Estimating Peer Effects Using Partial Network Data

Vincent Boucher[†] and Elysée Aristide Houndetoungan[‡]

March 2025

Abstract

We study the estimation of peer effects through social networks when researchers do not observe the entire network structure. Special cases include sampled networks, censored networks, and misclassified links. We assume that researchers can obtain a consistent estimator of the distribution of the network. We show that this assumption is sufficient for estimating peer effects using a linear-in-means model. We provide an empirical application to the study of peer effects on students' academic achievement using the widely used Add Health database and show that network data errors have a first-order downward bias on estimated peer effects.

JEL Codes: C31, C36, C51

Keywords: Social networks, Peer effects, Missing variables, Measurement errors

[†]Boucher: Department of Economics, Université Laval, CRREP, CREATE, CIRANO; email: vincent.boucher@ecn.ulaval.ca. [‡] Houndetoungan: Cy Cergy Paris Université and Thema; email: aristide.houndetoungan@cyu.fr. An R package, including all replication codes, is available at: https://github.com/ahoundetoungan/PartialNetwork.

Acknowledgements: We would like to thank Yann Bramoullé, and Bernard Fortin for their helpful comments and insights, as always. We would also like to thank Isaiah Andrews, Eric Auerbach, Arnaud Dufays, Stephen Gordon, Chih-Sheng Hsieh, Arthur Lewbel, Tyler McCormick, Angelo Mele, Francesca Molinari, Onur Özgür, Eleonora Patacchini, Xun Tang, and Yves Zenou for helpful comments and discussions. Thank you also to the participants of the many seminars at which we presented this research. This research uses data from Add Health, a program directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by Grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is given to Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). No direct support was received from Grant P01-HD31921 for this research.

1 Introduction

There is a large and growing literature on the impact of peer effects in social networks.[1] However, since eliciting network data is expensive (Breza et al., 2020), relatively few data sets contain comprehensive network information, and existing ones are prone to data errors. Despite some recent contributions, existing methodologies for the estimation of peer effects with incomplete or erroneous network data either focus on a specific kind of sampling or error, or they are highly computationally demanding.

In this paper, we propose a unifying framework that allows for the estimation of peer effects under the widely used linear-in-means model (e.g. Manski (1993); Bramoullé et al. (2009)) when the researcher does not observe the entire network structure. Our methodology is computationally attractive and sufficiently flexible to cover cases where, for example, network data are sampled (Chandrasekhar and Lewis, 2011; Liu, 2013; Lewbel et al., 2022), censored (Griffith, 2022), or misclassified (Hardy et al., 2024). Our central assumption is that the researcher is able to estimate a network formation model using some partial information about the network structure. Leveraging recent contributions on the estimation of network formation models, we show that this assumption is sufficient to identify and estimate peer effects.

We propose two estimators. First, we present a computationally attractive estimator based on a simulated generalized method of moments (SGMM). The moments are built using draws from the (estimated) network formation model. We study the finite sample properties of our SGMM estimator via Monte Carlo simulations. We show that the estimator

[1] For recent reviews, see Bramoullé et al. (2020), and De Paula (2017).

performs very well, even when a large fraction of the links are missing or misclassified. Second, we present a flexible likelihood-based (Bayesian) estimator allowing us to exploit the entire structure of the data-generating process. The Bayesian approach is flexible as it allows to cover cases for which the asymptotic framework of our SGMM fails. Although the computational cost is higher than that of the SGMM, we exploit recent computational advances in the literature, e.g. Mele (2017); Hsieh et al. (2019), and show that the estimator can be successfully implemented on common-sized data sets. In particular, we apply our estimator to study peer effects on academic achievement using the widely used Add Health database. We find that data errors have a first-order downward bias on the estimated endogenous effect.

Our SGMM estimator is built as a bias-corrected version of the instrumental strategy proposed by Bramoullé et al. (2009). Using a network formation model, we obtain a consistent estimator of the distribution of the true network. We then use this estimated distribution to obtain different draws from the distribution of the network. We show that our moment conditions are asymptotically valid and that the estimator is consistent and asymptotically normal, even with a finite number of draws from the estimated distribution of the network. This property significantly reduces the computational cost of the method compared to methods that rely on integrating the moment conditions (e.g., Chandrasekhar and Lewis, 2011).

Importantly, our SGMM strategy requires only the (partial) observation of a *single* cross-section, unlike, for example, the approach of Zhang (2024). The presence of this feature is because of two main properties of the model. First, we can consistently estimate the distribution of the mismeasured variable (i.e., the network) using a single (partial) observation of the variable. Second, in the absence of measurement error, valid instruments for the endogenous peer variable are available (Bramoullé et al., 2009).

Our Bayesian estimator is based on the likelihood function and therefore uses more information about the structure of the model, leading to more precise estimates. In the context of this estimator, the estimated distribution for the network acts as a prior distribution, and the inferred network structure is updated through a Markov chain Monte Carlo (MCMC) algorithm. Our approach relies on data augmentation (Tanner and Wong, 1987), which treats the network as an additional set of parameters to be estimated. In particular, our MCMC builds on recent developments from the empirical literature on network formation (e.g., Mele, 2017; Hsieh et al., 2019). We show that the computational cost of our estimator is reasonable and that it can easily be applied to standard data sets.

We study the impact of errors in adolescents' friendship network data for the estimation of peer effects in education (Calvó-Armengol et al., 2009). We show that the widely used Add Health database features many missing links—around 45% of the within-school friendship nominations are coded with error—and that these data errors strongly bias the estimated peer effects. Specifically, we estimate a model of peer effects on students' academic achievement. We probabilistically reconstruct the missing links, accounting for the potential censoring, and we obtain a consistent estimator of peer effects using both our estimators. The bias due to data errors is qualitatively important, even assuming that the network is exogenous. Our estimated endogenous peer effect coefficient is 1.5 times larger than that obtained by assuming the data contains no errors.

This paper contributes to the recent literature on the estimation of peer effects when the network is either not entirely observed or observed with noise. In particular, our framework is valid when network data are either sampled, censored, or misclassified.[2] We unify these strands in the literature and provide a flexible and computationally tractable framework for estimating peer effects with incomplete or erroneous network data.

Sampled networks and censoring: Chandrasekhar and Lewis (2011) show that models estimated using sampled networks are generally biased. They propose an analytical correction as well as a two-step general method of moment (GMM) estimator. Liu (2013) shows that when the interaction matrix is not row-normalized, instrumental variable estimators based on an out-degree distribution are valid, even with sampled networks. Hsieh et al. (2024) focus on a regression model that depends on global network statistics. They propose analytical corrections to account for nonrandom sampling of the network (see also Chen et al., 2013). Thirkettle (2019) also focuses on global network statistics, assuming that the researcher only observes a random sample of links. Using a structural network formation model, he derives bounds on the identified set for both the network formation model and the network statistic of interest. Lewbel et al. (2024b) develop a two-stage least squares estimator for the linear-in-means model when some links are potentially misclassified. They propose valid instruments under some restrictions on the observed and true interactions matrices, or when researchers observe at least two samples of the same true network. Finally, Zhang (2024) studies program evaluation in a context in which networks are sampled locally and where some links might be unobserved. Assuming that the researcher has access to two measurements of the network for each sampled unit, she presents a nonparametric estimator of the treatment and spillover effects.

[2]For related literature that studies the estimation of peer effects when researchers have no network data, see Manresa (2016); De Paula et al. (2024); Lewbel et al. (2022).

Relatedly, Griffith (2022) explores the impact of imposing an upper bound to the number of links when eliciting network data, e.g., "Name your five best friends." He presents a bias-correction method and explores the impact of censoring using two empirical applications. He finds that censoring underestimates peer effects. Griffith and Kim (2023) present a characterization of the analytic bias of censoring for the reduced-form parameters in the linear-in-means and linear-in-sums models under an Expectational Equivalence assumption.

We contribute to this literature by proposing two simple and flexible estimators for the estimation of peer effects based on a linear-in-means model. Our estimators do not require many observations of the sampled network. Similar to Griffith (2022) and Griffith and Kim (2023), we find—using the Add Health database—that sampling leads to an underestimation of peer effects, although we find that censoring has a negligible impact, in the context of peer effects, on academic achievement.

Our SGMM estimator does not suffer from the computational cost resulting from integrating the moment conditions (as in Chandrasekhar and Lewis, 2011) and can produce precise estimates with as little as three network simulations. While our Bayesian estimator is more computationally demanding, we exploit recent developments from the empirical literature on network formation (e.g., Mele, 2017; Hsieh et al., 2019) and show that it is computationally tractable. Moreover, the Bayesian estimator is valid in finite samples, which allows in particular to cover cases not covered by the asymptotic framework on which our SGMM relies.

Misclassification: Hardy et al. (2024) look at the estimation of (discrete) treatment effects when the network is observed noisily. Specifically, they assume that observed links are affected by iid errors and present an expectation maximization (EM) algorithm that allows for a consistent estimator of the treatment effect. Lewbel et al. (2024a) show that when the expected number of missing links grows at a rate strictly lower than the number of sampled individuals n, the 2SLS estimator in Bramoullé et al. (2009) is consistent.[3]

Our model allows for the misclassification of all links with positive probability, and we do not impose restrictions on the rate of misclassification. As in Hardy et al. (2024), we use a network formation model to estimate the probability of false positives and false negatives. However, our two-stage strategy—estimating the network formation model and then the peer effect model—allows for greater flexibility. In particular, our network formation model is allowed to flexibly depend on covariates. This is empirically important, as networks typically feature homophily on observed characteristics (e.g., Currarini et al., 2010; Bramoullé et al., 2012).[4]

The remainder of the paper is organized as follows. In Section 2, we present the econometric model as well as the main assumptions. In Section 3, we present our SGMM estimator and study its performance via Monte Carlo simulations. In Section 4, we present our likelihood-based estimation strategy. In Section 5, we present our application to peer effects on academic achievement. Section 6 concludes the paper.

[3]When the growth rate is strictly smaller than $\sqrt{n}$, the inference is also valid.

[4]Formally, Hardy et al. (2024) do not require the specification of distribution of the network, but only the distribution of the degree sequence, which is assumed not to depend on covariates.

2 The Model

We assume that the data are partitioned into $M > 1$ groups, where group m contains N_m individuals. A sample consists of the following:

$$\{\mathbf{y}_m, \mathbf{X}_m, \boldsymbol{\varepsilon}_m; \mathcal{A}_m, \mathbf{A}_m\}_{m=1}^{M}.$$

For individuals in group m, $\mathbf{y}_m$ is a vector of an observed outcome of interest (e.g., academic achievement), $\mathbf{X}_m$ is an observed matrix of individual characteristics (e.g., age and gender), and $\boldsymbol{\varepsilon}_m$ is a vector of *unobserved* individual heterogeneity.

The matrix $\mathbf{A}_m$ is the $N_m \times N_m$ *adjacency matrix* of the network between individuals in group m. We assume a directed network:[5] $a_{ij,m} \in \{0, 1\}$, where $a_{ij,m} = 1$ if i is linked to j. We normalize $a_{ii,m} = 0$ for all i and let $n_{i,m} = \sum_j a_{ij,m}$ denote the number of links of i within group m.

We assume that $\mathbf{A}_m$ is *not observed* but that researchers observe $\mathcal{A}_m$ instead. Informally, the idea is that $\mathcal{A}_m$ contains some information about the adjacency matrix $\mathbf{A}_m$. Our specific assumptions are presented in Section 2.2. The next assumptions formalize the above discussion.

Assumption 1. *The population is partitioned into $M > 1$ groups, where the size N_m of each group $m = 1, ..., M$ is bounded. The sequence $\{\mathbf{y}_m, \mathbf{X}_m, \boldsymbol{\varepsilon}_m; \mathcal{A}_m, \mathbf{A}_m\}$ is independent across m. Moreover, $\mathbf{X}_m$ is uniformly bounded in m.*[6]

Assumption 2. *For each group m, the variables $\mathbf{y}_m$, $\mathbf{X}_m$ and $\mathcal{A}_m$ are observed. The variables $\boldsymbol{\varepsilon}_m$ and $\mathbf{A}_m$ are not.*

[5] All of our results hold for undirected networks.

