Specification and Estimation of Social Interaction Models with Network Structures*

Lung-fei Lee
Department of Economics,
Ohio State University

Xiaodong Liu
Department of Economics,
University of Colorado at Boulder

Xu Lin
Department of Economics,
Wayne State University

First version: January 2009
Second revision: October 2009
Current revised version: December 2009

Abstract

This paper considers the specification and estimation of social interaction models with network structures and the presence of endogenous, contextual and correlated effects. With macro group settings, group-specific fixed effects are also incorporated in the model. The network structure provides information on the identification of the various interaction effects. We propose a quasi-maximum likelihood approach for the estimation of the model. We derive the asymptotic distribution of the proposed estimator, and provide Monte Carlo evidence on its small sample performance.

JEL classification: C13, C21

Keywords: social interactions, network, identification, maximum likelihood, peer effects.

*We appreciate valuable comments and suggestions from the editor and two anonymous referees to improve the presentation of this paper. Lee appreciates having financial support for the research from NSF grant SES-0519204. Lin appreciates having financial support for the research from NSFC grant 70701020. An earlier version of this paper is circulated under the title, "Specification and estimation of social interaction models with network structure, contextual factors, correlation and fixed effects". Corresponding address: Xiaodong Liu, Department of Economics, University of Colorado, Boulder, CO 80309. E-mail: xiaodong.liu@colorado.edu.

1 Introduction

Social interaction models study how interaction among individuals can lead to collective behavior and aggregate patterns (Anselin, 2006). Such models are subjects of interest in the new social economics (Durlauf and Young, 2001). Empirical studies on social interactions can be found in Case (1991; 1992) on consumption pattern and technology adoption; Bertrand et al. (2000) on welfare cultures; and Sacerdote (2001), Hanushek et al. (2003) and Lin (2005; 2008) on student achievement, to name a few. For these studies, an individual belongs to a social group. The individuals within a group may interact with each other.

A general social interaction model incorporates endogenous effects, contextual effects, and unobserved correlation effects. Identification of the endogenous interaction effect from the other effects is the main interest in social interaction models (see, eg., Manski, 1993; Moffitt, 2001). In his seminal work, Manski (1993) has shown that linear regression models where the endogenous effect is specified in terms of the group mean would suffer from the 'reflection problem'. The various interaction effects cannot be separately identified.

Lee (2004; 2007) recognizes that many of the empirical studies of social interactions in a group setting have their model specifications related to the spatial autoregressive (SAR) model in the spatial econometrics literature (see, eg., Case, 1991; Bertrand et al., 2000; Moffitt, 2001; Hanushek et al., 2003). Lee (2007) considers the SAR model in a group setting which allows endogenous group interactions, contextual factors, and group-specific fixed effects. Lee's (2007) group interaction model assumes that an individual is equally influenced by all others in that group, so that the endogenous effect and contextual effect are specified, respectively, as the average outcomes and characteristics of the peers. Lee (2007) shows that the identification of the various social interaction effects is possible if there are sufficient variations in group sizes in the sample. The identification, however, can be weak if all of the group sizes are large.

When there is no information on how individuals interact within a group, Lee's (2007) group interaction model is practical by assuming an individual is equally influenced by the peers. In some data sets which are designed for the study of social interactions, information on the network structure within a group may be available. An example is the Add Health data (Udry, 2003), where there is information on the 'named' friends within a grade or a school of each student in the sample. Such information on the connections of each individual (node) in a group (network) may be captured by the spatial weights matrix in a SAR model. Different from the equally weighted group interaction matrix in Lee (2007), the network weights matrix can be asymmetric and its off-diagonal entries may be zeros. Such a weights matrix introduces more nonlinearity for identification of various social interaction effects beyond the variation of group sizes.

Lin (2005) recognizes the value of the network structure and has estimated a network model on student academic achievement using the Add Health data. Lin's (2005) model has the specification of a SAR model, which includes group-specific fixed effects, in addition to endogenous and contextual effects. Lin (2005) has discussed the difference of the network model with the linear-in-mean model of Manski (1993) and argued that the information on network structure helps identification. However, formal identification conditions have not been explicitly derived in that paper. Subsequently, Bramoullé et al. (2009) investigate the identification of the network model in Lin (2005) by focusing on the network operator of the reduced form equation.

This paper discusses the specification, identification and estimation issues of the network model. The sample consists of many different groups and a network is formed among individuals within a group. To capture the group unobservables, a group dummy is included. As there are many groups in the sample, the joint estimation of the group fixed effects with the structural parameters will create the 'incidental parameter' problem (Neyman and Scott, 1948). For this reason, Lee (2007) considers the within estimation method for the group interaction model, and Lin (2005) takes the difference of an individual's outcome from the average outcomes of his/her named friends (or connections) to eliminate the group fixed effects. For the within equation, Lee (2007) discusses the 2SLS and (conditional) maximum likelihood (ML) methods for the model estimation. He shows that the ML method is efficient relative to the 2SLS. On the other hand, the empirical model in Lin (2005) is estimated by the 2SLS after the elimination of group fixed effects.

The model considered in this paper has a similar specification as the network model in Lin (2005). In addition, we allow the disturbances of connected individuals to be correlated, so that the selection effect in a network can be partially captured.[1] We characterize the identification conditions of the extended SAR model based on features of the network structure, the role of exogenous variables, and the presence of correlated disturbances. We propose an alternative method to eliminate the group

[1] If the network formation is endogenous due to the similar preference of connected individuals as argued in Moffitt (2001), disturbances of connected individuals may be correlated. Therefore, correlated disturbances shall be allowed in order to capture the endogenous network formation, which is regarded as an important selection issue in the empirical literature. Although network formation is assumed exogenous in this paper, such a specification of disturbances is in the right direction for a better model.

fixed effects. We compare the performance of the proposed elimination method with that of Lin (2005) in terms of estimation efficiency. For the estimation, we propose a quasi-maximum likelihood (QML) method which is computationally tractable and efficient relative to the 2SLS method. This likelihood is a partial likelihood in the terminology of Cox (1975).

The rest of the paper is organized as follows. Section 2 presents the SAR model with network structures. We interpret the specification of the model and discuss identification and estimation issues. Section 3 suggests a transformation of the model to eliminate group fixed effects. The implementation of the QML estimation of the transformed model is discussed in Section 4. Section 5 characterizes the identification conditions of the model and establishes the consistency of the QML estimator (QMLE). Section 6 derives the asymptotic distribution of the QMLE and compares the efficiency properties of the QMLE with the 2SLS estimator (2SLSE). Section 7 investigates the finite sample performance of the estimation methods, and consequences of model misspecifications via Monte Carlo experiments. Section 8 briefly concludes.[2]

2 The Network Model with Macro Groups

The model under consideration has the specification

$$Y_{nr} = \lambda_0 W_{nr} Y_{nr} + X_{nr}\beta_{10} + W_{nr}X_{nr}\beta_{20} + l_{m_r}\alpha_{r0} + u_{nr}, \tag{1}$$

where $u_{nr} = \rho_0 M_{nr} u_{nr} + \epsilon_{nr}$ for $r = 1, \cdots, \bar{r}$. $\bar{r}$ 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. $Y_{nr} = (y_{1r}, \cdots, y_{m_r,r})'$ is an m_r-dimensional vector of y_{ir}, where y_{ir} is the observed outcome of the ith member in the macro group r. W_{nr} and M_{nr} are nonstochastic $m_r \times m_r$ network weights matrices, which may or may not be the same.[3] X_{nr} is an $m_r \times k$ matrix of exogenous variables.[4] l_{m_r} is an m_r-dimensional vector of ones. $\epsilon_{nr} = (\epsilon_{nr,1}, \cdots, \epsilon_{nr,m_r})'$ is an m_r-dimensional vector of disturbances, where $\epsilon_{nr,i}$'s are i.i.d. with zero mean and variance σ_0^2.

The specification of the weights matrices W_{nr} and M_{nr} in (1) captures the network structure

[2] An empirical application to illustrate the practical use of the specified model and the proposed estimation method can be found in the working paper version of this paper and Lin (2008).

[3] Some empirical studies assume $M_r = W_r$, (see, e.g., Cohen, 2002; Fingleton, 2008). On the other hand, some discussions on the possibility that $M_r \neq W_r$ can be found in LeSage (1999, pp. 87-88).

[4] Sometimes, model (1) can be specified as $Y_{nr} = \lambda_0 W_{nr} Y_{nr} + X_{1nr}\beta_{10} + W_{nr}X_{2nr}\beta_{20} + l_{m_r}\alpha_{r0} + u_{nr}$ with $u_{nr} = \rho_0 M_{nr} u_{nr} + \epsilon_{nr}$. Here we assume $X_{1nr} = X_{2nr} = X_{nr}$ wlog. If X_{1nr} and X_{2nr} are not the same, they may be expanded to an X_{nr} which contains all the distinct columns of X_{1nr} and X_{2nr}. In that case, β_{10} and β_{20} will have zero restrictions in some of their entries.

of the macro group r.[5] In a group interaction model, with no information on how individuals interact within a group, it is typical to assume that each group member is equally affected by all the other members in that group, so that the weights matrix takes the special form $W^e_{nr} = M^e_{nr} = \frac{1}{m_r-1}(l_{m_r}l'_{m_r} - I_{m_r})$.[6] On the other hand, some data sets (e.g. the Add Health data as mentioned above) have information on the network structure. With such information, the (i,j) entry of the weights matrix is a non-zero constant if i is influenced by j, and zero otherwise. In principle, the influence is not necessarily reciprocal, and hence the weights matrices can be asymmetric. In the paper, we focus on the case that W_{nr} and M_{nr} are row-normalized such that the sum of each row is unity, i.e., $W_{nr}l_{m_r} = M_{nr}l_{m_r} = l_{m_r}$. Row normalization is popular in empirical studies of social interactions as $W_{nr}Y_{nr}$ can be then interpreted as the (weighted) average outcome (or behavior) of the peers.[7]

The network model (1) is an equilibrium model in the sense that the observed outcomes Y_{nr} are simultaneously determined through the network structure within a group under the assumption that $(I_{m_r} - \lambda_0 W_{nr})$ is invertible.[8] This model may have different economic contents under different contexts. One may interpret the equations in (1) as reaction functions in the industrial organization literature. Or, Y_{nr} may be regarded as the outcomes of the Nash equilibrium in a peer effect game (see Case et al., 1993; Calvó-Armengol et al., 2006). In the spatial econometrics literature, the model (1) is an extended SAR model with SAR disturbances.[9] A typical SAR model, however, does not have a macro group structure so group-specific effects are absent. As a model in the framework of social network, which is our main focus, $W_{nr}Y_{nr}$ captures the possible endogenous social interactions effect with the coefficient λ_0, $W_{nr}X_{nr}$ captures the contextual effect with the coefficient β_{20}. The endogenous effect refers to the contemporaneous influences of peers. The contextual effect includes characteristics of peers unaffected by the current behavior. The incorporation of the contextual

[5] In an empirical study, one might have different specifications of the network weights matrix. The model with a different network weights matrix would be a different model and we would have a model selection problem in practice. Some Monte Carlo studies in Lee (2008) provide evidence that the model selection based on the maximized likelihood values can be quite effective. Such a model selection issue is interesting and important but is not the focus of this paper.

[6] A list of frequently used notations is provided in the Appendix for the convenience of reference.

[7] In some cases, however, row-normalization is not plausible. For example, if a row has all zero elements, then it is impossible to normalize that row to one. Also, sometimes one may be interested in the aggregate influence rather than the average influence of the peers. Liu and Lee (2009) have proposed a GMM approach to estimate a social interaction model where the weights matrix is not row-normalized. The two models with or without row-normalization might address different empirical motivations and they can be complementary to each other.

[8] A sufficient condition for $(I_{m_r} - \lambda_0 W_{nr})$ to be invertible is that $||\lambda_0 W_{nr}|| < 1$ for some matrix norm $||\cdot||$. For the case where W_{nr}, with all entries being non-negative, is row-normalized, a sufficient condition is $|\lambda_0| < 1$.

[9] In the terminology of spatial econometrics, $W_{nr}X_{nr}$ is called a exogenous spatial lag (Florax and Folmer, 1992), and a model with such a term is referred to as a spatial Durbin model (Anselin, 1988).

variables, here $W_{nr}X_{nr}$, in addition to X_{nr}, has a long history in the social interaction literature in sociology before simultaneity is allowed in the model. α_{r0} captures the unobserved group-specific effect, and $M_{nr}u_{nr}$ captures the unobserved correlation effect among connected individuals with the coefficient ρ_0.[10]

Manski's (1993) reflection problem refers to the difficulty to distinguish between behavioral and contextual factors. Moffitt (2001) argues that the basic identification problem is how to distinguish correlations of outcomes that arise from social interactions from correlations that arise from correlated group unobservables. He believes in two generic sources of correlated unobservables – one from preferences or other forces that lead certain types of individuals to be grouped together, and the second from some unobservable common environmental factors. For our generalized network model with macro groups, while the second source is captured by the group-specific effect α_{r0}, the first source may be captured by the correlation effect parameter ρ_0.

Although we treat α_{r0} as the unobserved group effect of a macro group, such as a school-grade, this specification can be generalized if there are several network components in a macro group. In the terminology of networks, a component is formed by a maximal set of individuals directly or indirectly related to each other. A macro group may be regarded as the platform for a social network. A social network may have a single or several components. In some applications, one may prefer to introduce a separate dummy for each component within a group instead of a single group dummy. Such a generalization will be accommodated by model (1) as we might regard each component as a group instead.

[10] In the literature of spatial econometrics, several approaches have been suggested for the specification of the form of spatial error dependence. In model (1), the regression error term u_{nr} is assumed to follow a SAR process. Under this specification, all the observations in a group are related to each other, with a decreasing correlation with higher orders of contiguity. Hence, such a structure is desirable as it induces global spatial autocorrelation within a group (Anselin, 2006). As an alternative, one can model the structure of spatial correlation based on a moving average process. However, such a specification only represents a local pattern of autocorrelation. For example, with a first order moving average specification, there is no spatial covariance beyond the second neighbor (Anselin, 2006). We have shown, in Appendix E, that the proposed QML method can be extended to the model where the disturbances follow a more general spatial ARMA process. In some cases, one could model the spatial error dependence by assuming that the spatial correlation is a function of the distance between two observations (Cressie, 1993). Such a specification could be useful for geostatistic models but might be less so for social network models. For example, if the social network can be represented in a graph, the relationship between nodes could simply be represented by a binary indicator which is one for connected nodes and zero for unconnected ones. This is the case for the Add Health data (Udry, 2003) that we have applied the proposed method to the empirical studies in our previous version and that in Lin (2008). In addition to these specifications of the disturbances, another possibility is to leave the covariance structure unspecified such as those in Conly (1999) and Kelejian and Prucha (2007). For that alternative strategy, the main interest is to provide HAC covariance matrices for the 2SLS and/or GMM estimators. Our paper does not follow the latter strategy as our interest is to consider efficient estimation for the model as well as the variance structure of the disturbances.

3 Elimination of the Macro Group Fixed Effects

In this paper, we allow the distribution of α_{r0} to depend on X_{nr} and W_{nr}. We consider the estimation of the model conditioning on α_{r0}'s by treating α_{r0}'s as unknown parameters (as in the panel econometrics literature). To avoid the incidental parameter problem, we shall have the fixed effect parameter eliminated.

In a linear panel regression model or a logit panel regression model with fixed effects, the fixed effect parameter can be eliminated by the method of conditional likelihood when effective sufficient statistics can be found for each of the fixed effects. For those panel models, time average of the dependent variable provides the sufficient statistic (see Chamberlain, 1980; Hsiao, 2003). However, effective sufficient statistics might not be available for many other models. The well-known example is the probit panel regression model, where time average of the dependent variable does not provide the sufficient statistic, even though probit and logit models are close substitutes. For the group interaction model in Lee (2007), due to the specific structure of W_{nr}^e, the group average, i.e., $\bar{y}_r = \frac{1}{m_r}\sum_{i=1}^{m_r} y_{ir}$, does provide an effective sufficient statistic to eliminate the fixed effect parameter α_{r0}. The observation deviated from the group mean $(y_{ir} - \bar{y}_r)$ does not involve the fixed effect α_{r0} and hence can be used in the conditional likelihood function for the estimation of the structural parameters. For a general network weights matrix W_{nr}, $\bar{y}_r$ might not be a sufficient statistic for α_{r0}.[11] Even so, this paper suggests a method which eliminates the fixed effects and allows the estimation of the remaining parameters of interest via a QML framework by exploring the row-normalization property of the weights matrices.

To simplify repeated notations, let $S_{nr}(\lambda) = I_{m_r} - \lambda W_{nr}$, $S_{nr} = S_{nr}(\lambda_0)$, $R_{nr}(\rho) = I_{m_r} - \rho M_{nr}$, and $R_{nr} = R_{nr}(\rho_0)$. The reduced form equation of (1) is $Y_{nr} = S_{nr}^{-1}(Z_{nr}\beta_0 + l_{m_r}\alpha_{r0} + R_{nr}^{-1}\epsilon_{nr})$, where $Z_{nr} = (X_{nr}, W_{nr}X_{nr})$ and $\beta_0 = (\beta_{10}', \beta_{20}')'$. A Cochrane-Orcutt type transformation introduces i.i.d. disturbances so that

$$R_{nr}S_{nr}Y_{nr} = R_{nr}Z_{nr}\beta_0 + (1-\rho_0)l_{m_r}\alpha_{r0} + \epsilon_{nr}, \tag{2}$$

as $R_{nr}l_{m_r} = (1-\rho_0)l_{m_r}$. Let $J_{nr} = I_{m_r} - \frac{1}{m_r}l_{m_r}l_{m_r}'$ be the derivation from group mean projector.

[11] The model (1) implies that $\bar{y}_r = \frac{1}{m_r}l_{m_r}'Y_{nr} = \frac{\lambda_0}{m_r}l_{m_r}'W_{nr}Y_{nr} + \frac{1}{m_r}l_{m_r}'X_{nr}\beta_{10} + \frac{1}{m_r}l_{m_r}'W_{nr}X_{nr}\beta_{20} + \alpha_{r0} + \frac{1}{m_r}l_{m_r}'u_{nr}$. $\bar{y}_r$ does not provide a sufficient statistic for α_{r0} when $l_{m_r}'W_{nr}$ is not proportional to l_{m_r}' because $l_{m_r}'W_{nr}Y_{nr}$ may not be a function of $\bar{y}_r$.

