GMM Estimation of Social Interaction Models with Centrality*

Xiaodong Liu
University of Colorado,
Boulder, CO 80309

Lung-fei Lee
Ohio State University,
Columbus, OH 43210

First version, January 2009

Second revision, December 2009

Third revision, April 2010

Abstract

This paper considers the specification and estimation of social interaction models with network structures and the presence of endogenous, contextual, correlated, and group fixed effects. When the network structure in a group is captured by a graph in which the degrees of nodes are not all equal, the different positions of group members as measured by the Bonacich (1987) centrality provide additional information for identification and estimation. In this case, the Bonacich centrality measure for each group can be used as an instrument for the endogenous social effect, but the number of such instruments grows with the number of groups. We consider the 2SLS and GMM estimation for the model. The proposed estimators are asymptotically efficient, respectively, within the class of IV estimators and the class of GMM estimators based on linear and quadratic moments, when the sample size grows fast enough relative to the number of instruments.

JEL classification: C13, C21

Key words: social network, centrality, spatial autoregressive model, GMM, efficiency

*We thank John Chao and Ingmar Prucha for helpful discussions. We are also grateful to the editor, the associate editor, three referees, and seminar participants at the II World Conference of the Spatial Econometrics Association in New York, the Far East and South Asia Meeting of the Econometric Society in Tokyo, Yale University, and University of Maryland for valuable comments and suggestions.

1 Introduction

This paper studies social interaction models with network structures. The model considered has the specification of a spatial autoregressive (SAR) model but has features and implications directly relevant to social interaction issues. With such a specification, the information on network structures is usually summarized in the spatial weights matrix, also known as the sociomatrix (or adjacency matrix), in social interaction models.

A general social interaction model not only allows possible endogenous interactions, but also exogenous interactions, unobserved group effects, and correlation of unobservables. Identification of the endogenous interaction effect from the other effects is a main interest in social interaction models (see, eg., Manski, 1993; Moffitt, 2001). Linear regression models with endogenous interaction based on rational expectations of the group behavior would suffer from the 'reflection problem' of Manski (1993), and the various interaction effects cannot be separately identified. Lee (2007b) considers a group setting where an individual is equally influenced by all the other members in the group and the average outcome of peers represents the source of the endogenous effect. Lee's (2007b) social interaction model is identifiable only if there is variation in group sizes in the sample. The reason for the possible identification is that individuals in a small group will have stronger endogenous interactions than those in a larger group. The identification, however, can be weak if all of the groups have large sizes, even if there is group size variation. The sociomatrix in Lee's (2007b) model has zero diagonal and all of its off-diagonal entries take the value of $\frac{1}{m-1}$, where m is the group size. Such a sociomatrix represents a rather restrictive network structure, but may be practical when there is no information on how individuals interact with each other.

In some data sets, one may have information on network structures. Based on a specific network structure, the (i, j) entry of the sociomatrix is one if i is influenced by j, and zero otherwise. The corresponding sociomatrix represents a directed graph with a directed edge leading from j to i if j affects i.[1] Such a directed-graph sociomatrix has been considered in Lee et al. (2009), where it is row-normalized such that each row sums to unity. For SAR models in empirical studies, row-normalized spatial weights matrices are typical with a few exceptions,[2] because the spatial effect

[1] Note that the influence may or may not be reciprocal, so the sociomatrix could be asymmetric.

[2] An exception is argued in Bell and Bockstael (2000) for real estate problems with micro-level data. Kelejian and Robinson (1995) has argued that the parameter space should be free except some singularity points for the spatial matrix — and discussed not-row-normalized spatial matrices. More recently, Kelejian and Prucha (2007) consider implications on the parameter space of the SAR model when the spatial matrix is not row-normalized.

can be interpreted as a (weighted) average of neighborhood effects. The social interaction models based on expected group means in Manski (1993) and Brock and Durlauf (2001), and the one in Lee (2007b) all have the endogenous effect being an average of peers' outcomes.

The row-normalized sociomatrix in Lee et al. (2009) has some limitations. First, it implicitly rules out the possibility that an individual's outcome might affect peers' outcomes but he/she might not be affected by peers.[3] In addition, for social interaction studies, one may be interested in the aggregate influence of an individual's peers instead of the average influence. One may also be interested in how an individual's position in a network would influence peers' behavior. Notions such as prestige and centrality have received attention in network studies (Wasserman and Faust, 1994). When the social interaction is specified as a SAR model, the measure of centrality in Bonacich (1987) comes out naturally in the reduced form equation. If the sociomatrix represents the directed graph mentioned above, the sum of the ith row is the indegrees (the number of inward directed edges) of the node i in the graph. All group members (nodes) would have the same level of centrality by the Bonacich measure if and only if the indegrees of all nodes are equal. Thus, if the indegrees have a non-zero variation, so does the Bonacich centrality measure for the group members. The variation in the Bonacich centrality measure helps to identify the various interaction effects. Yet, row-normalization would eliminate the variation in the Bonacich centrality measure. So, for social network studies, sometimes a sociomatrix without row-normalization would be appropriate. In this paper, we study the identification and estimation of network effects without requiring row-normalization of the sociomatrix.

Similar to the model in Lee et al. (2009), the social interaction model in this paper has the specification of a SAR model and incorporates endogenous, exogenous, correlated, and unobserved group effects. The unobserved group effect is captured by a group dummy variable, which is allowed to have a conditional mean that depends on the exogenous variables and/or the sociomatrices (due to self-selection), and so it is treated as a fixed effect. With many groups in the sample, the group dummies may induce the incidental parameter problem as in Neyman and Scott (1948). Based on a transformed model that has the group dummies eliminated, Lee et al. (2009) has generalized the ML estimation approach in Lee (2007b) to the network model with a sociomatrix having constant row sums (including the special case of a row-normalized sociomatrix).[4] However, when the sociomatrix

[3] In this case, the corresponding row in the sociomatrix will have all zeros and can not be normalized to sum to unity.

[4] The resulted likelihood function can be shown to be a partial likelihood function under normal disturbances. The

does not have constant row sums, the likelihood function for the transformed model could not be derived, and alternative estimation approaches need to be considered.

This paper considers the 2SLS and generalized method of moments (GMM) estimation approaches. The 2SLS approach has been proposed for the estimation of SAR models in Kelejian and Prucha (1998). The GMM method has been considered for the estimation of a spatial process in Kelejian and Prucha (1999), and SAR models in Lee (2007c) and Lee and Liu (2010). The 2SLS and GMM approaches can be generalized for the estimation of social network models. When the sociomatrix is not row-normalized and the indegrees of its nodes are not all equal, the Bonacich centrality measure for each group can be used as an additional IV to improve estimation efficiency. The number of such instruments depends on the number of groups. If the number of groups grows with the sample size, so does the number of IVs. We show that the proposed 2SLS and GMM estimators can be consistent and asymptotically normal, and they can be efficient when the sample size grows fast enough relative to the number of instruments. We also suggest bias-correction procedures for both estimators based on the estimated leading order many-instrument biases.

Since Bekker's (1994) seminal work, the study of many-instrument asymptotics, where the number of instruments increases with the sample size, has attracted a lot of attention in the IV estimation literature. Some recent developments in this area include Donald and Newey (2001), Hansen et al. (2008), van Hasselt (2010) and Anderson et al. (2007), to name a few. In particular, Bekker and van der Ploeg (2005) has considered IV estimation of a model where group indicators are used as (dummy) instruments and the number of groups goes to infinity. In this paper, we also consider many-group asymptotics, where the number of instruments depends on the number of groups. However, the instruments based centrality measures are not dummy variables. Our model also relaxes the i.i.d. assumption for observations within a group in Bekker and van der Ploeg (2005) by allowing for possible spatial (or social) correlation among group members. Similar to Donald and Newey (2001), we focus on the case where the number of instruments grows with, but at a slower rate than, the sample size.[5] Another important direction of research in the IV estimation literature is on weak instruments or weak identification (see, e.g., Chao and Swanson, 2005 and 2007). In this paper, we

notion of partial likelihood is introduced in Cox (1975); see also Lancaster (2000).

[5]Under the asymptotic sequence that the number of instruments increases at the same rate as the same size, the asymptotic distribution of IV-based estimators has been established by Bekker (1994), Bekker and van der Ploeg (2005), Hansen et al. (2008), and van Hasselt (2010). However, their CLTs assume independent observations and might not be easy to modify for the case with (spatially) correlated observations without imposing strong regularity conditions.

assume the concentration parameter grows at the same rate as the sample size.[6] Hence, we restrict our attention to scenarios where instruments are stronger than assumed in the weak-instrument literature.

The rest of the paper is organized as follows. Section 2 presents the network model and suggests a transformation of the model to eliminate group fixed effects. Sections 3 and 4 propose the 2SLS and GMM approaches for the estimation of the model. We prove consistency of the proposed estimators, derive the asymptotic distributions, and suggest bias correction procedures for the many-instrument bias. The detailed proofs are given in the Appendix. Monte Carlo evidence on the small sample performance of the proposed estimators is given in Section 5. Section 6 briefly concludes.

2 The Network Model with Group Fixed Effects

The model considered has the specification

$$Y_r = \lambda_0 W_r Y_r + X_{1r}\beta_{01} + W_r X_{2r}\beta_{02} + l_{m_r}\alpha_{0r} + u_r, \tag{1}$$

and $u_r = \rho_0 M_r u_r + \epsilon_r$, for $r = 1, \cdots, \bar{r}$, where $\bar{r}$ is the total number of groups in the sample, m_r is the number of individuals in the rth group, and $n = \sum_{r=1}^{\bar{r}} m_r$ is the total number of sample observations. $Y_r = (y_{1r}, \cdots, y_{m_r r})'$ is an m_r-dimensional vector of y_{ir}'s, where y_{ir} is the observed outcome of the ith individual in the group r. W_r and M_r are $m_r \times m_r$ sociomatrices of known constants. In principle, W_r and M_r may or may not be the same.[7] λ_0 captures the endogenous effect, where outcomes of individuals influence those of their successors in the directed graph.[8] X_{1r} and X_{2r} are, respectively, $m_r \times k_1$ and $m_r \times k_2$ matrices of exogenous variables, which may or may not be the same. β_{01} represents the dependence of individuals' outcomes on their own characteristics. On the other hand, outcomes of individuals may also depend on the characteristics of their predecessors via the exogenous contextual effect β_{02}. l_{m_r} is an m_r-dimensional vector of ones and α_{0r} represents the unobserved group-specific effect. Aside from the group fixed effect, ρ_0 captures unobservable

[6] This condition on the concentration parameter is implied by Assumption 4 in Section 3. The assumption of independent observations is omnipresent in the literature of weak instruments. This model allows the observations within a group to be correlated. The analysis of asymptotic properties of IV estimators in the presence of weak instruments in a model with correlated observations is a difficult problem, which is beyond the scope of this paper.

[7] For models with row-normalized spatial matrices, some empirical studies assume $M_r = W_r$, (see, e.g., Cohen, 2002; Fingleton, 2008). Some discussions on the possibility that $M_r \neq W_r$ can be found in LeSage (1999, pp. 87-88). In this paper, we consider the case where W_r is not row-normalized. We think the use of row-normalized W_r as M_r instead of W_r itself might be more relevant in practice; otherwise, the variance of an individual's outcome would increase with the number of friends he has, and might be too large if he has a large number of friends.

[8] In a directed graph, if a edge leads from x to y, then y is said to be a successor of x, and x is said to be a predecessor of y.

correlated effects of individuals with their connections in the network. Whether or not all these various interaction or correlation effects can be identified or estimated is the main interest of the literature of social interaction models (Manski, 1993). $\epsilon_r = (\epsilon_{r,1}, \cdots, \epsilon_{r,m_r})'$ is an m_r-dimensional vector of disturbances, where $\epsilon_{r,i}$'s are i.i.d. with zero mean and variance σ^2 for all i and r.

Let $X_r = (X_{1r}, W_r X_{2r})$. For a sample with $\bar{r}$ groups, stack up the data by defining $Y = (Y_1', \cdots, Y_{\bar{r}}')'$, $X = (X_1', \cdots, X_{\bar{r}}')'$, $u = (u_1', \cdots, u_{\bar{r}}')'$, $\epsilon = (\epsilon_1', \cdots, \epsilon_{\bar{r}}')'$, $W = \mathrm{D}(W_1, \cdots, W_{\bar{r}})$, $M = \mathrm{D}(M_1, \cdots, M_{\bar{r}})$, $\iota = \mathrm{D}(l_{m_1}, \cdots, l_{m_{\bar{r}}})$ and $\alpha_0 = (\alpha_{01}, \cdots, \alpha_{0\bar{r}})'$, where $\mathrm{D}(A_1, \cdots, A_K)$ is a block diagonal matrix in which the diagonal blocks are $m_k \times n_k$ matrices A_k's.[9] For the entire sample, the model is $Y = \lambda_0 WY + X\beta_0 + \iota\alpha_0 + u$, where $u = \rho_0 Mu + \epsilon$. Denote $R(\rho) = I - \rho M$ and $R \equiv R(\rho_0)$, where I is the identity matrix. To filter the spatial correlation in u, a Cochrane-Orcutt type transformation by R gives

$$RY = \lambda_0 RZ\delta_0 + R\iota\alpha_0 + \epsilon, \tag{2}$$

where $Z = (WY, X)$ and $\delta_0 = (\lambda_0, \beta_0')'$.

In this paper, we treat α_0 as a vector of unknown parameters. When the number of groups $\bar{r}$ is large, we have the incidental parameter problem (Neyman and Scott, 1948).[10] For example, in the ML approach, the estimators of the common parameters for all the groups may become inconsistent due to the joint estimation of the many group-specific parameters. For this reason, various approaches have been proposed to eliminate the fixed effect parameters in order to consistently estimate the common parameters of interest (see, eg., Kalbfleisch and Sprott, 1970; Baltagi, 1995; Hsiao, 2003). For panel regression models, popular methods involve subtraction of the time-averaged model (the within estimator) or subtraction of the time-lagged model (the first-differences estimator).[11] In this paper, we subtract the (weighted) group-averaged model to eliminate the group fixed effects.

Let R_r be the rth diagonal block of the matrix R. For the rth group, if M_r's rows all sum to a constant c, then $R_r l_{m_r} = (1 - c\rho_0) l_{m_r}$ and one can eliminate the group effect parameter α_{0r} by subtracting the group average from (2). However, if row sums of M_r vary, so will the elements of $R_r l_{m_r}$. In that situation, the subtraction of the group average could not eliminate the

[9] For the convenience of reference, a list of frequently used notations is provided in the Appendix.

[10] Neyman and Scott (1948) considered a simple panel regression model with $y_{it} \sim N(\mu_i, \sigma^2)$ where μ_i is the mean of y_{it} which is invariant for the cross-sectional unit i over time t with $t = 1, \cdots, T$; but σ^2 is a common variance parameter. They show that in the presence of many fixed (incidental) parameters μ_i's but only a finite number of time periods T, the imprecise (ML) estimates of μ_i, $i = 1, \cdots, n$, will render the estimator of the common parameter σ^2 to be inconsistent.

[11] When the proper variance matrices of the resulted disturbances are taken into account in the estimation, the resulted estimators are equivalent as the samples of the two different transformations have the same degrees of freedom.

group effect, and an alternative transformation is needed. As $R_r l_{m_r} = (l_{m_r}, M_r l_{m_r})(1, -\rho_0)'$, the unobserved group-specific effect α_{0r} can be eliminated with the orthogonal projector $J_r = I_{m_r} - (l_{m_r}, M_r l_{m_r})[(l_{m_r}, M_r l_{m_r})'(l_{m_r}, M_r l_{m_r})]^{-}(l_{m_r}, M_r l_{m_r})'$, where A^{-} denotes a generalized inverse of a square matrix A. In general, J_r projects an m_r-dimensional vector to the space spanned by l_{m_r} and $M_r l_{m_r}$. If $M_r l_{m_r} \neq c l_{m_r}$ for any constant c, J_r has rank $(m_r - 2)$. Otherwise, $J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l'_{m_r}$, which is the deviation from group mean projector with rank $(m_r - 1)$. Let $J = \mathrm{D}(J_1, \cdots, J_{\bar{r}})$. Premultiplication of (2) by J gives a model without the group effect parameters,

$$JRY = JRZ\delta_0 + J\epsilon. \tag{3}$$

Let $S(\lambda) = I - \lambda W$ and $S \equiv S(\lambda_0)$. The model (1) represents an equilibrium equation, so S is assumed to be invertible. The equilibrium vector Y is given by the reduced form equation

$$Y = S^{-1}(X\beta_0 + \iota\alpha_0) + S^{-1}R^{-1}\epsilon. \tag{4}$$

It follows that $WY = G(X\beta_0 + \iota\alpha_0) + GR^{-1}\epsilon$, where $G = WS^{-1}$. WY is correlated with ϵ because $\mathrm{E}((GR^{-1}\epsilon)'\epsilon) = \sigma_0^2 \mathrm{tr}(GR^{-1}) \neq 0$. Hence, in general, (3) cannot be consistently estimated by OLS.[12] On the other hand, (3) may not be considered as a self-contained system where the transformed variable JRY can be expressed as a function of the exogenous variables and disturbances, and, hence, a partial likelihood type approach may not be feasible based on (3).[13] In this paper, we consider the estimation of (3) by 2SLS and GMM approaches.

3 2SLS Estimation of the Network Model

3.1 The Estimator

From the reduced form equation (4), $\mathrm{E}(Z) = [G(X\beta_0 + \iota\alpha_0), X]$ and $JRZ = JR\mathrm{E}(Z) + JRGR^{-1}\epsilon e_1'$, where e_1 is the first unit (column) vector of dimension $(k+1)$ with $k = k_1 + k_2$.[14] The best IV matrix for JRZ in (3) is given by

$$f = JR\mathrm{E}(Z) = JR[G(X\beta_0 + \iota\alpha_0), X], \tag{5}$$

[12] Lee (2002) has shown the OLS estimator can be consistent in the spatial scenario where each spatial unit is influenced by many neighbors whose influences are uniformly small.

[13] When both W and R are row normalized, Y can be transformed into JRY on both sides of the regression equation. It is based on such a transformation, Lee et al. (2009) can derive the likelihood function for the transformed model.

[14] In this paper, X and α_0 are treated as constants. See Assumption 2 in the next subsection for more discussions.

which is an $n \times (k+1)$ matrix (see, eg., Lee, 2003). However, this IV matrix is infeasible as it involves unknown parameters $\theta_0 = (\rho_0, \delta_0')'$ and α_0. As $R = I - \rho_0 M$, f can be considered as a linear combination of the IVs in $Q_\infty = J(Q_\infty^0, MQ_\infty^0)$, where $Q_\infty^0 = (GX, G\iota, X)$. As ι has $\bar{r}$ columns, the number of IVs in Q_∞ increases as the number of groups $\bar{r}$ increases. Furthermore, as $G = W(I - \lambda_0 W)^{-1} = \sum_{j=0}^{\infty} \lambda_0^j W^{j+1}$ when $\sup ||\lambda_0 W||_\infty < 1$, GX and $G\iota$ in Q_∞^0 can be replaced by linear combinations of $(WX, W^2X, \cdots)$ and $(W\iota, W^2\iota, \cdots)$ respectively, and, hence, $Q_\infty^0 = (WX, W^2X, \cdots, W\iota, W^2\iota, \cdots, X)$.[15] In this paper, we show that, when the sample size grows fast enough relative to the number of IVs, the asymptotic efficiency can be obtained by using a sequence of IVs approximating the best set of IVs Q_∞. On the other hand, the 2SLS with a fixed number of IVs would be consistent but, in general, not efficient.

Let Q_K^0 be a submatrix of Q_∞^0 including X,[16] and $Q_K = J[Q_K^0, MQ_K^0]$ be an $n \times K$ IV matrix with $K \geq (k+1)$. For simplicity, K serves as both the number of IVs and the index of the IV set, as in Donald and Newey (2001). In general, K may be a function of n such that K is allowed to increase with n. Let $\epsilon(\theta) = JR(\rho)(Y - Z\delta)$ with $\theta = (\rho, \delta')'$ and $\epsilon(\rho_0, \delta) = JR(Y - Z\delta)$. The moment function corresponding to the orthogonality condition of Q_K and $J\epsilon$ is $Q_K'\epsilon(\rho_0, \delta)$. Let $P_K = Q_K(Q_K'Q_K)^- Q_K'$. With a preliminary estimator $\tilde{\rho}$ of ρ_0 and $\tilde{R} = R(\tilde{\rho})$,[17] the 2SLS estimator is

$$\hat{\delta}_{2sls} = (Z'\tilde{R}'P_K\tilde{R}Z)^{-1}Z'\tilde{R}'P_K\tilde{R}Y. \tag{6}$$

3.2 Identification, Consistency and Asymptotic Distributions

To proceed, we assume the following regularity conditions. Henceforth, uniformly bounded in row (column) sums in absolute value of a sequence of square matrices $\{A\}$ will be abbreviated as UBR (UBC), and uniformly bounded in both row and column sums in absolute value as UB.[18]

Assumption 1 The elements of ϵ are i.i.d. with zero mean, variance σ_0^2 and that a moment of order

[15] See the paragraph following Assumption 5 for more details.

[16] A simple example for Q_K^0 would be $(WX, W\iota, X)$. Please see the Monte Carlo simulation part of the paper for more discussions on the selection of Q_K.

[17] A consistent initial estimator of $\tilde{\rho}$ may be obtained in various ways. One may follow a two step approach as in Kelejian and Prucha (1998), where a 2SLS with a finite number of IVs can be used to estimate the main equation (without the transformation by R) and then estimate ρ_0 in the SAR disturbances via some method of moments with estimated residuals. As an alternative, one may estimate jointly all the parameters with a finite number of linear and quadratic moments in a GMM setting in Lee (2007a). The resulted estimator $\tilde{\rho}$ can be $\sqrt{n}$-consistent by both methods.

