Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 140 (2007) 333–374
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Identification and estimation of econometric models with group interactions, contextual factors and fixed effects

Lung-fei Lee*

Department of Economics, The Ohio State University, 410 Arps Hall, 1945 North High St., Columbus, OH 43210-1172, USA

Available online 22 August 2006

Abstract

This paper considers identification and estimation of structural interaction effects in a social interaction model. 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.

JEL classification: C13; C21; R15

Keywords: Social interactions; Group structure; Fixed effects; Spatial autoregression; Conditional maximum likelihood

1. Introduction

The social interaction model considered in this paper has an important link with spatial econometric models. A typical spatial autoregressive (SAR) model is specified as

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

*Tel.: +1 614 292 5508; fax: +1 614 292 4192.
E-mail address: lflee@econ.ohio-state.edu.

0304-4076/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.jeconom.2006.07.001

where $\mathscr{E}_n$ 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 $\mathscr{W}_n$ is a specified constant spatial weights matrix with a zero diagonal (Cliff and Ord, 1973).

In urban and regional economic studies, a region, a district, or a county can be a spatial unit and its neighboring units in $\mathscr{W}_n$ are defined in terms of a certain physical or economic distance. The equation in (1.1) implies elements of Y_n shall be simultaneously determined given x's and the disturbances ε's as

$$Y_n = (I_n - \lambda_0 \mathscr{W}_n)^{-1} X_n \beta_0 + (I_n - \lambda_0 \mathscr{W}_n)^{-1} \mathscr{E}_n. \tag{1.2}$$

This model also has 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 within a group may interact with each other but are usually not interrelated with members in other groups. Suppose there are R groups and there are m_r units in the rth group. A typical group interactions model has $\mathscr{W}_n$ being a block diagonal matrix, i.e.,

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

where l_{m_r} is the m_r-dimensional column vector of ones, and I_{m_r} is the m_r-dimensional identity matrix. Empirical studies on group interactions are in Case (1991, 1992) in consumption pattern and technology adoption, Betrand et al. (2000) on welfare cultures, and Sacrerdote (2001) and Hanushek et al. (2003) on student achievement, among others.

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. As a group effect model with interactions may have policy implications, researchers have pointed out various important specification issues of a group effect model beyond those in a typical spatial model. 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. Manski (1993) and Brock and Durlauf (2001) separate interaction effects 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. Manski (1993) has considered a group effect model where 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 might be linearly depended on observed exogenous variables of a group in the model—the 'reflection' problem. The reflection problem refers to the difficulty to distinguish between behavioral and contextual factors. Another main concern is on possible unobservables in a group, as unobservables in a group may have direct effect on observed outcomes. The unobservables may also cause the total disturbances to be correlated across individuals in a group. 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.[1]

In this paper, we consider the SAR model as a group effect model with both endogenous group interaction and contextual factors and allow the existence of correlated unobservables as a fixed effect in a group. There is a close similarity of this model 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 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 $\mathrm{E}(Y_n|X_n) = (I_n - \lambda_0 \mathscr{W}_n)^{-1} X_n$ in (1.2) and another is due to the correlated distributions of disturbances in $(I_n - \lambda_0 \mathscr{W}_n)^{-1}\mathscr{E}_n$. 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.[2] 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.[3]

2. A SAR model with group interactions and fixed effects

The 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 place more structure on the model (1.1) with fixed effects α_r and an additional explanatory component $W_r X_{r2}$ for contextual effects:

$$Y_r = \lambda_0 W_r Y_r + X_{r1}\beta_{10} + W_r X_{r2}\beta_{20} + l_{m_r}\alpha_r + \varepsilon_r, \quad r = 1, \ldots, R, \tag{2.1}$$

where Y_r, X_{r1} and X_{r2} 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,

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

[1]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 preferences 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.

[2]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 the model cannot be identified.

[3]In Lee (2004), the asymptotic distribution of the (quasi-) maximum likelihood estimator is considered for the SAR model, which includes the group effect model but without exogenous interactions and group unobservables. The identification problem is less an issue for such a SAR model. But for the SAR model with a large group interaction structure, there are cases where the estimation of λ might also have a low rate of convergence. But its rate is still higher than that obtained for the model with fixed effects.

with $i = 1, \ldots, m_r$ and $r = 1, \ldots, R$, 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-dimensional row vectors of exogenous variables, and ε_{ri}'s are i.i.d. $(0, \sigma_0^2)$. In the model, the outcome of the unit i may be influenced by the outcomes of other units, which effect is captured by the parameter λ_0.[4] The α_r represents the unobservables of the rth group. As those unobservables may correlate with exogenous variables, they are treated as fixed effects. The vectors of all exogenous variables x_{ri}'s must vary across individuals in a 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 $\sum_{j=1, j\neq i}^{m_r} x_{ri,2}$, i.e., $W_r X_{r2}$, allow social interaction effect through observed neighborhood characteristics. Neighborhood characteristics have often been used in empirical studies of neighborhood effects, e.g.,Weinberg et al. (2004), in a regression setting. One may wonder whether this additional contextual factor would compound the identification of the structural interaction effect λ_0 in (2.1), especially, when $x_{ri,2}$ is identical to $x_{ri,1}$. We note that m_r is the number of members in the rth group.

It is revealing to decompose this equation into two parts:

$$(1-\lambda_0)\bar{y}_r = \bar{x}_{r1}\beta_{10} + \bar{x}_{r2}\beta_{20} + \alpha_r + \bar{\varepsilon}_r, \quad r = 1, \ldots, R \tag{2.2}$$

and

$$\left(1 + \frac{\lambda_0}{m_r - 1}\right)(y_{ri} - \bar{y}_r) = (x_{ri,1} - \bar{x}_{r1})\beta_{10} - \frac{1}{m_r - 1}(x_{ri,2} - \bar{x}_{r2})\beta_{20} + (\varepsilon_{ri} - \bar{\varepsilon}_r),$$
$$i = 1, \ldots, m_r, \quad r = 1, \ldots, R, \tag{2.3}$$

where $\bar{y}_r = \frac{1}{m_r}\sum_{i=1}^{m_r} y_{ri}$, $\bar{x}_{r1} = \frac{1}{m_r}\sum_{i=1}^{m_r} x_{ri,1}$ and $\bar{x}_{r2} = \frac{1}{m_r}\sum_{i=1}^{m_r} x_{ri,2}$ are means for the rth group.[5] The Eq. (2.2) may be called a 'between' equation and that in (2.3) is a 'within' equation as they have 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:

$$\bar{y}_r = \bar{x}_{r1}\frac{\beta_{10}}{(1-\lambda_0)} + \bar{x}_{r2}\frac{\beta_{20}}{(1-\lambda_0)} + \frac{\alpha_r}{(1-\lambda_0)} + \frac{\bar{\varepsilon}_r}{(1-\lambda_0)}, \quad r = 1, \ldots, R \tag{2.4}$$

and

$$(y_{ri} - \bar{y}_r) = (x_{ri,1} - \bar{x}_{r1})\frac{(m_r - 1)\beta_{10}}{(m_r - 1 + \lambda_0)} - (x_{ri,2} - \bar{x}_{r2})\frac{\beta_{20}}{(m_r - 1 + \lambda_0)} + \frac{(m_r - 1)}{(m_r - 1 + \lambda_0)}(\varepsilon_{ri} - \bar{\varepsilon}_r), \tag{2.5}$$

with $i = 1, \ldots, m_r$; $r = 1, \ldots, R$. Suppose that the interaction effect λ is positive. For the average group outcome $\bar{y}_r$, positive group interaction raises the regression effects of $\bar{x}_{r1}$ and $\bar{x}_{r2}$ on $\bar{y}_r$ in (2.4) by the factor $\frac{1}{(1-\lambda_0)}$. It also raises the variance of $\bar{y}_r$ (with the same $\bar{x}_r$) across different groups by a factor of $\frac{1}{(1-\lambda_0)^2}$.[6]

[4] It is meaningful to use the average $\frac{1}{m_r - 1}\sum_{j=1, j\neq i}^{m_r} y_{rj}$ for the ith unit by excluding y_{ri} instead of the average $\frac{1}{m_r}\sum_{j=1}^{m_r} y_{rj}$ because the outcome y_{ri} of i shall not be influenced by his own outcome.

[5] The detailed derivation of (2.2) and (2.3) from (2.1)$'$ can be found in the mathematical Appendix.

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

However, in the presence of the unobservables in α_r, the group interaction effect λ cannot be identified through the between equation (2.4) as it cannot 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 freedom 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 deviation 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 deviation 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 cannot be identifiable from the within equation. This is apparent as only the functions $\frac{(m-1)\beta_{10}}{(m-1+\lambda_0)}$, $\frac{\beta_{20}}{(m-1+\lambda_0)}$, and $\frac{(m-1)\sigma_0^2}{(m-1+\lambda_0)}$ may ever be identifiable from (2.5).[7]

The possible identification of the structural parameters in the group interaction model with fixed effects relies on various group sizes in a sample.[8] This identification can be weak, especially, when there are large group interactions. When m_r are all large, the factors $\left(1+\frac{\lambda_0}{m_r-1}\right)$ may be close to one and λ_0 may not be easily estimated from (2.3). In subsequent sections, we characterize possible consistent estimation of 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 $(y_{ri}-\bar{y}_r) = (x_{ri,1}-\bar{x}_{r1})\beta_{10} + (\varepsilon_{ri}-\bar{\varepsilon}_r)$ 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 find that this OLS estimator of β_{10} can be consistent but its rate of convergence differs from the usual $\sqrt{n}$-rate of convergence and its limiting distribution after rate normalization is degenerated. These features of the OLS estimate are rather surprising.

3. CML estimation

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

For analytical convenience, denote $z_{ri} = \left(x_{ri,1}, -\frac{m}{m_r-1}x_{ri,2}\right)$ where $m = \frac{1}{R}\sum_{r=1}^{R} m_r$ is the mean size of groups. Let $\delta_m = (\beta_1', \beta_2'/m)'$. The total sample size is $n = \sum_{r=1}^{R} m_r = Rm$. To simplify repeated notations, let $m_r(\lambda) = m_r - 1 + \lambda$. Thus, $m_r(0) = m_r - 1$. Under the

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

[8] 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.

assumption that ε's are normally distributed, the likelihood function for the within equation (2.3) as derived in the Appendix is

$$L_{w,n}(\theta) = \prod_{r=1}^{R} \frac{\sqrt{m_r}}{(2\pi)^{\frac{m_r(0)}{2}}} \left(\frac{m_r(\lambda)}{\sigma m_r(0)}\right)^{m_r(0)} \exp\left\{-\frac{1}{2\sigma^2}\left(\frac{m_r(\lambda)}{m_r(0)} Y_r^* - Z_r^*\delta_m\right)' \times \left(\frac{m_r(\lambda)}{m_r(0)} Y_r^* - Z_r^*\delta_m\right)\right\}, \tag{3.1}$$

where $\theta = (\lambda, \beta', \sigma^2)'$, $\beta = (\beta_1', \beta_2')'$, $Z_r^* = J_r Z_r$, and $Y_r^* = J_r Y_r$ with $J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l_{m_r}'$. The Z_r^* and Y_r^* are, respectively, matrices of elements of z_{ri} and y_{ri} deviated from their means, i.e., $Z_r^* = (z_{ri}^{*'}, \ldots, z_{r,m_r}^{*'})'$ and $Y_r^* = (y_{r1}^*, \ldots, y_{r,m_r}^*)'$, where $z_{ri}^* = z_{ri} - \bar{z}_r$ and $y_{ri}^* = y_{ri} - \bar{y}_r$. For any vector or matrix A_r conformable with J_r, denote $A_r^* = J_r A_r$. A list of often used notations is collected in the Appendix for convenience of reference.

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 $\bar{y}_r$, $r = 1, \ldots, R$. The whole sample y_{ri}'s can be transformed one-to-one into the observations $y_{ri} - \bar{y}_r$ with $i = 1, \ldots, m_r$ and $r = 1, \ldots, R$ for the within equation, and $\bar{y}_r$ with $r = 1, \ldots, R$ 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

$$\ln L_{w,n}(\theta) = c + \sum_{r=1}^{R} m_r(0) \ln(m_r(\lambda)) - \frac{(R(m-1))}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{r=1}^{R} \left(\frac{m_r(\lambda)}{m_r(0)} Y_r^* - Z_r^*\delta_m\right)' \left(\frac{m_r(\lambda)}{m_r(0)} Y_r^* - Z_r^*\delta_m\right), \tag{3.2}$$

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,

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

and

$$\hat{\sigma}_n^2(\lambda) = \frac{1}{R(m-1)} \left\{ \sum_{r=1}^{R} \left(\frac{m_r(\lambda)}{m_r(0)}\right)^2 Y_r^{*'} Y_r^* - \sum_{r=1}^{R} \left(\frac{m_r(\lambda)}{m_r(0)}\right) Y_r^{*'} Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^*\right)^{-1} \sum_{r=1}^{R} Z_r^{*'} Y_r^* \left(\frac{m_r(\lambda)}{m_r(0)}\right) \right\}. \tag{3.4}$$

The concentrated log likelihood function (3.2) at λ is[9]

$$\ln L_{c,n}(\lambda) = c_1 + \sum_{r=1}^{R} m_r(0)\ln(m_r(\lambda)) - \frac{(R(m-1))}{2}\ln \hat{\sigma}_n^2(\lambda). \tag{3.5}$$

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. Assume $m_r = a_r m \geqslant 2$ 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 $a_L \leqslant a_r \leqslant a_U$ for all $r = 1, 2, \ldots$ with $a_L m \geqslant 2$.

Assumption 3. The x_{ri}'s are assumed to be bounded (in absolute value) constants. The limit of $(1/n)\sum_{r=1}^{R} Z_r^{*'} Z_r^*$ exists and is a non-singular matrix.

Assumption 4. The parameter space Λ of λ is a connected compact subset with λ_0 in its interior satisfying the property that $(1 - \inf_{r=1,2,\ldots,m_r}) < \lambda$ for all $\lambda \in \Lambda$.

Assumption 1 is the basic distributional assumption of the disturbances in the model. The m in Assumption 2 is regarded as the empirical mean of m_r's. The positive lower bound a_L of the factors a_r and their upper bound a_U describe the possible scenario that if m_r's are large, they are uniformly so.

In a fixed effect model, statistical analysis is conditional on the unobservables α_r. One may regard that analysis shall also be conditional on x_{ri}. Thus, it seems natural to assume x_{ri}'s being non-stochastic in Assumption 3. The boundedness of x_{ri} is a convenient and simple assumption.[10] The components of z_{ri}^* consist of $x_{ri,1}^*$ and $-\frac{m}{m_r(0)} x_{ri,2}^*$. As $\frac{m}{m_r(0)} x_{ri,2}$ shall have approximately a similar magnitude as $x_{ri,2}$ even when m goes to infinity, it is reasonable to assume that the limit of $\frac{1}{n}\sum_{r=1}^{*} Z_r^{*'} Z_r^*$ has a finite limiting matrix.[11] For the case that $x_{ri,2} = x_{ri,1}$ for all r and i, as long as m_r's have variation, there would not be multicollinearity of $\frac{m}{m_r(0)} x_{ri,2}$ and $x_{ri,1}$. Thus, the finite limiting matrix of $\frac{1}{n}\sum_{r=1}^{R} Z_r^{*'} Z_r^*$ would be non-singular. This assumption has possible implications on the rates of convergence for

[9]The derivative of (3.5) with λ is $\frac{\partial \ln L_{c,n}(\lambda)}{\partial \lambda} = \sum_{r=1}^{R}\left(\frac{m_r(0)}{m_r(\lambda)}\right) - \frac{(R(m-1))}{2}\frac{\partial \ln \hat{\sigma}_n^2(\lambda)}{\partial \lambda}$. In the special case that $m_r (= m)$ are all equal, $\hat{\sigma}_n^2(\lambda) = \left(\frac{m-1+\lambda}{m-1}\right)^2 \frac{1}{R(m-1)}\left\{\sum_{r=1}^{R} Y_r^{*'} Y_r^* - \sum_{r=1}^{R} Y_r^{*'} Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^*\right)^{-1} \sum_{r=1}^{R} Z_r^{*'} Y_r^*\right\}$ 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.

[10]This condition is mainly used for the application of the simple Lyapounov central limit theorem for independent random variables. This boundedness condition can be replaced by more general conditions on third order empirical moment conditions.

