Estimation of Large Network Formation Games*

Geert Ridder† Shuyang Sheng‡

January 11, 2020

Abstract

This paper develops estimation methods for network formation models using observed data from a single large network. The model allows for utility externalities from friends of friends and friends in common, so the expected utility is nonlinear in the link choices of an agent. We propose a novel method that uses the Legendre transform to express the expected utility as a linear function of the individual link choices. This implies that the optimal link decision is that for an agent who myopically chooses to establish links or not to the other members of the network. The dependence between the agent's link choices is through an auxiliary variable. We propose a two-step estimation procedure that requires weak assumptions on equilibrium selection, is simple to compute, and has consistent and asymptotically normal estimators for the parameters. Monte Carlo results show that the estimation procedure performs well.

KEYWORDS: Network Formation, Large Games, Incomplete Information, Two-step Estimation, Legendre Transform

JEL Codes: C13, C31, C57, D85

*We thank Denis Chetverikov, Aureo de Paula, Bryan Graham, Jinyong Hahn, Bo Honore, Andres Santos, Kevin Song, Terrence Tao (for suggesting the Legendre transform) and participants of seminars at UC Berkeley, UCLA, UC Davis, UC Riverside, Princeton, U Colorado, Northwestern, Maryland, Georgetown, Duke, UBC, Ohio State, TAMU, Florida State, Penn State, U Chicago, Tinbergen Institute, Groningen U, FGV Rio, INSPER, PUC Rio, Vanderbilt, Econometric Society World Congress (2015), Econometric Society Winter Meeting and Summer Meeting (2016), INET conferences on econometrics of networks University of Cambridge (2015) and USC (2015), Berkeley/CeMMAP conference on networks Berkeley (2016), NYU conference (2017), Cowles Foundation conference (2017).

†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 growing literature on the estimation of game-theoretic models of network formation (Jackson, 2008). 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 forms links (friendship, business relation or some other type of link) with other members of the network. The preference for a link depends in general on the exogenous characteristics of the two members, and on their endogenous positions in the network, e.g., their number of links and their number of common links. It is the dependence of the link preference of an agent on the endogenous position of a potential partner in the network that complicates the analysis. The link preference of an agent also depends on unobservable features of the link. Assumptions on the nature of these unobservables play a key role in the empirical analysis.

Link formation models are discrete choice models where the choice is between alternatives that consist of the links to the other members. In a network with n members an agent chooses between 2^{n-1} overlapping sets of links. Because our analysis assumes that n grows large, this seems an intractable discrete choice problem. Our main contribution is to propose a method that for a general class of link preferences transforms this intractable discrete choice problem into a tractable sequence of related binary choice problems.

The first simplification comes from the assumption that agents have incomplete information on unobservables when making their link choices. We assume that agents know the unobserved (by the econometrician) utility shocks for their own potential links, but not the unobserved utility shocks for the potential links of the other agents. An alternative assumption is complete information under which agents know not just the unobserved utility shocks of their own potential links, but also those of the links for all other agents in the network. The complete information models are the hardest to estimate and the utility function parameters are in general partially identified (de Paula, Richards-Shubik, and Tamer, 2018; Miyauchi, 2016; Sheng, 2018). Leung (2015) considers an incomplete information model where the utility function is additively separable in one's own links. In that case, the optimal strategy is a set of independent binary link choices so that a link is established if the expected utility of the link is greater than the expected utility of not forming the link. If the utility function depends on one's own links in a nonseparable way, then the optimal strategy does not have this simple form. An important example of a non-additively separable utility function occurs if the utility of a link depends on links in common. If links in common have a positive utility then the network exhibits clustering which is a common feature of real-world networks.

We show that even if the utility function depends on the product of link choice indicators, the expected utility maximizing link choices are still equivalent to a set of (correlated) binary link choices. To obtain this equivalence we use the Legendre transform to linearize the expected utility function. This linearization introduces an auxiliary variable that depends on the unobserved utility shocks of the agent's links. This auxiliary variable is itself the solution to a (non-differentiable) optimization problem. After the inclusion of the auxiliary variable we can represent the optimal link decision as a set of binary link choices.

The parameters of the utility function can be estimated by a two-step procedure where in the first-step reduced-form link probabilities are estimated, and in the second step we estimate the utility function parameters. The asymptotic analysis of the two-step estimator has some complications. We assume that we have data on a single large network. A number of papers as Menzel (2017), Leung (2015), and de Paula, Richards-Shubik, and Tamer (2018) consider estimation using such data. In our model the link choices are dependent for each agent but not across agents. The dependence can be represented by the auxiliary variable introduced by the Legendre transform. If the number of network members n grows large the auxiliary variable converges to a constant that does not depend on the unobserved utility shocks so the link dependence vanishes. Our two-step estimator based on the observations on links is consistent even with this finite network dependence. The link dependence has to be accounted for in the asymptotic variance of the estimator.

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 optimal link choices of an agent. Section 4 discusses the two-step estimator. Section 5 introduces a number of extensions of the model and estimator. Section 6 reports the results of a simulation study.

2 Model

Consider n agents who choose to form links (or not) to each other. We introduce our model for friendships, but it applies to any kind of links or agents. The links form a network, which is represented by an $n \times n$ binary matrix $G \in \mathcal{G}$ with $\mathcal{G}$ the set of all $n \times n$ binary matrices with a 0 main diagonal. The (i, j) element $G_{ij} = 1$ if i and j are linked and 0 otherwise. The diagonal elements G_{ii} are set to 0. We consider directed links, i.e., G_{ij} and G_{ji} may be different. The case of undirected links is discussed later in Section 5.2.

Each individual i has a vector of observed characteristics $X_i \in \mathcal{X}$ and a vector of unobserved utility shocks $\varepsilon_i = (\varepsilon_{i1}, \ldots, \varepsilon_{i,i-1}, \varepsilon_{i,i+1}, \ldots, \varepsilon_{in})' \in \mathbb{R}^{n-1}$, where ε_{ij} is i's unobserved utility shock for link ij. We assume that the vector of characteristics $X = (X_1', \ldots, X_n')' \in \mathcal{X}^n$ is public information of all individuals, but the utility shock vector ε_i is the private information of i. We also assume that the utility shocks are i.i.d. and are independent of the observables.

Assumption 1 *(i) $\varepsilon_{ij}, \forall i \neq j$, are i.i.d. with cdf $F_\varepsilon(\cdot; \theta_\varepsilon)$ known up to the parameter vector $\theta_\varepsilon \in \Theta_\varepsilon \subset \mathbb{R}^{d_\varepsilon}$. The distribution is absolutely continuous with respect to the Lebesgue measure and has a density $f_\varepsilon(\cdot; \theta_\varepsilon)$ that is continuously differentiable in θ_ε and strictly positive and bounded on $\mathbb{R}$. (ii) The vector of utility shocks $\varepsilon = (\varepsilon_1', \ldots, \varepsilon_n')'$ and X are independent.*

Utility Given the vector of characteristics X and the private utility shocks ε_i, the utility of network G for i is

$$U_i(G, X, \varepsilon_i; \theta_u) = \sum_{j \neq i} G_{ij} \left(u_i(G_j, X; \beta) + \frac{1}{n-2} \sum_{k \neq i,j} G_{ik} v_i(G_j, G_k, X; \gamma) - \varepsilon_{ij} \right), \tag{2.1}$$

where $G_i = (G_{ij}, j \neq i)$ denotes the ith row of G, i.e., the links formed by i. We assume that the utility function is known up to the parameter vector $\theta_u = (\beta', \gamma')'$ in a compact set $\Theta_u \subset \mathbb{R}^{d_u}$.

In (2.1), $u_i(G_j, X; \beta)$ represents the part of the incremental utility from a link with j that does not depend on i's link decision G_i. An obvious specification is

$$u_i\left(G_j, X; \beta\right) = \beta_1 + X_i' \beta_2 + \left|X_i - X_j\right|' \beta_3 + G_{ji}\beta_4 + \frac{1}{n-2}\sum_{k \neq i,j} G_{jk}\beta_5\left(X_i, X_j, X_k\right). \tag{2.2}$$

The first four terms in (2.2) capture the direct utility from the link with j, which depends on the homophily effect (β_3) and the reciprocity effect (β_4). The last term in (2.2) captures the indirect utility from j's friends, which may vary with the characteristics of the individuals involved. This specification is similar to that in Leung (2015).

The utility function in (2.1) also accounts for the utility that i derives from simultaneously linking with j and k, denoted by $v_i\left(G_j, G_k, X; \gamma\right)$. An important example is the utility derived from friends in common

$$v_i\left(G_j, G_k, X; \gamma\right) = G_{jk}G_{kj}\gamma_1\left(X_i, X_j, X_k\right) + \frac{1}{n-3}\sum_{l \neq i,j,k} G_{jl}G_{kl}\gamma_2\left(X_i, X_j, X_k\right), \tag{2.3}$$

where the first term captures the utility of friends in common that are directly connected[1] and the second term captures the utility of friends in common that are indirectly connected. Allowing for such potential complementarities of links is crucial if we want to model networks that exhibit clustering, i.e., two individuals with friends in common are more likely to be friends (Jackson, 2008). The main difference between our model and that of Leung (2015) is that we allow for the complementarity of link decisions.

We normalize the sum terms in (2.1)-(2.3) by $n-2$ or $n-3$ to ensure that these terms remain bounded when n increases to infinity, the data scenario we consider in the asymptotic analysis.

Equilibrium Let $G_i\left(X, \varepsilon_i\right)$ denote individual i's link decision, which is a mapping from i's information $\left(X, \varepsilon_i\right)$ to a vector of links $G_i \in \mathcal{G}_i = \{0,1\}^{n-1}$. Write $G = (G_i, G_{-i})$, where G_{-i} denotes the submatrix of G with the ith row deleted, i.e., the links formed by individuals other than i.

Each individual i makes her optimal link decision by maximizing her expected utility $\mathbb{E}\left[U_i(g_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i\right]$ over $g_i \in \mathcal{G}_i$, where the expectation is taken with

[1] We can replace $G_{jk}G_{kj}$ by $G_{jk} + G_{kj}$.

respect to the link decisions of other individuals G_{-i}. Since G_{-i} is a function of X and $\varepsilon_{-i} = \left(\varepsilon_j', j \neq i\right)'$, and the private shocks ε_i are assumed to be independent across i (Assumption 1), individual i's belief about G_{-i} depends on her information (X, ε_i) only through the public information X. Let $\sigma_i\left(g_i|\, X\right) = \Pr\left(G_i\left(X, \varepsilon_i\right) = g_i|\, X\right)$ be the conditional probability that individual i chooses g_i given X. The independence of $\varepsilon_1, \ldots, \varepsilon_n$ implies that the link decisions G_i are independent across i given X, so individual i's belief about the link decisions of others is $\sigma_{-i}\left(g_{-i}|\, X\right) = \prod_{j \neq i} \sigma_j\left(g_j|\, X\right)$. Let $\sigma\left(X\right) = \left\{\sigma_i\left(g_i|\, X\right), g_i \in \mathcal{G}_i, i = 1, \ldots, n\right\}$ denote the belief profile. For a given belief profile σ, the expected utility of individual i is given by

$$
\begin{aligned}
&\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i)|\, X, \varepsilon_i, \sigma\right] \\
&= \sum_{j \neq i} G_{ij} \left(\mathbb{E}\left[u_i\left(G_j, X\right)|\, X, \sigma\right] + \frac{1}{n-2} \sum_{k \neq i,j} G_{ik} \mathbb{E}\left[v_i\left(G_j, G_k, X\right)|\, X, \sigma\right] - \varepsilon_{ij}\right).
\end{aligned}
\tag{2.4}
$$

For the specification in (2.2) and (2.3), we have

$$
\begin{aligned}
\mathbb{E}\left[u_i\left(G_j, X\right)|\, X, \sigma\right] = \beta_1 + X_i'\beta_2 + \left|X_i - X_j\right|' \beta_3 + \mathbb{E}\left[G_{ji}|\, X, \sigma\right] \beta_4 \\
+ \frac{1}{n-2} \sum_{k \neq i,j} \mathbb{E}\left[G_{jk}|\, X, \sigma\right] \beta_5\left(X_i, X_j, X_k\right),
\end{aligned}
\tag{2.5}
$$

and

$$
\begin{aligned}
\mathbb{E}\left[v_i\left(G_j, G_k, X\right)|\, X, \sigma\right] =& \mathbb{E}\left[G_{jk}|\, X, \sigma\right] \mathbb{E}\left[G_{kj}|\, X, \sigma\right] \gamma_1\left(X_i, X_j, X_k\right) \\
&+ \frac{1}{n-3} \sum_{l \neq i,j,k} \mathbb{E}\left[G_{jl}|\, X, \sigma\right] \mathbb{E}\left[G_{kl}|\, X, \sigma\right] \gamma_2\left(X_i, X_j, X_k\right),
\end{aligned}
\tag{2.6}
$$

with e.g.

$$
\mathbb{E}\left[G_{ji}|\, X, \sigma\right] = \sum_{g_j \in \mathcal{G}_j : g_{ji} = 1} \sigma_j(g_j|\, X).
$$

Given X and σ, the probability that individual i chooses g_i is

$$\begin{aligned}
&\Pr\left(G_i = g_i \middle| X, \sigma\right) \\
&= \Pr\left(\mathbb{E}\left[U_i(g_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right] \geq \max_{\tilde{g}_i \in \mathcal{G}_i} \mathbb{E}\left[U_i(\tilde{g}_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right] \middle| X, \sigma \right).
\end{aligned} \tag{2.7}$$

A Bayesian Nash equilibrium $\sigma^*(X) = \{\sigma_i^*(g_i|X), g_i \in \mathcal{G}_i . i = 1, \ldots, n\}$ is a belief profile that satisfies

$$\sigma_i^*(g_i|X) = \Pr(G_i = g_i| X, \sigma^*(X)) \tag{2.8}$$

for all link decisions $g_i \in \mathcal{G}_i$ and all $i = 1, \ldots, n$. There may be multiple belief profiles that satisfy (2.8).

3 Optimal Link Choices

The major challenge in estimating the model in Section 2 is that the expected utility of each agent i in (2.4) is nonseparable in her link choices, because the expected utility depends on $G_{ij}G_{ik}$. Solving for the optimal link choices is therefore a nonlinear integer programming problem that does not have a closed-form solution and has a problem size that grows with the number of agents. In this section, we develop an approach that overcomes this challenge and yields an expression for the link choice probability that is computationally convenient and can be used to derive asymptotic properties of parameter estimators. The idea is to find an auxiliary variable that captures the strategic interactions between i's link choices, so that after inclusion of this auxiliary variable the link choices become binary correlated choices, with the correlation between the link choices controlled by the auxiliary variable.

As a first step we observe that the expected utility from friends in common, i.e., the term $\mathbb{E}[v_i(G_j, G_k, X)| X, \sigma]$ in (2.6), is symmetric in j and k.[2] Later on we will focus on equilibria that are symmetric in individuals' observed characteristics. We say that an equilibrium $\sigma(X)$ is *symmetric* if for i and j with $X_i = X_j$, we have $\sigma_i(X) = \sigma_j(X)$, where $\sigma_i(X) = \{\sigma_i(g_i|X), g_i \in \mathcal{G}_i\}$ denotes the conditional choice probability profile of individual i.[3] In social networks, we typically do not observe

[2] An implicit assumption is that $\gamma_1(X_i, X_j, X_k)$ and $\gamma_2(X_i, X_j, X_k)$ are symmetric in X_j and X_k.

[3] Note that i and j have the same choice probabilities, but with probability 1 different expected

the identities of the agents and labels are arbitrary. It is therefore reasonable to assume that the equilibrium is symmetric because otherwise the conditional choice probabilities of an individual depend on how we label the observationally identical individuals. The symmetry of the equilibrium σ implies that agents j and k who have the same observed characteristics (i.e., $X_j = X_k$) have the same conditional choice probabilities, i.e., $\sigma_j = \sigma_k$, so $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\middle| X, \sigma\right]$ depends on j and k only through the values of X_j and X_k. Therefore $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\middle| X, \sigma\right]$ is a symmetric function of X_j and X_k.

To facilitate the presentation, we focus on the case where X_i is discrete. We assume that X_i takes a finite number of values, which are referred to as the types of the individuals.[4]

Assumption 2 (Discrete X) *X_i takes $T < \infty$ distinct values $x_1, \ldots, x_T$.*

Under Assumption 2, $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\middle| X, \sigma\right]$ takes T^2 possible values, depending on the types of j and k. For any $s, t = 1, \ldots, T$, let $V_{i,st}\left(X, \sigma\right)$ denote the value of $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\middle| X, \sigma\right]$ if j and k are of types s and t, respectively,

$$V_{i,st}\left(X, \sigma\right) = \mathbb{E}\left[v_i\left(G_j, G_k, X\right)\middle| X_j = x_s, X_k = x_t, X, \sigma\right].$$

Clearly $V_{i,st}\left(X, \sigma\right) = V_{i,ts}\left(X, \sigma\right)$. We arrange the type-specific expected utilities of friends in common in a $T \times T$ symmetric matrix

$$V_i\left(X, \sigma\right) = \begin{bmatrix} V_{i,11}\left(X, \sigma\right) & \cdots & V_{i,1T}\left(X, \sigma\right) \\ \vdots & & \vdots \\ V_{i,T1}\left(X, \sigma\right) & \cdots & V_{i,TT}\left(X, \sigma\right) \end{bmatrix}. \tag{3.1}$$

The expected utility in (2.4) can thus be represented as

$$\begin{aligned} &\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i; \theta_u)\middle| X, \varepsilon_i, \sigma\right] \\ &= \sum_{j \neq i} G_{ij}\left(U_{ij}\left(X, \sigma\right) - \varepsilon_{ij}\right) + \frac{1}{n-2} \sum_{j \neq i} \sum_{k \neq i} G_{ij} G_{ik} Z_j' V_i\left(X, \sigma\right) Z_k, \end{aligned} \tag{3.2}$$

utilities due to different random utility shocks.

[4] Potentially we can relax the assumption and allow for continuous X, but this will complicate the derivation of the optimal link choices because we need to replace the matrix notation with linear operators. For simplicity, we focus on discrete X in the paper and leave continuous X to future research.

where Z_j is a $T \times 1$ vector of binary variables that indicates the type of individual j

$$Z_j = \left(1\left\{X_j = x_1\right\}, \ldots, 1\left\{X_j = x_T\right\}\right)'$$

and

$$U_{ij}\left(X, \sigma\right) = \mathbb{E}\left[\left. u_i\left(G_j, X\right)\right| X, \sigma\right] - \frac{1}{n-2} Z_j' V_i\left(X, \sigma\right) Z_j. \tag{3.3}$$

The term $Z_j' V_i\left(X, \sigma\right) Z_k$ represents the additional expected utility that individual i receives if she links to both j and k and this additional utility depends on j and k's types.

We transform the expected utility in two steps, so that after the transformation the optimal decision can be obtained in closed form. First, since the matrix $V_i\left(X, \sigma\right)$ is real and symmetric, it has a real spectral decomposition

$$V_i\left(X, \sigma\right) = \Phi_i\left(X, \sigma\right) \Lambda_i\left(X, \sigma\right) \Phi_i\left(X, \sigma\right)', \tag{3.4}$$

where $\Lambda_i\left(X, \sigma\right) = diag\left(\lambda_{i1}\left(X, \sigma\right), \ldots, \lambda_{iT}\left(X, \sigma\right)\right)$ is the $T \times T$ diagonal matrix of eigenvalues $\lambda_{it}\left(X, \sigma\right) \in \mathbb{R}$, $t = 1, \ldots, T$, and $\Phi_i\left(X, \sigma\right) = \left(\phi_{i1}\left(X, \sigma\right), \ldots, \phi_{iT}\left(X, \sigma\right)\right)$ is the $T \times T$ orthogonal matrix of eigenvectors $\phi_{it}\left(X, \sigma\right) \in \mathbb{R}^T$, $t = 1, \ldots, T$. Using the spectral decomposition we can express the second term in the expected utility in (3.2) in a form that involves only the square of functions that are linear in the link choices, i.e., $\frac{1}{n-1}\sum_{j \neq i} G_{ij} Z_j' \phi_{it}\left(X, \sigma\right)$, $t = 1, \ldots, T$.

In the second step we "linearize" these squares of linear functions using the Legendre transform (Rockafellar, 1970). In particular, for any Y, we have the identity

$$Y^2 = \max_{\omega \in \mathbb{R}} \left\{2Y\omega - \omega^2\right\} \tag{3.5}$$

By choosing Y as the linear function $\frac{1}{n-1}\sum_{j \neq i} G_{ij} Z_j' \phi_{it}\left(X, \sigma\right)$, we can replace the square of this function by the maximization in (3.5). This maximization has an objective function that is linear in Y and therefore also linear in the link choices G_{ij}. The linearity will allow us to derive the optimal decision in closed form. The transformation of the expected utility is presented in Proposition 3.1.

Proposition 3.1 *Suppose that Assumptions 1-2 are satisfied. The expected utility in*

(3.2) is equal to

$$
\begin{aligned}
&\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\
&= \sum_{j \neq i} G_{ij}\left(U_{ij}\left(X, \sigma\right) - \varepsilon_{ij}\right) + \frac{\left(n-1\right)^2}{n-2} \sum_{t=1}^{T} \lambda_{it}\left(X, \sigma\right) \left(\frac{1}{n-1} \sum_{j \neq i} G_{ij} Z_j' \phi_{it}\left(X, \sigma\right)\right)^2 \\
&= \sum_{j \neq i} G_{ij}\left(U_{ij}\left(X, \sigma\right) - \varepsilon_{ij}\right) \\
&\quad + \frac{\left(n-1\right)^2}{n-2} \sum_{t=1}^{T} \lambda_{it}\left(X, \sigma\right) \max_{\omega_t \in \mathbb{R}} \left\{2\left(\frac{1}{n-1} \sum_{j \neq i} G_{ij} Z_j' \phi_{it}\left(X, \sigma\right)\right) \omega_t - \omega_t^2\right\}
\end{aligned} \tag{3.6}
$$

Proof. See the Supplemental Appendix. ■

To derive the optimal decision, recall that it is the link vector G_i that maximizes the expected utility. For the expected utility in (3.6), if the eigenvalues $\lambda_{it}\left(X, \sigma\right)$, $t = 1, \ldots, T$, are nonnegative (Assumption 3), we can interchange the maximization over $\omega = \left(\omega_1, \ldots, \omega_T\right)'$ and the maximization over G_i. Therefore, the optimal G_i is the solution to a simple maximization with an objective function that is linear in G_i. If we evaluate the optimal G_i at the optimal ω, we obtain the optimal link decision that maximizes the expected utility.

The solution is particularly simple if we assume the following.

