

Identification and Estimation of Peer Effects with Panel Data

Valentin Verdier*

October 30, 2015

Preliminary and Incomplete, please do not distribute

Abstract

I investigate how variations in peer groups across time can be used to identify and estimate peer effects in the presence of sorting and other confounding factors. I propose a simple estimator for the linear-in-means model of peer effects that can be extended to more general models of social interactions and to cases where there is feedback from past outcomes to current covariates or peer group assignment. In order to establish the asymptotic properties of the proposed estimators, I also develop an asymptotic framework for cross-sectional dependence characterized by time-varying cluster membership and for estimators that use non-nested two-way fixed effects transformations. As an empirical application, I evaluate the importance of accounting for peer-effects in models of student learning using administrative student achievement data from North Carolina.

Keywords: Peer effects, Panel data, Value-added models of student achievement

JEL codes: C23

*Assistant Professor, Department of Economics, University of North Carolina, Chapel Hill, NC 27599, United States. Tel.: +1 919-966-3962. E-mail address: vverdier@email.unc.edu.

1 Introduction

Peer effects are hard to identify with observational data because similarities in the behavior of peers could be due to peer effects or to selection into peer groups based on common unobserved characteristics (homophily), or to factors that affect all peers in a peer group (confounding factors). This issue was dubbed the reflexion problem by Manski (1993).

In this paper I show that variations in peer groups over time can be used to identify peer effects. I start with a model which has the model in Manski (1993) as a special case and show that when this model is observed over time and peer group assignment varies over time, peer effects are identified even in the presence of homophily and confounding factors, as long as some restrictions are imposed on these phenomena. I define a simple estimator for peer effects from this model which is consistent and asymptotically normal under an assumption of mean independence of the transitory shocks and efficient under an additional assumption of homoscedasticity and serial uncorrelation, hence this estimator can be dubbed parsimonious and locally efficient. The framework I use is relatively flexible and can accommodate the presence of feedback from past outcomes to current covariates or current peer group assignment. As an empirical application I evaluate the importance of accounting for peer effects in models of student learning using administrative data on student achievement.

Peer effects have generated a large amount of work both in the empirical literature and in the econometric literature. Because of the issues of identification pointed out in Manski (1993), experimental or quasi-experimental designs have often been used in empirical work. Among many other papers, Sacerdote (2001), Zimmerman (2003), Graham (2008), Duflo et al. (2011), Aral and Walker (2011) use data where peer groups are randomly assigned. Imberman et al. (2012) use hurricane Katrina as an exogenous source of variation in peer groups, Dahl et al. (2014) exploit a regression discontinuity design in outcome to identify peer effects. In one of the few papers using natural experiments and panel data that I am aware of, Aral et al. (2015) use variations in weather for peers who live in different geographical areas as sources of exogenous variation in outcome to identify peer effects.

With observational cross-sectional data, Lee (2007), Bramoullé et al. (2009), and Davezies

et al. (2009) showed that peer effects are identified in the presence of specific forms of homophily and confounding factors when a slight modification to the model of Manski (1993) is used (namely when leave-one-out averages are used instead of averages) and when the size of peer groups varies. I show that, with panel data, peer effects are identified with more general forms of homophily and confounding factors, independently of whether a model with leave-one-out averages or averages is used, which hopefully addresses some of the concerns of weak instrumental variables and identification by functional form voiced in Angrist (2014) for instance. Aral et al. (2009) used an estimator that accounts for general forms of homophily as long as it is based on observed covariates only, in the absence of confounding factors. In our model, no restriction is imposed on the relationship between peer group assignment and unobserved (and observed) characteristics. This is a desirable feature to study student learning since previous evidence has shown that unobserved characteristics account for a large part of the heterogeneity in student learning, which makes it likely that unobserved characteristics also play an important role in peer group assignment.

With panel data, Manresa (2015) studies the identification and estimation of peer effects using data on many observations over many time periods when peer relationships are not assumed to be observed, but are assumed to be stable over time and relatively few. Here I consider the case where observations span only a few time periods and the composition of peer groups changes over time, but the existence of a peer relationship is assumed to be observed. Hence the analysis presented in this paper can be seen as complementary to the work in Manresa (2015). For the empirical study of peer effects in student learning, a framework with few time periods is necessary since students are typically observed only for a few years while in school.

Arcidiacono et al. (2012) also consider a model of peer effects with panel data and few time periods. They consider a model without confounding factors and without observed covariates and propose a non-linear least square estimator that estimates simultaneously peer effects and individual unobserved heterogeneity. They provide conditions under which their estimator does not run into an incidental parameters issue. However, as is known from earlier

panel data literature (see Wooldridge (2010) for instance), so-called fixed effects estimators, which rely on the estimation of parameters with dimension proportional to sample size, are consistent for parameters of interest only in very specific special cases. In this paper I take a different approach which consists in transforming peer effects models so that a method of moments approach can be used. As a result, the estimator proposed in this paper can not only accommodate observed covariates and confounding factors, but is also easily extended to models where there is feedback from past outcomes to current covariates or peer group assignment. In addition, conditions for the efficiency of the simple estimators proposed in this paper are easily characterized. Feedback from past outcomes to current covariates is often very likely to be present when using observational panel data. In the case of models of student learning for instance, past educational inputs are likely to have an effect on current and future learning. Empirically, Andrabi et al. (2011) found that allowing for feedback in student learning yielded very different results from results obtained after imposing the ex-ante restriction that no feedback exists or that feedback does not decay with time. Dynamic peer group assignment is also likely to be present jointly with peer effects when using observational panel data. Indeed if peer effects have a significant role in determining outcomes, past outcomes might affect decisions on what peers an observation should be exposed to. Goldsmith-Pinkham and Imbens (2013) considered models where the dynamics in peer group assignment is assumed to be correctly specified. We will show that an additional advantage of our approach is that it allows us to be agnostic about the form of the dynamics in peer groups assignment when confounding factors are not included in the model, although some restrictions will have to be imposed when confounding factors are included in the model.

The estimators I propose take the general form of instrumental variable estimators where a transformation of the data is used to remove non-nested two-way fixed effects and where cross-sectional dependence is present. The cross-sectional dependence takes the form of cluster dependence in a given time period, but since cluster membership changes over time, asymptotic results for independent non-identically distributed cannot be applied. I show that asymptotic results for spatial data can be used for this case and present sufficient condi-

tions for the proposed estimators to be \sqrt{n} -asymptotically normal, where n is the number of cross-sectional observations. Regarding non-nested two-way fixed effects estimation, unlike in Abowd et al. (1999) I consider the case where an incidental parameter problem arises from estimating either of the fixed effects, not only from estimating one of the fixed effects (namely worker fixed effects in Abowd et al. (1999), as opposed to firm fixed effects which are assumed to be finite-dimensional in Abowd et al. (1999)). I show that conditions are needed on the assignment process to categories for the quantities entering commonly used estimators to satisfy laws of large numbers and central limit theorem. I also generalize the estimator to the case where dynamics are present in the covariates, which would invalidate existing estimators.

Section 2 presents the simplest form of the model and compares it to other models that have been used in the econometric literature. It also contains the basic result of lack of identification of peer effects in the absence of variations in peer groups over time. Section 3 shows how a simple estimator for peer effects is obtained, and provides conditions under which this estimator is consistent and asymptotically normal. It also extends the results to the presence of feedback from past outcomes to current covariates and peer group assignment. Finally it shows that the approach presented in this paper can be extended to any model in the so-called class of spatial models of social interactions. Section 4 applies the methods outlined in the previous section to the empirical study of peer effects in models of student learning.

2 The model

First consider a model for $i = 1, \dots, n$, $t = 1, \dots, T$ that can be written as:

$$y_{it} = c_i + x_{it}\beta_0 + (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} (c_j + x_{jt}\beta_0) + u_{it} \quad (2.1)$$

$$E(u_{it}|x, p) = 0 \quad (2.2)$$

where i is a cross-sectional observation, t a time period, c_i is unobserved heterogeneity, x_{it} are observed covariates (but not necessarily finite dimensional, see next section), u_{it} are unobserved shocks, p_{it} is the peer group of observation i in period t , $x = \{x_{it}\}_{i=1,2,\dots,t=1,\dots,T}$

and $p = \{p_{it}\}_{i=1,2,\dots,t=1,\dots,T}$. We also impose $\gamma_0 \neq 0$ as is standard in the rest of the literature on peer effects.

γ_0 is the social multiplier, i.e. the effect of a policy raising $x\beta_0$ for all observations by one unit on average outcomes. In the absence of peer-effects, the effect of such a policy on average outcomes would be one, with positive peer-effects it would be greater than one and with negative peer-effects it would be less than one.

Homophily is captured by the dependence between group membership p_{it} and unobserved heterogeneity c_i . Unobserved shocks that vary over time, u_{it} , are assumed to be mean independent of the past, current, and future values of the covariates and of peer group membership for now, but this is relaxed in later sections.

x contains covariates that might be specific to an observation i and affect observation i 's peers through peer effects and also contains confounding factors that might be common to all observations in a specific peer group, although we will see later that we will assume that these confounding factors are not an exact function of group membership, i.e. they either vary within groups or they are common to different groups over time. In our empirical application for instance, cofounding factors that are common to all students in a classroom will be interpreted as being the effect of teacher quality on student achievement and some restriction will be imposed on how teacher quality changes over time (teacher quality not changing over time being the simplest of such restrictions).

Comparison with the model of Manski 1993 In the absence of time variation, Manski (1993) showed that the model given by (2.1) and (2.2) is not identified.

Indeed, for a particular time period, the model given by (2.1) and (2.2) can be recast as in Manski (1993). (2.1) and (2.2) can be rewritten as:

$$E(y_{it}|x, p) = E(c_i|x, p) + x_{it}\beta_0 + (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} (E(c_j|x, p) + x_{jt}\beta_0) \quad (2.3)$$

If $E(c_i|x, p) = E(c_i|p_{it})$, which would be true for instance if x were independent of both

group membership and unobserved heterogeneity, then the the above equation becomes:

$$E(y_{it}|x, p) = \gamma_0 E(c_i|p_{it}) + x_{it}\beta_0 + (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} x_{jt}\beta_0 \quad (2.4)$$

which for any particular time period t corresponds to equation (5) in Manski (1993).

The purpose of this paper is to show that time variations in peer groups can be succesfully used to identify this model.

Comparison with the model of Graham (2008) Graham (2008) also considers the case where only cross-sectional data is available. In Graham (2008), homophily is assumed away, so that $E(c_i|p_{it}) = E(c_i)$. Systematic differences across peer groups other than those due to peer effects are still present in the form of group level confounding factors. The running example in Graham (2008) is that classroom level heterogeneity is due to differences in teacher quality, and the assignment of teachers to classrooms is assumed to be at random.

In our model we could have $x_{it} = \{1[d_{it} = d]\}_{d=1, \dots, D}$ where d_{it} is student i 's teacher in time period t , and D is the number of teachers observed in the data. Hence our equation (2.1) would be written:

$$y_{it} = \gamma_0 e_{d_{it}} + c_i + (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} c_j + u_{it} \quad (2.5)$$

where $e_{d_{it}}$ is the component of β_0 corresponding to d such that $1[d_{it} = d] = 1$.

No restriction other than mean independence is imposed on u_{it} in (2.1) and (2.2), so that a special case for u_{it} is:

$$u_{it} = v_{it} + (\gamma_0 - 1) \frac{\sum_{j \in P_{it}} v_{jt}}{n_{p_{it}}}$$

$$E(v_{it}|x, P) = 0$$

in which case, redefining $\epsilon_{it} = c_i + v_{it}$, we have:

$$y_{it} = e_{d_{it}} + \epsilon_{it} + (\gamma_0 - 1) \frac{\sum_{j \in p_{it}} \epsilon_{jt}}{n_{p_{it}}} \quad (2.6)$$

which, for any particular time period t , corresponds to equation (1) in Graham (2008).

Unlike in Graham (2008) however, the dependence between $e_{d_{it}}$, p_{it} , and c_i is not restricted, so that homophily (selection into peer groups) and correlated confounding factors (matching of teachers to classrooms) are present in the model. One important caveat with our approach is that some restriction has to be imposed on how teacher quality changes over time. Here and throughout the paper we use the assumption that teacher quality does not change over time, but this could be relaxed to a more flexible restriction.

Linear-in-means model with averages or leave-one-out averages Instead of the model given by (2.1) and (2.2), one could consider the model:

$$y_{it} = c_i + x_{it}\beta_0 + \eta_0 \frac{1}{n_{p_{it}} - 1} \sum_{j \in p_{it}, i \neq j} (c_j + x_{jt}\beta_0) + \lambda_0 \frac{1}{n_{p_{it}} - 1} \sum_{j \in p_{it}, i \neq j} E(y_{it}|x, p) + u_{it} \quad (2.7)$$

$$E(u_{it}|x, p) = 0 \quad (2.8)$$

which is considered in Arcidiacono et al. (2012) for $\lambda_0 = 0$ and $x_{it} = \emptyset$.

