Social Interactions with Endogeneity*

Zhongjian Lin
University of Georgia

Xun Tang
Rice University

August 30, 2023

Abstract

We identify and estimate peer and contextual effects in social interactions models with endogenous covariates (e.g., self-selected treatments). Our method uses individual instruments for endogenous covariates, but does *not* require additional instruments for simultaneity in outcomes, which are often hard to find in models with contextual effects. The method can be applied to relax the Stable Unit Treatment Value Assumption (SUTVA) in program evaluation, allowing individual treatments to influence the outcomes of others through peer and contextual effects. We apply our method to estimate peer effects in Grade 3 math scores of elementary school students in the State of Tennessee. Using lagged class sizes and teacher qualification as instruments for Grade 2 scores, we find significant evidence for positive peer effects and path dependence on G2 scores.

Keywords: Social Interactions, Endogeneity, Reflection Problem, Control Functions, Peer Effects, Academic Achievement.

JEL Codes: C31, C36, I21.

*We are grateful to Yingyao Hu, Kevin Lang, Ruixuan Liu, Xiaodong Liu, Áureo de Paula, Essie Maasoumi, Shuyang Sheng, Zhentao Shi and Matthew Shum for helpful comments. Xun Tang acknowledges financial support from National Science Foundation (SES-1919489). All errors are our own.

1 Introduction

Social interactions models have been used in a wide variety of environments for studying how members within a group influence each other's outcomes. Examples include Gaviria and Raphael (2001) on juvenile behavior; Hoxby (2000), Sacerdote (2001), Zimmerman (2003), Angrist and Lang (2004), Calvó-Armengol, Patacchini, and Zenou (2009), Carrell, Fullerton, and West (2009), Lavy, Paserman, and Schlosser (2012), Burke and Sass (2013) and Ross and Shi (2021) on students' academic achievements; Katz, Kling, and Liebman (2001), Sampson, Morenoff, and Gannon-Rowley (2002), Durlauf (2004), Kling, Liebman, and Katz (2007), and Chetty, Hendren, and Katz (2016) on neighborhood effects; Trogdon, Nonnemaker, and Pais (2008) on adolescent overweight; Bayer, Hjalmarsson, and Pozen (2009) on juvenile corrections; Bramoullé, Djebbari, and Fortin (2009) on recreational activities; Bollinger and Gillingham (2012) on diffusion of products; Waldinger (2012), Cornelissen, Dustmann, and Schönberg (2017) on productivity; Ahern, Duchin, and Shumway (2014) on risk aversion and trust; Bursztyn, Ederer, Ferman, and Yuchtman (2014) on financial decisions; Dahl, Løken, and Mogstad (2014) on paternity leave program participation, etc.

A popular specification for social interactions models takes a linear-in-means (LIM) form, in which each individual outcome is linear in the average group outcome, the individual's own characteristics, and possibly the average characteristics of group members. Manski (1993) specified such a LIM social interactions model, and used the term *contextual* effects to capture how an individual's characteristics directly impact other members' outcomes, and *endogenous peer* effects to reflect a structural simultaneity between member outcomes within a group. He showed that, without further restrictions such as homoskedastic errors or additional sources of exogenous variation, the peer and contextual effects can not be disentangled from the model's reduced form. This non-identification issue is commonly known as the "reflection problem". Manski (1993) proposed a solution to the reflection problem (in Proposition 2), which identified the peer effects using an exclusion restriction that some individual characteristics have a direct effect on an individual's own outcome but no contextual effect on others' outcomes. Such exclusion restrictions arise in some empirical contexts, but are not easy to motivate in others.

In this paper we propose a solution to the reflection problem that exploits individual-level instruments for endogenous covariates. Our method is related to the insight from Manski (1993) in the following sense. We use instruments to construct individual-specific control functions (CFs) to deal with endogeneity in the covariates in the structural form. Our main idea is that these control functions essentially function as generated regressors which satisfy the exclusion restriction in Manski (1993). Thus, we do not need to invoke further assumptions for identifying peer effects. Most importantly, our method does not require additional instruments for dealing with the simultaneity in individual outcomes.

Empirical studies either focused on settings where the contextual effects are absent, or chose to infer some form of a composite effect that combines peer and contextual effects. (See our discussion of the literature in Section 5.) At the same time, the econometrics literature have provided solutions for the reflection problem in many ways. Moffitt (2001) used an alternative exclusion restriction (that a randomly assigned policy variable affects some but not all individuals in a group) to identify a LIM social interactions model. Brock and Durlauf (2001a) and Blume, Brock, Durlauf, and Ioannides (2011) showed that a necessary condition for resolving this identification problem, in the absence of further assumptions or sources of exogenous variation, is to require at least one characteristic whose group-level average has no contextual effect.[1]

[1]Graham and Hahn (2005) made a similar observation in a different information setting where the actual group-level average of characteristics is replaced by its expectation conditional on common

Blume, Brock, Durlauf, and Ioannides (2011) pointed out that finding valid instrument for the simultaneity in the structural form is difficult, and "most likely requires the development of an auxiliary model of x_i (individual characteristics)".

Several papers used the second moments of observed individual outcomes to identify the peer effects. Graham (2008) exploited second-moment restrictions on unobserved errors to identify a social multiplier, which is a composite of peer and contextual effects. (In the presence of contextual effects, these social multipliers are reduced to peer effects.) Sacerdote (2001) identified endogenous peer effects, using the second moments of individual outcomes, as well as the homoskedasticity and uncorrelation of structural errors, which are assumed to be independent from explanatory variables. Liu (2017) proposed a root-estimator to estimate peer and contextual effects in a linear-in-means social interactions model without group size variation, exploiting the variance of heteroskedastic, uncorrelated structural errors that are independent from the regressors. In the context of social networks, Rose (2017) identified peer effects through the covariance of outcomes, assuming the structural errors are homoskedastic, uncorrelated, and independent from the regressors. Lee (2007a) identifies peer and contextual effects by exploiting exclusion restrictions (some covariates have no contextual effects) as well as group size variation. For models with general network structures, Bramoullé, Djebbari, and Fortin (2009), De Giorgi, Pellizzari, and Redaelli (2010), Lee, Liu, and Lin (2010), Lin (2010), Liu and Lee (2010) identified peer and contextual effects, using features derived from network structures as instruments. Such instruments are not available in the linear-in-means specification we consider.[2]

The method we propose in this paper is related to these papers above in the sense

information.

[2]Formally, this is because in a LIM social interactions model, the matrix $[X, GX, G^2X]$ (with X being individual covariates and G being a network that assigns equal weights to all peers) does not satisfy the rank condition for identification.

that we also exploit restrictions on the covariance of structural errors. In our case, the restrictions are invoked on the errors in the outcome equation and the auxiliary equation that models endogeneity. Our method does not use the second moment of observed outcomes, and is designed to solve the identification question in the presence of endogenous regressors. We do so by leveraging the identifying power from additional instruments on the individual level.

Several papers have studied social interactions models with selection bias, using exogenous instruments for identification. Brock and Durlauf (2001b) (Section 3.6) showed how to identify peer and contextual effects when each member's decision to join a group is determined in a Probit selection stage. They used instruments from the selection stage, and corrected the selection bias as in Heckman (1979). Ioannides and Zabel (2008), Hoshino (2019) also dealt with the sample selection issue in LIM social interactions models in the contexts of neighborhood housing demands and student friendships, respectively. Sheng and Sun (2021) estimated a social interactions model where endogenous group formation arises from a many-to-one matching model.

We contribute to the literature by resolving the reflection problem in a model where some individual characteristics (such as self-selected treatment) are endogenous. Unobserved individual characteristics are correlated with such endogenous covariates, and affect the peer outcomes of other group members in the reduced form. While we do not deal with sample selection issues, the method we use is related to the papers mentioned above in that we also require exogenous instrument variables (IVs) to deal with a different form of endogeneity bias that arises in individual covariates. Endogenous covariates are ubiquitous in empirical analyses in social sciences.

An important trait of our method is that it only requires one set of instruments for individual covariates with endogeneity, but does not need additional IVs for the simultaneity in outcomes, which are often hard to find in practice for linear-in-means models with contextual effects. Nor does it need exclusion restrictions that some observed covariates in the structural form have no contextual effects. This is perhaps a surprising feature, because it appears to counter the classical order conditions for identifying models with simultaneity. Nevertheless, the intuition of our method is that with social interactions, the control functions (CFs) constructed using individual-specific instruments have distinctive impacts on the reduced form of an individual's own outcome as well as those of the others. Such distinction between the marginal effects on one's own outcome and the effects on others' outcomes enables us to disentangle the peer effects from the contextual effects.

Our work also contributes to the treatment effects literature, by relaxing the Stable Unit Treatment Value Assumption (SUTVA). That is, self-selected treatments are allowed to influence the outcome of other group members, both through contextual and endogenous effects.

We propose a two-step estimator, which implements classical control function methods or Heckit correction for endogeneity bias in the context of social interactions. It consistently estimates peer and contextual effects in a model where endogenous covariates have both direct and contextual effects. We apply our method to estimate social effects in students' academic achievements in a sample from elementary schools in Tennessee (STAR). Our model allows students' current (Grade 3) achievements to have path dependence on the previous (Grade 2) achievements, which are possibly endogenous due to unobserved individual/family endowment or measurement errors. Using lagged class sizes, teacher qualifications and students' self-reported motivation scores as instruments, we find significant evidence for positive peer effects and path dependence on Grade 2 math scores.