Assumption 3 *Given X, for all $\theta_u \in \Theta_u$, all equilibria σ, and $i = 1, \ldots, n$, the smallest eigenvalue of the matrix $V_i\left(X, \sigma\right)$ is nonnegative.*

In Section 5.1 we show that the closed-form solution, with some modifications, remains valid if some or all of the eigenvalues are negative. Assumption 3 holds if link preferences have a large degree of homophily. If we define the type-specific link probability

$$
p_{st}\left(X, \sigma\right) = \Pr\left(G_{jk} = 1 \middle| X_j = x_s, X_k = x_t, X, \sigma\right)
$$

and assume in (2.3) that $\gamma_1\left(X_i, X_j, X_k\right) \equiv \gamma_1 > 0$, i.e., friends in common have positive utility, and $\gamma_2\left(X_i, X_j, X_k\right) \equiv 0$, then

$$
V_i\left(X, \sigma\right) = \gamma_1 \begin{bmatrix} p_{11}^2\left(X, \sigma\right) & \cdots & p_{1T}\left(X, \sigma\right) p_{T1}\left(X, \sigma\right) \\ \vdots & & \vdots \\ p_{1T}\left(X, \sigma\right) p_{T1}\left(X, \sigma\right) & \cdots & p_{T1}^2\left(X, \sigma\right) \end{bmatrix}. \tag{3.7}
$$

A sufficient condition for the eigenvalues to be nonnegative is that the matrix is diagonally dominant, i.e., for all types t

$$p_{tt}^2\left(X,\sigma\right) \geq \sum_{s\neq t} p_{st}\left(X,\sigma\right) p_{ts}\left(X,\sigma\right).$$

Our key result on the optimal link choices is as follows.

Theorem 3.2 *Suppose that Assumptions 1-3 are satisfied. For each i, the optimal decision $G_i\left(\varepsilon_i, X, \sigma\right) = \left(G_{ij}\left(\varepsilon_i, X, \sigma\right), j \neq i\right)' \in \{0,1\}^{n-1}$ is given by*

$$G_{ij}\left(\varepsilon_i, X, \sigma\right) = 1\left\{U_{ij}\left(X,\sigma\right) + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_i\left(X,\sigma\right) \Lambda_i\left(X,\sigma\right) \omega_i\left(\varepsilon_i, X, \sigma\right) - \varepsilon_{ij} \geq 0\right\}, \tag{3.8}$$

for all $j \neq i$, where the $T \times 1$ vector $\omega_i\left(\varepsilon_i, X, \sigma\right) = \left(\omega_{i1}\left(\varepsilon_i, X, \sigma\right), \ldots, \omega_{iT}\left(\varepsilon_i, X, \sigma\right)\right)'$ is a solution to the maximization problem

$$\begin{aligned} &\max_{\omega} \Pi_i\left(\omega, \varepsilon_i, X, \sigma\right) \\ =&\max_{\omega} \sum_{j\neq i} \left[U_{ij}\left(X,\sigma\right) + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_i\left(X,\sigma\right) \Lambda_i\left(X,\sigma\right) \omega - \varepsilon_{ij}\right]_+ - \frac{\left(n-1\right)^2}{n-2} \omega' \Lambda_i\left(X,\sigma\right) \omega \end{aligned} \tag{3.9}$$

with $[\cdot]_+ = \max\{\cdot, 0\}$. Set $\omega_{it}\left(\varepsilon_i, X, \sigma\right) = 0$ if $\lambda_{it}\left(X,\sigma\right) = 0$. Moreover, both $G_i\left(\varepsilon_i, X, \sigma\right)$ and $\omega_i\left(\varepsilon_i, X, \sigma\right)$ are unique almost surely.

Proof. See the Supplemental Appendix. ■

To understand the role and interpretation of $\omega_i\left(\varepsilon_i, X, \sigma\right)$ we consider the first-order condition of (3.9) derived in the Supplemental Appendix

$$\Lambda_i\left(X,\sigma\right) \omega_i\left(\varepsilon_i, X, \sigma\right) = \frac{1}{n-1} \Lambda_i\left(X,\sigma\right) \Phi_i'\left(X,\sigma\right) \sum_{k\neq i} G_{ik}\left(\varepsilon_i, X, \sigma\right) Z_k.$$

If we multiply both sides of this equation by $\frac{2(n-1)}{n-2} Z_j' \Phi_i\left(X,\sigma\right)$, we find

$$\frac{2\left(n-1\right)}{n-2} Z_j' \Phi_i\left(X,\sigma\right) \Lambda_i\left(X,\sigma\right) \omega_i\left(\varepsilon_i, X, \sigma\right) = \frac{2}{n-2} \sum_{k\neq i} G_{ik}\left(\varepsilon_i, X, \sigma\right) Z_j' V_i\left(X,\sigma\right) Z_k. \tag{3.10}$$

Note that the left-hand side is the component of the choice index in (3.8) associated with friends in common. The right-hand side is the expected marginal utility (times 2) from friends in common. To see this, note that if i contemplates a link with j, then i considers that her friends k can become friends in common with j. If j is of type s and i's friend k, a potential friend in common, is of type t, then the expected utility of i from the friend in common with j is $V_{i,st}(X,\sigma)$. Taking the sum over all friends k of i, we obtain the expected utility of friends in common if i links to j.

Corollary 3.3 suggests that the optimal decision $G_i(\varepsilon_i, X, \sigma)$ resembles the pure-stategy Nash equilibrium in an entry game (Ciliberto and Tamer, 2009; Tamer, 2003) where the entry decisions are the link choices.

Corollary 3.3 *Suppose that Assumptions 1-3 are satisfied. For each i, the optimal decision $G_i(\varepsilon_i, X, \sigma)$ satisfies*

$$G_{ij}(\varepsilon_i, X, \sigma) = 1\left\{U_{ij}(X,\sigma) + \frac{2}{n-2}\sum_{k\neq i} G_{ik}(\varepsilon_i, X, \sigma) Z_j' V_i(X,\sigma) Z_k \geq \varepsilon_{ij}\right\}, \tag{3.11}$$

for all $j \neq i$, and

$$\begin{aligned}&\sum_{j\neq i} G_{ij}(\varepsilon_i, X, \sigma)\left(U_{ij}(X,\sigma) + \frac{1}{n-2}\sum_{k\neq i} G_{ik}(\varepsilon_i, X, \sigma) Z_j' V_i(X,\sigma) Z_k - \varepsilon_{ij}\right)\\ &\geq \max_{\tilde{g}_i \text{ satisfies (3.11) a.s.}} \sum_{j\neq i} \tilde{g}_{ij}\left(U_{ij}(X,\sigma) + \frac{1}{n-2}\sum_{k\neq i} \tilde{g}_{ik} Z_j' V_i(X,\sigma) Z_k - \varepsilon_{ij}\right)\end{aligned} \tag{3.12}$$

with probability 1, where $\tilde{g}_i \in \{0,1\}^{n-1}$. Moreover, for each g_i, if we substitute $G_i(\varepsilon_i, X, \sigma) = g_i$ in (3.11) and (3.12) and define the set

$$\mathcal{E}(g_i, X, \sigma) = \left\{\varepsilon_i \in \mathbb{R}^{n-1} : g_i \text{ satisfies both (3.11) and (3.12)}\right\}. \tag{3.13}$$

Then $\mathcal{E}(g_i, X, \sigma)$, $g_i \in \{0,1\}^{n-1}$, is a partition of $\mathbb{R}^{n-1}$ with probability 1.

Proof. See the Supplemental Appendix. ■

From the corollary, the optimal decision $G_i(\varepsilon_i, X, \sigma)$ is a solution to the simultanous discrete choice model in (3.11), where an optimal link choice is determined by a random utility binary choice model and the latent utility includes the expected utility of friends in common as in (3.10). This makes the model in (3.11) and the model for an entry game (Ciliberto and Tamer, 2009; Tamer, 2003) similar if the potential entrants are the $n-1$ agents that i can link to. In this model the strategic interactions occur because the link utility depends on friends in common.

The system in (3.11) can have multiple solutions. Because i chooses links that maximize her expected utility, we have a natural equilibrium selection mechanism. That is, among the solutions to system (3.11), i chooses the $G_i(\varepsilon_i, X, \sigma)$ that gives the highest expected utility as stated in (3.12). The set $\mathcal{E}(g_i, X, \sigma)$ defined in (3.13) is the collection of $\varepsilon_i \in \mathbb{R}^{n-1}$ such that g_i is the optimal solution, i.e., $G_i(\varepsilon_i, X, \sigma) = g_i$. By Theorem 3.2 there is a unique optimal $G_i(\varepsilon_i, X, \sigma)$ that satisfies both (3.11) and (3.12) with probability 1. Therefore, the sets $\mathcal{E}(g_i, X, \sigma)$, $g_i \in \{0,1\}^{n-1}$, form a partition of the support of ε_i. These results are useful for establishing the properties of the conditional choice probabilities in Section 4.

The auxiliary variable $\omega_i(\varepsilon_i, X, \sigma)$ provides an explicit expression for the dependence of the link choices of an agent. Note that $\omega_i(\varepsilon_i, X, \sigma)$ is an optimal solution to the problem in (3.9), with an objective function that depends on ε_i, so the maximizer $\omega_i(\varepsilon_i, X, \sigma)$ is a function of ε_i. Under Assumption 1, two optimal link choices G_{ij} and G_{ik} are dependent because (i) they both depend on $\omega_i(\varepsilon_i, X, \sigma)$, as shown in (3.8), and (ii) $\omega_i(\varepsilon_i, X, \sigma)$ in G_{ij} is correlated with the utility shock ε_{ik} for G_{ik}, and symmetrically $\omega_i(\varepsilon_i, X, \sigma)$ in G_{ik} is correlated with the utility shock ε_{ij} for G_{ij}. This explicit characterization of the link dependence allows us to examine the dependence if n is large, a crucial step in the asymptotic analysis in Section 4.

If the matrix $V_i(X, \sigma)$ is singular, the $\omega_{it}(\varepsilon_i, X, \sigma)$ that correspond to the zero eigenvalues $\lambda_{it}(X, \sigma) = 0$ are indeterminate. Since it is $\Lambda_i(X, \sigma)\,\omega_i(\varepsilon_i, X, \sigma)$ that enters (3.8), the indeterminate $\omega_{it}(\varepsilon_i, X, \sigma)$ are irrelevant for the optimal decision. For that reason we can arbitrarily set $\omega_{it}(\varepsilon_i, X, \sigma) = 0$ if $\lambda_{it}(X, \sigma_i, \sigma) = 0$.

In the special case that friends in common have no effect, i.e., $\gamma_1, \gamma_2 \equiv 0$, the matrix $V_i(X, \sigma) \equiv 0$, so all the eigenvalues are equal to 0. In this case, the optimal link choice in (3.8) reduces to

$$G_{ij} = 1\left\{\mathbb{E}\left[\left. u_i(G_j, X)\right| X, \sigma\right] - \varepsilon_{ij} \geq 0\right\}, \ \forall j \neq i$$

This is the optimal link choice problem for a utility specification that is seperable in i's own links (Leung, 2015).

We focus on symmetric equilibria as discussed earlier. Applying Corollary 3.3, we can show that there exists a symmetric equilibrium. There may be multiple symmetric equilibria that satisfy (2.8). We assume that the observed equilibrium is symmetric.

Proposition 3.4 *Suppose that Assumptions 1-3 are satisfied. For any X, there exists a symmetric equilibrium $\sigma(X)$.*

Proof. See the Supplemental Appendix. ■

4 Estimation

In this section, we discuss how to estimate the structural parameter $\theta \in \mathbb{R}^{d_\theta}$. We propose a two-step procedure, where we estimate the conditional link probabilities nonparametrically in the first step, and estimate the parameter θ in the second step. When we analyze the properties of this estimator, a few complications arise. First, the model can have multiple equilibria. Second, the data are links in a single large network, where the links formed by an individual are dependent due to the preference for friends in common. We will discuss how these complications affect the estimation, and how we overcome them when we derive the properties of the estimator of θ.

Let us start with the data generating process. In this paper, we consider the scenario where we observe links from a single network, and in the asymptotic analysis we assume that the number of nodes of the network n increases to infinity.[5] To highlight the dependence of the network G on n we denote the network as G_n.

We think of the data as being generated by the following process. First, we draw a vector $X = (X_1', \dots X_n')'$ from a joint discrete distribution where X_i represents the observed characteristics of individual i. The characteristics need not be independent across individuals. Because X is ancillary, we treat X as fixed. Second, for each i we draw an $n-1$ vector of unobserved link preferences ε_i that are independent across individuals. Third, individuals form links that maximize their expected utility that depends on the equilibrium σ_n. There can be multiple equilibria, and nature selects one equilibrium σ_n among the fixed points in (2.8). We can think of σ_n as having a distribution over all the equilibria, i.e., the fixed points of (2.8). We condition on σ_n in addition to X to indicate that a particular equilibrium is selected.

[5]If we observe more than one network, we estimate the link probabilities in each network separately in the first step. In the second step we pool the links from the networks to estimate θ.

The expected utility in (2.4) and the optimal link choice in (3.8) depend on the equilibrium σ_n only through the link probabilities of each pair, so we can replace σ_n by the vector of link probabilities $p_n = (p_{n,ij}, i, j = 1, \ldots, n, i \neq j)$. The optimal link choice $G_{n,ij}(\varepsilon_i, X, \theta, p_n)$ in (3.8) implies the structural choice probability

$$P_{n,ij}(X, \theta, p_n) = \Pr\left(G_{n,ij}(\varepsilon_i, X, \theta, p_n) = 1 \,|\, X, p_n \right). \tag{4.1}$$

A Bayesian Nash equilibrium in (2.8) is that for all $i \neq j$

$$p_{n,ij} = P_{n,ij}(X, \theta, p_n). \tag{4.2}$$

Because of the symmetric equilibrium and the discrete X, each $p_{n,ij}$ depends on i and j only through their types. Therefore the link choice only depends on $p_{n,st}$, $s, t = 1, \ldots T$, the link probabilities between types. With abuse of notation we let $p_n = (p_{n,st}, s, t = 1, \ldots, T)$.

The equilibrium condition in (4.2) suggests an estimator of θ with the following two steps. In the first step, we estimate each $p_{n,st}$ by the relative frequency of links between pairs of types s and t

$$\hat{p}_{n,st} = \frac{\sum_i \sum_{j \neq i} G_{n,ij} 1\{X_i = x_s, X_j = x_t\}}{\sum_i \sum_{j \neq i} 1\{X_i = x_s, X_j = x_t\}}, \quad s, t = 1, \ldots, T. \tag{4.3}$$

In the second step, we estimate θ based on the moment condition (from now on we omit X in $P_{n,ij}(X, \theta, p_n)$)

$$\hat{\Psi}_n(\theta, p_n) = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \hat{W}_{n,ij} \left(G_{n,ij} - P_{n,ij}(\theta, p_n) \right), \tag{4.4}$$

where $\hat{W}_{n,ij} \in \mathbb{R}^{d_\theta}$, $i, j = 1, \ldots, n$, is a $d_\theta \times 1$ vector of instruments that may depend on X and p_n. The estimator $\hat{\theta}_n$ is a solution to the equation

$$\hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right) = 0. \tag{4.5}$$

The population moment function is

$$\Psi_n(\theta, p_n) = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} W_{n,ij} \left(\mathbb{E}\left[G_{n,ij} \,|\, X, p_n \right] - P_{n,ij}(\theta, p_n) \right), \tag{4.6}$$

where $W_{n,ij} \in \mathbb{R}^{d_\theta}$, $i, j = 1, \ldots, n$, is a $d_\theta \times 1$ vector of population instruments that are the probability limits of the components of $\hat{W}_{n,ij}$. Let θ_0 denote the true value of θ. Because $\mathbb{E}\left[G_{n,ij} \middle| X, p_n\right] = P_{n,ij}(\theta_0, p_n)$, θ_0 satisfies the population moment condition $\Psi_n(\theta_0, p_n) = 0$.

Our model can have multiple equilibria. Nevertheless, we do not need to specify an equilibrium selection mechanism, which in our model is the selection of a particular p_n. This is because in our two-step estimation procedure we estimate p_n in the first step and substitute the estimate in the moment condition to estimate the utility function parameters. Because we consider a single network instead of multiple networks, we do not need an assumption on equilibrium selection across networks either.[6]

Under the additional Assumption 4, $\hat{\theta}_n$ is a consistent estimator of θ_0.

Assumption 4 *(i) The parameter θ lies in a compact set $\Theta \subseteq \mathbb{R}^{d_\theta}$. (ii) For an equilibrium p_n and for all n, the system of equations $\Psi_n(\theta, p_n) = 0$ has a unique solution θ_0. (iii) The instruments $\hat{W}_{n,ij}$ and their population counterparts $W_{n,ij}$, $i, j = 1, \ldots, n$, satisfy $\max_{i,j=1,\ldots,n} \left\| \hat{W}_{n,ij} - W_{n,ij} \right\| \xrightarrow{p} 0$ and $\max_{i,j=1,\ldots,n} \left\| W_{n,ij} \right\| < \infty$. (iv) For $s, t = 1, \ldots, T$, $\lim_{n\to\infty} \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} 1\left\{X_i = x_s, X_j = x_t\right\}$ exists and is strictly positive.*

Assumption 4(iv) imposes a mild restriction that the fraction of pairs of all types s and t is positive as $n \to \infty$, so that the number of pairs of all types grows without bounds, and we can estimate the link probabilities $p_{n,st}$ consistently. This assumption is satisfied if X_i, $i = 1, \ldots, n$, are i.i.d. or have limited dependence.

Assumption 4(ii) is a local identification condition, local because it is an identification condition for θ for a given equilibrium p_n.[7] It requires that the solution θ_0 is invariant to the equilibrium p_n selected for all n.

Theorem 4.1 (Consistency) *If Assumptions 1-4 are satisfied, then*

$$\hat{\theta}_n - \theta_0 \xrightarrow{p} 0.$$

Proof. See the Supplemental Appendix. ■

[6] We need to assume that the equilibrium selection does not depend on the random utility shocks ε.

[7] Because of multiple equilibria we do not assume the global identification condition that $\Psi_n(\theta, p) = 0$ has a unique solution (θ_0, p_n).

We first show that $\hat{p}_n$ is consistent for p_n and next establish the consistency of $\hat{\theta}_n$. In addition to the identification condition, we need a uniform LLN for the sample moment $\hat{\Psi}_n(\theta, p_n)$. Note that links formed by different individuals in a single network are independent given X and p_n. This is crucial for a LLN to hold. Moreover, because link choices depend on the number of agents in the network, the distribution of the data on the links in a network with n nodes depends on n. Therefore, we prove the uniform LLN for a triangular array (Pollard, 1990).

Next we examine the asymptotic distribution of $\hat{\theta}_n$. The complication is that the links formed by an individual are correlated. By Theorem 3.2, the link choices $G_{n,ij}$ and $G_{n,ik}$ of individual i are correlated because they both depend on the auxiliary variable $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$ that maximizes the objective function

$$\begin{aligned}\Pi_{ni}(\omega, \varepsilon_i, \theta_0, p_n) = &\sum_{j \neq i} \left[U_{n,ij}(\theta_0, p_n) + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni}(\theta_0, p_n) \Lambda_{ni}(\theta_0, p_n) \omega - \varepsilon_{ij} \right]_+ \\ &- \frac{(n-1)^2}{n-2} \omega' \Lambda_{ni}(\theta_0, p_n) \omega, \end{aligned} \tag{4.7}$$

where we add subscript n to Π_i, U_{ij}, Λ_i, and Φ_i to indicate their dependence on n. To derive the asymptotic distribution of $\hat{\theta}_n$, we first derive the asymptotic properties of $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$, and then investigate how $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$ affects the asymptotic distribution of $\hat{\theta}_n$.

In particular, let $\Pi^*_{ni}(\omega, \theta_0, p_n)$ be the conditional expectation of $\Pi_{ni}(\omega, \varepsilon_i, \theta_0, p_n)$ given X and p_n

$$\begin{aligned}&\Pi^*_{ni}(\omega, \theta_0, p_n) \\ =&\sum_{j \neq i} \mathbb{E}\left[\left[U_{n,ij}(\theta_0, p_n) + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni}(\theta_0, p_n) \Lambda_{ni}(\theta_0, p_n) \omega - \varepsilon_{ij} \right]_+ \middle| X, p_n \right] \\ &- \frac{(n-1)^2}{n-2} \omega' \Lambda_{ni}(\theta_0, p_n) \omega. \end{aligned} \tag{4.8}$$

We make the following assumptions on the auxiliary variable.

Assumption 5 *(i) The auxiliary variable ω is in a compact set $\Omega \subseteq \mathbb{R}^T$, which contains a compact neighborhood of* 0*. (ii) The function $\Pi^*_{ni}(\omega, \theta_0, p_n)$ has a unique maximizer $\omega^*_{ni}(\theta_0, p_n)$. (iii) The gradient $\Gamma^*_{ni}(\omega, \theta_0, p_n)$ of $\Pi^*_{ni}(\omega, \theta_0, p_n)$ has the Ja-*

cobian matrix[8]

$$
\begin{aligned}
&\nabla_{\omega'}\Gamma^*_{ni}\left(\omega,\theta_0,p_n\right) \\
&=\left(\frac{2}{n-2}\sum_{j\neq i} f_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega\right)\Lambda_{ni}\Phi_{ni}'Z_jZ_j'\Phi_{ni}-I_T\right)\Lambda_{ni}
\end{aligned}
$$

with I_T being the $T\times T$ identity matrix, where the $T\times T$ matrix in parentheses

$$
\frac{2}{n-2}\sum_{j\neq i} f_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega\right)\Lambda_{ni}\Phi_{ni}'Z_jZ_j'\Phi_{ni}-I_T
$$

is nonsingular at $\omega^*_{ni}\left(\theta_0,p_n\right)$.

