Identification and Estimation of Large Network Games with Private Link Information

Hülya Eraslan and Xun Tang*

December 29, 2017

Abstract

We study the identification and estimation of large network games where each individual holds private information about its links *and* payoffs. Extending Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2010), we build a tractable empirical model of network games where the individuals are heterogenous with private link and payoff information, and characterize its unique, symmetric pure-strategy Bayesian Nash equilibrium. We then show that the parameters in individual payoffs are identified under "large market" asymptotics, whereby the number of individuals increases to infinity in a fixed and small number of networks. We also propose a consistent two-step m-estimator for individual payoffs. Our method is distribution-free in that it does not require parametrization of the distribution of shocks in individual payoffs. Monte Carlo simulation show that our estimator has good performance in moderate-sized samples.

1 Introduction

We study the identification and estimation of large Bayesian games on networks where individuals hold private information about their links as well as payoffs. Private information on links is prevalent in many empirical environments where the network involves a sizable population. In such cases, it is implausible to assume that each individual has complete information

*Eraslan: Rice University, Department of Economics, Houston, Texas. E-mail: eraslan@rice.edu. Tang: Rice University, Department of Economics, Houston, Texas. E-mail: xun.tang@rice.edu. We are grateful to Angelo Mele and Matt Shum for helpful comments. We thank seminar and conference participants at Vanderbilt University and "Third Annual Conference on Network Science in Economics" for helpful comments. İbrahim Emirahmetoğlu provided excellent research assistance.

about the *full* network structure. For example, Banerjee, Chandrasekhar, Duflo and Jackson (2014) presented evidence how members of rural communities in India had incomplete information about the network structure based on the diffusion of gossips. The impact of private information on links in network games has also been studied in theory. In a pioneering work, Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2010) established the existence of symmetric, monotone Bayesian Nash equilibria in such games. They also showed that shifts in the degree distribution due to increased connectivity have unambiguous effects on the equilibrium behavior of individuals on the network.

Despite their empirical relevance and theoretical importance, large network games with private information on links have not been investigated in the structural econometrics literature. Menzel (2015) studies the identification and estimation of large Bayesian games where all individuals are strategically related, and their types and actions are exchangeable conditional on observed characteristics, and where payoff-relevant private information are *conditionally independent across players*. In our model, individuals' private types are inherently correlated through their own links, thus inducing dependence between their actions even in equilibrium.

We provide a new econometric framework for the structural analysis of large network games with private information on links and payoffs. We extend Galeotti, Goyal, Jackson, Vega-Redondo and Yariv (2010) to build a tractable empirical model of network games where the individuals are heterogenous with such private information.[1] Under a quadratic payoff specification (often used in social interaction models), we show that the model has a unique, symmetric Bayesian Nash equilibrium. We then investigate the model identification under the concept of "large-market" asymptotics, whereby the number of networks is fixed and small but the sample size increases with the number of individuals. Such an asymptotic concept has been used in the estimation of large Bayesian games in different contexts. See, for example, Leung (2015), Menzel (2015), Lin and Xu (2017), Song (2015) and Xu (2016).[2]

Identification under such an asymptotics concept is known to be non-standard. This is because the sample units cannot be considered as independent draws from any "population" that can be treated as known when the sample size approaches infinity. Our identification strategy tackles this difficulty through two steps. First, we define a set of asymptotic moments that can be consistently estimated via sample analogs under the "large-market" asymptotics. Next, we

[1]The model in Galeotti et al (2010) focuses on homogenous individuals with private information about degrees alone. Each individual's payoff only depends on its degree and neighbors' actions. In comparison, our model accommodates a second source of private information in payoff shocks. This is because we consider a data environments where the links are reported in the data, and thus need to allow for unreported private information to avoid degeneracy in the econometric model.

[2]Menzel (2016) and Song (2015) study large Bayesian games in a general setup where all individuals are strategically interlinked. Leung (2015) estimates large games in network formation when individuals have private payoff shocks, rather than Bayesian games on a given network.

derive the structural link between these asymptotic moments and the model parameters, and use it as the basis for our identification. We then introduce an index sufficiency condition that reduces the dimension of arguments in the individuals' endogenous peer effects, and show that it is sufficient for recovering individual payoff parameters.

We propose a two-step estimator for individual payoffs. The first step estimates the asymptotic moments using kernel regressions. The second step estimates the payoff parameters through a minimum-distance approach. This extremum estimator is consistent under standard regularity conditions. The estimator has several desirable properties: It does not rely on any parametric assumption on payoff shocks and exploits some form of asymptotic uncorrelation between neighborhood profiles to achieve consistency; its implementation also does not involve any tuning parameters expect those used in the first-step kernel estimation. Monte Carlo simulation show that our estimator has good performance in moderate-sized samples.

▶ **Other Related Literature.** Bramoulle, Djebbari and Fortin (2009) and Blume, Brock, Durlauf and Jayaraman (2015) establish identification results in Bayesian games where individuals have private payoff shocks. Both papers maintain that the complete network structure is common knowledge among all individuals. In addition, the identification study in these papers is based on "many-market" asymptotics, which assumes the knowledge of reduced-form coefficients. Such knowledge typically requires the data to report a large number of repeated games on a fixed network structure. Xu (2016) and Lin and Xu (2017) estimate Bayesian games on large networks where individuals have private information about payoffs, but have common knowledge about the complete network structure. They require some form of near-epoch dependence between individual actions in equilibrium in order to estimate the model. In comparison, we accommodate a flexible information structure where individuals have private information about neighborhood characteristics in addition to payoffs. This leads to different forms of Bayesian Nash equilibrium, and hence qualitatively different structural equations that relate the asymptotic moments to model parameters. We also allow for richer contextual effects in payoffs than these two papers. Yang and Lee (2017) analyze social interactions where the conditional expectations about group members' behaviors are heterogeneous. The individuals have asymmetric private shocks but share common knowledge about the network/group structure. Canen, Schwartz and Song (2017) applies a behavioral approach to model games on networks where agents have partial observation of neighbor types.

Our paper does not study the strategic formation of networks. This is a related but different topic that has been studied extensively in the literature. Examples include Badev (2013), Chandrasekhar and Jackson (2016), Christakis, Fowler, Imbens and Kalyanaraman (2010), De Paula, Richard-Shubik and Tamer (2017), Mele (2017), Menzel (2016), Miyauchi (2016) and Sheng (2017). See Chandrasekhar (2016) for an extensive review.

▶ **Roadmap.** In Section 2 we present the model and establish the existence and uniqueness of pure-strategy Bayesian Nash equilibrium. In Section 3 we present our identification strategy as the number of individuals approaches infinity while the number of observed networks is fixed and small. In Section 4, we present a two-step consistent estimator. In Section 5, we provide Monte Carlo simulation results. In Section 6 we conclude. Proofs are collected in Appendix A.

2 The Model

Denote the finite set of individuals on a network by N, and let $n \equiv \#N$ denote its cardinality. The network structure is summarized by the n-by-n matrix $g \equiv (g_{ij})_{ij \in N}$. For any $i, j \in N$, let $g_{ij} \equiv 1$ if i's payoff is affected by j's action, and $g_{ij} \equiv 0$ otherwise. By convention in the literature, let $g_{ii} \equiv 0$. Define the set of neighbors for i by $N_i \equiv \{j \in N : g_{ij} = 1\}$. Each individual has a vector characteristics x_i, which has a discrete finite support X with $\#X \equiv K$. Let $x_{N_i} \equiv (x_j)_{j \in N_i}$ denote the vector of characteristics of i's neighbors.

Let $\tau_i \equiv (x_i, N_i, x_{N_i}, \varepsilon_i)$ summarize the information available to individual i, where $\varepsilon_i \in \mathbb{R}$ is an idiosyncratic shock to i's payoff. Each individual $i \in N$ chooses an action a_i from some interval $A \subset \mathbb{R}$. The payoff for each individual i from choosing a_i is:

$$u_i(a_i, a_{-i}, \tau_i) \equiv \tilde{h}_i(\tau_i)a_i - \frac{1}{2}a_i^2 - \tfrac{\phi}{2}\sum\nolimits_{j \in N_i} \tilde{w}_{ij}(\tau_i)(a_i - a_j)^2, \quad (1)$$

where $\phi > 0$ and $a_{-i} \equiv (a_j)_{j \in N \setminus \{i\}}$; and $\tilde{w}_{ij}(\tau_i)$ are the weights that i assigns to its own deviation from a neighbor j's choice. That is, $\tilde{w}_{ij}(\tau_i) \geq 0$ for all $j \in N_i$, and $\sum_{l \in N_i} \tilde{w}_{il}(\tau_i) = 1$ for any i and τ_i. In the terminology of Manski (1993), the function $\tilde{h}_i$ is the "contextual effect" and the last term is the "peer effect". The second term reflects the costs of the action.

We maintain the following assumption about the information available to each individual.

Assumption 1 *(Information) For each $i \in N$, $(N_i, x_{N_i}, \varepsilon_i)$ is privately known and unobserved by other individuals $j \neq i$.*

Let F denote the common prior for $(\tau_i)_{i \in N}$ that is known to all individuals and let T denote the support of each τ_i.[3] A pure strategy for individual i is a mapping from T to A. A pure-strategy Bayesian Nash equilibrium (p.s.BNE) is a profile of pure strategies $(s_i)_{i \in N}$ such that:

$$s_i(\tau_i) \in \arg\max_{a_i \in A} \mathbf{E}[u_i(a_i, \mathbf{s}_{-i}(\tau_{-i}), \tau_i) \mid \tau_i] \quad \forall i \in N,$$

where $\mathbf{s}_{-i}(\tau_{-i}) = (s_j(\tau_j))_{j \in N \setminus \{i\}}$. The expectation $\mathbf{E}$ integrates out $\tau_{-i} \equiv (\tau_j)_{j \in N \setminus \{i\}}$ with respect to its conditional distribution given τ_i, as implied by the common prior F. Assuming the order

[3]Theorem 1 continues to hold when the support of τ_i differs across $i \in N$.

of integration and differentiation can be swapped, we use the first-order condition for each i to derive the following best response to $\mathbf{s}_{-i} \equiv (s_j)_{j \in N \setminus \{i\}}$:

$$R_i(\tau_i; \mathbf{s}_{-i}) = \gamma \tilde{h}_i(\tau_i) + (1-\gamma) \sum\nolimits_{j \in N_i} \tilde{w}_{ij}(\tau_i) \mathbf{E}\left[s_j(\tau_j) \right| \tau_i] .$$

where $\gamma \equiv \frac{1}{1+\phi}$.

An alternative characterization of p.s.BNE is through the following fixed-point characterization:

$$s_i(\tau_i) = R_i(\tau_i; \mathbf{s}_{-i}) \text{ for all } i \in N \text{ and } \tau_i.$$

An argument similar to Blume et al (2015) implies there exists a unique p.s.BNE in this network game.

Theorem 1 *(Uniqueness of p.s.BNE) Under Assumption 1, there exists a unique p.s.BNE.*

The rest of this section focuses on an anonymous version of the game above, where only the profile of characteristics (rather than actual identities) of neighboring individuals affect an individual's ex post payoffs and interim beliefs. We show that in such a context, the unique p.s.BNE is symmetric (i.e., all individuals share the same pure strategy) and anonymous (i.e., the equilibrium strategy only depends on the profile of characteristics but not the actual identities).

For each $x \in X$, let $N_{i,x} \equiv \{j \in N_i : x_j = x\}$ denote the set of neighbors of i whose characteristic is x. Let $n_{i,x} \equiv \# N_{i,x}$ for each $x \in X$; and let a K-vector $\mathbf{n}_i \equiv (n_{i,x})_{x \in X}$ summarize the distribution of x_j in the neighborhood N_i. By construction, $\sum_{x \in X} n_{i,x} = \# N_i$. We consider the class of games where individuals' payoffs and interim beliefs depend on some sufficient statistics of $\mathbf{n}_i$. Specifically, let $\pi : \mathbb{N}^K \to M$ be a function that summarizes neighbor characteristics and is known to the researcher. For example, $\pi(\mathbf{n}_i) \in M$ may be the empirical distribution or mean of $(x_j)_{j \in N_i}$. It is important to note that $\pi(.)$ is a function of $\mathbf{n}_i$, but not of the specific identities of neighbors in N_i. In what follows we use $m_i \equiv \pi(\mathbf{n}_i)$ as shorthand.

Assumption 2 *(Symmetry and Sufficiency in Payoffs) (i) There exists $h : X \times M \times \mathbb{R} \to \mathbb{R}$ such that $\tilde{h}_i(\tau_i) = h(x_i, m_i, \varepsilon_i)$ for all i and τ_i, and $\mathbf{E}[h(x_i, m_i, \varepsilon_i) | x_i = x, m_i = m]$ exists for all $x \in X$, $m \in M$. (ii) For each $i \in N$, $\tilde{w}_{ij} = \tilde{w}_{ik}$ for any $j, k \in N_i$ such that $x_j = x_k$. (iii) For any $i \in N$ and $j \in N_i$, $\tilde{w}_{ij}(\tau_i)$ is determined by (x_i, m_i) and x_j only.*

Condition (i) in Assumption 2 states that the contextual effect for an individual i is determined by its own characteristics x_i, its private shocks ε_i, and the profile of neighbor characteristics $(x_j)_{j \in N_i}$. The specific identities of individuals in N_i do not matter for contextual effect. Condition (ii) states that the weights are divided equally among neighbors sharing the same characteristics.[4] Condition (iii) states that for each individual i, the distribution of weights across neighbors with different characteristics are solely determined by its own characteristic x_i and a summary of neighbor characteristics x_j in N_i, but not by specific identities of individuals in the neighborhood.

