n \zeta_n) = [\text{tr}(R_n G_n) \sigma_{v\epsilon} \gamma_0 + \text{tr}(G_n) \sigma_\epsilon^2, \sigma_{v\epsilon} \text{tr}(R_n)]$

(iv)  *$\frac{1}{\sqrt{n}}[\tilde{\Upsilon}_n(K) - \Upsilon_n(K)] = \frac{1}{n}(a_1', a_2')' + o_P(K/n) = O_P(K/n)$ , where*

$$a_1 = [\sigma_\epsilon^2 \text{tr}(\Gamma_{nK,3} M_n R_n^{-1}) - \sigma_\epsilon^2 \text{tr}(P_{K,n} M_n G_n R_n^{-1}) - \text{tr}(P_{K,n} M_n G_n) \sigma_{v\epsilon} \gamma_0] L_{1n} + [\text{tr}(\Gamma_{nK,2} G_n) \sigma_{v\epsilon} \gamma_0 + \sigma_\epsilon^2 \text{tr}(\Gamma_{nK,2} G_n R_n^{-1}), \sigma_{v\epsilon} \text{tr}(\Gamma_{nK,2})] L_{2n} + \text{tr}(\Gamma_{nK,2}) \gamma_0' L_{3n} + \text{tr}(\Gamma_{nK,3}) L_{4n}$$

and  $a_2 = -\sigma_{v\epsilon}' \text{tr}(P_{K,n} M_n) L_{1n} + \text{tr}(\Gamma_{nK,1}) L_{3n}$ .

*Proof.* (i) The  $\tilde{\delta}_n$  satisfies

$$\sqrt{n}(\tilde{\delta}_n - \delta_0) = \left[ \frac{1}{n} Z_n'(\tilde{\rho}_n) Q_{0,n} \left( \frac{1}{n} Q_{0,n}' Q_{0,n} \right)^{-1} \frac{1}{n} Q_{0,n}' Z_n(\tilde{\rho}_n) \right]^{-1} \frac{1}{n} Z_n'(\tilde{\rho}_n) Q_{0,n} \left( \frac{1}{n} Q_{0,n}' Q_{0,n} \right)^{-1} \frac{1}{\sqrt{n}} Q_{0,n}' u_n(\tilde{\rho}_n).$$

Note that  $\frac{1}{\sqrt{n}} Q_{0,n}' u_n(\tilde{\rho}_n) = \frac{1}{\sqrt{n}} Q_{0,n}' \epsilon_n + \frac{1}{n} Q_{0,n}' M_n R_n^{-1} \epsilon_n \sqrt{n}(\rho_0 - \tilde{\rho}_n)$  and

$$\frac{1}{n} Q_{0,n}' Z_n(\tilde{\rho}_n) = \frac{1}{n} Q_{0,n}' \bar{Z}_n(\rho_0) + \frac{1}{n} Q_{0,n}' R_n \zeta_n + \left[ \frac{1}{n} Q_{0,n}' M_n \bar{Z}_n + \frac{1}{n} Q_{0,n}' M_n \zeta_n \right] (\rho_0 - \tilde{\rho}_n).$$

By Lemma 6,  $\frac{1}{n} Q_{0,n}' R_n \zeta_n = o_P(1)$ ,  $\frac{1}{n} Q_{0,n}' M_n \zeta_n = o_P(1)$  and  $\frac{1}{n} Q_{0,n}' M_n R_n^{-1} \epsilon_n = o_P(1)$ . By Proposition 2,  $\sqrt{n}(\tilde{\rho}_n - \rho_0) = L_{1n} + o_P(1) = O_P(1)$ . Furthermore,  $\frac{1}{n} Q_{0,n}' \bar{Z}_n(\rho_0) = O(1)$  and  $\frac{1}{n} Q_{0,n}' M_n \bar{Z}_n = O(1)$ . Thus

$$\sqrt{n}(\tilde{\delta}_n - \delta_0) = \left[ \frac{1}{n} \bar{Z}_n'(\rho_0) P_{0,n} \bar{Z}_n(\rho_0) \right]^{-1} \frac{1}{\sqrt{n}} \bar{Z}_n'(\rho_0) P_{0,n} \epsilon_n + o_P(1)$$

(ii) Write  $\tilde{\sigma}'_{v\epsilon}$  as

$$\begin{aligned} \tilde{\sigma}'_{v\epsilon} &= \frac{1}{n} Z'_{2n} [R_n + (\rho_0 - \tilde{\rho}_n) M_n] [u_n + Z_n(\delta_0 - \tilde{\delta}_n)] \\ &= \frac{1}{n} Z'_{2n} \epsilon_n + \frac{1}{n} Z'_{2n} M_n u_n (\rho_0 - \tilde{\rho}_n) + \frac{1}{n} Z'_{2n} R_n Z_n (\delta_0 - \tilde{\delta}_n) + \frac{1}{n} Z'_{2n} M_n Z_n (\delta_0 - \tilde{\delta}_n) (\rho_0 - \tilde{\rho}_n) \end{aligned}$$

By Lemmas 5 and 6,

- (a)  $\frac{1}{n} Z'_{2n} M_n u_n = \frac{1}{n} E(v'_n M_n u_n) + o_P(1) = \frac{\sigma'_{v\epsilon}}{n} \text{tr}(M_n R_n^{-1}) + o_P(1) = O_P(1)$ ,
- (b)  $\frac{1}{n} Z'_{2n} R_n Z_n = \frac{1}{n} E(Z'_{2n} R_n Z_n) + o_P(1)$  with  $E(Z'_{2n} R_n Z_n) = \bar{Z}'_{2n} R_n \bar{Z}_n + [\text{tr}(R_n G_n) \Sigma_v \gamma_0 + \text{tr}(G_n) \sigma'_{v\epsilon}, \Sigma_v \text{tr}(R_n)] = O(n)$ , and
- (c)  $\frac{1}{n} Z'_{2n} M_n Z_n = O_P(1)$ .

Then

$$\begin{aligned} \sqrt{n}(\tilde{\sigma}_{v\epsilon} - \sigma_{v\epsilon})' &= \sqrt{n} \left( \frac{1}{n} v'_n \epsilon_n - \sigma'_{v\epsilon} \right) + \frac{1}{\sqrt{n}} \bar{Z}'_{2n} \epsilon_n + \frac{\sigma'_{v\epsilon}}{n} \text{tr}(M_n R_n^{-1}) \sqrt{n}(\rho_0 - \tilde{\rho}_n) \\ &\quad + \frac{1}{n} E(Z'_{2n} R_n Z_n) \sqrt{n}(\delta_0 - \tilde{\delta}_n) + o_P(1) \end{aligned}$$

The result follows as  $\sqrt{n}(\tilde{\rho}_n - \rho_0) = L_{1n} + o_P(1)$  by Proposition 2 and  $\sqrt{n}(\tilde{\delta}_n - \delta_0) = L_{2n} + o_P(1)$  by (i).

(iii) Note that

$$\begin{aligned} R_n(\tilde{\rho}_n)(y_n - Z_n \tilde{\delta}_n) &= [R_n + (\rho_0 - \tilde{\rho}_n) M_n] [u_n + Z_n(\delta_0 - \tilde{\delta}_n)] \\ &= \epsilon_n + (\rho_0 - \tilde{\rho}_n) M_n u_n + R_n Z_n (\delta_0 - \tilde{\delta}_n) + M_n Z_n (\delta_0 - \tilde{\delta}_n) (\rho_0 - \tilde{\rho}_n) \end{aligned}$$

then by an argument similar to that for (ii),

$$\sqrt{n}(\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2) = \sqrt{n} \left( \frac{1}{n} \epsilon'_n \epsilon_n - \sigma_\epsilon^2 \right) + \frac{2}{n} \epsilon'_n M_n R_n^{-1} \epsilon_n \sqrt{n}(\rho_0 - \tilde{\rho}_n) + \frac{2}{n} \epsilon'_n R_n Z_n \sqrt{n}(\delta_0 - \tilde{\delta}_n) + o_P(1)$$

where

- (a)  $\frac{1}{n} \epsilon'_n M_n R_n^{-1} \epsilon_n = \frac{1}{n} E(\epsilon'_n M_n R_n^{-1} \epsilon_n) + o_P(1)$  with  $E(\epsilon'_n M_n R_n^{-1} \epsilon_n) = \sigma_\epsilon^2 \text{tr}(M_n R_n^{-1}) = O(n)$ , and
- (b)  $\frac{1}{n} \epsilon'_n R_n Z_n = \frac{1}{n} E(\epsilon'_n R_n \zeta_n) + o_P(1)$  with  $E(\epsilon'_n R_n \zeta_n) = [\text{tr}(R_n G_n) \sigma_{v\epsilon} \gamma_0 + \text{tr}(G_n) \sigma_\epsilon^2, \Sigma_{v\epsilon} \text{tr}(R_n)] = O(n)$ .

The result follows by using the expressions for  $\sqrt{n}(\tilde{\delta}_n - \delta_0)$  and  $\sqrt{n}(\tilde{\rho}_n - \rho_0)$ .

