

Measurement Error and Peer Effects in Networks

YANN BRAMOULLÉ

SEBASTIAAN MAES

Aix-Marseille School of Economics

Department of Economics

Aix-Marseille University & CNRS

University of Antwerp

yann.bramouille@univ-amu.fr

sebastiaan.maes@uantwerpen.be

January 31, 2026

Abstract

In many practical applications, only noisy proxies for the true regressors are available, which is commonly believed to induce an attenuation bias. In the linear-in-means model, however, estimated peer effects might be inflated, potentially leading to false positives. This paper shows that the asymptotic bias depends on the interplay between individual characteristics and network links and demonstrates how the network structure can facilitate identification without the need for additional external information. Based on these identification results, we present consistent GMM and 2SLS estimators that are easily implementable. Our results are illustrated by means of a Monte Carlo simulation.

Keywords: peer effects, social interactions, linear-in-means model, measurement error, errors-in-variables, asymptotic bias

JEL codes: C31, C36

*Yann acknowledges financial support from the French government under the “France 2030” investment plan managed by the French National Research Agency (ANR-17-EURE-0020) and from the Excellence Initiative of Aix-Marseille University - A*MIDEX. Sebastiaan benefited from doctoral and postdoctoral fellowships from the Research Foundation Flanders (grants 11F8919N and 12C8623N). We thank the editor, Áureo de Paula, an anonymous associate editor, and two anonymous reviewers for useful comments and suggestions. We also thank Vincent Boucher, Nathan Canen, Sam Cosaert, Geert Dhaene, Marcel Fafchamps, Jan Feld, Bernard Fortin, Alan Griffith, as well as conference and seminar participants in Antwerp, Barcelona (ESEM), Manchester (EWMES), Orléans (IPDC), Oslo (IAAE), Québec, and Rome. Part of this research was conducted during visits to Aix-Marseille School of Economics, whose hospitality is gratefully acknowledged.

1 Introduction

How reliable are peer effect estimates in published studies? Not very, according to Josh Angrist’s critical assessment of the literature: “although correlation among peers is a reliable descriptive fact, the scope for incorrect or misleading attributions of causality in peer analysis is extraordinarily wide” ([Angrist, 2014](#), p. 98). Obtaining credible causal estimates of peer effects presents several challenges to researchers. To address peer endogeneity, they must find, or generate, exogenous variations in peer groups and develop models of peer selection.¹ To address mismeasurement in peers, they must collect detailed data on social networks and develop models of peer effects with unknown or mismeasured peers.² Building on [Manski \(1993\)](#), the methodological literature on peer effects has grown alongside the applied literature. Yet despite methodological advances, guaranteeing the reliability of peer effect estimates remains challenging. Addressing multiple endogeneity issues without resolving them all still leaves researchers some distance away from a causal interpretation.

This paper focuses on a critical yet understudied issue affecting the reliability of peer effect estimates: measurement error in individual characteristics. Such errors are a first-order empirical issue in survey-based ([Bound, Brown, & Mathiowetz, 2001](#)) and experimental work ([Gillen, Snowberg, & Yariv, 2019](#)). Errors-in-variables are especially salient when individual-level covariates are measured with noise due to survey misreporting, recall error, or imperfect measurement. In particular, self-reported income and assets, as well as noisy proxies for ability such as test scores, are commonly subject to classical measurement error.

Surprisingly, however, errors-in-variables remain a blind spot in the existing literature. In the applied literature on peer effects, problems raised by measurement error on covariates are almost never discussed or addressed. The methodological literature on this problem is scarce, and will be comprehensively reviewed below. In short, this issue was identified by [Moffitt \(2001\)](#), highlighted by [Angrist \(2014\)](#), and studied by [Ammermueller and Pischke \(2009\)](#); [de Paula \(2017\)](#); [Feld and Zölitz \(2017\)](#). These five papers, however, only consider group interactions: agents are partitioned in groups, such as classes within schools, and are affected by everyone in their group and by no one outside it. In contrast, many recent studies of peer

¹For studies of peer effects with random peers see, e.g., [Sacerdote \(2001\)](#); [Carrell, Sacerdote, and West \(2013\)](#); [Corno, La Ferrara, and Burns \(2022\)](#). For econometric approaches combining models of peer effects in networks with models of network formation see, e.g., [Goldsmith-Pinkham and Imbens \(2013\)](#); [Hsieh and Lee \(2016\)](#); [Griffith \(2022b\)](#).

²The literature on peer effects in networks has grown fast in the past fifteen years, see [Bramoullé, Djebbari, and Fortin \(2020\)](#) for a review. See, e.g., [Griffith \(2022a\)](#); [Boucher and Houndetoungan \(2025\)](#); [Lewbel, Qu, and Tang \(2023\)](#); [de Paula, Rasul, and Souza \(2024\)](#) for models of peer effects with imperfectly known peers.

effects consider richer network interactions. Our paper fills this gap by analyzing measurement error in network-based peer effects.

We provide the first analysis of measurement error when peer effects operate on a network. We consider the benchmark linear-in-means model with contextual and endogenous peer effects. An individual's outcome is affected by their individual characteristic, by the average characteristic of their peers, and by the average outcome of their peers. A significant challenge arises when researchers only observe a noisy proxy for the characteristic. Measurement error then occurs in both individual and average peer characteristic, and errors in the two variables are related through the structure of the model. The presence of related measurement errors on two regressors makes the problem non-standard. We adopt a many-networks asymptotic framework and consider econometric specifications with or without network fixed effects. Our investigation revolves around two central questions. First, under what conditions does measurement error cause asymptotic biases in peer effect estimates? Second, what strategies can researchers employ to mitigate these biases? Our identification and estimation results highlight that true peer effects can generally be recovered without additional data.

Consistent with existing research, our findings indicate that classical measurement error typically induces asymptotic bias in peer-effect estimates. Defying conventional wisdom, these estimates are not necessarily attenuated toward zero.³ We show that naive 2SLS estimates of peer effects, using mismeasured characteristics of peers at distance two as an instrument for the average outcome of peers, converge to a linear combination of the model's parameters, which we refer to as *smearing* or *weight shifting*. The magnitude and direction of the resulting asymptotic biases critically depend on the interplay between individual characteristics and network links. With measurement error, the estimated endogenous peer effect absorbs parts of the individual effect and of the contextual peer effect, because the instrument spuriously picks up variation in the true individual and average peer characteristic. Similarly, the estimated contextual peer effect absorbs part of the individual effect, because the average mismeasured peer characteristic partially proxies for the true individual characteristic. As a consequence, peer effect estimates may be significantly different from zero even in the absence of true peer effects. These cross-loadings vanish when characteristics and links are independent. In that case, the estimate of the endogenous peer effect is asymptotically unbiased, while estimates of individual and contextual peer effects display an asymptotic attenuation bias towards zero.

To build intuition we also study the bias in two simplified settings: a model with

³This conventional wisdom is based on the analysis of OLS estimates in the presence of one mismeasured variable. It is well known that little can be said, in general, about the nature of the asymptotic bias in OLS or 2SLS estimates if there is measurement error in multiple variables (see, e.g., [Levi, 1973](#); [Greene, 2003](#), p. 86).

contextual peer effects only and a model with endogenous peer effects only. In both cases, smearing only arises when an individual’s characteristic is correlated with their average peers’ characteristics. This correlation naturally emerges under *homophily*, which is widely documented in social networks ([McPherson, Smith-Lovin, & Cook, 2001](#)). For example, it is common for college students to form friendships with peers who have comparable academic abilities, similar socioeconomic backgrounds, and similar levels of parental education. Measures of these characteristics in survey data, such as *Add Health*, are notoriously noisy, suggesting a widespread risk of biased peer effect estimates.⁴

In the first part of our analysis, we thus clarify how relationships between characteristics and links give rise to an asymptotic bias in peer effect estimates. In the second part, we demonstrate how these relationships can be leveraged to solve the problem. We show that the econometric model can be identified through conditional mean and covariance restrictions, and without relying on external information.⁵ Identification depends on features of the measurement error. Under a weak assumption that measurement error has mean-zero, we characterize when the model is identified based on conditional mean restrictions (Theorem 1). We derive a necessary and sufficient rank condition for identification, which combines the interaction matrix and the characteristic’s first moments. This strategy exploits potential associations between network positions and the individual characteristic: for instance, when agents with more peers tend to have a higher value for the characteristic. Identification from first moments is generic if the network structure is rich enough (Proposition 2) and holds even when the measurement error is correlated and heteroscedastic.

Imposing more structure on the measurement error opens up more possibilities for identification. In a second step, we consider uncorrelated and homoscedastic measurement errors. We characterize when the model is identified based on conditional covariance restrictions (Theorem 2). This identification strategy takes advantage of variations in correlations across observed characteristics between two network positions. Overall, our results show that except in special cases, it is possible to eliminate the asymptotic bias resulting from measurement error.

To do so in practice, we propose generalized method of moments (GMM) and two-stage least squares (2SLS) estimators that are easy to implement. The GMM

⁴Although homophily disappears under complete randomization, it reappears in common quasi-experimental setups, such as peer randomization within stratified groups. Under complete randomization, inflated estimates might still occur when individuals are drawn from a finite pool, giving rise to the exclusion bias (e.g., see [Caeyers & Fafchamps, 2023](#)).

⁵External information, when available, can of course help better identify the parameters of interest. A standard method to address measurement error is to exploit multiple independent measurements of the noisy variable ([Reiersøl, 1941](#); [Schennach, 2007](#)). This can easily be combined with our internal identification strategies.

approach uses more information and is generally more efficient, albeit introducing non-linearities in the estimation procedure. Moments are directly built from the means and covariance restrictions highlighted in our identification results. In contrast, 2SLS estimation is linear but typically less efficient. We propose a variety of valid instruments, including network lags and network features. Monte Carlo simulations show that even modest measurement error can induce substantial bias in naive 2SLS estimates of peer effects. They also confirm that our proposed estimators are feasible and perform well in finite samples.

Related literature. Following the work of [Manski \(1993\)](#), the applied literature on peer effects has grown extensively.⁶ Peer effects have been studied in a wide variety of settings, ranging from the classroom ([Lavy & Schlosser, 2011](#)), through labor supply ([Nicoletti, Salvanes, & Tominey, 2018](#)), to consumption decisions ([De Giorgi, Frederiksen, & Pistaferri, 2019](#)). When it is suspected that characteristics may suffer from measurement error, our results and the tools proposed here can aid empirical researchers to address this problem and to obtain consistent peer effect estimates. Our analysis also clarifies the impact of randomization, a common strategy to mitigate peer endogeneity, on this bias.

Our analysis advances the sparse literature on measurement error and peer effects. A first strand of research concerns models with contextual peer effects. [Moffitt \(2001\)](#) was the first to show formally that errors-in-variables can give rise to an expansion bias in peer effect estimates. He discusses the type of policy interventions that can help address the problem. [Angrist \(2014\)](#) highlights the role played by measurement error in generating inflated peer effect estimates. He illustrates in Table 3 p.103 how adding noise to individual schooling leads to a large increase in the estimate of average state schooling, in a regression on log wage. Exploiting variation across classes within schools, [Ammermueller and Pischke \(2009\)](#) demonstrate that the inclusion of school fixed effects considerably reduces the magnitude of class peer effect estimates. This discrepancy is attributed to the interplay between errors-in-variables and homophily: when sorting into classes is random but sorting into schools is not, the inclusion of school fixed effects removes the school-level homophily that gives rise to the expansion bias. [Feld and Zölitz \(2017\)](#) show that when assignment to classes is completely random, errors-in-variables only lead to an attenuation bias. For the same setting, [Feld and Zölitz \(2022\)](#) propose a simple bias correction procedure based on multiple noisy measurements of the true characteristic.

Another strand of research concerns models with endogenous peer effects and

⁶For a theoretical and econometric discussion on linear social interaction models see, e.g., [Blume, Brock, Durlauf, and Jayaraman \(2015\)](#).

noisy outcomes. [de Paula \(2017, p. 310\)](#) shows that the covariance between peers' mismeasured outcomes identifies the endogenous peer effect in a linear-in-means model. The result operates under the assumption of homoscedastic and uncorrelated disturbances in the outcome equation. In the context of a game with misclassified binary actions, [Lin and Hu \(2024\)](#) develop a consistent estimator based on repeated measurements.

Taking a broader perspective, we also contribute to the literature on measurement error in dependent data. In a seminal contribution, [Griliches and Hausman \(1986\)](#) show how the errors-in-variables problem can be overcome in the standard linear panel data model without resorting to outside information. Our use of network-lagged characteristics as instruments resembles their use of time-lagged variables. More recently, [Evdokimov and Zeleneev \(2020\)](#) study errors-in-variables in general nonlinear semiparametric panel or network data models with fixed effects. In this more general setting, they show how the lagged values of covariates can still serve as instruments to overcome the bias. However, they assume that the variances of the measurement errors shrink with sample size, which is rather restrictive in our context.

Our work is also related to the literature on errors-in-variables in linear models through its use of higher moments. In early contributions, [Koopmans \(1937\)](#) and [Reiersøl \(1950\)](#) recognized that this approach fails if the observables are jointly normal distributed. [Cragg \(1997\)](#), [Dagenais and Dagenais \(1997\)](#), and [Erickson and Whited \(2002\)](#) therefore impose rank conditions on third and higher moments to ensure identification. [Klepper and Leamer \(1984\)](#) show that the first and second moments can be used to bound the coefficients. More recently, [Ben-Moshe \(2021\)](#) provides necessary and sufficient conditions for identification when there is measurement error in all variables. Alternatively, if some variables are known to be perfectly measured, the latter can be used to construct instruments. [Lewbel \(1997, 2012\)](#) and [Ben-Moshe, D'Haultfœuille, and Lewbel \(2017\)](#) construct valid instruments from perfectly measured variables without using additional outside information. Our approach differs from this literature in that we do not impose functional form assumptions on the distribution of measurement error, nor full independence of measurement error from the other variables in the model. Moreover, our results do not require the presence of perfectly measured covariates. We also allow for conditional heteroscedasticity in outcomes, which might be important in empirical applications.

Outline of the paper. The remainder of this paper is organized as follows. Section 2 introduces the linear-in-means model and details the associated naive 2SLS estimator. In Section 3, we show that in the presence of errors-in-variables, esti-

mated peer effects exhibit asymptotic biases and analyze how this bias depends on the interplay between characteristics and links. Section 4 provides formal conditions under which the linear-in-means model with errors-in-variables is identified. Based on these conditions, in Section 5, we propose GMM and 2SLS estimators that are straightforward to implement. A Monte Carlo simulation illustrates the applicability of our methods. Finally, Section 6 concludes. All proofs are in the Appendix. The Appendix and the Online Appendix contain extensions and additional results.

2 Setup

A researcher observes data on outcomes, characteristics, and peers, and wants to estimate the impact of peers' characteristics and peers' outcomes on individual outcomes. We consider a data-generating process where a sequence $\{\mathbf{y}_s, \mathbf{x}_s, \mathbf{e}_s, \mathbf{u}_s, \mathbf{A}_s\}_{s=1, \dots, S}$ of S i.i.d. network observations is drawn from a joint distribution. Network s has size N_s , \mathbf{y}_s is a $N_s \times 1$ vector of outcomes, \mathbf{x}_s is a $N_s \times 1$ vector of continuous characteristics, \mathbf{e}_s is a $N_s \times 1$ vector of disturbances, \mathbf{u}_s is a $N_s \times 1$ vector of measurement errors, and \mathbf{A}_s is the $N_s \times N_s$ adjacency matrix of network s where $(\mathbf{A}_s)_{ij} = 1$ if j is a peer of i and 0 otherwise. Networks may be directed, i.e. $(\mathbf{A}_s)_{ij}$ may differ from $(\mathbf{A}_s)_{ji}$. The researcher observes outcomes \mathbf{y}_s , networks \mathbf{A}_s and mismeasured characteristics $\tilde{\mathbf{x}}_s = \mathbf{x}_s + \mathbf{u}_s$.

We assume that the size of networks N_s is uniformly bounded and consider many-network asymptotics. The number of observations $N = \sum_{s=1}^S N_s \rightarrow \infty$ as the number of networks $S \rightarrow \infty$. Throughout, all probability limits are with respect to N and are assumed to exist and to be finite. Our setup resembles a cross-sectional setting where the researcher observes a large number of small networks (e.g., peers in classrooms or neighborhoods) at a single point in time. The assumption that outcomes, characteristics, and networks are jointly determined is fairly general. This notably covers setups with fixed networks as well as stochastic models of network formation.

Given two random variables a and b , we let $\mathbb{E}(a) := \text{plim } \frac{1}{N} \sum_s \sum_i a_{si}$ denote the expectation of a , $\mathbb{V}(a) := \mathbb{E}[a^2] - \mathbb{E}[a]^2$ its variance, and $\mathbb{C}(a, b) := \text{plim } \frac{1}{N} \sum_s \sum_i a_{si} b_{si} - \mathbb{E}(a)\mathbb{E}(b)$ the covariance between a and b .

2.1 Linear-in-means model

The neighborhood of individual i , \mathcal{N}_{si} , is the set of i 's peers, $j \in \mathcal{N}_{si} \iff (\mathbf{A}_s)_{ij} = 1$. We assume that every individual has at least one peer, $\forall i, \mathcal{N}_{si} \neq \emptyset$. The degree of individual i , d_i , is the number of i 's peers, $d_i = |\mathcal{N}_{si}| = \sum_j (\mathbf{A}_s)_{ij} \geq 1$. Introduce the *interaction* matrix, \mathbf{G}_s , as $(\mathbf{G}_s)_{ij} = (\mathbf{A}_s)_{ij}/d_i$. This matrix is row-normalized — every

row sums to one — and captures linear-in-means interactions.

As our baseline specification, we consider the standard linear-in-means model of social interactions,

$$y_{si} = \alpha + \gamma x_{si} + \delta \frac{1}{d_i} \sum_{j \in \mathcal{N}_{si}} x_{sj} + \beta \frac{1}{d_i} \sum_{j \in \mathcal{N}_{si}} y_{sj} + e_{si},$$

$$\tilde{x}_{si} = x_{si} + u_{si},$$

in which individual i 's outcome depends on her individual characteristic (captured by γ), her peers' average characteristic (captured by δ), and her peers' average outcome (captured by β). We assume that $|\beta| < 1$ such that the reduced form is well-defined. The disturbance in the outcome equation is assumed to satisfy the standard conditional mean independence condition: i.e., $\mathbb{E}(e_{si} \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{u}_s) = 0$.

Measurement error arises because the researcher only observes a noisy proxy \tilde{x}_{si} for the true characteristic x_{si} . Stacking observations, this model can be written compactly in matrix notation as

$$\mathbf{y}_s = \alpha \mathbf{1} + \gamma \mathbf{x}_s + \delta \mathbf{G}_s \mathbf{x}_s + \beta \mathbf{G}_s \mathbf{y}_s + \mathbf{e}_s, \quad (1a)$$

$$\tilde{\mathbf{x}}_s = \mathbf{x}_s + \mathbf{u}_s. \quad (1b)$$

Regressors without measurement error can be partialled out first using the Frisch-Waugh-Lovell (FWL) theorem.

To make the problem tractable, we impose some structure on the moments of the measurement error. Throughout, we assume that its conditional mean is unrelated to all other variables in the model and to the network structure.