[6] i.e., $\sup_{m\geq 1} \|\mathbf{X}_m\|_2 < \infty$, where $\|.\|_2$ is the Euclidean norm.

Assumption 1 implies a "many markets" asymptotic framework, meaning that the number of groups M goes to infinity as the number of individuals N goes to infinity. It is a standard assumption in the literature on the econometrics of games and the literature on peer effects.[7] For example, the data could consist of a collection of small villages (Banerjee et al., 2013) or schools (Calvó-Armengol et al., 2009). Assumption 2 implies in particular that the data are composed of group-level censuses for $\mathbf{y}_m$ and $\mathbf{X}_m$.[8] A similar assumption is made by Breza et al. (2020).

2.1 The Linear-in-Means Model

In this section, we present the linear-in-means model (Manski, 1993; Bramoullé et al., 2009), arguably the most widely used model for studying peer effects in networks (see Bramoullé et al., 2020, for a recent review).

Let $\mathbf{G}_m = f(\mathbf{A}_m)$, the $N_m \times N_m$ *interaction matrix* for some function f. Unless otherwise stated, we assume that $\mathbf{G}_m$ is a row-normalization of the adjacency matrix $\mathbf{A}_m$.[9] Most of our results hold for any function f.

We focus on the following model:

$$\mathbf{y}_m = c\mathbf{1}_m + \mathbf{X}_m\boldsymbol{\beta} + \alpha\mathbf{G}_m\mathbf{y}_m + \mathbf{G}_m\mathbf{X}_m\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_m, \tag{1}$$

where $\mathbf{1}_m$ is a N_m-dimensional vector of 1's. The parameter α therefore captures the impact of the average outcome of one's peers on their behavior (the endogenous peer effect). The parameter $\boldsymbol{\beta}$ captures the impact of one's characteristics on their behavior (the individual

[7] See for example Bramoullé et al. (2020), Breza (2016), and De Paula (2017).
[8] Contrary to Liu et al. (2017) or Wang and Lee (2013), for example.
[9] In such a case, $g_{ij,m} = a_{ij,m}/n_{i,m}$ whenever $n_{i,m} > 0$, whereas $g_{ij,m} = 0$ otherwise.

effects). The parameter $\boldsymbol{\gamma}$ captures the impact of the average characteristics of one's peers on their behavior (the contextual peer effects). For simplicity, we assume that the constant c does not vary across m. However, our results hold when considering group-level fixed effects.

We impose the following assumptions.

Assumption 3. *$|\alpha| < 1/\|\mathbf{G}_m\|$ for some submultiplicative norm $\|\cdot\|$, and all $m = 1, ..., M$.*

Assumption 4. *Exogeneity: $\mathbb{E}[\boldsymbol{\varepsilon}_m|\mathbf{X}_m, \mathbf{A}_m, \mathcal{A}_m] = \mathbf{0}$ for all $m = 1, ..., M$.*

Assumption 3 ensures that the model is coherent and that there exists a unique vector $\mathbf{y}_m$ compatible with (1). When $\mathbf{G}_m$ is row-normalized, $|\alpha| < 1$ is sufficient. Finally, Assumption 4 implies that individual characteristics and the network structure are exogenous. While the exogeneity of the network is a strong assumption, we consider it as a benchmark and focus on the case in which the network is not perfectly observed. We now describe the network sampling process in more detail.

2.2 Partial Network Information

In this paper, we relax the costly assumption that the adjacency matrix $\mathbf{A}_m$ is observed. We assume instead that sufficient information about the network (i.e., $\mathcal{A}_m$) is observed so that a network formation model can be estimated. The discussion below formalizes our assumptions about the relationship between $\mathcal{A}_m$ and $\mathbf{A}_m$. We start by describing the data-generating process for $\mathbf{A}_m$.

We assume that for any group m, $P(\mathbf{A}_m|\mathbf{X}_m) = \Pi_{ij} P(a_{ij,m}|\mathbf{X}_m)$, where

$$P(a_{ij,m}|\mathbf{X}_m) = \frac{\exp\{a_{ij,m} Q(\boldsymbol{\rho}_0, \mathbf{w}_{ij,m})\}}{1 + \exp\{Q(\boldsymbol{\rho}_0, \mathbf{w}_{ij,m})\}}, \tag{2}$$

and where Q is some known function that is twice continuously differentiable in $\boldsymbol{\rho}$, and $\mathbf{w}_{ij,m} = \mathbf{w}_{ij,m}(\mathbf{X}_m)$ is a vector of observed characteristics for the pair ij in group m.[10]

We focus on network formation models that are conditionally independent across links: $P(\mathbf{A}_m|\mathbf{X}_m) = \Pi_{ij} P(a_{ij,m}|\mathbf{X}_m)$. This notably excludes many models of strategic network formation such as the ones in Mele (2017) and De Paula et al. (2018). This is a strong assumption that deserves some discussion.

As we describe below (see Assumption 5), our strategy depends on our ability to estimate the network formation model without observing the entire network structure. In models that allow for dependence across links (e.g., Exponential Random Graph Models, henceforth ERGM), the feasibility of this strategy is highly context-dependent. Thus, for simplicity (and clarity) of the analysis in the main text, we restrict our attention to network formation models that are conditionally independent across links. We however note that our methodology can be adapted to more general network formation models.

In Online Appendix D, we further discuss how this can be done for a few specific network formation processes. First, we discuss the model in Graham (2017), which is conditionally independent across links but accounts for unobserved degree heterogeneity. We show that his model can be used, provided that the degree distribution is observed.[11] Second, we discuss ERGM. We explain how our setup can be adapted to the model in Boucher and Mourifié (2017), in which the probability of a link is a function of the individuals' degree. Similarly to Graham (2017), the estimation also requires the observation of the degree distribution. We

[10] Throughout, P refers to the probability notation. Note that by construction, links are only defined between individuals of the same group so the probability that individuals from different groups are linked is zero.

[11] Note that the degree distribution can be obtained from survey questions by simply asking individuals about the number of links they have.

also discuss the estimation of more general ERGM and their implication for survey design in the Online Appendix D.3. We now present our main assumption.

Assumption 5 (Partial Network Information)**.** *Given $\{\mathcal{A}_m, \mathbf{X}_m\}_{m=1}^M$ and the parametric model (2), there exists an estimator $\hat{\boldsymbol{\rho}}_M$, such that $\sqrt{M}(\hat{\boldsymbol{\rho}}_M - \boldsymbol{\rho}_0) \to_d N(\mathbf{0}, \boldsymbol{V}_{\boldsymbol{\rho}})$ as $M \to \infty$.*

Assumption 5 implies that the dependence between $\mathcal{A}_m$ and $\mathbf{A}_m$ is strong enough so that, using (2), the researcher can estimate the data generating process for $\mathbf{A}_m$. We can then use this information in order to obtain an estimator of the conditional distribution of $\mathbf{A}_m$.

Definition 1. *A consistent estimator of the distribution of the true network for some function κ is a probability distribution $\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m))$ such that $\sup_m \|\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m)) - P(\mathbf{A}_m|\mathbf{X}_m, \kappa(\mathcal{A}_m))\| \to_p 0$ as $M \to \infty$.*

Note that here, the partial information $\mathcal{A}_m$ is used twice. First, in order to estimate $\boldsymbol{\rho}$ (Assumption 5), and second, to construct a consistent estimator of the distribution of the true network (Definition 1). For this second use, we allow the researcher to consider only part of the information in $\mathcal{A}_m$.

Indeed, the function κ controls how much information in $\mathcal{A}_m$ is used in order to complement the information obtained by estimating the network formation process in Equation 2. Two important polar cases are the identity function $\kappa(\mathcal{A}_m) = \mathcal{A}_m$ implying that all the information in $\mathcal{A}$ is used, and the constant function $\kappa(\mathcal{A}_m) = \kappa_0$ for all $\mathcal{A}_m$ in which no information on $\mathcal{A}$ is used. Although our methodology is valid for any κ, the choice of κ may strongly affect the identification and precision of our estimators.

When κ is the identify function, the estimator is obtained from Bayes' rule:

$$\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \mathcal{A}_m) = \frac{P(\mathcal{A}_m|\mathbf{X}_m, \mathbf{A}_m)P(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m)}{P(\mathcal{A}_m|\mathbf{X}_m)}. \tag{3}$$

However, in some contexts, such a quantity may be hard to compute, depending on the nature of the information in $\mathcal{A}_m$. A solution, therefore, could be to disregard the information in $\mathcal{A}$ and use:

$$\hat{P}(a_{ij,m}|\hat{\boldsymbol{\rho}}, \mathbf{X}_m) = \frac{\exp\{a_{ij,m}Q(\hat{\boldsymbol{\rho}}, \mathbf{w}_{ij,m})\}}{1+\exp\{Q(\hat{\boldsymbol{\rho}}, \mathbf{w}_{ij,m})\}}.$$

In that case, the precision of the estimator strongly depends on the network formation process in (2). Thus, the loss in precision is context-dependent. In particular, it depends on the heterogeneity in the probability of link formation implied by (2), and on the specificity about $\mathbf{A}_m$ that is contained in $\mathcal{A}_m$.

We specifically discuss three leading examples in which Assumption 5 holds and focus on how $\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m))$ is constructed: *sampled networks* (Example 1), *censored networks* (Example 2), and *misclassified network links* (Example 3).

However, as discussed, the many-markets asymptotic framework (see Assumption 1) is restrictive and does not hold in some contexts such as when researchers have access to *aggregated relational data* (see the Online Appendix H). In Section 4, we present a Bayesian estimator which uses the same structure as the one presented here but is valid in finite samples.

Example 1 (Sampled Networks)**.** *Suppose that we observe the realizations of* a_{ij} *for a random sample of* m *pairs (e.g., Chandrasekhar and Lewis, 2011). Here* $\mathcal{A}_m$ *is simply a list of sampled pairs:* $\mathcal{A}_m = \{a_{ij,m}\}_{ij \text{ is sampled}}$ *(see e.g., Conley and Udry, 2010, for concrete example). Consider the following simple network formation model:*

$$P(a_{ij,m}=1|\mathbf{X}_m) = \frac{\exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}}{1+\exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}}.$$

In this case, a simple logistic regression on the subset of sampled pairs provides a consistent estimator of $\boldsymbol{\rho}$ *since pairs of individuals for which* $a_{ij,m}$ *is observed is random.*

In this simple framework, the linking status of sampled pairs of individuals is known. As such it is natural to define κ *as the identity map, which leads to the estimator* $\hat{P}(a_{ij,m}|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \mathcal{A}_m) = a_{ij,m}$ *for all sampled pairs* ij*, and* $\hat{P}(a_{ij,m}|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \mathcal{A}_m) = \exp\{\mathbf{w}_{ij,m}\hat{\boldsymbol{\rho}}\}/(1+\exp\{\mathbf{w}_{ij,m}\hat{\boldsymbol{\rho}}\})$ *otherwise. In essence, sampled pairs are used to estimate the network formation model, which is then used in order to predict the probability of a link for pairs that are not sampled.*

Example 2 (Censored Network Data)**.** *As discussed in Griffith (2022), network data is often censored. This typically arises when surveyed individuals are asked to name only* $T > 1$ *links (among the* N_m *possible links they may have). Here,* $\mathcal{A}_m$ *can be represented by an* $N_m \times N_m$ *binary matrix* $\mathbf{A}_m^{obs}$ *which takes value* $a_{ij,m} = 1$ *if* i *nominated* j*, and* 0 *otherwise. Consider the same simple model as in Example 1:*

$$P(a_{ij,m} = 1|\mathbf{X}_m) = \frac{\exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}}{1+\exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}}.$$

In Section 5 and the Online Appendix G.2, we present how to estimate $\boldsymbol{\rho}$ *in detail. Here, we discuss how to obtain the estimator* $\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m))$ *given* $\hat{\boldsymbol{\rho}}$*. Note that* $\hat{P}(a_{ij,m} = 1|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, a_{ij,m}^{obs} = 1) = 1$ *because observed links necessarily exist. Second, note also that for any individual* i*, such that* $n_{i,m} < T$*, we have* $\hat{P}(a_{ij,m}|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, a_{ij,m}^{obs}) = a_{ij}^{obs}$ *for all* j*, as their network data are not censored.*

Thus, the structural model is only used to obtain the probability of links that are not observed for individuals whose links are potentially censored, i.e., $\hat{P}(a_{ij,m} = 1|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, a_{ij,m}^{obs} = 0) = \exp\{\mathbf{w}_{ij,m}\hat{\boldsymbol{\rho}}\}/(1+\exp\{\mathbf{w}_{ij,m}\hat{\boldsymbol{\rho}}\})$ *for all* ij*, such that* $n_i \geq T$*.*

Example 3 (Misclassification)**.** *[Hardy et al. (2024)](#) study cases in which networks are observed but may include misclassified links (i.e., false positives and false negatives). Here, $\mathcal{A}_m$ can be represented by an $N_m \times N_m$ binary matrix $\mathbf{A}_m^{mis}$. Consider the same simple model as in Example 1 and 2:*

$$P(a_{ij,m} = 1|\mathbf{X}_m) = \frac{\exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}}{1 + \exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}}.$$

The (consistent) estimation $\boldsymbol{\rho}$ in such a context follows directly from the existing literature on misclassification in binary outcome models, e.g., Hausman et al. (1998). In this context, the simplicity of the sampling scheme allows to consider the identity map $\kappa(\mathcal{A}_m) = \mathcal{A}_m$. The estimator for the distribution of the true network can be obtained using Bayes' rule. We consider this case in our Monte Carlo simulations in Section 3.1.