(iv) By the mean value theorem,

$$\begin{aligned} &\frac{1}{\sqrt{n}} [\tilde{\Upsilon}_n(K) - \Upsilon_n(K)] \\ &= \frac{1}{n} \begin{pmatrix} \text{tr}(P_{K,n} \ddot{R}_n \ddot{G}_n^2) \ddot{\sigma}_{v\epsilon} \ddot{\gamma}_n + \ddot{\sigma}_\epsilon^2 \text{tr}(P_{K,n} \ddot{R}_n \ddot{G}_n^2 \ddot{R}_n^{-1}) & \ddot{\sigma}_{v\epsilon} \text{tr}(P_{K,n} \ddot{R}_n \ddot{G}_n) \\ 0 & 0 \end{pmatrix} \sqrt{n}(\tilde{\delta}_n - \delta_0) \\ &\quad + \frac{1}{n} \begin{pmatrix} -\text{tr}(P_{K,n} M_n \ddot{G}_n) \ddot{\sigma}_{v\epsilon} \ddot{\gamma}_n + \ddot{\sigma}_\epsilon^2 [-\text{tr}(P_{K,n} M_n \ddot{G}_n \ddot{R}_n^{-1}) + \text{tr}(P_{K,n} \ddot{R}_n \ddot{G}_n \ddot{R}_n^{-1} M_n \ddot{R}_n^{-1})] \\ -\ddot{\sigma}'_{v\epsilon} \text{tr}(P_{K,n} M_n) \end{pmatrix} \sqrt{n}(\tilde{\rho}_n - \rho_0) \\ &\quad + \frac{1}{n} \begin{pmatrix} \text{tr}(P_{K,n} \ddot{R}_n \ddot{G}_n \ddot{R}_n^{-1}) \\ 0 \end{pmatrix} \sqrt{n}(\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2) + \frac{1}{n} \begin{pmatrix} \text{tr}(P_{K,n} \ddot{R}_n \ddot{G}_n) \ddot{\gamma}'_n \\ \text{tr}(P_{K,n} \ddot{R}_n) I_m \end{pmatrix} \sqrt{n}(\tilde{\sigma}_{v\epsilon} - \sigma_{v\epsilon})' \end{aligned}$$



where  $\check{\sigma}_{v\epsilon}$  is between  $\tilde{\sigma}_{v\epsilon}$  and  $\sigma_{v\epsilon}$ ,  $\check{\sigma}_\epsilon^2$  is between  $\tilde{\sigma}_\epsilon^2$  and  $\sigma_\epsilon^2$ ,  $\check{\gamma}_n$  is between  $\tilde{\gamma}_n$  and  $\gamma_0$ ,  $\check{R}_n = R_n(\check{\rho}_n)$  with  $\check{\rho}_n$  being between  $\rho_0$  and  $\tilde{\rho}_n$ , and  $\check{G}_n = G_n(\check{\lambda}_n)$  with  $\check{\lambda}_n$  being between  $\lambda_0$  and  $\tilde{\lambda}_n$ . Let  $\text{tr}(P_{K,n}\check{A}_n)$  stand for a trace term that appeared in the above equation and  $\text{tr}(P_{K,n}A_n)$  be the term evaluated at the true  $\delta_0$  and  $\rho_0$ . Using the mean value theorem once again, then  $\frac{1}{n}[\text{tr}(P_{K,n}\check{A}_n) - \text{tr}(P_{K,n}A_n)] = o_P(K/n)$  by Lemmas 10 and 4. Thus by (ii), (iii) and Propositions 1 and 2,  $\frac{1}{\sqrt{n}}[\check{\Upsilon}_n(K) - \Upsilon_n(K)] = \frac{1}{n}(a'_1, a'_2)' + o_P(K/n)$ , where  $a_1 = O_P(K)$  and  $a_2 = O_P(K)$ .  $\square$

The following lemma, Lemma A.9 in [1], gives a sufficient condition that the chosen  $K$  by the minimization of  $\hat{S}_{n,\xi}(K)$ , say  $\hat{K}$ , is asymptotically optimal.

**Lemma 12.** *If  $\sup_K \frac{|\hat{S}_{n,\xi}(K) - S_{n,\xi}(K)|}{S_{n,\xi}(K)} \xrightarrow{p} 0$ , then  $\frac{S_{n,\xi}(\hat{K})}{\inf_K S_{n,\xi}(K)} \xrightarrow{p} 1$ .*

The following is a central limit theorem for linear-quadratic forms of disturbances from [17].

**Lemma 13.** *Suppose that  $\{A_n = [a_{n,ij}]\}$  is a sequence of symmetric  $n \times n$  matrices that are UB,  $b_{n,K} = (b_{nK,1}, \dots, b_{nK,n})'$  is a vector such that  $\sup_n n^{-1} \sum_{i=1}^n |b_{ni}|^{2+\eta_1} < \infty$  for some  $\eta_1 > 0$ , and  $\epsilon_{ni}$ 's in  $\epsilon_n = (\epsilon_{n1}, \dots, \epsilon_{nn})'$  are mutually independent, with mean zero, variance  $\sigma_{ni}^2$  and finite moment of order higher than four such that  $E(|\epsilon_{ni}|^{4+\eta_2})$  for some  $\eta_2 > 0$  are uniformly bounded for all  $n$  and  $i$ . Let  $\sigma_{Q_n}^2$  be the variance of  $Q_n$  where  $Q_n = \epsilon'_n A_n \epsilon_n + b'_{n,K} \epsilon_n - \sum_{i=1}^n a_{n,ii} \sigma_{ni}^2$ . Assume that  $\sigma_{Q_n}^2/n$  is bounded away from zero.*

Then,  $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0, 1)$ .

### C. Proofs

**Proof of Proposition 1.** As  $\check{\delta}_n = \delta_0 + (Z'_n P_{F_n} Z_n)^{-1} Z'_n P_{F_n} R_n^{-1} \epsilon_n$ ,

$$\sqrt{n}(\check{\delta}_n - \delta_0) = \left[ \frac{1}{n} Z'_n F_{0,n} \left( \frac{1}{n} F'_{0,n} F_{0,n} \right)^{-1} \frac{1}{n} F'_{0,n} Z_n \right]^{-1} \frac{1}{n} Z'_n F_{0,n} \left( \frac{1}{n} F'_{0,n} F_{0,n} \right)^{-1} \frac{1}{\sqrt{n}} F'_{0,n} R_n^{-1} \epsilon_n.$$

By Lemma 6,  $\frac{1}{n} F'_{0,n} \zeta_n = O_P(n^{-1/2})$  and  $\frac{1}{\sqrt{n}} F'_{0,n} R_n^{-1} \epsilon_n \xrightarrow{d} N(0, \lim_{n \rightarrow \infty} \frac{\sigma_\epsilon^2}{n} F'_{0,n} R_n^{-1} R_n^{-1} F_{0,n})$ . Hence,

$$\begin{aligned} \sqrt{n}(\check{\delta}_n - \delta_0) &= \left( \frac{1}{n} \bar{Z}'_n P_{F_n} \bar{Z}_n \right)^{-1} \frac{1}{\sqrt{n}} \bar{Z}'_n P_{F_n} R_n^{-1} \epsilon_n + O_P(n^{-1/2}) \\ &\xrightarrow{d} N\left(0, \lim_{n \rightarrow \infty} \left( \frac{1}{n} \bar{Z}'_n P_{F_n} \bar{Z}_n \right)^{-1} \frac{\sigma_\epsilon^2}{n} \bar{Z}'_n P_{F_n} R_n^{-1} R_n^{-1} P_{F_n} \bar{Z}_n \left( \frac{1}{n} \bar{Z}'_n P_{F_n} \bar{Z}_n \right)^{-1} \right) \end{aligned}$$

by Slutsky's lemma.  $\square$

**Proof of Proposition 2.** The consistency of  $\tilde{\rho}_n$  follows from the uniform convergence that

$$g'_n(\rho, \check{\delta}_n) g_n(\rho, \check{\delta}_n) - E g'_n(\rho, \delta_0) E g_n(\rho, \delta_0) = o_P(1) \text{ uniformly in } \rho \in [-a, a]$$

and the identification uniqueness condition ([23] [Theorem 3.4]).

To prove the uniform convergence, we first show that  $g_n(\rho, \check{\delta}_n) - g_n(\rho, \delta_0) = o_P(1)$  uniformly in  $\rho \in [-a, a]$ . As  $\epsilon_n(\rho, \check{\delta}_n) = R_n(\rho)(y_n - Z_n \check{\delta}_n) = R_n(\rho)[u_n + Z_n(\delta_0 - \check{\delta}_n)]$ ,

$$\begin{aligned} &\frac{1}{2n} \epsilon'_n(\rho, \check{\delta}_n) D_{nj}^s \epsilon_n(\rho, \check{\delta}_n) - \frac{1}{2n} \epsilon'_n(\rho, \delta_0) D_{nj}^s \epsilon_n(\rho, \delta_0) \\ &= \frac{1}{n} \epsilon'_n(\rho, \delta_0) D_{nj}^s R_n(\rho) Z_n (\delta_0 - \check{\delta}_n) + \frac{1}{2n} (\delta_0 - \check{\delta}_n)' Z'_n R'_n(\rho) D_{nj}^s R_n(\rho) Z_n (\delta_0 - \check{\delta}_n) \end{aligned}$$

Note that  $Z_n = \bar{Z}_n + \zeta_n$ ,  $\frac{1}{n}\epsilon'_n(\rho, \delta_0)D_{nj}^s R_n(\rho)Z_n = O_P(1)$  and  $\frac{1}{n}Z'_n R'_n(\rho)D_{nj}^s R_n(\rho)Z_n = O_P(1)$  by Lemmas 5 and 6. Then  $\frac{1}{2n}\epsilon'_n(\rho, \check{\delta}_n)D_{nj}^s \epsilon_n(\rho, \check{\delta}_n) - \frac{1}{2n}\epsilon'_n(\rho, \delta_0)D_{nj}^s \epsilon_n(\rho, \delta_0) = o_P(1)$ , as  $\check{\delta}_n - \delta_0 = o_P(1)$ . Since  $g_n(\rho, \delta)$  is quadratic in  $\rho$ , it follows that

$$g_n(\rho, \check{\delta}_n) - g_n(\rho, \delta_0) = o_P(1) \text{ uniformly in } \rho \in [-a, a]$$

By Lemma 5,  $g_n(\rho, \delta_0) - E g_n(\rho, \delta_0) = o_P(1)$  uniformly in  $\rho \in [-a, a]$ . Thus,

$$g_n(\rho, \check{\delta}_n) - E g_n(\rho, \delta_0) = o_P(1) \text{ uniformly in } \rho \in [-a, a]$$

Furthermore,  $E g_n(\rho, \delta_0) = O(1)$  uniformly in  $\rho \in [-a, a]$ . Hence,

$$\begin{aligned} &g'_n(\rho, \check{\delta}_n)g_n(\rho, \check{\delta}_n) - E g'_n(\rho, \delta_0) E g_n(\rho, \delta_0) \\ &= 2[g_n(\rho, \check{\delta}_n) - E g_n(\rho, \delta_0)]' E g_n(\rho, \delta_0) + [g_n(\rho, \check{\delta}_n) - E g_n(\rho, \delta_0)]'[g_n(\rho, \check{\delta}_n) - E g_n(\rho, \delta_0)] \\ &= o_P(1) \end{aligned}$$

uniformly in  $\rho \in [-a, a]$ .

