Social Networks with Heterogeneity and Group Fixed Effects: A Likelihood Approach*

Yanli Lin[†], Yang Yang[‡]

December 6, 2021

Abstract

This paper considers social interaction models with group fixed effects and observed heterogeneity among agents. By likelihood approach, with the control of group-level confounding effects of the common variables, both heterogeneous endogenous peer effects and exogenous contextual effects can be identified and estimated consistently. Under some regularity assumptions, we prove the consistency and asymptotic normality of the QMLE. Monte Carlo simulation results show that our QMLE has good finite sample performance. For an application, we investigate the China Education Panel Survey (CEPS) and focus on gender heterogeneity on academic achievement of Grade 8 students in junior high school. We capture significant gender disparities in peer effects from gender subgroups in a classroom. Besides, female students' test scores are more subject to both female and male peers' average achievement.

1 Introduction

Social interaction effects have received substantial attention since Coleman et al. (1966). Peer effects, as a typical example, have inherent externality (Hoxby, 2000), which provides justifications for policy intervention targeted at enhancing social welfare. However, the identification and estimation of social interaction effects is hard. Linear-in-means models might suffer from "reflection problem" described by Manski (1993)[1], omitted variable bias problem (or correlated effects in Manski (1993)) [2], and data limitation about an individual's reference group.

The spatial autoregressive (SAR) model with both endogenous peer effects and exogenous contextual effects combined with group fixed effects can confront the difficulties mentioned above, since

*We would like to thank Hanbat Jeong, Abolfazl Setayesh, and Professor Bruce Weinberg for their valuable comments.

†Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. Email: lin.3019@osu.edu.

‡Department of Economics, The Ohio State University, 1945 N. High Street, Columbus, Ohio. Email: yang.3524@osu.edu.

[1]It refers to the impossibility of separately identifying two distinct types of social effects - endogenous or behavioral effects and contextual or exogenous effects. The former one can generate a social multiplier while the latter one do not.

[2]It's difficult to disentangle social effects from other confounding effects. Peer group formation might not be random. The validity of strategies such as instrumental variable (IV) method, family fixed effect and experiment type strategy is open to question (Lin, 2010).

the nonlinearity introduced by variations in the measurements of the peer variables provides helpful information for identification. For instance, Lee (2007) incorporate endogenous and contextual effects and group fixed effects into a SAR model and consider a group interaction setting, which assumes that an individual is equally affected by all the other members in the same group. The proposed conditional maximum likelihood (CML) and IV methods rely on sufficient variations in group sizes for identification. Lin (2010) instead specify the spatial weight matrix based on the actual friendship network within each group and employ the "de-group-mean" approach for estimation. These two papers provide valuable framework for social interaction models, but they only consider homogeneous agents. Besides, one limitation of their studies is that they can only deal with interaction structure within a group, but not the case when an individual also interact with other individuals outside the group.

A more realistic social interaction scenario is that individuals' with disparate types, such as gender, races and education background, have different interaction patterns with other members, either from the same group or from different groups. Heterogeneous social interaction effects are investigated in some previous works (e.g., Yakusheva et al. (2014), Lu and Anderson (2015), Tincani (2018)), however, most of them consider experiment type or quasi - experimental strategy, as pointed out by Lin (2010), the validity of which depends heavily on the design and implementation of the experiment since it's possible that what's supposed to be random is actually a result of self-selection. The higher-order SAR (HSAR) models might provide a solution. Hsieh and Lin (2017) apply a HSAR model to study heterogeneous peer effects with endogenous network formation by using Bayesian approach for estimation. Gupta and Robinson (2018) develops pseudo maximum likelihood estimates for HSAR models with increasingly many parameters without considering fixed effects. Aquaro et al. (2020) considers estimation and inference of a spatiotemporal model with spatial lag coefficients being differed over the cross-section units.

This study further extend the HSAR model to incorporate both heterogeneous endogenous peer effects and exogenous contextual effects as well as group fixed effects in a social interaction setting. First, unlike the Bayesian approach in Hsieh and Lin (2017), we propose a quasi-maximum likelihood (QML) approach for identification and estimation. As will be discussed in Section 2.3, the approaches in Lee (2007) and Lin (2010) can not be applied to handle the group fixed effects in the presence of heterogeneity among agents. Instead, similar to Yu et al. (2008)[3], we consider a direct estimation approach by the joint estimation of the model parameters and the group-fixed effects. Since there exists asymptotic bias as the expectation of the first order derivatives evaluated at the true parameter values are not centered at zero, a bias correction procedure can be employed subsequently to greatly reduce the bias, which can be verified by the Monte Carlo simulation results. Second, the QML estimation can accommodate two commonly seen sampling methods in empirical applications: (i) more group members are included in the sample as the sample size increases, keeping the group number fixed; (ii) more groups enter the sample as the sample size increases, while keeping group members unaltered. We can obtain consistency and asymptotic normality for the estimates of the parameters of interest in both cases. For the first case, the limiting distribution

[3] They propose estimation methods for spatial dynamic panel data (SDPD) models with both time and individual fixed effects.

of each group fixed effect can be derived, while the limiting distribution doesn't exist for the second case. Third, our model also allow a more general scenario that an individual might interact with other individuals outside her/his group, which can not be handled by previous approaches.

We then apply our model to investigate China Education Panel Survey (CEPS) and focus on gender heterogeneity on academic achievement of Grade 8 students in junior high school. First, significant gender disparities in peer effects from gender subgroups in a classroom are captured. In general, a female student's test score is more subject to both her female and male peers' average achievement than a male student. Besides, female classmates' average education outcome has higher impact on a student's, either female or male, Chinese and total test scores, while male classmates contribute more to a student's Mathematics academic achievement. Second, we find evidence that the contextual effect of a student's relative age (measured by $\pm$ month(s) compared with sample median date of birth) has both competitive effects and complementary effects for a female student. To be more specific, being in a class with older male classmates would slightly reduce a female student's Chinese and Mathematics test scores, while staying with older female classmates can be beneficiary to a female student's English and total test scores. Third, by the contextual effect of mother's education, we detect the specific channel about how higher classmates' maternal education raises a students' test score (Chung and Zou, 2020), i.e., a female student's Mathematics and total test scores are positively affected when her male classmates have higher educated mothers. Those findings might provide evidence to support some low-cost ways to potentially improve students' academic performance within the world's largest school system. Last, the impact of a head teacher might be entangled with interaction patterns of within and across gender subgroups in the same class, but under current setting, we are unable to detect the channels, which might be an interesting topic for future studies.

In the following parts of this paper, Section 2 introduces the model setting, economic foundation and the quasi-maximum likelihood estimation method. Section 3 discusses the asymptotic properties of the QMLE, including conditions for identification, proofs for consistency and asymptotic distribution of the QMLE. Section 4 shows the finite sample performance of our QMLE by Monte Carlo simulations. Section 5 applies the model to data sets from CEPS and examines heterogeneous peer and contextual effect in Chinese student academic achievement.

2 Model Formation and the QMLE

2.1 Model Setting

Suppose there are n individual units in an economy. For regularity, we need the following assumptions:

Assumption 1 (Heterogeneity Source):

All individuals can be classified into K types $\left\{\mathcal{K}_n^k\right\}_{k=1}^K$, which satisfy $\cup_{k=1}^K \mathcal{K}_n^k = n$ and $\mathcal{K}_n^{k_1} \cap \mathcal{K}_n^{k_2} = 0$ for $k_1 \neq k_2$. K is a constant which does not depend on n.

Assumption 1 indicates that the heterogeneity is observed and generated by categorical characteristics of individuals. First, $\left\{\mathcal{K}_n^k\right\}_{k=1}^K$ may not be consecutive, which means that the division of types can be irrelevant to neighborhood relationship among individuals. For example, students coming from the same school/classroom can be divided into subgroups based on their characteristics. Second, we assume K is a constant, which is consistent with most empirical scenarios since a representative sample is needed regardless of the sample size and most of the categorical heterogeneities, such as gender, race and educational background, only have finitely many types.

Assumption 2 (Group Assignment):

The n individuals belong to G groups $\{\mathcal{G}_n^g\}_{g=1}^G$ which satisfy $\cup_{g=1}^G \mathcal{G}_n^g = n$ and $\mathcal{G}_n^{g_1} \cap \mathcal{G}_n^{g_2} = 0$ for $g_1 \neq g_2$. G can be a constant or growing with n.

Assumption 2 assumes the individuals are assigned to different groups and the group assignment may not be random. First, the assumption about the number of groups G can accommodate two sampling methods commonly seen in empirical studies: (i) G is fixed, while the group members grows with sample size n; (ii) group members are fixed, while G grows with n. Second, the groups in Assumption 2 and categories in Assumption 1 can be different. We treat the heterogeneity and group assignment as two abstract ways of segmentations of all the individuals n.

Let y_i be individual i's outcome, $x_i = (x_{i,1}, ..., x_{i,L})'$ be L univariate exogenous variables for i, and $w_{ij,n}$ be the link captures the impact from individual j to individual i ($w_{ii,n} = 0$). For heterogeneity, define $h_{i,k} = \begin{cases} 1 & i \in \mathcal{K}_n^k \\ 0 & i \notin \mathcal{K}_n^k \end{cases}$ as the indicator of type identity for individual i and category k. Similarly, for group assignment, define $\widetilde{h}_{i,g} = \begin{cases} 1 & i \in \mathcal{G}_n^g \\ 0 & i \notin \mathcal{G}_n^g \end{cases}$ as the indicator for whether individual i belongs to group $\mathcal{G}_n^g$ or not. For each individual i, we consider the following model:

$$
\begin{aligned}
y_i = & \sum_{k=1}^{K} \lambda_k h_{i,k} \left(\sum_{j=1}^{n} w_{ij,n} y_j \right) + \sum_{k=1}^{K} h_{i,k} \left(\sum_{j=1}^{n} w_{ij,n} x_i^{'} \right) \gamma_k \\
& + x_i^{'} \beta + \sum_{g=1}^{G} \widetilde{h}_{i,g} \alpha_g + u_i \qquad (1)
\end{aligned}
$$

where $u_i \overset{i.i.d}{\sim} \left(0, \sigma^2\right)$. In this model, λ_k and γ_k capture the spillover effects from neighbors received by an individual of category $\mathcal{K}_n^k$ directly or indirectly, i.e., peer and contextual effects respectively. β measure the effects from external regressors which are only associated with the individual himself/herself. α_g is a group-level fixed effect of group $\mathcal{G}_n^g$, which captures the common factors, either observed or unobserved, faced by all group members.

Let $Y_n = (y_1, \cdots, y_n)^{'}$, $X_n = (x_1, \cdots, x_n)^{'}$, $W_n = (w_{ij,n})_{n \times n}$, $H_{k,n} = diag\left(h_{1,k}, \cdots, h_{n,k}\right)$, $\alpha = (\alpha_1, \cdots, \alpha_G)^{'}$, $H_G = \left(\widetilde{h}_1, \cdots, \widetilde{h}_G\right)^{'}$ where $\widetilde{h}_g = \left(\widetilde{h}_{1,g}, \cdots, \widetilde{h}_{n,g}\right)^{'}$, and $u_n = (u_1, \cdots, u_n)^{'}$. We can rewrite the model into the following matrix form:

$$Y_n = \sum_{k=1}^{K} \lambda_k H_{k,n} W_n Y_n + \sum_{k=1}^{K} H_{k,n} W_n X_n \gamma_k + X_n \beta + H_G \alpha + u_n \tag{2}$$

Equation (1)(or (2)) is the single network specification, in Section 2.4, we consider a simple extension to the multiple networks specification.

2.2 Economic Foundation

The above model setting captures externalities from social interactions through a network, and allows the levels of externalities to be associated with individuals' types. When individual i's outcome y_i reflects his/her action, it can be regarded as a model of the Nash equilibrium of a static complete information game with different types of players processing the following linear-quadratic utility function:

$$u_i\left(y_i\right) = y_i \left(x_i^{'} \beta + \sum_{j=1}^{n} w_{ij,n} x_i^{'} \gamma_k + \alpha_g + v_i \right) - \frac{1}{2} \left(y_i - \lambda_k \sum_{j=1}^{n} w_{ij,n} y_j \right)^2 \tag{3}$$

for individual i who has type k and belongs to group $\mathcal{G}_n^g$. The first component represents private utility associated with his/her own action y_i which includes contextual effects from other individuals' characteristics, group fixed effects and random effects (notation changed from u_i to v_i to avoid ambiguity). The second component captures the interaction with neighbors, which is a conformity effect directly associated with neighbors' actions. The unique equilibrium outcome of individual i can be obtained by maximizing $u_i\left(y_i\right)$ with respect to y_i, and is exactly described by equation (1)[4]. Unlike the SAR situation in Brock and Durlauf (2001), the conformity effect here does depend on the type of i himself/herself. For example, female and male students will receive different peer effects from their classmates, and also students of different gender have distinct sensitive levels to average peer outcome.

2.3 QML Approach and Bias Correction

In existing literature, Lee (2007) demonstrates that in his model both endogenous and contextual effects are identifiable if group sizes are not constant, and that weak identification can occur in the case of large group sizes. Lin (2010)'s "de-group-mean" approach generalizes Lee (2007) by allowing the case when all the group sizes are identical. However, the "de-group-mean" approach has two limitations. First, it is only suitable to the scenario that individuals only interact with other individuals inside the same group. But it doesn't make sense in some empirical applications. For example, when studying the impact of peer effects on education outcomes, not only can students interact with their classmates, but they might also make friends with students in other classes by out-of-school activities via some sports and arts associations. Without considering the impact from friends outside the group, the peer and contextual effect might be misspecified. Second, even if a

[4]Note that $h_{i,k}$ and $\widetilde{h}_{i,g}$ are indicator variables.

group member only interact with other group members, when heterogeneity is introduced into the social interaction models, the "de-group-mean" transformation of data before forming the likelihood function in Lin (2010) is still not a good way to deal with group fixed effects, below is an illustration. Assume for each group $\mathcal{G}_n^g$, we have

$$Y_g = \sum_{k=1}^{K} \lambda_k H_{k,g} W_g Y_g + \sum_{k=1}^{K} H_{k,g} W_g X_g \gamma_k + \alpha_g + u_g \tag{4}$$

where Y_g and X_g are the vector and matrix of outcomes and characteristics of the members in group $\mathcal{G}_n^g$. W_g and $H_{k,g}$ are defined similar as W_n and $H_{k,n}$ for the adjacent matrix and heterogeneity source categories for group $\mathcal{G}_n^g$. Let m_g be the number of individuals in group $\mathcal{G}_n^g$, to eliminate α_g, the "de-group-mean" approach suggests that we can multiply $J_g = I_{m_g} - \frac{1}{m_g} l_g l_g^{'}$, where l_g is the m_g-dimensional vector of ones, on both RHS and LHS of (4). Then,

$$J_g Y_g = \sum_{k=1}^{K} \lambda_k J_g H_{k,g} W_g Y_g + \sum_{k=1}^{K} J_g H_{k,g} W_g X_g \gamma_k + J_g u_g$$

At first glance, it is similar to equation (6) in Lin (2010). However, due to the presence of the heterogeneity matrices $H_{k,g}$, $J_g H_{k,g} W_g Y_g \neq J_g H_{k,g} W_g J_g Y_g$ and $J_g H_{k,g} W_g X_g \neq J_g H_{k,g} W_g J_g X_g$, even when W_g is row-normalized because $H_{k,g} W_g$ will always have some rows with all zero elements. Thus, Lee (2007) and Lin (2010)'s approaches can not be applied when heterogeneity is introduced to the model. Due to both empirical and theoretical limitations discussed, in this paper, we consider another approach: direct estimation approach by jointly estimating the parameters and group fixed effect.

For notation simplicity, let $\Lambda = (\lambda_1, \cdots, \lambda_K)^{'}$, $\varphi = \left(\gamma_1^{'}, \cdots, \gamma_K^{'}, \beta^{'}\right)^{'}$ where $\gamma_k = (\gamma_{k,1}, ...\gamma_{k,L})'$ for $k = 1, ..., K$ and $\beta = (\beta_1, .., \beta_L)'$, $Z_n = (H_{1,n} W_n X_n, \cdots, H_{K,n} W_n X_n, X_n)^{'}$, and $S_n(\Lambda) = I_n - \sum_{k=1}^{K} \lambda_k H_{k,n} W_n$, the log-likelihood function of the model in Section 2.1 under $u_n \sim N\left(0, \sigma^2 I_n\right)$ is

$$\ln L_n(\theta, \alpha) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |S_n(\Lambda)| - \frac{1}{2\sigma^2} u_{\theta,\alpha}^{'} u_{\theta,\alpha} \tag{5}$$

where $u_{\theta,\alpha} = S_n(\Lambda) Y_n - Z_n \varphi - H_G \alpha$ and $\theta = \left(\Lambda^{'}, \varphi, \sigma^2\right)^{'}$. To make sure the likelihood function exist and the model can be identified, we need the following assumption:

Assumption 3:
$S_n(\Lambda)$ is invertible and Z_n is full rank.

Following existing literature (e.g., Yu et al. (2008)), when fixed effects appear in the likelihood function, by first order condition of α, i.e. $\frac{\partial \ln L_n(\theta,\alpha)}{\partial \alpha} = \frac{1}{\sigma^2} H_G^{'} u_{\theta,\alpha} = 0$, we can concentrate out group fixed effects and get the following concentrated log-likelihood function:

$$\ln L_n(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 + \ln |S_n(\Lambda)| - \frac{1}{2\sigma^2} \widetilde{u}_\theta^{'} \widetilde{H}_G^{'} \widetilde{H}_G \widetilde{u}_\theta \tag{6}$$

with $\widetilde{H}_G = I_n - H_G M_G^{-1} H_G^{'}$ and $\widetilde{u}_\theta = S_n(\Lambda) Y_n - Z_n \varphi$, where $M_G = H_G^{'} H_G = diag(m_1, \cdots, m_G)$, and m_g is the number of individuals in group $\mathcal{G}_n^g$. Note M_G is always invertible. In Section 3, we will discuss the identification and asymptotic properties of QMLE based on the concentrated log-likelihood function (6), since it is easier for us to discuss the asymptotic distributions of θ and α separately.

However, due to the existence of heterogeneity, there are still too many parameters in (6) which might be demanding to maximize it directly through numerical searches. To simplify the maximization problem, concentrating out as many parameters as possible further is the best way to numerically compute the QMLE. From (5), we can also get the following first order derivatives:

$$\frac{\partial \ln L_n(\theta, \alpha)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} u_{\theta,\alpha}^{'} u_{\theta,\alpha} \tag{7}$$

$$\frac{\partial \ln L_n(\theta, \alpha)}{\partial \varphi} = \frac{1}{\sigma^2} Z_n^{'} u_{\theta,\alpha} \tag{8}$$

When the log-likelihood function is maximized, (7) and (8) are all equal to zero and we can solve α, σ^2 and φ by representing them as functions of Λ:

$$\alpha = M_G^{-1} H_G^{'} (I_n - Z_n A_n) S_n(\Lambda) Y_n \tag{9}$$

$$\sigma^2 = \frac{1}{n} Y_n^{'} S_n^{'}(\Lambda) B_n^{'} B_n S_n(\Lambda) Y_n \tag{10}$$

$$\varphi = A_n S_n(\Lambda) Y_n \tag{11}$$

where $A_n = \left(Z_n^{'} \widetilde{H}_G Z_n\right)^{-1} Z_n^{'} \widetilde{H}_G$ and $B_n = \widetilde{H}_G (I_n - Z_n A_n)$. With Assumption 3, the invertibility of matrices showed in (9), (10) and (11) can be guaranteed. Then backing out the results to (6) yields the concentrated log-likelihood function which only contains parameter Λ:

$$Q_n(\Lambda) = -\frac{n}{2}[1 + \ln(2\pi)] - \frac{n}{2} \ln\left[\frac{1}{n} Y_n^{'} C_n(\Lambda) Y_n\right] + \ln|S_n(\Lambda)| \tag{12}$$

where $C_n(\Lambda) = S_n^{'}(\Lambda) B_n^{'} B_n S_n(\Lambda)$. As the number of parameters reduced from $(K+1)(L+1)$ [5] to K by further concentration, the computational time can be greatly reduced.