3 Simulated Generalized Method of Moment Estimators

In this section, we present an estimator based on a Simulated Generalized Method of Moments (SGMM). Our SGMM is constructed as a de-biased simulated version of the widely used linear GMM in Bramoullé et al. (2009).

Before presenting the estimator, we start with an informal discussion of how the moment function is built. A formal treatment is presented in Appendix A. Recall first the linear-in-means model presented in the previous section:

$$\mathbf{y}_m = \mathbf{V}_m\tilde{\boldsymbol{\theta}} + \alpha\mathbf{G}_m\mathbf{y}_m + \boldsymbol{\varepsilon}_m,$$

where we defined $\mathbf{V}_m = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m]$, and $\tilde{\boldsymbol{\theta}} = [c, \boldsymbol{\beta}', \boldsymbol{\gamma}']'$. A valid set of instruments for the endogenous variable $\mathbf{G}_m\mathbf{y}_m$ is: $\mathbf{Z}_m = [\mathbf{1}_m, \mathbf{X}_m, \mathbf{G}_m\mathbf{X}_m, \mathbf{G}_m^2\mathbf{X}_m, \mathbf{G}_m^3\mathbf{X}_m, ...]$.

(Bramoullé et al., 2009) This defines the following moment function: $\mathbf{m}(\boldsymbol{\theta}) = \mathbf{Z}_m'\boldsymbol{\varepsilon}_m$, where $\boldsymbol{\theta} = [\alpha, c, \boldsymbol{\beta}', \boldsymbol{\gamma}']'$, and one can easily show that $\mathbb{E}[\mathbf{m}(\boldsymbol{\theta})|\mathbf{A}_m, \mathbf{X}_m] = \mathbf{0}$ for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and that $\boldsymbol{\theta}_0$ is identified under the usual rank condition.[12]

Unfortunately, this approach is not feasible when $\mathbf{G}_m = f(\mathbf{A}_m)$ is not observed. As discussed, our strategy is to develop a simulated version of this simple linear GMM estimator. Indeed, equipped with a consistent estimator of the distribution of $\mathbf{A}_m$ (see Definition 1), we can draw network structures from that same distribution.

To simplify the notation, we denote $\dot{\mathbf{G}}_m = f(\dot{\mathbf{A}}_m)$, $\ddot{\mathbf{G}}_m = f(\ddot{\mathbf{A}}_m)$, and $\dddot{\mathbf{G}}_m = f(\dddot{\mathbf{A}}_m)$ as independent draws from the distribution $\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m))$. We will also note $\dot{\mathbf{Z}}_m$ and $\dot{\mathbf{V}}_m$, the versions of $\mathbf{Z}_m$ and $\mathbf{V}_m$ in which $\mathbf{G}_m$ is replaced with $\dot{\mathbf{G}}_m$ (and similarly for $\ddot{\mathbf{G}}_m$ and $\dddot{\mathbf{G}}_m$).

Now, suppose that we replace the unobserved $\mathbf{G}_m$ with $\dot{\mathbf{G}}_m$ everywhere in the expression $\mathbf{Z}_m'\boldsymbol{\varepsilon}_m$. This would lead to a moment function with an expectation given by:

$$\begin{aligned}
\mathbb{E}(\dot{\mathbf{m}}(\boldsymbol{\theta})|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m) &= \mathbb{E}(\dot{\mathbf{Z}}_m'[(\mathbf{I}_m - \alpha\dot{\mathbf{G}}_m)\mathbf{y}_m - \dot{\mathbf{V}}\tilde{\boldsymbol{\theta}}]|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m) \\
&= \mathbb{E}(\dot{\mathbf{Z}}_m'\dot{\boldsymbol{\varepsilon}}|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m),
\end{aligned}$$

where $\mathbf{I}_m$ is the identity matrix of dimension N_m. The expectation of the moment function does not generally equal $\mathbf{0}$ when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, even asymptotically.[13]

There are two issues with the previous moment function. First, the instruments and the explanatory variables are generated using the same network draw $\dot{\mathbf{G}}_m$, which introduces a correlation between the $\dot{\mathbf{Z}}_m$ and $\dot{\boldsymbol{\varepsilon}}$, conditionally on $\mathbf{A}_m$, $\mathcal{A}_m$, and $\mathbf{X}_m$. This can easily be

[12] As standard, we use the subscript 0 to denote the true value of the parameter. See e.g., Bramoullé et al. (2009) and Lee et al. (2010) for identification results when $\mathbf{G}_m$ is observed.

[13] Recall from Definition 1 that $\dot{\mathbf{G}}_m$ is drawn from the same distribution as $\mathbf{G}_m$ only as $M \to \infty$.

resolved by simply using different draws to construct the instruments and the explanatory variables. This leads to:

$$\begin{aligned}\mathbb{E}\left(\ddot{\mathbf{m}}(\boldsymbol{\theta})|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m\right) &= \mathbb{E}(\dot{\mathbf{Z}}_m'\ddot{\boldsymbol{\varepsilon}}|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m) \\ &= \mathbb{E}(\dot{\mathbf{Z}}_m'|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m)\mathbb{E}(\ddot{\boldsymbol{\varepsilon}}|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m),\end{aligned}$$

where $\ddot{\boldsymbol{\varepsilon}}_m = (\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)\mathbf{y}_m - \ddot{\mathbf{V}}\tilde{\boldsymbol{\theta}}$. However, in general, $\mathbb{E}(\ddot{\boldsymbol{\varepsilon}}|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m) \neq \mathbf{0}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. To see why, note that we can rewrite:

$$\ddot{\boldsymbol{\varepsilon}}_m = (\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)(\mathbf{I}_m - \alpha_0\mathbf{G}_m)^{-1}[\mathbf{V}_m\tilde{\boldsymbol{\theta}}_0 + \boldsymbol{\varepsilon}_m] - \ddot{\mathbf{V}}_m\tilde{\boldsymbol{\theta}}.$$

While we can show that $\mathbb{E}(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)(\mathbf{I}_m - \alpha_0\mathbf{G}_m)^{-1}\boldsymbol{\varepsilon}|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m) = \mathbf{0}$ from the law of iterated expectations and Assumption 4, we have:

$$\mathbb{E}[(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)(\mathbf{I}_m - \alpha_0\mathbf{G}_m)^{-1}\mathbf{V}_m\tilde{\boldsymbol{\theta}}_0|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m] - \mathbb{E}[\ddot{\mathbf{V}}_m\tilde{\boldsymbol{\theta}}|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m] \neq 0,$$

when $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, even asymptotically. This is due to the approximation error in using $\ddot{\mathbf{G}}_m$ instead of $\mathbf{G}_m$. This approximation error does not vanish asymptotically. In particular, because groups have bounded, the product $(\mathbf{I}_m - \alpha_0\ddot{\mathbf{G}}_m)(\mathbf{I}_m - \alpha_0\mathbf{G}_m)^{-1}$ does not converge to the identity matrix. If it did, consistency would follow.

Our SGMM presented below offers a bias-corrected version of this estimator. Specifically, consider the following (feasible) approximation of the bias of $\mathbb{E}(\ddot{\boldsymbol{\varepsilon}}|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m)$:

$$\boldsymbol{\delta}_m = (\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)(\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m)^{-1}\dddot{\mathbf{V}}_m\tilde{\boldsymbol{\theta}} - \ddot{\mathbf{V}}\tilde{\boldsymbol{\theta}}.$$

We obtain $\ddot{\boldsymbol{\varepsilon}}_m - \boldsymbol{\delta}_m = (\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)\mathbf{y}_m - (\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)(\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m)^{-1}\dddot{\mathbf{V}}_m\tilde{\boldsymbol{\theta}}$, and we can show that $\mathbb{E}(\ddot{\boldsymbol{\varepsilon}}_m - \boldsymbol{\delta}_m|\mathbf{A}_m, \mathcal{A}_m, \mathbf{X}_m) = \mathbf{0}$ for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ as $M \to \infty$.[14]

[14] In the bias approximation expression, we use a third independent draw, $\dddot{\mathbf{G}}_m$, as an approximation of $\mathbf{G}_m$ to ensure that it remains independent of $\dot{\mathbf{G}}_m$ and $\ddot{\mathbf{G}}_m$, just as $\mathbf{G}_m$ is.

The above discussion thus leads to the definition of our SGMM. Let $\{\dot{\mathbf{G}}_m^{(r)}\}_{r=1}^R$, $\{\ddot{\mathbf{G}}_m^{(s)}\}_{s=1}^S$, and $\{\dddot{\mathbf{G}}_m^{(t)}\}_{t=1}^T$ be sequences of independent draws from estimator of the network formation process (see Definition 1), we have the following.

Theorem 1 (SGMM)**.** *Suppose that Assumptions 1–5 and regularity conditions 6–10 and 12 hold (see Appendix A). Suppose also that the identification condition 11 holds (see Appendix A). Let* $\dot{\mathbf{Z}}_m^{(r)} = [\mathbf{1}_m, \mathbf{X}_m, \dot{\mathbf{G}}_m^{(r)}\mathbf{X}_m, (\dot{\mathbf{G}}_m^{(r)})^2\mathbf{X}_m, (\dot{\mathbf{G}}_m^{(r)})^3\mathbf{X}_m, ...]$ *and* $\dddot{\mathbf{V}}_m^{(t)} = [\mathbf{1}_m, \mathbf{X}_m, \dddot{\mathbf{G}}_m^{(t)}\mathbf{X}_m]$. *Consider also the following (simulated) moment function:*

$$\bar{\mathbf{m}}_M(\boldsymbol{\theta}) = \frac{1}{M}\sum_m \frac{1}{RST}\sum_{rst} \dot{\mathbf{Z}}_m^{(r)\prime}\left[(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m^{(s)})\left(\mathbf{y}_m - (\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m^{(t)})^{-1}\dddot{\mathbf{V}}_m^{(t)}\tilde{\boldsymbol{\theta}}\right)\right] \tag{4}$$

Then, for any positive integers R, S, and T, the (simulated) GMM estimator based on (4) is consistent and asymptotically normally distributed.

The identification condition is standard and ensures that the moment condition is uniquely solved at $\boldsymbol{\theta}_0$. We discuss it in more detail below.

