Identification and Inference of Network Formation Games with Misclassified Links

Luis E. Candelaria Takuya Ura

March 2020 No: 1258

Warwick Economics Research Papers

ISSN 2059-4283 (online)

ISSN 0083-7350 (print)

Identification and Inference of Network Formation Games with Misclassified Links[*]

Luis E. Candelaria[†] Takuya Ura[‡]

March 24, 2020

Abstract

This paper considers a network formation model when links are potentially measured with error. We focus on a game-theoretical model of strategic network formation with incomplete information, in which the linking decisions depend on agents' exogenous attributes and endogenous network characteristics. In the presence of link misclassification, we derive moment conditions that characterize the identified set for the preference parameters associated with homophily and network externalities. Based on the moment equality conditions, we provide an inference method that is asymptotically valid when a single network of many agents is observed. Finally, we apply our proposed method to study trust networks in rural villages in southern India.

Keywords: Misclassification, Network formation models, Strategic interactions, Incomplete information

JEL Codes: C13, C31

1 Introduction

Researchers across different disciplines have documented that measurement error of links is a pervasive problem in network data (e.g., Holland and Leinhardt 1973, Moffitt 2001, Kossinets 2006, Ammermueller and Pischke 2009, Wang, Shi, McFarland, and Leskovec 2012, Angrist 2014, de Paula 2017, Advani and Malde 2018). Although strategic network formation models provide essential information for learning about the creation of linking connections and peer effects when the network of interaction is endogenous, to the best of our knowledge, there has been no work addressing the effects of misclassifying links in strategic network formation models. In this paper, we consider identification and inference in a game-theoretical model of strategic network formation with potentially misclassified links.

We focus on a simultaneous game with imperfect information in which agents decide to form connections to maximize their expected utility (cf. Leung 2015, and Ridder and Sheng 2015). The agents' decisions are interdependent since the utility attached to creating a link depends on the agents' observed attributes and

[*]First version: April, 2017. We would like to thank Victor Aguirregabiria, Federico Bugni, Aureo de Paula, Bryan Graham, Shakeeb Khan, Michael Leung, Zhongjian Lin, Matt Masten, Arnaud Maurel, Konrad Menzel, Kenichi Nagasawa, Pedro Souza, Roland Rathelot, Eric Renault, Camilla Roncoroni, Adam Rosen, and seminar participants at Bristol Econometric Study Group, Warwick Summer Workshop, 2018 European Meeting of the Econometric Society at Naples, and 2019 Asian Meeting of the Econometric Society at Xiamen for very helpful comments. The usual disclaimer applies.

[†]Department of Economics, University of Warwick, Social Sciences Building, Coventry, CV4 7AL, U.K.; Email: L.Candelaria@warwick.ac.uk

[‡]Department of Economics, University of California, Davis, One Shields Avenue, Davis, CA 95616-5270; Email: takura@ucdavis.edu

network characteristics through link externalities (such as reciprocity, in-degree, and transitivity statistics). The misclassification problem will affect the link formation decisions in two different ways. First, the binary outcome variable representing an agent's optimal linking decision is misclassified. Second, the link misclassification problem prevents us from directly identifying the belief system that an agent uses to predict others' linking decisions. In this sense, the measurement error problem occurs on the left- and right-hand sides of the equation describing the optimal linking decisions (as shown in Lemma 1).

We propose a novel approach for analyzing network formation models, which is robust to misclassification. Specifically, in a setup that allows for the links to be potentially misclassified, we characterize the identified set for the structural parameters, including the preference parameters concerning homophily and network externalities. A notable innovation in our approach is that we derive the relationship between the choice probabilities of observed network connections and the belief system (as shown in Lemma 2). This result is crucial in allowing us to control for the endogeneity of the equilibrium beliefs and to reduce the model to a single-agent decision model in the presence of misclassification.

We also propose an inference method that is asymptotically valid when one network with a large number of agents is observed. Our proposed confidence interval is computationally feasible and controls the size even when the parameters are partially identified.

In an empirical illustration, we apply our inference method to examine trust networks in Karnataka, an area in southern India (see Banerjee, Chandrasekhar, Duflo, and Jackson 2013 and Jackson, Rodriguez-Barraquer, and Tan 2012). As a benchmark for our study, we consider the case when the network is assumed to be observed without measurement error, and we incorporate different degrees of misclassification in which some existing links are recorded as non-existent. This benchmark scenario corresponds to the analysis conducted by Leung (2015). Our results suggest that, even with misclassified links, the most important determinants driving the lending decisions of the individuals in the network are reciprocation, homophily on gender, and whether or not the individuals are relatives. For example, when we consider a scenario with up to 50% misclassification probability, the 95% confidence intervals proposed using our inference method are at most 50% larger in length than the confidence intervals that ignore the measurement error problem.

Our methodology contributes to the growing econometric literature studying the strategic formation of networks. Some recent surveys include Graham (2015), Chandrasekhar (2016), and de Paula (2017). The network formation model considered in this paper builds on the framework of strategic interactions with incomplete information introduced by Leung (2015) and extended by Ridder and Sheng (2015). The analysis in our paper addresses the problems arising due to link misclassification in their models.

This paper is also related to the literature of mismeasured discrete variables, e.g., misclassified binary outcome variable (Hausman, Abrevaya, and Scott-Morton, 1998), and misclassified discrete treatment variable (Mahajan, 2006; Lewbel, 2007; Chen, Hu, and Lewbel, 2008; Hu, 2008). Specifically, our approach to misclassified links is based on Molinari (2008), which offers a general bounding strategy with misclassified discrete variables.

There are a few papers in the literature of social interactions that have examined the presence of measurement error in network data (Chandrasekhar and Lewis, 2014; Kline, 2015; Lewbel, Norris, Pendakur, and Qu, 2017; de Paula, Rasul, and Souza, 2018; Hu and Lin, 2018; Lewbel, Qu, and Tang, 2019). However, the results in these papers cannot be applied directly to our framework since they have a different object of interest. In particular, they focus primarily on studying peer effects while taking as given an exogenous network of interactions, and they do not investigate directly the underlying process that drives the formation of links in the network. In contrast, our paper studies the effects of link misclassification in a model of strategic network formation.

The remainder of the paper is organized as follows. Section 2 describes the network formation model as a game of incomplete information. Section 3 characterizes the identified set of the structural parameters. Section 4 introduces an inference method based on the representation of the identified set. Sections 5 presents an empirical application using data on trust networks in rural villages in southern India. Section 6 provides concluding remarks. Appendix A collects all the proofs of the paper.

2 Network formation game with misclassification

We extend Leung (2015) and Ridder and Sheng (2015) to model a directed network with potentially misclassified links. In particular, we use a static game of incomplete information as a framework to model the formation of a directed network. For simplicity, our approach follows Leung (2015).

Consider a network determined by a set of n agents, which we denote by $\mathcal{N}_n = \{1, \ldots, n\}$. We assume that each pair of agents (i, j) with $i, j \in \mathcal{N}_n$ is endowed with a vector of exogenous attributes $X_{ij} \in \mathbb{R}^d$ and an idiosyncratic shock $\varepsilon_{ij} \in \mathbb{R}$. Let $X = \{X_{ij} : i, j \in \mathcal{N}_n\} \in \mathcal{X}^n$ be a profile of attributes that is common knowledge to all the agents in the network, $\varepsilon_i = \{\varepsilon_{ij} : j \in \mathcal{N}_n\}$ is a profile of idiosyncratic shocks that is agent i's private information, and $\varepsilon = \{\varepsilon_i : i \in \mathcal{N}_n\}$ collects all the profiles of idiosyncratic shocks.

The network is represented by an $n \times n$ adjacency matrix G_n^*, where the ijth element $G_{ij,n}^* = 1$ if agent i forms a direct link to agent j and $G_{ij,n}^* = 0$ otherwise. We assume that the network is directed, i.e., $G_{ij,n}^*$ and $G_{ji,n}^*$ may be different. The diagonal elements are normalized to be equal to zero, i.e., $G_{ii,n}^* = 0$. The network G_n^* is potentially misclassified and the researcher observes G_n, which is a proxy for G_n^*.

Given the network G_n^* and information (X, ε_i), agent i has utility

$$U_i(G_{i,n}^*, G_{-i,n}^*, X, \varepsilon_i) = \frac{1}{n}\sum_{j=1}^{n} G_{ij,n}^* \left[\left(G_{ji,n}^*, \frac{1}{n}\sum_{k\neq i} G_{kj,n}^*, \frac{1}{n}\sum_{k\neq i,j} G_{ki,n}^* G_{kj,n}^*, X_{ij}'\right)\beta_0 + \varepsilon_{ij}\right],$$

where $G_{i,n}^* = \{G_{ij,n}^* : j \in \mathcal{N}_n\}$, $G_{-i,n}^* = \{G_{j,n}^* : j \neq i\}$, and β_0 is an unknown finite dimensional vector in a parameter space $\mathcal{B}$.

Agent i's marginal utility of forming the link $G_{ij,n}^*$ depends on a vector of network statistics, the profile of exogenous attributes, and the link-specific idiosyncratic component.[1] The first component in the vector of network statistics captures the utility obtained from a reciprocated link with agent j, $G_{ji,n}^*$. The second network statistic represents the in-degree of agent j, $\frac{1}{n}\sum_{k\neq i} G_{kj,n}^*$, which captures the utility obtained from connecting with agents of high centrality in the network. The last network statistic captures the utility of being connected to the same agents, $\frac{1}{n}\sum_{k\neq i,j} G_{ki,n}^* G_{kj,n}^*$. The profile of exogenous attributes captures the preferences for homophily on observed characteristics. Finally, ε_{ij} is an unobserved link-specific component affecting agent i's decision whether to link with agent j.

Let $\delta_{i,n}(X, \varepsilon_i)$ denote a generic agent i's pure strategy, which maps the information available to agent i, (X, ε_i), to an action in $\mathcal{G}^n = \{0, 1\}^n$. Let $\sigma_{i,n}(g_{i,n}^* \mid X) = Pr(\delta_{i,n}(X, \varepsilon_i) = g_{i,n}^* \mid X)$ be the probability that agent i chooses $g_{i,n}^* \in \mathcal{G}^n$ given X, and $\sigma_n(X) = \{\sigma_{i,n}(g_{i,n}^* \mid X), i \in \mathcal{N}_n, g_{i,n}^* \in \mathcal{G}^n\}$. We call $\sigma_n(X)$ a belief profile. Given a belief profile σ_n and the information (X, ε_i), the agent i chooses $g_{i,n}^*$ from $\mathcal{G}^n$ to maximize the expected utility of $U_i(g_{i,n}^*, \delta_{-i,n}(X, \varepsilon_{-i}), X, \varepsilon_i)$ given $(X, \varepsilon_i, \sigma_n)$.

[1] For simplicity, we consider three different kinds of factors in the vector of network statistics. It is straightforward to generalize our results to a more flexible representation as in Leung (2015).

In an n-player game, a Bayesian Nash Equilibrium $\sigma_n(X)$ is a belief profile that satisfies the self-consistency condition

$$\sigma_{i,n}(g^*_{i,n} \mid X) = Pr(\delta_{i,n}(X, \varepsilon_i) = g^*_{i,n} \mid X, \sigma_n)$$

for all attribute profiles $X \in \mathcal{X}^n$, actions $g^*_{i,n} \in \mathcal{G}^n$, and agents $i \in \mathcal{N}_n$, where

$$\delta_{i,n}(X, \varepsilon_i) = \underset{g^*_{i,n} \in \mathcal{G}^n}{\arg\max}\, E\left[U_i(g^*_{i,n}, \delta_{-i,n}(X, \varepsilon_{-i}), X, \varepsilon_i) \mid X, \varepsilon_i, \sigma_n\right].$$

We impose the following assumptions, which are also used by Leung (2015) and Ridder and Sheng (2015).

Assumption 1. *The following holds for any n,*

(i) For any $A_1, A_2 \subset \mathcal{N}_n$ disjoint, $\{X_{ij} : i, j \in A_1\}$ and $\{X_{kl} : k, l \in A_2\}$ are independent.

(ii) $\{\varepsilon_{ij} : i, j \in \mathcal{N}_n\}$ are identically distributed with the standard normal distribution, cdf Φ, and pdf ϕ. Further, $\{\varepsilon_i : i \in \mathcal{N}_n\}$ are mutually independent.

(iii) ε and X are independent.

(iv) Attributes $\{X_{ij} : i, j \in \mathcal{N}_n\}$ are identically distributed with a probability mass function bounded away from zero.

Condition (i) allows for correlation across the pairs of attributes X_{ij} and X_{kl} if they have a common index (i.e., $i = k$). As a consequence, the attributes across all the dyads formed by one agent may be dependent. Condition (ii) assumes that the idiosyncratic shocks are identically distributed with known distribution. In Appendix B, we relax this assumption and consider a semi-parametric framework. This condition also implies that the components of ε_i may be arbitrarily correlated. Condition (iii) rules out the possibility of agents learning about others' private information from the observed profile of attributes that is common knowledge. Condition (iv) assumes that X_{ij} is a discrete random vector.

We focus on a symmetric equilibrium for our inference method. This approach is not restrictive in a setup where the identities of the individuals in the network are irrelevant. An equilibrium profile σ_n is symmetric if $\sigma_{i,n}(g^*_{i,n} \mid X) = \sigma_{\pi(i),n}(\pi(g^*_{\pi(i),n}) \mid \pi(X))$ for any $i \in \mathcal{N}_n$, $g^*_{i,n} \in \mathcal{G}^n$, and permutation function $\pi \in \Pi$.[2] Given Assumption 1, Leung (2015, Theorem 1) and Ridder and Sheng (2015, Proposition 1) show the existence of a symmetric equilibrium. The next assumption summarizes these results.

Assumption 2. *For any n, the agents play a symmetric equilibrium σ_n, i.e., there exists $\{\delta_{i,n} : i \in \mathcal{N}_n\}$ such that for any $i \in \mathcal{N}_n$ the following holds: (i) $G^*_{i,n} = \delta_{i,n}(X, \varepsilon_i)$, (ii) $\sigma_{i,n}(g^*_{i,n} \mid X) = Pr(\delta_{i,n}(X, \varepsilon_i) = g^*_{i,n} \mid X, \sigma_n)$, (iii) $\delta_{i,n}(X, \varepsilon_i) = \arg\max_{g^*_{i,n} \in \mathcal{G}^n} E\left[U_i(g^*_{i,n}, \delta_{-i,n}(X, \varepsilon_{-i}), X, \varepsilon_i) \mid X, \varepsilon_i, \sigma_n\right]$, and (iv) σ_n is symmetric.*

The next lemma characterizes the optimal decision rule for the formation of each link in the network.

[2]Define permutation functions as follows. Fix any $k, l \in \mathcal{N}_n$, and let $g^*_{i,n} \in \mathcal{G}^n$. Define $\pi_{kl} : \mathcal{N}_n \mapsto \mathcal{N}_n$ as a permutation of the indices k and l. Specifically, it maps the index k to the index l, l to k, and i to itself for any $i \neq k, l$. Define π^X_{kl} as a function that maps each component $X_{ij} \in \mathbb{R}^d$ to $X_{\pi_{kl}(i)\pi_{kl}(j)}$; π^a_{kl} as a function that permutes the kth and lth elements of any $g^*_{i,n} \in \mathcal{G}^n$. Hence, π^X_{kl} swaps the attributes of agents k and l; and π^a_{kl} swaps the links $G^*_{ik,n}$ and $G^*_{il,n}$ for any i. $\pi(\cdot)$ denotes a generic element of $\Pi = \left\{(\pi_{kl}, \pi^X_{kl}, \pi^a_{kl}); k, l \in \mathcal{N}_n\right\}$. In this paper, we abuse the notation $\pi(\cdot)$ so that it denotes any of the three components of an element in Π.