[18] A sequence of square matrices $\{A\}$, where $A = [A_{ij}]$, is said to be UBR (UBC) if the sequence of row sum matrix norm $||A||_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^n |A_{ij}|$ (column sum matrix norm $||A||_1 = \max_{j=1,\cdots,n} \sum_{i=1}^n |A_{ij}|$) is bounded. (Horn and Johnson, 1985)

higher than the fourth exists.[19]

Assumption 2 The elements of X are uniformly bounded constants, X has the full rank k, and $\lim_{n\to\infty} \frac{1}{n} X'X$ exists and is nonsingular.[20]

Assumption 3 The sequences of matrices $\{W\}$, $\{M\}$, $\{S^{-1}\}$ and $\{R^{-1}\}$ are UB.

The existence of moments higher than the fourth of ϵ is needed in order to apply a central limit theorem (CLT) of a linear and quadratic form due to Kelejian and Prucha (2001). The uniform boundedness condition for X in Assumptions 2 is for analytic tractability. The uniform boundedness of $\{W\}$, $\{M\}$, $\{S^{-1}\}$ and $\{R^{-1}\}$ in Assumption 3 limits spatial dependence among the units to a tractable degree and is originated in Kelejian and Prucha (1999). It rules out the unit root case (in time series as a special case).[21]

From (3), we have

$$\epsilon(\rho_0, \delta) = JR(Y - Z\delta) = f(\delta_0 - \delta) + JRS(\lambda)S^{-1}R^{-1}\epsilon, \tag{7}$$

where f is in (5) and $S(\lambda) = I - \lambda W$. The identification of δ_0 is based on the moment condition $\mathrm{E}(Q_K'\epsilon(\rho_0, \delta)) = 0$. For any feasible δ, $\mathrm{E}(Q_K'\epsilon(\rho_0, \delta)) = Q_K' f(\delta_0 - \delta) = 0$, which has the unique solution δ_0 if $Q_K' f$ has full column rank $k+1$. This rank condition requires necessarily that the best IV matrix f has a full rank $(k+1)$. For each group, $G_r l_{m_r} = W_r(I_{m_r} - \lambda_0 W_r)^{-1} l_{m_r}$ is the Bonacich measure of centrality of group members. When W_r does not have equal indegrees for all its nodes and W_r is not row-normalized, the elements of $G_r l_{m_r}$ are not all equal, which represent different centrality scores of group members. In this situation, J_r will not eliminate $R_r G_r l_{m_r}$, which, in turn, helps to identify δ_0. We see that, even when X is irrelevant in the model so that $\beta_0 = 0$, the best IV matrix $f = JR(G\iota\alpha_0, X)$ may still have the rank $(k+1)$ and identification is possible. This identification condition is weaker than the one when W_r has constant row sums. For the

[19] It may be possible to generalize the analysis to the case under heteroscedasticity, as we assume K increases slower than n. The analysis of 2SLS would be similar under heteroscedasticity even though the limiting variance of the estimate would change as usual. For the quadratic moments in the GMM estimation, as the number of quadratic moments is finite, based on a similar idea as in Lin and Lee (2006), it may be possible to modify the proposed GMM procedure so that the estimator is robust to unknown heteroscedasticity.

[20] If X is allowed to be stochastic, then appropriate moment conditions need to be imposed, and the results presented in this paper can be considered as conditional on X instead.

[21] For the limiting case $||\lambda_0 W_n|| \to 1$, Lee and Yu (2009) have some preliminary results. They have shown that the limiting distribution of the estimator of the endogenous interaction coefficient (λ_0 in their spatial model) can have a faster rate of convergence and the properly normalized estimator can still be asymptotically normal. In terms of the Bonacich centrality, this could imply a divergent sequence of such measures.

latter case, $J_r R_r G_r l_{m_r} = 0$ so that $f = JR(GX\beta_0, X)$ and the identification condition based on $\mathrm{E}(Q_K'\epsilon(\rho_0,\delta)) = 0$ could not be satisfied if $\beta_0 = 0$. Assumption 4 gives a sufficient identification condition for δ_0.

Assumption 4 $H = \lim_{n\to\infty} \frac{1}{n} f'f$ is a finite nonsingular matrix.

Assumption 4 also characterizes the quality of the instruments. Let $\bar{G} = RGR^{-1}$. The variance of the error term in the reduced form equation be denoted by $\Sigma = \frac{1}{n}\mathrm{E}(e_1\epsilon'\bar{G}'J\bar{G}\epsilon e_1') = \frac{1}{n}\sigma_0^2 \mathrm{tr}(\bar{G}'J\bar{G})e_1e_1'$. Because only the (1,1) entry of Σ is non-zero, with the corresponding (1,1) entry of $f'f$, the concentration parameter for our network model is $[G(X\beta_0 + \iota\alpha_0)]'R'JR[G(X\beta_0 + \iota\alpha_0)]/[\frac{1}{n}\sigma_0^2\mathrm{tr}(\bar{G}'J\bar{G})]$. The concentration parameter is a natural measure of the instrument strength. We assume the limit of $\frac{1}{n}f'f$ is a constant as the sample size goes to infinity in Assumption 4, which implies that the concentration parameter grows at the same rate as the sample size. Such a rate is assumed in Bekker (1994). Assumption 4 seems reasonable as it covers many practical cases where either X or the Bonacich measure has significant effects on Y. Hence, under Assumption 4, we focus on scenarios where instruments are stronger than assumed in the weak-instrument literature.

Assumption 5 There exists a $K \times (k+1)$ matrix π_K such that $||\mathrm{E}(Z) - Q_K^0\pi_K||_\infty \to 0$ as $n, K \to \infty$.[22]

Assumption 5 concerns approximation of the ideal IV matrix f. There are a few possible sequences of $\{Q_K^0\}$ that satisfy this assumption and we illustrate one of them as follows. $\mathrm{E}(Z)$ can be presented as a linear combination of $\bar{Q}_K^0 = (GX, G\iota, X)$ because $\mathrm{E}(Z) = \bar{Q}_K^0\bar{\pi}_K$ with $\bar{\pi}_K = \begin{pmatrix} \beta_0' & \alpha_0' & 0 \\ 0 & 0 & I_k \end{pmatrix}'$. Hence, the ideal IV matrix $f = JR\mathrm{E}(Z)$ can be presented as a linear combination of $\bar{Q}_K = J[\bar{Q}_K^0, M\bar{Q}_K^0]$. As G in $\bar{Q}_K$ involves the unknown parameter λ_0, we approximate G by a series expansion at λ_0, and express the approximated G as a linear combination of a sequence of known matrices. If $\sup||\lambda_0 W||_\infty < 1$, $G = WS^{-1} = \sum_{j=0}^{\infty}\lambda_0^j W^{j+1} = \sum_{j=0}^{p}\lambda_0^j W^{j+1} + (\lambda_0 W)^{p+1}G$. It follows that $||G - \sum_{j=0}^{p}\lambda_0^j W^{j+1}||_\infty \le ||\lambda_0 W||_\infty^{p+1}||G||_\infty = o(1)$ as $p \to \infty$. Here, as $\sup||\lambda_0 W||_\infty < 1$, the approximation error by series expansion diminishes very fast in a geometric rate, as long as the degree of approximation increases as sample size increases. This example is summarized in the following lemma.

[22]The assumption that the ideal instrument can be approximated by a certain linear combination has been used in the many-instruments literature (see, eg., Donald and Newey, 2001; Hansen et al., 2008; Hausman et al., 2008).

Lemma 3.1 *If* $\sup ||\lambda_0 W||_\infty < 1$*, for* $Q_K^{(p)} = J(G_X^{(p)}, MG_X^{(p)}, G_\iota^{(p)}, MG_\iota^{(p)}, X, MX)$ *where* $G_X^{(p)} = (WX, \cdots, W^{p+1}X)$*,* $G_\iota^{(p)} = (W\iota, \cdots, W^{p+1}\iota)$ *and* p *is an increasing integer-valued function of* K*, there exists a* $K \times (k+1)$ *matrix* $\pi_K^{(p)}$ *such that* $||f - Q_K^{(p)}\pi_K^{(p)}||_\infty \to 0$ *as* $n, K \to \infty$*.*

The 2SLS estimator with an increasing number of IVs approximating f can be asymptotically efficient under some conditions. However, when the number of instruments increases too fast, such an estimator could be asymptotically biased, which is known as the many-instrument problem (see, eg., Bekker, 1994; Donald and Newey, 2001; Alvarez and Arellano, 2003; Han and Phillips, 2006). It is possible to use a fixed number of IVs to avoid this problem. The 2SLS estimator with a fixed number of IVs will be consistent and asymptotically normal but may not be efficient. In this paper, we focus on IV matrices satisfying Assumption 5. The following proposition shows consistency and asymptotic normality of the 2SLS estimator in (6). Let $\Psi_K = P_K \bar{G}$.

Proposition 1 *Under Assumptions 1-5, if* $K/n \to 0$ *and* $\sqrt{n}(\tilde{\rho} - \rho_0) = O_p(1)$*, then* $\sqrt{n}(\hat{\delta}_{2sls} - \delta_0 - b_{2sls}) \xrightarrow{d} N(0, \sigma_0^2 H^{-1})$*, where* $b_{2sls} = \sigma_0^2 \text{tr}(\Psi_K)(Z'R'P_K RZ)^{-1} e_1 = O_p(K/n)$*.*

From this proposition, we see that $\sqrt{n}(\hat{\delta}_{2sls} - \delta_0)$ has the bias $\sqrt{n} b_{2sls}$ due to the increasing number of IVs. To understand this bias, let us look at the normal equation of the 2SLS, $\frac{1}{n} Z'R'P_K R(Y - Z\hat{\delta}_{2sls}) = 0$, as if ρ_0 were known. At δ_0, $\text{E}[\frac{1}{n} Z'R'P_K R(Y - Z\delta_0)] = \frac{1}{n}\sigma_0^2 \text{tr}(\Psi_K) e_1 = O(K/n)$ (see Lemma B.2 in the Appendix), which does not converge to zero when the number of IVs grows at the same rate of the sample size. The following corollary summarizes the asymptotic properties of the 2SLS for different divergent rates of K in terms of n.

Corollary 1 *Under Assumptions 1-5 and* $\sqrt{n}(\tilde{\rho} - \rho_0) = O_p(1)$*,* (i) *if* $K^2/n \to 0$*,* $\sqrt{n}(\hat{\delta}_{2sls} - \delta_0) \xrightarrow{d} N(0, \sigma_0^2 H^{-1})$*;* (ii) *if* $K^2/n \to c < \infty$ *and* $c \neq 0$*,* $\sqrt{n}(\hat{\delta}_{2sls} - \delta_0) \xrightarrow{d} N(\bar{b}_{2sls}, \sigma_0^2 H^{-1})$*, where* $\bar{b}_{2sls} = \lim_{n\to\infty} \sqrt{n} b_{2sls} = \sigma_0^2 H^{-1} \lim_{n\to\infty} \text{tr}(\Psi_K) e_1/\sqrt{n}$*; and* (iii) *if* $K^2/n \to \infty$ *but* $K^{1+\eta}/n \to 0$ *for* $0 < \eta < 1$*,* $K^\eta(\hat{\delta}_{2sls} - \delta_0) \xrightarrow{p} 0$*.*

The sequence of IV matrices $\{Q_K\}$ given in the previous subsection provides the asymptotically best IV estimator when $K^2/n \to 0$, as the variance matrix $\sigma_0^2 H^{-1}$ attains the efficiency lower bound for the class of IV estimators. With a fixed number of instruments given by $\bar{Q}$, the asymptotic distribution of the 2SLS estimator $\tilde{\delta}$ is given by $\sqrt{n}(\tilde{\delta} - \delta_0) \xrightarrow{d} N(0, \sigma_0^2(\lim_{n\to\infty} \frac{1}{n} f'\bar{P}f)^{-1})$, where $\bar{P} = \bar{Q}(\bar{Q}'\bar{Q})^{-1}\bar{Q}'$. Note that $H - \lim \frac{1}{n} f'\bar{P}f = \lim \frac{1}{n} f'(I - \bar{P})f$, which is positive semi-definite in general, unless f lies asymptotically on the space spanned by the columns of $\bar{P}$, i.e., $\lim \frac{1}{n}(f - \bar{P}f) = 0$. Hence, in general, the 2SLS with a fixed number of instruments is not efficient.

On the other hand, the condition that $K/n \to 0$ is crucial for the IV estimator to be consistent. The following corollary illustrates the inconsistency of the 2SLS estimator if K/n converges to a nonzero constant. This result is closely related to Bekker (1994), who is among the first to show the inconsistency of IV estimators when K grows as fast as n.

Proposition 2 *Under Assumptions 1-5, if* $K/n \to c \neq 0$ *and* $\tilde{\rho} - \rho_0 = o_p(1)$*, then* $\hat{\delta}_{2sls} - \delta_0 - b_{2sls} \xrightarrow{p} 0$*, where* b_{2sls} *might converge to a nonzero constant.*

Note $G\iota$ in the ideal IV set f has $\bar{r}$ columns and we use a series expansion to approximate G. So, if we include an approximated Bonacich centrality measure from each of the $\bar{r}$ groups as an IV in Q_K, then $\bar{r}/K \to 0$. Hence, $K/n \to 0$ implies $\bar{r}/n = 1/\bar{m} \to 0$, where $\bar{m}$ is the average group size. So for consistency of the 2SLS estimator using the centrality measures of all the groups as IVs, the average group size needs to be large. On the other hand, $K^2/n \to 0$ implies $\bar{r}^2/n = \bar{r}/\bar{m} \to 0$. So for asymptotic efficiency, the average group size needs to be large relative to the number of groups in order that the asymptotic distribution of the 2SLS estimator is properly centered. The asymptotic bias can be either eliminated or reduced by some bias correction procedures, which may relax the large average group size requirement to some extent.[23]

3.3 Bias Correction

To correct for the many-instrument bias in the 2SLS estimator, we can estimate the leading order bias b_{2sls} in Proposition 1 and adjust the 2SLS estimator by the estimated bias. With consistent initial estimators $\tilde{\sigma}^2, \tilde{\lambda}$ and $\tilde{\rho}$, the feasible bias-corrected 2SLS (FC2SLS) is given by

$$\hat{\delta}_{fc2sls} = (Z'\tilde{R}'P_K\tilde{R}Z)^{-1}[Z'\tilde{R}'P_K\tilde{R}Y - \tilde{\sigma}^2 \mathrm{tr}(P_K\tilde{R}\tilde{G}\tilde{R}^{-1})e_1], \tag{8}$$

where $\tilde{G} = G(\tilde{\lambda})$. The following result gives consistency and asymptotic normality of the FC2SLS estimator.

[23] The assumptions for asymptotic results in this paper do not rule out the case where the number of groups is fixed. So the analysis covers that case. However, this assertion relies on Assumption 4 in the paper. Section 6 of Lee (2004) has shown that the estimator could be inconsistent in a special case when $m_r \to \infty$ with a single group, i.e., $r = 1$. In that special case, an individual has "many neighbors/friends" such that weights matrix W_n has many nonnegative elements that are uniformly of order $O(1/h_n)$, where h_n increases to infinity at the same rate as the sample size n, and it generates a multicollinearity problem in the reduced form equation. Assumption 4 has ruled out that special case as multicollinearity is not allowed in the reduced form equation in this paper.

Proposition 3 *Under Assumptions 1-5, if $K/n \to 0$ and $\tilde{\sigma}^2, \tilde{\lambda}, \tilde{\rho}$ are $\sqrt{n}$-consistent initial estimators of $\sigma_0^2, \lambda_0, \rho_0$, then $\sqrt{n}(\hat{\delta}_{fc2sls} - \delta_0) \xrightarrow{d} N(0, \sigma_0^2 H^{-1})$.*

If the Bonacich centrality measures for all groups are included in Q_K as IVs, the properly centered and asymptotically efficient 2SLS estimator can be derived as long as the average group size, implicitly in n/K, is large.[24]

4 GMM Estimation of the Network Model

4.1 The Estimator, Identification and Consistency

The 2SLS can be generalized to the GMM with additional quadratic moment equations. The use of quadratic moments for the estimation of SAR models has been proposed in Kelejian and Prucha (1999) and Lee (2007a; 2007c). While the IV moments utilize the information of the mean regression function of the reduced form for estimation, the quadratic moments explore the correlation structure of the reduced form disturbances.[25] The IV moments $g_1(\theta) = Q_K'\epsilon(\theta)$ are linear in ϵ at θ_0. The additional quadratic moments are $g_2(\theta) = [U_1\epsilon(\theta), \cdots, U_q\epsilon(\theta)]'\epsilon(\theta)$, where U_j's are constant square matrices such that $\mathrm{tr}(JU_j) = 0$ and the number of quadratic moments q is fixed.[26] At θ_0, $\mathrm{E}(g_2(\theta_0)) = 0$, because $\epsilon(\theta_0) = J\epsilon$ and $\mathrm{E}(\epsilon' JU_j J\epsilon) = \sigma_0^2 \mathrm{tr}(JU_j) = 0$. As a normalization for notational simplicity, let U_j replace JU_jJ. The vector of combined linear and quadratic empirical moments for the GMM estimation is given by $g(\theta) = [g_1'(\theta), g_2'(\theta)]'$. For analytic tractability, the following assumption imposes uniform boundedness on the quadratic matrices U_j's.

Assumption 6 The sequences of matrices $\{U_j\}$ with $\mathrm{tr}\,(JU_j) = 0$ are UB for $j = 1, \cdots, q$.

Similar to (7), we have

$$\epsilon(\theta) = JR(\rho)(Y - Z\delta) = f(\rho)(\delta_0 - \delta) + JR(\rho)S(\lambda)S^{-1}R^{-1}\epsilon, \tag{9}$$

[24] The initial consistent estimator $\tilde{\lambda}$ is required to be $\sqrt{n}$-consistent. Such an estimator can be derived by using a fixed number of IVs. On the other hand, the 2SLS estimator $\hat{\lambda}$ with an increasing number of IVs in Proposition 1 is $\sqrt{n}/K$-consistent due to the bias. If such an estimator is used as the preliminary estimator, the FC2SLS estimator can be properly centered under the stronger restriction on K such that $K^{4/3}/n \to 0$. This can be seen from the proof of Proposition 3 in the appendix.

[25] For a typical SAR model with normal disturbances (without fixed effects), the best selection of linear and quadratic moments can provide a GMM estimator which is asymptotically as efficient as the maximum likelihood estimator (see, Lee, 2007a; Lee and Liu, 2010).

[26] For any $n \times n$ constant matrix B, define A as $A = B - \mathrm{tr}(JB)I/\mathrm{tr}(J)$, then $\mathrm{tr}(JA) = 0$. Some simple examples of U_j are $U_1 = M - \mathrm{tr}(JM)I/\mathrm{tr}(J)$ and $U_2 = W - \mathrm{tr}(JW)I/\mathrm{tr}(J)$. When the errors are normally distributed, the best quadratic moments, thus optimal q, is given in section 4.3.

where $f(\rho) = JR(\rho)\mathrm{E}(Z)$. The moment condition $\mathrm{E}(g_1(\theta)) = Q_K' f(\rho)(\delta_0 - \delta) = 0$ has the unique solution δ_0 under the rank condition in the following stronger version of Assumption 4.[27] So identification of δ_0 via the IV moments is possible, which is similar to the identification of the 2SLS.

Assumption 4' $\lim_{n\to\infty} \frac{1}{n} f'(\rho) f(\rho)$ is a nonsingular matrix for any ρ such that $R(\rho)$ is nonsingular.

With δ_0 identified, it follows from (9) that $\epsilon(\theta|_{\delta=\delta_0}) = JR(\rho)R^{-1}\epsilon$ and ρ_0 can be identified based on the quadratic moment condition

$$\mathrm{E}[g_2(\theta|_{\delta=\delta_0})] = [\sigma_0^2 \mathrm{tr}(R^{'-1}R'(\rho)U_1R(\rho)R^{-1}), \cdots, \sigma_0^2 \mathrm{tr}(R^{'-1}R'(\rho)U_qR(\rho)R^{-1})]' = 0.$$

As $R(\rho)\,R^{-1} = I + (\rho_0 - \rho)\,MR^{-1}$, it follows that

$$\mathrm{tr}(R^{'-1}R'(\rho)U_jR(\rho)R^{-1}) = (\rho_0 - \rho)\,[\mathrm{tr}(U_j^s MR^{-1}) + (\rho_0 - \rho)\,\mathrm{tr}(R^{-1\prime}M'U_jMR^{-1})],$$

for $i = 1, \cdots, q$. The following assumption gives a sufficient identification condition for ρ_0 via the unique solution of $\mathrm{E}[g_2(\theta|_{\delta=\delta_0})] = 0$ for a large enough n.

Assumption 7 $\lim_{n\to\infty} \frac{1}{n}\mathrm{tr}(U_j^s MR^{-1}) \neq 0$ for some j, and $\lim_{n\to\infty} \frac{1}{n}[\mathrm{tr}(U_1^s MR^{-1}), \cdots, \mathrm{tr}(U_q^s MR^{-1})]'$ is linearly independent of $\lim_{n\to\infty} \frac{1}{n}[\mathrm{tr}(R^{-1\prime}M'U_1^s MR^{-1}), \cdots, \mathrm{tr}(R^{-1\prime}M'U_q^s MR^{-1})]'$.

This identification condition is similar to the one for the SAR model with SAR disturbances in Lee and Liu (2010).