Assumption 1. *The measurement errors satisfy:*

$$\mathbb{E}(u_{si} \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = 0.$$

Despite its mildness, Assumption 1 alone can yield identification from first moments; Theorem 1 states the exact conditions. To exploit second moments, we can additionally restrict the conditional variance and pairwise covariance of the measurement error.

Assumption 2. *The measurement errors satisfy:*

$$\mathbb{E}(u_{si}^2 \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = \sigma_u^2, \quad \mathbb{E}(u_{si} u_{sj} \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = 0, \forall i \neq j.$$

Importantly, we make no identifying assumptions on higher moments of the variables in the model, nor do we require full independence of the measurement error

or the disturbance in the outcome equation. We also do not make distributional assumptions such as normality.

In Appendix B we develop two extensions of the baseline model. First, we incorporate network fixed effects. This relaxes the conditional mean independence assumption and is especially useful when unobserved factors cluster within networks. Second, we allow for measurement error in outcomes. In a linear-in-means setting this is nontrivial, since outcomes enter both sides of the equation. We show that, under mild conditions, our identification and estimation results extend to these cases.

2.2 2SLS estimator

Because individuals' outcomes appear on both sides of (1a), OLS estimation delivers biased estimates. Researchers therefore typically resort to a 2SLS estimator where the endogenous average outcomes of peers are instrumented with an instrumental variable z . A popular, model-based choice are the characteristics of peers at distance two (Bramoullé, Djebbari, & Fortin, 2009).

Define $\mathbf{y} := \text{vec}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_S)$, $\tilde{\mathbf{x}} := \text{vec}(\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \dots, \tilde{\mathbf{x}}_S)$, $\mathbf{e} := \text{vec}(\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_S)$, $\mathbf{u} := \text{vec}(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_S)$, $\mathbf{z} := \text{vec}(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_S)$, and let $\mathbf{G} := \text{diag}(\mathbf{G}_1, \mathbf{G}_2, \dots, \mathbf{G}_S)$ be the block-diagonal matrix that contains the interaction matrices. The *naïve* 2SLS estimator for the baseline specification can then be written as

$$\begin{bmatrix} \hat{\alpha}^{IV} & \hat{\gamma}^{IV} & \hat{\delta}^{IV} & \hat{\beta}^{IV} \end{bmatrix}^T := (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y}, \quad (2)$$

where $\mathbf{X} := \begin{bmatrix} \mathbf{1} & \tilde{\mathbf{x}} & \mathbf{G}\tilde{\mathbf{x}} & \mathbf{G}\mathbf{y} \end{bmatrix}$ and $\mathbf{Z} := \begin{bmatrix} \mathbf{1} & \tilde{\mathbf{x}} & \mathbf{G}\tilde{\mathbf{x}} & \mathbf{z} \end{bmatrix}$.

In the presence of measurement error, and as shown below, the 2SLS estimator (2) generally exhibits an asymptotic bias. The reason is that both the individual observed characteristic and average observed characteristics of peers are endogenous. To see why, substitute Equation (1b) in (1a), which gives that

$$\mathbf{y}_s = \alpha \mathbf{1} + \gamma \tilde{\mathbf{x}}_s + \delta \mathbf{G}_s \tilde{\mathbf{x}}_s + \beta \mathbf{G}_s \mathbf{y}_s + \underbrace{\mathbf{e}_s - \gamma \mathbf{u}_s - \delta \mathbf{G}_s \mathbf{u}_s}_{:= \boldsymbol{\eta}_s}. \quad (3)$$

The endogeneity problem arises because individual and peer characteristics are typically correlated with the composite error term: i.e., $\mathbb{C}(\tilde{x}, \eta) \neq 0$ and $\mathbb{C}(G\tilde{x}, \eta) \neq 0$ if $\gamma, \delta \neq 0$. At the core of the problem is an underidentification problem: there are three endogenous variables, but only one instrumental variable.

3 Asymptotic bias

We study the asymptotic bias in peer effect estimates in the presence of errors-in-variables. Our results highlight the role of the interplay between characteristics and links. We first present results for the general case, and then build intuition by focusing on two special cases that are relevant in applied work.

The general case. We derive analytical expressions for the probability limits of the naive 2SLS estimates of the individual effect $\hat{\gamma}^{IV}$, the contextual peer effect $\hat{\delta}^{IV}$, and the endogenous peer effect $\hat{\beta}^{IV}$. To focus the discussion, throughout this section we take the instrument z to be the average characteristics of peers at distance two.⁷ By Assumptions 1 and 2, z is uncorrelated with measurement error at distance zero and one, $\mathbb{C}(z, u) = \mathbb{C}(z, Gu) = 0$. In addition, $\mathbb{C}(z, e) = 0$, so the instrument is valid (for a formal discussion, see Proposition 5 in Section 5.2).

By substituting Equations (1a) and (1b) in (2), and applying the Frisch-Waugh-Lovell (FWL) theorem to partial out the intercept, we express the asymptotic biases of $\hat{\gamma}^{IV}$, $\hat{\delta}^{IV}$, and $\hat{\beta}^{IV}$ in terms of the true model parameters and the matrices \mathbf{S} and $\mathbf{\Sigma}$, which collect (co)variances in characteristics and errors.

Lemma 1. *Suppose that Assumptions 1 and 2 hold and that the matrix $(\mathbf{S} + \mathbf{\Sigma})$ is invertible, where*

$$\mathbf{S} := \begin{bmatrix} \mathbb{V}(x) & \mathbb{C}(x, Gx) & \mathbb{C}(x, Gy) \\ \mathbb{C}(Gx, x) & \mathbb{V}(Gx) & \mathbb{C}(Gx, Gy) \\ \mathbb{C}(z, x) & \mathbb{C}(z, Gx) & \mathbb{C}(z, Gy) \end{bmatrix}, \quad \mathbf{\Sigma} := \begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & h_0 \sigma_u^2 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and $h_0 := \frac{1}{N} \sum_s \sum_i \frac{1}{d_{si}}$. Then the IV estimates of γ, δ, β converge in probability to

$$\text{plim} \begin{bmatrix} \hat{\gamma}^{IV} \\ \hat{\delta}^{IV} \\ \hat{\beta}^{IV} \end{bmatrix} = \underbrace{(\mathbf{S} + \mathbf{\Sigma})^{-1} \mathbf{S}}_{:= \mathbf{M}} \begin{bmatrix} \gamma \\ \delta \\ \beta \end{bmatrix}.$$

The matrix $\mathbf{\Sigma}$ has a specific structure. First, under Assumption 2, measurement error is homoscedastic and uncorrelated across peers, so that $\mathbb{V}(u) = \sigma_u^2$ and $\mathbb{C}(u, Gu) = 0$. Moreover, the measurement error in the instrument is uncorrelated with the measurement error in both the individual characteristic and the average peer characteristic. Second, due to the averaging across peers, measurement error in the average

⁷The use of characteristics from peers at distance two is common practice in applied work on peer effects (see, e.g., Patacchini & Venanzoni, 2014; Nicoletti et al., 2018; De Giorgi et al., 2019). Note that the instrument $G^2 \tilde{x}$, which is valid in the absence of measurement error (Bramoullé et al., 2009) would introduce an additional source of bias. It is not valid in the presence of error-in-variables, since it is also built from links at distances zero and one and hence $\mathbb{C}(z, \eta) \neq 0$.

peer characteristic has smaller variance than measurement error in the individual characteristic. In particular, we have $\mathbb{V}(Gu) = h_0\sigma_u^2$, where $h_0 \leq 1$ is the arithmetic mean of the inverse degrees in the network. This network statistic captures the overall extent of the averaging across peers.

Note that in the absence of measurement error, $\Sigma = \mathbf{0}$ and $\mathbf{M} = \mathbf{I}$, so there is no asymptotic bias, as expected. In many settings, the presence of measurement error entails an attenuation bias toward zero, such that estimates are smaller than the corresponding true model parameters (e.g., see [Wansbeek & Meijer, 2000](#) for a detailed discussion). As argued above, however, this need not hold for the peer-effect coefficients in the linear-in-means model.

In the presence of measurement error, Lemma 1 implies that each estimate is a linear combination of the three true parameters. The off-diagonal elements of \mathbf{M} capture whether the (biased) estimate of one effect is picking up part of another effect. This phenomenon is referred to as *smearing* or *weight shifting* in the measurement-error literature. It can give rise to an expansion bias, falsely suggesting the presence of peer effects when there are none. For example, if $M_{31} \neq 0$, then $\hat{\beta}^{IV}$ inherits a smearing component from γ (and if $M_{32} \neq 0$ from δ). As a result, researchers may falsely detect endogenous peer effects even when $\beta = 0$, provided that $\gamma \neq 0$ or $\delta \neq 0$.

We study general properties of the probability limits in the next result. We find it useful to work with the following block decomposition of matrices \mathbf{S} and Σ :

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{s}_{12} \\ \mathbf{s}_{21} & s_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix},$$

where \mathbf{S}_{11} and Σ_{11} are 2×2 matrices and \mathbf{s}_{12} and \mathbf{s}_{21}^\top are 2×1 vectors. Define \mathbf{K} as the Schur complement of s_{22} in $\mathbf{S} + \Sigma$, i.e., $\mathbf{K} := \mathbf{S}_{11} + \Sigma_{11} - \frac{1}{s_{22}}\mathbf{s}_{12}\mathbf{s}_{21}$.

Proposition 1. *Suppose that Assumptions 1 and 2 hold, that the matrix $(\mathbf{S} + \Sigma)$ is invertible, and that $s_{22} \neq 0$. Then the IV estimates of γ , δ , β converge in probability to*

$$\text{plim} \begin{bmatrix} \hat{\gamma}^{IV} \\ \hat{\delta}^{IV} \\ \hat{\beta}^{IV} \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \mathbf{K}^{-1}\Sigma_{11} & \mathbf{0} \\ \frac{1}{s_{22}}\mathbf{s}_{21}\mathbf{K}^{-1}\Sigma_{11} & 1 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \\ \beta \end{bmatrix}.$$

When characteristics are independent of links, this further simplifies to

$$\text{plim} \begin{bmatrix} \hat{\gamma}^{IV} \\ \hat{\delta}^{IV} \\ \hat{\beta}^{IV} \end{bmatrix} = \begin{bmatrix} \frac{\mathbb{V}(x)}{\mathbb{V}(x) + \sigma_u^2} & 0 & 0 \\ 0 & \frac{\mathbb{V}(Gx)}{\mathbb{V}(Gx) + h_0\sigma_u^2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \\ \beta \end{bmatrix}.$$

Proposition 1 shows that, under classical measurement error, peer effect estimates are generally biased due to smearing. The estimated contextual peer effect $\hat{\delta}^{IV}$ picks up part of the individual effect, and the estimated endogenous peer effect $\hat{\beta}^{IV}$ picks up parts of the individual and contextual peer effects. Interestingly, there is no smearing in the opposite direction: $\hat{\gamma}^{IV}$ and $\hat{\delta}^{IV}$ do not pick up any part of the endogenous peer effect.

These smearing effects depend in complex ways on the relationship between characteristics and network structure and are therefore difficult to sign analytically. Indeed, we show below that even in the simpler model without contextual peer effects (i.e., with only endogenous peer effects), the smearing term can be negative or positive. Nevertheless, even in this general setup we can make some qualitative observations. As shown in Appendix A.2, the probability limit of β^{IV} can be written as

$$\text{plim } \hat{\beta}^{IV} = \frac{\sigma_u^2 \varphi_{z,\tilde{x}}}{D_1^{IV}} \gamma + \frac{h_0 \sigma_u^2 \varphi_{z,G\tilde{x}}}{D_1^{IV}} \delta + \beta,$$

where $D_1^{IV} = \mathbb{C}(z, Gy) - \varphi_{z,\tilde{x}} \mathbb{C}(x, Gy) - \varphi_{z,G\tilde{x}} \mathbb{C}(Gx, Gy)$, and where $\varphi_{z,\tilde{x}}$ and $\varphi_{z,G\tilde{x}}$ denote the population regression coefficients of \tilde{x} and $G\tilde{x}$, respectively, from regressing z on a constant, \tilde{x} , and $G\tilde{x}$. This representation shows that the smearing of γ and δ into β^{IV} is governed by the extent to which the observed individual and average peer characteristics are correlated with the instrument. In particular, γ and δ do not smear into $\hat{\beta}^{IV}$ when $\varphi_{z,\tilde{x}} = 0$ and $\varphi_{z,G\tilde{x}} = 0$, respectively.⁸ Intuitively, this highlights that correlation between the instrument and the individual and peer characteristics—which can naturally arise under homophily—is a key driver of the asymptotic bias in $\hat{\beta}^{IV}$. With measurement error, the estimated endogenous peer effect absorbs parts of the individual effect and of the contextual peer effect, because the instrument spuriously picks up variation in the true individual and average peer characteristic.

In addition, the asymptotic bias in $\hat{\beta}^{IV}$ tends to decrease with instrument strength, as captured by $\mathbb{C}(z, Gy)$. Ceteris paribus, stronger instruments reduce the noise-to-signal ratio and therefore mitigate the bias induced by errors-in-variables. This suggests that error-in-variables may exacerbate the consequences of weak instruments.

Proposition 1 also specializes the result to the case in which characteristics are independent of links, which is particularly relevant in experimental settings. In this case, the smearing effects disappear and $\hat{\gamma}^{IV}$ and $\hat{\delta}^{IV}$ exhibit attenuation bias. Interestingly, $\hat{\beta}^{IV}$ remains consistent.

⁸In Appendix A.2, we show that $\mathbb{C}(z, x) = \mathbb{C}(z, Gx) = 0$ implies $\varphi_{z,\tilde{x}} = \varphi_{z,G\tilde{x}} = 0$.

Two particular cases. To build intuition and further illustrate these effects, we next discuss two particular cases. The linear-in-means model accommodates two channels of peer effects, but applied work often estimates simpler variants that include only one channel.⁹ This naturally raises the question of whether the insights underlying Proposition 1 can be sharpened in these simpler models.

Suppose first that there is *no endogenous peer effect*, i.e., $\beta = 0$, and that this restriction is known by the researcher. In Online Appendix O.A.1, we derive the analogue of Proposition 1 for the restricted model estimated by OLS. We find that even in this case, $\hat{\delta}^{OLS}$ generally exhibits bias in the direction of γ . More precisely,

$$\text{plim } \hat{\delta}^{OLS} = \frac{\sigma_u^2 \mathbb{C}(x, Gx)}{D^{OLS}} \gamma + \frac{\mathbb{V}(x)\mathbb{V}(Gx) - \mathbb{C}(x, Gx)^2 + \sigma_u^2 \mathbb{V}(Gx)}{D^{OLS}} \delta, \quad (5)$$

where $D^{OLS} := \mathbb{V}(x)\mathbb{V}(Gx) - \mathbb{C}(x, Gx)^2 + h_0 \sigma_u^2 \mathbb{V}(x) + \sigma_u^2 \mathbb{V}(Gx) + h_0 \sigma_u^4$. Hence, estimates of contextual peer effects can exhibit asymptotic bias even when endogenous peer effects are absent.

In contrast to the general case, the smearing term can be signed here using the sign of $\mathbb{C}(x, Gx)$, since $D^{OLS} > 0$. In particular, expansion bias is positive when individual characteristics are positively correlated with the average characteristic of their peers. Such positive correlation arises naturally under homophily, when similar agents are more likely to be connected. The magnitude of the expansion bias is also decreasing in the inverse average degree h_0 .

The mechanism underlying the bias is as follows. When $\mathbb{C}(x, Gx) > 0$, $\hat{\delta}^{OLS}$ will spuriously pick-up part of the variation in x , and this tendency is more pronounced when the measurement error in $G\tilde{x}$ is smaller. This is because $G\tilde{x}$ then becomes a better proxy for x . Since $G\tilde{x}$ averages across peers, Lemma 1 implies that the variance of its measurement error is proportional to h_0 . Hence, a lower h_0 implies less measurement error and therefore more smearing: a larger share of the individual effect γ is loaded onto the estimated contextual peer effect $\hat{\delta}^{OLS}$.

Conversely, suppose next that there is *no contextual peer effect*, i.e., $\delta = 0$, and that, again, this restriction is known by the researcher. In this case, peers' characteristics $G\tilde{x}$ provide a natural instrument for peers' outcomes Gy . Computations in Online Appendix O.A.2 show that under 2SLS estimation, $\hat{\beta}^{IV}$ has an asymptotic bias in the direction of γ :

$$\text{plim } \hat{\beta}^{IV} = \frac{\sigma_u^2 \varphi_{G\tilde{x}, \tilde{x}}}{D_2^{IV}} \gamma + \beta, \quad (6)$$

where $D_2^{IV} := \mathbb{C}(Gx, Gy) - \varphi_{G\tilde{x}, \tilde{x}} \mathbb{C}(x, Gy)$, and where $\varphi_{G\tilde{x}, \tilde{x}} = \frac{\mathbb{C}(x, Gx)}{\mathbb{V}(x) + \sigma_u^2}$ denotes the

⁹Models with contextual peer effects only are estimated, for instance, in Carrell, Fullerton, and West (2009); Lavy and Schlosser (2011). Models with endogenous peer effects only are estimated in Gaviria and Raphael (2001); Trogdon, Nonnemaker, and Pais (2008); Paul and Nath (2024).

population regression coefficient on \tilde{x} from regressing $G\tilde{x}$ on a constant and \tilde{x} . Hence, estimates of endogenous peer effects can be asymptotically biased even when contextual peer effects are absent.

The smearing term is non-zero only when $\mathbb{C}(x, Gx) \neq 0$. As in the general case, the extent of smearing is driven by correlation between the instrument and the observed characteristics. The asymptotic bias tends to decrease with instrument strength, as captured by $\mathbb{C}(Gx, Gy)$.

When both types of peer effects are present, these sources of bias get combined, leading to the asymptotic biases highlighted in Proposition 1.

4 Identification

We now consider the identification of the linear-in-means model of peer effects, accounting for errors-in-variables. Our aim is to establish the conditions required for identification in this model, both necessary and sufficient. We demonstrate that the model exhibits generic identification properties, even in the absence of external information. In particular, inherent features of the network can serve as a valuable tool for mitigating measurement error and enabling the identification of peer effects.

We propose two complementary approaches to identify the parameters of interest $(\alpha, \beta, \gamma, \delta, \sigma_u^2)$, depending on how much structure is imposed on measurement error. The first strategy is based on conditional mean restrictions and only relies on Assumption 1 (mean-zero). It identifies $(\alpha, \beta, \gamma, \delta)$ by exploiting variation in average individual characteristics across network positions. This strategy is valid when, for instance, agents with more friends tend to have a higher value for the characteristic. The second identification strategy is based on conditional covariance restrictions and relies on Assumptions 1 and 2 (homoscedasticity and uncorrelatedness). It identifies $(\beta, \gamma, \delta, \sigma_u^2)$ by exploiting variation in covariances of individual characteristics across pairs of network positions.¹⁰ We maintain the weak assumptions on the disturbance term in Equation (1a) and on the measurement error in Equation (1b), and therefore focus exclusively on mean and covariance restrictions. Researchers willing to impose more structure could exploit additional moments for identification.¹¹

For identification purposes, we form a single pooled interaction matrix G_0 of dimensions $N_0 \times N_0$ by placing all networks in the support on the diagonal of a block-diagonal matrix. The setup is feasible because bounded network size implies a finite support of interaction matrices, even as we observe arbitrarily many net-

¹⁰In the absence of information on social connections, [de Paula et al. \(2024\)](#) implicitly leverage similar covariances to identify both social interactions and peer effects. In contrast to their approach, our results do not require panel data.

