Estimation of Large Network Formation Games

Geert Ridder[*]   Shuyang Sheng[†]

December 2015

(Preliminary and Incomplete)

Abstract

This paper provides estimation methods for network formation models using observed data of a single large network. We characterize network formation as a simultaneous-move game with incomplete information and allow for the effects of indirect friends such as friends in common, so the utility from direct friends can be nonseparable. Nonseparability poses a challenge in the estimation because each individual faces an interdependent multinomial discrete choice problem where the choice set increases with the number of individuals n. We propose a novel method to linearize the utility and derive the closed form of the conditional choice probability (CCP). With the closed form CCP, we show that the finite-player game converges to some limiting game as n goes to infinity. We propose a two-step estimation procedure using the equilibrium condition from the limiting game. The estimation procedure makes little assumption on equilibrium selection, is computationally simple, and provides consistent and asymptotic normal estimators for the structural parameters. Monte Carlo simulations show that the limiting game approximates finite-player games well and can provide accurate estimates.

[*]Department of Economics, University of Southern California, Los Angeles, CA 90089. E-mail: ridder@usc.edu.

[†]Department of Economics, University of California at Los Angeles, Los Angeles, CA 90095. E-mail: ssheng@econ.ucla.edu.

1 Introduction

This paper contributes to the as yet small literature on the estimation of game-theoretic models of network formation.[1] The purpose of the empirical analysis is to recover the preferences of the members of the network, in particular the preferences that determine whether a member of the network will form a link (friendship, business relation or some other type of link) with a specific other member of the network. The preference for a link depends in general on the characteristics of the two members, and on their position in the network, i.e., their number of friends, their number of common friends. It is the dependence of the link preference on the position in the network that complicates the analysis. The preference also depends on unobservable features of the members and the link and assumptions on the nature of these unobservables play a key role in the empirical analysis.

In principle link formation models are discrete choice models where multiple alternatives (links) can be chosen. If the agent has a myopic strategy s/he chooses to form a link if the utility of the link is larger than the utility of not forming the link. There are two reasons why such a strategy is suboptimal. First, the agent is better off considering the choice of links with the other members in the network simultaneously. Second all members in the network are making link choices and all these choices have to be consistent. In this paper the consistency is achieved by assuming that the observed network is a Bayesian Nash equilibrium.

In general there is no unique Bayesian Nash equilibrium. This implies that full-information methods, either have to specify an equilibrium selection mechanism or have to consider partial instead of point identification of the preference parameters. In this paper we propose a limited-information method that is valid even if we do not know the equilibrium selection mechanism.

Assumptions regarding the unobservables in the link preferences play an important role in the empirical analysis. The extreme assumptions are complete information where all members (but not the econometrician) know the unobservables in the preferences for links for all members and incomplete information where a member only knows its own link specific unobservables. The complete information models are the hardest to estimate and they achieve set and not point identification of the parameters (Miyauchi, 2013 and Sheng, 2014). Leung (2014) considers a model in which an agent only knows the link unobservable when considering to form that link. In this paper we consider a case in which the agent knows her/his own unobserved link preferences, but not those of other agents. So our assumption is between that in Leung (2014) and complete information, and our assumption is also in line with the usual assumption in discrete choice models.

---

[1] Jackson (2008) surveys game-theoretic models of network formation.

A further distinction in the empirical literature regards the data. We can have data on a large number of small networks, e.g., friendship networks in classrooms, or we can have data on a single large network. This paper considers the latter case (Menzel 2015, Leung, 2014, and De Paula, Richards-Shubik and Tamer, 2014 also consider large networks). An advantage of the large single network case is that for a fairly general utility function under a Bayesian Nash equilibrium the link choices converge to the myopic decision rule of a single agent, because the normalized preferences converge to the preferences for this case. This is true even though the link preferences depend on a non-trivial and non-vanishing way on the position of the agent in the network. This simplification only holds if we add a 'sufficient statistic' that captures network position and that is derived by letting the number of agents grow large to the random utility model for links. An implication of this result is that we can estimate preference parameters by a two-step MLE or GMM procedure. Since this procedure only uses the 'first-order condition' for optimal link choice it does not require an assumption on equilibrium selection.

The plan of the paper is as follows. In Section 2 we introduce the model and the specific utility function that we will use. We also discuss the Bayesian Nash equilibrium for the network. In Section 3 we obtain a closed-form expression for the link-formation probability that avoids the solution of an integer program. We also discuss the (uniform) convergence of the choice probabilities if the network size grows without bounds. Section 4 discusses the two-step GMM estimator and Section 5 presents a simulation study. Section 6 considers the extension to undirected networks.

2   Model

Suppose there are n individuals, denoted by $\mathcal{N} = \{1, \ldots, n\}$, who can form links. The links form a network, which we denote by $G \in \mathcal{G}$. This is an $n \times n$ binary matrix. Its (i, j) element $G_{ij} = 1$ if i, j are linked and 0 otherwise. We first consider directed networks, i.e., G_{ij} and G_{ji} may be different. The case of undirected links is discussed in Section 5. Let $G_i = (G_{ij})_{j \neq i} \in \mathcal{G}_i$ be the links of individual i and $G_{-i} = \{G_j\}_{j \neq i} \in \mathcal{G}_{-i}$ the links of individuals other than i.

Each individual i has a vector of observed characteristics $X_i \in \mathcal{X}$ and a vector of unobserved preference shocks $\varepsilon_i = (\varepsilon_{ij})_{j \neq i} \in \mathbb{R}^{n-1}$, where ε_{ij} is i's preference for link ij. We assume that X_i is publicly observed by all the n individuals, but ε_i is only privately observed by individual i, so the information each i knows is (X, ε_i), where $X = (X_i)_{i \in \mathcal{N}}$. Moreover, we assume that the private information is i.i.d. and is independent of the observables.

Assumption 1 *(i) ε_{ij}, $i \neq j \in \mathcal{N}$, are i.i.d. with a distribution $F(\theta_\varepsilon)$ that is absolutely continuous with respect to Lebesgue measure. $F(\theta_\varepsilon)$ is known up to $\theta_\varepsilon \in \Theta_\varepsilon \subset \mathbb{R}^{d_\varepsilon}$. (ii) X and $\varepsilon = (\varepsilon_i)_{i \in \mathcal{N}}$ are independent. (iii) The support of X has finite number of distinct values, $\mathcal{X} = \{x^1, \ldots, x^T\}$.*

Utility   Each individual i in network G has the utility $U_i(G, X, \varepsilon_i)$. We assume it takes the following form.

Assumption 2 *The utility function takes the form*

$$
\begin{aligned}
U_i(G, X, \varepsilon_i) &= \frac{1}{n-1} \sum_{j \neq i} G_{ij} \left(u(X_i, X_j; \beta) + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk} \gamma_1 - \varepsilon_{ij} \right) \\
&\quad + \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} G_{ij} G_{ik} G_{jk} G_{kj} \gamma_2
\end{aligned}
\tag{1}
$$

which is known up to $\theta_u = (\beta, \gamma_1, \gamma_2) \in \Theta_u \subset \mathbb{R}^{d_u}$. u is continuously differentiable in β and uniformly bounded.

In the utility specification in (1), $u(X_i, X_j; \beta)$ captures the utility from direct friend j. An example of $u(X_i, X_j; \beta)$ could be

$$
u(X_i, X_j; \beta) = \beta_0 + \beta_1' X_i + \beta_2' |X_i - X_j|
$$

where the second term is to capture the homophily effect. In addition to direct-friend effects, the utility in (1) also allows for the effects from indirect friends. The γ_1 term represents the effect of friends of friends, and the γ_2 term represents the effect of friends in common, both of which are well documented in the social network literature (Jackson, 2008).

Strategies   The strategies of the players can be modeled by extending the link announcement game in Myerson (1991) to the case of incomplete information. Under incomplete information, each individual i announces simultaneously a vector of links that he intends to form given his information (X, ε_i), namely

$$
S_i(X, \varepsilon_i) = (S_{ij}(X, \varepsilon_i))_{j \neq i} \in \mathcal{S}_i = \{0, 1\}^{n-1}
$$

The strategy profile $S = (S_i)_{i \in \mathcal{N}}$ induces a network $G(S)$. For directed networks, we have $G_{ij}(S) = S_{ij}$ for any $i \neq j \in \mathcal{N}$.

Equilibrium In the context of directed networks with incomplete information, a proper equilibrium concept is Bayesian Nash equilibrium, which we will assume in this paper. A directed network $G \in \mathcal{G}$ is Bayesian Nash equilibrium (BNE) if

$$G_i(X, \varepsilon_i) = \arg \max_{G_i' \in \mathcal{G}_i} \mathbb{E}\left[U_i(G_i', G_{-i}, X, \varepsilon_i) | X, \varepsilon_i\right], \ \forall i \in \mathcal{N}$$

Note that G_i is a function of (X, ε_i) and G_{-i} is a function of (X, ε_{-i}) where $\varepsilon_{-i} = (\varepsilon_j)_{j \neq i}$. Under Assumption 1, G_{-i} is independent of ε_i. Hence, the expected utility of i conditional on (X, ε_i) is given by

$$
\begin{aligned}
\mathbb{E}\left[U_i(G, X, \varepsilon_i) | X, \varepsilon_i\right] &= \mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) | X, \varepsilon_i\right] \\
&= \sum_{g_{-i} \in \mathcal{G}_{-i}} U_i(G_i, g_{-i}, X, \varepsilon_i) P\left(G_{-i} = g_{-i} | X, \varepsilon_i\right) \\
&= \sum_{g_{-i} \in \mathcal{G}_{-i}} U_i(G_i, g_{-i}, X, \varepsilon_i) P\left(G_{-i} = g_{-i} | X\right)
\end{aligned}
$$

The conditional choice probability of i for $g_i \in \mathcal{G}_i$ is then

$$
\begin{aligned}
&\Pr\left(G_i = g_i | X\right) \\
&= \Pr\left(\sum_{g_{-i}} U_i(g_i, g_{-i}, X, \varepsilon_i) \Pr\left(G_{-i} = g_{-i} | X\right) \geq \max_{g_i'} \sum_{g_{-i}} U_i(g_i', g_{-i}, X, \varepsilon_i) \Pr\left(G_{-i} = g_{-i} | X\right) \,\middle|\, X\right)
\end{aligned}
\tag{2}
$$

Define

$$
\begin{aligned}
P_i(X) &= \{\Pr\left(G_i = g_i | X\right), \forall g_i \in \mathcal{G}_i\} \\
P_{-i}(X) &= \{\Pr\left(G_{-i} = g_{-i} | X\right), \forall g_{-i} \in \mathcal{G}_{-i}\} \\
&= \left\{\prod_{j \neq i} \Pr\left(G_j = g_j | X\right), \forall g_j \in \mathcal{G}_j, \forall j \neq i\right\} \\
P(X) &= \{\Pr\left(G_i = g_i | X\right), \forall g_i \in \mathcal{G}_i, \forall i \in \mathcal{N}\}
\end{aligned}
$$

The equations in (2) define a nonlinear system over $P(X)$,

$$P_i(X) = \Gamma_i(X, P_{-i}(X)), \ \forall i \in \mathcal{N} \tag{3}$$

and a fixed point $P(X)$ is a BNE.