For any $n \times n$ matrix $A = [a_{ij}]$, let $A^s = A + A'$ and $\mathrm{vec}_D(A) = (a_{11}, \cdots, a_{nn})'$. In general, μ_3 and μ_4 denote, respectively, the third and fourth moments of the error term. The variance matrix of $g(\theta_0)$ is given by

$$\Omega = \mathrm{Var}(g(\theta_0)) = \begin{pmatrix} \sigma_0^2 Q_K' Q_K & \mu_3 Q_K' \omega \\ \mu_3 \omega' Q_K & (\mu_4 - 3\sigma_0^4)\omega'\omega + \sigma_0^4 \Upsilon \end{pmatrix},$$

where $\omega = [\mathrm{vec}_D(U_1), \cdots, \mathrm{vec}_D(U_q)]$ and $\Upsilon = \frac{1}{2}[\mathrm{vec}(U_1^s), \cdots, \mathrm{vec}(U_q^s)]'[\mathrm{vec}(U_1^s), \cdots, \mathrm{vec}(U_q^s)]$. By the generalized Schwartz inequality, the optimal weighting matrix of the GMM is Ω^{-1}. As the dimension of IV moments in $g(\theta)$ increases with the number of groups, the limit of $g(\theta)$ is not well

[27] For the feasible 2SLS approach, an initial consistent estimator of ρ_0 is utilized, so one needs only the local behavior of $\frac{1}{n} f'(\rho) f(\rho)$ at ρ_0 as in Assumption 4. On the other hand, the GMM approach estimates simultaneously δ_0 and ρ_0 in the model, Assumption 4 needs to be slightly generalized.

defined and the usual argument for consistency does not directly apply.[28] We find that if the optimal GMM objective function can be properly rewritten, asymptotic analysis of the GMM estimator for our case can become tractable. By the inverse of the partitioned matrix (Amemiya (1985), p.460), $\Omega^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, where $B_{11} = \sigma_0^{-2}(Q_K'Q_K)^{-1} + (\frac{\mu_3}{\sigma_0^2})^2(Q_K'Q_K)^{-1}Q_K'\omega B_{22}\omega' Q_K(Q_K'Q_K)^{-1}$, $B_{21} = B_{12}' = -\frac{\mu_3}{\sigma_0^2}B_{22}\omega' Q_K(Q_K'Q_K)^{-1}$ and $B_{22} = [(\mu_4 - 3\sigma_0^4)\omega'\omega + \sigma_0^4\Upsilon - \frac{\mu_3^2}{\sigma_0^2}\omega' P_K\omega]^{-1}$. Henceforth, we drop the subscript n on B_{ij} for simplicity. The optimal GMM minimizes

$$\begin{aligned}
& g'(\theta)\Omega^{-1}g(\theta) \\
= \;& g_1'(\theta)B_{11}g_1(\theta) + g_1'(\theta)B_{12}g_2(\theta) + g_2'(\theta)B_{21}g_1(\theta) + g_2'(\theta)B_{22}g_2(\theta) \\
= \;& \sigma_0^{-2}\epsilon'(\theta)P_K\epsilon(\theta) + (\frac{\mu_3}{\sigma_0^2})^2\epsilon'(\theta)P_K\omega B_{22}\omega' P_K\epsilon(\theta) - 2\frac{\mu_3}{\sigma_0^2}\epsilon'(\theta)P_K\omega B_{22}g_2(\theta) + g_2'(\theta)B_{22}g_2(\theta) \\
= \;& \sigma_0^{-2}\epsilon'(\theta)P_K\epsilon(\theta) + \bar{g}_2'(\theta)B_{22}\bar{g}_2(\theta),
\end{aligned}$$

where $\bar{g}_2(\theta) = \frac{\mu_3}{\sigma_0^2}\omega' P_K\epsilon(\theta) - g_2(\theta)$ is the error in the linear projection of $g_2(\theta)$ on $g_1(\theta)$. Note that

$$\mathrm{Var}(\bar{g}_2(\theta_0)) = \frac{\mu_3^2}{\sigma_0^2}\omega' P_K\omega - \frac{2\mu_3}{\sigma_0^2}\mathrm{E}[\omega' P_K\epsilon(U_1'\epsilon, \cdots, U_q'\epsilon)'\epsilon] + \mathrm{Var}[g_2(\theta_0)] = B_{22}^{-1}.$$

Hence, the optimal GMM objective function can be treated as a linear combination of the objective functions of the 2SLS and the optimal GMM based on moments $\bar{g}_2(\theta)$, which has a fixed dimension. Furthermore, as $\mathrm{E}[\bar{g}_2(\theta_0)g_1'(\theta_0)] = \mathrm{E}[\frac{\mu_3}{\sigma_0^2}\omega' P_K\epsilon\epsilon' Q_K] - \mathrm{E}[g_2(\theta_0)\epsilon' Q_K] = \mu_3\omega' Q_K - \mu_3\omega' Q_K = 0$, $g_1(\theta_0)$ and $\bar{g}_2(\theta_0)$ are uncorrelated.

The following assumption warrants the limit of nB_{22} exists. As usual for nonlinear estimation, the parameter space Θ of θ will be taken to be a bounded set with θ_0 in its interior.[29] Let $F = JM\mathrm{E}(Z)$.

Assumption 8 (i) $\lim_{n\to\infty} nB_{22}$ exists and is a nonsingular matrix. (ii) $\lim_{n\to\infty} \frac{1}{n}\omega' f$ and $\lim_{n\to\infty} \frac{1}{n}\omega' F$ exist.

Assumption 9 The θ_0 is in the interior of the bounded parameter space Θ.

The optimal weighting matrix Ω^{-1} involves unknown parameters σ_0^2, μ_3 and μ_4. In practice, with initial consistent estimators $\tilde{\sigma}^2$, $\tilde{\mu}_3$ and $\tilde{\mu}_4$, Ω can be estimated as $\tilde{\Omega} = \Omega(\tilde{\sigma}^2, \tilde{\mu}_3, \tilde{\mu}_4)$. The next

[28] With a high level of regularity conditions, Han and Phillips (2006) proposed an alternative framework for analysis. However, that framework does not easily incorporate general feasible optimal GMM weighting matrices.

[29] Here we do not need the parameter space to be compact because the GMM objective function is a quadratic function of the parameters of interest.

proposition gives the consistency of the feasible optimal GMM estimator.

Proposition 4 *Under Assumptions 1-3,4',5-9, if* $K/n \to 0$ *and* $\tilde{\Omega} = \Omega(\tilde{\sigma}^2, \tilde{\mu}_3, \tilde{\mu}_4)$ *where* $\tilde{\sigma}^2, \tilde{\mu}_3, \tilde{\mu}_4$ *are consistent initial estimators of* σ_0^2, μ_3, μ_4*, then the feasible optimal GMM estimator* $\hat{\theta}_{gmm} = \arg\min_{\theta \in \Theta} g'(\theta)\tilde{\Omega}^{-1}g(\theta)$ *is consistent.*

4.2 Asymptotic Normality and Bias Correction

Let

$$D_2 = \mathrm{E}[\frac{\partial}{\partial \theta'} g_2(\theta_0)] = -\sigma_0^2 \begin{pmatrix} \mathrm{tr}(U_1^s M R^{-1}) & \mathrm{tr}(U_1^s \bar{G}) & 0_{1\times k} \\ \vdots & \vdots & \vdots \\ \mathrm{tr}(U_q^s M R^{-1}) & \mathrm{tr}(U_q^s \bar{G}) & 0_{1\times k} \end{pmatrix}.$$

The next result gives the asymptotic distribution of the feasible optimal GMM estimator.[30]

Proposition 5 *Under Assumptions 1-3,4',5-9, if* $K^{3/2}/n \to 0$ *and* $\tilde{\Omega} = \Omega(\tilde{\sigma}^2, \tilde{\mu}_3, \tilde{\mu}_4)$ *where* $\tilde{\sigma}^2, \tilde{\mu}_3, \tilde{\mu}_4$ *are* $\sqrt{n}$*-consistent initial estimators of* σ_0^2, μ_3, μ_4*, then the feasible optimal GMM estimator* $\hat{\theta}_{gmm} = \arg\min_{\theta \in \Theta} g'(\theta)\tilde{\Omega}^{-1}g(\theta)$ *has the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{gmm} - \theta_0 - b_{gmm}) \xrightarrow{d} N(0, (\sigma_0^{-2}\mathrm{D}(0, H) + \lim_{n\to\infty} \frac{1}{n} \bar{D}_2' B_{22} \bar{D}_2)^{-1}),$$

where $b_{gmm} = (\sigma^{-2}\mathrm{D}(0, Z'R'P_K RZ) + \check{D}_2' B_{22} \check{D}_2)^{-1}[\mathrm{tr}(P_K M R^{-1}), \mathrm{tr}(\Psi_K) e_1']' = O(K/n)$*,* $\check{D}_2 = D_2 - \frac{\mu_3}{\sigma_0^2}(0, \omega' P_K RZ)$*, and* $\bar{D}_2 = D_2 - \frac{\mu_3}{\sigma_0^2}(0, \omega' f)$*.*

It is interesting to note that, although the identification of ρ_0 is based on a fixed number of quadratic moments, joint estimation of ρ_0 with other parameters introduces an asymptotic bias to its GMM estimator due to the large number of linear moments.[31] As the asymptotic bias is $O(K/n)$, the asymptotic distribution of the GMM estimator will be centered at θ_0 only when $K^2/n \to 0$. These are similar conditions to the 2SLS estimator in Proposition 1 and Corollary 1. Proposition 5 holds under the required regularity condition that $K^{3/2}/n \to 0$, so that the remainder terms in the asymptotic expansion vanish. The restriction on K via $K^{3/2}/n \to 0$ is stronger than that of

[30] When the number of instruments increases at a slower rate than the sample size, the asymptotic distribution of the 2SLS is not affected by non-normality of the error term. This is so, because the asymptotic approximation of the 2SLS estimator is linear in the error term when $K/n \to 0$. On the other hand, non-normality could affect the asymptotic distribution of the GMM estimator as the covariance matrix of the linear and quadratic moment conditions involves the third and fourth order moments of the error term.

[31] For the efficient estimation of a SAR model with SAR disturbances such as the method of maximum likelihood, the asymptotic variances of the estimators of ρ_0 and the other parameters such as λ_0 are not block diagonal. So the efficient estimates of ρ_0 and λ_0, etc., are not asymptotically uncorrelated.

$K/n \to 0$ in Proposition 1 for the 2SLS but is weaker than that of $K^2/n \to c$ for a finite nonzero constant c.

The bias term b_{gmm} in the Proposition 5 can be estimated by

$$\tilde{b}_{gmm} = [\tilde{\sigma}^{-2}\mathrm{D}(0, Z'\tilde{R}'P_K\tilde{R}Z) + \tilde{\tilde{D}}_2'\tilde{B}_{22}\tilde{\tilde{D}}_2]^{-1}[\mathrm{tr}(P_K M\tilde{R}^{-1}), \mathrm{tr}(P_K\tilde{R}\tilde{G}\tilde{R}^{-1})e_1']', \qquad (10)$$

where $\tilde{B}_{22} = [(\tilde{\mu}_4 - 3(\tilde{\sigma}^2)^2)\omega'\omega + (\tilde{\sigma}^2)^2\Upsilon - \frac{\tilde{\mu}_3}{\tilde{\sigma}^2}\omega'P_K\omega]^{-1}$ and $\tilde{\tilde{D}}_2 = \tilde{D}_2 - \frac{\tilde{\mu}_3}{\tilde{\sigma}^2}(0, \omega'P_K\tilde{R}Z)$ with

$$\tilde{D}_2 = -\tilde{\sigma}^2 \begin{pmatrix} \mathrm{tr}(U_1^s M\tilde{R}^{-1}) & \mathrm{tr}(U_1^s\tilde{R}\tilde{G}\tilde{R}^{-1}) & 0_{1\times k} \\ \vdots & \vdots & \vdots \\ \mathrm{tr}(U_q^s M\tilde{R}^{-1}) & \mathrm{tr}(U_q^s\tilde{R}\tilde{G}\tilde{R}^{-1}) & 0_{1\times k} \end{pmatrix}.$$

With the consistently estimated leading order bias, the feasible bias-corrected GMM (FCGMM) estimator is given by $\hat{\theta}_{fcgmm} = \hat{\theta}_{gmm} - \tilde{b}_{gmm}$.[32]

Proposition 6 *Under Assumptions 1-3,4',5-9, if* $K^{3/2}/n \to 0$, $\tilde{\lambda} - \lambda_0 = O_p(\max\{1/\sqrt{n}, K/n\})$, $\tilde{\rho} - \rho_0 = O_p(\max\{1/\sqrt{n}, K/n\})$ *and* $\tilde{\sigma}^2, \tilde{\mu}_3, \tilde{\mu}_4$ *are* $\sqrt{n}$*-consistent initial estimators of* σ_0^2, μ_3, μ_4, *then the FCGMM estimator* $\hat{\theta}_{fcgmm}$ *has the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_{fcgmm} - \theta_0) \xrightarrow{d} N(0, (\sigma_0^{-2}\mathrm{D}(0, H) + \lim_{n\to\infty}\frac{1}{n}\bar{D}_2'B_{22}\bar{D}_2)^{-1}).$$

The asymptotic variance matrix of the optimal GMM estimator can be compared with that of the 2SLS estimator. As $\bar{D}_2'B_{22}\bar{D}_2$ is nonnegative definite, the asymptotic variance of $\hat{\delta}_{gmm}$ is relatively smaller than that of $\hat{\delta}_{2sls}$. Thus, the optimum GMM estimator improves efficiency upon the 2SLS estimator by the joint estimation of ρ_0 and δ_0, with additional quadratic moments as expected.

4.3 The Best GMM under Normality

The preceding section has provided a general GMM estimation framework with many linear moments but a finite number of quadratic moments, where the finite number of quadratic moments and the quadratic matrices U_j's satisfying Assumption 6 can be arbitrary. So there remains an issue of the best selection of quadratic moments in this estimation framework. For the estimation of a SAR model, Lee (2007a) has derived the best GMM moments under a normality assumption. Here,

[32] Note that the GMM estimators of λ_0 and ρ_0, in Proposition 5 have the order $O_p(K/n)$, which can be used for the construction of the bias-corrected estimator according to the following proposition.

the best GMM moments can similarly be derived. Let $V = \mathrm{D}(Q_K'Q_K, \sigma_0^2\Upsilon)$. When ϵ is normally distributed, $\mu_3 = 0$ and $\mu_4 = 3\sigma_0^4$. It follows that $\Omega = \mathrm{Var}(g(\theta_0)) = \sigma_0^2 V$, $B_{22} = (\sigma_0^4\Upsilon)^{-1}$ and $\bar{D}_2 = D_2$. It follows Proposition 5 that, when $K^{3/2}/n \to 0$, the (infeasible) optimal GMM estimator $\hat{\theta}_{gmm} = \arg\min_{\theta\in\Theta} g'(\theta)V^{-1}g(\theta)$ has the asymptotic distribution

$$\sqrt{n}(\hat{\theta}_{gmm} - \theta_0 - b_{gmm}) \xrightarrow{d} N(0, \sigma_0^2(\mathrm{D}(0, H) + \sigma_0^{-2}\lim_{n\to\infty}\frac{1}{n}D_2'\Upsilon^{-1}D_2)^{-1}),$$

where the bias

$$b_{gmm} = \sigma_0^2[\mathrm{D}(0, Z'R'P_K RZ) + \sigma_0^{-2}D_2'\Upsilon^{-1}D_2]^{-1}[\mathrm{tr}(P_K MR^{-1}), \mathrm{tr}(\Psi_K)e_1']' = O(K/n).$$

For any $n\times n$ matrix A, let $A^t = A - \mathrm{tr}(A)J/\mathrm{tr}(J)$. Note that $\mathrm{tr}(U_j^s\bar{G}) = \frac{1}{2}\mathrm{tr}(U_j^s[(J\bar{G}J)^t]^s) = \frac{1}{2}\mathrm{vec}'(U_j^s)\mathrm{vec}([(J\bar{G}J)^t]^s)$ and, similarly, $\mathrm{tr}(U_j^s MR^{-1}) = \frac{1}{2}\mathrm{vec}'(U_j^s)\mathrm{vec}([(JMR^{-1}J)^t]^s)$. It follows from the generalized Schwartz inequality that $D_2'\Upsilon^{-1}D_2 \le \sigma_0^4\Sigma_{g_2}$, where

$$\Sigma_{g_2} = \begin{pmatrix} \mathrm{tr}([(JMR^{-1}J)^t]^s MR^{-1}) & \mathrm{tr}([(J\bar{G}J)^t]^s MR^{-1}) & 0 \\ * & \mathrm{tr}([(J\bar{G}J)^t]^s\bar{G}) & 0 \\ * & * & 0 \end{pmatrix}.$$

Hence, under normality, the best choice of U_j's for the quadratic moments are $U_1^* = (JMR^{-1}J)^t$ and $U_2^* = (J\bar{G}J)^t$.

Let $g^*(\theta) = [g_1'(\theta), \epsilon'(\theta)U_1^*\epsilon(\theta), \epsilon'(\theta)U_2^*\epsilon(\theta)]'$ and $V^* = \mathrm{D}(Q_K'Q_K, \sigma_0^2\Upsilon^*)$, where $\Upsilon^* = \frac{1}{2}[\mathrm{vec}(U_1^{*s}), \mathrm{vec}(U_2^{*s})]'\times[\mathrm{vec}(U_1^{*s}), \mathrm{vec}(U_2^{*s})]$. From Proposition 5, if $K^{3/2}/n \to 0$, the best GMM (BGMM) estimator $\hat{\theta}_{bgmm} = \arg\min_{\theta\in\Theta} g^{*\prime}(\theta)V^{*-1}g^*(\theta)$ has the limiting distribution

$$\sqrt{n}(\hat{\theta}_{bgmm} - \theta_0 - b_{bgmm}) \xrightarrow{d} N(0, \sigma_0^2(\mathrm{D}(0, H) + \sigma_0^2\lim_{n\to\infty}\frac{1}{n}\Sigma_{g_2})^{-1}), \tag{11}$$

where the bias

$$b_{bgmm} = \sigma_0^2[\mathrm{D}(0, Z'R'P_K RZ) + \sigma_0^2\Sigma_{g_2}]^{-1}[\mathrm{tr}(P_K MR^{-1}), \mathrm{tr}(\Psi_K)e_1']' = O(K/n). \tag{12}$$

In practice, with initial consistent estimators $\tilde{\lambda}$, $\tilde{\rho}$, and $\tilde{\sigma}^2$, U_1^* can be estimated as $\tilde{U}_1^* =$

$(JM\tilde{R}^{-1}J)^t$, U_2^* can be estimated as $\tilde{U}_2^* = (J\tilde{R}\tilde{G}\tilde{R}^{-1}J)^t$, and $\tilde{V}^* = V^*(\tilde{\sigma}^2, \tilde{\lambda}, \tilde{\rho},)$.[33] The following result shows that the feasible BGMM estimator has the same limiting distribution as $\hat{\theta}_{bgmm}$ given by (11).

Proposition 7 *Under Assumptions 1-3,4',5-9, if the disturbances are normally distributed,* $K^{3/2}/n \to 0$, $\tilde{\lambda} - \lambda_0 = O_p(\max\{1/\sqrt{n}, K/n\})$, $\tilde{\rho} - \rho_0 = O_p(\max\{1/\sqrt{n}, K/n\})$, *and* $\tilde{\sigma}^2$ *is a* $\sqrt{n}$*-consistent initial estimator of* σ_0^2*, then the feasible BGMM estimator* $\hat{\theta}_{fbgmm} = \arg\min_{\theta\in\Theta} \tilde{g}^{*\prime}(\theta)\tilde{V}^{*-1}\tilde{g}^*(\theta)$*, where* $\tilde{g}^*(\theta) = [g_1'(\theta), \epsilon'(\theta)\tilde{U}_1^*\epsilon(\theta), \epsilon'(\theta)\tilde{U}_2^*\epsilon(\theta)]'$*, has the limiting distribution given by (11).*

5 Monte Carlo Experiments

To investigate the finite sample performance of the 2SLS and GMM estimators, we conduct a limited simulation study based on the following model

$$Y = \lambda_0 WY + X\beta_{01} + WX\beta_{02} + \iota\alpha_0 + u, \tag{13}$$

where $u = \rho_0 Mu + \epsilon$. For the experiment, we consider four samples with different numbers of groups $\bar{r}$ and group sizes m_r. The first sample contains 30 groups with equal group sizes of $m_r = 10$. The second sample contains 60 groups with equal group sizes of $m_r = 10$. To study the effect of group sizes, we also consider, respectively, 30 and 60 groups with equal group sizes of $m_r = 15$.[34] For each group, the sociomatrix W_r is generated as follows.[35] First, for the ith row of W_r $(i = 1, \cdots, m_r)$, we generate an integer k_{ri} uniformly at random from the set of integers $[0, 1, 2, 3]$. Then we set the $(i+1)$th, $\cdots$, $(i+k_{ri})$th elements of the ith row of W_r to be ones and the rest elements in that row to be zeros, if $i + k_{ri} \le m_r$; otherwise the entries of ones will be wrapped around such that the first $(k_{ri} - m_r)$ entries of the ith row will be ones. In the case of $k_{ri} = 0$, the ith row of W_r will have all zeros.[36] M is the row-normalized W.

[33] Note that the best $\tilde{U}_1^*$ and $\tilde{U}_2^*$ in $\tilde{g}^*(\theta)$ involve estimates of λ_0 and ρ_0, while the U's in Proposition 5 are constant matrices.

[34] Note that the efficiency of 2SLS estimator needs the average groups size to be large relative to the number of groups, but the efficiency of bias-corrected 2SLS estimator only needs the average group size to be large (see the discussion after Propositions 2 and 3). To make the difference in the finite sample performance of these two estimators more prominent, we generate samples in the simulation study such that $\bar{r}/\bar{m}$ is not too small.

[35] Experiments have also been done based on the Addhealth sociomatrices. As the Add Health data has large groups, the bias is less pronounced. We did not to report those results here to save space.

[36] Note that the parameter space of λ depends on $||W||_\infty$. If the maximum number of direct connections changes with the group size m_r, so does the parameter space of λ. To facilitate comparison, we keep the maximum number of direct connections fixed at 3 across different sample sizes. We have tried different values for the maximum number of direct connections. The simulation results are similar to those reported here.

The number of repetitions is 500 for each case in this Monte Carlo experiment. For each repetition, X and α_{0r} $(r = 1, \cdots, \bar{r})$ are generated from $N(0, I)$ and $N(0, \sigma_\alpha^2)$ respectively. The error terms, $\epsilon_{r,i}$'s, are independently generated from the following 2 distributions: (a) normal, $\epsilon_{r,i} \sim N(0, 1)$ and (b) gamma, $\epsilon_{r,i} = \gamma_{r,i} - 1$ where $\gamma_{r,i} \sim gamma(1, 1)$. The $\epsilon_{r,i}$'s have mean zero and variance one. The skewness ($\eta_3 = \mu_3/\sigma^3$) and kurtosis ($\eta_4 = \mu_4/\sigma^4$) of these distributions are correspondingly: (a) $\eta_3 = 0$, $\eta_4 = 3$ and (b) $\eta_3 = 2$, $\eta_4 = 9$. The data are generated with $\beta_{10} = \beta_{20} = 0.2$.[37] Values of λ_0, ρ_0 and σ_α^2 are varied in the experiment. For each experimental design, we also report the ratio of concentration parameter to sample size as a measure of the quality of instruments to facilitate comparison.