¹¹For example, [Rose \(2017\)](#) identifies a linear-in-means model from second moments of outcomes under homoscedastic disturbances.

works. We likewise stack observed outcomes and characteristics into conformable vectors \mathbf{y}_0 and $\tilde{\mathbf{x}}_0$.

We say that two nodes i_0 and j_0 have symmetric network positions if they belong to the same network \mathbf{A} in \mathbf{G}_0 and if there exists a permutation of nodes' labels, π , that maps one node into another while preserving links. Formally, $\pi(i_0) = j_0$, $\pi(j_0) = i_0$ and for all pairs of network nodes k_0, l_0 , $A_{\pi(k_0)\pi(l_0)} = A_{k_0l_0}$. In this case, i_0 and j_0 are undistinguishable and hence have the same conditional distribution of characteristics and outcomes. We then obtain the number of *unique network positions* by counting symmetric positions only once. For instance, a star with N_0 nodes has two unique positions, center and periphery, while a circle with N_0 nodes has a unique network position. By contrast, all positions are generally unique in a large complex network.

4.1 Mean restrictions

To derive the conditional mean restrictions, note that using Equation (3) the reduced form can be written as

$$\mathbf{y}_s = (\mathbf{I} - \beta \mathbf{G}_s)^{-1} [\alpha \mathbf{1} + (\gamma \mathbf{I} + \delta \mathbf{G}_s) \tilde{\mathbf{x}} + \boldsymbol{\eta}_s].$$

Stacking networks, taking expectations on both sides, and using Assumption 1, we obtain

$$\mathbb{E}[\mathbf{y}_0 \mid \mathbf{G}_0] = \underbrace{(\mathbf{I} - \beta \mathbf{G}_0)^{-1} [\alpha \mathbf{1} + (\gamma \mathbf{I} + \delta \mathbf{G}_0) \mathbf{m}]}_{:= \mathbf{r}(\mathbf{m}, \mathbf{G}_0; \boldsymbol{\theta}_1)}, \quad (7)$$

where $\mathbf{m} := \mathbb{E}(\tilde{\mathbf{x}}_0 \mid \mathbf{G}_0)$ collects the expected characteristic by network position and $\boldsymbol{\theta}_1 := (\alpha, \beta, \gamma, \delta)$. Since there are N_0 network positions in total, this yields a nonlinear system of N_0 equations in four unknowns. Except for knife-edge cases, most networks should then deliver a substantial degree of overidentification. This identification strategy exploits variation in individual characteristics across network positions. By taking expectations by network position, we need no restriction on the covariance structure of measurement error in the network, and peers can make correlated mistakes.

We next derive the precise identification conditions. Say that the model is *identified from conditional mean restrictions* if there do not exist $\boldsymbol{\theta}_1, \boldsymbol{\theta}'_1$ with $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}'_1$ such that $\mathbf{r}(\mathbf{m}, \mathbf{G}_0; \boldsymbol{\theta}_1) = \mathbf{r}(\mathbf{m}, \mathbf{G}_0; \boldsymbol{\theta}'_1)$. We require $\beta\gamma + \delta \neq 0$; if this condition fails, it corresponds to a knife-edge case in which contextual and endogenous peer effects exactly offset one another (Bramoullé et al., 2009).

Theorem 1. *Suppose that Assumption 1 holds and that $\beta\gamma + \delta \neq 0$. The parameters $\alpha, \beta, \gamma, \delta$ are identified from conditional mean restrictions if and only if the vectors $\mathbf{1}, \mathbf{m}$,*

$\mathbf{G}_0\mathbf{m}$, $\mathbf{G}_0^2\mathbf{m}$ are linearly independent.

Identification based on mean restrictions requires the matrix $[\mathbf{1}, \mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}]$ to have full rank, that is, rank four. This condition captures restrictions on the network structure and on how the characteristic varies with network position. Identification can only hold if there is enough variation in conditional expectations m_{i_0} . Identification fails to hold, for instance, with homogeneous marginals, when for every pair i_0, j_0 , x_{i_0} and x_{j_0} have the same distributions. In that case, $m_{i_0} = m_{j_0}$ and $\mathbf{m} = \lambda\mathbf{1}$ for some scalar λ . It also fails to hold when the network has less than four unique network positions. This happens, for instance, if every network \mathbf{A}_s is a star with N_0 nodes (two unique network positions) or if every network \mathbf{A}_s is a line with five nodes (three unique network positions).

More generally, this rank condition is stronger than the condition in [Bramoullé et al. \(2009\)](#), who show identification in the linear-in-means model without measurement error when \mathbf{I} , \mathbf{G}_0 , \mathbf{G}_0^2 are linearly independent. Indeed, suppose that identification fails to hold in the absence of measurement error. Then there exist $\lambda_0, \lambda_1, \lambda_2$ not all equal to zero such that $\lambda_0\mathbf{I} + \lambda_1\mathbf{G}_0 + \lambda_2\mathbf{G}_0^2 = 0$. This implies that $\lambda_0\mathbf{m} + \lambda_1\mathbf{G}_0\mathbf{m} + \lambda_2\mathbf{G}_0^2\mathbf{m} = 0$ and hence the vectors $\mathbf{1}, \mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}$ are linearly dependent. Identification conditions are naturally more demanding with measurement error than without.

When there is variation in conditional expectations, identification is related to the spectral properties of the interaction matrix \mathbf{G}_0 . More precisely, suppose that identification fails to hold. There are two cases. Either there exist λ_0, λ_1 such that $\mathbf{G}_0\mathbf{m} = \lambda_0\mathbf{1} + \lambda_1\mathbf{m}$. Then if $\lambda_1 \neq 1$, we can see that $\mathbf{m} + \frac{\lambda_0}{\lambda_1-1}\mathbf{1}$ is an eigenvector of \mathbf{G}_0 for the eigenvalue λ_1 . Or there exist $\lambda_0, \lambda_1, \lambda_2$ such that $\mathbf{G}_0^2\mathbf{m} = \lambda_0\mathbf{1} + \lambda_1\mathbf{m} + \lambda_2\mathbf{G}_0\mathbf{m}$. In that case, and if $\lambda_1 + \lambda_2 \neq 1$, we can check that $\mathbf{m} + \frac{\lambda_0}{\lambda_1+\lambda_2-1}\mathbf{1}$ is an eigenvector of the matrix $\mathbf{G}_0^2 - \lambda_2\mathbf{G}_0$ for the eigenvalue λ_1 . In either case, the vector of characteristic expectations conditional of network positions must be precisely related to some eigenvector of a simple polynomial function of \mathbf{G}_0 .¹²

This suggests that non-identification based on mean restrictions is uncommon. We next derive a sufficient condition for generic identification. The following proposition shows that, if the network is sufficiently rich, the set of mean vectors \mathbf{m} for which identification fails is negligible. To formally define what it means for identification to be generic, we must first account for the fact that two nodes i and j with symmetric network positions are undistinguishable, and notably satisfy $m_i = m_j$.

¹²In graph theory, it is well known that structural symmetries in a network are often associated with higher algebraic multiplicities of the eigenvalues of its interaction matrix ([Biggs, 1993](#)). At one extreme lies the complete graph, whose interaction matrix has an eigenvalue with algebraic multiplicity $N_0 - 1$. At the other extreme, Erdős–Rényi random graphs typically exhibit simple spectra with high probability. Informally, the more “asymmetric” the network, the less likely it is that an arbitrary vector is an eigenvector of the associated interaction matrix.

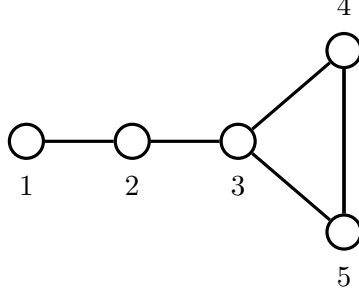


Figure 1: A 5-node network

Denote by \check{N}_0 the number of unique network positions, by $\check{\mathbf{m}}$ the $\check{N}_0 \times 1$ vector of conditional expectations and by $\check{\mathbf{G}}_0$ the $\check{N}_0 \times \check{N}_0$ interaction matrix defined over unique network positions.¹³ Note that Equation (7) and Theorem 1 still hold when replacing \mathbf{m} by $\check{\mathbf{m}}$ and \mathbf{G}_0 by $\check{\mathbf{G}}_0$.

Formally, for a fixed \mathbf{G}_0 , we say that the model is *generically identified from conditional mean conditions* if the set

$$\left\{ \check{\mathbf{m}} \in \mathbb{R}^{\check{N}_0} : \exists \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}'_1 \implies \mathbf{r}(\check{\mathbf{m}}, \check{\mathbf{G}}_0; \boldsymbol{\theta}_1) = \mathbf{r}(\check{\mathbf{m}}, \check{\mathbf{G}}_0; \boldsymbol{\theta}'_1) \right\}$$

has Lebesgue measure zero. Our definition of generic identification is conceptually related to the order (or rank) conditions in systems of linear equations being generically satisfied (Lewbel, 2019, 886). Note, however, that non-generic cases may still be of substantive interest.

Proposition 2. *Suppose that Assumption 1 holds and that $\beta\gamma + \delta \neq 0$. If $\check{\mathbf{G}}_0$ has at least 4 distinct eigenvalues, the parameters $\alpha, \beta, \gamma, \delta$ are generically identified from conditional mean restrictions.*

To illustrate Theorem 1 and Proposition 2, suppose that a researcher observes many networks of five nodes with the structure depicted in Figure 1. Here, positions 4 and 5 are symmetric, and hence the network structure has $\check{N}_0 = 4$ unique network positions. Parameters are identified from conditional mean restrictions if

¹³ $\check{\mathbf{m}}$ is obtained from \mathbf{m} by keeping a unique entry for each unique network position. $\check{\mathbf{G}}_0$ is obtained from \mathbf{G}_0 by keeping a unique row for each unique network position and by summing the columns over symmetric network positions.

the following system of linear equations in $(\alpha, \beta, \gamma, \delta)$ has a unique solution:

$$\begin{aligned} y_1 &= \alpha + \gamma m_1 + \delta m_2 + \beta y_2, \\ y_2 &= \alpha + \gamma m_2 + \delta \left(\frac{1}{2} m_1 + \frac{1}{2} m_3 \right) + \beta \left(\frac{1}{2} y_1 + \frac{1}{2} y_3 \right), \\ y_3 &= \alpha + \gamma m_3 + \delta \left(\frac{1}{3} m_2 + \frac{1}{3} m_4 + \frac{1}{3} m_5 \right) + \beta \left(\frac{1}{3} y_2 + \frac{1}{3} y_4 + \frac{1}{3} y_5 \right), \\ y_4 &= \alpha + \gamma m_4 + \delta \left(\frac{1}{2} m_3 + \frac{1}{2} m_5 \right) + \beta \left(\frac{1}{2} y_3 + \frac{1}{2} y_5 \right), \\ y_5 &= \alpha + \gamma m_5 + \delta \left(\frac{1}{2} m_3 + \frac{1}{2} m_4 \right) + \beta \left(\frac{1}{2} y_3 + \frac{1}{2} y_4 \right), \end{aligned}$$

where here y_{i_0} denote expected outcome conditional on position i_0 . By Theorem 1 parameters are identified when $\beta\gamma + \delta \neq 0$ if and only if $\mathbf{1}, \mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}$ are linearly independent. Here,

$$\mathbf{G}_0 = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}, \text{ and } \mathbf{G}_0^2 = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{2}{3} & 0 & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & 0 & \frac{1}{2} & \frac{1}{6} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{4} & \frac{5}{12} & \frac{1}{6} \\ 0 & \frac{1}{6} & \frac{1}{4} & \frac{1}{6} & \frac{5}{12} \end{bmatrix},$$

and identification from mean restrictions holds if and only if the following matrix has rank four:

$$[\mathbf{1}, \mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}] = \begin{bmatrix} 1 & m_1 & m_2 & \frac{1}{2}m_1 + \frac{1}{2}m_3 \\ 1 & m_2 & \frac{1}{2}m_1 + \frac{1}{2}m_3 & \frac{2}{3}m_2 + \frac{1}{6}m_4 + \frac{1}{6}m_5 \\ 1 & m_3 & \frac{1}{3}m_2 + \frac{1}{3}m_4 + \frac{1}{3}m_5 & \frac{1}{6}m_1 + \frac{1}{2}m_3 + \frac{1}{6}m_4 + \frac{1}{6}m_5 \\ 1 & m_4 & \frac{1}{2}m_3 + \frac{1}{2}m_5 & \frac{1}{6}m_2 + \frac{1}{4}m_3 + \frac{5}{12}m_4 + \frac{1}{6}m_5 \\ 1 & m_5 & \frac{1}{2}m_3 + \frac{1}{2}m_4 & \frac{1}{6}m_2 + \frac{1}{4}m_3 + \frac{1}{6}m_4 + \frac{5}{12}m_5 \end{bmatrix}.$$

Then, to apply Proposition 2, we merge the two symmetric positions 4 and 5. The interaction matrix defined over unique positions is now

$$\check{\mathbf{G}}_0 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}.$$

We can easily check that this matrix has four distinct eigenvalues. By Proposition 2, we know that for this network structure, the set of $\check{\mathbf{m}} \in \mathbb{R}^4$ for which identification fails has Lebesgue measure zero.

4.2 Covariance restrictions

Another source of identification is provided by conditional covariance restrictions. Consider $\mathbb{C}(\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0)$ the matrix that collects over all pairs (i_0, j_0) of network positions the covariance between outcome at position i_0 and the observed characteristic at position j_0 . Denote by $\mathbf{C} := \mathbb{V}(\tilde{\mathbf{x}}_0 \mid \mathbf{G}_0)$ the matrix of covariances of observed characteristics between network positions. Observe that under Assumption 1, $\mathbb{V}(\tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) = \mathbb{V}(\mathbf{x}_0 \mid \mathbf{G}_0) + \mathbb{V}(\mathbf{u} \mid \mathbf{G}_0)$. If one is willing to additionally impose Assumption 2, $\mathbb{V}(\mathbf{u} \mid \mathbf{G}_0) = \sigma_u^2 \mathbf{I}$. Replacing \mathbf{y}_0 by its reduced-form expression gives

$$\mathbb{C}(\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) = \underbrace{(\mathbf{I} - \beta \mathbf{G}_0)^{-1}(\gamma \mathbf{I} + \delta \mathbf{G}_0)(\mathbf{C} - \sigma_u^2 \mathbf{I})}_{:= \mathbf{R}(\mathbf{C}, \mathbf{G}_0; \boldsymbol{\theta}_2)}, \quad (8)$$

where $\boldsymbol{\theta}_2 := (\beta, \gamma, \delta, \sigma_u^2)$. This yields a nonlinear system of N_0^2 equations in four unknowns. As with mean restrictions, the model should then be substantially overidentified for most networks. This identification strategy exploits variation in covariances between observed characteristic across pairs of network positions.

We derive the exact identification conditions in the following result. We say the model is *identified from conditional covariance restrictions* if there do not exist $\boldsymbol{\theta}_2, \boldsymbol{\theta}'_2$ with $\boldsymbol{\theta}_2 \neq \boldsymbol{\theta}'_2$ such that $\mathbf{R}(\mathbf{C}, \mathbf{G}_0; \boldsymbol{\theta}_2) = \mathbf{R}(\mathbf{C}, \mathbf{G}_0; \boldsymbol{\theta}'_2)$.

Theorem 2. *Suppose that Assumptions 1 and 2 hold and that $\beta\gamma + \delta \neq 0$. The parameters $\beta, \gamma, \delta, \sigma_u^2$ are identified from conditional covariance restrictions if and only if the matrices $\mathbf{I}, \mathbf{G}_0, \mathbf{G}_0^2, \mathbf{C}, \mathbf{G}_0\mathbf{C}, \mathbf{G}_0^2\mathbf{C}$ are linearly independent.*

Identification based on covariance restrictions only holds if there is enough variation in characteristic covariance across network positions pairs. Indeed, Theorem 2 implies that identification fails when correlation in characteristics across pairs of network positions is constant. To see why, assume that $\mathbb{V}(x_{i_0}) = \sigma_x^2$ and $\mathbb{C}(x_{i_0}, x_{j_0}) = \rho\sigma_x^2$ for every pair of positions $i_0 \neq j_0$ who belong to the same network structure in \mathbf{G}_0 . Let \mathbf{J}_0 be a block-diagonal matrix where blocks correspond to networks in \mathbf{G}_0 and each block is a matrix of ones. Then, $\mathbf{C} = ((1 - \rho)\sigma_x^2 + \sigma_u^2)\mathbf{I} + \rho\sigma_x^2\mathbf{J}_0$. And note that since \mathbf{G}_0 is block-row stochastic, $\mathbf{G}_0\mathbf{J}_0 = \mathbf{J}_0$. This implies that $\mathbf{G}_0\mathbf{C} = ((1 - \rho)\sigma_x^2 + \sigma_u^2)\mathbf{G}_0 + \rho\sigma_x^2\mathbf{J}_0 = \mathbf{C} + ((1 - \rho)\sigma_x^2 + \sigma_u^2)(\mathbf{G}_0 - \mathbf{I})$. The linear independence condition of Theorem 2 is not satisfied, and hence the parameters are not identified from covariance restrictions.

By contrast, let us next illustrate how the identification conditions operate when characteristic covariance varies across pairs of network positions. Define, as usual, *network distance* $d(i_0, j_0)$ between two positions i_0, j_0 as the number of links in a shortest path between them and *diameter* as the largest finite network distance. Consider a setup where covariance between individuals' characteristics varies with network

distance, up to distance three, and where the diameter of the network is greater than or equal to five. Thus, $\mathbb{V}(x_{i_0}) = \sigma_x^2$, $\mathbb{C}(x_{i_0}, x_{j_0}) = \rho_1 \sigma_x^2$ if $d(i_0, j_0) = 1$, $\rho_2 \sigma_x^2$ if $d(i_0, j_0) = 2$, $\rho_3 \sigma_x^2$ if $d(i_0, j_0) = 3$ and 0 if $d(i_0, j_0) \geq 4$ with $\rho_1, \rho_2, \rho_3 > 0$. In that case, we can show that the linear independence condition of Theorem 2 is satisfied. Indeed, there exists a pair of positions i_0, j_0 such that $d(i_0, j_0) = 5$. For this pair, we see that $(\mathbf{G}_0^2 \mathbf{C})_{i_0 j_0} > 0$ while $(\mathbf{G}_0)_{i_0 j_0} = (\mathbf{G}_0^2)_{i_0 j_0} = \mathbf{C}_{i_0 j_0} = (\mathbf{G}_0 \mathbf{C})_{i_0 j_0} = 0$. Therefore, the matrix $\mathbf{G}_0^2 \mathbf{C}$ cannot be expressed as a linear combination of \mathbf{I} , \mathbf{G}_0 , \mathbf{G}_0^2 , \mathbf{C} , $\mathbf{G}_0 \mathbf{C}$. Repeating the argument with a pair at distance four, we see that $\mathbf{G}_0 \mathbf{C}$ is not a linear combination of \mathbf{I} , \mathbf{G}_0 , \mathbf{G}_0^2 , \mathbf{C} , and with a pair a distance three, that \mathbf{C} is not a linear combination of \mathbf{I} , \mathbf{G}_0 , \mathbf{G}_0^2 . Therefore, the matrices \mathbf{I} , \mathbf{G}_0 , \mathbf{G}_0^2 , \mathbf{C} , $\mathbf{G}_0 \mathbf{C}$, $\mathbf{G}_0^2 \mathbf{C}$ are linearly independent and hence parameters $(\beta, \gamma, \delta, \sigma_u^2)$ are identified from covariance restrictions.