We now show that the identification uniqueness condition holds. Note that  $E g_n(\rho, \delta_0) = \frac{\sigma_\epsilon^2}{2}\Xi_n[(\rho_0 - \rho), (\rho_0 - \rho)^2]'$ . Let  $\tau_{n1}$  and  $\tau_{n2}$  be the eigenvalues of  $\Xi'_n \Xi_n$ . Write  $\Xi'_n \Xi_n = \Xi'_{1n} \Xi_{2n} \Xi_{1n}$ , where  $\Xi_{2n}$  is a  $2 \times 2$  diagonal matrix with diagonal elements  $\tau_{n1}$  and  $\tau_{n2}$ , and  $\Xi_{1n}$  is an orthonormal matrix containing the eigenvectors of  $\Xi'_n \Xi_n$ . By Assumption 5, there exists some constant  $\eta > 0$  such that  $\tau_{n1} > \eta$  and  $\tau_{n2} > \eta$  for all  $n$ . Obviously,  $E g'_n(\rho_0, \delta_0) E g_n(\rho_0, \delta_0) = 0$ . Then,

$$\begin{aligned} &E g'_n(\rho, \delta_0) E g_n(\rho, \delta_0) - E g'_n(\rho_0, \delta_0) E g_n(\rho_0, \delta_0) \\ &= \frac{\sigma_\epsilon^4}{4}[\rho_0 - \rho, (\rho_0 - \rho)^2]\Xi'_{1n} \Xi_{2n} \Xi_{1n}[\rho_0 - \rho, (\rho_0 - \rho)^2]' \\ &\geq \frac{\eta\sigma_\epsilon^4}{4}[\rho_0 - \rho, (\rho_0 - \rho)^2]\Xi'_{1n} \Xi_{1n}[\rho_0 - \rho, (\rho_0 - \rho)^2]' \\ &= \frac{\eta\sigma_\epsilon^4}{4}[(\rho_0 - \rho)^2 + (\rho_0 - \rho)^4] \\ &> 0 \end{aligned}$$

for any  $\rho \neq \rho_0$ . Thus the identification uniqueness condition holds.

The consistency of  $\tilde{\rho}_n$  follows from the uniform convergence and identification uniqueness.

For the asymptotic distribution, by the mean value theorem, we have

$$0 = \frac{\partial g'_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho} g_n(\tilde{\rho}_n, \check{\delta}_n) = \frac{\partial g'_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho} [g_n(\rho_0, \check{\delta}_n) + \frac{\partial g_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho} (\tilde{\rho}_n - \rho_0)]$$

where  $\check{\rho}_n$  is between  $\tilde{\rho}_n$  and  $\rho_0$ . Then

$$\sqrt{n}(\tilde{\rho}_n - \rho_0) = -\left(\frac{\partial g'_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho} \frac{\partial g_n(\check{\rho}_n, \check{\delta}_n)}{\partial \rho}\right)^{-1} \frac{\partial g'_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho} \sqrt{n}g_n(\rho_0, \check{\delta}_n)$$

The  $i$ th element of  $\frac{\partial g_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho}$  is  $-\frac{1}{n}\epsilon'_n(\tilde{\rho}_n, \check{\delta}_n)D_{nj}^s M_n(y_n - Z_n \check{\delta}_n)$ , which can be expanded by using  $y_n - Z_n \check{\delta}_n = u_n + Z_n(\delta_0 - \check{\delta}_n)$  and  $\epsilon_n(\tilde{\rho}_n, \check{\delta}_n) = [R_n + (\rho_0 - \tilde{\rho}_n)M_n][u_n + Z_n(\delta_0 - \check{\delta}_n)]$ . By Lemmas 5 and 6, the terms involving  $(\delta_0 - \check{\delta}_n)$  or  $(\rho_0 - \tilde{\rho}_n)$  are  $O_P(n^{-1/2})$ . Therefore,

$$\begin{aligned} -\frac{1}{n}\epsilon'_n(\tilde{\rho}_n, \check{\delta}_n)D_{nj}^s M_n(y_n - Z_n \check{\delta}_n) &= -\frac{1}{n}\epsilon'_n D_{nj}^s M_n R_n^{-1} \epsilon_n + O_P(n^{-1/2}) \\ &= -\frac{1}{n}\sigma_\epsilon^2 \text{tr}(D_{nj}^s M_n R_n^{-1}) + O_P(n^{-1/2}) \\ &= O_P(1) \end{aligned}$$

by Lemma 5. Thus,  $\frac{\partial g_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho} = E \frac{\partial g_n(\rho_0, \delta_0)}{\partial \rho} + O_P(n^{-1/2}) = O_P(1)$ , where

$$E \frac{\partial g_n(\rho_0, \delta_0)}{\partial \rho} = -\frac{\sigma_\epsilon^2}{n} [\text{tr}(D_{n1}^s M_n R_n^{-1}), \dots, \text{tr}(D_{n, kd}^s M_n R_n^{-1})]'$$

Similarly,  $\frac{\partial g_n(\tilde{\rho}_n, \check{\delta}_n)}{\partial \rho} = E \frac{\partial g_n(\rho_0, \delta_0)}{\partial \rho} + O_P(n^{-1/2}) = O_P(1)$ . Thus,

$$\sqrt{n}(\tilde{\rho}_n - \rho_0) = -\left(E \frac{\partial g_n'(\rho_0, \delta_0)}{\partial \rho}\right) \left(E \frac{\partial g_n(\rho_0, \delta_0)}{\partial \rho}\right)^{-1} \left(E \frac{\partial g_n'(\rho_0, \delta_0)}{\partial \rho}\right) \sqrt{n}g_n(\rho_0, \check{\delta}_n) + O_P(n^{-1/2})$$

For  $\sqrt{n}g_n(\rho_0, \check{\delta}_n)$ , the  $i$ th element is

$$\frac{1}{\sqrt{n}}[\epsilon_n + R_n Z_n(\delta_0 - \check{\delta}_n)]' D_{nj} [\epsilon_n + R_n Z_n(\delta_0 - \check{\delta}_n)] = \frac{1}{\sqrt{n}} \epsilon_n' D_{nj} \epsilon_n + \frac{1}{n} \epsilon_n' D_{nj}^s R_n Z_n \sqrt{n}(\delta_0 - \check{\delta}_n) + O_P(n^{-1/2})$$

By Lemmas 5 and 6,

$$\frac{1}{n} \epsilon_n' D_{nj}^s R_n Z_n = \frac{1}{n} \epsilon_n' D_{nj}^s R_n (\bar{Z}_n + \zeta_n) = \frac{1}{n} E(\epsilon_n' D_{nj}^s R_n \zeta_n) + O_P(n^{-1/2}) = O_P(1)$$

where

$$E(\epsilon_n' D_{nj}^s R_n \zeta_n) = [\text{tr}(D_{nj}^s R_n G_n) \sigma_{v\epsilon} \gamma_0 + \sigma_\epsilon^2 \text{tr}(D_{nj}^s R_n G_n R_n^{-1}), \text{tr}(D_{nj}^s R_n) \sigma_{v\epsilon}]$$

By Proposition 1,

$$\sqrt{n}(\check{\delta}_n - \delta_0) = \left(\frac{1}{n} \bar{Z}_n' P_{F_n} \bar{Z}_n\right)^{-1} \frac{1}{\sqrt{n}} \bar{Z}_n' P_{F_n} R_n^{-1} \epsilon_n + O_P(n^{-1/2})$$

Then the  $i$ th element of  $\sqrt{n}g_n(\rho_0, \check{\delta}_n)$  is  $A_{ni} + O_P(n^{-1/2})$ , where

$$A_{ni} = \frac{1}{\sqrt{n}} \epsilon_n' D_{nj} \epsilon_n - \frac{1}{n} E(\epsilon_n' D_{nj}^s R_n \zeta_n) \left(\frac{1}{n} \bar{Z}_n' P_{F_n} \bar{Z}_n\right)^{-1} \frac{1}{\sqrt{n}} \bar{Z}_n' P_{F_n} R_n^{-1} \epsilon_n$$

Hence,

$$\begin{aligned} \sqrt{n}(\tilde{\rho}_n - \rho_0) &= \left(\frac{\sigma_\epsilon^2}{n^2} \sum_{j=1}^{k_d} \text{tr}^2(D_{nj}^s M_n R_n^{-1})\right)^{-1} \sum_{j=1}^{k_d} \frac{1}{n} \text{tr}(D_{nj}^s M_n R_n^{-1}) A_{ni} + O_P(n^{-1/2}) \\ &= \frac{1}{\sqrt{n}} (\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) + O_P(n^{-1/2}) \end{aligned}$$

where  $\mathbb{D}_n$  and  $\mathbb{F}_n$  are in, respectively, (6) and (7). Note that  $\text{tr}(\mathbb{D}_n) = 0$ , then  $\sqrt{n}(\tilde{\rho}_n - \rho_0)$  is asymptotically normal with a finite variance by Lemma 13.  $\square$

**Proof of Proposition 3.** The GS2SLS estimator  $\hat{\delta}_{2sls,n}$  satisfies

$$\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0 - b_{n,K}) = \left[\frac{1}{n} Z_n'(\tilde{\rho}_n) P_{K,n} Z_n(\tilde{\rho}_n)\right]^{-1} \frac{1}{\sqrt{n}} [Z_n'(\tilde{\rho}_n) P_{K,n} R_n(\tilde{\rho}_n) u_n - E(\zeta_n' R_n' P_{K,n} \epsilon_n)]$$

where

$$E(\zeta_n' R_n' P_{K,n} \epsilon_n) = [\text{tr}(\Gamma_{nK,2}) \sigma_{v\epsilon} \gamma_0 + \sigma_\epsilon^2 \text{tr}(\Gamma_{nK,3}), \text{tr}(\Gamma_{nK,1}) \sigma_{v\epsilon}]' = O(K)$$

by Lemma 4. Write  $Z_n(\tilde{\rho}_n) = Z_n(\rho_0) + (\rho_0 - \tilde{\rho}_n) M_n Z_n$ , then

$$\begin{aligned} \frac{1}{n} Z_n'(\tilde{\rho}_n) P_{K,n} Z_n(\tilde{\rho}_n) &= \frac{1}{n} Z_n'(\rho_0) P_{K,n} Z_n(\rho_0) + \frac{1}{n} [Z_n' M_n' P_{K,n} Z_n(\rho_0)]^s (\rho_0 - \tilde{\rho}_n) \\ &\quad + \frac{1}{n} Z_n' M_n' P_{K,n} M_n Z_n (\rho_0 - \tilde{\rho}_n)^2 \end{aligned}$$