In this paper, we will focus on BNE that are symmetric in agents' observables. We say

a BNE is *symmetric* if for i and j with $X_i = X_j$, we have $P_i(X) = P_j(X)$. In network data where individuals do not have identities, it makes sense to consider only symmetric equilibria because otherwise the equilibria and agents' choices may depend on how we label the individuals. More importantly, as we assume only one network is observed, the assumption of symmetric equilibria will be crucial for valid estimation and inference. It can be shown that there exists a symmetric BNE. We assume that the network observed in data is a symmetric BNE.

Lemma 1 *For any X, there exists a symmetric Bayesian Nash equilibrium $P(X)$.*

Proof. See the appendix. ■

3   Convergence of Games

In this section, we aim to show that the set of equilibria in the finite-player game will converge to the set of equilibria in some limit game as n goes to infinity. This is crucial if we want to apply a two-step procedure to estimate the parameter θ, where in the first step we are supposed to estimate the conditional choice probabilities directly from data. Because the equilibria in the finite-player game depend on the number of players n, if we only observe one network, it is impossible to estimate the conditional choice probabilities consistently. However, if the finite-player game converges to some limit game, for sufficiently large n we may approximate DGP well by the limit game, i.e., assuming the data are generated from the limit game rather than the finite-player game. Since the conditional choice probabilities in the limit no longer depend on n, by making use of the symmetry of equilibria we can estimate the conditional choice probabilities consistently using only one observation of the network. Once we have consistent estimates of the conditional choice probabilities in the first step, we can derive valid estimates of the parameter θ.

3.1   The closed form of optimal choices

The difficulty in analyzing the convergence of the conditional choice probabilities in the finite-player game lies in the fact that the expected utility is nonlinear in G_i, which is a binary-valued vector, so the optimal G_i may be solved only numerically. The insight of this paper is to provide an approach to linearize the expected utility function so that the optimal G_i can be solved analytically. To be specific, the expected utility of i given i's information

6

(X, ε_i) under Assumptions 1-2 is

$$\mathbb{E}\left[U_i(G, X, \varepsilon_i)\middle|\, X, \varepsilon_i\right] \;=\; \frac{1}{n-1}\sum_{j\neq i} G_{ij}\left(u(X_i, X_j) + \frac{1}{n-2}\sum_{k\neq i,j}\mathbb{E}\left[G_{jk}|\,X\right]\gamma_1 - \varepsilon_{ij}\right)$$

$$+\frac{1}{(n-1)(n-2)}\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}G_{ik}\mathbb{E}\left[G_{jk}|\,X\right]\mathbb{E}\left[G_{kj}|\,X\right]\gamma_2. \quad (4)$$

By the symmetry of BNE, in a network with a given X any pair (i,j) with the same (X_i, X_j) has the same choice probability $\mathbb{E}\left[G_{ij}|\,X\right]$. It is thus valid to denote

$$p(X_j, X_k; X) \;=\; \mathbb{E}\left[G_{jk}|\,X\right],$$
$$a(X_j, X_k; X) \;=\; p(X_j, X_k; X)\, p(X_k, X_j; X) = \mathbb{E}\left[G_{jk}|\,X\right]\mathbb{E}\left[G_{kj}|\,X\right].$$

Under the assumption that X has a finite support (Assumption 1(iii)), we may represent all possible values of $a(X_j, X_k)$ by a $T \times T$ matrix A

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1T} \\ \vdots & \ddots & \vdots \\ a_{T1} & \cdots & a_{TT} \end{bmatrix} = \begin{bmatrix} a(x^1, x^1) & \cdots & a(x^1, x^T) \\ \vdots & \ddots & \vdots \\ a(x^T, x^1) & \cdots & a(x^T, x^T) \end{bmatrix}. \quad (5)$$

Note that because we count j, k as a common pair of friends if $G_{jk} = 1$ and $G_{kj} = 1$, the friends-in-common term had an undirected flavor, so A is symmetric. If we count j, k as a common pair of friends if either $G_{jk} = 1$ or $G_{kj} = 1$, then $G_{jk}G_{kj}$ is replaced by $G_{jk}+G_{kj}-G_{jk}G_{kj}$ and $a(X_j, X_k)$ is replaced by $p(X_j, X_k)+p(X_k, X_j)-p(X_j, X_k)\,p(X_k, X_j)$, which is a symmetric function, so A is also symmetric. From linear algebra there exist $T \times T$ matrices $\Phi = (\phi_1, \ldots, \phi_T)$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_T)$ such that $A = \Phi\Lambda\Phi'$, where λ_t are the eigenvalues of A and ϕ_t are the eigenvectors that correspond to each λ_t for $t = 1, \ldots, T$. Using the eigenvalue decomposition of A and letting

$$D_t(X_i) \;=\; 1\left\{X_i = x^t\right\}, \quad t = 1, \ldots, T,$$
$$D(X_i) \;=\; (D_1(X_i), \ldots, D_T(X_i))',$$

we may transform the quadratic term in (4) into a form that involves only linear functions of G_{ij}.

Lemma 2 *Under Assumption 1(iii),*

$$\sum_{j\neq i}\sum_{k\neq i,j}G_{ij}G_{ik}a\left(X_j,X_k\right)=\sum_t\lambda_t\left(\sum_{j\neq i}G_{ij}D\left(X_j\right)'\phi_t\right)^2-\sum_{j\neq i}G_{ij}a\left(X_j,X_j\right).$$

Proof. See the appendix. ■

By Lemma 2 the expected utility in (4) can be written as

$$
\begin{aligned}
&\mathbb{E}\left[U_i(G,X,\varepsilon_i)|\,X,\varepsilon_i\right]\\
=\;&\frac{1}{n-1}\sum_{j\neq i}G_{ij}\left[u(X_i,X_j)+\frac{1}{n-2}\sum_{k\neq i,j}p\left(X_j,X_k\right)\gamma_1-\frac{1}{n-2}a\left(X_j,X_j\right)\gamma_2-\varepsilon_{ij}\right]\\
&+\frac{n-1}{n-2}\sum_t\lambda_t\left(\frac{1}{n-1}\sum_{j\neq i}G_{ij}D\left(X_j\right)'\phi_t\right)^2\gamma_2.
\end{aligned}
\tag{6}
$$

Our next objective is to linearize the expected utility, i.e., make it linear in G_i, so that the problem of maximizing $\mathbb{E}\left[U_i(G,X,\varepsilon_i)|\,X,\varepsilon_i\right]$ over G_i can be solved easily with a closed solution. A crucial step in the linearization is to make use of the identity[2]

$$
\left(\frac{1}{n-1}\sum_{j\neq i}G_{ij}D\left(X_j\right)'\phi_t\right)^2=\max_{\tilde{\omega}_t\in\mathbb{R}}\left\{2\left(\frac{1}{n-1}\sum_{j\neq i}G_{ij}D\left(X_j\right)'\phi_t\right)\tilde{\omega}_t-\tilde{\omega}_t^2\right\},\quad t=1,\ldots,T,
\tag{7}
$$

so that that the expected utility is equal to

$$
\begin{aligned}
&\mathbb{E}\left[U_i(G,X,\varepsilon_i)|\,X,\varepsilon_i\right]\\
=\;&\frac{1}{n-1}\sum_{j\neq i}G_{ij}\left[u(X_i,X_j)+\frac{1}{n-2}\sum_{k\neq i,j}p\left(X_j,X_k\right)\gamma_1-\frac{1}{n-2}a\left(X_j,X_j\right)\gamma_2-\varepsilon_{ij}\right]\\
&+\frac{n-1}{n-2}\sum_t\lambda_t\max_{\tilde{\omega}_t\in\mathbb{R}}\left\{2\left(\frac{1}{n-1}\sum_{j\neq i}G_{ij}D\left(X_j\right)'\phi_t\right)\tilde{\omega}_t-\tilde{\omega}_t^2\right\}\gamma_2
\end{aligned}
\tag{8}
$$

With the quadratic terms replaced with the maxima of linear functions, the expected utility in (8) becomes linear in G_i. Therefore, if we can move the maximization over $\tilde{\omega}=(\tilde{\omega}_1,\ldots,\tilde{\omega}_T)'\in\mathbb{R}^T$ to the very beginning of the expected utility (the first equivalence in the display below) and interchange the maximization over $\tilde{\omega}$ with the maximization over G_i (the second equivalence in the display below), the original problem of maximizing the expected utility over G_i can be transformed to an equivalent problem of which the maximization over

---

[2]The identity is a special case of Legendre transform (Rockafellar, 1970). We are grateful to Terrence Tao for suggesting the idea.

G_i part is linear and thus simple to solve. The first equivalence is guaranteed by Assumption 3, and the second equivalence follows by Lemma 3. Moreover, the maximization over $\tilde{\omega}$ is equivalent to the maximization over $\omega = (\omega_1, \ldots, \omega_T)' = \Phi\tilde{\omega} \in \mathbb{R}^T$ (the third equivalence in the display) because $A = \Phi\Lambda\Phi'$ and $\Phi^{-1} = \Phi'$. By the change of variables we can get around the eigenvalues and eigenvectors in the objective function, which may be discontinuous in A.

$$
\max_{G_i} \quad \mathbb{E}\left[U_i(G, X, \varepsilon_i)\mid X, \varepsilon_i\right]
$$

$$
\Leftrightarrow \max_{G_i} \max_{\tilde{\omega}} \quad \frac{1}{n-1} \sum_{j\neq i} G_{ij}\left[u(X_i, X_j) + \frac{1}{n-2}\sum_{k\neq i,j} p\left(X_j, X_k\right)\gamma_1 - \frac{1}{n-2}a\left(X_j, X_j\right)\gamma_2 - \varepsilon_{ij}\right.
$$

$$
\left. +\frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)'\sum_t \lambda_t\phi_t\tilde{\omega}_t\right] - \frac{n-1}{n-2}\gamma_2\sum_t \lambda_t\tilde{\omega}_t^2
$$

$$
\Leftrightarrow \max_{\tilde{\omega}} \max_{G_i} \quad \frac{1}{n-1} \sum_{j\neq i} G_{ij}\left[u(X_i, X_j) + \frac{1}{n-2}\sum_{k\neq i,j} p\left(X_j, X_k\right)\gamma_1 - \frac{1}{n-2}a\left(X_j, X_j\right)\gamma_2 - \varepsilon_{ij}\right.
$$

$$
\left. +\frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)'\Phi\Lambda\tilde{\omega}\right] - \frac{n-1}{n-2}\gamma_2\tilde{\omega}'\Lambda\tilde{\omega}
$$

$$
\Leftrightarrow \max_{\omega} \max_{G_i} \quad \frac{1}{n-1} \sum_{j\neq i} G_{ij}\left[u(X_i, X_j) + \frac{1}{n-2}\sum_{k\neq i,j} p\left(X_j, X_k\right)\gamma_1 - \frac{1}{n-2}a\left(X_j, X_j\right)\gamma_2 - \varepsilon_{ij}\right.
$$

$$
\left. +\frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)'A\omega\right] - \frac{n-1}{n-2}\gamma_2\omega'A\omega \tag{9}
$$