5 Estimation

While dependencies between characteristics and links may generate asymptotic bias, it simultaneously offers a valuable source of information that can be harnessed to develop consistent estimators for the parameters of interest. We next introduce straightforward and practical GMM and 2SLS estimators, directly built upon the conditional mean and covariance restrictions highlighted in the previous section.¹⁴

¹⁵ Importantly, these estimators are applicable to setups with fixed as well as stochastic networks.

5.1 GMM estimators

Fixed networks. First consider the case where a fixed interaction matrix \mathbf{G} is observed repeatedly by the analyst. It is applicable, for instance, in experimental setups where the researcher can control the network.

We first introduce a GMM estimator based on the conditional mean restrictions in Equation (7).¹⁶ The associated conditional moment conditions are $\mathbb{E}[\mathbf{v}_{1,s}(\boldsymbol{\theta}_1, \mathbf{G}) \mid \mathbf{G}] = \mathbf{0}$, where

$$\mathbf{v}_{1,s}(\boldsymbol{\theta}_1, \mathbf{G}) := (\mathbf{I} - \beta \mathbf{G}) \mathbf{y}_s - \alpha \mathbf{1} - (\gamma \mathbf{I} + \delta \mathbf{G}) \tilde{\mathbf{x}}_s.$$

¹⁴While the GMM approach can yield more efficient parameter estimates, it does introduce non-linearity into the estimation procedure. Appendix B.1 discusses how to apply the GMM estimators in the presence of network-specific fixed effects.

¹⁵In practice, researchers are advised to test for weak moments and weak instruments, since weak identification can lead to poor finite-sample properties and unreliable conventional inference.

¹⁶We construct the moment conditions from the *structural* rather than the *reduced* form of the model. These conditions are computationally more efficient because they avoid matrix inversion. Nevertheless, the two approaches are equivalent in the population and asymptotically equivalent when using the optimal weighting matrix.

Proposition 3. Suppose that Assumption 1 holds, that $\beta\gamma + \delta \neq 0$, and that the condition in Theorem 1 is satisfied. Under standard regularity conditions (see Appendix A.6), the GMM estimator

$$\hat{\theta}_1^{GMM} := \arg \min_{\theta_1} \left(\frac{1}{S} \sum_s \mathbf{v}_{1,s}(\theta_1, \mathbf{G}) \right)^\top \Omega_1 \left(\frac{1}{S} \sum_s \mathbf{v}_{1,s}(\theta_1, \mathbf{G}) \right),$$

delivers consistent parameter estimates (i.e., $\text{plim } \hat{\theta}_1^{GMM} = \theta_1$) for every positive definite weighting matrix Ω_1 , and is asymptotically normal.

Given that our estimator treats networks as the unit of observation, and under the assumption of independence and identically distributed networks in our framework, there is no need to adjust standard errors for cross-sectional dependence.

When utilizing solely the mean restrictions in the estimation process, the GMM estimator from Proposition 3 simplifies to a closed-form generalized least squares (GLS) estimator. This GLS estimator operates by using position-specific network-averaged outcomes, and individual and peer characteristics as its inputs.

Corollary 1. Suppose that Assumption 1 holds, that $\beta\gamma + \delta \neq 0$, and that the condition in Theorem 1 is satisfied. Then $\hat{\theta}_1^{GMM}$ is numerically equivalent to the GLS estimator $\hat{\theta}_1^{GLS} := (\bar{\mathbf{X}}^\top \Omega_1 \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}^\top \Omega_1 \bar{\mathbf{y}}$, where $\bar{\mathbf{y}} := \frac{1}{S} \sum_s \mathbf{y}_s$ and $\bar{\mathbf{X}} := \frac{1}{S} \sum_s \begin{bmatrix} \mathbf{1} & \tilde{\mathbf{x}}_s & \mathbf{G}\tilde{\mathbf{x}}_s & \mathbf{G}\mathbf{y}_s \end{bmatrix}$.

When the weighting matrix is the identity, GLS collapses to OLS on position-specific, network-averaged data: $\hat{\theta}_1^{OLS} := (\bar{\mathbf{X}}^\top \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}^\top \bar{\mathbf{y}}$. Averaging at the position level eliminates the sources of bias due to simultaneity and errors-in-variables, because both disturbances and measurement errors wash out within each position. This result is striking, and continues to hold even in the absence of errors-in-variables. It implies that, provided there is sufficient variation in characteristics across network positions, the parameters of a linear-in-means model with contextual and endogenous peer effects can be consistently estimated using OLS applied to the position-specific averages.

We now introduce a GMM estimator that instead exploits the conditional covariance restrictions in Equation (8). The associated conditional moment conditions are $\mathbb{E}[\mathbf{V}_{2,s}(\theta_2, \mathbf{G}) \mid \mathbf{G}] = \mathbf{0}$, where

$$\mathbf{V}_{2,s}(\theta_2, \mathbf{G}) := (\mathbf{I} - \beta\mathbf{G})\bar{\mathbf{y}}_s \bar{\tilde{\mathbf{x}}}_s^\top - (\gamma\mathbf{I} + \delta\mathbf{G}) \left(\bar{\tilde{\mathbf{x}}}_s \bar{\tilde{\mathbf{x}}}_s^\top - \sigma_u^2 \mathbf{I} \right),$$

with $\bar{\mathbf{y}}_s := \mathbf{y}_s - \frac{1}{S} \sum_s \mathbf{y}_s$ and $\bar{\tilde{\mathbf{x}}}_s := \tilde{\mathbf{x}}_s - \frac{1}{S} \sum_s \tilde{\mathbf{x}}_s$.

Proposition 4. Suppose that Assumptions 1 and 2 hold, that $\beta\gamma + \delta \neq 0$, and that the condition in Theorem 2 is satisfied. Let $\mathbf{v}_{2,s}(\theta_2, \mathbf{G}) := \text{vec}(\mathbf{V}_{2,s}(\theta_2, \mathbf{G}))$. Under standard

regularity conditions (see Appendix A.6), the GMM estimator

$$\hat{\boldsymbol{\theta}}_2^{GMM} := \arg \min_{\boldsymbol{\theta}_2} \left(\frac{1}{S} \sum_s \mathbf{v}_{2,s}(\boldsymbol{\theta}_2, \mathbf{G}) \right)^\top \boldsymbol{\Omega}_2 \left(\frac{1}{S} \sum_s \mathbf{v}_{2,s}(\boldsymbol{\theta}_2, \mathbf{G}) \right),$$

delivers consistent parameter estimates (i.e., $\text{plim } \hat{\boldsymbol{\theta}}_2^{GMM} = \boldsymbol{\theta}_2$) for every positive definite weighting matrix $\boldsymbol{\Omega}_2$, and is asymptotically normal.

Stochastic networks. We now consider the case where networks are stochastic. In this setting, aggregating the moment conditions across network positions (mean restrictions) and pairs of network positions (covariance restrictions) is often desirable. First, aggregation reduces the dimensionality of the system and mitigates noise in the moments.¹⁷ Second, aggregation can improve comparability of moment conditions derived from heterogeneous network structures.

Our approach follows Chamberlain (1987) and Newey (1993), who show that conditional moment restrictions can be transformed into unconditional ones by suitable linear transformations of the original moments. In our context, we start from the conditional moments

$$\mathbb{E}[\mathbf{v}_{m,s}(\boldsymbol{\theta}_m, \mathbf{G}) \mid \mathbf{G}] = \mathbf{0}, \quad m \in \{1, 2\},$$

which hold for each network \mathbf{G} . Let N_G denotes the number of individuals in \mathbf{G} and consider arbitrary aggregation matrices $\mathbf{H}_1(\mathbf{G})$ and $\mathbf{H}_2(\mathbf{G})$ of dimensions $N_G \times K_1$ and $N_G^2 \times K_2$, respectively. By the law of iterated expectations,

$$\mathbb{E}[\mathbf{H}_m(\mathbf{G})^\top \mathbf{v}_{m,s}(\boldsymbol{\theta}_m, \mathbf{G})] = \mathbb{E}[\mathbf{H}_m(\mathbf{G})^\top \mathbb{E}[\mathbf{v}_{m,s}(\boldsymbol{\theta}_m, \mathbf{G}) \mid \mathbf{G}]] = \mathbf{0}, \quad m \in \{1, 2\},$$

so $\mathbb{E}[\mathbf{H}_m(\mathbf{G})^\top \mathbf{v}_{m,s}(\boldsymbol{\theta}_m, \mathbf{G})] = \mathbf{0}$ provides valid unconditional moment conditions.

Aggregation occurs in two steps. Matrix $\mathbf{H}_1(\mathbf{G})$ first aggregates mean restrictions across positions within network \mathbf{G} into K_1 aggregated moments. These moments are then aggregated across networks. Similarly, Matrix $\mathbf{H}_2(\mathbf{G})$ first aggregates covariance restrictions across pairs of positions within network \mathbf{G} into K_2 aggregated moments, which are then aggregated across networks. In general, $\mathbf{H}_m(\mathbf{G})$ may depend on the interaction matrix \mathbf{G} , but it must not depend on observed individual characteristics, as this might introduce bias.¹⁸ In our Monte Carlo simulation in Sec-

¹⁷If a particular network position appears infrequently in the sample, the associated moments may be very noisy, so aggregation across positions is recommended. Moreover, for large or irregular networks, enumerating all positions can be computationally prohibitive.

¹⁸Chamberlain (1987) and Newey (1993) characterize optimal aggregation matrices, which depend on the inverse covariance matrix of the moment conditions and the Jacobian of those moments. In our setting, however, these objects are difficult to estimate reliably: they vary with the underlying

tion 5.3, we use aggregation matrices that are simple functions of \mathbf{G} .

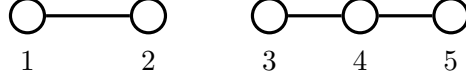


Figure 2: A pair and a 3-node line

To illustrate, suppose that a researcher observes many networks with the two structures depicted in Figure 2. In the population, the pair, denoted \mathbf{G}_1 , occurs with probability π , while the 3-node line, denoted \mathbf{G}_2 , occurs with probability $1 - \pi$. The researcher aggregates mean restrictions by individuals' degree and covariance restrictions by the distance between pairs. To aggregate the mean restrictions within and across structures, the researcher specifies the aggregation matrices

$$\mathbf{H}_1(\mathbf{G}_1) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}^\top, \quad \mathbf{H}_1(\mathbf{G}_2) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}^\top.$$

The aggregated moment conditions are then given by

$$\pi \mathbf{H}_1(\mathbf{G}_1)^\top \mathbb{E}[\mathbf{v}_{1,s}(\boldsymbol{\theta}_m, \mathbf{G}_1) \mid \mathbf{G}_1] + (1 - \pi) \mathbf{H}_1(\mathbf{G}_2)^\top \mathbb{E}[\mathbf{v}_{1,s}(\boldsymbol{\theta}_m, \mathbf{G}_2) \mid \mathbf{G}_2] = \mathbf{0},$$

where the first and second rows collect mean restrictions for individuals with degrees one and two, respectively. This procedure computes two aggregated moments, one that sums the mean restrictions over all individuals in positions 1, 2, 3 and 5 (degree one) and another that sums the mean restrictions over all individuals in position 4 (degree two).

Similarly, to aggregate the covariance restrictions, the researcher specifies

$$\mathbf{H}_2(\mathbf{G}_1) = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}^\top, \quad \mathbf{H}_2(\mathbf{G}_2) = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}^\top,$$

where the first, second, and third rows collect covariance restrictions for pairs at distances zero, one, and two, respectively. The aggregated moment conditions are then

$$\pi \mathbf{H}_2(\mathbf{G}_1)^\top \mathbb{E}[\mathbf{v}_{2,s}(\boldsymbol{\theta}_m, \mathbf{G}_1) \mid \mathbf{G}_1] + (1 - \pi) \mathbf{H}_2(\mathbf{G}_2)^\top \mathbb{E}[\mathbf{v}_{2,s}(\boldsymbol{\theta}_m, \mathbf{G}_2) \mid \mathbf{G}_2] = \mathbf{0}.$$

This yields three aggregated moments: a first one that sums the variance restrictions over all individuals, a second one that sums covariance restrictions over pairs of individuals in positions 12, 34 and 45 (distance one), and a third one that sums

network structure and are high dimensional, rendering their estimation impractical.

covariance restrictions over pair of individuals in position 35 (distance two). Although the two network structures differ in size and share no common positions, the aggregation renders the restrictions comparable across graphs.

5.2 2SLS estimators

The parameters of interest can also be recovered using a 2SLS estimator. This approach is simple to implement and readily accommodates network fixed effects and additional perfectly measured covariates. As discussed in Section 2, measurement error renders the individual observed characteristic and average observed characteristics of peers endogenous, while the average outcome of peers is endogenous by construction. With three endogenous regressors $(\tilde{x}, G\tilde{x}, Gy)$, we require at least three instruments.

Define the modified 2SLS estimator

$$\begin{bmatrix} \hat{\alpha}^{IV} & \hat{\gamma}^{IV} & \hat{\delta}^{IV} & \hat{\beta}^{IV} \end{bmatrix}^\top := (\mathbf{X}^\top \mathbf{P}_z \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{P}_z \mathbf{y}, \quad (9)$$

where \mathbf{Z}_m is the instrument matrix and $\mathbf{P}_z := \mathbf{Z}_m (\mathbf{Z}_m^\top \mathbf{Z}_m)^{-1} \mathbf{Z}_m^\top$ its projection matrix. Using \mathbf{P}_z accommodates overidentification (i.e., \mathbf{Z}_m may contain more than three instruments). It is consistent provided the instruments in \mathbf{Z}_m are valid and $\mathbf{X}^\top \mathbf{P}_z \mathbf{X}$ has full rank.

We explore two approaches to constructing relevant and valid instruments from within the model: using *network-lagged characteristics* and *network features*. These instruments' relevance relies on interdependence between characteristics and links. Alternatively, external information such as repeated measurements of the characteristics can be used to generate supplementary instruments. These three categories of instruments—network-lagged characteristics, network features, and external data—can be adopted interchangeably, provided that the rank condition is met.

Network-lagged characteristics. If the network is sufficiently sparse, one potential source of instrumental variables is network-lagged characteristics. The relevance and validity of these instruments rely on the correlation of individual characteristics across the network, under the assumption that measurement error is uncorrelated. When characteristics are correlated with links, the characteristics of peers at distances two or more can be used as instruments. This approach is reminiscent of [Griliches and Hausman \(1986\)](#), who employ time-lagged observations in the context of linear panel data models.

The interdependence of individual characteristics across the network can be harnessed through linear transformations of these characteristics. By restricting these

transformations to a specific class, the dependency between the instrument and the composite error term $\eta_s = \mathbf{e}_s - \gamma \mathbf{u}_s - \delta \mathbf{G}_s \mathbf{u}_s$ in Equation (3) is neutralized.

Proposition 5. *Suppose that Assumptions 1 and 2 hold. Then $\mathbf{W}_z \tilde{\mathbf{x}}$ is a valid instrument if $\text{trace}(\mathbf{W}_z) = 0$ and $\text{trace}(\mathbf{W}_z \mathbf{G}^\top) = 0$.*

This proposition confirms the validity of using of network-lagged characteristics as instruments. Let $\mathbf{A}^{(t)}$ denote the binary matrix indicating whether a pair is at distance t . For every distance $t \geq 2$, the transformation $\mathbf{W}_z = \mathbf{A}^{(t)}$ yields valid instruments, as we can see that diagonal elements of $\mathbf{A}^{(t)}$ and $\mathbf{A}^{(t)} \mathbf{G}^\top$ are all equal to zero.

Importantly, when regressors are measured with error, powers of the interaction matrix \mathbf{G} do not yield valid transformations of individual characteristics, in sharp contrast to the error-free case (Bramoullé et al., 2009). Instruments based on higher-order powers become invalid because paths with endpoints at distances zero and one induce a correlation between the instrument and the error term. For instance, $\mathbb{C}(G^2 \tilde{x}, \eta) = -\sigma_u^2 (\gamma \text{trace}(\mathbf{G}^2) + \delta \text{trace}(\mathbf{G}^2 \mathbf{G}^\top))$, which invalidates the instrument for most network structures whenever $\gamma \neq 0$ or $\delta \neq 0$.¹⁹

Network features. Another potential source of instruments comes from network structure. First consider instruments that exploit first-order moments. For example, if individuals with higher values of the underlying characteristic tend to have more connections, then degree is a relevant and valid instrument: $\mathbb{C}(d, \tilde{x}) \neq 0$ while $\mathbb{C}(d, \eta) = 0$. The same logic applies to other individual-level network statistics, such as local clustering, whenever these are correlated with the true characteristic. More generally, network structure can also generate instruments that work through higher-order moments, even when the associated network statistic is uncorrelated with the characteristic in levels. Consistent with this idea, we propose recentered, model-based instruments reminiscent of Lewbel (1997).

Proposition 6. *Suppose that Assumptions 1 and 2 hold. Let $t = t(\mathbf{G})$ be a network characteristic that varies at the individual level. Then instruments (i) t , (ii) $(t - \mathbb{E}[t])\tilde{x}$, and (iii) $d(t - \mathbb{E}[t])G\tilde{x}$ are valid. They are relevant if*

$$(i) \quad \mathbb{C}(t, x) \neq 0 \text{ or } \mathbb{C}(t, Gx) \neq 0,$$

$$(ii) \quad \mathbb{C}(t, x^2) - \mathbb{E}[x]\mathbb{C}(t, x) \neq 0 \text{ or } \mathbb{C}(t, xGx) - \mathbb{E}[Gx]\mathbb{C}(t, x) \neq 0,$$

$$(iii) \quad \mathbb{C}(t, d(Gx)^2) - \mathbb{E}[Gx]\mathbb{C}(t, dGx) \neq 0 \text{ or } \mathbb{C}(t, dxGx) - \mathbb{E}[x]\mathbb{C}(t, dGx) \neq 0,$$

respectively.

¹⁹For an undirected network, for example, $\text{trace}(\mathbf{G}^2) = \sum_{i,j} g_{ij}g_{ji} > 0$.

