Estimating Social Effects with Randomized and Observational Network Data

TszKin Julian Chan
Bates White Economic Consulting
Juan Estrada
Quantitative Theory and Methods Department, 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

February 23, 2024

Abstract

This paper proposes a novel method to identify and estimate the parameters of interest in the popular so-called linear-in-means regression model in situations where initial randomization of peers induces the observed network of interest. We argue that initially randomized peers do not generate social effects. However, after randomization, agents can endogenously form relevant connections that can create peer influences. We introduce a moment condition that aggregates local heterogeneous identifying information for all agents in the population. Assuming ψ-dependence in the endogenous network space, a Generalized Method of Moments (GMM) estimator is then proposed that is shown to be consistent, asymptotically normally distributed, and also easy to implement using widely used existing statistical software because of its closed form definition. Monte Carlo exercises confirm the good small-sample performance of the proposed GMM estimator, and an empirical application using data from high-school students in Hong Kong finds strong positive spillover effects of math test scores among study partners in our sample assuming that their observed seatmates were exogenously assigned by their teachers.

Keywords: Social Networks; Instrumental Variables; Causal Inference; ψ-dependence

*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.

†Corresponding Author

1 Introduction

In many observational studies in economics and other social settings, a unit of observation's outcome (e.g., purchase of an item, health well-being of an individual, performance of a firm, or test scores of a student) depends not only on this unit's own observed characteristics (direct effect), but also on the outcome (peer effects) and characteristics (contextual effects) of other observations (peers) in the sample with which they have a link. The workhorse model in economics and other social sciences for this type of setting is the so-called *linear-in-means* regression model (see Section 3.1 in de Paula, 2017, pp. 275-289) where the outcome variable for observation i (e.g., a person, a firm, or a country), $y_{n,i}$, is determined according to

$$y_{n;i} = \beta_0 \sum_{j \neq i} w_{n;i,j} y_{n;j} + \sum_{j \neq i} w_{n;i,j} \mathbf{x}_{n;j}^{\top} \boldsymbol{\delta}_0 + \widetilde{\mathbf{x}}_{n;i}^{\top} \boldsymbol{\gamma}_0 + \varepsilon_{n;i}, \tag{1.1}$$

where $i, j \in \{1, \ldots, n\}$ are also known as *nodes*, $\widetilde{\mathbf{x}}_{n;i} = [1, \mathbf{x}_{n;i}^{\top}]^{\top}$, and $\mathbf{x}_{n;i}$ is a vector of attributes that characterizes observations i, $w_{n;i,j} = 1$ if j is connected to i (an *edge*), and 0 otherwise, $\varepsilon_{n;i}$ represents an unobserved latent error, and n is the number of observations or nodes in the sample. The structure of the *social* network is fully characterized by the square $n \times n$ matrix, $\mathbf{W}_n$, with the entry (i, j) given by $w_{n;i,j}$, that is, the adjacency matrix. The structural parameters β_0 and $\boldsymbol{\delta}_0$ capture the peer and contextual effects, respectively, while $\boldsymbol{\gamma}_0$ captures the direct effects of the observation's own characteristics. They are jointly known as the social (or neighbor) effect parameters and the object of interest in empirical studies with network data; 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.

Sufficient conditions under which the social parameters in (1.1) can be uniquely recovered from the estimating sample $\left\{y_{n;i}, \mathbf{x}_{n;i}^{\top}, \{w_{n;i,j}\}_{j=1,j\neq i}^{n}\right\}_{i=1}^{n}$ is well understood under the assumption that the adjacency matrix, $\mathbf{W}_n$, is exogenous; see, e.g., de Paula (2017) and references therein. However, the endogenous case remains an active area of research among economists and social scientists due to simultaneity bias, measurement error in link information, or because there might be a natural correlation between covariates and the error term in (1.1) (homophily); see, e.g., Johnsson and Moon (2019) and references therein.

This article proposes the use of a type of multilayered (multidimensional) network data structure known as *multiplex* networks (Boccaletti et al., 2014; Kivela et al., 2014) to consistently estimate and perform correct inferences on the structural social parameters in (1.1) with a potentially endogenous network structure, $\mathbf{W}_n$. In particular, it is assumed that the researcher observes another set of social ties among the original observations, $\{w_{n,0;i,j}\}_{j=1,j\neq i}^{n}$, in the form of $n \times n$ adjacency matrix $\mathbf{W}_{n,0}$ that are exogenous in the usual sense.

The assumption of network exogeneity in $\mathbf{W}_{n,0}$ is motivated from the literature on experimental settings with interactions; see, e.g., Athey and Imbens (2017). Network randomization has been used in a wide range of applied settings, such as educational achievement (Sacerdote, 2001), entrepreneurship team performance (Hasan and Koning, 2019), and business performance (Cai and Szeidl, 2018), just to name a few. Furthermore, recent studies have emphasized the importance of endogenous choices of individuals to interact with or avoid their randomized peers as mediators of peer effects (Hasan and Koning, 2019). For example, Carrell, Sacerdote, and West (2013) finds that after assigning cadets to squadrons (initial group assignment), i.e., $\mathbf{W}_{n,0}$, they endogenously sorted into homogeneous groups based on their ability, i.e., $\mathbf{W}_n$. In the context of worker productivity, Mas and Moretti (2009), Hjort (2014), and Kato and Shu (2016) show that peer effects matter only when individuals interact frequently or share similar origins or ethnic divisions, even if they all belong to the same group.

Our approach has an advantage over previous methods to identify peer and contextual effects in that it does not rely on an explicit characterization of the network formation process; see, e.g., Goldsmith-Pinkham and Imbens (2013), Qu and Lee (2015), Johnsson and Moon (2019), Qu, Lee, and Yang (2021), and Auerbach (2022). In addition to avoiding estimating the parameters of a potentially complex nonlinear outcome, the flexibility of our method allows us to consider a wider range of potential endogeneity sources. One particularly important advantage is that our method can handle the presence of endogeneity caused by the simultaneous determination of the outcome and the network of interest. Existing methods that require modeling network formation explicitly assume that the networks are dependent on some endogenous variables and that the outcomes are determined after the endogenous network is formed. Extending those models to include a simultaneous system of equations that include both outcomes and dyadic linking choices requires one to fundamentally change the assumptions in the model, which can drastically affect the identification results. However, because our method does not directly model the formation of the links, we do not have to explicitly consider the potential simultaneous determination of the outcome and the network, making our simple approach powerful for dealing with different sources of endogeneity.

The paper uses the exogenous network, $\mathbf{W}_{n,0}$, to construct a set of valid moment conditions that point-identify parameters in (1.1). The resulting linear Generalized Method of Moments (GMM) estimator is very simple to implement using standard Instrumental Variable (IV) routines in existing statistical software like `Python`, `R` or `Stata`; see, e.g., Estrada et al. (2023). Furthermore, the estimator is shown to be asymptotically normal at the standard root–n rate of convergence, and a consistent estimator of the efficient asymptotic variance-covariance is proposed to perform asymptotically valid inference.

Our proof method allows us to characterize the asymptotic variance-covariance that takes into account the presence of network dependence between individuals generated by the endogenous network $\mathbf{W}_n$. In particular, we assume that the triangular array for the random vector of observed and unobserved characteristics is ψ-dependent (see, i.e., Doukhan and Louhichi, 1999), and the levels of dependence between individuals decrease with their distance in the network space spanned by $\mathbf{W}_n$. Moreover, our asymptotic normality result requires that the levels of dependence decrease at a rate that is fast enough to compensate for any potential increase in the asymptotic levels of network density. Finally, our large sample variance-covariance characterization also connects the asymptotic levels of sparsity in the exogenous network, $\mathbf{W}_{n,0}$, with the precision of the network effects parameters. In particular, we show that the precision of the parameters increases with the number of nodes for which we can find indirect paths of distance of at least two in $\mathbf{W}_{n,0}$. Of course, large values of network density are associated with a lower likelihood of finding indirect connections between individuals in that network, which will consequently impact the precision of our proposed estimator.

The structure of this paper is as follows. Section 2 provides various definitions and notation used throughout. Section 3 introduces the model and conditions for the parameters of the basic model to be uniquely recovered (point identification) from the estimation sample $\left\{y_i, \mathbf{x}_i^\top, \{w_{i,j}\}_{j=1,j\neq i}^n, \{w_{0;i,j}\}_{j=1,j\neq i}^n\right\}_{i=1}^n$. Section 4 describes the proposed linear GMM estimator, its asymptotic distribution, and how to calculate valid asymptotic standard errors. Section 5 provides the Monte Carlo exercises showing the small sample properties of the estimator, while an empirical illustration of the proposed methodology is discussed in Section 6. Finally, Section 7 concludes. The supplemental materials contain all mathematical proofs of the main results, as well as further details on the real data illustration.

2 Preliminaries

We assume the full observability of two types of networks over the same set of nodes. One is the endogenous network of interest that can create social externalities; the other is the instrumental network induced by random assignment. For $N \in \mathbb{N}_+ \equiv \{1, 2, \ldots\}$, let V be the set of N nodes. We denote by $\mathcal{G}_N = (V, E)$ the population network of interest and by $\mathcal{G}_{N,0} = (V, E_0)$ the instrumental population network, where E and E_0 are the corresponding sets of links (Edges). Similarly, as in Graham (2020), the observed networks of size $n < N$ are denoted as $\mathcal{G}_n$ and $\mathcal{G}_{n,0}$, 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 networks $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$ with their respective adjacency matrices; that is, $\mathbf{W}_N = [w_{N;i,j}]$ 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 $1 \times N$ vectors $\mathbf{w}_{N,i} = [w_{N;i,1}, \ldots, w_{N;i,N}] \in [0, 1]^N$ and $\mathbf{w}_{N,0;i} = [w_{N,0;i,1}, \ldots, w_{N,0;i,N}] \in [0, 1]^N$ to be the ith row of the adjacency matrices $\mathbf{W}_N$ and $\mathbf{W}_{N,0}$, respectively. Similarly $\mathbf{w}^p_{N,0;i}$ represents the ith row of powers $p \geq 1$ of the adjacency matrix $\mathbf{W}^p_{N,0}$. Adjacency matrices of 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 *endogenous* network of interest $\mathcal{G}_n$ induced by the sample size n. Let A and B be any two sets of individuals of size $a, b \in \mathbb{N}_+$. We define the distance between sets as $d_n(A, B) = \min_{i \in A} \min_{j \in B} d_n(i, j)$.

Based on the definition of set distance, we define the following group of nodes sets use in the asymptotic results presented in Section 4: (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 from node i is $\mathcal{P}_n^+(i, d) = \{j \in \mathcal{I}_n : d_n(i, j) \geq d\}$, $\mathcal{P}_n(i, d) = \{j \in \mathcal{I}_n : d_n = d\}$, and $\mathcal{P}_n^-(i, d) = \{j \in \mathcal{I}_n : d_n \leq d\}$.

For any random vector $\mathbf{r}_{N;i} \in \mathbb{R}^L$ for some $L \in \mathbb{N}_+$, we endow $\mathbb{R}^{L \times a}$ with the distance measure $\boldsymbol{d}_a(\mathbf{x}, \mathbf{y}) = \sum_{l=1}^{a} \|x_l - y_l\|_2$, for $a \in \mathbb{N}_+$, and where $\| \cdot \|_2$ denotes the Euclidean norm and $(\mathbf{x}, \mathbf{y}) \in \mathbb{R}^{L \times a}$. We let $\mathscr{L}_{L,a}$ denote the collection of bounded Lipschitz real functions that map values from $\mathbb{R}^{L \times a}$ to $\mathbb{R}$. For any set of individuals 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 such as $\{\lambda_n\}_{n \geq 1}$ as $\{\lambda_n\}$. As in Doukhan and Louhichi (1999) and Kojevnikov, Marmer, and Song (2021), we define $\psi-$dependence as follows.

Definition 2.1 (ψ-dependence) *A triangular array $\{\mathbf{r}_{n;i}\}$, $n \geq 1$, $\mathbf{r}_{n;i} \in \mathbb{R}^L$ is ψ-dependent if for each $n \in \mathbb{N}_+$ there exists a sequence $\{\lambda_n\} \equiv \{\lambda_{n,d}\}_{d \geq 0}, \lambda_{n,0} = 1$ and a collection of non-random functions $(\psi_{a,b})_{a,b \in \mathbb{N}}, \psi_{a,b} : \mathscr{L}_{v,a} \times \mathscr{L}_{v,b} \to [0, \infty)$, such that for all $A, B \in \mathcal{P}_N^+(a, b, d)$ for $d > 0$ and all $f \in \mathscr{L}_{L,a}$ and $g \in \mathscr{L}_{L,b}$,*

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

The sequence $\{\lambda_n\}$ is called the *dependence coefficients* of $\mathbf{r}_{n;i}$. The covariance of the nonlinear 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.

3 Peer Effects Model and Identification

There is an infinite population of agents in the set $\mathcal{I}_N$. Each agent $i \in \mathcal{I}_N$ 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, i.e., $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$. The data generation process characterizing the two observed networks adheres to a framework where the network $\mathcal{G}_N$ contains connections formed by two agents making endogenous decisions to participate in social or professional relationships, while $\mathcal{G}_{N,0}$ represents the network formed through the random (or quasirandom) allocation of individuals into groups. The endogenous network formation determining $\mathcal{G}_N$ induces a potential correlation between the individuals' decisions to connect and their observed and unobserved characteristics (see, e.g., Hasan and Koning, 2019). Conversely, by the properties of randomization, the network $\mathcal{G}_{N,0}$ is strictly exogenous (see, e.g., Athey and Imbens, 2017).

Let $\mathscr{G}$ be the discrete set of possible network configurations of size N, such that, for any possible realizations of the network of interest and the exogenous network, $\mathbf{g}$ and $\mathbf{g}_0$, it follows that $\{\mathbf{g}, \mathbf{g}_0\} \in \mathscr{G}$. Let $\mathbf{X}_N = [\mathbf{x}_{N;1}, \cdots, \mathbf{x}_{N;N}]^\top \in \mathscr{X}$. Here, $\mathscr{X}$ is the space that represents the support for the matrix of regressors, where we assume $\mathscr{X} \subseteq \mathbb{R}^{N\times K}$, without loss of generality. Finally, let $\boldsymbol{\varepsilon}_N = [\varepsilon_{N;1}, \cdots, \varepsilon_{N;N}]^\top \in \mathbb{R}^N$. We denote the realizations of $\mathbf{X}_N$ and $\boldsymbol{\varepsilon}_N$ by $\mathbf{X} \in \mathscr{X}$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^N$, respectively.

Formally, we assume that there is a population joint probability distribution function that determines the dependence patterns between the regressors, the networks, and the errors, i.e., the joint distribution is defined as $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g},\mathbf{g}_0,\boldsymbol{\varepsilon}) = \Pr(\mathcal{G}_N = \mathbf{g}, \mathcal{G}_{N,0} = \mathbf{g}_0 \mid \mathbf{X},\boldsymbol{\varepsilon}) f_{\mathbf{X}_N,\boldsymbol{\varepsilon}_N}(\mathbf{X},\boldsymbol{\varepsilon})$, where $\Pr(\mathcal{G}_N = \mathbf{g}, \mathcal{G}_{N,0} = \mathbf{g}_0 \mid \mathbf{X}_N, \boldsymbol{\varepsilon}_N)$ is the probability that $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$ take on particular structures $\{\mathbf{g},\mathbf{g}_0\} \in \mathscr{G}$ conditional on the regressors and the errors, while $f_{\mathbf{X}_N,\boldsymbol{\varepsilon}_N}$ represents the joint probability distribution function for $\mathbf{X}_N$ and $\boldsymbol{\varepsilon}_N$. Note that the joint distribution, $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g},\mathbf{g}_0,\boldsymbol{\varepsilon})$, is characterized by two main features: First, *validation*, the network $\mathcal{G}_{N,0}$ is independent of the agents' observed and unobserved characteristics because of randomization. Second, *relevance*, agents randomly assigned to the same group in $\mathcal{G}_{N,0}$ are more likely to form connections in $\mathcal{G}_N$ (see, e.g., Granovetter, 1973; Gargiulo and Benassi, 2000; Kim, Oh, and Swaminathan, 2006; Goldsmith-Pinkham and Imbens, 2013).

The following assumption imposes the primary validity condition that we call ex-ante exogeneity.

Assumption 1 (Ex-Ante Exogeneity) *Let* $f_{\mathbf{X}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g}_0,\boldsymbol{\varepsilon}) \equiv \sum_{\mathbf{g}\in\mathscr{G}} f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g},\mathbf{g}_0,\boldsymbol{\varepsilon})$ *be the resulting probability distribution from integrating* $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}$ *out with respect to the endogenous network* $\mathcal{G}_N$*. Thus, the probability distribution* $f_{\mathbf{X}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}$ *is such that* $\mathbb{E}[\mathbf{x}_{N;i}\varepsilon_{N;i}] = \mathbf{0}_K$ *and* $\mathbb{E}[(\mathbf{w}^p_{N,0;i}\mathbf{X}_N)^\top \varepsilon_{N;i}] = \mathbf{0}_K$, $\forall i \in \mathcal{I}_N$*, and any* $p \geq 1$*, where* $\mathbf{0}_K$ *is a* $K \times 1$ *vector of zeros. Moreover, we assume* $\mathbb{E}[\varepsilon_{N;i}] = 0$, $\forall i \in \mathcal{I}_N$*, where the expectation is taken with respect to the marginal distribution of* $\boldsymbol{\varepsilon}_N$*.*

This assumption imposes the restriction of zero correlation between the observed and unobserved characteristics of any individual $i \in \mathcal{I}_N$. For simplicity, we impose the noncorrelation assumption across the entire vector of observed characteristics $\mathbf{x}_{N;i}$; however, our results also hold under the non-correlation premise for a minimum of one variable within $\mathbf{x}_{N;i}$.[1] Furthermore, the constraint $\mathbb{E}[(\mathbf{w}^{p}_{N,0;i}\mathbf{X}_N)^\top \varepsilon_{N;i}] = \mathbf{0}_K$ implies that the unobserved characteristics of the individual $i \in \mathcal{I}_N$ remain uncorrelated with the observed characteristics of any other $j \in \mathcal{I}_N$, selected exogenously from the population of individuals. Analogous formulations of this assumption, employing randomized networks, have been previously used to substantiate the causal identification of peer effects (see, e.g., Sacerdote, 2001; Carrell, Fullerton, and West, 2009; Cai and Szeidl, 2018; Hasan and Koning, 2019).

Notice that Assumption 1 does not impose restrictions on the correlation between the vector $(\mathbf{w}_{N;i}\mathbf{X}_N)$ and the unobserved characteristics $\varepsilon_{N;i}$. Thus, we allow for correlation between the observed and unobserved characteristics of any two individuals engaging in endogenous linking formation within the network $\mathcal{G}_N$. The assumption 1 links the endogenous network formation process with the mechanism that induces correlation between the vectors of regressors and errors.

Under ex-ante exogeneity, a pair of agents (i, j), exogenously allocated in an initial network $\mathcal{G}_{N,0}$, establish connections in the network $\mathcal{G}_N$ based on observed and unobserved characteristics. This behavior creates correlation between $\mathbf{x}_{N,i}$ and $\varepsilon_{N,j}$ through observed and unobserved homophily. For example, in the context of Carrell, Sacerdote, and West (2013), students with similar abilities may be more inclined to form connections, and observable characteristics such as race or gender could correlate with both link formation and individuals' abilities. Moreover, as previously mentioned, the source of endogeneity can arise not only from unobserved factors, but also because an unspecified potential simulta-

[1] The ex-ante exogeneity assumption can also accommodate correlation between the network $G_{N,0}$ and individuals' characteristics. If the network $\mathcal{G}_{N,0}$ is determined by *observed* characteristics, one can control for them in the outcome equation and the ex-ante exogeneity assumption becomes a *conditional* ex-ante endogeneity assumption.

neous determination of the outcome and the network of interest. In the same context of educational achievement, for example, students may care about the educational achievement of potential connections, which can generate an issue of simultaneous determination in the outcome equation.