However, concentrated QMLE by maximizing (12) would have asymptotic bias. To see this, at true parameter (θ_0, α_0), we have

[5] $(K+1)(L+1) + G$ if group fixed effects are considered

$$\begin{aligned}
\frac{\partial Q_n\left(\Lambda_0\right)}{\partial \lambda_k} &= \frac{nY_n^{'}\left(H_{k,n}W_n\right)^{'}B_n^{'}B_nS_n\left(\Lambda_0\right)Y_n}{Y_n^{'}C_n\left(\Lambda_0\right)Y_n} - tr\left(H_{k,n}W_nS_n^{-1}\right) \\
&= \frac{\left[B_n\widetilde{W}_{n,k}S_n^{-1}\left(Z_n\varphi_0 + H_G\alpha_0 + u_n\right)\right]^{'}u_n}{\sigma_0^2} - tr\left(\widetilde{W}_{n,k}S_n^{-1}\right) \\
&= -tr\left(G_{n,k}\right) + \frac{1}{\sigma_0^2}\left[B_n\widetilde{W}_{n,k}S_n^{-1}\left(Z_n\varphi_0 + H_G\alpha_0\right)\right]^{'}u_n \\
&+ \frac{1}{\sigma^2}u_n^{'}\left(B_nG_{n,k}\right)^{\prime}u_n \qquad (13)
\end{aligned}$$

where $\widetilde{W}_{n,k} = H_{k,n}W_n$, $S_n = S_n\left(\Lambda_0\right) = I_n - \sum_{k=1}^{K}\lambda_{k,0}\widetilde{W}_{n,k}$ and $G_{n,k} = \widetilde{W}_{n,k}S_n^{-1}$ for $k = 1, \cdots, K$.

In (13), the expectation of linear term $\frac{1}{\sigma_0^2}\left[B_n\widetilde{W}_{n,k}S_n^{-1}\left(Z_n\varphi_0 + H_G\alpha_0\right)\right]^{'}u_n$ is zero, but for the remaining part, we have

$$\begin{aligned}
&E\left[\frac{1}{\sigma^2}u_n^{'}\left(B_nG_{n,k}\right)^{\prime}u_n - tr\left(G_{n,k}\right)\right] \\
=&tr\left(B_nG_{n,k}\right) - tr\left(G_{n,k}\right) \neq 0
\end{aligned}$$

Thus, let $\Delta_{k,n}\left(\hat{\Lambda}_n\right) = tr\left[\left(I_n - B_n\right)\widetilde{W}_{n,k}S_n^{-1}\left(\hat{\Lambda}_n\right)\right]$ and $\Sigma_{\hat{\Lambda}_n,n} = \frac{\partial^2 Q\left(\hat{\Lambda}_n\right)}{\partial\Lambda\partial\Lambda^{\prime}}$, by Taylor expansion $\frac{\partial Q}{\partial\Lambda}\left(\hat{\Lambda}\right) - \frac{\partial Q}{\partial\Lambda}\left(\Lambda_0\right) = \Sigma_{\hat{\Lambda}_n,n}^{-1}\left(\hat{\Lambda} - \Lambda_0\right) + o_p\left(\hat{\Lambda} - \Lambda_0\right)$, we have the following bias corrected estimator for Λ,

$$\hat{\Lambda}_{bc,n} = \hat{\Lambda}_n + \Sigma_{\hat{\Lambda}_n,n}^{-1}\Delta_n\left(\hat{\Lambda}_n\right) \qquad (14)$$

where $\Delta_n\left(\hat{\Lambda}_n\right) = \left(\Delta_{1,n}\left(\hat{\Lambda}_n\right), \cdots, \Delta_{K,n}\left(\hat{\Lambda}_n\right)\right)^{'}$. In Section 4, we will compare the performance of the concentrated QMLE with and without bias correction by Monte Carlo simulations.

2.4 A Simple Extension: Multiple Networks Situation

In many empirical applications, it's more natural to consider a multiple networks setting instead of a single network specification. For example, high school students may have multiple social links with different groups of people, like classmates, friends in student associations or sports clubs. So we may be interested in estimating heterogeneous peer effects and contextual effects in student academic achievement under multiple networks setting. In this case, the model described in (1) can be easily extended to accommodate this scenario:

$$y_i = \sum_{r=1}^{R}\sum_{k=1}^{K}\lambda_{r,k}h_{i,k}\left(\sum_{j=1}^{n}w_{ij,r,n}y_j\right) + \sum_{r=1}^{R}\sum_{k=1}^{K}h_{i,k}\left(\sum_{j=1}^{n}w_{ij,r,n}x_i^{\prime}\right)\gamma_{r,k} + x_i^{\prime}\beta + \sum_{g=1}^{G}\widetilde{h}_{i,g}\alpha_g + u_i \qquad (15)$$

where $w_{ij,r,n}$ is the impact from j to i in the network r ($w_{ii,r,n} = 0$). $\lambda_{r,k}$ and $\gamma_{r,k}$ capture the heterogeneous peer and contextual effects from network r for type k individuals. Denote $W_{r,n} = (w_{ij,r,n})$, where $W_{r,n}, \cdots, W_{R,n}$ are R different networks among the individuals, we can rewrite equation (15) in matrix form:

$$Y_n = \sum_{r=1}^{R}\sum_{k=1}^{K} \lambda_{r,k} H_{k,n} W_{r,n} Y_n + \sum_{r=1}^{R}\sum_{k=1}^{K} H_{k,n} W_{r,n} X_n \gamma_{r,k} + X_n \beta + H_G \alpha + u_n \tag{16}$$

Let $\Lambda = (\lambda_{1,1}, \cdots, \lambda_{1,K}, \cdots, \lambda_{R,1}, \cdots, \lambda_{R,K})^{'}$, $\varphi = \left(\gamma_{1,1}^{'}, \cdots, \gamma_{1,K}^{'}, \cdots, \gamma_{R,1}^{'}, \cdots, \gamma_{R,K}^{'}, \beta^{'}\right)^{'}$, $Z_n = (H_{1,n} W_{1,n} X_n, \cdots, H_{K,n} W_{1,n} X_n, \cdots, H_{1,n} W_{R,n} X_n, \cdots, H_{K,n} W_{R,n} X_n, X_n)^{'}$ and $S_n(\Lambda) = I_n - \sum_{r=1}^{R}\sum_{k=1}^{K} \lambda_{r,k} H_{k,n} W_{r,n}$, the maximum likelihood function is

$$\ln L_n(\theta, \alpha) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\Lambda)| - \frac{1}{2\sigma^2} u_{\theta,\alpha}^{'} u_{\theta,\alpha}$$

which has exactly the same form as the maximum likelihood function (5) for the single network specification. Since the major difference is that there are more parameters in Λ and φ, the concentrated approach discussed in Section 2.3 are also applicable as φ are linear parameters that can be concentrated out. Define $\widetilde{W}_{n,r,k} = H_{k,n} W_{r,n}$ and $G_{n,r,k} = \widetilde{W}_{n,r,k} S_n^{-1}$, by replacing $\widetilde{W}_{n,r,k}$ and $G_{n,r,k}$, the concentrated log-likelihood functions have the same form as (5) and (12)[6]. The biased correction approach is also the same as (14) with slight modification.

3 Asymptotic Properties of the QMLE

We focus on the single network specification for the discussion of asymptotic properties of the QMLE since the difference of the maximum likelihood functions and the bias correction procedure between the single network and multiple networks scenarios are modest, i.e., they only differ in the number of Λ and φ and some changes of notations.

3.1 Identification

As stated in the last section, we will discuss the property of QMLE based on the concentrated log-likelihood function (6) due to the convenience to separately discuss the properties of estimators for group fixed effects and other parameters. The identification conditions below are based on Rothenberg (1971). The expected log-likelihood function for equation (6) is

$$\begin{aligned} Q_n(\theta) &= \mathrm{E}\ln L_n(\theta) \\ &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 + \ln|S_n(\Lambda)| - \frac{1}{2\sigma^2}\mathrm{E}(\widetilde{u}_\theta^{'} \widetilde{H}_G^{'} \widetilde{H}_G \widetilde{u}_\theta) \end{aligned} \tag{17}$$

Then $\varphi_n(\Lambda) = \arg\max_\varphi Q_n(\theta) = [Z_n^{'} \widetilde{H}_G Z_n]^{-1} Z_n^{'} \widetilde{H}_G S_n(\Lambda) S_n^{-1} Z_n \varphi_0$ for each $\Lambda = (\lambda_1, ..., \lambda_K)^{'}$, and

[6] By the concentrated log-likelihood function for numerical searches, the number of parameters now can be reduced from $(KR+1)(L+1)$ (or $(KR+1)(L+1)+G$ if the group fixed effects are considered) to KR.

$$
\begin{aligned}
&\mathrm{E}\left(\widetilde{u}_\theta'\widetilde{H}_G'\widetilde{H}_G\widetilde{u}_\theta\right)\\
=&\mathrm{E}[(S_n(\Lambda)Y_n - Z_n\varphi_n(\Lambda))'\widetilde{H}_G'\widetilde{H}_G(S_n(\Lambda)Y_n - Z_n\varphi_n(\Lambda))]\\
=&[B_nS_n(\Lambda)S_n^{-1}Z_n\varphi_0]'(B_nS_n(\Lambda)S_n^{-1}Z_n\varphi_0) + \sigma_0^2\mathrm{tr}(S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1})
\end{aligned} \tag{18}
$$

Observe that $S_n(\Lambda)S_n^{-1} = I_n + (\lambda_{1,0} - \lambda_1)G_{n,1} + (\lambda_{2,0} - \lambda_2)G_{n,2} + \ldots + (\lambda_{K,0} - \lambda_K)G_{n,K}$, hence

$$
\begin{aligned}
&\mathrm{E}(\widetilde{u}_\theta'\widetilde{H}_G'\widetilde{H}_G\widetilde{u}_\theta)\\
=&(\Lambda_0 - \Lambda)'(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)'B_n'B_n(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)(\Lambda_0 - \Lambda)\\
+&\sigma_0^2\mathrm{tr}(S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1})
\end{aligned} \tag{19}
$$

by using the property that $B_nZ_n = \widetilde{H}_G(I_n - Z_nA_n)Z_n = 0$. Then we obtain

$$
\begin{aligned}
\sigma_n^2(\Lambda) &= \arg\max_{\sigma^2} Q_n(\Lambda, \varphi_n(\Lambda), \sigma^2)\\
&= \frac{1}{n}(\Lambda_0 - \Lambda)'(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)'B_n'B_n(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)(\Lambda_0 - \Lambda)\\
&+ \frac{\sigma_0^2}{n}\mathrm{tr}(S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1})
\end{aligned} \tag{20}
$$

Define $Q_n(\Lambda) = Q_n(\Lambda, \varphi_n(\Lambda), \sigma_n^2(\Lambda))$, we consider

$$
\begin{aligned}
&Q_n(\Lambda) - Q_n(\Lambda_0)\\
&= -\frac{1}{2}(\ln\sigma_n^2(\Lambda) - \ln\sigma_0^2) + \frac{1}{n}(\ln|S_n(\Lambda)| - \ln|S_n|)\\
&= \frac{1}{2}\Bigg[\ln|S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1}|^{\frac{1}{n}} - \ln\bigg(\frac{1}{n}\mathrm{tr}(S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1})\\
&+ \frac{1}{n\sigma_0^2}(\Lambda_0 - \Lambda)'(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)'B_n'B_n(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)(\Lambda_0 - \Lambda)\bigg)\Bigg]
\end{aligned}
$$

A unique identification condition requires that $Q_n(\Lambda) - Q_n(\Lambda_0) < 0$ when $\Lambda \neq \Lambda_0$ under large n. Above equation takes a form of $\frac{1}{2}\big[\ln|D|^{\frac{1}{n}} - \ln(\frac{1}{n}\mathrm{tr}(\mathbf{D}) + \mathbf{Q})\big]$, where $\mathbf{D}$ is a symmetric matrix and $\mathbf{Q}$ is a non-negative quadratic form.

There are two identification sources. Since all eigenvalues of $S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1}$ are real and positive, and each one is denoted by $\Psi_i(\Lambda)$, then $|S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1}|^{\frac{1}{n}} = (\prod_i^n \Psi_i(\Lambda))^{\frac{1}{n}}$ and $\frac{1}{n}\mathrm{tr}(S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1}) = \frac{1}{n}\sum_i^n \Psi_i(\Lambda)$. By the inequality of arithmetic and geometric means, $\frac{1}{n}\sum_i^n \Psi_i(\Lambda) \geq (\prod_i^n \Psi_i(\Lambda))^{\frac{1}{n}}$. The first identification source can be obtained if $\frac{1}{n}\sum_i^n \Psi_i(\Lambda) > (\prod_i^n \Psi_i(\Lambda))^{\frac{1}{n}}$ when $\Lambda \neq \Lambda_0$. It can be achieved when $S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1}$ is not proportional to I_n when $\Lambda \neq \Lambda_0$ (**Claim 1** provides a sufficient (not necessary) condition for this requirement). The second identification source is the "**Q**-term". Λ_0 can be identified if

$$
\lim_{n\to\infty}\frac{1}{n}(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)'B_n'B_n(G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0)
$$

exists and is nonsingular. This condition ensures that the set of regressors $[G_{n,1}Z_n\varphi_0, \ldots, G_{n,K}Z_n\varphi_0]$

have no multicollinearity.

Claim 1:

Since $\widetilde{W}_{n,k}$ is not symmetric (i.e., $H_k W_n \neq W_n^{'} H_k^{'}$), if I_n, $\widetilde{W}_{n,k} + \widetilde{W}_{n,k}^{'}$ for $k = 1, ..., K$, $\widetilde{W}_{n,1}^{'}\widetilde{W}_{n,1}, \ldots, \widetilde{W}_{n,1}^{'}\widetilde{W}_{n,K}, \ldots, \widetilde{W}_{n,K}^{'}\widetilde{W}_{n,1}, \ldots, \widetilde{W}_{n,K}^{'}\widetilde{W}_{n,K}$ are linearly independent, $S_n^{'-1}S_n^{'}(\Lambda)S_n(\Lambda)S_n^{-1}$ is not proportional to I_n when $\Lambda \neq \Lambda_0$.

Proof:

For some constant c, suppose $S_n^{'-1}S_n^{'}(\Lambda)S_n(\Lambda)S_n^{-1} = cI_n$, i.e., $S_n^{'}(\Lambda)S_n(\Lambda) = cS_n^{'}S_n$. Then we have

$$\begin{aligned}\mathbf{0} = (1-c)I_n - \sum_{k=1}^{K}(\lambda_k - c\lambda_{k,0})(\widetilde{W}_{n,k} + \widetilde{W}_{n,k}^{'}) \\ + (\lambda_1^2 - c\lambda_{1,0}^2)\widetilde{W}_{n,1}^{'}\widetilde{W}_{n,1} + \ldots + (\lambda_1\lambda_K - c\lambda_{1,0}\lambda_{K,0})\widetilde{W}_{n,1}^{'}\widetilde{W}_{n,K} \\ + \ldots \\ + (\lambda_K\lambda_1 - c\lambda_{K,0}\lambda_{1,0})\widetilde{W}_{n,K}^{'}\widetilde{W}_{n,1} + \ldots + (\lambda_K^2 - c\lambda_{K,0}^2)\widetilde{W}_{n,K}^{'}\widetilde{W}_{n,K}\end{aligned} \tag{21}$$

The linear independence assumption implies that $c = 1$, $\lambda_1 = \lambda_{1,0}$, $\lambda_2 = \lambda_{2,0}$,$\ldots$, $\lambda_K = \lambda_{K,0}$. $\square$

3.2 Consistency

For consistency, we need the following assumptions:

Assumption 4:

Denote $c_u = sup_n||W_n||_1$, $c_w = sup_n||W_n||_\infty$. The sequence $\{W_n\}$ satisfies $max\{c_u, c_w\} < \infty$, i.e., it's uniformly bounded in both row and column sum norms. Θ_Λ denotes a compact parameter space for Λ and assume that Λ_0 belongs to the interior of Θ_Λ. The sequence $\{S_n^{-1}(\Lambda)\}$ satisfies $max_{\Lambda\in\Theta_\Lambda}\{sup_n||S_n^{-1}(\Lambda)||_\infty, sup_n||S_n^{-1}(\Lambda)||_1\} < \infty$.

Assumption 5:

Elements of X_n have uniformly bounded constants. Or, if one wants to assume that X_n is stochastic, then $max_{l=1,\cdots,L}sup_{n,i}E|x_{i,l}|^{4+\eta} < \infty$ for some $\eta > 0$; and X_n and u_n are independent. Also, $lim_{n\to\infty}X_n^{'}X_n$ exists and is nonsingular.

Assumption 6:

The parameter space Θ of θ is compact. The true value θ_0 belongs to the interior of Θ.

Assumption 7:

$u_i \overset{i.i.d.}{\sim} (0, \sigma_0^2)$ with $\sigma_0^2 > 0$, and $sup_{n,i}E\left|u_i\right|^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption 8 (Identification):

At least, one of the two conditions holds:

(i) $S_n^{'-1}S_n^{'}(\Lambda)S_n(\Lambda)S_n^{-1}$ is not proportional to I_n when $\Lambda \neq \Lambda_0$;

(ii) $lim_{n\to\infty}\frac{1}{n}(G_{n,1}Z_n\varphi_0,\ldots,G_{n,K}Z_n\varphi_0)'B_n'B_n(G_{n,1}Z_n\varphi_0,\ldots,G_{n,K}Z_n\varphi_0)$ *exists and is non-singular.*

With global identification (Assumption 8), it suffices to show that $\sup_{\theta\in\Theta}\frac{1}{n}|\ln L_n(\theta)-E\ln L_n(\theta)|\xrightarrow{p}0$ and the uniform equicontinuity of $\{\frac{E\ln L_n(\theta)}{n}\}_{n=1}^{\infty}$. Then, we have:

Theorem 1:

Under Assumption 1-8, the QMLE of θ is consistent.

Proof:

With global identification (Assumption 8), it suffices to show that $\sup_{\theta\in\Theta}\frac{1}{n}|\ln L_n(\theta)-E\ln L_n(\theta)|\xrightarrow{p}0$ (uniform convergence) and the uniform equicontinuity of $\{\frac{E\ln L_n(\theta)}{n}\}_{n=1}^{\infty}$, then we can get consistency.

Proof of the uniform convergence:

Denote

$$\begin{aligned}\widetilde{V}_n(\Lambda,\varphi)&=\widetilde{H}_G\widetilde{u}_\theta\\&=\widetilde{H}_GY_n-\sum_{k=1}^{K}\lambda_k\widetilde{H}_G\widetilde{W}_{n,k}Y_n-\widetilde{H}_GZ_n\varphi\\&=\widetilde{V}_n-\sum_{k=1}^{K}(\lambda_k-\lambda_{k,0})\widetilde{H}_G\widetilde{W}_{n,k}Y_n-\widetilde{H}_GZ_n(\varphi-\varphi_0)\end{aligned}$$

where $\widetilde{V}_n=\widetilde{V}_n(\Lambda_0,\varphi_0)$.

