Identification and Estimation of Spatial Econometric Models

with Group Interactions, Contextual Factors and Fixed Effects

by Lung-fei Lee*

Department of Economics

The Ohio State University

Columbus, OH

Date: January 2004

Abstract

This paper considers identification and estimation of structural interaction effects in a spatial autoregression model with social economics context. The model allows unobservables in the group structure, which may be correlated with included regressors. We show that both the endogenous and exogenous interaction effects can be identified if there are sufficient variations in group sizes. We consider the estimation of the model by the conditional maximum likelihood and instrumental variables methods. For the case with large group sizes, the possible identification can be weak in the sense that the estimates converge in distribution at low rates.

Key Words Spatial econometrics, spatial autoregression, interactions, social economics, group structure, fixed effects, conditional maximum likelihood, rate of convergence, IV estimation.

Classification: C13, C21, R15

Correspondence address:

Lung-fei Lee, Department of Economics, The Ohio State University, Columbus, OH 43210-1172.

E-mail: lflee@econ.ohio-state.edu

1. Introduction

A typical spatial autoregressive (SAR) model is specified as

$$Y_n = \lambda_0 \mathcal{W}_n Y_n + X_n \beta_0 + \mathcal{E}_n, \tag{1.1}$$

where $\mathcal{E}_n$ is a n-dimensional vector consisting of i.i.d. disturbances with zero mean and a variance σ_0^2. In this model, n is the total number of spatial units, X_n is an $n \times k$ matrix of regressors, and $\mathcal{W}_n$ is a specified constant spatial weights matrix (Cliff and Ord 1973). The SAR equation in (1.1) implies elements of Y_n shall be simultaneously determined given x's and the disturbances ϵ's as

$$Y_n = S_n^{-1}(\lambda_0) X_n \beta_0 + V_n, \tag{1.2}$$

where $S_n(\lambda_0) = I_n - \lambda_0 \mathcal{W}_n$ and $V_n = S_n^{-1}(\lambda_0) \mathcal{E}_n$. The SAR model has useful applications in urban and regional economic studies. In those studies, a region, a district or a county can be a spatial unit and its neighboring units defined in terms of a certain physical or economic distance can be interrelated.

This model has also applications in labor economics and social studies – the so called new social economics (Durlauf and Young 2001). For those studies, a spatial unit can be an individual belonging to a social group. The individuals in a group may interact with each other. Earlier empirical studies on group interactions are in Case (1991, 1992) in consumption pattern and technology adoption. In Case (1991), there are distinct districts and 'neighbors' refer to farmers who live in the same district. Case assumes that in a district, each neighbor of a spatial unit is given equal weight. Suppose there are R districts and there are m_r units in the rth district. The sample size is $n = \sum_{r=1}^{R} m_r$. The observations can be arranged into blocks where each block includes all members in a single district. With this arrangement, $\mathcal{W}_n$ is a block diagonal matrix, i.e,

$$\mathcal{W}_n = Diag(W_1, \cdots, W_R), \quad W_r = \frac{1}{m_r - 1}(l_{m_r} l'_{mr} - I_{m_r}), \quad r = 1, \cdots, R \tag{1.3}$$

where l_{m_r} is the m_r-dimensional column vector of ones, and I_{m_r} is the m_r-dimensional identity matrix. This model has also found interesting applications in Betrand et al. (2000) on welfare cultures, and Sacrerdote (2001) and Hanushek et al. (2003) on student achievement. The SAR model with such a spatial structure is a model with group interactions. The effect of social interactions in a SAR model is directly modeled in terms of observed outcome y's in a group. The parameter λ in (1.1) captures the contemporaneous and reciprocal effect of peer achievement.

2

An alternative specification of group interactions is the endogenous social effect model formulated in Manski (1993). In Manski's model, social interaction is modeled with expected outcomes and the expected outcomes are solutions from social equilibrium. Manski has pointed out some difficult identification issues on his social effect model as the expected outcome from social equilibrium in a group might be linearly depended on observed exogenous variables of a group in the model – the 'reflection' problem. A review and generalization of Manski's model is in Brock and Durlauf (2001). Manski (1993) and Brock and Durlauf (2001) separate peer influences into endogenous and exogenous (contextual) effects. The endogenous effect refers to the contemporaneous and reciprocal influences of peers. The contextual effect includes measures of peers unaffected by current behavior. The reflection problem refers to the difficulty to distinguish between behavioral and contextual factors. There is a close similarity of the SAR model with group interactions to the Manski endogenous effect model when numbers of members in groups are large. Even so, the SAR model with group interactions does have a distinguishable feature which makes identification of social effect likely feasible. In Manski's social effect model, the identification of social effect is through the mean regression function and there are no correlations in disturbances generated by social interactions in his setting. For the SAR model, the identification of the spatial effect λ in (1.1) can be based on two sources. One is the mean regression function $E(Y_n|X_n) = S_n^{-1}(\lambda)X_n$ in (1.2) and another is the correlation across disturbances in V_n because $V_n = S_n^{-1}(\lambda)\mathcal{E}_n$.

Even so, one may concern about other possible specification issues which may have effects on identification in a SAR model. One main concern is on possible unobservables in a group, as unobservables in a group may have direct effect on observed outcomes. The unobservables also make the total disturbances to be correlated across individuals in a group. Manski (1993), Brock and Durlauf (2001) and Moffitt (2001) point out that empirical analyses of peer influences have been inhibited by both conceptual and data problems. Moffitt's criticism is, in particular, relevant as his discussions are presented for the SAR model in (1.1) with a group structure. Moffitt (2001) argues that the basic identification problem of group interaction effects is how to distinguish within group correlations of outcomes that arise from social interactions from correlations that arise for other reasons, in particular, correlated unobservables. Correlated unobservables may arise if there are group-specific components in disturbances that vary across groups and are correlated with exogenous characteristics of the individuals. He believes that there are two generic sources of correlated unobservables. The first is they may arise from sorting and endogenous group membership, and from prefer-

3

ences or other forces that lead certain types of individuals to be grouped together. The second source may be from some common environmental factors. For example, for the study of student achievement, Hanushek et al. (2003) indicates that one important and relevant example for common environmental factors in student achievement may be some systematic but unmeasured elements of teacher quality. Moffitt provides further a review and critique of studies in a set of housing experiments aimed at understanding neighborhood effects. Thus, possible correlated unobservable is the most genuine concern for empirical detection and measurement of social interaction effects with a SAR model in a group setting.

In this paper, we consider the SAR model with both endogenous group interaction and contextual factors and allow the existence of correlated unobservables as a fixed effect in a group. We show that the identification of the interaction effects is possible if there are sufficient variations in group sizes in the structural model. We characterize the identification conditions.[1] However, when identification is possible, it may be weak when there are large group interactions. We consider the estimation and the consequences of both strong and weak identification features on possible estimators.

2. A SAR model with group interactions and fixed effects

The SAR model (1.1) with the spatial scenario (1.3) has a well-defined group structure. The structural social interaction effect is captured by the parameter λ. In order to capture possible unobservables which may have common effects on the outcomes of y's in a group, we extend the SAR model (1.1) with fixed effects α_r and an additional explanatory component $W_n X_{r,2}$ for contextual effects:

$$Y_r = \lambda_0 W_r Y_r + X_{r,1}\beta_{10} + W_r X_{r,2}\beta_{20} + l_{m_r}\alpha_r + \epsilon_r, \qquad r = 1, \cdots, R, \tag{2.1}$$

where Y_r, $X_{r,1}$ and $X_{r,2}$ are the vector and matrices of the m_r observations in the rth group, or, equivalently in term of each unit i in a group r,

$$y_{ri} = \lambda_0 \left(\frac{1}{m_r - 1} \sum_{j=1, j\neq i}^{m_r} y_{rj} \right) + x_{ri,1}\beta_{10} + \left(\frac{1}{m_r - 1} \sum_{j=1, j\neq i}^{m_r} x_{rj,2} \right)\beta_{20} + \alpha_r + \epsilon_{ri}, \tag{2.1$'$}$$

with $i = 1, \cdots, m_r$ and $r = 1, \cdots, R$, where y_{ri} is the ith individual in the rth group, $x_{ri,1}$ and $x_{ri,2}$ are, respectively, k_1 and k_2 row vectors of exogenous variables, and ϵ_{ri}'s are i.i.d. $(0, \sigma_0^2)$. The α_r represents the unobservables of the r-th group. As those unobservables may correlate with exogenous variables, they are treated as fixed effects. The vectors of all exogenous variables x_{ri}'s may vary across individuals in a

---

[1] The illustration in Moffitt (2001) on underidentification of interaction effects is based on a model with two members in each group. As it will be shown below, if there is no variation in groups sizes, the interaction effects in a SAR model can not be identified.

group as any group invariant variables will be captured in α_r. In a general setting, $x_{ri,1}$ and $x_{ri,2}$ are subvectors of x_{ri}, which may or may not have common elements. The introduced variables $\sum_{j=1,j\neq i}^{m_r} x_{ri,2}$ of $W_n X_{r,2}$ allow social interaction effect through observed neighborhood characteristics, which has been termed 'contextual' effect in Manski (1993). Neighborhood characteristics have often been used in empirical studies of neighborhood effects, e.g., Weinberg, Reagan and Yankow (2003), in a regression setting. One may wonder whether this additional contextual effect would compound the identification of the structural interaction effect λ_0 in (2.1), especially, when $x_{ri,2}$ is identical to $x_{ri,1}$. Finally, we note that m_r is the number of members in the r-th group.

For this model, it is natural to require that $\lambda \in (-1,1)$. This follows from the equilibrium structure that $S_n(\lambda)$ shall be nonsingular for any value λ of interest. With W_r in (1.3), the determinant of $S_r(\lambda)$, where $S_r(\lambda) = I_{m_r} - \lambda W_r$, is $|S_r(\lambda)| = (1 - \lambda)(\frac{m_r - 1 + \lambda}{m_r - 1})^{m_r - 1}$ because $S_r(\lambda) = (\frac{m_r - 1 + \lambda}{m_r - 1})(I_{m_r} - \frac{\lambda}{m_r - 1 + \lambda} l_{m_r} l'_{m_r})$. This determinant is nonsingular if and only if $\lambda \neq 1$ and $(m_r - 1 + \lambda) \neq 0$ for all r. As $m_r \geq 2$, λ lies in the parameter space Λ, where Λ is a subset of $(-1,1)$, in order that $S_n(\lambda)$ does not become singular at any value in this interval.[2]

It is revealing to decompose this equation into two parts. Denote $J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l'_{m_r}$. Because of the group structure, (2.1) implies that

$$\frac{1}{m_r} l'_{m_r} Y_r = \lambda_0 \frac{1}{m_r} l'_{m_r} W_r Y_r + \frac{1}{m_r} l'_{m_r} X_{r,1} \beta_{10} + \frac{1}{m_r} l'_{m_r} W_r X_{r,2} \beta_{20} + \alpha_r + \frac{1}{m_r} l'_{m_r} \epsilon_r,$$

and

$$J_r Y_r = \lambda_0 J_r W_r Y_r + J_r X_{r,1} \delta_0 + J_r W_r X_{r,2} \beta_{20} + J_r \epsilon_r.$$

Because $W_r = \frac{1}{m_r - 1}(l_{m_r} l'_{m_r} - I_{m_r})$, $J_r W_r = -\frac{1}{m_r - 1} J_r$ and $l'_{m_r} W_r = l'_{m_r}$. Therefore, one has

$$(1 - \lambda_0)\bar{y}_r = \bar{x}_{r,1}\beta_{10} + \bar{x}_{r,2}\beta_{20} + \alpha_r + \bar{\epsilon}_r, \quad r = 1, \cdots, R, \tag{2.2}$$

and

$$(1 + \frac{\lambda_0}{m_r - 1})(y_{ri} - \bar{y}_r) = (x_{ri,1} - \bar{x}_{r,1})\beta_{10} - \frac{1}{m_r - 1}(x_{ri,2} - \bar{x}_{r,2})\beta_{20} + (\epsilon_{ri} - \bar{\epsilon}_r), \quad i = 1, \cdots, m_r; \ r = 1, \cdots, R, \tag{2.3}$$

where $\bar{y}_r = \frac{1}{m_r} \sum_{i=1}^{m_r} y_{ri}$, $\bar{x}_{r,1} = \frac{1}{m_r} \sum_{i=1}^{m_r} x_{ri,1}$ and $\bar{x}_{r,2} = \frac{1}{m_r} \sum_{i=1}^{m_r} x_{ri,2}$ are means for the r-th group. The equation (2.2) may be called a 'between' equation and that in (2.3) is a 'within' equation as they have

---

[2] The determinant $|S_r(\lambda)|$ shall not be zero or changes its sign on Λ. Because $\lambda = 0$ shall be in Λ, it follows that $1 > \lambda > 1 - m_r$ for all r. As m_r can be 2, therefore $1 > \lambda > -1$.

similarity with those of a panel data regression model (Hsiao 1986). The possible effects due to interactions are revealing in the reduced-form between and within equations:

$$\bar{y}_r = \bar{x}_{r,1}\frac{\beta_{10}}{(1-\lambda_0)} + \bar{x}_{r,2}\frac{\beta_{20}}{(1-\lambda_0)} + \frac{\alpha_r}{(1-\lambda_0)} + \frac{\bar{\epsilon}_r}{(1-\lambda_0)}, \quad r = 1, \cdots, R, \qquad (2.4)$$

and

$$(y_{ri} - \bar{y}_r) = (x_{ri,1} - \bar{x}_{r,1})\frac{(m_r - 1)\beta_{10}}{(m_r - 1 + \lambda_0)} - (x_{ri,2} - \bar{x}_{r,2})\frac{\beta_{20}}{(m_r - 1 + \lambda_0)} + \frac{(m_r - 1)}{(m_r - 1 + \lambda_0)}(\epsilon_{ri} - \bar{\epsilon}_r), \quad (2.5)$$

with $i = 1, \cdots, m_r$; $r = 1, \cdots, R$. Suppose that the interaction effect λ is positive. For the average group outcome $\bar{y}_r$, positive group interaction raises the regression effects of $\bar{x}_{r,1}$ and $\bar{x}_{r,2}$ on $\bar{y}_r$ in (2.2) by the factor $\frac{1}{(1-\lambda)}$. It raises also the variance of $\bar{y}_r$ (with the same $\bar{x}_r$) across different groups by a factor of $\frac{1}{(1-\lambda)^2}$.[3] However, in the presence of the unobservables in α_r, the group interaction effect λ in (2.4) can not be identified as it can not be isolated from α_r. In the presence of unobservables represented by the fixed effect α_r, the between equation (2.4) does not have any degree of freedoms to identify (and estimate) any of the unknown parameters. The possible identification will rest on the within equation (2.5). A positive interaction also diminishes the derivation of an individual outcome y_{ri} from the group average $\bar{y}_r$ through both its mean regression function and the disturbance of the within equation. The identification of λ_0 will rely on various degrees of derivations across groups. This is possible when different groups have different numbers of members. When all groups have the same number of members, i.e., m_r is a constant, say m, for all r, the effect λ can not be identifiable from the within equation. This is apparent as only the functions $\frac{(m-1)\delta_{10}}{(m-1+\lambda_0)}$, $\frac{(m-1)\delta_{20}}{(m-1+\lambda_0)}$, and $\frac{(m-1)\sigma_0^2}{(m-1+\lambda_0)}$ may ever be identifiable from (2.5).[4]

The possible identification of the structural parameters in the group interaction model with fixed effects relies on various group sizes in a sample.[5] This identification can be weak, especially, when there are large group interactions. When m_r are all large, the factors $(1 + \frac{\lambda_0}{m_r - 1})$ may be close to one and λ may not be easily estimated from (2.3). In subsequent sections, we characterize possible consistent estimation of

---

[3] The increasing group variance due to positive group interaction is the crucial observation in the analysis of criminal behavior in Glaeser et al. (1996).