Theorem 1 presents conditions for the consistency and asymptotic normality of our two-step estimator. In particular, similar to a standard simulated GMM (Gourieroux et al., 1996), consistency holds for a finite number of simulations. Our estimator therefore does not suffer from the curse of dimensionality faced by Chandrasekhar and Lewis (2011).[15]

Here, a few remarks regarding the consistency and asymptotic normality are in order. Note that the simulated moment function is based on network draws that depend on an *estimated* distribution. In particular, we have: $\dot{\mathbf{G}}_m = \dot{\mathbf{G}}_m(\hat{\boldsymbol{\rho}}) = f(\{\dot{a}_{m,ij}\}_{ij}) = f(\{\mathbb{1}[\hat{P}(\dot{a}_{m,ij} = 1|\hat{\boldsymbol{\rho}}; \mathbf{X}_m, \kappa(\mathcal{A}_m)) \geq \dot{u}_{m,ij}]\}_{ij})$, where $\dot{u}_{m,ij} \sim_{iid} U[0,1]$ and independent of $\boldsymbol{\varepsilon}_m$ (and similarly for $\ddot{\mathbf{G}}_m$ and $\dddot{\mathbf{G}}_m$), and $\mathbb{1}$ is the indicator function.

[15]The unconditional moment condition in Chandrasekhar and Lewis (2011) is based on the (Monte Carlo) integration of the moment condition $\mathbb{E}(\mathbf{Z}_m^{(s)\prime}\boldsymbol{\varepsilon}_m|\mathbf{G}_m)$ over $\mathbf{G}_m$.

This has two implications. First, it implies that our SGMM estimator is a two-stage estimator and therefore that the asymptotic variance-covariance matrix for $\boldsymbol{\theta}$ has to account for the first-stage sampling uncertainty. We show how to estimate the resulting asymptotic variance-covariance in the Online Appendix B.2.

Second, the indicator function $\mathbb{1}[\hat{P}(\dot{a}_{m,ij} = 1|\boldsymbol{\rho}; \mathbf{X}_m, \kappa(\mathcal{A}_m)) \geq \dot{u}_{m,ij}]$ implies that the objective function of the two-stage estimator is not everywhere continuous in $\boldsymbol{\rho}$. While this has a limited impact on consistency, it does complicate the proof of the asymptotic normality. Our proof builds on the argument in Andrews (1994), and we show that the stochastic equicontinuity condition holds using a bracketing argument.

Consistency and asymptotic normality also obviously depend on an identification condition. Here, the fact that the approximation of the bias $\boldsymbol{\delta}_m$ is non-linear in α implies that our SGMM is non-linear and that the identification condition cannot be simplified to a simple rank condition.

In Appendix B.1, we show that the objective function of our SGMM can be concentrated around α and that *conditional on* α, the identification conditions for $\tilde{\boldsymbol{\theta}}$ reduces to an asymptotic rank condition. Specifically, a sufficient condition for the identification of $\tilde{\boldsymbol{\theta}}$, conditional on α, is that the expected value of:

$$\frac{1}{RST}\sum_{rst}\dot{\mathbf{Z}}_m^{(r)\prime}(\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m^{(s)})(\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m^{(t)})^{-1}\dddot{\mathbf{V}}_m^{(t)}$$

converges in probability, to a full rank matrix for all α.[16] The last expression makes it clear that the non-linearity in our SGMM is sourced in the approximation of the asymptotic bias $\boldsymbol{\delta}_m = (\mathbf{I}_m - \alpha\ddot{\mathbf{G}}_m)(\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m)^{-1}\dddot{\mathbf{V}}_m\tilde{\boldsymbol{\theta}} - \ddot{\mathbf{V}}_m\tilde{\boldsymbol{\theta}}$.

[16]The identification of α then requires that the concentrated objective function is uniquely minimized.

When the matrix $\mathbf{G}_m$ is observed, we have $\dot{\mathbf{G}}_m^{(r)} = \ddot{\mathbf{G}}_m^{(s)} = \dddot{\mathbf{G}}_m^{(t)} = \mathbf{G}_m$ and the expression reduces to $\mathbf{Z}_m'\mathbf{V}_m$, (Bramoullé et al., 2009) which does not depend on α. This shows that the quality of $\hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}}, \mathbf{X}_m, \kappa(\mathcal{A}_m))$ has strong implications for identification since the dependence on α is weaker when the correlation between network draws is strong. We study the finite sample properties of our estimator using Monte Carlo simulation in Section 3.1.[17]

3.1 Monte Carlo Simulations

In this subsection, we study the performance of our SGMM estimator using Monte Carlo simulations. We consider cases where links are missing at random (see Example 1) and misclassified at random (see Example 3). The simulated individual characteristics (i.e., the matrix $\mathbf{X}$) include two characteristics similar to "age" and "female" in our empirical application.[18] The network formation process follows a logistic regression model:

$$P(a_{ij} = 1|\mathbf{X}) = \frac{\exp\{\rho_1 + \rho_2|x_{i1} - x_{j1}| + \rho_3\mathbb{1}\{x_{i2} = x_{j2}\}\}}{1 + \exp\{\rho_1 + \rho_2|x_{i1} - x_{j1}| + \rho_3\mathbb{1}\{x_{i2} = x_{j2}\}\}},$$

where x_{i1} represents "age" and x_{i2} represents "female".[19]

We analyze different proportions of randomly missing and misclassified entries in the network matrix. Figure 1 presents estimates for the endogenous peer effect coefficient α using our SGMM estimator. The left panel shows the peer effect estimates for the case of

[17] Theorem 1 assumes that the partial observability of $\mathbf{A}_m$ implies that $\mathbf{G}_m\mathbf{X}_m$ and $\mathbf{G}_m\mathbf{y}_m$ are both unobserved. However, in some cases, researchers can separately observe these quantities from survey questions. For example, one could simply obtain $\mathbf{G}_m\mathbf{y}_m$ from a question of the type: "What is the average value of your friends' y?" In these cases, it is possible to improve on our SGMM estimator by using this additional information. The resulting estimators are presented in Corollary 1 and Corollary 2 of the Online Appendix C.

[18] See Section 5. We simulate those variables from their empirical distributions in our sample. Parameter values are set to the estimates from our application: $(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma}) = (0.538, 3.806, -0.072, 0.132, 0.086, -0.003)$. We assume that ε is iid normally distributed with standard deviation of $\sigma = 0.707$.

[19] The parameter vector $\boldsymbol{\rho}$ is also set to its empirically estimated values: $\boldsymbol{\rho} = (-2.349, -0.700, 0.404)$.

missing links, while the right panel displays the estimates for the case of misclassified links.[20] Additionally, we report estimates obtained using the standard IV estimator of Bramoullé et al. (2009), treating the observed network with missing values or misclassified links as the true network.

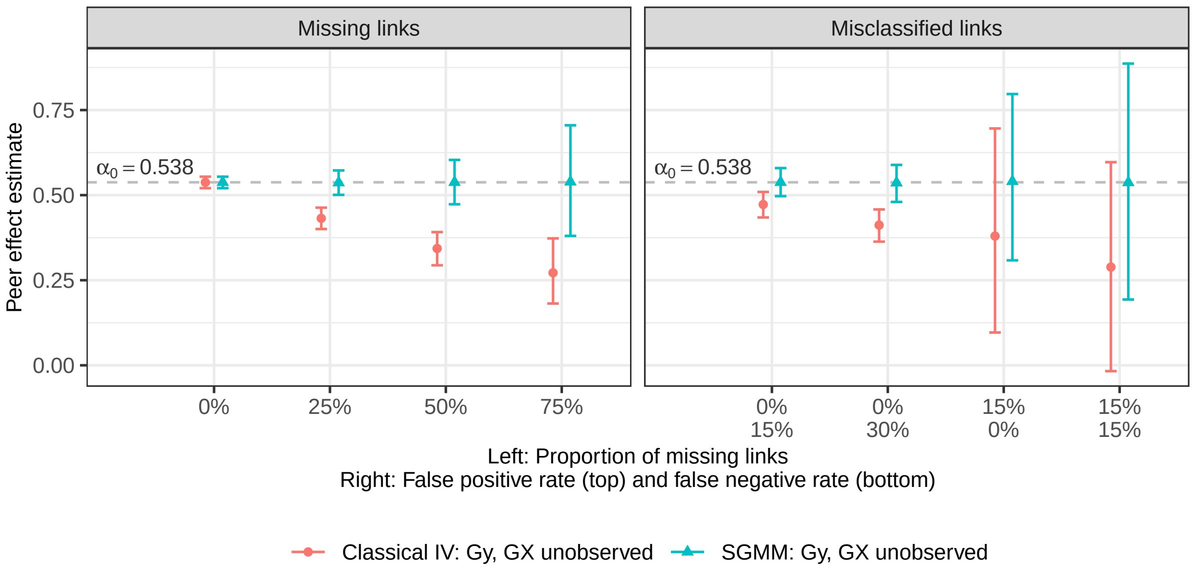

Figure 1: Estimated peer effects under missing links

Note: Dots represent the average estimated values of α, and bars indicate 95% confidence intervals. Tables E.1–E.4 in Online Appendix E provide the full set of estimated coefficients. The "Classical IV" refers to the standard estimator of Bramoullé et al. (2009). We simulate data for 100 groups of 30 individuals each. We assume that ε_i follows a normal distribution. We estimate $\boldsymbol{\rho}$ using a logit model based on the observed network entries (left panel) and a logit model with misclassification (right panel). The resulting estimates allow us to construct the network distribution (see Definition 1) and subsequently compute our SGMM estimator.

For the case of missing links, the estimates are centered around the true value. Although precision decreases as the fraction of missing links increases, our SGMM estimator maintains a reasonable level of accuracy, even when *half* of the links are missing. In contrast, the standard IV estimator significantly underestimates the peer effect coefficient α.

For the case of misclassified links, the estimator performs well when there are false nega-

[20]Tables E.1–E.4 in Online Appendix E provide the full set of estimated coefficients, including results that control for unobserved group heterogeneity through fixed effects.

tives only. Precision is affected when there are false positives, although the estimates remain centered around the true value. With false positives, the estimator for $\boldsymbol{\rho}$ loses precision since the network is simulated to match the one in our application: the density of the network is low.[21] With few links, the finite sample cost of false positives is thus more important.

4 Bayesian Estimator

In this section, we present a likelihood-based estimator. Accordingly, greater structure must be imposed on the errors $\boldsymbol{\varepsilon}_m$. Specifically, given parametric assumptions for $\boldsymbol{\varepsilon}_m$, one can write the log-likelihood of the outcome as:

$$\ln \mathcal{P}(\mathbf{y}|\mathbf{A}, \mathbf{X}, \boldsymbol{\theta}) = \sum_m \ln \mathcal{P}(\mathbf{y}_m|\mathbf{A}_m, \mathbf{X}_m; \boldsymbol{\theta}), \tag{5}$$

where notation without the index m denotes vectors and matrices at the sample level. We abuse notation by letting $\boldsymbol{\theta} = [\alpha, \boldsymbol{\beta}', \boldsymbol{\gamma}', \boldsymbol{\sigma}']'$, which now includes $\boldsymbol{\sigma}$, additional unknown parameters of the distribution of $\boldsymbol{\varepsilon}_m$. Recall that from Equation (1), we have: $\mathbf{y}_m = (\mathbf{I}_m - \alpha\mathbf{G}_m)^{-1}(c\mathbf{1}_m + \mathbf{X}_m\boldsymbol{\beta} + \mathbf{G}_m\mathbf{X}_m\boldsymbol{\gamma} + \boldsymbol{\varepsilon}_m)$ since $(\mathbf{I}_m - \alpha\mathbf{G}_m)^{-1}$ exists under our Assumption 3.