By Lemma 8 (i)–(iii),

- (a) if  $K/n \rightarrow c \neq 0$ ,  $\frac{1}{n}Z'_n(\rho_0)P_{K,n}Z_n(\rho_0) = H_n + \frac{1}{n}\zeta'_n R'_n P_{K,n} R_n \zeta_n + o_P(1) = O_P(1)$ ,  $\frac{1}{n}Z'_n M'_n P_{K,n} Z_n(\rho_0) = O_P(1)$  and  $\frac{1}{n}Z'_n M'_n P_{K,n} M_n Z_n = O_P(1)$ ;
- (b) if  $K/n \rightarrow 0$ ,  $\frac{1}{n}Z'_n(\rho_0)P_{K,n}Z_n(\rho_0) = H_n + o_P(1) = O_P(1)$ ,  $\frac{1}{n}Z'_n M'_n P_{K,n} Z_n(\rho_0) = O_P(1)$  and  $\frac{1}{n}Z'_n M'_n P_{K,n} M_n Z_n = O_P(1)$ .

By Lemma 5,  $\frac{1}{n}\zeta'_n R'_n P_{K,n} R_n \zeta_n - \frac{1}{n}\Omega_{n1}(K) = O_P(\sqrt{K}/n)$ , where  $\Omega_{n1}(K) = O(K)$  by Lemma 4. By Proposition 1,  $\sqrt{n}(\delta_0 - \check{\delta}_n) = O_P(1)$ . Hence,

- (c) if  $K/n \rightarrow c \neq 0$ ,  $\frac{1}{n}Z'_n(\tilde{\rho}_n)P_{K,n}Z_n(\tilde{\rho}_n) = H_n + \frac{1}{n}\Omega_{n1}(K) + o_P(1) = O_P(1)$  and  $b_{n,K} - \bar{b}_{nK,1} = o_P(1)$ ;
- (d) if  $K/n \rightarrow 0$ ,  $\frac{1}{n}Z'_n(\tilde{\rho}_n)P_{K,n}Z_n(\tilde{\rho}_n) = H_n + o_P(1) = O_P(1)$  and  $b_{n,K} - \bar{b}_{nK,2} = o_P(K/n)$ .

As  $R_n u_n = \epsilon_n$ ,  $Z_n(\tilde{\rho}_n) = Z_n(\rho_0) + (\rho_0 - \tilde{\rho}_n)M_n Z_n$  and  $R_n(\tilde{\rho}_n) = R_n + (\rho_0 - \tilde{\rho}_n)M_n$ ,

$$\begin{aligned} & \frac{1}{\sqrt{n}}[Z'_n(\tilde{\rho}_n)P_{K,n}R_n(\tilde{\rho}_n)u_n - E(\zeta'_n R'_n P_{K,n}\epsilon_n)] \\ &= \frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)\epsilon_n - \frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\epsilon_n + \frac{1}{\sqrt{n}}[\zeta'_n R'_n P_{K,n}\epsilon_n - E(\zeta'_n R'_n P_{K,n}\epsilon_n)] \\ & \quad + \frac{1}{n}Z'_n M'_n P_{K,n}\epsilon_n \sqrt{n}(\rho_0 - \tilde{\rho}_n) + \frac{1}{n}Z'_n(\rho_0)P_{K,n}M_n R_n^{-1}\epsilon_n \sqrt{n}(\rho_0 - \tilde{\rho}_n) \\ & \quad + \frac{1}{n}Z'_n M'_n P_{K,n}M_n R_n^{-1}\epsilon_n \sqrt{n}(\rho_0 - \tilde{\rho}_n)^2. \end{aligned}$$

The terms on the right hand side have the following properties:

- (1) By Lemma 8 (iv),  $\frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)\epsilon_n = O_P(1)$  and  $\frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)(I_n - P_{K,n})\epsilon_n = O_P(\Delta_{nK,1}^{1/2}) = o_P(1)$ ;
- (2) by Lemma 8 (viii),  $\frac{1}{\sqrt{n}}[\zeta'_n R'_n P_{K,n}\epsilon_n - E(\zeta'_n R'_n P_{K,n}\epsilon_n)] = O_P(\sqrt{K}/n)$ ;
- (3) by Lemma 8 (vi),  $\frac{1}{n}Z'_n M'_n P_{K,n}\epsilon_n = O_P(n^{-1/2}) + O_P(\sqrt{\Delta_{nK,2}/n}) + O_P(K/n)$ ;
- (4) by Lemma 8 (v),  $\frac{1}{n}Z'_n(\rho_0)P_{K,n}M_n R_n^{-1}\epsilon_n = O_P(n^{-1/2}) + O_P(\sqrt{\Delta_{nK,1}/n}) + O_P(K/n)$ ;
- (5) by Lemma 8 (vii),  $\frac{1}{n}Z'_n M'_n P_{K,n}M_n R_n^{-1}\epsilon_n = O_P(n^{-1/2}) + O_P(\sqrt{\Delta_{nK,2}/n}) + O_P(K/n)$ .

Therefore,

- (e) if  $K/n \rightarrow c \neq 0$ ,  $\frac{1}{n}[Z'_n(\tilde{\rho}_n)P_{K,n}R_n(\tilde{\rho}_n)u_n - E(\zeta'_n R'_n P_{K,n}\epsilon_n)] = o_P(1)$ , and  $\hat{\delta}_{2sls,n} - \delta_0 - \bar{b}_{nK,1} \xrightarrow{p} 0$ ;
- (f) if  $K/n \rightarrow 0$ ,  $\frac{1}{\sqrt{n}}[Z'_n(\tilde{\rho}_n)P_{K,n}R_n(\tilde{\rho}_n)u_n - E(\zeta'_n R'_n P_{K,n}\epsilon_n)] = \frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)\epsilon_n + o_P(1)$ , and  $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0 - b_{n,K}) \xrightarrow{d} N(0, \sigma_\epsilon^2 \bar{H}^{-1})$  by Lemma 6. □

**Proof of Proposition 4.** By Proposition 3, it is sufficient to show that  $\sqrt{n}(\tilde{b}_{n,K} - b_{n,K}) = o_P(1)$ . Furthermore, as  $\frac{1}{n}Z'_n(\tilde{\rho}_n)P_{K,n}Z_n(\tilde{\rho}_n) = O_P(1)$ , we only need to show that

$$\frac{1}{\sqrt{n}}\tilde{\Upsilon}_n(K) - \frac{1}{\sqrt{n}}\Upsilon_n(K) = o_P(1)$$

By Lemma 4, as  $\tilde{\rho}_n - \rho_0 = O_P(n^{-1/2})$ ,

$$\frac{1}{\sqrt{n}}[\text{tr}(\tilde{\Gamma}_{nK,1}) - \text{tr}(\Gamma_{nK,1})] = \frac{1}{\sqrt{n}}(\rho_0 - \tilde{\rho}_n) \text{tr}(P_{K,n}M_n) = O_P(K/n) = o_P(1)$$

By the mean value theorem,

$$\begin{aligned} \frac{1}{\sqrt{n}}[\text{tr}(\tilde{\Gamma}_{nK,2}) - \text{tr}(\Gamma_{nK,2})] &= \frac{1}{\sqrt{n}}(\rho_0 - \tilde{\rho}_n) \text{tr}[P_{K,n}M_nG_n(\ddot{\lambda}_n)] + \frac{1}{\sqrt{n}}(\tilde{\lambda}_n - \lambda_0) \text{tr}[P_{K,n}R_n(\ddot{\rho}_n)G_n^2(\ddot{\lambda}_n)] \\ &= O_P(K/n) \end{aligned}$$

as  $\tilde{\rho}_n - \rho_0 = O_P(n^{-1/2})$ ,  $\tilde{\lambda}_n - \lambda_0 = O_P(n^{-1/2})$ , and  $G_n(\ddot{\lambda}_n)$  is UB in probability by Lemma 10, where  $\ddot{\lambda}_n$  is between  $\lambda_0$  and  $\tilde{\lambda}_n$ , and  $\ddot{\rho}_n$  is between  $\rho_0$  and  $\tilde{\rho}_n$ . Similarly,

$$\frac{1}{\sqrt{n}}[\text{tr}(\tilde{\Gamma}_{nK,3}) - \text{tr}(\Gamma_{nK,3})] = O_P(K/n) = o_P(1)$$

By Lemma 11,  $\tilde{\sigma}_\epsilon^2 - \sigma_\epsilon^2 = O_P(n^{-1/2})$  and  $\tilde{\sigma}_{v\epsilon} - \sigma_{v\epsilon} = O_P(n^{-1/2})$ . Furthermore,  $\text{tr}(\Gamma_{nK,1}) = O(K)$ ,  $\text{tr}(\Gamma_{nK,2}) = O(K)$  and  $\text{tr}(\Gamma_{nK,3}) = O(K)$ . Then

$$\frac{1}{\sqrt{n}}[\text{tr}(\tilde{\Gamma}_{nK,2})\tilde{\sigma}_{v\epsilon}\tilde{\gamma}_n + \tilde{\sigma}_\epsilon^2 \text{tr}(\tilde{\Gamma}_{nK,3}), \text{tr}(\tilde{\Gamma}_{nK,1})\tilde{\sigma}_{v\epsilon}] - \frac{1}{\sqrt{n}}[\text{tr}(\Gamma_{nK,2})\sigma_{v\epsilon}\gamma_0 + \sigma_\epsilon^2 \text{tr}(\Gamma_{nK,3}), \text{tr}(\Gamma_{nK,1})\sigma_{v\epsilon}] = o_P(1)$$

and the result in the proposition holds. □

**Proof of Proposition 5.** We find a decomposition for  $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0)$  as in Lemma 2 and show that all the conditions in the lemma are satisfied.