Throughout this paper, we focus primarily on a linear-in-means social interactions model with a parametric distribution of errors, *even though* the main insight can be generalized to nonlinear models with nonparametric error distributions.[3] We choose to do so because of two reasons. First, even within the simple setting of linear parametric models, this insight – individual instruments for endogenous covariates can be used to solve the reflection problem, with no need for exclusion restrictions or additional IVs for simultaneity - - has not been appreciated and applied in the existing literature. Second, the LIM social interactions model is popular among empiricists, yet to the best of our knowledge no earlier works have dealt with the reflection problem *in combination with* endogenous covariates. Therefore, our goal is to propose a tractable method that addresses these two challenges together in a linear framework that is popular among practitioners.

The paper unfolds as follows. We introduce the linear-in-means model with endogeneity and discuss its identification in the next section. We propose a two-step estimator in Section 3, and study some extensions in Section 4. We illustrate our method by an empirical application of peer effects in classrooms in Section 5 and monte carlo experiments in Appendix A. Section 6 concludes.

2 The Model and Identification

We consider a data-generating process (DGP) that generates a large number of independent groups with fixed sizes. To simplify our exposition of the main idea, suppose each group has n members in this section, so that we can suppress the group index g in notation (e.g., individual outcomes $Y_{g,i}$ and their group means $\overline{Y}_g$).[4] Later in Section 3, which introduces a two-step estimator, we generalize by allowing the group sizes n_g to vary across g.

[3]We discuss in Section 2.2 how to extend the core idea when the auxiliary equation modeling endogenous variables is non-linear with a nonparametric distribution of errors.

[4]Lee (2007a) identifies peer and contextual effects using variation in group sizes. Our method allows variable as well as fixed group sizes (see Section 3). In this section, we focus on introducing another identification strategy for the reflection problem through a control function approach. Hence, we present on a setting with fixed group sizes only to simplify exposition of the main idea.

The structural form of individual outcomes is:

$$Y_i = \alpha\overline{Y} + \beta_0 + X_i'\beta_X + \overline{X}'\gamma_X + D_i\beta_D + \overline{D}\gamma_D + U_i, \text{ for } i = 1, \cdots, n, \tag{1}$$

where $\alpha \neq 1$, β_0 is a structural intercept, Y_i and X_i are the outcome and exogenous characteristics of individual i respectively, and U_i is a scalar unobserved error term. The endogenous covariate D_i can be either discrete or continuous. Let $\overline{Y}, \overline{X}, \overline{D}, \overline{U}$ be the average of Y_i, X_i, D_i, U_i among n members in the same group. We refer to α as the endogenous peer effect, γ_X, γ_D as contextual effects, and β_X, β_D as direct, individual effects. Note the model uses the overall mean of peer outcomes and characteristics on the right-hand side, as opposed to a "leave-one-out" average of *other* peers. This specification is used in a variety of empirical contexts. See, for example, Trogdon et al. (2008) and Mora and Gil (2013).

To fix ideas, let D_i be a scalar variable. (Generalization to the case with a multivariate D_i is conceptually straightforward but is algebraically more involved.) By construction, the group means are

$$\overline{Y} = \frac{\beta_0}{1-\alpha} + \overline{X}'\frac{\beta_X + \gamma_X}{1-\alpha} + \overline{D}\frac{\beta_D + \gamma_D}{1-\alpha} + \frac{\overline{U}}{1-\alpha}.$$

Substituting this in (1) gives the following reduced form:

$$Y_i = \tilde{\beta}_0 + X_i'\beta_X + \overline{X}'\tilde{\gamma}_X + D_i\beta_D + \overline{D}\tilde{\gamma}_D + \widetilde{U}_i, \tag{2}$$

where $\tilde{\beta}_0 \equiv \frac{\beta_0}{1-\alpha}$, $\tilde{\gamma}_X \equiv \frac{\alpha\beta_X+\gamma_X}{1-\alpha}$, $\tilde{\gamma}_D \equiv \frac{\alpha\beta_D+\gamma_D}{1-\alpha}$, and $\widetilde{U}_i \equiv U_i + \frac{\alpha\overline{U}}{1-\alpha}$. In this reduced form, each individual's outcome Y_i also depends on other members' unobserved errors U_j through the composite error $\widetilde{U}_i$.

We will use control function methods to deal with endogeneity in discrete or continuous D_i, and to resolve the reflection problem and identify all social effects. The control function method has been applied widely in theory and practice due to its simplicity and flexibility. Since its introduction by Heckman and Robb (1985), the control function approach has been used in a variety of settings. See, for example, Newey, Powell, and Vella (1999), Vella and Verbeek (1999), Chesher (2003), Das, Newey, and Vella (2003), Lee (2007b), Florens, Heckman, Meghir, and Vytlacil (2008), Carrell, Fullerton, and West (2009), Imbens and Newey (2009), Klein and Vella (2010), Petrin and Train (2010), Hahn and Ridder (2011), Kasy (2011), and Blundell and Matzkin (2014) among others.

2.1 Binary Endogenous Variable: A Baseline Case

Let D_i be binary and endogenous. For example, D_i can be a treatment for individual i. In this case, the model in (1) relaxes the SUTVA condition in treatment effects, because the outcome for a member i depends on the treatment of other individuals $j \neq i$ both directly through the contextual effect γ_D and indirectly through the peer effect α. Assume that for each $i \leq n$,

$$D_i = 1\left\{Z_i'\delta + V_i \geq 0\right\}. \tag{3}$$

We assume Z_i contains distinct elements not included in X_i. This assumption subsumes the case where X_i is a strict sub-vector of Z_i; it is necessary for rank conditions that identify the model. In Section 2.2, we show how to generalize (3) by allowing D_i to depend on other individual covariates Z_j for $j \neq i$ as well.

Let $D \equiv (D_i)_{i \leq n}$; and likewise for Z, V, X, U. Write (2) as

$$Y_i = \tilde{\beta}_0 + X_i'\beta_X + \overline{X}'\tilde{\gamma}_X + D_i\beta_D + \overline{D}\tilde{\gamma}_D + E\left(\widetilde{U}_i|X, D, Z\right) + \eta_i$$

where $E(\eta_i|X, D, Z) = 0$ by construction. Assume that (U_i, V_i) are identically and independently distributed across $i \leq n$, independent from (X, Z), and bivariate normal

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

where $\sigma_{uv} \neq 0$. Then

$$\begin{aligned} E\left(\widetilde{U}_i|X,D,Z\right) &= E\left(U_i|X,D,Z\right) + \frac{\alpha}{1-\alpha}E\left(\overline{U}|X,D,Z\right) \\ &= E\left(U_i|Z_i,D_i\right) + \frac{\alpha}{1-\alpha}\left[\frac{1}{n}\sum_{j\leq n}E\left(U_j\middle|Z_j,D_j\right)\right] \\ &= \sigma_{uv}R_i + \frac{\alpha}{1-\alpha}\sigma_{uv}\overline{R}, \end{aligned}$$

where $\overline{R} \equiv \frac{1}{n}\sum_{i\leq n} R_i$ and

$$R_i \equiv D_i\frac{\phi(Z_i'\delta)}{\Phi(Z_i'\delta)} - (1-D_i)\frac{\phi(Z_i'\delta)}{1-\Phi(Z_i'\delta)},$$

where ϕ and Φ denote the standard normal PDF and CDF respectively.

The second equality is due to independence of (U_i, V_i) across $i \leq n$ and their joint independence from (X, Z); the third equality uses the bivariate normality of (U_i, V_i). Note R_i is a typical correction term for the endogenous D_i, or a control function, when the latent errors in the structural and treatment equations are bivariate normal. (See Heckman, 1978, Gourieroux, Monfort, Renault, and Trognon, 1987, Vella, 1993, 1998). Denote $\tilde{\sigma}_{uv} \equiv \frac{\alpha}{1-\alpha}\sigma_{uv}$. Thus

$$E(Y_i|X,D,Z) = \tilde{\beta}_0 + X_i'\beta_X + \overline{X}'\tilde{\gamma}_X + D_i\beta_D + \overline{D}\tilde{\gamma}_D + \sigma_{uv}R_i + \tilde{\sigma}_{uv}\overline{R}. \quad (4)$$

With Z_i assumed independent from V_i and the standard deviation of V_i normalized to 1, we can identify and consistently estimate δ from (3) using a Probit step. Thus R_i can be treated as known in subsequent steps for identification. Assume the support of $(1, X_i, \overline{X}, D_i, \overline{D}, R_i, \overline{R})$ is not contained in a linear subspace for each $i \leq n$. This condition holds generically because Z_i has distinct elements excluded from X_i and because R_i is nonlinear in Z_i. Then regressing Y_i on $(1, X_i, \overline{X}, D_i, \overline{D}, R_i, \overline{R})$ in (4) identifies the reduced-form coefficients

$$\tilde{\beta}_0, \beta_X, \tilde{\gamma}_X, \beta_D, \tilde{\gamma}_D, \sigma_{uv}, \tilde{\sigma}_{uv}.$$

It then follows that $\alpha, \beta_0, \gamma_X, \gamma_D$ are also identified:

$$\alpha = \frac{\tilde{\sigma}_{uv}}{\sigma_{uv} + \tilde{\sigma}_{uv}},\ \beta_0 = (1-\alpha)\tilde{\beta}_0,\ \gamma_X = (1-\alpha)\tilde{\gamma}_X - \alpha\beta_X,\ \gamma_D = (1-\alpha)\tilde{\gamma}_D - \alpha\beta_D. \quad (5)$$

In this case, the reflection problem is solved thanks to the additional source of variation provided by the terms that correct endogeneity, i.e., $R_i, i = 1, ..., n$.

