Measurement Error and Peer Effects in Networks

Yann Bramoullé

Aix-Marseille School of Economics

Aix-Marseille University & CNRS

yann.bramoulle@univ-amu.fr

Sebastiaan Maes

Department of Economics

University of Antwerp

sebastiaan.maes@uantwerpen.be

January 11, 2024

[preliminary; do quote]

Abstract

In many practical applications, only noisy proxies for the true regressors are available, which typically induces an attenuation bias in OLS estimates. In the linear-in-means model, however, the estimates for the peer effect might be inflated, potentially leading to false positives. This paper explores how this expansion bias depends on the structure of the underlying social network and demonstrates how this network structure can facilitate identification without the need for additional external information. Based on these identification results, we present consistent GMM and IV 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, errors-in-variables, measurement error, expansion bias

JEL codes: C31, C36

*Yann acknowledges financial support from French National Research Agency Grant ANR-17-EURE-0020 and the Excellence Initiative of Aix-Marseille University, A*MIDEX. Sebastiaan benefited from doctoral and postdoctoral fellowships of the Research Foundation Flanders (grants 11F8919N and 12C8623N). Part of this research has been conducted in the course of a visit to Aix-Marseille School of Economics, whose hospitality is gratefully acknowledged. We thank Vincent Boucher, Sam Cosaert, Geert Dhaene, and Marcel Fafchamps, as well as conference and seminar participants in Antwerp, Barcelona (ESEM), Manchester (EWMES), Oslo (IAAE), and Québec for useful comments and suggestions.

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 for 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. Despite methodological advances, guaranteeing the reliability of peer effect estimates remains challenging. Addressing multiple endogeneity issues without resolving them all still leaves researchers distant from a causal interpretation.

This paper focuses on a critical but understudied issue affecting the reliability of peer effect estimates: measurement error in individual characteristics. This is likely a first-order empirical issue in applications based on survey data. Surprisingly, however, this issue has remained a blind spot in 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 we will comprehensively review it below. In short, this issue was identified by Moffitt (2001), emphasized 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 classrooms, and are affected by everyone in their group and by no one outside of it. By contrast, many recent studies of peer effects consider richer network interactions. Our paper aims to address this gap by studying measurement error in the context of 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 of peer effects. An individual's outcome is affected by their own characteristic and by the average char-

[1]For 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).

[2]The literature on peer effects in networks has grown fast in the past 15 years, see Bramoullé, Djebbari, and Fortin (2020) for a review. See, e.g., de Paula, Rasul, and Souza (2019); Griffith (2022a); Boucher and Houndetoungan (2023); Lewbel, Qu, and Tang (2023) for models of peer effects with imperfectly known peers.

acteristic of their peers. A significant challenge arises when researchers only observe a noisy proxy of the characteristic. There is then measurement error in both own and average peer characteristic and errors in the two variables are related through the structure of the model.[3] We adopt a many-networks asymptotic framework and consider econometric specifications without or with network fixed effects. Our investigation revolves around two central questions: First, under what conditions do we observe asymptotic biases in peer effect estimates? Second, what strategies can researchers employ to mitigate these biases?

Consistent with existing research, our findings indicate that measurement error typically leads to an *expansion bias* in peer effect estimates. Defying conventional wisdom, estimates are then artificially inflated rather than attenuated. The presence and extent of this expansion bias crucially depend on the structure of the network and on the interplay between network links and individual characteristics. We show, first, that this bias only appears in the presence of *homophily*, when an individual's characteristic is positively correlated with their peers' characteristics. Under homophily, the average observed peer characteristic acts as a proxy for the true individual characteristic and apparent estimates of the peer effect capture part of the individual effect. Due to averaging, the average peer characteristic of an individual is less noisy – and hence a better proxy for the individual characteristic – when this individual has more peers. Consequently, we find that the expansion bias tends to be larger in networks with higher degrees. Our analysis further reveals that, beyond degrees and correlation in friends' characteristics, the expansion bias tends to be lower in networks with higher clustering and higher correlation in friends' friends characteristics. Either feature is associated with a larger independent variation in average peer characteristics, reducing the bias. Overall, our study offers the first comprehensive examination of how the expansion bias in peer effect estimates is influenced by the structure of social interactions.

Homophily is a widely documented phenomenon 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 akin levels of parental education. Measures of these characteristics in survey data, such as *Add Health*, are notoriously noisy, suggesting a widespread risk of inflated peer effect estimates. Although homophily and the expansion bias disappear under complete randomization, they reappear in common quasi-experimental setups, such as the randomization of peers within stratified groups.[4]

[3]The presence of related measurement errors on two regressors makes the problem non-standard. In general, only little is known about the nature of asymptotic bias in OLS estimates if there is measurement error in multiple variables (Greene, 2003, 86).

[4]Under complete randomization, inflated estimates might still occur when individuals are drawn

Our main findings for the anatomy of the bias are robust to the addition of network fixed effects. These are commonly included to address concerns about network formation on unobservables. In simulations, we find that both the globally and locally demeaned estimators tend to exhibit larger asymptotic biases than the non-demeaned estimator – when the latter is also valid. The reason for this is that differencing disproportionately removes the actual signal compared to the underlying noise. This phenomenon mirrors what is observed in linear panel data models, where employing first-differences often results to amplified attenuation biases (Griliches & Hausman, 1986). We show that for larger networks, the bias in the globally demeaned estimates can be close to that in the non-demeaned estimates under common conditions.

In the first part of our analysis, we thus clarify how relationships between characteristics and networks give rise to an expansion bias in peer effect estimates. In the second part, we demonstrate how these relationships can actually be leveraged to solve the problem. We show that the econometric model is generically identified through mean and covariance restrictions, and without relying on external information.[5] Mean restrictions exploit potential associations between network positions and the individual characteristic. For instance, when agents with more peers tend to have a higher value of the characteristic. Covariance restrictions take advantage of natural variations in how outcomes at one network position relate to observed characteristics at another. We derive necessary and sufficient rank conditions for identification based on either the mean or the covariance restrictions. These conditions combine the interaction matrix and the characteristic's first two moments, and hold generically. In summary, we find that except in special cases, it is possible to eliminate the expansion bias resulting from measurement errors.

To do so in practice, we propose generalized method of moments (GMM) and instrumental variables (IV) estimators that are easy to implement. The GMM approach uses more information and is generally more efficient, at the cost of introducing non-linearities in the estimation procedure. Moments are directly built from the means and covariance restrictions highlighted in our identification results. IV estimation, by contrast, is linear but typically less efficient. We propose a variety of valid instruments, including network lags and network features. Monte-Carlo simulations demonstrate the feasibility and good small sample performance of these estimators.

from a finite pool, giving rise to the exclusion bias (e.g., see Caeyers & Fafchamps, 2023).

[5]External 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.

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), over labor supply (Nicoletti, Salvanes, & Tominey, 2018), to consumption decisions (De Giorgi, Frederiksen, & Pistaferri, 2019). When characteristics are suspected to suffer from measurement error, the results and tools developed in this paper can aid empirical researchers to assess the size and direction of the resulting bias, and to obtain consistent peer effects estimates. Our analysis also clarifies the impact of common strategies to mitigate peer endogeneity (e.g., peer randomization, network fixed effects) on this bias.

Our analysis advances the small literature on measurement error and 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 state average 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.

Another strand of research concerns models with endogenous peer effects and 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.

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

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 are shrinking 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 the 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 the measurement error, nor full independence of the 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 with and without network-specific fixed effects and details the associated OLS estimators. In Section 3, we show that in the presence of errors-in-variables, the OLS estimates for the peer effect might be inflated and analyze how this expansion bias depends on the underlying structure of the social network. 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 IV estimators that are straightforward to implement. Section 6 contains a Monte Carlo simulation to illustrate our main results. Finally, Section 7 concludes. All proofs are collected in the Appendix.

2 Setup

A researcher observes data on outcomes, characteristics, and peers, and wants to estimate the impact of peers' characteristics on individual outcomes. We consider a data generating process where a sequence $\{\mathbf{y}_s, \mathbf{x}_s, \mathbf{e}_s, \mathbf{u}_s, \mathbf{A}_s\}_{s=1,\ldots,S}$ 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. The researcher observes outcomes $\mathbf{y}_s$, networks $\mathbf{A}_s$ and mismeasured characteristics $\widetilde{\mathbf{x}}_s = \mathbf{x}_s + \mathbf{u}_s$. We assume that the size of the networks N_s is uniformly bounded and consider many-network asymptotics. The number of observations $N = \sum_{s=1}^{S} N_s \to \infty$ as the number of networks $S \to \infty$. Throughout, all probability limits are with respect to N and are assumed to exist and to be finite.

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. We say that the network is *fixed* when all network observations have the same non-stochastic network structure, i.e., $\forall s, \mathbf{A}_s = \mathbf{A}_0$. Note that in this case, individual characteristics can still depend on network positions.

A network is *connected* when there is a network path between any pair of individuals. It is *regular* if every individual has the same degree. The *network distance* between two individuals it the length of a shortest path between them. The *diameter* of a network is the maximum network distance between any pair.

2.1 Model

We assume that social networks are undirected and without isolated individuals.[7] The neighborhood of individual i, $\mathcal{N}_{si}$, is the set of i's peers, $j \in \mathcal{N}_{si} \iff (\mathbf{A}_s)_{ij} = 1$. Formally, all connections are reciprocal, $(\mathbf{A}_s)_{ij} = (\mathbf{A}_s)_{ji}$, and every individual has at least one peer, $\forall i, \mathcal{N}_{si} \neq \emptyset$. We say that the social network takes the form of a *group* if $\forall i, j : i \neq j$, we have that $(\mathbf{A}_s)_{ij} = 1$. 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

[7]Most of our results hold with directed networks. This applies, in particular, to Propositions 1 and 2.

social interactions,

$$\begin{aligned} y_{si} &= \alpha + \gamma x_{si} + \delta \frac{1}{d_i} \sum_{j \in \mathcal{N}_{si}} x_{sj} + e_{si}, \\ \widetilde{x}_{si} &= x_{si} + u_{si}, \end{aligned}$$

in which individual i's outcome depends on her own characteristic (captured by γ) and her peers' average characteristic (captured by δ). The disturbance in the outcome equation is assumed to satisfy the standard conditional mean independence condition $\mathbb{E}(e_{si} \mid \mathbf{x}_s, \mathbf{G}_s, \mathbf{u}_s) = 0$. Measurement error arises as the researcher only observes a noisy proxy $\widetilde{x}_{si}$ for the true characteristic x_{si}.[8] Stacking observations, this model can be written compactly in matrix notation as

$$\mathbf{y}_s = \alpha \boldsymbol{\iota} + \gamma \mathbf{x}_s + \delta \overline{\mathbf{x}}_s + \mathbf{e}_s, \tag{1a}$$

$$\widetilde{\mathbf{x}}_s = \mathbf{x}_s + \mathbf{u}_s, \tag{1b}$$

in which $\overline{\mathbf{x}}_s = \mathbf{G}_s \mathbf{x}_s$ and $\overline{\widetilde{\mathbf{x}}}_s = \mathbf{G}_s \widetilde{\mathbf{x}}_s$. Regressors without measurement error can be partialled out using the Frisch-Waugh-Lovell (FWL) theorem.

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

$$\mathbf{y}_s = \alpha_s \boldsymbol{\iota} + \gamma \mathbf{x}_s + \delta \overline{\mathbf{x}}_s + \mathbf{e}_s, \tag{2a}$$

$$\widetilde{\mathbf{x}}_s = \mathbf{x}_s + \mathbf{u}_s, \tag{2b}$$

with $\mathbb{E}(\mathbf{e}_s \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{u}_s) = \mathbf{0}$ and where $\mathbb{E}(\alpha_s \mid \mathbf{x}_s, \mathbf{G}_s)$ may be different from 0. This addition makes the conditional mean independence condition more palatable, especially in settings where there might be clustering on unobservables.

We will assume throughout this paper that the conditional mean, variance, and pairwise covariance of the measurement error do not depend on any of the other variables in the model, nor on the structure of the social network.

Assumption 1. *The measurement errors satisfy:*

$$\begin{aligned} \mathbb{E}(u_{si} \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) &= 0, \\ \mathbb{E}(u_{si}^2 \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) &= \sigma_u^2, \\ \mathbb{E}(u_{si} u_{sj} \mid \alpha_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) &= 0. \end{aligned}$$

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

[8] It is well known that measurement error in the dependent variable does not induce asymptotic bias in OLS estimates. We therefore only consider measurement error in the independent variables.

error or the disturbance in the outcome equation. Furthermore, we refrain from making distributional assumptions such as (non)normality.

2.2 OLS estimators

Define $\mathbf{y} = \text{vec}(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_S)$, $\widetilde{\mathbf{x}} = \text{vec}(\widetilde{\mathbf{x}}_1, \widetilde{\mathbf{x}}_2, \ldots, \widetilde{\mathbf{x}}_S)$, $\mathbf{e} = \text{vec}(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_S)$, $\mathbf{u} = \text{vec}(\mathbf{u}_1, \mathbf{u}_2\,, \ldots, \mathbf{u}_S)$, and let $\mathbf{G} = \text{diag}(\mathbf{G}_1, \mathbf{G}_2, \ldots, \mathbf{G}_S)$ be the block-diagonal matrix that collects the interaction matrices. The OLS estimator for the baseline specification can then be written as

$$\begin{bmatrix} \widehat{\alpha}^{OLS} \\ \widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS} \end{bmatrix} = \left(\begin{bmatrix} \boldsymbol{\iota} & \widetilde{\mathbf{x}} & \widetilde{\widetilde{\mathbf{x}}} \end{bmatrix}^\intercal \begin{bmatrix} \boldsymbol{\iota} & \widetilde{\mathbf{x}} & \widetilde{\widetilde{\mathbf{x}}} \end{bmatrix} \right)^{-1} \begin{bmatrix} \boldsymbol{\iota} & \widetilde{\mathbf{x}} & \widetilde{\widetilde{\mathbf{x}}} \end{bmatrix}^\intercal \mathbf{y}.$$

As a result of the FWL theorem, we can simplify the exposition by partialling out the intercept, which gives a numerically identical estimator for the two parameters of interest,

$$\begin{bmatrix} \widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{W}_b\widetilde{\mathbf{x}} & \mathbf{W}_b\widetilde{\widetilde{\mathbf{x}}} \end{bmatrix}^\intercal \begin{bmatrix} \mathbf{W}_b\widetilde{\mathbf{x}} & \mathbf{W}_b\widetilde{\widetilde{\mathbf{x}}} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{W}_b\widetilde{\mathbf{x}} & \mathbf{W}_b\widetilde{\widetilde{\mathbf{x}}} \end{bmatrix}^\intercal \mathbf{W}_b\mathbf{y}, \tag{3}$$

where $\mathbf{W}_b = \mathbf{I}_N - N^{-1}\mathbf{J}_N$, with $\mathbf{I}_N$ and $\mathbf{J}_N$ the $N \times N$ identity and all-ones matrices, respectively. In this reformulation, all variables are expressed in terms of deviations from their sample means.

The extended specification with network-specific fixed effects is typically estimated by including network-specific dummy variables in the OLS estimator:

$$\begin{bmatrix} \widehat{\boldsymbol{\alpha}}_g^{OLS} \\ \widehat{\gamma}_g^{OLS} \\ \widehat{\delta}_g^{OLS} \end{bmatrix} = \left(\begin{bmatrix} \mathbf{F} & \widetilde{\mathbf{x}} & \widetilde{\widetilde{\mathbf{x}}} \end{bmatrix}^\intercal \begin{bmatrix} \mathbf{F} & \widetilde{\mathbf{x}} & \widetilde{\widetilde{\mathbf{x}}} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mathbf{F} & \widetilde{\mathbf{x}} & \widetilde{\widetilde{\mathbf{x}}} \end{bmatrix}^\intercal \mathbf{y},$$

in which $\mathbf{F} = \text{diag}(\boldsymbol{\iota}_{N_1}, \boldsymbol{\iota}_{N_2}, \ldots, \boldsymbol{\iota}_{N_S})$. Again as a result of the FWL theorem, a numerically identically estimator for the two parameters of interest is given by replacing the differencing matrix $\mathbf{W}_b$ with $\mathbf{W}_g = \mathbf{I}_N - \text{diag}(N_1^{-1}\mathbf{J}_{N_1}, N_2^{-1}\mathbf{J}_{N_2}, \ldots, N_S^{-1}\mathbf{J}_{N_S})$ in Expression (3). The dummy variable approach is therefore equivalent to OLS regression on the network-demeaned variables. The resulting *globally demeaned* estimates $\widehat{\gamma}_g^{OLS}$ and $\widehat{\delta}_g^{OLS}$ resemble the within estimates from panel data analysis.

It is also possible to obtain valid estimates using other differencing operations.[9] For example, if the differencing matrix is set to $\mathbf{W}_l = \mathbf{I}_N - \mathbf{G}$, the demeaning takes

[9] Note that all $\mathbf{W} \neq \mathbf{0}$ such that $\mathbf{W}\boldsymbol{\iota} = \mathbf{0}$ can be used to difference out the network-specific intercept.

place among direct peers. The resulting *locally demeaned* estimates $\widehat{\gamma}_l^{OLS}$ and $\widehat{\delta}_l^{OLS}$ resemble the first-difference estimates from panel data analysis. As we will see below, both estimators can exhibit quite different asymptotic biases.

Given two random variables a and b, we let $\mathbb{E}(a)$ denote the expectation of a, $\mathbb{V}(a)$ the variance of a, and $\mathbb{C}(a,b)$ and $\rho(a,b)$ the covariance and correlation between a and b, respectively. In particular, $\mathbb{E}(a) = \text{plim } N^{-1}\sum_s \sum_i a_{si}$ and $\mathbb{C}(a,b) = \text{plim } N^{-1}\sum_s \sum_i a_{si}b_{si} - \mathbb{E}(a)\mathbb{E}(b)$.

3 Expansion bias

3.1 Baseline specification

We first study the asymptotic bias for a fixed interaction matrix $\mathbf{G}_0$ of size N_0 so that $\mathbf{G} = \mathbf{I}_S \otimes \mathbf{G}_0$, where $\otimes$ denotes the Kronecker product. We denote positions within this network with a 0 subscript. By substituting Equations (1a) and (1b) in (3), the asymptotic bias of $\widehat{\gamma}^{OLS}$ and $\widehat{\delta}^{OLS}$ can be derived in terms of the true model parameters and of the variance-covariance matrices $\mathbf{S}$ and $\mathbf{\Sigma}$ of the regressors and the measurement errors, respectively. Formally, we have that

$$\text{plim} \begin{bmatrix} \widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS} \end{bmatrix} = (\mathbf{S}+\mathbf{\Sigma})^{-1}\mathbf{S} \begin{bmatrix} \gamma \\ \delta \end{bmatrix}, \tag{4}$$

in which

$$\mathbf{S} = \text{plim } N^{-1} \begin{bmatrix} \mathbf{x}^\intercal \mathbf{W}_b \mathbf{x} & \mathbf{x}^\intercal \mathbf{W}_b \overline{\mathbf{x}} \\ \overline{\mathbf{x}}^\intercal \mathbf{W}_b \mathbf{x} & \overline{\mathbf{x}}^\intercal \mathbf{W}_b \overline{\mathbf{x}} \end{bmatrix} = \begin{bmatrix} \mathbb{V}(x) & \mathbb{C}(x,\overline{x}) \\ \mathbb{C}(x,\overline{x}) & \mathbb{V}(\overline{x}) \end{bmatrix}, \tag{5a}$$

$$\mathbf{\Sigma} = \text{plim } N^{-1} \begin{bmatrix} \mathbf{u}^\intercal \mathbf{W}_b \mathbf{u} & \mathbf{u}^\intercal \mathbf{W}_b \overline{\mathbf{u}} \\ \overline{\mathbf{u}}^\intercal \mathbf{W}_b \mathbf{u} & \overline{\mathbf{u}}^\intercal \mathbf{W}_b \overline{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbb{V}(u) & \mathbb{C}(u,\overline{u}) \\ \mathbb{C}(u,\overline{u}) & \mathbb{V}(\overline{u}) \end{bmatrix}. \tag{5b}$$

Note that if there is no measurement error, $\mathbf{\Sigma} = \mathbf{0}$ and $(\mathbf{S}+\mathbf{\Sigma})^{-1}\mathbf{S} = \mathbf{I}$, and therefore there is no asymptotic bias, as expected. In most settings, the presence of measurement error entails an attenuation bias, such that OLS estimates are smaller than the true model parameters (e.g., see Wansbeek & Meijer, 2000 for a detailed discussion). As argued above, however, that this is generally not the case for the peer effects coefficient in the linear-in-means model.

The linear-in-means model imposes additional structure on the variance-covariance matrix of the measurement error. First, as the measurement error is homoscedastic and uncorrelated across pairs of peers, we have that $\mathbb{V}(u) = \text{plim } N^{-1}\sum_s \sum_i u_{si}^2 = \sigma_u^2$ and $\mathbb{C}(u,\overline{u}) = \text{plim } N^{-1}\sum_s \sum_i u_{si}\overline{u}_{si} = 0$, which entails that the off-diagonal elements of $\mathbf{\Sigma}$ are zero. Second, due to the averaging across peers, the measurement error for the average peer characteristic has a smaller variance than that for the own characteristic. We have:

$$
\begin{aligned}
\mathbb{V}(\overline{u}) &= \text{plim } N^{-1} \sum_s \sum_i \overline{u}_{si}^2 \\
&= N_0^{-1} \sum_{i_0} \mathbb{E}((\mathbf{G}_0 \mathbf{u}_0)_{i_0}^2) \\
&= N_0^{-1} \sum_{i_0} \mathbb{E}\left(\left(\sum_{j_0} g_{i_0 j_0} u_{j_0}\right)^2\right) \\
&= N_0^{-1} \sum_{i_0, j_0} g_{i_0 j_0}^2 \mathbb{E}(u_{j_0})^2 \\
&= N_0^{-1} \sum_{i_0, j_0} a_{i_0 j_0} d_{i_0}^{-2} \sigma_u^2 \\
&= h_0 \sigma_u^2,
\end{aligned}
\tag{6}
$$

where

$$
h_0 = N_0^{-1} \sum_{i_0} d_{i_0}^{-1},
$$

is the arithmetic mean of the inverse degrees of the network. This first key network statistic captures the overall extent of the averaging across peers in the network. In particular, we have $0 < \frac{1}{d_{max}} \leq h_0 \leq \frac{1}{d_{min}} \leq 1$, where d_{max} and d_{min} denote the largest and lowest degree in the network, respectively.

Substituting these simplifications in Equation (5) and expanding (4), we can state the following result for the baseline model.

Proposition 1. *Suppose that Assumption 1 holds. OLS estimates of γ and δ converge to*

$$
plim \begin{bmatrix} \widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS} \end{bmatrix} = D^{-1} \begin{bmatrix} \Delta + h_0 \mathbb{V}(x)\sigma_u^2 & h_0 \mathbb{C}(x, \overline{x})\sigma_u^2 \\ \mathbb{C}(x, \overline{x})\sigma_u^2 & \Delta + \mathbb{V}(\overline{x})\sigma_u^2 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix} = \mathbf{B} \begin{bmatrix} \gamma \\ \delta \end{bmatrix},
$$

where

$$
\begin{aligned}
D &= \Delta + h_0 \mathbb{V}(x)\sigma_u^2 + \mathbb{V}(\overline{x})\sigma_u^2 + h_0 \sigma_u^4, \\
\Delta &= \det(\mathbf{S}) = \mathbb{V}(x)\mathbb{V}(\overline{x}) - \mathbb{C}(x, \overline{x})^2.
\end{aligned}
$$

Since $0 \leq b_{11} \leq 1$ and $0 \leq b_{22} \leq 1$, both $\widehat{\gamma}^{OLS}$ and $\widehat{\delta}^{OLS}$ exhibit an asymptotic attenuation effect with respect to their associated true parameter values. However, if $b_{12} \neq 0$ (or equivalently, $b_{21} \neq 0$) an expansion effect also occurs, due to weight shifting between both variables. In that case, the OLS estimates erroneously pick up some of the effect of the other variable. This weight shifting notably depends on $\mathbb{C}(x, \overline{x})$ and on $\mathbb{V}(\overline{x})$ which, in turn, depend on the data generating process and on the relation between characteristics and links. One can interpret Δ as a measure of the amount of independent variation between the individual characteristic and peer characteristic.[10] The expansion bias tends to be larger when this independent variation is lower.

For clarity of exposition, in the remainder of this section we focus on the case where there is no true peer effect, i.e., $\delta = 0$. The anatomy of the expansion bias is further explored in the following proposition.

