On the Estimation of Social Effects with Observational Network Data and Random Assignment

TszKin Julian Chan
Bates White Economic Consulting
Juan Estrada
Department of Economics, Emory University
Kim Huynh*
Currency Department, Bank of Canada
David Jacho-Chávez
Department of Economics, Emory University
Chungsang Tom Lam
Department of Economics, Clemson University
Leonardo Sánchez-Aragón
Facultad de Ciencias Sociales y Humanísticas, ESPOL University

August 5, 2022

Abstract

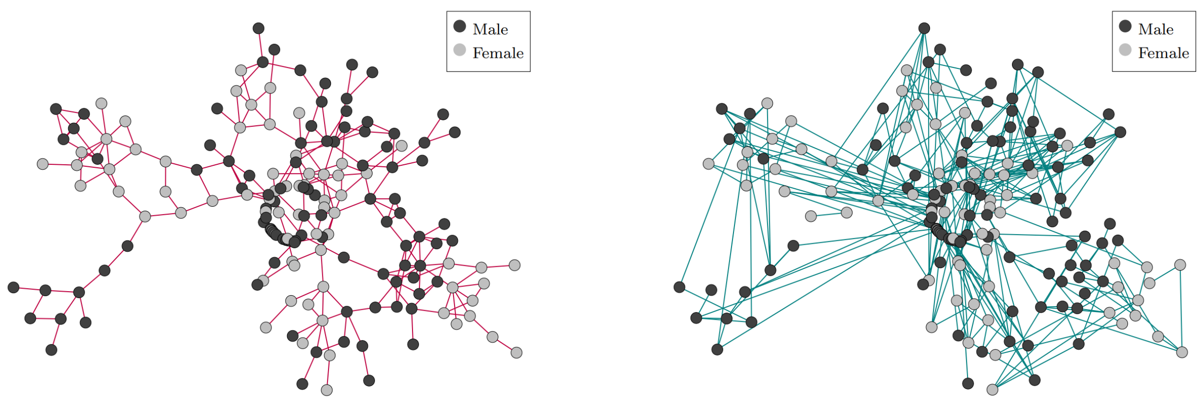
This paper proposes a new method to identify and estimate the parameters of the popular Linear-in-Means model of social effects in situations where initial randomization of peers induces the observed network of interest. We argue that the initially randomized peers do not generate social effects. However, after the randomization, agents can endogenously form relevant connections that can create peer influences. In this context, we introduce a moment condition that aggregates local heterogeneous identifying information for all the individuals in the population. We show that it is possible to identify the parameters of interest by using the exogenous variation in the randomized groups. We prove the consistency and root-n asymptotic normality of the resulting estimator of peer, contextual, and direct effects in the presence of ψ-dependence in the network space. The asymptotically efficient variance-covariance matrix is characterized and an estimator is proposed. A Monte Carlo experiment showcases the good small-sample properties of the estimator, while an application using networks of study partners and seatmates among high school students in Hong Kong is included. We find strong positive spillover effects of math test scores among study partners.

Keywords: Multidimensional Networks Data; ψ-dependence; Instrumental Variables; Causal Inference

*The views expressed in this article are those of the authors. No responsability for them should be attributed to the Bank of Canada. All remaining errors are the responsibility of the authors.

1 Introduction

This research provides a new way to estimate social effects from observational data consisting of characteristics and connections among individual-level observations in contexts in which individuals freely form the social relationships of interest after being exogenously assigned to peer groups, see, e.g., Carrell, Fullerton, and West (2009), Carrell, Hoekstra, and West (2011), and Carrell, Sacerdote, and West (2013). In our empirical application, we study the effects of study partners' performance on students' test scores for a sample of high school students in Hong Kong from March through May 2011. Using this data set, Figure 1a provides a visualization of the empirical challenge, while Figure 1b displays the information we propose here that can be used to solve the problem.



(a) Study partner network among 171 students

(b) Seatmate network among 171 students

Figure 1: Study partner and seatmate networks among a sample of 7th graders in Hong Kong in 2011

When individuals can freely choose who they associate themselves with (in this context, who they study with; i.e., Figure 1a), their decisions are influenced by observable characteristics (such as gender) as well as unobserved ones (such as latent homophily). This in turn can potentially create correlation between the average outcome and/or characteristics of peers with the error term in a linear model of social interactions; e.g., the so-called *linear-in-means* regression model (see Section 3.1 in de Paula, 2017, pp. 275-289). The network structure in Figure 1a presents evidence of gender homophily: male and female students are more likely to form study partner connections with other males and females, respectively. These homophily patterns can also exist when individuals connect based on unobserved attributes such as intrinsic abilities or extroversion, which creates an issue when estimating social effects.

On the other hand, let's suppose another set of connections among the same group of individuals is also observable (for instance, who they are assigned to share a desk with by their teacher; i.e., Figure 1b), and these connections are randomly imposed on the observational units (potentially, as part of an experimental design). In that case, this paper provides a new estimator that uniquely and consistently estimates social effects using these random peers as an *instrumental network*. As shown in Figure 1b, the exogenous assignment eliminates the clear gender homophily patterns, and by the virtues of randomization, the same should occur for unobserved homophily. The proposed estimator of the social effects turns out to be a linear Generalized-Method-of-Moments (GMM) estimator that has an explicit formula, converges at the standard rate, and has a multivariate normal limiting distribution with an efficient asymptotic variance-covariance matrix that takes into account the network dependence in the data. Specifically, the proposed inference here utilizes results in Kojevnikov, Marmer, and Song (2020)–that functionals of network data become uncorrelated as observations becomes distant in the network space; i.e., ψ–dependence. A Heteroskedasticity and Autocorrelation Consistent (HAC) estimator of the efficient asymptotic variance-covariance matrix is also proposed.

In the social sciences, randomly assigning peers has already been used to circumvent the endogeneity problem described above. The general approach is to assemble random groups such as classrooms, dorms, or squadrons and assume those are the relevant groups generating peer effects (Sacerdote, 2001; Carrell, Fullerton, and West, 2009; Carrell, Sacerdote, and West, 2013). It has been argued that these experimental designs produce peer effects estimators that measure the average changes in outcomes induced by randomly grouping people in settings of constant interpersonal interactions. However, as Carrell, Sacerdote, and West (2013) point out, this approach misses the possibility that people create relevant subnetworks of different types of connections, such as friendship or study partners, that are more relevant to generating network effects. This paper explicitly addresses this issue by demonstrating how an exogenous network can be used as an instrument. Our method does not require us to account for potentially unobserved network shocks in the randomized network in order to obtain consistent estimates of the social effects; therefore, we avoid the need to use pre-randomization values of outcomes and characteristics in the model specification (Advani and Malde, 2018). Also, since our method essentially uses the characteristics of observations at distance-p paths in the randomized network (seatmate of a seatmate when $p = 2$) as instruments for the endogenous social effects variables in the network of interest, the method can also be used when the source of endogeneity occurs because of a measurement error, such as misclassified or missing connections. Our simulation provides evidence of the usefulness of our proposed estimator in this setting.

Estimating social effects using a linear model of peer effects has been a fundamental approach to conducting empirical analysis in the social sciences see, e.g., Sacerdote (2001) in economics; Alexander et al. (2020) in public health; Kreager, Rulison, and Moody (2020) in criminology; and Salmivalli (2020) in sociology to mention just a few. Yet, network endogeneity continues to be a recurrent issue in empirical work evaluating social effects (Jackson, Rogers, and Zenou, 2017). Recent articles offering methods to alleviate the network endogeneity issue propose augmenting the standard linear-in-means model to include specific generating mechanisms for the network formation process, see e.g., Goldsmith-Pinkham and Imbens (2013), Qu and Lee (2015), Johnsson and Moon (2019) and Auerbach (2022). Specifying an additional network formation model generally involves imposing strong functional form and parametric assumptions (Graham and Pelican, 2020). This paper contributes by providing a computationally-simple estimator that does not require imposing extra structure to the model, which at the same time takes into account the potential resorting process after the initial exogenous configuration.

In our empirical application, we uncover strong positive spillover effects of math test scores among study partners–larger than alternative methods that ignore the potential endogeneity. The positive and significant peer effects in educational achievement are in line with previous results in the literature, such as Zimmerman (2003), Carrell, Fullerton, and West (2009), and Calvó-Armengol, Patacchini, and Zenou (2009). They all find strong peer effects based on different reference groups. The first two papers study social effects in the context of (quasi) random groups, while the third uses a friendship network. Although not directly comparable, our estimation results suggest a more significant level of peer effects in educational achievement. A potential reason is that we directly consider the network that endogenously results after randomization and our reference group is therefore likely to have a more direct impact on grades.

The remainder of this article is organized as follows. Section 2 introduces the necessary definitions, including distance metrics on graph space, as well as the general notation used throughout the paper. Section 3 then provides sufficient conditions under which social parameters can be uniquely recovered from observational data when information from an exogenous network is available. The proposed estimator and its asymptotic distributional properties under ψ–dependence are presented in Section 4. A simulation study is described in Section 5, while Section 6 provides an empirical illustration of the methods discussed by applying them to a data set from Hong Kong. Section 7 concludes with a discussion on various venues the proposed method can be used for experimental designs for example.

2 Preliminaries

We assume full observability of two types of networks. One is the endogenous network of interest that can create social externalities; the other is the instrumental network induced by random assignment. We use a *multilayered* network data structure to characterize our data-generating process (Boccaletti et al., 2014) and Kivela et al. (2014) for up-to-date comprehensive surveys and references therein. Following Boccaletti et al.'s (2014) definition, a multilayer network is a pair $\mathcal{M} = (\mathcal{N}, \mathcal{C})$, where $\mathcal{N} = \{\mathcal{G}_m; \quad m \in \{1, \ldots, M\}\}$ is a family of M graphs and $\mathcal{C}$ is the set of interconnections between nodes of different layers g_α and g_β with $\alpha \neq \beta$. When the same nodes are in each layer and there are no connections between different nodes in different layers except with themselves, these networks are called *multiplex* (Atkisson et al., 2020). The latter is the specific type of data structure used in this paper (see Figure 1). In our case, we have $M = 2$ and we denote $\mathcal{G}_N$ to be the population network of interest and $\mathcal{G}_{N,0}$ is the population instrumental network, where $N \subset N^+$ is the cardinality of the set $\mathcal{I}_n$ of individuals in the population, which is assumed to be arbitrarily large. Similarly, as in Graham (2020), the observed networks of size $n < N$ are denoted as $\mathcal{G}_n$ (see Figure 1a) and $\mathcal{G}_{n,0}$ (see Figure 1b), respectively, and are assumed to coincide with the subgraphs induced by a sample of nodes from their corresponding large population networks.

In the population, we represent the two layers $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$ in the multiplex network with the adjacency matrices of each layer; i.e., $\mathbf{W}_N = [w_{N;i,j}]$ and $\mathbf{W}_{N,0} = [w_{N,0;i,j}]$, where $w_{N;i,j}, w_{N,0;i,j} \in (0,1]$ are weights representing the importance of the (i,j) connection in each of the networks and $w_{N;i,j} = 0$ if i and j are not connected in $\mathcal{G}_N$. We do the same for $w_{N,0;i,j} = 0$ in $\mathcal{G}_{N,0}$. We also define the vectors $\mathbf{w}_{N,i} = [w_{N;i,1}, \ldots, w_{N;i,N}]^\top \in [0,1]^N$ and $\mathbf{w}_{N,0;i} = [w_{N,0;i,1}, \ldots, w_{N,0;i,N}]^\top \in [0,1]^N$ to be the ith row of the adjacency matrices $\mathbf{W}_N$ and $\mathbf{W}_{N,0}$, respectively. Adjacency matrices from the observed sample, $\mathbf{W}_n$ and $\mathbf{W}_{n,0}$, are defined and formed accordingly.

Section 4 uses the concept of $\psi-$dependence to bound the dependence among individuals as a function of their distance in the network space. Following the literature on graph theory, we use the shortest path length as our measure of distance; i.e., we denote $d_n(i,j)$ as the minimum path length connecting individuals i and j in the network $\mathcal{G}_n$ induced by the sample of size n. We define the following group of sets based on the geodesic distance $d_n(i,j)$. Let A and B be any two sets of individuals of sizes $a, b \in N_+$. Define the distance between sets as $d_n(A,B) = \min_{i \in A} \min_{j \in B} d_n(i,j)$. For a sample of size $n < N$, consider the following sets: (i) $\mathcal{P}_n^+(a,b,d) = \{(A,B) : A, B \subset \mathcal{I}_n, |A| = a, |B| = b, \text{ and } d_n(A,B) \geq d\}$ containing groups of nodes at a distance of at least d from each other; (ii) $\mathcal{P}_n^-(a,b,d) = \{(A,B) : A, B \subset \mathcal{I}_n, |A| = a, |B| = b, \text{ and } d_n(A,B) \leq d\}$ containing groups of nodes at a distance of at most d from each other; and (iii) $\mathcal{P}_n(a,b,d) = \{(A,B) : A, B \subset \mathcal{I}_n, |A| = a, |B| = b, \text{ and } d_n(A,B) = d\}$ the set associated with groups of nodes at distance d from each other. The associated set that contains all nodes at a certain distance of node i are $\mathcal{P}_n^+(i,d) = \{j \in \mathcal{I}_n : d_n(i,j) \geq d\}$, $\mathcal{P}_n(i,d) = \{j \in \mathcal{I}_n : d_n = d\}$, and $\mathcal{P}_n^-(i,d) = \{j \in \mathcal{I}_n : d_n \leq d\}$.

3 Peer Effects Model and Identification

The model in the population is formed by a set $\mathcal{I}_N$ of N agents, where $N \subset N^+$ can be arbitrarily large. In the population, each agent, i, is characterized by a set of K observable characteristics $\mathbf{x}_{N;i}$ and an unobserved idiosyncratic shock (error) $\varepsilon_{N;i}$. Agents in the population are connected by two types of networks $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$. In the data for our empirical study case, the network $\mathcal{G}_N$ represents the connections formed by two individuals who decide to study together, while $\mathcal{G}_{N,0}$ is the teacher-led exogenous seating assignment. Motivated by the empirical application, we assume that the network $\mathcal{G}_{N,0}$ is determined at random by an independent institution that is not trying to maximize either the agent's or any collective welfare function. However, the network $\mathcal{G}_N$ is formed endogenously by agents who are deciding whether or not to connect with each other to maximize some gain/loss function that potentially depends on others' observed and unobserved characteristics.

We assume that there is a population joint distribution determining the dependence patterns between the regressors, the networks, and the errors denoted by $\mathcal{F}(\mathbf{X}_N, \mathcal{G}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N)$, where $\mathbf{X}_N = [\mathbf{x}_{N;1}, \cdots, \mathbf{x}_{N;N}]^\top \in X$ and $\boldsymbol{\varepsilon}_N = [\varepsilon_{N;1}, \cdots, \varepsilon_{N;N}]^\top \in R^N$ represent the matrix of regressors and the vector of errors for all agents in $\mathcal{I}_N$. The joint distribution is characterized by three main features. First, by the properties of randomization, the network $\mathcal{G}_{N,0}$ is independent of the agents' observed and unobserved characteristics, $\mathbf{x}_{N;i} \in R^K$ and $\varepsilon_{N;i} \in R$, respectively. Second, agents randomly assign to the same group in $\mathcal{G}_{N,0}$ are more likely to form connections in $\mathcal{G}_N$ (Goldsmith-Pinkham and Imbens, 2013). Finally, in general, there could be agents in the population who do not have any connections in one or both networks. We call them *isolated* agents and we show here that they provide identifying information for both direct and social effects.

To incorporate the possibility that individual i can be isolated in one or both of the networks, let $\eta_{N;i}$ and $\eta_{N,0;i}$ be two Bernoulli random variables that equal one if individual i is non-isolated in the networks $\mathcal{G}_N$ or $\mathcal{G}_{N,0}$, respectively. It follows that $\kappa_{N;i} = E[\eta_{N;i}]$ and $\kappa_{N,0;i} = E[\eta_{N,0;i}]$ give the unconditional probability that individual i is non-isolated in the respective network, where the expectations are taken with respect to the marginal distributions of $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$, respectively. With this notation, the following assumption imposes restrictions on the joint distribution $\mathcal{F}$, which is consistent with the intuition in our empirical application.

Assumption 1 (Joint Distribution Characterization) *Consider the sets G, G_0, and X of all possible realizations of $\mathcal{G}_N$, $\mathcal{G}_{N,0}$ and $\mathbf{X}_N$ with a positive probability mass in $\mathcal{F}$, respectively. The following is true:*

(i) The distribution $\mathcal{H}(\mathbf{X}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N) = \int_{\mathcal{G}_N \in G} \mathcal{F}(\mathbf{X}_N, \mathcal{G}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N) d\mathcal{G}_N$ is such that $E[\mathbf{X}_N^\top \boldsymbol{\varepsilon}_N] = \mathbf{0}_K$, where $\mathbf{0}_K$ is a $K \times 1$ vector of zeros;

(ii) $\forall \mathcal{G}_{N,0} \in G_0$ and $\mathbf{X}_N \in X$, the conditional probability $\mathcal{F}(\mathcal{G}_N, \boldsymbol{\varepsilon}_N \mid \mathcal{G}_{N,0}, \mathbf{X}_N)$ is such that $\Pr(w_{N;i,j} > 0 | \mathcal{G}_{N,0}, \mathbf{X}_N) = \rho(w_{N,0;i,j}, \mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N)$, for some real-valued function $\rho : G_0 \times X \times R^N \to [0,1]$. Moreover, $\rho(0, \mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N) < \rho(1\{w_{N,0;i,j} > 0\}, \mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N)$ $\forall (\mathcal{G}_{N,0}, \mathbf{X}_N, \boldsymbol{\varepsilon}_N) \in G_0 \times X \times R^N$, where $1(\cdot)$ is the usual indicator function that equals one if its argument is true and zero otherwise;

(iii) $\forall \mathcal{G}_{N,0} \in G_0$ and associated marginal distributions, the random variables $\{\eta_{N,0;i}\}_{i=1}^N$ are such that the event $\eta_{N,0;i} = 0$, $\forall i \in \mathcal{I}_N$ happens with probability zero; and (iv) $\forall \mathcal{G}_N \in G$ and associated marginal distributions, the random variables $\{\eta_{N;i}\}_{i=1}^N$ are such that the event $\eta_{N;i} = 0$, $\forall i \in \mathcal{I}_N$ happens with probability zero.