Under Assumption 5, we show in the Supplemental Appendix that $\omega_{ni}\left(\varepsilon_i,\theta_0,p_n\right)$ is consistent for $\omega^*_{ni}\left(\theta_0,p_n\right)$ and has an asymptotically linear representation

$$
\omega_{ni}\left(\varepsilon_i,\theta_0,p_n\right)-\omega^*_{ni}\left(\theta_0,p_n\right)=\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}\left(\omega^*_{ni}\left(\theta_0,p_n\right),\varepsilon_{ij},\theta_0,p_n\right)+o_p\left(\frac{1}{\sqrt{n}}\right). \tag{4.9}
$$

In this representation $\varphi^{\omega}_{n,ij}\left(\omega,\varepsilon_{ij},\theta_0,p_n\right)\in\mathbb{R}^T$ is the influence function

$$
\varphi^{\omega}_{n,ij}\left(\omega,\varepsilon_{ij},\theta_0,p_n\right)=\nabla_{\omega'}\Gamma^*_{ni}\left(\omega,\theta_0,p_n\right)^{+}\varphi^{\pi}_{n,ij}\left(\omega,\varepsilon_{ij},\theta_0,p_n\right)
$$

where $\varphi^{\pi}_{n,ij}\left(\omega,\varepsilon_{ij},\theta_0,p_n\right)\in\mathbb{R}^T$ is defined by

$$
\varphi^{\pi}_{n,ij}\left(\omega,\varepsilon_{ij},\theta_0,p_n\right)=1\left\{U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega-\varepsilon_{ij}\geq 0\right\}\Lambda_{ni}\Phi_{ni}'Z_j-\Lambda_{ni}\omega.
$$

Since $\omega^*_i\left(\theta_0,p_n\right)$ is deterministic, the convergence of $\omega_{ni}\left(\varepsilon_i,\theta_0,p_n\right)$ to $\omega^*_{ni}\left(\theta_0,p_n\right)$ indicates that the correlation between links $G_{n,ij}$ and $G_{n,ik}$ vanishes as n approaches infinity. Moreover, the asymptotically linear representation in (4.9) implies that $\omega_{ni}\left(\varepsilon_i,\theta_0,p_n\right)$ converges to $\omega^*_{ni}\left(\theta_0,p_n\right)$ at the rate of $n^{-1/2}$. This rate is crucial in deriving the asymptotic distribution of the estimator.

Under the additional conditions in Assumption 6, we derive the asymptotic distribution of $\hat{\theta}_n$ in Theorem 4.2.

[8]We suppress the dependence of $U_{n,ij}$, Λ_{ni}, and Φ_{ni} on θ_0 and p_n hereafter.

Assumption 6 *(i) For any $i, j = 1, \ldots, n$, $P_{n,ij}(\theta, p)$ is continuously differentiable with respect to θ and p in a neighborhood of (θ_0, p_n). (ii) The $d_\theta \times d_\theta$ Jacobian matrix with respect to θ*

$$J_n^\theta(\theta_0, p_n) = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} W_{n,ij} \nabla_{\theta'} P_{n,ij}(\theta_0, p_n)$$

is nonsingular.

In the Supplemental Appendix (see Lemma S.4) we show that $P_{n,ij}(\theta, p)$ is continuous in θ (and p). Continuity follows from Corollary 3.3 that shows that there is a 1-1 relationship between the optimal link decision g_i and set $\mathcal{E}(g_i)$ that partition $\mathbb{R}^{n-1}$. The boundaries of the partition sets are continuous, but can have kinks if the set of inequalities in (3.12) that are binding depends on θ. We assume that there is a possibly small neighborhood of θ_0 without kinks, so that the choice probability is continuously differentiable in that neighborhood. Because we already established consistency, the estimator is in that neighborhood with probability approaching 1.

Theorem 4.2 (Asymptotic Distribution) *Suppose that Assumptions 1-6 are satisfied. Define the $d_\theta \times d_\theta$ matrix*

$$\begin{aligned}
&\Sigma_n(\theta_0, p_n) \\
&= \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} J_n^\theta(\theta_0, p_n)^{-1} \mathbb{E}\left[\varphi_{n,ij}^m(\varepsilon_{ij}, \theta_0, p_n) \varphi_{n,ij}^m(\varepsilon_{ij}, \theta_0, p_n)' \middle| X, p_n\right] \left(J_n^\theta(\theta_0, p_n)^{-1}\right)'
\end{aligned} \tag{4.10}$$

with $\varphi_{n,ij}^m(\varepsilon_{ij}, \theta_0, p_n) \in \mathbb{R}^{d_\theta}$ given by

$$\begin{aligned}
\varphi_{n,ij}^m(\varepsilon_{ij}, \theta_0, p_n) = &\tilde{W}_{n,ij}(\theta_0, p_n) \left(g_{n,ij}(\omega_{ni}^*(\theta_0, p_n), \varepsilon_{ij}, \theta_0, p_n) - P_{n,ij}^*(\omega_{ni}^*(\theta_0, p_n), \theta_0, p_n)\right) \\
&+ \tilde{J}_{ni}^\omega(\omega_{ni}^*(\theta_0, p_n), \theta_0, p_n) \varphi_{n,ij}^\omega(\omega_{ni}^*(\theta_0, p_n), \varepsilon_{ij}, \theta_0, p_n)
\end{aligned} \tag{4.11}$$

for all $i \neq j$, where $\omega_{ni}^*(\theta_0, p_n)$ is the maximizer of $\Pi_{ni}^*(\omega, \theta_0, p_n)$. In this expression $\tilde{W}_{n,ij}(\theta_0, p_n) \in \mathbb{R}^{d_\theta}$ is a $d_\theta \times 1$ vector of augmented instruments that include the contribution of the first-step estimates

$$\tilde{W}_{n,ij}(\theta_0, p_n) = W_{n,ij} - \left(\frac{1}{n(n-1)} \sum_k \sum_{l \neq k} W_{n,kl} \nabla_{p'} P_{n,kl}(\theta_0, p_n)\right) Q_{n,ij} \tag{4.12}$$

with $Q_{n,ij} = [Q_{n,ij,11}, \ldots, Q_{n,ij,1T}, \ldots, Q_{n,ij,T1}, \ldots, Q_{n,ij,TT}]' \in \mathbb{R}^{T^2}$ *and*

$$Q_{n,ij,st} = \frac{1\{X_i = x_s, X_j = x_t\}}{\frac{1}{n(n-1)}\sum_i \sum_{j\neq i} 1\{X_i = x_s, X_j = x_t\}}, \quad s,t = 1,\ldots,T.$$

Further $g_{n,ij}(\omega, \varepsilon_{ij}, \theta_0, p_n)$ *is the link choice indicator for a given* ω

$$g_{n,ij}(\omega, \varepsilon_{ij}, \theta_0, p_n) = 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega \geq \varepsilon_{ij}\right\},$$

and $P^*_{n,ij}(\omega, \theta_0, p_n)$ *is the probability of a link given* ω

$$P^*_{n,ij}(\omega, \theta_0, p_n) = F_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega\right).$$

Also $\tilde{J}^\omega_{ni}(\omega, \theta_0, p_n)$ *is the* $d_\theta \times T$ *Jacobian matrix of the moment function with link probabilities* $P^*_{n,ij}(\omega, \theta_0, p_n)$ *with respect to* ω

$$\tilde{J}^\omega_{ni}(\omega, \theta_0, p_n) = \frac{1}{n-1}\sum_{j\neq i} \tilde{W}_{n,ij}(\theta_0, p_n) \nabla_{\omega'} P^*_{n,ij}(\omega, \theta_0, p_n),$$

and finally $\varphi^\omega_{n,ij}(\omega, \varepsilon_{ij}, \theta_0, p_n) \in \mathbb{R}^T$ *is the influence function given in (4.9). Then*

$$\sqrt{n(n-1)}\Sigma_n^{-1/2}(\theta_0, p_n)\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0, I_{d_\theta})$$

as $n \to \infty$, *where* I_{d_θ} *is the* $d_\theta \times d_\theta$ *identity matrix.*

Proof. See the Supplemental Appendix. ■

The influence function $\varphi^m_{n,ij}$ has two components where the first captures the variability in the link choices. Note that $\varphi^m_{n,ij}$, $i,j = 1,\ldots,n$, is independent over j. The dependence between the link choices is through the auxiliary variable $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$ and this contributes to the second component of $\varphi^m_{n,ij}$.

In applications, we need to choose the instrument $\hat{W}_{n,ij}$. One option is to use the instrument from the quasi maximum likelihood estimation (QMLE). Let $\mathcal{L}_n(\theta, \hat{p}_n)$ be the single-link log likelihood function evaluated at the first-step estimate $\hat{p}_n$

$$\mathcal{L}_n(\theta, \hat{p}_n) = \sum_i \sum_{j\neq i} G_{n,ij} \ln P_{n,ij}(\theta, \hat{p}_n) + (1 - G_{n,ij}) \ln(1 - P_{n,ij}(\theta, \hat{p}_n)). \quad (4.13)$$

This is not the full-information likelihood, which requires a specification of the equilibrium selection mechanism. Because link choices are correlated there is also information on θ in the joint distribution of pairs of the link choices $G_{n,ij}$ and $G_{n,ik}$ (see also Section 5.2) that is not captured in $\mathcal{L}_n(\theta, \hat{p}_n)$.

Taking the derivative with respect to θ we obtain the quasi-likelihood equation

$$\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \frac{\nabla_\theta P_{n,ij}(\theta, \hat{p}_n)}{P_{n,ij}(\theta, \hat{p}_n)(1 - P_{n,ij}(\theta, \hat{p}_n))} (G_{n,ij} - P_{n,ij}(\theta, \hat{p}_n)) = 0.$$

Comparing this with the moment in (4.4) the instrument is

$$\hat{W}_{n,ij}(\theta) = \frac{\nabla_\theta P_{n,ij}(\theta, \hat{p}_n)}{P_{n,ij}(\theta, \hat{p}_n)(1 - P_{n,ij}(\theta, \hat{p}_n))}, \ i, j = 1, \ldots, n, \ i \neq j. \tag{4.14}$$

This instrument depends on θ. Therefore, we need a preliminary estimator of θ. For that purpose we can use powers and interactions of X_i and X_j as the d_θ instruments in the moment condition. A preliminary instrument is not needed if we use continuous updating as in Hansen, Heaton, and Yaron (1996).

In the discussion of Assumption 6(i) we noted that $P_{n,ij}(\theta, p)$ is not differentiable for a finite number of values of θ. That may be a problem for the instrument, because it has to be evaluated outside a small neighborhood of θ_0 during the search for the solution of the moment equation. We can avoid this problem, if we use the limiting choice probability derived in Section 5.3 in the instrument instead of the finite n probability. The limiting choice probability is differentiable in θ everywhere. We will come back to this after we discuss the limiting game.

The link choice probability $P_{n,ij}(\theta, p)$ is an $n-1$ dimensional integral without a closed form. It can be computed by simulation. We draw ε_i independently R times (and use these simulated utility shocks throughout the search for a solution of the moment equation), and for each simulated $\varepsilon_{i,r}$, $r = 1, \ldots, R$, we compute the auxiliary variable $\omega_{ni}(\varepsilon_{i,r}, \theta, \hat{p}_n)$ and the vector of link choices $G_{ni}(\varepsilon_{i,r}, \theta, \hat{p}_n)$ in (3.8). The vector of simulated link choice probabilities is the sample average of $G_{n,ij}(\varepsilon_{i,r}, \theta, \hat{p}_n)$, $r = 1, \ldots, R$. This simulation procedure does not affect the consistency, rate of convergence, and asymptotic normality of the estimator. The simulated GMM has an asymptotic variance equal to that in Theorem 4.2 multiplied by $1 + R^{-1}$ (Pakes and Pollard (1989)).

5 Extensions

5.1 General V_i

Assumption 3 that the matrix $V_i(X,\sigma)$ is positive semi-definite is not crucial to our approach. Without this assumption, the auxiliary variable is solved from a maximin problem.

Theorem 5.1 *Suppose that Assumptions 1-2 are satisfied. The optimal link decision* $G_i(\varepsilon_i, X, \sigma) = (G_{ij}(\varepsilon_i, X, \sigma), j \neq i)'$ *is given by*

$$G_{ij}(\varepsilon_i, X, \sigma) = 1\left\{U_{ij}(X,\sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_i(X,\sigma) \Lambda_i(X,\sigma) \omega_i(\varepsilon_i, X, \sigma) - \varepsilon_{ij} \geq 0\right\}, \tag{5.1}$$

almost surely, $\forall j \neq i$, *where the* $T \times 1$ *vector* $\omega_i(\varepsilon_i, X, \sigma)$ *is a solution to the maximin problem*

$$\begin{aligned}
&\max_{(\omega_t, t\in\mathcal{T}_{i+})} \min_{(\omega_t, t\in\mathcal{T}_{i-})} \Pi_i(\omega, \varepsilon_i, X, \sigma) \\
&= \max_{(\omega_t, t\in\mathcal{T}_{i+})} \min_{(\omega_t, t\in\mathcal{T}_{i-})} \sum_{j\neq i} \left[U_{ij}(X,\sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_i(X,\sigma) \Lambda_i(X,\sigma) \omega - \varepsilon_{ij}\right]_+ \\
&\qquad - \frac{(n-1)^2}{n-2} \omega' \Lambda_i(X,\sigma) \omega
\end{aligned} \tag{5.2}$$

with $\mathcal{T}_{i+} = \{t : \lambda_{it}(X,\sigma) > 0\}$ *and* $\mathcal{T}_{i-} = \{t : \lambda_{it}(X,\sigma) < 0\}$. *We set* $\omega_{it}(\varepsilon_i, X, \sigma) = 0$ *if* $\lambda_{it}(\varepsilon_i, X, \sigma) = 0$. *Moreover, both* $G_i(\varepsilon_i, X, \sigma)$ *and* $\omega_i(\varepsilon_i, X, \sigma)$ *are unique almost surely.*

Proof. See the Supplemental Appendix. ■

Note that the expected utility in (3.6) is separable in the maximizations over ω_t, $t = 1, \ldots, T$, so that a maximization over ω_t becomes a minimization if $\lambda_{it}(X,\sigma) < 0$. If $\lambda_{it}(X,\sigma) = 0$, the objective function does not depend on ω_t and we set $\omega_{it}(\varepsilon_i, X, \sigma) = 0$. The separability also implies that, unlike in a general maximin or minimax problem, the order of the maximizations and minimizations does not matter.

To gain some intuition regarding the role of the eigenvalues of $V_i(X,\sigma)$ we consider the case with 2 types ($T = 2$) and a utility specification as in (2.1)-(2.3) with γ_1 a positive constant and $\gamma_2 = 0$. We omit the arguments X and σ. The matrix V_i is as in (3.7) and has nonnegative components. Suppose $V_{i,12} > 0$.

Let λ_{i1} and λ_{i2} be the eigenvalues of V_i with $\lambda_{i1} \geq \lambda_{i2}$. We can see that $\lambda_{i1} > 0$.[9] Assume that $\lambda_{i2} \neq 0$, i.e., $V_{i,11}V_{i,22} \neq V_{i,12}^2$. Let $\phi_{i1} = (\phi_{i,11}, \phi_{i,12})'$ and $\phi_{i2} = (\phi_{i,21}, \phi_{i,22})'$ be the corresponding eigenvectors. It can be shown that the elements of ϕ_{i1} have the same sign, and the elements of ϕ_{i2} have opposite signs, i.e., $\phi_{i,11}\phi_{i,12} > 0$ and $\phi_{i,21}\phi_{i,22} < 0$.[10] Without loss of generality we take $\phi_{i,11}, \phi_{i,12} > 0$, and $\phi_{i,21} > 0$, $\phi_{i,22} < 0$.

From the first-order condition of the problem in (5.2), ω_i satisfies

$$\omega_{it} = \phi_{it}' \frac{1}{n-1} \sum_{j \neq i} G_{ij} Z_j, \text{ a.s., } t = 1, 2. \tag{5.3}$$

Note that $\sum_{j \neq i} G_{ij} Z_j$ is the 2×1 vector of the number of friends that i has of each type. Therefore, ω_{it} is a weighted sum of the numbers of friends of each type, with weights equal to the components of the eigenvector ϕ_{it}. By $V_i = \Phi_i \Lambda_i \Phi_i'$ and the first-order condition in (5.3), the expected utility in (3.2) evaluated at the optimal ω_i can be expressed as

$$\begin{aligned}
\mathbb{E}\left[U_i \middle| X, \varepsilon_i, \sigma\right] = &\sum_{j \neq i} G_{ij}\left(U_{ij} - \varepsilon_{ij}\right) \\
&+ \frac{(n-1)^2}{n-2}\left(\Phi_i' \frac{1}{n-1} \sum_{j \neq i} G_{ij} Z_j\right)' \Lambda_i \left(\Phi_i' \frac{1}{n-1} \sum_{k \neq i} G_{ik} Z_k\right) \\
= &\sum_{j \neq i} G_{ij}\left(U_{ij} - \varepsilon_{ij}\right) + \frac{(n-1)^2}{n-2} \omega_i' \Lambda_i \omega_i, \quad \text{a.s.}
\end{aligned}$$

Because $\lambda_{i1} > 0$, individual i prefers a larger ω_{i1}, which is a preference for many friends, with friends of the type that corresponds to the larger of $\phi_{i,11}$ and $\phi_{i,12}$ being preferred. If $\lambda_{i2} > 0$, i prefers a larger ω_{i2} (in absolute value), i.e., she prefers her circle of friends to be of one type. If $\lambda_{i2} < 0$, i prefers an ω_{i2} closer to 0, which is a preference for an integrated circle of friends.

---

[9] The eigenvalues are given by $\lambda_{i1}, \lambda_{i2} = \frac{1}{2}\left(V_{i,11} + V_{i,22} \pm \sqrt{V_{i,11}^2 + V_{i,22}^2 + 4V_{i,12}^2 - 2V_{i,11}V_{i,22}}\right)$. Since $V_{i,12} > 0$, they satisfy $\lambda_{i1} > \max\{V_{i,11}, V_{i,22}\} \geq 0$ and $\lambda_{i2} < \min\{V_{i,11}, V_{i,22}\}$.

[10] By definition $V_i \phi_{i1} = \lambda_{i1} \phi_{i1}$, so $(\lambda_{i1} - V_{i,11})\, \phi_{i,11} = V_{i,12} \phi_{i,12}$ and $V_{i,12} \phi_{i,11} = (\lambda_{i1} - V_{i,22})\, \phi_{i,12}$. Since $\lambda_{i1} > \max\{V_{i,11}, V_{i,22}\}$ and $V_{i,12} > 0$, these equations imply that $\phi_{i,11}$ and $\phi_{i,12}$ must have the same sign, i.e., $\phi_{i,11}\phi_{i,12} > 0$. Similarly, we can show $\phi_{i,21}\phi_{i,22} < 0$.

In the special case that $V_{i,11} = V_{i,22} = 0$, i.e., only agents of the opposite type link, the two eigenvalues are $\lambda_{i1} = V_{i,12}$ and $\lambda_{i2} = -V_{i,12}$, and the corresponding eigenvectors are $\phi_{i1} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)'$ and $\phi_{i2} = \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)'$. In this case,

$$\omega_{i1} = \frac{1}{\sqrt{2}\,(n-1)} \sum_{j \neq i} G_{ij} \left(Z_{j1} + Z_{j2}\right)$$

$$\omega_{i2} = \frac{1}{\sqrt{2}\,(n-1)} \sum_{j \neq i} G_{ij} \left(Z_{j1} - Z_{j2}\right).$$

Intuitively, if a network only allows for cross-type links, an agent has the most friends in common if she makes as many friends as she can (i.e., prefers ω_{i1} to be large) and chooses an equal number of friends of each type (i.e., prefers ω_{i2} to be close to 0).

5.2 Undirected Networks

In this section we show that our method that was derived for directed networks also can be applied to undirected networks. Let G_{ij} now denote an undirected link between i and j and G the adjacency matrix of the undirected network. In an undirected network $G_{ij} = G_{ji}$. To model the formation of undirected links, we follow the link-announcement framework (Jackson, 2008) and require mutual consent to form a link. Specifically, let S_{ij} indicate whether i proposes to link to j. A link is formed if both i and j propose to form it, so $G_{ij} = S_{ij} S_{ji}$. Our approach in Section 3 can be extended to undirected networks if we work with the proposals instead of the links. Because we observe the links but not the proposals, the estimation of the parameters is less straightforward. In this section we show that the extension is possible, but we leave the development of the extension to future research.

We consider the utility specification in (2.1), with G_{ij} an undirected link. In (2.2) we omit the reciprocity effect and in (2.3) k is a mutual friend of i and j if j and k have an undirected link, so that

$$u_i\left(G_j, X; \beta\right) = \beta_1 + X_i' \beta_2 + \left|X_i - X_j\right|' \beta_3 + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk} \beta_4 \left(X_i, X_j, X_k\right)$$

and

$$v_i\left(G_j, G_k, X; \gamma\right) = G_{jk}\gamma_1\left(X_i, X_j, X_k\right) + \frac{1}{n-3}\sum_{l \neq i,j,k} G_{jl}G_{kl}\gamma_2\left(X_i, X_j, X_k\right).$$

Since $G_{ij} = S_{ij}S_{ji}$, if S is the $n \times n$ matrix of proposed links, then G is a function of S, and we have $G = G(S) = G(S_i, S_{-i})$, with S_i the vector of link proposals of i and S_{-i} the matrix of link proposals of the other agents. We maintain the assumption that ε_i is private information of agent i, so each agent i forms a belief about the proposals of the other agents, S_{-i}, when choosing S_i. Given X, let $\sigma_i\left(\left. s_i \right| X\right)$ be the conditional probability that agent i proposes s_i given X and let $\sigma(X) = \left\{\sigma_i\left(\left. s_i\right| X\right), s_i \in \{0,1\}^{n-1}, i = 1, \ldots, n\right\}$ be the belief profile. For a belief profile σ, the expected utility of agent i is given by

$$\begin{aligned}\mathbb{E}\left[\left. U_i(G\left(S_i, S_{-i}\right), X, \varepsilon_i)\right| X, \varepsilon_i, \sigma\right] = \sum_{j \neq i} S_{ij} \Bigg(& \mathbb{E}\left[\left. S_{ji}u_i\left(G_j, X\right)\right| X, \sigma\right] - \mathbb{E}\left[\left. S_{ji}\right| X, \sigma\right]\varepsilon_{ij} \\ & + \frac{1}{(n-2)}\sum_{k \neq i,j} S_{ik}\mathbb{E}\left[\left. S_{ji}S_{ki}v_i\left(G_j, G_k, X\right)\right| X, \sigma\right]\Bigg),\end{aligned} \tag{5.4}$$

where

$$\begin{aligned}\mathbb{E}\left[\left. S_{ji}u_i\left(G_j, X\right)\right| X, \sigma\right] =& \mathbb{E}\left[\left. S_{ji}\right| X, \sigma\right]\left(\beta_1 + X_i'\beta_2 + \left|X_i - X_j\right|'\beta_3\right) \\ & + \frac{1}{n-2}\sum_{k \neq i,j}\mathbb{E}\left[\left. S_{ji}S_{jk}\right| X, \sigma\right]\mathbb{E}\left[\left. S_{kj}\right| X, \sigma\right]\beta_4\left(X_i, X_j, X_k\right)\end{aligned}$$

and

$$\begin{aligned}& \mathbb{E}\left[\left. S_{ji}S_{ki}v_i\left(G_j, G_k, X\right)\right| X, \sigma\right] \\ =& \mathbb{E}\left[\left. S_{ji}S_{jk}\right| X, \sigma\right]\mathbb{E}\left[\left. S_{ki}S_{kj}\right| X, \sigma\right]\gamma_1\left(X_i, X_j, X_k\right) \\ & + \frac{1}{n-3}\sum_{l \neq i,j,k}\mathbb{E}\left[\left. S_{ji}S_{jl}\right| X, \sigma\right]\mathbb{E}\left[\left. S_{ki}S_{kl}\right| X, \sigma\right]\mathbb{E}\left[\left. S_{lj}S_{lk}\right| X, \sigma\right]\gamma_2\left(X_i, X_j, X_k\right).\end{aligned}$$

In the derivation we have used that $G_{ij} = S_{ij}S_{ji}$ and S_i and S_j are independent given X and σ. Note that the expected utility depends on the probability that an agent proposes a link to two other agents simultaneously, as in $\mathbb{E}\left[\left. S_{ji}S_{jk}\right| X, \sigma\right]$, with S_{ji} and

S_{jk} dependent.