Proposition 6 highlights that a network statistic that is not a relevant instrument in the traditional sense may still be useful if it shifts second moments of the individual or average peer characteristic. To illustrate, consider an iso-correlational environment in which $\mathbb{E}[x_{i_0} \mid \mathbf{G}_0] = 0$, $\mathbb{V}(x_{i_0} \mid \mathbf{G}_0) = \sigma_x^2$, and $\mathbb{C}(x_{i_0}, x_{j_0} \mid \mathbf{G}_0) = \rho_{d(i_0, j_0)} \sigma_x^2$. Thus, the mean of x is zero and its variance is constant across network positions, while pairwise covariances depend only on network distance. It follows that t itself is not a relevant instrument, since $\mathbb{C}(t, x) = \mathbb{C}(t, Gx) = 0$. However, as shown in Online Appendix O.A.3, the first condition in (iii) of Proposition 6 yields

$$\mathbb{C}(d(t - \mathbb{E}[t])G\tilde{x}, G\tilde{x}) = \mathbb{C}(t, (d-1)c)\rho_1\sigma_x^2 + \mathbb{C}(t, (d-1)(1-c))\rho_2\sigma_x^2, \quad (10)$$

where $c_{i_0} = \frac{\sum_{j_0 \neq k_0 \in \mathcal{N}_{i_0}} A_{j_0 k_0}}{d_{i_0}(d_{i_0}-1)}$ is the local clustering coefficient.²⁰ If $\rho_1, \rho_2 \neq 0$, then $d(t - \mathbb{E}[t])G\tilde{x}$ may be relevant whenever $\mathbb{C}(t, (d-1)c) \neq 0$ or $\mathbb{C}(t, (d-1)(1-c)) \neq 0$. Because network statistics are often correlated in practice (Jackson & Rogers, 2007), the relevancy conditions in Proposition 6 are likely satisfied in many applications.

5.3 Monte Carlo evidence

We study the finite-sample performance of our GMM and 2SLS estimators using a Monte Carlo simulation. We consider a standard dyadic logit model of network formation in which links are conditionally independent across dyads. Despite its parsimony, this specification generates sufficient dependence between individual characteristics and network links to permit reliable recovery of the parameters of interest.

Setup. We consider a setup consisting of many small stochastic networks. In each simulation draw, we generate 500 networks according to a dyadic model of link formation. The number of individuals within every network is fixed at 20, a network size commonly encountered in empirical applications. The entire simulation procedure is repeated 500 times.

For each individual i in network s , we draw characteristics according to

$$x_{si} = \xi_s + \xi_{si},$$

where $\xi_s \sim N(10, 1)$ captures network-level heterogeneity and $\xi_{si} \sim N(0, 2)$ captures individual-level variation. This structure mimics selection into networks based on observable characteristics.

²⁰The local clustering coefficient measures how often an individual's friends are also friend with one another. It is the fraction of neighbor-pairs of i_0 that are themselves linked (triangles involving i_0 over potential triangles).

Links between individuals within a given network are formed through a directed dyadic logit specification.²¹ Specifically, the log odds of a i having an incoming link from j in network s is given by

$$\log \left(\frac{\Pr[a_{sij} = 1 \mid x_{si}, x_{sj}]}{\Pr[a_{sij} = 0 \mid x_{si}, x_{sj}]} \right) = \kappa + \mu x_{sj} - \nu |x_{si} - x_{sj}|,$$

where κ is an intercept governing the overall propensity to form links, μ captures the extent to which individuals with higher values of the characteristic attract more connections, and ν controls the degree of homophily. Larger values of ν imply that individuals with more dissimilar characteristics are less likely to form links, on average. In our simulations, we set $(\kappa, \mu, \nu) = (-9, 1, 1)$.

Finally, we assume that measurement error for characteristics is drawn from a normal distribution: i.e., $u_{si} \sim N(0, 0.3)$. This implies a noise-to-signal ratio of 10%.

Moments and instruments. To illustrate the performance of our GMM approach, we pool the moment restrictions from Propositions 3 and 4. Following the aggregation strategy in Section 5.1, we group conditional mean restrictions by degree into bins of width three (i.e., 0–3, 3–6, 6–9, and 9+) and group conditional covariance restrictions by network distance (0,1, and 2+). For the modified 2SLS estimator in Equation (9), we employ seven instruments to address the endogeneity of three regressors. In addition to average characteristics of individuals at network distance two, which are valid by Proposition 5, we construct six further instruments motivated by Proposition 6, based on degree and local clustering measures.²² Although richer sets of moments and instruments could in principle be constructed, our objective is not to optimize their choice, but to assess the performance of the theoretically motivated ones introduced above.

We contrast our estimates to those obtained from the naive, and biased, 2SLS estimator in Equation (2). This estimator is implemented by using only the average characteristics of individuals at network distance two as instrument. Because this specification employs a single instrument for three endogenous regressors, it is underidentified and therefore gives rise to an asymptotic bias, as discussed in Section 3.

Results. Figure 3 presents Monte Carlo evidence comparing the GMM and IV-based approaches. We set the model parameters to $(\alpha, \beta, \gamma, \delta) = (0, 1, 0.5, 0.5)$. The solid line corresponds to the naive 2SLS estimator, which is biased due to the presence of error-in-variables. The dotted line corresponds to the GMM estimator, while

²¹Logistic shocks are symmetric and reflect match-specific unobservables. If an individual is isolated, we assign that individual a random peer.

²²Specifically, we construct the following instruments: d , c , $(d - \mathbb{E}[d])\tilde{x}$, $(c - \mathbb{E}[c])\tilde{x}$, $d(d - \mathbb{E}[d])G\tilde{x}$, and $d(c - \mathbb{E}[c])G\tilde{x}$, where d denotes an individual's degree and c their local clustering coefficient.

the dashed line corresponds to the modified 2SLS estimator.

The top panel reports estimates of the contextual peer effect δ . The naive 2SLS estimator exhibits substantial upward bias, with an average estimate of 1.03 (s.d. 0.11), whereas the GMM and modified 2SLS the modified estimators are close to unbiased, with sample averages of 0.51 (s.d. 0.14) and 0.48 (s.d. 0.19), respectively. The bottom panel reports estimates of the endogenous peer effect β . In this case, the naive 2SLS estimator is biased downward, with an average of 0.34 (s.d. 0.04), while the GMM and modified 2SLS estimators yield average estimates of 0.50 (s.d. 0.04) and 0.51 (s.d. 0.06), respectively.

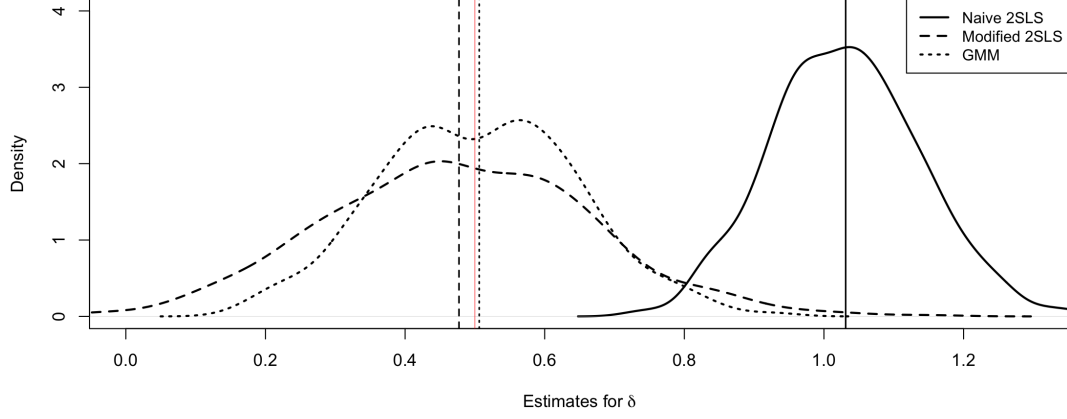
Overall, these results indicate that asymptotic bias in estimated peer effects is a first-order empirical concern. The naive estimate of the contextual peer effect is biased upward by more than 100%, while the naive estimate of the endogenous peer effect is biased downward by more than 30%. By contrast, our GMM and modified 2SLS estimators perform well in finite samples. Although the modified 2SLS estimator is more variable than the naive estimator—reflecting the treatment of three regressors as endogenous—its precision remains broadly comparable. The GMM estimator combines lower bias with higher precision.

6 Concluding remarks

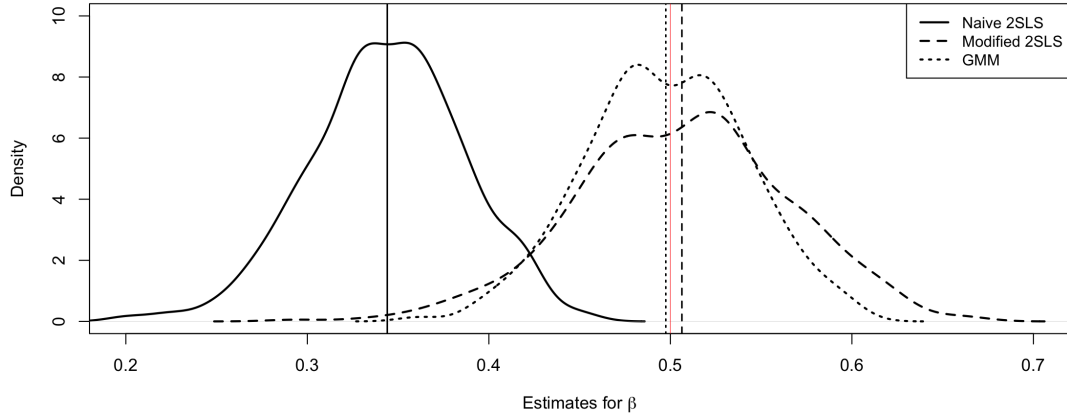
This paper studies errors-in-variables in the linear-in-means model of social interactions. We show that classical measurement error generally induces asymptotic bias in naive 2SLS estimates of peer effects. While this bias is driven by the interplay between individual characteristics and network formation, the same mechanism also offers an opportunity for identification and consistent estimation of the parameters of interest. We accordingly propose GMM and 2SLS estimators that are straightforward to implement, and demonstrate their performance in a Monte Carlo simulation.

Our analysis is a first step toward a general analysis of measurement error and peer effects in networks. We see three natural directions for future research. First, and given that the model is significantly overidentified, we believe that identification may be robust to relaxing Assumption 2. It would be interesting to extend our analysis to setups where the variance-covariance structure of measurement errors may depend on the network. Second, it would be interesting, and challenging, to extend our analysis to settings where measurement error is non-classical. Recent research by [Balestra, Eugster, and Puljic \(2023\)](#) has shown that when a binary characteristic is misclassified, an expansion bias can even arise under random group formation.²³

²³This additional expansion bias arises from the correlation between the average peer characteristic and the misclassification error in the individual characteristic.



(a) Contextual peer effect δ



(b) Endogenous peer effect β

Figure 3: Sampling distributions of the GMM and 2SLS estimators

Notes: Each panel displays kernel density estimates of Monte Carlo sampling distributions based on 500 replications for the GMM estimator and two 2SLS estimators. The GMM estimator follows Propositions 3 and 4 and exploits both conditional mean and covariance restrictions; it is implemented using a two-step GMM procedure. The naive 2SLS estimator is implemented using Equation (2), while the modified 2SLS estimator is implemented using Equation (9); the corresponding instrument set is described in Footnote 22. The horizontal axis reports estimated parameter values and the vertical axis reports density. Black vertical lines indicate the sample means of each estimator. For both peer-effects parameters shown, the true parameter value equals 0.5 and is indicated by a thin red vertical line. In each replication, outcomes are generated on 500 networks with 20 individuals each. Simulations are conducted under parameter values $(\alpha, \gamma, \delta, \beta, \sigma_u^2, \kappa, \mu, \nu) = (0, 1, 0.5, 0.5, 0.3, -9, 1, 1)$, which imply an average network degree of 8.22 and an average local clustering coefficient of 0.60. The implied noise-to-signal ratio is 10%.

We hypothesize that this additional bias will emerge in a network setting as well and that under non-random assignment, the usual asymptotic bias will reappear with qualitatively similar features to the bias studied here. Third, an exploration of the interplay between mismeasured characteristics and mismeasured links could contribute to a deeper and fuller understanding of the impact of measurement error on peer effect estimates. For instance, if homophily is correlated with friendship intensity, surveys that only sample a few best friends might deliver biased estimates of homophily. This, in turn, may interact with asymptotic bias due to mismeasured characteristics.

References

- Ammermueller, A., & Pischke, J. (2009). Peer effects in European primary schools: Evidence from the Progress in International Reading Literacy Study. *Journal of Labor Economics*, 27(3), 315–348.
- Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, 30, 98–108.
- Balestra, S., Eugster, B., & Puljic, F. (2023). *Misclassification in linear-in-means models: Theory and application to peer effects estimation* (Tech. Rep.). CEPR Press. (CEPR Discussion Paper No. 18342)
- Ben-Moshe, D. (2021). Identification of linear regressions with errors in all variables. *Econometric Theory*, 37(4), 633–663.
- Ben-Moshe, D., D’Haultfœuille, X., & Lewbel, A. (2017). Identification of additive and polynomial models of mismeasured regressors without instruments. *Journal of Econometrics*, 200(2), 207–222.
- Biggs, N. (1993). *Algebraic graph theory*. Cambridge: Cambridge University Press.
- Blume, L., Brock, W., Durlauf, S., & Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, 123(2), 444–496.
- Boucher, V., & Houndetoungan, A. (2025). Estimating peer effects using partial network data. *The Review of Economics and Statistics*, 1–48.
- Bound, J., Brown, C., & Mathiowetz, N. (2001). Measurement error in survey data. In J. Heckman & E. Leamer (Eds.), *Handbook of econometrics* (1st ed., Vol. 5, p. 3705–3843). Elsevier.
- Bramoullé, Y., Djebbari, H., & Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, 12(1), 603–629.
- Bramoullé, Y., Djebbari, H., & Fortin, B. (2009). Identification of peer effects through social networks. *Journal of Econometrics*, 150(1), 41–55.
- Caeyers, B., & Fafchamps, M. (2023). *Exclusion bias and the estimation of peer effects* (Tech. Rep.).
- Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge, UK: Cambridge University Press.
- Carrell, S., Fullerton, R. L., & West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3), 439–464.
- Carrell, S., Sacerdote, B., & West, J. (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica*, 81(3), 855–882.
- Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3), 305–334.
- Corno, L., La Ferrara, E., & Burns, J. (2022). Interaction, stereotypes, and performance: Evidence from south africa. *American Economic Review*, 112(12), 3848–75.
- Cragg, J. G. (1997). Using higher moments to estimate the simple errors-in-variables model. *The RAND Journal of Economics*, 28, S71–S91.
- Dagenais, M. G., & Dagenais, D. L. (1997). Higher moment estimators for linear regression models with errors in the variables. *Journal of Econometrics*, 76(1), 193–221.
- De Giorgi, G., Frederiksen, A., & Pistaferri, L. (2019). Consumption network effects. *The Review of Economic Studies*, 87(1), 130–163.
- de Paula, A. (2017). Econometrics of network models. In B. Honoré, A. Pakes, M. Piazzesi, & L. Samuelson (Eds.), *Advances in economics and econometrics: Eleventh world congress* (Vol. 1, p. 268–323). Cambridge University Press.
- de Paula, A., Rasul, I., & Souza, P. C. (2024). Identifying network ties from panel data: Theory and an application to tax competition. *The Review of Economic Studies*, rdae088.

- Erickson, T., & Whited, T. M. (2002). Two-step GMM estimation of the errors-in-variables model using high-order moments. *Econometric Theory*, 18, 776–799.
- Evdokimov, K. S., & Zeleneev, A. (2020). *Errors-in-variables in large nonlinear panel and network models* (Tech. Rep.).
- Feld, J., & Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2), 387–428.
- Feld, J., & Zölitz, U. (2022). The effect of higher-achieving peers on major choices and labor market outcomes. *Journal of Economic Behavior & Organization*, 196, 200–219.
- Gaviria, A., & Raphael, S. (2001, 05). School-based peer effects and juvenile behavior. *The Review of Economics and Statistics*, 83(2), 257–268.
- Gillen, B., Snowberg, E., & Yariv, L. (2019). Experimenting with measurement error: Techniques with applications to the caltech cohort study. *Journal of Political Economy*, 127(4), 1826–1863.
- Goldsmith-Pinkham, P., & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business Economic Statistics*, 31(3), 253–264.
- Greene, W. H. (2003). *Econometric analysis*. Upper Saddle River, NJ: Pearson Education Inc.
- Griffith, A. (2022a). Name your friends, but only five? The importance of censoring in peer effects estimates using social network data. *Journal of Labor Economics*, 40(4), 779–805.
- Griffith, A. (2022b). Random assignment with non-random peers: A structural approach to counterfactual treatment assessment. *The Review of Economics and Statistics*, 1–40.
- Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31, 93–118.
- Hsieh, C., & Lee, L. F. (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*, 31(2), 301–319.
- Jackson, M. O., & Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *The American Economic Review*, 97(3), 890–915.
- Klepper, S., & Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica*, 52(1), 163–183.
- Koopmans, T. C. (1937). *Linear regression analysis of economic time series*. Haarlem: De Erven F. Bohn N.V.
- Lavy, V., & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2), 1–33.
- Levi, M. (1973). Errors in the variables bias in the presence of correctly measured variables. *Econometrica*, 41(5), 985–86.
- Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and R&D. *Econometrica*, 65(5), 1201–1213.
- Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business Economic Statistics*, 30(1), 67–80.
- Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4), 835–903.
- Lewbel, A., Qu, X., & Tang, X. (2023). Ignoring measurement errors in social networks. *The Econometrics Journal*, 27(2), 171–187.
- Lin, Z., & Hu, Y. (2024). Binary choice with misclassification and social interactions, with an application to peer effects in attitude. *Journal of Econometrics*, 238(1), 105551.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, 60(3), 531–542.

- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- Moffitt, R. (2001). Policy interventions, low-level equilibria, and social interactions. In D. Steven & Y. Peyton (Eds.), *Social Dynamics*. The MIT Press.
- Newey, W. K. (1993). Efficient estimation of models with conditional moment restrictions. In (Vol. 11, p. 419-454). Elsevier.
- Nicoletti, C., Salvanes, K. G., & Tominey, E. (2018). The family peer effect on mothers' labor supply. *American Economic Journal: Applied Economics*, 10(3), 206-34.
- Patacchini, E., & Venanzoni, G. (2014). Peer effects in the demand for housing quality. *Journal of Urban Economics*, 83, 6-17.
- Paul, S., & Nath, S. (2024). *Spatial autoregressive model with measurement error in covariates*. Retrieved from <https://arxiv.org/abs/2402.04593>
- Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, 18(4), 375–389.
- Reiersøl, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, 9(1), 1–24.
- Rose, C. D. (2017). Identification of peer effects through social networks using variance restrictions. *The Econometrics Journal*, 20(3), S47-S60.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116(2), 681-704.
- Schennach, S. M. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica*, 75(1), 201–239.
- Trogon, J. G., Nonnemaker, J., & Pais, J. (2008). Peer effects in adolescent overweight. *Journal of Health Economics*, 27(5), 1388-1399.
- Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics* (Vol. 37). Amsterdam: Elsevier.

Appendix

Measurement Error and Peer Effects in Networks

A Proofs for results in the main text

A.1 Proof of Lemma 1

By the FWL theorem, we can partial out the intercept and work with demeaned variables. The resulting naive 2SLS estimator for (β, γ, δ) is

$$\begin{bmatrix} \hat{\gamma}^{IV} & \hat{\delta}^{IV} & \hat{\beta}^{IV} \end{bmatrix}^\top = (\mathbf{Z}_w^\top \mathbf{X}_w)^{-1} \mathbf{Z}_w^\top \mathbf{y}_w,$$

where $\mathbf{X}_w := \begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\tilde{\mathbf{x}})_w & (\mathbf{G}\mathbf{y})_w \end{bmatrix}$, $\mathbf{Z}_w := \begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\tilde{\mathbf{x}})_w & \mathbf{z}_w \end{bmatrix}$, and, for any $N \times 1$ vector \mathbf{a} , the demeaned version is $\mathbf{a}_w := (\mathbf{I}_N - \frac{1}{N}\mathbf{J}_N)\mathbf{a}$, with \mathbf{I}_N the $N \times N$ identity and \mathbf{J}_N the $N \times N$ matrix of ones. In this representation, all variables are expressed as deviations from their sample means.