The estimation methods considered are: (i) the 2SLS with a "few" IVs $Q_1 = J[X, WX, MX, MWX]$; (ii) the 2SLS in (6) with "many" IVs $Q_2 = [Q_1, JW\iota]$;[38] (iii) the FC2SLS; (iv) the optimal GMM with $g(\theta) = [Q_1, U_1^* \epsilon(\theta), U_2^* \epsilon(\theta)]' \epsilon(\theta)$, where $U_1^* = (JMR^{-1}J)^t$ and $U_2^* = (J\bar{G}J)^t$; (v) the optimal GMM with $g(\theta) = [Q_2, U_1^* \epsilon(\theta), U_2^* \epsilon(\theta)]' \epsilon(\theta)$; and (vi) the FCGMM. For the 2SLS, we need to estimate ρ_0 in a preliminary step by the MOM, $\tilde{\rho} = \arg\min_\rho \tilde{g}'(\rho)\tilde{g}(\rho)$, where $\tilde{g}(\rho) = [(JWJ)^t \tilde{\epsilon}(\rho), (JMJ)^t \tilde{\epsilon}(\rho), (JMWJ)^t \tilde{\epsilon}(\rho)]' \tilde{\epsilon}'(\rho)$, $\tilde{\epsilon}(\rho) = JR(\rho)(Y - Z'\tilde{\delta})$, and $\tilde{\delta} = [Z'Q_1(Q_1'Q_1)^{-1}Q_1'Z]^{-1} \times Z'Q_1(Q_1'Q_1)^{-1}Q_1'Y$. Under the distribution (a), $[U_1^* \epsilon(\theta), U_2^* \epsilon(\theta)]' \epsilon(\theta)$ are the best quadratic moments. The optimal GMM estimator is given in Proposition 7 and the bias correction procedure is based on the estimated bias given by (12). Under the distribution (b), the optimal GMM estimator is given in Proposition 5 and the estimated bias is given by (10). We report the mean and standard deviation (SD) of the empirical distributions of the estimates. To facilitate the comparison of various estimators, we also report their root mean square errors (RMSE).

We consider four experimental designs. Table 1 reports estimation results for the benchmark case with normally distributed errors and $\lambda_0 = \rho_0 = 0.1$; $\sigma_\alpha^2 = 1$. Table 2 reports the estimation results when $\lambda_0 = \rho_0 = 0.1$ and $\sigma_\alpha^2 = 0.04$, which represents the case that the additional IVs based on the centrality measure are less informative. To study how the estimators are affected by increasing the social dependence in the model, we consider the data generating process with $\lambda_0 = \rho_0 = 0.3$. Table 3 reports the estimation results when $\lambda_0 = \rho_0 = 0.3$ and $\sigma_\alpha^2 = 1$.[39] The quadratic moment

[37] In the previous version of the paper, we have also experimented with different values of β_0. When β_0 takes a larger value, the linear moment would be more informative. In that case, the additional quadratic moments used by GMM do not improve much of the efficiency of the 2SLS except for ρ_0.

[38] For the case with "many" IVs we have also tried to include additional IVs such as $JMW\iota$ and $JW^2\iota$. The estimation results with additional IVs are similar to those reported here with a slightly larger bias in the GMM estimator of ρ_0. And such a bias can be reduced by the proposed bias-correction procedure.

[39] The estimation results where $\lambda_0 = \rho_0 = 0.3$ and $\sigma_\alpha^2 = 0.04$ show a similar pattern.

conditions for the GMM estimators in Section 4.3 are the best under the normality assumption. To study the robustness of the GMM estimator, we consider the case where the disturbances follow the gamma distribution.[40] The estimation results when $\lambda_0 = \rho_0 = 0.1$ and $\sigma_\alpha^2 = 1$ are reported in Table 4. The simulation results are summarized as follows.

[Tables 1-4 approximately here]

(1) The additional linear moment conditions based on the centrality measure substantially reduce SDs in 2SLS estimators of λ_0, β_{02} and GMM estimators of $\lambda_0, \rho_0, \beta_{02}$. For the benchmark results reported in Table 1, SD reductions of 2SLS estimators of λ_0 and β_{02} with Q_2 relative to those with Q_1 are, respectively, about 69% and 23% when $m_r = 10$ and $\bar{r} = 30$. For the case in Table 2 when the additional IVs are less informative, SD reductions are small relative to those of the benchmark case. For the case when the social dependence is stronger, we find in Table 3 that SD reductions of both 2SLS and GMM estimators of λ_0 are larger than those in Table 1. For the case under non-normality in Table 4, we observe a similar pattern.

(2) The additional instruments in Q_2 introduce biases into 2SLS estimators of λ_0, β_{02} and GMM estimators of λ_0, ρ_0. For the benchmark case reported in Table 1, the 2SLS estimator of λ_0 with Q_2 is downwards biased by 38% and that of β_{02} is upwards biased by 7% when $m_r = 10$ and $\bar{r} = 30$. The magnitude of the bias reduces as the group size increases. The impact of the number of groups on the bias is less obvious. For the case in Table 2 where the additional instruments are less informative, biases introduced by the additional IVs are much larger than those reported in Table 1. For the case in Table 3 where the social dependence is stronger, the "many instrument" bias is smaller than the benchmark case in Table 1.

(3) The proposed bias-correction procedure substantially reduces the many-instrument bias for both the 2SLS and GMM estimators. In Table 1, bias reductions of 2SLS estimators of λ_0 and β_{02} with Q_2 are, respectively, about 79% and 57%, when $m_r = 10$ and $\bar{r} = 30$. For other cases considered in Tables 2-4, we observe a similar pattern.

(4) The optimal GMM improves the efficiency upon the 2SLS method. Those optimal GMM estimators given in (iv) have smaller SDs than corresponding 2SLS estimators with the IV matrix Q_1 in the estimation of λ_0, ρ_0 and β_{02}. In Table 1, SD reductions of GMM estimators of λ_0, ρ_0 and β_{02} relative to 2SLS estimators are, respectively, about 43%, 30% and 16%, when $m_r = 10$ and

[40] We also experimented with t distribution and mixed-normal distribution. The simulation results are similar to those reported here.

$\bar{r} = 30$. The optimal GMM estimators with "many" IVs given in (v) have smaller SDs than the 2SLS with the IV matrix Q_2 in the estimation of λ_0 and ρ_0. In Table 1, SD reductions of GMM estimators of λ_0 and ρ_0 relative to 2SLS estimators are, respectively, about 16% and 52%, when $m_r = 10$ and $\bar{r} = 30$. Similarly, for the case in Table 2, the optimal GMM estimators have smaller SDs than 2SLS estimators. Especially, the SD reduction in the GMM estimator of λ_0 with "many" IVs given in (v) relative to the corresponding 2SLS estimator with Q_2 is much larger than that in Table 1. Under non-normality, we find in Table 4 that the quadratic moments in GMM help to reduce SDs of 2SLS estimators. However, when the sample size is small, the magnitude of SD reduction in the estimation of λ_0 is smaller than that in the case under normality. Furthermore, for all cases considered, the optimal GMM estimator of λ_0 with "many" IVs given in (v) is less biased than the 2SLS estimator with the IV matrix Q_2.

6 Conclusion

This paper considers the specification, identification and estimation of social interaction models with network structures and the presence of endogenous, contextual, correlated and group fixed effects. We pay special attention to the role of centrality of members in a network on the identification and estimation of interaction effects on outcomes. The network structure in a group is captured by a sociomatrix W. For the case that W is not row-normalized and the indegrees of its nodes are not all equal, the different positions of members in the network as measured by the Bonacich (1987) centrality provide additional information for identification. In that case, the Bonacich centrality measure for each group can be used as an IV to improve estimation efficiency. However, the number of such IVs depends on the number of groups. If the number of groups grows with the sample size, so does the number of IVs. Taking into account the many possible IVs, we consider the 2SLS and GMM estimation for the model. We show that the proposed estimators can be consistent and asymptotically normal, and they can be efficient relative to those with a finite number of IV moment conditions when the sample size grows fast enough relative to the number of IVs. We also suggest bias-correction procedures for the proposed 2SLS and GMM estimators based on the estimated leading order biases due to the presence of many IVs.

In a social interaction model, sometimes, more than one sociomatrix needs to be introduced to capture different types of relationships. Following a similar strategy as in Lee and Liu (2010), the proposed 2SLS and GMM methods can be easily generalized to estimate a general multi-relational network.

APPENDICES

A Summary of Notations

- A^- denotes the generalized inverse of a square matrix A.

- $\mathrm{D}(A_1, \cdots, A_K)$ is a block diagonal matrix with $m_k \times n_k$ diagonal blocks A_k's.

- For an $n \times n$ matrix $A = [a_{ij}]$, $\mathrm{vec}_D(A) = (a_{11}, \cdots, a_{nn})'$.

- $Z = (WY, X)$; $\delta_0 = (\lambda_0, \beta_0')'$; $\theta_0 = (\rho_0, \delta_0')'$; $\alpha_0 = (\alpha_{01}, \cdots, \alpha_{0\bar{r}})'$; $\iota = \mathrm{D}(l_{m_1}, \cdots, l_{m_{\bar{r}}})$.

- e_j is the jth unit (column) vector.

- $S(\lambda) = I - \lambda W$; $S = S(\lambda_0)$; $R(\rho) = I - \rho M$; $R = R(\rho_0)$; $G = WS^{-1}$; $\bar{G} = RGR^{-1}$.

- If $M_r l_{m_r} \neq c l_{m_r}$ for any constant c, $J_r = I_{m_r} - (l_{m_r}, M_r l_{m_r})[(l_{m_r}, M_r l_{m_r})'(l_{m_r}, M_r l_{m_r})]^-(l_{m_r}, M_r l_{m_r})'$; Otherwise, $J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l_{m_r}'$. $J = \mathrm{D}(J_1, \cdots, J_{\bar{r}})$.

- For an $n \times n$ matrix A, $A^s = A + A'$ and $A^t = A - \mathrm{tr}(A)J/\mathrm{tr}(J)$.

- $f(\rho) = JR(\rho)\mathrm{E}(Z)$; $f = f(\rho_0) = JR\mathrm{E}(Z)$; $v = J\bar{G}\epsilon$; $H = \lim_{n\to\infty} \frac{1}{n} f'f$.

- $P_K = Q_K(Q_K'Q_K)^- Q_K'$; $\Psi_K = P_K \bar{G}$; $e_f(K) = \frac{1}{n} f'(I - P_K)f$; $\Delta_K = \mathrm{tr}(e_f(K))$.

- $\epsilon(\theta) = JR(\rho)(Y - Z\delta) = d(\theta) + r(\theta)\epsilon$, where $d(\theta) = f(\rho)(\delta_0 - \delta) = f(\delta_0 - \delta) + (\rho_0 - \rho)F(\delta_0 - \delta)$, $r(\theta) = JR(\rho)S(\lambda)S^{-1}R^{-1}$, and $F = JM\mathrm{E}(Z)$.

- $\omega = [\mathrm{vec}_D(U_1), \cdots, \mathrm{vec}_D(U_q)]$; $\Upsilon = \frac{1}{2}[\mathrm{vec}(U_1^s), \cdots, \mathrm{vec}(U_q^s)]'[\mathrm{vec}(U_1^s), \cdots, \mathrm{vec}(U_q^s)]$.

- $\bar{g}_2(\theta) = \frac{\mu_3}{\sigma_0^2}\omega' P_K \epsilon(\theta) - g_2(\theta)$, where $g_2(\theta) = [U_1\epsilon(\theta), \cdots, U_q\epsilon(\theta)]'\epsilon(\theta)$.

- $\bar{D}_2 = D_2 - \frac{\mu_3}{\sigma_0^2}(0, \omega' f)$, $\check{D}_2 = D_2 - \frac{\mu_3}{\sigma^2}(0, \omega' PRZ)$, where

$$D_2 = \mathrm{E}(\frac{\partial}{\partial \theta'} g_2(\theta_0)) = -\sigma_0^2 \begin{pmatrix} \mathrm{tr}(U_1^s M R^{-1}) & \mathrm{tr}(U_1^s \bar{G}) & 0 \\ \vdots & \vdots & \vdots \\ \mathrm{tr}(U_q^s M R^{-1}) & \mathrm{tr}(U_q^s \bar{G}) & 0 \end{pmatrix}.$$

B Some Useful Lemmas

To simplify notations, we drop the K subscript on P_K and Ψ_K. Let MI refer to the Markov Inequality, and CSI to the Cauchy-Schwartz inequality. Let $\|A\| = \sqrt{\text{tr}(A'A)}$ denote the Frobenius (Euclidean) norm for an $m \times n$ matrix A unless noted otherwise.

Lemma B.1 *Under Assumption 5, there exists a $K \times (k+1)$ matrix π_K such that $\frac{1}{n}||E(Z) - Q_K^0\pi_K||^2 \to 0$ as $n, K \to \infty$.*

Proof. For any n-dimensional vector x, the Frobenius norm $||x||$ and the row sum norm $||x||_\infty$ satisfy $\frac{1}{n}||x||^2 \leq (||x||_\infty)^2$. Hence, $\frac{1}{n}||\text{E}(Z) - Q_K^0\pi_K||^2 \leq (||\text{E}(Z) - Q_K^0\pi_K||_\infty)^2 \to 0$ as $n, K \to \infty$. ■

Lemma B.2 (i) $\text{tr}(P) = K$. (ii) *Suppose that $\{A\}$ is a sequence of $n \times n$ UB matrices. For $B = PA$, $\text{tr}(B) = O(K)$, $\text{tr}(B^2) = O(K)$, and $\sum_i (B_{ii})^2 = O(K)$, where B_{ii}'s are diagonal elements of B.*

Proof. (i) Trivial. (ii) First we show that $\text{tr}(B) = O(K)$. By eigenvalue decomposition, $AA' = \Gamma\Lambda\Gamma'$ where Γ is an orthonormal matrix and Λ is the eigenvalue matrix. It follows that $PAA'P \leq \lambda_{\max}P$, where $\lambda_{\max}$ is the largest eigenvalue. By the spectral radius theorem, $\text{tr}(PAA'P) \leq ||AA'||_1\text{tr}(P) = O(K)$. By CSI, $|\text{tr}(B)| \leq \text{tr}^{1/2}(P)\text{tr}^{1/2}(PAA'P) = O(K)$. By CSI, $\text{tr}(B^2) \leq \text{tr}(BB') = \text{tr}(PAA'P) = O(K)$. The last result holds because $\sum_i (B_{ii})^2 \leq \text{tr}(B'B) = O(K)$. ■

Lemma B.3 (i) $\Delta_K = o(1)$, $\frac{1}{n}\text{tr}[F'(I-P)F] = o(1)$ (ii) $f'(I-P)\epsilon/\sqrt{n} = O_p(\Delta_K^{1/2})$, $f'(I-P)v/\sqrt{n} = O_p(\Delta_K^{1/2})$, $\frac{1}{n}\omega'(I-P)f = O(\Delta_K^{1/2})$, $\frac{1}{n}\omega'(I-P)F = O(\sqrt{\frac{1}{n}\text{tr}[F'(I-P)F]})$, (iii) $\epsilon'B'PA\epsilon = O_p(K)$, $C'PA\epsilon = O_p(\sqrt{nK})$, $C'PD = O(n)$, *where $\{A\}$ and $\{B\}$ are sequences of $n \times n$ UB matrices and the elements of C and D are uniformly bounded constants, and* (iv) $E(v'P\epsilon\epsilon'Pv) = (\mu_4 - 3\sigma_0^4)\sum_i \Psi_{ii}^2 + \sigma_0^4[\text{tr}^2(\Psi) + \text{tr}(\Psi'\Psi) + \text{tr}(\Psi^2)] = \sigma_0^4\text{tr}^2(\Psi) + O_p(K)$.

Proof. The first part of (i) follows by Lemma A.3 (i) in Donald and Newey (2001) and Lemma B.1. For the second part of (i), by Assumption 5 and Lemma B.1, $\frac{1}{n}\text{tr}(F'(I-P)F) = \frac{1}{n}\text{tr}[(F - JMQ_K^0\pi_K)'(I-P)(F - JMQ_K^0\pi_K)] \leq \frac{1}{n}\text{tr}[(F - JMQ_K^0\pi_K)'(F - JMQ_K^0\pi_K)] \to 0$. The first part of (ii) is Lemma A.3 (ii) in Donald and Newey (2001). For the second half of (ii), $\text{Var}[\frac{1}{\sqrt{n}}f'(I-P)J\bar{G}\epsilon] = \frac{\sigma_0^2}{n}f'(I-P)J\bar{G}\bar{G}'J(I-P)f \leq \frac{\sigma_0^2}{n}f'(I-P)f \cdot ||J\bar{G}\bar{G}'J||_\infty = O(\Delta_K)$. So $\frac{1}{\sqrt{n}}f'(I-P)v = O_p(\Delta_K^{1/2})$. By CSI, $\frac{1}{n}|e_i'\omega'(I-P)fe_j| \leq \sqrt{\frac{1}{n}e_i'\omega'\omega e_i}\sqrt{\frac{1}{n}e_j'f'(I-P)fe_j} = O(\Delta_K^{1/2})$, which implies $\frac{1}{n}\omega'(I-P)f = O(\Delta_K^{1/2})$. By the same argument, $\frac{1}{n}\omega'(I-P)F = O(\sqrt{\frac{1}{n}\text{tr}[F'(I-P)F]})$. For (iii), as $\mathrm{E}|\epsilon' B' PA\epsilon| \leq [\mathrm{E}(\epsilon' B' PB\epsilon)]^{1/2}[\mathrm{E}(\epsilon' A' PA\epsilon)]^{1/2} = \sigma_0^2[\mathrm{tr}(B'PB)\mathrm{tr}(A'PA)]^{1/2} = O(K)$ by Lemma B.2 (ii), $\epsilon' B' PA\epsilon = O_p(K)$ by MI. By CSI, $|e_j' C' PA\epsilon| \leq \sqrt{e_j' C' C e_j}\sqrt{\epsilon' A' PA\epsilon} = O_p(\sqrt{nK})$. Hence, $C' PA\epsilon = O_p(\sqrt{nK})$. By CSI, $|e_i' C' PDe_j| \leq \sqrt{e_i' C' C e_i}\sqrt{e_j' D' PDe_j} = O(n)$, which implies $C'PD = O(n)$. For (iv), $\mathrm{E}(v' P\epsilon\epsilon' Pv) = \mathrm{E}(\epsilon' \Psi' \epsilon\epsilon' \Psi\epsilon) = (\mu_4 - 3\sigma_0^4)\sum_i \Psi_{ii}^2 + \sigma_0^4[\mathrm{tr}^2(\Psi) + \mathrm{tr}(\Psi'\Psi) + \mathrm{tr}(\Psi^2)] = \sigma_0^4 \mathrm{tr}^2(\Psi) + O_p(K)$ by Lemma B.2. ■

Lemma B.4 *Suppose that $\{A\}$ and $\{B\}$ are sequences $n \times n$ UB matrices and $K/n = O(1)$. Then* (i) $\frac{1}{n} Z' R' PRZ = \frac{1}{n} f' f + O_p(\sqrt{K/n}) + O(\Delta_K) = O_p(1)$, $\frac{1}{n}\omega' PRZ = \frac{1}{n}\omega' f + O_p(\sqrt{K/n}) + O(\Delta_K^{1/2}) = O_p(1)$, (ii) $[Z' R' P\epsilon - \sigma_0^2 \mathrm{tr}(\Psi) e_1]/\sqrt{n} = f'\epsilon/\sqrt{n} + O_p(\sqrt{K/n}) + O_p(\Delta_K^{1/2}) = O_p(1)$, (iii) $\frac{1}{n} Z' B' PA\epsilon = O_p(\sqrt{K/n})$, $\frac{1}{n} Z' B' PAZ = O_p(1)$, $\frac{1}{n}\omega' PAZ = O_p(1)$.

Proof. For (i), as $JRZ = f + ve_1'$, $\frac{1}{n} Z' R' PRZ = \frac{1}{n} f' f - e_f(K) + \frac{1}{n} v' Pve_1 e_1' + \frac{1}{n}(f' Pve_1')^s$, where $e_f(K) = O(\Delta_K)$, and $\frac{1}{n} v' Pv = O_p(K/n)$ and $\frac{1}{n} f' Pv = O_p(\sqrt{K/n})$ by Lemma B.3 (iii). $\frac{1}{n}\omega' PRZ = \frac{1}{n}\omega' Pf + \frac{1}{n}\omega' Pve_1'$, where $\frac{1}{n}\omega' Pf = \frac{1}{n}\omega' f - \frac{1}{n}\omega'(I-P)f = \frac{1}{n}\omega' f + O(\Delta_K^{1/2})$ and $\frac{1}{n}\omega' Pve_1' = O_p(\sqrt{K/n})$ by Lemma B.3 (iii). For (ii), as $JRZ = f + ve_1'$, $[Z' R' P\epsilon - \sigma_0^2 \mathrm{tr}(\Psi) e_1]/\sqrt{n} = f'\epsilon/\sqrt{n} - f'(I-P)\epsilon/\sqrt{n} + [v' P\epsilon - \sigma_0^2 \mathrm{tr}(\Psi)] e_1/\sqrt{n}$. By Lemma B.3 (ii), $f'(I-P)\epsilon/\sqrt{n} = O_p(\Delta_K^{1/2})$. As $\mathrm{E}(v' P\epsilon\epsilon' Pv) = \sigma_0^4 \mathrm{tr}^2(\Psi) + O(K)$ by Lemma B.3 (iv) and $\mathrm{E}(v' P\epsilon) = \sigma_0^2 \mathrm{tr}(\Psi)$, we have $[v' P\epsilon - \sigma_0^2 \mathrm{tr}(\Psi)]/\sqrt{n} = O_p(\sqrt{K/n})$. For (iii), as $Z = [G(X\beta_0 + \iota\alpha_0), X] + GR^{-1}\epsilon e_1'$, the result follows by Lemma B.3 (iii). ■

Lemma B.5 *Suppose $\tilde{\rho}$ is a consistent estimator of ρ_0 and $\tilde{R} = R(\tilde{\rho})$. Then,* (i) $\frac{1}{n} Z' \tilde{R}' P\tilde{R} Z = \frac{1}{n} Z' R' PRZ + O_p(\tilde{\rho} - \rho_0)$, (ii) $\frac{1}{n}\omega' P\tilde{R} Z = \frac{1}{n}\omega' PRZ + O_p(\tilde{\rho} - \rho_0)$, (iii) $\frac{1}{n} Z' \tilde{R}' P\tilde{R} R^{-1}\epsilon = \frac{1}{n} Z' R' P\epsilon + O((\tilde{\rho} - \rho_0)\sqrt{K/n})$.