[4] If members in a group are known to exert different effects on one another, one may expect that more structured W_r's rather than that in (1.3) might help identification in a certain way too.

[5] In two studies related to group interactions, group size is one of the interesting variables. Hoxby (2000) has investigated the effect of class size on student achievement. Rees et al. (2003) has investigated the effect of group size on workers' productivity. Their motivations are, however, different from ours. The class size in Hoxby (2000) is a factor in a school's production function. A larger group size in Rees et al. (2003) presents the difficulty for monitoring performance of workers.

this model and asymptotic properties of estimators for both small and large group interaction cases. The estimators that we consider are the conditional maximum likelihood (CML) estimator and the two-stage least squares (2SLS) estimators. The maximum likelihood and 2SLS are two popular approaches for the estimation of a SAR model (without fixed effects); see, e.g. Ord (1975) and Kelejian and Prucha (1998). When m_r are all large, it is intuitively appealing to approximate the within equation by the conventional equation $(y_{ri} - \bar{y}_r) = (x_{ri,1} - \bar{x}_{r,1})\beta_{10} + (\epsilon_{ri} - \bar{\epsilon}_r)$ and estimate the parameter β_{10} by the method of ordinary least squares (OLS). As this conventional within equation is slightly misspecified for models with large group interactions, it is of interest to investigate properties of the OLS estimator of β_{10} in this case. We discover a surprising result on the OLS estimate.

3. Conditional maximum likelihood estimation

3.1 The conditional likelihood function and the CML estimator (CMLE)

For analytical convenience, denote $z_{ri} = (x_{ri,1}, -\frac{m}{m_r-1}x_{ri,2})$ where $m = \frac{1}{R}\sum_{r=1}^R m_r$ is the mean size of groups. Let $\delta_m = (\beta_1', \beta_2'/m)'$. Under the assumption that ϵ's are normally distributed, the likelihood function for the within equation (2.3) is

$$L_{w,n}(\theta) = \prod_{r=1}^R \frac{\sqrt{m_r}}{(2\pi)^{\frac{m_r-1}{2}}}(\frac{1}{\sigma c_r(\lambda)})^{m_r-1}\exp\{-\frac{1}{2\sigma^2}(\frac{1}{c_r(\lambda)}Y_r - Z_r\delta_m)'J_r(\frac{1}{c_r(\lambda)}Y_r - Z_r\delta_m)\}, \qquad (3.1)$$

where $c_r(\lambda) = (\frac{m_r-1}{m_r-1+\lambda})$, $\theta = (\lambda, \beta', \sigma^2)'$, $\beta = (\beta_1', \beta_2')'$, $n = \sum_{r=1}^R m_r$, and $J_r = (I_{m_r} - \frac{1}{m_r}l_{m_r}l_{m_r}')$. A list of often used notations is collected in the Appendix for convenience of reference.

The (3.1) can be derived as follows. Because the components of the m_r-dimensional vector $(y_{r1} - \bar{y}_r, \cdots, y_{rm_r} - \bar{y}_r)$ are linearly dependent, it is sufficient to consider the first $(m_r - 1)$ linearly independent components. The variance matrix of the corresponding disturbance vector is

$$var(\epsilon_{r1} - \bar{\epsilon}_r, \cdots, \epsilon_{r(m_r-1)} - \bar{\epsilon}_r) = \sigma^2(I_{m_r-1} - \frac{1}{m_r}l_{m_r-1}l_{m_r-1}'),$$

which has the determinant $|I_{m_r-1} - \frac{1}{m_r}l_{m_r-1}l_{m_r-1}'| = \frac{1}{m_r}$ and $(I_{m_r-1} - \frac{1}{m_r}l_{m_r-1}l_{m_r-1}')^{-1} = I_{m_r-1} + l_{m_r-1}l_{m_r-1}'$. Furthermore,

$$(y_{r1} - \bar{y}_r, \cdots, y_{r(m_r-1)} - \bar{y}_r)'(I_{m_r-1} + l_{m_r-1}l_{m_r-1}')(y_{r1} - \bar{y}_r, \cdots, y_{r(m_r-1)} - \bar{y}_r)' = \sum_{i=1}^{m_r}(y_{ri} - \bar{y}_r)^2.$$

The likelihood function (3.1) follows. This likelihood function does not involve any fixed effects α's. It is the conditional likelihood function of the whole sample y_{ri}'s conditional on the sufficient statistics $\bar{y}_r$, $r = 1, \cdots, R$. The whole sample $y_{ri}'s$ can be transformed one-to-one into the observations $y_{ri} - \bar{y}_r$ with

7

$i = 1, \cdots, m_r$ and $r = 1, \cdots, R$ for the within equation, and $\bar{y}_r$ with $r = 1, \cdots, R$ for the between equation. Under normality, the disturbances of the within and between equations are independent. The likelihood function of these transformed observations is the product of the likelihood function of the within equation and that of the between equation. The likelihood function (3.1) is therefore the conditional likelihood function.

The log likelihood of (3.1) is

$$\ln L_{w,n}(\theta) = c + \sum_{r=1}^{R}(m_r - 1)\ln(m_r - 1 + \lambda) - \frac{(n-R)}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{r=1}^{R}(\frac{1}{c_r(\lambda)}Y_r - Z_r\delta_m)'J_r(\frac{1}{c_r(\lambda)}Y_r - Z_r\delta_m),$$
(3.2)

where c is a constant. This log likelihood function can be concentrated at λ, which has computational and analytical advantages over the whole function. Given a possible value λ, the CML estimates of β and σ^2 are, respectively,

$$\hat{\beta}_n(\lambda) = \begin{pmatrix} I_{k_1} & 0 \\ 0 & mI_{k_2} \end{pmatrix}\left(\sum_{r=1}^{R}Z_r'J_rZ_r\right)^{-1}\sum_{r=1}^{R}Z_r'J_rY_r\frac{1}{c_r(\lambda)},$$
(3.3)

and

$$\hat{\sigma}_n^2(\lambda) = \frac{1}{n-R}\left\{\sum_{r=1}^{R}\frac{1}{c_r^2(\lambda)}Y_r'J_rY_r - \sum_{r=1}^{R}\frac{1}{c_r(\lambda)}Y_r'J_rZ_r\left(\sum_{r=1}^{R}Z_r'J_rZ_r\right)^{-1}\sum_{r=1}^{R}Z_r'J_rY_r\frac{1}{c_r(\lambda)}\right\}.$$
(3.4)

The concentrated log likelihood function (3.2) at λ is[6]

$$\ln L_{c,n}(\lambda) = c_1 + \sum_{r=1}^{R}(m_r - 1)\ln(m_r - 1 + \lambda) - \frac{(n-R)}{2}\ln\hat{\sigma}_n^2(\lambda).$$
(3.5)

The derivative of (3.5) with λ is

$$\frac{\partial \ln L_{c,n}(\lambda)}{\partial \lambda} = \sum_{r=1}^{R}c_r(\lambda) - \frac{(n-R)}{2}\frac{\partial \ln\hat{\sigma}_n^2(\lambda)}{\partial \lambda}.$$
(3.6)

The following assumptions are some basic ones for the model:

Assumption 1. The ϵ_{ri}'s are i.i.d. $N(0, \sigma_0^2)$.

Assumption 2. The x_{ri}'s are assumed to be bounded. The probability limit of $\frac{1}{n}\sum_{r=1}^{R}Z_r'J_rZ_r$ exists and is a non-singular matrix.

---

[6] In the special case that $m_r(= m)$ are all equal, $c_r(\lambda)(= c(\lambda))$ is invariant with r. In this case, $\hat{\sigma}_n^2(\lambda) = \frac{1}{c^2(\lambda)}\frac{1}{n-R}\{\sum_{r=1}^{R}Y_r'J_rY_r - \sum_{r=1}^{R}Y_r'J_rZ_r(\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}Z_r'J_rY_r\}$ and $\frac{\partial \ln\hat{\sigma}_n^2(\lambda)}{\partial \lambda} = \frac{2}{m-1+\lambda}$. These imply $\frac{\partial \ln L_{c,n}(\lambda)}{\partial \lambda} = 0$ for all λ. That is, the conditional likelihood function does not provide information on λ_0 under such a circumstance.

8

Assumption 3. Suppose $m_r = a_r m \geq 2$ where a_r's are proportional factors with $\frac{1}{R} \sum_{r=1}^{R} a_r = 1$. There exist a lower bound $a_L > 0$ and an upper bound $a_U < \infty$ such that $a_L \leq a_r \leq a_U$ for all $r = 1, 2, \cdots$.

Assumption 4. The parameter space Λ is a connected compact subset of $(-1, 1)$ with λ_0 in its interior. This compact parameter space assumption is needed as the CM approach works with the concentrated likelihood (3.5), which is nonlinear in λ. One does not need to impose restricted parameter spaces for β and σ^2 as the CML estimates are naturally coming out from (3.3) and (3.4).

We are considering asymptotic properties of the estimates as the population size n goes to infinity. In the scenario with small group interactions, i.e., $\{m_r\}$ are bounded, it will correspond to the number of groups R tends to infinity. In order to allow large group interactions, its shall be understood that n goes to infinity refers to both R and m tend to infinity.[7] In the large group interactions case, the consistency of the estimates will require the following setting.

Assumption 5. As n goes to infinity, $\frac{R}{m}$ tends to infinity.

The Assumption 5 is equivalent to that $\frac{m^2}{n}$ tends to zero or $\frac{\sqrt{n}}{m}$ tends to infinity. Intuitively, this requires that whenever m goes to infinity, m does not go to infinity at a rate faster than or equal to R. For the scenario of small group interactions, that n goes to infinity is equivalent to that R tends to infinity. With large group interactions, one needs to have much larger R than m in order to achieve consistent estimates.

3.2 Uniform convergence of the concentrated log conditional likelihood function

Define a nonstochastic function

$$Q_{c,n}(\lambda) = c_1 + \sum_{r=1}^{R} (m_r - 1) \ln(m_r - 1 + \lambda) - \frac{(n-R)}{2} \ln \sigma_n^{*2}(\lambda), \tag{3.7}$$

where

$$
\begin{aligned}
\sigma_n^{*2}(\lambda) =& \frac{1}{n-R} \Big\{ \sum_{r=1}^{R} d_r^2(\lambda)(Z_r \delta_{m0})' J_r (Z_r \delta_{m0}) - \sum_{r=1}^{R} d_r(\lambda)(Z_r \delta_{m0})' J_r Z_r \big(\sum_{r=1}^{R} Z_r' J_r Z_r \big)^{-1} \\
& \cdot \sum_{r=1}^{R} d_r(\lambda) Z_r' J_r (Z_r \delta_{m0}) \Big\} + \frac{\sigma_0^2}{n-R} \sum_{r=1}^{R} (m_r - 1) d_r^2(\lambda) \\
=& \, (\frac{\lambda - \lambda_0}{m})^2 \frac{1}{n-R} \Big\{ \sum_{r=1}^{R} (\frac{m}{t_r})^2 (Z_r \delta_{m0})' J_r (Z_r \delta_{m0}) \\
& - \sum_{r=1}^{R} \frac{m}{t_r} (Z_r \delta_{m0})' J_r Z_r \big(\sum_{r=1}^{R} Z_r' J_r Z_r \big)^{-1} \sum_{r=1}^{R} \frac{m}{t_r} Z_r' J_r (Z_r \delta_{m0}) \Big\} + \frac{\sigma_0^2}{n-R} \sum_{r=1}^{R} (m_r - 1) d_r^2(\lambda),
\end{aligned}
\tag{3.8}
$$

---

[7] For precise notations, a subscript n may be attached to R, m and a_rs such that R_n and m_n may tend to infinity and a_{rn}'s remain bounded by a_L and a_U in Assumption 3 as n tends to infinity. The above convention simplifies the notations.

with $d_r(\lambda) = (\frac{m_r - 1 + \lambda}{m_r - 1 + \lambda_0})$ and $t_r = (m_r - 1 + \lambda_0)$. The $Q_{c,n}(\lambda)$ is motivated by $\max_{\beta,\sigma^2} E(\ln L_{w,n}(\theta))$. This optimization problem gives the optimum solutions

$$\delta_n^*(\lambda) = \begin{pmatrix} I_{k_1} & 0 \\ 0 & mI_{k_2} \end{pmatrix} (\sum_{r=1}^R Z_r' J_r Z_r)^{-1} \sum_{r=1}^R d_r(\lambda) Z_r' J_r Z_r \delta_{m0} (= E(\hat{\beta}_n(\lambda))),$$

and $\sigma_n^*(\lambda)$ in (3.8). Indeed, (3.7) is simply

$$Q_{c,n}(\lambda) = \max_{\beta,\sigma^2} E(\ln L_{w,n}(\beta, \sigma^2, \lambda)). \tag{3.7$'$}$$

It shall be shown that $\frac{m^2}{n}[(\ln L_{c,n}(\lambda) - \ln L_{c,n}(\lambda_0)) - (Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))] = o_P(1)$ uniformly on Λ.

Proposition 1. *Under Assumptions 1-5, $\frac{m^2}{n}[(\ln L_{c,n}(\lambda) - \ln L_{c,n}(\lambda_0)) - (Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))]$ converges in probability to zero uniformly in $\lambda \in \Lambda$.*

3.3 Identification and consistency of the CMLE

It shall be shown that $\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))$ satisfies the identification uniqueness condition (White 1994).

Define

$$\sigma_n^2(\lambda) = \frac{\sigma_0^2}{n-R} \sum_{r=1}^R (m_r - 1) d_r^2(\lambda), \tag{3.9}$$

and

$$D_n(\lambda) = \frac{m^2}{n} \left\{ \sum_{r=1}^R (m_r - 1)[\ln(m_r - 1 + \lambda) - \ln(m_r - 1 + \lambda_0)] - (\frac{n-R}{2})[\ln \sigma_n^2(\lambda) - \ln \sigma_0^2] \right\}. \tag{3.10}$$

Because $\sigma_n^{*2}(\lambda_0) = \sigma_0^2$ from (3.8), (3.7) implies

$$\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)) = D_r(\lambda) - \frac{m^2}{n}(\frac{n-R}{2})(\ln \sigma_n^{*2}(\lambda) - \ln \sigma_n^2(\lambda)). \tag{3.11.}$$

The $D_n(\lambda)$ in (3.10) is motivated by the following construction:

$$D_n(\lambda) = \frac{m^2}{n}[\max_{\beta,\sigma^2} E_{\beta_0=0}(\ln L_{w,n}(\beta, \sigma^2, \lambda)) - \max_{\beta,\sigma^2} E_{\beta_0=0}(\ln L_{w,n}(\beta, \sigma^2, \lambda_0))], \tag{3.10$'$}$$

where $E_{\beta_0=0}$ is the expectation operator under $\beta_0 = 0$, σ_0^2 and λ_0 being the 'true' parameters, from (3.7) and (3.7)$'$. This is because when $\beta_0 = 0$, $\sigma_n^{*2}(\lambda) = \sigma_n^2(\lambda)$ in (3.8) and, consequently, (3.7) and (3.7)$'$ imply (3.10)$'$. As the information inequality gives $E_{\beta_0=0}(\ln L_{w,n}(\beta, \sigma^2, \lambda)) \leq E_{\beta_0=0}(\ln L_{w,n}(\beta_0, \sigma_0^2, \lambda_0))$ for all β, σ^2 and λ,

$$D_n(\lambda) \leq \frac{m^2}{n}\{E_{\beta_0=0}(\ln L_{w,n}(\beta_0, \sigma_0^2, \lambda_0)) - \max_{\beta,\sigma^2} E_{\beta_0}(\ln L_{w,n}(\beta, \sigma^2, \lambda_0))\} = 0$$

for all λ. Eqs.(3.8) and (3.9) imply that

$$
\frac{m^2}{n}(n-R)(\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda))
$$

$$
= \frac{(\lambda-\lambda_0)^2}{n}\Big\{\sum_{r=1}^{R}(\frac{m}{t_r})^2(Z_r\delta_{m0})'J_r(Z_r\delta_{m0}) - \sum_{r=1}^{R}\frac{m}{t_r}(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}\frac{m}{t_r}Z_r'J_r(Z_r\delta_{m0})\Big\},
$$

$$(3.12)$$

which is positive for any $\lambda \neq \lambda_0$ by the generalized Schwartz inequality. In order that this positiveness will not be lost in the limit, we need the following assumption.