Proposition 2. *Suppose that Assumption 1 holds and that there is no peer effect, i.e., $\delta = 0$. If $\mathbb{C}(x, \overline{x}) > 0$, then $\widehat{\delta}^{OLS}$ has an asymptotic expansion bias in the direction of γ. It is zero when $\sigma_u^2 = 0$ or $+\infty$, and maximal when $\sigma_u^2 = \left(h_0^{-1}\Delta\right)^{\frac{1}{2}}$. If $\mathbb{C}(x, \overline{x}) = 0$, there is no asymptotic bias. The maximal bias is equal to*

$$\max_{\sigma_u^2} \left\{ plim\ \widehat{\delta}^{OLS} \right\} = \frac{\mathbb{C}(x, \overline{x})}{2\sqrt{h_0 \Delta} + \mathbb{V}(x)} \gamma.$$

Proposition 2 shows that the emergence of an expansion bias is crucially related to the presence of *homophily* in the network. A network displays homophily when similar individuals are relatively more likely to be connected. The expansion bias appears when $\mathbb{C}(x, \overline{x}) > 0$, which means that an individual's characteristic is positively correlated with their average peers' characteristic, a sign of homophily. By contrast, there is no expansion bias when $\mathbb{C}(x, \overline{x}) = 0$ and when an individual's characteristic and their average peer characteristic are uncorrelated. This happens, for instance, when links are randomly assigned and the individual characteristic is distributed in an i.i.d. way across individuals.[11]

Under homophily, peers' average observed characteristic acts as a proxy for the true individual characteristic. The apparent estimate of the peer effect then captures part of the individual effect, leading to an asymptotic bias. From the formulas presented in Propositions 1 and 2, we see that the expansion bias tends to be higher when homophily is higher and when degrees are higher (i.e., when average inverse degree is lower). In these cases, the average observed characteristic among peers becomes a better proxy for the true, unobserved individual characteristic, leading to a larger bias. These comparative statics are only valid when holding $\mathbb{V}(\overline{x})$ constant, however, which complicates interpretation. We address this issue below.

[10] As noted by Abel (2018), for any arbitrary covariance matrix $[\mathbf{V}]_{st} = \mathbb{C}(v_s, v_t)$, Hadamard's inequality implies that $0 \leq \det(\mathbf{V}) \leq \prod_t \mathbb{V}(v_t)$. The lower bound is attained when there is perfect linear dependence between variables; the upper bound is attained when there is no linear dependence. Therefore, $\det(\mathbf{V})$ acts as a measure of linear dependence between variables and is sometimes called the *generalized variance*.

[11] By contrast, when individuals are selected from a finite pool, random links may give rise to a negative correlation between the individual characteristic and average peer characteristic due to the exclusion bias (e.g., see Caeyers & Fafchamps, 2023). The estimate of the peer effect then displays an expansion bias in the opposite direction of γ.

Proposition 2 also shows that the expansion bias displays an inverse-U shape as a function of the variance of the measurement error, in the absence of true peer effects. This non-monotonicity is intuitive. There is no bias when there is no measurement error. And when measurement error is very large, observed characteristics essentially contain no information, leading to a valid zero estimate for peer effects. More generally, Proposition 1 shows that plim $\widehat{\gamma}^{OLS}, \widehat{\delta}^{OLS} \to 0$ when σ_u^2 becomes large.

Even simpler expressions than those in Proposition 1 can be obtained when measurement error is small.

Corollary 1. *Suppose that Assumption 1 holds and that $\sigma_u^2 \approx 0$. OLS estimates of γ and δ converge to*

$$plim \begin{bmatrix} \widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS} \end{bmatrix} \approx \begin{bmatrix} 1 - \frac{\mathbb{V}(\overline{x})}{\Delta}\sigma_u^2 & \frac{h_0\mathbb{C}(x,\overline{x})}{\Delta}\sigma_u^2 \\ \frac{\mathbb{C}(x,\overline{x})}{\Delta}\sigma_u^2 & 1 - \frac{h_0\mathbb{V}(x)}{\Delta}\sigma_u^2 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix}.$$

Under the conditions of Corollary 1, if $\delta = 0$ the expansion bias for $\widehat{\delta}^{OLS}$ is simply proportional to $\frac{\mathbb{C}(x,\overline{x})}{\Delta}\sigma_u^2$. Observing that $\Delta = (1-\rho(x,\overline{x})^2)\mathbb{V}(x)\mathbb{V}(\overline{x})$, it holds that

$$\text{plim } \widehat{\delta}^{OLS} \approx \frac{\mathbb{C}(x,\overline{x})}{\Delta}\sigma_u^2\gamma = \frac{\rho(x,\overline{x})}{1-\rho(x,\overline{x})^2}\frac{1}{\sqrt{\mathbb{V}(x)\mathbb{V}(\overline{x})}}\sigma_u^2\gamma.$$

The expansion bias is therefore zero when the correlation $\rho(x,\overline{x}) = 0$ and grows large when $\rho(x,\overline{x}) \to 1$; it is decreasing in the variances $\mathbb{V}(x)$ and $\mathbb{V}(\overline{x})$.

The extent of the expansion bias thus depends on statistics such as $\mathbb{C}(x,\overline{x})$ and $\mathbb{V}(\overline{x})$, which both depend on the underlying data generating process. To better understand how the expansion bias depends on network features, we now add some structure. Consider a fixed network $\mathbf{G}_0$ and let $d(i_0, j_0)$ denote the network distance between node i_0 and node j_0.

Assumption 2. *Consider a fixed network $\mathbf{G}_0$ and assume that for any two nodes i_0, j_0*

$$\begin{aligned} \mathbb{E}(x_{i_0}) &= \mu_x, \\ \mathbb{V}(x_{i_0}) &= \sigma_x^2, \\ \mathbb{C}(x_{i_0}, x_{j_0} \mid d(i_0,j_0) = 1) &= \rho_1\sigma_x^2, \\ \mathbb{C}(x_{i_0}, x_{j_0} \mid d(i_0,j_0) = 2) &= \rho_2\sigma_x^2, \end{aligned}$$

with $0 \leq \rho_2 < \rho_1$.

Under Assumption 2, the network is fixed and the expectation and the variance of the characteristic of an individual do not depend on this individual's position in the network. Moreover, the correlations between the characteristics of friends (at distance 1) and of friends of friends (at distance 2) are constant and greater than or equal to zero, and this correlation is lower at distance 2 than at distance 1.

A simple way to generate data consistent with Assumption 2 is as follows. Consider a fixed network of size N_0. Set, for instance, $\mathbb{C}(x_{i_0}, x_{j_0}) = 0$ if $d(i_0, j_0) \geq 3$. Let $\mathbf{V}$ be the $N_0 \times N_0$ variance-covariance matrix such that $V_{i_0 j_0} = \mathbb{C}(x_{i_0}, x_{j_0}) = \sigma_x^2$ if $i_0 = j_0$, $\rho_1 \sigma_x^2$ if $d(i_0, j_0) = 1$, $\rho_2 \sigma_x^2$ if $d(i_0, j_0) = 2$, and 0 otherwise. If $\mathbf{V}$ is positive semi-definite, pick i.i.d. draws from the multivariate normal distribution $N(\mu_x \boldsymbol{\iota}, \mathbf{V})$.

Given these assumptions, we can simplify the expressions for $\mathbb{C}(x, \overline{x})$ and $\mathbb{V}(\overline{x})$. Define the *clustering coefficient* of node i_0, c_{i_0}, as the likelihood that any two friends of i_0, picked at random, are also friends. Formally,

$$c_{i_0} = \frac{\sum_{j_0 \neq k_0 \in \mathcal{N}_{i_0}} a_{j_0 k_0}}{d_{i_0}(d_{i_0} - 1)}, \tag{7}$$

if $d_{i_0} \geq 2$ and $c_{i_0} = 0$ if $d_{i_0} = 1$. This is a standard index that captures the tendency of individuals with common friends to be connected. Clustering is typically high in real social networks, as two agent with a common friend are generally much more likely to be connected than two random agents.

Lemma 1. *Suppose that Assumption 2 holds. Then we have that*

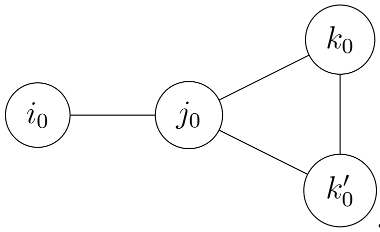
$$\begin{aligned} \mathbb{C}(x, \overline{x}) &= \rho_1 \sigma_x^2, \\ \mathbb{V}(\overline{x}) &= (h_0 + h_1 \rho_1 + h_2 \rho_2) \sigma_x^2, \end{aligned}$$

where $h_1 = N_0^{-1} \sum_{i_0} \frac{d_{i_0} - 1}{d_{i_0}} c_{i_0}$ *and* $h_2 = 1 - h_0 - h_1 \geq 0$.

Lemma 1 shows that the variance of average peers' characteristics is a weighted mean of the variances of the characteristics of the individual nodes and the covariances of the characteristics of nodes at distances 1 and 2. The first weight is the average inverse degree, h_0. The weight on the covariance of nodes at distance 1 is a new network index, h_1, equal to a linear combination of the nodes' clustering coefficients multiplied by 1 minus inverse degree. When all nodes' degrees are large, h_1 is approximately equal to the network's average clustering. In general, h_1 differs from average clustering by placing less weight on low degree nodes. Note, also, that $h_0 + h_1 \leq 1$ so that the residual weight on the covariance of nodes at distance 2, $h_2 = 1 - h_0 - h_1$ is greater than or equal to zero.

The variance of average peers' characteristics depends on correlations in characteristics for every pair of neighbors of a node. If these neighbors are connected themselves, this correlation is ρ_1, while it is ρ_2 if they are unconnected. And clustering computes the proportion of these pairs who are connected. Thus, when $\rho_1 > \rho_2$, $\mathbb{V}(\overline{x})$ tends to increase when clustering is higher. The following example illustrates Lemma 1.

Example 1. *Consider the following network with 4 nodes*



.

We have: $d_{i_0} = 1$, $d_{j_0} = 3$, $d_{k_0} = d_{k'_0} = 2$, *leading to*

$$h_0 = \frac{1}{4}\left(\frac{1}{1} + \frac{1}{3} + \frac{1}{2} + \frac{1}{2}\right) = \frac{7}{12},$$

and $c_{i_0} = 0$, $c_{j_0} = \frac{1}{3}$, $c_{k_0} = c_{k'_0} = 1$, *leading to*

$$h_1 = \frac{1}{4}\left(\frac{0}{1}0 + \frac{2}{3}\frac{1}{3} + \frac{1}{2}1 + \frac{1}{2}1\right) = \frac{11}{36},$$

such that

$$\mathbb{V}(\overline{x}) = \left(\frac{21}{36} + \frac{11}{36}\rho_1 + \frac{4}{36}\rho_2\right)\sigma_x^2.$$

Combining Lemma 1 and Proposition 1, we obtain the following expression for the expansion bias of the peer effect estimate. Denote by $\phi = \sigma_u^2/\sigma_x^2$ the noise-to-signal ratio.

Proposition 3. *Suppose that Assumptions 1 and 2 hold and that there is no peer effect, i.e.,* $\delta = 0$. *Then,* $\widehat{\delta}^{OLS}$ *has the asymptotic expansion bias*

$$\mathit{plim}\ \widehat{\delta}^{OLS} = \frac{\rho_1\phi}{(1+\phi)(h_0 + h_1\rho_1 + h_2\rho_2) - \rho_1^2 + h_0\phi + h_0\phi^2}\gamma.$$

This expansion bias (i) *increases in* ρ_1; (ii) *decreases in* ρ_2; (iii) *decreases in* h_0; *and* (iv) *decreases in* h_1.

The maximal bias is

$$\max_{\sigma_u^2}\left\{\mathit{plim}\ \widehat{\delta}^{OLS}\right\} = \frac{\rho_1\sigma_x^2}{2\sqrt{h_0(h_0 + h_1\rho_1 + h_2\rho_2)\sigma_x^2 + \sigma_x^2}}\gamma.$$

Proposition 3 clarifies how the expansion bias depends on the network structure. The expansion bias increases with ρ_1, an index of homophily, as this increases weight shifting. The expansion bias decreases with ρ_2 and with the clustering measure h_1. An increase in either measure increases $\mathbb{V}(\overline{x})$ without affecting $\mathbb{C}(x, \overline{x})$. This yields an increase in the independent variation in average peer characteristics which lowers the bias. The expansion bias decreases in average inverse degree h_0, as lower degrees imply that the mismeasured peer characteristic is more noisy and the estimated peer effect then captures less of the true individual effect.

Proposition 3 applies to any network structure and hence, in particular, when agents interact in groups. In a group of size N_0, $d_i = N_0 - 1$ and $h_0 = (N_0 - 1)^{-1}$. There are no pairs of agents at distance two, which means that $h_2 = 0$ and $h_1 = 1 - h_0$. In addition, in most settings with cross-sectional data the labels of the individuals do not matter under group interactions. Peers' characteristics are then exchangeable random variables. As a result, the requirement of homogeneous expectations, variances and covariances in Assumption 2 is innocuous when considering groups. The expression for the expansion bias in Proposition 3 simplifies as follows.

Corollary 2. *Suppose that Assumptions 1 and 2 hold and that there is no peer effect, i.e., $\delta = 0$. If the social graph takes the form of a group, then $\widehat{\delta}^{OLS}$ has the asymptotic bias*

$$\mathit{plim}\ \widehat{\delta}^{OLS} = \frac{\rho_1 \phi}{(1+\phi)(h_0 + (1-h_0)\rho_1) - \rho_1^2 + h_0\phi + h_0\phi^2}\gamma.$$

with $h_0 = (N_0 - 1)^{-1}$. The expansion bias (i) increases in ρ_1; and (ii) increases in N_0.

The maximal bias is

$$\max_{\sigma_u^2} \left\{ \mathit{plim}\ \widehat{\delta}^{OLS} \right\} = \frac{\rho_1 \sigma_x^2}{2\sqrt{h_0(h_0 + (1-h_0)\rho_1)\sigma_x^2} + \sigma_x^2}\gamma.$$

Corollary 2 complements earlier results obtained by Feld and Zölitz (2017) on group interactions. One difference between our setup and theirs is that they consider leave-in means, computed over every one in the group, and hence which double count the characteristic of the focal agent. By contrast, we consider leave-out means, computed over every one else in the group.[12] We show that the results that the expansion bias is increasing in ρ_1 and N_0 are robust to the way the mean is computed.

Up to this point, the results in this section were derived for a fixed network structure $\mathbf{G}_0$. In general, however, the data generating process may induce a distribution over network structures, where network $\mathbf{G}_s$ of size N_s is picked with probability p_s. One can show that the overall asymptotic variance-covariance matrices can be decomposed as

$$\mathbf{S} = \sum_s \frac{N_s p_s}{\sum_t N_t p_t} \mathbf{S}_s,$$
$$\mathbf{\Sigma} = \sum_s \frac{N_s p_s}{\sum_t N_t p_t} \mathbf{\Sigma}_s,$$

where the weights adjust for the differences in network sizes. This implies that the expansion bias can be expressed in terms of a weighted average of inverse degree

[12] We provide additional results for the setting with leave-in means in Appendix A.1.

and clustering

$$h_0 = \mathbb{E}(h_0(\mathbf{G}_k)) = \sum_s \frac{N_s p_s}{\sum_t N_t p_t} h_0(\mathbf{G}_s),$$
$$h_1 = \mathbb{E}(h_1(\mathbf{G}_k)) = \sum_s \frac{N_s p_s}{\sum_t N_t p_t} h_1(\mathbf{G}_s).$$

To sum up, the expansion bias generally increases with the correlation between individual and average peer characteristics and decreases with the variance of average peer characteristics. Our results show that the expansion bias is higher for networks with higher degrees, lower clustering coefficients, higher correlation in friends' characteristics, and lower correlation in friends' friends characteristics. Overall, we find that this expansion bias crucially depends on the network structure and on the relation between links and characteristics.

3.2 Extended specification

We now turn to the asymptotic bias in the extended specification that allows for network-specific fixed effects. It turns out that the globally and locally demeaned OLS estimators also typically exhibit expansion bias. The differencing operation make the analysis more complicated however: the nature and extent of the bias now depends on correlations between pairs at larger distances. Additional network features, besides average inverse degree and clustering, come into play. As in the baseline model, the presence of expansion bias can be connected to statistics that are estimable by the researcher, such as the covariance between the demeaned own and peer characteristic $\mathbb{C}_w(x, \overline{x})$ and the variance of the demeaned own characteristics $\mathbb{V}_w(x)$. As before, the subscript g refers to global demeaning, while l refers to local demeaning. We derive exact formulas for the probability limit of the peer effect estimates in the Appendix. In the following Proposition, we describe a simple sufficient condition for the emergence of expansion bias and we provide formulas for the asymptotic biases when measurement error is small.

Proposition 4. *Suppose that Assumption 1 holds and that there is no peer effect, i.e.,* $\delta = 0$. *If for* $w \in \{g, l\}$ *it holds that* $\mathbb{C}_w(x, \overline{x}) > 0$, *then* $\widehat{\delta}_w^{OLS}$ *has an asymptotic expansion bias in the direction of* γ. *Moreover, if* $\sigma_u^2 \approx 0$, *the asymptotic biases of* $\widehat{\delta}_g^{OLS}$ *and* $\widehat{\delta}_l^{OLS}$ *are*

$$plim\ \widehat{\delta}_g^{OLS} \approx \frac{(1 - N_0^{-1})\mathbb{C}_g(x, \overline{x}) + N_0^{-1}\mathbb{V}_g(x)}{\Delta_g} \sigma_u^2 \gamma,$$
$$plim\ \widehat{\delta}_l^{OLS} \approx \frac{(1 + h_0)\mathbb{C}_l(x, \overline{x}) + (h_0 + n_{10} - n_{20})\mathbb{V}_l(x)}{\Delta_l} \sigma_u^2 \gamma,$$

where $\Delta_w = \mathbb{V}_w(x)\mathbb{V}_w(\overline{x}) - \mathbb{C}_w(x, \overline{x})^2$, $n_{10} = N_0^{-1} \sum_{i_0, j_0} g_{i_0 j_0} g_{j_0 i_0}$, *and*

$n_{20} = N_0^{-1} \sum_{i_0,j_0,k_0} g_{i_0j_0}g_{i_0k_0}g_{k_0j_0}$.

Proposition 4 shows that correlations within the demeaned data, $\mathbb{C}_w(x, \overline{x}) > 0$, are a sufficient condition for the expansion bias to arise. Next, contrast the formulas for the asymptotic bias with and without network fixed effects. Recall, from Corollary 1 we know that plim $\widehat{\delta}^{OLS} = \frac{\mathbb{C}(x,\overline{x})}{\Delta}\sigma_u^2\gamma$ in the baseline model with small measurement error. The formula for plim $\widehat{\delta}_g^{OLS}$ is similar and, in particular, does not involve additional network statistics. We note two differences: covariances and variances are computed over the demeaned variables and there is an additional term scaling with N_0^{-1}. As network size N_0 increases, and if variations in variances and covariances are bounded, this new term becomes negligible. In fact, we show in Appendix A.2 that even when measurement error is large, plim $\widehat{\delta}_g^{OLS}$ − plim $\widehat{\delta}^{OLS} \rightarrow$ can tend to 0 when N_0 tends to infinity.

By contrast, the asymptotic bias of the locally demeaned estimator involves additional network parameters n_{10} and n_{20}. They are related to how degree and clustering are distributed throughout the network:

$$
\begin{aligned}
n_{10} &= N_0^{-1} \sum_{i_0} d_{i_0}^{-1} \sum_{j_0 \in \mathcal{N}_{i_0}} d_{j_0}^{-1}, \\
n_{20} &= N_0^{-1} \sum_{i_0} d_{i_0}^{-2} \sum_{j_0 \neq k_0 \in \mathcal{N}_{i_0}} d_{k_0}^{-1} a_{j_0k_0}.
\end{aligned}
$$

We see that n_{10} is equal to the average of peers' average inverse degrees, while n_{20} involves a combination of clustering and degree. For instance, when the network is regular and $\forall i, d_i = d$ we have $n_{10} = h_0 = d^{-1}$ and $n_{20} = d^{-1}h_1$.

To illustrate how the asymptotic bias depends on underlying network features, we consider the class of strongly regular graphs. A regular graph is strongly regular when every pair of connected agents has λ common neighbors. For strongly regular graphs, the key network statistics h_0 and h_1 are simply related to the parameters of the model: $h_0 = \frac{1}{d}$ and $h_1 = \frac{\lambda}{d}$. We also consider the following variation on Assumption 2.

Assumption 3. *Consider a fixed network* $\mathbf{G}_0$ *and assume that for any two nodes* i_0, j_0

$$
\begin{aligned}
\mathbb{E}(x_{i_0}) &= \mu_x, \\
\mathbb{V}(x_{i_0}) &= \sigma_x^2, \\
\mathbb{C}(x_{i_0}, x_{j_0} \mid d(i_0, j_0) = t) &= \rho_t\sigma_x^2, \quad t \in \{1, 2, \ldots, T\},
\end{aligned}
$$

with $0 \leq \rho_{t+1} \leq \rho_t \leq 1$ *and where* T *denotes the diameter of the network.*

We show in Appendix A.3 that when the graph is strongly regular, Assumptions 1 and 3 hold, $\sigma_u^2 \approx 0$ and $N_0 \to \infty$, then the peer effect estimate under global demeaning, $\widehat{\delta}_g^{OLS}$, has the following asymptotic expansion bias:

$$\text{plim } \widehat{\delta}_g^{OLS} \approx \frac{\rho_1 - \rho_2}{h_0(1-\rho_2)^2 + h_1(\rho_1 - \rho_2)(1-\rho_2) - (\rho_1 - \rho_2)^2} \sigma_u^2 \gamma.$$

In particular, this expansion bias (i) increases in ρ_1; (ii) decreases in ρ_2; (iii) decreases in h_0; and (iv) decreases in h_1. The main results from the baseline setting and Proposition 3 thus remain valid in this case.

Monte Carlo simulations from Section 6 suggest that the same holds true for the locally demeaned specification. These simulations also provide suggestive evidence that the expansion bias in the globally and locally demeaned estimators tends to be larger than that in the non-demeaned estimator – when the latter is also valid. The reason for this is that differencing disproportionately removes the actual signal compared to the underlying noise. This phenomenon mirrors what is observed in linear panel data models, where employing first-differences often results to amplified attenuation biases (Griliches & Hausman, 1986).

One important case where the expansion bias disapppears is when links are formed at random. More precisely, suppose that correlations between characteristics are constant across the network. Both differencing operations then remove all variation that induces the expansion bias.

Proposition 5. *Suppose that Assumptions 1 and 3 hold and that the correlation does not depend on network distance: i.e., $\rho = \rho_d$ for all d. Then there is only an attenuation bias in both the globally and locally demeaned OLS estimators, i.e.,*

$$\begin{aligned} \textit{plim } \widehat{\gamma}_g^{OLS} &= \textit{plim } \widehat{\gamma}_l^{OLS} = a_{\gamma\gamma}\gamma, \\ \textit{plim } \widehat{\delta}_g^{OLS} &= \textit{plim } \widehat{\delta}_l^{OLS} = a_{\delta\delta}\delta, \end{aligned}$$

where $a_{\gamma\gamma} = a_{\delta\delta} = \frac{(1-\rho)\sigma_x^2}{(1-\rho)\sigma_x^2 + \sigma_u^2}$.

Proposition 5 extends to networks results obtained by Ammermueller and Pischke (2009) for groups. An interesting implication is that in this case, the strength of the peer effect relative to the individual effect does not suffer from an asymptotic bias: i.e., plim $\frac{\widehat{\delta}_g^{OLS}}{\widehat{\gamma}_g^{OLS}} = $ plim $\frac{\widehat{\delta}_l^{OLS}}{\widehat{\gamma}_l^{OLS}} = \frac{\delta}{\gamma}$ (see Ammermueller & Pischke, 2009, p.335).

The assumption of distance-independent correlation covers, in particular, situations where links are formed at random within stratified groups.[13] Suppose that the characteristic of individual i in network s can be decomposed into the sum of two i.i.d. components: i.e., $x_{si} = \xi_s + \xi_{si}$. If the links within network s are formed

[13] As before, non-zero correlations under random assignment can also arise under the circumstances that give rise to the exclusion bias (Caeyers & Fafchamps, 2023).

randomly, the correlation between pairs of individuals at all distances is equal to

$$\rho = \frac{\mathbb{V}(\xi_s)}{\mathbb{V}(\xi_s) + \mathbb{V}(\xi_{si})}.$$

This represents situations where individuals sort themselves into different groups, and links are then random conditional on groups.

Even in the absence of network fixed effects, Proposition 2 shows that OLS estimates of the peer effect parameter suffer from an expansion bias with conditionally random links when $\rho > 0$. By contrast, Proposition 5 shows that the expansion bias disappears for the two demeaned estimators in this case. Demeaning thus might have an interest even in the absence of network fixed effects.

4 Identification

We now focus on 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.

Identification hinges on one of two complementary approaches. The first identification strategy, based on mean restrictions, exploits variation in the average individual characteristics across network positions. This strategy is valid when, for instance, agents with more friends tend to have a higher value of the characteristic. The second identification strategy, based on covariance restrictions, exploits variation in variances and covariances of individual characteristics across (pairs of) network positions. This strategy fails to work only in special cases.