Lemma 1. *Under Assumption 1 and 2 ,* $G^*_{ij,n} = 1\left\{(Z^*_{ij,n})'\beta_0 + \varepsilon_{ij} \geq 0\right\}$, *where*

$$\gamma^*_{ij,n} = E\left[\left(G^*_{ji,n}, \frac{1}{n}\sum_{k\neq i} G^*_{kj,n}, \frac{1}{n}\sum_{k\neq i,j} G^*_{ki,n}G^*_{kj,n}\right)' \mid X, \sigma_n\right]$$

and

$$Z^*_{ij,n} = \begin{pmatrix} \gamma^*_{ij,n} \\ X_{ij} \end{pmatrix}.$$

A direct implication of Lemma 1 is that each agent makes separate linking decisions for each of her potential links. Notice that given the misclassification problem, both the optimal action $G^*_{ij,n}$ and the equilibrium beliefs about the network statistics $\gamma^*_{ij,n}$ in the optimal decision rule will be misclassified. In other words, the misclassification problem affects both left and right-hand side variables in the optimal decision rule.

We assume that the conditional distribution of the observed network G_n is related to that of the true state of the network, G^*_n, as follows.

Assumption 3. *There are two non-negative real numbers* ρ_0 *and* ρ_1 *with* $\rho_0 + \rho_1 < 1$ *such that the following two statements hold for every* n *and every* $i, j, k \in \mathcal{N}_n$. *(i)* $G_{ki,n}$ *and* $G_{kj,n}$ *are independent given* $(G^*_{ki,n}, G^*_{kj,n}, X, \sigma_n)$. *(ii)* $Pr(G_{ij,n} \neq G^*_{ij,n} \mid G^*_{ij,n}, X, \sigma_n) = \rho_0 1\{G^*_{ij,n} = 0\} + \rho_1 1\{G^*_{ij,n} = 1\}$.

Condition (i) in Assumption 3 requires that the observed linking decisions $G_{ki,n}$ and $G_{kj,n}$ are conditional independent given the true state of the links $G^*_{ki,n}$ and $G^*_{kj,n}$, and information X, σ_n.[3] This condition is required to relate the nonlinear endogenous factor $\frac{1}{n}\sum_{k\neq i,j} G^*_{ki,n}G^*_{kj,n}$ to other observed network features; Lemma 2 provides the exact statement. Condition (ii) in Assumption 3 characterizes the misclassification probabilities. Furthermore, it states that the misclassification probabilities are conditionally independent from the information X, σ_n.[4] Hausman et al. (1998) has also used Condition (ii) but in a setting of a binary choice model with misclassification of the dependent variable.

The following statement is a key observation in our analysis, which relates the observed network statistics $\gamma_{ij,n}$ to the payoff relevant network statistics $\gamma^*_{ij,n}$.

Lemma 2. *If Assumptions 1-3 hold, then* $\gamma^*_{ij,n} = c(\rho_0, \rho_1) + C(\rho_0, \rho_1)\gamma_{ij,n}$ *for every* i, j, *where*

$$\gamma_{ij,n} = E\left[\left(G_{ji,n}, \frac{1}{n}\sum_{k\neq i} G_{kj,n}, \frac{1}{n}\sum_{k\neq i,j} G_{ki,n}G_{kj,n}, \frac{1}{n}\sum_{k\neq i,j}(G_{ki,n} + G_{kj,n})\right)' \mid X, \sigma_n\right],$$

and, for any $r_0, r_1 \geq 0$ *such that* $r_0 + r_1 < 1$,

$$c(r_0, r_1) = -\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} 1 - r_0 - r_1 & 0 & 0 & 0 \\ 0 & 1 - r_0 - r_1 & 0 & 0 \\ 0 & 0 & (1 - r_0 - r_1)^2 & r_0(1 - r_0 - r_1) \\ 0 & 0 & 0 & 1 - r_0 - r_1 \end{pmatrix}^{-1}\begin{pmatrix} r_0 \\ r_0 \\ r_0^2 \\ r_0 \end{pmatrix}$$

[3] The independence assumption is imposed for simplicity. It is possible to remove this condition by assuming that $\{\varepsilon_{ij} : i, j \in \mathcal{N}_n\}$ are i.i.d. (as in Ridder and Sheng (2015)) or by modeling the relationship between $G_{ki,n}G_{kj,n}$ and $G^*_{ki,n}G^*_{kj,n}$.

[4] A specification of Condition (ii) that allows the misclassification probabilities ρ_0 and ρ_1 to depend on the observed attributes X_{ij} can be analyzed within our inference method. We consider the current setup to simplify the exposition.

$$C(r_0, r_1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1-r_0-r_1 & 0 & 0 & 0 \\ 0 & 1-r_0-r_1 & 0 & 0 \\ 0 & 0 & (1-r_0-r_1)^2 & r_0(1-r_0-r_1) \\ 0 & 0 & 0 & 1-r_0-r_1 \end{pmatrix}^{-1}.$$

Notice that the first three components in $\gamma_{ij,n}$ are the observed analog to the statistics in $\gamma^*_{ij,n}$ since they are determined by the observed network G_n. The last component in $\gamma_{ij,n}$ is the sum of the in-degrees of agents i and j, and it is the result of controlling for the unobserved network statistics $\frac{1}{n}\sum_{k\neq i,j} G^*_{ki,n}G^*_{kj,n}$. To be precise, the last two statistics in $\gamma_{ij,n}$ control for the beliefs about the unobserved network statistic $\frac{1}{n}\sum_{k\neq i,j} G^*_{ki,n}G^*_{kj,n}$, which is the only nonlinear endogenous factor. The intuition behind this result is similar to the one found in polynomial regression models with mismeasured continuous covariates (Hausman, Newey, Ichimura, and Powell, 1991).

Assumptions 1-3 imply the following relationship between the distributions of $G_{ij,n}$ and $G^*_{ij,n}$, which will be used in our identification analysis. Since we observe $G_{ij,n}$ in the dataset but the outcome of interest is $G^*_{ij,n}$, it is crucial to connect these two objects.

Lemma 3. *Under Assumptions 1-3,* $Pr(G_{ij,n} = 1 \mid X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n}) = \rho_0 Pr(G^*_{ij,n} = 0 \mid X_{ij}, \gamma^*_{ij,n}) + (1-\rho_1)Pr(G^*_{ij,n} = 1 \mid X_{ij}, \gamma^*_{ij,n})$.

3 Identification Analysis

We characterize the identified set based on the joint distribution $P_{0,n}$ of the observed variables $(G_{ij,n}, X_{ij}, \gamma_{ij,n})$.[5] In this section, we treat $\gamma_{ij,n}$ as observed because it can be estimated from the data as follows. For a generic value x in the support of X_{ij}, we can define:

$$\begin{aligned} \hat{p}(x) &= \frac{1}{n^2}\sum_{i,j} 1\{X_{ij} = x\} \\ \hat{\gamma}(x) &= \frac{\frac{1}{n^2}\sum_{i,j}\left(G_{ji,n}, \frac{1}{n}\sum_k G_{kj,n}, \frac{1}{n}\sum_k G_{ki,n}G_{kj,n}, \frac{1}{n}\sum_k (G_{ki,n}+G_{kj,n})\right)' 1\{X_{ij} = x\}}{\hat{p}(x)}, \end{aligned}$$

where $\hat{p}(x)$ is an estimator for $Pr(X_{ij} = x)$ and $\hat{\gamma}_{ij} = \hat{\gamma}(X_{ij})$ is an estimator for $\gamma_{ij,n}$. Then we can estimate the distribution of $(G_{ij,n}, X_{ij}, \gamma_{ij,n})$ using the empirical distribution of $(G_{ij,n}, X_{ij}, \hat{\gamma}_{ij})$.

To formalize our identification analysis, we introduce the following notation. Denote by $\mathcal{P}^*$ the set of joint distributions of $(G_{ij,n}, G^*_{ij,n}, X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n}, \varepsilon_{ij})$. Define the parameter space $\Theta = \mathcal{B} \times \mathcal{R}$, where $\mathcal{B}$ is the parameter space for β_0 and $\mathcal{R}$ is a subset of $\{(r_0, r_1) : r_0, r_1 \geq 0, r_0 + r_1 < 1\}$. Denote by $\mathcal{P}$ the set of joint distributions of $(G_{ij,n}, X_{ij}, \gamma_{ij,n})$.

Given Assumptions 1-3 and based on the results summarized in Lemmas 1-3, we impose the following three conditions on the true joint distribution $P^*_{0,n}$ of the variables $(G_{ij,n}, G^*_{ij,n}, X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n}, \varepsilon_{ij})$ and the true parameter value $\theta_0 = (\beta, \rho_0, \rho_1)$.

Condition 1. *Under P^*, the following holds: (i) ε_{ij} is normally distributed with mean zero and variance one. (ii) ε_{ij} and $(X_{ij}, \gamma^*_{ij,n})$ are independent.*

[5] The identified set can be characterized based on the joint distribution of $\{(G_{ij,n}, X_{ij}, \gamma_{ij,n}) : i, j \in \mathcal{N}_n\}$. However, it is unfeasible to estimate the joint distribution of all the dyads in the network from a sample of n agents. Hence, the identified set based on $\{(G_{ij,n}, X_{ij}, \gamma_{ij,n}) : i, j \in \mathcal{N}_n\}$ is not immediately useful for inference. In contrast, $P_{0,n}$ can be estimated from our current sample.

Condition 2. *$G^*_{ij,n} = 1\left\{(Z^*_{ij,n})'b + \varepsilon_{ij} \geq 0\right\}$ a.s. P^*, where $Z^*_{ij,n} = \left((\gamma^*_{ij,n})', (X_{ij})'\right)'$.*

Condition 3. *(i) $P^*(G_{ij,n} = 1 \mid X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n}) = r_0 P^*(G^*_{ij,n} = 0 \mid X_{ij}, \gamma^*_{ij,n}) + (1 - r_1)P^*(G^*_{ij,n} = 1 \mid X_{ij}, \gamma^*_{ij,n})$. (ii) $\gamma^*_{ij,n} = c(r_0, r_1) + C(r_0, r_1)\gamma_{ij,n}$ a.s. P^*.*

For each element P of $\mathcal{P}$, we are going to define the identified set based on the three conditions above.

Definition 1. *For each distribution $P \in \mathcal{P}$, the identified set $\Theta_I(P)$ is defined as the set of all $\theta = (b, r_0, r_1)$ in Θ for which there is some joint distribution $P^* \in \mathcal{P}^*$ such that Conditions 1, 2, and 3 hold and the distribution of $(G_{ij,n}, X_{ij}, \gamma_{ij,n})$ induced from P^* is equal to P.*

Note that the definition of $\Theta_I(P)$ does not depend on n, but the identified set $\Theta_I(P_{0,n})$ under the data generating process $P_{0,n}$ can depend on the sample size when the data distribution $P_{0,n}$ depends on n.

Using Definition 1, we characterize the identified set as follows.

Theorem 1. *Given a joint distribution $P \in \mathcal{P}$, $\Theta_I(P)$ is equal to the set of $\theta \in \Theta$ satisfying*

$$E_P[G_{ij,n} \mid X_{ij}, \gamma_{ij,n}] = \Psi(\theta, X_{ij}, \gamma_{ij,n}), \tag{1}$$

where, for a generic value $(x, \check{\gamma}_{ij})$ of $(X_{ij}, \gamma_{ij,n})$, we define

$$\Psi(\theta, x, \check{\gamma}_{ij}) = r_0 + (1 - r_0 - r_1)\Phi((c(r_0, r_1) + C(r_0, r_1)\check{\gamma}_{ij})'b_1 + x'b_2).$$

If the links were measured without error, the moment equation in Eq. (1) would degenerate into the model in Leung (2015): $E_P[G_{ij,n} - \Phi([\gamma_{ij,n}]'_{123}b_1 + X'_{ij}b_2) \mid X_{ij}, \gamma_{ij,n}] = 0$, where $[\gamma_{ij,n}]_{123}$ is a vector composed by the first three components of $\gamma_{ij,n}$.

Our characterization of the identified set in Theorem 1 relies on ε_{ij} being normally distributed. In Appendix B, we characterize the identified set in a semi-parametric framework.

4 Inference

In this section, we construct confidence intervals for θ based on the identification analysis in Theorem 1 and derive its asymptotic coverage when we observe one single network with many agents. As in Leung (2015) and Ridder and Sheng (2015), we use the asymptotic arguments based on a symmetric equilibrium.

We derive two confidence intervals for a prespecified significance level $\alpha \in (0, 1)$, and we suggest using $\hat{\mathcal{C}}_n(\alpha)$ introduced in Section 4.2 rather than $CI_n(\alpha)$ introduced in Section 4.1, because the computation of $\hat{\mathcal{C}}_n(\alpha)$ is much less demanding. In particular, the computation of $\hat{\mathcal{C}}_n(\alpha)$ only requires us to calculate the quasi-maximum likelihood estimator and its confidence interval for the grid values of (r_0, r_1). On the other hand, the computation of $CI_n(\alpha)$ would require us to evaluate the test statistic that characterizes the confidence interval at every value of $\theta = (b, r_0, r_1)$, and therefore the computational cost of $CI_n(\alpha)$ can be exponential in the number of the (exogenous and endogenous) regressors.

4.1 Confidence Interval through Test Inversion

Consider the unconditional sample analog of the moment condition in Eq. (1):

$$\hat{m}_n(\theta) = \hat{m}_n(b, r_0, r_1) = \frac{1}{n^2}\sum_{i,j}\left(G_{ij,n} - \Psi(\theta, X_{ij}, \hat{\gamma}_{ij})\right)\zeta_{ij},$$

where $x_1, \ldots, x_J$ are all the support points for X_{ij} and $\zeta_{ij} = (1\{X_{ij} = x_1\}, \ldots, 1\{X_{ij} = x_J\})'$. Notice that $\hat{m}_n$ is different from the infeasible sample moment

$$m_n(\theta) = \frac{1}{n^2} \sum_{i,j} \left(G_{ij,n} - \Psi(\theta, X_{ij}, \gamma_{ij,n})\right) \zeta_{ij},$$

because $\gamma_{ij,n}$ is estimable but unknown. We estimate the variance of $\hat{m}_n(\theta)$ by

$$\hat{S}(\theta) = \hat{S}(b, r_0, r_1) = \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_i(\theta) \hat{\psi}_i(\theta)' - \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_i(\theta) \right) \left(\frac{1}{n} \sum_{i=1}^{n} \hat{\psi}_i(\theta) \right)' \right),$$

where $\hat{\psi}_i(\theta)$ denotes the estimated influence function of $\hat{m}_n(\theta)$:

$$\hat{\psi}_i(\theta) \quad = \quad \frac{1}{n} \sum_{j \neq i} G_{ij,n} \zeta_{ij} - \frac{1}{n^2} \sum_{l,j} \left(\frac{\partial}{\partial \tilde{\gamma}_{lj}'} \Psi(\theta, X_{lj}, \tilde{\gamma}_{lj}) \bigg|_{\tilde{\gamma}_{lj} = \hat{\gamma}_{lj,n}} \right) \hat{\psi}_{\gamma,i,n}(X_{lj}) \zeta_{lj}.$$

The component $\hat{\psi}_{\gamma,i,n}(x)$ in $\hat{\psi}_i(\theta)$ denotes the estimated influence function of the first stage estimator $\hat{\gamma}_{ij}$:

$$\hat{\psi}_{\gamma,k,n}(x) = \frac{1}{n^2} \sum_{i_1, j_1} \frac{1\{X_{i_1,j_1} = x\}}{\hat{p}(x)} \begin{pmatrix} 0 \\ G_{kj_1} \\ G_{ki_1} G_{kj_1} \\ G_{ki_1} + G_{kj_1} \end{pmatrix} + \frac{1}{n} \sum_{i_1} \frac{1\{X_{i_1,k} = x\}}{\hat{p}(x)} \begin{pmatrix} G_{ki_1} \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

We construct a confidence interval for θ as:

$$CI_n(\alpha) = \{\theta \in \Theta : n \hat{m}_n(\theta)' \hat{S}(\theta)^{-1} \hat{m}_n(\theta) \leq q(\chi_J^2, 1 - \alpha)\},$$

where $q(\chi_J^2, 1-\alpha)$ is the $(1-\alpha)$ quantile of a χ^2 distribution with J degrees of freedom. The degrees of freedom are determined by the number of points in the support of X_{ij}.

The following theorem demonstrates the asymptotic coverage for the confidence interval $CI_n(\alpha)$.