If the adjacency matrix $\mathbf{A}_m$ is observed, then $\boldsymbol{\theta}$ could be estimated using a simple maximum likelihood estimator (as in Lee et al., 2010) or using Bayesian inference (as in Goldsmith-Pinkham and Imbens, 2013). See in particular the identification conditions presented in Lee (2004) and Lee et al. (2010). Since $\mathbf{A}_m$ is not observed, but that $\mathcal{A}_m$ is observed, we focus on the following alternative likelihood:

$$\ln \mathcal{P}(\mathbf{y}|\mathcal{A}, \mathbf{X}; \boldsymbol{\theta}, \boldsymbol{\rho}) = \sum_m \ln \sum_{\mathbf{A}_m} \mathcal{P}(\mathbf{y}_m|\mathbf{A}_m, \boldsymbol{\theta}) P(\mathbf{A}_m|\boldsymbol{\rho}, \mathbf{X}_m, \mathcal{A}_m).$$

[21] This is typical of most network data: two randomly selected individuals are unlikely to be linked, even conditional on observables.

That is, we integrate the likelihood using posterior distribution obtained from the network formation model in Equation (2) after observing $\mathcal{A}_m$.[22]

One particular issue with estimating $\ln \mathcal{P}(\mathbf{y}|\mathcal{A}, \mathbf{X}; \boldsymbol{\theta}, \boldsymbol{\rho})$ is that the summations over the set of all possible network structures $\mathbf{A}_m$, for each group m is not tractable. Indeed, for a group of size N_m, the sum is over the set of possible adjacency matrices, which contain $2^{N_m(N_m-1)}$ elements. Then, simply simulating networks from $P(\mathbf{A}_m|\boldsymbol{\rho}, \mathbf{X}_m, \mathcal{A}_m)$ and taking the average likely lead to poor approximations. A classical way to address this issue is to use an EM algorithm (Hardy et al., 2024). Although valid, we found that the Bayesian estimator proposed in this section is less restrictive and numerically outperforms its classical counterpart. The Bayesian treatment also has the advantage of being valid in finite samples, allowing for a richer set of network formation models and partially observed network information $\mathcal{A}$.[23]

For concreteness, we will assume that $\boldsymbol{\varepsilon}_m \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m)$ for all m; however, it should be noted that our approach is valid for several alternative assumptions as long as it yields a computationally tractable likelihood. For each group m, and recalling that $\mathbf{G}_m = f(\mathbf{A}_m)$, we have:

$$\begin{aligned}\ln \mathcal{P}(\mathbf{y}_m|\mathbf{A}_m, \boldsymbol{\theta}) \quad = \quad & -N_m \ln(\sigma) + \ln|\mathbf{I}_m - \alpha\mathbf{G}_m| - \frac{N_m}{2}\ln(\pi) \\ & -\frac{1}{2\sigma^2}[(\mathbf{I}_m - \alpha\mathbf{G}_m)\mathbf{y}_m - c\mathbf{1}_m - \mathbf{X}_m\boldsymbol{\beta} - \mathbf{G}_m\mathbf{X}_m\boldsymbol{\gamma}]' \cdot \\ & [(\mathbf{I}_m - \alpha\mathbf{G}_m)\mathbf{y}_m - c\mathbf{1}_m - \mathbf{X}_m\boldsymbol{\beta} - \mathbf{G}_m\mathbf{X}_m\boldsymbol{\gamma}].\end{aligned}$$

Because $\mathbf{A}_m$ is not observed, we follow Tanner and Wong (1987), and we use data augmenta-

[22] See Equation (3). Note also that, conceptually, we could condition on $\kappa(\mathcal{A})$ instead of $\mathcal{A}$ as in Section 3. However, this is much less attractive from a Bayesian perspective and thus limit ourselves to this (more efficient) case.

[23] For example, models estimated using Aggregated Relational Data, see the Online Appendix H.

tion to evaluate the posterior distribution of $\boldsymbol{\theta}$. That is, instead of focusing on the posterior distribution of $\boldsymbol{\theta}$ (i.e., $P(\boldsymbol{\theta}|\mathbf{y},\mathbf{A},\mathbf{X})$) in the case in which the network was observed, we focus instead on the posterior distribution $P(\boldsymbol{\theta},\mathbf{A}|\mathbf{y},\mathcal{A},\mathbf{X})$, treating $\mathbf{A}$ as another set of unknown parameters.

Since the number of parameters to be estimated is larger than the number of observations,[24] the identification of the model rests on the a priori information on $\mathbf{A}$. A sensible prior for $\mathbf{A}$ is the consistent estimator of its distribution, i.e., $\Pi_m \hat{P}(\mathbf{A}_m|\hat{\boldsymbol{\rho}},\mathbf{X}_m,\mathcal{A}_m)$. Let $\pi(\boldsymbol{\rho}|\mathbf{X},\mathcal{A})$ be the prior density on $\boldsymbol{\rho}$. How to obtain $\pi(\boldsymbol{\rho}|\mathbf{X},\mathcal{A})$, depending on whether $\hat{\boldsymbol{\rho}}$ is obtained using a Bayesian or classical setting, is discussed in Examples 4 and 5 of the Online Appendix F.3. Given $\pi(\boldsymbol{\rho}|\mathbf{X},\mathcal{A})$, it is possible to obtain draws from the posterior distribution $P(\boldsymbol{\theta},\mathbf{A},\boldsymbol{\rho}|\mathbf{y},\mathcal{A})$ using the following Metropolis-Hastings MCMC:[25]

Algorithm 1. *The MCMC goes as follows for $t = 1,...,T$, starting from any $\mathbf{A}_0,\boldsymbol{\theta}_0$, and $\boldsymbol{\rho}_0$.*

1. *Draw $\boldsymbol{\rho}^*$ from the proposal distribution $q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})$ and accept $\boldsymbol{\rho}^*$ with probability*

$$\min\left\{1,\frac{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}^*,\mathcal{A})q_\rho(\boldsymbol{\rho}_{t-1}|\boldsymbol{\rho}^*)\pi(\boldsymbol{\rho}^*|\mathcal{A})}{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}_{t-1},\mathcal{A})q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})\pi(\boldsymbol{\rho}_{t-1}|\mathcal{A})}\right\}.$$

2. *Propose $\mathbf{A}^*$ from the proposal distribution $q_A(\mathbf{A}^*|\mathbf{A}_{t-1})$ and accept $\mathbf{A}^*$ with probability*

$$\min\left\{1,\frac{\mathcal{P}(\mathbf{y}|\boldsymbol{\theta}_{t-1},\mathbf{A}^*)q_A(\mathbf{A}_{t-1}|\mathbf{A}^*)P(\mathbf{A}^*|\boldsymbol{\rho}_{t-1},\mathcal{A})}{\mathcal{P}(\mathbf{y}|\boldsymbol{\theta}_{t-1},\mathbf{A}_{t-1})q_A(\mathbf{A}^*|\mathbf{A}_{t-1})P(\mathbf{A}_{t-1}|\boldsymbol{\rho}_{t-1},\mathcal{A})}\right\}.$$

[24] Each group contains N_m observations while the dimension of $\mathbf{A}_m$ is $N_m(N_m-1)$.

[25] As customary, for the rest of this section, we omit the dependence on $\mathbf{X}$ to lighten the notation. The notation with the index $t-1$ in this section refers to the $(t-1)$-th iteration of the MCMC, not the $(t-1)$-th group. Specifically, $\mathbf{A}_{t-1}$ denotes the adjacency matrix at the sample level in iteration $t-1$. Since the MCMC is a Metropolis-Hastings, the detailed balance and ergodicity conditions hold so the MCMC converges to $P(\boldsymbol{\theta},\mathbf{A},\boldsymbol{\rho}|\mathbf{y},\mathcal{A})$. See Cameron and Trivedi (2005), Section 13.5.4 for more details.

3. Draw α^* from the proposal $q_\alpha(\cdot|\alpha_{t-1})$ and accept α^* with probability

$$\min\left\{1, \frac{\mathcal{P}(\mathbf{y}|\mathbf{A}_t;\boldsymbol{\beta}_{t-1},\boldsymbol{\gamma}_{t-1},\alpha^*)q_\alpha(\alpha_{t-1}|\alpha^*)\pi(\alpha^*)}{\mathcal{P}(\mathbf{y}|\mathbf{A}_t;\boldsymbol{\theta}_{t-1})q_\alpha(\alpha^*|\alpha_{t-1})\pi(\alpha_{t-1})}\right\}.$$

4. Draw $[\boldsymbol{\beta},\boldsymbol{\gamma},\sigma]$ from their posterior conditional distributions (see Online Appendix F).

Step 1 allows to refine the estimation of $\boldsymbol{\rho}$. Indeed, in the first stage, $\boldsymbol{\rho}$ is inferred using the information provided by $\mathcal{A}$. In Step 1, however, $\boldsymbol{\rho}$ is updated conditional on $\mathcal{A}$ *and* $\mathbf{A}_{t-1}$. This provides additional information not available in the first stage since $\mathbf{A}_{t-1}$ uses information provided by the likelihood function (5).

Steps 3 and 4 are standard and detailed distributions can be found in the Online Appendix F. Step 2, however, requires some discussion. Indeed, the idea is the following: given the prior information $P(\mathbf{A}|\boldsymbol{\rho}_{t-1},\mathcal{A})$, one must be able to draw samples from the posterior distribution of $\mathbf{A}$, given $\mathbf{y}$. This is not a trivial task.

In particular, there is no general rule for selecting the network proposal distribution $q_A(\cdot|\cdot)$. A natural candidate is a Gibbs sampling algorithm for each link, i.e., change only one link ij at every step t and propose a_{ij} according to its marginal distribution, i.e., $a_{ij} \sim P(\cdot|\mathbf{A}_{-ij},\mathbf{y},\mathcal{A})$, where $\mathbf{A}_{-ij} = \{a_{kl}; k \neq i, l \neq j\}$. In this case, the proposal is always accepted.

However, it has been argued that Gibbs sampling could lead to slow convergence (e.g., Snijders, 2002; Chatterjee et al., 2013), especially when the network is *sparse* or exhibits a high level of *clustering*. For example, Mele (2017) and Bhamidi et al. (2008) propose different blocking techniques meant to improve convergence.

Here, however, achieving Step 2 involves an additional computational issue because evaluating the likelihood ratio in Step 1 requires comparing the determinants $|\mathbf{I} - \alpha f(\mathbf{A}^*)|$ for each proposed $\mathbf{A}^*$, which is computationally intensive.

Then, the appropriate blocking technique depends strongly on $P(\mathbf{A}|\boldsymbol{\rho}_{t-1}, \mathcal{A})$ and the assumed distribution for $\boldsymbol{\varepsilon}$. For the simulations and estimations presented in this paper, we use the Gibbs sampling algorithm for each link, adapting the strategy proposed by Hsieh et al. (2019) to our setting (see Proposition 3 in the Online Appendix F.2). This can be viewed as a *worst-case* scenario. Nonetheless, the Gibbs sampler performs reasonably well in practice however, we encourage researchers to try other updating schemes if Gibbs sampling performs poorly in their specific contexts. In particular, we present a blocking technique in the Online Appendix F that is also implemented in our R package `PartialNetwork`.[26]

Finally, note that for simple network formation models, it is possible to jointly estimate $\boldsymbol{\rho}$ and $\boldsymbol{\theta}$ within the same MCMC instead of using the two-step procedure described above. In this case, Step 1 can simply be replaced by:

1'. Draw $\boldsymbol{\rho}^*$ from the proposal distribution $q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})$ and accept $\boldsymbol{\rho}^*$ with probability

$$\min\left\{1, \frac{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}^*, \mathcal{A})P(\mathcal{A}|\boldsymbol{\rho}^*)q_\rho(\boldsymbol{\rho}_{t-1}|\boldsymbol{\rho}^*)\pi(\boldsymbol{\rho}^*)}{P(\mathbf{A}_{t-1}|\boldsymbol{\rho}_{t-1}, \mathcal{A})P(\mathcal{A}|\boldsymbol{\rho}_{t-1})q_\rho(\boldsymbol{\rho}^*|\boldsymbol{\rho}_{t-1})\pi(\boldsymbol{\rho}_{t-1})}\right\}.$$

Here, $P(\mathcal{A}|\boldsymbol{\rho}^*)$ is the likelihood of the network information $\mathcal{A}$ assuming the network formation model in (2). Note that $\pi(\boldsymbol{\rho})$, the prior density on $\boldsymbol{\rho}$, no longer depends on $\mathcal{A}$ and can be chosen arbitrarily (e.g., uniform).

