ABSTRACT

Title of dissertation: MEASURING PEER EFFECTS IN ACADEMIC OUTCOMES

Ying Zeng, Doctor of Philosophy, 2017

Dissertation directed by: Professors Guido Kuersteiner and Ingmar Prucha
Department of Economics

There is wide belief that a student's behavior and academic outcome are affected by her/his fellow students. This peer effect lies at the center of the debate about education policy. My dissertation focuses on measuring peer effects in academic outcomes. It consists of a theoretical part and an empirical part.

In the theoretical part, I study a peer effects model with group-wise equal interactions and random group effects. Identifying peer effects is notoriously challenging due to the reflection problem. Common shocks to the groups also generate spurious peer effects. My model, therefore, controls for the common shocks to the whole group with random group effects. My estimation strategy overcomes the identification problem with spatial econometrics techniques. I develop a quasi-maximum likelihood estimator of the model. Monte-Carlo simulations show that the bias of the estimator decreases with the number of groups and the variation in group size, and increases with group size. Finally, I prove the consistency and asymptotic normality of the estimator under standard assumptions.

In the empirical part, I apply the model to Project STAR data to study the peer effects among kindergarten students. Peers constitute an important context for children's academic development. This empirical study measures peer effects on math and reading scores of kindergarten children using data from Project STAR, an experiment in Tennessee that randomly assigned both children and teachers to classrooms of different sizes. It estimates the impact of peers' scores and characteristics on children's individual scores, controlling among other things for random class effects. In contrast to most existing studies, the estimated peer effects in the empirical part are small and insignificant. The results are robust when allowing peer effects to be heterogeneous by gender, using data from higher grades or considering alternative specifications. The findings of the empirical study cast doubt on the effectiveness of programs that manipulate peer groups for better educational outcomes.

MEASURING PEER EFFECTS IN ACADEMIC OUTCOMES

by

Ying Zeng

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professors Guido Kuersteiner and Ingmar Prucha, Co-Chairs
Professor John Chao
Professor Haluk Ünal
Professor Sergio Urzua

Acknowledgments

I would like to express my sincerest gratitude to my advisors, Professors Guido Kuersteiner and Ingmar Prucha. I am deeply indebted for their invaluable support, guidance and encouragement during my research and job market process.

I would also thank Professors John Chao, Haluk Ünal, and Sergio Urzua for serving on my committee and offering their helpful comments.

I am also thankful to Misty Heggeness for being such a good friend and mentor.

Special thanks to my advisors back in Peking University, Professors Cheng Yuan and Shuanglin Lin, for inspiring my interest in economics research.

My gratitude also goes to all the faculty, staff and students in the economics department at the University of Maryland. They make the department a friendly and supportive place.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The empirical literature has documented that students' behaviors and academic outcomes are influenced by their peers. This peer effect in education carries broad policy implications. It affects the evaluation of education programs and motivates regrouping policies that manipulate peer groups for better educational outcomes. However, Manski (1993) points out that a linear-in-means peer effects model, in which one's outcome is linear in the mean outcome of his peers, suffers from reflection problem. The reflection problem arises from the reciprocal nature of peer effects. Peer's outcomes affect one's own outcome, which in turn affects peers' outcomes. This gives the linear-in-means peer effects model a simultaneous equation feature and poses challenges to identification. Moreover, Angrist (2014) criticizes the existing (peer effects) literature for overestimating peer effects. He points out that peer effects model suffers from serious identification problem, and that spurious intra-group correlation can easily arise with problematic methods and confounding factors like common shocks to the whole group. Therefore, it's valuable for education policy to estimate peer effects with more reliable methods and better control of confounding factors. That's the goal of this dissertation.

This dissertation studies a peer effect model that controls for the common shocks with random group effects. It develops an estimation strategy of the model, which overcomes the reflection problem. It then applies the model and estimation method empirically to study peer effects among kindergarten students.

Chapter 1 is the introduction, describing the background and structure of the dissertation. Chapter 2 reviews the three lines of literature this dissertation is related to: empirical work on peer effects, studies with Project STAR data, spatial econometric and peer effects.

Chapter 3 is the theoretical part. It studies a peer effects model where peers interact in groups, one's outcome is linear in the mean outcome and characteristics of his peers in the group, and shocks to the whole groups are uncorrelated with characteristics of the group and group members. I study the model under the framework of spatial econometrics. Spatial econometrics models and identifies cross-sectional dependence in a space, not limited to geographic space. Existence and strength of social links can define the proximity in the social network space. Therefore spatial econometrics offers valuable tools and insights into modeling and identification of social network effects. I estimate the model with Quasi-maximum likelihood estimation (QMLE). Identification comes from the variation of group size. I conduct Monte-Carlo experiments to study the small sample properties of the QMLE estimator. Results show that the bias of the estimator is positively correlated with the group size and negatively correlated with the number of groups and variation of group size. Study of the large sample properties of the QMLE estimator shows that under standard assumptions, the QMLE estimator is consistent and has an asymptotically normal distribution. Formal proof for consistency and asymptotic normality is in Appendix E.

With enough variation of group size, the estimation strategy can overcome the reflection problem and identify peer effects. Further investigation of the first order moments reveals that identification is in fact based on both the within-group variance and between-group variance. The intuition behind my estimation strategy is that peer effects can reduce within-group variance and increase excess between-class variance. The magnitude of the changes in these variances incurred by peer effects depends on group sizes. For example, peer effects have smaller impacts on the within-group variance of larger groups because each person has a smaller impact on their peers. Identification is therefore possible with enough variation in class size.

Chapter 4 applies the model in Chapter 3 on Project STAR data. In kindergarten, Project STAR randomly assigned teachers and students to small classes and regular classes to study the effect of class size. This chapter measures the impact of peer's average score and average characteristics on kindergarten students' test score in Project STAR. The empirical study is a revisit to Graham (2008), which develops the conditional variance method to overcome the reflection problem and measure peer effects. Graham (2008) then applies the model to Project STAR data and find sizable peer effects in kindergarten children's test score. Both Graham's model and my model study the impact of average peer outcome on an individual's outcome, controlling for random group effects. But I also include additional control variables and relaxes some of the restrictions in his model. Likelihood ratio test also rejects the specification of Graham (2008). In all, the empirical part finds that the average peer outcome has an insignificant impact on an individual's outcome. The finding of insignificant peer effects differs from many. Such difference is driven by a new estimation strategy and a sufficient control of confounding forces. The finding echoes Angrist's argument that peer effects can easily arise from problematic methods and confounding factors like common shocks. Although the result does not necessarily generalize to other settings, it does urge researchers to be more cautious before attributing similarities in peer's outcomes to peer effects.

Chapter 2: Review of Literature

My dissertation contributes to three literature streams: empirical studies of peer effects, research with Project STAR data, and spatial econometric literature on group-wise interactions. This chapter reviews each of the three lines of literature.

2.1 Peer Effects in Education

Peer effects in education have drawn tremendous attention from researchers.[1] Sacerdote (2011) and Epple and Romano (2011) offer detailed reviews of the literature.

One line of the research examines peer effects in student's behaviors, e.g., risky behaviors (Eisenberg et al., 2014; Fletcher, 2011; Gaviria and Raphael, 2001), choosing major (De Giorgi et al., 2010), academic cheating (Carrell et al., 2008). Most of these studies find significant peer effects.

Another line of studies explores peer effects on academic performance, that is, how a student's academic achievement is affected by the academic ability of his or her peers. One limitation of these studies is that most of them cannot estimate

[1] Peer effects studies in other contexts, like technology adoption (Case, 1992; Dupas, 2014), welfare participation (Bertrand et al., 2000), insurance take-up (Cai et al., 2015), research productivity(Waldinger, 2012), charitable giving (Smith et al., 2015), workplace productivity (Mas and Moretti, 2009), etc.

the endogenous effects. Due to the reflection problem of the peer effects model, empirical studies typically measure the reduced form peer effects with regression methods. They either assume away endogenous effects (Carrell et al., 2009; Foster, 2006; Whitmore, 2005; Zimmerman, 2003) or acknowledge the distinction between endogenous and exogenous peer effects but do not try to separate them (Lyle, 2007; Sacerdote, 2001). Among these studies, some find significant positive peer effects (Carrell et al., 2009). Some only find moderate peer effects (Foster, 2006; Lyle, 2007; Angrist and Lang, 2004). For peer effects in academic achievement among children, results are mixed but in general find small but significant peer effects (Betts and Zau, 2004; Lefgren, 2004; Burke and Sass, 2013).

Some studies try to measure endogenous effects with the instrumental variable method (Boozer and Cacciola, 2001; Fletcher, 2011; Gaviria and Raphael, 2001). They find instruments for the mean peer outcome, as the reflection problem results from the endogeneity of it. One common choice of instrument is the mean of some individual characteristics. The validity of the instrumental variable requires that the mean of these individual characteristics does not have a direct impact on one's outcome, i.e., no exogenous peer effects for these variables. A plausible instrument variable is hard to find and its validity hard to justify. Angrist (2014) shows that instrumental variable method is very vulnerable to weak instrument problem.

So far there are two major strategies that can overcome the reflection problem and separate out the endogenous peer effects and exogenous peer effects. One is the conditional variance method of Graham (2008). The model assumes random group effects. It explores the relationship between within-group variance and between-group variance conditional on the difference in group size. The method does not measure endogenous peer effects directly but instead tries to detect the existence of peer effects.[2] The other is the spatial method, notably the conditional maximum likelihood estimation method of Lee (2007). Social networks or peer relationships can be characterized by spatial weights matrix in a spatial model. Spatial econometrics then provide a toolkit to estimate the model. Lee (2007) assumes fixed group effects and eliminates then with a transformation. Estimation relies on the distribution of deviations from the group mean conditional on fixed group effects. More details of both methods are discussed in Appendix H and Appendix I.

Empirical studies of endogenous peer effects in academic achievement are relatively rare but in general find significantly positive estimates for endogenous effects. Lin (2010) uses Add Health data of high school students in the United States and finds an endogenous peer effect of 0.27 for GPA. Boucher et al. (2014) use test scores of secondary school students in Canada. They find a significant endogenous peer effect in math scores of 0.83. But the endogenous peer effects are insignificant for other subjects. Both Boucher et al. (2014) and Lin (2010) use the estimation method in Lee (2007) and assume fixed group effects. Graham (2008) uses the conditional variance method to measure the endogenous effect indirectly with Project STAR data. His results are equivalent to an endogenous effect of 0.46 for math scores and 0.57 for reading scores in Kindergarten, though the estimates for reading is insignificant at the 5% level.

[2] For details of the method, see Appendix H.

I revisit Graham's paper and estimate peer effects in kindergarten with Project STAR data. Graham's model is estimated with a new methodology, which yields similar estimates. Moreover, an alternative specification that controls for school by class type fixed effects renders the endogenous peer effect insignificant.

2.2 Studies with Project STAR

The Project STAR data has been widely used to study the impact of class size, teacher quality, peer characteristics and early education on children's development. For a review of the literature, see Schanzenbach (2006) and Sohn (2014).

While studies of class size generate mixed results (Chingos, 2012; Jepsen and Rivkin, 2009), the studies with Project STAR data generally show that small classes improve students' test scores (Krueger, 1999).

There are two ways to measure teacher quality. One is to use observed teacher characteristics like education and experience. The other is to attribute the between-class variance to unobserved teacher effectiveness as in Rivkin et al. (2005). Chetty et al. (2011) and Nye et al. (2004) both use within-school between-class variance in Project STAR data to measure teacher quality and find significant teacher effects on students' academic performance.

The STAR data is also popular among peer effect studies (Boozer and Cacciola, 2001; Chetty et al., 2011; Graham, 2008; Sojourner, 2013; Whitmore, 2005). Among them, only Boozer and Cacciola (2001) and Graham (2008) estimate endogenous peer effects.

Among the studies of exogenous peer effects, Whitmore (2005) finds that the share of girls in class has a positive impact on students' performance. Chetty et al. (2011) show that peer's test scores are positively related to one's future outcome like earnings and college attendance. Sojourner (2013) avoids the reflection problem by using the leave-out-mean of lagged test scores as the regressor. He finds positive impacts of peer's previous scores on a student's current academic performance.

Boozer and Cacciola (2001) try to overcome the reflection problem with instrument variable method. They instrument the class mean score with the proportion of peers assigned to small classes, switching classes or newly joining the class. They find large endogenous peer effect, 0.86 in grade two and 0.92 in grade three, yet the method may suffer from the weak instrument problem.

2.3 Spatial Econometrics and Peer Effects

In the theoretical part of my dissertation, model specification and estimation are under the framework of spatial econometrics. The field of spatial econometrics models and measures spillovers among units. Study of the model has expanded considerably since the seminal work of Cliff and Ord (1973), Cliff and Ord (1981), and Anselin (1988).[3]

The Cliff-Ord type models specify spatial weights as proximity between units, including but not limited to geographic proximity, economic proximity, and social proximity, therefore extend the application of spatial econometrics beyond the scope

[3] Anselin (2010) offers a brief review of the development of spatial econometrics literature over the past thirty years.

of geographic space. Empirical work of spatial econometrics originally concentrates on the geographic space and has broad applications in regional and urban economics (Anselin, 1988; LeSage, 2008). The spatial econometrics method is also increasingly applied in a fuller range of economics topics like fiscal policy, household demand, technology change, etc.[4] Recently there is a growing literature using spatial methods to model social network effects, e.g., Lee (2007), Bramoullé et al. (2009), and Kuersteiner and Prucha (2015). The strength of social links can be characterized by proximity in the social network space.

The spatial models were traditionally estimated with maximum likelihood (ML) estimation method, e.g., Ord (1975). Kelejian and Prucha (1998, 1999) develop the generalized method of moments (GMM) estimator for spatial models, which are based on the linear and quadratic moments. While this paper utilizes a quasi-maximum likelihood estimation method, statistical analysis relies on the linear quadratic forms of the error terms. The properties of quadratic moment conditions were introduced by Kelejian and Prucha (1998, 1999) in the cross section case, and Kapoor et al. (2007) and Kuersteiner and Prucha (2015) in the panel setting. Moreover, Kelejian and Prucha (2001) and Kelejian and Prucha (2010) introduce central limit theorem for linear quadratic forms, which lays the foundation of asymptotic properties of the estimator of the theoretical part.

The linear-in-means peer effect model in Manski (1993) is akin to a spatial model with group-wise equal dependence,which has been studied by Kelejian and

[4]Some recent empirical applications of spatial models include Conley and Dupor (2003); Cohen and Paul (2004); Case (1991, 1992); Parent and LeSage (2008); Ertur and Koch (2007); Agrawal (2015); Case et al. (1993).

Prucha (2002) and Kelejian et al. (2006). Kelejian and Prucha (2002) was the first to study the group-wise equal dependence spatial model. They show that if there is one group in a single cross section and the model has equal spatial weights, two-stage least squares (2SLS) method, GMM and quasi-maximum likelihood estimation (QMLE) methods all yield inconsistent estimators. With panel data so that the model has blocks of equal weights, 2SLS and GMM estimator can be consistent. However, Kelejian et al. (2006) point out that if group fixed effects are incorporated and the panel is balanced, the estimators are inconsistent. The results in Kelejian et al. (2006) show the importance of variation in group size in identification of spatial models with blocks of equal weights. The quasi-maximum likelihood estimator in this chapter and the conditional maximum likelihood estimator in Lee (2007) both rely on group size variation for identification.

Lee (2007) works with a special case of the model in Kelejian et al. (2006) where there are blocks of equal weights and fixed block effects, but the sizes of the blocks are different. He interprets the model as a peer effects model. [5]He eliminates fixed group effects with a transformation and estimates the within-equations of the model with maximum likelihood method and 2SLS method. Identification requires variation in group size, echoing the conclusion in Kelejian et al. (2006). Lee (2007)'s model is later extended to allow for specific social structure within the group (Lee et al., 2010), non-row normalized weight matrix (Liu and Lee, 2010; Liu et al., 2014). It has also been applied to the empirical evaluation of peer effects by Lin (2010) and Boucher et al. (2014). Bramoullé et al. (2009) study a broader range

[5]For a detailed discussion of the model and estimation strategy of Lee (2007), see Appendix I.

of social interaction models with spatial econometrics method, with Lee (2007)'s model been a special case. In their models, the unobserved correlated effect is either absent or assumed to be fixed effect and eliminated. Interactions are allowed to be non-group-wise. They establish identification conditions for these models.

Chapter 3 extends the spatial model with blocks of equal weights in Kelejian et al. (2006) and Lee (2007) by assuming random group effects.

Chapter 3: A Peer Effects Model with Random Group Effects

In this chapter, I provide a theoretical discussion of a peer effects model with random group effects.

3.1 Model Setup

Suppose that there are R groups in the sample indexed by $c = 1, ..., R$. Let n_c be the size of group c. The total sample size is $N = \sum_{c=1}^{R} n_c$. Assume equal interaction within each group and no interaction across groups. Peer effects work through the mean outcome and mean characteristics of peers in the same group. The linear-in-means peer effects model in this chapter is

$$y_{ic} \;=\; \beta_0 + \lambda \bar{y}_{(-i)c} + \mathbf{x}'_{1,ic}\beta_1 + \bar{\mathbf{x}}'_{2,(-i)c}\beta_2 + \bar{\mathbf{x}}'_{3,c}\beta_3 + \alpha_c + \epsilon_{ic}, \tag{3.1}$$

where y_{ic} is the outcome variable of individual i in group c, $\bar{y}_{(-i)c} = \frac{1}{n_c - 1}\sum_{j \neq i}^{n_c} y_{ic}$ is the average outcome of i's peers, $\mathbf{x}_{1,ic}$ and $\mathbf{x}_{2,ic}$ are both vectors of predetermined characteristics of individual i in group c, $\bar{\mathbf{x}}_{2,(-i)c} = \frac{1}{n_c - 1}\sum_{j \neq i}^{n_c} \mathbf{x}_{2,jc}$ is a vector of average characteristics of i's peers, $\mathbf{x}_{3,c}$ is a vector of observed group characteristics. The variables in $\mathbf{x}_{1,ic}$ and $\mathbf{x}_{2,ic}$ can be non-overlapping, partially overlapping or totally overlapping. The error term consists of two components, the group effect

α_c and the disturbance term ϵ_{ic}. By assumption, $\mathbf{x}_{1,ic}$, $\bar{\mathbf{x}}_{2,(-i)c}$, and $\mathbf{x}_{3,c}$ are non-stochastic.

In this model, peer effects work through the mean peer outcome $\bar{y}_{(-i)c}$ and mean peer characteristics $\bar{\mathbf{x}}_{2,(-i)c}$. The two terms are also known as the leave-out-mean of y and $\mathbf{x}_2$, as they are means of the group leaving out oneself. In Manski's terminology, λ in equation (3.1) reflects endogenous peer effects, β_2 reflects exogenous peer effects.

Let $\mathbf{z}_{ic} = (1, \mathbf{x}'_{1,ic}, \bar{\mathbf{x}}'_{2,(-i)c}, \mathbf{x}'_{3,c})$ be the vector of all exogenous variables, $\gamma = (\beta_0, \beta'_1, \beta'_2, \beta'_3)'$ be the corresponding coefficients vector. Denote the number of columns in $\mathbf{z}_{ic}$ as k_Z. A compact form of model (3.1) is

$$y_{ic} = \lambda \bar{y}_{(-i)c} + \mathbf{z}'_{ic}\gamma + \alpha_c + \epsilon_{ic}. \tag{3.2}$$

The model can be further written as a Cliff-Ord type spatial model. Let W_c be the weights matrix of group c, whose off-diagonal elements are $\frac{1}{n_c-1}$ and diagonal elements are 0. Let $Y_c = (y_{1c}, ..., y_{n_c c})'$, $Z_c = (\mathbf{z}'_{1c}, ..., \mathbf{z}'_{n_c c})'$, $\epsilon_c = (\epsilon_{1c}, ..., \epsilon_{n_c c})'$. Let $\iota_c = (1, ..., 1)'$ be the n_c dimensional vector of ones. The model for group c in matrix form is

$$Y_c = \lambda W_c Y_c + Z_c \gamma + U_c, \tag{3.3}$$

where U_c is the composite error term:

$$U_c = \alpha_c \iota_c + \epsilon_c. \tag{3.4}$$

Let $Y = [Y'_1, Y'_2, ..., Y'_R]'$, $Z = [Z'_1, Z'_2, ..., Z'_R]'$, $U = [U'_1, U'_2, ..., U'_R]'$, and $W = diag_{c=1}^{R}\{W_c\}$. The model for the whole sample, in matrix form, is given as

$$Y = \lambda WY + Z\gamma + U. \tag{3.5}$$

Solving Y from equation (3.5) yields the reduced from:

$$Y = (I - \lambda W)^{-1} Z\gamma + (I - \lambda W)^{-1} U. \tag{3.6}$$

3.2 Assumptions

For estimation, I make the following assumptions.

Assumption 1. *For $c = 1, ..., R$, the group effects α_c are independently and identically distributed, with $E\alpha_c = 0, E\alpha_c^2 = \sigma_{\alpha 0}^2$, where $\sigma_{\alpha 0}^2 \geqslant 0$. There exists some $\eta_\alpha > 0$ such that $E|\alpha_c|^{4+\eta_\alpha} < \infty$.*

Under Assumption 1, the third and fourth moments of α_c exist. Let $E(\alpha_c)^3 = \mu_\alpha^{(3)}$ and $E\alpha_c^4 = \mu_\alpha^{(4)}$.

Assumption 2. *The disturbance terms ϵ_{ic} are independently and identically distributed across all i and c, with $E(\epsilon_{ic}) = 0$, $E(\epsilon_{ic}^2) = \sigma_{\epsilon 0}^2$, where $0 < \underline{a_\epsilon} \leqslant \sigma_{\epsilon 0}^2$. There some exists $\eta_\epsilon > 0$ such that $E|\epsilon_{ic}|^{4+\eta_\epsilon} < \infty$. Also, $\{\alpha_c, c = 1, ..., R\}$ are independent of $\{\epsilon_{ic} : i = 1, ..., n_c, c = 1, ..., R\}$.*

Under Assumption 2, the third and fourth moments of ϵ_c exists. Let $E(\epsilon_{ic})^3 = \mu_\epsilon^{(3)}$ and $E\epsilon_{ic}^4 = \mu_\epsilon^{(4)}$.

Denote the variance-covariance (VC) matrix of U_c in equation (3.4) as Ω_c and the variance-covariance matrix of $U = (U_1', U_2', ..., U_R')'$ as Ω. Under Assumptions 1 and 2,

$$\Omega_c = var(U_c) = \sigma_\alpha^2 J_c + \sigma_\epsilon^2 I_c, \tag{3.7}$$

where $J_c = \iota_c \iota_c'$ is the $n_c \times n_c$ matrix of ones, I_c is the n_c-dimensional identity matrix. Given the independence of α_c and ϵ_c across groups, the variance-covariance matrix for the whole sample is

$$\Omega = var(U) = diag_{c=1}^{R}\{\Omega_c\} = \sigma_\alpha^2 \tilde{J} + \sigma_\epsilon^2 I, \tag{3.8}$$

where $\tilde{J} = diag_{c=1}^{R}\{J_c\}$.

Let $\mathcal{I}_n$ be the index set of all groups with size equal to n. If $c \in \mathcal{I}_n$, then $n_c = n$. Let R_n be the cardinality of $\mathcal{I}_n$, or the number of groups with size equal to n. Denote the share of groups with size n as ω_n, $\omega_n = R_n/R$.

Assumption 3. *(a) For all $c = 1, 2, ..., R$, group size n_c is an integer greater than 2 and bounded above, $2 \leqslant \underline{a} \leqslant n_c \leqslant \bar{a} < \infty$; (b) The limit of ω_n exists, $lim_{R\to\infty}\omega_n = \omega_n^*$. There exists some $\varepsilon_\omega > 0$, such that for $\underline{a} \leqslant n \leqslant \bar{a}$, $0 \leqslant \omega_n < 1 - \varepsilon_\omega$ and hence $0 \leqslant \omega_n^* < 1 - \varepsilon_\omega$; (c) The number of groups R goes to infinity.*

The restriction that group size is no smaller than two rules out single-member groups. In such groups a student does not have any peers. Since group size $n_c \geqslant \underline{a} \geqslant 2$, the total sample size $N \geqslant \underline{a}R \geqslant 2R$. The sample size N goes to infinity as the number of groups R goes to infinity. In the following I will also use the notation N_R instead of N when it is important to stress the dependence of the total sample size on R.

Since $\sum_{n=\underline{a}}^{\bar{a}} R_n = R$, we have $\sum_{n=\underline{a}}^{\bar{a}} \omega_n = 1$ and $\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* = 1$. Condition $\omega_n^* < 1 - \varepsilon_\omega$ in Assumption 3(b) ensures that in the limit, group size is not the same for all groups. Assumption 3(b) also implies that the average group size converges to a constant. Denote the limit of average group size as n^*,

$$n^* = lim_{R\to\infty}\frac{N_R}{R} = lim_{R\to\infty}\sum_{n=\underline{a}}^{\bar{a}}\frac{R_n}{R}n$$
$$= \sum_{n=\underline{a}}^{\bar{a}}\omega_n^* n. \tag{3.9}$$

Since n_c is uniformly bounded, n^* is uniformly bounded, $2 \leqslant \underline{a} \leqslant n^* \leqslant \bar{a} < \infty$.

Assumption 4. *The endogenous peer effect $\lambda \in \Lambda$, where Λ is a compact subset of $(1-\underline{a}, 1)$.*

Assumption 4 ensures non-singularity of $I_c - \lambda W_c$ and hence non-singularity of $I - \lambda W = diag_{c=1}^{R}\{I_c - \lambda W_c\}$ in equation (3.6). To see this,

$$I_c - \lambda W_c = (1+\frac{\lambda}{n_c-1})I_c - \frac{1}{n_c-1}J_c$$
$$= (1+\frac{\lambda}{n_c-1})I_c^* + (1-\lambda)J_c^*, \tag{3.10}$$

where $I_c^* = I_c - J_c/n_c$, $J_c^* = J_c/n_c$. By Lemma A.3 in Appendix A, $I_c - \lambda W_c$ is nonsingular if $1 + \lambda/(n_c - 1) \neq 0$ and $1 - \lambda \neq 0$. The inverse of $I_c - \lambda W_c$ is in equation (B.4). Since $n_c \geqslant \underline{a}$ under Assumption 3, $n_c - 1 + \lambda > 0$ if $\lambda > 1 - \underline{a}$. Meanwhile, $1-\lambda < 0$ as $\lambda < 1$. Therefore, $I_c-\lambda W_c$ is invertible under Assumption 4.

The parameters of interest are $\lambda, \sigma_\epsilon^2, \sigma_\alpha^2$ and γ. Their true values are $\lambda_0, \sigma_{\epsilon 0}^2, \sigma_{\alpha 0}^2$ and γ_0 respectively. As will be shown later, γ can be concentrated out from the likelihood function and analysis will focuses on $\lambda, \sigma_\epsilon^2, \sigma_\alpha^2$. Let $\vartheta = (\lambda, \sigma_\epsilon^2, \sigma_\alpha^2)$ be the parameter vector, $\vartheta \in \Theta$, where Θ is the parameter space. From Assumptions 1, 2, and 4, Θ is a compact subset of the Euclidean space R^3 .

For an $N_R \times N_R$ matrix $A_N(\vartheta)$, denote the i,jth element as $A(\vartheta)_{ij,N}$. I call the elements of the sequence of matrices $A_N(\vartheta)$ are uniformly bounded in absolute value if $sup_{\vartheta\in\Theta,1\leqslant i,j\leqslant N}|A(\vartheta)_{ij,N}| < \bar{a}_A$, where $\bar{a}_A$ is a finite constant that does not depend on ϑ or N.

Let $\Omega(\vartheta) = \sigma_\epsilon^2 I + \sigma_\alpha^2 \tilde{J}$, where $\tilde{J} = diag_{c=1}^R\{J_c\}$. According to Lemma A.3, $\Omega(\vartheta)$ is nonsingular for all $\vartheta \in \Theta$ under Assumptions 1 and 2.

Let $\bar{Z}_c = \frac{1}{n_c}\iota_c' Z_c$ be the row vector of column means of Z_c, $Z_c^* = Z_c - \iota_c \bar{Z}_c$ be the deviations from the column means.

Assumption 5. *(a) The $N \times k_Z$ matrix Z is non-stochastic, with $rank(Z) = k_Z$, wherek_Z is a finite constant. The elements of Z are uniformly bounded in absolute value.*

(b)For $\underline{a} \leqslant n \leqslant \bar{a}$, $lim_{R\to\infty} R_n^{-1} \sum_{c\in\mathcal{I}_n} Z_c^{*\prime} Z_c^* = \varkappa_n^*$, $lim_{R\to\infty} R_n^{-1} \sum_{c\in\mathcal{I}_n} \bar{Z}_c' \bar{Z}_c = \bar{\varkappa}_n$, $lim_{R\to\infty} R_n^{-1} \sum_{c\in\mathcal{I}_n} Z_c' \iota_n = \underline{Z}_n$.

(c) For all $\vartheta \in \Theta$, $Z'\Omega(\vartheta)^{-1}Z$ is nonsingular, and the elements of $N_R[Z'\Omega(\vartheta)^{-1}Z]^{-1}$ are uniformly bounded in absolute value.

Assumption 5 is necessary for identification. Since the elements of Z are uniformly bounded in absolute value, $\varkappa_n^*$ and $\bar{\varkappa}_n$ are finite $k_Z \times k_Z$ matrices, and $\underline{Z}_n$ is a $k_Z \times 1$ vector of finite elements. With the closed form expression of $\Omega(\vartheta)^{-1}$ in equation (B.2) in Appendix B,

$$\frac{1}{N_R} Z'\Omega(\vartheta)^{-1}Z = \frac{1}{N_R}\frac{1}{\sigma_\epsilon^2}\sum_{c=1}^R Z_c^{*\prime} Z_c^* + \frac{1}{N_R}\sum_{c=1}^R \frac{n_c \bar{Z}_c' \bar{Z}_c}{\sigma_\epsilon^2 + n_c \sigma_\alpha^2}, \tag{3.11}$$

$$lim_{R\to\infty} \frac{1}{N_R} Z'\Omega(\vartheta)^{-1}Z = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* \left(\frac{\varkappa_n^*}{\sigma_\epsilon^2} + \frac{n\bar{\varkappa}_n}{\sigma_\epsilon^2 + n\sigma_\alpha^2}\right). \tag{3.12}$$

Under Assumption 3, $2 \leqslant \underline{a} \leqslant n \leqslant \bar{a} < \infty$, so $0 < 1/\bar{a} \leqslant 1/n^* \leqslant 1/\underline{a} < \infty$. By Assumption 2, $\sigma_\epsilon^2 > \underline{a}_\epsilon > 0$, so $1/(\sigma_\epsilon^2 + n\sigma_\alpha^2) \leqslant 1/\sigma_\epsilon^2 < 1/\underline{a}_\epsilon < \infty$. Therefore, the right hand side of equation (3.12) is finite, $lim_{R\rightarrow\infty}\frac{1}{N_R}Z'\Omega(\vartheta)^{-1}Z$ exists and is finite. By Assumption 5(c), $lim_{R\rightarrow\infty}\frac{1}{N_R}Z'\Omega(\vartheta)^{-1}Z$ are nonsingular.

3.3 Estimation Strategy

This chapter develops a quasi-Maximum likelihood estimation method for the peer effects model in equation (3.2). Assumptions 1 and 2 do not assume normality of the error terms. In this section, I will use a normality assumption to motivate the criterion function for maximum likelihood estimation. But normality is not essential for the consistency or the asymptotic normality of the estimator, as I will show in later sections.

From equation (3.6), the distribution of Y is

$$Y \sim ((I-\lambda_0 W)^{-1}Z\gamma_0, (I-\lambda_0 W)^{-1}\Omega_0(I-\lambda W)^{-1}). \tag{3.13}$$

If α_c and ϵ_{ic} follow normal distributions, the log likelihood function corresponding to the distribution of Y in equation (3.13) is

$$\begin{aligned} logL(\vartheta,\gamma) = & -\frac{N_R}{2}ln(2\pi) + ln|I-\lambda W| - \frac{1}{2}ln|\Omega(\vartheta)| \\ & -\frac{1}{2}(Y-\lambda WY - Z\gamma)'\Omega(\vartheta)^{-1}(Y-\lambda WY-Z\gamma), \end{aligned} \tag{3.14}$$

where $\Omega(\vartheta) = \sigma_\epsilon^2 I_N + \sigma_\alpha^2\tilde{J}$. The first order condition for γ is

$$\frac{\partial logL(\vartheta,\gamma)}{\partial\gamma} = (Y-\lambda WY-Z\gamma)'\Omega(\vartheta)^{-1}Z = 0. \tag{3.15}$$

Solving equation (3.15) yields

$$\hat{\gamma}(\vartheta) = (Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}(I-\lambda W)Y. \tag{3.16}$$

Plugging $\hat{\gamma}(\vartheta)$ back into equation (3.14), the concentrated log likelihood function is

$$Q_R(\vartheta) = -\frac{N}{2} ln(2\pi) + ln|I - \lambda W| - \frac{1}{2} ln|\Omega(\vartheta)| \\ - \frac{1}{2} Y'(I - \lambda W)' M_Z(\vartheta)(I - \lambda W)Y, \tag{3.17}$$

where

$$M_Z(\vartheta) = \Omega(\vartheta)^{-1} - P_Z(\vartheta), \tag{3.18}$$

$$P_Z(\vartheta) = \Omega(\vartheta)^{-1} Z(Z'\Omega(\vartheta)^{-1}Z)^{-1} Z'\Omega(\vartheta)^{-1}. \tag{3.19}$$

It is easy to see that $Q_R(\vartheta)$ is continuous in $\vartheta \in \Theta$.

3.4 Large Sample Properties

In this section, I describe the consistency and asymptotic distribution of the quasi-maximum likelihood estimator.

Let $\hat{\vartheta}_R = (\hat{\lambda}, \widehat{\sigma^2_\epsilon}, \widehat{\sigma^2_\alpha})_R$ be the maximum likelihood estimator for the concentrated log likelihood function $Q_R(\vartheta)$ in equation (3.17),

$$Q_R(\hat{\vartheta}_R) = max_{\vartheta \in \Theta} Q_R(\vartheta). \tag{3.20}$$

Given the assumptions outlined in Section 3.2, the quasi-maximum likelihood estimator is consistent.

Theorem 3.1. *Under Assumptions 1-5, the maximum likelihood estimator $\hat{\vartheta}_R$ is consistent, $\hat{\vartheta}_R \to_p \vartheta_0$ as $R \to \infty$.*

Proof of Theorem 3.1 is in Section E.1 of Appendix E.

Additional assumptions are needed for the asymptotic distribution of the quasi-maximum likelihood estimator. Define 3×3 symmetric matrices $\Xi_{2,n}$, $\Xi_{3,n}$, $\Xi_{4,n}$,

$\Xi_{0,n}$ and $\Xi_{1,n}$ as in equations (D.7), (D.9), (D.11), (D.3), and (D.5) in Appendix D. Note that $\Xi_{0,n}$, $\Xi_{1,n}$, $\Xi_{2,n}$, $\Xi_{3,n}$, and $\Xi_{4,n}$ are all symmetric so the dots in the upper right part represent corresponding symmetrical elements.

Assumption 6. *(a) Let* $\Gamma_0 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* \Xi_{0,n}$ *, where* $\Xi_{0,n}$ *is defined in equation* (D.3) *in Appendix D,* Γ_0 *is positive definite.*

(b) Let

$$\Psi_0 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* [\Xi_{0,n} + (\mu_\epsilon^{(4)} - 3\sigma_{\epsilon 0}^{(4)})\Xi_{1,n} + (\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^{(4)})\Xi_{2,n} + \mu_\epsilon^{(3)}\Xi_{3,n} + \mu_\alpha^{(3)}\Xi_{4,n}], \tag{3.21}$$

where $\Xi_{0,n}$*,* $\Xi_{1,n}$*,* $\Xi_{2,n}$*,* $\Xi_{3,n}$*, and* $\Xi_{4,n}$ *are defined in Appendix D. Let* $\rho_{min}(\Psi_0)$ *be the smallest eigenvalue of* Ψ_0*,* $0 < \underline{a}_\rho \leqslant \rho_{min}(\Psi_0)$*.*

Appendix E shows that the limit of the Hessian matrix (observed information matrix) at ϑ_0 is $-\Gamma_0$, and that the limit of the Fisher information matrix at ϑ_0 is Ψ_0.

Theorem 3.2. *Under Assumptions 1-7,* $\sqrt{N_R}(\hat{\vartheta}_R - \vartheta_0) \xrightarrow{D} N_R(0, \Phi_0)$ *as* R *goes to infinity, where* $\Phi_0 = \Gamma_0^{-1}\Psi_0\Gamma_0^{-1}$*.*

Note that if α_c and ϵ_{ic} follow normal distributions, $\mu_\alpha^{(3)} = 0$, $\mu_\epsilon^{(3)} = 0$, $\mu_\epsilon^{(4)} - 3\sigma_{\epsilon 0}^{(4)} = 0$, $\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^{(4)} = 0$. Therefore $\Psi_0 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* \Xi_{0,n} = \Gamma_0$. In that scenario, $\Phi_0 = \Gamma_0^{-1} = \Psi_0^{-1}$.

Proof of Theorem 3.2 is in Section E.2 of Appendix E.

3.5 Finite Sample Properties

I conduct Monte-Carlo experiments to study the finite sample properties of the quasi-maximum likelihood (QML) estimator $\hat{\vartheta}_R$.

3.5.1 Design of the Monte Carlo Experiments

The simulations are based on a simplified version of the main model in equation (3.2). I reduce the vectors $\mathbf{x}_{1,ic}$, $\mathbf{x}_{2,ic}$, $\mathbf{x}_{3,c}$ to be scalars $x_{1,ic}$, $x_{2,ic}$, $x_{3,c}$. The model is then

$$y_{ic} = \lambda \bar{y}_{(-i)c} + \beta_0 + \beta_1 x_{1,ic} + \beta_2 \bar{x}_{2,(-i)c} + \beta_3 x_{3,c} + \alpha_c + \epsilon_{ic}. \tag{3.22}$$

I set the true value of the parameters to $\lambda = 0.5$, $\beta_0 = 1$, $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 1$, $\sigma_\alpha^2 = 0.5^2$, and $\sigma_\epsilon^2 = 0.5^2$. The number of groups R is from the set $\{50, 100, 300\}$.

As later results will show, the type of distribution of group size affects the performance of the QML estimator. Therefore, I use different distributions of group sizes. Group sizes follow a discrete uniform distribution or discrete normal distribution, with small/large mean, and small/large standard deviation. Table 3.1 offers summary statistics of group sizes for both distributions.

I allow x_1 and x_2 to be the same or different in different experiments. In case one, $x_{1,ic} = x_{2,ic} \sim N(0,1)$, so one's characteristic x_1 impacts oneself as well as one's peers directly. In case two, $x_{1,ic} \sim N(0,1)$ and $x_{2,ic} \sim N(0,1)$ are independently drawn, so one's characteristic x_1 directly impact only oneself, and one's characteristic x_2 directly impacts only peers. Group characteristics $x_{3,c}$ and group effects α_c are independently drawn from normal distributions, with $x_{3,c} \sim$

$N(0, 1)$ and $\alpha_c \sim N(0, 0.5^2)$. The disturbance terms ϵ_{ic} are independently drawn from a normal distribution, $\epsilon_{ic} \sim N(0, 0.5^2)$. The dependent variable y_{ic} is generated as in equation (3.3).

I generate 300 repetitions for each of the experiments. A summary of the simulation results is in Tables 3.2, 3.3, 3.4, and 3.5. In both Tables 3.2 and 3.3, group sizes follow a discrete uniform distribution. In Tables 3.4 and 3.5, group sizes follow a normal distribution. In Tables 3.2 and 3.4, x_1 and x_2 overlap, so that the model includes only one individual characteristic, which affects both oneself and one's peers directly. In Tables 3.3 and 3.5, x_1 and x_2 are different and independently drawn, so that the model includes two types of individual characteristics, $\mathbf{x}_1$ directly affects oneself only and $\mathbf{x}_2$ directly affects only peers.

3.5.2 Results

The key parameter of interest is λ. First let's compare the results within each table, i.e., within the same distribution type of group size (discrete normal or discrete uniform) and the same relationship between x_1 and x_2 ($x_1 = x_2$ or x_1, x_2 are independent). One observation is that the bias and variance for the estimates of λ decrease as the number of group R increases. This can be shown by comparing the results as the size increases but the mean and standard deviation of group sizes are fixed within each table. Another observation is that in general, the bias and variance increase as the mean of group sizes increases or as the standard deviation of group sizes decreases. Therefore, the estimator does not necessarily converge to the true value as the sample size increases. If the number of groups is fixed but mean group size increases, the estimator deviates further away from the true value.

Second, the estimates for λ perform better when x_1 and x_2 are independent than when they are equal. To see this, compare between the results in Tables 3.2 and 3.3, and between the results in Tables 3.4 and 3.5. Note that the difference between Table 3.2 and Table 3.3, and between Table 3.4 and Table 3.5 is that in Tables 3.2 and 3.4 $x_1 = x_2$ while in Tables 3.3 and 3.5 x_1 and x_2 are independent. All other parameters are the same. The bias and variance of λ is smaller when x_1 and x_2 are independent.

Third, the QML estimator for λ has smaller bias and variance when the group sizes follow a normal distribution than when they follow a uniform distribution.

In all, the performance of the estimator for λ improves as (1) the number of groups increases; (2) the variance of group size increase; (3) average group size decrease; (4) the model includes individual characteristics that does not directly affect peers; (5) the group size follows a normal distribution rather than a uniform distribution.

The performance of the estimators for other parameters is parallel to that of λ. But estimates for σ_ϵ^2 and β_1 get close to the true value with smaller variance even under undesirable conditions: when the number of groups is small, group sizes have a large mean and a small variance, x_1 and x_2 are the same.

3.5.3 Comparison with Lee (2007)

The difference between my model and the model in Lee (2007) is that Lee (2007) allows for group effects to be correlated with exogenous variables while I assume random group effects. The conditional maximum likelihood estimator by Lee (2007) is still consistent when the group effects are in fact random. But the variance will be larger. To show this, I compare my simulation results with those in Lee (2007).