Assumption 3 (Positive externality from friends in common) *(i) $\gamma_2 \geq 0$. (ii) There is $\delta > 0$ such that the eigenvalues of $A(p)$ are nonnegative for any p in the δ-neighborhood of the equilibrium p_0 in the observed data.*

Lemma 3 *For any function $f(x, y) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$, $f(x, y) < \infty$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have*

$$
\max_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} f\left(x, y\right) = \max_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} f(x, y) \tag{10}
$$

This implies that if there is a unique (x^*, y^*) such that $f\left(x^*, y^*\right) = \max_{y\in\mathcal{Y}} \max_{x\in\mathcal{X}} f\left(x, y\right)$, then there is a unique $(\tilde{x}^*, \tilde{y}^*)$ such that $f\left(\tilde{x}^*, \tilde{y}^*\right) = \max_{x\in\mathcal{X}} \max_{y\in\mathcal{Y}} f(x, y)$. In particular, $(\tilde{x}^*, \tilde{y}^*) = (x^*, y^*)$.

Proof. See the appendix. ■

The optimal choice of i for a given (X, ε_i), denoted by

$$
G_{n,i} = (G_{n,ij}\left(X_i, X_j\right))_{j\neq i} = \left(G_{ij}\left(X_i, X_j; X_{-\{i,j\}}, \varepsilon_i\right)\right)_{j\neq i}
$$

can be solved easily from the equivalent problem in (9) since the objective function of (9) is linear in $G_{n,i}$. Furthermore, Lemma 3 ensures that the optimal $G_{n,i}$ solved from the equivalent problem is unique if the optimal $G_{n,i}$ solved from the original problem is also unique, which is true almost surely because of the assumption that ε_i has a continuous distribution.

Theorem 1 *Under Assumptions 1-3, the optimal choice of i is given by*

$$G_{n,ij}\left(X_i, X_j\right) = 1\left\{V_n\left(X_i, X_j\right) + \frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)' A\omega_n\left(X_i\right) - \varepsilon_{ij} \geq 0\right\}, \ \forall j \neq i \qquad (11)$$

where

$$\omega_n\left(X_i\right) = \omega_n\left(X_i; X_{-i}, \varepsilon_i\right)$$

is the optimal solution to

$$\max_{\omega} \frac{1}{n-1} \sum_{j \neq i}\left[V_n\left(X_i, X_j\right) + \frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)' A\omega - \varepsilon_{ij}\right]_+ - \frac{n-1}{n-2}\gamma_2\omega' A\omega \qquad (12)$$

with the notation $[x]_+ = \max\{x, 0\}$ and

$$
\begin{aligned}
V_n\left(X_i, X_j\right) &= V_n\left(X_i, X_j; X_{-\{i,j\}}\right) \\
&= u(X_i, X_j) + \frac{1}{n-2}\sum_{k \neq i,j} p\left(X_j, X_k\right)\gamma_1 - \frac{1}{n-2}a\left(X_j, X_j\right)\gamma_2
\end{aligned}
$$

Furthermore, the optimal $G_{n,ij}\left(X_i, X_j\right)$ and $\omega_n\left(X_i\right)$ are unique almost surely.

3.2 Convergence of conditional choice probabilities

With the closed-form optimal choices, it is straightforward to derive the conditional probability that a pair forms a link. Given $(x_i, x_j) \in \mathcal{X}^2$, the probability that pair i and j form link ij is given by

$$
\begin{aligned}
P_n\left(x_i, x_j; p, \theta\right) &= \mathbb{E}_{\varepsilon_i}\left(G_{n,ij}\left(x_i, x_j; \varepsilon_i\right)\middle| x_i, x_j\right) \\
&= \Pr_{\varepsilon_i}\left(V_n\left(x_i, x_j\right) + \frac{n-1}{n-2}2\gamma_2 D\left(x_j\right)' A\omega_n\left(x_i; \varepsilon_i\right) - \varepsilon_{ij} \geq 0\middle| x_i, x_j\right) \qquad (13)
\end{aligned}
$$

Note that the conditional choice probability is random even if ε_i is integrated out because both $V_n\left(x_i, x_j\right) = V\left(x_i, x_j; X_{-\{i,j\}}\right)$ and $\omega_n\left(x_i; \varepsilon_i\right) = \omega\left(x_i; X_{-\{i,j\}}, \varepsilon_i\right)$ depend on $X_{-\{i,j\}}$ which is random. The dependence of $P_n\left(x_i, x_j\right)$ on n makes it impossible to estimate $P_n\left(x_i, x_j\right)$, as needed in a two-step procedure, if we only assume a single observation of

10

the network. However, if $P_n(x_i, x_j)$ can be approximated sufficiently well by some limiting "conditional choice probability" $P(x_i, x_j)$ that does not depend on n, we may be able to estimate the limit $P(x_i, x_j)$ with a single large observation of the network and use the estimated $P(x_i, x_j)$ as an approximation of $P_n(x_i, x_j)$ for the estimation of the structural parameters. To be specific, define the conditional probability

$$P(x_i, x_j; p, \theta) = \Pr_{\varepsilon_{ij}} \left(V(x_i, x_j) + 2\gamma_2 D(x_j)' A\omega^*(x_i) - \varepsilon_{ij} \geq 0 \middle| x_i, x_j \right), \qquad (14)$$

where $\omega^*(x_i)$ is the optimal solution to

$$\max_{\omega} \mathbb{E}_{X_j, \varepsilon_{ij}} \left[V(x_i, X_j) + 2\gamma_2 D(X_j)' A\omega - \varepsilon_{ij} \right]_+ - \gamma_2 \omega' A\omega \qquad (15)$$

and

$$V(x_i, x_j) = u(x_i, x_j) + \mathbb{E}_{X_k} [p(x_j, X_k)] \gamma_1$$

The $P(x_i, x_j)$ in (14) may be understood as the conditional choice probability in a "limiting game" where each player chooses optimal links by making binary choices, i.e., $G_{ij} = 1$ if $V(x_i, x_j) + 2\gamma_2 D(x_j)' A\omega^*(x_i) - \varepsilon_{ij} \geq 0$ and 0 otherwise, with $\omega^*(x_i)$ controlling for the interactions from i's other links due to friends in common, and p controlling for the interactions from other players, both of which are determined in equilibrium. Our goal in this section is to prove that the conditional choice probability in the finite-player game converges to the conditional choice probability in the limiting game uniformly as n increases to infinity, namely,

$$\sup_{p, \theta} |P_n(x_i, x_j; p, \theta) - P(x_i, x_j; p, \theta)| \xrightarrow{p} 0$$

as $n \to \infty$.

Assumption 4 *(i) Normalize θ appropriately so that the density of ε, $f_\varepsilon(\varepsilon)$, satisfies $f_\varepsilon(\varepsilon) \leq 1$ for any $\varepsilon \in \mathbb{R}$. (ii) $\gamma_2 \in \left[0, \frac{1}{2} - \delta\right]$ for some $\delta > 0$.*

Proposition 1 *Under Assumptions 1-4(assume $\gamma_2 \neq 0$), for any $x_i \in X$, there is a unique $A\omega^*(x_i)$ that solves the problem in (15).*

Proof. See the appendix. ∎

Remark 1 *Note that ω^* may not be unique if A is singular. Nevertheless, the indeterminacy of ω^* does not cause an identification problem because only $A\omega$ enters the conditional choice probability P_n and P, and Proposition 1 ensures that there is a unique $A\omega^*$ that solves (15).*

Assumption 5 Θ *is compact.*

Lemma 4 *Under Assumptions 1-5, for any* $x_i \in \mathcal{X}$,

$$\sup_{p,\theta} |A\omega_n (x_i; p, \theta) - A\omega^* (x_i; p, \theta)| \xrightarrow{p} 0$$

Proof. See the appendix. ■

Theorem 2 *Under Assumptions 1-5, for any* $(x_i, x_j) \in \mathcal{X}^2$,

$$\sup_{p,\theta} |P_n (x_i, x_j; p, \theta) - P (x_i, x_j; p, \theta)| \xrightarrow{p} 0.$$

Proof. See the appendix. ■

4   Estimation

In this section, we discuss how to estimate parameter θ. The previous section says that the conditional choice probabilities in finite games can be well approximated by the conditional choice probabilities in the limiting game so long as the number of agents is large. While the true DGP is a finite game, that is, we sample n individuals at random from the population and let these n individuals play the formation game with n players, the convergence of finite games implies that we can approximate the true DGP by the limiting CCP. Such an approximation serves two roles in the estimation. First, the limiting CCP does not depend on n so that the model we need to estimate would not change with n. This is crucial if we assume only one network is observed in the data. Second, players in the limiting game make link decisions in the myopic way: a player decides to form a link as in a binary choice model, that is, he forms the link if the latent utility from the link is nonnegative, where the latent utility equals the expected marginal utility from the link plus a function of additional control variables $\omega (x; p, \theta)$ that take care of the interdependence between links of the player. This implies that links in the limiting game are independent and identically distributed. While links in the observed networks may be correlated, under the approximation we can view them as a i.i.d. sample generated from the distribution determined by the limiting CCP and the relative frequency of links for each given type of players could provide consistent estimates for the equilibrium p.

Our approach only requires one single large network. If more than one network is observed in the data, we can proceed network by network. That is, we treat links in each network as a sample generated from the limiting CCP $P (x_i, x_j; p, \theta)$ with p the equilibrium in that network and obtain from each network a set of restrictions for estimation.

We propose a two-step procedure to estimate θ. In the first stage, we estimate the equilibrium probabilities $p = \{p(x^s, x^t)\}_{s,t=1}^T$ by counting the relative frequency that two players with the observed characteristics (x^s, x^t) form a link. That is, we estimate $p(x^s, x^t)$ by

$$\hat{p}_n\left(x^s, x^t\right) = \frac{\sum_{i=1}^n \sum_{j>i} 1\{G_{ij} = 1, X_i = x^s, X_j = x^t\}}{\sum_{i=1}^n \sum_{j>i} 1\{X_i = x^s, X_j = x^t\}}$$

After we get the first-stage estimate $\hat{p}_n$, we estimate parameter θ from the restrictions that are derived from the equilibrium condition in the limiting game, with p replaced with the estimated $\hat{p}_n$. We can estimate θ by GMM or MLE as in the standard literature on two-step estimation. Note that while the limiting game may have more than one equilibrium, we do not need to impose any restrictions on equilibrium selection because the equilibrium p is directly estimated from the data.