Just as in Sections 3 and 5.1, we linearize the quadratic part of the expected utility in (5.4) using the Legendre transform. The linearized (in S_i) expected utility gives the optimal link proposals in closed form. We maintain Assumption 2, so that X takes T values. For $s, t = 1, \ldots, T$, we define $V^u_{i,st}(X, \sigma)$ as

$$V^u_{i,st}(X, \sigma) = \mathbb{E}\left[S_{ji} S_{ki} v_i (G_j, G_k, X) \middle| X_j = x_s, X_k = x_t, X, \sigma \right].$$

$V^u_{i,st}(X, \sigma)$ is the expected utility of friends in common if i proposes to link to both j (of type s) and k (of type t). The superscript u indicates an undirected network. Note that because the expected value is symmetric in j and k, $V^u_{i,st}(X, \sigma)$ is symmetric in s and t. Let $V^u_i(X, \sigma)$ denote the symmetric $T \times T$ matrix with components $V^u_{i,st}(X, \sigma)$, $s, t = 1, \ldots, T$. Let $\lambda^u_{it}(X, \sigma)$, $t = 1, \ldots, T$, be the eigenvalues of the matrix $V^u_i(X, \sigma)$ and $\phi^u_{it}(X, \sigma)$, $t = 1, \ldots, T$, the corresponding eigenvectors. Further, $\Lambda^u_i(X, \sigma) = diag(\lambda^u_{i1}(X, \sigma), \ldots, \lambda^u_{iT}(X, \sigma))$ and $\Phi^u_i(X, \sigma)$ the matrix of eigenvectors.

The optimal proposal decision can be represented as a set of related binary choices.

Corollary 5.2 *Suppose that Assumptions 1-2 are satisfied. The optimal proposal decision $S_i(\varepsilon_i, X, \sigma) = (S_{ij}(\varepsilon_i, X, \sigma), j \neq i)'$ is given by*

$$S_{ij}(\varepsilon_i, X, \sigma) = 1\left\{ U^u_{ij}(X, \sigma) + \frac{2(n-1)}{n-2} Z'_j \Phi^u_i(X, \sigma) \Lambda^u_i(X, \sigma) \omega^u_i(\varepsilon_i, X, \sigma) - \sigma_{ji}\varepsilon_{ij} \geq 0 \right\}, \tag{5.5}$$

almost surely, for all $j \neq i$, where the $T \times 1$ vector $\omega^u_i(\varepsilon_i, X, \sigma) = (\omega^u_{it}(\varepsilon_i, X, \sigma), t = 1, \ldots, T)'$ is a solution to the maximin problem

$$\begin{aligned} &\max_{(\omega_t)_{t \in \mathcal{T}_{i+}}} \min_{(\omega_t)_{t \in \mathcal{T}_{i-}}} \Pi^u_i(\omega; \varepsilon_i, X, \sigma) \\ &= \max_{(\omega_t)_{t \in \mathcal{T}_{i+}}} \min_{(\omega_t)_{t \in \mathcal{T}_{i-}}} \sum_{j \neq i} \left[U^u_{ij}(X, \sigma) + \frac{2(n-1)}{n-2} Z'_j \Phi^u_i(X, \sigma) \Lambda^u_i(X, \sigma) \omega - \sigma_{ji}\varepsilon_{ij} \right]_+ \\ &\qquad - \frac{(n-1)^2}{n-2} \omega' \Lambda^u_i(X, \sigma) \omega \end{aligned}$$

with

$$U^u_{ij}(X, \sigma) = \mathbb{E}\left[S_{ji} u_i(G_j, X) \middle| X, \sigma \right] - \frac{1}{n-2} Z'_j V^u_i(X, \sigma) Z_j,$$

$\sigma_{ji} = \mathbb{E}[S_{ji} | X, \sigma]$, *$\mathcal{T}_{i+} = \{t : \lambda^u_{it}(X, \sigma) > 0\}$, and $\mathcal{T}_{i-} = \{t : \lambda^u_{it}(X, \sigma) < 0\}$. We set*

$\omega_{it}^{u}(\varepsilon_i, X, \sigma) = 0$ if $\lambda_{it}^{u}(X, \sigma) = 0$. Moreover, both $S_i(\varepsilon_i, X, \sigma)$ and $\omega_i^u(\varepsilon_i, X, \sigma)$ are unique almost surely.

The directed and undirected cases are quite similar. Essentially only U_{ij} and V_i are replaced by U_{ij}^u and V_i^u. The two-step estimation must be adapted because we observe the links not the proposals. We leave the full development of our estimator for undirected networks to future research.

5.3 Limiting Game

In this section, we investigate the limit of the network formation game when the number of agents n grows large. We show that the link formation probability in the finite-n game converges to a limit as n approaches infinity. The limiting probability is in some aspects simpler than the finite-n probability. Because the limiting auxiliary variable does not depend on ε_i we do not need simulation to compute the link choice probability. The limiting link choice probability is also everywhere differentiable in the parameters.

By Theorem 3.2 the probability that i forms a link to j conditional on the characteristic profile X and the equilibrium σ is

$$\begin{aligned}
& P_{n,ij}(X, \sigma) \\
={}& \Pr\left(G_{n,ij}(\varepsilon_i, X, \sigma) = 1 \middle| X, \sigma\right) \\
={}& \Pr\left(U_{n,ij}(X, \sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni}(X, \sigma) \Lambda_{ni}(X, \sigma) \omega_{ni}(\varepsilon_i, X, \sigma) - \varepsilon_{ij} \geq 0 \middle| X, \sigma\right).
\end{aligned} \tag{5.6}$$

Note that we added a subscript n to U_{ij} and V_i to emphasize the dependence on the number of agents in the network.

Until now we have avoided an assumption on how the matrix of individual characteristics X is generated by conditioning on X. For the convergence of (5.6) it is convenient to assume that X_i, $i = 1, \ldots, n$, are i.i.d. and that the utility specification is such that $U_{n,ij}(X, \sigma)$ and $V_{ni}(X, \sigma)$ converge to limits $U_{ij}(X_i, X_j, \sigma)$ and $V_i(X_i, \sigma)$ as $n \to \infty$, as formally stated in Assumption 7 and verified below for the utility specified in Section 2. Under the assumptions, the link formation probability $P_{n,ij}(X, \sigma)$

converges to

$$
\begin{aligned}
& P_{ij}\left(X_i, X_j, \sigma\right) \\
= & \Pr\left(U_{ij}\left(X_i, X_j, \sigma\right) + 2Z_j'\Phi_i\left(X_i, \sigma\right)\Lambda_i\left(X_i, \sigma\right)\omega_i\left(X_i, \sigma\right) - \varepsilon_{ij} \geq 0 \middle| X_i, X_j, \sigma\right)
\end{aligned} \tag{5.7}
$$

as $n \to \infty$, where $\Lambda_i\left(X_i, \sigma\right)$ and $\Phi_i\left(X_i, \sigma\right)$ are the eigenvalue and eigenvector matrices of $V_i\left(X_i, \sigma\right)$, and $\omega_i\left(X_i, \sigma\right)$ solves

$$
\begin{aligned}
& \max_{\omega} \Pi_i\left(\omega, X_i, \sigma\right) \\
= & \max_{\omega} \mathbb{E}\left(\left[U_{ij}\left(X_i, X_j, \sigma\right) + 2Z_j'\Phi_i\left(X_i, \sigma\right)\Lambda_i\left(X_i, \sigma\right)\omega - \varepsilon_{ij}\right]_+ \middle| X_i, \sigma\right) - \omega'\Lambda_i\left(X_i, \sigma\right)\omega.
\end{aligned} \tag{5.8}
$$

The expectation in (5.8) is taken with respect to X_j and ε_{ij}.

Assumption 7 *(i) X_i, $i = 1, \ldots, n$, are i.i.d.. (ii) For $U_{n,ij}\left(X, \sigma\right)$ and $V_{ni}\left(X, \sigma\right)$ defined in (3.3) and (3.1), there exist $U_{ij}\left(X_i, X_j, \sigma\right)$ and $V_i\left(X_i, \sigma\right)$ such that for any i and any X_i and X_j, $\max_{j \neq i}\left|U_{n,ij}\left(X, \sigma\right) - U_{ij}\left(X_i, X_j, \sigma\right)\right| = o_p(1)$ and $V_{ni}\left(X, \sigma\right) - V_i\left(X_i, \sigma\right) = o_p(1)$. (iii) For any X_i and σ, $\Pi_i\left(\omega, X_i, \sigma\right)$ defined in (5.8) has a unique maximizer $\omega_i\left(X_i, \sigma\right)$.*

An implication of Assumption 7(i) is that in the limit of the choice probability we average over all X_k, $k \neq i, j$, so that the limiting choice probability only depends on X_i and X_j.[11] Assumption 7(ii) is an assumption on the utility function that we verify below. Assumption 7(iii) ensures that $\omega_i(X_i, \sigma)$ is well-defined.

Theorem 5.3 *Under Assumptions 1-3 and 7, we have that for all X_i and X_j and all σ*

$$
P_{n,ij}\left(X, \sigma\right) - P_{ij}\left(X_i, X_j, \sigma\right) = o_p(1) \tag{5.9}
$$

as $n \to \infty$.

Proof. See the Supplemental Appendix. ■

We refer to $P_{ij}\left(X_i, X_j, \sigma\right)$ defined in (5.7) as the limiting choice probability. It is the choice probability in the limiting game with a continuum of players, where each

[11] Assumption 7(i) could be relaxed to allow for non-i.i.d. X_i, but we do not pursue this here.

individual i forms a link with j following the limiting strategy

$$G_{ij} = 1\left\{U_{ij}\left(X_i, X_j, \sigma\right) + 2Z_j'\Phi_i\left(X_i, \sigma\right)\Lambda_i\left(X_i, \sigma\right)\omega_i\left(X_i, \sigma\right) - \varepsilon_{ij} \geq 0\right\}, \ \forall j \neq i.$$

The $\omega_i\left(X_i, \sigma\right)$ in the limiting strategy captures the expected utility of friends in common. With the inclusion of $\omega_i\left(X_i, \sigma\right)$ the optimal strategy of individual i is to myopically choose to form a link as in a binary choice problem.

We now verify Assumption 7(ii) for the utility function in Section 2.

Example 5.1 *Consider the expected utility in (2.5)-(2.6). Under the assumption that the equilibrium is symmetric, we can denote*

$$\sigma(X_j, X_k) = \mathbb{E}\left[G_{jk}\right| X, \sigma]$$

where the order of the arguments indicates that it is the probability that j links to k.[12] *Therefore*

$$\begin{aligned} U_{n,ij}\left(X, \sigma\right) =& \beta_1 + X_i'\beta_2 + \left|X_i - X_j\right|'\beta_3 + \sigma\left(X_j, X_i\right)\beta_4 \\ &+ \frac{1}{n-2}\sum_{k \neq i,j}\sigma\left(X_j, X_k\right)\beta_5\left(X_i, X_j, X_k\right) - \frac{1}{n-2}Z_j'V_{ni}\left(X, \sigma\right)Z_j \end{aligned}$$

By the law of large numbers the limit is

$$\begin{aligned} U_{ij}\left(X_i, X_j, \sigma\right) =& \beta_1 + X_i'\beta_2 + \left|X_i - X_j\right|'\beta_3 + \sigma\left(X_j, X_i\right)\beta_4 \\ &+ \mathbb{E}\left[\sigma\left(X_j, X_k\right)\beta_5\left(X_i, X_j, X_k\right)\right] \end{aligned}$$

where the expectation is over X_k.

Moreover, $V_{ni}\left(X, \sigma\right) = \left(V_{ni,st}\left(X, \sigma\right), s, t = 1, \ldots, T\right)$ with

$$\begin{aligned} V_{ni,st}\left(X, \sigma\right) =& \sigma\left(x_s, x_t\right)\sigma\left(x_t, x_s\right)\gamma_1\left(X_i, x_s, x_t\right) \\ &+ \frac{1}{n-3}\sum_{l \neq i,j,k}\sigma\left(x_s, X_l\right)\sigma\left(x_t, X_l\right)\gamma_2\left(X_i, x_s, x_t\right). \end{aligned}$$

[12] An equilibrium σ in general depends on the entire X. In this section, our focus is to approximate the finite-n choice probability at a particular σ, so we can treat σ as a fixed matrix and ignore its dependence on X. Any link probability in the expected utility can be viewed as a function of the characteristics of the involved agents only.

By the law of large numbers the limit of $V_{ni,st}(X,\sigma)$ is

$$V_{i,st}(X_i,\sigma) = \sigma(x_s,x_t)\,\sigma(x_t,x_s)\,\gamma_1(X_i,x_s,x_t) + \mathbb{E}\left[\sigma(x_s,X_l)\,\sigma(x_t,X_l)\,\gamma_2(X_i,x_s,x_t)\right],$$

where the expectation is over X_l. A detailed proof can be found in the Supplemental Appendix.

In Section 4 we proposed a two-step GMM estimator for the parameters of the utility function. That estimator requires an instrument in the second stage and we suggested the instrument derived from the quasi likelihood as in (4.14). This instrument involves the derivative of the choice probability that is not everywhere differentiable. Because the limiting choice probability is everywhere differentiable we can use its derivative in (4.14). The instruments based on the finite-n and limiting choice probabilities should behave similarly asymptotically.

A further simplification occurs if in the moment condition in (4.4) we replace the finite-n choice probability by the limiting choice probability that is simpler to compute since it does not require simulation. The formal theory that justifies the use of the limiting model for finite-n networks requires additional results. First, the convergence of the choice probabilities has to be uniform (over links), not just pointwise as established here. An additional complication is the presence of multiple equilibria. Unlike estimation based on the finite-n choice probability in Section 4 which requires no assumption on equilibrium selection, for estimation based on the limiting choice probability to be consistent, we need to assume that the sequence of equilibrium selection mechanisms converges to a limit, similar in spirit to the convergence of equilibria in Menzel (2016).

In the next section we provide simulation evidence on the use of the limiting choice probability in the instrument and/or the moment function.

6 Simulation

In this section, we conduct a simulation study of the two-step estimator proposed in Section 4 for a range of network sizes. We consider three cases: (i) second-step moments for the finite-n game, with instruments derived from the finite-n link probabilities, (ii) second-step moments for the finite-n game, but instruments derived from the limiting link probabilities, (iii) second-step moments for the limiting game, with instruments derived from the limiting link probabilities. For cases (i) and (ii) we established consistency in Section 4. We did not show consistency for case (iii), but our results are suggestive that the estimator is also consistent. It is computationally the easiest case.

In the simulation study we consider the utility specification

$$U_i(G,X,\varepsilon_i) = \textstyle\sum_{j\neq i} G_{ij}\left(\beta_1 + X_i\beta_2 + |X_i - X_j|\,\beta_3 + \dfrac{1}{n-2}\textstyle\sum_{k\neq i,j} G_{jk}\beta_4 \right.$$
$$\left. + \dfrac{1}{n-2}\textstyle\sum_{k\neq i,j} G_{ik}G_{jk}G_{kj}\gamma - \varepsilon_{ij}\right)$$

where X_i are i.i.d. binary variables with equal probability of being 0 or 1, and ε_{ij} are i.i.d. following the standard normal distribution $N(0,1)$. The true parameter values are set to $(\beta_1,\beta_2,\beta_3,\beta_4,\gamma) = (-1,1,-2,1,1)$. The network sizes are $n = 10, 25, 50, 100, 250, 500$.

For each value of n, we generate the links in a single directed network as follows. First, we compute a Bayesian Nash equilibrium by solving (2.8) for $\sigma^*(X)$ by iterating that equation from a starting value.[13] Second, using the equilibrium choice probabilities to compute $U_{n,ij}(X,\sigma)$ and $V_{ni}(X,\sigma)$ we generate the links by (3.8) after we calculate $\omega_{ni}(\varepsilon_i)$ for the simulated ε_i.[14] Each experiment is repeated 100 times. We report the means and standard errors of the estimated parameters.

Table 1 reports the two-step GMM estimates for case (i). The finite-n link probability in (4.4) is computed by simulation. We repeatedly draw ε_i, solve for $\omega_{ni}(\varepsilon_i)$ and substitute that solution in (3.8) to obtain the optimal link choices for that ε_i. The link probability is now approximated by the fraction of draws that result in a link. For the instrument we choose the instrument derived from QMLE in (4.14). The GMM estimator for θ is found by continous updating. The instrument is also

[13] We use an equilibrium in the limiting game as the initial value. This equilibrium is computed by iterating the limiting version of (2.8) where we replace the finite-n choice probability on the right-hand side by the limiting one.

[14] For small networks, generating the links by maximizing the expected utility in (2.4) with respect to G_i by quadratic integer programming (QIP) is computationally competitive. In our simulation study, we use QIP for $n \leq 100$ and (3.8) for $n > 100$. We solve QIP using the solver cplexmiqp provided in CPLEX. QIP is also a check on whether the link choices in (3.8) maximize the expected utility. We compare the simulated link probabilities based on QIP and to those based on (3.8) and find that they are the same.

Table 1: Two-Step GMM Estimates Using the Finite-n Game (Instruments from the Finite-n Game)

n	β_1	β_2	β_3	β_4	γ
10	-0.995	0.958	-1.937	0.981	0.979
	(0.206)	(0.191)	(0.402)	(0.185)	(0.182)
25	-1.010	1.060	-2.038	1.003	0.996
	(0.066)	(0.109)	(0.194)	(0.092)	(0.098)
50	-1.003	0.999	-2.001	1.020	0.988
	(0.042)	(0.065)	(0.097)	(0.083)	(0.072)
100	-0.996	0.993	-2.010	1.031	0.981
	(0.023)	(0.036)	(0.052)	(0.064)	(0.055)
250	-0.998	0.999	-2.000	1.027	0.987
	(0.008)	(0.017)	(0.020)	(0.035)	(0.033)
500	-1.001	1.007	-1.997	0.998	0.995
	(0.006)	(0.011)	(0.011)	(0.028)	(0.019)
DGP	-1	1	-2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the finite-n game, with the GMM instruments simulated independently from the finite-n game. The choice probabilities are computed from 500 simulations by either solving quadratic integer programming (for $n \leq 100$) or applying (3.8) (for $n > 100$).

calculated by simulation,[15] and the derivative in the numerator is approximated by a numerical derivative. Because the sample moment function is not everywhere differentiable we use a derivative-free optimization solver when searching for the estimate of θ.[16] The results in Table 1 show that the two-step GMM based on the finite-n game performs well. The mean estimates are close to the true values even for network sizes as small as $n = 25$. The standard errors also decrease as the network size increases, as expected.

Table 2 presents the results for case (ii), with the finite-n choice probability in the

[15] The instrument is simulated using ε_i that are drawn independently of those drawn to simulate the choice probabilities in the moment function.

[16] We use fminsearch provided in MATLAB.

Table 2: Two-Step GMM Estimates Using the Finite-n Game (Instruments from the Limiting Game)

n	β_1	β_2	β_3	β_4	γ
10	-1.008	1.273	-2.940	0.834	0.943
	(0.438)	(1.010)	(3.338)	(1.900)	(1.004)
25	-1.017	1.012	-2.065	1.016	0.986
	(0.097)	(0.185)	(0.268)	(0.232)	(0.146)
50	-1.010	0.995	-1.995	1.050	0.984
	(0.052)	(0.070)	(0.101)	(0.110)	(0.094)
100	-0.995	0.991	-2.010	1.034	0.979
	(0.023)	(0.040)	(0.050)	(0.073)	(0.062)
250	-0.998	1.000	-2.001	1.031	0.983
	(0.008)	(0.018)	(0.021)	(0.038)	(0.036)
500	-1.001	1.010	-2.000	0.999	0.989
	(0.005)	(0.013)	(0.012)	(0.034)	(0.025)
DGP	-1	1	-2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the finite-n game, with the GMM instruments calculated from the limiting game. The finite-n choice probabilities are computed from 500 simulations by either solving quadratic integer programming (for $n \leq 100$) or applying (3.8) (for $n > 100$).

moment function in (4.4), but in the instrument in (4.14) we replace the finite-n choice probability and its derivative by the limiting choice probability and its derivative. The limiting choice probability and its derivative need not be computed by simulation and the choice probability is differentiable everywhere with respect to θ. We find that for small networks (e.g. $n = 10$) the estimator is slightly biased. The standard errors are larger than those in case (i). For larger networks, the standard errors in cases (i) and (ii) are close, suggesting that in larger networks, the computationally convenient limiting game can be used without sacrificing precision.

In Table 3 we consider case (iii), where in both the moment function and the instrument we use the limiting choice probabilities. This yields a moment condition

Table 3: Two-Step QMLE Estimates Using the Limiting Game

n	β_1	β_2	β_3	β_4	γ
10	-1.152	2.806	-6.469	-2.626	-0.194
	(2.284)	(3.004)	(3.649)	(8.890)	(6.438)
25	-0.719	2.639	-3.899	-1.835	-0.887
	(0.447)	(2.029)	(2.152)	(3.710)	(3.948)
50	-0.986	1.058	-2.064	0.858	0.909
	(0.126)	(0.499)	(0.499)	(0.921)	(0.551)
100	-0.995	1.008	-2.007	0.985	0.959
	(0.034)	(0.084)	(0.084)	(0.165)	(0.208)
250	-1.001	1.004	-2.003	1.009	0.969
	(0.014)	(0.039)	(0.037)	(0.075)	(0.173)
500	-1.001	1.001	-2.000	1.006	0.986
	(0.010)	(0.022)	(0.022)	(0.047)	(0.103)
DGP	-1	1	-2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the limiting game, equivalent to GMM estimates with both the moment function and the instruments calculated based on the limiting game.

that is equal to the first-order condition from QMLE based on the limiting game. This case is computationally convenient, because no simulation is needed. The results show that the estimates are off in small networks. However, the bias disappears as the network size grows. This suggests that the estimator solved from the limiting moment condition is consistent. The standard errors in Table 3 are larger than those in Tables 1-2.