[11]Under Assumption 2,

$$\left(\frac{m}{a_U m - 1}\right)^2 \cdot \frac{1}{n}\sum_{r=1}^{R} X_{r2}^{*'} X_{r2}^* \leqslant \frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}\right)^2 X_{r2}^{*'} X_{r2}^* \leqslant \left(\frac{m}{a_L m - 1}\right)^2 \cdot \frac{1}{n}\sum_{r=1}^{R} X_{r2}^{*'} X_{r2}^*.$$

These show that the existence of a positive definite limiting matrix for $\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}\right)^2 X_{r2}^{*'} X_{r2}^*$ implied by Assumption 3 will be compatible with the possibility that $\frac{1}{n}\sum_{r=1}^{R} X_{r2}^{*'} X_{r2}^*$ converges to a finite matrix.

estimates of β_{10} and β_{20}. As $J_r W_r = -\frac{1}{m_r(0)} J_r$, one has

$$\frac{1}{n}\sum_{r=1}^{R} Z_r^{*'} Z_r^* = \frac{1}{n}\begin{pmatrix} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^* & -\sum_{r=1}^{R} \frac{m}{m_r(0)} X_{r1}^{*'} X_{r2}^* \\ -\sum_{r=1}^{R} \frac{m}{m_r(0)} X_{r2}^{*'} X_{r1}^* & \sum_{r=1}^{R} \left(\frac{m}{m_r(0)}\right)^2 X_{r2}^{*'} X_{r2}^* \end{pmatrix}$$
$$= \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{m}{\sqrt{n}} \end{pmatrix} \begin{pmatrix} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^* & \sum_{r=1}^{R} X_{r1}^{*'} (WX_{r2})^* \\ \sum_{r=1}^{R} (W_r X_{r2})^{*'} X_{r1}^* & \sum_{r=1}^{R} (W_r X_{r2})^{*'} (W_r X_{r2})^* \end{pmatrix}$$
$$\times \begin{pmatrix} \frac{1}{\sqrt{n}} & 0 \\ 0 & \frac{m}{\sqrt{n}} \end{pmatrix}.$$

The normalization factors $\frac{1}{\sqrt{n}}$ and $\frac{m}{\sqrt{n}}$ have implications for the possible rates of convergence of estimates of β_{10} and β_{20}.

This compact parameter space in Assumption 4 is needed as the CM approach works with the concentrated likelihood (3.5), which is nonlinear in λ. The condition on the lower bound of λ in Assumption 4 guarantees that $\ln(m_r(\lambda))$ is well defined and $m_r(\lambda)$ is bounded away from zero for all $\lambda \in \Lambda$ and for all r.[12] One does not need to impose any 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, it shall be understood that n goes to infinity refers to both R and m tend to infinity.[13] 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.

Assumption 5 is equivalent to that $\frac{m^2}{n}$ tends to zero or $\frac{\sqrt{n}}{m}$ tends to infinity because $n = Rm$. 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

[12]For the maximum likelihood estimation of the group effect model without fixed effects in Lee (2004), one needs to evaluate the determinant $|I_r - \lambda W_r|$ for each r at values of λ of interest. With W_r in (1.3), the determinant of $(I_{m_r} - \lambda W_r)$, is $(1-\lambda)\left(\frac{m_r(\lambda)}{m_r(0)}\right)^{m_r(0)}$ because $(I_{m_r} - \lambda W_r) = \left(\frac{m_r(\lambda)}{m_r(0)}\right)\left(I_{m_r} - \frac{\lambda}{m_r(\lambda)} l_{m_r} l'_{m_r}\right)$. This determinant is non-singular if and only if $\lambda \neq 1$ and $m_r(\lambda) \neq 0$ for all r. The determinant $|I_{m_r} - \lambda W_r|$ shall not be zero or change its sign on its parameter space 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, it may assume $1 > \lambda > -1$. However, the CML approach is focusing on the estimation of the within equation, which requires only to restrict $\lambda > 1 - m_r$ for all r in order the log likelihood in (3.2) to be well defined. Thus, for the asymptotic analysis of the CML estimates, one does not necessarily restrict λ to be less than one.

[13]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 2 as n tends to infinity. The above convention simplifies the notations.

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. Identification and consistency of the CMLE

Define a non-stochastic function

$$Q_{c,n}(\lambda) = \max_{\beta,\sigma^2} \mathrm{E}(\ln L_{w,n}(\beta, \sigma^2, \lambda)). \tag{3.6}$$

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* $\lambda \in \Lambda$.

The detailed proofs of Proposition 1 and subsequent propositions are collected in the Appendix.

Proposition 1 shows that the average difference $\frac{1}{n}[\ln L_{c,n}(\lambda) - \ln L_{c,n}(\lambda_0)]$ of the log concentrated likelihood can be asymptotically equivalent to the average difference $\frac{1}{n}[Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)]$ of the non-stochastic function (3.6). When m goes to infinity, the rate of convergence to zero can be fast with at least the rate m^2. The average difference $\frac{1}{n}[Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)]$ can be rather flat on Λ. In order to review its shape and the possible identification of λ_0, it is necessary to magnify this difference by the factor m^2. It is shown in the Appendix that $\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))$ satisfies the identification uniqueness condition (White, 1994) under the following Assumptions.

Assumption 6.1 (*Identification 1*). The limiting matrix

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{R} \left(-\frac{m}{m_r(\lambda_0)} Z_r^* \delta_{m0}, Z_r^* \right)' \left(-\frac{m}{m_r(\lambda_0)} Z_r^* \delta_{m0}, Z_r^* \right)$$

exits and is positive definite.

Note that under Assumption 3 that the limiting matrix of $\frac{1}{n}\sum_{r=1}^{R} Z_r^{*'} Z_r^*$ is non-singular, Assumption 6.1 is equivalent to

$$\lim_{n\to\infty} \frac{1}{n} \left\{ \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right)^2 (Z_r^* \delta_{m0})'(Z_r^* \delta_{m0}) - \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} Z_r^{*'} (Z_r^* \delta_{m0}) \right\} > 0$$

(e.g., Theil, 1971, p. 18 on the inverse of a partitioned symmetric matrix).

The identification condition in Assumption 6.1 shall be valid for cases with small group interactions; and, for large group interactions, when x_{r1} has relevant components which are not in x_{r2}. It can break down if $\beta_{10} = 0$ or $x_{ri,2} = x_{ri,1}$ for all r and i for the case with $m \to \infty$. If Assumption 6.1 breaks down, one has to resort to the covariance structure of the disturbance term in (2.5).

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

$$\limsup_{n\to\infty} m^2 \left[\sum_{r=1}^{R} \left(\frac{m_r(0)}{R(m-1)} \right) \ln \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2 - \ln \left(\sum_{r=1}^{R} \frac{m_r(0)}{R(m-1)} \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2 \right) \right] < 0.$$

More detailed motivation and justification of these identification conditions are in the Appendix (after the proof of Proposition 1).

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

Proposition 2. *Under Assumptions* 1–5, 6.1 *or* 6.2, *the identification uniqueness condition that, for any open neighborhood* $N_\varepsilon(\lambda)$ *of* λ_0 *in* Λ, $\limsup_{n\to\infty}\max_{\lambda\in\bar{N}_\varepsilon(\lambda_0)}\frac{m^2}{n}(Q_{c,n}(\lambda)-Q_{c,n}(\lambda_0))<0$, *where* $\bar{N}_\varepsilon(\lambda_0)$ *is the complement of* $N_\varepsilon(\lambda_0)$ *in* Λ, *will hold, and* $\hat{\lambda}_n$ *is a consistent estimator of* λ_0.

The global identification conditions in Assumptions 6.1 or 6.2 imply a local identification condition that $\lim_{n\to\infty}\frac{m^2}{n}\frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial\lambda^2}$ shall be negative definite. A sufficient condition for local identification induced by Assumption 6.2 is that the limiting weighted variance of $\frac{m}{m_r(\lambda_0)}$, $r=1,\ldots,R$, does not vanish, i.e.,

$$\lim_{n\to\infty}\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\left[\frac{m}{m_r(\lambda_0)}-\sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right]^2>0 \tag{3.7}$$

(see the Appendix).

3.3. Asymptotic distribution of the CMLE

For asymptotic distribution, it is important to investigate $\frac{\partial \ln L_{c,n}(\lambda)}{\partial\lambda}$ evaluated at λ_0. The following proposition shows that $\frac{1}{n}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial\lambda}$ may 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 $\mathscr{E}_n$.

Proposition 3. *Under Assumptions* 1–5, *and* 6.1 *or* 6.2,

$$\frac{\sqrt{n}}{m}(\hat{\lambda}_n-\lambda_0)\xrightarrow{D}\mathrm{N}(0,\Sigma_\lambda^{-1}), \tag{3.8}$$

where

$$\begin{aligned}\Sigma_\lambda = \lim_{n\to\infty}\Bigg\{&\frac{1}{n\sigma_0^2}\Bigg[\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2(Z_r^*\delta_{m0})'(Z_r^*\delta_{m0})\\ &-\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})'Z_r^*\left(\sum_{r=1}^{R}Z_r^{*'}Z_r^*\right)^{-1}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}(Z_r^*\delta_{m0})\Bigg]\\ &+2\left(\frac{m-1}{m}\right)\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\left[\frac{m}{m_r(\lambda_0)}-\sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right]^2\Bigg\}.\end{aligned} \tag{3.9}$$

A possible non-degenerated distribution for the estimator $\hat{\lambda}_n$ will depend on the identification conditions that either $\lim_{n\to\infty}\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\left[\frac{m}{m_r(\lambda_0)}-\sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right]^2\neq 0$ or Assumption 6.1 holds. The $\frac{1}{n}\left[\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2(Z_r^*\delta_{m0})'(Z_r^*\delta_{m0})-\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})'Z_r^*\right.$ $\left.\left(\sum_{r=1}^{R}Z_r^{*'}Z_r^*\right)^{-1}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}(Z_r^*\delta_{m0})\right]$ is the average sum of squared residuals of the regression of $\frac{m}{m_r(\lambda_0)} Z_r^* \delta_{m0}$ on Z_r^* with $r = 1, \ldots, R$. Its limit shall be strictly positive under Assumption 6.1.

From (3.8), the CMLE $\hat{\lambda}_n$ converges to λ_0 at the rate $\frac{\sqrt{n}}{m}$. For the case with small group interactions, the $\{m_r\}$ is bounded, its 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 precision of $\hat{\lambda}_n$ summarized in the precision matrix Σ_λ depends on the sum of squared residuals of the regression of $\frac{m}{m_r(\lambda_0)} Z_r^* \delta_{m0}$ on Z_r^* with $r = 1, \ldots, R$ and the weighted variations of $\frac{m}{m_r(\lambda_0)}$. Large variations will result in relatively more precise estimation of λ_0.

The CMLE of β_0 is

$$\hat{\beta}_n = \begin{pmatrix} I_{k_1} & 0 \\ 0 & mI_{k_2} \end{pmatrix} \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} \left(\frac{m_r(\hat{\lambda}_n)}{m_r(0)} \right) Z_r^{*'} Y_r^*. \tag{3.10}$$

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

Proposition 4. *Under Assumptions* 1–5, *and* 6.1 *or* 6.2,

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \dfrac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} \xrightarrow{D} \mathrm{N}(0, \Omega_\beta), \tag{3.11}$$

where

$$\Omega_\beta = \Sigma_\lambda^{-1} \lim_{n\to\infty} \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right) Z_r^{*'} (Z_r^* \delta_{m0}) \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right) (Z_r^* \delta_{m0})' Z_r^* \\ \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} + \sigma_0^2 \lim_{n\to\infty} \left(\frac{1}{n} \sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1}. \tag{3.12}$$

This proposition shows that the asymptotic distributions of $\hat{\beta}_{n1}$ and $\hat{\beta}_{n2}$ depend on the asymptotic distribution of $\hat{\lambda}_n$. However, 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 $\frac{1}{m_r(0)} x_{ri,2}$ 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, $\sqrt{\frac{R}{m}}$-rate). The contextual effect parameter is more difficult to be precisely estimated than that of the other regression coefficients.

The results of Propositions 3 and 4 can be combined to derive the joint distribution of $\hat{\lambda}_n$ and $\hat{\beta}_n$.

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

$$\begin{pmatrix} \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} \xrightarrow{D} \mathrm{N}(0, \Omega_{\lambda,\beta}), \tag{3.13}$$

where

$$\Omega_{\lambda,\beta} = \lim_{n\to\infty} \left[\frac{1}{\sigma_0^2 n} \sum_{r=1}^{R} \left(-\frac{m}{m_r(\lambda_0)} Z_r^* \delta_{m0}, Z_r^* \right)' \left(-\frac{m}{m_r(\lambda_0)} Z_r^* \delta_{m0}, Z_r^* \right) \right.$$
$$\left. + 2\left(\frac{m-1}{m}\right) \sum_{r=1}^{R} \left(\frac{m_r(0)}{R(m-1)} \right) \left[\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R} \left(\frac{m_s(0)}{R(m-1)} \right) \frac{m}{m_s(\lambda_0)} \right]^2 e_1 e_1' \right]^{-1} \tag{3.14}$$

and e_1 *is the first unit vector of dimension* $(k_1 + k_2 + 1)$.

4. Instrumental variables estimation

The within equation can be rewritten as $y_{ri} - \bar{y}_r = -\lambda_0 \frac{(y_{ri} - \bar{y}_r)}{m_r(0)} + (z_{ri} - \bar{z}_r)\delta_{m0} + (\varepsilon_{ri} - \bar{\varepsilon}_r)$, which is explicitly

$$y_{ri} - \bar{y}_r = -\lambda_0 \frac{(y_{ri} - \bar{y}_r)}{m_r(0)} + (x_{ri,1} - \bar{x}_{r,1})\beta_{10} - \frac{(x_{ri,2} - \bar{x}_{r,2})}{m_r(0)}\beta_{20} + (\varepsilon_{ri} - \bar{\varepsilon}_r). \tag{4.1}$$

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

$$\mathrm{E}\left[\frac{1}{m_r(0)}(y_{ri} - \bar{y}_r) \right] = \frac{1}{m_r(\lambda_0)}(z_{ri} - \bar{z}_r)\delta_{m0}, \tag{4.2}$$

the best IV vector is $\left(\frac{1}{m_r(\lambda_0)}(z_{ri} - \bar{z}_r)\delta_{m0}, z_{ri} - \bar{z}_r \right)$, or equivalently, $\left(\frac{1}{m_r(\lambda_0)}(z_{ri} - \bar{z}_r), z_{ri} - \bar{z}_r \right)$, as motivated by Amemiya (1985) and Lee (2003). The components of the best IV vector may 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 Eq. (4.1).

Let p_{ri} be an IV variable. After rescaling, $\frac{p_{ri}}{m_r(0)}$ can be used as an IV for $-\frac{(y_{ri} - \bar{y}_r)}{m_r(0)}$. Let P_r be the m_r-dimensional column vector of p_{ri} for the rth group. The IV estimator of $\theta_0 = (\lambda_0, \beta_{10}', \beta_{20}')'$ is

$$\hat{\theta}_{n,\mathrm{IV}} = \left[\sum_{r=1}^{R} \left(\frac{P_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)} \right)' \left(-\frac{Y_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)} \right) \right]^{-1}$$
$$\times \sum_{r=1}^{R} \left(\frac{P_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)} \right)' Y_r^*. \tag{4.3}$$

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 6. *Under Assumptions* 1–3, 5 *and* 6.1,

$$\begin{pmatrix} \frac{\sqrt{n}}{m}(\hat{\lambda}_{n,\mathrm{IV}} - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1,\mathrm{IV}} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2,\mathrm{IV}} - \beta_{20}) \end{pmatrix} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \Omega_{\mathrm{IV}}), \tag{4.4}$$

where

$$\begin{aligned}\Omega_{\mathrm{IV}} =& \sigma_0^2 \lim_{n\to\infty} \left[\frac{1}{n}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\left(-\frac{m}{m_r(\lambda_0)}Z_r^*\delta_{m0}, Z_r^*\right)\right]^{-1} \\ &\times \frac{1}{n}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)\left[\frac{1}{n}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\left(-\frac{m}{m_r(\lambda_0)}Z_r^*\delta_{m0}, Z_r^*\right)\right]'^{-1},\end{aligned}$$

which is assumed to exist.

The (*feasible*) *best* IV *will be* $\frac{1}{m_r(0)}P_r = \left(-\frac{X_{r,1}\hat{\beta}_{n1}}{m_r(\hat{\lambda}_n)} + \frac{X_{r,2}\hat{\beta}_{n2}}{m_r(0)m_r(\hat{\lambda}_n)}\right)$, *where* $(\hat{\lambda}_n, \hat{\beta}_{n1}, \hat{\beta}_{n2})$ *can be any initial* IV *consistent estimator. Its asymptotic distribution is*

$$\begin{pmatrix} \frac{\sqrt{n}}{m}(\hat{\lambda}_{n,\mathrm{BIV}} - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1,\mathrm{BIV}} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2,\mathrm{BIV}} - \beta_{20}) \end{pmatrix} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \Omega_{\mathrm{BIV}}), \tag{4.5}$$

where $\Omega_{\mathrm{BIV}} = \sigma_0^2 \lim_{n\to\infty}\left[\frac{1}{n}\sum_{r=1}^{R}\left(-\frac{m}{m_r(\lambda_0)}Z_r^*\delta_{m0}, Z_r^*\right)'\left(-\frac{m}{m_r(\lambda_0)}Z_r^*\delta_{m0}, Z_r^*\right)\right]^{-1}$.