Under the conditions in Assumption 2, the peer effect as the third term in (1) can be equivalently expressed as:

$$\frac{\phi}{2}\sum\nolimits_{x\in X} w(x,x_i,m_i)\frac{1}{n_{i,x}}\sum\nolimits_{j\in N_{i,x}}(a_i-a_j)^2,$$

where $w(x,x_i,m_i)\in[0,1]$ denotes the weight assigned to the group of neighbors with characteristics x by an individual i with characteristics x_i and neighborhood profile m_i. (That is, $\sum_{x\in X} w(x,x_i,m_i)=1$ for all x_i,m_i.)

Let $t_i \equiv (x_i,m_i,\varepsilon_i)$ denote the *anonymized information* for an individual i. We also maintain an anonymity condition on interim beliefs.

Assumption 3 *(Anonymity in Common Prior) The common prior F is exchangeable in the identities of individuals $i\in N$; and $F(t_j|\tau_i, j\in N_i)=F(t_j|x_j,x_i,m_i,j\in N_i)$ for all τ_i and j.*[5]

This condition states that for each $i\in N$, the interim belief of i about a neighbor j's anonymized information t_j only depends on (x_j,x_i,m_i). This conditional belief is the same across all individuals as F does not vary across individuals. Menzel (2015) used a similar exchangeability condition on the individuals' characteristics and private signals to estimate large Bayesian games. In our context, an individual's private information consists of both the payoff shocks and the neighborhood profile m_i, which is correlated across individuals in general.

In a *symmetric p.s.BNE* all individuals adopt the same pure strategy s, which maps from an individual's anonymized information t_i to A and solves the following fixed-point equation:

$$s(t_i)=r\circ s(t_i) \text{ for all } t_i, \tag{2}$$

where r is the best response function implied by the first-order condition, and $r\circ s$ is the composite of r and s. That is,

$$\begin{aligned} r\circ s(t_i) &\equiv \gamma h(t_i)+(1-\gamma)\sum\nolimits_{x\in X} w(x,x_i,m_i)\tfrac{1}{n_{i,x}}\sum\nolimits_{j\in N_{i,x}}\mathbf{E}\left[s(t_j)|x_i,x_j=x,m_i,j\in N_i\right] \\ &= \gamma h(t_i)+(1-\gamma)\sum\nolimits_{x\in X} w(x,x_i,m_i)\mathbf{E}\left[s(t_j)|x_i,x_j=x,m_i,j\in N_i\right]. \end{aligned} \tag{3}$$

[4] An intuitive interpretation of (ii) is that each node i first classifies its neighbors into K groups defined by the realized values of x_j, and then assigns positive weights to each group based on $\pi(\mathbf{n_i})$. These weights are then equally divided among nodes within the same group.

[5] A distribution of a random vector $(Y_1,Y_2,...,Y_n)$ is *exchangable* if its joint distribution is the same as that of $(Y_{\rho(1)},Y_{\rho(2)},...,Y_{\rho(n)})$ for any permutation $\rho(.)$.

Corollary 1 *(Unique symmetric p.s.BNE) Suppose Assumptions 1, 2 and 3 hold. Then the unique p.s.BNE is symmetric, and the equilibrium strategies depend only on the anonymized information.*

The proof of this corollary is similar to that of Theorem 1 and is omitted for brevity. The main idea is to show that r is a contraction mapping, and then apply the Banach Fixed-point Theorem to show that $s = r \circ s$ has a unique solution.

3 Identification

We now discuss the identification of parameters (γ, h, w) as the number of individuals approaches infinity while the number of observed networks is fixed and small. Such a dimension of asymptotics differs qualitatively from standard cross-sectional econometric models.

Suppose a researcher collects data from a *single* network with $\#N = n$ individuals. The data reports the choice a_i, the characteristics x_i and the neighbor information (N_i, x_{N_i}) for each single individual $i \in N$. Throughout this section, we maintain that such a data set with $\#N = n$ individuals is a single, random draw from some data-generating process (DGP) indexed by n. For each n, the DGP is summarized by a joint distribution F_n of the n-tuple $(\tau_i)_{i \in N_n}$, where F_n satisfies the exchangeability and anonymity conditions in Assumption 3.

We establish the identification of the model via two steps. In the first step, we show that the sample average of choices made by n observed individuals on the network converge in probability to certain asymptotic moments as $n \to \infty$. We then argue that the model can not be identified from these moments without any parametric or shape restrictions on the contextual effect, even when weights in the peer effects are known to the researcher. In the second step, we show that under a mild restriction of index sufficiency on the individual weights in peer effects the model parameters can be uniquely recovered from these asymptotic moments.

3.1 Asymptotic moments

Let N_n denote a sequence of sets such that $N_n \subset N_{n'}$ for all $n < n'$. For each n, let $\mathbf{E}_n(.)$ denote the expectation under F_n. First, we define the asymptotic moments to be used in our identification analysis. Throughout this section we use $\sum_i$ as shorthand for $\sum_{i \in N_n}$, which is a sum over the n individuals in N_n; and we use $\sum_{j \neq i}$ as shorthand for $\sum_{i \in N_n} \sum_{j \in N_n \setminus \{i\}}$, which is a sum over $n(n-1)$ ordered pairs from N_n. We maintain the following condition on idiosyncratic shocks in individual payoffs.

Assumption 4 *(Exogeneity) For all n and $i, j \in N_n$, the common prior F_n is such that the payoff shock ε_j is independent from $(\varepsilon_i, x_i, m_i, g_{ij})$ conditional on (x_j, m_j).*

This assumption states that conditioning on an individual j's characteristics x_j and neighborhood profile m_j purges any correlation between its payoff shock ε_j and neighbors' characteristics or links. This exogeneity assumption requires that ε_j be conditionally independent from any idiosyncratic noises that affect link formation. It fails, for example, if the network formation process depends on unobservable variables correlated with ε_i. We consider the case where the set of neighbor profiles M is discretized and finite.

Assumption 5 *(Existence of Limits) For any $x, x' \in X$, $m, m' \in M$,*

$$\begin{aligned} h^*(x,m) &\equiv \lim_{n\to\infty} \tfrac{1}{n} \sum_i \mathbf{E}_n[h(t_i)|x_i = x, m_i = m], \\ p^*(x,m) &\equiv \lim_{n\to\infty} \tfrac{1}{n} \sum_i \mathbf{E}_n\left[1\{x_i = x, m_i = m\}\right], \\ q^*(m'|x',x,m) &\equiv \lim_{n\to\infty} \tfrac{1}{n(n-1)} \sum_{j\neq i} \mathbf{E}_n(1\{m_j = m'\}|x_j = x', x_i = x, m_i = m, g_{ij} = 1) \end{aligned}$$

exist and $p^*(x,m) \neq 0$.

Existence of the limit $h^*(x,m)$ is a mild restriction on the conditional distribution of ε_i. It holds, for example, if $h(t_i)$ is additively separable in ε_i and $\mathbf{E}_n(\varepsilon_i|x_i = x, m_i = m) = 0$ for all n. In Appendix B we provide an example of a random link formation process in which p^*, q^* exist.

The following proposition relates the parameters γ, w, h to asymptotic moments λ^* and q^*.

Proposition 1 *Suppose Assumptions 1, 2, 3, 4 and 5 hold. Then*

$$\lambda^*(x,m) \equiv \lim_{n\to\infty} \tfrac{1}{n} \sum_i \mathbf{E}_n(a_i|x_i = x, m_i = m) \tag{4}$$

exists and

$$\lambda^*(x,m) = \gamma h^*(x,m) + (1-\gamma) \sum_{x'\in X} w(x',x,m) \left[\sum_{m'\in M} \lambda^*(x',m') q^*(m'|x',x,m) \right] \tag{5}$$

for all $x \in X$, $m \in M$.

Equation (5) is an empirical analog of the moments implied by p.s.BNE in (2), with individuals' interim expectation about others' actions replaced by an expression that only consists of estimable asymptotic moments. The double sum on the right-hand side of (5) is an expectation of λ^* with respect to a joint distribution over a neighbor's characteristics x', m', defined as $w(.,x,m) \times q^*(.|.,x,m)$. Such a distribution is an individual's *weighted interim belief in the limit*, because it applies the weights in peer effects to the limit of individual interim beliefs about neighbors' m_j.

Next, we show that the asymptotic moments λ^* and q^* in (4) can be consistently estimated using sample averages across individuals *from a single network* as $n \to \infty$ under certain condition of asymptotic uncorrelation. For any $x \in X$ and $m \in M$, let $\iota_i(x,m)$ be shorthand for $1\{x_i = x, m_i = m\}$. Let $\mathbf{C}_n$ and $\mathbf{V}_n$ denote covariance and variance under F_n respectively.

Assumption 6 *(Asymptotic Uncorrelation) For any $x, x' \in X$ and $m, m' \in M$,*

(i) $\mathbf{C}_n(\iota_i(x,m), \iota_j(x,m)) \to 0$ as $n \to \infty$;

(ii) $\mathbf{C}_n(\iota_i(x,m)\iota_j(x',m')g_{ij}, \iota_k(x,m)\iota_\ell(x',m')g_{k\ell}) \to 0$ as $n \to \infty$ if $\{i,j\} \cap \{k,\ell\} = \emptyset$;

(iii) $\mathbf{V}_n[a_i\iota_i(x,m)]$ and $\mathbf{V}_n[(\iota_i(x,m)\iota_j(x',m')g_{ij}]$ exist for all n, and are both $o(n)$.

This condition requires that as the network size increases, the correlation between the neighborhood profiles m_i and m_j for any two individuals i and j diminishes. In Appendix B we provide an example of how this condition holds in a simple random Poisson network.

The next proposition shows that λ^* and q^* can be estimated consistently as the number of individuals on the single network in data approaches infinity ($n \to \infty$).

Proposition 2 *Suppose Assumptions 1, 2, 3, 4, 5 and 6 hold. Then*

$$\frac{\sum_i a_i \iota_i(x,m)}{\sum_i \iota_i(x,m)} \xrightarrow{p} \lambda^*(x,m)$$

and

$$\frac{\sum_{j\neq i} \iota_j(x',m')\iota_i(x,m)g_{ij}}{\sum_{j\neq i} 1\{x_j = x'\}\iota_i(x,m)g_{ij}} \xrightarrow{p} q^*(m'|x',m,x)$$

as $n \to \infty$ for any $x, x' \in X, m, m' \in M$ on the support of p^*.

A generic vector of parameters (γ', h', w') is *observationally equivalent* to the actual parameter (γ, h, w) in the DGP based on the asymptotic moments λ^* and q^* if (γ', h', w') satisfies (5) almost surely $[p^*]$, where λ^*, q^* are identified as probability limits in Proposition 2. We say (γ, h, w) is *identified* based on these asymptotic moments if there exists no *other* element (γ', h', w') in the parameter space that is observationally equivalent to (γ, h, w).

It is clear from (5) that (γ, h, w) can not be identified using these asymptotic moments λ^* and q^* without further restrictions. To see this, note that for any generic weight function w' (not necessarily equal to the actual weight function w) there always exist γ' and h' such that (5) holds almost surely $[p^*]$, with λ^*, q^* fixed and identified from the data. Similarly, for other $(\gamma', h') \neq (\gamma, h)$, one may construct a weight function w' that satisfies (5) by redistributing weights across realized values of x' conditional on each (x, m).

3.2 Preview of identification strategy

In this subsection we illustrate the main idea for identification using a simple example. For the rest of Section 3, we maintain Assumption 1-6, so that the asymptotic moments are consistently estimable and are considered known for identification purpose.

The non-identification result mentioned above shows that the parameters can not be recovered from asymptotic moments without further restrictions on the contextual effects h *and* the weights in peer effects w. Thus we focus on a semiparametric model with conditional mean restriction on the contextual effects.

Assumption 7 *(Mean Contextual Effects)* $h(t_i) = \eta(x_i, m_i; \theta) + \varepsilon_i$ *where* η *is known up to a finite-dimension parameter* θ *and* $\mathbf{E}_n(\varepsilon_i | x_i, m_i) = 0$ *for all* n *and* $x_i \in X$, $m_i \in M$.

This assumption states that the conditional mean of contextual effects given (x_i, m_i) is known up to a finite-dimensional parameter θ. It is commonly used in econometric models estimated by Non-linear Least Squares (NLS) or Generalized Method of Momments (GMM). Under this condition, $h^*(x, m) = \eta(x, m; \theta)$. We show the model is identified under some exclusion restriction on the weights in the peer effects as well as some rank condition on the support of observables.

▶ **Exclusion restriction.** Suppose $x \equiv (z, v)$ and has a discrete support $X = Z \otimes V$ where $Z \equiv \{z_{(1)}, z_{(2)}\}$ and $V \equiv \{v_{(1)}, .., v_{(\kappa)}\}$. By construction $\#X = 2\kappa \geq 4$. Also suppose that $w(x', x, m)$ is a function of z and z' alone.