Example 1 *We can estimate θ by MLE. Let $P(X_i, X_j; p, \theta)$ be the limiting CCP. Because the limiting game is a good approximation of the true DGP, we may properly specify the log likelihood of the data as*

$$\ln L(\theta, p) = \sum_{i=1}^n \sum_{j>i} G_{ij} \ln P(X_i, X_j; p, \theta) + (1 - G_{ij}) \ln(1 - P(X_i, X_j; p, \theta))$$

The MLE estimate of θ is the maximizer of the log likelihood with p replaced by the estimate $\hat{p}_n$

$$\hat{\theta}_n = \arg\max_\theta \ln L(\theta, \hat{p}_n)$$

Example 2 *If we estimate θ by GMM, the equilibrium condition in the limiting game gives the conditional moment restrictions*

$$m_0(\theta, p; x, x') = \mathbb{E}[G_{ij} - P(x, x'; p, \theta)| x, x'] = 0, \quad \forall(x, x') \in \mathcal{X}^2$$

These restrictions should hold approximately in finite games with large n since the CCP in finite games converges to the limiting CCP as n increases to infinity. We estimate θ from the sample analogue of these moment restrictions with p replaced by the estimate $\hat{p}_n$

$$m_n(\theta, \hat{p}_n; x, x') = \frac{1}{\binom{n}{2}} \sum_{i=1}^n \sum_{j>i} (G_{ij} - P(x, x'; \hat{p}_n, \theta)) 1\{X_i = x, X_j = x'\}, \quad \forall(x, x') \in \mathcal{X}^2$$

Stack the T^2 sample moments into a vector

$$\mathbf{m}_n(\theta, \hat{p}_n) = \left[m_n(\theta, \hat{p}_n; x^1, x^1), m_n(\theta, \hat{p}_n; x^1, x^2), \ldots, m_n(\theta, \hat{p}_n; x^T, x^T)\right]'$$

The GMM estimate $\hat{\theta}_n$ minimizes the criterion function

$$\hat{\theta}_n = \arg\min_{\theta} \mathbf{m}_n \left(\theta, \hat{p}_n \right)' V \left(\theta \right) \mathbf{m}_n \left(\theta, \hat{p}_n \right)$$

with $V\left(\theta\right)$ a positive definite weighting matrix.

5 Undirected Networks

In this section we look at undirected networks. We show that, with modifications in the CCP and limiting game, the idea in Section 3 would work in undirected networks. Let S_{ij} indicate a directed link and $G_{ij} = S_{ij}S_{ji}$ an undirected link. We consider an undirected counterpart of the utility in (1). It depends on the formed links G rather than the proposals.

$$
\begin{aligned}
U_i\left(G, X, \varepsilon_i\right) = \quad & \frac{1}{n-1} \sum_{j \neq i} G_{ij} \left(u\left(X_i, X_j\right) + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk}\gamma_1 - \varepsilon_{ij} \right) \\
& + \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} G_{ij} G_{ik} G_{jk} \gamma_2 \\
= \quad & \frac{1}{n-1} \sum_{j \neq i} S_{ij}S_{ji} \left(u\left(X_i, X_j\right) + \frac{1}{n-2} \sum_{k \neq i,j} S_{jk}S_{kj}\gamma_1 - \varepsilon_{ij} \right) \\
& + \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} S_{ij}S_{ji}S_{ik}S_{ki}S_{jk}S_{kj} \gamma_2
\end{aligned}
$$

Assume Bayesian Nash equilibrium as before. We say an undirected network G is a Bayesian Nash Equilibrium if it is induced by a Bayesian Nash equilibrium strategy profile $\{S_i(X, \varepsilon_i)\}_{i \in \mathcal{N}}$, i.e.,

$$S_i\left(X, \varepsilon_i\right) = \arg\max_{S_i' \in \mathcal{S}_i} \mathbb{E}\left[U_i\left(G\left(S_i', S_{-i}\right), X, \varepsilon_i\right)\right) \middle| X, \varepsilon_i\right], \ \ \forall i \in \mathcal{N}$$

Note that the choice probability in (2) still applies if we replace G by S and g by s. We define $P_i(X) = \Pr(S_i = s_i|X)$. Then (3) still defines the equilibrium.

Remark 2 *A potential concern with Nash is that in undirected networks players may coordinate. This is reasonable under complete information, where pairwise stability (Jackson and Wolinsky (1996)) and Nash equilibrium are nonnested and neither of them implies the other. However, under incomplete information players won't be able to coordinate even in undirected networks; because i does not observe ε_{ji}, he cannot predict what j proposes and coordinate on that. In fact, if we define a Bayesian version of the pairwise stability, that is,*

a network G is Bayesian pairwise stable if

$$G_{ij} = 1 \Leftrightarrow \Delta_{ij}\mathbb{E}\left[U_i(G, X, \varepsilon_i)| X, \varepsilon_i\right] \geq 0 \ \& \ \Delta_{ji}\mathbb{E}\left[U_j(G, X, \varepsilon_j)| X, \varepsilon_j\right] \geq 0, \quad \forall i \neq j$$

where $\Delta_{ij}\mathbb{E}\left[U_i(G, X, \varepsilon_i)| X, \varepsilon_i\right]$ is the expected marginal utility if i proposes the link with j,

$$
\begin{aligned}
\Delta_{ij}\mathbb{E}\left[U_i(G, X, \varepsilon_i)| X, \varepsilon_i\right] &= \mathbb{E}\left[U_i(G\left(S_{ij} = 1, S_{i-ij}, S_{-i}\right), X, \varepsilon_i)| X, \varepsilon_i\right] \\
&\quad -\mathbb{E}\left[U_i(G\left(S_{ij} = 0, S_{i-ij}, S_{-i}\right), X, \varepsilon_i)| X, \varepsilon_i\right]
\end{aligned}
$$

and similar for $\Delta_{ji}\mathbb{E}\left[U_j(G, X, \varepsilon_j)| X, \varepsilon_j\right]$, then any undirected network that is Bayesian Nash must also be Bayesian pairwise stable. This is because for a Bayesian Nash G, $G_{ij} = 1$ if and only if $S_{ij} = S_{ji} = 1$ are optimal, so the expected marginal utility from the link must be nonnegative for both i and j. It is thus enough to consider Bayesian Nash equilibrium.

Define the probability i proposes to link to j by $p(X_i, X_j)$ and the probability i proposes to link to both j and k by $q(X_i, X_j, X_k)$

$$
\begin{aligned}
p(X_i, X_j) &= \Pr(S_{ij} = 1|X) \\
q(X_i, X_j, X_k) &= \Pr(S_{ij} = 1, S_{ik} = 1|X)
\end{aligned}
$$

The expected utility of i is given by

$$
\begin{aligned}
&\mathbb{E}\left[U_i\left(G, X, \varepsilon_i\right)| X, \varepsilon_i\right] \\
&= \frac{1}{n-1}\sum_{j\neq i} S_{ij}\left(\mathbb{E}\left[S_{ji}| X\right]\left(u\left(X_i, X_j\right) - \varepsilon_{ij}\right) + \frac{1}{n-2}\sum_{k\neq i,j}\mathbb{E}\left[S_{ji}S_{jk}S_{kj}| X\right]\gamma_1\right) \\
&\quad + \frac{1}{(n-1)(n-2)}\sum_{j\neq i}\sum_{k\neq i,j} S_{ij}S_{ik}\mathbb{E}\left[S_{ji}S_{jk}S_{ki}S_{kj}| X\right]\gamma_2 \\
&= \frac{1}{n-1}\sum_{j\neq i} S_{ij}\left(p(X_j, X_i)\left(u\left(X_i, X_j\right) - \varepsilon_{ij}\right) + \frac{1}{n-2}\sum_{k\neq i,j} q(X_j, X_i, X_k)p(X_k, X_j)\gamma_1\right) \\
&\quad + \frac{1}{(n-1)(n-2)}\sum_{j\neq i}\sum_{k\neq i,j} S_{ij}S_{ik}q(X_j, X_i, X_k)q(X_k, X_i, X_j)\gamma_2
\end{aligned}
$$

It takes the same forms as in the directed case (4), with $a\left(X_j, X_k\right)$ and A replaced by $a\left(X_i, X_j, X_k\right)$ and $A\left(X_i\right)$ defined by

$$a\left(X_i, X_j, X_k\right) = q(X_j, X_i, X_k)q(X_k, X_i, X_j)$$

and

$$A\left(X_i\right) = \begin{bmatrix} a\left(X_i, x^1, x^1\right) & \cdots & a\left(X_i, x^1, x^T\right) \\ \vdots & \ddots & \vdots \\ a\left(X_i, x^T, x^1\right) & \cdots & a\left(X_i, x^T, x^T\right) \end{bmatrix}$$

More important, like in the directed case, $A\left(X_i\right)$ is also a symmetric matrix, so the argument in Section 3.1 still applies if $A\left(X_i\right)$ satisfies Assumption 3 for any $X_i \in \mathcal{X}$. Theorem 1 holds with $G_{n,ij}$ replaced by $S_{n,ij}$ and slight modifications of the functions.

Corollary 1 *Under Assumptions 1-3 (with A replaced by $A\left(X_i\right)$), the optimal choice of i is given by*

$$S_{n,ij}\left(X_i, X_j\right) = 1\left\{V_n\left(X_i, X_j\right) + \frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)' A\left(X_i\right)\omega_n\left(X_i\right) - p(X_j, X_i)\varepsilon_{ij} \geq 0\right\}, \quad \forall j \neq i \tag{16}$$

where

$$\omega_n\left(X_i\right) = \omega_n\left(X_i; X_{-i}, \varepsilon_i\right)$$

is the optimal solution to

$$\max_\omega \frac{1}{n-1}\sum_{j \neq i}\left[V_n\left(X_i, X_j\right) + \frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)' A\left(X_i\right)\omega - p(X_j, X_i)\varepsilon_{ij}\right]_+ - \frac{n-1}{n-2}\gamma_2\omega' A\left(X_i\right)\omega \tag{17}$$

with

$$\begin{aligned} V_n\left(X_i, X_j\right) &= V_n\left(X_i, X_j; X_{-\{i,j\}}\right) \\ &= p(X_j, X_i)u(X_i, X_j) + \frac{1}{n-2}\sum_{k \neq i,j} q(X_j, X_i, X_k)p\left(X_k, X_j\right)\gamma_1 \\ &\quad - \frac{1}{n-2}a\left(X_i, X_j, X_j\right)\gamma_2 \end{aligned}$$

Furthermore, the optimal $G_{n,ij}\left(X_i, X_j\right)$ and $\omega_n\left(X_i\right)$ are unique almost surely.