Using the properties of probability limits, we have that

$$\text{plim} \begin{bmatrix} \hat{\gamma}^{IV} & \hat{\delta}^{IV} & \hat{\beta}^{IV} \end{bmatrix}^\top = \left(\text{plim} \frac{1}{N} \mathbf{Z}_w^\top \mathbf{X}_w \right)^{-1} \left(\text{plim} \frac{1}{N} \mathbf{Z}_w^\top \mathbf{y}_w \right),$$

and under Assumptions 1 and 2, we obtain

$$\begin{aligned} \text{plim} \frac{1}{N} \mathbf{Z}_w^\top \mathbf{X}_w &= \text{plim} \frac{1}{N} \begin{bmatrix} \mathbf{x}_w^\top \mathbf{x}_w & \mathbf{x}_w^\top (\mathbf{G}\mathbf{x})_w & \mathbf{x}_w^\top (\mathbf{G}\mathbf{y})_w \\ (\mathbf{G}\mathbf{x})_w^\top \mathbf{x}_w & (\mathbf{G}\mathbf{x})_w^\top (\mathbf{G}\mathbf{x})_w & (\mathbf{G}\mathbf{x})_w^\top (\mathbf{G}\mathbf{y})_w \\ \mathbf{z}_w^\top \mathbf{x}_w & \mathbf{z}_w^\top (\mathbf{G}\mathbf{x})_w & \mathbf{z}_w^\top (\mathbf{G}\mathbf{y})_w \end{bmatrix} \\ &\quad + \text{plim} \frac{1}{N} \begin{bmatrix} \mathbf{u}_w^\top \mathbf{u}_w & \mathbf{u}_w^\top (\mathbf{G}\mathbf{u})_w & 0 \\ (\mathbf{G}\mathbf{u})_w^\top \mathbf{u}_w & (\mathbf{G}\mathbf{u})_w^\top (\mathbf{G}\mathbf{u})_w & 0 \\ 0 & 0 & 0 \end{bmatrix}, \\ \text{plim} \frac{1}{N} \mathbf{Z}_w^\top \mathbf{y}_w &= \text{plim} \frac{1}{N} \begin{bmatrix} \mathbf{x}_w^\top \mathbf{x}_w & \mathbf{x}_w^\top (\mathbf{G}\mathbf{x})_w & \mathbf{x}_w^\top (\mathbf{G}\mathbf{y})_w \\ (\mathbf{G}\mathbf{x})_w^\top \mathbf{x}_w & (\mathbf{G}\mathbf{x})_w^\top (\mathbf{G}\mathbf{x})_w & (\mathbf{G}\mathbf{x})_w^\top (\mathbf{G}\mathbf{y})_w \\ \mathbf{z}_w^\top \mathbf{x}_w & \mathbf{z}_w^\top (\mathbf{G}\mathbf{x})_w & \mathbf{z}_w^\top (\mathbf{G}\mathbf{y})_w \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \\ \beta \end{bmatrix}, \end{aligned}$$

where the second equality follows by substituting Equation (1a). Evaluating these probability limits gives the desired result.

A.2 Proof of Proposition 1

The general case. Define \mathbf{K} as the Schur complement of s_{22} in $\mathbf{S} + \mathbf{\Sigma}$, i.e., $\mathbf{K} := \mathbf{S}_{11} + \mathbf{\Sigma}_{11} - \frac{1}{s_{22}}\mathbf{s}_{12}\mathbf{s}_{21}$. Using the formula for the inverse of a block matrix, we obtain

$$(\mathbf{S} + \mathbf{\Sigma})^{-1} = \begin{bmatrix} \mathbf{K}^{-1} & -\frac{1}{s_{22}}\mathbf{K}^{-1}\mathbf{s}_{12} \\ -\frac{1}{s_{22}}\mathbf{s}_{21}\mathbf{K} & \frac{1}{s_{22}} + \frac{1}{s_{22}^2}\mathbf{s}_{21}\mathbf{K}^{-1}\mathbf{s}_{12} \end{bmatrix},$$

such that for $\mathbf{M} := (\mathbf{S} + \mathbf{\Sigma})^{-1}\mathbf{S}$, we have

$$\mathbf{M} = \begin{bmatrix} \mathbf{K}^{-1} & -\frac{1}{s_{22}}\mathbf{K}^{-1}\mathbf{s}_{12} \\ -\frac{1}{s_{22}}\mathbf{s}_{21}\mathbf{K} & \frac{1}{s_{22}} + \frac{1}{s_{22}^2}\mathbf{s}_{21}\mathbf{K}^{-1}\mathbf{s}_{12} \end{bmatrix} \begin{bmatrix} \mathbf{S}_{11} & \mathbf{s}_{12} \\ \mathbf{s}_{21} & s_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{I} - \mathbf{K}^{-1}\mathbf{\Sigma}_{11} & \mathbf{0} \\ \frac{1}{s_{22}}\mathbf{s}_{21}\mathbf{K}^{-1}\mathbf{\Sigma}_{11} & 1 \end{bmatrix}.$$

Using the Sherman-Morrison formula, we obtain

$$\begin{aligned} \mathbf{K}^{-1} &= (\mathbf{S}_{11} + \mathbf{\Sigma}_{11})^{-1} + \frac{1}{s_{22} - \mathbf{s}_{21}(\mathbf{S}_{11} + \mathbf{\Sigma}_{11})^{-1}\mathbf{s}_{12}}(\mathbf{S}_{11} + \mathbf{\Sigma}_{11})^{-1}\mathbf{s}_{12}\mathbf{s}_{21}(\mathbf{S}_{11} + \mathbf{\Sigma}_{11})^{-1} \\ &= (\mathbf{S}_{11} + \mathbf{\Sigma}_{11})^{-1} + \frac{1}{s_{22} - \boldsymbol{\varphi}_z^\top \mathbf{s}_{12}}(\mathbf{S}_{11} + \mathbf{\Sigma}_{11})^{-1}\mathbf{s}_{12}\boldsymbol{\varphi}_z^\top, \end{aligned}$$

where we used $\boldsymbol{\varphi}_z^\top := \mathbf{s}_{21}(\mathbf{S}_{11} + \mathbf{\Sigma}_{11})^{-1}$. This vector can be interpreted as the probability limit of the coefficients on \tilde{x} and $G\tilde{x}$ obtained from regressing z on a constant, \tilde{x} , and $G\tilde{x}$. Therefore,

$$\begin{aligned} \frac{1}{s_{22}}\mathbf{s}_{21}\mathbf{K}^{-1}\mathbf{\Sigma}_{11} &= \frac{1}{s_{22} - \boldsymbol{\varphi}_z^\top \mathbf{s}_{12}}\boldsymbol{\varphi}_z^\top \mathbf{\Sigma}_{11} \\ &= \frac{1}{s_{22} - \boldsymbol{\varphi}_z^\top \mathbf{s}_{12}}\sigma_u^2 \begin{bmatrix} \varphi_{z,\tilde{x}} & h_0\varphi_{z,G\tilde{x}} \end{bmatrix}. \end{aligned}$$

If $\mathbb{C}(z, x) = \mathbb{C}(z, Gx) = 0$, then $\mathbf{s}_{21} = \mathbf{0}$, and hence $\varphi_{z,\tilde{x}} = \varphi_{z,G\tilde{x}} = 0$.

Independent characteristics and links. Under independence between characteristics and links, we have $\mathbf{s}_{21} = \mathbf{0}$, so that $\mathbf{K} = \mathbf{S}_{11} + \mathbf{\Sigma}_{11}$. In addition, $\mathbf{S}_{11} = \begin{bmatrix} \mathbb{V}(x) & 0 \\ 0 & \mathbb{V}(Gx) \end{bmatrix}$. We therefore obtain

$$\begin{aligned} \mathbf{I} - \mathbf{K}^{-1}\mathbf{\Sigma}_{11} &= \mathbf{I} - \begin{bmatrix} \mathbb{V}(x) + \sigma_u^2 & 0 \\ 0 & \mathbb{V}(Gx) + h_0\sigma_u^2 \end{bmatrix}^{-1} \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & h_0\sigma_u^2 \end{bmatrix} \\ &= \mathbf{I} - \begin{bmatrix} \frac{\sigma_u^2}{\mathbb{V}(x) + \sigma_u^2} & 0 \\ 0 & \frac{h_0\sigma_u^2}{\mathbb{V}(Gx) + h_0\sigma_u^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\mathbb{V}(x)}{\mathbb{V}(x) + \sigma_u^2} & 0 \\ 0 & \frac{\mathbb{V}(Gx)}{\mathbb{V}(Gx) + h_0\sigma_u^2} \end{bmatrix}. \end{aligned}$$

Similarly, $s_{21} = 0$ implies $\frac{1}{s_{22}}s_{21}\mathbf{K}^{-1}\Sigma_{11} = 0$.

A.3 Proof of Theorem 1

Consider two sets of parameters θ_1, θ'_1 leading to the conditional mean restrictions

$$\begin{aligned}\mathbb{E}[\mathbf{y}_0 \mid \mathbf{G}_0] &= (\mathbf{I} - \beta\mathbf{G}_0)^{-1}[\alpha\mathbf{1} + (\gamma\mathbf{I} + \delta\mathbf{G}_0)\mathbf{m}], \\ \mathbb{E}[\mathbf{y}_0 \mid \mathbf{G}_0] &= (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}[\alpha'\mathbf{1} + (\gamma'\mathbf{I} + \delta'\mathbf{G}_0)\mathbf{m}],\end{aligned}$$

or equivalently,

$$(\mathbf{I} - \beta\mathbf{G}_0)^{-1}[\alpha\mathbf{1} + (\gamma\mathbf{I} + \delta\mathbf{G}_0)\mathbf{m}] = (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}[\alpha'\mathbf{1} + (\gamma'\mathbf{I} + \delta'\mathbf{G}_0)\mathbf{m}].$$

Left-multiplying both sides by $(\mathbf{I} - \beta\mathbf{G}_0)(\mathbf{I} - \beta'\mathbf{G}_0)$ yields

$$(\mathbf{I} - \beta'\mathbf{G}_0)[\alpha\mathbf{1} + (\gamma\mathbf{I} + \delta\mathbf{G}_0)\mathbf{m}] = (\mathbf{I} - \beta\mathbf{G}_0)[\alpha'\mathbf{1} + (\gamma'\mathbf{I} + \delta'\mathbf{G}_0)\mathbf{m}],$$

where we made use of the push-through identity $\mathbf{G}_0(\mathbf{I} - \beta\mathbf{G}_0)^{-1} = (\mathbf{I} - \beta\mathbf{G}_0)^{-1}\mathbf{G}_0$. Rearranging gives $a_1\mathbf{1} + a_2\mathbf{m} + a_3\mathbf{G}_0\mathbf{m} + a_4\mathbf{G}_0^2\mathbf{m} = \mathbf{0}$, with coefficients

$$\begin{aligned}a_1 &= \alpha(1 - \beta') - \alpha'(1 - \beta), & a_2 &= \gamma - \gamma', \\ a_3 &= -(\beta'\gamma - \delta) + \beta\gamma' - \delta', & a_4 &= -\beta'\delta + \beta\delta'.\end{aligned}$$

(\implies) Suppose that $\mathbf{1}, \mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}$ are linearly independent. Using the fact that $a_2 = a_3 = a_4 = 0$, from the proof of Proposition 1 in [Bramoullé et al. \(2009\)](#), it follows that $(\beta, \gamma, \delta) = (\beta', \gamma', \delta')$ if $\beta\gamma + \delta \neq 0$. From $a_1 = 0$ it then follows that $\alpha = \alpha'$. (\impliedby) Suppose $\mathbf{1}, \mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}$ are linearly dependent. This implies that at least one of the coefficients a_1, a_2, a_3, a_4 is nonzero. One can confirm that $(\alpha, \beta, \gamma, \delta) = (2a_1, 0, 2a_2, 2a_3)$ and $(\alpha', \beta', \gamma', \delta') = (a_1, 0, a_2, a_3)$ is a feasible pairs of solutions.

A.4 Proof of Theorem 2

Consider two sets of parameters θ_2, θ'_2 leading to the conditional covariance restrictions

$$\begin{aligned}\mathbb{C}(\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) &= (\mathbf{I} - \beta\mathbf{G}_0)^{-1}(\gamma\mathbf{I} + \delta\mathbf{G}_0)(\mathbf{C} - \sigma_u^2\mathbf{I}), \\ \mathbb{C}(\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) &= (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}(\gamma'\mathbf{I} + \delta'\mathbf{G}_0)(\mathbf{C} - \sigma_u^{2'}\mathbf{I}),\end{aligned}$$

or equivalently,

$$(\mathbf{I} - \beta\mathbf{G}_0)^{-1}(\gamma\mathbf{I} + \delta\mathbf{G}_0)(\mathbf{C} - \sigma_u^2\mathbf{I}) = (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}(\gamma'\mathbf{I} + \delta'\mathbf{G}_0)(\mathbf{C} - \sigma_u^{2'}\mathbf{I}).$$

Left-multiplying both sides by $(\mathbf{I} - \beta \mathbf{G}_0)(\mathbf{I} - \beta' \mathbf{G}_0)$ gives

$$(\mathbf{I} - \beta' \mathbf{G}_0)(\gamma \mathbf{I} + \delta \mathbf{G}_0)(\mathbf{C} - \sigma_u^2 \mathbf{I}) = (\mathbf{I} - \beta \mathbf{G}_0)(\gamma' \mathbf{I} + \delta' \mathbf{G}_0)(\mathbf{C} - \sigma_u^{2'} \mathbf{I}).$$

Rearranging gives $a_1 \mathbf{I} + a_2 \mathbf{G}_0 + a_3 \mathbf{G}_0^2 + a_4 \mathbf{C} + a_5 \mathbf{G}_0 \mathbf{C} + a_6 \mathbf{G}_0^2 \mathbf{C} = \mathbf{0}$, with coefficients

$$\begin{aligned} a_1 &= -\gamma \sigma_u^2 + \gamma' \sigma_u^{2'}, & a_2 &= (\beta' \gamma - \delta) \sigma_u^2 - (\beta \gamma' - \delta') \sigma_u^{2'}, \\ a_3 &= \beta' \delta \sigma_u^2 - \beta \delta' \sigma_u^{2'}, & a_4 &= \gamma - \gamma', \\ a_5 &= -(\beta' \gamma - \delta) + \beta \gamma' - \delta', & a_6 &= -\beta' \delta + \beta \delta'. \end{aligned}$$

(\implies) Suppose that $\mathbf{I}, \mathbf{G}_0, \mathbf{G}_0^2, \mathbf{C}, \mathbf{G}_0 \mathbf{C}, \mathbf{G}_0^2 \mathbf{C}$ are linearly independent. Using the fact that $a_4 = a_5 = a_6 = 0$, from the proof of Proposition 1 in [Bramoullé et al. \(2009\)](#), it follows that $(\beta, \gamma, \delta) = (\beta', \gamma', \delta')$ if $\beta \gamma + \delta \neq 0$. From $a_1 = 0$ it then follows that $\sigma_u^2 = \sigma_u^{2'}$. (\impliedby) Suppose $\mathbf{I}, \mathbf{G}_0, \mathbf{G}_0^2, \mathbf{C}, \mathbf{G}_0 \mathbf{C}, \mathbf{G}_0^2 \mathbf{C}$ are linearly dependent. This implies that at least one of the coefficients $a_1, a_2, a_3, a_4, a_5, a_6$ is nonzero. One can confirm that $(\beta, \gamma, \delta, \sigma_u^2) = (0, 2a_4, 2a_5, 0)$ and $(\beta', \gamma', \delta', \sigma_u^{2'}) = (0, a_4, a_5, 0)$ is a feasible pair of solutions.

A.5 Proof of Proposition 2

We prove Proposition 2 through a series of lemmas. Let $\mathbf{T}(\check{\mathbf{m}}) := [\mathbf{1}, \check{\mathbf{m}}, \check{\mathbf{G}}_0 \check{\mathbf{m}}, \check{\mathbf{G}}_0^2 \check{\mathbf{m}}]$ for $\check{\mathbf{m}} \in \mathbb{C}^{\check{N}_0}$.

Lemma A.1. *If $\check{\mathbf{G}}_0$ has at least 4 distinct eigenvalues, there exists an $\check{\mathbf{m}}^* \in \mathbb{C}^{\check{N}_0}$ such that $\text{rank } \mathbf{T}(\check{\mathbf{m}}^*) = 4$.*

Proof. Consider 4 distinct eigenvalues $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ of $\check{\mathbf{G}}_0$, and denote their corresponding eigenvectors by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4 \in \mathbb{C}^{\check{N}_0}$. Since $\check{\mathbf{G}}_0 \mathbf{1} = \mathbf{1}$, $\lambda_1 = 1$ is an eigenvalue with eigenvector $\mathbf{1}$. Because the eigenvalues are pairwise distinct, their corresponding eigenvectors are linearly independent over \mathbb{C} .

Define $\mathcal{V} := \text{span}(\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$ and pick $\check{\mathbf{m}}^* = \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 \in \mathcal{V}$. Because $(\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$ forms a basis of \mathcal{V} , we can rewrite the $\check{N}_0 \times 3$ matrix $[\check{\mathbf{m}}^*, \check{\mathbf{G}}_0 \check{\mathbf{m}}^*, \check{\mathbf{G}}_0^2 \check{\mathbf{m}}^*]$ in terms of this basis. In coordinates relative to $(\mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4)$ we have:²⁴

$$V := \begin{bmatrix} 1 & \lambda_2 & \lambda_2^2 \\ 1 & \lambda_3 & \lambda_3^2 \\ 1 & \lambda_4 & \lambda_4^2 \end{bmatrix},$$

which is a Vandermonde matrix with $\det(V) = (\lambda_3 - \lambda_2)(\lambda_4 - \lambda_2)(\lambda_4 - \lambda_3) \neq 0$, since the eigenvalues are pairwise distinct. Therefore $\check{\mathbf{m}}^*, \check{\mathbf{G}}_0 \check{\mathbf{m}}^*, \check{\mathbf{G}}_0^2 \check{\mathbf{m}}^*$ are linearly

²⁴Note that $\check{\mathbf{G}}_0 \check{\mathbf{m}}^* = \lambda_2 \mathbf{v}_2 + \lambda_3 \mathbf{v}_3 + \lambda_4 \mathbf{v}_4$ and $\check{\mathbf{G}}_0^2 \check{\mathbf{m}}^* = \lambda_2^2 \mathbf{v}_2 + \lambda_3^2 \mathbf{v}_3 + \lambda_4^2 \mathbf{v}_4$

independent over \mathbb{C} and $\text{span}(\check{\mathbf{m}}^*, \check{\mathbf{G}}_0 \check{\mathbf{m}}^*, \check{\mathbf{G}}_0^2 \check{\mathbf{m}}^*) = \mathcal{V}$. Since $\mathbf{v}_1 = \mathbf{1} \notin \mathcal{V}$ we have that $\text{rank } \mathbf{T}(\check{\mathbf{m}}^*) = 4$. \square

Lemma A.2. *If there exists at least one $\check{\mathbf{m}}^* \in \mathbb{C}^{\tilde{N}_0}$ for which $\text{rank } \mathbf{T}(\check{\mathbf{m}}^*) = 4$ then the set of $\check{\mathbf{m}} \in \mathbb{R}^{\tilde{N}_0}$ for which $\text{rank } \mathbf{T}(\check{\mathbf{m}}) < 4$ has Lebesgue measure zero.*

Proof. Let $\Delta_1(\check{\mathbf{m}}), \dots, \Delta_K(\check{\mathbf{m}})$ be all 4×4 minors of $\mathbf{T}(\check{\mathbf{m}})$.²⁵ Then $\text{rank } \mathbf{T}(\check{\mathbf{m}}) < 4$ if and only if $\Delta_1(\check{\mathbf{m}}) = \dots = \Delta_K(\check{\mathbf{m}}) = 0$. Define $p(\check{\mathbf{m}}) := \sum_{k=1}^K |\Delta_k(\check{\mathbf{m}})|^2$, which is a real polynomial in the real and imaginary parts of $\check{\mathbf{m}}$.²⁶ Identification fails for the set of $\check{\mathbf{m}}$'s for which this polynomial is zero.