We mainly focus our analysis on the cross-sectional setting where a fixed interaction matrix $\mathbf{G}_0$ is observed repeatedly by the analyst. We maintain our weak assumptions on disturbances $\mathbf{e}$ and measurement error $\mathbf{u}$ and, as a consequence, consider identification based on the following means and covariances. Researchers willing to impose more structure on these variables could exploit other moments for identification. We consider, first, expected outcomes as functions of network positions:

$$\mathbb{E}(\mathbf{y}_s \mid \theta) = \text{plim}\, S^{-1} \sum_s \mathbf{y}_s(\theta).$$

We then consider how the outcome in one network position co-varies with the observed characteristic in another position:

$$\mathbb{C}(\mathbf{y}_s, \widetilde{\mathbf{x}}_s \mid \theta) = \text{plim } S^{-1} \sum_s \mathbf{y}_s(\theta)\widetilde{\mathbf{x}}_s^\intercal(\theta) - \left(\text{plim } S^{-1} \sum_s \mathbf{y}_s(\theta)\right) \left(\text{plim } S^{-1} \sum_s \widetilde{\mathbf{x}}_s^\intercal(\theta)\right).$$

The model is *identified* when there do not exist two parameter vectors θ_1 and θ_2 with $\theta_1 \neq \theta_2$ such that these means and covariances are identical, i.e., $\mathbb{E}(\mathbf{y}_s \mid \theta_1) = \mathbb{E}(\mathbf{y}_s \mid \theta_2)$ and $\mathbb{C}(\mathbf{y}_s, \widetilde{\mathbf{x}}_s \mid \theta_1) = \mathbb{C}(\mathbf{y}_s, \widetilde{\mathbf{x}}_s \mid \theta_2)$. The model is thus identified when the mapping from the parameters to these moments is injective.

We say that identification is *generic* if non-identification only occurs on a lower-dimensional subset. Let $\widetilde{\mathbf{m}} = \mathbb{E}(\widetilde{\mathbf{x}}_s) \in \mathscr{M} \subseteq \mathbb{R}^{N_0}$ be the the vector that collects the average characteristic by network position and $\widetilde{\mathbf{V}} = \mathbb{V}(\widetilde{\mathbf{x}}_s) \in \mathscr{V} \subseteq \mathbb{R}^{N_0 \times N_0}$ the matrix that collects the covariances of characteristics between these positions. Identification on means and covariances is generic if

$$\dim \left\{ \widetilde{\mathbf{m}} : \theta_1 \neq \theta_2 \implies \mathbb{E}(\mathbf{y}_s \mid \theta_1) = \mathbb{E}(\mathbf{y}_s \mid \theta_2) \right\} < \dim \mathscr{M},$$
$$\dim \left\{ \widetilde{\mathbf{V}} : \theta_1 \neq \theta_2 \implies \mathbb{C}(\mathbf{y}_s, \widetilde{\mathbf{x}}_s \mid \theta_1) = \mathbb{C}(\mathbf{y}_s, \widetilde{\mathbf{x}}_s \mid \theta_2) \right\} < \dim \mathscr{V},$$

respectively.

4.1 Baseline specification

We first consider identification in the baseline specification. One source of identification is provided by *mean restrictions*. Under Assumption 1, and using the structure of the model in Expressions (1a) and (1b), expected outcome for every network position i_0 is given by

$$\mathbb{E}(y_{i_0}) = \alpha + \gamma \mathbb{E}(\widetilde{x}_{i_0}) + \delta \mathbb{E}(\widetilde{\widetilde{x}}_{i_0}). \tag{8}$$

Since there are N_0 network positions in total, this yields a linear system of N_0 equations in three unknowns. In many cases, however, the number of relevant equations is lower, as those that belong to symmetric positions contain redundant information.[14] We therefore only consider the equations that belong to *unique* network positions.

Parameters α, γ and δ are identified when the system of equations (8) has full rank, that is, rank three. This can only happen when $\mathbb{E}(\widetilde{x}_{i_0})$ varies with i_0 and in the presence of three unique network positions. These two necessary conditions may not be sufficient however, as the following example illustrates.

[14]The presence of symmetric network positions reduces the informational content of the data, as the joint distribution is exchangeable in these positions. The network contains symmetric networks positions when there exists a permutation π of the nodes such that $(\mathbf{A}_s)_{ij} = 1$ if and only if $(\mathbf{A}_s)_{\pi(i)\pi(j)} = 1$ for all i, j.

Example 1 (continued)**.** *Again consider the network with 4 nodes from Example 1. In this example there are only three unique network positions as k_0 and k_0' are indistinguishable by the researcher. Identification based on mean restrictions holds iff the matrix*

$$\begin{bmatrix} 1 & \mathbb{E}(x_{i_0}) & \mathbb{E}(\overline{x}_{i_0}) \\ 1 & \mathbb{E}(x_{j_0}) & \mathbb{E}(\overline{x}_{j_0}) \\ 1 & \mathbb{E}(x_{k_0}) & \mathbb{E}(\overline{x}_{k_0}) \end{bmatrix} = \begin{bmatrix} 1 & \mathbb{E}(x_{i_0}) & \mathbb{E}(x_{j_0}) \\ 1 & \mathbb{E}(x_{j_0}) & \frac{1}{3}\mathbb{E}(x_{i_0}) + \frac{2}{3}\mathbb{E}(x_{k_0}) \\ 1 & \mathbb{E}(x_{k_0}) & \frac{1}{2}\mathbb{E}(x_{j_0}) + \frac{1}{2}\mathbb{E}(x_{k_0}) \end{bmatrix}$$

is invertible. Identification fails, for example, if the expectations satisfy

$$(\mathbb{E}(x_{i_0}), \mathbb{E}(x_{j_0}), \mathbb{E}(x_{k_0})) = \left(1, \frac{1}{2}\left(\sqrt{\frac{11}{3}} - 1\right), 0\right),$$

and there are infinitely many such configurations. These cases, however, are non-generic.

Note that the mean restriction for node k_0' is identical to that of k_0 and therefore does not add any information, since both nodes cannot distinguished from the point of view of the researcher.

Another source of identification is provided by *covariance restrictions*. For every pair (i_0, j_0) of network positions, they are defined as the association between the outcome at position i_0 and the observed characteristic at position j_0. We can distinguish three cases depending on the distance between the pair:

$$\mathbb{C}(y_{i_0}, \widetilde{x}_{j_0}) = \begin{cases} \gamma\mathbb{V}(\widetilde{x}_{i_0}^2) + \delta\mathbb{C}(\widetilde{x}_{i_0}, \overline{\widetilde{x}}_{i_0}) - \gamma\sigma_u^2, & i_0 = j_0, \\ \gamma\mathbb{C}(\widetilde{x}_{i_0}, \widetilde{x}_{j_0}) + \delta\mathbb{C}(\widetilde{x}_{j_0}, \overline{\widetilde{x}}_{i_0}) - \frac{1}{d_{i_0}}\delta\sigma_u^2, & d(i_0, j_0) = 1, \\ \gamma\mathbb{C}(\widetilde{x}_{i_0}, \widetilde{x}_{j_0}) + \delta\mathbb{C}(\widetilde{x}_{j_0}, \overline{\widetilde{x}}_{i_0}), & d(i_0, j_0) \geq 2. \end{cases} \tag{9}$$

Taken together, these equations yield a nonlinear system of N_0^2 equations in three unknowns. Similarly as before, however, only unique positions contribute towards identification.

Intuitively, we see that unless in knife-edge cases, most networks deliver a substantial degree of overidentification through both the mean and covariance restrictions. The following theorem makes this statement formal.

Theorem 1. *Suppose that Assumption 1 holds.*

1. *The parameters (α, γ, δ) are identified from the mean restrictions if and only if the vectors $\boldsymbol{\iota}$, $\widetilde{\mathbf{m}}$, and $\mathbf{G}_0\widetilde{\mathbf{m}}$ are linearly independent. This can only happen when there are at least three unique network positions with different means.*

2. *Suppose that either $\gamma \neq 0$ or $\delta \neq 0$. Then the parameters $(\gamma, \delta, \sigma_u^2)$ are identified from the covariance restrictions if and only if the matrices $\mathbf{I}$, $\mathbf{G}_0$, $\widetilde{\mathbf{V}}$, and $\mathbf{G}_0\widetilde{\mathbf{V}}$ are linearly independent.*

If there is a sufficient number of unique network positions, identification is generic both on the basis of mean and covariance restrictions. This means that the set of $\widetilde{\mathbf{m}}$'s and $\widetilde{\mathbf{V}}$'s for which identification fails is small relative to the entire space of admissible configurations.

Proposition 6. *Suppose that Assumption 1 holds and that the network is connected.*

1. *The parameters (α, γ, δ) are generically identified from the mean restrictions if there are at least three unique network positions.*

2. *The parameters (γ, δ) are generically identified from the covariance restrictions if there are at least two unique network positions.*

To illustrate how the identification conditions operate, assume that the network satisfies Assumption 2 and that, in addition, $\rho_2 > 0$ and $\mathbb{C}(x_{i_0}, x_{j_0}) = 0$ if $d(i_0, j_0) \geq 3$. Let $\mathbf{A}_0^{(2)}$ denote the matrix of two-step away connections, $(\mathbf{A}_0^{(2)})_{i_0 j_0} = 1$ if $d(i_0, j_0) = 2$ and 0 otherwise. Under these assumptions, $\widetilde{\mathbf{V}} = (\sigma_x^2 + \sigma_u^2)\mathbf{I} + \rho_1 \sigma_x^2 \mathbf{A}_0 + \rho_2 \sigma_x^2 \mathbf{A}_0^{(2)}$. Identification is then guaranteed when the diameter of the network is greater than or equal to 3. Indeed, in this case there exists a pair of network positions i_0, j_0 such that $d(i_0, j_0) = 3$. For this pair, $(\mathbf{G}_0 \widetilde{\mathbf{V}})_{i_0 j_0} > 0$ while $(\mathbf{G}_0)_{i_0 j_0} = \widetilde{V}_{i_0 j_0} = 0$, which implies that $\mathbf{G}_0 \widetilde{\mathbf{V}}$ can not be expressed as a linear combination of $\mathbf{I}$, $\mathbf{G}_0$ and $\widetilde{\mathbf{V}}$.

Identification may fail, however, in specific cases. Concerning mean restrictions, the necessity of having at least three distinct positions means that simple network structures such as complete networks, circles, and stars are excluded. Regarding covariance restrictions, Theorem 1 shows that identification fails when the matrices $\mathbf{I}$, $\mathbf{G}_0$, $\widetilde{\mathbf{V}}$, and $\mathbf{G}_0 \widetilde{\mathbf{V}}$ are linearly dependent. This happens, for instance, when there is no correlation in characteristics across network positions and when the variance of characteristics is homoscedastic: i.e., $\widetilde{\mathbf{V}} = (\sigma_x^2 + \sigma_u^2)\mathbf{I}$. This structure arises naturally when network formation is independent of individual characteristics. Identification from covariance restrictions also fails when the social network takes the form of a single group of size N_0. In that case, $\widetilde{\mathbf{V}} = (\sigma_x^2 + \sigma_u^2)\mathbf{I} + (N_0 - 1)\rho_1 \sigma_x^2 \mathbf{G}_0$ and $\mathbf{G}_0 \widetilde{\mathbf{V}} = \rho_1 \sigma_x^2 \mathbf{I} + [\sigma_x^2 + \sigma_u^2 + (N_0 - 2)\rho_1 \sigma_x^2]\mathbf{G}_0$, such that the linear independence condition is violated.

Finally, Theorem 1 also covers settings where the social network takes the form of multiple groups of different sizes. Identification fails when ρ_1 and σ_x^2 do not vary with group sizes, but holds, generically, otherwise.

Proposition 7. *Suppose that Assumption 1 holds and that there are groups of two different sizes N_0' and N_0''. Then the parameters (γ, δ) are not identified through covariance restrictions when $\sigma_x' = \sigma_x''$ and $\rho_1' = \rho_1''$. If $\sigma_x' = \sigma_x''$, the parameters are identified from covariance restrictions if and only if $\rho_1' \neq \rho_1''$ and $(N_0' - 1)'\rho_1' \neq (N_0'' - 1)\rho_1''$. If $\rho_1' = \rho_1'' = \rho_1$, the* *parameters are identified if and only if* $\sigma'_x \neq \sigma''_x$ *and* $(1-\rho_1)(\sigma_x^{2\prime\prime} - \sigma_x^{2\prime}) \neq (\rho_1^2 - \rho_1)[(N_0'' - 1)\sigma_x^{2\prime\prime} - (N_0' - 1)\sigma_x^{2\prime}]$.

4.2 Extended specification

For the extended specification with network-specific fixed effects, we obtain a similar result to that of Theorem 1, but applied to demeaned data instead.

Theorem 2. *Suppose that Assumption 1 holds and let* $\mathbf{W} \neq \mathbf{0}$ *be any differencing matrix for which* $\mathbf{W}\boldsymbol{\iota} = \mathbf{0}$.

1. *The parameters* (γ, δ) *are identified from the mean restrictions for the* $\mathbf{W}$*-demeaned data if and only if the vectors* $\mathbf{W}\widetilde{\mathbf{m}}$ *and* $\mathbf{W}\mathbf{G}_0\widetilde{\mathbf{m}}$ *are linearly independent.*

2. *Suppose that either* $\gamma \neq 0$ *or* $\delta \neq 0$. *Then the parameters* $(\gamma, \delta, \sigma_u^2)$ *are identified from the covariance restrictions for the* $\mathbf{W}$*-demeaned data if and only if the matrices* $\mathbf{W}$, $\mathbf{W}\mathbf{G}_0$ $\mathbf{W}\widetilde{\mathbf{V}}$, *and* $\mathbf{W}\mathbf{G}_0\widetilde{\mathbf{V}}$ *are linearly independent.*

The following corollary gives further results for when local differencing is used to account for the network-specific fixed effects.

Corollary 3. *Suppose that Assumption 1 holds and let* $\mathbf{W} = \mathbf{W}_l = \mathbf{I} - \mathbf{G}_0$.

1. *The parameters* (γ, δ) *are identified from the mean restrictions for the* $\mathbf{W}_l$*-demeaned data if and only if the vectors* $(\mathbf{I} - \mathbf{G}_0)\widetilde{\mathbf{m}}$ *and* $(\mathbf{I} - \mathbf{G}_0)\mathbf{G}_0\widetilde{\mathbf{m}}$ *are linearly independent.*

2. *Suppose that either* $\gamma \neq 0$ *or* $\delta \neq 0$. *Then the parameters* $(\gamma, \delta, \sigma_u^2)$ *are identified from the covariance restrictions for the* $\mathbf{W}_l$*-demeaned data if and only if the matrices* $\mathbf{I}$, $\mathbf{G}_0$, $\mathbf{G}_0^2$, $\widetilde{\mathbf{V}}$, $\mathbf{G}_0\widetilde{\mathbf{V}}$, *and* $\mathbf{G}_0^2\widetilde{\mathbf{V}}$ *are linearly independent.*

The first part of this result is similar to that of Theorem 1, apart from the fact the intercept is now omitted due to the demeaning operation. The second part now imposes more restrictions, by introducing the matrices $\mathbf{G}_0^2$ and $\mathbf{G}_0^2\widetilde{\mathbf{V}}$ in the linear independence condition. This highlights that the presence of network-specific fixed effects imposes higher demands on $\mathbf{G}_0$ and $\widetilde{\mathbf{V}}$ and their interplay.

However, the following proposition shows that even for the extended model identification on the basis of the covariance restrictions remains generic.

Proposition 8. *Suppose that Assumption 1 holds and that the network is connected. Suppose that either* $\gamma \neq 0$ *or* $\delta \neq 0$ *and that there are at least two unique network positions. Then the parameters* (γ, δ) *are generically identified from the covariance restrictions in the presence of network fixed effects.*

5 Estimation

While the presence of 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 introduce straightforward and practical GMM and IV estimators, built upon the mean and covariance restrictions highlighted in the previous section. Importantly, these estimators are also applicable to setups with stochastic networks.

While the GMM approach can yield more efficient parameter estimates by utilizing a broader set of information, it does introduce nonlinearity into the estimation procedure. This nonlinearity can become a practical challenge when dealing with a substantial number of perfectly measured covariates, such as network-specific fixed effects, within the model. In Appendix A.4, we address the treatment of perfectly measured covariates within a two-step GMM estimator, following the framework of Erickson and Whited (2002). Importantly, this approach incurs only minimal additional computational burden. For clarity, we focus on the estimation of the baseline specification for the remainder of this section.

5.1 GMM estimator

An efficient GMM estimator can be constructed by utilizing the mean and covariance restrictions, as detailed in Section 4. These moment equations yield a potentially extensive system of nonlinear equations involving only four unknown parameters, namely α, γ, δ, and σ_u^2.

Proposition 9. *Suppose that Assumption 1 holds. Write the stacked mean and covariance restrictions for a fixed network* $\mathbf{G}_0$ *as*

$$\begin{aligned}
&\mathbb{E}(\mathbf{m}_s^1(\boldsymbol{\theta})) = \mathbb{E}(\mathbf{y}_s) - \alpha\boldsymbol{\iota} - \gamma\mathbb{E}(\widetilde{\mathbf{x}}_s) - \delta\mathbb{E}(\overline{\widetilde{\mathbf{x}}}_s) = \mathbf{0},\\
&\mathbb{E}(\mathbf{m}_s^2(\boldsymbol{\theta})) = vec\left[\mathbb{C}(\mathbf{y}_s, \widetilde{\mathbf{x}}_s) - \gamma[\mathbb{V}(\widetilde{\mathbf{x}}_s) - \sigma_u^2\mathbf{I}] - \delta\mathbf{G}_0[\mathbb{V}(\widetilde{\mathbf{x}}_s) - \sigma_u^2\mathbf{I}]\right] = \mathbf{0},
\end{aligned}$$

where $\boldsymbol{\theta} = (\alpha, \gamma, \delta, \sigma_u^2)$. *Under standard regularity conditions (Hansen, 1982), the GMM estimator*

$$\widehat{\boldsymbol{\theta}}^{GMM} = \arg\min \left(S^{-1}\sum_s \begin{bmatrix}\mathbf{m}_s^1(\boldsymbol{\theta})\\ \mathbf{m}_s^2(\boldsymbol{\theta})\end{bmatrix}\right)^{\intercal} \boldsymbol{\Omega} \left(S^{-1}\sum_s \begin{bmatrix}\mathbf{m}_s^1(\boldsymbol{\theta})\\ \mathbf{m}_s^2(\boldsymbol{\theta})\end{bmatrix}\right),$$

delivers consistent parameter estimates (i.e., plim $\widehat{\boldsymbol{\theta}}^{GMM} = \boldsymbol{\theta}$*) for every positive definite weighting matrix* $\boldsymbol{\Omega}$.

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 simplifies to a closed-form generalized least squares (GLS) estimator. This GLS estimator operates by using position-specific network-averaged outcomes, and own and peer characteristics as its inputs.

Corollary 4. *The GMM estimator that only uses the mean restrictions, i.e.,*

$$\widehat{\boldsymbol{\theta}}_1^{GMM} = \arg\min \left(S^{-1} \sum_s \mathbf{m}_s^1(\boldsymbol{\theta}) \right)^{\intercal} \boldsymbol{\Omega}_1 \left(S^{-1} \sum_s \mathbf{m}_s^1(\boldsymbol{\theta}) \right),$$

is numerically equivalent to the GLS estimator

$$\widehat{\boldsymbol{\theta}}^{GLS} = \left(\overline{\mathbf{Q}}^{\intercal} \boldsymbol{\Omega}_1 \overline{\mathbf{Q}} \right)^{-1} \overline{\mathbf{Q}}^{\intercal} \boldsymbol{\Omega}_1 \overline{\mathbf{y}},$$

where $\overline{\mathbf{y}} = S^{-1} \sum_s \mathbf{y}_s$, $\mathbf{Q}_s = [\boldsymbol{\iota}, \widetilde{\mathbf{x}}_s, \overline{\widetilde{\mathbf{x}}}_s]$, *and* $\overline{\mathbf{Q}} = S^{-1} \sum_s \mathbf{Q}_s$. *The optimal weighting matrix is given by* $\boldsymbol{\Omega}_1^* = [\mathbb{V}(\mathbf{e}_s) + \sigma_u^2(\gamma^2 \mathbf{I} + \delta^2 \mathbf{G}_0 \mathbf{G}_0^{\intercal})]^{-1}$.

The GLS estimator simplifies further to the OLS estimator $\widehat{\boldsymbol{\theta}}^{OLS} = \left(\overline{\mathbf{Q}}^{\intercal} \overline{\mathbf{Q}} \right)^{-1} \overline{\mathbf{Q}}^{\intercal} \overline{\mathbf{y}}$ when the weighting matrix is set to the identity matrix. However, it is important to note that this estimator is typically not efficient, even under the assumption of homoscedastic and uncorrelated disturbances. This inefficiency arises due to the presence of both measurement error and peer effects, introducing an additional source of dependence between observations. However, if either σ_u^2 or δ is small, the OLS estimator approaches efficiency.

When employing only the covariance restrictions in the estimation process, the optimal weighting matrix takes the form of $\boldsymbol{\Omega}_2^* = \frac{1}{N_s \mathbb{V}(e)} [\mathbb{V}(\mathbf{x}_s) + \sigma_u^2 \mathbf{I}]^{-1}$. Notably, this weighting matrix is proportional to the identity matrix when characteristics exhibit homoscedasticity and are uncorrelated among peers. This covariance structure arises naturally when network formation is independent of individual characteristics.

In scenarios where the same network structure is observed frequently, such as in experimental setups, the estimators presented in Proposition 9 and Corollary 4 can be implemented directly. However, in cases involving stochastic networks, it may be advantageous to aggregate the moment restrictions from different network positions within networks. This aggregation reduces the number of equations and mitigates the noise in the moment conditions.[15] We provide an illustrative example

[15]For general networks, determining all unique network positions can be computationally demanding. Additionally, if a unique network position appears only infrequently in the sample, the associated moment restrictions can be noisy, making aggregation across positions a recommended approach.

of such moment aggregation in our simulation study, detailed in Section 6.

5.2 IV estimator

The parameters of interest can also be recovered by means of an IV estimator. A key advantage of this approach is its simplicity in estimation. Network fixed effects and additional perfectly measured characteristics can be easily included.

Due to the presence of measurement error, both the own characteristic and the average characteristics of peers are plagued with endogeneity. To see this, substitute $\mathbf{x}_s = \widetilde{\mathbf{x}}_s - \mathbf{u}_s$ in Expression (1a), which gives that

$$\begin{aligned}\mathbf{y}_s &= \alpha\boldsymbol{\iota} + \gamma\widetilde{\mathbf{x}}_s + \delta\overline{\widetilde{\mathbf{x}}}_s + \mathbf{e}_s - \gamma\mathbf{u}_s - \delta\overline{\mathbf{u}}_s,\\ &= \alpha\boldsymbol{\iota} + \gamma\widetilde{\mathbf{x}}_s + \delta\overline{\widetilde{\mathbf{x}}}_s + \widetilde{\mathbf{e}}_s.\end{aligned}$$

The endogeneity problem arises because the characteristics $\widetilde{x}$ and $\overline{\widetilde{x}}$ are correlated with the composite error term $\widetilde{e}$: i.e., $\mathbb{C}(\widetilde{x}, \widetilde{e}) \neq 0$ and $\mathbb{C}(\overline{\widetilde{x}}, \widetilde{e}) \neq 0$ if $\gamma, \delta \neq 0$.

Since there are two endogenous regressors, one needs at least two instrumental variables z_1 and z_2 to obtain consistent estimates for the parameters of interest. Besides being uncorrelated with the composite error term, these instruments should also satisfy the standard rank condition $\det \begin{bmatrix}\mathbb{C}(z_1, \widetilde{x}) & \mathbb{C}(z_1, \overline{\widetilde{x}})\\ \mathbb{C}(z_2, \widetilde{x}) & \mathbb{C}(z_2, \overline{\widetilde{x}})\end{bmatrix} \neq 0$. Under these conditions, the IV estimator

$$\begin{bmatrix}\widehat{\gamma}^{IV}\\ \widehat{\delta}^{IV}\end{bmatrix} = \left(\begin{bmatrix}\mathbf{W}_b\mathbf{z}_1 & \mathbf{W}_b\mathbf{z}_2\end{bmatrix}^\intercal \begin{bmatrix}\mathbf{W}_b\widetilde{\mathbf{x}} & \mathbf{W}_b\overline{\widetilde{\mathbf{x}}}\end{bmatrix}\right)^{-1} \begin{bmatrix}\mathbf{W}_b\mathbf{z}_1 & \mathbf{W}_b\mathbf{z}_2\end{bmatrix}^\intercal \mathbf{W}_b\mathbf{y},$$

provides consistent estimates for the parameters of interest.

We explore two approaches for 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, when external information is at hand, such as repeated measurements of the characteristics, this information 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 provided by network-lagged characteristics. The relevance and validity of these instruments relies on the fact that individual characteristics are correlated across the network, whereas measurement error is not. If characteristics are correlated with links, the characteristics of peers at distances two or beyond can be used as instruments. The usage of network-lagged characteristics is reminiscent of Griliches and Hausman (1986), who employ time-lagged observations in the context of the linear panel data model.