Assumption 1(i) does not impose any restrictions on the correlation among $\mathcal{G}_N$, $\mathbf{X}_N$, and $\boldsymbol{\varepsilon}_N$. In particular, this assumption allows for an endogenous network formation process where agents can create connections in $\mathcal{G}_N$ based on observed and unobserved characteristics, which could also induce a correlation between $\mathbf{X}_N$ and $\boldsymbol{\varepsilon}_N$ due to homophily. For instance, two students with similar levels of unobserved abilities may be more likely to want to study together. Importantly, the random assignment defining $\mathcal{G}_{N,0}$ guarantees that the network is independent of the regressors and the errors and that conditioning on it is not necessary. Assumption 1(ii) imposes the condition that the probability of agents i and j being connected in $\mathcal{G}_N$ increases when they are connected in $\mathcal{G}_{N,0}$. However, we are agnostic about the dependence structure in the implicit network formation process for which we do not impose any explicit functional form. In particular, this assumption can accommodate the pairwise independent network formation models as in Graham (2017), as well as the network formation models with strategic interactions as in de Paula, Richards-Shubik, and Tamer (2018) and Graham and Pelican (2020). Assumptions 1(iii) and 1(iv)

are necessary conditions for the model to be identified and exclude the possibility that $\mathbf{W}_{N,0} = \mathbf{W}_N = \mathbf{O}_N$ (the $N \times N$ matrix of zeros).

The following assumption corresponds to a linear model of peer effects, which is the workhorse specification in the social sciences (de Paula, 2017). It allows the network data to have a structural interpretation as the best response function of a Bayesian game of social interactions as shown by Blume et al. (2015). As discussed earlier, it encapsulates the idea that only the connections that are formed optimally by agents can create peer effects.

Assumption 2 (Linear Model) *The optimal choice (outcome), $y_{N;i}$, for agent i is characterized by*

$$y_{N;i} = \beta_0 \sum_{j \neq i} w_{N;i,j} y_{N;j} + \sum_{j \neq i} w_{N;i,j} \mathbf{x}_{N;j}^\top \boldsymbol{\delta}_0 + \widetilde{\mathbf{x}}_{N;i}^\top \boldsymbol{\gamma}_0 + \varepsilon_{N;i}, \tag{3.1}$$

where $\widetilde{\mathbf{x}}_{N;i} = [1, \mathbf{x}_{N;i}^\top]^\top$, $\boldsymbol{\theta}_0 = (\beta_0, \boldsymbol{\delta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$ belongs to the interior of the parameter space $\boldsymbol{\Theta} \subset R^{2K+2}$, which is assumed to be compact.

Assumption 2 is a linear model for social effects and it effectively imposes an exclusion restriction on the network $\mathcal{G}_{N,0}$. The intuitive motivation for this assumption is that agents tend to behave similarly to others with whom they have a close social relationship. In particular, we argue that a group of people randomly assigned together into a group are not likely to generate peer effects. After randomization, agents can form endogenous connections that create social effects. This intuition has found support in the literature. Carrell, Sacerdote, and West (2013) find that groups optimally designed to improve academic performance ended up having a negative effect because of the role of endogenous link formation. In our empirical application, Assumption 2 implies that randomly sitting next to another student does not generate knowledge spillovers, but having a study partner relationship does create such spillovers. The coefficients β_0 and $\boldsymbol{\delta}_0$ are known as the peer effects and the contextual effects, respectively, and are referred jointly as the social effects parameters hereafter. The assumption that the parameters $\boldsymbol{\theta}_0$ are in the interior of the parameter space is particularly relevant for the coefficient β_0, because (3.1) has a solution in terms of $w_{N;i}$, $\mathbf{X}_N$, and $\varepsilon_{N;i}$ only when $\beta_0 < 1/\lambda_{\max}$, where $\lambda_{\max}$ is the largest eigenvalue of $\mathbf{W}_N$. Assuming $K = 1$ and that there is no constant for the sake of illustration, Assumption 2 implies that the peer effects regressor can be written as

$$\mathbf{W}_N\mathbf{y}_N = \gamma_0\mathbf{W}_N\mathbf{x}_N + (\gamma_0\beta_0 + \delta_0)\sum_{p=0}^{\infty}\beta_0^p\mathbf{W}_N^{p+2}\mathbf{x}_N + \sum_{p=0}^{\infty}\beta_0^p\mathbf{W}_N^{p+1}\varepsilon_N, \tag{3.2}$$

which under the condition that $\gamma_0\beta_0+\delta_0 \neq 0$ shows that in principle, powers of the adjacency matrix $\mathbf{W}_N$ could be used to instrument $\mathbf{W}_N\mathbf{y}$ (Bramoullé, Djebbari, and Fortin, 2009; Manta et al., 2022). In the case at hand, this approach is not possible because the network $\mathcal{G}_N$ can be endogenous. However, note that from Assumptions 1 (i) and (ii), powers of the adjacency matrix $\mathbf{W}_{N,0}$ are natural candidates to replace $\mathbf{W}_N$.

We propose to use the random assignment embodied in $\mathbf{W}_{N,0}$ to identify the parameters of a linear model defined on the network space spanned by $\mathbf{W}_N$ in Assumption 2. To formalize our identification strategy, we define $\mathbf{D}_N = [\mathbf{W}_N\mathbf{y}_N, \mathbf{W}_N\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$ to be the matrix of regressors in the matrix-notation counterpart of Equation (3.1) and $\mathbf{Z}_N = [\mathbf{W}_{N,0}^p\mathbf{X}_N, \mathbf{W}_{N,0}^{p-1}\mathbf{X}_N, \ldots, \mathbf{W}_{N,0}\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$ to be the matrix producing the moment conditions formed based on Assumptions 1 (i) and (ii), where $p > 1$ is a constant parameter representing the powers of the adjacency matrix used as instruments. This framework allows for the option of using the characteristics of the so-called *connections of connections*' as instruments by letting $p = 2$. The flexibility of Assumption 1 also allows for the use of more indirect connections' characteristics that are at distance $p > 2$. For instance, in our empirical application, setting $p = 2$ means that we are using seatmates of seatmates as the instrumental network and setting $p = 4$ means that we use seatmates at a distance of four as the instrument. To take into account the fact that individuals without connections at distance p may not provide identifying information for the network effects parameters, define $\eta_{N,0;i}^p$ to be a random variable that equals one if individual i has at least one connection at distance p and zero otherwise. As before, $\kappa_{N,0;i}^p = E[\eta_{N,0;i}^p]$ is the unconditional probability that individual i has at least one connection at distance p.

Define $\boldsymbol{H}_{N,0;i} = \mathrm{diag}(\eta_{N,0;i}^p, \ldots, \eta_{N,0;i}, 1, \ldots, 1)$ to be a $((p+1)K+1) \times ((p+1)K+1)$ matrix where the first K elements contain the random variables determining whether or not individual i has at least one connection at distance p and the second K elements the random variables determining whether or not individual i has at least one connection at distance $p-1$, and so on, until the last K elements associated with $\mathbf{W}_{N,0}\mathbf{X}_N$, where $\eta_{N;i}$ is the random variable determining whether i is isolated in the network $\mathcal{G}_{N,0}$. Finally, the last $K+1$ elements in the lower-right sub-matrix, which coincide with the non-network regressors $\widetilde{\mathbf{x}}_{N;i}$, are ones. Define the $((p+1)K+1)\times((p+1)K+1)$ matrix $\boldsymbol{K}_{N,0;i} = \operatorname{diag}(\kappa^p_{N,0;i},\ldots,\kappa_{N,0;i},1,\ldots,1)$ to be $E[\boldsymbol{H}_{N,0;i}]$.

From Assumption 1(i), it follows that $E[\mathbf{z}_{N;i}\varepsilon_{N;i}] = 0$ for all i, where $\mathbf{z}_{N;i}$ be the ith row of the matrix $\mathbf{Z}_N$. Defining the function $\boldsymbol{m}_N(\boldsymbol{\theta}) := \sum_{i\in\mathcal{I}_N}\mathbf{z}_{N;i}\varepsilon_{N;i}$, it follows that $E[\boldsymbol{m}_N(\boldsymbol{\theta}_0)] = \mathbf{0}$. The following remark about the moment condition $E[\boldsymbol{m}_N(\boldsymbol{\theta}_0)]$ is noteworthy. By the law of total expectation, $E[\mathbf{z}_{N;i}\varepsilon_{N;i}] = \boldsymbol{K}_{N,0;i}E[\mathbf{z}_{N;i}\varepsilon_{N;i} \mid \boldsymbol{H}^*_{N,0;i} \neq \boldsymbol{O}_{pK}] + (\mathbf{I}_{(p+1)K+1} - \boldsymbol{K}_{N,0;i})E[\mathbf{z}_{N;i}\varepsilon_{N;i} \mid \boldsymbol{H}^*_{N,0;i} = \boldsymbol{O}_{pK}]$, where $\boldsymbol{H}^*_{N,0;i}$ contains the left top $(pK\times pK)$ upper matrix of $\boldsymbol{H}_{N,0;i}$ and $\boldsymbol{O}_{pK}$ is the $pK\times pK$ zero metrix. Note that when $\boldsymbol{H}^*_{N,0;i} = \boldsymbol{O}_{pK}$, the conditional expectation $E[\mathbf{z}_{N;i}\varepsilon_{N;i} \mid \boldsymbol{H}_{N,0;i} = \boldsymbol{O}_{pK}]$ is trivially zero for the first pK elements and are one for the last $k+1$ positions. Moreover, $\mathbf{I}_{(p+1)K+1} - \boldsymbol{H}_{N,0;i}$ equal zero in the last $k+1$ elements of its diagonal. Therefore, it follows that $E[\mathbf{z}_{N;i}\varepsilon_{N;i}] = \boldsymbol{K}_{N,0;i}E[\mathbf{z}_{N;i}\varepsilon_{N;i} \mid \boldsymbol{H}_{N,0;i} \neq \boldsymbol{O}]$ and thus $E[\boldsymbol{m}_N(\boldsymbol{\theta}_0)] = \sum_{i\in\mathcal{I}_N}\boldsymbol{K}_{N,0;i}E[\mathbf{z}_{N;i}\varepsilon_{N;i} \mid \boldsymbol{H}_{N,0;i} \neq \boldsymbol{O}]$. This remark is relevant because it gives an interpretation to the unconditional moment condition as a weighted sum where the weights are the probabilities of having distance-p connections and not being isolated. This scheme accommodates heterogeneity on the identifying power of each individual in the population and gives more importance for agents who are unlikely to not have distance-p connections or to be isolated in the randomized network in the population.

The use of the two networks $\mathcal{G}_{N,0}$ and $\mathcal{G}_N$ for identification requires a relevance condition that guarantees the two networks have enough overlap. The moment characterising the correlation between regressors and instruments can be written as $E[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N;i}\mathbf{d}^\top_{N;i}]$, where $\mathbf{d}_{N;i}$ and $\mathbf{z}_{N;i}$ represent the i rows of matrices $\mathbf{D}_N$ and $\mathbf{Z}_N$, respectively. Importantly, we do not assume that $E[\mathbf{z}_{N;i}\mathbf{d}^\top_{N;i}]$ are equal for all i. The following Assumption imposes a rank condition related with the strength of the correlation between $\mathbf{z}_{N;i}$ and $\mathbf{d}_{N;i}$. The conditions in Assumptions 1(ii) and 1(iii) are necessary, but not sufficient for the following assumption to hold.

Assumption 3 (Relevance) *The matrix* $E[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N;i}\mathbf{d}_{N;i}^{\top}]<\infty$ *has full column rank.*

Assumption 3 imposes restrictions on the product matrices $\mathbf{W}_{N,0}\mathbf{W}_N$, and $\mathbf{W}_{N,0}^p$. As mentioned before, if the product $\mathbf{W}_{N,0}\mathbf{W}_N=\mathbf{O}_N$, identification breaks down. This condition has three important empirical consequences: (1) There should exist nodes in the system that share connections in common from the two networks $\mathcal{G}_{N,0}$ and $\mathcal{G}_N$ (seatmates that are also study partners), (2) It should be possible to connect two nodes with a path formed by an edge from network $\mathcal{G}_{N,0}$ followed by one from the network $\mathcal{G}_N$ (or vice-versa), and (3) The matrices $\mathbf{I}_N$, $\mathbf{W}_{N,0},\ldots,\mathbf{W}_{N,0}^p$ have to be linearly independent (the network cannot be composed by fully connected groups of the same size). Under the conditions stated before, the following Theorem shows that identification of the parameters of interest in (3.1) is possible, even if the network $\mathbf{W}_N$ is endogenous.

Theorem 1 (Identification) *Let Assumptions 1, 2, and 3 hold, then* $E[\boldsymbol{m}_N(\boldsymbol{\theta})]=\mathbf{0}_K$ *if and only if* $\boldsymbol{\theta}=\boldsymbol{\theta}_0$, *where* $\boldsymbol{m}_N(\boldsymbol{\theta}):=\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N;i}\varepsilon_{N;i}$.

Appendix A presents the proof for Theorem 1. This Theorem shows that identification is possible in a context where the network of interest $\mathcal{G}_N$ is formed endogenously by taking advantage of the randomization and exclusion restrictions on the network $\mathcal{G}_{N,0}$. This approach allows us to attach a causal interpretation to the estimated parameters of a linear model of peer effects, which uses observational network data that emerges after an initial randomization. This method can be used to address research designs with randomized peers such as that in Carrell, Sacerdote, and West (2013).

4 Estimation

We propose a Generalized-Method-of-Moments (GMM) estimator based on the identifying moment condition in Theorem 1. We assume that the analyst observes a sample of size $n<N$ from the population described in the previous section. In our sample scheme, n agents are chosen at random without replacement and their observed characteristics, outcome, and connections in $\mathcal{G}$ and $\mathcal{G}_0$ are recorded. Therefore, the random sample consists of the observations $\{y_{n;i},\mathbf{x}_{n;i}^{\top},\{w_{n;i,j},w_{n,0;i,j}\}_{j=1,j\neq i}^{n}\}_{i=1}^{n}$ from where it is possible to calculate

the $n \times (2K+2)$ matrix of regressors $\mathbf{D}_n$ and the $n \times (3K+1)$ matrix of *instruments* $\mathbf{Z}_n$ (depending on the value of K, the system can be *just*- or *over*-identified). The population's GMM objective function is given by $J_N(\boldsymbol{\theta}) = E[\boldsymbol{m}_N(\boldsymbol{\theta})]^\top \mathbf{A}_N E[\boldsymbol{m}_N(\boldsymbol{\theta})]$, where $\mathbf{A}_N$ is a constant full rank weighting matrix $\mathbf{A}_N$. The GMM estimator of $\boldsymbol{\theta}$ is defined as $\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} J_n(\boldsymbol{\theta})$, where $J_n(\boldsymbol{\theta}) \equiv (n^{-1}\sum_{i\in\mathcal{I}_n} \mathbf{z}_{n;i}\varepsilon_{n;i})^\top \mathbf{A}_n (n^{-1}\sum_{i\in\mathcal{I}_n} \mathbf{z}_{n;i}^\top \varepsilon_{n;i})$, the $(3K+1) \times (3K+1)$ full rank weighting matrix $\mathbf{A}_n$ is assumed to converge in probability to $\mathbf{A}_N$. The linearity in (3.1) guarantees that the GMM estimator has a closed form solution given by

$$\widehat{\boldsymbol{\theta}}_{\text{GMM}} = [\mathbf{D}_n^\top \mathbf{Z}_n \mathbf{A}_n \mathbf{Z}_n^\top \mathbf{D}_n]^{-1} [\mathbf{D}_n^\top \mathbf{Z}_n \mathbf{A}_n \mathbf{Z}_n^\top \mathbf{y}_n]. \tag{4.1}$$

To allow for the possibility that individuals' observed and unobserved characteristics are correlated in the population's joint distribution $\mathcal{F}(\mathbf{X}_N, \mathcal{G}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N)$, we use the concept of Doukhan and Louhichi's (1999) $\psi-$dependence (Kojevnikov, Marmer, and Song, 2020; Estrada, 2021). That is, we bound the correlation between non-linear functions of random variables with the *dependence coefficients*, which are decreasing functions of the network distance. From Assumption 1, the network $\mathcal{G}_{N,0}$ is independent of the regressors and the errors. Thus, we can rule out a dependence structure based on that network. However, because individuals endogenously form connections in $\mathcal{G}_N$ based on observed and unobserved characteristics, we would expect relatively high levels of dependence between individuals close to each other in the network space spanned by $\mathcal{G}_N$. For instance, in our empirical application, we expect the observed and unobserved characteristics of students who study together or are indirectly connected by study partners to be more correlated than that of students who are not study partners and are not indirectly connected.

Following Estrada (2021), we define the random vector $\mathbf{r}_{N;i} \equiv [\mathbf{x}_{N;i}^\top, \varepsilon_{N;i}]^\top \in R^{K+1}$. For $K, a \in N_+$, we endow $R^{(K+1)\times a}$ with the distance measure $\boldsymbol{d}_a(\mathbf{x}, \mathbf{y}) = \sum_{l=1}^a \|x_l - y_l\|_2$, where $\|\cdot\|_2$ denotes the Euclidean norm and $\mathbf{x}, \mathbf{y} \in R^{(K+1)\times a}$. Let $L_{K,a}$ denote the collection of bounded Lipschitz real functions mapping values from $R^{(K+1)\times a}$ to R. For each set of nodes A, let $\mathbf{r}_{N,A} = (\mathbf{r}_{N;i})_{i\in A}$. Hereafter, we write triangular arrays simply as $\{\mathbf{r}_{N;i}\}$, and sequences like $\{\lambda_n\}_{n\geq 1}$ as $\{\lambda_n\}$. Following Doukhan and Louhichi (1999) and Kojevnikov, Marmer, and Song (2020), $\psi-$dependence can be defined.