In addition to ex-ante exogeneity, the following assumption imposes a linear model of peer effects, which has been shown to have a structural interpretation as the best response function of a Bayesian game of social interactions; see, i.e., Blume et al. (2015).

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 $\widetilde{\mathbf{x}}_{N;i} = [1, \mathbf{x}_{N;i}^\top]^\top$, $w_{N;i,j}$ is the ijth position in the adjacency matrix representing the endogenous network, $\mathbf{W}_N$, and $\boldsymbol{\theta}_0 \equiv (\beta_0, \boldsymbol{\delta}_0^\top, \boldsymbol{\gamma}_0^\top)^\top$ belongs to the interior of the parameter space $\boldsymbol{\Theta} \subset \mathbb{R}^{2K+2}$, which is assumed to be compact.[2]

Assumption 2 prescribes a linear model for social effects, effectively imposing an exclusion restriction on the network $\mathcal{G}_{N,0}$. The model posits that only optimally formed connections by agents can induce peer effects. Specifically, we contend that randomly grouped individuals are unlikely to generate peer effects, but, after randomization, agents can form endogenous connections that influence their behavior, see, for example, Hasan and Koning (2019). Similarly, Carrell, Sacerdote, and West (2013) shows that groups designed to improve academic performance can produce negative effects due to the role of endogenous study partners and the formation of friendship bonds after the initial allocation

[2] Our Linear Model assumption includes both the case when the adjacency matrices are row normalized and when they are not. Our results hold for both the local aggregate and the local average models (Liu, Patacchini, and Zenou, 2014).

of optimally designed to improve academic performance. Furthermore, Hjort (2014) and Kato and Shu (2016) find that the effects of peers on worker productivity manifest only between individuals of the same ethnic divisions and social origins, which could serve as proxies for closer social interactions. This evidence supports the plausibility of the exclusion restriction assumption of the initial exogenous network in a peer effects model.

Here, we implicitly impose a network *exclusion* restriction in the sense that potential peer effects from the exogenous network $\mathcal{G}_{N,0}$ are precisely zero in (3.1). Alternatively, it is possible to relax the exact exclusion restriction in the exogenous network by incorporating prior information where the effect of $\mathcal{G}_{N,0}$ on $y_{N;i}$ is proximate but not precisely zero, following the approach of Conley, Hansen, and Rossi (2012).[3]

Finally, 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, the powers of the adjacency matrix $\mathbf{W}_N$ could be used to instrument $\mathbf{W}_N \mathbf{y}_N$ (Bramoullé, Djebbari, and Fortin, 2009; Manta et al., 2022). This approach is not possible here because the network $\mathcal{G}_N$ is allowed to be endogenous. However, note that from Assumptions 1, the powers of the adjacency matrix $\mathbf{W}_{N,0}$ are natural candidates to replace $\mathbf{W}_N$ in this approach.

We propose to use the random assignment embodied in $\mathbf{W}_{N,0}$ to identify the parameters

[3] We thank the Associate Editor for pointing this out.

of a linear model defined on the network space spanned by $\mathbf{W}_N$ in Assumption 2. Formally, we define $\mathbf{D}_N = [\mathbf{W}_N\mathbf{y}_N, \mathbf{W}_N\mathbf{X}_N, \widetilde{\mathbf{X}}_N]$ as the matrix of regressors in the matrix notation counterpart of (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 Assumption 1, 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 the characteristics of more indirect connections that are at distance $p > 2$. One important aspect differentiating the use of ex-ante exogeneity in Assumption 1 and the standard validity of an instrumental variable, is that ex-ante exogeneity allows us to form a vector of potentially infinite number of instruments. In particular, as long as $\mathbf{I}_N$, $\mathbf{W}_{N,0}$,$\mathbf{W}_{N,0}^2$,$\ldots$, $\mathbf{W}_{N,0}^{p-1}$, and $\mathbf{W}_{N,0}^p$ are linearly independent, where $\mathbf{I}_N$ is the identity matrix of order N, we can form up to $K \cdot p$ different instruments using the K ex-ante exogenous variables in $\mathbf{x}_{N,i}$.

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 characterizing the correlation between the regressors and the instruments is the population average of the expected values of the random matrices $(\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top)$, which can be written as $\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$, where $\mathbf{d}_{N;i}$ and $\mathbf{z}_{N;i}$ contain the ith rows of the matrices $\mathbf{D}_N$ and $\mathbf{Z}_N$, respectively. Importantly, we do not assume that $\mathbb{E}[\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$ are equal for all i and work directly with the population average of those expectations. The following assumption imposes a rank condition related to the strength of the correlation between $\mathbf{z}_{N;i}$ and $\mathbf{d}_{N;i}$.

Assumption 3 (Relevance) *The matrix* $\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top] < \infty$ *has full column rank.*

Assumption 3 guarantees that the population average of expectations $\mathbb{E}[\mathbf{z}_{N;i}\mathbf{d}_{N;i}^{\top}]$ is finite, and it provides the necessary conditions to guarantee that a unique parameter value solves the identifying moment conditions. Unlike standard IV estimation, the relevance condition here imposes restrictions on the more primitive network architectures. Appendix A provides such a set of primitive conditions. They are empirically testable restrictions on the regressors and the networks. Importantly, unlike previous results, identification in our model does not require linear independence between the matrices $\mathbf{I}_N$, $\mathbf{W}_N$ and $\mathbf{W}_N^2$. Instead, we show that for Assumption 3 to hold, there should exists any two *different* numbers $(r, s) \in \mathbb{N}_+ \times \mathbb{N}_+$ such that $\mathbf{I}_N$, $\mathbf{W}_N^r$, and $\mathbf{W}_N^s$ are linearly independent. Thus, our primitive conditions for the network of interest are weaker in the sense that any two powers of the adjacency matrix have to be linearly independent, and not only the first and second powers. Furthermore, we show that, as mentioned above, relevance requires that the matrices $\mathbf{I}_N$, $\mathbf{W}_{N,0}$,$\mathbf{W}_{N,0}^2$,..., $\mathbf{W}_{N,0}^{p-1}$, and $\mathbf{W}_{N,0}^p$ be linearly independent for some $p \in \mathbb{N}_+$ and $(\gamma_{0,k}\beta_0 + \delta_{0,k}) \neq 0$ for any $k \in \{1, \ldots, K\}$. The first condition resembles the restrictions on the network structure imposed in previous literature, with the difference that in our model the linear independence condition is imposed on the excluded network. The second condition on the structural parameters of the linear model has also been used in previous literature, see, i.e., Bramoullé, Djebbari, and Fortin (2009). It excludes the possibility that peer and contextual effects cancel each other out. The following Theorem formalizes the identification result.

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

Appendix B presents the proof of 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 exogenously imposed 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 emerge after an initial randomization. This method can be used to address research designs with randomized peers such as in Carrell, Sacerdote, and West (2013).

4 Estimation

We propose a 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 $\mathcal{G}_N$ and $\mathcal{G}_{N,0}$ are recorded.[4] Therefore, the random sample consists of observations $\left\{y_i, \mathbf{x}_i^\top, \{w_{i,j}\}_{j=1,j\neq i}^n, \{w_{0;i,j}\}_{j=1,j\neq i}^n\right\}_{i=1}^n$ from which it is possible to calculate the $n\times(2K+2)$ matrix of regressors $\mathbf{D}_n$ and the $n \times ((p-1)K + 2K + 1)$ 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 $J_N(\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta})]^\top \mathbf{A}_N \mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta})]$, where $\mathbf{A}_N$ is a constant full rank weighting matrix $\mathbf{A}_N$. The GMM estimator of $\boldsymbol{\theta}_0$ is defined as $\widehat{\boldsymbol{\theta}}_{\text{GMM}} = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} J_n(\boldsymbol{\theta})$, where $J_n(\boldsymbol{\theta}) \equiv [n^{-1}\sum_{i\in\mathcal{I}_n} \mathbf{z}_{n;i}(y_{n;i} - \mathbf{d}_{n;i}^\top\boldsymbol{\theta})]^\top \mathbf{A}_n [n^{-1}\sum_{i\in\mathcal{I}_n} \mathbf{z}_{n;i}(y_{n;i} - \mathbf{d}_{n;i}^\top\boldsymbol{\theta})]$, the $((p-1)K + 2K + 1) \times ((p-1)K + 2K + 1)$ 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

[4] In our empirical illustration we observe the complete graph for the schools' finite populations of students. However, as in Graham (2020), we consider the sampling process as a thought experiment that is useful in characterizing limiting distributions.

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

To allow for the possibility that the observed and unobserved characteristics of individuals are correlated in the joint distribution of the population $f_{\mathbf{X}_N, \mathcal{G}_N, \mathcal{G}_{N,0}, \boldsymbol{\varepsilon}_N}$, we use the concept of $\psi-$dependence in Definition 2.1 above. As mentioned there, we bound the correlation between nonlinear functions of random variables with the *dependence coefficients*, which are decreasing functions of the network distance. We rule out a direct dependence structure based on the exogenous network $\mathcal{G}_{N,0}$, i.e., dependence is only generated through the endogenous network $\mathcal{G}_N$. Intuitively, this is justified here because individuals endogenously form connections in $\mathcal{G}_N$ based on observed and unobserved characteristics, and therefore we would expect relatively high levels of dependence between individuals close to each other in the network space spanned by $\mathcal{G}_N$. For example, in our empirical illustration, we expect the observed and unobserved characteristics of students who study together or are indirectly connected by study partners to be more correlated than those of students who are not study partners and are not indirectly connected. Given that the network $\mathcal{G}_N$ generates the dependence structure, when we talk about distance in the network space, we refer to the distances in the network space $\mathcal{G}_N$ hereafter.

Let $\mathbf{r}_{N;i} \equiv \left[\mathbf{x}_{N;i}^\top, \varepsilon_{N;i}\right]^\top \in \mathbb{R}^{K+1}$ be the vector that encompasses the observed and unobserved characteristics of the individual i. By choosing appropriate values for the functions f and g in Definition 2.1, the $\psi-$dependence framework allows us to bound the dependence between observed and unobserved characteristics among any set of individuals. Specifically, we impose the following assumption on the conditional population distribution $f_{\mathbf{X}_N, \varepsilon_N | \mathcal{G}_N}$.

Assumption 4 (Weak Dependence) *Consider the set $\mathscr{G}$ of all possible realizations of*

$\mathcal{G}_N$. *$\forall \mathcal{G}_N \in \mathscr{G}$, and $\mathcal{N}$ denoting either $N \in \mathbb{N}_+$ or $n \in \mathbb{N}_+$, assume that the conditional distribution $f_{\mathbf{X}_N, \boldsymbol{\varepsilon}_N | \mathcal{G}_N}$ is such that:*

(i) $\{\mathbf{r}_{\mathcal{N};i}\}$ is ψ-dependent with dependence coefficient $\lambda_{\mathcal{N}}$;

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

(iii) and for each $\mathcal{N} \in \mathbb{N}_+$, $\max_{d \geq 1} \lambda_{\mathcal{N},d} < \infty$ and $\lim_{d \to \infty} \lambda_{\mathcal{N},d} = 0$.

We impose Assumption 4 for both the population with conditional distribution $f_{\mathbf{X}_N, \boldsymbol{\varepsilon}_N | \mathcal{G}_N}$ and for the triangular array $\mathbf{r}_{n;i}$, where $n \geq 1$. This array is formed by randomly sampling networks $\mathcal{G}_N$ and the observed and unobserved characteristics that define $\mathbf{r}_{n;i}$. The 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 $\mathrm{Lip}(f)$ and $\mathrm{Lip}(g)$ increase, the values of the functions f and g can be higher 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 to the sup-norm. Finally, condition (iii) requires that the dependence coefficients be finite for any value of d and that they dissipate to zero for a sufficiently large 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 $\widehat{\boldsymbol{\theta}}_{\mathrm{GMM}}$, we now impose bounds to covariances of the form $\mathrm{cov}(r_{n;i,q} r_{n;j,\ell}, r_{n;h,q'} r_{n;s,\ell'})$, 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 $\mathrm{cov}(\varepsilon_{n;i}\varepsilon_{n;j}, \varepsilon_{n;h}\varepsilon_{n;s})$ or $\mathrm{cov}(x_{n;i,q} x_{n;j,\ell}, \varepsilon_{n;h}\varepsilon_{n;s})$, for example.

Assumption 5 (Bound Covariances) *Define functions* $f_{q,\ell}$ *and* $g_{q',\ell'}$ *mapping* $\mathbb{R}^{(K+1)\times 2}$ *into* $\mathbb{R}$ *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 to 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 (2021), 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* $\mathcal{G}_n$. *We assume that, for any realizations of the networks* $\mathcal{G}_n$ *for all* n, *it follows that* $n^{-1}\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d}\longrightarrow 0$ *as* $n \longrightarrow \infty$.

Assumption 6 is similar to Assumption 3.2 in Kojevnikov, Marmer, and Song (2021), using the notation in our paper. The key distinction lies in our use of the unconditional version of ψ-dependence, which results in the dependence coefficients in the sequence $\lambda_{n,d}$ being not random variables. Furthermore, we apply Assumption 6 conditionally to any realizations of the networks $\mathcal{G}_n$, ensuring that $n^{-1}\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d}$ is nonrandom. As emphasized by Kojevnikov, Marmer, and Song (2021), this assumption is implied by restrictions on dependence coefficients ($\lambda_{n,d} \leq \theta_{n,d}^{1-4/p}$, for $p > 4$) and the number of distance-d connections on $\mathcal{G}_n$ ($D_n(d) \leq 4c_n(d, m, k)$, where $c_n(d, m, k)$ is defined in (4.2)). The following assumption imposes the existence of moments for products of ψ-dependent random variables.

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

The previous assumptions are sufficient to guarantee that a Law of Large Numbers applies to products of ψ-dependent random variables. To show asymptotic normality, we again use the Central Limit Theorem result in Kojevnikov, Marmer, and Song (2021). 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 (2021), we define a measure of 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 $\bar{D}_n(d, m, k)^- = n^{-1}\sum_{i\in\mathcal{I}_n}\max_{j\in\mathcal{P}_n(i,d)}|\mathcal{P}_n^-(i, m)\setminus\mathcal{P}_n^-(j, d-1)|^k$, where $\mathcal{P}_n^-(j, d-1) = \{\emptyset\}$ 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 $\sigma_{n,q}^2 \equiv \text{var}(S_n)$, 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 network realizations $\mathcal{G}_n \in \mathscr{G}$, (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) , $\frac{n}{\sigma_{n,q}^{2+k}} \sum_{d\geq 0} c_n(d, m_n, k)\, \lambda_{n,d}^{1-\frac{2+k}{p}} \longrightarrow 0$ as $n \to \infty$, (iii) $\frac{n^2 \lambda_{n,m_n}^{1-(1/p)}}{\sigma_{n,q}} \longrightarrow 0$ as $n \to \infty$.*

These conditions impose a convergence rate of the dependence coefficients $\lambda_{n,d}$ that is related to the density of the network. 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. Formally, let $\mathbf{\Omega}_n = \mathrm{var}(\mathbf{Z}_n^\top \boldsymbol{\varepsilon}_n)$ be a variance defined in the set of individuals sampled. It converges (see Lemma C.3 in the supplemental materials) to the finite population variance,

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

where $\mathbf{\Gamma}_N(d) = \sum_{i\in\mathcal{I}_N} \sum_{j\in\mathcal{P}_N(i,d)} \mathbb{E}[\mathbf{z}_{N;i}\varepsilon_{N;i}\varepsilon_{N;j}\mathbf{z}_{N;j}^\top]$ 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 1–8 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 $\mathbf{\Sigma}_N \equiv (\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]^\top \mathbf{A}_N \mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top])^{-1} \times (\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]^\top \mathbf{A}_N \mathbf{\Omega}_N \mathbf{A}_N \times \mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top])(\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]^\top \mathbf{A}_N$ $\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top])^{-1}$, and when $\mathbf{A}_N = \mathbf{\Omega}_N^{-1}$, then*

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

We present the proof for Theorem 2 in Appendix B. The following subsections discuss the relationship between the precision of the estimator in (4.1) and the levels of sparsity of the endogenous population network of interest, $\mathcal{G}_N$.

4.1 Precision and Sparsity

Some individuals in the population may not have any connections at distance p, and may affect the identifying moment condition in Theorem 1. To consider the effect of changes in identifying information on the asymptotic variance-covariance matrix in (4.4), define $\eta^p_{N,0;i}$ to be a random variable equal to one if individual i has at least one connection at distance p and zero otherwise. Let $\kappa^p_{N,0;i} = \mathbb{E}[\eta^p_{N,0;i}]$ be the unconditional probability that the individual i has at least one connection at distance p. Define $\boldsymbol{H}_{N,0;i} \equiv \mathrm{diag}(\eta^p_{N,0;i}, \ldots, \eta_{N,0;i}, 1, \ldots, 1)$ to be a $[(p+1)K+1] \times [(p+1)K+1]$-matrix where the first K elements contain the random variables that determine whether or not the 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 submatrix, 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 $\boldsymbol{K}_{N,0;i} \equiv \mathrm{diag}(\kappa^p_{N,0;i}, \ldots, \kappa_{N,0;i}, 1, \ldots, 1)$ to be $\mathbb{E}[\boldsymbol{H}_{N,0;i}]$.

Note that when $\eta^p_{N,0;i} = 0$, the first K elements of $\mathbf{z}_{N;i}$ equal zero. Similarly, when $\eta^{p-1}_{N,0;i} = 0$, the second K elements of $\mathbf{z}_{N;i}$ equal zero. The same argument repeats until the K elements associated with $\mathbf{W}_{N,0}\mathbf{X}_N$, where if $\eta_{N;i} = 0$ (individual i is isolated), the elements of $\mathbf{z}_{N;i}$ associated with the component $(\mathbf{w}_{i;N,0}\mathbf{X}_N)^\top$ are equal to zero. Therefore, by the law of total expectation, $\mathbb{E}[\sum_{i\in\mathcal{I}_N}\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$ can be written as $\boldsymbol{K}_{N;0,i}\mathbb{E}[\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top \mid \boldsymbol{H}^*_{N;0,i}\neq\boldsymbol{O}_{pK}]$, 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 matrix. Therefore, (4.4) depends on the population probabilities that an individual provides identification information. For low values of these probabilities, the upper right submatrix of $\boldsymbol{K}_{N;0,i}$ approaches the zero matrix, and the variance-covariance matrix could grow arbitrarily large. In the extreme case of nonidentification, (4.4) diverges to infinity. Theorem 2 exposes a relationship between the precision of network parameters and the sparsity of the network.

4.2 Efficient Weight Matrix Estimation

To construct the efficient version of the proposed GMM estimator, we need a consistent estimator of $\boldsymbol{\Omega}_N$. Here we use Kojevnikov, Marmer, and Song's (2021) 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 (2021) 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 in our empirical study with $C = 1.8$ according to their suggestion. The proposed variance-covariance matrix estimator is then given by

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

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 to 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 $\mathbf{A}_n = \widetilde{\boldsymbol{\Omega}}_n^{-1}$ in (4.1), which we call $\widehat{\boldsymbol{\theta}}^{\star}_{\text{GMM}}$.

4.3 Standard Error Calculation

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

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

where $\widehat{\boldsymbol{\Omega}}_n^{\star}$ 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 $\left\{y_{n;i}, x_{n;i}, \{w_{n;i,j}\}_{j=1,j\neq i}^{n}, \{w_{n,0;i,j}\}_{j=1,j\neq i}^{n}\right\}_{i=1}^{n}$; 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. We set the true vector of the parameters at $\boldsymbol{\theta}_0 = [\beta_0, \delta_0, \gamma_0]^{\top} = [0.7, 1, 1]^{\top}$. The other data components are constructed using the following rules.