Let $\mathbf{A}^{(t)}$ denote the zero-ones matrix that indicates if a pair is at distance t and let $\mathbf{G}^{(t)}$ denote its row-normalized version. The set of instruments $\{\mathbf{G}^{(t)}\widetilde{\mathbf{x}}\}_{t=2,3,\ldots,T}$ is valid under Assumption 1 as

$$\mathbb{C}\left((\mathbf{G}^{(t)}\widetilde{\mathbf{x}})_i, \widetilde{e}_i\right) = \mathbb{C}\left((\mathbf{G}^{(t)}\widetilde{\mathbf{x}})_i, e_i\right) - \gamma\mathbb{C}\left((\mathbf{G}^{(t)}\widetilde{\mathbf{x}})_i, u_i\right) - \delta\mathbb{C}\left((\mathbf{G}^{(t)}\widetilde{\mathbf{x}})_i, \overline{u}_i\right) = 0,$$

for $t = 2, 3, \ldots, T$. Under Assumption 3, the strength of these instruments with respect to the own characteristic is equal to $\mathbb{C}\left((\mathbf{G}^{(t)}\widetilde{\mathbf{x}})_i, \widetilde{x}_i\right) = \rho_t\sigma_x^2$.

Alternatively, the following results allows to exploit the dependence of individual characteristics across the network through a differencing procedure. By considering differencing matrices of a specific class, the dependency between the instrument and the composite error term, which drives endogeneity, is cancelled out.

Proposition 10. *Suppose that Assumption 1 holds. For every differencing matrix $\mathbf{W}_z \neq \mathbf{0}$, such that $\mathbf{W}_z\boldsymbol{\iota} = \mathbf{0}$ and diag$(\mathbf{W}_z) = \mathbf{0}$, the variable $\mathbf{W}_z\widetilde{\mathbf{x}}$ is a valid instrument.*

A leading example is the differencing matrix $\mathbf{W}_z = \mathbf{G} - \mathbf{G}^{(2)}$. The resulting instrument $(\mathbf{G} - \mathbf{G}^{(2)})\widetilde{\mathbf{x}}$ leverages differences in correlations one and two steps away. For example, for the equi-correlational model, we obtain $\mathbb{C}\left(\widetilde{x}_i, ((\mathbf{G} - \mathbf{G}^{(2)})\widetilde{\mathbf{x}})_i\right) = (\rho_1 - \rho_2)\sigma_x^2$. That is, the strength of this instrument with respect to the own characteristic is determined by the difference between correlations one and two steps away. Intuitively, this strategy exploits discontinuities in the correlation of characteristics as a function of network distance.

Network features Another potential source of instruments is provided by the features of the network. First consider those that are related to the mean restrictions. For instance, when individuals who have large values for the characteristic also have more connections, the degree constitutes a relevant and valid instrument as $\mathbb{C}(d, \widetilde{x}) \neq 0$ and $\mathbb{C}(d, \widetilde{e}) = 0$. The same holds for other network features, like local clustering, that might be correlated with the (true) characteristic.

Such features of the social network may also be used to construct instruments that implicitly exploit the covariance restrictions.

Proposition 11. *Suppose that Assumption 1 holds. Let $c_i(\mathbf{G})$ be a network characteristic. The instruments (i) c_i, (ii) $(c_i - \mathbb{E}[c])\widetilde{x}_i$, and (iii) $d_i(c_i - \mathbb{E}[c])\overline{\widetilde{x}}_i$ are valid. They are relevant when*

- $\mathbb{C}(c, x) \neq 0$ *or* $\mathbb{C}(c, \overline{x}) \neq 0$,

- $\mathbb{C}(c, \mathbb{V}(x \mid c)) \neq 0$ *or* $\mathbb{C}(c, \mathbb{C}(x, \overline{x} \mid c)) \neq 0$,
- $\mathbb{C}(c, d\mathbb{V}(\overline{x} \mid c)) \neq 0$ *or* $\mathbb{C}(c, d\mathbb{C}(x, \overline{x} \mid c)) \neq 0$,

respectively.

From Proposition 11 it is apparent that a network characteristic that is not a relevant instrument in the traditional sense (i.e., $\mathbb{C}(c, x) = \mathbb{C}(c, \overline{x}) = 0$), might still be useful if it is correlated with the conditional variance of the individual characteristic (i.e., $\mathbb{C}(c, \mathbb{V}(x \mid c)) \neq 0$) or with conditional homophily (i.e., $\mathbb{C}(c, \mathbb{C}(x, \overline{x} \mid c)) \neq 0$). An instrument that is correlated with the degree-corrected conditional variance of the peer characteristic (i.e., $\mathbb{C}(c, d\mathbb{V}(\overline{x} \mid c)) \neq 0$) is also relevant.[16] For example, Lemma 1 suggest that local clustering induces heteroscedasticity in the average peer characteristics, as it affects the variance $\mathbb{V}(\overline{x})$ across positions.

Finally, although network-lagged characteristics and network features provide appealing and intuitive instruments, they might also exhibit weak first-stages. It is therefore advisable to test for the presence of weak instruments.

6 Simulation

We illustrate our main results for the expansion bias numerically through a Monte Carlo exercise. Our simulations suggest that even in settings where Assumption 2 is not satisfied, our expressions for the bias provide an adequate description of the comparative statics. They also illustrate that our GMM and IV estimators behave well in finite samples.

6.1 Setup

We consider a setting with a large number of small stochastic networks. In each simulation, S networks are generated at random by means of a dyadic model of network formation. We fix the number of individuals within every network at $N_s = 20$, as this corresponds to a common setting in empirical applications. We repeat this simulation procedure $R = 200$ times.

For every individual i within a network s, we draw characteristics at random. In particular, we assume that

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

where $\xi_s \sim \text{Normal}(10, 1)$ and $\xi_{si} \sim \text{Normal}(0, 2)$.

[16] The degree correction ensures that the instrument picks up conditional heteroscedasticity different from the one that operates through peer averaging (e.g., see Expression (6)).

Links between individuals within a given network are assumed to be formed through the undirected dyadic specification

$$a_{sij} = \mathbb{I}\left[\mu - \nu|x_{si} - x_{sj}| + f_i + f_j \geq w_{sij}\right], \tag{10}$$

where μ is an intercept that controls the overall propensity to form links and ν is a parameter that controls homophily. A large value of ν implies that individuals with dissimilar characteristics tends to link less often on average. We further assume that the individual-specific and pair-specific link shifters are distributed $f_i, f_j \sim$ Normal$(0, 1)$ and $w_{sij} \sim$ Logistic$(0, 1)$, respectively. The inclusion of the individual-specific random effects f_i and f_j allows for more realistic variation in the degree distribution. Note that the probability that a link is formed amounts to

$$\Pr[a_{sij} = 1 \mid x_{si}, x_{sj}, f_i, f_j] = \frac{1}{1 + \exp(-(\mu - \nu|x_{si} - x_{sj}| + f_i + f_j))}.$$

Individuals' outcomes are generated by setting the true peer effect to zero (i.e., $\delta = 0$) and the true own effect to one (i.e., $\gamma = 1$). In particular, we specify that

$$y_{si} = 1 + x_{si} + e_{si},$$

where $e_{si} \sim$ Normal$(0, 1)$. Measurement error is distributed $u_{si} \sim$ Normal$(0, 0.3)$, which implies that the noise-to-signal ratio amounts to $\phi = 0.1$.

6.2 Results

Expansion bias. We first assess how the asymptotic bias varies with μ and ν, the parameters that determine network formation through the dyadic specification. Table 1 reports the biases for the peer effect parameter δ using the non-demeaned as well as the globally and locally demeaned OLS estimators. Notice that in our setup, each of these three estimators are consistent in the absence of measurement error. For the purpose of the analysis of the expansion bias, if an individual remains isolated, we assign her a random peer.

Overall, we find that the expansion bias in the non-demeaned estimator is smaller than that of the globally and locally demeaned estimators, the latter two being rather similar. To assess the applicability of our theoretical expression for the bias in Proposition 3, we conduct a metaregression analysis. For each of the 30 possible data generating processes in Table 1, we regress the asymptotic bias for the non-demeaned estimator on the implied correlations at distances 1 and 2, average inverse degree and local clustering. The coefficients on ρ_1, ρ_2, and h_0 have the signs predicted by Proposition 3 , which suggests that the theoretical approximation provides an adequate description of the bias.

Table 1: Asymptotic bias in OLS estimates for δ

μ	ν					
	0.0	0.4	0.8	1.2	1.6	2.0
	Non-demeaned					
-2	0.05	0.07	0.07	0.07	0.07	0.06
-1	0.08	0.10	0.10	0.11	0.12	0.10
0	0.07	0.12	0.16	0.18	0.15	0.14
1	0.11	0.16	0.21	0.25	0.23	0.22
2	0.11	0.13	0.18	0.28	0.31	0.29
	Globally demeaned					
-2	-0.02	0.04	0.04	0.05	0.04	0.03
-1	0.03	0.08	0.09	0.11	0.10	0.07
0	-0.05	0.13	0.19	0.19	0.16	0.12
1	-0.02	0.24	0.34	0.32	0.28	0.22
2	0.07	0.38	0.45	0.54	0.46	0.39
	Locally demeaned					
-2	-0.03	0.04	0.02	0.05	0.05	0.03
-1	0.05	0.10	0.11	0.11	0.10	0.05
0	-0.07	0.11	0.18	0.18	0.13	0.11
1	-0.02	0.22	0.34	0.30	0.25	0.14
2	0.09	0.39	0.42	0.52	0.38	0.33

Results for parameter values $(\gamma, \delta) = (1, 0)$.

Estimation. We now assess the finite sample performance of the GMM and IV estimators proposed in Section 5. Our results suggest that even the parsimonious dyadic model in Expression (10) generates sufficient interdependence between characteristics and links to recover the parameters of interest. For simplicity, we set $(\mu, \nu) = (1, 1)$, which corresponds to moderate values.

As the setup entails stochastic networks, to operationalize the GMM estimator, we aggregate the moment conditions across network positions. This reduces the number of moment equations and increases their precision. We choose the mean and covariance restrictions

$$\overline{\mathbf{m}}_s^1(\boldsymbol{\theta}) = \sum_i m_{si}^1(\boldsymbol{\theta}),$$

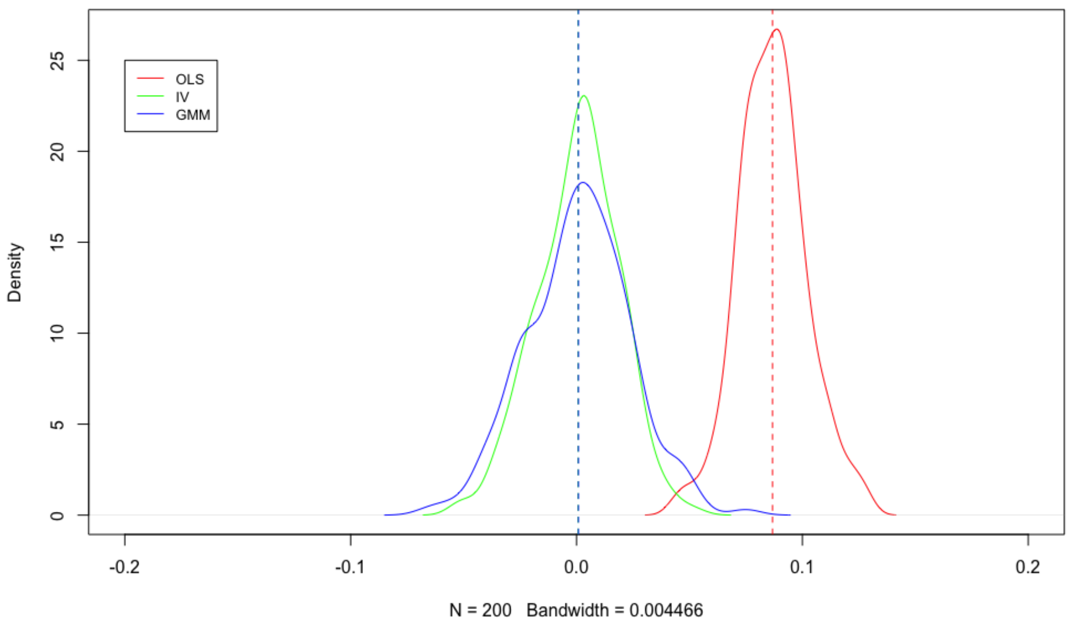
$$\overline{\mathbf{m}}_s^1(\boldsymbol{\theta}) = \begin{bmatrix} \sum_i m_{sii}^2(\boldsymbol{\theta}) \\ \sum_{i,j} \mathbb{I}[d(i,j) = 1] m_{sij}^2(\boldsymbol{\theta}) \\ \sum_{i,j} \mathbb{I}[d(i,j) = 2] m_{sij}^2(\boldsymbol{\theta}) \\ \sum_{i,j} \mathbb{I}[d(i,j) \geq 3] m_{sij}^2(\boldsymbol{\theta}) \end{bmatrix},$$

which are averaged across positions. Together, they yield a system of five equations and four unknowns. The estimates are obtained by means of two-step GMM. To operationalize the IV estimator, we choose two straightforward instruments based on lagged characteristics: $\mathbf{z}_1 = \mathbf{G}^{(2)}\widetilde{\mathbf{x}}$ and $\mathbf{z}_2 = (\mathbf{G} - \mathbf{G}^{(2)})\widetilde{\mathbf{x}}$.

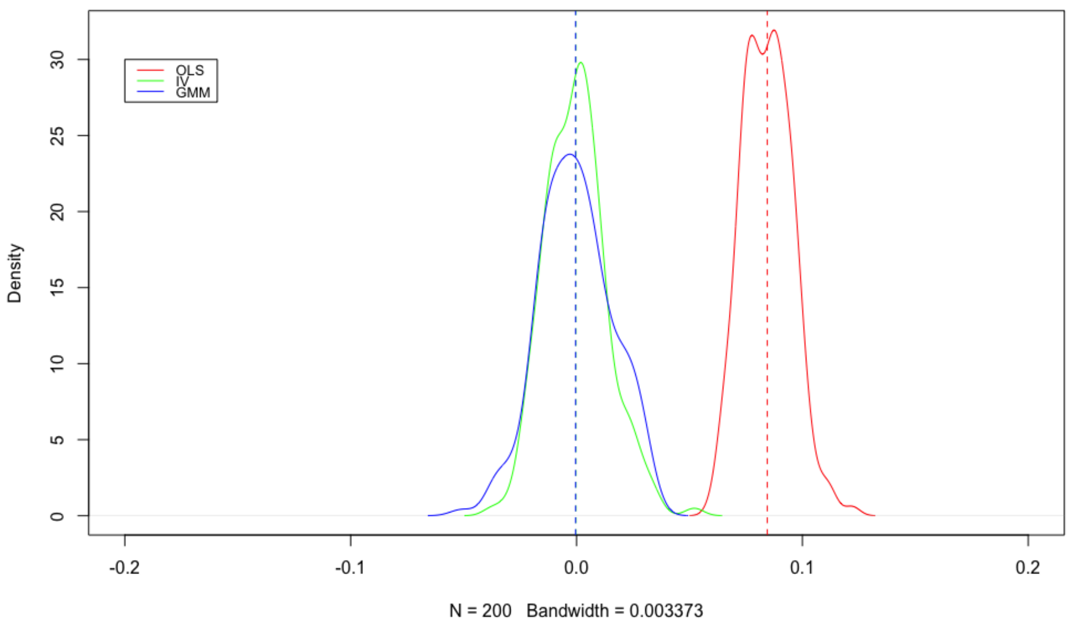
Figure 1 displays the empirical distribution of the OLS, IV, and GMM estimators for the peer effect parameter δ. First consider the results for samples containing $S = 250$ networks in Figure 1a. As before, the OLS estimates exhibit an expansion bias with a mean value of 0.0868 (indicated by a dotted vertical line) and a standard deviation of 0.0151. The IV and GMM estimates, however, are approximately centered around the true value with a mean value of 0.0007 and 0.0008 (again indicated by a dotted vertical line) and a standard deviation of 0.0182 and 0.0228, respectively. Although the sample variation of the IV and GMM estimators is slightly larger than that of the OLS estimator, both estimators are still quite precise. The IV and GMM estimators exhibit a comparable performance, which is not unusual since the latter only exploits a few moment conditions. As expected, the variation for samples with $S = 500$ networks (displayed in Figure 1b) is even smaller. For the OLS, IV, and GMM estimates we obtain standard deviations of 0.0108, 0.0137, and 0.0160, respectively.

7 Concluding remarks

In this paper, we study errors-in-variables in the linear-in-means model of social interactions. Due to the special structure of this model, the coefficient on the peer effect exhibits an asymptotic expansion bias in the presence of homophily. Homophily is a widely recognized phenomenon in social networks; false positives might therefore be a first-order concern in empirical studies. We demonstrate how the magnitude of this bias critically depends on network statistics, such as the average inverse degree and clustering. While the interdependence between individual characteristics and network formation contributes to this bias, it also offers an avenue for the consistent estimation of the parameters of interest, eliminating the need for external information. We propose GMM and IV estimators that are straightforward to implement.



(a) Sample size $S = 250$



(b) Sample size $S = 500$

Figure 1: Empirical distribution of the OLS, IV, and GMM estimates for δ

To illustrate the efficacy of these methods, we conduct a Monte Carlo simulation.

This paper constitutes a necessary first step in studying measurement error and peer effects in networks. Looking ahead, we envision at least three promising avenues for future research. Firstly, it would be worthwhile to extend our analysis to settings where measurement error is non-classical. Recent research by Balestra, Eugster, and Puljic (2023) has shown that under misclasification of a binary characteristic, an expansion bias can even arise under random group formation.[17] We hypothesise that (i) this additional expansion bias will emerge in a network setting as well and that (ii) under non-random assignment, the usual expansion bias will reappear with qualitatively similar features to the one studied in this paper.

Secondly, an extension of our analysis to models that accommodate endogenous peer effects will provide valuable insights. We conjecture that with the inclusion of endogenous peer effects, homophily remains an important driver of inflated peer effects estimates. Moreover, precursory results by de Paula (2017) suggest that covariance restrictions might continue to play a key role in the identification and estimation of this class of models.

Thirdly, an exploration of the interplay between mismeasured characteristics and mismeasured links could offer a deeper and more complete understanding of the impact of measurement error on social interaction modeling. 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, can interact with the expansion bias due to mismeasured characteristics.

[17]This additional expansion bias arises due to the correlation between the average peer characteristic and the misclassification error in the individual characteristic.

References

Abel, A. B. (2018). *Classical measurement error with several regressors* (Tech. Rep.). 11

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. 1, 4, 18

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics*, *30*, 98–108. 1, 4

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) 33

Ben-Moshe, D. (2021). Identification of linear regressions with errors in all variables. *Econometric Theory*, *37*(4), 633–663. 5

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. 5

Biggs, N. (1993). *Algebraic graph theory*. Cambridge: Cambridge University Press. A17

Blume, L., Brock, W., Durlauf, S., & Jayaraman, R. (2015). Linear social interactions models. *Journal of Political Economy*, *123*(2), 444-496. 4

Boucher, V., & Houndetoungan, E. A. (2023). *Estimating peer effects using partial network data* (Tech. Rep.). 1

Bramoullé, Y., Djebbari, H., & Fortin, B. (2020). Peer effects in networks: A survey. *Annual Review of Economics*, *12*(1), 603-629. 1

Caeyers, B., & Fafchamps, M. (2023). *Exclusion bias and the estimation of peer effects* (Tech. Rep.). 3, 11, 18

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. 1

Corno, L., La Ferrara, E., & Burns, J. (2022). Interaction, stereotypes, and performance: Evidence from south africa. *American Economic Review*, *112*(12), 3848-75. 1

Cragg, J. G. (1997). Using higher moments to estimate the simple errors-in-variables model. *The RAND Journal of Economics*, *28*, S71–S91. 5

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. 5

de Paula, A., Rasul, I., & Souza, P. C. (2019). Identifying network ties from panel data: theory and an application to tax competition. *CEMMAP Working Paper*, 1–78. 1

De Giorgi, G., Frederiksen, A., & Pistaferri, L. (2019). Consumption Network Effects. *The Review of Economic Studies*, *87*(1), 130-163. 4

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. 1, 4, 33

Erickson, T., & Whited, T. M. (2002). Two-step GMM estimatione of the errors-in-variables model using high-order moments. *Econometric Theory*, *18*, 776–799. 5, 24, A3, A5

Evdokimov, K. S., & Zeleneev, A. (2020). *Errors-in-variables in large nonlinear panel and network models* (Tech. Rep.). 5

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. 1, 4, 15

Goldsmith-Pinkham, P., & Imbens, G. W. (2013). Social networks and the identification of peer effects. *Journal of Business Economic Statistics*, *31*(3), 253–264. 1

Greene, W. H. (2003). *Econometric analysis*. Upple Saddle River, NJ: Pearson Education Inc. 2

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. 1

Griffith, A. (2022b). Random Assignment with Non-Random Peers: A Structural Approach to Counterfactual Treatment Assessment. *The Review of Economics and Statistics*, 1-40. 1

Griliches, Z., & Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, *31*, 93–118. 3, 4, 18, 27

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50*(4), 1029–1054. 24, A23

Hsieh, C., & Lee, L. F. (2016). A Social Interactions Model with Endogenous Friendship Formation and Selectivity. *Journal of Applied Econometrics*, *31*(2), 301-319. 1

Klepper, S., & Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica*, *52*(1), 163–183. 5

Koopmans, T. C. (1937). *Linear regression analysis of economic time series*. Haarlem: De Erven F. Bohn N.V. 5

Lavy, V., & Schlosser, A. (2011). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, *3*(2), 1–33. 4

Lewbel, A. (1997). Constructing instruments for regressions with measurement error when no additional data are available, with an application to patents and rd. *Econometrica*, *65*(5), 1201–1213. 5

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business Economic Statistics*, *30*(1), 67-80. 5

Lewbel, A., Qu, X., & Tang, X. (2023). *Ignoring measurement errors in social networks* (Tech. Rep.). 1

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. 4

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *Review of Economic Studies*, *60*(3), 531-542. 1, 4

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*, 415–444. 2

Mityagin, B. S. (2020). The zero set of a real analytic function. *Mathematical Notes*, *107*, 529-530. A18

Moffitt, R. (2001). Policy interventions, low-level equilibria, and social interactions. In D. Steven & Y. Peyton (Eds.), *Social Dynamics.* The MIT Press. 1, 4

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. 4

Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, *18*(4), 375–389. 5

Reiersøl, O. (1941). Confluence analysis by means of lag moments and other methods of confluence analysis. *Econometrica*, *9*(1), 1–24. 3

Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates. *The Quarterly Journal of Economics*, *116*(2), 681-704. 1

Schennach, S. M. (2007). Instrumental variable estimation of nonlinear errors-in-variables models. *Econometrica*, *75*(1), 201–239. 3

Wansbeek, T., & Meijer, E. (2000). *Measurement error and latent variables in econometrics* (Vol. 37). Amsterdam: Elsevier. 9

Online Appendix

Measurement Error and Peer Effects in Networks

A Additional results

A.1 Asymptotic bias for groups with leave-in means

We provide expressions for the asymptotic bias in the case that peer effects are defined on the basis as leave-in means. In this setting, we have that

$$\mathbf{\Sigma} = \sigma_u^2 \begin{bmatrix} 1 & N_0^{-1} \\ N_0^{-1} & N_0^{-1} \end{bmatrix}.$$

Similar to Proposition 1, one can show that

$$\text{plim } \widehat{\delta}^{OLS} = \frac{\mathbb{C}(x,\overline{x}) - N_0^{-1}\mathbb{V}(x)}{\Delta + (N_0^{-1}\mathbb{V}(x) + \mathbb{V}(\overline{x}) - 2N_0^{-1}\mathbb{C}(x,\overline{x}))\sigma_u^2 + (N_0^{-2}(N_0-1))\sigma_u^4}\sigma_u^2\gamma,$$

if $\delta = 0$. The counterparts to Lemma 1 and Corollary 2 are as follows.

