Peer Effects with Sample Selection*

Xin Gu† Haizheng Li‡ Zhongjian Lin § Xun Tang¶

November 1, 2022

Abstract

This paper studies peer effects in the presence of sample selection. Nonrandom selection is a fundamental aspect of many social processes. We By including individual correction terms for selection bias, we attain a sufficient rank condition for recovering the structural parameters from the reduced form, thus solving the reflection problem. We apply our method to study peer effects in an online training program for teachers in China, where the trainees participate in lectures through endogenous self-selection. We document significant peer effects and selection bias in the duration of lecture attendance among the trainees. Ignoring sample selection would yield misleading estimate and inference of the peer effects.

Keywords: Peer Effects, Sample Selection, Social Interactions, Reflection Problem, Online Training.

JEL Codes: C24, C31, C36, M53.

*We are grateful to Peng Shao, Nicholas Ziebarth, and audience at Auburn University for useful comments and suggestions. All errors are our own.

†School of Economics, Georgia Institute of Technology. 221 Bobby Dodd Way Atlanta, GA 30332. Email: xin.gu@gatech.edu.

‡School of Economics, Georgia Institute of Technology. 221 Bobby Dodd Way Atlanta, GA 30332. Email: haizheng.li@econ.gatech.edu.

§John E. Walker Department of Economics, Clemson University. 225 Walter T. Cox Blvd, Clemson, SC 26934. Email: zhongjl@clemson.edu.

¶Department of Economics, Rice University. 6100 Main St. Houston, TX 77005. Email: xun.tang@rice.edu.

1 Introduction

Sample selection is ubiquitous in social studies (Heckman, 1974, 1978, 1979). A typical example is the wage structure of female workers, where only wages of those who participate in the labor force are observed (Gronau, 1974, Heckman, 1974). A classical approach for dealing with the sample selection problem is the Heckit two-step correction with parametric assumptions on the data-generating process that determines sample selection and observed outcomes. This approach models the dependence between selection and outcomes within each independent observation, and does not allow for spillover effects in the outcome and/or selection across individuals.

In this paper, we investigate the sample selection problem in an alternative setting where individual outcomes are influenced by peer effects. Social interactions models with peer effects have proliferated in empirical research over the last three decades (Manski, 1993, 2000, Brock and Durlauf, 2001, Moffitt, 2001, Lee, 2007, Bramoullé, Djebbari, and Fortin, 2009). Evidence of peer effects has been found in many fields, including economics of education (Hoxby, 2000, Sacerdote, 2001, Calvó-Armengol, Patacchini, and Zenou, 2009), financial economics (Hong, Kubik, and Stein, 2004), health economics (Trogdon, Nonnemaker, and Pais, 2008), labor economics (Topa, 2001, Dahl, Løken, and Mogstad, 2014), and urban economics (Glaeser, Sacerdote, and Scheinkman, 1996).

Most existing papers that study peer effects take the formation of groups as exogenously given. We fill this gap in the literature by studying a social interactions model when the *potential* members of a group self-select themselves to join the *actual* group (e.g., attend the same lectures).[1] The outcome for each actual group member is then affected by, and determined simultaneously with, the outcomes of group peers.

We combine the identification of social interactions models with the solution to sample selection issues. Our work contributes to the econometrics literature in two ways. First, we show that in a social interactions model with peer effects, correcting for sample selection bias in the reduced form of individual outcomes involves instruments from other group members as well. This differs from the

[1] To the best of our knowledge, the only exception is Sheng and Sun (2021), who models group formation using a notion of stability from a matching model.

classical case with no social interactions, where each individual's own instrument suffices for correcting the sample selection bias. This is because the reduced form of a linear-in-means social interactions model contains an individual structural error, and a composite error that depends on the structural errors of all other peers in the group. Therefore, to correct for sample selection in the reduced form, we need to include not only individual-specific but also composite group-level correction terms.

Second, we show that the inclusion of individual instruments for sample selection provides enough exogenous variation for identifying the endogenous peer effects in the social interactions model. In other words, by introducing the selection instruments, we do not need to invoke additional assumptions for identifying these peer effects.

To put this in context, consider a linear-in-means social interactions model, which accommodates a contextual effect from average characteristics of group members, and an endogenous peer effect capturing a structural simultaneity between all individual outcomes within a group. Manski (1993) points out an identification problem (a.k.a. the "reflection problem") exists in such a model, because the peer effects can not be separated from the contextual effects from the reduced form without further restrictions. Manski (1993) proposes a solution to the problem using an exclusion restriction, i.e. there are covariates which have non-trivial direct effects but no contextual effects (Proposition 2). Moffitt (2001) uses similar exclusion restrictions to identify peer effects.

Our solution builds on this important insight from Manski (1993) in the following sense. We introduce individual correction terms for selection bias into the structural form of social interactions models. While doing so, we also show that these correction terms essentially function as *generated* regressors that satisfy the exclusion restriction in Manski (1993). Thus we do not need to invoke any additional assumptions for identifying peer effects.[2]

Building on our constructive identification strategy, we develop a multi-step estimator for the structural parameters. First, a Probit regression of the selection equation provides consistent estimates for the bias correction term. Second, a

[2]Apart from the exclusion restriction in (Manski, 1993), the econometrics literature has offered several alternatives to solve the reflection problem: second moment restrictions on the error terms (Lee, 2007, Graham, 2008, Sacerdote, 2001), variation in the size of networks/groups (Bramoullé, Djebbari, and Fortin, 2009, De Giorgi, Pellizzari, and Redaelli, 2010, Lin, 2010), and control functions for endogenous covariates (Lin and Tang, 2022), etc.

linear regression that includes the estimates of individual and composite bias correction terms as generated regressors consistently estimates the reduced-form parameters in the social interactions model, which are then used for recovering the structural parameters.

We apply our method to study the peer effects in an online training program for elementary and middle school teachers in China. In this environment, the teachers enrolled in the program decided to participate or skip each specific lecture, based on self-motivation and other factors. This results in a non-random sample of groups (lecture attendants from the same county) based on endogenous self-selection. The outcome of interest is the duration of lecture attendance by each individual. This depends on unobserved noises related to self-motivation, which may well be correlated with those determining lecture participation in the first place. Thus, this endogenous sample selection gives rise to new challenges in the identification and estimation of peer effects in the duration of lecture attendance. Using our method for addressing sample selection, we find significant peer effects among trainees attending the same lecture. Also, ignoring the sample selection issue in this context would result in an erroneous conclusion about the magnitude and significance of the peer effects.

The paper unfolds as follows. We introduce the peer effects model with sample selection and discuss its identification in the next section. We propose a multi-step estimator in Section 3. Section 4 then shows how to extend the method where there are unobserved group fixed effects. We show the finite-sample performance of the estimator via monte carlo simulations in Section 5. Finally, we apply our method in the empirical application of peer effects in the online training program for teachers in Section 6. Section 7 concludes.

2 The Model

We consider a data-generating process (DGP) which generates a large number of independent groups, indexed by $g = 1, 2, ..., G$. Each group has a set of *potential* members, denoted by $\mathcal{N}_g$. We suppress the group index g in notation in this section.

Our model extends conventional linear-in-means social interactions models by allowing individuals to join a group as *actual* members through endogenous self-selection. Specifically, for $i \in \mathcal{N}$, the decision to join a group is determined

as:

$$D_i = 1\{Z_i'\delta + V_i \geq 0\}. \tag{1}$$

Henceforth we refer to Z_i as individual *instruments*.

Within each group, let n denote the number of actual members, i.e., $\{i \in \mathcal{N} : D_i = 1\}$. The vector of outcomes of actual group members are determined simultaneously as:

$$Y = \alpha\overline{Y} + \beta_0 + X'\beta + \overline{X}'\gamma + U, \tag{2}$$

where X' is an $n \times K$ matrix of individual characteristics (which does not include a constant intercept), $\overline{X}'$ is an $n \times K$ matrix of n identical rows, each being a $1 \times K$ vector of average characteristics within the group, $\overline{Y}$ is the average outcome of individuals within the group, and $U \equiv (U_i)_{i\leq n}$ is an $n \times 1$ vector of individual structural errors.

The individual instruments Z_i contain elements that are not in X_i. While the data reports (D_i, Z_i) for all *potential* group members $i \in \mathcal{N}$, it only reports individual outcomes Y_i and demographics X_i for n *actual* group members.