Design 1: Unobserved Characteristics with Homophily

In this design, the outcome variable of the individual i' $y_{n,i}$ and the connections $\{w_{n;i,j}\}_{j=1,j\neq i}^{n}$ are jointly determined by a common idiosyncratic homophily-related unobserved variable $\varepsilon^*_{n,1;i}$. 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 $\boldsymbol{\varepsilon}^*_{n,1} = [\varepsilon^*_{n,1;1}, \ldots, \varepsilon^*_{n,1;n}]^\top$ from a multivariate standard normal distribution.[5] The elements of the endogenous adjacency matrix $\mathbf{W}_n = [w_{n;i,j}]$ are then calculated as

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

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 / maintain connections with those also endowed with large values of $\varepsilon_{n,1}$ and sever them with those with low values of this unusual unobserved characteristic. The $n \times 1$ vector of outcomes, $\mathbf{y}_n$, is then constructed from (3.1) by setting $\boldsymbol{\varepsilon}_n = m \times \boldsymbol{\varepsilon}_{n,1} + \boldsymbol{\varepsilon}_{n,2}$, where $m \in \{1, 3\}$, $\boldsymbol{\varepsilon}_{\boldsymbol{n},\mathbf{2}}$ is drawn from a multivariate standard normal distribution. The elements of $\boldsymbol{\varepsilon}_{n,1}$ are defined as

[5]The density of a network is the ratio between the total numbers of actual ties and of potential ties.

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

Design 2: Misclassified Links

This is a modified version of Lewbel, Qu, and Tang's (2023) 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 (2023), nonexisting links are never misclassified in $\mathbf{W}_n$, but misclassification of these nonexisting links is allowed 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; that is, the $n \times 1$ vector $\mathbf{y}$ is constructed following Equation (3.1), where $\varepsilon_n = \boldsymbol{\varepsilon}_{n,1} + \boldsymbol{\varepsilon}_{n,2}$, $\varepsilon_{n,1;i} = 1/n \sum_{j=1}^{n} w^*_{n,0;i,j}e_{n,1;i,j}$, and $\boldsymbol{\varepsilon}_{n,2}$ is drawn from a multivariate standard normal distribution independently of everything else.

Results

Figures 1 and 2 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 $\mathbf{\Omega}_n$. 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 estimated coefficients in the first step to calculate the network HAC variance estimator in (4.5), where we choose $K(\cdot)$ to be the Parzen kernel, we set the bandwidth $D_n = 1.8 \times [\log(\text{ average degree } \vee (1+0.05))]^{-1} \times \log n$ as in Kojevnikov, Marmer, and Song (2021).

Each panel in Figure 1 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 Monte Carlo replications of OLS (black), G2SLS (dark gray) and the two proposed GMM estimators for $p = 2$ (gray) and $p = 3$ (light gray) of the social effects. Peer effects (β), contextual effects (δ), and direct effects (γ) in (3.1) are shown with whiskers that show the empirical Monte Carlo quantiles 5% and 95%. Across the board for all parameters, designs, and sample sizes, the proposed both GMM estimators perform 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 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 for the case of misclassified links (Design 2).

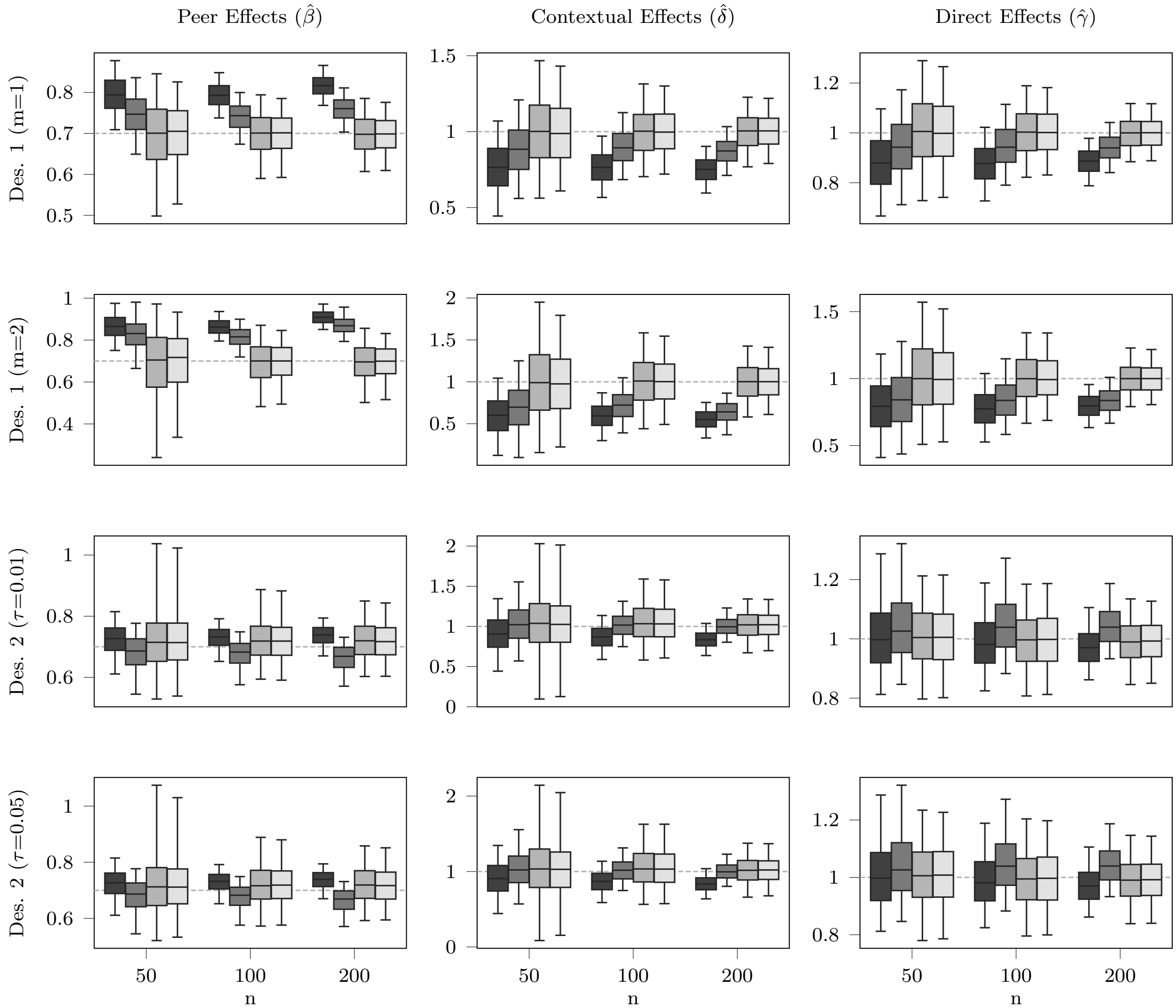
Similarly, Figure 2 displays the corresponding Q-Q plots for the GMM based on the standardized version of the Monte Carlo replications of the GMM estimator of the same social effects for sample sizes $n = 50$ (light gray), $n = 100$ (gray), and $n = 200$ (black). The blue dashed line depicts the 45-degree 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 the sample size increases, the approximation improves for all parameters and designs.

6 Empirical Illustration

We now provide an empirical illustration of the proposed methods in the framework of estimating peer effects on academic performance among high school students. The data set was collected between March and May 2011 as part of the Hong Kong Secondary Education Survey in Hong Kong (SESHK). The survey was carried out in the second semester before the final exams and involved three secondary schools with 868 students participating. The sample includes students in the seventh grade of all three schools and students in the eighth and ninth grades of one school ($\mathrm{g} \in \{7, 8, 9\}$). Each grade within a school is made up of five different sections ($\mathrm{cl} \in \{1, \ldots, 5\}$). Table 2 in the supplemental materials shows the summary statistics for the variables we use. Additional details about these variables can be found in Section D.1 of the supplemental material.

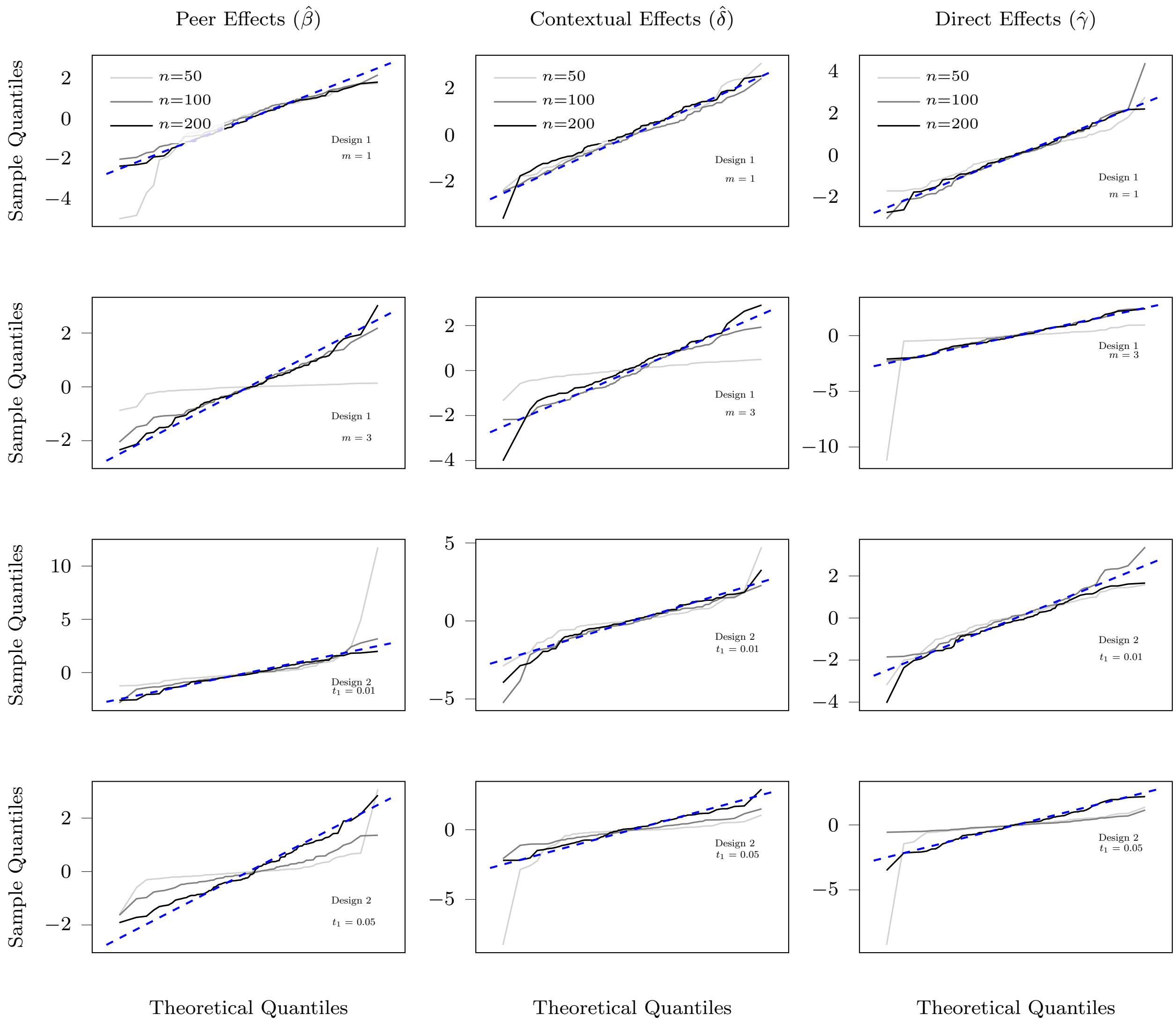
In the survey, students were asked to report lists of up to ten peers within their grade with whom they discussed schoolwork issues and who were sitting nearby in class during the first semester. We used this information to build the *study partner* and *seatmate* networks respectively. Classroom seat assignments undergo multiple changes throughout the semester, and the class teacher determines these adjustments. Unlike study partners, the seatmate network is based on proximity and is enforced by the school. Consequently,

Figure 1: Box Plots of the OLS, G2SLS, and GMM Estimators of Social Effects



Note: Box plots depict the Monte Carlo performance of 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 2: 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.

the seatmate network emerges as a compelling choice to serve as the instrumental network, $\mathbf{W}_{n,0}$, in our analysis, that is, we treat it ex-ante exogenous in this analysis. On the contrary, students have the autonomy to select their study partners, and this choice can be influenced by unobservable characteristics that also affect exam performance, rendering the study partner network $\mathbf{W}_n$, potentially endogenous. Table 3 in Section D.1 of the Supplementary Material reports the summary statistics of the network among all students by school.

We fit the Linear-in-Means model for social effects in (1.1). Specifically, for the individual i, we fit the following version of the model.

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

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; that is, $\mathbb{I}\{i \in \mathrm{s}\times\mathrm{g}\times\mathrm{cl}\} = 1$; $\texttt{personality}_{i,\mathrm{s}\times\mathrm{g}\times\mathrm{cl}}$ includes the 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 sibling help, parents help, whether they commute to school by car or taxi; whether they play music; and whether the student is a male. We also include interaction terms between the male indicator and the test scores. 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 before 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 (isolated). We set $\delta_{\texttt{personality}} = -\gamma_{\texttt{personality}}$ for all estimation routines to avoid potential collinearity problems. Behaviorally, this restriction implies that only deviations from the students' own personality characteristics from the average of their peers affect the students' tests scores; see, for example, Liu, Patacchini, and Zenou (2014).

The model (6.1) is estimated using simple Ordinary Least Squares (OLS) with standard errors clustered at the $\mathrm{s} \times \mathrm{g} \times \mathrm{cl}$ 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, and our proposed efficient GMM estimator with $p = 5$, $C = 1.8$ with the Tukey-Hanning kernel.

Table 1: Estimations results

Variables	OLS	G2SLS	GMM
Peer effect			
ln(Math Test)	0.2455***	0.4534***	0.6065***
	(0.0864)	(0.1027)	(0.1755)
Contextual effects			
Male	-0.0232	-0.0153	-0.2455***
	(0.0311)	(0.0276)	(0.0873)
ln(Height)	0.0794	0.0651	2.1682***
	(0.2229)	(0.1910)	(0.7400)
ln(Weight)	0.1012	0.0399	0.0329
	(0.0642)	(0.0600)	(0.1942)
Siblings Help	0.0263	0.0305	0.0294
	(0.0261)	(0.0217)	(0.0491)
Parents Help	0.0345	0.0064	-0.0106
	(0.0287)	(0.0189)	(0.0517)
Commute by Car/Taxi	0.0277	0.0250	0.1255**
	(0.0266)	(0.0224)	(0.0636)
Music	-0.0212	-0.0090	0.1187**
	(0.0166)	(0.0160)	(0.0480)
†			
ln(Cognitive)	0.0549	0.0734**	0.1086***
	(0.0357)	(0.0369)	(0.0286)
ln(Agreeableness)	-0.1145***	-0.1275***	-0.0712
	(0.0300)	(0.0356)	(0.0457)
ln(Conscientiousness)	0.0445	0.0537	0.0799**
	(0.0371)	(0.0396)	(0.0406)
ln(Extraversion)	-0.0998**	-0.0982**	-0.1164***
	(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)
Direct effects			
Male	-0.7922*	-0.8352**	-0.1648
	(0.4596)	(0.4224)	(0.5417)
ln(Height)	-0.3366**	-0.2635**	-0.9904***
	(0.1342)	(0.1210)	(0.2248)
ln(Weight)	-0.0088	-0.0106	-0.0494
	(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	0.0167
	(0.0159)	(0.0142)	(0.0139)
Commute by Car/Taxi	-0.0079	-0.0162	-0.0251*
	(0.0117)	(0.0111)	(0.0136)
Degree	0.0292***	0.0264***	0.0248***
	(0.0041)	(0.0039)	(0.0034)
Isolate Students	2.2787*	0.4720*	0.1425
	(1.2954)	(0.2813)	(0.2935)
n	868	868	868
Adjusted R^2	0.3372	0.3936	0.2675
RMSE	0.1854	0.1716	0.2002

Note: (i) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; (ii) Standard errors are in parentheses. (iii) † These regressors are measured as the deviation of students' personality from their peers' average.

Tables 1 show the results for a subset of all regressors included in (6.1) (Table 4 in the Supplement contains the results for the remaining set of regressors). 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 example, cognitive ability and noncognitive characteristics, such as conscientiousness, have a positive effect on math grade achievement (Heckman and Rubinstein, 2001). There is no gender gap between the grades of men and women. 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.

Our results suggest that estimators that do not control network endogeneity tend to underestimate peer effects. The observation of a negative bias in estimators that do not account for endogeneity aligns with findings in the literature, which highlight a similar negative bias in standard maximum likelihood estimators of spatial autoregressive models that overlook the reflection issue (Mizruchi and Neuman, 2008; Neuman and Mizruchi, 2010). This trend of negative bias, as observed between OLS and G2SLS, also extends to our GMM estimator, which incorporates control for network endogeneity.

All results are qualitatively robust to different choices of p and D_n, see, i.e. Section D.2 in the Supplementary Material. We perform empirical tests to validate the exclusion restriction in Assumption 2 and the relevance condition in Assumption 3. For the exclusion restriction, we perform a Least Squares (LS) regression that includes all variables in (6.1), but also includes our proposed instruments. The idea is to measure to what extent the instruments are good predictors of our variable of interest. We interpret the lack of predictability as evidence in favor of our exclusion restriction assumption. Table 11 in Appendix D presents the results. We find that all our instruments are not predictive of the outcome equation. For the relevance assumption, we perform a series of LS regressions in which the different outcomes are all different endogenous variables. We include all our instruments as regressors and calculate the F-statistic for each of the regressions. Consistent with the relevance condition, we reject the hypothesis that our instruments are jointly significant at the 1% significance level for all our endogenous variables (see Section D.3 in the Supplement).

7 Conclusion

This research adds to the literature on the identification and estimation of social effects with observational network data that often contain endogenous or mismeasured connections. Unlike current approaches, such as those in Johnsson and Moon (2019) and Auerbach (2022), our method does not require the specification and estimation of a model characterizing 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.

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; that is, ψ-dependence. We show that the resulting estimator can be easily implemented utilizing standard linear GMM estimation routines in popular software like Python, R, or Stata, see, e.g., Estrada et al. (2023). 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.

Empirically, an important aspect of the proposed methodology is that it recognizes that exogenously imposed connections on individuals 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 (with whom you study, for example) and what other type simply influences these connections, but are otherwise exogenous to the model (for instance, to whom you are randomly assigned to share a physical space). Our empirical and Monte Carlo 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 high-schoolers from study partners in Hong Kong. These results are in line with previous literature in that they show the existence of strong positive network effects, but suggest 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, the idea of using the initial random assignment of network connections as an instrument differs from using other sources of exogenous variation in two critical aspects. First, under restrictions on exogenous network density, the shape of the network structure together with K regressors can be used to form at least the $K + 1$ instruments required to identify endogenous peer effects and contextual effects in the linear-in-means model fitted above. Without the exogenous variation of the randomized network, a researcher would have a difficult task finding $K + 1$ different instrumental variables. Second, the standard relevance IV assumption imposes restrictions on the shape of the exogenous network and the process determining the formation of the network of interest. In particular, we have shown that relevance requires that a number of p powers of the adjacency matrix of the exogenous network to be linearly independent, and connections in the exogenous network need to have an effect on the decision of forming a connection on the endogenous network of interest. The need for linear independence imposes restrictions on the potential randomization of links, which researchers have to follow when designing their experiments. For example, researchers cannot randomly select individuals into groups of the same size if they want to estimate the effects of the network using our method. These are among important empirical considerations left for future research.