Definition 4.1 (ψ-dependence) *A triangular array* $\{\mathbf{r}_{n;i}\}$, $n \geq 1$, $\mathbf{r}_{n;i} \in R^{K+1}$ *is* ψ-*dependent if for each* $n \in N_+$ *there exists a sequence* $\{\lambda_n\} \equiv \{\lambda_{n,d}\}_{d \geq 0}, \lambda_{n,0} = 1$ *and a collection of non-random functions* $(\psi_{a,b})_{a,b \in N}, \psi_{a,b} : L_{v,a} \times L_{v,b} \to [0, \infty)$, *such that for all* $A, B \in \mathcal{P}_N^+(a, b, d)$ *for* $d > 0$ *and all* $f \in L_{Q+1,a}$ *and* $g \in L_{Q+1,b}$,

$$|\text{cov}\left(f\left(\mathbf{r}_{n,A}\right), g\left(\mathbf{r}_{n,B}\right)\right)| \leq \psi_{a,b}(f, g)\lambda_{n,d}.$$

The sequence $\{\lambda_n\}$ is called the *dependence coefficients* of $\mathbf{r}_{n;i}$. As mentioned before, the covariance of the non-linear functions of the random vectors $\mathbf{r}_{n,A}$ and $\mathbf{r}_{n,B}$ are bounded by the dependence coefficients $\lambda_{n,d}$ and a functional $\psi_{a,b}(f, g)$, which depends on the size of the sets A and B, and the aggregating non-linear functions f and g. Importantly by choosing appropriate values for the functions f and g, the ψ-dependence framework allows us to bound the dependence between observed and unobserved characteristics among any set of individuals. As in Estrada (2021), we impose a weak dependence assumption to control the network dependence between individuals. However, we do not impose a sharp bound on the decreasing pattern of the dependence coefficients with respect to the network distance.

Assumption 4 (Weak Dependence) *Consider the set* G *of all possible realizations of* $\mathcal{G}_N$ *with a positive probability mass in* $\mathcal{F}$. *For all networks* $\mathcal{G}_N \in G$, *the conditional distribution* $\mathcal{F}(\mathbf{X}_N, \boldsymbol{\varepsilon}_{\boldsymbol{N}} \mid \mathcal{G}_N)$ *is such that:*

(i) $\{\mathbf{r}_{N;i}\}$ *is* ψ-*dependent with dependence coefficients* λ_N;

(ii) for a generic constant $C > 0$, $\psi_{a,b}(f, g) \leq C \times ab\left(\|f\|_\infty + \text{Lip}(f)\right)\left(\|g\|_\infty + \text{Lip}(g)\right)$;

(iii) and for each $n \in N_+$, $\max_{d \geq 1} \lambda_{n,d} < \infty$ *and* $\lim_{d \to \infty} \lambda_{n,d} = 0$.

We impose Assumption 4(i) on the population, and it is inherited by the random vector $\mathbf{r}_{N;i} = [\mathbf{x}_{N;i}^\top, \varepsilon_{N;i}]^\top$. Given that any sample of size n is taken from the population characterized by the joint distribution $\mathcal{F}$, condition (i) applies for any $n \in N_+$. Condition (ii) bounds the functional $\psi_{a,b}(f, g)$ by an arbitrary constant C, the cardinality of the sets A and B, and the sup-norm and Lipschitz constants of the aggregating functions f and g. Intuitively, if the Lipschitz constants $\text{Lip}(f)$ and $\text{Lip}(g)$ increase, the values of the functions f and g can be larger for some values of $\mathbf{r}_{n,A}$ and $\mathbf{r}_{n,B}$, which requires larger constants to bound the covariance. This intuition is similar for the sup-norm. Finally, condition (iii) requires that the dependence coefficients are finite for any value of d and that they dissipate to zero for a large enough network distance between the random vectors $\mathbf{r}_{n,A}$ and $\mathbf{r}_{n,B}$.

The use of the $\psi-$dependence framework in modeling network dependence has the advantage that it does not impose functional form restrictions on the errors and it allows for correlation between indirectly connected nodes. However, transformations of ψ-dependent random variables are not necessarily ψ-dependent. Therefore, in order to analyze the asymptotic behavior of the estimator $\widehat{\boldsymbol{\theta}}_{\text{GMM}}$, we need to impose bounds to covariances of the form $\text{cov}(r_{n;i,q}r_{n;j,\ell}, r_{n;h,q'}r_{n;s,\ell'})$, where $i, j, h, s \in \mathcal{I}_n$, q, q', ℓ, and ℓ' are components of the vector $\mathbf{r}_{n;i}$. These include covariances such as $\text{cov}(\varepsilon_{n;i}\varepsilon_{n;j}, \varepsilon_{n;h}\varepsilon_{n;s})$ or $\text{cov}(x_{n;i,q}x_{n;j,\ell}, \varepsilon_{n;h}\varepsilon_{n;s})$, for example.

Assumption 5 (Bound Covariances) *Define functions $f_{q,\ell}$ and $g_{q',\ell'}$ mapping $R^{(Q+1)\times 2}$ into R to be such that $f_{q,\ell}(\mathbf{r}_{n;\{i,j\}}) = r_{n;i,q}r_{n;j,\ell}$ and $g_{q',\ell'}(\mathbf{r}_{n;\{h,s\}}) = r_{n;h,q'}r_{n;s,\ell'}$ for $i, j, h, s \in \mathcal{I}_n$, $i \neq j$, $h \neq s$, $q \neq \ell$ and $q' \neq \ell'$. The norms $\|f_{q,\ell}(\mathbf{r}_{n;\{i,j\}})\|_{p_f^*} + \|g_{q',\ell'}(\mathbf{r}_{n;\{h,s\}})\|_{p_g^*} < \infty$ for all q, ℓ where $p_f^* = \max\{p_{f,i}, p_{f,j}\}$ (analogous for p_g^*) and $1/p_{f,i}+1/p_{f,j}+1/p_{g,h}+1/p_{g,s} < 1$.*

Assumption 5 provides sufficient conditions for the functions $f_{q,\ell}$ and $g_{q',\ell'}$ of ψ-dependent random variables to have bounded covariances. The weak dependence in Assumption 4 guarantees that the dependence coefficients vanish to zero when the network distance increases. However, the network distance $d_n(i, j)$ between any two individuals i and j is also a function of the sample size. Therefore, the asymptotic behavior of the dependence coefficients $\lambda_{n,d}$ depends on the asymptotic behavior of the network features determining the distance between nodes. In particular, the density of the network is explicitly linked with the geodesic distance. When the network density is arbitrarily large, the geodesic distance is always one for any pair of nodes. Therefore, as noted by Kojevnikov, Marmer, and Song (2020), there is a trade-off between network density and the rate of convergence of the dependence coefficients. Networks with higher density would require the dependence to decrease faster (and vice-versa). The following Assumption provides a necessary condition on the dependence coefficients for a Law of Large Numbers to apply.

Assumption 6 (Dependence Rate of Decay) *Let $\bar{D}_n(d) \equiv n^{-1} \sum_{i \in \mathcal{I}_n} |\mathcal{P}_n(i,d)|$ be the average number of distance-d connections on the network $\mathcal{G}_n$, such that $n^{-1} \sum_{d \geq 1} \bar{D}_n(d)$ $\lambda_{n,d} \xrightarrow{a.s.} 0$ as $n \longrightarrow \infty$.*

This Assumption is similar in spirit to Assumption 3.2 in Kojevnikov, Marmer, and Song (2020), but it includes the fact that in our setting the analyst has access to two types of networks that can involve weighted adjacency matrices. The following Assumption imposes the existence of moments for products of ψ-dependent random variables and it is a regularity condition needed for a Law of Large Numbers to apply.

Assumption 7 (Existence of Moments) $\exists \epsilon > 0$, $\sup_{n \geq 1} \max_{i \in \mathcal{I}_n} \|R_{n;i,j}\|_{1+\epsilon} < \infty$ *a.s., where $R_{n;i,j} \equiv r_{n;i,q} r_{n;j,\ell}$ and $\|R_{n;i,j}\|_p \equiv (E[|R_{n;i,j}|^p])^{1/p}$.*

The previous Assumptions are enough to guarantee that the Law of Large Numbers applies for products of ψ-dependent random variables. To show asymptotic normality, we are using the Central Limit Theorem result in Kojevnikov, Marmer, and Song (2020). As mentioned before, for the asymptotic moments of network-dependent random variables to be well-defined, we need to control the level of asymptotic density. In particular, following Kojevnikov, Marmer, and Song (2020), we define a measure for the average neighborhood size as $\bar{D}_n(d,k) = 1/n \sum_{i \in \mathcal{I}_n} |\mathcal{P}_n(i,d)|^k$ and a measure for the average neighborhood shell size as

$$\bar{D}_n(d,m,k)^- = \frac{1}{n} \sum_{i \in \mathcal{I}_n} \max_{j \in \mathcal{P}_n(i,d)} \left|\mathcal{P}_n^-(i,m) \setminus \mathcal{P}_n^-(j,d-1)\right|^k,$$

were $\mathcal{P}_n^-(j,d-1) = \emptyset$ when $d = 0$. With these two measures of average density, construct the combined quantity

$$c_n(d,m,k) = \inf_{\alpha > 1} \left[\bar{D}_n(d,m,k\alpha)^-\right]^{\frac{1}{\alpha}} \left[\bar{D}_n\left(d, \frac{\alpha}{\alpha - 1}\right)\right]^{1-\frac{1}{\alpha}}. \tag{4.2}$$

For some arbitrary position q in the matrix $\mathbf{Z}_{n;i}$, let $S_n = \sum_{i \in \mathcal{I}_n} z_{n;i,q} \varepsilon_{n;i}$. Defining $\sigma_{n,q}^2 = \mathrm{var}(S_n)$, the following Assumption guarantees the existence of higher order moments, imposes asymptotic sparsity, and bounds the long-run variance.

Assumption 8 (Average Sparsity) *For all networks $\mathcal{G}_n \in G$, (i) for some $p > 4$, $\sup_{n\geq 1} \max_{i\in\mathcal{I}_n} \|z_{n;i,q}\varepsilon_{n;i}\|_p < \infty$. There exists a sequence $m_n \to \infty$, such that for $k = 1, 2$, (ii) , $\frac{n}{\sigma_{n,q}^{2+k}} \sum_{d\geq 0} c_n(d, m_n, k)\, \lambda_{n,d}^{1-\frac{2+k}{p}} \xrightarrow{a.s.} 0$ as $n \to \infty$, (iii) $\frac{n^2 \lambda_{n,m_n}^{1-(1/p)}}{\sigma_{n,q}} \xrightarrow{a.s.} 0$ as $n \to \infty$.*

These conditions impose a rate of convergence of the dependence coefficients $\lambda_{n,d}$ that is related to the network density. There is a trade-off in which a higher density requires a higher speed in the dependence-decreasing patterns. The previous Assumptions are sufficient to show that our GMM estimator is consistent and asymptotically normal. Moreover, Lemma B.3 in Appendix B shows that $\mathbf{\Omega}_n = \mathrm{var}(\mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n)$ converges to the finite population variance

$$\mathbf{\Omega}_N = \lim_{n\to\infty} n^{-1} \left[\sum_{i=1}^{n} \mathrm{var}(\mathbf{z}_{n;i}\varepsilon_{n;i}) + \sum_{i\neq j} \mathrm{cov}(\mathbf{z}_{n;i}\varepsilon_{n;i}, \mathbf{z}_{n;j}\varepsilon_{n;j}) \right] \equiv \sum_{d\geq 0} \mathbf{\Gamma}_N(d) < \infty, \tag{4.3}$$

where $\mathbf{\Gamma}_N(d) = \sum_{i\in\mathcal{I}_N} \sum_{j\in\mathcal{P}_N(i,d)} E[\mathbf{z}_{N;i}\varepsilon_{N;i}\varepsilon_{N;j}\mathbf{z}_{N;j}^\top]$ are the covariances between random variables of individuals at distance d. Therefore, the variance-covariance matrix $\mathbf{\Omega}_N$ can be calculated by summing the covariances for all possible distances $d \geq 0$. After characterizing $\mathbf{\Omega}_N$, Theorem 2 provides the asymptotic behavior of (4.1).

Theorem 2 *Let Assumptions 1 to 8 hold, then as $n \to \infty$, $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} = \boldsymbol{\theta} + o_p(1)$ and $\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}} - \boldsymbol{\theta}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_N)$, where $\mathbf{\Sigma}_N \equiv (E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]^\top \mathbf{A}_N E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top])^{-1} \times (E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]^\top \mathbf{A}_N \mathbf{\Omega}_N \mathbf{A}_N \times E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top])(E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]^\top \mathbf{A}_N E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top])^{-1}$, and when $\mathbf{A}_N = \mathbf{\Omega}_N^{-1}$, then*

$$\mathbf{\Sigma}_N = (E[\Sigma_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]^\top \mathbf{\Omega}_N^{-1} E[\Sigma_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top])^{-1}. \tag{4.4}$$

Recall that the expectation $E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$ can be written as $\boldsymbol{K}_{N;0,i} E[\mathbf{z}_{N;i}\varepsilon_{N;i} \mid \boldsymbol{H}_{N;0,i}^* \neq \boldsymbol{O}_{pK}]$. Therefore, Equation (4.4) depends on the population probabilities of an individual providing identification information. For low values of those probabilities, the upper-right sub-matrix of $\boldsymbol{K}_{N;0,i}$ approaches the zero matrix and the variance-covariance matrix could grow arbitrarily large. In the extreme case of non-identification, (4.4) diverges to infinity. Theorem 2 exposes a relationship between the network parameters' precision and network sparsity.

4.1 Efficient Weight Matrix Estimation

To construct the efficient version of the proposed GMM estimator we need a consistent estimator of $\mathbf{\Omega}_N$. Here we use Kojevnikov, Marmer, and Song's (2020) network heteroskedasticity and autocorrelation-consistent (HAC) variance estimator. Let D_n represent a bandwidth after which the dependence between individuals vanishes. For example, Kojevnikov, Marmer, and Song (2020) proposes $D_n = C \times [\log(\text{average degree} \vee (1 + 0.05))]^{-1} \times \log n$ and this rule-of-thumb is used in our Monte Carlo simulations and empirical study below with $C = 1.8$ The proposed variance-covariance matrix estimator is then given by

$$\widetilde{\mathbf{\Omega}}_n = \sum_{d \geq 0} K(d/D_n) \frac{1}{n} \sum_{i=1}^{n} \sum_{j \in \mathcal{P}_n(i,d)} \mathbf{z}_{n;i} \widetilde{\varepsilon}_{n;i} \widetilde{\varepsilon}_{n;j} \mathbf{z}_{n;j}^{\top}, \tag{4.5}$$

where $\widetilde{\varepsilon}_{n;i} = y_{n;i} - \mathbf{d}_{n;i}^{\top} \widetilde{\boldsymbol{\theta}}_{\text{GMM}}$; $K(\cdot)$ is a kernel (weighting) function, such that $K(0) = 1$ and $K(u) = 0$ for $u > 1$; and $\widetilde{\boldsymbol{\theta}}_{\text{GMM}}$ is a preliminary consistent estimator. In (4.1) with $\mathbf{A}_n$ equal the identity matrix or $n^{-1}\mathbf{Z}_n^{\top}\mathbf{Z}_n$, the latter was chosen in the Monte Carlo simulations and empirical study. In the second step, the feasible efficient GMM estimator is defined with $\mathbf{A}_n = \widetilde{\mathbf{\Omega}}_n^{-1}$ in (4.1), which we call $\widehat{\boldsymbol{\theta}}_{\text{GMM}}^{\star}$.

4.2 Standard Error Calculation

It follows that the efficient variance-covariance matrix (4.4) can be estimated by

$$\left[n^{-1} \mathbf{D}_n^{\top} \mathbf{Z}_n \widehat{\mathbf{\Omega}}_n^{\star -1} n^{-1} \mathbf{Z}_n^{\top} \mathbf{D}_n \right]^{-1}, \tag{4.6}$$

where $\widehat{\mathbf{\Omega}}_n^{\star}$ is calculated as in (4.5), but using $\widehat{\boldsymbol{\theta}}_{\text{GMM}}^{*}$ instead. The standard errors can then be calculated by taking the squared-root of the main diagonal elements of (4.6) after dividing them by n.

5 Monte Carlo Experiments

To showcase the versatility of the proposed estimator in this paper, this section documents its performance in two different data-generating processes (hereafter DGPs) where the endogeneity is generated by a simultaneous determination of network formation and outcomes–

also known as unobserved homophily–(Design 1) and measurement error in the connections (Design 2). A total of 1,500 data sets $\big\{y_{n;i}, x_{n;i}, \{w_{n;i,j}\}_{j=1,j\neq i}^{n}, \{w_{n,0;i,j}\}_{j=1,j\neq i}^{n}\big\}_{i=1}^{n}$; with $n \in \{50, 100, 200\}$, are generated from (3.1) by setting $k = 1$ and drawing $\{x_{n,i}\}_{i=1}^{n}$ as a random sample from a normal distribution with a mean of zero and variance of 3. The other data components are constructed using the following rules.