In the Monte Carlo experiments of Lee (2007), group sizes are integers between 2 to 11 for the small interaction case, i.e., $\{2, 3, ..., 11\}$; or these integers multiplied by 8 or 10 for the large interaction case, i.e., $\{16, 24, ..., 88\}$ or $\{20, 30, ..., 110\}$. The number of groups of each size is the same and equals to the total number of groups divided by 10. Therefore, Lee's simulation process fails to account for the impact of variation and distribution type of group sizes on the estimator. As results in Tables 3.2, 3.3, 3.4, and 3.5 show, the bias and variance of the estimator both decrease with the variance of group sizes and are affected by the distribution type of group sizes.

To better compare my results with Lee's results, I use the same distribution of group sizes as his. But the data generating process is different. There are no group effects in the data generating process of Lee's Monte Carlo experiments as his research assumes fixed group effects. In this study, the data generating process involves random group effects. I use my model in equation (3.22) as the true data generating process, and estimate the sample with both my methodology and Lee's conditional maximum likelihood estimator (CMLE). [1]The results are in Tables 3.6 and 3.7.

Comparing the estimates for λ, the QML estimator in this study are generally biased downward in small samples and the CMLE of Lee (2007) are biased upward. When the average group size is small, my estimates are less biased than Lee (2007). In all cases, the variance of Lee's CMLE is much larger.

---

[1]In Lee's data generating process,

$$y_{ic} \quad = \quad \lambda \bar{y}_{(-i)c} + \beta_1 x_{1,ic} + \beta_2 \bar{x}_{2,(-i)c} + \epsilon_{ic}. \tag{3.23}$$

It differs from equation (3.22) in that $\beta_0, x_{3,c}$ and α_c are dropped, because demeaning the equation by groups eliminates these variables.

3.6 Tables

Table 3.1: Summary Statistics of Group Sizes

	min	max	mean	sd
Normal Distribution				
small mean, small sd	6	13	9.5	1.76
small mean, large sd	2	17	9.5	3.50
large mean, small sd	16	23	19.5	1.76
large mean, large sd	12	27	19.5	3.50
Uniform Distribution				
small mean, small sd	7	14	10.5	2.29
small mean, large sd	3	18	10.5	4.61
large mean, small sd	17	24	20.5	2.29
large mean, large sd	13	28	20.5	4.61

Table 3.2: Monte Carlo Results: Uniform Distribution, $x_1 = x_2$

size mean	size sd	λ	σ_ϵ	σ_α	β_0	β_1	γ	π
		True value						
		0.500	1.000	0.500	1.000	1.000	1.000	1.000
		Number of Groups $R = 50$						
10.5	2.29	0.364	0.984	0.595	1.276	1.036	1.458	1.270
		(0.361)	(0.049)	(0.506)	(0.741)	(0.112)	(1.260)	(0.733)
10.5	4.61	0.470	0.995	0.495	1.061	1.008	1.094	1.061
		(0.164)	(0.036)	(0.253)	(0.345)	(0.064)	(0.562)	(0.348)
20.5	2.29	0.032	0.975	0.941	1.935	1.067	2.731	1.934
		(0.959)	(0.053)	(1.075)	(1.920)	(0.143)	(3.556)	(1.934)
20.5	4.61	0.287	0.989	0.669	1.426	1.029	1.728	1.423
		(0.525)	(0.035)	(0.612)	(1.064)	(0.079)	(1.866)	(1.051)
		Number of Groups $R = 100$						
10.5	2.29	0.426	0.991	0.544	1.148	1.022	1.259	1.151
		(0.267)	(0.036)	(0.388)	(0.544)	(0.085)	(0.963)	(0.542)
10.5	4.61	0.483	0.996	0.501	1.036	1.005	1.046	1.033
		(0.117)	(0.025)	(0.178)	(0.244)	(0.047)	(0.411)	(0.243)
20.5	2.29	0.143	0.982	0.842	1.716	1.052	2.342	1.703
		(0.716)	(0.039)	(0.822)	(1.447)	(0.107)	(2.667)	(1.428)
20.5	4.61	0.363	0.993	0.615	1.275	1.020	1.504	1.272
		(0.399)	(0.025)	(0.477)	(0.805)	(0.061)	(1.471)	(0.801)
		Number of Groups $R = 300$						
10.5	2.29	0.468	0.997	0.522	1.062	1.008	1.112	1.063
		(0.165)	(0.021)	(0.235)	(0.331)	(0.052)	(0.594)	(0.331)
10.5	4.61	0.496	0.999	0.500	1.010	1.001	1.014	1.009
		(0.068)	(0.015)	(0.103)	(0.140)	(0.028)	(0.240)	(0.141)
20.5	2.29	0.331	0.991	0.658	1.338	1.025	1.648	1.336
		(0.444)	(0.024)	(0.532)	(0.893)	(0.068)	(1.700)	(0.890)
20.5	4.61	0.444	0.997	0.547	1.111	1.007	1.194	1.114
		(0.253)	(0.015)	(0.306)	(0.505)	(0.039)	(0.951)	(0.509)

[1] Means and standard errors (in the parentheses) of estimates across 1000 replications.

[2] Simulation is based on model (3.22): $y_{ic} = \lambda\bar{y}_{(-i)c} + \beta_0 + \beta_1 x'_{1,ic} + \gamma\bar{x}_{2,(-i)c} + \pi x_{3,c} + \alpha_c + \epsilon_{ic}$, with the ture value of the parameters on the top panel of the table.

[3] Size of group c is drawn from a discrete uniform distribution. The mean and standard deviation of the group size are in column 1 and 2. Summary statistics of the group size is in Table 3.1. Sample is generated by: $x_{1,ic} = x_{2,ic} \sim N(0,1)$, $x_{3,c} \sim N(0,1)$, $\alpha_c \sim N(0, 0.5^2)$, and $\epsilon_{ic} \sim N(0, 0.5^2)$.

Table 3.3: Monte Carlo Results: Uniform Distribution, $x_1, x_2\, i.i.d$

size mean	size sd	λ	σ_ϵ	σ_α	β_0	β_1	γ	π
				True value				
		0.500	1.000	0.500	1.000	1.000	1.000	1.000
				Number of Groups $R = 50$				
10.5	2.29	0.477	0.995	0.492	1.048	0.997	1.016	1.041
		(0.133)	(0.038)	(0.202)	(0.278)	(0.046)	(0.281)	(0.274)
10.5	4.61	0.487	0.997	0.483	1.028	0.999	1.015	1.026
		(0.112)	(0.034)	(0.177)	(0.249)	(0.046)	(0.254)	(0.248)
20.5	2.29	0.451	0.996	0.525	1.101	0.998	1.036	1.097
		(0.256)	(0.027)	(0.335)	(0.516)	(0.034)	(0.397)	(0.534)
20.5	4.61	0.434	0.996	0.540	1.127	0.998	1.040	1.133
		(0.304)	(0.028)	(0.369)	(0.614)	(0.035)	(0.396)	(0.621)
				Number of Groups $R = 100$				
10.5	2.29	0.489	0.998	0.496	1.020	0.999	1.010	1.023
		(0.086)	(0.025)	(0.133)	(0.178)	(0.032)	(0.200)	(0.180)
10.5	4.61	0.490	0.997	0.497	1.022	1.001	1.009	1.019
		(0.072)	(0.024)	(0.114)	(0.159)	(0.032)	(0.172)	(0.158)
20.5	2.29	0.482	0.999	0.505	1.035	0.999	1.016	1.031
		(0.123)	(0.018)	(0.152)	(0.249)	(0.023)	(0.283)	(0.253)
20.5	4.61	0.479	0.999	0.507	1.041	0.999	1.025	1.040
		(0.135)	(0.017)	(0.161)	(0.274)	(0.023)	(0.282)	(0.278)
				Number of Groups $R = 300$				
10.5	2.29	0.497	0.999	0.498	1.006	0.999	1.005	1.007
		(0.046)	(0.014)	(0.071)	(0.101)	(0.019)	(0.107)	(0.099)
10.5	4.61	0.498	0.999	0.498	1.005	0.999	1.002	1.003
		(0.039)	(0.014)	(0.063)	(0.087)	(0.019)	(0.101)	(0.085)
20.5	2.29	0.494	0.999	0.502	1.012	1.000	1.002	1.011
		(0.068)	(0.010)	(0.084)	(0.143)	(0.013)	(0.154)	(0.140)
20.5	4.61	0.490	0.999	0.507	1.020	0.999	1.008	1.022
		(0.062)	(0.009)	(0.079)	(0.128)	(0.013)	(0.156)	(0.131)

[1] Means and standard errors (in the parentheses) of estimates across 1000 replications.

[2] Simulation is based on model (3.22): $y_{ic} = \lambda \bar{y}_{(-i)c} + \beta_0 + \beta_1 x'_{1,ic} + \gamma \bar{x}_{2,(-i)c} + \pi x_{3,c} + \alpha_c + \epsilon_{ic}$, with the ture value of the parameters on the top panel of the table.

[3] Size of group c is drawn from a discrete uniform distribution. The mean and standard deviation of the group size are in column 1 and 2. Summary statistics of the group size is in Table 3.1. Sample is generated by: $x_{1,ic}, x_{2,ic}\, i.i.d \sim N(0,1)$, $x_{3,c} \sim N(0,1)$, $\alpha_c \sim N(0, 0.5^2)$, and $\epsilon_{ic} \sim N(0, 0.5^2)$.

Table 3.4: Monte Carlo Results: Normal Distribution, $x_1 = x_2$

size mean	size sd	λ	σ_ϵ	σ_α	β_0	β_1	γ	π
				True value				
		0.500	0.500	0.500	1.000	1.000	1.000	1.000
				Number of Groups $R = 50$				
9.5	1.76	0.483	0.498	0.479	1.039	1.001	1.013	1.029
		(0.274)	(0.024)	(0.328)	(0.563)	(0.088)	(0.930)	(0.555)
9.5	3.5	0.494	0.498	0.483	1.015	0.997	0.993	1.017
		(0.134)	(0.020)	(0.161)	(0.280)	(0.046)	(0.433)	(0.291)
19.5	1.76	0.204	0.492	0.751	1.602	1.039	1.972	1.597
		(0.799)	(0.024)	(0.830)	(1.638)	(0.107)	(2.583)	(1.642)
19.5	3.5	0.382	0.496	0.587	1.231	1.013	1.335	1.230
		(0.468)	(0.016)	(0.492)	(0.916)	(0.064)	(1.538)	(0.927)
				Number of Groups $R = 100$				
9.5	1.76	0.502	0.499	0.475	0.993	0.999	0.979	0.989
		(0.206)	(0.016)	(0.243)	(0.417)	(0.068)	(0.725)	(0.413)
9.5	3.5	0.503	0.500	0.484	0.995	1.000	0.985	0.996
		(0.091)	(0.013)	(0.111)	(0.189)	(0.034)	(0.307)	(0.190)
19.5	1.76	0.327	0.496	0.649	1.348	1.023	1.570	1.351
		(0.570)	(0.018)	(0.610)	(1.142)	(0.082)	(2.016)	(1.149)
19.5	3.5	0.449	0.499	0.526	1.107	1.006	1.143	1.104
		(0.327)	(0.012)	(0.350)	(0.668)	(0.050)	(1.187)	(0.657)
				Number of Groups $R = 300$				
9.5	1.76	0.499	0.500	0.495	1.002	1.000	1.000	1.002
		(0.124)	(0.010)	(0.145)	(0.251)	(0.041)	(0.444)	(0.256)
9.5	3.5	0.496	0.499	0.501	1.008	1.001	1.018	1.007
		(0.051)	(0.007)	(0.061)	(0.108)	(0.019)	(0.173)	(0.106)
19.5	1.76	0.492	0.500	0.491	1.016	1.000	1.024	1.015
		(0.323)	(0.010)	(0.355)	(0.645)	(0.051)	(1.229)	(0.646)
19.5	3.5	0.492	0.500	0.499	1.015	1.000	1.012	1.016
		(0.200)	(0.007)	(0.214)	(0.399)	(0.031)	(0.736)	(0.406)

[1] Means and standard errors (in the parentheses) of estimates across 300 replications.

[2] Simulation is based on model (3.22): $y_{ic} = \lambda\bar{y}_{(-i)c} + \beta_0 + \beta_1 x'_{1,ic} + \gamma\bar{x}_{2,(-i)c} + \pi x_{3,c} + \alpha_c + \epsilon_{ic}$, with the ture value of the parameters on the top panel of the table.

[3] Size of group c is determined by $size_c = floor[meansize + expander * \Phi^{-1}(v_c)]$, where $v_c\, i.i.d \sim U(0.025, 0.975)$ and Φ is the Cumulative Distribution Function of the standard normal distribution. Summary statistics of the group size is in Table 3.1. Sample is generated by: $x_{1,ic} = x_{2,ic} \sim N(0,1)$, $x_{3,c} \sim N(0,1)$, $\alpha_c \sim N(0, 0.5^2)$, and $\epsilon_{ic} \sim N(0, 0.5^2)$.

Table 3.5: Monte Carlo Results: Normal Distribution, $x_1, x_2\, i.i.d$

size mean	size sd	λ	σ_ϵ	σ_α	β_0	β_1	γ	π
		True value						
		0.500	0.500	0.500	1.000	1.000	1.000	1.000
		Number of Groups $R = 50$						
9.5	1.76	0.483	0.498	0.498	1.037	0.999	1.022	1.028
		(0.094)	(0.019)	(0.114)	(0.204)	(0.027)	(0.169)	(0.213)
9.5	3.5	0.491	0.498	0.490	1.023	0.997	0.996	1.022
		(0.089)	(0.019)	(0.113)	(0.194)	(0.026)	(0.154)	(0.201)
19.5	1.76	0.440	0.498	0.544	1.128	0.998	0.989	1.119
		(0.417)	(0.016)	(0.455)	(0.848)	(0.027)	(0.261)	(0.779)
19.5	3.5	0.444	0.498	0.536	1.116	0.998	1.033	1.110
		(0.202)	(0.013)	(0.220)	(0.423)	(0.019)	(0.234)	(0.438)
		Number of Groups $R = 100$						
9.5	1.76	0.496	0.499	0.495	1.005	1.001	1.005	1.003
		(0.069)	(0.012)	(0.079)	(0.148)	(0.019)	(0.113)	(0.149)
9.5	3.5	0.496	0.499	0.492	1.009	1.001	1.003	1.010
		(0.057)	(0.013)	(0.074)	(0.125)	(0.017)	(0.104)	(0.126)
19.5	1.76	0.479	0.500	0.510	1.042	0.999	1.007	1.042
		(0.099)	(0.009)	(0.113)	(0.203)	(0.012)	(0.186)	(0.210)
19.5	3.5	0.483	0.500	0.506	1.036	1.000	1.001	1.039
		(0.095)	(0.009)	(0.107)	(0.204)	(0.012)	(0.160)	(0.202)
		Number of Groups $R = 300$						
9.5	1.76	0.499	0.500	0.499	1.001	0.999	0.998	1.000
		(0.036)	(0.007)	(0.046)	(0.076)	(0.011)	(0.063)	(0.080)
9.5	3.5	0.498	0.500	0.498	1.003	0.999	1.002	1.002
		(0.034)	(0.007)	(0.046)	(0.077)	(0.011)	(0.060)	(0.076)
19.5	1.76	0.497	0.500	0.501	1.005	0.999	0.998	1.004
		(0.055)	(0.005)	(0.063)	(0.111)	(0.007)	(0.099)	(0.114)
19.5	3.5	0.493	0.500	0.503	1.015	1.000	1.004	1.014
		(0.051)	(0.005)	(0.056)	(0.107)	(0.007)	(0.092)	(0.106)

[1] Means and standard errors (in the parentheses) of estimates across 300 replications.

[2] Simulation is based on model (3.22): $y_{ic} = \lambda\bar{y}_{(-i)c} + \beta_0 + \beta_1 x'_{1,ic} + \gamma\bar{x}_{2,(-i)c} + \pi x_{3,c} + \alpha_c + \epsilon_{ic}$, with the ture value of the parameters on the top panel of the table.

[3] Size of group c is determined by $size_c = floor[meansize + expander * \Phi^{-1}(v_c)]$, where $v_c\, i.i.d \sim U(0.025, 0.975)$ and Φ is the Cumulative Distribution Function of the standard normal distribution. Summary statistics of the group size is in Table 3.1. Sample is generated by: $x_{1,ic}, x_{2,ic}\, i.i.d \sim N(0,1)$, $x_{3,c} \sim N(0,1)$, $\alpha_c \sim N(0, 0.5^2)$, and $\epsilon_{ic} \sim N(0, 0.5^2)$.

Table 3.6: Monte Carlo Results: Case 1, $x_1 = x_2 \sim N(0,1)$

group size	Random effects model							Fixed effects model			
	λ	σ_ϵ	σ_α	β_0	β_1	γ	π	λ	σ_ϵ	β_1	γ
	True value										
	0.500	1.000	0.500	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000
	Number of Groups $R = 50$										
{2,3,...,11}	0.497	0.992	0.455	1.006	0.993	1.000	1.003	0.682	1.019	1.003	0.943
	(0.097)	(0.044)	(0.193)	(0.222)	(0.081)	(0.324)	(0.201)	(0.514)	(0.093)	(0.113)	(0.456)
{16,24,...,88}	0.340	0.996	0.635	1.313	1.002	1.424	1.313	0.612	1.001	0.996	0.844
	(0.476)	(0.017)	(0.503)	(0.945)	(0.027)	(1.423)	(0.950)	(1.288)	(0.029)	(0.037)	(2.045)
	Number of Groups $R = 100$										
{2,3,...,11}	0.490	0.997	0.491	1.021	1.005	1.028	1.022	0.561	1.007	1.011	1.015
	(0.079)	(0.032)	(0.149)	(0.170)	(0.055)	(0.259)	(0.169)	(0.320)	(0.059)	(0.078)	(0.327)
{16,24,...,88}	0.440	0.998	0.557	1.113	1.002	1.160	1.116	0.545	1.000	0.999	0.918
	(0.318)	(0.012)	(0.343)	(0.629)	(0.023)	(1.119)	(0.628)	(0.898)	(0.020)	(0.027)	(1.427)

[1] Means and standard errors (in the parentheses) of estimates across 300 replications.

[2] Simulation is based on the random effect model (3.22): $y_{ic} = \lambda \bar{y}_{(-i)c} + \beta_0 + \beta_1 x'_{1,ic} + \gamma \bar{x}_{2,(-i)c} + \pi \psi_c + \alpha_c + \epsilon_{ic}$, with the ture value of the parameters on the left of top panel.Sample is generated by: $x_{1,ic} = x_{2,ic} \sim N(0,1)$, $\psi_c \sim N(0,1)$, $\alpha_c \sim N(0, 0.5^2)$, and $\epsilon_{ic} \sim N(0,1)$.

[3] Estimation strategy is the Maximum Likelihood estimation for the random group effect model as described in this paper (left);or the Conditional Maximum Likelihood estimation for the fixed group effect model as described in Lee(2007)(right). In the fixed group effect model, all group level variables are absorbed by the group fixed effects, so β_0, π, and σ_α are not estimated.

[4] Group sizes are non-random. They are either $\{2, 3, ..., 11\}$ or $\{16, 24, ..., 88\}$.The number of groups of each size is the same.

Table 3.7: Monte Carlo Results: Case 2, $x_1, x_2\, i.i.d \sim N(0,1)$

group size	Random effects model							Fixed effects model			
	λ	σ_ϵ	σ_α	β_0	β_1	γ	π	λ	σ_ϵ	β_1	γ
					True value						
	0.500	1.000	0.500	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000
					Number of Groups $R = 50$						
{2,3,...,11}	0.493	0.991	0.467	1.014	0.991	1.005	1.015	0.591	1.004	1.006	1.028
	(0.087)	(0.044)	(0.169)	(0.202)	(0.064)	(0.191)	(0.196)	(0.375)	(0.075)	(0.087)	(0.306)
{16,24,...,88}	0.379	0.997	0.604	1.246	0.995	1.033	1.240	0.546	1.000	0.998	0.982
	(0.399)	(0.016)	(0.435)	(0.844)	(0.021)	(0.570)	(0.810)	(1.037)	(0.025)	(0.027)	(0.867)
					Number of Groups $R = 100$						
{2,3,...,11}	0.495	0.998	0.486	1.010	1.000	1.005	1.013	0.550	1.005	1.008	1.025
	(0.055)	(0.031)	(0.110)	(0.129)	(0.041)	(0.127)	(0.133)	(0.253)	(0.048)	(0.058)	(0.204)
{16,24,...,88}	0.469	0.999	0.526	1.059	1.000	1.040	1.062	0.513	1.000	1.000	1.010
	(0.166)	(0.010)	(0.186)	(0.338)	(0.016)	(0.364)	(0.334)	(0.723)	(0.017)	(0.020)	(0.582)

[1] Means and standard errors (in the parentheses) of estimates across 300 replications.

[2] Simulation is based on the random effect model (3.22): $y_{ic} = \lambda \bar{y}_{(-i)c} + \beta_0 + \beta_1 x'_{1,ic} + \gamma \bar{x}_{2,(-i)c} + \pi\psi_c + \alpha_c + \epsilon_{ic}$, with the ture value of the parameters on the left of top panel.Sample is generated by: $x_{1,ic}, x_{2,ic}\, i.i.d \sim N(0,1)$, $\psi_c \sim N(0,1)$, $\alpha_c \sim N(0, 0.5^2)$, and $\epsilon_{ic} \sim N(0,1)$.

[3] Estimation strategy is the Maximum Likelihood estimation for the random group effect model as described in this paper (left);or the Conditional Maximum Likelihood estimation for the fixed group effect model as described in Lee(2007)(right). In the fixed group effect model, all group level variables are absorbed by the group fixed effects, so β_0, π, and σ_α are not estimated.

[4] Group sizes are non-random. They are either $\{2,3,...,11\}$ or $\{16,24,...,88\}$.The number of groups of each size is the same.

Chapter 4: Do Peers Matter for Children's Academic Performance? Evidence from Project STAR

4.1 Introduction

Early educational outcomes have an enduring impact on lifetime educational and economic success (Heckman, 2006). Studies have shown that children with higher academic achievement are more likely to attend college and receive higher earnings (e.g., Chetty et al., 2011). Numerous policies are designed to improve early educational outcomes, especially for disadvantaged children. It is therefore of crucial interest to understand what factors impact children's educational achievement and how. In this chapter, I investigate how children's academic performance is affected by their peers in class as well as other factors such as class size and teacher quality.

There is widespread belief that peers matter for children's academic performance. According to theories in child psychology, the peer group is one of the most important contexts shaping children's behaviors and skills (Rubin et al., 2007).[1] Children develop their cognitive skills by learning from adults and more knowledge-

[1] The nurture assumption theory, for example, argues that peer environment is more important than parents for children's development.

able peers.[2] In the field of education policy, the groundbreaking Coleman (1956) report ascribes peer quality as a more valuable educational input than teacher quality and school facilities. Following his report, a large number of studies have examined peer effects in education, though with inconclusive results because of challenging identification issues.

Peer effects can work through pre-determined characteristics such as sex and race, therefore referred to as exogenous (peer) effects. For example, a higher share of girls in a class may increase the academic performance of everyone in the class because girls are less disruptive than boys (Lavy and Schlosser, 2011). Peer effects can also work through endogenously determined peer behaviors and outcomes, such as knowledge, effort, and motivation (Fruehwirth, 2013), known as endogenous (peer) effects. For example, Conley et al. (2015) find that high school students' study time is impacted by that of their peers. In- and out-of-class discussions lead to knowledge spillover among classmates. Although these standard mechanisms underlying peer effects may be invalid for children, children's imitation of each other's behavior and their compliance with the group norm can also generate peer effects. In practice, these endogenous characteristics are usually unobservable to researchers and proxied by test scores.

Identifying endogenous effects is notoriously challenging (Angrist, 2014; Moffitt, 2001). Manski (1993) highlights the "reflection problem" in identifying peer effects. The idea is that with endogenous effects, contemporaneous achievement of

[2]This is the argument of Lev Vygotskyl's "socialculture theory," one of the most recognized child development theories.

peers is simultaneously determined. This simultaneous equation feature of the peer effect model poses challenges for identification. Most studies identify the reduced form of peer effect models but are unable to disentangle the endogenous effects and exogenous effects in it. The few studies that achieve identification all find significant positive endogenous effects (Boucher et al., 2014; Graham, 2008; Lin, 2010).

Distinguishing endogenous effects and exogenous effects is important. Peer effects are informative to regrouping policies like busing and school choice programs. The reduced form estimates for peer effects are sufficient for regrouping policies in certain circumstances, e.g., the school choice program studied by Altonji et al. (2015). However, Fruehwirth (2013) points out that estimates of endogenous peer effects are necessary under reasonable assumptions. For example, when peer effects are heterogeneous and when there is matching between teachers and children. Second, endogenous peer effects generate social multiplier effect and confound the evaluation of education policies (Imbens and Wooldridge, 2009). If a policy applies to the whole group, e.g., assigning an effective teacher to class, the gross impact constitutes both the net teacher effect and the social multiplier effect. If the policy applies to part of the group, e.g., giving half of the class special instruction, then comparing the outcome of treated and untreated individuals underestimate the policy effect because those not receiving the instruction also benefit indirectly.

In this chapter, I estimate both endogenous and exogenous peer effects on academic performance among kindergarten students. I use data from Project STAR, a class size reduction experiment in Tennessee. Project STAR randomly assigned both students and teachers to classes of different sizes. The exogenous group formation facilitates the estimation of peer effects because sorting into the same groups based on similar characteristics can lead to spurious peer effects.

My model controls among other things for random group effects. I estimate the model with maximum likelihood (ML) method. A closer look at the first order conditions reveals that identification is based on second moments of within-class variance and between-class variance. What's more, identification comes from the exogenous and systematic variation of class size, a key feature of Project STAR. The model is developed and estimated under the framework of spatial econometrics.

In contrast to previous studies, this study finds insignificant endogenous peer effects and exogenous peer effects in kindergarteners' test scores. The results are robust across different specifications. The results urge caution against overstating the importance of peer effects and cast doubts on regrouping policies.

The difference from previous studies arises primarily from a relaxation of functional form restrictions. Using the same data as in this study, Graham (2008) finds significant endogenous peer effects with a model that controls for school fixed effects and class type fixed effects separately. Graham's specification relies on the assumption that the treatment effect of class size reduction is independent of school characteristics. Regression analysis in this study provides evidence that this assumption does not hold. A more flexible specification that controls for school by class type fixed effects, in combination with random group effects, is used instead and supported by specification tests.

While focusing on peer effects, this study also sheds light on the effectiveness of teacher quality and class size. The results also confirm with previous studies that class size reduction improves children's test scores in STAR schools.

The rest of the chapter is organized as follows. Section 4.2 describes Project STAR and summary statistics of the data. Section 4.3 outlines the model specification and identification strategy. Section 4.4 discusses the results. Section 4.5 concludes.

4.2 Data

This study uses data from Project STAR (student-teacher achievement ratio), a randomized experiment that assigned both students and teachers into classes of different sizes.[3]The goal of Project STAR is to study the effect of small classes on students' development. The experiment began in the 1985-1986 school year and ended in 1989. It followed a single cohort of students for four years, from kindergarten to third grade. Upon entry into the project, students and teachers were randomly assigned to small classes (13-17 students), regular classes (22-26 students) and regular classes with a full-time teacher's aid. Academic and nonacademic measurements of the students were taken at the end of each school year. Boyd-Zaharias et al. (2007) and Mosteller (1995) offer detailed descriptions of project STAR. Here I briefly describe the process of the experiment.

[3]The Project STAR data was originally posted on the project webiste (http://www.heros-inc.org/star.htm), which is no longer accessible. The data this chapter uses is downloaded from Harvard Dataverse (https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/10766).

4.2.1 Project STAR

Project STAR was implemented in the following steps.

In the preparation period, seventy-nine schools were selected to participate in the experiment. All elementary schools in Tennessee were invited, but only schools with enough students to have at least one class for each of the three class types were eligible.[4] Among the 180 schools expressing interest, 100 were eligible and 79 were selected. The selection of schools was nonrandom and over-sampled minority schools. Four schools withdrew in the process of the experiment.

In the school year of 1985, students entering kindergarten of the experiment schools were randomly assigned to small classes, regular classes and regular classes with aid within school.[5] Students were supposed to stay in the class type they were initially assigned to throughout the experiment. Kindergarten teachers were then randomly assigned to the three class types too.

In 1986, when the cohort of students entered first grade, students initially assigned to regular class and regular class with aid were reassigned randomly to these two class types. Students assigned to small classes remained in small classes. Meanwhile, a large influx of new students joined the program because kindergarten

[4]The minimum number of students is 57: 13 for small class, 22 for regular class and 22 for regular class with aid.

[5]No report describes the exact assigning process, a point criticized by Hanushek (2003). In the footnote, Krueger (1999) briefly shows that schools assign students into class types by picking every k-th student from the alphabeticalized enrollment list. The number k is calculated by some algorithm depending on the number of students and classes of each type. The initial point is generated randomly.

was not mandatory in Tennessee at the time of the experiment. They were randomly assigned to the three class types.

The experiment continued in grade two and grade three. No major change occurred. The project ended in grade four. All students returned to regular-sized classes. But students were followed through high school. Their participation ratings, achievement test score, high school GPA, high school graduation status, SAT/ACT score were recorded.

Students took the Stanford Achievement Tests (SAT) in spring each year. SAT is a nationally normed standardized achievement test, covering reading, math, spelling, and listening. The test was monitored by trained substitute teachers (Folger, 1989) and no special instruction was allowed before the tests. Other academic and nonacademic measurement were also taken each year.

In general, students enrolled in the class types they were assigned to. Krueger (1999) examines a subsample and finds that only 0.3% of the students enroll in a class different from the type they were assigned to. Students were assigned to the same class types if they migrated from one STAR school to another STAR school. Students left the sample if they migrated to a non-STAR school, repeated a grade or jumped a grade. New entrants to the participating schools were assigned randomly to the three types of schools as well.

In total, there are 11601 students involved in the experiment. Due to migration and grade repetition, only 26.57% of the students are in the experiment for all four years. In each grade, there are around 6000 students and 300 classes.

4.2.2 Randomness of Class Assignment

One merit of using project STAR data is the randomization of class assignment. Assignment of students and teachers into classes were carefully conducted and audited by Project STAR staff, leaving little room for parents or schools to manipulate (Boyd-Zaharias et al., 2007; Krueger, 1999). The randomness of class assignment in Project STAR has been examined carefully by a large number of studies, e.g., Chetty et al. (2011) and Krueger (1999).

Most studies with STAR data stress that students and teachers were assigned randomly to class types within schools, possibly because the goal of the project and their research interest is to analyze the impact of small classes. Studies whose primary interest is not the impact of class size generally assume that assignment into classes is random within school (Chetty et al., 2011; Graham, 2008).Chetty et al. (2011) get confirmation from original STAR designer that students and teachers were indeed randomly assigned into classrooms rather than class types. They also provide statistical evidence for the statement.[6]

While the initial assignment is carried out carefully, there are possible deviations from randomization in higher grades. The randomness of class assignment suffers from non-random switching and non-random attrition in higher grades. Although students were supposed to stay in the class types they were initially assigned to through grade three, around 10 percent of the students switched class types be-

[6] Chetty et al. (2011) are able to provide more pre-determined variables such as family background because they link the STAR data with the tax return information of the students.

tween grades "due to behavioral problems and parents complaints" (Krueger, 1999). Attrition happens if students were retained, jumped grades, or moved out of the experimental schools. Fewer students in small classes were retained than regular/aid classes (Finn et al., 2001). Kindergarten data is not affected by the non-random switching or non-random attrition and therefore more reliable. I focus my analysis on kindergarten students and also provide results for higher grades as a robustness check.

In this chapter, I assume that assignment of teachers and students into classes was random within the school in kindergarten. However, I control for school by class-type fixed effect as school effect may interact with class-type effect. For example, black students benefitted more from attending small classes than white students (Schanzenbach, 2006). Fifteen out of the 79 schools have all black students in kindergarten, while 15 schools have all white students in kindergarten. Class size reduction may impact students in the all-black schools differently from all-white schools. A likelihood ratio test supports using school by class type fixed effects.

4.2.3 Summary Statistics

The kindergarten sample includes 79 schools, 325 classes, and 6325 students. There are 127 small classes, 99 regular classes and 99 regular classes with aid. Tables 4.1 and 4.2 show summary statistics of student characteristics and class characteristics separately.

Over 98% of the minority students are black, so I use an indicator of black to control for race. I use receiving free lunch as an indicator of low-income family background. By this measure, 48.4% of the students are low-income (poor). Girls represent 48.6% of the sample. Around 4% of the students were in kindergarten last school year and repeating it in 1985. Age on September 1st, 1985 for each student is calculated using their date of birth. The average age for kindergarteners is 5.4 and 86.6% of the students' age is between 5 and 6.

The math score is between 288 and 626, with a mean of 485.38 and a standard deviation of 47.79. The reading score is between 315 and 627, with a mean of 436.725 and a standard deviation of 31.71.with For easier comparison of the estimates across exams and grades, I normalize the test scores with the mean and standard error of the test scores of the grade. Possible consequences of the normalization are discussed in the model specification part.

Class size ranges from 12 to 28, with the mean being 19.46. The class level summary statistics also show that 16% of the classes have black teachers and 35% percent of the kindergarten teachers have master's degree or higher. I do not control for teacher's gender because all teachers in Kindergarten are female. The average within-class standard deviation of the normalized math score is 0.82. The number is 0.80 for the normalized reading score. The standard deviation of the class mean is 0.58 and 0.57 respectively for the normalized math and reading score.

4.3 Model Specification and Estimation Strategy

4.3.1 Model Specification

Suppose that students interact equally within the class and that there is no interaction across classes. The main model of this study is

$$y_{ic,ps} = \beta_0 + \lambda \bar{y}_{(-i)c,ps} + x'_{ic,ps}\beta + \bar{x}'_{(-i)c,ps}\gamma + \psi'_{c,ps}\pi + f'_{c,ps}\phi + \alpha_{c,ps} + \epsilon_{ic,ps}, \quad (4.1)$$

where $y_{ic,ps}$ is the normalized Stanford Achievement Tests (SAT) math or reading score of the student i in class c. Class c is a type p class in school s, with $p = 1$ for small class, $p = 2$ for regular class, $p = 3$ for regular class with aid, $s = 1, ..., S$. For simplicity of notation, I drop the subscript for class type p and for school s from now on. The index for class c ranges from 1 to R, where R is the total number of classes in the sample. Class c has n_c students, indexed by $i = 1, ..., n_c$. The total sample size is $N = \sum_{c=1}^{R} n_c$. A cleaner form of the model is

$$y_{ic} = \beta_0 + \lambda \bar{y}_{(-i)c} + x'_{ic}\beta + \bar{x}'_{(-i)c}\gamma + \psi'_c\pi + f'_c\phi + \alpha_c + \epsilon_{ic}, \quad (4.2)$$

where $x_{ic} = (x_{1,ic}, x_{2,ic}, ..., x_{k_x,ic})'$ is a k_x dimensional vector of personal characteristics, including gender, race, poor (receiving free lunch), age, and repeat (retained from last year); ψ_c is a k_ψ dimensional vector of observed class characteristics, including teacher's education and race. α_c is the unobserved class effect, ϵ_{ic} is the disturbance term. The indicator vector $f_c = (d_{11,c}, d_{21,c}, d_{31,c}, ..., d_{1S,c}, d_{2S,c}, d_{3S,c})'$, with $d_{ml,c} = 1$ if class c is a type m class in school l and zero if not. The parameter for school by class type fixed effects is $\phi = (\phi_{11}, \phi_{21}, \phi_{31}, ..., \phi_{1S}, \phi_{2S}, \phi_{3S})'$, with ϕ_{ps}

being the fixed effect of class type p in school s. Therefore, $f_c'\phi$ is the school by class type fixed effect of class c.

A restricted version of this model can control for school by class type fixed effects separately. To see this, let $\bar{\phi}_p$ be the class type fixed effect for class type p, $p = 1, 2, 3$. Let ϕ_s^* be the school fixed effect for school s, $s = 1, ..., S$. Then

$$\phi_{ps} = \bar{\phi}_p + \phi_s^* + \mu_{ps}.$$

By imposing a restriction that $\mu_{ps} = 0$, the model controls for school fixed effects and class type fixed effects separately. Recall that the kindergarten sample of Project STAR includes 79 schools, each has three different class types. So the vector of school by class type fixed effects ϕ has 237 elements. Meanwhile, the sample includes 325 classes. If instead the model controls for school fixed effects and class type fixed effects separately, the vector of school fixed effects includes 79 elements.

The term $\bar{y}_{(-i)c} = \frac{1}{n_c-1}\sum_{j\neq i}^{n_c} y_{jc}$ is the average test score of peers. This is the mean score of the whole class leaving out oneself, therefore referred to as leave-out mean. Likewise , $\bar{x}'_{(-i)c} = \frac{1}{n_c-1}\sum_{j\neq i}^{n_c} x'_{jc}$ is the leave-out mean of personal characteristics.[7] The parameter λ captures the endogenous effect and γ captures the exogenous effect.

The dependent variable y is the normalized test score rather than the raw score. The normalization scales the individual test score and the mean peer score

[7]For simplicity, $\bar{x}_{(-i)c}$ is simply the leave-out mean of x_{ic}. The two are based on the same set of variables. Although it is possible to allow x_{ic} and $\bar{x}_{(-i)c}$ to include different variables, it is hard to justify that some characteristics only affect oneself but not peers. Otherwise, reflection problem is easily solved. To see this, suppose the $k-th$ individual characteristics x_k only has individual effect, i.e., $\beta_k \neq 0$ but $\gamma_k = 0$, then with reduced form of the peer effects one can identify both endogenous peer effect and exogenous peer effect. Conceptually, it is even harder to argue that some individual characteristics only affect peers but not oneself, i.e., $\beta_k = 0$ but $\gamma_k \neq 0$. Therefore, the specification of the model is plausible.

by the same factor. Hence it does not change the magnitude of the endogenous effect λ. The normalization scales exogenous effect γ and the coefficient of other control variables like β and π down by the standard deviation of y_{ic}.

A compressed form of the model is

$$y_{ic} = \lambda \bar{y}_{(-i)c} + z_{ic}'\delta + \alpha_c + \epsilon_{ic}, \tag{4.3}$$

where $z_{ic} = (1, x_{ic}', \bar{x}_{(-i)c}', \psi_c', f_c')'$ is the vector of all exogenous variables, $\delta = (\beta_0, \beta', \gamma', \pi', \phi')'$ is the vector of corresponding coefficients.

The model can be rewritten as a Cliff-Ord type spatial model. Let W_c be the weights matrix of class c, whose off-diagonal elements are $\frac{1}{n_c-1}$ and the diagonal elements are 0. $Y_c = (y_{1c}, ..., y_{n_c c})'$, $Z_c = (z_{1c}', ..., z_{n_c c}')'$, $\epsilon_c = (\epsilon_{1c}, ..., \epsilon_{n_c c})'$, $\iota_c = (1, ..., 1)'$ is the n_c dimensional vector of ones. The model for class c in matrix form is

$$Y_c = \lambda W_c Y_c + Z_c \delta + U_c, \tag{4.4}$$

$$U_c = \alpha_c \iota_c + \epsilon_c. \tag{4.5}$$

I identify the model with a quasi-maximum likelihood estimation (QMLE) strategy. Details of the assumptions, identification strategy and properties of the estimator are in Chapter 3. Peer effects are identified from variation in group size. The estimator is consistent if there are at least two different group sizes in the limit.

4.3.2 Estimation Strategy

Chapter 3 describes theoretical aspects of the QMLE. This section gives more intuitive explanations of the estimation strategy. I will demonstrate that the model is identified based on the within-class variance and between-class variance. It utilizes the exogenous variation of class size for identification.

For simplicity of illustration, I drop the covariates z except for the school by class type fixed effects ϕ_{ps}. Since both teachers and students were randomly assigned into classes within school, student and teacher characteristics are independent of class assignment conditional on school fixed effects. Excluding personal and teacher characteristics therefore makes little difference to the estimates for λ. A simplified model therefore is

$$y_{ic} = \lambda \bar{y}_{(-i)c} + \alpha_c + \phi_{ps} + \epsilon_{ic}. \tag{4.6}$$

Under Assumptions 1-3, the within-variance for class c is

$$E\frac{1}{(n_c - 1)}\sum_{i=1}^{n_c}[(y_{ic} - \bar{y}_c)]^2 = (1 - \frac{\lambda}{n_c - 1 + \lambda})^2\sigma_\epsilon^2. \tag{4.7}$$

If $\lambda = 0$, the right hand side becomes σ_ϵ^2. Without peer effects, all classes have the same within-class variance. If $\lambda > 0$, then the within-variance is $(1-\frac{\lambda}{n_c-1+\lambda})^2\sigma_\epsilon^2 < \sigma_\epsilon^2$ and increases with class size. The idea is that positive endogenous effects reduce within-class variance. But the reduction is smaller when class size is larger because with larger class, each person has a smaller impact on their peers. Therefore, as long as there is enough variation in class size, λ can be identified. This accord with the conclusion in Kelejian et al. (2006) that identification of spatial model with blocks of equal elements requires group size variation, which is also confirmed by Lee (2007) in the fixed group effects model.

Project STAR is a class size reduction experiment and hence generates enough class size variation for identification. Class level summary statistics in Table 4.2 show that the largest class has 28 students and the smallest class has 12 students. The distribution of class size in Figure 1 also shows that class size varies greatly in the sample. Class size also varies within small classes and regular classes.

If positive endogenous effects exist, within-class variance and class size are positively correlated. However, Figure 2 and Figure 3 show little correlation between within-variance and class size. This is an indication that there are no endogenous peer effects.

The between-class variance is

$$E(\bar{y}_c - \frac{\phi_{ps}}{1-\lambda})^2 = \frac{1}{(1-\lambda)^2}(\sigma_\alpha^2 + \frac{\sigma_\epsilon^2}{n_c}). \tag{4.8}$$

Note that $\frac{1}{(1-\lambda)^2}\sigma_\alpha^2$ is a constant term and does not change with class size. If $\lambda = 0$, then $E(\bar{y}_c - \frac{\phi_{ps}}{1-\lambda})^2 - \frac{\sigma_\alpha^2}{(1-\lambda)^2} = \frac{\sigma_\epsilon^2}{n_c}$. Without endogenous peer effects, the between-variance can be explained by the individual heterogeneity. If $\lambda > 0$, then $E(\bar{y}_c - \frac{\phi_{ps}}{1-\lambda})^2 - \frac{\sigma_\alpha^2}{(1-\lambda)^2} > \frac{\sigma_\epsilon^2}{n_c}$. With positive peer effects, there is excess variance. This is the idea that underlines Graham (2008).