The conditional probabilities of i proposing one link and two links are given by

$$\begin{aligned} P_n(x_i, x_j; p, q, \theta) &= \mathbb{E}_{\varepsilon_i}(S_{n,ij}\left(x_i, x_j; \varepsilon_i\right)|x_i, x_j) \\ &= \Pr_{\varepsilon_i}(V_n\left(x_i, x_j\right) + \frac{n-1}{n-2}2\gamma_2 D\left(x_j\right)' A\left(x_i\right)\omega_n\left(x_i; \varepsilon_i\right) - p(x_j, x_i)\varepsilon_{ij} \geq 0|x_i, x_j) \end{aligned}$$

16

and

$$Q_n(x_i, x_j, x_k; p, q, \theta) = \mathbb{E}_{\varepsilon_i}\left(S_{n,ij}(x_i, x_j; \varepsilon_i) S_{n,ik}(x_i, x_k; \varepsilon_i) \middle| x_i, x_j, x_k\right)$$

$$= \mathrm{Pr}_{\varepsilon_i}\left(V_n(x_i, x_j) + \frac{n-1}{n-2}2\gamma_2 D(x_j)' A(x_i) \omega_n(x_i; \varepsilon_i) - p(x_j, x_i)\varepsilon_{ij} \geq 0 \,\&\right.$$

$$\left. V_n(x_i, x_k) + \frac{n-1}{n-2}2\gamma_2 D(x_k)' A(x_i) \omega_n(x_i; \varepsilon_i) - p(x_k, x_i)\varepsilon_{ik} \geq 0 \middle| x_i, x_j, x_k\right)$$

They depend on ω_n in the way similar to choice probability in the directed case. We expect them to converge to the limiting probabilities

$$P(x_i, x_j; p, q, \theta) = \mathrm{Pr}_{\varepsilon_{ij}}(V(x_i, x_j) + 2\gamma_2 D(x_j)' A(x_i) \omega^*(x_i) - p(x_j, x_i)\varepsilon_{ij} \geq 0 | x_i, x_j)$$

and

$$Q(x_i, x_j, x_k; p, q, \theta) = P(x_i, x_j; p, q, \theta) \cdot P(x_i, x_k; p, q, \theta)$$

$$= \mathrm{Pr}_{\varepsilon_{ij}}\left(V(x_i, x_j) + 2\gamma_2 D(x_j)' A(x_i) \omega^*(x_i) - p(x_j, x_i)\varepsilon_{ij} \geq 0 \middle| x_i, x_j, x_k\right)$$

$$\mathrm{Pr}_{\varepsilon_{ik}}\left(V(x_i, x_k) + 2\gamma_2 D(x_k)' A(x_i) \omega^*(x_i) - p(x_k, x_i)\varepsilon_{ik} \geq 0 \middle| x_i, x_j, x_k\right)$$

where $\omega^*(x_i)$ is the optimal solution to

$$\max_{\omega} \mathbb{E}_{X_j, \varepsilon_{ij}}\left[V(x_i, X_j) + 2\gamma_2 D(X_j)' A(x_i) \omega - p(X_j, x_i)\varepsilon_{ij}\right]_+ - \gamma_2 \omega' A(x_i) \omega$$

and

$$V(x_i, x_j) = p(x_j, x_i)u(x_i, x_j) + \mathbb{E}_{X_k}q(x_j, x_i, X_k)p(X_k, x_j)\gamma_1$$

If the assumptions required in Theorem 2 hold in the undirected case, we can show similarly that

$$\sup_{p,q,\theta}|P_n(x_i, x_j; p, q, \theta) - P(x_i, x_j; p, q, \theta)| = o_p(1)$$

$$\sup_{p,q,\theta}|Q_n(x_i, x_j; p, q, \theta) - Q(x_i, x_j; p, q, \theta)| = o_p(1)$$

5.1 Estimation

For finite n the BNE choice probabilities are such that

$$\mathrm{Pr}(S_{ij} = 1, S_{ik} = 1|X) \neq \mathrm{Pr}(S_{ij} = 1|X)\mathrm{Pr}(S_{ik} = 1|X)$$

even though ε_{ij} and ε_{ik} are independent. The dependence between the choices is through ω that depends for finite n on ε_i. However if $n \to \infty$ then S_{ij} and S_{ik} are independent. So if we consider the estimation we can assume that

$$\Pr(S_{ij} = 1, S_{ik} = 1|X) = \Pr(S_{ij} = 1|X)\Pr(S_{ik} = 1|X)$$

This simplifies the estimation problem. Note that in the limit, because

$$q(x_j, x_i, x_k) = \mathbb{E}[S_{ji}S_{jk}|X]$$

we have

$$q(x_j, x_i, x_k) = p(x_j, x_i)p(x_j, x_k)$$

so that

$$
\begin{aligned}
a(x_i, x_j, x_k) &= p(x_j, x_i)p(x_k, x_i)p(x_j, x_k)p(x_k, x_j) \\
&= p(x_j, x_i)p(x_k, x_i)r(x_j, x_k)
\end{aligned}
$$

with $r(x_j, x_k)$ the probability of an undirected link between j and k that can be estimated from the data.

In the limit the BNE structural choice probabilities under Assumptions 1-3 are myopic with choice probability (choice is that i wants a link with j, i.e., $S_{ij} = 1$) if n is large

$$P(x_i, x_j; p, \theta) = F_\varepsilon \left(u(x_i, x_j) + \gamma_1 \mathbb{E}_{X_k}[r(x_j, X_k)] + 2\gamma_2 D(x_j)' \frac{A(x_i)}{p(x_j, x_i)}\omega^*(x_i) \right) \qquad (18)$$

with $\omega^*(x_i)$ the optimal solution to

$$\mathbb{E}_{X_j, \varepsilon_{ij}} p(X_j, x_i) \left[u(x_i, X_j) + \gamma_1 \mathbb{E}_{X_k}(r(X_j, X_k)) + 2\gamma_2 D(X_j)' \frac{A(x_i)}{p(X_j, x_i)}\omega - \varepsilon_{ij} \right]_+ - \gamma_2 \omega' A(x_i)\omega$$

Note that $r(x_j, x_k)$ can be directly estimated from data on undirected links. We also have

$$r(x_i, x_j) = p(x_i, x_j)p(x_j, x_i)$$

Since the left-hand side can be estimated, we have $\frac{1}{2}T(T + 1)$ equations in T^2 unknowns. The BNE structural probability of an undirected links is

$$R(x_i, x_j; p, \theta) = P(x_i, x_j; p, \theta)P(x_j, x_i; p, \theta)$$

18

So the restrictions we have is that for all $i < j$

$$
\begin{aligned}
r(x_i, x_j) &= P(x_i, x_j; p, \theta) P(x_j, x_i; p, \theta) \\
r(x_i, x_j) &= p(x_i, x_j) p(x_j, x_i)
\end{aligned}
\tag{19}
$$

With T types these are $T^2 + T$ restrictions on $T^2 + \dim(\theta)$ unknowns, so the order condition for identification is satisfied if $T \geq \dim(\theta)$. We may treat p as an additional parameter and estimate θ and p from (19) using one-step GMM.

6 Continuous X

In this section we consider the case when X has a continuous distribution. We show that for continuous X the analysis in Section 3 can still apply if we rewrite it in the language of infinitely dimensional functional spaces. We start with directed networks and the analysis of undirected networks is similar. Let $L_2(\mathcal{X}) = \left\{ h : \int_{\mathcal{X}} h^2(x)\, dx < \infty \right\}$ be the set of functions that are square integrable on $\mathcal{X}$. For any functions $h_1, h_2 \in L_2(\mathcal{X})$, define the inner product $\langle h_1, h_2 \rangle = \int_{\mathcal{X}} h_1(x) h_2(x)\, dx$. Define the operator $A : L_2(\mathcal{X}) \to L_2(\mathcal{X})$ by

$$
(Ah)(y) = \int_{\mathcal{X}} h(x)\, a(x, y)\, dx.
$$

Note that because we count j, k as a common pair of friends if $G_{jk} = G_{kj} = 1$, so $a(X_j, X_k) = a(X_k, X_j)$.[3] Symmetry of a implies that for any $h_1, h_2 \in L_2(\mathcal{X})$,

$$
\begin{aligned}
\langle Ah_1, h_2 \rangle &= \int_{\mathcal{X}^2} h_1(x) h_2(y)\, a(x, y)\, dxdy \\
&= \int_{\mathcal{X}^2} h_1(x) h_2(y)\, a(y, x)\, dxdy \\
&= \langle h_1, Ah_2 \rangle,
\end{aligned}
$$

so A is a self-adjoint operator. Since A is real and self adjoint, it has real eigenvalues and eigenvectors $\{\lambda_t, \phi_t, t = 1, 2, \ldots\}$. These are the solutions to $A\phi_t = \lambda_t \phi_t$, $t = 1, 2, \ldots$, and are analogous to the eigenvalues and eigenvectors of a real symmetric matrix (the case if X has a finite discrete support). Like in Lemma 2, we can transform the quadratic term in (4) into a canonical form that involves only squares of linear functions of G_{ij}.

---

[3]If we count j, k as a common pair of friends if either $G_{jk} = 1$ or $G_{kj} = 1$, then $G_{jk} G_{kj}$ is replaced by $G_{jk} + G_{kj} - G_{jk} G_{kj}$ and $a(X_j, X_k)$ is replaced by $p(X_j, X_k) + p(X_k, X_j) - p(X_j, X_k) p(X_k, X_j)$, which is also symmetric in X_j and X_k.