Design 1: Unobserved Characteristics with Homophily

In this design, individual i's outcome variable $y_{n,i}$ and connections $\{w_{n;i,j}\}_{j=1,j\neq i}^{n}$ are jointly determined through a common $\varepsilon^*_{n,1;i}$ idiosyncratic homophily-related unobserved feature. First, an exogenous adjacency matrix $\mathbf{W}_{n,0} = [w_{n,0;i,j}]$ from an Erdös and Rényi's (1959) random graph with a density of 0.01 is generated along with a $n \times 1$ vector $\boldsymbol{\varepsilon}^*_{n,1} = [\varepsilon^*_{n,1;1}, \dots, \varepsilon^*_{n,1;n}]^\top$ from a multivariate standard normal distribution. The elements of the endogenous adjacency matrix $\mathbf{W}_n = [w_{n;i,j}]$ are then calculated as

$$w_{n;i,j} = \begin{cases} I[|\varepsilon^*_{n,1;i} - \varepsilon^*_{n,1;j}| < \widehat{F}^{-1}_{\varepsilon^*_{n,1}}(0.95)] \times (1 - w_{n,0;i,j}) + w_{n,0;i,j} & \text{, if } \varepsilon^*_{n,1;i} > \Phi^{-1}(0.95); \\ I[|\varepsilon^*_{n,1;i} - \varepsilon^*_{n,1;j}| < \widehat{F}^{-1}_{\varepsilon^*_{n,1}}(0.95)] \times w_{n,0;i,j} & \text{if } \varepsilon^*_{n,1;i} < \Phi^{-1}(0.05); \\ w_{n,0;i,j} & \text{, otherwise;} \end{cases}$$

where $\widehat{F}^{-1}_{\varepsilon^*_{n,1}}(0.95)$, which represents the 95% empirical quantile of the elements of $\boldsymbol{\varepsilon}^*_{n,1}$, $\varepsilon^*_{n,1;k}$ represents its kth element, and $\Phi^{-1}(\cdot)$ represents the inverse of the cumulative distribution function of a standard normal random variable. This design captures the homophily idea; i.e., agents endowed with a large value of $\varepsilon_{n,1}$ will tend to create/mantain connections with those also endowed with large values of $\varepsilon_{n,1}$ and sever them with those with low values of this idiosyncratic unobserved feature. The $n \times 1$ vector of outcomes, $\mathbf{y}_n$, is then constructed from (3.1) by setting $\mathbf{v}_n = m \times \boldsymbol{\varepsilon}_{n,1} + \boldsymbol{\varepsilon_{n,2}}$, where $m \in \{1, 3\}$, $\boldsymbol{\varepsilon_{n,2}}$ is drawn from a multivariate standard normal distribution. The elements of $\boldsymbol{\varepsilon}_{n,1}$ are defined as

$$\varepsilon_{n,1;i} = \begin{cases} \varepsilon^*_{n,1;i} & \text{, if } \varepsilon^*_{n,1;i} < \Phi^{-1}(0.05) \text{ or } \varepsilon^*_{n,1;i} > \Phi^{-1}(0.95); \\ 0 & \text{, otherwise.} \end{cases}$$

Design 2: Misclassified Links

This is a modified version of Lewbel, Qu, and Tang's (2019) Monte Carlo design. While the true DGP involves an unobserved adjacency matrix $\mathbf{W}^*_{n,0} = [w^*_{n,0;i,j}]$ generated from a standard Erdös and Rényi's (1959) random network model with a density of 0.01 for size n, the empiricist is assumed to only have access to an adjacency matrix, $\mathbf{W}_n = [w_{n;i,j}]$, with randomly misclassified links; i.e., $w_{n;i,j} = w^*_{n,0;i,j} e_{n,1;i,j} + (1 - w^*_{n,0;i,j}) e_{n,2;i,j}$ for $i \neq j$ and an exogenous adjacency matrix $\mathbf{W}_{n,0} = [w_{n,0;i,j}]$, where $w_{n,0;i,j} = w^*_{n,0;i,j} b_{n,1;i,j} + (1 - w^*_{n,0;i,j}) b_{n,2;i,j}$ for $i \neq j$. The $e_{n,1;i,j}$, $e_{n,2;i,j}$, $b_{n,1;i,j}$, and $b_{n,2;i,j}$ are Bernoulli random variables drawn independently from each other $\forall i \neq j$ with parameters 0.5, 0, $1 - \tau$, and 0.002 respectively. The design parameter $\tau \in \{0.01, 0.05\}$ controls the probability of misclassification in $\mathbf{W}_{n,0}$. Notice that as in Lewbel, Qu, and Tang (2019), non-existing links are never misclassified in $\mathbf{W}_n$, but missclasification of these non-existing links are permitted in $\mathbf{W}_{n,0}$ with a very small probability of 0.2%. However, this design makes the vector of individual outcomes an explicit function of the proportion of misclassification in $\mathbf{W}_n$ for each i; i.e., the $n \times 1$ vector $\mathbf{y}$ is constructed following Equation (3.1), where $\mathbf{n}, \mathbf{v} = \boldsymbol{\varepsilon}_{n,1} + \boldsymbol{\varepsilon}_{n,2}$, $\varepsilon_{n,1;i} = 1/n \sum_{j=1}^{n} w^*_{n,0;i,j} e_{n,1;i,j}$, and $\boldsymbol{\varepsilon}_{n,2}$ is drawn from a multivariate standard normal distribution independently of everything else.

Results

Figures 2 and 3 show the results in terms of box plots and Q-Q plots of the Monte Carlo replications. Apart from implementing the proposed efficient GMM estimator described in Theorem 2 for $p \in 2, 3$, the performance of the standard Ordinary Least Squares (OLS) estimator and the Generalized Two Stage Least Squares (G2SLS) estimator are also included. All adjacency matrices in all designs are row-normalized prior to estimation (Liu, Patacchini, and Zenou, 2014). Computing the efficient GMM requires an estimator of the variance-covariance matrix $\boldsymbol{\Omega}_n$. We use the standard two-stage GMM procedure to calculate the efficient weighting matrix. In the first step, we calculate the GMM estimator for $\boldsymbol{\theta}$ setting $\mathbf{A}_n = (\mathbf{Z}_n^\top \mathbf{Z}_n)^{-1}$. We then use the first-step estimated coefficients to compute the network HAC variance estimator in (4.5), where we choose $K(\cdot)$ to be the Parzen kernel, we set the bandwidth $D_n = C \times [\log(\text{ average degree } \vee (1 + 0.05))]^{-1} \times \log n$ as in Kojevnikov, Marmer, and Song (2020), and $C = 1.8$.

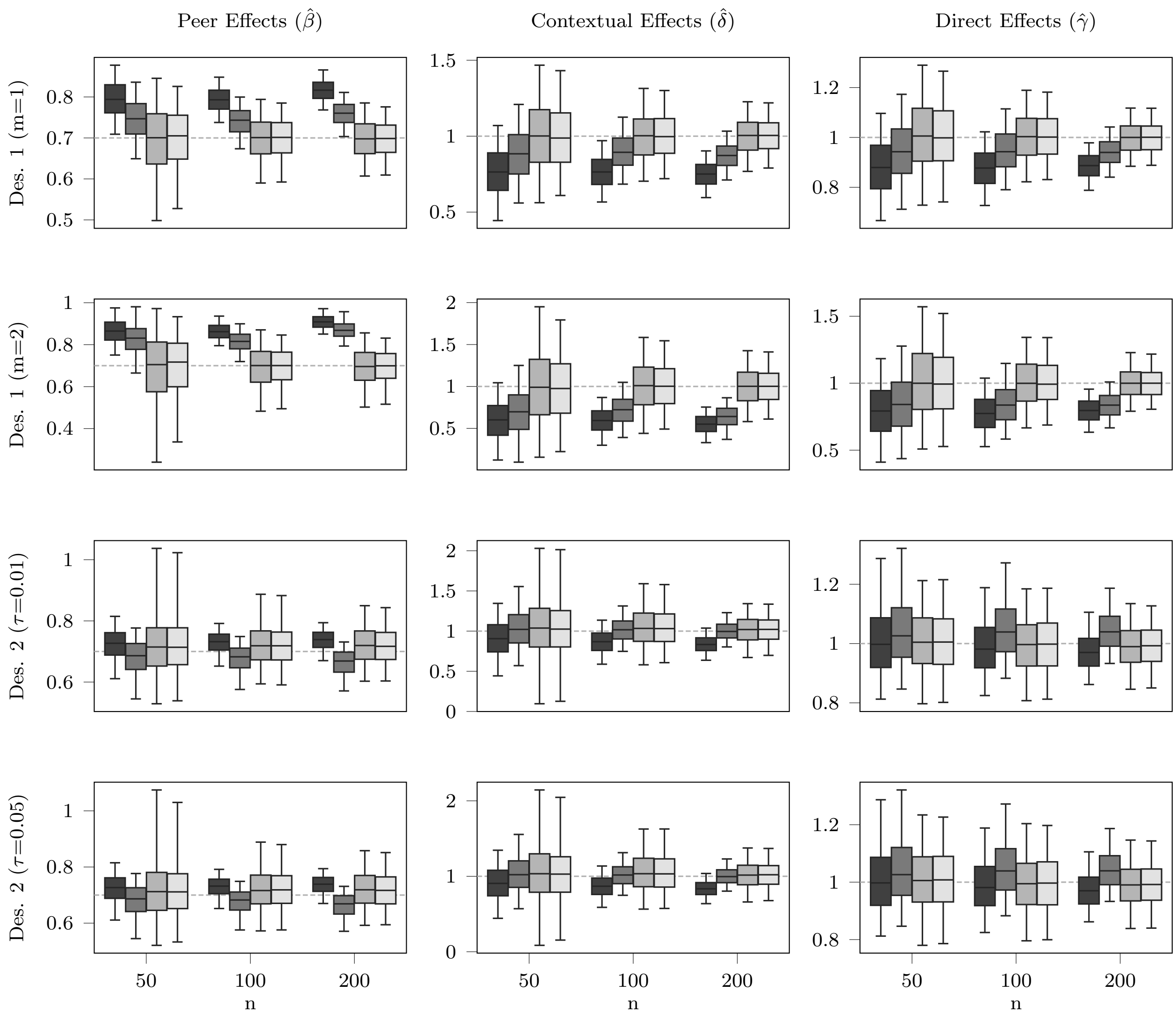
Each panel in Figure 2 displays the performance of the three estimators when the state of a design (Des.) changes by changing the relevant design parameter m or τ. The box plots are based on the Monte Carlo replications of the OLS (black), G2SLS (dark gray), and the two proposed GMM estimators for $p = 2$ (gray) and $p = 3$ (light gray) of the social effects. The peer effects (β), contextual effects (δ), and direct effects (γ) in (3.1) are shown with whiskers displaying the 5% and 95% empirical Monte Carlo quantiles. Across the board for all parameters, designs, and sample sizes the proposed both GMM estimators performs better than the OLS and G2SLS estimators in terms of bias and sampling variability. As expected, the estimation variability decreases when going from $p = 2$ to $p = 3$. In contrast, these results also show that the naive OLS and G2SLS could potentially lead to estimates with substantial biases in the presence of an endogenous network in a linear-in-means model. On average, the G2SLS underestimates the real value of the peer effects coefficient as shown in Chandrasekhar and Lewis (2016) for the case of missclassified links (Design 2).

Similarly, Figure 3 displays the corresponding Q-Q plots for the GMM based on the standardized version of the Monte Carlo replications of the GMM estimator of the same social effects for sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black). The blue dashed line depicts the 45 degree line. This plot shows that the asymptotic normal approximation in Theorem 2 works well even with a sample as small as 50 observations. Furthermore, as sample size increases, the approximation improves for all parameters and designs.

6 Data Analysis

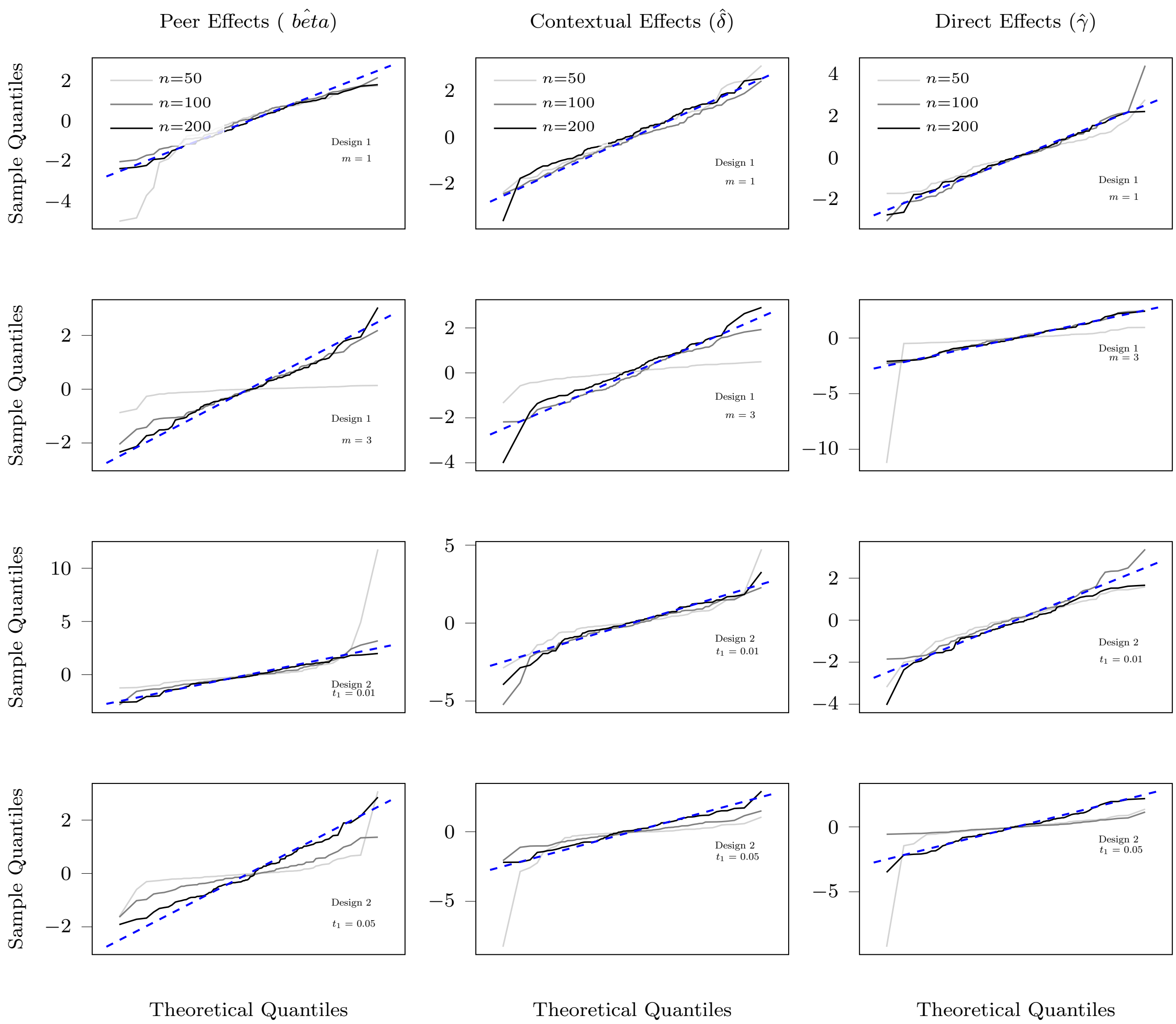
The data was collected between March and May 2011 as part of the Secondary Education Survey in Hong Kong (SESHK). The survey was conducted in the second semester before final exams. Three secondary schools (s $\in \{1, 2, 3\}$) are involved in the survey with 873 out of 898 students participating. The sample includes seventh grade students from all three schools (schools 1, 2, and 3 below) and eight and nine grade students from one school; i.e., g $\in \{7, 8, 9\}$. Each grade has five sections (cl $\in \{1, \ldots, 5\}$) with an average class size of about 35 students. The gathered information includes exam scores, student characteristics, and

Figure 2: Box Plots of the OLS, G2SLS, and GMM Estimators of Social Effects



Note: Box plots depict the Monte Carlo performance of the OLS (black), G2SLS (dark gray), and the proposed efficient GMM estimator for $p = 2$ (gray) and $p = 3$ (light gray). The box plots are based on 1,500 replications of Design 1 (Des.1) and Design 2 (Des. 2) for sample sizes $n \in \{50, 100, 200\}$. The whiskers display the 5% and 95% empirical quantiles. The parameters m and τ control the level of endogeneity and the probability of misclassification in $\mathbf{W}_n$, respectively.

Figure 3: Q-Q Plots for the GMM Estimator of Social Effects



Note: Q-Q plots are based on the standardized sample of 1,500 Monte Carlo replications of the proposed GMM estimator of the parameters in (3.1) for Design 1 (Des.1) and Design 2 (Des. 2) for sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black); $p = 3$. The blue dashed line shows the 45 degree line. The parameters m and τ control the level of endogeneity and the probability of misclassification in $\mathbf{W}_n$, respectively.

social networks. Our analysis focuses on the first mathematics exam score in the academic year–provided by the participating schools. For each student, a progressive matrix and the Big Five Inventory measures (based on a questionnaire with 44 items taken from John, Donahue, and Kentle, 1991) were also collected.

The progressive matrix test is a series of context-free logical deduction tests on spatial awareness and shapes. They are designed to measure students' cognitive abilities and are proxies for students' intelligence. The Big Five Inventory produces information on five personality measures. Conscientiousness includes elements like self-discipline, carefulness, and diligence. Agreeableness identifies whether a student is sympathetic, considerate, and kind. Extraversion indicates whether a student is outgoing or talkative. Openness refers to elements like imagination and curiosity. Neuroticism describes a student who worries or becomes anxious easily.

The SESHK also provides information on students' hobbies and other demographic information. Students were asked to write down their hobbies and other personal and family characteristics, such as their weight, height, sibling composition, and means of commuting to school. Table 1 contains the relevant descriptive statistics for all students by school.

Seatmate and Student Partner Networks

In the survey, students were asked to write down lists of up to ten peers from among their schoolmates within the same grade whom they discuss their problems with schoolwork and who sat next to them in class during the first semester. We use this information to build the study partner and seatmate networks using the following reciprocal peer rule: If students i and j named each other as study partners in the survey, we record an edge in the study partner network. We follow the same process for the seatmate network. Table 2 reports the network summary statistics among all students by school.

Seat assignments in the classrooms change several times over a semester and the changes are decided by the class teacher. Unlike study partners, the seatmate network is proximity-based and imposed by the school; i.e., 86% of the students in the original sample say that seat assignment is decided by the teachers and 16% of them perceived the assignment to be random. Therefore, the seatmate network is a prime candidate to be used as the instrumental network, $\mathbf{W}_{n,0}$, in our analysis. On the other hand, students can freely choose who they study with and this decision might be based on unobservable characteristics that also affect exam performance, making the study partner network, $\mathbf{W}_n$, likely endogenous. Further details about the data can be found in Appendix C in the supplement.