Since there exists an $\check{\mathbf{m}}^* \in \mathbb{C}^{\tilde{N}_0}$ such that $\text{rank } \mathbf{T}(\check{\mathbf{m}}^*) = 4$, we have $p(\check{\mathbf{m}}^*) > 0$. Hence p is a nontrivial (i.e., nonzero) polynomial. It is well known that the zero set of a nonzero real polynomial has Lebesgue measure zero. Viewing $\mathbb{C}^{\tilde{N}_0}$ as $\mathbb{R}^{2\tilde{N}_0}$, this implies that the zero set of p has Lebesgue measure zero in $\mathbb{R}^{2\tilde{N}_0}$. Its intersection with $\mathbb{R}^{\tilde{N}_0}$ therefore also has Lebesgue measure zero. \square

A.6 Regularity conditions for Proposition 3 and 4

Following [Cameron and Trivedi \(2005, p. 172–174\)](#), we state the main regularity conditions ensuring consistency and asymptotic normality of the GMM estimators.²⁷ Let $m \in \{1, 2\}$ index the two sets of moments, corresponding respectively to the mean and covariance restrictions, and let $\boldsymbol{\theta}_m^*$ denote the corresponding true parameter value. For each m , assume:

- (i) The model is correctly specified, i.e., $\mathbb{E}[\mathbf{v}_{m,s}(\boldsymbol{\theta}_m^*, \mathbf{G}) \mid \mathbf{G}] = \mathbf{0}$,
- (ii) The model is identified, i.e., the conditions in Theorem 1 (for $m = 1$) or Theorem 2 (for $m = 2$) hold,
- (iii) The Jacobian $\mathbb{E}[\nabla_{\boldsymbol{\theta}_m} \mathbf{v}_{m,s}(\boldsymbol{\theta}_m^*, \mathbf{G}) \mid \mathbf{G}]$ exists and is finite with full column rank,
- (iv) The moments are asymptotically normally distributed, i.e.,

$$\frac{1}{\sqrt{S}} \sum_s \mathbf{v}_{m,s}(\boldsymbol{\theta}_m^*, \mathbf{G}) \rightarrow_d N(\mathbf{0}, \boldsymbol{\Psi}_m(\boldsymbol{\theta}_m^*)),$$

where $\boldsymbol{\Psi}_m(\boldsymbol{\theta}_m^*) := \mathbb{E}[\mathbf{v}_{m,s}(\boldsymbol{\theta}_m^*, \mathbf{G}) \mathbf{v}_{m,s}(\boldsymbol{\theta}_m^*, \mathbf{G})^\top \mid \mathbf{G}]$,

- (v) The parameter space $\boldsymbol{\Theta}_m$ is compact and $\boldsymbol{\theta}_m^* \in \text{int}(\boldsymbol{\Theta}_m)$.

²⁵That is, the determinants of every combination of 4 rows of $\mathbf{T}(\check{\mathbf{m}})$.

²⁶We let $|\cdot|^2$ denote the squared modulus of a complex number. For instance, for $a + ib \in \mathbb{C}$, $|a + ib|^2 = a^2 + b^2$.

²⁷Additional standard regularity conditions guarantee a uniform law of large numbers for the sample moments and their Jacobian.

A.7 Proof of Corollary 1

The GMM estimate in Proposition 3 is numerically equivalent to the solution of the first order condition

$$\left(\frac{1}{S} \sum_s \nabla_{\theta_1} \mathbf{v}_{1,s}(\theta_1, \mathbf{G}) \right)^\top \Omega_1 \left(\frac{1}{S} \sum_s \mathbf{v}_{1,s}(\theta_1, \mathbf{G}) \right) = 0,$$

where $\mathbf{v}_{1,s}(\theta_1, \mathbf{G}) = \mathbf{y}_s - \mathbf{X}_s \theta_1$ and $\nabla_{\theta_1} \mathbf{v}_{1,s}(\theta_1, \mathbf{G}) = -\mathbf{X}_s$. By rearranging, it follows directly that $\hat{\theta}_1^{GMM} = (\bar{\mathbf{X}}^\top \Omega_1 \bar{\mathbf{X}})^{-1} \bar{\mathbf{X}}^\top \Omega_1 \bar{\mathbf{y}}$.

A.8 Proof of Proposition 5

A valid instrument satisfies $\mathbb{C}(W_z \tilde{x}, \eta) = \mathbb{C}(W_z \tilde{x}, e) - \gamma \mathbb{C}(W_z \tilde{x}, u) - \delta \mathbb{C}(W_z \tilde{x}, Gu) = 0$. Under conditional mean independence of the disturbance and Assumptions 1 and 2, it holds that

$$\begin{aligned} \mathbb{C}(W_z \tilde{x}, e) &= 0, \\ \mathbb{C}(W_z \tilde{x}, u) &= \mathbb{E}(\mathbf{u}^\top \mathbf{W}_z^\top \mathbf{u}) = \sigma_u^2 \text{trace}(\mathbf{W}_z), \\ \mathbb{C}(W_z \tilde{x}, Gu) &= \mathbb{E}(\mathbf{u}^\top \mathbf{W}_z^\top \mathbf{G} \mathbf{u}) = \sigma_u^2 \text{trace}(\mathbf{W}_z \mathbf{G}^\top). \end{aligned}$$

A.9 Proof of Proposition 6

By the mean independence condition of disturbances and by Assumption 1, $\mathbb{C}(t, \eta) = 0$ and t is a valid instrument. Next, we focus on the instruments $(t - \mathbb{E}[t])\tilde{x}$ and $d(t - \mathbb{E}[t])G\tilde{x}$.

Validity. We first show the validity of the proposed instruments. A valid instrument z should satisfy $\mathbb{C}(z, \eta) = \mathbb{C}(z, e - \gamma u - \delta Gu) = 0$. Consider the instrument $z = (t - \mathbb{E}[t])\tilde{x}$. We have that

$$\begin{aligned} \mathbb{C}(z, e) &= \mathbb{E}[z \mathbb{E}[e | z]] = 0, \\ \mathbb{C}(z, u) &= \mathbb{E}[(t - \mathbb{E}[t])(x + u)u] = \mathbb{E}[(t - \mathbb{E}[t])\mathbb{E}[u^2 | t]] = 0, \\ \mathbb{C}(z, Gu) &= \mathbb{E}[(t - \mathbb{E}[t])(x + u)Gu] = 0. \end{aligned}$$

Consider the instrument $z = d(t - \mathbb{E}[t])G\tilde{x}$. We have that

$$\begin{aligned} \mathbb{C}(z, e) &= \mathbb{E}[z \mathbb{E}[e | z]] = 0, \\ \mathbb{C}(z, u) &= \mathbb{E}[d(t - \mathbb{E}[t])(Gx + Gu)u] = 0, \\ \mathbb{C}(z, Gu) &= \mathbb{E}[d(t - \mathbb{E}[t])(Gx + Gu)Gu] = \mathbb{E}[d(t - \mathbb{E}[t])\mathbb{E}[(Gu)^2 | t, d]] = 0, \end{aligned}$$

where the last equality follows from the fact that $d\mathbb{E}[(Gu)^2 | t, d] = \sigma_u^2$.

Relevance. We now show the relevance of the proposed instruments. Instrument z is relevant if it satisfies $\mathbb{C}(z, \tilde{x}) \neq 0$ or $\mathbb{C}(z, G\tilde{x}) \neq 0$. Consider the instrument $z = (t - \mathbb{E}[t])\tilde{x}$. We have that

$$\begin{aligned}\mathbb{C}(z, \tilde{x}) &= \mathbb{C}(t\tilde{x}, \tilde{x}) - \mathbb{E}[t]\mathbb{V}(\tilde{x}) \\ &= \mathbb{E}[t\tilde{x}^2] - \mathbb{E}[t\tilde{x}]\mathbb{E}[\tilde{x}] - \mathbb{E}[t]\mathbb{E}[\tilde{x}^2] + \mathbb{E}[t]\mathbb{E}[\tilde{x}]^2 \\ &= \mathbb{C}(t, \tilde{x}^2) - \mathbb{E}[\tilde{x}]\mathbb{C}(t, \tilde{x}) \\ &= \mathbb{C}(t, x^2) - \mathbb{E}[x]\mathbb{C}(t, x),\end{aligned}$$

and

$$\begin{aligned}\mathbb{C}(z, G\tilde{x}) &= \mathbb{C}(t\tilde{x}, G\tilde{x}) - \mathbb{E}[t]\mathbb{C}(G\tilde{x}, G\tilde{x}) \\ &= \mathbb{E}[t\tilde{x}G\tilde{x}] - \mathbb{E}[t\tilde{x}]\mathbb{E}[G\tilde{x}] - \mathbb{E}[t]\mathbb{E}[\tilde{x}G\tilde{x}] + \mathbb{E}[t]\mathbb{E}[\tilde{x}]\mathbb{E}[G\tilde{x}] \\ &= \mathbb{C}(t, \tilde{x}G\tilde{x}) - \mathbb{E}[G\tilde{x}]\mathbb{C}(t, \tilde{x}) \\ &= \mathbb{C}(t, xGx) - \mathbb{E}[Gx]\mathbb{C}(t, x).\end{aligned}$$

Consider the instrument $z = d(t - \mathbb{E}[t])G\tilde{x}$. We have that

$$\begin{aligned}\mathbb{C}(z, \tilde{x}) &= \mathbb{C}(dtG\tilde{x}, \tilde{x}) - \mathbb{E}[t]\mathbb{C}(dG\tilde{x}, \tilde{x}) \\ &= \mathbb{C}(t, d\tilde{x}G\tilde{x}) - \mathbb{E}[\tilde{x}]\mathbb{C}(t, dG\tilde{x}) \\ &= \mathbb{C}(t, dxGx) - \mathbb{E}[x]\mathbb{C}(t, dGx),\end{aligned}$$

and

$$\begin{aligned}\mathbb{C}(z, G\tilde{x}) &= \mathbb{C}(dtG\tilde{x}, G\tilde{x}) - \mathbb{E}[t]\mathbb{C}(dG\tilde{x}, G\tilde{x}) \\ &= \mathbb{C}(t, d(G\tilde{x})^2) - \mathbb{E}[G\tilde{x}]\mathbb{C}(t, dG\tilde{x}) \\ &= \mathbb{C}(t, d(Gx)^2) - \mathbb{E}[Gx]\mathbb{C}(t, dGx).\end{aligned}$$

B Extensions

B.1 Network fixed effects

We consider an extended specification of the model that allows for network-specific fixed effects,

$$\mathbf{y}_s = \alpha_s \mathbf{1} + \gamma \mathbf{x}_s + \delta \mathbf{G}_s \mathbf{x}_s + \beta \mathbf{G}_s \mathbf{y}_s + \mathbf{e}_s, \quad (11a)$$

$$\tilde{\mathbf{x}}_s = \mathbf{x}_s + \mathbf{u}_s, \quad (11b)$$

with $\mathbb{E}(\mathbf{e}_s \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{u}_s) = \mathbf{0}$. The inclusion of the fixed effects α_s requires updated versions of Assumptions 1 and 2.

Assumption B.1. *The measurement errors satisfy:*

$$\mathbb{E}(u_{si} \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = 0.$$

Assumption B.2. *The measurement errors satisfy:*

$$\mathbb{E}(u_{si}^2 \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = \sigma_u^2, \quad \mathbb{E}(u_{si}u_{sj} \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = 0, \forall i \neq j.$$

Identification. We now state analogues of Theorems 1 and 2 for the model with network fixed effects. Akin to Bramoullé et al. (2009), the inclusion of network-specific fixed effects renders the identification conditions more demanding.

Theorem B.1. *Suppose that Assumption B.1 holds and that $\beta\gamma + \delta \neq 0$. The parameters β, γ, δ are identified from conditional mean restrictions if the vectors $\mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}, \mathbf{G}_0^3\mathbf{m}$ are linearly independent.*

Theorem B.2. *Suppose that Assumptions B.1 and B.2 hold and that $\beta\gamma + \delta \neq 0$. The parameters $\beta, \gamma, \delta, \sigma_u^2$ are identified from conditional covariance restrictions if the matrices $\mathbf{I}, \mathbf{G}_0, \mathbf{G}_0^2, \mathbf{G}_0^3, \mathbf{C}, \mathbf{G}_0\mathbf{C}, \mathbf{G}_0^2\mathbf{C}, \mathbf{G}_0^3\mathbf{C}$ are linearly independent.*

Estimation. 2SLS estimation proceeds analogously to the baseline specification without network-specific fixed effects. By contrast, GMM estimation with many network-specific fixed effects may be burdensome in practice. A simple alternative is to estimate the model using network-demeaned variables. This approach reduces the number of parameters entering the (nonlinear) objective functions and can substantially facilitate estimation.

Define the network-demeaning matrix $\mathbf{P}_s := \mathbf{I}_{N_s} - \frac{1}{N_s} \mathbf{J}_{N_s}$. Left-multiplying both

sides of Equation (11a) by \mathbf{P}_s yields the demeaned model

$$\mathbf{P}_s \mathbf{y}_s = \mathbf{P}_s [(\gamma \mathbf{I} + \delta \mathbf{G}_s) \mathbf{x}_s + \beta \mathbf{G}_s \mathbf{y}_s + \mathbf{e}_s],$$

which no longer contains the network-specific fixed effect.

From this demeaned model, under Assumption B.1, we obtain the conditional mean restriction

$$\mathbf{P}_0(\mathbf{I} - \beta \mathbf{G}_0) \mathbb{E}[\mathbf{y}_0 \mid \mathbf{G}_0] = \mathbf{P}_0(\gamma \mathbf{I} + \delta \mathbf{G}_0) \mathbb{E}[\tilde{\mathbf{x}}_0 \mid \mathbf{G}_0],$$

and hence the sample analogue for observation s

$$\mathbf{v}_{1,s}(\boldsymbol{\theta}_1, \mathbf{G}) := \mathbf{P}_s [(\mathbf{I} - \beta \mathbf{G}) \mathbf{y}_s - (\gamma \mathbf{I} + \delta \mathbf{G}) \tilde{\mathbf{x}}_s].$$

Likewise, under Assumptions B.1 and B.2, we obtain the conditional covariance restriction

$$\mathbf{P}_0(\mathbf{I} - \beta \mathbf{G}_0) \mathbb{C}(\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) \mathbf{P}_0^\top = \mathbf{P}_0(\gamma \mathbf{I} + \delta \mathbf{G}_0) (\mathbb{V}(\tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) - \sigma_u^2 \mathbf{I}) \mathbf{P}_0^\top,$$

and hence the sample analogue for observation s

$$\mathbf{V}_{2,s}(\boldsymbol{\theta}_2, \mathbf{G}) := \mathbf{P}_s \left[(\mathbf{I} - \beta \mathbf{G}) \bar{\mathbf{y}}_s \bar{\mathbf{x}}_s^\top - (\gamma \mathbf{I} + \delta \mathbf{G}) (\bar{\mathbf{x}}_s \bar{\mathbf{x}}_s^\top - \sigma_u^2 \mathbf{I}) \right] \mathbf{P}_s^\top.$$

Together, these alternative restrictions provide a potentially large system of equations in only four unknowns (i.e., $\beta, \gamma, \delta, \sigma_u^2$). GMM estimation then proceeds along the lines of Propositions 3 and 4 in Section 5.1.

As a final remark, researchers who wish to include a larger set of perfectly measured covariates can follow Erickson and Whited (2002), who propose a computationally convenient two-step procedure. In the first step, one partials out fixed effects and the additional perfectly measured covariates via auxiliary OLS regressions. In the second step, one applies GMM to the resulting residualised model.²⁸

B.2 Measurement error in outcomes

We now study an extended specification of the model that allows for mismeasured outcomes: i.e., the researcher observes $\tilde{\mathbf{y}}_s = \mathbf{y}_s + \mathbf{v}_s$. Together with Equations (1a)

²⁸This approach generalizes the demeaning procedure outlined above. In this setting, \mathbf{P}_s can be interpreted as the annihilator matrix from the first stage.

and (1b), this yields

$$\tilde{\mathbf{y}}_s = \alpha \mathbf{1} + \beta \mathbf{G}_s \tilde{\mathbf{y}}_s + \gamma \tilde{\mathbf{x}}_s + \delta \mathbf{G}_s \tilde{\mathbf{x}}_s + \underbrace{\mathbf{e}_s - \gamma \mathbf{u}_s - \delta \mathbf{G}_s \mathbf{u}_s + \mathbf{v}_s - \beta \mathbf{G}_s \mathbf{v}_s}_{:=\boldsymbol{\zeta}_s}.$$

The inclusion of measurement error in outcomes requires us to put structure on this additional source of measurement error, in the spirit of Assumptions 1 and 2.

Assumption B.3. *The measurement errors on outcomes satisfy:*

$$\mathbb{E}(v_{si} \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = 0.$$

Assumption B.4.

$$\mathbb{E}(u_{si}v_{si} \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = 0, \quad \mathbb{E}(u_{si}v_{sj} \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) = 0, \forall i \neq j.$$

Identification. We now state the extensions of Theorems 1 and 2. Under Assumptions B.3 and B.4, the same conditions for identification hold as in our baseline setup.

Theorem B.3. *Suppose that Assumptions 1 and B.3 hold and that $\beta\gamma + \delta \neq 0$. The parameters $\alpha, \beta, \gamma, \delta$ are identified from conditional mean restrictions if and only if the vectors $\mathbf{1}, \mathbf{m}, \mathbf{G}_0\mathbf{m}, \mathbf{G}_0^2\mathbf{m}$ are linearly independent.*

Theorem B.4. *Suppose that Assumptions 1, 2, B.3, and B.4 hold and that $\beta\gamma + \delta \neq 0$. The parameters $\beta, \gamma, \delta, \sigma_u^2$ are identified from conditional covariance restrictions if and only if the matrices $\mathbf{I}, \mathbf{G}_0, \mathbf{G}_0^2, \mathbf{C}, \mathbf{G}_0\mathbf{C}, \mathbf{G}_0^2\mathbf{C}$ are linearly independent.*

Estimation. Under Assumptions 1 and B.3, the GMM estimator based on the mean restrictions coincides with the estimator in Proposition 3. Under Assumptions 1, 2, B.3, and B.4, the GMM estimator based on the covariance restrictions coincides with the estimator in Proposition 4.