While for simplicity I concentrate on the specification with averages given by (2.1) and (2.2) in most of the paper, Section 3.3 shows that the results can easily be extended to the leave-one-out average specification to estimate β_0 , η_0 and λ_0 consistently, and efficiently under some additional conditions. In fact any model in the class of so-called spatial models of social interactions can be accomodated by the framework presented here.

Therefore, provided that peer groups vary over time, the approach presented in this paper addresses concerns of identification by functional form or of weak instrumental variables that were present when working with cross-sectional data, as expressed in Angrist (2014) for instance.

Non-identification in the absence of variation in peer groups

Proposition 1. For $\beta_0 = 0$, γ_0 is not identified in the model given by (2.1) and (2.2) if peer groups do not change over time or if $E(c_i|x, p)$ does not change across individuals.

Proof. If peer groups do not change over time, with $\beta_0 = 0$, one can redefine $c_i = c_i + (\gamma_0 - 1)\frac{1}{n_{p_{it}}}\sum_{j \in p_{it}} c_j$ and $\gamma_0 = 1$ so that a model with no peer effects is observationally equivalent to a model with non-zero peer effects. If $E(c_i|x, p)$ does not change across individuals, one can redefine c_i in the same fashion. \square

3 Identification and Estimation

Note that, for $\gamma_0 \neq 0$, from (2.1) we have:

$$\frac{\gamma_0 - 1}{\gamma_0} \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} y_{it} = (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} (c_j + x_{jt}\beta_0) + \frac{\gamma_0 - 1}{\gamma_0} \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} u_{jt} \quad (3.1)$$

Let $\bar{y}_{it} = \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} y_{it}$ and $\bar{u}_{it} = \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} u_{it}$, then:

$$y_{it} - \frac{\gamma_0 - 1}{\gamma_0} \bar{y}_{it} = c_i + x_{it}\beta_0 + u_{it} - \frac{\gamma_0 - 1}{\gamma_0} \bar{u}_{it} \quad (3.2)$$

$$E(u_{it} - \frac{\gamma_0 - 1}{\gamma_0} \bar{u}_{it} | x, p) = 0 \quad (3.3)$$

Note that, for $\gamma_0 \neq 0$, there is a one-to-one mapping from $\frac{\gamma_0 - 1}{\gamma_0}$ to γ_0 , so that our objective will be to estimate $\eta_0 \equiv \frac{\gamma_0 - 1}{\gamma_0}$ instead of γ_0 .

As mentioned above, x_{it} is not restricted to be finite dimensional. We write $x_{it} = [z_{it}, w_{it}]$ and $\beta'_0 = [e', \theta'_0]$ where θ_0 is finite-dimensional but the dimension of $e = [e_1, \dots, e_D]'$ can depend on sample size n . For simplicity, since our empirical objective here is to estimate a model of student achievement, we will restrict our attention to the case where z_{it} is a vector of mutually exclusive indicator functions, which in our application will be indicator functions for which teacher teaches student i in time period t among all teachers observed in the data, $z_{it} = [1[d_{it} = d]]^{d=1, \dots, D}$. However the following reasoning could be extended to the case where z_{it} is a more general vector of explanatory variables of growing dimension if needed.

We can rewrite:

$$y_{it} - \eta_0 \bar{y}_{it} = c_i + e_{d_{it}} + w_{it} \theta_0 + u_{it} - \eta_0 \bar{u}_{it} \quad (3.4)$$

$$E(u_{it} - \eta_0 \bar{u}_{it} | x, p) = 0 \quad (3.5)$$

Because of homophily, it is likely that peer group membership p_{it} is not independent of an observation's unobserved heterogeneity c_i . In addition, it is likely that w_{it} is directly related to $e_{d_{it}}$ or that c_i is related to $e_{d_{it}}$, which then would likely imply that p_{it} is not independent of $e_{d_{it}}$. Hence we cannot treat c_i and $e_{d_{it}}$ as unobserved shocks that would be mean independent of w and p .

On the other hand, we also cannot estimate e and $c = [c_1, \dots, c_n]'$ since, as their dimension grows proportionally with sample size, estimators for e and c would be inconsistent, resulting in an issue of incidental parameters.

Hence our objective is to estimate η_0 and θ_0 consistently in the presence of arbitrary dependence between x_{it}, p_{it} and $c_i, e_{d_{it}}$.¹

Let $g_{it} = [[1[d_{it} = d]]^{d=1, \dots, D}, [1[i = s]]^{s=1, \dots, n}]$, let $g = [g_{it}]_{i=1, \dots, n, t=1, \dots, T}$, $y = [y_{it}]_{i=1, \dots, n, t=1, \dots, T}$, $\bar{y} = [\bar{y}_{it}]_{i=1, \dots, n, t=1, \dots, T}$, $w = [w_{it}]_{i=1, \dots, n, t=1, \dots, T}$, $\ddot{u} = [u_{it} - \eta_0 \bar{u}_{it}]_{i=1, \dots, n, t=1, \dots, T}$.

Hence with this new notation we have:

$$y - \eta_0 \bar{y} = g[e', c']' + w\theta_0 + \ddot{u} \quad (3.6)$$

$$E(\ddot{u} | x, p) = 0 \quad (3.7)$$

Define $M_g = I - g(g'g)^-g'$, where A^- is a generalized inverse of A .²

¹In a second step we might also be interested in recovering information on the distribution of c_i and $e_{d_{it}}$, see the empirical application below.

²Lemma 1 in the appendix shows that the estimator defined here is algebraically identical no matter which generalized inverse of $g'g$ is used.

Then:

$$M_g(y - \eta_0 \bar{y} - w\theta_0) = M_g \ddot{u} \quad (3.8)$$

$$E(M_g \ddot{u} | x, p) = 0 \quad (3.9)$$

The unfeasible estimator for η_0 and θ_0 proposed in this paper is:

$$\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} = (z' M_g [\bar{y}, w])^{-1} z' M_g y$$

$$z = [z_{it}]_{i=1, \dots, n, t=1, \dots, T}$$

$$z_{it} = [E(\bar{y}_{it} | x, p), w_{it}]$$

From (3.1) and (2.2):

$$\begin{aligned} E(\bar{y}_{it} | x, p) &= \gamma_0(\bar{w}_{it}\theta_0 + \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} (E(e_{d_{jt}} | x, p) + E(c_j | x, p))) \\ &= \gamma_0(\bar{w}_{it}\theta_0 + E(e_{d_{it}} | x, p) + \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} E(c_j | x, p)) \end{aligned}$$

But $[1[d_{it} = d]]_{d=1, \dots, D} M_g = 0$. Hence redefine $z_{it} = [\bar{w}_{it}\theta_0 + E(\bar{c}_{it} | x, p), w_{it}]$, we still have:³

$$\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} = (z' M_g [\bar{y}, w])^{-1} z' M_g y$$

³Papers such as Lee and Yu (2010) considered spatial panel data models with fixed effects that are in some ways similar to the peer effects model considered in this paper, but where the location of observations (which would in some ways correspond to an observation's peer group in our model) is assumed to be fixed over time, so that identification is obtained from time-varying covariates that are assumed to have non-zero coefficients. It is interesting to note that, when the location of an observation is assumed to be fixed over time, $E(\bar{c}_{it} | x, p)$ in these models would be constant over time, i.e. a linear function of $[1[i = s]]_{s=1, \dots, n}$. Since $[1[i = s]]_{s=1, \dots, n} M_g = 0$, that quantity would also vanish from z_{it} . Here however, $E(\bar{c}_{it} | x, p)$ will be necessary to obtain an efficient estimator, as shown in Proposition 4.

in addition, from (3.6) we have:

$$\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} + (z' M_g [\bar{y}, w])^{-1} z' M_g \ddot{u} \quad (3.10)$$

Next we list assumptions that will guarantee that $\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix}$ is a consistent and \sqrt{n} -asymptotically normal estimator for $\begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix}$.

Let $\mathcal{S}_{it} = \{(j, s) : j = 1, \dots, n, s = 1, \dots, T, j = i \vee (t = s \wedge p_{it} = p_{jt})\}$ be the set of all observations that are either observations on i at different time periods or classmates of i in time period t .

Assumption 1. $\forall i, j = 1, \dots, n, \forall t, s = 1, \dots, T, 1 \leq n_{p_{it}} \leq M < \infty$, and, if $(j, s) \notin \mathcal{S}_{it}$, then $D(u_{it}, u_{js}|x, p) = D(u_{it}|x, p)D(u_{js}|x, p)$.

This first assumption implies that, conditional on the observed covariates, the transitory shocks u_{it} are independent across groups in a given time period or across time for different observations. Such dependence structure would exist for instance if u_{it} could be decomposed between a classroom component and a student specific component: $u_{it} = e_{p_{it}, t}^p + e_{it}^s$ and $\{e_{p,t}^p\}_{t,p} \perp \{e_{it}^s\}_{i,t}|x, p, e_{p,t}^p \perp e_{p,t'}^p|x, p$ for $t \neq t'$, $e_{it}^s \perp e_{i't'}^s$ for $i \neq i'$.

For any $d = 1, \dots, D$ and $d' = 1, \dots, D$, define $n_{dd'} = \sum_{i=1, \dots, n} \sum_{s, t=1, \dots, n} 1[d_{it} = d_{j-1}]1[d_{is} = d_j]$ to be the number of students that have had both d and d' as teachers. Then define $b_{d,d'} = \min\{k : \exists \{d_1, \dots, d_{k-1}\} \text{ s.t. } n_{d_{j-1}d_j} > 0 \forall j = 1, \dots, k, d_0 = d, d_k = d'\}$ where as a convention we set $\min\{\emptyset\} = 0$, to be the number of steps on the shortest path from d to d' , where two teachers are connected if they have had at least one student in common, and where we set $b_{dd'} = 0$ if there is no path linking d to d' . Finally define $b_n = \max_{d,d'=1, \dots, D} \{b_{d,d'}\}$ to be the maximum number of steps separating any two teachers in the data that are connected. If we see all of the teachers in the data as vertices, and whether two teachers have taught at least one student in common as edges, b_n is called the diameter of the resulting graph.

Assumption 2. For any random sequence $\{a_n\}_{n=1,2,\dots}$ where a_n is $(n + D) \times 1$:

a) $\lambda_{\min}(\frac{1}{n} \sum_{i,t} E((z_{it} - g_{it}a_n)'(z_{it} - g_{it}a_n))) > c > 0 \forall n > N$

b) $\lambda_{\min}(\frac{1}{n} Var(\sum_{i,t} (z_{it} - g_{it}a_n)' \ddot{u}_{it})) > c > 0 \forall n > N$

For some random variable $\mathbf{B} \in (0, +\infty)$ with $E(\mathbf{B}) < \infty$:

c) $\forall n, \forall i, j = 1, \dots, n, E(|c_i|^{4+\delta}) \leq C < \infty, E(|u_{it}|^{4+\delta} | x, p) \leq \mathbf{B}, |z_{it}[k]|^{4+\delta} \leq \mathbf{B}$ a.s.

d) $E(\mathbf{B}b^{4+\delta}(MT^2)^{(4+\delta)b}) < C < \infty$

Assumption 2a) implies that, after student and teacher level variations are removed, there are still linearly independent variations in the variables in z_{it} . For simplicity, consider the case where there is no time-varying covariates, $w_{it} = \emptyset$ (but x can include any time constant variable that can help predict c_i). Then $z_{it} = E(\bar{c}_{it} | x, p) = \frac{1}{n_{pit}} \sum_{j \in pit} E(c_j | x, p)$ and Assumption 2a) implies that the variations in $\frac{1}{n_{pit}} \sum_{j \in pit} E(c_j | x, p)$ cannot be perfectly predicted by teacher and student level variations. For Assumption 2a) to be satisfied, $\frac{1}{n_{pit}} \sum_{j \in pit} E(c_j | x, p)$ should change across time for many students and teachers. If peer groups do not change over time for instance, then $\frac{1}{n_{pit}} \sum_{j \in pit} E(c_j | x, p)$ does not change over time for students, even though it might change over time for teachers. If $E(c_i | x, p)$ does not change across individuals, then $\frac{1}{n_{pit}} \sum_{j \in pit} E(c_j | x, p)$ changes over time for neither students nor teachers. If teachers are always assigned students with the same average ability level, then $\frac{1}{n_{pit}} \sum_{j \in pit} E(c_j | x, p)$ does not change over time for teachers, even though it might change over time for students. These three situations would be cases where Assumption 2a) is violated when there is no time-varying covariates or $\theta_0 = 0$.

Assumption 2b) is a standard assumption of asymptotic ignorability of any particular summand $(z_{it} - g_{it}a_n)' \ddot{u}_{it}$. Assumption 2c) is also standard and bounds the higher moments of u_{it} and z_{it} . Assumption 2d) bounds the diameter of the graph formed by teachers as described above in expected value. This assumption is needed to guarantee the existence of higher moments of transformed variables from removing both student and teacher effects. With a model that contains only student effects c_i and finite dimensional covariates, Assumption 2d) would not be needed.