Then we have

$$\begin{aligned}&\widetilde{V}_n'(\Lambda,\varphi)\widetilde{V}_n(\Lambda,\varphi)\\=&\widetilde{V}_n'\widetilde{V}_n+\sum_{k=1}^{K}(\lambda_k-\lambda_{k,0})(\widetilde{H}_G\widetilde{W}_{n,k}Y_n)'\sum_{k=1}^{K}(\lambda_k-\lambda_{k,0})\widetilde{H}_G\widetilde{W}_{n,k}Y_n\\&+(\varphi-\varphi_0)'(\widetilde{H}_GZ_n)'\widetilde{H}_GZ_n(\varphi-\varphi_0)+2\sum_{k=1}^{K}(\lambda_k-\lambda_{k,0})(\widetilde{H}_G\widetilde{W}_{n,k}Y_n)'\widetilde{H}_GZ_n(\varphi-\varphi_0)\\&-2\sum_{k=1}^{K}(\lambda_k-\lambda_{k,0})(\widetilde{H}_G\widetilde{W}_{n,k}Y_n)'\widetilde{V}_n-2(\varphi-\varphi_0)'(\widetilde{H}_GZ_n)'\widetilde{V}_n\end{aligned}\tag{22}$$

Using $\widetilde{H}_G\widetilde{W}_{n,k}Y_n=\widetilde{H}_GG_{n,k}Z_n\varphi_0+\widetilde{H}_GG_{n,k}\widetilde{V}_n$, for $\forall k_1,k_2=1,\ldots,K$, we have

$$\begin{aligned}&(\widetilde{H}_G\widetilde{W}_{n,k_1}Y_n)'(\widetilde{H}_G\widetilde{W}_{n,k_2}Y_n)\\=&(\widetilde{H}_GG_{n,k_1}Z_n\varphi_0)'(\widetilde{H}_GG_{n,k_2}Z_n\varphi_0)+(\widetilde{H}_GG_{n,k_1}\widetilde{V}_n)'(\widetilde{H}_GG_{n,k_2}\widetilde{V}_n)\\&+(\widetilde{H}_GG_{n,k_1}Z_n\varphi_0)'(\widetilde{H}_GG_{n,k_2}\widetilde{V}_n)+(\widetilde{H}_GG_{n,k_1}\widetilde{V}_n)'(\widetilde{H}_GG_{n,k_2}Z_n\varphi_0)\end{aligned}\tag{23}$$

Under Assumptions 4 - 8, note that $\widetilde{H}_G'\widetilde{H}_G = \widetilde{H}_G$, for any $n \times n$ non-stochastic uniformly bounded matrices P_{1n} and P_{2n},

$$
\begin{aligned}
&\frac{1}{n}(\widetilde{H}_G P_{1n} Z_n)'(\widetilde{H}_G P_{2n} Z_n) - E\frac{1}{n}(\widetilde{H}_G P_{1n} Z_n)'(\widetilde{H}_G P_{2n} Z_n) = O_p(\frac{1}{\sqrt{n}}) \\
&\frac{1}{n}(\widetilde{H}_G P_{1n} Z_n)'(\widetilde{H}_G P_{2n} \widetilde{V}_n) - E\frac{1}{n}(\widetilde{H}_G P_{1n} Z_n)'(\widetilde{H}_G P_{2n} \widetilde{V}_n) = O_p(\frac{1}{\sqrt{n}}) \\
&\frac{1}{n}(\widetilde{H}_G P_{1n} \widetilde{V}_n)'(\widetilde{H}_G P_{2n} \widetilde{V}_n) - E\frac{1}{n}(\widetilde{H}_G P_{1n} \widetilde{V}_n)'(\widetilde{H}_G P_{2n} \widetilde{V}_n) = O_p(\frac{1}{\sqrt{n}})
\end{aligned} \tag{24}
$$

where $E\frac{1}{n}(\widetilde{H}_G P_{1n} Z_n)'(\widetilde{H}_G P_{2n} Z_n)$ is $O(1)$, $E\frac{1}{n}(\widetilde{H}_G P_{1n} Z_n)'(\widetilde{H}_G P_{2n} \widetilde{V}_n)$ is $O(\frac{1}{\sqrt{n}})$, $E\frac{1}{n}(\widetilde{H}_G P_{1n} \widetilde{V}_n)'(\widetilde{H}_G P_{2n} \widetilde{V}_n)$ is $O(1)$. Since Λ and φ are bounded in Θ, we have $\frac{1}{n}\widetilde{V}_n'(\Lambda,\varphi)\widetilde{V}_n(\Lambda,\varphi) - E\frac{1}{n}\widetilde{V}_n'(\Lambda,\varphi)\widetilde{V}_n(\Lambda,\varphi) \xrightarrow{p} 0$ uniformly in θ in Θ. Note that

$$
\frac{1}{n}\ln L_n(\theta) = -\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 + \frac{1}{n}\ln|S_n(\Lambda)| - \frac{1}{2\sigma^2 n}\widetilde{V}_n'(\Lambda,\varphi)\widetilde{V}_n(\Lambda,\varphi)
$$

since σ^2 is bounded away from zero in Θ, then

$$
\begin{aligned}
&\sup_{\theta\in\Theta}\frac{1}{n}\left|\ln L_n(\theta) - E\ln L_n(\theta)\right| \\
=&\sup_{\theta\in\Theta}\left|-\frac{1}{2\sigma^2}(\frac{1}{n}\widetilde{V}_n'(\Lambda,\varphi)\widetilde{V}_n(\Lambda,\varphi) - E\frac{1}{n}\widetilde{V}_n'(\Lambda,\varphi)\widetilde{V}_n(\Lambda,\varphi)\right| \\
\xrightarrow{p}&0
\end{aligned} \tag{25}
$$

Proof of the uniform equicontinuity:

Since $\widetilde{V}_n(\Lambda,\varphi) = \widetilde{H}_G S_n(\Lambda) S_n^{-1} Z_n\varphi_0 - \widetilde{H}_G Z_n\varphi + \widetilde{H}_G S_n(\Lambda) S_n^{-1}\widetilde{u}_{\theta_0}$, we have

$$
\begin{aligned}
E\frac{1}{n}\widetilde{V}_n'(\Lambda,\varphi)\widetilde{V}_n(\Lambda,\varphi) = &\underbrace{\frac{1}{n}E[\widetilde{H}_G S_n(\Lambda)S_n^{-1}Z_n\varphi_0 - \widetilde{H}_G Z_n\varphi]'[\widetilde{H}_G S_n(\Lambda)S_n^{-1}Z_n\varphi_0 - \widetilde{H}_G Z_n\varphi]}_{\textbf{Term 1}} \\
&+ \underbrace{\frac{1}{n}\sigma_0^2\mathrm{tr}(S_n'^{-1}S_n'(\Lambda)S_n(\Lambda)S_n^{-1})}_{\textbf{Term 2}} \\
&+ \underbrace{\frac{2}{n}E[\widetilde{H}_G S_n(\Lambda)S_n^{-1}Z_n\varphi_0 - \widetilde{H}_G Z_n\varphi]'[\widetilde{H}_G S_n(\Lambda)S_n^{-1}\widetilde{V}_n]}_{\textbf{Term 3}}
\end{aligned} \tag{26}
$$

Term 3 is a polynomial function in θ, Θ is bounded, then **Term 3** is $O(\frac{1}{\sqrt{n}})$ uniformly in θ in Θ. **Term 1** is equivalent to

$$\omega' \mathrm{E} \underbrace{\begin{bmatrix} Z_n' Z_n & (G_{n,1}Z_n)' Z_n & \dots & (G_{n,K}Z_n)' Z_n \\ Z_n' G_{n,1} Z_n & (G_{n,1}Z_n)' G_{n,1} Z_n & \dots & (G_{n,K}Z_n)' G_{n,1} Z_n \\ \vdots & \vdots & & \vdots \\ Z_n' G_{n,K} Z_n & (G_{n,1}Z_n)' G_{n,K} Z_n & \dots & (G_{n,K}Z_n)' G_{n,K} Z_n \end{bmatrix}}_{EJ_n} \omega$$

where $\omega = \left(\varphi' - \varphi_0', \lambda_1 - \lambda_{1,0}, \dots, \lambda_K - \lambda_{K,0}\right)'$.

Using $S_n(\Lambda)S_n^{-1} = I_n - (\lambda_1 - \lambda_{1,0})G_{n,1} - \dots - (\lambda_K - \lambda_{K,0})G_{n,K}$. **Term 2** are all polynomial functions of θ. To show $\{\frac{E \ln L_n(\theta)}{n}\}_{n=1}^{\infty}$, we need four sufficient conditions:

(a) $\ln \sigma^2$ is uniformly continuous, which is satisfied because σ^2 is bounded away from zero in Θ;

(b) $\frac{1}{n} \ln |S_n(\Lambda)|$ is uniformly equicontinuous. Note $\frac{1}{n} \ln |S_n(\Lambda_1)| - \frac{1}{n} \ln |S_n(\Lambda_2)| = \frac{1}{n} \mathrm{tr}(W_n S_n^{-1}(\bar{\Lambda}))(\Lambda_2 - \Lambda_1)$, where $\bar{\Lambda}$ lies between Λ_1 and Λ_2. As $\max_{\Lambda \in \Theta_\Lambda} \{\sup_n ||S_n^{-1}(\Lambda)||_\infty, \sup_n ||S_n^{-1}(\Lambda)||_1\} < \infty$, $\frac{1}{n} \mathrm{tr}(W_n S_n^{-1}(\bar{\Lambda}))$ is bounded, the condition is satisfied;

(c) $\omega' EJ_n \omega$ is uniformly equicontinuous since φ and λ are bounded and EJ_n is $O(1)$;

(d) $\sigma_n^2(\Lambda) = \frac{\sigma_0^2}{n} \mathrm{tr}(S_n'^{-1} S_n'(\Lambda) S_n(\Lambda) S_n^{-1})$ is uniformly equicontinuous. Note that

$$\begin{aligned} & \sigma_n^2(\Lambda_2) - \sigma_n^2(\Lambda_1) \\ = & \frac{\sigma_0^2}{n} \mathrm{tr}(S_n'^{-1} S_n'(\Lambda_2) S_n(\Lambda_2) S_n^{-1}) - \frac{\sigma_0^1}{n} \mathrm{tr}(S_n'^{-1} S_n'(\Lambda_1) S_n(\Lambda_1) S_n^{-1}) \\ = & \sigma_0^2 \Big[\sum_{k=1}^K \frac{(\lambda_k^{(1)} - \lambda_k^{(2)}) \mathrm{tr}(G_{n,k}' + G_{n,k})}{n} + \sum_{k=1}^K \frac{(\lambda_k^{(2)} - \lambda_k^{(1)})(\lambda_k^{(1)} + \lambda_k^{(2)} - 2\lambda_{k,0}) \mathrm{tr}(G_{n,k}' G_{n,k})}{n} \\ & + \sum_{k_1=1}^K \sum_{k_2=1}^K \frac{\big((\lambda_{k_1}^{(2)} \lambda_{k_2}^{(2)} - \lambda_{k_1}^{(1)} \lambda_{k_2}^{(1)}) + \lambda_{k_1,0}(\lambda_{k_2}^{(1)} - \lambda_{k_2}^{(2)}) + \lambda_{k_2,0}(\lambda_{k_1}^{(1)} - \lambda_{k_1}^{(2)})\big) \mathrm{tr}(G_{n,k_1}' G_{n,k_2})}{n} \Big] \end{aligned} \quad (27)$$

by using $S_n(\Lambda)S_n^{-1} = I_n - (\lambda_1 - \lambda_{1,0})G_{n,1} - \dots - (\lambda_K - \lambda_{K,0})G_{n,K}$. Since for any $k = 1, \dots, K; k_1 = 1, \dots, K; k_2 = 1, \dots, K$, $G_{n,k}$, $G_{n,k}' G_{n,k}$ and $G_{n,k_1}' G_{n,k_2}$ are uniformly bounded, then $\sigma_n^2(\Lambda)$ is uniformly equicontinuous. □

3.3 Asymptotic Distribution of the QMLE

To derive the asymptotic distribution, first, we need to decompose the first order derivatives (evaluated at true parameter values) as $\frac{1}{\sqrt{n}} \frac{\partial \ln L_n(\theta_0)}{\partial \theta} = \frac{1}{\sqrt{n}} \frac{\ln L_n^{(\nu)}(\theta_0)}{\partial \theta} - \widetilde{\Delta}_n$, where

$$\frac{1}{\sqrt{n}} \frac{\ln L_n^{(\nu)}(\theta_0)}{\partial \theta} = \begin{pmatrix} -\frac{1}{\sqrt{n}} \mathrm{tr}[(\widetilde{H}_G' \widetilde{H}_G) G_{n,1}] + \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} [(\widetilde{W}_{n,1} Y_n)' (\widetilde{H}_G' \widetilde{H}_G) u_n] \\ \vdots \\ -\frac{1}{\sqrt{n}} \mathrm{tr}[(\widetilde{H}_G' \widetilde{H}_G) G_{n,K}] + \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} [(\widetilde{W}_{n,K} Y_n)' (\widetilde{H}_G' \widetilde{H}_G) u_n] \\ \frac{1}{\sqrt{n}} \frac{1}{\sigma_0^2} Z_n' (\widetilde{H}_G' \widetilde{H}_G) u_n \\ -\frac{1}{\sqrt{n}} \frac{1}{2\sigma_0^2} \mathrm{tr}(\widetilde{H}_G' \widetilde{H}_G) + \frac{1}{\sqrt{n}} \frac{1}{2\sigma_0^4} u_n' (\widetilde{H}_G' \widetilde{H}_G) u_n \end{pmatrix} \quad (28)$$

and

$$
\frac{1}{\sqrt{\mathrm{n}}}\widetilde{\Delta}_n = \begin{pmatrix} \frac{1}{\sqrt{n}}\mathrm{tr}[(I_n - \widetilde{H}_G^{'}\widetilde{H}_G)G_{n,1}] \\ \vdots \\ \frac{1}{\sqrt{n}}\mathrm{tr}[(I_n - \widetilde{H}_G^{'}\widetilde{H}_G)G_{n,K}] \\ \mathbf{0} \\ \frac{1}{\sqrt{n}}\frac{1}{2\sigma_0^2}\mathrm{tr}(I_n - \widetilde{H}_G^{'}\widetilde{H}_G) \end{pmatrix} \tag{29}
$$

The second order conditions are as follows:

$$
\begin{aligned}
\frac{\partial \ln L_n^{\nu}(\theta_0)}{\partial\lambda_k\partial\lambda_k} &= -\mathrm{tr}[(\widetilde{H}_G^{'}\widetilde{H}_G)G_{n,k}^2] - \frac{1}{\sigma_0^2}(\widetilde{W}_{n,k}Y_n)^{'}(\widetilde{H}_G^{'}\widetilde{H}_G)(\widetilde{W}_{n,k}Y_n), \text{for } k = 1,\ldots,K \\
\frac{\partial \ln L_n^{\nu}(\theta_0)}{\partial\lambda_{k_1}\partial\lambda_{k_2}} &= -\mathrm{tr}[(\widetilde{H}_G^{'}\widetilde{H}_G)(G_{n,k_1}G_{n,k_2})] - \frac{1}{\sigma_0^2}(\widetilde{W}_{n,k_1}Y_n)^{'}(\widetilde{H}_G^{'}\widetilde{H}_G)(\widetilde{W}_{n,k_2}Y_n), \text{for } k_1,k_2 = 1,\ldots,K \\
\frac{\partial \ln L_n^{\nu}(\theta_0)}{\partial\lambda_k\partial\varphi} &= -\frac{1}{\sigma_0^2}(\widetilde{W}_{n,k}Y_n)^{'}(\widetilde{H}_G^{'}\widetilde{H}_G)Z_n \\
\frac{\partial \ln L_n^{\nu}(\theta_0)}{\partial\lambda_k\partial\sigma^2} &= -\frac{1}{\sigma_0^4}(\widetilde{W}_{n,k}Y_n)^{'}(\widetilde{H}_G^{'}\widetilde{H}_G)u_n \\
\frac{\partial \ln L_n^{\nu}(\theta_0)}{\partial\varphi\partial\varphi} &= -\frac{1}{\sigma_0^2}Z_n^{'}(\widetilde{H}_G^{'}\widetilde{H}_G)Z_n \\
\frac{\partial \ln L_n^{\nu}(\theta_0)}{\partial\varphi\partial\sigma^2} &= -\frac{1}{\sigma_0^4}Z_n^{'}(\widetilde{H}_G^{'}\widetilde{H}_G)u_n \\
\frac{\partial \ln L_n^{\nu}(\theta_0)}{\partial\sigma^2\partial\sigma^2} &= \frac{1}{2\sigma_0^4}\mathrm{tr}(\widetilde{H}_G^{'}\widetilde{H}_G) - \frac{1}{\sigma_0^6}u_n^{'}(\widetilde{H}_G^{'}\widetilde{H}_G)u_n
\end{aligned}
$$

As $\widetilde{H}_G^{'}\widetilde{H}_G = \widetilde{H}_G$, the variance matrix of $\frac{1}{\sqrt{n}}\frac{\ln L_n(\theta_0)}{\partial\theta}$ is equal to

$$E(\frac{1}{\sqrt{n}}\frac{\ln L_n(\theta_0)}{\partial\theta}\cdot\frac{1}{\sqrt{n}}\frac{\ln L_n(\theta_0)}{\partial\theta'})$$

$$= \underbrace{\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1K} & \frac{1}{\sigma_0^2 n}(\widetilde{G}_{n,1}(Z_n\varphi_0 + H_G\alpha_0))'\widetilde{Z}_n & \frac{1}{n\sigma_0^2}\mathrm{tr}(\widetilde{G}_{n,1}) \\ * & a_{22} & \dots & a_{2K} & \frac{1}{\sigma_0^2 n}(\widetilde{G}_{n,2}(Z_n\varphi_0 + H_G\alpha_0))'\widetilde{Z}_n & \frac{1}{n\sigma_0^2}\mathrm{tr}(\widetilde{G}_{n,2}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ * & * & \dots & a_{KK} & \frac{1}{\sigma_0^2 n}(\widetilde{G}_{n,K}(Z_n\varphi_0 + H_G\alpha_0))'\widetilde{Z}_n & \frac{1}{n\sigma_0^2}\mathrm{tr}(\widetilde{G}_{n,K}) \\ * & * & \dots & * & \frac{1}{\sigma_0^2 n}\widetilde{Z}_n'\widetilde{Z}_n & 0 \\ * & * & \dots & * & * & \frac{n-G}{2\sigma_0^4 n} \end{bmatrix}}_{\Sigma_{\theta_0,n}}$$

$$+ \underbrace{\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1K} & \frac{\mu_{3,0}}{\sigma_0^4 n}vec_D'\left(\widetilde{G}_{n,1}\right)\widetilde{Z}_n & \frac{\mu_{3,0}}{2\sigma_0^6 n}\left(\widetilde{G}_{n,1}\left(Z_n\varphi_0 + H_G\alpha_0\right)\right)' l_n + \frac{\mu_{4,0}-3\sigma_0^4}{2\sigma_0^6 n}\mathrm{tr}(\widetilde{G}_{n,1}) \\ * & b_{22} & \dots & b_{2K} & \frac{\mu_{3,0}}{\sigma_0^4 n}vec_D'\left(\widetilde{G}_{n,2}\right)\widetilde{Z}_n & \frac{\mu_{3,0}}{2\sigma_0^6 n}\left(\widetilde{G}_{n,2}\left(Z_n\varphi_0 + H_G\alpha_0\right)\right)' l_n + \frac{\mu_{4,0}-3\sigma_0^4}{2\sigma_0^6 n}\mathrm{tr}(\widetilde{G}_{n,2}) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ * & * & \dots & b_{KK} & \frac{\mu_{3,0}}{\sigma_0^4 n}vec_D'\left(\widetilde{G}_{n,K}\right)\widetilde{Z}_n & \frac{\mu_{3,0}}{2\sigma_0^6 n}\left(\widetilde{G}_{n,K}\left(Z_n\varphi_0 + H_G\alpha_0\right)\right)' l_n + \frac{\mu_{4,0}-3\sigma_0^4}{2\sigma_0^6 n}\mathrm{tr}(\widetilde{G}_{n,K}) \\ * & * & \dots & * & 0 & \frac{\mu_{3,0}}{2\sigma_0^6 n}\widetilde{Z}_n' l_n \\ * & * & \dots & * & * & \frac{\mu_{4,0}-3\sigma_0^4}{4\sigma_0^8 n}\cdot(n-G) \end{bmatrix}}_{\Omega_{\theta_0,n}-\Sigma_{\theta_0,n}}$$

with

$$\begin{aligned} a_{kk} &= \frac{1}{\sigma_0^2 n}\big(\widetilde{G}_{n,k}(Z_n\varphi_0 + H_G\alpha_0)\big)'\big(\widetilde{G}_{n,k}(Z_n\varphi_0 + H_G\alpha_0)\big) + \frac{1}{n}\big(\mathrm{tr}(\widetilde{G}_{n,k}'\widetilde{G}_{n,k}) + \mathrm{tr}(\widetilde{G}_{n,k}^2)\big), \\ a_{k_1k_2} &= \frac{1}{\sigma_0^2 n}\big(\widetilde{G}_{n,k_1}(Z_n\varphi_0 + H_G\alpha_0)\big)'\big(\widetilde{G}_{n,k_2}(Z_n\varphi_0 + H_G\alpha_0)\big) + \frac{1}{n}\left(\mathrm{tr}(\widetilde{G}_{n,k_1}'\widetilde{G}_{n,k_2}) + \mathrm{tr}\left(\widetilde{G}_{n,k_1}\widetilde{G}_{n,k_2}\right)\right), \\ b_{kk} &= \frac{2\mu_{3,0}}{\sigma_0^4 n}\big(\widetilde{G}_{n,k}(Z_n\varphi_0 + H_G\alpha_0)\big)' vec_D\left(\widetilde{G}_{n,k}\right) + \frac{\mu_{4,0}-3\sigma_0^4}{\sigma_0^4 n}vec_D'\left(\widetilde{G}_{n,k}\right)vec_D\left(\widetilde{G}_{n,k}\right) \\ b_{k_1k_2} &= \frac{\mu_{3,0}}{\sigma_0^4 n}\big(\widetilde{G}_{n,k_1}(Z_n\varphi_0 + H_G\alpha_0)\big)' vec_D\left(\widetilde{G}_{n,k_2}\right) + \frac{\mu_{3,0}}{\sigma_0^4 n}\big(\widetilde{G}_{n,k_2}(Z_n\varphi_0 + H_G\alpha_0)\big)' vec_D\left(\widetilde{G}_{n,k_1}\right) \\ &\quad + \frac{\mu_{4,0}-3\sigma_0^4}{\sigma_0^4 n}vec_D'\left(\widetilde{G}_{n,k_1}\right)vec_D\left(\widetilde{G}_{n,k_2}\right), \end{aligned}$$

for $\forall\ k, k_1, k_2 = 1, \dots, K$, where $\widetilde{G}_{n,k} = \widetilde{H}_G G_{n,k}$, $\widetilde{Z}_n = \widetilde{H}_G Z_n$. ($\mu_{4,0} - 3\sigma_0^4 = 0$ if u_n are normally distributed). Note that, for the multiple networks situation, we have more terms in the above matrix and need to replace a_{kk}, $a_{k_1k_2}$, b_{kk} and $b_{k_1k_2}$ with $a_{(r,k),(r,k)}$, $a_{(r_1,k_3)(r_2,k_4)}$, $b_{(r,k),(r,k)}$ and $b_{(r_1,k_3)(r_2,k_4)}$, where $\forall\ r, r_1, r_2 = 1, \dots, R$ and $\forall\ k, k_3, k_4 = 1, \dots, K$.