Lemma A.1. *Suppose that Assumption 2 holds. If the social graph takes the form of a group, then we have that*

$$\mathbb{C}(x,\overline{x}) = \left(\frac{1}{N_0} + \frac{N_0-1}{N_0}\rho_1\right)\sigma_x^2,$$
$$\mathbb{V}(\overline{x}) = \left(\frac{1}{N_0} + \frac{N_0-1}{N_0}\rho_1\right)\sigma_x^2.$$

Corollary A.1. *Suppose that Assumptions 1 and 2 hold and suppose that there is no peer effect, i.e., $\delta = 0$. If the social graph takes the form of a group, then $\widehat{\delta}^{OLS}$ has the asymptotic bias*

$$\textit{plim } \widehat{\delta}^{OLS} = \frac{\rho_1\phi}{(1+\phi)(\frac{1}{N_0} + (1-\frac{2}{N_0})\rho_1) - \frac{N_0-1}{N_0}\rho_1^2 + \frac{1}{N_0}\phi + \frac{1}{N_0}\phi^2}\gamma.$$

A.2 Convergence of the asymptotic bias in the baseline and globally demeaned estimates

From the results in Section B.6, we know that

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma} - N_0^{-1}\begin{bmatrix}\sigma_u^2 & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 N_0^{-1}(\boldsymbol{\iota}^\intercal \mathbf{G}\mathbf{G}^\intercal \boldsymbol{\iota})\end{bmatrix},$$
$$\mathbf{S}_g = \mathbf{S} - N_0^{-1}\begin{bmatrix}\mathbb{C}(x_{si}, (\mathbf{J}\mathbf{x})_{si}) & \mathbb{C}(x_{si}, (\mathbf{J}\overline{\mathbf{x}})_{si}) \\ \mathbb{C}(x_{si}, (\mathbf{J}\overline{\mathbf{x}})_{si}) & \mathbb{C}(\overline{x}_{si}, (\mathbf{J}\overline{\mathbf{x}})_{si})\end{bmatrix}.$$

If the degree distribution is stable as N_0 grows large, we have that

$$\begin{aligned} N_0^{-1}(\boldsymbol{\iota}^\intercal \mathbf{G}\mathbf{G}^\intercal \boldsymbol{\iota}) &= N_0^{-1}(\mathbf{G}^\intercal \boldsymbol{\iota})^\intercal \mathbf{G}^\intercal \boldsymbol{\iota} \\ &\leq N_0^{-1}(\mathbf{A}\boldsymbol{\iota})^\intercal \mathbf{A}\boldsymbol{\iota} \\ &= N_0^{-1}\mathbf{d}^\intercal \mathbf{d}, \end{aligned}$$

converges to a constant at most. In addition, if one assumes that

$$\begin{aligned} \lim_{N_0\to\infty} N_0^{-1}\mathbb{C}(x_{si}, (\mathbf{J}\mathbf{x})_{si}) &= 0, \\ \lim_{N_0\to\infty} N_0^{-1}\mathbb{C}(x_{si}, (\mathbf{J}\overline{\mathbf{x}})_{si}) &= 0, \\ \lim_{N_0\to\infty} N_0^{-1}\mathbb{C}(\overline{x}_{si}, (\mathbf{J}\overline{\mathbf{x}})_{si}) &= 0, \end{aligned} \tag{11}$$

by the sum and product rule for limits, it holds that

$$\begin{aligned} \lim_{N_0\to\infty} \operatorname{plim} \begin{bmatrix}\widehat{\gamma}_g^{OLS} \\ \widehat{\delta}_g^{OLS}\end{bmatrix} &= \lim_{N_0\to\infty} (\mathbf{S}_g + \boldsymbol{\Sigma}_g)^{-1}\mathbf{S}_g \begin{bmatrix}\gamma \\ \delta\end{bmatrix} \\ &= (\mathbf{S} + \boldsymbol{\Sigma})^{-1}\mathbf{S}\begin{bmatrix}\gamma \\ \delta\end{bmatrix} \\ &= \operatorname{plim}\begin{bmatrix}\widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS}\end{bmatrix}. \end{aligned}$$

Intuitively, the conditions in Expression (11) hold if correlations between characteristics decrease with network distance and if the average distance between nodes increases with N_0.

A.3 Asymptotic bias for strongly regular graphs

We now consider the global transformation's asymptotic bias for the class of strongly regular graphs. The following lemma will be useful to obtain a closed-form expression for the bias.

Lemma A.2. *Suppose that Assumption 3 holds and suppose the graph is strongly regular with parameters (N_0, d, λ). Then we have that*

$$h_0 = \frac{1}{d},$$
$$h_1 = \frac{\lambda}{d},$$
$$h_2 = 1 - \frac{\lambda + 1}{d},$$

and

$$m_{10} = m_{20} = m_{30} = \frac{1}{N_0},$$
$$m_{11} = m_{21} = m_{31} = \frac{d}{N_0},$$
$$m_{12} = m_{22} = m_{32} = 1 - \frac{d+1}{N_0},$$

for the globally demeaned differences (see Lemma A.3).

Using this lemma we can state the expression for the bias.

Proposition A.1. *Suppose that Assumptions 1 and 3 hold and that there is no peer effect, i.e., $\delta = 0$. If the graph is strongly regular, $\sigma_u^2 \approx 0$, and $N_0 \to \infty$, the asymptotic bias of $\widehat{\delta}_g^{OLS}$ is*

$$\text{plim } \widehat{\delta}_g^{OLS} \approx \frac{\rho_1 - \rho_2}{h_0(1-\rho_2)^2 + h_1(\rho_1 - \rho_2)(1-\rho_2) - (\rho_1 - \rho_2)^2} \sigma_u^2 \gamma.$$

A.4 GMM estimation with perfectly measured covariates

In this section, we discuss how to allow for a large number of additional perfectly measured covariates in GMM estimation. A two-step procedure similar to the one developed by Erickson and Whited (2002) provides a computationally attractive approach to estimate the parameters of interest. The two-step nature of this approach reduces the number of parameters that enter the nonlinear objective function, greatly facilitating estimation.

We consider the specification

$$\mathbf{y}_s = \mathbf{F}_s \boldsymbol{\alpha} + \gamma \mathbf{x}_s + \delta \overline{\mathbf{x}}_s + \mathbf{e}_s, \tag{12}$$

where $\mathbf{F}_s$ denotes a matrix that contains these perfectly measured covariates (e.g., network fixed effects). The disturbance is assumed to satisfy the conditional mean independence condition $\mathbb{E}(\mathbf{e}_s \mid, \mathbf{F}_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{u}_s) = 0$. As before, we assume that the conditional mean, variance and covariance of the measurement error do not depend on any of the other variables in the model, nor on the structure of the social network.

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

$$\begin{aligned}\mathbb{E}(u_{si} \mid \mathbf{F}_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) &= 0,\\ \mathbb{E}(u_{si}^2 \mid \mathbf{F}_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) &= \sigma_u^2,\\ \mathbb{E}(u_{si}u_{sj} \mid \mathbf{F}_s, \mathbf{x}_s, \mathbf{G}_s, \mathbf{e}_s) &= 0.\end{aligned}$$

The procedure makes use of the following two consecutive steps. In the first step, we partial out the perfectly measured variables by the use of auxiliary OLS regressions. We regress outcomes, own characteristics, and peer characteristics on the perfectly measured covariates and rewrite the model in terms of the associated residuals. Formally, the residuals of the OLS regression of a given vector $\mathbf{b}_s$ on the columns of $\mathbf{F}_s$ take the form $\mathbf{P}_s\mathbf{b}_s$, where $\mathbf{P}_s = \mathbf{I} - \mathbf{F}_s(\mathbf{F}_s^\intercal\mathbf{F}_s)^{-1}\mathbf{F}_s^\intercal$ is an annihilator matrix. Premultiplying all variables in Expression (12) by this annihilator matrix, we obtain

$$\mathbf{P}_s\mathbf{y}_s = \gamma\mathbf{P}_s\mathbf{x}_s + \delta\mathbf{P}_s\overline{\mathbf{x}}_s + \mathbf{P}_s\mathbf{e}_s, \tag{13}$$

which does not contain the possibly high-dimensional parameter vector $\boldsymbol{\alpha}$ anymore.

In the second step, we perform GMM estimation on the basis of the mean and covariance restrictions of the transformed model in Expression (13). Similar to Equations (8) and (9) for the baseline model, for a fixed network $\mathbf{G}_0$ it holds that

$$\begin{aligned}\mathbb{E}(\mathbf{P}_s\mathbf{y}_s) &= \gamma\mathbb{E}(\mathbf{P}_s\widetilde{\mathbf{x}}_s) + \delta\mathbb{E}(\mathbf{P}_s\overline{\widetilde{\mathbf{x}}}_s),\\ \mathbb{C}(\mathbf{P}_s\mathbf{y}_s, \mathbf{P}_s\widetilde{\mathbf{x}}_s) &= \gamma\Big[\mathbb{V}(\mathbf{P}_s\widetilde{\mathbf{x}}_s) - \sigma_u^2\mathbb{E}(\mathbf{P}_s)\Big] + \delta\Big[\mathbb{C}(\mathbf{P}_s\widetilde{\mathbf{x}}_s, \mathbf{P}_s\overline{\widetilde{\mathbf{x}}}_s) - \sigma_u^2\mathbb{E}(\mathbf{P}_s\mathbf{G}_0\mathbf{P}_s)\Big].\end{aligned}$$

The second and fourth terms in the convariance restrictions now depend on $\mathbf{P}_s$ to account for the fact that the data is transformed. Together, these restrictions provide a potentially large system of equations in only three unknowns (i.e., γ, δ, and σ_u^2).

The following proposition formalizes the asymptotic properties of this two-step GMM estimator.

Proposition A.2. *Suppose that Assumption A.1 holds. Write the stacked mean and covari-*

ance restrictions for a fixed network $\mathbf{G}_0$ as

$$
\begin{aligned}
\mathbb{E}(\mathbf{m}^1_{p,s}(\boldsymbol{\theta})) &= \mathbb{E}(\mathbf{P}_s\mathbf{y}_s) - \gamma\mathbb{E}(\mathbf{P}_s\widetilde{\mathbf{x}}_s) - \delta\mathbb{E}(\mathbf{P}_s\overline{\widetilde{\mathbf{x}}}_s) \\
&= \mathbf{0}, \\
\mathbb{E}(\mathbf{m}^2_{p,s}(\boldsymbol{\theta})) &= vec\left[\mathbb{C}(\mathbf{P}_s\mathbf{y}_s, \mathbf{P}_s\widetilde{\mathbf{x}}_s) - \gamma\left[\mathbb{V}(\mathbf{P}_s\widetilde{\mathbf{x}}_s) - \sigma_u^2\mathbb{E}(\mathbf{P}_s)\right] - \delta\left[\mathbb{C}(\mathbf{P}_s\widetilde{\mathbf{x}}_s, \mathbf{P}_s\overline{\widetilde{\mathbf{x}}}_s) - \sigma_u^2\mathbb{E}(\mathbf{P}_s\mathbf{G}_0\mathbf{P}_s)\right]\right] \\
&= \mathbf{0},
\end{aligned}
$$

where $\boldsymbol{\theta} = (\gamma, \delta, \sigma_u^2)$. Under standard regularity conditions (Erickson & Whited, 2002), the GMM estimator

$$
\widehat{\boldsymbol{\theta}}_p^{GMM} = \arg\min \left(S^{-1}\sum_s \begin{bmatrix}\mathbf{m}^1_{p,s}(\boldsymbol{\theta}) \\ \mathbf{m}^2_{p,s}(\boldsymbol{\theta})\end{bmatrix}\right)^{\intercal} \boldsymbol{\Omega} \left(S^{-1}\sum_s \begin{bmatrix}\mathbf{m}^1_{p,s}(\boldsymbol{\theta}) \\ \mathbf{m}^2_{p,s}(\boldsymbol{\theta})\end{bmatrix}\right),
$$

delivers consistent parameter estimates (i.e., plim $\widehat{\boldsymbol{\theta}}_p^{GMM} = \boldsymbol{\theta}$) for every positive definite weighting matrix $\boldsymbol{\Omega}$.

A.5 Additional results for the extended model

Lemma A.3. *Suppose that Assumption 3 holds. Then we have that*

$$
\begin{aligned}
\mathbb{V}_g(x_{si}) &= 1 - \sum_{t=0}^{T} m_{1t}\rho_t, \\
\mathbb{C}_g(x_{si}, \overline{x}_{si}) &= \rho_1 - \sum_{t=0}^{T} m_{2t}\rho_t, \\
\mathbb{V}_g(\overline{x}_{si}) &= \sum_{t=0}^{2} h_t\rho_t - \sum_{t=0}^{T} m_{3t}\rho_t,
\end{aligned}
$$

for the globally demeaned differences, and

$$
\begin{aligned}
\mathbb{V}_l(x_{si}) &= 1 + \sum_{t=0}^{2} h_t\rho_t - 2\rho_1, \\
\mathbb{C}_l(x_{si}, \overline{x}_{si}) &= \rho_1 - \sum_{t=0}^{2}(n_{1t} - h_t)\rho_t + \sum_{t=0}^{3} n_{2t}\rho_t, \\
\mathbb{V}_l(\overline{x}_{si}) &= \sum_{t=0}^{2} h_t\rho_t + \sum_{t=0}^{4} n_{3t}\rho_t - 2\sum_{t=0}^{3} n_{2t}\rho_t,
\end{aligned}
$$

for the locally demeaned differences, where

$$
\begin{aligned}
m_{1t} &= N_0^{-2}\boldsymbol{\iota}^\intercal \mathbf{A}_0^{(t)}\boldsymbol{\iota}, \\
m_{2t} &= N_0^{-2}\boldsymbol{\iota}^\intercal \mathbf{G}_0\mathbf{A}_0^{(t)}\boldsymbol{\iota}, \\
m_{3t} &= N_0^{-2}\boldsymbol{\iota}^\intercal \mathbf{G}_0\mathbf{A}_0^{(t)}\mathbf{G}_0^\intercal\boldsymbol{\iota}, \\
n_{1t} &= N_0^{-1}\mathit{trace}(\mathbf{G}_0^2\mathbf{A}_0^{(t)}), \\
n_{2t} &= N_0^{-1}\mathit{trace}(\mathbf{G}_0^\intercal\mathbf{G}_0^2\mathbf{A}_0^{(t)}), \\
n_{3t} &= N_0^{-1}\mathit{trace}((\mathbf{G}_0^\intercal)^2\mathbf{G}_0^2\mathbf{A}_0^{(t)}).
\end{aligned}
$$

B Proofs for Section 3

B.1 Proof of Propositions 1 and 2

Expanding the expression in (4), using (6), we get that

$$\operatorname{plim} \begin{bmatrix} \widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS} \end{bmatrix} = \frac{1}{D} \begin{bmatrix} \Delta + h_0 \mathbb{V}(x)\sigma_u^2 & h_0 \mathbb{C}(x, \overline{x})\sigma_u^2 \\ \mathbb{C}(x, \overline{x})\sigma_u^2 & \Delta + \mathbb{V}(\overline{x})\sigma_u^2 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix},$$

where

$$D = \Delta + h_0 \mathbb{V}(x)\sigma_u^2 + \mathbb{V}(\overline{x})\sigma_u^2 + h_0 \sigma_u^4,$$
$$\Delta = \det(\mathbf{S}) = \mathbb{V}(x)\mathbb{V}(\overline{x}) - \mathbb{C}(x, \overline{x})^2.$$

In particular, when there is no peer effect (i.e., $\delta = 0$), the expressions simplify to

$$\operatorname{plim} \widehat{\gamma}^{OLS} = \frac{\Delta + h_0 \mathbb{V}(x)\sigma_u^2}{\Delta + h_0 \mathbb{V}(x)\sigma_u^2 + \mathbb{V}(\overline{x})\sigma_u^2 + h_0 \sigma_u^4}\gamma,$$
$$\operatorname{plim} \widehat{\delta}^{OLS} = \frac{\mathbb{C}(x, \overline{x})\sigma_u^2}{\Delta + h_0 \mathbb{V}(x)\sigma_u^2 + \mathbb{V}(\overline{x})\sigma_u^2 + h_0 \sigma_u^4}\gamma.$$

Since the variance-covariance matrix $\mathbf{S}$ is positive semi-definite, we have that $\Delta \geq 0$, such that there will be an attenuation bias on $\widehat{\gamma}^{OLS}$ and an expansion bias on $\widehat{\delta}^{OLS}$ if $\mathbb{C}(x, \overline{x}) > 0$. Taking the partial derivative of the expansion bias with respect to σ_u^2, we find that

$$\frac{\partial(\operatorname{plim} \widehat{\delta}^{OLS}/\gamma)}{\partial \sigma_u^2} = \frac{\mathbb{C}(x, \overline{x})(\Delta - h_0 \sigma_u^4)}{(\Delta + h_0 \mathbb{V}(x)\sigma_u^2 + \mathbb{V}(\overline{x})\sigma_u^2 + h_0 \sigma_u^4)^2}.$$

If $\mathbb{C}(x, \overline{x}) > 0$, the expansion bias first increases and then decreases when measurement error increases. When $\sigma_u^2 = 0$ or $+\infty$ the expansion bias is equal to zero. It is maximal where

$$\frac{\partial(\operatorname{plim} \widehat{\delta}^{OLS}/\gamma)}{\partial \sigma_u^2} = 0,$$

which gives

$$\arg\max_{\sigma_u^2} \left\{ \operatorname{plim} \widehat{\delta}^{OLS}/\gamma \right\} = \sqrt{h_0^{-1}\Delta}.$$

The maximum takes the value

$$\max_{\sigma_u^2} \left\{ \operatorname{plim} \widehat{\delta}^{OLS}/\gamma \right\} = \frac{\mathbb{C}(x, \overline{x})}{2\sqrt{h_0 \Delta} + \mathbb{V}(x)}.$$

B.2 Proof of Corollary 1

For $\sigma_u^2 \approx 0$, we can approximate $(\mathbf{S}+\boldsymbol{\Sigma})^{-1}\mathbf{S} \approx \mathbf{I} - \mathbf{S}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_u^2}\sigma_u^2$. Therefore,

$$\begin{aligned}
\text{plim} \begin{bmatrix} \widehat{\gamma}^{OLS} \\ \widehat{\delta}^{OLS} \end{bmatrix} &\approx \left(\mathbf{I} - \mathbf{S}^{-1}\frac{\partial \boldsymbol{\Sigma}}{\partial \sigma_u^2}\sigma_u^2\right) \begin{bmatrix} \gamma \\ \delta \end{bmatrix} \\
&= \left(\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{\sigma_u^2}{\Delta} \begin{bmatrix} \mathbb{V}(\overline{x}) & -\mathbb{C}(x,\overline{x}) \\ -\mathbb{C}(x,\overline{x}) & \mathbb{V}(x) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & h_0 \end{bmatrix} \right) \begin{bmatrix} \gamma \\ \delta \end{bmatrix} \\
&= \begin{bmatrix} 1 - \frac{\mathbb{V}(\overline{x})}{\Delta}\sigma_u^2 & \frac{h_0 \mathbb{C}(x,\overline{x})}{\Delta}\sigma_u^2 \\ \frac{\mathbb{C}(x,\overline{x})}{\Delta}\sigma_u^2 & 1 - \frac{h_0 \mathbb{V}(x)}{\Delta}\sigma_u^2 \end{bmatrix} \begin{bmatrix} \gamma \\ \delta \end{bmatrix}.
\end{aligned}$$

B.3 Proof of Lemma 1

By Assumption 2, $\mathbb{E}(x) = \mathbb{E}(\overline{x}) = \mathbb{E}((\mathbf{G}^2\mathbf{x})_{si}) = \mu_x$, and $\mathbb{V}(x) = \sigma_x^2$. For the covariance of the own characteristic with the average characteristics of peers, we have that

$$\begin{aligned}
\mathbb{C}(x,\overline{x}) &= \mathbb{E}(x_{si}\overline{x}_{si}) - \mathbb{E}(x_{si})\mathbb{E}(\overline{x}_{si}) \\
&= N_0^{-1}\left(\sum_{i_0} \mathbb{E}(x_{i_0}(\mathbf{G}_0\mathbf{x})_{i_0})\right) - \mu_x^2 \\
&= N_0^{-1}\left(\sum_{i_0,j_0} g_{i_0,j_0}\mathbb{E}(x_{i_0}x_{j_0})\right) - \mu_x^2 \\
&= N_0^{-1}\left(\sum_{i_0,j_0} g_{i_0,j_0}\left(\mathbb{C}(x_{i_0},x_{j_0}) + \mu_x^2\right)\right) - \mu_x^2 \\
&= \rho_1\sigma_x^2.
\end{aligned}$$

For the variance of average characteristics of peers, we have that

$$\begin{aligned}
\mathbb{V}(\overline{x}) &= \mathbb{E}(\overline{x}_{si}^2) - \mathbb{E}(\overline{x}_{si})^2 \\
&= N_0^{-1}\sum_{i_0} \mathbb{E}((\mathbf{G}_0\mathbf{x})_{i_0}^2) - \mu_x^2 \\
&= N_0^{-1}\sum_{i_0,j_0,k_0} g_{i_0,j_0}g_{i_0,k_0}\mathbb{E}(x_{j_0}x_{k_0}) - \mu_x^2 \\
&= N_0^{-1}\sum_{i_0,j_0,k_0} g_{k_0,i_0}g_{k_0,j_0}\left[\mathbb{C}(x_{i_0},x_{j_0}) + \mu_x^2\right] - \mu_x^2 \\
&= \sum_{i_0,j_0} h(i_0,j_0)\mathbb{C}(x_{i_0},x_{j_0}),
\end{aligned}$$

where $h(i_0,j_0) = N_0^{-1}\sum_{k_0} g_{k_0 i_0}g_{k_0 j_0}$ and $\sum_{i_0,j_0} h(i_0,j_0) = 1$. Note that $h(i_0,j_0) = 0$ if individuals at positions i_0 and j_0 do not have a common friend. Therefore, i_0 and j_0

can only be at distances zero, one, or two. As such, we can write

$$\begin{aligned}
\mathbb{V}(\overline{x}) &= \sum_{i_0,j_0} h(i_0,j_0)\mathbb{C}_s(x_{i_0},x_{j_0}) \\
&= \sum_{i_0,j_0:d(i_0,j_0)=0} h(i_0,j_0)\sigma_x^2 + \sum_{i_0,j_0:d(i_0,j_0)=1} h(i_0,j_0)\kappa_1 + \sum_{i_0,j_0:d(i_0,j_0)=2} h(i_0,j_0)\kappa_2 \\
&= N_0^{-1}[\text{trace}(\mathbf{G}_0^\intercal\mathbf{G}_0) + \text{trace}(\mathbf{G}_0^\intercal\mathbf{F}_0\mathbf{G}_0)\rho_1 + (1-\text{trace}(\mathbf{G}_0^\intercal(\mathbf{I}+\mathbf{F}_0)\mathbf{G}_0))\rho_2]\sigma_x^2 \\
&= (h_0 + h_1\rho_1 + h_2\rho_2)\sigma_x^2.
\end{aligned}$$

B.4 Proof of Proposition 3

From the proof of Proposition 1, we know that for a fixed social network

$$\text{plim}\,\widehat{\delta}^{OLS} = \frac{\mathbb{C}(x,\overline{x})\sigma_u^2}{\mathbb{V}(x)\mathbb{V}(\overline{x}) - \mathbb{C}(x,\overline{x})^2 + h_0\mathbb{V}(x)\sigma_u^2 + \mathbb{V}(\overline{x})\sigma_u^2 + h_0\sigma_u^4}\gamma.$$

Substituting the results from Lemma 1 in this equation immediately gives

$$\begin{aligned}
\widehat{\delta}_p^{OLS} &= \frac{\rho_1\phi}{(1+\phi)(h_0+h_1\rho_1+h_2\rho_2) - \rho_1^2 + h_0\phi + h_0\phi^2}\gamma \\
&= \frac{\rho_1\phi}{(1+\phi)(h_0+h_1\rho_1+(1-h_0-h_1)\rho_2) - \rho_1^2 + h_0\phi + h_0\phi^2}\gamma,
\end{aligned}$$

where $\phi = \frac{\sigma_u^2}{\sigma_x^2}$ is the noise-to-signal ratio.

Let $D = (1+\phi)(h_0+h_1\rho_1+h_2\rho_2) - \rho_1^2 + h_0\phi + h_0\phi^2 \geq 0$ denote the denominator. For claim (i), we have that

$$\frac{\partial \frac{\rho_1\phi}{D}}{\partial\rho_1} = \frac{\phi(D-(1+\phi)h_{s1}\rho_1 + 2\rho_1^2)}{D^2},$$

by the quotient rule. The sign of this expression depends on the sign of

$$D - (1+\phi)h_1\rho_1 + 2\rho_1^2 = (1+\phi)(h_0 + \phi h_0 + (1-h_0-h_1)\rho_2) + \rho_1^2,$$

which is certainly positive if

$$\rho_2 \geq -\frac{(1+\phi)h_0}{1-h_0-h_1}.$$

For claim (ii), we have that

$$\frac{\partial \frac{\rho_1 \phi}{D}}{\partial \rho_2} = \frac{-\phi(1+\phi)(1-h_0-h_1)\rho_1}{D^2},$$
$$\text{sign}\left(\frac{\partial \frac{\rho_1 \phi}{D}}{\partial \rho_2}\right) = -\text{sign}(\rho_1) \leq 0,$$

where the second equality follows from the fact that for networks $(1-h_0-h_1) > 0$.
For claim (iii), we have that

$$\frac{\partial \frac{\rho_1 \phi}{D}}{\partial h_0} = \frac{-\phi((1+\phi)(1-\rho_2)+\phi+\phi^2)\rho_1}{D^2},$$
$$\text{sign}\left(\frac{\partial \frac{\rho_1 \phi}{D}}{\partial h_0}\right) = -\text{sign}(\rho_1) \leq 0.$$

Finally, for claim (iv), we have that

$$\frac{\partial \frac{\rho_1 \phi}{D}}{\partial h_1} = \frac{-\phi(1+\phi)(\rho_1-\rho_2)\rho_1}{D^2},$$
$$\text{sign}\left(\frac{\partial \frac{\rho_1 \phi}{D}}{\partial h_1}\right) = -\text{sign}(\rho_1-\rho_2)\text{sign}(\rho_1) \leq 0.$$

B.5 Proof of Corollary 2

The proof is analogous to that of Proposition 3 and is therefore omitted.