Let  $\rho_{K,n} = \text{tr}(S_n(K))$ , where  $S(K)$  is in (16). We first establish some order properties for  $\rho_{K,n}$ . The  $\rho_{K,n}$  is equal to

$$\begin{aligned} \rho_{K,n} &= \frac{\sigma_\epsilon^2}{n} \text{tr}[(I_n - P_{K,n})\bar{Z}_n(\rho_0)H_n^{-1}H_n^{-1}\bar{Z}'_n(\rho_0)(I_n - P_{K,n})] + \frac{1}{n}\Upsilon'_n(K)H_n^{-1}H_n^{-1}\Upsilon_n(K) \\ &\leq \frac{\sigma_\epsilon^2\tau_{H,\max}}{n} \text{tr}[(I_n - P_{K,n})\bar{Z}_n(\rho_0)\bar{Z}'_n(\rho_0)(I_n - P_{K,n})] + \frac{\tau_{H,\max}}{n}\Upsilon'_n(K)\Upsilon_n(K) \end{aligned}$$

where  $\tau_{H,\max}$  is the largest eigenvalue of  $H_n^{-1}$ , which is bounded from above because  $\lim_{n \rightarrow \infty} H_n$  is finite and nonsingular. Furthermore, as  $\Upsilon'_n(K)\Upsilon_n(K) \leq K^2c$  for some constant  $c > 0$  by Lemma 4,  $\rho_{K,n} = O(K^2/n + \Delta_{nK,1})$ . By a similar argument but with a lower bound for  $\rho_{K,n}$  by using the smallest eigenvalue of  $H_n^{-1}$  and by Assumption 9 (i), as  $\sigma_{v\epsilon} \neq 0$ ,  $\lim_{n \rightarrow \infty} \rho_{K,n}/(K^2/n + \Delta_{nK,1}) > c$  for some constant  $c > 0$ . These together mean that  $\rho_{K,n}$  has exactly the same order as  $(K^2/n + \Delta_{nK,1})$ . This order of  $\rho_{K,n}$  together with  $K^2/n \rightarrow 0$  are helpful to determine the orders of the terms in the decomposition of  $\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0)$ .

The  $\hat{\delta}_{2sls,n}$  satisfies

$$\sqrt{n}(\hat{\delta}_{2sls,n} - \delta_0) = \hat{H}_n^{-1}\hat{h}_n$$

where  $\hat{h}_n = \frac{1}{\sqrt{n}}Z'_n(\tilde{\rho}_n)P_{K,n}u_n(\tilde{\rho}_n)$ , and

$$\begin{aligned} \hat{H}_n &= \frac{1}{n}Z'_n(\rho_0)P_{K,n}Z_n(\rho_0) + \frac{1}{n}(\rho_0 - \tilde{\rho}_n)[Z'_n(\rho_0)P_{K,n}M_nZ_n + Z'_nM'_nP_{K,n}Z_n(\rho_0)] \\ &\quad + \frac{1}{n}(\rho_0 - \tilde{\rho}_n)^2Z'_nM'_nP_{K,n}M_nZ_n \end{aligned}$$

because  $Z_n(\tilde{\rho}_n) = Z_n(\rho_0) + (\rho_0 - \tilde{\rho}_n)M_nZ_n$ . By Lemma 8 (i)–(iii) and Lemma 7 (vi),

$$\frac{1}{n}Z'_n(\rho_0)P_{K,n}Z_n(\rho_0) = H_n + T_{1n}^H + T_{2n}^H + T_{3n}^H + T_{4n}^H$$

where  $H_n = O(1)$ ,  $T_{1n}^H = O(\Delta_{nK,1})$ ,  $T_{2n}^H = O_P(n^{-1/2})$ ,  $T_{3n}^H = O_P(K/n) = o_P(\rho_{K,n})$  and  $T_{4n}^H = o_P(K/n + \Delta_{nK,1}) = o_P(\rho_{K,n})$ ;

$$\begin{aligned} \frac{1}{n\sqrt{n}}Z'_n(\rho_0)P_{K,n}M_nZ_n &= \frac{1}{n\sqrt{n}}\bar{Z}'_n(\rho_0)M_n\bar{Z}_n + O(\sqrt{\Delta_{nK,1}\Delta_{nK,2}/n}) + O_P(n^{-1}) + O_P(Kn^{-3/2}) \\ &\quad + o_P(Kn^{-3/2} + n^{-1/2}\Delta_{nK,1}) + O_P(\sqrt{\Delta_{nK,2}/n}) \\ &= \frac{1}{n\sqrt{n}}\bar{Z}'_n(\rho_0)M_n\bar{Z}_n + o_P(\rho_{K,n}) \end{aligned}$$

and  $\frac{1}{n^2}Z'_nM'_nP_{K,n}M_nZ_n = O_P(1/n) + O_P(K/n^2) = o_P(\rho_{K,n})$ . As

$$\sqrt{n}(\tilde{\rho}_n - \rho_0) = \frac{1}{\sqrt{n}}(\epsilon'_n\mathbb{D}_n\epsilon_n + \mathbb{F}_n\epsilon_n) + O_P(n^{-1/2}) = O_P(1)$$

it follows that

$$\frac{1}{n}(\rho_0 - \tilde{\rho}_n)[Z'_n(\rho_0)P_{K,n}M_nZ_n + Z'_nM'_nP_{K,n}Z_n(\rho_0)] = T_{5n}^H + o_P(\rho_{K,n})$$

where  $T_{5n}^H = -\frac{1}{n^2}(\epsilon'_n\mathbb{D}_n\epsilon_n + \mathbb{F}_n\epsilon_n)[\bar{Z}'_n(\rho_0)M_n\bar{Z}_n + \bar{Z}'_nM'_n\bar{Z}_n(\rho_0)] = O_P(n^{-1/2})$ . Then,  $\hat{H}_n = H_n + T_n^H + o_P(\rho_{K,n})$  with  $T_n^H = T_{1n}^H + T_{2n}^H + T_{5n}^H$ .

For  $\hat{h}_n$ , we have

$$\hat{h}_n = \frac{1}{\sqrt{n}}Z'_n(\rho_0)P_{K,n}\epsilon_n + \frac{1}{\sqrt{n}}(\rho_0 - \tilde{\rho}_n)[Z'_nM'_nP_{K,n}\epsilon_n + Z'_n(\rho_0)P_{K,n}M_nu_n] + \frac{1}{\sqrt{n}}(\rho_0 - \tilde{\rho}_n)^2Z'_nM'_nP_{K,n}M_nu_n,$$

where, by Lemma 8 (iv)–(vii) and Lemma 7 (vi),

- (a)  $\frac{1}{\sqrt{n}}Z'_n(\rho_0)P_{K,n}\epsilon_n = h_n + T_{1n}^h + T_{2n}^h$  with  $h_n = O_P(1)$ ,  $T_{1n}^h = O_P(\Delta_{nK,1}^{1/2})$  and  $T_{2n}^h = O_P(K/\sqrt{n})$ ,
- (b)  $\frac{1}{n}Z'_nM'_nP_{K,n}\epsilon_n = \frac{1}{n}\bar{Z}'_nM'_nP_{K,n}\epsilon_n + O_P(K/n) = O_P(n^{-1/2})$ ,
- (c)  $\frac{1}{n}Z'_n(\rho_0)P_{K,n}M_nu_n = \frac{1}{n}\bar{Z}'_n(\rho_0)M_nu_n + o_P(\rho_{K,n})$ , and
- (d)  $n^{-3/2}Z'_nM'_nP_{K,n}M_nu_n = o_P(\rho_{K,n})$ .

Thus,  $\hat{h}_n = h_n + T_n^h + o_P(\rho_{K,n})$ , where  $T_n^h = T_{1n}^h + T_{2n}^h + T_{3n}^h + T_{4n}^h$  with

$$T_{3n}^h = -n^{-3/2}(\epsilon'_n\mathbb{D}_n\epsilon_n + \mathbb{F}_n\epsilon_n)\bar{Z}'_n(\rho_0)M_nu_n = O_P(1/\sqrt{n})$$

and

$$T_{4n}^h = -n^{-3/2}(\epsilon'_n\mathbb{D}_n\epsilon_n + \mathbb{F}_n\epsilon_n)\bar{Z}'_nM'_nP_{K,n}\epsilon_n = O_P(1/\sqrt{n})$$

Corresponding to the terms of the decomposition in Lemma 2, we have  $Z_n^h = \hat{h}_n - h_n - T_n^h$ ,  $Z_n^H = \hat{H}_n - H_n - T_n^H$ ,

$$\hat{A}_n(K) = (h_n + T_{1n}^h + T_{2n}^h)(h_n + T_{1n}^h + T_{2n}^h)' + [(T_{3n}^h + T_{4n}^h)h_n']^s - (h_nh_n'H_n^{-1}T_n^{H'})^s$$

and

$$Z_n^A(K) = [(T_{1n}^h + T_{2n}^h)(T_{3n}^h + T_{4n}^h)']^s + (T_{3n}^h + T_{4n}^h)(T_{3n}^h + T_{4n}^h)' = o_P(\rho_{K,n})$$

We shall check that all conditions in Lemma 2 are satisfied and derive the explicit expression for  $E[\hat{A}_n(K)]$ . As  $h_n + T_{1n}^h + T_{2n}^h = \frac{1}{\sqrt{n}}\bar{Z}'_n(\rho_0)P_{K,n}\epsilon_n + \frac{1}{\sqrt{n}}\zeta'_nR'_nP_{K,n}\epsilon_n$ , then under the assumption that  $\mu_3 = E(\epsilon_{ni}^2v_{ni}) = 0$ , we have

$$E[(h_n + T_{1n}^h + T_{2n}^h)(h_n + T_{1n}^h + T_{2n}^h)'] = \frac{\sigma_\epsilon^2}{n}\bar{Z}'_n(\rho_0)P_{K,n}\bar{Z}_n(\rho_0) + \frac{1}{n}E(\zeta'_nR'_nP_{K,n}\epsilon_n\epsilon'_nP_{K,n}R_n\zeta_n)$$