Now, we need an assumption for the information matrix:

Assumption 9:

$\Sigma_{\theta_0} = lim_{n\to\infty}\Sigma_{\theta_0,n}$ is nonsingular where $\Sigma_{\theta_0,n} = E(-\frac{1}{n}\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'})$.

Then, we have

Theorem 2:
With Assumption 1-9, we have $\sqrt{n}(\hat{\theta}_n^c - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1})$ where $\hat{\theta}_n^c = \hat{\theta}_n + \Sigma_{\hat{\theta},n}^{-1}\widetilde{\Delta}_n(\hat{\theta}_n)$.

Proof:
Based on Taylor expansion, $\sqrt{n}(\hat{\theta}_n^c - \theta_0) = (-\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'})^{-1} \cdot (\frac{1}{\sqrt{n}}\frac{\partial \ln L_n^{(\nu)}(\theta_0)}{\partial\theta} - \widetilde{\Delta}_n)$, where $\bar{\theta}_n$ lies between $\hat{\theta}^c$ and θ_0, and $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n(\theta_0)}{\partial\theta} = \frac{1}{\sqrt{n}}\frac{\partial \ln L_n^{(\nu)}(\theta_0)}{\partial\theta} - \widetilde{\Delta}_n$, $\widetilde{\Delta}_n = O(1)$. First, we show the asymptotic distribution of $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n^{(\nu)}(\theta_0)}{\partial\theta}$. Note that $\frac{1}{\sqrt{n}}\frac{\partial \ln L_n^{(\nu)}(\theta_0)}{\partial\theta}$ generally takes the following linear-quadratic form:

$$S_n = \frac{1}{\sigma_0^2\sqrt{n}}T_n^{'}u_n + \frac{1}{\sigma_0^2\sqrt{n}}\left(u_n^{'}\mathcal{R}_n u_n - \sigma_0^2\mathrm{tr}(\mathcal{R}_n)\right)$$

where $T_n = (t_1, \ldots, t_n)^{'}$ is a vector of constants, $\mathcal{R}_n$ is an n-dimensional uniformly bounded symmetric matrix in both row and column sum (see Assumption 4). S_n can be represented by a single summation $S_n = \frac{1}{\sqrt{n}}\sum_{i=1}^n q_{i,n}$ where $q_{i,n} = t_i u_i + \mathbf{r}_i(u_i^2 - \sigma_0^2) + 2u_i\sum_{i'=1}^{i-1}\mathbf{r}_{ii'}u_i u_i^{'}$. We can define a σ-field $\mathcal{F}_{i,n} = \sigma(u_1, \ldots, u_i)$ for $i = 1, \ldots, n$ and $\mathcal{F}_{0,n} = \{\phi, \Omega\}$, a martingale difference double array $\{(q_{i,n}, \mathcal{F}_{i,n})|1 \leq i \leq n\}$ can be formulated, then $E(q_{i,n}|\mathcal{F}_{i-1,n}) = 0$. Then we can apply the martingale central limit theorem to S_n, $\frac{S_n}{\sigma_{S_n}} \xrightarrow{d} N(0,1)$ as $n \to \infty$, and $\sigma_{S_n}^2 = \sum_{i=1}^n E(q_{i,n}^2)$ (refer to Lemma 13 and its proof in Yu et al. (2008)). As $\frac{\sigma_{S_n}^2}{n}$ is bounded away from zero, we have

$$\frac{1}{\sqrt{n}}\frac{\partial \ln L_n^{(\nu)}(\theta_0)}{\partial\theta} \xrightarrow{d} N(0, \Omega_{\theta_0})$$

where $\Omega_{\theta_0} = \lim_{n\to\infty}\Omega_{\theta_0,n}$ and $\Omega_{\theta_0,n} = E(\frac{1}{\sqrt{n}}\frac{\ln L_n^{(\nu)}(\theta_0)}{\partial\theta} \cdot \frac{1}{\sqrt{n}}\frac{\ln L_n^{(\nu)}(\theta_0)}{\partial\theta})$.

Second, we need to show that $\frac{1}{n}|\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'} - E\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}| \xrightarrow{p} 0$. By Assumption 4 and the consistency proof, we can establish $\frac{1}{n}|\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'} - E\frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}| \xrightarrow{p} 0$ and $\frac{1}{n}|\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'} - \frac{\partial^2 \ln L_n(\theta_0)}{\partial\theta\partial\theta'}| \xrightarrow{p} 0$ separately and have the desired result. (For more details, refer to the proof of Theorem 3.2 in Lee (2004)). Then, by Assumption 9, we obtain $(-\frac{1}{n}\frac{\partial^2 \ln L_n(\bar{\theta}_n)}{\partial\theta\partial\theta'})^{-1} - \Sigma_{\theta_0,n}^{-1} = o_p(1)$. As a result, we finally have $\sqrt{n}(\hat{\theta}_n^c - \theta_0) \xrightarrow{d} N(0, \Sigma_{\theta_0}^{-1}\Omega_{\theta_0}\Sigma_{\theta_0}^{-1})$, where $\hat{\theta}_n^c = \hat{\theta}_n + \Sigma_{\hat{\theta},n}^{-1}\widetilde{\Delta}_n(\hat{\theta}_n)$. $\square$

After deriving the asymptotic distribution of $\hat{\theta}_n^c$, finally we can investigate the asymptotic distribution of the group fixed effects $\hat{\alpha} = (\hat{\alpha}_1, ..., \hat{\alpha}_G)'$.

Theorem 3:
With Assumption 1-9, when G is fixed, m_g grows with n, $\sqrt{m_g}(\hat{\alpha}_g - \alpha_{g,0}) \xrightarrow{d} N(0, \sigma_0^2)$ for $g = 1, ..., G$. When m_g is fixed, while G grows with n, the limiting distribution of $\hat{\alpha}_g$ doesn't exist.

Proof:

By first order condition of α of the original log-likelihood equation, we have $\hat{\alpha} = M_G^{-1} H_G' \widetilde{u}_{\hat{\theta}_n}$. Then, using $S_n(\Lambda)S_n^{-1} = I_n + (\lambda_{1,0} - \lambda_1)G_{n,1} + \ldots + (\lambda_{K,0} - \lambda_K)G_{n,K}$, we can get

$$\begin{aligned}
\widetilde{u}_{\hat{\theta}_n} &= S_n(\hat{\Lambda}_n)Y_n - Z_n\hat{\varphi}_n \\
&= S_n(\hat{\Lambda}_n)[S_n^{-1}(Z_n\varphi_0 + H_G\alpha_0 + u_n)] - Z_n\hat{\varphi}_n \\
&= Z_n(\varphi_0 - \hat{\varphi}_n) + H_G\alpha_0 + u_n \\
&+ (\lambda_{1,0} - \hat{\lambda}_{1,n})G_{n,1}(Z_n\varphi_0 + H_G\alpha_0 + u_n) \\
&+ \ldots\ldots \\
&+ (\lambda_{K,0} - \hat{\lambda}_{K,n})G_{n,K}(Z_n\varphi_0 + H_G\alpha_0 + u_n)
\end{aligned}$$

Thus,

$$\begin{aligned}
\hat{\alpha} - \alpha_0 = M_G^{-1}H_G' \cdot \Bigg[\big(G_{n,1}(Z_n\varphi_0 + H_G\alpha_0), ..., G_{n,K}(Z_n\varphi_0 + H_G\alpha_0)\big) \times \begin{pmatrix} \lambda_{1,0} - \hat{\lambda}_{1,n} \\ \vdots \\ \lambda_{K,0} - \hat{\lambda}_{K,n} \\ \varphi_0 - \hat{\varphi}_n \end{pmatrix} \\
+ \big(I_n - \sum_{k=1}^{K}(\lambda_k - \lambda_{k,0})G_{n,k}\big)u_n \Bigg]
\end{aligned}$$

As proved in previous section, we have $\hat{\theta}_n - \theta_0 = O_p(\frac{1}{\sqrt{n}})$ and elements of $\big(G_{n,1}(Z_n\varphi_0 + H_G\alpha_0), ..., G_{n,K}(Z_n\varphi_0 + H_G\alpha_0)\big)$ are $O_p(1)$. Then, the dominant term of $\hat{\alpha} - \alpha_0$ is $M_G^{-1}H_G' u_n$, which is equivalent to

$$\begin{pmatrix} \frac{1}{m_1}\sum_{l=1}^{m_1} u_{l,1} \\ \vdots \\ \vdots \\ \frac{1}{m_G}\sum_{l=1}^{m_G} u_{l,G} \end{pmatrix}$$

where m_g is the number of individuals in group g, $g = 1, \ldots, G$. So for each fixed effect, when m_g grows with n, $\sqrt{m_g}(\hat{\alpha}_g - \alpha_{g,0}) \xrightarrow{d} N(0, \sigma_0^2)$ and they are independent from each other asymptotically. However, when m_g is fixed regardless of G, the limiting distribution of each fixed effect does not exist. $\square$

4 Monte Carlo Simulations

4.1 Basic Settings

In this section, we run Monte Carlo simulations to investigate the finite sample performance of the QMLE we proposed in Section 2. As we stated before, for computational convenience, Λ is estimated by maximizing the concentrated likelihood function $Q(\Lambda)$ showed by (12), and other estimators can then be backed out by (9) $\sim$ (11).

To generate the model, we first simulate the networks among individuals which satisfy our assumptions. In this simulation exercise, we consider the situation when there exist two different networks among agents, and both of them are row-stochastic nearest neighbor spatial weight matrices. For such a matrix $W_n = (w_{ij,n})$, we use the procedure provided by LeSage's econometrics toolbox[7]:

1. Generate two random vectors of coordinates as the geographic location for each observation;
2. Find l nearest neighbors for each observation according to their spatial distances and denote the corresponding $w_{ij,n} = 1$, otherwise $w_{ij,n} = 0$;
3. Row-normalize W_n.

In our application, we consider two networks $W_{1,n}$ and $W_{2,n}$ with l_1 and l_2. The neighborhood relationships are allowed to overlap between these two networks, i.e. an individual may be another individual's neighbor in both networks. For each simulation round, with different sample size, we do 1,000 times replications with the same spatial weighting matrices which are randomly generated.

For heterogeneity source, we consider two different types of individuals and fix the ratio of the two types as 3:2. For external regressors, x_1 is the dummy variable for whether the individual belongs to the first type, and x_2 is a random draw from uniform distribution on $[0, 1]$.

For group settings, we consider two different scenarios as the following:

Scenario 1: fixed number of groups with growing members, 10 groups with $n/10$ members in each group

Scenario 2: fixed membership for each group with growing number of groups, 20 members in each group with totally $n/20$ groups

For both scenarios, group fixed effects $\{\alpha_g\}$ are set as random draws from uniform distribution on $[0, 1]$. To compare the performance under different situations, we also consider two different settings for network density and true parameter values:

Setting 1: $l_1 = 30, l_2 = 20$, $\lambda_1 = (-0.3, 0.7)'$, $\lambda_2 = (0.5, 0.2)'$, $\gamma_1 = (-3, 2, 4, 5)'$, $\gamma_2' = (-1, 2, 2, 3)'$, $\beta = (2, -6)'$, $\sigma^2 = 1$

Setting 2: $l_1 = l_2 = 10$, $\lambda_1 = (0.6, 0.2)'$, $\lambda_2 = (-0.3, -0.5)'$, $\gamma_1 = (1, -2, 2, 3)'$, $\gamma_2' = (2, -2, 1, 3)'$, $\beta = (-1, 3)'$, $\sigma^2 = 4$

[7] See https://www.spatial-econometrics.com

We evaluate the performance under both scenarios and settings, with three different sample sizes: $n = 200$, $n = 500$ and $n = 1000$. Additionally, besides normal distribution, we also evaluate the performance under two other residual distributions: continuous uniform distribution $U\left[-\frac{2}{\sqrt{3}}\sigma, \frac{2}{\sqrt{3}}\sigma\right]$ and rescaled Gamma distribution $\frac{1}{\sqrt{3}}\sigma\left[\Gamma\left(2.25, 2\right) - 4.5\right]$.

4.2 Simulation Results for Scenario 1

Table 1-7 show the performance for Scenario 1 when the number of groups is fixed. Table 1 reports both the raw estimators and bias corrected estimators for peer effects, which are directly estimated by the concentrated likelihood approach. For both parameter settings, the average biases are reduced for all the parameters, and the performance is robust for different residual distributions.

Table 2-7 report the performance for all the parameters except the group fixed effects. For all the three residual distributions and parameter settings, the average biases shrink quickly as sample size grows. Meanwhile, the medians of estimates also get closer to the true value. When sample size is $n = 500$, the average biases are reduced to be less than 10%.

For standard deviation, the performance is not as good as those of the means and medians. Although it shrinks with the sample size for all the parameters, when some contextual effects and effects from own characteristics are relatively small (Setting 2), the standard deviation of their QMLE may still be large. Thus, our QML method is more suitable to be applied in large samples, with at least 1000 observations.

Table 1: Comparison between Raw and Bias Corrected QMLE (Scenario 1)

		Setting 1: $l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)^{'}$											
		Normal				Uniform				Gamma			
mean of estimator		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$
$n = 200$	raw	-.3525	.6278	.4548	.1454	-.3775	.6167	.4548	.1465	-.3533	.6228	.4709	.1502
	bias corrected	-.3080	.6459	.4771	.1703	-.3403	.6954	.4673	.1694	-.3075	.6952	.4836	.1743
$n = 500$	raw	-.3164	.6729	.4861	.1891	-.3313	.6894	.4902	.1811	-.3254	.6777	.4809	.1734
	bias corrected	-.3040	.6920	.4910	.1997	-.3070	.7003	.4932	.1899	-.3100	.6866	.4836	.1847
$n = 1000$	raw	-.3150	.6859	.4891	.1910	-.3198	.6890	.4902	.1935	-.3212	.6872	.4929	.1850
	bias corrected	-.3022	.6966	.4983	.1986	-.3097	.6967	.4948	.1962	-.3051	.6983	.4957	.1943
		Setting 2: $l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)^{'}$											
		Normal				Uniform				Gamma			
mean of estimator		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$
$n = 200$	raw	.5535	.1156	-.3463	-.5640	.5440	.1494	-.3820	-.5382	.5460	.1478	-.3523	-.5845
	bias corrected	.6055	.1933	-.2360	-.4010	.5940	.2023	-.2774	-.4422	.5910	.2063	-.2551	-.4463
$n = 500$	raw	.5773	.1751	-.3391	-.5186	.5775	.1726	-.3309	-.5246	.5777	.1835	-.3316	-.5212
	bias corrected	.5987	.1997	-.2877	-.4613	.6001	.1975	-.2749	-.4692	.5992	.2098	-.2810	-.4638
$n = 1000$	raw	.5876	.1917	-.3111	-.5202	.5874	.1901	-.3209	-.5062	.5897	.1985	-.3146	-.5122
	bias corrected	.5975	.2018	-.2871	-.4971	.5983	.2024	-.2936	-.4783	.5989	.2086	-.2935	-.4872

Table 2: Performance of QMLE with Normal Residuals for Scenario 1 (bias corrected, Setting 1)

		$l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	-.3080	.6459	.4771	.1703	-2.9758	1.9143	3.8506	5.0424
	std	.2957	.4017	.1491	.1881	1.2585	1.5682	1.5864	1.6049
$n = 200$	med	-.3020	.6510	.4909	.1765	-2.9557	1.9555	3.8881	4.9873
	$q_{0.25}$	-.5007	.3779	.3837	.0508	-3.7691	.7920	2.7981	3.9536
	$q_{0.75}$	-.1015	.9140	.5818	.2984	-2.1461	2.9898	4.8867	6.1083
	mean	-.3040	.6920	.4910	.1997	-3.0383	2.0239	3.9443	5.0382
	std	.1349	.1589	.0967	.1098	.7953	.9020	.8846	1.0497
$n = 500$	med	-.3036	.7029	.4906	.1993	-3.0354	2.0211	3.9499	5.0599
	$q_{0.25}$	-.3924	.5846	.4286	.1268	-3.5645	1.4210	3.3669	4.3447
	$q_{0.75}$	-.2059	.8053	.5566	.2797	-2.5127	2.6234	4.5561	5.7033
	mean	-.3022	.6966	.4983	.1986	-2.9753	1.9833	3.9843	4.9877
	std	.0886	.1099	.0582	.0808	.5523	.7596	.6791	.9158
$n = 1000$	med	-.3018	.6957	.4983	.1988	-2.9695	1.9802	3.9857	5.0434
	$q_{0.25}$	-.3613	.6275	.4585	.1439	-3.3755	1.4531	3.5110	4.3568
	$q_{0.75}$	-.2501	.7764	.5396	.2543	-2.6044	2.4987	4.4135	5.6429

		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2
	mean	-1.1083	1.9369	1.9204	2.9682	2.0232	-5.9842	.8782
	std	.9235	1.2154	1.4181	1.8862	1.2497	.2650	.0894
$n = 200$	med	-1.0742	1.9447	1.9091	2.9782	2.0683	-5.9780	.8746
	$q_{0.25}$	-1.6870	1.1365	1.1545	1.7642	1.1596	-6.1538	.8184
	$q_{0.75}$	-.4639	2.6869	2.7309	4.2078	2.8634	-5.8063	.9381
	mean	-1.0377	1.9657	1.9665	3.0841	2.0577	-5.9967	.9528
	std	.5818	.7574	.6481	.8949	.7765	.1503	.0626
$n = 500$	med	-1.0185	2.0049	1.9772	3.0837	2.0548	-5.9988	.9525
	$q_{0.25}$	-1.4632	1.4632	1.5259	2.4757	1.5559	-6.0957	.9105
	$q_{0.75}$	2.4434	2.4434	2.4412	3.6533	2.6080	-5.8910	.9941
	mean	-1.0083	2.0056	2.0080	3.0047	1.9925	-5.9955	.9766
	std	.4401	.5674	.5028	.6253	.6047	.1053	.0441
$n = 1000$	med	-.9922	2.0127	2.0177	3.0044	1.9642	-5.9955	.9782
	$q_{0.25}$	-1.3015	1.6439	1.6440	2.5951	1.5913	-6.0671	.9467
	$q_{0.75}$	-.7036	2.3713	2.3512	3.4209	2.4010	-5.9294	1.0059

Table 3: Performance of QMLE with Uniform Residuals for Scenario 1 (bias corrected, Setting 1)