Theorem 2. *Suppose that (i) the minimum eigenvalue of $Var(\psi_i(\theta_0) \mid X, \sigma_n)$ is bounded away from zero, and (ii) $\liminf \min_x \hat{p}(x) > 0$. Under Assumptions 1-3,*

$$\liminf_{n \to \infty} Pr(\theta_0 \in CI_n(\alpha) \mid X, \sigma_n) \geq 1 - \alpha.$$

Condition (i) guarantees that $\hat{S}(\theta)$ is non-singular. A similar condition is used in Leung (2015, Theorem 3). Condition (ii) is required to ensure uniform consistency of the first-stage estimator $\hat{\gamma}_{ij}$.

4.2 Confidence Interval based on Quasi-Maximum Likelihood Estimator

In this section, we construct a more computationally feasible (but potentially larger) confidence interval for β. In addition to the assumptions in the previous section, we assume that $\{((\gamma_{ij,n}^*)', X_{ij}')' : i, j\}$ is not contained in any proper linear subspace of $\mathbb{R}^{d+3}$. By this assumption, the parameter β_0 would be identified if we knew the true value of (ρ_0, ρ_1).[6]

[6] The exact statement and its proof are found in Lemma 12 in the appendix.

If we knew $(\rho_0, \rho_1) = (r_0, r_1)$ for a given value $(r_0, r_1) \in \mathcal{R}$, we could construct a confidence interval $\mathcal{C}_n(\alpha; r_0, r_1)$ for β by computing the quasi-maximum likelihood estimator $\hat{\beta}(r_0, r_1)$ and its estimated asymptotic variance $\widehat{AV}(r_0, r_1)$ in the following way. We consider the following quasi-maximum likelihood estimator:

$$\hat{\beta}(r_0, r_1) = \arg\max_{b \in \mathcal{B}} \hat{\boldsymbol{Q}}_n(b, r_0, r_1)$$

where the feasible objective function is:

$$\hat{\boldsymbol{Q}}_n(b, r_0, r_1) = \frac{1}{n^2} \sum_{i,j} \log \left(\Psi(b, r_0, r_1, X_{ij}, \hat{\gamma}(X_{ij}))^{G_{ij,n}} (1 - \Psi(b, r_0, r_1, X_{ij}, \hat{\gamma}(X_{ij})))^{1-G_{ij,n}} \right).$$

Define the estimated influence function of $\hat{\boldsymbol{Q}}_n(b, r_0, r_1)$ by

$$\begin{aligned} \hat{\psi}_{\boldsymbol{Q},k,n}(\theta) &= \frac{1}{n} \sum_j (G_{kj,n} - \Psi(\theta, X_{kj}, \hat{\gamma}(X_{kj}))) \boldsymbol{C}_1(\theta, X_{ij}, \hat{\gamma}(X_{kj})) \\ &\quad + \frac{1}{n^2} \sum_{i,j} \left(\Psi(\theta, X_{ij}, \hat{\gamma}(X_{ij})) \boldsymbol{D}_1(\theta, X_{ij}, \hat{\gamma}(X_{ij})) - \boldsymbol{D}_2(\theta, X_{ij}, \hat{\gamma}(X_{ij})) \right) \hat{\psi}_{\gamma,k,n}(X_{ij}), \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{C}_1(\theta, X_{ij}, \check{\gamma}_{ij}) &= \frac{\frac{\partial}{\partial b} \Psi(\theta, X_{ij}, \check{\gamma}_{ij})}{\Psi(\theta, X_{ij}, \check{\gamma}_{ij})(1 - \Psi(\theta, X_{ij}, \check{\gamma}_{ij}))} \\ \boldsymbol{C}_2(\theta, X_{ij}, \check{\gamma}_{ij}) &= \frac{\frac{\partial}{\partial b} \Psi(\theta, X_{ij}, \check{\gamma}_{ij})}{1 - \Psi(\theta, X_{ij}, \check{\gamma}_{ij})} \\ \boldsymbol{D}_1(\theta, X_{ij}, \check{\gamma}_{ij}) &= \frac{\partial}{\partial \check{\gamma}_{ij}'} \boldsymbol{C}_1(\theta, X_{ij}, \check{\gamma}_{ij}) \\ \boldsymbol{D}_2(\theta, X_{ij}, \check{\gamma}_{ij}) &= \frac{\partial}{\partial \check{\gamma}_{ij}'} \boldsymbol{C}_2(\theta, X_{ij}, \check{\gamma}_{ij}). \end{aligned}$$

The asymptotic variance for $\hat{\beta}(r_0, r_1)$ is estimated by:

$$\begin{aligned} \widehat{AV}(r_0, r_1) &= \left(\frac{\partial^2}{\partial b \partial b'} \hat{\boldsymbol{Q}}_n(b, r_0, r_1) \Big|_{b = \hat{\beta}(r_0, r_1)} \right)^{-1} \left(\frac{1}{n} \sum_{k=1}^n \hat{\psi}_{\boldsymbol{Q},k,n}(\hat{\beta}(r_0, r_1), r_0, r_1) \hat{\psi}_{\boldsymbol{Q},k,n}(\hat{\beta}(r_0, r_1), r_0, r_1)' \right) \\ &\quad \times \left(\frac{\partial^2}{\partial b \partial b'} \hat{\boldsymbol{Q}}_n(b, r_0, r_1) \Big|_{b = \hat{\beta}(r_0, r_1)} \right)^{-1}, \end{aligned}$$

and a confidence interval for β is:

$$\mathcal{C}_n(\alpha; r_0, r_1) = \left\{ b \in \mathcal{B} : n(\hat{\beta}(r_0, r_1) - b)' \widehat{AV}(r_0, r_1)^{-1} (\hat{\beta}(r_0, r_1) - b) \le q(\chi^2_{d+3}, 1 - \alpha) \right\},$$

where $q(\chi^2_{d+3}, 1 - \alpha)$ is the $(1 - \alpha)$ quantile of a χ^2 distribution with $d + 3$ degrees of freedom.[7]

[7] We can construct a confidence interval for a subvector $\eta'\beta$ of a given vector η:

$$\eta' \hat{\beta}(r_0, r_1) \pm q(N(0,1), 1 - \alpha/2) \sqrt{\frac{\eta' \widehat{AV}(r_0, r_1) \eta}{n}}$$

where $q(N(0,1), 1 - \alpha/2)$ is the $(1 - \alpha/2)$ quantile of the standard normal distribution. In the same way as β, we can also take the union over $(r_0, r_1) \in \mathcal{R}$ and construct a feasible confidence interval for $\eta'\beta$.

Since we do not know the true value of (ρ_0, ρ_1), we construct a confidence interval for β by taking the union of $\mathcal{C}_n(\alpha; r_0, r_1)$ over $(r_0, r_1) \in \mathcal{R}$:

$$\hat{\mathcal{C}}_n(\alpha) = \bigcup_{(r_0,r_1)\in\mathcal{R}} \mathcal{C}_n(\alpha; r_0, r_1).$$

This confidence interval contains the true parameter value with correct asymptotic size. The next theorem formalizes this result.

Theorem 3. *Suppose that (i)* $\liminf \min_x \hat{p}(x) > 0$*, (ii)* β_0 *is in the interior of a compact subset* $\mathcal{B}$ *of the Euclidean space, (iii)* $\{((\gamma^*_{ij,n})', X'_{ij})' : i, j\}$ *is not contained in any proper linear subspace of* $\mathbb{R}^{d+3}$*, and (iv) the minimum eigenvalue of* $E\left[\frac{1}{n}\sum_{k=1}^n \psi_{\boldsymbol{Q},k,n}\psi'_{\boldsymbol{Q},k,n}\right]$ *is bounded away from zero, where:*

$$\begin{aligned}\psi_{\boldsymbol{Q},k,n} \quad &= \quad \frac{1}{n}\sum_j (G_{kj,n} - \Psi(\theta_0, X_{kj}, \gamma_{kj,n}))\boldsymbol{C}_1(\theta_0, X_{ij}, \gamma_{ij,n}) \\ &\quad + \frac{1}{n^2}\sum_{i,j}\left(E\left[G_{ij,n} \mid X, \sigma_n\right]\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n}) - \boldsymbol{D}_2(\theta_0, X_{ij}, \gamma_{ij,n})\right)\psi_{\gamma,k,n}(X_{ij}).\end{aligned}$$

Under Assumptions 1-3,

$$\liminf_{n\to\infty} Pr(\beta_0 \in \hat{\mathcal{C}}_n(\alpha) \mid X, \sigma_n) \geq 1 - \alpha.$$

Condition (i) is required to ensure uniform consistency of the first-stage estimator $\hat{\gamma}_{ij}$. Condition (ii) is a regularity condition and is used to derive the asymptotic distribution of $\hat{\beta}(\rho_0, \rho_1)$. Condition (iii) is a rank condition and guarantees that β_0 is the unique maximizer of the limiting objective function. This assumption guarantees that the equilibrium beliefs about the network statistics $\gamma^*_{ij,n}$ have sufficient exogenous variations for any finite n. Condition (iv) ensures that $\widehat{AV}(\rho_0, \rho_1)$ is asymptotically invertible.

The size property of $\hat{\mathcal{C}}_n(\alpha)$ in Theorem 3 follows from:

$$\sqrt{n}(\widehat{AV}(\rho_0, \rho_1))^{-1/2}(\hat{\beta}(\rho_0, \rho_1) - \beta_0) \to_d N(0, I), \tag{2}$$

because $Pr(\beta_0 \in \hat{\mathcal{C}}_n(\alpha) \mid X, \sigma_n) \geq Pr(\beta_0 \in \mathcal{C}_n(\alpha; \rho_0, \rho_1) \mid X, \sigma_n)$. Although Eq. (2) is proved in a similar manner to Leung (2015, Theorem 3), it is not a direct implication of Leung (2015, Theorem 3) since we do not directly observe the true underlying network G^*_n.

5 Empirical Illustration

In this section, we implement the confidence interval introduced in Section 4.2 using data on social networks from 75 rural villages in southern India. In particular, we investigate the robustness of the empirical results in Leung (2015) to the presence of misclassified links. The data was collected in 2006 to study the introduction of a microfinance program (see Banerjee et al. 2013 and Jackson et al. 2012). This dataset contains household characteristics that were collected using full village censuses, and individual and network data that were obtained using follow-up surveys conducted on random samples of individuals in each village.

Among the different dimensions of social relationships contained in the dataset, we follow Leung (2015) and focus on trust networks. These networks measure the individuals' willingness to lend money. The direct links observed in our dataset are obtained from using the following survey question: "Who do you trust enough that if he/she needed to borrow Rs. 50 for a day you would lend it to him/her?"

Table 1: Descriptive statistics

	mean	sd	min	max
# villagers	225.667	67.446	98.000	303.000
in-degree	0.951	1.275	0	10
out-degree	0.951	0.807	0	4
average age	38.470	1.405	35.775	40.599
share female	0.561	0.016	0.547	0.592
share Hindu	0.793	0.106	0.581	0.918
share OBC	0.621	0.133	0.429	0.762
share scheduled	0.295	0.085	0.206	0.439

Jackson et al. (2012) discuss concerns about measurement error issues that might be present in this dataset. Potential sources include (i) individuals forgetting to mention connections, (ii) people getting fatigued by interviews, and (iii) top-censoring the number of social connections that individuals could report.[8] Under the structure of the survey questions, which ask individuals about actual actions (such as lending or borrowing money) rather than perceived relationships, individuals are more likely to forget existing interactions than to imagine non-existent ones. Hence, the most likely type of measurement error to appear in this dataset is the misclassification of true links as non-existent (i.e., false negative). Therefore, in this empirical illustration, we focus on the false-negative case. For completeness, we consider the false-positive case at the end of the empirical application.

We examine the relative importance that homophily on observed attributes and endogenous beliefs about trustworthiness have in the formation of trust networks. Regarding the preferences for homophily, we study homophily relations on gender, caste, language, religion, and family relationships. The villages are primarily homogeneous in language and religion but heterogeneous in caste. In terms of religion, Hinduism represents the large majority. Due to multicollinearity concerns and in order to study homophily on religion, we restrict our sample to 9 villages where the non-Hindu minorities have at least a 10% representation. The total sample then consists of $2,031$ individuals in those 9 villages.

Table 1 provides descriptive statistics about the observed attributes.[9] The average number of individuals across the 9 villages is 225, where the largest network has 303 individuals. The average degree across the 9 networks is equal to 0.951. The largest in-degree value observed for an individual within our sample is equal to 10, which means that, at least in one case, ten people in our sample listed a specific individual as someone to whom they would be willing to lend money. Meanwhile, the maximum out-degree observed across the 9 networks is equal to 4. In other words, the maximum number of people to whom a single individual within our sample was willing to lend money is equal to 4. On average, 56% percent of the individuals surveyed are female and 79.8% Hindu. The religious minorities, composed of Christians and Muslims, are aggregated into a general non-Hindu category. Scheduled castes are at the bottom of the hierarchy and represent 62% of the sample. OBC castes are second to bottom and account for 29%. The remaining 8.4% is composed of different castes that are aggregated into a general category at the top of the hierarchy.

The direct link $G_{ij,n}$, for any distinct individuals i and j, is recorded to be equal to 1 if individual i lists j as someone to whom he or she is willing to lend Rs. 50, and 0 otherwise. Notice that we allow for $G_{ij,n}$ to

[8] Network subsampling is an interesting and challenging setting that is beyond the scope of this paper. Hence, we assume that we observe the full list of nodes in the networks.

[9] Hindu is a dummy variable that is equal to 1 if the respondent reports being Hindu and equal to 0 if he or she reports being Christian or Muslim. Scheduled and OBC castes are respectively the bottom and second to bottom caste of the hierarchy. The remaining castes are aggregated into a general category at the top of the hierarchy.

be potentially misclassified. In the vector of observed attributes X_{ij}, we include individual i- and individual j-specific regressors, such as age, caste, gender, religion, and an indicator for whether or not i and j are heads of their household, as well as the controls for homophily described above.

In the vector of endogenous network statistics $\gamma^*_{ij,n}$, we consider the conditional expectation of the following factors: (i) $G^*_{ji,n}$, which accounts for the value of reciprocation; (ii) $n^{-1}\sum_{k\neq i} G^*_{kj,n}$, which measures the share of people willing to lend money to j; and (iii) $n^{-1}\sum_{k\neq i,j} G^*_{ki,n}G^*_{kj,n}$, which is the supported trust or share of individuals that are willing to lend to both i and j. We account for the misclassification on the endogenous network statistics via Lemma 2. As a first stage estimator, we use the frequency estimator described in Section 3.

In order to examine the effects of potentially misclassified links on the estimation of the structural parameters of a network formation model, we allow for the possibility of missing links to exist in the true underlying network. In particular, we consider six scenarios for the probability of misclassifying a link as missing when it is present in the true underlying network, i.e., $r_1 \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$. Given that we focus on false negatives, we assume that $r_0 = 0$, i.e., the links that are not present in the true underlying network are not misclassified. Consequently, by setting $r_0 = 0$ and $r_1 = 0$, we can analyze the scenario when there is no measurement error in the network. This scenario corresponds to the empirical analysis in Leung (2015), and we use it as our baseline case.[10]

Table 2 presents 95% confidence intervals for the estimates of the network statistics, the homophily parameters, and the constant term. Point estimates are reported in Table 3. Column 1 of Table 2 presents the confidence intervals for the "no misclassification" case, i.e., $r_0 = r_1 = 0$. These results indicate that reciprocation is an important endogenous factor in determining the willingness of an individual to lend money. In other words, an individual within the network is more willing to lend money to someone else if that trust is reciprocated. There is also evidence that individuals present preferences for homophily on gender and for being close relatives when lending money to another individual. There is no conclusive evidence regarding the relative importance of the remaining covariates considered in the specification for the network formation model.

Columns 2 to 6 of Table 2 present 95% confidence intervals for the parameter estimates under different misclassification probabilities, $r_1 \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$. Reciprocation and homophily on gender and family relationships remain significant factors driving the formation of a link on trust networks. Most of the confidence intervals become wider as the size of the measurement error problem becomes more pervasive in the data. Nonetheless, this increase remains relatively small for the cases of misclassification considered. Table 4 provides further evidence about this insight.