My estimator is in fact based on both (4.7) and (4.8). To see this, define

$$\xi_c^w = \frac{1}{(n_c-1)}\sum_{i=1}^{n_c}[(1+\frac{\lambda}{n_c-1})(y_{ic}-\bar{y}_c)]^2 - \sigma_{\epsilon c}^2,$$
$$\xi_c^b = (\bar{y}_c - \frac{\phi_{ps}}{1-\lambda})^2 - \frac{1}{(1-\lambda)^2}(\sigma_\alpha^2 + \frac{\sigma_{\epsilon c}^2}{n_c}),$$

then $E\xi_c^w = 0$ and $E\xi_c^b = 0$ are two moment conditions that underly my estimator.

In most spatial models, calculation of $(I - \lambda W)^{-1}$ and $|I - \lambda W|$ is demanding. But the special form of weights matrix and variance-covariance matrix allows for explicit calculation of these qualities.[8] The log likelihood function can be factored

[8]See Appendix B

into a component of within-class variance and a component between class variance,

$$
\begin{aligned}
lnL = &-\frac{N}{2}ln(2\pi) + Rln|1-\lambda| + \sum_{c=1}^{R}[(n_c-1)ln(\frac{n_c-1+\lambda)}{n_c-1}] \\
&-\sum_{c=1}^{R}\frac{n_c-1}{2}ln(\sigma_\epsilon^2) - \frac{1}{2}\sum_{c=1}^{R}ln(\sigma_\epsilon^2+n_c\sigma_\alpha^2) \qquad (4.9)\\
&-\frac{1}{2}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\frac{1}{\sigma_\epsilon^2}[(\frac{n_c-1+\lambda}{n_c-1}(y_{ic}-\bar{y}_c)]^2 - \sum_{c=1}^{R}\frac{[(1-\lambda)\bar{y}_c-\phi_{ps}]^2}{2(\sigma_\alpha^2+\frac{1}{n_c}\sigma_\epsilon^2)}.
\end{aligned}
$$

The first order conditions for maximum likelihood are:

$$
\frac{\partial lnL}{\partial\lambda} = \sum_{c=1}^{R}\frac{(1-\lambda)}{(\sigma_\alpha^2+\sigma_\epsilon^2/n_c)}\xi_c^b - \sum_{c=1}^{R}\frac{n_c-1}{(n_c-1+\lambda)\sigma_\epsilon^2}\xi_c^w = 0, \qquad (4.10)
$$

$$
\frac{\partial lnL}{\partial\sigma_\epsilon^2} = \sum_{c=1}^{R}\frac{1}{2(n_c-1)\sigma_\epsilon^4}\xi_c^w - \sum_{c=1}^{R}\frac{(1-\lambda)^2}{2n_c(\sigma_\epsilon^2/n+\sigma_\alpha^2)^2}\xi_c^b = 0, \qquad (4.11)
$$

$$
\frac{\partial lnL}{\partial\sigma_\alpha^2} = \sum_{c=1}^{R}\frac{(1-\lambda)^2}{2(\sigma_{\epsilon c}^2/n+\sigma_\alpha^2)^2}\xi_c^b = 0. \qquad (4.12)
$$

In this sense, estimation is based on finding the λ that best fits $E(\xi_c^w) = 0$ and $E(\xi_c^b) = 0$. There are several things to note here.

First, the moment conditions (4.7) and (4.8) hold regardless of the distribution of ϵ_{ic} and α_c. Therefore, while the log likelihood function is derived under the assumption of Gaussian distributions, this assumption is not necessary for identification. The maximum likelihood estimators are efficient if the true distribution is Gaussian and consistent even if it is not.

Second, the moment conditions (4.7) and (4.8) can uniquely identify λ with enough variation in class size as long as n_c does not go to infinity.[9]

Third, the moment conditions also show that the model is identifiable even if it incorporates class fixed effects for some classes. My main specification includes

[9] Combining (4.7) and (4.8),

$$
E\bar{y}_c{}^2 = \frac{\sigma_\alpha^2}{(1-\lambda)^2} + \frac{(n_c-1+\lambda)^2}{(n_c-1)^3(1-\lambda)^2}E\sum_{i=1}^{n_c}(y_{ic}-\bar{y}_c)^2. \qquad (4.13)
$$

school by class type fixed effects. For classes which are the only one of a particular type within a school, the model essentially adds a class fixed effect for it. However, adding class fixed effects does not change the within-class variance in (4.7). So the model is identified even if all class fixed effects are added.

Consider a case when there are only two classes, indexed by c_1 and c_2. Suppose the two classes are different in size, $n_{c_1} \neq n_{c_2}$ and the class fixed effects are included for both. Then λ is still identified by

$$\frac{\frac{1}{(n_{c_1}-1)}\sum_{i=1}^{n_{c_1}}[(y_{ic_1}-\bar{y}_{c_1})]^2}{\frac{1}{(n_{c_2}-1)}\sum_{j=1}^{n_{c_2}}[(y_{jc_2}-\bar{y}_{c_2})]^2} = \frac{(n_{c_1}-1)^2}{(n_{c_1}-1+\lambda)^2}\frac{(n_{c_2}-1+\lambda)^2}{(n_{c_2}-1)^2}. \tag{4.15}$$

Meanwhile, among the 325 classes in my sample, 153 are the only one of a particular type within the school. For these classes, school by class type fixed effect is equivalent to class fixed effect. Therefore, between-class variance conditional on school by class type fixed effect is 0 for them. But still, the between-class variance from the rest 172 classes gives additional moment conditions. In this sense, using school by class type fixed effects can control heterogeneous responses to class size reduction across schools while offering additional moment conditions and increasing efficiency of estimation.

If the leave-out-mean $\bar{y}_{(-i)c}$ in model (4.1) is replaced with the full mean $\bar{y}_c = \frac{1}{n_c}\sum_{i=1}^{n_c} y_{ic}$, as Graham (2008) does,

$$E\bar{y}_c{}^2 = \frac{\sigma_\alpha^2}{(1-\lambda)^2} + \frac{1}{(1-\lambda)^2}\frac{1}{(n_c-1)}E\sum_{i=1}^{n_c}(y_{ic}-\bar{y}_c)^2. \tag{4.14}$$

Equation (4.14) is the key equation underlying Graham(2008)'s estimation strategy. $\frac{1}{(1-\lambda)^2}$ is identified by replacing both $E\bar{y}_c{}^2$ and $\frac{1}{(n_c-1)}E\sum_{i=1}^{n_c}(y_{ic}-\bar{y}_c)^2$ with their sample analog.

4.3.3 Interpretation of the Coefficients

With the closed form of $(I - \lambda W_c)^{-1}$ in equation (B.4), the reduced form of model (4.1) is

$$y_{ic} = \frac{\beta_0}{1-\lambda} + x'_{ic}\frac{(n_c - 1)\beta - \gamma}{n_c - 1 + \lambda} + \bar{x}'_c\frac{n_c(\lambda\beta + \gamma)}{(1-\lambda)(n_c - 1 + \lambda)} \tag{4.16}$$
$$+ \psi'_c\frac{\pi}{1-\lambda} + f'_c\frac{\phi}{1-\lambda} + \frac{\alpha_c}{1-\lambda} + \frac{\lambda n_c}{(1-\lambda)(n_c - 1 + \lambda)}\bar{\epsilon}_c + \frac{n_c - 1}{n_c - 1 + \lambda}\epsilon_{ic},$$

where $\bar{x}_c = \frac{1}{n_c}\sum_{i=1}^{n_c} x_{ic}$, $\bar{\epsilon}_c = \frac{1}{n_c}\sum_{i=1}^{n_c}\epsilon_{ic}$ are the full mean of personal characteristics and disturbance term respectively.

The total direct impact[10] of one-unit increase in x_{ic} on y_{ic}, holding others constant, is therefore

$$\frac{(n_c - 1)\beta - \gamma}{n_c - 1 + \lambda} + \frac{(\lambda\beta + \gamma)}{(1-\lambda)(n_c - 1 + \lambda)} = \frac{[(n_c - 1) + \lambda(2 - n_c)]\beta + \lambda\gamma}{(1-\lambda)(n_c - 1 + \lambda)}. \tag{4.17}$$

There are multiple channels through which a student's characteristics x_{ic} can affect his or her own score y_{ic}. First is the direct impact, as captured by coefficient β in equation (4.1). The second effect is through affecting the peer score $\bar{y}_{(-i)c}$, which in turn affects student's own score y_{ic}. For the second channel, x_{ic} can affect peer scores either by affecting the own score y_{ic} and hence $\bar{y}_{(-i)c}$ through the endogenous effects. Moreover, x_{ic} can affect $\bar{x}_{(-j)c}$ and hence y_{jc} through exogenous peer effects, where $j \neq i$. The gross effect of x_{ic} is therefore the combination of the individual effect β, the endogenous peer effect λ and the exogenous peer effect γ.

[10] "Total direct impact" is a terminology used in LeSage and Pace (2009) and Drukker et al. (2013). It refers to the marginal effect of a change in exogenous variable on the dependent variable.

The total direct impact of a one-unit increase in x_{jc} on y_{ic}, holding everything else constant, is

$$\frac{\lambda\beta + \gamma}{(1-\lambda)(n_c - 1 + \lambda)}.$$

Peer j's characteristics x_j, where $j \neq i$, can affect student i's score y_{ic} directly through exogenous peer effects γ, or indirectly through affecting individual j's own score y_{jc} and hence y_{ic} through the endogenous peer effect.

The impact of teacher characteristics ψ_c is $\frac{\pi}{1-\lambda}$. The impact of school by class type fixed effects is $\frac{\phi}{1-\lambda}$. Both constitute the net effects and the magnification of the endogenous peer effects. The coefficient λ for the mean peer score $\bar{y}_{(-i)c}$ in model (4.1) has no easy interpretation (Fruehwirth, 2013). It is a structural rather than causal parameter. One caveat is to draw causal inference from it. An analog is the coefficient of the endogenous variable in a simultaneous equation system. The coefficient λ reflects the magnitude in which a student's score is affected by the behavior and outcome of peers. In this study, I will simply refer to it as endogenous effect.

4.3.4 Linear-in-means Model and Comparison with Manski (1993)

This chapter focuses on a linear-in-means model, in which peer effects work through peers' mean test score and mean characteristics. There are theoretical models suggesting alternative specifications. For example, the spotlight model suggests that the best students in the class matter. The bad apple model suggests that the low-achieving students in the class matter. These alternative models are not the focus of this study. The linear-in-means model is the most popular among all peer effect models. It is also a good starting point for further exploring other specifications. The linear-in-means model can be motivated by "students conform to mean behavior" of the group (Blume et al., 2015). Also, without specific information on group structure, the mean peer achievement is a good proxy for the average academic ability of the people an individual is actually interacting with.

Model (4.1) is different from Manski's model. The Manski model is

$$y_{ic} = \beta_0 + \lambda E(y_c|c) + x_{ic}'\beta + E(x_c|c)'\gamma + \psi_c'\pi + v_{ic}. \tag{4.18}$$

In my model in equation (4.1), endogenous peer effects work through the actual mean outcome of peers $\bar{y}_{(-i)c}$. In Manski (1993), peer effects work through $E(y_{ic}|c)$, the "(population) mean outcome of the reference group". The actual means specification applies to a setting where people interact in small groups and each member knows each other (Manski, 1993). Moreover, identification of endogenous effects is difficult with regression methods based on population mean, e.g., ordinary least squares (OLS) or instrumental variable (IV) method (Angrist, 2014). The actual mean of one's specific peer group reflects properties of the social network structure. Identification of endogenous peer effects is possible through social networks (Bramoullé et al., 2009).

4.3.5 Comparison with Graham (2008)

The main model in equation (4.1) is similar to the model in Graham (2008).[11] Graham's model is equivalent to

$$y_{ic} = \lambda \bar{y}_c + z_{ic}^{*\prime}\delta^* + \alpha_c + \epsilon_{ic}, \tag{4.19}$$

where $\bar{y}_c = \frac{1}{n_c}\sum_{c=1}^{n_c} y_{ic}$ is the mean score of the whole class, including individual i, therefore referred to as full mean. The term $z_{ic}^{*}\delta$ controls for school fixed effects and class type fixed effects. The vector of exogenous variables

$$z_{ic} = (1, reg, aid, sch_1, sch_2, ..., sch_{S-1}),$$

with $reg = 1$ if class c is a regular class and zero if not, $aid = 1$ if class c is a regular class with aid and zero if not, $sch_s = 1$ if class c is in school s and zero if not, δ^* is the corresponding parameter vector. The class effect α_c and the disturbance ϵ_{ic} are the same as in (4.1). Both Assumption 1 and Assumption 2 are maintained in Graham (2008).

Graham's model differs from the main model in this study in two major ways. First of all, Graham's model uses a full-mean specification, and endogenous peer effects work through mean class score $\bar{y}_c$. Model (4.1) uses a leave-out-mean specification, and endogenous peer effects work through mean peer score $\bar{y}_{(-i)c}$. Putting y_{ic} on both sides of the equation is conceptually inappropriate as it makes the dependent variable also an independent variable, though it has a negligible impact on the estimates. Second, in Graham (2008), the vector of exogenous variables z_{ic}^{*}

[11]For a detailed discussion of Graham's model and estimation strategy, see Appendix H.

only includes school fixed effects and class type fixed effects separately.[12] The main model in (4.1) includes school by class type fixed effects f_c as well as personal characteristics x_{ic}, mean peer characteristics $\bar{x}_{(-i)c}$ and teacher characteristics ψ_c. The inclusion of personal, peer and teacher characteristics does not have a large impact on the estimates of λ because of the random assignment of teachers and students within the school.

Graham's model can be estimated with maximum likelihood method under the framework of spatial econometrics, in a similar manner as model (4.1) is estimated. As a matter of fact, Graham's conditional variance method utilize the first order conditions of the maximum likelihood estimation, as shown in Appendix H. In this respect, the estimation strategy in this study is a generalization of Graham's method. One limitation of Graham's method is that it only identifies $1/(1-\lambda)^2$ and is therefore unable to identify the sign of λ.

4.4 Results and Discussion

4.4.1 Main Results

Results of the main model in (4.1) are in Columns 1 and 2 of Table 4.3. Column 1 is for math score, and Column 2 is for reading score. The top panel shows the estimates for the endogenous peer effect λ. The estimates are 0.040 for the math score and 0.035 for the reading score. The estimates are small in magnitude and insignificant.

[12] Graham does not estimate δ but first regress y on z and then use the residual for estimation.

The second panel shows the estimates for the exogenous peer effects γ, the impact of average peer characteristics on individual scores. For example, if the share of girls in peers increases by one percentage point, the net direct effect is that individual math scores increase by 0.00172 standard deviation, which amount to 0.082 point while the average math score is 485.3 and the maximum score is 626. If the average age of peers increases by one year, the net direct effect is a drop of 0.042 standard deviation in individual math score, which amounts to around 2 points. But none of the exogenous peer effects are significant.

The third panel shows the impact of own characteristics on the individual score β and the impact of class characteristics ψ. Academic achievement is higher for older students and girls, and lower for black students, students receiving free lunch and students repeating a grade. All estimates are significant at the 1% level. If the age of a student increases by one year, the net direct effects are an increase in the math score by 0.388 standard deviation (18.51 points) and an increase in reading scores by 0.265 standard deviations (8.40 points). Note that math score ranges from 288 to 626 and the reading score ranges from 315 to 627. The net direct impact of being black is having math score 0.377 standard deviation (17.98 points) lower. Girls perform better than boys in both math and reading tests. Teacher's race and education have an insignificant impact on students' achievements.

These net direct effects are combined with endogenous effects and exogenous effects and generate a total effect as shown in Section 4.3.3. For example, to calculate the total effect of gender, plug $\lambda = 0.04$, $\beta = 0.14$, and $\gamma = 0.172$ into equation (4.17). For a class of 20 students, the total effect of being a girl on one's math score is 0.14 standard deviation (or 6.70 points) .

All models in this chapter control for individual characteristics and class level characteristics, unless otherwise specified. The estimates for coefficients of these characteristics are very stable in magnitude and significance level across all specifications. Therefore, I will suppress their estimates in following tables.

My main model allows the disturbance terms ϵ_{ic} to be heteroscedastic across class types. The term σ_{ϵ_1} is the standard deviation of ϵ_{ic} in small classes, and σ_{ϵ_2} is the standard deviation of ϵ_{ic} in regular/aid classes. The term σ_α is the estimates for the standard deviation of random class effects. It is close to 0. Note that my model controls for school by class type fixed effects. Therefore, most of the class level heterogeneity is absorbed by these fixed effects. Many studies interpret σ_α as unobserved teacher effectiveness. Under this interpretation, the results show no unobserved teacher effects. In general, my estimation finds few peer effects, either endogenous or exogenous.

Model reported in Columns 3 and 4 are different in that they control for school fixed effects and class type fixed effect separately. Endogenous peer effects for reading and some of the exogenous peer effects become significant in this model. But such significant peer effects are likely to be the result of class specific effects that are correlated with student or class characteristics. A likelihood ratio test rejects the hypothesis that class type effects are homogeneous across schools and advocates using school by class type fixed effects. I will further show that class type effects are heterogeneous across schools. Note that σ_α is again insignificant in this specification. The reason is that endogenous peer effects λ and individual heterogeneity σ_ϵ are sufficient to explain the between-class variation.

Table 4.4 provides the estimates for a model without endogenous peer effects as a comparison. Columns 1 and 2 control for school by class type fixed effects. The model for Columns 3 and 4 controls for school fixed effects and class type fixed effects separately. Comparing Columns 1 and 2 of Table 4.4 with Columns 1 and 2 of Table 4.3, the coefficients for other variables do not change much after imposing a restriction that the endogenous effect is zero. Also, a likelihood ratio test fails to reject the null hypothesis that the endogenous effect is zero. These two facts further support that endogenous effect is 0.

The difference between my results and Graham (2008)'s significant positive endogenous effect is mainly due to the inclusion of school by class type fixed effects. If school fixed effects and class type fixed effects are additive, as in Columns 3 and 4 of Table 4.3, the estimate for λ is 0.306 for math and 0.363 for the reading score, where the latter is significant. This is close to Graham's estimate of 0.45 for math score and 0.56 for the reading score in Columns 1 and 2 of Table 4.5. It is also close to a maximum likelihood estimator for Graham's full mean model in Columns 1 and 2 of Table 4.6. Once school by class type fixed effects are controlled for, Graham's full mean model yields an insignificant λ, either with the conditional variance method (See Columns 3 and 4 of Table 4.5) or with maximum likelihood estimation (see Columns 3 and 4 of Table 4.6).

I also apply Graham's method to data in grade one to three. If school fixed effects and class type fixed effects are controlled for separately, as shown in Columns

1 and 2 of Table 4.5, Graham's conditional variance will yield positive and significant peer effects for both reading and math scores in grade one, and reading scores in grade three. The magnitude is greater than 0.5. If school by class type fixed effects are controlled for, as shown in Columns 3 and 4 of Table 4.5, the estimator for endogenous peer effects by Graham's method is only significant for math scores in grade one.

4.4.2 Missing Observations and Outliers

The initial kindergarten sample includes 6325 students and 325 classes. Students with missing information are dropped out of the sample, mainly due to missing test scores. Classes with less than two students with complete information or missing teacher information are dropped. The final sample includes 323 classes, 5804 students for math scores and 5723 students for reading scores. That is, about 10% of the observations are missing.

Since identification comes from the within-class variance and between class variance, how the missing observations affect the estimator for λ depends on how they affect these two variances. The estimate for λ increases with between class variance and decreases with within-class variances. Missing observations would bias λ downward if they bias the within-class variance upward or between-class variance downward. The within-class variance is biased upward if observations close to the class mean is missing. Between-class variance is biased downward if high performers in high-achieving classes or low performers in low-achieving classes are more likely to be missing. In practice, though, students with higher absent days and potentially lower test scores are more likely to be missing across all classes.[13] Therefore, the missing observations can potentially bias the estimates upward. Given that the estimates for λ is slightly above 0 and insignificant, the bias does not change the conclusion of no peer effects.

Section 4.3.2 shows that identification of the model comes from within-class variance and between-class variance. Theses variances may be sensitive to some extreme values. Also, the spotlight model and the bad apple model suggest that peers with very high scores or very low scores have special impacts on others. Therefore, I exclude outliers and reestimate the main model to check robustness of the results.

Results after excluding the outliers are in Table 4.7. Columns 1 and 2 exclude outliers for whole sample, whose score is outside three standard deviations of the sample mean. Columns 3 and 4 exclude outliers of the class, whose score is outside three standard deviations of the class mean. While the magnitude of the estimates increases, endogenous peer effects remain insignificant. The exogenous effects, too, are insignificant. Therefore, the results are robust to excluding outliers.

4.4.3 Alternative Assumptions on the Variance Structure

Graham's method for identifying peer effects is variance-based rather than regression-based. Therefore, assumptions on the structure of variances are key to identification. A major critic on Graham (2008) is that its assumptions are unjustified and may be too restrict (Durlauf and Tanaka, 2008; Blume et al., 2011).

[13] This is the result from regressing the indicator for missing test score on absent days.

As demonstrated in Section 4.3.2, identification in this study is also variance-based as it relies on the within-class variance and between-class variance. But my model is more flexible than Graham (2008) in that it can relax some of the assumptions on class effect α_c and the disturbance terms ϵ_{ic} . Of course, certain restrictions are still necessary for identification. Firstly, the assumptions that disturbance terms are independent, class effects are independent, and disturbance terms are independent of class effects are indispensable for identification. These assumptions are justified by the random assignment of teachers and students. Second, the endogenous peer effect λ is unidentifiable when both ϵ_{ic} and α_c are heteroscedastic across classes. In that scenario, the number of moment conditions from between-class variance and within-class variance is less than the number of parameters to be estimated. But my model can accommodate alternative assumptions on the homoscedasticity of ϵ_{ic}.

Following Graham (2008), I assume that the disturbance terms ϵ_{ic} are heteroscedastic across class types in model (4.1), i.e., $\sigma^2_{\epsilon,ic} = \sigma^2_{\epsilon 1}$ if class c is a small class and $\sigma^2_{\epsilon,ic} = \sigma^2_{\epsilon,2}$ if class c is a regular/aid class. Disturbance terms reflect unobserved heterogeneity in student ability. Given the random assignments of students within the school, a more plausible assumption is that ϵ_{ic} is homoscedastic within schools and heteroscedastic across schools.

I check the robustness of the results to alternative assumptions on ϵ_{ic} by examining two extreme cases. In all cases, ϵ_{ic} are independently distributed across i and c. In the first case ϵ_{ic} is homoscedastic across all classes, i.e., $\sigma^2_{ic} = \sigma^2$ for all i and c. In the second case ϵ_{ic} is homoscedastic within the class and heteroscedastic across classes, i.e., $\sigma^2_{ic} = \sigma^2_{\epsilon,c}$ for any c. The estimates under these alternative assumptions are in Table 4.8. A key takeaway from the results is that endogenous effect is insignificant under all specifications. Therefore, the insignificance of endogenous peer effects is robust to alternative assumptions on variance structure.

4.4.4 Heterogeneous Peer Effects

Numerous authors have demonstrated that peer effects can be heterogeneous within a group. The impact of a peer on a student can depend on the peer's characteristics and the student's characteristics. For example, the influence of a boy on another boy might be different from that of a boy on a girl. Peer effect studies find that peer effects may change with student ability. For example, Carrell et al. (2009) find that low-ability students benefit more from having high ability peers. Hoxby and Weingarth (2005) discover that peer effects vary with one's position in the ability distribution. Meanwhile, peer effects may be underestimated if students tend to interact more with students with similar ability (Marmaros and Sacerdote, 2006). Unfortunately, the Project STAR data does not have pre-experiment test score to serve as an indicator of ability for kindergarten students. The data does not have precise friendship structure inside classes, either.

However, observed characteristics like sex and race are strongly correlated with one's academic performance. Therefore, I set up an extended model that allow peer effects to vary with student's own and peer's characteristics. Such models also have implications for racial segregation and single-sex education. For example, if peer effects are larger among same-sex peers than among opposite-sex peers, then single-sex classes can potentially promote performance.

The main model in equation (4.1) is extended to allow for higher order spatial autocorrelation, thus enables estimation of heterogeneous endogenous peer effects. Consider the case when each class is divided into two subgroups based on observed characteristics, e.g, each class can be divided into girls and boys. The number of boys and girls in class c is n_c^b and n_c^g respectively. The outcome of boys is determined by

$$y_{ic} = \beta_0 + \lambda_1 \bar{y}^b_{(-i)c} + \lambda_2 \bar{y}^g_c + x'_{ic}\beta + \bar{x}'^b_{(-i)c}\gamma_1 + \bar{x}'^g_c\gamma_2 + \psi'_c\pi + f'_c\phi + \alpha_c + \epsilon_{ic}, \quad (4.20)$$

and the outcome of girls is determined by

$$y_{ic} = \beta_0 + \lambda_3 \bar{y}^b_c + \lambda_4 \bar{y}^g_{(-i)c} + x'_{ic}\beta + \bar{x}'^b_c\gamma_3 + \bar{x}'^g_{(-i)c}\gamma_4 + \psi'_c\pi + f'_c\phi + \alpha_c + \epsilon_{ic}, \quad (4.21)$$

where $\bar{y}^b_{(-i)c}$ is the leave-out mean of outcome of boys in class,

$$\bar{y}^b_{(-i)c} = \frac{1}{n_c^b - 1}\sum_{j\neq i}^{n_c}[(1 - \mathbb{1}^g_{jc})y_{jc}], \quad (4.22)$$

dummy variable $\mathbb{1}^g_{ic}$ is an indicator for individual i in class c being girl, $\bar{y}^g_c$ is the mean outcome of girls in class

$$\bar{y}^g_c = \frac{1}{n_c^g}\sum_{j=1}^{n_c}\mathbb{1}^g_{jc}y_{jc}, \quad (4.23)$$

$\bar{y}^g_{(-i)c}$ is the leave-out mean of outcome of girls in class,

$$\bar{y}^g_{(-i)c} = \frac{1}{n_c^g - 1}\sum_{j\neq i}^{n_c}[\mathbb{1}^g_{jc}y_{jc}], \quad (4.24)$$

$\bar{y}^b_c$ is the mean outcome of boys in class,

$$\bar{y}^b_c = \frac{1}{n_c^b}\sum_{j=1}^{n_c}(1 - \mathbb{1}^g_{jc})y_{jc}. \quad (4.25)$$

The parameters λ_1, λ_2, λ_3, λ_4 are the endogenous peer effects of boys on boys, girls on boys, boys on girls and girls on girls respectively. Coefficients γ_1, γ_2, γ_3, γ_4 are the exogenous peer effects of boys on boys, girls on boys, boys on girls and girls on girls respectively.

An alternative specification is to make the weight reciprocal of total number of peers in the class, i.e., $1/(n_c - 1)$ in class c. Then the outcome of boys is determined as

$$y_{ic} = \beta_0 + \lambda_1 \tilde{y}^b_{(-i)c} + \lambda_2 \tilde{y}^g_c + x'_{ic}\beta + \tilde{x}'^b_{(-i)c}\gamma_1 + \tilde{x}'^g_c\gamma_2 + \psi'_c\pi + f'_c\phi + \alpha_c + \epsilon_{ic}, \quad (4.26)$$

and the outcome of girls is determined as

$$y_{ic} = \beta_0 + \lambda_3 \tilde{y}^{boys}_c + \lambda_4 \tilde{y}^{girls}_{(-i)c} + x'_{ic}\beta + \tilde{x}'^{boys}_c\gamma_3 + \tilde{x}'^{girls}_{(-i)c}\gamma_4 + \psi'_c\pi + f'_c\phi + \alpha_c + \epsilon_{ic}, \quad (4.27)$$

where

$$\tilde{y}^b_{(-i)c} = \frac{1}{n_c - 1}\sum_{j \neq i}^{n_c}[(1 - \mathbb{1}^g_{jc})y_{jc}],$$
$$\tilde{y}^g_c = \frac{1}{n_c - 1}\sum_{j=1}^{n_c}\mathbb{1}^g_{jc}y_{jc}$$
$$\tilde{y}^{boy}_c = \frac{1}{n_c - 1}\sum_{j \neq i}^{n_c}[(1 - \mathbb{1}^{girl}_{jc})y_{jc}],$$
$$\tilde{y}^{girl}_{(-i)c} = \frac{1}{n_c - 1}\sum_{j \neq 1}^{n_c}\mathbb{1}^{girl}_{jc}y_{jc}.$$

If $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda$, then the model is the same as model (4.1).

Models (4.20) and (4.21) are plausible if both interaction intensity and influence vary with gender, as both the weight (reflecting interaction intensity) and the endogenous peer effects λ (representing peer influence) depend on the gender. Models (4.26) and (4.27) are plausible if interaction intensity is the same across gender but influences differs by gender because weight only depends on class size and is not impacted by gender. Both models can be estimated under the framework of spatial econometrics like the main model, see Appendix G.

Results are in Columns 1 and 2 of Table 4.9. An alternative model uses $1/(n_c - 1)$ as spatial weights as characterized by equations (4.26) and (4.27). Results are in Columns 3 and 4 of Table 4.9. The estimate for λ is insignificant for both math and reading score in both models, demonstrating the robustness of the result to alternative weights matrix.

4.4.5 Peer Effects in Higher Grades

Due to deviation from the randomization scheme in higher grades, this chapter analyzes academic performance of kindergarten students. The results show insignificant peer effects in kindergarten. However it is possible that peer effects are significant in higher grades. Peer effects may grow stronger as children get older and more socialized. In this section, I discuss results for higher grades.

I estimate the main model in equation (4.1) by test and grade with kindergarten to grade three data. Since grade repetition status is not available for all grades, the individual characteristics only include age, race, gender, and poor (free lunch status). All else is the same. Results for math scores are in Table 4.10 and results for reading score are in Table 4.11. The model for Column 1 of Table 4.10 and Table 4.11 is the same as the model for Column 1 and 2 of Table 4.3 except that repeat (the variable indicating grade repetition) is dropped. The estimates change little after dropping the variable.

The first observation is that the endogenous effects remain insignificant for both math score and reading score across all grades. For math score in grade one, the estimate for λ is 0.135, but only significant at the 10% level.

Some of the exogenous peer effects in reading scores become significant in higher grades. For example, in grade three, if the mean peer age increases by one year, the individual reading scores decreases by 0.69 standard deviation as a net direct effect. In grade two, individual reading score increases with the share of black students in class. One has to keep in mind that these estimates are subject to the problem of endogenous group formation due to non-random switching and attrition of students.

4.4.6 Class Size Effects

Endogenous peer effects generate social multiplier effects and affect policy evaluation. Since Project STAR is a class size reduction experiment, this study studies how incorporating peer effects will influences the assessment and interpretation of class size effect.

In the reduced form of the peer effect model, the total effect of class characteristics ψ is $\frac{\pi}{1-\lambda}$. If $0 < \lambda < 1$, the net direct impact of ψ on the dependent variable is magnified by $1/(1 - \lambda)$. Boozer and Cacciola (2001) point out that the reduced-form class size effect "constitute a 'black box' of underlying components". Their estimated "pure class size effect" net of the peer effect is much smaller than previous studies and mostly insignificant. But their estimates for endogenous peer effect may be upward biased due to weak instrument problem.

In the model controlling for school fixed effects and class type fixed effects separately, the class size effect is measured directly. As Columns 3 and 4 of Table 4.3 shows, students in regular/aid classes perform worse than students in small classes.

In the model controlling for school by class type fixed effects, I regress the estimates for school by class type fixed effects on school fixed effects and class type fixed effects as well as the interaction term between class type fixed effects and percent of black students in school. Results are in Table 4.12. Columns 1 and 2 use the school by class type fixed effects obtained from the main model (4.1). They correspond to the estimates in Columns 1 and 2 of Table 4.3. Columns 3 and 4 use the estimates for school by class type fixed effects of the model without peer effects, corresponding to Columns 1 and 2 and of Table 4.4. Estimates in Columns 1 and 2 are only slightly smaller than those in Column 3 and 4 as peer effects are close to 0 and insignificant.

The results confirm previous studies that students in smaller classes have better academic performance than those in regular sized classes. Compared to students in small classes, math scores are 0.114 standard deviation lower for students in regular classes and 0.123 standard deviation lower for students in the regular classes with aid. The results are significant at the 10% level. Class size effect on reading scores is stronger. Students in regular classes have a 0.136 standard deviation lower reading score than students in small classes. Regular/aid class students have a 0.166 standard deviation lower reading score. Both estimates are significant at the 5% level.

The table also shows that the class size effect is heterogeneous across schools with different shares of black students. If the proportion of black students in school increases by one percentage point, the achievement gap between small and regular classes increases by 0.261 standard deviation. The estimate is significant at the 5% level. This illustrates the importance of controlling for school by class type fixed effects.

Results above show that at the class level, the class type effect vary with the percent of black students in schools. One may wonder if such variation persists if the interaction term between students' race and class type effect is is controlled for. The answer is yes, because the heterogeneous responses to class reduction cannot be controlled simply by heterogeneous response to class reduction at the individual level. I reestimate the model by adding the interaction term of individual's race and the dummies for class types. The results are in Table 4.13. The interaction term between regular class and percent of black students in school is still significant.

4.5 Conclusion

This chapter assesses peer effects in kindergarten student's reading and math scores, along with the effects of teacher quality, class size, and other factors. Data is from Project STAR, which randomly assigned teachers and students to small classes, regular aid classes and regular classes with aid within the school. This chapter models and measures peer effects under the framework of spatial econometrics. Identification is based on within-class variance and between-class variance and hinges on variation in class size. It relaxes the functional assumptions of Graham (2008) by replacing the school plus class type fixed effects with the more flexible school by class type fixed effects, so that Graham's finding of sizable endogenous peer effects does not hold.

By and large, this study finds that peer characteristics (age, gender, race, poverty status, and grade repetition) and peer scores have insignificant impacts on individual scores. While the results do not necessarily generalize to other settings, like college students or teenagers, they do cast doubts on the effectiveness of policies that try to manipulate peer groups for better outcomes. This study also does not find any teacher effects, either measured by observed teacher characteristics or unobserved class effects. In accordance with previous studies of the class size effect using Project STAR data, the results show that small classes improve children's academic performance.

While finding different results from the peer effects literature, particularly Graham (2008), the finding of few peer effects in this chapter may not be too surprising. Empirically, there is evidence that regrouping programs do not bring the anticipated change to the academic performance of students. For example, assigning low-ability students to high-ability peers even leads to a decrease in the score of the low-ability students (Carrell et al., 2013). Hurricane-induced reallocation of students does not lower the average score of incumbent students in receiving schools (Imberman et al., 2012). Theoretically, Angrist (2014) cautions against "spurious correlation" that is usually mistaken as peer effects and can easily arise from problematic identification strategy or common shocks to the groups. This study accounts for common shocks by including the random class effects and school by class type fixed effects. After taking into account these "perils of peer effects," the endogenous effects do fade away.

While it is possible that previous findings of significant peer effects are contaminated by insufficient control of confounding factors or reflection problems, the finding of insignificant peer effects in this study may also be due to that kindergarten students are just too young to interact academically. A majority of empirical studies finding strong peer effects use samples for high school or college students, when individuals are more socialized. Another explanation is that Project Star over-sampled schools in poor areas (Boyd-Zaharias et al., 2007). Rescores are more limited in such schools. The competition for recourses can generate negative peer effects, offsetting the positive peer effects generated from knowledge sharing, conforming to group norms, etc. For example, Antecol et al. (2016) find negative peer effects in disadvantaged primary schools and show that the results are likely to be driven by using the sample of disadvantaged schools.

4.6 Figures and Tables

Figure 4.1: Distribution of class size

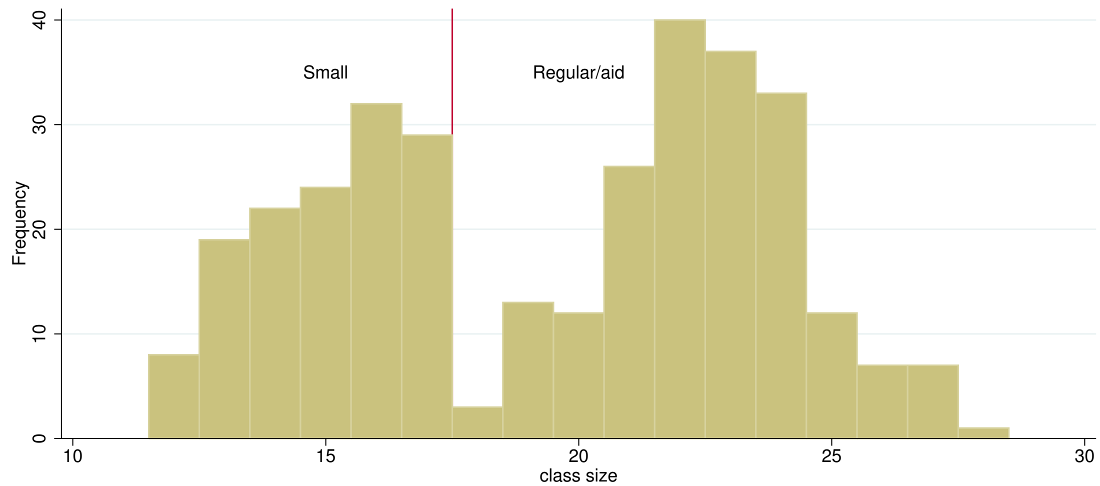
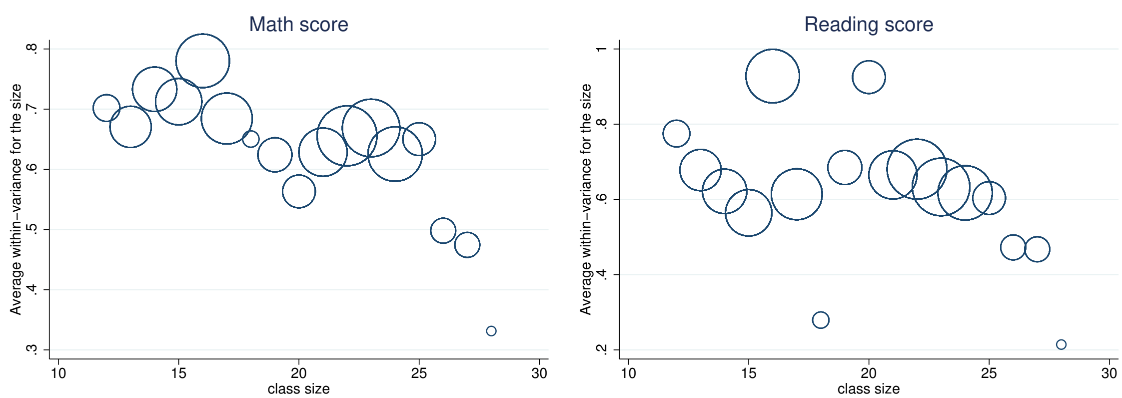


Figure 4.2: Within-class variance v.s. class size



Note: Size of the circle is proportional to the number of classes for that size.

Table 4.1: Summary Statistics for Students

	observations	mean	sd
poor	6300	0.484	0.500
black	6322	0.326	0.469
girl	6325	0.486	0.500
repeat	6297	0.040	0.196
age	6317	5.428	0.349
math score	5871	485.377	47.698
reading score	5789	436.725	31.706
black teacher	6282	0.165	0.371
teacher has M.A.	6304	0.347	0.476

[1] The sample size is 6325.

[2] Repeat means grade repetition. Poor indicates receiving free lunch.

Table 4.2: Summary Statistics for Classes

	obs	mean	sd	min	max
class size	325	19.46	4.14	12.00	28.00
within-class sd of mathnorm	325	0.82	0.22	0.41	2.01
within-class sd of readnorm	325	0.80	0.31	0.24	1.91
class mean of mathnorm	325	0.02	0.58	-1.58	2.00
class mean of readnorm	325	0.02	0.57	-1.39	1.79

[1] There are 325 classes in kindergarten.

[2] Mathnorm is the normalized math score. Readnorm is the normalized reading score.

Table 4.3: Results for the Main Model

	School by Class Type FE		School FE+Class Type FE	
	math	reading	math	reading
	Endogenous peer effect λ			
λ	0.040	0.035	0.306	0.363***
	(0.080)	(0.084)	(0.961)	(0.042)
	Exogenous peer effect γ			
age	-0.042	-0.526	0.318	-0.073
	(0.383)	(0.382)	(0.196)	(0.169)
black	-0.615	-0.628	0.190	-0.240
	(0.642)	(0.521)	(0.333)	(0.256)
girl	0.172	0.173	0.254**	0.210*
	(0.264)	(0.304)	(0.119)	(0.113)
poor	0.068	-0.050	0.090	0.181*
	(0.283)	(0.283)	(0.128)	(0.106)
repeat	-0.746	0.720	-0.505	0.206
	(0.693)	(0.705)	(0.337)	(0.291)
	Coefficients of personal and class characteristics: β and π			
age	0.388***	0.265***	0.414***	0.298***
	(0.045)	(0.044)	(0.041)	(0.041)
black	-0.377***	-0.261***	-0.337***	-0.243***
	(0.059)	(0.057)	(0.052)	(0.051)
girl	0.140***	0.168***	0.147***	0.173***
	(0.026)	(0.029)	(0.023)	(0.024)
poor	-0.404***	-0.458***	-0.409***	-0.453***
	(0.031)	(0.032)	(0.028)	(0.028)
repeat	-0.464***	-0.400***	-0.456***	-0.440***
	(0.078)	(0.070)	(0.066)	(0.060)
tchblack	0.012	0.098	0.022	0.024
	(0.105)	(0.105)	(0.054)	(0.053)
tchms	0.065	0.055	-0.007	0.012
	(0.064)	(0.058)	(0.029)	(0.026)
regular class			-0.106***	-0.123***
			(0.033)	(0.029)
regular aid			-0.107***	-0.101***
			(0.031)	(0.028)
σ_{ϵ_1}	0.781***	0.803***	0.795***	0.820***
	(0.014)	(0.022)	(0.043)	(0.023)
σ_{ϵ_2}	0.866***	0.855***	0.876***	0.870***
	(0.024)	(0.031)	(0.063)	(0.033)
σ_α	1.8e-05	7.6e-06	0.111	6.0e-05
	(4.0e-05)	(1.5e-05)	(0.629)	(2.4e-04)
R	323	323	323	323
N	5804	5723	5804	5723
lnf	-6982.9	-6974.8	-7138.9	-7116.3

[1] Columns 1 and 2 show estimation results of the main model (3.2), which controls for school by class type fixed effects.The model for Columns 3 and 4 differs from the model 3.2 only in that it controls for school fixed effects and class type fixed effects separately. The dependent variable is normalized SAT score. Both models control for personal

Table 4.4: Model with No Endogenous Effects

	School × class type FE		School FE+ class type FE	
	Math	Reading	Math	Reading
	Estimates for average peer characteristics			
age	-0.029	-0.535	0.660**	-0.031
	(0.399)	(0.395)	(0.283)	(0.267)
black	-0.654	-0.658	0.174	-0.437
	(0.668)	(0.540)	(0.468)	(0.369)
girl	0.183	0.184	0.394**	0.371**
	(0.275)	(0.315)	(0.171)	(0.187)
poor	0.049	-0.069	-0.075	-0.041
	(0.295)	(0.293)	(0.181)	(0.168)
repeat	-0.787	0.730	-0.904*	0.209
	(0.720)	(0.730)	(0.477)	(0.440)
tchblack	0.012	0.101	0.046	0.063
	(0.109)	(0.109)	(0.078)	(0.085)
tchms	0.068	0.057	-0.007	0.011
	(0.067)	(0.060)	(0.042)	(0.040)
regular class			-0.150***	-0.195***
			(0.047)	(0.046)
regular aid			-0.152***	-0.156***
			(0.045)	(0.044)
σ_{ϵ_1}	0.780***	0.803***	0.783***	0.805***
	(0.013)	(0.023)	(0.013)	(0.023)
σ_{ϵ_2}	0.867***	0.855***	0.855***	0.846***
	(0.025)	(0.032)	(0.024)	(0.033)
σ_α	0.027	0.022	0.262***	0.248***
	(0.106)	(0.121)	(0.016)	(0.022)
R	323	323	323	323
N	5804	5723	5804	5723
lnf	-6983.5	-6975.3	-7140.2	-7120.0

[1] Estimates from the model without endogenous peer effect.