Since  $\zeta_n = [G_n v_n \gamma_0 + G_n R_n^{-1} \epsilon_n, v_n]$ , the matrix  $E(\zeta_n' R_n' P_{K,n} \epsilon_n \epsilon_n' P_{K,n} R_n \zeta_n) - \Omega_{n2}(K)$ , where

$$\Omega_{n2}(K) = E(\zeta_n' R_n' P_{K,n} \epsilon_n) E(\epsilon_n' P_{K,n} R_n \zeta_n)$$

can be expanded as a  $4 \times 4$  block matrix, with each block being of the order  $O(K)$  by Lemmas 3 and 4. Thus,

$$E(\zeta_n' R_n' P_{K,n} \epsilon_n \epsilon_n' P_{K,n} R_n \zeta_n) = \Omega_{n2}(K) + O(K)$$

Then,

$$E[(h_n + T_{1n}^h + T_{2n}^h)(h_n + T_{1n}^h + T_{2n}^h)'] = \frac{\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0) P_{K,n} \bar{Z}_n(\rho_0) + \Omega_{n2}(K) + o_P(\rho_{K,n})$$

Note that

$$E(T_{3n}^h h'_n) = -\frac{1}{n^2} \bar{Z}'_n(\rho_0) M_n R_n^{-1} E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n'] \bar{Z}_n(\rho_0)$$

and

$$E(T_{4n}^h h'_n) = -\frac{1}{n^2} \bar{Z}'_n M_n' P_{K,n} E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n'] \bar{Z}_n(\rho_0)$$

where  $E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n']$  is UB by Lemma 9 (i), then  $E(T_{3n}^h h'_n) = O(1/n) = o(\rho_{K,n})$ , and

$$\begin{aligned} & E(T_{4n}^h h'_n) \\ &= -\frac{1}{n^2} \bar{Z}'_n M_n' E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n'] \bar{Z}_n(\rho_0) + \frac{1}{n^2} \bar{Z}'_n M_n' (I_n - P_{K,n}) E[(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n) \epsilon_n \epsilon_n'] \bar{Z}_n(\rho_0) \\ &= O(1/n) + O(\sqrt{\Delta_{nK,2}}/n) \\ &= o(\rho_{K,n}) \end{aligned}$$

by Lemma 7 (ii). As  $E(h_n h'_n H_n^{-1} T_{1n}^{H'}) = -\frac{\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0) (I_n - P_{K,n}) \bar{Z}_n(\rho_0)$ , and, by Lemma 9,  $E(h_n h'_n H_n^{-1} T_{2n}^{H'}) = O(1/n) = o_P(\rho_{K,n})$  and  $E(h_n h'_n H_n^{-1} T_{5n}^{H'}) = O(1/n) = o_P(\rho_{K,n})$ , we have

$$\begin{aligned} E[\hat{A}_n(K)] &= \frac{\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0) P_{K,n} \bar{Z}_n(\rho_0) + \frac{1}{n} \Omega_{n2}(K) + \frac{2\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0) (I_n - P_{K,n}) \bar{Z}_n(\rho_0) + o_P(\rho_{K,n}) \\ &= \sigma_\epsilon^2 H_n + \frac{\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0) (I_n - P_{K,n}) \bar{Z}_n(\rho_0) + \frac{1}{n} \Omega_{n2}(K) + o_P(\rho_{K,n}) \end{aligned}$$

Let  $S_n(K)$  be given by (16), then all conditions of Lemma 2 are satisfied. □

**Proof of Proposition 6.** The proof follows by modifying that of Proposition 5. Now  $\rho_{K,n} = \text{tr}(S_n(K)) = O(K/n + \Delta_{nK,1})$ . The  $\hat{\delta}_{c2sls,n}$  satisfies

$$\sqrt{n}(\hat{\delta}_{c2sls,n} - \delta_0) = \hat{H}_n^{-1} \hat{h}_n \text{ with } \hat{H}_n = \frac{1}{n} Z'_n(\tilde{\rho}_n) P_{K,n} Z_n(\tilde{\rho}_n) \text{ and } \hat{h}_n = \frac{1}{\sqrt{n}} [Z'_n(\tilde{\rho}_n) P_{K,n} u_n(\tilde{\rho}_n) - \tilde{Y}_n(K)]$$

By Lemma 8 (viii),  $T_{3n}^H - E(T_{3n}^H) = o_P(K/n)$ , where  $T_{3n}^H = \frac{1}{n} \zeta_n' R_n' P_{K,n} R_n \zeta_n$  is defined in Lemma 8 and  $E(T_{3n}^H) = \frac{1}{n} \Omega_{n1}(K)$  with  $\Omega_{n1}(K)$  given in (11).

Define  $T_{6n}^H = E(T_{3n}^H)$ . Then, from the proof of Proposition 5,

$$\hat{H}_n = H_n + T_n^H + o_P(\rho_{K,n}), \text{ where } T_n^H = T_{1n}^H + T_{2n}^H + T_{5n}^H + T_{6n}^H$$

For  $\hat{h}_n$ , we have

$$\hat{h}_n = h_n + T_{1n}^h + T_{5n}^h - \frac{1}{\sqrt{n}}[\tilde{\Upsilon}_n(K) - \Upsilon_n(K)] + \frac{1}{\sqrt{n}}(\rho_0 - \tilde{\rho}_n)[Z_n' M_n' P_{K,n} \epsilon_n + Z_n'(\rho_0) P_{K,n} M_n u_n] + \frac{1}{\sqrt{n}}(\rho_0 - \tilde{\rho}_n)^2 Z_n' M_n' P_{K,n} M_n u_n$$

where  $T_{5n}^h = \frac{1}{\sqrt{n}}[\zeta_n' R_n' P_{K,n} \epsilon_n - E(\zeta_n' R_n' P_{K,n} \epsilon_n)] = O_P(\sqrt{K/n})$  by Lemma 8 (viii). By Lemma 11 (iv),

$$-\frac{1}{\sqrt{n}}[\tilde{\Upsilon}_n(K) - \Upsilon_n(K)] = T_{6n}^h + o_P(\rho_{K,n}) \text{ with } T_{6n}^h = -\frac{1}{n}(a_1', a_2')' = O_P(K/n)$$

where  $a_1$  and  $a_2$  are defined in Lemma 11 (iv). By Lemma 8 (v)–(viii),  $\frac{1}{n} Z_n' M_n' P_{K,n} \epsilon_n = \frac{1}{n} \bar{Z}_n' M_n' P_{K,n} \epsilon_n + \frac{1}{n} E(\zeta_n' M_n' P_{K,n} \epsilon_n) + o_P(\rho_{K,n})$ ,

$$\frac{1}{n} Z_n'(\rho_0) P_{K,n} M_n u_n = \frac{1}{n} \bar{Z}_n'(\rho_0) M_n u_n + \frac{1}{n} E(\zeta_n' R_n' P_{K,n} M_n u_n) + o_P(\rho_{K,n})$$

and  $\frac{1}{n^{3/2}} Z_n' M_n' P_{K,n} M_n u_n = o_P(\rho_{K,n})$ . Therefore,

$$\hat{h}_n = h_n + T_n^h + o_P(\rho_{K,n}), \text{ where } T_n^h = T_{1n}^h + T_{5n}^h + T_{6n}^h + T_{3n}^h + T_{4n}^h + T_{7n}^h$$

with  $T_{3n}^h$  and  $T_{4n}^h$  defined in the proof of Proposition 5; and

$$T_{7n}^h = -n^{-3/2}(\epsilon_n' \mathbb{D}_n \epsilon_n + \mathbb{F}_n \epsilon_n)[E(\zeta_n' M_n' P_{K,n} \epsilon_n) + E(\zeta_n' R_n' P_{K,n} M_n u_n)] = O_P(K/n)$$

For the decomposition in Lemma 2, take  $Z_n^A(K) = (h_n + T_n^h)(h_n + T_n^h)' - (h_n h_n' H_n^{-1} T_n^{H'})^s - \hat{A}_n(K)$ , and

$$\hat{A}_n(K) = (h_n + T_{1n}^h)(h_n + T_{1n}^h)' + [h_n(T_{3n}^h + T_{4n}^h + T_{5n}^h + T_{6n}^h + T_{7n}^h)]^s + (T_{1n}^h T_{5n}^h)' + T_{5n}^h T_{5n}^h - (h_n h_n' H_n^{-1} T_n^{H'})^s$$

Then  $Z_n^A(K) = o_P(\rho_{K,n})$ . To check that the conditions in Lemma 2 are satisfied, we now investigate  $E(\hat{A}_n(K))$ . First,

$$E[(h_n + T_{1n}^h)(h_n + T_{1n}^h)'] = \frac{1}{n} E(\bar{Z}_n'(\rho_0) P_{K,n} \epsilon_n \epsilon_n' P_{K,n} \bar{Z}_n(\rho_0)) = \frac{\sigma_\epsilon^2}{n} \bar{Z}_n'(\rho_0) P_{K,n} \bar{Z}_n(\rho_0)$$

By the proof of Proposition 5,  $E[h_n(T_{3n}^h + T_{4n}^h)] = o_P(\rho_{K,n})$ . Under the assumption that  $E(\epsilon_{ni}^3) = E(\epsilon_{ni}^2 v_{ni}) = 0$ , we have  $E(h_n T_{5n}^h) = 0$ ,  $E(T_{1n}^h T_{5n}^h) = 0$ ,

$$E(h_n T_{7n}^h) = -\frac{\sigma_\epsilon^2}{n^2} \bar{Z}_n'(\rho_0) \mathbb{F}_n' [E(\zeta_n' M_n' P_{K,n} \epsilon_n) + E(\zeta_n' R_n' P_{K,n} M_n u_n)]'$$

and

$$E(h_n T_{6n}^h) = -\frac{1}{n} [E(h_n a_1'), E(h_n a_2')] = -\frac{1}{n} [\Pi_{n2,1}(K), \Pi_{n2,2}(K)]$$

where  $\Pi_{n2,1}(K)$  and  $\Pi_{n2,2}(K)$  are given in (23) and (24) respectively. The expression for  $E(T_{5n}^h T_{5n}^h)$  can be derived by Lemma 3. Under Assumption 9 (ii),  $\text{vec}_D'(\Gamma_{nK,i}) \text{vec}_D(\Gamma_{nK,j}) = o(K)$  for  $i, j = 1, 2, 3$ . Then  $E(T_{5n}^h T_{5n}^h) = \frac{1}{n} \Pi_{n1}(K) + o(K/n)$ , where  $\Pi_{n1}(K)$  is given in (21). By the proof of Proposition 5,  $E[h_n h_n' H_n^{-1} (T_{2n}^H + T_{5n}^H)'] = o_P(\rho_{K,n})$ . Furthermore,

$$E[h_n h_n' H_n^{-1} (T_{1n}^H + T_{6n}^H)'] = -\frac{\sigma_\epsilon^2}{n} \bar{Z}_n'(\rho_0) (I_n - P_{K,n}) \bar{Z}_n(\rho_0) + \sigma_\epsilon^2 T_{6n}^H$$