Premultiplication of (2) by J_{nr} eliminates α_{r0}'s, so we have

$$J_{nr}R_{nr}S_{nr}Y_{nr} = J_{nr}R_{nr}Z_{nr}\beta_0 + J_{nr}\epsilon_{nr}. \tag{3}$$

The transformed disturbances $J_{nr}\epsilon_{nr}$ are linearly dependent because its variance matrix $\sigma_0^2 J_{nr}$ is singular. For an essentially equivalent but more effective transformation, we consider the orthonormal matrix of J_{nr} given by $[F_{nr}, l_{m_r}/\sqrt{m_r}]$. The columns in F_{nr} are eigenvectors of J_{nr} corresponding to the eigenvalue one, such that $F'_{nr}l_{m_r} = 0$, $F'_{nr}F_{nr} = I_{m_r^*}$ and $F_{nr}F'_{nr} = J_{nr}$, where $m_r^* = m_r - 1$. Premultiplication of (2) by F_{nr} leads to a transformed model without α_{r0}'s, $F'_{nr}R_{nr}S_{nr}Y_{nr} = F'_{nr}R_{nr}Z_{nr}\beta_0 + F'_{nr}\epsilon_{nr}$. By Lemma C.1,[12] this implies that

$$(F'_{nr}R_{nr}F_{nr})(F'_{nr}S_{nr}F_{nr})F'_{nr}Y_{nr} = (F'_{nr}R_{nr}F_{nr})F'_{nr}Z_{nr}\beta_0 + F'_{nr}\epsilon_{nr}. \tag{4}$$

Denote $Y^*_{nr} = F'_{nr}Y_{nr}$, $Z^*_{nr} = F'_{nr}Z_{nr}$, $W^*_{nr} = F'_{nr}W_{nr}F_{nr}$, $M^*_{nr} = F'_{nr}M_{nr}F_{nr}$, $S^*_{nr}(\lambda) = F'_{nr}S_{nr}(\lambda)F_{nr} = I_{m_r^*} - \lambda W^*_{nr}$, and $R^*_{nr}(\rho) = F'_{nr}R_{nr}(\rho)F_{nr} = I_{m_r^*} - \rho M^*_{nr}$. Furthermore, denote $S^*_{nr} = S^*_{nr}(\lambda_0)$ and $R^*_{nr} = R^*_{nr}(\rho_0)$ for simplicity. The transformed model (4) can be rewritten more compactly as

$$R^*_{nr}S^*_{nr}Y^*_{nr} = R^*_{nr}Z^*_{nr}\beta_0 + \epsilon^*_{nr}, \tag{5}$$

where $\epsilon^*_{nr} = F'_{nr}\epsilon_{nr}$ is an m_r^*-dimensional disturbance vector with zero mean and variance matrix $\sigma_0^2 I_{m_r^*}$. Equation (5) is used for the estimation of the structural parameters in the model.

Some features in (5) may not conform to a typical SAR model. A spatial weights matrix in a conventional SAR model is specified to have zero diagonal elements. Such a specification facilitates the interpretation of spatial effects of neighboring units on a spatial unit and excludes self-influence. A zero diagonal spatial weights matrix is also utilized in Moran's test of spatial independence so as the test statistic has zero mean under the null hypothesis of spatial independence. Many articles on spatial econometrics maintain this assumption. While W_{nr} and M_{nr} have zero diagonals, the transformed W^*_{nr} and M^*_{nr} do not.[13] Also even though W_{nr} and M_{nr} are row-normalized, the transformed W^*_{nr} and M^*_{nr} do not preserve this feature. However, these do not turn out to be difficult issues for understanding asymptotic properties of estimators. The difficulty from the analytic point

[12] See Appendix C for some useful lemmas.

[13] As $\mathrm{tr}(W_{nr}) = 0$ and $W_{nr}l_{m_r} = l_{m_r}$, W^*_{nr} has nonzero diagonal elements because $\mathrm{tr}(W^*_{nr}) = \mathrm{tr}(W_{nr}F_{nr}F'_{nr}) = \mathrm{tr}(W_{nr}J_{nr}) = \mathrm{tr}(W_{nr}) - \frac{1}{m_r}\mathrm{tr}(W_{nr}l_{m_r}l'_{m_r}) = -1$.

of view is on the uniform boundedness properties of the transformed spatial matrices. Furthermore, when elements of ϵ_{nr} are i.i.d., the elements of ϵ^*_{nr} are only uncorrelated but, in general, not necessarily independent. So asymptotic results which are developed for the estimation of a typical SAR model, e.g., the QMLE in Lee (2004), may not directly apply. The following section discusses the implementation of the QML method for the transformed model.

4 Quasi Maximum Likelihood Estimation

Let $\epsilon^*_{nr}(\delta) = R^*_{nr}(\rho)[S^*_{nr}(\lambda)Y^*_{nr} - Z^*_{nr}\beta]$, where $\delta = (\beta', \lambda, \rho)'$. For a sample with $\bar{r}$ macro groups, the log likelihood function is

$$\ln L_n(\theta) = -\frac{n^*}{2}\ln(2\pi\sigma^2) + \sum_{r=1}^{\bar{r}} \ln|S^*_{nr}(\lambda)| + \sum_{r=1}^{\bar{r}} \ln|R^*_{nr}(\rho)| - \frac{1}{2\sigma^2}\sum_{r=1}^{\bar{r}} \epsilon^{*\prime}_{nr}(\delta)\epsilon^*_{nr}(\delta), \quad (6)$$

where $\theta = (\delta', \sigma^2)'$ and $n^* = \sum_{r=1}^{\bar{r}} m^*_r = n - \bar{r}$ is the number of effective sample observations. The likelihood function has a partial likelihood (Cox, 1975) interpretation as showed in Appendix D.

In order to implement the QML, the determinant and inverse of $S^*_{nr}(\lambda)$ and $R^*_{nr}(\rho)$ are needed. As $|S^*_{nr}(\lambda)| = \frac{1}{1-\lambda}|S_{nr}(\lambda)|$ and $|R^*_{nr}(\rho)| = \frac{1}{1-\rho}|R_{nr}(\rho)|$ by Lemma C.1, the tractability of evaluating $|S^*_{nr}(\lambda)|$ and $|R^*_{nr}(\rho)|$ is exactly that of $|S_{nr}(\lambda)|$ and $|R_{nr}(\rho)|$.[14] Furthermore, as $S^*_{nr}(\lambda)^{-1} = F'_{nr}S_{nr}(\lambda)^{-1}F_{nr}$ and $R^*_{nr}(\rho)^{-1} = F'_{nr}R_{nr}(\rho)^{-1}F_{nr}$ by Lemma C.1, $S^*_{nr}(\lambda)$ and $R^*_{nr}(\rho)$ are invertible as long as the original matrices $S_{nr}(\lambda)$ and $R_{nr}(\rho)$ are invertible.

Let $\epsilon_{nr}(\delta) = R_{nr}(\rho)[S_{nr}(\lambda)Y_{nr} - Z_{nr}\beta]$. As $\epsilon^*_{nr}(\delta) = F'_{nr}\epsilon_{nr}(\delta)$ by Lemma C.2, it follows that $\epsilon^{*\prime}_{nr}(\delta)\epsilon^*_{nr}(\delta) = \epsilon'_{nr}(\delta)J_{nr}\epsilon_{nr}(\delta)$ because $J_{nr} = F_{nr}F'_{nr}$. Denote $Y_n = (Y'_{n1}, \cdots, Y'_{n\bar{r}})'$, $X_n = (X'_{n1}, \cdots, X'_{n\bar{r}})'$, $Z_n = (Z'_{n1}, \cdots, Z'_{n\bar{r}})'$, $\epsilon_n = (\epsilon'_{n1}, \cdots, \epsilon'_{n\bar{r}})'$, $W_n = \mathrm{Diag}\{W_{n1}, \cdots, W_{n\bar{r}}\}$, $M_n = \mathrm{Diag}\{M_{n1}, \cdots, M_{n\bar{r}}\}$, and $J_n = \mathrm{Diag}\{J_{n1}, \cdots, J_{n\bar{r}}\}$. The log likelihood function can be evaluated without F_{nr}'s as

$$\begin{aligned}
&\ln L_n(\theta) \\
= {} & -\frac{n^*}{2}\ln(2\pi\sigma^2) + \sum_{r=1}^{\bar{r}} \ln\frac{|S_{nr}(\lambda)|}{1-\lambda} + \sum_{r=1}^{\bar{r}} \ln\frac{|R_{nr}(\rho)|}{1-\rho} - \frac{1}{2\sigma^2}\sum_{r=1}^{\bar{r}} \epsilon'_{nr}(\delta)J_{nr}\epsilon_{nr}(\delta) \\
= {} & -\frac{n^*}{2}\ln(2\pi\sigma^2) + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - \bar{r}\ln[(1-\lambda)(1-\rho)] - \frac{1}{2\sigma^2}\epsilon'_n(\delta)J_n\epsilon_n(\delta), \quad (7)
\end{aligned}$$

[14] When W_{nr} and M_{nr} are constructed as row normalized weights matrices from original symmetric matrices, Ord (1975) suggests a computational tractable method for the evaluation of $|S_{nr}(\lambda)|$ and $|R_{nr}(\rho)|$. This will also be useful for evaluating $|S^*_{nr}(\lambda)|$ and $|R^*_{nr}(\rho)|$, even though the row sums of the transformed spatial weights matrices W^*_{nr} and M^*_{nr} may not be unity.

where $\epsilon_n(\delta) = R_n(\rho)[S_n(\lambda)Y_n - Z_n\beta]$, $S_n(\lambda) = I_n - \lambda W_n$ and $R_n(\rho) = I_n - \rho M_n$. For simplicity, let $S_n = S_n(\lambda_0)$ and $R_n = R_n(\rho_0)$.

In Lee's (2007) group interaction model, $\rho_0 = 0$ and $W_{nr} = W^e_{nr} = \frac{1}{m^*_r}(l_{m_r}l'_{m_r} - I_{m_r})$. Hence, the likelihood function (7) becomes

$$\ln L_n(\theta) = -\frac{n^*}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{r=1}^{\bar{r}} m^*_r[\ln(m^*_r + \lambda) - \ln m^*_r] - \frac{1}{2\sigma^2}\sum_{r=1}^{\bar{r}} \epsilon'_{nr}(\delta)J_{nr}\epsilon_{nr}(\delta),$$

where $J_{nr}\epsilon_{nr}(\delta) = \frac{m^*_r+\lambda}{m^*_r}J_{nr}Y_{nr} - (J_{nr}X_{nr}, -\frac{1}{m^*_r}J_{nr}X_{nr})\beta$, because $|I_{m_r} - \lambda W_{nr}| = (1-\lambda)(\frac{m^*_r+\lambda}{m^*_r})^{m^*_r}$, $J_{nr}W_{nr} = -\frac{1}{m^*_r}J_{nr}$. This is exactly the one derived in Lee (2007). Thus, the proposed estimation approach in this paper generalizes Lee (2007).

For computational and analytical simplicity, the concentrated log likelihood can be derived by concentrating out β and σ^2. From (7), given $\gamma = (\lambda, \rho)'$, the QMLE of β_0 is given by $\hat{\beta}_n(\gamma) = [Z'_nR'_n(\rho)J_nR_n(\rho)Z_n]^{-1}Z'_nR'_n(\rho)J_nR_n(\rho)S_n(\lambda)Y_n$, and the QMLE of σ^2_0 is given by

$$\begin{aligned}\hat{\sigma}^2_n(\gamma) &= \frac{1}{n^*}[S_n(\lambda)Y_n - Z_n\hat{\beta}_n(\gamma)]'R'_n(\rho)J_nR_n(\rho)[S_n(\lambda)Y_n - Z_n\hat{\beta}_n(\gamma)] \\ &= \frac{1}{n^*}Y'_nS'_n(\lambda)R'_n(\rho)P_n(\rho)R_n(\rho)S_n(\lambda)Y_n,\end{aligned}$$

where $P_n(\rho) = J_n - J_nR_n(\rho)Z_n[Z'_nR'_n(\rho)J_nR_n(\rho)Z_n]^{-1}Z'_nR'_n(\rho)J_n$ and $P_n = P_n(\rho_0)$ for simplicity. The concentrated log likelihood function of γ is

$$\ln L_n(\gamma) = -\frac{n^*}{2}(\ln(2\pi) + 1) - \frac{n^*}{2}\ln\hat{\sigma}^2_n(\gamma) + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - \bar{r}\ln[(1-\lambda)(1-\rho)]. \tag{8}$$

The QMLE $\hat{\gamma}_n = (\hat{\lambda}_n, \hat{\rho}_n)'$ is the maximizer of the concentrated log likelihood (8). The QMLEs of β_0 and σ^2_0 are, respectively, $\hat{\beta}_n(\hat{\gamma}_n)$ and $\hat{\sigma}^2_n(\hat{\gamma}_n)$. For asymptotic analysis, we assume the following regularity conditions.

Assumption 1 The $\{\epsilon_{nr,i}\}_{i=1,\cdots,m_r,r=1,\cdots,\bar{r}}$ are i.i.d. with mean zero and variance σ^2_0.[15] The moment $\mathrm{E}(|\epsilon_{nr,i}|^{4+\eta})$ for some $\eta > 0$ exists.

[15] Homoskedasticity might be a restrictive assumption, but it is beyond the scope of this paper to incorporate heteroskedasticity. Under unknown heteroskedasticity, one might need consider an alternative estimation strategy like an IV-based method (see, eg., Lin and Lee, 2006; Kelejian and Prucha, 2007). However, the IV (or, in general, moment-based) estimation method can be sensitive in non-obvious ways to various implementation issues such as the interaction between the choice of instruments and the specification of the model (LeSage and Pace, 2009, p.56). Furthermore, the IV estimates can be imprecise when instruments are weak. For these reasons, we focused on likelihood-based techniques in this paper.

Assumption 2 The elements of Z_n are uniformly bounded constants for all n.[16] Z_n has the full rank $2k$, and $\lim_{n\to\infty} \frac{1}{n} Z_n' R_n' J_n R_n Z_n$ exists and is nonsingular.

Assumption 3 The sequences of row-normalized spatial weights matrices $\{W_n\}$ and $\{M_n\}$ are uniformly bounded in both row and column sums in absolute value.[17]

Assumption 4 $\{S_n^{-1}(\lambda)\}$ and $\{R_n^{-1}(\rho)\}$ are uniformly bounded in both row and column sums in absolute value uniformly in γ in a compact parameter space Γ, with the true $\gamma_0 = (\lambda_0, \rho_0)'$ in the interior of Γ.

The higher than the fourth moment condition in Assumption 1 is needed in order to apply a central limit theorem due to Kelejian and Prucha (2001). The nonstochastic Z_n and its uniform boundedness conditions in the first half of Assumption 2 are for analytical simplicity. The $R_n Z_n$ are regressors transformed by the spatial filter R_n, and the $J_n R_n Z_n$ are those transformed by the deviation form group means projector J_n. The second half of Assumption 2 assumes that the exogenous regressors $J_n R_n Z_n$ in the transformed model (3) are not multicollinear. Assumption 3 limits the spatial dependence among the units to a tractable degree and is originated by Kelejian and Prucha (1999). It rules out the unit root case (in time series as a special case). Assumption 4 deals with the parameter space of γ to make sure that $\ln|S_n(\lambda)|$, $\ln|R_n(\rho)|$, $\ln[(1-\lambda)(1-\rho)]$, and their related derivatives are well behaved. As shown in Lee (2004), if $\|W_n\| \le 1$ and $\|M_n\| \le 1$ where $\|\cdot\|$ is a matrix norm, then $\{||S_n^{-1}(\lambda)||\}$ and $\{||R_n^{-1}(\rho)||\}$ are uniformly bounded in any subset of $(-1, 1)$ bounded away from the boundary.

5 Identification and Consistency

There is a fundamental identification issue for the network model different from the reflection problem in Manski (1993) if $\lambda_0\beta_{10} + \beta_{20} = 0$ and $W_n = M_n$, as summarized in the following lemma.

Lemma 5.1 *If $\lambda_0\beta_{10} + \beta_{20} = 0$ and $W_n = M_n$, then the endogenous effect parameter λ_0, the contextual effect parameter β_{20} and the correlated effect parameter ρ_0 cannot be separately identified.*

[16] If Z_{nr} 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 Z_{nr} instead. Furthermore, if Z_{nr} is allowed to be correlated with ϵ_{nr}, then we have an endogenous regressor problem. In that case, estimation methods such as IV, etc., which takes into account the endogeneity issue, would be needed.

[17] A sequence of square matrices $\{A_n\}$, where $A_n = [a_{n,ij}]$, is said to be uniformly bounded in row sums (column sums) in absolute value if the sequence of row sum matrix norm $\|A_n\|_\infty = \max_{i=1,\cdots,n} \sum_{j=1}^n |a_{n,ij}|$ (column sum matrix norm $\|A_n\|_1 = \max_{j=1,\cdots,n} \sum_{i=1}^n |a_{n,ij}|$) is bounded. (Horn and Johnson, 1985)

This problem is revealing from the reduced form equation of (5), which is

$$Y_{nr}^* = S_{nr}^{*-1} Z_{nr}^* \beta_0 + S_{nr}^{*-1} R_{nr}^{*-1} \epsilon_{nr}^*. \tag{9}$$

With the restriction $\lambda_0\beta_{10} + \beta_{20} = 0$, $Z_{nr}^*\beta_0 = X_{nr}^*\beta_{10} + W_{nr}^* X_{nr}^* \beta_{20} = (I_{m_r} - \lambda_0 W_{nr}^*) X_{nr}^* \beta_{10} = S_{nr}^* X_{nr}^* \beta_{10}$, and, hence, the reduced form equation (9) becomes $Y_{nr}^* = X_{nr}^*\beta_{10} + v_{nr}$, where $v_{nr} = S_{nr}^{*-1} R_{nr}^{*-1} \epsilon_{nr}^*$. While β_{10} can be identified from the mean regression $E(Y_{nr}^*|X_{nr}^*) = X_{nr}^*\beta_{10}$, both λ_0 and β_{20} can not be identified as they are not in the mean regression equation. On the other hand, the disturbances v_{nr} follows a high-order SAR process, $v_{nr} = \rho_0 M_{nr}^* v_{nr} + \lambda_0 W_{nr}^* v_{nr} - \rho_0\lambda_0 M_{nr}^* W_{nr}^* v_{nr} + \epsilon_{nr}^*$, where the identification conditions have been considered in Lee and Liu (2008). If $W_{nr} \neq M_{nr}$ so that $W_{nr}^* \neq M_{nr}^*$, ρ_0 and λ_0 can be identified from the correlation structure of v_{nr}. β_{20} can then be identified via the restriction $\beta_{20} = -\lambda_0\beta_{10}$ once λ_0 is identified. However, when $M_{nr} = W_{nr}$, $v_{nr} = (\rho_0 + \lambda_0) W_{nr}^* v_{nr} - \rho_0\lambda_0 W_{nr}^{*2} v_{nr} + \epsilon_{nr}^*$, and hence ρ_0 and λ_0 can only be locally identified but can not be separately identified.