[26] The complexity of Step 2 is not limited to our Bayesian approach. Classical estimators, such as GMM estimators, face a similar challenge in requiring the integration over the entire set of networks. The strategy used here is to rely on a Metropolis-Hastings algorithm, a strategy that has also been successfully implemented in the related literature on ERGMs (e.g., Snijders, 2002; Mele, 2017, 2020; Badev, 2021; Hsieh et al., 2019).

5 Application

In this section, we assume that the econometrician has access to network data but that the data may contain errors due to both *sampling* (links coded with errors) and *censoring*. To show how our method can be used to address these issues, we consider a simple example where we are interested in estimating peer effects on adolescents' academic achievements.

We use the widely used AddHealth database and show that network data errors have a first-order impact on the estimated peer effects. Specifically, we focus on a subset of schools from the Wave I "In School" sample that have less than 200 students (33 schools). Table G.1 in the Online Appendix G.3 presents the summary statistics.

Most papers estimating peer effects that use this particular database have taken the network structure as given. One notable exception is Griffith (2022), looking at censoring: students can only report up to five male and five female friends. We also allow for censoring but show that censoring is not the most important issue with the Add Health data. To understand why, we discuss the organization of the data.

Each adolescent is assigned a unique identifier. The data includes ten variables for the ten potential friendships (maximum of five male and five female friends). These variables can contain missing values (no friendship was reported), an error code (the named friend could not be found in the database), or an identifier for the reported friends. These data are then used to generate the network's adjacency matrix $\mathbf{A}$.

Of course, error codes cannot be matched to any particular adolescent. Moreover, even in the case where the friendship variable refers to a valid identifier, the referred adolescent may still be absent from the database. A prime example is when the referred adolescent has been removed from the database by the researcher, perhaps because of other missing variables for these particular individuals. These missing links are quantitatively important as they account for roughly 45% of the total number of links (7,830 missing for 10,163 observed links). Figure 2 displays the distribution of the number of "unmatched named friends."[27]

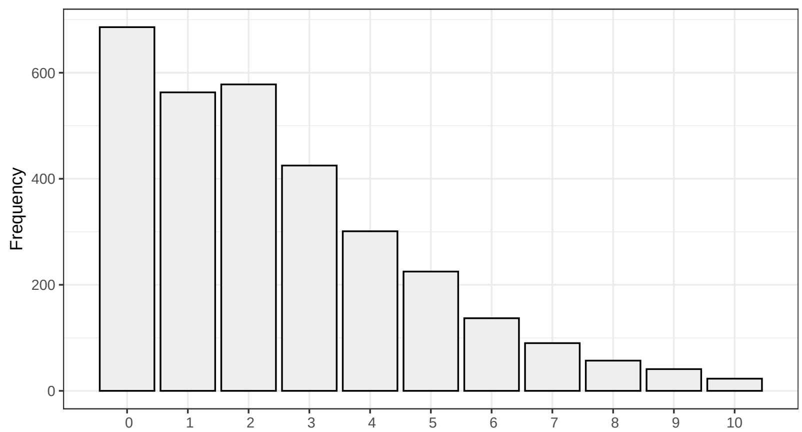

Figure 2: Frequencies of the number of missing links per adolescent

To use the methodology developed in sections 3 and 4, we first need to estimate a network formation model using the observed network data. In this section, we assume that links are generated using a simple logistic framework, i.e.,

$$P(a_{ij,m} = 1) = \frac{\exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}}{1 + \exp\{\mathbf{w}_{ij,m}\boldsymbol{\rho}\}},$$

where $\mathbf{w}_{ij,m}$ is built to capture homophily on the observed characteristics of i and j (see Tables G.2 and G.3 in the Online Appendix G.3).

We estimate the network formation model on the set of individuals for which we observe no "unmatched friends." For these students, we know for sure that their friendship data are complete. However, even under a missing at-random assumption, the estimation of $\boldsymbol{\rho}$ on

[27]We focus on within-school friendships; thus, nominations outside of school are not treated as "unmatched friends." Note also that these data errors could be viewed as a special case of censoring (Griffith, 2022) in which researchers know exactly how many links are censored. The attenuation bias is thus expected.

this subsample is affected by a selection bias: individuals with more friends have a higher probability of being censored, or of having a friendship nomination coded with error.[28]

We control for this selection bias by weighting the log-likelihood of the network following Manski and Lerman (1977). The details are presented in the Online Appendix G.1 and Online Appendix G.2. Intuitively, individuals in our restricted sample have fewer links. Therefore, the likelihood of $a_{i,j}$ when i is selected in our restricted sample is weighted by the inverse selection probability. When accounting for missing data due to error codes only, we estimate the selection probability for an individual i who declares n_i friends as the proportion of individuals without missing network data who declare n_i friends.

We use the same approach when controlling for missing data due to both error codes and censoring. However, in this case, the individual's censored number of friends has to be replaced with the (unobserved) true number of friends. We estimate individuals' true number of friends using a censored Poisson regression, where the observed number of friends in the network is used as the censored dependent variable: the variable is censored when individual i nominates five male friends or five female friends.

We present the estimation results for the SGMM and Bayesian estimator. Figure 3 summarizes the results for the endogenous peer effect coefficient α, whereas the full set of results is presented in the Online Appendix G.3. The first two estimations (*Obsv.Bayes* and *Obsv.SGMM*) assume that the observed network is the true network for both estimators. The third and fourth estimations (*Miss.Bayes* and *Miss.SGMM*) account for missing data due to error codes but not for censoring. The last two estimations (*TopMiss.Bayes* and

[28] Note that this is different from the random sampling discussed in our Example 1 and closer to the misclassification in Example 3, with only false-negative type of errors.

TopMiss.SGMM) account for missing data due to error codes and censoring.

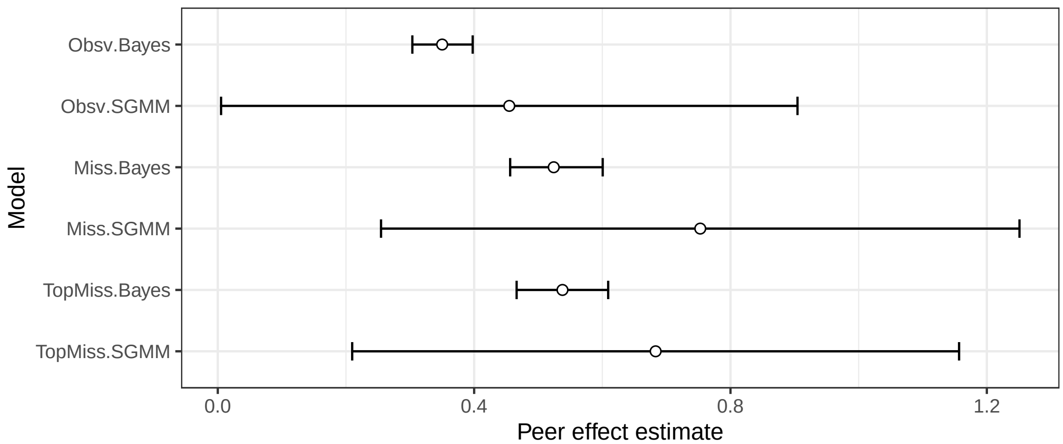

Figure 3: Peer effect estimate

Note: Dots represent estimated values (and posterior mean) of α, and bars represent 95% confidence intervals (and 95% credibility intervals). Tables G.2 and G.3 in Online Appendix G.3 present the full set of estimated coefficients.

We first see that the SGMM estimator is less efficient than the Bayesian estimator. This should not be surprising since the Bayesian estimator uses more structure (in particularity homoscedastic, normally distributed errors). When we compare the estimations *Obsv.SGMM* and *Miss.SGMM*, the observed differences imply that the efficiency loss is because of the relative inefficiency of the GMM approach, and not of the missing links or specifically of our SGMM estimator.[29]

Importantly, we see that the bias due to the assumption that the network is fully observed is quantitatively and qualitatively important. Using either estimator, the estimated endogenous peer effect using the reconstructed network is 1.5 times larger than that estimated assuming the observed network is the true network.[30] Almost all of the bias is produced by the presence of error codes and not because of potential censoring.

[29]Recall that when the network is observed, our SGMM uses the same moment conditions as, for example, those suggested by Bramoullé et al. (2009).

[30]The difference is "statistically significant" for the Bayesian estimator.

This exercise shows that data errors are a first-order concern when using the Add Health database. Not only does the bias in the endogenous peer effect coefficient α have an impact on the social multiplier (Glaeser et al., 2003), but it can also affect the anticipated effect of targeted interventions, i.e., the identity of the key player (Ballester et al., 2006). We include a more detailed discussion in Appendix G.4.

However, we would like to stress that we do not argue that our estimated coefficients are causal, because the friendship network is likely endogenous (e.g., Goldsmith-Pinkham and Imbens, 2013; Hsieh and Van Kippersluis, 2018; Hsieh et al., 2020). While previous literature has focused on the impact of network endogeneity, it has done so by assuming that the network is fully observed, despite the fact that roughly 45% of the links are missing. Above, we showed that errors in the observed network have a first-order impact on the estimated peer effect, even when one assumes that the network is exogenous.

6 Conclusion

In this paper, we propose two estimators for which peer effects can be estimated without observing the entire network structure. We find, perhaps surprisingly, that even very partial information on network structure is sufficient. By specifying a network formation model, researchers can probabilistically reconstruct the true network and base the estimation of peer effects on this reconstructed network. Importantly, we provide computationally tractable and flexible estimators to do so, all of which are available in our R package `PartialNetwork`. We apply our methodology to the widely used Add Health data and find that missing links due to noise in the data have first-order effects on the estimated peer effect coefficient.

References

Andrews, D. W. (1994): "Empirical process methods in econometrics," *Handbook of Econometrics*, 4, 2247–2294.

Badev, A. (2021): "Nash equilibria on (un) stable networks," *Econometrica*, 89, 1179–1206.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006): "Who's who in networks. Wanted: The key player," *Econometrica*, 74, 1403–1417.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013): "The diffusion of microfinance," *Science*, 341, 1236498.

Bhamidi, S., G. Bresler, and A. Sly (2008): "Mixing time of exponential random graphs," in *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, IEEE, 803–812.

Boucher, V. and I. Mourifié (2017): "My friend far, far away: a random field approach to exponential random graph models," *The Econometrics Journal*, 20, S14–S46.

Bramoullé, Y., S. Currarini, M. O. Jackson, P. Pin, and B. W. Rogers (2012): "Homophily and long-run integration in social networks," *Journal of Economic Theory*, 147, 1754–1786.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of Econometrics*, 150, 41–55.

——— (2020): "Peer effects in networks: A survey," *Annual Review of Economics*, 12, 603–629.

Breza, E. (2016): "Field experiments, social networks, and development," *The Oxford Handbook on the Economics of Networks*, 412–439.

Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2020): "Using aggregated relational data to feasibly identify network structure without network data," *American Economic Review*, 110, 2454–84.

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009): "Peer effects and social networks in education," *The Review of Economic Studies*, 76, 1239–1267.

Cameron, A. C. and P. K. Trivedi (2005): *Microeconometrics: methods and applications*, Cambridge University Press.

Chandrasekhar, A. and R. Lewis (2011): "Econometrics of sampled networks," *Unpublished manuscript, MIT.[422]*.

Chatterjee, S., P. Diaconis, et al. (2013): "Estimating and understanding exponential random graph models," *The Annals of Statistics*, 41, 2428–2461.

Chen, X., Y. Chen, and P. Xiao (2013): "The impact of sampling and network topology on the estimation of social intercorrelations," *Journal of Marketing Research*, 50, 95–110.

Conley, T. G. and C. R. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *American Economic Review*, 100, 35–69.

Currarini, S., M. O. Jackson, and P. Pin (2010): "Identifying the roles of race-based choice and chance in high school friendship network formation," *Proceedings of the National Academy of Sciences*, 107, 4857–4861.