From this proposition, the IV estimate $\hat{\beta}_{n1,\mathrm{IV}}$ converges in probability to β_{10} at the usual $\sqrt{n}$-rate, but, the IV estimates $\hat{\lambda}_{n,\mathrm{IV}}$ and $\hat{\beta}_{n2,\mathrm{IV}}$ of λ_0 and β_{20} converge at the $\frac{\sqrt{n}}{m}$-rate.[14]

One may compare the asymptotic relative efficiency of the IV estimators with that of the CMLE. From Propositions 5 and 6, it is apparent that $\Omega_{\lambda,\beta} \leqslant \Omega_{\mathrm{BIV}}$. Indeed, in terms of their precision matrices $\Omega_{\lambda,\beta}^{-1} - \Omega_{\mathrm{BIV}}^{-1} = 2\left(\frac{m-1}{m}\right)\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\left[\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right]^2 e_1e_1'$. Thus, the main efficient gain of the CMLE is due to the interaction effect on the reduced form disturbances of the within Eq. (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., $\beta_{10} = 0$. Another case is when $X_{r2} = X_{r1}$ for all r and $m \to \infty$. The following proposition illustrates the asymptotic property of the IV estimator for the case $\beta_{10} = 0$. 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.

[14]For the large groups case, the best IV can simply be $\frac{P_r}{m_r(0)} = \left(-\frac{X_{r1}\hat{\beta}_{n,1}}{m_r(\hat{\lambda}_n)}\right)$ for the rth group because the term $\frac{m}{m_r(0)m_r(\lambda_0)}X_{r2}\beta_{20}$ of $Z_r\delta_{m0}$ is relatively small and can be ignored asymptotically.

Proposition 7. *In the event that* $\beta_{10} = 0$ *in* (4.1), *under Assumptions* 1–3, 5 *and that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{R} \left(\frac{mP_r^*}{m_r(0)}, Z_r^* \right)' \left(\frac{m^2}{m_r(\lambda_0)m_r(0)} X_{r2}^* \beta_{20}, Z_r^* \right)$$

exists and is a non-singular matrix, *then*

$$\begin{pmatrix} \frac{\sqrt{n}}{m^2}(\hat{\lambda}_{n,\mathrm{IV}} - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1,\mathrm{IV}} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2,\mathrm{IV}} - \beta_{20}) \end{pmatrix} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \Omega_{\mathrm{IV}}), \tag{4.6}$$

where

$$\Omega_{\mathrm{IV}} = \sigma_0^2 \lim_{n\to\infty} \left[\frac{1}{n} \sum_{r=1}^{R} \left(\frac{mP_r^*}{m_r(0)}, Z_r^* \right)' \left(\frac{m^2}{m_r(\lambda_0)m_r(0)} X_{r2}^* \beta_{20}, Z_r^* \right) \right]^{-1}$$
$$\times \frac{1}{n} \sum_{r=1}^{R} \left(\frac{mP_r^*}{m_r(0)}, Z_r^* \right)' \left(\frac{mP_r^*}{m_r(0)}, Z_r^* \right) \left[\frac{1}{n} \sum_{r=1}^{R} \left(\frac{mP_r^*}{m_r(0)}, Z_r^* \right)' \right.$$
$$\left. \times \left(\frac{m^2}{m_r(\lambda_0)m_r(0)} X_{r2}^* \beta_{20}, Z_r^* \right) \right]^{\prime -1},$$

which is assumed to exist.

The best IV *will be* $\frac{1}{m_r(0)} P_r = \left(-\frac{X_{r,1}\hat{\beta}_{n1}}{m_r(\hat{\lambda}_n)} + \frac{X_{r,2}\hat{\beta}_{n2}}{m_r(0)m_r(\hat{\lambda}_n)} \right)$, *where* $(\hat{\lambda}_n, \hat{\beta}_{n1}, \hat{\beta}_{n2})$ *can be any initial* IV *consistent estimator. Its asymptotic distribution is*

$$\begin{pmatrix} \frac{\sqrt{n}}{m^2}(\hat{\lambda}_{n,\mathrm{BIV}} - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1,\mathrm{BIV}} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2,\mathrm{BIV}} - \beta_{20}) \end{pmatrix} \xrightarrow{\mathrm{D}} \mathrm{N}(0, \Omega_{\mathrm{BIV}}), \tag{4.7}$$

where $\Omega_{\mathrm{BIV}} = \sigma_0^2 \lim_{n\to\infty} \left[\frac{1}{n} \sum_{r=1}^{R} \left(\frac{m^2}{m_r(\lambda_0)m_r(0)} X_{r2}^* \beta_{20}, Z_r^* \right)' \left(\frac{m^2}{m_r(\lambda_0)m_r(0)} X_{r2}^* \beta_{20}, Z_r^* \right) \right]^{-1}$.

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. 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 7 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 in Proposition 5. Proposition 5 is valid without Assumption 6.1 as long as Assumption 6.2 is satisfied. This is so because, in this situation, 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 with large group interactions, as m is large, 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 $Y_r^* \doteq X_{r1}^* \beta_{10} + J_r \varepsilon_r$ and estimate β_{10} by the OLS:

$$\hat{\beta}_{n1,\mathrm{L}} = \left[\sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*} \right]^{-1} \sum_{r=1}^{R} X_{r1}^{*'} Y_r^{*}. \tag{5.1}$$

Under the setting of Assumption 5, $m(\hat{\beta}_{n1,\mathrm{L}} - \beta_{10}) = mb_n + \mathrm{O_P}\left(\sqrt{\frac{m}{R}}\right)$ as shown in the Appendix. As $\frac{m}{R} \to 0$, $m(\hat{\beta}_{n,\mathrm{L}} - \beta_0)$ converges in probability to the limit of mb_n. The OLS (within) estimate of $\hat{\beta}_{n1,\mathrm{L}}$ is consistent but its rate of convergence is of order $\mathrm{O}(m)$, which is lower than the $\sqrt{n}$-rate, and the limiting distribution of $m(\hat{\beta}_{n1,\mathrm{L}} - \beta_0)$ is degenerated. Thus, one can conclude that if the within equation is the proper model, the seemingly minor misspecification of the regression equation, which ignores the structural spatial interaction, seems to have damaging effects on the conventional within estimates.

In the scenario that $m \to \infty$ but $\frac{m}{R} \to c$ where c is a finite positive constant, one has from (A.85) in the Appendix that $m(\hat{\beta}_{n1,\mathrm{L}} - \beta_{10}) \xrightarrow{\mathrm{D}} \mathrm{N}\Big(\lim_{n\to\infty} mb_n, \sigma_0^2(\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*})^{-1}\Big)$, which has an asymptotic normal distribution with a non-zero asymptotic bias.

The within OLS would behave properly only under the circumstance that $\frac{m}{R} \to \infty$, i.e., the sizes of groups are much larger than the total number of groups. Under this situation which corresponds to $\frac{\sqrt{n}}{m} \to 0$, (A.85) implies $\sqrt{n}(\hat{\beta}_{n1,\mathrm{L}} - \beta_{10}) \xrightarrow{\mathrm{D}} \mathrm{N}\Big(0, \sigma^2(\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*})^{-1}\Big)$.

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 $\mathrm{N}(0,1)$. 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. $\mathrm{N}(0,\sigma^2)$ 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 10 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 and 10. With the magnification factor 8, the smallest group size is 16, the largest group size is 88, and the average group size is 52 per group. With the large magnification factor 10, the smallest group size is 20, the largest group size is 110, and the average group size is 65 per group. For the CML, we have experimented with various numbers of groups R from 50 to 3000 for the small group interaction cases, and from 50 to 6000 for the large group interaction cases.

The study in Hanushek et al. (2003) has analyzed a panel data set of school operations constructed by the UTD Texas Schools Project. This data track the universe of three successive cohorts of Texas public elementary students. For each cohort, there are over 200,000 students in over 3000 public schools. Thus if a grade in a school is considered as a group, there would be approximately 3000 groups for a cohort and the average group size is about 66 students in a grade. The Monte Carlo design with average group size 65 per group and a total of 3000 or 6000 groups is attempting to calibrate the size of the UTD data.[15] 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 for models with small group interactions are reported in Table 1. 'SG-IX' denotes the small group interactions model where x_1 and x_2 are independent. 'SG-SX' is the small group interactions model with $x_1 = x_2$. These models are estimated by the CML method with the parametric space of λ restricted to $(-1, 1)$. As the CML estimates can also be derived without imposing an upper limit, those estimates are reported under the columns 'SG-IX(UC)' and 'SG-SX(UC)'. The main entries in all the tables are means and standard errors (in bracket) of estimates across the replications. Table 2 reports CML estimates for models with large group interactions. 'LG8' denotes the scenario with an average of 52 members per group, and 'LG10' denotes the case with an average of 65 members per group. In the large group interaction cases, x_1 and x_2 are independent. The CML estimates for LG8 and LG10 are derived with the parameter space of λ restricted to $(-1, 1)$. Without imposing the upper limit, the (unconstrained) estimates are reported under the columns 'LG8(UC)' and 'LG10(UC)'.

For the small group interactions model SG-IX, 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 R's are not large, there are cases where the CML estimates $\hat{\lambda}_n$ occur at the boundary value 1. For $R = 50$ or 100, the CML estimates of λ with constrained value on $(-1, 1)$ have smaller variances than those under the column SG-IX(UC). For $R = 200$ or more, there are little differences between the CML estimates with or without restricted parameter spaces.[16] When $x_2 = x_1$, the CMLE's of SG-SX and SG-SX(UC) have similar properties as those of SG-IX and SG-IX(UC), except that these estimates have larger standard deviations.

For the models with large group interactions LG8 and LG10, the behavior of the estimates of λ is overall poor, even though the estimates of the other parameters β_1, β_2 and σ are satisfactory. The estimates of both LG8 and LG10 are biased downward when $R = 50$ and 100. But for larger R, they tend to be biased upward instead. These biases might not be reduced with very large $R = 3000$ or 6000. 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. There are many cases with estimates of λ having the boundary value 1. Without imposing the parameter being

[15]The UTD data set is not available for a general public user. A relative smaller data set which is available for the public use is the National Longitudinal Study of Adolescent Health (Add Health). This is a nationally representative school-based study of adolescents in grades 7 through 12 in 132 schools. An in-school questionnaire was given to all the students attending the sampled schools from September 1994 to April 1995, resulting in a total sample of over 90,000 students. The adolescents were asked to nominate friends. Friends's identification numbers make it possible to construct friendship networks in school. This data set is useful for the study of small group interactions. Information on this data set is available in Bearman et al. (1997). One may consider a grade as a group. However, the model in this paper does not directly applicable without modification when friendship networks are taken into account. A student at OSU has applied an extended model and the IV estimation method to student academic attainments after controlling for the grade fixed effect and friendship network. Some preliminary results are reported in Lin (2004).

[16]For $R = 400$, with one exception, all the CML estimates are less than one.

Table 1
CML estimates—small group interactions

R		Models			
		SG-IX	SG-IX(UC)	SG-SX	SG-SX(UC)
50	λ	0.5678 (.2876)	0.6202 (.4067)	0.5748 (.3316)	0.6668 (.5198)
	β_1	1.0105 (.0749)	1.0184 (.0866)	0.9960 (.1077)	1.0039 (.1138)
	β_2	1.0142 (.3183)	1.0298 (.3279)	0.9761 (.4883)	0.9585 (.5294)
	σ	1.0046 (.0639)	1.0123 (.0758)	1.0048 (.0696)	1.0185 (.0909)
100	λ	0.5527 (.2339)	0.5682 (.2770)	0.5590 (.2745)	0.5828 (.3358)
	β_1	1.0115 (.0551)	1.0138 (.0596)	0.9996 (.0737)	1.0019 (.0761)
	β_2	1.0052 (.2147)	1.0097 (.2190)	0.9802 (.3161)	0.9773 (.3207)
	σ	1.0055 (.0467)	1.0078 (.0515)	1.0052 (.0519)	1.0088 (.0595)
200	λ	0.5366 (.1745)	0.5356 (.1791)	0.5263 (.2046)	0.5294 (.2196)
	β_1	1.0070 (.0396)	1.0068 (.0399)	0.9982 (.0522)	0.9985 (.0527)
	β_2	1.0098 (.1460)	1.0094 (.1466)	0.9806 (.2188)	0.9802 (.2190)
	σ	1.0049 (.0330)	1.0047 (.0334)	1.0015 (.0377)	1.0020 (.0395)
400	λ	0.5216 (.1239)	0.5182 (.1255)	0.5213 (.1535)	0.5166 (.1522)
	β_1	1.0059 (.0277)	1.0054 (.0278)	0.9994 (.0399)	0.9989 (.0398)
	β_2	1.0055 (.1019)	1.0042 (.1024)	0.9869 (.1613)	0.9872 (.1608)
	σ	1.0031 (.0249)	1.0026 (.0249)	1.0016 (.0291)	1.0009 (.0288)
800	λ	0.5118 (.0890)	0.5076 (.0898)	0.5143 (.1119)	0.5067 (.1124)
	β_1	1.0031 (.0211)	1.0024 (.0212)	1.0004 (.0301)	0.9995 (.0301)
	β_2	1.0013 (.0697)	0.9997 (.0698)	0.9958 (.1141)	0.9963 (.1135)
	σ	1.0019 (.0175)	1.0013 (.0176)	1.0012 (.0208)	1.0000 (.0209)
1600	λ	0.5088 (.0589)	0.4995 (.0614)	0.5178 (.0827)	0.5082 (.0802)
	β_1	1.0028 (.0142)	1.0012 (.0144)	1.0003 (.0199)	0.9992 (.0194)
	β_2	1.0001 (.0518)	0.9965 (.0518)	0.9959 (.0759)	0.9963 (.0756)
	σ	1.0014 (.0119)	0.9999 (.0125)	1.0024 (.0148)	1.0008 (.0146)
3000	λ	0.5025 (.0427)	0.4941 (.0452)	0.5105 (.0626)	0.4977 (.0602)
	β_1	1.0008 (.0104)	0.9995 (.0108)	1.0003 (.0145)	0.9989 (.0147)
	β_2	1.0013 (.0377)	0.9980 (.0378)	0.9976 (.0540)	0.9983 (.0534)
	σ	1.0001 (.0087)	0.9988 (.0094)	1.0018 (.0110)	0.9997 (.0110)

Remarks:
(1) SG-IX: a model with small group interactions with an average of 6.5 members per group; x_1 and x_2 are independent. Corresponding to $R = 50, 100, 200, 400, 800, 1600, 3000$, the total sample observations are, respectively, $NT = 325, 650, 1300, 2600, 5200, 10{,}400$ and 19,500.
(2) SG-SX: the model with small group interactions in (1) but $x_2 = x_1$.
(3) SG-SX and SG-SX are estimated by imposing the constraints $(-1, 1)$. SG-IX(UC) and SG-SX(UC) are estimated without imposing the upper limit 1.
(4) R: the number of groups.
(5) The main entries are means and standard errors (in bracket) of estimates across the replications.

less than one, the estimates of λ under the columns LG8(UC) and LG10(UC) are now biased upwards for small R. Again, for large $R = 1600, 3000$ or 6000, the magnitudes of the biases are not stable. The estimates of λ have still quite large variances even R is large as 3000 or 6000. Both the interaction effects $\hat{\lambda}_n$ and $\hat{\beta}_{n2}$ have obviously much larger standard deviations than those in Table 1 with small group interactions. We note that this occurs