An interpretation of the situation $\lambda_0\beta_{10} + \beta_{20} = 0$ is that (1) does not represent a reaction function with simultaneity but a model with spurious social correlation among peers. This is because, under the restriction $\lambda_0\beta_{10} + \beta_{20} = 0$, (1) can be generated from the panel regression model $Y_{nr}^* = X_{nr}^*\beta_{10} + v_{nr}$ with SAR disturbances. Let $\beta_{10,j}$ and $\beta_{20,j}$ be the jth element of β_{10} and β_{20} respectively. The spurious social correlation model can be ruled out when $\beta_{20,j} \neq 0$ and $\beta_{10,j} = 0$ for some j, or, in another word, there is a relevant variable in X_n that affects Y_n only through the contextual effect $W_n X_n$. For the linear-in-mean model of Manski (1993), the identification of endogenous and exogenous interaction effects depends crucially on the existence of relevant variables in X_{nr} that directly affect Y_n. For the network model, it is the behavioral interpretation of the parameters that can be problematic when $\lambda_0\beta_{10} + \beta_{20} = 0$.[18]

The transformed equilibrium vector $J_n R_n(\rho) Y_n$ for any ρ in its parameter space can be represented as

$$J_n R_n(\rho) Y_n = \lambda_0 J_n R_n(\rho) G_n Z_n \beta_0 + J_n R_n(\rho) Z_n \beta_0 + J_n R_n(\rho) S_n^{-1} R_n^{-1} \epsilon_n, \tag{10}$$

[18] This restriction can be tested even λ_0 and β_{20} were not identifiable. One may test the significance of the added regressor vector in the expanded equation $Y_{nr}^* = X_{nr}^*\beta_{10} + W_{nr}^* X_{nr}^* \zeta + v_{nr}$ by testing that $\zeta = 0$.

because $S_n^{-1} = \lambda_0 G_n + I_n$ where $G_n = W_n S_n^{-1}$. A sufficient condition for global identification of θ_0 is that the generated regressors $J_n R_n(\rho) G_n Z_n \beta_0$ and $J_n R_n(\rho) Z_n$ are not asymptotically multicollinear, and the variance of matrix of $J_n R_n^{-1} \epsilon_n$ is unique. Let $\sigma_{a,n}^2(\rho) = \frac{\sigma_0^2}{n^*} \mathrm{tr}(R_n'^{-1} R_n'(\rho) J_n R_n(\rho) R_n^{-1})$ and $\sigma_n^2(\gamma) = \frac{\sigma_o^2}{n^*} \mathrm{tr}([R_n(\rho) S_n(\lambda) S_n^{-1} R_n^{-1}]' J_n [R_n(\rho) S_n(\lambda) S_n^{-1} R_n^{-1}])$.

Assumption 5 Either (a) $\lim_{n\to\infty} \frac{1}{n^*}[G_n Z_n \beta_0, Z_n]' R_n'(\rho) J_n R_n(\rho) [G_n Z_n \beta_0, Z_n]$ exists and is non-singular for each possible ρ in its parameter space and $\lim_{n\to\infty} \frac{1}{n^*}\Big\{ \ln |\sigma_{a,n}^2(\rho) R_n^{-1}(\rho) J_n R_n'^{-1}(\rho)| - \ln |\sigma_0^2 R_n^{-1} J_n R_n'^{-1}| \Big\} \neq 0$, for any $\rho \neq \rho_0$; or (b) for any $\gamma \neq \gamma_0$,

$$\lim_{n\to\infty} \frac{1}{n^*}\Big\{ \ln |\sigma_n^2(\gamma)[S_n^{-1}(\lambda) R_n^{-1}(\rho)] J_n [S_n^{-1}(\lambda) R_n^{-1}(\rho)]'| - \ln |\sigma_0^2 (S_n^{-1} R_n^{-1}) J_n (S_n^{-1} R_n^{-1})'| \Big\} \neq 0.$$

The rank condition on $J_n R_n(\rho)[G_n Z_n \beta_0, Z_n]$ in Assumption 5(a) is for the identification of λ_0 and β_0 from the deterministic component of the reduced form equation (10). The following Lemmas provide some sufficient conditions which imply this rank condition.

Lemma 5.2 *If $\beta_{20} + \lambda_0 \beta_{10} \neq 0$ and $[X_{nr}, W_{nr} X_{nr}, W_{nr}^2 X_{nr}, l_{m_r}]$ has full column rank for some group r, then $J_n R_n(\rho)[G_n Z_n \beta_0, Z_n]$ has full column rank.*

Lemma 5.2 gives a sufficient condition for the rank condition in Assumption 5(a) based on the network structure of a single group, which is feasible only if the size of that group is greater than $3k+1$, where k is the column rank of X_{nr}. If there are not enough members in any of the groups in the sample, information across groups need to be explored to achieve identification. A sufficient condition for the rank condition in Assumption 5(a) based on the whole sample is given as follows.

Lemma 5.3 *If $\beta_{20} + \lambda_0 \beta_{10} \neq 0$ and $J_n[X_n, W_n X_n, W_n^2 X_n]$ has full column rank, then $J_n R_n(\rho)[G_n Z_n \beta_0, Z_n]$ has full column rank.*

The group interaction model in Lee (2007) has the spatial weights matrix $W_{nr}^e = \frac{1}{m_r - 1}(l_{m_r} l_{m_r}' - I_{m_r})$. As $J_{nr} W_{nr}^e = -\frac{1}{(m_r-1)} J_{nr}$, $J_{nr}[X_{nr}, W_{nr}^e X_{nr}, (W_{nr}^e)^2 X_{nr}] = J_{nr}[X_{nr}, -\frac{1}{(m_r-1)} X_{nr}, \frac{1}{(m_r-1)^2} X_{nr}]$ does not have full column rank. Identification is not possible only with a single group. On the other hand, let c_1, c_2 and c_3 be conformable vectors such that $J_{nr} X_{nr} c_1 + J_{nr} W_{nr}^e X_{nr} c_2 + J_{nr}(W_{nr}^e)^2 X_{nr} c_3 = 0$, or more explicitly, $J_{nr} X_{nr}[c_1 - \frac{1}{(m_r-1)} c_2 + \frac{1}{(m_r-1)^2} c_3] = 0$ by $J_{nr} W_{nr}^e = -\frac{1}{(m_r-1)} J_{nr}$. As $J_{nr} X_{nr} \neq 0$, if there are at least three distinct values for m_r's in the sample, the equality holds only if $c_1 = c_2 = c_3 = 0$. Hence, if there is sufficient group size variations, then

$J_n[X_n, W_n X_n, W_n^2 X_n]$ has full column rank, which implies the rank condition in Assumption 5(a) holds by Lemma 5.3.

The identification of the endogenous effect, and hence, the exogenous effect, may be intuitively illustrated via the reduced form. For a group r,

$$\begin{aligned} Y_{nr}^* &= S_{nr}^{*-1}(X_{nr}^*\beta_{10} + W_{nr}^* X_{nr}^*\beta_{20}) + S_{nr}^{*-1}R_{nr}^{*-1}\epsilon_{nr}^* \\ &= X_{nr}^*\beta_{10} + \sum_{j=0}^{\infty}\lambda_0^j W_{nr}^{*j+1}X_{nr}^*(\beta_{20}+\lambda_0\beta_{10}) + S_{nr}^{*-1}R_{nr}^{*-1}\epsilon_{nr}^*, \end{aligned}$$

because $S_{nr}^{*-1} = \sum_{j=0}^{\infty}\lambda_0^j W_{nr}^{*j}$ when $\sup||\lambda_0 W_{nr}^*||_\infty < 1$. The effects of X_{nr}^* on Y_{nr}^* can be decomposed in layers. The direct effect of X_{nr}^* is captured by β_{10}, the effect due to immediate neighbors is captured by $(\beta_{20}+\lambda_0\beta_{10})$, and that due to neighbors of neighbors in the second layer is captured by $(\beta_{20}+\lambda_0\beta_{10})\lambda_0$ with the discount factor λ_0. So if the immediate neighbors can be distinguished from the second layer neighbors, the discount factor provides the identification of the endogenous effect λ_0. For the case with $W_{nr}^e = \frac{1}{m_r-1}(l_{m_r}l_{m_r}' - I_{m_r})$, as $(F_{nr}'W_{nr}^e F_{nr})^2 = -\frac{1}{m_r-1}F_{nr}'W_{nr}^e F_{nr}$, the net effect of X_{nr}^* on Y_{nr}^* through W_{nr}^e in the group r is captured by the coefficient $(\beta_{20}+\lambda_0\beta_{10})\sum_{j=0}^{\infty}(-\frac{\lambda_0}{m_r-1})^j$. Hence, the endogenous effect λ_0 can be identified only by comparing these net effects across groups with different sizes.

An example where the rank condition above fails is the complete bipartite network, where individuals in a group are divided into two blocks such that each individual in one block is connected to all individuals in the other block but none in the same block, and vice versa. These include the star network where one individual is connected to all other individuals in a group and all the others in the group connect only to him. This example is due to Bramoullé et al. (2009) for a different transformation.[19] For the complete bipartite network, $W_{nr} = \begin{pmatrix} 0 & \frac{1}{m_{r2}}l_{m_{r1}}l_{m_{r2}}' \\ \frac{1}{m_{r1}}l_{m_{r2}}l_{m_{r1}}' & 0 \end{pmatrix}$ with $m_{r1}+m_{r2} = m_r$. It implies that $W_{nr}^2 = \begin{pmatrix} \frac{1}{m_{r1}}l_{m_{r1}}l_{m_{r1}}' & 0 \\ 0 & \frac{1}{m_{r2}}l_{m_{r2}}l_{m_{r2}}' \end{pmatrix}$. Consequently, $W_{nr}+W_{nr}^2 = [\frac{1}{m_{r1}}l_{m_r}l_{m_{r1}}', \frac{1}{m_{r2}}l_{m_r}l_{m_{r2}}']$ with all its columns proportional to l_{m_r}. This implies, in particular, the column space spanned by the columns of $J_{nr}[W_{nr}X_{nr}, W_{nr}^2 X_{nr}]$ contains l_{m_r}. So if all groups in a sample consist of bipartite networks, the rank condition in Assumption 5(a) may not hold.

[19] Bramoullé et al. (2009) point out this underidentification case for the model with the transformation $I_{m_r} - W_{nr}$, which has been utilized in Lin (2005), to eliminate the group effect.

We have discussed the rank condition in Assumption 5 (a) for the identification of λ_0 and β_0 in the mean regression function of the reduced form equation. The second part of Assumption 5 (a) is for the identification of ρ_0 in the SAR error process. It is clear that ρ_0 can not be identified from the mean regression function, as $R_n(\rho)$ only plays the role of weighting sample observations for efficient estimation. So, ρ_0 needs to be identified from the disturbances $S_n^{-1}R_n^{-1}\epsilon_n$. On the other hand, when $J_nR_n(\rho)G_nZ_n\beta_0$ and $J_nR_n(\rho)Z_n$ are linearly dependent or asymptotically multicollinear as n goes to infinity, a global identification condition would be related to the uniqueness of the variance matrix of J_nY_n, which is given by Assumption 5(b).[20]

Finally, we would like to point out that the division by the effective sample size n^* in the limiting conditions in Assumption 5 (as well as in the following Assumption 6) has ruled out the case of large group interactions, which have been considered in Lee (2004; 2007). In that case, both the endogenous and exogenous interaction effects would be weakly identified and their rates of convergence can be quite low (Lee, 2004; 2007). But for network models, one has emphasized on 'small world', as the main interest in the network literature. This is also for the empirical application in this paper.

Let $Q_n(\gamma) = \max_{\beta,\sigma^2} \mathrm{E}(\ln L_n(\theta))$. The solutions of this maximization problem are $\beta_n^*(\gamma) = [Z_n'R_n'(\rho)J_nR_n(\rho)Z_n]^{-1}Z_n'R_n'(\rho)J_nR_n(\rho)S_n(\lambda)S_n^{-1}Z_n\beta_0$, and

$$\begin{aligned}\sigma_n^{*2}(\gamma) &= \frac{1}{n^*}\mathrm{E}\{[S_n(\lambda)Y_n - Z_n\beta_n^*(\gamma)]'R_n'(\rho)J_nR_n(\rho)[S_n(\lambda)Y_n - Z_n\beta_n^*(\gamma)]\} \\ &= \frac{1}{n^*}(\lambda_0-\lambda)^2(G_nZ_n\beta_0)'R_n'(\rho)P_n(\rho)R_n(\rho)G_nZ_n\beta_0 \\ &\quad + \frac{\sigma_0^2}{n^*}\mathrm{tr}[(S_n^{-1}R_n^{-1})'S_n'(\lambda)R_n'(\rho)J_nR_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}].\end{aligned}$$

Hence,

$$Q_n(\gamma) = -\frac{n^*}{2}(\ln(2\pi)+1) - \frac{n^*}{2}\ln\sigma_n^{*2}(\gamma) + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - \bar{r}\ln[(1-\lambda)(1-\rho)]. \qquad (11)$$

Identification of γ_0 can be based on the maximum of $\frac{1}{n^*}Q_n(\gamma)$. With identification and uniform convergence of $\frac{1}{n^*}\ln L_n(\gamma) - \frac{1}{n^*}Q_n(\gamma)$ to zero on Γ, consistency of $\hat{\theta}_n$ follows.

Proposition 1 *Under Assumptions 1-5, θ_0 is globally identifiable and $\hat{\theta}_n$ is a consistent estimator*

[20] The identification of λ_0 and/or ρ_0 via the variance structure is exactly those for SAR models in Lee (2004) and Lee and Liu (2008).

of θ_0.

6 Asymptotic Distributions

From the Taylor expansion of $\frac{\partial \ln L_n(\hat{\theta}_n)}{\partial \theta} = 0$, it follows that $\sqrt{n}(\hat{\theta}_n - \theta_0) = (\frac{1}{n}\frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'})^{-1}\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$, for some $\tilde{\theta}_n$ between $\hat{\theta}_n$ and θ_0. The first order derivatives of the log likelihood function at θ_0 given in Appendix B are linear or quadratic functions of ϵ_n. The asymptotic distribution of the first order derivatives may be derived from central limit theorems in Kelejian and Prucha (2001).

The variance matrix of $\frac{1}{\sqrt{n^*}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}$ is $\mathrm{E}[\frac{1}{\sqrt{n^*}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{n^*}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta'}] = \Sigma_{\theta,n} + \Omega_{\theta,n}$, where $\Sigma_{\theta,n} = -\mathrm{E}[\frac{1}{n^*}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}]$ is the symmetric average Hessian matrix, and $\Omega_{\theta,n}$ is a symmetric matrix such that $\Omega_{\theta,n} = 0$ when $\epsilon_{nr,i}$'s are normally distributed.[21] Assumption 5(a) is sufficient to guarantee that the limiting average Hessian matrix is nonsingular. If γ_0 is a regular point (Rothenberg, 1971), as Assumption 5(b) is a global identification condition which implies local identification, the limiting average Hessian matrix will also be nonsingular. The sufficient condition which complements Assumption 5(b) for this purpose is given as follows. Let $A^s = A + A'$ for a square matrix A. Let $C_n = J_nR_nG_nR_n^{-1} - \frac{1}{n}\mathrm{tr}(J_nR_nG_nR_n^{-1})I_n$ and $D_n = J_nH_n - \frac{1}{n}\mathrm{tr}(J_nH_n)I_n$.

Assumption 6 $\lim_{n\to\infty}(\frac{1}{n^*})^2[\mathrm{tr}(D_n^sD_n^s)\mathrm{tr}(C_n^sC_n^s) - \mathrm{tr}^2(C_n^sD_n^s)] > 0$.

Proposition 2 *Under Assumptions 1-4 and 5(a); or 1-4, 5(b) and 6,* $\sqrt{n^*}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \Sigma_\theta^{-1} + \Sigma_\theta^{-1}\Omega_\theta\Sigma_\theta^{-1})$, *where* $\Omega_\theta = \lim_{n\to\infty}\Omega_{\theta,n}$ *and* $\Sigma_\theta = \lim_{n\to\infty}\Sigma_{\theta,n}$, *which are assumed to exist. If* $\epsilon_{nr,i}$*'s are normally distributed, then* $\sqrt{n^*}(\hat{\theta}_n - \theta_0) \xrightarrow{D} N(0, \Sigma_\theta^{-1})$.

For the transformed model $J_nY_n = \lambda_0J_nW_nY_n + J_nZ_n\beta_0 + J_nR_n^{-1}\epsilon_n$, a computational convenient estimation method is the generalized 2SLS (G2SLS) by Kelejian and Prucha (1998). In the first step of G2SLS, $\zeta_0 = (\beta_0', \lambda_0)'$ will be estimated by the 2SLS with an IV matrix Q_{1n}, $\hat{\zeta}_{2sls,n} = [(Z_n, W_nY_n)'J_nP_{1n}J_n(Z_n, W_nY_n)]^{-1}(Z_n, W_nY_n)'J_nP_{1n}J_n'Y_n$ where $P_{1n} = Q_{1n}(Q_{1n}'Q_{1n})^{-1}Q_{1n}'$. With the initial 2SLSE $\hat{\zeta}_{2sls,n}$, $J_nR_n^{-1}\epsilon_n$ can be estimated as a residual and ρ_0 will be estimated by a method of moments (MOM) in Kelejian and Prucha (1999). Let $\hat{\rho}_{mom,n}$ be the consistent MOM estimate of ρ_0 and $\hat{R}_n = R_n(\hat{\rho}_{mom,n})$. The feasible G2SLS estimator (G2SLSE) of ζ_0 in the model is

$$\hat{\zeta}_{g2sls,n} = [(Z_n, W_nY_n)'\hat{R}_n'J_nP_{2n}J_n\hat{R}_n(Z_n, W_nY_n)]^{-1}(Z_n, W_nY_n)'\hat{R}_n'J_nP_{2n}J_n\hat{R}_nY_n,$$

[21] The explicit expressions of $\Sigma_{\theta,n}$ and $\Omega_{\theta,n}$ are given in Appendix B.

where $P_{2n} = Q_{2n}(Q'_{2n}Q_{2n})^{-1}Q'_{2n}$ for some IV matrix Q_{2n}. The G2SLSE is consistent and asymptotically normal with

$$\sqrt{n^*}(\hat{\zeta}_{g2sls,n} - \zeta_0) \xrightarrow{D} N(0, \sigma_0^2[\lim_{n\to\infty} \frac{1}{n^*}(Z_n, G_nZ_n\beta_0)'R'_nJ_nP_{2n}J_nR_n(Z_n, G_nZ_n\beta_0)]^{-1}).$$

It follows from the generalized Schwarz inequality that the best selection of Q_{2n} is $J_n\hat{R}_n[Z_n, G_n(\hat{\lambda}_n)Z_n\hat{\beta}_n]$, where $G_n(\lambda) = W_nS_n^{-1}(\lambda)$, and the variance matrix of the best G2SLS estimator $\hat{\zeta}_{b2sls,n}$ is $\frac{1}{n^*}\Sigma_{\zeta,n}^{-1}$, where $\Sigma_{\zeta,n} = \frac{1}{\sigma_0^2 n^*}(Z_n, G_nZ_n\beta_0)'R'_nJ_nR_n(Z_n, G_nZ_n\beta_0)$. When $\epsilon_{nr,i}$'s are normally distributed, the variance matrix of $\hat{\zeta}_{b2sls,n}$ can be easily compared with that of the MLE.

Proposition 3 *When $\epsilon_{nr,i}$'s are normally distributed, the MLE is more efficient than the best G2SLS estimator.*

7 Monte Carlo Results

To investigate the finite sample performance of the MLE, we consider the following model

$$Y_{nr} = \lambda_0 W_{nr}Y_{nr} + X_{nr}\beta_{10} + W_{nr}X_{nr}\beta_{20} + l_{m_r}\alpha_{r0} + u_{nr},$$

where $u_{nr} = \rho_0 W_{nr}u_{nr} + \epsilon_{nr}$ and $\epsilon_{nr} \sim N(0, \sigma_0^2 I_{m_r})$, for $r = 1, \cdots, \bar{r}$. The weights matrix W_{nr} is based on the Add Health data (see Udry, 2003). For the Monte Carlo study, we consider 4 samples. The first sample consists of groups with the group size less than or equal to 30. There are 102 such groups in the data with 1344 observations and average group size being 13.1. We also consider a sub-sample with the group size less than or equal to 15. In the data, there are 67 such small groups with 557 observations and average group size being 8.3. To facilitate comparison, we also randomly pick 67 groups with the group size less than or equal to 30 and 102 groups with the group size less than or equal to 50 from the data. For the first sample of randomly picked groups, the sample size is 877 with the average group size being 13.1. For the second one, the sample size is 2279 with the average group size being 22.3. This allows us to inspect the effect of increasing the number of groups $\bar{r}$ and increasing the average group size separately. The number of repetitions is 400 for each case in this Monte Carlo experiment. For each repetition, X_{nr} and α_{r0} are generated from $N(0, I_m)$ and $N(0, 2)$ respectively, for $r = 1, \cdots, \bar{r}$. The data are generated with $\lambda_0 = \rho_0 = 0.5$ and $\sigma_0^2 = 1$. β_{10} and β_{20} are varied in the experiments.