De Paula, A. (2017): "Econometrics of network models," in *Advances in Economics and Econometrics: Theory and Applications: Eleventh World Congress (Econometric Society Monographs*, ed. by M. P. B. Honore, A. Pakes and L. Samuelson, Cambridge: Cambridge University Press, 268–323.

De Paula, A., I. Rasul, and P. C. Souza (2024): "Identifying network ties from panel data: Theory and an application to tax competition," *Review of Economic Studies*, rdae088.

De Paula, Á., S. Richards-Shubik, and E. Tamer (2018): "Identifying preferences in networks with bounded degree," *Econometrica*, 86, 263–288.

Glaeser, E. L., B. I. Sacerdote, and J. A. Scheinkman (2003): "The social multiplier," *Journal of the European Economic Association*, 1, 345–353.

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

Gourieroux, M., C. Gourieroux, A. Monfort, D. A. Monfort, et al. (1996): *Simulation-based econometric methods*, Oxford University Press.

Graham, B. S. (2017): "An econometric model of network formation with degree heterogeneity," *Econometrica*, 85, 1033–1063.

Griffith, A. (2022): "Name your friends, but only five? the importance of censoring in peer effects estimates using social network data," *Journal of Labor Economics*, 40, 779–805.

Griffith, A. and J. Kim (2023): "The Impact of Missing Links on Linear Reduced-form Network-Based Peer Effects Estimates," *Working Paper*.

Hardy, M., R. M. Heath, W. Lee, and T. H. McCormick (2024): "Estimating spillovers using imprecisely measured networks," *arXiv preprint arXiv:1904.00136*.

Hausman, J. A., J. Abrevaya, and F. M. Scott-Morton (1998): "Misclassification of the dependent variable in a discrete-response setting," *Journal of Econometrics*, 87, 239–269.

Hsieh, C.-S., Y.-C. Hsu, S. I. Ko, J. Kovářík, and T. D. Logan (2024): "Non-representative sampled networks: Estimation of network structural properties by weighting," Tech. Rep. 1.

Hsieh, C.-S., M. D. König, and X. Liu (2019): "A structural model for the coevolution of networks and behavior," *Review of Economics and Statistics*, 1–41.

Hsieh, C.-S., L.-F. Lee, and V. Boucher (2020): "Specification and estimation of network formation and network interaction models with the exponential probability distribution," *Quantitative Economics*, 11, 1349–1390.

Hsieh, C.-S. and H. Van Kippersluis (2018): "Smoking initiation: Peers and personality," *Quantitative Economics*, 9, 825–863.

Lee, L.-F. (2004): "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 72, 1899–1925.

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

Lewbel, A., X. Qu, and X. Tang (2022): "Social networks with unobserved links," *Journal of Political Economy, Forthcoming.*

——— (2024a): "Ignoring measurement errors in social networks," *The Econometrics Journal*, 27, 171–187.

——— (2024b): "stimating Social Network Models with Link Misclassification," *Working Paper.*

Liu, X. (2013): "Estimation of a local-aggregate network model with sampled networks," *Economics Letters*, 118, 243–246.

Liu, X., E. Patacchini, and E. Rainone (2017): "Peer effects in bedtime decisions among adolescents: a social network model with sampled data," *The Econometrics Journal*, 20, S103–S125.

Manresa, E. (2016): "Estimating the structure of social interactions using panel data," *Working paper.*

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

Manski, C. F. and S. R. Lerman (1977): "The estimation of choice probabilities from choice based samples," *Econometrica: Journal of the Econometric Society*, 1977–1988.

Mele, A. (2017): "A structural model of Dense Network Formation," *Econometrica*, 85, 825–850.

——— (2020): "Does school desegregation promote diverse interactions? An equilibrium model of segregation within schools," *American Economic Journal: Economic Policy*, 12, 228–57.

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

Snijders, T. A. (2002): "Markov chain Monte Carlo estimation of exponential random graph models," *Journal of Social Structure*, 3, 1–40.

Tanner, M. A. and W. H. Wong (1987): "The calculation of posterior distributions by data augmentation," *Journal of the American Statistical Association*, 82, 528–540.

Thirkettle, M. (2019): "Identification and estimation of network statistics with missing link data," *Working Paper*.

Van der Vaart, A. W. (2000): *Asymptotic Statistics*, vol. 3, Cambridge University Press.

Wang, W. and L.-F. Lee (2013): "Estimation of spatial autoregressive models with randomly missing data in the dependent variable," *The Econometrics Journal*, 16, 73–102.

Zhang, L. (2024): "Spillovers of program benefits with mismeasured networks," *arXiv preprint arXiv:2009.09614*.

A Appendix: Proof of Theorem 1

For the sake of clarity, we often write objects that depend on simulated networks as functions of $\boldsymbol{\rho}$; e.g., we write $\dot{\mathbf{Z}}_m(\boldsymbol{\rho})$ and $\dot{\mathbf{G}}_m(\boldsymbol{\rho})$ instead of $\dot{\mathbf{Z}}_m$ and $\dot{\mathbf{G}}_m$, unless this precision is unnecessary for the exposition. We define:

$$\mathbf{m}_{m,rst}(\boldsymbol{\theta},\boldsymbol{\rho}) = \dot{\mathbf{Z}}_m^{(r)\prime}(\boldsymbol{\rho})(\mathbf{I}-\alpha\ddot{\mathbf{G}}_m^{(s)}(\boldsymbol{\rho}))\left(\mathbf{y}_m - (\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m^{(t)}(\boldsymbol{\rho}))^{-1}\dddot{\mathbf{V}}_m^{(t)}(\boldsymbol{\rho})\tilde{\boldsymbol{\theta}}\right).$$

Let also $\mathbf{m}_m(\boldsymbol{\theta},\boldsymbol{\rho}) = \dfrac{1}{RST}\sum_{rst}\mathbf{m}_{m,rst}(\boldsymbol{\theta},\boldsymbol{\rho})$ and $\bar{\mathbf{m}}_M(\boldsymbol{\theta},\boldsymbol{\rho}) = \dfrac{1}{M}\sum_m \mathbf{m}_m(\boldsymbol{\theta},\boldsymbol{\rho})$. The objective function of the SGMM is given by:

$$\mathcal{Q}_M(\boldsymbol{\theta}) = [\bar{\mathbf{m}}_M(\boldsymbol{\theta},\hat{\boldsymbol{\rho}})]'\mathbf{W}_M[\bar{\mathbf{m}}_M(\boldsymbol{\theta},\hat{\boldsymbol{\rho}})],$$

where $\mathbf{W}_M$ is a weighing matrix. The SGMM estimator is $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}}\mathcal{Q}_M(\boldsymbol{\theta})$.

We impose the following regularity assumptions.

Assumption 6. *$\boldsymbol{\rho}_0$ and $\boldsymbol{\theta}_0$ are interior points of Θ and $\mathcal{R}$, respectively, where both Θ and $\mathcal{R}$ are compact subsets of the Euclidean space.*

Assumption 7. *(i) For all $m = 1,...,M$, $r = 1,...,R$, $s = 1,...,S$, and $t = 1,...,T$, $(\mathbf{I}_m - \alpha\mathbf{G}_m)$ and $(\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m^{(t)})$ are non-singular. (ii) The (i,j)-th entries of $\mathbf{G}_m$ (so $\dot{\mathbf{G}}_m^{(r)}$, $\ddot{\mathbf{G}}_m^{(s)}$, and $\dddot{\mathbf{G}}_m^{(t)}$), $(\mathbf{I}_m - \alpha\mathbf{G}_m)^{-1}$, and $(\mathbf{I}_m - \alpha\dddot{\mathbf{G}}_m^{(t)})^{-1}$ are bounded uniformly in i, j, and m.*

In particular, when $\mathbf{G}_m$ is row-normalized (so $\dot{\mathbf{G}}_m^{(r)}$, $\ddot{\mathbf{G}}_m^{(s)}$, and $\dddot{\mathbf{G}}^{(t)}$ are also row-normalized), Assumption 3 implies Assumption 7.

Assumption 8. *$\sup_{m\geq 1}\mathbb{E}\{\|\boldsymbol{\varepsilon}_m\|_2^{\mu}|\mathbf{X}_m,\mathcal{A}_m\}$ exists and is bounded, for some $\mu > 2$, where $\|.\|_2$ is the Euclidean norm.*

Assumption 9. *The derivative of $\hat{P}(a_{ij,m}|\boldsymbol{\rho},\mathbf{X}_m,\kappa(\mathcal{A}_m))$ with respect to $\boldsymbol{\rho}$ is bounded uniformly in i, j, and m.*

Assumption 10. $\mathbf{W}_M$ *is positive definite and* $\text{plim}\,\mathbf{W}_M = \mathbf{W}_0$*, where* plim *denotes the probability limit as* M *goes to infinity and* $\mathbf{W}_0$ *is a non-stochastic and positive definite matrix.*

Assumption 11 (Identification)**.** *For any* $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$*,* $\lim \mathbb{E}(\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)) \neq \mathbf{0}$*, where* lim *denotes the standard limit as* M *goes to infinity.*

While assumptions 6–10 are quite weak and standard, Assumption 11 is more substantial in nature. We discuss identification in more detail in Section B.1.

The proof of Theorem 1 proceeds as follows. In Section A.1, we show that the estimator is consistent. In Section A.2, we show that the estimator is asymptotically normal.

A.1 Proof of the consistency of the SGMM

We proceed to show that Theorem 2.1 in Newey and McFadden (1994) applies to our SGMM estimator. The proof relies on the following Lemmatta.

Lemma 1 (Validity of the moment function)**.** *The moment condition is verified for* $(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)$*; that is,* $\mathbb{E}(\mathbf{m}_m(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)) = \mathbf{0}$ *for all* m*.*

Proof. Let us substitute $\mathbf{y}_m = (\mathbf{I}_m - \alpha_0 \mathbf{G}_m)^{-1}(\mathbf{V}_m \tilde{\boldsymbol{\theta}}_0 + \boldsymbol{\varepsilon}_m)$ in the moment function. We have

$$\begin{aligned}\mathbf{m}_{m,rst}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) = \quad & \dot{\mathbf{Z}}_m^{(r)\prime}(\boldsymbol{\rho}_0)(\mathbf{I}_m - \alpha_0 \ddot{\mathbf{G}}_m^{(s)}(\boldsymbol{\rho}_0)) \left[(\mathbf{I}_m - \alpha_0 \mathbf{G}_m)^{-1} \mathbf{V}_m \right. \\ & \left. -(\mathbf{I}_m - \alpha_0 \dddot{\mathbf{G}}_m^{(t)}(\boldsymbol{\rho}_0))^{-1} \dddot{\mathbf{V}}_m^{(t)}(\boldsymbol{\rho}_0) \right] \tilde{\boldsymbol{\theta}}_0 \\ + \quad & \dot{\mathbf{Z}}_m^{(r)\prime}(\boldsymbol{\rho}_0)(\mathbf{I}_m - \alpha_0 \ddot{\mathbf{G}}_m^{(s)}(\boldsymbol{\rho}_0))(\mathbf{I}_m - \alpha_0 \mathbf{G}_m)^{-1} \boldsymbol{\varepsilon}.\end{aligned} \tag{6}$$

Consider the last part first. We have, for any r and s:

$$\mathbb{E}\left(\dot{\mathbf{Z}}_m^{(r)\prime}(\boldsymbol{\rho}_0)(\mathbf{I}_m - \alpha_0 \ddot{\mathbf{G}}_m^{(s)}(\boldsymbol{\rho}_0))(\mathbf{I}_m - \alpha_0 \mathbf{G}_m)^{-1} \boldsymbol{\varepsilon}_m | \mathbf{X}_m, \kappa(\mathcal{A}_m) \right) = \mathbf{0},$$

from Assumption 4.