By solving for the reduced form of $\overline{Y}$ and substituting it back into the structural form in (2), we have

$$Y = \tilde{\beta}_0 + X'\beta + \overline{X}'\tilde{\gamma} + \widetilde{U}, \tag{3}$$

where $\widetilde{U} \equiv U + \frac{\alpha}{1-\alpha}\overline{U}$ with $\overline{U} \equiv \frac{1}{n}\sum_{i\leq n} U_i$, $\tilde{\beta}_0 \equiv \frac{\beta_0}{1-\alpha}$, and $\tilde{\gamma} \equiv \frac{\alpha\beta+\gamma}{1-\alpha}$. Let $\mathcal{S}$ be shorthand for the selection event that "$D_i = 1$ for all $i = 1, 2, ..., n$ and $D_j = 0$ for all other $j \in \mathcal{N}$". Define $\varepsilon \equiv \widetilde{U} - E(\widetilde{U}|X, Z, \mathcal{S})$, where $Z \equiv (Z_i)_{i\in\mathcal{N}}$ denotes the vector of all instruments associated with all individual members. Then (3) can be written as

$$Y = \tilde{\beta}_0 + X'\beta + \overline{X}'\tilde{\gamma} + E(\widetilde{U}|X, Z, \mathcal{S}) + \varepsilon, \tag{4}$$

where (X, Z) are exogenous in the sense that $E(\varepsilon|X, Z, \mathcal{S}) = 0$. Thus the conditional mean of $\widetilde{U}$ in (4) serves as a correction for the sample selection bias, which is due to the correlation between U and $V \equiv (V_i)_{i\in\mathcal{N}}$. We maintain the following assumptions, which allows us to derive the correction term.

Assumption 1. *(i)* $E[U_i|V, Z, (X_i)_{i\in\mathcal{N}}] = E(U_i|V_i)$ *for each* $i \in \mathcal{N}$. *(ii)* V *is independent from* $(X_i, Z_i)_{i\in\mathcal{N}}$, *and* V_i*'s are independent across* $i \in \mathcal{N}$. *(iii) For each i, the vector*

(U_i, V_i) follows a bivariate normal distribution with $\sigma_{uv} \neq 0$:

$$\begin{pmatrix} U_i \\ V_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma_{uv} \\ \sigma_{uv} & 1 \end{pmatrix}\right).$$

Remark 1. *Assumption 1 allows the n-vector of structural errors U to be correlated among individual members, possibly through group fixed effects. It only restricts the correlation between U_i and V_i for each i, as well as the independence between the selection errors V_i across $i \in \mathcal{N}$.*[3] *Note that the issue of endogenous sample selection arises because the structural errors in the outcomes and participation decisions are correlated, i.e., $\sigma_{uv} \neq 0$. If $\sigma_{uv} = 0$ in Assumption 1, there would be no sample selection bias at all.*

Under Assumption 1, we have

$$E(U_i|X, Z, \mathcal{S}) = E(U_i|V_i \geq -Z_i'\delta) = \sigma_{uv}\lambda(Z_i'\delta), \tag{5}$$

where $\lambda(Z_i'\delta) = \frac{\phi(-Z_i'\delta)}{1-\Phi(-Z_i'\delta)} \equiv \lambda_i$ is the inverse Mills ratio (as in Heckman, 1979). Since its introduction by Heckman (1979), this method for correcting sample selection bias has been applied widely in theory and practice. After correcting for the selection bias, we write the reduced form for each i as

$$Y_i = \tilde{\beta}_0 + X_i'\beta + \overline{X}'\tilde{\gamma} + \sigma_{uv}\lambda_i + \tilde{\sigma}_{uv}\bar{\lambda} + \varepsilon_i \tag{6}$$

where $\tilde{\sigma}_{uv} \equiv \frac{\alpha}{1-\alpha}\sigma_{uv}$, and $\bar{\lambda}$ is the average of λ_i over all $i \leq n$ in a group.

Let $W_i \equiv (1, X_i', \overline{X}', \lambda_i, \bar{\lambda})$ denote a row-vector in $\boldsymbol{R}^{2K+3}$. Provided $E(W_i'W_i)$ is non-singular, we can consistently estimate $(\tilde{\beta}_0, \beta, \tilde{\gamma}, \sigma_{uv}, \tilde{\sigma}_{uv})$ by regressing Y_i on W_i. Then we can recover the structural parameters in (2) as:

$$\alpha = \frac{\tilde{\sigma}_{uv}}{\tilde{\sigma}_{uv} + \sigma_{uv}};\ \beta_0 = (1-\alpha)\tilde{\beta}_0;\ \gamma = (1-\alpha)\tilde{\gamma} - \alpha\beta. \tag{7}$$

As noted above, if $\sigma_{uv} = 0$, there would be no sample selection bias, and the model of outcomes in (2) would reduce to a standard linear-in-means model. In that case, we have $2K + 1$ parameters, $\{\tilde{\beta}_0, \beta, \tilde{\gamma}\}$ in the reduced form and $2K + 2$ parameters $\{\alpha, \beta_0, \beta, \gamma\}$ in the structural form. The lack of a full rank between the two forms leads to the reflection problem (Manski, 1993). With $\sigma_{uv} \neq 0$, we introduce one additional parameter σ_{uv} in the structural form, but generate two parameters $\{\sigma_{uv}, \tilde{\sigma}_{uv}\}$ in the reduced form. This retains a full rank condition that allows us to recover the structural parameters from the reduced form ones,

[3] For example, suppose (U, V) is multivariate normal and independent from (X, Z). Assumption 1 does not restrict the off-diagonal entries in the upper-left quadrant of its covariance matrix. On the other hand, it requires all three of the remaining quadrants to be diagonal.

because the number of parameters is $2K + 3$ in both cases. Without sample selection, we would not be able to use individual instruments from (1) as a source of exogenous variation to help us resolve the reflection problem.

There is another intuitive interpretation of our method. We have introduced individual-level correction terms for the selection bias in the structural form of this model. These correction terms then conveniently function as *generated* regressors which satisfy the exclusion restrictions in Manski (1993). Thus we are able to solve the "reflection problem" without imposing further assumptions.

3 Estimation

We define a multi-step estimator using the constructive identification strategy above. For simplicity, we present the estimator when X is a strict sub-vector of Z; generalization to cases where Z contains distinct elements from X is straightforward.

Let the sample contain G independent groups. Each group g is formed out of a finite superset of *potential* members, which is denoted by $\mathcal{N}_g$. Each potential member $i \in \mathcal{N}_g$ chooses to join the group g or not, $D_{g,i} \in \{0, 1\}$, by following the rule in Equation (1). We refer to those who choose $D_{g,i} = 1$ and self-select into the group as the *actual* group members. Let $n_g = \sum_{i \in \mathcal{N}_g} D_{g,i}$ denote the actual size of group g. For each group $g \leq G$, the sample reports $(D_{g,i}, Z_{g,i})$ for all potential members $i \in \mathcal{N}_g$, but only reports $Y_{g,i}$ for actual group members who self-select to join the group ($D_{g,i} = 1$). Similar to Section 2, let $\mathcal{S}_g$ denote the sample selection event in potential group g.

It is worth pointing out that the identification strategy in Section 2 applies to groups with at least two actual members. Formally, this means the sample correction term in Equation (4) needs to condition on $n_g \geq 2$. The identification strategy in Section 2 applies because under Assumption 1 the individual correction term takes the form in Equation (5). That is, $E(U_{g,i}|X_g, Z_g, \mathcal{S}_g, n_g \geq 2) = E(U_{g,i}|V_{g,i} \geq -Z'_{g,i}\delta)$.