Proof. As $\tilde{R} = R - (\tilde{\rho} - \rho_0)M$, we have $\frac{1}{n} Z' \tilde{R}' P\tilde{R} Z = \frac{1}{n} Z' R' PRZ - (\tilde{\rho} - \rho_0)\frac{1}{n}(Z' M' PRZ)^s + (\tilde{\rho} - \rho_0)^2 \frac{1}{n} Z' M' PMZ$, $\frac{1}{n}\omega' P\tilde{R} Z = \frac{1}{n}\omega' PRZ - (\tilde{\rho} - \rho_0)\frac{1}{n}\omega' PMZ$, and $\frac{1}{n} Z' \tilde{R}' P\tilde{R} R^{-1}\epsilon = \frac{1}{n} Z' R' P\epsilon - (\tilde{\rho} - \rho_0)\frac{1}{n}(Z' M' P\epsilon + Z' R' PMR^{-1}\epsilon) + (\tilde{\rho} - \rho_0)^2 \frac{1}{n} Z' M' PMR^{-1}\epsilon$. The results hold as $\frac{1}{n} Z' M' PRZ = O_p(1)$, $\frac{1}{n} Z' M' PMZ = O_p(1)$, $\frac{1}{n}\omega' PMZ = O_p(1)$, $\frac{1}{n}(Z' M' P\epsilon + Z' R' PMR^{-1}\epsilon) = O_p(\sqrt{K/n})$, and $\frac{1}{n} Z' M' PMR^{-1}\epsilon = O_p(\sqrt{K/n})$ by Lemma B.4. ■

Lemma B.6 *$\epsilon(\theta) = JR(\rho)(Y - Z\delta) = d(\theta) + r(\theta)\epsilon$, where $d(\theta) = f(\delta_0 - \delta) + (\rho_0 - \rho)F(\delta_0 - \delta)$, $F = JME(Z)$, and $r(\theta) = JR(\rho)S(\lambda)S^{-1}R^{-1} = J + (\rho_0 - \rho)JMR^{-1} + (\lambda_0 - \lambda)J\bar{G} + (\rho_0 - \rho)(\lambda_0 - \lambda)JMR^{-1}\bar{G}$.*

Proof. As $JR(\rho)\iota = 0$, $S(\lambda)S^{-1} = I + (\lambda_0 - \lambda)G$ and $R(\rho) = R + (\rho_0 - \rho)M$, it follows that $\epsilon(\theta) = JR(\rho)(Y - Z\delta) = JR(\rho)[X(\beta_0 - \beta) + (\lambda_0 - \lambda)G(X\beta_0 + \iota\alpha_0) + S(\lambda)S^{-1}u] = d(\theta) + r(\theta)\epsilon$, where $d(\theta) = JR(\rho)\mathrm{E}(Z)(\delta_0 - \delta) = JR\mathrm{E}(Z)(\delta_0 - \delta) + (\rho_0 - \rho)JM\mathrm{E}(Z)(\delta_0 - \delta)$, and $r(\theta) = JR(\rho)S(\lambda)S^{-1}R^{-1}$. ■

Lemma B.7 *If $\bar{\theta} - \theta_0 = O_p(K/n)$, where $K/n \to c < \infty$ and $c \geq 0$, then* (i) $\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'} = \mathrm{D}(0, \frac{1}{n}Z'R'PRZ) + O_p(\sqrt{K/n})$, (ii) $\frac{1}{n}\frac{\partial\bar{g}_2(\bar{\theta})}{\partial\theta'} = \frac{1}{n}\check{D}_2 + O_p(\sqrt{K/n})$, (iii) $\frac{1}{n}\frac{\partial^2\epsilon'(\bar{\theta})}{\partial\theta'\partial\theta}P\epsilon(\bar{\theta}) = O_p(\sqrt{K/n})$, $\frac{1}{n}\omega'P\frac{\partial^2\epsilon(\bar{\theta})}{\partial\theta\partial\theta'} = O_p(1)$, $\frac{1}{n}\frac{\partial^2 g_2(\bar{\theta})}{\partial\theta\partial\theta'} = O_p(1)$, *and* (iv) $\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}\frac{1}{n}\omega'P\epsilon(\bar{\theta}) - \frac{1}{n}g_2(\bar{\theta}) = O(\sqrt{K/n})$.

If $\bar{\theta} - \theta_0 = o_p(1)$, then (v) $\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'} = \mathrm{D}(0, H) + o_p(1)$, (vi) $\frac{1}{n}\frac{\partial\bar{g}_2(\bar{\theta})}{\partial\theta'} = \lim_{n\to\infty}\frac{1}{n}\bar{D}_2 + o_p(1)$, (vii) $\frac{1}{n}\frac{\partial^2\epsilon'(\bar{\theta})}{\partial\theta'\partial\theta}P\epsilon(\bar{\theta}) = o_p(1)$, $\frac{1}{n}\omega'P\frac{\partial^2\epsilon(\bar{\theta})}{\partial\theta\partial\theta'} = O_p(1)$, $\frac{1}{n}\frac{\partial^2 g_2(\bar{\theta})}{\partial\theta\partial\theta'} = O_p(1)$, *and* (viii) $\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}\frac{1}{n}\omega'P\epsilon(\bar{\theta}) - \frac{1}{n}g_2(\bar{\theta}) = o_p(1)$.

Proof. Here we show the first set of results holds when $\bar{\theta} - \theta_0 = O_p(K/n)$. The second set of results under $\bar{\theta} - \theta_0 = o_p(1)$ is a straightforward extension.

For (i), as $JM\iota = 0$ and $R(\rho) = R + (\rho_0 - \rho)M$, we have $\frac{\partial\epsilon(\theta)}{\partial\theta'} = -J[Mu + MZ(\delta_0 - \delta), RZ + (\rho_0 - \rho)MZ]$. As $u = R^{-1}\epsilon$, by Lemma B.3 and B.4, $\frac{1}{n}u'M'PMu = O_p(K/n)$, $\frac{1}{n}u'M'P^sMZ = O_p(\sqrt{K/n})$, and $\frac{1}{n}Z'M'PMZ = O_p(1)$. As $\bar{\theta} - \theta_0 = O_p(K/n)$, we have $\frac{1}{n}[Mu + MZ(\delta_0 - \bar{\delta})]'P[Mu + MZ(\delta_0 - \bar{\delta})] = O_p(K/n)$. Similarly, $\frac{1}{n}[RZ + (\rho_0 - \bar{\rho})MZ]'P[Mu + MZ(\delta_0 - \bar{\delta})] = O_p(\sqrt{K/n})$, and $\frac{1}{n}[RZ + (\rho_0 - \bar{\rho})MZ]'P[RZ + (\rho_0 - \bar{\rho})MZ] = \frac{1}{n}Z'R'PRZ + O_p(K/n)$. Hence, $\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'} = \mathrm{D}(0, \frac{1}{n}Z'R'PRZ) + O_p(\sqrt{K/n})$.

For (ii), $\frac{1}{n}\frac{\partial\bar{g}_2(\bar{\theta})}{\partial\theta'} = \frac{\mu_3}{n\sigma_0^2}\omega'P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'} + \frac{1}{n}\frac{\partial g_2(\bar{\theta})}{\partial\theta'}$. By Lemma B.3 (iii) $\frac{1}{n}\omega'PMu = O_p(\sqrt{K/n})$, and by Lemma B.4 $\frac{1}{n}\omega'PMZ = O_p(1)$. It follows that $\frac{1}{n}\omega'P[Mu + MZ(\delta_0 - \bar{\delta})] = O_p(\sqrt{K/n})$, and $\frac{1}{n}\omega'P[RZ + (\rho_0 - \rho)MZ] = \frac{1}{n}\omega'PRZ + O_p(K/n)$. Hence $\frac{1}{n}\omega'P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'} = -(0, \frac{1}{n}\omega'PRZ) + O_p(\sqrt{K/n})$. On the other hand, $\frac{1}{n}\frac{\partial g_2(\bar{\theta})}{\partial\theta'} = \frac{1}{n}[U_1^s\epsilon(\bar{\theta}), \cdots, U_q^s\epsilon(\bar{\theta})]'\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'}$, where $\frac{1}{n}\epsilon'(\bar{\theta})U_j\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'} = -\frac{1}{n}\epsilon'(\bar{\theta})U_j[Mu + MZ(\delta_0 - \bar{\delta}), RZ + (\rho_0 - \bar{\rho})MZ]$. By Lemma B.6 and Lee (2004), $\frac{1}{n}d'(\bar{\theta})U_jMu = \frac{1}{n}(\delta_0 - \bar{\delta})'f'U_jMu + \frac{1}{n}(\rho_0 - \rho)(\delta_0 - \bar{\delta})'F'U_jMu = O_p((\bar{\delta} - \delta_0)/\sqrt{n})$, $\frac{1}{n}\epsilon'r'(\bar{\theta})U_jMu = \frac{1}{n}\epsilon'U_jMu + (\rho_0 - \rho)\frac{1}{n}\epsilon'(MR^{-1})'U_jMu + (\lambda_0 - \lambda)\frac{1}{n}\epsilon'\bar{G}'U_jMu + (\rho_0 - \rho)(\lambda_0 - \lambda)\frac{1}{n}\epsilon'(MR^{-1}\bar{G})'U_jMu = \frac{\sigma_0^2}{n}\mathrm{tr}(U_jMR^{-1}) + O_p(\max\{(\bar{\theta} - \theta_0), 1/\sqrt{n}\})$, $\frac{1}{n}d'(\bar{\theta})U_jMZ(\delta_0 - \bar{\delta}) = O_p((\bar{\delta} - \delta_0)^2)$, and $\frac{1}{n}\epsilon'r'(\bar{\theta})U_jMZ(\delta_0 - \bar{\delta}) = O_p((\bar{\delta} - \delta_0))$. It follows that $\epsilon'(\bar{\theta})U_j[Mu + MZ(\delta_0 - \bar{\delta})] = \frac{1}{n}\mathrm{tr}(U_jMR^{-1}) + O_p(\max\{1/\sqrt{n}, K/n\})$. Similarly, as $\frac{1}{n}d'(\bar{\theta})U_jRZ = O_p((\bar{\delta} - \delta_0))$, $\frac{1}{n}\epsilon'r'(\bar{\theta})U_jRZ = \frac{\sigma_0^2}{n}\mathrm{tr}(U_j\bar{G}) + O_p(\max\{(\bar{\theta} - \theta_0), 1/\sqrt{n}\})$, $(\rho_0 - \bar{\rho})\frac{1}{n}d'(\bar{\theta})U_jMZ = O_p((\rho_0 - \bar{\rho})(\bar{\delta} - \delta_0))$, and $(\rho_0 - \bar{\rho})\frac{1}{n}\epsilon'r'(\bar{\theta})U_jMZ = O_p((\rho_0 - \bar{\rho}))$, we have $-\frac{1}{n}[d(\bar{\theta}) + r(\bar{\theta})\epsilon]'U_j[RZ + (\rho_0 - \bar{\rho})MZ] = \frac{1}{n}\mathrm{tr}(U_j\bar{G}) + O_p(\max\{1/\sqrt{n}, K/n\})$. Hence, $\frac{1}{n}\frac{\partial g_2(\bar{\theta})}{\partial\theta'} =$

$\frac{1}{n}D_2 + O_p(\max\{1/\sqrt{n}, K/n\})$. In summary, $\frac{1}{n}\frac{\partial \bar{g}_2(\bar{\theta})}{\partial \theta'} = \frac{1}{n}\check{D}_2 + O_p(\sqrt{K/n})$.

For (iii), $\frac{\partial^2}{\partial \delta' \partial \rho}\epsilon(\theta) = JMZ$. By Lemma B.4 $\frac{1}{n}Z'M'Pd(\bar{\theta}) = \frac{1}{n}Z'M'Pf(\delta_0 - \bar{\delta}) + (\rho_0 - \bar{\rho})\frac{1}{n}Z'M'PF(\delta_0 - \bar{\delta}) = O_p(\bar{\delta} - \delta_0)$, and $\frac{1}{n}Z'M'Pr(\bar{\theta})\epsilon = \frac{1}{n}Z'M'P\epsilon + (\lambda_0 - \bar{\lambda})\frac{1}{n}Z'M'P\bar{G}\epsilon + (\rho_0 - \bar{\rho})\frac{1}{n}Z'M'PMR^{-1}\epsilon + (\rho_0 - \bar{\rho})(\lambda_0 - \bar{\lambda})\frac{1}{n}Z'M'PMR^{-1}\bar{G}\epsilon = O_p(\sqrt{K/n})$. Hence, $\frac{1}{n}Z'M'P\epsilon(\bar{\theta}) = O_p(\sqrt{K/n})$, which implies that $\frac{1}{n}\frac{\partial^2 \epsilon'(\bar{\theta})}{\partial \theta' \partial \theta}P\epsilon(\bar{\theta}) = O_p(\sqrt{K/n})$. Similarly, $\frac{1}{n}\omega' P\frac{\partial^2 \epsilon(\bar{\theta})}{\partial \theta \partial \theta'} = O_p(1)$ by Lemma B.4. On the other hand, $\frac{1}{n}\frac{\partial^2 g_2(\bar{\theta})}{\partial \theta \partial \theta'} = \frac{1}{n}[U_1^s\frac{\partial \epsilon(\bar{\theta})}{\partial \theta'}, \cdots, U_q^s\frac{\partial \epsilon(\bar{\theta})}{\partial \theta'}]'\frac{\partial \epsilon(\bar{\theta})}{\partial \theta'} + \frac{1}{n}[U_1^s\epsilon(\bar{\theta}), \cdots, U_q^s\epsilon(\bar{\theta})]'\frac{\partial^2 \epsilon(\bar{\theta})}{\partial \theta \partial \theta'} = O_p(1)$.

For (iv), as $\frac{1}{n}\omega' Pd(\bar{\theta}) = O(\bar{\delta} - \delta_0)$ and $\frac{1}{n}\omega' Pr(\bar{\theta})\epsilon = O(\sqrt{K/n})$, it follows that $\frac{1}{n}\omega' P\epsilon(\bar{\theta}) = O(\sqrt{K/n})$. On the other hand, by Lemma B.6 and Lee (2004), $\frac{1}{n}d'(\bar{\theta})U_j d(\bar{\theta}) = O_p((\delta_0 - \bar{\delta})^2)$, $\frac{1}{n}d'(\bar{\theta})U_j r(\bar{\theta})\epsilon = O_p((\delta_0 - \bar{\delta})/\sqrt{n})$ and $\frac{1}{n}\epsilon' r'(\bar{\theta})U_j r(\bar{\theta})\epsilon = O_p(1/\sqrt{n})$. It follows that $\frac{1}{n}g_2(\bar{\theta}) = \frac{1}{n}[U_1^s\epsilon(\bar{\theta}), \cdots, U_q^s\epsilon(\bar{\theta})]'\epsilon(\bar{\theta}) = O_p(\max\{(\delta_0 - \bar{\delta})^2, 1/\sqrt{n}\})$. In summary, $\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}\frac{1}{n}\omega' P\epsilon(\bar{\theta}) - \frac{1}{n}g_2(\bar{\theta}) = O(\sqrt{K/n})$. ■

Lemma B.8 *Suppose that $K^{3/2}/n \to 0$, z_1 and z_2 are n-dimensional column vectors of uniformly bounded constants, the sequence of $n \times n$ constant matrices $\{A\}$ is UBC, and $\{B_1\}$ and $\{B_2\}$ are UB, and $\epsilon_1, \cdots, \epsilon_n$ are i.i.d. with zero mean and finite second moment. $(\hat{\alpha} - \alpha_0) = O_p(\max\{1/\sqrt{n}, K/n\})$ where α_0 is a p-dimensional vector in the interior of its convex parameter space. For notational simplicity, denote $(\hat{\alpha} - \alpha_0)^{<i>} = \sum_{j_1=1}^{p} \cdots \sum_{j_i=1}^{p}(\hat{\alpha}_{j_1} - \alpha_{j_1 0}) \cdots (\hat{\alpha}_{j_i} - \alpha_{j_i 0})$. The matrix $C(\hat{\alpha})$ has the expansion that*

$$C(\hat{\alpha}) - C(\alpha_0) = \sum_{i=1}^{m-1}(\hat{\alpha} - \alpha_0)^{<i>}K_i(\alpha_0) + (\hat{\alpha} - \alpha_0)^{<m>}K_m(\hat{\alpha}), \tag{14}$$

for some $m \geq 2$, where $\{C(\alpha_0)\}$ and $\{K_i(\alpha_0)\}$ are UB for $i = 1, \cdots, m-1$, and $\{K_m(\alpha)\}$ is UB uniformly in a small neighborhood of α_0. Then, for $\Delta_1 = C(\hat{\alpha}) - C(\alpha_0)$, (i) *$\frac{1}{n}z_1'\Delta_1 z_2 = o_p(1)$;* (ii) *$\frac{1}{\sqrt{n}}z_1'\Delta_1 A\epsilon = o_p(1)$;* (iii) *if (14) holds for $m > 2$, $\frac{1}{n}\epsilon' B_1'\Delta_1 B_2\epsilon = o_p(1)$; and* (iv) *if (14) holds for $m > 4$ with $\mathrm{tr}(K_i(\alpha_0)) = 0$ for $i = 1, \cdots, m-1$, $\frac{1}{\sqrt{n}}\epsilon'\Delta_1\epsilon = o_p(1)$.*

Furthermore, suppose another matrix $D(\hat{\gamma})$ has the expansion that

$$D(\hat{\gamma}) - D(\gamma_0) = \sum_{i=1}^{m-1}(\hat{\gamma} - \gamma_0)^{<i>}L_i(\gamma_0) + (\hat{\gamma} - \gamma_0)^{<m>}L_m(\hat{\gamma}), \tag{15}$$

for some $m \geq 2$, where all the components on the right hand side have the same properties of corresponding ones in (14). Then, for $\Delta_2 = (C(\hat{\alpha}) - C(\alpha_0))(D(\hat{\gamma}) - D(\gamma_0))$, (v) *$\frac{1}{n}z_1'\Delta_2 z_2 = o_p(1)$;* (vi) *$\frac{1}{\sqrt{n}}z_1'\Delta_2 A\epsilon = o_p(1)$;* (vii) *if (14) and (15) hold for $m > 2$, $\frac{1}{n}\epsilon' B_1'\Delta_2 B_2\epsilon = o_p(1)$; and* (viii) *if*

(14) and (15) hold for $m > 4$ with $\mathrm{tr}(K_i(\alpha_0) L_j(\gamma_0)) = 0$ *for* $i, j = 1, \cdots, m-1$, $\frac{1}{\sqrt{n}}\epsilon' \Delta_2 \epsilon = o_p(1)$.

Proof. If $K^2/n \to c < \infty$, then $(\hat{\alpha} - \alpha_0) = O_p(1/\sqrt{n})$, and, so, the results follow Lemma D.10 in Liu et al. (2009). For the case that $K^2/n \to \infty$ and $K^{3/2}/n \to 0$ such that $(\hat{\alpha} - \alpha_0) = O_p(K/n)$, the proof follows a similar argument as that in Lemma D.10 in Liu et al. (2009). Here, we only detail the proof of (ii). Let $U = \frac{1}{\sqrt{n}} z_1'(C(\hat{\alpha}) - C(\alpha_0))A\epsilon$. Then, with (14), $U = U_1 + U_2$ where $U_1 = \sum_{i=1}^{m-1}(\hat{\alpha} - \alpha_0)^{<i>} \frac{1}{\sqrt{n}} z_1' K_i(\alpha_0) A\epsilon = o_p(1)$, because $\frac{1}{\sqrt{n}} z_1' K_i(\alpha_0) A\epsilon = O_p(1)$ by Lee (2004), and $U_2 = (\hat{\alpha} - \alpha_0)^{<m>} \frac{1}{\sqrt{n}} z_1' K_m(\hat{\alpha}) A\epsilon$. Because the product of UBC matrices is UBC, $||K_m(\hat{\alpha})A||_1 \le c_1$ for some constant c_1 for all n. As elements of z_1 are uniformly bounded, $\|z_1'\|_1 \le c_2$ for some constant c_2. It follows that $||U_2||_1 \le n^{1/2}||\hat{\alpha} - \alpha_0||_1^m \cdot ||z_1'||_1 \cdot ||K_m(\hat{\alpha})A||_1 \cdot \frac{1}{n}||\epsilon||_1 \le c_1 c_2 n^{1/2} ||\hat{\alpha} - \alpha_0||_1^m \cdot (\frac{1}{n}\sum_{i=1}^n |\epsilon_i|)$. As $K^{3/2}/n \to 0$, $U_2 = o_p(1)$ for $m \ge 2$ because $(\hat{\alpha} - \alpha_0) = O_p(K/n)$ and $\frac{1}{n}\sum_{i=1}^n |\epsilon_i| = O_p(1)$ by the strong law of large numbers. The desired result follows. ■

Lemma B.9 *Suppose that $K^{3/2}/n \to 0$, z_1 and z_2 are n-dimensional column vectors of constants which are uniformly bounded, the sequence of $n \times n$ constant matrices $\{A\}$ is UBC, $\{B_1\}$ and $\{B_2\}$ are UB, and $\epsilon_1, \cdots, \epsilon_n$ are i.i.d. with zero mean and finite fourth moment.* $(\hat{\theta} - \theta_0) = O_p(\max\{1/\sqrt{n}, K/n\})$. *Let C be either MR^{-1} or $\bar{G}$, and let $\hat{C}$ be its estimated counterpart with $\hat{\theta}$ for θ_0. For $\Delta = \hat{C} - C$, Δ^L represents its linearly transformed matrix which preserves the UB property. Then, we have* (i) $\frac{1}{n} z_1' \Delta^L z_2 = o_p(1)$, $\frac{1}{\sqrt{n}} z_1' \Delta^L A\epsilon = o_p(1)$, $\frac{1}{n}\epsilon' B_1' \Delta^L B_2 \epsilon = o_p(1)$, $\frac{1}{\sqrt{n}}\epsilon' \Delta^t \epsilon = o_p(1)$; (ii) $\frac{1}{n}\mathrm{tr}(A'\Delta^L) = o_p(1)$. *In addition, if $\{D(\gamma)\}$ is UB uniformly in a small neighborhood of γ_0 that is in the interior of its parameter space, then* (iii) $\frac{1}{n}\mathrm{tr}[D'(\hat{\gamma})\Delta^L] = o_p(1)$, *where* $\hat{\gamma} - \gamma_0 = o_p(1)$.