Lemma 5

$$\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}G_{ik}a\left(X_j,X_k\right) = \sum_{t=1}^{\infty}\lambda_t\left(\sum_{j\neq i}G_{ij}\phi_t\left(X_j\right)\right)^2 - \sum_{j\neq i}G_{ij}a\left(X_j,X_j\right).$$

Proof. See the appendix. ∎

Rewriting the double sum as in the lemma and applying Legendre transform again to the square terms

$$\left(\frac{1}{n-1}\sum_{j\neq i}G_{ij}\phi_t\left(X_j\right)\right)^2 = \max_{\tilde\omega_t\in\mathbb{R}}\left\{2\left(\frac{1}{n-1}\sum_{j\neq i}G_{ij}\phi_t\left(X_j\right)\right)\tilde\omega_t - \tilde\omega_t^2\right\}, \quad t=1,2,\dots$$

we obtain the expected utility

$$
\begin{aligned}
&\mathbb{E}\left[U_i(G,X,\varepsilon_i)\,\middle|\,X,\varepsilon_i\right]\\
&= \frac{1}{n-1}\sum_{j\neq i}G_{ij}\left[u(X_i,X_j)+\frac{1}{n-2}\sum_{k\neq i,j}p\left(X_j,X_k\right)\gamma_1-\frac{1}{n-2}a\left(X_j,X_j\right)\gamma_2-\varepsilon_{ij}\right]\\
&\quad +\frac{n-1}{n-2}\sum_{t=1}^{\infty}\lambda_t\max_{\tilde\omega_t\in\mathbb{R}}\left\{2\left(\frac{1}{n-1}\sum_{j\neq i}G_{ij}\phi_t\left(X_j\right)\right)\tilde\omega_t-\tilde\omega_t^2\right\}\gamma_2
\end{aligned}
$$

As in the discrete case, under Assumption 3 we can interchange the maximization over $\tilde\omega$ with the maximization over G_i. Moreover, we can define $\omega\left(x\right)=\sum_{t=1}^{\infty}\phi_t\left(x\right)\tilde\omega_t$, so $\tilde\omega_t = \langle\omega,\phi_t\rangle$, $t=1,2,\dots$, and maximizing over $\tilde\omega$ is then equivalent to maximizing over ω as shown below because $\sum_{t=1}^{\infty}\lambda_t\phi_t\left(X_j\right)\tilde\omega_t = \sum_{t=1}^{\infty}\left(A\phi_t\right)\left(X_j\right)\tilde\omega_t = \left(A\omega\right)\left(X_j\right)$, and $\sum_{t=1}^{\infty}\lambda_t\tilde\omega_t^2 = \langle\sum_{t=1}^{\infty}\phi_t\tilde\omega_t,\sum_{t=1}^{\infty}\lambda_t\phi_t\tilde\omega_t\rangle = \langle\omega,A\omega\rangle$.

$$
\begin{aligned}
&\max_{G_i}\quad \mathbb{E}\left[U_i(G,X,\varepsilon_i)\,\middle|\,X,\varepsilon_i\right]\\[4pt]
\Leftrightarrow\;&\max_{\tilde\omega\in\mathbb{R}^{\mathbb{N}}}\max_{G_i}\quad \frac{1}{n-1}\sum_{j\neq i}G_{ij}\left[u(X_i,X_j)+\frac{1}{n-2}\sum_{k\neq i,j}p\left(X_j,X_k\right)\gamma_1-\frac{1}{n-2}a\left(X_j,X_j\right)\gamma_2-\varepsilon_{ij}\right.\\
&\qquad\qquad\qquad\left. +\frac{n-1}{n-2}2\gamma_2\sum_{t=1}^{\infty}\lambda_t\phi_t\left(X_j\right)\tilde\omega_t\right]-\frac{n-1}{n-2}\gamma_2\sum_{t=1}^{\infty}\lambda_t\tilde\omega_t^2\\[4pt]
\Leftrightarrow\;&\max_{\omega\in L_2(\mathcal{X})}\max_{G_i}\quad \frac{1}{n-1}\sum_{j\neq i}G_{ij}\left[u(X_i,X_j)+\frac{1}{n-2}\sum_{k\neq i,j}p\left(X_j,X_k\right)\gamma_1-\frac{1}{n-2}a\left(X_j,X_j\right)\gamma_2-\varepsilon_{ij}\right.\\
&\qquad\qquad\qquad\left. +\frac{n-1}{n-2}2\gamma_2\left(A\omega\right)\left(X_j\right)\right]-\frac{n-1}{n-2}\gamma_2\left\langle\omega,A\omega\right\rangle
\end{aligned}
$$

The optimal choice of an individual has a simple closed form that is similar to that in the discrete case.

Corollary 2 *Under Assumptions 1-3, the optimal choice of i is given by*

$$G_{n,ij}(X_i, X_j) = 1\left\{V_n(X_i, X_j) + \frac{n-1}{n-2}2\gamma_2(A\omega_n(X_i))(X_j) - \varepsilon_{ij} \geq 0\right\}, \ \forall j \neq i \quad (20)$$

where

$$(\omega_n(X_i))(x) = (\omega_n(X_i; X_{-i}, \varepsilon_i))(x) \in L_2(\mathcal{X})$$

is the optimal solution to

$$\max_\omega \frac{1}{n-1}\sum_{j\neq i}\left[V_n(X_i, X_j) + \frac{n-1}{n-2}2\gamma_2(A\omega)(X_j) - \varepsilon_{ij}\right]_+ - \frac{n-1}{n-2}\gamma_2\langle\omega, A\omega\rangle \quad (21)$$

and

$$\begin{aligned} V_n(X_i, X_j) &= V_n\left(X_i, X_j; X_{-\{i,j\}}\right) \\ &= u(X_i, X_j) + \frac{1}{n-2}\sum_{k\neq i,j}p(X_j, X_k)\gamma_1 - \frac{1}{n-2}a(X_j, X_j)\gamma_2 \end{aligned}$$

Furthermore, the optimal $G_{n,ij}(X_i, X_j)$ and $(\omega_n(X_i))(x)$ are unique almost surely.

7   Simulation

To be completed.

8   Conclusion

To be completed

9   Appendix: Proofs

Proof of Lemma 1. Define the set of symmetric $P(X)$

$$\mathcal{P}_s(X) = \left\{P(X) \in [0,1]^{n2^{n-1}} : P_i(X) = P_j(X) \text{ if } X_i = X_j\right\}$$

It is clear that $\mathcal{P}_s(X)$ is a convex, closed and bounded subset of $[0,1]^{n2^{n-1}}$. Equations in (3) forms a mapping from $\mathcal{P}_s(X)$ to $\mathcal{P}_s(X)$, because if $P(X) \in \mathcal{P}_s(X)$, then $\Gamma_i(X, P_{-i}(X)) = \Gamma_j(X, P_{-j}(X))$, so $\Gamma(X, P_{-i}(X))$ is also symmetric. The mapping is continuous in $P(X)$ because the expected utilities are continuous in $P_{-i}(X)$ and ε_i has a continuous distribution under Assumption 1. By Brouwer's fixed point theorem there is a fixed point. ∎

Proof of Lemma 2. For simplicity of notation write $D_{it} = D_t(X_i)$ and $D_i = D(X_i)$.

$$
\begin{aligned}
\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}G_{ik}a(X_j,X_k) &= \sum_{j\neq i}\sum_{k\neq i}G_{ij}G_{ik}a(X_j,X_k) - \sum_{j\neq i}G_{ij}a(X_j,X_j) \\
&= \sum_{j\neq i}\sum_{k\neq i}G_{ij}G_{ik}\sum_s\sum_t a_{st}D_{js}D_{kt} - \sum_{j\neq i}G_{ij}a(X_j,X_j) \\
&= \sum_s\sum_t a_{st}\sum_{j\neq i}G_{ij}D_{js}\sum_{k\neq i}G_{ik}D_{kt} - \sum_{j\neq i}G_{ij}a(X_j,X_j) \\
&= \left(\sum_{j\neq i}G_{ij}D_j'\right)A\left(\sum_{j\neq i}G_{ij}D_j\right) - \sum_{j\neq i}G_{ij}a(X_j,X_j)
\end{aligned}
$$

By eigenvalue decomposition of A,

$$
\begin{aligned}
\left(\sum_{j\neq i}G_{ij}D_j'\right)A\left(\sum_{j\neq i}G_{ij}D_j\right) &= \left(\sum_{j\neq i}G_{ij}D_j'\right)\Phi\Lambda\Phi'\left(\sum_{j\neq i}G_{ij}D_j\right) \\
&= \left(\sum_{j\neq i}G_{ij}D_j'\Phi\right)\Lambda\left(\sum_{j\neq i}G_{ij}\Phi'D_j\right) \\
&= \sum_t \lambda_t\left(\sum_{j\neq i}G_{ij}D_j'\phi_t\right)^2
\end{aligned}
$$

Therefore,

$$
\sum_{j\neq i}\sum_{k\neq i,j}G_{ij}G_{ik}a(X_j,X_k) = \sum_t\lambda_t\left(\sum_{j\neq i}G_{ij}D_j'\phi_t\right)^2 - \sum_{j\neq i}G_{ij}a(X_j,X_j)
$$

∎

Proof of Lemma 3. Note that

$$
\begin{aligned}
\max_{x\in\mathcal{X}} f(x,y) &\geq f(x,y), \quad \forall(x,y)\in\mathcal{X}\times\mathcal{Y} \\
\Rightarrow \max_{y\in\mathcal{Y}}\max_{x\in\mathcal{X}} f(x,y) &\geq \max_{y\in\mathcal{Y}} f(x,y), \quad \forall x\in\mathcal{X} \\
\Rightarrow \max_{y\in\mathcal{Y}}\max_{x\in\mathcal{X}} f(x,y) &\geq \max_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}} f(x,y)
\end{aligned}
$$

Similarly,

$$\max_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y) \geq \max_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y)$$

Hence (10) is proved. From (10), any $(\tilde{x}^*, \tilde{y}^*)$ that maximizes the RHS of (10) should satisfy

$$f(\tilde{x}^*, \tilde{y}^*) = \max_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y) = \max_{y \in \mathcal{Y}} \max_{x \in \mathcal{X}} f(x, y) = f(x^*, y^*)$$

By uniqueness of (x^*, y^*), we have $(\tilde{x}^*, \tilde{y}^*) = (x^*, y^*)$. $\blacksquare$

Proof of Proposition 1. Let f_X denote the density of X_i.

$$\mathbb{E}_{X_j, \varepsilon_{ij}} \left[V(x_i, X_j) + 2\gamma_2 D(X_j)' A\omega - \varepsilon_{ij} \right]_+ = \sum_{t=1}^{T} \mathbb{E}_{\varepsilon_{ij}} \left[V(x_i, x^t) + 2\gamma_2 a_t' \omega - \varepsilon_{ij} \right]_+ f_X(x^t)$$

where a_t' is the t-th row of A. Note that for any $u \in \mathbb{R}$, $\mathbb{E}_{\varepsilon_{ij}} [u - \varepsilon_{ij}]_+ = \int_{-\infty}^{u} (u - \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon$, soς

$$\frac{\partial}{\partial u} \mathbb{E}_{\varepsilon_{ij}} [u - \varepsilon_{ij}]_+ = \int_{-\infty}^{u} f_\varepsilon(\varepsilon) d\varepsilon = F_\varepsilon(u)$$

