Estimating Spillovers using Panel Data*

Peter Arcidiacono, Duke University
Gigi Foster, University of New South Wales
Natalie Goodpaster, Analysis Group, Inc.
Josh Kinsler, University of Rochester

May 19, 2010

Abstract

Obtaining consistent estimates of spillovers in an educational context is hampered by at least two issues: selection into peer groups and peer effects emanating from unobservable characteristics. We develop an algorithm for estimating spillovers using panel data that addresses both of these problems. The key innovation is to allow the spillover to operate through the fixed effects of a student’s peers. The only data requirements are multiple outcomes per student and heterogeneity in the peer group over time. We first show that the non-linear least squares estimate of the spillover parameter is consistent and asymptotically normal as $N \to \infty$ with $T$ fixed. We then provide an iterative estimation algorithm that is easy to implement and converges to the non-linear least squares solution. The estimator and algorithm readily adapt to cases where there are correlated effects, heterogeneous spillovers, and accumulation.

*We thank Joe Altonji, Pat Bayer, Jane Cooley, Paul Frijters, John Haltiwanger, Caroline Hoxby, Tom Nechyba, Bruce Sacerdote, seminar participants at Duke University, University of Georgia, Yale University, and participants at the 2004 Labour Econometrics Workshop held at the University of Auckland and the NBER’s Higher Education Workshop for helpful comments, as well as the University of Maryland administrators who have provided us with the data used in this paper. Shakeeb Khan, Nese Yildiz, and the comments of the referees were particularly helpful. We also thank Graham Gower for superior research assistance.
1 Introduction

The question of how peers affect student achievement underlies many debates in applied economics. Peer effects are relevant to the estimation of the impact of affirmative action, school quality, and public school improvement initiatives such as school vouchers, and are central to more immediate concerns such as how best to group students to maximize learning. However, despite this wide field of potential relevance, the empirical estimation of spillovers—whether in the education context or elsewhere—is not straightforward.

There are at least two barriers that must be overcome when estimating spillovers on student achievement. The first is the selection problem. When individuals choose their peer groups, high ability\textsuperscript{1} students may sort themselves into peer groups with other high ability students. With ability only partially observable, positive estimates of peer effects may result even when no peer effects are present because of a positive correlation between the student’s unobserved ability and the observed ability of his peers. Researchers have undertaken a variety of estimation strategies to try to overcome the selection problem,\textsuperscript{2} but significant empirical problems linger, both because researchers only have access to incomplete measures of ability and because peer effects may operate differently when peers are chosen rather than assigned.

A second barrier is that spillovers may work in part through characteristics or actions that are not observed by the econometrician. The importance of peer effects may be significantly understated if the primary channel through which they operate is unobserved. Peer effects through unobservables has received little attention outside of Altonji et al. (2004). Random assignment is able to circumvent the selection problem but the obstacle to estimation posed by peer effects through unobservables remains.

We introduce a new algorithm for estimating spillovers using panel data that overcomes both these obstacles. Our key innovation is that the peer effects are captured through a linear combination of individual fixed effects. Constructing the spillover as a linear combination of

\textsuperscript{1}For ease of exposition we refer to the bundle of individuals’ performance-relevant characteristics as ‘ability.’

\textsuperscript{2}One set of papers uses proxy variables to break the link between unobserved and peer ability (Arcidiacono and Nicholson (2005), Hanushek et al. (2003), and Betts and Morell (1999)). Another set of papers relies on some form of random assignment (Sacerdote (2001), Zimmerman (2003), Winston and Zimmerman (2003), Foster (2006), Lehrer and Ding (2007), and Hoxby (2001)). Finally, researchers have tried to circumvent the endogeneity problem with instrumental variables (Evans et al. (1992)).
individual fixed effects results in a non-linear optimization problem. Estimating individual 
unobserved heterogeneity in non-linear panel data models often results in biased estimates of 
the key parameters of interest—the incidental parameters problem.\(^3\) As \(N\) goes to infinity for 
a fixed \(T\), the estimation error for the fixed effects often does not vanish as the sample size 
grows, contaminating the estimates of the parameters of interest.\(^4\) We show, however, that 
the nonlinear least squares estimate of the spillover parameter is consistent and asymptotically 
normal as \(N \to \infty\) with \(T\) fixed, even though the fixed effects themselves are not consistent.\(^5\)

While nonlinear least squares yields consistent estimates of the spillover parameters, the 
dimensionality of the problem renders nonlinear least squares infeasible. We develop an it-

terative algorithm that, under certain conditions, produces the same estimates as nonlinear 
least squares. The algorithm toggles between estimating the individual fixed effects and the 
spillover parameters. Each iteration lowers the sum of squared errors, with a fixed point 
reached at the nonlinear least squares solution to the full problem.

The original framing of the model is one where only exogenous effects are present: it 
is only the characteristics of the individual and their peers that matter, not their shared 
environment (correlated effects) or their choices (endogenous effects). We show that the 
model and estimation algorithm can easily be adapted to include correlated effects, and that 
estimating the model can often be viewed as solving the reduced form of a structural model 
that has both exogenous and endogenous effects. In a wide class of cases, the methods used in 
previous research to disentangle endogenous effects from exogenous effects\(^6\) can also be applied 
here. Further, our model allows for spillovers to be heterogeneous, with some individuals 
being more susceptible to peer influence or more susceptible to influence from those with 
similar characteristics to their own. Finally, we show that when peer effects persist all the 
identification results hold and the estimation procedures in the transitory case still apply.

The remainder of the paper proceeds as follows. Section 2 presents the baseline model, the

\(^3\)Neyman and Scott (1948) were the first to document the incidental parameters problem. 
\(^4\)Hahn and Newey (2004) provide two methods of bias correction: a panel jackknife and an analytical 
correction. Woutersen (2002) and Fernandez-Val (forthcoming) consider estimators from bias-corrected moment 
conditions. In a similar vein, Arellano and Hahn (2006) and Bester and Hansen (forthcoming) consider bias-
correcting the initial objective function. 
\(^5\)Other special cases where the incidental parameters problem does not require a bias correction are Manski 
\(^6\)See Cooley (2009a) for a discussion.

3
identification result, and the solution algorithm. Section 3 extends the model to incorporate correlated effects, endogenous effects, heterogeneity in peer spillovers, and persistent peer effects. Monte Carlo evidence on the performance of the algorithm is presented in Section 4. Section 5 concludes.

2 Estimating Spillovers with Panel Data

In this section we present a model and estimation strategy for measuring achievement spillovers using student fixed effects. We first consider a case where one’s outcome depends only on one’s own fixed effect and the fixed effects of the other individuals in a pre-defined peer group. We show that it is possible to obtain consistent estimates of the spillover and that there is a computationally cheap way of obtaining the solution. All proofs appear in the appendix.

2.1 Identifying Spillovers Using Panel Data

Our baseline model has individual $i$’s outcome at time $t$ in peer group $n$, $Y_{itn}$, depending upon his own observed and unobserved permanent characteristics, $X_i$ and $u_i$, the unobserved permanent characteristics of each of the other students in his peer group, and a transitory error, $\epsilon_{itn}$. Denote as $M_{tn} + 1$ the total number of individuals in peer group $n$ at time $t$. Each member of peer group $n$ at time $t$ then has $M_{tn}$ peers. Denote as $M_{tn-n}$ the set of individuals (numbering $M_{tn}$) in peer group $n$ at time $t$ with individual $i$ removed. Our baseline specification can then be written:

$$Y_{itn} = X_i\beta_1 + u_i\beta_2 + \frac{1}{M_{tn}} \sum_{j \in M_{tn-n}} (X_j\gamma_1 + u_j\gamma_2) + \epsilon_{itn} \tag{1}$$

The specification in (1) is restrictive along a number of dimensions. There are no endogenous effects, as peer choices do not enter the outcome equation. There are also no correlated effects, as there are no variables to capture the commonality of the environment faced by all members of student $i$’s time-$t$ peer group. Finally, this specification does not allow for heterogeneity in the susceptibility to peer influence.

While each of these restrictions is relaxed in the next section, even in this special case estimation is problematic when peer groups are chosen. In particular, there may be correlation between $u_i$ and the sum of observed peer characteristics, leading to biased estimates of $\gamma_1$. 
Also, we will not be able to capture the peer influence through unobservables, meaning that $\gamma_2$ is inestimable without further assumptions. While random assignment can remove the correlation between $u_i$ and observed peer characteristics, the inability to capture spillovers through unobservables remains.\textsuperscript{7}

We next make one additional assumption: the relevance to outcomes of peer characteristics is proportional to that of own characteristics, meaning that we can write $\gamma_1$ and $\gamma_2$ as\textsuperscript{8}

$$\gamma_1 = \gamma_o \beta_1$$
$$\gamma_2 = \gamma_o \beta_2$$

This implies, for example, that if two dimensions of an individual’s ability are equally important in their effect on $Y_{itn}$, then those two dimensions of peer ability will also be equally important in determining $Y_{itn}$. This same assumption is used in Altonji et al. (2004).

Now define $\alpha_{io}$ as:

$$\alpha_{io} = X_i \beta_1 + u_i \beta_2$$

We can then rewrite equation (1) as:

$$Y_{itn} = \alpha_{io} + \frac{\gamma_o}{M_{tn}} \sum_{j \in M_{tn-i}} \alpha_{jo} + \epsilon_{itn}$$

(2)

An individual’s outcome is then a function of the individual’s fixed effect plus the mean of the fixed effects of the other students in the peer group. Using fixed effects in this way allows us to abstract from many other covariates that may affect student outcomes. All of the heterogeneity in fixed student characteristics that might affect student outcomes, such as birth cohort, sex, IQ, or race is captured with this one measure.

Our goal is then to show the properties of the solution to the non-linear least squares problem:

$$\min_{\alpha, \gamma} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( Y_{itn} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn-i}} \alpha_j \right)^2$$

(3)

Given the non-linearities present in the problem, one may suspect that for fixed $T$, the estimate of $\gamma_o$ that solves the above least squares problem, $\hat{\gamma}$, will be biased as a result of the incidental

\textsuperscript{7}Using random assignment to identify the spillover also disregards the possibility that spillovers operate differently in selected versus randomized contexts.

\textsuperscript{8}For the remainder of the paper, we designate population parameters with an ‘o’ subscript ($\gamma_o$) and estimates of the population parameters with a hat ($\hat{\gamma}$).
parameters problem. However, we show that under mild assumptions this is not the case as long as the peer group changes over time.

**Theorem 1.** Let $N$ denote the number of individuals that a) are observed at least two times and b) satisfy $\sum_{j \in M_{tn-i}} \frac{\alpha_{i}}{M_{tn-i}} \neq \sum_{j \in M_{tn-i}'} \frac{\alpha_{i}}{M_{tn-i}'}$ for some $t, t'$ and for all $i \in \{1, ..., N\}$. If:

1. $E(\epsilon_{itn}\epsilon_{jsk}) = 0 \forall j \neq i, t \neq s, n \neq k$
2. $E(\epsilon_{itn}\alpha_{jo}) = 0 \forall i, j, t, n$
3. $E(\alpha_{io}^4) < \infty \forall i, t, n$
4. $E(\epsilon_{itn}) = 0$ and $E(\epsilon_{itn}^4) < \infty \forall i, t, n$
5. Either $E(\epsilon_{itn}^2|n, t) = E(\epsilon_{jtn}^2|n, t) \forall i, j, t, n$ or $\text{Cov}(\epsilon_{itn}^2, N_i) = 0$ where $N_i$ is the number of observations for individual $i$.
6. $\gamma_o$ lies in the interior of a compact parameter space $\Gamma$, where the largest element of $\Gamma$ is given by $\bar{\gamma}$. Further, $\bar{\gamma} < M$ where $M$ is the smallest class size.

then $\hat{\gamma}$ is $\sqrt{N}$ consistent and asymptotically normal estimator of $\gamma_o$ for fixed $T$.

While most of the above assumptions are standard, a few non-standard assumptions deserve closer inspection. Assumption 1 requires that the residuals across any two observations are uncorrelated. Any correlation across outcomes for the same individual is captured by the individual fixed effect, while correlation in outcomes across individuals in the same peer group is entirely captured by the peer effect. Assumption 5 requires that either the residuals within a peer-group have equal variance, implying only heteroskedasticity across peer groups can be accommodated, or, if heteroskedasticity operates at the individual level, it is uncorrelated with the number times the individual is observed in the data. This assumption is generally applied in virtually all papers in the peer effects literature as standard errors are typically clustered at the class level.

The structure of the proof relies on solving for each of the $\alpha$’s as a function of the data and $\gamma$ and then substituting these functions in for the $\alpha$’s in (3). Minimizing with respect to $\gamma$ alone then yields the result.

---

9. We introduce a procedure for capturing correlated effects that do not work through the peer effect in the next section.
Because the estimator of the $\alpha$’s is inconsistent for fixed $T$, one would expect the estimator of $\gamma_o$ to be downward biased as a result of measurement error. To provide some intuition for why this does not happen, we simplify the model here and assume that there exist $N$ independent blocks of individuals, each containing three students. Within a block, individual 1 is observed twice, first paired with individual 2 in period one, and then paired with individual 3 in period two. Individuals 2 and 3 are only observed when paired with individual 1. Given the simple outcome generation process just described, we then have:

\[
Y_{11n} = \alpha_{1on} + \gamma_o \alpha_{2on} + \epsilon_{11n}
\]
\[
Y_{21n} = \alpha_{2on} + \gamma_o \alpha_{1on} + \epsilon_{21n}
\]

and

\[
Y_{12n} = \alpha_{1on} + \gamma_o \alpha_{3on} + \epsilon_{12n}
\]
\[
Y_{32n} = \alpha_{3on} + \gamma_o \alpha_{1on} + \epsilon_{32n}
\]

where the $n$ subscript denotes a particular block of three students.

The parameter of interest is $\gamma_o$, for which we might try to recover an estimate using a conventional panel-data strategy. Differencing the outcomes of individual 1 across the two time periods yields

\[
Y_{12n} - Y_{11n} = \gamma_o(\alpha_{3on} - \alpha_{2on}) + \epsilon_{12n} - \epsilon_{11n}
\]

If $\alpha_{3on} - \alpha_{2on}$ were observable, estimating the above equation by ordinary least squares would yield a consistent estimate of $\gamma_o$. When $\alpha_{3on} - \alpha_{2on}$ is not observed, a natural, albeit noisy, proxy for $\alpha_{3on} - \alpha_{2on}$ is available since the difference in outcomes between students 2 and 3 is given by

\[
Y_{32n} - Y_{21n} = \alpha_{3on} - \alpha_{2on} + \epsilon_{32n} - \epsilon_{21n}
\]

Hence, we could regress $Y_{12n} - Y_{11n}$ on $Y_{32n} - Y_{21n}$ to obtain an estimate of $\gamma_o$. However, as $N$ goes to infinity with $T$ set at 2, the $\hat{\gamma}$ resulting from the regression of $Y_{12n} - Y_{11n}$ on $Y_{32n} - Y_{21n}$ suffers from the standard measurement error problem:

\[
\text{plim } \hat{\gamma} = \gamma_o \left( \frac{E[(\alpha_{3o} - \alpha_{2o})^2]}{E[(\alpha_{3o} - \alpha_{2o})^2] + E[\epsilon_{32}^2 + \epsilon_{21}^2]} \right)
\]

Since $E[\epsilon_{21}^2 + \epsilon_{32}^2] > 0$, $\hat{\gamma}$ will be inconsistent with a downward bias for fixed $T$, a standard result when a regressor is measured with error.
Our estimator differs from the above approach in that when we concentrate the $\alpha$’s out of the least squares problem, we use all available information. In particular, when taking the first order condition with respect to any $\alpha_i$, there is an effect on $i$’s outcomes, but there is also an effect on all individuals who happen to be paired with $i$. Ignoring this latter effect in the first-order condition would essentially be identical to the exercise just completed. However, when the full first-order condition is written down, the estimated difference between $\alpha_{3n}$ and $\alpha_{2n}$ is not proxied by $Y_{32n} - Y_{21n}$, but is written as a function of $\gamma$ and all observed outcomes:

$$\alpha_{3n} - \alpha_{2n} = \frac{Y_{32n} - Y_{21n} + \gamma(Y_{12n} - Y_{11n})}{1 + \gamma^2}$$

where the above equation can be derived using Lemma 1 in the appendix. This difference in $\alpha$’s is a function of $Y_{12n} - Y_{11n}$, which when used to predict $Y_{12n} - Y_{11n}$ induces a mechanical positive correlation between the regressor and the error term. This correlation exactly offsets the measurement error bias highlighted in the conventional differencing strategy discussed above, allowing us to achieve a consistent estimator of our target, $\gamma_0$.