Assumption 6.1.(Identification 1)

$$
\lim_{n\to\infty}\frac{1}{n}\Big\{\sum_{r=1}^{R}(\frac{m}{t_r})^2(Z_r\delta_{m0})'J_r(Z_r\delta_{m0}) - \sum_{r=1}^{R}\frac{m}{t_r}(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}\frac{m}{t_r}Z_r'J_r(Z_r\delta_{m0})\Big\} > 0.
$$

The identification assumption 6.1 can break down if $\beta_0 = 0$. If $\beta_{10} = 0$ but $\beta_{20} \neq 0$, i.e., z_{ri} will effectively contain only the contextual effect variables, $-\frac{1}{m_r-1}(x_{ri,2} - \bar{x}_{r,2})$, (3.12) is of order $O(\frac{1}{m})$. For the latter, when m goes to infinity, the identification Assumption 6.1 will not be applicable. As long as $\beta_{10} \neq 0$, this identification assumption will likely be satisfied under the setting in Assumption 3, because $\frac{m}{t_r} \to \frac{1}{a_r}$ as $m \to \infty$. Note that under Assumption 2 that the limiting matrix of $\frac{1}{n}\sum_{r=1}^{R}Z_r'J_rZ_r$ is nonsingular, Assumption 6.1 is equivalent to that the limiting matrix of $\frac{1}{n}\sum_{r=1}^{R}(-\frac{m}{t_r}Z_r\delta_{m0}, Z_r)'J_r(-\frac{m}{t_r}Z_r\delta_{m0}, Z_r)$ is nonsingular (e.g., Theil 1971, p.18 on the inverse of a partitioned symmetric matrix).

If Assumption 6.1 breaks down, one has to resort to the covariance structure of the disturbance term in (2.5) captured by the function $D_n(\lambda)$ in (3.10). The $D_n(\lambda)$ can be rewritten as

$$
D_n(\lambda) = \frac{n-R}{2n}m^2\left[\sum_{r=1}^{R}(\frac{m_r-1}{n-R})\ln d_r^2(\lambda) - \ln\left(\sum_{r=1}^{R}\frac{m_r-1}{n-R}d_r^2(\lambda)\right)\right],
$$

which is strictly negative for any $\lambda \neq \lambda_0$ by Jensen's inequality applied to the logarithmic function as long as m_r's are not identical to each other so that $d_r(\lambda)$'s vary across r. The following assumption is needed to guarantee that the negativeness of $D_n(\lambda)$ does not vanish in the limit.

Assumption 6.2. (Identification 2) For any $\lambda \neq \lambda_0$,

$$
\limsup_{n\to\infty} m^2\left[\sum_{r=1}^{R}(\frac{m_r-1}{n-R})\ln d_r^2(\lambda) - \ln(\sum_{r=1}^{R}\frac{m_r-1}{n-R}d_r^2(\lambda))\right] < 0.
$$

Under the identification condition in Assumption 6.2, λ_0 is the unique maximum of $\lim_{n\to\infty} D_n(\lambda)$ in Λ. The global identification conditions in Assumptions 6.1 or 6.2 imply a local identification condition that

11

$\lim_{n\to\infty} \frac{m^2}{n} \frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial\lambda^2}$ shall be negative definite. It can be shown from (3.7) and (3.10) that

$$\frac{m^2}{n} \frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial\lambda^2} = \frac{\partial^2 D_n(\lambda_0)}{\partial\lambda^2} - \frac{1}{n\sigma_0^2} \Big\{ \sum_{r=1}^{R} (\frac{m}{t_r})^2 (Z_r\delta_{m0})' J_r (Z_r\delta_{m0})$$
$$- \sum_{r=1}^{R} \frac{m}{t_r} (Z_r\delta_{m0})' J_r Z_r (\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} \frac{m}{t_r} Z_r' J_r (Z_r\delta_{m0}) \Big\},$$

(3.13)

where

$$\frac{\partial^2 D_n(\lambda_0)}{\partial\lambda^2} = -2\frac{m^2}{n} \left[\sum_{r=1}^{R} \frac{c_r(\lambda_0)}{t_r} - \frac{1}{n-R} (\sum_{r=1}^{R} c_r(\lambda_0))^2 \right] = -2\frac{m-1}{m} \sum_{r=1}^{R} (\frac{m_r-1}{n-R})\omega_r^2, \qquad (3.14)$$

with $\omega_r = (\frac{m}{t_r} - \sum_{s=1}^{R}(\frac{m_s-1}{n-R})\frac{m}{t_s})$, which is the deviation of $\frac{m}{t_r}$ from its weighted mean of the R groups. A sufficient condition for local identification induced by Assumption 6.2 is that the limiting weighted variance of $\frac{m}{t_r}$ does not vanish, i.e.,

$$\lim_{n\to\infty} \sum_{r=1}^{R} (\frac{m_r-1}{n-R})\omega_r^2 > 0. \qquad (3.15)$$

The following proposition summarizes the consistency of the CMLE $\hat{\lambda}_n$.

Proposition 2. *Under the Assumption 1-5, 6.1 or 6.2, the identification uniqueness condition that, for any open neighborhood $N_\epsilon(\lambda)$ of λ_0 in Λ,*

$$\limsup_{n\to\infty} \max_{\lambda\in\bar{N}_\epsilon(\lambda_0)} \frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)) < 0,$$

where $\bar{N}_\epsilon(\lambda_0)$ is the complement of $N_\epsilon(\lambda_0)$ in Λ, will hold, and $\hat{\lambda}_n$ is a consistent estimator of λ_0.

3.4 Asymptotic distribution of the CMLE

For asymptotic distribution, it is important to investigate $\frac{\partial \ln L_{w,c}(\lambda)}{\partial\lambda}$ evaluated at λ_0. Define

$$l_{n1} = \sum_{r=1}^{R} \frac{1}{t_r} (Z_r\delta_{m0})' J_r \epsilon_r, \quad l_{n2} = \sum_{r=1}^{R} \frac{1}{t_r} (Z_r\delta_{m0})' J_r Z_r \left(\sum_{r=1}^{R} Z_r' J_r Z_r \right)^{-1} \sum_{r=1}^{R} Z_r' J_r \epsilon_r,$$

and

$$q_{n1} = \sum_{r=1}^{R} \frac{1}{t_r} \epsilon_r' J_r \epsilon_r, \quad q_{n2} = \sum_{r=1}^{R} \frac{1}{t_r} \epsilon_r' J_r Z_r \left(\sum_{r=1}^{R} Z_r' J_r Z_r \right)^{-1} \sum_{r=1}^{R} Z_r' J_r \epsilon_r.$$

One has

$$\frac{\partial\hat{\sigma}_n^2(\lambda_0)}{\partial\lambda} = \frac{2}{n-R}(l_{n1} + q_{n1} - l_{n2} - q_{n2}),$$

and

$$\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial\lambda} = \frac{1}{\hat{\sigma}_n^2(\lambda_0)} \left\{ (\hat{\sigma}_n^2(\lambda_0) - \sigma_0^2) \sum_{r=1}^{R} \frac{1}{c_r(\lambda_0)} - l_{n1} + l_{n2} - \left(q_{n1} - \sum_{r=1}^{R} \frac{\sigma_0^2}{c_r(\lambda_0)} \right) + q_{n2} \right\}. \qquad (3.16)$$

12

The following proposition shows that $\frac{1}{n}\frac{\partial \ln L_{w,c}(\lambda_0)}{\partial \lambda}$ does not have the usual $\sqrt{n}$-rate of convergence in distribution. Instead, its rate of convergence is of the higher order of $m\sqrt{n}$. Asymptotically, this score is a sum of a linear term and a quadratic term of $\mathcal{E}_n$.

Proposition 3. *Under Assumptions 1-3,*

$$\frac{m}{\sqrt{n}}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda} = \frac{1}{\sigma_0^2}\frac{m}{\sqrt{n}}(Q_n - L_n) + O_P(\frac{1}{\sqrt{n}}), \tag{3.17}$$

where

$$Q_n = (\sum_{r=1}^{R}\frac{1}{c_r(\lambda_0)})\left(\frac{1}{n-R}\sum_{r=1}^{R}\epsilon_r'J_r\epsilon_r - \sigma_0^2\right) - \left(q_{n1} - \sum_{r=1}^{R}\frac{\sigma_0^2}{c_r(\lambda_0)}\right) = \frac{1}{m}\sum_{r=1}^{R}\omega_r(\epsilon_r'J_r\epsilon_r - \sigma_0^2(m_r-1)), \tag{3.18}$$

with $\omega_r = -[\frac{m}{t_r} - \sum_{s=1}^{R}(\frac{m_s-1}{n-R})\frac{m}{t_s}]$, and

$$L_n = \sum_{r=1}^{R}\frac{1}{t_r}(Z_r\delta_{m0})'J_r\epsilon_r - \sum_{r=1}^{R}\frac{1}{t_r}(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}Z_r'J_r\epsilon_r. \tag{3.19}$$

Under the normality for $\mathcal{E}_n$,

$$var(\frac{m}{\sqrt{n}}Q_n) = 2\sigma_0^4(\frac{m-1}{m})\sum_{r=1}^{R}(\frac{m_r-1}{n-R})\omega_r^2 \tag{3.20}$$

and

$$var(\frac{m}{\sqrt{n}}L_n) = \frac{\sigma_0^2}{n}\{\sum_{r=1}^{R}(\frac{m}{t_r})^2(Z_r\delta_{m0})'J_r(Z_r\delta_{m0}) - \sum_{r=1}^{R}\frac{m}{t_r}(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}\frac{m}{t_r}Z_r'J_r(Z_r\delta_{m0})\}, \tag{3.21}$$

with $cov(Q_n, L_n) = 0$.

A possible non-degenerated distributed for the estimator $\hat{\lambda}_n$ will depend on the identification conditions that either $\lim_{n\to\infty}\sum_{r=1}^{R}(\frac{m_r-1}{n-R})\omega_r^2 \neq 0$ or Assumption 6.1 holds.

Next, we shall consider the behavior of $\frac{\partial^2 \ln L_{c,n}(\lambda)}{\partial \lambda}$ around neighborhoods of λ_0.

Proposition 4. *Under Assumptions 1-5, and 6.1 or 6.2, for any consistent estimate $\bar{\lambda}_n$ of λ_0,*

$$\frac{m^2}{n}\frac{\partial^2 \ln L_{c,n}(\bar{\lambda}_n)}{\partial \lambda^2} = \frac{m^2}{n}\frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial \lambda^2} + o_P(1). \tag{3.22}$$

The asymptotic distribution of $\hat{\lambda}_n$ is

$$\frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0) = -(\frac{m^2}{n}\frac{\partial^2 \ln L_{c,n}(\bar{\lambda}_n)}{\partial \lambda^2})^{-1}\frac{m}{\sqrt{n}}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda} \xrightarrow{D} N(0, \Omega_\lambda), \tag{3.23}$$

13

where

$$\Omega_\lambda^{-1} = -\lim_{n\to\infty} \frac{m^2}{n} \frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial\lambda^2} = \lim_{n\to\infty} \Big\{ \frac{1}{n\sigma_0^2} \Big[\sum_{r=1}^{R} (\frac{m}{t_r})^2 (Z_r\delta_{m0})' J_r (Z_r\delta_{m0}) - \sum_{r=1}^{R} \frac{m}{t_r} (Z_r\delta_{m0})' J_r Z_r$$

$$\cdot (\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} \frac{m}{t_r} Z_r' J_r (Z_r\delta_{m0}) \Big] + 2(\frac{m-1}{m}) \sum_{r=1}^{R} (\frac{m-1}{n-R}) \omega_r^2 \Big\},$$

with $\omega_r = (\frac{m}{t_r} - \sum_{s=1}^{R} (\frac{m_s-1}{n-R}) \frac{m}{t_s})$.

From (3.23), the CMLE $\hat{\lambda}_n$ converges to λ_0 in distribution at the rate $\frac{\sqrt{n}}{m}$. For the case with small group interactions, the $\{m_r\}$ is bounded, the convergence rate is the usual $\sqrt{n}$-rate. With large group interactions, the convergence rate is scaled down by m and results in a slower rate of convergence, which is equivalent to the $\sqrt{\frac{R}{m}}$-rate of convergence. The explicit expression of $\frac{m^2}{n} \frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial\lambda^2}$ are in (3.13) and (3.14), which equals to the minus of the variance of $\sqrt{\frac{m}{R}} \frac{\partial \ln L_{c,n}(\lambda_0)}{\partial\lambda}$ from (3.17) as expected. The precision of $\hat{\lambda}_n$ depends on the sum of squared residuals of the regression of $\frac{m}{t_r} J_r Z_r \delta_{m0}$ on $J_r Z_r$ with $r = 1, \cdots, R$ and the weighted variations of $\frac{m}{t_r}$. Large variations will result in relatively more precise estimation of λ_0 as summarized in the precision matrix Ω_λ^{-1}.

The CMLE of β_0 is

$$\hat{\beta}_n = \begin{pmatrix} I_{k_1} & 0 \\ 0 & mI_{k_2} \end{pmatrix} (\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} Z_r' J_r Y_r \frac{1}{c_r(\hat{\lambda}_n)}. \tag{3.24}$$

The asymptotic distribution of $\hat{\beta}_n$ is in the following proposition.

Proposition 5. *Under Assumptions 1-5, and 6.1 or 6.2,,*

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} \xrightarrow{D} N(0, \Omega_\beta), \tag{3.25}$$

where

$$\Omega_\beta = \sigma_0^2 \lim_{n\to\infty} (\frac{1}{n} \sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} + \Omega_\lambda \lim_{n\to\infty} h_n h_n'$$

with $h_n = (\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} (\frac{m}{t_r}) Z_r' J_r (Z_r\delta_{m0})$.

This proposition shows that the asymptotic distributions of $\hat{\beta}_{n1}$ and $\hat{\beta}_{n2}$ depend on the asymptotic distribution of $\hat{\lambda}_n$. The h_n in Proposition 5 is the regression coefficient of $\frac{m}{t_r} J_r (Z_r\delta_{m0})$ on $J_r Z_r$ for $r = 1, \cdots, R$. But, the lower rate of convergence $\hat{\lambda}_n$ does not dominate the other random components involving ϵ_r's. As the magnitudes of the explanatory variables $x_{ri,1}$ and $\frac{1}{m_r-1} x_{ri,2}$ of (2.3) are different when m tends to infinity, the CML estimates $\hat{\beta}_{n1}$ and $\hat{\beta}_{n2}$ of the coefficients β_{10} and β_{20} in (2.3) have different rates of convergence. The proper rate of convergence for $\hat{\beta}_{n1}$ is $\sqrt{n}$ and that for $\hat{\beta}_{n2}$ is $\frac{\sqrt{n}}{m}$-rate (equivalently,

14

$\sqrt{\frac{R}{m}}$-rate). The contextual effect parameter is more difficult to be precisely estimated than that of the other regression coefficients.