[2] *,0.1,**,0.05,***, 0.01. Standard errors in the parentheses. The dependent variable is the normalized test score. The model controls for individual characteristics(age, gender,race, poverty status and whether one is repeating a grade)

[3] The model for in Columns 1 and 2 controls for school by class type fixed effects. The model for Columns 3 and 4 controls for school fixed effects and class type fixed effects. Both models allow disturbance terms to be heteroscedastic by class types.

Table 4.5: Estimates of Endogenous Effects with Graham's Method

	Original grham model		Modified Graham model	
	γ_0^2	corresponding λ	γ_0^2	corresponding λ
	Original Graham (2008) data: Kindergarten			
math	3.469(1.033)**	0.463	1.895(0.622)	0.274
reading	5.282(2.481)*	0.565	2.913(1.684)	0.414
	New data:Kindergarten			
math	3.241(1.038)**	0.445	2.107(0.666)*	0.311
reading	5.058(2.221)*	0.555	3.223(1.600)	0.443
	New data:Grade One			
math	7.045(2.094)***	0.623	3.316(1.123)**	0.451
reading	4.589(1.209)***	0.533	1.727(0.540)	0.239
	New data:Grade Two			
math	4.355(1.811)*	0.521	1.698(0.530)	0.233
reading	4.435(2.099)	0.525	1.438(0.547)	0.166
	New data: Grade Three			
math	2.516(1.401)	0.370	1.474(0.786)	0.176
reading	4.024(1.086)***	0.501	1.742(0.620)	0.242

[1] *,0.1,**,0.05,***, 0.01 for test $\gamma_0^2 = 1$, with the endogenous peer effect being $\lambda = 1 - \frac{1}{\gamma_0}$, so $\gamma_0^2 = 1$ corresponds to $\lambda = 0$ if $|\lambda| < 1$. Robust standard errors in the parentheses.

[2] This table shows estiamtes for peer effects using the conditional variance method in Graham (2008). Model for Columns 1 and 2 uses the original model in Graham(2008) as shown in (H.3) and controls for school fixed effects and class type fixed effects separately. Results in Columns 1 and 2 of the top panel are the replication of the Graham(2008) results. Model for Columns 3 and 4 controls for school by class type fixed effects.

[3] Graham's data is slightly different from mine. The data in this paper is published in 2008, long after Graham (2008) was written. Besides offering more varialbes and slightly more observations, the data this paper uses also gives explicit class assignment information. Graham (2008) constructs the class assignment variable from teacher characteristics.

Table 4.6: ML Estimates for Graham's Full Mean Model

	School FE+class type FE		School by class type FE	
	math	reading	math	reading
	Endogenous peer effect			
λ	0.373	0.377***	0.022	0.027
	(0.615)	(0.042)	(0.093)	(0.090)
σ_{ϵ_1}	0.821***	0.840***	0.818***	0.837***
	(0.015)	(0.023)	(0.014)	(0.023)
σ_{ϵ_2}	0.889***	0.880***	0.897***	0.888***
	(0.027)	(0.036)	(0.025)	(0.035)
σ_α	0.073	4.4e-06	2.1e-04	6.8e-09
	(0.573)	(1.2e-05)	(5.9e-04)	(1.0e-08)
R	323	323	323	323
N	5828	5747	5828	5747
lnf	-7431.8	-7388.6	-7267.7	-7245.2

[1] *,0.1,**,0.05,***, 0.01. Robust standard erros in the parentheses. λ is the endogenous effect, σ_{ϵ_1} and σ_{ϵ_2} are the standard deviations of the disturbance terms for small classes and regular/aid classes respectively, σ_α is the standard the deviation of randome class effect α_c. R is the number of classes, N is the number of students, lnf is the log likelihood.

[2] The results are the maximum likelihood estimates for Graham's full mean model in (4.19). The first model for Columns 1 and 2 has exactly the same specification as Graham (2008) and controls for school fixed effects and class type fixed effects separately. The second model for Columns 3 and 4 is the same as model (4.19) except that it controls for school by class type fixed effects.

[3] Following Graham(2008), both models do not control for personal characteristics, peer characteristics and teacher characteristics.

Table 4.7: Robustness to Excluding Outliers

	Excluding outliers of the sample		Excluding outleirs of the class	
	math	reading	math	reading
		Endogenous peer effect		
λ	0.060	0.112	0.052	0.117
	(0.080)	(0.076)	(0.078)	(0.076)
		Exogenous peer effect		
age	-0.071	-0.456	-0.083	-0.489
	(0.374)	(0.324)	(0.382)	(0.356)
black	-0.645	-0.669*	-0.622	-0.734
	(0.653)	(0.404)	(0.633)	(0.448)
girl	0.227	0.251	0.190	0.126
	(0.269)	(0.223)	(0.262)	(0.270)
poor	0.083	0.175	0.091	-0.072
	(0.279)	(0.227)	(0.281)	(0.262)
repeat	-0.648	0.578	-0.739	0.777
	(0.676)	(0.588)	(0.688)	(0.686)
tchblack	0.023	0.094	0.014	0.079
	(0.103)	(0.092)	(0.103)	(0.101)
tchms	0.062	0.089*	0.064	0.069
	(0.063)	(0.050)	(0.063)	(0.052)
σ_{ϵ_1}	0.778***	0.679***	0.773***	0.724***
	(0.014)	(0.013)	(0.014)	(0.020)
σ_{ϵ_2}	0.856***	0.706***	0.866***	0.791***
	(0.020)	(0.018)	(0.024)	(0.029)
σ_α	4.0e-07	1.2e-04	5.3e-05	1.8e-04
	(8.1e-07)	(5.9e-04)	(1.3e-04)	(4.6e-04)
R	323	323	323	323
N	5797	5630	5796	5675
lnf	-6937.6	-5877.1	-6931.6	-6372.4

[1] Columns 1 and 2 exclude outliers of the whole sample, that is above or below 3 standard deviations of the sample mean. Columns 3 and 4 exlude outliers of the classe, that is above or below 3 standard deviations of the class mean.

[2] *,0.1,**,0.05,***, 0.01. Robust standard errors in the parentheses.

[3] Both models control for individual characteristics(gender, race, age, poverty status,repeat), teacher characteristics (teacher's highest education and race)

[4] Observations with missing information are dropped. N is the number of students and R is the number of classes in the final sample. In the original sample, N=6325, R=325.

Table 4.8: Homoscadastic and Heteroscadastic Disturbance Terms

	Homoscedastic ϵ_{ic}		Heteroscadastic ϵ_{ic}	
	math	reading	math	reading
		Endogenous peer effect		
λ	0.054	0.042	-0.130	-0.112
	(0.078)	(0.084)	(0.128)	(0.142)
		Exogenous peer effect		
age	-0.072	-0.573	-0.232	-0.500
	(0.374)	(0.372)	(0.385)	(0.398)
black	-0.548	-0.590	-1.279*	-0.514
	(0.641)	(0.510)	(0.689)	(0.429)
girl	0.145	0.173	0.098	0.211
	(0.254)	(0.304)	(0.229)	(0.226)
poor	0.026	-0.067	-0.033	-0.297
	(0.282)	(0.283)	(0.291)	(0.239)
repeat	-0.632	0.787	-0.556	0.660
	(0.670)	(0.690)	(0.726)	(0.752)
tchblack	0.009	0.095	-0.058	0.145*
	(0.106)	(0.104)	(0.088)	(0.084)
tchms	0.070	0.055	0.068	0.155***
	(0.064)	(0.058)	(0.060)	(0.053)
σ_ϵ	0.808***	0.819***		
	(0.012)	(0.018)		
σ_α	1.3e-08	6.0e-08	2.3e-06***	1.1e-07***
	(2.4e-04)	(2.1e-04)	(4.9e-07)	(2.6e-08)
R	323	323	323	323
N	5804	5723	5804	5723
lnf	-6996.3	-6979.5	-6604.2	-6173.1

[1] The model for Columns 1 and 2 assume homoscedestic disturbance terms. The model for Columns 3 and 4 allows for heteroscedastic disturbance terms.

[2] *,0.1,**,0.05,***, 0.01. Robust standard errors in the parentheses.

[3] Both models control for individual characteristics(gender,race, age, poverty status,repeat), teacher characteristics (teacher's highest education and race)

[4] Observations with missing information are dropped. N is the number of students and R is the number of classes in the final sample. In the original sample, N=6325, R=325.

Table 4.9: Peer Effects by Gender

	Model one		Model two	
	math	reading	math	reading
	Endogenous peer effect			
boy on boy	0.177	0.211	0.033	0.140
	(0.169)	(0.231)	(0.211)	(0.257)
girl on boy	-0.229	-0.294	-0.063	-0.277
	(0.196)	(0.259)	(0.268)	(0.337)
boy on girl	0.219	0.246	0.025	0.156
	(0.196)	(0.282)	(0.246)	(0.328)
girl on girl	-0.132	-0.128	0.015	-0.073
	(0.120)	(0.151)	(0.196)	(0.280)
	Exogenous peer effect:Boy on boy			
age	-0.030	-0.421	-0.536	-0.952**
	(0.283)	(0.258)	(0.462)	(0.453)
black	-0.847*	-0.435	-2.086***	-1.755**
	(0.440)	(0.467)	(0.745)	(0.725)
poor	0.144	-0.030	0.353	0.038
	(0.171)	(0.157)	(0.325)	(0.311)
repeat	-0.588	0.180	-0.411	0.941
	(0.429)	(0.435)	(0.750)	(0.761)
	Exogenous peer effect:Girl on boy			
age	0.274	0.244	-0.178	-0.571
	(0.260)	(0.250)	(0.439)	(0.450)
black	-0.181	-0.415	-0.040	0.034
	(0.428)	(0.398)	(0.730)	(0.608)
poor	0.076	0.149	0.123	0.280
	(0.220)	(0.221)	(0.421)	(0.416)
repeat	-0.957	-0.162	-0.750	1.316
	(0.596)	(0.660)	(1.222)	(1.252)

to be continued

Table 4.9: Peer Effects by Gender:Continued

	Model one		Model two	
	math	reading	math	reading
	Exogenous peer effect:Boy on girl			
age	-0.272	-0.699**	-0.621	-1.110**
	(0.264)	(0.273)	(0.429)	(0.442)
black	-0.662	-0.664	-1.805**	-1.870***
	(0.454)	(0.491)	(0.745)	(0.708)
poor	0.239	0.190	0.622**	0.534*
	(0.182)	(0.182)	(0.313)	(0.315)
repeat	-0.632	0.439	-0.773	0.995
	(0.449)	(0.455)	(0.730)	(0.761)
	Exogenous peer effect:Girl on girl			
age	0.232	0.234	-0.173	-0.671
	(0.260)	(0.247)	(0.438)	(0.459)
black	-0.210	-0.118	0.109	0.401
	(0.400)	(0.363)	(0.723)	(0.614)
poor	-0.086	-0.102	-0.403	-0.359
	(0.200)	(0.194)	(0.415)	(0.407)
repeat	-0.307	-0.145	0.349	1.320
	(0.529)	(0.606)	(1.151)	(1.332)
tchblack	-0.005	0.122	0.023	0.173
	(0.106)	(0.100)	(0.106)	(0.106)
tchms	0.080	0.054	0.106*	0.102*
	(0.068)	(0.068)	(0.063)	(0.059)
σ_{ϵ_1}	0.778***	0.800***	0.777***	0.798***
	(0.014)	(0.022)	(0.014)	(0.022)
σ_{ϵ_2}	0.869***	0.855***	0.868***	0.852***
	(0.024)	(0.031)	(0.024)	(0.031)
σ_α	8.2e-05	8.7e-06	2.5e-05	4.3e-05
	(1.4e-04)	(2.4e-05)	(3.4e-05)	(9.2e-05)
R	323	322	323	322
N	5804	5719	5804	5719
lnf	-6968.6	-6940.5	-6964.0	-6936.1

[1] *,0.1,**,0.05,***, 0.01. Robust standard errors in the parentheses. The dependent variable is the normalized score.

[2] In model one as shown in equation (4.20) and (4.21), each weights matrix is row-normalized. Weight is $1/(n_c^b - 1)$, $1/n_c^g$, $1/n_c^b$ and $1/(n_c^g - 1)$ for effect of boys on boys, girls on boys, boys on girls and girls on girls. n_c^b and n_c^g are the number of boys and girls in class c. In model two as shown in equation (4.26) and (4.27), weight is $1/(n_c - 1)$ for all regardless of gender. n_c is the number of students in class c.

[3] Both models control for individual characteristics(race, age, poverty status,repeat,), teacher's highest education and race and school by class type fixed effect.

[4] Observations with missing information are dropped. N and R are the number of students and classes in the final sample. In the original sample, N=6325, R=325.

Table 4.10: Peer Effects by Grade: Math

	Kindergarten	Grade 1	Grade 2	Grade 3
		Endogenous peer effect		
λ	0.044	0.135*	-0.081	0.060
	(0.081)	(0.081)	(0.108)	(0.081)
		Exogenous peer effect		
age	-0.180	-0.201	-0.227	-0.271*
	(0.358)	(0.178)	(0.173)	(0.159)
black	-0.517	0.229	-0.561	0.222
	(0.634)	(0.379)	(0.544)	(0.485)
girl	0.216	-0.006	0.414*	-0.145
	(0.263)	(0.283)	(0.251)	(0.178)
poor	0.072	-0.095	-0.344	-0.105
	(0.285)	(0.189)	(0.211)	(0.177)
tchblack	0.022	0.140**	0.121	-0.120
	(0.108)	(0.070)	(0.091)	(0.083)
tchms	0.057	0.130**	0.018	0.028
	(0.064)	(0.064)	(0.050)	(0.050)
σ_{ϵ_1}	0.786***	0.778***	0.761***	0.774***
	(0.014)	(0.012)	(0.012)	(0.011)
σ_{ϵ_2}	0.866***	0.819***	0.791***	0.793***
	(0.024)	(0.019)	(0.018)	(0.015)
σ_{α}	8.3e-07	1.1e-04	1.8e-06	1.1e-06
	(2.2e-06)	(1.7e-04)	(3.0e-06)	(1.4e-06)
R	323	336	321	323
N	5806	6419	5755	5830
lnf	-7010.0	-7591.0	-6661.0	-6824.5

[1] *,0.1,**,0.05,***, 0.01. Robust standard errors in the parentheses. The dependent variable is the normalized Math score.

[2] Estimator for model 3.2 by grade. The first column is estimator for kindergarten and comparable to column 1 in table 4.3 excepet that the exogeneous variables change.

[3] The model controls for individual characteristics(gender,race, poverty status). It does not control for retention status because it is not available for certain grade. The model also controls for teacher characteristics (teacher's highest education and race) and school by class type fixed effect.

[4] Observations with missing information are dropped. N is the number of students and R is the number of classes in the final sample.

Table 4.11: Peer Effects by Grade: Reading

	Kindergarten	Grade 1	Grade 2	Grade 3
		Endogenous peer effect		
λ	0.031	-0.066	-0.158	-0.151
	(0.084)	(0.096)	(0.117)	(0.111)
		Exogenous peer effect		
age	-0.368	0.043	-0.195	-0.690***
	(0.360)	(0.166)	(0.172)	(0.187)
black	-0.649	0.870**	-0.676	0.730
	(0.523)	(0.377)	(0.531)	(0.518)
girl	0.159	0.180	0.476*	-0.080
	(0.303)	(0.302)	(0.244)	(0.243)
poor	-0.030	-0.323	-0.492**	-0.342
	(0.283)	(0.197)	(0.206)	(0.208)
tchblack	0.095	0.099	0.041	-0.067
	(0.106)	(0.071)	(0.085)	(0.068)
tchms	0.056	0.030	0.019	0.008
	(0.059)	(0.062)	(0.050)	(0.054)
σ_{ϵ_1}	0.806***	0.781***	0.762***	0.774***
	(0.022)	(0.012)	(0.013)	(0.011)
σ_{ϵ_2}	0.860***	0.824***	0.782***	0.822***
	(0.032)	(0.016)	(0.019)	(0.015)
σ_α	2.3e-07	8.8e-07	8.6e-05	1.3e-05
	(4.5e-07)	(1.5e-06)	(1.1e-04)	(1.2e-05)
R	323	332	321	318
N	5725	6237	5764	5752
lnf	-7000.5	-7405.9	-6660.4	-6802.9

1 *,0.1,**,0.05,***, 0.01. Robust standard errors in the parentheses. The dependent variable is the normalized Reading score.

2 Estimator for model 3.2 by grade. The first column is estimator for kindergarten and comparable to column 2 in table 4.3 excepet that the exogeneous variables change.

3 The model controls for individual characteristics(gender,race, poverty status). It does not control for retention status because it is not available for certain grade. The model also controls for teacher characteristics (teacher's highest education and race) and school by class type fixed effect.

4 Observations with missing information are dropped. N is the number of students and R is the number of classes in the final sample.

Table 4.12: Class Type Effects

	Peer effect model		Linear model	
	math	reading	math	reading
regular class	-0.114*	-0.136**	-0.119*	-0.141**
	(0.065)	(0.061)	(0.068)	(0.063)
regular aid	-0.123*	-0.166***	-0.128*	-0.172***
	(0.065)	(0.061)	(0.068)	(0.063)
reg*pctblack	-0.130	-0.261**	-0.135	-0.271**
	(0.134)	(0.125)	(0.140)	(0.130)
regaid*pctblack	-0.083	0.003	-0.086	0.003
	(0.132)	(0.123)	(0.138)	(0.128)

[1] Results from regressing the school by class type fixed effects on class type fixed effects and school fixed effects separately.The school by class type fixed effects are obtained from the main model in equation (4.1).

[2] *,0.1,**,0.05,***, 0.01. Standard errors in the parentheses.

[3] The first two columns use school by class type fixed effects from the main model, corresponding to Columns 1 and 2 of Table 4.3. Columns 3 and 4 use school by class type fixed effects from the linear model, corresponding to Columns 1 and 2 of Table 4.4.

Table 4.13: Class Type Effects: Controls for Students' Race by Class Type

	Peer effect model		Linear model	
	math	reading	math	reading
regular class	-0.113*	-0.136**	-0.118*	-0.141**
	(0.065)	(0.061)	(0.068)	(0.063)
regular aid	-0.121*	-0.166***	-0.127*	-0.172***
	(0.065)	(0.061)	(0.068)	(0.063)
reg*pctblack	-0.357***	-0.275**	-0.362**	-0.284**
	(0.134)	(0.125)	(0.140)	(0.130)
regaid*pctblack	-0.215	-0.050	-0.218	-0.050
	(0.132)	(0.123)	(0.138)	(0.128)

[1] Results from regressing the school by class type fixed effects on class type fixed effects and school fixed effects separately.The school by class type fixed effects are obtained after controlling for the interaction term between students' race and class type fixed effects, in addition to other variables in model (4.1).

[2] *,0.1,**,0.05,***, 0.01. Standard errors in the parentheses.

[3] The first two columns use school by class type fixed effects from the main model, corresponding to Columns 1 and 2 of Table 4.3. Columns 3 and 4 use school by class type fixed effects from the linear model, corresponding to Columns 1 and 2 of Table 4.4.

Appendix A: Analytical Properties of Matrices $pI_m + qJ_m$ in General

Many matrices in this paper can be written in the form of blocks of $pI_m + qJ_m$ type of matrices, where I_m is the identity matrix of size m, $J_m = \iota_m \iota_m'$ is the $m \times m$ matrix of ones, p and q are some real numbers. Let $I_m^* = I_m - J_m/m$, $J_m^* = J_m/m$, then

$$pI_m + qJ_m = pI_m^* + sJ_m^*,$$

where $s = p + mq$. The $pI_m + qJ_m$, or $pI_m^* + sJ_m^*$ type of matrices bear some special properties, which can facilitate the calculation and proof in this research and may be of interest more generally. This appendix lists properties of $pI_m^* + sJ_m^*$ (and $pI_m + qJ_m$) in general.

Both I_m^* and J_m^* are projection matrices, with $I_m^* + J_m^* = I_m$, and $I_m^* J_m^* = 0$. For any $m \times k$ matrix X, let $\bar{X} = \iota_m' X/m$ be the $k-$dimensional row vector of column means of X. Let $X^* = X - \iota_m \bar{X}$ be the deviations from column means of X. Then $I_m^* X = X^*$, $J_m^* X = \iota_m \bar{X}$. Also, $X'X = X^{*\prime} X^* + m\bar{X}'\bar{X}$.

While matrix multiplication obeys distributive and associative rules, the multiplication of $pI_m^* + sJ_m^*$ type of matrices further obeys the commutative properties.

Lemma A.1. *If $A = p_A I_m^* + s_A J_m^*$, $B = p_B I_m^* + s_B J_m^*$, then $AB = BA$.*

Remark A.1. Equivalently, if $A = p_A I_m + q_A J_m$, $B = p_B I_m + q_B J_m$, then $AB = BA$.

Proof. First note that $pI_m^* + sJ_m^*$ type of matrices are symmetric. Since I_m^* and J_m^* are projection matrices and $I_m^* J_m^* = 0$, the product of $pI_m^* + sJ_m^*$ type matrices are also in the format of $pI_m^* + sJ_m^*$ and hence symmetric. Therefore $AB = (AB)' = B'A' = BA$. □

Lemma A.2. *The determinant of* $pI_m^* + sJ_m^*$ *is*

$$|pI_m^* + sJ_m^*| = p^{m-1}s. \tag{A.1}$$

Remark A.2. Since $pI_m + qJ_m = pI_m^* + (p + mq)J_m^*$, the lemma implies that

$$|pI_m + qJ_m| = p^{m-1}(p + mq). \tag{A.2}$$

Proof. Using Proposition 31 in the appendix of Dhrymes (1978)[1],

$$\begin{aligned}
|pI_m^* + sJ_m^*| &= |pI_m + \frac{(s-p)/m}{p} J_m| \\
&= p^m |I_m + \frac{(s-p)/m}{p} \iota_m \iota_m'| \\
&= p^m |I_m| (1 + \frac{(s-p)/m}{p} \iota_m' I_m^{-1} \iota_m) \\
&= p^m (\frac{s}{p} m) = p^{m-1} s.
\end{aligned}$$

□

Lemma A.3. *The matrix* $pI_m^* + sJ_m^*$ *is nonsingular if* $p \neq 0$ *and* $s \neq 0$. *Its inverse matrix is*

$$(pI_m^* + sJ_m^*)^{-1} = \frac{1}{p} I_m^* + \frac{1}{s} J_m^*. \tag{A.3}$$

[1]The proposition states that, suppose A is an $m \times m$ matrix, a is a scalar, α and β are m dimensional vector, then

$$|A + a\alpha\beta'| = |A|(1 + a\alpha' A^{-1}\beta).$$

Remark A.3. Since $pI_m + qJ_m = pI_m^* + (p+mq)J_m^*$. The lemma is equivalent to : The matrix $pI_m + qJ_m$ is nonsingular if $p \neq 0$ and $p + mq \neq 0$. Its inverse matrix is

$$(pI_m + qJ_m)^{-1} = \frac{1}{p} I_m - \frac{q}{p(p+mq)} J_m. \tag{A.4}$$

Proof. According to Lemma A.2, $|pI_m^* + sJ_m^*| = p^{m-1}s$. So $|pI_m^* + sJ_m^*| \neq 0$ if $p \neq 0$ and $s \neq 0$. Besides,

$$[\frac{1}{p} I_m^* + \frac{1}{s} J_m^*](pI_m^* + sJ_m^*) = (pI_m^* + sJ_m^*)[\frac{1}{p} I_m^* + \frac{1}{s} J_m^*] = I_m^* + J_m^* = I_m.$$

□

Lemma A.4. *Product of matrices:*

$$\prod_{l=1}^{L} (p_l I_m^* + s_l J_m^*)^{k_l} = (\prod_{l=1}^{L} {p_l}^{k_l}) I_m^* + (\prod_{l=1}^{L} {s_l}^{k_l}) J_m^*,$$

where k_s can be any integer. If $k_l < 0$, $p_l \neq 0$ and $s_l \neq 0$.

Remark A.4. Since $pI_m + qJ_m = pI_m^* + (p+mq)J_m^*$, the lemma is equivalent to

$$\prod_{l=1}^{L} (p_l I_m + q_l J_m)^{k_l}] = \prod_{l=1}^{L} ({p_l}^{k_l}) I_m + [\prod_{l=1}^{L} (p_l + q_l m)^{k_l} - \prod_{l=1}^{L} ({p_l}^{k_l})] \frac{J_m}{m}.$$

Proof. By Lemma A.3, if $k_l < 0$,

$$\begin{aligned}
(p_l I_m^* + s_l J_m^*)^{k_l} &= [(p_l I_m^* + s_l J_m^*)^{-1}]^{|k_l|} \\
&= [\frac{1}{p_l} I_m^* + \frac{1}{s_l} J_m^*]^{|k_l|} \\
&= [p_l^{sgn(k_l)} I_m^* + s_l^{sgn(k_l)} J_m^*]^{|k_l|} \qquad \text{(A.5)}
\end{aligned}$$

where $sgn(k_l)$ is the sign function, $sgn(k_l) = 1$ if $k_l > 0$, $sgn(k_l) = 0$ if $k_l = 0$, and $sgn(k_l) = -1$ if $k_l < 0$. Equation (A.5) clearly holds when $k_l \geqslant 0$. Since I_m^* and

$J_m^*$are projection matrices, and $I_m^* J_m^* = 0$,

$$\begin{aligned}\prod_{l=1}^{L}(p_l I_m^* + s_l J_m^*)^{k_l} &= \prod_{l=1}^{L}[p_l^{sgn(k_l)} I_m^* + s_l^{sgn(k_l)} J_m^*]^{|k_l|} \\ &= (\prod_{l=1}^{L} {p_l}^{sgn(k_l)|k_l|}) I_m^* + (\prod_{l=1}^{L} {s_l}^{sgn(k_l)|k_l|}) J_m^*, \\ &= (\prod_{l=1}^{L} {p_l}^{k_l}) I_m^* + (\prod_{l=1}^{L} {s_l}^{k_l}) J_m^*.\end{aligned}$$

□

Lemma A.5. *Trace of the product of matrices:*

$$tr[\prod_{l=1}^{L}(p_l I_m^* + s_l J_m^*)^{k_l}]] = (m-1)\prod_{l=1}^{L}({p_l}^{k_l}) + \prod_{l=1}^{L} s_l^{k_l}.$$

where k_l can any integer. If $k_l < 0$, $p_l \neq 0$ and $s_l \neq 0$.

Remark A.5. Equivalently, $tr[\prod_{l=1}^{L}(p_l I_m + q_l J_m)^{k_l}] = \prod_{l=1}^{L}(p_l + q_l m)^{k_l} + (m - 1)\prod_{l=1}^{L}({p_l}^{k_l})$, where k_l can any integer.

Proof. Since $tr(I_m^*) = m-1$, $tr(J_m^*) = 1$, the lemma follows from Lemma A.4. □

Appendix B: Analytical Properties of Matrices in the Theoretical Part

In the theoretical part , most matrices are formulas of two types of matrices. The first type is block diagonal matrix with the diagonal blocks being $p_c I_c^* + s_c J_c^*$, where p_c and s_c are uniformly bounded in absolute value. For example, $W = diag_{c=1}^{R}\{-\frac{1}{n_c-1} I_c^* + J_c^*\}$. This type of matrices includes W, $\Omega(\vartheta)$, $\Omega(\vartheta)^{-1}$, $I - \lambda W$, $(I - \lambda W)^{-1}$, $\tilde{J}$ and I. The second type is $P_Z(\vartheta) = \Omega(\vartheta)^{-1} Z[Z'\Omega(\vartheta)^{-1}Z]^{-1}Z'\Omega(\vartheta)^{-1}$. Note that $M_Z(\vartheta) = \Omega(\vartheta)^{-1} - P_Z(\vartheta)$ is a formula of these two types of matrices. In this appendix, I discuss analytical properties of these matrices.

A majority of the matrices in this paper can be written as the products of some or all of the four matrices in the table below.

Table B.1: Matrices in the Form of $pI_c^* + sJ_c^*$

	$pI_c^* + sJ_c^*$	p	s
J_c	$0I_c^* + n_c J_c^*$	0	n_c
W_c	$-\frac{1}{n_c-1} I_c^* + J_c^*$	$-\frac{1}{n_c-1}$	1
Ω_c	$\sigma_\epsilon^2 I_c^* + (\sigma_\epsilon^2 + n_c\sigma_\alpha^2) J_c^*$	σ_ϵ^2	$\sigma_\epsilon^2 + n_c\sigma_\alpha^2$
$I_c - \lambda W_c$	$(1 + \frac{\lambda}{n_c-1}) I_c^* + (1-\lambda) J_c^*$	$1 + \frac{\lambda}{n_c-1}$	$1 - \lambda$

Applying the formulas for trace and determinant from Lemma A.2 and Lemma A.3, the closed form expressions for $|\Omega_c(\vartheta)|$, $\Omega_c^{-1}(\vartheta)$, $|I_c - \lambda W_c|$ and $(I_c - \lambda W_c)^{-1}$ in the concentrated log likelihood function $Q_R(\vartheta)$ in equation (3.17) are

$$|\Omega_c(\vartheta)| = (\sigma_\epsilon^2)^{n_c-1}(\sigma_\epsilon^2 + n_c\sigma_\alpha^2), \tag{B.1}$$

$$\Omega_c^{-1}(\vartheta) = \frac{1}{\sigma_\epsilon^2}I_c^* + \frac{1}{(\sigma_\epsilon^2 + n_c\sigma_\alpha^2)}J_c^*, \tag{B.2}$$

$$|I_c - \lambda W_c| = (1-\lambda)(1+\frac{\lambda}{n_c-1})^{n_c-1}, \tag{B.3}$$

$$(I_c - \lambda W_c)^{-1} = (\frac{n_c-1}{n_c-1+\lambda})I_c^* + \frac{1}{1-\lambda}J_c^*. \tag{B.4}$$

Lemma B.1. *If an $n_c \times n_c$ matrix G_c can be written in the form of*

$$G_c = J_c^{m_J}W_c^{m_W}\Omega_c(\vartheta)^{-m_\Omega}\Omega_{c0}^{m_{\Omega 0}}(I_c - \lambda W_c)^{m_S}(I_c - \lambda_0 W_c)^{-m_{S0}}, \tag{B.5}$$

where m_J, m_W, $m_{\Omega 0}$, m_Ω , m_{S0} and m_S are non-negative integers. Then

$$G_c = p_{G,c}I_c^* + s_{G,c}J_c^*, \tag{B.6}$$

$$tr(G_c) = s_{G,c} + (n_c - 1)p_{G,c}, \tag{B.7}$$

where

$$s_{G,c} = n^{m_J}1^{m_W}(\sigma_\epsilon^2 + n_c\sigma_\alpha^2)^{-m_\Omega}(\sigma_{\epsilon 0}^2 + n_c\sigma_{\alpha 0}^2)^{m_{\Omega 0}}(1-\lambda)^{m_S}(1-\lambda_0)^{-m_{S0}}, \tag{B.8}$$

$$p_{G,c} = 0^{m_J}(-\frac{1}{n_c-1})^{m_W}(\sigma_\epsilon^2)^{-m_\Omega}(\sigma_{\epsilon 0}^2)^{m_{\Omega 0}}(1+\frac{\lambda}{n_c-1})^{m_S}(1+\frac{\lambda_0}{n_c-1})^{-m_{S0}}. \tag{B.9}$$

Remark B.1. Under Assumptions 1, 2, and 3, $p_{G,c}$ and $s_{G,c}$ are uniformly bounded in absolute value.

Proof. The lemma follows from Lemma A.4, Lemma A.5 and Table B.1. □

Lemma B.2. *Let $B_{N\times m}(\vartheta) = (B_{ij,N\times m}(\vartheta))$ be an $N \times m$ matrix, where m is a finite positive integer. The elements of $B_{N\times m}(\vartheta)$ are uniformly bounded in absolute value by a finite constant $\bar{a}_B$, $sup_{\vartheta\in\Theta,1\leqslant i\leqslant N,1\leqslant j\leqslant m}|B_{ij,N\times m}(\vartheta)| \leqslant \bar{a}_B$. Let $A_N(\vartheta)$ be an $N \times N$ matrix, $A_N(\vartheta) = diag_{c=1}^R\{p_c(\vartheta)I_c^* + s_c(\vartheta)J_c^*\}$, where $p_c(\vartheta)$ and $s_c(\vartheta)$ are* *uniformly bounded in absolute value. Then under Assumptions 1, 2, and 3, the elements of $A_N(\vartheta)B_{N\times m}(\vartheta)$ are uniformly bounded in absolute value.*

Remark B.2. Here $B_{N\times m}(\vartheta)$ can be an $N\times N$ matrix like $\Omega(\vartheta)$, W and $I-\lambda W$. It can also be the $N\times k_Z$ matrix Z or the $N\times 1$ vector $Z\gamma_0$. The matrix $A_N(\vartheta)$ can be $I-\lambda W$, $(I-\lambda W)^{-1}$, W, $\Omega(\vartheta)$, $\Omega(\vartheta)^{-1}$, $\tilde{J}$ and their products and sums.

Proof. Partition $B_{N\times m}(\vartheta)$ into R blocks, $B_{N\times m}(\vartheta)=[B'_{1,N\times m}(\vartheta),...,B'_{R,N\times m}(\vartheta)]'$, where $B_{c,N}(\vartheta)$ is an $n_c\times m$ matrix, then

$$A_N(\vartheta)B_{N\times m}(\vartheta)=\begin{pmatrix}(p_1(\vartheta)I_1^*+s_1(\vartheta)J_1^*)B_{1,N\times m}(\vartheta)\\ \vdots\\ (p_R(\vartheta)I_R^*+q_R(\vartheta)J_R^*)B_{R,N\times m}(\vartheta)\end{pmatrix}. \tag{B.10}$$

For any $1\leqslant c\leqslant R$, since $p_c(\vartheta)$ and $s_c(\vartheta)$ are uniformly bounded, n_c is bounded, and the elements of $B_{c,N\times m}(\vartheta)$ are uniformly bounded in absolute value, the elements of $(p_c(\vartheta)I_c^*+s_c(\vartheta)J_c^*)B_{c,N\times m}(\vartheta)$ are uniformly bounded. Therefore, the elements of $A_N(\vartheta)B_{N\times m}(\vartheta)$ are uniformly bounded in absolute value. □

Lemma B.3. *For $l=1,2$, $A_N^{(l)}(\vartheta)=diag_{c=1}^R\{p_c^{(l)}(\vartheta)I_c^*+s_c^{(l)}(\vartheta)J_c^*\}$, where $p_c^{(l)}(\vartheta)$ and $s_c^{(l)}(\vartheta)$ are uniformly bounded in absolute value. Under Assumptions 1, 2, 3, and 5,*

$$sup_{\vartheta\in\Theta}|lim_{R\to\infty}N_R^{-1/2}tr[A_N^{(1)}(\vartheta)P_Z(\vartheta)]|=0,$$

$$sup_{\vartheta\in\Theta}|lim_{R\to\infty}N_R^{-1/2}tr[A_N^{(1)}(\vartheta)P_Z(\vartheta)A_N^{(2)}P_Z(\vartheta)]|=0.$$

Remark B.3. Since $M_Z(\vartheta)=\Omega(\vartheta)^{-1}-P_Z(\vartheta)$, the lemma implies that

$$\begin{aligned} lim_{R\to\infty}N_R^{-1}tr[A_N^{(1)}(\vartheta)M_Z(\vartheta)] &= lim_{R\to\infty}N_R^{-1}tr[A_N^{(1)}(\vartheta)\Omega(\vartheta)^{-1}],\\ lim_{R\to\infty}N_R^{-1}tr[A_N^{(1)}(\vartheta)M_Z(\vartheta)A_N^{(2)}M_Z(\vartheta)] &= lim_{R\to\infty}N_R^{-1}tr[A_N^{(1)}(\vartheta)\Omega(\vartheta)^{-1}A_N^{(2)}\Omega(\vartheta)^{-1}]|. \end{aligned}$$

Proof. By Assumption 5, the elements of $[N_R^{-1}Z'\Omega(\vartheta)^{-1}Z]^{-1}$ are uniformly bounded in absolute value. Since the number of columns in Z is $k_Z < \infty$ and the elements of Z are uniformly bounded in absolute value, the elements of $Z[\frac{1}{N_R}Z'\Omega(\vartheta)^{-1}Z]^{-1}Z'$ are uniformly bounded in absolute value. Since $\Omega(\vartheta)^{-1} = diag_{c=1}^{R}\{\sigma_\epsilon^{-1}I_c^* + (\sigma_\epsilon^2 + n_c\sigma_\alpha^2)^{-1}J_c^*\}$, σ_ϵ^{-1}, $(\sigma_\epsilon^2+n_c\sigma_\alpha^2)^{-1}$ are uniformly bounded in absolute value, by Lemma B.2 the elements of $N_R P_Z = \Omega(\vartheta)^{-1}Z[\frac{1}{N_R}Z'\Omega(\vartheta)^{-1}Z]^{-1}Z'\Omega(\vartheta)^{-1}$ are uniformly bounded in absolute value.

Since $A_N^{(l)}(\vartheta) = diag_{c=1}^{R}\{p_c^{(l)}(\vartheta)I_c^* + s_c^{(l)}(\vartheta)J_c^*\}$, $p_c^{(l)}(\vartheta)$ and $s_c^{(l)}(\vartheta)$ are uniformly bounded in absolute value, the elements of $N_R A_N^{(1)}(\vartheta)P_Z(\vartheta)$, $N_R A_N^{(1)}(\vartheta)P_Z(\vartheta)A_N^{(2)}(\vartheta)$ and $N_R A_N(\vartheta)P_Z(\vartheta)A_N^{(2)}(\vartheta)P_Z(\vartheta)$ are uniformly bounded in absolute value (Lemma B.2). Therefore $tr[(A_N(\vartheta)P_Z(\vartheta)]$, $tr[A_N(\vartheta)P_Z(\vartheta)B_N(\vartheta)]$ and $tr[A_N(\vartheta)P_Z(\vartheta)B_N(\vartheta)P_Z(\vartheta)]$ are uniformly bounded in absolute value, the lemma then follows. $\square$

Appendix C: Asymptotic Properties of Linear Quadratic Forms

In this section, I describe the asymptotic properties of linear quadratic forms of U. First, I describe the expected value and variance of the linear quadratic forms in Lemma C.1. Then I describe the central limit theorem for vectors of linear quadratic forms in Theorem C.2.

Let $S_N(\vartheta)$ be a linear quadratic form of U,

$$S_N(\vartheta) = U'A_N(\vartheta)U + U'B_N(\vartheta)Z\gamma_0. \tag{C.1}$$

Note that either $A_N(\vartheta)$ or $B_N(\vartheta)$ can to be 0. If $A_N(\vartheta) = 0_{N\times N}$, then $S_N(\vartheta)$ becomes a linear form of U. For notational simplicity, let $\eta_N(\vartheta) = B_N(\vartheta)Z\gamma_0$, then $S_N(\vartheta) = U'A_N(\vartheta)U + U'\eta_N(\vartheta)$.

Partition the $N_R \times N_R$ matrix A_N into $R \times R$ blocks, with the c, rth block being an $n_c \times n_r$ matrix. Denote the c, rth block of A_N as $A_{cr,N}$, the c, cth block of A_N as $A_{c,N}$. Denote the i, jth element of $A_{cr,N}$ as $A_{ij,cr,N}$, $1 \leqslant i, j \leqslant n_c$. Denote the $\tilde{i}, \tilde{j}$ th element of A_N as $A_{(\tilde{i},\tilde{j}),N}$, $1 \leqslant \tilde{i}, \tilde{j} \leqslant N$. I put parentheses over $\tilde{i}, \tilde{j}$ to avoids confusion of the i, jth block. I call the row sums of $A_N(\vartheta)$ are uniformly (in R) bounded in absolute value if $sup_{1\leqslant \tilde{i} \leqslant N, R\geqslant 2} \sum_{\tilde{j}=1}^{N} |A_N(\vartheta)_{(\tilde{i}\tilde{j},N)}| < \infty$. Partition the $N_R \times 1$ vector η_N into R components, with the cth component being an $n_c \times 1$ vector. Denote the cth component of η as η_c, the ith element of η_c as η_{ic}.

The following lemma describes the expected value of linear quadratic forms.