There is another intuitive interpretation of our method. We have used instruments Z_i to construct individual-level control functions (CFs) R_i. Thus we can substitute $U_i = \sigma_{uv}R_i + \zeta_i$ in the structural form of (1), where $\zeta_i \equiv [U_i - E(U_i|X, D, Z)]$. Then these CFs as *generated* regressors, which satisfy the exclusion restriction in Manski (1993). Moreover, in the reduced form for each i in (4), the instruments for the other individuals, Z_j for $j \neq i$, contribute to a *second* CF, i.e., $\overline{R}$, that is associated with the endogenous peer average $\overline{D}$. Thanks to these two contributions by individual and peer instruments respectively, we are able to solve the reflection problem without further assumptions.

Some remarks about the assumption that "(U_i, V_i) are independent across $i \leq n$" are in order. This condition is stronger than necessary for identification; yet it is essential for tractable estimation in the linear parametric framework we consider. Without such independence, an individual control function $E(U_i|D, Z)$ (where D, Z consists of D_j, Z_j for all j in the group) would be a nonlinear function of *multiple indices* $(Z_j'\delta)_{j\leq n}$, even when (U, V) follows a multivariate normal distribution. Such a control function involves a $(n+1)$-dimensional integral over the support of (U_i, V), and needs to be approximated through numerical integration in estimation. Nevertheless, our identification strategy remains valid, provided there is enough joint variation in $(R_i, \overline{R})$ so that the support of $(1, X_i, \overline{X}, D_i, \overline{D}, R_i, \overline{R})$ is not contained in a linear subspace.[5] We acknowledge that the independence of (U_i, V_i) across $i \leq n$ may be restrictive in that

[5] In contrast, when (U_i, V_i) are bivariate normal and independent across $i \leq n$, the individual CF $E(U_i|D, Z)$ takes a simple form of R_i, which only involves a *single index* $Z_i\delta$ in (inverse) mills ratios.

it rules out correlated effects within each group. Later in Section 4.3, we show how to relax this independence assumption and allow for correlated effects in a "leave-one-out" social interactions model where individual structural errors are correlated through unobserved group heterogeneity.

Note our method in this section also relies on the existence of endogeneity, i.e. $\sigma_{uv} \neq 0$. Without such endogeneity, the CFs R_i would not enter the structural or reduced form of the model. As a result, we would not be able to recover the peer effect α as in (5). It is also important to note that a researcher can conveniently test the null hypothesis "$\sigma_{uv} = 0$" by doing an F-test for a null hypothesis that the two reduced-form coefficients for R_i and $\overline{R}$ in (4) are jointly zeros. A direction for future research is to account for such a pre-test in the inference of peer and contextual effects.

2.2 Binary Endogenous Variable: Extensions

In this subsection we discuss how to extend our method in more general settings that either relax the functional form or distributional assumption, or allow for strategic interactions between multiple decision makers in the first-stage model in (3).

First of all, the linearity in Z_i in the first-stage model for D_i in (3) is not necessary and can be relaxed. To see this, suppose $D_i = 1\{g(Z_i) + V_i \geq 0\}$ with V_i distributed as standard normal. In this case, we can recover $g(Z_i)$ by inverting the conditional choice probabilities $\Pr\{D_i = 1|Z_i\} = \Phi(g(Z_i))$, where Φ denotes the standard normal CDF. Control functions can then be constructed similarly, with $Z_i'\delta$ replaced by $g(Z_i)$ in the definition of R_i.

Second, the bivariate normality of (U_i, V_i) is also stronger than necessary, and can be replaced by weaker conditions. In particular, our method applies when the conditional mean of U_i is linear in V_i, i.e., $E(U_i|V_i) = V_i'\varrho$, and the marginal distribution of V_i is unrestricted. In this case, one may first recover $g(\cdot)$ and the marginal distribution of V_i

from an additive, nonparametric model $D_i = 1\{g(Z_i) + V_i \geq 0\}$, using shape restrictions and methods from Matzkin (1992) and Matzkin (1994). The independence between (U_i, V_i) and (X_i, Z_i) then implies $E(U_i|Z_i, D_i = 1) = \varrho R_i$, where $R_i \equiv \int_{\{v \geq -g(Z_i)\}} v dF_{V_i}(v)$ with F_{V_i} being the distribution of V_i identified from the first step. Thus we can use a CF method to identify and estimate the structural parameters as in the baseline case.

Last but not the least, we can in principle extend the method above to a setting where the endogenous D_i is determined through a Bayesian Nash equilibrium (BNE) in a static game with incomplete information. That is,

$$D_i = 1\{Z_i'\delta + \psi E(\overline{D}_{-i}|Z) + V_i \geq 0\}, \text{ where } Z \equiv (Z_i)_{i \leq n}.$$

Then the method above can be applied with $R_i = E(U_i|Z, D) = E(U_i|Z, D_i)$. Note that the conditional mean for U_i depends on $Z_j, j \neq i$ if they are publicly observed by all members. Consequently, R_i and $\overline{R}$ both depend on (D, Z). This is in contrast with the case above, where R_i is a function of (D_i, Z_i) and $\overline{R}$ is function of (D, Z). It is worth noting that in such an extended model, where D is determined in BNE, the rank condition required for identification would fail if the individual members are ex ante symmetric, i.e., $R_i = R_j = \overline{R}$ for all (D, Z).

2.3 Continuous Endogenous Variable

In this section we use the control function (CF) approach to estimate a linear-in-means social interactions model when the endogenous covariates are *continuous*. In this case, an analogous CF method applies under weaker conditions. Specifically, the structural errors (U_i, V_i) only need to be *uncorrelated* with the covariates and instruments (X_i, Z_i) and are therefore allowed to be heteroskedastic. Furthermore, the CF method in this case does not require the joint distribution of (U_i, V_i) to belong to any parametric family.

Consider a DGP that has the same structural and reduced form as in (1) and (2) with $E(X_iU_j) = 0$ for all $i, j \leq n$, but with a *continuous* endogenous D_i:

$$D_i = Z_i'\delta + V_i, \text{ where } E(Z_iV_j) = 0, E(Z_iU_j) = 0 \text{ for all } i, j \leq n. \tag{6}$$

Let Z_i contain distinct elements that are not in X_i. We show how to use control functions to deal with endogenous D_i and solve the reflection problem.

Assume (a) $E(X_iV_j) = 0$ for all $i, j \leq n$ (which is already implied by (6) if X_i is a sub-vector of Z_i), and (b) (U_i, V_i) is uncorrelated with (U_j, V_j) for all $j \neq i$. (Later in Section 4.1 we generalize our method after removing condition (b) and allowing for correlation between (U_i, V_i) across $i \leq n$.)

Write the linear projection of U_i on V_i as:

$$U_i = \rho V_i + e_i. \tag{7}$$

Suppose $\rho \neq 0$ so that the endogeneity in D_i is due to the correlation between U_i and V_i. By construction, $E(V_ie_i) = 0$. Besides, $E(V_ie_j) = 0$ for $i \neq j$ because V_i is uncorrelated with (U_j, V_j). Moreover, $E(X_ie_j) = 0$ and $E(Z_ie_j) = 0$ for all i, j, because e_j is a linear function of U_j and V_j. Thus we can write the reduced form in (2) as

$$Y_i = \tilde{\beta}_0 + X_i'\beta_X + \overline{X}'\tilde{\gamma}_X + D_i\beta_D + \overline{D}\tilde{\gamma}_D + \rho V_i + \tilde{\rho}\overline{V} + \tilde{e}_i, \tag{8}$$

where $\tilde{\rho} \equiv \frac{\alpha\rho}{1-\alpha}$ and $\tilde{e}_i \equiv e_i + \frac{\alpha\bar{e}}{1-\alpha}$.

Because D_i is a linear function of Z_i and V_i, it then follows that the error terms $\tilde{e}_i$ in Equation (8) are also uncorrelated with Z_i, D_i, V_i and their respective group means. By regressing D_i on Z_i, we can consistently estimate V_i and its group mean $\overline{V}$. Thus for identification, we can treat V_i and $\overline{V}$ as "observable". Hoshino (2023) uses a similar strategy to control for endogenous treatment and summed treatments of friends in a potential outcome framework, i.e., including own controlled variable and summed control variables of friends. Assume the support of $(1, X_i, \overline{X}, D_i, \overline{D}, V_i, \overline{V})$

is not contained in a linear subspace for each $i \leq n$. We identify

$$\tilde{\beta}_0, \beta_X, \tilde{\gamma}_X, \beta_D, \tilde{\gamma}_D, \rho, \tilde{\rho}$$

from OLS regression of Y_i on $(1, X_i, \overline{X}, D_i, \overline{D}, V_i, \overline{V})$. Then we can recover the structural parameters as:

$$\alpha = \frac{\tilde{\rho}}{\rho + \tilde{\rho}},\ \beta_0 = (1-\alpha)\tilde{\beta}_0,\ \gamma_X = (1-\alpha)\tilde{\gamma}_X - \alpha\beta_X,\ \gamma_D = (1-\alpha)\tilde{\gamma}_D - \alpha\beta_D. \quad (9)$$

In this case, additional sources of exogenous variation from the control function variables V_i helps us to solve the reflection problem. Note that the endogeneity of D_i is in fact necessary for our method, i.e., our method requires $\rho \neq 0$ in order to recover α from the reduced-form coefficients as in (9). As in the case with σ_{uv} in Section 2.1, a researcher can conveniently test the null hypothesis "$\rho = 0$" by doing a simple t-test on the reduced-form coefficient for V_i in (8).