Assumption 3. All observations can be split into R groups and $D(\{\{x_{rj}, p_{rj}\}_{j=1,\dots,N_r}\}_{r=1,\dots,R}) =$

$\prod_{r=1}^R D(\{x_{rj}, p_{rj}\}_{j=1, \dots, N_r})$. In addition, $\frac{R \max\{n_r\}_{r=1, \dots, R}}{n} = O(1)$ and $R \rightarrow \infty$ as $n \rightarrow \infty$.

This last assumption imposes a relatively mild restriction on the cross-sectional dependence of the covariates ... It is composed of three parts: independence across groups, the number of groups grows unboundedly, the size of each group in general grows at the same rate. This assumption is relatively mild because no restriction is imposed on the dependence of observations inside each group and on the rate at which R grows compared to n . R could increase very slowly compared to n , so that groups with unrestricted within-dependence could be very large, but the asymptotic rate of convergence of the estimator we propose will still be \sqrt{n} . This is because Assumption 1 and Assumption 3 can be combined to show that our estimator has \sqrt{n} rate of convergence. Assumption 3 fits our empirical application well since our data can naturally be split into 115 school districts, with a maximum enrollment of around 47,000 students and an average enrollment of around 4,000 students.

Under these three assumptions, the estimator defined above is consistent and asymptotically normal.

Proposition 2. *Under (2.1) and (2.2) and Assumptions 1-3, we have:*

$$\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} \quad (3.11)$$

and:

$$\begin{aligned} \sqrt{n} V_n^{-\frac{1}{2}} A_n \left(\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} \right) &\xrightarrow{d} N(0, I) \\ A_n &= \frac{1}{n} E(z' M_g z) \\ V_n &= \frac{1}{n} E(z' M_g \ddot{u} \ddot{u}' M_g z) \end{aligned}$$

The next proposition is concerned with the conditions under which $\hat{\eta}, \hat{\theta}$ is efficient for estimating η_0, θ_0 . We can first note that, with regard to the estimation of θ_0, η_0 , the model given by (3.6) and (3.7) is equivalent to the model given by (3.8) and (3.9) since there is no

restriction imposed on the conditional distribution of c, e . Indeed (3.6) and (3.7) are implied by:

$$E(P_g(y - w\theta_0 - \eta_0\bar{y}) - g \begin{bmatrix} c \\ e \end{bmatrix} | x, p) = 0 \quad (3.12)$$

$$E(M_g(y - w\theta_0 - \eta_0\bar{y}) | x, p) = 0 \quad (3.13)$$

and the first equation can be written:

$$gE\left(\begin{bmatrix} c \\ e \end{bmatrix} | x, p\right) = g(g'g)^{-1}g'E(y - w\theta_0 - \eta_0\bar{y} | x, p) \quad (3.14)$$

For any θ, η ,

$$g\xi = g(g'g)^{-1}g'E(y - w\theta - \eta\bar{y} | x, p) \quad (3.15)$$

holds by setting $\xi = (g'g)^{-1}g'E(y - w\theta - \eta\bar{y} | x, p)$, and $E\left(\begin{bmatrix} c \\ e \end{bmatrix} | x, p\right)$ is an $n + D \times 1$ vector of unrestricted unknown functions. Hence (3.12) does not provide any restrictions on η_0, θ_0 that could be used for estimation, and we can restrict our attention to estimating θ_0, η_0 from (3.13).

Assumption 4 considers a specific covariance structure for u_{it} obtained from homoscedastic, serially uncorrelated shocks with the same peer effects structure as the one that applies to the observed covariates and the unobserved heterogeneity.

Assumption 4. $u_{it} = v_{it} + (\gamma_0 - 1)\frac{1}{n_{pit}} \sum_{j \in p_{it}} v_{jt}$, $Cov(v_{it}, v_{js} | x, p) = \sigma_v^2 1[i = j, t = s]$.

Then under Assumption 4, the next proposition shows that the instruments z are the best choice of instruments in terms of efficiency

Proposition 3. Consider any $(K + 1) \times nT$ function Z of x, p , define $B_n = E(\frac{1}{n}Z' M_g[\bar{y}, w])$ and $W_n = (\frac{1}{n}Var(Z' M_g \ddot{u}))$.

If $\lambda_{\min} B_n > c > 0$ for all $n > N$ and Assumption 4 holds then:

$$D = B_n^{-1} W_n B_n^{-1} - A_n^{-1} V_n A_n^{-1} \quad (3.16)$$

is positive semi-definite.

Since $B_n^{-1} W_n B_n^{-1}$ would be the asymptotic variance of an estimator defined by $(Z' M_g[\bar{y}, w])^{-1} Z' M_g y$ provided that regularity conditions apply so that this estimator is \sqrt{n} -asymptotically normal, the result of this proposition implies that z is the vector of optimal instruments for estimating η_0, θ_0 from (2.1) and (2.2) under Assumption 4. Note that considering exactly identifying instruments Z is without loss of generality since it accounts for the case where over-identifying instruments are used with optimal weighting already applied.

The last proposition shows that standard errors are consistent under Assumptions 1-3.

Proposition 4. Define $\hat{V}_n = \frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j,s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} \hat{u}_{it} \hat{u}_{js}$, $\hat{u}_{it} = y_{it} - \hat{\eta} \bar{y}_{it} - w_{it} \hat{\theta}$.

Under (2.1) and (2.2) and Assumptions 1-3 we have:

$$\sqrt{n} \hat{V}_n^{-\frac{1}{2}} \left(\frac{1}{n} z' M_g[\bar{y}, w] \right) \left(\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} \right) \xrightarrow{d} N(0, I) \quad (3.17)$$

The estimator $\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix}$ is unfeasible since $E(c|x, p)$ is unknown. In order to implement the estimator, one should replace $E(c|x, p)$ with an estimable predictor of c based on x, p . As in Verdier (2015), one can use an auxiliary model that makes use of the cross-sectional dependence likely induced by sorting on unobservables to obtain a stronger predictor of c based on x, p . A simple auxiliary model for $E(c_i|x, p)$ could be chosen to be $\mathcal{E}_T(c_i|x, p, \delta)$ from:

$$\mathcal{E}_1(c_i|x, p, \delta) = \delta_{10} + w_{i1} \delta_{11} + \bar{w}_{i1} \delta_{12}$$

$$\mathcal{E}_t(c_i|x, p, \delta) = \delta_{t0} + w_{it} \delta_{t1} + \bar{w}_{it} \delta_{t2} + \delta_{t3} \mathcal{E}_{t-1}(c_i|x, p, \delta) + \delta_{t4} \frac{1}{n_{pit}} \sum_{j \in pit} \mathcal{E}_{t-1}(c_j|x, p, \delta)$$

$\{\delta_{tj}\}_{t=1, \dots, T, j=0, \dots, 4}$ could be estimated by a non-linear regression of $M_{g_2}(y - \hat{\eta}_{\text{prelim}} \bar{y} -$

$w\hat{\theta}_{prelim}$) on $M_{g_2}[\mathcal{E}_T(c_i|x, p, \delta)]_{i=1,2,\dots,n,t=1,2,\dots,T}$, where $\begin{bmatrix} \hat{\eta}_{prelim} \\ \hat{\theta}_{prelim} \end{bmatrix}$ is a consistent preliminary estimator that could be obtained for instance by using $\{\hat{\delta}_{tj,prelim}\}_{t=1,\dots,T,j=0,\dots,4}$ obtained from a non-linear regression of $M_{g_2}y$ on $M_{g_2}[\mathcal{E}_T(c_i|x, p, \delta)]_{i=1,2,\dots,n,t=1,2,\dots,T}$.

As in Verdier (2015), the resulting estimator will be efficient if Assumption 4 holds and the auxiliary model $\mathcal{E}_T(c_i|x, p, \delta)$ actually coincides with $E(c|x, p)$ for some δ , but even if one or both of this conditions are violated, the resulting estimator will be consistent and asymptotically normal, and will be close to being efficient if Assumption 4 and the auxiliary assumption are good approximations for the true form of the conditional variance matrix of \ddot{u} and the true form of $E(c|x, p)$.

3.1 Extension to sequentially exogenous covariates

The above analysis can be extended to models where feedback from past outcomes to current covariates is present. In this section we consider models that can be written:

$$y_{it} = c_i + e_{d_{it}} + w_{it}\theta_0 + (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} (c_j + w_{jt}\theta_0) + u_{it} \quad (3.18)$$

$$E(u_{it}|z^t, g, p) = 0 \quad (3.19)$$

where $z^t = \{z_1, \dots, z_t\}$ and $z_t = \{z_{it}\}_{i=1,2,\dots}$. Here again we also impose $\gamma_0 \neq 0$.

A simple example would be $w_{it} = z_{it} = y_{it-1}$ for instance, which would lead to the dynamic model with peer effects:

$$y_{it} = c_i + e_{d_{it}} + \theta_0 y_{it-1} + (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} (c_j + \theta_0 y_{jt-1}) + u_{it}$$

$$E(u_{it}|y^{t-1}, g, p) = 0$$

The model can be transformed as in the previous section:

$$y_{it} - \eta_0 \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} y_{jt} = c_i + e_{d_{it}} + w_{it} \theta_0 + \ddot{u}_{it} \quad (3.20)$$

$$E(\ddot{u}_{it} | z^t, g, p) = 0 \quad (3.21)$$

However since the set of instruments is increasing with time rather than being constant over time as in the previous section, we cannot use the transformation used in the last section to obtain moment conditions that do not involve c_i or $e_{d_{it}}$.

Since $E(c_i | z^T, g, p)$ is unrestricted, the information found in (3.20) and (3.21) for estimating η_0 and θ_0 is equivalent to the information found in:

$$\begin{aligned} \dot{y}_{it} - \eta_0 \dot{y}_{it} - \dot{w}_{it} \theta_0 &= \dot{g}_{2it} e + \dot{\ddot{u}}_{it} \\ E(\dot{\ddot{u}}_{it} | z^t, g, p) &= 0 \end{aligned}$$

where $\dot{x}_t = \left(\frac{T-t+1}{T-t}\right)^{\frac{1}{2}} \left(x_t - \frac{1}{T-t+1} \sum_{s=t}^T x_s\right)$ for any variable x_t .

Stacking over cross-sectional observations, define $m_{t,0} = \dot{y}_t - \eta_0 \dot{y}_t - \dot{w}_t \theta_0$, so that for any t , (3.20) and (3.21) can be written:

$$E(m_{t,0} - \dot{g}_{2t} e | z^t, g, p) = 0 \quad (3.22)$$

The information on $E(e | z^t, g, p)$ that is obtained from (3.20) and (3.21) for $s = t+1, \dots, T-1$ is contained in:

$$E(m_0^{t+1} - \dot{g}_2^{t+1} e | z^{t+1}, g, p) = 0 \quad (3.23)$$

This is equivalent to:

$$\begin{aligned} E(e | z^{t+1}, g, p) &= (\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-1} \dot{g}_2^{t+1'} E(m_0^{t+1} | z^{t+1}, g, p) \\ &\quad + (I - (\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-1} \dot{g}_2^{t+1'} \dot{g}_2^{t+1}) \xi_t \end{aligned}$$

where ξ_t is a function of z^{t+1}, g, p that is unrestricted.

Hence the information for estimating η_0, θ_0 found in (3.22) given (3.20) and (3.21) for $s = t + 1, \dots, T - 1$ is the same information found in:

$$\begin{aligned} & E(m_{t,0} - \dot{g}_{2t}((\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-\dot{g}_2^{t+1'}} m_0^{t+1} \\ & + (I - (\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-\dot{g}_2^{t+1'}} \xi_t) | z^t, g, p) = 0 \end{aligned} \quad (3.24)$$

which in turn is equivalent to:

$$E(M_{\dot{g}_{2t}(I - (\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-\dot{g}_2^{t+1'}} \xi_t)}(m_{t,0} - \dot{g}_{2t}(\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-\dot{g}_2^{t+1'}} m_0^{t+1}) | z^t, g, p) = 0 \quad (3.25)$$

since ξ_t is unrestricted.

Let $A_t = [-M_{\dot{g}_{2t}(I - (\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-\dot{g}_2^{t+1'}} \xi_t)} \dot{g}_{2t}(\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-\dot{g}_2^{t+1'}}, M_{\dot{g}_{2t}(I - (\dot{g}_2^{t+1'} \dot{g}_2^{t+1})^{-\dot{g}_2^{t+1'}} \xi_t)}]$, $X_t = [\dot{y}_s, \dot{w}_{is}]_{s=T-1, \dots, t}$, $Y_t = [\dot{y}_s]_{s=T-1, \dots, t}$, $U_t = [\dot{u}_s]_{s=T-1, \dots, t}$, and $Z_t = A_t[E(\dot{y}_s | z^t, g, p), E(\dot{w}_s | z^t, g, p)]_{s=T-1, \dots, t}$.