Proof. Based on Lemma B.8, with an estimator of order $O_p(\max\{1/\sqrt{n}, K/n\})$, Lemma B.9 follows the same argument as in the proof of Lemma D.11 in Liu et al. (2009). ■

C Proofs

Proof of Lemma 3.1. Let $\lambda_0^{(p)} = (1, \lambda_0, \cdots, \lambda_0^p)$, and $\otimes$ denote the Kronecker product. For

$$\pi_K^{(p)} = \begin{pmatrix} \lambda_0^{(p)} \otimes \beta_0' & -\lambda_0^{(p)} \otimes (\rho_0 \beta_0') & \lambda_0^{(p)} \otimes \alpha_0' & -\lambda_0^{(p)} \otimes (\rho_0 \alpha_0') & 0 & 0 \\ 0 & 0 & 0 & 0 & I_k & -\rho_0 I_k \end{pmatrix}',$$

$Q_K^{(p)} \pi_K^{(p)} = JR[(\sum_{j=0}^p \lambda_0^j W^{j+1})(X\beta_0 + \iota\alpha_0), X]$. Hence, when $\sup ||\lambda_0 W||_\infty < 1$, it follows that $||f - Q_K^{(p)}\pi_K^{(p)}||_\infty = ||JR[\lambda_0^{p+1} W^{p+2} S^{-1}(X\beta_0 + \iota\alpha_0), 0]||_\infty = ||JR||_\infty \cdot ||\lambda_0 W||_\infty^{p+1} \cdot ||G(X\beta_0 + \iota\alpha_0)||_\infty = o(1)$ as $p \to \infty$. ■

Proof of Proposition 1. As $J\tilde{R}\iota = 0$, $\sqrt{n}(\hat{\delta}_{2sls} - \delta_0) = \sqrt{n}(Z'\tilde{R}'P\tilde{R}Z)^{-1}Z'\tilde{R}'P\tilde{R}R^{-1}\epsilon$. By Lemmas B.5 and B.4, $\frac{1}{n}Z'\tilde{R}'P\tilde{R}Z = \frac{1}{n}Z'R'PRZ + O_p(1/\sqrt{n}) = H + o_p(1)$. As $\sqrt{n}(\tilde{\rho} - \rho_0) = O_p(1)$, by Lemmas B.5 and B.4, $[Z'\tilde{R}'P\tilde{R}R^{-1}\epsilon - \sigma_0^2\text{tr}(\Psi)e_1]/\sqrt{n} = [Z'R'P\epsilon - \sigma_0^2\text{tr}(\Psi)e_1]/\sqrt{n} + O(\sqrt{n}(\tilde{\rho} - \rho_0)\sqrt{K/n}) = f'\epsilon/\sqrt{n} + o_p(1) \xrightarrow{d} N(0, \sigma_0^2 H)$. The conclusion follows by the Slutzky theorem. ■

Proof of Corollary 1. (i) and (ii) are trivial. For (iii), as $0 < \eta < 1$, $K^\eta/\sqrt{n} < \sqrt{K^{1+\eta}/n} \to 0$. As $\sqrt{n}(\hat{\delta}_{2sls} - \delta_0 - b_{2sls}) = O_p(1)$, $(K^\eta/\sqrt{n})\sqrt{n}(\hat{\delta}_{2sls} - \delta_0 - b_{2sls}) = o_p(1)$. Note that $b_{2sls} = O_p(K/n)$ and $K^\eta b_{2sls} = O_p(K^{1+\eta}/n) = o_p(1)$. It follows that $K^\eta(\hat{\delta}_{2sls} - \delta_0) = K^\eta(\hat{\delta}_{2sls} - \delta_0 - b_{2sls}) + K^\eta b_{2sls} = o_p(1)$. ■

Proof of Proposition 2. The 2SLS estimator satisfies $\hat{\delta}_{2sls} - \delta_0 - b_{2sls} = (Z'\tilde{R}'P\tilde{R}Z)^{-1}(Z'\tilde{R}'P\tilde{R}R^{-1}\epsilon - \sigma_0^2\text{tr}(\Psi)e_1) + [(\frac{1}{n}Z'\tilde{R}'P\tilde{R}Z)^{-1} - (\frac{1}{n}Z'R'PRZ)^{-1}]\sigma_0^2\frac{1}{n}\text{tr}(\Psi)e_1$. By Lemmas B.5 and B.4, $(\frac{1}{n}Z'\tilde{R}'P\tilde{R}Z)^{-1} = (\frac{1}{n}Z'R'PRZ)^{-1} + o_p(1) = O_p(1)$. Also $\frac{1}{n}\text{tr}(\Psi) = O(K/n) = O(1)$. By Lemma B.5, $\frac{1}{n}[Z'\tilde{R}'P\tilde{R}R^{-1}\epsilon - \sigma_0^2\text{tr}(\Psi)e_1] = \frac{1}{n}[Z'R'P\epsilon - \sigma_0^2\text{tr}(\Psi)e_1] + O_p((\tilde{\rho} - \rho_0)\sqrt{K/n}) = \frac{1}{n}[Z'R'P\epsilon - \sigma_0^2\text{tr}(\Psi)e_1] + o_p(1)$, where $\frac{1}{n}[Z'R'P\epsilon - \sigma_0^2\text{tr}(\Psi)e_1] = O_p(1/\sqrt{n})$ by Lemma B.4. The conclusion follows by the Slutzky theorem. ■

Proof of Proposition 3. From the proof of Proposition 1, it is sufficient to show that $[\tilde{\sigma}^2\text{tr}(P\tilde{R}\tilde{G}\tilde{R}^{-1}) - \sigma_0^2\text{tr}(\Psi)]/\sqrt{n} = o_p(1)$. As $\tilde{G} - G = (\tilde{\lambda} - \lambda_0)\tilde{G}G$, $\tilde{R} - R = -(\tilde{\rho} - \rho_0)M$, and $\tilde{R}^{-1} - R^{-1} = (\tilde{\rho} - \rho_0)\tilde{R}^{-1}MR^{-1}$, it follows that $\tilde{R}\tilde{G}\tilde{R}^{-1} - RGR^{-1} = (\tilde{R} - R)\tilde{G}\tilde{R}^{-1} + R(\tilde{G} - G)\tilde{R}^{-1} + RG(\tilde{R}^{-1} - R^{-1}) = -(\tilde{\rho} - \rho_0)M\tilde{G}\tilde{R}^{-1} + (\tilde{\lambda} - \lambda_0)R\tilde{G}G\tilde{R}^{-1} + (\tilde{\rho} - \rho_0)RG\tilde{R}^{-1}MR^{-1}$. Hence, $[\tilde{\sigma}^2\text{tr}(P\tilde{R}\tilde{G}\tilde{R}^{-1}) - \sigma_0^2\text{tr}(\Psi)]/\sqrt{n} = \sqrt{n}(\tilde{\sigma}^2 - \sigma_0^2)\text{tr}(P\tilde{G})/n + \sqrt{n}\sigma_0^2\text{tr}[P(\tilde{R}\tilde{G}\tilde{R}^{-1} - RGR^{-1})]/n = O_p(K/n) = o_p(1)$. ■

Proof of Proposition 4. The GMM minimizes $\frac{1}{n}g'(\theta)\tilde{\Omega}^{-1}g(\theta) = \frac{1}{n}g'(\theta)\Omega^{-1}g(\theta) + \frac{1}{n}g'(\theta)(\tilde{\Omega}^{-1} - \Omega^{-1})g(\theta)$, where $\frac{1}{n}g'(\theta)\Omega^{-1}g(\theta) = \sigma_0^{-2}\frac{1}{n}\epsilon'(\theta)P\epsilon(\theta) + [\mu_3\sigma_0^{-2}\frac{1}{n}\omega'P\epsilon(\theta) - \frac{1}{n}g_2(\theta)]'nB_{22}[\mu_3\sigma_0^{-2}\frac{1}{n}\omega'P\epsilon(\theta) - \frac{1}{n}g_2(\theta)]$. First, we show that the minimizer of $\frac{1}{n}g'(\theta)\Omega^{-1}g(\theta)$ in Θ is a consistent estimator of θ_0. As $\epsilon(\theta) = d(\theta) + r(\theta)\epsilon$ by Lemma B.6, $\frac{1}{n}\epsilon'(\theta)P\epsilon(\theta) = \frac{1}{n}d'(\theta)Pd(\theta) + \frac{2}{n}\gamma(\theta) + \frac{1}{n}q(\theta)$, where $\gamma(\theta) = d'(\theta)Pr(\theta)\epsilon$ and $q(\theta) = \epsilon'r'(\theta)Pr(\theta)\epsilon$. $\gamma(\theta)$ is linear in ϵ and is a third degree polynomial of θ. By Lemma B.3 (iii), $\frac{1}{n}\gamma(\theta) = O_p(\sqrt{K/n}) = o_p(1)$ and, similarly, $\frac{1}{n}q(\theta) = O_p(K/n) = o_p(1)$ uniformly in $\theta \in \Theta$. The uniform convergence in probability follows because $\gamma(\theta)$ and $q(\theta)$ are simply polynomial functions of θ and Θ is a bounded set. It follows that

$$
\begin{aligned}
& \frac{1}{n}\epsilon'(\theta)P\epsilon(\theta) = \frac{1}{n}d'(\theta)Pd(\theta) + o_p(1) \\
= \quad & \frac{1}{n}(\delta_0 - \delta)'\{f'f - f'(I-P)f + (\rho_0 - \rho)[f'F - f'(I-P)F]^s \\
& +(\rho_0 - \rho)^2[F'F - F'(I-P)F]\}(\delta_0 - \delta) + o_p(1) \\
= \quad & (\delta_0 - \delta)'[H + (\rho_0 - \rho)\frac{1}{n}(f'F)^s + (\rho_0 - \rho)^2\frac{1}{n}F'F](\delta_0 - \delta) + o_p(1) \\
\rightarrow \quad & (\delta_0 - \delta)' \lim_{n\to\infty} \frac{1}{n} f'(\rho)f(\rho)(\delta_0 - \delta). \qquad (16)
\end{aligned}
$$

By Assumption 4, H is nonsingular, so $(\delta_0 - \delta)' \lim_{n\to\infty} \frac{1}{n} f'(\rho)f(\rho)(\delta_0 - \delta) \geq 0$, with equality iff $\delta = \delta_0$. Similarly, as $\frac{1}{n}\omega' Pr(\theta)\epsilon = O_p(\sqrt{K/n}) = o_p(1)$ uniformly in $\theta \in \Theta$,

$$
\begin{aligned}
& \frac{1}{n}\omega' P\epsilon(\theta) = \frac{1}{n}\omega' Pd(\theta) + o_p(1) \\
= \quad & [\frac{1}{n}\omega' f - \frac{1}{n}\omega'(I-P)f](\delta_0 - \delta) + (\rho_0 - \rho)[\frac{1}{n}\omega' F - \frac{1}{n}\omega'(I-P)F](\delta_0 - \delta) + o_p(1) \\
\rightarrow \quad & \lim_{n\to\infty} \frac{1}{n}\omega' f(\delta_0 - \delta) + (\rho_0 - \rho) \lim_{n\to\infty} \frac{1}{n}\omega' F(\delta_0 - \delta). \qquad (17)
\end{aligned}
$$

Hence, $\frac{1}{n}\omega' P\epsilon(\theta)$ has a limit of zero at $\delta = \delta_0$. Lastly, because the number of quadratic moments does not depend on sample size, by a similar argument in the proof of Proposition 1 of Lee (2007a), $\frac{1}{n}g_2(\theta) - \frac{1}{n}\mathrm{E}[g_2(\theta)] \xrightarrow{p} 0$ uniformly in θ in any bounded set for θ. The consistency follows from the uniform convergence in probability and the identification uniqueness of the limiting function.

It remains to show that $\frac{1}{n}g'(\theta)(\tilde{\Omega}^{-1} - \Omega^{-1})g(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$. $\frac{1}{n}g'(\theta)(\tilde{\Omega}^{-1} - \Omega^{-1})g(\theta) = (\tilde{\sigma}^{-2} - \sigma_0^{-2})\frac{1}{n}\epsilon'(\theta)P\epsilon(\theta) + \frac{1}{n}\epsilon'(\theta)P\omega[(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2})^2 n\tilde{B}_{22} - (\frac{\mu_3}{\sigma_0^2})^2 nB_{22}]\frac{1}{n}\omega' P\epsilon(\theta) - \frac{2}{n}\epsilon'(\theta)P\omega(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{B}_{22} - \frac{\mu_3}{\sigma_0^2}nB_{22})\frac{1}{n}g_2(\theta) + \frac{1}{n}g_2'(\theta)(n\tilde{B}_{22} - nB_{22})\frac{1}{n}g_2(\theta)$. From (16), as $\frac{1}{n}f'PF$ and $\frac{1}{n}F'PF$ are $O(1)$ by Lemma B.3 (iii), $\frac{1}{n}\epsilon'(\theta)P\epsilon(\theta) = O_p(1)$ uniformly in $\theta \in \Theta$. From (17), as $\frac{1}{n}\omega' PF = O(1)$ by Lemma B.3 (iii), $\frac{1}{n}||\omega' P\epsilon(\theta)|| = O_p(1)$ uniformly in $\theta \in \Theta$. By a similar argument in the proof of Proposition 2 in Lee (2007a), $\frac{1}{n}||g_2(\theta)|| = O_p(1)$ uniformly in $\theta \in \Theta$. As $\frac{1}{n}\omega' P\omega = O(1)$ by Lemma B.3 (iii), $\frac{1}{n}\tilde{B}_{22}^{-1} - \frac{1}{n}B_{22}^{-1} = [(\tilde{\mu}_4 - 3(\tilde{\sigma}^2)^2) - (\mu_4 - 3\sigma_0^4)]\frac{1}{n}\omega'\omega + ((\tilde{\sigma}^2)^2 - \sigma_0^4)\frac{1}{n}\Upsilon - (\frac{\tilde{\mu}_3^2}{\tilde{\sigma}^2} - \frac{\mu_3^2}{\sigma_0^2})\frac{1}{n}\omega' P\omega = o_p(1)$. It follows that $n\tilde{B}_{22} - nB_{22} = o_p(1)$ by the continuous mapping theorem. Hence, $||\frac{1}{n}\epsilon'(\theta)P\omega[(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2})^2 n\tilde{B}_{22} - (\frac{\mu_3}{\sigma_0^2})^2 nB_{22}]\frac{1}{n}\omega' P\epsilon(\theta)|| \leq (\frac{1}{n}||\omega' P\epsilon(\theta)||)^2 ||(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2})^2 n\tilde{B}_{22} - (\frac{\mu_3}{\sigma_0^2})^2 nB_{22}|| = o_p(1)$, $||\frac{1}{n}\epsilon'(\theta)P\omega(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{B}_{22} - \frac{\mu_3}{\sigma_0^2}nB_{22})\frac{1}{n}g_2(\theta)|| \leq ||\frac{1}{n}\epsilon'(\theta)P\omega|| \cdot ||\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{B}_{22} - \frac{\mu_3}{\sigma_0^2}nB_{22}|| \cdot ||\frac{1}{n}g_2(\theta)|| = o_p(1)$, and $||\frac{1}{n}g_2'(\theta)(n\tilde{B}_{22} - nB_{22})\frac{1}{n}g_2(\theta)|| \leq (\frac{1}{n}||g_2(\theta)||)^2 ||n\tilde{B}_{22} - nB_{22}|| = o_p(1)$ uniformly in $\theta \in \Theta$. It follows that $\frac{1}{n}g'(\theta)(\tilde{\Omega}^{-1}-\Omega^{-1})g(\theta)=o_p(1)$ uniformly in $\theta\in\Theta$. ■

Proof of Proposition 5. By the Taylor expansion of $\frac{\partial g'(\hat{\theta})}{\partial\theta}\tilde{\Omega}^{-1}g(\hat{\theta})=0$ at θ_0,

$$\sqrt{n}(\hat{\theta}-\theta_0)=-[\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\tilde{\Omega}^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'}+\frac{1}{n}\frac{\partial^2 g'(\bar{\theta})}{\partial\theta'\partial\theta}\tilde{\Omega}^{-1}g(\bar{\theta})]^{-1}\frac{1}{\sqrt{n}}\frac{\partial g'(\theta_0)}{\partial\theta}\tilde{\Omega}^{-1}g(\theta_0). \qquad (18)$$

First, $\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\tilde{\Omega}^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'}=\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\Omega^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'}+\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}(\tilde{\Omega}^{-1}-\Omega^{-1})\frac{\partial g(\bar{\theta})}{\partial\theta'}$. By Lemma B.7 (v) and (vi), $\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\Omega^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'}=\frac{\sigma_0^{-2}}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'}+\frac{1}{n}\frac{\partial\bar{g}_2'(\bar{\theta})}{\partial\theta}B_{22}\frac{\partial\bar{g}_2(\bar{\theta})}{\partial\theta'}=\mathrm{D}(0,H)+\lim\frac{1}{n}\bar{D}_2'B_{22}\bar{D}_2+o_p(1)$. As $\frac{1}{n}(\tilde{B}_{22}^{-1}-B_{22}^{-1})=[(\tilde{\mu}_4-3(\tilde{\sigma}^2)^2)-(\mu_4-3\sigma_0^4)]\frac{1}{n}\omega'\omega+((\tilde{\sigma}^2)^2-\sigma_0^4)\frac{1}{n}\Upsilon-(\frac{\tilde{\mu}_3^2}{\tilde{\sigma}^2}-\frac{\mu_3^2}{\sigma_0^2})\frac{1}{n}\omega'P\omega=O_p(1/\sqrt{n})$, $n(\tilde{B}_{22}-B_{22})=O_p(1/\sqrt{n})$. As $\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'}=O_p(1)$, we have $\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\omega=O_p(1)$, and $\frac{1}{n}\frac{\partial g_2(\bar{\theta})}{\partial\theta'}=O_p(1)$ by Lemma B.7 (v) and (vi), it follows that

$$\begin{aligned}&\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}(\tilde{\Omega}^{-1}-\Omega^{-1})\frac{\partial g(\bar{\theta})}{\partial\theta'}\\ =\;&(\tilde{\sigma}^{-2}-\sigma_0^{-2})\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'}+\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\omega[(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2})^2n\tilde{B}_{22}-(\frac{\mu_3}{\sigma_0^2})^2nB_{22}]\frac{1}{n}\omega'P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'}\\ &-2\frac{1}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\omega(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{B}_{22}-\frac{\mu_3}{\sigma_0^2}nB_{22})\frac{1}{n}\frac{\partial g_2(\bar{\theta})}{\partial\theta'}+\frac{1}{n}\frac{\partial g_2'(\bar{\theta})}{\partial\theta}(n\tilde{B}_{22}-nB_{22})\frac{1}{n}\frac{\partial g_2(\bar{\theta})}{\partial\theta'}\\ =\;&O_p(1/\sqrt{n}).\end{aligned} \qquad (19)$$

Next,

$$\begin{aligned}&\frac{1}{n}\frac{\partial^2 g'(\bar{\theta})}{\partial\theta'\partial\theta}\tilde{\Omega}^{-1}g(\bar{\theta})\\ =\;&\frac{1}{n\tilde{\sigma}^2}\frac{\partial^2\epsilon'(\bar{\theta})}{\partial\theta'\partial\theta}P\epsilon(\bar{\theta})+[\frac{\tilde{\mu}_3}{n\tilde{\sigma}^2}\omega'P\frac{\partial^2\epsilon(\bar{\theta})}{\partial\theta\partial\theta'}-\frac{1}{n}\frac{\partial^2 g_2(\bar{\theta})}{\partial\theta\partial\theta'}]'n\tilde{B}_{22}[\frac{\tilde{\mu}_3}{n\tilde{\sigma}^2}\omega'P\epsilon(\bar{\theta})-\frac{1}{n}g_2(\bar{\theta})].\end{aligned} \qquad (20)$$

It follows from Lemma B.7 (vii) that $\frac{1}{n}\frac{\partial^2\epsilon'(\bar{\theta})}{\partial\theta'\partial\theta}P\epsilon(\bar{\theta})=o_p(1)$. As $\frac{\tilde{\mu}_3}{n\tilde{\sigma}^2}\omega'P\epsilon(\bar{\theta})-\frac{1}{n}g_2(\bar{\theta})=o_p(1)$ by Lemma B.7 (viii), the remaining term in (20) is $o_p(1)$, noting that by Lemma B.7 (vii) $\frac{1}{n}\omega'P\frac{\partial^2\epsilon(\bar{\theta})}{\partial\theta\partial\theta'}=O_p(1)$, $\frac{1}{n}\frac{\partial^2 g_2(\bar{\theta})}{\partial\theta\partial\theta'}=O_p(1)$, and $n\tilde{B}_{22}=O_p(1)$. In summary, $[\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\tilde{\Omega}^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'}+\frac{1}{n}\frac{\partial^2 g'(\bar{\theta})}{\partial\theta'\partial\theta}\tilde{\Omega}^{-1}g(\bar{\theta})]^{-1}=[\sigma_0^{-2}\mathrm{D}(0,H)+\lim\frac{1}{n}\bar{D}_2'B_{22}\bar{D}_2]^{-1}+o_p(1)$, as $\sigma_0^{-2}\mathrm{D}(0,H)+\lim\frac{1}{n}\bar{D}_2'B_{22}\bar{D}_2$ is nonsingular implied by Assumptions 4' and 8.