For $k, \ell = 1, 2$, let $\omega_{k\ell}$ denote the weights for $z = z_{(k)}$ and $z' = z_{(\ell)}$, and let $\lambda_k(v, m)$ and $\eta_k(v, m; \theta)$ be shorthand for $\lambda^*(z_{(k)}, v, m)$ and $\eta(z_{(k)}, v, m; \theta)$. Under this exclusion restriction, the structural link in Proposition 1 is reduced to

$$\lambda_k^*(v, m) = \gamma \eta_k(v, m; \theta) + (1 - \gamma) \sum\nolimits_{\ell=1,2} \omega_{k\ell} \sum\nolimits_{v', m'} Q_{k\ell}^*(v', m', v, m) \lambda_\ell^*(v', m') \tag{6}$$

for $k = 1, 2$, where the summation $\sum_{v', m'}$ is over the supports V and M, and

$$Q_{k\ell}^*(v', m', v, m) \equiv \frac{q^*(m' \mid z' = z_{(\ell)}, v', z = z_{(k)}, v, m)}{\#V} \text{ for } \ell, k = 1, 2.$$

That is, $Q_{k\ell}^*$ summarizes an individual i's belief about a neighbor j's own neighborhood profile m_j, based on i's information (x_j, x_i, m_i) and adjusted by the weights that i assigns in peer effects. By construction, $\omega_{k\ell} Q_{k\ell}^*(., ., v, m)$ is an individual i's weighted interim belief about v', m' and $z' = z_{(\ell)}$ conditional on $(z_i, v_i, m_i) = (z_{(k)}, v, m)$, as explained after Proposition 1.

▶ **Rank condition.** To fix ideas, suppose (x, m) are discrete and denote $\lambda_k^*(.)$, $\eta_k(., .; \theta)$ and $Q_{k\ell}^*(., .)$ respectively by two column vectors λ_k^*, $\eta_k(\theta)$ and a square matrix $Q_{k\ell}^*$, with each component in λ_k^* and $\eta_k(\theta)$ and each row and column in $Q_{k\ell}^*$ corresponding to an element (v, m)

on the joint support $V \otimes M$. Let $\lambda^* \equiv [\lambda_1^{*\prime}, \lambda_2^{*\prime}]'$, $\eta(\theta) \equiv [\eta_1(\theta)', \eta_2(\theta)']'$. Then we can write (6) as:

$$\lambda^* = \gamma\eta(\theta) + (1-\gamma)Q_\omega^*\lambda^*, \tag{7}$$

where

$$Q_\omega^* \equiv \begin{pmatrix} \omega_{11}Q_{11}^*, & \omega_{12}Q_{12}^* \\ \omega_{21}Q_{21}^*, & \omega_{22}Q_{22}^* \end{pmatrix}.$$

Equation (7) consists of $2 \times \#V \times \#M$ equalities and involves unknown parameters γ, θ, $\omega \equiv (\omega_{k\ell})_{k,\ell=1,2}$. as well as the identified asymptotic moments $\lambda_k^*(.)$ and $q^*(.|.)$ (in $Q_{k\ell}^*$). These equalites are "quasi-structural" in that they depend on the expected choices of actions λ_k^*, which themselves are endogenous objects arising from the equilibrium. However, by Proposition 2, both λ^*, q^* are consistently estimable from sample averages and can be considered known in identification.[6]

We now derive the rank conditions needed for uniquely recovering γ, θ, ω from (7). Suppose there exists some other $(\gamma', \theta', \omega') \neq (\gamma, \theta, \omega)$ that is observationally equivalent to (γ, θ, ω). Then the right-hand side of (7) must remain the same when (γ, θ, ω) is replaced by $(\gamma', \theta', \omega')$ based on the asymptotic moments λ^* and q^*. This implies there exists a column vector $\upsilon_k \in \mathbb{R}^4 \backslash \{0\}$ such that $[\eta_k(\theta'), \eta_k(\theta), Q_{k1}^*\lambda_1^*, Q_{k2}^*\lambda_2^*]\upsilon_k = 0$ for $k = 1, 2$. An intuitive condition that prevent this from happening is:

$$\text{"For any } \theta' \neq \theta, \ [\eta_k(\theta'), \eta_k(\theta), Q_{k1}^*\lambda_1^*, Q_{k2}^*\lambda_2^*] \text{ has full rank for } k = 1, 2\text{"}. \tag{8}$$

Thus this is a sufficient condition for identifying (γ, θ, ω) from the asymptotic moments. In general it holds when the conditional mean contextual effects η is nonlinear in θ and (x, m), and there is enough variation of (x, m) on the support. Also note that because the functional form of η_k is known (up to θ) and $\lambda_k^*, Q_{k\ell}^*$ are identified from Proposition 2, the rank condition in (8) can be tested.

► **Special cases: linearity in parameters.** The rank condition in (8) does not hold when η is linear in θ. That is, $\eta_k(\theta) = \boldsymbol{\zeta}_k\theta_k$ for all $k = 1, 2$, where ζ_k is a $(\#V \times \#M)$-by-dim(θ_k) matrix of known functions of (v, m), and $\theta_k \neq 0$ is a vector of constant coefficients. In such cases, $[\boldsymbol{\zeta}_k\theta_k', \boldsymbol{\zeta}_k\theta_k]$ can not have full rank for any θ_k' that is proportional to θ_k. Nevertheless it is relatively straight-forward to adjust the argument above to derive the following rank conditions for identifying (γ, θ, ω):

$$\text{"}[\zeta_k, Q_{k1}^*\lambda_1^*, Q_{k2}^*\lambda_2^*] \text{ has full rank for } k = 1, 2\text{"}. \tag{9}$$

[6]The econometrics literature abounds in examples where structural models are identified using quasi-structural equations which involve equilibrium outcomes. For instance, see Bajari, Hong, Krainer and Nekipelov (2009) in the context of static discrete games with incomplete information; and Aguirregabiria and Mira (2010) in dynamic games with Markovian perfect equilibria.

To see why, suppose there exists some $(\gamma', \theta', \omega')$ that is observationally equivalent to (γ, θ, ω). Then there exists a column vector $\tau_k \in \mathbb{R}^3 \backslash \{0\}$ such that $[\zeta_k, Q^*_{k1}\lambda^*_1, Q^*_{k2}\lambda^*_2]\tau_k = 0$ for $k = 1, 2$. Thus the rank condition (9) is sufficient for identification.

It follows from (7) that the reduced-form for asymptotic moments in equilibrium is:

$$\lambda^* = \gamma[I - (1-\gamma)Q^*_\omega]^{-1} \begin{pmatrix} \boldsymbol{\zeta}_1\theta_1 \\ \boldsymbol{\zeta}_2\theta_2 \end{pmatrix}.$$

Hence the rank condition in (9) can be expressed in terms of model primitives:

$$\text{“}[\zeta_k,\ \gamma Q^*_{k1}(M_{11}\boldsymbol{\zeta}_1\theta_1 + M_{12}\boldsymbol{\zeta}_2\theta_2),\ \gamma Q^*_{k2}(M_{21}\boldsymbol{\zeta}_1\theta_1 + M_{22}\boldsymbol{\zeta}_2\theta_2)] \text{ has full rank for } k = 1, 2\text{”}, \quad (10)$$

where $M_{k\ell}$ for $k, \ell = 1, 2$ are conformable submatrices that partitition the inverse of $I - (1-\gamma)Q^*_\omega$.

► **A Numerical Example.** We conclude this preview with a numerical example that illustrates the rank conditions. Let $x_i \equiv (v_i, z_i)$, where v_i and z_i are binary with supports $\{1, 2\}$ and $\{z_{(1)}, z_{(2)}\}$ respectively. Define $m_i \equiv 1\left\{\frac{\#\{j \in N_i : z_j = z_{(1)}\}}{\# N_i} \geq \frac{1}{2}\right\}$. That is, the contextual effect depends on the network structure only through the proportion of neighbors with $z_j = z_{(1)}$. Suppose $\eta(x_i, m_i; \theta) \equiv \theta_d v_i m_i$ if $z_i = z_{(d)}$ for $d = 1, 2$, and suppose that the weights in peer effects only depend on z_j, z_i alone. Let $\theta_1 = 0.8$, $\theta_2 = 1.3$, $\gamma = 0.7$, $\omega_{11} = 0.6$, $\omega_{12} = 0.4$, $\omega_{21} = 0.3$, $\omega_{22} = 0.7$. Let

$$Q^*_{k\ell} = \begin{pmatrix} 0.15 & 0.40 & 0.25 & 0.20 \\ 0.15 & 0.40 & 0.25 & 0.20 \\ 0.40 & 0.10 & 0.30 & 0.20 \\ 0.40 & 0.10 & 0.30 & 0.20 \end{pmatrix} \forall k, \ell = 1, 2,$$

where the (i, j)-entry in $Q^*_{k\ell}$ corresponds to $Q^*_{k\ell}(v', m', v, m)$ with (v, m), (v', m') being the i-th and j-th element in $\{(1, 1), (1, 0), (2, 1), (2, 0)\}$. It is worth emphasizing that in our specification of $Q^*_{k\ell}$ above, we intentionally minimize the source of exogenous variation by restricting $q^*(m' \mid z', v', z, v, m)$ to be invariant in (z', z, m). Yet even in this scenario it is straighforward to verify that the rank condition in (10) holds.

3.3 Formal results: index sufficiency

In this subsection we generalize and formalize the identification argument in Section 3.2. Our method requires an index sufficiency condition on the weights in peer effects.

Assumption 8 *(Index Sufficiency) There exist known indexes* $\psi : X \to \Psi$ *and* $\varphi : X \times M \to \Phi$, *where* $\dim(\Psi) < \dim(X)$ *and* $\dim(\Phi) \leq \dim(X \times M)$, *such that* $w(x', x, m) = w(y', y, \tilde{m})$ *whenever* $\psi(x') = \psi(y')$ *and* $\varphi(x, m) = \varphi(y, \tilde{m})$ *for all* $x, x', y, y' \in X$ *and* $m, \tilde{m} \in M$.

Index sufficiency is used frequently in semiparametric econometric models. (See Powell (1994) for further discussion.) In our context, Assumption 8 is decomposed into two substantive restrictions. First, individuals with the same index $\varphi(x_i, m_i)$ assigns weights to neighbors in the same way. Second, neighbors with the same index $\psi(x_j)$ always receive the same weight. Index sufficiency subsumes the exclusion restriction mentioned in Section 3.2 with with the indexes $\psi(x')$ and $\varphi(x, m)$ being subvectors of x' and (x, m) respectively.

Under this condition, we can reparametrize the weight function w as a function defined over the lower-dimensional support of indexes. That is, there exists $\omega : \Psi \times \Phi \to [0, 1]$ with $\sum_{c\in\Psi} \omega(c, d) = 1$ for all $d \in \Phi$ such that:

$$w(x', x, m) = \frac{\omega(\psi(x'), \varphi(x, m))}{\#\{\tilde{x} : \psi(\tilde{x}) = \psi(x')\}} \text{ for all } x', x \in X \text{ and } m \in M.$$

Under Assumptions 7 and 8, the link between asymptotic moments and parameters in (5) is

$$\lambda^*(x, m) = \gamma\eta(x, m; \theta) + (1 - \gamma) \sum\nolimits_{c\in\Psi} \omega(c, \varphi(x, m))\mu^*(c, x, m), \tag{11}$$

where

$$\mu^*(c, x, m) \equiv \frac{\sum_{\{x':\psi(x')=c\}} \sum_{m'\in M} \lambda^*(x', m')q^*(m'|x', x, m)}{\#\{\tilde{x} : \psi(\tilde{x}) = c\}}.$$

Because the index function ψ is known, Proposition 2 implies that μ^* is identified (and consistently estimable) as $n \to \infty$.[7]

By definition, the true model elements (γ, θ, ω) in the data-generating process is identified if the equality in (11) fails at least for some set of (x, m) with positive measure in p^* whenever (γ, θ, ω) is replaced by a different vector of parameters $(\tilde{\gamma}, \tilde{\theta}, \tilde{\omega}) \neq (\gamma, \theta, \omega)$.

To fix ideas, suppose $\#M < \infty$ so that $\#\Psi < \infty$ and $\#\Phi < \infty$. For each (x, m), let $\boldsymbol{\mu}^*(x, m) \equiv (\mu^*(c, x, m))_{c\in\Psi}$ be a row-vector. We say a random row-vector $\mathbf{v}$ *has full rank conditional on some event* $\mathcal{E}$ (under a probability measure p) if there exists no column-vector $\boldsymbol{\alpha} \neq \mathbf{0}$ such that $p\{\mathbf{v}\boldsymbol{\alpha} = 0 \mid \mathcal{E}\} = 1$.

Assumption 9 *(Rank Condition) For any* $\tilde{\theta} \neq \theta$ *and any* $d \in \Phi$*,* $[\eta(x, m; \tilde{\theta}),\ \eta(x, m; \theta),\ \boldsymbol{\mu}^*(x, m)]$ *has full rank conditional on* $\varphi(x, m) = d$ *under* p^**.*

This condition requires there be sufficient variation over an individual's private information (x, m). Generalization to the case with $\#M{=}\infty$ would involve some form of "completeness" condition on linear operators defined by integrals.