While concentrating out the $\alpha$’s is useful for proving consistency, the formulas are quite cumbersome and difficult to calculate. Directly solving (3) is also generally not possible because of the dimensionality of the problem. Instead, we consider an iterative estimation strategy that both circumvents the dimensionality problem and yields the same solution as direct maximization. The next section introduces the computational procedure and discusses how it relates to the broader literature regarding estimation of high-dimensional problems.

### 2.2 Computing Spillovers with Panel Data

Before moving directly to the computation of the spillover model outlined in the previous section, we illustrate how our proposed procedure can ameliorate a somewhat simpler computational problem prominently discussed in the literature. An outstanding problem in applied microeconomics is how to estimate models containing multiple types of fixed effects where each set of fixed effects is of a large dimension. We begin with this econometric problem since the iterative method we employ solves the issue of multiple fixed effects en route to estimating spillovers. We focus on two papers in particular: Rivkin et al. (2005) and Abowd et al. (1999), to illustrate the difficulties in estimating large numbers of fixed effects.
Rivkin et al. (2005) model gains in test scores as a function of the observed characteristics of the students, $X_i$, and teacher fixed effects, $\pi_{jo}$, where $i$ indexes individuals and $j$ indexes teachers. The change in test scores from time $t-1$ to $t$ given that the individual has teacher $j$ at time $t$, $\Delta Y$, is then modeled as:

$$\Delta Y = \beta_o X_i + \pi_{jo} + \epsilon_{it}$$  \hspace{1cm} (4)$$

Note that $X_i$ includes characteristics of the students that do not vary over time. However, $X_i$ may not include the full set of individual characteristics that are relevant for achievement gains, and the omitted variables may be correlated with the $\pi_{jo}$’s due to streaming of students and/or systematic selection of certain teachers into classrooms with higher- or lower-ability students. As an alternative, we could estimate the model with both student and teacher fixed effects:

$$\Delta Y = \theta_{io} + \pi_{jo} + \epsilon_{it}$$  \hspace{1cm} (5)$$

However, estimating both sets of fixed effects simultaneously would be infeasible given the large number of students and teachers in their data.

Abowd et al. (1999) are interested in modeling wages as a function of both firm and worker fixed effects. The most basic model they are interested in estimating contains just individual and firm-specific effects in a regression of log earnings. A more interesting case occurs when there are tenure effects that vary across firms. For simplicity, assume that the effects of tenure are linear. Labeling $X_{ijt}$ as the amount of tenure individual $i$ has in firm $j$ at time $t$, the outcome equation is:

$$Y_{ijt} = \theta_{io} + \pi_{jo} + \phi_{jo} X_{ijt} + \epsilon_{ijt}$$  \hspace{1cm} (6)$$

where $\phi_{jo}$ is the firm-specific return to tenure. Abowd and Kramarz (1999) recognize that with over 1 million workers and 500,000 firms, they cannot estimate the above equation directly. Instead, they consider a number of estimation techniques, none of which results in least squares estimates of the firm and worker fixed effects without imposing additional assumptions on the data generating process.\(^\text{(10)}\)

\(^{10}\)Abowd et al. (2002) provide one way to recover the exact least squares estimates of the firm and worker effects when both vectors are of a high dimension. Using the code provided on the authors’ website, we compared the performance of our estimator to the new estimator in Abowd et al. (2002). With 500,000 firms, 1,000,000 workers, and a linear returns-to-tenure parameter, our algorithm produced the same parameter estimate and
Our approach yields least squares estimates of both firm and worker effects in a computationally feasible way without imposing any extraneous orthogonality conditions. Estimating the firm-worker model by OLS solves:

\[
\min_{\theta, \pi, \phi} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{ijt} - \theta_i - \pi_j - \phi_j X_{ijt})^2
\] (7)

Minimizing this function in one step remains infeasible as a result of the large number of firms and workers. Instead, we propose an iterative method that yields OLS estimates of the parameters of interest while circumventing the dimensionality problem. Given starting values for the \( \theta \)'s the algorithm iterates on two steps with the \( q \)th iteration following:

- **Step 1:** Conditional on \( \theta^{q-1} \), estimate \( \pi^q \) and \( \phi^q \) by OLS.

- **Step 2:** Conditional on \( \pi^q \) and \( \phi^q \), estimate \( \theta^q \) by OLS.

The process continues until the parameters converge. Because the sum of squared errors is decreased at each step, we will eventually converge to the parameter values that minimize the least squares problem in (7), regardless of which pair of parameters we guess first to start the algorithm. The primary advantage of our method in applications such as those described above is that it is capable of estimating extremely large sets of fixed effects in a reasonable amount of time.

The model becomes slightly more complicated when the outcomes are allowed to depend on functions of the fixed effects themselves. The iterative estimation strategy we employ involves toggling between estimating the spillover parameter by OLS and estimating the individual fixed effects. The additional complexity arises in the second step. In the firm-worker example, the \( q \)th iteration estimate for \( \theta_i \) does not depend on the \( q \)th iteration estimate of \( \theta_j \). However, in the spillover model, \( i \)'s outcome is a function of \( \alpha_i \) and \( \alpha_j \) for all \( j \in M_{tn} \). This suggests that we need to minimize the conditional likelihood function over all of the \( \alpha \)'s directly. We are able to avoid this by instead repeatedly updating \( \alpha_i \) using the first order condition from the least squares problem.

Consider the first order condition of the nonlinear least squares problem with respect to

---

reduced the required computational time by 25%.
\[ 0 = \sum_{t=1}^{T} \left[ \left( Y_{itn} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_j \right) + \sum_{j \in M_{tn \sim i}} \frac{\gamma}{M_{tn}} \left( Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k \in M_{tn \sim i \sim j}} \alpha_k \right) \right] \]  

(8)

Solving for \( \alpha_i \) yields:

\[ \alpha_i = \frac{\sum_t \left[ Y_{itn} - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_j + \sum_{j \in M_{tn \sim i \sim j}} \frac{\gamma}{M_{tn}} \left( Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k \in M_{tn \sim i \sim j}} \alpha_k \right) \right]}{T + \sum_t \frac{\gamma^2}{M_{tn}}} \]  

(9)

Note that we have pulled out the \( \alpha_i \) terms from the last term in (8) to derive (9). We establish in Theorem 2 the conditions under which, given any initial set of \( \alpha \)'s, repeatedly updating the \( \alpha \)'s using (9) yields a fixed point.

**Theorem 2.** Denote \( f(\alpha) \) as a function mapping from \( \mathbb{R}^N \rightarrow \mathbb{R}^N \) where the \( i \)th element of \( f(\alpha) \) is given by the right hand side of (9) \( \forall i \in N \). A sufficient condition for \( f(\alpha) \) to be a contraction mapping is that the maximum of \( \gamma \) is less than 0.4.

The restriction on the maximum value \( \gamma \) is needed to ensure that the feedback effects are not too strong. With Theorem 2 giving a solution method for the \( \alpha \)'s conditional on the \( \gamma \)'s, our algorithm iterates on estimating the \( \alpha \)'s using \( f(\alpha) \) (taking the \( \gamma \)'s as given), and then estimating the \( \gamma \)'s taking the \( \alpha \)'s as given. Each of these two steps lowers the sum of squared errors and converges to the nonlinear least squares solution. In practice, we have found that the algorithm performs substantially faster if the \( \alpha \)'s are only updated until the sum of squared errors falls before moving on to re-estimating \( \gamma \).\(^{11}\) To summarize, the algorithm is started with an initial guess for the \( \alpha \)'s and iterates on two steps until convergence, with the \( q \)th iteration given by:

- **Step 1:** Conditional on \( \alpha^{q-1} \), estimate \( \gamma^q \) by OLS.
- **Step 2:** Conditional on \( \gamma^q \), update \( \alpha^q \) according to (9).

### 3 Model Extensions

The baseline model makes a number of simplifying assumptions regarding the channel through which the spillover operates, the shared group environment, the form of the spillover effect,\(^{11}\) For most iterations of our models updating the \( \alpha \)'s just once led to a decrease in the sum of squared errors.
and the persistence of the spillover effect. The following sections discuss extensions of the model to address these complications.

3.1 Endogenous Effects

Until this point we have ignored how individual and peer choices may affect outcomes: endogenous effects. The peer effects literature that allows for endogenous effects can be broken out into two classes of models. In the first class, the outcome of interest is itself a choice, and this choice is directly affected by the actual or expected choices of an individual’s peers. Examples of this situation, which illustrates what is commonly known as the “reflection problem”, are the choice of college major and the choice to use drugs.\(^{12}\) In the second class of models, the outcome of interest is not completely within the individual’s control. However, choices by both the individual and the individual’s peers directly affect the outcome. Examples of this situation are wages and final examination scores.\(^{13}\) In these two cases, it is own effort and the effort that other individuals exert in the office or the classroom that affect own outcomes, but others’ outcomes \textit{per se} do not appear in the own-outcome equation. Cooley (2009a) shows that identification is much more complicated in this second class of models.

In Appendix B, we consider the complications introduced by endogenous effects by setting up a structural representation of each of these classes of models and showing what our estimator is able to recover using reduced-form estimation in each case. We illustrate the obstacles to estimation confronted by both observables-based approaches and our fixed-effects-based approach when the underlying structural model contains endogenous effects. We also show what structural parameters can be estimated using our model, and we provide conditions under which we can separate out exogenous from endogenous effects using the reduced form. The key result is that, in a wide variety of endogenous effects settings, the fixed-effects-based approach itself does not restrict one’s ability to separately identify the various peer effect channels when compared with the standard observables-based approaches.

\(^{12}\)See, for example, Giorgi et al. (2007).

\(^{13}\)See, for example, Cooley (2009b).
3.2 Correlated Effects

We next discuss an extension to our baseline model in which common shocks, correlated effects, influence outcomes. If each individual peer group is exposed to a different environment, it is impossible to separate the correlated effects from the exogenous effects without further parameterizations. A restriction that is easily imposed in many data sets is a component of one’s outcome that is teacher-specific. Teachers observed with multiple peer groups then permits the identification of correlated effects. Denote individual i’s outcome at time t, in peer group n, and teacher c as $Y_{itnc}$, and the set of students in teacher c at time t by $M_{tc}$. Letting $\delta_{co}$ indicate the effect of teacher c, adding teacher effects to the achievement equation yields:

$$Y_{itnc} = \alpha_{io} + \frac{\gamma_o}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_{jo} + \delta_{co} + \epsilon_{itnc} \quad (10)$$

The non-linear least squares problem we are now interested in solving takes the following form:

$$\min_{\alpha, \gamma, \delta} \sum_{i} \sum_{t=1}^{T} \left( Y_{itnc} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_j - \delta_c \right)^2 \quad (11)$$

While the consistency of $\hat{\gamma}$ is unchanged regardless of whether we include other time-varying regressors, it is particularly clear here since we can re-write the above expression without $\delta_c$ by demeaning the dependent variable at the teacher level.

The first order condition with respect to $\alpha_i$ changes to reflect the presence of the teacher fixed effects:

$$\alpha_i = \frac{\sum_{t} \left[ Y_{itnc} - \delta_c - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_j + \sum_{j \in M_{tn \sim i}} \frac{\gamma}{M_{tn}} \left( Y_{jtn} - \alpha_j - \delta_c - \frac{\gamma}{M_{tn}} \sum_{k \in M_{tn \sim j-i}} \alpha_k \right) \right]}{T + \sum_{t} \frac{\gamma^2}{M_{tn}}} \quad (12)$$

The updating rule for the $\alpha$’s then follows directly from (12), once the $\alpha$ terms are collected for each student.

For a given set of $\alpha$’s and $\gamma$’s, the teacher fixed effects can be calculated according to:

$$\delta_c = \frac{\sum_{i \in M_{tc}} \left( Y_{itnc} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_j \right)}{\sum_{i \in M_{tc}} 1} \quad (13)$$

The estimation strategy is then the same as without correlated effects, with one additional step. As before, each step of the estimation decreases the sum of squared errors. The algorithm is set by an initial guess of the $\alpha$’s and $\delta$’s and then iterates, with the qth iteration following:
• Step 1: Conditional on $\alpha^{q-1}$ and $\delta^{q-1}$, estimate $\gamma^q$ by OLS.

• Step 2: Conditional on $\delta^{q-1}$ and $\gamma^q$, update $\alpha^q$ according to (12).

• Step 3: Conditional on $\alpha^q$ and $\gamma^q$, update $\delta^q$ according to (13).

Adding other types of fixed effects simply adds additional steps to the estimation, with each set of fixed effects updated in an additional step.

3.3 Heterogenous Effects

The assumption that each student is affected in the same manner by their classmates is restrictive and, as pointed out by Hoxby and Weingarth (2005), not particularly interesting from a policy perspective. In particular, the linear-in-means model implies that, in terms of grades, any winners from reshuffling peers are perfectly balanced by those who lose from the reshuffling. We now relax this assumption by extending our spillover framework to allow for either heterogeneity in the response to peers or heterogeneity in the impact of peers.

The first extension allows the effect of peers to vary with an individual’s own characteristics, a model we refer to as heterogeneity in responsiveness to peers. A simple example would be if female students are influenced more by peer ability than male students. We can express a spillover model that incorporates heterogeneity in the responsiveness to peers as follows:

$$Y_{itn} = \alpha_{io} + \sum_{j \in M_{tn~i}} \frac{\alpha_{jo}}{M_{tn}} (X_i * \gamma_o) + \delta_{co} + \epsilon_{itn}$$

where $X_i$ denotes the observable characteristics of individual $i$.

The second model, which we refer to as heterogeneity in peer influence, allows the strength of the peer effect to depend on the interaction between own and peer characteristics. For example, male students may be affected more by other male students than they are by female students. For ease of exposition, assume that students can be assigned to one of two groups, such as male or female, or black or white. Heterogeneity in peer influence can then be easily incorporated as follows:

$$Y_{itn} = \alpha_{io} + \frac{1}{M_{tn}} \left( \gamma_{1o} \sum_{j \in M_{tn~i}} \alpha_{jo} + \gamma_{2o} \sum_{j \in M_{tn'~i}} \alpha_{jo} \right) + \delta_{co} + \epsilon_{itn}$$
where \( M_{tn}^{\sim i} \) is the set of all students in peer group \( n \) at time \( t \) who are in the same group as \( i \), excluding individual \( i \), and \( M_{tn}^{\prime} \) are all individuals in peer group \( n \) at time \( t \) who are not in the same group as \( i \). This simple model can be extended to allow for interaction-specific spillovers (as opposed to own and other group), or matching based on continuous regressors using a distance measure.

The steps required to estimate either of the above models are identical to the those outlined in the previous section, though each step becomes slightly more complicated. Rather than estimating a single \( \gamma \) by OLS in Step 1, multiple \( \gamma \)'s will need to be estimated. Computationally, Step 2 is also more cumbersome since the first order condition for \( \alpha \) will likely depend on \( i \)'s type and the type of peers with whom \( i \) is grouped.

### 3.4 Accumulation

Until now we have only considered cases where peer effects were transitory. However, our model can accommodate a much larger class of peer interactions. Here we extend the model to allow for accumulation effects. We establish that identification and consistency hold when peer effects persist and that our estimation procedure can easily be adapted to this case.

We consider the case when peer effects operate through permanently increasing one’s own ability.\(^{14}\) Denote \( \alpha_{it0}^* \) as individual \( i \)'s ability at time \( t \) where at time 1 individual ability is \( \alpha_{i10}^* = \alpha_{i1o} \). For dates after \( t = 1 \), \( \alpha_{it0}^* \) is defined by adding the peer effect in each period until \( t - 1 \) to \( \alpha_{i10}^* \):

\[
\alpha_{it0}^* = \alpha_{i1o}^* + \sum_{t'=1}^{t-1} \frac{\gamma_{t'0}}{M_{t'n}} \sum_{j \in M_{t'n}^{\sim i}} \alpha_{jt0}^*
\]

The outcome equation is then:

\[
Y_{itn} = \alpha_{it0}^* + \frac{\gamma_{t0}}{M_{tn}} \sum_{j \in M_{tn}^{\sim i}} \alpha_{jt0}^* + \epsilon_{itn}
\]  

(14)

We can express the non-linear least squares problem as:

\[
\min_{\alpha^*, \gamma} \sum_{i=1}^{N} \sum_{t=1}^{T} \left( Y_{itn} - \alpha_{it0}^* - \frac{\gamma_{t0}}{M_{tn}} \sum_{j \in M_{tn}^{\sim i}} \alpha_{jt0}^* \right)^2
\]  

(15)

\(^{14}\)We do not allow persistence in either the correlated effect or the error term, though the former could easily be accommodated.
subject to $\alpha_{i_0}^j = \alpha_i + \sum_{t'=1}^{t-1} \frac{\gamma}{M_{t' \gamma}} \sum_{j \in M_{t' \gamma} \sim i} \alpha_{j_{t'}}^*$ for all $t > 1$.

We next establish that, given virtually the same conditions in Theorem 1, $\gamma_0$ is identified and that our estimator is consistent and asymptotic normal.