Next, $\frac{1}{\sqrt{n}}\frac{\partial g'(\theta_0)}{\partial\theta}\tilde{\Omega}^{-1}g(\theta_0)=\frac{1}{\sqrt{n}}\frac{\partial g'(\theta_0)}{\partial\theta}\Omega^{-1}g(\theta_0)+\frac{1}{\sqrt{n}}\frac{\partial g'(\theta_0)}{\partial\theta}(\tilde{\Omega}^{-1}-\Omega^{-1})g(\theta_0)$. As $-\frac{\sigma_0^{-2}}{\sqrt{n}}\frac{\partial\epsilon'(\theta_0)}{\partial\theta}P\epsilon=\frac{\sigma_0^{-2}}{\sqrt{n}}(Mu,RZ)'P\epsilon=\frac{\sigma_0^{-2}}{\sqrt{n}}(MR^{-1}\epsilon,f+ve_1')'P\epsilon=\frac{1}{\sqrt{n}}[\mathrm{tr}(PMR^{-1}),\mathrm{tr}(\Psi)e_1']'+\frac{\sigma_0^{-2}}{\sqrt{n}}(0,f'\epsilon)+o_p(1)$, we have $-\frac{1}{\sqrt{n}}\frac{\partial g'(\theta_0)}{\partial\theta}\Omega^{-1}g(\theta_0)=-\frac{\sigma_0^{-2}}{\sqrt{n}}\frac{\partial\epsilon'(\theta_0)}{\partial\theta}P\epsilon+\frac{1}{\sqrt{n}}\frac{\partial\bar{g}_2'(\theta_0)}{\partial\theta}B_{22}\bar{g}_2(\theta_0)=\frac{1}{\sqrt{n}}(\mathrm{tr}(PMR^{-1}),\mathrm{tr}(\Psi)e_1')'+\frac{\sigma_0^{-2}}{\sqrt{n}}(0,f'\epsilon)+\frac{1}{\sqrt{n}}\bar{D}_2'B_{22}\bar{g}_2(\theta_0)+o_p(1)$. As $\mathrm{E}[\bar{g}_2(\theta_0)\epsilon'Pf]=\mathrm{E}[\frac{\mu_3}{\sigma_0^2}\omega'P\epsilon\epsilon'Pf]-\mathrm{E}[g_2(\theta_0)\epsilon'Pf]=\mu_3\omega'Pf-$

$\mu_3\omega' Pf = 0$, and $\mathrm{E}[\bar{g}_2(\theta_0)\epsilon'(I-P)f] = -\mu_3\omega'(I-P)f$, it follows that $\frac{1}{n}\mathrm{E}[\bar{g}_2(\theta_0)\epsilon' f] = \frac{1}{n}\mathrm{E}[\bar{g}_2(\theta_0)\epsilon' Pf] + \frac{1}{n}\mathrm{E}[\bar{g}_2(\theta_0)\epsilon'(I-P)f] = -\frac{1}{n}\mu_3\omega'(I-P)f = o(1)$. Hence $\frac{1}{\sqrt{n}}f'\epsilon$ and $\frac{1}{\sqrt{n}}\bar{g}_2(\theta_0)$ are asymptotically uncorrelated. By the CLT in Kelejian and Prucha (2001), $\frac{\sigma_0^{-2}}{\sqrt{n}}(0, f'\epsilon) + \frac{1}{\sqrt{n}}\bar{D}_2' B_{22}\bar{g}_2(\theta_0) \xrightarrow{d} N(0, \sigma_0^{-2}\mathrm{D}(0,H) + \lim_{n\to\infty}\frac{1}{n}\bar{D}_2' B_{22}\bar{D}_2)$. It remains to show that $\frac{1}{\sqrt{n}}\frac{\partial g'(\theta_0)}{\partial\theta}(\tilde{\Omega}^{-1} - \Omega^{-1})g(\theta_0) = o_p(1)$. As $\frac{\partial\epsilon(\theta_0)}{\partial\theta'} = -J(Mu, RZ)$, $\frac{1}{n}\frac{\partial\epsilon'(\theta_0)}{\partial\theta}P\epsilon = o_p(1)$ and $\frac{1}{n}\frac{\partial\epsilon'(\theta_0)}{\partial\theta}P\omega = O_p(1)$. As shown above, $n(\tilde{B}_{22} - B_{22}) = o_p(1)$, and $\frac{1}{\sqrt{n}}\omega' P\epsilon = O_p(1)$ and $\frac{1}{\sqrt{n}}g_2(\theta_0) = O_p(1)$, it follows that

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\frac{\partial g'(\theta_0)}{\partial\theta}(\tilde{\Omega}^{-1} - \Omega^{-1})g(\theta_0) \\
= \; & \sqrt{n}(\tilde{\sigma}^{-2} - \sigma_0^{-2})\frac{1}{n}\frac{\partial\epsilon'(\theta_0)}{\partial\theta}P\epsilon + \frac{1}{n}\frac{\partial\epsilon'(\theta_0)}{\partial\theta}P\omega[(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2})^2 n\tilde{B}_{22} - (\frac{\mu_3}{\sigma_0^2})^2 nB_{22}]\frac{1}{\sqrt{n}}\omega' P\epsilon \\
& -\frac{1}{n}\frac{\partial\epsilon'(\theta_0)}{\partial\theta}P\omega(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{B}_{22} - \frac{\mu_3}{\sigma_0^2}nB_{22})\frac{1}{\sqrt{n}}g_2(\theta_0) - \frac{1}{n}\frac{\partial g_2'(\theta_0)}{\partial\theta}(\frac{\tilde{\mu}_3}{\tilde{\sigma}^2}n\tilde{B}_{22} - \frac{\mu_3}{\sigma_0^2}nB_{22})'\frac{1}{\sqrt{n}}\omega' P\epsilon \\
& +\frac{1}{n}\frac{\partial g_2'(\theta_0)}{\partial\theta}(n\tilde{B}_{22} - nB_{22})\frac{1}{\sqrt{n}}g_2(\theta_0) = o_p(1).
\end{aligned}
$$

From (18), $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(K/\sqrt{n})$, which implies that $\bar{\theta} - \theta_0 = O_p(K/n)$. $\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}(\tilde{\Omega}^{-1} - \Omega^{-1})\frac{\partial g(\bar{\theta})}{\partial\theta'} = O_p(1/\sqrt{n})$ by (19). As $\frac{1}{n}\frac{\partial^2\epsilon'(\bar{\theta})}{\partial\theta'\partial\theta}P\epsilon(\bar{\theta}) = O_p(\sqrt{K/n})$, $\frac{1}{n}\omega' P\frac{\partial^2\epsilon(\bar{\theta})}{\partial\theta\partial\theta'} = O_p(1)$ and $\frac{1}{n}\frac{\partial^2 g_2(\bar{\theta})}{\partial\theta\partial\theta'} = O_p(1)$ by Lemma B.7 (vii), $n\tilde{B}_{22} = O_p(1)$, and $\frac{\tilde{\mu}_3}{n\tilde{\sigma}^2}\omega' P\epsilon(\bar{\theta}) - \frac{1}{n}g_2(\bar{\theta}) = O_p(\sqrt{K/n})$ by Lemma B.7 (viii), it follows from (20) that $\frac{1}{n}\frac{\partial^2 g'(\bar{\theta})}{\partial\theta'\partial\theta}\tilde{\Omega}^{-1}g(\bar{\theta}) = O_p(\sqrt{K/n})$. By Lemma B.7 (v) and (vi), $\frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\Omega^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'} = \frac{\sigma_0^{-2}}{n}\frac{\partial\epsilon'(\bar{\theta})}{\partial\theta}P\frac{\partial\epsilon(\bar{\theta})}{\partial\theta'} + \frac{1}{n}\frac{\partial\bar{g}_2'(\bar{\theta})}{\partial\theta}B_{22}\frac{\partial\bar{g}_2(\bar{\theta})}{\partial\theta'} = \frac{1}{n}[\sigma^{-2}\mathrm{D}(0, Z'R'PRZ) + \check{D}_2' B_{22}\check{D}_2] + O_p(\sqrt{K/n})$. It follows that

$$
\begin{aligned}
& \frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\tilde{\Omega}^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'} + \frac{1}{n}\frac{\partial^2 g'(\bar{\theta})}{\partial\theta'\partial\theta}\tilde{\Omega}^{-1}g(\bar{\theta}) \\
= \; & \frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}\Omega^{-1}\frac{\partial g(\bar{\theta})}{\partial\theta'} + \frac{1}{n}\frac{\partial g'(\bar{\theta})}{\partial\theta}(\tilde{\Omega}^{-1} - \Omega^{-1})\frac{\partial g(\bar{\theta})}{\partial\theta'} + \frac{1}{n}\frac{\partial^2 g'(\bar{\theta})}{\partial\theta'\partial\theta}\tilde{\Omega}^{-1}g(\bar{\theta}) \\
= \; & \frac{1}{n}\sigma^{-2}\mathrm{D}(0, Z'R'PRZ) + \frac{1}{n}\check{D}_2' B_{22}\check{D}_2 + O_p(\sqrt{K/n}).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\sqrt{n}(\hat{\theta} - \theta_0) \; = \; & \{[\frac{1}{n}\sigma^{-2}\mathrm{D}(0, Z'R'PRZ) + \frac{1}{n}\check{D}_2' B_{22}\check{D}_2]^{-1} + O_p(\sqrt{K/n})\} \\
& \times[\frac{1}{\sqrt{n}}(\mathrm{tr}(PMR^{-1}), \mathrm{tr}(\Psi)e_1')' + \frac{\sigma_0^{-2}}{\sqrt{n}}(0, f'\epsilon) + \frac{1}{\sqrt{n}}\bar{D}_2' B_{22}\bar{g}_2(\theta_0) + o_p(1)], \\
= \; & [\sigma_0^{-2}\mathrm{D}(0,H) + \lim\frac{1}{n}\bar{D}_2' B_{22}\bar{D}_2]^{-1}[\frac{\sigma_0^{-2}}{\sqrt{n}}(0, f'\epsilon) + \frac{1}{\sqrt{n}}\bar{D}_2' B_{22}\bar{g}_2(\theta_0)] \\
& +\sqrt{n}b_{gmm} + O_p(K^{3/2}/n) + o_p(1). \qquad (21)
\end{aligned}
$$

When $K^{3/2}/n \to 0$, the conclusion then follows by the Slutzky theorem. ■

Proof of Proposition 6. Based on (21) in the proof of Proposition 5, it is sufficient to show that $\sqrt{n}(\tilde{b}_{gmm} - b_{gmm}) = o_p(1)$. $\sqrt{n}(\tilde{b}_{gmm} - b_{gmm}) = \{[\tilde{\sigma}^{-2}\frac{1}{n}\mathrm{D}(0, Z'\tilde{R}'P\tilde{R}Z) + \frac{1}{n}\tilde{\tilde{D}}_2'\tilde{B}_{22}\tilde{\tilde{D}}_2]^{-1} - (\sigma^{-2}\frac{1}{n}\mathrm{D}(0, Z'R'PRZ)+\frac{1}{n}\check{D}_2'B_{22}\check{D}_2)^{-1}\}[\mathrm{tr}(PM\tilde{R}^{-1}), \mathrm{tr}(P\tilde{R}\tilde{G}\tilde{R}^{-1})e_1']'/\sqrt{n}+(\sigma^{-2}\frac{1}{n}\mathrm{D}(0, Z'R'PRZ)+\frac{1}{n}\check{D}_2'B_{22}\check{D}_2)^{-1}[\mathrm{tr}(PM(\tilde{R}^{-1} - R^{-1})), \mathrm{tr}(P\tilde{R}\tilde{G}\tilde{R}^{-1} - \Psi)e_1']'/\sqrt{n}$. By Lemma B.5 and our assumption that $\tilde{\rho} - \rho_0 = O_p(K/n)$, $\frac{1}{n}Z'\tilde{R}'P\tilde{R}Z - \frac{1}{n}Z'R'PRZ = O_p(\tilde{\rho} - \rho_0) = O_p(K/n)$, and $\frac{1}{n}\omega'P\tilde{R}Z - \frac{1}{n}\omega'PRZ = O_p(\tilde{\rho} - \rho_0) = O_p(K/n)$. Also, $\frac{1}{n}(\tilde{D}_2 - D_2) = O_p(1/\sqrt{n})$ and $n(\tilde{B}_{22} - B_{22}) = O_p(1/\sqrt{n})$. Hence, $\frac{1}{n}[\tilde{\sigma}^{-2}\mathrm{D}(0, Z'\tilde{R}'P\tilde{R}Z)+\tilde{\tilde{D}}_2'\tilde{B}_{22}\tilde{\tilde{D}}_2]-\frac{1}{n}[\sigma_0^{-2}\mathrm{D}(0, Z'R'PRZ)+\check{D}_2'B_{22}\check{D}_2] = O_p(\max\{K/n, 1/\sqrt{n}\})$. On the other hand, as $\tilde{G}-G = (\tilde{\lambda}-\lambda_0)\tilde{G}G$, $\tilde{R}-R = -(\tilde{\rho}-\rho_0)M$, and $\tilde{R}^{-1}-R^{-1} = (\tilde{\rho}-\rho_0)\tilde{R}^{-1}MR^{-1}$, we have $\mathrm{tr}[PM(\tilde{R}^{-1} - R^{-1})]/\sqrt{n} = (\tilde{\rho} - \rho_0)\mathrm{tr}(PM\tilde{R}^{-1}MR^{-1})/\sqrt{n} = O_p(K^2/n^{3/2}) = o_p(1)$, and, similarly, $\mathrm{tr}(P\tilde{R}\tilde{G}\tilde{R}^{-1} - \Psi)/\sqrt{n} = O_p(K^2/n^{3/2}) = o_p(1)$, because $(K^2/n^{3/2})/(K^{3/2}/n) = \sqrt{K/n} \to 0$. The desired result follows since $\frac{1}{n}[\sigma_0^{-2}\mathrm{D}(0, Z'R'PRZ) + \check{D}_2'B_{22}\check{D}_2] = O_p(1)$ and $[\mathrm{tr}(PM\tilde{R}^{-1}), \mathrm{tr}(P\tilde{R}\tilde{G}\tilde{R}^{-1})e_1']'/\sqrt{n} = O(K/\sqrt{n})$. ■

Proof of Proposition 7. It follow from Proposition 5, the infeasible BGMM estimator $\hat{\theta}_{bgmm} = \arg\min_{\theta\in\Theta}\mathcal{L}(\theta)$, where $\mathcal{L}(\theta) = g^{*\prime}(\theta)V^{*-1}g^*(\theta) = \epsilon'(\theta)P\epsilon(\theta)+g_2^{*\prime}(\theta)(\sigma_0^2\Upsilon^*)^{-1}g_2^*(\theta)$, has the limiting distribution given by (11). The objective function of the feasible BGMM estimator is given by $\mathcal{L}^*(\theta) = \tilde{g}^{*\prime}(\theta)\tilde{V}^{*-1}\tilde{g}^*(\theta) = \epsilon'(\theta)P\epsilon(\theta) + \tilde{g}_2^{*\prime}(\theta)(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1}\tilde{g}_2^*(\theta)$, where $\tilde{g}_2^*(\theta) = [\tilde{U}_1^*\epsilon(\theta), \tilde{U}_2^*\epsilon(\theta)]'\epsilon(\theta)$ and $\tilde{\Upsilon}^* = \begin{pmatrix} \mathrm{tr}(\tilde{U}_1^{*s}\tilde{U}_1^*) & \mathrm{tr}(\tilde{U}_1^{*s}\tilde{U}_2^*) \\ \mathrm{tr}(\tilde{U}_2^{*s}\tilde{U}_1^*) & \mathrm{tr}(\tilde{U}_2^{*s}\tilde{U}_2^*) \end{pmatrix}$. We shall show that the objective functions of the feasible and infeasible GMM will satisfy the conditions of the CLT in Lee (2004). If so, the estimator from the minimization of the objective functions of the feasible and infeasible GMM will have the same limiting distribution.

First, $\frac{1}{n}(\tilde{g}_2^*(\theta) - g_2^*(\theta))' = \frac{1}{n}[(\tilde{U}_1^* - U_1^*)\epsilon(\theta), (\tilde{U}_2^* - U_2^*)\epsilon(\theta)]'\epsilon(\theta) = o_p(1)$ uniformly in $\theta \in \Theta$. The derivatives of $\tilde{g}_2^*(\theta)$ and $g_2^*(\theta)$ are

$$\frac{\partial g_2^*(\theta)}{\partial\theta'} = \begin{pmatrix} \epsilon'(\theta)U_1^{*s}\frac{\partial\epsilon(\theta)}{\partial\theta'} \\ \epsilon'(\theta)U_2^{*s}\frac{\partial\epsilon(\theta)}{\partial\theta'} \end{pmatrix}, \text{ and } \frac{\partial^2 g_2^*(\theta)}{\partial\theta\partial\theta'} = \begin{pmatrix} \frac{\partial\epsilon'(\theta)}{\partial\theta}U_1^{*s}\frac{\partial\epsilon(\theta)}{\partial\theta'} + \epsilon'(\theta)U_1^{*s}\frac{\partial^2\epsilon(\theta)}{\partial\theta\partial\theta'} \\ \frac{\partial\epsilon'(\theta)}{\partial\theta}U_2^{*s}\frac{\partial\epsilon(\theta)}{\partial\theta'} + \epsilon'(\theta)U_2^{*s}\frac{\partial^2\epsilon(\theta)}{\partial\theta\partial\theta'} \end{pmatrix}.$$

$\frac{\partial\epsilon(\theta)}{\partial\theta'} = -J[MS(\lambda)Y - MX\beta, R(\rho)WY, R(\rho)X]$ where $Y = S^{-1}(X\beta_0 + \iota\alpha_0) + S^{-1}R^{-1}\epsilon$. $\frac{\partial^2\epsilon(\theta)}{\partial\rho\partial\theta'} = J[0, MWY, MX]$, $\frac{\partial^2\epsilon(\theta)}{\partial\lambda\partial\theta'} = J[MWY, 0, 0]$, and $\frac{\partial^2\epsilon(\theta)}{\partial\beta\partial\theta'} = J[MX, 0, 0]$. It follows from Lemma B.9 that $\frac{1}{n}(\frac{\partial\tilde{g}_2^*(\theta)}{\partial\theta'} - \frac{\partial g_2^*(\theta)}{\partial\theta'}) = o_p(1)$ and $\frac{1}{n}(\frac{\partial^2\tilde{g}_2^*(\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2 g_2^*(\theta)}{\partial\theta\partial\theta'}) = o_p(1)$ uniformly in $\theta \in \Theta$. Consider

$\frac{1}{n}(\tilde{\sigma}^2\tilde{\Upsilon}^* - \sigma_0^2\Upsilon^*)$. As $\{\tilde{U}_j^*\}$ is UBC in probability, it follows from Lemma B.9 that $\frac{1}{n}\text{tr}(\tilde{U}_i^{*s}\tilde{U}_j^*) - \frac{1}{n}\text{tr}(U_i^{*s}U_j^*) = o_p(1)$. Hence, $\frac{1}{n}(\tilde{\sigma}^2\tilde{\Upsilon}^* - \sigma_0^2\Upsilon^*) = o_p(1)$, as $\hat{\sigma}^2 - \sigma_0^2 = o_p(1)$. As the limit of $\frac{1}{n}\Upsilon^*$ exists and is a nonsingular matrix, $(\frac{1}{n}\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1} - (\frac{1}{n}\sigma_0^2\Upsilon^*)^{-1} = o_p(1)$ by the continuous mapping theorem. Furthermore, because $\frac{1}{n}(\tilde{g}_2^*(\theta) - g_2^*(\theta)) = o_p(1)$, and $\frac{1}{n}[g_2^*(\theta) - \text{E}(g_2^*(\theta))] = o_p(1)$ uniformly in $\theta \in \Theta$, and $\sup_{\theta\in\Theta}\frac{1}{n}|\text{E}(g_2^*(\theta))| = O(1)$ (Lee, 2007a, p. 21), $\frac{1}{n}g_2^*(\theta)$ and $\frac{1}{n}\tilde{g}_2^*(\theta)$ are $O_p(1)$ uniformly in $\theta \in \Theta$. Similarly, $\frac{1}{n}\frac{\partial g_2^*(\theta)}{\partial\theta'}$, $\frac{1}{n}\frac{\partial \tilde{g}_2^*(\theta)}{\partial\theta'}$, $\frac{1}{n}\frac{\partial^2 g_2^*(\theta)}{\partial\theta\partial\theta'}$ and $\frac{1}{n}\frac{\partial^2 \tilde{g}_2^*(\theta)}{\partial\theta\partial\theta'}$ are $O_p(1)$ uniformly in $\theta \in \Theta$.

With the uniform convergence in probability and uniformly stochastic boundedness properties, the difference of $\mathcal{L}^*(\theta)$ and $\mathcal{L}(\theta)$ can be investigated. By expansion, $\frac{1}{n}(\mathcal{L}^*(\theta) - \mathcal{L}(\theta)) = \frac{1}{n}\tilde{g}_2^{*\prime}(\theta)(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1}[\tilde{g}_2^*(\theta) - g_2^*(\theta)] + \frac{1}{n}\tilde{g}_2^{*\prime}(\theta)[(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1} - (\sigma_0^2\Upsilon^*)^{-1}]g_2^*(\theta) + \frac{1}{n}[\tilde{g}_2^*(\theta) - g_2^*(\theta)]'(\sigma_0^2\Upsilon^*)^{-1}g_2^*(\theta)$, which is $o_p(1)$ uniformly in $\theta \in \Theta$. Similarly, for each component θ_l of θ, $\frac{1}{n}\frac{\partial^2\mathcal{L}^*(\theta)}{\partial\theta_l\partial\theta'} - \frac{1}{n}\frac{\partial^2\mathcal{L}(\theta)}{\partial\theta_l\partial\theta'} = \frac{2}{n}[\frac{\partial\tilde{g}_2^{*\prime}(\theta)}{\partial\theta_l}(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1}\frac{\partial\tilde{g}_2^*(\theta)}{\partial\theta'} + \tilde{g}_2^{*\prime}(\theta)(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1}\frac{\partial^2\tilde{g}_2^*(\theta)}{\partial\theta_l\partial\theta'} - \frac{\partial g_2^{*\prime}(\theta)}{\partial\theta_l}(\sigma_0^2\Upsilon^*)^{-1}\frac{\partial g_2^*(\theta)}{\partial\theta'} - g_2^{*\prime}(\theta)(\sigma_0^2\Upsilon^*)^{-1}\frac{\partial^2 g_2^*(\theta)}{\partial\theta_l\partial\theta'}] = o_p(1)$.