Therefore,

$$E(\hat{A}_n(K)) = \sigma_\epsilon^2 H_n + \frac{\sigma_\epsilon^2}{n} \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) + \frac{1}{n} \Pi_{n1}(K) + \frac{1}{n} \Pi_{n2}(K) + \frac{1}{n} \Pi_{n3}(K) + o_P(\rho_{K,n})$$

where  $\frac{1}{n} \Pi_{n2}(K) = E[(h_n T_{6n}^{h'})^s] - 2\sigma_\epsilon^2 T_{6n}^H$  and  $\frac{1}{n} \Pi_{n3}(K) = E[(h_n T_{7n}^{h'})^s]$ . Let  $S_n(K)$  be given by (17), then all conditions in Lemma 2 are satisfied and the result in the proposition holds.  $\square$

**Proof of Proposition 7.** As  $Z_n(\hat{\rho}_n) = Z_n(\rho_0) + (\rho_0 - \hat{\rho}_n)M_n Z_n$  and  $Z_n = \bar{Z}_n + \zeta_n$ , we have

$$\begin{aligned} Z'_n(\hat{\rho}_n)(I_n - P_{K,n})Z_n(\hat{\rho}_n) &= Z'_n(\rho_0)(I_n - P_{K,n})Z_n(\rho_0) + (\rho_0 - \hat{\rho}_n)[Z'_n M'_n(I_n - P_{K,n})Z_n(\rho_0)]^s \\ &\quad + (\rho_0 - \hat{\rho}_n)^2 Z'_n M'_n(I_n - P_{K,n})M_n Z_n \end{aligned}$$

where

$$\begin{aligned} &Z'_n(\rho_0)(I_n - P_{K,n})Z_n(\rho_0) \\ &= \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) + [\bar{Z}'_n(\rho_0)(I_n - P_{K,n})R_n \zeta_n]^s + \zeta'_n R'_n R_n \zeta_n - \zeta'_n R'_n P_{K,n} R_n \zeta_n \end{aligned}$$

Let

$$\tilde{S}_{n,\xi}(K) = \hat{S}_{n,\xi}(K) - \frac{\hat{\sigma}_\epsilon^2}{n} \xi' \hat{H}_n^{-1} [\zeta'_n R'_n R_n \zeta_n + (\rho_0 - \hat{\rho}_n)(\zeta'_n M'_n R_n \zeta_n)^s + (\rho_0 - \hat{\rho}_n)^2 \zeta'_n M'_n M_n \zeta_n] \hat{H}_n^{-1} \xi$$

As  $[\hat{S}_{n,\xi}(K) - \tilde{S}_{n,\xi}(K)]$  does not depend on  $K$ ,  $\arg \min_K \hat{S}_{n,\xi}(K) = \arg \min_K \tilde{S}_{n,\xi}(K)$ . By Lemma 12, we only need to show that  $\sup_K \frac{|\tilde{S}_{n,\xi}(K) - S_{n,\xi}(K)|}{S_{n,\xi}(K)} \xrightarrow{p} 0$ .

Let  $e_i$  be the  $i$ th column of the  $(m + 1) \times (m + 1)$  identity matrix. Since  $\hat{\sigma}_\epsilon^2 = \sigma_\epsilon^2 + o_P(1)$  and  $\hat{H}_n = H_n + o_P(1) = O_P(1)$ , for the GS2SLS, by the triangular inequality, it is sufficient to show the following:

- (i)  $\sup_K |e'_i \Omega_{n2}(K) e_j| / [n S_{n,\xi}(K)] < c$  for some constant  $c > 0$  and  $\sup_K |e'_i [\hat{\Omega}_{n2}(K) - \Omega_{n2}(K)] e_j| / [n S_{n,\xi}(K)] \rightarrow 0$ ;
- (ii)  $\sup_K |e'_i \bar{Z}'_n(\rho_0)(I_n - P_{K,n})\bar{Z}_n(\rho_0) e_j| / [n S_{n,\xi}(K)] \rightarrow 0$ ;
- (iii)  $\sup_K |e'_i \Omega_{n1}(K) e_j| / [n S_{n,\xi}(K)] < c$  for some constant  $c > 0$ ,  $\sup_K |e'_i [\hat{\Omega}_{n1}(K) - \Omega_{n1}(K)] e_j| / [n S_{n,\xi}(K)] \xrightarrow{p} 0$  and  $\sup_K |e'_i [\zeta'_n R'_n P_{K,n} R_n \zeta_n - \Omega_{n1}(K)] e_j| / [n S_{n,\xi}(K)] \xrightarrow{p} 0$ ;
- (iv)  $\sup_K |e'_i \bar{Z}'_n(\rho_0)(I_n - P_{K,n})R_n \zeta_n e_j| / [n S_{n,\xi}(K)] \xrightarrow{p} 0$ , and
- (v)  $\sup_K |e'_i \{(\rho_0 - \hat{\rho}_n)[Z'_n M'_n(I_n - P_{K,n})Z_n(\rho_0) - \zeta'_n M'_n R_n \zeta_n]^s + (\rho_0 - \hat{\rho}_n)^2 [Z'_n M'_n(I_n - P_{K,n})M_n Z_n - \zeta'_n M'_n M_n \zeta_n]\} e_j| / [n S_{n,\xi}(K)] \xrightarrow{p} 0$ .

For the CGS2SLS, we need to show (ii)–(v) and

- (i')  $\sup_K |e'_i [\Pi_{n1}(K) + \Pi_{n2}(K) + \Pi_{n3}(K)] e_j| / [n S_{n,\xi}(K)] < c$  for some constant  $c > 0$  and  $\sup_K |e'_i [\hat{\Pi}_{n1}(K) + \hat{\Pi}_{n2}(K) + \hat{\Pi}_{n3}(K) - \Pi_{n1}(K) - \Pi_{n2}(K) - \Pi_{n3}(K)] e_j| / [n S_{n,\xi}(K)] \rightarrow 0$ .

We first show (i) and (i'). By Lemma 4,

$$\sup_K |e'_i \Omega_{n2}(K) e_j| / [n S_{n,\xi}(K)] < c_1 \sup_K K^2 / [n S_{n,\xi}(K)]$$

for some constant  $c_1 > 0$ . By Assumption 10 (ii), for the GS2SLS,  $S_{n,\xi}(K) > K^2 c_2/n$  for some  $c_2 > 0$ . Then  $\sup_K |e'_i \Omega_{n2}(K) e_j| / [n S_{n,\xi}(K)] < c$  for some constant  $c > 0$ . For  $\text{tr}[P_{K,n} R_n(\hat{\rho}_n) G_n(\hat{\lambda}_n)]$ , by the mean value theorem,

$$\begin{aligned} & |\text{tr}[P_{K,n} R_n(\hat{\rho}_n) G_n(\hat{\lambda}_n)] - \text{tr}(\Gamma_{nK,2})| \\ &= |(\hat{\lambda}_n - \lambda_0) \text{tr}[P_{K,n} R_n(\hat{\rho}_n) G_n^2(\hat{\lambda}_n)] - (\hat{\rho}_n - \rho_0) \text{tr}[P_{K,n} M_n G_n(\hat{\lambda}_n)]| \\ &\leq Kc \sqrt{(\hat{\lambda}_n - \lambda_0)^2 + (\hat{\rho}_n - \rho_0)^2} \end{aligned}$$

in probability for some constant  $c > 0$ , by Lemmas 10 and 4. As all parameter estimates used in  $\hat{\Omega}_{n2}(K)$  are consistent, applying similarly the mean value theorem to other terms in  $\hat{\Omega}_{n2}(K)$ , we can see that  $\|\hat{\Omega}_{n2}(K) - \Omega_{n2}(K)\| \leq K^2 c_n$  in probability, where  $c_n = o_P(1)$  does not depend on  $K$ . Thus (i) holds. For the CGS2SLS, (i') holds similarly.

The (ii) holds because  $|e'_i \bar{Z}'_n(\rho_0)(I_n - P_{K,n}) \bar{Z}_n(\rho_0) e_j| \leq c_1 \Delta_{nK,1}$  for some  $c_1 > 0$ , and  $n S_{n,\xi}(K) > c \Delta_{nK,1}$  for some constant  $c > 0$  by Assumption 10 (ii).

For (iii), the first two results are similar to those in (i), thus we only show that

$$\sup_K |e'_i [\zeta'_n R'_n P_{K,n} R_n \zeta_n - \Omega_{n1}(K)] e_j| / [n S_{n,\xi}(K)] \xrightarrow{P} 0$$

By Chebyshev's inequality, for any  $\eta > 0$ ,

$$\begin{aligned} & P(\sup_K |e'_i [\zeta'_n R'_n P_{K,n} R_n \zeta_n - \Omega_{n1}(K)] e_j| / [n S_{n,\xi}(K)] \geq \eta) \\ &\leq \sum_K E\{e'_i [\zeta'_n R'_n P_{K,n} R_n \zeta_n - \Omega_{n1}(K)] e_j e'_j [\zeta'_n R'_n P_{K,n} R_n \zeta_n - \Omega_{n1}(K)] e_i\} / [\eta^2 n^2 S_{n,\xi}^2(K)] \\ &\leq \sum_K E\{e'_i [\zeta'_n R'_n P_{K,n} R_n \zeta_n - \Omega_{n1}(K)] [\zeta'_n R'_n P_{K,n} R_n \zeta_n - \Omega_{n1}(K)] e_i\} / [\eta^2 n^2 S_{n,\xi}^2(K)] \\ &\leq \sum_K Kc / [\eta^2 n^2 S_{n,\xi}^2(K)] \leq \sum_K c_1 / [\eta^2 n S_{n,\xi}(K)] \end{aligned}$$

for some constants  $c > 0$  and  $c_1 > 0$ , where the third inequality follows by Lemmas 3 and 4, and the last inequality holds since  $n S_{n,\xi}(K) \geq K c_2$  for some constant  $c_2 > 0$  by Assumption 10 (ii). The result then follows by Assumption 11.

