

# Efficient Peer Effects Estimators with Group Effects

Guido M. Kuersteiner, Ingmar R. Prucha, and Ying Zeng\*

April 15, 2022

## Abstract

We study linear peer effects models where peers interact in groups and individual's outcomes are linear in the group mean outcome and characteristics. We allow for unobserved random group effects as well as observed fixed group effects. The specification is in part motivated by the moment conditions imposed in Graham (2008). We show that these moment conditions can be cast in terms of a linear random group effects model and that they lead to a class of GMM estimators with parameters generally identified as long as there is sufficient variation in group size or group types. We also show that our class of GMM estimators contains a Quasi Maximum Likelihood estimator (QMLE) for the random group effects model, as well as the Wald estimator of Graham (2008) and the within estimator of Lee (2007) as special cases. Our identification results extend insights in Graham (2008) that show how assumptions about random group effects, variation in group size and certain forms of heteroscedasticity can be used to overcome the reflection problem in identifying peer effects. Our QMLE and GMM estimators accommodate additional covariates and are valid in situations with a large but finite number of different group sizes or types. Because our estimators are general moment based procedures, using instruments other than binary group indicators in estimation is straight forward. Our QMLE estimator accommodates group level covariates in the spirit of Mundlak and Chamberlain and offers an alternative to fixed effects specifications. This model feature significantly extends the applicability of Graham's identification strategy to situations where group assignment may not be random but correlation of group level effects with peer effects can be controlled for with observable group level characteristics. Monte-Carlo simulations show that the bias of the QMLE estimator decreases with the number of groups and the variation in group size, and increases with group size. We also prove the consistency and asymptotic normality of the estimator under reasonable assumptions.

---

\*Kuersteiner: Department of Economics, University of Maryland, College Park, MD 20742, United States (gkuerste@umd.edu); Prucha: Department of Economics, University of Maryland, College Park, MD 20742, United States (prucha@umd.edu); Zeng: Department of Public Finance, School of Economics, Xiamen University, Xiamen 361005, China (zengying17@xmu.edu.cn).

# 1 Introduction

Peer effects are of great interest to empirical researchers and policy makers. The idea that individuals are affected by their peers motivates policies that try to manipulate peer composition for better outcomes. Peer effects are often confounded by group level effects. An example are teacher effects in a class room setting. Identifying peer effects is notoriously challenging due to the reflection problem (Manski, 1993; Angrist, 2014) as well as due to spurious peer effects originating from group level effects. Random group allocation may be one way to overcome these identification problems. With groups formed at random, a random effects specification for group level characteristics can be adopted. An alternative approach consists in postulating that, conditional on observed group level characteristics, group level effects can be viewed as randomly assigned. Regression control techniques based on observed group characteristics then lead to a similar random effects specification, but without the need to appeal to random group assignment. We propose estimators that can accommodate both scenarios.

Random group assignment plays a prominent role in the empirical peer effects literature in a number of fields including education, labor, firm, finance and development studies. Recent examples from this literature include Sacerdote (2001); Duflo and Saez (2003); Zimmerman (2003); Stinebrickner and Stinebrickner (2006); Kang (2007); Graham (2008); Guryan et al. (2009); Carrell et al. (2009, 2013); Duflo et al. (2011); Sojourner (2013); Booij et al. (2017); Garlick (2018); Fafchamps and Quinn (2018); Cai and Szeidl (2018); Frijters et al. (2019). Assuming group effects to be independent of observed individual and group characteristics is plausible when groups are formed at random. Ignoring group effects or assuming fixed group effects (Lee, 2007) leads to consistent but less efficient estimators. Random group effects themselves have important empirical interpretations. For example, researchers in education policy often treat random class effects as unobserved teacher effects (e.g., Nye et al. 2004; Rivkin et al. 2005; Chetty et al. 2011). Absent random group assignment, the estimators we propose can accommodate observed group level effects that can come from information about group characteristics such as the training and experience of teachers, or averages of individual group member characteristics. Group level characteristics can be interpreted as parametrizations of group effects in the spirit of Mundlak (1978) and Chamberlain (1980). The choice between a random effects or fixed effects estimator then depends less on random group assignment but more on whether group specific effects are believed to be observable or not. In some cases there may be independent interest in the effects of group specific covariates. An example is the effect of teacher training on student performance. In such cases a random effects estimator is the preferred choice because fixed effects estimators are often unable to identify these types of group level effects.

We use the moment restrictions proposed by Graham (2008) as the starting point for our analysis. Random effects assumptions for the group level effects combined with assumptions of cross-sectional independence for idiosyncratic errors are the basis for a set of moment conditions we impose. We give an interpretation of the conditional variance estimator (CVE) of Graham (2008) in terms of a GMM estimator based on moment conditions for the within-group variance

and between-group variance. We also show that the moment conditions underlying Graham (2008) are the score function of a quasi maximum likelihood estimator (QMLE) for a random group effects model. The QMLE can be shown to be the best GMM estimator in the class of estimators using the moment conditions utilized by Graham (2008). One limitation of the conditional variance estimator proposed by Graham is the fact that it amounts to a difference in difference identification strategy for the variances that requires groups to fall into two size categories. As shown in Graham (2008) the resulting procedure takes the form of a Wald estimator for a set of binary instruments. This setting is restrictive in applications where groups may not be easily separated into two categories or where a more general set of instruments needs to be considered. The estimators that we propose are general GMM based procedures that accommodate additional covariates as well as offer flexibility in terms of the instruments and the number of moment conditions that are being used. We illustrate these points by explicitly considering moment based estimators that exploit exogenous variation in group size as well as general group level heteroscedasticity, as opposed to a binary group size indicator, as instruments. This leads us to study a general linear random group effects model estimated using QML.

The QMLE we develop in this paper uses moment conditions for the within and between variances of outcomes individually, rather than combining them into a single moment condition as is the case for the CV estimator. This leads to a more flexible procedure that is able to identify the endogenous peer effects parameter in a broader class of settings. Our setup also facilitates the inclusion of additional covariates in a unified joint estimation framework which is important for statistical inference. In contrast to the CMLE of Lee (2007) our procedure is based on both the within and between variance. This leads to efficiency gains under correct specification but comes at the cost of potential miss-specification bias if the random group effect assumption is incorrect. The trade-offs are similar to related results for fixed and random effects in the panel literature.

Our work is also related to the literature in spatial econometrics started by the work of Cliff and Ord (1973, 1981) and Anselin (1988).<sup>1</sup> 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 (2020). The strength of social links can be characterized by proximity in the social network space. We extend Kelejian et al. (2006) and Lee (2007) by considering a random group effects specification. Spatial models were traditionally estimated with maximum likelihood (ML), e.g., Ord (1975). Kelejian and Prucha (1998, 1999) develop generalized method of moments (GMM) estimators based on linear and quadratic moments. While this paper utilizes a quasi-maximum likelihood estimation method, the score function depends on linear quadratic forms of the error terms. 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 (2020) in a panel setting. Moreover, Kelejian and Prucha (2001) and Kelejian and Prucha (2010) develop a central limit theorem for linear quadratic forms, which is the basis for the asymptotic analysis in

---

<sup>1</sup>Anselin (2010) offers a brief review of the development of spatial econometrics literature over the past thirty years.

this paper.

The linear-in-means peer effect model in Manski (1993) is a special case of a spatial model with group-wise equal dependence, see Kelejian and Prucha (2002) and Kelejian et al. (2006). Kelejian and Prucha (2002) were 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), GMM and QMLE methods all yield inconsistent estimators, although consistent estimation with 2SLS and GMM is possible for panel data. 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 QMLE developed in this paper and the conditional maximum likelihood estimator in Lee (2007) both rely on group size variation for identification although we show that identification exploiting heteroscedastic errors is also possible. Extensions include Lee et al. (2010) who allow for specific social structure within each group and Liu and Lee (2010) and Liu et al. (2014) who allow for non-row normalized weight matrices. The linear spatial model 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 of social interaction models and give conditions for identification.

The paper is organized as follows. In Section 2 we consider identification of endogenous peer effects in a simple setting without covariates for the CV, CML and QML estimators. Section 3 presents the full model that allows for covariates and general variation in group size. Section 4 summarizes the technical conditions we impose and presents theoretical results for the QMLE. Section 5 contains a small Monte Carlo experiment. Proofs are collected in an appendix.

## 2 Peer Effects with Random Group Effects

We start the discussion by presenting a simple model without covariates, to introduce and discuss basic features of our new quasi-maximum likelihood estimator (QMLE), and connect it to the conditional variance (CV) estimator in Graham (2008) and the conditional maximum likelihood (CMLE) estimator in Lee (2007). The model decomposes variation in outcomes of a cross-section of individuals into idiosyncratic noise, group level random effects and correlation that is due to group level interaction.

Quadratic moment conditions implied by this random effects specification lead to efficient GMM, quasi maximum likelihood, and under additional distributional assumptions, maximum likelihood estimators. Estimators based on these moment conditions include the CV estimator of Graham (2008), the QMLE as well as the CMLE of Lee (2007) as special cases.

Let  $y_{ir}$  be an observed outcome of individual  $i$  in group  $r$  which has  $m_r$  members, let  $\alpha_r$  be an unobserved group level effect and let  $\epsilon_{ir}$  be unobserved individual specific characteristics. We observe data for  $R$  groups as well as a categorical variable  $D_r$  which determines group type. An example is when there are three group sizes such that  $D_r \in \{ 'small', 'medium', 'large' \}$ . However,

$D_r$  could be a characteristic that is not necessarily related to group size. An example is when groups are defined by classrooms of schools in urban, suburban or rural districts and  $D_r$  is used to denote urbanicity. Classes could also be categorized by sociodemographic composition such as whether English or other languages are the native language spoken by students in the class. We allow for type-dependent heteroscedasticity. Types add flexibility to the specification by relaxing the constraints the model imposes on the relationship between group variance and group size. In some cases type specific heteroscedasticity provides identifying variation that is separate from group size variation.

The peer effects model is stated in terms of a structural equation

$$y_{ir} = \lambda \bar{y}_{(-i)r} + \alpha_r + \epsilon_{ir}, \quad (1)$$

where  $\bar{y}_{(-i)r} = \frac{1}{m_r-1} \sum_{j \neq i}^{m_r} y_{jr}$  is the leave-out-mean of the outcome variable. The parameter  $\lambda$  captures the endogenous peer effects, see Manski (1993). The structural form emphasizes the decomposition of  $y_{ir}$  into a social interaction term  $\lambda \bar{y}_{(-i)r}$ , a group level effect  $\alpha_r$  and an idiosyncratic error term  $\epsilon_{ir}$ . For example, when  $y_{ir}$  is a measure of student performance and  $r$  is a class-room index then  $\alpha_r$  can be interpreted as a class-room or teacher effect while  $\epsilon_{ir}$  are unobserved student characteristics for student  $i$  in classroom  $r$ . Cross-sectional independence of  $\epsilon_{ir}$  can be justified by random group assignment such as in the application of Graham (2008). The assumptions we impose on  $\epsilon_{ir}$  and  $\alpha_r$  are in line with the random effects panel literature where group level dependence of unobservables is modeled with the common factor  $\alpha_r$ . We leave possible generalizations of this framework to cases where  $\epsilon_{ir}$  is allowed to be dependent for future work.

Following Graham (2008) who emphasizes random assignments of individuals to groups, we assume that  $\alpha_r$  is a random effect independent of  $\epsilon_{ir}$ . As shown by Graham (2008) for a slightly different model based on full rather than leave-out means, the random effects nature of the model leads to a set of quadratic moment conditions that can be exploited for identification. We expand on these ideas by showing that the implied moment conditions are related to the moment conditions of a random effects pseudo likelihood estimator. Transformations of these moments turn out to coincide with moments used by Graham (2008) as well as Lee (2007) who considers a fixed effects version of the model. Lee (2007) focuses on identification of  $\lambda$  based on group size variation. Here we emphasize a random effects specification where identification is driven by heterogeneity at the group level that could result from sources including but not limited to class size variation. A literature on linear instrumental variables methods gives conditions under which  $\lambda$  can be identified in models that have additional exogenous covariates  $Z_r$ , e.g., Angrist (2014) or Bramoullé et al. (2009).<sup>2</sup> Besides the conventional instrumental variables strategies, alternative strategies are also available, see Lee (2007), Graham (2008) for a modified model or Kuersteiner and Prucha (2020). Letting  $Y_r = (y_{1r}, \dots, y_{m_r r})'$ ,  $\epsilon_r = (\epsilon_{1r}, \dots, \epsilon_{m_r r})'$ ,  $\iota_{m_r} = (1, \dots, 1)'$  and  $W_{m_r} = \frac{1}{m_r-1}(\iota_{m_r} \iota_{m_r}' - I_{m_r})$ ,

---

<sup>2</sup>The leave-out-mean  $\bar{y}_{(-i)r}$  can be viewed as a special case of a Cliff-Ord-type (Cliff and Ord 1973, 1981) spatial lag. Kelejian and Prucha (1998) give an early basic condition for identification by IV.

the model can be written in matrix notation as

$$Y_r = \lambda W_{m_r} Y_r + \alpha_r \iota_{m_r} + \epsilon_r. \quad (2)$$

To isolate or identify the social interaction effect, we impose the following restrictions on unobservables.

**Assumption 1.** For  $r = 1, \dots, R$  the  $r$ -th group is associated with a categorical variable  $D_r \in \{1, 2, \dots, J\}$  with  $J \geq 1$  being fixed and finite, and for each category  $j \in \{1, 2, \dots, J\}$  there is at least one group  $r$  with  $D_r = j$ . For  $r = 1, \dots, R$  and  $i = 1, \dots, m_r$  the disturbance terms  $\epsilon_{ir}$  are independently distributed across all  $i$  and  $r$ , with  $E[\epsilon_{ir}|D_r, m_r] = 0$  and  $E[\epsilon_{ir}^2|D_r, m_r] = \sigma_{\epsilon 0, D_r}^2$ ,  $0 < \underline{a}_\epsilon \leq \sigma_{\epsilon 0, D_r}^2 \leq \bar{a}_\epsilon < \infty$  and where  $\sigma_{\epsilon 0, D_r}^2$  is a function only of  $D_r$ . There exists some  $\eta_\epsilon > 0$  such that  $E[|\epsilon_{ir}|^{4+\eta_\epsilon}] < \infty$ .

Note that the variance  $E[\epsilon_{ir}^2|D_r, m_r] = \sigma_{\epsilon 0, D_r}^2$  has the representation  $\sigma_{\epsilon 0, D_r}^2 = \sigma_{\epsilon 0, 1}^2 1\{D_r = 1\} + \dots + \sigma_{\epsilon 0, J}^2 1\{D_r = J\}$  where  $\sigma_{\epsilon 0, 1}^2, \dots, \sigma_{\epsilon 0, J}^2$  are fixed parameters to be estimated.

**Assumption 2.** For  $r = 1, \dots, R$ , the group effects  $\alpha_r$  are independently and identically distributed, with  $E[\alpha_r|D_r, m_r] = 0$  and  $E[\alpha_r^2|D_r, m_r] = \sigma_{\alpha 0}^2$ , where  $0 \leq \sigma_{\alpha 0}^2 \leq \bar{a}_\alpha < \infty$ . There exists some  $\eta_\alpha > 0$  such that  $E[|\alpha_r|^{4+\eta_\alpha}] < \infty$ . Also,  $\{\alpha_r : r = 1, \dots, R\}$  are independent of  $\{\epsilon_{ir} : i = 1, \dots, m_r; r = 1, \dots, R\}$ .

Assumption 1 implies in particular that individuals do not self select into groups based on unobserved characteristics and Assumption 2 suggests that there is no matching between group characteristics and individual characteristics. This no sorting or matching assumption can sometimes be motivated by specific empirical designs. For example, in the Project STAR experiment that Graham (2008) considers, kindergarten students and teachers are randomly assigned to classrooms. This random assignment mechanism justifies interpreting  $\alpha_r$  as the classroom or teacher effect. It also justifies assuming that  $\alpha_r$  and  $\epsilon_{ir}$  are mutually independent random variables, see Graham (2008) Assumption 1.1. Assumption 1 allows  $\epsilon_{ir}$  to be homoskedastic across all groups when  $J = 1$  or heteroscedastic across different categories of  $D_r$  when  $J \geq 2$ . This formulation contains the case considered by Graham (2008) where  $J = 2$  as a special case.

Assumptions 1 and 2 above imply moment conditions. These moment conditions take the form of restrictions on the within and between group variance. As discussed in more detail below, these moment conditions are fundamental to the ML estimator. In particular, we show that the score of the ML estimator is a weighted average of those fundamental moment conditions.

To derive the moment conditions, define the composite error term  $U_r = \alpha_r \iota_{m_r} + \epsilon_r$  where  $U_r$  is an  $m_r \times 1$  vector with elements  $u_{ir} = \alpha_r + \epsilon_{ir}$ . Let  $\bar{u}_r$  and  $\bar{\epsilon}_r$  be the mean of  $u_{ir}$  and  $\epsilon_{ir}$  in group  $r$ . Let  $\check{U}_r = U_r - \bar{u}_r \iota_{m_r}$  be the vector of within-group deviations from the mean of  $U_r$  and let  $\check{Y}_r$  and  $\check{\epsilon}_r$  be defined in a similar manner. It can be shown that  $\bar{y}_r = \bar{u}_r / (1 - \lambda) = (\alpha_r + \bar{\epsilon}_r) / (1 - \lambda)$  with  $\bar{u}_r = \alpha_r + \bar{\epsilon}_r$ , and  $\check{Y}_r = \frac{m_r - 1}{m_r - 1 + \lambda} \check{U}_r = \frac{m_r - 1}{m_r - 1 + \lambda} \check{\epsilon}_r$ . Two conditional moment conditions, one for the within-group variance, the other for the between group variance, arise for the model in (2) under

Assumptions 1 and 2. The expected value of the within-group and between-group squares of group  $r$  are

$$var_r^w = E \left[ \frac{\ddot{Y}_r' \ddot{Y}_r}{m_r - 1} | m_r, D_r \right] = E \left[ \frac{(m_r - 1) \ddot{U}_r' \ddot{U}_r}{(m_r - 1 + \lambda)^2} | m_r, D_r \right] = \frac{(m_r - 1)^2}{(m_r - 1 + \lambda)^2} \sigma_{\epsilon, D_r}^2, \quad (3)$$

$$var_r^b = E \left[ \bar{y}_r^2 | m_r, D_r \right] = E \left[ \frac{\bar{u}_r^2}{(1 - \lambda)^2} | m_r, D_r \right] = \frac{1}{(1 - \lambda)^2} \left( \sigma_\alpha^2 + \frac{\sigma_{\epsilon, D_r}^2}{m_r} \right), \quad (4)$$

where  $\sigma_{\epsilon, D_r}^2 = \sigma_{\epsilon, 1}^2 1 \{D_r = 1\} + \dots + \sigma_{\epsilon, J}^2 1 \{D_r = J\}$ .

To see how these moment conditions can achieve the identification of  $\lambda$  consider the case where  $\sigma_{\epsilon, D_r}^2 = \sigma_{\epsilon, D_s}^2$  but  $m_r \neq m_s$ . Then, Equation (3) implies that

$$\left( \frac{m_r - 1 + \lambda}{m_s - 1 + \lambda} \right)^2 = \frac{(m_r - 1)^3 E \left[ \ddot{Y}_s' \ddot{Y}_s | m_s, D_s \right]}{(m_s - 1)^3 E \left[ \ddot{Y}_r' \ddot{Y}_r | m_r, D_r \right]}. \quad (5)$$

Alternatively consider the case where  $m_r = m_s = m$  and  $\sigma_{\epsilon, D_r}^2 \neq \sigma_{\epsilon, D_s}^2$ , then combining (3) and (4) gives

$$\frac{(m - 1 + \lambda)^2}{(1 - \lambda)^2} = \frac{E \left[ \bar{y}_r^2 | m, D_r \right] - E \left[ \bar{y}_s^2 | m, D_s \right]}{E \left[ \ddot{Y}_r' \ddot{Y}_r / [m(m - 1)^3] | m, D_r \right] - E \left[ \ddot{Y}_s' \ddot{Y}_s / [m(m - 1)^3] | m, D_s \right]}. \quad (6)$$

Expressions on the left hand side of both (5) and (6) in principle can be solved for  $\lambda$  if we restrict  $\lambda \in (-1, 1)$  and  $m_r \geq 2$ , as both expressions are monotonic functions of  $\lambda$ . Equation (6) is a modified version of Equation (9) in Graham (2008) that accounts for the leave-out-mean specification we consider. The numerator differences out the variance of  $\alpha_r$  which is assumed constant across types. This restriction is also imposed by Graham (2008) in his Assumption 1.2. In Lemma 2.1 below we outline the exact conditions under which identification is possible.

The discussion above shows that under additional assumptions on  $\lambda$  and group size, identification of  $\lambda$  is possible through moment conditions related to within and between variance when there is variation in either group size  $m_r$  or idiosyncratic error variance  $\sigma_{\epsilon, D_r}^2$ . We now formalize the discussion into Lemma 2.1 below. Let the parameter vector be  $\theta = \left( \lambda, \sigma_\alpha^2 \sigma_{\epsilon, 1}^2, \dots, \sigma_{\epsilon, J}^2 \right)'$  and, for clarity, let the true parameter vector be denoted by  $\theta_0 = \left( \lambda_0, \sigma_{\alpha 0}^2, \sigma_{\epsilon 0, 1}^2, \dots, \sigma_{\epsilon 0, J}^2 \right)'$ . For identification, we further assume that group size  $m_r \geq 2$  and impose the following assumption on  $\lambda$ .

**Assumption 3.** *The parameter of the endogenous peer effects  $\lambda_0 \in \Lambda$ , where  $\Lambda$  is a compact subset of  $(-1, 1)$ . Assume that  $\theta_0 \in \Theta$  with  $\Theta = \Lambda \times [0, \bar{a}_\alpha] \times [\underline{a}_\epsilon, \bar{a}_\epsilon] \times \dots \times [\underline{a}_\epsilon, \bar{a}_\epsilon]$  compact.*

The estimation procedures we propose in this paper can be implemented with the availability of a general set of valid instruments and are valid for cases where  $J \geq 1$  as long as  $J$  is fixed and finite. In the simple model without covariates the available instruments are group size  $m_r$  and categorical variable  $D_r$ . These instruments are valid if assignment to groups is random in a way that generates random variation in group size or category. Utilizing Equation (3) and (4), and using group size  $m_r$  and the categorical variable  $D_r$  as instruments yields the following conditional

moment restriction  $E[\chi_r(\theta_0)|m_r, D_r] = 0$  with

$$\chi_r(\theta) = \begin{bmatrix} \chi_r^w(\theta) \\ \chi_r^b(\theta) \end{bmatrix} = \begin{bmatrix} \frac{(m_r-1+\lambda)^2 \ddot{Y}_r' \ddot{Y}_r}{(m_r-1)^2} - (m_r-1)\sigma_{\epsilon, D_r}^2 \\ (1-\lambda)^2 \bar{y}_r^2 - \sigma_\alpha^2 - \frac{\sigma_{\epsilon, D_r}^2}{m_r} \end{bmatrix}. \quad (7)$$

Identification of the parameter  $\theta$  is possible with variation in group size for a given category or variation in the idiosyncratic variance over categories for the same group size. This is summarized in the following lemma. The proof of the lemma is given in Appendix C.

**Lemma 2.1.** *Suppose Assumptions 1-3 hold. Then the parameter  $\theta_0$  is identified under the following two scenarios:*

(i) *There are two groups  $r$  and  $s$  such that  $m_r \neq m_s$  and  $D_r = D_s$ , and therefore  $\sigma_{\epsilon_0, D_r}^2 = \sigma_{\epsilon_0, D_s}^2$ . Then the parameter  $\theta_0$  is identified in  $\Theta$ . In particular, the moment conditions  $E[\chi_q^w(\theta)|m_q, D_q] = 0$  and  $E[\chi_q^b(\theta)|m_q, D_q] = 0$  for  $q = r, s$  with  $\chi_r^w(\theta)$  and  $\chi_r^b(\theta)$  defined in (7) identify  $\lambda_0$ ,  $\sigma_{\epsilon_0, D_r}^2$ , and  $\sigma_{\alpha_0}^2$ . The remaining parameters  $\sigma_{\epsilon_0, j}^2$  are identified by  $E(\chi_q^w(\theta)|m_q, D_q) = 0$  for  $q \neq r$  or  $s$ .*

(ii) *There are two groups  $r$  and  $s$ , such that  $m_r = m_s$  and  $\sigma_{\epsilon_0, D_r}^2 \neq \sigma_{\epsilon_0, D_s}^2$ . Then the parameter  $\theta_0$  is identified in  $\Theta$ . In particular, the moment condition  $E[\nu_q(\theta)|m_q, D_q] = 0$ ,  $q = r, s$  uniquely identifies  $\lambda_0$  and  $\sigma_{\alpha_0}^2$ , where*

$$\nu_q(\theta) = \chi_q^b(\theta) - \frac{\chi_q^w(\theta)}{m_q(m_q-1)} = (1-\lambda)^2 \bar{y}_q^2 - \sigma_\alpha^2 - \frac{(m_q-1+\lambda)^2 \ddot{Y}_q' \ddot{Y}_q}{m_q(m_q-1)^3} \quad (8)$$

with  $\chi_q^w(\theta)$  and  $\chi_q^b(\theta)$  defined in (7). The remaining parameters  $\sigma_{\epsilon_0, j}^2$  are identified by  $E(\chi_q^w(\theta)|m_q, D_q) = 0$ .

Full identification is achieved in Scenario (i) with group size variation in at least one category. As an example, consider types that describe urbanicity such that  $D_r = D_s = 1$  denotes two classrooms  $r$  and  $s$  that are both located in an urban school but where  $m_r \neq m_s$  such that the classrooms differ in size, while the remaining categories  $d = 2, \dots, J$  may have the same group sizes. In this setting  $\theta_0$  is identified without any further constraints on the variances  $\sigma_{\epsilon, j}^2$ . If the number of distinct group sizes exceeds the number of categories  $J$  then it automatically must be the case that there exist some category that is associated with at least two distinct group sizes. Note that the result holds irrespective of whether the constraint of homoscedastic errors  $\sigma_{\epsilon_0, D_r}^2 = \sigma_{\epsilon_0, D_s}^2$  is imposed on the model or not. From Scenario (i) we see that variation in group size alone can provide variation that is sufficient for identification. Furthermore, in the homoscedastic case where only a common variance parameter  $\sigma_\epsilon^2$  is specified, two distinct group sizes are sufficient for identification by the result in Scenario (i). This corresponds to the identification result of the conditional maximum likelihood estimator (CMLE) in Lee (2007), the score function of which can be written as  $\varphi(m_r)\chi_r^w(\theta)$ , where  $\varphi(m_r)$  is a function of  $m_r$ .

While variation in group size serves as the source of identification in Scenario (i), identification based on the moment condition  $E[\chi_r(\theta)|m_r, D_r] = 0$  is also possible without group size variation as long as there is some other form of group heterogeneity. As is shown in the proof for Scenario (ii) of



Lemma 2.1, utilizing  $m_q = m$  and  $E[\nu_q(\theta)|m_q, D_q] = 0$  for  $q = r, s$  yields (6). From (6) we see that the endogenous peer effect parameter  $\lambda$  is identified if there is heteroscedasticity across groups of the same size for at least one size, and that  $\lambda$  can be estimated from the sample analog of (6). The intuition of identification in Scenario (ii) echoes that of the conditional variance (CV) estimator of Graham (2008). Similar to Graham, (8) is based on the relationship between the within-group and between-group variance as captured by  $\nu_r(\theta)$ , and can be used to construct a Wald type moment condition like in (6) using the categorical variable as the instrument.

The above discussion focused on identification based on the moment vector  $\chi_r(\theta)$ . We next discuss the importance of these moment conditions for efficient estimation, and their relationship to the score of the Gaussian ML estimator. The optimal moment function corresponding to  $\chi_r(\theta)$  is given by  $\chi_r^*(\theta) = \varphi^*(m_r, D_r)\chi_r(\theta)$  where, focusing on the case with  $J = 2$  for exposition,<sup>3</sup>

$$\begin{aligned} \varphi^*(m_r, D_r) &= E\left[\frac{\partial}{\partial\theta'}\chi_r(\theta_0)|m_r, D_r\right] \left(E[\chi_r(\theta_0)\chi_r(\theta_0)'|m_r, D_r]\right)^{-1} \\ &= \begin{pmatrix} \frac{1}{(m_r-1+\lambda)\sigma_{\epsilon_0, D_r}^2} & -\frac{m_r}{(1-\lambda)(\sigma_{\epsilon_0, D_r}^2 + m_r\sigma_{\alpha_0}^2)} \\ 0 & -\frac{m_r^2}{2(\sigma_{\epsilon_0, D_r}^2 + m_r\sigma_{\alpha_0}^2)^2} \\ -\frac{1\{D_r=1\}}{2\sigma_{\epsilon_0, 1}^4} & -\frac{1\{D_r=1\}m_r}{2(\sigma_{\epsilon_0, 1}^2 + m_r\sigma_{\alpha_0}^2)^2} \\ -\frac{1\{D_r=2\}}{2\sigma_{\epsilon_0, 2}^4} & -\frac{1\{D_r=2\}m_r}{2(\sigma_{\epsilon_0, 2}^2 + m_r\sigma_{\alpha_0}^2)^2} \end{pmatrix}. \end{aligned} \quad (9)$$

Clearly, it follows that  $E[\chi_r^*(\theta_0)] = 0$  by iterated expectations. We note that the moment condition in (6) underlying the CV estimator is based on a linear transformation of  $\varphi^*(m_r, D_r)$ . Furthermore, as we shall see in the next section, under the additional assumption that  $\alpha$  and  $\epsilon$  follow a Gaussian distribution, the score function of the log likelihood for group  $r$  is exactly the negative of  $\chi_r^*(\theta)$ , that is

$$\partial \ln L_r(\theta_0)/\partial\theta = -\chi_r^*(\theta_0),$$

where  $\ln L_r(\theta)$  denotes the log likelihood function for group  $r$  conditional on  $(m_1, \dots, m_R, D_1, \dots, D_R)$ . From these observations we see that the matrices  $\varphi^*(m_r, D_r)$  can be viewed to provide the optimal weighting for the basic moment functions  $\chi_r(\delta)$ ; compare also the corresponding discussion for the general model for more details.

The result that  $\partial \ln L_r(\theta_0)/\partial\theta = -\chi_r^*(\theta_0)$  for the score function under Gaussianity establishes the asymptotic efficiency of the GMM estimator based on  $E[\chi_r(\theta)|m_r, D_r] = 0$  under the assumption of Gaussian distributions for the unobservables. When the unobservables are not Gaussian then the GMM estimator has the interpretation of a quasi maximum likelihood estimator (QMLE). Similarly, in Lee (2007) the score function of the conditional maximum likelihood estimator (CMLE) for group  $r$  is the optimal moment function corresponding to  $E[\chi_r^w(\theta)|m_r] = 0$  under the assumption of homoscedastic and normally distributed errors  $\epsilon_{ir}$ . While the CMLE of Lee (2007) is not

<sup>3</sup>See our Online Appendix for details. The derivation uses Lemma B.1 and the special properties of matrices  $\Omega(\theta)$ ,  $I - \lambda W$  and  $W$  described in Appendix B.1. In the Online Appendix we also give an explicit expression for the variance covariance matrix of  $\chi_r(\delta)$ .

efficient under the assumptions we postulate in this paper, it shares robustness properties of within group panel estimators in cases where the group effects are possibly correlated with covariates in the model. Under those circumstances, random effects quasi maximum likelihood estimators are generally not expected to be consistent.

Our discussion so far highlights variance as the source of identification, with variation in either size  $m_r$  or variance of the idiosyncratic error terms  $\sigma_{\epsilon, D_r}^2$  across groups as conditions. We show that variation in group size and error term variance is a source of identification in the QMLE, CVE and CMLE. In all, the CMLE utilizes how within-group variance changes with  $\lambda$  and size when error terms are homoscedastic, while the CVE exploits the relationship between the within-group variance and between-group variance in relation to  $\lambda$  and size when there is either variation in group size or heteroscedasticity across groups. Our QMLE uses both pieces of information. All three estimators remain valid without covariates, and may achieve identification as long as there are at least two different group sizes in the limit in the case of homoscedasticity. This complements other results in the literature. For example, Proposition 4 in Bramoullé et al. (2009) states that in the setting of Lee (2007),  $\lambda$  is identified by instrumenting  $(I - W)WY$  with  $(I - W)W^2Z$ ,  $(I - W)W^3Z$ , etc., in line with the spatial literature on the estimation of Cliff-Ord type models. Their result is due to the fact that they only exploit restrictions for the conditional mean of  $\epsilon$ . In Graham (2008) as well as in this paper additional constraints on the distribution of  $\alpha$  and  $\epsilon$  are imposed and shown to be useful in the identification of peer effects. Under these conditions including  $Z$  offers additional sources of variation, but identification is possible with or without it.

Adding covariates is critically important in empirical applications. Consider adding the covariate matrix  $Z$ . This leads to two additional moment conditions  $E[\ddot{Z}'_r \ddot{U}_r | m_r, D_r] = 0$  and  $E[\bar{z}'_r \bar{u}_r | m_r, D_r] = 0$ , where  $\bar{z}_r = \iota'_{m_r} Z_r / m_r$  is the group mean of  $Z_r$  and  $\ddot{Z}_r = Z_r - \iota_{m_r} \bar{z}_r$  is the deviation from group mean. Moreover,  $\check{Y}_r$  and  $\bar{y}_r$  now need to be replaced by  $\check{Y}_r - \frac{m_r - 1}{m_r - 1 + \lambda} \ddot{Z}_r \beta$  and  $\bar{y}_r - \frac{\bar{z}_r \beta}{1 - \lambda}$  respectively. The score function of the QMLE then is the same as the moment conditions of the best GMM corresponding to these two moment functions in addition to the moments  $E[\chi_r(\theta) | m_r, D_r] = 0$ . In the same way, in the presence of covariates and assuming homoscedasticity of  $\epsilon$ , Lee's CMLE estimator is based on  $E[\ddot{Z}'_r \ddot{U}_r] = 0$  in addition to  $E[\chi_r^w(\theta) | m_r] = 0$  and the relative efficiency considerations discussed in this section continue to apply to the situation with covariates.