References

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

Athey, Susan and Guido W Imbens. 2017. "The Econometrics of Randomized Experiments." In *Handbook of economic field experiments*, vol. 1. Elsevier, 73–140.

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.

Cai, Jing and Adam Szeidl. 2018. "Interfirm Relationships and Business Performance." *The Quarterly Journal of Economics* 133 (3):1229–1282.

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., 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.

Conley, Timothy G, Christian B Hansen, and Peter E Rossi. 2012. "Plausibly Exogenous." *Review of Economics and Statistics* 94 (1):260–272.

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

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, Pablo, Juan Estrada, Kim P. Huynh, David T. Jacho-Chávez, and Leonardo Sánchez-Aragón. 2023. "`netivreg`: Instrumental Variable Estimation in the Linear-in-Means Model." Unpublished Manuscript.

Gargiulo, Martin and Mario Benassi. 2000. "Trapped in Your Own Net? Network Cohesion, Structural Holes, and the Adaptation of Social Capital." *Organization Science* 11 (2):183–196.

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. 2020. "Network Data." In *Handbook of Econometrics*, vol. 7. Elsevier, 111–218.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6):1360–1380.

Hasan, Sharique and Rembrand Koning. 2019. "Prior Ties and the Limits of Peer Effects on Startup Team Performance." *Strategic Management Journal* 40 (9):1394–1416.

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.

Hjort, Jonas. 2014. "Ethnic Divisions and Production in Firms." *The Quarterly Journal of Economics* 129 (4):1899–1946.

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.

Kato, Takao and Pian Shu. 2016. "Competition and Social Identity in the Workplace: Evidence from a Chinese Textile Firm." *Journal of Economic Behavior & Organization* 131:37–50.

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.

Kim, Tai-Young, Hongseok Oh, and Anand Swaminathan. 2006. "Framing interorganizational network change: A network inertia perspective." *Academy of Management Review* 31 (3):704–720.

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. 2021. "Limit Theorems for Network Dependent Random Variables." *Journal of Econometrics* 222 (2):882–908.

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. 2023. "Social Networks with Unobserved Links." *Journal of Political Economy* 131 (4):898–946.

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.

Mas, Alexandre and Enrico Moretti. 2009. "Peers at Work." *American Economic Review* 99 (1):112–145.

Mizruchi, Mark S and Eric J Neuman. 2008. "The Effect of Density on the Level of Bias in the Network Autocorrelation Model." *Social Networks* 30 (3):190–200.

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.

Neuman, Eric J. and Mark S. Mizruchi. 2010. "Structure and Bias in the Network Autocorrelation Model." *Social Networks* 32 (4):290–300.

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.

Qu, Xi, Lung-fei Lee, and Chao Yang. 2021. "Estimation of a SAR model with endogenous spatial weights constructed by bilateral variables." *Journal of Econometrics* 221 (1):180–197.

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.

Estimating Social Effects with Randomized and Observational Network Data – 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

February 23, 2024

Appendix A Primitive Conditions for Relevance

This section provides a set of primitive conditions that imply the relevance condition in Assumption 3. This is formally proven in Proposition 1 below.

Assumption A.9 $\mathbf{X}$ *is full column rank for any realization of the matrix of regressors* $\mathbf{X} \in \mathscr{X}$ *with positive probability in* $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X}, \mathbf{g}, \mathbf{g}_0, \boldsymbol{\varepsilon})$.

Assumption A.10 *For any two regressors* k *and* ℓ *and a number* $p \geq 2$*, the expectation* $\mathbb{E}[\mathbf{w}^p_{N,0,i}\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{x}_{N,\ell}]$ *exist for all* i.

[‖]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.

Assumption A.11 *There exists two* different *numbers* $(r,s) \in \mathbb{N}_+ \times \mathbb{N}_+$ *such that* $\mathbf{I}_N$, $\mathbf{W}^r$ *and* $\mathbf{W}^s$ *are linearly independent for any realization of the network of interest* $\mathbf{g} \in \mathscr{G}$ *with a positive probability in* $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g},\mathbf{g}_0,\boldsymbol{\varepsilon})$. *Furthermore,* $(\gamma_{0,k}\beta_0 + \delta_{0,k}) \neq 0$ *for all* $k \in \{1,\ldots,K\}$.

Assumption A.12 *The equation* $x_{i,k} \neq \mathbf{w}_{N,i}^m \mathbf{x}_k$ *holds for some number* $m \in \mathbb{N}_+$, *for all individuals* i, *any regressor* k, *and any realization of the matrix of regressors* $\mathbf{X} \in \mathscr{X}$ *and the network of interest* $\mathbf{g} \in \mathscr{G}$ *with positive probability in* $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g},\mathbf{g}_0,\boldsymbol{\varepsilon})$.

Assumption A.13 *There exist a number* $p \geq 2$ *such that* $\mathbf{I}_N$, $\mathbf{W}_0$, $\mathbf{W}_0^2$, $\ldots$, $\mathbf{W}_0^p$ *are linearly independent for any realization of the exogenous network* $\mathbf{g}_0 \in \mathscr{G}_0$ *with positive probability in* $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g},\mathbf{g}_0,\boldsymbol{\varepsilon})$.

Assumption A.14 *The joint probability distribution* $\Pr(\mathcal{G}_N = \mathbf{g}, \mathcal{G}_{N,0} = \mathbf{g}_0)$ *is such that* $\mathbb{E}[\mathbf{w}_{0,N,i}^p \mathbf{x}_\ell \mathbf{w}_i \mathbf{x}_{N,k}] \neq 0$ *for at least* $p = 2$, *and any realizations of* $\mathbf{x}_\ell$ *and* $\mathbf{x}_k$ *with positive probability in* $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X},\mathbf{g},\mathbf{g}_0,\boldsymbol{\varepsilon})$ *and all* i.

Assumption A.15 $\mathbb{E}[\sum_{r=0}^{\infty} \beta_0^r \mathbf{w}_{N,i} \mathbf{W}_N^r \varepsilon_N \mid \mathbf{W}_{0,N}, \mathbf{X}_{N,k}] = 0$ *for any* $k \in \{1,\ldots,K\}$.

Proposition 1 *Let Assumptions A.9, A.10, A.11, A.13, A.12, A.14 and A.15 hold. It follows that the matrix* $\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$ *has full column rank.*

Proof. Let $\mathbf{w}_{N,i}$ and $\mathbf{w}_{N,0,i}$ be the ith row of the adjacency matrices representing the network of interest and the exogenous network. Similarly let $\mathbf{x}_{N,k}$ be the $N \times 1$ vector of the regressor k for all N individuals in the population, and $x_{N,k,i}$ be the value of the regressor k for individual i. Therefore, it follows that for individual i, $\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top$ equals

$$\begin{bmatrix}
\mathbf{w}_{N,0,i}^p \mathbf{x}_{N,1}\mathbf{w}_{N,i}\mathbf{y}_N & \mathbf{w}_{N,0,i}^p \mathbf{x}_{N,1}\mathbf{w}_{N,i}\mathbf{x}_{N,1} & \cdots & \mathbf{w}_{N,0,i}^p \mathbf{x}_{N,1}\mathbf{w}_{N,i}\mathbf{x}_{N,k} & \cdots & \mathbf{w}_{N,0,i}^p \mathbf{x}_{N,1} x_{N,K,i} \\
\mathbf{w}_{N,0,i}^p \mathbf{x}_{N,2}\mathbf{w}_{N,i}\mathbf{y}_N & \mathbf{w}_{N,0,i}^p \mathbf{x}_{N,2}\mathbf{w}_{N,i}\mathbf{x}_{N,1} & \cdots & \mathbf{w}_{N,0,i}^p \mathbf{x}_{N,2}\mathbf{w}_{N,i}\mathbf{x}_{N,K} & \cdots & \mathbf{w}_{N,0,i}^p \mathbf{x}_{N,2} x_{N,K,i} \\
\vdots & \vdots & & \vdots & & \vdots \\
\mathbf{w}_{N,0,i}^{p-1} \mathbf{x}_{N,K}\mathbf{w}_{N,i}\mathbf{y}_N & \mathbf{w}_{N,0,i}^{p-1} \mathbf{x}_{N,K}\mathbf{w}_{N,i}\mathbf{x}_{N,1} & \cdots & \mathbf{w}_{N,0,i}^{p-1} \mathbf{x}_{N,K}\mathbf{w}_{N,i}\mathbf{x}_{N,K} & \cdots & \mathbf{w}_{N,0,i}^{p-1} \mathbf{x}_{N,K} x_{N,K,i} \\
\vdots & \vdots & & \vdots & & \vdots \\
\mathbf{w}_{N,0,i} \mathbf{x}_{N,K}\mathbf{w}_{N,i}\mathbf{y}_N & \mathbf{w}_{N,0,i} \mathbf{x}_{N,K}\mathbf{w}_{N,i}\mathbf{x}_{N,1} & \cdots & \mathbf{w}_{N,0,i} \mathbf{x}_{N,K}\mathbf{w}_{N,i}\mathbf{x}_{N,K} & \cdots & \mathbf{w}_{N,0,i} \mathbf{x}_{N,K} x_{N,K,i} \\
\vdots & \vdots & & \vdots & & \vdots \\
x_{N,K,i}\mathbf{w}_{N,i}\mathbf{y}_N & x_{N,K,i}\mathbf{w}_{N,i}\mathbf{x}_{N,1} & \cdots & x_{N,K,i}\mathbf{w}_{N,i}\mathbf{x}_{N,K} & \cdots & x_{N,K,i}^2
\end{bmatrix}. \tag{A-1}$$

If, in expectation, the columns of the matrix $\mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top$ are linearly independent for all i, it follows that $\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top]$ has full column rank. We check the linear independence across columns for a generic row from the matrix in (A-1). Firs, note that for all the components involving the outcome variable $\mathbf{y}_N$, the linearity assumption in 2 implies that, for $K = 1$,

$$\mathbf{W}_N\mathbf{y}_N = \gamma_0\mathbf{W}_N\mathbf{x}_N + \pi_0\sum_{r=0}^{\infty}\beta_0^r\mathbf{W}_N^{r+2}\mathbf{x}_N + \sum_{r=0}^{\infty}\beta_0^r\mathbf{W}_N^{r+1}\varepsilon_N,$$

where $\pi_0 = (\gamma_0\beta_0 + \delta_0)$. Thus, all the elements in the first column involve infinite powers of the endogenous adjacency matrix $\mathbf{W}_N$. In particular, it follows that for an arbitrary individual i and any number of regressors K,

$$\begin{aligned}\mathbf{w}_{N,i}\mathbf{y}_N =& \gamma_{0,1}\mathbf{w}_{N,i}\mathbf{x}_{N,1} + \cdots + \gamma_{0,K}\mathbf{w}_{N,i}\mathbf{x}_{N,K} + \pi_{0,1}\sum_{r=0}^{\infty}\beta_0^r\mathbf{w}_{N,i}\mathbf{W}_N^{r+1}\mathbf{x}_{N,1} + \dots \\ &+ \pi_{0,K}\sum_{r=0}^{\infty}\beta_0^r\mathbf{w}_{N,i}\mathbf{W}_N^{r+1}\mathbf{x}_{N,K} + e_i,\end{aligned} \tag{A-2}$$

where $e_i = \sum_{r=0}^{\infty}\beta_0^r\mathbf{w}_{N,i}\mathbf{W}_N^r\varepsilon_N$ and $\pi_{0,k} = (\gamma_{0,k}\beta_0 + \delta_{0,k})$ for some $k \in \{1, \dots, K\}$. Choose an arbitrary row k from the first K rows in equation (A-1). Taking expectations with respect to the joint distribution $f_{\mathbf{X}_N,\mathcal{G}_N,\mathcal{G}_{N,0},\boldsymbol{\varepsilon}_N}(\mathbf{X}, \mathbf{g}, \mathbf{g}_0, \boldsymbol{\varepsilon})$, the expected value for row k is given by the vector

$$[\mathbb{E}[\mathbf{w}_{N,0,i}^p\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{y}_N], \dots, \mathbb{E}[\mathbf{w}_{N,0,i}^p\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{x}_{N,K}], \dots, \mathbb{E}[\mathbf{w}_{N,0,i}^p\mathbf{x}_{N,k}x_{N,K,i}]].$$

By definition of conditional expectations, and considering the discrete nature of the random networks, for any two aggressors k and ℓ, we can write

$$\begin{aligned}\mathbb{E}[\mathbf{w}_{N,0,i}^p\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{x}_{N,\ell}] = \sum_{\mathbf{w}_0:\mathbf{g}_0\in\mathscr{G}_0}\int_{\mathbf{x}_k:\mathbf{X}\in\mathscr{X}}&\mathbf{w}_0^p\mathbf{x}_k\mathbb{E}[\mathbf{w}_{N,i}\mathbf{x}_{N,\ell} \mid \mathbf{w}_{N,0,i} = \mathbf{w}_0, \mathbf{x}_{N,k} = \mathbf{x}_k] \\ &f_{\mathbf{x}_{N,k}}(\mathbf{x}_k)\Pr(\mathcal{G}_{N,0} = \mathbf{g}_0)d\mathbf{x}_k,\end{aligned} \tag{A-3}$$

where $f_{\mathbf{x}_{N,k}}(\mathbf{x}_k)\Pr(\mathcal{G}_{N,0} = \mathbf{g}_0)$ represents the product of the marginal distributions of the kth regressors and the exogenous network. We can represent the distribution of the regressors and the exogenous network by the products of the marginals because of the independence guaranteed by the properties of randomization. From Assumption A.10, $\mathbb{E}[\mathbf{w}_{N,0,i}^p\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{x}_{N,\ell}]$ exists, which implies that the conditional expectations define in (A-3) also exist. Choose arbitrary values of $\mathbf{x}_k$ and $\mathbf{g}_0$ such that $\mathbb{E}[\mathbf{w}_{N,i}\mathbf{x}_{N,\ell} \mid \mathbf{w}_{0,i} = \mathbf{w}_0, \mathbf{x}_k = \mathbf{x}_k] \neq 0$, and that happen with positive probability in $f_{\mathbf{x}_{N,k}}(\mathbf{x}_k)$ and $\Pr(\mathcal{G}_{N,0} = \mathbf{g}_0)$. We can collect all the values related with the same arbitrary regressor $\mathbf{x}_\ell$ from the vector of expectations $[\mathbb{E}[\mathbf{w}^p_{N,0,i}\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{y}_N], \ldots, \mathbb{E}[\mathbf{w}^p_{N,0,i}\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{x}_{N,K}], \ldots, \mathbb{E}[\mathbf{w}^p_{N,0,i}\mathbf{x}_{N,k}x_{N,K,i}]]$, which, after replacing $\mathbf{w}_{N,i}\mathbf{y}_N$ with the expression in (A-2) and considering Assumption A.15, is given by

$$\begin{aligned}&[\gamma_{0,\ell}\mathbf{w}^p_0\mathbf{x}_k\mathbb{E}[\mathbf{w}_{N,i}\mathbf{x}_{N,\ell} \mid \mathbf{w}_0, \mathbf{x}_k] + \pi_{0,\ell}\mathbf{w}^p_0\mathbf{x}_k\sum_{r=0}^{\infty}\beta^r\mathbb{E}[\mathbf{w}_{N,i}\mathbf{W}^{r+1}_N\mathbf{x}_{N,\ell} \mid \mathbf{w}_0, \mathbf{x}_k], \\ &\mathbf{w}^p_0\mathbf{x}_k\mathbb{E}[\mathbf{w}_{N,i}\mathbf{x}_{N,\ell} \mid \mathbf{w}_0, \mathbf{x}_k], \mathbf{w}^p_0\mathbf{x}_k\mathbb{E}[x_{N,i,\ell} \mid \mathbf{w}_0, \mathbf{x}_k]].\end{aligned} \tag{A-4}$$

The three components of the vector in (A-4) are linearly dependent if and only if there exists three constants a, b and c different from zero such that

$$\begin{aligned}&\mathbb{E}[(a\gamma_{0,\ell} + b)\mathbf{w}_{N,i}\mathbf{x}_{N,\ell} + cx_{N,i,\ell} + a\pi_{0,\ell}\beta\mathbf{w}_{N,i}\mathbf{W}_N\mathbf{x}_{N,\ell} + \\ &a\pi_{0,\ell}\beta^2\mathbf{w}_{N,i}\mathbf{W}^2_N\mathbf{x}_{N,\ell} + \cdots \mid \mathbf{w}_0, \mathbf{x}_k] = 0,\end{aligned} \tag{A-5}$$

where the dots represent the infinite sum on r. Under the assumption that $\pi_{0,\ell} \neq 0$, the only way in which equation (A-5) holds for constant a, b and c different from zero is if the matrices $\mathbf{I}_N, \mathbf{W}_N, \mathbf{W}^2_N, \ldots$ are linearly dependent, and there exists $\mathbf{W}^r_N$ such that $x_{i,\ell} = \mathbf{w}^r_{N,i}\mathbf{x}_\ell$. If $\pi_{0,\ell} = 0$, clearly the first and the second component of the vector are linearly dependent. Therefore, Assumptions A.11 and A.12 imply that the component of the vector in A-4 are linearly independent. We chose the regressors k and ℓ arbitrarily. Then, under Assumption A.9 all the regressors are linearly independent, which implies that all the component of the vector

$$[\mathbb{E}[\mathbf{w}^p_{N,0,i}\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{y}_N], \ldots, \mathbb{E}[\mathbf{w}^p_{N,0,i}\mathbf{x}_{N,k}\mathbf{w}_{N,i}\mathbf{x}_{N,K}], \ldots, \mathbb{E}[\mathbf{w}^p_{N,0,i}\mathbf{x}_{N,k}x_{N,K,i}]]$$

are linearly independent. The same result follow for the vectors of the type

$$[\mathbb{E}[x_{N,k,i}\mathbf{w}_{N,i}\mathbf{y}_N], \ldots, \mathbb{E}[x_{N,k,i}\mathbf{w}_{N,i}\mathbf{x}_{N,k}], \ldots, \mathbb{E}[x^2_{N,k,i}]]$$

by conditioning on the arbitrary regressor $x_{N,k,i}$ for non-zero rows. To show that the rows are linearly independent, we can use an analogous approach by considering arbitrary values of $\mathbf{w}_i$ and $\mathbf{x}_K$. It is straightforward to see that under the column rank assumption on all matrices of regressors, Assumption A.13 implies the result. Finally, given that we showed linear independence conditions for non-zero rows, we need to show that there are at least $2K + 1$ rows different from zero. The rank condition on the matrix of regressors implies that we only need to focus on combinations of connections in $\mathbf{W}_{N,0}$ and $\mathbf{W}_N$ that can make $\mathbb{E}[\mathbf{w}_{0,N,i}^p \mathbf{x}_{N,\ell} \mathbf{w}_{N,i} \mathbf{x}_{N,k}] = 0$ for any value of $\mathbf{x}_\ell$ and $\mathbf{x}_k$ and some value of p. First note that

$$\mathbb{E}[\mathbf{w}_{0,N,i}^p \mathbf{x}_{N,\ell} \mathbf{w}_{N,i} \mathbf{x}_{N,k}] = \int_{\mathbf{x}_k:\mathbf{X}\in\mathscr{X}} \int_{\mathbf{x}_\ell:\mathbf{X}\in\mathscr{X}} \mathbb{E}[\mathbf{w}_{0,N,i}^p \mathbf{x}_\ell \mathbf{w}_i \mathbf{x}_N] f_{\mathbf{x}_{N,k}\mathbf{x}_{N,\ell}}(\mathbf{x}_k, \mathbf{x}_\ell) d\mathbf{x}_k d\mathbf{x}_\ell, \quad \text{(A-6)}$$

where $f_{\mathbf{x}_{N,k}\mathbf{x}_{N,\ell}}$ is the joint probability of $\mathbf{x}_{N,k}$ and $\mathbf{x}_{N,\ell}$. Take some arbitrary values $\mathbf{x}_k \neq 0$ and $\mathbf{x}_\ell \neq 0$ with positive probability in $f_{\mathbf{x}_{N,k}\mathbf{x}_{N,\ell}}$. It follows that

$$\mathbb{E}[\mathbf{w}_{0,N,i}^p \mathbf{x}_\ell \mathbf{w}_i \mathbf{x}_{N,k}] = \sum_{\mathbf{w}_{0,i}:\mathbf{g}_0\in\mathscr{G}_0} \sum_{\mathbf{w}_i:\mathbf{g}\in\mathscr{G}} \mathbf{w}_{0,i}^p \mathbf{x}_\ell \mathbf{w}_i \mathbf{x}_k \Pr(\mathcal{G}_N = \mathbf{g}, \mathcal{G}_{N,0} = \mathbf{g}_0). \quad \text{(A-7)}$$