The first step of our estimator is to construct individual correction terms $\lambda_{g,i}$'s for $i \leq n_g$ by running a Probit regression of $D_{g,i}$ on $Z_{g,i}$ in Equation (1) using *all* potential group members $i \in \mathcal{N}_g$. Let $\hat{\delta}$ denote the Probit estimator for δ from this

step. For each actual member $i \leq n_g$ in group g, calculate

$$\hat{\lambda}_{g,i} \equiv \phi(Z'_{g,i}\hat{\delta})/\Phi(Z'_{g,i}\hat{\delta}), \text{ and } \hat{\bar{\lambda}}_g \equiv \frac{1}{n_g}\sum_{i=1}^{n_g}\hat{\lambda}_{g,i}.$$

The second step is an OLS regression of $Y_{g,i}$ on $X_{g,i}, \overline{X}_g, \hat{\lambda}_{g,i}$ and $\hat{\bar{\lambda}}_g$ using the actual group members. Let $\theta \equiv \left(\tilde{\beta}_0, \beta', \tilde{\gamma}', \sigma_{uv}, \tilde{\sigma}_{uv}\right)'$ be a column-vector that collects all reduced-form parameters to be estimated in this step. For each group g and actual group member $i \leq n_g$, define a row-vector of generated regressors:

$$W_{g,i}(\delta) \equiv \left(1, X'_{g,i}, \overline{X}'_g, \hat{\lambda}_{g,i}, \hat{\bar{\lambda}}_g\right).$$

Denote the total number of actual group members in the sample by $N = \sum_{g\leq G} n_g$. Let $W(\hat{\delta})$ be an N-by-dim(θ) matrix that stacks the row-vector of generated regressors $W_{g,i}(\hat{\delta})$ from all groups and actual members, and Y be an N-by-1 vector that stacks the column-vectors Y_g from all groups in the sample. Our estimator for θ in this step is constructed by regressing Y on $W(\hat{\delta})$:

$$\hat{\theta} \equiv \left[\sum_{g,i} W_{g,i}(\hat{\delta})'W_{g,i}(\hat{\delta})\right]^{-1}\left[\sum_{g,i} W_{g,i}(\hat{\delta})'Y_{g,i}\right] = [W(\hat{\delta})'W(\hat{\delta})]^{-1}W(\hat{\delta})Y,$$

where $\sum_{g,i}$ is shorthand for the double summation $\sum_{g\leq G}\sum_{i\leq n_g}$. By definition, $\hat{\theta}$ provides estimators for the reduced-form parameters $(\hat{\tilde{\beta}}_0, \hat{\beta}, \hat{\tilde{\gamma}}, \hat{\sigma}_{uv}, \hat{\tilde{\sigma}}_{uv})$.

The last step is to calculate the structural parameters from elements in $\hat{\theta}$ using Equation (7). We denoted the estimators by $(\hat{\alpha}, \hat{\beta}_0, \hat{\gamma})$.

We sketch a proof for the asymptotic property of $\hat{\theta}$ as a two-step m-estimator as follows. Let $A \equiv \lim_{G\to\infty}\frac{1}{G}\sum_{g\leq G} E\left(W'_g W_g\right)$, where W_g is shorthand for $W_g(\delta)$, which stacks $W_{g,i}(\delta)$ over all $i \leq n_g$, and is evaluated at the true parameter in the selection equation (1). First, under standard regularity conditions, e.g., including finite, non-singular A and those in Lemma 4.3 of Newey and McFadden (1994), $\frac{1}{G}W(\hat{\delta})'W(\hat{\delta})$ and $\frac{1}{G}W(\hat{\delta})'Y$ converge in probability to A and $A\theta$ respectively as $G \to \infty$. This establishes the consistency of our estimator: $\hat{\theta} \xrightarrow{p} \theta$.

Next, it can be established under standard regularity conditions that the first-order condition in the second-step regression implies:

$$\sqrt{G}\left(\hat{\theta} - \theta\right) = A^{-1}\left\{-G^{-1/2}\sum_g s_g(\theta;\hat{\delta})\right\} + o_p(1),$$

where $s_g(\theta;\hat{\delta}) \equiv W_g(\hat{\delta})'\left[Y_g - W_g(\hat{\delta})\theta\right]$, with $W_g(\hat{\delta})$ being n_g-by-dim(θ) and stacking $W_{g,i}(\hat{\delta})$ across i in each group g. A mean-value expansion of $s_g(\theta;\hat{\delta})$

around δ implies

$$G^{-1/2}\sum\nolimits_g s_g(\theta;\hat{\delta}) = G^{-1/2}\sum\nolimits_g s_g(\theta;\delta) + F_{g,0}\sqrt{G}(\hat{\delta}-\delta) + o_p(1),$$

where $F_{g,0} \equiv E\left[\nabla_\delta s_g(\theta;\delta)\right] \in \mathbb{R}^{dim(\theta)\times dim(\delta)}$ is a matrix of expected Jacobian. Let $r_g(\delta)$ denote the influence function in the asymptotic linear representation of the first-step estimator $\hat{\delta}$. That is, $\sqrt{G}(\hat{\delta}-\delta) = G^{-1/2}\sum_g r_g(\delta) + o_p(1)$. It then follows that the limiting distribution of $\hat{\theta}$ is

$$\sqrt{G}\left(\hat{\theta}-\theta\right) \xrightarrow{d} \mathcal{N}(0, A^{-1}BA^{-1}),$$

where $B \equiv \lim_{G\to\infty}\frac{1}{G}\sum_{g\le G} E[m_g(\theta;\delta)m_g(\theta;\delta)']$, with $m_g(\theta;\delta) \equiv s_g(\theta;\delta) + F_{g,0}r_g(\delta)$.

The components in asymptotic variance A, B can both be consistently estimated by plugging parameter estimates into their respective sample analogs. In our empirical application, we use bootstrap resampling methods to calculate the standard errors.

Recall that in the last step, the remaining structural parameters, i.e., the peer effect α, the contextual effects γ and the intercept β_0 are estimated by plugging $\hat{\theta}$ in the formulas in (7). Asymptotic variance of these structural parameters can be obtained by a direct application of the Delta Method.

4 Extensions

4.1 Group Fixed Effects in Outcomes

We extend our method to allow for unobserved group fixed effects in the outcomes. That is, the outcomes are given by:

$$Y = \alpha\overline{Y} + \beta_0 + X'\beta + \overline{X}'\gamma + \eta + U, \tag{8}$$

where η is an unobserved group fixed effect. As before, individual selection into the sample is based on Equation (1).

Using the same restrictions on (U, V) as Assumption 1 in Section 2, we get

$$Y = \tilde{\beta}_0 + X\beta + \overline{X}\tilde{\gamma} + \sigma_{uv}\lambda + \tilde{\sigma}_{uv}\overline{\lambda} + \tilde{\eta} + \varepsilon, \tag{9}$$

where $\tilde{\eta} \equiv \eta/(1-\alpha)$ and $(\tilde{\beta}_0, \tilde{\gamma}, \varepsilon)$ are defined as in (4) so that $E(\varepsilon|X, Z, \mathcal{S}) = 0$, which implies (X, Z) are exogenous in the sample. Without further restrictions, $\tilde{\eta}$ is generally correlated with (X, Z, U, V) and therefore ε.

There are several ways to estimate the parameters in (9), depending on the assumption on how η is correlated with the other variables. First, in the simplest

case, suppose η is assumed to be independent from (U, V, Z, X) and therefore ε. Then we can write $\tilde{\eta} + \varepsilon = E(\tilde{\eta}) + \tilde{\varepsilon}$ with $\tilde{\varepsilon} \equiv \varepsilon + \tilde{\eta} - E(\tilde{\eta})$, so that $E(\tilde{\varepsilon}|X, Z, \mathcal{S}) = 0$. By regressing y on $(1, X, \overline{X}, \lambda, \overline{\lambda})$ and using a necessary location normalization $E(\tilde{\eta}) = 0$, we can consistently estimate $(\tilde{\beta}_0, \beta', \tilde{\gamma}', \sigma_{uv}, \tilde{\sigma}_{uv})$. This in turn implies identification of $(\alpha, \beta_0, \gamma)$.

In a more general setting where η is correlated with (U, V, X, Z), one can estimate the model using additional instruments. That is, one can look for additional variables in the data available that are uncorrelated with η but are correlated with (X, Z) in the sense of satisfying instrument exogeneity and relevance. Using those instruments, one can consistently estimate the intercept $\tilde{\beta}_0$ and the slope coefficients $(\beta', \tilde{\gamma}', \sigma_{uv}, \tilde{\sigma}_{uv})$ in (9) through two-stage least squares.

Yet another option for estimating the model with correlated fixed effects is to parameterize the dependence between η and (U, V), and construct a correction term. In the next section, we elaborate on this solution in more general contexts.