**Theorem 3.** Let $N$ denote the number of individuals that a) are observed at least two times and b) satisfy $\sum_{j \in M_{t' \gamma} \sim i} (\alpha_{j_{t'}}^*) / M_{t' \gamma} \neq \sum_{j \in M_{t' \gamma} \sim i} (\alpha_{j_{t'}}^*) / M_{t' \gamma}$ for some $t, t'$ and all $i \in \{1, \ldots, N\}$. If:

1. $E(\varepsilon_{i_{t'n}}\varepsilon_{j_{t'sk}}) = 0 \forall j \neq i, t \neq s, n \neq k$
2. $E(\varepsilon_{i_{t'n}}\alpha_{j_{t'o}}) = 0 \forall i, j, t, n$
3. $E(\alpha_{i_0}^4) < \infty \forall i, t, n$
4. $E(\varepsilon_{i_{t'n}}) = 0$ and $E(\varepsilon_{i_{t'n}}^4) < \infty \forall i, t, n$
5. Either $E(\varepsilon_{i_{t'n}}^2|n, t) = E(\varepsilon_{j_{t'n}}^2|n, t) \forall i, j, t, n$ or $\text{Cov}(\varepsilon_{i_{t'n}}^2, N_i) = 0$ where $N_i$ is the number of observations for individual $i$. Further, $E(\varepsilon_{i_{t'n}}^2|t) = E(\varepsilon_{j_{t'n}}^2|t')$.

6. $\gamma_0$ lies in the interior of a compact parameter space $\Gamma$, where the largest element of $\Gamma$ is given by $\overline{\gamma}$. Further, $\overline{\gamma} < M$ where $M$ is the smallest class size.

then $\hat{\gamma}$ is $\sqrt{N}$ consistent and asymptotically normal estimator of $\gamma_0$ for fixed $T$.

The only difference between these assumptions and those of Theorem 1 is that assumption 5 is stricter. Namely, we need heteroskedasticity to not be related with time.\(^{15}\)

Although identification, consistency, and asymptotic normality are as easy to show as in the case when there is no accumulation, estimation can become more complicated as we have to keep track of the accumulation of the $\alpha$'s. Here we show that we can express the non-linear least squares problem in such a way that the estimation strategy suggested in section 2.2 applies directly, with weaker conditions on the value of $\gamma_0$.

Consider the two-period case.\(^{16}\) The key insight is that, rather than have $\alpha_{i_0}$ be the individual effect of interest, we can rewrite the non-linear least squares problem such that $\alpha_{i_0}^*$
is the individual effect of interest. Recall that \( \alpha_{i_{2o}} = \alpha_{i_{1o}} + \gamma_o \sum_{j \in M_{1n} \sim i} \alpha_{j_o} \). Our outcome equations in the two periods for individual \( i \) are then:

\[
Y_{i1n} = \alpha_{i_{2o}} + \epsilon_{i1n} \\
Y_{i2n} = \alpha_{i_{2o}} + \gamma_o \sum_{j \in M_{2n} \sim i} \alpha_{j_2} + \epsilon_{i2n}
\]

The non-linear least squares problem is then:

\[
\min_{\alpha^*, \gamma} \sum_{i=1}^{N} \left[ (Y_{i1n} - \alpha_{i2})^2 + \left( Y_{i2n} - \alpha_{i2} - \frac{\gamma_o}{M_{2n}} \sum_{j \in M_{2n} \sim i} \alpha_{j2} \right)^2 \right]
\]

The first order condition of the nonlinear least squares problem with respect to \( \alpha_{i2}^* \):

\[
0 = \sum_{t=1}^{2} (Y_{itn} - \alpha_{i2}^*) - \frac{\gamma}{M_{2n}} \sum_{j \in M_{2n} \sim i} \alpha_{j2} + \sum_{j \in M_{2n} \sim i} \frac{\gamma}{M_{2n}} \left( Y_{j2n} - \alpha_{j2} - \frac{\gamma}{M_{2n}} \sum_{k \in M_{2n} \sim j} \alpha_{k2} \right)
\]

where note that peer characteristics in the first period no longer enter into the first order condition. Solving for \( \alpha_{i2}^* \) yields:

\[
\alpha_{i2}^* = \frac{\sum_{t=1}^{2} Y_{itn} - \frac{\gamma}{M_{2n}} \sum_{j \in M_{2n} \sim i} \alpha_{j2} + \sum_{j \in M_{2n} \sim i} \frac{\gamma}{M_{2n}} \left( Y_{j2n} - \alpha_{j2} - \frac{\gamma}{M_{2n}} \sum_{k \in M_{2n} \sim j} \alpha_{k2} \right)}{2 + \frac{\gamma^2}{M_{2n}}}
\]

(16)

It is useful to compare this expression to the corresponding expression when peer effects are transitory, which is equation (9) evaluated at \( T = 2 \):

\[
\alpha_i = \frac{\sum_{t=1}^{2} Y_{itn} - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn} \sim i} \alpha_{j} + \sum_{j \in M_{tn} \sim i} \frac{\gamma}{M_{tn}} \left( Y_{jtn} - \alpha_{j} - \frac{\gamma}{M_{tn}} \sum_{k \in M_{tn} \sim j} \alpha_{k} \right)}{2 + \sum_{t=1}^{2} \frac{\gamma^2}{M_{tn}}}
\]

Note that the numerator here contains twice as many terms involving \( \gamma \) times peer abilities due to the presence of the \( \alpha \)’s for peers in the first period. Hence the conditions on \( \gamma \) necessary for iteration on (16) to yield a unique fixed point are weaker.

4 Monte Carlo Simulations

To investigate the properties of our iterative estimator, we now run simulations using different assumptions about the composition and selection into the peer groups. In each setting, the model is simulated using 10,000 students. We simulate the model 100 times under various states of the world constructed by varying four dimensions of the problem:
1. *Observations per student*- The number of outcomes observed per student varies across simulations between 2, 5, and 10. 5 is the maximum number of observations a researcher may have when analyzing grade school or high school test score data, and 10 observations is more likely when analyzing grades achieved in university-level teachers. More observations per student implies more accurate measurement of the $\alpha_o$'s.

2. *Students per peer group*- The number of students per peer group varies across simulations between 2 and 15. 2 is the minimum number of observations required to identify a spillover in this type of model. 15 students per peer group is in the range of what might be observed in typical classroom-based data sets.

3. *Selection into classes*- To show that our estimator solves the selection problem, we simulate the model under alternative assignment rules. Under random assignment, the average standard deviation of the $\alpha_o$'s within a peer group equals the standard deviation of $\alpha_o$ in the population. We also simulate the model with selection such that the average standard deviation of the $\alpha_o$'s within a peer group is 75% of the population standard deviation.

4. *Transitory component*- The noisier the outcome measure, the noisier the estimates of the $\alpha_o$’s will be. The distribution of the $\alpha_o$’s is set at $N(0,1)$. The $\epsilon$’s are distributed with mean zero and standard deviation ($\sigma_\epsilon$) equal to 1.15 or 1.95. We also investigate the impact of heteroskedasticity at the peer group level.

The common group-level shock used to model the correlated effect is not statistically associated with the abilities of students in the classes. However, students are sorted into classes based upon ability. Thus, the average standard deviation of abilities within a class is smaller than standard deviation of abilities in the population.

Table 1 documents the model’s performance when the true value of $\gamma_o$ is 0.15. Again, regardless of assignment procedure or section size, $\hat{\gamma}$ is centered around the truth. However, two interesting patterns emerge in the estimates and standard errors of $\hat{\gamma}$. First, $\hat{\gamma}$ is more precisely measured when students are randomly assigned to sections. Selection in this case can be thought of as occurring at two levels: the class level and the section level. These results reflect the fact that selection at the class level confounds the estimate of the correlated
effect and reduces the precision of the section-level peer effect estimate. In fact, if selection occurred only at the section level, the peer effect estimates would be more precise than in the random assignment case (ceteris paribus), since there would be greater variation in peer ability. Second, as the peer group size increases, the precision of $\hat{\gamma}$ decreases.\(^{17}\) This is again related to the variation in peer ability across sections. With smaller section sizes, other things equal, there is greater variation in peer ability across sections which yields more precise estimates of the spillover.

Many applications involving peer effects may want to allow for heteroskedasticity at the class level. Table 2 shows the performance of the algorithm in the presence of heteroskedasticity, including the case when heteroskedasticity is a function of the size of the class. The first panel of results illustrates that heterogenous class size does not affect the performance of the peer effects estimator. $\hat{\gamma}$ remains centered around 0.15 as peer group size varies uniformly between 5 and 15. The second and third panels add heteroskedasticity to the heterogenous class size case. In the second panel, $\sigma_\epsilon$ is drawn from a Normal distribution with a mean of 1.15 or 1.95 and a standard deviation of 0.3. It is assumed that each peer group draws from the same distribution. In the third panel, the mean of the distribution of $\sigma_\epsilon$ shifts according the size of the peer group. The peer effects estimator continues to perform quite well regardless of the type of heteroskedasticity. Across the various distributions of $\sigma_\epsilon$ and sorting scenarios, we estimate a peer effect centered on the truth.

As noted previously, the linear-in-means model may not be the most interesting case from the policy maker’s perspective. We suggested two extensions to the baseline framework that would relax this assumption. Table 3 illustrates the performance of the heterogenous effect models, where the basic structure of the Monte Carlos is kept intact. In each case we assume students are characterized by one binary variable. The results indicate that the estimation framework previously outlined is amenable to heterogenous peer effects.

\(^{17}\)This can be seen in Table 1 since as the number of observations per student increases, we should naturally see an increase in precision—yet we do not, because peer group size increases as well, driving standard errors up. The negative association of peer group size and precision, as well as all other relationships discussed here, have also been verified in numerous additional Monte Carlo exercises; results are available upon request from the authors.
Finally, the potential impact of peers is quite limited if the spillover effect is assumed to be transitory. The last extension to the baseline model allows for peer effects to accumulate over time. Table 4 documents the proposed estimator’s performance when peer effects accumulate in a two period framework across various class sizes. Unobserved student ability in the second period incorporates the spillover effect from the first period. Again, the peer effects estimator yields estimates centered around the truth.

5 Conclusion

Accurate estimation of peer effects in the classroom is plagued by at least two issues, both of which have to do with ability not being fully observed. First, there is selection into the peer group which leads to a positive correlation between unobserved individual ability and observed peer ability. If ignored, this correlation leads to biased-upward estimates of peer effect parameters. On the other hand, underestimation of the effects of peers may result from ignoring peer effects that operate through unobservables.

We present a new iterative method for estimating educational peer effects that overcomes both these obstacles. All that is required is that there are multiple observations per student, with the peer group changing over time. We control for individual effects and allow the peer effect to operate through a linear combination of the other individual fixed effects. We show that our estimator is consistent and asymptotically normal for fixed $T$ as $N$ goes to infinity. We also develop an iterative algorithm that is computationally much cheaper than direct non-linear least squares minimization yet produces the non-linear squares results upon convergence. Monte Carlo results suggest that the model performs quite well even when the number of observations per student is small.

---

18 The simulations in this section differ from the previous simulations in that there are no correlated effects.
References


A Proofs

A.1 Proof of Theorem 1

Proof. For ease of exposition, we illustrate the proof assuming that students are grouped with at most one other student at any point in time. The proof for general class sizes is given in the attached appendix. Keeping with the literature, we also assume a homogeneous peer effect that is proportional to the ability of a student’s peer. The proof can be readily expanded to multiple γ’s.

We consider the limiting case where

1. We observe students for at most two time periods.

2. Within each class there is only one student that is observed for two periods. The other student is observed for only one time period.

Remark 1: Clearly if the estimator is consistent for \( T = 2 \), it is also consistent for \( T > 2 \). The second simplification is equivalent to allowing all of the individual effects in a class but one to vary over time. For example, suppose there were \( 2N \) students observed for two periods, implying that \( 2N \) individual effects would be estimated. We could, however, allow the individual effect to vary over time for one student in each group, making sure to choose these students in such a way that they are matched with someone in both periods whose individual effect does not vary over time.\(^{19}\) \( 3N \) individual effects would then be estimated. Having one individual whose effect varies over time is equivalent to estimating two individual effects—it is the same as having two different individuals who were each observed once. If the estimator is consistent in this case, then it is also consistent under the restricted case when all of the individual effects are time invariant (fixed effects).

Consider the set of students that are observed for two time periods. Each of these students has one peer in period one and one peer in period two. Denote a student block as one student observed for two periods plus his two peers. There are then \( N \) blocks of students, one block...
for each student observed twice, with three students in each block. Denote the first student in each block as the student who is observed twice where $\alpha_{1n}$ is the individual effect. The individual effect for the first classmate in block $n$ is $\alpha_{2n}$, while the individual effect for the second classmate in block $n$ is $\alpha_{3n}$.

The optimization problem is then

$$\min_{\alpha, \gamma} \frac{1}{N} \sum_{n=1}^{N} \left( (y_{11n} - \alpha_{1n} - \gamma \alpha_{2n})^2 + (y_{12n} - \alpha_{1n} - \gamma \alpha_{3n})^2 + \sum_{i=2}^{3} (y_{in} - \alpha_{in} - \gamma \alpha_{1n})^2 \right)$$

(17)

Within each block there are four terms, two residuals for the student observed twice, and a residual for the peer in each period.

Remark 2: Note that, conditional on $\gamma$, the estimates of individual effects in one block will not affect the estimates of the individual effects in another block. Hence, we are able to focus on individual blocks in isolation from one another when concentrating out the $\alpha$’s as a function of $\gamma$.

Our proof then consists of the following five lemmas, each of which is proven later in this appendix.

We first show that the $\alpha$’s can be written as closed form expressions of $\gamma$ and the data.

**Lemma 1**

*The vector of unobserved student abilities, $\alpha$, can be concentrated out of the least squares problem and written strictly as a function of $\gamma$ and $y$. Ability for the student in block $n$ observed in both periods is given by*

$$\alpha_{1n} = \frac{y_{11n} + y_{12n} - \gamma (y_{2n} + y_{3n})}{2(1 - \gamma^2)}$$

*while the abilities for the peers in block $n$ are given by*

$$\alpha_{2n} = \frac{y_{2n} + \gamma^2 y_{3n} - \gamma y_{12n} - \gamma^3 y_{11n}}{1 - \gamma^4}$$

*and*

$$\alpha_{3n} = \frac{y_{3n} + \gamma^2 y_{2n} - \gamma y_{11n} - \gamma^3 y_{12n}}{1 - \gamma^4}$$
We then show the form of the minimization problem when the \( \alpha \)'s are concentrated out.

**Lemma 2**

Concentrating the \( \alpha \)'s out of the original least squares problem results in an optimization problem over \( \gamma \) that takes the following form

\[
\min_{\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{(y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n}))^2}{2(1 + \gamma^2)}
\]

Our nonlinear least squares problem now has only one parameter, \( \gamma \). We are now in a position to investigate the properties of our estimator of \( \gamma_0 \). For ease of notation, define \( q(w, \gamma) \) as

\[
q(w, \gamma) = \frac{(y_{11} - y_{12} + \gamma(y_{3} - y_{2}))^2}{2(1 + \gamma^2)}
\]

where \( w \equiv y \). We let \( W \) denote the subset of \( \mathbb{R}^4 \) representing the possible values of \( w \). Our key result is then Lemma 3, which establishes identification.

**Lemma 3**

\[
E[q(w, \gamma_0)] < E[q(w, \gamma)], \quad \forall \gamma \in \Gamma, \gamma \neq \gamma_0
\]

Theorem 12.2 of Wooldridge (2002) establishes that sufficient conditions for consistency are identification and uniform convergence. Having already established identification, Lemma 4 shows uniform convergence.

**Lemma 4**

\[
\max_{\gamma \in \Gamma} \left| \frac{1}{N} \sum_{n=1}^{N} q(w_n, \gamma) - E[q(w, \gamma)] \right| \overset{p}{\to} 0
\]

Consistency then follows from Theorem 12.2 of Wooldridge: \( \hat{\gamma} \overset{p}{\to} \gamma_0 \).

Finally, we establish asymptotic normality of \( \hat{\gamma} \). Denote \( s(w, \gamma_0) \) and \( H(w, \gamma_0) \) as the first and second derivative of \( q(w, \gamma) \) evaluated at \( \gamma_0 \). Then, Lemma 5 completes the proof.

**Lemma 5**

\[
\sqrt{N}(\hat{\gamma} - \gamma_0) \overset{d}{\to} N(0, A_0^{-1} B_\theta A_0^{-1})
\]
where

\[ A_o \equiv E[H(w, \gamma_o)] \]

and

\[ B_o \equiv E[s(w, \gamma_o)^2] = Var[s(w, \gamma_o)] \]

QED.