For (iv), by Chebyshev's inequality, for any  $\eta > 0$ ,

$$\begin{aligned} & P(\sup_K |e'_i \bar{Z}'_n(\rho_0)(I_n - P_{K,n}) R_n \zeta_n e_j| / [n S_{n,\xi}(K)] > \eta) \\ &\leq \sum_K E[e'_i \bar{Z}'_n(\rho_0)(I_n - P_{K,n}) R_n \zeta_n e_j e'_j \zeta'_n R'_n (I_n - P_{K,n}) \bar{Z}_n(\rho_0)] e_i / [\eta^2 n^2 S_{n,\xi}^2(K)] \\ &\leq \sum_K e'_i \bar{Z}'_n(\rho_0)(I_n - P_{K,n}) R_n E(\zeta_n \zeta'_n) R'_n (I_n - P_{K,n}) \bar{Z}_n(\rho_0) e_i / [\eta^2 n^2 S_{n,\xi}^2(K)] \\ &\leq \sum_K \tau_{\zeta, \max} e'_i \bar{Z}'_n(\rho_0)(I_n - P_{K,n}) \bar{Z}_n(\rho_0) e_i / [\eta^2 n^2 S_{n,\xi}^2(K)] \\ &\leq \sum_K c / [\eta^2 n S_{n,\xi}(K)] \end{aligned}$$

for some constant  $c$ , where  $\tau_{\zeta, \max}$  denotes the largest eigenvalue of  $R'_n E(\zeta_n \zeta'_n) R_n$ , and the last inequality holds because  $R'_n E(\zeta_n \zeta'_n) R_n$  is UB and  $S_{n,\xi}(K) > c_1 \Delta_{nK,1}$  for some  $c_1 > 0$ . Thus the result holds.

For (v), as  $Z_n = \bar{Z}_n + \zeta_n$  and  $\sqrt{n}(\hat{\rho}_n - \rho_0) = O_P(1)$ , we show the following:

- (1)  $\sup_K |e'_i \bar{Z}'_n M'_n (I_n - P_{K,n}) \bar{Z}_n(\rho_0) e_j| / [n\sqrt{n} S_{n,\xi}(K)] \rightarrow 0$  and  $\sup_K |e'_i \bar{Z}'_n M'_n (I_n - P_{K,n}) M_n \bar{Z}_n e_j| / [n^2 S_{n,\xi}(K)] \rightarrow 0$ ;
- (2)  $\sup_K |e'_i A_n \zeta_n e_j| / [n\sqrt{n} S_{n,\xi}(K)] \xrightarrow{p} 0$ , where  $A_n = \bar{Z}'_n M'_n (I_n - P_{K,n}) R_n$ ,  $\bar{Z}'_n(\rho_0) (I_n - P_{K,n}) M_n$  or  $\bar{Z}'_n M'_n (I_n - P_{K,n}) M_n$ ;
- (3)  $\sup_K |e'_i E[\zeta'_n M'_n P_{K,n} R_n \zeta_n] e_j| / [n\sqrt{n} S_{n,\xi}(K)] \rightarrow 0$ ,  $\sup_K |e'_i E[\zeta'_n M'_n P_{K,n} M_n \zeta_n] e_j| / [n\sqrt{n} S_{n,\xi}(K)] \rightarrow 0$ ,  $\sup_K |e'_i \{\zeta'_n M'_n P_{K,n} R_n \zeta_n - E[\zeta'_n M'_n P_{K,n} R_n \zeta_n]\} e_j| / [n\sqrt{n} S_{n,\xi}(K)] \xrightarrow{p} 0$ , and  $\sup_K |e'_i \{\zeta'_n M'_n P_{K,n} M_n \zeta_n - E[\zeta'_n M'_n P_{K,n} M_n \zeta_n]\} e_j| / [n\sqrt{n} S_{n,\xi}(K)] \xrightarrow{p} 0$ .

By Lemma 7 (iii), we have

$$\sup_K |e'_i \bar{Z}'_n M'_n (I_n - P_{K,n}) \bar{Z}_n(\rho_0) e_j| / [n\sqrt{n} S_{n,\xi}(K)] \leq \sup_K c \sqrt{\Delta_{nK,1} / S_{n,\xi}(K)} \sqrt{\Delta_{nK,2} / [n S_{n,\xi}(K)]}$$

for some  $c > 0$ . Since  $\sup_K \Delta_{nK,2} / [n S_{n,\xi}(K)] \leq \sum_K \Delta_{nK,2} / [n S_{n,\xi}(K)]$  and  $\Delta_{nK,2} = o(1)$ , the first result in (1) holds by Assumption 10 (ii). The second result in (1) holds since

$$\sup_K |e'_i \bar{Z}'_n M'_n (I_n - P_{K,n}) M_n \bar{Z}_n e_j| / [n^2 S_{n,\xi}(K)] \leq \sup_K c \Delta_{nK,2} / [n S_{n,\xi}(K)]$$

for some  $c > 0$ . For (2), similar to (iv), for any  $\eta > 0$ , we have

$$P(\sup_K |e'_i \bar{Z}'_n M'_n (I_n - P_{K,n}) R_n \zeta_n e_j| / [n\sqrt{n} S_{n,\xi}(K)] > \eta) \leq c\eta^{-2} \sum_K (\Delta_{nK,2} / [n S_{n,\xi}(K)]) [n S_{n,\xi}(K)]^{-1},$$

$$P(\sup_K |e'_i \bar{Z}'_n(\rho_0) (I_n - P_{K,n}) M_n \zeta_n e_j| / [n\sqrt{n} S_{n,\xi}(K)] > \eta) \leq c\eta^{-2} \sum_K (\Delta_{nK,1} / [n S_{n,\xi}(K)]) [n S_{n,\xi}(K)]^{-1},$$

and

$$P(\sup_K |e'_i \bar{Z}'_n M'_n (I_n - P_{K,n}) M_n \zeta_n e_j| / [n\sqrt{n} S_{n,\xi}(K)] > \eta) \leq c\eta^{-2} \sum_K (\Delta_{nK,2} / [n S_{n,\xi}(K)]) [n S_{n,\xi}(K)]^{-1}$$

for some  $c > 0$ . (3) is similar to (i). □

## References

1. Donald, S.G.; Newey, W.K. Choosing the number of instruments. *Econometrica* **2001**, *69*, 1161–1191.
2. Nagar, A.L. The bias and moment matrix of the general k-class estimators of the parameters in simultaneous equations. *Econometrica* **1959**, *27*, 575–595.
3. Liu, X.; Lee, L.F. Two stage least squares estimation of spatial autoregressive models with endogenous regressors and many instruments. *Econ. Rev.* **2013**, *32*, 734–753.
4. Benirschka, M.; Binkley, J.K. Land price volatility in a geographically dispersed market. *Am. J. Agric. Econ.* **1994**, *76*, 185–195.
5. Anselin, L.; Bera, A. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In *Handbook of Applied Economic Statistics*; Ullah, A., Giles, D.E., Eds.; Marcel Dekker: New York, NY, USA, 1998; pp. 237–289.
6. Case, A. On the use of spatial autoregressive models in demand analysis. Discussion Paper 135, Research Program in Development Studies, Woodrow Wilson School, Princeton University: Princeton, NJ, USA, 1987.

7. Case, A. Spatial patterns in household demand. *Econometrica* **1991**, *59*, 953–965.
8. Case, A. Neighborhood influence and technological change. *Reg. Sci. Urban Econ.* **1992**, *22*, 491–508.
9. Case, A.; Hines, J., Jr.; Rosen, H. Budget spillovers and fiscal policy independence: Evidence from the states. *J. Public Econ.* **1993**, *52*, 285–307.
10. Besley, T.; Case, A. Incumbent behavior: Vote-seeking, tax-setting, and yardstick competition. *Am. Econ. Rev.* **1995**, *85*, 951–963.
11. Kelejian, H.H.; Prucha, I.R. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *J. Real Estate Financ. Econ.* **1998**, *17*, 99–121.
12. Drukker, D.M.; Egger, P.; Prucha, I.R. On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. *Econ. Rev.* **2013**, *32*, 686–733.
13. Kelejian, H.H.; Prucha, I.R. Estimation of simultaneous systems of spatially interrelated cross sectional equations. *J. Econom.* **2004**, *118*, 27–50.
14. Kelejian, H.H.; Prucha, I.R. HAC estimation in a spatial framework. *J. Econom.* **2007**, *140*, 131–154.
15. Lee, L.F.; Yu, J. Efficient GMM estimation of spatial dynamic panel data models with fixed effects. Working paper, 2012.
16. Chamberlain, G. Asymptotic efficiency in estimation with conditional moment restrictions. *J. Econom.* **1987**, *34*, 305–334.
17. Kelejian, H.H.; Prucha, I.R. On the asymptotic distribution of the Moran *I* test statistic with applications. *J. Econom.* **2001**, *104*, 219–257.
18. Kelejian, H.H.; Prucha, I.R. A generalized moments estimator for the autoregressive parameter in a spatial model. *Int. Econ. Rev.* **1999**, *40*, 509–533.
19. Hahn, J.; Hausman, J. A new specification test for the validity of instrumental variables. *Econometrica* **2002**, *70*, 163–189.
20. Anselin, L. *Spatial Econometrics: Methods and Models*; Kluwer Academic Publishers: Boston, MA, USA, 1988.
21. Liu, X.; Lee, L.F. GMM estimation of social interaction models with centrality. *J. Econom.* **2010**, *159*, 99–115.
22. Lee, L.F. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* **2004**, *72*, 1899–1925.
23. White, H. *Estimation, Inference and Specification Analysis*; Cambridge University Press: New York, NY, USA, 1994.