4.2 Group Fixed Effects in Sample Selection

We now generalize the model in Section 4.1 by accommodating a second unobserved group fixed effect in the sample selection equation. Suppose individuals self-select into a group in the sample as follows:

$$D_i = 1\{Z_i\delta + \zeta + V_i \geq 0\}, \tag{10}$$

for $i \in \mathcal{N}$, where ζ is an unobserved group fixed effect. The structural form of outcomes within each group is the same as (8), which includes a group fixed effect η.

For convenience, define $V_i^* \equiv \zeta + V_i$ and $U_i^* = \eta + U_i$; let $V^* \equiv (V_i^*)_{i\in\mathcal{N}}$ and $Z \equiv (Z_i)_{i\in\mathcal{N}}$; let $U^* \equiv (U_i^*)_{i\leq n}$ and $X \equiv (X_i)_{i\leq n}$, where n denotes the number of actual group members i with $D_i = 1$. We maintain the following assumptions:

Assumption 1′. *(i) V is independent from (ζ, Z), and V_i are independently distributed as standard normal across potential group members. (ii) $(U_i^*)_{i\in\mathcal{N}}$ and V^* are independent from $(X_i)_{i\in\mathcal{N}}$ and Z. (iii) $E(U_i^*|V, \zeta) = E(U_i^*|V_i, \zeta) = \pi_1\zeta + \pi_2 V_i$.*

Conditions (i) and (ii) and the first equality in (iii) are analogous to the case with no fixed effects in Section 2. The second equality in (iii) holds if (U_i, V_i, η, ζ) is multivariate normal.

As in Section 4.1, plugging in the reduced form of $\overline{Y}$ gives:

$$Y = \tilde{\beta}_0 + X\beta + \overline{X}\tilde{\gamma} + E\left(U^* + \frac{\alpha}{1-\alpha}\overline{U}^* \middle| X, Z, \mathcal{S}\right) + \tilde{\varepsilon}, \tag{11}$$

where $E(\tilde{\varepsilon}|X, Z, \mathcal{S}) = 0$ by construction. We can estimate the model using the following steps:

Step 1. Use a *correlated random effect* model, as proposed in Chamberlain (1980), to estimate the selection equation in Equation (10). Specifically, let $F(\zeta|Z)$ denote the distribution of ζ conditional on Z, which is parameterized up to some unknown parameters. For example, following Chamberlain (1980), we may adopt the specification below for the fixed effects in group participation decisions:

Assumption 1′ (continued). *(iv) $\zeta = \overline{Z}\tau + e$, where $\overline{Z}$ is the average of individual Z_i's within the group, and $e \perp Z$ with $e \sim N(0, \sigma_e^2)$.*

Under Assumption 1′ (iv), the distribution $F(\zeta|Z)$ is parameterized up to (τ, σ_e). We estimate them jointly with δ using maximum likelihood:

$$(\hat{\delta}, \hat{\tau}, \hat{\sigma}_e) = \arg\max_{\delta,\tau,\sigma_e} \sum\nolimits_{g \le G} \log \int \prod\nolimits_{i \le n} \tilde{\Phi}_{g,i}(e; \delta, \tau)^{D_{g,i}} \left[1 - \tilde{\Phi}_{g,i}(e; \delta, \tau)\right]^{1-D_{g,i}} \frac{1}{\sigma_e}\phi\left(\frac{e}{\sigma_e}\right) de,$$

where $\tilde{\Phi}_{g,i}(e; \delta, \tau) \equiv \Phi(Z_{g,i}\delta + \overline{Z}_g\tau + e)$. Here we add subscripts g to index groups in the sample, and use $D_{g,i}, Z_{g,i}$ to denote the variables for a potential member i in group g.

Step 2. Apply a generalized method to correct the bias due to sample selection, using estimates for (δ, τ, σ_e) from the previous step.

To do so, let $\overline{\mathcal{S}}$ denote the event that "$V_j^* \ge -Z_j\delta$ for all $j \le n$ and $V_k^* < -Z_k\delta$ for all other $k \in \mathcal{N}$". Then note

$$E(U_i^*|X, Z, D = 1) \;=\; E(U_i^*|\overline{\mathcal{S}}) = \int E(U_i^*|\zeta, V_i \ge -Z_i\delta - \zeta) dF(\zeta|\overline{\mathcal{S}}), \tag{12}$$

where the second equality is due to Assumption 1′ (iii) above. The integrand on the right-hand side of (12) is:

$$\begin{aligned} E(U_i^*|\zeta, V_i \ge -Z_i\delta - \zeta) = & \int E(U_i^*|\zeta, V_i) dF(V_i|V_i \ge -Z_i\delta - \zeta) \\ = & \int (\pi_1\zeta + \pi_2 V_i)\, dF(V_i|V_i \ge -Z_i\delta - \zeta) \\ = & \; \pi_1\zeta + \pi_2\lambda(Z_i\delta + \zeta). \end{aligned}$$

Under Assumption 1′-(ii), (iii) and (iv), we can write the right-hand side of (12)

as:

$$\int \left[\pi_1(\overline{Z}\tau + e) + \pi_2\lambda(Z_i\delta + \overline{Z}\tau + e)\right] dF(e|\mathcal{S}^*).$$

where $\mathcal{S}^*$ denotes "$e + V_j \geq -Z_j\delta - \overline{Z}\tau$ for all $j \leq n$, and $e + V_k < -Z_k\delta - \overline{Z}\tau$ for all other $k \in \mathcal{N}$". Under conditions (i) and (iv) in Assumption 1', $e = \zeta - \overline{Z}\tau$ is independent from the vector of selection errors V. It then follows that

$$e \mid e + V_1 = t_1, e + V_2 = t_2, ...e + V_{(\#\mathcal{N})} = t_{(\#\mathcal{N})}$$

is normally distributed with variance $\tilde{\sigma}^2 \equiv \left[\sigma_e^{-2} + (\#\mathcal{N})\right]^{-1}$ and mean $\tilde{\sigma}^2(\sum_{i\in\mathcal{N}} t_i)$ (where $\#\mathcal{N}$ denotes the cardinality of $\mathcal{N}$). Therefore, we can write (12) in the form of

$$\pi_1 \underbrace{\int (\overline{Z}\tau + e) d\tilde{F}(e|Z;\theta)}_{\psi(Z)} + \pi_2 \underbrace{\int \lambda(Z_i\delta + \overline{Z}\tau + e) d\tilde{F}(e|Z;\theta)}_{\varphi_i(Z)},$$

where $\tilde{F}(e|Z;\theta)$ is the distribution of e conditional on $\mathcal{S}^*$. This conditional distribution is known up to (δ, τ, σ_e), which can be consistently estimated from Step 1. The quantities ψ, φ_i can be constructed using these estimates of (δ, τ, σ_e). Note φ_i varies across individual members in each group while ψ does not.

Step 3. Using the estimates from Steps 1 and 2, we can write the individual outcome Y_i in (11) as

$$Y_i = \tilde{\beta}_0 + X_i\beta + \overline{X}\tilde{\gamma} + \tilde{\pi}_1\psi + \pi_2\varphi_i + \tilde{\pi}_2\overline{\varphi} + \tilde{\varepsilon}_i,$$

where $\overline{\varphi}$ denotes the group mean of φ_i, and $\tilde{\pi}_1 \equiv \frac{\pi_1}{1-\alpha}$, $\tilde{\pi}_2 \equiv \frac{\alpha\pi_2}{1-\alpha}$. Let $\widetilde{W}_i \equiv (1, X_i, \overline{X}, \psi, \varphi_i, \overline{\varphi})$. Provided $E(\widetilde{W}_i'\widetilde{W}_i|\mathcal{S}^*)$ has a full rank, we can consistently estimate $(\tilde{\beta}_0, \beta', \tilde{\gamma}', \tilde{\pi}_1, \pi_2, \tilde{\pi}_2)$ by regressing Y_i on $\widetilde{W}_i$. This in turn allows us to construct consistent estimators for α, β_0 and γ as before.