Lemma C.1. *(1) For $l = 1, 2$, let*

$$S_N^{(l)}(\vartheta) = U' A_N^{(l)}(\vartheta) U + \eta_N^{(l)\prime}(\vartheta) U.$$

Under Assumptions 1 and 2,

$$E[S_N^{(l)}(\vartheta)] = tr(\Omega_0 A_N^{(l)}(\vartheta)), l = 1, 2 \tag{C.2}$$

$$\begin{aligned} E(S_N^{(l)}(\vartheta) S_N^{(m)}(\vartheta)) = {} & tr[\Omega_0 A_N^{(l)} \Omega_0 (A_N^{(m)} + A_N^{(m)\prime})] + tr(\Omega_0 A_N^{(l)}) tr(\Omega_0 A_N^{(m)}) + \eta^{(l)\prime} \Omega_0 \eta^{(m)} \\ & + (\mu_\epsilon^{(4)} - 3\sigma_{\epsilon 0}^4) \sum_{c=1}^{R} \sum_{i=1}^{n_c} (A_{ii,c}^{(l)} A_{ii,c}^{(m)}) \\ & + (\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^4) \sum_{c=1}^{R} tr(A_c^{(l)} J_c) tr(A_c^{(m)} J_c) \\ & + \mu_\epsilon^{(3)} \sum_{c=1}^{R} \sum_{i=1}^{n_c} \eta_{ic}^{(l)} A_{ii,c}^{(m)} + \mu_\epsilon^{(3)} \sum_{c=1}^{R} \sum_{i=1}^{n_c} \eta_{ic}^{(m)} A_{ii,c}^{(l)} \\ & + \mu_\alpha^{(3)} \sum_{c=1}^{R} \eta_c^{(l)\prime} J_c A_{cc}^{(m)} \iota_c + \mu_\alpha^{(3)} \sum_{c=1}^{R} \eta_c^{(m)\prime} J_c A_c^{(l)} \iota_c, \; m, l = 1, 2 \end{aligned} \tag{C.3}$$

where $A_c^{(l)}$ is the $c-th$ diagonal block of matrix $A^{(l)}$, and is an $n_c \times n_c$ matrix, $A_{ii,c}^{(l)}$ is the $ii-th$ entry of matrix $A_{cc}^{(l)}$, $\eta_c^{(l)}$ is the cth component of $\eta^{(l)}$ and is a $n_c \times 1$ vector, $\eta_{ic}^{(l)}$ is the ith element of $\eta_c^{(l)}$.

Proof. The proof is adapted from the appendix of Kelejian and Prucha (2001) and Kelejian and Prucha (2010). For notational simplicity, I will drop the argument ϑ and subscript N from $A_N(\vartheta)$, $\eta_N(\vartheta)$, $S_N(\vartheta)$ in the proof. Under Assumptions 1 and 2, U_c and U_r are independent if $c \neq r$. Let $\alpha = (\alpha_1, ..., \alpha_R)'$ be a $R \times 1$ vector of the group effect, ϵ be the $N \times 1$ vector of individual effects $\epsilon = (\epsilon_1', \epsilon_2', ..., \epsilon_R')'$. Note that under Assumption 1 and Assumption 2, the elements of α are identically and independently distributed with $E(\alpha_c) = 0$, $Var(\alpha_c) = \sigma_{\alpha 0}^2$, $E(\alpha_c^3) = \mu_\alpha^{(3)}$, $E(\alpha_c^4) = \mu_\alpha^{(4)}$. The elements of ϵ are identically and independently distributed, with

$E(\epsilon_{ic}) = 0$, $Var(\epsilon_{ic}) = \sigma^2_{\epsilon 0}$, $E(\epsilon^3_{ic}) = \mu^{(3)}_\epsilon$, $E(\epsilon^4_{ic}) = \mu^{(4)}_\epsilon$. Also α is independent of ϵ.

Define a $N \times R$ matrix F as

$$F_{N\times R} = \begin{pmatrix} \iota_1 & 0_{n_1\times 1} & \cdots & 0_{n_1\times 1} \\ 0_{n_2\times 1} & \iota_2 & \cdots & 0_{n_2\times 1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{n_R\times 1} & 0_{n_R\times 1} & \cdots & \iota_R \end{pmatrix}. \tag{C.4}$$

Then $FF' = \tilde{J}$,

$$U = F\alpha + \epsilon. \tag{C.5}$$

For $l = 1, 2$, The linear quadratic form $S^{(l)}$ then can be divided into three linear quadratic forms of α and ϵ as

$$\begin{aligned} S^{(l)} &= U'A^{(l)}U + U'\eta^{(l)} = \alpha' F' A^{(l)} F\alpha + \epsilon' A^{(l)}\epsilon + \alpha' F'(A^{(l)} + A^{(l)\prime})\epsilon + \alpha' F'\eta^{(l)} + \epsilon'\eta^{(l)} \\ &= (\alpha' F' A^{(l)} F\alpha + \alpha' F'\eta^{(l)}) + (\epsilon' A^{(l)}\epsilon + \epsilon'\eta) + (\alpha' F'(A^{(l)} + A^{(l)\prime})\epsilon) \\ &= S^{(l)}_\alpha + S^{(l)}_\epsilon + S^{(l)}_{\alpha,\epsilon}, \end{aligned} \tag{C.6}$$

where $S^{(l)}_\alpha = \alpha' F' A^{(l)} F\alpha + \alpha' F'\eta^{(l)}$, $S_\epsilon = \epsilon' A^{(l)}\epsilon + \epsilon'\eta^{(l)}$ and $S^{(l)}_{\alpha,\epsilon} = \alpha' F'(A^{(l)} + A^{(l)\prime})\epsilon$. Since α is independent of ϵ, $S^{(l)}_\alpha$, $S^{(l)}_\epsilon$ and $S^{(l)}_{\alpha,\epsilon}$ are uncorrelated with each other for a given l. Using Lemma A.1 in appendix A of Kelejian and Prucha (2010), $E(S^{(l)}_\alpha) = \sigma^2_{\alpha 0} tr(F' A^{(l)} F) = tr(\sigma^2_{\alpha 0}\tilde{J}A^{(l)})$, $E(S^{(l)}_\epsilon) = \sigma^2_{\epsilon 0} tr(A^{(l)})$, $E(S^{(l)}_{\alpha,\epsilon}) = 0$. Therefore,

$$\begin{aligned} E(S^{(l)}) &= tr(\sigma^2_{\alpha 0}\tilde{J}A^{(l)} + \sigma^2_{\epsilon 0}A^{(l)}) \\ &= tr[(\sigma^2_{\alpha 0}\tilde{J} + \sigma^2_{\epsilon 0}I)A^{(l)}] = tr(\Omega_0 A^{(l)}). \end{aligned} \tag{C.7}$$

Since α is independent of ϵ,

$$cov(S^{(l)}, S^{(m)}) = cov(S^{(l)}_\alpha, S^{(m)}_\alpha) + cov(S^{(l)}_\epsilon, S^{(m)}_\epsilon). \tag{C.8}$$

Using Lemma A.1 in appendix A of Kelejian and Prucha (2010),

$$
\begin{aligned}
cov(S_\alpha^{(l)}, S_\alpha^{(m)}) &= \sigma_{\alpha 0}^4 tr[F'A^{(l)}FF'(A^{(m)} + A^{(m)\prime})F] + \sigma_{\alpha 0}^2 \eta^{(l)\prime} FF'\eta^{(m)} \\
&+ \sum_{c=1}^{R} [F'A^{(l)}F]_{cc}[F'A^{(m)}F]_{cc}(\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^4) \\
&+ \sum_{c=1}^{R} [(F'A^{(l)}F)_{cc}(F'\eta^{(m)})_c + (F'A^{(m)}F)_{cc}(F'\eta^{(l)})_c]\mu_\alpha^{(3)} \\
&= \sigma_{\alpha 0}^4 tr[A^{(l)}\tilde{J}(A^{(m)} + A^{(m)\prime})\tilde{J}] + \eta^{(l)\prime}(\sigma_{\alpha 0}^2 \tilde{J})\eta^{(m)} \\
&+ \sum_{c=1}^{R} tr(A_{cc}^{(l)} J_c) tr(A_{cc}^{(m)} J_c)(\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^4) \\
&+ \sum_{c=1}^{R} [(\iota_c' A_{cc}^{(l)} \iota_c \iota_c' \eta_c^{(m)}) + (\iota_c' A_{cc}^{(m)} \iota_c \iota_c' \eta_c^{(l)})]\mu_\alpha^{(3)}, \qquad \text{(C.9)}
\end{aligned}
$$

$$
\begin{aligned}
cov(S_\epsilon^{(l)}, S_\epsilon^{(m)}) &= \sigma_{\epsilon 0}^4 tr[A^{(l)}(A^{(m)} + A^{(m)\prime})] + \sigma_{\epsilon 0}^2 \eta^{(l)\prime}\eta^{(m)} \\
&+ \sum_{c=1}^{R}\sum_{i=1}^{n_c} [A_{ii,c}^{(l)} A_{ii,c}^{(m)}](\mu_\epsilon^{(4)} - 3\sigma_{\epsilon 0}^4) + \sum_{c=1}^{R}\sum_{i=1}^{n_c} [A_{ii,c}^{(l)}\eta_{ic}^{(m)} + A_{ii,c}^{(m)}\eta_{ic}^{(l)}]\mu_\epsilon^{(3)}.
\end{aligned}
\qquad \text{(C.10)}
$$

Plugging equations equation (C.7), (C.9), and (C.10) into the following equation,

$$
\begin{aligned}
E(S^{(l)}S^{(m)}) &= cov(S^{(l)}, S^{(m)}) + E(S^{(l)})E(S^{(m)}) \\
&= cov(S_\alpha^{(l)}, S_\alpha^{(m)}) + cov(S_\epsilon^{(l)}, S_\epsilon^{(m)}) + tr(\Omega_0 A^{(l)}) tr(\Omega_0 A^{(m)}).
\end{aligned}
$$

So the lemma follows.

The lemma below describes the central limit theory for vectors of linear quadratic form. $\square$

Lemma C.2. *Let $S_N(\vartheta)$ be an $L \times 1$ vector of linear quadratic forms,*

$$
S_N(\vartheta) = (S_N^{(1)}(\vartheta), S_N^{(2)}(\vartheta), .., S_N^{(L)}(\vartheta))',
$$

where $L < \infty$ is a constant, $S_N^{(l)}(\vartheta)$ is a linear quadratic form of U:

$$S_N^{(l)}(\vartheta) = U'A_N^{(l)}(\vartheta)U + U'B_N^{(l)}Z\gamma_0,$$

$A_N^{(l)}(\vartheta) = diag_{c=1}^R\{p_{A,c}^{(l)}(\vartheta)I_c^* + s_{A,c}^{(l)}(\vartheta)J_c^*\}$, $B_N^{(l)}(\vartheta) = diag_{c=1}^R\{p_{B,c}^{(l)}(\vartheta)I_c^* + s_{B,c}^{(l)}(\vartheta)J_c^*\}$, *$p_{A,c}^{(l)}$, $s_{A,c}^{(l)}$, $p_{B,c}^{(l)}$, $s_{B,c}^{(l)}$ are uniformly bounded in absolute value. If for some $\vartheta^* \in \vartheta$, $lim_{R\to\infty}N_R^{-1}Var(S_N(\vartheta^*)) = \Sigma_{\bar{S}(\vartheta^*)}$, with $0 < \underline{a}_\rho \leqslant \rho_{min}(\Sigma_{\bar{S}(\vartheta^*)})$, where $\rho_{min}(\Sigma_{\bar{S}(\vartheta^*)})$ is the smallest eigenvalue of $\Sigma_{\bar{S}(\vartheta^*)}$. Then under Assumptions 1,2, and 3, there exists a symmetric nonsingular real matrix $\Sigma_{\bar{S}(\vartheta^*)}^{1/2}$ such that $\Sigma_{\bar{S}(\vartheta^*)} = (\Sigma_{\bar{S}(\vartheta^*)}^{1/2})(\Sigma_{\bar{S}(\vartheta^*)}^{1/2})$, and $N_R^{-1}\Sigma_{\bar{S}(\vartheta^*)}^{-1/2}(S_N(\vartheta^*) - E(S_N(\vartheta^*))) \xrightarrow{D} N(0, I_L)$ as R goest to infinity.*

Proof. The proof builds on Theorem A.1 in Appendix A of Kelejian and Prucha (2010). Let

$$\xi = (\alpha_1/\sigma_{\alpha_0}, ..., \alpha_R/\sigma_{\alpha_0}, \epsilon_1'/\sigma_{\epsilon_0}, ..., \epsilon_R'/\sigma_{\epsilon 0})'. \tag{C.11}$$

The error term U is then $U = H\xi$,where $H = [\sigma_{\alpha 0}F, \sigma_{\epsilon 0}I_{N\times N}]$, F is the $N \times R$ matrix defined in equation (C.4). For $l = 1, ..., L$, let $\eta_N^{(l)}(\vartheta) = B_N^{(l)}Z\gamma_0$, the linear quadratic form $S^{(l)}(\vartheta)$ is

$$\begin{aligned} S_N^{(l)}(\vartheta) &= U'A_N^{(l)}(\vartheta)U + U'\eta_N^{(l)}(\vartheta) \\ &= \xi'[H'A_N^{(l)}(\vartheta)H]\xi + \xi'[H'\eta_N^{(l)}(\vartheta)]. \end{aligned} \tag{C.12}$$

Under Assumptions 1 and 2, ξ is a $(N + R)$ dimensional vector of independent random variables, $E(\xi) = 0$, $Var(\xi) = I_{N+R}$. There exists some $\eta_\xi > 0$ such that $E|\xi_i|^{4+\eta_\xi} < \infty$. Therefore, Assumption A.1 and A.3.(b) of Kelejian and Prucha (2010) hold. Since $A_N^{(l)}(\vartheta) = diag_{c=1}^R\{p_{A,c}^{(l)}(\vartheta)I_c^* + s_{A,c}^{(l)}(\vartheta)J_c^*$, $H'A^{(l)}H$ is symmetric. It remains to verify Assumption 2 of Kelejian and Prucha (2010): (a) the row sums of $H'A_N^{(l)}(\vartheta)H$ are uniformly (in R) bounded in absolute value,

$sup_{1\leqslant\tilde{i}\leqslant N+R, R\geqslant 2}\sum_{\tilde{j}=1}^{N+R}|(H'A_N(\vartheta)H)_{(\tilde{i},\tilde{j})}|<\infty$; (b) For all $\vartheta\in\Theta$, there exists some $\eta_b>0$ such that $sup_R\frac{1}{N_R+R}\sum_{\tilde{i}=1}^{N+R}|H'B_NZ\gamma_0|_{(\tilde{i})}^{2+\eta_b}<\infty$. Since $H=[\sigma_{\alpha 0}F,\sigma_{\epsilon 0}I_{N\times N}]$,

$$H'A_N(\vartheta)H=\begin{pmatrix}\sigma_{\alpha 0}^2F'A(\vartheta)F & \sigma_{\epsilon 0}\sigma_{\alpha 0}F'A(\vartheta)\\ \sigma_{\alpha 0}\sigma_{\epsilon 0}A(\vartheta)F & \sigma_{\epsilon 0}^2A(\vartheta)\end{pmatrix}. \tag{C.13}$$

First of all, since n_c is bounded, the row sums of A are uniformly bounded in absolute value. Second,

$$F'A(\vartheta)=\begin{pmatrix}\iota_1'A_{11} & 0 & \cdots & 0\\ 0 & \iota_2'A_{22} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \iota_R'A_{RR}\end{pmatrix}. \tag{C.14}$$

The row sums of the $F'A$ are uniformly bounded in absolute value, as the elements of A are uniformly bounded and n_c is bounded. Similarly, the row sums of the AF are uniformly bounded in absolute value. Third, $F'A^{(1)}(\vartheta)F=diag_{c=1}^{R}\{\iota_c'A_{cc}\iota_c\}$, where A_{cc} is the $c-th$ block along the diagonal of matrix A, $\iota_c'A_{cc}\iota_c$ is uniformly bounded as the elements of A_{cc} are uniformly bounded and n_c is bounded. In all, such that for all $\vartheta\in\Theta$, $sup_{\tilde{i}}\sum_{\tilde{j}=1}^{N+R}|H'A(\vartheta)H|_{(\tilde{i}\tilde{j})}<\infty$. Note that

$$H'\eta_N=[\sigma_{\alpha 0}\iota_1'\eta_{1,N},...\sigma_{\alpha 0}\iota_R'\eta_{R,N},\sigma_{\epsilon 0}\eta_N']',$$

and that n_c is bounded. Since the elements of $\eta(\vartheta)$ are uniformly bounded (Lemma B.2), the elements of $H'\eta_N$ are uniformly bounded, and for all $\vartheta\in\Theta$, $\frac{1}{N+R}\sum_{\tilde{i}=1}^{N+R}|H'B_NZ\gamma_0|_{\tilde{i}}^3<\infty$.

Using Theorem A.1 of Kelejian and Prucha (2010), $\frac{1}{\sqrt{N_R}}\Sigma_{\tilde{S}}^{-1/2}(S-\mu_S)\xrightarrow{D}N(0,I_L)$ as R goest to infinity. □

Appendix D: Asymptotic Variance-Covariance Matrix

This appendix derives the closed forms of Ψ_0 and Γ_0, which are components of the asymptotic variance-covariance matrix of $\hat{\vartheta}_R$ as described in Theorem 3.2. As shown in Section E.2, the score function the limiting matrix of the information matrix is $lim_{R\to\infty} N_R^{-1} E[f_{1R}(\vartheta_0) f_{1R}(\vartheta_0)'] = \Psi_0$, where $f_{1R}(\vartheta_0)$ is in equation (E.44). Let $G^{(1)} = \Omega_0^{-1} W(I - \lambda_0 W)^{-1}$, $G^{(2)} = \frac{1}{2}\Omega_0^{-2}$, $G^{(3)} = \frac{1}{2}\Omega_0^{-2}\tilde{J}$, $\eta^{(1)} = G^{(1)} Z\gamma_0$, $\eta^{(2)} = \eta^{(3)} = 0$, then

$$f_{1R}(\vartheta_0) = \begin{bmatrix} U'G^{(1)}U + U'\eta^{(1)} \\ U'G^{(2)}U \\ U'G^{(3)}U \end{bmatrix} - \begin{bmatrix} tr[\Omega_0 G^{(1)}] \\ tr(\Omega_0 G^{(2)}) \\ tr(\Omega_0 G^{(3)}) \end{bmatrix}.$$

Note that $G^{(1)}$, $G^{(2)}$, $G^{(3)}$ are all block diagonal matrices. For $l = 1, 2, 3$, denote the c th block of $G^{(l)}$as $G_c^{l)}$, the i, j the element of $G_c^{(l)}$ as $G_{ij,c}^{(l)}$. The $N \times 1$ vector $\eta^{(l)}$can be partitioned into R vectors, with the c the being a $n_c \times 1$ vector, $\eta^{(l)} = (\eta_1^{(l)}, ..., \eta_R^{(l)})$. Therefore, $\eta_c^{(1)} = G_c^{(1)} Z_c \gamma_0$. Denote the ith elment of $\eta_c^{(l)}$ as $\eta_{ic}^{(l)}$. By Lemma C.1, for $i, j = 1, 2, 3$,

$$
\begin{aligned}
& E(\frac{\partial Q_R(\vartheta)}{\partial \vartheta_i}\frac{\partial Q_R(\vartheta)}{\partial \vartheta_j}) \\
& = 2tr[\Omega_0 G^{(i)}\Omega_0 G^{(j)}] + \eta^{(i)\prime}\Omega_0\eta^{(j)} \\
& + (\mu_\epsilon^{(4)} - 3\sigma_{\epsilon 0}^4)\sum_{c=1}^{R}\sum_{l=1}^{n_c}(G_{ll,c}^{(i)}G_{ll,c}^{(j)}) + (\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^4)\sum_{c=1}^{R} tr(G_c^{(i)}J_c)tr(G_c^{(j)}J_c) \\
& + \mu_\epsilon^{(3)}\sum_{c=1}^{R}\sum_{l=1}^{n_c}\eta_{lc}^{(i)}G_{ll,c}^{(j)} + \mu_\epsilon^{(3)}\sum_{c=1}^{R}\sum_{l=1}^{n_c}\eta_{lc}^{(j)}G_{ll,c}^{(i)} \\
& + \mu_\alpha^{(3)}\sum_{c=1}^{R}\eta_c^{(i)\prime}J_c G_c^{(j)}\iota_c + \mu_\alpha^{(3)}\sum_{c=1}^{R}\eta_c^{(j)\prime}J_c G_c^{(i)}\iota_c. \qquad \text{(D.1)}
\end{aligned}
$$

Therefore,

$$
E[f_{1R}(\vartheta_0)f_{1R}(\vartheta_0)'] = \Xi_0 + (\mu_\epsilon^{(4)} - 3\sigma_\epsilon^4)\Xi_1 + (\mu_\alpha^{(4)} - 3\sigma_\alpha^{(4)})\Xi_2 + \mu_\epsilon^{(3)}\Xi_3 + \mu_\alpha^{(3)}\Xi_4,
$$

where Ξ_0, Ξ_1, Ξ_2, Ξ_3, Ξ_4 are defined in equations (D.2), (D.4), (D.6), (D.8), and (D.10). Their closed forms are calculated with Lemma B.1.

Under Assumptions 3 and 5, for $l = 0, 1, ..., 4$,

$$
lim_{R\to\infty}\frac{1}{N_R}\Xi_l = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*\mathit{\Xi}_{l,n},
$$

where $\mathit{\Xi}_{0,n}$, $\mathit{\Xi}_{1,n}$, $\mathit{\Xi}_{2,n}$, $\mathit{\Xi}_{3,n}$, and $\mathit{\Xi}_{4,n}$ are defined in equations (D.3), (D.5), (D.7), (D.9), (D.11). Therefore,

$$
\begin{aligned}
\Psi_0 =& lim_{R\to\infty}N_R^{-1}E[f_{1R}(\vartheta_0)f_{1R}(\vartheta_0)'] \\
=& \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*[\mathit{\Xi}_{0,n} + (\mu_\epsilon^{(4)} - 3\sigma_{\epsilon 0}^{(4)})\mathit{\Xi}_{1,n} + (\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^{(4)})\mathit{\Xi}_{2,n} + \mu_\epsilon^{(3)}\mathit{\Xi}_{3,n} + \mu_\alpha^{(3)}\mathit{\Xi}_{4,n}].
\end{aligned}
$$

Below are the expressions for Ξ_0, Ξ_1, Ξ_2, Ξ_3, Ξ_4 , $\mathit{\Xi}_{0,n}$, $\mathit{\Xi}_{1,n}$, $\mathit{\Xi}_{2,n}$, $\mathit{\Xi}_{3,n}$, and $\mathit{\Xi}_{4,n}$:

$$\Xi_0 = \begin{pmatrix} 2tr[W^2(I-\lambda_0 W)^{-2}] + \gamma_0' Z'(I-\lambda_0 W)^{-1} W \Omega_0^{-1} W (I-\lambda_0 W)^{-1} Z\gamma_0 & ... & ... \\ tr[W(I-\lambda_0 W)^{-1}\Omega_0^{-1}] & \frac{1}{2}tr(\Omega_0^{-2}) & ... \\ tr[W(I-\lambda_0 W)^{-1}\Omega_0^{-1}\tilde{J}] & \frac{1}{2}tr[\Omega_0^{-2}\tilde{J}] & \frac{1}{2}tr[\Omega_0^{-2}\tilde{J}^2] \end{pmatrix}$$

$$= \begin{pmatrix} \begin{array}{c} 2\sum_{c=1}^R \frac{n_c(n_c-1+\lambda^2)}{(1-\lambda)^2(n_c-1+\lambda)^2} + \frac{1}{\sigma_\epsilon^2}\sum_{c=1}^R (\frac{n_c-1}{n_c-1+\lambda_0})^2 \gamma_0' Z_c^{*\prime} Z_c^* \gamma_0 \\ +\frac{1}{(1-\lambda_0)^2}\sum_{c=1}^R \frac{n_c \gamma_0' \bar{Z}_c' \bar{Z}_c \gamma_0}{\sigma_\epsilon^2 + n_c \sigma_\alpha^2} \end{array} & ... & ... \\ \sum_{c=1}^R \frac{\lambda_0 n_c - (n_c-1)n_c(1-\lambda_0)\sigma_{\alpha 0}^2}{(1-\lambda)(n_c-1+\lambda)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2 + n_c\sigma_{\alpha 0}^2)} & \frac{1}{2}\sum_{c=1}^R \frac{n_c \sigma_\epsilon^4 + 2n_c(n_c-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2 + (n_c-1)n_c^2\sigma_{\alpha 0}^2}{(\sigma_{\epsilon 0}^2 + n_c\sigma_{\alpha 0}^2)^2 \sigma_{\epsilon 0}^4} & ... \\ \sum_{c=1}^R \frac{n_c}{(1-\lambda_0)(\sigma_{\epsilon 0}^2 + n_c\sigma_{\alpha 0}^2)} & \frac{1}{2}\sum_{c=1}^R \frac{n_c}{(\sigma_{\epsilon 0}^2 + n_c\sigma_{\alpha 0}^2)^2} & \frac{1}{2}\sum_{c=1}^R \frac{n_c^2}{(\sigma_{\epsilon 0}^2 + n_c\sigma_{\alpha 0}^2)^2} \end{pmatrix}, \tag{D.2}$$

$$lim_{R\to\infty}\frac{1}{N_R}\Xi_0 \quad = \quad \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* \Xi_{0,n},$$

$$\Xi_{0,n} = \begin{pmatrix} 2\frac{n(n-1+\lambda_0^2)}{(1-\lambda_0)^2(n_c-1+\lambda_0)^2} + \frac{1}{\sigma_\epsilon^2}(\frac{n-1}{n-1+\lambda_0})^2\gamma_0'\varkappa_n^*\gamma_0 + \frac{1}{(1-\lambda_0)^2}\frac{n\gamma_0'\bar{\varkappa}_n\gamma_0}{\sigma_\epsilon^2 + n\sigma_\alpha^2} & & \\ \frac{\lambda_0 n - (n-1)n(1-\lambda_0)\sigma_{\alpha 0}^2}{(1-\lambda_0)(n-1+\lambda)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)} & \frac{1}{2}\frac{n\sigma_{\epsilon 0}^4 + 2n(n-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2 + (n_c-1)n^2\sigma_{\alpha 0}^2}{(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)^2\sigma_{\epsilon 0}^4} & ... \\ \frac{n}{(1-\lambda_0)(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)} & \frac{1}{2}\frac{n}{(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)^2} & \frac{1}{2}\frac{n^2}{(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)^2} \end{pmatrix}, \tag{D.3}$$

$$
\Xi_1 = \begin{pmatrix} \sum_{c=1}^{R}\sum_{i=1}^{n_c}\{G_{ii,c}^{(1)}\}^2 & ... & ... \\ \sum_{c=1}^{R}\sum_{i=1}^{n_c} G_{ii,c}^{(1)}G_{ii,c}^{(2)} & \sum_{c=1}^{R}\sum_{i=1}^{n_c}[G_{ii,c}^{(2)}]^2 & ... \\ \sum_{c=1}^{R}\sum_{i=1}^{n_c} G_{ii,c}^{(1)}G_{ii,c}^{(3)} & \sum_{c=1}^{R}\sum_{i=1}^{n_c} G_{ii,c}^{(2)}G_{ii,c}^{(3)} & \sum_{c=1}^{R}\sum_{i=1}^{n_c}[G_{ii,c}^{(3)}]^2 \end{pmatrix},
$$

$$
= \begin{pmatrix} \sum_{c=1}^{R} n_c[\frac{\lambda_0\sigma_{\epsilon 0}^2-(n_c-1)(1-\lambda_0)\sigma_{\alpha 0}^2}{(1-\lambda_0)(n_c-1+\lambda_0)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)}]^2 & ... & ... \\ \sum_{c=1}^{R}\frac{n_c[\lambda_0\sigma_{\epsilon 0}^2-(n_c-1)(1-\lambda_0)\sigma_{\alpha 0}^2][\sigma_{\epsilon 0}^4+(n_c^2-n_c)\sigma_{\alpha 0}^4+2(n_c-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2]}{2(1-\lambda_0)(n_c-1+\lambda_0)\sigma_{\epsilon 0}^6(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^3} & \sum_{c=1}^{R}\frac{n_c[\sigma_{\epsilon 0}^4+(n_c^2-n_c)\sigma_{\alpha 0}^4+2(n_c-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2]^2}{4\sigma_{\epsilon 0}^8(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^4} & ... \\ \sum_{c=1}^{R}\frac{n_c[\lambda_0\sigma_{\epsilon 0}^2-(n_c-1)(1-\lambda_0)\sigma_{\alpha 0}^2]}{2(1-\lambda_0)(n_c-1+\lambda_0)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^3} & \sum_{c=1}^{R}\frac{n_c[\sigma_{\epsilon 0}^4+(n_c^2-n_c)\sigma_{\alpha 0}^4+2(n_c-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2]}{4\sigma_{\epsilon 0}^4(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^4} & \sum_{c=1}^{R}\frac{n_c}{4(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^4} \end{pmatrix}, \tag{D.4}
$$

$$
lim_{R\to\infty}\frac{1}{N_R}\Xi_1 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*\Xi_{1,n},
$$

$$
\Xi_{1,n} = \begin{pmatrix} n[\frac{\lambda_0\sigma_{\epsilon 0}^2-(n-1)(1-\lambda_0)\sigma_{\alpha 0}^2}{(1-\lambda_0)(n-1+\lambda_0)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)}]^2 & ... & ... \\ \frac{n[\lambda_0\sigma_{\epsilon 0}^2-(n-1)(1-\lambda_0)\sigma_{\alpha 0}^2][\sigma_{\epsilon 0}^4+(n^2-n)\sigma_{\alpha 0}^4+2(n-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2]}{2(1-\lambda_0)(n-1+\lambda_0)\sigma_{\epsilon 0}^6(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^3} & \frac{n[\sigma_{\epsilon 0}^4+(n^2-n)\sigma_{\alpha 0}^4+2(n-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2]^2}{4\sigma_{\epsilon 0}^8(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^4} & ... \\ \frac{n[\lambda_0\sigma_{\epsilon 0}^2-(n-1)(1-\lambda_0)\sigma_{\alpha 0}^2]}{2(1-\lambda_0)(n-1+\lambda_0)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^3} & \frac{n[\sigma_{\epsilon 0}^4+(n^2-n)\sigma_{\alpha 0}^4+2(n-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2]}{4\sigma_{\epsilon 0}^4(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^4} & \frac{n}{4(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^4} \end{pmatrix}. \tag{D.5}
$$

$$\Xi_2 = \begin{pmatrix} \sum_{c=1}^R tr(G_c^{(1)}J_c)^2 & ... & ... \\ \sum_{c=1}^R tr(G_c^{(1)}J_c)tr(G_c^{(2)}J_c) & \sum_{c=1}^R tr(G_c^{(2)}J_c)^2 & ... \\ \sum_{c=1}^R tr(G_c^{(1)}J_c)tr(G_c^{(3)}J_c) & \sum_{c=1}^R tr(G_c^{(2)}J_c)tr(G_c^{(3)}J_c) & \sum_{c=1}^R tr(G_c^{(3)}J_c)^2 \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{c=1}^R \frac{n_c^2}{(1-\lambda_0)^2(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^2} & ... & ... \\ \sum_{c=1}^R \frac{1}{2}\frac{n_c^2}{(1-\lambda_0)(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^3} & \sum_{c=1}^R \frac{1}{2}\frac{n_c^2}{(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^4} & ... \\ \sum_{c=1}^R \frac{1}{2}\frac{n_c^3}{(1-\lambda_0)(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^3} & \sum_{c=1}^R \frac{1}{4}\frac{n_c^3}{(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^4} & \sum_{c=1}^R \frac{1}{4}\frac{n_c^4}{(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^4} \end{pmatrix}, \quad \text{(D.6)}$$

$$lim_{R\to\infty}\frac{1}{N_R}\Xi_2 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*\Xi_{2,n},$$

$$\Xi_{2,n} = \begin{pmatrix} \frac{n^2}{(1-\lambda_0)^2(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^2} & ... & ... \\ \frac{n_c^2}{(1-\lambda_0)(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^3} & \frac{1}{2}\frac{n^2}{(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^4} & ... \\ \frac{n_c^3}{(1-\lambda_0)(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^3} & \frac{1}{4}\frac{n^3}{(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^4} & \frac{1}{4}\frac{n^4}{(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^4} \end{pmatrix}, \quad \text{(D.7)}$$

$$\Xi_3 = \begin{pmatrix} 2\sum_{c=1}^R\sum_{i=1}^{n_c}[G^{(1)}Z\gamma_0]_{ic}G_{ii,c}^{(1)} & ... & ... \\ \sum_{c=1}^R\sum_{i=1}^{n_c}[G^{(1)}Z\gamma_0]_{ic}G_{ii,c}^{(2)} & 0 & ... \\ \sum_{c=1}^R\sum_{i=1}^{n_c}[G^{(1)}Z\gamma_0]_{ic}G_{ii,c}^{(3)} & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 2\sum_{c=1}^R \frac{[\lambda_0\sigma_{\epsilon 0}^2-(n_c-1)(1-\lambda_0)\sigma_{\alpha 0}^2](\iota_c' Z_c\gamma_0)}{(1-\lambda_0)^2(n_c-1+\lambda_0)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^2} & ... & ... \\ \sum_{c=1}^R \frac{[\sigma_{\epsilon 0}^4+(n_c^2-n_c)\sigma_{\alpha 0}^4+2(n_c-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2](\iota_c' Z_c\gamma_0)}{2(1-\lambda_0)\sigma_{\epsilon 0}^4(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^3} & 0 & ... \\ \sum_{c=1}^R \frac{\iota_c' Z_c\gamma_0}{2(1-\lambda_0)(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)^3} & 0 & 0 \end{pmatrix}, \quad \text{(D.8)}$$

$$lim_{R\to\infty}\frac{1}{N_R}\Xi_3 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*\Xi_{3,n},$$

$$\Xi_{3,n} = \begin{pmatrix} 2\frac{[\lambda_0\sigma_{\epsilon 0}^2-(n-1)(1-\lambda_0)\sigma_{\alpha 0}^2]n\underline{Z}_n\gamma_0}{(1-\lambda_0)^2(n-1+\lambda_0)\sigma_{\epsilon 0}^2(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^2} & ... & ... \\ \frac{[\sigma_{\epsilon 0}^4+(n^2-n)\sigma_{\alpha 0}^4+2(n-1)\sigma_{\epsilon 0}^2\sigma_{\alpha 0}^2]n(\underline{Z}_n\gamma_0)}{2(1-\lambda_0)\sigma_{\epsilon 0}^4(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^3} & 0 & ... \\ \frac{n\underline{Z}_n\gamma_0}{2(1-\lambda_0)(\sigma_{\epsilon 0}^2+n\sigma_{\alpha 0}^2)^3} & 0 & 0 \end{pmatrix}, \quad \text{(D.9)}$$

$$\Xi_4 = \begin{pmatrix} 2\sum_{c=1}^{R}\gamma_0' Z_c' G_c^{(1)} J_c G_c^{(1)} \iota_c & ... & ... \\ \sum_{c=1}^{R}\gamma_0' Z_c' G_c^{(1)} J_c G_c^{(2)} \iota_c & 0 & ... \\ \sum_{c=1}^{R}\gamma_0' Z_c' G_c^{(1)} J_c G_c^{(3)} \iota_c & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 2\sum_{c=1}^{R}\frac{n_c \iota_c' Z_c \gamma_0}{(1-\lambda_0)^2(\sigma_{\epsilon 0}^2 + n_c \sigma_{\alpha 0}^2)^2} & ... & ... \\ \sum_{c=1}^{R}\frac{n_c \iota_c' Z_c \gamma_0}{2(1-\lambda_0)(\sigma_{\epsilon 0}^2 + n_c \sigma_{\alpha 0}^2)^3} & 0 & ... \\ \sum_{c=1}^{R}\frac{n_c^2 \iota_c' Z_c \gamma_0}{2(1-\lambda_0)(\sigma_{\epsilon 0}^2 + n_c \sigma_{\alpha 0}^2)^3} & 0 & 0 \end{pmatrix}, \tag{D.10}$$

$$lim_{R\to\infty}\frac{1}{N_R}\Xi_4 \quad = \quad \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* \Xi_{4,n},$$

$$\Xi_{4,n} = \begin{pmatrix} 2\frac{n^2 \underline{Z}_n \gamma_0}{(1-\lambda_0)^2(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)^2} & ... & ... \\ \frac{n^2 \underline{Z}_n \gamma_0}{2(1-\lambda_0)(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)^3} & 0 & ... \\ \frac{n_c^3 \underline{Z}_n \gamma_0}{2(1-\lambda_0)(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)^3} & 0 & 0 \end{pmatrix}. \tag{D.11}$$

Appendix E: Proof of Theorem 3.1 and Theorem 3.2

E.1 Proof of Theorem 3.1

To prove the consistency of $\hat{\vartheta}_R$ in Theorem 3.1, if suffices to prove Proposition E.1 on this page, Proposition E.2 on page 108 and Proposition E.3 on page 114. [1]

Proposition E.1. $\bar{Q}(\vartheta) = lim_{R\to\infty} N_R^{-1} E Q_R(\vartheta)$ *exists and is finite and continuous on* Θ.

Proof. Using Lemma C.1, the expected value of the the concentrated log likelihood function $Q_R(\vartheta)$ in equation (3.17) is

$$E(Q_R(\vartheta)) = Q_R^{(1)}(\vartheta) + Q_R^{(2)}(\vartheta),$$

where[2]

$$\begin{aligned} Q_R^{(1)}(\vartheta) &= -\frac{N}{2} ln(2\pi) + ln|I - \lambda W| - \frac{1}{2} ln|\Omega(\vartheta)| \\ &- \frac{1}{2} tr[(I - \lambda_0 W)^{-2} (I - \lambda W)^2 M_Z(\vartheta) \Omega_0], \end{aligned} \tag{E.1}$$

$$Q_R^{(2)}(\vartheta) = -\frac{1}{2} \gamma_0' Z' (I - \lambda_0 W)^{-1} (I - \lambda W) M_Z(\vartheta) (I - \lambda W)(I - \lambda_0 W)^{-1} Z \gamma_0 \tag{E.2}$$

[1] See Lemma 3.1 in Pötscher and Prucha (1991).

[2] Note that by the quasi-commutative properties of matrix trace and the commutative property of $pI_m + sJ_m$type of matrices,

$$tr[(I - \lambda_0 W)^{-1} (I - \lambda W) M_Z(\vartheta) \Omega_0 (I - \lambda W)(I - \lambda_0 W)^{-1}] = tr[(I - \lambda_0 W)^{-2} (I - \lambda W)^2 M_Z(\vartheta) \Omega_0].$$

Next I will show that $lim_{R\to\infty} N_R^{-1} Q_R^{(1)}(\vartheta)$ and $lim_{R\to\infty} N_R^{-1} Q_R^{(2)}(\vartheta)$ both exist and are finite. To calculate $lim_{R\to\infty} N_R^{-1} Q_R^{(1)}(\vartheta)$, first note that by Lemma B.3,

$$\begin{aligned} & lim_{R\to\infty}\frac{1}{N_R}tr[(I-\lambda_0 W)^{-2}(I-\lambda W)^2 M_Z(\vartheta)\Omega_0] \\ = \quad & lim_{R\to\infty}\frac{1}{N_R}tr[(I-\lambda_0 W)^{-2}(I-\lambda W)^2 \Omega(\vartheta)^{-1}\Omega_0]. \end{aligned}$$

Therefore,

$$\begin{aligned} lim_{R\to\infty}\frac{1}{N_R}Q_R^{(1)}(\vartheta) \quad = \quad & lim_{R\to\infty}\frac{1}{N_R}\{-\frac{N}{2}ln(2\pi)+ln|I-\lambda W|-\frac{1}{2}ln|\Omega(\vartheta)| \\ - \quad & \frac{1}{2}tr[(I-\lambda_0 W)^{-2}(I-\lambda W)^2\Omega(\vartheta)^{-1}\Omega_0]\}. \end{aligned}$$

Calculating the closed form expression for the trace above with Lemma B.1, and using the closed form expression for $|I-\lambda W|$, $(I-\lambda W)$, $|\Omega(\vartheta)|$, and $\Omega(\vartheta)^{-1}$ in equations (B.1), (B.2), (B.3), and (B.4),

$$\begin{aligned} & lim_{R\to\infty}\frac{1}{N_R}Q_R^{(1)}(\vartheta) \\ =& lim_{R\to\infty}\{-\frac{1}{2}ln(2\pi)+\frac{R}{N_R}ln(1-\lambda)+\frac{1}{N_R}\sum_{c=1}^{R}[(n_c-1)ln(n_c-1+\lambda)] \\ -&\frac{N_R-R}{2N_R}ln(\sigma_\epsilon^2)-\frac{1}{2N_R}\sum_{c=1}^{R}ln(\sigma_\alpha^2+\frac{1}{n_c}\sigma_\epsilon^2)-\frac{(1-\lambda)^2}{2N_R(1-\lambda_0)^2}\sum_{c=1}^{R}\frac{(\sigma_{\epsilon 0}^2+n_c\sigma_{\alpha 0}^2)}{(\sigma_\epsilon^2+n_c\sigma_\alpha^2)} \\ -&\frac{\sigma_{\epsilon 0}^2}{2N_R\sigma_\epsilon^2}\sum_{c=1}^{R}\frac{(n_c-1)(n_c-1+\lambda)^2}{(n_c-1+\lambda_0)^2}\}. \end{aligned}$$

Recall that the number of groups with size n is R_n. Under Assumption 3, $\underline{a} \leqslant n_c \leqslant \bar{a}$ and $lim_{R\to\infty} R_n/R = \omega_n^*$. The limit of average group size is $lim_{R\to\infty} N_R/R =$

n^*. For $n = \underline{a}, ..., \bar{a}$, $lim_{R\to\infty} N_R/R_n = \omega_n^*/n^*$.In all,

$$
\begin{aligned}
& lim_{R\to\infty}\frac{1}{N_R}E(Q_R^{(1)}(\vartheta)) \\
=& lim_{R\to\infty}\{-\frac{1}{2}ln(2\pi) + \frac{R}{N_R}ln(1-\lambda) + \frac{1}{N_R}\sum_{n=\underline{a}}^{\bar{a}} R_n[(n-1)ln(n-1+\lambda)] \\
& -\frac{1}{2N_R}\sum_{n=\underline{a}}^{\bar{a}} R_n(n-1)ln(\sigma_\epsilon^2) - \frac{1}{2N_R}\sum_{n=\underline{a}}^{\bar{a}} R_n ln(\sigma_\alpha^2 + \frac{1}{n}\sigma_\epsilon^2) \\
& -\frac{(1-\lambda)^2}{2N_R(1-\lambda_0)^2}\sum_{n=\underline{a}}^{\bar{a}} R_n\frac{(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)}{(\sigma_\epsilon^2 + n\sigma_\alpha^2)} - \frac{\sigma_{\epsilon 0}^2}{2N_R\sigma_\epsilon^2}\sum_{n=\underline{a}}^{\bar{a}} R_n\frac{(n-1)(n-1+\lambda)^2}{(n-1+\lambda_0)^2}\}. \\
=& -\frac{1}{2}ln(2\pi) + \frac{1}{n^*}ln(1-\lambda) + \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^*[(n-1)ln(n-1+\lambda)] \\
& -\frac{n^*-1}{2n^*}ln(\sigma_\epsilon^2) - \frac{1}{2n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* ln(\sigma_\alpha^2 + \frac{1}{n}\sigma_\epsilon^2) - \frac{(1-\lambda)^2}{2n^*(1-\lambda_0)^2}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^*\frac{(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)}{(\sigma_\epsilon^2 + n\sigma_\alpha^2)} \\
& -\frac{\sigma_{\epsilon 0}^2}{2n^*\sigma_\epsilon^2}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^*\frac{(n-1)(n-1+\lambda)^2}{(n-1+\lambda_0)^2}\} \\
=& -\frac{1}{2}ln(2\pi) + \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* g(n,\vartheta),
\end{aligned}
\tag{E.3}
$$

where

$$
\begin{aligned}
g(n,\vartheta) =& ln(1-\lambda) + (n-1)ln(n-1+\lambda) - \frac{n-1}{2}ln(\sigma_\epsilon^2) - \frac{1}{2}ln(\sigma_\alpha^2 + \frac{1}{n}\sigma_\epsilon^2) \\
& - \frac{1}{2}\frac{(1-\lambda)^2}{(1-\lambda_0)^2}\frac{(\sigma_{\epsilon 0}^2 + n\sigma_{\alpha 0}^2)}{(\sigma_\epsilon^2 + n\sigma_\alpha^2)} - \frac{(n-1)}{2}\frac{(n-1+\lambda)^2}{(n-1+\lambda_0)^2}\frac{\sigma_{\epsilon 0}^2}{\sigma_\epsilon^2}.
\end{aligned}
\tag{E.4}
$$

Under Assumptions 1, 2, 3, and 4, $g(n,\vartheta)$ is finite and continuous, so $lim_{R\to\infty} N_R^{-1}E(Q_R^{(1)}(\vartheta))$ is finite and continuous.