Our proposed estimator is:

$$\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} = \left(\sum_{t=1}^{T-1} Z_t'(A_t A_t')^{-1} A_t X_t \right)^{-1} \sum_{t=1}^{T-1} Z_t'(A_t A_t')^{-1} A_t Y_t \quad (3.26)$$

Note that $E(\dot{y}_s | z^t, g, p) = \gamma_0(E(\dot{w}_s | z^t, g, p)\theta_0 + E(\dot{c}_s | z^t, g, p) + \dot{g}_{2s}E(e | z^t, g, p))$ but $A_t[\dot{g}_{2s}]_{s=T-1, \dots, t} = 0$. Hence we can redefine $Z_t = A_t[(E(\dot{w}_s | z^t, g, p)\theta_0 + E(\dot{c}_s | z^t, g, p)) E(\dot{w}_s | z^t, g, p)]_{s=T-1, \dots, t}$ and still have:

$$\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} = \left(\sum_{t=1}^{T-1} Z_t'(A_t A_t')^{-1} A_t X_t \right)^{-1} \sum_{t=1}^{T-1} Z_t'(A_t A_t')^{-1} A_t Y_t \quad (3.27)$$

Note that from (3.18):

$$\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} = \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} + \left(\sum_{t=1}^{T-1} Z_t'(A_t A_t')^{-1} A_t X_t \right)^{-1} \sum_{t=1}^{T-1} Z_t'(A_t A_t')^{-1} A_t U_t \quad (3.28)$$

Let $\mathcal{R}_{it} = \{z^t, u^{t-1}, \{u_{jt}\}_{j \neq p_{it}}, g, p\}$.

Assumption 5. $\forall i = 1, \dots, n, \forall t = 1, \dots, T, 1 \leq n_{p_{it}} \leq M < \infty$, and $E(u_{it} | \mathcal{R}_{it}) = 0$.

Assumption 5 is consistent with sequentially exogenous instruments z^t , and allows for arbitrary correlation for the transitory shocks u_{it} within peer groups. However Assumption 5 also imposes serial uncorrelation of the transitory shocks, which was not assumed in the previous section. This serial uncorrelation of the transitory shocks will allow one to use asymptotic results for martingale difference sequences.

THIS SECTION TO BE COMPLETED IN FUTURE VERSIONS OF THE DRAFT.

3.2 Extension to Dynamic Peer Group and Teacher Assignment

A test for the strict exogeneity of peer group and teacher assignment could be obtained by testing the significance of characteristics of future peers and teachers in the model estimated as in the previous section. Such a test is presented and used in the empirical application in Section 4.

If the covariates in the model are finite dimensional, dynamics in peer group assignment can easily be accomodated in our framework.

Consider the model:

$$y_{it} = c_i + w_{it}\theta_0 + (\gamma_0 - 1) \frac{1}{n_{p_{it}}} \sum_{j \in p_{it}} (c_j + w_{jt}\theta_0) + u_{it}$$

$$E(u_{it}|z^t, p^t) = 0$$

The same transformation as in the previous sections can be applied:

$$y_{it} - \eta_0 \bar{y}_{it} = c_i + w_{it}\theta_0 + \ddot{u}_{it}$$

$$E(\ddot{u}_{it}|z^t, p^t) = 0$$

Hence, for $t = 1, \dots, T - 1$:

$$\dot{y}_{it} - \eta_0 \dot{\bar{y}}_{it} = \dot{w}_{it}\theta_0 + \dot{\ddot{u}}_{it}$$

$$E(\dot{\ddot{u}}_{it}|z^t, p^t) = 0$$

and a consistent locally efficient estimator could be obtained for this model in a similar way as in the previous section.

Hence we see that with our approach, the dynamics of peer group assignment do not have to be fully specified to estimate peer effects consistently, unlike with the approach found in Goldsmith-Pinkham and Imbens (2013) where the form of the dynamics in peer group formation is assumed to be correctly specified.

If unrestricted teacher effects, in addition to student effects, were included in the model, there would be to my knowledge no transformation of the data that could yield moment conditions to estimate η_0 , θ_0 without additional restrictions. Identification and estimation of the parameters of interest could be achieved under additional restrictions on the dynamics of teacher assignment for instance, but this is left for future research.

3.3 Extension to a more general model of social interactions

For simplicity here I consider the case of strictly exogenous covariates and peer groups assignment, but the same extensions to the presence of feedback as in subsections 3.1 and 3.2 above could be made.

Consider the model:

$$y_t = (I + G_t(\theta_0))(x_t\beta_0 + c) + u_t$$

$$E(u_t|x, G) = 0$$

where $G_t(\theta_0)$ is a function of x, G and governs the magnitude of peer effects.

This model is subject to the same identification issue as the linear-in-means model of the previous section. Indeed if $\beta_0 = 0$ and $G_t(\theta_0) = G_s(\theta_0)$ for any $t, s = 1, \dots, T$, c could be redefined to be $(I + G_1(\theta_0))c$ so that a model without peer effects would be observationally equivalent to a model that has peer effects.

As long as $(I + G_t(\theta_0))$ is invertible, let $\ddot{u}_t = (I + G_t(\theta_0))^{-1}u_t$:

$$(I + G_t(\theta_0))^{-1}y_t = x_t\beta_0 + c + \ddot{u}_t$$

$$E(\ddot{u}_t|x) = 0$$

Hence, as in the previous section, one can stack the above model across time, let $g_{it} = [1[i = j]]_{j=1,\dots,n}$. We have:

$$M_{x,g}(I + G(\theta_0))^{-1}y = M_{x,g}\ddot{u}$$

$$E(M_{x,g}\ddot{u}|x, G) = 0$$

Hence, as long as the above conditions identify θ_0 , and proper regularity conditions hold, a consistent, asymptotically normal and locally efficient estimator can be proposed as in Section 3.

Example: Leave-one-out linear in means model (To be included in future versions of the draft) With endogenous and exogenous effects.

4 Application to estimating the importance of teacher quality for student achievement

4.1 Why accounting for peer effects matters when evaluating the importance of teacher quality for student achievement

To be included in future versions of the draft

4.2 Empirical evaluation of the importance of teacher quality in the presence of peer effects with data from North Carolina

To be included in future versions of the draft

5 Conclusion

References

- Abowd, J. M., F. Kramarz, and D. N. Margolis (1999, March). High Wage Workers and High Wage Firms. *Econometrica* 67(2), 251–333.
- Andrabi, T., J. Das, A. Ijaz Khwaja, and T. Zajonc (2011, July). Do Value-Added Estimates Add Value? Accounting for Learning Dynamics. *American Economic Journal. Applied Economics* 3(3), 29–54.
- Angrist, J. D. (2014, October). The perils of peer effects. *Labour Economics* 30, 98–108.
- Aral, S., L. Muchnik, and A. Sundararajan (2009, December). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* 106(51), 21544–21549.
- Aral, S. and D. Walker (2011, August). Creating Social Contagion Through Viral Product Design: A Randomized Trial of Peer Influence in Networks. *Management Science* 57(9), 1623–1639.
- Arcidiacono, P., G. Foster, N. Goodpaster, and J. Kinsler (2012, November). Estimating spillovers using panel data, with an application to the classroom. *Quantitative Economics* 3(3), 421–470.
- Bramoullé, Y., H. Djebbari, and B. Fortin (2009, May). Identification of peer effects through social networks. *Journal of Econometrics* 150(1), 41–55.
- Dahl, G. B., K. V. Løken, and M. Mogstad (2014). Peer Effects in Program Participation. *American Economic Review* 104(7), 2049–74.
- Davezies, L., X. D’Haultfoeuille, and D. Fougère (2009, November). Identification of peer effects using group size variation. *Econometrics Journal* 12(3), 397–413.

- Duflo, E., P. Dupas, and M. Kremer (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review* 101(5), 1739–74.
- Goldsmith-Pinkham, P. and G. W. Imbens (2013, May). Social Networks and the Identification of Peer Effects. *Journal of Business & Economic Statistics* 31(3), 253–264.
- Graham, B. S. (2008, May). Identifying Social Interactions Through Conditional Variance Restrictions. *Econometrica* 76(3), 643–660.
- Imberman, S. A., A. D. Kugler, and B. I. Sacerdote (2012). Katrina’s Children: Evidence on the Structure of Peer Effects from Hurricane Evacuees. *American Economic Review* 102(5), 2048–82.
- Lee, L.-f. (2007, October). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics* 140(2), 333–374.
- Lee, L.-f. and J. Yu (2010, February). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* 154(2), 165–185.
- Manresa, E. (2015). Estimating the Structure of Social Interactions Using Panel Data. *Working Paper*.
- Manski, C. F. (1993, July). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies* 60(3), 531–542.
- Sacerdote, B. (2001, May). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics* 116(2), 681–704.
- Verdier, V. (2015). Estimation of Dynamic Panel Data Models with Cross-Sectional Dependence: Using Cluster Dependence for Efficiency. *Journal of Applied Econometrics* (Forthcoming).
- Wooldridge, J. M. (2010, October). *Econometric Analysis of Cross Section and Panel Data* (second ed.). The MIT Press.

Zimmerman, D. J. (2003, February). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics* 85(1), 9–23.

Identification and Estimation of Peer Effects with Panel Data

Appendix and Supplemental Material

Valentin Verdier*

September 18, 2015

*Assistant Professor, Department of Economics, University of North Carolina, Chapel Hill, NC 27599, United States. Tel.: +1 919-966-3962. E-mail address: vverdier@email.unc.edu.

Appendix

To simplify notation, we will use C as a generic large constant, so that if we can show that a is less than some arbitrary large constant, we will always write $a \leq C$, independently of what the arbitrary large constant actually is.

A Results for the model with strict exogeneity

Lemma 1 shows that any generalized inverse of $g'g$ used to compute M_g yields the same estimator. Hence we pick the generalized inverse that consists in selecting all teachers such that the student indicators and selected teacher are linearly independent and taking the inverse of the resulting cross-product matrix. Let $\mathcal{A} \subseteq \{1, 2, \dots, D\}$ be a set of indexes of teachers such that the matrix $[1[i = s]]_{i=1, \dots, n, t=1, \dots, T}^{s=1, \dots, n}, [1[d_{it} = d]]_{i=1, \dots, n, t=1, \dots, T}^{d \in \mathcal{A}}$ has full rank and all of the vectors in $[1[d_{it} = d]]_{i=1, \dots, n, t=1, \dots, T}^{d=1, \dots, D, d \notin \mathcal{A}}$ are linear functions of the vectors in $[1[i = s]]_{i=1, \dots, n, t=1, \dots, T}^{s=1, \dots, n}, [1[d_{it} = d]]_{i=1, \dots, n, t=1, \dots, T}^{d \in \mathcal{A}}$.

Let $g_1 = [1[i = s]]_{i=1, \dots, n, t=1, \dots, T}^{s=1, \dots, n}$, $g_2 = [1[d_{it} = d]]_{i=1, \dots, n, t=1, \dots, T}^{d \in \mathcal{A}}$, $g_{2it} = [1[d_{it} = d]]^{d \in \mathcal{A}}$ and $\bar{g}_{2i} = \frac{1}{T} \sum_{s=1}^T g_{2is}$.

$$\text{We set } M_g = I - [g_1, g_2] \begin{pmatrix} g_1' g_1 & g_1' g_2 \\ g_2' g_1 & g_2' g_2 \end{pmatrix}^{-1} \begin{bmatrix} g_1' \\ g_2' \end{bmatrix}.$$

By the Frisch and Waugh theorem, $M_g = (M_{g_1} - M_{g_1} g_2 (g_2' M_{g_1} g_2)^{-1} g_2' M_{g_1})$.

Hence:

$$z' M_g \ddot{u} = \sum_{i=1, \dots, n} \sum_{t=1, \dots, T} (z'_{it} - \bar{z}'_i - a_n (g'_{2it} - \bar{g}'_{2i})) \ddot{u}_{it} \quad (\text{A.1})$$

where $a_n = z' M_{g_1} g_2 (g_2' M_{g_1} g_2)^{-1}$ and $\bar{z}'_i = \frac{1}{T} \sum_{s=1}^T z'_{is}$.