**Proof of Lemma 1**

Our objective is to show that the system of equations obtained by differentiating Equation (17) with respect to \( \alpha \) can be expressed as a series of equations in terms of \( \gamma \) and \( y \), and that these expressions are as given in Lemma 1. Again, conditional on \( \gamma \), the estimates of individual effects in one block will not affect the estimates of the individual effects in another block. Thus, we can work with the system of first-order conditions within one block and then generalize the results to the full system of equations. The first-order condition for \( \alpha_{1n} \) (student in each block who is observed in both time periods) is given by

\[
0 = -\frac{2}{N} \left[ (y_{11n} - \alpha_{1n} - \gamma \alpha_{2n}) + (y_{12n} - \alpha_{1n} - \gamma \alpha_{3n}) + \gamma \sum_{i=2}^{3} (y_{in} - \alpha_{in} - \gamma \alpha_{1n}) \right]
\]

while the first-order condition for \( \alpha_{2n} \) and \( \alpha_{3n} \) are respectively given by

\[
0 = -\frac{2}{N} \left[ (y_{2n} - \alpha_{2n} - \gamma \alpha_{1n}) + \gamma (y_{11n} - \alpha_{1n} - \gamma \alpha_{2n}) \right]
\]

and

\[
0 = -\frac{2}{N} \left[ (y_{3n} - \alpha_{3n} - \gamma \alpha_{1n}) + \gamma (y_{12n} - \alpha_{1n} - \gamma \alpha_{3n}) \right]
\]

Within each block, this yields a relatively simple system of 3 equations and 3 unknown abilities. The first order conditions for \( \alpha_{2n} \) and \( \alpha_{3n} \) can be re-arranged such that

\[
\alpha_{2n} = \frac{y_{2n} + \gamma y_{11n} - 2 \gamma \alpha_{1n}}{1 + \gamma^2}
\]

and

\[
\alpha_{3n} = \frac{y_{3n} + \gamma y_{12n} - 2 \gamma \alpha_{1n}}{1 + \gamma^2}
\]

Notice that the equation for \( \alpha_{2n} \) depends only on the own outcome, the outcome of individual one when grouped with individual two, and the ability of individual one. A similar result occurs for \( \alpha_{3n} \). Thus, the only thing linking individuals two and three within a block is the ability of individual one.
Re-arranging the first order condition for $\alpha_{1n}$ such that the $\alpha_{1n}$ are grouped on the left hand side of the equation results in

$$
\alpha_{1n}(2 + 2\gamma^2) = y_{11n} + y_{12n} + \gamma(y_{2n} + y_{3n}) - 2\gamma(\alpha_{2n} + \alpha_{3n})
$$

substituting for $\alpha_{2n}$ and $\alpha_{3n}$ using the previously derived formulas yields

$$
\alpha_{1n}(2 + 2\gamma^2) = y_{11n} + y_{12n} + \gamma(y_{2n} + y_{3n}) - \frac{2\gamma}{1 + \gamma^2}(y_{2n} + y_{3n} + \gamma(y_{11n} + y_{12n}) - 4\gamma\alpha_{1n})
$$

Moving all the $\alpha_{1n}$ terms to the left side and finding common denominators on both sides of the equation results in

$$
\alpha_{1n}(2 + 2\gamma^2)(1 + \gamma^2) - 8\gamma^2 = \frac{(1 + \gamma^2)(y_{11n} + y_{12n} + \gamma(y_{2n} + y_{3n})) - 2\gamma(y_{2n} + y_{3n} + \gamma(y_{11n} + y_{12n}))}{1 + \gamma^2}
$$

Canceling out the denominators and simplifying both sides of the equation yields

$$
\alpha_{1n}(2(1 - \gamma^2)^2) = (1 - \gamma^2)(y_{11n} + y_{12n}) - \gamma(1 - \gamma^2)(y_{2n} + y_{3n})
$$

Dividing both sides of the equation by $2(1 - \gamma^2)^2$ yields the desired result that

$$
\alpha_{1n} = \frac{y_{11n} + y_{12n} - \gamma(y_{2n} + y_{3n})}{2(1 - \gamma^2)}
$$

The solution for $\alpha_{1n}$ can now be substituted back into the first-order conditions for $\alpha_{2n}$ and $\alpha_{3n}$ to yield solutions strictly as functions of $\gamma$ and $y$. Substituting $\alpha_{1n}$ into the equation for $\alpha_{2n}$ and finding a common denominator yields

$$
\alpha_{2n} = \frac{2(1 - \gamma^2)(y_{2n} + \gamma y_{11n}) - 2\gamma(y_{11n} + y_{12n} - \gamma(y_{2n} + y_{3n}))}{2(1 - \gamma^2)(1 + \gamma^2)}
$$

Factoring out the 2 in the numerator and expanding the resulting expression yields

$$
\alpha_{2n} = \frac{(1 - \gamma^2 + \gamma^2)y_{2n} + (\gamma(1 - \gamma^2) - \gamma)y_{11n} - \gamma y_{12n} + \gamma^2 y_{3n}}{(1 - \gamma^2)(1 + \gamma^2)}
$$

Some simple manipulation leads to the final result that

$$
\alpha_{2n} = \frac{y_{2n} + \gamma^2 y_{3n} - \gamma y_{12n} - \gamma^3 y_{11n}}{1 - \gamma^4}
$$

Obtaining the solution for $\alpha_{3n}$ proceeds in exactly the same way, and yields a formula that mirrors the solution for $\alpha_{2n}$ with the appropriate indices changed to reflect when individual three is grouped with individual one. The result is given below

$$
\alpha_{3n} = \frac{y_{3n} + \gamma^2 y_{2n} - \gamma y_{11n} - \gamma^3 y_{12n}}{1 - \gamma^4}
$$
Proof of Lemma 2

Lemma 1 provides a solution for \( \alpha \) strictly as a function of \( y \) and \( \gamma \). We can substitute this solution back into the original optimization problem to derive the result in Lemma 2.

Consider minimizing the sum of squared residuals within a particular block \( n \). There are four residuals within each block, two for the student observed twice, and one each for the corresponding peer. We begin by simplifying the residual for the first observation of the student observed twice, which is given by the expression below

\[
e_{11n} = y_{11n} - \alpha_{1n} - \gamma \alpha_{2n}
\]

Substituting for \( \alpha_{1n} \) and \( \alpha_{2n} \) in \( e_{11n} \) with the results from Lemma 1 results in

\[
e_{11n} = y_{11n} - \frac{y_{11n} + y_{12n} - \gamma(y_{2n} + y_{3n})}{2(1 - \gamma^2)} - \frac{\gamma(y_{2n} + \gamma^2 y_{3n} - \gamma y_{12n} - \gamma^3 y_{11n})}{1 - \gamma^4}
\]

Finding a common denominator and combining like terms in the numerator yields

\[
e_{11n} = \frac{(2(1 - \gamma^4) - (1 + \gamma^2) + 2\gamma^4)y_{11n} - ((1 + \gamma^2) - 2\gamma^2)y_{12n} + (\gamma(1 + \gamma^2) - 2\gamma)y_{2n} + (\gamma(1 + \gamma^2) - 2\gamma^3)y_{3n}}{2(1 - \gamma^4)}
\]

Simplifying the numerators on each of the \( y \) terms and factoring the denominator yields

\[
e_{11n} = \frac{(1 - \gamma^2)y_{11n} - (1 - \gamma^2)y_{12n} - \gamma(1 - \gamma^2)y_{2n} + \gamma(1 - \gamma^2)y_{3n}}{2(1 - \gamma^2)(1 + \gamma^2)}
\]

Finally, we can cancel all the \( (1 - \gamma^2) \) terms to arrive at

\[
e_{11n} = \frac{y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n})}{2(1 + \gamma^2)}
\]

The expression for \( e_{12n} \) as a function of \( \gamma \) and \( y \) can be similarly derived by substituting in \( \alpha_{1n} \) and \( \alpha_{3n} \). However, the expressions for \( e_{12n} \) and \( \alpha_{3n} \) are mirror images of the expressions for \( e_{11n} \) and \( \alpha_{2n} \). Thus, \( e_{12n} \) will take the exact same form as \( e_{11n} \) except the subscripts denoting the period or classmate are swapped. The expression is given below.

\[
e_{12n} = \frac{y_{12n} - y_{11n} + \gamma(y_{2n} - y_{3n})}{2(1 + \gamma^2)}
\]

The residuals for the one observation individuals in each block, \( e_{2n} \) and \( e_{3n} \), are given by

\[
e_{2n} = y_{2n} - \alpha_{2n} - \gamma \alpha_{1n}
\]

\[
e_{3n} = y_{3n} - \alpha_{3n} - \gamma \alpha_{1n}
\]
and

\[ e_{3n} = y_{3n} - \alpha_{3n} - \gamma \alpha_{1n} \]

To write these strictly as functions of \( \gamma \) and \( y \), we again use the results of Lemma 1. Substituting for \( \alpha_{1n} \) and \( \alpha_{2n} \) in \( e_{2n} \) yields

\[
e_{2n} = y_{2n} - \frac{y_{2n} + \gamma^2 y_{3n} - \gamma y_{12n} - \gamma^3 y_{11n}}{1 - \gamma^4} - \frac{\gamma (y_{11n} + y_{12n} - \gamma (y_{2n} + y_{3n}))}{2(1 - \gamma^2)}
\]

Finding a common denominator and simplifying the resulting expressions yields

\[
e_{2n} = \frac{\gamma (y_{12n} - y_{11n} + \gamma (y_{2n} - y_{3n}))}{2(1 + \gamma^2)}
\]

The expression for \( e_{3n} \) is similar to that of \( e_{2n} \), except the subscripts differ to reflect the time period in which individual three is grouped with one. Thus the solution for \( e_{3n} \) will mirror the solution for \( e_{2n} \), except that the appropriate subscripts are swapped across terms. The final expression for \( e_{3n} \) is given below.

\[
e_{3n} = \frac{\gamma (y_{11n} - y_{12n} + \gamma (y_{3n} - y_{2n}))}{2(1 + \gamma^2)}
\]

The original optimization problem written as a function of the residuals in each block \( n \) takes the following form

\[
\min_{\alpha, \gamma} \frac{1}{N} \sum_{n=1}^{N} \left( e_{11n}^2 + e_{12n}^2 + e_{2n}^2 + e_{3n}^2 \right)
\]

Now we can substitute in for each residual using the formulas previously derived. However, a cursory glance at the formulas for \( e_{11n} \), \( e_{12n} \), \( e_{2n} \), and \( e_{3n} \) reveals that

\[
e_{11n} = -e_{12n} = -\gamma e_{2n} = \gamma e_{3n}
\]

Using these relationships we can re-write the least squares problem as

\[
\min_{\alpha, \gamma} \frac{1}{N} \sum_{n=1}^{N} \left( 2 + 2\gamma^2 \right) e_{11n}^2
\]

Substituting in with our solution for \( e_{11n} \) yields

\[
\min_{\gamma} \frac{1}{N} \sum_{n=1}^{N} \left( 2 + 2\gamma^2 \right) \frac{(y_{11n} - y_{12n} + \gamma (y_{3n} - y_{2n}))^2}{(2(1 + \gamma^2))^2}
\]
Canceling terms results in the following optimization problem

\[
\min_{\gamma} \frac{1}{N} \sum_{n=1}^{N} \left( y_{11n} - y_{12n} + \gamma(y_{3n} - y_{2n}) \right)^2 / 2(1 + \gamma^2)
\]

QED

**Proof of Lemma 3**

The population objective function as a function of \( \gamma \) is given by

\[
E[q(w, \gamma)] = E \left[ \frac{(y_{11} - y_{12} + \gamma(y_3 - y_2))^2}{2(1 + \gamma^2)} \right]
\]

Substituting for \( y \) with the data generating process yields

\[
E[q(w, \gamma)] = E \left[ \frac{(\alpha_{10} + \gamma_0 \alpha_{20} + \epsilon_{11} - (\alpha_{10} + \gamma_0 \alpha_{30} + \epsilon_{12}) + \gamma(\alpha_{30} + \gamma_0 \alpha_{10} + \epsilon_3 - (\alpha_{20} + \gamma_0 \alpha_{10} + \epsilon_2)))^2}{2(1 + \gamma^2)} \right]
\]

Canceling the appropriate terms and combining like terms in the numerator leaves

\[
E[q(w, \gamma)] = E \left[ \frac{((\gamma_0 - \gamma)(\alpha_{20} - \alpha_{30}) + (\epsilon_{11} - \epsilon_{12}) + \gamma(\epsilon_3 - \epsilon_2))^2}{2(1 + \gamma^2)} \right]
\]

Opening up the square term leaves

\[
E[q(w, \gamma)] = E \left[ \frac{1}{2(1 + \gamma^2)} \left( (\gamma_0 - \gamma)^2(\alpha_{20} - \alpha_{30})^2 + (\epsilon_{11} - \epsilon_{12})^2 + \gamma^2(\epsilon_3 - \epsilon_2)^2 + 2(\gamma_0 - \gamma)(\alpha_{20} - \alpha_{30})(\epsilon_{11} - \epsilon_{12}) + 2\gamma(\epsilon_3 - \epsilon_2)(\epsilon_{11} - \epsilon_{12})(\epsilon_3 - \epsilon_2) \right) \right]
\]

By assumptions 1 and 2, the final 3 terms in the numerator all have expectation 0. Similarly, any covariance terms associated with the first three terms in the numerator will have expectation 0. The final simplified expression is given by

\[
E[q(w, \gamma)] = \frac{(\gamma_0 - \gamma)^2 E[(\alpha_{20} - \alpha_{30})^2] + E[\epsilon_{11}^2] + E[\epsilon_{12}^2] + \gamma^2(E[\epsilon_3^2] + E[\epsilon_2^2])}{2(1 + \gamma^2)}
\]

which we can re-write in the following manner

\[
E[q(w, \gamma)] = \frac{(\gamma_0 - \gamma)^2 E[(\alpha_{20} - \alpha_{30})^2] + E[\epsilon_{11}^2] + E[\epsilon_{12}^2] + \gamma^2(E[\epsilon_3^2] + E[\epsilon_2^2])}{2(1 + \gamma^2)}
\]
Note that by assumption 5, $E[\epsilon_{11}^2] = E[\epsilon_2^2]$ and $E[\epsilon_{12}^2] = E[\epsilon_3^2]$ implying that we can rewrite the above equation as

$$E[q(w, \gamma)] = \frac{(\gamma_o - \gamma)^2 E[(\alpha_{2o} - \alpha_{3o})^2]}{2(1 + \gamma^2)} + \frac{(E[\epsilon_{11}^2] + E[\epsilon_{12}^2])}{2}$$

The first term in the above expression is strictly greater than 0 for all $\gamma \neq \gamma_o$ and the second term does not depend upon $\gamma$. As a result, $E[q(w, \gamma_o)] < E[q(w, \gamma)]$ for all $\gamma \in \Gamma$ when $\gamma \neq \gamma_o$. QED.

**Proof of Lemma 4**

Uniform convergence, requires that

$$\max_{\gamma \in \Gamma} \left| \frac{1}{N} \sum_{n=1}^{N} q(w_n, \gamma) - E[q(w, \gamma)] \right| \overset{P}{\to} 0$$

Theorem 12.1 in Wooldridge states four conditions that the data and $q$ must satisfy in order for the above condition to hold.

1. $\Gamma$ is compact
   This condition is satisfied by assumption 6.

2. For each $\gamma \in \Gamma$, $q(\cdot, \gamma)$ is Borel measurable on $\mathcal{W}$
   Since $q(\cdot, \gamma)$ is a continuous function of $w$, it is also Borel measurable.

3. For each $w \in \mathcal{W}$, $q(w, \cdot)$ is continuous on $\Gamma$
   Our concentrated objective function is continuous in $\gamma$.

4. $|q(w, \gamma)| \leq b(w)$ for all $\gamma \in \Gamma$, where $b$ is a nonnegative function on $\mathcal{W}$ such that $E[b(w)] < \infty$
   Note that $q(w, \gamma)$ is always positive so we can ignore the absolute value. We derive a bounding function $b(w)$ in the following manner

$$q(w, \gamma) = \frac{(y_{11} - y_{12} + \gamma(y_3 - y_2))^2}{2(1 + \gamma^2)}$$

$$= \frac{(y_{11} - y_{12})^2 + \gamma^2(y_3 - y_2)^2 + 2\gamma(y_3 - y_2)(y_{11} - y_{12})}{2(1 + \gamma^2)}$$

$$\leq \frac{2(y_{11} - y_{12})^2}{2(1 + \gamma^2)} + \frac{2\gamma^2(y_3 - y_2)^2}{2(1 + \gamma^2)}$$

$$\leq (y_{11} - y_{12})^2 + (y_3 - y_2)^2$$

32
Where the third line follows from the triangle inequality. Our bounding function is then
\[ b(w) = (y_{11} - y_{12})^2 + (y_{3} - y_{2})^2 \]
where we have shown that \( b(w) \geq q(w, \gamma) \) for all \( y \).