Similar to the case with discrete endogeneity, our method essentially uses the CFs (V_i) as generated regressors that satisfy the exclusion restriction in Manski (1993). In other words, if we substitute (7) into (1), then V_i serves as an additional regressor in the structural form, which has no contextual effects. Thus the reflection problem is resolved thanks to Proposition 2 in Manski (1993).

To reiterate, the CF method for dealing with continuous endogenous variables under social interactions only requires $E(X_iU_j), E(X_iV_j), E(Z_iU_j), E(Z_iV_j)$ to be zero for all i, j. This allows (U_i, V_i) to be heteroskedastic, e.g., to have conditional variances that depend on the values of (X_i, Z_i). Moreover, it does not require the distribution of (U_i, V_i) to belong to any parametric family. These conditions are much weaker than those used for the case with dummy endogenous covariates in Section 2.1.

3 Two-Step Estimation

For simplicity, we present estimators when X_i is a strict sub-vector of Z_i; generalization to cases where X_i contains distinct elements from Z_i is straightforward. Let the sample contain G independent groups, indexed by $g = 1, ..., G$. Group sizes may vary across the groups, with n_g denote the number of individuals in a group g. For each group, the sample reports $\{Y_g, D_g, Z_g\}_{g\leq G}$, where $Y_g \equiv (Y_{g,i})_{i\leq n_g}$ and likewise for D_g, Z_g.

Consider the baseline case with binary endogenous variable $D_{g,i} \in \{0,1\}$ in Section 2.1. Let $\hat{\delta}$ denote the first-step Probit estimator for δ in (3). For each individual i in group g, calculate

$$\widehat{R}_{g,i} \equiv \left[D_{g,i}\lambda(Z_{g,i}\hat{\delta}) - (1 - D_{g,i})\lambda(-Z_{g,i}\hat{\delta})\right],$$

where $\lambda(\cdot) \equiv \phi(\cdot)/\Phi(\cdot)$. Let $\widehat{\overline{R}}_g \equiv \frac{1}{n_g}\sum_{i=1}^{n_g}\widehat{R}_{g,i}$, and define a row-vector of generated regressors as

$$W_{g,i}(\hat{\delta}) \equiv \left(1, X_{g,i}, \overline{X}_g, D_{g,i}, \overline{D}_g, \widehat{R}_{g,i}, \widehat{\overline{R}}_g\right).$$

Let $\theta \equiv \left(\beta_0, \beta_X', \tilde{\gamma}_X', \beta_D', \tilde{\gamma}_D', \sigma_{uv}, \tilde{\sigma}_{uv}\right)' \in \mathbb{R}^{\dim(\theta)}$ be a column-vector that collects all reduced-form parameters. Let $W(\hat{\delta})$ be a $(\sum_{g=1}^{G} n_g)$-by-$\dim(\theta)$ matrix that stacks the row-vector of generated regressors $W_{g,i}(\hat{\delta})$ from all groups and individuals, and Y be $(\sum_{g=1}^{G} n_g)$-by-1 vector that stacks the column-vectors $Y_g \in \mathbb{R}^{n_g}$ for all $g = 1, ..., G$. Our two-step estimator for θ is constructed by regressing Y on $W(\hat{\delta})$:

$$\hat{\theta} \equiv \left[W(\hat{\delta})'W(\hat{\delta})\right]^{-1}\left[W(\hat{\delta})'Y\right].$$

We derive the asymptotic property of $\hat{\theta}$ as a two-step m-estimator as follows. First, under regularity conditions, e.g., as 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 their population counterparts as $G \to \infty$. This establishes the consistency of our estimator: $\hat{\theta} \xrightarrow{p} \theta$. Next, let

$A \equiv \lim_{G\to\infty} E\left(\frac{1}{G}\sum_{g=1}^{G} W_g' W_g\right)$, where W_g is shorthand for the n_g-by-$dim(\theta)$ matrix $W_g(\delta)$, evaluated at the true parameter. Under standard regularity conditions, the first-order condition in the second-step regression implies:

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

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

$$G^{-1/2}\sum\nolimits_g s_g(\theta;\hat{\delta}) = G^{-1/2}\sum\nolimits_g s_g(\theta;\delta) + F_0\sqrt{G}(\hat{\delta}-\delta) + o_p(1),$$

where $F_0 \equiv \lim_{g\to\infty} E\left[\frac{1}{G}\sum_{g=1}^{G}\nabla_\delta s_g(\theta;\delta)\right]$. 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 $B \equiv \lim_{g\to\infty} E\left[\frac{1}{G}\sum_{g=1}^{G} m_g(\theta;\delta)m_g(\theta;\delta)'\right]$, with $m_g(\theta;\delta) \equiv s_g(\theta;\delta) + F_0 r_g(\delta)$.

The components in asymptotic variance A, B can both be consistently estimated by their sample analogs. In our empirical application, we use bootstrap resampling methods to calculate the standard errors. Our estimator $\hat{\theta}$ is a smooth function of sample averages that are constructed using a standard MLE (probit) estimator $\hat{\delta}$ in a first step. Hence we employ the bootstrap algorithm as a means to approximate sampling uncertainties in the application. A formal proof of bootstrap validity is outside the scope of the paper, and left for future research.

To estimate remaining structural parameters, i.e., the peer effect α, the contextual effects γ_X, γ_D and the intercept β_0, simply plug $\hat{\theta}$ in the formulas in (5). Asymptotic variance of these parameters can be obtained from the limiting distribution of the reduced-form parameters in (4) using the Delta Method.

The model with continuous endogenous D_i is estimated using a similar two-step procedure. First, regress $D_{g,i}$ on $Z_{g,i}$ for all individual i and group g. Calculate individual residuals $\widehat{V}_{g,i}$ and group means $\widehat{\overline{V}}_g \equiv \frac{1}{n_g}\sum_{i=1}^{n_g}\widehat{V}_{g,i}$. Next, use pooled OLS to regress $Y_{g,i}$ on $(1, X_{g,i}, \overline{X}_g, D_{g,i}, \overline{D}_g, \widehat{V}_{g,i}, \widehat{\overline{V}}_g)$ and get estimates for $\hat{\tilde{\beta}}_0, \hat{\beta}_X, \hat{\tilde{\gamma}}_X, \hat{\beta}_D, \hat{\tilde{\gamma}}_D, \hat{\rho}, \hat{\tilde{\rho}}$. Then plug them in (9) to estimate remaining structural parameters $\alpha, \gamma_X, \gamma_D$. Asymptotic properties are similar to the case with discrete D_i, with $V_{g,i}, \rho$ playing the roles that are analogous to those of $R_{g,i}, \delta$ in the former case.

4 Extensions

4.1 Continuous Endogenous Variables with Correlated Errors

In this subsection we extend the method for continuous endogenous variables in Section 2.3 to more general settings where the structural errors (U_i, V_i) are correlated across individual members $i = 1, 2, ..., n$ in the same group. We maintain the same conditions as in Section 2.3, except for the uncorrelation between (U_i, V_i) and (U_j, V_j) in (b).

In the first step, project U_i over V_i and the average of V_j with $j \neq i$. That is,

$$U_i = \rho V_i + \varphi \overline{V}_{-i} + e_i.$$

By substituting this adapted linear projection into the reduced form (2), and using the facts that $\overline{V}_{-i} = \frac{n\overline{V} - V_i}{n-1}$ and $\frac{1}{n}\sum_i \overline{V}_{-i} = \overline{V}$, we get

$$Y_i = \tilde{\beta}_0 + X_i'\beta_X + \overline{X}'\tilde{\gamma}_X + D_i\beta_D + \overline{D}\tilde{\gamma}_D + \rho_n^* V_i + \varphi_n^*\overline{V} + \tilde{e}_i, \quad (10)$$

for each $i = 1, 2, ..., n$, where $\rho_n^* \equiv \rho - \frac{\varphi}{n-1}$ and $\varphi_n^* \equiv \frac{\alpha(\rho+\varphi)}{1-\alpha} + \frac{n}{n-1}\varphi$ while $\tilde{\beta}_0, \tilde{\gamma}_X, \tilde{\gamma}_D$ are the same as in the reduced form (2).

Under our maintained assumptions and the property of linear projection, the composite errors $\tilde{e}_i$ are uncorrelated with all explanatory variables on the right-hand side of equation (10) (including the generated regressors V_i and $\overline{V}$). Thus we can use

OLS to consistently estimate ρ_n^* and φ_n^*.