Define $\tilde{z}'_{it} = z'_{it} - \bar{z}'_i - a_n (g'_{2it} - \bar{g}'_{2i})$ so that we can write:

$$z' M_g \ddot{u} = \sum_{i=1}^n \sum_{t=1}^T \tilde{z}'_{it} \ddot{u}_{it} \quad (\text{A.2})$$

and:

$$z' M_g z = \sum_{i=1}^n \sum_{t=1}^T \tilde{z}'_{it} \tilde{z}_{it} \quad (\text{A.3})$$

Lemma 1. *For any two generalized inverses of $g'g$, A and B , let $M_{gA} = I - gAg'$ and $M_{gB} = I - gBg'$, then $z' M_{gA} = z' M_{gB}$.*

Proof. We can always write $z = g\delta + e$ where $g'e = 0$. Then:

$$\begin{aligned} z' M_{gA} &= \delta' g' (I - gAg') + e' (I - gAg') \\ &= e' \\ &= \delta' g' (I - gBg') + e' (I - gBg') \\ &= z' M_{gB} \end{aligned}$$

□

Proof of Proposition 2:

a) Proof of consistency of $\hat{\eta}$ and $\hat{\theta}$.

a-i) $\frac{1}{n} z' M_g \ddot{u} = o_p(1)$

Fix t , since $\frac{1}{n} \sum_i \tilde{z}_{it} \ddot{u}_{it} = o_p(1)$ implies $\frac{1}{n} \sum_{t=1}^T \sum_i \tilde{z}_{it} \ddot{u}_{it} = o_p(1)$ for fixed T , and consider the case where z_{it} is a scalar, since consistency can be shown element by element in the case

where z_{it} is a vector. From (...), $E(\frac{1}{n} \sum_i \tilde{z}_{it} \ddot{u}_{it}) = 0$. In addition:

$$\begin{aligned}
\text{Var}(\frac{1}{n} \sum_i \tilde{z}_{it} \ddot{u}_{it}) &= \frac{1}{n^2} \sum_i \sum_j E(\tilde{z}_{it} \tilde{z}_{jt} \ddot{u}_{it} \ddot{u}_{jt}) \\
&= \frac{1}{n^2} \sum_i \sum_j E(1[p_{it} = p_{jt}] \tilde{z}_{it} \tilde{z}_{jt} \ddot{u}_{it} \ddot{u}_{jt}) \\
&\leq \frac{1}{n^2} \sum_i \sum_j P(p_{it} = p_{jt})^{\frac{\delta}{1+\delta}} E(|\tilde{z}_{it} \tilde{z}_{jt} \ddot{u}_{it} \ddot{u}_{jt}|^{1+\delta})^{\frac{1}{1+\delta}} \\
&\leq C \frac{1}{n^2} \sum_i \sum_j P(p_{it} = p_{jt})^{\frac{\delta}{1+\delta}} \\
&\leq C (\frac{1}{n^2} \sum_i \sum_j P(p_{it} = p_{jt}))^{\frac{\delta}{1+\delta}}
\end{aligned}$$

where the second equality follows from Assumption 1, the first inequality follows from Hölder's inequality, the second inequality follows from Lemma 3, and the third inequality follows from Jensen's inequality.

But since $\sum_j 1[p_{it} = p_{jt}] \leq M$ by Assumption 1, then $\sum_i \sum_j 1[p_{it} = p_{jt}] \leq Mn$, so that:

$$\sum_i \sum_j P(p_{it} = p_{jt}) \leq Mn \tag{A.4}$$

Therefore:

$$\text{Var}(\frac{1}{n} \sum_i \tilde{z}_{it} \ddot{u}_{it}) = o(1) \tag{A.5}$$

so that $\frac{1}{n} \sum_i \tilde{z}_{it} \ddot{u}_{it} = o_p(1)$, so that $z' M_g \ddot{u} = o_p(1)$ since $\sum_{t=1}^T o_p(1) = o_p(1)$ for T fixed.

$$\text{a-ii) } \frac{1}{n} z' M_g [\bar{y}, w] - E(\frac{1}{n} z' M_g z) = o_p(1)$$

From (...):

$$\bar{y} = \bar{w} \gamma_0 \theta_0 + \gamma_0 \bar{g}_1 c + g_2 e + \bar{u} \tag{A.6}$$

Hence:

$$\frac{1}{n}z' M_g[\bar{y}, w] = \frac{1}{n}z' M_g z + \frac{1}{n}z' M_g[\gamma_0 \bar{g}_1(c - E(c|x, p)), 0] + \frac{1}{n}z' M_g[\gamma_0 \bar{u}, 0] \quad (\text{A.7})$$

$\frac{1}{n}z' M_g \bar{u} = o_p(1)$ can be shown as in part a-i) of this proof.

To show $\frac{1}{n}z' M_g z - E(\frac{1}{n}z' M_g z) = o_p(1)$, again fix t and consider the case where z_{it} is a scalar, since the same reasoning can be applied to any cross-product in the elements of z_{it} when z_{it} is a vector.

We want to show that $\frac{1}{n} \sum_i (\tilde{z}_{it}^2 - E(\tilde{z}_{it}^2)) = o_p(1)$. Let $b_{it} = \tilde{z}_{it}^2 - E(\tilde{z}_{it}^2)$

$$\begin{aligned} E\left(\left(\frac{1}{n} \sum_i b_{it}\right)^2\right) &= \frac{1}{n^2} \sum_i \sum_{j \in r_i} E(b_{it} b_{jt}) \\ &\leq \frac{1}{n^2} \sum_i \sum_{j \in r_i} C \\ &= C \frac{1}{n^2} \sum_i n_{r_i} \\ &= C \frac{\sum_r n_r^2}{n^2} \\ &\leq C \frac{\max\{n_r : r = 1, \dots, R\}}{n} \end{aligned}$$

where the first inequality follows from Lemma 3.

From Assumption 3, the last quantity on the right hand side is $o(1)$. Hence $\frac{1}{n} \sum_i (\tilde{z}_{it}^2 - E(\tilde{z}_{it}^2)) = o_p(1)$ and with the reasoning extended to multivariate z_{it} we have $\frac{1}{n} \sum_i (\tilde{z}'_{it} \tilde{z}_{it} - E(\tilde{z}'_{it} \tilde{z}_{it})) = o_p(1)$, hence $\frac{1}{n} z' M_g z - E(\frac{1}{n} z' M_g z) = \frac{1}{n} \sum_i \sum_{t=1}^T (\tilde{z}'_{it} \tilde{z}_{it} - E(\tilde{z}'_{it} \tilde{z}_{it})) = o_p(1)$.

Finally we show that $\frac{1}{n} z' M_g \bar{g}_1(c - E(c|x, p)) = o_p(1)$. $\frac{1}{n} z' M_g \bar{g}_1(c - E(c|x, p)) = \frac{1}{n} \sum_{i,t} \tilde{z}'_{it} (\bar{c}_{it} - E(\bar{c}_{it}|x, p))$. Fix t , consider the case where z_{it} is a scalar, and let $b_{it} = \tilde{z}_{it} (\bar{c}_{it} - E(\bar{c}_{it}|x, p))$.

Then:

$$\begin{aligned} E\left(\left(\frac{1}{n} \sum_i b_{it}\right)^2\right) &= \frac{1}{n^2} \sum_i \sum_{j \in r_i} E(b_{it} b_{jt}) \\ &\leq \frac{1}{n^2} \sum_i \sum_{j \in r_i} C \end{aligned}$$

since $E(z_{it}^{4+\delta}) < C$ is shown in Lemma 3, $E(\bar{c}_{it}^{4+\delta}) < C$ is shown in Lemma 3, and $E(E(\bar{c}_{it}|x, p)^{4+\delta}) < C$ is implied by Jensen's inequality and the law of iterated expectations. Hence since we have already shown that $\frac{1}{n^2} \sum_i \sum_{j \in r_i} C = o(1)$, we have $\frac{1}{n} z' M_g \bar{g}_1 (c - E(c|x, p)) = o_p(1)$.

Hence we have $\frac{1}{n} z' M_g [\bar{y}, w] - E(\frac{1}{n} z' M_g z) = o_p(1)$

$$\text{a-iii)} \quad \begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} \xrightarrow{p} \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix}$$

Since $\tilde{z}'_{it} = z'_{it} - a_n g'_{it}$ and by Assumption 2, $\lambda_{\min}(\frac{1}{n} \sum_i \sum_t E((z'_{it} - a_n g'_{it})(z_{it} - g_{it} a'_n))) > c > 0 \forall n > N$, we have $\lambda_{\min}(\frac{1}{n} z' M_g [\bar{y}, w]) > c > 0$ w.p.a. one as $n \rightarrow \infty$. Hence we have $\lambda_{\max}((\frac{1}{n} z' M_g [\bar{y}, w])^{-1}) < \frac{1}{c} < \infty$ w.p.a. one, so that $(\frac{1}{n} z' M_g [\bar{y}, w])^{-1} = O_p(1)$.

Hence combining this result with $\frac{1}{n} z' M_g \ddot{u} = o_p(1)$, we have $\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} = O_p(1) o_p(1) = o_p(1)$.

$$\text{b)} \text{ Proof of } \sqrt{n} V_n^{-1/2} A_n \left(\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} \right) \xrightarrow{d} N(0, I)$$

b-i) Proof that $\sqrt{n} \Sigma_n^{-\frac{1}{2}} \frac{1}{n} \sum_{i,t} \tilde{z}_{it} \ddot{u}_{it} \xrightarrow{d} N(0, I)$, where $\Sigma_n = \frac{1}{n} \text{Var}(\sum_{i,t} \tilde{z}_{it} \ddot{u}_{it} | x, p)$

Conditioning on x, p , from Assumption 1, we have independence across students who have never been in the same classroom. Hence observations could be located on a lattice in \mathcal{R}^T by defining an observation's location as $l_i = \{p_{it}\}_{t=1, \dots, T}$. We can also define the distance between two observations as $\delta_{ij} = \min\{p_{it} - p_{jt}\}_{t=1, \dots, T}$. We have independence for $\delta_{ij} > 0$, and for any observation i , $\#(\{j : \delta_{ij} < 1\}) \leq MT$, so we have so called increasing domain

asymptotics. (Leung 2015 showed that the increasing domain asymptotics framework can be applied as long as any observation has less than some finite number of neighbors in a neighborhood of radius $\delta_{min} > 0$, which is a slight relaxation of the assumption in Jenish and Prucha 2009 that any observation has no neighbor in a neighborhood of radius $\delta_{min} > 0$.)

Hence, conditional on x, p , since independence is stronger than α -mixing, we can apply the central limit theorem found in Jenish and Prucha 2009 as long as:

$$E\left(\left|\frac{1}{T} \sum_t \tilde{z}_{it} \ddot{u}_{it}\right|^{2+\delta} | x, p\right) \leq \mathbf{B}$$

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \lambda_{min}(\text{Var}(\sum_{i,t} \tilde{z}_{it} \ddot{u}_{it} | x, p)) > 0$$

where $\mathbf{B} \in (0, +\infty)$ is a function of x, p .

In the proof of Lemma 3 we find:

$$\tilde{z}_{it} \leq 2\mathbf{B}^{\frac{1}{4+\delta}} + 2\mathbf{B}^{\frac{1}{4+\delta}} b(MT^2)^b$$

$$E(\ddot{u}_{it}^{4+\delta} | x, p)^{\frac{1}{4+\delta}} \leq \mathbf{B} + M\mathbf{B}$$

which satisfies the first condition.

The second condition is proved in Lemma 4.

Hence we have:

$$\sqrt{n} \Sigma_n^{-\frac{1}{2}} \frac{1}{n} \sum_{i,t} \tilde{z}_{it} \ddot{u}_{it} | x, p \xrightarrow{d} N(0, I) \tag{A.8}$$

Hence, since $f_{x,p}$ is integrable, we can apply the Lebesgues dominated convergence theorem to obtain:¹

$$\sqrt{n} \Sigma_n^{-\frac{1}{2}} \sum_{i,t} \tilde{z}_{it} \ddot{u}_{it} \xrightarrow{d} N(0, I) \tag{A.9}$$

¹By the dominated convergence theorem, $F_n | x, p \rightarrow \Phi$ implies $\int F_n | x, p f_{x,p} dx, p \rightarrow \Phi$ if $f_{x,p}$ is integrable since $F_n | x, p \in [0, 1]$.

b-ii) Proof that $\sqrt{n}(V_n)^{-\frac{1}{2}}A_n \begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \sqrt{n}\Sigma_n^{-\frac{1}{2}} \sum_{i,t} \tilde{z}_{it}\ddot{u}_{it} = o_p(1)$

$\sqrt{n}\Sigma_n^{-\frac{1}{2}} \sum_{i,t} \tilde{z}_{it}\ddot{u}_{it} = O_p(1)$, so that if $(V_n)^{-\frac{1}{2}}A_n(\frac{1}{n}z' M_g[\bar{y}, w])^{-1}\Sigma_n^{\frac{1}{2}} - I = o_p(1)$ we have the desired result.

$$(V_n)^{-\frac{1}{2}}A_n(\frac{1}{n}z' M_g[\bar{y}, w])^{-1}\Sigma_n^{\frac{1}{2}} - I = (V_n)^{-\frac{1}{2}}\Sigma_n^{\frac{1}{2}} - I + (V_n)^{-\frac{1}{2}}(A_n(\frac{1}{n}z' M_g[\bar{y}, w])^{-1} - I)\Sigma_n^{\frac{1}{2}} \quad (\text{A.10})$$

$A_n(\frac{1}{n}z' M_g[\bar{y}, w])^{-1} - I = o_p(1)$ is shown in part a) of the proof of this proposition.