We now show that \( E[b(w)] < \infty \), completing the proof. Note that \( E[b(w)] \) is given by
\[ E[b(w)] = E \left[ (y_{11} - y_{12})^2 + (y_{3} - y_{2})^2 \right] \]

Using the triangle inequality, we can re-write the above expression as
\[ E[b(w)] \leq E[2y_{11}^2 + 2y_{12}^2 + 2y_{3}^2 + 2y_{2}^2] \]
\[ \leq 2(E[y_{11}^2] + E[y_{12}^2] + E[y_{3}^2] + E[y_{2}^2]) \]

Next we substitute in for \( y \) using the data generating process. Consider \( E[y_{11}^2] \), which is given by
\[ E[y_{11}^2] = E[(\alpha_{1o} + \gamma o \alpha_{2o} + \epsilon_{11})^2] \]

Applying the triangle inequality again yields
\[ E[y_{11}^2] \leq 3(E[\alpha_{1o}^2] + \gamma_o^2 E[\alpha_{2o}^2] + E[\epsilon_{11}^2]) \]

Assumptions 3 and 4 ensure that all of the terms on the right hand side of the inequality in the above equation are finite. Thus, \( E[y_{11}^2] \) is finite. By a similar argument, it can be shown that all the terms in \( E[b(w)] \) are finite.

QED

Proof of Lemma 5

Theorem 12.3 in Wooldridge(2002) states six conditions that must hold in order for \( \hat{\gamma} \) to be distributed asymptotically normal.

Many of these conditions involve the first and second derivatives of \( q(w, \gamma) \). We begin our proof of asymptotic normality by deriving the first and second derivatives of the objective function.

The first derivative of the objective function, or the score, is given by
\[ s(w, \gamma) = \frac{1}{4(1 + \gamma^2)^2} \left[ 2(1 + \gamma^2) \left( 2(y_3 - y_2)(y_{11} - y_{12} + \gamma(y_3 - y_2)) \right) - 4\gamma \left( (y_{11} - y_{12} + \gamma(y_3 - y_2))^2 \right) \right] \]
Expanding the square and grouping on the $\gamma$ terms yields

$$s(w, \gamma) = \frac{1}{(1 + \gamma^2)^2} \left[ (1 - \gamma^2)(y_{11} - y_{12})(y_3 - y_2) + \gamma \left( - (y_{11} - y_{12})^2 + (y_3 - y_2)^2 \right) \right]$$

The Hessian of the objective function is simply the derivative of the score, $\frac{\partial s(y, \gamma)}{\partial \gamma}$, and is given below

$$H(w, \gamma) = \frac{1}{(1 + \gamma^2)^3} \left( (1 - 3\gamma^2)((y_3 - y_2)^2 - (y_{11} - y_{12})^2) - 2\gamma(3 - \gamma^2)(y_3 - y_2)(y_{11} - y_{12}) \right)$$

Factoring out a $(1 + \gamma^2)$ and combining like terms greatly simplifies the above expression, leaving

$$H(w, \gamma) = \frac{1}{(1 + \gamma^2)^3} \left( (1 - 3\gamma^2)((y_3 - y_2)^2 - (y_{11} - y_{12})^2) - 2\gamma(3 - \gamma^2)(y_3 - y_2)(y_{11} - y_{12}) \right)$$

We now show that the six conditions of Theorem 12.3 in Wooldridge(2002) are satisfied. We will refer to the above formulations of the score and Hessian throughout.

1. $\gamma_o$ must be in the interior of $\Gamma$
   This condition is satisfied by assumption 6.

2. $s(w, \cdot)$ is continuously differentiable on the interior of $\Gamma$ for all $w \in W$
   Since $H(w, \gamma)$ is continuous in $\gamma$, $s(w, \cdot)$ is continuously differentiable.

3. Each element of $H(w, \gamma)$ is bounded in absolute value by a function $b(w)$ where $E[b(w)] < \infty$

We derive a bounding function $b(w)$ in the following manner

$$H(w, \gamma) = \frac{(1 - 3\gamma^2)((y_3 - y_2)^2 - (y_{11} - y_{12})^2) - 2\gamma(3 - \gamma^2)(y_3 - y_2)(y_{11} - y_{12})}{(1 + \gamma^2)^3}$$

$$|H(w, \gamma)| \leq \left| (1 - 3\gamma^2)((y_3 - y_2)^2 - (y_{11} - y_{12})^2) - 2\gamma(3 - \gamma^2)(y_3 - y_2)(y_{11} - y_{12}) \right|$$

$$|H(w, \gamma)| \leq \left| (1 - 3\gamma^2)((y_3 - y_2)^2 - (y_{11} - y_{12})^2) \right| + \left| 2\gamma(3 - \gamma^2)(y_3 - y_2)(y_{11} - y_{12}) \right|$$

$$|H(w, \gamma)| \leq (1 + 3\gamma^2)((y_3 - y_2)^2 + (y_{11} - y_{12})^2) + (3 + \gamma^2) |2\gamma(y_3 - y_2)(y_{11} - y_{12})|$$

$$|H(w, \gamma)| \leq (1 + 3\gamma^2)((y_3 - y_2)^2 + (y_{11} - y_{12})^2) + (3 + \gamma^2) |\gamma((y_3 - y_2)^2 + (y_{11} - y_{12})^2)|$$

$$|H(w, \gamma)| \leq (1 + 3\gamma^2)((y_3 - y_2)^2 + (y_{11} - y_{12})^2) + (3 + \gamma^2) |\gamma((y_3 - y_2)^2 + (y_{11} - y_{12})^2)|$$
where the second to last line utilizes the fact that \((y_3 - y_2)^2 + (y_{11} - y_{12})^2 > 2(y_3 - y_2)(y_{11} - y_{12})\) as \((y_3 - y_2) - ((y_{11} - y_{12}))^2 > 0\). Let \(\gamma\) denote the largest and smallest elements of the set \(\Gamma\). The \(\gamma\) that maximizes the right hand side is given by \(\gamma^* = \max\{\gamma, -\gamma\} < \infty\). Our bounding function is then

\[
b(w) = (1 + 3\gamma^*2)\left((y_3 - y_2)^2 + (y_{11} - y_{12})^2\right) + \gamma^*(3 + \gamma^2)\left((y_3 - y_2)^2 + (y_{11} - y_{12})^2\right)
\]

where we have shown that \(b(w) \geq H(w, \gamma)\) for all \(w\). Notice that the absolute value of \(\gamma\) is no longer necessary since by definition \(\gamma^*\) is always positive.

We now show that \(E[b(w)] < \infty\), completing the proof.

\[
E[b(w)] = (1 + \gamma^*(\gamma^*2 + 3\gamma^* + 3))E\left[(y_3 - y_2)^2 + (y_{11} - y_{12})^2\right]
\]

When deriving the bounding function for \(q(w, \gamma)\), we showed that \(E[(y_3 - y_2)^2 + (y_{11} - y_{12})^2] < \infty\). Since \(\gamma^*\) is also finite, \(E[b(w)] < \infty\).

4. \(A_o \equiv E[H(w, \gamma_o)]\) is positive definite

We first note that we can interchange the expectations and the partial derivatives:

\[
E[H(w, \gamma)] = \partial^2 E[q(w, \gamma)] / \partial \gamma^2.
\]

From Lemma 3, we know that \(E[q(w, \gamma)]\) can be written as

\[
E[q(w, \gamma)] = \frac{\gamma - \gamma_o)^2E[(\alpha_{2o} - \alpha_{3o})^2]}{2(1 + \gamma^2)} + \left(E[\epsilon_{11}^2] + E[\epsilon_{12}^2]\right) / 2.
\]

Note that \(\gamma\) affects two terms: \((\gamma - \gamma_o)^2\) and the denominator. However, because we are going to evaluate the expected Hessian at \(\gamma_o\), we only need the second derivative of the first term, \((\gamma - \gamma_o)^2\). All of the other partial derivatives will either be multiplied by \((\gamma - \gamma_o)^2\) or \((\gamma - \gamma_o)\), both of which are zero when \(\gamma = \gamma_o\). The second derivative of \((\gamma - \gamma_o)^2\) with respect to \(\gamma\) is positive. This second derivative is then multiplied by the expectation of a squared object in the numerator and divided by the sum of squared objects in the denominator. Thus, the expectation of the Hessian evaluated at \(\gamma_o\) is strictly positive.

5. \(E[s(w, \gamma_o)] = 0\)

Note that \(E[s(w, \gamma)] = \partial E[q(w, \gamma)] / \partial \gamma\). Differentiating \(E[q(w, \gamma)]\) with respect to \(\gamma\)
leaves terms that are multiplied by \((\gamma - \gamma_0)\) or by \((\gamma - \gamma_0)^2\), implying that if we evaluate the derivative at \(\gamma = \gamma_0\) then the expected score is zero.

6. Each element of \(s(w, \gamma_0)\) has finite second moment.

Given that the score has only one element, this condition boils down to \(E[s(w, \gamma_0)^2] < \infty\).

To show this we square the score function, repeatedly apply the triangle equality, and evaluate the expected value at the true \(\gamma\).

\[
E[s(w, \gamma_0)^2] = E\left(\frac{1}{(1 + \gamma_0^2)^4}\left[(1 - \gamma_0^2)(y_{11} - y_{12})(y_3 - y_2) + \gamma_0(-(y_{11} - y_{12})^2 + (y_3 - y_2)^2)\right]^2\right)
\]

Repeatedly applying the triangle inequality yields

\[
E[s(w, \gamma_0)^2] \leq E\left(\frac{1}{(1 + \gamma_0^2)^4}\left[2(1 - \gamma_0^2)^2(y_{11} - y_{12})^2(y_3 - y_2)^2 + 2\gamma_0^2(-(y_{11} - y_{12})^2 + (y_3 - y_2)^2)\right]\right)
\]

\[
\leq E\left(\frac{4}{(1 + \gamma_0^2)^4}\left[2(1 - \gamma_0^2)^2(y_{11}^2 + y_{12}^2)(y_3^2 + y_2^2) + \gamma_0^2((y_{11} - y_{12})^4 + (y_3 - y_2)^4)\right]\right)
\]

\[
\leq E\left(\frac{8}{(1 + \gamma_0^2)^4}\left[1 - \gamma_0^2)(y_{11}^4 + y_{12}^4 + y_3^4 + y_2^4) + 4\gamma_0^2((y_{11}^2 + y_{12}^2)^2 + (y_3^2 + y_2^2)^2)\right]\right)
\]

\[
\leq E\left(\frac{8}{(1 + \gamma_0^2)^4}\left[y_{11}^4 + y_{12}^4 + y_3^4 + y_2^4\right]\right)
\]

\[
\leq \frac{8}{(1 + \gamma_0^2)^2}E\left(y_{11}^4 + y_{12}^4 + y_3^4 + y_2^4\right)
\]

Now we substitute for \(y\) with the DGP. Consider \(E[y_{11}^4]\) which is given by

\[
E[y_{11}^4] = E[(\alpha_{1o} + \gamma_0\alpha_{2o} + \epsilon_{11})^4]
\]

Repeatedly applying the triangle inequality yields

\[
E[y_{11}^4] \leq 9E[\alpha_{1o}^2 + \gamma_0^2\alpha_{2o}^2 + \epsilon_{11}^2]\]

\[
\leq 27(E[\alpha_{1o}^4] + \gamma_0^4E[\alpha_{2o}^4] + E[\epsilon_{11}^4])
\]

Assumptions 3 and 4 ensure that all of the terms on the right hand side of the inequality in the above equation are finite. Thus, \(E[y_{11}^4]\) is finite. By a similar argument, it can be shown that all the terms in the expectation of the squared score are finite.

QED
A.2 Proof of Theorem 2

Proof. The first order condition for $\alpha_i$ can be written as

$$0 = \sum_{t=1}^{T} \left( Y_{itn} - \alpha_i - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_j \right) + \sum_{t=1}^{T} \sum_{j \in M_{tn \sim i}} \frac{\gamma}{M_{tn}} \left( Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k \in M_{tn \sim j \sim i}} \alpha_k \right)$$

(18)

Solving for $\alpha_i$ and collecting terms, we have

$$\alpha_i = \frac{\sum_{t=1}^{T} \left[ Y_{itn} - \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \alpha_j + \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \left( Y_{jtn} - \alpha_j - \frac{\gamma}{M_{tn}} \sum_{k \in M_{tn \sim j \sim i}} \alpha_k \right) \right]}{T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}}}$$

(19)

Now we stack these equations, such that the $N \times 1$ vector of $\alpha$'s runs down the left hand side of the stack. To apply our iterative method, we make a first guess at this vector, and then use this guess to generate OLS-derived estimates of the other parameters appearing in the model. Once obtained, these estimates are then plugged into the right-hand side of these equations and we update our guess of the $\alpha$ vector. Let the first of any two consecutive guesses of the $\alpha$ vector be called simply $\alpha$, and let the second (updated) guess be called $\alpha'$. We would like to show that our mapping, call it $f$, from $\alpha \rightarrow \alpha'$ is a contraction mapping. That is, $\rho(f(\alpha), f(\alpha')) < \beta \rho(\alpha, \alpha')$ for some $\beta < 1$ and where $\rho$ is a valid distance function. Using a Euclidean distance function for $\rho$, our task is then to show under what conditions, for a chosen $\beta < 1$, the following

$$\left( \sum_{i=1}^{N} \left( \frac{\sum_{t=1}^{T} \left[ \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \tilde{\alpha}_j + \frac{\gamma}{M_{tn}} \sum_{j \in M_{tn \sim i}} \left( \tilde{\alpha}_j + \frac{\gamma}{M_{tn}} \sum_{k \in M_{tn \sim j \sim i}} \tilde{\alpha}_k \right) \right]}{T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}}} \right)^2 \right)^{1/2} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{1/2}$$

(20)

will be less than

$$\beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{1/2}$$

(21)

where $\tilde{\alpha} = \alpha - \alpha'$ and $N$ again refers to the total student population. Factoring out the $\alpha$'s, this requirement can be rewritten as

$$\left( \sum_{i=1}^{N} \left( \frac{\sum_{t=1}^{T} \left[ \frac{2\gamma}{M_{tn}} + \frac{\gamma^2(M_{tn} - 1)}{M_{tn}^2} \sum_{j \in M_{tn \sim i}} \tilde{\alpha}_j \right]}{T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}}} \right)^2 \right)^{1/2} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{1/2}$$

(22)
Expanding the inner square on the left hand side of the inequality and repeatedly applying the triangle inequality yields

\[
\left( \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} T \left[ \left( \frac{2\gamma}{M_{tn}} + \frac{\gamma^2 (M_{tn} - 1)}{M_{tn}^2} \right)^2 \left( \sum_{j \in M_{tn-1}} \tilde{\alpha}_j \right)^2 \right]}{\left( T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}} \right)^2} \right)^{\frac{1}{2}} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{\frac{1}{2}} \tag{23}
\]

Expanding the square on the sum of the \(\tilde{\alpha}_j\)'s and applying the triangle inequality leaves

\[
\left( \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} T \left[ \left( \frac{2\gamma}{M_{tn}} + \frac{\gamma^2 (M_{tn} - 1)}{M_{tn}^2} \right)^2 M_{tn} \sum_{j \in M_{tn-1}} \tilde{\alpha}_j^2 \right]}{\left( T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}} \right)^2} \right)^{\frac{1}{2}} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{\frac{1}{2}} \tag{24}
\]

Inside the square brackets of equation (25) there are no \(\tilde{\alpha}_i\) since this term reflects the purged first order condition from individual \(i\). However, \(\tilde{\alpha}_i\) will be present in the first order condition from all of \(i\)'s classmates over time. Because the \(M_{tn}\) in the denominator reflects the peer group sizes experienced by individual \(i\) over time, all the terms on the left hand side of the inequality containing an \(\tilde{\alpha}_i\) will have different denominators. Substituting \(\overline{M}\) for \(M_{tn}\) in the denominator will ensure a common denominator across the terms containing an \(\tilde{\alpha}_i\).

\[
\left( \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} T \left[ \left( \frac{2\gamma}{\overline{M}} + \frac{\gamma^2 (M_{tn} - 1)}{\overline{M}^2} \right)^2 M_{tn} \sum_{j \in M_{tn-1}} \tilde{\alpha}_j^2 \right]}{\left( T + \sum_{t=1}^{T} \frac{\gamma^2}{\overline{M}} \right)^2} \right)^{\frac{1}{2}} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{\frac{1}{2}} \tag{25}
\]

This substitution is valid since it shrinks the denominator for every term on the left hand side of the inequality, making it less likely to hold. Now we can easily collect all the terms containing an \(\tilde{\alpha}_i\), yielding

\[
\left( \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} T \left[ \left( \frac{2\gamma}{M_{tn}} + \frac{\gamma^2 (M_{tn} - 1)}{M_{tn}^2} \right)^2 M_{tn}^2 \tilde{\alpha}_i^2 \right]}{\left( T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}} \right)^2} \right)^{\frac{1}{2}} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{\frac{1}{2}} \tag{26}
\]