Table 2
CML estimates—models with large group interactions

R		Models			
		LG8	LG8(UC)	LG10	LG10(UC)
50	λ	0.4262 (.6640)	0.7097 (1.0323)	0.3661 (.6968)	0.6902 (1.1489)
	β_1	1.0003 (.0234)	1.0057 (.0279)	0.9993 (.0197)	1.0043 (.0246)
	β_2	0.9808 (.8228)	0.9897 (.8280)	0.9820 (.9664)	0.9929 (.9733)
	σ	0.9988 (.0204)	1.0042 (.0257)	0.9982 (.0174)	1.0032 (.0223)
100	λ	0.4903 (.5570)	0.6181 (.7457)	0.4416 (.6041)	0.5943 (.8135)
	β_1	1.0005 (.0181)	1.0030 (.0202)	0.9999 (.0155)	1.0022 (.0177)
	β_2	0.9812 (.5974)	0.9856 (.6006)	0.9755 (.6770)	0.9805 (.6815)
	σ	1.0002 (.0150)	1.0027 (.0179)	0.9993 (.0135)	1.0016 (.0159)
200	λ	0.5139 (.4298)	0.5405 (.5310)	0.4952 (.4814)	0.5410 (.5791)
	β_1	1.0013 (.0136)	1.0018 (.0146)	1.0007 (.0116)	1.0014 (.0126)
	β_2	0.9862 (.4230)	0.9868 (.4237)	0.9792 (.5287)	0.9805 (.5297)
	σ	1.0004 (.0116)	1.0009 (.0130)	0.9998 (.0101)	1.0005 (.0110)
400	λ	0.5328 (.3481)	0.4893 (.3559)	0.5537 (.4077)	0.5130 (.4129)
	β_1	1.0010 (.0096)	1.0001 (.0096)	1.0011 (.0087)	1.0005 (.0087)
	β_2	0.9877 (.2966)	0.9861 (.2959)	0.9728 (.3515)	0.9718 (.3517)
	σ	1.0005 (.0087)	0.9997 (.0089)	1.0008 (.0078)	1.0002 (.0077)
800	λ	0.5803 (.2694)	0.5171 (.2770)	0.5664 (.2994)	0.5549 (.3091)
	β_1	1.0014 (.0075)	1.0002 (.0076)	1.0010 (.0068)	1.0008 (.0069)
	β_2	0.9923 (.2130)	0.9901 (.2125)	0.9815 (.2468)	0.9811 (.2466)
	σ	1.0015 (.0065)	1.0003 (.0065)	1.0011 (.0054)	1.0009 (.0067)
1600	λ	0.5590 (.2693)	0.5200 (.2592)	0.8059 (.3908)	0.4042 (.2710)
	β_1	1.0011 (.0062)	1.0004 (.0061)	1.0046 (.0070)	0.9984 (.0055)
	β_2	0.9976 (.1484)	0.9963 (.1481)	0.9968 (.1680)	0.9856 (.1659)
	σ	1.0012 (.0058)	1.0004 (.0056)	1.0048 (.0065)	0.9986 (.0045)
3000	λ	0.5010 (.1532)	0.3616 (.2006)	0.8435 (.2939)	0.4689 (.2598)
	β_1	1.0001 (.0041)	0.9974 (.0046)	1.0053 (.0052)	0.9995 (.0047)
	β_2	0.9994 (.1147)	0.9943 (.1142)	1.0049 (.1319)	0.9942 (.1297)
	σ	1.0001 (.0035)	0.9974 (.0043)	1.0053 (.0048)	0.9995 (.0044)
6000	λ	0.8040 (.2275)	0.3892 (.2160)	0.5729 (.2456)	0.6890 (.1977)
	β_1	1.0059 (.0049)	0.9979 (.0047)	1.0011 (.0043)	1.0029 (.0036)
	β_2	1.0129 (.0746)	0.9980 (.0729)	1.0017 (.0875)	1.0050 (.0874)
	σ	1.0059 (.0046)	0.9978 (.0044)	1.0011 (.0040)	1.0029 (.0033)

(1) LG8 is a model with large group interactions with an average of 52 members per group; LG10 has an average of 65 members per groups; x_1 and x_2 are independent in these models.
(2) For LG8, corresponding to $R = 50, 100, 200, 400, 800, 1600, 3000$, and 6000, the total sample observations are, respectively, $NT = 2,600, 5,200, 10,400, 20,800, 41,600, 83,200, 156,000$ and 312,000.
(3) ForLG10, corresponding to $R = 50, 100, 200, 400, 800, 1,600, 3,000$, and 6,000, the total sample observations are, respectively, $NT = 3250, 6500, 13,000, 26,000, 52,000, 104,000, 195,000$ and 390,000.
(4) LG8 and LG10 are estimated by imposing the constraints $(-1, 1)$. LG8(UC) and LG10(UC) are estimated without imposing the upper limit 1.

even though the total sample sizes NT in LG8 and LG10 are, respectively, 8 and 10 times larger than the corresponding one in SG-IX. The $\hat{\beta}_{n1}$ and $\hat{\sigma}_n$ have smaller standard errors than those of SG-IX because of its overall larger sample size. The large standard errors in

Table 3
IV estimates

R		Models			
		SG-IX		SG-SX	
		IV	BIV	IV	BIV
50	λ	0.6478 (.7360)	0.5996 (.5990)	–[a]	–[a]
	β_1	1.0204 (.1238)	1.0134 (.1089)	–	–
	β_2	1.0309 (.3867)	1.0159 (.3598)	–	–
100	λ	0.6305 (.4695)	0.6034 (.4256)	–[b]	–[a]
	β_1	1.0228 (.0838)	1.0186 (.0783)	–	–
	β_2	1.0275 (.2615)	1.0180 (.2482)	–	–
200	λ	0.5768 (.3282)	0.5641 (.3071)	–[a]	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) [a]indicates that the empirical means of the estimated λ are greater than one and corresponding standard errors are large (greater than 10).
(2) [b]indicates that the empirical mean of the estimated λ is negative.

the estimates of $\hat{\lambda}_n$ (and $\hat{\beta}_{n2}$) confirm our theoretical implications on the difficulty of estimating those effects. These Monte Carlo results illustrate both the statistical and numerical difficulties for the estimation of λ for the large group interactions case.[17]

Table 3 reports some 2SLS estimates of the models with small group interactions. For an (initial) IV estimation, an IV for $\frac{(y_{ri}-\bar{y}_r)}{m_r(0)}$ in (4.1) is constructed as follows. First, $(y_{ri}-\bar{y}_r)$ is regressed on $(x_{ri,1}-\bar{x}_{r,1})$ and $\frac{(x_{ri,2}-\bar{x}_{r,2})}{m_r(0)}$. The IV variable is constructed as the predicted value of this regression divided by $m_r(0)$ 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-IX and SG-SX. With this IV estimate as the initial one, the best IV can be derived as in

[17]For large $R = 3000$, the total sample observations are 156 thousands for LG8, and they are 195 thousands for LG10. The estimation might suffer from numerical stability or round error problems, which are hard to be quantified.

Table 4
Within OLS estimates

R		LG8 model		
		Mean SD	90%CI	95%CI
50	β_1	0.9921 (.0191)	0.887	0.940
100	β_1	0.9910 (.0142)	0.820	0.907
200	β_1	0.9914 (.0097)	0.760	0.847
400	β_1	0.9907 (.0067)	0.607	0.723
800	β_1	0.9902 (.0051)	0.387	0.560
1600	β_1	0.9903 (.0037)	0.177	0.247
3000	β_1	0.9904 (.0026)	0.023	0.040

Proposition 6. The best IV estimates are reported under the columns 'BIV'. For SG-IX 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-IX and SG-IX(UC) 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. The IV estimates of the SG-SX 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 $R = 800$. For larger R, the biases can be reduced. The standard errors are much larger than those of the SG-IX model. This is so, because the IV or best IV variables in the SG-SX cases are more likely to be highly correlated.

Table 4 reports the conventional OLS (within) estimate of β_{10} by ignoring the endogenous and exogenous interaction effects for the large group interactions scenario as discussed in Section 5. The sample observations are generated from the LG8 model. As R increases from 50 to 3000, the within estimates change a little and their values are about 0.99. As the true value of β_{10} is 1, these point estimates seem reasonably good. However, as the sample variances decrease with increasing R, the asymptotic bias of 0.001 may be a problem for statistical inference as argued in Section 5. To reveal the possible difficulty in statistical inference, the 90% and 95% confidence intervals (CI) are constructed.[18] The coverage probabilities for the 90% CIs are reported under the column '90% CI'. Those for the 95% CIs are reported under the column '95% CI'. Even though the point estimates of β_1 seems close to one, the coverage probabilities are very poor with moderate or large groups. With $R = 400$, the coverage probabilities for the 90% and 95% CIs are, respectively, only 0.607 and 0.723. With $R = 1600$, the coverage probabilities are only 0.177 and 0.247. Table 5 provides coverage probabilities for the CML estimates for both small group and large group interactions cases for comparison purpose. The coverage probabilities seem mostly adequate. The relatively poor coverage probability cases occur

[18]The empirical standard deviations of the estimates are used for the construction of the CIs reported in Tables 4 and 5. For the within estimates, we have also used the conventional least squares standard error formulae, the corresponding 90% and 95% coverage probabilities are similar.

Table 5
Coverage probabilities of CML estimates

R		Models							
		SG-IX		SG-IX(UC)		LG8		LG8(UC)	
		90%CI	95%CI	90%CI	95%CI	90%CI	95%CI	90%CI	95%CI
50	λ	.840	.993	.910	.930	.887	.933	.917	.947
	β_1	.903	.950	.903	.940	.910	.963	.890	.930
	β_2	.907	.973	.917	.970	.893	.953	.893	.953
100	λ	.873	.917	.917	.940	.930	.950	.903	.947
	β_1	.893	.947	.887	.940	.900	.967	.897	.953
	β_2	.893	.957	.890	.967	.897	.947	.897	.950
200	λ	.893	.950	.893	.950	.920	.960	.893	.947
	β_1	.913	.963	.917	.963	.920	.963	.913	.963
	β_2	.897	.963	.887	.967	.897	.957	.897	.953
400	λ	.920	.950	.930	.953	.963	.973	.893	.943
	β_1	.890	.950	.893	.943	.913	.957	.897	.953
	β_2	.880	.947	.883	.947	.903	.937	.900	.933
800	λ	.870	.960	.897	.967	.940	.963	.973	.987
	β_1	.917	.940	.900	.940	.890	.950	.913	.943
	β_2	.887	.957	.877	.957	.910	.947	.913	.947
1600	λ	.890	.933	.910	.957	.973	.987	.960	.990
	β_1	.903	.947	.887	.943	.900	.950	.910	.963
	β_2	.887	.960	.873	.960	.910	.960	.917	.957
3000	λ	.913	.950	.900	.940	.907	.933	.840	.943
	β_1	.910	.943	.880	.940	.913	.940	.830	.920
	β_2	.893	.957	.910	.953	.913	.960	.927	.963

for the CML estimates of λ with restricted parameter space on $(-1, 1)$, when R is small as $R = 50$ or 100. For large group interactions, the poor cases occur when $R = 800$ or larger. The coverage probabilities for all the estimates of β_1 and β_2 seem adequate except a case with $R = 3000$ for β_1.

7. Conclusion

In this paper, we consider identification and estimation of social interaction models with fixed effects, which has an SAR structure. 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 model.[19]

[19]This model has more structures than models based on a cluster modeling approach. In the cluster approach, sample observations are assumed to be dependent only due to unobservables within clusters (see, e.g., Trivedi (2003, Chapter 24.5)).

With fixed group effect specification, the group unobservables are allowed to correlate 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 cannot 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 interaction 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 may have the same stochastic rates of convergence. The IV estimates are, however, relatively inefficient to the CML estimates. When the regressors in the model consist of only contextual factors, the IV estimate of the endogenous 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 on another occasion.[20]

[20]Strong identification and relatively fast rate of convergence of estimators may be 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. For the MLE of the model without group unobservables, the results in Lee (2004) show that the endogenous effect λ can either be $\sqrt{n}$ or $\sqrt{R}$ rates of convergence. 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).

Acknowledgments

I am grateful to the National Science Foundation, Economics Program for financial support of my research under the grant numbers SES-0111380 and SES-0519204. I appreciate having valuable comments and suggestions from three anonymous referees, Robert Moffitt, and seminar participants at Ohio State University, London School of Economics, and University College London. An earlier version of this paper is entitled 'Identification and Estimation of Spatial Econometric Models with Group Interactions, Contextual Factors and Fixed Effects'.

Appendix A. Mathematical appendix

A.1. List of often used notations

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
a_r a proportional factor such that $m_r = a_r m$
l_{m_r} the m_r-dimensional column vector of ones
$m_r(\lambda) = m_r - 1 + \lambda$
$m_r(0) = m_r - 1$
$z_{ri} = \left(x_{ri,1}, -\frac{m}{m_r(0)} x_{ri,2}\right)$
$\theta = (\lambda, \beta_1', \beta_2')'$
$\delta_m = (\beta_1', \beta_2'/m)'$
$J_r = I_{m_r} - \frac{1}{m_r} l_{m_r} l_{m_r}'$
$A_r^* = J_r A_r$ for any matrix A_r conformable with J_r

A.2. Derivation of the within and between equations in (2.2) and (2.3)

Because of the group structure, (2.1) implies that

$$\frac{1}{m_r} l_{m_r}' Y_r = \lambda_0 \frac{1}{m_r} l_{m_r}' W_r Y_r + \frac{1}{m_r} l_{m_r}' X_{r1} \beta_{10} + \frac{1}{m_r} l_{m_r}' W_r X_{r2} \beta_{20} + \alpha_r + \frac{1}{m_r} l_{m_r}' \varepsilon_r,$$

and $J_r Y_r = \lambda_0 J_r W_r Y_r + J_r X_{r1} \beta_{10} + J_r W_r X_{r2} \beta_{20} + J_r \varepsilon_r$. Because $W_r = \frac{1}{m_r(0)} (l_{m_r} l_{m_r}' - I_{m_r})$, $J_r W_r = -\frac{1}{m_r(0)} J_r$ and $l_{m_r}' W_r = l_{m_r}'$, one has the within and between equations in (2.2) and (2.3).

A.3. Derivation of the likelihood function

The likelihood function in (3.1) can be derived as follows. Because the components of the vector $(y_{r1}^*, \ldots, y_{r,m_r}^*)$ are linearly dependent, it is sufficient to consider the first $m_r(0)$

linearly independent components. The variance matrix of the corresponding disturbance vector of $\varepsilon^*_{ri} = \varepsilon_{ri} - \bar{\varepsilon}_r$, $i = 1, \ldots, m_r(0)$, is $var(\varepsilon^*_{r1}, \ldots, \varepsilon^*_{r,m_r(0)}) = \sigma^2(I_{m_r(0)} - \frac{1}{m_r} l_{m_r(0)} l'_{m_r(0)})$, which has the determinant $|I_{m_r(0)} - \frac{1}{m_r} l_{m_r(0)} l'_{m_r(0)}| = \frac{1}{m_r}$ and $(I_{m_r(0)} - \frac{1}{m_r} l_{m_r(0)} l'_{m_r(0)})^{-1} = I_{m_r(0)} + l_{m_r(0)} l'_{m_r(0)}$. Furthermore,

$$(y^*_{r1}, \ldots, y^*_{rm_r(0)})(I_{m_r(0)} + l_{m_r(0)} l'_{m_r(0)})(y^*_{r1}, \ldots, y^*_{rm_r(0)})' = Y^{*'}_r Y^*_r.$$

The likelihood function (3.1) follows.

Proof of Proposition 1. Because the within equation is $\left(\frac{m_r(\lambda_0)}{m_r(0)}\right) Y^*_r = Z^*_r \delta_{m0} + J_r \varepsilon_r$, and $\mathrm{E}(\varepsilon'_r J_r \varepsilon_r) = \sigma_0^2 \operatorname{tr}(J_r) = \sigma_0^2 m_r(0)$, it follows that

$$\begin{aligned}
&\mathrm{E}(\ln L_{w,n}(\beta, \sigma^2, \lambda)) \\
&\quad = c + \sum_{r=1}^{R} m_r(0) \ln(m_r(\lambda)) - \frac{(R(m-1))}{2} \ln \sigma^2 \\
&\quad - \frac{1}{2\sigma^2} \sum_{r=1}^{R} \left[\left(\frac{m_r(\lambda)}{m_r(\lambda_0)} Z^*_r \delta_{m0} - Z^*_r \delta_m \right)' \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} Z^*_r \delta_{m0} - Z^*_r \delta_m \right) \right. \\
&\quad \left. + \sigma_0^2 m_r(0) \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2 \right].
\end{aligned} \tag{A.1}$$

From (A.1), the optimization problem $\max_{\beta,\sigma^2} \mathrm{E}(\ln L_{w,n}(\beta, \sigma^2, \lambda))$ has the optimum solutions

$$\beta^*_n(\lambda) = \begin{pmatrix} I_{k_1} & 0 \\ 0 & m I_{k_2} \end{pmatrix} \left(\sum_{r=1}^{R} Z^{*'}_r Z^*_r \right)^{-1} \sum_{r=1}^{R} \frac{m_r(\lambda)}{m_r(\lambda_0)} Z^{*'}_r Z^*_r \delta_{m0} (= \mathrm{E}(\hat{\beta}_n(\lambda))), \tag{A.2}$$

where $\hat{\beta}_n(\lambda)$ is in (3.3), and

$$\begin{aligned}
\sigma^{*2}_n(\lambda) = {} & \left(\frac{\lambda - \lambda_0}{m} \right)^2 \frac{1}{R(m-1)} \left\{ \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right)^2 (Z^*_r \delta_{m0})'(Z^*_r \delta_{m0}) \right. \\
& \left. - \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} (Z^*_r \delta_{m0})' Z^*_r \left(\sum_{r=1}^{R} Z^{*'}_r Z^*_r \right)^{-1} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} Z^{*'}_r (Z^*_r \delta_{m0}) \right\} \\
& + \frac{\sigma_0^2}{R(m-1)} \sum_{r=1}^{R} m_r(0) \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2.
\end{aligned} \tag{A.3}$$

Let $Q_{c,n}(\lambda) = \max_{\beta,\sigma^2} \mathrm{E}(\ln L_{w,n}(\beta, \sigma^2, \lambda))$. It follows that

$$Q_{c,n}(\lambda) = c_1 + \sum_{r=1}^{R} m_r(0) \ln(m_r(\lambda)) - \frac{(R(m-1))}{2} \ln \sigma^{*2}_n(\lambda), \tag{A.4}$$

where c_1 is a constant.