In Table 4, we compare the length of 95% confidence intervals that take misclassification into account, with the length of the confidence intervals computed under the assumption of no measurement error, i.e., $|\hat{\mathcal{C}}_n(\alpha)|/|\mathcal{C}_n(\alpha, 0, 0)|$. Given a misclassification probability of up to 10%, the coefficient for reciprocation presents the largest increase in interval length, which is 7.9% larger than the baseline confidence interval. With a misclassification probability of up to 30%, the largest increase in the length of the confidence intervals is 24.3% larger than the confidence intervals that do not take measurement error into account. Finally, at the extreme case of at most 50% misclassification probability, the largest increase in the length of confidence intervals is 49.8% of the benchmark confidence intervals, and the second-largest increase is 26%.

Overall, this empirical illustration suggests that reciprocation, homophily on gender, and family relation-

[10]Unlike the empirical specification in Leung (2015), we do not include in-degree or out-degree statistics weighted by caste or religion as part of the factors that capture network externalities.

ships are the most important factors driving the lending decisions of the individuals in the network. These results are robust to the misclassification of existent links as non-existent in network data. Furthermore, we use different scenarios for the misclassification probabilities to compare the length of 95% confidence intervals computed under our method to the length of confidence intervals that assume no measurement error. The analysis suggests that with a misclassification probability of up to 50%, the 95% confidence intervals are at most 49.8% larger than the length of the baseline confidence intervals.

5.1 Alternative Design: False Positives

As an alternative design, we consider the case when non-existent links are misclassified as existent (i.e., false positives). In particular, we consider $r_0 \in \{0, 0.05, 0.1, 0.2\}$ and $r_1 = 0$. Table 5 presents 95% confidence intervals for the estimates of the parameters of interest. In Column 1 of Table 5, we report the confidence intervals for the "no misclassification" case to simplify comparisons. Columns 2 to 4 of Table 5 present 95% confidence intervals for the parameter estimates under $r_0 \in \{0.05, 0.1, 0.2\}$. At a 5% misclassification probability, reciprocation is no longer a significant factor in explaining the formation of links on trust networks. Homophily on gender loses significance at a 10% misclassification probability, and none of the factors considered are significant at a 20% misclassification probability.

Table 6 compares the length of 95% confidence intervals that take misclassification of non-existent links into account with the length of the confidence intervals computed under the assumption of no measurement error. The evidence suggests that even at 5% small misclassification probability, the confidence intervals can be 65% wider in length than the baseline confidence intervals. The length of the confidence intervals can be severely affected by 10% and higher misclassification probabilities. These results suggest that the estimates of the parameters in the network formation model might be sensitive to misclassifying non-existent links as real links, even at small probabilities. One explanation behind these outcomes could be the sparsity of the networks considered in our empirical application. In other words, due to the reduced number of links formed in the network, the effects of misclassifying links are asymmetric and more pervasive for the false-positive case.

6 Conclusion

We study a network formation model with potentially misclassified links. Specifically, we focus on a strategic game of network formation with incomplete information. We propose a novel approach for analyzing network formation models, which is robust to misclassification. In particular, we characterize the identified set for the structural parameters, including the preference parameters concerning homophily and network externalities. Based on the identification result, we develop an inference method which is asymptotically valid when a single large network is available. We apply the proposed inference method to examine trust networks in Karnataka, India. Using different scenarios for the misclassification probabilities, our 95% confidence intervals demonstrate the statistical significance of the key factors driving the lending decisions of the individuals in the network.

Table 2: 95% Confidence intervals $\hat{\mathcal{C}}_n(\alpha)$ with $n = 2,031$ and $r_0 = 0$

	$r_1 = 0$ (1)	$r_1 \leq 0.1$ (2)	$r_1 \leq 0.2$ (3)	$r_1 \leq 0.3$ (4)	$r_1 \leq 0.4$ (5)	$r_1 \leq 0.5$ (6)
Reciprocation	[0.833, 2.184]	[0.727, 2.184]	[0.618, 2.184]	[0.505, 2.184]	[0.365, 2.184]	[0.160, 2.184]
In degree	[-60.108, 119.573]	[-60.108, 119.573]	[-60.108, 119.573]	[-60.108, 119.573]	[-60.108, 119.573]	[-60.108, 119.573]
Supported trust	[-133.290, 302.937]	[-133.290, 302.937]	[-133.290, 302.937]	[-133.290, 302.937]	[-133.290, 302.937]	[-133.290, 302.937]
Constant	[-11.073, 3.641]	[-11.153, 3.776]	[-11.240, 3.919]	[-11.350, 4.089]	[-11.502, 4.311]	[-11.722, 4.612]
Same religion	[-0.147, 0.987]	[-0.149, 0.998]	[-0.152, 1.012]	[-0.155, 1.030]	[-0.158, 1.053]	[-0.166, 1.087]
Same sex	[0.481, 0.789]	[0.481, 0.803]	[0.481, 0.817]	[0.481, 0.830]	[0.481, 0.846]	[0.481, 0.868]
Same caste	[-0.136, 0.641]	[-0.147, 0.659]	[-0.159, 0.680]	[-0.174, 0.704]	[-0.191, 0.734]	[-0.211, 0.770]
Same language	[-0.792, 0.860]	[-0.810, 0.878]	[-0.832, 0.898]	[-0.857, 0.922]	[-0.889, 0.953]	[-0.932, 0.993]
Same family	[0.303, 2.541]	[0.303, 2.617]	[0.303, 2.681]	[0.303, 2.741]	[0.303, 2.828]	[0.303, 2.975]

Table 3: Estimates for the parameter coefficients given a value of r_1

	$r_1 = 0$ (1)	$r_1 = 0.1$ (2)	$r_1 = 0.2$ (3)	$r_1 = 0.3$ (4)	$r_1 = 0.4$ (5)	$r_1 = 0.5$ (6)
Reciprocation	1.509	1.444	1.367	1.277	1.176	1.066
	(0.345)	(0.366)	(0.382)	(0.394)	(0.414)	(0.462)
In degree	29.733	27.202	24.673	22.114	19.508	16.837
	(45.838)	(42.387)	(38.823)	(35.174)	(31.439)	(27.606)
Supported trust	84.823	73.997	61.288	49.072	37.911	27.995
	(111.285)	(92.788)	(75.330)	(59.840)	(46.266)	(34.573)
Constant	-3.716	-3.688	-3.661	-3.630	-3.595	-3.555
	(3.753)	(3.808)	(3.867)	(3.939)	(4.034)	(4.167)
Same religion	0.420	0.424	0.430	0.438	0.447	0.461
	(0.289)	(0.293)	(0.297)	(0.302)	(0.309)	(0.319)
Same sex	0.635	0.645	0.657	0.669	0.684	0.701
	(0.078)	(0.081)	(0.082)	(0.082)	(0.083)	(0.085)
Same caste	0.252	0.256	0.260	0.265	0.272	0.279
	(0.198)	(0.206)	(0.214)	(0.224)	(0.236)	(0.250)
Same language	0.034	0.034	0.033	0.033	0.032	0.031
	(0.421)	(0.431)	(0.441)	(0.454)	(0.470)	(0.491)
Same family	1.422	1.488	1.559	1.637	1.724	1.828
	(0.571)	(0.576)	(0.572)	(0.563)	(0.563)	(0.585)

Note: Standard errors are in parenthesis. Sample size is $n = 2,031$.

Table 4: Ratio of lengths of 95% confidence intervals, $|\hat{\mathcal{C}}_n(\alpha)|/|\mathcal{C}_n(\alpha, 0, 0)|$, with $n = 2,031$

	$r_1 \leq 0.1$ (1)	$r_1 \leq 0.2$ (2)	$r_1 \leq 0.3$ (3))	$r_1 \leq 0.4$ (4)	$r_1 \leq 0.5$ (5)
Reciprocation	1.079	1.159	1.243	1.347	1.498
In degree	1.000	1.000	1.000	1.000	1.000
Supported trust	1.000	1.000	1.000	1.000	1.000
Constant	1.015	1.030	1.049	1.075	1.110
Same religion	1.013	1.028	1.045	1.069	1.105
Same sex	1.046	1.089	1.134	1.186	1.256
Same caste	1.038	1.080	1.130	1.190	1.262
Same language	1.022	1.047	1.077	1.115	1.165
Same family	1.034	1.062	1.089	1.128	1.194

Table 5: 95% Confidence intervals $\hat{\mathcal{C}}_n(\alpha)$ with $n = 2,031$ and $r_1 = 0$

	$r_0 = 0$ (1)	$r_0 \leq 0.05$ (2)	$r_0 \leq 0.1$ (3)	$r_0 \leq 0.2$ (4)
Reciprocation	[0.833, 2.184]	[-0.056, 2.184]	[-0.415, 2.184]	[-0.415, 2.184]
In degree	[-60.107, 119.567]	[-71.059, 119.567]	[-71.059, 119.567]	[-71.059, 119.567]
Supported trust	[-133.213, 302.953]	[-133.213, 302.953]	[-133.213, 302.953]	[-133.213, 302.953]
Constant	[-11.072, 3.640]	[-11.072, 3.640]	[-11.072, 3.640]	[-11.072, 3.640]
Same religion	[-0.147, 0.986]	[-2.076, 0.986]	[-5.939, 0.986]	[-5.939, 0.986]
Same sex	[0.481, 0.789]	[0.124, 1.257]	[-0.205, 4.160]	[-0.425, 4.160]
Same caste	[-0.136, 0.641]	[-1.019, 0.968]	[-1.019, 1.777]	[-1.019, 1.777]
Same language	[-0.792, 0.860]	[-0.792, 0.860]	[-0.792, 0.860]	[-0.792, 0.860]
Same family	[0.303, 2.541]	[0.303, 6.223]	[0.303, 10.519]	[-0.359, 10.519]

Table 6: Ratio of lengths of 95% confidence intervals, $|\hat{\mathcal{C}}_n(\alpha)|/|\mathcal{C}_n(\alpha,0,0)|$, with $n = 2,031$

	$r_0 \leq 0.05$ (1)	$r_0 \leq 0.1$ (2)	$r_0 \leq 0.2$ (3))
Reciprocation	1.658	1.924	1.924
In degree	1.061	1.061	1.061
Supported trust	1.000	1.000	1.000
Constant	1.000	1.000	1.000
Same religion	2.703	6.113	6.113
Same sex	3.680	14.187	14.904
Same caste	2.558	3.600	3.600
Same language	1.000	1.000	1.000
Same family	2.645	4.565	4.860

A Proofs

A.1 Proof of Lemmas in Section 2

Proof of Lemma 1. By Assumption 2,

$$G^*_{i,n} = \underset{g^*_{i,n}\in\mathcal{G}^n}{\arg\max}\, E\left[U_i(g^*_{i,n}, G^*_{-i,n}, X, \varepsilon_i) \mid X, \varepsilon_i, \sigma_n\right] = \underset{g^*_{i,n}\in\mathcal{G}^n}{\arg\max}\, \frac{1}{n}\sum_{j=1}^{n} g^*_{ij,n}\left[(Z^*_{ij,n})'\beta_0 + \varepsilon_{ij}\right].$$

Therefore, $G^*_{ij,n} = 1\left\{(Z^*_{ij,n})'\beta_0 + \varepsilon_{ij} \geq 0\right\}$. □

Proof of Lemma 2. Define

$$D(r_0, r_1) = \begin{pmatrix} 1-r_0-r_1 & 0 & 0 & 0 \\ 0 & 1-r_0-r_1 & 0 & 0 \\ 0 & 0 & (1-r_0-r_1)^2 & r_0(1-r_0-r_1) \\ 0 & 0 & 0 & 1-r_0-r_1 \end{pmatrix}.$$

By Assumption 3, we can derive:

$$E\left[G_{ki,n}G_{kj,n} \mid X, \sigma_n\right] = \rho_0^2 + (1-\rho_0-\rho_1)^2 E\left[G^*_{ki,n}G^*_{kj,n} \mid X, \sigma_n\right] + \rho_0(1-\rho_0-\rho_1)E\left[G^*_{ki,n} + G^*_{kj,n} \mid X, \sigma_n\right].$$

Therefore:

$$\begin{aligned} \gamma_{ij,n} &= \begin{pmatrix} E\left[G_{ji,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G_{kj,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G_{ki,n}G_{kj,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G_{ki,n} + G_{kj,n} \mid X, \sigma_n\right] \end{pmatrix} \\ &= \begin{pmatrix} \rho_0 \\ \rho_0 \\ \rho_0^2 \\ \rho_0 \end{pmatrix} + D(\rho_0, \rho_1)\begin{pmatrix} E\left[G^*_{ji,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G^*_{kj,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G^*_{ki,n}G^*_{kj,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G^*_{ki,n} + G^*_{kj,n} \mid X, \sigma_n\right] \end{pmatrix}. \end{aligned}$$

Since $D(\rho_0, \rho_1)$ is invertible given $1-\rho_0-\rho_1 \neq 0$, it follows that:

$$\begin{pmatrix} E\left[G^*_{ji,n} \mid X\right] \\ \frac{1}{n}\sum_k E\left[G^*_{kj,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G^*_{ki,n}G^*_{kj,n} \mid X, \sigma_n\right] \\ \frac{1}{n}\sum_k E\left[G^*_{ki,n} + G^*_{kj,n} \mid X, \sigma_n\right] \end{pmatrix} = D(\rho_0, \rho_1)^{-1}\left(\gamma_{ij,n} - \begin{pmatrix} \rho_0 \\ \rho_0 \\ \rho_0^2 \\ \rho_0 \end{pmatrix}\right).$$

The first three components of the right-hand side of the above equation are: $\gamma^*_{ij,n}$, so

$$\gamma^*_{ij,n} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} D(\rho_0, \rho_1)^{-1}\left(\gamma_{ij,n} - \begin{pmatrix} \rho_0 \\ \rho_0 \\ \rho_0^2 \\ \rho_0 \end{pmatrix}\right) = c(\rho_0, \rho_1) + C(\rho_0, \rho_1)\gamma_{ij,n}.$$

□

Proof of Lemma 3. It suffices to show that $Pr(G_{ij,n} = 1 \mid X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n}, X, \sigma_n) = \rho_0 Pr(G^*_{ij,n} = 0 \mid X_{ij}, \gamma^*_{ij,n}) + (1-\rho_1)Pr(G^*_{ij,n} = 1 \mid X_{ij}, \gamma^*_{ij,n})$. Since $(X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n})$ are a function of (X, σ_n), it follows that:

$$Pr(G_{ij,n} = 1 \mid X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n}, X, \sigma_n) = Pr(G_{ij,n} = 1 \mid X, \sigma_n).$$

Using Assumptions 1-3:

$$\begin{aligned}
Pr(G_{ij,n} = 1 \mid X, \sigma_n) &= \rho_0 Pr(G^*_{ij,n} = 0 \mid X, \sigma_n) + (1-\rho_1)Pr(G^*_{ij,n} = 1 \mid X, \sigma_n) \\
&= \rho_0 Pr((Z^*_{ij,n})'b + \varepsilon_{ij} < 0 \mid X, \sigma_n) + (1-\rho_1)Pr((Z^*_{ij,n})'b + \varepsilon_{ij} \geq 0 \mid X, \sigma_n) \\
&= \rho_0 Pr((Z^*_{ij,n})'b + \varepsilon_{ij} < 0 \mid Z^*_{ij,n}) + (1-\rho_1)Pr((Z^*_{ij,n})'b + \varepsilon_{ij} \geq 0 \mid Z^*_{ij,n}),
\end{aligned}$$

where the first equality follows from Assumption 3, the second follows from Lemma 1, and the last follows from the independence between ε and X. □

A.2 Proof of Theorem 1

Proof. To show that every element θ of $\Theta_I(P)$ satisfies Eq. (1), we can derive the following equalities:

$$\begin{aligned}
P(G_{ij,n} = 1 \mid X_{ij}, \gamma_{ij,n}) &= P^*(G_{ij,n} = 1 \mid X_{ij}, \gamma_{ij,n}) \\
&= P^*(G_{ij,n} = 1 \mid X_{ij}, \gamma_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1)P^*(G^*_{ij,n} = 1 \mid X_{ij}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1)P^*((Z^*_{ij,n})'b + \varepsilon_{ij} \geq 0 \mid X_{ij}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1)\Phi((\gamma^*_{ij,n})'b_1 + X'_{ij}b_2) \\
&= r_0 + (1 - r_0 - r_1)\Phi((c(r_0, r_1) + C(r_0, r_1)\gamma_{ij,n})'b_1 + X'_{ij}b_2),
\end{aligned}$$

where the first equality follows from $P = P^*$ for the observables $(G_{ij,n}, X_{ij}, \gamma_{ij,n})$, the second equality follows because $\gamma^*_{ij,n}$ is a function of $\gamma_{ij,n}$ in Condition 3(ii), the third equality follows from Condition 3(i), the fourth equality follows from Condition 2, the fifth equality follows from Condition 1, and the last equality follows from Condition 3(ii). The rest of the proof will show that every element θ of Θ satisfying Eq. (1) belongs to $\Theta_I(P)$.