5 Monte Carlo

We present two monte carlo experiments, with different sizes of potential groups, $\#\mathcal{N} = 10$ and $\#\mathcal{N} = 50$. For each group g and member i, let $X_{g,i}$ and $Z_{g,i}$, be two distinctive scalar variables, which are drawn from the standard normal distribution independently.[4] Let $(U_{g,i}, V_{g,i})$ be drawn from the bivariate normal

[4]With a slight abuse of notation, in this section we use $Z_{g,i}$ to denote the instrument variable that enters the selection equation, but not directly in the outcome equation.

with mean $(0,0)$, unit variance and covariance σ_{uv}. These error terms are independent across individuals and groups. The sample selection is governed by

$$D_{g,i} = 1\{\delta_0 + \delta_1 X_{g,i} + \delta_2 Z_{g,i} + V_{g,i} \geq 0\},\ i \in \mathcal{N}_g; g = 1, \cdots, G,$$

with true parameter being $\delta = (0,1,1)$. Among those individuals with $D_{g,i} = 1$, the outcomes are generated through the reduced form:

$$Y_{g,i} = \frac{\beta_0}{1-\alpha} + \beta X_{g,i} + \overline{X}_g \frac{\alpha\beta + \gamma}{1-\alpha} + U_{g,i} + \frac{\alpha}{1-\alpha}\overline{U}_g$$

We set the true parameters as $(\alpha, \beta_0, \beta, \gamma, \sigma_{uv}) = (1/2, 1, 1, 1, 2/3)$. We experiment with sample sizes $G = 250, 500, 1000, 2000$. We report average biases and mean-squared error (MSE) with 1,000 replications in Tables 1 and 2.

Table 1: Monte Carlo Results: #$\mathcal{N}$=10

G	Average Bias				
	α	β_0	β	γ	σ_{uv}
250	-0.040	0.112	0.001	0.141	-0.002
500	-0.015	0.043	0.000	0.055	0.001
1,000	-0.007	0.020	0.000	0.024	-0.001
2,000	-0.006	0.015	0.000	0.021	0.000
	MSE				
	α	β_0	β	γ	σ_{uv}
250	0.042	0.315	0.002	0.557	0.006
500	0.011	0.075	0.001	0.143	0.003
1,000	0.005	0.036	0.000	0.068	0.002
2,000	0.002	0.018	0.000	0.034	0.001

In Tables 1 and 2, both the average bias and MSE decrease at the same rate as the sample size increases. This confirms our asymptotic theory that the two-step estimator is root-G consistent. Convergence of the squared average bias at a rate faster than the increase in sample sizes indicates the dominant component in MSE is the estimator variance. Meanwhile, the size of groups does not have an obvious impact on estimation precision, especially in larger samples.

Table 2: Monte Carlo Results: #$\mathcal{N}$=50

G	Average Bias				
	α	β_0	β	γ	σ_{uv}
250	-0.032	0.088	0.000	0.113	0.000
500	-0.014	0.039	0.000	0.054	0.000
1,000	-0.006	0.016	0.000	0.022	0.001
2,000	-0.003	0.009	0.000	0.011	0.000
	MSE				
	α	β_0	β	γ	σ_{uv}
250	0.026	0.199	0.000	0.343	0.001
500	0.010	0.077	0.000	0.135	0.001
1,000	0.004	0.033	0.000	0.056	0.000
2,000	0.002	0.015	0.000	0.026	0.000

6 Peer Effects in Online Training

In this section, we apply the model to estimate peer effects in a large online teacher training program in China, known as the Young Teacher Empowerment Program (YTEP). [5] The YTEP is an annual training program designed to boost the morale and improve the skills of young teachers in elementary and middle schools in rural China. To participate in the YTEP program, applicants must be chosen by participating rural schools and the education bureau in the local county government coordinating the training.

Our data was collected from the Training Year of 2019-2020, which consists of two semesters (Fall 2019 and Spring 2020). The YTEP consists of two phases, mandatory general courses in Fall 2019 and elective field courses in Spring 2020. All trainees are automatically enrolled in two mandatory courses, *Career Development* and *Teacher Ethics*. We investigate how peer effects affect the trainees' lecture attendance in these mandatory courses, which have a much larger number of participants than elective field courses.

The sample contains 8,627 trainees across 63 counties in 17 provinces of China. The career development and teacher ethics courses have 17 and 12 independent and synchronous lectures, respectively. Instructors and contents differ across lectures in each course. We pool 29 lectures in the sample. For each lecture, trainees first decide whether to attend the lecture and then decide how long to

[5] YTEP details: http://www.youcheng.org/news_detail.php?id=645

stay. We define a potential group by a pair of county and lecture. That is, all teachers from a county are treated as potential group members for a specific lecture. We model the first-step decision to attend the lecture as self-selection into participation according to Equation (1). We model the duration of lecture attendance, measured by the number of minutes a trainee spent in a lecture, as the outcomes determined in the second stage governed by Equation (2).

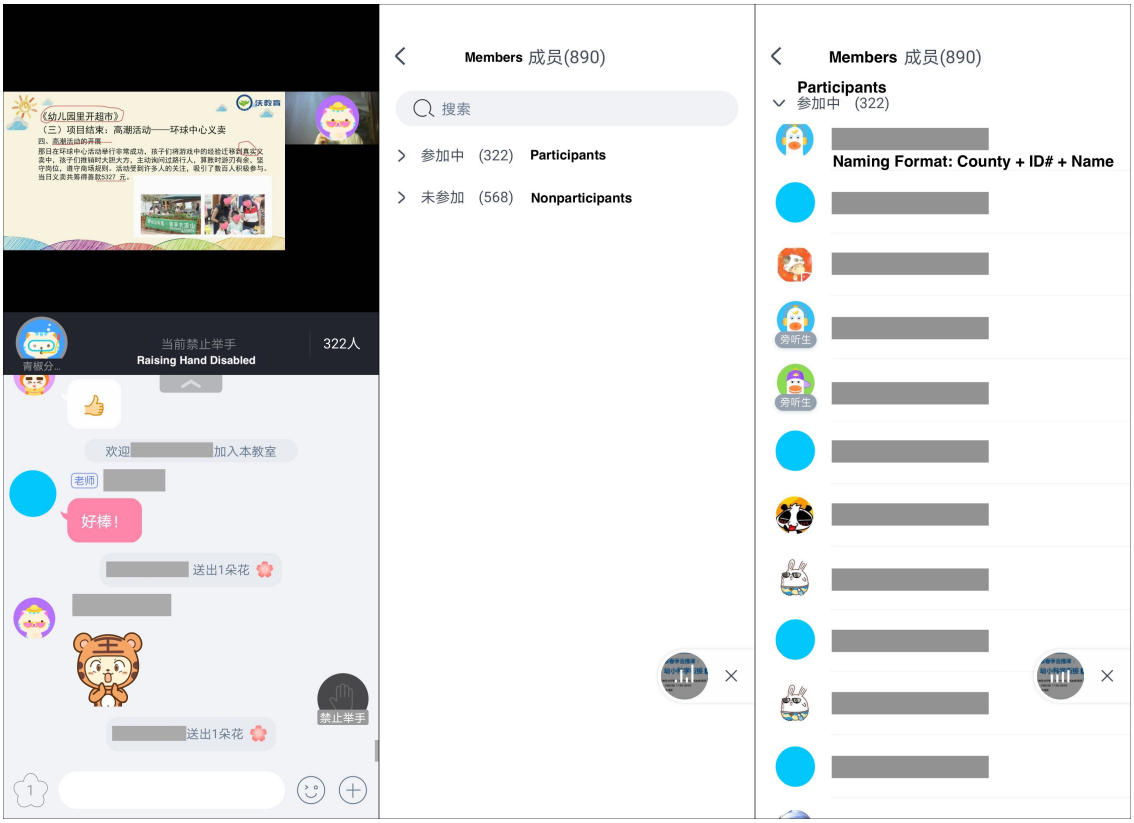

Figure 1: Interface of Instructional Platform

The lectures are held synchronously online during the evening on a weekly basis. Figure 1 shows the user interface of the instructional platform. Upon entering the meeting room, a lecture participant can watch the live broadcast of the lecture. Below the presentation window, there is a chatroom where participants can communicate with the instructor, TAs, and other trainees. More importantly, the participants can observe the total number of other participants and nonparticipants (those who enrolled in the course but did not attend this particular lecture) as well as the detailed list of participants. The list refreshes

whenever there is an entry or exit.