[7]To derive (11), use the reparametrized weights ω to write $\lambda^*(x, m)$ as

$$\begin{aligned} & \gamma h(x, m) + (1 - \gamma) \sum\nolimits_{c\in\Psi} \sum\nolimits_{\{x':\psi(x')=c\}} \tfrac{\omega(c,\varphi(x,m))}{\#\{\tilde{x}:\psi(\tilde{x})=c\}} \left[\sum\nolimits_{m'} \lambda^*(x', m')q^*(m'|x', x, m)\right] \\ = \quad & \gamma h(x, m) + (1 - \gamma) \sum\nolimits_{c\in\Psi} \omega(c, \varphi(x, m)) \tfrac{\sum_{\{x':\psi(x')=c\}} \sum_{m'} \lambda^*(x',m')q^*(m'|x',x,m)}{\#\{\tilde{x}:\psi(\tilde{x})=c\}}, \end{aligned}$$

where the equality holds because $\omega(c, \varphi(x, m))$ is constant over $\{x' : \psi(x') = c\}$ and $m' \in M$.

Proposition 3 *Suppose Assumptions 7, 8 and 9 hold. Then* (γ, θ, ω) *are identified from the asymptotic moments.*

Assumption 9 is often satisfied when η is nonlinear in θ as well as (x, m). However, the rank condition in Assumption 9 does not hold in a class of models that are linear in parameters. These are models with $\eta(x, m; \theta) = \boldsymbol{\zeta}_d(x, m)'\theta_d$ whenever $\varphi(x, m) = d$, where the vector of functions $\boldsymbol{\zeta}_d : X \times M \to \mathbb{R}^{L_d}$, $L_d < \infty$ are known up to a finite-dimensional parameter θ_d.

For example, $\boldsymbol{\zeta}_d(x, m) \equiv [1, x, f(x, m)]$ for all $d \in \Phi$, where $x \in \mathbb{R}^{D_X}$ and $f(x, m) : X \times M \to \mathbb{R}^{D_f}$ is a vector of component-wise squared differences between x and the mean of neighbor characteristics. Then $\eta(x, m; \theta) = \theta_{d,0} + x\theta_{d,X} + f(x, m)\theta_{d,f}$ whenever $\varphi(x, m) = d$, with $\theta_d \equiv (\theta_{d,0}, \theta_{d,X}, \theta_{d,f})$ and $L = 1 + D_X + D_f$. For such a class of models, Assumption 9 does not hold because for any $\tilde{\theta}_d \neq \theta_d$ that is proportional to θ_d, $[\boldsymbol{\zeta}_d(x, m)'\tilde{\theta}_d, \boldsymbol{\zeta}_d(x, m)'\theta_d, \boldsymbol{\mu}^*(x, m)]$ can not have full rank conditional on $\varphi(x, m) = d$ regardless of the value of d. Our next proposition shows that for such models (θ, γ, ω) is identified under different and yet intuitive rank conditions.

Proposition 4 *Suppose Assumptions 7 and 8 hold with* $\eta(x, m; \theta) = \boldsymbol{\zeta}_d(x, m)'\theta_d$ *for all* (x, m) *such that* $\varphi(x, m) = d$*, where* $\boldsymbol{\zeta}_d : X \times M \to \mathbb{R}^{L_d}$ *is known,* $\theta_d \neq 0$ *and* $L_d < \infty$ *for each* $d \in \Phi$*. Then* (γ, θ, ω) *are identified from the asymptotic moments if for each* $d \in \Phi$, $[\boldsymbol{\zeta}_d(x, m), \boldsymbol{\mu}^*(x, m)]$ *has full rank conditional on* $\varphi(x, m) = d$ *under* p^**.*

4 Two-Step M-Estimator

We propose a two-step estimator for parameters in individual payoffs. First, estimate the asymptotic moments. Then estimate payoff parameters by matching the implied asymptotic moments with empirical analogs from the data. Throughout this section, we maintain Assumptions 1 to 7 in Section 2 and 3.

To fix ideas, we present the estimator for the example in Section 3.2, where the observed characteristics are discrete (i.e., $X = Z \otimes V$ with $Z \equiv \{z_{(1)}, z_{(2)}\}$ and $V \equiv \{v_{(1)}, .., v_{(\kappa)}\}$ for $\kappa \geq 2$), and the weights in peer effects only depend on the binary characteristics z_i of an individual and its neighbors. Generalization to the cases with continuous covariates in $Z \otimes V$ is complex and left for future research.

To simplify notation, we reparametrize the model as

$$\beta_{k\ell} \equiv (1 - \gamma)\omega_{k\ell} \text{ for } k, \ell \in \{1, 2\}.$$

In what follows, we use $\delta_0 \equiv [\gamma_0; \beta_0; \theta_0]$ to denote the true parameters in the data-generating process and let $\delta \equiv [\gamma; \beta; \theta]$ denote a generic element in the parameter space D.

Partition the set of individuals N_n into... into $N_{n,(k)} \equiv \{i \in N_n : z_i = z_{(k)}\}$ for $k = 1, 2$, and let $\hat{\lambda}, \hat{q}$ be the non-parametric kernel (frequency) estimators for λ^*, q^* in Proposition 2. Let $\eta_i(\theta) \equiv \eta(x_i, m_i; \theta)$, $\hat{\lambda}_i \equiv \hat{\lambda}(z_i, v_i, m_i)$, $\hat{q}_{\ell,i}(m'|v') \equiv \hat{q}(m'|z' = z_{(\ell)}, v', x_i, m_i)$ and

$$\hat{\chi}_{\ell,i} \equiv \tfrac{1}{\#V} \sum\nolimits_{v',m'} \hat{q}_{\ell,i}(m'|v') \hat{\lambda}(z_{(\ell)}, v', m').$$

Our two-step estimator is:

$$\hat{\delta}_n \equiv \arg\min_{\delta \in D} \hat{G}_n(\delta)$$

with

$$\hat{G}_n(\delta) \equiv n^{-1} \sum\nolimits_i \left[\hat{\lambda}_i - \gamma \eta_i(\theta) - \sum\nolimits_k 1\{z_i = z_{(k)}\} \sum\nolimits_\ell \hat{\chi}_{\ell,i} \beta_{k\ell} \right]^2.$$

This estimator is consistent under the following conditions.

T1 (Parameter Space) δ_0 is in the interior of a compact parameter space D.

T2 (Identificaton) For any $\delta \neq \delta_0$ in D, (6) does not hold for a set of (x, m) with positive measure under p^*.

T3 (Regular contextual effects) The mean contextual effect $\eta(.,.;\theta)$ is continuously differentiable in θ with a bounded gradient almost surely under p^*.

Compactness of D in T1 ensures that a maximum of the probability limit of the objective function exists; interiority of the maximum allows for Taylor series expansion around the true parameter in the proof of consistency in Proposition 5 below. Condition T2 maintains that the parameters in the model are point identified. Sufficient conditions for T2 are provided in Section 3. The smoothness and the boundedness conditions in T3 are used for showing the uniform convergence of the objective function in the proof of consistency below.

Proposition 5 *Suppose Assumptions 1 to 7 hold. Then* $\hat{\delta}_n \xrightarrow{p} \delta_0$ *under the conditions T1, T2, T3.*

An alternative approach for two-step estimation would be as follows. First, use (5) and iterative, forward-substitution to express λ^* as an infinite series that depends on the parameters and asymptotic moments that can be estimated from the data. Then use a minimum-distance method to estimate the parameters by matching the estimated infinite series with the estimate $\hat{\lambda}$. Compared with our current approach, this alternative estimator involves higher computation costs, and more complex conditions are needed in order to show identification and consistency using its objective function.

5 Simulation

In this section we present simulation evidence for the performance of our two-step m-estimator, using simulated samples under various designs of the DGP. The contextual effect in (3) is parametrized as

$$\gamma h(t_i) = x_i\beta_x + m_i\beta_m + \epsilon_i,$$

where $\beta_x = 3.0$, $\beta_m = 1.5$, the error ϵ_i follows a zero-mean truncated normal distribution, and the individual characteristic x_i is uniformly distributed over a discrete support. The peer effects in (3) is parametrized as $\tau \equiv (1-\gamma) = 0.8$, with weights allocated equally among different types of neighbors. We experiment with two distinctive definitions of the sufficient statistic for neighbor characteristics: (i) m_i is a discretized value of $\frac{1}{\#N_i}\sum_{j\in N_i}|x_j - x_i|$; and (ii) the number of same-type neighbors censored at 10.

For every design and sample size n considered, we simulate $S = 200$ independent samples, each of which consists of observable characteristics x_i and choices a_i by n individuals on a *single* network. In each sample, the individual characteristics are drawn independently from a specified support X. The links between individuals are undirected, and are formed independently with some probability that depends on individual characteristics. We study two scenarios for each pair x_i, x_j: (a) the link formation probability $p_n(x_i, x_j)$ decreases as the sample size n increases, and $np_n(x_i, x_j)$ converge to a constant; and (b) the link formation probability is fixed $p(x_i, x_j)$ and invariant as $n \to \infty$.

The individual choices under the symmetric pure-strategy Bayesian Nash equilibrium are simulated using the following steps. First, use our specification of the DGP to calculate individuals' interim belief about m_j conditional on x_j, m_i, x_i and $g_{ij} = 1$. Next, plug in this belief into the fixed-point characterization of equilibrium and solve for the endogenous moment $\mathbf{E}_n(a_i|x_i, m_i)$. (See the proof of Proposition 1 in Appendix A for details.) Then, draw individual noises ϵ_i from the distribution specified and set $a_i = \mathbf{E}_n(a_i|x_i, m_i) + \epsilon_i$.

For each sample, we calculate our two-step m-estimator $\hat{\beta}_x, \hat{\beta}_m, \hat{\tau}$. No smoothing parameter is required because of the discrete support for x_i, m_i. The tables below report the empirical bias, variance and mean-squard errors (MSE) from $S = 200$ estimates.

Table 1. Discretized m_i; $X = \{0, 1\}$; $np_n \to (10, 5, 10)$

Case 1: $\epsilon \sim$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0164	0.0135	0.0138	0.1930	0.2098	0.2470	-0.0907	0.0314	0.0396
400	0.0106	0.0055	0.0056	0.1651	0.0953	0.1225	-0.0880	0.0189	0.0264
800	-0.0007	0.0036	0.0036	0.0369	0.0210	0.0224	-0.0432	0.0096	0.0115

Case 2: $\epsilon \sim N(0, 1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0193	0.0251	0.0255	0.1886	0.2435	0.2791	-0.0910	0.0357	0.0439
400	0.0148	0.0111	0.0113	0.1046	0.1169	0.1279	-0.0738	0.0223	0.0277
800	-0.0082	0.0061	0.0061	0.0154	0.0434	0.0436	-0.0290	0.0171	0.0179

Note: The bias, variance and mean squared errors in this table are calculated using $S = 200$ independent samples of single networks with n individuals. Neighborhood profile m_i is defined as $(\#N_i)^{-1} \sum_{j \in N_i} |x_j - x_i|$ rounded to the nearest multiple of $\frac{1}{5}$.

Table 1 reports the results for a design where x_i is Bernoulli with equal probability and m_i is the discretization of the average neighbor characteristics $(\#N_i)^{-1} \sum_{j \in N_i} |x_j - x_i|$, defined by rounding this average to the nearest multiple of $\frac{1}{5}$. As the sample size increases, the independent link-formation probability diminishes but converges to nonzero constants. For simplicity in implementation, we set $nP_n\{g_{ij} = 1 | x_i = x_j = k\} = 10$ for $k \in \{0, 1\}$ and $nP_n\{g_{ij} = 1 | x_i \neq x_j\} = 5$ for all n. The two panels in Table 1 show how the estimator's performance vary with the support and variance of individual noises.

The MSE for three parameters decrease as $n \to \infty$, as the consistency in Proposition 5 implies. The MSEs are quite small even for a moderate sample size $n = 800$. The estimation error in the contextual effect of m_i appears to be greater than that of the peer effects τ and individual effect of x_i. The distribution of individual payoff noises ϵ_i affects the estimation accuracy, as the MSEs are greater for models with error terms that have higher variances.

Table 2: Censored m_i; $X = \{0, 1\}$; $np_n \to (10, 5, 10)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0460	0.0687	0.0708	0.0190	0.0047	0.0051	-0.1498	0.0127	0.0351
400	-0.0451	0.0267	0.0287	0.0111	0.0019	0.0020	-0.1130	0.0065	0.0193
800	0.0146	0.0152	0.0154	0.0089	0.0009	0.0010	-0.0890	0.0036	0.0116
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0505	0.0762	0.0787	0.0214	0.0060	0.0064	-0.1534	0.0162	0.0397
400	-0.0375	0.0346	0.0360	0.0147	0.0017	0.0019	-0.1172	0.0075	0.0212
800	0.0047	0.0176	0.0176	0.0109	0.0009	0.0011	-0.0898	0.0041	0.0122

Note: The bias, variance and mean squared errors in this table are calculated using $S = 200$ independent samples of single networks with n individuals. For each i, the neighbor profile m_i is defined as the minimum of same-type neighbors and 10.

Table 2 reports the results for a similar model where the neighbor profile is defined as the censored number of same-type neighbors. As before, the MSEs converge to zero as $n \to 0$. In contrast, the estimator for the contextual effect of m_i is more accurate than the case where m_i is the discretization of average neighbor characteristics. This may be explained in part by a richer variation in m_i under this new specification with $np_n \xrightarrow{p} (10, 5, 20)$ relative to that in Table 1. While we do not provide a formal result about the rate of convergence of our two-step estimator, the rate of convergence in the MSE in both Table 1 and Table 2 appears to be reasonably close to $\sqrt{n}$ with few exceptions in the finite sample.