In sum, the simulation results suggest that the two-step estimation procedure based on the finite-n game gives good estimates for the parameters even in relatively small networks. For large networks, estimators based on the limiting game are as good as those based on the finite-n game. This is encouraging given that the limiting choice probabilities are much easier to compute than the finite-n ones.

7 Conclusions

In this paper, we develop a new method for the estimation of network formation games using data from a single large network. We consider a network formation game with incomplete information, where the utility of an agent in a network can be nonseparable in her own link choices to accommodate the utility from friends in common. We propose a new approach in which the optimal link decision of each agent is a set of binary link choices if an auxiliary variable is included. Based on this representation we analyze the dependence between the link choices of an agent, and its effect on the estimation of the parameters of the utility function. We propose a two-step estimation procedure where we estimate the link choice probabilities nonparameterically in the first step and estimate the utility function parameters in the second step. This two-step procedure requires weak assumptions about equilibrium selection, is simple to compute, and provides consistent and asymptotically normal estimators for the parameters that account for the link dependence.

Some extensions of our approach are discussed in Section 5. We consider an unrestricted expected utility of friends in common, undirected networks and the limit of the link choice probabilities. Another important extension is to relax the i.i.d. assumption on the unobservables and introduce an individual effect, similar to that in Graham (2017). This creates stronger and non-vanishing link dependence within an agent. Ridder and Sheng (2017) discuss the estimation and inference in this case.

The Legendre transform may be useful in other applications, where an agent chooses between a large number of overlapping alternatives that exhibit strategic complementarity. Such models are challenging to analyze because the optimal decision of an agent generally does not have a closed form. Our approach may facilitate the econometric analysis of these models.

References

Andrews, D. W. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 37, pp. 2247–2294. Elsevier Science.

Boucheron, S., G. Lugosi, and P. Massart (2013): *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

CILIBERTO, F., AND E. TAMER (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

DE PAULA, A., S. RICHARDS-SHUBIK, AND E. TAMER (2018): "Identifying Preferences in Networks With Bounded Degree," *Econometrica*, 86(1), 263–288.

GRAHAM, B. S. (2017): "An Econometric Model of Network Formation With Degree Heterogeneity," *Econometrica*, 85(4), 1033–1063.

HANSEN, L. P., J. HEATON, AND A. YARON (1996): "Finite-Sample Properties of Some Alternative GMM Estimators," *Journal of Business & Economic Statistics*, 14(3), 262–280.

JACKSON, M. O. (2008): *Social and Economic Networks*. Princeton University Press.

JENNRICH, R. I. (1969): "Asymptotic Properties of Non-Linear Least Squares Estimators," *The Annals of Mathematical Statistics*, 40(2), 633–643.

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

MENZEL, K. (2016): "Inference for Games with Many Players," *The Review of Economic Studies*, 83(1), 306–337.

——— (2017): "Strategic Network Formation with Many Players," *NYU Working Paper*.

MIYAUCHI, Y. (2016): "Structural Estimation of Pairwise Stable Networks with Nonnegative Externality," *Journal of Econometrics*, 195(2), 224–235.

NEWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. IV, chap. 36, pp. 2111–2245. Elsevier Science.

POLLARD, D. (1990): *Empirical Processes: Theory and Applications*, vol. 2 of *NSF-CBMS Regional Conference Series in Probability and Statistics*. Institute of Mathematical Statistics.

RIDDER, G., AND S. SHENG (2017): "Estimation of Social Interactions in Endogenous and Strategically Formed Networks," *UCLA Working Paper*.

Rockafellar, R. T. (1970): *Convex Analysis*. Princeton University Press.

Sheng, S. (2018): "A Structural Econometric Analysis of Network Formation Games Through Subnetworks," *Econometrica (forthcoming)*.

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70(1), 147–165.

van der Vaart, A. W., and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag.

Supplemental Appendix

This supplemental appendix contains all the proofs in the paper. We use $\|.\|$ to denote the Euclidean norm. If the argument is a matrix A the norm is the matrix Euclidean norm $\|A\| = \sqrt{tr\,(A'A)}$.

S.1 Proofs in Section 3

Proof of Proposition 3.1. It suffices to show the first equality as the second equality follows immediately from (3.5). By the real spectral decomposition of $V_i\,(X,\sigma)$, the double-sum term in the expected utility in (3.2) satisfies

$$
\begin{aligned}
&\sum_{j\neq i}\sum_{k\neq i} G_{ij}G_{ik}Z_j' V_i\,(X,\sigma)\,Z_k \\
&= \left(\sum_{j\neq i} G_{ij}Z_j'\right) V_i\,(X,\sigma) \left(\sum_{k\neq i} G_{ik}Z_k\right) \\
&= \left(\sum_{j\neq i} G_{ij}Z_j'\right) \Phi_i\,(X,\sigma)\,\Lambda_i\,(X,\sigma)\,\Phi_i'\,(X,\sigma) \left(\sum_{k\neq i} G_{ik}Z_k\right) \\
&= \left(\sum_{j\neq i} G_{ij}Z_j'\Phi_i\,(X,\sigma)\right) \Lambda_i\,(X,\sigma) \left(\sum_{k\neq i} G_{ik}\Phi_i'\,(X,\sigma)\,Z_k\right) \\
&= (n-1)^2 \sum_{t=1}^{T} \lambda_{it}\,(X,\sigma) \left(\frac{1}{n-1}\sum_{j\neq i} G_{ij}Z_j'\phi_{it}\,(X,\sigma)\right)^2
\end{aligned}
$$

Combining this with (3.2) yields the first equality in (3.6). The proof is complete. ■

Proof of Theorem 3.2. From Proposition 3.1, the expected utility can be written as

$$\begin{aligned}
&\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\
=&\sum_{j\neq i} G_{ij}\left(U_{ij}\left(X,\sigma\right) - \varepsilon_{ij}\right) \\
&+\sum_{t=1}^{T} \lambda_{it}\left(X,\sigma\right) \max_{\omega_t\in\mathbb{R}} \left\{\frac{2\left(n-1\right)}{n-2} \sum_{j\neq i} G_{ij} Z_j' \phi_{it}\left(X,\sigma\right)\omega_t - \frac{\left(n-1\right)^2}{n-2}\omega_t^2\right\} \\
=&\max_{\omega\in\mathbb{R}^T} \sum_{j\neq i} G_{ij}\left(U_{ij}\left(X,\sigma\right) + \frac{2\left(n-1\right)}{n-2} Z_j' \sum_{t=1}^{T}\phi_{it}\left(X,\sigma\right)\lambda_{it}\left(X,\sigma\right)\omega_t - \varepsilon_{ij}\right) \\
&-\frac{\left(n-1\right)^2}{n-2}\sum_{t=1}^{T}\lambda_{it}\left(X,\sigma\right)\omega_t^2 \\
=&\max_{\omega\in\mathbb{R}^T} \sum_{j\neq i} G_{ij}\left(U_{ij}\left(X,\sigma\right) + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_i\left(X,\sigma\right)\Lambda_i\left(X,\sigma\right)\omega - \varepsilon_{ij}\right) \\
&-\frac{\left(n-1\right)^2}{n-2}\omega'\Lambda_i\left(X,\sigma\right)\omega
\end{aligned} \tag{S.1}$$

where $\omega = (\omega_t)_{\forall t} \in \mathbb{R}^T$. The second equality holds because $\lambda_{it}\left(X,\sigma\right) \geq 0$, $t = 1,\ldots,T$, by Assumption 3.

Denote by $\tilde{\Pi}_i\left(G_i,\omega,\varepsilon_i,X,\sigma\right)$ the objective function of the maximization problem in (S.1)

$$\begin{aligned}
\tilde{\Pi}_i\left(G_i,\omega,\varepsilon_i,X,\sigma\right) = &\sum_{j\neq i} G_{ij}\left(U_{ij}\left(X,\sigma\right) + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_i\left(X,\sigma\right)\Lambda_i\left(X,\sigma\right)\omega - \varepsilon_{ij}\right) \\
&-\frac{\left(n-1\right)^2}{n-2}\omega'\Lambda_i\left(X,\sigma\right)\omega
\end{aligned}$$

From (S.1), the maximized expected utility can be derived from

$$\begin{aligned}
&\max_{G_i}\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\
=&\max_{G_i}\max_{\omega}\tilde{\Pi}_i\left(G_i,\omega,\varepsilon_i,X,\sigma\right) \\
=&\max_{\omega}\max_{G_i}\tilde{\Pi}_i\left(G_i,\omega,\varepsilon_i,X,\sigma\right) \\
=&\max_{\omega}\Pi_i\left(\omega,\varepsilon_i,X,\sigma\right)
\end{aligned} \tag{S.2}$$

where $\Pi_i(\omega, \varepsilon_i, X, \sigma)$ is defined in (3.9). The second equality in (S.2) follows because $\max_\omega \tilde{\Pi}_i(G_i, \omega) \leq \max_\omega \max_{G_i} \tilde{\Pi}_i(G_i, \omega)$ for all G_i, so $\max_{G_i} \max_\omega \tilde{\Pi}_i(G_i, \omega) \leq \max_\omega \max_{G_i} \tilde{\Pi}_i(G_i, \omega)$, and similarly we can prove the other direction. The last equality follows from the definition of $\Pi_i(\omega, \varepsilon_i, X, \sigma)$. The result in (S.2) shows that the maximum expected utility can be obtained by solving the last maximization problem in (S.2) or equivalently (3.9).

By the definition of $G_i(X, \varepsilon_i, \sigma)$ and $\omega_i(X, \varepsilon_i, \sigma)$, we have

$$\begin{aligned}
&\max_\omega \Pi_i(\omega, \varepsilon_i, X, \sigma) \\
=&\tilde{\Pi}_i(G_i(X, \varepsilon_i, \sigma), \omega_i(X, \varepsilon_i, \sigma); X, \varepsilon_i, \sigma) \\
\leq&\max_\omega \tilde{\Pi}_i(G_i(X, \varepsilon_i, \sigma), \omega, \varepsilon_i, X, \sigma) \\
=&\mathbb{E}\left[U_i(G_i(X, \varepsilon_i, \sigma), G_{-i}, X, \varepsilon_i) | X, \varepsilon_i, \sigma\right]
\end{aligned} \tag{S.3}$$

where the last equality comes from (S.1). Combining (S.2) and (S.3) we see that the inequality in (S.3) becomes an equality. Therefore, $G_i(X, \varepsilon_i, \sigma)$ is an optimal solution.

As for the uniqueness, $G_i(X, \varepsilon_i, \sigma)$ is unique almost surely because ε_i has a continuous distribution by Assumption 1, so two link decisions achieve the same expected utility with probability zero. To show the uniqueness of $\Lambda_i(X, \sigma)\omega_i(X, \varepsilon_i, \sigma)$, note that (S.3) implies that $\omega_i(X, \varepsilon_i, \sigma)$ is an optimal solution to the maximization problem $\max_\omega \tilde{\Pi}_i(G_i, \omega, \varepsilon_i, X, \sigma)$ evaluated at $G_i = G_i(X, \varepsilon_i, \sigma)$, so $\omega_i(X, \varepsilon_i, \sigma)$ satisfies the first-order condition

$$\Lambda_i(X, \sigma)\omega_i(X, \varepsilon_i, \sigma) = \frac{1}{n-1}\Lambda_i(X, \sigma)\Phi_i'(X, \sigma)\sum_{j\neq i} G_{ij}(X, \varepsilon_i, \sigma) Z_j \tag{S.4}$$

Since $G_i(X, \varepsilon_i, \sigma)$ is unique almost surely, so is $\Lambda_i(X, \sigma)\omega_i(X, \varepsilon_i, \sigma)$. The proof is complete. ■

Lemma S.1 *Suppose Assumption 1-3 are satisfied. An $\omega_i(\varepsilon_i, X, \sigma)$ that solves the maximization problem in (3.9) satisfies the first-order condition*

$$\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij}(X, \sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_i(X, \sigma)\Lambda_i(X, \sigma)\omega - \varepsilon_{ij} \geq 0\right\}\Lambda_i(X, \sigma)\Phi_i'(X, \sigma) Z_j$$
$$= \Lambda_i(X, \sigma)\omega$$

almost surely.

Proof. Omit X and σ in the notation. Since $\Pi_i(\omega, \varepsilon_i)$ is sub-differentiable at all ω,[17] by optimality of $\omega_i(\varepsilon_i)$, $\Pi_i(\omega, \varepsilon_i)$ has subgradient 0 at $\omega_i(\varepsilon_i)$, that is, $\omega_i(\varepsilon_i)$ satisfies the first-order condition

$$\begin{aligned}
&\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega - \varepsilon_{ij} > 0\right\}\Lambda_i\Phi_i'Z_j - \Lambda_i\omega \\
&= -\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega - \varepsilon_{ij} = 0\right\} diag(\tau)\Lambda_i\Phi_i'Z_j, \qquad \text{(S.5)}
\end{aligned}$$

for some $\tau = (\tau_1, \ldots, \tau_T) \in [0,1]^T$. Define the right-hand side of (S.5) as $\Delta_n(\omega, \varepsilon_i)$. For any ω,

$$\begin{aligned}
&\Pr\left(\|\Delta_n(\omega, \varepsilon_i)\| > 0 \middle| X, \sigma\right) \\
&\leq \Pr\left(\exists j \neq i, U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega = \varepsilon_{ij} \middle| X, \sigma\right) \\
&\leq \sum_{j\neq i}\Pr\left(U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega = \varepsilon_{ij} \middle| X, \sigma\right) = 0, \qquad \text{(S.6)}
\end{aligned}$$

because ε_{ij} has a continuous distribution. Hence the first-order condition (S.5) holds with $\Delta_n(\omega, \varepsilon_i)$ replaced by 0 with probability 1. By (S.6) again, we obtain

$$\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega_i(\varepsilon_i) - \varepsilon_{ij} \geq 0\right\}\Lambda_i\Phi_i'Z_j - \Lambda_i\omega_i(\varepsilon_i) = 0, \text{ a.s.}$$

■

Proof of Corollary 3.3. Omit X and σ in the notation. By Lemma S.1, $\omega_i(\varepsilon_i)$ is a solution to the first-order condition

$$\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega \geq \varepsilon_{ij}\right\}\Lambda_i\Phi_i'Z_j = \Lambda_i\omega, \text{ a.s..} \qquad \text{(S.7)}$$

Note that the first-order condition could have multiple solutions, and among these local solutions, $\omega_i(\varepsilon_i)$ is the unique maximizer of $\Pi_i(\omega, \varepsilon_i)$. For this reason we refer

[17] Notice that the function $\max\{x, 0\}$ is differentiable for $x \neq 0$ and sub-differentiable for $x = 0$ with subderivatives in $[0, 1]$.

to $\omega_i(\varepsilon_i)$ as the global solution.

For any $\omega \in \mathbb{R}^T$, define the choice function $g_i(\omega;\varepsilon_i) = (g_{ij}(\omega;\varepsilon_{ij}), j \neq i) : \mathbb{R}^T \to \{0,1\}^{n-1}$ by

$$g_{ij}(\omega;\varepsilon_{ij}) = 1\left\{U_{ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_i \Lambda_i \omega \geq \varepsilon_{ij}\right\}, \ \forall j \neq i. \tag{S.8}$$

The first-order condition (S.7) defines a system of equations over ω

$$\Lambda_i \omega = \frac{1}{n-1} \sum_{j \neq i} g_{ij}(\omega;\varepsilon_{ij}) \Lambda_i \Phi_i' Z_j, \ \text{a.s.}. \tag{S.9}$$

On the other hand, for any $g_i = (g_{ij}, j \neq i) \in \{0,1\}^{n-1}$, define the function $\omega_i(g_i) : \{0,1\}^{n-1} \to \mathbb{R}^T$ by

$$\omega_i(g_i) = \frac{1}{n-1} \sum_{j \neq i} g_{ij} \Phi_i' Z_j. \tag{S.10}$$

We derive a system of equations over g_i

$$g_{ij} = 1\left\{U_{ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_i \Lambda_i \omega_i(g_i) \geq \varepsilon_{ij}\right\}, \ \forall j \neq i. \tag{S.11}$$

We show that with probability 1 there is a one-to-one mapping between the solutions to (S.9) and the solutions to (S.11).

First, for any local solution $\omega_i^l(\varepsilon_i)$ that solves system (S.9), the choice function $g_i(\omega;\varepsilon_i)$ evaluated at $\omega_i^l(\varepsilon_i)$, i.e., $g_i\left(\omega_i^l(\varepsilon_i);\varepsilon_i\right)$, is a solution to system (S.11) with probability 1. To see this, note that by the first-order condition in (S.9) and the definition of $\omega_i(g_i)$ in (S.10) $\omega_i^l(\varepsilon_i)$ satisfies

$$\begin{aligned}\Lambda_i \omega_i^l(\varepsilon_i) &= \frac{1}{n-1} \sum_{j \neq i} g_{ij}\left(\omega_i^l(\varepsilon_i);\varepsilon_{ij}\right) \Lambda_i \Phi_i' Z_j, \ \text{a.s.} \\ &= \Lambda_i \omega_i\left(g_i\left(\omega_i^l(\varepsilon_i);\varepsilon_i\right)\right). \end{aligned} \tag{S.12}$$

Then by the definition of $g_i(\omega;\varepsilon_i)$ in (S.8), for any $j \neq i$,

$$\begin{aligned}
& g_{ij}\left(\omega_i^l(\varepsilon_i);\varepsilon_{ij}\right) \\
=& 1\left\{U_{ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_i \Lambda_i \omega_i^l(\varepsilon_i) \geq \varepsilon_{ij}\right\} \\
=& 1\left\{U_{ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_i \Lambda_i \omega_i\left(g_i\left(\omega_i^l(\varepsilon_i);\varepsilon_i\right)\right) \geq \varepsilon_{ij}\right\}, \text{ a.s..}
\end{aligned}$$

This shows that $g_{ij}\left(\omega_i^l(\varepsilon_i);\varepsilon_{ij}\right)$ satisfies system (S.11) with probability 1. Second, for two distinct $\omega_i^{l_1}(\varepsilon_i) \neq \omega_i^{l_2}(\varepsilon_i)$, by (S.12), with probability 1, we have $g_i\left(\omega_i^{l_1}(\varepsilon_i);\varepsilon_i\right) \neq g_i\left(\omega_i^{l_2}(\varepsilon_i);\varepsilon_i\right)$. Therefore, there is a one-to-one mapping between the solutions to (S.9) and (S.11) with probability 1.

The equivalence between systems (S.9) and (S.11) motivates us to analyze the relationship between $\omega_i(\varepsilon_i)$ and ε_i through the relationship between the solutions to (S.11) and ε_i. By the definition of $\omega_i(g_i)$ in (S.10), write system (S.11) explicitly as

$$g_{ij} = 1\left\{U_{ij} + \frac{2}{n-2}\sum_{k\neq i} g_{ik} Z_j' V_i Z_k \geq \varepsilon_{ij}\right\},\ \forall j \neq i,\ \text{ a.s.} \tag{S.13}$$

For any $g_i \in \{0,1\}^{n-1}$, define the set

$$\mathcal{E}_i^l(g_i) = \left\{\varepsilon_i \in \mathbb{R}^{n-1} : g_i \text{ satisfies (S.13)}\right\}. \tag{S.14}$$

This set can be regarded as the collection of ε_i that support g_i as a solution to (S.13). Note that since ε_i has support on $\mathbb{R}^{n-1}$, the set $\mathcal{E}_i^l(g_i)$ is nonempty for all $g_i \in \{0,1\}^{n-1}$.

As discussed, system (S.13) may have multiple solutions, resembling the presence of multiple equilibria in entry games (Ciliberto and Tamer, 2009; Tamer, 2003). In particular, it is possible that the sets in (S.14) for two different g_i and g_i' overlap and in the overlapping area both g_i and g_i' satisfy (S.13). For example, assume that all elements in V_i are positive so link choices are strategic complements. In the region of ε_i where

$$U_{ij} < \varepsilon_{ij} \leq U_{ij} + \frac{2}{n-2}\sum_{k\neq i} Z_j' V_i Z_k,\ \forall j \neq i,$$

we find that both $(g_{ij}=1, j\neq i)$ and $(g_{ij}=0, j\neq i)$ are solutions to (S.13).

Unlike in entry games where equilibrium selection mechanisms are typically unknown, in our case we have a natural selection mechanism. Recall that from Proposition 3.2 the optimal link decision $G_i(\varepsilon_i) = (G_{ij}(\varepsilon_i), j \neq i) \in \{0,1\}^{n-1}$ is given by the choice function (S.8) evaluated at the global solution $\omega_i(\varepsilon_i)$, i.e.,

$$G_{ij}(\varepsilon_i) = g_{ij}(\omega_i(\varepsilon_i); \varepsilon_{ij}), \ \forall j \neq i.$$

From our earlier discussion we can see that $G_{ij}(\varepsilon_i)$ satisfies (S.13). For a given $\varepsilon_i \in \mathbb{R}^{n-1}$, system (S.13) could have multiple solutions, and among all such solutions $G_{ij}(\varepsilon_i)$ is selected because it is the choice function evaluated at $\omega_i(\varepsilon_i)$, the global maximizer of the objective function

$$\Pi_i(\omega, \varepsilon_i) = \sum_{j\neq i}\left[U_{ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_i \Lambda_i \omega - \varepsilon_{ij}\right]_+ - \frac{(n-1)^2}{n-2}\omega' \Lambda_i \omega.$$

To further characterize the selection mechanism, we examine the objective function $\Pi_i(\omega, \varepsilon_i)$ evaluated at the local solutions. Note that for any $\omega^l \in \mathbb{R}^T$ that solves the system (S.9), we have

$$\Lambda_i \omega^l = \frac{1}{n-1}\sum_{j\neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right)\Lambda_i \Phi_i' Z_j, \text{ a.s.}$$

Hence, $\Pi_i\left(\omega^l, \varepsilon_i\right)$ can be represented as

$$\begin{aligned}
&\Pi_i\left(\omega^l, \varepsilon_i\right) \\
=&\sum_{j\neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right)\left(U_{ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_i \Lambda_i \omega^l - \varepsilon_{ij}\right) \\
&-\frac{(n-1)^2}{n-2}\left(\frac{1}{n-1}\sum_{j\neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right)\Lambda_i \Phi_i' Z_j\right)' \omega^l, \text{ a.s.} \\
=&\sum_{j\neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right)\left(U_{ij} + \frac{n-1}{n-2} Z_j' \Phi_i \Lambda_i \omega^l - \varepsilon_{ij}\right) \\
=&\sum_{j\neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right)\left(U_{ij} + \frac{1}{n-2}\sum_{k\neq i} g_{ik}\left(\omega^l; \varepsilon_{ik}, \theta, p\right) Z_j' V_i Z_k - \varepsilon_{ij}\right), \text{ a.s.} \qquad \text{(S.15)}
\end{aligned}$$

This indicates that $\Pi_i\left(\omega^l,\varepsilon_i\right)$ can be regarded as a function of the link decision $g_i\left(\omega^l;\varepsilon_i\right)$ that corresponds to ω^l.