### 3 General Model

In this section we generalize the model to allow for individual characteristics, average individual characteristics of peers and group level covariates. We assume that we have access to observations on  $R$  groups belonging to  $J$  categories, where  $1 \leq J < \infty$  is fixed. We consider asymptotics where the number of groups  $R$  tends to infinity and where the number of group sizes is finite. For the asymptotic identification of  $\lambda_0$  and  $\sigma_{\alpha_0}^2$  this setup assumes that in the limit we observe infinitely many groups for at least two group sizes or two categories, echoing the requirement of

variation in either group sizes or categories for identification in Section 2. In designs that allow for heteroscedasticity, we also need infinitely many groups for each category  $j \in \{1, \dots, J\}$  to identify the remaining variance parameters  $\sigma_{\epsilon,j}^2$ . Let  $r = 1, \dots, R$  denote the group index, let  $D_r$  denote the category of group  $r$ , and let  $m_r$  denote the size of group  $r$ . The total sample size is then given by  $N = \sum_{r=1}^R m_r$ . Suppose further that interactions occur within each group, but not across groups, and that peer effects work through the mean outcome and mean characteristics of peers in the same group. The linear-in-means peer effects model that includes endogenous as well as exogenous peer effects then is given by

$$y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir} \beta_2 + \bar{x}_{2,(-i)r} \beta_3 + x_{3,r} \beta_4 + \alpha_r + \epsilon_{ir}, \quad (10)$$

where  $y_{ir}$  is the outcome variable of individual  $i$  in group  $r$ ,  $\bar{y}_{(-i)r} = \frac{1}{m_r-1} \sum_{j \neq i}^{m_r} y_{jr}$  is the average outcome of  $i$ 's peers,  $x_{1,ir}$  and  $x_{2,ir}$  are both row vectors of predetermined characteristics of individual  $i$  in group  $r$ ,  $\bar{x}_{2,(-i)r} = \frac{1}{m_r-1} \sum_{j \neq i}^{m_r} x_{2,jr}$  is a vector of average characteristics of  $i$ 's peers,  $x_{3,r}$  is a vector of observed group characteristics. The variables in  $x_{1,ir}$  and  $x_{2,ir}$  can be non-overlapping, partially overlapping or totally overlapping. The error term consists of two components, the group effect  $\alpha_r$  and the disturbance term  $\epsilon_{ir}$ . We treat  $x_{1,ir}$ ,  $x_{2,ir}$ ,  $x_{3,r}$ ,  $D_r$  and  $m_r$  as non-stochastic, while noting that at the expense of more complex notation we could also think of the analysis as being conditional on these variables. In this model, peer effects work through the mean peer outcome  $\bar{y}_{(-i)r}$  and mean peer characteristics  $\bar{x}_{2,(-i)r}$ . The two terms are also known as the leave-out-mean of  $y$  and  $x_2$ , as they are means of the group leaving out oneself. In Manski's terminology,  $\lambda \bar{y}_{(-i)r}$  in (10) reflects endogenous peer effects, and  $\bar{x}_{2,(-i)r} \beta_3$  is the exogenous peer effect, also referred to as contextual peer effects. The covariates  $\bar{x}_{2,(-i)r}$  and  $x_{3,r}$  contain group level information and can be interpreted as parametrizations of group level fixed effects in the spirit of Mundlak (1978) and Chamberlain (1980). For example,  $x_{3,r}$  can contain full group averages of individual characteristics or be composed of other characteristics that only vary at the group level. The CMLE, as in the conventional panel case, cannot account for this group level information. This can be a limitation in cases where the effects of group level characteristics are of independent interest in the analysis. An example are the effects of teacher education and training on class test scores.

Let  $z_{ir} = (1, x_{1,ir}, \bar{x}_{2,(-i)r}, x_{3,r})$  be the row vector of all exogenous variables, let  $\beta = (\beta_1, \beta_2', \beta_3', \beta_4)'$  be the corresponding coefficients vector, and let  $k_Z$  denote the number of columns in  $z_{ir}$ . A compact form of model (10) is

$$y_{ir} = \lambda \bar{y}_{(-i)r} + z_{ir} \beta + \alpha_r + \epsilon_{ir}. \quad (11)$$

The model can be further written as a Cliff-Ord type spatial model. To see this let  $I_m$  denote the  $m$ -dimensional identity matrix, let  $\iota_m$  denote the  $m$ -dimensional column vector of ones, and define the weight matrix  $W_{m_r}$  for group  $r$  as  $W_{m_r} = \frac{1}{m_r-1} (\iota_{m_r} \iota_{m_r}' - I_{m_r})$ . The off-diagonal elements of this matrix are all equal to  $\frac{1}{m_r-1}$  and diagonal elements are 0. Let  $Y_r = (y_{1r}, \dots, y_{m_r r})'$ ,  $Z_r = (z'_{1r}, \dots, z'_{m_r r})'$ ,  $\epsilon_r = (\epsilon_{1r}, \dots, \epsilon_{m_r r})'$ , then the model for group  $r$  can be expressed in matrix

form as

$$Y_r = \lambda W_{m_r} Y_r + Z_r \beta + U_r, \quad (12)$$

where  $U_r = \alpha_r \iota_{m_r} + \epsilon_r$ . Let  $Y = [Y'_1, Y'_2, \dots, Y'_R]'$ ,  $Z = [Z'_1, Z'_2, \dots, Z'_R]'$ ,  $U = [U'_1, U'_2, \dots, U'_R]'$ , and  $W = \text{diag}_{r=1}^R \{W_{m_r}\}$  such that the model for the whole sample is given by

$$Y = \lambda W Y + Z \beta + U. \quad (13)$$

In the spatial literature  $W$  is referred to as a spatial weight matrix and  $WY$  as a spatial lag. In analyzing the model in (13) we maintain the random effects specification detailed in Assumptions 1 and 2 of Section 2, which imply that  $\alpha_r \sim (0, \sigma_\alpha^2)$  and  $\epsilon_{ir} \sim (0, \sigma_{\epsilon, D_r}^2)$ , where  $D_r \in \{1, \dots, J\}$  with  $J \geq 1$  fixed and finite. The specification allows for heteroscedasticity at the group level as long as there are only a finite number of different parameters. For example, we could allow for  $\sigma_{\epsilon, D_r}^2$  to be different for small and large groups, or more generally for all groups of a certain size  $m_r$ . On the other hand we do not cover the case where  $\sigma_{\epsilon, D_r}^2$  differs for each individual group  $r$ , as this would lead to an infinite dimensional parameter space.

The parameters of interest are  $\lambda, \sigma_\alpha^2, \sigma_{\epsilon, 1}^2, \dots, \sigma_{\epsilon, J}^2$  and  $\beta$ . Their respective true values are  $\lambda_0, \sigma_{\alpha 0}^2, \sigma_{\epsilon 0, 1}^2, \dots, \sigma_{\epsilon 0, J}^2$  and  $\beta_0$ . In analyzing the model it will be convenient to concentrate the log-likelihood function with respect to  $\beta$  for given values of  $\theta = (\lambda, \sigma_\alpha^2, \sigma_{\epsilon 1}^2, \dots, \sigma_{\epsilon J}^2)'$ . Let  $\Theta$  denote the parameter space for  $\theta$ , and let  $\delta = (\theta', \beta')'$  denote the vector of all parameters.

Under Assumptions 1 and 2 the expression for the variance covariance matrix  $\Omega_0$  of  $U$  is

$$\Omega_0 = \Omega(\theta_0) = \text{diag}_{r=1}^R \Omega_r(\theta_0) = \text{diag}_{r=1}^R \{\sigma_{\epsilon 0, D_r}^2 I_{m_r} + \sigma_{\alpha 0}^2 \iota_{m_r} \iota'_{m_r}\}.$$

To define the quasi-maximum likelihood estimator (QMLE) for the peer effects model in (11) note that solving  $Y$  from (13) yields the reduced form:

$$Y = (I - \lambda W)^{-1} Z \beta + (I - \lambda W)^{-1} U. \quad (14)$$

If  $\alpha_r$  and  $\epsilon_{ir}$  follow normal distributions,

$$Y \sim N((I - \lambda W)^{-1} Z \beta, (I - \lambda W)^{-1} \Omega(\theta) (I - \lambda W')^{-1}). \quad (15)$$

The corresponding log likelihood function is

$$\begin{aligned} \ln L_N(\theta, \beta) &= -\frac{N}{2} \ln(2\pi) + \frac{1}{2} \ln |(I - \lambda W)^2 \Omega(\theta)^{-1}| \\ &\quad - \frac{1}{2} (Y - \lambda W Y - Z \beta)' \Omega(\theta)^{-1} (Y - \lambda W Y - Z \beta). \end{aligned} \quad (16)$$

and the corresponding QMLE is given by

$$\hat{\delta}_N = (\hat{\theta}'_N, \hat{\beta}'_N)' = \text{argmax}_{\theta, \beta} \ln L_N(\theta, \beta). \quad (17)$$

It is convenient to concentrate out  $\beta$  and to obtain the QMLE for  $\theta$  first. The first order condition for  $\beta$  is

$$\frac{\partial \ln L_N(\theta, \beta)}{\partial \beta} = (Y - \lambda WY - Z\beta)' \Omega(\theta)^{-1} Z = 0, \quad (18)$$

which leads to

$$\hat{\beta}_N(\theta) = (Z' \Omega(\theta)^{-1} Z)^{-1} Z' \Omega(\theta)^{-1} (I - \lambda W) Y. \quad (19)$$

Plugging  $\hat{\beta}_N(\theta)$  back into (16) yields the following concentrated log likelihood function,

$$\begin{aligned} Q_N(\theta) &= \frac{1}{N} \ln L_N(\theta, \hat{\beta}_N(\theta)) \\ &= -\frac{\ln(2\pi)}{2} + \frac{1}{2N} \ln |(I - \lambda W)^2 \Omega(\theta)^{-1}| - \frac{1}{2N} Y' (I - \lambda W)' M_Z(\theta) (I - \lambda W) Y, \end{aligned} \quad (20)$$

where

$$M_Z(\theta) = \Omega(\theta)^{-1} - \Omega(\theta)^{-1} Z (Z' \Omega(\theta)^{-1} Z)^{-1} Z' \Omega(\theta)^{-1}. \quad (21)$$

Then the QMLE for  $\theta$ ,  $\hat{\theta}_N = (\hat{\lambda}_N, \hat{\sigma}_{\alpha, N}^2, \hat{\sigma}_{\epsilon_1, N}^2, \dots, \hat{\sigma}_{\epsilon_J, N}^2)'$  is given by

$$\hat{\theta}_N = \operatorname{argmax}_{\theta} Q_N(\theta). \quad (22)$$

Plugging  $\hat{\theta}_N$  back into (19), the QMLE for  $\beta$  is

$$\hat{\beta}_N = \hat{\beta}_N(\hat{\theta}_N) = (Z' \Omega(\hat{\theta}_N)^{-1} Z)^{-1} Z' \Omega(\hat{\theta}_N)^{-1} (I - \hat{\lambda}_N W) Y. \quad (23)$$

A formal result regarding the asymptotic identification of the model parameters is given in the next section. We next provide some intuition for that result, by extending our earlier discussion of identification for the canonical model without covariates to our model (13) with covariates. Let  $\|\cdot\|$  be the Euclidean norm on  $\mathbb{R}^k$ . Using the relationships  $\ddot{U}_r = \frac{(m_r - 1 + \lambda_0)}{m_r - 1} \ddot{Y}_r - \ddot{Z}_r \beta_0$  and  $\bar{u}_r = (1 - \lambda_0) \bar{y}_r - \bar{z}_r \beta_0$ , the moment functions related to the full model can be written as follows

$$\chi_r(\theta) = \begin{bmatrix} \chi_r^w(\theta) \\ \chi_r^b(\theta) \\ \chi_r^{zw}(\delta) \\ \chi_r^{zb}(\delta) \end{bmatrix} = \begin{bmatrix} \left\| \frac{m_r - 1 + \lambda}{m_r - 1} \ddot{Y}_r - \ddot{Z}_r \beta \right\|^2 - (m_r - 1) \sigma_{\epsilon, D_r}^2 \\ [(1 - \lambda) \bar{y}_r - \bar{z}_r \beta]^2 - \sigma_{\alpha}^2 - \frac{\sigma_{\epsilon, D_r}^2}{m_r} \\ \ddot{Z}_r' \left( \frac{m_r - 1 + \lambda}{m_r - 1} \ddot{Y}_r - \ddot{Z}_r \beta \right) \\ \bar{z}_r' ((1 - \lambda) \bar{y}_r - \bar{z}_r \beta) \end{bmatrix}$$

where  $\chi_r^w(\delta)$  and  $\chi_r^b(\delta)$  summarize the restrictions on the unobservables, and are natural extensions of the moment conditions considered before in (7) for the model without covariates. The additional moment restrictions  $\chi_r^{zw}(\delta)$  and  $\chi_r^{zb}(\delta)$  relate to the exogeneity of  $Z_r$  relative to  $\epsilon_r$  and  $\alpha_r$ . A formal asymptotic identification result will be given in the next section. Intuitively, for given  $\lambda$  the last two moment conditions identify  $\beta$ , while the first two identify  $\lambda, \sigma_{\alpha}^2, \sigma_{\epsilon, j}^2, j = 1, \dots, J$  in an analogous manner as described in the discussion of Lemma 2.1 for the model without covariates.

As for the model without covariates there is a representation of the score of the log-likelihood

in terms of the fundamental moment conditions. To describe the relationship between moments and the score we define the matrix

$$\varphi(m_r, D_r) = \begin{pmatrix} \frac{1}{(m_r-1+\lambda_0)\sigma_{\epsilon_0, D_r}^2} & -\frac{m_r}{(1-\lambda_0)(\sigma_{\epsilon_0, D_r}^2+m_r\sigma_{\alpha_0}^2)} & \frac{1}{(m_r-1+\lambda_0)\sigma_{\epsilon_0, D_r}^2}\beta_0' & -\frac{m_r}{(1-\lambda_0)(\sigma_{\epsilon_0, D_r}^2+m_r\sigma_{\alpha_0}^2)}\beta_0' \\ 0 & -\frac{m_r^2}{2(\sigma_{\epsilon_0, D_r}^2+m_r\sigma_{\alpha_0}^2)^2} & 0 & 0 \\ -\frac{1(D_r=1)}{2\sigma_{\epsilon_0, 1}^4} & -\frac{m_r 1(D_r=1)}{2(\sigma_{\epsilon_0, 1}^2+m_r\sigma_{\alpha_0}^2)^2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{1(D_r=J)}{2\sigma_{\epsilon_0, J}^4} & -\frac{m_r 1(D_r=J)}{2(\sigma_{\epsilon_0, J}^2+m_r\sigma_{\alpha_0}^2)^2} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sigma_{\epsilon_0, D_r}^2}I_{k_Z} & -\frac{m_r}{\sigma_{\epsilon_0, D_r}^2+m_r\sigma_{\alpha_0}^2}I_{k_Z} \end{pmatrix}. \quad (24)$$

Furthermore observe that the log-likelihood function can be written as  $\ln L_N(\delta) = -\frac{N}{2} \ln(2\pi) + \sum_{r=1}^R \ln L_r(\delta)$  where

$$\begin{aligned} \ln L_r(\delta) &= \frac{1}{2} \ln |(I_{m_r} - \lambda W_{m_r})^2 \Omega_r(\theta)^{-1}| \\ &\quad - \frac{1}{2} (Y_r - \lambda W_{m_r} Y_r - Z_r \beta)' \Omega_r(\theta)^{-1} (Y_r - \lambda W_{m_r} Y_r - Z_r \beta). \end{aligned}$$

is the log-likelihood function for group  $r$ . Then it can be shown that<sup>4</sup>

$$\frac{\partial \ln L_r(\delta_0)}{\partial \delta} = -\chi_r^*(\delta_0) = -\varphi(m_r, D_r) \chi_r(\delta_0).$$

As is well known, the score of the log-likelihood function,  $S(\delta) = -\sum_{r=1}^R \frac{\partial \ln L_r(\delta)}{\partial \delta}$  can be interpreted as a moment function corresponding to the moments  $E[S(\delta_0)] = -\sum_{r=1}^R E\left[\frac{\partial \ln L_r(\delta_0)}{\partial \delta}\right] = 0$ . Furthermore, under a Gaussian assumption the score is an optimal moment function.<sup>5</sup> From this we see that the matrices  $\varphi(m_r, D_r)$  can be viewed to provide the optimal weighting for the basic moment functions  $\chi_r(\delta)$ . Under Gaussian assumptions the optimal GMM estimator coincides with the maximum likelihood estimator and is asymptotically efficient under the stated assumptions.

<sup>4</sup>See our Online Appendix for details. The derivation uses Lemma B.1 and the special properties of matrices  $\Omega(\theta)$ ,  $I - \lambda W$  and  $W$  described in Appendix B.1. In the Online Appendix we also give an explicit expression for the variance covariance matrix of  $\chi_r(\delta)$ .

<sup>5</sup>Observe that

$$\begin{aligned} E\left[\frac{\partial S(\delta_0)}{\partial \delta'}\right] [Var(S(\delta_0))]^{-1} S(\delta_0) &= \left(-\sum_{r=1}^R E\left[\frac{\partial^2 \ln L_r(\delta_0)}{\partial \delta \partial \delta'}\right]\right) \left(\sum_{r=1}^R E\left[\frac{\partial \ln L_r(\delta_0)}{\partial \delta} \frac{\partial \ln L_r(\delta_0)}{\partial \delta'}\right]\right)^{-1} S(\delta_0) \\ &= S(\delta_0). \end{aligned}$$

in light of the information matrix equality.

## 4 Theoretical Results

We next state our assumptions for the general model. We maintain Assumptions 1-3 on  $\epsilon$ ,  $\alpha$  and  $\lambda$ . In the following we add assumptions regarding the exogenous variables, and the sizes and relative magnitudes of groups in the sample. Let  $\mathcal{I}_{m,j} \subset \{1, \dots, R\}$  be the index set of all groups in category  $j$  with size equal to  $m$ . Thus if  $r \in \mathcal{I}_{m,j}$ , then  $D_r = j$  and  $m_r = m$ . Let  $R_{m,j}$  be the cardinality of  $\mathcal{I}_{m,j}$ , in other words  $R_{m,j}$  is the number of groups in category  $j$  with size equal to  $m$ , and let  $R_j$  be the number of groups in category  $j$ , that is  $R_j = \sum_{r=1}^R 1(D_r = j) = \sum_{m=2}^{\bar{M}} R_{m,j}$ , where the upper bound  $\bar{M}$  on the group size is specified in the next assumption below. Furthermore let  $\omega_{m,j} = R_{m,j}/R$  denote the share of groups in category  $j$  with size equal to  $m$ , and let  $\omega_j = R_j/R = \sum_{m=2}^{\bar{M}} \omega_{m,j}$  be the share of groups in category  $j$ . Below we maintain the following assumption regarding the group sizes and their relative magnitudes.

**Assumption 4.** (a) *The sample size  $N$  goes to infinity;* (b) *The group size is bounded in the sense that there exists some positive constant  $\bar{M}$  such that  $2 \leq m_r \leq \bar{M} < \infty$  for  $r = 1, 2, \dots, R$ ;* (c) *The limit  $\omega_{m,j}^* = \lim_{N \rightarrow \infty} \omega_{m,j}$  exists and  $\omega_{m,j}^* < 1$  for all  $2 \leq m \leq \bar{M}$  and  $j$ , and  $\omega_j^* = \lim_{N \rightarrow \infty} \omega_j = \sum_{m=2}^{\bar{M}} \omega_{m,j}^* > 0$  for all  $j$ .*

The restriction that the minimal group size is 2 rules out singleton groups. A member of such a group has no peers. Assumption 4(b) imposes a fixed upper bound on group size. In many applications this is not a serious constraint. The assumption is more restrictive than Lee (2007) who allows for group size to grow with sample size. It is worth pointing out that increasing group sizes generally reduce the convergence rates for estimators of peer effects parameters, and as demonstrated by Kelejian and Prucha (2002) in some cases lead to inconsistency of these estimators.

Assumption 4(c) states that asymptotically, no single type-group size combination can dominate the sample by requiring that  $\omega_{m,j}^* < 1$  for all  $2 \leq m \leq \bar{M}$  and  $j$ . In addition, all types  $j$  occur in the sample in an asymptotically non-negligible way because  $\omega_j^* > 0$  for all  $j$ . On the other hand, we do allow that for certain combinations of  $j$  and  $m$  the limit  $\omega_{m,j}^*$  is zero, allowing for some group sizes of type  $j$  to occur infrequently or not at all in the sample.

Observe that  $N = \sum_{r=1}^R m_r = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} m R_{m,j}$ . Since group size is bounded, the number of groups  $R$  goes to infinity as  $N$  goes to infinity. Since  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} R_{m,j} = R$ , we have  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j} = \sum_{j=1}^J \omega_j = 1$  and thus  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* = \sum_{j=1}^J \omega_j^* = 1$ . Since  $\omega_j^* > 0$  by Assumption 4(c) it follows that also  $R_j$  goes to infinity, which is needed to facilitate the consistent estimation of  $\sigma_{\epsilon,j}^2$ . Assumption 4(c) implies that the limit of the average group size is given by

$$m^* = \lim_{N \rightarrow \infty} \frac{N}{R} = \lim_{N \rightarrow \infty} \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \frac{R_{m,j}}{R} m = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* m. \quad (25)$$

Clearly  $2 \leq m^* \leq \bar{M}$ , since  $2 \leq m_r \leq \bar{M}$ .

Observe that in light of Assumptions 1, 2, and 3 the parameter space  $\Theta$  for  $\theta = (\lambda, \sigma_\alpha^2 \sigma_{\epsilon,1}^2, \dots, \sigma_{\epsilon,J}^2)'$

is a compact subset of the Euclidean space  $\mathbb{R}^{2+J}$ . Observe further that

$$I_{m_r} - \lambda W_{m_r} = \left(1 + \frac{\lambda}{m_r - 1}\right) I_{m_r}^* + (1 - \lambda) J_{m_r}^*, \quad (26)$$

where  $I_{m_r}^* = I_{m_r} - \iota_{m_r} \iota'_{m_r} / m_r$  and  $J_{m_r}^* = \iota_{m_r} \iota'_{m_r} / m_r$  are symmetric, idempotent, orthogonal, and sum to the identity matrix. Furthermore from the results in Appendix B.1 we have  $|I_{m_r} - \lambda W_{m_r}| = [1 + \lambda/(m_r - 1)]^{m_r - 1} (1 - \lambda)$ .<sup>6</sup> Thus the matrix  $I_{m_r} - \lambda W_{m_r}$  is nonsingular if  $1 + \lambda/(m_r - 1) \neq 0$  and  $1 - \lambda \neq 0$ . Assumption 3 ensures the non-singularity of  $I_{m_r} - \lambda W_{m_r}$ , and hence the non-singularity of  $I - \lambda W = \text{diag}_{r=1}^R \{I_{m_r} - \lambda W_{m_r}\}$ , since for  $m_r \geq 2$  and  $\lambda < 1$  we have  $1 + \lambda/(m_r - 1) > 0$  and  $1 - \lambda > 0$ .

Let  $\bar{z}_r = \frac{1}{m_r} \iota'_{m_r} Z_r$  be the row vector of column means of  $Z_r$ , and let  $\ddot{Z}_r = Z_r - \iota_{m_r} \bar{z}_r$  be the deviations from the column means. Then  $Z_r' I_{m_r}^* Z_r = \ddot{Z}_r' \ddot{Z}_r$ ,  $Z_r' J_{m_r}^* Z_r = m_r \bar{z}_r' \bar{z}_r$ .

**Assumption 5.** (a) *The  $N \times k_Z$  matrix  $Z$  is non-stochastic, with  $\text{rank}(Z) = k_Z > 0$  for  $N$  sufficiently large. The elements of  $Z$  are uniformly bounded in absolute value.*

(b) *For  $2 \leq m \leq \bar{M}$ , and  $1 \leq j \leq J$  the following limits exist:*

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r = \ddot{\varkappa}_{m,j},$$

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r = \bar{\varkappa}_{m,j},$$

$$\lim_{N \rightarrow \infty} N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \bar{z}_r = \bar{z}_{m,j}.$$

(c) *For at least one pair of  $(m, j)$  such that  $\omega_{m,j}^* > 0$ , and  $N$  sufficiently large, the smallest eigenvalues of  $N^{-1} \sum_{r \in \mathcal{I}_{m,j}} Z_r' Z_r = N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r + N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r$  are bounded away from zero, uniformly in  $N$ , by some finite constant  $\xi_Z > 0$ .*

Suppose we have some  $N \times N$  matrix  $A_N(\theta) = \text{diag}_{r=1}^R \{p(m_r, D_r, \theta) I_{m_r}^* + s(m_r, D_r, \theta) J_{m_r}^*\}$ , where  $p(m_r, D_r, \theta)$  and  $s(m_r, D_r, \theta)$  are positive, uniformly continuous and bounded on  $\Theta$ . An example of an expression of this form is  $\Omega(\theta)^{-1}$  which is obtained in closed form in Equation (B.3) in Appendix B.1. Then under Assumption 5(b), the limiting matrix of  $N^{-1} Z' A_N(\theta) Z$  always exists, is continuous in  $\theta$  and takes the form

$$\lim_{N \rightarrow \infty} \frac{1}{N} Z' A_N(\theta) Z = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} [p(m, j, \theta) \ddot{\varkappa}_{m,j} + s(m, j, \theta) \bar{\varkappa}_{m,j}].$$

Furthermore,  $N^{-1} Z' A_N(\theta) Z$  converges to its limiting matrix uniformly on  $\Theta$ . With  $p(m_r, D_r, \theta) > 0$  and  $s(m_r, D_r, \theta) > 0$ , Assumption 5(a) ensures that  $N^{-1} Z' A_N(\theta) Z$  and its limiting matrix are

<sup>6</sup>In Appendix B.1 we review additional properties of matrices of the form  $pI_m^* + sJ_m^*$ , which will be used repeatedly in this paper. In particular, their multiplication is commutative. The products of such matrices are also of the form of  $pI_m^* + sJ_m^*$ , and  $|pI_m^* + sJ_m^*| = p^{m-1}s$ ,  $(pI_m^* + sJ_m^*)^{-1} = \frac{1}{p}I_m^* + \frac{1}{s}J_m^*$ .



invertible, with the elements of the inverse matrix uniformly bounded in absolute value. In the special case when  $A_N(\theta)$  is the identity matrix,  $\lim_{N \rightarrow \infty} \frac{1}{N} Z'Z = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} [\dot{\omega}_{m,j} + \bar{\omega}_{m,j}]$ , which has the smallest eigenvalue bounded above zero by some finite constant  $\xi_Z > 0$ . See Lemma B.5 for details and a proof.

As shown by Lemma 2.1 in Section 2, identification of  $\lambda$  and  $\sigma_\alpha^2$  requires variation in the group size or variance of the error terms. The following assumption ensures this so that in the limit we have non-negligible samples for at least two different group sizes or two different categories with different variances of the idiosyncratic errors  $\epsilon_{ir}$ .

**Assumption 6.** *For some sizes  $m$  and  $m'$ , and some categories  $j$  and  $j'$  we have  $\omega_{m,j}^* > 0$  and  $\omega_{m',j'}^* > 0$ , and either of the following two scenarios hold,*

(a)  $m \neq m'$ , and  $\sigma_{\epsilon_0,j}^2 = \sigma_{\epsilon_0,j'}^2$  for some  $j, j' \in \{1, \dots, J\}$ . (b)  $m = m'$ , and  $\sigma_{\epsilon_0,j}^2 \neq \sigma_{\epsilon_0,j'}^2$  for some  $j, j' \in \{1, \dots, J\}$  with  $j \neq j'$ .

The conditions in Assumption 6 are the asymptotic analogs of identification conditions imposed in Lemma 2.1. Assumption 4 by itself is not sufficient for identification because it only implies that no single pair  $(m, j)$  asymptotically dominates the sample. Assumption 4 alone does not guarantee that there is enough variation in the underlying group sizes  $m$  or the variances  $\sigma_{\epsilon_0,j}^2$ . For example, it is possible under Assumption 4 that all groups are of the same size and that all variances  $\sigma_{\epsilon_0,j}^2$  are the same. Assumption 6 rules out such cases. Assumption 6(a) is related to Assumption 6.1 and Footnote 9 of Lee (2007) which requires group size variation to achieve identification for the case where group sizes are bounded, which is the only case we consider. Assumption 6(b) has no analog in Lee (2007) because of his Assumption 1 which imposes homoscedasticity on the errors  $\epsilon_{ir}$ . We show that identification is possible purely based on group level heteroscedasticity even if all group sizes are the same. This insight also extends the analysis of Graham (2008) where types and class sizes are linked.

Below we give results on the consistency and asymptotic normality of the QMLE  $\hat{\delta}_N = (\hat{\theta}'_N, \hat{\beta}'_N)'$  defined in (17).

**Theorem 4.1.** *Suppose Assumptions 1-6 hold, then*

(a) *The parameter  $\delta_0$  is asymptotically identified in the sense that it is the unique maximizer of the criterion  $\bar{R}(\theta, \beta) = \lim_{N \rightarrow \infty} E \left[ \frac{1}{N} \ln L(\theta, \beta) \right]$ .*

(b) *The QMLE  $\hat{\delta}_N$  is consistent, i.e.,  $\hat{\delta}_N \xrightarrow{p} \delta_0$  as  $N \rightarrow \infty$ .*

A detailed proof of the theorem is given in Appendices E.1 and E.2. As can be seen from the proof, the argumentation that ensures part (a) of the theorem is analogous to the argumentation used in establishing Lemma 2.1. Here is a sketch of the proof to provide some intuition. The limiting expected value of the concentrated log likelihood function  $Q_N(\theta)$  is

$$\bar{Q}^*(\theta) = C^* + \frac{1}{2m^*} \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta) + Q^{(2)*}(\theta),$$

where  $C^*$  is a constant term,  $g(m, j, \theta) = \ln |G(m, j, \theta)| - \text{tr} G(m, j, \theta)$  with

$$G(m, j, \theta) = \frac{\sigma_{\epsilon 0, j}^2}{\sigma_{\epsilon, j}^2} \left( \frac{m-1+\lambda}{m-1+\lambda_0} \right)^2 I_m^* + \frac{(\sigma_{\epsilon 0, j}^2 + m\sigma_{\alpha 0}^2)}{(\sigma_{\epsilon, j}^2 + m\sigma_{\alpha}^2)} \left( \frac{1-\lambda}{1-\lambda_0} \right)^2 J_m^*,$$

and  $Q^{(2)*}(\theta) = \lim_{N \rightarrow \infty} \bar{Q}_N^{(2)}(\theta)$  where  $\bar{Q}_N^{(2)}(\theta) = -\frac{1}{2N} \tilde{\eta}_Z(\theta)' \tilde{M}_Z(\theta) \tilde{\eta}_Z(\theta)$  with  $\tilde{M}_Z(\theta) = I - \Omega(\theta)^{-1/2} Z(Z'\Omega(\theta)^{-1}Z)^{-1} Z'\Omega(\theta)^{-1/2}$  and

$$\tilde{\eta}_Z(\theta) = \Omega(\theta)^{-1/2} (I - \lambda W) (I - \lambda_0 W)^{-1} Z \beta_0.$$

It is easy to see that  $\theta_0$  is a global maximizer of  $Q^{(2)*}(\theta)$ , given that  $-\bar{Q}_N^{(2)}(\theta)$  is the quadratic form of an idempotent and thus positive semi-definite matrix, and  $Q^{(2)*}(\theta_0) = 0$ . However, this does not ensure that  $\theta_0$  is a unique global maximizer. Identification thus comes from  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta)$ . Note that for any symmetric positive definite  $m \times m$  matrix  $A$ ,  $\ln |A| - \text{tr}(A) \leq -m$  with equality if and only if  $A$  is an identity matrix.<sup>7</sup> For any  $m$  and  $j$ ,  $g(m, j, \theta)$  is maximized if and only if  $G(m, j, \theta) = I_m$ , which is equivalent to  $E[\chi_r(\theta) | m_r = m, D_r = j] = 0$  with  $\chi_r(\theta) = (\chi_r^w(\theta), \chi_r^b(\theta))$  defined in (7). It now follows from an asymptotic analogue of Lemma 2.1 that in either case (i) or (ii) of Assumption 6,  $\theta_0$  is the only solution to  $E[\chi_r(\theta) | m_r = m, D_r = j] = 0$  and  $E[\chi_r(\theta) | m_r = m', D_r = j'] = 0$ . Thus for any  $\theta \neq \theta_0$ ,  $\min(g(m, j, \theta_0) - g(m, j, \theta), g(m', j', \theta_0) - g(m', j', \theta)) > 0$ . As a result,  $\theta_0$  is the unique global maximizer of  $\bar{Q}^*(\theta)$  when one of the two scenarios holds true for some  $\omega_{m,j}^* > 0$  and  $\omega_{m',j'}^* > 0$ .

To study the asymptotic distribution of the estimator, first note that under Assumptions 1 and 2, the third and fourth moments of  $\epsilon_{ir}$  and  $\alpha_r$  exist. Let  $E[\epsilon_{ir}^3 | D_r = j] = \mu_{\epsilon 0, j}^{(3)}$ ,  $E[\epsilon_{ir}^4 | D_r = j] = \mu_{\epsilon 0, j}^{(4)}$ ,  $E[\alpha_r^3] = \mu_{\alpha 0}^{(3)}$  and  $E[\alpha_r^4] = \mu_{\alpha 0}^{(4)}$ . Also, define  $\Gamma_0$  and  $\Upsilon_0$  as

$$\begin{aligned} \Gamma_0 &= \lim_{N \rightarrow \infty} N^{-1} E \left[ -\frac{\partial^2 \ln L_N(\delta_0)}{\partial \delta \partial \delta'} \right], \\ \Upsilon_0 &= \lim_{N \rightarrow \infty} N^{-1} E \left[ \frac{\partial \ln L_N(\delta_0)}{\partial \delta} \frac{\partial \ln L_N(\delta_0)}{\partial \delta'} \right]. \end{aligned}$$

As shown in Appendix E.3, the two limiting matrices exist. Specific expressions are given in Appendix F. When  $\epsilon_{ir}$  and  $\alpha_r$  both follow normal distributions,  $\Upsilon_0 = \Gamma_0$ .