Each trainee has a unique identifier whose naming format is "*County* + *ID#* + *Name*". The list is sorted by the characters of an identifier, and participants from the same county are placed adjacently. Thus, a lecture participant could easily observe peers from the same county sitting in the lecture, generating potential peer effects. In addition, each county has a coordinator who helps the program administration and communicates with the trainees. County coordinators are usually local education administrators. They organize trainees from the same county into a group via online social platforms, such as an online WeChat group.[6] The online platform is used to send program announcements and facilitates communication between trainees from the same county. Therefore, all trainees from the same county naturally form a potential peer group.

We define a (potential) peer group as a specific county-lecture combination in the two mandatory courses. The size of an actual group is equal to the number of trainees from a county who actually attended the particular lecture. These groups are indexed by g, as in Section 3. For each group g, the set of *potential* participants, denoted by N_g in Section 3, is defined as all trainees from that county who have enrolled in the courses. For each trainee attending a lecture, we observe the duration of her or his attendance.

Additionally, information for trainees is obtained from their registration records, as well as their responses to routine program surveys during the training period (designed to collect feedback). The program surveys consist of two waves, one at the end of the first semester and the other upon completion of the program at the end of the second semester. Each wave has a response rate of about 40%. The registration and survey provide demographic features about the trainees and characteristics of their school and the county. We impute missing values of the school and county characteristics based on their formation available.[7] Table 3 summarizes the related variables (variables with no designated units are dummies).

[6]WeChat is a popular Chinese instant messaging smartphone application, similar to WhatsApp.

[7]For school sizes, we use the median of values self-reported by survey respondents from the same school to replace their original answers and impute the variable for non-respondents from the same school. For the dummy variables at the county level, we use the mode to do so. Missing values are assigned in the imputation if there are multiple modes.

Table 3: Summary Statistics

Variable	Obs	Mean	SD	Min	Max
Selection Stage					
Lecture Attendance	250,183	0.404	0.491	0	1
Outcome Stage					
Participation Duration (mins)	100,954	65.083	24.120	1	165
Personal Characteristics					
Married	5,357	0.295	0.456	0	1
Teaching Experience (yrs)	5,357	2.145	3.030	0	37
Slow Internet Speed	5,357	0.200	0.400	0	1
Gender (male)	8,573	0.230	0.421	0	1
Ethnicity (han)	7,667	0.671	0.470	0	1
Party Affiliation	7,533	0.578	0.494	0	1
Bachelor (or above)	8,233	0.800	0.400	0	1
Teachers College	8,135	0.715	0.451	0	1
Tenured Teacher	8,452	0.452	0.498	0	1
Village School	8,453	0.415	0.493	0	1
School Size (number of students)	7,237	847.761	1,031.370	4	10,001
County Characteristics					
Impoverished County	8,475	0.884	0.320	0	1
Encouraging County Coordinator	8,000	0.255	0.436	0	1
Other Statistics					
County Enrollment	63	136.778	157.221	5	1,020
Group Size	1,517	48.238	81.738	2	797

Note: Married, Teaching Experience, Slow Internet Speed, School Size, Impoverished County, and Encouraging County Coordinator are reported in a program survey. Other characteristics are reported in the program registration. For missing values, School Size uses the median of respondent self-reported values for the variable imputation at the school level. Impoverished County and Encouraging County Coordinator use the mode for the imputation at the county level. Missing values are assigned in the imputation if there are multiple modes.

We apply our multi-step method to estimate the peer and contextual effects on the trainees' duration of lecture attendance, while accounting for self-selection in lecture participation. Our group definition has two overlays, namely county and lecture. Thus, self-selection may occur at both levels, joining a school/county upon employment and attending a lecture after training course enrollment. In the first (participation/selection) stage, we model the self-selection into the group

with the following specification:

$$Participation_{g,i} = 1\{Z_i'\delta + V_{g,i} \geq 0\}, \tag{13}$$

where Z_i contains personal and school/county characteristics.

More specifically, among trainees, males account for 23%, and nearly 30% of them are married. 80% of them obtained a bachelor's degree or above with about 70% graduating from teachers colleges. Trainees are mostly young teachers with about 2 years of teaching experience on average, and over 45% of the trainees hold tenured positions. Those characteristics and their ethnicity and party affiliation may affect their incentive to join a particular school/county and their participation in a training lecture. Some variables may capture constraints they face in attending a lecture.

Additionally, approximately 20% of the trainees report slow internet speed, which could deter their participation. More than 40% of the trainees work at rural village schools, and 88% of the schools are in impoverished counties. Therefore, those trainees may have a higher incentive to join a lecture than their urban counterparts. We also include the behavior of county coordinators, because such behavior can generate common shocks for the group.

For the variables that affect the participation but not the duration of the lecture, we use the marital status dummy, *Married*, and its interaction with other covariates. [8] It is likely that the decision of finding a job in a particular county as a teacher is affected by marital status. Moreover, to attend a lecture in the evening generally requires planning and coordinating with the spouses. Therefore, this variable affects the participation but to a much lesser degree on how long to stay in a lecture.

The first-stage estimation results in (13) are reported in Table A1. The signs of coefficient estimates in this probit model are also largely consistent with intuition. For instance, slow internet connectivity imposes significant restrictions on lecture participation. Males are less likely than their female peers to attend a lecture. Tenured trainees are less active than their untenured colleagues in lecture participation. Working at rural village schools incentivizes individuals to join a lecture with a higher probability, while employment in an impoverished county

[8]The F-statistic for testing the joint significance of all interaction terms involving *Married* is large enough to suggest these instruments are strong. Given that the other covariates included in the structural form for the duration of attendance are assumed to be exogenous, these interaction terms also satisfy the exclusion requirement for the structural model.

limits the chance of lecture attendance. Having a coordinator who encourages participation enhances the likelihood of participation. The results show that *Married* is statistically significant in the decision to attend a lecture. Furthermore, *Married* and its interaction terms with other covariates are jointly significant in this participation stage. These results indicate that the IV relevance condition holds for our choices of instrumental variables.

In the second (outcome) stage, we adopt the following specification to estimate the reduced form based on (6) in the theoretical part:

$$Duration_{g,i} = \tilde{\beta}_0 + X'_{g,i}\beta + \overline{X}'_g\tilde{\gamma} + \sigma_{uv}\lambda_{g,i} + \tilde{\sigma}_{uv}\bar{\lambda}_g + \varepsilon_{g,i}, \tag{14}$$

where $X_{g,i}$ contains all individual demographics in Table 3 except *Married*, which is used as the instrument discussed above. We denote g as a peer group defined by the county-lecture pair and i as trainees who self-select into the group g.

Our method for identification and estimation applies to data-generating processes (DGPs) where the number of groups (county-lecture combinations) is large relative to the size of each group. To allow for contextual effects, we restrict our groups to have at least two trainees from the same county attending the same lecture. In the data, we have 63 counties and 29 lectures, so a total of 1,827 county-lecture pairs. After excluding the pair with fewer than 2 participants, we have 1,517 groups with an average group size of 48 in our sample for estimation. The max group size is 797, still relatively smaller than the number of groups. Therefore, our requirement for the DGP is satisfied.

The vector $\overline{X}'_g$ consists of the group averages of $X_{g,i}$ with each county-lecture combination, as well as the county characteristics. The variable $\lambda_{g,i}$ is the inverse Mills ratio constructed from the estimates in Equation (13), and $\bar{\lambda}_g$ is the group average of $\lambda_{g,i}$ within a county-lecture pair. The inclusion of $\lambda_{g,i}$ and $\bar{\lambda}_g$ helps us deal with two sources of endogeneity at the same time, self-selection into participation and simultaneity in the determination of structural outcomes. Specifically, the reduced form contains an individual structural error and a composite error including the structural errors of all group members. To correct for sample selection, $\lambda_{g,i}$ takes care of the correlation between V_i and U_i, and $\bar{\lambda}_g$ addresses the correlation between V_i and $\overline{U}$. Moreover, the presence of $\lambda_{g,i}$ and $\bar{\lambda}_g$ in the reduced form provides additional variations for structural form identification and thus solves the reflection problem.