Hence, the first order condition of (15) with respect to $\omega = (\omega_1, \ldots, \omega_T)'$ is given by

$$\begin{aligned}
0 &= \sum_{t=1}^{T} F_\varepsilon \left(V(x_i, x^t) + 2\gamma_2 a_t' \omega \right) f_X(x^t) 2\gamma_2 a_t - 2\gamma_2 A\omega \\
&= \sum_{t=1}^{T} F_\varepsilon \left(V(x_i, x^t) + 2\gamma_2 a_t' \omega \right) f_X(x^t) a_t - A\omega
\end{aligned}$$

Let $y = A\omega$ and

$$\Psi(y) = \sum_{t=1}^{T} F_\varepsilon \left(V(x_i, x^t) + 2\gamma_2 y_t \right) f_X(x^t) a_t$$

We want to show that $\Psi(y)$ is a contraction mapping from $\mathbb{R}^T$ to $\mathbb{R}^T$, i.e., there is $\kappa \in [0, 1)$ such that $\|\Psi(y + h) - \Psi(y)\| \leq \kappa \|h\|$ for any $y, h \in \mathbb{R}^T$. By mean value theorem, $\Psi(y + h) - \Psi(y) = \nabla\Psi(\tilde{y}) \cdot y$, with $\tilde{y} = y + \xi h$ for some $\xi \in [0, 1]$, where

$$\nabla\Psi(\tilde{y}) = 2\gamma_2 \begin{bmatrix} a_{11} f_\varepsilon^1(\tilde{y}) f_X(x^1) & \cdots & a_{1T} f_\varepsilon^T(\tilde{y}) f_X(x^T) \\ \vdots & \ddots & \vdots \\ a_{T1} f_\varepsilon^1(\tilde{y}) f_X(x^1) & \cdots & a_{TT} f_\varepsilon^T(\tilde{y}) f_X(x^T) \end{bmatrix}$$

with notation $f_\varepsilon^t(\tilde{y}) = f_\varepsilon(V(x_i, x^t) + 2\gamma_2 \tilde{y}_t)$. Consider the maximum row sum matrix norm

of $\nabla\Psi\left(\tilde{y}\right)$

$$\left\Vert \nabla\Psi\left(\tilde{y}\right)\right\Vert _{\infty} = 2\gamma_{2}\max_{1\leq s\leq T}\sum_{t=1}^{T}\left|a_{st}f_{\varepsilon}^{t}\left(\tilde{y}\right)f_{X}\left(x^{t}\right)\right| \leq 2\gamma_{2}\max_{1\leq s\leq T}\sum_{t=1}^{T}f_{X}\left(x^{t}\right) = 2\gamma_{2} < 1$$

The second inequality is because $a_{st} \in [0,1]$ and $f_{\varepsilon}^{t}\left(\tilde{y}\right) \in [0,1]$ for any $1 \leq s,t \leq T$. Hence, the matrix norm $\left\Vert \nabla\Psi\left(\tilde{y}\right)\right\Vert _{2}$ induced by the Euclidean norm on $\mathbb{R}^{T}$, which equals the largest singular value of $\nabla\Psi\left(\tilde{y}\right)$, satisfies

$$\left\Vert \nabla\Psi\left(\tilde{y}\right)\right\Vert _{2} \leq \left\Vert \nabla\Psi\left(\tilde{y}\right)\right\Vert _{\infty} < 1$$

We conclude that $\Psi(y)$ is a contraction mapping of y, so by contraction mapping theorem there is a unique $y^{*} \in \mathbb{R}^{T}$ such that
$$\Psi(y^{*}) = y^{*}$$

or equivalently, there is a unique $A\omega^{*} = y^{*}$ that solves the first order condition. Furthermore, if $\omega^{*} \neq \tilde{\omega}^{*}$ satisfies $A\omega^{*} = A\tilde{\omega}^{*}$, we have $(\omega^{*} + \tilde{\omega}^{*})' A (\omega^{*} - \tilde{\omega}^{*}) = 0$, so $\omega^{*\prime}A\omega^{*} = \tilde{\omega}^{*\prime}A\tilde{\omega}^{*}$. Hence, there is a unique $A\omega^{*}$ that achieves the maximum of problem (15). ∎

Proof of Lemma 4. Let

$$\Pi_{n}\left(x_{i},\omega;p,\theta\right) = \frac{1}{n-1}\sum_{j\neq i}\left[V_{n}\left(x_{i},X_{j}\right) + \frac{n-1}{n-2}2\gamma_{2}D\left(X_{j}\right)'A\omega - \varepsilon_{ij}\right]_{+} - \gamma_{2}\omega'A\omega$$

$$\Pi\left(x_{i},\omega;p,\theta\right) = \mathbb{E}_{X_{j},\varepsilon_{ij}}\left[V\left(x_{i},X_{j}\right) + 2\gamma_{2}D\left(X_{j}\right)'A\omega - \varepsilon_{ij}\right]_{+} - \gamma_{2}\omega'A\omega$$

Note that by Proposition 1 for any (p,θ) that satisfies the assumptions there is a unique $A\omega^{*}\left(p,\theta\right)$. Because $\Pi\left(x_{i},\omega;p,\theta\right)$ is continuous in (p,θ), the unique $A\omega^{*}\left(p,\theta\right)$ must also be continuous in (p,θ) and thus bounded since Θ and $\mathcal{P} = [0,1]^{T}$ are compact. Hence we can assume without loss of generality that ω is in a compact set Ω. Moreover, the unique maximizer $A\omega^{*}\left(p,\theta\right)$ of continuous function $\Pi\left(x_{i},\omega;p,\theta\right)$ over the compact set Ω must also be well separated. If we can further show that

$$\sup_{\omega,p,\theta}\left|\Pi_{n}\left(x_{i},\omega;p,\theta\right) - \Pi\left(x_{i},\omega;p,\theta\right)\right| = o_{p}(1) \tag{22}$$

then it follows that $\sup_{p,\theta}\Pi_{n}\left(x_{i},\omega_{n};p,\theta\right) \geq \sup_{p,\theta}\Pi_{n}\left(x_{i},\omega^{*};p,\theta\right) \geq \sup_{p,\theta}\Pi\left(x_{i},\omega^{*};p,\theta\right) -$

$o_p(1)$, whence,

$$\sup_{p,\theta} \Pi\left(x_i, \omega^*; p, \theta\right) - \sup_{p,\theta} \Pi\left(x_i, \omega_n; p, \theta\right) \leq \sup_{p,\theta} \Pi_n\left(x_i, \omega_n; p, \theta\right) - \sup_{p,\theta} \Pi\left(x_i, \omega_n; p, \theta\right) + o_p\left(1\right)$$

$$\leq \sup_{\omega,p,\theta} \left|\Pi_n\left(x, \omega; p, \theta\right) - \Pi\left(x, \omega; p, \theta\right)\right| + o_p\left(1\right) = o_p\left(1\right).$$

Well-separateness of $A\omega^*\left(p, \theta\right)$ implies that for any $\varepsilon > 0$, there is $\eta > 0$ such that, for any $\left(p, \theta\right)$, $\Pi\left(x, \omega; p, \theta\right) < \Pi\left(x, \omega^*; p, \theta\right) - \eta$ for every ω with $\left\|A\omega - A\omega^*\left(p, \theta\right)\right\| \geq \varepsilon$. Therefore,

$$\Pr\left(\sup_{p,\theta} \left\|A\omega_n\left(p, \theta\right) - A\omega^*\left(p, \theta\right)\right\| \geq \varepsilon\right) \leq \Pr\left(\sup_{p,\theta}\left[\Pi\left(x, \omega_n; p, \theta\right) - \Pi\left(x, \omega^*; p, \theta\right)\right] < -\eta\right)$$

$$\leq \Pr\left(\sup_{p,\theta} \Pi\left(x, \omega_n; p, \theta\right) - \sup_{p,\theta} \Pi\left(x, \omega^*; p, \theta\right) < -\eta\right)$$

$$\rightarrow 0$$

in view of the preceding display and we are done.

Next we prove (22). Let

$$M_n\left(x_i, X_j, \varepsilon_{ij}; \omega, p, \theta\right) = V_n\left(x_i, X_j\right) + \frac{n-1}{n-2}2\gamma_2 D\left(X_j\right)' A\omega$$

$$M\left(x_i, X_j, \varepsilon_{ij}; \omega, p, \theta\right) = V\left(x_i, X_j\right) + 2\gamma_2 D\left(X_j\right)' A\omega$$

The left hand side of (22) equals

$$\sup_{\omega,p,\theta} \left|\frac{1}{n-1}\sum_{j\neq i}\left[M_n\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+ - \mathbb{E}_{X_j,\varepsilon_{ij}}\left[M\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+\right|$$

$$\leq \sup_{\omega,p,\theta}\left|\frac{1}{n-1}\sum_{j\neq i}\left[M_n\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+ - \left[M\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+\right|$$

$$+ \sup_{\omega,p,\theta}\left|\frac{1}{n-1}\sum_{j\neq i}\left[M\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+ - \mathbb{E}_{X_j,\varepsilon_{ij}}\left[M\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+\right|$$

For the second term in the last display, because

$$\left[M\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+ = \left[u(x_i, X_j) + \mathbb{E}_{X_k}p\left(X_j, X_k\right)\gamma_1 + 2\gamma_2 D\left(X_j\right)' A\omega - \varepsilon_{ij}\right]_+$$

$$\leq \left[\sup_{\omega,p,\theta}\left(u(x_i, X_j) + \mathbb{E}_{X_k}p\left(X_j, X_k\right)\gamma_1 + 2\gamma_2 D\left(X_j\right)' A\omega\right) - \varepsilon_{ij}\right]_+$$

which is absolute integrable, and $\left[M\left(x_i, X_j; \omega, p, \theta\right) - \varepsilon_{ij}\right]_+$ is continous in $\left(\omega, p, \theta\right)$, uniform

25

law of large numbers holds, i.e,

$$\sup_{\omega,p,\theta} \left| \frac{1}{n-1} \sum_{j\neq i} [M(x_i, X_j; \omega, p, \theta) - \varepsilon_{ij}]_+ - \mathbb{E}_{X_j, \varepsilon_{ij}} [M(x_i, X_j; \omega, p, \theta) - \varepsilon_{ij}]_+ \right| = o_p(1)$$