The estimation methods considered are the 2SLS with the IV matrix $Q_{1n} = J_n(Z_n, W_n Z_n, W_n^2 Z_n, W_n^3 Z_n)$, the G2SLS with the IV matrix Q_{1n} in the first step and $Q_{2n} = J_n \hat{R}_n(Z_n, W_n Z_n, W_n^2 Z_n, W_n^3 Z_n)$ in the last step,[22] and the ML approach proposed in this paper (labeled ML1 in the following tables). Lin (2005) suggests an alternative elimination method of the fixed effects by the transformation using $(I_{m_r} - W_{nr})$. (See Appendix F for more details.)[23] However, as the rank of $(I_{m_r} - W_{nr})$ may be less than m_r^*, more linear dependence is induced when eliminating the fixed effects. Hence, this alternative elimination method may be less efficient. We also report the MLE (labeled ML2 in the following tables) based on the alternative elimination method using $(I_{m_r} - W_{nr})$ in the Monte Carlo experiments. We report the mean 'Mean' and standard deviation 'SD' of the empirical distributions of the estimates. To facilitate the comparison of various estimators, their root mean square errors 'RMSE' are also reported.

Table 1 reports the results in the case with $\beta_{10} = \beta_{20} = 1$, i.e., the regressors are 'strong'. For all sample sizes considered, the G2SLS estimates of ρ_0 are downward biased. The bias reduces as the average group size increases. The other estimates are essentially unbiased. In terms of the SD, G2SLS improves 2SLS upon the estimates of λ_0, β_{10} and β_{20} and ML improves G2SLS upon estimates of ρ_0. For the same estimator, the SDs decrease as either $\bar{r}$ or the average group size increases.

In the case when the regressors are 'weak' with $\beta_{10} = \beta_{20} = 0.2$, the estimation results are summarized in Table 2. When $\bar{r} = 67$, the 2SLS and G2SLS estimates of λ_0 are upward biased. The G2SLS estimates of ρ_0 and the 2SLS and G2SLS estimates of β_{20} are downward biased. When $\bar{r}$ increases to 102, the 2SLS estimates of λ_0 become downward biased with a smaller magnitude, and the other biases also reduce. The MLE of ρ_0 is slightly downward biased with the sample of small groups. The bias reduces as the sample size increases. The MLEs also have smaller SDs and RMSEs than the other estimates for all sample sizes considered. For example, when $n = 2279$, the percentage reduction in SD of the MLEs of λ_0, ρ_0 and β_{20} relative to the G2SLS estimates is, respectively, 42.0%,

[22] In the finite sample, the best G2SLS with $Q_{2n} = J_n \hat{R}_n(Z_n, G_n(\hat{\lambda}_n) Z_n \hat{\beta}_n)$ is quite sensitive to the initial estimates. As the initial 2SLSEs are obtained with no restrictions on the parameter space, the initial estimate of $\hat{\lambda}_n$ could have an absolute value greater than one. This causes the estimated best IV used in the second step problematic. In the case when $\beta_{10} = \beta_{20} = 0.2$, and $n = 557$, about 1/10 of the replications had an initial estimate with $|\hat{\lambda}_n| > 1$. In the Monte Carlo experiments, we use the above simpler Q_{2n} instead to avoid the effect of bad initial estimates.

[23] We have experimented with the iterated G2SLS. However, for many replications, the iterated estimator failed to converge. For example, when $\beta_{10} = \beta_{20} = 0.2$, and $n = 557$, the iterated estimator failed to converge in about 1/4 of the replications. This issue tends to occur especially when some estimates of λ_0 are out of bound, i.e., with an absolute value greater than one, during the iterations. Note that 2SLS approach does not impose restrictions on $|\lambda| < 1$. Even for the converged iterated G2SLS estimates, there is no evidence that iteration procedure improves the performance of the G2SLS estimator in this specific simulation experiment. Hence, we choose not to report the simulation results on the iterated G2SLS in this paper.

47.1% and 15.9%. The percentage reduction is even larger with smaller samples. For both cases with 'strong' and 'weak' regressors, MLEs based on the alternative elimination method of the fixed effects by the transformation using $(I_{m_r} - W_{nr})$ have larger SDs than those of the MLEs proposed in this paper.

Results in Table 3 inspect the effects of model misspecification on the MLEs using the sample with 102 moderate size groups. When positive endogenous effects captured by λ_0 is ignored in the estimation, $\hat{\rho}_n$ and $\hat{\beta}_{2n}$ will be upward biased. When positive exogenous effects captured by β_{20} is ignored in the estimation, $\hat{\lambda}_n$ will be upward biased, and $\hat{\beta}_{1n}$ and $\hat{\rho}_n$ will be downward biased. When a positive spatial correlation with ρ_0 in the disturbances fails to be modeled, $\hat{\lambda}_n$ will be upward biased and $\hat{\beta}_{2n}$ will be downward biased. The bias of $\hat{\lambda}_n$ can be large enough to change its sign in the case when $\lambda_0 < 0$. The opposite occurs when the omitted ρ_0 has a negative value. The bottom panel of Table 3 studies the effects of misspecified weights matrices in a model with i.i.d. disturbances ($\rho_0 = 0$). The weights matrices W_{nr} in the data generating process is specified as above. However, suppose, when estimating this model, we don't have the information on the network structure and put equal weight on each member of a group as in the model with group interaction so that $W_{nr}^e = \frac{1}{m_r^*}(l_{m_r} l_{m_r}' - I_{m_r})$ is used. With the misspecified W_{nr}, $\hat{\lambda}_n$, $\hat{\beta}_{1n}$ and $\hat{\beta}_{2n}$ are upward biased by 65.2%, 31.6% and 83.3%, respectively. SD of $\hat{\lambda}_n$ also dramatically increases relative to the estimate with correctly specified W_{nr}. We also compare the likelihood values of ML estimation of the correctly specified model with those of the misspecified models. We find a larger likelihood value indicates a better specified model in most cases.

8 Conclusion

This paper considers model specification, identification and estimation of a social interaction model. The social interaction model generalizes the group interaction model in Lee (2007), where an individual in a group interacts with all other members with equal weights, to the situation that each individual may have their own connected peers. This model extends the SAR model with SAR errors to incorporate contextual variables and group unobservables. The social interactions are rich in that endogenous interaction effects, contextual effects, group-specific effects, and correlations among connected individuals in a network can all be captured in the model. The incorporation of possible correlations among connected individuals may partially capture the endogeneity of network formation.

The identification of endogenous and contextual effects in Manski's (1993) linear-in-mean model requires the inclusion of some individual exogenous characteristics but the exclusion of their corresponding contextual effects. In the group interaction model in Lee (2007), identification requires variation in group sizes in the sample. For the network model, identification is in general feasible even when groups have the same size because of additional nonlinearity due to the network structure. The identification issue is similar to that of the SAR model but with a slight complication due to the presence of contextual variables and group unobservables. Identification can be based on the mean regression function as well as correlation structure of the dependent variables. In general, all the social interaction effects of interest can be identified in a network model.

We consider the estimation of the network model. As a model with endogeneity, it can in general be estimated by the 2SLS method as instrumental variables can be generated from the network structure with the presence of relevant exogenous variables. The 2SLS method is simple but not efficient. This paper considers a possible extension of the QML method for the group interaction model in Lee (2007) to the general network model. It generalizes the QML method for a SAR model with SAR errors in that there are incidental parameters due to group-specific dummy variables. The QML method is designed after the elimination of group dummies. This strategy may have applications in other models, e.g., the spatial panel data models with time dummies in Lee and Yu (2007). We establish analytically the consistency and asymptotic normality of the estimators and show that the QMLE is asymptotically efficient relative to the G2SLS estimator.

Monte Carlo studies are designed to investigate the finite sample performance of the estimators. The QMLE has better finite sample properties than the 2SLS and G2SLS estimators as confirmed by the Monte Carlo results. We also pay special attention to possible consequences of omitting some social or correlation effects on the estimates of the remaining effects. Furthermore, we provide some limited evidence on possible consequences with the misspecification on network connections and the usefulness of the maximized log likelihood as a model selection criterion.

Appendices

A Summary of Notations

- $\beta = (\beta_1', \beta_2')'$, $\gamma = (\lambda, \rho)'$, $\delta = (\beta', \gamma')'$, $\theta = (\delta', \sigma^2)'$, $\zeta = (\beta', \lambda)'$.
- $n = \sum_{r=1}^{\bar{r}} m_r$, $n^* = \sum_{r=1}^{\bar{r}} m_r^* = n - \bar{r}$, $m_r^* = m_r - 1$.
- l_{m_r} is an m_r-dimensional vector of ones.

- $W_{nr}^e = M_{nr}^e = \frac{1}{m_r - 1}(l_{m_r} l_{m_r}' - I_{m_r})$.

- $Z_{nr} = (X_{nr}, W_{nr} X_{nr})$; $S_{nr}(\lambda) = I_{m_r} - \lambda W_{nr}$, $S_{nr} = S_{nr}(\lambda_0)$; $R_{nr}(\rho) = I_{m_r} - \rho M_{nr}$, $R_{nr} = R_{nr}(\rho_0)$; $G_{nr} = W_{nr} S_{nr}^{-1}$.

- $J_{nr} = I_{m_r} - \frac{1}{m_r} l_{m_r} l_{m_r}'$. $[F_{nr}, l_{m_r}/\sqrt{m_r}]$ is the orthonormal matrix of J_{nr} where F_{nr} corresponds to the eigenvalue one.

- $Y_{nr}^* = F_{nr}' Y_{nr}$, $Z_{nr}^* = F_{nr}' Z_{nr}$, $\epsilon_{nr}^* = F_{nr}' \epsilon_{nr}$; $W_{nr}^* = F_{nr}' W_{nr} F_{nr}$, $M_{nr}^* = F_{nr}' M_{nr} F_{nr}$.

- $S_{nr}^*(\lambda) = F_{nr}' S_{nr}(\lambda) F_{nr} = I_{m_r^*} - \lambda W_{nr}^*$, $S_{nr}^* = S_{nr}^*(\lambda_0)$; $R_{nr}^*(\rho) = F_{nr}' R_{nr}(\rho) F_{nr} = I_{m_r^*} - \rho M_{nr}^*$, $R_{nr}^* = R_{nr}^*(\rho_0)$.

- $\epsilon_{nr}(\delta) = R_{nr}(\rho)[S_{nr}(\lambda) Y_{nr} - Z_{nr}\beta]$, $\epsilon_{nr}^*(\delta) = R_{nr}^*(\rho)[S_{nr}^*(\lambda) Y_{nr}^* - Z_{nr}^*\beta]$

- $Y_n = (Y_{n1}', \cdots, Y_{n\bar{r}}')'$, $X_n = (X_{n1}', \cdots, X_{n\bar{r}}')'$, $Z_n = (Z_{n1}', \cdots, Z_{n\bar{r}}')'$, $\epsilon_n = (\epsilon_{n1}', \cdots, \epsilon_{n\bar{r}}')'$, $W_n = \mathrm{Diag}\{W_{n1}, \cdots, W_{n\bar{r}}\}$, $M_n = \mathrm{Diag}\{M_{n1}, \cdots, M_{n\bar{r}}\}$, and $J_n = \mathrm{Diag}\{J_{n1}, \cdots, J_{n\bar{r}}\}$.

- $H_n = M_n R_n^{-1}$; $\tilde{Z}_n = R_n Z_n$, $\tilde{G}_n = R_n G_n R_n^{-1}$.

- $C_n = J_n \tilde{G}_n - \frac{1}{n}\mathrm{tr}(J_n \tilde{G}_n) I_n$, $D_n = J_n H_n - \frac{1}{n}\mathrm{tr}(J_n H_n) I_n$.

- $P_n(\rho) = J_n - J_n R_n(\rho) Z_n [Z_n' R_n'(\rho) J_n R_n(\rho) Z_n]^{-1} Z_n' R_n'(\rho) J_n$ and $P_n = P_n(\rho_0)$.

- Let $A^s = A + A'$ for a square matrix A. Let $vec_D(A)$ denote the column vector formed with the diagonal elements of a square matrix A.

B The Score Vector and Information Matrix

- The first order derivatives of the log likelihood function at θ_0 are

$$
\begin{aligned}
\frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial \lambda} &= \frac{1}{\sigma_0^2 \sqrt{n^*}} \epsilon_n' J_n \tilde{G}_n \tilde{Z}_n \beta_0 + \frac{1}{\sigma_0^2 \sqrt{n^*}} [\epsilon_n' J_n \tilde{G}_n \epsilon_n - \sigma_0^2 \mathrm{tr}(J_n \tilde{G}_n)], \\
\frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial \rho} &= \frac{1}{\sigma_0^2 \sqrt{n^*}} [\epsilon_n' J_n H_n \epsilon_n - \sigma_0^2 \mathrm{tr}(J_n H_n)], \\
\frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial \beta} &= \frac{1}{\sigma_0^2 \sqrt{n^*}} \tilde{Z}_n' J_n \epsilon_n, \quad \frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial \sigma^2} = \frac{1}{2\sigma_0^4 \sqrt{n^*}} (\epsilon_n' J_n \epsilon_n - n^* \sigma_0^2),
\end{aligned}
$$

- The second order derivatives of the log likelihood function are

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda^2} = -\mathrm{tr}[J_n G_n^2(\lambda)] - \frac{1}{\sigma^2} Y_n' W_n' R_n'(\rho) J_n R_n(\rho) W_n Y_n,$$
$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\beta} = -\frac{1}{\sigma^2} Z_n' R_n'(\rho) J_n R_n(\rho) W_n Y_n, \quad \frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\sigma^2} = -\frac{1}{\sigma^4} Y_n' W_n' R_n'(\rho) J_n \epsilon_n(\delta),$$
$$\frac{\partial^2 \ln L_n(\theta)}{\partial\lambda\partial\rho} = -\frac{1}{\sigma^2} Y_n' W_n' M_n' J_n \epsilon_n(\delta) - \frac{1}{\sigma^2} Y_n' W_n' R_n'(\rho) J_n M_n [S_n(\lambda) Y_n - Z_n\beta],$$
$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\beta'} = -\frac{1}{\sigma^2} Z_n' R_n'(\rho) J_n R_n(\rho) Z_n, \quad \frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4} Z_n' R_n'(\rho) J_n \epsilon_n(\delta),$$
$$\frac{\partial^2 \ln L_n(\theta)}{\partial\beta\partial\rho} = -\frac{1}{\sigma^2} Z_n' M_n' J_n \epsilon_n(\delta) - \frac{1}{\sigma^2} Z_n' R_n'(\rho) J_n M_n [S_n(\lambda) Y_n - Z_n\beta],$$

$$\frac{\partial^2 \ln L_n(\theta)}{\partial\rho^2} = -\mathrm{tr}(J_n [M_n R_n^{-1}(\rho)]^2) - \frac{1}{\sigma^2} [S_n(\lambda) Y_n - Z_n\beta]' M_n' J_n M_n [S_n(\lambda) Y_n - Z_n\beta],$$
$$\frac{\partial^2 \ln L_n(\theta)}{\partial\rho\partial\sigma^2} = -\frac{1}{\sigma^4} \epsilon_n'(\delta) J_n M_n [S_n(\lambda) Y_n - Z_n\beta], \quad \frac{\partial^2 \ln L_n(\theta)}{\partial(\sigma^2)^2} = \frac{n^*}{2\sigma^4} - \frac{1}{\sigma^6} \epsilon_n'(\delta) J_n \epsilon_n(\delta).$$

- The variance matrix of $\frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial\theta}$ is $\mathrm{E}[\frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial\theta} \cdot \frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial\theta'}] = \Sigma_{\theta,n} + \Omega_{\theta,n}$.

$$\Sigma_{\theta,n} = -\mathrm{E}[\frac{1}{n^*} \frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}]$$
$$= \begin{pmatrix} \frac{1}{\sigma_0^2 n^*} \tilde{Z}_n' J_n \tilde{Z}_n & * & * & * \\ \frac{1}{\sigma_0^2 n^*} (\tilde{G}_n \tilde{Z}_n \beta_0)' J_n \tilde{Z}_n & \frac{1}{\sigma_0^2 n^*} (\tilde{G}_n \tilde{Z}_n \beta_0)' J_n (\tilde{G}_n \tilde{Z}_n \beta_0) + \frac{1}{n^*} \mathrm{tr}(\tilde{G}_n^s J_n \tilde{G}_n) & * & * \\ 0_{1\times k} & \frac{1}{n^*} \mathrm{tr}(H_n^s J_n \tilde{G}_n) & \frac{1}{n^*} \mathrm{tr}(H_n^s J_n H_n) & * \\ 0_{1\times k} & \frac{1}{\sigma_0^2 n^*} \mathrm{tr}(J_n \tilde{G}_n) & \frac{1}{\sigma_0^2 n^*} \mathrm{tr}(J_n H_n) & \frac{1}{2\sigma_0^4} \end{pmatrix},$$

and

$$\Omega_{\theta,n} = \begin{pmatrix} 0_{k\times k} & * & * & * \\ \frac{\mu_3}{\sigma_0^4 n^*} vec_D'(J_n \tilde{G}_n) J_n \tilde{Z}_n & \Omega_{\theta,n;22} & * & * \\ \frac{\mu_3}{\sigma_0^4 n^*} vec_D'(J_n H_n) J_n \tilde{Z}_n & \Omega_{\theta,n;32} & \frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4 n^*} vec_D'(J_n H_n) vec_D(J_n H_n) & * \\ 0_{1\times k} & \frac{\mu_4 - 3\sigma_0^4}{2\sigma_0^6 n} \mathrm{tr}(J_n \tilde{G}_n) & \frac{\mu_4 - 3\sigma_0^4}{2\sigma_0^6 n} \mathrm{tr}(J_n H_n) & \frac{(\mu_4 - 3\sigma_0^4) n^*}{4\sigma_0^8 n} \end{pmatrix},$$

where

$$\Omega_{\theta,n;22} = \frac{2\mu_3}{\sigma_0^4 n^*} vec'_D(J_n\tilde{G}_n)J_n\tilde{G}_n\tilde{Z}_n\beta_0 + \frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4 n^*} vec'_D(J_n\tilde{G}_n)vec_D(J_n\tilde{G}_n),$$
$$\Omega_{\theta,n;32} = \frac{\mu_3}{\sigma_0^4 n^*} vec'_D(J_n H_n)J_n\tilde{G}_n\tilde{Z}_n\beta_0 + \frac{\mu_4 - 3\sigma_0^4}{\sigma_0^4 n^*} vec'_D(J_n H_n)vec_D(J_n\tilde{G}_n),$$

and μ_3 and μ_4 being the third and fourth moments of $\epsilon_{nr,i}$ respectively.