The next lemma shows that  $\Gamma_0$  is p.d. under the maintained assumptions. The lemma also provides a sufficient condition on the moments of  $\epsilon$  under which  $\Upsilon_0$  is p.d..

**Lemma 4.1.** *Suppose Assumptions 1-6 hold, then  $\Gamma_0$  is positive definite. Under the additional assumption that  $\mu_{\epsilon 0, j}^{(4)} - \sigma_{\epsilon 0, j}^4 > (\mu_{\epsilon 0, j}^{(3)})^2 / \sigma_{\epsilon 0, j}^2$  for all  $j \in \{1, \dots, J\}$ ,  $\Upsilon_0$  is also positive definite.*

The proof of the lemma is in Appendix F. Note that from Holder's inequality we have  $\mu_{\epsilon 0, j}^{(4)} - \sigma_{\epsilon 0, j}^4 \geq (\mu_{\epsilon 0, j}^{(3)})^2 / \sigma_{\epsilon 0, j}^2$ . The sufficient condition is mild in that it only postulates that the inequality

<sup>7</sup>To see this, note that under the maintained assumptions the eigenvalues of  $A$ , say,  $\lambda_i$ , are positive and  $\ln |A| - \text{tr}[A] = \sum_{i=1}^m [\ln(\lambda_i) - \lambda_i]$ . The claim is seen to hold by observing that the function  $f(x) = \ln(x) - x \leq -1$  for  $x \in (0, \infty)$  with a unique maximum at  $x = 1$ , and observing that  $A = I_m$  if and only if  $\lambda_i = 1$  for  $i = 1, \dots, m$ .

holds strongly. Of course, the condition holds, e.g., for the Gaussian distribution.

With both  $\Upsilon_0$  and  $\Gamma_0$  ensured to be positive definite, we have the following theorem.

**Theorem 4.2.** *Under Assumptions 1-6, and assuming that  $\delta_0$  is in the interior of the parameter space  $\Theta$  defined in Assumption 3 and that  $\mu_{\varepsilon 0, j}^{(4)} - \sigma_{\varepsilon 0, j}^4 > (\mu_{\varepsilon 0, j}^{(3)})^2 / \sigma_{\varepsilon 0, j}^2$  for  $j \in \{1, \dots, J\}$ , we have  $\sqrt{N}(\hat{\delta}_N - \delta_0) \xrightarrow{d} N(0, \Gamma_0^{-1} \Upsilon_0 \Gamma_0^{-1})$  as  $N \rightarrow \infty$ .*

The proof of the theorem is given in Appendix E.3. We next discuss consistent estimators for the matrices  $\Gamma_0$  and  $\Upsilon_0$  composing the asymptotic variance covariance matrix. An inspection shows that  $\Gamma_0 = \Gamma(\delta_0, s_0)$  and  $\Upsilon_0 = \Upsilon(\delta_0, \mu_{\alpha 0}^{(3)}, \mu_{\alpha 0}^{(4)}, \mu_{\varepsilon 0, 1}^{(3)}, \dots, \mu_{\varepsilon 0, J}^{(3)}, \mu_{\varepsilon 0}^{(4)}, \dots, \mu_{\varepsilon 0, J}^{(4)}, s_0)$ , with  $s_0 = (s_{0,1}, \dots, s_{0,J}, m^*)$  and

$$s_{0,j} = [\check{z}_{2,j}, \dots, \check{z}_{\bar{M},j}, \bar{z}_{2,j}, \dots, \bar{z}_{\bar{M},j}, \bar{z}_{2,j}, \dots, \bar{z}_{(\bar{M}),j}, \omega_{2,j}^*, \dots, \omega_{\bar{M},j}^*],$$

and where the functions  $\Gamma(\cdot)$  and  $\Upsilon(\cdot)$  are continuous. Since the functions  $\Gamma(\cdot)$  and  $\Upsilon(\cdot)$  are continuous, consistent estimators for  $\Gamma_0$  and  $\Upsilon_0$  can be readily obtained by replacing the arguments of those functions by consistent estimators thereof. Let  $\hat{s}_N$  be the sample analogue of  $s_0$ , then clearly  $\hat{s}_N \xrightarrow{p} s_0$  in light of Assumptions 4 and 5. Recall further that by Theorem 4.1 the QMLE estimator  $\hat{\delta}_N$  is consistent for  $\delta_0$ , and suppose we have consistent estimators for  $\mu_{\alpha 0}^{(3)}, \mu_{\alpha 0}^{(4)}, \mu_{\varepsilon 0, 1}^{(3)}, \dots, \mu_{\varepsilon 0, J}^{(3)}$ , and  $\mu_{\varepsilon 0, 1}^{(4)}, \dots, \mu_{\varepsilon 0, J}^{(4)}$ , denoted as  $\hat{\mu}_{\alpha}^{(3)}, \hat{\mu}_{\alpha}^{(4)}, \hat{\mu}_{\varepsilon, 1}^{(3)}, \dots, \hat{\mu}_{\varepsilon, J}^{(3)}, \hat{\mu}_{\varepsilon, 1}^{(4)}, \dots, \hat{\mu}_{\varepsilon, J}^{(4)}$ . Now define  $\hat{\Gamma}_N$  and  $\hat{\Upsilon}_N$  as

$$\hat{\Gamma}_N = \Gamma(\hat{\delta}_N, \hat{s}_N), \tag{27}$$

$$\hat{\Upsilon}_N = \Upsilon(\hat{\delta}_N, \hat{\mu}_{\alpha}^{(3)}, \hat{\mu}_{\alpha}^{(4)}, \hat{\mu}_{\varepsilon, 1}^{(3)}, \dots, \hat{\mu}_{\varepsilon, J}^{(3)}, \hat{\mu}_{\varepsilon, 1}^{(4)}, \dots, \hat{\mu}_{\varepsilon, J}^{(4)}, \hat{s}_N), \tag{28}$$

then it follows from Slutsky's theorem that  $\hat{\Gamma}_N$  and  $\hat{\Upsilon}_N$  are consistent estimators for  $\Gamma_0$  and  $\Upsilon_0$ . A consistent estimator for the variance covariance matrix of the limiting distribution is given by  $\hat{\Gamma}_N^{-1} \hat{\Upsilon}_N \hat{\Gamma}_N^{-1}$ .

The above discussion assumed the availability of consistent estimators for the third and fourth moment of the error components. In the following we now define consistent estimators for  $\mu_{\alpha 0}^{(3)}$ ,  $\mu_{\alpha 0}^{(4)}$  and  $\mu_{\varepsilon 0, j}^{(3)}, \mu_{\varepsilon 0, j}^{(4)}$ ,  $j = 1, \dots, J$ . To motivate the estimators consider the composite error term for individual  $i$  in group  $r$ ,  $u_{ir} = \alpha_r + \varepsilon_{ir}$ , and let  $\bar{u}_r = \frac{1}{m_r} \sum_{i=1}^{m_r} u_{ir}$ , and  $\bar{\varepsilon}_r = \bar{u}_r - \alpha_r$ . Then  $\bar{u}_r = \alpha_r + \bar{\varepsilon}_r$  and  $\bar{\varepsilon}_r = \bar{u}_r - \alpha_r$ , where  $\bar{\varepsilon}_r$  is the group mean of  $\varepsilon_{ir}$ . It is readily verified that under Assumptions 1 and 2, we have

$$E[\bar{\varepsilon}_{ir}^3] = \left(1 - \frac{3}{m_r} + \frac{2}{m_r^2}\right) \mu_{\varepsilon 0, D_r}^{(3)},$$

$$E[\bar{\varepsilon}_{ir}^2 \bar{u}_r] = \frac{(m_r - 1)}{m_r^2} \mu_{\varepsilon 0, D_r}^{(3)},$$

$$E[\bar{u}_r^3] = \mu_{\alpha 0}^{(3)} + \frac{\mu_{\varepsilon 0, D_r}^{(3)}}{m_r^2},$$

$$E \left[ \ddot{u}_{ir}^4 \right] = \frac{m_r^3 - 4m_r^2 + 6m_r - 3}{m_r^3} \mu_{\epsilon 0, D_r}^{(4)} + \frac{3(m_r - 1)(2m_r - 3)}{m_r^3} \sigma_{\epsilon 0, D_r}^4,$$

$$E \left[ \bar{u}_r^4 \right] = \mu_{\alpha 0}^{(4)} + \frac{1}{m_r^3} \mu_{\epsilon 0, D_r}^{(4)} + 3 \frac{m_r - 1}{m_r^3} \sigma_{\epsilon 0, D_r}^4 + \frac{6}{m_r} \sigma_{\alpha 0}^2 \sigma_{\epsilon 0, D_r}^2.$$

Next define for group  $r$ ,

$$f_{\epsilon, r}^{(3)} = \begin{cases} \frac{1}{m_r} \sum_{i=1}^{m_r} \ddot{u}_{ir}^3 / (1 - \frac{3}{m_r} + \frac{2}{m_r^2}) & m_r \geq 3 \\ \frac{1}{m_r} \sum_{i=1}^{m_r} \ddot{u}_{ir}^2 \bar{u}_r / (\frac{1}{m_r} - \frac{1}{m_r^2}) & m_r = 2 \end{cases},$$

$$f_{\alpha, r}^{(3)} = \bar{u}_r^3 - f_{\epsilon, r}^{(3)} / m_r^2,$$

$$f_{\epsilon, r}^{(4)} = \frac{m_r^3}{m_r^3 - 4m_r^2 + 6m_r - 3} \left[ \left( \frac{1}{m_r} \sum_{i=1}^{m_r} \ddot{u}_{ir}^4 \right) - \frac{3(m_r - 1)(2m_r - 3)}{m_r^3} \sigma_{\epsilon 0, D_r}^4 \right],$$

$$f_{\alpha, r}^{(4)} = \bar{u}_r^4 - f_{\epsilon, r}^{(4)} / m_r^3 - \frac{3(m_r - 1)}{m_r^3} \sigma_{\epsilon 0, D_r}^4 - \frac{6}{m_r} \sigma_{\alpha 0}^2 \sigma_{\epsilon 0, D_r}^2.$$

Then  $E \left[ f_{\epsilon, r}^{(3)} \right] = \mu_{\epsilon 0, D_r}^{(3)}$ ,  $E \left[ f_{\alpha, r}^{(3)} \right] = \mu_{\alpha 0}^{(3)}$ ,  $E \left[ f_{\epsilon, r}^{(4)} \right] = \mu_{\epsilon 0, D_r}^{(4)}$ ,  $E \left[ f_{\alpha, r}^{(4)} \right] = \mu_{\alpha 0}^{(4)}$ . By Lemma B.4(a),  $\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) f_{\epsilon, r}^{(l)} \xrightarrow{p} \mu_{\epsilon 0, j}^{(l)}$  and  $\frac{1}{R} \sum_{r=1}^R f_{\alpha, r}^{(l)} \xrightarrow{p} \mu_{\alpha 0}^{(l)}$  for  $l = 3, 4$  and  $j = 1, \dots, J$  as  $R$  goes to infinity.

To construct feasible counterparts of these estimates, consider the estimated disturbances  $\hat{u}_{ir} = y_{ir} - \hat{\lambda} \bar{y}_{(-i)r} - z_{ir} \hat{\beta}$ , where  $\hat{\lambda}$  and  $\hat{\beta}$  denote the QML estimators, and let  $\hat{\bar{u}}_r = \frac{1}{m_r} \sum_{i=1}^{m_r} \hat{u}_{ir}$  and  $\hat{\ddot{u}}_{ir} = \hat{u}_{ir} - \hat{\bar{u}}_r$ . Feasible counterparts, say,  $\hat{f}_{\epsilon, r}^{(3)}$ ,  $\hat{f}_{\alpha, r}^{(3)}$ ,  $\hat{f}_{\epsilon, r}^{(4)}$ ,  $\hat{f}_{\alpha, r}^{(4)}$  of  $f_{\epsilon, r}^{(3)}$ ,  $f_{\alpha, r}^{(3)}$ ,  $f_{\epsilon, r}^{(4)}$ ,  $f_{\alpha, r}^{(4)}$  can now be defined by replacing  $\bar{u}_r$  and  $\ddot{u}_{ir}$  with  $\hat{\bar{u}}_r$  and  $\hat{\ddot{u}}_{ir}$ , and  $\sigma_{\alpha 0}^2$  and  $\sigma_{\epsilon 0, j}^2$  with their QML estimators. Now consider the following estimators for the third and fourth moments of the error components:  $\hat{\mu}_{\alpha}^{(3)} = \sum_{r=1}^R \hat{f}_{\alpha, r}^{(3)} / R$ ,  $\hat{\mu}_{\alpha}^{(4)} = \sum_{r=1}^R \hat{f}_{\alpha, r}^{(4)} / R$ ,  $\hat{\mu}_{\epsilon, j}^{(3)} = \sum_{r=1}^R 1(D_r = j) \hat{f}_{\epsilon, r}^{(3)} / R_j$ ,  $\hat{\mu}_{\epsilon, j}^{(4)} = \sum_{r=1}^R 1(D_r = j) \hat{f}_{\epsilon, r}^{(4)} / R_j$ ,  $j = 1, \dots, J$ .

The next theorem establishes that valid inference based on standardized statistics is possible. At the core of this result is the fact that  $\hat{\Gamma}_N \xrightarrow{p} \Gamma_0$  and  $\hat{\Upsilon}_N \xrightarrow{p} \Upsilon_0$  as shown in Appendix E.

**Theorem 4.3.** *Under Assumptions 1-6, and assuming that  $\mu_{\epsilon 0, j}^{(4)} - \sigma_{\epsilon 0, j}^4 > (\mu_{\epsilon 0, j}^{(3)})^2 / \sigma_{\epsilon 0, j}^2$  for  $j \in \{1, \dots, J\}$ , and  $\hat{\Gamma}_N$ ,  $\hat{\Upsilon}_N$  defined in (27) and (28) we have  $\sqrt{N} \left( \hat{\Gamma}_N^{-1} \hat{\Upsilon}_N \hat{\Gamma}_N^{-1} \right)^{-1/2} (\hat{\delta}_N - \delta_0) \xrightarrow{d} N(0, I)$  as  $N \rightarrow \infty$ .*

The proof of the theorem is in Appendix E.

## 5 Monte Carlo Results

We conduct Monte-Carlo (MC) experiments to assess the finite sample properties of the quasi-maximum likelihood (QML) estimator  $\hat{\delta}_N$ . The data generating mechanism is determined by the main model in (11). For simplicity,  $x_{1,ir}$ ,  $x_{2,ir}$  and  $x_{3,ir}$  each only includes a scalar variable. We

set the true value of the parameters to  $\lambda_0 = 0.5$ ,  $\sigma_{\alpha_0}^2 = 0.25$ ,  $\beta_{00} = 1$ ,  $\beta_{10} = 1$ ,  $\beta_{20} = 1$  and  $\beta_{30} = 1$ , while  $\sigma_{\epsilon_0}^2 = 1$  in the case of homoscedasticity. The model for the data generating process (DGP) is thus

$$y_{ir} = 0.5\bar{y}_{(-i)r} + 1 + x_{1,ir} + \bar{x}_{2,(-i)r} + x_{3,r} + \alpha_r + \epsilon_{ir}. \quad (29)$$

The inputs  $x_{j,ir}$ ,  $\alpha_r$  and  $\epsilon_{ir}$  are generated as follows. In the case when  $x_1 = x_2$ ,  $x_{1,ir} = x_{2,ir} \sim$  i.i.d.  $N(0, 1)$ . In the case when  $x_1 \neq x_2$ ,  $x_{1,ir}$  and  $x_{2,ir}$  are generated mutually independently, each drawn from an i.i.d.  $N(0, 1)$ . We then calculate the leave-out-mean  $\bar{x}_{2,(-i)r} = \frac{1}{m_r} \sum_{j \neq i} x_{2j,r}$ . Group characteristics are drawn as  $x_{3,r} \sim$  i.i.d.  $N(0, 1)$ . In the case of homoscedastic normal errors in Tables 1 to 5, the idiosyncratic error terms  $\epsilon_{ir}$  are i.i.d.  $N(0, 1)$  and group effects  $\alpha_r$  are i.i.d.  $N(0, 0.25)$ . Both  $\epsilon_{ir}$  and  $\alpha_r$  are drawn independently of  $x_{1,ir}$ ,  $x_{2,ir}$ ,  $x_{3,r}$ , and of each other. The dependent variable  $y_{ir}$  is calculated using Equation (14). In Table 6, we use homoscedastic but nonnormal errors. In the case of the Skew normal distribution, we set the location parameter to 0, scale to 1 and shape to  $0.9/\sqrt{1 - 0.9^2}$ . Therefore, Skewness is 0.472 and Kurtosis is 3.321. In the case of the student distribution, degrees of freedom are set to 6. Therefore, Skewness is 0 and Kurtosis is 6. In both cases,  $\alpha_r$  and  $\epsilon_{ir}$  are independently drawn from identical distributions and then standardized to have mean 0 and variance 0.25 and 1 respectively. In Table 7, group effects  $\alpha_r$  are still i.i.d.  $N(0, 0.25)$ ,  $\epsilon_{ir}$  follow normal distributions but are allowed to be heteroscedastic. In the first case (Columns 1-2), we randomly select half of the groups into category 1, with  $\epsilon_{ir}$  i.i.d.  $N(0, 0.5)$ . The other half of the groups have  $\epsilon_{ir}$  i.i.d.  $N(0, 1.5)$ . In the second case (Columns 3-4),  $\epsilon_{ir}$  are i.i.d.  $N(0, 1)$ . But we randomly divide the groups into two categories and allow for heteroscedasticity of  $\epsilon_{ir}$  between categories in estimation. In the third case (Columns 5-6), groups are randomly divided into two categories, with  $\sigma_{\epsilon_r}^2 \in \{0.5, 1.5\}$  and  $\epsilon_{ir}$  i.i.d.  $N(0, \sigma_{\epsilon_r}^2)$ . In the fourth case, groups are randomly divided into four categories with  $\sigma_{\epsilon_r}^2 \in \{0.4, 0.8, 1.2, 1.6\}$  and  $\epsilon_{ir}$  i.i.d.  $N(0, \sigma_{\epsilon_r}^2)$ .

The number of groups  $R$  is selected from the set  $\{50, 100, 200, 400, 800, 1600\}$ . In Tables 1, 2 and 6, group size  $m_r$  is drawn from a discrete uniform distribution  $\mathcal{U}\{2, 6\}$  so that the average group size is 4. Small group sizes are motivated by applications to college room mates, friendship networks in the Add Health data set or golf tournaments, see Sacerdote (2001), Goldsmith-Pinkham and Imbens (2013) and Guryan et al. (2009). In Tables 3 and 4, group size is drawn from  $\mathcal{U}\{13, 25\}$ . The distribution is motivated by Project STAR where class size ranges from 13 to 25. We also consider the case when  $m_r$  is drawn from  $\mathcal{U}\{3, 5\}$ ,  $\mathcal{U}\{4, 8\}$ ,  $\mathcal{U}\{8, 30\}$  and  $\mathcal{U}\{10, 22\}$  in Table 5 to examine how the distribution of group size affects the performance of the estimator. Note that  $\mathcal{U}\{3, 5\}$  has the same mean as  $\mathcal{U}\{2, 6\}$  but smaller variance,  $\mathcal{U}\{4, 8\}$  has the same variance as  $\mathcal{U}\{2, 6\}$  but larger mean. Meanwhile  $\mathcal{U}\{8, 30\}$  has the same mean as  $\mathcal{U}\{13, 25\}$  but larger variance,  $\mathcal{U}\{10, 22\}$  has the same variance as  $\mathcal{U}\{13, 25\}$  but smaller mean.

In Tables 1-6, we compare our QML estimator with the conditional maximum likelihood (CML) estimator of Lee (2007). Table 7 does not present CMLE estimates as it does not allow for heteroscedasticity. Lee (2007) assumes normality of the error terms. Our discussion suggests that the CMLE is in fact consistent under nonnormal errors, as it can be viewed as a GMM estimator based

on the moment conditions from the within equation. When group effects are in fact independent of the observed characteristics, the CML estimator is still consistent but less efficient than our QML estimator. The comparison thus helps to evaluate the efficiency gain of our estimator over the CML estimator in finite samples. The CML estimator is based on the within-group variation hence  $\sigma_\alpha^2$ ,  $\beta_0$  and  $\beta_3$  are not identified.

We generate 5000 repetitions for each of the experiments. Tables 1-7 summarize the results of the Monte Carlo (MC) experiments. Each panel displays the MC median, MC robust standard errors (Rob.Std.Dev), MC sample standard deviation (Std.Dev.), MC median of the estimated standard deviation (est.Std.Dev), and the mean rejection rate of the Wald test with significance level 0.05 of our QMLE and Lee's CMLE across 5000 repetitions. The robust standard errors are defined as  $IQ/1.35$ . where  $IQ$  denotes the inter-quantile range, that is  $IQ = C_{0.75} - C_{0.25}$  with  $C_{0.75}$  and  $C_{0.25}$  being the 75th and 25th percentile respectively. If the distribution of the estimate is normal,  $IQ/1.35$  is (apart from rounding errors) equal to the standard deviation. The null hypothesis for the Wald test is that the estimate equals its true value. Critical values for the test are obtained at 5% significance level and are based on the asymptotic approximation in Theorem 4.3.

Identification of our models is more challenging, the larger group sizes are, all else equal. This follows from work of Kelejian and Prucha (2002). Identification is also more difficult when there is less variation in group sizes, or less variation in type specific variances or both. Finally, identification is more difficult in designs where  $x_{1,ir} = x_{2,ir}$  because the implied correlation between  $x_{1,ir}$  and  $\bar{x}_{2,(-i)r}$  reduces the overall variation in the covariates. Standard finite sample theory for the Gaussian regression model shows that maximum likelihood estimators for the variance parameters are biased in finite samples. In fixed effects panel regressions this finite sample bias can lead to inconsistent estimates of the variance parameter due to incidental parameter bias, as demonstrated by Neyman and Scott (1948). In the current context, we expect the CML estimator to suffer from such incidental parameter bias because the moment conditions that identify  $\lambda$  depend on the estimated variances. We also expect Wald type statistics, such as the t-ratio, to perform poorly in designs where identification is problematic, in line with insights from Dufour (1997).

Tables 1 and 2 contain results for small groups and homoscedastic Gaussian errors. In Table 1 where  $x_1 \neq x_2$ , both the QMLE and CMLE perform well, with the CMLE being more biased for the parameter  $\lambda$  in sample sizes where  $R$  is below 200. The QMLE is generally less biased and significantly more precise than CMLE, demonstrating the expected efficiency gains of QMLE. Size is better controlled for CMLE but the size distortions for the parameters  $\lambda$  and  $\beta$  do not exceed 7% in the smallest sample sizes even for the QMLE. Size distortions for the t-ratios of the two estimated variance parameters are somewhat larger, reaching 11.6% for the t-ratio for  $\sigma_\alpha^2$  when  $R = 50$ . The size distortion seems to be due both to some estimator bias as well as standard errors that are a bit too small. Size distortions for all parameters disappear in the larger samples. In Table 2 where  $x_1 = x_2$  the CMLE for  $\lambda$  is even more biased in small samples, and considerably more volatile than in the design in Table 1. The performance of the QMLE is not very different

from the case with  $x_1 \neq x_2$ . The standard deviation measured by  $\text{IQ}/1.35$  is somewhat larger than when  $x_1 \neq x_2$ , as are size distortions, confirming the intuition that this design is more difficult to identify.

Tables 3 and 4 differ from Tables 1 and 2 in that they consider the same designs but with larger group sizes, now drawn from the uniform distribution on the interval  $[13, 25]$ . In Table 3 we consider the case with  $x_1 \neq x_2$ . The QMLE remains roughly unbiased across all sample sizes. The robust standard deviation roughly doubles relative to the small group size case and the size properties for t-ratios of the parameters  $\lambda$  and  $\beta$  deteriorate in samples where  $R \leq 100$  with size reaching around 10% in some cases. Size remains well controlled in larger samples with  $R \geq 200$ . The size distortions for the variance parameters are not much affected by the larger class sizes. The CMLE is even more biased when  $R = 50$  but less biased for larger sample sizes compared to Tables 1 and 2. This is consistent with incidental parameter bias which is expected to decrease with increasing group size. In addition the CMLE now is significantly less precise. This is in line with results by Lee (2007). Table 4 contains results for the case  $x_1 = x_2$  and large group sizes. The QMLE remains largely unbiased across all sample sizes but there is notable loss in estimator precision as measured by  $\text{IQ}/1.35$ , indicating the more challenging estimation environment. In line with theoretical predictions, estimator precision increases monotonically with sample size. Size distortions are now pronounced with empirical size reaching more than 20% in the smaller samples. The CMLE controls size well across all four designs. This comes at the cost of much less precisely and sometimes more biased estimated parameters.

Table 5 explores the effects that variation in group size has on both estimators. The case with  $\mathcal{U}\{3, 5\}$  maintains the same mean group size as in Table 2 but reduces the group size variance. We only report results for  $\lambda$ . The bias of the QMLE is not affected while the CMLE is somewhat less biased. The variance of both estimators increases. For the QMLE size distortions are somewhat larger than in Table 1. The design with  $\mathcal{U}\{4, 8\}$  increases the mean while leaving the variance of class sizes unchanged relative to Table 2. Overall, the results for this case are quite similar to the scenario with  $\mathcal{U}\{3, 5\}$ . The designs with  $\mathcal{U}\{8, 30\}$  and  $\mathcal{U}\{10, 22\}$  both improve identification relative to the design in Table 4. For the QMLE this results in unchanged good bias properties except when  $R = 50$  where we now see a small amount of bias, somewhat lower variance and slightly improved size properties. For the CMLE bias increases while variance somewhat improves relative to the results in Table 4 and the size properties remain similar. The larger bias for the CMLE may be related to a larger fraction of smaller classes in both designs. Smaller group sizes tend to amplify incidental parameter bias.

Table 6 explores the effects that non-Gaussian error distributions have on the estimators. For the Skew Normal distribution we see little difference to the results in Table 2 both for the QMLE and the CMLE estimator. The QMLE is also robust to the second design which uses a t-distribution with 6 degrees of freedom. The CMLE is more sensitive to this fat-tailed distribution. It is somewhat more biased and has higher variance compared to the Gaussian case. In addition, we now observe size distortions for the t-ratio related to the parameter  $\lambda$ . These size distortions don't disappear in

larger samples and seem to be due to the fact that the standard errors show a significant downward bias. This is most likely due to the fact that Lee (2007) bases standard errors on Gaussian error distributions. The final set of results we discuss are in Table 7 where we examine the effects of heteroscedasticity on the QMLE. We do not report results for the CMLE since this estimator was designed for the homoscedastic case only. The first set of results are based on a design where class size varies according to a  $\mathcal{U}\{2, 6\}$  distribution and where we maintain  $x_1 = x_2$ . Compared to a homoscedastic design the QMLE is somewhat less variable with no change in bias. The size properties of the t-ratio are overall comparable between the two cases, with slightly smaller size distortions in the heteroscedastic case when  $R = 50$ . We also consider a scenario where group size is fixed at  $m = 4$  while the type specific variances vary. While the QMLE continues to be nearly unbiased it has a higher variance. The size properties of t-ratios are slightly worse than in the homoscedastic case. For larger sample sizes both standard errors and t-ratios are well behaved.

## 6 Conclusion

In this paper, we show that moment conditions underlying the conditional variance method of Graham (2008) can be related to and motivated from a general class of linear peer effects models with random group effects. When augmented with group specific covariates our specification of the peer effects model is appropriate for settings where people are randomly assigned to groups or where group level heterogeneity is credibly controlled for with observed group level characteristics. We show that the quasi maximum likelihood estimator (QMLE) related to a linear Gaussian specification, as well as Graham’s estimator and the fixed effects estimator of Lee (2007) are contained in the class of GMM estimators we consider. Under Gaussian error assumptions the QMLE is the most efficient estimator in this class. We study conditions of identification, extending results in Graham (2008) and Lee (2007) for a simple model without covariates and a general model with covariates estimated by QML. We also establish that our QMLE is asymptotically normal and we construct consistent standard error formulas. Monte Carlo results show that our QML estimator has good small sample properties.



## References

- Angrist, J.D., 2014. The perils of peer effects. *Labour Economics* 30, 98–108.
- Anselin, L., 1988. *Spatial Econometrics: Methods and Models*. volume 4. Springer Science & Business Media.
- Anselin, L., 2010. Thirty years of spatial econometrics. *Papers in Regional Science* 89, 3–25.
- Booij, A.S., Leuven, E., Oosterbeek, H., 2017. Ability Peer Effects in University: Evidence from a Randomized Experiment. *The Review of Economic Studies* 84, 547–578.
- Boucher, V., Bramoullé, Y., Djebbari, H., Fortin, B., 2014. Do Peers Affect Student Achievement? Evidence from Canada Using Group Size Variation. *Journal of Applied Econometrics* 29, 91–109.
- Bramoullé, Y., Djebbari, H., Fortin, B., 2009. Identification of peer effects through social networks. *Journal of Econometrics* 150, 41–55.
- Cai, J., Szeidl, A., 2018. Interfirm Relationships and Business Performance\*. *The Quarterly Journal of Economics* 133, 1229–1282.
- Carrell, S.E., Fullerton, R.L., West, J.E., 2009. Does Your Cohort Matter? Measuring Peer Effects in College Achievement. *Journal of Labor Economics* 27, 439–464.
- Carrell, S.E., Sacerdote, B.I., West, J.E., 2013. From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation. *Econometrica* 81, 855–882.
- Chamberlain, G., 1980. Analysis of Covariance with Qualitative Data. *The Review of Economic Studies* 47, 225–238.
- Chetty, R., Friedman, J.N., Hilger, N., Saez, E., Schanzenbach, D.W., Yagan, D., 2011. How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star. *The Quarterly Journal of Economics* 126, 1593–1660.
- Chung, K.L., 2001. *A Course in Probability Theory*. Academic press.
- Cliff, A.D., Ord, J.K., 1973. *Spatial Autocorrelation*. volume 5. Pion London.
- Cliff, A.D., Ord, J.K., 1981. *Spatial Processes: Models & Applications*. volume 44. Pion London.
- Dhrymes, P.J., 1978. *Mathematics for Econometrics*. Technical Report. Springer.
- Duflo, E., Dupas, P., Kremer, M., 2011. Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *The American Economic Review* 101, 1739–1774.

- Duflo, E., Saez, E., 2003. The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment\*. *The Quarterly Journal of Economics* 118, 815–842.
- Dufour, J.M., 1997. Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models. *Econometrica* 65, 1365–1387.
- Fafchamps, M., Quinn, S., 2018. Networks and Manufacturing Firms in Africa: Results from a Randomized Field Experiment. *The World Bank Economic Review* 32, 656–675.
- Frijters, P., Islam, A., Pakrashi, D., 2019. Heterogeneity in peer effects in random dormitory assignment in a developing country. *Journal of Economic Behavior & Organization* 163, 117–134.
- Garlick, R., 2018. Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment. *American Economic Journal: Applied Economics* 10, 345–369.
- Goldsmith-Pinkham, P., Imbens, G.W., 2013. Social Networks and the Identification of Peer Effects. *Journal of Business & Economic Statistics* 31, 253–264.
- Graham, B.S., 2008. Identifying Social Interactions Through Conditional Variance Restrictions. *Econometrica* 76, 643–660.
- Guryan, J., Kroft, K., Notowidigdo, M.J., 2009. Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments. *American Economic Journal: Applied Economics* 1, 34–68.
- Johnson, C.R., Horn, R.A., 1985. *Matrix Analysis*. Cambridge university press Cambridge.
- Kang, C., 2007. Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics* 61, 458–495.
- Kapoor, M., Kelejian, H.H., Prucha, I.R., 2007. Panel data models with spatially correlated error components. *Journal of Econometrics* 140, 97–130.
- Kelejian, H.H., Prucha, I.R., 1998. 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, 99–121.
- Kelejian, H.H., Prucha, I.R., 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40, 509–533.
- Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219–257.

- Kelejian, H.H., Prucha, I.R., 2002. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics* 32, 691–707.
- Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157, 53–67.
- Kelejian, H.H., Prucha, I.R., Yuzefovich, Y., 2006. Estimation Problems in Models with Spatial Weighting Matrices Which Have Blocks of Equal Elements\*. *Journal of Regional Science* 46, 507–515.
- Kuersteiner, G.M., Prucha, I.R., 2020. Dynamic Spatial Panel Models: Networks, Common Shocks, and Sequential Exogeneity. *Econometrica* 88, 2109–2146.
- Lee, L.F., 2007. Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140, 333–374.
- Lee, L.F., Liu, X., Lin, X., 2010. Specification and estimation of social interaction models with network structures. *The Econometrics Journal* 13, 145–176.
- Lin, X., 2010. Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables. *Journal of Labor Economics* 28, 825–860.
- Liu, X., Lee, L.f., 2010. GMM estimation of social interaction models with centrality. *Journal of Econometrics* 159, 99–115.
- Liu, X., Patacchini, E., Zenou, Y., 2014. Endogenous peer effects: Local aggregate or local average? *Journal of Economic Behavior & Organization* 103, 39–59.
- Manski, C.F., 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* 60, 531–542.
- Mundlak, Y., 1978. On the Pooling of Time Series and Cross Section Data. *Econometrica* 46, 69–85.
- Newey, W.K., 1991. Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica* 59, 1161–1167.
- Neyman, J., Scott, E.L., 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica* 16, 1–32.
- Nye, B., Konstantopoulos, S., Hedges, L.V., 2004. How Large Are Teacher Effects? *Educational Evaluation and Policy Analysis* 26, 237–257.
- Ord, K., 1975. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association* 70, 120–126.

- Pötscher, B.M., Prucha, I.R., 1991. Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part i: Consistency and approximation concepts. *Econometric Reviews* 10, 125–216.
- Pötscher, B.M., Prucha, I.R., 1994. Generic uniform convergence and equicontinuity concepts for random functions. *Journal of Econometrics* 60, 23–63.
- Rivkin, S.G., Hanushek, E.A., Kain, J.F., 2005. Teachers, Schools, and Academic Achievement. *Econometrica* 73, 417–458.
- Sacerdote, B., 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics* 116, 681–704.
- Sojourner, A., 2013. Identification of Peer Effects with Missing Peer Data: Evidence from Project STAR. *The Economic Journal* 123, 574–605.
- Stinebrickner, R., Stinebrickner, T.R., 2006. What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics* 90, 1435–1454.
- Zimmerman, D.J., 2003. Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics* 85, 9–23.