Next I will calculate $lim_{R\to\infty} N_R^{-1}Q_R^{(2)}(\vartheta)$. With $Q_R^{(2)}$ in equation (E.2), and $M_Z(\vartheta)$ in equation (3.18),

$$
\begin{aligned}
& lim_{R\to\infty}\frac{1}{N_R}Q_R^{(2)}(\vartheta) \\
=& lim_{R\to\infty}\gamma_0' Z'(I-\lambda_0 W)^{-1}(I-\lambda W)\Omega(\vartheta)^{-1}(I-\lambda W)(I-\lambda_0 W)^{-1}Z\gamma_0 \\
-& \gamma_0'\{[lim_{R\to\infty}\frac{1}{N_R}Z'(I-\lambda_0 W)^{-1}(I-\lambda W)\Omega(\vartheta)^{-1}Z][lim_{R\to\infty}\frac{1}{N_R}Z'\Omega(\vartheta)^{-1}Z]^{-1} \\
\times & [lim_{R\to\infty}\frac{1}{N_R}Z'\Omega(\vartheta)^{-1}(I-\lambda W)(I-\lambda_0 W)^{-1}Z]\}\gamma_0. \qquad \text{(E.5)}
\end{aligned}
$$

Using Lemma B.1,

$$
(I_c-\lambda_0 W_c)^{-1}(I_c-\lambda W_c)\Omega_c(\vartheta)^{-1} = \frac{1}{\sigma_\epsilon^2}(\frac{n_c-1+\lambda}{n_c-1+\lambda_0})I_c^* + \frac{1-\lambda}{1-\lambda_0}\frac{1}{\sigma_\epsilon^2+n_c\sigma_\alpha^2}J_c^*,
$$

$$
(I_c-\lambda_0 W_c)^{-2}(I_c-\lambda W_c)^2\Omega_c(\vartheta)^{-1} = \frac{1}{\sigma_\epsilon^2}(\frac{n_c-1+\lambda}{n_c-1+\lambda_0})^2 I_c^* + (\frac{1-\lambda}{1-\lambda_0})^2\frac{1}{\sigma_\epsilon^2+n_c\sigma_\alpha^2}J_c^*.
$$

Also note that $Z_c'I_c^*Z_c = Z_c^{*\prime}Z_c^*$ and $Z_c'J_c^*Z_c = n_c\bar{Z}_c'\bar{Z}_c$. Therefore,

$$
Z'(I-\lambda_0 W)^{-1}(I-\lambda W)\Omega(\vartheta)^{-1}Z = \frac{1}{\sigma_\epsilon^2}\sum_{c=1}^{R}(\frac{n_c-1+\lambda}{n_c-1+\lambda_0})Z_c^{*\prime}Z_c^* + \frac{1-\lambda}{1-\lambda_0}\sum_{c=1}^{R}\frac{n_c\bar{Z}_c'\bar{Z}_c}{\sigma_\epsilon^2+n_c\sigma_\alpha^2},
$$

$$
Z'(I-\lambda W)_0^{-2}(I-\lambda W)^2\Omega(\vartheta)^{-1}Z = \frac{1}{\sigma_\epsilon^2}\sum_{c=1}^{R}(\frac{n_c-1+\lambda}{n_c-1+\lambda_0})^2 Z_c^{*\prime}Z_c^* + (\frac{1-\lambda}{1-\lambda_0})^2\sum_{c=1}^{R}\frac{n_c\bar{Z}_c'\bar{Z}_c}{\sigma_\epsilon^2+n_c\sigma_\alpha^2}.
$$

I have shown in Section 3.2 that $lim_{R\to\infty}N_R^{-1}Z'\Omega(\vartheta)^{-1}Z$ exists and has a closed form expression in equation (3.12). By Assumption 3, $lim_{R\to\infty}R_n/N_R = \omega_n^*/n^*$. By Assumption 5, $lim_{R\to\infty}R_n^{-1}\sum_{c\in\mathcal{I}_n}Z_c^{*\prime}Z_c^* = \varkappa_n^*$, $lim_{R\to\infty}R_n^{-1}\sum_{c\in\mathcal{I}_n}\bar{Z}_c'\bar{Z}_c = \bar{\varkappa}_n$, where $\varkappa_n^*$ and $\bar{\varkappa}_n$ are both finite. In all,

$$
\begin{aligned}
& lim_{R\to\infty}\frac{1}{N_R}Z'(I-\lambda W)_0^{-1}(I-\lambda W)\Omega(\vartheta)^{-1}Z \\
=& lim_{R\to\infty}\frac{1}{N_R}\frac{1}{\sigma_\epsilon^2}\sum_{n=\underline{a}}^{\bar{a}}R_n(\frac{n-1+\lambda}{n-1+\lambda_0})(\frac{1}{R_n}\sum_{c\in\mathcal{I}_n}Z_c^{*\prime}Z_c^*) \\
+& lim_{R\to\infty}\frac{1}{N_R}\frac{1-\lambda}{1-\lambda_0}\sum_{n=\underline{a}}^{\bar{a}}R_n\frac{n}{\sigma_\epsilon^2+n\sigma_\alpha^2}(\frac{1}{R_n}\sum_{c\in\mathcal{I}_n}\bar{Z}_c'\bar{Z}_c) \\
=& \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*[(\frac{n-1+\lambda}{n-1+\lambda_0})\frac{\varkappa_n^*}{\sigma_\epsilon^2} + (\frac{1-\lambda}{1-\lambda_0})\frac{n\bar{\varkappa}_n}{\sigma_\epsilon^2+n\sigma_\alpha^2}], \qquad \text{(E.6)}
\end{aligned}
$$

$$
\begin{aligned}
&lim_{R\to\infty}\frac{1}{N_R}Z'(I-\lambda W)_0^{-2}(I-\lambda W)^2\Omega(\vartheta)^{-1}Z\\
=&lim_{R\to\infty}\frac{1}{N_R}\frac{1}{\sigma_\epsilon^2}\sum_{n=\underline{a}}^{\bar{a}}R_n(\frac{n-1+\lambda}{n-1+\lambda_0})^2(\frac{1}{R_n}\sum_{c\in\mathcal{I}_n}Z_c^{*\prime}Z_c^*)\\
+&lim_{R\to\infty}\frac{1}{N_R}(\frac{1-\lambda}{1-\lambda_0})^2\sum_{n=\underline{a}}^{\bar{a}}R_n\frac{n}{\sigma_\epsilon^2+n\sigma_\alpha^2}(\frac{1}{R_n}\sum_{c\in\mathcal{I}_n}\bar{Z}_c'\bar{Z}_c)\\
=&\frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*[(\frac{n-1+\lambda}{n-1+\lambda_0})^2\frac{\varkappa_n^*}{\sigma_\epsilon^2}+(\frac{1-\lambda}{1-\lambda_0})^2\frac{n\bar{\varkappa}_n}{\sigma_\epsilon^2+n\sigma_\alpha^2}].
\end{aligned}
\tag{E.7}
$$

Plugging equations (3.12), (E.6), and (E.7) into equation (E.5),

$$
\begin{aligned}
&lim_{R\to\infty}\frac{1}{N_R}Q_R^{(2)}(\vartheta)\\
=&\frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*[(\frac{n-1+\lambda}{n-1+\lambda_0})^2\frac{\gamma_0'\varkappa_n^*\gamma_0}{\sigma_\epsilon^2}+(\frac{1-\lambda}{1-\lambda_0})^2\frac{n_c\gamma_0'\bar{\varkappa}_n\gamma_0}{\sigma_\epsilon^2+n_c\sigma_\alpha^2}]\\
-&\frac{1}{n^*}\gamma_0'\{\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*[(\frac{n-1+\lambda}{n-1+\lambda_0})\frac{\varkappa_n^*}{\sigma_\epsilon^2}+(\frac{1-\lambda}{1-\lambda_0})\frac{n_c\bar{\varkappa}_n}{\sigma_\epsilon^2+n_c\sigma_\alpha^2}]\}\{\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*(\frac{\varkappa_n^*}{\sigma_\epsilon^2}+\frac{n_c\bar{\varkappa}_n}{\sigma_\epsilon^2+n_c\sigma_\alpha^2})\}^{-1}\\
\times&\{\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*[(\frac{n-1+\lambda}{n-1+\lambda_0})\frac{\varkappa_n^*}{\sigma_\epsilon^2}+(\frac{1-\lambda}{1-\lambda_0})\frac{n_c\bar{\varkappa}_n}{\sigma_\epsilon^2+n_c\sigma_\alpha^2}]\}\gamma_0.
\end{aligned}
\tag{E.8}
$$

Since $\varkappa_n^*$, $\bar{\varkappa}_n$ and n are finite, $lim_{R\to\infty}N_R^{-1}E(Q_R^{(2)}(\vartheta))$ is finite and continous. In all, $\bar{Q}(\vartheta)=lim_{R\to\infty}N_R^{-1}E(Q_R(\vartheta))$ exists and is finite and continuous on Θ. □

Proposition E.2. *The parameter space Θ is compact and ϑ_0 is the unique maximizer of $\bar{Q}(\vartheta)$ on Θ.*

The compactness of Θ follows from Assumptions 1, 2, and 4. Since $\bar{Q}(\vartheta)=lim_{R\to\infty}N_R^{-1}[Q_R^{(1)}(\vartheta)+Q_R^{(2)}(\vartheta)]$, and the closed form for $lim_{R\to\infty}N_R^{-1}Q_R^{(1)}(\vartheta)$ is in equation (E.3),

$$
\bar{Q}(\vartheta)=-\frac{1}{2}ln(2\pi)+\frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*g(n,\vartheta)+lim_{R\to\infty}N_R^{-1}Q_R^{(2)}(\vartheta).
\tag{E.9}
$$

To prove Proposition E.2, it then suffices to show Lemma E.1, Lemma E.2 and Lemma E.3 below.

Lemma E.1. *Under Assumptions 1, 2, 3, and 5, ϑ_0 is a global maximizer of $Q_R^{(2)}(\vartheta)$ on Θ, $Q_R^{(2)}(\vartheta)) \leqslant Q_R^{(2)}(\vartheta_0))$ for any $\vartheta \in \Theta$.*

Remark E.1. As shown above, $lim_{R\to\infty} N_R^{-1} E(Q_R^{(2)}(\vartheta))$ exists and is finite. The lemma therefore implies that $lim_{R\to\infty} N_R^{-1} Q_R^{(2)}(\vartheta)) \leqslant lim_{R\to\infty} N_R^{-1} Q_R^{(2)}(\vartheta_0))$ for all $\vartheta \in \Theta$.

Proof. Define a projection matrix $\tilde{M}_Z(\vartheta)$ as

$$\tilde{M}_Z(\vartheta) = I - \Omega^{-1/2} Z'(Z'\Omega(\vartheta)Z)^{-1} Z\Omega^{-1/2}.$$

Then $M_Z(\vartheta) = \Omega(\vartheta)^{-1/2}\tilde{M}_Z(\vartheta)\Omega^{-1/2}(\vartheta)$, $Q_R^{(2)}(\vartheta)$ can be rewritten as

$$Q_R^{(2)}(\vartheta) = -\frac{1}{2}\eta_Z(\vartheta)'\tilde{M}_Z(\vartheta)\eta_Z(\vartheta),$$

where

$$\eta_Z(\vartheta) = \Omega(\vartheta)^{-1/2}(I - \lambda W)(I - \lambda_0 W)^{-1} Z\gamma_0.$$

Since $\tilde{M}_Z(\vartheta)$ is a projection matrix, it is positive semidefinite,

$$\begin{aligned}\eta(\vartheta)'\tilde{M}_Z(\vartheta)\eta(\vartheta) &= \eta(\vartheta)'\tilde{M}_Z'(\vartheta)\tilde{M}_Z(\vartheta)\eta(\vartheta)\\ &= [\tilde{M}_Z(\vartheta)\eta(\vartheta)]'[\tilde{M}_Z(\vartheta)\eta(\vartheta)]\\ &\geqslant 0.\end{aligned}$$

Therefore $Q_R^{(2)}(\vartheta) \leqslant 0$. Also note that

$$Q_R^{(2)}(\vartheta_0) = \gamma_0' Z' M_Z(\vartheta_0) Z\gamma_0 = 0.$$

Therefore, ϑ_0 is a global maximizer of $Q_R^{(2)}(\vartheta)$ on Θ. □

Note that ϑ_0 is not the unique global maximizer of $Q_R^{(2)}(\vartheta_0)$. For $\vartheta^* = (\lambda^*, \sigma_\epsilon^{2*}, \sigma_\alpha^{2*}) \in \Theta$, $Q_R^{(2)}(\vartheta^*) = 0$ as long as $\lambda^* = \lambda_0$.

Lemma E.2. *Under Assumptions 1-5, ϑ_0 is a global maximizer of $g(n, \vartheta)$ on Θ for $n = \underline{a}, ..., \bar{a}$.*

Proof. It can be shown that the Hessian matrix $\partial^2 g(n,\vartheta)/\partial\vartheta\partial\vartheta'$ is singular, $g(n,\vartheta)$ is flat on a two-dimensional surface $(\zeta_1(n,\vartheta), \zeta_2(n,\vartheta))$ defined on the three dimensional parameter space Θ, where

$$\zeta_1(n,\vartheta) = \frac{(1-\lambda)^2}{(\sigma_\epsilon^2 + n\sigma_\alpha^2)}, \tag{E.10}$$

$$\zeta_2(n,\vartheta) = \frac{(n-1+\lambda)^2}{\sigma_\epsilon^2}. \tag{E.11}$$

Note that $\zeta_1(n,\vartheta) > 0$ and $\zeta_2(n,\vartheta) > 0$,

$$g(n,\vartheta) = \frac{1}{2} ln\zeta_1(n,\vartheta) + \frac{(n-1)}{2} ln\zeta_2(n,\vartheta) - \frac{1}{2}\frac{\zeta_1(n,\vartheta)}{\zeta_1(n,\vartheta_0)} - \frac{n-1}{2}\frac{\zeta_2(n,\vartheta)}{\zeta_2(n,\vartheta_0)}. \tag{E.12}$$

The first order derivates of $g(n,\vartheta)$ with respect to $\zeta_1(n,\vartheta)$ and $\zeta_2(n,\vartheta)$ are

$$\frac{\partial g(n,\vartheta)}{\partial\zeta_1(n,\vartheta)} = \frac{1}{2}(\frac{1}{\zeta_1(n,\vartheta)} - \frac{1}{\zeta_1(n,\vartheta_0)}), \tag{E.13}$$

$$\frac{\partial g(n,\vartheta)}{\partial\zeta_2(n,\vartheta)} = \frac{n-1}{2}(\frac{1}{\zeta_2(n,\vartheta)} - \frac{1}{\zeta_2(n,\vartheta_0)}). \tag{E.14}$$

The first order conditions are $\zeta_1(n,\vartheta) = \zeta_1(n,\vartheta_0)$ and $\zeta_2(n,\vartheta) = \zeta_2(n,\vartheta_0)$.

The second derivatives are

$$\frac{\partial^2 g(n,\vartheta)}{\partial\zeta_1(n,\vartheta)^2} = -\frac{1}{2}\frac{1}{\zeta_1(n,\vartheta)^2} < 0, \tag{E.15}$$

$$\frac{\partial^2 g(n,\vartheta)}{\partial\zeta_2(n,\vartheta)^2} = -\frac{n-1}{2}\frac{1}{\zeta_2(n,\vartheta)^2} < 0, \tag{E.16}$$

$$\frac{\partial^2 g(n,\vartheta)}{\partial\zeta_1(n,\vartheta)\partial\zeta_2(n,\vartheta)} = 0. \tag{E.17}$$

The Hessian matrix is negative definite. Any $\vartheta^* \in \Theta$ satisfying the first order conditions $\zeta_1(n,\vartheta^*) = \zeta_1(n,\vartheta_0)$ and $\zeta_2(n,\vartheta^*) = \zeta_2(n,\vartheta_0)$ is a global maximizer of $g(n,\vartheta)$. Therefore, the global maximizer of $g(n,\vartheta)$ on Θ is defined by:

$$\zeta_1(n,\vartheta^*) = \zeta_1(n,\vartheta_0) \Longleftrightarrow \frac{(1-\lambda^*)^2}{(1-\lambda_0)^2} = \frac{(n\sigma_\alpha^{2*} + \sigma_\epsilon^{2*})}{(n\sigma_{\alpha 0}^2 + \sigma_{\epsilon 0}^2)}, \tag{E.18}$$

$$\zeta_2(n,\vartheta^*) = \zeta_2(n,\vartheta_0) \Longleftrightarrow \frac{(n-1+\lambda^*)^2}{(n-1+\lambda_0)^2} = \frac{\sigma_\epsilon^{2*}}{\sigma_{\epsilon 0}^2} \tag{E.19}$$

If equations (E.18) and (E.18) hold, ϑ^* is a global maximizer of $g(n,\vartheta)$ on Θ. It's easy to see that ϑ_0 satisfies the two conditions for any n, so ϑ_0 is a global maximizer of $g(n,\vartheta)$ for any n. Hence ϑ_0 is a global maximizer of $\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* g(n,\vartheta)$. $\square$

Lemma E.3. *Under Assumptions 1-5, for any $\vartheta^* \in \Theta$ and $\vartheta^* \neq \vartheta_0$, $g(n_1,\vartheta_0) = g(n_1,\vartheta^*)$ and $g(n_2,\vartheta_0) = g(n_2,\vartheta^*)$ do not hold simultaneously if $n_1 \neq n_2$.*

The idea behind Lemma E.2 and Lemma E.3 goes as follows. For each n, there is a two-dimensional surface $(\zeta_1(n,\vartheta),\zeta_2(n,\vartheta))$ in the three dimensional space Θ that globally maximizes $g(n,\vartheta)$. The true parameter ϑ_0 is on this surface for any n. Therefore ϑ_0 is a global maximizer of $g(n,\vartheta)$ for any n. With variation in group size, the two-dimensional maximization surface $(\zeta_1(n,\vartheta),\zeta_2(n,\vartheta))$ for each $g(n,\vartheta)$ is different, and ϑ_0 is the unique interception of these surfaces. Therefore ϑ_0 is the unique global maximizer of $\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* g(n,\vartheta)$.

Proof of Lemma E.3. By Lemma E.2, ϑ_0 is a global maximizer of $g(n,\vartheta)$ on Θ for any n. To have $g(n_1,\vartheta_0) = g(n_1,\vartheta^*)$ and $g(n_2,\vartheta_0) = g(n_2,\vartheta^*)$, we need ϑ^* to be a global maximizer for both $g(n_1,\vartheta)$ and $g(n_2,\vartheta)$. Recall that $\vartheta^* = (\lambda^*, \sigma_\epsilon^{2*}, \sigma_\alpha^{2*})$ is a global maximizer of of $g(n,\vartheta)$ if it satisfies equations (E.18) and (E.19). From equation (E.19),

$$\sigma_\epsilon^{2*} = \frac{(n-1+\lambda^*)^2}{(n-1+\lambda_0)^2}\sigma_{\epsilon 0}^2. \tag{E.20}$$

Plugging equation (E.20) into equation (E.18), the global maximizer should satisfy equation (E.20) and the following condition:

$$(\lambda^* - \lambda_0)\varrho(n,\lambda^*) = [\frac{(1-\lambda^*)^2}{(1-\lambda_0)^2}\sigma_{\alpha 0}^2 - \sigma_\alpha^{2*}]\frac{(1-\lambda_0)^2}{\sigma_{\epsilon 0}^2}, \tag{E.21}$$

where

$$\varrho(n, \lambda^*) = \frac{n[(1-\lambda_0)+(1-\lambda^*)] - 2(1-\lambda_0)(1-\lambda^*)}{(n-1+\lambda_0)^2}. \tag{E.22}$$

Given equations (E.20) and (E.21), $\vartheta^* \neq \vartheta_0$ is equivalent to $\lambda^* \neq \lambda_0$. [3]Note that the right hand side of equation (E.21) does not depend on group size n. With $\lambda^* \neq \lambda_0$, a necessary condition for equation (E.21) to hold for both n_1 and n_2 is that $\varrho(n_1, \lambda^*) = \varrho(n_2, \lambda^*)$. If $\varrho(n_1, \lambda^*) \neq \varrho(n_2, \lambda^*)$, then $g(n_1, \vartheta_0) = g(n_1, \vartheta^*)$ and $g(n_2, \vartheta_0) = g(n_2, \vartheta^*)$ do not hold simultaneously. Therefore, it suffices to show that if $n_1 \neq n_2$, $\varrho(n_1, \lambda^*) \neq \varrho(n_2, \lambda^*)$.

Let $l_0 = 1 - \lambda_0$ and $l^* = 1 - \lambda^*$. Then under Assumption 4, $0 < l_0 < \underline{\mathrm{a}}$, $0 < l^* < \underline{\mathrm{a}}$, where $\underline{\mathrm{a}}$ is the lower bound of n_c and $\underline{\mathrm{a}} \geqslant 2$.Let m be the continuous counterpart of n, $m \geqslant \underline{\mathrm{a}}$,

$$\tilde{\varrho}(m, l^*) = \frac{m(l_0 + l^*) - 2l_0 l^*}{(m - l_0)^2}. \tag{E.23}$$

Therefore $\varrho(n, \lambda^*) = \tilde{\varrho}(m, l^*)$ if $n = m$. Below I will show that $\tilde{\varrho}(m, l^*)$ is decreasing in m. The first derivative of $\tilde{\varrho}(m, l^*)$ with respect to m is

$$\frac{\partial \tilde{\varrho}(m, l^*)}{\partial m} = \frac{-{l_0}^2 - m(l_0 + l^*) + 3l_0 l^*}{(m - l_0)^3}. \tag{E.24}$$

[3]Clearly $\vartheta^* \neq \vartheta_0$ if $\lambda^* \neq \lambda_0$. If $\lambda^* = \lambda_0$, then equation (E.20) gives $\sigma_\epsilon^{2*} = \sigma_{\epsilon 0}^2$, and equation (E.21) gives $\sigma_\alpha^{2*} = \sigma_{\alpha 0}^2$. Therefore if $\vartheta^* \neq \vartheta_0$, $\lambda^* \neq \lambda_0$. So $\lambda^* = \lambda_0$ is equivalent to $\vartheta^* = \vartheta_0$.

The denominator $(m-l_0)^3>0$ as $m \geqslant \underline{a}$ and $l_0<\underline{a}$. Since $l_0+l^*>0$ and $m \geqslant \underline{a}$, the numerator

$$
\begin{aligned}
-{l_0}^2 - m(l_0+l^*) + 3l_0l^* &\leqslant -{l_0}^2 - 2(l_0+l^*) + 3l_0l^* \\
&= -l_0^2 - 2l_0 + (3l_0-2)l^* \\
&\begin{cases} \leqslant -l_0^2 - 2l_0 + 2*(3l_0-2) & if\, 2/3 \leqslant l_0 < \underline{a} \\ \leqslant -l_0^2 - 2l_0 & if\, 0 < l_0 < 2/3 \end{cases}
\end{aligned}
\tag{E.25}
$$

When $2/3 \leqslant l_0 < \underline{a}$, $-l_0^2-2l_0+2*(3l_0-2) = -(l_0-2)^2 < 0$. When $0<l_0<2/3$, $-l_0^2-2l_0 < 0$. In all, the numerator is negative. Therefore $\partial\tilde{\varrho}(m,l^*)/\partial m < 0$ and $\tilde{\varrho}(m,l^*)$ is decreasing in m. If $n_1 \neq n_2$, then $\tilde{\varrho}(n_1,l^*) \neq \tilde{\varrho}(n_2,l^*)$ and hence $\varrho(n_1,\lambda^*) \neq \varrho(n_2,\lambda^*)$. Therefore if $\lambda^* \neq \lambda_0$, equation (E.21) does not hold for both n_1 and n_2 if $n_1 \neq n_2$. As a result, $g(n_1,\vartheta_0)=g(n_1,\vartheta^*)$ and $g(n_2,\vartheta_0)=g(n_2,\vartheta^*)$ cannot hold simultaneously if $n_1 \neq n_2$. □

Identification uniqueness follows from Lemma E.1, Lemma E.2, and Lemma E.3.

Proof of Proposition E.2. Given $\bar{Q}(\vartheta)$ in equation (E.9), and that for any $\vartheta^* \in \Theta$ and $\vartheta^* \neq \vartheta_0$, $lim_{R\to\infty} N^{-1}[Q_R^{(2)}(\vartheta_0) - Q^{(2)}(\vartheta^*)] \geqslant 0$ (Lemma E.1),

$$\bar{Q}(\vartheta_0) - \bar{Q}(\vartheta^*) \geqslant \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^*[g(n,\vartheta_0)-g(n,\vartheta^*)]. \tag{E.26}$$

Under Assumption 3, $1/n^* \geqslant 1/\bar{a} > 0$. It suffices to show that for any for any $\vartheta^* \in \Theta$ and $\vartheta^* \neq \vartheta_0$, $\sum_{n=\underline{a}}^{\bar{a}} \omega_n^*[g(n,\vartheta_0)-g(n,\vartheta^*)] > 0$.

Under Assumption 3, there exists some $\xi_\omega^* > 0$ and $n_1 \neq n_2$, such that $\omega_{n_1}^* > \xi_\omega^*$ and $\omega_{n_2}^* > \xi_\omega^*$. Since $g(n, \vartheta_0) - g(n, \vartheta^*) \geqslant 0$ for any $\underline{a} \leqslant n \leqslant \bar{a}$ (Lemma E.2),

$$\sum_{n=\underline{a}}^{\bar{a}} \omega_n^*[g(n, \vartheta_0) - g(n, \vartheta^*)] \geqslant \xi_\omega^*[g(n_1, \vartheta_0) - g(n_1, \vartheta^*) + g(n_2, \vartheta_0) - g(n_2, \vartheta^*)]. \tag{E.27}$$

By Lemma E.3, if $\vartheta^* \neq \vartheta_0$ and $n_1 \neq n_2$, $g(n_1, \vartheta_0) - g(n_1, \vartheta^*) > 0$ or $g(n_2, \vartheta_0) - g(n_2, \vartheta^*) > 0$ or both. Therefore, for any $\xi_\vartheta > 0$, there exists some $\xi_g > 0$ such that

$$inf_{\vartheta^* \in \Theta, ||\vartheta^* - \vartheta_0|| > \xi_\vartheta} \{[g(n_1, \vartheta_0) - g(n_1, \vartheta^*)] + [g(n_2, \vartheta_0) - g(n_2, \vartheta^*)]\} > \xi_g. \tag{E.28}$$

Combining equations (E.26), (E.27) and (E.28), for any $\xi_\vartheta > 0$, there exists some $\xi_g > 0$ and $\xi_\omega^* > 0$, such that

$$inf_{\vartheta^* \in \Theta, ||\vartheta^* - \vartheta_0|| > \xi_\vartheta} [\bar{Q}(\vartheta_0) - \bar{Q}(\vartheta^*)] \geqslant \frac{1}{\bar{a}} \xi_\omega^* \xi_g^* > 0.$$

In sum, ϑ_0 is the unique global maximizer of $\bar{Q}(\vartheta) = lim_{R \to \infty} N_R^{-1} E[Q_R(\vartheta)]$. $\square$

Proposition E.3. *$sup_{\vartheta \in \Theta} |N_R^{-1} Q_R(\vartheta) - \bar{Q}(\vartheta)| \to_p 0$ as R goes to infinity.*

Proof of this proposition is in Appendix F on page 119.

E.2 Proof of Theorem 3.2

To prove Theorem 3.2, I first derive the score function here. Then I show the convergence of the Hessian matrix in Proposition E.4. Next I show the asymptotic normality of the score function at ϑ_0 in Proposition E.5. Proof of Theorem 3.2 is based on Proposition E.4 and Proposition E.5, and is at the end of this section.

Since $\vartheta = (\lambda, \sigma_\epsilon^2, \sigma_\alpha^2)$, I denote λ as ϑ_1, σ_ϵ^2 as ϑ_2 and σ_α^2 as ϑ_3. Note that $\partial\Omega(\vartheta)/\partial\vartheta_2 = I$ and $\partial\Omega(\vartheta)/\partial\vartheta_3 = \tilde{J}$. The first derivatives of $Q_R(\vartheta)$ with respect to

ϑ are

$$\frac{\partial Q_R(\vartheta)}{\partial \lambda} = -tr[(I-\lambda W)^{-1}W] + Y'(I-\lambda W)M_Z(\vartheta)WY, \tag{E.29}$$

For $i = 2, 3$,

$$\frac{\partial Q_R(\vartheta)}{\partial \vartheta_i} = -\frac{1}{2}tr(\Omega(\vartheta)^{-1}\frac{\partial \Omega}{\partial \vartheta_i}) + \frac{1}{2}Y(I-\lambda W)M_Z(\vartheta)\frac{\partial \Omega(\vartheta)}{\partial \vartheta_i}M_Z(\vartheta)(I-\lambda W)Y. \tag{E.30}$$

Since $Y = (I-\lambda W_0)^{-1}(Z\gamma_0 + U)$ and $Z'M_{Z0} = 0$, first order derivatives at ϑ_0 are

$$\frac{\partial Q_R(\vartheta)}{\partial \lambda}|_{\vartheta_0} = -tr[(I-\lambda_0 W)^{-1}W] + U'M_{Z0}W(I-\lambda_0 W)^{-1}(Z\gamma_0 + U). \tag{E.31}$$

$$\frac{\partial Q_R(\vartheta)}{\partial \vartheta_i}|_{\vartheta_0} = -\frac{1}{2}tr(\Omega_0^{-1}\frac{\partial \Omega}{\partial \vartheta_i}|_{\vartheta_0}) + \frac{1}{2}U'M_{Z0}\frac{\partial \Omega}{\partial \vartheta}|_{\vartheta_0}M_{Z0}U,\ i = 1, 2. \tag{E.32}$$

Using Lemma C.1 and the commutative properties of $diag_{c=1}^R\{p_c I_c + q_c J_c\}$ type of matrices in Lemma A.1, the expected value of the score function at ϑ_0 is

$$\begin{aligned} E[\frac{\partial Q_R(\vartheta)}{\partial \lambda}]_{\vartheta_0} &= -tr[(I-\lambda_0 W)^{-1}W] + tr[M_{Z0}W(I-\lambda_0 W)^{-1}\Omega_0] \\ &= -tr[P_{Z0}W(I-\lambda_0 W)^{-1}\Omega_0], \\ E[\frac{\partial Q_R(\vartheta)}{\partial \vartheta_i}]_{\vartheta_0} &= -\frac{1}{2}tr(\Omega_0^{-1}\frac{\partial \Omega}{\partial \vartheta_i}|_{\vartheta_0}) + \frac{1}{2}tr(M_{Z0}\frac{\partial \Omega}{\partial \vartheta}|_{\vartheta_0}M_{Z0}\Omega_0) \\ &= -tr(P_{Z0}\frac{\partial \Omega_0}{\partial \vartheta_i}) + \frac{1}{2}tr(P_{Z0}\frac{\partial \Omega_0}{\partial \vartheta_i}P_{Z0}\Omega_0), i = 2, 3. \end{aligned}$$

Note that $W(I-\lambda_0 W)^{-1}\Omega_0$, $\partial\Omega_0/\partial\vartheta_i$, and Ω_0 are all in the form of $diag_{c=1}^R\{p_c I_c^* + s_c J_c^*\}$, with p_c and s_c uniformly bounded in absolute value. Using Lemma B.3,

$$lim_{R\to\infty}N_R^{-1}tr[P_{Z0}W(I-\lambda_0 W)^{-1}\Omega_0] = 0, \tag{E.33}$$

$$lim_{R\to\infty}N_R^{-1}tr[P_{Z0}\frac{\partial \Omega_0}{\partial \vartheta_i}] = 0, \tag{E.34}$$

$$lim_{R\to\infty}N_R^{-1}tr[P_{Z0}\frac{\partial \Omega_0}{\partial \vartheta_i}P_{Z0}\Omega_0] = 0. \tag{E.35}$$

Therefore, $lim_{R\to\infty}E[N_R^{-1}\partial Q_R(\vartheta)/\partial\vartheta_i]_{\vartheta_0} = 0$ but $E[N_R^{-1}\partial Q_R(\vartheta)/\partial\vartheta_i]_{\vartheta_0} \neq 0$. The expected value of the score function is not 0 because we are using the concentrated maximum likelihood estimator here. Concentrating out γ leads to loss of degree of freedom.

Proposition E.4. *$N_R^{-1}\partial^2 Q_R(\hat{\vartheta}_R)/\partial\hat{\vartheta}\partial\hat{\vartheta}'$ converges to matrix Γ_0 uniformly as R goes to infinity and $\hat{\vartheta}_R \to \vartheta_0$, where $\Gamma_0 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}} \omega_n^* \Xi_{0,n}$, $\Xi_{0,n}$ is in D of Appendix D.*

Proof. Recall that $\lambda = \vartheta_1$, $\sigma_\epsilon^2 = \vartheta_2$, $\sigma_\alpha^2 = \vartheta_3$. Note that $\Omega(\vartheta) = \sigma_\epsilon^2 I + \sigma_\alpha^2 \tilde{J}$, for $i, j = 1, 2, 3$, we have $\partial^2\Omega(\vartheta)/\partial\vartheta_i\partial\vartheta_j = 0$. Therefore, the second derivatives are

$$\frac{\partial^2 Q_R(\vartheta)}{\partial\lambda^2} = -tr[(I-\lambda W)^{-2}W^2] - Y'WM_Z(\vartheta)WY, \quad \text{(E.36)}$$

$$\frac{\partial^2 Q_R(\vartheta)}{\partial\lambda\partial\vartheta_i} = -Y'(I-\lambda W)M_Z(\vartheta)\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}M_Z(\vartheta)WY,\ i = 2,3, \quad \text{(E.37)}$$

$$\frac{\partial^2 Q_R(\vartheta)}{\partial\vartheta_i\partial\vartheta_j} = \frac{1}{2}tr(\Omega^{-2}\frac{\partial\Omega}{\partial\vartheta_i}\frac{\partial\Omega}{\partial\vartheta_j}) - Y'(I-\lambda W)M_Z(\vartheta)\frac{\partial\Omega}{\partial\vartheta_i}M_Z(\vartheta)\frac{\partial\Omega}{\partial\vartheta_j}M_Z(\vartheta)(I-\lambda W)Y,\ i,j \neq 1. \quad \text{(E.38)}$$

Therefore $\partial^2 Q_R(\vartheta)/\partial\vartheta\partial\vartheta'$ exits and is continuous in $\vartheta \in \Theta$.

Using Lemma C.1 for the expected value of linear quadratic forms,

$$E[\frac{\partial^2 Q_R(\vartheta)}{\partial\lambda^2}]_{\vartheta_0} = -tr[(I-\lambda_0 W)^{-2}W^2] - \gamma_0' Z'(I-\lambda_0 W)^{-1}WM_{Z0}W(I-\lambda_0 W)^{-1}Z\gamma_0,$$
$$-\ tr[\Omega_0(I-\lambda_0 W)^{-1}WM_{Z0}W(I-\lambda_0 W)^{-1}] \quad \text{(E.39)}$$

$$E[\frac{\partial^2 Q_N}{\partial\lambda\partial\vartheta_i}]_{\vartheta_0} = -tr[\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}M_{Z0}W(I-\lambda_0 W)],\ i = 2,3, \quad \text{(E.40)}$$

$$E[\frac{\partial^2 Q_R(\vartheta)}{\partial\vartheta_i\partial\vartheta_j}]_{\vartheta_0} = \frac{1}{2}tr(\Omega^{-2}\frac{\partial\Omega}{\partial\vartheta_i}\frac{\partial\Omega}{\partial\vartheta_j})|_{\vartheta_0} - tr(\frac{\partial\Omega}{\partial\vartheta_i}M_{Z0}\frac{\partial\Omega}{\partial\vartheta_j}\Omega_0^{-1}), i,j \neq 1. \quad \text{(E.41)}$$

Using Lemma B.3,

$$lim_{R\to\infty}\frac{1}{N_R}tr[\Omega_0(I-\lambda_0 W)^{-1}WM_{Z0}W(I-\lambda_0 W)^{-1}] = lim_{R\to\infty}\frac{1}{N_R}tr[(I-\lambda_0 W)^{-2}W^2],$$
$$lim_{R\to\infty}\frac{1}{N_R}tr[\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}M_{Z0}W(I-\lambda_0 W)] = lim_{R\to\infty}\frac{1}{N_R}tr[\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\Omega_0^{-1}W(I-\lambda_0 W)],$$
$$lim_{R\to\infty}\frac{1}{N_R}tr(\frac{\partial\Omega}{\partial\vartheta_i}M_{Z0}\frac{\partial\Omega}{\partial\vartheta_j}\Omega_0^{-1}) = lim_{R\to\infty}\frac{1}{N_R}tr(\Omega_0^{-2}\frac{\partial\Omega}{\partial\vartheta_i}\frac{\partial\Omega}{\partial\vartheta_j}).$$

Therefore,

$$lim_{R\to\infty}\frac{1}{N_R}E[\partial Q_R(\vartheta)/\partial\vartheta\partial\vartheta']_{\vartheta_0} = -lim_{R\to\infty}\frac{1}{N_R}\Xi_0,$$

where Ξ_0 is defined in equation (D.2) in Appendix D. Under Assumptions 3 and 5,

$$lim_{R\to\infty}\frac{1}{N_R}\Xi_0 = \frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*\mathit{\Xi}_{0,n},$$

where $\mathit{\Xi}_{0,n}$ is in equation (D.3) in Appendix D. For more details of derivation, see Appendix D. Under Assumption 6,

$$lim_{R\to\infty}\frac{1}{N_R}E(\frac{\partial^2 Q_R(\vartheta)}{\partial\vartheta\partial\vartheta'})_{\vartheta_0} = -lim_{R\to\infty}\frac{1}{N_R}\Xi_0 = -\frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*\mathit{\Xi}_{0,n} = -\Gamma_0. \quad \text{(E.42)}$$

Finally, $\frac{1}{N_R}\frac{\partial^2 Q_R(\hat{\vartheta})}{\partial\vartheta\partial\vartheta'}$ converges to $lim_{R\to\infty}\frac{1}{N_R}E(\frac{\partial^2 Q_R(\vartheta)}{\partial\vartheta\partial\vartheta'})_{\vartheta_0}$ uniformly in probability as $\hat{\vartheta}_R \to_p \vartheta_0$ and R goes to infinity (See proof in Appendix F). In all, we have $\frac{1}{N_R}E(\frac{\partial^2 Q_R(\hat{\vartheta}_R)}{\partial\vartheta\partial\vartheta'}) \to_p -\Gamma_0$ as $\hat{\vartheta}_R \to_p \vartheta_0$ and $R \to \infty$. □

Proposition E.5. *$N^{-1/2}(\partial Q_R(\vartheta)/\partial\vartheta)_{\vartheta_0} \xrightarrow{D} N(0,\Psi_0)$, where Ψ_0 is defined in equation (3.21).*

Proof. From equation (E.31) and equation (E.32), the score function at ϑ_0 is

$$(\frac{\partial Q_R(\vartheta)}{\partial\vartheta})_{\vartheta_0} = f_{1R}(\vartheta_0) - f_{2R}(\vartheta_0), \quad \text{(E.43)}$$

where

$$f_{1R}(\vartheta_0) = \begin{bmatrix} U'\Omega_0^{-1}W(I-\lambda_0 W)^{-1}(U+Z\gamma_0) \\ \frac{1}{2}U'\Omega_0^{-2}U \\ \frac{1}{2}U'\Omega_0^{-2}\tilde{J}U \end{bmatrix} - \begin{bmatrix} tr[\Omega_0 W(I-\lambda_0 W)^{-1}] \\ \frac{1}{2}tr(\Omega_0^{-1}) \\ \frac{1}{2}tr(\Omega_0\tilde{J}) \end{bmatrix} \quad \text{(E.44)}$$

$$f_{2R}(\vartheta_0) = \begin{bmatrix} U'P_{Z0}W(I-\lambda_0 W)^{-1}(U+Z\gamma_0) \\ \frac{1}{2}U'[P_{Z0}\Omega_0^{-1}+\Omega_0^{-1}P_{Z0}-P_{Z0}P_{Z0}]U \\ \frac{1}{2}U'[P_{Z0}\tilde{J}\Omega_0^{-1}+\Omega_0^{-1}P_{Z0}\tilde{J}-P_{Z0}\tilde{J}P_{Z0}]U \end{bmatrix}.$$

Using Lemma C.1,

$$Ef_{2R}(\vartheta_0) = \begin{bmatrix} tr[P_{Z0}W(I-\lambda_0 W)^{-1}\Omega_0] \\ \frac{1}{2}tr[2P_{Z0}-\Omega_0 P_{Z0}P_{Z0}] \\ \frac{1}{2}tr[2P_{Z0}\tilde{J}-P_{Z0}\tilde{J}P_{Z0}\Omega_0] \end{bmatrix}.$$

By Lemma B.3, $lim_{R\to\infty}\frac{1}{\sqrt{N_R}}Ef_{2R}(\vartheta_0) = 0$, and $lim_{R\to\infty}\frac{1}{N_R}Var[f_{2R}(\vartheta_0)] = 0$. By Chebychev's inequality, $lim_{R\to\infty}\frac{1}{\sqrt{N_R}}f_{2R}(\vartheta) \to_p 0$. Therefore, it suffices to show that $N^{-1/2}f_{1R}(\vartheta_0) \xrightarrow{D} N(0,\Psi_0)$.