Define the joint distribution P^* in the following way. The marginal distribution of ε_{ij} is standard normal. The conditional distribution of $(\gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij})$ given ε_{ij} is

$$P^*((\gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}) \in B \mid \varepsilon_{ij}) = P((\gamma_{ij,n}, c(r_0, r_1) + C(r_0, r_1)\gamma_{ij,n}, X_{ij}) \in B) \tag{3}$$

for all the measurable sets B. The conditional distribution of $G^*_{ij,n}$ given $(\gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij})$ is:

$$P^*(G^*_{ij,n} = 1 \mid \gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij}) = 1\{(Z^*_{ij,n})'b + \varepsilon_{ij} \geq 0\}. \tag{4}$$

The conditional distribution of $G_{ij,n}$ given $(G^*_{ij,n}, \gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij})$ is:

$$P^*(G_{ij,n} = 1 \mid G^*_{ij,n}, \gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij}) = \begin{cases} r_0 & \text{if } G^*_{ij,n} = 0 \\ 1 - r_1 & \text{if } G^*_{ij,n} = 1. \end{cases} \tag{5}$$

Note that (P^*, θ) satisfies Conditions 1-3, because Condition 1(i) follows because ε_{ij} is normally distributed under P^*, Condition 1(ii) follows from Eq. (3). Condition 2 follows from Eq. (4). Condition 3(i) follows from Eq. (4) and (5), and Condition 3(ii) follows from Eq. (3).

The distribution of $(G_{ij,n}, X_{ij}, \gamma_{ij,n})$ induced from P^* is equal to P. The distribution of $(X_{ij}, \gamma_{ij,n})$ induced from P^* is equal to that from P, by the construction of $P^*((\gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}) \in B \mid \varepsilon_{ij})$. The equality of $P^*(G_{ij,n} = 1 \mid Z_{ij,n}) = P(G_{ij,n} = 1 \mid Z_{ij,n})$ a.s. under P^* is shown as follows. Note that:

$$\gamma^*_{ij,n} = c(r_0, r_1) + C(r_0, r_1)\gamma_{ij,n} \text{ a.s. under } P^*. \tag{6}$$

Then:

$$\begin{aligned}
P^*(G_{ij,n} = 1 \mid Z_{ij,n}) &= P^*(G_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 P^*(G^*_{ij,n} = 0 \mid Z_{ij,n}, \gamma^*_{ij,n}) + (1 - r_1) P^*(G^*_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1) P^*(G^*_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1) E_{P^*}[P^*(G^*_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}, \varepsilon_{ij}) \mid Z_{ij,n}, \gamma^*_{ij,n}] \\
&= r_0 + (1 - r_0 - r_1) P^*((Z^*_{ij,n})' b + \varepsilon_{ij} \geq 0 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1) \Phi((Z^*_{ij,n})' b) \\
&= r_0 + (1 - r_0 - r_1) \Phi((c(r_0, r_1) + C(r_0, r_1)\gamma_{ij,n})' b_1 + X'_{ij} b_2) \\
&= P(G_{ij,n} = 1 \mid Z_{ij,n}),
\end{aligned}$$

where the first and seventh equalities follow from Eq. (6), the second follows from Eq. (5), the fifth follows from Eq. (4), and the last follows from Eq. (1). □

A.3 Proof of Theorem 2

Theorem 2 follows because Lemma 10 and 11 in this appendix imply that, conditional on (X, σ_n), $n\hat{m}_n(\theta)' \hat{S}(\theta)^{-1} \hat{m}_n(\theta)$ converges in distribution to the χ^2_J distribution.

In the proof of this theorem, all the statements are conditional on (X, σ_n). For any vector, the norm is understood as the Euclidean norm, and for any matrix the norm is induced by the Euclidean norm.

Define:

$$\begin{aligned}
u_{ij}(\theta) &= (c(r_0, r_1) + C(r_0, r_1)\gamma_{ij,n})' b_1 + X'_{ij} b_2 \\
\hat{u}_{ij}(\theta) &= (c(r_0, r_1) + C(r_0, r_1)\hat{\gamma}_{ij,n})' b_1 + X'_{ij} b_2.
\end{aligned}$$

For a generic random variable RV, define:

$$RV^\dagger = RV - E[RV \mid X, \sigma_n],$$

and note that $E[RV^\dagger \mid X, \sigma_n] = 0$. Define:

$$
\begin{aligned}
\psi_{\gamma,k,n}(x) &= \frac{1}{n^2}\sum_{i,j}\left(\frac{1\{X_{i,j}=x\}}{\hat{p}(x)}\right)\begin{pmatrix} 0 \\ G^\dagger_{kj,n} \\ (G_{ki,n}G_{kj,n})^\dagger \\ (G_{ki,n}+G_{kj,n})^\dagger \end{pmatrix} + \frac{1}{n}\sum_i \left(\frac{1\{X_{i,k}=x\}}{\hat{p}(x)}\right)\begin{pmatrix} G^\dagger_{ki,n} \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
\psi_k(\theta_0) &= \frac{1}{n}\sum_{j\neq k}\left(G_{kj,n}-\rho_0-(1-\rho_0-\rho_1)\Phi(u_{kj}(\theta_0))\right)\zeta_{kj} \\
&\quad -(1-\rho_0-\rho_1)\frac{1}{n^2}\sum_{i,j}\left(\phi(u_{kj}(\theta_0))\beta_1' C(\rho_0,\rho_1)\psi_{\gamma,k,n}(X_{ij})\right)\zeta_{ij} \\
\tilde{\psi}_k(\theta_0) &= \frac{1}{n}\sum_{j\neq k} G_{kj,n}\zeta_{kj} - (1-\rho_0-\rho_1)\frac{1}{n^2}\sum_{i,j}\left(\phi(u_{ij}(\theta_0))\beta_1' C(\rho_0,\rho_1)\hat{\psi}_{\gamma,k,n}(X_{ij})\right)\zeta_{ij}.
\end{aligned}
$$

Lemma 4.

$$
1\{X_{i_1,j_1}=X_{ij}\}\begin{pmatrix} E[G^*_{j_1 i_1,n}\mid X,\sigma_n]-E[G^*_{ji,n}\mid X,\sigma_n] \\ \frac{1}{n}\sum_k (E[G^*_{kj_1,n}\mid X,\sigma_n]-E[G^*_{kj,n}\mid X,\sigma_n]) \\ \frac{1}{n}\sum_k (E[G^*_{ki_1,n}G^*_{kj_1,n}\mid X,\sigma_n]-E[G^*_{ki,n}G^*_{kj,n}\mid X,\sigma_n]) \\ \frac{1}{n}\sum_k \left(E\left[G^*_{ki_1,n}+G^*_{kj_1,n}\mid X,\sigma_n\right]-E\left[G^*_{ki,n}+G^*_{kj,n}\mid X,\sigma_n\right]\right) \end{pmatrix} = 0. \tag{7}
$$

Proof. This result follows from symmetry of the equilibrium and it is shown in a similar way to Lemma 1 in Leung (2015). □

Lemma 5.

$$
\max\{\|\hat{\psi}_{\gamma,k,n}(X_{ij})\|, \|\psi_{\gamma,k,n}(X_{ij})\|\} \le \frac{4}{\min_x \hat{p}(x)}
$$

$$
\max_i\{\|\tilde{\psi}_i(\theta_0)\|, \|\hat{\psi}_i(\theta_0)\|, \|\psi_i(\theta_0)\|\} \le 1+(1-\rho_0-\rho_1)\phi(0)\|\beta_1' C(\rho_0,\rho_1)\|\frac{4}{\min_x \hat{p}(x)}.
$$

Proof. The bound for $\|\hat{\psi}_{\gamma,k,n}(X_{ij})\|$ is derived from

$$
\begin{aligned}
\|\hat{\psi}_{\gamma,k,n}(x)\| &\le \frac{1}{n^2}\sum_{i_1,j_1}\frac{1\{X_{i_1,j_1}=x\}}{\hat{p}(x)}\left\|\begin{pmatrix} 0 \\ G_{kj_1} \\ G_{ki_1}G_{kj_1} \\ G_{ki_1}+G_{kj_1}\end{pmatrix}\right\| + \frac{1}{n}\sum_{i_1}\frac{1\{X_{i_1,k}=x\}}{\hat{p}(x)}\left\|\begin{pmatrix} G_{ki_1} \\ 0 \\ 0 \\ 0 \end{pmatrix}\right\| \\
&\le \frac{\sqrt{6}}{\min_x \hat{p}(x)} + \frac{1}{\min_x \hat{p}(x)} \\
&\le \frac{4}{\min_x \hat{p}(x)}.
\end{aligned}
$$

The bound for $\|\psi_{\gamma,k,n}(X_{ij})\|$ is similarly derived.

The bound for $\|\tilde{\psi}_i(\theta)\|$ is derived from

$$
\begin{aligned}
\|\tilde{\psi}_i(\theta)\| &\le \max_{j\neq i}|G_{ij,n}| + \max_{l,j}\left|\phi(u_{lj}(\theta_0))\beta_1' C(\rho_0,\rho_1)\hat{\psi}_{\gamma,i,n}(X_{lj})\right| \\
&\le 1+(1-\rho_0-\rho_1)\phi(0)\|\beta_1' C(\rho_0,\rho_1)\|\frac{4}{\min_x \hat{p}(x)}.
\end{aligned}
$$

The bound for $\|\hat{\psi}_i(\theta)\|$ is similarly derived.

The bound for $\|\psi_i(\theta)\|$ is derived from:

$$\begin{aligned}
\|\psi_i(\theta_0)\| &\leq \max_{j\neq i}|G_{ij,n}-\rho_0-(1-\rho_0-\rho_1)\Phi(u_{ij}(\theta_0))| \\
&\quad +(1-\rho_0-\rho_1)\max_{l,j}\|\phi(u_{ij}(\theta_0))\beta_1' C(\rho_0,\rho_1)\|\|\psi_{\gamma,i,n}(X_{lj})\| \\
&\leq 1+(1-\rho_0-\rho_1)\phi(0)\|\beta_1' C(\rho_0,\rho_1)\|\frac{4}{\min_x \hat{p}(x)}.
\end{aligned}$$

$\square$

Lemma 6.

$$\hat{\gamma}_{ij}-\gamma_{ij,n}=\frac{1}{n}\sum_k \psi_{\gamma,k,n}(X_{ij})$$

and

$$\sup_{i,j}\|\hat{\gamma}_{ij}-\gamma_{ij,n}\| = O_p(n^{-1/2}) \text{ given } (X,\sigma_n).$$

Proof. First, from Lemma 4 and Assumption 3, we can derive:

$$1\{X_{i_1,j_1}=X_{ij}\}\begin{pmatrix} E[G_{j_1 i_1,n}\mid X,\sigma_n]-E[G_{ji,n}\mid X,\sigma_n] \\ \frac{1}{n}\sum_k(E[G_{kj_1,n}\mid X,\sigma_n]-E[G_{kj,n}\mid X,\sigma_n]) \\ \frac{1}{n}\sum_k(E[G_{ki_1,n}G_{kj_1,n}\mid X,\sigma_n]-E[G_{ki,n}G_{kj,n}\mid X,\sigma_n]) \\ \frac{1}{n}\sum_k(E[(G_{ki_1,n}+G_{kj_1,n})\mid X,\sigma_n]-E[(G_{ki,n}+G_{kj,n})\mid X,\sigma_n]) \end{pmatrix}=0. \tag{8}$$

Using Eq. (8), we have:

$$\begin{aligned}
\hat{\gamma}_{ij}-\gamma_{ij,n} &= \frac{1}{n^2}\sum_{i_1,j_1}\frac{1\{X_{i_1,j_1}=X_{ij}\}}{\frac{1}{n^2}\sum_{i_1,j_1}1\{X_{i_1,j_1}=X_{ij}\}}\begin{pmatrix} G_{j_1 i_1,n}-E[G_{ji,n}\mid X,\sigma_n] \\ \frac{1}{n}\sum_k(G_{kj_1,n}-E[G_{kj,n}\mid X,\sigma_n]) \\ \frac{1}{n}\sum_k(G_{ki_1,n}G_{kj_1,n}-E[G_{ki,n}G_{kj,n}\mid X,\sigma_n]) \\ \frac{1}{n}\sum_k((G_{ki_1,n}+G_{kj_1,n})-E[(G_{ki,n}+G_{kj,n})\mid X,\sigma_n]) \end{pmatrix} \\
&= \frac{1}{n^2}\sum_{i_1,j_1}\frac{1\{X_{i_1,j_1}=X_{ij}\}}{\frac{1}{n^2}\sum_{i_1,j_1}1\{X_{i_1,j_1}=X_{ij}\}}\begin{pmatrix} G^\dagger_{j_1 i_1,n} \\ \frac{1}{n}\sum_k G^\dagger_{kj_1,n} \\ \frac{1}{n}\sum_k(G_{ki_1,n}G_{kj_1,n})^\dagger \\ \frac{1}{n}\sum_k((G_{ki_1,n}+G_{kj_1,n}))^\dagger \end{pmatrix} \\
&= \frac{1}{n}\sum_k \psi_{\gamma,k,n}(X_{ij}).
\end{aligned}$$

Since X_{ij} has a finite support, the uniform convergence over i,j follows from the point convergence for every i,j. By Lyapunov's central limit theorem, it suffices to show that $E[\psi_{\gamma,k,n}(X_{ij})\mid X,\sigma_n]=0$ and that

$\psi_{\gamma,k,n}(X_{ij})$ is independent across k given (X, σ_n). The equality $E[\psi_{\gamma,k,n}(X_{ij}) \mid X, \sigma_n] = 0$ follows from

$$
\begin{aligned}
E[\psi_{\gamma,k,n}(X_{ij}) \mid X, \sigma_n] &= \frac{1}{n^2}\sum_{i_1,j_1}\left(\frac{1\{X_{i_1,j_1} = X_{ij}\}}{\hat{p}(X_{ij})}\right)\begin{pmatrix} 0 \\ E\left[G^{\dagger}_{kj_1,n} \mid X, \sigma_n\right] \\ E\left[(G_{ki_1,n}G_{kj_1,n})^{\dagger} \mid X, \sigma_n\right] \\ E\left[(G_{ki_1,n} + G_{kj_1,n})^{\dagger} \mid X, \sigma_n\right] \end{pmatrix} \\
&\quad + \frac{1}{n}\sum_{i_1}\left(\frac{1\{X_{i_1,k} = X_{ij}\}}{\hat{p}(X_{ij})}\right)\begin{pmatrix} E\left[G^{\dagger}_{ki_1,n} \mid X, \sigma_n\right] \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&= 0
\end{aligned}
$$

since $E\left[RV^{\dagger} \mid X, \sigma_n\right] = 0$ by definition of $RV^{\dagger}$. The conditional independence of $\psi_{\gamma,k,n}(X_{ij})$ across k is shown as follows. Note that $\psi_{\gamma,k,n}(X_{ij})$ does not depend on $G_{-k,n}$, so it is a function of ε_k, (X, σ_n). Therefore, it follows from Assumptions 1 that $\psi_{\gamma,k,n}(X_{ij})$ is independent across k given (X, σ_n). $\square$

Lemma 7. $\max_i \|\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0)\| = o_p(1)$ *given* (X, σ_n).