In this application we are interested in fitting a model for student i's performance in the math test. In particular, the following Linear-in-Means model for social effects for individual i is fitted as

$$
\begin{aligned}
\texttt{math}_{i,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}} \quad = \quad & \alpha + \beta \sum_{j\neq i}^{n} w_{n;i,j} \texttt{math}_{j,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}} \\
& + \sum_{j\neq i}^{n} w_{n;i,j} \texttt{characteristics}'_{j,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}} \delta_{\texttt{characteristics}} \\
& + \texttt{personality}'_{i,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}} \gamma_{\texttt{personality}} + \texttt{characteristics}'_{i,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}} \gamma_{\texttt{characteristics}} \\
& + \sum_{\mathrm{s}=1}^{3} \sum_{\mathrm{g}=7}^{9} \sum_{\mathrm{cl}=1}^{5} f_{\mathrm{s}\times\mathrm{g}\times\mathrm{cl}} \times I\{i \in \mathrm{s}\times\mathrm{g}\times\mathrm{cl}\} + \epsilon_i,
\end{aligned} \tag{6.1}
$$

where $\texttt{math}_{i,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}}$ is the natural logarithm of the first math test score of student i in class cl, grade g, and school s; i.e., $I\{i \in \mathrm{s}\times\mathrm{g}\times\mathrm{cl}\} = 1$; $\texttt{personality}_{i,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}}$ includes natural logarithm of cognitive ability, agreeableness, conscientiousness, extraversion, neuroticism, and openness test scores.

Similarly, $\texttt{characteristics}_{i,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}}$ includes variables such as height, weight, indicator variables such as whether sibling help and parents help–which equal one if the students report that they tend to seek help from their siblings or parents; whether they commute to school by car or taxi; whether they play music; and whether the student is a male. We emphasize that including the variable height in both direct and contextual effects is also relevant from the identification perspective. The reason is that height can be correlated with whether or not students sit in front or the back of the classroom. We also include interaction terms between the male indicator and test scores. The parameters $f_{\mathrm{s}\times\mathrm{g}\times\mathrm{cl}}$ are jointly estimated with the social effects after setting $f_{3\times 9\times 5} = 0$. The adjacency matrices are row-normalized prior to estimating the model as permitted by our theory. The effect of having more peers is captured by including the network degree (number of study partners).

We also control for the fact that some students study alone (isolate).

Table 3 shows the results. Model (6.1) is fitted by simple Ordinary Least Squares (OLS) with standard errors clustered at the s $\times$ g $\times$ cl level, the Generalized Two-Stage Least Squares (G2SLS) of Kelejian and Prucha (1998, 1999), Lee (2003), and Bramoullé, Djebbari, and Fortin (2009) with clustered standard errors as in the OLS estimator, the proposed efficient GMM estimator with $p = 5$, $C = 1.7$, and the Tukey-Hanning kernel. We set $\delta_{\texttt{personality}} = -\gamma_{\texttt{personality}}$ across estimation routines in order to avoid potential issues of collinearity.

All estimators show positive spillover effects; however, both OLS and G2SLS are significantly smaller than what the proposed GMM estimator uncovers. The direct effect coefficients are relatively consistent across different estimators, both in terms of size and direction. Importantly, the signs of most coefficients match previous findings in the literature. For instance, cognitive ability and non-cognitive features such as conscientiousness have a positive effect in math grade achievement (Heckman and Rubinstein, 2001). There is not a gender gap in the grades of males and females. Interestingly, the estimated coefficient for the Degree variable suggests that having a larger number of study partners has a positive and significant effect on achievement and that result is robust across different estimators.

Regarding network effects, our results suggest that the estimators that do not control for network endogeneity tend to underestimate peer effects. Based on our simulation results in Section 5, one possible explanation is the existence of unobserved homophily. If the unobserved variables driving the choice of study partners are negatively correlated with the outcome, we expect the endogeneity bias to underestimate the actual peer effects value. Students can choose to study with others that they find fun for reasons other than learning the test material. If students select study partners that can distract them from schoolwork, estimators that do not take that sorting process into account can be downward-biased. These results suggest that policies that strengthen the collaboration between students inside and outside of the classroom can generate benefits that have the potential to generate positive social multipliers. Finally, all results are qualitatively robust to different choices of p and D_n; see Appendix C in the supplemental materials.

Table 1: Summary Statistics

		School 1 ($n = 133$)				School 2 ($n = 171$)				School 3 ($n = 564$)			
Variables	Scale	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max
Student-related variables													
Math Test	[0,100]	61.88	13.56	33.57	92.33	61.44	13.12	31.00	92.00	68.38	14.01	24.35	100.00
Male		0.37	0.48	0.00	1.00	0.58	0.50	0.00	1.00	0.42	0.49	0.00	1.00
Height (cm)		156.50	7.48	139.00	176.00	157.41	7.80	133.00	175.00	161.05	9.18	100.00	208.30
Weight (kg)		46.56	9.15	27.00	72.00	46.86	11.27	28.00	99.00	48.33	10.64	26.30	130.00
Siblings Help		0.47	0.50	0.00	1.00	0.43	0.50	0.00	1.00	0.46	0.50	0.00	1.00
Parents Help		0.61	0.49	0.00	1.00	0.63	0.48	0.00	1.00	0.65	0.48	0.00	1.00
Music		0.53	0.58	0.00	2.00	0.66	0.62	0.00	3.00	0.85	0.56	0.00	3.00
Commute by car/taxi		0.68	0.47	0.00	1.00	0.53	0.50	0.00	1.00	0.58	0.49	0.00	1.00
Cognitive and Personality Tests													
Cognitive	[0,16]	7.80	1.66	3.00	12.00	7.94	1.78	4.00	12.00	8.92	1.88	2.00	14.00
Agreeableness	[9,40]	27.12	4.26	14.00	39.00	27.09	3.91	15.00	37.00	27.04	3.99	12.00	40.00
Conscientiousness	[9,45]	26.71	5.87	14.00	40.00	27.90	4.97	18.00	43.00	25.88	5.47	12.00	45.00
Extraversion	[8,40]	27.65	4.86	16.00	38.00	27.62	4.85	16.00	38.00	26.35	5.10	10.00	39.00
Neuroticism	[8,40]	22.31	5.83	9.00	36.00	21.95	5.36	8.00	35.00	23.45	5.57	9.00	38.00
Openness	[10,55]	37.45	5.45	24.00	50.00	36.65	5.09	19.00	51.00	35.26	5.48	18.00	51.00

Note: Descriptive statistics such as sample mean (Mean), standard deviation (SD), minimum (Min), maximum (Max), and sample size (n) for all variables and for each school are presented here. Course grades, personality trait measures, and cognitive ability tests are scored on the scale indicated.

Table 2: Summary Network Statistics

	School 1		School 2		School 3	
Variables	Study Partners	Seatmates	Study Partners	Seatmates	Study Partners	Seatmates
Number of Nodes	133	133	171	171	564	564
Number of Edges	171	175	228	275	819	799
Density $\times$ 100	1.948	1.994	1.569	1.892	0.516	0.503
Average Degree	2.571	2.632	2.667	3.216	2.904	2.833
Average Clustering	0.213	0.068	0.153	0.066	0.129	0.063
Assortativity Measure	0.190	0.266	0.039	0.194	0.177	0.177
Number of Isolated Node	15	7	19	3	69	13
Number of Subgraph	21	14	27	9	76	30
Transitivity	0.281	0.094	0.207	0.092	0.180	0.081

Note: Degree is multiplied by 100 in order to increase the scale.

7 Discussion

This research adds to the literature of the identification and estimation of social effects with observational network data that often contains endogenous or mismeasured connections. Unlike current approaches, such as in, Johnsson and Moon (2019) and Auerbach (2022), our method does not require the specification and estimation of a model of how connections are created or misclassified. Our method circumvents the imposition of these modeling requirements (along with its potential misspecification issues) by showing how a fully observed set of exogenous connections can be used as an *instrumental* network to uniquely identify and estimate parameters of interest in a widely-used linear model of social interactions. Therefore, our approach is semiparametric in nature and hence avoids the usual drawbacks of strong modeling assumptions in this literature.

An important aspect of the proposed methodology is that it recognizes that exogenously-imposed connections on individuals (such as randomization) do not necessarily cause social effects. However, they can generate new types of freely-formed connections that do so; i.e., resorting. The correlation between these two networks is at the heart of our identification and estimation strategy. In this sense, our approach provides an explicit solution to the resorting issue in random network identification strategies, such as in Moffitt (2001), by distinguishing what type of network creates peer effects (who you study with, for example) and what other type simply influences these connections, but are otherwise exogenous to the model (for instance, to whom your are randomly assigned to share a physical space). Our empirical results show that ignoring the potential network endogeneity can severely bias the network effects estimators. We find significant positive network effects of math test scores among study partners. These results are in line with previous literature in that they show the existence of strong positive network effects, but it suggests that the magnitude of the effects can be larger. We postulate that this could be due to the fact that we focus directly on networks that endogenously emerge after an initial exogenous network assignment.

Finally, another contribution of this research is technical in nature. A byproduct of acknowledging potential network mismeasurement or endogeneity is that it explicitly permits the observed and unobserved characteristics of individuals to be correlated; i.e., creating network dependence across observations in the sample. Our asymptotic results utilize the idea that dependence among observations decreases as a function of their distance in the network; i.e, $\psi-$dependence. Our results show that the resulting estimator can be easily implemented utilizing standard linear GMM estimation routines in popular software like Python, R, or Stata. The estimator is consistent and asymptotically normally-distributed at the standard parametric convergence rate. We characterize the form of the asymptotic variance-covariance matrix that accounts for the network dependence and illustrate how standard errors can be calculated in an empirical application.

References

Advani, Arun and Bansi Malde. 2018. "Credibly Identifying Social Effects: Accounting for Network Formation and Measurement Error." *Journal of Economic Surveys* 32 (4):1016–1044.

Alexander, Cheryl, Marina Piazza, Debra Mekos, and Thomas Valente. 2020. "Peers, Schools, and Adolescent Cigarette Smoking." *Journal of Adolescent Health* 29:22–30.

Atkisson, Curtis, Piotr J. Górski, Matthew O. Jackson, Janusz A. Hołyst, and Raissa M. D'Souza. 2020. "Why Understanding Multiplex Social Network Structuring Processes Will Help Us Better Understand the Evolution of Human Behavior." *Evolutionary Anthropology* 29 (3):102–107.

Auerbach, Eric. 2022. "Identification and Estimation of a Partially Linear Regression Model Using Network Data." *Econometrica* 90 (1):347–365.

Blume, Lawrence E., William A. Brock, Steven N. Durlauf, and Rajshri Jayaraman. 2015. "Linear Social Interactions Models." *Journal of Political Economy* 123 (2):444–496.

Boccaletti, S., G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang, and M. Zanin. 2014. "The Structure and Dynamics of Multilayer Networks." *Physics Reports* 544:1–122.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. "Identification of Peer Effects through Social Networks." *Journal of Econometrics* 150 (1):41–55.

Calvó-Armengol, Antoni, Eleonora Patacchini, and Yves Zenou. 2009. "Peer effects and social networks in education." *The review of economic studies* 76 (4):1239–1267.

Carrell, Scott E., Richard L. Fullerton, and James E. West. 2009. "Does Your Cohort Matter? Measuring Peer Effects in College Achievement." *Journal of Labor Economics* 27 (3):439–464.

Carrell, Scott E., Mark Hoekstra, and James E. West. 2011. "Is Poor Fitness Contagious?. Evidence from Randomly Assigned Friends." *Journal of Public Economics* 95 (7-8):657–663.

Carrell, Scott E., Bruce I. Sacerdote, and James E. West. 2013. "From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation." *Econometrica* 81 (3):855–882.

Chandrasekhar, Arun G. and Randall Lewis. 2016. "Econometrics of Sampled Networks." Unpublished Manuscript.

de Paula, Áureo. 2017. *Econometrics of Network Models*, *Econometric Society Monographs*, vol. 1, chap. 8. Cambridge University Press, 268–323.

de Paula, Áureo, Seth Richards-Shubik, and Elie Tamer. 2018. "Identifying preferences in networks with bounded degree." *Econometrica* 86 (1):263–288.

Doukhan, Paul and Sana Louhichi. 1999. "A new weak dependence condition and applications to moment inequalities." *Stochastic processes and their applications* 84 (2):313–342.

Erdös, P and A Rényi. 1959. "On Random Graphs." *Publicationes Mathematicae Debrecen* 6:290–297.

Estrada, Juan. 2021. "Causal Inference in Multilayer Social Networks." Unpublished Manuscript.

Goldsmith-Pinkham, Paul and Guido W. Imbens. 2013. “Social Networks and the Identification of Peer Effects.” *Journal of Business and Economic Statistics* 31 (3):253–264.

Graham, Bryan S. 2017. “An Econometric Model of Network Formation With Degree Heterogeneity.” *Econometrica* 85 (4):1033–1063.

———. 2020. “Network data.” *Handbook of Econometrics* 7A.

Graham, Bryan S. and Andrin Pelican. 2020. “Chapter 4 - Testing for externalities in network formation using simulation.” In *The Econometric Analysis of Network Data*, edited by Bryan Graham and Áureo de Paula. Academic Press, 63–82.

Heckman, James J and Yona Rubinstein. 2001. “The importance of noncognitive skills: Lessons from the GED testing program.” *American Economic Review* 91 (2):145–149.

Jackson, Matthew O., Brian W. Rogers, and Yves Zenou. 2017. “The Economic Consequences of Social-Network Structure.” *Journal of Economic Literature* 55 (1):49–95.

John, O., E. Donahue, and R. Kentle. 1991. “The Big Five Inventory - Versions 4a and 54.” University of California, Berkeley, Institute of Personality and Social Research.

Johnsson, Ida and Hyungsik Roger Moon. 2019. “Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach.” *The Review of Economics and Statistics* :1–51.

Kelejian, Harry H. and Ingmar R. Prucha. 1998. “A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances.” *Journal of Real Estate Finance and Economics* 17 (1):99–121.

———. 1999. “A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model.” *International Economic Review* 40 (2):509–533.

Kivela, M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. 2014. “Multilayer Networks.” *Journal of Complex Networks* 2 (3):203–271.

Kojevnikov, Denis, Vadim Marmer, and Kyungchul Song. 2020. “Limit theorems for network dependent random variables.” *Journal of Econometrics* .

Kreager, Derek a., Kelly Rulison, and James Moody. 2020. “Delinquency and the Structure of Adolescent Peer Groups.” *Criminology* 49:95–127.

Lee, Lung Fei. 2003. “Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances.” *Econometric Reviews* 22 (4):307–335.

Lewbel, Arthur, Xi Qu, and Xun Tang. 2019. “Social Networks with Misclassified or Unobserved Links.” Unpublished Manuscript.

Liu, Xiaodong, Eleonora Patacchini, and Yves Zenou. 2014. “Endogenous Peer Effects: Local Aggregate or Local Average?” *Journal of Economic Behavior and Organization* 103:39 – 59.

Manta, Alexandra, Anson T.Y. Ho, Kim P. Huynh, and David T. Jacho-Chavez. 2022. “Estimating Social Effects in a Multilayered Linear-in-Means Model with Network Data.” *Statistics & Probability Letters* 183:109331.

Moffitt, Robert A. 2001. “Policy Interventions, Low-Level Equilibria And Social Interactions.” In *Social Dynamics*, edited by Steven Durlauf and Peyton Young. MIT Press, 45–82.

Qu, Xi and Lung Fei Lee. 2015. “Estimating a Spatial Autoregressive Model with an Endogenous Spatial Weight Matrix.” *Journal of Econometrics* 184 (2):209–232.

Sacerdote, Bruce. 2001. “Peer Effects with Random Assignment: Results for Dartmouth Roommates.” *Quaterly Journal of Economics* 116 (2):681–704.

Salmivalli, Christina. 2020. “Bullying and the Peer Group: A Review.” *Aggression and Violent Behavior* 15:112–120.

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

Table 3: Estimation Results

Variables	OLS	G2SLS	GMM
Peer effect			
ln(Math Test)	0.2455***	0.4534***	0.6054***
	(0.0864)	(0.1027)	(0.1757)
Contextual effects			
Male	-0.0232	-0.0153	-0.2455***
	(0.0311)	(0.0276)	(0.0873)
ln(Height)	0.0794	0.0651	2.1787***
	(0.2229)	(0.1910)	(0.7404)
ln(Weight)	0.1012	0.0399	0.0325
	(0.0642)	(0.0600)	(0.1943)
Siblings Help	0.0263	0.0305	0.0296
	(0.0261)	(0.0217)	(0.0491)
Parents Help	0.0345	0.0064	-0.0103
	(0.0287)	(0.0189)	(0.0517)
Commute by Car/Taxi	0.0277	0.0250	0.1253**
	(0.0266)	(0.0224)	(0.0636)
Music	-0.0212	-0.0090	0.1187**
	(0.0166)	(0.0160)	(0.0481)
Direct effects			
ln(Cognitive)	0.0549	0.0734**	0.1086***
	(0.0357)	(0.0369)	(0.0286)
ln(Agreeableness)	-0.1145***	-0.1275***	-0.0711
	(0.0300)	(0.0356)	(0.0458)
ln(Conscientiousness)	0.0445	0.0537	0.0799**
	(0.0371)	(0.0396)	(0.0406)
ln(Extraversion)	-0.0998**	-0.0982**	-0.1165***
	(0.0387)	(0.0384)	(0.0405)
ln(Neuroticism)	-0.0386	-0.0409	-0.0113
	(0.0378)	(0.0373)	(0.0257)
ln(Openness)	0.0461	0.0358	0.0272
	(0.0485)	(0.0489)	(0.0414)
Male	-0.7922*	-0.8352**	-0.1642
	(0.4596)	(0.4224)	(0.5423)
ln(Height)	-0.3366**	-0.2635**	-0.9926***
	(0.1342)	(0.1210)	(0.2251)
ln(Weight)	-0.0088	-0.0106	-0.0496
	(0.0293)	(0.0298)	(0.0453)
Siblings Help	-0.0099	-0.0172	-0.0395*
	(0.0176)	(0.0165)	(0.0204)
Parents Help	0.0278*	0.0225	0.0168
	(0.0159)	(0.0142)	(0.0139)
Commute by Car/Taxi	-0.0079	-0.0162	-0.0250*
	(0.0117)	(0.0111)	(0.0136)
Degree	0.0292***	0.0264***	0.0249***
	(0.0041)	(0.0039)	(0.0034)
Isolate Students	2.2787*	0.4720*	0.1425
	(1.2954)	(0.2813)	(0.2933)
Constant	3.8210***	2.9464***	-4.6230*
	(1.2635)	(0.9048)	(2.5819)
n	868	868	868
Adjusted R^2	0.3372	0.3936	0.2672
RMSE	0.1854	0.1716	0.2002

Note: (i) * p-value < 0.10, ** p-value < 0.05, and *** p-value < 0.01; (ii) Standard errors are in parentheses; and (iii) Other controls are available in the supplemental material.