By Lemma C.1, $E(U'\Omega_0^{-1}W(I-\lambda_0 W)^{-1}(U+Z\gamma_0)) = tr[\Omega_0 W(I-\lambda_0 W)^{-1}]$, $E(U'\Omega_0^{-2}U) = tr(\Omega_0^{-1})$, and $E(U'\Omega_0^{-2}\tilde{J}U) = tr(\Omega_0\tilde{J})$. Therefore $Ef_{1R}(\vartheta_0) = 0$. Using Lemma C.1, Lemma B.1 and Assumption 5,

$$
\begin{aligned}
&lim_{R\to\infty}\frac{1}{N_R}E[f_{1R}(\vartheta_0)f_{1R}(\vartheta_0)] \\
=&\frac{1}{n^*}\sum_{n=\underline{a}}^{\bar{a}}\omega_n^*[\Xi_{0,n} + (\mu_\epsilon^{(4)} - 3\sigma_{\epsilon 0}^{(4)})\Xi_{1,n} + (\mu_\alpha^{(4)} - 3\sigma_{\alpha 0}^{(4)})\Xi_{2,n} + \mu_\epsilon^{(3)}\Xi_{3,n} + \mu_\alpha^{(3)}\Xi_{4,n}],
\end{aligned}
$$

where $\Xi_{0,n}$, $\Xi_{1,n}$, $\Xi_{2,n}$, $\Xi_{3,n}$, and $\Xi_{4,n}$are defined in equations (D.3), (D.5), (D.7), (D.9), (D.11) in Appendix D respectively. Under Assumption 6, the limiting matrix of the information matrix is $lim_{R\to\infty}\frac{1}{N_R}E[f_{1R}(\vartheta_0)f_{1R}(\vartheta_0)]]_{\vartheta_0} = \Psi_0$. For more details, see Appendix D.

Note that $f_{1R}(\vartheta_0) = [S^{(1)}, S_2^{(2)}, S_3^{(3)}]$, with for $i = 1,2,3$, $S^{(l)}$ can be written in the form of $S^{(l)} = U'A^{(l)}U + U'B^{(l)}Z\gamma_0$, where $A^{(l)}, B^{(l)}$ can all be written in the form of $diag_{c=1}^R\{p_c I^* + q_c J_c\}$. Using the CLT for vectors of linear quadratic forms in Theorem C.2, $\frac{1}{\sqrt{N_R}}f_{1R}(\vartheta_0) \xrightarrow{D} N(0,\Psi_0)$, and hence $\frac{1}{\sqrt{N_R}}(\frac{\partial Q_R}{\partial\vartheta})_{\vartheta_0} \xrightarrow{D} N(0,\Psi_0)$. □

With Proposition E.4 and Proposition E.5, Theorem 3.2 follows.

Proof of Theorem 3.2. From equations (E.36),(E.37), and (E.38), the second derivative $\partial^2 Q_R(\vartheta)/\partial\vartheta\partial\vartheta'$ exists and is continuous in $\vartheta \in \Theta$. By Proposition E.4, $\frac{1}{N_R}\frac{\partial^2 Q_R(\hat{\vartheta}_R)}{\partial\hat{\vartheta}\partial\hat{\vartheta}'}$ converges to matrix $-\Gamma_0$ as R goes to infinity and $\hat{\vartheta}_R \to \vartheta_0$. By Proposition E.5, $N^{-1/2}(\partial Q_R(\vartheta)/\partial\vartheta)_{\vartheta_0} \xrightarrow{D} N(0,\Psi_0)$. Therefore, $\sqrt{N}(\hat{\vartheta}_R - \vartheta_0) \xrightarrow{D} N_R(0,\Phi_0)$ as R goes to infinity, where $\Phi_0 = \Gamma_0^{-1}\Psi_0\Gamma_0^{-1}$. □

Appendix F: Proof of Uniform convergence

In this section, I will prove the following two arguments.

1. $sup_{\vartheta\in\Theta}|N_R^{-1}Q_R(\vartheta)-\bar{Q}(\vartheta)|\rightarrow_p 0$ as R goes to infinity.

2. $sup_{\vartheta\in\Theta}N_R^{-1}|\frac{\partial^2 Q_N}{\partial\vartheta_i\partial\vartheta_j}-E\frac{\partial^2 Q_N}{\partial\vartheta_i\partial\vartheta_j}|\rightarrow_p 0$ as R goes to infinity.

Note that $lim_{R\rightarrow\infty}N_R^{-1}EQ_R(\vartheta)=\bar{Q}(\vartheta)$,

$$Q_R(\vartheta)-EQ_R(\vartheta) \;=\; U'A_1(\vartheta)U-tr(A_1(\vartheta))-2U'A_1(\vartheta)Z\gamma_0,$$

where

$$A_1(\vartheta) \;=\; -\frac{1}{2}(I-\lambda_0W)^{-1}I-\lambda W)M_Z(\vartheta)(I-\lambda W)(I-\lambda_0W)^{-1}.$$

Also, the hessian matrix can be written as

$$\begin{aligned}
\frac{\partial^2 Q_R(\vartheta)}{\partial\lambda^2} &= (Z\gamma_0+U)'A_2(\vartheta)(Z\gamma_0+U)-tr[(I-\lambda W)^{-2}W^2],\\
\frac{\partial^2 Q_R(\vartheta)}{\partial\lambda\partial\vartheta_i} &= (Z\gamma_0+U)'A_3^{(i)}(\vartheta)(Z\gamma_0+U)-tr[(I-\lambda W)^{-2}W^2],\; i=2,3,\\
\frac{\partial^2 Q_R(\vartheta)}{\partial\vartheta_i\partial\vartheta_j} &= \frac{1}{2}tr(\Omega^{-2}\frac{\partial\Omega}{\partial\vartheta_i}\frac{\partial\Omega}{\partial\vartheta_j})-(Z\gamma_0+U)'A_4^{(i,j)}(\vartheta)(Z\gamma_0+U), i\neq 1, j\neq 1
\end{aligned}$$

where

$$\begin{aligned}
A_2(\vartheta) &= (I-\lambda_0W)^{-1}WM_Z(\vartheta)W(I-\lambda_0W)^{-1},\\
A_3^{(i)}(\vartheta) &= (I-\lambda_0W)^{-1}(I-\lambda W)M_Z(\vartheta)\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}M_Z(\vartheta)(I-\lambda W)(I-\lambda_0W)^{-1},\\
A_4^{(i,j)}(\vartheta) &= (I-\lambda_0W)^{-1}(I-\lambda W)M_Z(\vartheta)\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}M_Z(\vartheta)\frac{\partial\Omega}{\partial\vartheta_j}M_Z(\vartheta)(I-\lambda W)(I-\lambda_0W)^{-1}.
\end{aligned}$$

Let $\mathscr{G}$ be the set of matrices which can be written in the form of $diag_{c=1}^R\{p_c(\vartheta)I_c^*+s_c(\vartheta)J_c^*\}$, where both $p_c(\vartheta)$ and $s_c(\vartheta)$ are uniformly bounded in absolute value, continuously differentiable in the interior of Θ, and the derivatives are uniformly bounded,

$$\mathscr{G} = \{diag_{c=1}^{R}\{p_c(\vartheta)I_c + s_c(\vartheta)J_c\}|\}, \tag{F.1}$$

where $sup_{\vartheta\in\Theta}|p_c(\vartheta)| < \bar{a}_g$, $sup_{\vartheta\in\Theta}s_c(\vartheta) < \bar{a}_g$, $sup_{\vartheta\in int(\Theta)}|\nabla_\vartheta p_c(\vartheta)| < \bar{a}_g$,$sup_{\vartheta\in int(\Theta)}|s_c(\vartheta)| < \bar{a}_g$, $\bar{a}_g$ is a finite constant.

Note $A_1(\vartheta)$, $A_2(\vartheta)$, $A_3^{(i)}(\vartheta)$ and $A_4^{(i,j)}(\vartheta)$ can all be written as the products of matrices in the following matrix set

$$\mathscr{A} = \{0_{N\times N}, I_N, \tilde{J}, W, (I-\lambda W), (I-\lambda_0 W)^{-1}, \Omega(\vartheta)^{-1}, \Omega_0, M_Z(\vartheta)\} \tag{F.2}$$

$$= \mathscr{A}^* \cup \{M_Z(\vartheta)\}, \tag{F.3}$$

where

$$\mathscr{A}^* = \{0_{N\times N}, I_N, \tilde{J}, W, (I-\lambda W), (I-\lambda_0 W)^{-1}, \Omega(\vartheta)^{-1}, \Omega_0\}. \tag{F.4}$$

Note that $\mathscr{A}^* \subset \mathscr{G}$, and $\mathscr{A}^* \cap \{M_Z(\vartheta)\} = \emptyset$.

If the elements of matrix $A(\vartheta)$ is uniformly bounded in absolute value, $sup_{\vartheta\in\Theta}|A(\vartheta)|_{ij} < \bar{a}_A$, where $\bar{a}_A$ is a finite constant, I denote $A(\vartheta) = O_U(1)$. If the elements of matrix $A(\vartheta)$ is uniformly of order $1/N_R$, $sup_{\vartheta\in\Theta}N_R|A(\vartheta)|_{ij} < \bar{a}_A$, I denote $A_N(\vartheta) = O_U(1/N)$.

In this section, I will demonstrate that products of matrices in $\mathscr{A}$ have some special properties in Lemma F.1, with such properties, the quadratic forms associated with these products converges uniformly to the expected value.

Lemma F.1. *Suppose matrix* $A(\vartheta) = \prod_{k=1}^{K} A_k(\vartheta)$*, where* $1 \leqslant K < \infty$*,* $A_k(\vartheta) \in \mathscr{A}$*,* $\mathscr{A}$ *is the matrix set defined in equation* (F.2)*. Then*

(a) $A(\vartheta) = \tilde{A}(\vartheta) + \breve{A}(\vartheta)$*, where* $\tilde{A}(\vartheta) \in \mathscr{G}$*, and* $\breve{A}(\vartheta) = O_U(1/N)$*.*

(b) $A(\vartheta)$ *is continuous in* ϑ *and continuously differentiable on* $\vartheta \in \Theta$*, with* $\nabla_\vartheta \tilde{A}(\vartheta) \in \mathscr{G}$*, and* $\nabla_\vartheta \breve{A}(\vartheta) = O_U(1/N)$.

(c) The elements of $A(\vartheta)$ *and* $A(\vartheta)Z\gamma_0$ *and are uniformly bounded in absolute value.*

(d) There exists some $K_A < \infty$ *and* $K_\eta < \infty$*, such that for all* $\vartheta_1, \vartheta_2 \in \Theta$, $|\tilde{A}_{ij,N}(\vartheta_1) - \tilde{A}_{ij,N}(\vartheta_2)| \leqslant K_A||\vartheta_1 - \vartheta_2||$, $N|\breve{A}_{ij,N}(\vartheta_1) - \breve{A}_{ij,N}(\vartheta_2)| \leqslant K_A||\vartheta_1 - \vartheta_2||$, $|A(\vartheta_1)Z\gamma_0 - A(\vartheta_2)Z\gamma_0| < K_\eta||\vartheta_1 - \vartheta_2||$.

(e)$\frac{1}{N_R}tr(A(\vartheta)\Omega_0)$ *is uniformly continuous.*

Proof. (a) Since $\mathscr{A} = \mathscr{A}^* \cup \{M_Z(\vartheta)\}$ and $\mathscr{A}^* \cap \{M_Z(\vartheta)\} = \emptyset$, if $A_k(\vartheta) \in \mathscr{A}$, then either $A_k(\vartheta) \in \mathscr{A}^*$ or $A_k(\vartheta) = M_Z(\vartheta)$. Define $A_k^*(\vartheta)$ and $A^\diamond(\vartheta)$ as

$$A_k^*(\vartheta) = \begin{cases} A_k(\vartheta) & if\ A_k(\vartheta) \in \mathscr{A}^* \\ \\ \Omega(\vartheta)^{-1} & if\ A_k(\vartheta) = M_Z(\vartheta) \end{cases}. \tag{F.5}$$

$$A_k^\diamond(\vartheta) = \begin{cases} 0_{N\times N} & if\ A_k(\vartheta) \in \mathscr{A}^* \\ \\ -\tilde{P}_Z(\vartheta) & if\ A_k(\vartheta) = M_Z(\vartheta) \end{cases}. \tag{F.6}$$

Clearly, $0_{N\times N}$ is $O_U(1/N)$. By Lemma B.3, $\tilde{P}_Z = O_U(1/N)$. If $A_k(\vartheta) \in \mathscr{A}$, then $A_k = A_k^*(\vartheta) + A_k^\diamond(\vartheta)$, where $A_k^*(\vartheta) \in \mathscr{A}^*$, and $A_k^\diamond(\vartheta) = O_U(1/N)$. By Lemma B.2, $A_k^*(\vartheta)O_U(1/N) = O_U(1/N)$, and $O_U(1/N)O_U(1/N) = O_U(1/N)$. Therefore,

$$\begin{aligned} A(\vartheta) &= \prod_{k=1}^{K} A_k(\vartheta) = \prod_{k=1}^{K}[A_k^*(\vartheta) + O_U(1/N)] \\ &= \prod_{k=1}^{K} A_k^*(\vartheta) + O_U(1/N). \end{aligned} \tag{F.7}$$

Note that$\prod_{k=1}^K A_k^*(\vartheta) \in \mathscr{G}$. Let

$$\tilde{A}(\vartheta) = \prod_{k=1}^{K} A_k^*(\vartheta), \tag{F.8}$$

and $\breve{A}(\vartheta) = A(\vartheta) - \tilde{A}(\vartheta)$. Then $A(\vartheta) = \tilde{A}(\vartheta) + \breve{A}(\vartheta)$, with $\tilde{A}(\vartheta) \in \mathscr{G}$ and $\breve{A}(\vartheta) = O_U(1/N)$.

(b) From Table B.1, if $A_k(\vartheta) \in \mathscr{A}^*$, $A_k(\vartheta)$ can be written in the form of $diag_{c=1}^R\{p_c(\vartheta)I_c + q_c(\vartheta)J_c\}$, where $p_c(\vartheta)$ and $q_c(\vartheta)$ are uniformly bounded and continuous in Θ. Since $(Z'\Omega(\vartheta)^{-1}Z)^{-1}$ is continuous in $\vartheta \in \Theta$, $\Omega(\vartheta)^{-1}$ is continuous in $\vartheta \in \Theta$, $M_Z(\vartheta)$ defined in (3.18) is continuous in $\vartheta \in \Theta$. In all, if $A_k(\vartheta) \in \mathscr{A}$, A_k is continuous in $\vartheta \in \Theta$. Therefore $A(\vartheta) = \prod_{k=1}^R A_k(\vartheta)$ is continuous in $\vartheta \in \Theta$.

Next I will show the continuous differentiability of $A(\vartheta)$. Denote λ as ϑ_1, σ_ϵ^2 as ϑ_2, σ_α^2 as ϑ_3. Note that

$$\partial\Omega(\vartheta)/\lambda = 0, \tag{F.9}$$

$$\partial\Omega(\vartheta)/\partial\sigma_\epsilon^2 = I, \tag{F.10}$$

$$\partial\Omega(\vartheta)/\partial\sigma_\alpha^2 = \tilde{J}, \tag{F.11}$$

For $i = 1, 2, 3$, $\partial\Omega(\vartheta)/\partial\vartheta_i \in \mathscr{A}^*$, and

$$\partial\Omega(\vartheta)^{-1}/\partial\vartheta_i = -\Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\Omega(\vartheta)^{-1}, \tag{F.12}$$

Since $\Omega(\vartheta)^{-1}, \partial\Omega(\vartheta)/\partial\vartheta_i \in \mathscr{A}^*$, $\partial\Omega(\vartheta)^{-1}/\partial\vartheta_i \in \mathscr{G}$. Note also that

$$\partial(I - \lambda W)/\partial\lambda = -W, \tag{F.13}$$

$$\partial(I-\lambda W)/\vartheta_i = 0, i = 2,3. \tag{F.14}$$

Besides, $0_{N\times N}$, I_N, $\tilde{J}$,W, $(I-\lambda_0 W)^{-1}$, and Ω_0 are constant matrices that do not depend on ϑ. So their derivatives are 0. In all, if $A_k(\vartheta)\in A^*$, $\nabla_{\vartheta_i}A_k(\vartheta)=\nabla_{\vartheta_i}A_k^*(\vartheta)\in \mathscr{G}$ defined in equation (F.1). When $A_k(\vartheta)=M_Z(\vartheta)$,

$$\begin{aligned}
&\partial M_Z(\vartheta)/\partial\vartheta_i\\
&= -\Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\Omega(\vartheta)^{-1} + \Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}
\end{aligned} \tag{F.15}$$

$$\begin{aligned}
&+\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\Omega(\vartheta)^{-1}\\
&-\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}\\
&= -\Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}[\Omega(\vartheta)^{-1}-\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}]\\
&+\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}[\Omega(\vartheta)^{-1}-\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}]\\
&= [\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}-\Omega(\vartheta)^{-1}]\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}[\Omega(\vartheta)^{-1}-\Omega(\vartheta)^{-1}Z(Z'\Omega(\vartheta)^{-1}Z)^{-1}Z'\Omega(\vartheta)^{-1}]\\
&= -M_Z(\vartheta)\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}M_Z(\vartheta).
\end{aligned} \tag{F.16}$$

Since $M_Z(\vartheta)=\Omega(\vartheta)^{-1}+O_U(1/N)$, $\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\in\mathscr{G}$ for $i=1,2,3$. By Lemma B.2, $\Omega(\vartheta)^{-1}O_U(1/N)=O_U(1/N)$ and $\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}O_U(1/N)=O_U(1/N)$. Therefore,

$$\begin{aligned}
M_Z(\vartheta)\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}M_Z(\vartheta) &= -\Omega(\vartheta)^{-1}\frac{\partial\Omega(\vartheta)}{\partial\vartheta_i}\Omega(\vartheta)^{-1}+O_U(1/N)\\
&= \frac{\partial\Omega(\vartheta)^{-1}}{\partial\vartheta_i}+O_U(1/N).
\end{aligned} \tag{F.17}$$

To sum up, if $A_k(\vartheta)=M_Z(\vartheta)$, $\nabla_\vartheta A_k(\vartheta)=\nabla_\vartheta A_k^*(\vartheta)+O_U(1/N)$.

In all, if $A_k(\vartheta)\in\mathscr{A}$ as defined in equation (F.2), $\nabla_\vartheta A_k(\vartheta)=\nabla_\vartheta A_k^*(\vartheta)+O_U(1/N)$, with $\nabla_\vartheta A_k^*(\vartheta)\in\mathscr{G}$. By Lemma B.2, $[\nabla_{\vartheta_i}A_k^*(\vartheta)]O_U(1/N)=O_U(1/N)$ and

$A_k(\vartheta)O_U(1/N) = O_U(1/N)$, $O_U(1/N)O_U(1/N) = O_U(1/N)$. Therefore, if $A(\vartheta) = \prod_{k=1}^{R} A_k(\vartheta)$,

$$\nabla_{\vartheta_i} A(\vartheta) = \nabla_{\vartheta_i}(\prod_{k=1}^{R} A_k^*(\vartheta)) + O_U(1/N),$$

$$= \nabla_{\vartheta_i}\tilde{A}(\vartheta) + O_U(1/N). \tag{F.18}$$

Since $\tilde{A}(\vartheta) \in \mathscr{G}$, $\nabla_{\vartheta_i}\tilde{A}(\vartheta) \in \mathscr{G}$. Since $A(\vartheta) = \tilde{A}(\vartheta) + \breve{A}(\vartheta)$, $\nabla_{\vartheta_i}\breve{A}(\vartheta) = O_U(1/N)$.

(c) From (a), $A(\vartheta) = \tilde{A}(\vartheta) + \breve{A}(\vartheta)$, where $\tilde{A}(\vartheta) \in \mathscr{G}$, and $\breve{A}(\vartheta) = O_U(1/N)$. So the elements of $A(\vartheta)$ are uniformly bounded in absolute value. Decompose $A(\vartheta)Z\gamma_0$ as

$$A(\vartheta)Z\gamma_0 = \tilde{A}(\vartheta)Z\gamma_0 + \breve{A}(\vartheta)Z\gamma_0. \tag{F.19}$$

Under Assumption 5, the elements of Z are bounded in absolute value. Using Lemma B.2, the elements of $\tilde{A}(\vartheta)Z\gamma_0$ and $\breve{A}(\vartheta)Z\gamma_0$ are uniformly bounded in absolute value. Therefore the elements of $A(\vartheta)Z\gamma_0$ are uniformly bounded in absolute value.

(d) Using the mean value theorem, for any $\vartheta_1, \vartheta_2 \in \Theta$, there exists some $t_1, t_2, t_3 \in [0,1]$, such that for $\vartheta^*_{(i)} = t_i\vartheta_1 + (1-t_i)\vartheta_2$, $i = 1, 2, 3$,

$$|\tilde{A}_{ij}(\vartheta_1) - A_{ij}(\vartheta_2)| \leqslant ||\nabla_\vartheta \tilde{A}_{ij}(\vartheta^*_{(1)})||||\vartheta_1 - \vartheta_2||, \tag{F.20}$$

$$N|\breve{A}_{ij}(\vartheta_1) - \breve{A}_{ij}(\vartheta_2)| \leqslant ||N\nabla_\vartheta \breve{A}_{ij}(\vartheta^*_{(2)})||||\vartheta_1 - \vartheta_2||, \tag{F.21}$$

$$|A(\vartheta)Z\gamma_0 - A(\vartheta)Z\gamma_0|_{ij} \leqslant ||\nabla_\vartheta A(\vartheta^*_{(3)}))Z\delta||_{ij}||\vartheta_1 - \vartheta_2||. \tag{F.22}$$

By (b), there exists some $K_A < \infty$ such that $||\nabla_\vartheta \tilde{A}_{ij}(\vartheta^*_{(1)})|| < K_A$ and $||N\nabla_\vartheta \breve{A}_{ij}(\vartheta^*_{(2)})|| < K_A$ uniformly. Therefore $|\tilde{A}_{ij,N}(\vartheta_1) - \tilde{A}_{ij,N}(\vartheta_2)| \leqslant K_A||\vartheta_1 - \vartheta_2||$, $N|\breve{A}_{ij,N}(\vartheta_1) - \breve{A}_{ij,N}(\vartheta_2)| \leqslant K_A||\vartheta_1 - \vartheta_2||$.

Since $\nabla_\vartheta A(\vartheta^*_{(3)}) = \nabla_\vartheta \tilde{A}(\vartheta^*_{(3)} + \nabla_\vartheta \breve{A}(\vartheta^*_{(3)})$, $\nabla_\vartheta \tilde{A}(\vartheta^*_{(3)} \in \mathscr{G}$ and $\nabla_\vartheta \breve{A}(\vartheta^*_{(3)}) = O_U(1/N)$. By Lemma B.2, the elements of $\nabla_\vartheta A(\vartheta^*_{(3)})Z\delta$ are uniformly bounded in absolute value. Therefore, there exists some $K_\eta < \infty$ such that $|A(\vartheta)Z\gamma_0 - A(\vartheta)Z\gamma_0|_{ij} \leqslant K_\eta||\vartheta_1 - \vartheta_2||$.

(e) Let $A^\star = A(\vartheta)\Omega_0$. Since $\Omega_0 \in \mathscr{A}$, $A^\star(\vartheta) = \prod_{k=1}^{K+1} A_k(\vartheta)$, with $A_k(\vartheta) \in \mathscr{A}$. Therefore, $A^\star(\vartheta) = \tilde{A}^\star(\vartheta) + \breve{A}^\star(\vartheta)$. By (d), there exists some $K_{A^\star} < \infty$ such that for all $\vartheta_1, \vartheta_2 \in \mathit{\Theta}$, $|\tilde{A}^\star_{ij,N}(\vartheta_1) - \tilde{A}^\star_{ij,N}(\vartheta_2)| \leqslant K_{A^\star}||\vartheta_1 - \vartheta_2||$, $N|\breve{A}^\star_{ij,N}(\vartheta_1) - \breve{A}^\star_{ij,N}(\vartheta_2)| \leqslant K_{A^\star}||\vartheta_1 - \vartheta_2||$. Therefore,

$$\begin{aligned}
\frac{1}{N_R}|tr(A^\star(\vartheta_1)) - tr(A^\star(\vartheta_2))| &= \frac{1}{N_R}|tr[\tilde{A}^\star(\vartheta_1) - \tilde{A}^\star(\vartheta_2)]| + \frac{1}{N_R}|tr[\breve{A}^\star(\vartheta_1) - \breve{A}^\star(\vartheta_2)]| \\
&\leqslant \frac{1}{N_R}\sum_{l=1}^{N} K_{A^\star}||\vartheta_1 - \vartheta_2|| + \frac{1}{N_R}\sum_{l=1}^{N}\frac{1}{N_R}K_{A^\star}||\vartheta_1 - \vartheta_2|| \\
&\leqslant K_{A^\star}(1 + \frac{1}{N_R})||\vartheta_1 - \vartheta_2|| < 2K_{A^\star}||\vartheta_1 - \vartheta_2||. \quad \text{(F.23)}
\end{aligned}$$

Therefore, $\frac{1}{N_R}tr(A(\vartheta)\Omega_0)$ is uniformly continuous .

The lemma above shows that the products with matrices in $\mathscr{A}$ can be divided into two types of matrices, one from $\mathscr{G}$, the other has elements that are uniformly of order $1/N$. Below I will show the uniform convergence of the quadratic forms associated with each type. $\square$

Theorem F.1. *For a quadratic form $S_N(\vartheta) = U'A_N(\vartheta)U + U'B_N(\vartheta)Z\gamma_0$, where $A_N(\vartheta), B_N(\vartheta) \in \mathscr{G}$. Under Assumptions 1, 2, and 3, the quadratic form $S_N(\vartheta)$ converge to its mean uniformly in ϑ, $sup_{\vartheta\in\Theta}\frac{1}{N_R}|S_N(\vartheta) - E[S_N(\vartheta)]| \to_p 0$ as $R \to \infty$.*

Proof. First I will prove the point-wise convergence. Since $A_N(\vartheta) \in \mathscr{G}$, $A_N(\vartheta) = diag_{c=1}^{R}\{A_{cc}(\vartheta)\}$. By Lemma C.1, the expected value of the quadratic form is

$$E(S_N(\vartheta)) = \sum_{c=1}^{R} tr(\Omega_{c0}A_{cc,N}(\vartheta)). \tag{F.24}$$

The deviation of the quadratic form from its mean is the sum of R independent quadratic forms as follows:

$$S_N(\vartheta) - E(S_N(\vartheta)) = \sum_{c=1}^{R} v_{c,N}(\vartheta), \tag{F.25}$$

$$v_{c,N}(\vartheta) = [U_c'A_{cc,N}(\vartheta)U_c - tr(\Omega_{c0}A_{cc,N}(\vartheta))] + \eta_c'(\vartheta)U_c. \tag{F.26}$$

Since U_c are independent across all c, v_c are independently distributed across c, with $E(v_c) = 0$. By Lemma C.1, the variance of $v_{c,N}(\vartheta)$ is

$$\begin{aligned} Var(v_c(\vartheta)) = {} & 2tr[A_{cc}(\vartheta)\Omega_{c0}A_{cc}(\vartheta)\Omega_{c0}] + \eta_c'(\vartheta)\Omega_{c0}\eta_c(\vartheta) \\ & + \sum_{i=1}^{n_c}[A_{ii,cc}(\vartheta)]^2(\mu_\epsilon^{(4)} - 3\sigma_\epsilon^4) + tr(A_{cc}(\vartheta)J_cA_{cc}(\vartheta)J_c)(\mu_\alpha^{(4)} - 3\sigma_{\alpha0}^2) \\ & + 2[\mu_\alpha^{(3)}\iota_c'A_{cc}(\vartheta)\iota_c\eta_c'(\vartheta)\iota_c + \mu_\epsilon^{(3)}\sum_{i=1}^{n_c}A_{ii,cc}(\vartheta)\eta_{ic}]. \end{aligned} \tag{F.27}$$

By Lemma F.1, the elements of $A(\vartheta)$ and $\eta(\vartheta)$ are uniformly bounded. Under Assumptions 1-5, n_c is bounded, $\mu_\alpha^{(3)}$, $\mu_\epsilon^{(3)}$, $\mu_\alpha^{(4)}$, and $\mu_\epsilon^{(4)}$ are bounded. Therefore $Var(v_c)$ is uniformly bounded. Using law of large numbers, for all $\vartheta \in \Theta$, $\frac{1}{N_R}(S(\vartheta) - \mu_S(\vartheta)) \to_p 0$ as R goes to infinity.

Above has proven point-wise convergence on Θ. Meanwhile, Θ is compact under Assumptions 1, 1, and 4. According to Lemma F.1(b), $\frac{1}{N}E(S(\vartheta))$ is uniformly continuous. Therefore, according to Corollary 2.2 of Newey (1991), it suffices to show that for any $\vartheta_1, \vartheta_2 \in \Theta$, $\frac{1}{N_R}|S(\vartheta_1) - S(\vartheta_2)| \leqslant C_N h(d(\vartheta_1, \vartheta_2))$, where C_N is $O_p(1)$, $h : [0, \infty) \to [0, \infty)$, with $h(0) = 0$.

Using Holder's inequality,

$$
\begin{aligned}
&\frac{1}{N_R}|S(\vartheta_1) - S(\vartheta_2)| \\
&= \frac{1}{N_R}|\sum_{c=1}^{R}\{U_c'[A_{cc}(\vartheta_1) - A_{cc}(\vartheta_2)]U_c + U_c'[(\eta_c(\vartheta_1) - \eta_c(\vartheta_2)]| \\
&\leqslant \frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}|[\alpha_c^2 + \alpha_c\epsilon_{ic} + \alpha\epsilon_{jc} + \epsilon_{ic}\epsilon_{jc}][A_{ij,cc}(\vartheta_1) - A_{ij,cc}(\vartheta_2)]| \\
&+ \frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}|\alpha_c[(\eta_c(\vartheta_1) - \eta_c(\vartheta_2)]_i + \epsilon_{ic}[(\eta_{ic}(\vartheta_1) - \eta_{ic}(\vartheta_2)]| \\
&\leqslant [(\frac{1}{N_R}\sum_{c=1}^{R}n_c^2\alpha_c^4)^{1/2} + 2(\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}n_c\alpha_c^2\epsilon_{ic}^2)^{1/2} + (\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2)^{1/2}] \\
&\times \{\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}[A_{ij,cc}(\vartheta_1) - A_{ij,cc}(\vartheta_2)]^2\}^{1/2} \\
&+ [(\frac{1}{N_R}\sum_{c=1}^{R}n_c\alpha_c^2)^{1/2} + (\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2}][\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}(\eta_{ic}(\vartheta_1) - \eta_{ic}(\vartheta_2)^2]^{1/2} \\
&\leqslant [(\frac{\bar{a}}{N}\sum_{c=1}^{R}\alpha_c^4)^{1/2} + 2(\frac{\bar{a}}{N}\sum_{c=1}^{R}\alpha_c^2\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2} + (\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^4 + \frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2)^{1/2}] \\
&\times \{\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}[A_{cc}(\vartheta_1) - A_{cc}(\vartheta_2)]_{ij}^2\}^{1/2} \\
&+ [(\frac{\bar{a}}{N}\sum_{c=1}^{R}\alpha_c^2)^{1/2} + (\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2}][\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}(\eta_{ic}(\vartheta_1) - \eta_{ic}(\vartheta_2))^2]^{1/2} \\
&\leqslant C_N\tilde{g}(\vartheta_1, \vartheta_2), \qquad \text{(F.28)}
\end{aligned}
$$

where

$$C_N = [(\frac{\bar{a}R}{N}\frac{1}{R}\sum_{c=1}^{R}\alpha_c^4)^{1/2} + 2(\frac{\bar{a}}{N}\sum_{c=1}^{R}\alpha_c^2\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2} + (\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^4 + \frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2)^{1/2}] \tag{F.29}$$

$$+ [(\frac{\bar{a}R}{N}\frac{1}{R}\sum_{c=1}^{R}\alpha_c^2)^{1/2} + (\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2}][\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}(\eta_{ic}(\vartheta_1) - \eta_{ic}(\vartheta_2))^2]^{1/2},$$

$$\tilde{g}(\vartheta_1, \vartheta_2) = \{\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}[A_{ij,cc}(\vartheta_1) - A_{ij,cc}(\vartheta_2)]^2\}^{1/2} + [\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}(\eta_{ic}(\vartheta_1) - \eta_{ic}(\vartheta_2))^2]^{1/2}. \tag{F.30}$$

Next I will show that C_N is $O_p(1)$. By Assumption 1, α_c are identically and independently distributed with $E(\alpha_c^2) = \sigma_{\alpha_0}^2$ and $E\alpha_c^4 = \mu_\alpha^{(4)} < \infty$, therefore $\frac{1}{R}\sum_{c=1}^{R}\alpha_c^2 \to_p \sigma_{\alpha 0}^2$, and $\frac{1}{R}\sum_{c=1}^{R}\alpha_c^4 \to_p \mu_\alpha^{(4)}$. By Assumption 2, ϵ_{ic} are identically and independently distributed across i and c with $E(\epsilon_{ic}^2) = \sigma_{\epsilon_0}^2$ and $E\epsilon_c^4 = \mu_\epsilon^{(4)} < \infty$, therefore $\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2 \to_p \sigma_{\epsilon 0}^2$, and $\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^4 \to_p \mu_\epsilon^{(4)}$. Since $\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2$ are independently distributed across c, with $E[\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2] = (n_c-1)n_c\sigma_{\epsilon 0}^4$,

$$\begin{aligned}
Var[\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2] &= E\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\sum_{s=1}^{n_c}\sum_{k\neq s}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2] - [(n_c-1)n_c\sigma_{\epsilon 0}^4]^2 \\
&= E[2\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^4\epsilon_{jc}^4 + 4\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\sum_{k\neq i,,j}^{n_c}\epsilon_{ic}^4\epsilon_{jc}^2\epsilon_{kc}^2 \\
&+ \sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\sum_{s\neq i,j}^{n_c}\sum_{k\neq i,js}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2\epsilon_{sc}^2\epsilon_{kc}^2] - (n_c-1)^2 n_c^2\sigma_{\epsilon 0}^8 && \text{(F.31)} \\
&= 2n_c(n_c-1)(\mu_\epsilon^{(4)})^2 + 4n_c(n_c-1)(n_c-2)(\mu_\epsilon^4\sigma_{\epsilon 0}^4) && \text{(F.32)} \\
&+ n_c(n_c-1)(n_c-2)(n_c-3)\sigma_{\epsilon 0}^8. && \text{(F.33)}
\end{aligned}$$

Since $\mu_\epsilon^{(4)} < \infty$ and n_c is bounded, $Var(\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2) < \infty$. Therefore $\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{j\neq i}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2 \to_p \frac{1}{N_R}\sum_{c=1}^{R}n_c(n_c-1)\sigma_{\epsilon 0}^4$. Finally, $\alpha_c^2\sum_{i=1}^{n_c}\epsilon_{ic}^2$ are independently distributed with $E\alpha_c^2 \sum_{i=1}^{n_c} \epsilon_{ic}^2 = n_c\sigma_{\alpha0}^2\sigma_{\epsilon0}^2$, and

$$Var[\alpha_c^2 \sum_{i=1}^{n_c} \epsilon_{ic}^2] = E(\alpha_c^4)E(\sum_{i=1}^{n_c} \epsilon_{ic}^2)^2 - (n_c\sigma_{\alpha0}^2\sigma_{\epsilon0}^2)^2$$

$$= \mu_\alpha^{(4)}[n_c\mu_\epsilon^{(4)} + n_c(n_c-1)\sigma_{\epsilon0}^4] - n_c^2\sigma_{\alpha0}^4\sigma_{\epsilon0}^4.$$

Therefore, $\frac{1}{N}\sum_{c=1}^R \alpha_c^2 \sum_{i=1}^{n_c} \epsilon_{ic}^2 \to_p \frac{1}{N}\sum_{c=1}^R n_c\sigma_{\alpha0}^2\sigma_{\epsilon0}^2 = \sigma_{\alpha0}^2\sigma_{\epsilon0}^2$. In all, $C_N = O_p(1)$.

By Lemma F.1(d), there exists some $K_A < \infty$ and $K_\eta < \infty$ such that for all i, j, c and $\vartheta_1, \vartheta_2 \in \Theta$,

$$|A(\vartheta_1) - A(\vartheta_2)|_{ij,cc} \leqslant K_A||\vartheta_1 - \vartheta_2||, \tag{F.34}$$

$$([A(\vartheta_1) - A(\vartheta_2)]Z\gamma_0)_{ic} \leqslant K_\eta||\vartheta_1 - \vartheta_2||. \tag{F.35}$$

Therefore,

$$\tilde{g}(\vartheta_1, \vartheta_2) \leqslant \{\frac{1}{N_R}\sum_{c=1}^R\sum_{i=1}^{n_c}\sum_{j=1}^{n_c} K_A^2||\vartheta_1 - \vartheta_2||^2\}^{1/2} + [\frac{1}{N_R}\sum_{c=1}^R\sum_{i=1}^{n_c} K_\eta^2||\vartheta_1 - \vartheta_2||]^{1/2} \tag{F.36}$$

$$= (K_A + K_\eta)||\vartheta_1 - \vartheta_2||. \tag{F.37}$$

□

The following theorem describes the central limit theorem of the quadratic form $S_N(\vartheta)$ defined in equation (C.1).

Theorem F.2. *For a quadratic form $S_N(\vartheta) = U'A_N(\vartheta)U + U'B_NZ\gamma_0$, if*

(a) The elements of $N_RA(\vartheta)$ and $N_RB(\vartheta)$are uniformly bounded in absolute value;

(b) There exist some$K_A < \infty$ and $K_\eta < \infty$ such that for all $\vartheta_1, \vartheta_2 \in \Theta$, $N|A_{ij,N}(\vartheta_1) - A_{ij,N}(\vartheta_2)| \leqslant K_A|\vartheta_1 - \vartheta_2|$ and $|\eta_{i,N}(\vartheta) - \eta_{j,N}(\vartheta)| \leqslant K_\eta|\vartheta_1 - \vartheta_2|$;

(c) The trace $\frac{1}{N}tr[\Omega_0 A_N(\vartheta)]$ is uniformly continuous in ϑ.

Then $sup_{\vartheta\in\Theta}\frac{1}{N_R}|S(\vartheta)-E(S_N(\vartheta))|\rightarrow_p 0$ as $R\rightarrow\infty$.

Proof. Let $\eta(\vartheta)=U'B_N Z\gamma_0$, using Lemma C.1,

$$\begin{aligned}Var(S(\vartheta)) &= tr[\Omega_0 A(\vartheta)\Omega_0(A(\vartheta)+A'(\vartheta))]+\eta'(\vartheta)\Omega_0\eta(\vartheta)\\ &\quad+(\mu_\epsilon^{(4)}-3\sigma_\epsilon^4)\sum_{c=1}^{R}\sum_{i=1}^{n_c}(A_{ii,cc}(\vartheta))^2+(\mu_\alpha^{(4)}-3\sigma_{\alpha 0}^2)\sum_{c=1}^{R}tr(A_{cc}(\vartheta)J_c)^2\\ &\quad+2\mu_\epsilon^{(3)}\sum_{c=1}^{R}\sum_{i=1}^{n_c}A_{ii,cc}(\vartheta)\eta_{ic}(\vartheta)+2\mu_\alpha^{(3)}\sum_{c=1}^{R}\eta_c'(\vartheta)\iota_c\iota_c'A_{cc}(\vartheta)\iota_c. \qquad (F.38)\end{aligned}$$

Since the elements of $A_N(\vartheta)$ and $B_N(\vartheta)$ are uniformly of order $1/N$, using Lemma B.2, $\Omega_0 A(\vartheta)\Omega_0[A(\vartheta)+A'(\vartheta)]$ is $O_U(1/N)$ and, so $tr[\Omega_0 A(\vartheta)\Omega_0(A(\vartheta)+A'(\vartheta))]=O_U(1)$. Since A is $O_U(1/N)$, $\sum_{c=1}^{R}\sum_{i=1}^{n_c}(A_{ii,cc}(\vartheta))^2$ is $O_U(1/N)$, $\sum_{c=1}^{R}tr(A_{cc}(\vartheta)J_c)^2$ is $O_U(1/N)$. Since $B_N(\vartheta)$ is $O_U(1/N)$, $\eta(\vartheta)$ is $O_U(1)$, (Lemma B.2) by Lemma B.2, $\eta'(\vartheta)\Omega_0\eta(\vartheta)=\sum_{c=1}^{R}\eta_c'\Omega_{c0}\eta_c$ is $O(N)$ as η_c, Ω_{c0} and n_c are bounded. Moreover, $A_{ii,cc}(\vartheta)\eta_{ic}(\vartheta)$ and $\eta_c'(\vartheta)\iota_c\iota_c'A_{cc}(\vartheta)\iota_c$ are $O_U(1/N)$, $\sum_{c=1}^{R}\sum_{i=1}^{n_c}A_{ii,cc}(\vartheta)\eta_{ic}(\vartheta)=O_U(1)$ and $\sum_{c=1}^{R}\eta_c'(\vartheta)\iota_c\iota_c'A_{cc}(\vartheta)\iota_c=O_U(1)$.