Table 3: Censored m_i; $X = \{0, 1\}$; $np_n \to (20, 10, 20)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0349	0.0261	0.0273	0.0094	0.0020	0.0020	-0.0708	0.0033	0.0083
400	0.0255	0.0158	0.0165	0.0054	0.0011	0.0011	-0.0662	0.0021	0.0065
800	-0.0193	0.0093	0.0097	-0.0037	0.0004	0.0004	-0.0459	0.0010	0.0031
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0446	0.0367	0.0387	0.0116	0.0037	0.0038	-0.0758	0.0062	0.0119
400	0.0313	0.0214	0.0224	0.0076	0.0015	0.0016	-0.0708	0.0037	0.0087
800	-0.0248	0.0113	0.0119	-0.0037	0.0007	0.0007	-0.0450	0.0015	0.0035

Table 3 reports results under a design almost identical to that in Table 2, except that the sequence of link-formation probability now converges to a higher level $(20, 10, 20)$. In comparison with Table 2, the MSE in this case are slightly smaller. This pattern is related to the fact that a higher link formation probability tends to increase the variation in neighbor profile definded the censored number of same-type neighbors. Similar to Tables 1 and 2, the results in Table 3 demonstrates that an increase in the variance of noises leads to slightly worse performance of the estimators.

Table 4: Censored m_i; $X = \{0, 1, 2\}$; $np_n \to (10, 8, 5, 12, 8, 10)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0368	0.0228	0.0241	0.0587	0.0086	0.0121	-0.1501	0.0121	0.0346
400	0.0286	0.0120	0.0129	0.0362	0.0038	0.0051	-0.1381	0.0058	0.0249
800	0.0222	0.0044	0.0049	0.0244	0.0012	0.0018	-0.1146	0.0034	0.0165
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0338	0.0259	0.0270	0.0611	0.0093	0.0130	-0.1517	0.0127	0.0357
400	0.0223	0.0138	0.0143	0.0375	0.0040	0.0054	-0.1394	0.0065	0.0259
800	0.0185	0.0049	0.0053	0.0237	0.0015	0.0021	-0.1149	0.0031	0.0163

Next, to see how an increase in the variation of individual characteristics could impact the estimator performance, we consider a design where x_i is uniformly distributed over $\{0, 1, 2\}$.

As in Table 1-3, we let the link formation probability to diminish as $n \to \infty$. For simplicity in implementation, we set link formation probability as follows: $nP_n\{g_{ij} = 1|x_i = x_j = 0\} = 10$, $nP_n\{g_{ij} = 1|x_i + x_j = 1\} = 8$, $nP_n\{g_{ij} = 1 \mid |x_i - x|_j = 2\} = 5$, $nP_n\{g_{ij} = 1|x_i = x_j = 1\} = 12$, $nP_n\{g_{ij} = 1|x_i + x_j = 3\} = 8$ and $nP_n\{g_{ij} = 1|x_i = x_j = 2\} = 10$ for all n.

Table 4 reports the results for such a design with m_i defined as the censored number of same-type neighbors. Compared with Table 3, the MSEs for β_m and τ, or the contextual and endogenous effects of m_i, are both higher while that for β_x, the marginal effect of individual characteristics, is slightly smaller. We interpret such a pattern as the result of an interaction of two immediate consequences of a larger support of X: on the one hand, a richer support for individual characteristics provides more sources of variation for recovering the parameter; on the other hand this increases the curse of dimensionality in that for a given sample size n there are fewer observations (individuals) that can be used to estimate the asymptotic moments in the first step (both of which condition on individual characteristics).

Table 5: Discretized m_i; $X = \{0, 1\}$; fixed $p = (0.6, 0.4, 0.6)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0286	0.0572	0.0580	0.2388	1.0166	1.0736	0.1810	0.0510	0.0838
400	0.0178	0.0094	0.0097	-0.1048	0.6867	0.6977	0.0980	0.0263	0.0359
800	0.0080	0.0045	0.0046	-0.0083	0.2713	0.2782	0.0260	0.0177	0.0184
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0475	0.0940	0.0963	-0.0622	2.2340	2.2378	0.2771	0.0826	0.1593
400	-0.0300	0.0140	0.0149	-0.0117	1.1734	1.1736	0.1503	0.0467	0.0692
800	-0.0155	0.0090	0.0092	0.0053	0.8007	0.8007	0.0201	0.0303	0.0307

Up to now we have only considered the designs where the link formation probabality varies with the sample size n. The last two tables of this section, Table 5 and 6, report the simulation results in designs where the links are conditionally independently formed with fixed probability that is invariant with n. In both designs, the neighbor profiles are defined as the discretized average neighbor characteristic (as in Table 1). The two designs differ in the support of x_i.

Both Table 5 and 6 demonstrate evidence of convergence of MSE for all three estimators under a fixed probability design. While the estimation error for $\hat{\beta}_x$ and $\hat{\tau}$ is comparable to their counterparts under convergent link formation probability (in Table 1), the MSEs for $\hat{\beta}_m$ is greater. We conjecture that this distinction happens because an increasing sample size n has different implication for the distribution of neighbor profile m_i under two paradigms of fixed or convergent link formation probabilities.

Table 6: Discretized m_i; $X = \{0, 1, 2\}$; fixed $p = (0.6, 0.4, 0.3, 0.6, 0.4, 0.6)$

Case 1: $\epsilon \sim N(0,1)$ truncated at $[-3/2, 3/2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	0.0124	0.0250	0.0251	-0.0646	0.8372	0.8414	0.3398	0.0312	0.1467
400	0.0108	0.0026	0.0027	-0.0754	0.4497	0.4554	0.1202	0.0077	0.0221
800	-0.0046	0.0012	0.0012	0.0312	0.3069	0.3079	-0.0633	0.0053	0.0093
Case 2: $\epsilon \sim N(0,1.5)$ truncated at $[-2, 2]$									
	$\hat{\beta}_x$			$\hat{\beta}_m$			$\hat{\tau}$		
n	Bias	Var	MSE	Bias	Var	MSE	Bias	Var	MSE
200	-0.0206	0.0265	0.0269	-0.1845	1.2083	1.2423	0.3431	0.0335	0.1512
400	-0.0113	0.0057	0.0058	-0.0968	0.8013	0.8106	0.1403	0.0128	0.0325
800	0.0045	0.0025	0.0025	0.0191	0.6374	0.6377	-0.1225	0.0092	0.0242

6 Concluding Remarks

Directions for future research include: the use of higher moments of individual actions in identification and estimation; implementation of estimation when state variables are continuous; and richer models where network formation is endogenized along with individual actions.

The model and the method we propose in this article can be used to analyze the individual incentives in a variety of environments. Examples include individuals' choices of education or consumption in the context of large social networks. Yet another interesting direction for future research would be the empirical analyses of individual preferences and interaction under such scenarios.

REFERENCES

Badev, A. "Discrete Games with Endogenous Networks: Theory and Policy," Working paper, University of Pennsylvania, 2013.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. Jackson. "Gossip: Identifying Central Individuals in a Social Network," Working paper, NBER, 2014.

Blume, L., W. Brock, S. Durlauf, and R. Jayaraman. "Linear Social Interactions Models," Journal of Political Economy, 2015, 123 (2), 444-496.

Bramoulle, Y., H. Djebbari, and B. Fortin. "Identification of Peer Effects through Social Networks," Journal of Econometrics, 2009, 150 (1), 44-55.

Canen, N., Schwartz, J. and K. Song. "Estimating Local Interactions Among Many Agents Who Observe Their Neighbors," UBC working paper, 2017.

Chandrasekhar, A. "Econometrics of Network Formation," Oxford Handbook on the Economics of Networks, 2016.

Chandrasekhar, A. and M. Jackson. "A Network Formation Model based on Subgraphs," Stanford working paper, 2016.

Christakis, N., J. Fowler, G. Imbens and K. Kalyanaraman. "An Empirical Model for Strategic Network Formation," Working paper, NBER, 2010.

De Paula, A., S. Richards-Shubik and E. Tamer: "Identification of Preferences in Network Formation Games," Econometrica, forthcoming, 2017.

Galeotti, A., S. Goyal, M. Jackson, F. Vega-Redondo, and L. Yariv. "Network Games," Review of Economic Studies, 2010, 77 (1), 218-244.

Leung, M., "Two-Step Estimation of Network-Formation Models with Incomplete Information," Journal of Econometrics, 2015, 188 (1), 182-195.

Lin, Z. and H. Xu, "Identification and Estimation of Hierarchy Effects in Social Interactions," Econometrics Journal, forthcoming, 2017.

Manski, C., "Identification of Endogenous Social Effects: The Reflection Problem," Review of Economic Studies, 1993, 60 (3), 531-542.

Mele, A., "A Structural Model of Segregation in Social Networks," Econometrica, forthcoming, 2017.

Menzel, K., "Inference for Games with Many Players," Review of Economic Studies, 2015, 88 (1), 306-337.

Menzel, K., "Strategic Network Formation with Many Agents", working paper, NYU, 2016.

Miyauchi, Y. "Structural Estimation of a Pairwise Stable Network with Nonnegative Externality," Journal of Econometrics, 2016, 195 (2), 224-235.

Newey, W. and D. McFadden, "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. New York: Elsevier, 1994.

Powell, J. L. "Estimation of Semiparametric Models," *Handbook of Econometrics*, Vol. IV, ed. by R. F. Engle and D. L. McFadden. New York: Elsevier, 1994.

Sheng, S., "A Structural Econometric Analysis of Network Formation Games," Working paper, UCLA, 2017.

Song, K., "Econometric Inference on a Large Bayesian Game," Working paper, Vancouver School of Economics, 2015.

Xu, H., "Social Interactions on Large Networks: A Game Theoretic Approach," International Economic Review, forthcoming, 2016.

Xu, X. and L. Lee, "Estimation of a Binary Choice Game Model with Network Links," Working paper, Ohio State University, 2015.

Yang, C. and L. Lee, "Social Interactions under Incomplete Information with Heterogeneous Expectations," Journal of Econometrics, forthcoming, 2017.

Appendix

A. Proofs in Section 3

Proof of Theorem 1. Let $\mathcal{S}$ denote the set of bounded functions on T^n taking values in A^n. For $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$, let $\mathbf{s} \leq \mathbf{s}'$ denote $\mathbf{s}(\tau) \leq \mathbf{s}'(\tau)$ for all $\tau \in T^n$. For any $\mathbf{s} \in \mathcal{S}$, let $||\mathbf{s}|| = \max_i \sup\{|s_i(\tau_i)| : \tau_i \in T\}$, i.e., $||.||$ is the supremum norm. Note that $\mathcal{S}$ with the supremum norm is a complete metric space.

Define a mapping $\mathbf{R} : \mathcal{S} \to \mathcal{S}$ as

$$\mathbf{R}(\mathbf{s})_i(\tau_i) \equiv R_i(\tau_i; \mathbf{s}_{-i}) \text{ for all } i \in N.$$

First note that $\mathbf{R}(\mathbf{s}) \in \mathcal{S}$ for any $\mathbf{s} \in \mathcal{S}$, and so $\mathbf{R}$ maps $\mathcal{S}$ to itself. To establish existence of a unique p.s.BNE, we show that $\mathbf{R}$ satisfies the contraction property. That is, there exists $c \in (0,1)$ such that $||\mathbf{R}(\mathbf{s}) - \mathbf{R}(\mathbf{s}')|| \leq c||\mathbf{s} - \mathbf{s}'||$ for any $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$.

To do so, fix any $\mathbf{s}, \mathbf{s}' \in \mathcal{S}$. Note that for any $i \in N$ and $\tau_i \in T$, we have

$$\begin{aligned} |R_i(\tau_i; \mathbf{s}_{-i}) - R_i(\tau_i; \mathbf{s}'_{-i})| &= (1-\gamma) \sum\nolimits_{j \in N_i} \tilde{w}_{ij}(\tau_i) \left| \mathbf{E}\left[s_j(\tau_j) - s'_j(\tau_j) \middle| \tau_i \right] \right| \\ &\leq (1-\gamma) \sum\nolimits_{j \in N_i} \tilde{w}_{ij}(\tau_i) \left| \mathbf{E}\left[||\mathbf{s} - \mathbf{s}'|| \,|\, \tau_i \right] \right| \\ &\leq (1-\gamma) ||\mathbf{s} - \mathbf{s}'||. \end{aligned} \tag{12}$$

By definition, $||\mathbf{R}(\mathbf{s}) - \mathbf{R}(\mathbf{s}')|| = \sup\{|R_i(\tau_i; \mathbf{s}_{-i}) - R_i(\tau_i; \mathbf{s}'_{-i})| : \tau_i \in T\}$ for some $i \in N$. Hence the desired result follows from (12). $\square$

Proof of Proposition 1. Under Assumption 1, 2 and **??**, a unique symmetric pure-strategy Bayesian Nash equilibrium exists in each data-generating process indexed by n, and

$$a_i = s(t_i) = \gamma h(t_i) + (1-\gamma) \sum\nolimits_{x' \in X} w(x', x_i, m_i) \mathbf{E}_n[s(t_j) | x_j = x', x_i, m_i, g_{ij} = 1].$$

Define $\lambda_n(x, m) \equiv \frac{1}{n} \sum_i \mathbf{E}_n(a_i | x_i = x, m_i = m) = \mathbf{E}_n(a_i | x_i = x, m_i = m)$, which does not vary with the specific identity of an individual i because of the symmetry in Assumption 2 and exchangeability in Assumption 3. Likewise, define $\bar{h}_n(x, m) \equiv \frac{1}{n} \sum_i \mathbf{E}_n\left[h(t_i) | x_i = x, m_i = m\right] = \mathbf{E}_n\left[h(t_i) | x_i = x, m_i = m\right]$. Thus by construction,

$$\lambda_n(x, m) = \gamma \bar{h}_n(x, m) + (1-\gamma) \sum_{x' \in X} w(x', x, m) \mathbf{E}_n[s(t_j) | x_j = x', m_i = m, x_i = x, g_{ij} = 1]. \tag{13}$$