B.6 Proof of Proposition 4

After applying the differencing matrix $\mathbf{W}_w$ for $w = \{g, l\}$ to $\mathbf{x}$ and $\overline{\mathbf{x}}$, the expressions for the variance-covariance matrices of the regressors and the measurement error become

$$\mathbf{S}_w = \begin{bmatrix} \mathbb{V}_w(x_{si}) & \mathbb{C}_w(x_{si}, \overline{x}_{si}) \\ \mathbb{C}_w(x_{si}, \overline{x}_{si}) & \mathbb{V}_w(\overline{x}_{si}) \end{bmatrix}, \quad \mathbf{\Sigma}_w = \begin{bmatrix} \mathbb{V}_w(u_{si}) & \mathbb{C}_w(u_{si}, \overline{u}_{si}) \\ \mathbb{C}_w(u_{si}, \overline{u}_{si}) & \mathbb{V}_w(\overline{u}_{si}) \end{bmatrix}.$$

Straightforward calculations give the following expression for the asymptotic bias when $\delta = 0$,

$$\text{plim } \widehat{\delta}_w^{OLS} = \det(\mathbf{S}_w + \mathbf{\Sigma}_w)^{-1}[\mathbb{V}_w(u_{si})\mathbb{C}_w(x_{si}, \overline{x}_{si}) - \mathbb{C}_w(u_{si}, \overline{u}_{si})\mathbb{V}_w(x_{si})]\gamma, \tag{14}$$

which is strictly positive if $\mathbb{C}_w(x_{si}, \overline{x}_{si}) > 0$ and $\mathbb{C}_w(u_{si}, \overline{u}_{si}) < 0$ for $w = \{g, l\}$.

B.6.1 Global transformation

We first consider the global transformation. Let $\mathbf{Q} = \mathbf{I}_S \otimes \mathbf{J}_{N_0}$ such that $\mathbf{W}_g = (\mathbf{I}_N - N_0^{-1}\mathbf{Q})$. Note that $\mathbf{Q}$ is symmetric and that $\mathbf{Q}^2 = N_0\mathbf{Q}$. For the variance-covariance matrix $\boldsymbol{\Sigma}_g$, we therefore find that

$$\boldsymbol{\Sigma}_g = \boldsymbol{\Sigma} - N_0^{-1}\begin{bmatrix} \mathbb{C}(u_{si}, (\mathbf{Qu})_{si}) & \mathbb{C}(u_{si}, (\mathbf{Q\overline{u}})_{si}) \\ \mathbb{C}(u_{si}, (\mathbf{Q\overline{u}})_{si}) & \mathbb{C}(\overline{u}_{si}, (\mathbf{Q\overline{u}})_{si}) \end{bmatrix}.$$

We can expand the covariances $\mathbb{C}(u_{si}, (\mathbf{Qu})_{si})$, $\mathbb{C}(u_{si}, (\mathbf{Q\overline{u}})_{si})$, and $\mathbb{C}(\overline{u}_{si}, (\mathbf{Q\overline{u}})_{si})$ as

$$\begin{aligned}
\mathbb{C}(u_{si}, (\mathbf{Qu})_{si}) &= N_0^{-1}\sum_{i_0}\mathbb{E}(u_{i_0}(\mathbf{Qu})_{i_0}) \\
&= N_0^{-1}\sum_{i_0,j_0}\mathbb{E}\left(u_{i_0}u_{j_0}\right) \\
&= N_0^{-1}\sum_{i_0}\mathbb{E}\left(u_{i_0}^2\right) \\
&= N_0 m_{10}\sigma_u^2 \\
&= \sigma_u^2,
\end{aligned}$$

$$\begin{aligned}
\mathbb{C}(u_{si}, (\mathbf{Q\overline{u}})_{si}) &= N_0^{-1}\sum_{i_0}\mathbb{E}(u_{i_0}(\mathbf{Q}\overline{u})_{i_0}) \\
&= N_0^{-1}\sum_{i_0,j_0,k_0} g_{j_0,k_0}\mathbb{E}\left(u_{i_0}u_{k_0}\right) \\
&= N_0^{-1}\sum_{i_0,j_0} g_{j_0,i_0}\mathbb{E}(u_{i_0}^2) \\
&= N_0 m_{20}\sigma_u^2 \\
&= \sigma_u^2,
\end{aligned}$$

$$\begin{aligned}
\mathbb{C}(\overline{u}_{si}, (\mathbf{Q\overline{u}})_{si}) &= N_0^{-1}\sum_{i_0}\mathbb{E}(\overline{u}_{i_0}(\mathbf{Q\overline{u}})_{i_0}) \\
&= N_0^{-1}\sum_{i_0,j_0,k_0,l_0} g_{i_0,j_0}g_{k_0,l_0}\mathbb{E}\left(u_{j_0}u_{l_0}\right) \\
&= N_0^{-1}\sum_{i_0,j_0,k_0} g_{i_0,j_0}g_{k_0,j_0}\mathbb{E}\left(u_{j_0}^2\right) \\
&= N_0^{-1}(\boldsymbol{\iota}^\intercal\mathbf{G}\mathbf{G}^\intercal\boldsymbol{\iota})\sigma_u^2 \\
&= N_0 m_{30}\sigma_u^2.
\end{aligned}$$

Putting things together, we have that

$$\boldsymbol{\Sigma}_g = \sigma_u^2\begin{bmatrix} 1 - m_{10} & -m_{20} \\ -m_{20} & h_0 - m_{30} \end{bmatrix}.$$

Since $\mathbb{C}_g(u_{si}, \overline{u}_{si}) = -N_0^{-1} < 0$, we have that the global transformation entails an expansion bias if $\mathbb{C}_g(x_{si}, \overline{x}_{si}) > 0$.

Moreover, as

$$\text{plim } \widehat{\delta}_g^{OLS} = \frac{[(1-m_{10})\mathbb{C}_g(x_{si}, \overline{x}_{si}) + m_{20}\mathbb{V}_g(x_{si})]\sigma_u^2}{[(1-m_{10})(h_0 - m_{30}) - m_{20}^2]\sigma_u^4 + O(\sigma_u^2)}\gamma,$$

it follows directly from L'Hôpital's rule that $\lim_{\sigma_u^2 \to \infty} \text{plim } \widehat{\delta}_g^{OLS} = 0$.[18]

The approximation of the expansion bias for $\sigma_u^2 \approx 0$ follows directly from the first-order Taylor linearization of Expression (14) around $\sigma_u^2 = 0$.

B.6.2 Local transformation

We now consider the local transformation for which $\mathbf{W}_l = (\mathbf{I}_N - \mathbf{G})$. For the variance-covariance matrix $\boldsymbol{\Sigma}_l$, we have that

$$\begin{aligned}\boldsymbol{\Sigma}_l &= \boldsymbol{\Sigma} \\ &+ \begin{bmatrix} \mathbb{V}(\overline{u}_{si}) & -\mathbb{C}(u_{si}, (\mathbf{G}^2\mathbf{u})_{si}) + \mathbb{C}(\overline{u}_{si}, (\mathbf{G}^2\mathbf{u})_{si}) - \mathbb{V}(\overline{u}_{si}) \\ -\mathbb{C}(u_{si}, (\mathbf{G}^2\mathbf{u})_{si}) + \mathbb{C}(\overline{u}_{si}, (\mathbf{G}^2\mathbf{u})_{si}) - \mathbb{V}(\overline{u}_{si}) & \mathbb{V}((\mathbf{G}^2\mathbf{u})_{si}) - 2\mathbb{C}(\overline{u}_{si}, (\mathbf{G}^2\mathbf{u})_{si}) \end{bmatrix}.\end{aligned}$$

We can expand the covariances $\mathbb{C}(u_{si}, (\mathbf{G}^2\mathbf{u})_{si})$ and $\mathbb{C}(\overline{u}_{si}, (\mathbf{G}^2\mathbf{u})_{si})$, and the variance $\mathbb{V}((\mathbf{G}^2\mathbf{u})_{si})$ as

$$\begin{aligned}\mathbb{C}(u_{si}, (\mathbf{G}^2\mathbf{u})_{si}) &= N_0^{-1}\sum_{i_0}\mathbb{E}(u_{i_0}(\mathbf{G}^2\mathbf{u})_{i_0}) \\ &= N_0^{-1}\sum_{i_0,j_0,k_0} g_{i_0j_0}g_{j_0k_0}\mathbb{E}(u_{i_0}u_{k_0}) \\ &= N_0^{-1}\sum_{i_0,j_0} g_{i_0j_0}g_{j_0i_0}\mathbb{E}(u_{i_0}^2) \\ &= N_0^{-1}\text{trace}(\mathbf{G}^2)\sigma_u^2 \\ &= n_{10}\sigma_u^2,\end{aligned}$$

$$\begin{aligned}\mathbb{C}(\overline{u}_{si}, (\mathbf{G}^2\mathbf{u})_{si}) &= N_0^{-1}\sum_{i_0}\mathbb{E}(\overline{u}_{i_0}(\mathbf{G}^2\mathbf{u})_{i_0}) \\ &= N_0^{-1}\sum_{i_0,j_0,k_0,l_0} g_{i_0j_0}g_{i_0k_0}g_{k_0l_0}\mathbb{E}(u_{j_0}u_{l_0}) \\ &= N_0^{-1}\sum_{i_0,j_0,k_0} g_{i_0j_0}g_{i_0k_0}g_{k_0j_0}\mathbb{E}(u_{j_0}^2) \\ &= N_0^{-1}\text{trace}(\mathbf{G}^\intercal\mathbf{G}^2)\sigma_u^2 \\ &= n_{20}\sigma_u^2,\end{aligned}$$

[18] We let $O(\cdot)$ denote Landau's Big O.

and

$$\begin{aligned}
\mathbb{V}((\mathbf{G}^2\mathbf{u})_{si}) &= N_0^{-1}\sum_{i_0}\mathbb{E}((\mathbf{G}^2\mathbf{u})_{i_0}^2) \\
&= N_0^{-1}\sum_{i_0,j_0,k_0,l_0,m_0} g_{i_0j_0}g_{j_0k_0}g_{i_0l_0}g_{l_0m_0}\mathbb{E}(x_{k_0}x_{m_0}) \\
&= N_0^{-1}\sum_{i_0,j_0,k_0,l_0} g_{i_0j_0}g_{j_0k_0}g_{i_0l_0}g_{l_0k_0}\mathbb{E}(x_{k_0}^2) \\
&= N_0^{-1}\text{trace}((\mathbf{G}^2)^\intercal\mathbf{G}^2)\sigma_u^2 \\
&= n_{30}\sigma_u^2.
\end{aligned}$$

Putting thing together, we have that

$$\mathbf{\Sigma}_l = \sigma_u^2\begin{bmatrix} 1+h_0 & -n_{10}+n_{20}-h_0 \\ -n_{10}+n_{20}-h_0 & h_0+n_{30}-2n_{20}\end{bmatrix}.$$

If $\mathbb{C}_l(u_{si},\overline{u}_{si}) = N_0^{-1}(-\text{trace}(\mathbf{G}^2)+\text{trace}(\mathbf{G}^\intercal\mathbf{G}^2)-\text{trace}(\mathbf{G}^\intercal\mathbf{G})) < 0$, we have that the local transformation entails an expansion bias if $\mathbb{C}_l(x_{si},\overline{x}_{si}) > 0$. This is true for every network because $\text{trace}(\mathbf{G}^2) \geq \text{trace}(\mathbf{G}^\intercal\mathbf{G}^2)$ always holds.[19]

Moreover, as

$$\text{plim }\widehat{\delta}_l^{OLS} = \frac{[(1+h_0)\mathbb{C}_l(x_{si},\overline{x}_{si}) - (-n_{10}+n_{20}-h_0)\mathbb{V}_l(x_{si})]\sigma_u^2}{[(1+h_0)(h_0+n_{30}-2n_{20}) - (-n_{10}+n_{20}-h_0)^2]\sigma_u^4 + O(\sigma_u^2)}\gamma,$$

it follows by L'Hôpital's rule that $\lim_{\sigma_u^2\to\infty}\text{plim }\widehat{\delta}_l^{OLS} = 0$.

The approximation of the expansion bias for $\sigma_u^2 \approx 0$ follows directly from the first-order Taylor linearization of Expression (14) around $\sigma_u^2 = 0$.

[19]This inequality relies on three well-known results from linear algebra. Let $\lambda(\mathbf{A}) = \{\lambda_i\}_{i=1,2,\ldots,N_0}$ denote the set of eigenvalues of the $N_0 \times N_0$ matrix $\mathbf{A}$. Then (i) $\text{trace}(\mathbf{A}) = \sum_i \lambda_i$; (ii) $\lambda(\mathbf{A}) = \lambda(\mathbf{A}^\intercal)$; and (iii) $\lambda(\mathbf{A}^k) = \lambda(\mathbf{A}^\intercal)^k$, where the power on the set of eigenvalues is element-wise. We know by the Perron-Frobenius theorem that all eigenvalues of $\mathbf{G}$ are within the unit circle, such that the eigenvalues of $\mathbf{G}^2$ are all positive. Since also the eigenvalues of $\mathbf{G}^\intercal$ are within the unit circle, and by using the von Neumann trace inequality, we therefore immediately have that $\text{trace}(\mathbf{G}^2) \geq \text{trace}(\mathbf{G}^\intercal\mathbf{G}^2)$.

B.7 Proof of Proposition 5

Using the results from Lemma A.3, and the fact that

$$
\begin{aligned}
\sum_{t=0}^{2} h_t \rho &= \rho + h_0 - h_0 \rho, \\
\sum_{t=0}^{T} n_{kt} \rho &= \rho + m_{k0} - m_{k0} \rho, \quad \forall k \in \{1,2,3\}, \\
\sum_{t=0}^{k+1} n_{kt} \rho &= \rho + n_{k0} - n_{k0} \rho, \quad \forall k \in \{1,2,3\},
\end{aligned}
$$

we have that

$$
\begin{aligned}
\mathbf{S}_g &= (1-\rho)\sigma_x^2 \begin{bmatrix} 1 - m_{10} & -m_{20} \\ -m_{20} & h_{s0} - m_{30}, \end{bmatrix}, \\
\mathbf{S}_l &= (1-\rho)\sigma_x^2 \begin{bmatrix} 1 + h_0 & -n_{10} + n_{20} - h_0 \\ -n_{10} + n_{20} - h_0 & h_0 + n_{30} - 2n_{20} \end{bmatrix},
\end{aligned}
$$

such that it holds that

$$
(\mathbf{S}_g + \mathbf{\Sigma}_g)^{-1} \mathbf{S}_g = (\mathbf{S}_l + \mathbf{\Sigma}_l)^{-1} \mathbf{S}_l = \frac{(1-\rho)\sigma_x^2}{(1-\rho)\sigma_x^2 + \sigma_u^2} \mathbf{I}.
$$

C Proofs for Section 4

C.1 Proof of Theorem 1

Part 1: Mean restrictions Let $\mathbf{M} = [\boldsymbol{\iota}, \widetilde{\mathbf{m}}, \mathbf{G}_0\widetilde{\mathbf{m}}]$ be the $N_0 \times 3$ matrix that collects the mean restrictions. The parameter vector (α, γ, δ) is identified whenever $\text{rank}(\mathbf{M}) = 3$.

We first show that $\text{rank}(\mathbf{M}) = 1$ if and only if $\widetilde{\mathbf{m}} = c\boldsymbol{\iota}$ for some $c \in \mathbb{R}$. It holds that $\text{rank}(\mathbf{M}) = 1$ if and only if there exist $c_1, c_2 \in \mathbb{R}$ such that $\widetilde{\mathbf{m}} = c_1\boldsymbol{\iota}$ and $\mathbf{G}_0\widetilde{\mathbf{m}} = c_2\boldsymbol{\iota}$. And if $\widetilde{\mathbf{m}} = c_1\boldsymbol{\iota}$, then $\mathbf{G}_0\widetilde{\mathbf{m}} = c_1\mathbf{G}_0\boldsymbol{\iota} = c_1\boldsymbol{\iota}$.

We now show that $\text{rank}(\mathbf{M}) = 2$ if and only if $\widetilde{\mathbf{m}} + c\boldsymbol{\iota}$ is an eigenvector of $\mathbf{G}_0$ for some $c \in \mathbb{R}$. It holds that $\text{rank}(\mathbf{M}) = 2$ if and only if there exist $c_1, c_2 \in \mathbb{R}$ such that $\mathbf{G}_0\widetilde{\mathbf{m}} = c_1\boldsymbol{\iota} + c_2\widetilde{\mathbf{m}}$. ($\Longrightarrow$) We first establish that $c_2 \neq 1$, by proceeding towards a contradiction. Suppose $c_2 = 1$, we have that $(\mathbf{I} - \mathbf{G}_0)\widetilde{\mathbf{m}} = -c_1\boldsymbol{\iota}$ such that $(\mathbf{I} - \mathbf{G}_0)^2\widetilde{\mathbf{m}} = 0$ should hold. However, since the null space of $(\mathbf{I} - \mathbf{G}_0)^2$ is 1-dimensional and spanned by the vector $\boldsymbol{\iota}$, it should also hold that $\widetilde{\mathbf{m}} = c_3\boldsymbol{\iota}$ for some $c_3 \in \mathbb{R}$.[20] But then $\mathbf{M}$ should be of rank 1, which is a contradiction. Therefore $c_2 \neq 1$, and it is straightforward to verify that $\widetilde{\mathbf{m}} + \frac{c_1}{c_2-1}\boldsymbol{\iota}$ is an eigenvector of $\mathbf{G}_0$ for the eigenvalue c_2:

$$\begin{aligned}
\mathbf{G}_0\left(\widetilde{\mathbf{m}} + \frac{c_1}{c_2 - 1}\boldsymbol{\iota}\right) &= c_1\boldsymbol{\iota} + c_2\widetilde{\mathbf{m}} + \frac{c_1}{c_2 - 1}\boldsymbol{\iota} \\
&= c_2\left(\widetilde{\mathbf{m}} + \frac{c_1}{c_2 - 1}\boldsymbol{\iota}\right).
\end{aligned}$$

($\Longleftarrow$) Let $\widetilde{\mathbf{m}} + c\boldsymbol{\iota}$ be an eigenvector of $\mathbf{G}_0$ with eigenvalue $c_3 \in \mathbb{R}$. Then, it holds that $\mathbf{G}_0(\widetilde{\mathbf{m}} + c\boldsymbol{\iota}) = c_3(\widetilde{\mathbf{m}} + c\boldsymbol{\iota})$, or alternatively that $\mathbf{G}_0\widetilde{\mathbf{m}} = c_3\widetilde{\mathbf{m}} + (c_3 - 1)c\boldsymbol{\iota}$.

The conditions stated in the theorem therefore ensure that $\text{rank}(\mathbf{M}) \neq \{1, 2\}$ such that $\text{rank}(\mathbf{M}) = 3$, which was required. $\square$

Part 2: Covariance restrictions Rewriting the covariance restrictions in Expression (9) in terms of matrices, we have that

$$\mathbf{C} = \gamma(\widetilde{\mathbf{V}} - \sigma_u^2\mathbf{I}) + \delta\mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^2\mathbf{I}). \tag{15}$$

Note that $\gamma, \delta = 0$ implies that $\mathbf{C} = \mathbf{0}$ for any value of σ_u^2, so we will assume that at least one of both parameters is non-zero. The model is identified from these restric-

[20] Since $\mathbf{G}_0$ is row-stochastic and irreducible, we know from the Perron-Frobenius theorem that it has a spectral radius of 1 with algebraic multiplicity one. This implies that $\dim(\ker((\mathbf{I} - \mathbf{G}_0)^k)) = 1$ for every $k \in \mathbb{N}_{>0}$, as the algebraic multiplicity of an eigenvalue λ is defined as the dimension of the generalized eigenspace associated to that eigenvalue: i.e., the vector space of all generalized λ-eigenvectors v_λ for which $(\mathbf{G}_0 - \lambda\mathbf{I})^n v_\lambda = 0$.

tions when

$$\gamma(\widetilde{\mathbf{V}}-\sigma_u^2\mathbf{I})+\delta\mathbf{G}_0(\widetilde{\mathbf{V}}-\sigma_u^2\mathbf{I}) = \gamma'(\widetilde{\mathbf{V}}-\sigma_u^{2\prime}\mathbf{I})+\delta'\mathbf{G}_0(\widetilde{\mathbf{V}}-\sigma_u^{2\prime}\mathbf{I}) \implies (\gamma,\delta,\sigma_u^2) = (\gamma',\delta',\sigma_u^{2\prime}).$$

The left-hand side of this expression can be rewritten as

$$(\gamma'\sigma_u^{2\prime} - \gamma\sigma_u^2)\mathbf{I} + (\delta'\sigma_u^{2\prime} - \delta\sigma_u^2)\mathbf{G}_0 + (\gamma-\gamma')\widetilde{\mathbf{V}} + (\delta-\delta')\mathbf{G}_0\widetilde{\mathbf{V}} = \mathbf{0}.$$

($\Longrightarrow$) Suppose $\mathbf{I}$, $\mathbf{G}_0$, $\widetilde{\mathbf{V}}$, and $\mathbf{G}_0\widetilde{\mathbf{V}}$ are linearly independent. By independence, $\gamma = \gamma'$ and $\delta = \delta'$. In addition, $\gamma'\sigma_u^{2\prime} = \gamma\sigma_u^2$ and $\delta'\sigma_u^{2\prime} - \delta\sigma_u^2$, such that $\sigma_u^2 = \sigma_u^{2\prime}$. ($\Longleftarrow$) Suppose $\mathbf{I}$, $\mathbf{G}_0$, $\widetilde{\mathbf{V}}$, and $\mathbf{G}_0\widetilde{\mathbf{V}}$ are not linearly independent. This implies that there exist constants $c_1, c_2, c_3, c_4 \in \mathbb{R}$, not all equal to zero, such that

$$c_1\mathbf{I} + c_2\mathbf{G}_0 + c_3\widetilde{\mathbf{V}} + c_4\mathbf{G}_0\widetilde{\mathbf{V}} = \mathbf{0}.$$

Identification fails if there exist $(\gamma,\delta,\sigma_u^2) \neq (\gamma',\delta',\sigma_u^{2\prime})$ with ($\gamma \neq 0$ or $\delta \neq 0$) and ($\gamma' \neq 0$ or $\delta' \neq 0$) such that

$$\begin{aligned}
c_1 &= \gamma'\sigma_u^{2\prime} - \gamma\sigma_u^2,\\
c_2 &= \delta'\sigma_u^{2\prime} - \delta\sigma_u^2,\\
c_3 &= \gamma - \gamma',\\
c_4 &= \delta - \delta',
\end{aligned}$$

holds. Direct substitution shows that this holds for $(\gamma,\delta,\sigma_u^2) = (c_1 + 2c_3, c_2 + 2c_4, 1)$ and $(\gamma',\delta',\sigma_u^{2\prime}) = (c_1 + c_3, c_2 + c_4, 2)$.

C.2 Proof of Proposition 6

We first proof the following auxiliary lemma.

Lemma C.1. *For $|\phi^{-1}| \leq 1$, $ker((\mathbf{G}_0 - \mathbf{I})(\mathbf{G}_0 - \phi^{-1}\mathbf{I})) = ker(\mathbf{G}_0 - \mathbf{I}) \cup ker(\mathbf{G}_0 - \phi^{-1}\mathbf{I})$.*

Proof. Let $\mathbf{A} := (\mathbf{G}_0 - \mathbf{I})$ and $\mathbf{B} := (\mathbf{G}_0 - \phi^{-1}\mathbf{I})$. If $\phi = 1$, we know from Footnote 20 that $\ker(\mathbf{A}^2) = \boldsymbol{\iota}$, so we therefore focus on the more interesting case where $\phi \neq 1$.

Since $\mathbf{A}$ and $\mathbf{B}$ commute, we have that $\ker(\mathbf{AB}) \supseteq \ker(\mathbf{A}) \cup \ker(\mathbf{B})$. As eigenvectors belonging to different eigenvalues are linearly independent, we know that $\text{rank}(\ker(\mathbf{A}) \cup \ker(\mathbf{B})) = 1 + k$, where k is the algebraic multiplicity of $\mathbf{G}$ for eigenvalue ϕ^{-1}. Therefore $\text{rank}(\ker(\mathbf{AB})) \geq 1 + k$, such that $\text{rank}(\mathbf{AB}) \leq N_0 - k - 1$.