The only way in which equation (A-7) can equal zero for different values of p, even when $\mathbf{x}_k \neq 0$ and $\mathbf{x}_\ell \neq 0$, is if the linear combination of $\mathbf{w}_{0,i}^p \mathbf{x}_\ell \mathbf{w}_i \mathbf{x}_k$ for different values of $\mathbf{w}_{0,i}^p$ and $\mathbf{w}_i$ weighted by their respective probabilities equals zero. Therefore, Assumption A.14 guarantees the existence of at least $3K$ rows different from zero. ■

B Proofs of Main Results

Proof of Theorem 1. First, note that Assumption 2 guarantees that the solution for model (3.1) exists. Assumption 3 guarantees that the system of equations $\mathbb{E}[\boldsymbol{m}_N(\boldsymbol{\theta})] = \mathbf{0}_K$ are not trivially satisfied by making all individuals $i \in \mathcal{I}_N$ isolated. 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. Choose an arbitrary vector of parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, such that $\mathbb{E}[\mathbf{m}(\boldsymbol{\theta})] = 0$. Assumptions 1 and 2 imply that $\mathbb{E}[N^{-1}\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}(y_{N;i} - \mathbf{d}_{N;i}^\top \boldsymbol{\theta})] = \mathbf{0}_K$. It follows that $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top](\boldsymbol{\theta}_0 - \boldsymbol{\theta}) + \mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\varepsilon_{N;i}] = \mathbf{0}_K$ and $\mathbb{E}[\sum_{i\in\mathcal{I}_N} \mathbf{z}_{N;i}\mathbf{d}_{N;i}^\top](\boldsymbol{\theta}_0 - \boldsymbol{\theta}) = \mathbf{0}_K$, given that N is arbitrarily large, but finite. Under Assumption 3, it follows that $\mathbb{E}[\mathbf{m}(\boldsymbol{\theta})] = \mathbf{0}_K$ 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

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

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

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

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

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

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

C Auxiliary Results

All the results in this section, and consequently in section B, are derived conditional on the sequence of networks $\{\mathcal{G}_n\}$. For simplicity in notation, we have omitted explicit conditioning on expectations, but it's important to note that all expectations are taken with respect to the conditional distribution $f_{\mathbf{X}_N,\boldsymbol{\varepsilon}_N|\mathcal{G}_N}$.

Lemma C.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*

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

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 (2021, 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 $\mathscr{L}_{Q+1,2}$. 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 (2021, Appendix A, pp. 899-907) applies to this setting (see also Corollary A.2. in Appendix A in Kojevnikov, Marmer, and Song, 2021, pp. 899-907). ■

Lemma C.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 $\mathcal{I}_i \subset \mathcal{I}_n$ is a set of indexes defined for each $i \in \mathcal{I}_n$, which can either be empty, equal to the union of individual i's connections in the networks $\mathcal{G}$ and $\mathcal{G}_0$, or equal to $\mathcal{P}_n(i,1)$. Then, as $n \to \infty$,*

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

Proof. Using the same approach of Jenish and Prucha (2009) and Kojevnikov, Marmer, and Song (2021), 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 $\tilde{R}^{(k)}_{n;i,j} = R_{n;i,j} - \varphi_k(R_{n;i,j}) = (R_{n;i,j} - \text{sgn}(R_{n;i,j})k)\mathbb{1}\{|R_{n;i,j}| > k\}$. Let $\|X\|_k = (\mathbb{E}[|X|^k)^{1/k}$ for $k \in [1, \infty)$. Therefore, following the previous definition, we apply the triangle inequality to get

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

From Assumption 7, note that the expectation on the second term of the previous equation is bounded by $\mathbb{E}[|\tilde{R}^{(k)}_{n;i,j}|] = \mathbb{E}[|\tilde{R}^{(k)}_{n;i,j}|\mathbb{1}\{|R_{n;i,j}| > k\}] \le 2\mathbb{E}[|R_{n;i,j}|\mathbb{1}\{|R_{n;i,j}| > k\}]$. Following the arguments as in Kojevnikov, Marmer, and Song (2021), the second component of the right-hand side in the equation above is bounded by $\sup_{n\ge1} \max_{i\in\mathcal{I}_n} \mathbb{E}\left[|R_{n;i,j}| \, 1\{|R_{n;i}| > k\}\right]$, where $\lim_{k\to\infty} \sup_{n\ge1} \max_{i\in\mathcal{I}_n} \mathbb{E}\left[|R_{n;i,j}| \, 1\{|R_{n;i,j}| > k\}\right] = 0$. Focusing on the first component of the right-hand side, by Lyapunov's inequality, it follows that

$$\begin{aligned}\left\| \frac{1}{n} \sum_{i\in\mathcal{I}_n} \sum_{j\in\mathcal{I}_i} w^*_{i,j} \left(R^{(k)}_{n;i,j} - \mathbb{E}\left[R^{(k)}_{n;i,j}\right] \right) \right\|_1 &\le \left\| \frac{1}{n} \sum_{i\in\mathcal{I}_n} \sum_{j\in\mathcal{I}_i} w^*_{i,j} \left(R^{(k)}_{n;i,j} - \mathbb{E}\left[R^{(k)}_{n;i,j}\right] \right) \right\|_2 \\ &= \frac{1}{n} \sqrt{\text{var}\left(\sum_{i\in\mathcal{I}_n} \sum_{j\in\mathcal{I}_i} w^*_{i,j} R^{(k)}_{n;i,j} \right)},\end{aligned} \tag{C-2}$$

where (C-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

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

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

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

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 C.1 under Assumptions 4 and 5, $|\operatorname{cov}(R^{(k)}_{n;i,j}R^{(k)}_{n;i,s})| \leq \psi_{1,1}\left(\varphi_k,\varphi_k\right)\lambda_{n,d}$ for $d_n(i,j) = d$ and φ_k is a bounded function with $\operatorname{Lip}\left(\psi_k\right) = 1$. 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

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

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

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

$$= C\sum_{i\in\mathcal{I}_n}\sum_{j\in\mathcal{I}_i} w^{*2}_{i,j} + 2\psi_{1,1}\left(\varphi_k,\varphi_k\right)\sum_{d\geq 1}\lambda_{n,d}\sum_{i\in\mathcal{I}_n}\left|\mathcal{P}_n(j,d)\right|,$$

$$\leq n\left(C\bar{\mathcal{I}}_n + 2\psi_{1,1}\left(\varphi_k,\varphi_k\right)\sum_{d\geq 1}\bar{D}_n(d)\lambda_{n,d}\right),$$

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 connections in the networks $\mathcal{G}$ and $\mathcal{G}_0$, or equal to $\mathcal{P}_n(i,1)$ (individual i's connections in network $\mathcal{G}$). Note that, for any of the three cases, $|\mathcal{I}_i| \leq |\mathcal{P}_n(i,1)|$ 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 in probability to zero by Assumption 6. It follows that $n^{-1}\bar{\mathcal{I}}_n \xrightarrow{\text{p}} 0$. Therefore,

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

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

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

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

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 C.2 by choosing $\mathcal{I}_i = \emptyset$ for (1), $\mathcal{I}_i$ as the union of individual i's connections 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

$$\mathbb{E}[\mathbf{W}_N\mathbf{y}] = \gamma_0\mathbf{W}_N\mathbf{x}_N + (\gamma_0\beta_0 + \delta_0)\sum_{p=0}^{\infty}\beta^k_0\mathbf{W}^{p+2}_N\mathbf{x}_N. \quad \text{(C-7)}$$

By choosing $\mathcal{I}_i$ to be the union of individual i's connections in the network $\mathcal{G}$ and the set of individuals at distance p from i (for all $p \in \mathbb{R}_+$), Lemma C.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 (C-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 C.2 (LLN for Instruments and Errors) *Let Assumptions 4 to 7 hold, then*

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

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 C.1 (1) and (3). ■

Corollary C.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* $n \to \infty$, $\boldsymbol{\Omega}_n \longrightarrow \boldsymbol{\Omega}_N < \infty$.

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 C.1 combined with the arguments in Lemma C.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 (C-6), using the appropriate values for $R_{n;i,j}$ and $\mathcal{I}_i$ (see Corollary C.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 $\boldsymbol{\Omega}_n \longrightarrow \boldsymbol{\Omega}_N$, 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 C.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,* $\mathbb{E}[z_{n;i,q}\varepsilon_{n;i}] = 0$. *As* $n \to \infty$,

$$\sup_{t\in\mathbf{R}}\left|\mathbf{P}\left\{\frac{S_n}{\sigma_n} \leq t\right\} - \Phi(t)\right| \longrightarrow 0,$$

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 C.1, the covariance of any two $Y_{n;i}$ and $Y_{n;j}$ is bounded. The proof then follows from applying the unconditional version of Lemmas A.2 and A.3 in Kojevnikov, Marmer, and Song (2021, Appendix A, pp. 899-907) to $Y_{n;i}$ and S_n/σ_n, respectively. ■

Lemma C.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 C.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 C.3, it follows that $\boldsymbol{\Omega}_N$ exists. Therefore, the result follows from an application of the Cramér-Wold device. ■

D Empirical Application

D.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 868 students participating. 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 2 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. The dataset also includes a cognitive ability test information on five personality measures: `Agreeableness`, `Conscientiousness`, `Extraversion`, `Neuroticism` and `Openness`.[1] In our empirical application, we also include the following variables: `Male` 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` are count variables, and `Commute by car/taxi` equals 1 when a student goes to school either by car or by taxi.

We also include indicator variables capturing students' engagement at school. For instance, the indicator variables `Siblings' Help` and `Parents' Help` take the value of 1 if students receive help from either their siblings or parents. We label those variables as 0 otherwise. To capture extracurricular school-related activities, we include the indicator variable `Music`, which equals 1 if students play music, and 0 otherwise.

[1]The numbers in parenthesis in Scale column of Table 2 represent the scales for the tests.

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 with whom they discussed 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 3 reports the summary statistics of the network 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. 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.

As mentioned in Section 6 in the main manuscript, we follow the Linear-in-Means model for social effects. To improve readability, we rewrite the equation 6.1 below,

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

Estimations results are shown in Table 1, in Section 6 in the main text, and Table 4 (in this section). As mentioned in the manuscript, estimators that do not control for network endogeneity tend to underestimate peer effects. In the context of our simulation results, one possible explanation for the negative bias 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 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 collaboration between students inside and outside the classroom can generate benefits that have the potential to generate positive social multipliers. All results are qualitatively robust to different choices of p and D_n; see section D.2 in the supplemental material.

D.2 Supplementary Estimation Results

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

D.3 Assessing Assumptions

To validate the Assumption 2, we perform a Least Square (LS) regression that includes all variables in equation 6.1 and our proposed instruments. Namely, we run the following equation,

$$\begin{aligned} \texttt{math} \;=\;& \beta \mathbf{W}_n \texttt{math} + \texttt{personality}\,\boldsymbol{\gamma}_p + \texttt{characteristics}\,\boldsymbol{\gamma}_{ch} \\ &+ \mathbf{W}_n\,\texttt{characteristics}\,\boldsymbol{\delta} + \mathbf{W}_{n,0}\,\texttt{characteristics}\,\boldsymbol{\delta}_1 \\ &+ \mathbf{W}_{n,0}^2\,\texttt{characteristics}\,\boldsymbol{\delta}_2 + \text{error}, \end{aligned} \tag{D-1}$$

where `characteristics`, `personality`, and the adjacency matrices $\mathbf{W}_n$, $\mathbf{W}_{n,0}$ are defined above. Estimation results for equation D-1 are shown in Table 11. They suggest no correlation between the output variable `math` and the vectors $\mathbf{W}_{n,0}$ `characteristics` and $\mathbf{W}_{n,0}^2$ `characteristics`, which are the proposed instruments. The estimated coefficients $\boldsymbol{\delta}_1$ and $\boldsymbol{\delta}_2$ are statistically non-significant[2], which allow us to use the random assignment embodied in $\mathbf{W}_{n,0}$ to identify the parameters of the linear model (6.1).

To validate the assumption 3, we run a series of LS regressions in which the outcome is the endogenous variable $\mathbf{W}_n$`math`. Namely, we estimate the following models,

[2]The estimated coefficient of $\mathbf{W}_{n,0}^2$ln(Weight) is significant at 10%.

$$\begin{aligned}
\mathbf{W}_n\texttt{math} &= \mathbf{W}_{n,0}\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-2)}\\
\mathbf{W}_n\texttt{math} &= \mathbf{W}_{n,0}^2\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}^2\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-3)}\\
\mathbf{W}_n\texttt{math} &= \mathbf{W}_{n,0}\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}\ \texttt{personality}\ \boldsymbol{\gamma} && \text{(D-4)}\\
&\quad + \mathbf{W}_{n,0}^2\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}^2\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}.
\end{aligned}$$

The estimation results for these specifications are shown in Table 12. F-Statistics suggest that our proposed instruments are relevant to describe the endogenous variable. Furthermore, we also run LS regressions where the dependent variable $\mathbf{W}_n\mathbf{x}$ is the average of characteristics x between the study partners. $\mathbf{x}$ can be any of the characteristics mentioned above, say height, weight, siblings help, parents help, commute to school by car or taxi; playing music; and whether the student is a male.

$$\begin{aligned}
\mathbf{W}_n\mathbf{x} &= \mathbf{W}_{n,0}\ \texttt{characteristics}_{-\mathbf{x}}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-5)}\\
\mathbf{W}_n\mathbf{x} &= \mathbf{W}_{n,0}^2\ \texttt{characteristics}_{-\mathbf{x}}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}^2\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-6)}\\
\mathbf{W}_n\mathbf{x} &= \mathbf{W}_{n,0}\ \texttt{characteristics}_{-\mathbf{x}}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-7)}\\
&\quad + \mathbf{W}_{n,0}^2\ \texttt{characteristics}_{-\mathbf{x}}\ \boldsymbol{\delta} + \mathbf{W}_{n,0}^2\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}.
\end{aligned}$$

Here $\texttt{characteristics}_{-\mathbf{x}}$ means that the characteristic includes all previous variables except $\mathbf{x}$, used in $\mathbf{W}_n\mathbf{x}$. Tables 13-19 show OLS estimates, and F-Statistics also suggest that our proposed instruments are relevant to describe endogenous variables, $\mathbf{W}_n\mathbf{x}$.

Finally, we present an innovative network architecture that integrates students based on both study partnerships and shared seating arrangements. Specifically, we establish a new set of connections, denoted $\mathbf{W}_{n,*}$, between students i and k following a defined rule: If students i and j mutually identified each other as study partners in the survey, and students j and k reciprocated as seatmates, we establish an edge between i and k. This connection is formed when neither students i and k are study partners, nor are students j and k. Table 1 reports the summary statistics of this new network in the column named Extra. Based on this new set of connections, we rewrite previous equations as follows.

$$
\begin{aligned}
\mathbf{W}_n\texttt{math} &= \mathbf{W}_{n,*}\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,*}\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-8)}\\
\mathbf{W}_n\texttt{math} &= \mathbf{W}_{n,*}^2\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,*}^2\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-9)}\\
\mathbf{W}_n\texttt{math} &= \mathbf{W}_{n,*}\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,*}\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}, && \text{(D-10)}\\
&\quad + \mathbf{W}_{n,*}^2\ \texttt{characteristics}\ \boldsymbol{\delta} + \mathbf{W}_{n,*}^2\ \texttt{personality}\ \boldsymbol{\gamma} + \text{error}.
\end{aligned}
$$

The LS estimates for all these three specifications are shown in Table 20. F-Statistics suggest that our proposed instruments are relevant to describe the endogenous variable.

Table 2: Summary Statistics

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
Cognitiva 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 each school are presented here. Course grades, personality trait measures, and cognitive ability tests are scored on the scale indicated.

Table 3: Summary Network Statistics

	School 1			School 2			School 3		
Variables	Studymates	Seatmates	Extra	Studymates	Seatmates	Extra	Studymates	Seatmates	Extra
Number of nodes	133	133	133	171	171	171	564	564	564
Number of edges	171	175	454	228	275	748	819	799	2974
Density × 100	1.948	1.994	5.172	1.569	1.892	5.146	0.516	0.503	1.873
Average degree	2.571	2.632	6.827	2.667	3.216	8.749	2.904	2.833	10.546
Average clustering	0.213	0.068	0.209	0.153	0.066	0.228	0.129	0.063	0.160
Assortativity measure	0.190	0.266	0.187	0.039	0.194	0.256	0.177	0.177	0.067
Number of isolated node	15	7	3	19	3	5	69	13	5
Number of Subgraph	21	14	4	27	9	6	76	30	8
Transitivity	0.281	0.094	0.244	0.207	0.092	0.234	0.180	0.081	0.153

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

Table 4: Estimations results, cont.