# Appendix

## A Monte Carlo Simulation Results

Table 1: Simulation Results:  $m_r \sim \mathcal{U}\{2, 6\}$ , Homoscedastic Normal Errors,  $x_1 \neq x_2$

	QMLE							CMLE			
	$\lambda$	$\sigma_\alpha^2$	$\sigma_\epsilon^2$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\lambda$	$\sigma_\epsilon^2$	$\beta_2$	$\beta_3$
	0.500	0.250	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000
	<i>True value</i>										
	<i>50 groups, 200 observations</i>										
Median	0.500	0.214	0.979	0.999	0.998	0.998	1.001	0.541	1.004	1.018	1.025
Rob.Std.Dev.	(0.070)	(0.177)	(0.118)	(0.173)	(0.076)	(0.162)	(0.178)	(0.379)	(0.239)	(0.132)	(0.274)
Std.Dev.	[0.074]	[0.198]	[0.119]	[0.184]	[0.076]	[0.162]	[0.184]	[0.428]	[0.275]	[0.139]	[0.294]
Est.Std.Dev.	0.067	0.163	0.110	0.166	0.075	0.153	0.168	0.370		0.131	0.272
Rej.	0.070	0.116	0.095	0.074	0.050	0.063	0.064	0.036		0.039	0.041
	<i>100 groups, 400 observations</i>										
Median	0.498	0.233	0.990	1.002	0.998	0.997	0.999	0.524	1.007	1.008	1.012
Rob.Std.Dev.	(0.049)	(0.128)	(0.084)	(0.122)	(0.053)	(0.110)	(0.121)	(0.257)	(0.162)	(0.091)	(0.190)
Std.Dev.	[0.050]	[0.131]	[0.085]	[0.126]	[0.053]	[0.109]	[0.125]	[0.280]	[0.176]	[0.095]	[0.195]
Est.Std.Dev.	0.048	0.120	0.081	0.119	0.053	0.109	0.120	0.256		0.091	0.191
Rej.	0.057	0.099	0.074	0.059	0.052	0.051	0.059	0.046		0.043	0.044
	<i>200 groups, 800 observations</i>										
Median	0.500	0.241	0.995	1.000	1.000	1.001	0.998	0.512	1.003	1.004	1.004
Rob.Std.Dev.	(0.035)	(0.089)	(0.060)	(0.085)	(0.037)	(0.080)	(0.087)	(0.182)	(0.116)	(0.066)	(0.134)
Std.Dev.	[0.035]	[0.092]	[0.060]	[0.087]	[0.038]	[0.078]	[0.087]	[0.188]	[0.119]	[0.066]	[0.138]
Est.Std.Dev.	0.034	0.087	0.059	0.085	0.038	0.077	0.085	0.178		0.064	0.134
Rej.	0.051	0.078	0.062	0.054	0.048	0.054	0.057	0.052		0.051	0.054
	<i>400 groups, 1600 observations</i>										
Median	0.500	0.245	0.998	1.000	1.000	0.999	1.001	0.504	1.001	1.002	1.001
Rob.Std.Dev.	(0.024)	(0.061)	(0.041)	(0.059)	(0.027)	(0.056)	(0.060)	(0.129)	(0.081)	(0.045)	(0.093)
Std.Dev.	[0.024]	[0.063]	[0.042]	[0.060]	[0.027]	[0.054]	[0.060]	[0.129]	[0.082]	[0.045]	[0.094]
Est.Std.Dev.	0.024	0.062	0.042	0.060	0.027	0.055	0.060	0.125		0.045	0.094
Rej.	0.049	0.061	0.054	0.048	0.050	0.046	0.051	0.046		0.047	0.049
	<i>800 groups, 3200 observations</i>										
Median	0.500	0.247	0.999	0.999	0.999	0.999	0.999	0.504	1.001	1.001	1.000
Rob.Std.Dev.	(0.017)	(0.045)	(0.031)	(0.043)	(0.019)	(0.038)	(0.043)	(0.090)	(0.056)	(0.032)	(0.066)
Std.Dev.	[0.017]	[0.046]	[0.030]	[0.043]	[0.019]	[0.039]	[0.043]	[0.088]	[0.056]	[0.031]	[0.067]
Est.Std.Dev.	0.017	0.044	0.030	0.043	0.019	0.039	0.043	0.089		0.032	0.066
Rej.	0.056	0.064	0.053	0.053	0.048	0.054	0.054	0.044		0.043	0.051
	<i>1600 groups, 6400 observations</i>										
Median	0.500	0.249	0.999	1.001	1.000	0.999	1.000	0.503	1.002	1.001	1.002
Rob.Std.Dev.	(0.012)	(0.032)	(0.021)	(0.031)	(0.013)	(0.027)	(0.030)	(0.063)	(0.040)	(0.022)	(0.046)
Std.Dev.	[0.012]	[0.032]	[0.021]	[0.031]	[0.013]	[0.027]	[0.031]	[0.063]	[0.041]	[0.023]	[0.047]
Est.Std.Dev.	0.012	0.032	0.021	0.030	0.013	0.027	0.030	0.063		0.022	0.047
Rej.	0.055	0.057	0.052	0.054	0.047	0.046	0.057	0.050		0.050	0.048

1. Median value, robust standard deviation (IQ/1.35), standard deviation, median of estimated standard deviation and mean rejection rate of the Wald test of our QMLE and Lee's CMLE across 5000 repetitions. The CMLE is based on the within-group variation hence  $\sigma_\alpha^2, \beta_1, \beta_4$  are not estimated. Also, Lee(2007) does not offer estimate of the variance for  $\sigma_\epsilon^2$ .

2. Data generating process is based on model (10):  $y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir}\beta_2 + \bar{x}_{2,(-i)r}\beta_3 + x_{3,r}\beta_4 + \alpha_r + \epsilon_{ir}$ , with the true parameter values given in the top panel of the table. Group size  $m_r$  is drawn from a discrete uniform distribution  $\mathcal{U}\{2, 6\}$ . Sample is generated by:  $x_{1,ir} \sim N(0, 1)$ ,  $x_{2,ir} \sim N(0, 1)$ , and  $\bar{x}_{2,(-i)r}$  is the leave out mean of  $x_{2,ir}$ ,  $x_{3,r} \sim N(0, 1)$ ,  $\alpha_r \sim N(0, 0.25)$ , and  $\epsilon_{ir} \sim N(0, 1)$ . All variables are independent of each other across  $i$  and  $r$ .

Table 2: Simulation Results:  $m_r \sim \mathcal{U}\{2, 6\}$ , Homoscedastic Normal Errors,  $x_1 = x_2$

	QMLE							CMLE			
	$\lambda$	$\sigma_\alpha^2$	$\sigma_\epsilon^2$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\lambda$	$\sigma_\epsilon^2$	$\beta_2$	$\beta_3$
	<i>True value</i>										
	0.500	0.250	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000
	<i>50 groups, 200 observations</i>										
Median	0.501	0.202	0.976	0.992	0.999	0.994	0.989	0.592	1.036	1.013	0.995
Rob.Std.Dev.	(0.114)	(0.241)	(0.130)	(0.251)	(0.122)	(0.338)	(0.246)	(0.514)	(0.313)	(0.198)	(0.455)
Std.Dev.	[0.114]	[0.275]	[0.129]	[0.250]	[0.122]	[0.336]	[0.253]	[0.632]	[0.417]	[0.222]	[0.533]
Est.Std.Dev.	0.106	0.230	0.116	0.236	0.115	0.313	0.235	0.488		0.193	0.444
Rej.	0.090	0.120	0.100	0.084	0.081	0.091	0.082	0.042		0.040	0.036
	<i>100 groups, 400 observations</i>										
Median	0.499	0.230	0.988	0.999	1.001	0.998	1.003	0.539	1.019	1.011	1.008
Rob.Std.Dev.	(0.081)	(0.179)	(0.090)	(0.179)	(0.084)	(0.243)	(0.180)	(0.335)	(0.197)	(0.136)	(0.304)
Std.Dev.	[0.080]	[0.190]	[0.092]	[0.177]	[0.085]	[0.239]	[0.178]	[0.375]	[0.237]	[0.141]	[0.326]
Est.Std.Dev.	0.077	0.172	0.085	0.170	0.083	0.227	0.170	0.327		0.133	0.303
Rej.	0.077	0.104	0.078	0.079	0.070	0.077	0.074	0.040		0.044	0.044
	<i>200 groups, 800 observations</i>										
Median	0.501	0.239	0.995	0.999	0.999	0.996	0.995	0.528	1.013	1.005	1.000
Rob.Std.Dev.	(0.055)	(0.128)	(0.065)	(0.119)	(0.060)	(0.164)	(0.119)	(0.231)	(0.142)	(0.092)	(0.206)
Std.Dev.	[0.055]	[0.129]	[0.064]	[0.122]	[0.059]	[0.163]	[0.120]	[0.241]	[0.147]	[0.094]	[0.213]
Est.Std.Dev.	0.055	0.126	0.062	0.121	0.059	0.163	0.121	0.229		0.093	0.208
Rej.	0.061	0.082	0.062	0.060	0.054	0.058	0.057	0.042		0.043	0.044
	<i>400 groups, 1600 observations</i>										
Median	0.500	0.243	0.999	0.998	1.000	0.998	0.998	0.509	1.005	1.002	1.000
Rob.Std.Dev.	(0.040)	(0.091)	(0.045)	(0.087)	(0.042)	(0.117)	(0.085)	(0.159)	(0.097)	(0.064)	(0.143)
Std.Dev.	[0.039]	[0.090]	[0.045]	[0.087]	[0.042]	[0.117]	[0.086]	[0.163]	[0.098]	[0.065]	[0.147]
Est.Std.Dev.	0.039	0.090	0.044	0.086	0.042	0.117	0.086	0.159		0.065	0.146
Rej.	0.055	0.068	0.058	0.058	0.048	0.055	0.055	0.046		0.046	0.049
	<i>800 groups, 3200 observations</i>										
Median	0.501	0.248	0.999	0.999	0.999	0.999	0.999	0.509	1.004	1.001	0.999
Rob.Std.Dev.	(0.027)	(0.062)	(0.031)	(0.060)	(0.030)	(0.080)	(0.060)	(0.111)	(0.069)	(0.047)	(0.103)
Std.Dev.	[0.028]	[0.064]	[0.031]	[0.061]	[0.030]	[0.083]	[0.061]	[0.115]	[0.070]	[0.046]	[0.104]
Est.Std.Dev.	0.028	0.064	0.032	0.061	0.030	0.083	0.061	0.113		0.046	0.103
Rej.	0.049	0.059	0.043	0.050	0.050	0.053	0.054	0.049		0.048	0.050
	<i>1600 groups, 6400 observations</i>										
Median	0.500	0.249	1.000	1.000	1.000	0.999	1.000	0.503	1.001	1.001	1.000
Rob.Std.Dev.	(0.020)	(0.047)	(0.023)	(0.043)	(0.021)	(0.060)	(0.043)	(0.078)	(0.047)	(0.032)	(0.073)
Std.Dev.	[0.020]	[0.046]	[0.023]	[0.043]	[0.021]	[0.059]	[0.043]	[0.078]	[0.048]	[0.032]	[0.072]
Est.Std.Dev.	0.020	0.046	0.022	0.043	0.021	0.059	0.043	0.079		0.032	0.072
Rej.	0.052	0.056	0.054	0.048	0.049	0.052	0.053	0.042		0.048	0.048

1. Median value, robust standard deviation (IQ/1.35), standard deviation, median of estimated standard deviation and mean rejection rate of the Wald test of our QMLE and Lee's CMLE across 5000 repetitions. The CMLE is based on the within-group variation hence  $\sigma_\alpha^2, \beta_1, \beta_4$  are not estimated. Also, Lee(2007) does not offer estimate of the variance for  $\sigma_\epsilon^2$ .

2. Data generating process is based on model (10):  $y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir} \beta_2 + \bar{x}_{2,(-i)r} \beta_3 + x_{3,r} \beta_4 + \alpha_r + \epsilon_{ir}$ , with the true parameter values given in the top panel of the table. Group size  $m_r$  is drawn from a discrete uniform distribution  $\mathcal{U}\{2, 6\}$ . Sample is generated by:  $x_{1,ir} \sim N(0, 1)$ ,  $x_{2,ir} = x_{1,ir}$ , and  $\bar{x}_{2,(-i)r}$  is the leave out mean of  $x_{2,ir}$ ,  $x_{3,r} \sim N(0, 1)$ ,  $\alpha_r \sim N(0, 0.25)$ , and  $\epsilon_{ir} \sim N(0, 1)$ . All variables are independent of each other across  $i$  and  $r$ .

Table 3: Simulation Results:  $m_r \sim \mathcal{U}\{13, 25\}$ , Homoscedastic Normal Errors,  $x_1 \neq x_2$

	QMLE							CMLE			
	$\lambda$	$\sigma_\alpha^2$	$\sigma_\epsilon^2$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\lambda$	$\sigma_\epsilon^2$	$\beta_2$	$\beta_3$
	<i>True value</i>										
	0.500	0.250	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000
	<i>50 groups, 950 observations</i>										
Median	0.500	0.224	0.994	0.999	0.998	0.992	0.999	0.599	1.007	1.005	0.991
Rob.Std.Dev.	(0.151)	(0.184)	(0.050)	(0.311)	(0.035)	(0.370)	(0.323)	(1.781)	(0.198)	(0.101)	(0.597)
Std.Dev.	[0.255]	[0.970]	[0.054]	[0.519]	[0.035]	[0.384]	[0.518]	[1.840]	[0.214]	[0.104]	[0.611]
Est.Std.Dev.	0.140	0.167	0.050	0.291	0.034	0.362	0.292	1.732		0.099	0.601
Rej.	0.097	0.166	0.054	0.096	0.054	0.080	0.094	0.044		0.046	0.046
	<i>100 groups, 1900 observations</i>										
Median	0.501	0.235	0.997	0.999	0.999	0.996	1.000	0.510	0.999	1.001	0.988
Rob.Std.Dev.	(0.105)	(0.132)	(0.036)	(0.219)	(0.024)	(0.258)	(0.217)	(1.216)	(0.136)	(0.071)	(0.422)
Std.Dev.	[0.121]	[0.185]	[0.036]	[0.250]	[0.024]	[0.265]	[0.249]	[1.264]	[0.143]	[0.073]	[0.427]
Est.Std.Dev.	0.101	0.125	0.035	0.210	0.024	0.256	0.210	1.206		0.069	0.422
Rej.	0.072	0.119	0.053	0.071	0.045	0.067	0.070	0.058		0.057	0.049
	<i>200 groups, 3800 observations</i>										
Median	0.500	0.243	0.999	0.999	1.000	0.994	1.000	0.517	1.001	1.002	0.997
Rob.Std.Dev.	(0.073)	(0.092)	(0.026)	(0.151)	(0.017)	(0.182)	(0.151)	(0.849)	(0.094)	(0.050)	(0.302)
Std.Dev.	[0.077]	[0.107]	[0.025]	[0.159]	[0.017]	[0.184]	[0.160]	[0.847]	[0.095]	[0.049]	[0.301]
Est.Std.Dev.	0.072	0.091	0.025	0.149	0.017	0.181	0.149	0.853		0.049	0.298
Rej.	0.060	0.084	0.049	0.056	0.051	0.060	0.057	0.046		0.048	0.050
	<i>400 groups, 7600 observations</i>										
Median	0.500	0.246	0.999	1.000	1.000	0.999	1.001	0.490	0.999	0.999	1.000
Rob.Std.Dev.	(0.052)	(0.068)	(0.018)	(0.109)	(0.012)	(0.127)	(0.112)	(0.606)	(0.068)	(0.034)	(0.211)
Std.Dev.	[0.054]	[0.071]	[0.018]	[0.111]	[0.012]	[0.130]	[0.111]	[0.605]	[0.068]	[0.035]	[0.212]
Est.Std.Dev.	0.051	0.065	0.018	0.106	0.012	0.128	0.106	0.600		0.034	0.210
Rej.	0.051	0.065	0.051	0.055	0.054	0.056	0.050	0.051		0.050	0.052
	<i>800 groups, 15200 observations</i>										
Median	0.499	0.249	1.000	1.001	1.000	1.000	1.001	0.499	0.999	1.000	1.003
Rob.Std.Dev.	(0.036)	(0.047)	(0.013)	(0.075)	(0.008)	(0.093)	(0.074)	(0.426)	(0.047)	(0.024)	(0.152)
Std.Dev.	[0.037]	[0.049]	[0.012]	[0.077]	[0.008]	[0.092]	[0.077]	[0.434]	[0.048]	[0.025]	[0.151]
Est.Std.Dev.	0.036	0.047	0.012	0.075	0.008	0.091	0.075	0.424		0.024	0.149
Rej.	0.056	0.062	0.050	0.054	0.052	0.052	0.055	0.054		0.054	0.055
	<i>1600 groups, 30400 observations</i>										
Median	0.500	0.249	1.000	1.001	1.000	1.000	1.000	0.500	1.001	1.000	1.002
Rob.Std.Dev.	(0.025)	(0.033)	(0.009)	(0.052)	(0.006)	(0.063)	(0.053)	(0.306)	(0.033)	(0.017)	(0.108)
Std.Dev.	[0.026]	[0.033]	[0.009]	[0.053]	[0.006]	[0.064]	[0.053]	[0.300]	[0.033]	[0.017]	[0.106]
Est.Std.Dev.	0.026	0.033	0.009	0.053	0.006	0.064	0.053	0.300		0.017	0.105
Rej.	0.051	0.054	0.054	0.055	0.051	0.049	0.049	0.047		0.043	0.050

1. Median value, robust standard deviation (IQ/1.35), standard deviation, median of estimated standard deviation and mean rejection rate of the Wald test of our QMLE and Lee's CMLE across 5000 repetitions. The CMLE is based on the within-group variation hence  $\sigma_\alpha^2, \beta_1, \beta_4$  are not estimated. Also, Lee(2007) does not offer estimate of the variance for  $\sigma_\epsilon^2$ .

2. Data generating process is based on model (10):  $y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir} \beta_2 + \bar{x}_{2,(-i)r} \beta_3 + x_{3,r} \beta_4 + \alpha_r + \epsilon_{ir}$ , with the true parameter values given in the top panel of the table. Group size  $m_r$  is drawn from a discrete uniform distribution  $\mathcal{U}\{13, 25\}$ . Sample is generated by:  $x_{1,ir} \sim N(0, 1)$ ,  $x_{2,ir} \sim N(0, 1)$ , and  $\bar{x}_{2,(-i)r}$  is the leave out mean of  $x_{2,ir}$ ,  $x_{3,r} \sim N(0, 1)$ ,  $\alpha_r \sim N(0, 0.25)$ , and  $\epsilon_{ir} \sim N(0, 1)$ . All variables are independent of each other across  $i$  and  $r$ .

Table 4: Simulation Results:  $m_r \sim \mathcal{U}\{13, 25\}$ , Homoscedastic Normal Errors,  $x_1 = x_2$

	QMLE							CMLE			
	$\lambda$	$\sigma_\alpha^2$	$\sigma_\epsilon^2$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\lambda$	$\sigma_\epsilon^2$	$\beta_2$	$\beta_3$
	<i>True value</i>										
	0.500	0.250	1.000	1.000	1.000	1.000	1.000	0.500	1.000	1.000	1.000
	<i>50 groups, 950 observations</i>										
Median	0.500	0.225	0.987	1.001	1.005	0.995	0.995	0.545	1.003	0.999	0.898
Rob.Std.Dev.	(0.553)	(0.788)	(0.073)	(1.121)	(0.091)	(2.123)	(1.109)	(2.141)	(0.241)	(0.163)	(3.344)
Std.Dev.	[0.592]	[1.868]	[0.078]	[1.198]	[0.093]	[2.076]	[1.194]	[2.301]	[0.272]	[0.167]	[3.482]
Est.Std.Dev.	0.634	0.836	0.079	1.251	0.109	2.517	1.265	2.120		0.162	3.336
Rej.	0.230	0.238	0.066	0.231	0.114	0.242	0.229	0.045		0.045	0.043
	<i>100 groups, 1900 observations</i>										
Median	0.486	0.253	0.991	1.032	1.004	1.055	1.025	0.490	0.996	1.001	1.036
Rob.Std.Dev.	(0.439)	(0.625)	(0.055)	(0.888)	(0.070)	(1.669)	(0.870)	(1.463)	(0.160)	(0.112)	(2.286)
Std.Dev.	[0.422]	[0.931]	[0.056]	[0.849]	[0.070]	[1.566]	[0.851]	[1.512]	[0.171]	[0.116]	[2.366]
Est.Std.Dev.	0.473	0.599	0.059	0.940	0.082	1.883	0.931	1.475		0.113	2.321
Rej.	0.221	0.231	0.067	0.218	0.129	0.224	0.220	0.049		0.052	0.049
	<i>200 groups, 3800 observations</i>										
Median	0.500	0.241	0.996	0.994	1.001	0.994	0.999	0.515	1.000	0.999	0.947
Rob.Std.Dev.	(0.320)	(0.414)	(0.040)	(0.636)	(0.052)	(1.218)	(0.639)	(1.062)	(0.119)	(0.079)	(1.659)
Std.Dev.	[0.308]	[0.572]	[0.040]	[0.619]	[0.051]	[1.153]	[0.620]	[1.055]	[0.118]	[0.080]	[1.643]
Est.Std.Dev.	0.340	0.416	0.043	0.677	0.059	1.341	0.678	1.044		0.079	1.638
Rej.	0.195	0.208	0.066	0.190	0.132	0.189	0.189	0.046		0.052	0.047
	<i>400 groups, 7600 observations</i>										
Median	0.495	0.252	0.997	1.011	1.002	1.019	1.009	0.480	0.998	1.000	1.024
Rob.Std.Dev.	(0.248)	(0.320)	(0.031)	(0.495)	(0.041)	(0.939)	(0.498)	(0.712)	(0.078)	(0.057)	(1.154)
Std.Dev.	[0.236]	[0.381]	[0.030]	[0.474]	[0.039]	[0.889]	[0.473]	[0.735]	[0.081]	[0.057]	[1.178]
Est.Std.Dev.	0.249	0.307	0.031	0.494	0.043	0.968	0.497	0.735		0.056	1.153
Rej.	0.171	0.181	0.073	0.169	0.134	0.172	0.170	0.048		0.056	0.052
	<i>800 groups, 15200 observations</i>										
Median	0.498	0.250	0.998	1.004	1.000	1.010	1.004	0.505	1.000	1.000	0.992
Rob.Std.Dev.	(0.175)	(0.219)	(0.022)	(0.351)	(0.030)	(0.672)	(0.351)	(0.517)	(0.058)	(0.040)	(0.821)
Std.Dev.	[0.168]	[0.242]	[0.022]	[0.336]	[0.029]	[0.641]	[0.337]	[0.523]	[0.058]	[0.040]	[0.826]
Est.Std.Dev.	0.178	0.218	0.022	0.356	0.030	0.684	0.356	0.520		0.040	0.817
Rej.	0.129	0.144	0.073	0.130	0.113	0.132	0.127	0.052		0.051	0.052
	<i>1600 groups, 30400 observations</i>										
Median	0.497	0.253	0.999	1.005	1.001	1.007	1.007	0.498	1.000	1.001	1.004
Rob.Std.Dev.	(0.127)	(0.158)	(0.016)	(0.254)	(0.021)	(0.484)	(0.253)	(0.369)	(0.041)	(0.028)	(0.586)
Std.Dev.	[0.123]	[0.165]	[0.016]	[0.246]	[0.021]	[0.468]	[0.246]	[0.368]	[0.041]	[0.028]	[0.577]
Est.Std.Dev.	0.126	0.155	0.016	0.253	0.021	0.484	0.253	0.368		0.028	0.577
Rej.	0.111	0.120	0.076	0.110	0.093	0.108	0.111	0.047		0.050	0.050

1. Median value, robust standard deviation (IQ/1.35), standard deviation, median of estimated standard deviation and mean rejection rate of the Wald test of our QMLE and Lee's CMLE across 5000 repetitions. The CMLE is based on the within-group variation hence  $\sigma_\alpha^2, \beta_1, \beta_4$  are not estimated. Also, Lee(2007) does not offer estimate of the variance for  $\sigma_\epsilon^2$ .

2. Data generating process is based on model (10):  $y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir} \beta_2 + \bar{x}_{2,(-i)r} \beta_3 + x_{3,r} \beta_4 + \alpha_r + \epsilon_{ir}$ , with the true parameter values given in the top panel of the table. Group size  $m_r$  is drawn from a discrete uniform distribution  $\mathcal{U}\{13, 25\}$ . Sample is generated by:  $x_{1,ir} \sim N(0, 1)$ ,  $x_{2,ir} = x_{1,ir}$ , and  $\bar{x}_{2,(-i)r}$  is the leave out mean of  $x_{2,ir}$ ,  $x_{3,r} \sim N(0, 1)$ ,  $\alpha_r \sim N(0, 0.25)$ , and  $\epsilon_{ir} \sim N(0, 1)$ . All variables are independent of each other across  $i$  and  $r$ .



Table 5: Simulation Results for  $\lambda$  : Alternative Group Size Distributions, Homoscedastic Normal Errors,  $x_1 = x_2$

	$m_r \sim \mathcal{U}\{3, 5\}$		$m_r \sim \mathcal{U}\{4, 8\}$		$m_r \sim \mathcal{U}\{8, 30\}$		$m_r \sim \mathcal{U}\{10, 22\}$	
	QMLE	CMLE	QMLE	CMLE	QMLE	CMLE	QMLE	CMLE
	$\lambda$	$\lambda$	$\lambda$	$\lambda$	$\lambda$	$\lambda$	$\lambda$	$\lambda$
	<i>True value</i>							
	0.500	0.500	0.500	0.500	0.500	0.500	0.500	0.500
	<i>50 Groups</i>							
Median	0.502	0.561	0.494	0.575	0.487	0.545	0.481	0.567
Rob.Std.Dev.	(0.222)	(0.916)	(0.257)	(0.993)	(0.331)	(1.090)	(0.457)	(1.682)
Std.Dev.	[0.201]	[1.580]	[0.241]	[1.193]	[0.340]	[1.136]	[0.451]	[1.777]
Est.Std.Dev.	0.205	0.887	0.243	0.953	0.330	1.088	0.483	1.603
Rej.	0.163	0.064	0.175	0.058	0.175	0.040	0.209	0.051
	<i>100 Groups</i>							
Median	0.500	0.542	0.497	0.549	0.493	0.531	0.496	0.550
Rob.Std.Dev.	(0.164)	(0.629)	(0.184)	(0.667)	(0.243)	(0.758)	(0.329)	(1.140)
Std.Dev.	[0.151]	[0.734]	[0.175]	[0.736]	[0.237]	[0.785]	[0.325]	[1.184]
Est.Std.Dev.	0.152	0.614	0.179	0.659	0.242	0.761	0.348	1.129
Rej.	0.140	0.054	0.145	0.049	0.151	0.050	0.188	0.046
	<i>200 Groups</i>							
Median	0.502	0.534	0.503	0.518	0.498	0.534	0.500	0.537
Rob.Std.Dev.	(0.110)	(0.437)	(0.127)	(0.459)	(0.171)	(0.529)	(0.249)	(0.812)
Std.Dev.	[0.108]	[0.464]	[0.127]	[0.481]	[0.169]	[0.534]	[0.240]	[0.811]
Est.Std.Dev.	0.110	0.430	0.129	0.460	0.173	0.538	0.254	0.793
Rej.	0.113	0.046	0.119	0.049	0.121	0.047	0.168	0.047
	<i>400 Groups</i>							
Median	0.497	0.507	0.500	0.517	0.499	0.511	0.497	0.501
Rob.Std.Dev.	(0.081)	(0.298)	(0.095)	(0.333)	(0.124)	(0.383)	(0.183)	(0.552)
Std.Dev.	[0.079]	[0.312]	[0.093]	[0.332]	[0.124]	[0.381]	[0.173]	[0.561]
Est.Std.Dev.	0.080	0.299	0.093	0.325	0.125	0.379	0.185	0.557
Rej.	0.085	0.046	0.093	0.050	0.105	0.046	0.135	0.049
	<i>800 Groups</i>							
Median	0.501	0.508	0.500	0.505	0.498	0.501	0.501	0.499
Rob.Std.Dev.	(0.058)	(0.211)	(0.067)	(0.229)	(0.088)	(0.265)	(0.129)	(0.388)
Std.Dev.	[0.057]	[0.216]	[0.065]	[0.232]	[0.087]	[0.266]	[0.127]	[0.388]
Est.Std.Dev.	0.057	0.211	0.066	0.228	0.089	0.267	0.130	0.394
Rej.	0.067	0.051	0.072	0.049	0.080	0.047	0.107	0.046
	<i>1600 Groups</i>							
Median	0.499	0.506	0.500	0.499	0.501	0.504	0.499	0.497
Rob.Std.Dev.	(0.040)	(0.150)	(0.046)	(0.161)	(0.062)	(0.188)	(0.093)	(0.274)
Std.Dev.	[0.040]	[0.152]	[0.047]	[0.163]	[0.062]	[0.187]	[0.091]	[0.278]
Est.Std.Dev.	0.040	0.149	0.047	0.161	0.063	0.189	0.093	0.278
Rej.	0.059	0.048	0.061	0.052	0.066	0.044	0.077	0.050

1. Median value, robust standard deviation (IQ/1.35), standard deviation, median of estimated standard deviation and mean rejection rate of the Wald test of our QMLE and Lee's CMLE across 5000 repetitions. For simplicity, we only present estimates of the endogeneous peer effects ( $\lambda$ ).
2. Data generating process is based on model (10):  $y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir} \beta_2 + \bar{x}_{2,(-i)r} \beta_3 + x_{3,r} \beta_4 + \alpha_r + \epsilon_{ir}$ , with the  $\lambda = 0.5$  and all  $\beta$  s being 1. Sample is generated by:  $x_{1,ir} \sim N(0, 1)$ ,  $x_{2,ir} = x_{1,ir}$  and  $\bar{x}_{2,(-i)r}$  is the leave out mean of  $x_{2,ir}$ ,  $x_{3,r} \sim N(0, 1)$ ,  $\alpha_r \sim N(0, 0.25)$ , and  $\epsilon_{ir} \sim N(0, 1)$ . All variables are independent of each other across  $i$  and  $r$ .
3. Group size  $m_r$  is drawn from  $\mathcal{U}\{3, 5\}$  (Case 1),  $\mathcal{U}\{4, 8\}$  (Case 2),  $\mathcal{U}\{8, 30\}$  (Case 3),  $\mathcal{U}\{10, 22\}$  (Case 4).

Table 6: Simulation Results: Homoscedastic but Nonnormal Errors,  $m_r \sim \mathcal{U}\{2, 6\}$ ,  $x_1 = x_2$

	Skew Normal				Student Distribution			
	QMLE		CMLE		QMLE		CMLE	
	$\lambda$	$\beta_3$	$\lambda$	$\beta_3$	$\lambda$	$\beta_3$	$\lambda$	$\beta_3$
	<i>True value</i>							
	0.500	1.000	0.500	1.000	0.500	1.000	0.500	1.000
	<i>50 Groups</i>							
Median	0.500	0.993	0.580	0.991	0.499	1.004	0.610	1.010
Rob.Std.Dev.	(0.119)	(0.345)	(0.525)	(0.443)	(0.106)	(0.303)	(0.633)	(0.386)
Std.Dev.	[0.114]	[0.339]	[0.634]	[0.516]	[0.109]	[0.315]	[0.862]	[0.470]
Est.Std.Dev.	0.107	0.314	0.483	0.446	0.099	0.288	0.487	0.390
Rej.	0.091	0.089	0.047	0.033	0.084	0.084	0.078	0.034
	<i>100 Groups</i>							
Median	0.501	1.002	0.539	0.993	0.500	1.001	0.553	1.007
Rob.Std.Dev.	(0.080)	(0.241)	(0.346)	(0.309)	(0.074)	(0.213)	(0.422)	(0.268)
Std.Dev.	[0.080]	[0.236]	[0.380]	[0.325]	[0.075]	[0.219]	[0.468]	[0.284]
Est.Std.Dev.	0.077	0.228	0.326	0.301	0.070	0.207	0.328	0.262
Rej.	0.082	0.070	0.046	0.043	0.070	0.071	0.085	0.042
	<i>200 Groups</i>							
Median	0.501	1.000	0.520	0.998	0.499	1.001	0.534	1.002
Rob.Std.Dev.	(0.057)	(0.169)	(0.235)	(0.209)	(0.051)	(0.152)	(0.295)	(0.186)
Std.Dev.	[0.058]	[0.171]	[0.254]	[0.217]	[0.052]	[0.154]	[0.313]	[0.189]
Est.Std.Dev.	0.055	0.164	0.228	0.208	0.051	0.149	0.229	0.182
Rej.	0.074	0.068	0.058	0.046	0.065	0.062	0.121	0.047
	<i>400 Groups</i>							
Median	0.500	1.000	0.515	0.999	0.501	0.998	0.516	0.998
Rob.Std.Dev.	(0.039)	(0.115)	(0.169)	(0.143)	(0.036)	(0.107)	(0.204)	(0.128)
Std.Dev.	[0.039]	[0.116]	[0.172]	[0.148]	[0.036]	[0.109]	[0.219]	[0.131]
Est.Std.Dev.	0.039	0.117	0.160	0.146	0.036	0.106	0.159	0.127
Rej.	0.054	0.053	0.059	0.047	0.056	0.056	0.131	0.052
	<i>800 Groups</i>							
Median	0.500	1.000	0.504	0.998	0.501	0.998	0.510	0.998
Rob.Std.Dev.	(0.028)	(0.086)	(0.116)	(0.104)	(0.026)	(0.076)	(0.152)	(0.091)
Std.Dev.	[0.028]	[0.084]	[0.119]	[0.104]	[0.026]	[0.077]	[0.155]	[0.093]
Est.Std.Dev.	0.028	0.083	0.112	0.102	0.026	0.075	0.112	0.089
Rej.	0.052	0.055	0.062	0.052	0.056	0.052	0.137	0.058
	<i>1600 Groups</i>							
Median	0.500	1.000	0.504	0.999	0.501	1.000	0.504	1.001
Rob.Std.Dev.	(0.019)	(0.057)	(0.084)	(0.071)	(0.018)	(0.052)	(0.103)	(0.064)
Std.Dev.	[0.019]	[0.058]	[0.085]	[0.071]	[0.018]	[0.054]	[0.106]	[0.064]
Est.Std.Dev.	0.020	0.059	0.079	0.072	0.018	0.054	0.079	0.063
Rej.	0.045	0.045	0.068	0.043	0.054	0.053	0.136	0.051

1. Median value, robust standard deviation (IQ/1.35), standard deviation, median of estimated standard deviation and mean rejection rate of the Wald test of our QMLE and Lee's CMLE across 5000 repetitions. For simplicity, we only present estimates of the endogeneous peer effects ( $\lambda$ ) and exogenous peer effects ( $\beta_3$ ).