From Sylvester's rank inequality, we also know that

$$
\begin{aligned}
\operatorname{rank}(\mathbf{AB}) &\geq \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(\mathbf{B}) - N_0 \\
&\geq (N_0 - 1) + (N_0 - k) - N_0 \\
&\geq N_0 - k - 1.
\end{aligned}
$$

Combining these two inequalities, we have that $\operatorname{rank}(\mathbf{AB}) = N_0 - k - 1$. This implies that $\ker(\mathbf{AB}) = \ker(\mathbf{A}) \cup \ker(\mathbf{B})$, which was required. $\square$

From the covariance restrictions in Expression (15), it follows that the parameter vector (γ, δ) is not identified if and only if $\operatorname{rank}(\operatorname{vec}(\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I}), \operatorname{vec}(\mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I}))) = 0, 1$. The latter will only be of rank 0 when $\widetilde{\mathbf{V}} = \sigma_u^2 \mathbf{I}$, which is a degenerate event. We therefore focus on the case where the rank is 1, or equivalently $(\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I}) = \phi \mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I})$ for some $\phi \in \mathbb{R}$. From this, we have that

$$
\begin{aligned}
(\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I}) = \phi \mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I}) &\implies (\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I})\boldsymbol{\iota} = \phi \mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^2 \mathbf{I})\boldsymbol{\iota} \\
&\iff \widetilde{\mathbf{V}}\boldsymbol{\iota} - \sigma_u^2 \boldsymbol{\iota} = \phi \mathbf{G}_0 \widetilde{\mathbf{V}}\boldsymbol{\iota} - \phi \sigma_u^2 \boldsymbol{\iota} \\
&\implies \mathbf{W}(\widetilde{\mathbf{V}}\boldsymbol{\iota} - \sigma_u^2 \boldsymbol{\iota}) = \mathbf{W}(\phi \mathbf{G}_0 \widetilde{\mathbf{V}}\boldsymbol{\iota} - \phi \sigma_u^2 \boldsymbol{\iota}) \\
&\implies \mathbf{W}\widetilde{\mathbf{V}}\boldsymbol{\iota} = \phi \mathbf{W}\mathbf{G}_0 \widetilde{\mathbf{V}}\boldsymbol{\iota} \\
&\iff \mathbf{W}(\mathbf{I} - \phi \mathbf{G}_0)\widetilde{\mathbf{V}}\boldsymbol{\iota} = \mathbf{0},
\end{aligned}
$$

where $\mathbf{W}$ is a differencing matrix such that $\mathbf{W}\boldsymbol{\iota} = \mathbf{0}$. For analytical convenience, take local differences $\mathbf{W}_l = (\mathbf{G}_0 - \mathbf{I})$, which yields the condition

$$
(\mathbf{G}_0 - \mathbf{I})(\mathbf{G}_0 - \phi^{-1}\mathbf{I})\widetilde{\mathbf{V}}\boldsymbol{\iota} = \mathbf{0}.
$$

By using Lemma C.1, we have that (γ, δ) is identified if $\widetilde{\mathbf{V}}\boldsymbol{\iota}$ is not an eigenvector of $\mathbf{G}_0$.

Since $\mathbf{G}_0$ does not take the form of a group, it has at least three distinct eigenvalues, which implies that $N_0 - 2$ is an upper-bound for the algebraic multiplicity of its eigenvalues. By Lemma C.1, we therefore know that the kernel of $(\mathbf{G}_0 - \mathbf{I})(\mathbf{G}_0 - \phi^{-1}\mathbf{I})$ is at most a vector space of dimension $N_0 - 2 + 1 = N_0 - 1$. Since this is a lower-dimensional vector space, non-identification is non-generic and identification is generic.

It follows that the dimensionality of the nonidentified set is related to the size of the largest algebraic multiplicity of $\mathbf{G}_0$. In graph theory, it is well known that an adjacency matrix's algebraic multiplicities tends to increase with the network's isomorphisms (Biggs, 1993).[21] So informally, the more "asymmetric" the network,

[21] At one extreme we have the group, for which the largest algebraic multiplicity is $N_0 - 1$, and at the other extreme we have the random graph, for which the largest algebraic multiplicity is 1 almost

the more "likely" the parameters of interest are identified.

C.3 Proof of Proposition 7

Let $\mathbf{G}_0'$ and $\widetilde{\mathbf{V}}'$ be $N_0' \times N_0'$ matrices where $(\mathbf{G}_0')_{ii} = 0$, $(\mathbf{G}_0')_{ij} = \frac{1}{d'}$, $(\widetilde{\mathbf{V}}')_{ii} = \sigma_x^{2\prime} + \sigma_u^2$, and $(\widetilde{\mathbf{V}}')_{ij} = \rho_1' \sigma_x^{2\prime}$ for $j \neq i$. Define $\mathbf{G}_0''$ and $\widetilde{\mathbf{V}}''$ similarly, mutatis mutandis. The linear independence condition from Theorem 1 implies that there do not exist constants $c_1, c_2, c_3, c_4 \neq 0$ such that

$$c_1 \begin{bmatrix} \mathbf{I}_{N_0'} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{N_0''} \end{bmatrix} + c_2 \begin{bmatrix} \mathbf{G}_0' & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0'' \end{bmatrix} + c_3 \begin{bmatrix} \widetilde{\mathbf{V}}' & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{V}}'' \end{bmatrix} + c_4 \begin{bmatrix} \mathbf{G}_0'\widetilde{\mathbf{V}}' & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_0''\widetilde{\mathbf{V}}'' \end{bmatrix} = \mathbf{0}.$$

Since there are only four relevant equations to test (one for the diagonal and one for the off-diagonal elements for each group size), this condition reduces to testing whether the following matrix is of full rank:

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & \sigma_x^{2\prime} + \sigma_u^2 & \rho' \sigma_x^{2\prime} \\ 0 & \frac{1}{d'} & \rho' \sigma_x^{2\prime} & \frac{1}{d'}(\sigma_x^{2\prime} + \sigma_u^2 + (d'-1)\rho'\sigma_x^{2\prime}) \\ 1 & 0 & \sigma_x^{2\prime\prime} + \sigma_u^2 & \rho'' \sigma_x^{2\prime\prime} \\ 0 & \frac{1}{d''} & \rho'' \sigma_x^{2\prime\prime} & \frac{1}{d''}(\sigma_x^{2\prime\prime} + \sigma_u^2 + (d''-1)\rho''\sigma_x^{2\prime\prime}) \end{bmatrix}.$$

Visual inspection shows that there are no obvious linear dependencies and therefore this matrix is generically of full rank.[22]

However, it is apparent that only variation in group size is not enough as the first and third row would become identical if $\rho_1' = \rho_1''$ and $\sigma_x^{2\prime} = \sigma_x^{2\prime\prime}$. We therefore study identification for two worst-case scenarios and show that non-identification is also for those a degenerate event.

surely.

[22]This point can be made more formally. Let $v = \{\rho_1', \rho_1'', \sigma_x^{2\prime}, \sigma_x^{2\prime\prime}\} \subseteq \mathscr{V}$, where $\mathscr{V}$ is a connected open subset of $\mathbb{R}_+^4$. Since $\det \mathbf{R}(v)$ is a nonzero analytic function of v, it holds that the set $\{v \in \mathscr{V} \mid \det \mathbf{R}(v) = 0\}$ has Lebesgue measure zero (e.g., see Mityagin, 2020).

Case 1: $\sigma_x^2 = \sigma_x^{2\prime} = \sigma_x^{2\prime\prime}$ First suppose there is variation in group size and homophily, but not in the variance. Using elementary row operations we have that

$$\begin{aligned}
\det \mathbf{R} &= \det \begin{bmatrix} 1 & 0 & \sigma_x^2 + \sigma_u^2 & \rho_1' \sigma_x^2 \\ 0 & \frac{1}{d'} & \rho_1' \sigma_x^2 & \frac{1}{d'}(\sigma_x^2 + \sigma_u^2 + (d'-1)\rho_1'\sigma_x^2) \\ 0 & 0 & 0 & (\rho_1'' - \rho_1')\sigma_x^2 \\ 0 & \frac{1}{d''} & \rho_1''\sigma_x^2 & \frac{1}{d''}(\sigma_x^2 + \sigma_u^2 + (d''-1)\rho_1''\sigma_x^2) \end{bmatrix} \\
&= -(\rho_1'' - \rho_1')\sigma_x^2 \det \begin{bmatrix} \frac{1}{d'} & \rho_1'\sigma_x^2 \\ \frac{1}{d''} & \rho_1''\sigma_x^2 \end{bmatrix} \\
&= -(\rho_1'' - \rho_1')(\rho_1'' d'' - \rho_1' d')\sigma_x^4 \frac{1}{d'd''}.
\end{aligned}$$

This implies that if $\rho_1' \neq \rho_1''$ and $\rho_1' d' \neq \rho_1'' d''$, the parameters of interest are identified.

Case 2: $\rho_1 = \rho_1' = \rho_1''$ Now suppose there is variation in group size and in the variance, but not in homophily. Again using elementary row operations, we have that

$$\begin{aligned}
\det \mathbf{R} &= \det \begin{bmatrix} 1 & 0 & \sigma_x^{2\prime} + \sigma_u^2 & \rho_1 \sigma_x^{2\prime} \\ 0 & \frac{1}{d'} & \rho_1 \sigma_x^{2\prime} & \frac{1}{d'}(\sigma_x^{2\prime} + \sigma_u^2 + (d'-1)\rho_1\sigma_x^{2\prime}) \\ 0 & 0 & \sigma_x^{2\prime\prime} - \sigma_x^{2\prime} & \rho(\sigma_x^{2\prime\prime} - \sigma_x^{2\prime}) \\ 0 & \frac{1}{d''} & \rho_1\sigma_x^{2\prime\prime} & \frac{1}{d''}(\sigma_x^{2\prime\prime} + \sigma_u^2 + (d''-1)\rho_1\sigma_x^{2\prime\prime}) \end{bmatrix} \\
&= \frac{1}{d'} \det \begin{bmatrix} \sigma_x^{2\prime\prime} - \sigma_x^{2\prime} & \rho_1(\sigma_x^{2\prime\prime} - \sigma_x^{2\prime}) \\ \rho_1\sigma_x^{2\prime\prime} & \frac{1}{d''}(\sigma_x^{2\prime\prime} + \sigma_u^2 + (d''-1)\rho_1\sigma_x^{2\prime\prime}) \end{bmatrix} \\
&\quad + \frac{1}{d''} \det \begin{bmatrix} \rho_1\sigma_x^{2\prime} & \frac{1}{d'}(\sigma_x^{2\prime} + \sigma_u^2 + (d'-1)\rho_1\sigma_x^{2\prime}) \\ \sigma_x^{2\prime\prime} - \sigma_x^{2\prime} & \rho_1(\sigma_x^{2\prime\prime} - \sigma_x^{2\prime}) \end{bmatrix} \\
&= (\sigma_x^{2\prime\prime} - \sigma_x^{2\prime})[(1-\rho_1)(\sigma_x^{2\prime\prime} - \sigma_x^{2\prime}) + (\rho_1 - \rho_1^2)(\sigma_x^{2\prime\prime} d'' - \sigma_x^{2\prime} d')]\frac{1}{d'd''}.
\end{aligned}$$

This implies that if $\sigma_x^{2\prime} \neq \sigma_x^{2\prime\prime}$ and $(1-\rho)(\sigma_x^{2\prime\prime} - \sigma_x^{2\prime}) \neq -(\rho - \rho^2)(\sigma_x^{2\prime\prime} d'' - \sigma_x^{2\prime} d')$, the parameters of interst are identified.

C.4 Proof of Theorem 2

Part 1: Mean restrictions The proof of this part is similar to that of Theorem 1. In particular, note that $\mathbf{WM} = [\mathbf{0}, \mathbf{W}\widetilde{\mathbf{m}}, \mathbf{WG}_0\widetilde{\mathbf{m}}]$ is the $N_0 \times 3$ matrix that collects the demeaned mean restrictions. The parameter vector (γ, δ) is identified whenever $\text{rank}(\mathbf{WM}) = 2$.

Part 2: Covariance restrictions Similarly to the proof of Theorem 1, we can rewrite the covariance restrictions as

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

Assuming $\gamma \neq 0$ or $\delta \neq 0$, we have that the model is identified from these moments when

$$\begin{matrix} \gamma\mathbf{W}(\widetilde{\mathbf{V}} - \sigma_u^2\mathbf{I}) + \delta\mathbf{W}\mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^2\mathbf{I}) \\ = \\ \gamma'\mathbf{W}(\widetilde{\mathbf{V}} - \sigma_u^{2\prime}\mathbf{I}) + \delta'\mathbf{W}\mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^{2\prime}\mathbf{I}) \end{matrix} \quad \Longrightarrow \quad (\gamma, \delta, \sigma_u^2) = (\gamma', \delta', \sigma_u^{2\prime}).$$

The left-hand side of this expression can be rewritten as

$$c_1\mathbf{W} + c_2\mathbf{W}\mathbf{G}_0 + c_3\mathbf{W}\widetilde{\mathbf{V}} + c_4\mathbf{W}\mathbf{G}_0\widetilde{\mathbf{V}} = \mathbf{0},$$

where

$$\begin{aligned} c_1 &= \gamma'\sigma_u^{2\prime} - \gamma\sigma_u^2, \\ c_2 &= \delta'\sigma_u^{2\prime} - \delta\sigma_u^2 \\ c_3 &= \gamma - \gamma', \\ c_4 &= \delta - \delta'. \end{aligned} \tag{16}$$

($\Longrightarrow$) Suppose $\mathbf{W}$, $\mathbf{W}\mathbf{G}_0$, $\mathbf{W}\widetilde{\mathbf{V}}$, and $\mathbf{W}\mathbf{G}_0\widetilde{\mathbf{V}}$ are linearly independent. By independence, we have that $(\gamma, \delta, \sigma_u^2) = (\gamma', \delta', \sigma_u^{2\prime})$. ($\Longleftarrow$) Suppose $\mathbf{W}$, $\mathbf{W}\mathbf{G}_0$, $\mathbf{W}\widetilde{\mathbf{V}}$, and $\mathbf{W}\mathbf{G}_0\widetilde{\mathbf{V}}$ are not linearly independent. This implies there exist constants $c_1, c_2, c_3, c_4 \in \mathbb{R}$ such that $c_1\mathbf{W} + c_2\mathbf{W}\mathbf{G}_0 + c_3\mathbf{W}\widetilde{\mathbf{V}} + c_4\mathbf{W}\mathbf{G}_0\widetilde{\mathbf{V}} = \mathbf{0}$. Identification fails if there exist $(\gamma, \delta, \sigma_u^2) \neq (\gamma', \delta', \sigma_u^{2\prime})$ with ($\gamma \neq 0$ or $\delta \neq 0$) and ($\gamma' \neq 0$ or $\delta' \neq 0$) such that Expression (16) holds. Direct substitution shows that this holds for $(\gamma, \delta, \sigma_u^2) = (-c_1 - c_3, c_5 + c_6, 2)$ and $(\gamma', \delta', \sigma_u^{2\prime}) = (-c_1 - 2c_3, 2c_5 + c_6, 1)$.

C.5 Proof of Corollary 3

Part 1: Mean restrictions Identification fails if and only if $\text{rank}((\mathbf{I} - \mathbf{G}_0)\widetilde{\mathbf{m}}, (\mathbf{I} - \mathbf{G}_0)\mathbf{G}_0\widetilde{\mathbf{m}}) = 0, 1$. The latter has rank 0 if and only if $\widetilde{\mathbf{m}}$ is a scalar multiple of $\boldsymbol{\iota}$. It has rank 1 when there exists a $c_1 \in \mathbb{R}$ such that $c_1(\mathbf{I} - \mathbf{G}_0)\widetilde{\mathbf{m}} = (\mathbf{I} - \mathbf{G}_0)\mathbf{G}_0\widetilde{\mathbf{m}}$ or equivalently $\mathbf{G}_0^2\widetilde{\mathbf{m}} = (1 + c_1)\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}}$. This happens if and only if $(\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}})$ is an eigenvector of $\mathbf{G}_0$ associated with eigenvalue 1. ($\Longrightarrow$) If $(\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}})$ is an eigenvector associated with eigenvalue 1, we have that

$$\mathbf{G}_0(\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}}) = \mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}} \Longrightarrow \mathbf{G}_0^2\widetilde{\mathbf{m}} = (1 + c_1)\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}}.$$

($\Longleftarrow$) On the other hand, consider

$$\begin{aligned}\mathbf{G}_0(\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}}) &= \mathbf{G}_0^2\widetilde{\mathbf{m}} - c_1\mathbf{G}_0\widetilde{\mathbf{m}}\\ &= (1 + c)\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}} - c_1\mathbf{G}_0\widetilde{\mathbf{m}}\\ &= \mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}}.\end{aligned}$$

From the Perron-Frobenius theorem, we know that the eigenvector associated with eigenvalue 1 must be a multiple of $\boldsymbol{\iota}$. Therefore, when $\operatorname{rank}((\mathbf{I}-\mathbf{G}_0)\widetilde{\mathbf{m}}, (\mathbf{I}-\mathbf{G}_0)\mathbf{G}_0\widetilde{\mathbf{m}}) = 1$, it should hold that $(\mathbf{G}_0\widetilde{\mathbf{m}} - c_1\widetilde{\mathbf{m}}) = c_2\boldsymbol{\iota}$ for some $c_2 \in \mathbb{R}$. From the first part of the proof of Theorem 1, it is clear that happens if an only if $\widetilde{\mathbf{m}} + c\boldsymbol{\iota}$ is an eigenvector of $\mathbf{G}_0$ for some $c \in \mathbb{R}$.

Part 2: Covariance restrictions Similarly to the proof of Theorem 1, we can rewrite the covariance restrictions as

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

Assuming $\gamma \neq 0$ or $\delta \neq 0$, we have that the model is identified from these restrictions when

$$\begin{matrix}\gamma(\mathbf{I} - \mathbf{G}_0)(\widetilde{\mathbf{V}} - \sigma_u^2\mathbf{I}) + \delta(\mathbf{I} - \mathbf{G}_0)\mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^2\mathbf{I})\\ = \\ \gamma'(\mathbf{I} - \mathbf{G}_0)(\widetilde{\mathbf{V}} - \sigma_u^{2\prime}\mathbf{I}) + \delta'(\mathbf{I} - \mathbf{G}_0)\mathbf{G}_0(\widetilde{\mathbf{V}} - \sigma_u^{2\prime}\mathbf{I})\end{matrix} \quad \Longrightarrow \quad (\gamma, \delta, \sigma_u^2) = (\gamma', \delta', \sigma_u^{2\prime}).$$

The left-hand side of this expression can be rewritten as

$$c_1\mathbf{I} + c_2\mathbf{G}_0 + c_3\widetilde{\mathbf{V}} + c_4\mathbf{G}_0\widetilde{\mathbf{V}} + c_5\mathbf{G}_0^2\widetilde{\mathbf{V}} + c_6\mathbf{G}_0^2 = \mathbf{0},$$

where

$$\begin{aligned}c_1 &= \gamma'\sigma_u^{2\prime} - \gamma\sigma_u^2,\\ c_2 &= (\gamma - \delta)\sigma_u^2 - (\gamma' - \delta')\sigma_u^{2\prime},\\ c_3 &= \gamma - \gamma',\\ c_4 &= -\gamma + \delta + \gamma' - \delta',\\ c_5 &= -\delta + \delta',\\ c_6 &= \delta\sigma_u^2 - \delta'\sigma_u^{2\prime}.\end{aligned} \tag{17}$$

($\Longrightarrow$) Suppose $\mathbf{I}$, $\mathbf{G}_0$, $\mathbf{G}_0^2$, $\widetilde{\mathbf{V}}$, $\mathbf{G}_0\widetilde{\mathbf{V}}$, and $\mathbf{G}_0^2\widetilde{\mathbf{V}}$ are linearly independent. By independence, we have that $(\gamma, \delta, \sigma_u^2) = (\gamma', \delta', \sigma_u^{2\prime})$. ($\Longleftarrow$) Suppose $\mathbf{I}$, $\mathbf{G}_0$, $\mathbf{G}_0^2$, $\widetilde{\mathbf{V}}$, $\mathbf{G}_0\widetilde{\mathbf{V}}$, and $\mathbf{G}_0^2\widetilde{\mathbf{V}}$ are not linearly independent. This implies there exist constants $c_1, c_2, c_3, c_4, c_5, c_6 \in \mathbb{R}$ such that $c_1\mathbf{I} + c_2\mathbf{G}_0 + c_3\widetilde{\mathbf{V}} + c_4\mathbf{G}_0\widetilde{\mathbf{V}} + c_5\mathbf{G}_0^2\widetilde{\mathbf{V}} + c_6\mathbf{G}_0^2 = \mathbf{0}$.[23] Identification fails if there exist $(\gamma, \delta, \sigma_u^2) \neq (\gamma', \delta', \sigma_u^{2\prime})$ with ($\gamma \neq 0$ or $\delta \neq 0$) and ($\gamma' \neq 0$ or $\delta' \neq 0$) such that Expression (17) holds. Direct substitution shows that this holds for $(\gamma, \delta, \sigma_u^2) = (-c_1 - c_3, c_5 + c_6, 2)$ and $(\gamma', \delta', \sigma_u^{2\prime}) = (-c_1 - 2c_3, 2c_5 + c_6, 1)$.

C.6 Proof of Proposition 8

Since the proof of Proposition 6 already makes use of local differencing, it is also valid for the extended specification.

[23]Note that the second and fourth equation are a linear combination of the other equations. These can therefore be ignored.

D Proofs for Section 5

D.1 Proof of Proposition 9

The consistency of the GMM estimator follows from Hansen (1982).

D.2 Proof of Corollary 4

The GMM estimate based on mean restrictions

$$\widehat{\boldsymbol{\theta}}_1^{GMM} = \arg\min \left(S^{-1} \sum_{s=1}^{S} \mathbf{m}_s^1(\boldsymbol{\theta}) \right)^{\intercal} \boldsymbol{\Omega}_1 \left(S^{-1} \sum_{s=1}^{S} \mathbf{m}_s^1(\boldsymbol{\theta}) \right),$$

is equal to the solution to the first order condition

$$\left(S^{-1} \sum_{s=1}^{S} \nabla_{\theta} \mathbf{m}_s^1(\boldsymbol{\theta}) \right)^{\intercal} \boldsymbol{\Omega}_1 \left(S^{-1} \sum_{s=1}^{S} \mathbf{m}_s^1(\boldsymbol{\theta}) \right) = \mathbf{0}.$$

Since $\mathbf{m}_s^1(\boldsymbol{\theta}) = \mathbf{y}_s - \mathbf{Q}_s\boldsymbol{\theta}$ and $\nabla_{\theta}\mathbf{m}_s^1(\boldsymbol{\theta}) = -\mathbf{Q}_s$, it follows directly from the first order condition that

$$\widehat{\theta}^{GMM} = \left(\overline{\mathbf{Q}}^{\intercal}\boldsymbol{\Omega}_1\overline{\mathbf{Q}}\right)^{-1}\overline{\mathbf{Q}}^{\intercal}\boldsymbol{\Omega}_1\overline{\mathbf{y}},$$

which has the form of a GLS estimator.

The optimal weighting matrix is equal to the precision matrix of the moments. Therefore,

$$\begin{aligned}
[\boldsymbol{\Omega}_1^*(\boldsymbol{\theta})]^{-1} &= \mathbb{V}(m_s^1(\boldsymbol{\theta})) \\
&= \mathbb{V}(\mathbf{y}_s - \mathbf{Q}_s\boldsymbol{\theta}) \\
&= \mathbb{V}(\mathbf{e}_s - \gamma\mathbf{u}_s - \delta\overline{\mathbf{u}}_s) \\
&= \mathbb{V}(\mathbf{e}_s) + \sigma_u^2(\gamma^2\mathbf{I} + \delta^2\mathbf{G}\mathbf{G}^{\intercal}).
\end{aligned}$$

D.3 Proof of Proposition 10

We need to show that $\mathbb{C}(\mathbf{W}_z\widetilde{\mathbf{x}}, \mathbf{u}) = 0$ and $\mathbb{C}(\mathbf{W}_z\widetilde{\mathbf{x}}, \overline{\mathbf{u}}) = 0$. For the former, we have that

$$\begin{aligned}
\mathbb{C}(\mathbf{W}_z\widetilde{\mathbf{x}}, \mathbf{u}) &= \mathbb{C}(\mathbf{W}_z\mathbf{u}, \mathbf{u}) \\
&= 0,
\end{aligned}$$

since $\text{diag}(\mathbf{W}_z) = \mathbf{0}$. For the latter, we have that

$$\begin{aligned}
\mathbb{C}(\mathbf{W}_z\widetilde{\mathbf{x}}, \overline{\mathbf{u}}) &= \mathbb{C}(\mathbf{W}_z\mathbf{u}, \overline{\mathbf{u}}) \\
&= N^{-1}\sum_i \mathbb{E}\left[(\mathbf{W}_z\mathbf{u})_i(\overline{\mathbf{u}})_i\right] \\
&= N^{-1}\sum_{i,j,k} (\mathbf{W}_z)_{ij} g_{ik} \mathbb{E}[u_j u_k] \\
&= N^{-1}\sum_{i,j} (\mathbf{W}_z)_{ij} g_{ik} \sigma_u^2 \\
&= 0,
\end{aligned}$$

since $\mathbf{W}_z\boldsymbol{\iota} = \mathbf{0}$ implies that $\sum_j (\mathbf{W}_z)_{ij} = 0$.