In all, $Var(S(\vartheta)$ is $O(N)$, hence $Var\{\frac{1}{N_R}S(\vartheta)]\}=\frac{1}{N_R^2}Var(S_N(\vartheta))\rightarrow 0$ as $R\rightarrow\infty$. By Chebychev's inequality, for all $\vartheta\in\Theta$, $\frac{1}{N_R}(S(\vartheta)-E(S(\vartheta)))\rightarrow_p 0$ as R goes to infinity.

Note that $\frac{1}{\sqrt{N_R}}(S(\vartheta)-\mu_S(\vartheta))\rightarrow_p 0$ implies $\frac{1}{N_R}(S(\vartheta)-E(S(\vartheta)))\rightarrow_p 0$. Above has proven point-wise convergence of $\frac{1}{N_R}(S(\vartheta)-E(S(\vartheta)))$ to 0 on Θ. Meanwhile, Θ is compact under Assumptions 1, 1, and 4. From Lemma C.1, $E(S(\vartheta))=tr(\Omega_0 A(\vartheta))$. By assumption, $E(\frac{1}{N}S(\vartheta))$ is uniformly continuous in Θ. Therefore, according to Corollary 2.2 of Newey (1991), it suffices to show that for any $\vartheta_1,\vartheta_2\in\Theta$, $\frac{1}{N_R}|S(\vartheta_1)-S(\vartheta_2)|\leqslant B_N h(d(\vartheta_1,\vartheta_2))$, where B_N is $O_p(1)$, $h:[0,\infty)\rightarrow[0,\infty)$, with $h(0)=0$.

Using Holder's inequality,

$$\frac{1}{N_R}|S(\vartheta_1)-S(\vartheta_2)|$$

$$= \frac{1}{N_R}|\sum_{c=1}^{R}\sum_{r=1}^{R}U_c'[A_{cr}(\vartheta_1)-A_{cr}(\vartheta_2)]U_r+\sum_{c=1}^{R}U_c'(\eta_c(\vartheta_1)-\eta_c(\vartheta_2))]| \tag{F.39}$$

$$= \frac{1}{N_R}|\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}[(\alpha_c\alpha_r+\alpha_c\epsilon_{jr}+\alpha_r\epsilon_{ic}+\epsilon_{ic}\epsilon_{jr})(A_{ij,cr}(\vartheta_1)-A_{ij,cr}(\vartheta_2))]$$

$$+\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}[\alpha_c(\eta_{ic}(\vartheta_1)-\eta_{ic}(\vartheta_2)+\epsilon_{ic}(\eta_{ic}(\vartheta_1)-\eta_{ic}(\vartheta_2))]|$$

$$\leqslant [(\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}n_cn_r\alpha_c^2\alpha_r^2)^{1/2}+(\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{j=1}^{n_r}n_c\alpha_c^2\epsilon_{jr}^2)^{1/2}+(\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{i=1}^{n_c}n_r\alpha_r^2\epsilon_{ic}^2)^{1/2}$$

$$+(\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}\epsilon_{ic}^2\epsilon_{jc}^2)^{1/2}]\{\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}(NA_{ij,cr}(\vartheta_1)-NA_{ij,cr}(\vartheta_2))^2\}^{1/2}$$

$$+[(\frac{1}{N_R}\sum_{c=1}^{R}n_c\alpha_c^2)^{1/2}+(\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2}][\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}(\eta_{ic}(\vartheta_1)-\eta_{ic}(\vartheta_2)^2]^{1/2}$$

$$\leqslant B_N g(\vartheta_1,\vartheta_2), \tag{F.40}$$

where

$$B_N = (\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}n_cn_r\alpha_c^2\alpha_r^2)^{1/2}+2(\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{j=1}^{n_r}n_c\alpha_c^2\epsilon_{jr}^2)^{1/2}+(\frac{1}{N^2}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\sum_{r=1}^{R}\sum_{j=1}^{n_r}\epsilon_{ic}^2\epsilon_{jr}^2)^{1/2} \tag{F.41}$$

$$+(\frac{1}{N_R}\sum_{c=1}^{R}n_c\alpha_c^2)^{1/2}+(\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2},$$

$$=(\frac{1}{N}\sum_{c=1}^{R}n_c\alpha_c^2)+2(\frac{1}{N}\sum_{c=1}^{R}n_c\alpha_c^2)^{1/2}(\frac{1}{N}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2}+(\frac{1}{N}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2) \tag{F.42}$$

$$+(\frac{1}{N_R}\sum_{c=1}^{R}n_c\alpha_c^2)^{1/2}+(\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2)^{1/2}, \tag{F.43}$$

$$g(\vartheta_1,\vartheta_2)=\{\frac{1}{N^2}\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c}(NA_{ij,cr}(\vartheta_1)-NA_{ij,cr}(\vartheta_2))^2\}^{1/2}+[\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}(\eta_{ic}(\vartheta_1)-\eta_{ic}(\vartheta_2)^2]^{1/2}. \tag{F.44}$$

By assumption, $|NA_{ij,cr}(\vartheta_1) - NA_{ij,cr}(\vartheta_2)| \leqslant K_A|\vartheta_1 - \vartheta_2|$, and $|\eta_{ic}(\vartheta_1) - \eta_{ic}(\vartheta_2)| \leqslant K_\eta|\vartheta_1 - \vartheta_2|$. Therefore,

$$g(\vartheta_1, \vartheta_2) \leqslant \{\frac{1}{N_R^2}\sum_{c=1}^{R}\sum_{r=1}^{R}\sum_{i=1}^{n_c}\sum_{j=1}^{n_c} K_A^2||\vartheta_1 - \vartheta_2||^2\}^{1/2} + [\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c} K_\eta^2||\vartheta_1 - \vartheta_2||^2]^{1/2}.$$

$$= K_A||\vartheta_1 - \vartheta_2|| + K_\eta||\vartheta_1 - \vartheta_2|| = (K_A + K_\eta)||\vartheta_1 - \vartheta_2||. \tag{F.45}$$

Therefore, $h(||\vartheta_1 - \vartheta_2||) = (K_A + K_\eta)||\vartheta_1 - \vartheta_2||$. Next I will demonstrate that $B_R = O_p(1)$.

Since α_c^2 are $i.i.d$ across c, ϵ_{jr}^2 are $i.i.d$ across j and r, by law of large numbers, $\frac{1}{R}\sum_{c=1}^{R}\alpha_c^2 \to_p \sigma_{\alpha 0}^2$, and $\frac{1}{N_R}\sum_{c=1}^{R}\sum_{i=1}^{n_c}\epsilon_{ic}^2 \to_p \sigma_{\epsilon 0}^2$. Therefore, $B_R = O_p(1)$. □

Corollary F.1. *Suppose* $S_N(\vartheta) = U'A_N(\vartheta)U + U'B_N(\vartheta)Z\gamma_0$, *where* $A_N(\vartheta) = \prod_{k=1}^{K} A_{k,N}(\vartheta)$, $B_N(\vartheta) = \prod_{j=1}^{J} B_{j,N}(\vartheta)$, $A_{k,N}(\vartheta) \in \mathscr{A}$, *and* $B_{j,N}(\vartheta) \in \mathscr{A}$, $\mathscr{A}$ *is a matrix set defined in equation* (F.2). *Then under Assumptions 1-5,* $sup_{\vartheta\in\Theta}\frac{1}{N_R}|S(\vartheta) - E(S(\vartheta))| \to_p 0$ *as* R *goes to infinity.*

Proof. From Lemma F.1(a),$A(\vartheta) = \tilde{A}(\vartheta) + \breve{A}(\vartheta)$, where $\tilde{A}(\vartheta) \in \mathscr{G}$ is a block diagonal matrix, $\breve{A}(\vartheta) = O_U(1/N)$. Rewrite $S(\vartheta)$ as

$$S(\vartheta) = [U'\tilde{A}(\vartheta)U + U'B(\vartheta)Z\gamma_0] + U'\breve{A}(\vartheta)U. \tag{F.46}$$

By Lemma C.1,

$$E(S(\vartheta)) = tr(\Omega_0\tilde{A}(\vartheta)) + tr(\Omega_0\breve{A}(\vartheta)). \tag{F.47}$$

By Lemma F.1(c), $B(\vartheta)Z\gamma_0$ is $O_U(1)$. From Lemma F.1(d), there exists some $K_A < \infty$, $K_\eta < \infty$, such that for all $\vartheta_1, \vartheta_2 \in \Theta$, $|\tilde{A}_{ij,N}(\vartheta_1) - \tilde{A}_{ij,N}(\vartheta_2)| \leqslant K_A||\vartheta_1 - \vartheta_2||$, $N|\breve{A}_{ij,N}(\vartheta_1) - \breve{A}_{ij,N}(\vartheta_2)| \leqslant K_A||\vartheta_1 - \vartheta_2||$, $|B(\vartheta_1)Z\gamma_0 - B(\vartheta_2)Z\gamma_0| < K_\eta||\vartheta_1 - \vartheta_2||$. From Lemma F.1(e), $\frac{1}{N_R}tr(\Omega_0\tilde{A}(\vartheta))$ and $\frac{1}{N_R}tr(\Omega_0\breve{A}(\vartheta))$ are uniformly continuous.

By Theorem F.2,

$$sup_{\vartheta\in\Theta}\frac{1}{N_R}[U'\breve{A}(\vartheta)U - tr(\Omega_0\breve{A}(\vartheta))] \rightarrow_p \infty. \tag{F.48}$$

By Theorem F.1,

$$sup_{\vartheta\in\Theta}\frac{1}{N_R}|U'\tilde{A}(\vartheta)U + U'B(\vartheta)Z\gamma - tr(\Omega_0\tilde{A}(\vartheta)]| \rightarrow_p 0. \tag{F.49}$$

Therefore, $sup_{\vartheta\in\Theta}\frac{1}{N_R}|S(\vartheta) - E(S(\vartheta))| \rightarrow_p 0$ as R goes to infinity. □

Appendix G: Estimation of the Heterogeneous Peer Effects Model

This appendix discusses about the estimation strategy of the heterogeneous peer effects model in Section 4.4.4. Let $Y_c = (y_{1c}, ..., y_{n_c c})'$, $X_c = (x_{1c}, ..., x_{n_c c})'$, and $\epsilon_c = (\epsilon_{1c}, ..., \epsilon_{n_c c})'$. The matrix form of the heterogeneous peer effects model in equations (4.20) and (4.21) is

$$\begin{aligned} Y_c &= \iota_c \beta_0 + \sum_{p=1}^{4} \lambda_p W_{p,c} Y_c + X_c \beta + \sum_{p=1}^{4} W_{p,c} X_c \gamma_p + \iota_c \psi_c' \pi + \iota_c f_c' \phi + \iota_c \alpha_c + \epsilon_c \\ &= \sum_{p=1}^{4} \lambda_p W_{p,c} Y_c + Z_c \delta + U_c, \end{aligned} \tag{G.1}$$

where $W_{1,c}$, $W_{2,c}$, $W_{3,c}$ and $W_{4,c}$ are weight matrices. If i and j are both boys in class c and $i \neq j$, $(W_{1,c})_{ij} = 1/(n_c^b - 1)$, otherwise $(W_{1,c})_{ij} = 0$. If i is a boy and j is a girl,$(W_{2,c})_{ij} = 1/n_c^g$, otherwise $(W_{2,c})_{ij} = 0$. If i is a girl and j is a boy, $(W_{3,c})_{ij} = 1/n_c^b$, otherwise $(W_{3,c})_{ij} = 0$. If both i and j are girls 0 and $i \neq j$, $(W_{4,c})_{ij} = 1/(n_c^g - 1)$, otherwise $(W_{4,c})_{ij} = 0$. The matrix Z_c includes all the exogenous variables,

$$Z_c = (\iota_c, X_c, W_{1,c}X_c, W_{2,c}X_c, W_{3,c}X_c, W_{4,c}X_c, \iota_c\psi_c', \iota_c f_c').$$

The vector $\delta = (\beta_0, \beta', \gamma_1', \gamma_2', \gamma_3', \gamma_4', \pi', \phi')'$ is the vector of corresponding coefficients. The model for the whole sample is

$$Y = \sum_{p=1}^{4} \lambda_p W_p Y + Z\delta + U, \tag{G.2}$$

where $Y = (Y_1', ..., Y_R')'$, $Z = (Z_1', ..., Z_R')'$, $U = (U_1', ..., U_R')'$, $W_p = diag(W_{1,p}, ..., W_{R,p})$.

The variance-covariance matrix is the same as in the main model, and is defined in equation (3.7). The distribution of Y is

$$Y \sim ((I - \sum_{p=1}^{4} \lambda_p W_p)^{-1} Z\delta, (I - \sum_{p=1}^{4} \lambda_p W_p)^{-1} \Omega (I - \sum_{p=1}^{4} \lambda_p W_p')^{-1}).$$

The model can be estimated with maximum likelihood method, with the log likelihood function being

$$\begin{aligned} lnL &= -\frac{N}{2} ln(2\pi) + ln|I - \sum_{p=1}^{4} \lambda_p W_p| - \frac{1}{2} ln|\Omega| \\ &\quad -\frac{1}{2}(Y - \sum_{p=1}^{4} \lambda_p W_p Y - Z\delta)' \Omega^{-1} (Y - \sum_{p=1}^{4} \lambda_p W_p Y - Z\delta). \end{aligned} \tag{G.3}$$

Appendix H: Graham's Conditional Variance Method

Graham (2008) directly explores the relationship between within-class variance and between-class variance conditional on different class sizes. The original Graham (2008) model is

$$y_{ic} = \nu_c + (\gamma_0 - 1)\bar{\epsilon}_c + \epsilon_{ic}. \tag{H.1}$$

where ϵ_{ic} are unobserved individual characteristics, $\bar{\epsilon}_c = \sum_{i=1}^{n_c} \epsilon_{ic}$ is the full mean of ϵ, and ν_c is the random class effect. The model can be written as a linear-in-means model with endogenous effects and random class effects as

$$\begin{aligned} y_{ic} &= (1 - \frac{1}{\gamma_0})\bar{y}_c + \frac{\nu_c}{\gamma_0} + \epsilon_{ic} \qquad \text{(H.2)} \\ &= \lambda \bar{y}_c + \alpha_c + \epsilon_{ic}, \end{aligned}$$

where $\lambda = 1 - \frac{1}{\gamma_0}$, $\alpha_c = \nu_c/\gamma_0$.

Equation (H.2) is similar to the main model in equation (3.2) except that: (a) equation (H.2) uses full-means rather than leave-out means. Hence I will refer the model as the full-mean model; (b) it does not include individual or class characteristics.[1] Equivalently, equation (H.1) is comparable to the reduced form of the main model in equation (4.16).

[1] The Graham (2008) model can include individual characteristics and characteristics, but only if peers effect work through the full-mean rather than the leave-out mean. An extended version of model H.1 is

$$y_{ic} = \theta_0 + x'_{ic}\theta_1 + \bar{x}'_c\theta_2 + \psi'_c\theta_3 + \nu_c + (\gamma_0 - 1)\bar{\epsilon}_c + \epsilon_{ic}$$

Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be the index set of small classes and regular/aid classes respectively. If $c \in \mathcal{C}_1$, class c is a small class. If $c \in \mathcal{C}_2$, class c is a regular/aid class. Graham (2008) makes three assumptions:

(1) The disturbance terms ϵ_{ic} are independent of class effect ν_c , and ϵ_{ic} are independently distributed within each class. Denote the variance of ϵ_{ic} as $\sigma^2_{\epsilon c}$. It represents "heterogeneity of peer quality". It varies across class types: $\sigma^2_{\epsilon c} = \sigma^2_1$ if $c \in \mathcal{C}_1$ and class c is a small class, $\sigma^2_{\epsilon c} = \sigma^2_2$ if $c \in \mathcal{C}_2$ and class c is a regular/aid class.

(2) The variance of random class effects is $var(\nu_c) = \sigma^2_\nu$ for all c.

These two assumptions are similar Assumptions 1 and 2, expect that I assume homoscedasity of ϵ_{ic} across all classes while Graham (2008) assume homoscedasity within each class type. The between class variance is $var(\bar{y}_c) = \sigma^2_\nu + \frac{\gamma_0^2 \sigma^2_{\epsilon c}}{n_c}$ and the within-class variance is $var(y_{ic} - \bar{y}_c) = \frac{n_c - 1}{n_c}\sigma^2_{\epsilon c}$. Therefore,

$$var(\bar{y}_c) = \sigma^2_\nu + \gamma_0^2 \frac{var(y_{ic} - \bar{y}_c)}{n_c - 1}. \tag{H.3}$$

Since $\frac{var(y_{ic} - \bar{y}_c)}{n_c - 1} = \frac{\sigma^2_{\epsilon c}}{n_c}$ is a function of class size, and assignment into small classes and large classes is random, Graham (2008) instruments $\frac{var(y_{ic} - \bar{y}_c)}{n_c - 1}$ with an indicator for small class and estimates γ_0^2. The sample within-class variance and

The model can be written as a linear-in-means model with endogenous effect and random class effect as:

$$\begin{aligned} y_{ic} &= \frac{\theta_0}{\gamma_0} + (1 - \frac{1}{\gamma_0})\bar{y}_c + x'_{ic}\theta_1 + \bar{x}'_c \frac{\theta_2 - (\gamma_0 - 1)\theta_1}{\gamma_0} + \psi_c \frac{\theta_3}{\gamma_0} + \frac{\nu_c}{\gamma_0} + \epsilon_{ic} \\ &= \beta_0 + \lambda \bar{y}_c + x'_{ic}\beta + \bar{x}'_c\gamma + \psi_c\pi + \alpha_c + \epsilon_{ic} \end{aligned}$$

where $\beta_0 = \theta_0/\gamma_0$, $\lambda = 1 - \frac{1}{\gamma_0}$, $\beta = \theta_1$, $\pi = \frac{\theta_3}{\gamma_0}$, $\gamma = \frac{\theta_2 - (\gamma_0 - 1)\theta_1}{\gamma_0}$, $\pi = \frac{\theta_3}{\gamma_0}$, $\alpha_c = \nu_c/\gamma_0$.

between class variance are constructed from residuals from equation (H.1). For the validity of the method, Graham (2008) further assumes

(3) $[\frac{var(y_{ic}-\bar{y}_c)}{n_c-1}|small\, class] \neq [\frac{var(y_{ic}-\bar{y}_c)}{n_c-1}|large\, class]$.

Assumption (3) is the rank condition for equation (H.3). Note that $\frac{var(y_{ic}-\bar{y}_c)}{n_c-1} = \frac{1}{n_c}\sigma^2_{\epsilon c}$, and n_c is smaller for small classes than for regular/aid classes. If ϵ_{ic} is homoscedastic across classes, i.e., $\sigma^2_{\epsilon c} = \sigma^2_1 = \sigma^2_2$ for all c, then condition is automatically satisfied. Note that this assumption implicitly assumes that there is variation in class size and class size n_c does not go to infinity.

Graham (2008)'s innovative method captures the essence of peer effect identification: utilization of between-class variance and within-class variance. But the method has its limitation. It identifies γ_0^2 but not λ directly. It works only for the full-mean specification. Most importantly, applying the method to other datasets is difficult. The method works specifically for the Tennessee Project STAR, where teachers and students were randomly assigned to two types of classes. Graham's method does not work if classes cannot be grouped into two types. In contrast, my methodology works as long as the assignment of students of students into classes is random.

The method proposed in this paper can be viewed as a generalization of the Graham (2008)'s conditional variance method. It shares the model specification and the idea of using between-class variance and within-class variance for identification of Graham (2008). It also overcomes the three limitations of Graham (2008) and can be adapted for different specifications and applied to more general settings.

Graham's full-mean model in equation (H.2) can be estimated by quasi-maximum likelihood. The log likelihood function is the same as equation (3.14) of my model except that the weight matrix is now $W_c^* = \frac{1}{n_c} J_c$, and

$$\begin{aligned} lnL^* = &-\frac{N}{2} ln(2\pi) + ln|I - \lambda W| - ln|\Omega| \\ &- \frac{1}{2}(Y - \lambda W^* Y - Z\delta)'\Omega^{-1}(Y - \lambda W^* Y - Z\delta). \end{aligned} \tag{H.4}$$

Given the special form of weights matrix and variance-covariance matrix[2], the log likelihood function can be rewritten as

$$\begin{aligned} lnL^* = &-\frac{N}{2} ln(2\pi) + Rln|1-\lambda| - \sum_{c=1}^{R} \frac{n_c - 1}{2} ln(\sigma_{\epsilon c}^2) - \frac{1}{2} \sum_{c=1}^{R} ln(\sigma_\alpha^2 + \frac{1}{n_c}\sigma_{\epsilon c}^2) \\ &- \frac{1}{2} \sum_{c=1}^{R} \sum_{i=1}^{n_c} \frac{[(y_{ic} - \bar{y}_c)]^2}{\sigma_{\epsilon c}^2} - \sum_{c=1}^{R} \frac{[(1-\lambda)\bar{y}]^2}{2(\sigma_\alpha^2 + \frac{1}{n_c}\sigma_{\epsilon c}^2)}. \end{aligned} \tag{H.5}$$

The first order conditions are

$$\frac{\partial lnL^*}{\partial \lambda} = -\frac{R}{1-\lambda} + \sum_{c=1}^{R} \frac{(1-\lambda)\bar{y}_c^2}{(\sigma_\alpha^2 + \frac{1}{n_c}\sigma_{\epsilon c})} = 0, \tag{H.6}$$

$$\frac{\partial lnL^*}{\partial \sigma_p^2} = \sum_{c \in \mathcal{C}_P} -\frac{n_c - 1}{2\sigma_p^2} + \frac{1}{2} \frac{\sum_{c \in \mathcal{C}_P}^{R} \sum_{i=1}^{n_c} [(y_{ic} - \bar{y}_c)]^2}{(\sigma_p^2)^2} + \sum_{c \in \mathcal{C}_P}^{R} \frac{n_c[(1-\lambda)\bar{y}_c]^2}{2(\sigma_p^2 + n_c\sigma_\alpha^2)^2} = 0,\ p = 1, 2 \tag{H.7}$$

$$\frac{\partial lnL^*}{\partial \sigma_\alpha^2} = -\frac{1}{2} \sum_{c=1}^{R} \frac{n_c}{\sigma_{\epsilon c}^2 + n_c \sigma_\alpha^2} + \sum_{c=1}^{R} \frac{n_c^2[(1-\lambda)\bar{y}_c]^2}{2(\sigma_{\epsilon c}^2 + n_c\sigma_\alpha^2)^2} = 0. \tag{H.8}$$

Let

$$\begin{aligned} \varsigma_{1c} &= \sigma_\alpha^2 + \frac{\sigma_{\epsilon c}^2}{n_c}, & \text{(H.9)} \\ \xi_c &= \bar{y}_c^2 - \frac{\varsigma_{ic}}{(1-\lambda)^2}, & \text{(H.10)} \\ \chi_c &= n_c \sum_{c=1}^{n_c} (y_{ic} - \bar{y}_c)^2 - n_c(n_c - 1)\sigma_{\epsilon c}^2. & \text{(H.11)} \end{aligned}$$

[2]See Appendix B.

The first order conditions are equivalent to

$$\begin{aligned}\sum_{c=1}^{R}\frac{\xi_c}{\varsigma_c} &= 0,\\ \sum_{c\in\mathcal{C}_p}\frac{\chi_c}{\sigma^4_{\epsilon c}} &= 0, p=1,2,\\ \sum_{c=1}^{R}\frac{\xi_c}{\varsigma_c^2} &= 0.\end{aligned}$$

The first order conditions for the maximum likelihood are equivalent to that some weighted average of ξ_c and χ_c are 0.

Combining equations (H.9), (H.10), (H.11),

$$\bar{y}_c^2 = \frac{\sigma_\alpha^2}{(1-\lambda)^2} + \gamma_0^2\frac{\sum_{c=1}^{n_c}(y_{ic}-\bar{y}_c)^2}{n_c(n_c-1)} + \xi_c - \frac{\chi_c}{n_c^2(n_c-1)(1-\lambda)^2}. \tag{H.12}$$

Since $\alpha_c = \nu_c/\gamma_0$, $\sigma_\alpha^2/(1-\lambda)^2 = \sigma_\nu^2$. Since $var(\bar{y}_c) = \sigma_\nu^2 + \frac{\gamma_0^2\sigma_{\epsilon c}^2}{n_c}$, $E(\xi_c) = 0$. Since $E(y_{ic}-\bar{y}_c)^2 = \frac{n_c-1}{n_c}\sigma_{\epsilon c}^2$, $E(\chi_c) = 0$. Taking expectation of equation (H.12) leads to equation (H.3), the main model for Graham (2008). In all, Graham (2008)'s method is based on the moment condition that $E(\xi_c - \frac{\chi_c}{n_c^2(n_c-1)(1-\lambda)^2}) = 0$, while the maximum likelihood estimation is based $E\xi_c = 0$ and $E(\chi_c) = 0$.

Appendix I: Model and Estimation Strategy of Lee (2007)

The model in my dissertation is closely related to Lee (2007), who studies a conditional maximum likelihood estimator for the peer effect model with fixed group effects. The model in Lee (2007) is different from the one in Chapter 3 in that it assumes fixed group effect. He eliminates the group effect α_c and estimates uses within group equation.

The writeup of Lee's model is the same as equation (3.2) and the reduced form is equation (4.16). The within equation is

$$\frac{n_c - 1 + \lambda}{n_c - 1}(y_{ic} - \bar{y}_c) = (x_{ic} - \bar{x}_c)(\beta - \frac{\gamma}{n_c - 1}) + (\epsilon_{ic} - \bar{\epsilon}_c) \tag{I.1}$$

The between equation is

$$(1 - \lambda)\bar{y}_c = \beta_0 + \bar{x}_c(\beta + \gamma) + \psi_c'\pi + \alpha_c + \bar{\epsilon}_c. \tag{I.2}$$

Note that $cov(y_{ic} - \bar{y}_c, \bar{y}_c) = 0$. So $y_{ic} - \bar{y}_c$ is independent of $\bar{y}_c$ under Gaussianity. The log likelihood function for the within equation is

$$\begin{aligned} lnL_c^w = & -\frac{n_c - 1}{2}ln(2\pi) + \frac{1}{2}ln(n_c) - \frac{n_c - 1}{2}ln(\sigma_{\epsilon_c}^2) + (n_c - 1)ln(\frac{n_c - 1 + \lambda}{n_c - 1}) \\ & - \frac{1}{2\sigma_{\epsilon c}^2}\sum_{i=1}^{n_c}[\frac{n_c - 1 + \lambda}{n_c - 1}(y_{ic} - \bar{y}_c) - (x_{ic} - \bar{x}_c)(\beta - \frac{\gamma}{n_c - 1})]^2 \end{aligned} \tag{I.3}$$

The log likelihood function for the between equation is

$$lnL_c^b = -\frac{1}{2}ln(2\pi) - \frac{1}{2}ln(\frac{\sigma_{\epsilon c}^2}{n_c} + \sigma_\alpha^2) - \frac{1}{2}ln(n_c) + ln|1-\lambda| \tag{I.4}$$
$$- \frac{1}{2(\sigma_\alpha^2 + \sigma_{\epsilon c}^2/n_c)}[(1-\lambda)\bar{y}_c - \beta_0 - \bar{x}_c(\beta+\gamma) - \psi_c'\pi]^2$$

The sum of lnL_c^w and lnL_c^b over all classes is the log likelihood function in equation (4.9) of the main model.

Lee (2007) allows α_c to be correlated with exogenous variables. His conditional maximum likelihood method is based on equation (I.3), the distribution of $y_{ic} - \bar{y}_c$ conditional on $\bar{y}_c$. Identification comes from within-group variation. The method in Chapter 3 takes into account the distribution of $\bar{y}_c$ and includes equation (I.4). Identification comes from both the within-group variance and between-group variance.

Lee's fixed group effect model is plausible if the difference of group averages $\bar{y}_c$ is driven mainly by unobserved factors correlated with member and group characteristics. In the case when unobserved group characteristics are randomly distributed, i.e., in the random group effect setting, the distribution of group averages is revealing for the presence of peer effect. For example, Glaeser et al. (1996) shows that the large between-city variance of crime rates can be attributed to social interactions. The random effects model can help explain how peer effects lead to high variance of outcomes across groups. In that scenario, incorporating the distribution of $\bar{y}_c$ improves the efficiency of estimation. In the setting of my dissertation, where people are randomly assigned into groups, it is reasonable to assume random group effects.

Bibliography

[1] David R. Agrawal. The Tax Gradient: Spatial Aspects of Fiscal Competition. *American Economic Journal: Economic Policy*, 7(2):1–29, 2015.

[2] Joseph G. Altonji, Ching-I Huang, and Christopher R. Taber. Estimating the Cream Skimming Effect of School Choice. *Journal of Political Economy*, 123 (2):266–324, 2015.

[3] Joshua D. Angrist. The perils of peer effects. *Labour Economics*, 30:98–108, 2014.

[4] Joshua D. Angrist and Kevin Lang. Does School Integration Generate Peer Effects? Evidence from Boston's Metco Program. *The American Economic Review*, 94(5):1613–1634, 2004.

[5] L Anselin. *Spatial Econometrics: Methods and Models*, volume 4. Springer Science & Business Media, 1988.

[6] Luc Anselin. Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1):3–25, 2010.

[7] Heather Antecol, Ozkan Eren, and Serkan Ozbeklik. Peer Effects in Disadvantaged Primary Schools Evidence from a Randomized Experiment. *Journal of Human Resources*, 51(1):95–132, 2016.

[8] Marianne Bertrand, Erzo F. P. Luttmer, and Sendhil Mullainathan. Network Effects and Welfare Cultures. *The Quarterly Journal of Economics*, 115(3): 1019–1055, 2000.

[9] Julian R. Betts and Andrew Zau. Peer groups and academic achievement: Panel evidence from administrative data. 2004.

[10] Lawrence E. Blume, William A. Brock, Steven N. Durlauf, and Yannis M. Ioannides. Chapter 18 - Identification of Social Interactions§. In Alberto Bisin Jess Benhabib and Matthew O. Jackson, editors, *Handbook of Social Economics*, volume 1, pages 853–964. North-Holland, 2011.

[11] Lawrence E. Blume, William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman. Linear Social Interactions Models. *Journal of Political Economy*, 123(2):

444–496, 2015.

[12] Michael Boozer and Stephen E. Cacciola. Inside the 'Black Box' of Project Star: Estimation of Peer Effects Using Experimental Data. SSRN Scholarly Paper ID 277009, Social Science Research Network, Rochester, NY, 2001.

[13] Vincent Boucher, Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation. *Journal of Applied Econometrics*, 29(1):91–109, 2014.

[14] J Boyd-Zaharias, J Finn, R Fish, and S Gerber. Project STAR and Beyond: Database Users Guide. Technical report, Tech. rep, 2007.

[15] Yann Bramoullé, Habiba Djebbari, and Bernard Fortin. Identification of peer effects through social networks. *Journal of Econometrics*, 150(1):41–55, 2009.

[16] Mary A Burke and Tim R Sass. Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1):51–82, 2013.

[17] Jing Cai, Alain De Janvry, and Elisabeth Sadoulet. Social networks and the decision to insure. *American Economic Journal: Applied Economics*, 7(2):81–108, 2015.

[18] Scott E. Carrell, Frederick V. Malmstrom, and James E. West. Peer Effects in Academic Cheating. *Journal of Human Resources*, 43(1):173–207, 2008.

[19] Scott E. Carrell, Richard L. Fullerton, and James E. West. Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics*, 27(3):439–464, 2009.

[20] Scott E. Carrell, Bruce I. Sacerdote, and James E. West. From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica*, 81(3):855–882, 2013.

[21] Anne Case. Special Issue Space and Applied EconometricsNeighborhood influence and technological change. *Regional Science and Urban Economics*, 22(3): 491–508, 1992.

[22] Anne C. Case. Spatial Patterns in Household Demand. *Econometrica*, 59(4): 953–965, 1991.

[23] Anne C. Case, Harvey S. Rosen, and James R. Hines. Budget spillovers and fiscal policy interdependence. *Journal of Public Economics*, 52(3):285–307, 1993.

[24] Raj Chetty, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Jour-*

nal of Economics, 126(4):1593–1660, 2011.

[25] Matthew M. Chingos. The impact of a universal class-size reduction policy: Evidence from Florida's statewide mandate. *Economics of Education Review*, 31(5):543–562, 2012.

[26] Andrew David Cliff and J Keith Ord. *Spatial Autocorrelation*, volume 5. Pion London, 1973.

[27] Andrew David Cliff and J Keith Ord. *Spatial Processes: Models & Applications*, volume 44. Pion London, 1981.

[28] C.M. Achilles, Helen Pate Bain, Fred Bellott, Jayne Boyd-Zaharias, Jeremy Finn, John Folger, John Johnston, and Elizabeth Word. Tennessee's Student Teacher Achievement Ratio (STAR) project, 2008.

[29] Jeffrey P. Cohen and Catherine J. Morrison Paul. Public Infrastructure Investment, Interstate Spatial Spillovers, and Manufacturing Costs. *Review of Economics and Statistics*, 86(2):551–560, 2004.

[30] Timothy Conley, Nirav Mehta, Ralph Stinebrickner, and Todd Stinebrickner. Social Interactions, Mechanisms, and Equilibrium: Evidence from a Model of Study Time and Academic Achievement. Working Paper 21418, National Bureau of Economic Research, 2015.

[31] Timothy G. Conley and Bill Dupor. A Spatial Analysis of Sectoral Complementarity. *Journal of Political Economy*, 111(2):311–352, 2003.

[32] Giacomo De Giorgi, Michele Pellizzari, and Silvia Redaelli. Identification of Social Interactions through Partially Overlapping Peer Groups. *American Economic Journal: Applied Economics*, 2(2):241–275, 2010.

[33] Phoebus J. Dhrymes. *Introductory Econometrics*. Springer New York, New York, NY, 1978.

[34] D. M. Drukker, I. R. Prucha, and R. Raciborski. Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. *Stata Journal*, 13(2):221–241, 2013.

[35] Pascaline Dupas. Short-Run Subsidies and Long-Run Adoption of New Health Products: Evidence From a Field Experiment. *Econometrica*, 82(1):197–228, 2014.

[36] Steven N. Durlauf and Hisatoshi Tanaka. Understanding Regression Versus Variance Tests for Social Interactions. *Economic Inquiry*, 46(1):25–28, 2008.

[37] Daniel Eisenberg, Ezra Golberstein, and Janis L. Whitlock. Peer effects on risky behaviors: New evidence from college roommate assignments. *Journal of Health Economics*, 33:126–138, 2014.

[38] Dennis Epple and Richard Romano. Peer effects in education: A survey of the theory and evidence. *Handbook of social economics*, 1(11):1053–1163, 2011.

[39] Cem Ertur and Wilfried Koch. Growth, technological interdependence and spatial externalities: Theory and evidence. *Journal of Applied Econometrics*, 22(6):1033–1062, 2007.

[40] Jeremy Finn, Susan Gerber, Charles Achilles, and Jayne Boyd-Zaharias. The enduring effects of small classes. *The Teachers College Record*, 103(2):145–183, 2001.

[41] Jason M. Fletcher. Peer influences on adolescent alcohol consumption: Evidence using an instrumental variables/fixed effect approach. *Journal of Population Economics*, 25(4):1265–1286, 2011.

[42] John Folger. Project star and class size policy. *Peabody Journal of Education*, 67(1):1–16, 1989.

[43] Gigi Foster. It's not your peers, and it's not your friends: Some progress toward understanding the educational peer effect mechanism. *Journal of Public Economics*, 90(8–9):1455–1475, 2006.

[44] Jane C. Fruehwirth. Can Achievement Peer Effect Estimates Inform Policy? A View from Inside the Black Box. *Review of Economics and Statistics*, 96(3): 514–523, 2013.

[45] Alejandro Gaviria and Steven Raphael. School-Based Peer Effects and Juvenile Behavior. *Review of Economics and Statistics*, 83(2):257–268, 2001.

[46] Edward L. Glaeser, Bruce Sacerdote, and José A. Scheinkman. Crime and Social Interactions. *The Quarterly Journal of Economics*, 111(2):507–548, 1996.

[47] Bryan S. Graham. Identifying Social Interactions Through Conditional Variance Restrictions. *Econometrica*, 76(3):643–660, 2008.

[48] Eric A. Hanushek. The Failure of Input-based Schooling Policies*. *The Economic Journal*, 113(485):F64–F98, 2003.

[49] James J. Heckman. Skill Formation and the Economics of Investing in Disadvantaged Children. *Science*, 312(5782):1900–1902, 2006.

[50] Caroline M Hoxby and Gretchen Weingarth. Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, 2005.

[51] Guido W. Imbens and Jeffrey M. Wooldridge. Recent Developments in the Econometrics of Program Evaluation. *Journal of Economic Literature*, 47(1): 5–86, 2009.

[52] Scott A. Imberman, Adriana D. Kugler, and Bruce I. Sacerdote. Katrina's Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *The American Economic Review*, 102(5):2048–2082, 2012.

[53] Christopher Jepsen and Steven Rivkin. Class Size Reduction and Student Achievement The Potential Tradeoff between Teacher Quality and Class Size. *Journal of Human Resources*, 44(1):223–250, 2009.

[54] Mudit Kapoor, Harry H. Kelejian, and Ingmar R. Prucha. Panel data models with spatially correlated error components. *Journal of Econometrics*, 140(1): 97–130, 2007.

[55] Harry H. Kelejian and Ingmar R. Prucha. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121, 1998.

[56] Harry H. Kelejian and Ingmar R. Prucha. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, 40(2):509–533, 1999.

[57] Harry H. Kelejian and Ingmar R. Prucha. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*, 104(2): 219–257, 2001.

[58] Harry H. Kelejian and Ingmar R. Prucha. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics*, 32(6):691–707, 2002.

[59] Harry H. Kelejian and Ingmar R. Prucha. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, 157(1):53–67, 2010.

[60] Harry H. Kelejian, Ingmar R. Prucha, and Yevgeny Yuzefovich. Estimation Problems in Models with Spatial Weighting Matrices Which Have Blocks of Equal Elements*. *Journal of Regional Science*, 46(3):507–515, 2006.

[61] Alan B. Krueger. Experimental Estimates of Education Production Functions. *The Quarterly Journal of Economics*, 114(2):497–532, 1999.

[62] Guido M Kuersteiner and Ingmar R Prucha. Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity. 2015.

[63] Victor Lavy and Analía Schlosser. Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2):1–33, 2011.

[64] Lung-Fei Lee. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics*, 140 (2):333–374, 2007.

[65] Lung-fei Lee, Xiaodong Liu, and Xu Lin. Specification and estimation of social interaction models with network structures. *Econometrics Journal*, 13(2):145–176, 2010.

[66] Lars Lefgren. Educational peer effects and the Chicago public schools. *Journal of Urban Economics*, 56(2):169–191, 2004.

[67] James LeSage. An Introduction to Spatial Econometrics. *Revue d'économie industrielle*, n° 123(3):19–44, 2008.

[68] James P LeSage and R Kelley Pace. *Introduction to Spatial Econometrics (Statistics, Textbooks and Monographs)*. CRC Press, 2009.

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

[70] Xiaodong Liu and Lung-fei Lee. GMM estimation of social interaction models with centrality. *Journal of Econometrics*, 159(1):99–115, 2010.

[71] Xiaodong Liu, Eleonora Patacchini, and Yves Zenou. Endogenous peer effects: Local aggregate or local average? *Journal of Economic Behavior & Organization*, 103:39–59, 2014.

[72] David S Lyle. Estimating and Interpreting Peer and Role Model Effects from Randomly Assigned Social Groups at West Point. *Review of Economics and Statistics*, 89(2):289–299, 2007.

[73] Charles F. Manski. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3):531–542, 1993.

[74] David Marmaros and Bruce Sacerdote. How Do Friendships Form? *The Quarterly Journal of Economics*, 121(1):79–119, 2006.

[75] Alexandre Mas and Enrico Moretti. Peers at Work. *The American Economic Review*, 99(1):112–145, 2009.

[76] Robert A Moffitt. Policy interventions, low-level equilibria, and social interactions. *Social dynamics*, 4(45-82):6–17, 2001.

[77] Frederick Mosteller. The Tennessee Study of Class Size in the Early School Grades. *The Future of Children*, 5(2):113–127, 1995.

[78] Whitney K. Newey. Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica*, 59(4):1161–1167, 1991.

[79] Barbara Nye, Spyros Konstantopoulos, and Larry V. Hedges. How Large Are Teacher Effects? *Educational Evaluation and Policy Analysis*, 26(3):237–257, 2004.

[80] Keith Ord. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association*, 70(349):120–126, 1975.

[81] Olivier Parent and James P. LeSage. Using the variance structure of the conditional autoregressive spatial specification to model knowledge spillovers. *Journal of Applied Econometrics*, 23(2):235–256, 2008.

[82] Benedikt M. Pötscher and Ingmar R. Prucha. Basic structure of the asymptotic theory in dynamic nonlineaerco nometric models, part i: Consistency and approximation concepts. *Econometric Reviews*, 10(2):125–216, 1991.

[83] Steven G. Rivkin, Eric A. Hanushek, and John F. Kain. Teachers, Schools, and Academic Achievement. *Econometrica*, 73(2):417–458, 2005.

[84] Kenneth H. Rubin, William M. Bukowski, and Jeffrey G. Parker. Peer Interactions, Relationships, and Groups. In *Handbook of Child Psychology*. John Wiley & Sons, Inc., 2007.

[85] Bruce Sacerdote. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, 116(2):681–704, 2001.

[86] Bruce Sacerdote. Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In Stephen Machin Eric A. Hanushek and Ludger Woessmann, editors, *Handbook of the Economics of Education*, volume Volume 3, pages 249–277. Elsevier, 2011.

[87] Diane Whitmore Schanzenbach. What Have Researchers Learned from Project STAR? *Brookings Papers on Education Policy*, (9):205–228, 2006.

[88] Sarah Smith, Frank Windmeijer, and Edmund Wright. Peer Effects in Charitable Giving: Evidence from the (Running) Field. *The Economic Journal*, 125 (585):1053–1071, 2015.

[89] Kitae Sohn. A review of research on Project STAR and path ahead. *School Effectiveness and School Improvement*, pages 1–19, 2014.

[90] Aaron Sojourner. Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR. *The Economic Journal*, 123(569):574–605, 2013.

[91] Fabian Waldinger. Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. *The Review of Economic Studies*, 79(2):838–861, 2012.

[92] Diane Whitmore. Resource and Peer Impacts on Girls' Academic Achievement: Evidence from a Randomized Experiment. *The American Economic Review*, 95(2):199–203, 2005.

[93] David J. Zimmerman. Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics*, 85(1):9–23, 2003.