Next, suppose there is exogenous variation in the group size n in the sample. That is, the reported groups have different sizes such as n and n', but share the same structural parameters. Then we can recover φ from the difference of identified reduced-form parameters $\varphi_{n'}^* - \varphi_n^*$. This in turn allows us to sequentially (over-)identify ρ from ρ_n^*, and then α from φ_n^* respectively.

Note the method above allows for flexible correlation between (U_i, V_i) across all members $i = 1, 2, ..., n$ in the same group. In fact, it exploits a nonzero φ when such correlation exists. This differs qualitatively from earlier papers that use second moments of observed outcomes to identify peer effects, e.g., Sacerdote (2001) and Liu (2017). In those papers, the structural errors are assumed to be uncorrelated across individual members within the same group.

4.2 Binary Endogeneity with Group Heterogeneity: Random Effects

Suppose the structural form of DGP is similar to (1), except that now it contains a group-level *unobserved* heterogeneity c. The reduced form of such a model is:

$$Y_i = \tilde{\beta}_0 + X'\beta_X + \overline{X}'\tilde{\gamma}_x + D\beta_D + \overline{D}\tilde{\gamma}_D + \tilde{c} + \widetilde{U}_i \text{ for } i \leq n,$$

where $\tilde{c} \equiv c/(1-\alpha)$ and $\tilde{\beta}_0, \tilde{\gamma}_x, \tilde{\gamma}_D, \widetilde{U}_i$ are as defined in Section 2. Let D_i be determined as in (3). To fix ideas, suppose X_i is a strict sub-vector of Z_i. Let $Z \equiv (Z_i)_{i \leq n}$; likewise define two n-vectors U, V. Assume $E(c|Z, V) = 0$; (U, V) is independent from Z; and (U_i, V_i) is bivariate normal and independent across $i \leq n$. Under these assumptions, $E(Y_i|Z, D)$ takes the same form as (4) in Section 2:

$$E(Y_i|Z, D) = \tilde{\beta}_0 + X'\beta_X + \overline{X}'\tilde{\gamma}_x + D\beta_D + \overline{D}\tilde{\gamma}_D + \sigma_{uv}R_i + \tilde{\sigma}_{uv}\overline{R}.$$

Therefore we can apply the same method as in Section 2.1 to identify and estimate all structural parameters.

The main identifying condition is the "random effect" assumption that $E(c|Z,V)=0$. This rules out endogeneity in D_i due to correlation between group heterogeneity c and individual noises V. However, both (U_i,V_i) and (c,U) are allowed to be correlated respectively. In a multivariate normal case, this means

$$\begin{pmatrix} c \\ U_i \\ V_i \\ U_j \\ V_j \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_c^2 & \sigma_{cu} & 0 & \sigma_{cu} & 0 \\ & \sigma_u^2 & \sigma_{uv} & 0 & 0 \\ & & 1 & 0 & 0 \\ & & & \sigma_u^2 & \sigma_{uv} \\ & & & & 1 \end{pmatrix}\right).$$

4.3 Binary Endogeneity with Group Heterogeneity: Fixed Effects

In this subsection, we remove the random-effect assumption $E(c|Z,V)=0$, and apply a fixed-effect approach to recover peer and contextual effects in a social interactions model. We allow group sizes to vary across observations, and assume the structural parameters are the same across groups with different sizes. This assumption was maintained in other papers that used group size variation to identify models with social interactions or social networks, such as Lee (2007a) and Davezies, d'Haultfoeuille, and Fougère (2009). In this section we extend these papers by allowing for endogenous covariates. We hope to make the following point here: When there is additional complication due to endogenous regressors, researchers can use the variation in individual instruments, as well as that in group sizes, to identify the model.

Let the group means take a "leave-one-out" form. That is,

$$Y_i = \alpha\overline{Y}_{-i} + X_i'\beta_X + \overline{X}_{-i}'\gamma_X + D_i\beta_D + \overline{D}_{-i}\gamma_D + c + U_i \text{ for } i \le n,$$

where $\overline{Y}_{-i} \equiv \frac{1}{n-1}\sum_{j\neq i} Y_j$ and likewise for $\overline{X}_{-i}$ and $\overline{D}_{-i}$. Lee (2007) showed that for all

$i \leq n$,

$$Y_i - \overline{Y} = (X_i - \overline{X})\pi_{X,n} + (D_i - \overline{D})\pi_{D,n} + \pi_{0,n}(U_i - \overline{U}),$$

where $\overline{Y} \equiv \frac{1}{n-1}\sum_{j=1}^{n} Y_i$ and $\overline{X}, \overline{D}, \overline{U}$ are respective group means, and

$$\pi_{X,n} = \frac{(n-1)\beta_X - \gamma_X}{n-1+\alpha}, \pi_{D,n} = \frac{(n-1)\beta_D - \gamma_D}{n-1+\alpha}, \pi_{0,n} = \frac{n-1}{n-1+\alpha}. \tag{11}$$

The endogenous covariate D_i is determined as in (3), with X_i being a strict sub-vector of Z_i.

As before, assume that (U_i, V_i) are independent from Z, and are i.i.d. bivariate normal across group members $i \leq n$. Normalize the standard deviation of V_i to 1. As before, $E(U_i|Z, D) = E(U_i|Z_i, D_i) = \sigma_{uv}R_i$ and $E(\overline{U}|Z, D) = \sigma_{uv}\overline{R}$ with $\overline{R} \equiv \sum_i R_i/n$. Thus by regressing $Y_i - \overline{Y}$ on $(X_i - \overline{X}, D_i - \overline{D}, R_i - \overline{R})$ and their interaction with group size dummies, we can consistently estimate $\pi_{X,n}$, $\pi_{D,n}$, and $\tilde{\pi}_{0,n} \equiv \pi_{0,n}\sigma_{uv}$ for $n \geq 2$. For each group size n represented in the data-generating process, we can use (11) to construct a linear system:

$$\underbrace{\begin{pmatrix} -\pi_{X,n} & n-1 & -1 & 0 & 0 & 0 \\ -\pi_{D,n} & 0 & 0 & n-1 & -1 & 0 \\ -\tilde{\pi}_{0,n} & 0 & 0 & 0 & 0 & n-1 \end{pmatrix}}_{\equiv \mathbf{M}(n)} \underbrace{\begin{pmatrix} \alpha \\ \beta_X \\ \gamma_X \\ \beta_D \\ \gamma_D \\ \sigma_{uv} \end{pmatrix}}_{\equiv \tau} = \underbrace{\begin{pmatrix} (n-1)\pi_{X,n} \\ (n-1)\pi_{D,n} \\ (n-1)\tilde{\pi}_{0,n} \end{pmatrix}}_{\equiv \mathbf{b}(n)}.$$

Stacking two linear systems with group sizes $n \neq n'$, we get

$$\begin{pmatrix} \mathbf{M}(n) \\ \mathbf{M}(n') \end{pmatrix} \tau = \begin{pmatrix} \mathbf{b}(n) \\ \mathbf{b}(n') \end{pmatrix}.$$

The structural parameters are identified as long as the coefficient matrix on the left-hand side has full rank. This holds generically over the parameter space of $(\alpha, \beta_X, \gamma_X, \beta_D, \gamma_D, \sigma_{uv})$. With more variation in the group sizes, we can append the linear system above with more equations to achieve identification under proper rank conditions. Note the assumption on the unobserved errors here is weaker than that in the previous section. Namely, in a multivariate normal case, this means the covariance between c and V_i as well as U_i are allowed to be both nonzero.

5 Peer Effects in Academic Achievements

In this section, we study peer effects in student academic achievements using data from elementary schools in the State of Tennessee. The data comes from the Student/Teacher Achievement Ratio (STAR) Project, which was a four-year longitudinal study funded by the Tennessee General Assembly, and conducted by the Tennessee State Department of Education. The project was designed to study the relation between class sizes and student academic performance through randomized experiments.

We apply our method to infer peer and contextual effects in Grade 3 math test scores. In particular, we include students' lagged test scores from Grade 2 as an explanatory variable, in order to account for previous educational inputs and heritable endowments, and obtain an "value-added" interpretation. Other papers that used lagged scores as covariates include Todd and Wolpin (2003) and Hanushek, Kain, Markman, and Rivkin (2003). Our analysis takes into account the endogeneity in lagged scores, which could be due to unobserved heterogeneity in student ability and family influence that persisted over time or due to the interpretation of Grade 2 score as a proxy for a student's ability.

Papers that studied general peer effects in student academic achievements include Hoxby (2000), Sacerdote (2001), Zimmerman (2003), Angrist and Lang (2004), Hoxby and Weingarth (2005), Kang (2007), Ammermueller and Pischke (2009), Calvó-

Armengol, Patacchini, and Zenou (2009), Carrell, Fullerton, and West (2009), Duflo, Dupas, and Kremer (2011), Lin (2010), Lavy, Paserman, and Schlosser (2012), Burke and Sass (2013), Hong and Lee (2017), Ross and Shi (2021). More specifically, a series of earlier papers had used the same source of STAR data to study the relation between students' academic achievements and class or peer characteristics. Word et al. (1990) and Krueger (1999) found evidence that on average small classes had positive effects on student achievements; Krueger and Whitmore (2001) analyzed the effect of students' past attendance in small classes; Dee (2004) investigated the effect of exposure to an own-race teacher; Whitmore (2005) documented that both genders showed similar gains from being assigned to small classes, and noted that such gains depend on the gender composition of classrooms. Graham (2008) and Sojourner (2013) estimated the impact of peer characteristics on student outcomes, by exploiting second moment restrictions and pre-assignment achievement measures, respectively. In the terminology of Manski (1993), the impact they recovered is a composite of peer and contextual effects. Boozer and Cacciola (2001) estimated the peer effects in a model with no contextual effect, using experimental variation in class quality (fraction of students exposed in the previous year to small classes) as an instrument for peer achievement. Hanushek, Kain, Markman, and Rivkin (2003) studied a model where the vector of student/family characteristics with direct effects differed from those with contextual effects. They also used lagged rather than contemporaneous peer achievement in the structural form. Our work differs from these papers, and contributes to the literature by estimating the peer, contextual and individual effects in a general specification that allows all individual characteristics to have contextual effects, and accommodates endogeneity in students' lagged test scores.