The results of Propositions 4 and 5 can be combined to derive the joint distribution of $\hat{\lambda}_n$ and $\hat{\beta}_n$.

Proposition 6. *Under Assumptions 1-5, and 6.1 or 6.2,*

$$\begin{pmatrix} \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} \xrightarrow{D} N(0, \Omega_{\lambda,\beta}), \tag{3.26}$$

where $\Omega_{\lambda,\beta} = \lim_{n\to\infty} \left[\frac{1}{\sigma_0^2 n} \sum_{r=1}^{R}(-\frac{m}{t_r} Z_r \delta_{m0}, Z_r)' J_r(-\frac{m}{t_r} Z_r \delta_{m0}, Z_r) + 2(\frac{m-1}{m}) \sum_{r=1}^{R}(\frac{m-1}{n-R})\omega_r^2 e_1 e_1' \right]^{-1}$ *and* e_1 *is the first unit vector.*

4. Instrumental variables estimation

The within equation can be rewritten as

$$y_{ri} - \bar{y}_r = -\lambda_0 \frac{(y_{ri} - \bar{y}_r)}{m_r - 1} + (z_{ri} - \bar{z}_r)\delta_{m0} + (\epsilon_{ri} - \bar{\epsilon}_r), \tag{4.1}$$

which is explicitly

$$y_{ri} - \bar{y}_r = -\lambda_0 \frac{(y_{ri} - \bar{y}_r)}{m_r - 1} + (x_{ri,1} - \bar{x}_{r,1})\beta_{10} - \frac{(x_{ri,2} - \bar{x}_{r,2})}{m_r - 1}\beta_{20} + (\epsilon_{ri} - \bar{\epsilon}_r). \tag{4.1$'$}$$

This equation can be estimated by the method of IV. As the reduced form equation (2.5) of (4.1) implies that

$$E[\frac{1}{m_r - 1}(y_{ri} - \bar{y}_r)] = \frac{1}{t_r}(z_{ri} - \bar{z}_r)\delta_{m0}, \tag{4.2}$$

the best IV vector for the estimation of (4.1) is $(\frac{1}{t_r}(z_{ri} - \bar{z}_r)\delta_{m0}, z_{ri} - \bar{z}_r)$, or equivalently, $(\frac{1}{t_r}(z_{ri} - \bar{z}_r), z_{ri} - \bar{z}_r)$, as motivated by Amemiya (1985) and Lee (2003). The components of the best IV vector will not be perfectly multicollinear if m_r's vary across different groups. As any IV estimates for the coefficients of (4.1)$'$ may have different rates of convergence, we shall explicitly consider the estimation of equation (4.1)$'$.

Let q_{ri} be a IV variable. After rescaling, $\frac{q_{ri}}{m_r - 1}$ can be used as an IV for $-\frac{(y_{ri} - \bar{y}_r)}{m_r - 1}$. Let Q_r be the corresponding m_r-dimensional column vector of q_{ri} for the rth group. The IV estimator of $\theta_0 = (\lambda_0, \beta_{10}', \beta_{20}')'$ is

$$\hat{\theta}_{n,IV} = \left[\sum_{r=1}^{R}(\frac{Q_r}{m_r - 1}, X_{r,1}, -\frac{X_{r,2}}{m_r - 1})' J_r(-\frac{Y_r}{m_r - 1}, X_{r,1}, -\frac{X_{r,2}}{m_r - 1}) \right]^{-1} \sum_{r=1}^{R}(\frac{Q_r}{m_r - 1}, X_{r,1}, -\frac{X_{r,2}}{m_r - 1})' J_r Y_r. \tag{4.3}$$

The following proposition provides the rate of convergence and the asymptotic distribution of the IV estimator. It delivers the best IV matrix for the estimation.

Proposition 7. *Under the assumptions 1-3, 5 and 6.1,*

$$
\begin{pmatrix}
\frac{\sqrt{n}}{m}(\hat{\lambda}_{n,IV} - \lambda_0) \\
\sqrt{n}(\hat{\beta}_{n1,IV} - \beta_{10}) \\
\frac{\sqrt{n}}{m}(\hat{\beta}_{n2,IV} - \beta_{20})
\end{pmatrix}
= \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(-\frac{mY_r}{m_r - 1}, Z_r) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r \epsilon_r
$$
$$
\xrightarrow{D} N(0, \Omega_{IV}),
$$

(4.4)

where

$$
\Omega_{IV} = \sigma_0^2 \lim_{n \to \infty} \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(-\frac{m}{t_r} Z_r \delta_{m0}, Z_r) \right]^{-1}
$$
$$
\cdot \frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(\frac{mQ_r}{m_r - 1}, Z_r) \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(-\frac{m}{t_r} Z_r \delta_{m0}, Z_r) \right]'^{-1},
$$

which is assumed to exist.

The best IV will be $\frac{1}{m_r - 1} Q_r = (-\frac{X_{r,1}\hat{\beta}_{n1}}{m_r - 1 + \hat{\lambda}_n} + \frac{X_{r,2}\hat{\beta}_{n2}}{(m_r - 1)(m_r - 1 + \hat{\lambda}_n)})$, *where* $(\hat{\lambda}_n, \hat{\beta}_{n1}, \hat{\beta}_{n2})$ *can be any initial*

IV consistent estimator. Its asymptotic distribution is

$$
\begin{pmatrix}
\frac{\sqrt{n}}{m}(\hat{\lambda}_{n,BIV} - \lambda_0) \\
\sqrt{n}(\hat{\beta}_{n1,BIV} - \beta_{10}) \\
\frac{\sqrt{n}}{m}(\hat{\beta}_{n2,BIV} - \beta_{20})
\end{pmatrix}
= \left[\frac{1}{n} \sum_{r=1}^{R} (-\frac{m}{t_r} Z_r \delta_{m0}, Z_r)' J_r(-\frac{m}{m_r - 1} Y_r, Z_r) \right]^{-1}
$$
$$
\cdot \frac{1}{\sqrt{n}} \sum_{r=1}^{R} (-\frac{m}{t_r} Z_r \delta_{m0}, Z_r)' J_r \epsilon_r + o_P(1) \xrightarrow{D} N(0, \Omega_{BIV}),
$$

(4.5)

where $\Omega_{BIV} = \sigma_0^2 \lim_{n \to \infty} \left[\frac{1}{n} \sum_{r=1}^{R} (-\frac{m}{t_r} Z_r \delta_{m0}, Z_r)' J_r(-\frac{m}{t_r} Z_r \delta_{m0}, Z_r) \right]^{-1}.$

From this proposition, the IV estimate $\hat{\beta}_{n1,IV}$ converges in probability to β_{10} at the usual $\sqrt{n}$-rate, but, the IV estimates $\hat{\lambda}_{n,IV}$ and $\hat{\beta}_{n2,IV}$ of λ_0 and β_{20} converge at the $\frac{\sqrt{n}}{m}$-rate.[8]

One may compare the asymptotic relative efficiency of the IV estimators with that of the CMLE. From Propositions 6 and 7, it is apparent that $\Omega_{\lambda,\beta} \leq \Omega_{BIV}$. Indeed, in terms of their precision matrices $\Omega_{\lambda,\beta}^{-1} - \Omega_{BIV}^{-1} = 2(\frac{m-1}{m}) \sum_{r=1}^{R} (\frac{m-1}{n-R}) \omega_r^2 e_1 e_1'$. Thus, the main efficient gain of the CMLE is due to the interaction effect on the reduced form disturbances of the within equation (2.5).

As we have pointed out that Assumption 6.1 will not be satisfied if only contextual factors matter but not the regressors X_{r1}, i.e., $\beta_{10} = 0$. In this case, the consistency of the IV estimates of the interaction effects will require a stronger setting and their rates of convergence may also be lower.

---

[8]   For the case of large group interactions, as $m \to \infty$, the term $\frac{m}{(m_r - 1)t_r} X_{r,2}\beta_{20}$ of $Z_r \delta_{m0}$ in (4.4) and (4.5) are relatively small and can be ignored asymptotically. For the large groups case, the best IV can simply be $\frac{Q_r}{m_r - 1} = (-\frac{X_{r,1}\hat{\beta}_{n,1}}{m_r - 1 + \hat{\lambda}_n})$ for the rth group.

Proposition 8. *In the event that $\beta_{10} = 0$ in (4.1)', under the assumptions 1-3, 5 and that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(\frac{m^2}{t_r(m_r - 1)} X_{r2}\beta_{20}, Z_r)$$

exists and is a nonsingular matrix, then

$$\begin{pmatrix} \frac{\sqrt{n}}{m^2}(\hat{\lambda}_{n,IV} - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1,IV} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2,IV} - \beta_{20}) \end{pmatrix} = \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(-\frac{m^2 Y_r}{m_r - 1}, Z_r) \right]^{-1} \frac{1}{\sqrt{n}} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r \epsilon_r \tag{4.6}$$

$$\xrightarrow{D} N(0, \Omega_{IV}),$$

where

$$\Omega_{IV} = \sigma_0^2 \lim_{n\to\infty} \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(\frac{m^2}{t_r(m_r - 1)} X_{r2}\beta_{20}, Z_r) \right]^{-1}$$

$$\cdot \frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(\frac{mQ_r}{m_r - 1}, Z_r) \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{mQ_r}{m_r - 1}, Z_r)' J_r(\frac{m^2}{t_r(m_r - 1)} X_{r2}\beta_{20}, Z_r) \right]'^{-1},$$

which is assumed to exist.

The best IV will be $\frac{1}{m_r-1}Q_r = (-\frac{X_{r,1}\hat{\beta}_{n1}}{m_r-1+\hat{\lambda}_n} + \frac{X_{r,2}\hat{\beta}_{n2}}{(m_r-1)(m_r-1+\hat{\lambda}_n)})$, where $(\hat{\lambda}_n, \hat{\beta}_{n1}, \hat{\beta}_{n2})$ can be any initial IV consistent estimator. Its asymptotic distribution is

$$\begin{pmatrix} \frac{\sqrt{n}}{m^2}(\hat{\lambda}_{n,BIV} - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1,BIV} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2,BIV} - \beta_{20}) \end{pmatrix} = \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{m^2}{t_r(m_r - 1)} X_{r2}\beta_{20}, Z_r)' J_r(-\frac{m^2}{(m_r - 1)} Y_r, Z_r) \right]^{-1}$$

$$\cdot \frac{1}{\sqrt{n}} \sum_{r=1}^{R} (\frac{m^2}{t_r(m_r - 1)} X_{r2}\beta_{20}, Z_r)' J_r \epsilon_r + o_P(1) \tag{4.7}$$

$$\xrightarrow{D} N(0, \Omega_{BIV}),$$

where $\Omega_{BIV} = \sigma_0^2 \lim_{n\to\infty} \left[\frac{1}{n} \sum_{r=1}^{R} (\frac{m^2}{t_r(m_r-1)} X_{r2}\beta_{20}, Z_r)' J_r(\frac{m^2}{t_r(m_r-1)} X_{r2}\beta_{20}, Z_r) \right]^{-1}$.

From the above results, we see that if there are only contextual factors but no other regressors in the spatial model, consistency of an IV estimate of the endogenous interaction effect λ will require that the number of groups R shall be much greater than m^3, and its rate of convergence can only be $\frac{\sqrt{n}}{m^2}$. It is of interest to note that the exogenous interaction effect β_2 has the relatively better $\frac{\sqrt{n}}{m}$-rate of convergence. In the small group interactions case, this is not much a requirement. However, for the large group interactions case, this can be an excessive requirement on a sample. One can easily check that if the constraint $\beta_1 = 0$ is imposed on an IV estimation, the rates of convergence of the constrained IV estimates will not be improved. They have the same rates as those in Proposition 8 even though the limiting variance matrix may be smaller. For the CML approach, even $\beta_{10} = 0$, the CMLE of the endogenous interaction effect λ remains to have

the $\frac{\sqrt{n}}{m}$-rate of convergence. Proposition 6 is valid without Assumption 6.1 as long as Assumption 6.2 is satisfied. This is so in this situation because the information in the reduced form disturbance of the within equation dominates that in the mean reduced form regression function.

5. An OLS (within) approach for models with large group interactions

For the case that $m \to \infty$, one may be interested in the OLS (the conventional within) estimate of β by approximating the within equation (2.3) by the simplified equation that $J_r Y_r \doteq J_r X_{r,1}\beta_{10} + J_r \epsilon_r$ and estimate β_{10} by the OLS:

$$\hat{\beta}_{n1,L} = [\sum_{r=1}^{R} X'_{r,1} J_r X_{r,1}]^{-1} \sum_{r=1}^{R} X'_r J_r Y_r. \tag{5.1}$$

It follows that

$$\hat{\beta}_{n1,L} - \beta_{10} = b_n + [\sum_{r=1}^{R} X'_{r,1} J_r X_{r,1}]^{-1} \sum_{r=1}^{R} c_r(\lambda_0) X'_{r,1} J_r \epsilon_r,$$

where

$$b_n = -\frac{1}{m}[\sum_{r=1}^{R} X'_{r,1} J_r X_{r,1}]^{-1} \sum_{r=1}^{R} \frac{m}{t_r} X'_{r,1} J_r (\lambda_0 X_{r,1}\beta_{10} + X_{r,2}\beta_{20}).$$

The b_n is the bias of $\hat{\beta}_{n1,L}$ for a finite sample with size n, and it has the order $O(\frac{1}{m})$. As $m \to \infty$, the bias tends to zero and $\hat{\beta}_{n1,L}$ is a consistent estimate of β_{10}, but

$$m(\hat{\beta}_{n1,L} - \beta_{10}) = mb_n + O_P(\frac{m}{\sqrt{n}}) \tag{5.2}$$

and the limit of mb_n can be finite:

$$\lim_{n\to\infty} mb_n = - \lim_{n\to\infty} [\frac{1}{n}\sum_{r=1}^{R} X'_{r,1} J_r X_{r,1}]^{-1}\frac{1}{n}\sum_{r=1}^{R}\frac{1}{a_r} X'_{r,1} J_r (\lambda_0 X_{r,1}\beta_{10} + X_{r,2}\beta_{20}).$$

The $m(\hat{\beta}_{n,L} - \beta_0)$ has a degenerate distribution at the limit of mb_n.

One may also be interested in the estimation of the following simplified equation which includes the contextual variables

$$J_r Y_r \doteq J_r X_{r,1}\beta_{10} - J_r \frac{X_{r,2}}{m_r - 1}\beta_{20} + J_r \epsilon_r, \tag{5.3}$$

by the OLS method. The following proposition summarizes the main asymptotic feature of the OLS estimates $(\hat{\beta}_{n1,L}, \hat{\beta}_{n2,L})$ of (β_{10}, β_{20}) from the regression of $J_r Y_r$ on $J_r Z_r$ as in (5.3).

Proposition 9. *Under Assumptions 1-3 and 5, as $m \to \infty$,*

$$\text{plim}_{n\to\infty} \begin{pmatrix} m(\hat{\beta}_{n1,L} - \beta_{10}) \\ (\hat{\beta}_{n2,L} - \beta_{20}) \end{pmatrix} = -\lambda_0 \lim_{n\to\infty} (\frac{1}{n}\sum_{r=1}^{R} Z'_r J_r Z_r)^{-1}\frac{1}{n}\sum_{r=1}^{R}\frac{m}{t_r} Z'_r J_r Z_r \delta_{m0}$$
$$= -\lambda_0 \lim_{n\to\infty} (\frac{1}{n}\sum_{r=1}^{R} Z'_r J_r Z_r)^{-1}\frac{1}{n}\sum_{r=1}^{R}\frac{m}{t_r} Z'_r J_r X_{r,1}\beta_{10}, \tag{5.4}$$

18

which is assumed to exist.