2. Data generating process is based on model (10):  $y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir} \beta_2 + \bar{x}_{2,(-i)r} \beta_3 + x_{3,r} \beta_4 + \alpha_r + \epsilon_{ir}$ , with  $\lambda = 0.5$  and all  $\beta$  s being 1. Group size  $m_r$  is drawn from a discrete uniform distribution  $\mathcal{U}\{2, 6\}$ . Sample is generated by:  $x_{1,ir} \sim N(0, 1)$ ,  $x_{2,ir} = x_{1,ir}$ , and  $\bar{x}_{2,(-i)r}$  is the leave out mean of  $x_{2,ir}$ ,  $x_{3,r} \sim N(0, 1)$ . All variables are independent of each other across  $i$  and  $r$ .

3. In the case of Skew normal distribution, location is 0, scale is 1 and shape is  $0.9/\sqrt{1 - 0.9^2}$ . In the case of student distribution, degree of freedom is 6. In all cases,  $\alpha_r$  and  $\epsilon_{ir}$  are independently drawn from identical distribution and standardized to have mean 0 and variance 0.25, 1 respectively.

Table 7: Simulation Results: Heteroscedastic Normal Errors,  $x_1 = x_2$

	$m_r \sim \mathcal{U}\{2, 6\}$				$m_r = 4$ for all			
	$\sigma_\epsilon^2 \in \{0.5, 1.5\}$		$\sigma_\epsilon = 1$ for all		$\sigma_\epsilon^2 \in \{0.5, 1.5\}$		$\sigma_\epsilon^2 \in \{0.4, 0.8, 1.2, 1.6\}$	
	$\lambda$	$\beta_3$	$\lambda$	$\beta_3$	$\lambda$	$\beta_3$	$\lambda$	$\beta_3$
	<i>True value</i>							
	0.500	1.000	0.500	1.000	0.500	1.000	0.500	1.000
	<i>50 Groups</i>							
Median	0.502	0.994	0.504	0.990	0.485	1.047	0.499	1.003
Rob.Std.Dev.	(0.097)	(0.298)	(0.116)	(0.352)	(0.246)	(0.794)	(0.305)	(0.989)
Std.Dev.	[0.100]	[0.299]	[0.118]	[0.352]	[0.783]	[2.493]	[0.969]	[3.083]
Est.Std.Dev.	0.090	0.265	0.106	0.311	0.173	0.566	0.172	0.545
Rej.	0.079	0.085	0.093	0.098	0.104	0.103	0.144	0.141
	<i>100 Groups</i>							
Median	0.502	0.996	0.501	0.999	0.500	1.003	0.500	1.002
Rob.Std.Dev.	(0.067)	(0.196)	(0.080)	(0.239)	(0.147)	(0.479)	(0.182)	(0.591)
Std.Dev.	[0.067]	[0.200]	[0.079]	[0.235]	[0.479]	[1.509]	[0.639]	[2.012]
Est.Std.Dev.	0.064	0.191	0.077	0.227	0.125	0.410	0.130	0.422
Rej.	0.077	0.076	0.074	0.074	0.085	0.081	0.122	0.114
	<i>200 Groups</i>							
Median	0.500	1.004	0.500	0.998	0.496	1.011	0.498	1.007
Rob.Std.Dev.	(0.046)	(0.140)	(0.055)	(0.168)	(0.099)	(0.321)	(0.114)	(0.365)
Std.Dev.	[0.047]	[0.140]	[0.056]	[0.166]	[0.187]	[0.593]	[0.327]	[1.018]
Est.Std.Dev.	0.045	0.136	0.055	0.163	0.092	0.300	0.096	0.312
Rej.	0.068	0.065	0.068	0.067	0.070	0.067	0.093	0.086
	<i>400 Groups</i>							
Median	0.500	1.002	0.500	1.000	0.498	1.002	0.500	1.000
Rob.Std.Dev.	(0.033)	(0.098)	(0.040)	(0.122)	(0.068)	(0.221)	(0.072)	(0.236)
Std.Dev.	[0.032]	[0.097]	[0.040]	[0.119]	[0.088]	[0.283]	[0.128]	[0.405]
Est.Std.Dev.	0.032	0.097	0.039	0.117	0.066	0.214	0.069	0.222
Rej.	0.056	0.057	0.059	0.059	0.057	0.055	0.067	0.061
	<i>800 Groups</i>							
Median	0.500	0.999	0.500	1.000	0.500	1.002	0.500	1.001
Rob.Std.Dev.	(0.023)	(0.071)	(0.028)	(0.082)	(0.049)	(0.161)	(0.050)	(0.160)
Std.Dev.	[0.023]	[0.069]	[0.028]	[0.083]	[0.053]	[0.171]	[0.056]	[0.180]
Est.Std.Dev.	0.023	0.069	0.028	0.083	0.047	0.152	0.049	0.159
Rej.	0.055	0.050	0.059	0.054	0.050	0.050	0.056	0.055
	<i>1600 Groups</i>							
Median	0.500	1.001	0.500	1.001	0.499	1.003	0.500	0.998
Rob.Std.Dev.	(0.016)	(0.048)	(0.020)	(0.060)	(0.033)	(0.109)	(0.035)	(0.116)
Std.Dev.	[0.016]	[0.049]	[0.020]	[0.059]	[0.034]	[0.111]	[0.037]	[0.120]
Est.Std.Dev.	0.016	0.049	0.020	0.059	0.033	0.108	0.035	0.113
Rej.	0.053	0.053	0.055	0.059	0.046	0.044	0.050	0.045

1. Median value, robust standard deviation (IQ/1.35), standard deviation, median of estimated standard deviation and mean rejection rate of the Wald test of our QMLE across 5000 repetitions. For simplicity, we only present estimates of the endogeneous peer effects ( $\lambda$ ) and exogenous peer effects ( $\beta_3$ ).

2. Data generating process is based on model (10):  $y_{ir} = \beta_1 + \lambda \bar{y}_{(-i)r} + x_{1,ir} \beta_2 + \bar{x}_{2,(-i)r} \beta_3 + x_{3,r} \beta_4 + \alpha_r + \epsilon_{ir}$ , with  $\lambda = 0.5$  and all  $\beta$  s being 1. Sample is generated by:  $x_{1,ir} \sim N(0, 1)$ ,  $x_{2,ir} = x_{1,ir}$  and  $\bar{x}_{2,(-i)r}$  is the leave out mean of  $x_{2,ir}$ ,  $x_{3,r} \sim N(0, 1)$ ,  $\alpha_r \sim N(0, 0.25)$ . When there are more than one category of  $\sigma_\epsilon$ , groups are equally distributed into different categories. All variables are independent of each other across  $i$  and  $r$ .

3. In the first case (Columns 1-2), the model has both heteroscedasticity and group size variation. In the second case (Columns 3-4), the DGP has homoscedastic  $\sigma_\epsilon^2$  and group size variation. But the estimation process assumes two categories of  $\sigma_\epsilon^2$ . In both case 1 and case 2, group size  $m_r$  is drawn from  $\mathcal{U}\{2, 6\}$ . In Cases 3 and 4, group size is 4 for all.

## B Preliminaries

In proving consistency and asymptotic normality of the QMLE estimator we encounter linear quadratic forms of the form

$$S_N(\theta) = U' A_N(\theta) U + U' a_N(\theta) \quad (\text{B.1})$$

where  $A_N(\theta)$  is an  $N \times N$  non-stochastic matrix,  $a_N(\theta)$  is an  $N$ -dimensional non-stochastic column vector, and where  $A_N(\theta)$  and  $a_N(\theta)$  exhibit some special structures. In the following we describe that structure in more detail, and collect some basic lemmata used in proving the consistency and asymptotic normality of the QMLE.

We adopt the following notation: Partition an  $N \times N$  matrix  $A_N$  into  $R \times R$  submatrices, with the  $(r, r')$ -th submatrix being an  $m_r \times m_{r'}$  matrix,  $r, r' = 1, \dots, R$ . We then denote the  $(r, r')$ -th submatrix of  $A_N$  as  $A_{(r,r'),N}$ , and the  $(i, j)$ -th element of  $A_N$  as  $a_{ij,N}$ ,  $1 \leq i \leq N$ ,  $1 \leq j \leq N$ . Partition an  $N \times 1$  vector  $a_N$  into  $R$  subvectors, with the  $r$ -th subvector being an  $m_r \times 1$  vector. We then denote the  $r$ -th subvector of  $a_N$  as  $a_{(r),N}$  and the  $i$ -th element of  $a_N$  as  $a_{i,N}$ . In line with Kelejian and Prucha (2001), we call the column and row sums of an  $N \times N$  matrix  $A_N(\theta)$  uniformly bounded in absolute value if there exists some finite constant  $C$  (which does not depend on  $N$  or  $\theta$ ) such that

$$\sup_{\theta \in \Theta} \sum_{i=1}^N |a(\theta)_{ij,N}| \leq C, \quad \sup_{\theta \in \Theta} \sum_{j=1}^N |a(\theta)_{ij,N}| \leq C.$$

A corresponding definition applies to rectangular matrices. Of course, if the row sums of  $A_N(\theta)$  are uniformly bounded in absolute value, and the elements of  $a_N(\theta)$  are uniformly bounded in absolute value, then the elements of  $A_N(\theta)a_N(\theta)$  are uniformly bounded in absolute value. Note that if the row and column sums of  $A_N(\theta)$  and  $B_N(\theta)$  are uniformly bounded in absolute value, then  $A_N(\theta) + B_N(\theta)$  and  $A_N(\theta)B_N(\theta)$  (if dimension permits addition or multiplication) also have row and column sums uniformly bounded in absolute value.<sup>8</sup>

### B.1 Basic Properties of Matrices Forming the Log-Likelihood Function

Recall that  $\theta = (\theta_1, \dots, \theta_{J+2})$ , with  $\theta_1 = \lambda$ ,  $\theta_2 = \sigma_\alpha^2$ , and  $\theta_{j+2} = \sigma_{\epsilon,j}^2$  for  $j = 1, \dots, J$ , and that in light of Assumptions 1, 2, and 3 the parameter space  $\Theta$  is compact. An inspection of the expression of the log-likelihood function shows that it depends on the following set of matrices:  $I - \lambda W$ ,  $(I - \lambda W)^{-1}$ ,  $\Omega(\theta)$ ,  $\Omega(\theta)^{-1}$ ,  $W$ . For generic functions  $p(m_r, D_r, \theta)$  and  $s(m_r, D_r, \theta)$  all these matrices are symmetric block diagonal matrices of the form

$$A_N(\theta) = \text{diag}_{r=1}^R \{p(m_r, D_r, \theta) I_{m_r}^* + s(m_r, D_r, \theta) J_{m_r}^*\}. \quad (\text{B.2})$$

---

<sup>8</sup>This is readily seen by argumentation in line with Kelejian and Prucha (1999).

In particular, by replacing  $p(\cdot)$  and  $s(\cdot)$  with specific functions  $\phi_S, \phi_\Omega, \phi_W, \psi_S, \psi_\Omega$  and  $\psi_W$  defined below, one obtains

$$\begin{aligned}
I - \lambda W &= \text{diag}_{r=1}^R \{ \phi_S(m_r, \theta) I_{m_r}^* + \psi_S(m_r, \theta) J_{m_r}^* \} \\
(I - \lambda_0 W)^{-1} &= \text{diag}_{r=1}^R \{ \phi_S^{-1}(m_r, \theta_0) I_{m_r}^* + \psi_S^{-1}(m_r, \theta_0) J_{m_r}^* \}, \\
\Omega_0 &= \text{diag}_{r=1}^R \{ \phi_\Omega(m_r, D_r, \theta_0) I_{m_r}^* + \psi_\Omega(m_r, D_r, \theta_0) J_{m_r}^* \}, \\
\Omega(\theta)^{-1} &= \text{diag}_{r=1}^R \{ \phi_\Omega^{-1}(m_r, D_r, \theta) I_{m_r}^* + \psi_\Omega^{-1}(m_r, D_r, \theta) J_{m_r}^* \}, \\
W &= \text{diag}_{r=1}^R \{ \phi_W(m_r, \theta) I_{m_r}^* + \psi_W(m_r, \theta) J_{m_r}^* \},
\end{aligned} \tag{B.3}$$

where

$$\begin{aligned}
\phi_S(m_r, \theta) &= \frac{m_r - 1 + \lambda}{m_r - 1}, & \psi_S(m_r, \theta) &= 1 - \lambda, \\
\phi_\Omega(m_r, D_r, \theta) &= \sigma_{\epsilon, D_r}^2, & \psi_\Omega(m_r, D_r, \theta) &= \sigma_{\epsilon, D_r}^2 + m_r \sigma_\alpha^2. \\
\phi_W(m_r, \theta) &= -\frac{1}{m_r - 1}, & \psi_W(m_r, \theta) &= 1
\end{aligned} \tag{B.4}$$

It is readily seen that there exists an open bounded set  $\Theta_o$  such that  $\Theta \subset \Theta_o \subset (-1, 1) \times R^{J+1}$  such that the placeholder functions  $p(m_r, D_r, \theta)$  and  $s(m_r, D_r, \theta)$ , explicitly defined in (B.4), are continuously differentiable on  $\Theta_o$ . Thus, by Bolzano-Weierstrass' extreme value theorem there exists a positive constant  $C$ , which does not depend on  $\theta$ , such that

$$0 \leq |p(m_r, D_r, \theta)|, |s(m_r, D_r, \theta)|, |\partial p(m_r, D_r, \theta)/\partial \theta_i|, |\partial s(m_r, D_r, \theta)/\partial \theta_i| \leq C < \infty, \tag{B.5}$$

for all  $\theta \in \Theta$ .<sup>9</sup> This implies that  $p(m_r, D_r, \theta)$  and  $s(m_r, D_r, \theta)$  are both uniformly continuous on  $\Theta$ . Observing that  $\phi_\Omega(m_r, D_r, \theta)$  and  $\psi_\Omega(m_r, D_r, \theta)$  are positive on  $\Theta$  it follows further that there exists a positive constant  $c$ , which does not depend on  $\theta$ , such that

$$0 < c \leq \phi_\Omega(m_r, D_r, \theta), \psi_\Omega(m_r, D_r, \theta) \leq C < \infty. \tag{B.6}$$

Since  $I_{m_r}^*$  and  $J_{m_r}^*$  are orthogonal, the multiplication of block diagonal matrices, where the blocks are of the form  $p(m_r, D_r, \theta) I_{m_r}^* + s(m_r, D_r, \theta) J_{m_r}^*$ , yields a matrix with the same structure. Furthermore the multiplication of those matrices is commutative. More specifically, let

$$A_N(\theta) = \text{diag}_{r=1}^R \{ p(m_r, D_r, \theta) I_{m_r}^* + s(m_r, D_r, \theta) J_{m_r}^* \}$$

and  $\mathring{A}_N(\theta) = \text{diag}_{r=1}^R \{ \mathring{p}(m_r, D_r, \theta) I_{m_r}^* + \mathring{s}(m_r, D_r, \theta) J_{m_r}^* \}$ , then

$$A_N(\theta) \mathring{A}_N(\theta) = \text{diag}_{r=1}^R \{ p(m_r, D_r, \theta) \mathring{p}(m_r, D_r, \theta) I_{m_r}^* + s(m_r, D_r, \theta) \mathring{s}(m_r, D_r, \theta) J_{m_r}^* \}.$$

---

<sup>9</sup>Of course, since the  $m_r$  only takes on finitely many values, the constants can also be taken such that they do not depend on  $m_r$ . We note, although not stated explicitly, all subsequent uniformity results also hold uniformly for  $m_r \in \{m : 2 \leq m \leq \bar{M}\}$ .

In addition,  $A_N(\theta)$  and  $\mathring{A}_N(\theta)$  commute,  $A_N(\theta)\mathring{A}_N(\theta) = \mathring{A}_N(\theta)A_N(\theta)$ . Also,

$$|A_N(\theta)| = \prod_{j=1}^J \prod_{m=2}^{\bar{M}} |p(m, j, \theta)I_m^* + s(m, j, \theta)J_m^*|^{R_{m,j}} = \prod_{j=1}^J \prod_{m=2}^{\bar{M}} [p(m, j, \theta)^{m-1}s(m, j, \theta)]^{R_{m,j}}, \quad (\text{B.7})$$

as is readily checked observing that  $pI_m^* + sJ_m^* = p\{I_m + [(s-p)/(pm)]\iota_m\iota_m'\}$  and applying Proposition 31 in Dhrymes (1978), Section 2.7 on p. 38, to compute the determinant of the matrix in curly brackets. Furthermore,

$$\text{tr}(A_N(\theta)) = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} R_{m,j} ((m-1)p(m, j, \theta) + s(m, j, \theta)).$$

and

$$\begin{aligned} \frac{1}{N}Z'A_N(\theta)Z &= \frac{1}{N} \sum_{r=1}^R (p(m_r, D_r, \theta)Z_r'I_{m_r}^*Z_r + s(m_r, D_r, \theta)Z_r'J_{m_r}^*Z_r) \\ &= \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \left( p(m, j, \theta) \frac{1}{N} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r'\ddot{Z}_r + s(m, j, \theta) \frac{1}{N} \sum_{r \in \mathcal{I}_{m,j}} m\ddot{z}'_r\ddot{z}_r \right). \end{aligned} \quad (\text{B.8})$$

We note that the row and column sums of any matrix  $A_N(\theta)$  of the form (B.2) are uniformly bounded in absolute value, if  $p(m_r, D_r, \theta)$  and  $s(m_r, D_r, \theta)$  are uniformly bounded in absolute value (observing that  $m_r$  is bounded by Assumption 4). We note further that in light of Assumption 5 the elements of  $N^{-1}Z'A_N(\theta)Z$  are uniformly bounded in absolute value. If additionally  $p(m_r, D_r, \theta)$  and  $s(m_r, D_r, \theta)$  are positive and bounded away from zero, then also the elements of  $(N^{-1}Z'A_N(\theta)Z)^{-1}$  are uniformly bounded; see Lemma B.5. Consequently the elements of  $(N^{-1}Z'\Omega(\theta)Z)^{-1}$  are uniformly bounded, and the row and column sums of  $Z(Z'\Omega(\theta)Z)^{-1}Z' = N^{-1}Z(N^{-1}Z'\Omega(\theta)Z)^{-1}Z'$ ,  $\Omega(\theta)^{-1}Z(N^{-1}Z'\Omega(\theta)Z)^{-1}Z'\Omega(\theta)^{-1}$  and  $M_Z(\theta)$  are uniformly bounded in absolute value. As a result,  $M_Z(\theta)$  and  $\partial M_Z(\theta)/\partial\theta_i = -M_Z(\theta)(\partial\Omega(\theta)/\partial\theta_i)M_Z(\theta)$  have row and column sums uniformly bounded in absolute value.

In all, if a matrix  $A_N(\theta)$  is the product of  $I - \lambda W$ ,  $(I - \lambda W)^{-1}$ ,  $\Omega(\theta)$ ,  $\Omega(\theta)^{-1}$ ,  $W$ ,  $\partial\Omega(\theta)/\partial\theta_i$ , and  $M_Z(\theta)$ , then both  $A_N(\theta)$  and  $\partial A_N(\theta)/\partial\theta_i$  have row and column sums uniformly bounded in absolute value, and the elements of  $A_N(\theta)Z\beta_0$  are uniformly bounded in absolute value over  $\theta \in \Theta$  and  $N$ .

## B.2 Limit Theorems for Linear Quadratic Forms in $U$

The following result follows trivially from Lemma A.1 in Kelejian and Prucha (2010), and is only given for the convenience of the reader.

**Lemma B.1.** *[Mean and Covariance] Let  $A$  and  $B$  be  $N \times N$  nonstochastic symmetric matrices, which are partitioned into  $R^2$  submatrices and let  $a$  and  $b$  be  $N \times 1$  vectors, which are conformably partitioned into  $R$  subvectors. Let  $a_{ij}$  and  $b_{ij}$  denote the  $(i, j)$ -th element of  $A$  and  $B$ , let  $A_{(r,r')}$  and*

$B_{(r,r')}$  denote the  $(r, r')$ -th block of dimension  $m_r \times m_{r'}$ , let  $\text{vec}_D(A_{(r,r)})$  and  $\text{vec}_D(B_{(r,r)})$  denote the column vectors of the diagonal elements of  $A_{(r,r)}$  and  $B_{(r,r)}$ , let  $a_i$  and  $b_i$  denote the  $i$ -th element of  $a$  and  $b$ , and  $a_{(r)}$  and  $b_{(r)}$  denote the  $r$ -th subvectors of dimension  $m_r \times 1$ . Then, under Assumptions 1 and 2,

$$\begin{aligned} E(U'AU + U'a) &= \text{tr}(\Omega_0 A), \\ \text{Cov}(U'AU + U'a, U'BU + U'b) &= 2 \text{tr}(A\Omega_0 B\Omega_0) + a'\Omega_0 b \\ &+ \sum_{r=1}^R \text{vec}_D(A_{(r,r)})' \text{vec}_D(B_{(r,r)}) (\mu_{\epsilon_0, D_r}^{(4)} - 3\sigma_{\epsilon_0, D_r}^4) + \sum_{r=1}^R \left[ \text{tr}(A_{(rr)} J_{m_r}) \right] \left[ \text{tr}(B_{(rr)} J_{m_r}) \right] (\mu_{\alpha_0}^{(4)} - 3\sigma_{\alpha_0}^4) \\ &+ \sum_{r=1}^R \left( \text{vec}_D(A_{(r,r)})' b_{(r)} + \text{vec}_D(B_{(r,r)})' a_{(r)} \right) \mu_{\epsilon_0, D_r}^{(3)} + \sum_{r=1}^R \left[ l'_{m_r} A_{(rr)} J_{m_r} b_{(r)} + l'_{m_r} B_{(rr)} J_{m_r} a_{(r)} \right] \mu_{\alpha_0}^{(3)}. \end{aligned}$$

*Proof.* Let  $H = [\sigma_{\alpha_0} \text{diag}_{r=1}^R \{l_{m_r}\}, \text{diag}_{r=1}^R \{\sigma_{\epsilon_0, D_r} I_{m_r}\}]$ . Consider the  $(N + R) \times 1$  dimensional vector

$$\xi = (\alpha_1/\sigma_{\alpha_0}, \dots, \alpha_R/\sigma_{\alpha_0}, \epsilon'_1/\sigma_{\epsilon_0, D_1}, \dots, \epsilon'_R/\sigma_{\epsilon_0, D_R})',$$

Then  $U = H\xi$ , and

$$U'AU + U'a = \xi'(H'AH)\xi + \xi'(H'a). \quad (\text{B.9})$$

Note that by Assumptions 1 and 2, the elements of  $\xi$  are independently distributed with  $E[\xi] = 0_{(N+R) \times 1}$ ,  $\text{Var}(\xi) = I_{N+R}$ . Denote the  $i$ -th entry of  $\xi$  as  $\xi_i$ ,  $1 \leq i \leq N + R$ . Denote the third and fourth moments of  $\xi_i$  as  $\mu_{\xi_i}^{(3)}$  and  $\mu_{\xi_i}^{(4)}$  respectively. Under Assumptions 1 and 2, when  $1 \leq i \leq R$ ,  $\mu_{\xi_i}^{(3)} = \mu_{\alpha_0}^{(3)}/\sigma_{\alpha_0}^3$  and  $\mu_{\xi_i}^{(4)} = \mu_{\alpha_0}^{(4)}/\sigma_{\alpha_0}^4$ . When  $R + m_1 + \dots + m_{r-1} + 1 \leq i \leq R + m_1 + \dots + m_r$ ,  $\mu_{\xi_i}^{(3)} = \mu_{\epsilon_0, D_r}^{(3)}/\sigma_{\epsilon_0, D_r}^3$  and  $\mu_{\xi_i}^{(4)} = \mu_{\epsilon_0, D_r}^{(4)}/\sigma_{\epsilon_0, D_r}^4$ . Furthermore, there exists some  $\eta_\xi > 0$  such that  $E|\xi_i|^{4+\eta_\xi} < \infty$ .

Using the transformation of linear quadratic forms in (B.9) and applying Lemma A.1 in Kelejian and Prucha (2010) yields,

$$E[U'AU + U'a] = E[\xi'(H'AH)\xi + \xi'(H'a)] = \text{tr}(H'AH) = \text{tr}(A\Omega_0),$$

observing that

$$HH' = \sigma_{\alpha_0}^2 \text{diag}_{r=1}^R \{l_{m_r} l'_{m_r}\} + \text{diag}_{r=1}^R \{\sigma_{\epsilon_0, D_r}^2 I_{m_r}\} = \Omega_0.$$

Furthermore the variance of the linear quadratic forms in  $U$  is given by

$$\begin{aligned}
& \text{Cov}(U'AU + U'a, U'BU + U'b) \\
&= \text{Cov}(\xi'(H'AH)\xi + \xi'(H'a), \xi'(H'BH)\xi + \xi'(H'b)) \\
&= 2 \text{tr}(H'AHH'BH) + a'HH'b \\
&+ \sum_{i=1}^{N+R} (H'AH)_{ii}(H'BH)_{ii}(\mu_{\xi_i}^{(4)} - 3) + \sum_{i=1}^{N+R} [(H'AH)_{ii}(H'b)_i + (H'BH)_{ii}(H'a)_i] \mu_{\xi_i}^{(3)} \\
&= 2 \text{tr}(A\Omega_0B\Omega_0) + a'\Omega_0b \\
&+ \sum_{r=1}^R \text{vec}_D(A_{(r,r)})' \text{vec}_D(B_{(r,r)}) (\mu_{\epsilon_0, D_r}^{(4)} - 3\sigma_{\epsilon_0, D_r}^4) + \sum_{r=1}^R \left[ \text{tr}(A_{(rr)}J_{m_r}) \right] \left[ \text{tr}(B_{(rr)}J_{m_r}) \right] (\mu_{\alpha_0}^{(4)} - 3\sigma_{\alpha_0}^4) \\
&+ \sum_{r=1}^R (\text{vec}_D(A_{(r,r)})'b_{(r)} + \text{vec}_D(B_{(r,r)})'a_{(r)}) \mu_{\epsilon_0, D_r}^{(3)} + \sum_{r=1}^R \left( \iota'_{m_r} A_{(rr)} J_{m_r} b_{(r)} + \iota'_{m_r} B_{(rr)} J_{m_r} a_{(r)} \right) \mu_{\alpha_0}^{(3)}.
\end{aligned}$$

□

**Lemma B.2.** [Central Limit Theorem] Suppose Assumptions 1, 2, 4 hold. For  $l = 1, \dots, L$  let  $A_N^{(l)}$  be non-stochastic  $N \times N$  matrices where the row and column sums of the absolute elements are uniformly bounded in  $N$ , and let  $a_N^{(l)}$  be  $N \times 1$  non-stochastic vectors where the absolute elements are uniformly bounded in  $N$ . Let  $S_N = [S_N^{(1)}, S_N^{(2)}, \dots, S_N^{(L)}]'$  be an  $L \times 1$  vector of linear quadratic forms of  $U$ , with

$$S_N^{(l)} = U' A_N^{(l)} U + U' a_N^{(l)}, \quad l = 1, \dots, L.$$

Let  $\Sigma_{S,N}$  denote variance covariance matrix of  $S_N$ , where explicit expressions for the elements of  $\Sigma_{S,N}$  are readily obtained from Lemma B.1, and assume that  $\rho_{\min}(\Sigma_{S,N}) \geq c$  for some constant  $c > 0$ . Let  $\Sigma_{S,N} = \Sigma_{S,N}^{1/2} \Sigma_{S,N}^{1/2}$ , then

$$\Sigma_{S,N}^{-1/2} (S_N - E[S_N]) \xrightarrow{d} N(0, I_L)$$

as  $N \rightarrow \infty$ .

(Note that under Assumption 5 the conditions postulated for  $a_N^{(l)}$  hold if  $a_N^{(l)} = B_N^{(l)} Z \beta_0$ , and the  $B_N^{(l)}$  are non-stochastic  $N \times N$  matrices where the row and column sums of the absolute elements are uniformly bounded in  $N$ .)

*Proof.* Let  $H$  and  $\xi$  be defined as in the proof of Lemma B.1, so that  $U = H\xi$ . Upon substitution of this expression for  $U$  we have

$$S_N^{(l)} = \xi' \tilde{A}_{N+R}^{(l)} \xi + \xi' \tilde{a}_{N+R}^{(l)}$$

where  $\tilde{A}_{N+R}^{(l)} = (1/2)H'(A_N^{(l)} + A_N^{(l)'})H$ ,  $\tilde{a}_{N+R}^{(l)} = H'a_N^{(l)}$ . Clearly, in light of Assumptions 1, 2,  $\xi$  satisfies Assumptions A.1 and A.3 in Kelejian and Prucha (2010). Furthermore, given the maintained assumptions on  $A_N^{(l)}$  and  $a_N^{(l)}$ , and since the row and column sums of  $H$  are uniformly bounded in



absolute value, it follows that the row and column sums of  $\tilde{A}_{N+R}^{(l)}$  and the elements of  $\tilde{a}_{N+R}^{(l)}$  are uniformly bounded in absolute value. This verifies that those matrices and vectors satisfy the conditions of Assumption A.2 in Kelejian and Prucha (2010). The lemma now follows from Theorem A.1 in Kelejian and Prucha (2010).  $\square$

As above, let  $\Theta_o$  be an open bounded set with  $\Theta \subset \Theta_o \subset (-1, 1) \times R_+^{J+1}$ .

**Lemma B.3.** *[Uniform Convergence] Let  $\Theta_0$  be an open set containing  $\Theta$ . Let  $A_N(\theta)$  and  $B_N(\theta)$  be  $N \times N$  matrices and let  $S_N(\theta)$  be a linear-quadratic form of  $U$ :*

$$S_N(\theta) = U' A_N(\theta) U + U' B_N(\theta) Z \beta_0$$

where  $A_N(\theta)$  and  $B_N(\theta)$  are differentiable  $N \times N$  matrices defined for  $\theta \in \Theta_0$ . Suppose Assumptions 1-5 hold, and suppose the row and column sums of  $A_N(\theta)$ ,  $B_N(\theta)$ ,  $\partial A_N(\theta)/\partial \theta_i$  and  $\partial B_N(\theta)/\partial \theta_i$ ,  $i = 1, \dots, J+2$ , are bounded in absolute value uniformly in  $N$  and  $\theta$ . Then  $N^{-1}S_N(\theta) - N^{-1}E[S_N(\theta)]$  converges uniformly to zero i.p., i.e.,

$$\text{plim}_{N \rightarrow \infty} \sup_{\theta \in \Theta} |N^{-1}S_N(\theta) - N^{-1}E[S_N(\theta)]| = 0.$$

*Remark B.1.* Given the uniform convergence in probability of  $N^{-1}S_N(\theta)$  to its mean and the equicontinuity of  $N^{-1}S_N(\theta)$ , we have  $\text{plim}_{N \rightarrow \infty} |N^{-1}S_N(\hat{\theta}_N) - N^{-1}E[S_N(\theta_0)]|$  as  $N$  goes to infinity if  $\hat{\theta}_N \rightarrow \theta_0$ .

*Proof.* To prove the lemma we verify that  $N^{-1}S_N(\theta)$  and  $N^{-1}\bar{S}_N(\theta) = N^{-1}E[S_N(\theta)]$  satisfy the conditions postulated by Corollary 2.2 of Newey (1991); cp., also Theorem 3.1(a) and the discussion after eq. (2.7) in Pötscher and Prucha (1994).

The parameter space  $\Theta$  is compact by assumption. We next verify that  $N^{-1}\bar{S}_N(\theta)$  is uniformly equicontinuous. By Lemma B.1,  $N^{-1}\bar{S}_N(\theta) = N^{-1}\text{tr}(\Omega_0 A(\theta))$ . Let  $\theta, \theta' \in \Theta$ , then by the mean value theorem

$$\text{tr}(\Omega_0 A_N(\theta)) = \text{tr}(\Omega_0 A_N(\theta')) + [\text{tr}(\Omega_0 \frac{\partial A_N(\theta^*)}{\partial \theta_1}), \dots, \text{tr}(\Omega_0 \frac{\partial A_N(\theta^*)}{\partial \theta_{J+2}})](\theta - \theta').$$

where  $\theta^*$  is a “vector of between values”. Note that the row and column sums of  $A_N(\theta)$ ,  $\nabla_{\theta_i} A_N(\theta) = \partial A_N(\theta)/\partial \theta_i$ ,  $\Omega_0$ , and consequently the row and column sums of  $\Omega_0 A_N(\theta)$  and  $(\Omega_0 \nabla_{\theta_i} A_N(\theta))$ , are uniformly (in  $\theta$  and  $N$ ) bounded in absolute value. Consequently there exists a constant  $C_A$  which does not depend on  $\theta$ ,  $\theta'$ , or  $N$  such that

$$\left| N^{-1} \text{tr}(\Omega_0 A_N(\theta)) - N^{-1} \text{tr}(\Omega_0 A_N(\theta')) \right| \leq C_A \|\theta - \theta'\|, \quad (\text{B.10})$$

which establishes that  $N^{-1}\bar{S}_N(\theta) = N^{-1}\text{tr}(\Omega_0 A_N(\theta))$  is uniformly equicontinuous on  $\Theta$ .