C Some Basic Properties

In this appendix, we list some properties which are useful for the proofs of the results in the text. The results in Lemmas 3 to 9 are either straightforward or can been found in Kelejian and Prucha (2001) and Lee (2004). They are listed here for easy reference. Throughout this appendix, the elements v_i's in $V_n = (v_1, \cdots, v_n)'$ are assumed to be i.i.d. with zero mean, finite variance σ^2 and finite fourth moment μ_4.

Lemma C.1 *Suppose W_{nr} is a row-normalized $m_r \times m_r$ matrix, $J_{nr} = I_{m_r} - \frac{1}{m_r} l_{m_r} l'_{m_r}$, and $[F_{nr}, l_{m_r}/\sqrt{m_r}]$ is the orthonormal matrix of J_{nr} where F_{nr} corresponds to the eigenvalue one. Let $W^*_{nr} = F'_{nr} W_{nr} F_{nr}$ and $m^*_r = m_r - 1$. Then (1) $F'_{nr}(I_{m_r} - \lambda W_{nr}) = F'_{nr}(I_{m_r} - \lambda W_{nr})F_{nr}F'_{nr}$, (2) $|I_{m^*_r} - \lambda W^*_{nr}| = \frac{1}{1-\lambda}|I_{m_r} - \lambda W_{nr}|$, (3) $(I_{m^*_r} - \lambda W^*_{nr})^{-1} = F'_{nr}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr}$, and (4) $W^*_{nr}(I_{m^*_r} - \lambda W^*_{nr})^{-1} = (I_{m^*_r} - \lambda W^*_{nr})^{-1}W^*_{nr} = F'_{nr}W_{nr}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr}$.*

Proof. As $F_{nr}F'_{nr} = J_{nr} = I_{m_r} - l_{m_r}l'_{m_r}/m_r$, we have $F'_{nr}(I_{m_r} - \lambda W_{nr}) = F'_{nr}(I_{m_r} - \lambda W_{nr})(F_{nr}F'_{nr} + l_{m_r}l'_{m_r}/m_r) = F'_{nr}(I_{m_r} - \lambda W_{nr})F_{nr}F'_{nr} + F'_{nr}(I_{m_r} - \lambda W_{nr})l_{m_r}l'_{m_r}/m_r$. As W_{nr} is a row-normalized, $F'_{nr}W_{nr}l_{m_r} = F'_{nr}l_{m_r} = 0$. Hence, (1) holds.

To show (2), we note that $(I_{m^*_r} - \lambda W^*_{nr}) = F'_{nr}(I_{m_r} - \lambda W_{nr})F_{nr}$. As

$$[F_{nr}, l_{m_r}/\sqrt{m_r}]'(I_{m_r} - \lambda W_{nr})[F_{nr}, l_{m_r}/\sqrt{m_r}]$$
$$= \begin{pmatrix} F'_{nr}(I_{m_r} - \lambda W_{nr})F_{nr} & F'_{nr}(I_{m_r} - \lambda W_{nr})l_{m_r}/\sqrt{m_r} \\ l'_{m_r}(I_{m_r} - \lambda W_{nr})F_{nr}/\sqrt{m_r} & l'_{m_r}(I_{m_r} - \lambda W_{nr})l_{m_r}/m_r \end{pmatrix} = \begin{pmatrix} F'_{nr}(I_{m_r} - \lambda W_{nr})F_{nr} & 0 \\ 0 & 1 - \lambda \end{pmatrix},$$

because $F'_{nr}W_{nr}l_{m_r} = F'_{nr}l_{m_r} = 0$ and $l'_{m_r}W_{nr}l_{m_r} = m_r$. Hence $|I_{m^*_r} - \lambda W^*_{nr}| = |F'_{nr}(I_{m_r} - \lambda W_{nr})F_{nr}| = \frac{1}{1-\lambda}|I_{m_r} - \lambda W_{nr}|$.

Since $F'_{nr}W_{nr}l_{m_r} = F'_{nr}l_{m_r} = 0$, (3) and (4) can be verified as $(I_{m^*_r} - \lambda W^*_{nr}) \cdot F'_{nr}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr} = F'_{nr}(I_{m_r} - \lambda W_{nr})F_{nr} \cdot F'_{nr}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr} = F'_{nr}(I_{m_r} - \lambda W_{nr})(I_{m_r} - l_{m_r}l'_{m_r}/m_r)(I_{m_r} -$

$\lambda W_{nr})^{-1}F_{nr} = F'_{nr}F_{nr} - F'_{nr}(I_{m_r} - \lambda W_{nr})l_{m_r}l'_{m_r}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr}/m_r = I_{m_r^*}$, and $W^*_{nr}(I_{m_r^*} - \lambda W^*_{nr})^{-1} = F'_{nr}W_{nr}F_{nr} \cdot F'_{nr}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr} = F'_{nr}W_{nr}(I_{m_r} - l_{m_r}l'_{m_r}/m_r)(I_{m_r} - \lambda W_{nr})^{-1}F_{nr} = F'_{nr}W_{nr}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr}$, and $(I_{m_r^*} - \lambda W^*_{nr})^{-1}W^*_{nr} = F'_{nr}(I_{m_r} - \lambda W_{nr})^{-1}W_{nr}F_{nr} = F'_{nr}W_{nr}(I_{m_r} - \lambda W_{nr})^{-1}F_{nr}$. ■

Lemma C.2 $\epsilon^*_{nr}(\delta) = R^*_{nr}(\rho)[S^*_{nr}(\lambda)Y^*_{nr} - Z^*_{nr}\beta] = F'_{nr}R_{nr}(\rho)[S_{nr}(\lambda)Y_{nr} - Z_{nr}\beta]$.

Proof. $\epsilon^*_{nr}(\delta) = R^*_{nr}(\rho)[S^*_{nr}(\lambda)Y^*_{nr} - Z^*_{nr}\beta] = F'_{nr}R_{nr}(\rho)F_{nr} \cdot F'_{nr}[S_{nr}(\lambda)F_{nr} \cdot F'_{nr}Y_{nr} - Z_{nr}\beta] = F'_{nr}R_{nr}(\rho)(I_{m_r} - \frac{1}{m_r}l_{m_r}l'_{m_r})[S_{nr}(\lambda)(I_{m_r} - \frac{1}{m_r}l_{m_r}l'_{m_r})Y_{nr} - Z_{nr}\beta] = F'_{nr}R_{nr}(\rho)[S_{nr}(\lambda)Y_{nr} - Z_{nr}\beta]$ because $F'_{nr}l_{m_r} = F'_{nr}W_{nr}l_{m_r} = F'_{nr}M_{nr}l_{m_r} = 0$. ■

Lemma C.3 *Suppose that* $\{\|W_n\|\}$, $\{\|M_n\|\}$, $\{\|S_n^{-1}\|\}$, *and* $\{\|R_n^{-1}\|\}$, *where* $\|\cdot\|$ *is a matrix norm, are bounded. Then* $\{\|S_n^{-1}(\lambda)\|\}$ *and* $\{\|R_n^{-1}(\rho)\|\}$ *are uniformly bounded in a neighborhood of* λ_0 *and* ρ_0 *respectively.*

Lemma C.4 *Suppose that elements of the* $n \times k$ *matrices* Z_n *are uniformly bounded for all* n*; and the* $\lim_{n\to\infty} \frac{1}{n}Z'_nR'_nJ_nR_nZ_n$ *exists and is nonsingular, then the projectors* P_n *and* $(J_n - P_n)$*, where* $P_n = J_n - J_nR_nZ_n[Z'_nR'_nJ_nR_nZ_n]^{-1}Z'_nR'_nJ_n$*, are uniformly bounded in both row and column sums in absolute value.*

Lemma C.5 *Suppose that the elements of the sequences of vectors* $P_n = (p_{n1}, \cdots, p_{nn})'$ *and* $Q_n = (q_{n1}, \cdots, q_{nn})'$ *are uniformly bounded for all* n*. (1) If* $\{A_n\}$ *are uniformly bounded in either row or column sums in absolute value, then* $|Q'_nA_nP_n| = O(n)$*. (2) If the row sums of* $\{A_n\}$ *and* $\{Z_n\}$ *are uniformly bounded,* $|z_{i,n}A_nP_n| = O(1)$ *uniformly in* i*, where* $z_{i,n}$ *is the* i*th row of* Z_n*.*

Lemma C.6 *Suppose that the elements of the* $n \times n$ *matrices* $\{A_n\}$ *are uniformly bounded, and the* $n \times n$ *matrices* $\{B_n\}$ *are uniformly bounded in column sums (respectively, row sums) in absolute value. Then, the elements of* A_nB_n *(respectively,* B_nA_n*) are uniformly bounded. For both cases,* $tr(A_nB_n) = tr(B_nA_n) = O(n)$*.*

Lemma C.7 *Suppose that* A_n *is an* $n \times n$ *matrix with its column sums being uniformly bounded in absolute value and elements of the* $n \times k$ *matrix* Z_n *are uniformly bounded. Elements* v_i*'s of* $V_n = (v_1, \cdots, v_n)'$ *are i.i.d.*$(0, \sigma^2)$*. Then,* $\frac{1}{\sqrt{n}}Z'_nA_nV_n = O_p(1)$*, Furthermore, if the limit of* $\frac{1}{n}Z'_nA_nA'_nZ_n$ *exists and is positive definite, then* $\frac{1}{\sqrt{n}}Z'_nA_nV_n \xrightarrow{D} N(0, \sigma_0^2 \lim_{n\to\infty} \frac{1}{n}Z'_nA_nA'_nZ_n)$*.*

Lemma C.8 *Let A_n be an $n \times n$ matrix. Then $E(V_n' A_n V_n) = \sigma^2 tr(A_n)$ and $Var(V_n' A_n V_n) = (\mu_4 - 3\sigma^4) vec_D'(A_n) vec_D(A_n) + \sigma^4 [tr(A_n A_n') + tr(A_n^2)]$.*

Lemma C.9 *Suppose that $\{A_n\}$ is a sequence of $n \times n$ matrices uniformly bounded in either row or column sums in absolute value. Then, $E(V_n' A_n V_n) = O(n)$, $Var(V_n' A_n V_n) = O(n)$, $V_n' A_n V_n = O_p(n)$, and $\frac{1}{n}[V_n' A_n V_n - E(V_n' A_n V_n)] = o_p(1)$.*

Lemma C.10 *Suppose that $\{A_n\}$ is a sequence of symmetric $n \times n$ matrices with row and column sums uniformly bounded in absolute value and $\{b_n\}$ is a sequence of n-dimensional constant vectors such that $\sup_n \frac{1}{n} \sum_{i=1}^{n} |b_{ni}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$. The moment $E(|v|^{4+2\eta})$ of v for some $\eta > 0$ exists. Let $\sigma_{Q_n}^2$ be the variance of Q_n where $Q_n = b_n' V_n + V_n' A_n V_n - \sigma^2 tr(A_n)$. Assume that the variance $\sigma_{Q_n}^2$ is bounded away from zero at the rate n. Then $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0,1)$.*

D A Partial Likelihood Justification

The likelihood function (6) is for $F_{nr} R_{nr} Y_{nr}$ given in equation (5). It remains to consider the remaining component $\frac{1}{m_r} l_{m_r}' R_{nr} Y_{nr}$, which is

$$\frac{1}{m_r} l_{m_r}' R_{nr} Y_{nr} = \frac{1}{m_r} l_{m_r}' R_{nr} (\lambda_0 W_{nr} Y_{nr} + Z_{nr} \beta_0) + (1 - \rho_0) \alpha_{r0} + \bar{\epsilon}_r, \tag{12}$$

where $\bar{\epsilon}_r = \frac{1}{m_r} l_{m_r}' \epsilon_{nr}$. As $F_{nr} F_{nr}' = J_{nr}$ and $\frac{1}{m_r} l_{m_r}' R_{nr} l_{m_r} = (1 - \rho_0)$, it follows that

$$\begin{aligned}
\frac{1}{m_r} l_{m_r}' R_{nr} Y_{nr} &= \frac{1}{m_r} l_{m_r}' R_{nr} (F_{nr} F_{nr}' + \frac{1}{m_r} l_{m_r} l_{m_r}') Y_{nr} = \frac{1}{m_r} l_{m_r}' R_{nr} F_{nr} Y_{nr}^* + (1 - \rho_0) \bar{y}_r, && (13) \\
\frac{1}{m_r} l_{m_r}' R_{nr} Z_{nr} &= \frac{1}{m_r} l_{m_r}' R_{nr} (F_{nr} F_{nr}' + \frac{1}{m_r} l_{m_r} l_{m_r}') Z_{nr} = \frac{1}{m_r} l_{m_r}' R_{nr} F_{nr} Z_{nr}^* + (1 - \rho_0) \bar{z}_r, && (14)
\end{aligned}$$

where $\bar{y}_r = \frac{1}{m_r} l_{m_r}' Y_{nr}$ and $\bar{z}_r = \frac{1}{m_r} l_{m_r}' Z_{nr}$. Similarly,

$$\begin{aligned}
\frac{1}{m_r} l_{m_r}' R_{nr} W_{nr} Y_{nr} &= \frac{1}{m_r} l_{m_r}' R_{nr} (F_{nr} F_{nr}' + \frac{1}{m_r} l_{m_r} l_{m_r}') W_{nr} Y_{nr} \\
&= \frac{1}{m_r} l_{m_r}' R_{nr} F_{nr} W_{nr}^* Y_{nr}^* + (1 - \rho_0) \frac{1}{m_r} l_{m_r}' W_{nr} Y_{nr}, && (15)
\end{aligned}$$

where $\frac{1}{m_r} l_{m_r}' W_{nr} Y_{nr} = \frac{1}{m_r} l_{m_r}' W_{nr} (F_{nr} F_{nr}' + \frac{1}{m_r} l_{m_r} l_{m_r}') Y_{nr} = \frac{1}{m_r} l_{m_r}' W_{nr} F_{nr} Y_{nr}^* + \bar{y}_r$. Substitution of (13)-(15) in (12) gives $\frac{1}{m_r} l_{m_r}' R_{nr} F_{nr} Y_{nr}^* + (1 - \rho_0) \bar{y}_r = \lambda_0 [\frac{1}{m_r} l_{m_r}' R_{nr} F_{nr} W_{nr}^* Y_{nr}^* + (1 - \rho_0) \frac{1}{m_r} l_{m_r}' W_{nr} F_{nr} Y_{nr}^* + (1 - \rho_0) \bar{y}_r] + \frac{1}{m_r} l_{m_r}' R_{nr} F_{nr} Z_{nr}^* \beta_0 + (1 - \rho_0) \bar{z}_r \beta_0 + (1 - \rho_0) \alpha_{r0} + \bar{\epsilon}_r$, , or equivalently

$$\bar{y}_r = \frac{\lambda_0}{(1-\lambda_0)(1-\rho_0)}\frac{1}{m_r}l'_{m_r}R_{nr}F_{nr}W^*_{nr}Y^*_{nr} - \frac{1}{(1-\lambda_0)(1-\rho_0)}\frac{1}{m_r}l'_{m_r}R_{nr}F_{nr}Y^*_{nr}$$
$$+\frac{\lambda_0}{(1-\lambda_0)}\frac{1}{m_r}l'_{m_r}W_{nr}F_{nr}Y^*_{nr} + \frac{1}{(1-\lambda_0)(1-\rho_0)}\frac{1}{m_r}l'_{m_r}R_{nr}F_{nr}Z^*_{nr}\beta_0$$
$$+\frac{1}{(1-\lambda_0)}\bar{z}_r\beta_0 + \frac{1}{(1-\lambda_0)}\alpha_{r0} + \frac{1}{(1-\lambda_0)(1-\rho_0)}\bar{\epsilon}_r. \tag{16}$$

As $\bar{\epsilon}_r$ is independent of Y^*_{nr}, conditional on X^*_{nr}, (16) can be regarded as a nonlinear regression equation.

The joint likelihood function of Y^*_{nr} and $\bar{y}_r$ can thus be decomposed into a product of the conditional likelihood of $\bar{y}_r$ given Y^*_{nr} from (16) and the likelihood function of Y^*_{nr} from (5). Therefore, the likelihood function of Y^*_{nr} from the transformation method for (5) is a partial likelihood function. (Cox, 1975; Lancaster, 2000).

E The Likelihood Function of a Network Model with a Spatial ARMA Disturbances

In this Appendix, we show that the proposed QML approach can be generalized to the case where the disturbances follow a more general spatial ARMA process. The generalized model has the specification that $Y_{nr} = \lambda_0 W_{nr}Y_{nr} + Z_{nr}\beta_0 + l_{m_r}\alpha_{r0} + u_{nr}$, where $u_{nr} = \rho_{10}M_{1nr}u_{nr} + \rho_{20}M_{2nr}\epsilon_{nr} + \epsilon_{nr}$ for $r = 1, \cdots, \bar{r}$. In this model, W_{nr}, M_{1nr} and M_{2nr} are row-normalized such that the sum of each row is unity, i.e., $W_{nr}l_{m_r} = M_{1nr}l_{m_r} = M_{2nr}l_{m_r} = l_{m_r}$. As before, let $R_{nr}(\rho) = I_{m_r} - \rho M_{1nr}$ and $R_{nr} = R_{nr}(\rho_{10})$. A Cochrane-Orcutt type transformation gives $R_{nr}S_{nr}Y_{nr} = R_{nr}Z_{nr}\beta_0 + (1-\rho_{10})l_{m_r}\alpha_{r0} + (I_{m_r} + \rho_{20}M_{2nr})\epsilon_{nr}$. Note that $(I_{m_r} + \rho_{20}M_{2nr})^{-1}l_{m_r} = (1+\rho_{20})^{-1}l_{m_r}$ under the assumption that $(I_{m_r} + \rho_{20}M_{2nr})$ is invertible and M_{2nr} is row normalized. It follows that

$$(I_{m_r} + \rho_{20}M_{2nr})^{-1}R_{nr}S_{nr}Y_{nr} = (I_{m_r} + \rho_{20}M_{2nr})^{-1}R_{nr}Z_{nr}\beta_0 + (\frac{1-\rho_{10}}{1+\rho_{20}})l_{m_r}\alpha_{r0} + \epsilon_{nr}.$$

As $F'_{nr}(I_{m_r} + \rho_{20}M_{2nr})^{-1} = F'_{nr}(I_{m_r} + \rho_{20}M_{2nr})^{-1}F_{nr}F'_{nr} = (I_{m^*_r} + \rho_{20}F'_{nr}M_{2nr}F_{nr})^{-1}F'_{nr}$, premultiplication by F_{nr} leads to a transformed model without α_{r0}'s, i.e.,

$$(I_{m^*_r} + \rho_{20}F'_{nr}M_{2nr}F_{nr})^{-1}R^*_{nr}S^*_{nr}Y^*_{nr} = (I_{m^*_r} + \rho_{20}F'_{nr}M_{2nr}F_{nr})^{-1}R^*_{nr}Z^*_{nr}\beta_0 + \epsilon^*_{nr}.$$

Let $\epsilon^*_{nr}(\delta) = (I_{m^*_r} + \rho_{20} F'_{nr} M_{2nr} F_{nr})^{-1} R^*_{nr}(\rho)[S^*_{nr}(\lambda) Y^*_{nr} - Z^*_{nr}\beta]$, where $\delta = (\beta', \lambda, \rho_1, \rho_2)'$. For a sample with $\bar{r}$ macro groups, the log likelihood function is

$$\begin{aligned}\ln L_n(\theta) \quad = \quad & -\frac{n^*}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{r=1}^{\bar{r}} \ln |S^*_{nr}(\lambda)| + \sum_{r=1}^{\bar{r}} \ln |R^*_{nr}(\rho_1)| - \sum_{r=1}^{\bar{r}} \ln |I_{m^*_r} + \rho_2 F'_{nr} M_{2nr} F_{nr}| \\ & -\frac{1}{2\sigma^2} \textstyle\sum_{r=1}^{\bar{r}} \epsilon^{*\prime}_{nr}(\delta)\epsilon^*_{nr}(\delta),\end{aligned}$$

where $\theta = (\delta', \sigma^2)'$. As $|S^*_{nr}(\lambda)| = \frac{1}{1-\lambda}|S_{nr}(\lambda)|$, $|R^*_{nr}(\rho_1)| = \frac{1}{1-\rho_1}|R_{nr}(\rho_1)|$, $|I_{m^*_r} + \rho_2 F'_{nr} M_{2nr} F_{nr}| = \frac{1}{1+\rho_2}|I_{m_r} + \rho_2 M_{2nr}|$, and $\epsilon^*_{nr}(\delta) = F'_{nr}\epsilon_{nr}(\delta)$, where $\epsilon_{nr}(\delta) = (I_{m_r} + \rho_{20} M_{2nr})^{-1} R_{nr}(\rho)[S_{nr}(\lambda) Y_{nr} - Z_{nr}\beta]$, the log likelihood function can be evaluated without F_{nr}'s as

$$\begin{aligned}\ln L_n(\theta) \quad = \quad & -\frac{n^*}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{r=1}^{\bar{r}} \ln \dfrac{|S_{nr}(\lambda)|}{1-\lambda} + \sum_{r=1}^{\bar{r}} \ln \dfrac{|R_{nr}(\rho_1)|}{1-\rho_1} - \sum_{r=1}^{\bar{r}} \ln \dfrac{|I_{m_r} + \rho_2 M_{2nr}|}{1+\rho_2} \\ & -\frac{1}{2\sigma^2} \textstyle\sum_{r=1}^{\bar{r}} \epsilon'_{nr}(\delta) J_{nr} \epsilon_{nr}(\delta).\end{aligned}$$

F An Alternative Elimination Method of the Macro Group Fixed Effects

Lin (2005) suggests an alternative method to eliminate the fixed effects by the transformation using $(I_{m_r} - W_{nr})$, the deviation from the weighted average of an individual's connections. Although this transformation is not very convenient when $\rho_0 \neq 0$ with an arbitrary M_{nr} such that $M_{nr} \neq W_{nr}$, it can be used as an alternative approach in the special case when $M_{nr} = W_{nr}$.