The $\sigma_n^{*2}(\lambda)$ in (A.3) is related to $\hat{\sigma}_n^2(\lambda)$ in (3.4) by replacing $Y_r^{*'}Y_r^*$ and Y_r^* by its expected values. Eq. (3.4) can be rewritten as

$$\hat{\sigma}_n^2(\lambda) = \sigma_n^{*2}(\lambda) + \frac{2}{R(m-1)}[q_{n1}(\lambda) - q_{n2}(\lambda) + h_n(\lambda)], \tag{A.5}$$

where

$$q_{n1}(\lambda) = \frac{1}{2}\sum_{r=1}^{R}\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2 [\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0)], \tag{A.6}$$

$$q_{n2}(\lambda) = \frac{1}{2}\sum_{r=1}^{R}\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)\varepsilon_r' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'}Z_r^*\right)^{-1} \sum_{r=1}^{R}\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right) Z_r^{*'}\varepsilon_r, \tag{A.7}$$

and

$$\begin{aligned} h_n(\lambda) = (\lambda - \lambda_0)\Bigg\{ &\sum_{r=1}^{R} \frac{m_r(\lambda)}{(m_r(\lambda_0))^2}(Z_r^*\delta_{m0})'\varepsilon_r \\ &- \sum_{r=1}^{R}\frac{1}{m_r(\lambda_0)}(Z_r^*\delta_{m0})' Z_r^*\left(\sum_{r=1}^{R} Z_r^{*'}Z_r^*\right)^{-1}\sum_{r=1}^{R}\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right) Z_r^{*'}\varepsilon_r\Bigg\}. \end{aligned} \tag{A.8}$$

Their derivatives are

$$\begin{aligned} \frac{\partial \sigma_n^{*2}(\lambda)}{\partial\lambda} = \frac{2}{R(m-1)}\Bigg\{ &\frac{(\lambda-\lambda_0)}{m^2}\Bigg[\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 (Z_r^*\delta_{m0})'(Z_r^*\delta_{m0}) \\ &- \sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})' Z_r^* \\ &\times \left(\sum_{r=1}^{R} Z_r^{*'}Z_r^*\right)^{-1}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)} Z_r^{*'}(Z_r^*\delta_{m0})\Bigg] + \sigma_0^2\sum_{r=1}^{R}\frac{m_r(0)m_r(\lambda)}{(m_r(\lambda_0))^2}\Bigg\} \end{aligned} \tag{A.9}$$

from (A.3), and

$$\frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial\lambda} = \frac{\partial \sigma_n^{*2}(\lambda)}{\partial\lambda} + \frac{2}{R(m-1)}\left(\frac{\partial q_{n1}(\lambda)}{\partial\lambda} - \frac{\partial q_{n2}(\lambda)}{\partial\lambda} + \frac{\partial h_n(\lambda)}{\partial\lambda}\right), \tag{A.10}$$

from (A.5), where

$$\frac{\partial q_{n1}(\lambda)}{\partial\lambda} = \sum_{r=1}^{R}\frac{m_r(\lambda)}{(m_r(\lambda_0))^2}(\varepsilon_r' J_r\varepsilon_r - \sigma_0^2 m_r(0)), \tag{A.11}$$

$$\frac{\partial q_{n2}(\lambda)}{\partial\lambda} = \sum_{r=1}^{R}\frac{1}{m_r(\lambda_0)}\varepsilon_r' Z_r^*\left(\sum_{r=1}^{R} Z_r^{*'}Z_r^*\right)^{-1}\sum_{r=1}^{R}\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right) Z_r^{*'}\varepsilon_r \tag{A.12}$$

from (A.6) and (A.7), and

$$\begin{aligned}
&\frac{\partial h_n(\lambda)}{\partial \lambda} \\
&= \left\{ \sum_{r=1}^{R} \frac{m_r(\lambda)}{(m_r(\lambda_0))^2} (Z_r^* \delta_{m0})' \varepsilon_r - \sum_{r=1}^{R} \frac{1}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \right. \\
&\quad \left. \times \sum_{r=1}^{R} \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right) Z_r^{*'} \varepsilon_r \right\} \\
&\quad + (\lambda - \lambda_0) \left[\sum_{r=1}^{R} \frac{1}{(m_r(\lambda_0))^2} (Z_r^* \delta_{m0})' \varepsilon_r \right. \\
&\quad \left. - \sum_{r=1}^{R} \frac{1}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} \left(\frac{1}{m_r(\lambda_0)} \right) Z_r^{*'} \varepsilon_r \right]
\end{aligned} \tag{A.13}$$

from (A.8).

The stochastic orders of the terms (A.5)–(A.13) can be derived from central limit theorems. Those orders shall be uniform in $\lambda \in \Lambda$. Under Assumptions 1 and 3, the Lyapounov CLT gives that $\frac{1}{\sqrt{n}} \sum_{r=1}^{R} Z_r^{*'} \varepsilon_r = O_P(1)$. Under Assumption 3, Lyapounov's CLT also implied that $\frac{1}{\sqrt{n}} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} Z_r^{*'} \varepsilon_r = O_P(1)$ and $\frac{1}{\sqrt{n}} \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right)^2 Z_r^{*'} \varepsilon_r = O_P(1)$, because $\left\{ \frac{m}{m_r(\lambda_0)} \right\}$ is bounded. For uniform order, we note that

$$\sum_{r=1}^{R} \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right) Z_r^{*'} \varepsilon_r = \sum_{r=1}^{R} Z_r^{*'} \varepsilon_r + \frac{(\lambda - \lambda_0)}{m} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} Z_r^{*'} \varepsilon_r \tag{A.14}$$

and

$$\sum_{r=1}^{R} \frac{m_r(\lambda)}{(m_r(\lambda_0))^2} Z_r^{*'} \varepsilon_r = \frac{1}{m} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} Z_r^{*'} \varepsilon_r + \frac{(\lambda - \lambda_0)}{m^2} \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right)^2 Z_r^{*'} \varepsilon_r. \tag{A.15}$$

As λ appears linearly in (A.14) and (A.15) and is in the compact set Λ, uniform orders for those terms (A.5)–(A.13) follow. Therefore, under Assumptions 1–4, $\frac{q_{n2}(\lambda)}{n} = O_p\left(\frac{1}{n}\right)$, $\frac{h_n(\lambda)}{n} = O_P\left(\frac{1}{m\sqrt{n}}\right)$, $\frac{1}{n} \frac{\partial q_{n2}(\lambda)}{\partial \lambda} = O_P\left(\frac{1}{mn}\right)$, and $\frac{1}{n} \frac{\partial h_n(\lambda)}{\partial \lambda} = O_P\left(\frac{1}{m\sqrt{n}}\right)$, uniformly on Λ.[21] For quadratic forms, the CLT of Kelejian and Prucha (2001) is applicable. It implies that $\frac{1}{\sqrt{n}} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} (\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0)) = O_P(1)$ and $\frac{1}{\sqrt{n}} \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right)^2 (\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0)) = O_P(1)$. For uniform order of $\frac{q_{n1}(\lambda)}{n}$ and $\frac{1}{n} \frac{\partial q_{n1}(\lambda)}{\partial \lambda}$ in (A.6) and (A.11), we note that

$$\sum_{r=1}^{R} \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2 (\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0))$$

[21] A sequence $\{f_n(\lambda)\}$ of random functions is said to be of order $O_p(b_n)$ if $\frac{1}{b_n} \sup_{\lambda \in \Lambda} |f_n(\lambda)| = O_p(1)$.

$$= \sum_{r=1}^{R}(\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0)) + 2\frac{(\lambda-\lambda_0)}{m}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0))$$
$$+ \left(\frac{\lambda-\lambda_0}{m}\right)^2 \sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2(\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0)), \quad \text{(A.16)}$$

and

$$\sum_{r=1}^{R}\frac{m_r(\lambda)}{(m_r(\lambda_0))^2}(\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0)) = \frac{1}{m}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0))$$
$$+ \frac{(\lambda-\lambda_0)}{m^2}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2(\varepsilon_r' J_r \varepsilon_r - \sigma_0^2 m_r(0)). \quad \text{(A.17)}$$

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

With the uniform stochastic orders for these terms, it follows from (A.5) and (A.10) that

$$\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda) = O_p\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \frac{\partial\hat{\sigma}_n^2(\lambda)}{\partial\lambda} - \frac{\partial\sigma_n^{*2}(\lambda)}{\partial\lambda} = O_p\left(\frac{1}{m\sqrt{n}}\right), \quad \text{(A.18)}$$

uniformly in $\lambda \in \Lambda$. From (A.3) and (A.9),

$$\sigma_n^{*2}(\lambda) = O(1), \quad \frac{\partial\sigma_n^{*2}(\lambda)}{\partial\lambda} = O\left(\frac{1}{m}\right) \quad \text{(A.19)}$$

uniformly on Λ. Note that under Assumption 5, $O\left(\frac{1}{m\sqrt{n}}\right)$ is smaller than $O\left(\frac{1}{m^2}\right)$ because $\frac{1}{m\sqrt{n}} = \frac{1}{m^2}\sqrt{\frac{m}{R}}$. With respect to leading terms, (A.3) and (A.10) have

$$\sigma_n^{*2}(\lambda) = \frac{\sigma_0^2}{R(m-1)}\sum_{r=1}^{R} m_r(0)\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2 + O\left(\frac{1}{m^2}\right) \quad \text{(A.20)}$$

and

$$\frac{\partial\hat{\sigma}_n^2(\lambda)}{\partial\lambda} = \frac{2\sigma_0^2}{R(m-1)}\sum_{r=1}^{R}\frac{m_r(0)m_r(\lambda)}{(m_r(\lambda_0))^2} + O\left(\frac{1}{m^2}\right), \quad \text{(A.21)}$$

uniformly in $\lambda \in \Lambda$ under Assumption 5. We note that (A.3) implies that

$$\sigma_n^{*2}(\lambda) \geqslant \frac{\sigma_0^2}{R(m-1)}\sum_{r=1}^{R} m_r(0)\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2 > c_b^2\sigma_0^2, \quad \text{(A.22)}$$

where $c_b > 0$ is a lower bound of $\frac{m_r(\lambda)}{m_r(\lambda_0)}$ on Λ for all r. The $\frac{m_r(\lambda)}{m_r(\lambda_0)}$ is increasing in λ. If $\{m_r\}$ is a bounded sequence, $\frac{m_L-1+\lambda_L}{m_U-1+\lambda_0} \leqslant \frac{m_r(\lambda)}{m_r(\lambda_0)} \leqslant \frac{m_U-1+\lambda_U}{m_L-1+\lambda_0}$ where m_L and m_U are, respectively, the largest lower and least upper bounds of m_r, and λ_L and λ_U are, respectively, the largest lower and least upper bounds of Λ. On the other hand, if m tends to infinity, $\frac{m_r(\lambda)}{m_r(\lambda_0)}$ converges to 1 uniformly in λ. Therefore, (A.22) shows that $\sigma_n^{*2}(\lambda)$ is uniformly bounded away from zero on Λ.[22] As $\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda) = o_P(1)$ uniformly in $\lambda \in \Lambda$ from (A.18), $\hat{\sigma}_n^2(\lambda)$ is also uniformly bounded away from zero on Λ in probability. By the mean value theorem, (3.5) and (A.4) imply

$$
\begin{aligned}
&\frac{m^2}{n}\{[\ln L_{c,n}(\lambda) - \ln L_{c,n}(\lambda_0)] - [Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)]\} \\
&\quad = -\left(\frac{R(m-1)}{2}\right)\left(\frac{m^2}{n}\right)\left[\frac{\partial \ln \hat{\sigma}_n^2(\bar{\lambda})}{\partial \lambda} - \frac{\partial \ln \sigma_n^{*2}(\bar{\lambda})}{\partial \lambda}\right](\lambda - \lambda_0) \\
&\quad = -\left(\frac{R(m-1)}{2}\right)\left(\frac{m^2}{n}\right)\frac{(\lambda-\lambda_0)}{\hat{\sigma}_n^2(\bar{\lambda})\sigma_n^{*2}(\bar{\lambda})}\left[\sigma_n^{*2}(\bar{\lambda})\left(\frac{\partial \hat{\sigma}_n^2(\bar{\lambda})}{\partial \lambda} - \frac{\partial \sigma_n^{*2}(\bar{\lambda})}{\partial \lambda}\right)\right. \\
&\qquad \left. - \frac{\partial \sigma_n^{*2}(\bar{\lambda})}{\partial \lambda}(\hat{\sigma}_n^2(\bar{\lambda}) - \sigma_n^{*2}(\bar{\lambda}))\right],
\end{aligned} \tag{A.23}
$$

where $\bar{\lambda}$ lies between λ and λ_0. As $\frac{R(m-1)m^2}{n}\sigma_n^{*2}(\lambda)\left(\frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda} - \frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda}\right) = \mathrm{O_P}\left(\frac{m}{\sqrt{n}}\right)$ and

$$
\frac{R(m-1)m^2}{n}\frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda}(\hat{\sigma}_n^2(\lambda) - \sigma_n^{*2}(\lambda)) = \mathrm{O_P}\left(\frac{m}{\sqrt{n}}\right)
$$

uniformly in $\lambda \in \Lambda$ from (A.18) and (A.19), it follows that

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

uniformly on Λ. The second equality holds under Assumption 5 because $\frac{m}{\sqrt{n}} = \sqrt{\frac{m}{R}} \to 0$. $\square$

A.4. Motivation of identification conditions in Assumptions 6.1 and 6.2

Assumptions 6.1 and 6.2 are needed to guarantee the limit of $\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))$ satisfies the identification uniqueness condition (White, 1994). As shown in the subsequent proof of Proposition 2, $\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))$ can be bounded:

$$
\begin{aligned}
&\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)) \\
&\quad \leqslant -\frac{(\lambda-\lambda_0)^2}{2n\sigma_n^{*2}(\lambda)}\left\{\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 (Z_r^*\delta_{m0})'(Z_r^*\delta_{m0})\right. \\
&\qquad \left. - \sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})'Z_r^*\left(\sum_{r=1}^{R} Z_r^{*'}Z_r^*\right)^{-1}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}Z_r^*\delta_{m0}\right\} + D_r(\lambda),
\end{aligned} \tag{A.24}
$$

where

$$
D_n(\lambda) = \frac{m^2}{n}\left[\max_{\beta,\sigma^2}\mathrm{E}_{\beta_0=0}(\ln L_{w,n}(\beta,\sigma^2,\lambda)) - \max_{\beta,\sigma^2}\mathrm{E}_{\beta_0=0}(\ln L_{w,n}(\beta,\sigma^2,\lambda_0))\right], \tag{A.25}
$$

where $E_{\beta_0=0}$ is the expectation operator under $\beta_0 = 0$, σ_0^2 and λ_0 being the 'true' parameters. The identification of the parameters can be based on the two components of the upper bound in (A.24). Each of two components is negative. The first term is negative for any $\lambda \neq \lambda_0$ by the generalized Schwartz inequality. The second component, $D_n(\lambda)$, can

[22] A sequence of (positive) random functions $\{h_n(\lambda)\}$ is said to be uniformly bounded away from zero in probability if there exists a positive constant $c>0$ such that $\lim_{n\to\infty} \mathrm{P}(\inf_{\lambda\in\Lambda} f_n(\lambda) > c) = 1$.

be negative from the information inequality. Each of them provides a source of identification. In order that the positiveness will not be lost in the limit, we need the Assumptions 6.1 and 6.2.

Assumption 6.1 can break down if $\beta_{10} = 0$ or $x_{ri,2} = x_{ri,1}$ for all r and i for the case with $m \to \infty$. If $\beta_{10} = 0$, $z_{ri}\delta_{m0}$ will effectively consist only of the contextual component $-\frac{1}{m_r(0)}(x_{ri,2} - \bar{x}_{r,2})\beta_{20}$, which is of order $O(\frac{1}{m})$ and goes to zero as $m \to \infty$. If $x_{r2} = x_{r1}$, under the setting of Assumption 2, because $\frac{m}{m_r(\lambda_0)} \to \frac{1}{a_r}$ as $m \to \infty$, $\left(\frac{m}{m_r(\lambda_0)}\right) Z_r \delta_{m0}$ will be approximated by $\frac{1}{a_r} X_{r1}\beta_{10}$. On the other hand, the second component of Z_r will be approximated by $\frac{1}{a_r} X_{r1}$ as $X_{r2} = X_{r1}$. Hence, under such circumstances, Assumption 6.1 may fail.