Our sample contains 4,821 students in 327 classes from 79 schools in the State of Tennessee. The students were randomly assigned into one of three class types (interventions): small class, regular class, and regular-with-aide class. Teachers were also randomly assigned into classes. The longitudinal study followed a cohort of students from kindergarten to the third grade. For each student, the sample reports his/her race, gender, days of absence and presence in the third grade, as well as the type of class he/she belonged to in each grade. The data also contains students' Self-Concept and Motivation Scores in each grade. [6] In addition, the data reports the experience and qualification of teachers assigned to each class, as well as whether the school is located in an urban area.

We study the peer effects in classroom on Grade 3 (G3) math test score. Due to the unobserved heterogeneity or measurement error (if we take Grade 2 math performance as proxy for ability), the Grade 2 math performance is considered endogenous. Table 1 summarizes student and class characteristics in the sample. G3/G2 math scores report the percent of learning objectives mastered by a student, which are measured on a standardized scale of 100, a.k.a. BSF (Basic Skill Factor); other covariates with no designated units are dummy variables.

We adopt the following econometric specification for peer and contextual effects in classroom, with β_D, γ_D reflecting the "path dependence" on a student's own and classmates' test scores in G2 math:

$$G3M_{g,i} = \alpha_0 + \alpha_1 \overline{G3M}_g + X'_{g,i}\beta_X + \overline{X}'_g \gamma_X + \beta_D G2M_{g,i} + \gamma_D \overline{G2M}_g + \rho V_{g,i} + e_{g,i}. \qquad (12)$$

In this specification, $X_{g,i}$ contains both student and class characteristics. These include Gender, White (a dummy variable for race), Days of Absence, self-reported Motivation Score, Self-Concept Score, Class Type (small or regular), School Urbanicity, Teacher

[6] Each student participating in STAR was asked to complete a self-concept and motivation inventory, a.k.a. SCAMIN (Milchus, Farrah, and Reitz, 1968), which asked students to indicate pictorially their responses to 24 situations. These responses are then condensed into a continuous measure of motivation and self-concept on a scale between 0-60.

Table 1: Summary Statistics

Variable	Mean	SD
G3 Math	83.859	19.614
G2 Math	89.540	12.966
Female	0.491	0.500
White	0.677	0.468
Absence (days)	6.568	6.047
G3 Motivation	49.214	3.978
G3 Self-Concept	44.163	4.773
G2 Motivation	49.618	3.732
G2 Self-Concept	46.786	4.655
Class Characteristics		
G3 Small Class	0.416	0.494
G2 Small Class	0.382	0.487
School Urbanicity	0.526	0.500
G3 Teacher Bachelor Degree	0.557	0.498
G2 Teacher Bachelor Degree	0.606	0.489
G3 Teacher Experience (yrs)	13.981	8.606
G2 Teacher Experience (yrs)	13.281	8.787
G3 Teacher STAR training	0.153	0.360
G2 Teacher STAR training	0.147	0.354

Bachelor Degree,[7] Teacher Experience (in years) and a dummy for whether the teacher had received STAR training. We let $\overline{X}_g$ consist of the proportion of females and whites in each class, the class averages of days of absence and presence, and the class average of motivation and self-concept scores.

As noted in Section 2.3, $V_{g,i}$ is the residual from a first step that regresses G2 math scores on a vector of exogenous covariates, which include our choice of instruments: lagged motivation score of each student in the second grade, lagged self-concept score of each student in the second grade, lagged values of teacher experiences and qualifications in the second grade, and the class type the student belonged to in the second grade. These instruments do not directly contribute to G3 math scores, once the path dependence through G2 scores are taken into account. On the other hand, they do

[7] We define Teacher Bachelor Degree to be 1 if the teacher is with a bachelor degree and 0 for those with master and higher degree.

directly contribute to the endogenous G2 math scores. Table A5 in the appendix reports estimates, R-square and F-statistic from the first-stage regression. It provides evidence that such lagged information about teacher and students has statistically significant impact on G2 math scores.

Table 2: Estimates of Social Effects in Structural Equation

Variable	Estimate	Standard Error
G2 Math	1.142**	0.212
Average G2 Math	−0.558**	0.164
Female	0.506	0.501
Average Female	0.449	1.813
White	2.371**	1.128
Average White	−0.162	2.729
Absence	−0.185**	0.041
Average Absence	−0.045	0.182
Motivation	0.062	0.064
Average Motivation	−0.194	0.165
Self-Concept	0.067	0.050
Average Self-Concept	0.037	0.148
Small Class	0.005	1.022
School Urbanicity	1.426	1.018
Teacher Degree	−1.711*	0.891
Teacher Experience	−0.052	0.055
Teacher STAR training	−1.076	1.055
Intercept	−16.740	15.048
Peer effects	0.598**	0.267
$\widehat{V}$	−0.369*	0.213

R^2 = 0.348, F-statistic=134.9 from the reduced form regression.

Table 2 reports the results from our two-step estimator that uses the control function approach to deal with endogeneity in G2 math scores. The peer effects is estimated to be 0.598, and is statistically significant at 5% level. A student's G2 math score is shown to have a significant direct positive effect on his/her own G3 score; the point estimate of the size of this effect is larger than that of the peer effect. Teachers with master degrees (and higher) are shown to have a positive effect on students' G3 math scores.

The days of absence from school has a small yet significant negative effect (per day) on G3 math scores. Moreover, statistical significance of the control function variable $V_{g,i}$ suggests there is correlation between U and V, which corroborates endogeneity in G2 math scores. The negative sign for the coefficient of $V_{g,i}$ can be attributed to measurement errors in the structural equation. That is, the lagged math score $G2M$ can be interpreted as a noisy proxy for a student's unobservable ability, which is an actual covariate in (12) in place of $G2M$.[8]

For comparison, we estimate two alternative models that ignore endogeneity or path dependence in G2 scores. The first one has no peer effects but includes the endogenous G2 math scores and its contextual effects. In this case, we use control function to deal with endogeneity in G2 math scores, using $\widehat{V}$ from the first stage as in Table A5. The second is even more simplistic, with no peer effects or G2 math scores in the covariates.

Comparing the results in Table 2 with those in Table 3 (no peer effect) and Table 4 (no peer effect or path dependence) illustrates the consequence of ignoring peer effects in empirical analysis. When peer effects are not accommodated in estimation, school urbanicity, teacher experience and teacher training are shown to have significant direct effects on G3 math scores while white race, days of absence and self-reported motivation have significant contextual effects. Yet these effects turn out to be not statistically significant once peer effects are incorporated in the analysis in Table 2. Also, without peer effects, the extent of path dependence, as reflected by the coefficient for G2 Math, is over-estimated at 1.404, in comparison with that reported in Table 2.

Comparison of Table 4 with Tables 2 and 3 indicates that ignoring the endogenous

[8]To see this, suppose the right-hand side of (12) includes a student's unobserved ability, Abt_i instead of $G2M$, where $G2M_{g,i} = Abt_i + \epsilon_{g,i}$. Then the econometric error in the structural equation absorbs $-\beta_{Abt}\epsilon_{g,i} - \gamma_{Abt}\overline{\epsilon}_g$, where β_{Abt} and γ_{Abt} are positive.

Table 3: Estimation Results: No Peer Effects

Variable	Estimate	Standard Error
G2 Math	1.404**	0.132
Average G2 Math	−0.219**	0.046
Female	0.327	0.489
Average Female	1.953	2.214
White	1.446	1.068
Average White	5.900**	1.348
Absence	−0.187**	0.040
Average Absence	−0.390**	0.128
Motivation	0.054	0.065
Average Motivation	−0.354*	0.195
Self-Concept	0.068	0.054
Average Self-Concept	0.197	0.155
Small Class	0.435	0.537
School Urbanicity	1.202*	0.618
Teacher Degree	−1.688**	0.476
Teacher Experience	−0.053*	0.028
Teacher STAR training	−0.965	0.644
$\widehat{V}$	−0.634**	0.132
Intercept	−20.417	13.922

R^2 = 0.347, F-statistic=142.0.

G2 math scores would suggest small classes have significant positive effect on G3 math scores. However, such significance disappears once we accommodate path dependence on G2 math scores in the model, and use the control function method to deal with the endogenous lagged math scores. That small classes have no significant effect on student achievements is consistent with findings in Lazear (2001), Hanushek (2003).