$V_n^{-\frac{1}{2}} = O(1)$ by Assumption 2b), and $V_n^{\frac{1}{2}} = O(1)$ by Assumption 2c).

Lemma 4 shows $\Sigma_n - V_n \rightarrow 0$ a.s., so that $\Sigma_n^{\frac{1}{2}} = O_p(1)$ and $(V_n)^{-\frac{1}{2}}\Sigma_n^{\frac{1}{2}} - I = o_p(1)$ by the continuous mapping theorem.

Hence $(V_n)^{-\frac{1}{2}}\Sigma_n^{\frac{1}{2}} - I = o_p(1)$ and $(V_n)^{-\frac{1}{2}}(A_n(\frac{1}{n}z' M_g[\bar{y}, w])^{-1} - I)\Sigma_n^{\frac{1}{2}} = o_p(1)$ and the desired result follows.

Hence combining part b-i) and b-ii) of this proof, we have:

$$\sqrt{n}(V_n)^{-\frac{1}{2}}A_n \begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} \xrightarrow{d} N(0, I) \quad (\text{A.11})$$

Proof of Proposition 3:

When $u_{it} = v_{it} + (\gamma_0 - 1)\frac{1}{n_{pit}} \sum_{j \in p_{it}} v_{jt}$, $Cov(v_{it}, v_{js}|x, p) = \sigma_v^2 1[i = j, t = s]$, we have $\ddot{u}_{it} = v_{it}$ and $Var(\ddot{u}|x, p) = I$. Hence:

$$\begin{aligned} V_n &= \frac{1}{n}E(z' M_g z) \\ &= A_n \end{aligned}$$

so that:

$$A_n^{-1}V_n A_n^{-1} = E(\frac{1}{n}z' M_g z)^{-1} \quad (\text{A.12})$$

and:

$$W_n = Var(Z' M_g \ddot{u}) = E(Z' M_g Z) \quad (\text{A.13})$$

Hence:

$$\begin{aligned} D &= E\left(\frac{1}{n} Z' M_g [\bar{y}, w]\right)^{-1} \frac{1}{n} Var(Z' M_g \ddot{u}) E\left(\frac{1}{n} Z' M_g [\bar{y}, w]\right)^{-1'} - E\left(\frac{1}{n} z' M_g z\right)^{-1} \\ &= n E(Z' M_g z)^{-1} E(Z' M_g Z) E(z' M_g Z)^{-1} - n E(z' M_g z)^{-1} \\ &= n E(a a') \end{aligned}$$

where $a = E(Z' M_g z)^{-1} Z' M_g - E(z' M_g z)^{-1} z' M_g$, so that D is positive semi-definite.

Proof of Proposition 4:

Part b) of the proof of Proposition 2 shows that $\sqrt{n} V_n^{-\frac{1}{2}} \left(\frac{1}{n} z' M_g [\bar{y}, w]\right) \begin{pmatrix} \hat{\eta} \\ \hat{\theta} \end{pmatrix} - \begin{pmatrix} \eta_0 \\ \theta_0 \end{pmatrix} = \sqrt{n} V_n^{-\frac{1}{2}} \frac{1}{n} \sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it} \xrightarrow{d} N(0, I)$.

Hence we simply have to show that $(\hat{V}_n^{-\frac{1}{2}} V_n^{\frac{1}{2}} - I) = o_p(1)$.

$$\begin{aligned} \hat{V}_n &= \frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} \hat{u}_{it} \hat{u}_{js} \\ &= \frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} \ddot{u}_{it} \ddot{u}_{js} \\ &\quad + \frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} [\bar{y}_{it}, w_{it}] \begin{pmatrix} \hat{\eta} \\ \hat{\theta} \end{pmatrix} - \begin{pmatrix} \eta_0 \\ \theta_0 \end{pmatrix} [\bar{y}_{js}, w_{js}] \begin{pmatrix} \hat{\eta} \\ \hat{\theta} \end{pmatrix} - \begin{pmatrix} \eta_0 \\ \theta_0 \end{pmatrix} \end{aligned}$$

First we show that $\frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} \ddot{u}_{it} \ddot{u}_{js} - V_n = o_p(1)$. Since Lemma 4 shows that $\Sigma_n - V_n \rightarrow 0$ a.s., where $\Sigma_n = \frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} E(\ddot{u}_{it} \ddot{u}_{js} | x, p)$, we only have to show $\frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} \ddot{u}_{it} \ddot{u}_{js} - \Sigma_n = o_p(1)$. Without loss of generality consider the case where \tilde{z}_{it} is a scalar, since convergence can be shown element by element.

Define $\hat{\Sigma}_n = \frac{1}{n} \sum_{i,t} \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}_{it} \tilde{z}_{js} \ddot{u}_{it} \ddot{u}_{js}$.

Define $b_{it} = \sum_{j,s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}_{it} \tilde{z}_{js} \ddot{u}_{it} \ddot{u}_{js} = \sum_s \tilde{z}_{it} \tilde{z}_{is} \ddot{u}_{it} \ddot{u}_{is} + \sum_j 1[j \neq i, p_{it} = p_{jt}] \tilde{z}_{it} \tilde{z}_{jt} \ddot{u}_{it} \ddot{u}_{jt}$.

Under Assumption 1, conditional on x, p , b_{it} is independent of $b_{i't}$ if $p_{is} \neq p_{i's}$ for $s = 1, \dots, T$. Hence as in part b) of the proof of Proposition, we can locate observations on a lattice in \mathcal{R}^T and apply the strong law of large numbers found in Jenish and Prucha 2009 conditional on x, p as long as:

$$E(|b_{it}|^{1+\delta}|x, p) < \mathbf{B} \quad (\text{A.14})$$

where $\mathbf{B} < \infty$ can be a function of x, p .

By Minkowski's inequality:

$$E(|b_{it}|^{1+\delta})^{\frac{1}{1+\delta}} \leq \sum_s E(|\tilde{z}_{it}\tilde{z}_{is}\ddot{u}_{it}\ddot{u}_{is}|^{1+\delta}|x, p) + \sum_j 1[j \neq i, p_{it} = p_{jt}]E(|\tilde{z}_{it}\tilde{z}_{jt}\ddot{u}_{it}\ddot{u}_{jt}|^{1+\delta}) \quad (\text{A.15})$$

The proof of Lemma 3 implies that $E(|\tilde{z}_{it}\tilde{z}_{is}\ddot{u}_{it}\ddot{u}_{is}|^{1+\delta}|x, p) \leq \mathbf{B}$. Hence, from Assumption 1:

$$E(|b_{it}|^{1+\delta})^{\frac{1}{1+\delta}} \leq \mathbf{B}(T + MT - 1) \leq \mathbf{B} \quad (\text{A.16})$$

Hence we have $\frac{1}{n} \sum_i b_{it} \rightarrow \frac{1}{n} \sum_i E(b_{it}|x, p)$ a.s. conditional on x, p , hence we have $\frac{1}{n} \sum_i b_{it} - \frac{1}{n} \sum_i E(b_{it}|x, p) = o_p(1)$ and $\hat{\Sigma}_n - \Sigma_n = \frac{1}{n} \sum_{i,t} b_{it} - \frac{1}{n} \sum_{i,t} E(b_{it}|x, p) = o_p(1)$ conditional on x, p .

Hence, since $P(|\hat{\Sigma}_n - \Sigma_n| > \epsilon|x, p) \in [0, 1]$ and $f_{x,p}$ is integrable, we can use the Lebesgues dominated convergence theorem to obtain $\hat{\Sigma}_n - \Sigma_n = o_p(1)$.

Hence $\frac{1}{n} \sum_{i,t} \sum_{js} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} \ddot{u}_{it} \ddot{u}_{js} - V_n = o_p(1)$.

Next we show that $\frac{1}{n} \sum_{i,t} \sum_{js} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}'_{it} \tilde{z}_{js} [\bar{y}_{it}, w_{it}] \left(\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} \right) [\bar{y}_{js}, w_{js}] \left(\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} \right) = o_p(1)$. Consider one specific summand of $[\bar{y}_{it}, w_{it}] \left(\begin{bmatrix} \hat{\eta} \\ \hat{\theta} \end{bmatrix} - \begin{bmatrix} \eta_0 \\ \theta_0 \end{bmatrix} \right)$ denoted $b_{it}(\hat{\delta} - \delta_0)$, then to prove the last statement we have to prove:

$$(\hat{\delta} - \delta_0)^2 \frac{1}{n} \sum_i \sum_{js} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}_{it} \tilde{z}_{js} b_{it} b_{js} = o_p(1) \quad (\text{A.17})$$

where \tilde{z}_{it} can be treated as a scalar since we can prove consistency separately for each element

of the matrix and each time period as previously.

From Proposition 2, $(\hat{\delta} - \delta_0)^2 = O(\frac{1}{n})$.

In addition, from Assumption 2 and Lemma 3:

$$\begin{aligned} \left| \frac{1}{n} \sum_i \sum_{js} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}_{it} \tilde{z}_{js} b_{it} b_{js} \right| &\leq \frac{1}{n} \sum_i \sum_{js} 1[(j, s) \in \mathcal{S}_{it}] |\tilde{z}_{it} \tilde{z}_{js} b_{it} b_{js}| \\ &\leq \mathbf{B} \frac{1}{n} \sum_i \sum_{js} 1[(j, s) \in \mathcal{S}_{it}] \\ &\leq \mathbf{BMT} \end{aligned}$$

where the last inequality follows from Assumption 1. Hence we have $\frac{1}{n} \sum_i \sum_{js} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}_{it} \tilde{z}_{js} b_{it} b_{js} = O_p(1)$.

Hence we have $\hat{V}_n - V_n = o_p(1)$ and by the continuous mapping theorem we have $\hat{V}_n^{-\frac{1}{2}} V_n - I = o_p(1)$ which is the result that was needed.

Lemma 2. For any $d_0 \in \mathcal{A}$, define $a = j' [g_{2it} + \bar{g}_{2i}]_{i=1, \dots, n, t=1, \dots, T} (g_2' M_{g_1} g_2)^{-1} [1[d_{it} = d_0]]_{i=1, \dots, n, t=1, \dots, T}$, where $j = [1]_{i=1, \dots, n, t=1, \dots, T}$, then $a \leq Cb(MT^2)^b$, where b is defined in section 3 of the paper.

Proof. We have:

$$M_{g_1} g_2 = [1[d_{it} = d]] - \frac{1}{T} \sum_{s=1}^T 1[d_{is} = d]_{i=1, \dots, n, t=1, \dots, T}^{d \in \mathcal{A}} \quad (\text{A.18})$$

so that:

$$\begin{aligned} g_2' M_{g_1} g_2 &= \text{diag}(\{n_d\}_{d \in \mathcal{A}}) - \frac{1}{T} [n_{dd'}]_{d \in \mathcal{A}}^{d' \in \mathcal{A}} \\ &= \text{diag}(\{n_d^{\frac{1}{2}}\}_{d \in \mathcal{A}}) (I_{nT} - \frac{1}{T} [\frac{n_{dd'}}{n_d^{\frac{1}{2}} n_{d'}^{\frac{1}{2}}}]_{d \in \mathcal{A}}^{d' \in \mathcal{A}}) \text{diag}(\{n_d^{\frac{1}{2}}\}_{d \in \mathcal{A}}) \end{aligned}$$

where $n_d = \sum_{i=1, \dots, n, t=1, \dots, T} 1[d_{it} = d]$ and $n_{dd'} = \sum_{i=1, \dots, n, t, s=1, \dots, T} 1[d_{it} = d] 1[d_{is} = d']$.

Hence:

$$(g_2' M_{g_1} g_2)^{-1} = \text{diag}(\{n_d^{-\frac{1}{2}}\}_{d \in \mathcal{A}}) (I_{nT} - \frac{1}{T} [\frac{n_{dd'}}{\frac{1}{2} n_d \frac{1}{2} n_{d'}}]_{d \in \mathcal{A}})^{-1} \text{diag}(\{n_d^{-\frac{1}{2}}\}_{d \in \mathcal{A}}) \quad (\text{A.19})$$

Define $A = \frac{1}{T} [\frac{n_{dd'}}{\frac{1}{2} n_d \frac{1}{2} n_{d'}}]_{d \in \mathcal{A}}$.