		$l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
$n = 200$	mean	-.3403	.6954	.4673	.1694	-.3090	1.9757	4.0018	4.9854
	std	.2852	.3002	.1855	.2023	1.3265	1.8731	1.6860	2.5538
	med	-.3356	.7039	.4848	.1751	-3.0027	1.9857	4.0265	4.8883
	$q_{0.25}$	-.5234	.4987	.3439	.0332	-3.9363	.6649	2.9022	3.1526
	$q_{0.75}$	-.1416	.8987	.5995	.3097	-2.1225	3.2871	5.1690	6.7563
$n = 500$	mean	-.3070	.7003	.4932	.1899	-3.0021	2.0117	3.9994	4.9259
	std	.1451	.1728	.0827	.1038	.7632	.9347	.9534	1.1365
	med	-.3009	.6974	.4942	.1962	-3.0021	2.0189	3.9870	4.9203
	$q_{0.25}$	-.4047	.5885	.4410	.1234	-3.4981	1.3816	3.3733	4.2215
	$q_{0.75}$	-.2091	.8166	.5532	.2604	-2.4786	2.6350	4.5876	5.7306
$n = 1000$	mean	-.3097	.6967	.4948	.1962	-3.0273	2.0389	3.9692	5.0167
	std	.1015	.1208	.0607	.0769	.5240	.7307	.6137	.7826
	med	-.3092	.6951	.4973	.1975	-3.0454	2.0270	3.9514	5.0651
	$q_{0.25}$	-.3833	.6152	.4538	.1428	-3.3810	1.5487	3.5340	4.4457
	$q_{0.75}$	-.2400	.7773	.5374	.2461	-2.6530	2.5602	4.3794	5.5571
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
$n = 200$	mean	-1.0306	1.9244	1.9326	2.9914	2.0353	-5.9910	.8781	
	std	1.0376	1.3274	1.2179	1.5805	1.3946	.2483	.0644	
	med	-.9888	1.9456	1.9478	2.9400	2.0275	-5.9943	.8763	
	$q_{0.25}$	-1.7167	1.0285	1.1579	1.9259	1.1003	-6.1571	.8319	
	$q_{0.75}$	-.3277	2.8675	2.7664	4.1391	2.9271	-5.8301	.9233	
$n = 500$	mean	-.9898	2.0003	1.9831	2.9619	1.9409	-6.0014	.9538	
	std	.5644	.7790	.6859	.9743	.8391	.1496	.0403	
	med	-1.0050	2.0359	1.9642	2.9753	1.9610	-6.0063	.9539	
	$q_{0.25}$	-1.3781	1.4977	1.5482	2.3199	1.3665	-6.1014	.9270	
	$q_{0.75}$	-.6142	2.5189	2.4179	3.6396	2.4753	-5.9040	.9813	
$n = 1000$	mean	-1.0280	2.0279	1.9935	2.9929	1.9949	-5.9974	.9758	
	std	.4156	.5659	.5645	.7012	.5596	.1122	.0282	
	med	-1.0335	2.0565	1.9905	3.0144	2.0141	-6.0006	.9761	
	$q_{0.25}$	-1.3233	1.6561	1.6056	2.5321	1.6248	-6.0694	.9561	
	$q_{0.75}$	-.7428	2.3997	2.3697	3.4702	2.3765	5.9226	.9942	

Table 4: Performance of QMLE with Gamma Residuals for Scenario 1 (bias corrected, Setting 1)

		$l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	-.3075	.6952	.4836	.1743	-2.9908	1.9894	4.0652	4.8682
	std	.2671	.3443	.1580	.2145	1.3715	1.5945	1.6821	2.0365
$n = 200$	med	-.3010	.7132	.4855	.1777	-2.9647	2.0512	4.1263	4.8639
	$q_{0.25}$	-.4800	.4855	.3786	.0391	-3.9594	.9993	2.9885	3.5777
	$q_{0.75}$	-.1141	.9156	.5868	.3207	-2.0820	3.0516	5.1341	6.2370
	mean	-.3100	.6866	.4836	.1847	-3.0138	2.0017	3.9517	5.0570
	std	.1504	.1840	.0964	.1205	.7566	.9838	.9354	1.1931
$n = 500$	med	-.3081	.6905	.4886	.1867	-3.0019	1.9717	3.9396	5.0362
	$q_{0.25}$	-.4030	.5651	.4244	.1033	-3.5148	1.3237	3.3266	4.2818
	$q_{0.75}$	-.2060	.8070	.5496	.2638	-2.5001	2.6913	4.6060	5.8178
	mean	-.3051	.6983	.4957	.1943	-3.0355	2.0437	4.0309	4.9834
	std	.0928	.1102	.0597	.0745	.4759	.6615	.5683	.7448
$n = 1000$	med	-.3043	.6990	.4971	.1943	-3.0297	2.0091	4.0128	4.9968
	$q_{0.25}$	-.3665	.6215	.4558	.1475	-3.3469	1.6036	3.6295	4.4954
	$q_{0.75}$	-.2421	.7777	.5367	.2458	-2.7246	2.5025	4.4031	5.4708
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
	mean	-1.0529	1.9574	1.9239	2.9911	1.9878	-5.9944	.8693	
	std	1.0957	1.2094	1.4193	1.6482	1.4973	.2728	.1327	
$n = 200$	med	-1.0202	1.9566	1.9201	3.0489	1.9423	-5.9969	.8592	
	$q_{0.25}$	-1.7757	1.1504	.9904	1.9369	.9936	-6.1744	.7761	
	$q_{0.75}$	-.3524	2.8310	2.8918	4.1072	2.9933	-5.8161	.9512	
	mean	-1.0476	1.9735	1.9457	3.0069	2.0142	-6.0037	.9520	
	std	.6346	.8030	.7962	1.0089	.8651	.1534	.0952	
$n = 500$	med	-1.0349	1.9682	1.9087	3.0177	2.0034	-6.0014	.9520	
	$q_{0.25}$	-1.4838	1.4230	1.3897	2.3471	1.4522	-6.1086	.8854	
	$q_{0.75}$	-.6007	2.5177	2.4839	3.6681	2.5790	-5.8948	1.0124	
	mean	-1.0144	1.9875	1.9765	3.0149	2.0126	-5.9957	.9804	
	std	.4233	.5106	.4627	.6549	.5832	.1090	.0688	
$n = 1000$	med	-1.0022	1.9744	1.9739	3.0470	2.0119	-5.9987	.9786	
	$q_{0.25}$	-1.2997	1.6381	1.6769	2.5719	1.6270	-6.0701	.9314	
	$q_{0.75}$	-.7287	2.3616	2.2893	2.4217	2.4217	-5.9201	1.0270	

Table 5: Performance of QMLE with Normal Residuals for Scenario 1 (bias corrected, Setting 2)

		$l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	.6055	.1933	-.2360	-.4010	1.0165	-1.9192	1.9525	3.1740
	std	.1971	.2664	.2710	.3605	1.6550	2.4120	2.2392	3.5716
$n = 200$	med	.6207	.2048	-.2334	-.4071	1.1091	-1.7020	2.0673	3.1570
	$q_{0.25}$	.4855	.0373	-.4178	-.6370	-.0923	-3.5573	.5263	.7570
	$q_{0.75}$	.7472	.3832	-.0466	-.1642	2.1371	-.3455	3.4536	5.5290
	mean	.5987	.1997	-.2877	-.4613	1.0004	-1.9885	1.9989	2.9464
	std	.1036	.1332	.1609	.2197	.9191	1.2391	1.1534	1.6120
$n = 500$	med	.6051	.2024	-.2895	-.4596	1.0067	-2.0598	2.0111	2.9712
	$q_{0.25}$	.5344	.1199	-.3971	-.6018	.3644	-2.8669	1.1939	1.9045
	$q_{0.75}$	.6654	.2849	-.1846	-.3120	1.6551	-1.1459	2.7746	3.9288
	mean	.5975	.2018	-.2871	-.4971	.9630	-1.9671	1.9932	3.0039
	std	.0731	.0915	.1107	.1392	.6612	.9867	.7927	1.2081
$n = 1000$	med	.6048	.2035	-.2892	-.5006	1.0005	-2.0233	2.0048	3.0036
	$q_{0.25}$	.5503	.1421	-.3610	-.5878	.5260	-2.6514	1.4266	2.1789
	$q_{0.75}$	.6445	.2654	-.2130	-.4055	1.4271	-1.3380	2.4949	3.8154

		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2
	mean	2.1339	-2.1068	1.3460	2.6500	-.9285	3.0241	3.5180
	std	1.6491	2.3739	2.0216	2.1542	4.0403	.5554	.3756
$n = 200$	med	2.1393	-2.0926	1.3579	2.5100	-.8684	3.0217	3.4931
	$q_{0.25}$	1.0348	-3.7728	.0100	.5121	-3.8310	2.6211	3.2570
	$q_{0.75}$	3.2257	-.4490	2.6533	4.8468	1.7462	3.4169	3.7548
	mean	1.9953	-2.0778	1.0887	2.9052	-.8939	3.0078	3.8014
	std	.8764	1.0856	1.2502	1.4767	2.1761	.3191	.2439
$n = 500$	med	1.9904	-2.0645	1.1071	2.9271	-.7851	3.0106	3.8116
	$q_{0.25}$	1.4083	-2.7925	.2590	1.9178	-2.2734	2.7809	3.6370
	$q_{0.75}$	2.6191	-1.3324	1.9467	3.8772	.4917	3.2277	3.9652
	mean	2.0479	-2.0800	.9966	3.0007	-1.0019	3.0245	3.8998
	std	.5926	1.0020	.7977	1.2091	1.4969	.2182	.1749
$n = 1000$	med	2.0567	-2.1162	1.0231	3.0047	-1.0221	3.0181	3.8893
	$q_{0.25}$	1.6522	-2.7632	.4297	2.1531	-1.9770	2.8800	3.7825
	$q_{0.75}$	2.4511	-1.3985	1.5245	2.7923	.0528	3.1725	4.0246

Table 6: Performance of QMLE with Uniform Residuals for Scenario 1 (bias corrected, Setting 2)

		$l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	.5940	.2023	-.2774	-.4422	.9060	-1.9537	1.9988	3.0043
	std	.1919	.2390	.2753	.3090	1.4634	2.4911	2.0735	3.1673
$n = 200$	med	.5991	.2126	-.2798	-.4282	.9808	-1.9174	2.0583	2.9532
	$q_{0.25}$	.4623	.0490	-.4616	-.6539	-.0158	-3.6430	.5194	.7385
	$q_{0.75}$	.7291	.3706	-.0878	-.2323	1.9149	-.4283	3.4797	5.0076
	mean	.6001	.1975	-.2749	-.4602	1.0107	-2.0010	1.9875	2.9693
	std	.1072	.1352	.1625	.2091	.9505	1.3254	.9751	1.6658
$n = 500$	med	.6059	.2052	-.2763	-.4743	1.0436	-1.9952	1.9999	2.8898
	$q_{0.25}$	.5251	.1117	-.3830	-.6071	.3560	-2.9189	1.3826	1.8382
	$q_{0.75}$	.6799	.2896	-.1662	-.3275	1.6529	-1.1046	2.7017	4.1258
	mean	.5983	.2024	-.2936	-.4783	1.0191	-1.9360	2.0025	2.9605
	std	.0730	.1000	.1103	.1446	.6952	.9972	.8332	1.2086
$n = 1000$	med	.6051	.2067	-.2913	-.4833	1.0289	-1.9486	1.9684	2.9406
	$q_{0.25}$	.5515	.1387	-.3700	-.5753	.5471	-2.6199	1.4337	2.1042
	$q_{0.75}$	.6498	.2769	-.2165	-.3788	1.4931	-1.2480	2.5832	3.8171
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
	mean	2.0697	-1.9757	1.0925	2.8696	-.9115	3.0058	3.5236	
	std	1.3153	2.6258	1.5687	2.9268	3.2401	.4979	.2575	
$n = 200$	med	2.1139	-1.9821	1.1070	2.8259	-.9511	3.0173	3.5149	
	$q_{0.25}$	1.2128	-3.7532	.0063	.9085	-3.1388	2.6661	3.3481	
	$q_{0.75}$	2.9717	-.2329	2.1601	4.8785	1.2507	3.3469	3.7068	
	mean	2.0936	-2.0294	1.0561	2.9067	-1.0915	2.9924	3.8100	
	std	.9019	1.5227	.8873	1.8312	1.9261	.3160	.1647	
$n = 500$	med	2.0960	-2.0051	1.0769	2.8952	-1.0802	2.9883	3.8165	
	$q_{0.25}$	1.4939	-3.0585	.4490	1.6711	-2.4025	2.7856	3.6957	
	$q_{0.75}$	2.7054	-1.0334	1.6542	4.0821	.1046	3.2046	3.9191	
	mean	1.9978	-2.0195	1.0353	2.9057	-1.0271	3.0063	3.9004	
	std	.6096	.8817	.8134	1.1170	1.5558	.2186	.1210	
$n = 1000$	med	1.9926	-2.0505	1.0517	2.8754	-1.0318	3.0140	3.9010	
	$q_{0.25}$	1.5936	-2.6235	.5042	2.0953	-2.0377	2.8558	3.8181	
	$q_{0.75}$	2.4007	-1.4208	1.6041	3.6695	-.0076	3.1603	3.9807	

Table 7: Performance of QMLE with Gamma Residuals for Scenario 1 (bias corrected, Setting 2)

		$l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	.5910	.2063	-.2551	-.4463	.9686	-1.9384	1.9506	2.7567
	std	.1851	.2359	.2723	.3561	1.4045	2.2364	1.8998	2.7472
$n = 200$	med	.6053	.2205	-.2449	-.4348	1.0151	-2.0041	1.9439	2.5536
	$q_{0.25}$	.4796	.0579	-.4416	-.6873	.0616	-3.4500	.7567	.8920
	$q_{0.75}$	.7169	.3688	-.0696	-.2012	1.9320	-.4094	3.1793	4.5093
	mean	.5992	.2098	-.2810	-.4638	1.0250	-1.9725	2.0355	2.9699
	std	.1056	.1389	.1627	.1921	.9057	1.5080	1.1406	1.7682
$n = 500$	med	.6062	.2172	-.2870	-.4579	1.0302	-1.9421	2.0559	2.9248
	$q_{0.25}$	.5296	.1201	-.3873	-.5937	.4023	-2.9942	1.2420	1.7857
	$q_{0.75}$	.6730	.3027	-.1692	-.3324	1.6549	-.9227	2.8270	4.1612
	mean	.5980	.2086	-.2935	-.4873	1.0134	-1.9423	2.0446	2.9641
	std	.0671	.0931	.1065	.1365	.6477	.9949	.8373	1.1726
$n = 1000$	med	.6016	.2097	-.2898	-.4873	1.0239	-1.8973	2.0674	2.9955
	$q_{0.25}$	.5566	.1490	-.3647	-.5856	.5872	-2.6284	1.4925	2.1381
	$q_{0.75}$	.6441	.2720	-.2190	-.3979	1.4454	-1.2744	2.6090	3.7976
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
	mean	2.0378	-2.2064	1.1276	2.8142	-1.0054	3.0184	3.5441	
	std	1.6758	2.2132	1.8930	2.4518	4.1456	.4986	.5454	
$n = 200$	med	2.0250	-2.3017	1.1220	2.7623	-.9857	3.0098	3.4792	
	$q_{0.25}$	.9669	-3.6629	-.1304	1.2000	-3.8386	2.6743	3.1895	
	$q_{0.75}$	3.1943	-.7494	2.4922	4.3763	1.6749	3.3660	3.8594	
	mean	2.0571	-2.0698	1.1260	2.9358	-.8917	3.0113	3.7960	
	std	.9223	1.5855	1.0089	1.8054	2.4528	.3194	.3821	
$n = 500$	med	2.0551	-2.0770	1.1674	2.9192	-.8977	3.0199	3.7764	
	$q_{0.25}$	1.4233	-3.1660	.4609	1.7033	-2.5276	2.7916	3.5149	
	$q_{0.75}$	2.7189	-.9551	1.8025	4.1032	.6982	3.2219	4.0336	
	mean	2.0450	-2.0383	1.0161	2.9344	-1.0167	3.0192	3.8805	
	std	.5758	.9023	.7290	1.1407	1.4420	.2207	.2561	
$n = 1000$	med	2.0635	-2.0375	1.0138	2.9162	-1.0762	3.0264	3.8845	
	$q_{0.25}$	1.6454	-2.6233	.5537	2.1555	-1.9912	2.8743	3.7105	
	$q_{0.75}$	2.4192	-1.4539	1.5216	3.6758	-.0044	3.1690	4.0629	

4.3 Simulation Results for Scenario 2

Table 8-14 show the performance for Scenario 2 when the membership of each group is fixed. Table 8 reports both the raw estimates and bias corrected estimates for peer effects, which are directly estimated by the concentrated likelihood approach. Unlike the results showed in Table 1 for Scenario 1, the performance of our bias correction method seems to depend on the value of true parameters and the networks. For Setting 1 with more dense networks, the bias corrected QMLE dominate the raw estimates on average bias level for all the peer effects. However, for Setting 2 with relatively more sparse networks, although the performance of the bias corrected estimators improves as sample size gets larger, for $\lambda_{2,1}$ and $\lambda_{2,2}$, the performance does not dominate the raw estimates. The raw estimates overestimate the peer effects through $W_{2,n}$ in general and the bias correction method tends to over shoot and results in underestimated QMLE. But in general, when sample size is large enough ($n \geq 1000$), the average bias level is less than 10% which is acceptable.

Table 9-14 report the performance for all the parameters except group fixed effects. The performance for the mean, median and standard deviations are similar to those under Scenario 1. Average biases for $\lambda_{2,1}$, $\lambda_{2,2}$ and σ^2 are slightly larger for Setting 2, but are still acceptable. For similar reasons, we suggest to use our method when one has a relatively large sample size in order to make more precise inference for contextual effects.

Table 8: Comparison between Raw and Bias Corrected QMLE (Scenario 2)

		Setting 1: $l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)^{'}$											
		Normal				Uniform				Gamma			
mean of estimator		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$
$n = 200$	raw	-.3557	.6371	.4686	.1654	-.3296	.6534	.4753	.1531	-.3430	.6373	.4602	.1697
	bias corrected	-.3151	.6843	.4827	.1710	-.3254	.6879	.4817	.1628	-.3174	.6647	.4718	.1829
$n = 500$	raw	−.3189	.6825	.4752	.1802	-.3224	.6777	.4821	.1841	-.3203	.6870	.4862	.1903
	bias corrected	-.3006	.6981	.4885	.1925	-.3074	.6933	.4957	.1921	-.3077	.6950	.4899	.1934
$n = 1000$	raw	-.3093	.6910	.4931	.1859	-.3131	.6959	.4944	.1937	-.3128	.6873	.4967	.1943
	bias corrected	-.2966	.6936	.4960	.1908	-.3059	.7038	.4981	.1969	-.3011	.6988	.4982	.1985
		Setting 2: $l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)^{'}$											
		Normal				Uniform				Gamma			
mean of estimator		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$
$n = 200$	raw	.5435	.1573	-.3586	-.5346	.5452	.1331	-.3646	-.5701	.5407	.1313	-.4111	-.5531
	bias corrected	.5839	.2014	-.2626	-.4531	.5918	.1901	-.2656	-.4581	.5887	.2031	-.3068	-.4000
$n = 500$	raw	.5738	.1825	-.3286	-.5262	.5810	.1750	-.3283	-.5260	.5659	.1819	-.3218	-.5308
	bias corrected	.5993	.2079	-.2687	-.4692	.6055	.2049	-.2699	-.4568	.5919	.2065	-.2614	-.4703
$n = 1000$	raw	.5856	.1889	-.3183	-.5099	.5892	.1908	-.3112	-.5126	.5898	.1915	-.3179	-.5146
	bias corrected	.6027	.2087	-.2780	-.4767	.6040	.2070	-.2754	-.4761	.6044	.2086	-.2799	-.4739

Table 9: Performance of QMLE with Normal Residuals for Scenario 2 (bias corrected, Setting 1)

$l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)'$

		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	-.3151	.6843	.4827	.1710	-3.0693	2.1132	3.9907	5.0279
	std	.2039	.2257	.1387	.1714	1.4607	1.7738	1.3169	1.6850
$n = 200$	med	-.3080	.6845	.4925	.1761	-3.0318	2.0239	4.0070	4.9817
	$q_{0.25}$	-.4440	.5346	.3902	.0591	-4.0569	.7870	3.0893	3.9415
	$q_{0.75}$	-.1739	.8410	.5746	.2909	-2.0886	3.3251	4.8765	6.1599
	mean	-.3006	.6981	.4885	.1925	-2.9721	2.0032	3.9948	5.0405
	std	.1521	.1520	.0972	.1114	.7803	1.0120	.9241	1.2421
$n = 500$	med	-.3014	.7004	.4890	.1941	-2.9727	2.0033	4.0262	5.0311
	$q_{0.25}$	-.3990	.5982	.4252	.1207	-3.4898	1.3241	3.3486	4.1477
	$q_{0.75}$	-.1973	.7964	.5544	.2664	-2.4734	2.7361	4.6130	5.9357
	mean	-.2966	.6936	.4960	.1908	-2.9779	2.0110	3.9447	5.0328
	std	.1052	.1327	.0676	.0884	.5724	.7342	.7154	.9152
$n = 1000$	med	-.2947	.6956	.4998	.1911	-2.9557	2.0294	3.9571	5.0022
	$q_{0.25}$	-.3682	.6054	.4550	.1301	-3.3517	1.5196	3.4627	4.4230
	$q_{0.75}$	-.2275	.7846	.5402	.2507	-2.5899	2.4837	4.4090	5.6288