The second component $D_n(\lambda)$ in (A.25) captures the covariance structure of the disturbance term in (2.5). The $D_n(\lambda) \leqslant 0$ follows because the information inequality gives $\mathrm{E}_{\beta_0=0}(\ln L_{w,n}(\beta, \sigma^2, \lambda)) \leqslant \mathrm{E}_{\beta_0=0}(\ln L_{w,n}(\beta_0, \sigma_0^2, \lambda_0))$ for all β, σ^2 and λ, and, hence

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

for all λ. Alternatively, as shown in the proof of Proposition 2, $D_n(\lambda)$ can be explicitly written as

$$\begin{aligned} D_n(\lambda) = \frac{(R(m-1))}{2n} m^2 \Bigg[& \sum_{r=1}^{R} \left(\frac{m_r(0)}{R(m-1)} \right) \ln \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2 \\ & - \ln \left(\sum_{r=1}^{R} \frac{m_r(0)}{R(m-1)} \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2 \right) \Bigg], \end{aligned} \tag{A.26}$$

which is strictly negative for any $\lambda \neq \lambda_0$. This is so because the Jensen inequality can be applied to the logarithmic function as long as m_r's are not identical to each other so that $\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)$'s vary across r. Assumption 6.2 is needed to guarantee that the negativeness of $D_n(\lambda)$ does not vanish in the limit. Under the identification condition in Assumption 6.2, λ_0 is the unique maximum of $\lim_{n\to\infty} D_n(\lambda)$ in Λ.

Proof of Proposition 2. Define

$$\sigma_n^2(\lambda) = \frac{\sigma_0^2}{R(m-1)} \sum_{r=1}^{R} m_r(0) \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2, \tag{A.27}$$

which is the leading term of $\sigma_n^{*2}(\lambda)$ in (A.20).

Because $\sigma_n^{*2}(\lambda_0) = \sigma_0^2$ from (A.3), (A.4) implies

$$\begin{aligned} Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0) = & \sum_{r=1}^{R} m_r(0) \ln(m_r(\lambda)) - \frac{(R(m-1))}{2} \ln \sigma_n^{*2}(\lambda) \\ & - \sum_{r=1}^{R} m_r(0) \ln(m_r(\lambda_0)) + \frac{(R(m-1))}{2} \ln \sigma_0^2. \end{aligned} \tag{A.28}$$

Define

$$D_n(\lambda) = \frac{m^2}{n}\left\{\sum_{r=1}^{R} m_r(0)[\ln(m_r(\lambda)) - \ln(m_r(\lambda_0))] - \left(\frac{R(m-1)}{2}\right)[\ln\sigma_n^2(\lambda) - \ln\sigma_0^2]\right\}. \tag{A.29}$$

Hence,

$$\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)) = D_n(\lambda) - \frac{m^2}{n}\left(\frac{R(m-1)}{2}\right)(\ln\sigma_n^{*2}(\lambda) - \ln\sigma_n^2(\lambda)). \tag{A.30}$$

The $D_n(\lambda)$ has the relation with the concentrated log likelihood in (A.25). This is so as follows. When $\beta_0 = 0$, $\delta_{m0} = 0$. Therefore, from (A.3), $\sigma_n^{*2}(\lambda)|_{\beta_0=0} = \sigma_n^2(\lambda)$ in (A.27). Furthermore, from (3.6),

$$Q_{c,n}(\lambda)|_{\beta_0=0} = c + \sum_{r=1}^{R} m_r(0)\ln(m_r(\lambda)) - \frac{(R(m-1))}{2}\ln\sigma_n^2(\lambda). \tag{A.31}$$

Note that $\sigma_n^2(\lambda_0) = \sigma_0^2$ and $Q_{c,n}(\lambda)|_{\beta_0=0} = \max_{\beta,\sigma^2} \mathrm{E}_{\beta_0=0}(\ln L_{w,n}(\beta,\sigma^2,\lambda))$. Hence,

$$D_n(\lambda) = \frac{m^2}{n}\left[\max_{\beta,\sigma^2} \mathrm{E}_{\beta_0=0}(\ln L_{w,n}(\beta,\sigma^2,\lambda)) - \max_{\beta,\sigma^2} \mathrm{E}_{\beta_0=0}(\ln L_{w,n}(\beta,\sigma^2,\lambda_0))\right]. \tag{A.32}$$

The $D_n(\lambda)$ can be rewritten as

$$D_n(\lambda) = \frac{R(m-1)}{2n}m^2\left[\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\ln\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2 - \ln\left(\sum_{r=1}^{R}\frac{m_r(0)}{R(m-1)}\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2\right)\right]. \tag{A.33}$$

Eqs. (A.3) and (A.27) imply that

$$\begin{aligned}\frac{m^2}{n}(R(m-1))(\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda)) = \frac{(\lambda-\lambda_0)^2}{n}&\left\{\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 (Z_r^*\delta_{m0})'(Z_r^*\delta_{m0})\right.\\ &- \sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})'Z_r^*\left(\sum_{r=1}^{R} Z_r^{*'}Z_r^*\right)^{-1}\\ &\left.\times\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}(Z_r^*\delta_{m0})\right\},\end{aligned} \tag{A.34}$$

which is non-negative for any $\lambda \neq \lambda_0$ by the generalized Schwartz inequality. This implies, in particular, $\sigma_n^{*2}(\lambda) \geqslant \sigma_n^2(\lambda)$. By the mean value theorem, $\ln\sigma_n^{*2}(\lambda) - \ln\sigma_n^2(\lambda) = (\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda))/\bar{\sigma}_n^2(\lambda)$, where $\sigma_n^{*2}(\lambda) \geqslant \bar{\sigma}_n^2(\lambda) \geqslant \sigma_n^2(\lambda)$. As $\bar{\sigma}_n^2(\lambda) \leqslant \sigma_n^{*2}(\lambda)$ and $\sigma_n^{*2}(\lambda) \geqslant \sigma_n^2(\lambda)$, it follows that

$$\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0)) \leqslant D_n(\lambda) - \frac{m^2}{n}\left(\frac{R(m-1)}{2}\right)(\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda))/\sigma_n^{*2}(\lambda). \tag{A.35}$$

Under Assumption 6.1, (A.34) shows that, for any open neighborhood $N_\varepsilon(\lambda_0)$ of λ_0,

$$\liminf_{n\to\infty} \min_{\lambda\in\bar{N}_\varepsilon(\lambda_0)} \frac{m^2}{n}(R(m-1))(\sigma_n^{*2}(\lambda) - \sigma_n^2(\lambda)) > 0. \tag{A.36}$$

That implies $\frac{m^2}{n}(Q_{c,n}(\lambda) - Q_{c,n}(\lambda_0))$ in (A.30) satisfies the identification uniqueness condition because $D_n(\lambda) \leqslant 0$ for all λ from (A.33) by Jensen's inequality and $\sigma_n^{*2}(\lambda) = O(1)$ uniformly in $\lambda \in \Lambda$.

In the case of Assumption 6.2, as $D_n(\lambda_0) = 0$, the identification uniqueness condition for $D_n(\lambda)$ will be satisfied if $D_n(\lambda)$ is uniformly equicontinuous on Λ. From (A.27) and (A.29), the derivative of $D_n(\lambda)$ is

$$\frac{\partial D_n(\lambda)}{\partial \lambda} = \frac{m^2}{n}\left\{\sum_{r=1}^{R}\frac{m_r(0)}{m_r(\lambda)} - \frac{R(m-1)}{\sum_{r=1}^{R} m_r(0)\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2}\sum_{r=1}^{R}\frac{m_r(0)m_r(\lambda)}{(m_r(\lambda_0))^2}\right\}, \tag{A.37}$$

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

$$\begin{aligned}\frac{\partial^2 D_n(\lambda)}{\partial \lambda^2} = & -\frac{1}{R}\sum_{r=1}^{R}\frac{mm_r(0)}{(m_r(\lambda))^2} - \frac{1}{\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2}\frac{1}{R}\sum_{r=1}^{R}\frac{mm_r(0)}{(m_r(\lambda_0))^2} \\ & + \frac{2}{\left[\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\left(\frac{m_r(\lambda)}{m_r(\lambda_0)}\right)^2\right]^2}\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\frac{mm_r(\lambda)}{(m_r(\lambda_0))^2}\cdot\frac{1}{R}\sum_{r=1}^{R}\frac{m_r(0)m_r(\lambda)}{(m_r(\lambda_0))^2},\end{aligned} \tag{A.38}$$

which is uniformly bounded on Λ. By the mean value theorem, $\frac{\partial D_n(\lambda)}{\partial \lambda} = \frac{\partial^2 D_n(\bar{\lambda}_n)}{\partial \lambda^2}(\lambda - \lambda_0)$. Thus, $\sup_{\lambda\in\Lambda}|\frac{\partial D_n(\lambda)}{\partial \lambda}| = O(1)$. In turn, for any λ_1 and $\lambda_2 \in \Lambda$, this and the mean value theorem imply $|D_n(\lambda_1) - D_n(\lambda_2)| = |\frac{\partial D_n(\lambda^*)}{\partial \lambda}(\lambda_1 - \lambda_2)| = O(1)|\lambda_1 - \lambda_2|$. Hence, the uniform equicontinuity of $D_n(\lambda)$ follows. 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). □

A.5. Local identification

From the expression of $Q_{c,n}(\lambda)$ in (A.30), it follows that

$$\frac{m^2}{n}\frac{\partial Q_{c,n}(\lambda)}{\partial \lambda} = \frac{\partial D_n(\lambda)}{\partial \lambda} - \frac{m^2}{n}\left(\frac{R(m-1)}{2}\right)\left(\frac{1}{\sigma_n^{*2}(\lambda)}\frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda} - \frac{1}{\sigma_n^2(\lambda)}\frac{\partial \sigma_n^2(\lambda)}{\partial \lambda}\right) \tag{A.39}$$

and

$$\begin{aligned}\frac{m^2}{n}\frac{\partial^2 Q_{c,n}(\lambda)}{\partial \lambda^2} = & \frac{\partial^2 D_n(\lambda)}{\partial \lambda^2} - \frac{m^2}{n}\left(\frac{R(m-1)}{2}\right)\left\{\frac{1}{\sigma_n^{*2}(\lambda)}\frac{\partial^2 \sigma_n^{*2}(\lambda)}{\partial \lambda^2} - \frac{1}{\sigma_n^2(\lambda)}\frac{\partial^2 \sigma_n^2(\lambda)}{\partial \lambda^2}\right. \\ & \left. - \frac{1}{\sigma_n^{*4}(\lambda)}\left(\frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda}\right)^2 + \frac{1}{\sigma_n^4(\lambda)}\left(\frac{\partial \sigma_n^2(\lambda)}{\partial \lambda}\right)^2\right\}.\end{aligned} \tag{A.40}$$

From (A.27), the derivatives of $\sigma_n^2(\lambda)$ are

$$\frac{\partial \sigma_n^2(\lambda)}{\partial \lambda} = \frac{2\sigma_0^2}{R(m-1)}\sum_{r=1}^{R}\frac{m_r(0)m_r(\lambda)}{(m_r(\lambda_0))^2} \tag{A.41}$$

and

$$\frac{\partial^2 \sigma_n^2(\lambda)}{\partial \lambda^2} = \frac{2\sigma_0^2}{R(m-1)} \sum_{r=1}^{R} \frac{m_r(0)}{(m_r(\lambda_0))^2}. \tag{A.42}$$

The $\frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda}$ is in (A.9), and

$$\begin{aligned} \frac{\partial^2 \sigma_n^{*2}(\lambda)}{\partial \lambda^2} = \frac{2}{R(m-1)} \Bigg\{ \frac{1}{m^2} \Bigg[& \sum_{r=1}^{r} \left(\frac{m}{m_r(\lambda_0)} \right)^2 (Z_r^* \delta_{m0})'(Z_r^* \delta_{m0}) \\ & - \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \\ & \times \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} Z_r^{*'} (Z_r^* \delta_{m0}) \Bigg] + \sigma_0^2 \sum_{r=1}^{R} \frac{m_r(0)}{(m_r(\lambda_0))^2} \Bigg\}. \end{aligned} \tag{A.43}$$

At λ_0, $\sigma_n^2(\lambda_0) = \sigma_n^{*2}(\lambda_0) = \sigma_0^2$ and $\frac{\partial \sigma_n^2(\lambda_0)}{\partial \lambda} = \frac{\partial \sigma_n^{*2}(\lambda_0)}{\partial \lambda} = \frac{2\sigma_0^2}{R(m-1)} \sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda_0)}$. The second order derivatives of $\sigma_n^2(\lambda)$ and $\sigma_n^{*2}(\lambda)$ do not depend on λ. Therefore,

$$\frac{m^2}{n} \frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial \lambda^2} = \frac{\partial^2 D_n(\lambda_0)}{\partial \lambda^2} - \frac{m^2}{n} \left(\frac{R(m-1)}{2} \right) \left[\frac{1}{\sigma_n^{*2}(\lambda_0)} \frac{\partial^2 \sigma_n^{*2}(\lambda_0)}{\partial \lambda^2} - \frac{1}{\sigma_n^2(\lambda)} \frac{\partial \sigma_n^2(\lambda_0)}{\partial \lambda^2} \right]. \tag{A.44}$$

The expression of $\frac{\partial^2 D_n(\lambda)}{\partial \lambda^2}$ is in (A.38). These together give the expressions that

$$\begin{aligned} \frac{m^2}{n} \frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial \lambda^2} = \frac{\partial^2 D_n(\lambda_0)}{\partial \lambda^2} - \frac{1}{n\sigma_0^2} \Bigg\{ & \sum_{r=1}^{R} \left(\frac{m}{m_r(\lambda_0)} \right)^2 (Z_r^* \delta_{m0})'(Z_r^* \delta_{m0}) \\ & - \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} Z_r^{*'} (Z_r^* \delta_{m0}) \Bigg\}, \end{aligned} \tag{A.45}$$

where

$$\begin{aligned} \frac{\partial^2 D_n(\lambda_0)}{\partial \lambda^2} &= -2 \frac{m^2}{n} \left[\sum_{r=1}^{R} \frac{m_r(0)}{(m_r(\lambda_0))^2} - \frac{1}{R(m-1)} \left(\sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda_0)} \right)^2 \right] \\ &= -2 \frac{m-1}{m} \sum_{r=1}^{R} \left(\frac{m_r(0)}{R(m-1)} \right) \left[\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R} \left(\frac{m_s(0)}{R(m-1)} \right) \frac{m}{m_s(\lambda_0)} \right]^2. \end{aligned} \tag{A.46}$$

The $\left(\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R} \left(\frac{m_s(0)}{R(m-1)} \right) \frac{m}{m_s(\lambda_0)} \right)$ is the deviation of $\frac{m}{m_r(\lambda_0)}$, $r = 1, \ldots, R$, from its weighted mean of the R groups. These provide the sufficient condition for local identification in (3.7).

Proof of Proposition 3. The score vector $\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda}$ shall first be derived.

The derivative of $\ln L_{c,n}(\lambda)$ in (3.5) is

$$\frac{\partial \ln L_{c,n}(\lambda)}{\partial \lambda} = \sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda)} - \frac{(R(m-1))}{2\hat{\sigma}_n^2(\lambda)} \frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda}. \tag{A.47}$$

The derivative of $\hat{\sigma}_n^2(\lambda)$ in (A.5) is

$$\frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda} = \frac{2}{R(m-1)} \left\{ \sum_{r=1}^{R} \frac{m_r(\lambda)}{(m_r(\lambda_0))^2} (Z_r \delta_{m0} + \varepsilon_r)' J_r (Z_r \delta_{m0} + \varepsilon_r) \right.$$
$$\left. - \sum_{r=1}^{R} \frac{1}{m_r(\lambda_0)} (Z_r^* \delta_{m0} + \varepsilon_r)' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} \frac{m_r(\lambda)}{m_r(\lambda_0)} Z_r^{*'} (Z_r^* \delta_{m0} + \varepsilon_r) \right\}. \tag{A.48}$$

At $\lambda = \lambda_0$, $\frac{\partial \hat{\sigma}_n^2(\lambda_0)}{\partial \lambda} = \frac{2}{R(m-1)}(h_{n1} + q_{n1} - h_{n2} - q_{n2})$, where

$$h_{n1} = \sum_{r=1}^{R} \frac{1}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' \varepsilon_r^*, \quad h_{n2} = \sum_{r=1}^{R} \frac{1}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} Z_r^{*'} \varepsilon_r^* \tag{A.49}$$

and

$$q_{n1} = \sum_{r=1}^{R} \frac{1}{m_r(\lambda_0)} \varepsilon_r^{*'} \varepsilon_r^*, \quad q_{n2} = \sum_{r=1}^{R} \frac{1}{m_r(\lambda_0)} \varepsilon_r^{*'} Z_r^* \left(\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right)^{-1} \sum_{r=1}^{R} Z_r^{*'} \varepsilon_r^*. \tag{A.50}$$

Hence,

$$\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda} = \frac{1}{\hat{\sigma}_n^2(\lambda_0)} \left\{ (\hat{\sigma}_n^2(\lambda_0) - \sigma_0^2) \sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda_0)} - h_{n1} + h_{n2} \right.$$
$$\left. - \left(q_{n1} - \sigma_0^2 \sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda_0)} \right) + q_{n2} \right\}. \tag{A.51}$$

The order of $\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda}$ will depend on the orders of the various items in (A.51). Under Assumptions 1,2,3, Lyapounov's CLT gives

(i) $h_{n1} = \frac{\sqrt{n}}{m} \frac{1}{\sqrt{n}} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' \varepsilon_r^* = O_P\left(\frac{\sqrt{n}}{m}\right)$,
(ii) $h_{n2} = \frac{\sqrt{n}}{m} \frac{1}{n} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} (Z_r^* \delta_{m0})' Z_r^* \left[\frac{1}{n} \sum_{r=1}^{R} Z_r^{*'} Z_r^*\right]^{-1} \frac{1}{\sqrt{n}} \sum_{r=1}^{R} Z_r^{*'} \varepsilon_r^* = O\left(\frac{\sqrt{n}}{m}\right)$,
(iii) $q_{n2} = \frac{1}{m\sqrt{n}} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} \varepsilon_r^{*'} Z_r^* \left[\frac{1}{n} \sum_{r=1}^{R} Z_r^{*'} Z_r^*\right]^{-1} \frac{1}{\sqrt{n}} \sum_{r=1}^{R} Z_r^{*'} \varepsilon_r^* = O_P\left(\frac{1}{m}\right)$.