By the global optimality of $\omega_i\left(\varepsilon_i\right)$ we have $\Pi_i\left(\omega_i\left(\varepsilon_i\right),\varepsilon_i\right)\geq\Pi_i\left(\omega^l,\varepsilon_i\right)$ for all ω^l that solves the first-order condition (S.9). By the represetation of $\Pi_i\left(\omega^l,\varepsilon_i\right)$ in (S.15) and the equivalence between (S.9) (thus (S.7)) and (S.13), if $G_i\left(\varepsilon_i\right)$ takes a value $g_i\in\{0,1\}^{n-1}$, then g_i must satisfy both (S.13) and

$$\sum_{j\neq i}g_{ij}\left(U_{ij}+\frac{1}{n-2}\sum_{k\neq i}g_{ik}Z_j'V_iZ_k-\varepsilon_{ij}\right)\geq\sum_{j\neq i}g_{ij}^l\left(U_{ij}+\frac{1}{n-2}\sum_{k\neq i}g_{ik}^lZ_j'V_iZ_k-\varepsilon_{ij}\right)\tag{S.16}$$

almost surely, for all g_i^l that solve (S.13). We can view (S.16) as a selection criterion that determines which solution to (S.13) is selected.

Therefore, for any $g_i\in\{0,1\}^{n-1}$, we can define the set

$$\mathcal{E}_i\left(g_i\right)=\left\{\varepsilon_i\in\mathbb{R}^{n-1}:g_i\text{ satisfies both (S.13) and (S.16)}\right\}.\tag{S.17}$$

This set is the collection of ε_i that support g_i as the unique optimal decision, i.e., for any $\varepsilon_i\in\mathcal{E}_i\left(g_i\right)$, we have $G_i\left(\varepsilon_i\right)=g_i$. The uniqueness implies that if $g_i\neq g_i'$, the sets $\mathcal{E}_i\left(g_i\right)$ and $\mathcal{E}_i\left(g_i'\right)$ are disjoint. The collection of such sets $\mathcal{E}_i\left(g_i\right)$ for all $g_i\in\{0,1\}^{n-1}$ thus forms a partition of the space of ε_i, with each region in the partition corresponding to a unique optimal link decision, similarly as in entry games (Ciliberto and Tamer, 2009; Tamer, 2003). The proof is complete. ■

Proof of Proposition 3.4. We follow the proof in Leung (2015, Theorem 1). We organize the choice probabilities in an $n\times 2^{n-1}$ matrix $\sigma\left(X\right)$. The ith row has individual i's choice probabilities $\sigma_i\left(X\right)=\left\{\sigma_i\left(g_i|\,X\right),g_i\in\mathcal{G}_i\right\}$. The entries in the row sum to 1. The set of such matrices is $\Sigma\left(X\right)$. With row i of $\sigma\left(X\right)$ we associate X_i.

Let $\Sigma^s\left(X\right)\subset\Sigma\left(X\right)$ be the subset of matrices of choice probabilities, such that if $X_i=X_j$ then $\sigma_i\left(X\right)=\sigma_j\left(X\right)$, i.e., rows i and j are identical.

If we organize the choice probabilities in (2.7) in an $n\times 2^{n-1}$ matrix $P\left(X,\sigma\right)$, it maps the matrix σ to a matrix of choice probabilities in $\Sigma\left(X\right)$. An equilibrium is a fixed point of this mapping. Because we focus on symmetric equilibria in $\Sigma^s\left(X\right)$, we have to show that $P\left(X,\sigma\right)$ is a continuous mapping from $\Sigma^s\left(X\right)$ to $\Sigma^s\left(X\right)$ and that $\Sigma^s\left(X\right)$ is convex and compact.

First, the mapping $P(X,\sigma)$ maps $\Sigma^s(X)$ to itself. Let $\sigma(X) \in \Sigma^s(X)$. If $X_i = X_j$, then in the expected utilities of i and j, (2.5) and (2.6) are equal for i and j. Because ε_i and ε_j have the same distribution, rows i and j of $P(X,\sigma(X))$ are identical, so indeed $P(X,\sigma(X)) \in \Sigma^s(X)$.

Second, a convex combination of matrices $\sigma(X), \tilde{\sigma}(X) \in \Sigma^s(X)$ is a matrix with rows that sum to 1 and that rows i and j are identical if $X_i = X_j$. The convex combination is therefore in $\Sigma^s(X)$.

Third, $\Sigma^s(X)$ is bounded. It is also closed. Let $\left\{\sigma^k(X), k = 1, 2, \ldots\right\}$ be a sequence in $\Sigma^s(X)$ that converges to a limit. Then for all k the rows of $\sigma^k(X)$ sum to 1 and rows i and j are identical if $X_i = X_j$. So the limit has the same properties and is therefore in $\Sigma^s(X)$.

Finally, the mapping $P(X,\sigma)$ is continuous on $\Sigma^s(X)$. This is shown in Lemma S.4 in the Supplemental Appendix.

We conclude that by Brower's fixed point theorem, $P(X,\sigma)$ has a fixed point in $\Sigma^s(X)$. The proof is complete. ■

S.2 Proofs in Section 4

S.2.1 Consistency

Proof of Theorem 4.1. We follow the consistency proof in Newey and McFadden (1994). A complication is the presence of the first-stage parameter p_n. Fix $\delta > 0$. Let $\mathcal{B}_\delta(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| < \delta\}$ be an open δ-ball centered at θ_0. If $\left\|\Psi_n\left(\hat{\theta}_n, p_n\right)\right\| < \inf_{\theta \in \Theta \setminus \mathcal{B}_\delta(\theta_0)} \|\Psi_n(\theta, p_n)\|$, then $\hat{\theta}_n \notin \Theta \setminus \mathcal{B}_\delta(\theta_0)$, or equivalently, $\hat{\theta}_n \in \mathcal{B}_\delta(\theta_0)$. Therefore,

$$\begin{aligned}
&\Pr\left(\left\|\hat{\theta}_n - \theta_0\right\| < \delta \middle| X, p_n\right) \\
&\geq \Pr\left(\left\|\Psi_n\left(\hat{\theta}_n, p_n\right)\right\| < \inf_{\theta \in \Theta \setminus \mathcal{B}_\delta(\theta_0)} \|\Psi_n(\theta, p_n)\| \middle| X, p_n\right).
\end{aligned} \tag{S.18}$$

Because by Assumption 4(i)-(ii) and Lemma S.4

$$\inf_{\theta \in \Theta \setminus \mathcal{B}_\delta(\theta_0)} \|\Psi_n(\theta, p_n)\| > 0,$$

the right-hand side in (S.18) goes to 1, if

$$\left\|\Psi_n\left(\hat{\theta}_n, p_n\right)\right\| = o_p(1). \tag{S.19}$$

Now by the triangle inequality

$$\begin{aligned}\left\|\Psi_n\left(\hat{\theta}_n, p_n\right)\right\| &\leq \left\|\hat{\Psi}_n\left(\hat{\theta}_n, p_n\right)\right\| + \left\|\hat{\Psi}_n\left(\hat{\theta}_n, p_n\right) - \Psi_n\left(\hat{\theta}_n, p_n\right)\right\| \\ &\leq \left\|\hat{\Psi}_n\left(\hat{\theta}_n, p_n\right)\right\| + \sup_{\theta\in\Theta}\left\|\hat{\Psi}_n\left(\theta, p_n\right) - \Psi_n\left(\theta, p_n\right)\right\|.\end{aligned}$$

By the uniform LLN in Lemma S.3 the second term of the last inequality is $o_p(1)$, so we need to show that $\left\|\hat{\Psi}_n\left(\hat{\theta}_n, p_n\right)\right\| = o_p(1)$.

We have

$$\begin{aligned}\left\|\hat{\Psi}_n\left(\hat{\theta}_n, p_n\right)\right\| &\leq \left\|\hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right)\right\| + \left\|\hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right) - \hat{\Psi}_n\left(\hat{\theta}_n, p_n\right)\right\| \\ &\leq \left\|\hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right)\right\| + \sup_{\theta\in\Theta}\left\|\hat{\Psi}_n\left(\theta, \hat{p}_n\right) - \hat{\Psi}_n\left(\theta, p_n\right)\right\|,\end{aligned}$$

and $\left\|\hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right)\right\| = o_p(1)$ by (4.5), so we need to show that the second term is also $o_p(1)$.

For any $p \in [0,1]^{T^2}$, we have

$$\begin{aligned}&\sup_{\theta\in\Theta}\left\|\hat{\Psi}_n\left(\theta, p\right) - \hat{\Psi}_n\left(\theta, p_n\right)\right\| \\ \leq &\frac{1}{n\left(n-1\right)}\sum_i\sum_{j\neq i}\sup_{\theta\in\Theta}\left\|\hat{W}_{n,ij}\left(P_{n,ij}\left(\theta, p\right) - P_{n,ij}\left(\theta, p_n\right)\right)\right\| \\ \leq &\max_{i,j=1,\ldots,n}\left\|\hat{W}_{n,ij}\right\|\max_{i,j=1,\ldots,n}\sup_{\theta\in\Theta}\left\|P_{n,ij}\left(\theta, p\right) - P_{n,ij}\left(\theta, p_n\right)\right\| \\ = &\max_{i,j=1,\ldots,n}\left\|\hat{W}_{n,ij}\right\|\max_{s,t=1,\ldots,T}\sup_{\theta\in\Theta}\left\|P_{n,(st)}\left(\theta, p\right) - P_{n,(st)}\left(\theta, p_n\right)\right\|,\end{aligned}$$

where $P_{n,(st)}\left(\theta, p\right)$ represents the value of $P_{n,ij}\left(\theta, p\right)$ if $X_i = x_s$ and $X_j = x_t$. By Lemma S.4 $P_{n,(st)}(\theta, p)$ is continuous in θ and p at any $\theta \in \Theta$ and p_n. Since Θ is a compact set, this function is uniformly continuous in θ on Θ and pointwise continuous in p at p_n.

If a function $f(\theta, p)$ is uniformly continuous in θ on Θ and pointwise continuous in p at p_n, then $\sup_{\theta\in\Theta}\|f(\theta, p) - f(\theta, p_n)\|$ is continuous in p at p_n. This is true because for any $\eta > 0$ there is a δ such that $\|(\theta', p) - (\theta, p_n)\| < \delta$ implies that $\|f(\theta', p) - f(\theta, p_n)\| < \eta$ where δ does not depend on θ, θ', and p. Now if $\|p - p_n\| < \delta$, we have also $\|(\theta, p) - (\theta, p_n)\| < \delta$ for all θ, so

$$\sup_{\theta \in \Theta} \|f(\theta, p) - f(\theta, p_n)\| < \eta.$$

By letting $f(\theta, p) = P_{n,(st)}(\theta, p)$, we derive that $\sup_{\theta \in \Theta} \left\| P_{n,(st)}(\theta, p) - P_{n,(st)}(\theta, p_n) \right\|$ is continuous in p at p_n. This together with Assumption (4)(iii) implies that the function $\sup_{\theta \in \Theta} \left\| \hat{\Psi}_n(\theta, p) - \hat{\Psi}_n(\theta, p_n) \right\|$ is continuous in p at p_n.

By the continuous mapping theorem and the consistency of $\hat{p}_n$ in Lemma S.2,

$$\sup_{\theta \in \Theta} \left\| \hat{\Psi}_n(\theta, \hat{p}_n) - \hat{\Psi}_n(\theta, p_n) \right\| \xrightarrow{p} 0,$$

as $n \to \infty$, so (S.19) holds and weak consistency is proven. ■

Lemma S.2 (Consistency of $\hat{p}_n$) *Suppose that Assumptions 1-3 and 4(iv) are satisfied. The first-step estimator $\hat{p}_n$ is consistent for p_n, i.e., for any $\delta > 0$,*

$$\Pr\left(\|\hat{p}_n - p_n\| > \delta \mid X, p_n\right) \to 0$$

as $n \to \infty$.

Proof. Recall that $\hat{p}_n = (\hat{p}_{n,st}, s, t = 1, \ldots, T)$ and $p_n = (p_{n,st}, s, t = 1, \ldots, T)$, where $\hat{p}_{n,st}$ is the link frequency of pairs with the characteristics x_s and x_t

$$\hat{p}_{n,st} = \frac{\sum_i \sum_{j \neq i} G_{n,ij} 1\{X_i = x_s, X_j = x_t\}}{\sum_i \sum_{j \neq i} 1\{X_i = x_s, X_j = x_t\}}$$

and $p_{n,st}$ is the population link probability of such pairs

$$p_{n,st} = \mathbb{E}\left[G_{n,ij} \mid X_i = x_s, X_j = x_t, X, p_n\right],$$

so

$$\mathbb{E}\left[(G_{n,ij} - p_{n,st}) \mid X, p_n\right] 1\{X_i = x_s, X_j = x_t\} = 0 \tag{S.20}$$

By Chebyshev's inequality, for any $\delta > 0$,

$$\Pr\left(\|\hat{p}_n - p_n\| > \delta \middle| X, p_n \right) \leq \frac{1}{\delta^2} \mathbb{E}\left[\|\hat{p}_n - p_n\|^2 \middle| X, p_n \right].$$

It suffices to show that $\mathbb{E}\left[\|\hat{p}_n - p_n\|^2 \middle| X, p_n \right] \to 0$ as $n \to \infty$.

Observe that

$$\begin{aligned}
\mathbb{E}\left[\|\hat{p}_n - p_n\|^2 \middle| X, p_n \right] &= \mathbb{E}\left[\sum_s \sum_t (\hat{p}_{n,st} - p_{n,st})^2 \middle| X, p_n \right] \\
&= \sum_s \sum_t \mathbb{E}\left[(\hat{p}_{n,st} - p_{n,st})^2 \middle| X, p_n \right]. \qquad \text{(S.21)}
\end{aligned}$$

We can write

$$\hat{p}_{n,st} - p_{n,st} = \frac{\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} (G_{n,ij} - p_{n,st}) 1\{X_i = x_s, X_j = x_t\}}{\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} 1\{X_i = x_s, X_j = x_t\}}$$

Therefore, the conditional variance of $\hat{p}_{n,st} - p_{n,st}$ given X and p_n has a numerator

$$\begin{aligned}
&\frac{1}{n^2 (n-1)^2} \sum_i \sum_{j \neq i} \mathbb{E}\left[(G_{n,ij} - p_{n,st})^2 \middle| X, p_n \right] 1\{X_i = x_s, X_j = x_t\} \\
&+ \frac{1}{n^2 (n-1)^2} \sum_i \sum_{j \neq i} \sum_{k \neq i,j} \mathbb{E}\left[(G_{n,ij} - p_{n,st}) (G_{n,ik} - p_{n,st}) \middle| X, p_n \right] \\
&\qquad \cdot 1\{X_i = x_s, X_j = x_t, X_k = x_t\} \\
&+ \frac{1}{n^2 (n-1)^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \mathbb{E}\left[(G_{n,ij} - p_{n,st}) (G_{n,kl} - p_{n,st}) \middle| X, p_n \right] \\
&\qquad \cdot 1\{X_i = x_s, X_j = x_t, X_k = x_s, X_l = x_t\} \qquad \text{(S.22)}
\end{aligned}$$

Because the link choices are independent between individuals, the last term in (S.22) is 0 by (S.20). Further,

$$\mathbb{E}\left[(G_{n,ij} - p_{n,st})^2 \middle| X, p_n \right] 1\{X_i = x_s, X_j = x_t\} \leq 1$$

and

$$\mathbb{E}\left[(G_{n,ij} - p_{n,st}) (G_{n,ik} - p_{n,st}) \middle| X, p_n \right] 1\{X_i = x_s, X_j = x_t, X_k = x_t\} \leq 1$$

so the numerator in (S.22) is bounded by

$$\frac{n\left(n-1\right)}{n^{2}\left(n-1\right)^{2}}+\frac{n\left(n-1\right)\left(n-2\right)}{n^{2}\left(n-1\right)^{2}}=\frac{1}{n}$$

Because $\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}1\left\{X_i=x_s,X_j=x_t\right\}$ converges to a strictly positive limit by Assumption 4(iv), the denominator of the conditional variance of $\hat{p}_{n,st}-p_{n,st}$ converges to the square of that limit. Therefore, the conditional variance of $\hat{p}_{n,st}-p_{n,st}$ is $o\left(1\right)$ for each s and t. This implies

$$\mathbb{E}\left[\left\|\hat{p}_n-p_n\right\|^2\middle|X,p_n\right]\to 0$$

and $\hat{p}_n$ is consistent for p_n. ∎

Lemma S.3 (Uniform LLN for Sample Moments) *Suppose that Assumptions 1-3 and 4(iii) are satisfied. For any $\delta>0$,*

$$\Pr\left(\sup_{\theta\in\Theta}\left\|\hat{\Psi}_n\left(\theta,p_n\right)-\Psi_n\left(\theta,p_n\right)\right\|>\delta\middle|X,p_n\right)\to 0 \tag{S.23}$$

as $n\to\infty$.

Proof. By the definition of $\hat{\Psi}_n$ and Ψ_n

$$\begin{aligned}&\hat{\Psi}_n\left(\theta,p_n\right)-\Psi_n\left(\theta,p_n\right)\\ =&\frac{1}{n\left(n-1\right)}\sum_i\sum_{j\neq i}\hat{W}_{n,ij}\left(G_{n,ij}-P_{n,ij}\left(\theta,p_n\right)\right)-W_{n,ij}\left(\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]-P_{n,ij}\left(\theta,p_n\right)\right)\\ =&\frac{1}{n\left(n-1\right)}\sum_i\sum_{j\neq i}\left(\hat{W}_{n,ij}-W_{n,ij}\right)\left(G_{n,ij}-P_{n,ij}\left(\theta,p_n\right)\right)\\ &+\frac{1}{n\left(n-1\right)}\sum_i\sum_{j\neq i}W_{n,ij}\left(G_{n,ij}-\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]\right)\end{aligned} \tag{S.24}$$

The first term in the last expression in (S.24) is $o_p\left(1\right)$ uniformly over $\theta\in\Theta$ because

$$\begin{aligned}\sup_{\theta\in\Theta}\left\|\frac{1}{n\left(n-1\right)}\sum_i\sum_{j\neq i}\left(\hat{W}_{n,ij}-W_{n,ij}\right)\left(G_{n,ij}-P_{n,ij}\left(\theta,p_n\right)\right)\right\|&\leq\max_{i,j=1,\ldots,n}\left\|\hat{W}_{n,ij}-W_{n,ij}\right\|\\ &=o_p\left(1\right)\end{aligned}$$

by Assumption 4(iii). Write the last term in (S.24) as

$$\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}W_{n,ij}\left(G_{n,ij}-\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]\right)=\frac{1}{n}\sum_i Y_{ni}$$

with

$$Y_{ni}=\frac{1}{n-1}\sum_{j\neq i}W_{n,ij}\left(G_{n,ij}-\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]\right)$$

Note that $\frac{1}{n}\sum_i Y_{ni}$ does not depend on θ. We prove that it is $o_p(1)$ following the proof for a pointwise LLN. By Chebyshev's inequality, for any $\delta>0$,

$$\Pr\left(\left\|\frac{1}{n}\sum_i Y_{ni}\right\|>\delta\middle|X,p_n\right)\leq\frac{1}{\delta^2}\mathbb{E}\left[\left\|\frac{1}{n}\sum_i Y_{ni}\right\|^2\middle|X,p_n\right].$$

Note that conditional on X and p_n, the random variables Y_{ni}, $i=1,\ldots,n$, are independent with mean 0. Therefore, $\frac{1}{n}\sum_i Y_{ni}$ has the conditional variance

$$\begin{aligned}
&\mathbb{E}\left[\left\|\frac{1}{n}\sum_i Y_{ni}\right\|^2\middle|X,p_n\right]\\
=&\frac{1}{n^2}\sum_i\mathbb{E}\left[\|Y_{ni}\|^2\middle|X,p_n\right]\\
=&\frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i}W_{n,ij}'\mathbb{E}\left[\left(G_{n,ij}-\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]\right)^2\middle|X,p_n\right]W_{n,ij}\\
&+\frac{1}{n^2(n-1)^2}\sum_i\sum_{j\neq i}\sum_{k\neq i,j}W_{n,ij}'\mathbb{E}\left[\left(G_{n,ij}-\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]\right)\right.\\
&\qquad\left.\cdot\left(G_{n,ik}-\mathbb{E}\left[G_{n,ik}\middle|X,p_n\right]\right)\middle|X,p_n\right]W_{n,ik}
\end{aligned}$$

Since

$$\mathbb{E}\left[\left(G_{n,ij}-\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]\right)^2\middle|X,p_n\right]\leq 1$$

and

$$\left|\mathbb{E}\left[\left(G_{n,ij}-\mathbb{E}\left[G_{n,ij}\middle|X,p_n\right]\right)\left(G_{n,ik}-\mathbb{E}\left[G_{n,ik}\middle|X,p_n\right]\right)\middle|X,p_n\right]\right|\leq 1$$

the conditional variance is bounded by

$$\frac{1}{n(n-1)}\max_{i,j=1,\ldots,n}\|W_{n,ij}\|^2+\frac{n-2}{n(n-1)}\max_{i,j,k=1,\ldots,n}\|W_{n,ij}\|\,\|W_{n,ik}\|=o(1)$$

by Assumption 4(iii), so

$$\frac{1}{n}\sum_{i} Y_{ni} = o_p(1).$$

Combining the results we obtain

$$\sup_{\theta}\left\|\hat{\Psi}_n(\theta, p_n) - \Psi_n(\theta, p_n)\right\| = o_p(1)$$

as $n \to \infty$. ■

Lemma S.4 (Continuity of CCP) *Suppose that Assumptions 1-3 are satisfied. Given X, the conditional choice probability $P_{n,ij}(\theta, p)$ is continuous in θ and p.*

Proof. Recall that

$$\begin{aligned}
&P_{n,ij}(\theta, p) \\
&= \int 1\left\{U_{n,ij}(\theta, p) + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni}(\theta, p)\Lambda_{ni}(\theta, p)\omega_{ni}(\varepsilon_i, \theta, p) \geq \varepsilon_{ij}\right\} f_{\varepsilon_i}(\varepsilon_i; \theta)\, d\varepsilon_i,
\end{aligned}$$

where f_{ε_i} represents the density of ε_i. By (2.5), (2.6), (3.3) and Assumption 1, $U_{n,ij}(\theta, p)$ and $f_{\varepsilon_i}(\varepsilon_i; \theta)$ are continuous in θ and p. The challenge is that $\omega_{ni}(\varepsilon_i, \theta, p)$ is a function of ε_i and it depends on θ and p. To establish the continuity of $P_{n,ij}(\theta, p)$, we need to investigate how $\omega_{ni}(\varepsilon_i, \theta, p)$ varies with θ and p.