For this Proposition, we see that the OLS (within) estimate of $\hat{\beta}_{n1,L}$ is consistent. But its rate of convergence is of order $O(m)$, which is lower than the $\sqrt{n}$-rate of convergence (under the setting of Assumption 5), and the limiting distribution of $m(\hat{\beta}_{n1,L} - \beta_0)$ is degenerated. The OLS estimate $\hat{\beta}_{n2,L}$ will be inconsistent. In conclusion, if the SAR within model is the proper model, the seemingly minor misspecification of the regression equation (5.3), which ignores the structural spatial interaction, seems to have damaging effects on the conventional within estimates, even though contextual effect has been taken into account.

6. Some Monte Carlo results

For the finite sample performance of the CML and 2SLS estimation, some Monte Carlo experiments are conducted. The sample data are generated with two regressors $x_{ri,1}$ and $x_{ri,2}$ which are both $N(0,1)$. We consider cases that $x_{ri,1}$ and $x_{ri,2}$ can be distinct as well as they can be identical variables. In all cases, $(x_{ri,1}, x_{ri,2})$'s are i.i.d. for all r and i. The ϵ_{ri}'s are i.i.d. $N(0, \sigma^2)$ and are independent of $x_{ri,1}$ and $x_{ri,2}$. For cases with small group interactions, the group sizes are from 2, 3 to 11 for each ten groups in cycle. The average group size is 6.5 for the small interaction case by design. For cases with large group interactions, the group sizes are magnified by 8. That is, for large group interactions, the smallest group size is 16, the largest group size is 88, and the average group size is 52 per group.

For the CML, we have experimented with various numbers of groups R from 50 to 800 for the small group interaction cases, and from 50 to 1600 for the large group interaction cases. The true parameters are $\lambda_0 = 0.5$, $\beta_{10} = 1.0$, $\beta_{20} = 1.0$ and $\sigma_0 = 1.0$. The observations on regressors in Monte Carlo repetitions are independently drawn. The number of Monte Carlo repetitions is 300. The CML estimates are reported in Table 1. 'SG-1' denotes the small group interactions model where x_1 and x_2 are independent. 'SG-2' is the small group interactions model with $x_1 = x_2$. 'LG' denotes the large group interactions model where x_1 and x_2 are independent. We report the empirical mean of the CMLE and its empirical standard deviation is in a bracket.

For the small group interactions model SG-1, the CMLE $\hat{\lambda}_n$ is biased upward when $R = 50$. This bias decreases as R increases. The CMLE's for β_1, β_2 and σ are unbiased for all R. As expected, the CMLE's $\hat{\lambda}_n$ and $\hat{\beta}_{n2}$ have larger standard deviations than those of $\hat{\beta}_{n1}$ and $\hat{\sigma}_n$. When $x_2 = x_1$, the CMLE's of SG-2 have similar properties as those of SG-1, except that these estimates have larger standard deviations. For the model with large group interactions LG, both the interaction effects $\hat{\lambda}_n$ and $\hat{\beta}_{n2}$ have obviously much

19

larger standard deviations than those in SG-1. We note that this occurs even though the total sample size NT in LG is 8 times larger than the corresponding one in SG-1 (or SG-2). The $\hat{\beta}_{n1}$ and $\hat{\sigma}_n$ have smaller standard errors than those of SG-1 because of its overall larger sample size. There is also a bias problem for the CML $\hat{\lambda}_n$ even for $R = 1600$. On the other hand, the bias of $\hat{\beta}_{n2}$ is rather small. It seems that the exogenous interaction effect β_2 can be better estimated than the endogenous effect λ within the large group interactions case. The large standard errors in the estimates of $\hat{\lambda}_n$ and $\hat{\beta}_{n2}$ confirm our theoretical implications on the difficulty on estimating those effects.[9]

Table 2 reports some 2SLS estimates of the models with small group interactions. For an (initial) IV estimation, an IV for $\frac{(y_{ri} - \bar{y}_r)}{m_r - 1}$ in $(4.1)'$ is constructed as follows. First, $(y_{ri} - \bar{y}_r)$ is regressed on $(x_{ri,1} - \bar{x}_{r,1})$ and $\frac{(x_{ri,2} - \bar{x}_{r,2})}{m_r - 1}$. The IV variable is constructed as the predicted value of this regression divided by $(m_r - 1)$ for the rth group. This is justified as if λ_0 were zero in $(4.1)'$. These IV estimates are reported under the columns 'IV' for the models SG-1 and SG-2 in Table 2. With this IV estimate as the initial one, the best IV can be derived as in Proposition 7. The best IV estimates are reported under the columns 'BIV' in Table 2. For SG-1 where x_2 is independent of x_1, both the initial IV and best IV estimates of β_1 and β_2 are nearly unbiased. Both these IV estimates are biased upwards but the biases decrease as R increases. For $R = 800$, the biases are small. The BIV estimates have relatively smaller standard deviations than those of the initial IV estimates. But with larger R's, the efficient gains are rather small. These 2SLS estimates can be compared with the CML estimates of SG-1 in Table 1. The CML estimates are relatively more efficient than those of 2SLS. This is so, especially for the estimates of λ. For $R = 400$ and 800, the standard errors of the BIV estimates of λ are, respectively, 62% and 52% larger than those of the CML estimates.[10] The IV estimates of the SG-2 model, where $x_2 = x_1$, are rather poor when R is not large enough. The biases for the λ estimates still have large biases even when $R = 800$. The standard errors are much larger than those of the SG-1 model. This is so, because the IV or best IV variables in the SG-2 cases are more likely to be highly correlated. A much larger number of groups is needed for the IV estimates to be precise for the SG-2 model.

---

[9] For the LG case with $R = 1600$, this represents a large sample demand on observations. With an average of 52 members per group, this R implies a large sample size of $NT = 83200$.

[10] The estimates of λ in the CML procedure is restricted to lie on $(-1, 1)$, but the IV estimates of λ do not. For $R = 800$, the initial IV and BIV estimates are all lying in the unit interval. The range of the 300 BIV estimates of λ is $(.1687, .9771)$. So the relative comparison of efficiency is valid with or without constraints for the $R = 800$ case. For the CML estimates, there are cases where the CML estimates occur at the boundary value 1 when R are not large. For $R = 400$, with one exception, all the CML estimates are less than one.

Table 1: CML Estimates

	Models	SG-1	SG-2	LG
R				
50	λ	0.5678 (.2876)	0.5748 (.3316)	0.4272 (.6638)
	β_1	1.0105 (.0749)	0.9960 (.1077)	1.0003 (.0234)
	β_2	1.0142 (.3183)	0.9761 (.4883)	0.9810 (.8229)
	σ	1.0046 (.0639)	1.0048 (.0696)	0.9989 (.0203)
100	λ	0.5527 (.2339)	0.5590 (.2745)	0.4914 (.5544)
	β_1	1.0115 (.0551)	0.9996 (.0737)	1.0005 (.0180)
	β_2	1.0052 (.2147)	0.9802 (.3161)	0.9813 (.5974)
	σ	1.0055 (.0467)	1.0052 (.0519)	1.0003 (.0149)
200	λ	0.5366 (.1745)	0.5263 (.2046)	0.5096 (.4298)
	β_1	1.0070 (.0396)	0.9982 (.0522)	1.0013 (.0136)
	β_2	1.0098 (.1460)	0.9806 (.2188)	0.9861 (.4232)
	σ	1.0049 (.0330)	1.0015 (.0377)	1.0003 (.0116)
400	λ	0.5216 (.1239)	0.5213 (.1535)	0.5274 (.3507)
	β_1	1.0059 (.0277)	0.9994 (.0399)	1.0009 (.0097)
	β_2	1.0055 (.1019)	0.9869 (.1613)	0.9874 (.2966)
	σ	1.0031 (.0249)	1.0016 (.0291)	1.0004 (.0087)
800	λ	0.5118 (.0890)	0.5143 (.1119)	0.5740 (.2724)
	β_1	1.0031 (.0211)	1.0004 (.0301)	1.0012 (.0077)
	β_2	1.0013 (.0697)	0.9958 (.1141)	0.9921 (.2130)
	σ	1.0019 (.0175)	1.0012 (.0208)	1.0014 (.0066)
1600	λ	—	—	0.5697 (.2533)
	β_1	—	—	1.0014 (.0060)
	β_2	—	—	0.9981 (.1486)
	σ	—	—	1.0014 (.0055)

Remarks:

1) SG-1: a model with small group interactions with an average of 6.5 members per group; x_1 and x_2 are independent. Corresponding to R=50, 100, 200, 400, and 800, the total sample observations are, respectively, NT=325, 650, 1,300, 2,600, and 5,200.

2) SG-2: the model with small group interactions in 1) but $x_2 = x_1$.

3) LG: a model with large group interactions with an average of 52 members per group; x_1 and x_2 are independent. Corresponding to R=50, 100, 200, 400, 800, and 1,600, the total sample observations are, respectively, NT=2,600, 5,200, 10,400, 20,800, 41,600, and 83,200.

4) R: the number of groups.

21

Table 2: IV Estimates

Models		SG-1		SG-2	
R		IV	BIV	IV	BIV
50	λ	0.6478 (.7360)	0.5996 (.5990)	—*	—*
	β_1	1.0204 (.1238)	1.0134 (.1089)	—	—
	β_2	1.0309 (.3867)	1.0159 (.3598)	—	—
100	λ	0.6305 (.4695)	0.6034 (.4256)	—**	—*
	β_1	1.0228 (.0838)	1.0186 (.0783)	—	—
	β_2	1.0275 (.2615)	1.0180 (.2482)	—	—
200	λ	0.5768 (.3282)	0.5641 (.3071)	—*	0.8711 (1.3514)
	β_1	1.0129 (.0595)	1.0109 (.0567)	1.1368 (1.5448)	1.0233 (.1261)
	β_2	1.0197 (.1717)	1.0154 (.1666)	0.4488 (6.4365)	0.8931 (.4181)
400	λ	0.5374 (.2066)	0.5322 (.2009)	0.6727 (.9459)	0.6280 (.7161)
	β_1	1.0082 (.0370)	1.0073 (.0362)	1.0114 (.0961)	1.0074 (.0812)
	β_2	1.0098 (.1146)	1.0079 (.1128)	0.9550 (.2366)	0.9623 (.1983)
800	λ	0.5211 (.1382)	0.5187 (.1352)	0.5854 (.4746)	0.5782 (.4599)
	β_1	1.0045 (.0274)	1.0041 (.0270)	1.0066 (.0584)	1.0059 (.0574)
	β_2	1.0045 (.0805)	1.0036 (.0798)	0.9849 (.1236)	0.9855 (.1224)
1600	λ	—	—	0.5470 (.3188)	0.5417 (.3082)
	β_1	—	—	1.0026 (.0377)	1.0021 (.0371)
	β_2	—	—	0.9902 (.0814)	0.9907 (.0804)
3200	λ	—	—	0.5300 (.2193)	0.5289 (.2177)
	β_1	—	—	1.0025 (.0261)	1.0024 (.0260)
	β_2	—	—	0.9960 (.0568)	0.9960 (.0566)

Remark:

1) * indicates that the empirical means of the estimated λ are greater than one and corresponding standard errors are large (greater than 10).
2) The ** indicates that the empirical mean of the estimated λ is negative.

7. Conclusion

In this paper, we consider identification and estimation of SAR models with group interactions. The identification and estimation of the structural group interaction effects are clouded by group unobservables, as unobservables may cause spurious effects which may be confused with group interaction effects. In our analysis, we allow both endogenous and exogenous group interactions in the presence of fixed group effects in the SAR model.

With fixed group effect specification, the group unobservables are allowed to correlated with included explanatory variables. We show that for the familiar group interaction specification with equal weights among group members, the SAR model can be decomposed into a within and a between group equations. The fixed group effects have been eliminated in the within equation. The between equation provides sufficient statistics for the fixed effects, but does not provide information for the identification of the structural interaction effects. With the fixed group effects specification, the identification and estimation of the structural interaction effects may only be revealed by the within equation. The identification of the structural interaction effects may be possible only when there are various group sizes. The presence of the endogenous interaction effect reduces the within group variations via both the responses of regressors and disturbances. When groups have the same size, one can not make inference about the interaction effects as there are no distinctions among groups due to interactions. When groups have different sizes, inference may be possible because of different degrees of interactions. We provide characterized conditions for the identification and estimation of the interaction effects.

We consider the estimation of the within equation by the method of CML. Consistency of estimates requires that the number of groups in the sample is much larger than the average size of groups. For cases with large group interactions, identification can be weak in the sense the CML estimates of the interaction effects converge in distribution at a low rate. The introduction of contextual factors in addition to valid individual regressors does not create additional identification and estimation problems for the endogenous iteraction effect. But, for the case with large group interactions, the estimate of the contextual (exogenous) effect will also have the same low rate of convergence as that of the endogenous effect. We compare also the efficiency gain of the CMLE over IV estimators. The corresponding CML and IV estimates have the same stochastic rates of convergence. The IV estimates are, in general, relatively inefficient to the CML estimates. When the regressors in the model consist of only contextual factors, the IV estimate of the endogenous

23

interaction effect may become worse as it will have a much lower rate of convergence than that of the CML estimates. The rates of the CML estimates do not change. This is so, because the IV approach does not take into account the correlation information of the reduced form disturbances, while the CML does. Our Monte Carlo results confirm such implications.

If the fixed group effects are uncorrelated with the included regressors, the between group equation can provide valuable information in addition to the within group equation for the estimation of the structural interaction effects. The implication of group random effect on estimation has not been considered in this paper. It shall be investigated in another occasion.[11]

---

[11] Strong identification and relatively fast rate of convergence of estimators are possible when either fixed group effects are uncorrelated with included exogenous variables or valid IVs exist, and there are regressive effects which are not completely contextual. The model $(2.1)'$ with a random effect specification, i.e., α_r's are random and are uncorrelated with x's, have been used in a recent empirical study on housing demand in Ioannides and Zabel (2003).