Finally, because $[\frac{\partial\tilde{g}_2^{*\prime}(\theta)}{\partial\theta}(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1} - \frac{\partial g_2^{*\prime}(\theta)}{\partial\theta}(\sigma_0^2\Upsilon^*)^{-1}] = o_p(1)$ as above, and $\frac{1}{\sqrt{n}}g_2^*(\theta_0) = O_p(1)$,

$$\begin{aligned}
& \frac{1}{\sqrt{n}}(\frac{\partial\mathcal{L}^*(\theta_0)}{\partial\theta} - \frac{\partial\mathcal{L}(\theta_0)}{\partial\theta}) \\
= \; & 2\{\frac{\partial\tilde{g}_2^{*\prime}(\theta)}{\partial\theta}(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1}\frac{1}{\sqrt{n}}(\tilde{g}_2^*(\theta_0) - g_2^*(\theta_0)) + [\frac{\partial\tilde{g}_2^{*\prime}(\theta)}{\partial\theta}(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1} - \frac{\partial g_2^{*\prime}(\theta)}{\partial\theta}(\sigma_0^2\Upsilon^*)^{-1}]\frac{1}{\sqrt{n}}g_2^*(\theta_0)\} \\
= \; & 2\frac{\partial\tilde{g}_2^{*\prime}(\theta)}{\partial\theta}(\tilde{\sigma}^2\tilde{\Upsilon}^*)^{-1}\frac{1}{\sqrt{n}}(\tilde{g}_2^*(\theta_0) - g_2^*(\theta_0)) + o_p(1).
\end{aligned}$$

As $\frac{1}{\sqrt{n}}(\tilde{g}_2^*(\theta_0) - g_2^*(\theta_0)) = o_p(1)$ by Lemma B.9, $\frac{1}{\sqrt{n}}(\frac{\partial\mathcal{L}^*(\theta_0)}{\partial\theta} - \frac{\partial\mathcal{L}(\theta_0)}{\partial\theta}) = o_p(1)$. The desired result follows from the CLT in Lee (2004). ■

References

Alvarez, J. and Arellano, M. (2003). The time series and cross-section asymptotics of dynamic panel data estimators, *Econometrica* **71**: 1121–1159.

Amemiya, T. (1985). *Advanced Econometrics*, Harvard University Press, Cambridge.

Anderson, T., Kunitomo, N. and Matsushita, Y. (2007). On the asymptotic optimality of the liml estimator with possibly many instruments. Working paper, University of Tokyo.

Baltagi, B. (1995). *Econometric Analysis of Panel Data*, John Wiley and Sons, New York.

Bekker, P. (1994). Alternative approximations to the distributions of instrumental variable estimators, *Econometrica* **62**: 657–681.

Bekker, P. and van der Ploeg, J. (2005). Instrumental variable estimation based on grouped data, *Statistica Neerlandica* **59**: 239–267.

Bell, K. P. and Bockstael, N. E. (2000). Applying the generalized method of moments approach to spatial problems involving micro-level data, *Review of Economics and Statistics* **82**: 72–82.

Bonacich, P. (1987). Power and centrality: a family of measures, *American Journal of Sociology* **92**: 1170–1182.

Brock, W. A. and Durlauf, S. N. (2001). Interactions-based models, *in* J. J. Heckman and E. E. Leamer (eds), *Handbook of Econometrics*, North-Holland, Amsterdam.

Chao, J. and Swanson, N. (2005). Consistent estimation with a large number of weak instruments, *Econometrica* **73**: 1673–1692.

Chao, J. and Swanson, N. (2007). Alternative approximations of the bias and mse of the iv estimator under weak identification with an application to bias correction, *Journal of Econometrics* **137**: 515–555.

Cohen, J. (2002). Reciprocal state and local airport spending spillovers and symmetric responses to cuts and increases in federal airport grants, *Public Finance Review* **30**: 41–55.

Cox, D. R. (1975). Partial likelihood, *Biometrika* **62**: 269–276.

Donald, S. G. and Newey, W. K. (2001). Choosing the number of instruments, *Econometrica* **69**: 1161–1191.

Fingleton, B. (2008). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors, *Spatial Economic Analysis* **3**: 27–44.

Han, C. and Phillips, P. C. (2006). Gmm with many moment conditions, *Econometrica* **74**: 147–192.

Hansen, C., Hausman, J. and Newey, W. (2008). Estimation with many instrumental variables, *Journal of Business and Economic Statistics* **26**: 398–422.

Hausman, J., Newey, W., Woutersen, T., Chao, J. and Swanson, N. (2008). Instrumental variable estimation with heteroscedasticity and many instruments. Working paper.

Horn, R. and Johnson, C. (1985). *Matrix Analysis*, Cambridge Univsersity Press, Cambridge.

Hsiao, C. (2003). *Analysis of Panel Data*, Cambridge University Press, Cambridge.

Kalbfleisch, J. D. and Sprott, D. A. (1970). Application of likelihood methods to models involving large numbers of parameters, *Journal of the Royal Statistical Society. Series B (Methodological)* **32**: 175–208.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbance, *Journal of Real Estate Finance and Economics* **17**: 99–121.

Kelejian, H. H. and Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* **40**: 509–533.

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

Kelejian, H. H. and Prucha, I. R. (2007). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Working paper, Department of Economics, University of Maryland.

Kelejian, H. H. and Robinson, D. P. (1995). Spatial correlation: a suggested alternative to autoregressive models, *in* L. Anselin and R. Florax (eds), *New Directions in Spatial Econometrics*, Springer, Berlin.

Lancaster, T. (2000). The incidental parameter problem since 1948, *Journal of Econometrics* **95**: 391–413.

Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models, *Econometric Theory* **18**: 252–277.

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

Lee, L. F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial econometric models, *Econometrica* **72**: 1899–1926.

Lee, L. F. (2007a). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics* **137**: 489–514.

Lee, L. F. (2007b). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects, *Journal of Econometrics* **140**: 333–374.

Lee, L. F. (2007c). The method of elimination and substitution in the gmm estimation of mixed regressive, spatial autoregressive models, *Journal of Econometrics* **140**: 155–189.

Lee, L. F. and Liu, X. (2010). Efficient gmm estimation of high order spatial autoregressive models with autoregressive disturbances, *Econometric Theory* **26**: 187–230.

Lee, L. F., Liu, X. and Lin, X. (2009). Specification and estimation of social interaction models with network structures. forthcoming in *The Econometrics Journal*.

Lee, L. F. and Yu, J. (2009). Near unit root in the spatial autoregressive model. Working paper, Ohio State University.

LeSage, J. (1999). The theory and practice of spatial econometrics. Manuscript, Department of Economics, University of Toledo.

Lin, X. and Lee, L. F. (2006). Gmm estimation of spatial autoregressive models with unknown heteroskedasticity. Working paper, Department of Economics, Ohio State University.

Liu, X., Lee, L. F. and Bollinger, C. (2009). An efficient gmm estimator of spatial autoregressive models. Working paper, Department of Economics, University of Colorado at Boulder.

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

Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions, *in* S. N. Durlauf and H. P. Young (eds), *Social Dynamics*, MIT press, Cambridge.

Neyman, J. and Scott, E. L. (1948). Consistent estimates based on partially consistent observations, *Econometrica* **16**: 1–32.

van Hasselt, M. (2010). Many instruments asymptotic approximations under nonnormal error distributions, *Econometric Theory* **26**: 633–645.

Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York.

Table 1: 2SLS and GMM estimation (normal)

$\sigma^2_\alpha = 1$	$\lambda_0 = 0.1$	$\rho_0 = 0.1$	$\beta_{01} = 0.2$	$\beta_{02} = 0.2$
$m_r = 10$	$\bar{r} = 30$		$c_n/n = 0.609$	
2SLS (few IVs)	.099(.219)[.219]	.148(.309)[.313]	.201(.072)[.072]	.208(.074)[.074]
2SLS (many IVs)	.062(.068)[.078]	–	.194(.065)[.066]	.214(.057)[.059]
FC2SLS	.108(.082)[.082]	–	.198(.066)[.066]	.206(.058)[.059]
GMM (few IVs)	.096(.125)[.125]	.120(.215)[.216]	.197(.068)[.068]	.206(.062)[.062]
GMM (many IVs)	.085(.057)[.059]	.065(.146)[.150]	.196(.066)[.066]	.207(.058)[.058]
FCGMM	.099(.064)[.064]	.121(.169)[.170]	.197(.067)[.067]	.204(.058)[.058]
	$\bar{r} = 60$		$c_n/n = 0.660$	
2SLS (few IVs)	.104(.151)[.151]	.126(.241)[.242]	.203(.048)[.048]	.205(.051)[.051]
2SLS (many IVs)	.069(.046)[.055]	–	.198(.047)[.047]	.212(.041)[.043]
FC2SLS	.105(.056)[.056]	–	.201(.047)[.047]	.205(.042)[.042]
GMM (few IVs)	.100(.089)[.089]	.105(.144)[.144]	.201(.047)[.047]	.203(.044)[.044]
GMM (many IVs)	.087(.040)[.042]	.070(.094)[.099]	.199(.046)[.046]	.205(.041)[.041]
FCGMM	.098(.044)[.044]	.120(.111)[.112]	.200(.046)[.046]	.203(.041)[.041]
$m_r = 15$	$\bar{r} = 30$		$c_n/n = 0.615$	
2SLS (few IVs)	.100(.166)[.166]	.124(.270)[.271]	.204(.052)[.052]	.203(.056)[.056]
2SLS (many IVs)	.091(.056)[.056]	–	.201(.052)[.052]	.206(.042)[.043]
FC2SLS	.108(.063)[.063]	–	.202(.052)[.052]	.203(.044)[.044]
GMM (few IVs)	.101(.097)[.097]	.104(.160)[.160]	.202(.051)[.051]	.201(.046)[.046]
GMM (many IVs)	.098(.046)[.046]	.089(.116)[.116]	.201(.051)[.051]	.200(.042)[.042]
FCGMM	.102(.048)[.048]	.103(.120)[.120]	.201(.051)[.051]	.199(.042)[.042]
	$\bar{r} = 60$		$c_n/n = 0.629$	
2SLS (few IVs)	.095(.106)[.106]	.112(.190)[.190]	.199(.037)[.037]	.199(.041)[.041]
2SLS (many IVs)	.083(.038)[.042]	–	.198(.038)[.038]	.203(.033)[.033]
FC2SLS	.099(.040)[.040]	–	.199(.038)[.038]	.199(.033)[.033]
GMM (few IVs)	.094(.057)[.057]	.107(.104)[.104]	.198(.037)[.037]	.199(.035)[.035]
GMM (many IVs)	.093(.033)[.034]	.090(.078)[.079]	.198(.037)[.037]	.198(.032)[.032]
FCGMM	.097(.034)[.034]	.104(.080)[.080]	.198(.037)[.037]	.197(.033)[.033]

Mean(SD)[RMSE] c_n : average empirical concentration parameter; n : sample size

Table 2: 2SLS and GMM estimation (normal)

$\sigma_\alpha^2 = 0.04$	$\lambda_0 = 0.1$	$\rho_0 = 0.1$	$\beta_{01} = 0.2$	$\beta_{02} = 0.2$
$m_r = 10$	$\bar{r} = 30$		$c_n/n = 0.149$	
2SLS (few IVs)	.089(.220)[.220]	.155(.329)[.334]	.201(.070)[.070]	.210(.072)[.072]
2SLS (many IVs)	−.009(.120)[.162]	–	.189(.067)[.067]	.226(.060)[.065]
FC2SLS	.108(.151)[.151]	–	.199(.068)[.068]	.206(.065)[.065]
GMM (few IVs)	.095(.122)[.122]	.121(.219)[.220]	.197(.068)[.068]	.206(.062)[.062]
GMM (many IVs)	.062(.084)[.092]	.100(.169)[.169]	.195(.067)[.067]	.211(.058)[.060]
FCGMM	.088(.099)[.100]	.144(.203)[.208]	.197(.067)[.068]	.206(.060)[.060]
	$\bar{r} = 60$		$c_n/n = 0.154$	
2SLS (few IVs)	.102(.149)[.149]	.130(.249)[.251]	.203(.048)[.048]	.205(.050)[.050]
2SLS (many IVs)	.013(.094)[.128]	–	.195(.047)[.047]	.222(.043)[.048]
FC2SLS	.103(.117)[.117]	–	.202(.048)[.048]	.205(.046)[.046]
GMM (few IVs)	.100(.097)[.097]	.107(.148)[.148]	.201(.047)[.047]	.203(.044)[.044]
GMM (many IVs)	.071(.060)[.066]	.095(.113)[.113]	.198(.046)[.046]	.208(.041)[.042]
FCGMM	.091(.071)[.072]	.134(.139)[.143]	.200(.047)[.047]	.204(.042)[.042]
$m_r = 15$	$\bar{r} = 30$		$c_n/n = 0.162$	
2SLS (few IVs)	.098(.167)[.167]	.125(.286)[.287]	.203(.052)[.052]	.203(.056)[.056]
2SLS (many IVs)	.056(.120)[.128]	–	.200(.052)[.052]	.212(.047)[.048]
FC2SLS	.105(.135)[.135]	–	.203(.052)[.053]	.202(.051)[.051]
GMM (few IVs)	.100(.097)[.097]	.104(.164)[.164]	.202(.051)[.051]	.201(.047)[.047]
GMM (many IVs)	.089(.076)[.077]	.103(.150)[.151]	.201(.051)[.051]	.202(.044)[.044]
FCGMM	.097(.082)[.082]	.113(.160)[.161]	.201(.051)[.051]	.201(.044)[.044]
	$\bar{r} = 60$		$c_n/n = 0.160$	
2SLS (few IVs)	.095(.108)[.108]	.114(.197)[.197]	.199(.037)[.037]	.199(.041)[.041]
2SLS (many IVs)	.046(.083)[.099]	–	.196(.037)[.038]	.210(.037)[.038]
FC2SLS	.097(.091)[.092]	–	.199(.037)[.037]	.199(.039)[.039]
GMM (few IVs)	.095(.056)[.057]	.106(.105)[.105]	.198(.037)[.037]	.199(.035)[.035]
GMM (many IVs)	.085(.050)[.052]	.103(.096)[.096]	.198(.037)[.038]	.200(.034)[.034]
FCGMM	.093(.053)[.053]	.112(.101)[.101]	.198(.037)[.037]	.198(.034)[.034]

Mean(SD)[RMSE] c_n : average empirical concentration parameter; n : sample size

Table 3: 2SLS and GMM estimation (normal)

$\sigma_\alpha^2 = 1$	$\lambda_0 = 0.3$	$\rho_0 = 0.3$	$\beta_{01} = 0.2$	$\beta_{02} = 0.2$
$m_r = 10$	$\bar{r} = 30$		$c_n/n = 1.182$	
2SLS (few IVs)	.308(.198)[.198]	.356(.253)[.259]	.201(.076)[.076]	.206(.065)[.065]
2SLS (many IVs)	.278(.042)[.048]	–	.194(.065)[.065]	.206(.055)[.055]
FC2SLS	.309(.064)[.065]	–	.198(.067)[.067]	.203(.057)[.057]
GMM (few IVs)	.304(.123)[.123]	.311(.200)[.201]	.199(.069)[.069]	.205(.060)[.060]
GMM (many IVs)	.288(.038)[.040]	.238(.136)[.149]	.195(.066)[.066]	.203(.056)[.056]
FCGMM	.297(.042)[.042]	.319(.157)[.158]	.196(.066)[.066]	.203(.056)[.056]
	$\bar{r} = 60$		$c_n/n = 1.300$	
2SLS (few IVs)	.310(.138)[.138]	.330(.186)[.189]	.204(.050)[.051]	.206(.045)[.045]
2SLS (many IVs)	.282(.028)[.033]	–	.198(.046)[.046]	.206(.040)[.041]
FC2SLS	.303(.045)[.045]	–	.201(.046)[.046]	.205(.041)[.041]
GMM (few IVs)	.306(.091)[.091]	.301(.123)[.123]	.202(.048)[.049]	.204(.042)[.043]
GMM (many IVs)	.289(.026)[.029]	.249(.087)[.101]	.199(.046)[.046]	.203(.040)[.040]
FCGMM	.297(.028)[.028]	.321(.102)[.104]	.200(.046)[.046]	.203(.040)[.041]
$m_r = 15$	$\bar{r} = 30$		$c_n/n = 1.051$	
2SLS (few IVs)	.314(.141)[.142]	.333(.224)[.226]	.206(.053)[.054]	.203(.049)[.049]
2SLS (many IVs)	.293(.034)[.035]	–	.202(.051)[.051]	.204(.042)[.042]
FC2SLS	.307(.045)[.045]	–	.203(.051)[.051]	.202(.043)[.043]
GMM (few IVs)	.301(.086)[.086]	.310(.144)[.144]	.203(.052)[.052]	.202(.044)[.044]
GMM (many IVs)	.296(.031)[.031]	.290(.104)[.104]	.201(.050)[.050]	.200(.042)[.042]
FCGMM	.301(.032)[.032]	.307(.107)[.108]	.201(.050)[.050]	.199(.042)[.042]
	$\bar{r} = 60$		$c_n/n = 1.101$	
2SLS (few IVs)	.296(.089)[.089]	.320(.153)[.154]	.200(.037)[.037]	.199(.036)[.036]
2SLS (many IVs)	.288(.024)[.026]	–	.198(.037)[.038]	.200(.032)[.032]
FC2SLS	.300(.025)[.025]	–	.199(.037)[.037]	.199(.032)[.032]
GMM (few IVs)	.297(.049)[.050]	.305(.089)[.089]	.198(.037)[.037]	.198(.033)[.033]
GMM (many IVs)	.294(.022)[.023]	.286(.069)[.070]	.198(.037)[.037]	.197(.032)[.032]
FCGMM	.299(.023)[.023]	.304(.070)[.070]	.199(.037)[.037]	.197(.032)[.032]

Mean(SD)[RMSE] c_n : average empirical concentration parameter; n : sample size

Table 4: 2SLS and GMM estimation (gamma)

$\sigma_\alpha^2 = 1$	$\lambda_0 = 0.1$	$\rho_0 = 0.1$	$\beta_{01} = 0.2$	$\beta_{02} = 0.2$
$m_r = 10$	$\bar{r} = 30$		$c_n/n = 0.616$	
2SLS (few IVs)	.097(.262)[.263]	.163(.330)[.336]	.205(.071)[.071]	.211(.073)[.074]
2SLS (many IVs)	.065(.073)[.081]	–	.198(.068)[.068]	.215(.053)[.055]
FC2SLS	.113(.096)[.097]	–	.203(.068)[.068]	.208(.055)[.056]
GMM (few IVs)	.096(.168)[.168]	.134(.235)[.238]	.202(.066)[.066]	.208(.061)[.062]
GMM (many IVs)	.085(.059)[.061]	.067(.162)[.165]	.199(.067)[.067]	.206(.053)[.053]
FCGMM	.100(.067)[.067]	.137(.189)[.193]	.201(.067)[.067]	.203(.053)[.053]
	$\bar{r} = 60$		$c_n/n = 0.642$	
2SLS (few IVs)	.091(.157)[.157]	.142(.241)[.245]	.204(.048)[.048]	.207(.047)[.047]
2SLS (many IVs)	.065(.050)[.061]	–	.199(.046)[.046]	.213(.039)[.041]
FC2SLS	.099(.056)[.056]	–	.202(.046)[.046]	.206(.040)[.040]
GMM (few IVs)	.094(.099)[.099]	.124(.157)[.159]	.203(.047)[.047]	.206(.043)[.043]
GMM (many IVs)	.085(.041)[.044]	.074(.105)[.108]	.200(.046)[.046]	.206(.040)[.040]
FCGMM	.095(.047)[.047]	.133(.124)[.129]	.202(.046)[.046]	.204(.040)[.040]
$m_r = 15$	$\bar{r} = 30$		$c_n/n = 0.620$	
2SLS (few IVs)	.093(.164)[.164]	.126(.252)[.253]	.206(.054)[.054]	.206(.054)[.054]
2SLS (many IVs)	.085(.054)[.056]	–	.204(.053)[.054]	.209(.043)[.044]
FC2SLS	.102(.060)[.060]	–	.205(.053)[.054]	.206(.043)[.044]
GMM (few IVs)	.096(.091)[.091]	.111(.158)[.159]	.205(.052)[.053]	.204(.045)[.045]
GMM (many IVs)	.095(.045)[.045]	.089(.110)[.111]	.205(.053)[.053]	.203(.042)[.042]
FCGMM	.099(.047)[.047]	.104(.114)[.114]	.205(.053)[.053]	.202(.042)[.042]
	$\bar{r} = 60$		$c_n/n = 0.627$	
2SLS (few IVs)	.094(.105)[.105]	.117(.195)[.195]	.198(.037)[.037]	.203(.038)[.038]
2SLS (many IVs)	.085(.040)[.042]	–	.197(.036)[.036]	.206(.031)[.032]
FC2SLS	.101(.042)[.042]	–	.198(.036)[.036]	.203(.032)[.032]
GMM (few IVs)	.098(.062)[.062]	.103(.115)[.115]	.198(.036)[.036]	.201(.033)[.033]
GMM (many IVs)	.096(.033)[.033]	.087(.080)[.081]	.197(.036)[.036]	.201(.031)[.031]
FCGMM	.100(.035)[.035]	.102(.084)[.084]	.198(.036)[.036]	.200(.031)[.031]

Mean(SD)[RMSE] c_n : average empirical concentration parameter; n : sample size