	OLS		G2SLS		GMM	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	-0.0196**	(0.0099)	-0.0031	(0.0130)	-0.0019	(0.0126)
Elder Siblings Help	0.0087	(0.0116)	-0.0030	(0.0126)	0.0001	(0.0138)
Younger Siblings Help	0.0435***	(0.0159)	0.0403**	(0.0201)	0.0382*	(0.0212)
Male $\times$ ln(Cognitive)	0.1275***	(0.0486)	0.1528**	(0.0680)	0.2008***	(0.0576)
Male $\times$ ln(Agreeableness)	-0.1318**	(0.0668)	-0.0318	(0.0622)	-0.0407	(0.0667)
Male $\times$ ln(Conscientiousness)	0.1216**	(0.0520)	0.1612**	(0.0639)	0.1582**	(0.0727)
Male $\times$ ln(Extraversion)	0.0266	(0.0404)	0.0565	(0.0559)	0.0434	(0.0641)
Male $\times$ ln(Neuroticism)	0.0441	(0.0518)	0.1033	(0.0630)	0.1003	(0.0658)
Male $\times$ ln(Openness)	-0.0157	(0.0566)	-0.0968	(0.0805)	-0.1089	(0.0945)
School 1, grade 7, class 1	0.1081**	(0.0516)	0.0417***	(0.0155)	0.0293*	(0.0176)
School 1, grade 7, class 2	0.0243	(0.0533)	-0.0682***	(0.0247)	-0.1288***	(0.0238)
School 1, grade 7, class 3	0.0854***	(0.0283)	0.0658***	(0.0106)	0.0779***	(0.0111)
School 1, grade 7, class 4	0.1583***	(0.0525)	0.0889***	(0.0148)	0.1177***	(0.0132)
School 1, grade 7, class 5	0.1897***	(0.0493)	0.1225***	(0.0154)	0.1394***	(0.0124)
School 2, grade 7, class 1	0.0938***	(0.0275)	0.0204	(0.0125)	0.0162	(0.0135)
School 2, grade 7, class 2	0.0516	(0.0326)	-0.0120	(0.0132)	-0.0172	(0.0159)
School 2, grade 7, class 3	0.1350***	(0.0428)	0.0538***	(0.0153)	0.0649***	(0.0155)
School 2, grade 7, class 4	0.0720**	(0.0316)	0.0472***	(0.0141)	0.0390**	(0.0161)
School 2, grade 7, class 5	0.1312***	(0.0449)	0.1126***	(0.0180)	0.1480***	(0.0164)
School 3, grade 7, class 1	0.0948**	(0.0433)	0.1207***	(0.0159)	0.1541***	(0.0131)
School 3, grade 7, class 2	0.0983**	(0.0434)	0.1335***	(0.0165)	0.1755***	(0.0117)
School 3, grade 7, class 3	0.0900**	(0.0438)	0.1488***	(0.0172)	0.1916***	(0.0170)
School 3, grade 7, class 4	0.0719*	(0.0385)	0.1034***	(0.0153)	0.1294***	(0.0129)
School 3, grade 7, class 5	0.0612	(0.0453)	0.0944***	(0.0202)	0.1384***	(0.0151)
School 3, grade 8, class 1	0.1364***	(0.0379)	0.1232***	(0.0144)	0.1547***	(0.0096)
School 3, grade 8, class 2	0.1223**	(0.0520)	0.1482***	(0.0235)	0.2016***	(0.0162)
School 3, grade 8, class 3	0.0982***	(0.0334)	0.1308***	(0.0150)	0.1619***	(0.0116)
School 3, grade 8, class 4	0.1069**	(0.0484)	0.1493***	(0.0219)	0.1997***	(0.0172)
School 3, grade 8, class 5	0.0770*	(0.0411)	0.1087***	(0.0175)	0.1545***	(0.0115)
School 3, grade 9, class 1	0.0409*	(0.0246)	0.0782***	(0.0075)	0.0693***	(0.0072)
School 3, grade 9, class 2	0.0384	(0.0263)	0.0557***	(0.0066)	0.0575***	(0.0090)
School 3, grade 9, class 3	0.0374	(0.0295)	0.0543***	(0.0082)	0.0518***	(0.0080)
School 3, grade 9, class 4	0.0382	(0.0291)	0.0798***	(0.0072)	0.0819***	(0.0084)
Constant	-4.5872*	(2.5808)	2.9464***	(0.9048)	3.8210***	(1.2635)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Estimations results with $p = 3$

Variables	$C = 1.5$	$C = 1.6$	$C = 1.7$
Peer effect			
ln(Math Test)	0.7002*	0.8889***	0.8919***
	(0.3737)	(0.3157)	(0.3069)
Contextual effects			
Male	-0.2958	-0.2395*	-0.2403*
	(0.1836)	(0.1436)	(0.1417)
ln(Height)	3.8548*	2.2241**	2.2012**
	(2.0045)	(1.0261)	(1.0108)
ln(Weight)	0.0245	-0.0452	-0.0558
	(0.4489)	(0.3267)	(0.3123)
Siblings Help	-0.0049	0.0674	0.0630
	(0.1390)	(0.0833)	(0.0812)
Parents Help	-0.0303	-0.1154	-0.1114
	(0.1479)	(0.0841)	(0.0823)
Commute by Car/Taxi	0.0857	0.1395	0.1515
	(0.1307)	(0.1126)	(0.1110)
Music	0.0575	0.0842	0.0782
	(0.0989)	(0.0841)	(0.0822)
†			
ln(Cognitive)	0.1119**	0.1318***	0.1322***
	(0.0494)	(0.0437)	(0.0421)
ln(Agreeableness)	-0.0953	-0.1254**	-0.1204*
	(0.0720)	(0.0632)	(0.0618)
ln(Conscientiousness)	0.0738	0.0649	0.0647
	(0.0740)	(0.0718)	(0.0713)
ln(Extraversion)	-0.1552**	-0.1290**	-0.1259**
	(0.0768)	(0.0614)	(0.0608)
ln(Neuroticism)	-0.0003	-0.0333	-0.0331
	(0.0447)	(0.0396)	(0.0389)
ln(Openness)	0.0229	0.0115	0.0068
	(0.0741)	(0.0654)	(0.0645)
Direct effects			
Male	-0.2902	-0.6800	-0.6580
	(0.8509)	(0.8366)	(0.8331)
ln(Height)	-1.3506**	-0.9101***	-0.9102***
	(0.5345)	(0.3046)	(0.2987)
ln(Weight)	-0.0692	-0.0238	-0.0187
	(0.0966)	(0.0759)	(0.0725)
Siblings Help	-0.0407	-0.0640**	-0.0627**
	(0.0444)	(0.0282)	(0.0273)
Parents Help	0.0229	0.0378*	0.0371*
	(0.0295)	(0.0203)	(0.0200)
Commute by Car/Taxi	-0.0229	-0.0315*	-0.0335*
	(0.0190)	(0.0176)	(0.0171)
Degree	0.0258***	0.0209***	0.0208***
	(0.0067)	(0.0058)	(0.0057)
Isolate Students	0.3446	0.4858	0.4755
	(0.4991)	(0.4822)	(0.4766)
n	868	868	868
Adjusted R^2	0.2100	0.2617	0.2615
RMSE	0.2205	0.2092	0.2094

Note: (i) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; (ii) Standard errors are in parentheses. (iii) † These regressors are measured as the deviation of students' personality from their peers' average.

Table 6: Estimations results with $p = 3$, cont.

	$C = 1.5$		$C = 1.6$		$C = 1.7$	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	-0.0164	(0.0189)	-0.0115	(0.0163)	-0.0102	(0.0158)
Elder Siblings Help	0.0171	(0.0207)	0.0086	(0.0173)	0.0073	(0.0169)
Younger Siblings Help	0.0357	(0.0296)	0.0470**	(0.0218)	0.0470**	(0.0216)
Male × ln(Cognitive)	0.0823	(0.0951)	0.0802	(0.0902)	0.0788	(0.0875)
Male × ln(Agreeableness)	-0.0804	(0.1019)	-0.0349	(0.1017)	-0.0354	(0.0997)
Male × ln(Conscientiousness)	0.1095	(0.0821)	0.1334*	(0.0751)	0.1324*	(0.0736)
Male × ln(Extraversion)	-0.0066	(0.0830)	0.0313	(0.0632)	0.0286	(0.0626)
Male × ln(Neuroticism)	0.0384	(0.0846)	0.0700	(0.0777)	0.0660	(0.0776)
Male × ln(Openness)	0.0417	(0.0925)	0.0239	(0.0821)	0.0261	(0.0798)
School 1, grade 7, class 1	0.1649**	(0.0771)	0.1201**	(0.0604)	0.1136*	(0.0592)
School 1, grade 7, class 2	0.0582	(0.1200)	0.0671	(0.1028)	0.0607	(0.0998)
School 1, grade 7, class 3	0.1181**	(0.0460)	0.0711*	(0.0367)	0.0687*	(0.0354)
School 1, grade 7, class 4	0.2019**	(0.0891)	0.1097	(0.0687)	0.1033	(0.0671)
School 1, grade 7, class 5	0.2346***	(0.0808)	0.1595**	(0.0639)	0.1525**	(0.0624)
School 2, grade 7, class 1	0.1464**	(0.0568)	0.1190***	(0.0414)	0.1146***	(0.0392)
School 2, grade 7, class 2	0.1105	(0.0674)	0.0621	(0.0532)	0.0609	(0.0509)
School 2, grade 7, class 3	0.1861***	(0.0704)	0.1174**	(0.0564)	0.1107**	(0.0544)
School 2, grade 7, class 4	0.1138*	(0.0600)	0.1154**	(0.0463)	0.1113**	(0.0444)
School 2, grade 7, class 5	0.1701**	(0.0767)	0.1049	(0.0644)	0.1016	(0.0624)
School 3, grade 7, class 1	0.1646**	(0.0769)	0.0778	(0.0569)	0.0737	(0.0558)
School 3, grade 7, class 2	0.1504*	(0.0808)	0.0727	(0.0623)	0.0688	(0.0611)
School 3, grade 7, class 3	0.1408*	(0.0744)	0.0634	(0.0600)	0.0609	(0.0587)
School 3, grade 7, class 4	0.1212*	(0.0631)	0.0701	(0.0488)	0.0665	(0.0475)
School 3, grade 7, class 5	0.0904	(0.0741)	0.0268	(0.0611)	0.0242	(0.0600)
School 3, grade 8, class 1	0.1560**	(0.0616)	0.1077**	(0.0508)	0.1055**	(0.0496)
School 3, grade 8, class 2	0.1342	(0.0906)	0.0715	(0.0770)	0.0690	(0.0751)
School 3, grade 8, class 3	0.0943	(0.0590)	0.0611	(0.0492)	0.0586	(0.0480)
School 3, grade 8, class 4	0.1327	(0.0845)	0.0601	(0.0714)	0.0568	(0.0700)
School 3, grade 8, class 5	0.0750	(0.0640)	0.0438	(0.0592)	0.0407	(0.0574)
School 3, grade 9, class 1	0.0446	(0.0355)	0.0515*	(0.0301)	0.0486	(0.0298)
School 3, grade 9, class 2	0.0319	(0.0427)	0.0269	(0.0338)	0.0238	(0.0331)
School 3, grade 9, class 3	0.0239	(0.0554)	0.0308	(0.0394)	0.0270	(0.0383)
School 3, grade 9, class 4	0.0428	(0.0380)	0.0500	(0.0336)	0.0488	(0.0334)
Constant	-11.5128*	(6.3535)	-6.1593*	(3.3041)	-6.0328*	(3.2568)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Estimations results with $p = 4$

Variables	$C = 1.5$	$C = 1.6$	$C = 1.7$
Peer effect			
ln(Math Test)	0.6729***	0.6762***	0.6791***
	(0.2236)	(0.2159)	(0.2090)
Contextual effects			
Male	-0.2057*	-0.2111*	-0.2161**
	(0.1115)	(0.1080)	(0.1046)
ln(Height)	2.2455***	2.2006***	2.1608***
	(0.8568)	(0.8471)	(0.8359)
ln(Weight)	0.0443	0.0460	0.0486
	(0.2703)	(0.2653)	(0.2607)
Siblings Help	0.0387	0.0399	0.0419
	(0.0567)	(0.0546)	(0.0525)
Parents Help	-0.0205	-0.0133	-0.0072
	(0.0658)	(0.0639)	(0.0620)
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)
†			
ln(Cognitive)	0.1098***	0.1107***	0.1118***
	(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)	-0.1227***	-0.1206**	-0.1185**
	(0.0474)	(0.0469)	(0.0462)
ln(Neuroticism)	-0.0150	-0.0136	-0.0122
	(0.0293)	(0.0284)	(0.0275)
ln(Openness)	0.0221	0.0179	0.0138
	(0.0493)	(0.0475)	(0.0458)
Direct effects			
Male	-0.4717	-0.4442	-0.4138
	(0.6411)	(0.6298)	(0.6167)
ln(Height)	-0.9817***	-0.9767***	-0.9699***
	(0.2697)	(0.2625)	(0.2553)
ln(Weight)	-0.0685	-0.0668	-0.0656
	(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	0.2403	0.2186	0.1963
	(0.3495)	(0.3417)	(0.3327)
n	868	868	868
Adjusted R^2	0.2800	0.2813	0.2821
RMSE	0.1980	0.1979	0.1980

Note: (i) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; (ii) Standard errors are in parentheses. (iii) † These regressors are measured as the deviation of students' personality from their peers' average.

Table 8: Estimations results with $p = 4$, cont.

	$C = 1.5$		$C = 1.6$		$C = 1.7$	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	-0.0232*	(0.0119)	-0.0225*	(0.0115)	-0.0219*	(0.0112)
Elder Siblings Help	0.0052	(0.0129)	0.0055	(0.0125)	0.0060	(0.0121)
Younger Siblings Help	0.0372*	(0.0192)	0.0380**	(0.0190)	0.0389**	(0.0188)
Male × ln(Cognitive)	0.1055	(0.0675)	0.1049	(0.0640)	0.1044*	(0.0609)
Male × ln(Agreeableness)	-0.0848	(0.0879)	-0.0866	(0.0857)	-0.0885	(0.0836)
Male × ln(Conscientiousness)	0.1311**	(0.0634)	0.1297**	(0.0614)	0.1276**	(0.0599)
Male × ln(Extraversion)	0.0488	(0.0488)	0.0473	(0.0461)	0.0460	(0.0437)
Male × ln(Neuroticism)	0.0720	(0.0620)	0.0703	(0.0610)	0.0682	(0.0600)
Male × ln(Openness)	-0.0239	(0.0706)	-0.0242	(0.0672)	-0.0246	(0.0641)
School 1, grade 7, class 1	0.1284**	(0.0519)	0.1232**	(0.0510)	0.1189**	(0.0502)
School 1, grade 7, class 2	0.0504	(0.0677)	0.0469	(0.0656)	0.0439	(0.0636)
School 1, grade 7, class 3	0.0897***	(0.0324)	0.0879***	(0.0312)	0.0867***	(0.0301)
School 1, grade 7, class 4	0.1606***	(0.0563)	0.1568***	(0.0555)	0.1540***	(0.0547)
School 1, grade 7, class 5	0.1954***	(0.0528)	0.1902***	(0.0518)	0.1859***	(0.0507)
School 2, grade 7, class 1	0.0985***	(0.0343)	0.0960***	(0.0320)	0.0935***	(0.0299)
School 2, grade 7, class 2	0.0584	(0.0450)	0.0571	(0.0430)	0.0570	(0.0412)
School 2, grade 7, class 3	0.1376***	(0.0457)	0.1335***	(0.0443)	0.1304***	(0.0429)
School 2, grade 7, class 4	0.0852**	(0.0343)	0.0823**	(0.0328)	0.0802**	(0.0312)
School 2, grade 7, class 5	0.1242**	(0.0494)	0.1226**	(0.0479)	0.1219***	(0.0465)
School 3, grade 7, class 1	0.1041**	(0.0477)	0.0995**	(0.0468)	0.0955**	(0.0460)
School 3, grade 7, class 2	0.1009**	(0.0498)	0.0988**	(0.0486)	0.0968**	(0.0476)
School 3, grade 7, class 3	0.0940*	(0.0497)	0.0909*	(0.0483)	0.0882*	(0.0470)
School 3, grade 7, class 4	0.0791*	(0.0431)	0.0757*	(0.0416)	0.0729*	(0.0401)
School 3, grade 7, class 5	0.0560	(0.0507)	0.0542	(0.0497)	0.0527	(0.0488)
School 3, grade 8, class 1	0.1302***	(0.0396)	0.1290***	(0.0383)	0.1280***	(0.0373)
School 3, grade 8, class 2	0.1121*	(0.0590)	0.1101*	(0.0572)	0.1085*	(0.0555)
School 3, grade 8, class 3	0.0903**	(0.0382)	0.0879**	(0.0371)	0.0858**	(0.0361)
School 3, grade 8, class 4	0.1058*	(0.0550)	0.1025*	(0.0536)	0.0996*	(0.0523)
School 3, grade 8, class 5	0.0644	(0.0465)	0.0635	(0.0450)	0.0629	(0.0437)
School 3, grade 9, class 1	0.0475*	(0.0256)	0.0442*	(0.0256)	0.0413	(0.0253)
School 3, grade 9, class 2	0.0424	(0.0297)	0.0390	(0.0292)	0.0359	(0.0289)
School 3, grade 9, class 3	0.0483	(0.0342)	0.0434	(0.0332)	0.0394	(0.0323)
School 3, grade 9, class 4	0.0358	(0.0309)	0.0339	(0.0307)	0.0319	(0.0302)
Constant	-5.2394*	(2.7824)	-5.0676*	(2.7405)	-4.9306*	(2.6995)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Estimations results with $p = 5$

Variables	$C = 1.5$	$C = 1.6$	$C = 1.7$
Peer effect			
ln(Math Test)	0.5943***	0.5987***	0.6065***
	(0.1923)	(0.1836)	(0.1755)
Contextual effects			
Male	-0.2494***	-0.2466***	-0.2455***
	(0.0949)	(0.0909)	(0.0873)
ln(Height)	2.3137***	2.2570***	2.1682***
	(0.7917)	(0.7660)	(0.7400)
ln(Weight)	0.0491	0.0399	0.0329
	(0.2123)	(0.2031)	(0.1942)
Siblings Help	0.0263	0.0274	0.0294
	(0.0521)	(0.0505)	(0.0491)
Parents Help	-0.0326	-0.0207	-0.0106
	(0.0590)	(0.0554)	(0.0517)
Commute by Car/Taxi	0.1147*	0.1205*	0.1255**
	(0.0683)	(0.0659)	(0.0636)
Music	0.1243**	0.1206**	0.1187**
	(0.0531)	(0.0505)	(0.0480)
†			
ln(Cognitive)	0.1093***	0.1087***	0.1086***
	(0.0317)	(0.0300)	(0.0286)
ln(Agreeableness)	-0.0866*	-0.0787	-0.0712
	(0.0501)	(0.0479)	(0.0457)
ln(Conscientiousness)	0.0752*	0.0774*	0.0799**
	(0.0435)	(0.0418)	(0.0406)
ln(Extraversion)	-0.1229***	-0.1202***	-0.1164***
	(0.0420)	(0.0412)	(0.0405)
ln(Neuroticism)	-0.0132	-0.0124	-0.0113
	(0.0278)	(0.0266)	(0.0257)
ln(Openness)	0.0353	0.0311	0.0272
	(0.0456)	(0.0435)	(0.0414)
Direct effects			
Male	-0.2089	-0.1901	-0.1648
	(0.5776)	(0.5609)	(0.5417)
ln(Height)	-1.0302***	-1.0141***	-0.9904***
	(0.2436)	(0.2343)	(0.2248)
ln(Weight)	-0.0578	-0.0537	-0.0494
	(0.0516)	(0.0483)	(0.0453)
Siblings Help	-0.0394*	-0.0391*	-0.0395*
	(0.0225)	(0.0214)	(0.0204)
Parents Help	0.0224	0.0194	0.0167
	(0.0155)	(0.0147)	(0.0139)
Commute by Car/Taxi	-0.0216	-0.0234*	-0.0251*
	(0.0145)	(0.0140)	(0.0136)
Degree	0.0251***	0.0250***	0.0248***
	(0.0037)	(0.0036)	(0.0034)
Isolate Students	0.2024	0.1752	0.1425
	(0.3118)	(0.3022)	(0.2935)
n	868	868	868
Adjusted R^2	0.2593	0.2636	0.2675
RMSE	0.2020	0.2010	0.2002

Note: (i) * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$; (ii) Standard errors are in parentheses. (iii) † These regressors are measured as the deviation of students' personality from their peers' average.

Table 10: Estimations results with $p = 5$, cont.