Proof. Note that

$$\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0) \quad = \quad -(1-\rho_0-\rho_1)\frac{1}{n^2}\sum_{l,j}\left(\phi(\hat{u}_{lj}(\theta_0)) - \phi(u_{lj}(\theta_0))\right)\beta_1' C(\rho_0,\rho_1)\hat{\psi}_{\gamma,i,n}(X_{lj})\zeta_{lj}.$$

Then

$$
\begin{aligned}
\|\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0)\| &\leq \|\beta_1' C(\rho_0,\rho_1)\| \max_{l,j} |\phi(\hat{u}_{lj}(\theta_0)) - \phi(u_{lj}(\theta_0))| \, \|\hat{\psi}_{\gamma,i,n}(X_{lj})\| \\
&\leq \phi(0)\|\beta_1' C(\rho_0,\rho_1)\| \max_{l,j} \max\{|\hat{u}_{lj}(\theta_0)|, |u_{lj}(\theta_0)|\}|\hat{u}_{lj}(\theta_0) - u_{lj}(\theta_0)|\|\hat{\psi}_{\gamma,i,n}(X_{lj})\|,
\end{aligned}
$$

where the last inequality follows from the mean value expansion of the normal pdf ϕ: $|\phi(u_1) - \phi(u_2)| \leq \max_{u_1 \leq u \leq u_2} |\phi'(u)||u_1 - u_2| \leq \phi(0)\max\{|u_1|, |u_2|\}|u_1 - u_2|$. Since

$$
\begin{aligned}
|u_{lj}(\theta_0)| &\leq (\|c(\rho_0,\rho_1)\| + \|C(\rho_0,\rho_1)\|\|\gamma_{lj,n})\|\|\beta_1\| + \|X_{lj}\|\|\beta_2\| \\
&\leq (\|c(\rho_0,\rho_1)\| + 4\|C(\rho_0,\rho_1)\|)\|\beta_1\| + \max_x \|x\|\|\beta_2\| \\
|\hat{u}_{lj}(\theta_0)| &\leq (\|c(\rho_0,\rho_1)\| + 4\|C(\rho_0,\rho_1)\|)\|\beta_1\| + \max_x \|x\|\|\beta_2\| \\
|\hat{u}_{lj}(\theta_0) - u_{lj}(\theta_0)| &= |C(\rho_0,\rho_1)(\hat{\gamma}_{lj} - \gamma_{lj,n})'\beta_1| \\
&\leq \|C(\rho_0,\rho_1)\|\|\beta_1\| \max_{lj} \|\hat{\gamma}_{lj} - \gamma_{lj,n}\|,
\end{aligned}
$$

it follows that

$$\max_i \|\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0)\| = O_p(\max_{lj} \|\hat{\gamma}_{lj} - \gamma_{lj,n})\|) = o_p(1).$$

$\square$

Lemma 8. $\psi_i(\theta_0)$ *is independent across* i *given* (X, σ_n).

Proof. $\psi_i(\theta_0)$ does not depend on $G_{-i,n}$, so it is a function of $(\varepsilon_i, X, \sigma_n)$. By the independence of ε_i across i, it implies the statement of this lemma. $\square$

Lemma 9. *$\hat{m}_n(\theta_0) = \frac{1}{n}\sum_{i=1}^n \psi_i(\theta_0) + o_p(n^{-1/2})$ given (X, σ_n).*

Proof. Note that

$$\hat{m}_n(\theta_0) - \frac{1}{n}\sum_{i=1}^n \psi_i(\theta_0) = (1-\rho_0-\rho_1)\frac{1}{n^2}\sum_{i,j}\left(\Phi(\hat{u}_{ij}(\theta_0)) - \Phi(u_{ij}(\theta_0)) - \phi(u_{ij}(\theta_0))\beta_1' C(\rho_0,\rho_1)(\hat{\gamma}_{ij} - \gamma_{ij,n})\right)\zeta_{ij}$$

By the second-order Taylor expansion of the normal cdf Φ,

$$\Phi(u_1) = \Phi(u_2) + \phi(u_2)(u_2 - u_1) + R(u_1, u_2)$$

where

$$|R_{ij}| \leq \frac{1}{2}\max_{u_1 \leq u \leq u_2} \phi'(u)|u_1 - u_2|^2 \leq \frac{1}{2}\phi(0)\max\{|u_1|, |u_2|\}|u_1 - u_2|^2.$$

Since

$$\begin{aligned}
\max\{|u_{lj}(\theta_0)|, |\hat{u}_{lj}(\theta_0)|\} &\leq (\|c(\rho_0,\rho_1)\| + 4\|C(\rho_0,\rho_1)\|)\|\beta_1\| + \max_x \|x\|\|\beta_2\| \\
|\hat{u}_{lj}(\theta_0) - u_{lj}(\theta_0)| &\leq \|C(\rho_0,\rho_1)\|\|\beta_1\| \max_{lj} \|\hat{\gamma}_{lj} - \gamma_{lj,n}\|,
\end{aligned}$$

it follows that

$$\|\hat{m}_n(\theta_0) - \frac{1}{n}\sum_{i=1}^n \psi_i(\theta_0)\| = O_p(\max_{lj}\|\hat{\gamma}_{lj} - \gamma_{lj,n})\|^2) = O_p(n^{-1}).$$

$\square$

Lemma 10. *Conditional on (X, σ_n),*

$$\hat{m}_n(\theta_0) = o_P(1)$$

and

$$Var(\psi_i(\theta_0) \mid X, \sigma_n)^{-1/2}\sqrt{n}\hat{m}_n(\theta_0) \to_d N(0, I).$$

Proof. By Lemmas 5 and 8 and Lyapunov's central limit theorem, it suffices to show $E[\psi_i(\theta_0) \mid X, \sigma_n] = 0$. We can derive:

$$\begin{aligned}
E[\psi_i(\theta_0) \mid X, \sigma_n] &= \frac{1}{n}\sum_{j \neq i}\left(E[G_{ij,n} \mid X, \sigma_n] - \rho_0 - (1-\rho_0-\rho_1)\Phi(u_{ij}(\theta_0))\right)\zeta_{ij} \\
&\quad -(1-\rho_0-\rho_1)\frac{1}{n^2}\sum_{l,j}\left(\phi(u_{lj}(\theta_0))\beta_1' C(\rho_0,\rho_1) E[\psi_{\gamma,i,n}(X_{lj}) \mid X, \sigma_n]\right)\zeta_{lj} \\
&= 0,
\end{aligned}$$

because

$$
\begin{aligned}
E[G_{ij,n} \mid X, \sigma_n] &= \rho_0 + (1-\rho_0-\rho_1)\Phi(u_{ij}(\theta_0)) \\
E[\psi_{\gamma,i,n}(X_{lj}) \mid X, \sigma_n] &= \frac{1}{n^2}\sum_{i_1,j_1}\left(\frac{1\{X_{i_1,j_1} = X_{ij}\}}{\hat{p}(X_{ij})}\right)\begin{pmatrix} 0 \\ E[G^{\dagger}_{kj_1,n} \mid X, \sigma_n] \\ E[(G_{ki_1,n}G_{kj_1,n})^{\dagger} \mid X, \sigma_n] \\ E[(G_{ki_1,n} + G_{kj_1,n})^{\dagger} \mid X, \sigma_n] \end{pmatrix} \\
&\quad + \frac{1}{n}\sum_{i_1}\left(\frac{1\{X_{i_1,k} = X_{ij}\}}{\hat{p}(X_{ij})}\right)\begin{pmatrix} E[G^{\dagger}_{ki_1,n} \mid X, \sigma_n] \\ 0 \\ 0 \\ 0 \end{pmatrix} \\
&= 0.
\end{aligned}
$$

Note that $E[RV^{\dagger} \mid X, \sigma_n] = 0$ by the definition of $RV^{\dagger}$. □

Lemma 11. *$\hat{S}(\theta_0) = Var(\psi_i(\theta_0) \mid X, \sigma_n) + o_p(1)$ given (X, σ_n).*

Proof. First, we are going to show that $\hat{S}(\theta_0) = \frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\tilde{\psi}_i(\theta_0)' - \left(\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\right)\left(\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\right)' + o_p(1)$. Since

$$
\begin{aligned}
&\hat{S}(\theta_0) - \frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\tilde{\psi}_i(\theta_0)' + \left(\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\right)\left(\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\right)' \\
= \ &\frac{1}{n}\sum_{i=1}^n (\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0))(\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0))' \\
&+ \frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)(\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0))' \\
&+ \frac{1}{n}\sum_{i=1}^n (\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0))\tilde{\psi}_i(\theta_0)' \\
&- \left(\frac{1}{n}\sum_{i=1}^n (\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0))\right)\left(\frac{1}{n}\sum_{i=1}^n (\hat{\psi}_i(\theta_0))\right)' \\
&- \left(\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\right)\left(\frac{1}{n}\sum_{i=1}^n (\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0))\right)',
\end{aligned}
$$

it follows that:

$$
\begin{aligned}
&\left\| \hat{S}(\theta_0) - \frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\tilde{\psi}_i(\theta_0)' + \left(\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\right)\left(\frac{1}{n}\sum_{i=1}^n \tilde{\psi}_i(\theta_0)\right)' \right\| \\
&\quad \leq \max_i \|\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0)\|^2 + 3\max_i \|\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0)\| \max_i \|\tilde{\psi}_i(\theta_0)\| + \max_i \|\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0)\| \max_i \|\hat{\psi}_i(\theta_0)\|.
\end{aligned}
$$

Thus it suffices to show that $\max_i \|\hat{\psi}_i(\theta_0) - \tilde{\psi}_i(\theta_0)\| = o_p(1)$ and $\max_i\{\|\tilde{\psi}_i(\theta_0)\|, \|\hat{\psi}_i(\theta_0)\|\} = O_p(1)$. They are shown in Lemmas 5 and 7.

Second, we are going to show that $\hat{S}(\theta_0) = Var(\tilde{\psi}_i(\theta_0) \mid X, \sigma_n) + o_p(1)$. It suffices to show $E[\|\tilde{\psi}_i(\theta_0)\|^4 \mid$

$X, \sigma_n] < \infty$. By the triangle inequality,

$$
\begin{aligned}
E[\|\tilde{\psi}_i(\theta_0)\|^4 \mid X, \sigma_n]^{1/4} &\leq \frac{1}{n}\sum_{j\neq i} E\left[\|G_{ij,n}\|^4 \mid X, \sigma_n\right]^{1/4} \\
&\quad + \frac{1}{n^2}\sum_{l,j} E\left[\left\|\phi(u_{ij}(\theta_0))\beta_1' C(\rho_0,\rho_1)\hat{\psi}_{\gamma,i,n}(X_{lj})\right\|^4 \mid X, \sigma_n\right]^{1/4} \\
&\leq \frac{1}{n}\sum_{j\neq i}\left(E[\|G_{ij,n}\|^4 \mid X, \sigma_n]^{1/4}\right) \\
&\quad + \frac{1}{n^2}\sum_{l,j}\phi(u_{ij}(\theta_0))\beta_1' C(\rho_0,\rho_1) E\left[\left\|\hat{\psi}_{\gamma,i,n}(X_{lj})\right\|^4 \mid X, \sigma_n\right]^{1/4} \\
&\leq 1 + \frac{1}{n^2}\sum_{l,j}\phi(u_{ij}(\theta_0))\beta_1' C(\rho_0,\rho_1) E\left[\left\|\hat{\psi}_{\gamma,i,n}(X_{lj})\right\|^4 \mid X, \sigma_n\right]^{1/4} \\
&< \infty,
\end{aligned}
$$

where the last inequality follows from Lemma 5.

Third, we show that $Var(\tilde{\psi}_i(\theta_0) \mid X, \sigma_n) = Var(\psi_i(\theta_0) \mid X, \sigma_n)$. Note that $\tilde{\psi}_i(\theta_0) - \psi_i(\theta_0)$ is a function of (X, σ_n), so the conditional variances are the same. □

A.4 Proof of Theorem 3

As in the previous section, all the statements in this appendix are conditional on (X, σ_n). Theorem 3 follows from Lemma 18.

Lemma 12. *β is the unique maximizer of $E[\boldsymbol{Q}_n(b) \mid X, \sigma_n]$, where*

$$\boldsymbol{Q}_n(b) = \frac{1}{n^2}\sum_{i,j}\log\left(\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})^{G_{ij,n}}(1-\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n}))^{1-G_{ij,n}}\right).$$

Proof. Applying Jensen's inequality to the logarithm function, we have:

$$
\begin{aligned}
& E[\boldsymbol{Q}_n(b) \mid X, \sigma_n] - E[\boldsymbol{Q}_n(\beta) \mid X, \sigma_n] \\
= \ & \frac{1}{n^2}\sum_{i,j}\left(\Psi(\theta_0,X_{ij},\gamma_{ij,n})\log\frac{\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})} + (1-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))\log\frac{1-\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})}{1-\Psi(\theta_0,X_{ij},\gamma_{ij,n})}\right) \\
\leq \ & \log\left(\frac{1}{n^2}\sum_{i,j}\left(\Psi(\theta_0,X_{ij},\gamma_{ij,n})\frac{\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})} + (1-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))\frac{1-\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})}{1-\Psi(\theta_0,X_{ij},\gamma_{ij,n})}\right)\right) \\
= \ & 0.
\end{aligned}
$$

It suffices to show that the equality holds only when $b = \beta$. By Jensen's inequality, the equality holds if and only if:

$$\frac{\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})} = 1 \text{ for every } i,j. \tag{9}$$

Eq. (9) implies $((\gamma^*_{ij,n})', X'_{ij})\beta = ((\gamma^*_{ij,n})', X'_{ij})b$ for every i,j. Since $\{((\gamma^*_{ij,n})', X'_{ij})' : i,j\}$ is not contained in any proper linear subspace of $\mathbb{R}^{d+3}$, we have $\beta = b$. □

Lemma 13. *Conditional on* (X, σ_n),

$$\sup_{b\in\mathcal{B}} |\boldsymbol{Q}_n(b) - E\left[\boldsymbol{Q}_n(b) \mid X, \sigma_n\right]| = o_p(1)$$

$$\sup_{b\in\mathcal{B}} \| \frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) - E\left[\frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) \mid X, \sigma_n\right] \| = o_p(1).$$

Proof. They follow from Jenish and Prucha (2009, Proposition 1) as in the proof of Leung (2015, Theorem 2). □

Lemma 14. $\hat{\beta}(\rho_0, \rho_1) \to_{a.s.} \beta$.