On the Estimation of Social Effects with Observational Network Data and Random Assignment
– Supplemental Materials –

TszKin Julian Chan
Bates White Economic Consulting
Juan Estrada
Department of Economics, Emory University
Kim Huynh*
Currency Department, Bank of Canada
David Jacho-Chávez
Department of Economics, Emory University
Chungsang Tom Lam
Department of Economics, Clemson University
Leonardo Sánchez-Aragón
Facultad de Ciencias Sociales y Humanísticas, ESPOL University

August 5, 2022

Appendix A Proofs of Main Results

Proof of Theorem 1. First, note that Assumption 2 guarantees that the solution for model (3.1) exists. Assumption 1(*iii*) guarantees that the system of equations $E[\boldsymbol{m}_N(\boldsymbol{\theta})] = \mathbf{0}_K$ are not trivially satisfied by making $\eta_{N,0;i} = 0$ for all $i \in \mathcal{I}_N$. We show that the moment condition equation has a unique root at $\boldsymbol{\theta}_0 = (\alpha_0, \beta_0, \boldsymbol{\delta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$. In particular, we show that there cannot be any other $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ different from $\boldsymbol{\theta}_0$ for which the moment condition holds.

*The views expressed in this article are those of the authors. No responsibility for them should be attributed to the Bank of Canada. All remaining errors are the responsibility of the authors.

Choose an arbitrary vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, such that $E[\mathbf{m}(\boldsymbol{\theta})] = 0$. Assumption 2 implies that $E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}(y_{N;i} - \mathbf{d}_{N;i}^\top\boldsymbol{\theta})] = \mathbf{0}_K$. It follows that $E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top](\boldsymbol{\theta}_0 - \boldsymbol{\theta}) + E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\varepsilon_{N;i}] = \mathbf{0}_K$ and $E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]\,(\boldsymbol{\theta}_0 - \boldsymbol{\theta}) = \mathbf{0}_K$. Under Assumption 3, it follows that $E[\mathbf{m}(\boldsymbol{\theta})] = \mathbf{0}_K$ if, and only if, $\boldsymbol{\theta}_0 = \boldsymbol{\theta}$. ■

Proof of Theorem 2. The GMM estimator in (4.1) in the main text can be written as

$$\widehat{\boldsymbol{\theta}}_{\text{GMM}} = \boldsymbol{\theta} + (n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\mathbf{A}_n n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n)^{-1} n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\mathbf{A}_n n^{-1}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n. \tag{A-1}$$

By construction, the matrix $\mathbf{A}_n$ is assumed to converge to the full rank matrix $\mathbf{A}_N$ as $n \to \infty$. From Corollary B.1, $n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n$ converges to the population quantity $E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$, which is finite given Assumption 3. Finally, Corollary B.2 shows that $n^{-1}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n(\boldsymbol{\theta})$ converges to $E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\varepsilon_{N;i}(\boldsymbol{\theta})] = 0$. It then follows that $\widehat{\boldsymbol{\theta}}_{\text{GMM}} = \boldsymbol{\theta} + o_p(1)$ as $n \to \infty$. For asymptotic normality, note that, from, (A-1)

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{\text{GMM}} - \boldsymbol{\theta}) = (n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\mathbf{A}_n n^{-1}\mathbf{Z}_n^\top\mathbf{D}_n)^{-1} n^{-1}\mathbf{D}_n^\top\mathbf{Z}_n\mathbf{A}_n \times n^{-1/2}\mathbf{Z}_n^\top\boldsymbol{\varepsilon}_n.$$

Let $\mathbf{Q}_{zx} = E[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$. Then from Corollary B.1 and Lemma B.3, it follows that

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_{\text{GMM}} - \boldsymbol{\theta}\right) \xrightarrow{d} \left[\mathbf{Q}_{zx}^\top\mathbf{A}_N\mathbf{Q}_{zx}\right]^{-1}\mathbf{Q}_{zx}^\top\mathbf{A}_N \times \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_N).$$

The result then follows. The efficient variance-covariance matrix in (4.4) follows from standard matrix algebra calculations. ■

Appendix B Auxiliary Results

Lemma B.1 *Let Assumption 4 hold for* $\{\mathbf{r}_{n;i}\}_{n\geq 1}$, $i \in \mathcal{I}_n$ *and define* $R_{n;i,j} = f_{q,\ell}(\mathbf{r}_{n,\{i,j\}}) \equiv r_{n;i,q}r_{n;j,\ell}$ *and* $R_{n;h,s} = g_{q',\ell'}(\mathbf{r}_{n,\{h,s\}}) \equiv r_{n;h,q'}r_{n;s,\ell'}$ *for* $i,j,h,s \in \mathcal{I}_n$, *where* q, q', ℓ, *and* ℓ' *are components of the vector* $\mathbf{r}_{n;i}$. *Let Assumption 5 hold for* $R_{i,j}$ *and* $R_{h,s}$; *then*

$$|\text{cov}\,(R_{n;i,j}, R_{n;h,s})| \leq 2\bar{\lambda}_{n,d}(C+16) \times 4\,(\pi_1 + \tilde{\gamma}_1)\,(\pi_2 + \tilde{\gamma}_2)\,\underline{\lambda}_{n,d}^{1-p_f-p_g}, \tag{B-1}$$

where $\underline{\lambda}_{n,d} = \lambda_{n,d} \wedge 1$, $\bar{\lambda}_{n,d} = \lambda_{n,d} \vee 1$, $\pi_1 = \|\mathbf{r}_{n;i}\|_{p_{f,i}}\|\mathbf{r}_{n;j}\|_{p_{f,j}}$, $\pi_2 = \|\mathbf{r}_{n;h}\|_{p_{f,h}}\|\mathbf{r}_{n;s}\|_{p_{f,s}}$, $\widetilde{\gamma}_1 = \max\{\|\mathbf{r}_{n;i}\|_{p_{f,i}+p_{f,j}}, \|\mathbf{r}_{n;j}\|_{p_{f,i}+p_{f,j}}\}$; $\widetilde{\gamma}_2 = \max\{\|\mathbf{r}_{n;h}\|_{p_f}, \|\mathbf{r}_{n;s}\|_{p_g}\}$, *where* $p_f = 1/p_{f,i} + 1/p_{f,j}$ *and* $p_g = 1/p_{g,h} + 1/p_{g,s}$, *where the constant* C *is the same as in Assumption 4. The indexes* i,j,h,s, *and components* q, q', ℓ, ℓ' *may or may not be the same.*

Proof. Define the increasing continuous functions $h_1(x)$ and $h_2(x)$ as in Theorem A.2 in Kojevnikov, Marmer, and Song (2020, Appendix A, pp. 899-907) to be $h_1(x) = h_2(x) = x$. Note that the functions $f_{q,\ell}$ and $g_{q',\ell'}$ are continuous, and their truncated version of the form $\varphi_{K_1} \circ f \circ \varphi_{h_1}(K_2)$ and $\varphi_{K_1} \circ g \circ \varphi_{h_1}(K_2)$ for all $K \in (0,\infty)^2$ are in $L_{Q+1,2}$. Assumption 5 guarantees the existence of the moments defining $\widetilde{\gamma}_1$ and $\widetilde{\gamma}_2$. Then, Theorem A.2 in Kojevnikov, Marmer, and Song (2020, Appendix A, pp. 899-907) applies to this setting (see also Corollary A.2. in Appendix A in Kojevnikov, Marmer, and Song, 2020, pp. 899-907). ■

Lemma B.2 (LLN for Products of ψ-dependent Random Variables) *Let Assumptions 4 – 7 hold, define $R_{n;i,j} \equiv r_{n;i,q} r_{n;j,\ell}$, and let $w^*_{i,j}$ be weights between zero and one. Form $\{R_{n;i,j}\}_{i\in\mathcal{I}_n, j\in\mathcal{I}_i}$, where $\mathcal{I}_i$ is a set of indexes defined for each $i \in \mathcal{I}_n$. Then, as $n \to \infty$,*

$$\left\| \frac{1}{n} \sum_{i\in\mathcal{I}_n} \sum_{j\in\mathcal{I}_i} w^*_{i,j} \left(R_{n;i,j} - E\left[R_{n;i,j}\right]\right) \right\|_1 \xrightarrow{a.s.} 0.$$

Proof. Using the same approach of Jenish and Prucha (2009) and Kojevnikov, Marmer, and Song (2020), let the censoring function $\varphi_k(x) = (-K) \vee (K \wedge x)$ be such that, for some $k > 0$,

$$R_{n;i,j} = R^{(k)}_{n;i,j} + \tilde{R}^{(k)}_{n;i,j},$$

where $R^{(k)}_{n;i,j} = \varphi_k(R_{n;i,j})$ and $\tilde{R}^{(k)}_{n;i,j} = R_{n;i,j} - \varphi_k(R_{n;i,j}) = (R_{n;i,j} - \operatorname{sgn}(R_{n;i,j})k)1\{|R_{n;i,j}| > k\}$. Let $\|X\|_k = (E[|X|^k)^{1/k}$ for $k \in [1,\infty)$. Therefore, following the previous definition, we apply the triangle inequality to get

$$\begin{aligned} \left\| \frac{1}{n} \sum_{i\in\mathcal{I}_n} \sum_{j\in\mathcal{I}_i} w^*_{i,j} \left(R_{n;i,j} - E\left[R_{n;i,j}\right]\right) \right\|_1 &\leq \left\| \frac{1}{n} \sum_{i\in\mathcal{I}_n} \sum_{j\in\mathcal{I}_i} w^*_{i,j} \left(R^{(k)}_{n;i,j} - E\left[R^{(k)}_{n;i,j}\right]\right) \right\|_1 \\ &+ \left\| \frac{1}{n} \sum_{i\in\mathcal{I}_n} \sum_{j\in\mathcal{I}_i} w^*_{i,j} \left(\tilde{R}^{(k)}_{n;i,j} - E\left[\tilde{R}^{(k)}_{n;i,j}\right]\right) \right\|_1 . \end{aligned}$$

From Assumption 7, note that the expectation on the second term of the previous equation is bounded by $E[|\tilde{Y}^{(k)}_{n;i}|] = E[|\tilde{Y}^{(k)}_{n;i}|1\{|Y_{n;i}| > k\}] \leq 2E[|Y_{n;i}|1\{|Y_{n;i}| > k\}]$. Following the arguments as in Kojevnikov, Marmer, and Song (2020), the second component of the right-hand side in the equation above is bounded by $\sup_{n\geq 1} \max_{i\in\mathcal{I}_n} E\left[|R_{n;i,j}|\, 1\left\{|Y_{n;i}| > k\right\}\right]$, where $\lim_{k\to\infty} \sup_{n\geq 1} \max_{i\in\mathcal{I}_n} E\left[|R_{n;i,j}|\, 1\left\{|Y_{n;i}| > k\right\}\right] = 0$ a.s. Focusing on the first component of the right-hand side, by Lyapunov's inequality, it follows that

$$\left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w^*_{i,j} \left(R^{(k)}_{n;i,j} - E\left[R^{(k)}_{n;i,j} \right] \right) \right\|_1 \leq \left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w^*_{i,j} \left(R^{(k)}_{n;i,j} - E\left[R^{(k)}_{n;i,j} \right] \right) \right\|_2, \quad \text{(B-2)}$$

where (B-2) is an expression for the standard deviation of $\sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w_{i,j} R^{(k)}_{n;i,j}$. Note that

$$\text{var}\left(\sum_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j} \right) = \sum_{i \in \mathcal{I}_n} \text{var}\left(\sum_{j \in \mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j} \right) + \sum_{i \neq h \in \mathcal{I}_n} \text{cov}\left(\sum_{j \in \mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j}, \sum_{s \in \mathcal{I}_h} w^*_{h,s} R^{(k)}_{n;h,s} \right).$$

The variance part of the previous equation can be further expressed as

$$\begin{aligned}
\text{var}\left(\sum_{j \in \mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j} \right) &= \sum_{j \in \mathcal{I}_i} w^{*2}_{i,j} \text{var}(R^{(k)}_{n;i,j}) + \sum_{j \neq s \in \mathcal{I}_i} w^*_{i,j} w^*_{i,s} \text{cov}(R^{(k)}_{n;i,j}, R^{(k)}_{n;i,s}), && \text{(B-3)} \\
&\leq C \sum_{j \in \mathcal{I}_i} w^{*2}_{i,j} + \sum_{j \in \mathcal{I}_i} \sum_{d \geq 1} \sum_{s \in \mathcal{P}_n(j,d) \cap \mathcal{I}_i} |\text{cov}(R^{(k)}_{n;i,j} R^{(k)}_{n;i,s})|, \\
&\leq C \sum_{j \in \mathcal{I}_i} w^{*2}_{i,j} + \psi_{1,1}(\varphi_k, \varphi_k) \sum_{d \geq 1} \lambda_{n,d} \sum_{j \in \mathcal{I}_i} |\mathcal{P}_n(j,d)|,
\end{aligned}$$

where the second inequality follows from $w^*_{i,j}, w^*_{i,s} \in [0,1]$. In the first term of the second inequality, C represents any generic constant from the fact that after the initial partition of $R_{n;i,j}$, the variance of $R^{(k)}_{n;i,j}$ is bounded. The last inequality follows from two reasons. First, from Lemma B.1 under Assumptions 4 and 5, $|\text{cov}(R^{(k)}_{n;i,j} R^{(k)}_{n;i,s})| \leq \psi_{1,1}(\varphi_k, \varphi_k) \lambda_{n,d}$ for $d_n(i,j) = d$ and φ_k is a bounded function with $\text{Lip}(\psi_k) = 1$. Second, the set of indexes $\mathcal{P}_n(j,d)$ are such that $\mathcal{P}_n(j,d) \cap \mathcal{I}_i \subset \mathcal{P}_n(j,d)$. The covariance component can be written as

$$\begin{aligned}
\text{cov}\left(\sum_{j \in \mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j}, \sum_{s \in \mathcal{I}_h} w^*_{h,s} R^{(k)}_{n;h,s} \right) &= \sum_{j \in \mathcal{I}_i} \sum_{s \in \mathcal{I}_h} w^*_{i,j} w^*_{h,s} \text{cov}(R^{(k)}_{n;i,j}, R^{(k)}_{n;h,s}), && \text{(B-4)} \\
&\leq \sum_{j \in \mathcal{I}_i} \sum_{d \geq 1} \sum_{s \in \mathcal{P}_n(j,d) \cap \mathcal{I}_h} |\text{cov}(R^{(k)}_{n;i,j} R^{(k)}_{n;h,s})|, \\
&\leq \psi_{1,1}(\varphi_k, \varphi_k) \sum_{d \geq 1} \lambda_{n,d} \sum_{j \in \mathcal{I}_i} |\mathcal{P}_n(j,d)|,
\end{aligned}$$

where the second and third inequalities follow from the same principles already discussed in the previous paragraph. It follows from Equations (B-3) and (B-4) that the total variance of $\sum_{j \in \mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j}$ can be bounded by

$$\operatorname{var}\left(\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j}\right) = C\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^{*2}_{i,j} + 2\psi_{1,1}\left(\varphi_k,\varphi_k\right)\sum_{i\in\mathcal{I}_n}\sum_{d\geq 1}\lambda_{n,d}\sum_{j\in\mathcal{I}_i}\left|\mathcal{P}_n(j,d)\right|, \quad \text{(B-5)}$$
$$= C\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^{*2}_{i,j} + 2\psi_{1,1}\left(\varphi_k,\varphi_k\right)\sum_{d\geq 1}\lambda_{n,d}\sum_{i\in\mathcal{I}_n}\left|\mathcal{P}_n(j,d)\right|,$$
$$\leq n\left(C\bar{\mathcal{I}}_n + 2\psi_{1,1}\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d}\right),$$

where $\bar{\mathcal{I}}_n = n^{-1}\sum_{i\in\mathcal{I}_n}|\mathcal{I}_i|$ and the inequality follows because $w^{*2}_{i,j}\in[0,1]$. The set $\mathcal{I}_i$ can either be empty, equal to the union of individual i's degree in the networks $\mathcal{G}$ and $\mathcal{G}_0$, or equal to $\mathcal{P}_n(i,1)$ (individual i's degree in network $\mathcal{G}$). Note that $|\mathcal{I}_i|\leq|\mathcal{P}_n(i,1)|$ for all i. Also, $\sum_{i\in\mathcal{I}_n}|\mathcal{P}_n(i,1)|\lambda_{n,1}\leq\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d}$, which converges to zero almost surely by Assumption 6. It follows that $n^{-1}\bar{\mathcal{I}}_n \xrightarrow{\text{a.s.}} 0$. Therefore,

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_{n,m}}\sum_{j\in\mathcal{I}_i} w^*_{i,j}\left(R^{(k)}_{n;i,j} - E\left[R^{(k)}_{n;i,j}\right]\right)\right\|_1 \leq \left(n^{-1}C\bar{\mathcal{I}}_n + 2\psi_{1,1}n^{-1}\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d}\right)^{1/2}. \quad \text{(B-6)}$$

The result follows from $n^{-1}\bar{\mathcal{I}}_n \xrightarrow{\text{a.s.}} 0$ and $n^{-1}\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d} \xrightarrow{\text{a.s.}} 0$ under Assumption 6. ■

Corollary B.1 (LLN for Instruments and Regressors) *Let Assumptions 4 to 7 hold. Then,*

$$\left\|\frac{1}{n}\sum_{i\in\mathcal{I}_n}\left(\mathbf{z}_{n;i}\mathbf{d}^\top_{n;i} - E[\mathbf{z}_{n;i}\mathbf{d}^\top_{n;i}]\right)\right\|_1 \xrightarrow{a.s.} 0.$$