For any n, the Law of Total Expectation and Assumption 4 imply

$$\begin{aligned} &\mathbf{E}_n[s(t_j) | x_j = x', x_i = x, m_i = m, g_{ij} = 1] \\ = &\sum_{m' \in M} \mathbf{E}_n(a_j | m_j = m', x_j = x') \mathbf{E}_n(1\{m_j = m'\} | x_j = x', x_i = x, m_i = m, g_{ij} = 1) \\ = &\sum_{m' \in M} \lambda_n(x', m') q_n(m' | x', m, x), \end{aligned} \tag{14}$$

where $q_n(m'|x',x,m) \equiv \frac{1}{n(n-1)}\sum_{j\neq i}\mathbf{E}_n(1\{m_j = m'\}|x_j = x', x_i = x, m_i = m, g_{ij} = 1)$. The second equality holds because $\mathbf{E}_n(1\{m_j = m'\}|x_j = x', x_i = x, m_i = m, g_{ij} = 1)$ does not depend on specific identities of individuals i and j under Assumptions 2 and 3. Combining (13) and (14), we write λ_n as the solution to a fixed-point equation that depends on $(\bar{h}_n, q_n)$. That is,

$$\lambda_n = \Gamma(\lambda_n; \bar{h}_n, q_n), \tag{15}$$

where $\Gamma(.; \bar{h}_n, q_n)$ is a self-map over the set of bounded and continuous functions with domain $X \otimes M$, and

$$\Gamma(\lambda_n; \bar{h}_n, q_n)(x,m) \equiv \gamma\bar{h}_n(x,m) + (1-\gamma)\sum\nolimits_{x'\in X} w(x',x,m)\sum\nolimits_{m'\in M}\lambda_n(x',m')q_n(m'|x',x,m).$$

The solution is unique because, for any $\bar{h}_n$ and q_n, the map $\Gamma(.; \bar{h}_n, q_n)$ has a contraction property under the sup norm.[8]

Next, let q and $\tilde{q}$ denote generic density (probability mass) functions of m' given x', x, m. We now show that for any $\bar{c} > 0$ there exists $c_1, c_2 > 0$ so that

$$\|\tilde{h} - h\| \leq c_1 \text{ and } \|\tilde{q} - q\| \leq c_2 \quad \text{implies } \|\tilde{\lambda} - \lambda\| \leq \bar{c}, \tag{16}$$

where $\|.\|$ denotes the sup norm over the respective domains, and λ and $\tilde{\lambda}$ are the unique solutions in $\lambda = \Gamma(\lambda; h, q)$ and $\tilde{\lambda} = \Gamma(\tilde{\lambda}; \tilde{h}, \tilde{q})$ respectively. To verify (16), recursively substitute λ and $\tilde{\lambda}$ into $\Gamma(\lambda; h, q)$ and $\Gamma(\tilde{\lambda}; \tilde{h}, \tilde{q})$ and use the fact that "$\gamma \in (0,1)$, $w(., x, m) \geq 0$ and $\sum_{x'\in X} w(x', x, m) = 1$ for all $x \in X, m \in M$".

It then follows from (16) that the solution to the fixed point problem $\lambda = \Gamma(\lambda; h, q)$ is continuous in (h, q). Under Assumption 5, $\bar{h}_n \to h^*$ and $q_n \to q^*$ given the sup norm. Consequently, the sequence of solutions $\lambda_n = \Gamma(\lambda_n; \bar{h}_n, q_n)$ converges to the unique λ^* that solves $\lambda^* = \Gamma(\lambda^*; h^*, q^*)$. $\square$

Proof of Proposition 2. Fix some $n \in \mathbb{N}_{++}$, $x \in X$ and $m \in M$. The Chebychev's Inequality implies that for any constant $c > 0$,

$$\mathbf{P}_n\left\{\left|\frac{1}{n}\sum\nolimits_i \iota_i - \mathbf{E}_n\left(\frac{1}{n}\sum\nolimits_i \iota_i\right)\right| \geq c\right\} \leq c^{-2}\mathbf{V}_n\left(\frac{1}{n}\sum\nolimits_i \iota_i\right), \tag{17}$$

where $\mathbf{P}_n$ is the probability measure associated with F_n, and $\mathbf{V}_n$ denotes the variance under F_n. In what follows, let $\sigma^2_{n,i}$ be a shorthand for $\mathbf{V}_n(\iota_i)$; and let $\mathbf{C}_{n,i,j}$ be a shorthand for $\mathbf{C}_n(\iota_i, \iota_j)$, which is the covariance between ι_i and ι_j under F_n. By the exchangeability and anonymity of F_n in Assumption 3, $\sigma_{n,i}$ does not depend on i and $\mathbf{C}_{n,i,j}$ does not depend on i and j. Therefore the right-hand side of (17) can be written as

$$c^{-2}n^{-2}\left[\sum\nolimits_i \sigma^2_{n,i} + \sum\nolimits_{j\neq i}\mathbf{C}_{n,i,j}\right] = \frac{n\sigma^2_{n,i}}{n^2c^2} + \frac{n(n-1)\mathbf{C}_{n,i,j}}{n^2c^2}. \tag{18}$$

[8] The proof of the contraction property of $\Gamma(.; \bar{h}_n, q_n)$ is similar to that of Theorem 1, and omitted for brevity.

By the first asymptotic uncorrelation condition in Assumption 6, the two terms on the right-hand of (18) converge to 0 as $n \to \infty$. Thus $\frac{1}{n}\sum_i \iota_i - \frac{1}{n}\sum_i \mathbf{E}_n(\iota_i) \xrightarrow{p} 0$ as $n \longrightarrow \infty$. Under Assumption 5, this implies $\frac{1}{n}\sum_i \iota_i(x,m) \xrightarrow{p} p^*(x,m)$ as $n \longrightarrow \infty$.

Next, recall that a_i is a function of $(x_i, m_i, \varepsilon_i)$ in p.s.BNE. By the law of total covariance, $\mathbf{C}_n(a_i\iota_i(x,m), a_j\iota_j(x,m)) \to 0$ under conditional independence in Assumption 4 and the first asymptotic uncorrelation condition in Assumption 6. It follows from a similar argument using Chebychev's Inequality that $\frac{1}{n}\sum_i a_i\iota_i - \frac{1}{n}\sum_i \mathbf{E}_n(a_i\iota_i) \xrightarrow{p} 0$. Under Assumption 5 and Proposition 1, the limit of $\frac{1}{n}\sum_i \mathbf{E}_n(a_i\iota_i)$ as $n \to \infty$ exists. By exchangeability of F_n in Assumption 3 both $\mathbf{E}_n(\iota_i)$ and $\mathbf{E}_n(a_i\iota_i)$ do not vary across the identities of individuals i. It then follows from the Slutsky's Theorem that

$$\frac{\frac{1}{n}\sum_i a_i\iota_i(x,m)}{\frac{1}{n}\sum_i \iota_i(x,m)} \xrightarrow{p} \frac{\lim_{\tilde{n}\to\infty}\mathbf{E}_{\tilde{n}}[a_i\iota_i(x,m)]}{\lim_{\tilde{n}\to\infty}\mathbf{E}_{\tilde{n}}[\iota_i(x,m)]} = \lim_{\tilde{n}\to\infty}\frac{\mathbf{E}_{\tilde{n}}[a_i\iota_i(x,m)]}{\mathbf{E}_{\tilde{n}}[\iota_i(x,m)]} = \lambda^*(x,m) \tag{19}$$

for all x, m on the support of p^*.

We now prove the second claim in the proposition. Fix $x, x' \in X$ and $m, m' \in M$. In what follows, let ι_j', ι_i be shorthand for $\iota_j(x',m'), \iota_i(x,m)$ respectively. Define $\xi_{ij} \equiv (\iota_j'\iota_i g_{ij} + \iota_i'\iota_j g_{ji})/2$ so that $\xi_{ij} = \xi_{ji}$ and $\frac{1}{n(n-1)}\sum_{j\neq i}\iota_j'\iota_i g_{i,j} = \frac{2}{n(n-1)}\sum_i\sum_{j>i}\xi_{ij}$ by construction. By the Chebychev Inequality, for any constant $c > 0$,

$$\mathbf{P}_n\left\{\left|\tfrac{2}{n(n-1)}\sum\nolimits_{j\neq i}\xi_{ij} - \mathbf{E}_n\left(\tfrac{2}{n(n-1)}\sum\nolimits_{j\neq i}\xi_{ij}\right)\right| \geq c\right\} \leq c^{-2}\mathbf{V}_n\left(\tfrac{2}{n(n-1)}\sum\nolimits_{j\neq i}\xi_{ij}\right),$$

where the right-hand side is

$$\tfrac{4}{c^2n^2(n-1)^2}\sum\nolimits_{j>i}\sum\nolimits_{t>l}\mathbf{C}_n(\xi_{ij}, \xi_{lt}).$$

By construction, this quadruple sum consists of $\binom{n}{2} \times \binom{n}{2} = \frac{1}{4}(n^4 - 2n^3 + n^2)$ terms. These include $\binom{n}{2} = \frac{1}{2}(n^2 - n)$ variance terms $\mathbf{V}_n(\xi_{ij})$, $\binom{n}{2} \times \binom{n-2}{2} = \frac{1}{4}(n^4 - 6n^3 + 11n^2 - 6n)$ covariance terms $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ in which the unordered pairs $\{i,j\}$ and $\{l,t\}$ do not overlap, and $\binom{n}{2} \times \left(\binom{n}{2} - \binom{n-2}{2} - 1\right) = n^3 - 3n^2 + 2n$ covariance terms $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ in which the two pairs $\{i,j\}$ and $\{l,t\}$ share *exactly one* individual in common. Note that $\mathbf{V}_n(\xi_{ij})$ and $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ are bounded for all $\{i,j\}$ and $\{l,t\}$. Furthermore, the covariance term $\mathbf{C}_n(\xi_{ij}, \xi_{lt})$ with $\{i,j\} \cap \{l,t\} = \emptyset$ does not vary with the identities $\{i,j,l,t\}$ due to the anonymity of common prior in Assumption 3. Under the asymptotic uncorrelation condition in Assumption 6, $\mathbf{C}_n(\xi_{ij}, \xi_{lt}) \to 0$ as $n \to \infty$ if $\{i,j\} \cap \{l,t\} = \emptyset$. Therefore $\frac{4}{c^2n^2(n-1)^2}\sum_{j>i}\sum_{l>t}\mathbf{C}_n(\xi_{ij}, \xi_{lt}) \to 0$ as $n \to 0$. Hence $\frac{1}{n(n-1)}\sum_{j\neq i}[\iota_j'\iota_i g_{ij} - \mathbf{E}_n(\iota_j'\iota_i g_{ij})] \xrightarrow{p} 0$. By a similar argument, $\frac{1}{n(n-1)}\sum_{j\neq i}[1\{x_j = x'\}\iota_i g_{ij} - \mathbf{E}_n(1\{x_j = x'\}\iota_i g_{ij})] \xrightarrow{p} 0$. Under our condition in Assumption 5, $\lim_{\tilde{n}\to\infty}\mathbf{E}_{\tilde{n}}\left(\iota_j'\iota_i g_{ij}\right)$ exists and $\lim_{\tilde{n}\to\infty}\mathbf{E}_{\tilde{n}}(1\{x_j = x'\}\iota_i g_{ij})$ is non-zero. The second convergence result in the proposition follows from an argument analogous to (19) under the exchangeability in Assumption 3 and the existence of the limits q^* in Assumption 5. □

Proof of Proposition 3. Suppose (θ, γ, ω) is observationally equivalent to some different vector of parameters $(\tilde{\theta}, \tilde{\gamma}, \tilde{\omega})$ based on asymptotic moments in (11). This means (11) holds almost surely p^* when (γ, θ, ω) is replaced by $(\tilde{\gamma}, \tilde{\theta}, \tilde{\omega})$. For each $d \in \Phi$, let $\boldsymbol{\omega}(d) \equiv (\omega(c, d))_{c \in \Psi}$, which is a column-vector of weights assigned over Ψ conditional on $\varphi(x, m) = d$. Likewise, define $\tilde{\boldsymbol{\omega}}$ using $\tilde{\omega}(., .)$. That is, for every $d \in \Phi$,

$$\gamma \eta(x, m; \theta) + (1 - \gamma) \boldsymbol{\mu}^*(x, m) \boldsymbol{\omega}(d) = \tilde{\gamma} \eta(x, m; \tilde{\theta}) + (1 - \tilde{\gamma}) \boldsymbol{\mu}^*(x, m) \tilde{\boldsymbol{\omega}}(d) \tag{20}$$

whenever $\varphi(x, m) = d$.