Appendix: A List of Notations and Proofs

A List of Often Used Notations in the Text:

m_r member size of the rth group

R total number of groups

$n = \sum_{r=1}^{R} m_r$ total sample size

$m = \frac{n}{R}$ the mean group size

l_{m_r} the m_r-dimensional column vector of ones

$J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l'_{m_r}$

$z_{ri} = (x_{ri,1}, -\frac{m}{m_r - 1} x_{ri,2})$

$c_r(\lambda) = (\frac{m_r - 1}{m_r - 1 + \lambda})$

$d_r(\lambda) = (\frac{m_r - 1 + \lambda}{m_r - 1 + \lambda_0})$

$t_r = (m_r - 1 + \lambda_0)$

$\theta = (\lambda, \beta_1', \beta_2')'$

$\delta_m = (\beta_1', \beta_2'/m)'$

$\omega_r = (\frac{m}{t_r} - \sum_{s=1}^{R} (\frac{m_s - 1}{n - R}) \frac{m}{t_s})$

Proof of Proposition 1. The $\sigma_n^{*2}(\lambda)$ in (3.8) is related to $\hat{\sigma}_n^2(\lambda)$ in (3.4) by replacing $Y_r' J_r Y_r$ and $J_r Y_r$ by its expected values. The within equation (2.3) has $\frac{1}{c_r(\lambda_0)} J_r Y_r = J_r Z_r \delta_{m0} + J_r \epsilon_r$. Therefore, (3.4) can be rewritten as

$$
\begin{aligned}
\hat{\sigma}_n^2(\lambda) &= \frac{1}{n - R} \Big\{ \sum_{r=1}^{R} d_r^2(\lambda)(Z_r \delta_{m0} + \epsilon_r)' J_r (Z_r \delta_{m0} + \epsilon_r) \\
&\quad - \sum_{r=1}^{R} d_r(\lambda)(Z_r \delta_{m0} + \epsilon_r)' J_r Z_r (\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} d_r(\lambda) Z_r' J_r (Z_r \delta_{m0} + \epsilon_r) \Big\} \\
&= \sigma_n^{*2}(\lambda) + \frac{2}{n - R} [q_{n1}(\lambda) - q_{n2}(\lambda) + l_n(\lambda)],
\end{aligned}
\tag{A1}
$$

where

$$
q_{n1}(\lambda) = \frac{1}{2} \left(\sum_{r=1}^{R} d_r^2(\lambda) \epsilon_r' J_r \epsilon_r - \sigma_0^2 \sum_{r=1}^{R} d_r^2(\lambda)(m_r - 1) \right),
\tag{A2}
$$

$$
q_{n2}(\lambda) = \frac{1}{2} \sum_{r=1}^{R} d_r(\lambda) \epsilon_r' J_r Z_r (\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} d_r(\lambda) Z_r' J_r \epsilon_r,
\tag{A3}
$$

and

$$l_n(\lambda) = \sum_{r=1}^{R} d_r^2(\lambda)(Z_r\delta_{m0})'J_r\epsilon_r - \sum_{r=1}^{R} d_r(\lambda)(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R} Z_r'J_rZ_r)^{-1} \sum_{r=1}^{R} d_r(\lambda)Z_r'J_r\epsilon_r$$

$$= (\lambda - \lambda_0)\Big\{ \sum_{r=1}^{R} \frac{d_r(\lambda)}{t_r}(Z_r\delta_{m0})'J_r\epsilon_r - \sum_{r=1}^{R} \frac{1}{t_r}(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R} Z_r'J_rZ_r)^{-1} \sum_{r=1}^{R} d_r(\lambda)Z_r'J_r\epsilon_r \Big\}. \tag{A4}$$

Their derivatives are

$$\frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda} = \frac{2}{n-R}\Big\{ \frac{(\lambda - \lambda_0)}{m^2}\Big[\sum_{r=1}^{R} (\frac{m}{t_r})^2 (Z_r\delta_{m0})'J_r(Z_r\delta_{m0}) - \sum_{r=1}^{R} \frac{m}{t_r}(Z_r\delta_{m0})'J_rZ_r$$

$$\cdot (\sum_{r=1}^{R} Z_r'J_rZ_r)^{-1} \sum_{r=1}^{R} \frac{m}{t_r} Z_r'J_r(Z_r\delta_{m0})\Big] + \sigma_0^2 \sum_{r=1}^{R} c_r(\lambda_0)d_r(\lambda) \Big\} \tag{A5}$$

from (3.8), and

$$\frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda} = \frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda} + \frac{2}{n-R}\Big(\frac{\partial q_{n1}(\lambda)}{\partial \lambda} - \frac{\partial q_{n2}(\lambda)}{\partial \lambda} + \frac{\partial l_n(\lambda)}{\partial \lambda}\Big), \tag{A6}$$

from (A1), where

$$\frac{\partial q_{n1}(\lambda)}{\partial \lambda} = \sum_{r=1}^{R} \frac{d_r(\lambda)}{t_r}(\epsilon_r'J_r\epsilon_r - \sigma_0^2(m_r - 1)), \tag{A7}$$

$$\frac{\partial q_{n2}(\lambda)}{\partial \lambda} = \sum_{r=1}^{R} \frac{1}{t_r}\epsilon_r'J_rZ_r(\sum_{r=1}^{R} Z_r'J_rZ_r)^{-1} \sum_{r=1}^{R} d_r(\lambda)Z_r'J_r\epsilon_r \tag{A8}$$

from (A2) and (A3), and

$$\frac{\partial l_n(\lambda)}{\partial \lambda} = \Big\{ \sum_{r=1}^{R} \frac{d_r(\lambda)}{t_r}(Z_r\delta_{m0})'J_r\epsilon_r - \sum_{r=1}^{R} \frac{1}{t_r}(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R} Z_r'J_rZ_r)^{-1} \sum_{r=1}^{R} d_r(\lambda)Z_r'J_r\epsilon_r \Big\}$$

$$+ (\lambda - \lambda_0)\Big[\sum_{r=1}^{R} \frac{1}{t_r^2}(Z_r\delta_{m0})'J_r\epsilon_r - \sum_{r=1}^{R} \frac{1}{t_r}\epsilon_r'J_rZ_r(\sum_{r=1}^{R} Z_r'J_rZ_r)^{-1} \sum_{r=1}^{R} \frac{1}{t_r}Z_r'J_r(Z_r\delta_{m0})\Big] \tag{A9}$$

from (A4).

The stochastic orders of the terms (A1)-(A9) can be derived from central limit theorems. Those orders shall be uniform in $\lambda \in \Lambda$. Under Assumptions 1 and 2, the Liapounov CLT gives that $\frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r'J_r\epsilon_r = O_P(1)$. Under Assumption 3, as $\{\frac{m}{t_r}\}$ is bounded, Liapounov's CLT implies also that $\frac{1}{\sqrt{n}}\sum_{r=1}^{R} \frac{m}{t_r}Z_r'J_r\epsilon_r = O_P(1)$ and $\frac{1}{\sqrt{n}}\sum_{r=1}^{R} (\frac{m}{t_r})^2 Z_r'J_r\epsilon_r = O_P(1)$. For uniform order, we note that

$$\sum_{r=1}^{R} d_r(\lambda)Z_r'J_r\epsilon_r = \sum_{r=1}^{R} Z_r'J_r\epsilon_r + \frac{(\lambda - \lambda_0)}{m} \sum_{r=1}^{R} \frac{m}{t_r}Z_r'J_r\epsilon_r,$$

and

$$\sum_{r=1}^{R} \frac{d_r(\lambda)}{t_r}Z_r'J_r\epsilon_r = \frac{1}{m}\sum_{r=1}^{R} \frac{m}{t_r}Z_r'J_r\epsilon_r + \frac{(\lambda - \lambda_0)}{m^2} \sum_{r=1}^{R} (\frac{m}{t_r})^2 Z_r'J_r\epsilon_r.$$

26

As λ appears linearly and is in the compact set Λ, uniform orders for those terms follow. Therefore, under Assumptions 1-4, $\frac{q_{n2}(\lambda)}{n} = O_p(\frac{1}{n})$, $\frac{l_n(\lambda)}{n} = O_P(\frac{1}{m\sqrt{n}})$, $\frac{1}{n}\frac{\partial q_{n2}(\lambda)}{\partial \lambda} = O_P(\frac{1}{mn})$, and $\frac{1}{n}\frac{\partial l_n(\lambda)}{\partial \lambda} = O_P(\frac{1}{m\sqrt{n}})$, uniformly on Λ. For quadratic forms, the CLT of Kelejian and Prucha (2001) shows that $\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)) = O_P(1)$ and $\frac{1}{\sqrt{n}}\sum_{r=1}^{R}(\frac{m}{t_r})^2(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)) = O_P(1)$. For uniform order, we note

$$\sum_{r=1}^{R} d_r^2(\lambda)(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)) = \sum_{r=1}^{R}(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)) + 2\frac{(\lambda - \lambda_0)}{m}\sum_{r=1}^{R}\frac{m}{t_r}(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1))$$
$$+ (\frac{\lambda - \lambda_0}{m})^2 \sum_{r=1}^{R}(\frac{m}{t_r})^2(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)),$$

and

$$\sum_{r=1}^{R}\frac{d_r(\lambda)}{t_r}(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)) = \frac{1}{m}\sum_{r=1}^{R}\frac{m}{t_r}(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)) + \frac{(\lambda - \lambda_0)}{m^2}\sum_{r=1}^{R}(\frac{m}{t_r})^2(\epsilon_r' J_r \epsilon_r - \sigma_0^2(m_r - 1)).$$

Therefore, $\frac{q_{n1}(\lambda)}{n} = O_P(\frac{1}{\sqrt{n}})$ and $\frac{1}{n}\frac{\partial q_{n1}(\lambda)}{\partial \lambda} = O_P(\frac{1}{m\sqrt{n}})$.

With the uniform stochastic orders for these terms, it follows from (A1) and (A6) that

$$\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda) = O_p(\frac{1}{\sqrt{n}}), \quad \text{and} \quad \frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda} - \frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda} = O_p(\frac{1}{m\sqrt{n}}), \tag{A10}$$

uniformly in $\lambda \in \Lambda$. From (3.8) and (A5), $\sigma_n^{*2}(\lambda) = O(1)$ and $\frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda} = O(\frac{1}{m})$ uniformly on Λ. With respect to leading terms, (3.8) and (A6) have

$$\sigma_n^{*2}(\lambda) = \frac{\sigma_0^2}{n - R}\sum_{r=1}^{R}(m_r - 1)d_r^2(\lambda) + O(\frac{1}{m^2}), \quad \text{and} \quad \frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda} = \frac{2\sigma_0^2}{n - R}\sum_{r=1}^{R}c_r(\lambda_0)d_r(\lambda) + O(\frac{1}{m^2}). \tag{A11}$$

under Assumption 5. We note that (3.8) implies that

$$\sigma_n^{*2}(\lambda) \geq \frac{\sigma_0^2}{n - R}\sum_{r=1}^{R}(m_r - 1)d_r^2(\lambda) > c_b^2 \sigma_0^2, \tag{A12}$$

where $c_b > 0$ is a lower bound of $d_r(\lambda)$ on Λ for all r. The (A12) shows that $\sigma_n^{*2}(\lambda)$ is uniformly bounded away from zero on Λ.[12] As $\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda) = o_P(1)$ from (A10), $\hat{\sigma}_n^2(\lambda)$ is also uniformly bounded away from zero on Λ in probability. By the mean value theorem, (3.5) and (3.7) imply

$$\frac{m^2}{n}\Big\{[\ln L_{c,n}(\lambda) - \ln L_{c,n}(\lambda_0)] - [Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)]\Big\}$$
$$= -(\frac{n - R}{2})(\frac{m^2}{n})\Big[\frac{\partial \ln \hat{\sigma}_n^2(\bar{\lambda})}{\partial \lambda} - \frac{\partial \ln \sigma_n^{*2}(\bar{\lambda})}{\partial \lambda}\Big](\lambda - \lambda_0) \tag{A13}$$
$$= -(\frac{n - R}{2})(\frac{m^2}{n})\frac{(\lambda - \lambda_0)}{\hat{\sigma}_n^2(\bar{\lambda})\sigma_n^{*2}(\bar{\lambda})}\Big[\sigma_n^{*2}(\bar{\lambda})\Big(\frac{\partial \hat{\sigma}_n^2(\bar{\lambda})}{\partial \lambda} - \frac{\partial \sigma_n^{*2}(\bar{\lambda})}{\partial \lambda}\Big) - \frac{\partial \sigma_n^{*2}(\bar{\lambda})}{\partial \lambda}(\hat{\sigma}_n^2(\bar{\lambda}) - \sigma_n^{*2}(\bar{\lambda}))\Big],$$

----

[12] The $d_r(\lambda)$ is increasing in $\lambda \in (-1, 1)$. When $\{m_r\}$ is bounded as n tends to infinity, $\frac{m_L - 2}{m_U - 1 + \lambda_0} \leq d_r(\lambda) \leq \frac{m_U}{m_L - 1 + \lambda_0}$ where m_L and m_U are, respectively, the lower and upper bounds of m_r. If m tends to infinity, $d_r(\lambda)$ converges to 1 uniformly in λ.

where $\bar{\lambda}$ lies between λ and λ_0. As $\frac{(n-R)m^2}{n}\sigma_n^{*2}(\lambda)(\frac{\partial\hat{\sigma}_n^2(\lambda)}{\partial\lambda} - \frac{\partial\sigma_n^{*2}(\lambda)}{\partial\lambda}) = O_P(\frac{m}{\sqrt{n}})$ and $\frac{(n-R)m^2}{n}\frac{\partial\sigma_n^{*2}(\lambda)}{\partial\lambda}(\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda)) = O_P(\frac{m}{\sqrt{n}})$ from (A10) and (A12), it follows that

$$\frac{m^2}{n}\{[\ln L_{c,n}(\lambda) - \ln L_{c,n}(\lambda_0)] - [Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)]\} = O_P(\frac{m}{\sqrt{n}}) = o_P(1)$$

uniformly on Λ under Assumption 5 because $\frac{m}{\sqrt{n}} = \sqrt{\frac{m}{R}} \to 0$. **Q.E.D.**

Proof of Proposition 2. Under Assumption 6.1, (3.12) shows that, for any open neighborhood $N_\epsilon(\lambda_0)$ of λ_0,

$$\liminf_{n\to\infty} \min_{\lambda\in\bar{N}_\epsilon(\lambda_0)} \frac{m^2}{n}(n-R)(\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda)) > 0.$$

That implies $\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))$ in (3.11) satisfies the identification uniqueness condition because $D_n(\lambda) \le 0$ for all λ from (3.10)′.

In the case of Assumption 6.2, the identification uniqueness condition for $D_n(\lambda)$ will be satisfied if $D_n(\lambda)$ is uniformly equicontinuous on Λ. At λ_0, $D_n(\lambda_0) = 0$. From (3.9) and (3.10), the derivative of $D_n(\lambda)$ is

$$\frac{\partial D_n(\lambda)}{\partial\lambda} = \frac{m^2}{n}\left\{\sum_{r=1}^R c_r(\lambda) - \frac{n-R}{\sum_{r=1}^R(m_r-1)d_r^2(\lambda)}\sum_{r=1}^R c_r(\lambda_0)d_r(\lambda)\right\},$$

which has $\frac{\partial D_n(\lambda_0)}{\partial\lambda} = 0$. The second order derivative is

$$\frac{\partial^2 D_n(\lambda)}{\partial\lambda^2} = -\frac{1}{R}\sum_{r=1}^R(\frac{m}{m_r-1})c_r^2(\lambda) - \frac{1}{\sum_{r=1}^R(\frac{m_r-1}{n-R})d_r^2(\lambda)}\frac{1}{R}\sum_{r=1}^R\frac{m}{t_r}c_r(\lambda_0)$$

$$+ \frac{2}{[\sum_{r=1}^R(\frac{m_r-1}{n-R})d_r^2(\lambda)]^2}\sum_{r=1}^R(\frac{m_r-1}{n-R})\frac{m}{t_r}d_r(\lambda)\cdot\frac{1}{R}\sum_{r=1}^R c_r(\lambda_0)d_r(\lambda).$$

As $c_r(\lambda)$ and $d_r(\lambda)$ are uniformly bounded on Λ for all r and $d_r(\lambda)$ is uniformly bounded away from zero, $\frac{\partial^2 D_n(\lambda)}{\partial\lambda^2}$ are uniformly bounded on Λ. By the mean value theorem, $D_n(\lambda) = \frac{\partial^2 D_n(\bar{\lambda}_n)}{\partial\lambda^2}(\lambda - \lambda_0)^2$ and the uniform equicontinuity of $D_n(\lambda)$ follows.