Regarding 2SLS estimation, we obtain the following result, which generalizes Proposition 5.

Proposition B.1. *Suppose that Assumptions 1, 2, B.3, and B.4 hold. Then $\mathbf{W}_z\tilde{\mathbf{x}}$ is a valid instrument if $\text{trace}(\mathbf{W}_z) = 0$ and $\text{trace}(\mathbf{W}_z\mathbf{G}^\top) = 0$.*

Online Appendix

Measurement Error and Peer Effects in Networks

O.A Additional results

O.A.1 Expansion bias without endogenous peer effects

Consider the model without endogenous peer effect,

$$\begin{aligned} \mathbf{y}_s &= \alpha \mathbf{1} + \gamma \mathbf{x}_s + \delta \mathbf{G}_s \mathbf{x}_s + \mathbf{e}_s, \\ \tilde{\mathbf{x}}_s &= \mathbf{x}_s + \mathbf{u}_s, \end{aligned}$$

whose parameters are estimated using the OLS estimator

$$\begin{bmatrix} \hat{\gamma}^{OLS} & \hat{\delta}^{OLS} \end{bmatrix}^\top := \left(\begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\tilde{\mathbf{x}})_w \end{bmatrix}^\top \begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\tilde{\mathbf{x}})_w \end{bmatrix} \right)^{-1} \begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\tilde{\mathbf{x}})_w \end{bmatrix}^\top \mathbf{y}_w,$$

where, for any $N \times 1$ vector \mathbf{a} , the demeaned version is $\mathbf{a}_w := (\mathbf{I}_N - \frac{1}{N} \mathbf{J}_N) \mathbf{a}$, with \mathbf{I}_N the $N \times N$ identity and \mathbf{J}_N the $N \times N$ matrix of ones.

Similar to Lemma 1, under Assumptions 1 and 2 we have that

$$\text{plim} \begin{bmatrix} \hat{\gamma}^{OLS} \\ \hat{\delta}^{OLS} \end{bmatrix} = (\mathbf{S} + \mathbf{\Sigma})^{-1} \mathbf{S} \begin{bmatrix} \gamma \\ \delta \end{bmatrix},$$

where

$$\mathbf{S} := \begin{bmatrix} \mathbb{V}(x) & \mathbb{C}(x, Gx) \\ \mathbb{C}(Gx, x) & \mathbb{V}(Gx) \end{bmatrix}, \quad \mathbf{\Sigma} := \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & h_0 \sigma_u^2 \end{bmatrix}.$$

Therefore, it holds that

$$\begin{aligned} \text{plim} \begin{bmatrix} \hat{\gamma}^{OLS} \\ \hat{\delta}^{OLS} \end{bmatrix} &= \frac{1}{D^{OLS}} \begin{bmatrix} \mathbb{V}(Gx) + h_0 \sigma_u^2 & -\mathbb{C}(Gx, x) \\ -\mathbb{C}(x, Gx) & \mathbb{V}(x) + \sigma_u^2 \end{bmatrix} \begin{bmatrix} \mathbb{V}(x) & \mathbb{C}(x, Gx) \\ \mathbb{C}(Gx, x) & \mathbb{V}(Gx) \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix} \\ &= \frac{1}{D^{OLS}} \begin{bmatrix} \det(\mathbf{S}) + h_0 \sigma_u^2 \mathbb{V}(x) & h_0 \sigma_u^2 \mathbb{C}(x, Gx) \\ \sigma_u^2 \mathbb{C}(x, Gx) & \det(\mathbf{S}) + \sigma_u^2 \mathbb{V}(Gx) \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix}, \end{aligned}$$

where $D^{OLS} := \det(\mathbf{S}) + h_0 \sigma_u^2 \mathbb{V}(x) + \sigma_u^2 \mathbb{V}(Gx) + h_0 \sigma_u^4$. Matrix multiplication yields Equation (5). Observe that $\det(\mathbf{S}) = \mathbb{V}(x) \mathbb{V}(Gx) - \mathbb{C}(x, Gx)^2 \geq 0$.

O.A.2 Expansion bias without contextual peer effects

Consider the model without contextual peer effect,

$$\begin{aligned} \mathbf{y}_s &= \alpha \mathbf{1} + \gamma \mathbf{x}_s + \beta \mathbf{G}_s \mathbf{y}_s + \mathbf{e}_s, \\ \tilde{\mathbf{x}}_s &= \mathbf{x}_s + \mathbf{u}_s, \end{aligned}$$

whose parameters are estimated using the IV estimator

$$\begin{bmatrix} \hat{\gamma}^{IV} & \hat{\beta}^{IV} \end{bmatrix}^\top := \left(\begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\tilde{\mathbf{x}})_w \end{bmatrix}^\top \begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\mathbf{y})_w \end{bmatrix} \right)^{-1} \begin{bmatrix} \tilde{\mathbf{x}}_w & (\mathbf{G}\tilde{\mathbf{x}})_w \end{bmatrix}^\top \mathbf{y}_w,$$

where, for any $N \times 1$ vector \mathbf{a} , the demeaned version is $\mathbf{a}_w := (\mathbf{I}_N - \frac{1}{N} \mathbf{J}_N) \mathbf{a}$, with \mathbf{I}_N the $N \times N$ identity and \mathbf{J}_N the $N \times N$ matrix of ones.

Similar to Lemma 1, under Assumptions 1 and 2 we have that

$$\text{plim} \begin{bmatrix} \hat{\gamma}^{IV} \\ \hat{\beta}^{IV} \end{bmatrix} = (\mathbf{S} + \mathbf{\Sigma})^{-1} \mathbf{S} \begin{bmatrix} \gamma \\ \beta \end{bmatrix},$$

where

$$\mathbf{S} := \begin{bmatrix} \mathbb{V}(x) & \mathbb{C}(x, Gy) \\ \mathbb{C}(Gx, x) & \mathbb{C}(Gx, Gy) \end{bmatrix}, \quad \mathbf{\Sigma} := \begin{bmatrix} \sigma_u^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

Therefore, it holds that

$$\begin{aligned} \text{plim} \begin{bmatrix} \hat{\gamma}^{IV} \\ \hat{\beta}^{IV} \end{bmatrix} &= \frac{1}{D^{IV}} \begin{bmatrix} \mathbb{C}(Gx, Gy) & -\mathbb{C}(x, Gy) \\ -\mathbb{C}(Gx, x) & \mathbb{V}(x) + \sigma_u^2 \end{bmatrix} \begin{bmatrix} \mathbb{V}(x) & \mathbb{C}(x, Gy) \\ \mathbb{C}(Gx, x) & \mathbb{C}(Gx, Gy) \end{bmatrix} \begin{bmatrix} \gamma \\ \beta \end{bmatrix} \\ &= \frac{1}{D^{IV}} \begin{bmatrix} \mathbb{V}(x)\mathbb{C}(Gx, Gy) - \mathbb{C}(Gx, x)\mathbb{C}(x, Gy) & 0 \\ \sigma_u^2 \mathbb{C}(Gx, x) & D^{IV} \end{bmatrix} \begin{bmatrix} \gamma \\ \beta \end{bmatrix}, \end{aligned}$$

where $D^{IV} = (\mathbb{V}(x) + \sigma_u^2)\mathbb{C}(Gx, Gy) - \mathbb{C}(Gx, x)\mathbb{C}(x, Gy)$. Matrix multiplication yields Equation (6).

O.A.3 Proof of Equation (10)

The condition $\mathbb{E}[x_{i_0} \mid \mathbf{G}_0] = 0$ implies that $\mathbb{C}(t, x) = \mathbb{C}(t, Gx) = \mathbb{E}[Gx] = 0$. Therefore, by Proposition 6, we have that $\mathbb{C}(d(t - \mathbb{E}[t])G\tilde{x}, G\tilde{x}) = \mathbb{C}(t, d(Gx)^2)$. Observe that

$$\begin{aligned} \mathbb{C}(t, d(Gx)^2) &= \left(\text{plim} \frac{1}{N} \sum_i t_i d_i \sum_{j,k} g_{ij} g_{ik} x_j x_k \right) \\ &\quad - \left(\text{plim} \frac{1}{N} \sum_i t_i \right) \left(\text{plim} \frac{1}{N} \sum_i d_i \sum_{j,k} g_{ij} g_{ik} x_j x_k \right), \end{aligned}$$

where, due to $\mathbb{V}(x_{i_0} \mid \mathbf{G}_0) = \sigma_x^2$ and $\mathbb{C}(x_{i_0}, x_{j_0} \mid \mathbf{G}_0) = \rho_{d(i_0, j_0)} \sigma_x^2$,

$$\begin{aligned} \text{plim } \frac{1}{N} \sum_i t_i d_i \sum_{j,k} g_{ij} g_{ik} x_j x_k &= \frac{1}{N_0} \sum_{i_0} t_{i_0} d_{i_0} \left(\frac{1}{d_{i_0}} + \frac{(d_{i_0} - 1)c_{i_0}}{d_{i_0}} \rho_1 + \frac{(d_{i_0} - 1)(1 - c_{i_0})}{d_{i_0}} \rho_2 \right) \sigma_x^2, \\ \text{plim } \frac{1}{N} \sum_i d_i \sum_{j,k} g_{ij} g_{ik} x_j x_k &= \frac{1}{N_0} \sum_{i_0} d_{i_0} \left(\frac{1}{d_{i_0}} + \frac{(d_{i_0} - 1)c_{i_0}}{d_{i_0}} \rho_1 + \frac{(d_{i_0} - 1)(1 - c_{i_0})}{d_{i_0}} \rho_2 \right) \sigma_x^2. \end{aligned}$$

Combined with $\text{plim } \frac{1}{N} \sum_i t_i = \frac{1}{N_0} \sum_{i_0} t_{i_0}$, collecting terms yields Equation (10).

O.B Omitted proofs

O.B.1 Network fixed effects

O.B.1.1 Proof of Theorem B.1

Left-multiplying both sides of the conditional mean restrictions by the local differencing matrix $(\mathbf{I} - \mathbf{G}_0)$ gives

$$\begin{aligned}\mathbb{E}[(\mathbf{I} - \mathbf{G}_0)\mathbf{y}_0 \mid \mathbf{G}_0] &= (\mathbf{I} - \beta\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma\mathbf{I} + \delta\mathbf{G}_0)\mathbf{m}, \\ \mathbb{E}[(\mathbf{I} - \mathbf{G}_0)\mathbf{y}_0 \mid \mathbf{G}_0] &= (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma'\mathbf{I} + \delta'\mathbf{G}_0)\mathbf{m},\end{aligned}$$

where we made use of the push-through identity. Equivalently,

$$(\mathbf{I} - \beta\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma\mathbf{I} + \delta\mathbf{G}_0)\mathbf{m} = (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma'\mathbf{I} + \delta'\mathbf{G}_0)\mathbf{m}.$$

Left-multiplying both sides by $(\mathbf{I} - \beta\mathbf{G}_0)(\mathbf{I} - \beta'\mathbf{G}_0)$ yields

$$(\mathbf{I} - \beta'\mathbf{G}_0)(\mathbf{I} - \mathbf{G}_0)(\gamma\mathbf{I} + \delta\mathbf{G}_0)\mathbf{m} = (\mathbf{I} - \beta\mathbf{G}_0)(\mathbf{I} - \mathbf{G}_0)(\gamma'\mathbf{I} + \delta'\mathbf{G}_0)\mathbf{m}.$$

Rearranging gives $a_1\mathbf{m} + a_2\mathbf{G}_0\mathbf{m} + a_3\mathbf{G}_0^2\mathbf{m} + a_4\mathbf{G}_0^3\mathbf{m} = \mathbf{0}$, with coefficients

$$\begin{aligned}a_1 &= \gamma - \gamma', & a_2 &= -(1 + \beta')\gamma + \delta + (1 + \beta)\gamma' - \delta', \\ a_3 &= \beta'\gamma - \delta(1 + \beta') - \beta\gamma' + \delta'(1 + \beta), & a_4 &= \beta'\delta - \beta\delta'.$$

(\implies) Suppose that \mathbf{m} , $\mathbf{G}_0\mathbf{m}$, $\mathbf{G}_0^2\mathbf{m}$, $\mathbf{G}_0^3\mathbf{m}$ are linearly independent. Using the fact that $a_1 = a_2 = a_4 = 0$, from the proof of Proposition 1 in [Bramoullé et al. \(2009\)](#), it follows that $(\beta, \gamma, \delta) = (\beta', \gamma', \delta')$ if $\beta\gamma + \delta \neq 0$.

O.B.1.2 Proof of Theorem B.2

Left-multiplying both sides of the conditional covariance restrictions by the local differencing matrix $(\mathbf{I} - \mathbf{G}_0)$ gives

$$\begin{aligned}\mathbb{C}((\mathbf{I} - \mathbf{G}_0)\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) &= (\mathbf{I} - \beta\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma\mathbf{I} + \delta\mathbf{G}_0)(\mathbf{C} - \sigma_u^2\mathbf{I}), \\ \mathbb{C}((\mathbf{I} - \mathbf{G}_0)\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) &= (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma'\mathbf{I} + \delta'\mathbf{G}_0)(\mathbf{C} - \sigma_u^2\mathbf{I}),\end{aligned}$$

where we made use of the push-through identity. Equivalently,

$$(\mathbf{I} - \beta\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma\mathbf{I} + \delta\mathbf{G}_0)(\mathbf{C} - \sigma_u^2\mathbf{I}) = (\mathbf{I} - \beta'\mathbf{G}_0)^{-1}(\mathbf{I} - \mathbf{G}_0)(\gamma'\mathbf{I} + \delta'\mathbf{G}_0)(\mathbf{C} - \sigma_u^2\mathbf{I}).$$

Left-multiplying both sides by $(\mathbf{I} - \beta \mathbf{G}_0)(\mathbf{I} - \beta' \mathbf{G}_0)$ gives

$$(\mathbf{I} - \beta' \mathbf{G}_0)(\mathbf{I} - \mathbf{G}_0)(\gamma \mathbf{I} + \delta \mathbf{G}_0)(\mathbf{C} - \sigma_u^2 \mathbf{I}) = (\mathbf{I} - \beta \mathbf{G}_0)(\mathbf{I} - \mathbf{G}_0)(\gamma' \mathbf{I} + \delta' \mathbf{G}_0)(\mathbf{C} - \sigma_u^{2'} \mathbf{I}).$$

Rearranging gives $a_1 \mathbf{I} + a_2 \mathbf{G}_0 + a_3 \mathbf{G}_0^2 + a_4 \mathbf{G}_0^3 + a_5 \mathbf{C} + a_6 \mathbf{G}_0 \mathbf{C} + a_7 \mathbf{G}_0^2 \mathbf{C} + a_8 \mathbf{G}_0^3 \mathbf{C} = \mathbf{0}$, with coefficients

$$\begin{aligned} a_1 &= -\sigma_u^2 \gamma + \sigma_u^{2'} \gamma', \\ a_2 &= -\sigma_u^2 (\delta - \gamma(1 + \beta')) + \sigma_u^{2'} (\delta' - \gamma'(1 + \beta)), \\ a_3 &= -\sigma_u^2 (\gamma \beta' - (1 + \beta') \delta) + \sigma_u^{2'} (\gamma' \beta - (1 + \beta) \delta'), \\ a_4 &= -\sigma_u^2 \beta' \delta + \sigma_u^{2'} \beta \delta', \\ a_5 &= \gamma - \gamma', \\ a_6 &= \delta - \gamma(1 + \beta') - \delta' + \gamma'(1 + \beta), \\ a_7 &= \gamma \beta' - (1 + \beta') \delta - \gamma' \beta + (1 + \beta) \delta', \\ a_8 &= \beta' \delta - \beta \delta'. \end{aligned}$$

(\implies) Suppose that $\mathbf{I}, \mathbf{G}_0, \mathbf{G}_0^2, \mathbf{G}_0^3, \mathbf{C}, \mathbf{G}_0 \mathbf{C}, \mathbf{G}_0^2 \mathbf{C}, \mathbf{G}_0^3 \mathbf{C}$ are linearly independent. Using the fact that $a_5 = a_6 = a_8 = 0$, from the proof of Proposition 1 in [Bramoullé et al. \(2009\)](#), it follows that $(\beta, \gamma, \delta) = (\beta', \gamma', \delta')$ if $\beta\gamma + \delta \neq 0$. From $a_1 = 0$ it then follows that $\sigma_u^2 = \sigma_u^{2'}$.

O.B.2 Mismeasured outcome variable

O.B.2.1 Proof of Theorem B.3

Under Assumption B.3, we have

$$\mathbb{E}[\tilde{\mathbf{y}}_0 \mid \mathbf{G}_0] = \mathbb{E}[\mathbf{y}_0 \mid \mathbf{G}_0] + \mathbb{E}[\mathbf{v}_0 \mid \mathbf{G}_0] = \mathbb{E}[\mathbf{y}_0 \mid \mathbf{G}_0],$$

since $\mathbb{E}[\mathbf{v}_0 \mid \mathbf{G}_0] = \mathbf{0}$ by Assumption B.3. Identification then proceeds as in Theorem 1, based on the conditional mean restriction

$$\mathbb{E}[\tilde{\mathbf{y}}_0 \mid \mathbf{G}_0] = (\mathbf{I} - \beta \mathbf{G}_0)^{-1} [\alpha \mathbf{1} + (\gamma \mathbf{I} + \delta \mathbf{G}_0) \mathbf{m}].$$

O.B.2.2 Proof of Theorem B.4

Under Assumptions B.3 and B.4, we have

$$\mathbb{C}(\tilde{\mathbf{y}}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) = \mathbb{C}(\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) + \mathbb{C}(\mathbf{v}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) = \mathbb{C}(\mathbf{y}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0),$$

since $\mathbb{C}(\mathbf{v}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) = \mathbf{0}$ by Assumption B.4. Identification then proceeds as in Theorem 2, based on the conditional covariance restriction

$$\mathbb{C}(\tilde{\mathbf{y}}_0, \tilde{\mathbf{x}}_0 \mid \mathbf{G}_0) = (\mathbf{I} - \beta \mathbf{G}_0)^{-1}(\gamma \mathbf{I} + \delta \mathbf{G}_0)(\mathbf{C} - \sigma_u^2 \mathbf{I}).$$

O.B.2.3 Proof of Proposition B.1

The instrument $W_z \tilde{x}$ is valid when

$$\mathbb{C}(W_z \tilde{x}, \zeta) = \mathbb{C}(W_z \tilde{x}, e - \gamma u - \delta Gu + v - \beta Gv) = 0.$$

From the proof of Proposition 5, we have already established that, under Assumptions 1 and 2, $\mathbb{C}(W_z \tilde{x}, e) = \mathbb{C}(W_z \tilde{x}, u) = \mathbb{C}(W_z \tilde{x}, Gu) = 0$. Assumptions B.3 and B.4 additionally imply $\mathbb{C}(W_z \tilde{x}, v) = \mathbb{C}(W_z \tilde{x}, Gv) = 0$. It follows that $\mathbb{C}(W_z \tilde{x}, \zeta) = 0$.