Consider the following cases. Case (i): $\theta = \tilde{\theta}$ and $(\gamma, \omega) \neq (\tilde{\gamma}, \tilde{\omega})$. Then (20) implies that for each $d \in \Phi$, $[\eta(x, m; \theta), \boldsymbol{\mu}^*(x, m)] \boldsymbol{\alpha}(d) = 0$ whenever $\varphi(x, m) = d$, where $\boldsymbol{\alpha}(d) \equiv [\gamma - \tilde{\gamma}, (1 - \gamma) \boldsymbol{\omega}(d)' - (1 - \tilde{\gamma}) \tilde{\boldsymbol{\omega}}(d)']'$. Because $(\gamma, \omega) \neq (\tilde{\gamma}, \tilde{\omega})$, the vector $\boldsymbol{\alpha}(d)$ must be nonzero at least for some $d \in \Phi$. This implies that *at least for some* $d \in \Phi$, $[\eta(x, m; \theta), \boldsymbol{\mu}^*(x, m)]$ does not have full rank conditional on $\varphi(x, m) = d$. Case (ii): $\theta \neq \tilde{\theta}$ and $(\gamma, \omega) = (\tilde{\gamma}, \tilde{\omega})$. Then (20) implies $\eta(x, m; \theta) = \eta(x, m; \tilde{\theta})$ almost everywhere p^*. Case (iii): $\theta \neq \tilde{\theta}$ and $(\gamma, \omega) \neq (\tilde{\gamma}, \tilde{\omega})$. Then (20) implies that for every $d \in \Phi$,

$$[\eta(x, m; \theta), \eta(x, m; \tilde{\theta}), \boldsymbol{\mu}^*(x, m)] \mathbf{b}(d) = 0$$

whenever $\varphi(x, m) = d$, where $\mathbf{b}(d) \equiv [\gamma, -\tilde{\gamma}, (1 - \gamma) \boldsymbol{\omega}(d)' - (1 - \tilde{\gamma}) \tilde{\boldsymbol{\omega}}(d)']'$. By construction, $\mathbf{b}(d)$ is non-zero for all d. Thus (20) implies that *for each* $d \in \Phi$, $[\eta(x, m; \theta),\ \eta(x, m; \tilde{\theta}),\ \boldsymbol{\mu}^*(x, m)]$ does not have full rank conditional on $\varphi(x, m) = d$.

Each of these cases of observational equivalence implies the following condition: "There exists $\tilde{\theta}$ such that *at least for some* $d \in \Phi$, $[\eta(x, m; \theta),\ \eta(x, m; \tilde{\theta}),\ \boldsymbol{\mu}^*(x, m)]$ does not have full rank conditonal on $\varphi(x, m) = d$ under p^*". It then follows that under Assumption 9, (θ, γ, ω) is not observationally equivalent to any $(\tilde{\theta}, \tilde{\gamma}, \tilde{\omega}) \neq (\theta, \gamma, \omega)$. $\square$

Proof of Proposition 4. Suppose (θ, γ, ω) is observationally equivalent to some other $(\tilde{\theta}, \tilde{\gamma}, \tilde{\omega})$. This implies that for each $d \in \Phi$, $[\boldsymbol{\zeta}_d(x, m), \boldsymbol{\mu}^*(x, m)] \boldsymbol{\alpha}(d) = 0$ when $\varphi(x, m) = d$, where $\boldsymbol{\alpha}(d) \equiv [\gamma \theta_d - \tilde{\gamma} \tilde{\theta}_d, (1 - \gamma) \boldsymbol{\omega}(d)' - (1 - \tilde{\gamma}) \tilde{\boldsymbol{\omega}}(d)']'$. Consider two cases. Case (i): $\omega = \tilde{\omega}$. In this case, either $\gamma \neq \tilde{\gamma}$ or $\theta_d \neq \tilde{\theta}_d$ at least for some $d \in \Phi$. Otherwise (θ, γ, ω) would be identical to $(\tilde{\theta}, \tilde{\gamma}, \tilde{\omega})$. Hence at least for some d, the two terms $\gamma \theta_d - \tilde{\gamma} \tilde{\theta}_d$ and $(1 - \gamma) \boldsymbol{\omega}(d)' - (1 - \tilde{\gamma}) \tilde{\boldsymbol{\omega}}(d)'$ can not be zero simultaneously. Thus $\boldsymbol{\alpha}(d)$ is a non-zero vector for all d. This implies that *at least for some* $d \in \Phi$, $[\boldsymbol{\zeta}_d(x, m), \boldsymbol{\mu}^*(x, m)]$ does not have full rank conditional on $\varphi(x, m) = d$. Case (ii): $\omega \neq \tilde{\omega}$. In this case, $\omega(d) \neq \tilde{\omega}(d)$ at least for some d. It then follows that $(1 - \gamma) \boldsymbol{\omega}(d) - (1 - \tilde{\gamma}) \tilde{\boldsymbol{\omega}}(d)$ is a non-zero vector at least for some d, regardless of whether $\gamma \neq \tilde{\gamma}$. This implies that *at least for some* $d \in \Phi$, $[\boldsymbol{\zeta}_d(x, m), \boldsymbol{\mu}^*(x, m)]$ does not have full rank conditional on $\varphi(x, m) = d$. Therefore, if $[\zeta_d(x, m), \boldsymbol{\mu}^*(x, m)]$ has full rank conditional on $\varphi(x, m) = d$ for all $d \in \Phi$, then (θ, γ, ω) is not observationally equivalent to any other $(\tilde{\theta}, \tilde{\gamma}, \tilde{\omega})$. $\square$

Proof of Proposition 5. Let $\sum_{x,m}$, $\sum_k$, $\sum_\ell$ be shorthand for $\sum_{x\in X,m\in M}$, $\sum_{k=1,2}$, $\sum_{\ell=1,2}$ respectively. For each $\delta \in D$, let

$$G_0(\delta) \equiv \sum_{x,m} p^*(x,m)\left[\lambda^*(x,m) - \gamma\eta(x,m;\theta) - \sum_k 1\{z = z_{(k)}\}\left(\sum_\ell \chi^*_\ell(x,m)\beta_{k\ell}\right)\right]^2,$$

where

$$\chi^*_\ell(x,m) \equiv \tfrac{1}{\#V}\sum_{v',m'} \lambda^*(v', z_{(\ell)}, m')q^*(m'|v', z' = z_{(\ell)}, x, m).$$

By Proposition 2 and an application of the Weak Law of Large Numbers, $\hat{G}_n(\delta) \xrightarrow{p} G_0(\delta)$ for each $\delta \in D$. For notational convenience, let $1_{k,i} \equiv 1\{z_i = z_{(k)}\}$,

$$\hat{\mathbf{H}}_i \equiv [1_{1,i}\hat{\chi}_{1,i}, 1_{1,i}\hat{\chi}_{2,i}, 1_{2,i}\hat{\chi}_{1,i}, 1_{2,i}\hat{\chi}_{2,i}]$$

and

$$\bar{q}_i(\theta) \equiv \hat{\lambda}_i - \gamma\eta_i(\theta) - \sum_k \mathbf{1}_{k,i}\sum_\ell \hat{\chi}_{\ell,i}\beta_{k\ell}.$$

For any $\delta \neq \delta'$, the mean value approximation implies $\hat{G}_n(\delta') - \hat{G}_n(\delta) = \nabla_\delta\hat{G}_n(\tilde{\delta})(\delta' - \delta)$, where $\tilde{\delta}$ is an intermediate value between δ, δ' and the gradient $\nabla_\delta\hat{G}_n(\delta)$ is

$$\frac{2}{n}\sum_i \bar{q}_i(\delta)[\eta_i(\theta), \hat{\mathbf{H}}_i, \gamma\nabla_\theta\eta_i(\theta)].$$

Let $\|.\|$ denote the Euclidean norm. By the Cauchy-Schwartz inequality,

$$|\nabla_\delta\hat{G}_n(\tilde{\delta})(\delta' - \delta)| \leq \left\|\nabla_\delta\hat{G}_n(\tilde{\delta})\right\| \times \|\delta' - \delta\|,$$

where $\left\|\nabla_\delta\hat{G}_n(\tilde{\delta})\right\|$ is $O_p(1)$ under T3. Then by Lemma 2.9 in Newey and McFadden (1994),

$$\sup_{\delta\in D}|\hat{G}_n(\delta) - G_0(\delta)| \xrightarrow{p} 0.$$

In addition, G_0 is continuous in δ over D under T3. Under the condition in T2, δ_0 is a unique maximizer of $G_0(.)$ over D. By Theorem 2.1 in Newey and McFadden (1994), $\hat{\delta}_n \xrightarrow{p} \delta_0$. □

B. An Examples of Random Network

This section presents an example of networks with *undirected* links (i.e., $g_{ij} = g_{ji}$ for any $i, j \in N$) that satisfy the conditions in Assumption 5 (existence of limits as $n \to \infty$) and Assumption 6 (asymptotic uncorrelation as $n \to \infty$).

As in Section 2, let N denote the set of individuals and let $n \equiv \#N \in \mathbb{N}_{++} \equiv \{1, 2, 3, .., \infty\}$. Each individual is characterized by a binary characteristic $x_i \in X \equiv \{1, 2\}$. (Generalization to the case $\#X \geq 3$ is straightforward.) Individual characteristics $x_i, i = 1, ..., n$ are independently drawn from a fixed multinomial distribution with $p_{(k)} \equiv \Pr\{x_i = k\}$ with $k = 1, 2$. Let

$n_{(k)} \equiv \#\{i \in N : x_i = k\}$ for $k = 1, 2$. By an application of the Weak Law of Large Numbers, $n_{(k)} \to \infty$ and $n_{(k)}/n \to p_{(k)} \in (0, 1)$ as $n \to \infty$ for $k = 1, 2$. As in the text, let $\mathbf{E}_n(.)$ denote the expectation under F_n, or the distribution of $(\tau_i)_{i \in N_n}$ in the data-generating process (DGP) indexed by network size n, and let P_n denote the probability measure in the DGP. Consider a random Poisson network that satisfy the following conditions on link formation.

(R1)**.** *For each n and $k, \ell = 1, 2$, $P_n\{g_{ij} = 1 | x_i = k, x_j = l\} = q_{(k\ell),n}$, where $q_{(k\ell),n} n_{(\ell)} \to \rho_{(k\ell)} < \infty$ as $n \to \infty$.*

(R2). *For each n, the joint distribution of all links conditional on all individual characteristics is*

$$\prod\nolimits_{j>i} \mathbf{E}_n(g_{ij} | x_i, x_j)^{g_{ij}} [1 - \mathbf{E}_n(g_{ij} | x_i, x_j)]^{1-g_{ij}}.$$

Under R2, the links are independent once conditional on the characteristics of individuals. Recall from Section 2 that an individual i's neighborhood profile is summarized by a vector of integers $\mathbf{n}_i \equiv (n_{i,1}, n_{i,2})$ with $n_{i,1} + n_{i,2} = \#N_i$, where $n_{i,k} \equiv \#\{j : g_{ij} = 1, \ x_j = k\}$ and $N_i \equiv \{j \in N : g_{ij} = 1\}$. Let $m_i \equiv (m_{i1}, m_{i2}) \equiv (\min\{n_{i,1}, \bar{n}_1\}, \min\{n_{i,2}, \bar{n}_2\})$. That is, for any fixed n, each component in $m_i \in \mathbb{N}_+^2$ follows a binomial distribution censored at a maximum number of friends $\bar{n}_k$ for $k = 1, 2$. By construction $M \equiv \{0, 1, ..., \bar{n}_1\} \otimes \{0, 1, ..., \bar{n}_2\}$ is finite and invariant as $n \to \infty$. For any n and any $\bar{m} \equiv (\bar{m}_1, \bar{m}_2) \in M$,

$$\begin{aligned} \mathbf{E}_n\left[1\{x_i = k, m_i = \bar{m}\}\right] &= p_{(k)} P_n\{m_{i1} = \bar{m}_1, m_{i2} = \bar{m}_2 | x_i = k\} \\ &= p_{(k)} P_n\{m_{i1} = \bar{m}_1 | x_i = k\} P_n\{m_{i2} = \bar{m}_2 | x_i = k\}, \end{aligned} \tag{21}$$

where the second equality follows from Condition R2. By definition, for any $\bar{m}_k < \bar{n}_k$,

$$\begin{aligned} P_n\{m_{ik} &= \bar{m}_k | x_i = k\} = P_n\{n_{ik} = \bar{m}_k | x_i = k\} \\ &= \binom{n_{(k)} - 1}{\bar{m}_k} \left[q_{(kk),n}\right]^{\bar{m}_k} \left[1 - q_{(kk),n}\right]^{n_{(k)} - \bar{m}_k - 1}. \end{aligned}$$

Under Condition R1, the Poisson Limit Theorem applies and the expression on the right-hand side converges to

$$\left[\rho_{(kk)}\right]^{\bar{m}_k} \exp\{-\rho_{(kk)}\}/(\bar{m}_k!),$$

which is the probability mass function (p.m.f.) of a random variable distributed as Poisson with mean $\rho_{(kk)}$. Furthermore,

$$P_n\{m_{ik} = \bar{n}_k | x_i = k\} = P_n\{n_{ik} \geq \bar{n}_k | x_i = k\}.$$

Under Condition R1, $q_{(k\ell),n} = O(n^{-1})$. This implies $n q^2_{(k\ell),n} \to 0$. Thus by the Le Cam's Theorem and an application of triangular inequality,

$$\left| P_n\{n_{ik} \geq \bar{n}_k | x_i = k\} - \sum\nolimits_{\bar{m}_k = \bar{n}_k}^{\infty} \left[\rho_{(kk)}\right]^{\bar{m}_k} \exp\{-\rho_{(kk)}\}/(\bar{m}_k!) \right| \to 0$$

as $n \to \infty$. That is, as the network size increases to infinity, $P_n\{n_{ik} \geq \bar{n}_k | x_i = k\}$ converges to the survival function (evaluated at $\bar{n}_k$) of a Poisson variable with mean $\rho_{(kk)}$. By a symmetric argument, we can show similar results with $l \neq k$: $P_n\{m_{il} = \bar{m}_\ell | x_i = k\}$ converges to a Poisson p.m.f. with mean $\rho_{(k\ell)}$ for $\bar{m}_\ell < \bar{n}_\ell$; and converges to the survival function at $\bar{n}_\ell$ of a Poisson variable with mean $\rho_{(k\ell)}$ for $\bar{m}_\ell = \bar{n}_\ell$. (To show this, replace $n_{(k)}-1$, $q_{(kk),n}$, $\rho_{(kk)}$ in the argument above with $n_{(\ell)}$, $q_{(k\ell),n}$, $\rho_{(k\ell)}$ respectively, and apply the Poisson Limit Theorem and the Le Cam Theorem.) Thus the right-hand side of (21), and consequently $\mathbf{E}_n\left[1\{x_i = k, m_i = \bar{m}\}\right]$, converges to some non-zero limits $p^*(x, m)$ for all $\bar{m} \in M$ and $k = 1, 2$.