As $\sigma_n^{*2}(\lambda) \ge \sigma_n^2(\lambda)$ from (3.12), $\sigma_n^{*2}(\lambda)$ is uniformly bounded away from zero in probability on Λ as $n \to \infty$. Furthermore, $\sigma_n^{*2}(\lambda) \ge \sigma_n^2(\lambda)$ and (3.11) together imply that

$$\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)) \le D_r(\lambda) - \frac{m^2}{n}(\frac{n-R}{2})(\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda))/\sigma_n^{*2}(\lambda).$$

The identification uniqueness condition holds for $Q_{c,n}(\lambda)$.

The uniform convergence in probability of the concentrated log likelihood function in Proposition 1 and the identification uniqueness condition above for $Q_{c,n}(\lambda)$ imply the consistency of $\hat{\lambda}_n$ (White 1994).

Q.E.D.

28

Proof of Proposition 3. The order of $\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda}$ will depend on the orders of the various items in (3.16). Under Assumptions 1-3, Liapounov's CLT gives

(i) $l_{n1} = \frac{\sqrt{n}}{m}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}(Z_r\delta_{m0})'J_r\epsilon_r = O_P(\frac{\sqrt{n}}{m})$,

(ii) $l_{n2} = \frac{\sqrt{n}}{m}\frac{1}{n}\sum_{r=1}^{R}\frac{m}{t_r}(Z_r\delta_{m0})'J_rZ_r)[\frac{1}{n}\sum_{r=1}^{R}Z_r'J_rZ_r]^{-1}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}Z_r'J_r\epsilon_r = O(\frac{\sqrt{n}}{m})$,

(iii) $q_{n2} = \frac{1}{m\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}\epsilon_r'J_rZ_r[\frac{1}{n}\sum_{r=1}^{R}Z_r'J_rZ_r]^{-1}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}Z_r'J_r\epsilon_r = O_P(\frac{1}{m})$.

Note that at λ_0,

$$\hat{\sigma}_n^2(\lambda_0) = \frac{1}{n-R}\{\sum_{r=1}^{R}\epsilon_r'J_r\epsilon_r - \sum_{r=1}^{R}\epsilon_r'J_rZ_r[\sum_{r=1}^{R}Z_r'J_rZ_r]^{-1}\sum_{r=1}^{R}Z_r'J_r\epsilon_r\} = \frac{1}{n-R}\sum_{r=1}^{R}\epsilon_r'J_r\epsilon_r + O_P(\frac{1}{n}),$$

from (3.4). As $\frac{1}{n-R}\sum_{r=1}^{R}\epsilon_r'J_r\epsilon_r - \sigma_0^2 = \frac{\sqrt{n}}{n-R}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}(\epsilon_r'J_r\epsilon_r - \sigma_0^2(m_r-1)) = O_P(\frac{1}{\sqrt{n}})$, it follows that $\hat{\sigma}_n^2(\lambda_0) \xrightarrow{p} \sigma_0^2$. For quadratic forms, the CLT of Kelejian and Prucha (2001) is applicable. As $\frac{\sqrt{n}}{n-R} = (\frac{m}{m-1})\frac{1}{\sqrt{n}}$, one has

(iv) $\sqrt{n}(\hat{\sigma}_n^2(\lambda_0) - \sigma_0^2) = (\frac{m}{m-1})\frac{1}{\sqrt{n}}\sum_{r=1}^{R}[\epsilon_r'J_r\epsilon_r - (m_r-1)\sigma_0^2] + O(\frac{1}{\sqrt{n}})$, and

(v) $q_{n1} - \sum_{r=1}^{R}\frac{\sigma_0^2}{c_r(\lambda_0)} = \frac{\sqrt{n}}{m}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}[\epsilon_r'J_r\epsilon_r - (m_r-1)\sigma_0^2] = O_P(\frac{\sqrt{n}}{m})$.

From (3.16), as

$$\frac{m}{\sqrt{n}}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda} = \frac{1}{\hat{\sigma}_n^2(\lambda_0)}\{\frac{m}{\sqrt{n}}(\hat{\sigma}_n^2(\lambda_0) - \sigma_0^2)\sum_{r=1}^{R}\frac{1}{c_r(\lambda_0)} - \frac{m}{\sqrt{n}}(q_{n1} - \sum_{r=1}^{r}\frac{\sigma_0^2}{c_r(\lambda_0)}) - \frac{m}{\sqrt{n}}(l_{n1} - l_{n2}) + \frac{m}{\sqrt{n}}q_{n2}\},$$

the stochastic orders above imply (3.17).

Under normality of ϵ_r's, $var(\epsilon_r'J_r\epsilon_r) = 2(m_r-1)\sigma_0^4$. The variance of Q_n in (3.20) follows. For the variance of L_n, we note that (3.19) can be rewritten as

$$L_n = l_{n1} - l_{n2} = \delta_{m0}'(\frac{1}{t_1}Z_1', \cdots, \frac{1}{t_R}Z_R')'\mathcal{J}_n[I_n - \mathcal{J}_n\mathcal{Z}_n(\mathcal{Z}_n'\mathcal{J}_n\mathcal{Z}_n)^{-1}\mathcal{Z}_n'\mathcal{J}_n]\mathcal{E}_n, \tag{A14}$$

where $\mathcal{J}_n = diag\{J_1, \cdots, J_R\}$ is a block diagonal idempotent matrix, and $\mathcal{Z}_n = (Z_1', \cdots, Z_R')'$. With this expression for L_n, it is immediate that

$$var(\frac{m}{\sqrt{n}}L_n) = \frac{\sigma_0^2}{n}\delta_{m0}'(\frac{m}{t_1}Z_1', \cdots, \frac{m}{t_R}Z_R')'\mathcal{J}_n[I_n - \mathcal{J}_n\mathcal{Z}_n(\mathcal{Z}_n'\mathcal{J}_n\mathcal{Z}_n)^{-1}\mathcal{Z}_n'\mathcal{J}_n]\mathcal{J}_n(\frac{m}{t_1}Z_1, \cdots, \frac{m}{t_R}Z_R)\delta_{m0},$$

and $cov(Q_n, L_n) = 0$ as the third moment of ϵ_{ri} is zero. **Q.E.D.**

Proof of Proposition 4. From (3.6),

$$\frac{m^2}{n}\frac{\partial^2 \ln L_{c,n}(\lambda)}{\partial \lambda^2} = -\frac{1}{R}\sum_{r=1}^{R}\frac{m}{(m_r-1)}c_r^2(\lambda) - \frac{m(m-1)}{2}\left[\frac{1}{\hat{\sigma}_n^2(\lambda)}\frac{\partial^2\hat{\sigma}_n^2(\lambda)}{\partial\lambda^2} - \frac{1}{\hat{\sigma}_n^4(\lambda)}(\frac{\partial\hat{\sigma}_n^2(\lambda)}{\partial\lambda})^2\right]. \tag{A15}$$

29

From (A10) and (3.8),

$$\hat{\sigma}_n^2(\lambda) = \sigma_n^{*2}(\lambda) + O_P(\frac{1}{\sqrt{n}}) = \frac{\sigma_0^2}{n-R}\sum_{r=1}^{R}(m_r - 1)d_r^2(\lambda) + (\frac{\lambda - \lambda_0}{m})^2 \cdot O(1) + O_P(\frac{1}{\sqrt{n}}),$$

and, from (A10) and (A5),

$$\frac{\partial\hat{\sigma}_n^2(\lambda)}{\partial\lambda} = \frac{\partial\sigma_n^{*2}(\lambda)}{\partial\lambda} + O_P(\frac{1}{m\sqrt{n}}) = \frac{2\sigma_0^2}{n-R}\sum_{r=1}^{R}c_r(\lambda_0)d_r(\lambda) + (\frac{\lambda - \lambda_0}{m^2}) \cdot O(1) + O_P(\frac{1}{m\sqrt{n}}),$$

uniformly on Λ. Thus, for any consistent estimate $\bar{\lambda}_n$ of λ_0, $\hat{\sigma}_n^2(\bar{\lambda}_n) \xrightarrow{p} \sigma_0^2$ and

$$m\frac{\partial\hat{\sigma}_n^2(\bar{\lambda}_n)}{\partial\lambda} - 2\sigma_0^2(\frac{m}{m-1})\frac{1}{R}\sum_{r=1}^{R}c_r(\lambda_0) = o_P(1).$$

From (3.4), as

$$\frac{\partial\hat{\sigma}_n^2(\lambda)}{\partial\lambda} = \frac{2}{n-R}\left\{\sum_{r=1}^{R}\frac{Y_r'J_rY_r}{(m_r-1)c_r(\lambda)} - \sum_{r=1}^{R}\frac{Y_r'J_rZ_r}{(m_r-1)}(\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}\frac{Z_r'J_rY_r}{c_r(\lambda)}\right\},$$

which is linear in λ, hence

$$m(m-1)\frac{\partial^2\hat{\sigma}_n^2(\lambda)}{\partial\lambda^2} = 2\left\{\frac{1}{R}\sum_{r=1}^{R}\frac{m}{(m_r-1)^2}Y_r'J_rY_r - \frac{1}{R}\sum_{r=1}^{R}\frac{Y_r'J_rZ_r}{(m_r-1)}(\frac{1}{n}\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\frac{1}{R}\sum_{r=1}^{R}\frac{Z_r'J_rY_r}{(m_r-1)})\right\},$$

which does not depend on λ. The probability limit of the latter can be derived by using

$$\frac{1}{R}\sum_{r=1}^{R}\frac{1}{(m_r-1)}Z_r'J_rY_r = \frac{1}{n}\sum_{r=1}^{R}\frac{m}{t_r}Z_r'J_rZ_r\delta_{m0} + \frac{1}{n}\sum_{r=1}^{R}\frac{m}{t_r}Z_r'J_r\epsilon_r = \frac{1}{n}\sum_{r=1}^{R}\frac{m}{t_r}Z_r'J_rZ_r\delta_{m0} + o_P(1),$$

and

$$\frac{1}{R}\sum_{r=1}^{R}\frac{m}{(m_r-1)^2}Y_r'J_rY_r = \frac{1}{n}\sum_{r=1}^{R}(\frac{m}{t_r})^2(Z_r\delta_{m0})'J_r(Z_r\delta_{m0}) + \frac{1}{n}\sum_{r=1}^{R}(\frac{m}{t_r})^2\epsilon_r'J_r\epsilon_r + \frac{2}{n}\sum_{r=1}^{R}(\frac{m}{t_r})^2(Z_r\delta_{m0})'J_r\epsilon_r$$

$$= \frac{1}{n}\sum_{r=1}^{R}(\frac{m}{t_r})^2(Z_r\delta_{m0})'J_r(Z_r\delta_{m0}) + \frac{\sigma_0^2}{n}\sum_{r=1}^{R}(\frac{m}{t_r})^2(m_r-1) + o_P(1).$$

By combining the above results into (A15),

$$\frac{m^2}{n}\frac{\partial^2\ln L_{c,n}(\bar{\lambda}_n)}{\partial\lambda^2}$$

$$= -\frac{1}{n\sigma_0^2}\{\sum_{r=1}^{R}(\frac{m}{t_r})^2(Z_r\delta_{m0})'J_r(Z_r\delta_{m0}) - \sum_{r=1}^{R}\frac{m}{t_r}(Z_r\delta_{m0})'J_rZ_r(\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}(\sum_{r=1}^{R}\frac{m}{t_r}Z_r'J_rZ_r\delta_{m0})\}$$

$$- \frac{2}{R}\sum_{r=1}^{R}\frac{m}{t_r}c_r(\lambda_0) + 2\frac{m}{m-1}(\frac{1}{R}\sum_{r=1}^{R}c_r(\lambda_0))^2 + o_P(1)$$

$$= \frac{m^2}{n}\frac{\partial^2Q_{c,n}(\lambda_0)}{\partial\lambda^2} + o_P(1),$$

as in (3.13) and (3.14). This proves (3.22).

The asymptotic normal distribution of $\frac{m}{\sqrt{n}}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda}$ follows from Proposition 3 with the CLT for linear and quadratic forms in Kelejian and Prucha (2001). **Q.E.D.**

Proof of Proposition 5. The (3.3) and (3.24) imply that

$$\begin{pmatrix} I_{k_1} & 0 \\ 0 & \frac{I_{k_2}}{m} \end{pmatrix}\hat{\beta}_n = (\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} Z_r' J_r (Z_r \delta_{m0} + \epsilon_r) d_r(\hat{\lambda}_n).$$

Hence,

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} = (\frac{1}{n}\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1}[\frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r' J_r \epsilon_r + \frac{1}{n}\sum_{r=1}^{R}(Z_r' J_r Z_r \delta_{m0} + Z_r' J_r \epsilon_r)\frac{m}{t_r} \cdot \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0)], \tag{A16}$$

because $d_r(\hat{\lambda}_n) - 1 = (\hat{\lambda}_n - \lambda_0)/t_r$.

The term $\frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r' J_r \epsilon_r$ is uncorrelated with $\frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0)$. The latter depends on $\frac{m}{\sqrt{n}}(Q_n - L_n)$ in (3.17) of Proposition 3. Under normality, because the third order moment of ϵ is zero, $\frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r' J_r \epsilon_r$ is uncorrelated with $\frac{m}{\sqrt{n}}Q_n$. It is uncorrelated with $\frac{m}{\sqrt{n}}L_n$. That becomes apparent from the expression of L_n in (A14) and that $\sum_{r=1}^{R} Z_r' J_r \epsilon_r = \mathcal{Z}_n' \mathcal{J}_n \mathcal{E}_n$. The detailed asymptotic variance Ω_β of $\hat{\beta}_n$ follows from (A16) and Proposition 4. **Q.E.D.**

Proof of Proposition 6. The (3.23) of Proposition 4 and (A16) in the proof of Proposition 5 imply that

$$A_n \begin{pmatrix} \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r' J_r \epsilon_r \\ \frac{m}{\sqrt{n}}(Q_n - L_n)/\sigma_0^2 \end{pmatrix} + o_P(1),$$

where

$$A_n = \begin{pmatrix} -\frac{1}{n}\sum_{r=1}^{R}\frac{m}{t_r} Z_r' J_r Z_r \delta_{m0} & \frac{1}{n}\sum_{r=1}^{R} Z_r' J_r Z_r \\ \Omega_\lambda^{-1} & 0 \end{pmatrix}.$$

Therefore,

$$\begin{pmatrix} \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} = A_n^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r' J_r \epsilon_r \\ \frac{m}{\sqrt{n}}(Q_n - L_n)/\sigma_0^2 \end{pmatrix} + o_P(1) \xrightarrow{D} N(0, \Omega_{\lambda,\beta}),$$

where

$$\Omega_{\lambda,\beta} = A_n^{-1} \begin{pmatrix} \frac{\sigma_0^2}{n}\sum_{r=1}^{R} Z_r' J_r Z_r & 0 \\ 0 & \Omega_\lambda^{-1} \end{pmatrix} A_n'^{-1}$$

$$= \begin{pmatrix} \frac{1}{n\sigma_0^2}\sum_{r=1}^{R}(\frac{m}{t_r})^2 (Z_r \delta_{m0})' J_r (Z_r \delta_{m0}) + 2(\frac{m-1}{m})\sum_{r=1}^{R}(\frac{m-1}{n-R})\omega_r^2 & -\frac{1}{n\sigma_0^2}\sum_{r=1}^{R}\frac{m}{t_r}(Z_r \delta_{m0})' J_r Z_r \\ -\frac{1}{n\sigma_0^2}\sum_{r=1}^{R}\frac{m}{t_r} Z_r' J_r (Z_r \delta_{m0}) & \frac{1}{n\sigma_0^2}\sum_{r=1}^{R} Z_r' J_r Z_r \end{pmatrix}^{-1}.$$

Q.E.D.