Note that at λ_0, (A.5) can be simplified to

$$\hat{\sigma}_n^2(\lambda_0) = \frac{1}{R(m-1)} \left\{ \sum_{r=1}^{R} \varepsilon_r^{*'} \varepsilon_r^* - \sum_{r=1}^{R} \varepsilon_r^{*'} Z_r^* \left[\sum_{r=1}^{R} Z_r^{*'} Z_r^* \right]^{-1} \sum_{r=1}^{R} Z_r^{*'} \varepsilon_r^* \right\}$$
$$= \frac{1}{R(m-1)} \sum_{r=1}^{R} \varepsilon_r^{*'} \varepsilon_r^* + O_P\left(\frac{1}{n}\right), \tag{A.52}$$

from (3.4), because $\frac{R(m-1)}{n} = \frac{m-1}{m} = O(1)$, which means that $(R(m-1))$ and n are of the same order. As $\frac{1}{R(m-1)}\sum_{r=1}^{R}\varepsilon_r^{*'}\varepsilon_r^{*} - \sigma_0^2 = \frac{\sqrt{n}}{R(m-1)}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}(\varepsilon_r^{*'}\varepsilon_r^{*} - \sigma_0^2 m_r(0)) = O_P\left(\frac{1}{\sqrt{n}}\right)$, 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 $\frac{\sqrt{n}}{R(m-1)} = \left(\frac{m}{m-1}\right)\frac{1}{\sqrt{n}}$, one has

(iv) $\sqrt{n}(\hat{\sigma}_n^2(\lambda_0) - \sigma_0^2) = \sqrt{n}\left(\frac{1}{R(m-1)}\sum_{r=1}^{R}\varepsilon_r' J_r\varepsilon_r - \sigma_0^2\right) + O_P\left(\frac{1}{\sqrt{n}}\right) = \left(\frac{m}{m-1}\right)\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\left[\varepsilon_r^{*'}\varepsilon_r^{*} - m_r(0)\sigma_0^2\right] + O_P\left(\frac{1}{\sqrt{n}}\right)$, and

(v) $q_{n1} - \sigma_0^2\sum_{r=1}^{R}\frac{\sigma_0^2 m_r(0)}{m_r(\lambda_0)} = \frac{\sqrt{n}}{m}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}\left[\varepsilon_r^{*'}\varepsilon_r^{*} - m_r(0)\sigma_0^2\right] = O_P\left(\frac{\sqrt{n}}{m}\right)$.

(iv) and (v) together imply that

$$\sqrt{n}(\hat{\sigma}_n^2(\lambda_0) - \sigma_0^2)\frac{1}{R}\sum_{r=1}^{R}\frac{m_r(0)}{m_r(\lambda_0)} - \frac{m}{\sqrt{n}}\left(q_{n1} - \sigma_0^2\sum_{r=1}^{R}\frac{m_r(0)}{m_r(\lambda_0)}\right)$$
$$= -\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\left[\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right](\varepsilon_r' J_r\varepsilon_r - m_r(0)\sigma_0^2) + O_P\left(\frac{1}{\sqrt{n}}\right). \quad (A.53)$$

From (A.51) and the above results,

$$\frac{m}{\sqrt{n}}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial\lambda} = \frac{1}{\hat{\sigma}_n^2(\lambda_0)}\left\{\sqrt{n}(\hat{\sigma}_n^2(\lambda_0) - \sigma_0^2)\frac{1}{R}\sum_{r=1}^{R}\frac{m_r(0)}{m_r(\lambda_0)} - \frac{m}{\sqrt{n}}\left(q_{n1} - \sigma_0^2\sum_{r=1}^{R}\frac{m_r(0)}{m_r(\lambda_0)}\right) - \frac{m}{\sqrt{n}}(h_{n1} - h_{n2}) + O\left(\frac{1}{\sqrt{n}}\right)\right\}$$
$$= -(A_n + B_n) + o_P(1), \quad (A.54)$$

where

$$A_n = \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\left[\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right](\varepsilon_r' J_r\varepsilon_r - m_r(0)\sigma_0^2) \quad (A.55)$$

and

$$B_n = \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})'\varepsilon_r - \left(\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})'Z_r^*\right)\left(\sum_{r=1}^{R}Z_r^{*'}Z_r^*\right)^{-1} \times\frac{1}{\sqrt{n}}\sum_{r=1}^{R}Z_r^{*'}\varepsilon_r. \quad (A.56)$$

Under normality of ε_r's, $var(\varepsilon_r' J_r\varepsilon_r) = 2m_r(0)\sigma_0^4$. The variance of A_n is

$$2\sigma_0^2\left(\frac{R(m-1)}{n}\right)\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\left[\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right]^2.$$

Note that $\frac{(R(m-1))}{n}$ can also be written as $\frac{m-1}{m}$. The B_n can be rewritten as

$$B_n = \frac{1}{\sqrt{n}} \delta'_{m0} \left(\frac{m}{m_1(\lambda_0)} Z_1^{*'}, \ldots, \frac{m}{m_R(\lambda_0)} Z_R^{*'} \right)' \left[I_n - \mathscr{Z}_n^* (\mathscr{Z}_n^{*'} \mathscr{Z}_n^*)^{-1} \mathscr{Z}_n^{*'} \right] \mathscr{E}_n, \tag{A.57}$$

where $\mathscr{Z}_n^* = (Z_1^{*'}, \ldots, Z_R^{*'})'$. With this expression, it is immediate that

$$\begin{aligned} var(B_n) = \frac{\sigma_0^2}{n} \delta'_{m0} \left(\frac{m}{m_1(\lambda_0)} Z_1^{*'}, \ldots, \frac{m}{m_R(\lambda_0)} Z_R^{*'} \right)' [I_n - \mathscr{Z}_n^* (\mathscr{Z}_n'^{*} \mathscr{Z}_n^*)^{-1} \mathscr{Z}_n'^{*}] \\ \times \left(\frac{m}{m_1(\lambda_0)} Z_1^*, \ldots, \frac{m}{m_R(\lambda_0)} Z_R^* \right) \delta_{m0}. \end{aligned} \tag{A.58}$$

The $cov(A_n, B_n) = 0$ because the third moment of ε_{ri} is zero. The A_n is a quadratic function of $\mathscr{E}_n$, and B_n is a linear function of $\mathscr{E}_n$. The $\frac{m}{\sqrt{n}} \frac{\partial \ln L_{c,n}(\lambda_0)}{\partial \lambda} \xrightarrow{D} N(0, \Sigma_\lambda)$ follows from the central limit theorem of Kelejian and Prucha (2001) for the linear and quadratic forms.

From (3.5),

$$\frac{m^2}{n} \frac{\partial^2 \ln L_{c,n}(\lambda)}{\partial \lambda^2} = -\frac{1}{R} \sum_{r=1}^R \frac{m m_r(0)}{(m_r(\lambda))^2} - \frac{m(m-1)}{2} \left[\frac{1}{\hat{\sigma}_n^2(\lambda)} \frac{\partial^2 \hat{\sigma}_n^2(\lambda)}{\partial \lambda^2} - \frac{1}{\hat{\sigma}_n^4(\lambda)} \left(\frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda} \right)^2 \right]. \tag{A.59}$$

From (A.18) and (A.3),

$$\begin{aligned} \hat{\sigma}_n^2(\lambda) = \sigma_n^{*2}(\lambda) + O_P\left(\frac{1}{\sqrt{n}}\right) = \frac{\sigma_0^2}{R(m-1)} \sum_{r=1}^R m_r(0) \left(\frac{m_r(\lambda)}{m_r(\lambda_0)} \right)^2 \\ + \left(\frac{\lambda - \lambda_0}{m} \right)^2 \cdot O(1) + O_P\left(\frac{1}{\sqrt{n}}\right) \end{aligned} \tag{A.60}$$

and, from (A.18) and (A.9),

$$\begin{aligned} \frac{\partial \hat{\sigma}_n^2(\lambda)}{\partial \lambda} = \frac{\partial \sigma_n^{*2}(\lambda)}{\partial \lambda} + O_P\left(\frac{1}{m\sqrt{n}}\right) = \frac{2\sigma_0^2}{R(m-1)} \sum_{r=1}^R \frac{m_r(0) m_r(\lambda)}{(m_r(\lambda_0))^2} \\ + \left(\frac{\lambda - \lambda_0}{m^2} \right) \cdot O(1) + O_P\left(\frac{1}{m\sqrt{n}}\right), \end{aligned} \tag{A.61}$$

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

$$m \frac{\partial \hat{\sigma}_n^2(\bar{\lambda}_n)}{\partial \lambda} = 2\sigma_0^2 \left(\frac{m}{m-1} \right) \frac{1}{R} \sum_{r=1}^R \frac{m_r(0)}{m_r(\lambda_0)} + o_P(1). \tag{A.62}$$

It remains to consider the second order derivative $\frac{\partial^2 \hat{\sigma}_n^2(\lambda)}{\partial \lambda^2}$. From (3.4),

$$\begin{aligned} & m(m-1) \frac{\partial^2 \hat{\sigma}_n^2(\lambda)}{\partial \lambda^2} \\ & = 2 \left\{ \frac{1}{R} \sum_{r=1}^R \frac{m}{m_r(0)^2} Y_r^{*'} Y_r^* - \frac{1}{R} \sum_{r=1}^R \frac{Y_r^{*'} Z_r^*}{m_r(0)} \left(\frac{1}{n} \sum_{r=1}^R Z_r^{*'} Z_r^* \right)^{-1} \frac{1}{R} \sum_{r=1}^R \frac{Z_r^{*'} Y_r^*}{m_r(0)}) \right\}, \end{aligned} \tag{A.63}$$

which does not depend on λ. As

$$\frac{1}{R}\sum_{r=1}^{R}\frac{1}{m_r(0)}Z_r^{*'}Y_r^* = \frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}Z_r^*\delta_{m0} + \frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}\varepsilon_r$$
$$= \frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}Z_r^*\delta_{m0} + o_P(1) \tag{A.64}$$

and

$$\frac{1}{R}\sum_{r=1}^{R}\frac{m}{m_r(0)^2}Y_r^{*'}Y_r^*$$
$$= \frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 (Z_r^*\delta_{m0})'(Z_r^*\delta_{m0}) + \frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 \varepsilon_r' J_r\varepsilon_r$$
$$+ \frac{2}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 (Z_r^*\delta_{m0})'\varepsilon_r$$
$$= \frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 (Z_r^*\delta_{m0})'(Z_r^*\delta_{m0}) + \frac{\sigma_0^2}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 m_r(0) + o_P(1), \tag{A.65}$$

it follows that

$$m(m-1)\frac{\partial^2\hat{\sigma}_n^2(\lambda)}{\partial\lambda^2}$$
$$= 2\left\{\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 (Z_r^*\delta_{m0})'(Z_r^*\delta_{m0}) + \frac{\sigma_0^2}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(\lambda_0)}\right)^2 m_r(0)\right.$$
$$\left. -\left(\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})'Z_r^*\right)\left(\sum_{r=1}^{R}Z_r^{*'}Z_r^*\right)^{-1}\frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}Z_r^{*'}(Z_r^*\delta_{m0})\right\} + o_P(1). \tag{A.66}$$

These results and (A.59) together imply

$$\frac{m^2}{n}\frac{\partial^2 \ln L_{c,n}(\bar{\lambda}_n)}{\partial\lambda^2} = \frac{m^2}{n}\frac{\partial^2 Q_{c,n}(\lambda_0)}{\partial\lambda^2} + o_P(1), \tag{A.67}$$

as in (A.45) and (A.46).

By the mean value theorem, $\frac{\sqrt{n}}{m}(\hat{\lambda}_n-\lambda_0) = -\left(\frac{m^2}{n}\frac{\partial^2 \ln L_{c,n}(\bar{\lambda}_n)}{\partial\lambda^2}\right)^{-1}\frac{m}{\sqrt{n}}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial\lambda}$, where $\bar{\lambda}_n$ is a consistent estimate lying between $\hat{\lambda}_n$ and λ_0. The limiting distribution of $\frac{\sqrt{n}}{m}(\hat{\lambda}_n-\lambda_0)$ follows. □

Proof of Proposition 4. Eq. (3.10) implies that

$$\begin{pmatrix}\sqrt{n}(\hat{\beta}_{n1}-\beta_{10})\\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2}-\beta_{20})\end{pmatrix}$$

$$= \left(\frac{1}{n}\sum_{r=1}^{R} Z_r^{*'} Z_r^*\right)^{-1} \left[\frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r^{*'}\varepsilon_r + \frac{1}{n}\sum_{r=1}^{R}(Z_r^{*'} Z_r^*\delta_{m0} + Z_r^{*'}\varepsilon_r)\frac{m}{m_r(\lambda_0)}\cdot\frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0)\right]. \tag{A.68}$$

The term $\frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r^{*'}\varepsilon_r$ is uncorrelated with $\frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0)$. The latter depends on $\frac{m}{\sqrt{n}}\frac{\partial \ln L_{c,n}(\lambda_0)}{\partial\lambda} = -(A_n + B_n) + o_P(1)$ in (A.54) of the proof of Proposition 3. Under normality, because the third order moment of ε is zero, $\frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r^{*'}\varepsilon_r$ is uncorrelated with the quadratic moment A_n. It is uncorrelated with the linear moment B_n because of the orthogonality of B_n with $\mathscr{Z}_n^{*'}\mathscr{E}_n$ in expectation. The detailed asymptotic variance Ω_β of $\hat{\beta}_n$ follows from (A.67) and the asymptotic variance of $\frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0)$ in Proposition 3. □

Proof of Proposition 5. Let

$$C_n = \begin{pmatrix} -\frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)} Z_r^{*'} Z_r^*\delta_{m0} & \frac{1}{n}\sum_{r=1}^{R} Z_r^{*'} Z_r^* \\ \Omega_\lambda^{-1} & 0. \end{pmatrix}. \tag{A.69}$$

Eqs. (A.67) and (A.54) imply that

$$\begin{pmatrix} \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2} - \beta_{20}) \end{pmatrix} = C_n^{-1}\begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{r=1}^{R} Z_r^{*'}\varepsilon_r \\ -(A_n + B_n) + o_P(1) \end{pmatrix} \xrightarrow{D} N(0, \Omega_{\lambda,\beta}), \tag{A.70}$$

where

$$\begin{aligned}\Omega_{\lambda,\beta} &= C_n^{-1}\begin{pmatrix} \frac{\sigma_0^2}{n}\sum_{r=1}^{R} Z_r^{*'} Z_r^* & 0 \\ 0 & \Omega_\lambda^{-1} \end{pmatrix} C_n^{'-1} \\ &= \frac{1}{n\sigma_0^2}\begin{pmatrix} \begin{array}{l}\sum_{r=1}^{R}\frac{m^2}{(m_r(\lambda_0))^2}(Z_r^*\delta_{m0})'(Z_r^*\delta_{m0}) \\ \quad +2(\frac{m-1}{m})\sum_{r=1}^{R}\left(\frac{m_r(0)}{R(m-1)}\right)\omega_r^2\end{array} & -\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(Z_r^*\delta_{m0})' Z_r^* \\ -\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)} Z_r^{*'}(Z_r^*\delta_{m0}) & \sum_{r=1}^{R} Z_r^{*'} Z_r^* \end{pmatrix}^{-1}\end{aligned} \tag{A.71}$$

with $\omega_r = \left(\frac{m}{m_r(\lambda_0)} - \sum_{s=1}^{R}\left(\frac{m_s(0)}{R(m-1)}\right)\frac{m}{m_s(\lambda_0)}\right)$. □