The additional \(M_{tn}\) term in the numerator comes from the fact that \(\tilde{\alpha}_i\) will show up once for each the \(M_{tn}\) peers at time \(t\). Bringing the \(M_{tn}^2\) inside the parentheses in the numerator yields

\[
\left( \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} T \left[ \left( \frac{2\gamma}{M_{tn}} + \frac{\gamma^2 (M_{tn} - 1)}{M_{tn}^2} \right)^2 \right]}{\left( T + \sum_{t=1}^{T} \frac{\gamma^2}{M_{tn}} \right)^2} \right)^{\frac{1}{2}} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^{\frac{1}{2}} \tag{27}
\]
Notice that we can again substitute for $M_{tn}$ with $\overline{M}$ since this will strictly increase the coefficient on $\tilde{\alpha}_i$, making it less likely that the inequality is satisfied. Making this substitution and canceling the $T^2$ terms leaves

$$\left( \sum_{i=1}^{N} \left( \frac{\left( 2\gamma + \gamma^2 - \frac{\gamma^2}{\overline{M}} \right)^2}{1 + \frac{\gamma^2}{\overline{M}}} \right) \tilde{\alpha}_i^2 \right)^\frac{1}{2} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^\frac{1}{2}$$  \hspace{1cm} (28)$$

which can be re-written as

$$\frac{2\gamma + \gamma^2 - \frac{\gamma^2}{\overline{M}}}{1 + \frac{\gamma^2}{\overline{M}}} \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^\frac{1}{2} < \beta \left( \sum_{i=1}^{N} \tilde{\alpha}_i^2 \right)^\frac{1}{2}$$  \hspace{1cm} (29)$$

As long as the $\gamma$’s are such that (29) is satisfied, we have a contraction mapping. The denominator of the leading term is strictly greater than one, implying that if the numerator is strictly less than one than the contraction holds for some $\beta < 1$. If $2\gamma + \gamma^2 < 1$ the numerator will be strictly less than one, which is true for $\gamma \leq .20$ \textbf{QED}

\textbf{A.3 Proof of Theorem 3}

\textit{Proof.} In the case of accumulation, we face the following least squares problem

$$\min_{\alpha^*,\gamma} \sum_{i=1}^{N} \left[ (Y_{1in} - \alpha^*_{12})^2 + \left( Y_{2in} - \alpha^*_{2} - \frac{\gamma}{M_{2n}} \sum_{j \in M_{2n} \sim i} \alpha^*_j \right)^2 \right]$$  \hspace{1cm} (30)$$

where $\alpha^*_t$ is defined in the text as the initial ability plus the accumulated peer effects through period $t$. Similar to the proof of Theorem 1, we illustrate the proof assuming that students are grouped with at most one other student at any point in time and that students are observed for at most two time periods. In addition, within each class there is only one student that is observed for two periods.\textsuperscript{21}

Consider the set of students that are observed for two time periods. Each of these students has one peer in period one and one peer in period two. Denote a student block as one student observed for two periods plus his two peers. There are then $N$ blocks of students, one block

\textsuperscript{20}An identical restriction on $\gamma$ is required in the case of an unbalanced panel. To derive this simply define $\rho$ as a weighted Euclidean distance where the individual weights are given by the number of observations for student $i$, $T_i$.

\textsuperscript{21}See the proof of Theorem 1 for further discussion of these simplifications.
for each student observed twice, with three students in each block. Denote the first student in each block as the student who is observed twice where \( \alpha_{1n} \) is the initial individual effect. The initial individual effect for the first classmate in block \( n \) is \( \alpha_{2n} \), while the initial individual effect for the second classmate in block \( n \) is \( \alpha_{3n} \).

Define \( \alpha_{1n}^* = \alpha_{1n} + \gamma \alpha_{2n} \) and \( \alpha_{2n}^* = \alpha_{2n} + \gamma \alpha_{1n} \). The optimization problem is then

\[
\min_{\alpha, \gamma} \frac{1}{N} \sum_{n=1}^{N} \left( (y_{11n} - \alpha_{1n}^*)^2 + (y_{12n} - \alpha_{1n}^* - \gamma \alpha_{3n})^2 + (y_{2n} - \alpha_{2n}^*)^2 + (y_{3n} - \alpha_{3n} - \gamma \alpha_{1n}^*)^2 \right)
\]

Within each block there are four terms, two residuals for the student observed twice, and a residual for the peer in each period. Our proof then consists of the following five lemmas, each of which is proven later in this appendix.

We first show that the \( \alpha \)'s can be written as closed form expressions of \( \gamma \) and the data.

**Lemma 1**

The vector of unobserved student abilities, \( \alpha \), can be concentrated out of the least squares problem and written strictly as a function of \( \gamma \) and \( y \). Ability for the student in block \( n \) observed in both periods is given by

\[
\alpha_{1n}^* = \frac{(1 + \gamma^2)y_{11n} + (1 - \gamma^2)y_{12n} - \gamma(1 - \gamma^2)y_{31n}}{2 - \gamma^2 + \gamma^4}
\]

while the abilities for the peers in block \( n \) are given by

\[
\alpha_{2n}^* = y_{2n}
\]

and

\[
\alpha_{3n}^* = \frac{-2\gamma y_{11n} + \gamma^3 y_{12n} + (2 - \gamma^2)y_{31n}}{2 - \gamma^2 + \gamma^4}
\]

We then show the form of the minimization problem when the \( \alpha \)'s are concentrated out.

**Lemma 2**

Concentrating the \( \alpha \)'s out of the original least squares problem results in an optimization problem over \( \gamma \) that takes the following form

\[
\min_{\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{(\gamma^2 - 1)y_{11n} + y_{12n} - \gamma y_{3n})^2}{2 - \gamma^2 + \gamma^4}
\]
Our nonlinear least squares problem now has only one parameter, \( \gamma \). We are now in a position to investigate the properties of our estimator of \( \gamma_o \). For ease of notation, define \( q(w, \gamma) \) as

\[
q(w, \gamma) = \frac{(\gamma - 1)y_{11} + y_{12} - \gamma y_3)^2}{2 - \gamma^2 + \gamma^4}
\]

where \( w \equiv y \). We let \( \mathcal{W} \) denote the subset of \( \mathbb{R}^4 \) representing the possible values of \( w \). Our key result is then Lemma 3, which establishes identification.

**Lemma 3**

\[ E[q(w, \gamma_o)] < E[q(w, \gamma)], \quad \forall \gamma \in \Gamma, \gamma \neq \gamma_o \]

Theorem 12.2 of Wooldridge (2002) establishes that sufficient conditions for consistency are identification and uniform convergence. Having already established identification, Lemma 4 shows uniform convergence.

**Lemma 4**

\[
\max_{\gamma \in \Gamma} \left| \frac{1}{N} \sum_{n=1}^{N} q(w_n, \gamma) - E[q(w, \gamma)] \right| \xrightarrow{p} 0
\]

Consistency then follows from Theorem 12.2 of Wooldridge: \( \hat{\gamma} \xrightarrow{p} \gamma_o \).

Finally, we establish asymptotic normality of \( \hat{\gamma} \). Denote \( s(w, \gamma_o) \) and \( H(w, \gamma_o) \) as the first and second derivative of \( q(w, \gamma) \) evaluated at \( \gamma_o \). Then, Lemma 5 completes the proof.

**Lemma 5**

\[
\sqrt{N} (\hat{\gamma} - \gamma_o) \overset{d}{\to} N(0, A_o^{-1}B_oA_o^{-1})
\]

where

\[
A_o \equiv E[H(w, \gamma_o)]
\]

and

\[
B_o \equiv E[s(w, \gamma_o)^2] = Var[s(w, \gamma_o)]
\]

QED.

**Proof of Lemma 1**

Our objective is to show that the system of equations obtained by differentiating Equation (A.3) with respect to \( \alpha \) can be expressed as a series of equations in terms of \( \gamma \) and \( y \), and
that these expressions are as given in Lemma 1. Again, conditional on \( \gamma \), the estimates of individual effects in one block will not affect the estimates of the individual effects in another block. Thus, we can work with the system of first-order conditions within one block and then generalize the results to the full system of equations. The first-order condition for \( \alpha_{1n}^* \) (student in each block who is observed in both time periods) is given by

\[
0 = \frac{-2}{N} \left[ (y_{11n} - \alpha_{1n}^*) + (y_{12n} - \alpha_{1n}^* - \gamma \alpha_{3n}) + \gamma (y_{3n} - \alpha_{3n} - \gamma \alpha_{1n}^*) \right]
\]

while the first-order condition for \( \alpha_{2n}^* \) and \( \alpha_{3n} \) are respectively given by

\[
0 = \frac{-2}{N} (y_{2n} - \alpha_{2n}^*)
\]

and

\[
0 = \frac{-2}{N} \left[ (y_{3n} - \alpha_{3n} - \gamma \alpha_{1n}^*) + \gamma (y_{12n} - \alpha_{1n}^* - \gamma \alpha_{3n}) \right]
\]

Within each block, this yields a system of 3 equations and 3 unknown abilities. The solution for \( \alpha_{2n}^* \) is simply given by \( y_{2n} \). The first order condition \( \alpha_{3n} \) can be re-arranged such that

\[
\alpha_{3n} = \frac{y_{3n} + \gamma y_{12n} - 2 \gamma \alpha_{1n}}{1 + \gamma^2}
\]

and the first order condition for \( \alpha_{1n}^* \) can be re-written as

\[
\alpha_{1n}^* = \frac{y_{11n} + y_{12n} + \gamma y_{3n} - 2 \gamma \alpha_{3n}}{2 + \gamma^2}
\]

Substituting the equation for \( \alpha_{3n} \) into the equation for \( \alpha_{1n}^* \) and combining terms yields

\[
\alpha_{1n}^* = \frac{(1 + \gamma^2)y_{11n} + (1 - \gamma^2)y_{12n} - \gamma(1 - \gamma^2)y_{3n} + 4\gamma^2 \alpha_{1n}^*}{(2 + \gamma^2)(1 + \gamma^2)}
\]

Moving all terms containing an \( \alpha_{1n} \) to the left hand side of the equation and simplifying results in

\[
\alpha_{1n}^* \left( \frac{2 - \gamma^2 + \gamma^4}{(2 + \gamma^2)(1 + \gamma^2)} \right) = \frac{(1 + \gamma^2)y_{11n} + (1 - \gamma^2)y_{12n} - \gamma(1 - \gamma^2)y_{3n}}{(2 + \gamma^2)(1 + \gamma^2)}
\]

Multiplying both sides of the equation by \( \frac{(2 + \gamma^2)(1 + \gamma^2)}{2 - \gamma^2 + \gamma^4} \) yields the desired result,

\[
\alpha_{1n}^* = \frac{(1 + \gamma^2)y_{11n} + (1 - \gamma^2)y_{12n} - \gamma(1 - \gamma^2)y_{3n}}{2 - \gamma^2 + \gamma^4}
\]

Substituting the solution for \( \alpha_{1n}^* \) into the first order condition for \( \alpha_{3n} \) gives

\[
\alpha_{3n} = \frac{y_{3n} + \gamma y_{12n}}{1 + \gamma^2} - \frac{2 \gamma}{1 + \gamma^2} \left[ \frac{(1 + \gamma^2)y_{11n} + (1 - \gamma^2)y_{12n} - \gamma(1 - \gamma^2)y_{3n}}{2 - \gamma^2 + \gamma^4} \right]
\]
Finding a common denominator and combining like terms yields

$$\alpha_{3n} = \frac{-2\gamma(1 + \gamma^2)y_{11n} + \gamma^3(1 + \gamma^2)y_{12n} + (2 - \gamma^2)(1 + \gamma^2)y_{3n}}{(1 + \gamma^2)(2 - \gamma^2 + \gamma^4)}$$

All the \((1 + \gamma^2)\) terms cancel, leaving the desired solution,

$$\alpha_{3n} = \frac{-2\gamma y_{11n} + \gamma^3 y_{12n} + (2 - \gamma^2)y_{3n}}{2 - \gamma^2 + \gamma^4}$$

QED

Proof of Lemma 2

Lemma 1 provides a solution for \(\alpha\) strictly as a function of \(y\) and \(\gamma\). We can substitute this solution back into the original optimization problem to derive the result in Lemma 2.

Consider minimizing the sum of squared residuals within a particular block \(n\). There are four residuals within each block, two for the student observed twice, and one each for the corresponding peer. We begin by simplifying the residual for the first observation of the student observed twice, which is given by the expression below

$$e_{11n} = y_{11n} - \alpha^*_1 n$$

Substituting for \(\alpha^*_1 n\) with the results from Lemma 1 and finding a common denominator results in

$$e_{11n} = \frac{(2 - \gamma^2 + \gamma^4)y_{11n} - (1 + \gamma^2)y_{11n} - (1 - \gamma^2)y_{12n} + \gamma(1 - \gamma^2)y_{3n}}{2 - \gamma^2 + \gamma^4}$$

Combining like terms in the numerator yields

$$e_{11n} = \frac{(1 - \gamma^2)^2 y_{11n} - (1 - \gamma^2)y_{12n} + \gamma(1 - \gamma^2)y_{3n}}{2 - \gamma^2 + \gamma^4}$$

Next we want to find an expression for \(e_{12n}\), the residual for individual one in period two, as a function of the data and \(\gamma\). Recall that

$$e_{12n} = y_{12n} - \alpha^*_1 n - \gamma\alpha_{3n}$$

Substituting for \(\alpha^*_1 n\) and \(\alpha_{3n}\) and finding a common denominator yields

$$e_{12n} = \frac{(2 - \gamma^2 + \gamma^4)y_{12n} - (1 + \gamma^2)y_{11n} - (1 - \gamma^2)y_{12n} + \gamma(1 - \gamma^2)y_{3n} + 2\gamma^2 y_{11n} - \gamma^4 y_{12n} - \gamma(2 - \gamma^2)y_{3n}}{2 - \gamma^2 + \gamma^4}$$
Combining like terms yields

\[ e_{12n} = \frac{(\gamma^2 - 1)y_{11n} + y_{12n} - \gamma y_{3n}}{2 - \gamma^2 + \gamma^4} \]

The residual for individual two is zero since \( \alpha^*_{2n} = y_{2n} \) and

\[ e_{2n} = y_{2n} - \alpha^*_{2n} \]

The residual for individual three is given by

\[ e_{3n} = y_{3n} - \alpha_{3n} - \gamma \alpha^*_{1n} \]

Substituting for \( \alpha_{3n} \) and \( \alpha^*_{1n} \) from Lemma 1 and finding a common denominator leaves

\[ e_{3n} = \frac{(2 - \gamma^2 + \gamma^4)y_{3n} + 2\gamma y_{11n} - \gamma^3 y_{12n} - (2 - \gamma^2)y_{3n} - \gamma(1 + \gamma^2)y_{11n} - \gamma(1 - \gamma^2)y_{12n} + \gamma^2(1 - \gamma^2)y_{3n}}{2 - \gamma^2 + \gamma^4} \]

Combining like terms yields

\[ e_{3n} = \frac{\gamma(1 - \gamma^2)y_{11n} - \gamma y_{12n} + \gamma^2 y_{3n}}{2 - \gamma^2 + \gamma^4} \]

Now we return to the optimization problem. The original optimization problem written as a function of the residuals in each block \( n \) takes the following form

\[
\min_{\alpha, \gamma} \frac{1}{N} \sum_{n=1}^{N} \left( e_{11n}^2 + e_{12n}^2 + e_{2n}^2 + e_{3n}^2 \right)
\]

We can substitute into the above formulation for each residual using the formulas previously derived. However, a cursory glance at the formulas for \( e_{11n} \), \( e_{12n} \), and \( e_{3n} \) reveals that

\[ e_{11n} = -(1 - \gamma^2) e_{12n} \]

\[ e_{3n} = -\gamma e_{12n} \]

Using these relationships along with the fact that \( e_{2n} = 0 \), we can re-write the least squares problem as

\[
\min_{\alpha, \gamma} \frac{1}{N} \sum_{n=1}^{N} \left( (2 - \gamma^2 + \gamma^4) e_{12n}^2 \right)
\]

Substituting in with our solution for \( e_{12n} \) yields

\[
\min_{\gamma} \frac{1}{N} \sum_{n=1}^{N} \left( (2 - \gamma^2 + \gamma^4) \frac{(\gamma^2 - 1)y_{11n} + y_{12n} - \gamma y_{3n})^2}{(2 - \gamma^2 + \gamma^4)^2} \right)
\]

44
Canceling terms results in the following optimization problem

\[
\min_{\gamma} \frac{1}{N} \sum_{n=1}^{N} \frac{(\gamma^2 - 1)y_{11n} + y_{12n} - \gamma y_{3n})^2}{2 - \gamma^2 + \gamma^4}
\]