We also have:

$$\begin{aligned} j' [g_{2it} + \bar{g}_{2i}]_{i=1, \dots, n, t=1, \dots, T} &= j' [1[d_{it} = d] + \frac{1}{T} \sum_{s=1}^T 1[d_{is} = d]]_{i=1, \dots, n, t=1, \dots, T}^{d \in \mathcal{A}} \\ &= 2[n_d]^{d \in \mathcal{A}} \end{aligned}$$

and we show a few lines below that all of the elements in $(I - A)^{-1}$ are positive, so that:

$$a \leq 2M^{\frac{1}{2}} [1]^{d \in \mathcal{A}} (I - A)^{-1} [1[d_{it} = d_0]]_{i,t=1, \dots, n} \quad (\text{A.20})$$

since $n_d^{\frac{1}{2}} \leq M^{\frac{1}{2}}$ and $n_d^{-\frac{1}{2}} \leq 1$. Hence redefine:

$$a = [1]^{d \in \mathcal{A}} (I - A)^{-1} [1[d_{it} = d_0]]_{i,t=1, \dots, n} \quad (\text{A.21})$$

we just need to show that $a \leq b(MT^2)^b$.

a) $\lambda_{\max}(A) < 1$ so that $(I - A)^{-1} = \sum_{j=0}^{\infty} A^j$:

Since A is symmetric, the maximum eigenvalue of A is found by maximizing over $x \in$

$\mathcal{R}^{\#(\mathcal{A})}$ such that $x'x = 1$:

$$\begin{aligned}
x'Ax &= \frac{1}{T} \sum_{d \in \mathcal{A}} \sum_{d' \in \mathcal{A}} x[d]x[d'] \frac{n_{dd'}}{n_d^{\frac{1}{2}}n_{d'}^{\frac{1}{2}}} \\
&= \frac{1}{T} \sum_{i=1, \dots, n} \sum_{t,s=1, \dots, T} \sum_{d \in \mathcal{A}} \sum_{d' \in \mathcal{A}} x[d]x[d'] \frac{1}{n_d^{\frac{1}{2}}n_{d'}^{\frac{1}{2}}} 1[d_{it} = d]1[d_{is} = d'] \\
&\leq \frac{1}{T} \sum_{i=1, \dots, n} \sum_{t,s=1, \dots, T} \sum_{d \in \mathcal{A}} \sum_{d' \in \mathcal{A}} \frac{x[d]^2}{n_d} 1[d_{it} = d]1[d_{is} = d'] \\
&= \frac{1}{T} \sum_{i=1, \dots, n} \sum_{t,s=1, \dots, T} \sum_{d \in \mathcal{A}} \frac{x[d]^2}{n_d} 1[d_{it} = d] \sum_{d' \in \mathcal{A}} 1[d_{is} = d'] \\
&\leq \frac{1}{T} \sum_{i=1, \dots, n} \sum_{t,s=1, \dots, T} \sum_{d \in \mathcal{A}} \frac{x[d]^2}{n_d} 1[d_{it} = d] \\
&= \frac{1}{T} \sum_{s=1, \dots, T} \sum_{d \in \mathcal{A}} \frac{x[d]^2}{n_d} n_d \\
&= \frac{1}{T} \sum_{s=1, \dots, T} 1 \\
&= 1
\end{aligned}$$

where the first inequality follows from the Cauchy-Schwarz inequality, the second inequality follows from $\sum_{d'=1}^D 1[d_{is} = d'] = 1$.

Hence $\lambda_{max}(A) \leq 1$. However the definition of \mathcal{A} implies that $I - A$ is invertible, which implies $\lambda_{max}(A) < 1$. Hence we have:

$$(I - A)^{-1} = \sum_{j=0}^{\infty} A^j \tag{A.22}$$

b) For any $d_0 \in \mathcal{A}$, d_0 is “connected” to some $d \in \mathcal{A}^c$, i.e. there is d_1, \dots, d_{b-1} such that $n_{d_0 d_1} n_{d_1 d_2} \dots n_{d_{b-1} d} > 0$:

Pick any $d_0 \in \mathcal{A}$. There must exist $d \in \{1, \dots, D\}$ such that $n_{d_0 d} > 0$, since otherwise $[1[d_{it} = d_0]]_{i=1, \dots, n, t=1, \dots, T} = [1[i = s]]_{i=1, \dots, nt=1, \dots, T}$ where $s \in \{1, 2, \dots, n\}$ is such that $d_{st} = d_0$ for some t (and hence all $t = 1, \dots, T$ since d_0 teaches the same students over time), and $d_0 \notin \mathcal{A}$ by the definition of \mathcal{A} , leading to a contradiction.

Case 1: All $d \in \{1, \dots, D\}$ such that $n_{d_0d} > 0$ are in \mathcal{A}^c , so that the statement above is true.

Case 2: There is $d \in \mathcal{A}$ such that $n_{d_0d} > 0$. Then consider $x \in \mathcal{R}^{\#(\mathcal{A})}$ with $x[d] = \frac{n_d^{\frac{1}{2}}}{(\sum_{d' \in \mathcal{A}: d' \leftrightarrow d_0} n_{d'})^{\frac{1}{2}}} 1[d \leftrightarrow d_0]$ for all $d \in \mathcal{A}$, where $d \leftrightarrow d'$ means that there is $\{d_1, \dots, d_{b-1}\} \in \mathcal{A}^{b-1}$ such that $n_{dd_1} n_{d_1 d_2} \dots n_{d_{b-2} d_{b-1}} n_{d_{b-1} d'} > 0$. Then, for any $d, d' \in \mathcal{A}$, $i = 1, \dots, n$, $t, s = 1, \dots, T$ we have:

$$x[d] \frac{1}{n_d^{\frac{1}{2}}} 1[d_{it} = d] 1[d_{is} = d'] = x[d'] \frac{1}{n_{d'}^{\frac{1}{2}}} 1[d_{it} = d] 1[d_{is} = d'] \quad (\text{A.23})$$

since $1[d \leftrightarrow d_0] 1[d_{it} = d] 1[d_{is} = d'] = 1$ if and only if $1[d' \leftrightarrow d_0] 1[d_{it} = d] 1[d_{is} = d'] = 1$.

Hence the Cauchy-Schwarz inequality holds with equality for this choice of x :

$$\begin{aligned} x'Ax &= \frac{1}{T} \sum_{i=1, \dots, n} \sum_{t, s=1, \dots, T} \sum_{d \in \mathcal{A}} \sum_{d' \in \mathcal{A}} \frac{x[d]^2}{n_d} 1[d_{it} = d] 1[d_{is} = d'] \\ &= \frac{1}{T} \sum_{i=1, \dots, n} \sum_{t, s=1, \dots, T} \sum_{d \in \mathcal{A}} \sum_{d' \in \mathcal{A}} \frac{1}{\sum_{d'' \in \mathcal{A}: d'' \leftrightarrow d_0} n_{d''}} 1[d \leftrightarrow d_0] 1[d_{it} = d] 1[d_{is} = d'] \\ &= \frac{\sum_{d \in \mathcal{A}: d \leftrightarrow d_0} \sum_{d' \in \mathcal{A}} n_{dd'}}{\sum_{d \in \mathcal{A}: d \leftrightarrow d_0} \sum_{d'=1, \dots, D} n_{dd'}} \\ &= 1 - \frac{\sum_{d \in \mathcal{A}: d \leftrightarrow d_0} \sum_{d' \in \mathcal{A}^c} n_{dd'}}{\sum_{d \in \mathcal{A}: d \leftrightarrow d_0} \sum_{d'=1, \dots, D} n_{dd'}} \end{aligned}$$

where \mathcal{A}^c is defined by $\mathcal{A}^c = \{1, \dots, D\} \setminus \mathcal{A}$ and the third equality holds because

$$\begin{aligned} Tn_d &= \sum_{s=1}^T n_d \\ &= \sum_{s=1}^T \sum_{i, t} 1[d_{it} = d] \\ &= \sum_{s=1}^T \sum_{i, t} 1[d_{it} = d] \sum_{d'=1}^D 1[d_{is} = d'] \\ &= \sum_{d'=1}^D n_{dd'} \end{aligned}$$

since $\sum_{d'=1}^D 1[d_{is} = d'] = 1$.

Hence $\lambda_{max}(A) \geq 1 - \frac{\sum_{d \in \mathcal{A}: d \leftrightarrow d_0} \sum_{d' \in \mathcal{A}^c} n_{dd'}}$, and for $\lambda_{max}(A) < 1$ we need that for any $d_0 \in \mathcal{A}$, there is $d' \in \mathcal{A}^c$ and $d \in \mathcal{A}$ such that $d_0 \leftrightarrow d$ and $n_{dd'} > 0$. So the statement under b) holds in this case as well.

$$c) a \leq b(MT^2)^b$$

By recursion, we can show:

$$A^j = \frac{1}{T^j} \left[\sum_{d_1, \dots, d_{j-1} \in \mathcal{A}} \frac{n_{dd_1} (\prod_{k=1}^{j-2} n_{d_k d_{k+1}}) n_{d_{j-1} d'}}{n_d^{\frac{1}{2}} \prod_{k=1}^{j-1} n_{d_k} n_{d'}^{\frac{1}{2}}} \right]_{d \in \mathcal{A}} \quad (A.24)$$

Hence we can write:

$$a = \sum_{d \in \mathcal{A}} \sum_{j=0}^{\infty} \frac{1}{T^j} \sum_{d_1, \dots, d_{j-1} \in \mathcal{A}} \frac{n_{dd_1} (\prod_{k=1}^{j-2} n_{d_k d_{k+1}}) n_{d_{j-1} d_0}}{n_d^{\frac{1}{2}} \prod_{k=1}^{j-1} n_{d_k} n_{d_0}^{\frac{1}{2}}} \quad (A.25)$$

From Fatou's lemma, since all of the summands are positive, we have:

$$\begin{aligned} a &= \sum_{j=0}^{\infty} \frac{1}{T^j} \sum_{d_1, \dots, d_{j-1} \in \mathcal{A}} \sum_{d \in \mathcal{A}} \frac{n_{dd_1} (\prod_{k=1}^{j-2} n_{d_k d_{k+1}}) n_{d_{j-1} d_0}}{n_d^{\frac{1}{2}} \prod_{k=1}^{j-1} n_{d_k} n_{d_0}^{\frac{1}{2}}} \\ &\leq M^{\frac{1}{2}} \sum_{j=0}^{\infty} \frac{1}{T^{jn_{d_0}}} \sum_{d_{j-1} \in \mathcal{A}} n_{d_{j-1} d_0} \frac{1}{T^{n_{d_{j-1}}}} \sum_{d_{j-2} \in \mathcal{A}} n_{d_{j-1} d_{j-2}} \cdots \sum_{d_1 \in \mathcal{A}} n_{d_2 d_1} \frac{1}{T^{n_{d_1}}} \sum_{d \in \mathcal{A}} n_{dd_1} \end{aligned}$$

where the inequality simply follows from $n_{d_0}^{\frac{1}{2}} \leq M^{\frac{1}{2}}$, $n_d \geq 1$, and a reordering of the summation.

$$\text{Define } a_j = \frac{1}{T^{n_{d_0}}} \sum_{d_{j-1} \in \mathcal{A}} n_{d_{j-1} d_0} \frac{1}{T^{n_{d_{j-1}}}} \sum_{d_{j-2} \in \mathcal{A}} n_{d_{j-1} d_{j-2}} \cdots \sum_{d_1 \in \mathcal{A}} n_{d_2 d_1} \frac{1}{T^{n_{d_1}}} \sum_{d \in \mathcal{A}} n_{dd_1}.$$

From part b) of this proof:

$$\frac{1}{T^{n_d}} \sum_{d_1 \in \mathcal{A}} n_{dd_1} \frac{1}{T^{n_{d_1}}} \sum_{d_2 \in \mathcal{A}} n_{d_1 d_2} \cdots n_{d_{b-1} d_b} \frac{1}{T^{n_{d_{b-1}}}} \sum_{d_b \in \mathcal{A}} n_{d_{b-1} d_b} \leq 1 - \frac{1}{(MT^2)^b} \quad (A.26)$$

Indeed, let $b_{d, \mathcal{A}^c} = \min\{k : \exists d_1, \dots, d_{k-1} \ni n_{dd_1} n_{d_1 d_2} \cdots n_{d_{k-1} d'}, d' \in \mathcal{A}^c\}$, from part b) we have $b_{d, \mathcal{A}^c} > 0$, so that $b_{d, \mathcal{A}^c} \in \{1, 2, \dots, b\}$.

Suppose $b_{d,\mathcal{A}^c} = 1$, then:

$$\frac{1}{Tn_d} \sum_{d_1 \in \mathcal{A}} n_{dd_1} \frac{1}{Tn_{d_1}} \sum_{d_2 \in \mathcal{A}} n_{d_1 d_2} \dots n_{d_{b-1} d_b} \frac{1}{Tn_{d_{b-1}}} \sum_{d_b \in \mathcal{A}} n_{d_{b-1} d_b} \leq 1 - \frac{1}{(MT^2)} \quad (\text{A.27})$$

since $\frac{1}{Tn_{d_1}} \sum_{d_2 \in \mathcal{A}} n_{d_1 d_2} \dots n_{d_{b-1} d_b} \frac{1}{Tn_{d_{b-1}}} \sum_{d_b \in \mathcal{A}} n_{d_{b-1} d_b} \leq 1$ and $\frac{1}{Tn_d} \sum_{d_1 \in \mathcal{A}} n_{dd_1} = \frac{Tn_d - n_{dd'}}{Tn_d} \leq 1 - \frac{1}{MT^2}$ where $d' \in \mathcal{A}^c$.