We next prove point-wise convergence i.p. of  $N^{-1}S_N(\theta) - N^{-1}E[S_N(\theta)]$  to zero. In light of Chebychev’s inequality it suffices to show that the variance of  $N^{-1}S_N(\theta)$  converges to 0 for any

$\theta \in \Theta$ . Let  $\bar{A}_N = (A_N(\theta) + A'_N(\theta)) / 2$  and  $\bar{a}_N = B_N(\theta)Z\beta_0$ , then by Lemma B.1, the variance of  $S_N(\theta)$  is

$$\begin{aligned} \text{Var} \left( N^{-1/2} S_N(\theta) \right) &= N^{-1} \text{tr} \left( \bar{A}_N \Omega_0 \bar{A}_N \Omega_0 \right) + N^{-1} \bar{a}'_N \Omega_0 \bar{a}_N \\ &\quad + N^{-1} \sum_{i=1}^N (\bar{a}_{ii,N})^2 (\mu_{\epsilon 0, D_{r(i)}}^{(4)} - 3\sigma_{\epsilon 0, D_{r(i)}}^4) + N^{-1} \sum_{r=1}^R \left( \text{tr}(\bar{A}_{(rr),N} J_{m_r}) \right)^2 (\mu_{\alpha}^{(4)} - 3\sigma_{\alpha 0}^4) \\ &\quad + 2N^{-1} \sum_{i=1}^N \bar{a}_{ii,N} \bar{a}_{i,N} \mu_{\epsilon 0, D_{r(i)}}^{(3)} + 2N^{-1} \sum_{r=1}^R \left( \iota'_{m_r} \bar{A}_{(rr),N} J_{m_r} \bar{a}_{(r),N} \right) \mu_{\alpha}^{(3)}. \end{aligned}$$

Under our assumptions the row and column sums of  $\Omega_0$ ,  $\bar{A}_N$ , and thus of  $\bar{A}_N \Omega_0 \bar{A}_N \Omega_0$  are uniformly bounded. Furthermore, it is readily seen that the elements of  $\bar{a}_N$  are uniformly bounded in absolute value. Consequently  $N^{-1} \text{tr} \left( \bar{A}_N \Omega_0 \bar{A}_N \Omega_0 \right)$ ,  $N^{-1} \bar{a}'_N \Omega_0 \bar{a}_N$ , and all sums in the above expression are seen to be bounded by a finite constant uniformly in  $N$ . In turn this implies that  $\text{Var}(N^{-1} S_N(\theta)) \rightarrow 0$ .

Finally we prove that  $N^{-1} S_N(\theta)$  satisfies the following Lipschitz condition:

$$\left| N^{-1} S_N(\theta) - N^{-1} S_N(\theta') \right| \leq C_N \|\theta - \theta'\|$$

for all  $\theta, \theta' \in \Theta$  and some nonnegative random variable  $C_N$  that does not depend on  $\theta, \theta'$  and where  $C_N = O_p(1)$ . It prove again convenient to rewrite as  $S_N(\theta)$  as  $S_N(\theta) = \xi' \tilde{A}_N(\theta) \xi + \xi' \tilde{a}_N(\theta)$  with  $\tilde{A}_N(\theta) = H' A_N(\theta) H$  and  $\tilde{a}_N(\theta) = H' B_N(\theta) Z \beta_0$ , where  $H$  and  $\xi$  are defined as in the proof of Lemma B.1. Under the maintained assumptions  $\tilde{A}_N(\theta)$  and  $\tilde{a}_N(\theta)$  are differentiable for  $\theta \in \Theta_o$ , and the row and column sums of  $\tilde{A}_N(\theta)$ ,  $\partial \tilde{A}_N(\theta) / \partial \theta_i$  and the elements of  $\tilde{a}_N(\theta)$ ,  $\partial \tilde{a}_N(\theta) / \partial \theta_i$  are uniformly bounded in absolute value in  $\theta$  and  $N$ , with  $i = 1, \dots, J+2$ . Consequently, for some finite constant, say  $K$ , we have  $|\tilde{a}_{i,N}(\theta)| \leq K/2$  and, using the mean value theorem,

$$\sum_{j=1}^{N+R} |\tilde{a}_{ij,N}(\theta) - \tilde{a}_{ij,N}(\theta')| \leq \sum_{j=1}^{N+R} \|\partial \tilde{a}_{ij,N}(\theta_*) / \partial \theta\| \|\theta - \theta'\| \leq K \|\theta - \theta'\|,$$

with  $\theta_*$  a ‘‘between value’’. Observing further that  $|\xi_i \xi_j| \leq (\xi_i^2 + \xi_j^2) / 2$  we have for any  $\theta, \theta' \in \Theta$

$$\begin{aligned} &\left| N^{-1} S_N(\theta) - N^{-1} S_N(\theta') \right| = \left| N^{-1} \xi' \left[ \tilde{A}_N(\theta) - \tilde{A}_N(\theta') \right] \xi + N^{-1} \xi' [\tilde{a}_N(\theta) - \tilde{a}_N(\theta')] \right| \\ &\leq \frac{2}{N+R} \sum_{i=1}^{N+R} \sum_{j=1}^{N+R} |\tilde{a}_{ij,N}(\theta) - \tilde{a}_{ij,N}(\theta')| (\xi_i^2 + \xi_j^2) / 2 + \frac{2}{N+R} \sum_{i=1}^{N+R} [|\tilde{a}_{i,N}(\theta)| + |\tilde{a}_{i,N}(\theta')|] |\xi_i| \\ &\leq \frac{1}{N+R} \sum_{i=1}^{N+R} \xi_i^2 \sum_{j=1}^{N+R} |\tilde{a}_{ij,N}(\theta) - \tilde{a}_{ij,N}(\theta')| + \frac{1}{N+R} \sum_{j=1}^{N+R} \xi_j^2 \sum_{i=1}^{N+R} |\tilde{a}_{ij,N}(\theta) - \tilde{a}_{ij,N}(\theta')| \\ &\quad + \frac{2K}{N+R} \sum_{i=1}^{N+R} |\xi_i| \end{aligned}$$

and consequently  $|N^{-1}S_N(\theta) - N^{-1}S_N(\theta')| \leq C_N \|\theta - \theta'\|$  where

$$C_N = \frac{2K}{N+R} \sum_{i=1}^{N+R} (\xi_i^2 + |\xi_i|).$$

Since  $E\xi_i^2 = 1$ , it follows that  $EC_N \leq 4KE(\xi_i^2 + |\xi_i|) \leq 8K$ , and thus  $C_N$  is  $O_p(1)$ . Having verified all conditions of Corollary 2.2 of Newey (1991), this concludes the proof of the lemma.  $\square$

The following Lemma is helpful in proving Theorem E.1 later. Motivated by the proof of Theorem E.1, we define the following variables:  $\bar{\phi} = \frac{(1-\hat{\lambda})\beta_0}{1-\lambda_0} - \hat{\beta}$ ,  $\bar{\varphi} = \frac{1-\hat{\lambda}}{1-\lambda_0}$ ,  $\bar{z}_r = \frac{1}{m_r} \sum_{i=1}^{m_r} z_{ir}$ ,  $\ddot{\phi}_r = \frac{m_r-1+\hat{\lambda}}{m_r-1+\lambda_0}\beta_0 - \hat{\beta}$ ,  $\ddot{\varphi}_r = 1 + \frac{\hat{\lambda}-\lambda_0}{m_r-1+\lambda_0}$ ,  $\ddot{z}_{ir} = z_{ir} - \bar{z}_r$ .

**Lemma B.4.** *Suppose Assumptions 1-5 hold, and let  $\psi(\cdot)$  be a finite positive scalar function of  $m_r$ ,  $0 < c_\psi \leq \psi(m_r) \leq C_\psi < \infty$  for  $r = 1, \dots, R$ . Assume that  $\hat{\lambda} \xrightarrow{p} \lambda_0$  and  $\hat{\beta} \xrightarrow{p} \beta_0$ . Let  $p_1$  and  $p_2$  be integers such that  $p_1 \geq 0$ ,  $p_2 \geq 0$  and  $p_1 + p_2 \leq 4$ , then:*

(a) *The term  $\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \psi(m_r) \sum_{i=1}^{m_r} \ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}$  has a finite expected value, and its deviation from the expected value converges in probability to zero as  $R \rightarrow \infty$ .*

(b) *For integers  $0 \leq s_1 \leq p_1$ ,  $0 \leq s_2 \leq p_2$ , and  $s_1 + s_2 \geq 1$ ,*

$$\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \psi(m_r) \sum_{i=1}^{m_r} (\ddot{z}_{ir} \ddot{\phi}_r)^{s_1} (\ddot{\varphi}_r \ddot{u}_{ir})^{p_1 - s_1} (\bar{z}_r \bar{\phi})^{s_2} (\bar{\varphi} \bar{u}_r)^{p_2 - s_2} \rightarrow_p 0.$$

(c) *As  $R$  goes to infinity,*

$$\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \psi(m_r) \sum_{i=1}^{m_r} [(\ddot{\varphi}_r \ddot{u}_{ir})^{p_1} (\bar{\varphi} \bar{u}_r)^{p_2} - \ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}] \rightarrow_p 0.$$

*Proof.* (a) With both  $m_r$  and  $1(D_r = j)\psi(m_r)$  being finite and  $0 \leq p_1 + p_2 \leq 4$ , Assumptions 1 and 2 imply that  $E[|1(D_r = j)\psi(m_r) \sum_{i=1}^{m_r} \ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}|^{1+\eta_\mu}] \leq C_\mu < \infty$  uniformly in  $r$  for some constant  $C_\mu$  and some  $\eta_\mu > 0$ , and that  $1(D_r = j)\psi(m_r) \sum_{i=1}^{m_r} \ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}$  are independently distributed across  $r$ . The claim thus follows from Theorem 5.4.1 and Corollary(ii) to that theorem in Chung (2001).

(b) Under Assumption 5, the elements of  $\bar{z}_r$  and  $\ddot{z}_{ir}$  are uniformly bounded in absolute value by some constants  $0 < C_Z < \infty$ . Under Assumptions 3 and 4,  $|\ddot{\varphi}|$  and  $|\bar{\varphi}|$  are uniformly bounded by some constants  $0 < C_\varphi < \infty$ . Let  $|\ddot{\phi}_r|_1$ ,  $|\beta_0|_1$ ,  $|\beta_0 - \hat{\beta}|_1$  be the  $\ell_1$  norm of  $\ddot{\phi}_r$ ,  $\beta_0$  and  $\beta_0 - \hat{\beta}$  respectively. Observe that  $m_r - 1 + \lambda_0 \geq \epsilon_\lambda$  for some  $\epsilon_\lambda > 0$ , and thus

$$\begin{aligned} |\ddot{\phi}_r|_1 &= \left| \frac{m_r - 1 + \hat{\lambda}}{m_r - 1 + \lambda_0} \beta_0 - \hat{\beta} \right|_1 \\ &= \left| \frac{(\hat{\lambda} - \lambda_0)\beta_0}{m_r - 1 + \lambda_0} + (\beta_0 - \hat{\beta}) \right|_1 \\ &\leq |\hat{\lambda} - \lambda_0| |\beta_0|_1 \frac{1}{\epsilon_\lambda} + |\beta_0 - \hat{\beta}|_1. \end{aligned}$$

Therefore,

$$\begin{aligned} & |\psi(m_r)| |\ddot{z}_{ir} \ddot{\phi}_r|^{s_1} |\ddot{\varphi}_r \ddot{u}_{ir}|^{p_1 - s_1} |\bar{z}_r \bar{\phi}|^{s_2} |\bar{\varphi} \bar{u}_r|^{p_2 - s_2} \\ & \leq C_\psi C_Z^{s_1 + s_2} C_\varphi^{(p_1 + p_2 - s_1 - s_2)} (|\hat{\lambda} - \lambda_0| |\beta_0|_1 \frac{1}{\epsilon_\lambda} + |\beta_0 - \hat{\beta}|_1)^{s_1} |\bar{\phi}|^{s_2} |\ddot{u}_{ir}|^{p_1 - s_1} |\bar{u}_r|^{p_2 - s_2}. \end{aligned}$$

and

$$\begin{aligned} & \left| \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \psi(m_r) \sum_{i=1}^{m_r} (\ddot{z}_{ir} \ddot{\phi}_r)^{s_1} (\ddot{\varphi}_r \ddot{u}_{ir})^{p_1} (\bar{z}_r \bar{\phi})^{s_2} (\bar{\varphi} \bar{u}_r)^{p_2} \right| \\ & \leq \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \sum_{i=1}^{m_r} |\psi(m_r)| |\ddot{z}_{ir} \ddot{\phi}_r|^{s_1} |\ddot{\varphi}_r \ddot{u}_{ir}|^{p_1 - s_1} |\bar{z}_r \bar{\phi}|^{s_2} |\bar{\varphi} \bar{u}_r|^{p_2 - s_2} \\ & \leq C_\psi C_Z^{s_1 + s_2} C_\varphi^{(p_1 + p_2 - s_1 - s_2)} (|\hat{\lambda} - \lambda_0| |\beta_0|_1 \frac{1}{\epsilon_\lambda} + |\beta_0 - \hat{\beta}|_1)^{s_1} |\bar{\phi}|^{s_2} \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \sum_{i=1}^{m_r} |\ddot{u}_{ir}|^{p_1 - s_1} |\bar{u}_r|^{p_2 - s_2}. \end{aligned}$$

Note that  $\bar{\phi} \rightarrow_p 0$ . With  $s_1 \geq 0$ ,  $s_2 \geq 0$  and  $s_1 + s_2 \geq 1$ ,  $(|\hat{\lambda} - \lambda_0| |\beta_0|_1 \frac{1}{\epsilon_\lambda} + |\beta_0 - \hat{\beta}|_1)^{s_1} |\bar{\phi}|^{s_2} \rightarrow_p 0$ . By part (a) of the Lemma,  $\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \sum_{i=1}^{m_r} |\ddot{u}_{ir}|^{p_1 - s_1} |\bar{u}_r|^{p_2 - s_2}$  is bounded in probability. Consequently the equation above converges to 0 in probability.

(c) We can rewrite  $\ddot{\varphi}_r = 1 + \zeta_r$ , and  $\bar{\varphi} = 1 + \bar{\zeta}$ , where  $\zeta_r = \frac{\hat{\lambda} - \lambda_0}{m_r - 1 + \lambda_0}$ ,  $\bar{\zeta} = -\frac{\hat{\lambda} - \lambda_0}{1 - \lambda_0}$ . Since  $m_r - 1 + \lambda_0 > \epsilon_\lambda$  for some  $\epsilon_\lambda > 0$  and  $|\hat{\lambda} - \lambda_0| < 2$ , both  $\zeta_r$  and  $\bar{\zeta}$  are uniformly bounded in absolute value and there exists some constant  $0 < C_\zeta < \infty$  such that  $|\zeta_r| \leq C_\zeta |\hat{\lambda} - \lambda_0|$  and  $|\bar{\zeta}| \leq C_\zeta |\hat{\lambda} - \lambda_0|$ . Next observe that by the mean-value theorem, for  $p \geq 1$  we have  $(1+x)^p = 1 + px(1+\tilde{x})^{p-1}$  where  $\tilde{x}$  lies between  $x$  and 0. The equation also holds trivially for  $p = 0$ . Consequently,

$$\begin{aligned} |\ddot{\varphi}_r^{p_1} \bar{\varphi}^{p_2} - 1| &= \left| \left(1 + p_1 \zeta_r (1 + \tilde{\zeta})^{p_1 - 1}\right) \left(1 + p_2 \bar{\zeta} (1 + \tilde{\zeta})^{p_2 - 1}\right) - 1 \right| \\ &= \left| p_1 p_2 \zeta_r (1 + \tilde{\zeta})^{p_1 - 1} \bar{\zeta} (1 + \tilde{\zeta})^{p_2 - 1} + p_1 \zeta_r (1 + \tilde{\zeta})^{p_1 - 1} + p_2 \bar{\zeta} (1 + \tilde{\zeta})^{p_2 - 1} \right| \\ &\leq p_1 p_2 |\zeta_r \bar{\zeta}| |(1 + \tilde{\zeta})^{p_1 - 1} (1 + \tilde{\zeta})^{p_2 - 1}| + p_1 |\zeta_r| |(1 + \tilde{\zeta})^{p_1 - 1}| + p_2 |\bar{\zeta}| |(1 + \tilde{\zeta})^{p_2 - 1}| \end{aligned}$$

where  $\tilde{\zeta}$  lies between  $\zeta$  and 0,  $\tilde{\zeta}$  lies between  $\bar{\zeta}$  and 0, and thus  $\tilde{\zeta}$  and  $\tilde{\zeta}$  are both uniformly bounded in absolute value. Therefore  $|(1 + \tilde{\zeta})^{p_1 - 1} (1 + \tilde{\zeta})^{p_2 - 1}|$ ,  $|(1 + \tilde{\zeta})^{p_1 - 1}|$  and  $|(1 + \tilde{\zeta})^{p_2 - 1}|$  are all uniformly bounded. Therefore there exists some constant  $0 < C_p < \infty$  such that

$$|\ddot{\varphi}_r^{p_1} \bar{\varphi}^{p_2} - 1| \leq C_p |\hat{\lambda} - \lambda_0|,$$

and

$$\begin{aligned}
& \left| \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \psi(m_r) \sum_{i=1}^{m_r} ((\ddot{\varphi}_r \ddot{u}_{ir})^{p_1} (\bar{\varphi} \bar{u}_r)^{p_2} - \ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}) \right| \\
& \leq \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \sum_{i=1}^{m_r} |\psi(m_r)| |\ddot{\varphi}_r^{p_1} \bar{\varphi}^{p_2} - 1| |\ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}| \\
& \leq C_\varphi C_p |\hat{\lambda} - \lambda_0| \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \sum_{i=1}^{m_r} |\ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}|.
\end{aligned}$$

By part (a) of the lemma,  $\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \sum_{i=1}^{m_r} |\ddot{u}_{ir}^{p_1} \bar{u}_r^{p_2}|$  is bounded in probability. The lemma then follows.  $\square$

**Lemma B.5.** *Suppose Assumption 5 holds. Let*

$$A_N(\theta) = \text{diag}_{r=1}^R \{p(m_r, \theta) I_{m_r}^* + s(m_r, \theta) J_{m_r}^*\},$$

where, for  $2 \leq m_r \leq \bar{M}$ , the scalar functions  $p(m_r, \theta)$  and  $s(m_r, \theta)$  are positive and continuous on the compact parameter space  $\Theta$ . Let  $S_N(\theta) = N^{-1} Z' A_N(\theta) Z$ , then there exist positive constants  $\underline{c}$  and  $\underline{C}$  that do not depend on  $\theta$  and  $N$  such that

$$0 < \underline{c} \leq \lambda_{\min} [S_N(\theta)] \leq \lambda_{\max} [S_N(\theta)] \leq \underline{C} < \infty. \quad (\text{B.11})$$

Furthermore

$$\sup_{\theta \in \Theta} |S_N(\theta) - S(\theta)| \rightarrow 0 \text{ as } N \rightarrow \infty, \quad (\text{B.12})$$

where  $S(\theta) = \sum_{m=2}^{\bar{M}} \sum_{j=1}^J [p(m, \theta) \ddot{z}_{m,j} + s(m, \theta) \bar{z}_{m,j}]$ . The elements of  $S(\theta)$  are continuous on  $\Theta$ , and

$$0 < \underline{c} \leq \lambda_{\min} [S(\theta)] \leq \lambda_{\max} [S(\theta)] \leq \underline{C} < \infty. \quad (\text{B.13})$$

*Remark B.2.* It follows from the uniform convergence of  $S_N(\theta)$  and the continuity of  $S(\theta)$  that if  $\hat{\theta}_N \rightarrow_p \theta_0$ , then  $|S_N(\hat{\theta}_N) - S(\theta_0)| \rightarrow_p 0$  as  $N \rightarrow \infty$ .

*Proof.* Observe that by the Bolzano-Weierstrass' extreme value theorem there exist positive constants  $c$  and  $C$ , which do not depend on  $\theta$ , such that

$$0 < c \leq p(m_r, \theta), s(m_r, \theta) \leq C < \infty.$$

Since  $m_r$  only takes on finitely many values the constants  $c$  and  $C$  can be chosen such that the above inequalities hold for all  $m$ . By Assumption 5 we have  $0 < \underline{\xi}_Z \leq \lambda_{\min}[N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}'_r \ddot{Z}_r + N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}'_r \bar{z}_r]$  for some pair of  $(m, j)$ . Since the elements of  $Z$  are bounded in absolute value it follows further that there exists a finite constant  $\bar{\xi}_Z$  such that for all pairs of  $(m, j)$ ,  $\lambda_{\max}[N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}'_r \ddot{Z}_r] \leq \bar{\xi}_Z < \infty$  and  $\lambda_{\max}[N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}'_r \bar{z}_r] \leq \bar{\xi}_Z < \infty$ ; see, e.g., Johnson

and Horn (1985), Lemma 5.6.10 and the equivalence of matrix norms. Next observe that

$$\begin{aligned} S_N(\theta) &= N^{-1} \sum_{r=1}^R (p(m_r, \theta) Z_r' I_{m_r}^* Z_r + s(m_r, \theta) Z_r' J_{m_r}^* Z_r) \\ &= \sum_{m=2}^{\bar{M}} \sum_{j=1}^J \left( p(m, \theta) N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r + s(m, \theta) N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r \right). \end{aligned}$$

Consequently

$$\begin{aligned} \lambda_{\min} [S_N(\theta)] &= \inf_{\phi \in R^{k_Z}} \frac{\phi' S_N(\theta) \phi}{\phi' \phi} \\ &\geq \sum_{m=2}^{\bar{M}} \sum_{j=1}^J \inf_{\phi \in R^{k_Z}} \frac{\phi' \left( p(m, \theta) N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r + s(m, \theta) N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r \right) \phi}{\phi' \phi} \\ &\geq c \sum_{m=2}^{\bar{M}} \sum_{j=1}^J \inf_{\phi \in R^{k_Z}} \frac{\phi' \left( N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r + N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r \right) \phi}{\phi' \phi} \\ &\geq c \xi_Z > 0 \end{aligned}$$

and

$$\begin{aligned} \lambda_{\max} [S_N(\theta)] &\leq \sup_{\phi \in R^{k_Z}} \frac{\phi' S_N(\theta) \phi}{\phi' \phi} \\ &\leq \sum_{m=2}^{\bar{M}} \sum_{j=1}^J \left( p(m, \theta) \sup_{\phi \in R^{k_Z}} \frac{\phi' [N^{-1} \sum_{r \in \mathcal{I}_m} \ddot{Z}_r' \ddot{Z}_r] \phi}{\phi' \phi} + s(m, \theta) \sup_{\phi \in R^{k_Z}} \frac{\phi' [N^{-1} \sum_{r \in \mathcal{I}_m} m \bar{z}_r' \bar{z}_r] \phi}{\phi' \phi} \right) \\ &\leq 2 \sum_{m=2}^{\bar{M}} \sum_{j=1}^J C \bar{\xi}_Z < \infty. \end{aligned}$$

This proves the first part of the lemma. Next observe that

$$\begin{aligned} &\sup_{\theta \in \Theta} |S_N(\theta) - S(\theta)| \\ &\leq \sup_{\theta \in \Theta} \sum_{m=2}^{\bar{M}} \sum_{j=1}^J \left( p(m, \theta) \left| N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r - \ddot{\mathcal{Z}}_{m,j} \right| + s(m, \theta) \left| N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r - \bar{\mathcal{Z}}_{m,j} \right| \right) \\ &\leq C \sum_{m=2}^{\bar{M}} \sum_{j=1}^J \left( \left| N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r - \ddot{\mathcal{Z}}_{m,j} \right| + \left| N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r - \bar{\mathcal{Z}}_{m,j} \right| \right) \rightarrow 0 \end{aligned}$$

by Assumption 5. Clearly  $S(\theta)$  is continuous given the assumptions maintained w.r.t.  $p(m_r, \theta)$  and  $s(m_r, \theta)$ . Recall that by Assumption 5 we have  $0 < \xi_Z \leq \lambda_{\min} [N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}_r' \ddot{Z}_r + N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}_r' \bar{z}_r]$  for some pair of  $(m, j)$ . Therefore and since the eigenvalues of a matrix are continuous functions

of the elements of the matrix we have for some pair of  $(m, j)$

$$\begin{aligned}
0 < \underline{\xi}_Z &\leq \lim_{N \rightarrow \infty} \lambda_{\min} \left[ N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}'_r \ddot{Z}_r + N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}'_r \bar{z}_r \right] \\
&= \lambda_{\min} \left[ \lim_{N \rightarrow \infty} N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}'_r \ddot{Z}_r + \lim_{N \rightarrow \infty} N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}'_r \bar{z}_r \right] \\
&= \lambda_{\min} [\ddot{\mathcal{Z}}_{m,j} + \bar{\mathcal{Z}}_{m,j}],
\end{aligned}$$

and we have for all pairs of  $(m, j)$ ,

$$\begin{aligned}
\lambda_{\max}(\ddot{\mathcal{Z}}_{m,j}) &= \lambda_{\max} \left[ \lim_{N \rightarrow \infty} N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}'_r \ddot{Z}_r \right] = \lim_{N \rightarrow \infty} \lambda_{\max} \left[ N^{-1} \sum_{r \in \mathcal{I}_{m,j}} \ddot{Z}'_r \ddot{Z}_r \right] \leq \bar{\xi}_Z < \infty, \\
\lambda_{\max}(\bar{\mathcal{Z}}_{m,j}) &= \lambda_{\max} \left[ \lim_{N \rightarrow \infty} N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}'_r \bar{z}_r \right] = \lim_{N \rightarrow \infty} \lambda_{\max} \left[ N^{-1} \sum_{r \in \mathcal{I}_{m,j}} m \bar{z}'_r \bar{z}_r \right] \leq \bar{\xi}_Z < \infty.
\end{aligned}$$

The remainder of the proof of (B.13) is analogous to the proof of (B.11).

□

## C Proof of Lemma 2.1

We first consider Scenario (i) . By assumption there are two groups  $r$  and  $s$  such that  $m_r \neq m_s$  and  $E[\epsilon_{ir}^2|m_r, D_r] / E[\epsilon_{is}^2|m_s, D_s] = \sigma_{\epsilon_0, D_r}^2 / \sigma_{\epsilon_0, D_s}^2 = 1$  . Now consider

$$E[\chi_r^w(\theta)|m_r, D_r] = \frac{(m_r - 1 + \lambda)^2 E[\ddot{Y}_r' \ddot{Y}_r]}{(m_r - 1)^2} - \sigma_{\epsilon, D_r}^2 (m_r - 1) = 0 \quad (\text{C.1})$$

Using  $E[\ddot{Y}_r' \ddot{Y}_r] = (m_r - 1)^2 / (m_r - 1 + \lambda_0)^2 \sigma_{\epsilon_0, D_r}^2$  gives

$$E[\chi_r^w(\theta)|m_r, D_r] = (m_r - 1) \left[ \frac{(m_r - 1 + \lambda)^2 \sigma_{\epsilon_0, D_r}^2}{(m_r - 1 + \lambda_0)^2} - \sigma_{\epsilon, D_r}^2 \right] = 0$$

which leads to the equation

$$(m_r - 1 + \lambda)^2 \sigma_{\epsilon_0, D_r}^2 = \sigma_{\epsilon, D_r}^2 (m_r - 1 + \lambda_0)^2. \quad (\text{C.2})$$

Now use the moment condition for groups  $r$  and  $s$  and noting that  $\sigma_{\epsilon_0, D_r}^2 / \sigma_{\epsilon_0, D_s}^2 = \sigma_{\epsilon, D_r}^2 / \sigma_{\epsilon, D_s}^2 = 1$  It follows that

$$\frac{(m_r - 1 + \lambda)^2}{(m_s - 1 + \lambda)^2} = \frac{(m_r - 1 + \lambda_0)^2}{(m_s - 1 + \lambda_0)^2}. \quad (\text{C.3})$$

Clearly the equation in (C.3) holds for  $\lambda = \lambda_0$ . The RHS is constant in  $\lambda$ . The LHS is a monotonic function in  $\lambda$ . To see this, compute the derivative  $\partial h(\lambda) / \partial \lambda$  of

$$h(\lambda) = \frac{(m_r - 1 + \lambda)^2}{(m_s - 1 + \lambda)^2}$$

given by

$$\begin{aligned} \frac{\partial h(\lambda)}{\partial \lambda} &= \frac{2(m_r - 1 + \lambda)}{(m_s - 1 + \lambda)^2} - \frac{2(m_r - 1 + \lambda)^2 (m_s - 1 + \lambda)}{(m_s - 1 + \lambda)^4} \\ &= \frac{2(m_r - 1 + \lambda)(m_s - 1 + \lambda)(m_s - m_r)}{(m_s - 1 + \lambda)^4}. \end{aligned}$$

Since  $\lambda > -1$ ,  $m_r > 1$  and  $m_s > 1$  then  $\text{sign}(\partial h(\lambda) / \partial \lambda) = \text{sign}(m_s - m_r)$ . This implies that (C.3) can only have one solution when  $m_r \neq m_s$ . Thus  $E[\chi_r^w(\theta)|m_q, D_q] = 0$ ,  $q = r, s$  alone identifies  $\lambda$  under scenario (i). Plugging  $\lambda = \lambda_0$  into (C.2) and noting that  $m_r - 1 + \lambda_0 > 0$  shows that  $\sigma_{\epsilon_0, D_r}^2$  is identified. By assumption,  $\sigma_{\epsilon_0, D_r}^2 = \sigma_{\epsilon_0, D_s}^2$  such that  $\sigma_{\epsilon_0, D_s}^2$  is also identified. The remaining moments  $E[\chi_r^w(\theta)|m_q, D_q] = 0$ ,  $q \neq r, s$  now determine the remaining parameters  $\sigma_{\epsilon, j}^2$ .



Now consider .

$$\begin{aligned} E \left[ \chi_r^b(\theta) | m_r, D_r \right] &= (1 - \lambda)^2 E \bar{y}_r^2 - \sigma_\alpha^2 - \frac{\sigma_{\epsilon, D_r}^2}{m_r} \\ &= \frac{(1 - \lambda)^2}{(1 - \lambda_0)^2} \left( \sigma_{\alpha, 0}^2 + \frac{\sigma_{\epsilon_0, D_r}^2}{m_r} \right) - \left( \sigma_\alpha^2 + \frac{\sigma_{\epsilon, D_r}^2}{m_r} \right) \end{aligned}$$

We already established that  $E \left[ \chi_q^w(\theta) | m_q, D_q \right] = 0$ ,  $q = r, s$  imply that  $\lambda = \lambda_0$ ,  $\sigma_{\epsilon, D_q}^2 = \sigma_{\epsilon_0, D_q}^2$  for  $q = r, s$ . Then, for  $\lambda = \lambda_0$ ,  $\sigma_{\epsilon, D_r}^2 = \sigma_{\epsilon_0, D_r}^2$

$$E \left[ \chi_r^b(\theta) | m_r, D_r \right] = \frac{(1 - \lambda)^2}{(1 - \lambda_0)^2} \left( \sigma_{\alpha, 0}^2 + \frac{\sigma_{\epsilon_0, D_r}^2}{m_r} \right) - \left( \sigma_\alpha^2 + \frac{\sigma_{\epsilon, D_r}^2}{m_r} \right) = 0 \quad (\text{C.4})$$

reduces to  $\sigma_\alpha^2 = \sigma_{\alpha, 0}^2$ , and thus also  $\sigma_{\alpha, 0}^2$  is identified. If there are additional groups with sizes different from  $m_r$  and  $m_s$  then moment conditions related to these groups constitute overidentifying restrictions.

Now consider Scenario (ii). By assumption,  $m_r = m_s = m$  and  $\sigma_{\epsilon_0, D_r}^2 \neq \sigma_{\epsilon_0, D_s}^2$ . Thus

$$\begin{aligned} &E [\nu_r(\theta) | m_r, D_r] - E [\nu_s(\theta) | m_s, D_s] \\ &= (1 - \lambda)^2 \left( E [\bar{y}_r^2 | m, D_r] - E [\bar{y}_s^2 | m, D_s] \right) \\ &\quad - (m - 1 + \lambda)^2 \left( E \left[ \frac{\ddot{Y}_r' \ddot{Y}_r}{m(m - 1)^3} | m, D_r \right] - E \left[ \frac{\ddot{Y}_s' \ddot{Y}_s}{m(m_s - 1)^3} | m, D_s \right] \right). \\ &= \left[ \frac{(1 - \lambda)^2}{(1 - \lambda_0)^2} - \frac{(m - 1 + \lambda)^2}{(m - 1 - \lambda_0)^2} \right] \frac{\sigma_{\epsilon_0, D_r}^2 - \sigma_{\epsilon_0, D_s}^2}{m} \end{aligned}$$

observing that,

$$E [\bar{y}_r^2 | m, D_r] - E [\bar{y}_s^2 | m, D_s] = \frac{\sigma_{\alpha_0}^2 + \frac{\sigma_{\epsilon_0, D_r}^2}{m}}{(1 - \lambda_0)^2} - \frac{\sigma_{\alpha_0}^2 + \frac{\sigma_{\epsilon_0, D_s}^2}{m}}{(1 - \lambda_0)^2} = \frac{\sigma_{\epsilon_0, D_r}^2 - \sigma_{\epsilon_0, D_s}^2}{m(1 - \lambda_0)^2},$$

and

$$E \left[ \frac{\ddot{Y}_r' \ddot{Y}_r}{m(m - 1)^3} | m, D_r \right] - E \left[ \frac{\ddot{Y}_s' \ddot{Y}_s}{m(m - 1)^3} | m, D_s \right] = \frac{1}{m(m - 1 - \lambda_0)^2} (\sigma_{\epsilon_0, D_r}^2 - \sigma_{\epsilon_0, D_s}^2).$$

Since  $\sigma_{\epsilon_0, D_r}^2 - \sigma_{\epsilon_0, D_s}^2 \neq 0$  it follows that  $E [\nu_r(\theta) | m_r, D_r] - E [\nu_s(\theta) | m_s, D_s] = 0$  implies

$$\frac{(m - 1 + \lambda)^2}{(1 - \lambda)^2} = \frac{(m - 1 + \lambda_0)^2}{(1 - \lambda_0)^2}.$$

Define  $c = \frac{(m - 1 + \lambda_0)^2}{(1 - \lambda_0)^2}$ , then the equation is equivalent to

$$(m - 1 + \lambda)^2 = c(1 - \lambda)^2$$

which in turn is equivalent to the following polynomial in  $\lambda$ :

$$(m-1)^2 - c + 2((m-1) + c)\lambda + (1-c)\lambda^2 = 0$$

Clearly,  $\lambda = \lambda_0$  is a solution. Consider the derivative

$$\begin{aligned} \frac{\partial \left( \frac{(m-1+\lambda)^2}{(1-\lambda)^2} \right)}{\partial \lambda} &= \frac{2(m-1+\lambda)}{(1-\lambda)^2} + \frac{2(m-1+\lambda)^2}{(1-\lambda)^3} \\ &= \frac{2((m-1+\lambda)(1-\lambda) + (m-1+\lambda)^2)}{(1-\lambda)^3} \\ &= 2(m-1+\lambda) \frac{((1-\lambda) + (m-1+\lambda))}{(1-\lambda)^3} \\ &= \frac{2(m-1+\lambda)m}{(1-\lambda)^3} > 0. \end{aligned}$$

Since  $m \geq 2$  and  $\lambda \in (-1, 1)$  it follows that  $(m-1+\lambda)m > 0$  and  $1-\lambda > 0$  such that the derivative is positive for all values of  $\lambda$  on the parameter space. This implies that  $\lambda = \lambda_0$  is the only solution to the moment condition. Once  $\lambda$  is identified,  $E[\nu_r(\theta)] = 0$  identifies  $\sigma_\alpha^2$  as

$$\sigma_\alpha^2 = E \left[ (1-\lambda_0)^2 \bar{y}_r^2 - \frac{\bar{Y}_r' \bar{Y}_r}{m_r(m_r-1)^3} \right] = \sigma_{\alpha 0}^2.$$

Finally, note that  $\nu_r(\theta)$  is a function of  $\chi_r(\theta)$ , and thus the moment conditions  $E[\chi_r(\theta_0) | m_r, D_r] = 0$  are sufficient to identify the parameter  $\lambda$  and  $\sigma_\alpha^2$ .