		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2
	mean	-1.0615	1.9640	1.9956	2.9940	2.0229	-6.0010	.8820
	std	.8970	1.2285	.8485	1.5082	1.3464	.2648	.0966
$n = 200$	med	-1.0520	1.9146	2.0095	2.9213	2.0457	-5.9976	.8782
	$q_{0.25}$	-1.6526	1.1060	1.4127	1.8896	1.0685	-6.1794	.8114
	$q_{0.75}$	-.4245	2.7959	2.5833	4.0192	2.9467	-5.8204	.9466
	mean	-1.0095	1.9882	1.9614	3.0118	2.0001	-5.9954	.9207
	std	.6341	.8600	.7712	.9365	.8246	.1641	.0588
$n = 500$	med	-.9945	1.9868	1.9703	3.0565	1.9744	-5.9920	.9202
	$q_{0.25}$	-1.4225	1.4287	1.4414	2.3867	1.4553	-6.1071	.8767
	$q_{0.75}$	-.5814	2.5940	2.4846	3.6351	2.5571	-5.8879	.9617
	mean	-1.0167	2.0156	1.9820	2.9677	1.9499	-5.9904	.9353
	std	.3982	.5409	.5117	.7276	.5803	.1121	.0445
$n = 1000$	med	-1.0056	2.0174	2.0115	2.9729	1.9580	-5.9873	.9341
	$q_{0.25}$	-1.2836	1.6293	1.6502	2.4512	1.5572	-6.0650	.9062
	$q_{0.75}$	-.7359	2.3630	2.3221	3.4409	2.3424	-5.9130	.9638

Table 10: Performance of QMLE with Uniform Residuals for Scenario 2 (bias corrected, Setting 1)

		$l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
$n = 200$	mean	-.3254	.6879	.4817	.1628	-3.0851	2.1024	4.0201	4.9982
	std	.1816	.2617	.1696	.2345	1.2399	1.5377	1.4947	2.1480
	med	-.3261	.6897	.4903	.1747	-3.1151	2.1279	4.0263	5.0040
	$q_{0.25}$	-.4480	.5125	.3752	.0043	-3.9277	1.0068	2.9560	3.5721
	$q_{0.75}$	-.1961	.8711	.5987	.3292	-2.2307	3.1203	5.0106	6.4308
$n = 500$	mean	-.3074	.6933	.4957	.1921	-2.9893	1.9857	3.9987	5.0566
	std	.1351	.1655	.0872	.1108	.9008	1.1180	1.0129	1.4476
	med	-.3011	.6953	.4970	.1911	-2.9710	2.0070	4.0078	5.0560
	$q_{0.25}$	-.3952	.5797	.4370	.1163	-3.6123	1.2297	3.3149	4.1449
	$q_{0.75}$	-.2160	.8054	.5589	.2680	-2.3574	2.7627	4.6947	5.9601
$n = 1000$	mean	-.3059	.7038	.4981	.1969	-3.0033	1.9945	3.9803	4.9875
	std	.0866	.1120	.0622	.0723	.5542	.7569	.7392	.9681
	med	-.3041	.7068	.4990	.1993	-2.9663	1.9888	3.9637	5.0005
	$q_{0.25}$	-.3654	.6288	.4578	.1500	-3.3726	1.4564	3.4691	4.3098
	$q_{0.75}$	-.2471	.7786	.5398	.2472	-2.6232	2.4851	4.4745	5.6174
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
$n = 200$	mean	-1.0511	2.0617	1.9561	2.9572	1.9508	-6.0059	.8802	
	std	1.0013	1.2686	1.1349	1.7722	1.3544	.2563	.0662	
	med	-1.0399	2.0570	1.9570	2.9776	1.9225	-5.9997	.8810	
	$q_{0.25}$	-1.7364	1.2321	1.2038	1.8087	1.0343	-6.1890	.8351	
	$q_{0.75}$	-.3688	2.8798	2.7451	4.0525	2.9005	-5.8296	.9246	
$n = 500$	mean	-1.0361	1.9941	1.9730	3.0287	2.0419	-6.0033	.9215	
	std	.5616	.7601	.7396	.9239	.8322	.1661	.0399	
	med	-1.0397	1.9789	1.9737	3.0423	2.0987	-6.0000	.9209	
	$q_{0.25}$	-1.4161	1.4722	1.4701	2.4025	1.4848	-6.1163	.8940	
	$q_{0.75}$	-.6607	2.5195	2.4519	3.6562	2.6204	-5.8976	.9499	
$n = 1000$	mean	-1.0081	1.9820	1.9861	2.9890	1.9968	-6.0046	.9352	
	std	.4712	.5695	.5138	.7095	.5848	.1099	.0285	
	med	-.9993	1.9677	2.0107	2.9860	1.9873	-6.0079	.9352	
	$q_{0.25}$	-1.3424	1.5973	1.6465	2.4930	1.5932	-6.0801	.9150	
	$q_{0.75}$	-.6710	2.3796	2.3280	3.4916	2.4194	5.9321	.9547	

Table 11: Performance of QMLE with Gamma Residuals for Scenario 2 (bias corrected, Setting 1)

		$l_1 = 30, l_2 = 20, \theta_0 = (-.3, .7, .5, .2, -3, 2, 4, 5, -1, 2, 2, 3, 2, -6, 1)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
$n = 200$	mean	-.3174	.6647	.4718	.1829	-3.0175	2.0976	4.0353	4.9178
	std	.2532	.2775	.1530	.1837	1.4360	2.0822	1.9943	2.6931
	med	-.3153	.6682	.4757	.1825	-2.9648	2.0860	3.9564	4.9086
	$q_{0.25}$	-.4761	.4883	.3699	.0628	-3.9524	.6993	2.6779	3.0477
	$q_{0.75}$	-.1478	.8475	.5805	.3023	-2.0840	3.5060	5.3434	6.7846
$n = 500$	mean	-.3077	.6950	.4899	.1934	-2.9904	1.9898	3.9853	4.9902
	std	.1268	.1481	.0847	.1127	.6133	.8885	.7666	1.0967
	med	-.3055	.6946	.4923	.1957	-2.9779	1.9846	3.9741	4.9906
	$q_{0.25}$	-.3857	.5924	.4363	.1242	-3.4228	1.3839	3.4335	4.2622
	$q_{0.75}$	-.2260	.7996	.5448	.2633	-2.5701	2.6118	4.5514	5.7389
$n = 1000$	mean	-.3011	.6988	.4982	.1985	-3.0243	2.0235	4.0153	4.9751
	std	.0933	.1278	.0614	.0836	.5596	.7279	.7391	1.0403
	med	-.2957	.7057	.5004	.1992	-3.0140	2.0153	4.0321	4.9961
	$q_{0.25}$	-.3642	.6051	.4562	.1430	-3.4106	1.5079	3.5191	4.2496
	$q_{0.75}$	-.2416	.7899	.5407	.2567	-2.6702	2.5181	4.5318	5.7034
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
$n = 200$	mean	-1.0579	2.0747	1.9862	2.9419	1.9294	-5.9880	.8831	
	std	1.0629	1.3194	1.3342	1.5989	1.3943	.2674	.1366	
	med	-.9958	2.0886	1.9771	2.9480	1.9469	-5.9906	.8691	
	$q_{0.25}$	-1.7923	1.2108	1.0705	1.9000	1.0439	-6.1648	.7862	
	$q_{0.75}$	-.2862	2.9776	2.8443	4.0161	2.9098	-5.8078	.9683	
$n = 500$	mean	-1.0354	2.0157	1.9515	2.9908	1.8734	-5.9960	.9204	
	std	.5781	.9326	.7873	1.0959	.8894	.1583	.0882	
	med	-1.0386	2.0542	1.9444	3.0023	1.9496	-5.9980	.9177	
	$q_{0.25}$	-1.4274	1.3648	1.4123	2.2204	1.3456	-6.1061	.8584	
	$q_{0.75}$	-.6614	2.6789	2.5416	3.7369	2.5787	-5.8885	.9752	
$n = 1000$	mean	-.9985	1.9933	2.0047	3.0066	2.0078	-6.0009	.9356	
	std	.4419	.5830	.5445	.7104	.6067	.1111	.0659	
	med	-.9724	2.0060	1.9920	2.9843	2.0255	-6.0050	.9349	
	$q_{0.25}$	-1.2928	1.5890	1.6288	2.5405	1.5967	-6.0726	.8878	
	$q_{0.75}$	-.6818	2.3818	2.3848	3.4964	2.3909	-5.9261	.9769	

Table 12: Performance of QMLE with Normal Residuals for Scenario 2 (bias corrected, Setting 2)

		$l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	.5839	.2014	-.2626	-.4531	.9176	-1.8879	2.0046	3.0055
	std	.1649	.2112	.2441	.2833	1.4518	2.1314	1.7818	2.5138
$n = 200$	med	.5956	.2155	-.2525	-.4526	1.0098	-2.0374	1.9667	2.9081
	$q_{0.25}$	.4813	.0736	-.4183	-.6467	-.0316	-3.2992	.9077	1.2318
	$q_{0.75}$	.6973	.3483	-.0997	-.2486	1.9439	-.4596	3.2523	4.6239
	mean	.5993	.2079	-.2687	-.4692	.9915	-1.9532	2.0166	2.9959
	std	.0950	.1471	.1493	.2120	.9280	1.3854	1.3247	2.1520
$n = 500$	med	.6048	.2146	-.2702	-.4721	.9989	-2.0168	3.0257	2.1554
	$q_{0.25}$	.5389	.1052	-.3690	-.6115	.4024	-2.9384	1.1157	1.5734
	$q_{0.75}$	.6666	.3133	-.1682	-.3225	1.5992	-1.0275	2.9226	4.4140
	mean	.6027	.2087	-.2780	-.4767	.9740	-2.0167	1.9968	2.9928
	std	.0782	.1037	.1122	.1386	.6785	.9684	.7681	1.2115
$n = 1000$	med	.6060	.2096	-.2829	-.4788	.9718	-2.0145	2.0075	2.9608
	$q_{0.25}$	.5481	.1375	-.3562	-.5695	.5125	-2.7283	1.5058	2.1495
	$q_{0.75}$	.6576	.2782	-.2014	-.3588	1.4376	-1.3506	2.5100	3.8102
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
	mean	2.1086	-2.0787	1.1773	2.8766	-.9089	2.9814	3.5341	
	std	1.3879	2.2844	1.9446	2.5703	3.1769	.4780	.3794	
$n = 200$	med	2.1809	-2.1548	1.1897	2.8986	-.7939	2.9691	3.5167	
	$q_{0.25}$	1.2124	-3.6398	-.1450	1.1396	-3.0565	2.6807	3.2615	
	$q_{0.75}$	3.0897	-.5616	2.4986	4.5463	1.1900	3.3078	3.7708	
	mean	2.1264	-2.0933	1.0767	2.9758	-.9903	2.9999	3.6865	
	std	.8471	1.3049	1.2596	1.8329	2.5311	.3243	.2488	
$n = 500$	med	2.1554	-2.0905	1.0866	2.9797	-.9170	3.0006	3.6760	
	$q_{0.25}$	1.5588	-2.9378	.2145	1.6858	-2.6178	2.7849	3.5237	
	$q_{0.75}$	2.6735	-1.2532	1.9409	4.2828	.6884	3.2282	3.8502	
	mean	2.0817	-2.0581	1.0909	2.9012	-.9421	3.0189	3.7392	
	std	.6732	.9602	.6879	1.1371	1.6235	.2333	.1728	
$n = 1000$	med	2.0581	-2.0712	1.0823	2.9266	-.9985	3.0175	3.7380	
	$q_{0.25}$	1.6463	-2.7488	.6160	2.1385	-1.9822	2.8619	3.6284	
	$q_{0.75}$	2.5329	-1.3980	1.5791	3.6783	.1520	3.1743	3.8574	

Table 13: Performance of QMLE with Uniform Residuals for Scenario 2 (bias corrected, Setting 2)

		$l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
	mean	.5918	.1901	-.2656	-.4581	.9096	-1.8560	1.9252	2.9467
	std	.1939	.2203	.2493	.3018	2.0125	2.0426	1.5648	2.8164
$n = 200$	med	.6138	.2053	-.2659	-.4545	1.0069	-1.8783	1.9855	2.8763
	$q_{0.25}$	.4723	.0485	-.4241	-.6635	-.4099	-3.2707	.8723	1.0080
	$q_{0.75}$	.7245	.3452	-.0987	-.2619	2.2302	-.4492	3.0266	4.8484
	mean	.6055	.2049	-.2699	-.4569	1.0162	-2.0153	2.0363	2.9826
	std	.1041	.1353	.1554	.2013	1.0225	1.2521	1.0534	1.7014
$n = 500$	med	.6145	.2113	-.2658	-.4544	1.0513	-2.0843	2.0610	2.9821
	$q_{0.25}$	.5343	.1186	-.3752	-.5920	.3305	-2.9219	1.3229	1.7808
	$q_{0.75}$	.6783	.2980	-.1693	-.3255	1.7033	-1.1488	2.7800	4.1709
	mean	.6040	.2070	-.2754	-.4761	.9980	-1.9462	2.0113	2.9882
	std	.0732	.0886	.1121	.1335	.6680	.9598	.8353	1.1072
$n = 1000$	med	.6048	.2117	-.2725	-.4759	1.0286	-1.9664	2.0289	2.9668
	$q_{0.25}$	.5581	.1480	-.3477	-.5605	.5744	-2.5604	1.5063	2.2559
	$q_{0.75}$	.6573	.2680	-.2077	-.3890	1.4828	-1.3150	2.5600	3.7620
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
	mean	2.0122	-2.0884	1.0700	2.9585	-1.0291	3.0012	3.5318	
	std	1.3354	2.1269	1.8286	2.7592	3.3595	.5628	.2630	
$n = 200$	med	2.0087	-2.0899	1.1152	3.0405	-1.1427	3.0164	3.5375	
	$q_{0.25}$	1.1115	-3.4610	-.1045	1.0232	-3.3260	2.6356	3.3585	
	$q_{0.75}$	2.9521	-.6803	2.2514	4.8781	1.2655	3.3684	3.7032	
	mean	2.0577	-2.0978	1.1208	3.0365	-.8452	2.9978	3.6885	
	std	1.0072	1.1759	1.2495	1.4664	2.1465	.3324	.1672	
$n = 500$	med	2.0642	-2.0934	1.1626	3.0018	-.8014	3.0059	3.6856	
	$q_{0.25}$	1.3344	-2.8947	.2730	2.0508	-2.2436	2.7716	3.5733	
	$q_{0.75}$	2.7503	-1.3022	1.9953	4.0651	.6278	3.2191	3.8033	
	mean	2.0564	-2.0158	1.0569	2.9611	-1.0403	3.0046	3.7410	
	std	.6232	1.0171	.7600	1.0586	1.4719	.2108	.1171	
$n = 1000$	med	2.0835	-2.0045	1.0604	2.9908	-.9799	3.0009	3.7418	
	$q_{0.25}$	1.6432	-2.7013	.5558	2.2173	-1.9985	2.8653	3.6597	
	$q_{0.75}$	2.5136	-1.3223	1.5588	3.6685	-.0858	3.1488	3.8242	

Table 14: Performance of QMLE with Gamma Residuals for Scenario 2 (bias corrected, Setting 2)

		$l_1 = l_2 = 10, \theta_0 = (.6, .2, -.3, -.5, 1, -2, 2, 3, 2, -2, 1, 3, -1, 3, 4)'$							
		$\lambda_{1,1}$	$\lambda_{1,2}$	$\lambda_{2,1}$	$\lambda_{2,2}$	$\gamma_{1,1,1}$	$\gamma_{1,1,2}$	$\gamma_{1,2,1}$	$\gamma_{1,2,2}$
$n = 200$	mean	.5887	.2031	-.3068	-.4000	1.0026	-2.0184	2.0154	3.1025
	std	.1962	.2450	.2846	.3591	1.6946	2.6374	2.1178	2.9387
	med	.5991	.2146	-.3013	-.3930	1.1019	-2.0195	2.0777	3.0808
	$q_{0.25}$	.4669	.0484	-.4923	-.6273	-.1585	-3.7195	.6858	1.1298
	$q_{0.75}$	.7281	.3724	-.1073	-.1652	2.2410	-.1708	3.4678	5.0518
$n = 500$	mean	.5919	.2065	-.2614	-.4703	.9554	-1.9856	2.0380	3.0512
	std	.1054	.1409	.1601	.2111	1.0111	1.5751	1.1938	1.9597
	med	.5971	.2137	-.2544	-.4666	.9788	-1.9977	1.9983	3.0710
	$q_{0.25}$	.5258	.1116	-.3636	-.6164	.3148	-3.0373	1.2601	1.6701
	$q_{0.75}$	.6651	.3068	-.1573	-.3260	1.6291	-.9459	2.8206	4.3658
$n = 1000$	mean	.6044	.2086	-.2799	-.4739	1.0038	-2.0431	2.0561	2.9433
	std	.0701	.0987	.1083	.1353	.6880	1.0303	.8973	1.2742
	med	.6089	.2127	-.2765	-.4744	1.0245	-2.0136	2.0900	2.9159
	$q_{0.25}$	.5600	.1443	-.3553	-.5636	.5625	-2.7358	1.4623	2.1296
	$q_{0.75}$	.6531	.2775	-.2078	-.3821	1.4722	-1.3509	2.7052	3.7960
		$\gamma_{2,1,1}$	$\gamma_{2,1,2}$	$\gamma_{2,2,1}$	$\gamma_{2,2,2}$	β_1	β_2	σ^2	
$n = 200$	mean	2.0525	-2.0310	1.2970	2.6537	-.6421	3.0251	3.5102	
	std	1.6251	2.5823	1.7897	3.3350	3.4801	.5223	.5529	
	med	2.0889	-2.1437	1.4184	2.5535	-.5428	3.0395	3.4769	
	$q_{0.25}$	1.0064	-3.7498	.1990	.2913	-3.0189	2.6804	3.1180	
	$q_{0.75}$	3.2134	-.3665	2.5193	4.9282	1.7693	2.2746	3.8547	
$n = 500$	mean	2.1256	-2.1281	1.0653	2.8783	-.9531	3.0108	3.6955	
	std	.9334	1.4631	1.1594	1.8699	2.5994	.3453	.3627	
	med	2.1594	-2.1340	1.1200	2.8556	-.9285	3.0239	3.6757	
	$q_{0.25}$	1.4968	-3.1059	.3177	1.6368	-2.7728	2.7700	3.4488	
	$q_{0.75}$	2.7069	-1.1189	1.8640	4.0512	.7866	3.2473	3.9209	
$n = 1000$	mean	2.0552	-2.0284	1.1169	2.9000	-.9425	3.0125	3.7392	
	std	.6823	.9342	.8280	1.1654	1.6763	.2309	.2582	
	med	2.0785	-2.0161	1.1539	2.8661	-.8790	3.0050	3.7390	
	$q_{0.25}$	1.5955	-2.6586	.5672	2.0942	-2.0247	2.8503	3.5658	
	$q_{0.75}$	2.5368	-1.4423	1.6601	3.6846	.1855	3.1645	3.9023	

5 Application: Heterogeneous Peer Effects in Chinese Student Academic Achievement

5.1 Data Description

The Chinese pre-university education system generally includes 6 years in primary school (Grade 1 to 6), 3 years in junior high school (Grade 7 to 9), and 3 years in senior high school (Grade 10 to 12) before college. To understand the heterogeneous peer effects on students' academic achievement, we focus on the junior high school period, as required by the nine-year compulsory education in China, this is the phase that students must complete in order to decide whether to continue their education in senior high school. Upon the beginning of Grade 7, i.e., the first year of junior high school, students are assigned to classrooms either randomly or non-randomly[8] and then usually stay in the same class throughout the 3 years of junior high school. Students are required to take 3 core subjects – Chinese, mathematics, and English – and a set of subsidiary subjects. A head teacher, which can be one of the core-subject teachers, is responsible for students' social lives and providing feedback to students and their parents about academic performance and behavior.

We obtain data from China Education Panel Survey (CEPS), which is the first large-scale, nationwide, and longitudinal survey dataset hosted by the National Survey Research Center (NSRC) at Renmin University of China. So far, available data covers two waves in CEPS. In Wave I, CEPS surveys 19,487 students from both Grade 7 and Grade 9 in 438 classrooms of 112 schools in 28 county-level units in mainland China in the 2013-2014 academic year. The survey contains 5 different questionnaires regarding topics about the students, parents, head teacher, subject teacher, and school administrators. In Wave II, CEPS conducts a follow-up survey[9] for 10,279 students in Grade 8 in the 2014-2015 academic year, which was Grade 7 in Wave I[10]. The response rate is 91.9%. The missing observations are due to reasons such as transferring to another school, dropping out, long-term sick leave, etc. Our targets are the 9,449 students who appear in both Wave I and Wave II.

In this application, "group" refers to a grade level (Grade 8) in the same school as in Lin (2010). However, instead of considering friendship networks, we construct networks based on class assignment for two reasons. First, we are unable to identify a student's peers as her friends since the CEPS dataset only contains the number of a student's best friends and the conditions, e.g., sex, residence status, whether in the same school/class or not, of up to 5 best friends without providing their names/ID information for matching. Second, due to the fact that generally students will be in the same class during the 3-year junior high school stage, their education outcomes might be more influenced by classmates rather than by their friends. In order to mitigate endogenous network

[8] Various methods for assigning students to classrooms are implemented in China. Nonrandom assignments can be based on students' entrance exam scores, residency, etc. To ensure equal and fair opportunities for all students during their compulsory education years, randomized assignments are heavily promoted by the Ministry of Education, which can rely on computer program that incorporate desired multidimensional information or drawing lots to determine students' placement.