Proof. There are four different types of components in the matrix $\mathbf{Z}^\top_n\mathbf{D}_n$ formed by summation of products of: (1) Non-network regressors of the form $x_{n;i,q}x_{n;i,\ell}$; (2) Network regressors of the form $\mathbf{w}_{n,0;i}\mathbf{x}_{n,q}\mathbf{w}_{n;i}\mathbf{x}_{n;\ell}$; (3) Network and non-network regressors of the form $\mathbf{w}_{n,0;i}\mathbf{x}_{n,q}\mathbf{x}_{n;i,\ell}$; and (4) Network regressors and network outcomes of the form $\mathbf{w}_{n,0;i}\mathbf{x}_{n,q}\mathbf{w}_{n;i}\mathbf{y}_n$ [and the versions of (2) and (3) with $\mathbf{w}^p_{n,0;i}$ instead of $\mathbf{w}_{n,0;i}$]. The LLN follows from Lemma B.2 by choosing $\mathcal{I}_i=\emptyset$ for (1), $\mathcal{I}_i$ as the union of individual i's degree in the networks $\mathcal{G}$ and $\mathcal{G}_0$ in (2), and $\mathcal{I}_i=\mathcal{P}_n(i,1)$ for (3). For (4), note that

$$E[\mathbf{W}_N\mathbf{y}] = \gamma_0\mathbf{W}_N\mathbf{x}_N + (\gamma_0\beta_0+\delta_0)\sum_{p=0}^{\infty}\beta_0^k\mathbf{W}_N^{p+2}\mathbf{x}_N. \quad \text{(B-7)}$$

By choosing $\mathcal{I}_i$ to be the union of individual i's degree in the network $\mathcal{G}$ and the set of individuals at distance p from i (for all $p \in R_+$), Lemma B.2 applies for all the values in the infinite sum formed by $\mathbf{w}_{n,0;i}\mathbf{x}_{n,q}\mathbf{w}_{n;i}\mathbf{y}_n$ after replacing $\mathbf{w}_{n;i}\mathbf{y}_n$ from Equation (B-7) [the same argument holds for (2) and (3) when using $\mathbf{w}^p_{n,0;i}$ instead of $\mathbf{w}_{n,0;i}$]. Given that each component of the sum converges to a finite expectation, the infinite sum of finite expectations is also finite given the restriction on the parameters β_0 from Assumption 2, thus completing the proof. ■

Corollary B.2 (LLN for Instruments and Errors) *Let Assumptions 4 to 7 hold, then*

$$\left\| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \left(\mathbf{z}_{n;i}\varepsilon_{n;i}^\top - E[\mathbf{z}_{n;i}\varepsilon_{n;i}^\top] \right) \right\|_1 \xrightarrow{a.s.} 0.$$

Proof. Given that $\mathbf{r}_{n;i} = [\mathbf{x}_{n;i}, \varepsilon_{n;i}]$ and $\mathbf{z}_{n;i}$ can be divided into both network and non-network components, the proof of this result is analogous to that of Corollary B.1 (1) and (3). ■

Corollary B.3 (Finite Variance) *Define* $\mathbf{S}_n = \mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n$ *and* $\boldsymbol{\Omega}_n = \mathrm{var}(n^{-1/2}\mathbf{S}_n)$ *and let Assumptions 4 to 7 hold, then as* $n \to \infty$, $\boldsymbol{\Omega}_n \xrightarrow{a.s.} \boldsymbol{\Omega}_N < \infty$.

Proof. As before, $n^{-1/2}\mathbf{S}_n \equiv n^{-1/2}\sum_{i=1}^n \mathbf{z}_{n;i}\varepsilon_{n;i}$. The bounded covariance assumptions from Lemma B.1 combined with the arguments in Lemma B.2 guarantee that the following limit $\lim_{n\to\infty} n^{-1}\mathrm{var}\left(\sum_{i=1}^n \mathbf{z}_{n;i}\varepsilon_{n;i}\right)$ is finite. In particular, from Equation (B-6), using the appropriate values for $R_{n;i,j}$ and $\mathcal{I}_i$ (see Corollary B.1), it follows that $\mathrm{var}(\sum_{i=1}^n \mathbf{z}_{n;i}\varepsilon_{n;i}) = O_p(1)$. Given that $\boldsymbol{\Omega}_n$ converges to a finite quantity, it follows that $\boldsymbol{\Omega}_n \xrightarrow{\text{a.s.}} \boldsymbol{\Omega}_N$, where

$$\boldsymbol{\Omega}_N = \lim_{n\to\infty} n^{-1} \left[\sum_{i=1}^n \mathrm{var}(\mathbf{z}_{n;i}\varepsilon_{n;i}) + \sum_{i \neq j} \mathrm{cov}(\mathbf{z}_{n;i}\varepsilon_{n;i}, \mathbf{z}_{n;j}\varepsilon_{n;j}) \right] < \infty.$$

■

Lemma B.3 (Central Limit Theorem) *Let Assumptions 1 and 4-8 hold and define* $S_n \equiv \sum_{i \in \mathcal{I}_n} z_{n;i,q}\varepsilon_{n;i}$, *where* $z_{n;i,q}$ *is the* qth *entrance of the vector* $\mathbf{z}_{n;i}$. *Then, by definition of* $\mathbf{z}_{n;i}$ *and Assumption 1,* $E[z_{n;i,q}\varepsilon_{n;i}] = 0$. *As* $n \to \infty$,

$$\sup_{t \in \mathbf{R}} \left| \mathbf{P}\left\{ \frac{S_n}{\sigma_n} \leq t \mid \mathcal{C}_n \right\} - \Phi(t) \right| \xrightarrow{a.s.} 0,$$

where $\sigma_n \equiv var(S_n)$ *and* $\Phi(\cdot)$ *denotes the cumulative distribution function of a standard normal random variable.*

Proof. Let $Y_{n;i} = z_{n;i,q}\varepsilon_{n;i}$. From Lemma B.1, the covariance of any two $Y_{n;i}$ and $Y_{n;j}$ is bounded. The proof then follows from applying Lemmas A.2 and A.3 in Kojevnikov, Marmer, and Song (2020, Appendix A, pp. 899-907) to $Y_{n;i}$ and S_n/σ_n, respectively. ■

Lemma B.4 (Multivariate Central Limit Theorem) *Let Assumptions 1 and 4-8 hold. Then, as $n \to \infty$, $n^{-1/2}\sum_{i=1}^{n} \mathbf{z}_{n;i}\varepsilon_{n;i} \xrightarrow{d} \mathcal{N}(0, \Omega_N)$.*

Proof. From Lemma B.3, it follows that $n^{-1/2}\sum_{i=1}^{n} z_{n;i,q}\varepsilon_{n;i} \xrightarrow{d} \mathcal{N}(0, \sigma_n^2)$, while from Lemma B.3, it follows that $\mathbf{\Omega}_N$ exists. Therefore, the result follows from an application of the Cramér-Wold device. ■

C Empirical Application

C.1 Data Description

Our dataset was collected between March and May 2011 as part of the Secondary Education Survey in Hong Kong (SESHK). The survey was conducted in the second semester before the final exams and it involved three secondary schools with 873 out 898 students participating. The difference in participants corresponds to students that did not attend school when the survey was taken. For privacy protection, both schools' and students' names are labeled by unique identifiers. The sample includes 7th grade students from all three schools and 8th and 9th grade students from one school (g $\in \{7, 8, 9\}$). Each grade within a school is composed of five different sections (cl $\in \{1, \ldots, 5\}$).

Table 1 shows the summary statistics for the variables we use in our empirical application. `Math test` corresponds to the first mathematics exam score for each student i, on a scale of 0 to 100. The dataset also includes a cognitive ability test (0-16) and information on five personality measures: `Agreeableness` (9-40), `Conscientiousness` (9-45), `Extraversion` (8-40), `Neuroticism` (8-40) and `Openness` (10-55), where the numbers in parenthesis represent the scales for the tests. We preset the values in levels for our variables of interest; however, in our empirical model, we use the log transformation to represent these tests. The dataset also includes gender, family characteristics, and physical-related variables. In our empirical application, we categorize those variables as follows: `Male` is an indicator variable that equals 1 if the student is male, and 0 otherwise. The `Height` for each student is measured in centimeters (cm) and the `Weight` in kilograms (kg). Both the `number of elder and younger siblings` is a count variable, and `Commute by car/taxi` takes the value of 1 when a student goes to school either by car or by taxi.

Our dataset also contains information that allows us to construct indicator variables capturing students' engagement at school. The survey includes a series of questions that students can answer using a scale from strongly agree to strongly disagree. Some of these school engagement questions also contain information about the levels of involvement of family members in students' educational development. For instance, students were asked whether they get help from siblings or parents in their homework. We use those questions to generate the indicator variables `Siblings' Help` and `Parents' Help`, which take the value of 1 if students either somewhat agree, agree, or strongly agree with the statements of whether they receive help from their siblings or parents. We label those variables as 0 otherwise. To capture extracurricular school-related activities, we include the indicator variable `Play Anything`, which equals 1 if students play music, sports, or belong to any school-organized clubs, and 0 otherwise.

C.2 Supplementary Estimation Results

Table 4 shows the rest of the estimated coefficients from Table 3. Specifically, it includes class, school, grade fixed effects, interaction terms between male and logs of personality, number of elder and younger siblings, interaction terms between type of sibling and the indicator variable `Siblings' help`, and the indicator variable `Play Anything`.

For robustness purposes, Tables 5-10 show empirical estimation of Equation (6.1) with the kernel Tukey-Hanning, constant $C \in \{1.5, 1.6, 1.7\}$, and $p \in \{3, 4, 5\}$.

Table 4: Estimation Results, Continuation Table 3 (Tukey-Hanning, $p = 5$ and $C = 1.7$)

	OLS		G2SLS		GMM	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	0.0001	(0.0138)	-0.0030	(0.0126)	0.0087	(0.0116)
Elder Siblings Help	0.0382*	(0.0212)	0.0403**	(0.0201)	0.0434***	(0.0159)
Younger Siblings Help	0.2008***	(0.0576)	0.1528**	(0.0680)	0.1276***	(0.0486)
Male × ln(Cognitive)	-0.0407	(0.0667)	-0.0318	(0.0622)	-0.1319**	(0.0670)
Male × ln(Agreeableness)	0.1582**	(0.0727)	0.1612**	(0.0639)	0.1216**	(0.0520)
Male × ln(Conscientiousness)	0.0434	(0.0641)	0.0565	(0.0559)	0.0265	(0.0405)
Male × ln(Extraversion)	0.1003	(0.0658)	0.1033	(0.0630)	0.0440	(0.0518)
Male × ln(Neuroticism)	-0.1089	(0.0945)	-0.0968	(0.0805)	-0.0157	(0.0566)
Male × ln(Openness)	0.0293*	(0.0176)	0.0417***	(0.0155)	0.1084**	(0.0517)
School 1, grade 7, class 1	-0.1288***	(0.0238)	-0.0682***	(0.0247)	0.0242	(0.0533)
School 1, grade 7, class 2	0.0779***	(0.0111)	0.0658***	(0.0106)	0.0857***	(0.0283)
School 1, grade 7, class 3	0.1177***	(0.0132)	0.0889***	(0.0148)	0.1587***	(0.0525)
School 1, grade 7, class 4	0.1394***	(0.0124)	0.1225***	(0.0154)	0.1902***	(0.0493)
School 1, grade 7, class 5	0.0162	(0.0135)	0.0204	(0.0125)	0.0940***	(0.0275)
School 2, grade 7, class 1	-0.0172	(0.0159)	-0.0120	(0.0132)	0.0518	(0.0326)
School 2, grade 7, class 2	0.0649***	(0.0155)	0.0538***	(0.0153)	0.1354***	(0.0428)
School 2, grade 7, class 3	0.0390**	(0.0161)	0.0472***	(0.0141)	0.0721**	(0.0317)
School 2, grade 7, class 4	0.1480***	(0.0164)	0.1126***	(0.0180)	0.1315***	(0.0449)
School 2, grade 7, class 5	0.1541***	(0.0131)	0.1207***	(0.0159)	0.0952**	(0.0434)
School 3, grade 7, class 1	0.1755***	(0.0117)	0.1335***	(0.0165)	0.0986**	(0.0434)
School 3, grade 7, class 2	0.1916***	(0.0170)	0.1488***	(0.0172)	0.0902**	(0.0438)
School 3, grade 7, class 3	0.1294***	(0.0129)	0.1034***	(0.0153)	0.0721*	(0.0385)
School 3, grade 7, class 4	0.1384***	(0.0151)	0.0944***	(0.0202)	0.0615	(0.0453)
School 3, grade 7, class 5	0.1547***	(0.0096)	0.1232***	(0.0144)	0.1367***	(0.0380)
School 3, grade 8, class 1	0.2016***	(0.0162)	0.1482***	(0.0235)	0.1226**	(0.0520)
School 3, grade 8, class 2	0.1619***	(0.0116)	0.1308***	(0.0150)	0.0984***	(0.0335)
School 3, grade 8, class 3	0.1997***	(0.0172)	0.1493***	(0.0219)	0.1072**	(0.0485)
School 3, grade 8, class 4	0.1545***	(0.0115)	0.1087***	(0.0175)	0.0772*	(0.0411)
School 3, grade 8, class 5	0.0693***	(0.0072)	0.0782***	(0.0075)	0.0408*	(0.0246)
School 3, grade 9, class 1	0.0575***	(0.0090)	0.0557***	(0.0066)	0.0385	(0.0263)
School 3, grade 9, class 2	0.0518***	(0.0080)	0.0543***	(0.0082)	0.0374	(0.0296)
School 3, grade 9, class 3	0.0819***	(0.0084)	0.0798***	(0.0072)	0.0381	(0.0291)
School 3, grade 9, class 4	3.8210***	(1.2635)	2.9464***	(0.9048)	-4.6230*	(2.5819)

Note: (i) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$; and (ii) Standard errors (SE) in parentheses.

Table 5: Estimation Results with $p = 3$

Variables	$C = 1.5$	$C = 1.6$	$C = 1.7$
Peer effect			
ln(Math Test)	0.8850***	0.8888***	0.8920***
	(0.3243)	(0.3157)	(0.3069)
Contextual effects			
Male	-0.2380	-0.2395*	-0.2402*
	(0.1452)	(0.1436)	(0.1417)
ln(Height)	2.2428**	2.2246**	2.2002**
	(1.0373)	(1.0261)	(1.0108)
ln(Weight)	-0.0312	-0.0453	-0.0558
	(0.3415)	(0.3267)	(0.3122)
Siblings Help	0.0722	0.0674	0.0630
	(0.0855)	(0.0833)	(0.0812)
Parents Help	-0.1202	-0.1154	-0.1114
	(0.0857)	(0.0841)	(0.0823)
Commute by Car/Taxi	0.1254	0.1395	0.1515
	(0.1142)	(0.1126)	(0.1110)
Music	0.0898	0.0842	0.0782
	(0.0860)	(0.0841)	(0.0822)
Direct effects			
ln(Cognitive)	0.1314***	0.1318***	0.1322***
	(0.0452)	(0.0437)	(0.0421)
ln(Agreeableness)	-0.1306**	-0.1254**	-0.1205*
	(0.0647)	(0.0632)	(0.0618)
ln(Conscientiousness)	0.0656	0.0649	0.0646
	(0.0721)	(0.0718)	(0.0713)
ln(Extraversion)	-0.1316**	-0.1290**	-0.1259**
	(0.0618)	(0.0614)	(0.0608)
ln(Neuroticism)	-0.0330	-0.0334	-0.0331
	(0.0405)	(0.0396)	(0.0389)
ln(Openness)	0.0160	0.0115	0.0068
	(0.0662)	(0.0654)	(0.0645)
Male	-0.6978	-0.6800	-0.6582
	(0.8374)	(0.8365)	(0.8331)
ln(Height)	-0.9044***	-0.9102***	-0.9099***
	(0.3100)	(0.3046)	(0.2987)
ln(Weight)	-0.0302	-0.0238	-0.0187
	(0.0796)	(0.0759)	(0.0725)
Siblings Help	-0.0654**	-0.0640**	-0.0627**
	(0.0291)	(0.0282)	(0.0273)
Parents Help	0.0386*	0.0378*	0.0371*
	(0.0205)	(0.0203)	(0.0200)
Commute by Car/Taxi	-0.0291	-0.0315*	-0.0335*
	(0.0180)	(0.0176)	(0.0171)
Degree	0.0210***	0.0209***	0.0208***
	(0.0059)	(0.0058)	(0.0057)
Isolate Students	0.4938	0.4858	0.4755
	(0.4872)	(0.4821)	(0.4766)
Constant	-6.2953*	-6.1603*	-6.0293*
	(3.3383)	(3.3041)	(3.2567)
n	868	868	868
Adjusted R^2	0.2619	0.2617	0.2615
RMSE	0.2089	0.2092	0.2094

Note: (i) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$; and (ii) Standard errors (SE) in parentheses.