If $b_{d,\mathcal{A}^c} = 2$, then:

$$\frac{1}{Tn_d} \sum_{d_1 \in \mathcal{A}} n_{dd_1} \frac{1}{Tn_{d_1}} \sum_{d_2 \in \mathcal{A}} n_{d_1 d_2} \dots n_{d_{b-1} d_b} \frac{1}{Tn_{d_{b-1}}} \sum_{d_b \in \mathcal{A}} n_{d_{b-1} d_b} \leq 1 - \frac{1}{(MT^2)^2} \quad (\text{A.28})$$

since $\frac{1}{Tn_{d_2}} \sum_{d_3 \in \mathcal{A}} n_{d_2 d_3} \dots n_{d_{b-1} d_b} \frac{1}{Tn_{d_{b-1}}} \sum_{d_b \in \mathcal{A}} n_{d_{b-1} d_b} \leq 1$ and $\frac{1}{Tn_d} \sum_{d_1 \in \mathcal{A}} n_{dd_1} \frac{1}{Tn_{d_1}} \sum_{d_2 \in \mathcal{A}} n_{d_1 d_2} \leq \frac{Tn_d - \frac{1}{(MT^2)}}{Tn_d} \leq 1 - \frac{1}{(MT^2)^2}$.

Hence by recursion we see that:

$$\frac{1}{Tn_d} \sum_{d_1 \in \mathcal{A}} n_{dd_1} \frac{1}{Tn_{d_1}} \sum_{d_2 \in \mathcal{A}} n_{d_1 d_2} \dots n_{d_{b-1} d_b} \frac{1}{Tn_{d_{b-1}}} \sum_{d_b \in \mathcal{A}} n_{d_{b-1} d_b} \leq 1 - \frac{1}{(MT^2)^b} \quad (\text{A.29})$$

Hence:

$$a_j \leq \left(1 - \frac{1}{(MT^2)^b}\right)^{\lfloor \frac{j}{b} \rfloor} \quad (\text{A.30})$$

Hence:

$$\begin{aligned} \sum_{j=0}^{\infty} a_j &\leq b \sum_{j=0}^{\infty} \left(1 - \frac{1}{(MT^2)^b}\right)^j \\ &= b(MT^2)^b \end{aligned}$$

□

Lemma 3. Under Assumption 2, $E(\|\tilde{z}_{it}\|^{4+\delta}) < C < \infty$, $E(\|\ddot{u}_{it}\|^{4+\delta}) < C < \infty \forall i = 1, \dots, n$, and $E(\|\bar{c}_{it}\|^{4+\delta}) < C < \infty$.

Proof. a) Proof that $E(\|\tilde{z}_{it}\|^{4+\delta}) < C < \infty$.

By Minkowski's inequality:

$$E(\|\tilde{z}_{it}\|^{4+\delta})^{\frac{1}{4+\delta}} \leq E(\|z_{it}\|^{4+\delta})^{\frac{1}{4+\delta}} + \frac{1}{T} \sum_{s=1}^T E(\|z_{is}\|^{4+\delta})^{\frac{1}{4+\delta}} + E(\|a_n g_{2it}\|^{4+\delta})^{\frac{1}{4+\delta}} + \frac{1}{T} \sum_{s=1}^T E(\|a_n g_{2is}\|^{4+\delta})^{\frac{1}{4+\delta}} \quad (\text{A.31})$$

From Assumption 2 and the law of iterated expectations, $E(\|z_{it}\|^{4+\delta}) < C < \infty$, so we only have to show that $E(\|a_n g_{2it}\|^{4+\delta}) < C < \infty$

From previous derivations in this Appendix, we have:

$$a_n g_{2it} = \left(\sum_{j,s} z'_{js} (g_{2js} - \bar{g}_{2j}) \right) (g'_2 M_{g_1} g_2)^{-1} [1[d_{js} = d_{it}]]_{j=1,\dots,n,s=1,\dots,T} \quad (\text{A.32})$$

But we have shown in the proof of Lemma 2 that all of the elements of $(g'_2 M_{g_1} g_2)^{-1}$ are positive. Hence, by Minkowski's inequality:

$$\begin{aligned} E(\|a_n g_{2it}\|^{4+\delta} | p, g)^{\frac{1}{4+\delta}} &\leq \sum_{j,s} E(\|z'_{js}\|^{4+\delta} | p, g)^{\frac{1}{4+\delta}} (g_{2js} + \bar{g}_{2j}) (g'_2 M_{g_1} g_2)^{-1} [1[d_{js} = d_{it}]]_{j=1,\dots,n,s=1,\dots,T} \\ &\leq \mathbf{B}^{\frac{1}{4+\delta}} \sum_{j,s} (g_{2js} + \bar{g}_{2j}) (g'_2 M_{g_1} g_2)^{-1} [1[d_{js} = d_{it}]]_{j=1,\dots,n,s=1,\dots,T} \\ &\leq \mathbf{B}^{\frac{1}{4+\delta}} b (MT^2)^b \end{aligned}$$

where the second inequality follows from Assumption 2 and the third inequality follows from Lemma 2.

Hence:

$$\begin{aligned} E(\|a_n g_{2it}\|^{4+\delta}) &= E(\|a_n g_{2it}\|^{4+\delta} | p, g) \\ &\leq E(\mathbf{B}^{4+\delta} (MT^2)^{(4+\delta)b}) \\ &\leq C \end{aligned}$$

where the last inequality follows from Assumption 2.

b) Proof that $E(\ddot{u}_{it}^{4+\delta}) < C < \infty$:

First note that:

$$\begin{aligned} \left| \frac{1}{n_{p_{it}}} \sum_j 1[p_{it} = p_{jt}] u_{jt} \right|^{4+\delta} &\leq \frac{1}{n_{p_{it}}} \sum_j 1[p_{it} = p_{jt}] |u_{jt}|^{4+\delta} \\ &\leq \sum_j 1[p_{it} = p_{jt}] |u_{jt}|^{4+\delta} \end{aligned}$$

where the first inequality follows from Jensen's inequality and the second inequality follows from Assumption 1.

Then:

$$\begin{aligned} E\left(\left|\frac{1}{n_{p_{it}}} \sum_j 1[p_{it} = p_{jt}] u_{jt}\right|^{4+\delta} | p\right) &\leq \sum_j 1[p_{it} = p_{jt}] E(|u_{jt}|^{4+\delta} | p) \\ &\leq \mathbf{B} \sum_j 1[p_{it} = p_{jt}] \end{aligned}$$

where the second inequality follows from Assumption 2.

By Assumption 1, $\sum_j 1[p_{it} = p_{jt}] \leq M$, so that $E(|\frac{1}{n_{p_{it}}} \sum_j 1[p_{it} = p_{jt}] u_{jt}|^{4+\delta} | p) \leq \mathbf{B}M$, and, since $E(\mathbf{B}) < \infty$, by the law of iterated expectations:

$$E\left(\left|\frac{1}{n_{p_{it}}} \sum_j 1[p_{it} = p_{jt}] u_{jt}\right|^{4+\delta}\right) \leq C \tag{A.33}$$

where we wrote $ME(\mathbf{B})$ as C .

Hence, by Minkowski's inequality:

$$\begin{aligned} \|\ddot{u}_{it}\|_{4+\delta} &\leq \|u_{it}\|_{4+\delta} + C \\ &\leq C \end{aligned}$$

since $\|u_{it}\|_{4+\delta} \leq C$ by Assumption 2 and the law of iterated expectations, and where we abused notation as described at the beginning of this appendix by writing $C = \sqrt[4+\delta]{C}$ and $C = C + C$.

c) Proof that $E(\bar{c}_{it}^{4+\delta}) < \infty$.

As in part b) of this proof:

$$\begin{aligned} \left| \frac{1}{n_{p_{it}}} \sum_j 1[p_{it} = p_{jt}] c_j \right|^{4+\delta} &\leq \frac{1}{n_{p_{it}}} \sum_j 1[p_{it} = p_{jt}] |c_j|^{4+\delta} \\ &\leq \sum_j 1[p_{it} = p_{jt}] |c_j|^{4+\delta} \end{aligned}$$

so that, under Assumption 2:

$$E(|\bar{c}_{it}|^{4+\delta}) \leq E(\mathbf{BM}) < C \quad (\text{A.34})$$

□

Lemma 4. *Under Assumptions 1-3, we have:*

$$\liminf_{n \rightarrow \infty} \lambda_{\min} \left(\frac{1}{n} \text{Var} \left(\sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it} | x, p \right) \right) > 0 \text{ a.s.} \quad (\text{A.35})$$

Proof. $\frac{1}{n} E(\text{Var}(\sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it} | x, p)) = \frac{1}{n} \text{Var}(\sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it})$ since $E(\sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it} | x, p) = 0$. In addition, $\lambda_{\min}(\frac{1}{n} \text{Var}(\sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it})) > c > 0 \forall n > N$ by Assumption 2. So we simply have to show that $\frac{1}{n} \text{Var}(\sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it} | x, p) - \frac{1}{n} \text{Var}(\sum_{i,t} \tilde{z}'_{it} \ddot{u}_{it}) \rightarrow 0$ a.s. Without loss of generality, consider the case where \tilde{z}_{it} is a scalar, since convergence can be proved element by element.

Under Assumption 3, $\frac{1}{n} \text{Var}(\sum_{i,t} \tilde{z}_{it} \ddot{u}_{it} | x, p) = \frac{1}{n} \sum_{r=1}^R \text{Var}(\sum_{i \in r, t} \tilde{z}_{it} \ddot{u}_{it} | x^r, p^r) = \frac{1}{R} \sum_{r=1}^R \frac{n}{R} \text{Var}(\sum_{i \in r, t} \tilde{z}_{it} \ddot{u}_{it} | x^r, p^r)$

Under Assumption 1:

$$\text{Var} \left(\sum_{i \in r, t} \tilde{z}_{it} \ddot{u}_{it} | x^r, p^r \right) = \sum_{i \in r, t} \sum_{j \in r, s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}_{it} \tilde{z}_{js} E(\ddot{u}_{it} \ddot{u}_{js} | x^r, p^r) \quad (\text{A.36})$$

Hence, under Assumption 2, applying Jensen's inequality twice:

$$\begin{aligned}
\text{Var}\left(\sum_{i \in r, t} \tilde{z}_{it} \ddot{u}_{it} | x^r, p^r\right)^{1+\delta} &\leq \left(\sum_{i \in r, t} \sum_{j \in r, s} 1[(j, s) \in \mathcal{S}_{it}]\right)^{1+\delta} \frac{\sum_{i \in r, t} \sum_{j \in r, s} 1[(j, s) \in \mathcal{S}_{it}] \tilde{z}_{it}^{1+\delta} \tilde{z}_{js}^{1+\delta} E(\ddot{u}_{it}^{1+\delta} \ddot{u}_{js}^{1+\delta} | x^r, p^r)}{\sum_{i \in r, t} \sum_{j \in r, s} 1[(j, s) \in \mathcal{S}_{it}]} \\
&\leq \mathbf{B} \left(\sum_{i \in r, t} \sum_{j \in r, s} 1[(j, s) \in \mathcal{S}_{it}]\right)^{1+\delta} \\
&\leq \mathbf{B} (MT n_r)^{1+\delta}
\end{aligned}$$

Hence:

$$\begin{aligned}
\frac{R^{1+\delta}}{n^{1+\delta}} \text{Var}\left(\sum_{i \in r, t} \tilde{z}_{it} \ddot{u}_{it} | x^r, p^r\right)^{1+\delta} &\leq \mathbf{B} (MT)^{1+\delta} \frac{R^{1+\delta} n_r^{1+\delta}}{n^{1+\delta}} \\
&\leq \mathbf{B} (MT)^{1+\delta} \left(\frac{R n_r}{n}\right)^{1+\delta} \\
&\leq \mathbf{B} C
\end{aligned}$$

since $\frac{R \max\{n_r\}}{n} = O(1)$ by Assumption 3.

Hence:

$$E\left(\frac{R^{1+\delta}}{n^{1+\delta}} \text{Var}\left(\sum_{i \in r, t} \tilde{z}_{it} \ddot{u}_{it} | x^r, p^r\right)^{1+\delta}\right) < C \quad (\text{A.37})$$

Hence Markov's condition is satisfied, and one can apply the strong law of large numbers with independent heterogeneously distributed observations to show:

$$\frac{1}{n} \text{Var}\left(\sum_{i, t} \tilde{z}'_{it} \ddot{u}_{it} | x, p\right) - \frac{1}{n} \text{Var}\left(\sum_{i, t} \tilde{z}'_{it} \ddot{u}_{it}\right) \rightarrow 0 \text{ a.s.} \quad (\text{A.38})$$

as was to be shown. □

B Results for the model with sequential exogeneity

To be included in future versions of the draft.

Proof. □