In Corollary 3.3, we show that $\omega_{ni}(\varepsilon_i, \theta, p)$ satisfies

$$\Phi_{ni}(\theta, p)\Lambda_{ni}(\theta, p)\omega_{ni}(\varepsilon_i, \theta, p) = \frac{1}{n-1}\sum_{k\neq i} G_{n,ik}(\varepsilon_i, \theta, p) V_{ni}(\theta, p) Z_k, \text{ a.s.},$$

where $G_{ni}(\varepsilon_i, \theta, p) = (G_{n,ij}(\varepsilon_i, \theta, p), j \neq i) \in \{0,1\}^{n-1}$ is the optimal decision given in Theorem 3.2. Therefore $P_{n,ij}(\theta, p)$ can be expressed as

$$\begin{aligned}
&P_{n,ij}(\theta, p) \\
&= \int 1\left\{U_{n,ij}(\theta, p) + \frac{2}{n-2}\sum_{k\neq i} G_{n,ik}(\varepsilon_i, \theta, p) Z_j' V_{ni}(\theta, p) Z_k \geq \varepsilon_{ij}\right\} f_{\varepsilon_i}(\varepsilon_i; \theta)\, d\varepsilon_i.
\end{aligned}$$

From Corollary 3.3, the optimal decision $G_{ni}(\varepsilon_i, \theta, p) = g_{ni}$ for some $g_{ni} \in \{0,1\}^{n-1}$ if and only if $\varepsilon_i \in \mathcal{E}_i(g_{ni}, \theta, p)$, where the set $\mathcal{E}_i(g_{ni}, \theta, p)$ is defined in (S.17)

$$\mathcal{E}_i(g_{ni}, \theta, p) = \left\{\varepsilon_i \in \mathbb{R}^{n-1} : g_{ni} \text{ satisfies both (S.13) and (S.16)}\right\}.$$

For any $g_{ni} \in \{0,1\}^{n-1}$, the equations in (S.13) define an orthant in $\mathbb{R}^{n-1}$

$$\varepsilon_{ij} \begin{cases} < U_{n,ij}(\theta, p) + \frac{2}{n-2} \sum_{k \neq i} g_{n,ik} Z_j' V_{ni}(\theta, p) Z_k & \text{if } g_{n,ij} = 1 \\ \geq U_{n,ij}(\theta, p) + \frac{2}{n-2} \sum_{k \neq i} g_{n,ik} Z_j' V_{ni}(\theta, p) Z_k & \text{if } g_{n,ij} = 0 \end{cases}, \ \forall j \neq i. \tag{S.25}$$

Both $U_{n,ij}(\theta, p)$ and $V_{ni}(\theta, p)$ are continuous in θ and p, so the boundary of this orthant is continuous in θ and p. Moreover, the inequalities in (S.16) define half-spaces in $\mathbb{R}^{n-1}$ given by the hyperplanes

$$\begin{aligned} &\sum_{j \neq i} \left(g_{n,ij} - g_{n,ij}^l\right) \varepsilon_{ij} \\ \leq &\sum_{j \neq i} \left(g_{n,ij} - g_{n,ij}^l\right) U_{n,ij}(\theta, p) + \frac{1}{n-2} \sum_{k \neq i} \left(g_{n.ik}^l - g_{n,ik}^l\right) Z_j' V_{ni}(\theta, p) Z_k, \end{aligned} \tag{S.26}$$

for all g_{ni}^l that solve (S.13) with probability 1. While the set of solutions to (S.13) for a given ε_i could be discontinuous in θ and p (i.e., some link choices in an optimal g_{ni}^l may switch from 0 to 1 or the opposite as θ or p changes), this occurs with probability zero because ε_i follows a continuous distribution by Assumption 1(i). Since the right-hand side of (S.26) is continuous in θ and p, the boundaries of such half-spaces are also continuous in θ and p.

The set $\mathcal{E}_i(g_{ni}, \theta, p)$ is the intersection of the orthant in (S.25) and the half-spaces defined by (S.26). Because continuity is preserved under max and min operations, if two sets have boundaries that are continuous in θ and p, their intersection must also have a boundary that is continuous in θ and p. Therefore, the set $\mathcal{E}_i(g_{ni}, \theta, p)$ has a boundary that is continuous in θ and p.

Partitioning the space of ε_i into a collection of the sets $\mathcal{E}_i(g_{ni}, \theta, p)$ for all $g_{ni} \in \{0,1\}^{n-1}$, we can write $P_{n,ij}(\theta, p)$ as

$$\begin{aligned}
&P_{n,ij}(\theta, p) \\
&= \sum_{\substack{g_{ni} \in \{0,1\}^{n-1} \\ g_{n,ij}=1}} \int_{\mathcal{E}_i(g_{ni},\theta,p)} 1\left\{U_{n,ij}(\theta, p) + \frac{2}{n-2}\sum_{k\neq i} g_{n,ik} Z_j' V_{ni}(\theta, p) Z_k \geq \varepsilon_{ij}\right\} f_{\varepsilon_i}(\varepsilon_i; \theta)\, d\varepsilon_i \\
&= \sum_{\substack{g_{ni} \in \{0,1\}^{n-1} \\ g_{n,ij}=1}} \int_{\mathcal{E}_i(g_{ni},\theta,p)} f_{\varepsilon_i}(\varepsilon_i; \theta)\, d\varepsilon_i. \qquad \text{(S.27)}
\end{aligned}$$

For each g_{ni}, the set $\mathcal{E}_i(g_{ni}; \theta, p)$ has a boundary that is continuous in θ and p, so each integral in the summation in (S.27) is continuous in θ and p by Assumption 1(i). The proof is complete. ■

S.2.2 Asymptotic Distribution

In this section, we prove that the asymptotic distribution of $\hat{\theta}_n$ is as in Theorem 4.2. We first derive the asymptotic properties of ω_{ni} in a sequence of lemmas. Then we use these lemmas to prove Theorem 4.2.

Asymptotic Properties of $\omega_{ni}(\varepsilon_i)$ In the derivation of the asymptotic properties of $\omega_{ni}(\varepsilon_i)$ we suppress the dependence on θ_0 and p_n to simplify the notation. Recall that $\omega_{ni}(\varepsilon_i)$ maximizes $\Pi_{ni}(\omega, \varepsilon_i)$

$$\omega_{ni}(\varepsilon_i) = \arg\max_{\omega \in \mathbb{R}^T} \Pi_{ni}(\omega, \varepsilon_i),$$

where

$$\Pi_{ni}(\omega, \varepsilon_i) = \sum_{j\neq i}\left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij}\right]_+ - \frac{(n-1)^2}{n-2}\omega' \Lambda_{ni} \omega.$$

Let $\Pi^*_{ni}(\omega)$ denote the conditional expectation of $\Pi_{ni}(\omega, \varepsilon_i)$ given X and p_n

$$\Pi^*_{ni}(\omega) = \sum_{j\neq i} \mathbb{E}\left[\left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij}\right]_+ \middle| X, p_n\right] - \frac{(n-1)^2}{n-2}\omega' \Lambda_{ni} \omega$$

and ω_{ni}^{*} is a maximizer of $\Pi_{ni}^{*}(\omega)$

$$\omega_{ni}^{*} = \arg\max_{\omega \in \mathbb{R}^{T}} \Pi_{ni}^{*}(\omega).$$

In the subsequent lemmas, we establish that $\omega_{ni}(\varepsilon_i)$ is consistent for ω_{ni}^{*} (Lemma S.5). Moreover, $\omega_{ni}(\varepsilon_i)$ has an asymptotically linear representation (Lemma S.7) and satisfies certain uniformity properties (Lemma S.8). Additional results that are needed to prove these lemmas are in Lemma S.6 and S.9.

Remark S.1 *By Lemma S.1 we have*

$$\Lambda_{ni}\omega_{ni}(\varepsilon_i) = \frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega_{ni}(\varepsilon_i) - \varepsilon_{ij} \geq 0\right\}\Lambda_{ni}\Phi_{ni}'Z_j$$

almost surely. We set $\omega_{ni,t}(\varepsilon_i) = 0$ *if* $\lambda_{ni,t} = 0$, $t = 1, \ldots, T$, *so*

$$\omega_{ni}(\varepsilon_i) = \frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega_{ni}(\varepsilon_i) - \varepsilon_{ij} \geq 0\right\}\Lambda_{ni}^{+}\Lambda_{ni}\Phi_{ni}'Z_j$$

where Λ_{ni}^{+} *is the generalized inverse of* Λ_{ni}. *Then*

$$\|\omega_{ni}(\varepsilon_i)\| \leq \max_{j\neq i}\left\|\Lambda_{ni}^{+}\Lambda_{ni}\Phi_{ni}'Z_j\right\| \leq \max_{j\neq i}\left\|\Lambda_{ni}^{+}\Lambda_{ni}\right\|\left\|\Phi_{ni}'\right\|\left\|Z_j\right\| \leq T < \infty$$

Therefore $\omega_{ni}(\varepsilon_i)$ *is bounded, and without loss of generality we can assume that* ω *lies in a compact set* $\Omega \subseteq \mathbb{R}^{T}$ *as in Assumption 5(i).*

Lemma S.5 (Consistency of ω_{ni}) *Suppose that Assumptions 1-3 and 5(i)-(ii) are satisfied. For* $i = 1, \ldots, n$, $\omega_{ni}(\varepsilon_i)$ *is consistent for* ω_{ni}^{*}, *i.e., for any* $\delta > 0$

$$\Pr\left(\left\|\omega_{ni}(\varepsilon_i) - \omega_{ni}^{*}\right\| > \delta \middle| X, p_n\right) \to 0$$

as $n \to \infty$.

Proof. We follow the proof in Newey and McFadden (1994). Fix $\delta > 0$. Let $\mathcal{B}_\delta(\omega_{ni}^{*}) = \{\omega \in \Omega : \|\omega - \omega_{ni}^{*}\| < \delta\}$ be an open δ-ball centered at ω_{ni}^{*}. If $\Pi_{ni}^{*}(\omega_{ni}(\varepsilon_i)) > \sup_{\omega \in \Omega\setminus\mathcal{B}_\delta(\omega_{ni}^{*})}\Pi_{ni}^{*}(\omega)$, $\omega_{ni}(\varepsilon_i) \notin \Omega\setminus\mathcal{B}_\delta(\omega_{ni}^{*})$, or equivalently, $\omega_{ni}(\varepsilon_i) \in \mathcal{B}_\delta(\omega_{ni}^{*})$. Therefore,

$$
\begin{aligned}
&\Pr\left(\left\|\omega_{ni}\left(\varepsilon_{i}\right)-\omega_{ni}^{*}\right\|<\delta\right| X, p_{n}) \\
&\geq \Pr\left(\Pi_{ni}^{*}\left(\omega_{ni}\left(\varepsilon_{i}\right)\right)>\sup_{\omega\in\Omega\backslash\mathcal{B}_{\delta}\left(\omega_{ni}^{*}\right)}\Pi_{ni}^{*}\left(\omega\right)\right| X, p_{n}\Bigg) \\
&=\Pr\left(\Pi_{ni}^{*}\left(\omega_{ni}^{*}\right)-\Pi_{ni}^{*}\left(\omega_{ni}\left(\varepsilon_{i}\right)\right)<\Pi_{ni}^{*}\left(\omega_{ni}^{*}\right)-\sup_{\omega\in\Omega\backslash\mathcal{B}_{\delta}\left(\omega_{ni}^{*}\right)}\Pi_{ni}^{*}\left(\omega\right)\right| X, p_{n}\Bigg). \quad \text{(S.28)}
\end{aligned}
$$

By Assumption 5(i)-(ii)

$$
\frac{1}{n-1}\left(\Pi_{ni}^{*}\left(\omega_{ni}^{*}\right)-\sup_{\omega\in\Omega\backslash\mathcal{B}_{\delta}\left(\omega_{ni}^{*}\right)}\Pi_{ni}^{*}\left(\omega\right)\right)>0,
$$

so the right-hand size of (S.28) goes to 1 if

$$
\frac{1}{n-1}\left(\Pi_{ni}^{*}\left(\omega_{ni}^{*}\right)-\Pi_{ni}^{*}\left(\omega_{ni}\left(\varepsilon_{i}\right)\right)\right)\leq o_{p}\left(1\right). \quad \text{(S.29)}
$$

By the optimality of $\omega_{ni}\left(\varepsilon_{i}\right)$ we have

$$
\begin{aligned}
0\leq\Pi_{ni}^{*}\left(\omega_{ni}^{*}\right)-\Pi_{ni}^{*}\left(\omega_{ni}\left(\varepsilon_{i}\right)\right)=&\Pi_{ni}^{*}\left(\omega_{ni}^{*}\right)-\Pi_{ni}\left(\omega_{ni}^{*},\varepsilon_{i}\right)+\Pi_{ni}\left(\omega_{ni}^{*},\varepsilon_{i}\right)-\Pi_{ni}^{*}\left(\omega_{ni}\left(\varepsilon_{i}\right)\right) \\
\leq&\Pi_{ni}^{*}\left(\omega_{ni}^{*}\right)-\Pi_{ni}\left(\omega_{ni}^{*},\varepsilon_{i}\right)+\Pi_{ni}\left(\omega_{ni}\left(\varepsilon_{i}\right),\varepsilon_{i}\right)-\Pi_{ni}^{*}\left(\omega_{ni}\left(\varepsilon_{i}\right)\right) \\
\leq&2\sup_{\omega\in\Omega}\left|\Pi_{ni}\left(\omega,\varepsilon_{i}\right)-\Pi_{ni}^{*}\left(\omega\right)\right|.
\end{aligned}
$$

By the uniform LLN for $\Pi_{ni}\left(\omega,\varepsilon_{i}\right)$ in Lemma S.6,

$$
\sup_{\omega\in\Omega}\frac{1}{n-1}\left|\Pi_{ni}\left(\omega,\varepsilon_{i}\right)-\Pi_{ni}^{*}\left(\omega\right)\right|=o_{p}\left(1\right).
$$

so (S.29) holds and the consistency is proved. ■

Lemma S.6 (Uniform LLN for Π_{ni}) *Suppose that Assumptions 1-3 and 5 are satisfied. Then for any $\delta>0$,*

$$
\Pr\left(\sup_{\omega\in\Omega}\frac{1}{n-1}\left|\left(\Pi_{ni}\left(\omega,\varepsilon_{i}\right)-\Pi_{ni}^{*}\left(\omega\right)\right)\right|>\delta\right| X, p_{n}\Bigg)\to 0 \quad \text{(S.30)}
$$

as $n \to \infty$.

Proof. Recall that

$$\Pi_{ni}(\omega, \varepsilon_i) = \sum_{j \neq i} \left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \right]_+ - \frac{(n-1)^2}{n-2} \omega' \Lambda_{ni} \omega$$

and

$$\Pi_{ni}^*(\omega) = \sum_{j \neq i} \mathbb{E}\left(\left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \right]_+ \middle| X, p_n \right) - \frac{(n-1)^2}{n-2} \omega' \Lambda_{ni} \omega.$$

Define

$$\pi_{n,ij}(\omega, \varepsilon_{ij}) = \left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \right]_+ .$$

Hence,

$$\frac{1}{n-1} \left(\Pi_{ni}(\omega, \varepsilon_i) - \Pi_{ni}^*(\omega) \right) = \frac{1}{n-1} \sum_{j \neq i} \left(\pi_{n,ij}(\omega, \varepsilon_{ij}) - \mathbb{E}\left[\pi_{n,ij}(\omega, \varepsilon_{ij}) \middle| X, p_n \right] \right).$$

By Assumption 5(i)

$$\left| Z_j' \Phi_{ni} \Lambda_{ni} \omega \right| \leq \| \Phi_{ni} \| \, \| \Lambda_{ni} \| \sup_{\omega \in \Omega} \| \omega \| \leq \sqrt{T} \max_{t=1,\ldots,T} \lambda_{ni,t} \sup_{\omega \in \Omega} \| \omega \| \leq M < \infty$$

Therefore for all $\omega \in \Omega$

$$\pi_{n,ij}(\omega, \varepsilon_{ij})^2 \leq \left(U_{n,ij} + \frac{2(n-1)}{n-2} M - \varepsilon_{ij} \right)^2$$

with

$$\mathbb{E}\left[\left(U_{n,ij} + \frac{2(n-1)}{n-2} M - \varepsilon_{ij} \right)^2 \middle| X, p_n \right] < \infty$$

Also $\pi_{n,ij}(\omega, \varepsilon_{ij})$ is continuous in ω on a compact set Ω. Therefore the conditions of the uniform LLN for triangular arrays are satisfied (Jennrich, 1969) and (S.30) follows. ■

Lemma S.7 (Asymptotically linear representation of $\omega_{ni}(\varepsilon_i)$) *Suppose that Assumptions 1-3 and 5 are satisfied. For each $i = 1, \ldots, n$, $\omega_{ni}(\varepsilon_i)$ has an asymptotically* *linear representation*

$$\omega_{ni}(\varepsilon_i) - \omega_{ni}^* = \frac{1}{n-1}\sum_{j\neq i}\varphi_{n,ij}^{\omega}(\omega_{ni}^*, \varepsilon_{ij}) + r_{ni}^{\omega}(\varepsilon_i) \tag{S.31}$$

as $n \to \infty$*, with the influence function* $\varphi_{n,ij}^{\omega}(\omega_{ni}^*, \varepsilon_{ij}) \in \mathbb{R}^T$ *given by*

$$\varphi_{n,ij}^{\omega}(\omega_{ni}^*, \varepsilon_{ij}) = -\nabla_{\omega'}\Gamma_{ni}^*(\omega_{ni}^*)^{+}\varphi_{n,ij}^{\pi}(\omega_{ni}^*, \varepsilon_{ij}), \tag{S.32}$$

where the function $\varphi_{n,ij}^{\pi}(\omega, \varepsilon_{ij}) \in \mathbb{R}^T$ *is defined by*

$$\varphi_{n,ij}^{\pi}(\omega, \varepsilon_{ij}) = 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega - \varepsilon_{ij} \geq 0\right\}\Lambda_{ni}\Phi_{ni}'Z_j - \Lambda_{ni}\omega, \tag{S.33}$$

and

$$\nabla_{\omega'}\Gamma_{ni}^*(\omega_{ni}^*) = \frac{2}{n-2}\sum_{j\neq i} f_{\varepsilon}\left(U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega_{ni}^*\right)\Lambda_{ni}\Phi_{ni}'Z_jZ_j'\Phi_{ni}\Lambda_{ni} - \Lambda_{ni}, \tag{S.34}$$

which by Assumption 5(iii) has the generalized inverse

$$\begin{aligned}&\nabla_{\omega'}\Gamma_{ni}^*(\omega_{ni}^*)^{+}\\ =&\Lambda_{ni}^{+}\left(\frac{2}{n-2}\sum_{j\neq i} f_{\varepsilon}\left(U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega_{ni}^*\right)\Lambda_{ni}\Phi_{ni}'Z_jZ_j'\Phi_{ni} - I_T\right)^{-1}.\end{aligned}$$

Moreover, the remainder $r_{ni}^{\omega}(\varepsilon_i)$ *in (S.31) satisfies*

$$r_{ni}^{\omega}(\varepsilon_i) = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{S.35}$$

Proof. Define $\Gamma_{ni}(\omega, \varepsilon_i) \in \mathbb{R}^T$

$$\begin{aligned}\Gamma_{ni}(\omega, \varepsilon_i) &= \frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega - \varepsilon_{ij} \geq 0\right\}\Lambda_{ni}\Phi_{ni}'Z_j - \Lambda_{ni}\omega\\ &= \frac{1}{n-1}\sum_{j\neq i}\varphi_{n,ij}^{\pi}(\omega, \varepsilon_{ij}),\end{aligned}$$

where $\varphi^{\pi}_{n,ij}(\omega, \varepsilon_{ij}) \in \mathbb{R}^T$ is defined in (S.33). By Lemma S.1 $\omega_{ni}(\varepsilon_i)$ satisfies the first-order condition

$$\Gamma_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) = 0, \text{ a.s.} \tag{S.36}$$

Let $\Gamma^*_{ni}(\omega) \in \mathbb{R}^T$ be the conditional expectation of $\Gamma_{ni}(\omega, \varepsilon_i)$

$$\Gamma^*_{ni}(\omega) = \mathbb{E}\left[\Gamma_{ni}(\omega, \varepsilon_i) \middle| X, p_n\right] = \frac{1}{n-1} \sum_{j \neq i} \mathbb{E}\left[\varphi^{\pi}_{n,ij}(\omega, \varepsilon_{ij}) \middle| X, p_n\right],$$

where

$$\mathbb{E}\left[\varphi^{\pi}_{n,ij}(\omega, \varepsilon_{ij}) \middle| X, p_n\right] = F_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z'_j \Phi_{ni} \Lambda_{ni} \omega\right) \Lambda_{ni} \Phi'_{ni} Z_j - \Lambda_{ni}\omega.$$

By Assumption 5(ii), $\Pi^*_{ni}(\omega)$ is maximized at ω^*_{ni}, so ω^*_{ni} satisfies the first-order condition

$$\Gamma^*_{ni}(\omega^*_{ni}) = 0. \tag{S.37}$$

By a Taylor expansion of $\Gamma^*_{ni}(\omega)$ at ω^*_{ni} and the consistency of $\omega_{ni}(\varepsilon_i)$ in Lemma S.5, we have

$$\Gamma^*_{ni}(\omega_{ni}(\varepsilon_i)) = \nabla_{\omega'} \Gamma^*_{ni}(\omega^*_{ni})(\omega_{ni}(\varepsilon_i) - \omega^*_{ni}) + O_p\left(\|\omega_{ni}(\varepsilon_i) - \omega^*_{ni}\|^2\right), \tag{S.38}$$

where $\nabla_{\omega'} \Gamma^*_{ni}(\omega^*_{ni})$ is the Jacobian matrix of $\Gamma^*_{ni}(\omega)$ at ω^*_{ni} defined in (S.34) that we rewrite as

$$\nabla_{\omega'} \Gamma^*_{ni}(\omega^*_{ni}) = H_{ni}(\omega^*_{ni}) \Lambda_{ni}$$

with

$$H_{ni}(\omega^*_{ni}) = \frac{2}{n-2} \sum_{j \neq i} f_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z'_j \Phi_{ni} \Lambda_{ni} \omega^*_{ni}\right) \Lambda_{ni} \Phi'_{ni} Z_j Z'_j \Phi_{ni} - I_T.$$