	$C = 1.5$		$C = 1.6$		$C = 1.7$	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
Music	-0.0204*	(0.0110)	-0.0199*	(0.0105)	-0.0196**	(0.0099)
Elder Siblings Help	0.0088	(0.0125)	0.0087	(0.0121)	0.0087	(0.0116)
Younger Siblings Help	0.0410**	(0.0169)	0.0421**	(0.0164)	0.0435***	(0.0159)
Male × ln(Cognitive)	0.1349**	(0.0538)	0.1313**	(0.0510)	0.1275***	(0.0486)
Male × ln(Agreeableness)	-0.1317*	(0.0730)	-0.1321*	(0.0702)	-0.1318**	(0.0668)
Male × ln(Conscientiousness)	0.1256**	(0.0563)	0.1240**	(0.0539)	0.1216**	(0.0520)
Male × ln(Extraversion)	0.0337	(0.0472)	0.0298	(0.0438)	0.0266	(0.0404)
Male × ln(Neuroticism)	0.0425	(0.0538)	0.0435	(0.0527)	0.0441	(0.0518)
Male × ln(Openness)	-0.0163	(0.0633)	-0.0154	(0.0598)	-0.0157	(0.0566)
School 1, grade 7, class 1	0.1198**	(0.0543)	0.1142**	(0.0530)	0.1081**	(0.0516)
School 1, grade 7, class 2	0.0304	(0.0583)	0.0268	(0.0557)	0.0243	(0.0533)
School 1, grade 7, class 3	0.0877***	(0.0309)	0.0869***	(0.0296)	0.0854***	(0.0283)
School 1, grade 7, class 4	0.1716***	(0.0540)	0.1652***	(0.0533)	0.1583***	(0.0525)
School 1, grade 7, class 5	0.2042***	(0.0521)	0.1973***	(0.0508)	0.1897***	(0.0493)
School 2, grade 7, class 1	0.1028***	(0.0324)	0.0985***	(0.0299)	0.0938***	(0.0275)
School 2, grade 7, class 2	0.0573	(0.0370)	0.0541	(0.0347)	0.0516	(0.0326)
School 2, grade 7, class 3	0.1482***	(0.0456)	0.1418***	(0.0443)	0.1350***	(0.0428)
School 2, grade 7, class 4	0.0801**	(0.0356)	0.0759**	(0.0337)	0.0720**	(0.0316)
School 2, grade 7, class 5	0.1387***	(0.0478)	0.1352***	(0.0463)	0.1312***	(0.0449)
School 3, grade 7, class 1	0.1047**	(0.0460)	0.1005**	(0.0446)	0.0948**	(0.0433)
School 3, grade 7, class 2	0.1046**	(0.0458)	0.1022**	(0.0446)	0.0983**	(0.0434)
School 3, grade 7, class 3	0.0969**	(0.0471)	0.0940**	(0.0454)	0.0900**	(0.0438)
School 3, grade 7, class 4	0.0816**	(0.0409)	0.0770*	(0.0397)	0.0719*	(0.0385)
School 3, grade 7, class 5	0.0682	(0.0468)	0.0652	(0.0461)	0.0612	(0.0453)
School 3, grade 8, class 1	0.1428***	(0.0403)	0.1398***	(0.0391)	0.1364***	(0.0379)
School 3, grade 8, class 2	0.1314**	(0.0551)	0.1271**	(0.0535)	0.1223**	(0.0520)
School 3, grade 8, class 3	0.1049***	(0.0352)	0.1016***	(0.0343)	0.0982***	(0.0334)
School 3, grade 8, class 4	0.1167**	(0.0509)	0.1124**	(0.0496)	0.1069**	(0.0484)
School 3, grade 8, class 5	0.0818*	(0.0440)	0.0795*	(0.0425)	0.0770*	(0.0411)
School 3, grade 9, class 1	0.0450*	(0.0254)	0.0428*	(0.0251)	0.0409*	(0.0246)
School 3, grade 9, class 2	0.0415	(0.0279)	0.0401	(0.0271)	0.0384	(0.0263)
School 3, grade 9, class 3	0.0476	(0.0324)	0.0421	(0.0310)	0.0374	(0.0295)
School 3, grade 9, class 4	0.0430	(0.0306)	0.0407	(0.0300)	0.0382	(0.0291)
Constant	-5.0956*	(2.7565)	-4.8916*	(2.6662)	-4.5872*	(2.5808)

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: LS Estimation results of the model $y = \beta\mathbf{W}\mathrm{y} + \boldsymbol{\delta}\mathrm{X} + \boldsymbol{\gamma}\mathbf{W}\mathrm{X} + \boldsymbol{\theta}_1\mathbf{W}_0\mathrm{X} + \boldsymbol{\theta}_2\mathbf{W}_0^2\mathrm{X}$

Variables	Coef.	SE	Variables	Coef.	SE
ln(Math Test)	0.2204***	(0.0784)	$\mathbf{W}_{n,0}^2$ × Male	-0.0512	(0.0466)
Male	-0.3938	(0.4050)	$\mathbf{W}_{n,0}^2$ × ln(Height)	-0.3393	(0.2333)
ln(Height)	-0.2541*	(0.1477)	$\mathbf{W}_{n,0}^2$ × ln(Weight)	0.1710*	(0.0903)
ln(Weight)	-0.0669*	(0.0392)	$\mathbf{W}_{n,0}^2$ × Siblings Help	0.0571	(0.0470)
Siblings Help	-0.0262	(0.0223)	$\mathbf{W}_{n,0}^2$ × Parents Help	-0.0117	(0.0557)
Parents Help	0.0290	(0.0248)	$\mathbf{W}_{n,0}^2$ × Commute by Car/Taxi	0.0111	(0.0508)
Commute by Car/Taxi	-0.0104	(0.0200)	$\mathbf{W}_{n,0}^2$ × Music	0.0251	(0.0363)
Music	-0.0091	(0.0132)	School 1, grade 7, class 1	0.0105	(0.0232)
Degree	0.0281***	(0.0040)	School 1, grade 7, class 2	-0.1337***	(0.0251)
Isolate Students	1.2398	(1.1350)	School 1, grade 7, class 3	0.0768***	(0.0127)
Elder Siblings Help	-0.0002	(0.0140)	School 1, grade 7, class 4	0.1075***	(0.0206)
Younger Siblings Help	0.0364*	(0.0191)	School 1, grade 7, class 5	0.1462***	(0.0167)
Male × ln(Cognitive)	0.2563***	(0.0452)	School 2, grade 7, class 1	0.0233	(0.0150)
Male × ln(Agreeableness)	-0.1395**	(0.0588)	School 2, grade 7, class 2	-0.0049	(0.0244)
Male × ln(Conscientiousness)	0.2086***	(0.0624)	School 2, grade 7, class 3	0.0742***	(0.0175)
Male × ln(Extraversion)	-0.0559	(0.0543)	School 2, grade 7, class 4	0.0487***	(0.0177)
Male × ln(Neuroticism)	0.0632	(0.0537)	School 2, grade 7, class 5	0.1526***	(0.0227)
Male × ln(Openness)	-0.0795	(0.0802)	School 3, grade 7, class 1	0.1593***	(0.0188)
$\mathbf{W}_n$ × Male	-0.0330	(0.0299)	School 3, grade 7, class 2	0.1710***	(0.0187)
$\mathbf{W}_n$ × ln(Height)	-0.0563	(0.2103)	School 3, grade 7, class 3	0.1997***	(0.0270)
$\mathbf{W}_n$ × ln(Weight)	0.1383**	(0.0673)	School 3, grade 7, class 4	0.1355***	(0.0249)
$\mathbf{W}_n$ × Siblings Help	0.0321	(0.0259)	School 3, grade 7, class 5	0.1486***	(0.0184)
$\mathbf{W}_n$ × Parents Help	0.0438*	(0.0264)	School 3, grade 8, class 1	0.1516***	(0.0167)
$\mathbf{W}_n$ × Commute by Car/Taxi	0.0334	(0.0264)	School 3, grade 8, class 2	0.2023***	(0.0183)
$\mathbf{W}_n$ × Music	-0.0236	(0.0201)	School 3, grade 8, class 3	0.1563***	(0.0129)
$\mathbf{W}_{n,0}$ × Male	-0.0248	(0.0149)	School 3, grade 8, class 4	0.1964***	(0.0215)
$\mathbf{W}_{n,0}$ × ln(Height)	0.2446	(0.2356)	School 3, grade 8, class 5	0.1676***	(0.0112)
$\mathbf{W}_{n,0}$ × ln(Weight)	-0.0355	(0.0528)	School 3, grade 9, class 1	0.0651***	(0.0125)
$\mathbf{W}_{n,0}$ × Siblings Help	-0.0151	(0.0213)	School 3, grade 9, class 2	0.0449***	(0.0067)
$\mathbf{W}_{n,0}$ × Parents Help	-0.0106	(0.0206)	School 3, grade 9, class 3	0.0508***	(0.0085)
$\mathbf{W}_{n,0}$ × Commute by Car/Taxi	-0.0118	(0.0245)	School 3, grade 9, class 4	0.0668***	(0.0159)
$\mathbf{W}_{n,0}$ × Music	-0.0036	(0.0169)	Constant	4.2484***	(1.2059)
n	868				
F-Statistics	12.9000				
p-value	0.0000				