The OLS estimates and standard errors are reported in Table A2. The standard

errors are estimated using bootstrap resampling on the groups with replacement for 1,000 replications. Generally, individual characteristics have statistically significant effects on the duration of lecture attendance in the reduced form, with signs that are consistent with intuition. For instance, slow internet speed has a negative impact on attendance duration. The ratio of group members having difficulty connecting to the internet also negatively affects the length of attendance significantly at 1% level. Male participants stay for shorter periods than female participants. Also, we find evidence that the gender composition of peers in the same lecture has a significant reduced-form effect in this online setting. Analogous effects were reported for face-to-face environments in (Hoxby, 2000, Hill, 2017, Gong, Lu, and Song, 2021). The statistical significance of $\lambda_{g,i}$ and $\bar{\lambda}_g$ indicates a high correlation between the error terms in the selection stage and those in the outcome stage.

By plugging the OLS estimates of the reduced-form parameters from Table A2 into (7), we report the estimates of structural parameters in Table 4. The standard errors are estimated using bootstrap resampling on the groups with replacement for 1,000 replications. The estimate of peer effects, the parameter α in Equation (2), is 0.842, and the coefficient is statistically significant at the 1% level. It implies that a 10-minute increase in the peer group average duration of attendance leads to an 8.42-minute increase in own participation. Our estimate of peer effects is within the range of (0, 1), which is comparable with those reported for other contexts in the literature (Calvó-Armengol, Patacchini, and Zenou, 2009, Bramoullé, Djebbari, and Fortin, 2009, Lin, 2010).

The contextual effects of some variables, such as *SlowInternet*, *Gender*, and *Bachelor*, are statistically not significant in the structural form. For other characteristics, the direct, own effect (β) and the corresponding contextual effect (γ) are both significant with opposite signs. While the sign of the direct effect often conforms with intuition, the sign of the contextual effect depends on whether trainees with certain characteristics are substitutes or complements in the participation decisions (Blume, Brock, Durlauf, and Jayaraman, 2015).

For comparison, we also estimate the model of peer effects in Equation (2) without taking into account the sample selection in Equation (1). However, in this case, without exogenous variation from instruments in the selection equation, we need other exclusion restrictions to solve the reflection problem.[9] In particular

[9]We have $2K+2$ parameters in the structural form (2). Neglecting sample selection would lead

we exploit a variable that has direct effect on an individual's own duration of attendance, but has no contextual effect on others' in the structural form. Such an exclusion restriction is known to help solve the reflection problem in the social interactions model (Manski, 1993, Moffitt, 2001). We can not use the instrument in the selection framework, i.e., whether a trainee is married (*Married*), for this purpose. This is because *Married* is already excluded from the structural form of the equation that determines the duration of attendance.

Instead, we posit the internet speed has a direct effect on an individual's own outcome (duration of lecture attendance) but has no contextual effect on others' outcomes. The latter property means an individual's duration of attendance is not immediately affected by the proportion of peers who have slow internet access. This is confirmed by results in Table 4, which suggest that the proportion of group peers with slow internet has no significant contextual effect on an individual's duration of attendance. Hence, to estimate the peer effects in Equation (2) without taking into account the selective participation in Equation (1), we exploit this exclusion restriction on the internet speed variable. [10]

Formally, let I denote the vector of dummy variables indicating slow internet speed for each member in a group, and let X denote the vector of all other covariates. The structural form is

$$Y = \alpha\overline{Y} + \beta_0 + X'\beta + \overline{X}'\gamma + \beta_I I + U, \tag{15}$$

which implies the following reduced form:

$$Y = \tilde{\beta}_0 + X'\beta + \overline{X}'\tilde{\gamma} + \beta_I I + \tilde{\gamma}_I \bar{I} + \widetilde{U}, \tag{16}$$

where $\widetilde{U}, \tilde{\beta}_0, \tilde{\gamma}$ are defined as in Equation (3) and $\tilde{\gamma}_I = \frac{\alpha\beta_I}{1-\alpha}$. With covariates in Equation (16) satisfying the rank condition for OLS, we identify $\tilde{\beta}_0, \beta, \tilde{\gamma}, \beta_I, \tilde{\gamma}_I$. We can then identify all structural parameters in Equation (15) by

$$\alpha = \frac{\tilde{\gamma}_I}{\beta_I + \tilde{\gamma}_I};\ \beta_0 = (1-\alpha)\tilde{\beta}_0;\ \gamma = (1-\alpha)\tilde{\gamma} - \alpha\beta. \tag{17}$$

to a reduced form that drops λ_i and $\bar{\lambda}$ in Equation (6) and thus leave us with $2K+1$ parameters to estimate. Hence, we cannot identify $2K+2$ structural parameters from $2K+1$ estimated coefficients of the reduced form.

[10] Apart from internet speed, two other variables, *Gender* and *Bachelor*, also show no statistically significant contextual effects in Table 4. We did not consider them as candidates satisfying exclusion restrictions, because the literature has documented evidence of gender peer effects (Hoxby, 2000, Hill, 2017, Gong, Lu, and Song, 2021) and education contextual effects (Harmon, Fisman, and Kamenica, 2019, Laliberté, 2021).

For estimation, we regress $Y_{g,i}$ on $X_{g,i}, \overline{X}_g, I_{g,i}, \bar{I}_g$ and an intercept, using the same sample for the model with sample selection. We then use Equation (17) to recover all parameter estimates in the structural form.

Table 4: Estimates of Social Effects in Structural Equation

Variable	Estimate	Standard Error
Peer Effects	0.842^{***}	0.027
Intercept	17.967^{***}	1.776
Slow Internet Speed	-1.074^{***}	0.009
Gender (male)	-0.708^{***}	0.012
Teaching Experience (yrs)	0.065^{***}	0.001
Teachers College	-0.381^{***}	0.007
Tenured Teacher	1.273^{***}	0.011
Bachelor (or above)	-0.184^{***}	0.008
Ethnicity (han)	1.382^{***}	0.008
Party Affiliation	-0.598^{***}	0.007
Village School	0.179^{***}	0.008
School Size	-0.001^{***}	0.000
Average Slow Internet Speed	-0.005	0.221
Average Gender (male)	0.310	0.369
Average Teaching Experience	-0.141^{***}	0.014
Average Teachers College	0.757^{***}	0.168
Average Tenured Teacher	-0.442^{***}	0.026
Average Bachelor (or above)	0.021	0.080
Average Ethnicity (han)	-0.900^{***}	0.081
Average Party Affiliation	0.812^{***}	0.076
Average Village School	-0.376^{***}	0.024
Average School Size	0.001^{***}	0.000
Impoverished County	1.717^{***}	0.063
Encouraging County Coordinator	-0.203^{***}	0.017

The standard errors are estimated by bootstrap resampling on the groups with replacement for 1,000 replications. Significance Level: *** 1%, ** 5%, * 10%.

Table A3 in the appendix shows the estimates of the reduced form in Equation (16); Table 5 shows estimates of structural parameters using (17). The

results show that ignoring the sample selection in participation, we estimate the peer effects to be 0.813. However, this estimated effect is statistically insignificant at the 10% level.

The no-selection model above is the closest alternative to our selection model and thus the results from both models are comparable. Firstly, if we write I and $\bar{I}$ into X and $\overline{X}$ in Equation (16), the specification becomes identical to the selection model in Equation (6) when dropping the selection terms λ_i and $\bar{\lambda}$. Secondly, the sample used in estimating model (6) and (16) are the same. Therefore, the difference in the estimation is caused by the sample selection in the decision to participation in lectures.

The distinction between the peer effect estimates in Table 4 and Table 5 illustrates the consequence of failing to account for sample selection bias due to endogenous lecture participation. Specifically, ignoring the sample selection issue would result in an erroneous conclusion that the peer effects are statistically insignificant in the empirical context we consider. On the other hand, once sample selection is properly taken into account and addressed, the peer effects turn out to be statistically significant in this context.

7 Conclusion

This paper studies the estimation of peer effects in self-selected groups that are formed out of endogenous individual participation decisions. We correct the sample selection bias using individual instruments that affect the participation decisions, but do not directly affect the outcomes. In the context of social interactions, dealing with the sample selection issue requires insertion of both individual and group-level composite correction terms. The inclusion of this latter, group-level correction term provides additional sources of exogenous variation that help us to resolve the reflection problem. We apply our method to study peer effects in an online teacher training program in China, where the trainees endogenously decide to participate in online lectures. The outcome of interest is the duration of attendance in each lecture. We find significant peer effects among trainees in the same lecture after accounting for sample selection bias. On the other hand, ignoring the sample selection issue in this context would result in an erroneous conclusion about the magnitude and significance of the peer effects.