Consider now the first part. Since network draws are independent, we have:

$$\hat{\mathbb{E}}_m[\dot{\mathbf{Z}}_m^{(r)\prime}(\boldsymbol{\rho}_0)]\hat{\mathbb{E}}_m[(\mathbf{I}_m - \alpha_0\ddot{\mathbf{G}}_m^{(s)}(\boldsymbol{\rho}_0))]\Big(\mathbb{E}_m^{(0)}[(\mathbf{I}_m - \alpha_0\mathbf{G}_m)^{-1}\mathbf{V}_m] - \hat{\mathbb{E}}_m[(\mathbf{I}_m - \alpha_0\dddot{\mathbf{G}}_m^{(t)}(\boldsymbol{\rho}_0))^{-1}\dddot{\mathbf{V}}_m^{(t)}(\boldsymbol{\rho}_0)]\Big)\tilde{\boldsymbol{\theta}}_0,$$

where $\hat{\mathbb{E}}_m$ denotes the expectation with respect to the distribution of the simulated networks, conditional on $\mathbf{X}_m, \kappa(\mathcal{A}_m)$, and where $\mathbb{E}_m^{(0)}$ is the expectation with respect to the distribution of the true network $\mathbf{G}_m$, conditional on $\mathbf{X}_m, \kappa(\mathcal{A}_m)$. Since, at $\boldsymbol{\rho}_0$, these are the same distributions, the terms in the big parenthesis cancel out, and thus, $\mathbb{E}[\mathbf{m}_{m,rst}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)|\mathbf{X}_m, \kappa(\mathcal{A}_m)] = \mathbf{0}$. As a result, $\mathbb{E}[\mathbf{m}_m(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)] = \frac{1}{RST}\sum_{rst}\mathbb{E}[\mathbf{m}_{m,rst}(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)] = \mathbf{0}$ by the law of iterated expectations. □

Lemma 2 (Differentiability)**.** *$\mathbb{E}[\mathbf{m}_m(\boldsymbol{\theta}, \boldsymbol{\rho})]$ is continuously differentiable in $(\boldsymbol{\theta}, \boldsymbol{\rho})$.*

Proof. See the Online Appendix B. □

Lemma 3 (Uniform convergence)**.** *We establish the following results.*

(a) $\mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})] - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ converges uniformly to $\mathbf{0}$ in $\boldsymbol{\theta}$ as $M \to \infty$.

(b) $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}}) - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})]$ converges uniformly in probability to $\mathbf{0}$ in $\boldsymbol{\theta}$ as $M \to \infty$.

(c) $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}}) - \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ converges uniformly in probability to $\mathbf{0}$ in $\boldsymbol{\theta}$ as $M \to \infty$.

Proof. See the Online Appendix B. □

The needed result from Lemma 3 is Statement (c), which allows us to replace $\hat{\boldsymbol{\rho}}$ with its limit $\boldsymbol{\rho}_0$ to show the consistency of $\hat{\boldsymbol{\theta}}$. However, this result is not trivial because the moment function is not continuous for all $\boldsymbol{\rho}$. We thus first show Statements (a) and (b), which together imply Statement (c).

Proof of Theorem 1

We define:

$$\mathcal{Q}_0(\boldsymbol{\theta}) = \Big[\lim \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]\Big]' \mathbf{W}_0 \Big[\lim \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]\Big].$$

As $\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \hat{\boldsymbol{\rho}})$ converges uniformly in probability to $\lim \mathbb{E}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}_0)]$ in $\boldsymbol{\theta}$ (Lemma 3, Statement (c)) and $\operatorname{plim} \mathbf{W}_M = \mathbf{W}_0$ (Assumption 10), by Cauchy-Schwartz (see e.g., Theorem 2.6 in Newey and McFadden, 1994), $\mathcal{Q}_M(\boldsymbol{\theta})$ converges uniformly in probability to $\mathcal{Q}_0(\boldsymbol{\theta})$.

From Theorem 2.1 in Newey and McFadden (1994), consistency of $\hat{\boldsymbol{\theta}}$ requires: (i) $\mathcal{Q}_0(\boldsymbol{\theta})$ is uniquely minimized at $\boldsymbol{\theta}_0$ (which holds from Lemma 1 and Assumption 11), (ii) the parameter space for $\boldsymbol{\theta}$ is compact (which holds by Assumption 6), (iii) $\mathcal{Q}_0(\boldsymbol{\theta})$ is continuous (which holds from Lemma 2), (iv) $\mathcal{Q}_M(\boldsymbol{\theta})$ converges uniformly in probability to $\mathcal{Q}_0(\boldsymbol{\theta})$ (which holds by Lemma 3 and Assumption 10 as pointed out above). Therefore, $\hat{\boldsymbol{\theta}}$ is consistent.

A.2 Proof of the Asymptotic Normality of the SGMM

We show that the SGMM estimator is asymptotically normally distributed. Recall that

$$\bar{\mathbf{m}}_M(\boldsymbol{\theta}, \boldsymbol{\rho}) = \frac{1}{M}\sum_m \bar{\mathbf{m}}_m(\boldsymbol{\theta}, \boldsymbol{\rho}),$$

and let

$$\bar{\mathbf{m}}^*_M(\boldsymbol{\theta}, \boldsymbol{\rho}) = \frac{1}{M}\sum_m \mathbb{E}\left(\bar{\mathbf{m}}_m(\boldsymbol{\theta}, \boldsymbol{\rho})\right).$$

Our proof relies on the following stochastic equicontinuity condition, which is formally shown in Lemma 4 in Online Appendix B:

CD1. $\sqrt{M}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}}) - \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0, \hat{\boldsymbol{\rho}})] - \sqrt{M}[\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) - \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)] = o_p(1)$

The first order condition of the empirical objective function $\mathcal{Q}_M$ with respect to $\boldsymbol{\theta}$ is

$\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\bar{\mathbf{m}}_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}}) = \mathbf{0}$. As $\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0) - \bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0) = \mathbf{0}$, this implies:

$$\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\left[\bar{\mathbf{m}}_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}}) - \bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0) + \bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0)\right] = \mathbf{0},$$

Given that $\bar{\mathbf{m}}_M(\boldsymbol{\theta},\hat{\boldsymbol{\rho}})$ is differentiable in $\boldsymbol{\theta}$, we replace $\bar{\mathbf{m}}_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})$ in the previous equation with its mean value expansion. After rearranging the terms we obtain:

$$\begin{aligned}&\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\frac{\partial \bar{\mathbf{m}}_M(\boldsymbol{\theta}^+,\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) = \\ &-\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\left[\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\hat{\boldsymbol{\rho}}) - \bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0) + \bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0)\right] = \mathbf{0},\end{aligned} \tag{7}$$

for some $\boldsymbol{\theta}^+$ lying between $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$.

From Lemma 2, $\bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\boldsymbol{\rho})$ is continuously differentiable in $\boldsymbol{\rho}$. Thus, it can be replaced by its mean value expansion, this time with respect to $\boldsymbol{\rho}$. We obtain:

$$\bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\hat{\boldsymbol{\rho}}) = \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0) + \frac{\partial \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}^+)}{\partial \boldsymbol{\rho}'}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0),$$

for some $\boldsymbol{\rho}^+$ lying between $\boldsymbol{\rho}_0$ and $\hat{\boldsymbol{\rho}}$. By premultiplying the last equation by $\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M$, we obtain:

$$\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\frac{\partial \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}^+)}{\partial \boldsymbol{\rho}'}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0) = \frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\left[\bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\hat{\boldsymbol{\rho}}) - \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0)\right].$$

By adding the previous equation to (7) and rearranging the terms, we have:

$$\begin{aligned}&\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\frac{\partial \bar{\mathbf{m}}_M(\boldsymbol{\theta}^+,\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) + \frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\frac{\partial \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}^+)}{\partial \boldsymbol{\rho}'}(\hat{\boldsymbol{\rho}}-\boldsymbol{\rho}_0) \\ = \; &-\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\left[\{\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\hat{\boldsymbol{\rho}}) - \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\hat{\boldsymbol{\rho}})\} - \{\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0) - \bar{\mathbf{m}}^*_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0)\}\right] \\ &-\frac{\partial \bar{\mathbf{m}}'_M(\hat{\boldsymbol{\theta}},\hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}}\mathbf{W}_M\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0,\boldsymbol{\rho}_0).\end{aligned} \tag{8}$$

As for the empirical moment in Lemma 3, $\frac{\partial \bar{\mathbf{m}}_M(\boldsymbol{\theta},\boldsymbol{\rho})}{\partial \boldsymbol{\theta}'}$ converges uniformly in $\boldsymbol{\theta}$ and $\boldsymbol{\rho}$ because it can be written as an average of independent elements that are differentiable with bounded derivatives. Thus, since $\text{plim}\,\hat{\boldsymbol{\theta}} = \text{plim}\,\boldsymbol{\theta}^{+} = \boldsymbol{\theta}_0$ and $\text{plim}\,\hat{\boldsymbol{\rho}} = \boldsymbol{\rho}_0$, we have:

$$\text{plim}\,\frac{\partial \bar{\mathbf{m}}_M(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}'} = \text{plim}\,\frac{\partial \bar{\mathbf{m}}_M(\boldsymbol{\theta}^{+}, \hat{\boldsymbol{\rho}})}{\partial \boldsymbol{\theta}'} = \text{plim}\,\frac{\partial \bar{\mathbf{m}}^{*}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\theta}'} \equiv \mathbf{H}_0. \tag{9}$$

As usual, we also impose the following assumption so that, under Assumption 10, the matrix $\mathbf{H}_0'\mathbf{W}_0\mathbf{H}_0$ is not singular.

Assumption 12. *The matrix $\mathbf{H}_0$ has full rank.*

Equation (8) implies that:

$$\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = -(\mathbf{H}_0'\mathbf{W}_0\mathbf{H}_0)^{-1}\mathbf{H}_0'\mathbf{W}_0\Big[\sqrt{M}\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) + \\ \frac{\partial \bar{\mathbf{m}}^{*}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}^{+})}{\partial \boldsymbol{\rho}'}\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)\Big] + o_p(1)$$

provided that the stochastic equicontinuity condition CD1 holds (see Lemma 4 in Online Appendix B).

Under Assumption 5, $\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$ converges in distribution to a $N(\mathbf{0}, \mathbf{V}_{\boldsymbol{\rho}})$, and

$$\text{plim}\,\frac{\partial \bar{\mathbf{m}}^{*}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}^{+})}{\partial \boldsymbol{\rho}'} \equiv \boldsymbol{\Gamma}_0,$$

exists by the uniform law or large numbers. Thus, $\dfrac{\partial \bar{\mathbf{m}}^{*}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}^{+})}{\partial \boldsymbol{\rho}'}\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$ converges in distribution to a $N(\mathbf{0}, \boldsymbol{\Gamma}_0\mathbf{V}_{\boldsymbol{\rho}}\boldsymbol{\Gamma}_0')$.

We now apply the Lyapunov CLT to $\sqrt{M}\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)$ since it is a normalized sum of independent elements. However, as we need the joint asymptotic distribution of $\sqrt{M}\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)$ and $\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$ to be normal, we apply the Lyapunov CLT conditional on $\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$. The Lyapunov condition (conditional $\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$) is verified by Assumption 8.[31] Thus the asymptotic distribution of $\sqrt{M}\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)$, conditional on $\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$ is normal, which implies that the joint asymptotic distribution of $\sqrt{M}\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0)$ and $\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$ is nor-

[31] See for example, Van der Vaart (2000), Section 23.4

mal. Consequently $\sqrt{M}\bar{\mathbf{m}}_M(\boldsymbol{\theta}_0, \boldsymbol{\rho}_0) + \frac{\partial \bar{\mathbf{m}}_M^*(\boldsymbol{\theta}_0, \boldsymbol{\rho}^+)}{\partial \boldsymbol{\rho}'}\sqrt{M}(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho}_0)$ is asymptotically normally distributed. As a result, $\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normally distributed.

Estimating the asymptotic variance of $\sqrt{M}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ requires an estimate of $\boldsymbol{\Gamma}_0$, which can be complex. In Online Appendix B.2, we present an approach to estimate this asymptotic variance without requiring an estimate of $\boldsymbol{\Gamma}_0$.