Identification of the remaining parameters  $\sigma_{\epsilon, j}^2$  follows trivially from an inspection of  $\chi_r^w(\theta)$  as once  $\lambda$  is identified,  $\sigma_{\epsilon, j}^2$  is identified from  $\chi_r^w(\theta)$  recalling that by Assumption 1 there exists some  $r$  such that  $D_r = j$ .

## D The CV estimator of Graham (2008)

In this appendix we interpret the CV estimator of Graham (2008) as based on moment conditions developed in Section 2. Specifically, we show that the identification results in Graham (2008) can be seen as an adapted version of Scenario (ii) of Lemma 2.1.

The peer effects model of Graham (2008) can be written as

$$y_{ir} = v_r + \epsilon_{ir} + (\gamma - 1) \bar{\epsilon}_r, \tag{D.1}$$

where  $\bar{\epsilon}_r = m_r^{-1} \sum_{i=1}^{m_r} \epsilon_{ir}$  is the group average of unobserved characteristics. The parameter  $\gamma$  captures the peer effect. Taking group averages on both sides of (D.1), we get  $\bar{\epsilon}_r = \frac{1}{\gamma} (\bar{y}_r - v_r)$ . Plugging back into (D.1), and letting  $\tilde{\lambda} = 1 - 1/\gamma$  as well as  $\alpha_r = v_r/\gamma$ , we get the following structural model

$$y_{ir} = \tilde{\lambda} \bar{y}_r + \alpha_r + \epsilon_{ir}. \tag{D.2}$$

The specification differs from our main model in (1) in that it uses the full group mean  $\bar{y}_r$  rather than the leave-out-mean  $\bar{y}_{(-i)r}$ . The leave-out-mean specifications are often preferred in the literature, see for example Angrist (2014) for a discussion. Defining  $\tilde{W}_{m_r} = \frac{1}{m_r} \iota_{m_r} \iota'_{m_r}$ , we can rewrite (D.2) in matrix form as

$$Y_r = \tilde{\lambda} \tilde{W}_{m_r} Y_r + \alpha \iota_{m_r} + \epsilon_r.$$

Using the same notation as in Section 2, it can be shown that  $\bar{y}_r = \bar{u}_r / (1 - \tilde{\lambda})$  with  $\bar{u}_r = \alpha_r + \bar{\epsilon}_r$ , and  $\ddot{Y}_r = \ddot{U}_r = \ddot{\epsilon}_r$ . Note that in the context of Model (D.2) the results in Manski (1993), Kelejian et al. (2006) or Bramoullé et al. (2009) show that  $\tilde{\lambda}$  cannot be identified by instrumenting  $\tilde{W}_{m_r} Y_r$  with  $\tilde{W}_{m_r}^2 Z_r$  when  $\tilde{W}_{m_r} Z_r$  is included as a covariate, observing that  $\tilde{W}_{m_r}^2 = \tilde{W}_{m_r}$ .

To isolate or identify the social interaction effect Graham (2008) imposes restrictions on the unobservables and group size. Graham considers the case when  $J = 2$  and  $D_r$  is a categorical variable for small/regular classes, which are coded as  $D_r = 1$  whenever  $m_r \geq \bar{m}$  for some constant  $\bar{m}$  and  $D_r = 2$  otherwise. Assumptions 1.1 and 1.2 in Graham (2008) amount to Assumptions 1 and 2 in Section 2. Assumption 1.3 in Graham (2008) imposes that  $E[\ddot{Y}_r' \ddot{Y}_r | D_r = 1] \neq E[\ddot{Y}_r' \ddot{Y}_r | D_r = 2]$ . Observing that  $E[\ddot{Y}_r' \ddot{Y}_r | D_r = d] = \sigma_{\epsilon, d}^2 (m_r - 1)$  the latter condition is seen to be satisfied if Scenario (ii) in Lemma 2.1 holds true. In addition, the condition is also true if  $m_r$  varies and the idiosyncratic errors  $\epsilon_{ir}$  are homoscedastic as in Scenario (i).

The parameter  $\gamma^2 = 1 / (1 - \tilde{\lambda})^2$  is identified under Assumptions 1.1-1.3, see Proposition 1.1 in Graham (2008). Below we prove the proposition by adapting the proof for Scenario (ii) in Lemma 2.1 to the full-mean specification, thus verifying that the CV estimator is a special case of the moment based estimators studied in Section 2.

Under the full-mean specification of Graham (2008), our moment conditions in (7) change correspondingly to

$$\tilde{\chi}_r^w(\theta) = \ddot{Y}_r' \ddot{Y}_r - (m_r - 1) \sigma_{\epsilon, D_r}^2, \quad (\text{D.3})$$

$$\tilde{\chi}_r^b(\theta) = (1 - \tilde{\lambda})^2 \bar{y}_r^2 - \sigma_\alpha^2 - \frac{\sigma_{\epsilon, D_r}^2}{m_r}. \quad (\text{D.4})$$

The combined moment condition in (8) is now

$$\tilde{\nu}_r(\tilde{\lambda}, \sigma_\alpha^2) = \tilde{\chi}_r^b(\theta) - \frac{\tilde{\chi}_r^w(\theta)}{m_r(m_r - 1)} = (1 - \tilde{\lambda})^2 \bar{y}_r^2 - \sigma_\alpha^2 - \frac{\ddot{Y}_r' \ddot{Y}_r}{m_r(m_r - 1)}, \quad (\text{D.5})$$

Note that by design  $E(\tilde{\nu}_r(\tilde{\lambda}, \sigma_\alpha^2) | m_r, D_r) = 0$  at the true parameter vector  $\theta_0$ .

The restriction that group effect variances are homoscedastic can be exploited by taking differences  $E[\nu_r(\tilde{\lambda}, \sigma_\alpha^2) | m_r, D_r = 1] - E[\nu_r(\tilde{\lambda}, \sigma_\alpha^2) | m_r, D_r = 2]$  to eliminate  $\sigma_\alpha^2$ . This implies the following population equation for  $(1 - \tilde{\lambda})^2$

$$\frac{1}{(1 - \tilde{\lambda})^2} = \frac{E[\bar{y}_r^2 | m_r, D_r = 1] - E[\bar{y}_r^2 | m_r, D_r = 2]}{E\left[\frac{\ddot{Y}_r' \ddot{Y}_r}{m_r(m_r - 1)} | m_r, D_r = 1\right] - E\left[\frac{\ddot{Y}_r' \ddot{Y}_r}{m_r(m_r - 1)} | m_r, D_r = 2\right]}, \quad (\text{D.6})$$

which is a modified version of (6). The Wald estimate for  $\tilde{\lambda}$  can then be calculated from the sample analog of the right-hand side above. Two points are worth noting. First, identification is possible under Scenario (ii) of Lemma 2.1, i.e., when there exists some  $m_r = m_s$  and  $\sigma_{\epsilon, D_r}^2 \neq \sigma_{\epsilon, D_s}^2$ . This confirms the applicability of Lemma 2.1 to the full-mean specification. Second, due to the full-mean specification, identification is also possible even when  $\sigma_{\epsilon, 1}^2 = \sigma_{\epsilon, 2}^2$  as long as there is variation in  $m_r$ . So, in the case of homoscedasticity, group size variation alone is enough for the CV estimator to identify  $\tilde{\lambda}$ . It can be shown that the score of a Gaussian likelihood estimator is a function of  $\tilde{\chi}_r(\theta) = (\tilde{\chi}_r^w(\theta), \tilde{\chi}_r^b(\theta))'$  and therefore the Gaussian maximum likelihood estimator shares the same identification properties.

It is well known that identification for the case of peer effects captured by full group means is difficult, see Manski (1993), Bramoullé et al. (2009) or Angrist (2014). In the case of the conditional variance restrictions or likelihood approaches considered here, this manifests itself in the fact that  $(1 - \tilde{\lambda})^2$  but not  $\tilde{\lambda}$  is identified without additional constraints on the parameter space. An inspection of (D.6) shows that while  $\gamma^2 = 1/(1 - \tilde{\lambda})^2$  is identified, the sign of  $\gamma = 1/(1 - \tilde{\lambda})$  is not identified, unless  $\tilde{\lambda}$  is constrained to take values in  $(-\infty, 1)$ . The reason is that the function  $1/(1 - \tilde{\lambda})^2$  is monotonically increasing on the interval  $(-\infty, 1)$  and  $(1 - \tilde{\lambda}) > 0$  for  $\tilde{\lambda} \in (-\infty, 1)$ . The implied range for  $\gamma$  then is  $(0, \infty)$  and the permissible parametrizations of the term  $\bar{\epsilon}_r$  in (D.1) is  $(-1, \infty)$ . For the latter, the positive values are most relevant in peer effects applications.

## E Proofs of Theorems 4.1, 4.2, and 4.3

In this section, we collect the proof of Theorem 4.1 in Sections E.1 and E.2. This theorem establishes the identification and the consistency of the quasi-maximum likelihood estimator. Then we provide a proof of Theorem 4.2 and Theorem 4.3 in Section E.3 and Section E.4, respectively. These theorems establish the asymptotic distribution of the QMLE and the consistency of our estimators for the third and fourth moments.

### E.1 Proof of Theorem 4.1(a)

For the un-concentrated log likelihood function in (16), let

$$\bar{R}(\theta, \beta) = \lim_{N \rightarrow \infty} E \left[ \frac{1}{N} \ln L(\theta, \beta) \right].$$

Let  $\bar{\beta}(\theta)$  be the maximizer of  $\bar{R}(\theta, \beta)$  with respect to  $\beta$ ,

$$\bar{R}(\theta, \bar{\beta}(\theta)) = \max_{\beta} \bar{R}(\theta, \beta),$$

and let

$$\bar{Q}^{**}(\theta) = \bar{R}(\theta, \bar{\beta}(\theta)).$$

For the concentrated log likelihood function in (20), let  $\bar{Q}^*(\theta) = \lim_{N \rightarrow \infty} E [Q_N(\theta)]$ . To prove that  $\theta_0$  is identifiably unique, it suffices to show that Condition 1(a) and Condition 2 below hold; cp., e.g., Definition 3.1 of identifiable uniqueness and the subsequent discussion in Pötscher and Prucha (1991). In fact, under Condition 1(a) and Condition 2 the identifiable uniqueness of the parameter vector  $\theta_0$  is equivalent with  $\theta_0$  being asymptotically identified in the sense that it is the unique maximizer of  $\bar{Q}^*(\theta)$ .

**Condition 1.** (a) The non-stochastic functions  $\bar{Q}^*(\theta)$  and  $\bar{Q}^{**}(\theta)$  exist, and  $\bar{Q}^*(\theta) = \bar{Q}^{**}(\theta)$  are continuous and finite ;

(b) As  $N$  goes to infinity,  $\sup_{\theta \in \Theta} |E [Q_N(\theta)] - \bar{Q}^*(\theta)| \rightarrow 0$

**Condition 2.** The parameter space  $\Theta$  is compact, the true value  $\theta_0$  is the unique maximizer of  $\bar{Q}^*(\theta)$  (and hence  $\bar{Q}^{**}(\theta)$ ) on  $\Theta$  and  $\bar{\beta}(\theta_0) = \beta_0$ .

Condition 1(b) is used for the proof of consistency that is presented in Section E.2 below. Given Condition 1(b) and the identifiable uniqueness of the true parameter vector consistency of the QMLE follows immediately from, e.g., Lemma 3.1 in Poetscher and Prucha (1997), p. 16.

We combine conditions 1(a) and 1(b) as they can be established together.

■ **Verification of Condition 1:** To prove that  $\bar{Q}^*(\theta) = \lim_{N \rightarrow \infty} E [Q_N(\theta)]$  exists, it is readily

seen that

$$\begin{aligned}
E [Q_N(\theta)] &= -\frac{\ln(2\pi)}{2} + \frac{1}{2N} \ln |(I - \lambda W)^2 \Omega(\theta)^{-1}| \\
&\quad - \frac{1}{2N} \text{tr} \{ (I - \lambda W)' M_Z(\theta) (I - \lambda W) (E [YY']) \} \\
&= -\frac{\ln(2\pi)}{2} + \frac{1}{2N} \ln |(I - \lambda W)^2 \Omega(\theta)^{-1}| \\
&\quad - \frac{1}{2N} \text{tr} \left[ (I - \lambda W)' M_Z(\theta) (I - \lambda W) (I - \lambda_0 W)^{-1} (\Omega_0 + Z \beta_0 \beta_0' Z') (I - \lambda_0 W)^{-1} \right] \\
&= \bar{Q}_N^{(1)}(\theta) + \bar{Q}_N^{(2)}(\theta) + \bar{Q}_N^{(3)}(\theta),
\end{aligned} \tag{E.1}$$

with

$$\begin{aligned}
\bar{Q}_N^{(1)}(\theta) &= -\frac{\ln(2\pi)}{2} + \frac{1}{2N} \ln |(I - \lambda W)^2 \Omega(\theta)^{-1}| - \frac{1}{2N} \text{tr} \left[ (I - \lambda_0 W)^{-2} (I - \lambda W)^2 \Omega(\theta)^{-1} \Omega_0 \right], \\
\bar{Q}_N^{(2)}(\theta) &= -\frac{1}{2N} \text{tr} \left[ \beta_0' Z' (I - \lambda_0 W)^{-1} (I - \lambda W) M_Z(\theta) (I - \lambda W) (I - \lambda_0 W)^{-1} Z \beta_0 \right], \\
\bar{Q}_N^{(3)}(\theta) &= \frac{1}{2N} \text{tr} \left[ (I - \lambda_0 W)^{-2} (I - \lambda W)^2 \Omega(\theta)^{-1} Z (Z' \Omega(\theta)^{-1} Z)^{-1} Z' \Omega(\theta)^{-1} \Omega_0 \right],
\end{aligned}$$

recalling that  $M_Z(\theta) = \Omega(\theta)^{-1} - \Omega(\theta)^{-1} Z (Z' \Omega(\theta)^{-1} Z)^{-1} Z' \Omega(\theta)^{-1}$  and that the matrices  $(I - \lambda_0 W), (I - \lambda W), \Omega(\theta)^{-1}$  and  $\Omega_0$  all commute.

We show that the limits of  $\bar{Q}_N^{(1)}(\theta)$ ,  $\bar{Q}_N^{(2)}(\theta)$  and  $\bar{Q}_N^{(3)}(\theta)$  exist, in reverse order. Observe that

$$\bar{Q}_N^{(3)}(\theta) = \frac{1}{2N} \text{tr} \left[ \left( \frac{1}{N} Z' \Omega(\theta)^{-1} \Omega_0 (I - \lambda_0 W)^{-2} (I - \lambda W)^2 \Omega(\theta)^{-1} Z \right) \left( \frac{1}{N} Z' \Omega(\theta)^{-1} Z \right)^{-1} \right].$$

Both matrices in square brackets are of the form considered in (B.8) with  $p(m_r, D_r, \theta)$  and  $s(m_r, D_r, \theta)$  satisfying the assumptions of Lemma B.5. Thus their elements, and in turn the trace, are bounded in absolute value by respective constants that do not depend on  $\theta$  and  $N$ . Consequently  $\sup_{\theta \in \Theta} \bar{Q}_N^{(3)}(\theta) \leq \text{const}/N \rightarrow 0$  as  $N \rightarrow \infty$  and  $\lim_{N \rightarrow \infty} \bar{Q}_N^{(3)}(\theta) = 0$ .

Second, observe that

$$\begin{aligned}
2\bar{Q}_N^{(2)}(\theta) &= \beta_0' \left( \frac{1}{N} Z' (I - \lambda_0 W)^{-2} (I - \lambda W)^2 \Omega(\theta)^{-1} Z \right) \beta_0 \\
&\quad - \beta_0' \left\{ \left( \frac{1}{N} Z' (I - \lambda_0 W)^{-1} (I - \lambda W) \Omega(\theta)^{-1} Z \right) \left( \frac{1}{N} Z' \Omega(\theta)^{-1} Z \right)^{-1} \right. \\
&\quad \left. \times \left( \frac{1}{N} Z' \Omega(\theta)^{-1} (I - \lambda W) (I - \lambda_0 W)^{-1} Z \right) \right\} \beta_0.
\end{aligned} \tag{E.3}$$

In light of (B.3) and (B.4) and using Lemma B.5 we see that  $\sup_{\theta \in \Theta} |\bar{Q}_N^{(2)}(\theta) - \bar{Q}^{(2)*}(\theta)| \rightarrow_p 0$ ,

where

$$\begin{aligned}\bar{Q}^{(2)*}(\theta) &= \frac{1}{2}\beta_0' \left( \Upsilon_1(\theta) - \Upsilon_2(\theta)\Upsilon_3^{-1}(\theta)\Upsilon_2(\theta) \right) \beta_0, \\ \Upsilon_1(\theta) &= \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \left( \frac{\phi_W^2(m, \theta)}{\phi_W^2(m, \theta_0)\phi_\Omega(m, j, \theta)} \check{z}_{m,j} + \frac{\psi_W^2(m, \theta)}{\psi_W^2(m, \theta_0)\psi_\Omega(m, j, \theta)} \bar{z}_{m,j} \right),\end{aligned}\quad (\text{E.4})$$

$$\Upsilon_2(\theta) = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \left( \frac{\phi_W(m, \theta)}{\phi_W(m, \theta_0)\phi_\Omega(m, j, \theta)} \check{z}_{m,j} + \frac{\psi_W(m, \theta)}{\psi_W(m, \theta_0)\psi_\Omega(m, j, \theta)} \bar{z}_{m,j} \right), \quad (\text{E.5})$$

$$\Upsilon_3(\theta) = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \left( \frac{1}{\phi_\Omega(m, j, \theta)} \check{z}_{m,j} + \frac{1}{\psi_\Omega(m, j, \theta)} \bar{z}_{m,j} \right), \quad (\text{E.6})$$

and where  $\bar{Q}^{(2)*}(\theta)$  is finite and continuous on  $\Theta$  by Lemma B.5.

Third, observe that

$$\begin{aligned}\bar{Q}_N^{(1)}(\theta) &= -\frac{\ln(2\pi)}{2} - \frac{1}{2N} \ln |(I - \lambda_0 W)^{-2} \Omega_0| \\ &\quad + \frac{1}{2N} \ln |(I - \lambda_0 W)^{-2} (I - \lambda W)^2 \Omega(\theta)^{-1} \Omega_0| - \frac{1}{2N} \text{tr} \left[ (I - \lambda_0 W)^{-2} (I - \lambda W)^2 \Omega(\theta)^{-1} \Omega_0 \right] \\ &= C_N + \frac{1}{2} \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \frac{R_{m,j}}{N} \ln |G(m, j, \theta)| - \frac{1}{2} \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \frac{R_{m,j}}{N} \text{tr}[G(m, j, \theta)]\end{aligned}$$

with

$$\begin{aligned}G(m, j, \theta) &= (I_m - \lambda W_m)^2 \Omega_{m,j}(\theta)^{-1} (I_m - \lambda_0 W_m)^{-2} \Omega_{m,j0} \\ &= \frac{\phi_W^2(m, \theta)\phi_\Omega(m, j, \theta_0)}{\phi_W^2(m, \theta_0)\phi_\Omega(m, j, \theta)} I_m^* + \frac{\psi_W^2(m, \theta)\psi_\Omega(m, j, \theta_0)}{\psi_W^2(m, \theta_0)\psi_\Omega(m, j, \theta)} J_m^*, \\ &= \frac{\sigma_{\epsilon_0,j}^2}{\sigma_{\epsilon,j}^2} \left( \frac{m-1+\lambda}{m-1+\lambda_0} \right)^2 I_m^* + \frac{(\sigma_{\epsilon_0,j}^2 + m\sigma_{\alpha_0}^2)}{(\sigma_{\epsilon,j}^2 + m\sigma_{\alpha}^2)} \left( \frac{1-\lambda}{1-\lambda_0} \right)^2 J_m^*,\end{aligned}\quad (\text{E.7})$$

and  $C_N = -\frac{\ln(2\pi)}{2} - \frac{1}{2} \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \frac{R_{m,j}}{N} \ln |(I - \lambda_0 W_m)^{-2} \Omega_{m,j0}|$ . Under Assumption 4,  $R_{m,j}/N \rightarrow \omega_{m,j}^*/m^*$ . Let  $C^* = \lim_{N \rightarrow \infty} C_N$  and

$$\bar{Q}^{(1)*}(\theta) = C^* + \frac{1}{2m^*} \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta) \quad (\text{E.8})$$

with

$$g(m, j, \theta) = \ln |G(m, j, \theta)| - \text{tr}[G(m, j, \theta)],$$

then clearly  $\sup_{\theta \in \Theta} |\bar{Q}_N^{(1)}(\theta) - \bar{Q}^{(1)*}(\theta)| \rightarrow 0$  with  $\bar{Q}^{(1)*}(\theta)$  finite and continuous on  $\Theta$ . In all,

$\sup_{\theta \in \Theta} |E[Q_N(\theta)] - \bar{Q}^*(\theta)| \rightarrow 0$ , where

$$\bar{Q}^*(\theta) = \bar{Q}^{(1)*}(\theta) + \bar{Q}^{(2)*}(\theta) = C^* + \frac{1}{2m^*} \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta) + \bar{Q}^{(2)*}(\theta),$$

and where  $\bar{Q}^*(\theta)$  is continuous and finite.

For the un-concentrated likelihood function,

$$\begin{aligned} \bar{R}(\theta, \beta) &= \lim_{N \rightarrow \infty} \frac{1}{N} E[\ln L_n(\theta, \beta)] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \left\{ -\frac{N}{2} \ln(2\pi) + \frac{1}{2} \ln |(I - \lambda W)^2 \Omega(\theta)^{-1}| \right. \\ &\quad \left. - \frac{1}{2} \text{tr} \left[ (I - \lambda_0 W)^{-2} (I - \lambda W)^2 \Omega(\theta)^{-1} \Omega_0 \right] \right. \\ &\quad \left. - \frac{1}{2} \left( (I - \lambda W)(I - \lambda_0 W)^{-1} Z \beta_0 - Z \beta \right)' \Omega(\theta)^{-1} \left( (I - \lambda W)(I - \lambda_0 W)^{-1} Z \beta_0 - Z \beta \right) \right\} \\ &= \lim_{N \rightarrow \infty} \bar{Q}_N^{(1)}(\theta) - \lim_{N \rightarrow \infty} \frac{1}{2N} \left( \beta_0' Z' (I - \lambda W)^2 (I - \lambda_0 W)^2 \Omega(\theta)^{-1} Z \beta_0 \right) \\ &\quad + \lim_{N \rightarrow \infty} \frac{1}{N} \beta_0' Z' (I - \lambda W)(I - \lambda_0 W) \Omega(\theta)^{-1} Z \beta - \lim_{N \rightarrow \infty} \frac{1}{2N} \beta' Z' \Omega(\theta)^{-1} Z \beta \\ &= \bar{Q}^{(1)*}(\theta) - \frac{1}{2} \beta_0' \Upsilon_1(\theta) \beta_0 + \beta_0' \Upsilon_2(\theta) \beta - \frac{1}{2} \beta' \Upsilon_3(\theta) \beta, \end{aligned}$$

where  $\Upsilon_1(\theta)$ ,  $\Upsilon_2(\theta)$ ,  $\Upsilon_3(\theta)$  are defined in Equations (E.4), (E.5), and (E.6). Taking the derivative of  $\bar{R}(\theta, \beta)$  with respect to  $\beta$ ,

$$\frac{\partial \bar{R}(\theta, \bar{\beta})}{\partial \beta} = \beta_0' \Upsilon_2(\theta) - \bar{\beta}' \Upsilon_3(\theta) = 0.$$

Since  $\Upsilon_3(\theta)$  is non-singular by Assumption 5 and Lemma B.5,

$$\bar{\beta}(\theta) = \Upsilon_3(\theta)^{-1} \Upsilon_2(\theta) \beta_0. \tag{E.9}$$

Let  $\bar{Q}^{**}(\theta) = \bar{R}(\theta, \bar{\beta}(\theta))$  and plug  $\bar{\beta}(\theta)$  above back into  $\bar{R}(\theta, \beta)$ ,

$$\begin{aligned} \bar{Q}^{**}(\theta) &= \bar{Q}^{(1)*}(\theta) - \frac{1}{2} \beta_0' \left( \Upsilon_1(\theta) - \Upsilon_2(\theta) \Upsilon_3^{-1}(\theta) \Upsilon_2(\theta) \right) \beta_0 \\ &= \bar{Q}^{(1)*}(\theta) + \bar{Q}^{(2)*}(\theta) = \bar{Q}^*(\theta). \end{aligned}$$

Note that the second order derivative

$$\frac{\partial^2 \bar{R}(\theta, \beta)}{\partial \beta \partial \beta'} = -\Upsilon_3(\theta) = - \lim_{N \rightarrow \infty} Z' \Omega(\theta)^{-1} Z$$

is negative definite by Assumption 5 and Lemma B.5 uniformly in  $\theta$ , thus  $\bar{\beta}(\theta)$  is the unique maximizer of  $\bar{R}(\theta, \beta)$  over  $\beta$ . In all, we have  $\bar{Q}^*(\theta)$  and  $\bar{Q}^{**}(\theta)$  both exist and  $\bar{Q}^*(\theta) = \bar{Q}^{**}(\theta)$ .

■ **Verification of Condition 2** Since  $\Upsilon_3(\theta_0) = \Upsilon_2(\theta_0)$ ,  $\bar{\beta}(\theta_0) = \beta_0$  is readily seen. Next we show that  $\theta_0$  is the unique global maximizer of  $\bar{Q}^*(\theta)$  on  $\Theta$ . We first show that  $\theta_0$  is a global maximizer of



$\bar{Q}^{(2)*}(\theta)$ . To see this observe that we can rewrite  $Q_N^{(2)}(\theta)$  as  $Q_N^{(2)}(\theta) = -\frac{1}{2N}\tilde{\eta}_Z(\theta)'\tilde{M}_Z(\theta)\tilde{\eta}_Z(\theta)$ , where  $\tilde{M}_Z(\theta) = I - \Omega^{-1/2}Z'(Z'\Omega(\theta)Z)^{-1}Z\Omega^{-1/2}$  is idempotent and positive semidefinite, and  $\tilde{\eta}_Z(\theta) = \Omega(\theta)^{-1/2}(I - \lambda W)(I - \lambda_0 W)^{-1}Z\beta_0$ . Thus  $Q_N^{(2)}(\theta) \leq 0$  and consequently also  $\bar{Q}^{(2)*}(\theta) \leq 0$ . Next observe, as is readily checked, that  $Q^{(2)*}(\theta_0) = 0$ . Therefore  $\bar{Q}^{(2)*}(\theta) \leq \bar{Q}^{(2)*}(\theta_0)$  for all  $\theta \in \Theta$ .

To show that  $\theta_0$  is the unique global maximizer of  $\bar{Q}^*(\theta)$  it thus suffices to show that  $\theta_0$  is the unique maximizer of  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta)$ . Observe that

$$\sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta) = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* (\ln |G(m, j, \theta)| - \text{tr}[G(m, j, \theta)]) \leq -m^*, \quad (\text{E.10})$$

where  $m^* = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* m$ .<sup>10</sup> The equality holds if and only if  $g(m, j, \theta) = -1$  or equivalently  $G(m, j, \theta) = I_m$  for all  $m$  and  $j$  with  $\omega_{m,j}^* > 0$ . Under the two scenarios described by Assumption 6 this is the case if and only if  $\theta = \theta_0$ , which establishes that  $\theta_0$  is the unique maximizer of  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta)$ . To see this, observe that in light of (E.7) the equality  $G(m, j, \theta) = I_m$  only holds if

$$\left( \frac{m-1+\lambda}{m-1+\lambda_0} \right)^2 \frac{\sigma_{\epsilon_0,j}^2}{\sigma_{\epsilon,j}^2} = 1, \quad (\text{E.11})$$

$$\frac{(\sigma_{\epsilon_0,j}^2 + m\sigma_{\alpha_0}^2)}{(\sigma_{\epsilon,j}^2 + m\sigma_{\alpha}^2)} \left( \frac{1-\lambda}{1-\lambda_0} \right)^2 = 1. \quad (\text{E.12})$$

Note that (E.11) and (E.12) are equivalent to  $E[\chi_r^w(\theta)|m_r = m, D_r = j] = 0$  and  $E[\chi_r^b(\theta)|m_r = m, D_r = j] = 0$  respectively, where  $\chi_r^w(\theta)$  and  $\chi_r^b(\theta)$  are defined in (7). See Equations (C.2) and (C.4) in Appendix C. Thus mathematically  $G(m, j, \theta) = I_m$  is equivalent to  $E[\chi_r(\theta)|m_r = m, D_r = j] = 0$ , with  $\chi_r(\theta) = (\chi_r^w(\theta), \chi_r^b(\theta))$ . Utilizing Lemma 2.1,  $\theta = \theta_0$  is the only solution to  $E[\chi_r(\theta)|m_r = m, D_r = j] = 0$  and  $E[\chi_r(\theta)|m_r = m', D_r = j'] = 0$  under Scenarios (i) or (ii). Therefore,  $\theta_0$  is the unique global maximizer of  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} \omega_{m,j}^* g(m, j, \theta)$  and thus of  $\bar{Q}^*(\theta)$ .

## E.2 Proof of Theorem 4.1(b)

To prove the consistency of the QMLE estimator  $\hat{\theta}_N$  we utilize Lemma 3.1 of Pötscher and Prucha (1991). Previously we have shown that  $\theta_0$  is the unique maximizer of  $\bar{Q}^*(\theta)$  on  $\Theta$ , where  $\bar{Q}^*(\theta)$  is finite and continuous. The compactness of  $\Theta$  follows from Assumptions 1, 2, and 3. To prove consistency of  $\hat{\theta}$ , it then suffices to have Condition 3. Since  $\bar{\beta}(\theta_0) = \beta_0$ , once we have shown that  $\hat{\theta}_N \rightarrow_p \theta_0$ , consistency of  $\hat{\beta}_N(\hat{\theta}_N)$  follows from Condition 4.

**Condition 3.** As  $N \rightarrow \infty$ ,  $\sup_{\theta \in \Theta} |Q_N(\theta) - \bar{Q}^*(\theta)| \rightarrow_p 0$ .

**Condition 4.** As  $N \rightarrow \infty$ ,  $\sup_{\theta \in \Theta} |\hat{\beta}_N(\theta) - \bar{\beta}(\theta)| \rightarrow_p 0$ .

■ **Verification of Condition 3:** Verification of Condition 1 has shown that  $\sup_{\theta \in \Theta} |E[Q_N(\theta)] - \bar{Q}^*(\theta)| \rightarrow 0$  as  $N \rightarrow \infty$ . It remains to show that as  $N$  goes to infinity,  $\sup_{\theta \in \Theta} |Q_N(\theta) - E[Q_N(\theta)]| \rightarrow_p 0$

<sup>10</sup>See Footnote 7 for details.

0 . Upon substitution of  $Y = (I - \lambda_0 W)^{-1}(Z\beta_0 + U)$  into (20) we have

$$Q_N(\theta) - E[Q_N(\theta)] = \frac{1}{N} (U' A_{Q_N}(\theta) U + 2U' A_{Q_N}(\theta) Z\beta_0 - \text{tr}[A_{Q_N}(\theta)\Omega_0]),$$

where

$$A_{Q_N}(\theta) = -\frac{1}{2}(I - \lambda_0 W)^{-1}(I - \lambda W)' M_Z(\theta)(I - \lambda W)(I - \lambda_0 W)^{-1}.$$

The row and column sums in absolute value of  $(I - \lambda_0 W)^{-1}$ ,  $(I - \lambda W)$ ,  $M_Z(\theta)$  and their first derivatives are all uniformly bounded in absolute value. It now follows from Lemma B.3 that  $Q_N(\theta) - E[Q_N(\theta)] \rightarrow_p 0$  uniformly in  $\theta$ .

■ **Verification of Condition 4:**

In light of (23) we have

$$\begin{aligned} \hat{\beta}_N(\theta) &= (Z'\Omega(\theta)^{-1}Z)^{-1}Z'\Omega(\theta)^{-1}(I - \lambda W)Y \\ &= \left(N^{-1}Z'\Omega(\theta)^{-1}Z\right)^{-1} \left(N^{-1}Z'\Omega(\theta)^{-1}(I - \lambda W)(I - \lambda_0 W)^{-1}Z\right) \beta_0 \\ &\quad + \left(N^{-1}Z'\Omega(\theta)^{-1}Z\right)^{-1} \left(N^{-1}Z'\Omega(\theta)^{-1}(I - \lambda W)(I - \lambda_0 W)^{-1}U\right). \end{aligned}$$

By Lemma B.3,

$$\sup_{\theta \in \Theta} \left(N^{-1}Z'\Omega(\theta)^{-1}(I - \lambda W)(I - \lambda_0 W)^{-1}U\right) \rightarrow_p 0.$$

Also  $(N^{-1}Z'\Omega(\theta_N)^{-1}Z)^{-1}$  is uniformly bounded in absolute value. By Lemma B.5,

$$\sup_{\theta \in \Theta} \left(\left(N^{-1}Z'\Omega(\theta)^{-1}Z\right)^{-1} - \mathcal{R}_3(\theta)^{-1}\right) \rightarrow 0$$

and

$$\sup_{\theta \in \Theta} \left(N^{-1}Z'\Omega(\theta)^{-1}(I - \lambda W)(I - \lambda_0 W)^{-1}Z - \mathcal{R}_2(\theta)\right) \rightarrow 0.$$

In all, we have  $\sup_{\theta \in \Theta} |\hat{\beta}_N(\theta) - \bar{\beta}(\theta)| \rightarrow_p 0$ .