QED

Proof of Lemma 3

The population objective function as a function of \(\gamma\) is given by

\[
E[q(w, \gamma)] = E \left[ \frac{((\gamma^2 - 1)y_{11} + y_{12} - \gamma y_{3})^2}{2 - \gamma^2 + \gamma^4} \right]
\]

Substituting for \(y\) with the data generating process, canceling the appropriate terms and combining like terms in the numerator leaves

\[
E[q(w, \gamma)] = E \left[ \frac{(\gamma^2 - \gamma \gamma_o)\alpha_{1o} + (\gamma^2 \gamma_o - \gamma \gamma_o^2)\alpha_{2o} + (\gamma_o - \gamma)\alpha_{3o} + (\gamma^2 - 1)\epsilon_{11} + \epsilon_{12} - \gamma \epsilon_{13})^2}{2 - \gamma^2 + \gamma^4} \right]
\]

Assuming that the \(\epsilon\)'s are uncorrelated with the \(\alpha\)'s we can re-write the above as

\[
E[q(w, \gamma)] = E \left[ \frac{(\gamma^2 - \gamma \gamma_o)\alpha_{1o} + (\gamma^2 \gamma_o - \gamma \gamma_o^2)\alpha_{2o} + (\gamma_o - \gamma)\alpha_{3o} + (\gamma^2 - 1)\epsilon_{11} + \epsilon_{12} - \gamma \epsilon_{13})^2}{2 - \gamma^2 + \gamma^4} \right]
\]

\[+ E \left[ \frac{(\gamma^2 - 1)\epsilon_{11} + \epsilon_{12} - \gamma \epsilon_{13})^2}{2 - \gamma^2 + \gamma^4} \right]
\]

Assuming that the \(\epsilon\)'s are uncorrelated with each other

\[
E[q(w, \gamma)] = E \left[ \frac{(\gamma^2 - \gamma \gamma_o)\alpha_{1o} + (\gamma^2 \gamma_o - \gamma \gamma_o^2)\alpha_{2o} + (\gamma_o - \gamma)\alpha_{3o} + (\gamma^2 - 1)\epsilon_{11} + \epsilon_{12} - \gamma \epsilon_{13})^2}{2 - \gamma^2 + \gamma^4} \right]
\]

\[+ E \left[ \frac{(\gamma^2 - 1)\epsilon_{11} + \epsilon_{12} + \gamma^2 \epsilon_{13})^2}{2 - \gamma^2 + \gamma^4} \right]
\]

Factoring out a \(\gamma - \gamma_o\) from the first term leaves

\[
E[q(w, \gamma)] = E \left[ \frac{(\gamma - \gamma_o)^2(\gamma \alpha_{1o} + \gamma \gamma_o \alpha_{2o} + \alpha_{3o})^2}{2 - \gamma^2 + \gamma^4} \right]
\]

\[+ E \left[ \frac{(\gamma^2 - 1)\epsilon_{11} + \epsilon_{12} + \gamma^2 \epsilon_{13})^2}{2 - \gamma^2 + \gamma^4} \right]
\]
Finally, assumption 5 implies we can express the above equation as:

\[
E[q(w, \gamma)] = E \left[ \frac{(\gamma - \gamma_0)^2 (\gamma_0 \alpha_{10} + \gamma_0 \alpha_{20} + \alpha_{30})^2}{2 - \gamma^2 + \gamma^4} \right] + E[\epsilon^2]
\]

For any \( \gamma \neq \gamma_0 \) the first term is always positive and the second term is not a function of \( \gamma \).

QED.

Proof of Lemma 4

Uniform convergence, requires that

\[
\max_{\gamma \in \Gamma} \left| \frac{1}{N} \sum_{n=1}^{N} q(w_n, \gamma) - E[q(w, \gamma)] \right| \xrightarrow{p} 0
\]

Theorem 12.1 in Wooldridge states four conditions that the data and \( q \) must satisfy in order for the above condition to hold.

1. \( \Gamma \) is compact
   
   This condition is satisfied by assumption 6.

2. For each \( \gamma \in \Gamma \), \( q(\cdot, \gamma) \) is Borel measurable on \( \mathcal{W} \)
   
   Since \( q(\cdot, \gamma) \) is a continuous function of \( w \), it is also Borel measurable.

3. For each \( w \in \mathcal{W} \), \( q(w, \cdot) \) is continuous on \( \Gamma \)
   
   Our concentrated objective function is continuous in \( \gamma \).

4. \( |q(w, \gamma)| \leq b(w) \) for all \( \gamma \in \Gamma \), where \( b \) is a nonnegative function on \( \mathcal{W} \) such that \( E[b(w)] < \infty \)
   
   Note that \( q(w, \gamma) \) is always positive so we can ignore the absolute value. We derive a bounding function \( b(w) \) in the following manner

\[
q(w, \gamma) = \frac{((\gamma^2 - 1)y_{11} + y_{12} - \gamma y_3)^2}{2 - \gamma^2 + \gamma^4} \leq \frac{3(\gamma^2 - 1)^2 y_{11}^2 + 3y_{12}^2 + 3\gamma^2 y_3^2}{2 - \gamma^2 + \gamma^4}
\]
where the last line follows from the triangle inequality. Since \( \gamma^4 - \gamma^2 \) is always greater than \(-1\), the following inequality must also be satisfied:

\[
q(w, \gamma) \leq 3(\gamma^4 + 1)y_{11}^2 + 3y_{12}^2 + 3\gamma^2y_3^2
\]

Let \( \gamma \) and \( \gamma \) denote the largest and smallest elements of the set \( \Gamma \) and denote \( \gamma^* = \max\{\gamma, -\gamma\} \). Our bounding function is then:

\[
b(w) = 3(\gamma^*4 + 1)y_{11}^2 + 3y_{12}^2 + 3\gamma^*2y_3^2
\]

We now show that \( E[b(w)] < \infty \), completing the proof. Since \( \gamma^* \) is finite, this amounts to establishing that \( E[y_{11}^2], E[y_{12}^2], \) and \( E[y_3^2] \) are all finite. First consider \( E[y_{12}^2] \):

\[
E[y_{12}^2] = E[(\alpha_{10} + \gamma_o\alpha_{20} + \gamma_o\alpha_{30} + \epsilon_{12})^2]
\]

Repeatedly applying the triangle inequality yields:

\[
E[y_{12}^2] \leq 4E[\alpha_{10}^2] + \gamma_o^2E[\alpha_{20}^2] + \gamma_o^2E[\alpha_{30}^2] + E[\epsilon_{12}^2])
\]

Assumption 3 and 4 ensure that all the terms of the right hand side are finite. By a similar argument, it can be shown that all the terms in \( E[b(w)] \) are finite.

QED

Proof of Lemma 5

Theorem 12.3 in Wooldridge(2002) states six conditions that must hold in order for \( \hat{\gamma} \) to be distributed asymptotically normal.

Many of these conditions involve the first and second derivatives of \( q(w, \gamma) \). The first and second derivatives are given by:

\[
s(w, \gamma) = 2\left[ (-3\gamma + 2\gamma^3 + \gamma^5) y_{11}^2 + (\gamma - 2\gamma^3) y_{12}^2 + (2\gamma - \gamma^5) y_3^2 + (2\gamma + 4\gamma^3 - 2\gamma^5) y_{11}y_{12} + (2 - 5\gamma^2 - 2\gamma^4 + \gamma^6) y_{11}y_{12} + (-2 - \gamma^2 + 3\gamma^4) y_{12}y_3 \right] / (2 - \gamma^2 + \gamma^4)^2
\]

\[
H(w, \gamma) = -2\left[ (6 - 3\gamma^2 - 3\gamma^4 + 11\gamma^6 + 3\gamma^8)y_{11}^2 + (-2 + 9\gamma^2 + 9\gamma^4 - 10\gamma^6)y_{12}^2 - (4 + 6\gamma^2 - 24\gamma^4 + \gamma^6 + 3\gamma^8)y_3^2 - 2(2 + 15\gamma^2 - 15\gamma^4 - 9\gamma^6 + 3\gamma^8)y_{11}y_{12} + 2\gamma(6 + 21\gamma^2 - 21\gamma^4 - 3\gamma^6 + \gamma^8)y_{11}y_3 + 2\gamma(6 - 19\gamma^2 - 3\gamma^4 + 6\gamma^6)y_{12}y_3 \right] / (2 - \gamma^2 + \gamma^4)^3
\]
We now show that the six conditions of Theorem 12.3 in Wooldridge(2002) are satisfied. We will refer to the above formulations of the score and Hessian throughout.

1. $\gamma_o$ must be in the interior of $\Gamma$
   
   This condition is satisfied by assumption 6.

2. $s(w, \cdot)$ is continuously differentiable on the interior of $\Gamma$ for all $w \in W$
   
   Since $H(w, \gamma)$ is continuous in $\gamma$, $s(w, \cdot)$ is continuously differentiable.

3. Each element of $H(w, \gamma)$ is bounded in absolute value by a function $b(w)$ where $E[b(w)] < \infty$

   Taking absolute values through and noting that the denominator is always greater than one implies:

   $$|H(w, \gamma)| \leq 2 \left[ (6 + 3\gamma^2 + 33\gamma^4 + 11\gamma^6 + 3\gamma^8)y_{11}^2 + (2 + 9\gamma^2 + 9\gamma^4 + 10\gamma^6)y_{12}^2 
   + (4 + 6\gamma^2 + 24\gamma^4 + \gamma^6 + 3\gamma^8)y_{12}^2 + 2(2 + 15\gamma^2 + 15\gamma^4 + 9\gamma^6 + 3\gamma^8)y_{11}y_{12} 
   + 2\gamma (6 + 21\gamma^2 + 21\gamma^4 + 3\gamma^6 + \gamma^8)y_{11}y_{12} + 2\gamma (6 + 19\gamma^2 + 3\gamma^4 + 6\gamma^6)y_{12}y_{12} \right]$$

   Applying the triangle inequality to the last three terms and collecting terms yields:

   $$|H(w, \gamma)| \leq 2 \left[ (8 + 18\gamma^2 + 48\gamma^4 + 20\gamma^6 + 6\gamma^8 + |\gamma| [6 + 21\gamma^2 + 21\gamma^4 + 3\gamma^6 + \gamma^8])y_{11}^2 
   + (4 + 24\gamma^2 + 24\gamma^4 + 19\gamma^6 + 3\gamma^8 + |\gamma| [6 + 19\gamma^2 + 3\gamma^4 + 6\gamma^6])y_{12}^2 
   + (4 + 6\gamma^2 + 24\gamma^4 + \gamma^6 + 3\gamma^8 + |\gamma| [12 + 30\gamma^2 + 24\gamma^4 + 9\gamma^6 + \gamma^8])y_{12}y_{12} \right]$$

   Our bounding function can then be found by setting $\gamma$ to $\gamma^*$ where $\gamma^* = \max\{-\gamma, \gamma\}$:

   $$b(w) = 2 \left[ (8 + 18\gamma^*^2 + 48\gamma^*^4 + 20\gamma^*^6 + 6\gamma^*^8 + \gamma^*[6 + 21\gamma^*^2 + 21\gamma^*^4 + 3\gamma^*^6 + \gamma^*^8])y_{11}^2 
   + (4 + 24\gamma^*^2 + 24\gamma^*^4 + 19\gamma^*^6 + 3\gamma^*^8 + |\gamma^*| [6 + 19\gamma^*^2 + 3\gamma^*^4 + 6\gamma^*^6])y_{12}^2 
   + (4 + 6\gamma^*^2 + 24\gamma^*^4 + \gamma^*^6 + 3\gamma^*^8 + \gamma^*[12 + 30\gamma^*^2 + 24\gamma^*^4 + 9\gamma^*^6 + \gamma^*^8])y_{12}y_{12} \right]$$

   Since $\gamma^*$ is finite and we have already established that $E[y_{11}^2], E[y_{12}^2],$ and $E[y_{12}^2]$ are all finite, then $E[b(w)] < \infty$, completing the proof.

4. $A_o \equiv E[H(w, \gamma_o)]$ is positive definite

   We first note that we can interchange the expectations and the partial derivatives:
$$E[H(w, \gamma)] = \partial^2 E[q(w, \gamma)]/\partial \gamma^2.$$ From Lemma 3, we know that $E[q(w, \gamma)]$ can be written as

$$E[q(w, \gamma)] = \frac{(\gamma - \gamma_o)^2 E \left[(\gamma \alpha_{1o} + \gamma \gamma_2 \alpha_{2o} - \alpha_3)^2 \right]}{2 - \gamma^2 + \gamma^4} + E[\epsilon^2]$$

Note that $\gamma$ operates through three terms: $(\gamma - \gamma_o)^2$ in the numerator, through the expectation in the numerator, and the denominator. However, because we are going to evaluate the expected Hessian at $\gamma_o$, we only need the second derivative of the first term, $(\gamma - \gamma_o)^2$. All of the other partial derivatives will either be multiplied by $(\gamma - \gamma_o)^2$ or $(\gamma - \gamma_o)$, both of which are zero when $\gamma = \gamma_o$. The second derivative of $(\gamma - \gamma_o)^2$ with respect to $\gamma$ is positive. This second derivative is then multiplied by the expectation of a squared object in the numerator and divided by the sum of squared objects in the denominator. Thus, the expectation of the Hessian evaluated at $\gamma_o$ is strictly positive.

5. $E[s(w, \gamma_o)] = 0$

Note that $E[s(w, \gamma)] = \partial E[q(w, \gamma)]/\partial \gamma$. Differentiating $E[q(w, \gamma)]$ with respect to $\gamma$ leaves terms that are multiplied by $(\gamma - \gamma_o)$ or by $(\gamma - \gamma_o)^2$, implying that if we evaluate the derivative at $\gamma = \gamma_o$ then the expected score is zero.

6. Each element of $s(w, \gamma_o)$ has finite second moment.

Given that the score has only one element, this condition boils down to $E[s(w, \gamma_o)^2] < \infty$.

To show this we square the score function and evaluate at the true $\gamma$.

$$E \left[s(w, \gamma_o)^2 \right] = E \left[(-6\gamma_o + 4\gamma_3^2 + 2\gamma_5^5) y_{11}^2 + (2\gamma_o - 4\gamma_3^3) y_{12}^2 + (4\gamma_o - 2\gamma_5^5) y_3^2 + (4\gamma_o + 8\gamma_3^3 - 4\gamma_5^5) y_{11} y_{12} + (4 - 10\gamma_o^2 - 4\gamma_3^4 + 2\gamma_5^6) y_{11} y_{13} + (-4 - 2\gamma_o^2 + 6\gamma_3^4) y_{12} y_{13} \right]^2 / (2 - \gamma_o^2 + \gamma_o^4)^4$$

Note that the denominator is greater than one. Further, making all terms in the numerator positive will result in an increase in the left hand side:

$$E \left[s(w, \gamma_o)^2 \right] \leq E \left[|\gamma_o| (6 + 4\gamma_o^2 + 2\gamma_5^4) y_{11}^2 + |\gamma_o| (2 + 4\gamma_3^2) y_{12}^2 + |\gamma_o| (4 + 2\gamma_5^4) y_3^2 + |\gamma_o| (4 + 8\gamma_o^2 + 4\gamma_3^4) y_{11} y_{12} + (4 + 10\gamma_o^2 + 4\gamma_3^4 + 2\gamma_5^6) y_{11} y_{13} + (4 + 2\gamma_o^2 + 6\gamma_3^4) y_{12} y_{13} \right]^2$$
Applying the triangle inequality to remove the cross-\( y \) terms and collecting terms yields:
\[
E[s(w, \gamma_o)^2] \leq E\left[ (2 + 5\gamma_o^2 + 2\gamma_o^4 + \gamma_o^6 + |\gamma_o| [8 + 8\gamma_o^2 + 4\gamma_o^4]) y_{11}^2 \right. \\
+ (2 + \gamma_o^2 + 3\gamma_o^4 + |\gamma_o| [4 + 8\gamma_o^2 + 2\gamma_o^4]) y_{12}^2 \\
+ (4 + 6\gamma_o^2 + 5\gamma_o^4 + \gamma_o^6 + |\gamma_o| [4 + 2\gamma_o^4]) y_3^2 \right]^2
\]

Label the coefficients on \( y_{11}^2, y_{12}^2, \) and \( y_3^2 \) as \( A, B, \) and \( C, \) where all are positive and finite:
\[
E[s(w, \gamma_o)^2] \leq E[Ay_{11}^2 + By_{12}^2 + Cy_3^2] \]

Taking the square through and again applying the triangle inequality yields:
\[
E[s(w, \gamma_o)^2] \leq 3A^2 E[y_{11}^4] + 3B^2 E[y_{12}^4] + 3C^2 E[y_3^4]
\]

With \( A, B, \) and \( C \) positive and finite, we now need to establish \( E[y_{11}^4], E[y_{12}^4], \) and \( E[y_3^4] \) are finite. Consider \( E[y_{12}^4] \) and substitute in for the DGP:
\[
E[y_{12}^4] = E\left[ (\alpha_{1o} + \gamma_o\alpha_{2o} + \gamma_o\alpha_{3o} + \epsilon_{11})^2 \right]
\]

Repeatedly applying the triangle inequality yields
\[
E[y_{12}^4] \leq 16E\left[ (\alpha_{1o}^2 + \gamma_o^2\alpha_{2o}^2 + \gamma_o^2\alpha_{3o}^2 + \epsilon_{11}^2) \right]^2 \\
\leq 64 \left( E[\alpha_{1o}^4] + \gamma_o^4 E[\alpha_{2o}^4] + \gamma_o^4 E[\alpha_{3o}^4] + E[\epsilon_{11}^4] \right)
\]

Assumptions 3 and 4 ensure that all of the terms on the right hand side of the inequality in the above equation are finite. Thus, \( E[y_{12}^4] \) is finite. By a similar argument, it can be shown that all the terms in the expectation of the squared score are finite.