As for the first term, because $\left| [x]_+ - [y]_+ \right| \leq |x - y|$, we have

$$\sup_{\omega,p,\theta} \left| \frac{1}{n-1} \sum_{j\neq i} [M_n(x_i, X_j; \omega, p, \theta) - \varepsilon_{ij}]_+ - [M(x_i, X_j; \omega, p, \theta) - \varepsilon_{ij}]_+ \right|$$

$$\leq \sup_{\omega,p,\theta} \frac{1}{n-1} \sum_{j\neq i} |M_n(x_i, X_j; \omega, p, \theta) - M(x_i, X_j; \omega, p, \theta)|$$

$$\leq \sup_{\omega,p,\theta} \frac{1}{(n-1)(n-2)} \sum_{j\neq i} \left| \sum_{k\neq i,j} p(X_j, X_k) \gamma_1 - \mathbb{E}_{X_k}[p(X_j, X_k)] \gamma_1 \right|$$

$$+ \sup_{\omega,p,\theta} \frac{1}{(n-1)(n-2)} \sum_{j\neq i} \left| 2D(X_j)' A\omega\gamma_2 - a(X_j, X_j)\gamma_2 \right|$$

where the last inequality follows from the definition of M_n and M and triangular inequality. Uniform law of large numbers is satisfied for the second term in the last display, so

$$\sup_{\omega,p,\theta} \frac{1}{(n-1)(n-2)} \sum_{j\neq i} \left| 2D(X_j)' A\omega\gamma_2 - a(X_j, X_j)\gamma_2 \right| \leq O\left(\frac{1}{n}\right) + o_p(1) = o_p(1)$$

It suffices if

$$\sup_{\omega,p,\theta} \frac{1}{(n-1)(n-2)} \sum_{j\neq i} \left| \sum_{k\neq i,j} p(X_j, X_k) \gamma_1 - \mathbb{E}_{X_k}[p(X_j, X_k)] \gamma_1 \right| = o_p(1) \qquad (23)$$

Let

$$h(X_j, X_k; p, \theta) = p(X_j, X_k) \gamma_1 - \mathbb{E}_{X_k}[p(X_j, X_k)] \gamma_1$$

By Cauchy-Schwarz inequality,

$$\left(\frac{1}{(n-1)(n-2)} \sum_{j \neq i} \left| \sum_{k \neq i,j} h\left(X_j, X_k; p, \theta\right) \right| \right)^2$$

$$\leq \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \left(\sum_{k \neq i,j} h\left(X_j, X_k; p, \theta\right) \right)^2$$

$$= \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} h\left(X_j, X_k; p, \theta\right)^2$$

$$+ \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} h\left(X_j, X_k; p, \theta\right) h\left(X_j, X_l; p, \theta\right)$$

The two terms in the last display are U-processes. For each of them, we calculate the Hoeffding decomposition and apply the results in Sherman (1994) for degenerate U-processes. It is not difficult to show that

$$\sup_{p,\theta} \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} h\left(X_j, X_k; p, \theta\right)^2 - \mathbb{E}h\left(X_j, X_k; p, \theta\right)^2 = O_p\left(\frac{1}{\sqrt{n}}\right)$$

which implies

$$\sup_{p,\theta} \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} h\left(X_j, X_k; p, \theta\right)^2 \leq \sup_{p,\theta} \frac{1}{n-2}\mathbb{E}h\left(X_j, X_k; p, \theta\right)^2 + O_p\left(\frac{1}{n\sqrt{n}}\right) = O_p\left(\frac{1}{n}\right)$$

and

$$\sup_{p,\theta} \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} h\left(X_j, X_k; p, \theta\right) h\left(X_j, X_l; p, \theta\right) = O_p\left(\frac{1}{n}\right)$$

where the rate of convergence follows because its first-order projections (i.e., projection onto $\{X_j\}$) are identically zero. Combining the two convergence results yields

$$\sup_{p,\theta} \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \left(\sum_{k \neq i,j} h\left(X_j, X_k; p, \theta\right) \right)^2 = O_p\left(\frac{1}{n}\right)$$

which implies (23) and thus (22). The proof is complete. ∎

Denote

$$M_n\left(x_i, x_j, \omega; p, \theta\right) = V_n\left(x_i, x_j\right) + \frac{n-1}{n-2} 2\gamma_2 D\left(x_j\right)' A\omega$$

$$M\left(x_i, x_j, \omega; p, \theta\right) = V\left(x_i, x_j\right) + 2\gamma_2 D\left(x_j\right)' A\omega$$

27

From the definition of the P_n and P

$$\sup_{p,\theta} |P_n(x_i, x_j; p, \theta) - P(x_i, x_j; p, \theta)|$$

$$\leq \sup_{p,\theta} \int |1\{M_n(x_i, x_j, \omega_n(x_i; \varepsilon_i); p, \theta) \geq \varepsilon_{ij}\} - 1\{M(x_i, x_j, \omega^*(x_i); p, \theta) \geq \varepsilon_{ij}\}| \, dF_{\varepsilon_i}(\varepsilon_i)$$

$$\leq \sup_{p,\theta} \int [1\{M(x_i, x_j, \omega^*(x_i); p, \theta) < \varepsilon_{ij} \leq M_n(x_i, x_j, \omega_n(x_i; \varepsilon_i); p, \theta)\}$$

$$+ 1\{M_n(x_i, x_j, \omega_n(x_i; \varepsilon_i); p, \theta) < \varepsilon_{ij} \leq M(x_i, x_j, \omega^*(x_i); p, \theta)\}] \, dF_{\varepsilon_i}(\varepsilon_i)$$

Uniform law of large numbers applied to $V_n(x_i, x_j)$ and Lemma 4 imply that there is $c_n = o_p(1)$ independent of ε_i such that

$$\sup_{p,\theta} \left| M_n(x_i, x_j, \omega_n(x_i; \varepsilon_i); p, \theta) - M(x^s, x^t, \omega^*(x_i); p, \theta) \right| \leq c_n$$

Thus the preceding display satisfies

$$\leq \sup_{p,\theta} \int [1\{M(x_i, x_j, \omega^*(x_i); p, \theta) < \varepsilon_{ij} \leq M(x_i, x_j, \omega^*(x_i); p, \theta) + c_n\}$$

$$+ 1\{M(x_i, x_j, \omega^*(x_i); p, \theta) - c_n < \varepsilon_{ij} \leq M(x_i, x_j, \omega^*(x_i); p, \theta)\}] \, dF_{\varepsilon_{ij}}(\varepsilon_{ij})$$

$$= \sup_{p,\theta} F_{\varepsilon_{ij}}(M(x_i, x_j, \omega^*(x_i); p, \theta) + c_n) - F_{\varepsilon_{ij}}(M(x_i, x_j, \omega^*(x_i); p, \theta) - c_n)$$

$$= \sup_{p,\theta} f_{\varepsilon_{ij}}(M(x_i, x_j, \omega^*(x_i); p, \theta) + \xi c_n) 2c_n$$

$$\leq o_p(1)$$

where the second last equality follows from mean value theorem with some $\xi \in [-1, 1]$. We conclude that

$$\sup_{p,\theta} |P_n(x_i, x_j; p, \theta) - P(x_i, x_j; p, \theta)| = o_p(1).$$

Proof of Lemma 5. Note that

$$\sum_{j \neq i} \sum_{k \neq i,j} G_{ij} G_{ik} a(X_j, X_k) = \sum_{j \neq i} \sum_{k \neq i} G_{ij} G_{ik} a(X_j, X_k) - \sum_{j \neq i} G_{ij} a(X_j, X_j)$$

It suffices to show

$$a(X_j, X_k) = \sum_{t=1}^{\infty} \lambda_t \phi_t(X_j) \phi_t(X_k)$$

28

because

$$\sum_{j \neq i} \sum_{k \neq i} G_{ij} G_{ik} \sum_{t=1}^{\infty} \lambda_t \phi_t (X_j) \phi_t (X_k) = \sum_{t=1}^{\infty} \lambda_t \left(\sum_{j \neq i} G_{ij} \phi_t (X_j) \right)^2$$

Let $\delta_{X_i}(x)$ be the dirac delta function at X_i

$$\delta_{X_i}(x) = \begin{cases} \infty, & X_i = x \\ 0, & X_i \neq x \end{cases} \quad \text{and} \quad \int_{\mathcal{X}} \delta_{X_i}(x)\, dx = 1.$$

Because the eigenvectors form an orthonormal basis, function $\delta_{X_j}(x) \in L_2(\mathcal{X})$ has the representation $\delta_{X_j}(x) = \sum_{t=1}^{\infty} \langle \delta_{X_j}, \phi_t \rangle \phi_t(x)$. Then we can obtain

$$
\begin{aligned}
a(X_j, X_k) &= \int_{\mathcal{X}} \delta_{X_j}(x)\, a(x, X_k)\, dx \\
&= \int_{\mathcal{X}} \sum_{t=1}^{\infty} \langle \delta_{X_j}, \phi_t \rangle \phi_t(x)\, a(x, X_k)\, dx \\
&= \sum_{t=1}^{\infty} \phi_t(X_j) \int_{\mathcal{X}} \phi_t(x)\, a(x, X_k)\, dx \quad \text{(by dominated convergence theorem)} \\
&= \sum_{t=1}^{\infty} \phi_t(X_j)\, (A\phi_t)(X_k) \quad \text{(by definition of operator } A) \\
&= \sum_{t=1}^{\infty} \lambda_t \phi_t(X_j) \phi_t(X_k) \quad \text{(by definition of } \lambda_t \text{ and } \phi_t)
\end{aligned}
$$

■

References

[1] Aradillas-Lopez, A. (2010) Sepmiparametric estimation of a simultaneous game with incomplete infomraiton. *Journal of Econometrics*, 157, 409-431.

[2] Bajari, P. Hong, H., Krainer, J., & Nekipelov, D. (2010) Computing equilibria in static games of incomplete information using the all-solution homotopy, mimeo, University of Minnesota.

[3] Bajari, P., Hahn, J., Hong, H., & Ridder, G. (2011). A note on semiparametric estimation of finite mixtures of discrete choice models with application to game theoretic models. *International Economic Review*, 53(3), 807-824.

[4] De Paula, Richards-Shubik and Tamer (2014) Identification of Preferences in Network Formation Games, working paper.

[5] Dube, J., Fox, J. T., & C. Su (2012) Improving the numerical performance of static and dynamic aggregated discrete choice random coefficients demand estimation. *Econometrica*, 80(5), 2231-2267.

[6] Jackson, M. O., & Wolinsky, A. (1996) A strategic model of social and economic networks. *Journal of Economic Theory*, 71, 44-74.

[7] Leung, M. (2014) Two-Step Estimation of Network-Formation Models with Incomplete Information, Stanford University, working paper.

[8] Miyauchi (2013) Structural Estimation of a Pairwise Stable Network with Nonnegative Externality, working paper.

[9] Menzel, K. (2014) Inference for Games with Many Players, NYU, working paper.

[10] Menzel, K. (2015) Strategic Network Formation with Many Players, NYU, working paper.

[11] Sheng, S. (2014) A Structural Econometric Analysis of Network Formation Games, UCLA, working paper.

[12] Sherman, R. (1994) Maximal Inequalities for Degenerate U-processes with Applications to Optimization Estimators. *Annals of Statistics*, 1994(22), 439-459.

[13] Su, C., & K. L. Judd (2012) Constrained optimization approaches to estimation of structural models. *Econometrica*, 80(5), 2213-2230.