Table 6: Estimation Results, Continuation Table 5 ($p = 3$)

	$C = 1.5$		$C = 1.6$		$C = 1.7$	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	0.0098	(0.0176)	0.0086	(0.0173)	0.0072	(0.0169)
Elder Siblings Help	0.0471**	(0.0220)	0.0470**	(0.0218)	0.0470**	(0.0216)
Younger Siblings Help	0.0818	(0.0929)	0.0802	(0.0902)	0.0788	(0.0875)
Male × ln(Cognitive)	-0.0347	(0.1036)	-0.0349	(0.1017)	-0.0354	(0.0997)
Male × ln(Agreeableness)	0.1338*	(0.0768)	0.1334*	(0.0751)	0.1324*	(0.0736)
Male × ln(Conscientiousness)	0.0341	(0.0640)	0.0313	(0.0632)	0.0286	(0.0626)
Male × ln(Extraversion)	0.0731	(0.0776)	0.0700	(0.0776)	0.0661	(0.0776)
Male × ln(Neuroticism)	0.0216	(0.0846)	0.0239	(0.0821)	0.0261	(0.0798)
Male × ln(Openness)	0.1272**	(0.0615)	0.1202**	(0.0604)	0.1135*	(0.0592)
School 1, grade 7, class 1	0.0735	(0.1056)	0.0670	(0.1028)	0.0607	(0.0998)
School 1, grade 7, class 2	0.0734*	(0.0380)	0.0711*	(0.0367)	0.0687*	(0.0354)
School 1, grade 7, class 3	0.1168*	(0.0703)	0.1097	(0.0687)	0.1033	(0.0671)
School 1, grade 7, class 4	0.1672**	(0.0652)	0.1595**	(0.0639)	0.1524**	(0.0624)
School 1, grade 7, class 5	0.1226***	(0.0437)	0.1190***	(0.0414)	0.1145***	(0.0392)
School 2, grade 7, class 1	0.0643	(0.0554)	0.0621	(0.0532)	0.0609	(0.0509)
School 2, grade 7, class 2	0.1247**	(0.0582)	0.1174**	(0.0564)	0.1107**	(0.0544)
School 2, grade 7, class 3	0.1200**	(0.0482)	0.1154**	(0.0463)	0.1113**	(0.0444)
School 2, grade 7, class 4	0.1086	(0.0664)	0.1049	(0.0644)	0.1016	(0.0624)
School 2, grade 7, class 5	0.0827	(0.0579)	0.0778	(0.0569)	0.0737	(0.0558)
School 3, grade 7, class 1	0.0771	(0.0636)	0.0727	(0.0623)	0.0688	(0.0611)
School 3, grade 7, class 2	0.0663	(0.0612)	0.0634	(0.0600)	0.0609	(0.0587)
School 3, grade 7, class 3	0.0741	(0.0500)	0.0701	(0.0488)	0.0665	(0.0475)
School 3, grade 7, class 4	0.0300	(0.0621)	0.0268	(0.0611)	0.0241	(0.0600)
School 3, grade 7, class 5	0.1101**	(0.0520)	0.1078**	(0.0508)	0.1055**	(0.0496)
School 3, grade 8, class 1	0.0744	(0.0790)	0.0716	(0.0770)	0.0690	(0.0751)
School 3, grade 8, class 2	0.0639	(0.0503)	0.0611	(0.0492)	0.0586	(0.0480)
School 3, grade 8, class 3	0.0642	(0.0726)	0.0602	(0.0714)	0.0567	(0.0700)
School 3, grade 8, class 4	0.0470	(0.0609)	0.0438	(0.0592)	0.0407	(0.0574)
School 3, grade 8, class 5	0.0543*	(0.0304)	0.0515*	(0.0301)	0.0486	(0.0298)
School 3, grade 9, class 1	0.0297	(0.0346)	0.0269	(0.0338)	0.0238	(0.0331)
School 3, grade 9, class 2	0.0351	(0.0406)	0.0308	(0.0394)	0.0270	(0.0383)
School 3, grade 9, class 3	0.0513	(0.0337)	0.0500	(0.0336)	0.0488	(0.0334)
School 3, grade 9, class 4	-6.2953*	(3.3383)	-6.1603*	(3.3041)	-6.0293*	(3.2567)

Note: (i) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$; and (ii) Standard errors (SE) in parentheses.

References

Jenish, Nazgul and Ingmar R Prucha. 2009. "Central limit theorems and uniform laws of large numbers for arrays of random fields." *Journal of econometrics* 150 (1):86–98.

Table 7: Estimation Results with $p = 4$

Variables	$C = 1.5$	$C = 1.6$	$C = 1.7$
Peer effect			
ln(Math Test)	0.6733***	0.6764***	0.6798***
	(0.2236)	(0.2159)	(0.2090)
Contextual effects			
Male	-0.2057*	-0.2111*	-0.2161**
	(0.1115)	(0.1080)	(0.1046)
ln(Height)	2.2427***	2.2003***	2.1566**
	(0.8567)	(0.8471)	(0.8358)
ln(Weight)	0.0446	0.0461	0.0493
	(0.2703)	(0.2653)	(0.2608)
Siblings Help	0.0388	0.0399	0.0419
	(0.0567)	(0.0546)	(0.0525)
Parents Help	-0.0205	-0.0133	-0.0073
	(0.0657)	(0.0639)	(0.0619)
Commute by Car/Taxi	0.0649	0.0733	0.0799
	(0.0688)	(0.0669)	(0.0653)
Music	0.1161*	0.1143**	0.1131**
	(0.0600)	(0.0582)	(0.0565)
Direct effects			
ln(Cognitive)	0.1098***	0.1107***	0.1119***
	(0.0378)	(0.0361)	(0.0345)
ln(Agreeableness)	-0.0963*	-0.0902*	-0.0844*
	(0.0500)	(0.0483)	(0.0466)
ln(Conscientiousness)	0.0915*	0.0930*	0.0946*
	(0.0517)	(0.0501)	(0.0487)
ln(Extraversion)	-0.1226***	-0.1206**	-0.1184**
	(0.0474)	(0.0469)	(0.0462)
ln(Neuroticism)	-0.0150	-0.0136	-0.0122
	(0.0294)	(0.0284)	(0.0275)
ln(Openness)	0.0220	0.0179	0.0137
	(0.0493)	(0.0475)	(0.0458)
Male	-0.4718	-0.4443	-0.4137
	(0.6411)	(0.6298)	(0.6167)
ln(Height)	-0.9811***	-0.9766***	-0.9690***
	(0.2696)	(0.2625)	(0.2553)
ln(Weight)	-0.0685	-0.0668	-0.0657
	(0.0675)	(0.0654)	(0.0637)
Siblings Help	-0.0347	-0.0351	-0.0360*
	(0.0235)	(0.0227)	(0.0218)
Parents Help	0.0210	0.0186	0.0168
	(0.0168)	(0.0163)	(0.0158)
Commute by Car/Taxi	-0.0160	-0.0175	-0.0188
	(0.0156)	(0.0153)	(0.0150)
Degree	0.0248***	0.0247***	0.0247***
	(0.0044)	(0.0042)	(0.0041)
Isolate Students	0.2402	0.2185	0.1960
	(0.3496)	(0.3417)	(0.3328)
Constant	-5.2308*	-5.0675*	-4.9186*
	(2.7821)	(2.7405)	(2.6993)
n	868	868	868
Adjusted R^2	0.2801	0.2813	0.2821
RMSE	0.1980	0.1979	0.1980

Note: (i) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$; and (ii) Standard errors (SE) in parentheses.

Table 8: Estimation Results, Continuation Table 7 ($p = 4$)

	$C = 1.5$		$C = 1.6$		$C = 1.7$	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	0.0052	(0.0129)	0.0055	(0.0125)	0.0060	(0.0121)
Elder Siblings Help	0.0372*	(0.0192)	0.0380**	(0.0190)	0.0390**	(0.0188)
Younger Siblings Help	0.1055	(0.0675)	0.1049	(0.0640)	0.1043*	(0.0609)
Male × ln(Cognitive)	-0.0848	(0.0879)	-0.0866	(0.0857)	-0.0885	(0.0836)
Male × ln(Agreeableness)	0.1310**	(0.0634)	0.1296**	(0.0614)	0.1275**	(0.0599)
Male × ln(Conscientiousness)	0.0489	(0.0488)	0.0473	(0.0461)	0.0460	(0.0437)
Male × ln(Extraversion)	0.0721	(0.0620)	0.0703	(0.0610)	0.0683	(0.0601)
Male × ln(Neuroticism)	-0.0238	(0.0706)	-0.0241	(0.0672)	-0.0246	(0.0641)
Male × ln(Openness)	0.1284**	(0.0519)	0.1232**	(0.0510)	0.1188**	(0.0502)
School 1, grade 7, class 1	0.0505	(0.0677)	0.0469	(0.0656)	0.0441	(0.0636)
School 1, grade 7, class 2	0.0896***	(0.0324)	0.0879***	(0.0312)	0.0866***	(0.0301)
School 1, grade 7, class 3	0.1605***	(0.0563)	0.1568***	(0.0555)	0.1538***	(0.0547)
School 1, grade 7, class 4	0.1953***	(0.0528)	0.1902***	(0.0518)	0.1858***	(0.0507)
School 1, grade 7, class 5	0.0985***	(0.0343)	0.0960***	(0.0320)	0.0934***	(0.0299)
School 2, grade 7, class 1	0.0584	(0.0450)	0.0571	(0.0430)	0.0570	(0.0412)
School 2, grade 7, class 2	0.1376***	(0.0457)	0.1335***	(0.0443)	0.1303***	(0.0428)
School 2, grade 7, class 3	0.0852**	(0.0343)	0.0824**	(0.0328)	0.0802**	(0.0312)
School 2, grade 7, class 4	0.1242**	(0.0494)	0.1226**	(0.0479)	0.1218***	(0.0464)
School 2, grade 7, class 5	0.1040**	(0.0477)	0.0994**	(0.0468)	0.0954**	(0.0460)
School 3, grade 7, class 1	0.1008**	(0.0498)	0.0988**	(0.0486)	0.0967**	(0.0476)
School 3, grade 7, class 2	0.0939*	(0.0497)	0.0909*	(0.0483)	0.0881*	(0.0469)
School 3, grade 7, class 3	0.0791*	(0.0431)	0.0757*	(0.0416)	0.0728*	(0.0401)
School 3, grade 7, class 4	0.0559	(0.0507)	0.0541	(0.0497)	0.0526	(0.0487)
School 3, grade 7, class 5	0.1301***	(0.0396)	0.1290***	(0.0383)	0.1279***	(0.0373)
School 3, grade 8, class 1	0.1120*	(0.0590)	0.1101*	(0.0572)	0.1084*	(0.0555)
School 3, grade 8, class 2	0.0903**	(0.0382)	0.0879**	(0.0371)	0.0857**	(0.0361)
School 3, grade 8, class 3	0.1057*	(0.0550)	0.1025*	(0.0536)	0.0994*	(0.0523)
School 3, grade 8, class 4	0.0643	(0.0465)	0.0635	(0.0450)	0.0628	(0.0437)
School 3, grade 8, class 5	0.0475*	(0.0256)	0.0442*	(0.0256)	0.0413	(0.0253)
School 3, grade 9, class 1	0.0424	(0.0296)	0.0390	(0.0292)	0.0359	(0.0289)
School 3, grade 9, class 2	0.0483	(0.0342)	0.0434	(0.0332)	0.0394	(0.0323)
School 3, grade 9, class 3	0.0358	(0.0309)	0.0339	(0.0307)	0.0319	(0.0302)
School 3, grade 9, class 4	-5.2308*	(2.7821)	-5.0675*	(2.7405)	-4.9186*	(2.6993)

Note: (i) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$; and (ii) Standard errors (SE) in parentheses.

Kojevnikov, Denis, Vadim Marmer, and Kyungchul Song. 2020. “Limit theorems for network dependent random variables.” *Journal of Econometrics* .

Table 9: Estimation Results with $p = 5$

Variables	$C = 1.5$	$C = 1.6$	$C = 1.7$
Peer effect			
ln(Math Test)	0.5945***	0.5997***	0.6054***
	(0.1923)	(0.1835)	(0.1757)
Contextual effects			
Male	-0.2493***	-0.2465***	-0.2455***
	(0.0949)	(0.0909)	(0.0873)
ln(Height)	2.3103***	2.2549***	2.1787***
	(0.7915)	(0.7658)	(0.7404)
ln(Weight)	0.0490	0.0395	0.0325
	(0.2122)	(0.2031)	(0.1943)
Siblings Help	0.0263	0.0274	0.0296
	(0.0521)	(0.0505)	(0.0491)
Parents Help	-0.0326	-0.0207	-0.0103
	(0.0590)	(0.0554)	(0.0517)
Commute by Car/Taxi	0.1148*	0.1204*	0.1253**
	(0.0683)	(0.0659)	(0.0636)
Music	0.1243**	0.1206**	0.1187**
	(0.0531)	(0.0505)	(0.0481)
Direct effects			
ln(Cognitive)	0.1093***	0.1087***	0.1086***
	(0.0317)	(0.0300)	(0.0286)
ln(Agreeableness)	-0.0866*	-0.0787	-0.0711
	(0.0501)	(0.0479)	(0.0458)
ln(Conscientiousness)	0.0752*	0.0775*	0.0799**
	(0.0435)	(0.0418)	(0.0406)
ln(Extraversion)	-0.1229***	-0.1202***	-0.1165***
	(0.0420)	(0.0412)	(0.0405)
ln(Neuroticism)	-0.0133	-0.0124	-0.0113
	(0.0278)	(0.0266)	(0.0257)
ln(Openness)	0.0353	0.0311	0.0272
	(0.0456)	(0.0435)	(0.0414)
Male	-0.2091	-0.1906	-0.1642
	(0.5774)	(0.5607)	(0.5423)
ln(Height)	-1.0290***	-1.0132***	-0.9926***
	(0.2435)	(0.2342)	(0.2251)
ln(Weight)	-0.0577	-0.0537	-0.0496
	(0.0516)	(0.0483)	(0.0453)
Siblings Help	-0.0394*	-0.0391*	-0.0395*
	(0.0224)	(0.0214)	(0.0204)
Parents Help	0.0224	0.0193	0.0168
	(0.0155)	(0.0147)	(0.0139)
Commute by Car/Taxi	-0.0216	-0.0234*	-0.0250*
	(0.0145)	(0.0140)	(0.0136)
Degree	0.0251***	0.0250***	0.0249***
	(0.0037)	(0.0036)	(0.0034)
Isolate Students	0.2024	0.1751	0.1425
	(0.3118)	(0.3022)	(0.2933)
Constant	-5.0848*	-4.8886*	-4.6230*
	(2.7561)	(2.6654)	(2.5819)
n	868	868	868
Adjusted R^2	0.2595	0.2637	0.2672
RMSE	0.2019	0.2010	0.2002

Note: (i) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$; and (ii) Standard errors (SE) in parentheses.

Table 10: Estimation Results, Continuation Table 9 ($p = 5$)

	$C = 1.5$		$C = 1.6$		$C = 1.7$	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	0.0088	(0.0125)	0.0086	(0.0121)	0.0087	(0.0116)
Elder Siblings Help	0.0410**	(0.0169)	0.0421**	(0.0164)	0.0434***	(0.0159)
Younger Siblings Help	0.1349**	(0.0538)	0.1312**	(0.0510)	0.1276***	(0.0486)
Male × ln(Cognitive)	-0.1317*	(0.0730)	-0.1321*	(0.0702)	-0.1319**	(0.0670)
Male × ln(Agreeableness)	0.1256**	(0.0563)	0.1240**	(0.0538)	0.1216**	(0.0520)
Male × ln(Conscientiousness)	0.0337	(0.0471)	0.0298	(0.0437)	0.0265	(0.0405)
Male × ln(Extraversion)	0.0426	(0.0538)	0.0436	(0.0527)	0.0440	(0.0518)
Male × ln(Neuroticism)	-0.0164	(0.0633)	-0.0152	(0.0598)	-0.0157	(0.0566)
Male × ln(Openness)	0.1197**	(0.0543)	0.1141**	(0.0530)	0.1084**	(0.0517)
School 1, grade 7, class 1	0.0304	(0.0583)	0.0270	(0.0557)	0.0242	(0.0533)
School 1, grade 7, class 2	0.0876***	(0.0309)	0.0869***	(0.0296)	0.0857***	(0.0283)
School 1, grade 7, class 3	0.1715***	(0.0540)	0.1650***	(0.0533)	0.1587***	(0.0525)
School 1, grade 7, class 4	0.2040***	(0.0521)	0.1971***	(0.0507)	0.1902***	(0.0493)
School 1, grade 7, class 5	0.1028***	(0.0324)	0.0984***	(0.0298)	0.0940***	(0.0275)
School 2, grade 7, class 1	0.0572	(0.0369)	0.0541	(0.0347)	0.0518	(0.0326)
School 2, grade 7, class 2	0.1480***	(0.0456)	0.1416***	(0.0443)	0.1354***	(0.0428)
School 2, grade 7, class 3	0.0800**	(0.0356)	0.0759**	(0.0336)	0.0721**	(0.0317)
School 2, grade 7, class 4	0.1386***	(0.0478)	0.1349***	(0.0463)	0.1315***	(0.0449)
School 2, grade 7, class 5	0.1046**	(0.0460)	0.1003**	(0.0446)	0.0952**	(0.0434)
School 3, grade 7, class 1	0.1045**	(0.0458)	0.1020**	(0.0446)	0.0986**	(0.0434)
School 3, grade 7, class 2	0.0968**	(0.0471)	0.0938**	(0.0454)	0.0902**	(0.0438)
School 3, grade 7, class 3	0.0816**	(0.0409)	0.0768*	(0.0397)	0.0721*	(0.0385)
School 3, grade 7, class 4	0.0681	(0.0468)	0.0650	(0.0461)	0.0615	(0.0453)
School 3, grade 7, class 5	0.1427***	(0.0403)	0.1396***	(0.0391)	0.1367***	(0.0380)
School 3, grade 8, class 1	0.1313**	(0.0551)	0.1268**	(0.0535)	0.1226**	(0.0520)
School 3, grade 8, class 2	0.1048***	(0.0352)	0.1015***	(0.0343)	0.0984***	(0.0335)
School 3, grade 8, class 3	0.1167**	(0.0509)	0.1122**	(0.0496)	0.1072**	(0.0485)
School 3, grade 8, class 4	0.0817*	(0.0439)	0.0793*	(0.0425)	0.0772*	(0.0411)
School 3, grade 8, class 5	0.0450*	(0.0254)	0.0429*	(0.0251)	0.0408*	(0.0246)
School 3, grade 9, class 1	0.0415	(0.0279)	0.0400	(0.0271)	0.0385	(0.0263)
School 3, grade 9, class 2	0.0476	(0.0324)	0.0421	(0.0309)	0.0374	(0.0296)
School 3, grade 9, class 3	0.0430	(0.0306)	0.0407	(0.0299)	0.0381	(0.0291)
School 3, grade 9, class 4	-5.0848*	(2.7561)	-4.8886*	(2.6654)	-4.6230*	(2.5819)

Note: (i) * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$; and (ii) Standard errors (SE) in parentheses.