QED

B Endogenous Effects

In this section we show how our framework can be incorporated to allow for endogenous effects. We introduce a new variable, \( Z_{itn} \), that affects the choices of the individual but affects his peers only through the individual’s choice. Throughout, we assume that \( Z_{itn} \) is uncorrelated
with all the $\epsilon$’s. For ease of notation, we also focus on the case where peer groups consist of only two individuals.\footnote{Results for larger peer groups are available upon request.} We first consider the case where individuals have total control of the outcome: the outcome of interest is a choice. We then consider the case that is most relevant to our empirical work, where individuals only have partial control over the outcome.

**B.1 Total Control**

We first consider the case where $Y_{itn}$ is directly affected by $Y_{jtn}$. In this case, the linear model is:

$$Y_{itn} = \alpha_{io} + \gamma_o \alpha_{jo} + \phi_o Y_{jtn} + \theta_o Z_{itn} + \epsilon_{itn}$$  \hspace{1cm} (31)

Substituting into (31) the expression for $Y_{jtn}$ and solving for $Y_{itn}$ yields:

$$Y_{itn} = \left( \frac{1 + \phi_o \gamma_o}{1 - \phi_o^2} \right) \alpha_{io} + \left( \frac{\gamma_o + \phi_o}{1 - \phi_o^2} \right) \alpha_{jo} + \frac{\theta_o Z_{itn}}{1 - \phi_o^2} + \frac{\phi_o \theta_o Z_{jtn}}{1 - \phi_o^2} + \frac{\epsilon_{itn} + \phi_o \epsilon_{jtn}}{1 - \phi_o^2}$$  \hspace{1cm} (32)

Note that the last term, the reduced form error, has both $\epsilon_{itn}$ and $\epsilon_{jtn}$. The reduced form errors will then be correlated between individuals who share a peer group, violating assumption 3 of Theorem 1. In estimation, this correlation is partially absorbed by the peer fixed effects, which in turn prohibits consistent estimation of the coefficient on $\alpha_{jo}$. Our conclusion is that when the outcome variable is a choice that is affected by the actual choices of one’s peers, we cannot obtain a consistent estimate of the parameter on the peer fixed effects for fixed $T$.

Note, however, that if the spillovers only operated through observables, which would imply replacing the $\alpha_{io}$’s with $X\beta$, then all of the structural parameters would be identified.

We now consider the case where individuals only have expectations about what their peers will choose. This situation maps well to a wide variety of outcomes where the behavior of others is either not perfectly observed, or occurs at exact the same time as own behavior and therefore cannot be a direct input to own behavior. In particular, suppose that $\epsilon_{jt}$ is unknown to individual $i$ and has mean zero. The outcome equation is then:

$$Y_{itn} = \alpha_{io} + \gamma_o \alpha_{jo} + \phi_o E(Y_{jtn}) + \theta_o Z_{itn} + \epsilon_{itn}$$  \hspace{1cm} (33)
Again substituting in for $Y_{jtn}$ and solving for $Y_{itn}$ yields:

$$Y_{itn} = \left( \frac{1 + \phi_o \gamma_o}{1 - \phi_o^2} \right) \alpha_{io} + \left( \frac{\gamma_0 + \phi_o}{1 - \phi_o^2} \right) \alpha_{jo} + \frac{\theta_o Z_{itn}}{1 - \phi_o^2} + \frac{\phi_o \theta_o Z_{jtn}}{1 - \phi_o^2} + \frac{\epsilon_{itn}}{1 - \phi_o^2} \tag{34}$$

Assumption 3 of Theorem 1, that the reduced form error is uncorrelated between peer group members, is no longer violated by the model. We can then write (34) as:

$$Y_{itn} = \alpha^*_{io} + \gamma^* o \alpha_{jo} + \theta^* o Z_{itn} + \phi^* o Z_{jtn} + \epsilon^*_{itn} \tag{35}$$

where:

$$\alpha^*_{io} = \left( \frac{1 + \phi_o \gamma_o}{1 - \phi_o^2} \right) \alpha_{io}$$

$$\gamma^* o = \left( \frac{\gamma_0 + \phi_o}{1 + \phi_o \gamma_o} \right)$$

$$\theta^* o = \frac{\theta_o}{1 - \phi_o^2}, \quad \phi^* o = \phi_o \theta^* o, \quad \epsilon^* = \frac{\epsilon_{itn}}{1 - \phi_o^2}$$

Estimating the reduced form then makes it possible to recover all the structural parameters, as would also hold in the standard case where the $\alpha_{io}$’s were replaced by a set of observables multiplied by a vector of coefficients. We can recover $\hat{\phi}$ and $\hat{\theta}$ from $\hat{\phi}^*$ and $\hat{\theta}^*$. Next, given $\hat{\phi}$, we can obtain $\hat{\gamma}$ using $\hat{\gamma}^*$ as $\hat{\gamma} = (\hat{\gamma}^* - \hat{\phi})/(1 - \hat{\gamma}^* \hat{\phi})$.

One key identifying assumption in this case is that the expected choices of the individual’s peers are formed on the basis of observed characteristics and the peer fixed effects, both of which are uncorrelated with the structural errors. Identification of the underlying parameters using our fixed-effects approach also requires $Z_{itn}$ to be time-varying. If it is not, then $Z_{itn}$ would be absorbed into the reduced-form individual effect, and we would be back to using two coefficients to recover three parameters. We would be left with the same estimating equation as the baseline model, and the reduced form would be a linear combination of own and peer fixed effects plus the $Z$ values of the peers, but we could not separate out the endogenous effects from the exogenous effects. Note that in the case that spillovers operated only through observable characteristics, $Z_{itn}$ is only required to vary across individuals, not within-person.

### B.2 Partial Control

As pointed out by Cooley (2009b) and Cooley (2009a), the estimation issues become much more complicated when individuals only have partial control over their outcomes. For example,
in educational settings where grades are the outcome of interest, it is not the grades of the
other students in the class that affect the student’s grades, but the effort the other students
exert. Moreover, students cannot directly choose their grades but can only choose effort levels
which in turn combine with other forces (including peer effort) to determine their grades.
Separating out endogenous and exogenous effects is much harder in this case.

We now show what we can identify when individuals make choices that only partially affect
their outcome, and where the choices of others influence both own choices and own outcomes.
Let \( e_{itn} \) indicate the continuous choice individuals make to affect outcome \( Y_{itn} \). Adding \( e_{itn} \)
and \( e_{jtn} \) to the baseline model as direct influences on outcomes yields:

\[
Y_{itn} = \alpha_{io} + \phi_{1o} e_{itn} + \gamma_{jo} \alpha_{jo} + \phi_{2o} e_{jtn} + \epsilon_{itn} \tag{36}
\]

The utility associated with choosing a particular value of \( e_{itn} \) depends on the individual’s
fixed effect, \( \alpha_{io} \), as well as on the choices of the other individual and their individual effect.
Similar to the previous case, we assume that there is an additional variable, \( Z_{itn} \), that affects
the choice of effort. We assume that the utility function takes the following form:

\[
U(e_{itn}, E(Y_{itn})) = E(Y_{itn}) + e_{itn}(\lambda_{1o} \alpha_{io} + \lambda_{2o} Z_{itn} + \lambda_{3o} e_{jtn} + \lambda_{4o} \alpha_{jo}) - e_{itn}^2/2 \tag{37}
\]

where we have normalized the coefficient on the squared term. The first order condition from
maximizing (37) with respect to \( e_{itn} \) and solving for \( e_{itn} \) implies that own optimal effort can
be written as:

\[
e_{itn} = \phi_{1o} + \lambda_{1o} \alpha_{io} + \lambda_{2o} Z_{itn} + \lambda_{3o} e_{jtn} + \lambda_{4o} \alpha_{jo} \tag{38}
\]

Substituting in for \( e_{jtn} \) from \( j \)'s maximization problem into (38) yields:

\[
e_{itn} = \frac{(1 + \lambda_{3o}) \phi_{1o} + (\lambda_{1o} + \lambda_{3o} \lambda_{4o}) \alpha_{io} + (\lambda_{4o} + \lambda_{3o} \lambda_{1o}) \alpha_{jo} + \lambda_{2o} Z_{itn} + \lambda_{3o} \lambda_{2o} Z_{jtn}}{(1 - \lambda_{3o}^2)} \tag{39}
\]

Substituting in for \( e_{itn} \) and \( e_{jtn} \) in equation (36) and collecting terms implies we can rewrite
(36) as:

\[
Y_{itn} = \alpha_{io}^* + \phi_{1o}^* Z_{itn} + \gamma_{jo}^* \alpha_{jo}^* + \phi_{2o}^* Z_{jtn} + \epsilon_{itn}^* \tag{40}
\]

where:

\[
\alpha_{io}^* = C + \left(1 + \frac{\phi_{1o}(\lambda_{1o} + \lambda_{3o} \lambda_{4o}) + \phi_{2o}(\lambda_{4o} + \lambda_{3o} \lambda_{1o})}{1 - \lambda_{3o}^2}\right) \alpha_{i}
\]

\[
\gamma_{jo}^* = \frac{(1 - \lambda_{3o}^2) \gamma_{jo} + \phi_{2o}(\lambda_{1o} + \lambda_{3o} \lambda_{4o}) + \phi_{1o}(\lambda_{4o} + \lambda_{3o} \lambda_{1o})}{(1 - \lambda_{3o}^2 + \phi_{1o}(\lambda_{1o} + \lambda_{3o} \lambda_{4o}) + \phi_{2o}(\lambda_{4o} + \lambda_{3o} \lambda_{1o}))}
\]
\[ \phi_{1'o}^* = \frac{\lambda_2(\phi_{1'o} + \lambda_3\phi_{2'o})}{1 - \lambda_{3'o}^2}, \quad \phi_{2'o}^* = \frac{\lambda_2(\phi_{2'o} + \lambda_3\phi_{1'o})}{1 - \lambda_{3'o}^2}, \quad \epsilon_{itn}^* = \frac{\epsilon_{itn}}{1 - \lambda_{3'o}^2}, \]

and where \( C \) is the adjustment to \( \alpha_{i'o}^* \) coming from the \( \phi_o \) terms that are not multiplying a regressor.

Reduced-form estimation will then yield estimates of three coefficients, \( \hat{\phi}_{1'o}^* \), \( \hat{\phi}_{2'o}^* \), and \( \hat{\gamma}^* \), that are functions of six underlying parameters. What we can say is that \( \hat{\phi}_{1'o}^* \) being greater than zero implies that individual effort either directly affects the outcome or affects the outcome through the other individual’s effort, which in turn affects the individual’s outcome. Similarly, if the coefficient on \( Z_{jtn} \), \( \hat{\phi}_{2'o}^* \), is greater than zero, we can conclude that peer effort matters in some form, either directly or through affecting the individual’s own effort. Once again, these results are essentially identical to those in Cooley (2009b), subject to replacing observable characteristics with individual effects.
Table 1: Baseline Model, $\gamma_0 = .15$

<table>
<thead>
<tr>
<th>Obs. Per Peer Group</th>
<th>Random Assignment</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Size</td>
<td>$\sigma_\varepsilon=1.95$</td>
<td>$\sigma_\varepsilon=1.15$</td>
</tr>
<tr>
<td>2 2</td>
<td>$\hat{\gamma}$</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.034)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.706</td>
</tr>
<tr>
<td>5 10</td>
<td>$\hat{\gamma}$</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.041)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.482</td>
</tr>
<tr>
<td>10 10</td>
<td>$\hat{\gamma}$</td>
<td>0.150</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.025)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.415</td>
</tr>
</tbody>
</table>

Note: The R-squared values reported in this table are those pertaining to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student in order to manipulate the amount of variation in performance that is explained by the ability measure. Parameter values are averages over 100 simulations on a population of 10,000 students.
Table 2: Varying Class Size and Heteroskedasticity, $\gamma_0 = .15$

<table>
<thead>
<tr>
<th>Obs. Per Peer Group</th>
<th>Random Assignment</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Size</td>
<td>$\sigma_\epsilon$</td>
<td>$\sigma_\epsilon$</td>
</tr>
<tr>
<td>5 U[5,15]</td>
<td>1.95</td>
<td>1.15</td>
</tr>
<tr>
<td>$\hat{\gamma}$</td>
<td>0.158</td>
<td>0.150</td>
</tr>
<tr>
<td>(0.041)</td>
<td>(0.022)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.484</td>
<td>0.686</td>
</tr>
</tbody>
</table>

| 5 U[5,15]           | N(1.95,.09)       | N(1.15,.09) | N(1.95,.09)       | N(1.15,.09)       |
| $\hat{\gamma}$     | 0.148             | 0.151      | 0.146             | 0.149             |
| (0.037)             | (0.022)           | (0.059)    | (0.026)           |
| $R^2$               | 0.480             | 0.67       | 0.497             | 0.692             |

| 5 U[5,15]           | N(.95+size/10,.09) | N(.15+size/10,.09) | N(.95+size/10,.09) | N(.15+size/10,.09) |
| $\hat{\gamma}$     | 0.153             | 0.149      | 0.150             | 0.149             |
| (0.049)             | (0.029)           | (0.061)    | (0.027)           |
| $R^2$               | 0.459             | 0.633      | 0.474             | 0.649             |

Note: The R-squared values reported in this table are those pertaining to the regression of grades onto the constructed fixed effect values. We alter the random error added on to the constructed grade for each student in order to manipulate the amount of variation in performance that is explained by the ability measure. For the bottom two-thirds of the table, the standard deviation of the random error varies across peer groups according to the distributions listed. The variable size refers to the size of the peer group. Parameter values are averages over 100 simulations on a population of 10,000 students.
Table 3: Heterogenous Gamma Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Random Assignment</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heterogeneity in Responsiveness to Peers</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{1o} = .15$</td>
<td>0.151</td>
<td>0.146</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.029)</td>
</tr>
<tr>
<td>$\gamma_{2o} = .1$</td>
<td>0.100</td>
<td>0.094</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.032)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.683</td>
<td>0.699</td>
</tr>
<tr>
<td>Heterogeneity in Peer Influence</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_{1o} = .15$</td>
<td>0.150</td>
<td>0.151</td>
</tr>
<tr>
<td></td>
<td>(0.029)</td>
<td>(0.037)</td>
</tr>
<tr>
<td>$\gamma_{2o} = .1$</td>
<td>0.102</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>(.030)</td>
<td>(.039)</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.687</td>
<td>0.684</td>
</tr>
</tbody>
</table>

Note: The R-squared values reported in this table are those pertaining to the regression of grades onto the constructed fixed effect values. Parameter values are averages over 100 simulations on a population of 10,000 students. Each student is observed 5 times with a total group size of 10 students.
Table 4: Accumulation Model, $\gamma_0 = .15$

<table>
<thead>
<tr>
<th>Obs. Per Peer Group</th>
<th>Random Assignment</th>
<th>Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student Size</td>
<td>$\sigma_\epsilon = 1.95$</td>
<td>$\sigma_\epsilon = 1.5$</td>
</tr>
<tr>
<td>2</td>
<td>$\hat{\gamma}$</td>
<td>0.149</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.046)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.608</td>
</tr>
<tr>
<td>2</td>
<td>$\hat{\gamma}$</td>
<td>0.154</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.091)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.607</td>
</tr>
<tr>
<td>2</td>
<td>$U[2,8]$</td>
<td>$\hat{\gamma}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.083)</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>0.606</td>
</tr>
</tbody>
</table>

Note: The R-squared values reported in this table are those pertaining to the regression of grades onto the constructed fixed effect values. In practice, the fixed effects change over time as individuals internalize previous spillover effects. We alter the random error added on to the constructed grade for each student in order to manipulate the amount of variation in performance that is explained by the ability measure. Parameter values are averages over 100 simulations on a population of 10,000 students.