Proof of Proposition 7. To capture the possible different rates of convergence of components of $\hat{\theta}_{n,IV}$, consider the normalizing matrix

$$\Lambda_n = \begin{pmatrix} m & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & m \end{pmatrix}.$$

With this normalizing matrix, the components of

$$\frac{1}{n}\Lambda_n \sum_{r=1}^{R}(\frac{Q_r}{m_r-1}, X_{r,1}, -\frac{X_{r,2}}{m_r-1})'J_r(-\frac{Y_r}{m_r-1}, X_{r,1}, -\frac{X_{r,2}}{m_r-1})\Lambda_n = \frac{1}{n}\sum_{r=1}^{R}(\frac{mQ_r}{m_r-1}, Z_r)'J_r(-\frac{mY_r}{m_r-1}, Z_r)$$

have a similar order of magnitude. As the reduced form equation is $J_rY_r = (\frac{m_r-1}{t_r})J_r(Z_r\delta_{m0}+\epsilon_r)$ and, by a law of large numbers, $\frac{1}{n}\sum_{r=1}^{R}(\frac{m_rQ_r}{m_r-1}, Z_r)'J_r(\frac{m_r-1}{t_r})\epsilon_r = o_P(1)$,

$$\frac{1}{n}\sum_{r=1}^{R}(\frac{mQ_r}{m_r-1}, Z_r)'J_r(-\frac{mY_r}{m_r-1}, Z_r) = \frac{1}{n}\sum_{r=1}^{R}(\frac{mQ_r}{m_r-1}, Z_r)'J_r(-\frac{m}{t_r}Z_r\delta_{m0}, Z_r) + o_P(1),$$

which converges to a well-defined nonsingular matrix. Similarly,

$$\frac{1}{\sqrt{n}}\Lambda_n \sum_{r=1}^{R}(\frac{Q_r}{m_r-1}, X_{r,1}, -\frac{X_{r,2}}{m_r-1})'J_r\epsilon_r = \frac{1}{\sqrt{n}}\sum_{r=1}^{R}(\frac{mQ_r}{m_r-1}, Z_r)'J_r\epsilon_r$$

may converge in distribution to a non-degenerate normal distribution. It follows that

$$\sqrt{n}\Lambda_n^{-1}(\hat{\theta}_{n,IV} - \theta_0) = [\frac{1}{n}\sum_{r=1}^{R}(\frac{mQ_r}{m_r-1}, Z_r)'J_r(-\frac{mY_r}{m_r-1}, Z_r)]^{-1}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}(\frac{mQ_r}{m_r-1}, Z_r)'J_r\epsilon_r$$

$$\xrightarrow{D} N(0, \Omega_{IV}),$$

as in (4.4)

The distribution of the initial IV estimates in (4.4) does not have effect on the asymptotic distribution of the best IV estimator. This is so because

$$\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1+\hat{\lambda}_n)}(X_{r,1}\hat{\beta}_{n1})'J_r\epsilon_r$$

$$= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}(X_{r,1}\beta_{10})'J_r\epsilon_r - \frac{1}{n}\sum_{r=1}^{R}\frac{m}{(m_r-1+\bar{\lambda}_n)^2}(X_{r,1}\bar{\beta}_{n1})'J_r\epsilon_r \cdot \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0)$$

$$+ \frac{1}{n}\sum_{r=1}^{R}\frac{m}{(m_r-1+\bar{\lambda}_n)}\epsilon_r'J_rX_{r,1} \cdot \sqrt{n}(\hat{\beta}_{n1} - \beta_{10})$$

$$= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}(X_{r,1}\beta_{10})'J_r\epsilon_r + o_P(1),$$

32

and

$$\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1)(m_r-1+\hat{\lambda}_n)}(X_{r,2}\hat{\beta}_{n2})'J_r\epsilon_r$$

$$=\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1)t_r}(X_{r,2}\beta_{20})'J_r\epsilon_r-\frac{1}{n}\sum_{r=1}^{R}\frac{m^2}{(m_r-1)(m_r-1+\bar{\lambda}_n)^2}(X_{r,2}\bar{\beta}_{n2})'J_r\epsilon_r\cdot\frac{\sqrt{n}}{m}(\hat{\lambda}_n-\lambda_0)$$

$$+\frac{1}{n}\sum_{r=1}^{R}\frac{m^2}{(m_r-1)(m_r-1+\bar{\lambda}_n)}\epsilon_r'J_rX_{r,2}\cdot\frac{\sqrt{n}}{m}(\hat{\beta}_{n2}-\beta_{20})$$

$$=\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1)t_r}(X_{r,2}\beta_{20})'J_r\epsilon_r+o_P(1),$$

hence,

$$\sqrt{n}\Lambda_n^{-1}(\hat{\theta}_{n,BIV}-\theta_0)$$
$$=[\frac{1}{n}\sum_{r=1}^{R}(-\frac{m}{t_r}Z_r\delta_{m0},Z_r)'J_r(-\frac{m}{m_r-1}Y_r,Z_r)]^{-1}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}(-\frac{m}{t_r}Z_r\delta_{m0},Z_r)'J_r\epsilon_r+o_P(1)\xrightarrow{D}N(0,\Omega_{BIV}),$$

as in (4.5).

The limiting variances Ω_{BIV} and Ω_{IV} can be compared. Denote $\mathcal{S}_n=((-\frac{m}{t_1}Z_1\delta_{m0})',\cdots,(-\frac{m}{t_R}Z_R\delta_{m0})')'$
and $\mathcal{Q}_n=(\frac{m}{m_1-1}Q_1',\cdots,\frac{m}{m_R-1}Q_R')'$. It follows that their limiting precision matrices are

$$\Omega_{IV}^{-1}=\frac{1}{\sigma_0^2}\lim_{n\to\infty}\frac{1}{n}(\mathcal{S}_n,\mathcal{Z}_n)'\mathcal{J}_n(\mathcal{Q}_n,\mathcal{Z}_n)[(\mathcal{Q}_n,\mathcal{Z}_n)'\mathcal{J}_n(\mathcal{Q}_n,\mathcal{Z}_n)]^{-1}(\mathcal{Q}_n,\mathcal{Z}_n)'\mathcal{J}_n(\mathcal{S}_n,\mathcal{Z}_n),$$

and $\Omega_{BIV}^{-1}=\frac{1}{\sigma_0^2}\lim_{n\to\infty}(\mathcal{S}_n,\mathcal{Z}_n)'\mathcal{J}_n(\mathcal{S}_n,\mathcal{Z}_n)$. The $\Omega_{BIV}^{-1}\geq\Omega_{IV}^{-1}$ follows from the generalized Schwartz inequality. This justifies the best selection of IVs. **Q.E.D.**

Proof of Proposition 8. When $\beta_{10}=0$, the within equation is $J_rY_r=J_r(-\frac{\lambda_0}{m_r-1}Y_r-\frac{1}{m_r-1}X_{r,2}\beta_{20}+\epsilon_r)$
and its reduced form equation is $J_rY_r=(\frac{m_r-1}{t_r})J_r(-\frac{X_{r,2}\beta_{20}}{m_r-1}+\epsilon_r)$. Therefore,

$$-\frac{m^2}{m_r-1}J_rY_r=\frac{m^2}{t_r(m_r-1)}J_rX_{r,2}\beta_{20}-\frac{m^2}{t_r}J_r\epsilon_r.$$

Denote $\Lambda_n=\begin{pmatrix}m^2&0&0\\0&1&0\\0&0&m\end{pmatrix}$ and $\Gamma_n=\begin{pmatrix}m&0&0\\0&1&0\\0&0&m\end{pmatrix}$. It follows that

$$\frac{1}{n}\Gamma_n\sum_{r=1}^{R}(\frac{Q_r}{m_r-1},X_{r,1},-\frac{X_{r,2}}{m_r-1})'J_r(-\frac{Y_r}{m_r-1},X_{r,1},-\frac{X_{r,2}}{m_r-1})\Lambda_n$$

$$=\frac{1}{n}\sum_{r=1}^{R}(\frac{m}{m_r-1}Q_r,Z_r)'J_r(\frac{m^2}{t_r(m_r-1)}X_{r,2}\beta_{20},Z_r)+\frac{1}{n}\sum_{r=1}^{R}(\frac{m}{m_r-1}Q_r,Z_r)'J_r(-\frac{m^2}{t_r}\epsilon_r,0,0)$$

$$=\frac{1}{n}\sum_{r=1}^{R}(\frac{m}{m_r-1}Q_r,Z_r)'J_r(\frac{m^2}{t_r(m_r-1)}X_{r,2}\beta_{20},Z_r)+o_P(1),$$

because $\frac{1}{n}\sum_{r=1}^{R}(\frac{m}{m_r-1}Q_r, Z_r)'J_r\frac{m^2}{t_r}\epsilon_r = O_P(\frac{m}{\sqrt{n}}) = o_P(1)$ under Assumption 5. Therefore,

$$\sqrt{n}\Lambda_n^{-1}(\hat{\theta}_{n,IV} - \theta_0)$$

$$= \left[\frac{1}{n}\Gamma_n\sum_{r=1}^{R}(\frac{Q_r}{m_r-1}, Z_r)'J_r(-\frac{Y_r}{m_r-1}, Z_r)\Lambda_n\right]^{-1}\frac{\Gamma_n}{\sqrt{n}}\sum_{r=1}^{R}(\frac{Q_r}{m_r-1}, Z_r)'J_r\epsilon_r$$

$$= \left[\frac{1}{n}\sum_{r=1}^{R}(\frac{mQ_r}{m_r-1}, Z_r)'J_r(\frac{m^2}{t_r(m_r-1)}X_{r,2}\beta_{20}, Z_r) + o_P(1)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}(\frac{m}{m_r-1}Q_r, Z_r)'J_r\epsilon_r$$

$$\xrightarrow{D} N(0, \Omega_{IV}),$$

where Ω_{IV} is in the proposition.

The distribution of the initial IV estimates does not have effect on the asymptotic distribution of the best IV estimator in this case, because

$$\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1+\hat{\lambda}_n)}(X_{r,1}\hat{\beta}_{n1})'J_r\epsilon_r$$

$$= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}(X_{r,1}\beta_{10})'J_r\epsilon_r + \frac{1}{n}\sum_{r=1}^{R}\frac{m}{(m_r-1+\bar{\lambda}_n)}\epsilon_r'J_rX_{r,1}\cdot\sqrt{n}(\hat{\beta}_{n1} - \beta_{10})$$

$$- \frac{1}{n}\sum_{r=1}^{R}\frac{m^2}{(m_r-1+\bar{\lambda}_n)^2}(X_{r,1}\bar{\beta}_{n1})'J_r\epsilon_r\cdot\frac{\sqrt{n}}{m^2}(\hat{\lambda}_n - \lambda_0)$$

$$= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{t_r}(X_{r,1}\beta_{10})'J_r\epsilon_r + o_P(1),$$

and

$$\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1)(m_r-1+\hat{\lambda}_n)}(X_{r,2}\hat{\beta}_{n2})'J_r\epsilon_r$$

$$= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1)t_r}(X_{r,2}\beta_{20})'J_r\epsilon_r + \frac{1}{n}\sum_{r=1}^{R}\frac{m^2}{(m_r-1)(m_r-1+\bar{\lambda}_n)}\epsilon_r'J_rX_{r,2}\cdot\frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20})$$

$$- \frac{1}{n}\sum_{r=1}^{R}\frac{m^3}{(m_r-1)(m_r-1+\bar{\lambda}_n)^2}(X_{r,2}\bar{\beta}_{n2})'J_r\epsilon_r\cdot\frac{\sqrt{n}}{m^2}(\hat{\lambda}_n - \lambda_0)$$

$$= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{(m_r-1)t_r}(X_{r,2}\beta_{20})'J_r\epsilon_r + o_P(1).$$

The result for the best IV estimator follows.   **Q.E.D.**

Proof of Proposition 9. The OLS estimates $(\hat{\beta}_{n1,L}, \hat{\beta}_{n2,L})$ of (β_{10}, β_{20}) would correspond to a rescaled OLS $\hat{\delta}_{m,L}$ of δ_{m0} with $\hat{\delta}_{m,L} = (\hat{\beta}'_{n1,L}, \frac{\hat{\beta}'_{n2,L}}{m})'$, where $\hat{\delta}_{m,L} - \delta_{m0} = (\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}Z_r'J_rY_r$. It follows that

$$\hat{\delta}_{m,L} - \delta_{m0} = (\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}[\sum_{r=1}^{R}(c_r(\lambda_0) - 1)Z_r'J_rZ_r\delta_{m0}] + (\sum_{r=1}^{R}Z_r'J_rZ_r)^{-1}\sum_{r=1}^{R}c_r(\lambda_0)Z_r'J_r\epsilon_r.$$

This implies that $m(\hat{\delta}_{m,L} - \delta_{m0}) = -\lambda_0(\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \sum_{r=1}^{R} \frac{m}{t_r} Z_r' J_r Z_r \delta_{m0} + O_P(\frac{m}{\sqrt{n}})$. Hence, as $m \to \infty$ and $\frac{\sqrt{n}}{m} \to \infty$,

$$\text{plim}_{n\to\infty}(m(\hat{\beta}_{n1,L} - \beta_{10})', (\hat{\beta}_{n2,L} - \beta_{20})')' = -\lambda_0 \lim_{n\to\infty} (\frac{1}{n}\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \frac{1}{n}\sum_{r=1}^{R} \frac{m}{t_r} Z_r' J_r Z_r \delta_{m0}$$

$$= -\lambda_0 \lim_{n\to\infty} (\frac{1}{n}\sum_{r=1}^{R} Z_r' J_r Z_r)^{-1} \frac{1}{n}\sum_{r=1}^{R} \frac{m}{t_r} Z_r' J_r X_{r,1} \beta_{10}.$$

Q.E.D.

References

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

Betrand, M., E.F.P. Luttmer and S. Mullainathan (2000), "Net-work effects and welfare cultures." *Quarterly Journal of Economics* 115, 1019-1055.

Brock, W.A. and S.N. Durlauf (2001), "Interactions-base models." In *Handbook of Econometrics*, J.J. Heckman and E.E. Leamer (eds). North-Holland: Amsterdam, 3297-3380.

Case, A.C. (1991). 'Spatial patterns in household demands." *Econometrica* 59: 953-965.

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

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

Durlauf, S., and H.P. Young H. (2001). *Social Dynamics*, MIT Press: Cambridge, MA.

Glaeser. E.L., B. Sacerdote, and J.A. Scheinkman (1996). "Crime and social interactions." *The Quarterly Journal of Economics* 111: 507-548.

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

Hoxby, C.M. (2000). "The effects of class size on student achievement: new evidence from population variation", *Quarterly Journal of Economics* 115, 1239-1285.

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge University Press, Cambridge.

Ioannides, Y. and J.E. Zabel (2003). "Neighbourhood effects and housing demand." *Journal of Applied Econometrics* 18: 563-584.

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

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

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

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

Moffitt, R.A. (2001). "Policy interventions, low-level equilibria, and social interactions." In *Social Dynamics*, Durlauf, S., Peyton Young, H. (eds). MIT Press: Cambridge, MA, 45-82.

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

Rees, D.I., J.S. Zax and J. Herries (2003). "Interdependence in worker productivity", *Journal of Applied Econometrics* 18, 585-604.

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

Theil, H. (1971). *Principles of Econometrics*. New York: John Wiley and Sons.

Weinberg, B., P.B. Reagan, and J.J. Yankow (2003). "Do neighborhoods affect work behavior? Evidence from the NLSY 79." Forthcoming in *Journal of Labor Economics*.

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