Proof. By Lemma 12 and Gallant and White (1988, Theorem 3.3), it suffices to show that

$$\sup_{b\in\mathcal{B}} |\hat{\boldsymbol{Q}}_n(b, \rho_0, \rho_1) - E\left[\boldsymbol{Q}_n(b) \mid X, \sigma_n\right]| \to_p 0.$$

By Lemma 13, we need to show that $\sup_{b\in\mathcal{B}} |\hat{\boldsymbol{Q}}_n(b, \rho_0, \rho_1) - \boldsymbol{Q}_n(b)| \to_p 0$. Some calculations yield:

$$\begin{aligned}
& |\hat{\boldsymbol{Q}}_n(b, \rho_0, \rho_1) - \boldsymbol{Q}_n(b)| \\
= & \left| \frac{1}{n^2} \sum_{i,j} \log \left(\left(\frac{\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij}))}{\Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})} \right)^{G_{ij,n}} \left(\frac{1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij}))}{1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})} \right)^{1-G_{ij,n}} \right) \right| \\
\leq & \max_{i,j} \max \left\{ \left| \log \left(\frac{\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij}))}{\Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})} \right) \right|, \left| \log \left(\frac{1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij}))}{1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})} \right) \right| \right\} \\
\leq & \max_{i,j} \frac{|\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij})) - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})|}{\min\{\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij})), \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n}), 1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij})), 1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})\}} \\
\leq & \max_{i,j} \frac{|\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij})) - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})|}{\min\{\Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n}), 1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})\} - |\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij})) - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})|} \\
\leq & \frac{\text{term1}}{\text{term2} - \text{term1}}
\end{aligned}$$

where the second inequality follows from $|\log(x)| \leq \max\{|x - 1|, |x - 1|/x\}$ for $x > 0$ and the last equation uses

$$\begin{aligned}
\text{term1} &= \max_{i,j} |\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij})) - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})| \\
\text{term2} &= \min_{i,j} \min\{\Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n}), 1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})\}.
\end{aligned}$$

Since $\min_{i,j} \min\{\Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n}), 1 - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})\}$ is bounded away from zero (because the support of X_{ij} is finite), the uniform convergence of $\hat{\boldsymbol{Q}}_n(b, \rho_0, \rho_1) - \boldsymbol{Q}_n(b)$ follows from:

$$\begin{aligned}
& \sup_{b\in\mathcal{B}} \max_{i,j} |\Psi(b, \rho_0, \rho_1, X_{ij}, \hat{\gamma}(X_{ij})) - \Psi(b, \rho_0, \rho_1, X_{ij}, \gamma_{ij,n})| \\
& \quad = (1 - \rho_0 - \rho_1) \sup_{b\in\mathcal{B}} \max_{i,j} |\Phi(\hat{u}_{ij}(b, \rho_0, \rho_1)) - \Phi(u_{ij}(b, \rho_0, \rho_1))| \\
& \quad = O_p\left(\| \max_{ij} \|\hat{\gamma}_{ij} - \gamma_{ij,n}\| \right).
\end{aligned}$$

□

Lemma 15. *The minimum eigenvalue of* $\{ E\left[\frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) \mid X, \sigma_n\right]\Big|_{b=\beta}\}$ *is bounded away from zero.*

Proof. We have the following equalities:

$$
\begin{aligned}
E\left[\frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) \mid X, \sigma_n\right]\Bigg|_{b=\beta} &= \frac{1}{n^2}\sum_{i,j} \frac{\frac{\partial}{\partial b}\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})\big|_{b=\beta}\ \frac{\partial}{\partial b'}\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})\big|_{b=\beta}}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})(1-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))} \\
&= \frac{1}{n^2}\sum_{i,j} \frac{\frac{\partial}{\partial b}\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})\big|_{b=\beta}\ \frac{\partial}{\partial b'}\Psi(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})\big|_{b=\beta}}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})(1-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))} \\
&= \frac{(1-\rho_0-\rho_1)^2}{n^2}\sum_{i,j} \frac{\phi((Z^*_{ij,n})'\beta_0)^2}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})(1-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))} Z^*_{ij,n}(Z^*_{ij,n})'.
\end{aligned}
$$

Note that the minimum eigenvalue of $\sum_{i,j} Z^*_{ij,n}(Z^*_{ij,n})'$ is bounded away from zero. Since

$$
\frac{\phi((Z^*_{ij,n})'\beta_0)^2}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})(1-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))}
$$

is bounded from zero uniformly over i, j, n, the minimum eigenvalue of

$$
\sum_{i,j} \frac{\phi((Z^*_{ij,n})'\beta_0)^2}{\Psi(\theta_0,X_{ij},\gamma_{ij,n})(1-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))} Z^*_{ij,n}(Z^*_{ij,n})'
$$

is bounded away from zero. □

Lemma 16. $\sup_{b\in\mathcal{B}} \left\| E\left[\frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) \mid X, \sigma_n\right] - \frac{\partial^2}{\partial b \partial b'} \hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1)\right\| = o_p(1)$ *given* (X, σ_n).

Proof. By Lemma 13, we need to show that

$$
\sup_{b\in\mathcal{B}} \left\| \frac{\partial^2}{\partial b \partial b'} \hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1) - \frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) \right\| = o_p(1),
$$

that is,

$$
\sup_{b\in\mathcal{B}} \left\| \boldsymbol{u}'\left(\frac{\partial^2}{\partial b \partial b'} \hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1) - \frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) \right) \right\| = o_p(1) \text{ for every vector } \boldsymbol{u}.
$$

Since

$$
\begin{aligned}
\boldsymbol{u}'\left(\frac{\partial^2}{\partial b \partial b'} \hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1) - \frac{\partial^2}{\partial b \partial b'} \boldsymbol{Q}_n(b) \right) &= \frac{1}{n^2}\sum_{i,j} G_{ij,n}\boldsymbol{u}'\left(\frac{\partial}{\partial b'}(\boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij})) - \boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n}))\right) \\
&\quad - \frac{1}{n^2}\sum_{i,j} \boldsymbol{u}'\left(\frac{\partial}{\partial b'}(\boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij})) - \boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n}))\right) \\
&= \frac{1}{n^2}\sum_{i,j} G_{ij,n}\frac{\partial}{\partial b'}(\boldsymbol{u}'\boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij})) - \boldsymbol{u}'\boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})) \\
&\quad - \frac{1}{n^2}\sum_{i,j} \frac{\partial}{\partial b'}(\boldsymbol{u}'\boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij})) - \boldsymbol{u}'\boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n})),
\end{aligned}
$$

we have:

$$\begin{aligned}
\left\|\boldsymbol{u}'\left(\frac{\partial^2}{\partial b\partial b'}\hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1)-\frac{\partial^2}{\partial b\partial b'}\boldsymbol{Q}_n(b)\right)\right\| &\le \frac{1}{n^2}\sum_{i,j}\left\|\frac{\partial}{\partial b'}(\boldsymbol{u}'\boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij}))-\boldsymbol{u}'\boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n}))\right\| \\
&\quad +\frac{1}{n^2}\sum_{i,j}\left\|\frac{\partial}{\partial b'}(\boldsymbol{u}'\boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij}))-\boldsymbol{u}'\boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\gamma_{ij,n}))\right\| \\
&\le \frac{1}{n^2}\sum_{i,j}\sup_{\check{\gamma}_{ij}}\left\|\frac{\partial^2}{\partial b\partial\check{\gamma}_{ij}'}\boldsymbol{u}'\boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\check{\gamma}_{ij})\right\|\|\hat{\gamma}(X_{ij})-\gamma_{ij,n}\| \\
&\quad +\frac{1}{n^2}\sum_{i,j}\sup_{\check{\gamma}_{ij}}\left\|\frac{\partial^2}{\partial b\partial\check{\gamma}_{ij}'}\boldsymbol{u}'\boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\check{\gamma}_{ij})\right\|\|\hat{\gamma}(X_{ij})-\gamma_{ij,n}\| \\
&\le \sup_{i,j}\sup_{\check{\gamma}_{ij}}\left\|\frac{\partial^2}{\partial b\partial\check{\gamma}_{ij}'}\boldsymbol{u}'\boldsymbol{C}_1(b,\rho_0,\rho_1,X_{ij},\check{\gamma}_{ij})\right\|\sup_{i,j}\|\hat{\gamma}(X_{ij})-\gamma_{ij,n}\| \\
&\quad +\sup_{i,j}\sup_{\check{\gamma}_{ij}}\left\|\frac{\partial^2}{\partial b\partial\check{\gamma}_{ij}'}\boldsymbol{u}'\boldsymbol{C}_2(b,\rho_0,\rho_1,X_{ij},\check{\gamma}_{ij})\right\|\sup_{i,j}\|\hat{\gamma}(X_{ij})-\gamma_{ij,n}\|.
\end{aligned}$$

Since $\frac{\partial^2}{\partial b\partial\check{\gamma}_{ij}'}\boldsymbol{u}'\boldsymbol{C}_1$ and $\frac{\partial^2}{\partial b\partial\check{\gamma}_{ij}'}\boldsymbol{u}'\boldsymbol{C}_2$ have bounded supports, we have:

$$\sup_{b\in\mathcal{B}}\left\|\boldsymbol{u}'\left(\frac{\partial^2}{\partial b\partial b'}\hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1)-\frac{\partial^2}{\partial b\partial b'}\boldsymbol{Q}_n(b)\right)\right\| = O_p\left(\sup_{i,j}\|\hat{\gamma}(X_{ij})-\gamma_{ij,n}\|\right)=o_p(1).$$

□

Lemma 17.

$$\sqrt{n}E\left[\frac{1}{n}\sum_{k=1}^{n}\psi_{\boldsymbol{Q},k,n}\psi_{\boldsymbol{Q},k,n}'\mid X,\sigma_n\right]^{-1/2}E\left[\frac{\partial^2}{\partial b\partial b'}\boldsymbol{Q}_n(b)\mid X,\sigma_n\right]\Big|_{b=\beta}(\hat{\beta}(\rho_0,\rho_1)-\beta)\to_d N(0,I)\ \textit{given}\ (X,\sigma_n).$$

Proof. By Lemma 14, 15, 16 and Gallant and White (1988, Theorem 5.1), it suffices to prove the following statements:

- $E\left[\frac{\partial^2}{\partial b\partial b'}\boldsymbol{Q}_n(b)\mid X,\sigma_n\right]\Big|_{b=\beta}$ and $E\left[\frac{1}{n}\sum_{k=1}^{n}\psi_{\boldsymbol{Q},k,n}\psi_{\boldsymbol{Q},k,n}'\right]$ are $O(1)$;
- $E\left[\frac{\partial^2}{\partial b\partial b'}\boldsymbol{Q}_n(b)\mid X,\sigma_n\right]$ is continuous in $b\in\mathcal{B}$ uniformly in n; and
- $$\sqrt{n}E\left[\frac{1}{n}\sum_{k=1}^{n}\psi_{\boldsymbol{Q},k,n}\psi_{\boldsymbol{Q},k,n}'\mid X,\sigma_n\right]^{-1/2}\frac{\partial}{\partial b}\hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1)\Big|_{b=\beta}\to_d N(0,I) \tag{10}$$

The first two statements follow from the error being normally distributed.

Before proving Eq. (10), we will show that

$$\frac{1}{n^2}\sum_{i,j}(G_{ij,n}-E\left[G_{ij,n}\mid X,\sigma_n\right])\boldsymbol{D}_1(\theta_0,X_{ij},\gamma_{ij,n})(\hat{\gamma}_{ij}-\gamma_{ij,n})=o_p(n^{-1/2}). \tag{11}$$

Note that:

$$\begin{aligned}
& \frac{1}{n^2}\sum_{i,j}(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right])\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})(\hat{\gamma}_{ij} - \gamma_{ij,n}) \\
= & \frac{1}{n^2}\sum_{i,j}(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right])\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})\frac{1}{n}\sum_{k}\psi_{\gamma,k,n}(X_{ij}) \\
= & \frac{1}{n^2}\sum_{i,k}\text{term}(i,k),
\end{aligned}$$

where:

$$\text{term}(i,k) = \frac{1}{n}\sum_{j}(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right])\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})\psi_{\gamma,k,n}(X_{ij}).$$

We will demonstrate the L^2 convergence of $\frac{1}{n^2}\sum_{i,k}\text{term}(i,k)$. Regarding the expectation of $\frac{1}{n^2}\sum_{i,k}\text{term}(i,k)$, we have

$$\begin{aligned}
& E\left[\frac{1}{n^2}\sum_{i,k}\text{term}(i,k) \mid X, \sigma_n\right] \\
= & \frac{1}{n^2}\sum_{i\neq k}\frac{1}{n}\sum_{j}E\left[(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right])\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})\psi_{\gamma,k,n}(X_{ij}) \mid X, \sigma_n\right] \\
& +\frac{1}{n^2}\sum_{i}\frac{1}{n}\sum_{j}E\left[(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right])\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})\psi_{\gamma,i,n}(X_{ij}) \mid X, \sigma_n\right] \\
= & \frac{1}{n^2}\sum_{i\neq k}\frac{1}{n}\sum_{j}E\left[(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right]) \mid X, \sigma_n\right]\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})E\left[\psi_{\gamma,k,n}(X_{ij}) \mid X, \sigma_n\right] \\
& +\frac{1}{n^2}\sum_{i}\frac{1}{n}\sum_{j}E\left[(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right])\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})\psi_{\gamma,i,n}(X_{ij}) \mid X, \sigma_n\right] \\
= & \frac{1}{n^2}\sum_{i}\frac{1}{n}\sum_{j}E\left[(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right])\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})\psi_{\gamma,i,n}(X_{ij}) \mid X, \sigma_n\right] \\
= & O(n^{-1}).
\end{aligned}$$

where the second equality comes from the independence of $\{G_{ij,n} : j\}$ across i, the third equality follows from $E\left[(G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right]) \mid X, \sigma_n\right] = 0$, and the last equality follows because $G_{ij,n}$, $\boldsymbol{D}_1(\theta_0, X_{ij}, \gamma_{ij,n})$, and $\psi_{\gamma,i,n}(X_{ij})$ are bounded. Regarding the variance of $\frac{1}{n^2}\sum_{i,k}\text{term}(i,k)$, we use:

$$Cov\left(\text{term}(i_1,k_1), \text{term}(i_2,k_2) \mid X, \sigma_n\right) = 0 \text{ if } k_2 \neq i_1, k_1, i_2 \tag{12}$$

$$Cov\left(\text{term}(i_1,k_1), \text{term}(i_2,k_2) \mid X, \sigma_n\right) = 0 \text{ if } k_1 \neq i_1, k_2, i_2 \tag{13}$$

$$Cov\left(\text{term}(i_1,k_1), \text{term}(i_2,k_1) \mid X, \sigma_n\right) = 0 \text{ if } i_1 \neq k_1, i_2, \tag{14}$$

where they result from the fact that $G_{ij,n} - E\left[G_{ij,n} \mid X, \sigma_n\right]$ and $\psi_{\gamma,k,n}(X_{ij})$ are mean-zero and independent across (i,k). We have:

$$\begin{aligned}
Var\left(\frac{1}{n^2}\sum_{i,k}\text{term}(i,k)\mid X,\sigma_n\right) &= \frac{1}{n^4}\sum_{(i_1,k_1,i_2,k_2)} Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_2)\mid X,\sigma_n\right)\\
&= \frac{1}{n^4}\sum_{(i_1,k_1,i_2)}\sum_{k_2=i_1,k_1,i_2} Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_2)\mid X,\sigma_n\right)\\
&= \frac{1}{n^4}\sum_{(i_1,i_2)}\sum_{k_2=i_1,i_2}\sum_{k_1} Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_2)\mid X,\sigma_n\right)\\
&\quad +\frac{1}{n^4}\sum_{(k_1,i_2)}\sum_{i_1} Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_1)\mid X,\sigma_n\right)\\
&= \frac{1}{n^4}\sum_{(i_1,i_2)}\sum_{k_2=i_1,i_2}\sum_{k_1=i_1,i_2} Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_2)\mid X,\sigma_n\right)\\
&\quad +\frac{1}{n^4}\sum_{(k_1,i_2)}\sum_{i_1=k_1,i_2} Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_1)\mid X,\sigma_n\right)\\
&\leq \frac{6}{n^2}\max_{(i_1,k_1,i_2,k_2)}\left|Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_2)\mid X,\sigma_n\right)\right|\\
&= O(n^{-2}),
\end{aligned}$$

where the third equality uses Eq.(12), the fifth equality uses Eq.(13) and Eq.(14), and the last equality follows from $\sup_{(i_1,k_1,i_2,k_2)}\left|Cov\left(\text{term}(i_1,k_1),\text{term}(i_2,k_2)\mid X,\sigma_n\right)\right| = O(1)$.