When $M_{nr} = W_{nr}$, as $(I_{m_r} - W_{nr})W_{nr} = W_{nr}(I_{m_r} - W_{nr})$ and $(I_{m_r} - W_{nr})R_{nr} = R_{nr}(I_{m_r} - W_{nr})$, premultiplication of (2) by $(I_{m_r} - W_{nr})$ gives

$$R_{nr}(I_{m_r} - W_{nr})Y_{nr} = \lambda_0 R_{nr} W_{nr}(I_{m_r} - W_{nr})Y_{nr} + R_{nr}(I_{m_r} - W_{nr})Z_{nr}\beta_0 + (I_{m_r} - W_{nr})\epsilon_{nr}. \quad (17)$$

The fixed effect α_{r0} is eliminated because $(I_{m_r} - W_{nr})l_{nr} = 0$ as $W_{nr} l_{nr} = l_{nr}$. The variance of the transformed disturbances $(I_{m_r} - W_{nr})\epsilon_{nr}$ is $\sigma^2 \ddot{\Sigma}_{nr}$ where $\ddot{\Sigma}_{nr} = (I_{m_r} - W_{nr})(I_{m_r} - W_{nr})'$. The elements of $(I_{m_r} - W_{nr})\epsilon_{nr}$ may be correlated and heteroskedastic. There are also linear dependence among its elements because $(I_{m_r} - W_{nr})$ does not have full row rank. Suppose that the rank of $(I_{m_r} - W_{nr})$ is $m^\star_r$, which, in principle, can be empirically evaluated as W_{nr} is a given matrix. As $(I_{m_r} - W_{nr})l_{m_r} = 0$, $m^\star_r \leq m_r - 1$, the transformation using $(I_{m_r} - W_{nr})$ to eliminate the fixed effects may leave the number of independent sample observations less than $\sum_{r=1}^{\bar{r}}(m_r - 1)$.

As $\ddot{\Sigma}_{nr}$ is positive semidefinite, there exists some orthonormal matrix $[\ddot{F}_{nr}, \ddot{H}_{nr}]$ where $\ddot{F}_{nr}$

is an $m_r \times m_r^\star$ matrix of normalized eigenvectors corresponding to the positive eigenvalues and $\ddot{H}_{nr}$ is an $m_r \times (m_r - m_r^\star)$ matrix of normalized eigenvectors with zero eigenvalues. Let Λ_{nr} be the $m_r^\star \times m_r^\star$ diagonal matrix consisting of all the positive eigenvalues. Thus, $\ddot{\Sigma}_{nr}\ddot{F}_{nr} = \ddot{F}_{nr}\Lambda_{nr}$, $\ddot{\Sigma}_{nr}\ddot{H}_{nr} = 0$, $\ddot{F}_{nr}'\ddot{F}_{nr} = I_{m_r^\star}$, $\ddot{F}_{nr}'\ddot{H}_{nr} = 0$, $\ddot{F}_{nr}\ddot{F}_{nr}' + \ddot{H}_{nr}\ddot{H}_{nr}' = I_{m_r}$, and $\ddot{\Sigma}_{nr} = \ddot{F}_{nr}\Lambda_{nr}\ddot{F}_{nr}'$. Denote $Y_{nr}^\star = \Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'(I_{m_r} - W_{nr})Y_{nr}$, $Z_{nr}^\star = \Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'(I_{m_r} - W_{nr})Z_{nr}$, and $\epsilon_{nr}^\star = \Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'(I_{m_r} - W_{nr})\epsilon_{nr}$. To eliminate heteroskedasticity and linear dependence in $(I_{m_r} - W_{nr})\epsilon_{nr}$, premultiplication of (17) by $\Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'$ yields

$$R_{nr}^\star Y_{nr}^\star = \lambda_0 R_{nr}^\star W_{nr}^\star Y_{nr}^\star + R_{nr}^\star Z_{nr}^\star \beta_0 + \epsilon_{nr}^\star. \tag{18}$$

where $W_{nr}^\star = \Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'W_{nr}\ddot{F}_{nr}\Lambda_{nr}^{\frac{1}{2}}$ and $R_{nr}^\star = \Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'R_{nr}\ddot{F}_{nr}\Lambda_{nr}^{\frac{1}{2}} = I_{m_r^\star} - \rho_0 W_{nr}^\star$. The variance matrix of the transformed disturbances $\epsilon_{nr}^\star$ is $\sigma^2 I_{m_r^\star}$.

Under the normality assumption, the log likelihood function of the sample with $\bar{r}$ macro groups is

$$\begin{aligned}\ln L_n(\theta) \quad = \quad & -\frac{n_r^\star}{2}\ln(2\pi\sigma^2) + \textstyle\sum_{r=1}^{\bar{r}} \ln|I_{m_r^\star} - \lambda W_{nr}^\star| + \sum_{r=1}^{\bar{r}} \ln|I_{m_r^\star} - \rho W_{nr}^\star| \\ & -\frac{1}{2\sigma^2}\textstyle\sum_{r=1}^{\bar{r}}[(I_{m_r^\star} - \lambda W_{nr}^\star)Y_{nr}^\star - Z_{nr}^\star\beta]' R_{nr}^{\star\prime} R_{nr}^\star [(I_{m_r^\star} - \lambda W_{nr}^\star)Y_{nr}^\star - Z_{nr}^\star\beta],\end{aligned}$$

where $n_r^\star = \sum_{r=1}^{\bar{r}} m_r^\star$. To implement the ML estimation, one needs to evaluate the determinants $|I_{m_r^\star} - \lambda W_{nr}^\star|$ and $|I_{m_r^\star} - \rho W_{nr}^\star|$ for each macro group r. The evaluation of this determinant is equivalent to the evaluation of the determinants $|I_{m_r} - \lambda W_{nr}|$ and $|I_{m_r} - \rho W_{nr}|$, which can be shown as follows. As

$$\begin{aligned}&[\ddot{F}_{nr}, \ddot{H}_{nr}]'(I_{m_r} - \lambda W_{nr})[\ddot{F}_{nr}, \ddot{H}_{nr}] = I_{m_r} - \lambda[\ddot{F}_{nr}, \ddot{H}_{nr}]'W_{nr}[\ddot{F}_{nr}, \ddot{H}_{nr}] \\ &= \begin{pmatrix} I_{m_r^\star} - \lambda\ddot{F}_{nr}'W_{nr}\ddot{F}_{nr} & -\lambda\ddot{F}_{nr}'W_{nr}\ddot{H}_{nr} \\ -\lambda\ddot{H}_{nr}'W_{nr}\ddot{F}_{nr} & I_{(m_r - m_r^\star)} - \lambda\ddot{H}_{nr}'W_{nr}\ddot{H}_{nr} \end{pmatrix} = \begin{pmatrix} I_{m_r^\star} - \lambda\ddot{F}_{nr}'W_{nr}\ddot{F}_{nr} & -\lambda\ddot{F}_{nr}'W_{nr}\ddot{H}_{nr} \\ 0 & (1-\lambda)I_{(m_r - m_r^\star)} \end{pmatrix},\end{aligned}$$

because $\ddot{H}_{nr}'W_{nr} = \ddot{H}_{nr}'$, $\ddot{H}_{nr}'\ddot{F}_{nr} = 0$ and $\ddot{H}_{nr}'\ddot{H}_{nr} = I_{(m_r - m_r^\star)}$. It follows that $|I_{m_r} - \lambda W_{nr}| = |I_{m_r} - \lambda[\ddot{F}_{nr}, \ddot{H}_{nr}]'W_{nr}[\ddot{F}_{nr}, \ddot{H}_{nr}]| = |I_{m_r^\star} - \lambda\ddot{F}_{nr}'W_{nr}\ddot{F}_{nr}| \cdot |(1-\lambda)I_{(m_r - m_r^\star)}|$. Therefore,

$$|I_{m_r^\star} - \lambda W_{nr}^\star| = |I_{m_r^\star} - \lambda\Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'W_{nr}\ddot{F}_{nr}\Lambda_{nr}^{\frac{1}{2}}| = |I_{m_r^\star} - \lambda\ddot{F}_{nr}'W_{nr}\ddot{F}_{nr}| = (1-\lambda)^{-(m_r - m_r^\star)}|I_{m_r} - \lambda W_{nr}|.$$

Similarly, $|I_{m_r^\star} - \rho W_{nr}^\star| = (1-\rho)^{-(m_r - m_r^\star)}|I_{m_r} - \rho W_{nr}|$. As $R_{nr}^\star[(I_{m_r^\star} - \lambda W_{nr}^\star)Y_{nr}^\star - Z_{nr}^\star\beta] =$

$\Lambda_{nr}^{-\frac{1}{2}}\ddot{F}_{nr}'(I_{m_r}-W_{nr})R_{nr}[(I_{m_r}-\lambda W_{nr})Y_{nr}-Z_{nr}\beta]$, the log likelihood can also be expressed in terms of Y_{nr}, Z_{nr} and W_{nr} as

$$\ln L_n(\theta) = -\frac{n_r^\star}{2}\ln(2\pi\sigma^2) - (n_r - n_r^\star)\ln[(1-\lambda)(1-\rho)] + \textstyle\sum_{r=1}^{\bar{r}}\ln|S_{nr}(\lambda)| + \sum_{r=1}^{\bar{r}}\ln|R_{nr}(\rho)|$$
$$-\frac{1}{2\sigma^2}\textstyle\sum_{r=1}^{\bar{r}}[S_{nr}(\lambda)Y_{nr}-Z_{nr}\beta]'R_{nr}'(I_{m_r}-W_{nr})'\ddot{\Sigma}_{nr}^{+}(I_{m_r}-W_{nr})R_{nr}[S_{nr}(\lambda)Y_{nr}-Z_{nr}\beta], \quad (19)$$

where $\ddot{\Sigma}_{nr}^{+} = \ddot{F}_{nr}\Lambda_{nr}^{-1}\ddot{F}_{nr}'$ is the generalized inverse of $(I_{m_r}-W_{nr})(I_{m_r}-W_{nr})'$. The MLE is derived from the maximization of (19).

G Proofs

Proof of Lemma 5.1. With the restriction $\lambda_0\beta_{10}+\beta_{20}=0$, $Z_{nr}^*\beta_0 = S_{nr}^*X_{nr}^*\beta_{10}$, and, hence, (9) becomes $Y_{nr}^* = X_{nr}^*\beta_{10}+v_{nr}$, where $v_{nr} = S_{nr}^{*-1}R_{nr}^{*-1}\epsilon_{nr}^*$. As λ_0 and ρ_0 are not in the mean regression equation, they could only be identified via the disturbances, $v_{nr} = \rho_0 M_{nr}^*v_{nr}+\lambda_0 W_{nr}^*v_{nr} - \rho_0\lambda_0 M_{nr}^*W_{nr}^*v_{nr}+\epsilon_{nr}^*$, when $M_{nr}\neq W_{nr}$. With λ_0 and β_{10} identified, β_{20} can be identified from the restriction $\lambda_0\beta_{10}+\beta_{20}=0$. However, when $M_{nr}=W_{nr}$, $v_{nr} = (\rho_0+\lambda_0)W_{nr}^*v_{nr} - \rho_0\lambda_0 W_{nr}^{*2}v_{nr}+\epsilon_{nr}^*$. In this case, ρ_0 and λ_0 may only be identified locally but not globally, and hence β_{20} can not be separately identified. ■

Proof of Lemma 5.2. For the group r, let c_1, c_2 and c_3 be conformable scalar and column vectors such that

$$J_{nr}R_{nr}(\rho)G_{nr}(X_{nr}\beta_{10}+W_{nr}X_{nr}\beta_{20})c_1 + J_{nr}R_{nr}(\rho)X_{nr}c_2 + J_{nr}R_{nr}(\rho)W_{nr}X_{nr}c_3 = 0, \quad (20)$$

where $G_{nr} = W_{nr}S_{nr}^{-1}$. We are interested in sufficient conditions so that $c_1=c_2=c_3=0$. Denote $\mu_{1r} = \frac{1}{m_r}l_{m_r}'R_{nr}(\rho)G_{nr}(X_{nr}\beta_{10}+W_{nr}X_{nr}\beta_{20})$, $\mu_{2r} = \frac{1}{m_r}l_{m_r}'R_{nr}(\rho)X_{nr}$ and $\mu_{3r} = \frac{1}{m_r}l_{m_r}'R_{nr}(\rho)W_{nr}X_{nr}$. As $W_{nr}S_{nr}^{-1} = S_{nr}^{-1}W_{nr}$,

$$\begin{aligned}
& [J_{nr}R_{nr}(\rho)G_{nr}(X_{nr}\beta_{10}+W_{nr}X_{nr}\beta_{20}), J_{nr}R_{nr}(\rho)X_{nr}, J_{nr}R_{nr}(\rho)W_{nr}X_{nr}] \\
= \; & R_{nr}(\rho)S_{nr}^{-1}\{[W_{nr}(X_{nr}\beta_{10}+W_{nr}X_{nr}\beta_{20}), S_{nr}X_{nr}, S_{nr}W_{nr}X_{nr}] - S_{nr}R_{nr}^{-1}(\rho)l_{m_r}(\mu_{1r},\mu_{2r},\mu_{3r})\} \\
= \; & R_{nr}(\rho)S_{nr}^{-1}\{[W_{nr}(X_{nr}\beta_{10}+W_{nr}X_{nr}\beta_{20}), S_{nr}X_{nr}, S_{nr}W_{nr}X_{nr}] - l_{m_r}(\mu_{1r}^*,\mu_{2r}^*,\mu_{3r}^*)\},
\end{aligned}$$

where $\mu_l^* = (\frac{1-\lambda_0}{1-\rho})\mu_l$, for $l=1,2,3$, because $S_{nr}R_{nr}^{-1}(\rho)l_{m_r} = (\frac{1-\lambda_0}{1-\rho})l_{m_r}$. As $R_{nr}(\rho)$ and S_{nr} are nonsingular, (20) is equivalent to

$$
\begin{aligned}
& W_{nr}(X_{nr}\beta_{10} + W_{nr}X_{nr}\beta_{20})c_1 + S_{nr}X_{nr}c_2 + S_{nr}W_{nr}X_{nr}c_3 - l_{m_r}(\mu_{1r}^*c_1 + \mu_{2r}^*c_2 + \mu_{3r}^*c_3) \\
= \;\; & X_{nr}c_2 + W_{nr}X_{nr}(c_1\beta_{10} - \lambda_0 c_2 + c_3) + W_{nr}^2 X_{nr}(c_1\beta_{20} - \lambda_0 c_3) - l_{m_r}(\mu_{1r}^*c_1 + \mu_{2r}^*c_2 + \mu_{3r}^*c_3) = 0.
\end{aligned}
$$

As $[X_{nr}, W_{nr}X_{nr}, W_{nr}^2 X_{nr}, l_{m_r}]$ has the full column rank, it follows that $c_2 = 0$, $c_3 + c_1\beta_{10} = 0$ and $\beta_{20}c_1 - \lambda_0 c_3 = 0$. These imply, in turn, that $c_1 = 0$ and $c_3 = -c_1\beta_{10} = 0$ under the assumption $\beta_{20} + \lambda_0\beta_{10} \neq 0$. The desired result follows. ■

Proof of Lemma 5.3. By Lemma C.1, $F_{nr}'R_{nr}(\rho)[G_{nr}(X_{nr}\beta_{10} + W_{nr}X_{nr}\beta_{20}), X_{nr}, W_{nr}X_{nr}] = R_{nr}^*(\rho)S_{nr}^{*-1}[W_{nr}^*(X_{nr}^*\beta_{10} + W_{nr}^*X_{nr}^*\beta_{20}), S_{nr}^*X_{nr}^*, S_{nr}^*W_{nr}^*X_{nr}^*]$. As $R_{nr}^*(\rho)$ and S_{nr}^* are nonsingular, a sufficient identification condition derived from a similar argument in the proof of Lemma 5.2 (but without l_{m_r} term) is that the stacked matrix with its rth row block being $[X_{nr}^*, W_{nr}^*X_{nr}^*, W_{nr}^{*2}X_{nr}^*]$ has full column rank as long as $\beta_{20}+\lambda_0\beta_{10} \neq 0$. By a pre-multiplication with F_{nr}, a sufficient condition is that the stacked matrix with its rth row block being $[J_{nr}X_{nr}, (J_{nr}W_{nr})(J_{nr}X_{nr}), (J_{nr}W_{nr})^2(J_{nr}X_{nr})] = J_{nr}[X_{nr}, W_{nr}X_{nr}, W_{nr}^2 X_{nr}]$ has full column rank. ■

Proof of Proposition 1. We shall prove that $\frac{1}{n^*}[\ln L_n(\gamma) - Q_n(\gamma)]$ converges in probability to zero uniformly on Γ, and the identification uniqueness condition holds, i.e., for any $\varepsilon > 0$, $\limsup_{n\to\infty} \max_{\gamma \in \bar{N}_\varepsilon(\gamma_0)} \frac{1}{n^*}[Q_n(\gamma) - Q_n(\gamma_0)] < 0$ where $\bar{N}_\varepsilon(\gamma_0)$ is the complement of an open neighborhood of γ_0 in Γ with radius ε. The following arguments extend those in Lee (2004) for the SAR model with i.i.d. disturbances to our transformed equation model.