Table 4: Estimation Results: No Peer Effects and No Path Dependence

Variable	Estimate	Standard Error
Female	1.404**	0.557
Average Female	−0.985	2.558
White	6.501**	1.113
Average White	8.844**	1.458
Absence	−0.264**	0.046
Average Absence	−0.626**	0.146
Motivation	0.067	0.076
Average Motivation	−0.069	0.225
Self-Concept	0.173**	0.062
Average Self-Concept	−0.073	0.179
Small Class	2.378**	0.573
School Urbanicity	−0.082	0.704
Teacher Degree	−2.280**	0.550
Teacher Experience	−0.081**	0.033
Teacher STAR training	−0.353	0.743
Intercept	76.402**	10.659

$R^2 = 0.123$, F-statistic=44.75.

6 Conclusion

We use control function methods to identify and estimate the peer and contextual effects in linear-in-means social interactions models with endogenous covariates. We resolve the "reflection problem" as noted in Manski (1993) by exploiting the variation in individual-level instruments for endogenous covariates. Our approach has several desirable features. It applies classical control function methods or Heckit correction for endogeneity bias in the new context of social interactions, and it is easy to implement in practice. It eliminates the need for additional instruments for simultaneity in individual outcomes, which are often hard to find in practice. Nor does it require any exclusion restriction that some covariates have no contextual effect. Our work also contributes to the treatment effects literature by relaxing the Stable Unit Treatment Value Assumption (SUTVA). That is, we allow individuals' self-selected treatments to influence the outcomes of other group members through peer and contextual effects. Applying our method to estimate social effects in Grade 3 math scores of elementary school students in Tennessee, we find significant evidence for positive peer effects and path dependence on G2 scores.

References

Ahern, K. R., R. Duchin, and T. Shumway (2014). Peer effects in risk aversion and trust. *The Review of Financial Studies 27*(11), 3213–3240.

Ammermueller, A. and J.-S. Pischke (2009). Peer effects in european primary schools: Evidence from the progress in international reading literacy study. *Journal of Labor Economics 27*(3), 315–348.

Angrist, J. D. and K. Lang (2004). Does school integration generate peer effects? evidence from boston's metco program. *American Economic Review 94*(5), 1613–1634.

Bayer, P., R. Hjalmarsson, and D. Pozen (2009). Building criminal capital behind bars: Peer effects in juvenile corrections. *The Quarterly Journal of Economics 124*(1), 105–147.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2011). Identification of social interactions. In *Handbook of Social Economics*, Volume 1, pp. 853–964. Elsevier.

Blundell, R. and R. L. Matzkin (2014). Control functions in nonseparable simultaneous equations models. *Quantitative Economics 5*(2), 271–295.

Bollinger, B. and K. Gillingham (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science 31*(6), 900–912.

Boozer, M. and S. E. Cacciola (2001). Inside the'black box'of project star: Estimation of peer effects using experimental data. *Available at SSRN 277009*.

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 (2001a). Discrete choice with social interactions. *The Review of Economic Studies 68*(2), 235–260.

Brock, W. A. and S. N. Durlauf (2001b). Interactions-based models. In *Handbook of Econometrics*, Volume 5, pp. 3297–3380. Elsevier.

Burke, M. A. and T. R. Sass (2013). Classroom peer effects and student achievement. *Journal of Labor Economics 31*(1), 51–82.

Bursztyn, L., F. Ederer, B. Ferman, and N. Yuchtman (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica 82*(4), 1273–1301.

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.

Carrell, S. E., R. L. Fullerton, and J. E. West (2009). Does your cohort matter? measuring peer effects in college achievement. *Journal of Labor Economics 27*(3), 439–464.

Chesher, A. (2003). Identification in nonseparable models. *Econometrica 71*(5), 1405–1441.

Chetty, R., N. Hendren, and L. F. Katz (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review 106*(4), 855–902.

Cornelissen, T., C. Dustmann, and U. Schönberg (2017). Peer effects in the workplace. *American Economic Review 107*(2), 425–56.

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

Das, M., W. K. Newey, and F. Vella (2003). Nonparametric estimation of sample selection models. *The Review of Economic Studies 70*(1), 33–58.

Davezies, L., X. d'Haultfoeuille, and D. Fougère (2009). Identification of peer effects using group size variation. *The Econometrics Journal 12*(3), 397–413.

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.

Dee, T. S. (2004). Teachers, race, and student achievement in a randomized experiment. *Review of Economics and Statistics 86*(1), 195–210.

Duflo, E., P. Dupas, and M. Kremer (2011). Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in kenya. *American Economic Review 101*(5), 1739–74.

Durlauf, S. N. (2004). Neighborhood effects. *Handbook of Regional and Urban Economics 4*, 2173–2242.

Florens, J.-P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica 76*(5), 1191–1206.

Gaviria, A. and S. Raphael (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics 83*(2), 257–268.

Gourieroux, C., A. Monfort, E. Renault, and A. Trognon (1987). Generalised residuals. *Journal of Econometrics 34*(1-2), 5–32.

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

Graham, B. S. and J. Hahn (2005). Identification and estimation of the linear-in-means model of social interactions. *Economics Letters 88*(1), 1–6.

Hahn, J. and G. Ridder (2011). Conditional moment restrictions and triangular simultaneous equations. *Review of Economics and Statistics 93*(2), 683–689.

Hanushek, E. A. (2003). The failure of input-based schooling policies. *The Economic Journal 113*(485), F64–F98.

Hanushek, E. A., J. F. Kain, J. M. Markman, and S. G. Rivkin (2003). Does peer ability affect student achievement? *Journal of Applied Econometrics 18*(5), 527–544.

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.

Heckman, J. J. and R. Robb (1985). Alternative methods for evaluating the impact of interventions: An overview. *Journal of Econometrics 30*(1-2), 239–267.

Hong, S. C. and J. Lee (2017). Who is sitting next to you? peer effects inside the classroom. *Quantitative Economics 8*(1), 239–275.

Hoshino, T. (2019). Two-step estimation of incomplete information social interaction models with sample selection. *Journal of Business & Economic Statistics 37*(4), 598–612.

Hoshino, T. (2023). Estimating a continuous treatment model with spillovers: A control function approach. *Journal of Business & Economic Statistics* (Accepted).

Hoxby, C. (2000). Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research.

Hoxby, C. M. and G. Weingarth (2005). Taking race out of the equation: School reassignment and the structure of peer effects. Technical report, Citeseer.

Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

Ioannides, Y. M. and J. E. Zabel (2008). Interactions, neighborhood selection and housing demand. *Journal of Urban Economics 63*(1), 229–252.

Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from south korea. *Journal of Urban Economics 61*(3), 458–495.

Kasy, M. (2011). Identification in triangular systems using control functions. *Econometric Theory 27*(3), 663–671.

Katz, L. F., J. R. Kling, and J. B. Liebman (2001). Moving to opportunity in boston: Early results of a randomized mobility experiment. *The Quarterly Journal of Economics 116*(2), 607–654.

Klein, R. and F. Vella (2010). Estimating a class of triangular simultaneous equations models without exclusion restrictions. *Journal of Econometrics 154*(2), 154–164.

Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica 75*(1), 83–119.

Krueger, A. B. (1999). Experimental estimates of education production functions. *The Quarterly Journal of Economics 114*(2), 497–532.

Krueger, A. B. and D. M. Whitmore (2001). The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project star. *The Economic Journal 111*(468), 1–28.

Lavy, V., M. D. Paserman, and A. Schlosser (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal 122*(559), 208–237.

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics 116*(3), 777–803.

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

Lee, L.-f., X. Liu, and X. Lin (2010). Specification and estimation of social interaction models with network structures. *The Econometrics Journal 13*(2), 145–176.

Lee, S. (2007b). Endogeneity in quantile regression models: A control function approach. *Journal of Econometrics 141*(2), 1131–1158.

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.

Liu, X. (2017). Identification of peer effects via a root estimator. *Economics Letters 156*, 168–171.

Liu, X. and L.-f. Lee (2010). Gmm estimation of social interaction models with centrality. *Journal of Econometrics 159*(1), 99–115.

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

Matzkin, R. L. (1992). Nonparametric and distribution-free estimation of the binary threshold crossing and the binary choice models. *Econometrica: Journal of the Econometric Society*, 239–270.

Matzkin, R. L. (1994). Restrictions of economic theory in nonparametric methods. *Handbook of econometrics 4*, 2523–2558.

Milchus, N. J., G. A. Farrah, and W. Reitz (1968). The self-concept and motivation inventory: what face would you wear? dearborn heights, michigan: Person-o-metrics.

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

Mora, T. and J. Gil (2013). Peer effects in adolescent bmi: evidence from spain. *Health Economics 22*(5), 501–516.

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

Newey, W. K., J. L. Powell, and F. Vella (1999). Nonparametric estimation of triangular simultaneous equations models. *Econometrica 67*(3), 565–603.

Petrin, A. and K. Train (2010). A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research 47*(1), 3–13.

Rose, C. D. (2017). Identification of peer effects through social networks using variance restrictions. *The Econometrics Journal 20*(3), S47–S60.

Ross, S. L. and Z. Shi (2021). Measuring social interaction effects when instruments are weak. *Journal of Business & Economic Statistics*, 1–12.

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

Sampson, R. J., J. D. Morenoff, and T. Gannon-Rowley (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual Review of Sociology 28*(1), 443–478.