D.4 Proof of Proposition 11

Validity We first show the validity of the proposed instruments. A valid instrument z_i should satisfy $\mathbb{C}(z_i, e_i - \gamma u_i - \delta\overline{u}_i) = 0$. First consider the instrument $z_i = c_i$. We have that

$$\begin{aligned}
\mathbb{C}(z_i, e_i) &= 0, \\
\mathbb{C}(z_i, u_i) &= 0, \\
\mathbb{C}(z_i, \overline{u}_i) &= 0.
\end{aligned}$$

Now consider the instrument $z_i = (c_i - \mathbb{E}[c_i])\widetilde{x}_i$. We have that

$$\begin{aligned}
\mathbb{C}(z_i, e_i) &= \mathbb{E}[z_i e_i] = \mathbb{E}[z_i \mathbb{E}[e_i \mid z_i]] = 0, \\
\mathbb{C}(z_i, u_i) &= \mathbb{E}[z_i u_i] = \mathbb{E}[(c_i - \mathbb{E}[c_i])(x_i + u_i)u_i] = \mathbb{E}[(c_i - \mathbb{E}[c_i])\mathbb{E}[u_i^2 \mid c_i]] = 0, \\
\mathbb{C}(z_i, \overline{u}_i) &= \mathbb{E}[z_i \overline{u}_i] = \mathbb{E}[(c_i - \mathbb{E}[c_i])(x_i + u_i)\overline{u}_i] = 0.
\end{aligned}$$

Finally consider the instrument $z_i = d_i(c_i - \mathbb{E}[c_i])\overline{\widetilde{x}}_i$. We have that

$$\begin{aligned}
\mathbb{C}(z_i, e_i) &= \mathbb{E}[z_i e_i] = \mathbb{E}[z_i \mathbb{E}[e_i \mid z_i]] = 0, \\
\mathbb{C}(z_i, u_i) &= \mathbb{E}[z_i u_i] = \mathbb{E}[d_i(c_i - \mathbb{E}[c_i])(\overline{x}_i + \overline{u}_i)u_i] = 0, \\
\mathbb{C}(z_i, \overline{u}_i) &= \mathbb{E}[z_i \overline{u}_i] = \mathbb{E}[d_i(c_i - \mathbb{E}[c_i])(\overline{x}_i + \overline{u}_i)\overline{u}_i] = \mathbb{E}[d_i(c_i - \mathbb{E}[c_i])\mathbb{E}[\overline{u}_i^2 \mid c_i, d_i]] = 0,
\end{aligned}$$

where the last equality follows from the fact that $d_i\mathbb{E}[\overline{u}_i^2 \mid c_i, d_i] = \sigma_u^2$.

Relevance We now show the relevance of the proposed instruments. A relevant instrument z_i should satisfy $\mathbb{C}(z_i, \widetilde{x}_i) \neq 0$ or $\mathbb{C}(z_i, \overline{\widetilde{x}}_i) \neq 0$. First consider the instrument $z_i = c_i$. We have that

$$
\begin{aligned}
\mathbb{C}(z_i, \widetilde{x}_i) &= \mathbb{C}(z_i, x_i), \\
\mathbb{C}(z_i, \widetilde{\overline{x}}_i) &= \mathbb{C}(z_i, \overline{x}_i).
\end{aligned}
$$

Now consider the instrument $z_i = (c_i - \mathbb{E}[c_i])\widetilde{x}_i$. We have that

$$
\begin{aligned}
\mathbb{C}(z_i, \widetilde{x}_i) &= \mathbb{C}((c_i - \mathbb{E}[c_i])\widetilde{x}_i, \widetilde{x}_i) \\
&= \mathbb{C}(c_i\widetilde{x}_i, \widetilde{x}_i) - \mathbb{E}[c_i]\mathbb{V}(\widetilde{x}_i) \\
&= \mathbb{E}[c_i\widetilde{x}_i^2] - \mathbb{E}[c_i\widetilde{x}_i]\mathbb{E}[\widetilde{x}_i] - \mathbb{E}[c_i]\mathbb{E}[\widetilde{x}_i^2] + \mathbb{E}[c_i]\mathbb{E}[\widetilde{x}_i]^2, \\
&= \mathbb{C}(c_i, \widetilde{x}_i^2) - \mathbb{E}[\widetilde{x}_i]\mathbb{C}(c_i, \widetilde{x}_i) \\
&= \mathbb{C}(c_i, x_i^2) - \mathbb{E}[x_i]\mathbb{C}(c_i, x_i) \\
&= \mathbb{C}(c, \mathbb{V}(x \mid c)),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{C}(z_i, \widetilde{\overline{x}}_i) &= \mathbb{C}((c_i - \mathbb{E}[c_i])\widetilde{x}_i, \widetilde{\overline{x}}_i) \\
&= \mathbb{C}(c_i\widetilde{x}_i, \widetilde{\overline{x}}_i) - \mathbb{E}[c_i]\mathbb{C}(\widetilde{x}_i, \widetilde{\overline{x}}_i) \\
&= \mathbb{E}[c_i\widetilde{x}_i\widetilde{\overline{x}}_i] - \mathbb{E}[c_i\widetilde{x}_i]\mathbb{E}[\widetilde{\overline{x}}_i] - \mathbb{E}[c_i]\mathbb{E}[\widetilde{x}_i\widetilde{\overline{x}}_i] + \mathbb{E}[c_i]\mathbb{E}[\widetilde{x}_i]\mathbb{E}[\widetilde{\overline{x}}_i], \\
&= \mathbb{C}(c_i, \widetilde{x}_i\widetilde{\overline{x}}_i) - \mathbb{E}[\widetilde{\overline{x}}_i]\mathbb{C}(c_i, \widetilde{x}_i) \\
&= \mathbb{C}(c_i, x_i\overline{x}_i) - \mathbb{E}[\overline{x}_i]\mathbb{C}(c_i, x_i) \\
&= \mathbb{C}(c, \mathbb{C}(x, \overline{x} \mid c)).
\end{aligned}
$$

Finally, consider the instrument $z_i = d_i(c_i - \mathbb{E}[c_i])\widetilde{\overline{x}}_i$. We have that

$$
\begin{aligned}
\mathbb{C}(z_i, \widetilde{x}_i) &= \mathbb{C}(d_i(c_i - \mathbb{E}[c_i])\widetilde{\overline{x}}_i, \widetilde{x}_i) \\
&= \mathbb{C}(d_ic_i\widetilde{\overline{x}}_i, \widetilde{x}_i) - \mathbb{E}[c_i]\mathbb{C}(d_i\widetilde{\overline{x}}_i, \widetilde{x}_i) \\
&= \mathbb{C}(c_i, d_i\widetilde{x}_i\widetilde{\overline{x}}_i) - \mathbb{E}[\widetilde{x}_i]\mathbb{C}(c_i, d_i\widetilde{\overline{x}}_i) \\
&= \mathbb{C}(c_i, d_ix_i\overline{x}_i) - \mathbb{E}[x_i]\mathbb{C}(c_i, d_i\overline{x}_i) \\
&= \mathbb{C}(c, d\mathbb{C}(x, \overline{x} \mid c)),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{C}(z_i, \widetilde{\overline{x}}_i) &= \mathbb{C}(d_i(c_i - \mathbb{E}[c_i])\widetilde{\overline{x}}_i, \widetilde{\overline{x}}_i) \\
&= \mathbb{C}(d_ic_i\widetilde{\overline{x}}_i, \widetilde{\overline{x}}_i) - \mathbb{E}[c_i]\mathbb{C}(d_i\widetilde{\overline{x}}_i, \widetilde{\overline{x}}_i) \\
&= \mathbb{C}(c_i, d_i\widetilde{\overline{x}}_i^2) - \mathbb{E}[\widetilde{\overline{x}}_i]\mathbb{C}(c_i, d_i\widetilde{\overline{x}}_i) \\
&= \mathbb{C}(c_i, d_i\overline{x}_i^2) - \mathbb{E}[\overline{x}_i]\mathbb{C}(c_i, d_i\overline{x}_i) \\
&= \mathbb{C}(c, d\mathbb{V}(\overline{x} \mid c)).
\end{aligned}
$$

E Proofs for Appendix A

E.1 Proof of Lemma A.2

For strongly regular networks, it holds that

$$\begin{aligned}\mathbf{A}_0\mathbf{J} = \mathbf{J}\mathbf{A}_0 = d\mathbf{J},\\ \mathbf{A}_0^2 = d\mathbf{I} + \lambda\mathbf{A}_0 + \mu(\mathbf{J} - \mathbf{I} - \mathbf{A}_0),\end{aligned}$$

where $\mu = \frac{d(d-\lambda-1)}{N_0-d-1}$. With straightforward calculations, one can show that $\boldsymbol{\iota}^\intercal \mathbf{A}_0^k \boldsymbol{\iota} = N_0 d^k$ for all $k \in \mathbb{N}_+$. The expressions for the parameters of interest follow immediately:

$$\begin{aligned}
m_{10} &= N_0^{-2}\boldsymbol{\iota}^\intercal\boldsymbol{\iota} = N_0^{-1},\\
m_{20} &= N_0^{-2}\boldsymbol{\iota}^\intercal\mathbf{G}_0\boldsymbol{\iota} = N_0^{-1},\\
m_{30} &= N_0^{-2}\boldsymbol{\iota}^\intercal\mathbf{G}_0\mathbf{G}_0^\intercal\boldsymbol{\iota} = N_0^{-1},\\
m_{10} &= N_0^{-2}\boldsymbol{\iota}^\intercal\mathbf{A}_0\boldsymbol{\iota} = dN_0^{-1},\\
m_{20} &= N_0^{-2}\boldsymbol{\iota}^\intercal\mathbf{G}_0\mathbf{A}_0\boldsymbol{\iota} = d^{-1}N_0^{-2}\boldsymbol{\iota}^\intercal\mathbf{A}_0^2\boldsymbol{\iota} = dN_0^{-1},\\
m_{30} &= N_0^{-2}\boldsymbol{\iota}^\intercal\mathbf{G}_0\mathbf{A}_0\mathbf{G}_0^\intercal\boldsymbol{\iota} = d^{-2}N_0^{-2}\boldsymbol{\iota}^\intercal\mathbf{A}_0^3\boldsymbol{\iota} = dN_0^{-1},\\
m_{12} &= 1 - m_{10} - m_{11},\\
m_{22} &= 1 - m_{20} - m_{21},\\
m_{32} &= 1 - m_{30} - m_{31}.
\end{aligned}$$

E.2 Proof of Proposition A.1

Using the approximation for $\sigma_u^2 \approx 0$ from Proposition 4, the bias is approximately

$$\operatorname{plim} \widehat{\delta}_g^{OLS} \approx \frac{(1 - N_0^{-1})\mathbb{C}_g(x, \overline{x}) + N_0^{-1}\mathbb{V}_g(x)}{\Delta_g}\sigma_u^2\gamma.$$

By Lemma A.3, we have that

$$\begin{aligned}
\mathbb{C}_g(x, \overline{x}) &= \rho_1 - m_{20} - m_{21}\rho_1 - m_{22}\rho_2,\\
\mathbb{V}_g(x) &= 1 - m_{10} - m_{11}\rho_1 - m_{12}\rho_2,\\
\mathbb{V}_g(\overline{x}) &= h_0 + h_1\rho_1 + h_2\rho_2 - m_{30} - m_{31}\rho_1 - m_{32}\rho_2.
\end{aligned}$$

The assumption of strongly regular graphs simplifies things further. Using the expressions from Lemma A.2, we obtain

$$\begin{aligned}
\mathbb{C}_g(x, \overline{x}) &= N_0^{-1}(\rho_1 - 1) + t_1(\rho_1 - \rho_2), \\
\mathbb{V}_g(x) &= \left(N_0^{-1} - 1\right)(\rho_1 - 1) + t_1(\rho_1 - \rho_2), \\
\mathbb{V}_g(\overline{x}) &= (t_1 - t_2)(\rho_1 - \rho_2) + \left(N_0^{-1} - h_0\right)(\rho_1 - 1),
\end{aligned}$$

where

$$\begin{aligned}
t_1 &= 1 - \frac{d+1}{N_0} \\
t_2 &= 1 - \frac{\lambda + 1}{d} (= h_2).
\end{aligned}$$

Now suppose that $N_0 \to \infty$. It follows immediately that

$$\text{plim } \widehat{\delta}_g^{OLS} \approx \frac{\rho_1 - \rho_2}{h_0(1-\rho_2)^2 + h_1(\rho_1 - \rho_2)(1-\rho_2) - (\rho_1 - \rho_2)^2} \sigma_u^2 \gamma.$$

To establish the directions of the partial derivatives, it is convenient to denote the denominator as $D = h_0(1-\rho_2)^2 + h_1(\rho_1 - \rho_2)(1-\rho_2) - (\rho_1 - \rho_2)^2$. By the quotient rule, we have that

$$\text{sign}\left(\frac{\partial \frac{\rho_1 - \rho_2}{D}}{\partial \rho_1}\right) = \text{sign}\left(h_0(1-\rho_2)^2 + (\rho_1 - \rho_2)^2\right) \geq 0,$$

for claim (i),

$$\text{sign}\left(\frac{\partial \frac{\rho_1 - \rho_2}{D}}{\partial \rho_2}\right) = \text{sign}(-D - [\rho_1 - \rho_2][-2h_0(1-\rho_2) + h_1(2\rho_2 - \rho_1 - 1) + 2(\rho_1 - \rho_2)]) \leq 0,$$

for claim (ii),

$$\text{sign}\left(\frac{\partial \frac{\rho_1 - \rho_2}{D}}{\partial h_0}\right) = \text{sign}(-(\rho_1 - \rho_2)(1-\rho_2)^2) \leq 0,$$

for claim (iii), and

$$\text{sign}\left(\frac{\partial \frac{\rho_1 - \rho_2}{D}}{\partial h_1}\right) = \text{sign}(-(\rho_1 - \rho_2)(\rho_1 - \rho_2)(1-\rho_2)) \leq 0,$$

for claim (iv).

E.3 Proof of Lemma A.3

The expressions for the variance-covariance matrices of the measurement error are in the proof of Proposition 4. In this section, we focus on the variance-covariance matrices of the regressors.

E.3.1 Global transformation

For the variance-covariance matrix $\mathbf{S}_g$, we have that

$$\mathbf{S}_g = \mathbf{S} - N_0^{-1} \begin{bmatrix} \mathbb{C}(x_{si}, (\mathbf{Q}\mathbf{x})_{si}) & \mathbb{C}(x_{si}, (\mathbf{Q}\overline{\mathbf{x}})_{si}) \\ \mathbb{C}(x_{si}, (\mathbf{Q}\overline{\mathbf{x}})_{si}) & \mathbb{C}(\overline{x}_{si}, (\mathbf{Q}\overline{\mathbf{x}})_{si}) \end{bmatrix}.$$

For the covariances $\mathbb{C}(x_{si}, (\mathbf{Q}\mathbf{x})_{si})$, $\mathbb{C}(x_{si}, (\mathbf{Q}\overline{\mathbf{x}})_{si})$, and $\mathbb{C}(\overline{x}_{si}, (\mathbf{Q}\overline{\mathbf{x}})_{si})$, we have that

$$\begin{aligned}
\mathbb{C}(x_{si}, (\mathbf{Q}\mathbf{x})_{si}) &= \mathbb{E}(x_{si}(\mathbf{Q}\mathbf{x})_{si}) - \mathbb{E}(x_{si})\mathbb{E}((\mathbf{Q}\mathbf{x})_{si}) \\
&= N_0^{-1} \sum_{i_0, j_0} \mathbb{E}(x_{i_0} x_{j_0}) - N_0 \mu_x^2 \\
&= N_0^{-1} \sum_{i_0, j_0} (\mathbb{C}(x_{i_0}, x_{j_0}) + \mu_x^2) - N_0 \mu_x^2 \\
&= N_0^{-1} \left(\sum_{t=0}^{T} (\boldsymbol{\iota}^\intercal \mathbf{A}_0^{(t)} \boldsymbol{\iota}) \rho_t \right) \sigma_x^2 \\
&= N_0 \left(\sum_{t=0}^{T} m_{1t} \rho_t \right) \sigma_x^2,
\end{aligned}$$

$$\begin{aligned}
\mathbb{C}(x_{si}, (\mathbf{Q}\overline{\mathbf{x}})_{si}) &= \mathbb{E}(x_{si}(\mathbf{Q}\overline{\mathbf{x}})_{si}) - \mathbb{E}(x_{si})\mathbb{E}((\mathbf{Q}\overline{\mathbf{x}})_{si}) \\
&= N_0^{-1} \sum_{i_0, j_0, k_0} g_{j_0 k_0} \mathbb{E}(x_{i_0} x_{k_0}) - N_0 \mu_x^2 \\
&= N_0^{-1} \sum_{i_0, j_0, k_0} g_{j_0 k_0} (\mathbb{C}(x_{i_0}, x_{k_0}) + \mu_x^2) - N_0 \mu_x^2 \\
&= N_0^{-1} \left(\sum_{t=0}^{T} (\boldsymbol{\iota}^\intercal \mathbf{G}_0 \mathbf{A}_0^{(t)} \boldsymbol{\iota}) \rho_t \right) \sigma_x^2 \\
&= N_0 \left(\sum_{t=0}^{T} m_{2t} \rho_t \right) \sigma_x^2,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{C}(\overline{x}_{si}, (\mathbf{Q}\overline{\mathbf{x}})_{si}) &= \mathbb{E}(\overline{x}_{si}(\mathbf{Q}\overline{\mathbf{x}})_{si}) - \mathbb{E}(\overline{x}_{si})\mathbb{E}((\mathbf{Q}\overline{\mathbf{x}})_{si}) \\
&= N_0^{-1} \sum_{i_0,j_0,k_0,l_0} g_{i_0j_0} g_{k_0l_0} \mathbb{E}(x_{j_0} x_{l_0}) - N_0 \mu_x^2 \\
&= N_0^{-1} \sum_{i_0,j_0,k_0,l_0} g_{i_0j_0} g_{k_0l_0} (\mathbb{C}(x_{j_0} x_{l_0}) + \mu_x^2) - N_0 \mu_x^2 \\
&= N_0^{-1} \left(\sum_{t=0}^{T} (\boldsymbol{\iota}^\intercal \mathbf{G}_0 \mathbf{A}_0^{(t)} \mathbf{G}_0^\intercal \boldsymbol{\iota}) \rho_t \right) \sigma_x^2 \\
&= N_0 \left(\sum_{t=0}^{T} m_{3t} \rho_t \right) \sigma_x^2,
\end{aligned}$$

where $\sum_{t=0}^{T} m_{1d} = \sum_{t=0}^{T} m_{2d} = \sum_{t=0}^{T} m_{3d} = 1$. Putting things together, we have that

$$\mathbf{S}_g = \mathbb{V} \begin{bmatrix} 1 - \sum_{t=0}^{T} m_{1t}\rho_t & \rho_1 - \sum_{t=0}^{T} m_{2t}\rho_t \\ \rho_1 - \sum_{t=0}^{T} m_{2t}\rho_t & \sum_{t=0}^{2} h_t \rho_t - \sum_{t=0}^{T} m_{3t}\rho_t \end{bmatrix}.$$

E.3.2 Local transformation

For the variance-covariance matrix $\mathbf{S}_l$, we have that

$$\mathbf{S}_l = \mathbf{S} + \begin{bmatrix} \mathbb{V}(\overline{x}) - 2\mathbb{C}(x, \overline{x}) & -\mathbb{C}(x_{si}, (\mathbf{G}^2\mathbf{x})_{si}) + \mathbb{C}(\overline{x}_{si}, (\mathbf{G}^2\mathbf{x})_{si}) - \mathbb{V}(\overline{x}) \\ -\mathbb{C}(x_{si}, (\mathbf{G}^2\mathbf{x})_{si}) + \mathbb{C}(\overline{x}_{si}, (\mathbf{G}^2\mathbf{x})_{si}) - \mathbb{V}(\overline{x}) & \mathbb{V}((\mathbf{G}^2\mathbf{x})_{si}) - 2\mathbb{C}(\overline{x}_{si}, (\mathbf{G}^2\mathbf{x})_{si}) \end{bmatrix}.$$

For the covariances $\mathbb{C}(x_{si}, (\mathbf{G}^2\mathbf{x})_{si})$ and $\mathbb{C}(\overline{x}_{si}, (\mathbf{G}^2\mathbf{x})_{si})$, and the variance $\mathbb{V}((\mathbf{G}^2\mathbf{x})_{si})$, we have that

$$\begin{aligned}
\mathbb{C}(x_{si}, (\mathbf{G}^2\mathbf{x})_{si}) &= \mathbb{E}(x_{si}(\mathbf{G}^2\mathbf{x})_{si}) - \mathbb{E}(x_{si})\mathbb{E}((\mathbf{G}^2\mathbf{x})_{si}) \\
&= N_0^{-1} \sum_{i_0} \mathbb{E}(x_{i_0}(\mathbf{G}^2\mathbf{x})_{i_0}) - \mu_x^2 \\
&= N_0^{-1} \sum_{i_0,j_0,k_0} g_{i_0j_0} g_{j_0k_0} \mathbb{E}(x_{i_0} x_{k_0}) - \mu_x^2 \\
&= N_0^{-1} \left(\sum_{t=0}^{2} \text{trace}(\mathbf{G}_0^2 \mathbf{A}_0^{(t)}) \rho_t \right) \sigma_x^2 \\
&= \left(\sum_{t=0}^{2} n_{1t} \rho_t \right) \sigma_x^2,
\end{aligned}$$

$$
\begin{aligned}
\mathbb{C}(\overline{x}_{si}, (\mathbf{G}^2\mathbf{x})_{si}) &= \mathbb{E}(\overline{x}_{si}(\mathbf{G}^2\mathbf{x})_{si}) - \mathbb{E}(\overline{x}_{si})\mathbb{E}((\mathbf{G}^2\mathbf{x})_{si}) \\
&= N_0^{-1}\sum_{i_0}\mathbb{E}(\overline{x}_{i_0}(\mathbf{G}^2\mathbf{x})_{i_0}) - \mu_x^2 \\
&= N_0^{-1}\sum_{i_0,j_0,k_0,l_0} g_{i_0j_0}g_{i_0k_0}g_{k_0l_0}\mathbb{E}(x_{j_0}x_{l_0}) - \mu_x^2 \\
&= N_0^{-1}\left(\sum_{t=0}^{3}\text{trace}(\mathbf{G}_0^{\intercal}\mathbf{G}_0^2\mathbf{A}_0^{(t)})\rho_t\right)\sigma_x^2 \\
&= \left(\sum_{t=0}^{3} n_{2t}\rho_t\right)\sigma_x^2,
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbb{V}((\mathbf{G}^2\mathbf{x})_{si}) &= \mathbb{E}((\mathbf{G}^2\mathbf{x})_{si}^2) - \mathbb{E}((\mathbf{G}^2\mathbf{x})_{si})^2 \\
&= N_0^{-1}\sum_{i_0}(\mathbf{G}^2\mathbf{x})_{i_0}^2) - \mu_x^2 \\
&= N_0^{-1}\sum_{i_0,j_0,k_0,l_0,m_0} g_{i_0j_0}g_{j_0k_0}g_{i_0l_0}g_{l_0,m_0}x_{k_0}x_{m_0} - \mu_x^2 \\
&= N_0^{-1}\left(\sum_{t=0}^{4}\text{trace}((\mathbf{G}_0^{\intercal})^2\mathbf{G}_0^2\mathbf{A}_0^{(t)})\rho_t\right)\sigma_x^2 \\
&= \left(\sum_{t=0}^{4} n_{3t}\rho_t\right)\sigma_x^2,
\end{aligned}
$$

where $\sum_{t=0}^{2} n_{1t} = \sum_{t=0}^{3} n_{2t} = \sum_{t=0}^{4} n_{3t} = 1$. Putting things together, we have that

$$
\mathbf{S}_l = \mathbb{V}\begin{bmatrix} 1 + \sum_{t=0}^{2} h_t\rho_t - 2\rho_1 & \rho_1 - \sum_{t=0}^{2} n_{1t}\rho_t + \sum_{t=0}^{3} n_{2t}\rho_t - \sum_{t=0}^{2} h_t\rho_t \\ \rho_1 - \sum_{t=0}^{2} n_{1t}\rho_t + \sum_{t=0}^{3} n_{2t}\rho_t - \sum_{t=0}^{2} h_t\rho_t & \sum_{t=0}^{2} h_t\rho^t + \sum_{t=0}^{4} n_{3t}\rho_t - 2\sum_{t=0}^{3} n_{2t}\rho_t \end{bmatrix}.
$$