Proof of Proposition 6. 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,

$$\frac{1}{n}\Lambda_n \sum_{r=1}^{R}\left(\frac{P_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)'\left(-\frac{Y_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)\Lambda_n$$
$$= \frac{1}{n}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\left(-\frac{mY_r^*}{m_r(0)}, Z_r^*\right) \tag{A.72}$$

of which its components have a similar order of magnitude. As the reduced form equation is $Y_r^* = \left(\frac{m_r(0)}{m_r(\lambda_0)}\right)(Z_r^*\delta_{m0} + J_r\varepsilon_r)$ and, by a law of large numbers, $\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m_rP_r^*}{m_r(0)}, Z_r^*\right)'\left(\frac{m}{m_r(\lambda_0)}\right)\varepsilon_r = o_P(1)$,

$$\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}P_r^*, Z_r^*\right)'\left(-\frac{mY_r^*}{m_r(0)}, Z_r^*\right) = \frac{1}{n}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'$$
$$\times\left(-\frac{m}{m_r(\lambda_0)}Z_r^*\delta_{m0}, Z_r^*\right) + o_P(1), \tag{A.73}$$

which converges to a well-defined non-singular matrix. Similarly,

$$\frac{1}{\sqrt{n}}\Lambda_n \sum_{r=1}^{R}\left(\frac{P_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)'\varepsilon_r = \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\varepsilon_r \tag{A.74}$$

may converge to a normal distribution. It follows that

$$\sqrt{n}\Lambda_n^{-1}(\hat{\theta}_{n,\mathrm{IV}} - \theta_0)$$
$$= \left[\frac{1}{n}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\left(-\frac{mY_r^*}{m_r(0)}, Z_r^*\right)\right]^{-1}\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\varepsilon_r \xrightarrow{\mathrm{D}} \mathrm{N}(0, \Omega_{\mathrm{IV}}), \tag{A.75}$$

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

$$\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\hat{\lambda}_n)}(X_{r1}^*\hat{\beta}_{n1})'\varepsilon_r = \frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\hat{\lambda}_n)}\varepsilon_r' X_{r1}^* \cdot \sqrt{n}(\hat{\beta}_{n1} - \beta_{10}) + \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(X_{r1}^*\beta_{10})'\varepsilon_r$$
$$- \frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\hat{\lambda}_n)m_r(\lambda_0)}(X_{r1}^*\beta_{10})'\varepsilon_r \cdot \frac{\sqrt{n}}{m}(\hat{\lambda}_n - \lambda_0)$$
$$= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(X_{r1}^*\beta_{10})'\varepsilon_r + o_P(1) \tag{A.76}$$

and

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(0)m_r(\hat{\lambda}_n)}(X_{r2}^{*}\hat{\beta}_{n2})'\varepsilon_r \\
&=\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(0)m_r(\lambda_0)}(X_{r2}^{*}\beta_{20})'\varepsilon_r-\frac{1}{n}\sum_{r=1}^{R}\frac{m^2}{m_r(0)m_r(\hat{\lambda}_n)m_r(\lambda_0)}(X_{r2}^{*}\beta_{20})'\varepsilon_r\cdot\frac{\sqrt{n}}{m}(\hat{\lambda}_n-\lambda_0) \\
&\quad+\frac{1}{n}\sum_{r=1}^{R}\frac{m^2}{m_r(0)m_r(\hat{\lambda}_n)}\varepsilon_r'X_{r2}^{*}\cdot\frac{\sqrt{n}}{m}(\hat{\beta}_{n2}-\beta_{20}) \\
&=\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(0)m_r(\lambda_0)}(X_{r2}^{*}\beta_{20})'\varepsilon_r+o_P(1), \qquad (A.77)
\end{aligned}$$

hence,

$$\begin{aligned}
&\sqrt{n}\Lambda_n^{-1}(\hat{\theta}_{n,\mathrm{BIV}}-\theta_0) \\
&=\left[\frac{1}{n}\sum_{r=1}^{R}\left(-\frac{m}{m_r(\lambda_0)}Z_r^{*}\delta_{m0},Z_r^{*}\right)'\left(-\frac{m}{m_r(0)}Y_r^{*},Z_r^{*}\right)\right]^{-1} \\
&\quad\times\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\left(-\frac{m}{m_r(\lambda_0)}Z_r^{*}\delta_{m0},Z_r^{*}\right)'\varepsilon_r+o_P(1) \\
&\xrightarrow{D} N(0,\Omega_{\mathrm{BIV}}), \qquad (A.78)
\end{aligned}$$

as in (4.5).

The limiting variances Ω_{BIV} and Ω_{IV} can be compared. Denote $\mathscr{P}_n^{*}=\left(\frac{m}{m_1(0)}P_1^{*'},\ldots,\frac{m}{m_R(0)}P_R^{*'}\right)'$ and $\mathscr{S}_n^{*}=\left(\left(-\frac{m}{m_1(\lambda_0)}Z_1^{*}\delta_{m0}\right)',\ldots,\left(-\frac{m}{m_R(\lambda_0)}Z_R^{*}\delta_{m0}\right)'\right)'$. It follows that their limiting precision matrices are

$$\Omega_{\mathrm{IV}}^{-1}=\frac{1}{\sigma_0^2}\lim_{n\to\infty}\frac{1}{n}(\mathscr{S}_n^{*},\mathscr{Z}_n^{*})'(\mathscr{P}_n^{*},\mathscr{Z}_n^{*})[(\mathscr{P}_n^{*},\mathscr{Z}_n^{*})'(\mathscr{P}_n^{*},\mathscr{Z}_n^{*})]^{-1}(\mathscr{P}_n^{*},\mathscr{Z}_n^{*})'(\mathscr{S}_n^{*},\mathscr{Z}_n^{*}),$$

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

Proof of Proposition 7. When $\beta_{10}=0$, the within equation is $Y_r^{*}=\left(-\frac{\lambda_0}{m_r(0)}Y_r^{*}-\frac{1}{m_r(0)}X_{r2}^{*}\beta_{20}+J_r\varepsilon_r\right)$ and its reduced form equation is $Y_r^{*}=\left(\frac{m_r(0)}{m_r(\lambda_0)}\right)\left(-\frac{X_{r2}^{*}\beta_{20}}{m_r(0)}+J_r\varepsilon_r\right)$. Therefore,

$$-\frac{m^2}{m_r(0)}Y_r^{*}=\frac{m^2}{m_r(\lambda_0)m_r(0)}X_{r2}^{*}\beta_{20}-\frac{m^2}{m_r(\lambda_0)}J_r\varepsilon_r. \qquad (A.79)$$

Denote

$$\Lambda_n=\begin{pmatrix} m^2 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & m\end{pmatrix} \quad \text{and} \quad \Gamma_n=\begin{pmatrix} m & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & m\end{pmatrix}.$$

It follows that

$$\begin{aligned}
&\frac{1}{n}\Gamma_n\sum_{r=1}^{R}\left(\frac{P_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)'\left(-\frac{Y_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)\Lambda_n\\
&\quad=\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}P_r^*, Z_r^*\right)'\left(\frac{m^2}{m_r(\lambda_0)m_r(0)}X_{r2}^*\beta_{20}, Z_r^*\right)-\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}P_r^*, Z_r^*\right)'\\
&\quad\times\left(\frac{m^2}{m_r(\lambda_0)}\varepsilon_r, 0, 0\right)\\
&\quad=\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}P_r^*, Z_r^*\right)'\left(\frac{m^2}{m_r(\lambda_0)m_r(0)}X_{r2}^*\beta_{20}, Z_r^*\right)+o_P(1),
\end{aligned}\tag{A.80}$$

because $\frac{1}{n}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}P_r^*, Z_r^*\right)'\frac{m^2}{m_r(\lambda_0)}\varepsilon_r = O_P\left(\frac{m}{\sqrt{n}}\right) = o_P(1)$ under Assumption 5. Therefore,

$$\begin{aligned}
&\sqrt{n}\Lambda_n^{-1}(\hat{\theta}_{n,\mathrm{IV}}-\theta_0)\\
&\quad=\left[\frac{1}{n}\Gamma_n\sum_{r=1}^{R}\left(\frac{P_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)'\left(-\frac{Y_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)\Lambda_n\right]^{-1}\\
&\qquad\times\frac{\Gamma_n}{\sqrt{n}}\sum_{r=1}^{R}\left(\frac{P_r^*}{m_r(0)}, X_{r1}^*, -\frac{X_{r2}^*}{m_r(0)}\right)'\varepsilon_r\\
&\quad=\left[\frac{1}{n}\sum_{r=1}^{R}\left(\frac{mP_r^*}{m_r(0)}, Z_r^*\right)'\left(\frac{m^2}{m_r(\lambda_0)m_r(0)}X_{r2}^*\beta_{20}, Z_r^*\right)+o_P(1)\right]^{-1}\\
&\quad\times\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\left(\frac{m}{m_r(0)}P_r^*, Z_r^*\right)'\varepsilon_r\\
&\quad\xrightarrow{D} N(0,\Omega_{\mathrm{IV}}),
\end{aligned}\tag{A.81}$$

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

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\hat{\lambda}_n)}(X_{r1}^*\hat{\beta}_{n1})'\varepsilon_r &= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(X_{r1}^*\beta_{10})'\varepsilon_r+\frac{1}{n}\sum_{r=1}^{R}\frac{m}{m_r(\hat{\lambda}_n)}\varepsilon_r'X_{r1}^*\cdot\sqrt{n}(\hat{\beta}_{n1}-\beta_{10})\\
&\quad-\frac{1}{n}\sum_{r=1}^{R}\frac{m^2}{m_r(\hat{\lambda}_n)m_r(\lambda_0)}(X_{r1}^*\beta_{10})'\varepsilon_r\cdot\frac{\sqrt{n}}{m^2}(\hat{\lambda}_n-\lambda_0)\\
&= \frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(\lambda_0)}(X_{r1}^*\beta_{10})'\varepsilon_r+o_P(1)
\end{aligned}\tag{A.82}$$

and

$$\frac{1}{\sqrt{n}}\sum_{r=1}^{R}\frac{m}{m_r(0)m_r(\hat{\lambda}_n)}(X_{r2}^*\hat{\beta}_{n2})'\varepsilon_r$$

$$= \frac{1}{\sqrt{n}} \sum_{r=1}^{R} \frac{m}{m_r(0)m_r(\lambda_0)} (X_{r2}^{*}\beta_{20})'\varepsilon_r + \frac{1}{n} \sum_{r=1}^{R} \frac{m^2}{m_r(0)m_r(\hat{\lambda}_n)} \varepsilon_r' X_{r2}^{*} \cdot \frac{\sqrt{n}}{m} (\hat{\beta}_{n2} - \beta_{20})$$
$$- \frac{1}{n} \sum_{r=1}^{R} \frac{m^3}{m_r(0)m_r(\hat{\lambda}_n)m_r(\lambda_0)} (X_{r2}^{*}\beta_{20})'\varepsilon_r \cdot \frac{\sqrt{n}}{m^2} (\hat{\lambda}_n - \lambda_0)$$
$$= \frac{1}{\sqrt{n}} \sum_{r=1}^{R} \frac{m}{m_r(0)m_r(\lambda_0)} (X_{r2}^{*}\beta_{20})'\varepsilon_r + o_P(1). \tag{A.83}$$

Thus,

$$\begin{pmatrix} \frac{\sqrt{n}}{m^2}(\hat{\lambda}_{n,\mathrm{BIV}} - \lambda_0) \\ \sqrt{n}(\hat{\beta}_{n1,\mathrm{BIV}} - \beta_{10}) \\ \frac{\sqrt{n}}{m}(\hat{\beta}_{n2,\mathrm{BIV}} - \beta_{20}) \end{pmatrix}$$
$$= \left[\frac{1}{n} \sum_{r=1}^{R} \left(\frac{m^2}{m_r(\lambda_0)m_r(0)} X_{r2}^{*}\beta_{20}, Z_r^{*} \right)' \left(-\frac{m^2}{m_r(0)} Y_r^{*}, Z_r^{*} \right) \right]^{-1}$$
$$\times \frac{1}{\sqrt{n}} \sum_{r=1}^{R} \left(\frac{m^2}{m_r(\lambda_0)m_r(0)} X_{r2}^{*}\beta_{20}, Z_r^{*} \right)' \varepsilon_r + o_P(1) \xrightarrow{D} N(0, \Omega_{\mathrm{BIV}}), \tag{A.84}$$

for the best IV estimator. $\square$

A.6. Asymptotic bias of the OLS estimator in (5.1)

From (2.3), because $(1 + \frac{\lambda_0}{m_r(0)})Y_r^{*} = X_{r1}^{*}\beta_{10} - \frac{1}{m_r(0)} X_{r2}^{*}\beta_{20} + \varepsilon_r^{*}$, it follows that

$$\hat{\beta}_{n1,L} - \beta_{10} = b_n + \left[\sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*} \right]^{-1} \sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda_0)} X_{r1}^{*'} \varepsilon_r,$$

where $b_n = -\frac{1}{m}[\sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*}]^{-1} \sum_{r=1}^{R} \frac{m}{m_r(\lambda_0)} X_{r1}^{*'}(\lambda_0 X_{r1}^{*}\beta_{10} + X_{r2}^{*}\beta_{20})$. The b_n is the bias of $\hat{\beta}_{n1,\mathrm{L}}$ 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 $\hat{\beta}_{n1,\mathrm{L}}$ is a consistent estimate of β_{10}, but

$$m(\hat{\beta}_{n1,\mathrm{L}} - \beta_{10}) = mb_n + \sqrt{\frac{m}{R}} \left(\frac{1}{n} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda_0)} X_{r1}^{*'} \varepsilon_r, \tag{A.85}$$

where the limit of mb_n can be finite as

$$\lim_{n\to\infty} mb_n = -\lim_{n\to\infty} \left[\frac{1}{n} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*} \right]^{-1} \frac{1}{n} \sum_{r=1}^{R} \frac{1}{a_r} X_{r1}^{*'}(\lambda_0 X_{r1}^{*}\beta_{10} + X_{r2}^{*}\beta_{20}),$$

and $\left(\frac{1}{n} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*} \right)^{-1} \frac{1}{\sqrt{n}} \sum_{r=1}^{R} \frac{m_r(0)}{m_r(\lambda_0)} X_{r1}^{*'} \varepsilon_r \xrightarrow{D} N\left(0, \sigma_0^2 \lim_{n\to\infty} \left[\frac{1}{n} \sum_{r=1}^{R} X_{r1}^{*'} X_{r1}^{*} \right]^{-1} \right)$. Thus, $m(\hat{\beta}_{n1,\mathrm{L}} - \beta_{10}) = mb_n + O_P\left(\sqrt{\frac{m}{R}}\right)$.

References

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge, MA.

Bearman, P.S., Jones, J., Udry, J.R., 1997. The national longitudinal study of adolescent health: research design, ⟨http://www.cpc.unc.edu/projects/addhealth/design⟩.

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

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

Cameron, A.C., Trivedi, P.K., 2003. Microeconometrics: methods and applications. Monograph, Department of Economics, Indiana University.

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., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.

Durlauf, S., Young, H., 2001. Social Dynamics. MIT Press, Cambridge, MA.

Glaeser, E.L., Sacerdote, B., Scheinkman, J.A., 1996. Crime and social interactions. Quarterly Journal of Economics 111, 507–548.

Hanushek, E.A., Kain, J.F., Markman, J.M., Rivkin, S.G., 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, MA.

Ioannides, Y., Zabel, J.E., 2003. Neighborhood effects and housing demand. Journal of Applied Econometrics 18, 563–584.

Kelejian, H.H., Prucha, I.R., 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., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219–257.

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

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

Lin, X., 2004. Peer effects and student academic achievement: an application of spatial autoregressive model with group unobservables. Manuscript, Department of Economics, The Ohio State University.

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: Durlauf, S., Young, H.P. (Eds.), Social Dynamics. MIT Press, Cambridge, MA, pp. 45–82.

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

Rees, D.I., Zax, J.S., Herries, J., 2003. Interdependence in worker productivity. Journal of Applied Econometrics 18, 585–604.

Sacrerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. Quarterly Journal of Economics 116, 681–704.

Theil, H., 1971. Principles of Econometrics. Wiley, New York, NY.

Weinberg, B., Reagan, P.B., Yankow, J.J., 2004. Do neighborhoods affect work behavior? Evidence from the NLSY 79. Journal of Labor Economics 24(4), 891–924.

White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press, New York, NY.