For the proof of these properties, it is useful to establish some properties for $\ln|S_n(\lambda)|$, $\ln|R_n(\rho)|$, and $\sigma_n^2(\gamma) = \frac{\sigma_0^2}{n^*}\mathrm{tr}([R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}]'J_n[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}])$, where $J_n[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}] = J_n[I_n + (\rho_0 - \rho)H_n + (\lambda_0 - \lambda)R_nG_nR_n^{-1} + (\rho_0 - \rho)(\lambda_0 - \lambda)H_nR_nG_nR_n^{-1}]$.

There is also an auxiliary model which has useful implications. Denote $Q_{p,n}(\gamma) = -\frac{n^*}{2}(\ln(2\pi) + 1) - \frac{n^*}{2}\ln\sigma_n^2(\gamma) + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - \bar{r}\ln[(1-\lambda)(1-\rho)]$. The log likelihood function of a transformed SAR process $R_{nr}^*Y_{nr}^* = \lambda_0 R_{nr}^*W_{nr}^*Y_{nr}^* + \epsilon_{nr}^*$, where $\epsilon_{nr}^* \sim N(0, \sigma_0^2 I_{m_r^*})$ for $r = 1, \cdots, \bar{r}$, is $\ln L_{p,n}(\gamma, \sigma^2) = -\frac{n^*}{2}\ln(2\pi) - \frac{n^*}{2}\ln\sigma^2 + \ln|S_n(\lambda)| + \ln|R_n(\rho)| - \bar{r}\ln[(1-\lambda)(1-\rho)] - \frac{1}{2\sigma^2}Y_n'S_n'(\lambda)R_n'(\rho) \times J_nR_n(\rho)S_n(\lambda)Y_n$. It is apparent that $Q_{p,n}(\gamma) = \max_{\sigma^2}\mathrm{E}_p[\ln L_{p,n}(\gamma, \sigma^2)]$, where E_p is the expectation under this SAR process. By the Jensen inequality, $Q_{p,n}(\gamma) \leq \mathrm{E}_p[\ln L_{p,n}(\gamma_0, \sigma_0^2)] = Q_{p,n}(\gamma_0)$ for all γ. This implies that $\frac{1}{n^*}[Q_{p,n}(\gamma) - Q_{p,n}(\gamma_0)] \leq 0$ for all γ.

Let (λ_1, ρ_1) and (λ_2, ρ_2) be in Γ. By the mean value theorem, $\frac{1}{n^*}(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) =$

$\frac{1}{n^*}\mathrm{tr}(G_n(\bar{\lambda}_n))(\lambda_2-\lambda_1)$ where $\bar{\lambda}_n$ lies between λ_1 and λ_2. By the uniform boundedness of Assumption 4, Lemma C.6 implies that $\frac{1}{n^*}\mathrm{tr}(G_n(\bar{\lambda}_n)) = O(1)$. Thus, $\frac{1}{n^*}\ln|S_n(\lambda)|$ is uniformly equicontinuous in λ in Γ. As Γ is a bounded set, $\frac{1}{n^*}(\ln|S_n(\lambda_2)| - \ln|S_n(\lambda_1)|) = O(1)$ uniformly in λ_1 and λ_2 in Γ. Similarly, $\frac{1}{n^*}\ln|R_n(\rho)|$ is uniformly equicontinuous in ρ in Γ, and $\frac{1}{n^*}(\ln|R_n(\rho_2)|-\ln|R_n(\rho_1)|) = O(1)$ uniformly in ρ_1 and ρ_2 in Γ.

The $\sigma_n^2(\gamma)$ is uniformly bounded away from zero on Γ. This can be established by a counter argument. Suppose that $\sigma_n^2(\gamma)$ were not uniformly bounded away from zero on Γ. Then, there would exist a sequence $\{\gamma_n\}$ in Γ such that $\lim_{n\to\infty}\sigma_n^2(\gamma_n) = 0$. We have shown that $\frac{1}{n^*}[Q_{p,n}(\gamma) - Q_{p,n}(\gamma_0)] \leq 0$ for all γ, which implies that $-\frac{1}{2}\ln\sigma_n^2(\gamma) \leq -\frac{1}{2}\ln\sigma_0^2 + \frac{1}{n^*}(\ln|S_n| - \ln|S_n(\lambda)|) + \frac{1}{n^*}(\ln|R_n| - \ln|R_n(\rho)|) - \frac{\bar{r}}{n^*}(\ln[(1-\lambda_0)(1-\rho_0)] - \ln[(1-\lambda)(1-\rho)]) = O(1)$, because $\frac{1}{n^*}(\ln|S_n| - \ln|S_n(\lambda)|) = O(1)$ and $\frac{1}{n^*}(\ln|R_n| - \ln|R_n(\rho)|) = O(1)$ uniformly on Γ. That is, $-\ln\sigma_n^2(\gamma_n)$ is bounded from above, a contradiction. Therefore, $\sigma_n^2(\gamma)$ must be bounded always from zero uniformly on Γ.

(uniform convergence) We will show that $\sup_{\gamma\in\Gamma}|\frac{1}{n^*}\ln L_n(\gamma) - \frac{1}{n^*}Q_n(\gamma)| = \sup_{\gamma\in\Gamma}\frac{1}{2}|\ln\hat{\sigma}_n^2(\gamma) - \ln\sigma_n^{*2}(\gamma)| = o_p(1)$. As $P_n(\rho)R_n(\rho)S_n(\lambda)Y_n = (\lambda_0-\lambda)P_n(\rho)R_n(\rho)G_nZ_n\beta_0 + P_n(\rho)R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n$, $\hat{\sigma}_n^2(\gamma) = \frac{1}{n^*}Y_n'S_n'(\lambda)R_n'(\rho)P_n(\rho)R_n(\rho)S_n(\lambda)Y_n = \frac{(\lambda_0-\lambda)^2}{n^*}[R_n(\rho)G_nZ_n\beta_0]'P_n(\rho)[R_n(\rho)G_nZ_n\beta_0] + 2(\lambda_0-\lambda)K_{1n}(\gamma) + K_{2n}(\gamma)$, where $K_{1n}(\gamma) = \frac{1}{n^*}[R_n(\rho)G_nZ_n\beta_0]'P_n(\rho)[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n]$ and $K_{2n}(\gamma) = \frac{1}{n^*}[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n]'P_n(\rho)[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n]$. $\hat{\sigma}_n^2(\gamma) - \sigma_n^{*2}(\gamma) = 2(\lambda_0-\lambda)K_{1n}(\gamma) + K_{2n}(\gamma) - \sigma_n^2(\gamma)$, since $\sigma_n^{*2}(\gamma) = \frac{(\lambda_0-\lambda)^2}{n^*}[R_n(\rho)G_nZ_n\beta_0]'P_n(\rho)[R_n(\rho)G_nZ_n\beta_0] + \sigma_n^2(\gamma)$. Lemma C.7 implies $K_{1n}(\gamma) = o_p(1)$. The convergence is uniform on Γ as λ and ρ appears simply as polynomial factors. On the other hand, $K_{2n}(\gamma) - \sigma_n^2(\gamma) = \frac{1}{n^*}[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n]'J_n[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n] - \frac{\sigma_0^2}{n^*}\mathrm{tr}([R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}]'J_n[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}]) - T_n(\gamma)$, where

$$T_n(\gamma) = \frac{1}{n^*}[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n]'J_nR_n(\rho)Z_n[Z_n'R_n'(\rho)J_nR_n(\rho)Z_n]^{-1}Z_n'R_n'(\rho)J_n[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n].$$

As $\frac{1}{\sqrt{n^*}}Z_n'R_n'(\rho)J_nR_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n = O_p(1)$ uniformly on Γ by Lemma C.7, it follows that

$$\begin{aligned}T_n(\gamma) =& \frac{1}{n^*}[\frac{1}{\sqrt{n^*}}Z_n'R_n'(\rho)J_nR_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n]'[\frac{1}{n^*}Z_n'R_n'(\rho)J_nR_n(\rho)Z_n]^{-1}\\ &\times[\frac{1}{\sqrt{n^*}}Z_n'R_n'(\rho)J_nR_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n] = o_p(1).\end{aligned}$$

By Lemma C.9, we have $\frac{1}{n^*}\Big\{[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n]'J_n[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}\epsilon_n] - \sigma_0^2\mathrm{tr}([R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}]'$

$\times J_n[R_n(\rho)S_n(\lambda)S_n^{-1}R_n^{-1}])\Big\} = o_p(1)$. These convergences are uniform on Γ because λ and ρ appears simply as polynomial factors in those terms. That is, $K_{2n}(\gamma) - \sigma_n^2(\gamma) = o_p(1)$ uniformly on Γ. Therefore, $\hat{\sigma}_n^2(\gamma) - \sigma_n^{*2}(\gamma) = o_p(1)$ uniformly on Γ. By the Taylor expansion, $|\ln \hat{\sigma}_n^2(\gamma) - \ln \sigma_n^{*2}(\gamma)| = |\hat{\sigma}_n^2(\gamma) - \sigma_n^{*2}(\gamma)|/\tilde{\sigma}_n^2(\gamma)$, where $\tilde{\sigma}_n^2(\gamma)$ lies between $\hat{\sigma}_n^2(\gamma)$ and $\sigma_n^{*2}(\gamma)$. As $\sigma_n^{*2}(\gamma) \geq \sigma_n^2(\gamma)$ and $\sigma_n^2(\gamma)$ is uniformly bounded away from zero on Γ, $\sigma_n^{*2}(\gamma)$ will be so too. It follows that, because $\hat{\sigma}_n^2(\gamma) - \sigma_n^{*2}(\gamma) = o_p(1)$ uniformly on Γ, $\hat{\sigma}_n^2(\gamma)$ will be bounded away from zero uniformly on Γ in probability. Hence, $|\ln \hat{\sigma}_n^2(\gamma) - \ln \sigma_n^{*2}(\gamma)| = o_p(1)$ uniformly on Γ. Consequently, $\sup_{\gamma\in\Gamma} |\frac{1}{n^*}\ln L_n(\gamma) - \frac{1}{n^*}Q_n(\gamma)| = o_p(1)$.

(uniform equicontinuity) We will show that $\frac{1}{n^*}\ln Q_n(\gamma) = -\frac{1}{2}(\ln(2\pi)+1) - \frac{1}{2}\ln \sigma_n^{*2}(\gamma) + \frac{1}{n^*}(\ln|S_n(\lambda)| + \ln|R_n(\rho)|) - \frac{\bar{r}}{n^*}\ln[(1-\lambda)(1-\rho)]$ is uniformly equicontinuous on Γ. The $\sigma_n^{*2}(\gamma)$ is uniformly continuous on Γ. This is so, because $\sigma_n^{*2}(\gamma)$ is a polynomial of λ and ρ, with bounded coefficients by Lemmas C.5 and C.6. The uniform continuity of $\ln \sigma_n^{*2}(\gamma)$ on Γ follows because $\frac{1}{\sigma_n^{*2}(\gamma)}$ is uniformly bounded on Γ. Hence $\frac{1}{n^*}\ln Q_n(\gamma)$ is uniformly equicontinuous on Γ.

(identification uniqueness) At γ_0, $\sigma_n^{*2}(\gamma_0) = \sigma_0^2$. Therefore, $\frac{1}{n^*}Q_n(\gamma) - \frac{1}{n^*}Q_n(\gamma_0) = -\frac{1}{2}[\ln \sigma_n^2(\gamma) - \ln \sigma_0^2] + \frac{1}{n^*}(\ln|S_n(\lambda)| - \ln|S_n|) + \frac{1}{n^*}(\ln|R_n(\rho)| - \ln|R_n|) - \frac{\bar{r}}{n^*}(\ln[(1-\lambda)(1-\rho)] - \ln[(1-\lambda_0)(1-\rho_0)]) - \frac{1}{2}[\ln \sigma_n^{*2}(\gamma) - \ln \sigma_n^2(\gamma)] = \frac{1}{n^*}(Q_{p,n}(\gamma) - Q_{p,n}(\gamma_0)) - \frac{1}{2}[\ln \sigma_n^{*2}(\gamma) - \ln \sigma_n^2(\gamma)]$. Suppose that the identification uniqueness condition would not hold. Then, there would exist an $\varepsilon > 0$ and a sequence $\{\gamma_n\}$ in $\bar{N}_\varepsilon(\gamma_0)$ such that $\lim_{n\to\infty}[\frac{1}{n^*}Q_n(\gamma_n) - \frac{1}{n^*}Q_n(\gamma_0)] = 0$. Because $\bar{N}_\varepsilon(\lambda_0)$ is a compact set, there would exist a convergent subsequence $\{\gamma_{n_m}\}$ of $\{\gamma_n\}$. Let γ_+ be the limit point of $\{\gamma_{n_m}\}$ in Γ. As $\frac{1}{n^*}Q_n(\gamma)$ is uniformly equicontinuous in γ, $\lim_{n_m\to\infty}\frac{1}{n_m}[Q_{n_m}(\gamma_+) - Q_{n_m}(\gamma_0)] = 0$. Because $(Q_{p,n}(\gamma) - Q_{p,n}(\gamma_0)) \leq 0$ and $-[\ln \sigma_n^{*2}(\gamma) - \ln \sigma_n^2(\gamma)] \leq 0$, this is possible only if $\lim_{n_m\to\infty}(\sigma_{n_m}^{*2}(\gamma_+) - \sigma_{n_m}^2(\gamma_+)) = 0$ and $\lim_{n_m\to\infty}\frac{1}{n_m}[Q_{p,n_m}(\gamma_+) - Q_{p,n_m}(\gamma_0)] = 0$. The $\lim_{n_m\to\infty}(\sigma_{n_m}^{*2}(\gamma_+) - \sigma_{n_m}^2(\gamma_+)) = 0$ is a contradiction when $\lim_{n\to\infty}\frac{1}{n^*}[R_n(\rho)G_nZ_n\beta_0]'P_n(\rho)[R_n(\rho)G_nZ_n\beta_0] \neq 0, \forall\rho$. In the event that $\lim_{n\to\infty}\frac{1}{n^*}[R_n(\rho)G_nZ_n\beta_0]'P_n(\rho)[R_n(\rho)G_nZ_n\beta_0] = 0$ for some ρ, the contradiction follows from the relation $\lim_{n\to\infty}\frac{1}{n_m}[Q_{p,n_m}(\gamma_+) - \frac{1}{n_m}Q_{p,n_m}(\gamma_0)] = 0$ under Assumption 5(b). This is so, because, in this event, Assumption 5(b) is equivalent to that $\lim_{n\to\infty}[\frac{1}{n^*}(\ln|S_n(\lambda)| - \ln|S_n|) + \frac{1}{n^*}(\ln|R_n(\rho)| - \ln|R_n|) - \frac{\bar{r}}{n^*}(\ln[(1-\lambda)(1-\rho)] - \ln[(1-\lambda_0)(1-\rho_0)]) - \frac{1}{2}(\ln \sigma_n^2(\gamma) - \ln \sigma_0^2)] = \lim_{n\to\infty}\frac{1}{n^*}[Q_{p,n}(\gamma) - Q_{p,n}(\gamma_0)] \neq 0$ for $\gamma \neq \gamma_0$. Therefore, the identification uniqueness condition must hold.

The consistency of $\hat{\gamma}_n$ and, hence, $\hat{\theta}_n$ follow from this identification uniqueness and uniform convergence (White, 1994, Theorem 3.4). ■

Proof of Proposition 2. (Show that $\frac{1}{n^*}\frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial\theta\partial\theta'} - \frac{1}{n^*}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} \xrightarrow{p} 0$.) The second-order derivatives are given in Appendix B. By the mean value theorem, $\mathrm{tr}(J_nG_n^2(\tilde{\lambda}_n)) = \mathrm{tr}(J_nG_n^2) + 2\mathrm{tr}(J_nG_n^3(\bar{\lambda}_n))(\tilde{\lambda}_n - \lambda_0)$. Note that $G_n(\bar{\lambda}_n)$ is uniformly bounded in row and column sums uniformly in a neighborhood of λ_0 by Lemma C.3 under Assumption 4. As $R_n(\tilde{\rho}_n) = R_n + (\rho_0 - \tilde{\rho}_n)M_n$, it follows that $\frac{1}{n^*}[\frac{\partial^2}{\partial\lambda^2}\ln L_n(\tilde{\theta}_n) - \frac{\partial^2}{\partial\lambda^2}\ln L_n(\theta_0)] = -2\frac{1}{n^*}\mathrm{tr}[J_nG_n^3(\bar{\lambda}_n)](\tilde{\lambda}_n - \lambda_0) - \frac{1}{n^*}(\frac{1}{\tilde{\sigma}_n^2} - \frac{1}{\sigma_0^2})Y_n'W_n'R_n'J_nR_nW_nY_n + \frac{2(\tilde{\rho}_n-\rho_0)}{n^*\tilde{\sigma}_n^2}Y_n'W_n'R_n'J_nM_nW_nY_n - \frac{(\tilde{\rho}_n-\rho_0)^2}{n^*\tilde{\sigma}_n^2}Y_n'W_n'M_n'J_nM_nW_nY_n = o_p(1)$, because $\frac{1}{n^*}\mathrm{tr}(J_nG_n^3(\bar{\lambda}_n)) = O(1)$, $\frac{1}{n^*}Y_n'W_n'R_n'J_nR_nW_nY_n = O_p(1)$, $\frac{1}{n^*}Y_n'W_n'R_n'J_nM_nW_nY_n = O_p(1)$, and $\frac{1}{n^*}Y_n'W_n'M_n'J_nM_nW_nY_n = O_p(1)$. The convergence in probability of the other second order derivatives follows similar or more straightforward arguments.

(Show $\frac{1}{n^*}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E(\frac{1}{n^*}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}) \xrightarrow{p} 0$.) As $\frac{1}{n^*}(\tilde{G}_n\tilde{Z}_n\beta_0)'J_n\tilde{G}_n\epsilon_n = o_p(1)$ by Lemma C.7, it follows that $\frac{1}{n^*}Y_n'W_n'R_n'J_nR_nW_nY_n = \frac{1}{n^*}(\tilde{G}_n\tilde{Z}_n\beta_0)'J_n\tilde{G}_n\tilde{Z}_n\beta_0 + \frac{1}{n^*}\epsilon_n'\tilde{G}_n'J_n\tilde{G}_n\epsilon_n + o_p(1)$. Lemmas C.8 and C.6 imply $E(\epsilon_n'\tilde{G}_n'J_n\tilde{G}_n\epsilon_n) = \sigma_0^2\mathrm{tr}(\tilde{G}_n'J_n\tilde{G}_n)$ and

$$\mathrm{Var}(\frac{1}{n}\epsilon_n'\tilde{G}_n'J_n\tilde{G}_n\epsilon_n) = \frac{(\mu_4 - 3\sigma_0^4)}{n^2}\sum_{i=1}^n vec_D'(\tilde{G}_n'J_n\tilde{G}_n)vec_D(\tilde{G}_n'J_n\tilde{G}_n) + \frac{2\sigma_0^4}{n^2}\mathrm{tr}[J_n(\tilde{G}_n'\tilde{G}_n)^2] = O(\frac{1}{n}).$$

Hence $\frac{1}{n^*}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda^2} - E(\frac{1}{n^*}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\lambda^2}) \xrightarrow{p} 0$ follows from the law of large numbers. The convergence of the other terms can be derived by similar arguments.