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

Sojourner, A. (2013). Identification of peer effects with missing peer data: Evidence from project star. *The Economic Journal 123*(569), 574–605.

Todd, P. E. and K. I. Wolpin (2003). On the specification and estimation of the production function for cognitive achievement. *The Economic Journal 113*(485), F3–F33.

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

Vella, F. (1993). A simple estimator for simultaneous models with censored endogenous regressors. *International Economic Review 34*(2), 441–457.

Vella, F. (1998). Estimating models with sample selection bias: a survey. *Journal of Human Resources 33*(1), 127–169.

Vella, F. and M. Verbeek (1999). Estimating and interpreting models with endogenous treatment effects. *Journal of Business & Economic Statistics 17*(4), 473–478.

Waldinger, F. (2012). Peer effects in science: Evidence from the dismissal of scientists in nazi germany. *The Review of Economic Studies 79*(2), 838–861.

Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement:

Evidence from a randomized experiment. *American Economic Review 95*(2), 199–203.

Word, E. R. et al. (1990). The state of tennessee's student/teacher achievement ratio (star) project: Technical report (1985-1990).

Zimmerman, D. J. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *Review of Economics and Statistics 85*(1), 9–23.

Appendix

A Monte Carlo Evidence

We consider two monte carlo experiments, with the endogenous variable being binary and continuous respectively.

A.1 Binary Endogenous Variable

We let X_i, D_i both be scalar variables. For each group, let $Z_i \equiv (1, X_i, Z_{i1})$ be independent across $i \leq n$, and X_i, Z_{i1} be independent, standard normal. For each i, let (U_i, V_i) be drawn from the bivariate normal with mean $(0, 0)$, unit variance and covariance matrix σ_{uv}. We set the true parameters as $(\alpha, \beta_X, \gamma_X, \beta_D, \gamma_D, \sigma_{uv}) = (1/2, 1, 1, 1, 1, 2/3)$. The vector of outcomes $Y_i, i \leq n$ is generated by the reduced form in (2), and D_i is generated by (3) with $\delta = (0, 1, 2)$. We experiment with group sizes $n = 5$ and $n = 10$, and sample sizes $G = 250, 500, 1,000$. We report average biases and MSE with 1,000 replications in Tables A1 and A2.

Table A1: Monte Carlo Results: Binary D (n=5)

G	Average Bias						
	α	β_0	β_X	β_D	γ_X	γ_D	σ_{uv}
250	-0.028	-0.045	0.001	0.001	0.109	0.151	-0.001
500	-0.015	-0.026	0.000	0.002	0.054	0.076	-0.001
1,000	-0.007	-0.012	-0.001	0.001	0.029	0.036	0.000
	MSE						
	α	β_0	β_X	β_D	γ_X	γ_D	σ_{uv}
250	0.025	0.123	0.001	0.007	0.365	0.652	0.005
500	0.010	0.053	0.001	0.004	0.141	0.255	0.002
1,000	0.004	0.023	0.000	0.002	0.060	0.103	0.001

In Tables A1 and A2, both the average bias and the mean-squared error decrease at the same rate as the increase in sample size. This confirms our asymptotic theory that the two-step estimator is root-n consistent. That the squared average bias converge

Table A2: Monte Carlo Results: Binary D (n=10)

G	Average Bias						
	α	β_0	β_X	β_D	γ_X	γ_D	σ_{uv}
250	-0.033	-0.051	-0.001	0.003	0.125	0.163	-0.001
500	-0.017	-0.029	-0.001	0.001	0.067	0.084	0.000
1,000	-0.008	-0.014	0.000	0.000	0.029	0.039	0.000
	MSE						
	α	β_0	β_X	β_D	γ_X	γ_D	σ_{uv}
250	0.032	0.132	0.001	0.003	0.461	0.844	0.002
500	0.009	0.049	0.000	0.002	0.134	0.226	0.001
1,000	0.004	0.021	0.000	0.001	0.055	0.094	0.001

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 obvious impact on estimation precision, especially in larger samples.

A.2 Continuous Endogenous Variable

In this subsection, we use a DGP with continuous endogenous covariates D. We adopt the same specification for the distribution of (X_i, Z_i) as in Appendix A.1. But Y_i is now generated by the reduced form in (8) and D_i is generated by (6). Both V_i and e_i are drawn from standard normal and $U_i = \rho V_i + e_i$. The true parameters are: $(\alpha, \beta_X, \gamma_X, \beta_D, \gamma_D, \rho) = (1/2, 1, 1, 1, 1, 2/3)$ and $\delta = (0, 1, 2)$. As before, we experiment with group sizes $n = 5$ and $n = 10$, and sample sizes $G = 250, 500, 1000$. We report average biases and MSE with 1,000 replications in Tables A3 and A4.

Table A3: Monte Carlo Results: Continuous D ($n = 5$)

G	Average Bias						
	α	β_0	β_X	β_D	γ_X	γ_D	ρ
250	-0.007	0.014	0.002	0.000	0.022	0.028	0.000
500	-0.002	0.006	0.001	0.000	0.005	0.009	0.001
1,000	-0.002	0.005	-0.001	0.000	0.007	0.009	0.000
	MSE						
	α	β_0	β_X	β_D	γ_X	γ_D	ρ
250	0.004	0.018	0.002	0.000	0.062	0.074	0.001
500	0.002	0.008	0.001	0.000	0.027	0.033	0.001
1,000	0.001	0.004	0.000	0.000	0.014	0.016	0.000

Table A4: Monte Carlo Results: Continuous D (n = 10)

G	Average Bias						
	α	β_0	β_X	β_D	γ_X	γ_D	ρ
250	-0.007	0.015	0.001	0.000	0.023	0.029	0.002
500	-0.004	0.010	-0.001	0.001	0.019	0.018	0.000
1,000	-0.002	0.005	0.000	0.000	0.009	0.010	0.000
	MSE						
	α	β_0	β_X	β_D	γ_X	γ_D	ρ
250	0.003	0.014	0.001	0.000	0.051	0.060	0.001
500	0.002	0.007	0.000	0.000	0.026	0.031	0.000
1,000	0.001	0.004	0.000	0.000	0.013	0.016	0.000

Similar to the case with discrete D_i, the simulation results are consistent with root-n convergence of the two-step estimator, and the group size does not seem to have obvious impact on estimation precision. Again, there is evidence that estimator variance is the dominating component in MSE. Overall, both the average bias and the mean-squared errors appear to be lower than those reported for the DGP with discrete D_i. This comparison is more obvious with smaller samples. We conclude that in this setup, the richer variation in the endogenous D_i has helped to increase estimation precision, once such endogeneity is dealt with using control functions.

B First-stage OLS Result

Table A5: First-stage OLS Results: Grade 2 Math Scores

Variable	Estimate	Standard Error
Female	−2.774	6.051
Average Female	−1.029	1.673
White	18.613**	7.209
Average White	2.875**	1.135
G2 Small Class	2.779	6.057
School Urbanicity	3.012	6.710
G2 Teacher Degree	−2.162	5.614
G2 Teacher Experience	−0.254	0.326
G2 Teacher STAR training	−13.244*	7.533
G2 Motivation	−0.091	0.338
Average G2 Motivation	0.679**	0.180
G2 Self-Concept	−0.576*	0.340
Average G2 Self-Concept	−0.184	0.132
Female × White	0.569	0.945
Female × G2 Small	−1.057	0.795
Female × Urbanicity	0.546	0.879
Female × G2 Degree	−1.192	0.764
Female × G2 Experience	−0.047	0.043
Female × G2 Training	0.006	1.038
Female × G2 Motivation	0.126	0.113
Female × G2 Self-Concept	−0.039	0.088
White × G2 Small	−0.084	1.012
White × Urbanicity	0.117	1.375
White × G2 Degree	1.192	0.988
White × G2 Experience	−0.012	0.056
White × G2 Training	0.413	1.331
White × G2 Motivation	−0.416**	0.139
White × G2 Self-Concept	0.099	0.111
G2 Small × Urbanicity	0.743	0.936
G2 Small × G2 Degree	−1.102	0.846
G2 Small × G2 Experience	−0.061	0.048
G2 Small × G2 Training	−1.169	1.150
G2 Small × G2 Motivation	−0.053	0.117
G2 Small × G2 Self-Concept	0.076	0.096
Urbanicity × G2 Degree	−0.183	0.921
Urbanicity × G2 Experience	0.114**	0.055
Urbanicity × G2 Training	2.812**	1.206
Urbanicity × G2 Motivation	0.064	0.132
Urbanicity × G2 Self-Concept	−0.072	0.104
G2 Degree × G2 Experience	−0.004	0.051
G2 Degree × G2 Training	0.073	1.082
G2 Degree × G2 Motivation	0.014	0.111
G2 Degree × G2 Self-Concept	0.041	0.088
G2 Experience × G2 Training	0.106*	0.057
G2 Experience × G2 Motivation	0.003	0.006
G2 Experience × G2 Self-Concept	0.001	0.005
G2 Training × G2 Motivation	0.066	0.152
G2 Training × G2 Self-Concept	0.172	0.121
G2 Motivation × G2 Self-Concept	0.009	0.007
Intercept	71.380**	17.279

R^2 = 0.081, F-statistic=8.527 with p-value 0.000.
**: 5% Significant; *: 10% Significant.