[9] Follow-up surveys are annual as the sample adolescents matriculate throughout the junior-high stage and in the 1st, 3rd, 7th, 8th, 17th and 27th year after they graduate from junior-high. CEPS will last more than 30 years, during which a new cohort of 7th graders will be started in a 10-year interval.

[10] There are 471 newcomers in Wave II, but we exclude them since we examine the peer effects of classmates on the relative change/progress of academic outcomes as students proceed from Grade 7 to Grade 8.

formation and potential selection problem, we focus on schools that use random assignments of students to classrooms, which are determined by the similar criteria in Gong et al. (2018)[11]. We then further drop observations with at least one missing midterm exam score for the three core subjects, and the students who choose to transfer to other classes. The remaining sample consists of 3,944 students across 97 classrooms in 56 schools. On average, there are 41 students in each class with a minimum of 14, a maximum of 74 and a standard deviation 13.02. 15 schools have only one classroom, and 41 schools have two classrooms included in the sample.

Table 15 displays our variable settings. The heterogeneous source for networks in this study is students' gender, which is excluded from students' characteristics in order to avoid multicollinearity. Outcome variables are students' academic performance, which are measured by the Fall 2014 midterm scores of Chinese, Mathematics, English, and the total score by summing up the scores of the three core subjects, as the cohort goes from Grade 7 to Grade 8 and stays in the same classroom for more than one year. The raw data of scores for the 2014-2015 academic year has been standardized for all classes within the same school with mean 70 and standard deviation 10 to have comparable results. The variables of interest are students' characteristics, such as whether they are the only child, relative age (compared with the median level), ethnicity (whether belong to the Han nationality or not), local residency status, whether they attended kindergarten, and their parents' education levels. Those pre-determined variables barely change and are also used in Gong et al. (2018) and Gong et al. (2019)[12]. Note that family income information, which might be an important factor affecting students' academic achievement, is not available in the CEPS dataset. But parents' education levels might capture most, if not all, of the effects of family income. The dataset also contains head teachers' information, such as their gender, teaching experience and whether they are one of the core subject teachers, in both 2013-2014 and 2014-2015 academic years. The 2014 midterm exams usually occur in October, since class begins in September, we choose head teachers' characteristics from 2013-2014 academic year because they might have more impact on the academic outcomes of the cohort as students have interactions with their head teachers for a longer time period. We exclude school level information, such as school quality and teacher/student ratio, due to our control of group fixed effects.

The summary statistics for the whole sample are provided in Table 16. In the whole sample, 48% of the students are female with a standard deviation 0.50. Scores for Chinese, Math and English have mean around 70 due to standardization and a standard deviation near 10. Average total score is 74.13 with higher variation. For students' characteristics, 48% are the only child in their family, 10% are minority students, 77% are local residents based on the Chinese Hukou System and 85% attended kindergarten. Although our selected sample differs from the one employed in Gong et al. (2018), the differences in those characteristics are modest, which implies sample selection bias is not

[11] The three conditions should be met: (i) the school principal reports that students are randomly assigned to classrooms; (ii) the school doesn't rearrange their classes for grade 8; (iii) all head teachers report that students are not assigned by test scores.

[12] The only difference is that they also consider whether the students skipped/repeated a grade in primary school. We find that corresponding data are quite noisy as a student might skip a grade for up to 9 times and repeat a grade for up to 7 times. Besides, as we show below, relative age is a good approximate for these two variables. Moreover, some (baseline) pre-noncognitive measures are included in their paper since they consider different outcome variables which include students' non-cognitive outcomes.

Table 15: Variable Settings

	Variables	Settings
Heterogeneous source	students' gender	"0" for "male student"; "1" for "female student"
Outcome variables	Chinese score Mathematics score English score total score	standardized for all classes within the same school with mean 70 and standard deviation 10
Student's characteristics	only child in family	"0" for "no"; "1" for "yes"
	relative age	± month, compared with sample median value (February, 2001)
	minority	"0" for "no"; "1" for "yes"
	local resident	"0" for "no"; "1" for "yes"
	attend kindergarten	"0" for "no"; "1" for "yes"
	parents' education	"1" for "none"; "2" for "finished elementary school"; "3" for "junior high school degree"; "4" for "technical secondary school or technical school degree"; "5" for "vocational high school degree"; "6" for "senior high school degree"; "7" for "junior college degree"; "8" for "bachelor's degree"; "9" for "master's degree or higher"
Head teacher's characteristics	female head teacher	"0" for "no"; "1" for "yes"
	teaching experience	year
	whether teach Chinese/ Mathematics/English	"0" for "no"; "1" for "yes"

a concern in this study. The average relative age (month) is 1.29 with a standard deviation 7.29. Parents' education level has a mean slightly above 4, which is between "technical secondary school or technical school degree" and "vocational high school degree", and standard deviations around 2. For head teachers' information, the sample covers 71% female head teachers, 29% head teachers who teach Chinese, 32% head teachers who teach Math, and 25% head teachers who teach English. The average teaching experience of head teachers is 14.39 year with a standard deviation 7.89.

Besides, to investigate the gender disparity in the variables, we further decompose our sample into female and male students subgroups and the corresponding summary statistics are presented in Table 17. For academic performance, female students outperform male students for the three core subjects and also the total score, while male students' scores have large variations than those of female students. For students' characteristics, there are more only child and higher relative age for the male students compared with those of female students, while parents' education levels are slightly higher for female students. For head teachers' characteristics, the difference for females students and males students are small in terms of sample mean and standard deviations.

Table 16: Summary Statistics for the Whole Sample

	Mean	SD	Observations
Female student	.48	.50	3,944
A. *Outcome Variables*:			
Chinese Score	70.56	9.55	3,944
Math Score	70.65	9.78	3,944
English Score	70.60	9.74	3,944
Total Score	74.13	16.84	3,944
B. *Student's characteristics*:			
Only child in family	.48	.50	3,944
Relative age (month)	1.29	7.29	3,944
Minority	.10	.30	3,931
Local resident (Hukou System)	.77	.42	3,944
Attend kindergarten	.85	.36	3,906
Father's Education	4.41	2.06	3,943
Mother's Education	4.09	2.46	3,943
C. *Head teacher' characteristics*:			
Female head teacher	.71	.45	3,944
Teaching experience of head teacher (year)	14.39	7.89	3,944
Chinese head teacher	.29	.21	3,944
Math head teacher	.32	.22	3,944
English head teacher	.25	.19	3,944

Table 17: Summary Statistics for Female and Male Students Subgroups

	Female		Male	
	Mean	SD	Mean	SD
A. Outcome Variables:				
Chinese Score	73.44	7.68	67.93	10.24
Math Score	71.43	9.35	70.00	10.11
English Score	73.17	8.75	68.22	9.95
Total Score	76.72	16.31	71.68	16.97
B. Student's characteristics:				
Only child in family	.48	.50	.55	.50
Relative age (month)	.60	7.10	1.88	7.41
Minority	.10	.30	.09	.29
Local resident (Hukou System)	.78	.41	.76	.43
Attend kindergarten	.85	.36	.84	.36
Father's Education	4.48	2.08	4.35	2.04
Mother's Education	4.11	2.03	4.09	2.06
C. Head teacher's characteristics:				
Female head teacher	.72	.45	.70	.46
Teaching experience of head teacher (year)	14.55	7.86	14.23	7.93
Chinese head teacher	.30	.46	.29	.45
Math head teacher	.32	.47	.32	.47
English head teacher	.26	.44	.25	.43
Sample Size: 3893, including 1890 females and 2003 males[13]				

5.2 Empirical Strategy

5.2.1 Specification 1: Single Network

To estimate the heterogeneous peer and contextual effects for male and female students based on the single network constructed by classmates, the following model can be used:

$$y_i = \sum_{g \in \{F,M\}} d_{g,i} \lambda_g \bar{y}_{-i,c_i} + \sum_{g \in \{F,M\}} d_{g,i} \bar{x}'_{-i,c_i} \gamma_g + x'_i \beta_1 + t'_{c_i} \beta_2 + \alpha_{s_i} + u_i \tag{30}$$

where c_i and s_i denote the class and school identity associated with individual i. y_i is individual i's academic achievement showed in Table 16 Section A, including midterm scores of Chinese, Mathematics, English and the total score. $d_{F,i}$ and $d_{M,i}$ are dummy variables for female and male students, so for each individual i, $d_{F,i} + d_{M,i} \equiv 1$. $\bar{y}_{-i,c_i}$ is the average score of individual i's classmates excluding her/him-self. x'_{i,c_i} include pre-determined characteristics of individual i showed in Table 16 Section B. $\bar{x}'_{-i,c}$ include the average characteristics of individual i's classmates excluding her/him-self. t'_{c_i} contain the characteristics of the head teacher of the class c_i, which is showed in Table 16 Section C. α_{s_i} captures the effects of common variables, either observable or more commonly unobservable, which are identical for all the students in the same grade of the same school. u_i is the residual term. Then, in this model, $\{\lambda_F, \lambda_M\}$ and $\{\gamma_F, \gamma_M\}$ capture the peer and contextual effects for students of different genders.

Denote $Y_n = (y_1, \cdots, y_n)'$, $H_F = diag\{d_{F,1}, \ldots, d_{F,n}\}$, $H_M = diag\{d_{M,1}, \ldots, d_{M,n}\}$, $X_n = (x_1, \cdots, x_n)'$, $T_n = (t_{c_1}, \cdots, t_{c_n})'$, $u_n = (u_1, \cdots, u_n)'$, and define h_s as the dummy variable for each school s, then we can rewrite the model into the following vector form:

$$Y_n = \sum_{g \in \{F,M\}} \lambda_g H_g W_n Y_n + \sum_{g \in \{F,M\}} H_g W_n X_n \gamma_g + X_n \beta_1 + T_n \beta_2 + \sum_{s=1}^{S} \alpha_s h_s + u_n \tag{31}$$

where n is the total number of observations and S is the total number of schools. h_s is the school dummy variable for school s. W_n is the row normalized $n \times n$ matrix with each element

$$w_{ij,n} = \begin{cases} 1/\left(|c_i| - 1\right) & c_i = c_j \\ 0 & c_i \neq c_j \end{cases}$$

which defines the social network in the same classroom. $|c_i|$ is the total number of students in class c_i. The parameters of interest are $\theta_0 = (\lambda_{F,0}, \lambda_{M,0}, \gamma_{F,0}, \gamma_{M,0}, \beta_{1,0}, \beta_{2,0}, \sigma_0^2)'$. By QML method developed in Section 2.3, we can estimate this model.

5.2.2 Specification 2: Multiple Networks

In the study of heterogeneous peer/contextual effects with gender being the heterogeneity source in student academic achievement, it would be more interesting to separately identify both the within-gender peer/contextual effects and the cross-gender peer/contextual effects for female students and male students respectively, and examine their heterogeneous interaction patterns. The empirical specification for this purpose can be

$$\begin{aligned} y_i = & \sum_{g \in \{F,M\}} \sum_{p \in \{F,M\}} d_{g,i} \lambda_{g,p} \bar{y}_{-i,p,c_i} + \sum_{g \in \{F,M\}} \sum_{p \in \{F,M\}} d_{g,i} \bar{x}'_{-i,p,c_i} \gamma_{g,p} \\ & + x'_i \beta_1 + t'_{c_i} \beta_2 + \alpha_{s_i} + u_i \end{aligned} \tag{32}$$

where y_i, x_i, t_{c_i} and α_{s_i} are the same as the single network setting. $d_{g,i}$ is the dummy variable for student ith gender, $d_{F,i} + d_{M,i} = 1$. $\bar{y}_{-i,p,c_i} = \begin{cases} \text{average score of female classmates} & \text{if } p = F \\ \text{average score of male classmates} & \text{if } p = M \end{cases}$ and $\bar{x}_{-i,p,c_i} = \begin{cases} \text{average characteristics of female classmates} & \text{if } p = F \\ \text{average characteristics of male classmates} & \text{if } p = M \end{cases}$. Therefore, $\{\lambda_{F,F}, \lambda_{M,M}\}$ and $\{\gamma_{F,F}, \gamma_{M,M}\}$ capture the within-gender peer and contextual effects, while $\{\lambda_{M,F}, \lambda_{F,M}\}$ and $\{\gamma_{M,F}, \gamma_{F,M}\}$ capture cross-gender peer and contextual effects.

In matrix/vector form, the model can be rewritten as

$$\begin{aligned} Y_n = &\sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} \lambda_{g,p} H_g W_{p,n} Y_n + \sum_{g\in\{F,M\}} \sum_{p\in\{F,M\}} H_g W_{p,n} X_n \gamma_{p,g} \\ &+ X_n\beta_1 + T_n\beta_2 + \sum_{s=1}^{S} \alpha_s h_s + u_n \end{aligned} \tag{33}$$

$W_{F,n}$ and $W_{M,n}$ are row normalized $n \times n$ matrix with each element

$$w_{ij,F,n} = \begin{cases} 1/\left(|F_{c_i}| - 1\right) & c_i = c_j \\ 0 & c_i \neq c_j \end{cases} \quad \text{and} \quad w_{ij,M,n} = \begin{cases} 1/\left(|M_{c_i}| - 1\right) & c_i = c_j \\ 0 & c_i \neq c_j \end{cases}$$

where $|F_{c_i}|$ is the total number of female students in class c_i, and $|M_{c_i}|$ is the total number of male students in class c_i.

5.2.3 The control of school-grade fixed effects

Here some people may wonder why we control the school-grade level instead of classroom level fixed effects. On one hand, there are some technical limitations to include smaller group level fixed effects. Since we include the characteristics of head teachers which should be identical for students in the same class, those variables are perfectly multi-collinear with classroom level fixed effects obviously. Thus, we are not able to separately identify the potential pre-determined classroom selection effect and the effect from the head teachers. Similarly, gender-classroom or gender-school level fixed effects can not be included in the model. Since they are perfectly co-linear with the female dummy, they can not be separately identified from the gender effect. Thus, the smallest unit of group which we can control for fixed effects is the school assignment.

On the other hand, it is due to our sample features. In fact, as we described in Section 5.1, for each school, the classes are randomly assigned, there is no need to control classroom level fixed effects. But at school level, since we have both public schools and private schools included, they might have different selection standards for students. Even if they did not, the different levels of tuition fees would potentially differentiate the students. Besides, different regions in China have different ways to allocate students into junior high schools. For example, in 2013 which is the year our sampling students entered their junior high schools, the capital city Beijing just allocate the students into public schools nearby their homes randomly. However, in the same year, Tianjin, which is another municipality in China close to Beijing, used application basis just like college admission, and the schools selected the students based on their performance in elementary schools, their scores in junior high school entrance exams hold by local governments, or even private tests/interviews for top schools. Thus, the school allocation in our sample is definitely non-random. Considering both technical and sample reasons, controlling the school-grade level fixed effects is our best choice.

5.3 Estimation Results

Table 18a and 18b summarize our estimation results with the single network constructed by classmates, and those with multiple networks constructed by gender subgroups in a class are reported from Table 19a to Table 19e. First, the log-likelihood values exhibit slight improvement for all the four models with multiple networks specification (Table 19e) than those with single network (Table 18b). By the *Efron's* R^2, the models using total score as the dependent variable provides the best model fit. It might arise from the fact that in China, for junior high school students, the goal of study is to maximize one's total score rather than the test score of a specific subject. Students might strategically make some trade-offs for their study time among the three score subjects. As a result, the estimation results for the total score columns might be more meaningful as it's less noisy, while the results in other columns can be used as comparisons.

Second, we consider the estimation results for peer effects (Table 18a and Table 19a-19d). Under the single network specification, the gender disparity in peer effects from classmates are modest. 1 standard deviation increase in classmates' average achievement raises a male student's total score by 6.532 points and that of a female student by 5.456 point. However, we find strong evidence for heterogeneous gender peer effects from female and male classmates as significant gender disparities are detected under the multiple networks specification. For all the subjects, female students' performances are more subject to both female and male peers' influences. The finding is consistent with some previous studies, for instance, Yakusheva et al. (2014) and Trogdon et al. (2008) find that females are subject to peer influence in weight gain. Besides, female peers' average achievement contribute more to a student's Chinese and total test scores, while for Mathematics, male peers' average achievement have more impacts. For the English subject, the impacts of female and male classmates are not significantly statistically different from each other. [14] If we focus on peer effect estimates of the total score, the within-gender effects are stronger than cross-gender effects for female students, while for male students, the opposite is true. 1 standard deviation increase in female classmates' average achievement lead to 10.838 points increment in a female student's total score, and raises a male student's total score by 8.245 points. For the 1 standard deviation increase in male classmates' average achievement, a female student's total score can be raised by 7.411 points and that of a male student is increased by 5.853. These results show that the social multiplier effects exist in Chinese junior high school learning and the magnitudes differ by heterogeneous social interaction patterns among gender subgroups.

Third, we consider the impacts of individual characteristics (Table 18b and Table 19e). The

[14] The estimated peer effect coefficients from the impacts of female classmates on females and males are $\hat{\lambda}_{F,F} = .5140$ and $\hat{\lambda}_{M,F} = .4489$, and those from male classmates on females and males are $\hat{\lambda}_{F,M} = .7130$ and $\hat{\lambda}_{M,M} = .4762$, by simple test statistics, we can not reject the null hypothesis that $H_0 : \hat{\lambda}_{F,F,0} = \hat{\lambda}_{F,M,0}$, and $\hat{\lambda}_{M,F,0} = \hat{\lambda}_{M,M,0}$.

estimates are robust under both specifications with same signs and close magnitudes. We do not find significant impacts of minority students and students who are local residents on the test scores. We detect that being the only child in the family slightly raises a student's test scores, as depicted in previous literature such as Poston and Falbo (1990), Falbo and Poston (1993), Li and Zhang (2017)[15]. We show that having attended kindergarten helps to increase students' academic achievement, which offers the evidence for the importance of early childhood education.[16] Furthermore, a student's test scores for all subjects are positively correlated with his/her parents' education levels, similar results can be found in Davis-Kean (2005) and Dickson (2016).[17]

We capture a counterintuitive result that an older student performs a little bit worse in all subjects, which seems to violate the famous "relative age effect (RAE)". However, RAE are more commonly seen in the sports field, for instance, Musch and Grondin (2001), Helsen et al. (2005), and Delorme et. al. (2010), whereas we focus on the academic outcomes. Moreover, since a student's relative age is the $\pm$ month(s) compared with the sample median value of students' date of birth (February, 2001) by our construction, due to the fact that there is a cutoff date regulating the precise age for entry into primary school in China and that our sample has excluded students who have skipped or repeated grades in the junior high school, relative age is a good approximate for whether a student has repeated or skipped grades in primary school because the data for repeated or skipped grades is quite noisy and with some missing values due to students' self-report in the questionnaire. To see this, we decompose our sample into three subgroups: the delayed range group (19.16%, might have repeated grades in primary school) that is at least 5 months older than the sample median, the regular range group (71.10%) that is at most (or exactly) 5 months older and at least (or exactly) 6 months younger than the sample median value, the earlier range group (9.74%, might have skipped grades in primary school) that is at least 6 months younger than the sample median. As in Table 20a and Table 20b, the regular range group has lower correlation coefficients with all the subjects, while the other two groups have higher negative correlations. The regular range group has test scores around the sample averages, but the delayed range group has lower average scores, and the earlier range group has the opposite outcomes. Based on these findings, we might safely conjecture that the extreme performance of the delayed range and the earlier range groups has dragged the sign of the estimated coefficients to the slightly negative side.

Fourth, for contextual effects, some variables show significant impact. Under the single network specification (Table 18a), the contextual variables that show negative effects include relative age (for both female and male students)[18], minority (for a female student's Chinese score)[19], and

[15] Poston and Falbo (1990) find that those without siblings score higher academically than those with siblings. Falbo and Poston (1993) show that onlies are more likely to outscore others in verbal tests in terms of academics. Li and Zhang (2017) provide new evidence of the causal effect of child quantity on child quality.

[16] One related finding in Chetty et al. (2011) is that kindergarten test scores are highly correlated with outcomes such as later earnings and college attendance.

[17] Parents' years of schooling was found to be an important socioeconomic factor for students' academic outcomes (Davis-Kean, 2005) and increasing parental education has a positive causal effect on children's outcomes (Dickson, 2016).

[18] Being in a class with older classmates decreases a student's Math and total scores and reduces the English score if the student is male. In other words, given a student's age ranking in the sample, being with higher percentage of older classmates puts the student at an unfavorable academic status. The finding is consistent with Bedard and Dhuey (2006), which state that youngest members of each cohort score lower than the oldest members in grade 4 and 8, although they didn't formally use "contextual effects" to describe their finding.