(Show that Σ_θ is nonsingular.) Let $\alpha = (\alpha_1', \alpha_2, \alpha_3, \alpha_4)'$ be a column vector of constants such that $\Sigma_\theta\alpha = 0$. It is sufficient to show that $\alpha = 0$. From the first row block of the linear equation system $\Sigma_\theta\alpha = 0$, one has $\alpha_1 = -\lim_{n\to\infty}(\tilde{Z}_n'J_n\tilde{Z}_n)^{-1}\tilde{Z}_n'J_n\tilde{G}_n\tilde{Z}_n\beta_0\cdot\alpha_2$. From the last equation of the linear system, one has $\alpha_4 = -\lim_{n\to\infty}\frac{2\sigma_0^2}{n^*}\mathrm{tr}(J_n\tilde{G}_n)\cdot\alpha_2 - \lim_{n\to\infty}\frac{2\sigma_0^2}{n^*}\mathrm{tr}(J_nH_n)\cdot\alpha_3$. Substitution in the the third equation of the linear system gives $\lim_{n\to\infty}\frac{1}{n^*}[\mathrm{tr}(H_n^sJ_n\tilde{G}_n) - \frac{2}{n^*}\mathrm{tr}(J_nH_n)\mathrm{tr}(J_n\tilde{G}_n)]\alpha_2 + \lim_{n\to\infty}\frac{1}{n^*}[\mathrm{tr}(H_n^sJ_nH_n) - \frac{2}{n^*}\mathrm{tr}^2(J_nH_n)]\alpha_3 = 0$. By eliminating α_1, α_3 and α_4, the remaining equation becomes $\lim_{n\to\infty}\frac{1}{(n^*)^2}\Big\{\frac{1}{\sigma_0^2}[\mathrm{tr}(H_n^sJ_nH_n) - \frac{2}{n^*}\mathrm{tr}^2(J_nH_n)](\tilde{G}_n\tilde{Z}_n\beta_0)'P_n(\tilde{G}_n\tilde{Z}_n\beta_0) + \phi_n\Big\}\cdot\alpha_2 = 0$, where $\phi_n = [\mathrm{tr}(H_n^sJ_nH_n) - \frac{2}{n^*}\mathrm{tr}^2(J_nH_n)][\mathrm{tr}(\tilde{G}_n^sJ_n\tilde{G}_n) - \frac{2}{n^*}\mathrm{tr}^2(J_n\tilde{G}_n)] - [\mathrm{tr}(H_n^sJ_n\tilde{G}_n) - \frac{2}{n^*}\mathrm{tr}(J_nH_n)\mathrm{tr}(J_n\tilde{G}_n)]^2$. Let $C_n = J_n\tilde{G}_n - \frac{1}{n}\mathrm{tr}(J_n\tilde{G}_n)I_n$ and $D_n = J_nH_n - \frac{1}{n}\mathrm{tr}(J_nH_n)I_n$. Then, $\lim_{n\to\infty}\frac{1}{n^*}[\mathrm{tr}(H_n^sJ_nH_n) - \frac{2}{n^*}\mathrm{tr}^2(J_nH_n)] = \lim_{n\to\infty}\frac{1}{2n^*}\mathrm{tr}(D_n^sD_n^s) \geq 0$ and

$$\phi_n = \frac{1}{4}[\mathrm{tr}(D_n^sD_n^s)\mathrm{tr}(C_n^sC_n^s) - \mathrm{tr}^2(C_n^sD_n^s)] \geq 0.$$

As Assumption 5(a) implies that $\lim_{n\to\infty}\frac{1}{n^*}(R_nG_nZ_n\beta_0)'P_n(R_nG_nZ_n\beta_0)$ is positive definite and $\lim_{n\to\infty}\frac{1}{n^*}[\mathrm{tr}(H_n^sJ_nH_n) - \frac{2}{n^*}\mathrm{tr}^2(J_nH_n)] > 0$, it follows that $\alpha_2 = 0$ and, so, $\alpha = 0$. On the other hand, if $\lim_{n\to\infty} \frac{1}{n^*}(R_n G_n Z_n \beta_0)' P_n (R_n G_n Z_n \beta_0) = 0$, $\lim_{n\to\infty}(\frac{1}{n^*})^2 \phi_n > 0$ by Assumption 6. it follows that $\alpha_2 = 0$ and, so, $\alpha = 0$. Hence Σ_θ is nonsingular.

(the limiting distribution of $\frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta}$) The matrices $J_n R_n$, $J_n H_n$ and $J_n \tilde{G}_n$ are uniformly bounded in both row and column sums in absolute value. As the elements of Z_n are bounded, the elements of $J_n \tilde{Z}_n$ and $J_n \tilde{G}_n \tilde{Z}_n \beta_0$ for all n are uniformly bounded by Lemma C.5. With the existence of high order moments of ϵ in Assumption 1, the central limit theorem for quadratic forms of double arrays of Kelejian and Prucha (2001) can be applied and the limiting distribution of the score vector follows.

Finally, from the expansion $\sqrt{n^*}(\hat{\theta}_n - \theta_0) = -\left(\frac{1}{n^*} \frac{\partial^2 \ln L_n(\tilde{\theta}_n)}{\partial \theta \partial \theta'}\right)^{-1} \frac{1}{\sqrt{n^*}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta}$, the asymptotic distribution of $\hat{\theta}_n$ follows. ■

Proof of Proposition 3. Let $B_n = \begin{pmatrix} B_{n,11} & B_{n,12} \\ B_{n,21} & B_{n,22} \end{pmatrix}$, where $B_{n,11} = \frac{1}{n^*}\mathrm{tr}(\tilde{G}_n^s J_n \tilde{G}_n)$, $B_{n,21} = B'_{n,12} = (\frac{1}{n^*}\mathrm{tr}(H_n^s J_n \tilde{G}_n), \frac{1}{\sigma_0^2 n^*}\mathrm{tr}(J_n \tilde{G}_n))'$, and

$$B_{n,22} = \begin{pmatrix} \frac{1}{n^*}\mathrm{tr}(H_n^s J_n H_n) & * \\ \frac{1}{\sigma_0^2 n^*}\mathrm{tr}(J_n H_n) & \frac{1}{2\sigma_0^4} \end{pmatrix}.$$

Under normality assumption, the variance matrix of the MLE of θ_0 is

$$\frac{1}{n^*}\Sigma_{\theta,n}^{-1} = \frac{1}{n^*}\left(\begin{pmatrix} \Sigma_{\zeta,n} & 0_{(k+1)\times 2} \\ 0_{2\times(k+1)} & 0_{2\times 2} \end{pmatrix} + \begin{pmatrix} 0_{k\times k} & 0_{k\times 3} \\ 0_{3\times k} & B_n \end{pmatrix}\right)^{-1}.$$

The variance matrix of the MLE of ζ_0 is

$$\frac{1}{n^*}\left(\Sigma_{\zeta,n} + \begin{pmatrix} 0_{k\times k} & 0_{k\times 1} \\ 0_{1\times k} & B_{n,11} - B_{n,12}B_{n,22}^{-1}B_{n,21} \end{pmatrix}\right)^{-1},$$

by the inversion of the partitioned matrix. As B_n is nonnegative definite, the variance matrix of the MLE is relatively smaller than that of $\hat{\zeta}_{b2sls,n}$. ■

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht.

Anselin, L. (2006). Spatial econometrics, *in* T. C. Mills and K. Patterson (eds), *Palgrave Handbook of Econometrics*, Vol. 1, Palgrave MacMillan, New York.

Bertrand, M., Luttmer, E. and Mullainathan, S. (2000). Network effects and welfare cultures, *Quarterly Journal of Economics* **115**: 1019–1055.

Bramoullé, Y., Djebbari, H. and Fortin, B. (2009). Identification of peer effects though social networks, *Journal of Econometrics* **150**: 41–55.

Calvó-Armengol, A., Patacchini, E. and Zenou, Y. (2006). Peer effects and social networks in education. Manuscript, Universitat Autonoma de Barcelona.

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

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

Case, A., J. R. Hines, J. and Rosen, H. S. (1993). Interstate tax competition after tra 86, *Journal of Policy Analysis and Management* **12**: 136–148.

Chamberlain, G. (1980). Analysis of covariance with qualitative data, *Review of Economic Studies* **47**: 225–238.

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.

Conly, T. G. (1999). Gmm estimation with cross sectional dependence, *Journal of Econometrics* **92**: 1–45.

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

Cressie, N. (1993). *Statistics for Spatial Data*, John Wiley and Sons, New York.

Durlauf, S. N. and Young, H. P. (2001). The new social economics, *in* S. N. Durlauf and H. P. Young (eds), *Social Dynamics*, MIT press, Cambridge.

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.

Florax, R. and Folmer, H. (1992). Specification and estimation of spatial linear regression models: Monte carlo evaluation of pre-test estimators, *Regional science and urban economics* **22**: 405–432.

Hanushek, E. A., Kain, J. F., Markman, J. M. and Rivkin, S. G. (2003). Does peer ability affect student achievement?, *Journal of Applied Econometrics* **18**: 527–544.

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

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

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). Hac estimation in a spatial framework, *Journal of Econometrics* **140**: 131–154.

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

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

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

Lee, L. F. and Liu, X. (2008). Efficient gmm estimation of high order spatial autoregressive models with autoregressive disturbances. Forthcoming in Econometric Theory.

Lee, L. F. and Yu, J. (2007). A spatial dynamic panel data model with both time and individual fixed effects. working paper, Department of Economics, Ohio State University.

Lee, S. Y. (2008). Three essays on spatial autoregressive models and empirical organization. PhD thesis, Department of Economics, Ohio State University.

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

LeSage, J. and Pace, R. (2009). *Introduction to Spatial Econometrics*, CRC Press, Boca Raton, FL.

Lin, X. (2005). Peer effects and student academic achievement: an application of spatial autoregressive model with group unobservables. Manuscript, Ohio State University.

Lin, X. (2008). Identifying peer effects in student academic achievement by a spatial autoregressive model with group unobservables. Manuscript, Tsinghua Unversity, Beijing.

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. and Lee, L. F. (2009). Gmm estimation of social interaction models with centrality. Manuscript, 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.

Ord, J. (1975). Estimation methods for models of spatial interaction, *Journal of the American Statistical Association* **70**: 120–126.

Rothenberg, T. J. (1971). Indentification in parametric models, *Econometrica* **39**: 577–591.

Sacerdote, B. (2001). Peer effects with random assignment: results for dartmouth roommates, *Quarterly Journal of Economics* **116**: 681–704.

Udry, J. R. (2003). The national longitudinal study of adolescent health (add health), waves 1.

White, H. (1994). *Estimation, Inference and Specification Analysis*, Cambridge Univsersity Press, New York.

Table 1: 2SLS, G2SLS and ML Estimation with Strong X's

	$\lambda_0 = 0.5$	$\rho_0 = 0.5$	$\beta_{10} = 1$	$\beta_{20} = 1$
	small size groups: $\bar{r} = 67$, $n = 557$, $\sum_{r=1}^{\bar{r}} \text{rank}(I_{m_r} - W_{nr}) = 466$			
2SLS	0.489(.085)[.086]	–	1.004(.056)[.056]	1.003(.092)[.092]
G2SLS	0.498(.069)[.069]	0.346(.107)[.187]	1.003(.053)[.054]	1.001(.083)[.083]
ML1	0.495(.069)[.069]	0.495(.081)[.081]	1.004(.053)[.053]	1.004(.082)[.082]
ML2	0.495(.089)[.089]	0.490(.110)[.111]	1.005(.053)[.053]	1.005(.083)[.083]
	moderate size groups: $\bar{r} = 67$, $n = 877$, $\sum_{r=1}^{\bar{r}} \text{rank}(I_{m_r} - W_{nr}) = 761$			
2SLS	0.494(.059)[.059]	–	1.004(.041)[.041]	1.008(.078)[.078]
G2SLS	0.499(.048)[.048]	0.424(.072)[.105]	1.003(.037)[.037]	1.004(.064)[.064]
ML1	0.496(.048)[.048]	0.497(.058)[.058]	1.003(.037)[.037]	1.006(.063)[.063]
ML2	0.497(.060)[.060]	0.498(.073)[.073]	1.003(.037)[.037]	1.006(.065)[.065]
	moderate size groups: $\bar{r} = 102$, $n = 1344$, $\sum_{r=1}^{\bar{r}} \text{rank}(I_{m_r} - W_{nr}) = 1166$			
2SLS	0.495(.047)[.047]	–	1.003(.034)[.035]	1.003(.059)[.059]
G2SLS	0.498(.038)[.039]	0.428(.058)[.092]	1.002(.032)[.032]	1.001(.049)[.049]
ML1	0.496(.038)[.039]	0.501(.046)[.046]	1.002(.032)[.032]	1.003(.048)[.049]
ML2	0.496(.050)[.050]	0.504(.063)[.063]	1.002(.032)[.032]	1.003(.050)[.050]
	large size groups: $\bar{r} = 102$, $n = 2279$, $\sum_{r=1}^{\bar{r}} \text{rank}(I_{m_r} - W_{nr}) = 2076$			
2SLS	0.497(.038)[.038]	–	1.002(.028)[.028]	1.006(.050)[.050]
G2SLS	0.500(.031)[.031]	0.460(.046)[.061]	1.001(.025)[.025]	1.002(.041)[.041]
ML1	0.499(.031)[.031]	0.499(.038)[.038]	1.001(.025)[.025]	1.003(.040)[.040]
ML2	0.502(.036)[.036]	0.497(.044)[.044]	1.001(.025)[.025]	1.002(.041)[.041]

Mean(SD)[RMSE]

Table 2: 2SLS, G2SLS and ML Estimation with Weak X's

	$\lambda_0 = 0.5$	$\rho_0 = 0.5$	$\beta_{10} = 0.2$	$\beta_{20} = 0.2$
	small size groups: $\bar{r} = 67$, $n = 557$, $\sum_{r=1}^{r} \text{rank}(I_{m_r} - W_{nr}) = 466$			
2SLS	0.591(.499)[.507]	–	0.194(.057)[.058]	0.177(.101)[.103]
G2SLS	0.633(.508)[.525]	0.226(.348)[.443]	0.198(.057)[.057]	0.178(.108)[.110]
ML1	0.517(.149)[.150]	0.453(.159)[.166]	0.204(.052)[.052]	0.201(.074)[.074]
ML2	0.506(.168)[.168]	0.469(.177)[.179]	0.205(.052)[.053]	0.204(.075)[.075]
	moderate size groups: $\bar{r} = 67$, $n = 877$, $\sum_{r=1}^{r} \text{rank}(I_{m_r} - W_{nr}) = 761$			
2SLS	0.527(.409)[.410]	–	0.197(.045)[.046]	0.189(.090)[.090]
G2SLS	0.548(.324)[.328]	0.356(.277)[.312]	0.199(.040)[.040]	0.190(.079)[.079]
ML1	0.507(.117)[.117]	0.473(.123)[.126]	0.203(.037)[.037]	0.203(.058)[.058]
ML2	0.513(.143)[.144]	0.471(.148)[.150]	0.203(.037)[.037]	0.203(.060)[.060]
	moderate size groups: $\bar{r} = 102$, $n = 1344$, $\sum_{r=1}^{r} \text{rank}(I_{m_r} - W_{nr}) = 1166$			
2SLS	0.481(.398)[.398]	–	0.199(.045)[.045]	0.195(.088)[.088]
G2SLS	0.508(.220)[.220]	0.396(.225)[.248]	0.199(.033)[.033]	0.194(.056)[.056]
ML1	0.503(.106)[.106]	0.483(.110)[.111]	0.201(.031)[.031]	0.200(.046)[.046]
ML2	0.512(.133)[.133]	0.479(.138)[.140]	0.202(.032)[.032]	0.200(.047)[.047]
	large size groups: $\bar{r} = 102$, $n = 2279$, $\sum_{r=1}^{r} \text{rank}(I_{m_r} - W_{nr}) = 2076$			
2SLS	0.470(.212)[.214]	–	0.201(.030)[.030]	0.204(.054)[.054]
G2SLS	0.515(.169)[.170]	0.455(.191)[.197]	0.200(.026)[.026]	0.198(.044)[.044]
ML1	0.502(.098)[.098]	0.487(.101)[.102]	0.201(.025)[.025]	0.202(.037)[.037]
ML2	0.514(.107)[.108]	0.478(.112)[.114]	0.201(.025)[.025]	0.201(.037)[.037]

Mean(SD)[RMSE]

Table 3: ML Estimation of Misspecified Models (R=102, n=1344)

	$\lambda_0 = 0.5$	$\rho_0 = 0.5$	$\beta_{10} = 1$	$\beta_{20} = 1$	likelihood value
correct model	.496(.038)[.039]	.501(.046)[.046]	1.002(.032)[.032]	1.003(.048)[.049]	$-1749.0(-)$
imposing $\hat{\lambda}_n = 0$	$-$	.805(.012)[.305]	1.000(.033)[.033]	1.144(.046)[.151]	$-1804.5(6.0\%)$
imposing $\hat{\beta}_{2n} = 0$	.799(.011)[.300]	$-.018(.048)[.521]$	0.821(.037)[.183]	$-$	$-1907.1(0.0\%)$
imposing $\hat{\rho}_n = 0$	.718(.016)[.218]	$-$	0.897(.032)[.108]	0.677(.050)[.327]	$-1803.9(7.0\%)$
	$\lambda_0 = -0.3$	$\rho_0 = 0.5$			
correct model	$-.295(.044)[.045]$	.496(.042)[.042]	1.001(.031)[.031]	0.997(.047)[.047]	$-1797.2(-)$
imposing $\hat{\rho}_n = 0$	.166(.033)[.467]	$-$	0.920(.032)[.086]	0.603(.056)[.401]	$-1824.5(0.0\%)$
	$\lambda_0 = 0.5$	$\rho_0 = -0.3$			
correct model	.497(.024)[.024]	$-.295(.038)[.038]$	1.003(.031)[.031]	1.003(.060)[.061]	$-1797.1(-)$
imposing $\hat{\rho}_n = 0$	.369(.023)[.133]	$-$	1.064(.031)[.071]	1.192(.056)[.200]	$-1820.8(0.0\%)$
	$\lambda_0 = 0.5$	$\rho_0 = 0$			
correct model	.499(.019)[.019]	$-$	1.002(.030)[.030]	1.000(.048)[.048]	$-1755.0(-)$
misspecified W_r	.826(.328)[.462]	$-$	1.316(.050)[.320]	1.833(.084)[.838]	$-1978.1(0.0\%)$

parameter estimates: Mean(SD)[RMSE]; likelihood value: Mean(frequency of exceeding likelihood value of the correct model)