Consider an uncensored vectors $\tilde{m}$. Then

$$\begin{aligned}
& \mathbf{E}_n(1\{m_j = \tilde{m}\} | x_j = 2, x_i = 1, m_i = \bar{m}, g_{ij} = 1) \\
= & \binom{n_{(1)} - 1}{\tilde{m}_1 - 1} \left[q_{(21),n}\right]^{\tilde{m}_1 - 1} \left[1 - q_{(21),n}\right]^{n_{(1)} - \tilde{m}_1} \binom{n_{(2)} - 1}{\tilde{m}_2} \left[q_{(22),n}\right]^{\tilde{m}_2} \left[1 - q_{(22),n}\right]^{n_{(2)} - \tilde{m}_2 - 1} \\
\to & \rho_{(21)}^{\tilde{m}_1 - 1} \frac{\exp\{-\rho_{(21)}\}}{(\tilde{m}_1 - 1)!} \rho_{(22)}^{\tilde{m}_2} \frac{\exp\{-\rho_{(22)}\}}{\tilde{m}_2!},
\end{aligned}$$

where the second equality follows from conditional independence in link formation under Condition R2, and the convergence is due to Poisson approximation of a binomial distribution. Similar derivation for the other case with $\tilde{m} = (\bar{n}_1, \bar{n}_2)$ implies similar results, only with probability mass functions replaced by survival functions in the limit. Thus $q^*(m'|x', x, m)$ exists for all $x, x' \in X$ and $m, m' \in M$. Note that in this example, $q^*(.|x', x, m)$ depends on x', x but not m, which is consitent with the rank condition for identification presented in Section 3.2.

To show asymptotic uncorrelation conditions in Assumption 6, first note that for any $k, \ell \in \{1, 2\}$ and $\bar{m}, \tilde{m} \in M$,

$$\begin{aligned}
& P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l\} \\
= & P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 1\} q_{(k\ell),n} \\
+ & P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\}(1 - q_{(k\ell),n})
\end{aligned}$$

Because $q_{(k\ell),n} \to 0$ as $n \to \infty$, the difference between the right-hand above and the sequence $P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\}$ diminishes to zero as $n \to \infty$. For example, for $k \neq l$ and any uncensored values of $\bar{m}, \tilde{m}$ in M,

$$\begin{aligned}
& P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\} \\
= & \binom{n_{(k)} - 1}{\bar{m}_k} \left[q_{(kk),n}\right]^{\bar{m}_k} \left[1 - q_{(kk),n}\right]^{n_{(k)} - \bar{m}_k - 1} \binom{n_{(\ell)} - 1}{\bar{m}_\ell} \left[q_{(k\ell),n}\right]^{\bar{m}_\ell} \left[1 - q_{(k\ell),n}\right]^{n_{(\ell)} - \bar{m}_\ell - 1} \\
& \binom{n_{(k)} - 1}{\tilde{m}_k} \left[q_{(lk),n}\right]^{\tilde{m}_k} \left[1 - q_{(lk),n}\right]^{n_{(k)} - \tilde{m}_k - 1} \binom{n_{(\ell)} - 1}{\tilde{m}_\ell} \left[q_{(\ell\ell),n}\right]^{\tilde{m}_\ell} \left[1 - q_{(\ell\ell),n}\right]^{n_{(\ell)} - \tilde{m}_\ell - 1}.
\end{aligned}$$

Again, by an application of the Poisson approximation of a Binomial distribution and the Le Cam's Theorem, we have:

$$P_n\{m_i = \bar{m}, m_j = \tilde{m} | x_i = k, x_j = l, g_{ij} = 0\} - P_n\{m_i = \bar{m} | x_i = k\} P_n\{m_j = \tilde{m} | x_j = l\} \to 0$$

for all $\bar{m}, \tilde{m} \in M$. Thus by the law of total covariance,

$$\mathbf{C}_n(\iota_i(x,m), \iota_j(x,m)) \to 0 \ \forall x \in X, m \in M.$$

Finally, that

$$\mathbf{C}_n(\iota_j(x',m')\iota_i(x,m)g_{ij}, \iota_\ell(x',m')\iota_t(x,m)g_{lt}) \to 0 \ \forall x, x' \in X, m, m' \in M$$

as $n \to \infty$ for $\{i,j\} \cap \{t,l\} = \emptyset$ follows from similar derivation.

C. Asymptotic Distribution of Two-Step Estimator

In this part of the appendix, we sketch a heuristic discussion about the asymptotic distribution of our two-step m-estimator. This discussion requires some admittedly high-level assumptions (D1-D6 below). Let $\hat{\rho}$ denotes the vector of all first-stage estimates in $\hat{\lambda}_i$ and $\hat{\chi}_{\ell,i}$. That is,

$$\hat{\rho} = \begin{pmatrix} n^{-1} \sum_i \iota_i(x,m) \\ n^{-1} \sum_i a_i \iota_i(x,m) \\ \left(\frac{1}{n(n-1)} \sum_{j \neq i} \iota_j(x',m') \iota_i(x,m) g_{ij} \right)_{x',m'} \\ \left(\frac{1}{n(n-1)} \sum_{j \neq i} 1\{x_j = x'\} \iota_i(x,m) g_{ij} \right)_{x'} \end{pmatrix}_{x,m}.$$

Let $\nabla_\delta \hat{G}_n(\delta) = n^{-1} \sum_i \Gamma_i(\delta; \hat{\rho})$, with

$$\Gamma_i(\delta; \hat{\rho}) = 2\bar{q}_i(\delta; \hat{\rho})[\eta_i(\theta), \hat{\mathbf{c}}_i, \gamma \nabla_\theta \eta_i(\theta)]$$

where

$$\bar{q}_i(\delta; \hat{\rho}) = \hat{\lambda}_i - \gamma \eta_i(\theta) - \sum\nolimits_k \mathbf{1}_{k,i} \sum\nolimits_\ell \hat{\chi}_{\ell,i} \beta_{k\ell}$$

and $\hat{\mathbf{c}}_i$ is such that $2\bar{q}_i(\delta; \hat{\rho})\hat{\mathbf{c}}_i$ is the deriative of the objective function with respect to the vector of coefficients β. By the first-order condition and a mean-value expansion,

$$\sqrt{n} \nabla_\delta \hat{G}_n(\delta) + \nabla^2_{\delta,\delta} \hat{G}_n(\tilde{\delta}) \sqrt{n}(\hat{\delta} - \delta) = o_p(1),$$

where $\tilde{\delta}$ is some intermediate value between $\hat{\delta}$ and δ. Assume:

D1. $\sup_{x \in X, m \in M} \|\hat{\rho} - \rho\| = o_p(n^{-1/4})$.

D2. $n^{-1} \sum_i \nabla_\delta \Gamma_i(\delta; \rho) - \Psi_n = o_p(1)$, where $\Psi_n \equiv \mathbf{E}_n[\frac{1}{n} \sum_i \nabla_\delta \Gamma_i(\delta; \rho)]$ has full-rank for each n.

D3. The Hessian $\nabla^2_{\rho,\rho} \Gamma_i(\delta; \rho)$ exists and is bounded over an open neighborhood around ρ almost surely under p^*.

D4. $n^{-1}\sum_i \nabla_\rho \Gamma_i(\delta;\rho) - \Phi_n = O_p(n^{-1/4})$, where $\Phi_n \equiv \mathbf{E}_n[\frac{1}{n}\sum_i \nabla_\rho \Gamma_i(\delta;\rho)]$ has full-rank for each n.

D5. $\hat{\rho} - \rho = n^{-1}\sum_i \psi_{n,i} + o_p(n^{-1/2})$ for some $\psi_{n,i}$ determined by (δ, ρ) such that $\mathbf{E}_n(\psi_{n,i}) = 0$ for all n.

D6. $\Lambda_n^{-1}\left\{\frac{1}{\sqrt{n}}\sum_i [\Gamma_i(\delta;\rho) + \Phi_n \times \psi_{n,i}]\right\} \xrightarrow{d} N(0, I)$, where I is the identity matrix and Λ_n a squence of positive semidefinite matrices.

Lemma C1. Under D1 and D2, $\nabla^2_{\delta,\delta}\hat{G}_n(\tilde{\delta}) - \Psi_n \xrightarrow{p} 0$ whenever $\tilde{\delta} \xrightarrow{p} \delta$.

Proof of Lemma C1. Let $\Gamma_i(\delta, \rho)$ be defined in a way that is similar to $\Gamma_i(\delta;\hat{\rho})$, only with the first-stage estimates $\hat{\rho}$ replaced by the probability limit ρ. That is,

$$\begin{aligned}\Gamma_i(\delta;\rho) &= 2\left(\lambda^*(x_i, m_i) - \gamma\eta(x_i, m_i;\theta) - \sum\nolimits_k \mathbf{1}_{k,i}\sum\nolimits_\ell \chi^*_\ell(x_i, m_i)\beta_{k\ell}\right)\\ &\quad\times[\eta(x_i, m_i;\theta), \mathbf{c}_i, \gamma\nabla_\theta\eta(x_i, m_i;\theta)].\end{aligned}$$

By definition,

$$\nabla^2_{\delta,\delta}\hat{G}_n(\tilde{\delta}) - \Psi_n = \underbrace{n^{-1}\sum\nolimits_i \left[\nabla_\delta\Gamma_i(\tilde{\delta};\hat{\rho}) - \nabla_\delta\Gamma_i(\delta;\rho)\right]}_{A} + \underbrace{n^{-1}\sum\nolimits_i \nabla_\delta\Gamma_i(\delta;\rho) - \Psi_n}_{B}.$$

The absolute value of the first term A on the right-hand side is bounded above by

$$\sup\nolimits_{x,m}|\nabla_\delta\Gamma(x, m;\tilde{\delta};\hat{\rho}) - \nabla_\delta\Gamma(x, m;\delta;\rho)|.$$

Note that the consistency of $\hat{\delta}$ implies $\tilde{\delta} \xrightarrow{p} \delta$. Hence it is bounded above by a term that is $o_p(1)$ because $\sup_{x\in X, m\in M}\|\hat{\rho} - \rho\| \xrightarrow{p} 0$. The second term B is $o_p(1)$ under condition D2. □

Lemma C2. Under D1, D3, D4 and D5,

$$\sqrt{n}\nabla_\delta\hat{G}_n(\delta) = n^{-1}\sum\nolimits_i [\Gamma_i(\rho) + \Phi_n \times \psi_{n,i}] + o_p(n^{-1/2}).$$

Proof of Lemma C2. Let $\Gamma_i(\hat{\rho})$ be shorthand for $\Gamma(x_i, m_i;\delta;\hat{\rho})$. By a second-order Taylor expansion, we can write

$$\begin{aligned}\nabla_\delta\hat{G}_n(\delta) &= n^{-1}\sum\nolimits_i \Gamma_i(\hat{\rho}) \qquad (22)\\ &= n^{-1}\sum\nolimits_i \left[\Gamma_i(\rho) + \nabla_\rho\Gamma_i(\rho)(\hat{\rho} - \rho) + \tfrac{1}{2}(\hat{\rho} - \rho)'\nabla^2_{\rho,\rho}\Gamma_i(\tilde{\rho})(\hat{\rho} - \rho)\right]\end{aligned}$$

for some $\tilde{\rho}$ between $\hat{\rho}$ and ρ. Under D3, the absolute value of $n^{-1}\sum_i(\hat{\rho} - \rho)'\nabla^2_{\rho,\rho}\Gamma_i(\tilde{\rho})(\hat{\rho} - \rho)$ on the right-hand side is bounded above by the product of a constant and $\sup_{x\in X, m\in M}\|\hat{\rho} - \rho\|^2$, where $\|.\|$ denotes the sup-norm. Under D1, this upper bound is $o_p(n^{-1/2})$. Under D4, $[n^{-1}\sum_i \nabla_\rho\Gamma_i(\rho) - \Phi_n] \times (\hat{\rho} - \rho) = o_p(n^{-1/2})$. The claim in the lemma follows from D5. □

It then follows that $(\Psi_n^{-1}\Lambda_n\Psi_n^{-1})^{-1/2}\sqrt{n}(\hat{\delta} - \delta) \xrightarrow{d} N(0, I)$ under D1-D5.