[19] There might be some trade-off effects for the time that a female student can spend on using minority language

attended kindergarten (for a male student)[20]. On the other hand, having classmates who are only child in family will help improve a female student's total score, and having classmates who are local resident will raise a male students' Math score, having classmates who have mothers with higher education level will improve a male student's Chinese score, which is a similar result as in Chung and Zou (2020) and Bifulco et al. (2011)[21]. Under the multiple networks specification (Table 19a-19d), relative age, local resident, attended kindergarten, and parents' education levels show some significant impacts, but whether the impacts are positive or negative, and which ones are stronger between the within-gender effects and cross-gender effects vary across different gender groups and depend on the subjects. Two contextual variables that are worth noting are relative age and mother's education. Relative age demonstrates both competitive effects and complementary effects for a female student, i.e., having older male classmates will deteriorate a female student's Chinese and Math scores, while having older female classmates will improve a female student's English and total scores. By the contextual effect of mother's education, we detect the specific channel about how higher classmates' maternal education raises a students' test score (Chung and Zou, 2020), i.e., a female student's Mathematics and total test scores are positively affected when her male classmates have higher educated mothers.

Last, the roles of head teachers' characteristics are investigated. Under the single network specification (Table 18b), we detect that having a female head teacher can raise a student's Chinese and total test score.[22] Additionally, similar with Rockoff (2004) and Ladd and Sorensen (2017)[23], we capture significant positive influences of an experienced head teacher on a student's Math score, but evidence about its impact on Chinese, English and the total scores are not found. Then, when a head teacher teaches Math or English, a student's corresponding test scores rise.[24] However, under the multiple networks specification (Table 19e), the effects of head teachers' characteristics are not significant. The impact of a head teacher on a student's academic achievement might be entangled with interaction patterns of within and across gender subgroups in the same class, but under current model design, we are unable to identify the specific channel, which might be an interesting topic for future research.

and learning Chinese when her classmates are minority.

[20] Although early childhood education benefits later cognitive outcomes, it worsens a male student' total score when surrounding peers also have this advantage.

[21] Chung and Zou (2020) find that higher classmates' maternal education raises students test score, and Bifulco et al. (2011) states that increases in the percent of classmates with college-educated mothers decreases the likelihood of dropping out and increases the likelihood of attending college.

[22] A finding that is somewhat consistent with Gong et al. (2018), which show that the gender of teacher matters, i.e., having a female teacher raises girls' test scores and improves their mental status and social acclimation relative to those of boys.

[23] Rockoff (2004) presents evidence that teaching experience significantly raises student test scores. Ladd and Sorensen (2017) find large returns to experience for middle school teachers in the form of higher test scores.

[24] The possible explanation might be that a Math or English subject teacher, who is also the head teacher, is more likely to provide positive feedback to boost students' confidence in studying and enforces students' beliefs about the importance of the corresponding subject.

Table 18a: Results for Peer and Contextual Effects (Single Network)

	Chinese		Mathematics		English		Total	
	Female	Male	Female	Male	Female	Male	Female	Male
Peer Effects	.3656***	.3726***	.3658***	.3321***	.3901***	.2923***	.3240***	.3879***
	(.0721)	(.0715)	(.0746)	(.0742)	(.0749)	(.0744)	(.0709)	(.0709)
Contextual Effects:								
Only child in family	4.7833	2.0666	4.8465	3.3977	2.2819	1.0572	5.6077*	3.6862
	(2.9539)	(2.9431)	(3.2501)	(3.2352)	(3.2440)	(3.2355)	(3.1959)	(3.1831)
Relative age	.1100	-.2556	-.7967***	-1.0025***	-.3340	-.6954***	-.4241*	-.8817***
	(.2158)	(.2123)	(.2534)	(.2496)	(.2275)	(.2237)	(.2398)	(.2363)
Minority	-14.7965*	-12.1904	-10.3601	-7.4425	-9.1319	-5.9387	-12.0838	-9.1474
	(8.1058)	(8.0748)	(8.5368)	(8.5005)	(8.2613)	(8.2194)	(8.8552)	(8.8052)
Local resident	.6041	.9761	2.9473	4.8338*	-3.7354	-1.1683	1.2820	1.9565
	(2.4794)	(2.3379)	(2.7114)	(2.5654)	(2.5465)	(2.3932)	(2.6205)	(2.5071)
Attend kindergarten	4.8920	-3.4412	-5.6608	-6.8057	-2.8181	-6.0426	-.8018	-11.3979***
	(3.6721)	(3.6713)	(3.8240)	(3.8275)	(3.7896)	(3.7965)	(3.6347)	(3.6357)
Father's Education	-.6310	-1.3140	-1.3689	.1082	-.8856	.1847	-.9275	-.8498
	(1.0761)	(1.0809)	(1.1663)	(1.1661)	(1.1272)	(1.1336)	(1.1484)	(1.1493)
Mother's Education	1.5897	2.8957**	1.2113	-.0611	1.1679	.8801	.5024	.6502
	(1.1318)	(1.1277)	(1.1467)	(1.1413)	(1.1019)	(1.0962)	(1.1573)	(1.1595)

Sample Size: 3893, including 1890 females and 2003 males
Significance Level: * <10%, **< 5%, *** <1%

Table 18b: Results for Students and Head Teachers' Characteristics (Single Network)

	Chinese	Mathematics	English	Total
Students' Characteristics:				
Only child in family	.6831**	.8476**	.5880*	.7883**
	(.3387)	(.3628)	(.3470)	(.3631)
Relative age	-.0653***	-.1377***	-.1208***	-.1411***
	(.0229)	(.0247)	(.0234)	(.0246)
Minority	-.1849	-.9118	-.3147	-.7075
	(.6976)	(.7451)	(.7118)	(.7487)
Local resident	-.1869	.0863	-.1700	-.0376
	(.3974)	(.4225)	(.4056)	(.4259)
Attend kindergarten	1.7546***	1.4655***	1.5599***	1.7002***
	(.4116)	(.4400)	(.4207)	(.4409)
Father's Education	.2002**	.2589**	.4361***	.3539***
	(.0993)	(.1063)	(.1015)	(.1064)
Mother's Education	.3528***	.1811*	.2539**	.2700**
	(.1024)	(.1091)	(.1041)	(.1094)
Head teachers' Characteristics:				
Female	1.1822**	.9150	.1980	1.0761*
	(.5978)	(.6901)	(.7080)	(.6367)
Teaching experience	.0195	.1207**	-.0004	.0577
	(.0362)	(.0408)	(.0376)	(.0397)
Teach this course	-.7817	1.5422**	2.8613***	.7996
	(.5160)	(.5447)	(.8189)	(.7755)
Log Likelihood	-13982	-14243	–14059	-14251
Efron's R^2	.1467	.0788	.1492	.6885

Table 19a: Results for Peer and Contextual Effects for Chinese (Multiple Networks)

	Female Classmates		Male Classmates	
	Female	Male	Female	Male
Peer Effects	.7734***	.5315***	.5125***	.4112***
	(.0854)	(.0666)	(.1088)	(.1008)
Contexual Effects:				
Only child in family	-.2905	-.7255	-2.4732	-3.9304
	(2.1785)	(2.2159)	(3.6477)	(3.5172)
Relative age	.1823	-.0467	-.3581*	.1224
	(.1702)	(.1675)	(.1962)	(.1901)
Minority	9.2257	-.9966	-8.7721	2.0019
	(5.8687)	(5.9554)	(6.7361)	(6.4996)
Local resident	.8898	-1.8995	-3.8662	3.3407
	(2.6673)	(2.6811)	(2.8180)	(2.6604)
Attend kindergarten	-3.0196	3.9821	-6.3429*	1.3857
	(3.6818)	(3.5969)	(3.6933)	(3.6009)
Father's Education	.1306	.1437	-1.1901	.2700
	(.8162)	(.8456)	(.8486)	(.8292)
Mother's Education	-.6086	-1.1002	.6998	.4899
	(.9664)	(1.0023)	(.9991)	(.9620)

Sample Size: 3893, including 1890 females and 2003 males
Significance Level: * <10%, **< 5%, *** <1%

Table 19b: Results for Peer and Contextual Effects for Mathematics (Multiple Networks)

	Female Classmates		Male Classmates	
	Female	Male	Female	Male
Peer Effects	.3989***	.2482***	.5652***	.4683***
	(.0730)	(.0839)	(.0988)	(.0967)
Contexual Effects:				
Only child in family	1.7972	-2.8156	-4.4454	1.5916
	(2.4011)	(2.4220)	(3.9466)	(3.7976)
Relative age	.2740	-.0730	-.4849**	.0683
	(.1806)	(.1790)	(.2123)	(.2015)
Minority	4.2831	-1.2726	-7.5487	.7558
	(6.1626)	(6.2631)	(7.1075)	(6.8799)
Local resident	4.1093	-2.2392	-5.5803*	5.1577*
	(2.8141)	(2.8532)	(3.0723)	(2.8911)
Attend kindergarten	-1.0687	3.3024	-4.7658	1.1933
	(3.9099)	(3.7246)	(3.8439)	(3.5472)
Father's Education	-.5004	1.3676	-1.4728	.2082
	(.8834)	(.9075)	(.9237)	(.8807)
Mother's Education	-.4686	-.6195	2.2530*	-.9135
	(1.0204)	(1.0587)	(1.0950)	(1.0423)

Sample Size: 3893, including 1890 females and 2003 males
Significance Level: * <10%, **< 5%, *** <1%

Table 19c: Results for Peer and Contextual Effects for English (Multiple Networks)

	Female Classmates		Male Classmates	
	Female	Male	Female	Male
Peer Effects	.6140***	.4489**	.7130***	.4762
	(.1197)	(.1781)	(.1606)	(.3031)
Contexual Effects:				
Only child in family	1.7173	-1.2912	-6.3287	-3.7488
	(3.4419)	(3.8040)	(3.9272)	(3.5656)
Relative age	.3245*	-.0599	-.2713	.2511
	(.1732)	(.1688)	(.2212)	(.2531)
Minority	6.0672	5.8218	-3.8457	-1.1949
	(6.2754)	(6.3311)	(7.2545)	(6.8532)
Local resident	3.8016	-1.8771	-8.1032**	3.7163
	(2.6681)	(2.8193)	(2.9436)	(3.0092)
Attend kindergarten	-7.5587**	5.6609	.6911	1.1286
	(3.6176)	(3.7019)	(4.5833)	(5.1422)
Father's Education	-1.1618	.5321	-1.2378	.7179
	(1.0101)	(1.0908)	(.9188)	(1.0457)
Mother's Education	-.0599	-.8382	1.2656	-.4033
	(.9700)	(1.0413)	(1.0477)	(1.1543)

Sample Size: 3893, including 1890 females and 2003 males
Significance Level: * <10%, **< 5%, *** <1%

Table 19d: Results for Peer and Contextual Effects for Total Score (Multiple Networks)

	Female Classmates		Male Classmates	
	Female	Male	Female	Male
Peer Effects	.6645***	.5055***	.4367***	.3449***
	(.0854)	(.0731)	(.1107)	(.0949)
Contexual Effects:				
Only child in family	.7100	-2.1874	-5.1136	2.3158
	(2.3602)	(2.3880)	(4.1073)	(3.9398)
Relative age	.4468**	-.2151	-.3330	.0388
	(.1872)	(.1848)	(.2096)	(.2001)
Minority	4.2213	-4.3866	-11.9775	4.4193
	(6.2655)	(6.3663)	(7.5096)	(7.2233)
Local resident	4.6416	-5.5168*	-5.4366*	7.4535**
	(2.9250)	(2.9339)	(3.0552)	(2.8492)
Attend kindergarten	-1.8706	1.2642	-2.2291	-.5564
	(3.9767)	(3.8201)	(3.8645)	(3.6747)
Father's Education	-.7261	.7579	-1.6946*	.3593
	(.8840)	(.9119)	(.9415)	(.9040)
Mother's Education	-.3528	-.5139	1.8081*	-.5879
	(1.0699)	(1.1102)	(1.0876)	(1.0458)

Sample Size: 3893, including 1890 females and 2003 males
Significance Level: * <10%, **< 5%, *** <1%

Table 19e: Results for Students and Head Teachers' Characteristics (Multiple Networks)

	Chinese	Mathematics	English	Total
Students' Characteristics:				
Only child in family	.5946*	.7728***	.5065	.8019**
	(.3486)	(.3702)	(.3534)	(.3762)
Relative age	-.0627***	-.1142**	-.0993***	-.1201***
	(.0233)	(.0247)	(.0236)	(.0251)
Minority	.2142	-.7511	-.0197	-.3891
	(.7077)	(.7495)	(.7102)	(.7640)
Local resident	-.3122	-.0315	-.1521	-.0121
	(.4046)	(.4291)	(.4016)	(.4364)
Attend kindergarten	1.6937***	1.6219***	1.6241***	1.8566***
	(.4192)	(.4425)	(.4256)	(.4499)
Father's Education	.2367**	.2782**	.4416***	.3761***
	(.1008)	(.1071)	(.1034)	(.1088)
Mother's Education	.2983***	.1610	.2304**	.2504**
	(.1041)	(.1105)	(.1042)	(.1124)
Head teachers' Characteristics:				
Female	-.1180	.4767	-.3119	.2077
	(.6354)	(.7766)	(1.4381)	(.6936)
Teaching experience	.0435	.0477	.0232	.0411
	(.0414)	(.0439)	(.0484)	(.0451)
Teach this course	.0803	.3344	-.4607	.0652
	(.5618)	(.5857)	(1.8734)	(.8861)
Log Likelihood	-13955	-14215	-14041	-14233
Efron's R^2	.1251	.0739	.1252	.6771

Table 20a: Correlation Coefficients Between Grades and Relative Ages

	All Range	Delayed Range (Relative Age>5)	Regular Range (-6≤Relative Age≤5)	Earlier Range (Relative Age<-6)
Chinese	-.0885	-.0224	.0013	-.0603
Mathematics	-.0886	-.0200	-.0333	-.1298
English	-.1124	-.0826	-.0150	-.0764
Total	-.1990	-.0576	-.0275	-.0766
# observations	3893	746	2768	379

Table 20b: Average Grades for Different Age Ranges

	All Range	Delayed Range (Relative Age>5)	Regular Range (-6≤Relative Age≤5)	Earlier Range (Relative Age<-6)
Chinese	70.6049	68.8624	70.8442	72.2879
Mathematics	70.6940	69.3231	70.8073	72.5650
English	70.6258	68.7782	70.8314	72.7606
Total	74.1286	66.0787	75.9426	76.7250
# observations	3893	746	2768	379

6 Conclusion

This paper considers higher-order spatial autoregressive models with group fixed effects to confront some conceptual problems in social interaction estimation. The heterogeneous peer and contextual effects can be separately identified and the peer effects can be disentangled from other confounding effects captured by the group fixed effects term. We show consistency and asymptotic normality of the proposed QMLE and verify its finite sample performance by Monte Carlo simulations. We detect significant gender disparities in peer effects from gender subgroups in a classroom for Chinese junior high school students, which provides justification for some policy related interventions aimed at improving social welfare in school learning. As in Lin (2010), the limitation of the group fixed effect model is that it can not deal with possible unobservable factors in common within groups.

References

[1] Aquaro M., N. Bailey, and Pesaran M.H., 2020. Estimation and inference for spatial models with heterogenous coefficients: An application to US house prices. *Journal of Applied Econometrics* 36: 18-44.

[2] Bedard, K., and Dhuey, E., 2006. The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects. *The Quarterly Journal of Economics* 121(4): 1437-1472.

[3] Bifulco R., Fletcher, J.M., and Ross S.L., 2011. The Effect of Classmate Characteristics on Post-Secondary Outcomes: Evidence from the Add Health. *American Economic Journal: Economic Policy* 3(1): 25-53.

[4] Brock, W.A., and Durlauf, S.N., 2001. Interaction-based models. *Handbook of Econometrics*, Volume 5, Chapter 54, 3297-3380.

[5] Chetty, R., Friedman, J.N., Hilger N., Saez, E., Schanzenbach, D.W., and Yagan, D., 2011. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics* 126(4): 1593-1660.

[6] Coleman, J.S., et al. 1966. Equality of educational opportunity. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

[7] Chung, B.W., and Zou, J., 2020. Understanding Spillover of Peer Parental Education: Randomization Evidence and Mechanisms. Working Papers 2020-045, Human Capital and Economic Opportunity Working Group.

[8] Davis-Kean, P.E., 2005. The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology* 19(2): 294-304.

[9] Delorme N., Boiché, J., and Raspaud, M., 2010. Relative Age Effect in Elite Sports: Methodological Bias or Real Discrimination? *European Journal of Sport Science* 10(2): 91-96.

[10] Dickson, M., Gregg, P., and Robinson, H., 2017. Early, Late or Never? When Does Parental Education Impact Child Outcomes? *The Economic Journal* 126(596): F184-F231.

[11] Falbo, T., and Poston, D.L., 1993. The Academic, Personality, and Physical Outcomes of Only Children in China. *Child Development* 64(1): 18-35.

[12] Fortin N. M., Oreopoulos P., and Phipps S., 2015. Leaving Boys Behind: Gender Disparities in High Academic Achievement. *Journal of Human Resources* 50(3): 549-579.

[13] Gong, J., Lu, Y., Song, H., 2018. The Effect of Teacher Gender on Students' Academic and Noncognitive Outcomes. *Journal of Labor Economics* 36 (3): 743-778.

[14] Gong, J., Lu, Y., Song, H., 2019. Gender Peer Effects on Students' Academic and Noncognitive Outcomes: Evidence and Mechanisms. *Journal of Human Resources* 56 (3): 686-710.

[15] Gupta A., and P.M. Robinson, 2018. Pseudo maximum likelihood estimation of spatial autoregressive models with increasing dimension. *Journal of Econometrics* 202(2018): 92-107.

[16] Helsen W.F., Winckel, J.V., and Williams A.M., 2005. The relative age effect in youth soccer across Europe. *Journal of Sports Sciences* 23(6): 629-636.

[17] Hoxby, C., 2000. Peer effects in the classroom, learning from gender and race variation. Working Paper no. 7867, National Bureau of Economic Research, Cambridge, MA.

[18] Hsieh, C.S., and X. Lin, 2017. Gender and racial peer effects with endogenous network formation. *Regional Science and Urban Economics* 67(2017): 135-147.

[19] Ladd H.F., and Sorensen, L., 2017. Returns to Teacher Experience: Student Achievement and Motivation in Middle School. *Education Finance and Policy* 12(2): 241-279.

[20] Lee, L.F., 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* 72 (6), 1899-1925.

[21] Lee, L.F. 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140, 333-374

[22] Li, B.J., and Zhang, H.L., 2017. Does population control lead to better child quality? Evidence from China's one-child policy enforcement. *Journal of Comparative Economics* 45(2017): 246-260.

[23] Lin, Xu., 2010. Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics* 28(4): 825-860.

[24] Lu, F., and M.L., Anderson, 2015. Peer effects in microenvironments: the benefits of homogenous classroom groups. *Journal of Labor Economics* 33(1): 91-122.

[25] Manski, C.F., 1993. Identification of endogenous social effects: The reflection problem. *Review of Economic Studies* 60 (3): 531-542.

[26] Musch, J., and Grondin, S., 2001. Unequal Competition as an Impediment to Personal Development: A Review of the Relative Age Effect in Sport. *Developmental Review* 21(2): 147-167.

[27] Poston, D.L., and Falbo, T., 1990. Academic Performance and Personality Traits of Chinese Children: "Onlies" Versus Others. *American Journal of Sociology* 96(2): 433-451.

[28] Rockoff, J.E., 2004. The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *American Economic Review: Papers & Proceedings* 94(2): 247-252.

[29] Rothenberg, T. J., 1971. Identification in parametric models. *Econometrica* 39 (3), 577-591.

[30] Tincani, M. M., 2018. Heterogeneous Peer Effects in the Classroom. *Working paper.*

[31] Trogdon, J.G., J. Nonnemaker, J. Pais, 2008. Peer effects in adolescent overweight. *Journal of Health Economics*, 27(2008), 1388-1399.

[32] Yakusheva O., Kapinos K., and D. Eisenberg. Estimating Heterogeneous and Hierarchical Peer Influences on Body Weight Using Roommate Assignment as a Natural Experiment. *Journal of Human Resources*, 49(1), January 2014, pp: 234-261.

[33] Yu, J., R. de Jong, and L.F. Lee, 2008. Quasi-Maximum Likelihood Estimators for Spatial Dynamic Panel Data With Fixed Effects When Both n and T Are Large. *Journal of Econometrics* 146, 118-134.