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: LS Estimations using $\mathbf{W}y$ as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0}$ × Male	0.1277	(0.1521)			0.2244	(0.1505)
$\mathbf{W}_{n,0}$ × ln(Height)	1.1821**	(0.5783)			1.6369	(2.0158)
$\mathbf{W}_{n,0}$ × ln(Weight)	-0.5142	(0.3457)			-0.2787	(0.4065)
$\mathbf{W}_{n,0}$ × Siblings Help	0.1925	(0.1588)			0.1315	(0.1354)
$\mathbf{W}_{n,0}$ × Parents Help	0.1572	(0.1427)			0.2092	(0.1405)
$\mathbf{W}_{n,0}$ × Commute by Car/Taxi	0.0588	(0.1407)			0.0455	(0.1457)
$\mathbf{W}_{n,0}$ × Music	0.2277**	(0.1077)			0.0790	(0.1280)
$\mathbf{W}_{n,0}$ × ln(Cognitive)	-0.1062	(0.1635)			0.3246	(0.2969)
$\mathbf{W}_{n,0}$ × ln(Agreeableness)	0.1843	(0.2616)			-0.0557	(0.4539)
$\mathbf{W}_{n,0}$ × ln(Conscientiousness)	0.6115***	(0.2035)			0.0661	(0.3383)
$\mathbf{W}_{n,0}$ × ln(Extraversion)	0.1699	(0.1471)			0.2083	(0.2455)
$\mathbf{W}_{n,0}$ × ln(Neuroticism)	0.0704	(0.1952)			-0.0816	(0.2820)
$\mathbf{W}_{n,0}$ × ln(Openness)	0.2039	(0.2609)			0.2749	(0.3080)
$\mathbf{W}^2_{n,0}$ × Male			-0.2862	(0.2443)	-0.3681	(0.2556)
$\mathbf{W}^2_{n,0}$ × ln(Height)			1.1908	(0.8774)	-0.5011	(1.9841)
$\mathbf{W}^2_{n,0}$ × ln(Weight)			-0.2088	(0.6695)	0.0118	(0.7225)
$\mathbf{W}^2_{n,0}$ × Siblings Help			0.5321***	(0.1883)	0.5106***	(0.1875)
$\mathbf{W}^2_{n,0}$ × Parents Help			0.0411	(0.2204)	0.0127	(0.2210)
$\mathbf{W}^2_{n,0}$ × Commute by Car/Taxi			-0.1939	(0.2112)	-0.1865	(0.2059)
$\mathbf{W}^2_{n,0}$ × Music			0.4272***	(0.1628)	0.4009**	(0.1825)
$\mathbf{W}^2_{n,0}$ × ln(Cognitive)			-0.3517	(0.2263)	-0.7223	(0.4509)
$\mathbf{W}^2_{n,0}$ × ln(Agreeableness)			0.3012	(0.4596)	0.4606	(0.7711)
$\mathbf{W}^2_{n,0}$ × ln(Conscientiousness)			1.1246***	(0.2488)	1.0246**	(0.4803)
$\mathbf{W}^2_{n,0}$ × ln(Extraversion)			0.1388	(0.3130)	-0.1316	(0.5399)
$\mathbf{W}^2_{n,0}$ × ln(Neuroticism)			0.0915	(0.3264)	0.1898	(0.5169)
$\mathbf{W}^2_{n,0}$ × ln(Openness)			0.1003	(0.4454)	-0.2281	(0.5347)
Constant	-0.7597	(1.8614)	-1.8886	(2.8393)	-1.6844	(2.9717)
n	868		868		868	
F-Statistics	4.1245		9.2074		12.3000	
p-value	0.0013		0.0000		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: LS Estimations using $\mathbf{W} \times$ Male as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0} \times$ ln(Height)	-0.1938*	(0.1152)			-0.8460*	(0.4750)
$\mathbf{W}_{n,0} \times$ ln(Weight)	0.1995**	(0.0950)			0.1754*	(0.0922)
$\mathbf{W}_{n,0} \times$ Siblings Help	0.0570	(0.0516)			0.0652	(0.0486)
$\mathbf{W}_{n,0} \times$ Parents Help	-0.0086	(0.0494)			-0.0142	(0.0496)
$\mathbf{W}_{n,0} \times$ Commute by Car/Taxi	-0.0495	(0.0511)			-0.0470	(0.0510)
$\mathbf{W}_{n,0} \times$ Music	-0.0159	(0.0496)			0.0092	(0.0491)
$\mathbf{W}_{n,0} \times$ ln(Cognitive)	0.2337***	(0.0448)			0.1023	(0.0700)
$\mathbf{W}_{n,0} \times$ ln(Agreeableness)	-0.1969***	(0.0622)			0.1203	(0.1030)
$\mathbf{W}_{n,0} \times$ ln(Conscientiousness)	0.1031	(0.0627)			0.0781	(0.1071)
$\mathbf{W}_{n,0} \times$ ln(Extraversion)	0.0437	(0.0367)			0.0723	(0.0714)
$\mathbf{W}_{n,0} \times$ ln(Neuroticism)	-0.1813***	(0.0502)			-0.0531	(0.0784)
$\mathbf{W}_{n,0} \times$ ln(Openness)	-0.0243	(0.0876)			-0.0344	(0.0896)
$\mathbf{W}^2_{n,0} \times$ ln(Height)			-0.3383*	(0.1797)	0.3152	(0.4610)
$\mathbf{W}^2_{n,0} \times$ ln(Weight)			0.2465	(0.1653)	0.2330	(0.1827)
$\mathbf{W}^2_{n,0} \times$ Siblings Help			0.0158	(0.0582)	0.0296	(0.0621)
$\mathbf{W}^2_{n,0} \times$ Parents Help			0.0624	(0.0618)	0.0802	(0.0579)
$\mathbf{W}^2_{n,0} \times$ Commute by Car/Taxi			-0.0127	(0.0620)	-0.0032	(0.0651)
$\mathbf{W}^2_{n,0} \times$ Music			-0.0940*	(0.0505)	-0.0906*	(0.0483)
$\mathbf{W}^2_{n,0} \times$ ln(Cognitive)			0.4096***	(0.0722)	0.2789**	(0.1191)
$\mathbf{W}^2_{n,0} \times$ ln(Agreeableness)			-0.5443***	(0.1188)	-0.7011***	(0.1642)
$\mathbf{W}^2_{n,0} \times$ ln(Conscientiousness)			0.2130**	(0.0943)	0.1059	(0.1780)
$\mathbf{W}^2_{n,0} \times$ ln(Extraversion)			0.0224	(0.0805)	-0.0703	(0.1309)
$\mathbf{W}^2_{n,0} \times$ ln(Neuroticism)			-0.2919***	(0.0801)	-0.2427*	(0.1313)
$\mathbf{W}^2_{n,0} \times$ ln(Openness)			0.0448	(0.1575)	0.0866	(0.1824)
Constant	0.5878	(0.4326)	1.1516*	(0.6258)	1.4789**	(0.6419)
n	868		868		868	
F-Statistics	10.9248		19.3011		126.4900	
p-value	0.0000		0.0000		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: LS Estimations using $\mathbf{W} \times \ln(\text{Height})$ as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0} \times$ Male	0.1180	(0.1872)			0.2803	(0.1927)
$\mathbf{W}_{n,0} \times$ ln(Weight)	0.2625	(0.1916)			0.0643	(0.3955)
$\mathbf{W}_{n,0} \times$ Siblings Help	0.2883	(0.1941)			0.1956	(0.1622)
$\mathbf{W}_{n,0} \times$ Parents Help	0.1967	(0.1748)			0.2461	(0.1683)
$\mathbf{W}_{n,0} \times$ Commute by Car/Taxi	0.1375	(0.1625)			0.1052	(0.1700)
$\mathbf{W}_{n,0} \times$ Music	0.2489*	(0.1306)			0.0600	(0.1596)
$\mathbf{W}_{n,0} \times$ ln(Cognitive)	-0.2424	(0.1828)			0.4494	(0.3957)
$\mathbf{W}_{n,0} \times$ ln(Agreeableness)	-0.1306	(0.2771)			-0.1050	(0.5708)
$\mathbf{W}_{n,0} \times$ ln(Conscientiousness)	0.5010**	(0.2017)			0.0665	(0.4104)
$\mathbf{W}_{n,0} \times$ ln(Extraversion)	0.0614	(0.1853)			0.2481	(0.2555)
$\mathbf{W}_{n,0} \times$ ln(Neuroticism)	-0.1489	(0.1897)			-0.0746	(0.3510)
$\mathbf{W}_{n,0} \times$ ln(Openness)	0.1441	(0.3229)			0.3960	(0.3747)
$\mathbf{W}^2_{n,0} \times$ Male			-0.4456	(0.3082)	-0.5613*	(0.3149)
$\mathbf{W}^2_{n,0} \times$ ln(Weight)			0.6125	(0.5231)	0.5680	(0.6976)
$\mathbf{W}^2_{n,0} \times$ Siblings Help			0.5837**	(0.2417)	0.5906**	(0.2335)
$\mathbf{W}^2_{n,0} \times$ Parents Help			0.0234	(0.2432)	-0.0045	(0.2465)
$\mathbf{W}^2_{n,0} \times$ Commute by Car/Taxi			-0.2487	(0.2473)	-0.2697	(0.2421)
$\mathbf{W}^2_{n,0} \times$ Music			0.5045***	(0.1665)	0.4822**	(0.1975)
$\mathbf{W}^2_{n,0} \times$ ln(Cognitive)			-0.5994**	(0.2816)	-1.1088*	(0.5745)
$\mathbf{W}^2_{n,0} \times$ ln(Agreeableness)			0.0437	(0.4733)	0.3197	(0.8356)
$\mathbf{W}^2_{n,0} \times$ ln(Conscientiousness)			1.0995***	(0.2510)	1.0402*	(0.5467)
$\mathbf{W}^2_{n,0} \times$ ln(Extraversion)			-0.0054	(0.3580)	-0.2854	(0.5586)
$\mathbf{W}^2_{n,0} \times$ ln(Neuroticism)			-0.1436	(0.3153)	-0.0041	(0.5480)
$\mathbf{W}^2_{n,0} \times$ ln(Openness)			-0.0027	(0.5465)	-0.4294	(0.6521)
Constant	2.8956***	(0.8051)	1.8026	(1.9623)	1.3428	(1.7616)
n	868		868		868	
F-Statistics	2.8743		8.3756		267.8900	
p-value	0.0134		0.0000		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: LS Estimations using $\mathbf{W} \times \ln(\text{Weight})$ as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0} \times$ Male	0.0836	(0.1408)			0.1835	(0.1335)
$\mathbf{W}_{n,0} \times$ ln(Height)	0.7174**	(0.3325)			1.2811	(1.4793)
$\mathbf{W}_{n,0} \times$ Siblings Help	0.1808	(0.1492)			0.1313	(0.1349)
$\mathbf{W}_{n,0} \times$ Parents Help	0.1202	(0.1302)			0.1585	(0.1339)
$\mathbf{W}_{n,0} \times$ Commute by Car/Taxi	0.0749	(0.1223)			0.0626	(0.1324)
$\mathbf{W}_{n,0} \times$ Music	0.1526	(0.0957)			0.0207	(0.1181)
$\mathbf{W}_{n,0} \times$ ln(Cognitive)	-0.1177	(0.1470)			0.2401	(0.2855)
$\mathbf{W}_{n,0} \times$ ln(Agreeableness)	0.1349	(0.2465)			-0.1997	(0.4257)
$\mathbf{W}_{n,0} \times$ ln(Conscientiousness)	0.5276***	(0.1917)			-0.0289	(0.3189)
$\mathbf{W}_{n,0} \times$ ln(Extraversion)	0.1552	(0.1399)			0.1207	(0.2145)
$\mathbf{W}_{n,0} \times$ ln(Neuroticism)	0.0531	(0.1750)			-0.1283	(0.2683)
$\mathbf{W}_{n,0} \times$ ln(Openness)	0.2478	(0.2398)			0.3701	(0.2943)
$\mathbf{W}^2_{n,0} \times$ Male			-0.2662	(0.2209)	-0.3274	(0.2211)
$\mathbf{W}^2_{n,0} \times$ ln(Height)			1.1151**	(0.5044)	-0.2210	(1.6553)
$\mathbf{W}^2_{n,0} \times$ Siblings Help			0.3890**	(0.1887)	0.3742**	(0.1821)
$\mathbf{W}^2_{n,0} \times$ Parents Help			-0.0328	(0.1957)	-0.0597	(0.1952)
$\mathbf{W}^2_{n,0} \times$ Commute by Car/Taxi			-0.2155	(0.1877)	-0.2338	(0.1884)
$\mathbf{W}^2_{n,0} \times$ Music			0.3344**	(0.1344)	0.3274**	(0.1590)
$\mathbf{W}^2_{n,0} \times$ ln(Cognitive)			-0.3279	(0.2107)	-0.6054	(0.4380)
$\mathbf{W}^2_{n,0} \times$ ln(Agreeableness)			0.3474	(0.4004)	0.6676	(0.6866)
$\mathbf{W}^2_{n,0} \times$ ln(Conscientiousness)			1.0425***	(0.2321)	1.0599**	(0.4446)
$\mathbf{W}^2_{n,0} \times$ ln(Extraversion)			0.1596	(0.2918)	0.0006	(0.4765)
$\mathbf{W}^2_{n,0} \times$ ln(Neuroticism)			0.1233	(0.2948)	0.2734	(0.4803)
$\mathbf{W}^2_{n,0} \times$ ln(Openness)			0.1558	(0.4027)	-0.2917	(0.4938)
Constant	-0.5989	(1.6537)	-2.4319	(2.4581)	-2.3797	(2.4980)
n	868		868		868	
F-Statistics	3.2110		7.3745		28.3500	
p-value	0.0072		0.0000		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: LS Estimations using $\mathbf{W} \times$ Siblings Help as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0} \times$ Male	-0.0436	(0.0379)			-0.0275	(0.0345)
$\mathbf{W}_{n,0} \times$ ln(Height)	-0.0999	(0.0919)			-0.6092	(0.3855)
$\mathbf{W}_{n,0} \times$ ln(Weight)	0.0392	(0.0863)			0.1094	(0.0930)
$\mathbf{W}_{n,0} \times$ Parents Help	0.0451	(0.0332)			0.0330	(0.0301)
$\mathbf{W}_{n,0} \times$ Commute by Car/Taxi	0.0639*	(0.0326)			0.0656*	(0.0372)
$\mathbf{W}_{n,0} \times$ Music	0.0158	(0.0352)			-0.0173	(0.0335)
$\mathbf{W}_{n,0} \times$ ln(Cognitive)	-0.0248	(0.0345)			-0.0716	(0.0642)
$\mathbf{W}_{n,0} \times$ ln(Agreeableness)	-0.0670	(0.0598)			-0.0254	(0.0959)
$\mathbf{W}_{n,0} \times$ ln(Conscientiousness)	0.0854**	(0.0399)			0.0177	(0.0724)
$\mathbf{W}_{n,0} \times$ ln(Extraversion)	-0.0416	(0.0547)			0.0135	(0.0898)
$\mathbf{W}_{n,0} \times$ ln(Neuroticism)	-0.0215	(0.0378)			-0.0045	(0.0657)
$\mathbf{W}_{n,0} \times$ ln(Openness)	-0.0335	(0.0614)			-0.0775	(0.0799)
$\mathbf{W}^2_{n,0} \times$ Male			-0.1188**	(0.0597)	-0.1127*	(0.0644)
$\mathbf{W}^2_{n,0} \times$ ln(Height)			0.0725	(0.1319)	0.6017	(0.4002)
$\mathbf{W}^2_{n,0} \times$ ln(Weight)			-0.1291	(0.0842)	-0.1695*	(0.1014)
$\mathbf{W}^2_{n,0} \times$ Parents Help			0.0610	(0.0602)	0.0652	(0.0592)
$\mathbf{W}^2_{n,0} \times$ Commute by Car/Taxi			0.0416	(0.0510)	0.0415	(0.0519)
$\mathbf{W}^2_{n,0} \times$ Music			0.0947**	(0.0417)	0.1039***	(0.0402)
$\mathbf{W}^2_{n,0} \times$ ln(Cognitive)			0.0496	(0.0587)	0.1322	(0.1107)
$\mathbf{W}^2_{n,0} \times$ ln(Agreeableness)			-0.1213	(0.1071)	-0.1176	(0.1763)
$\mathbf{W}^2_{n,0} \times$ ln(Conscientiousness)			0.1726***	(0.0634)	0.1475	(0.1139)
$\mathbf{W}^2_{n,0} \times$ ln(Extraversion)			-0.0354	(0.0873)	-0.0551	(0.1422)
$\mathbf{W}^2_{n,0} \times$ ln(Neuroticism)			-0.0623	(0.0651)	-0.0665	(0.1185)
$\mathbf{W}^2_{n,0} \times$ ln(Openness)			-0.0080	(0.1083)	0.0776	(0.1576)
Constant	0.7144**	(0.3385)	0.4611	(0.5473)	0.5524	(0.6007)
n	868		868		868	
F-Statistics	1.7506		4.3328		10.4200	
p-value	0.1174		0.0011		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 17: LS Estimations using $\mathbf{W} \times$ Parents Help as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0} \times$ Male	0.0252	(0.0534)			0.0333	(0.0564)
$\mathbf{W}_{n,0} \times$ ln(Height)	0.2073	(0.1268)			0.5553	(0.4865)
$\mathbf{W}_{n,0} \times$ ln(Weight)	-0.1920**	(0.0948)			-0.1684*	(0.0909)
$\mathbf{W}_{n,0} \times$ Siblings Help	0.0617	(0.0435)			0.0520	(0.0438)
$\mathbf{W}_{n,0} \times$ Commute by Car/Taxi	0.0236	(0.0373)			0.0162	(0.0412)
$\mathbf{W}_{n,0} \times$ Music	0.0308	(0.0399)			-0.0005	(0.0379)
$\mathbf{W}_{n,0} \times$ ln(Cognitive)	0.0143	(0.0390)			0.0399	(0.0814)
$\mathbf{W}_{n,0} \times$ ln(Agreeableness)	-0.0304	(0.0722)			-0.0502	(0.1190)
$\mathbf{W}_{n,0} \times$ ln(Conscientiousness)	0.1347**	(0.0646)			0.0392	(0.0911)
$\mathbf{W}_{n,0} \times$ ln(Extraversion)	0.0377	(0.0427)			0.1117	(0.0750)
$\mathbf{W}_{n,0} \times$ ln(Neuroticism)	-0.0500	(0.0554)			-0.1029	(0.0751)
$\mathbf{W}_{n,0} \times$ ln(Openness)	0.0034	(0.0716)			-0.0387	(0.0908)
$\mathbf{W}^2_{n,0} \times$ Male			0.0146	(0.0651)	-0.0097	(0.0735)
$\mathbf{W}^2_{n,0} \times$ ln(Height)			0.0980	(0.1905)	-0.3670	(0.5233)
$\mathbf{W}^2_{n,0} \times$ ln(Weight)			-0.0499	(0.1486)	0.0217	(0.1424)
$\mathbf{W}^2_{n,0} \times$ Siblings Help			0.1268***	(0.0433)	0.0972**	(0.0477)
$\mathbf{W}^2_{n,0} \times$ Commute by Car/Taxi			0.0172	(0.0634)	0.0378	(0.0602)
$\mathbf{W}^2_{n,0} \times$ Music			0.0976**	(0.0454)	0.0989**	(0.0454)
$\mathbf{W}^2_{n,0} \times$ ln(Cognitive)			-0.0012	(0.0618)	-0.0401	(0.1282)
$\mathbf{W}^2_{n,0} \times$ ln(Agreeableness)			-0.0425	(0.1254)	0.0448	(0.1980)
$\mathbf{W}^2_{n,0} \times$ ln(Conscientiousness)			0.2338**	(0.1002)	0.1852	(0.1538)
$\mathbf{W}^2_{n,0} \times$ ln(Extraversion)			-0.0281	(0.0860)	-0.1573	(0.1540)
$\mathbf{W}^2_{n,0} \times$ ln(Neuroticism)			-0.0400	(0.0898)	0.0910	(0.1330)
$\mathbf{W}^2_{n,0} \times$ ln(Openness)			0.0027	(0.1347)	0.0405	(0.1839)
Constant	0.2101	(0.4627)	0.1457	(0.6395)	0.0231	(0.6233)
n	868		868		868	
F-Statistics	3.0899		1.8041		25.4400	
p-value	0.0090		0.1057		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18: LS Estimations using $\mathbf{W} \times$ Commute by Car/Taxi as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0} \times$ Male	0.0167	(0.0482)			0.0494	(0.0512)
$\mathbf{W}_{n,0} \times$ ln(Height)	0.1140	(0.1299)			-0.1766	(0.3776)
$\mathbf{W}_{n,0} \times$ ln(Weight)	0.0206	(0.0859)			0.1109	(0.0949)
$\mathbf{W}_{n,0} \times$ Siblings Help	0.0464	(0.0444)			0.0344	(0.0441)
$\mathbf{W}_{n,0} \times$ Parents Help	0.0028	(0.0290)			0.0043	(0.0314)
$\mathbf{W}_{n,0} \times$ Music	0.0211	(0.0360)			-0.0033	(0.0404)
$\mathbf{W}_{n,0} \times$ ln(Cognitive)	-0.0589	(0.0596)			0.0425	(0.0971)
$\mathbf{W}_{n,0} \times$ ln(Agreeableness)	0.0527	(0.0563)			-0.0520	(0.1007)
$\mathbf{W}_{n,0} \times$ ln(Conscientiousness)	0.1208***	(0.0422)			-0.0152	(0.0892)
$\mathbf{W}_{n,0} \times$ ln(Extraversion)	0.0581	(0.0519)			-0.0253	(0.0891)
$\mathbf{W}_{n,0} \times$ ln(Neuroticism)	0.0323	(0.0419)			-0.0002	(0.0748)
$\mathbf{W}_{n,0} \times$ ln(Openness)	-0.0344	(0.0811)			0.0570	(0.1415)
$\mathbf{W}^2_{n,0} \times$ Male			-0.1128*	(0.0684)	-0.1296**	(0.0587)
$\mathbf{W}^2_{n,0} \times$ ln(Height)			0.3602*	(0.2035)	0.4488	(0.4245)
$\mathbf{W}^2_{n,0} \times$ ln(Weight)			-0.1240	(0.1436)	-0.1462	(0.1429)
$\mathbf{W}^2_{n,0} \times$ Siblings Help			0.1113***	(0.0398)	0.1211***	(0.0415)
$\mathbf{W}^2_{n,0} \times$ Parents Help			0.0816	(0.0527)	0.0838	(0.0510)
$\mathbf{W}^2_{n,0} \times$ Music			0.0342	(0.0376)	0.0350	(0.0381)
$\mathbf{W}^2_{n,0} \times$ ln(Cognitive)			-0.0991	(0.0765)	-0.1547	(0.1215)
$\mathbf{W}^2_{n,0} \times$ ln(Agreeableness)			0.1499	(0.1091)	0.2124	(0.1820)
$\mathbf{W}^2_{n,0} \times$ ln(Conscientiousness)			0.2543***	(0.0743)	0.2695*	(0.1472)
$\mathbf{W}^2_{n,0} \times$ ln(Extraversion)			0.1712***	(0.0663)	0.1959*	(0.1155)
$\mathbf{W}^2_{n,0} \times$ ln(Neuroticism)			0.0603	(0.0730)	0.0494	(0.1346)
$\mathbf{W}^2_{n,0} \times$ ln(Openness)			-0.1452	(0.0956)	-0.2129	(0.1814)
Constant	-0.1795	(0.4529)	-0.9086	(0.6702)	-0.8397	(0.6981)
n	868		868		868	
F-Statistics	2.8197		6.3966		21.4400	
p-value	0.0148		0.0001		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 19: LS Estimations using $\mathbf{W} \times$ Music as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,0} \times$ Male	-0.0666	(0.0557)			-0.0043	(0.0557)
$\mathbf{W}_{n,0} \times$ ln(Height)	0.3127**	(0.1515)			-0.1204	(0.5234)
$\mathbf{W}_{n,0} \times$ ln(Weight)	-0.3884***	(0.1083)			-0.2920***	(0.1108)
$\mathbf{W}_{n,0} \times$ Siblings Help	0.0005	(0.0389)			-0.0089	(0.0356)
$\mathbf{W}_{n,0} \times$ Parents Help	0.0761	(0.0708)			0.0791	(0.0626)
$\mathbf{W}_{n,0} \times$ Commute by Car/Taxi	0.0057	(0.0499)			-0.0038	(0.0485)
$\mathbf{W}_{n,0} \times$ ln(Cognitive)	-0.1652***	(0.0637)			-0.0564	(0.0934)
$\mathbf{W}_{n,0} \times$ ln(Agreeableness)	0.0980	(0.0797)			0.0883	(0.1335)
$\mathbf{W}_{n,0} \times$ ln(Conscientiousness)	0.0190	(0.0478)			-0.0632	(0.0964)
$\mathbf{W}_{n,0} \times$ ln(Extraversion)	-0.0369	(0.0520)			-0.1098	(0.0681)
$\mathbf{W}_{n,0} \times$ ln(Neuroticism)	0.0477	(0.0569)			-0.0444	(0.0774)
$\mathbf{W}_{n,0} \times$ ln(Openness)	-0.0273	(0.0824)			-0.0120	(0.1017)
$\mathbf{W}^2_{n,0} \times$ Male			-0.2865***	(0.0669)	-0.2900***	(0.0804)
$\mathbf{W}^2_{n,0} \times$ ln(Height)			0.3415	(0.2258)	0.6147	(0.5420)
$\mathbf{W}^2_{n,0} \times$ ln(Weight)			-0.2853*	(0.1620)	-0.1608	(0.2271)
$\mathbf{W}^2_{n,0} \times$ Siblings Help			0.1136*	(0.0687)	0.0990**	(0.0445)
$\mathbf{W}^2_{n,0} \times$ Parents Help			0.0645	(0.0680)	0.0614	(0.0753)
$\mathbf{W}^2_{n,0} \times$ Commute by Car/Taxi			-0.0844	(0.0652)	-0.0823	(0.0755)
$\mathbf{W}^2_{n,0} \times$ ln(Cognitive)			-0.2039**	(0.0905)	-0.1334	(0.1251)
$\mathbf{W}^2_{n,0} \times$ ln(Agreeableness)			0.0838	(0.1486)	-0.0304	(0.2709)
$\mathbf{W}^2_{n,0} \times$ ln(Conscientiousness)			0.0887	(0.1102)	0.1801	(0.1672)
$\mathbf{W}^2_{n,0} \times$ ln(Extraversion)			0.0247	(0.1097)	0.1649	(0.1593)
$\mathbf{W}^2_{n,0} \times$ ln(Neuroticism)			0.0749	(0.0914)	0.1379	(0.1367)
$\mathbf{W}^2_{n,0} \times$ ln(Openness)			-0.0119	(0.1453)	-0.0084	(0.2091)
Constant	0.5647	(0.5552)	0.1232	(0.8391)	-0.0420	(0.8874)
n	868		868		868	
F-Statistics	3.7731		3.65		13.4900	
p-value	0.0027		0.0000		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 20: LS Estimations using $\mathbf{W}y$ as dependent variable

	(1)		(2)		(3)	
Variables	Coef.	SE	Coef.	SE	Coef.	SE
$\mathbf{W}_{n,*} \times$ Male	-0.0504	(0.0487)			-0.0496	(0.0413)
$\mathbf{W}_{n,*} \times$ ln(Height)	0.8742***	(0.2639)			0.3636**	(0.1728)
$\mathbf{W}_{n,*} \times$ ln(Weight)	-0.1477	(0.1316)			0.0612	(0.0988)
$\mathbf{W}_{n,*} \times$ Siblings Help	0.0403	(0.0444)			0.0544	(0.0394)
$\mathbf{W}_{n,*} \times$ Parents Help	0.0065	(0.0348)			-0.0557	(0.0400)
$\mathbf{W}_{n,*} \times$ Commute Car/Taxi	0.0122	(0.0539)			-0.0089	(0.0474)
$\mathbf{W}_{*} \times$ Music	0.0651	(0.0436)			-0.0012	(0.0365)
$\mathbf{W}_{n,*} \times$ ln(Cognitive)	-0.2676**	(0.1081)			-0.2059***	(0.0729)
$\mathbf{W}_{n,*} \times$ ln(Agreeableness)	-0.0154	(0.2255)			-0.2280*	(0.1287)
$\mathbf{W}_{n,*} \times$ ln(Conscientious)	0.3370**	(0.1402)			0.1333*	(0.0731)
$\mathbf{W}_{n,*} \times$ ln(Extraversion)	0.0306	(0.1665)			0.1601*	(0.0826)
$\mathbf{W}_{n,*} \times$ ln(Neuroticism)	-0.0098	(0.1254)			0.0122	(0.0625)
$\mathbf{W}_{n,*} \times$ ln(Openness)	0.2343	(0.2068)			-0.0649	(0.1187)
$\mathbf{W}^2_{n,*} \times$ Male			-0.2182	(0.1743)	-0.0939	(0.1271)
$\mathbf{W}^2_{n,*} \times$ ln(Height)			0.9572*	(0.4962)	0.5604	(0.4129)
$\mathbf{W}^2_{n,*} \times$ ln(Weight)			-0.0553	(0.4378)	-0.1974	(0.2979)
$\mathbf{W}^2_{n,*} \times$ Siblings Help			0.2420	(0.2819)	0.2458	(0.1603)
$\mathbf{W}^2_{n,*} \times$ Parents Help			0.0493	(0.1802)	0.0326	(0.0824)
$\mathbf{W}^2_{n,*} \times$ Commute Car/Taxi			0.5050***	(0.1750)	0.0522	(0.1192)
$\mathbf{W}^2_{n,*} \times$ Music			-0.1076	(0.0977)	0.1880***	(0.0484)
$\mathbf{W}^2_{n,*} \times$ ln(Cognitive)			-0.3588**	(0.1464)	-0.0596	(0.1343)
$\mathbf{W}^2_{n,*} \times$ ln(Agreeableness)			0.0570	(0.2684)	0.2458	(0.2486)
$\mathbf{W}^2_{n,*} \times$ ln(Conscientious)			0.4525**	(0.2023)	0.2621*	(0.1535)
$\mathbf{W}^2_{n,*} \times$ ln(Extraversion)			-0.0010	(0.1893)	-0.1465	(0.1558)
$\mathbf{W}^2_{n,*} \times$ ln(Neuroticism)			0.0569	(0.1904)	-0.0245	(0.1524)
$\mathbf{W}^2_{n,*} \times$ ln(Openness)			0.4369	(0.2845)	0.3608*	(0.1930)
Constant	0.2689	(1.1172)	-0.3995	(1.6556)	-0.1877	(1.4016)
n	868		868		868	
F-Statistics	20.2968		22.4297		40.1600	
p-value	0.0000		0.0000		0.0000	

Note: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

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.

Kojevnikov, Denis, Vadim Marmer, and Kyungchul Song. 2021. "Limit Theorems for Network Dependent Random Variables." *Journal of Econometrics* 222 (2):882–908.