### E.3 Proof of Theorem 4.2

To derive the limiting distribution of the QMLE  $\hat{\delta}_N = (\hat{\theta}'_N, \hat{\beta}'_N)'$  it proves more convenient to work with the unconcentrated log-likelihood function defined in (16). Applying the mean value theorem, the first order condition for the QMLE can be written as

$$0 = \frac{1}{N^{1/2}} \frac{\partial \ln L_N(\hat{\delta}_N)}{\partial \delta} = \frac{1}{N^{1/2}} \frac{\partial \ln L_N(\delta_0)}{\partial \delta} + \frac{1}{N} \frac{\partial \ln L_N(\check{\delta}_N)}{\partial \delta \partial \delta} N^{1/2}(\hat{\delta}_N - \delta_0),$$

where  $\check{\delta}_N$  denotes a ‘‘between’’ value vector. Given that  $\hat{\delta}_N$  was shown to be consistent, it follows that also the ‘‘between’’ value  $\check{\delta}_N$  is consistent for  $\delta_0$ . It is not difficult to see that

$$\frac{\partial \ln L_N(\delta)}{\partial \delta} = \begin{bmatrix} -\text{tr}[(I - \lambda W)^{-1}W] + U(\delta)' \Omega^{-1} W Y \\ -\frac{1}{2} \text{tr}[\Omega^{-1} \text{diag}_{r=1}^R \{m_r J_{m_r}^*\}] + \frac{1}{2} U(\delta)' \Omega^{-1} \text{diag}_{r=1}^R \{m_r J_{m_r}^*\} \Omega^{-1} U(\delta) \\ -\frac{1}{2} \text{tr}[\Omega^{-1} \text{diag}_{r=1}^R \{1(D_r = 1)I_{m_r}\}] + \frac{1}{2} U(\delta)' \Omega^{-2} \text{diag}_{r=1}^R \{1(D_r = 1)I_{m_r}\} U(\delta) \\ \vdots \\ -\frac{1}{2} \text{tr}[\Omega^{-1} \text{diag}_{r=1}^R \{1(D_r = J)I_{m_r}\}] + \frac{1}{2} U(\delta)' \Omega^{-2} \text{diag}_{r=1}^R \{1(D_r = J)I_{m_r}\} U(\delta) \\ Z' \Omega^{-1} U(\delta) \end{bmatrix}$$

with  $U(\delta) = Y - \lambda W Y - Z\beta$ , and thus

$$\frac{\partial \ln L_N(\delta_0)}{\partial \delta} = \begin{bmatrix} -\text{tr}[(I - \lambda_0 W)^{-1}W] + U' \Omega_0^{-1} W (I - \lambda_0 W)^{-1} (Z\beta_0 + U) \\ -\frac{1}{2} \text{tr}[\Omega_0^{-1} \text{diag}_{r=1}^R \{J_{m_r}^* m_r\}] + \frac{1}{2} U' \Omega_0^{-1} \text{diag}_{r=1}^R \{J_{m_r}^* m_r\} \Omega_0^{-1} U \\ -\frac{1}{2} \text{tr}[\Omega_0^{-1} \text{diag}_{r=1}^R \{1(D_r = 1)I_{m_r}\}] + \frac{1}{2} U' \Omega_0^{-2} \text{diag}_{r=1}^R \{1(D_r = 1)I_{m_r}\} U \\ \vdots \\ -\frac{1}{2} \text{tr}[\Omega_0^{-1} \text{diag}_{r=1}^R \{1(D_r = J)I_{m_r}\}] + \frac{1}{2} U' \Omega_0^{-2} \text{diag}_{r=1}^R \{1(D_r = J)I_{m_r}\} U \\ Z' \Omega_0^{-1} U \end{bmatrix}. \quad (\text{E.13})$$

Furthermore, it is not difficult to see that with  $\theta_2 = \sigma_\alpha^2$ ,  $\theta_{2+j} = \sigma_{\epsilon,j}^2$ ,  $j = 1, \dots, J$ , the elements of the Hessian matrix are

$$\frac{\partial^2 \ln L_N(\delta)}{\partial \lambda^2} = -\text{tr}[(I - \lambda W)^{-2} W^2] - Y' W' \Omega(\theta)^{-1} W Y, \quad (\text{E.14})$$

$$\frac{\partial^2 \ln L_N(\delta)}{\partial \lambda \partial \theta_i} = -U(\delta)' \Omega(\theta)^{-1} \frac{\partial \Omega(\theta)}{\partial \theta_i} \Omega(\theta)^{-1} W Y, \quad (\text{E.15})$$

$$\frac{\partial^2 \ln L_N(\delta)}{\partial \theta_i \partial \theta_j} = \frac{1}{2} \text{tr}[\Omega(\theta)^{-2} \frac{\partial \Omega}{\partial \theta_i} \frac{\partial \Omega}{\partial \theta_j}] \quad (\text{E.16})$$

$$-U(\delta)' \Omega(\theta)^{-1} \frac{\partial \Omega(\theta)}{\partial \theta_i} \Omega(\theta)^{-1} \frac{\partial \Omega(\theta)}{\partial \theta_j} \Omega(\theta)^{-1} U(\delta), \quad (\text{E.17})$$

$$\frac{\partial^2 \ln L_N(\delta)}{\partial \theta_i \partial \beta} = -Z' \Omega(\theta)^{-1} \frac{\partial \Omega(\theta)}{\partial \theta_i} \Omega(\theta)^{-1} U(\delta), \quad (\text{E.18})$$

$$\frac{\partial^2 \ln L_N(\delta)}{\partial \lambda \partial \beta} = -Z' \Omega(\theta)^{-1} W Y, \quad (\text{E.19})$$

$$\frac{\partial^2 \ln L_N(\delta)}{\partial \beta \partial \beta'} = -Z' \Omega(\theta)^{-1} Z, \quad (\text{E.20})$$

with  $i, j = 2, 3, \dots, 2 + J$  and

$$\frac{\partial \Omega(\theta)}{\partial \theta_2} = \text{diag}_{r=1}^R \{J_{m_r}^* m_r\}, \quad \frac{\partial \Omega(\theta)}{\partial \theta_{2+j}} = \text{diag}_{r=1}^R \{1(D_r = j)I_{m_r}\}.$$

Since  $Y = (I - \lambda_0 W)^{-1} (Z\beta_0 + U)$  and  $U(\delta) = (I - \lambda W)(I - \lambda_0 W)^{-1} (Z\beta_0 + U) - Z\beta$ , each

element of  $N^{-1}\partial^2 \ln L_N(\delta)/\partial\delta\partial\delta'$  is a linear quadratic form of  $U$  or  $Z$  in the form of  $\frac{1}{N}U'A(\theta)U$ ,  $\frac{1}{N}Z'A(\theta)U$ ,  $\frac{1}{N}Z'A(\theta)Z$ , and their products with  $\beta$ :  $\frac{1}{N}\beta'Z'A(\theta)U$ ,  $\frac{1}{N}Z'A(\theta)Z\beta$ ,  $\frac{1}{N}\beta'Z'A(\theta)Z\beta$ , etc., where  $A(\theta) = \text{diag}_{r=1}^R \{p(m_r, D_r, \theta)I_{m_r}^* + s(m_r, D_r, \theta)J_{m_r}^*\}$  satisfies the conditions in Lemma B.3 and Lemma B.5. By these two lemmas, if  $\hat{\theta}_N \rightarrow_p \theta_0$ , all three types of linear quadratic forms converge to the limit of their expected value at  $\theta_0$  in probability. That is as  $N \rightarrow \infty$ ,

$$\begin{aligned} \left| \frac{1}{N}U'A(\hat{\theta}_N)U - \lim_{N \rightarrow \infty} \frac{1}{N} \text{tr}[A(\theta_0)\Omega_0] \right| &\rightarrow_p 0, \\ \left| \frac{1}{N}Z'A(\hat{\theta}_N)U \right| &\rightarrow_p 0, \\ \left| \frac{1}{N}Z'A(\hat{\theta}_N)Z - \lim_{N \rightarrow \infty} \frac{1}{N}Z'A(\theta_0)Z \right| &\rightarrow_p 0. \end{aligned}$$

Also, we have  $\hat{\beta}_N \rightarrow_p \beta_0$ . Thus by Slutsky's theorem, the products of the linear quadratic forms with  $\hat{\beta}_N$  converge in probability to the products of the expected values with  $\beta_0$ . Therefore, as  $\check{\delta}_N \rightarrow_p \delta_0$ ,

$$\frac{1}{N} \frac{\partial \ln L_N(\check{\delta}_N)}{\partial\delta\partial\delta'} \rightarrow_p \lim_{N \rightarrow \infty} \frac{1}{N} E \left[ \frac{\partial^2 \ln L_N(\delta_0)}{\partial\delta_0\partial\delta_0'} \right] = -\Gamma_0,$$

where the specific structure of  $\Gamma_0$  is given in Appendix F.

We next show that  $N^{1/2}\partial \ln L_N(\delta_0)/\partial\delta \xrightarrow{d} N(0, \Upsilon_0)$ . Each element of the score function in (E.13) can be written as a linear quadratic form of  $U$  in the form of  $U'A_N(\theta_0)U + U'B_N(\theta_0)Z\beta_0 + C(\delta_0)$ , which has zero mean and where the row and column sums of  $A_N(\theta_0)$  and  $B_N(\theta_0)$  are uniformly bounded in absolute value and where  $C(\delta_0)$  are constants. Using Theorem B.2,  $N^{-1/2}\partial \ln L_N(\delta_0)/\partial\delta \xrightarrow{d} N(0, \Upsilon_0)$  with  $\Upsilon_0 = \lim_{N \rightarrow \infty} \frac{1}{N} E \left[ \frac{\partial \ln L_N(\delta_0)}{\partial\delta} \frac{\partial \ln L_N(\delta_0)}{\partial\delta'} \right]$ , whose expression is given in Appendix F. In all,  $\sqrt{N}(\hat{\delta}_N - \delta_0) \xrightarrow{d} N(0, \Gamma_0^{-1}\Upsilon_0\Gamma_0^{-1})$  as  $N$  goes to infinity.

#### E.4 Proof of Theorem 4.3

The proof of the theorem follows from the definition of  $\hat{\Gamma}_N$  and  $\hat{\Upsilon}_N$  in (27) and (28), Lemma E.1 which is stated below in this section, Assumptions 4 and 5 and Theorem 4.1 which together imply that  $\hat{\Gamma}_N \xrightarrow{p} \Gamma_0$  and  $\hat{\Upsilon}_N \xrightarrow{p} \Upsilon_0$ . Since by Lemma 4.1 the matrices  $\Upsilon_0$  and  $\Gamma_0$  are full rank, and thus  $\Gamma_0^{-1}\Upsilon_0\Gamma_0^{-1}$  is full rank, it follows that  $(\hat{\Gamma}_N^{-1}\hat{\Upsilon}_N\hat{\Gamma}_N^{-1})^{-1/2} \xrightarrow{p} (\Gamma_0^{-1}\Upsilon_0\Gamma_0^{-1})^{-1/2}$ . The result then follows from the continuous mapping theorem and Theorem 4.2.

**Lemma E.1.** *Suppose Assumptions 1-5 hold, then  $\hat{\mu}_\alpha^{(3)} \xrightarrow{p} \mu_{\alpha 0}^{(3)}$ ,  $\hat{\mu}_\alpha^{(4)} \xrightarrow{p} \mu_{\alpha 0}^{(4)}$ , and  $\hat{\mu}_{\epsilon, j}^{(3)} \xrightarrow{p} \mu_{\epsilon 0, j}^{(3)}$ ,  $\hat{\mu}_{\epsilon, j}^{(4)} \xrightarrow{p} \mu_{\epsilon 0, j}^{(4)}$ , for any  $j \in \{1, \dots, J\}$ .*

First note that  $f_{\alpha, r}^{(3)}$ ,  $f_{\alpha, r}^{(4)}$ ,  $f_{\epsilon, r}^{(3)}$  and  $f_{\epsilon, r}^{(4)}$  can all be rewritten as a linear combination of finitely many terms in the form of  $\psi(m_r) \sum_{i=1}^{m_r} \check{u}_{ir}^{p_1} \check{u}_{ir}^{p_2}$  and some nonstochastic function  $f(m_r, \theta)$ , with  $\psi(m_r)$  being a finite function,  $p_1 \geq 0, p_2 \geq 0$  and  $p_1 + p_2 \leq 4$ , and  $f(m, \theta)$  being continuous in  $\theta$ .

Also note that

$$\frac{1}{R} \sum_{r=1}^R f_{\alpha,r}^{(l)} = \sum_{j=1}^J \frac{R_j}{R} \left( \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) f_{\alpha,r}^{(l)} \right)$$

for  $l = 3$  and  $4$ . Thus our estimates for the third and fourth moments,  $\frac{1}{R} \sum_{r=1}^R f_{\alpha,r}^{(l)}$  and  $\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) f_{\alpha,r}^{(l)}$ ,  $l = 3, 4$  can all be written as a weighted sum of finitely many terms of the form

$$\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \psi(m_r) \sum_{i=1}^{m_r} \hat{u}_{ir}^{p_1} \bar{u}_r^{p_2}. \quad (\text{E.21})$$

and

$$\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) f(m_r, \theta). \quad (\text{E.22})$$

The former converges to its mean by Lemma B.4 (a) and the latter is nonstochastic. Consequently,  $\frac{1}{R} \sum_{r=1}^R f_{\alpha,r}^{(l)} \rightarrow_p \mu_{\alpha 0}^{(l)}$  and  $\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) f_{\alpha,r}^{(l)} \rightarrow_p \mu_{\epsilon 0, j}^{(l)}$  for  $l = 3, 4$  and  $j = 1, \dots, J$  as  $R$  goes to infinity. For the feasible counterparts of the terms in (E.21) and (E.22), note that  $\hat{\theta} - \theta_0 \rightarrow_p 0$  and  $f(m, \theta)$  is continuous in  $\theta$ . We have

$$\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) f(m_r, \hat{\theta}) - \frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) f(m_r, \theta_0) \rightarrow_p 0.$$

by the continuous mapping theorem as. It thus remains to show that

$$\frac{1}{R_j} \sum_{r=1}^R 1(D_r = j) \psi(m_r) \sum_{i=1}^{m_r} (\hat{u}_{ir}^{p_1} \hat{u}_r^{p_2} - \hat{u}_{ir}^{p_1} \bar{u}_r^{p_2}) \rightarrow_p 0. \quad (\text{E.23})$$

Let  $\hat{U}_r = (\hat{u}_{1r}, \dots, \hat{u}_{m_r, r})'$ , then

$$\begin{aligned} \hat{U}_r &= (I - \hat{\lambda}W)Y_r - Z_r \hat{\beta} = (I - \hat{\lambda}W)(I - \lambda_0 W)^{-1}(Z_r \beta_0 + U_r) - Z_r \hat{\beta} \\ &= \left( (I - \hat{\lambda}W)(I - \lambda_0 W)^{-1} Z_r \beta_0 - Z_r \hat{\beta} \right) + (I - \hat{\lambda}W)(I - \lambda_0 W)^{-1} U_r. \end{aligned}$$

Note that

$$(I - \hat{\lambda}W)(I - \lambda_0 W)^{-1} = \frac{m_r - 1 + \hat{\lambda}}{m_r - 1 + \lambda_0} I_{m_r}^* + \frac{1 - \hat{\lambda}}{1 - \lambda_0} J_{m_r}^*,$$

where  $J_{m_r}^* = \iota_{m_r} \iota_{m_r}' / m_r$  and  $I_{m_r}^* = I_{m_r} - J_{m_r}^*$  are two orthogonal idempotent matrices that generate vectors of group means and vectors of deviations from the group means. Thus

$$\hat{u}_r = \iota_{m_r}' \hat{U}_r / m_r = \bar{z}_r \left( \frac{(1 - \hat{\lambda}) \beta_0}{1 - \lambda_0} - \hat{\beta} \right) + \frac{1 - \hat{\lambda}}{1 - \lambda_0} \bar{u}_r = \bar{z}_r \bar{\phi} + \bar{\varphi} \bar{u}_r,$$

where  $\bar{\phi} = \frac{(1-\hat{\lambda})\beta_0}{1-\lambda_0} - \hat{\beta}$ ,  $\bar{\varphi} = \frac{1-\hat{\lambda}}{1-\lambda_0}$ ,  $\bar{z}_r = \frac{1}{m_r} \sum_{i=1}^{m_r} z_{ir}$ . Let  $\hat{U}_r = (\hat{u}_{1r}, \dots, \hat{u}_{m_r r})'$ , then

$$\hat{U}_r = I_{m_r}^* \hat{U}_r = \ddot{Z}_r \left( \frac{m_r - 1 + \hat{\lambda}}{m_r - 1 + \lambda_0} \beta_0 - \hat{\beta} \right) + \frac{m_r - 1 + \hat{\lambda}}{m_r - 1 + \lambda_0} \ddot{U}_r,$$

and

$$\ddot{u}_{ir} = \ddot{z}_{ir} \ddot{\phi}_r + \ddot{\varphi}_r \ddot{u}_{ir},$$

where  $\ddot{\phi}_r = \frac{m_r - 1 + \hat{\lambda}}{m_r - 1 + \lambda_0} \beta_0 - \hat{\beta}$ ,  $\ddot{\varphi}_r = 1 + \frac{\hat{\lambda} - \lambda_0}{m_r - 1 + \lambda_0}$ ,  $\ddot{z}_{ir} = z_{ir} - \bar{z}_r$ .

In all,

$$\hat{u}_{ir}^{p_1} \hat{u}_r^{p_2} - \ddot{u}_{ir}^{p_1} \ddot{u}_r^{p_2} = (\ddot{z}_{ir} \ddot{\phi}_r + \ddot{\varphi}_r \ddot{u}_{ir})^{p_1} (\bar{z}_r \bar{\phi} + \bar{\varphi} \bar{u}_r)^{p_2} - \ddot{u}_{ir}^{p_1} \ddot{u}_r^{p_2}.$$

Given that  $p_1$  and  $p_2$  are nonnegative integers with  $p_1 + p_2 \leq 4$ , the above equation can be written as a linear combination of terms of the form  $(\ddot{z}_{ir} \ddot{\phi}_r)^{s_1} (\ddot{\varphi}_r \ddot{u}_{ir})^{p_1 - s_1} (\bar{z}_r \bar{\phi})^{s_2} (\bar{\varphi} \bar{u}_r)^{p_2 - s_2}$ , and  $(\ddot{\varphi}_r \ddot{u}_{ir})^{p_1} (\bar{\varphi} \bar{u}_r)^{p_2} - \ddot{u}_{ir}^{p_1} \ddot{u}_r^{p_2}$  with  $0 \leq s_1 \leq p_1$ ,  $0 \leq s_2 \leq p_2$  and  $s_1 + s_2 \geq 1$ . The claim in (E.23) now follows immediately from Lemma B.4(b)-(c).

## F Variance-Covariance Matrix and Proof of Lemma 4.1

### F.1 Variance-Covariance Matrix

Recall that  $\Gamma_0 = \lim_{N \rightarrow \infty} -\frac{1}{N} E \left[ \frac{\partial^2 \ln L_N(\delta_0)}{\partial \delta \partial \delta'} \right]$  and  $\Upsilon_0 = \lim_{N \rightarrow \infty} \frac{1}{N} E \left[ \frac{\partial \ln L_N(\delta_0)}{\partial \delta} \frac{\partial \ln L_N(\delta_0)}{\partial \delta'} \right]$ . These matrices are of dimension  $(2 + J + k_Z) \times (2 + J + k_Z)$ , symmetric, and underlie the expression for the limiting variance covariance matrix of the QMLE estimator for  $\delta_0$ . In the following we give explicit expressions for  $\Gamma_0$  and  $\Upsilon_0$ . Detailed derivations are provided in the Online Appendix. We have

$$\Upsilon_0 = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \varphi(m, j) \bar{\Psi}(m, j) \varphi(m, j)', \quad (\text{F.1})$$

and

$$\Gamma_0 = \sum_{j=1}^J \sum_{m=2}^{\bar{M}} \varphi(m, j) \bar{\Psi}_G(m, j) \varphi(m, j)', \quad (\text{F.2})$$

where

$$\varphi(m, j) = \begin{pmatrix} \frac{1}{(m-1+\lambda_0)\sigma_{\epsilon 0, j}^2} & -\frac{m}{(1-\lambda_0)(\sigma_{\epsilon 0, j}^2+m\sigma_{\alpha 0}^2)} & \frac{1}{(m-1+\lambda_0)\sigma_{\epsilon 0, j}^2}\beta'_0 & -\frac{m}{(1-\lambda_0)(\sigma_{\epsilon 0, j}^2+m\sigma_{\alpha 0}^2)}\beta'_0 \\ 0 & -\frac{m^2}{2(\sigma_{\epsilon 0, j}^2+m\sigma_{\alpha 0}^2)^2} & 0 & 0 \\ -\frac{1(j=1)}{2\sigma_{\epsilon 0, j}^4} & -\frac{m1(j=1)}{2(\sigma_{\epsilon 0, j}^2+m\sigma_{\alpha 0}^2)^2} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ -\frac{1(j=J)}{2\sigma_{\epsilon 0, j}^4} & -\frac{m1(j=J)}{2(\sigma_{\epsilon 0, j}^2+m\sigma_{\alpha 0}^2)^2} & 0 & 0 \\ 0 & 0 & -\frac{1}{\sigma_{\epsilon 0, j}^2}I_{kZ} & -\frac{m}{\sigma_{\epsilon 0, j}^2+m\sigma_{\alpha 0}^2}I_{kZ} \end{pmatrix}, \quad (\text{F.3})$$

which is given in (24) and repeated here for the convenience of the reader,

$$\bar{\Psi}_G(m, j) = \text{diag}\{2(m-1)\sigma_{\epsilon 0, j}^4 \frac{\omega_{m, j}^*}{m^*}, 2(\sigma_{\alpha 0}^2 + \frac{\sigma_{\epsilon 0, j}^2}{m})^2 \frac{\omega_{m, j}^*}{m^*}, \sigma_{\epsilon 0}^2 \ddot{\chi}_{m, j}, (\sigma_{\alpha 0}^2 + \frac{\sigma_{\epsilon 0, j}^2}{m}) \frac{\bar{\chi}_{m, j}}{m}\}, \quad (\text{F.4})$$

$$\bar{\Psi}(m, j) = \begin{bmatrix} \bar{\Psi}_{11}(m, j) & \bar{\Psi}_{12}(m, j) \\ \bar{\Psi}_{21}(m, j) & \bar{\Psi}_{22}(m, j) \end{bmatrix}, \quad (\text{F.5})$$

with

$$\begin{aligned} \bar{\Psi}_{11}(m, j) &= \frac{\omega_{m, j}^*}{m^*} \begin{bmatrix} 2(m-1)\sigma_{\epsilon 0, j}^4 & 0 \\ 0 & 2(\sigma_{\alpha 0}^2 + \frac{\sigma_{\epsilon 0, j}^2}{m})^2 + (\mu_{\alpha 0}^{(4)} - 3\sigma_{\alpha 0}^4) \end{bmatrix} \\ &+ (\mu_{\epsilon 0, j}^{(4)} - 3\sigma_{\epsilon 0, j}^4) \frac{\omega_{m, j}^*}{m^*} \begin{bmatrix} \frac{(m-1)^2}{m} & \frac{(m-1)}{m^2} \\ \frac{(m-1)}{m^2} & \frac{1}{m^3} \end{bmatrix}, \\ \bar{\Psi}_{21}(m, j) &= \begin{bmatrix} 0 & 0 \\ \frac{m-1}{m} \mu_{\epsilon 0, j}^{(3)} \bar{z}'_{m, j} & [\mu_{\alpha 0}^{(3)} + \frac{1}{m^2} \mu_{\epsilon 0, j}^{(3)}] \bar{z}'_{m, j} \end{bmatrix} = \Psi'_{12}(m, j), \\ \bar{\Psi}_{22}(m, j) &= \begin{bmatrix} \sigma_{\epsilon 0, j}^2 \ddot{\chi}_{m, j} & 0 \\ 0 & (\sigma_{\alpha 0}^2 + \sigma_{\epsilon 0, j}^2/m) \frac{\bar{\chi}_{m, j}}{m} \end{bmatrix}. \end{aligned}$$

Note that  $\bar{\Psi}_G(m, j)$  can be obtained by setting  $\mu_{\epsilon 0, j}^{(4)} - 3\sigma_{\epsilon 0, j}^4 = \mu_{\alpha 0}^{(4)} - 3\sigma_{\alpha 0}^4 = \mu_{\alpha 0}^{(3)} = \mu_{\epsilon 0, j}^{(3)} = 0$  in  $\bar{\Psi}(m, j)$ . When  $\epsilon$  and  $\alpha$  are both Gaussian,  $\Upsilon_0 = \Gamma_0$ , consistent with what is expected from the information matrix equality.

## F.2 Proof of the Positive Definiteness of $\Upsilon_0$ and $\Gamma_0$

Let  $\varphi(m, j)$ ,  $\bar{\Psi}_G(m, j)$  and  $\bar{\Psi}(m, j)$  be as defined in (F.3), (F.4) and (F.5) respectively. We can partition  $\varphi(m, j)$  as  $\varphi(m, j) = \begin{pmatrix} A_{m,j} & B_{m,j} \\ 0 & C_{m,j} \end{pmatrix}$ , where

$$A_{m,j} = \begin{pmatrix} \frac{1}{(m-1+\lambda_0)\sigma_{\epsilon_0,j}^2} & -\frac{m}{(1-\lambda_0)(\sigma_{\epsilon_0,j}^2+m\sigma_{\alpha_0}^2)} \\ 0 & -\frac{m^2}{2(\sigma_{\epsilon_0,j}^2+m\sigma_{\alpha_0}^2)^2} \\ -\frac{1(j=1)}{2\sigma_{\epsilon_0,j}^4} & -\frac{m1(j=1)}{2(\sigma_{\epsilon_0,j}^2+m\sigma_{\alpha_0}^2)^2} \\ \vdots & \vdots \\ -\frac{1(j=J)}{2\sigma_{\epsilon_0,j}^4} & -\frac{m1(j=J)}{2(\sigma_{\epsilon_0,j}^2+m\sigma_{\alpha_0}^2)^2} \end{pmatrix} \quad (\text{F.6})$$

is a  $(2+J) \times 2$  matrix,  $B_{m,j}$  is the upper right block,

$$C_{m,j} = \left[ -\frac{1}{\sigma_{\epsilon_0,j}^2} I_{k_Z}, -\frac{m}{\sigma_{\epsilon_0,j}^2 + m\sigma_{\alpha_0}^2} I_{k_Z} \right]. \quad (\text{F.7})$$

Let  $\ell = (\ell_1, \ell_2, \ell'_3, \ell'_4)'$  be a  $(2+2k_Z)$  dimensional vector, where  $\ell_1$  and  $\ell_2$  are scalars and  $\ell_3$  and  $\ell_4$  are both  $k_Z$  dimensional vectors. To prove Lemma 4.1, we introduce the three lemmas below.

**Lemma F.1.** *Suppose Assumptions 1-5 hold and  $\omega_{m,j}^* > 0$ , then  $\ell' \bar{\Psi}_G(m, j) \ell = 0$  if and only if  $\ell_1 = \ell_2 = 0$ ,  $\ell_3 \bar{\mathcal{Z}}_{m,j} \ell_3 = 0$  and  $\ell_4 \bar{\mathcal{Z}}_{m,j} \ell_4 = 0$ .*

**Lemma F.2.** *Suppose Assumptions 1-5 hold and assume further that  $\mu_{\epsilon_0,j}^{(4)} - \sigma_{\epsilon_0,j}^4 > (\mu_{\epsilon_0,j}^{(3)})^2 / \sigma_{\epsilon_0,j}^2$  and  $\omega_{m,j}^* > 0$ , then  $\ell' \bar{\Psi}(m, j) \ell = 0$  if and only if  $\ell_1 = \ell_2 = 0$ ,  $\ell_3 \bar{\mathcal{Z}}_{m,j} \ell_3 = 0$  and  $\ell_4 \bar{\mathcal{Z}}_{m,j} \ell_4 = 0$ .*

Let  $\{(m, j) | \omega_{m,j}^* > 0\}$  be the set of all pairs of  $(m, j)$  such that  $\omega_{m,j}^* > 0$ , and index its elements with  $p = 1, \dots, \bar{P}$ . We therefore have  $\omega_{m_p, j_p}^* > 0$  for  $p = 1, \dots, \bar{P}$ . Note that for all  $j = 1, \dots, J$ , there exists some  $p$  such that  $j_p = j$ . This is because for each  $j$  there exists some  $m$  such that  $\omega_{m,j}^* > 0$ , observing that  $\omega_j^* = \sum_{m=2}^{\bar{M}} \omega_{m,j}^* > 0$  all  $j$ , and  $m \leq \bar{M}$  is bounded. The set of all  $A_{m,j}$  defined in (F.6) with  $\omega_{m,j}^* > 0$  is  $\{A_{m,j} | \omega_{m,j}^* > 0\} = \{A_{m_1, j_1}, \dots, A_{m_{\bar{P}}, j_{\bar{P}}}\}$ . The Lemma below states that the column by column concatenation of all matrices in this set has full row rank.

**Lemma F.3.** *Suppose Assumptions 1-6 hold, then the matrix  $\Phi = [A_{m_1, j_1}, \dots, A_{m_{\bar{P}}, j_{\bar{P}}}]$  has full row rank.*

Lemma F.1 follows easily from (F.4), observing that  $\omega_{m,j}^* > 0$ ,  $\sigma_{\epsilon_0,j}^2 > 0$ . The proofs of Lemma F.2 and Lemma F.3 are given in the Online Appendix.

We can now utilize the above lemmas to prove that under the maintained assumptions  $\Gamma_0$  is positive definite, and that the matrix  $\Upsilon_0$  is positive definite for  $\mu_{\epsilon_0,j}^{(4)} - \sigma_{\epsilon_0,j}^4 > (\mu_{\epsilon_0,j}^{(3)})^2 / \sigma_{\epsilon_0,j}^2$ . We present a proof for the positive definiteness of  $\Upsilon_0$ . The proof for the positive definiteness of  $\Gamma_0$  is analogous.



Let  $\alpha = (\alpha'_1, \alpha'_2)$  be a  $(2 + J + k_Z)$  vector, where  $\alpha_1$  is a  $(2 + J) \times 1$  vector and  $\alpha_2$  is a  $k_Z \times 1$  vector. To show that  $\Upsilon_0$  is positive definite is equivalent to showing that  $\alpha' \Upsilon \alpha = 0$  if and only if  $\alpha_1 = 0$  and  $\alpha_2 = 0$ . Observe that

$$\begin{aligned} \alpha' \Upsilon_0 \alpha &= \sum_{j=1}^J \sum_{m=2, \omega_{m,j}^* > 0}^{\bar{M}} \alpha' \varphi(m, j) \bar{\Psi}(m, j) \varphi(m, j)' \alpha \\ &= \sum_{j=1}^J \sum_{m=2, \omega_{m,j}^* > 0}^{\bar{M}} \ell'_{m,j} \bar{\Psi}(m, j) \ell_{m,j}, \end{aligned}$$

where

$$\ell_{m,j} = \varphi(m, j)' \alpha = \begin{pmatrix} A'_{m,j} \alpha_1 \\ B'_{m,j} \alpha_1 + C'_{m,j} \alpha_2 \end{pmatrix},$$

with  $A_{m,j}$  and  $C_{m,j}$  defined in Equations (F.6) and (F.7). It thus suffices to show that  $\ell'_{m,j} \bar{\Psi}(m, j) \ell_{m,j} = 0$  for all  $m, j$  with  $\omega_{m,j}^* > 0$  if and only if  $\alpha_1 = 0$  and  $\alpha_2 = 0$ . Given Lemma F.2, if  $\omega_{m,j}^* > 0$  and  $\ell'_{m,j} \bar{\Psi}(m, j) \ell_{m,j} = 0$  then  $A'_{m,j} \alpha_1 = 0$ . Lemma F.3 indicates that for  $A'_{m,j} \alpha_1 = 0$  to hold for all  $m$  and  $j$ , we must have  $\alpha_1 = 0$ . With  $\alpha_1 = 0$ ,

$$\ell_{m,j} = \begin{pmatrix} 0 \\ C'_{m,j} \alpha_2 \end{pmatrix} = - \begin{pmatrix} 0 \\ \frac{1}{\sigma_{\epsilon 0, j}^2} \alpha_2 \\ \frac{m}{\sigma_{\epsilon 0, j}^2 + m \sigma_{\alpha 0}^2} \alpha_2 \end{pmatrix}.$$

Utilizing Lemma F.2 again and noting that then  $\frac{1}{\sigma_{\epsilon 0, j}^2} > 0$  and  $\frac{m}{\sigma_{\epsilon 0, j}^2 + m \sigma_{\alpha 0}^2} > 0$ , we have  $\alpha'_2 \check{z}_{m,j} \alpha_2 = 0$  and  $\alpha'_2 \bar{z}_{m,j} \alpha_2 = 0$  for all  $m$  and  $j$ . Consequently,

$$\alpha'_2 \sum_{j=1}^J \sum_{m=2}^{\bar{M}} (\check{z}_{m,j} + \bar{z}_{m,j}) \alpha_2 = 0.$$

This gives  $\alpha_2 = 0$  as  $\sum_{j=1}^J \sum_{m=2}^{\bar{M}} (\check{z}_{m,j} + \bar{z}_{m,j})$  is positive definite under Assumption 5. In all,  $\alpha' \Upsilon_0 \alpha = 0$  if and only if  $\alpha = (\alpha'_1, \alpha'_2) = 0$  hence  $\Upsilon_0$  is positive definite. The proof of the positive definiteness of  $\Gamma_0$  follows similarly.