Now we show that Eq. (11) implies Eq. (10). The first-order Taylor expansions yield

$$\begin{aligned}
\sup_{i,j}\|\boldsymbol{C}_1(\theta_0,X_{ij},\hat{\gamma}_{ij})-\boldsymbol{C}_1(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{D}_1(\theta_0,X_{ij},\gamma_{ij,n})(\hat{\gamma}_{ij}-\gamma_{ij,n})\| &= o_p\left(\sup_{i,j}\|\hat{\gamma}_{ij}-\gamma_{ij,n}\|\right)\\
\sup_{i,j}\|\boldsymbol{C}_2(\theta_0,X_{ij},\hat{\gamma}_{ij})-\boldsymbol{C}_2(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{D}_2(\theta_0,X_{ij},\gamma_{ij,n})(\hat{\gamma}_{ij}-\gamma_{ij,n})\| &= o_p\left(\sup_{i,j}\|\hat{\gamma}_{ij}-\gamma_{ij,n}\|\right),
\end{aligned}$$

and

$$
\begin{aligned}
&\frac{\partial}{\partial b}\hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1)\Big|_{b=\beta} \\
&= \frac{1}{n^2}\sum_{i,j}(G_{ij,n}-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))\boldsymbol{C}_1(\theta_0,X_{ij},\gamma_{ij,n}) \\
&\quad+\frac{1}{n^2}\sum_{i,j}\left(G_{ij,n}\left(\boldsymbol{C}_1(\theta_0,X_{ij},\hat{\gamma}(X_{ij}))-\boldsymbol{C}_1(\theta_0,X_{ij},\gamma_{ij,n})\right)-\left(\boldsymbol{C}_2(\theta_0,X_{ij},\hat{\gamma}(X_{ij}))-\boldsymbol{C}_2(\theta_0,X_{ij},\gamma_{ij,n})\right)\right) \\
&= \frac{1}{n^2}\sum_{i,j}(G_{ij,n}-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))\boldsymbol{C}_1(\theta_0,X_{ij},\gamma_{ij,n}) \\
&\quad+\frac{1}{n^2}\sum_{i,j}\left(G_{ij,n}\boldsymbol{D}_1(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{D}_2(\theta_0,X_{ij},\gamma_{ij,n})\right)(\hat{\gamma}_{ij}-\gamma_{ij,n})+o_p(n^{-1/2}) \\
&= \frac{1}{n^2}\sum_{i,j}(G_{ij,n}-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))\boldsymbol{C}_1(\theta_0,X_{ij},\gamma_{ij,n}) \\
&\quad+\frac{1}{n^2}\sum_{i,j}\left(E\left[G_{ij,n}\mid X,\sigma_n\right]\boldsymbol{D}_1(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{D}_2(\theta_0,X_{ij},\gamma_{ij,n})\right)(\hat{\gamma}_{ij}-\gamma_{ij,n})+o_p(n^{-1/2}) \\
&= \frac{1}{n^2}\sum_{i,j}(G_{ij,n}-\Psi(\theta_0,X_{ij},\gamma_{ij,n}))\boldsymbol{C}_1(\theta_0,X_{ij},\gamma_{ij,n}) \\
&\quad+\frac{1}{n^2}\sum_{i,j}\left(E\left[G_{ij,n}\mid X,\sigma_n\right]\boldsymbol{D}_1(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{D}_2(\theta_0,X_{ij},\gamma_{ij,n})\right)\frac{1}{n}\sum_{k}\psi_{\gamma,k,n}(X_{ij})+o_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{k}\psi_{\boldsymbol{Q},k,n}+o_p(n^{-1/2}),
\end{aligned}
$$

where the third equality uses Eq. (11). We can apply Lyapunovs central limit theorem to uniformly-bounded random variables $\psi_{\boldsymbol{Q},k,n}$, and we have Eq. (10). □

Lemma 18. *$\sqrt{n}\widehat{AV}(\rho_0,\rho_1)^{-1/2}(\hat{\beta}(\rho_0,\rho_1)-\beta)\to_d N(0,I)$ given (X,σ_n).*

Proof. By Lemma 17, it is sufficient to show that

$$
\begin{aligned}
\frac{\partial^2}{\partial b\partial b'}\hat{\boldsymbol{Q}}_n(b,\rho_0,\rho_1)\Big|_{b=\hat{\beta}(\rho_0,\rho_1)}-E\left[\frac{\partial^2}{\partial b\partial b'}\boldsymbol{Q}_n(b)\mid X,\sigma_n\right]\Big|_{b=\beta} &= o_p(1) \\
\frac{1}{n}\sum_{k=1}^{n}\hat{\psi}_{\boldsymbol{Q},k,n}(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1)\hat{\psi}_{\boldsymbol{Q},k,n}(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1)'-E\left[\frac{1}{n}\sum_{k=1}^{n}\psi_{\boldsymbol{Q},k,n}\psi'_{\boldsymbol{Q},k,n}\mid X,\sigma_n\right] &= o_p(1).
\end{aligned}
$$

The first statement follows from Lemma 16. For the rest of the proof, we will show the second statement.

First, we show that

$$
\frac{1}{n}\sum_{k=1}^{n}\hat{\psi}_{\boldsymbol{Q},k,n}(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1)\hat{\psi}_{\boldsymbol{Q},k,n}(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1)'-\frac{1}{n}\sum_{k=1}^{n}\psi_{\boldsymbol{Q},k,n}\psi'_{\boldsymbol{Q},k,n}=o_p(1).
$$

Since $\psi_{\boldsymbol{Q},k,n}$ is uniformly bounded, it suffices to show that $\max_k\|\hat{\psi}_{\boldsymbol{Q},k,n}(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1)-\psi_{\boldsymbol{Q},k,n}\|=o_p(1)$.

This convergence follows from:

$$\max_{i,j,k}\left\|\psi_{\gamma,k,n}(X_{ij})-\hat{\psi}_{\gamma,k,n}(X_{ij})\right\|=o_p(1)$$
$$\max_{i,j}\left\|\Psi(\theta_0,X_{ij},\gamma_{ij,n})-\Psi(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij}))\right\|=o_p(1)$$
$$\max_{i,j}\left\|\boldsymbol{C}_1(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{C}_1(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{kj}))\right\|=o_p(1)$$
$$\max_{i,j}\left\|\boldsymbol{D}_1(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{D}_1(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij}))\right\|=o_p(1)$$
$$\max_{i,j}\left\|\boldsymbol{D}_2(\theta_0,X_{ij},\gamma_{ij,n})-\boldsymbol{D}_2(\hat{\beta}(\rho_0,\rho_1),\rho_0,\rho_1,X_{ij},\hat{\gamma}(X_{ij}))\right\|=o_p(1).$$

Note that the uniform convergence over (i,j,k) is equivalent to the pointwise convergence, since X_{ij} has a finite support.

Then, we can obtain $\frac{1}{n}\sum_{k=1}^{n}\psi_{\boldsymbol{Q},k,n}\psi'_{\boldsymbol{Q},k,n}-E\left[\frac{1}{n}\sum_{k=1}^{n}\psi_{\boldsymbol{Q},k,n}\psi'_{\boldsymbol{Q},k,n}\mid X,\sigma_n\right]=o_p(1)$, by applying the weak law of large numbers to uniformly-bounded random variables $\psi_{\boldsymbol{Q},k,n}$. □

B Semiparametric Identification Analysis

Given $P\in\mathcal{P}$, we will characterize the identified set in the semiparametric model.

Definition 2. *For each distribution $P\in\mathcal{P}$, the identified set $\Theta_{I,SP}(P)$ is defined as the set of all $\theta=(b,r_0,r_1)$ in Θ for which there is some joint distribution $P^*\in\mathcal{P}^*$ such that Condition 1, 2(ii), and 3 holds, and that the distribution of $(G_{ij,n},X_{ij},\gamma_{ij,n})$ induced from P^* is equal to P.*

Theorem 4. *Given $P\in\mathcal{P}$, $\Theta_{I,SP}(P)$ is equal to the set of $\theta\in\Theta$ satisfying the following statements a.s. for some $r_0,r_1\geq 0$ such that $r_0+r_1<1$ and some weakly increasing and right-continuous function Λ:*

$$r_0\leq E_P\left[G_{ij,n}\mid Z_{ij,n}\right]\ \text{ and } r_1\leq E_P\left[1-G_{ij,n}\mid Z_{ij,n}\right]\tag{15}$$
$$E_P\left[G_{ij,n}\mid Z_{ij,n}\right]=\Lambda\left((c(r_0,r_1)+\gamma'_{ij,n}C(r_0,r_1))'b_1+X'_{ij}b_2\right).\tag{16}$$

Proof. First, we will show that every element θ of $\Theta_{I,SP}(P)$ satisfies the conditions in (15)-(16). Let (r_0,r_1) denote the misclassification probabilities. Denote by Λ^* the cdf of $-\varepsilon_{ij}$ and define $\Lambda(v)=r_0+(1-r_0-r_1)\Lambda^*(v)$. By Lemma 2 and 3:

$$\begin{aligned}E_{P^*}\left[G_{ij,n}\mid Z_{ij,n}\right] &= r_0+(1-r_0-r_1)E_{P^*}\left[G^*_{ij,n}\mid Z_{ij,n}\right]\\ &= r_0+(1-r_0-r_1)\Lambda^*((c(r_0,r_1)+\gamma'_{ij,n}C(r_0,r_1))'b_1+X'_{ij}b_2),\end{aligned}$$

and we have the condition (16). Note that Λ is weakly increasing and right-continuous, because Λ^* is weakly increasing and right-continuous. The two inequalities in (15) are shown as follows:

$$\begin{aligned}E_{P^*}\left[G_{ij,n}\mid Z_{ij,n}\right] &= r_0+(1-r_0-r_1)E_{P^*}\left[G^*_{ij,n}\mid Z_{ij,n}\right]\geq r_0\\ E_{P^*}\left[1-G_{ij,n}\mid Z_{ij,n}\right] &= r_1+(1-r_0-r_1)E_{P^*}\left[1-G^*_{ij,n}\mid Z_{ij,n}\right]\geq r_1,\end{aligned}$$

where the inequalities follow from $1-r_0-r_1\geq 0$.

Now, the rest of the proof will show that every element $\theta\in\Theta$ satisfying (15)-(16), belongs to $\Theta_{I,SP}(P)$. By the condition (16) as well as Condition (15), there is a weakly increasing and right-continuous function

$\Lambda : \mathbb{R} \to [r_0, 1-r_1]$ such that:

$$E_P\left[G_{ij,n} \mid Z_{ij,n}\right] = \Lambda\left((c(r_0,r_1) + \gamma'_{ij,n}C(r_0,r_1))'b_1 + X'_{ij}b_2\right). \tag{17}$$

Denote by Λ^* the cdf satisfying $\Lambda(v) = r_0 + (1 - r_0 - r_1)\Lambda^*(v)$.

Define the joint distribution P^* in the following way. Define the cdf of ε_{ij} such that Λ^* is the cdf of $-\varepsilon_{ij}$. The conditional distribution of $(\gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij})$ given ε_{ij} is

$$P^*((\gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}) \in B \mid \varepsilon_{ij}) = P((\gamma_{ij,n}, c(r_0,r_1) + C(r_0,r_1)\gamma_{ij,n}, X_{ij}) \in B) \tag{18}$$

for all the measurable sets B. The conditional distribution of $G^*_{ij,n}$ given $(\gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij})$ is:

$$P^*(G^*_{ij,n} = 1 \mid \gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij}) = 1\{(Z^*_{ij,n})'b + \varepsilon_{ij} \geq 0\}. \tag{19}$$

The conditional distribution of $G_{ij,n}$ given $(G^*_{ij,n}, \gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij})$ is:

$$P^*(G_{ij,n} = 1 \mid G^*_{ij,n}, \gamma_{ij,n}, \gamma^*_{ij,n}, X_{ij}, \varepsilon_{ij}) = \begin{cases} r_0 & \text{if } G^*_{ij,n} = 0 \\ 1 - r_1 & \text{if } G^*_{ij,n} = 1. \end{cases} \tag{20}$$

Note that (P^*, θ) satisfies Conditions 1(ii), 2 and 3, because Condition 1(ii) follows from Eq. (18), Condition 2 follows from Eq. (19), Condition 3(i) follows from Eq. (19) and (20), and Condition 3(ii) follows from Eq. (18).

To conclude this proof, we now show that the distribution of $(G_{ij,n}, X_{ij}, \gamma_{ij,n})$ induced from P^* is equal to P. The distribution of $(X_{ij}, \gamma_{ij,n})$ induced from P^* is equal to that from P, by Eq. (18). The equality of $P^*(G_{ij,n} = 1 \mid Z_{ij,n}) = P(G_{ij,n} = 1 \mid Z_{ij,n})$ a.s. under P^* is shown as follows. Note that

$$\gamma^*_{ij,n} = c(r_0,r_1) + C(r_0,r_1)\gamma_{ij,n} \text{ a.s. under } P^* \tag{21}$$

Then

$$\begin{aligned}
P^*(G_{ij,n} = 1 \mid Z_{ij,n}) &= P^*(G_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 P^*(G^*_{ij,n} = 0 \mid Z_{ij,n}, \gamma^*_{ij,n}) + (1 - r_1)P^*(G^*_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1)P^*(G^*_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1)E_{P^*}[P^*(G^*_{ij,n} = 1 \mid Z_{ij,n}, \gamma^*_{ij,n}, \varepsilon_{ij}) \mid Z_{ij,n}, \gamma^*_{ij,n}] \\
&= r_0 + (1 - r_0 - r_1)P^*((Z^*_{ij,n})'b + \varepsilon_{ij} \geq 0 \mid Z_{ij,n}, \gamma^*_{ij,n}) \\
&= r_0 + (1 - r_0 - r_1)\Lambda^*((Z^*_{ij,n})'b) \\
&= r_0 + (1 - r_0 - r_1)\Lambda^*((c(r_0,r_1) + C(r_0,r_1)\gamma_{ij,n})'b_1 + X'_{ij}b_2) \\
&= P(G_{ij,n} = 1 \mid Z_{ij,n}),
\end{aligned}$$

where the first and seventh equalities follow from Eq. (21), the second follows from Eq. (20), the fifth follows from Eq. (19), and the last follows from Eq. (17). □

References

Advani, A. and B. Malde (2018): "Credibly identifying social effects: Accounting for network formation and measurement error," *Journal of Economic Surveys*, 32, 1016–1044.

Ammermueller, A. and J. Pischke (2009): "Peer Effects in European Primary Schools: Evidence from the Progress in International Reading Literacy Study," *Journal of Labor Economics*, 27, 315–348.

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

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

Chandrasekhar, A. (2016): "Econometrics of network formation," *The Oxford Handbook of the Economics of Networks*, 303–357.

Chandrasekhar, A. and R. Lewis (2014): "Econometrics of Sampled Networks," *Working Paper*.

Chen, X., Y. Hu, and A. Lewbel (2008): "A note on the closed-form identification of regression models with a mismeasured binary regressor," *Statistics & Probability Letters*, 78, 1473–1479.

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 B. Honoré, A. Pakes, M. Piazzesi, and L. Samuelson, Cambridge: Cambridge University Press, 268–323.

de Paula, A., I. Rasul, and P. Souza (2018): "Recovering social networks from panel data: identification, simulations and an application," .

Gallant, R. and H. White (1988): *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*, Oxford: Basil Blackwell.

Graham, B. S. (2015): "Methods of identification in social networks," *Annu. Rev. Econ.*, 7, 465–485.

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.

Hausman, J. A., W. K. Newey, H. Ichimura, and J. L. Powell (1991): "Identification and Estimation of Polynomial Errors-in-Variables Models," *Journal of Econometrics*, 50, 273–295.

Holland, P. W. and S. Leinhardt (1973): "The structural implications of measurement error in sociometry," *Journal of Mathematical Sociology*, 3, 85–111.

Hu, Y. (2008): "Identification and Estimation of Nonlinear Models with Misclassification Error using Instrumental Variables: A General Solution," *Journal of Econometrics*, 144, 27–61.

Hu, Y. and Z. Lin (2018): "Misclassification and the hidden silent rivalry," Working paper.

Jackson, M. O., T. Rodriguez-Barraquer, and X. Tan (2012): "Social capital and social quilts: Network patterns of favor exchange," *American Economic Review*, 102, 1857–97.

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

Kline, B. (2015): "Identification of complete information games," *Journal of Econometrics*, 189, 117–131.

Kossinets, G. (2006): "Effects of missing data in social networks," *Social networks*, 28, 247–268.

Leung, M. (2015): "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics*, 188, 182–195.

Lewbel, A. (2007): "Estimation of Average Treatment Effects with Misclassification," *Econometrica*, 75, 537–551.

Lewbel, A., S. Norris, K. Pendakur, and X. Qu (2017): "Keeping up with peers in India: A new social interactions model of perceived needs," Working paper.

Lewbel, A., X. Qu, and X. Tang (2019): "Social Networks with Misclassified or Unobserved Links," Working paper.

Mahajan, A. (2006): "Identification and Estimation of Regression Models with Misclassification," *Econometrica*, 74, 631–665.

Moffitt, R. A. (2001): "Policy interventions, low-level equilibria, and social interactions," *Social Dynamics*, 4, 6–17.

Molinari, F. (2008): "Partial Identification of Probability Distributions with Misclassified Data," *Journal of Econometrics*, 144, 81–117.

Ridder, G. and S. Sheng (2015): "Estimation of Large Network Formation Games," Working paper.

Wang, D. J., X. Shi, D. A. McFarland, and J. Leskovec (2012): "Measurement error in network data: A re-classification," *Social Networks*, 34, 396–409.