Table 5: Structural Estimates With No Sample Selection

Variable	Estimate	Standard Error
Peer Effects	0.813	1.526
Intercept	12.697	98.263
Slow Internet Speed	-1.491^{***}	0.007
Gender (male)	-1.722^{***}	0.007
Teaching Experience (yrs)	0.084^{***}	0.001
Teachers College	-0.006	0.006
Tenured Teacher	0.498^{***}	0.009
Bachelor (or above)	-0.227^{***}	0.008
Ethnicity (han)	1.213^{***}	0.007
Party Affiliation	-0.411^{***}	0.007
Village School	0.398^{***}	0.007
School Size	-0.001^{***}	0.000
Average Slow Internet Speed	–	–
Average Gender (male)	-0.118	18.234
Average Teaching Experience	-0.156	0.263
Average Teachers College	0.760	2.373
Average Tenured Teacher	-0.473	1.988
Average Bachelor (or above)	0.029	4.048
Average Ethnicity (han)	-0.918	2.679
Average Party Affiliation	0.894	4.547
Average Village School	-0.403	1.669
Average School Size	0.001	0.001
Impoverished County	-0.505	3.405
Encouraging County Coordinator	0.104	0.331

References

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear social interactions models. *Journal of Political Economy 123*(2), 444–496.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

Brock, W. A. and S. N. Durlauf (2001). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies 47*(1), 225–238.

Dahl, G. B., K. V. Løken, and M. Mogstad (2014). Peer effects in program participation. *American Economic Review 104*(7), 2049–74.

De Giorgi, G., M. Pellizzari, and S. Redaelli (2010). Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics 2*(2), 241–75.

Glaeser, E. L., B. Sacerdote, and J. A. Scheinkman (1996). Crime and social interactions. *The Quarterly Journal of Economics 111*(2), 507–548.

Gong, J., Y. Lu, and H. Song (2021). Gender peer effects on students' academic and noncognitive outcomes evidence and mechanisms. *Journal of Human Resources 56*(3), 686–710.

Graham, B. S. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica 76*(3), 643–660.

Gronau, R. (1974). Wage comparisons–a selectivity bias. *Journal of Political Economy 82*(6), 1119–1143.

Harmon, N., R. Fisman, and E. Kamenica (2019). Peer effects in legislative voting. *American Economic Journal: Applied Economics 11*(4), 156–80.

Heckman, J. (1974). Shadow prices, market wages, and labor supply. *Econometrica 42*(4), 679–694.

Heckman, J. J. (1978). Dummy endogenous variables in a simultaneous equation system. *Econometrica 46*(4), 931–959.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Hill, A. J. (2017). The positive influence of female college students on their male peers. *Labour Economics 44*, 151–160.

Hong, H., J. D. Kubik, and J. C. Stein (2004). Social interaction and stock-market participation. *The Journal of Finance 59*(1), 137–163.

Hoxby, C. M. (2000). Peer effects in the classroom: Learning from gender and race variation.

Laliberté, J.-W. (2021). Long-term contextual effects in education: Schools and neighborhoods. *American Economic Journal: Economic Policy 13*(2), 336–77.

Lee, L.-F. (2007). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 140*(2), 333–374.

Lin, X. (2010). Identifying peer effects in student academic achievement by spatial autoregressive models with group unobservables. *Journal of Labor Economics 28*(4), 825–860.

Lin, Z. and X. Tang (2022). Solving the reflection problem in social interactions models with endogeneity. *Working paper*.

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

Manski, C. F. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives 14*(3), 115–136.

Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions. *Social Dynamics 4*(45-82), 6–17.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Sacerdote, B. (2001). Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics 116*(2), 681–704.

Sheng, S. and X. Sun (2021). Identification and estimation of social interactions in endogenous peer groups. *Working paper, UCLA and Simon Frasier U*.

Topa, G. (2001). Social interactions, local spillovers and unemployment. *The Review of Economic Studies 68*(2), 261–295.

Trogdon, J. G., J. Nonnemaker, and J. Pais (2008). Peer effects in adolescent overweight. *Journal of Health Economics 27*(5), 1388–1399.

Appendix

Table A1: Selection Stage Probit Estimation Results

Variable	Estimate	Standard Error
Intercept	0.344***	0.021
Slow Internet Speed	−0.048***	0.010
Gender (male)	−0.131***	0.010
Teaching Experience (yrs)	−0.004	0.003
Teachers College	0.032***	0.009
Tenured Teacher	−0.102***	0.009
Bachelor (or above)	0.011	0.010
Ethnicity (han)	−0.021**	0.009
Party Affiliation	0.045***	0.009
Village School	0.047***	0.010
School Size	0.000	0.000
Impoverished County	−0.359***	0.013
Encouraging County Coordinator	0.052***	0.010
Teaching Experience Squared	0.000	0.000
School Size Squared	0.000	0.000
Married	−0.099***	0.037
Slow Internet Speed × Married	−0.020	0.019
Gender (male) × Married	−0.018	0.019
Teaching Experience × Married	0.004	0.003
Teachers College × Married	0.053***	0.016
Tenured Teacher × Married	−0.007	0.016
Bachelor (or above) × Married	−0.040**	0.020
Ethnicity (han) × Married	−0.000	0.017
Party Affiliation × Married	−0.066***	0.015
Village School × Married	−0.072***	0.017
School Size × Married	0.000	0.000
Impoverished County × Married	0.171***	0.022
Encouraging County Coordinator × Married	−0.054***	0.017

Wald Test on IVs (Married and its interactions): $\chi^2 = 144.5$.

Significance Level: *** 1%, ** 5%, * 10%.

Table A2: Results of Outcome Stage Regression
(Dependent variable: lecture attendance in minutes)

Variable	Estimate	Standard Error
Intercept	113.705***	0.572
Slow Internet Speed	−1.074***	0.009
Gender (male)	−0.708***	0.012
Teaching Experience (yrs)	0.065***	0.001
Teachers College	−0.381***	0.007
Tenured Teacher	1.273***	0.011
Bachelor (or above)	−0.184***	0.008
Ethnicity (han)	1.382***	0.008
Party Affiliation	−0.598***	0.007
Village School	0.179***	0.008
School Size	−0.001***	0.000
Average Slow Internet Speed	−5.751***	0.119
Average Gender (male)	−1.811***	0.154
Average Teaching Experience	−0.550***	0.005
Average Teachers College	2.764***	0.081
Average Tenured Teacher	3.983***	0.070
Average Bachelor (or above)	−0.848***	0.067
Average Ethnicity (han)	1.670***	0.047
Average Party Affiliation	1.957***	0.043
Average Village School	−1.426***	0.041
Average School Size	−0.000	0.000
Impoverished County	10.864***	0.168
Encouraging County Coordinator	−1.286***	0.032
$\hat{\lambda}$	−11.840***	0.110
$\hat{\hat{\lambda}}$	−63.094***	0.929

Note: the standard errors are estimated by bootstrap resampling on the groups with replacement for 1,000 replications. Significance Level: *** 1%, ** 5%, * 10%.

Table A3: OLS Estimation Results With No Sample Selection

Variable	Estimate	Standard Error
Intercept	67.838***	1.126
Slow Internet Speed	−1.491***	0.228
Gender (male)	−1.722***	0.233
Teaching Experience (yrs)	0.084**	0.036
Teachers College	−0.006	0.203
Tenured Teacher	0.498*	0.281
Bachelor (or above)	−0.227	0.250
Ethnicity (han)	1.213***	0.228
Party Affiliation	−0.411*	0.219
Village School	0.398*	0.239
School Size	−0.001***	0.000
Average Slow Internet Speed	−6.474***	1.460
Average Gender (male)	−8.111***	1.342
Average Teaching Experience	−0.472***	0.066
Average Teachers College	4.037***	0.847
Average Tenured Teacher	−0.363	0.427
Average Bachelor (or above)	−0.829	0.796
Average Ethnicity (han)	0.365	0.439
Average Party Affiliation	2.990***	0.476
Average Village School	−0.426	0.464
Average School Size	0.001***	0.000
Impoverished County	−2.696***	0.275
Encouraging County Coordinator	0.556**	0.228

Significance Level: *** 1%, ** 5%, * 10%.