By Assumption 5(iii), $H_{ni}(\omega^*_{ni})$ is nonsingular. There exists a constant $c > 0$ such that

$$\|\nabla_{\omega'} \Gamma^*_{ni}(\omega^*_{ni})(\omega - \omega^*_{ni})\| \geq c \|\omega - \omega^*_{ni}\|$$

for every ω. This is because

$$\begin{aligned}
\left\|\nabla_{\omega'}\Gamma^*_{ni}\left(\omega^*_{ni}\right)\left(\omega-\omega^*_{ni}\right)\right\|^2 &= \left(\omega-\omega^*_{ni}\right)'\left(\nabla_{\omega'}\Gamma^*_{ni}\left(\omega^*_{ni}\right)\right)'\nabla_{\omega'}\Gamma^*_{ni}\left(\omega^*_{ni}\right)\left(\omega-\omega^*_{ni}\right)\\
&= \left(\omega-\omega^*_{ni}\right)'\Lambda_{ni}H_{ni}\left(\omega^*_{ni}\right)'H_{ni}\left(\omega^*_{ni}\right)\Lambda_{ni}\left(\omega-\omega^*_{ni}\right)\\
&\geq \lambda_{\min}\left(H_{ni}\left(\omega^*_{ni}\right)'H_{ni}\left(\omega^*_{ni}\right)\right)\left(\omega-\omega^*_{ni}\right)'\Lambda^2_{ni}\left(\omega-\omega^*_{ni}\right)\\
&\geq \lambda_{\min}\left(H_{ni}\left(\omega^*_{ni}\right)'H_{ni}\left(\omega^*_{ni}\right)\right)\lambda_{\min}\left(V'_{ni}V_{ni}\right)\left\|\omega-\omega^*_{ni}\right\|^2,
\end{aligned}$$

where $\lambda_{\min}\left(H_{ni}\left(\omega^*_{ni}\right)'H_{ni}\left(\omega^*_{ni}\right)\right)$ is the smallest eigenvalue of $H_{ni}\left(\omega^*_{ni}\right)'H_{ni}\left(\omega^*_{ni}\right)$, which is positive because $H_{ni}\left(\omega^*_{ni}\right)$ is nonsingular, and $\lambda_{\min}\left(V'_{ni}V_{ni}\right)$ is the smallest among the eigenvalues of $V'_{ni}V_{ni}$ that are not zero, which is also positive. Combining this with the Taylor expansion of $\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)$, we obtain

$$\left\|\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)\right\| \geq \left\|\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}\right\|\left(c+o_p\left(1\right)\right). \tag{S.39}$$

By (S.36) and (S.37), we can write $\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)$ as

$$\begin{aligned}
&\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)\\
=&-\Gamma_{ni}\left(\omega^*_{ni},\varepsilon_i\right)-\left(\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)\right)-\left(\Gamma_{ni}\left(\omega^*_{ni},\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega^*_{ni}\right)\right)\right),\ \text{a.s.}
\end{aligned} \tag{S.40}$$

We apply the Lindeberg-Feller CLT to show that the first term on the right-hand side satisfies

$$\Gamma_{ni}\left(\omega^*_{ni},\varepsilon_i\right)=O_p\left(\frac{1}{\sqrt{n}}\right). \tag{S.41}$$

To verify the Lindeberg condition, define the mean 0 random vector

$$Y^{\gamma}_{n,ij}=\frac{1}{\sqrt{n-1}}1\left\{U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z'_j\Phi_{ni}\Lambda_{ni}\omega^*_{ni}-\varepsilon_{ij}\geq 0\right\}\Lambda_{ni}\Phi'_{ni}Z_j-\Lambda_{ni}\omega^*_{ni},$$

so that

$$\Gamma_{ni}\left(\omega^*_{ni},\varepsilon_i\right)=\frac{1}{\sqrt{n-1}}\sum_{j\neq i}Y^{\gamma}_{n,ij}.$$

By the Cramer-Wold device it suffices to show that $a'\sum_{j\neq i}Y^{\gamma}_{n,ij}$ satisfies the Lindeberg condition for any $T\times 1$ vector of constants a. The Lindeberg condition is that for any $\xi > 0$

$$\lim_{n\to\infty} \frac{1}{a'\Sigma_{ni}^{\gamma}a} \sum_{j\neq i} \mathbb{E}\left[\left(a'Y_{n,ij}^{\gamma}\right)^2 1\left\{ \left|a'Y_{n,ij}^{\gamma}\right| \geq \xi\sqrt{a'\Sigma_{ni}^{\gamma}a} \right\} \middle| X, p_n \right] = 0,$$

with

$$\begin{aligned}
\Sigma_{ni}^{\gamma} &= \sum_{j\neq i} Var\left(Y_{n,ij}^{\gamma}\middle| X, p_n\right) \\
&= \frac{1}{n-1}\sum_{j\neq i} F_{\varepsilon}\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j'\Phi_{ni}\Lambda_{ni}\omega_{ni}^{*}\right) \\
&\qquad \cdot \left(1 - F_{\varepsilon}\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j'\Phi_{ni}\Lambda_{ni}\omega_{ni}^{*}\right)\right) \Lambda_{ni}\Phi_{ni}' Z_j Z_j'\Phi_{ni}\Lambda_{ni}.
\end{aligned}$$

We have

$$\begin{aligned}
&\sum_{j\neq i} \mathbb{E}\left[\left(a'Y_{n,ij}^{\gamma}\right)^2 1\left\{ \left|a'Y_{n,ij}^{\gamma}\right| \geq \xi\sqrt{a'\Sigma_{ni}^{\gamma}a} \right\} \middle| X, p_n \right] \\
&\leq \mathbb{E}\left[\sum_{j\neq i} \left(a'Y_{n,ij}^{\gamma}\right)^2 1\left\{ \frac{\max_{j\neq i}\left|a'Y_{n,ij}^{\gamma}\right|}{\sqrt{a'\Sigma_{ni}^{\gamma}a}} \geq \xi \right\} \middle| X, p_n \right].
\end{aligned}$$

Note that $\sum_{j\neq i}\left(a'Y_{n,ij}^{\gamma}\right)^2$ has a finite expectation and is therefore $O_p(1)$. Hence if

$$\frac{\max_{j\neq i}\left|a'Y_{n,ij}^{\gamma}\right|}{\sqrt{a'\Sigma_{ni}^{\gamma}a}} = o_p(1), \tag{S.42}$$

then

$$\sum_{j\neq i} \left(a'Y_{n,ij}^{\gamma}\right)^2 1\left\{ \frac{\max_{j\neq i}\left|a'Y_{n,ij}^{\gamma}\right|}{\sqrt{a'\Sigma_{ni}^{\gamma}a}} \geq \xi \right\} = O_p(1)\, o_p(1) = o_p(1).$$

Finally, this random variable is bounded by $\sum_{j\neq i}\left(a'Y_{n,ij}^{\gamma}\right)^2$ that has a finite expectation. We conclude that by dominated convergence the Lindeberg condition is satisfied if (S.42) holds.

By Chebyshev's inequality,

$$\Pr\left(\frac{\max_{j\neq i}\left|a'Y_{n,ij}^{\gamma}\right|}{\sqrt{a'\Sigma_{ni}^{\gamma}a}} \geq \xi \middle| X, p_n \right) \leq \frac{1}{\xi^2 a'\Sigma_{ni}^{\gamma}a} \mathbb{E}\left[\max_{j\neq i} \left(a'Y_{n,ij}^{\gamma}\right)^2 \middle| X, p_n \right].$$

The random variable $a'Y_{n,ij}^{\gamma}$ has a support bounded by

$$\left|a'Y_{n,ij}^{\gamma}\right| \leq \frac{\|a\| \, \|\Lambda_{ni}\| \left(\sqrt{T} + \|\omega_{ni}^{*}\|\right)}{\sqrt{n-1}} \leq \frac{M_i}{\sqrt{n-1}}$$

with $M_i < \infty$. Let $\|Z\|_{\psi|X,p_n}$ be the conditional Orlicz norm of a random variable Z given X and p_n for the convex function $\psi(z) = e^z - 1$.[18] Then $\mathbb{E}\left[|Z|\, X, p_n\right] \leq \|Z\|_{\psi|X,p_n}$[19] so that

$$\mathbb{E}\left[\max_{j\neq i}\left(a'Y_{n,ij}^{\gamma}\right)^2 \middle| X, p_n\right] \leq \left\|\max_{j\neq i}\left(a'Y_{n,ij}^{\gamma}\right)^2\right\|_{\psi|X,p_n}.$$

By the maximal inequality in Lemma 2.2.2 in van der Vaart and Wellner (1996) we have the bound

$$\left\|\max_{j\neq i}\left(a'Y_{n,ij}^{\gamma}\right)^2\right\|_{\psi|X,p_n} \leq K\ln(n+1)\max_{j\neq i}\left\|\left(a'Y_{n,ij}^{\gamma}\right)^2\right\|_{\psi|X,p_n}.$$

By the Hoeffding's inequality for bounded random variables (Boucheron, Lugosi, and Massart, 2013, Theorem 2.8)

$$\begin{aligned}\Pr\left(\left(a'Y_{n,ij}^{\gamma}\right)^2 \geq t \middle| X, p_n\right) &= \Pr\left(a'Y_{n,ij}^{\gamma} \geq \sqrt{t}\middle| X, p_n\right) + \Pr\left(-a'Y_{n,ij}^{\gamma} \geq \sqrt{t}\middle| X, p_n\right) \\ &\leq 2\exp\left(-\frac{(n-1)\,t}{2M_i^2}\right)\end{aligned}$$

so that by Lemma 2.2.1 in van der Vaart and Wellner (1996)

$$\left\|\left(a'Y_{n,ij}^{\gamma}\right)^2\right\|_{\psi|X,p_n} \leq \frac{6M_i^2}{n-1}.$$

Combining these results

$$\frac{1}{\xi^2 a'\Sigma_{ni}^{\gamma}a}\mathbb{E}\left[\max_{j\neq i}\left(a'Y_{n,ij}^{\gamma}\right)^2\middle| X, p_n\right] \leq \frac{1}{\xi^2 a'\Sigma_{ni}^{\gamma}a}\frac{6K\ln(n+1)\,M_i^2}{n-1} = o(1)$$

[18] The conditional Orlicz norm is defined by $\|Z\|_{\psi|X,p_n} = \inf\left\{C>0 : \mathbb{E}\left(\psi\left(\frac{|Z|}{C}\right)\middle| X, p_n\right) \leq 1\right\}$.

[19] This is true because $z \leq \psi(z)$, we have $\mathbb{E}\left(\psi\left(\frac{|Z|}{\|Z\|_{\psi}}\right)\middle| X_n, p_n\right) \leq 1 \leq \mathbb{E}\left(\psi\left(\frac{|Z|}{\mathbb{E}(|Z||X_n,p_n)}\right)\middle| X_n, p_n\right)$.

so the Lindeberg condition holds.

As for the second term on the right-hand side of (S.40), note that

$$\Gamma_{ni}\left(\omega,\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega\right)=\frac{1}{n-1}\sum_{j\neq i}\left(\varphi^{\gamma}_{n,ij}\left(\omega,\varepsilon_{ij}\right)-\mathbb{E}\left[\varphi^{\gamma}_{n,ij}\left(\omega,\varepsilon_{ij}\right)\middle|X,p_n\right]\right)$$

with $\varphi^{\gamma}_{n,ij}\left(\omega,\varepsilon_{ij}\right)$ defined by

$$\varphi^{\gamma}_{n,ij}\left(\omega,\varepsilon_{ij}\right)=1\left\{U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega-\varepsilon_{ij}\geq 0\right\}\Lambda_{ni}\Phi_{ni}'Z_j. \tag{S.43}$$

Define the empirical process

$$\begin{aligned}\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega,\varepsilon_i\right)&=\sqrt{n-1}\left(\Gamma_{ni}\left(\omega,\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega\right)\right)\\&=\frac{1}{\sqrt{n-1}}\sum_{j\neq i}\left(\varphi^{\gamma}_{n,ij}\left(\omega,\varepsilon_{ij}\right)-\mathbb{E}\left[\varphi^{\gamma}_{n,ij}\left(\omega,\varepsilon_{ij}\right)\middle|X,p_n\right]\right),\ \omega\in\Omega\end{aligned} \tag{S.44}$$

so the second term on the right-hand side of (S.40) can be written as

$$\begin{aligned}&\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)-\left(\Gamma_{ni}\left(\omega^*_{ni},\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega^*_{ni}\right)\right)\\=&\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega^*_{ni},\varepsilon_i\right)\right).\end{aligned} \tag{S.45}$$

In Lemma S.9(i) we show that

$$\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega^*_{ni},\varepsilon_i\right)=o_p\left(1\right). \tag{S.46}$$

Hence, the second term on the right-hand side of (S.40) is $o_p\left(n^{-1/2}\right)$

$$\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)-\left(\Gamma_{ni}\left(\omega^*_{ni},\varepsilon_i\right)-\Gamma^*_{ni}\left(\omega^*_{ni}\right)\right)=o_p\left(\frac{1}{\sqrt{n}}\right). \tag{S.47}$$

Applying (S.39), (S.41) and (S.47) to (S.40) we obtain

$$\left\|\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}\right\|\left(c+o_p\left(1\right)\right)\leq O_p\left(\frac{1}{\sqrt{n}}\right)+o_p\left(\frac{1}{\sqrt{n}}\right). \tag{S.48}$$

This implies that

$$\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}=O_p\left(\frac{1}{\sqrt{n}}\right), \tag{S.49}$$

i.e., $\omega_{ni}(\varepsilon_i)$ converges to ω^*_{ni} at the rate of $n^{-\frac{1}{2}}$.

Combining (S.38), (S.40), and (S.47) yields

$$\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})(\omega_{ni}(\varepsilon_i)-\omega^*_{ni}) = -\Gamma_{ni}(\omega^*_{ni},\varepsilon_i) + o_p\left(\frac{1}{\sqrt{n}}\right).$$

By Assumption 5(iii), $\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})(\omega-\omega^*_{ni}) = \Lambda^+_{ni}\Lambda_{ni}(\omega-\omega^*_{ni}) = \omega - \omega^*_{ni}$. Multiplying both sides by $\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+$ we obtain

$$\begin{aligned}\omega_{ni}(\varepsilon_i)-\omega^*_{ni} &= -\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+\Gamma_{ni}(\omega^*_{ni},\varepsilon_i) + r^\omega_{ni}(\varepsilon_i)\\ &= -\frac{1}{n-1}\sum_{j\neq i}\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+\varphi^\pi_{n,ij}(\omega^*_{ni},\varepsilon_{ij}) + r^\omega_{ni}(\varepsilon_i) \qquad \text{(S.50)}\end{aligned}$$

with $r^\omega_{ni}(\varepsilon_i) = o_p\left(\frac{1}{\sqrt{n}}\right)$. The proof is complete. ■

Lemma S.8 (Uniform Properties of $\omega_{ni}(\varepsilon_i)$) *Suppose that Assumptions 1-3 and 5 are satisfied. Then (i) $\omega_{ni}(\varepsilon_i)$ satisfies*

$$\max_{1\le i\le n}\|\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\|^2 = o_p\left(\frac{1}{\sqrt{n}}\right).$$

(ii) The remainder $r^\omega_{ni}(\varepsilon_i)$ defined in Lemma S.7 satisfies

$$\max_{1\le i\le n}\|r^\omega_{ni}(\varepsilon_i)\| = o_p\left(\frac{1}{\sqrt{n}}\right).$$

Proof. Part (i): By Markov's inequality, for any $\delta > 0$,

$$\Pr\left(\max_{1\le i\le n}\sqrt{n}\|\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\|^2 > \delta\middle| X, p_n\right) \le \frac{\sqrt{n}}{\delta}\mathbb{E}\left[\max_{1\le i\le n}\|\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\|^2\middle| X, p_n\right].$$

Let $\|\cdot\|_{\psi|X,p_n}$ be the conditional Orlicz norm given X and p_n for the convex function $\psi(z) = e^z - 1$. By $\mathbb{E}[|Z||X,p_n] \le \|Z\|_{\psi|X,p_n}$ for any random variable Z and the maximal inequality in Lemma 2.2.2 in van der Vaart and Wellner (1996) we derive

$$\begin{aligned}\mathbb{E}\left[\max_{1\le i\le n}\|\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\|^2\middle| X, p_n\right] &\le \left\|\max_{1\le i\le n}\|\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\|^2\right\|_{\psi|X,p_n}\\ &\le K\ln(n+1)\max_{1\le i\le n}\left\|\|\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\|^2\right\|_{\psi|X,p_n},\end{aligned}$$

where K is a constant. Let $\|\cdot\|_{\psi_2|X,p_n}$ be the conditional Orlicz norm given X and p_n for the convex function $\psi_2(z) = e^{z^2} - 1$. For any random variable Z and constant $C > 0$, we have $\mathbb{E}\left[\psi\left(\frac{|Z|^2}{C^2}\right)\middle| X, p_n\right] = \mathbb{E}\left[\psi_2\left(\frac{|Z|}{C}\right)\middle| X, p_n\right]$, so $\|Z^2\|_{\psi|X,p_n} = \|Z\|^2_{\psi_2|X,p_n}$. Hence,

$$\max_{1\leq i\leq n}\left\|\|\omega_{ni}(\varepsilon_i) - \omega^*_{ni}\|^2\right\|_{\psi|X,p_n} = \max_{1\leq i\leq n}\|\omega_{ni}(\varepsilon_i) - \omega^*_{ni}\|^2_{\psi_2|X,p_n}.$$

From (S.38), (S.40), (S.45), and (S.50) in Lemma S.7, the remainder $r^{\omega}_{ni}(\varepsilon_i)$ is

$$\begin{aligned} r^{\omega}_{ni}(\varepsilon_i) =& \nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+ \\ & \cdot \left(O_p\left(\|\omega_{ni}(\varepsilon_i) - \omega^*_{ni}\|^2\right) + \frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\varphi^{\gamma}_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - \mathbb{G}_n\varphi^{\gamma}_{ni}(\omega^*_{ni}, \varepsilon_i)\right)\right). \end{aligned} \tag{S.51}$$

Therefore, from (S.50) and (S.51), we obtain

$$\begin{aligned} & \omega_{ni}(\varepsilon_i) - \omega^*_{ni} \\ =& \frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}(\omega^*_{ni}, \varepsilon_{ij}) + \nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+ \\ & \cdot \left(o_p\left(\|\omega_{ni}(\varepsilon_i) - \omega^*_{ni}\|\right) + \frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\varphi^{\gamma}_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - \mathbb{G}_n\varphi^{\gamma}_{ni}(\omega^*_{ni}, \varepsilon_i)\right)\right). \end{aligned}$$

By the triangle inequality for the Orlicz norm[20] and the boundedness of the inverse Jacobian $\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+$ we obtain

$$
\begin{aligned}
&\left\|\left(\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\right)(1+o_p(1))\right\|_{\psi_2|X,p_n}\\
\leq&\left\|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}\left(\omega^*_{ni},\varepsilon_{ij}\right)\right\|_{\psi_2|X,p_n}\\
&+\frac{1}{\sqrt{n-1}}\left\|\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+\right\|\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega_{ni}(\varepsilon_i),\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega^*_{ni},\varepsilon_i\right)\right\|_{\psi_2|X,p_n}. \quad \text{(S.52)}
\end{aligned}
$$

Note that $\left\|\left(\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\right)(1+o_p(1))\right\|_{\psi_2|X,p_n}=\left\|\left(\omega_{ni}(\varepsilon_i)-\omega^*_{ni}\right)\right\|_{\psi_2|X,p_n}(1+o(1))$.[21]

Consider the first term on the right-hand side. Recall that the influence function $\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})$ is a $T\times 1$ vector given by

$$
\begin{aligned}
&\varphi^{\omega}_{n,ij}\left(\omega^*_{ni},\varepsilon_{ij}\right)\\
=&-\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+\left(1\left\{U_{n,ij}+\frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega^*_{ni}-\varepsilon_{ij}\geq 0\right\}\Lambda_{ni}\Phi'_{ni}Z_j-\Lambda_{ni}\omega^*_{ni}\right).
\end{aligned}
$$

---

[20] Take two random variables X and Y. For any $\varepsilon>0$, there exist u and v such that $u<\|X\|_{\psi_2|X,p_n}+\varepsilon$, $v<\|Y\|_{\psi_2|X,p_n}+\varepsilon$, and $\max\left\{\mathbb{E}\left[\frac{|X|}{u}\middle| X,p_n\right],\mathbb{E}\left[\frac{|Y|}{v}\middle| X,p_n\right]\right\}\leq 1$. Because ψ_2 is non-decreasing and convex, we have

$$
\psi_2\left(\frac{|X+Y|}{u+v}\right)\leq\psi_2\left(\frac{|X|+|Y|}{u+v}\right)=\psi_2\left(\frac{u}{u+v}\frac{|X|}{u}+\frac{v}{u+v}\frac{|Y|}{v}\right)\leq\frac{u}{u+v}\psi_2\left(\frac{|X|}{u}\right)+\frac{v}{u+v}\psi_2\left(\frac{|Y|}{v}\right).
$$

Hence u and v satisfy $u+v<\|X\|_{\psi_2|X,p_n}+\|Y\|_{\psi_2|X,p_n}+2\varepsilon$ and $\mathbb{E}\left[\psi_2\left(\frac{|X+Y|}{u+v}\right)\middle| X,p_n\right]\leq 1$. By definition of the Orlicz norm $\|X+Y\|_{\psi_2|X,p_n}\leq u+v<\|X\|_{\psi_2|X,p_n}+\|Y\|_{\psi_2|X,p_n}+2\varepsilon$. This proves $\|X+Y\|_{\psi_2|X,p_n}\leq\|X\|_{\psi_2|X,p_n}+\|Y\|_{\psi_2|X,p_n}$.

[21] For any bounded random variable Z, we have $\|Zo_p(1)\|_{\psi_2|X,p_n}=o\left(\|Z\|_{\psi_2|X,p_n}\right)$. This is because for any sequence $\delta_n\downarrow 0$

$$
1<\mathbb{E}\left[\psi_2\left(\frac{|Zo_p(1)|}{\|Zo_p(1)\|_{\psi_2|X,p_n}-\delta_n}\right)\middle| X,p_n\right]=\mathbb{E}\left[\psi_2\left(\frac{|Zo_p(1)|}{\|Z\|_{\psi_2|X,p_n}}\cdot\frac{\|Z\|_{\psi_2|X,p_n}}{\|Zo_p(1)\|_{\psi_2|X,p_n}-\delta_n}\right)\middle| X,p_n\right]
$$

If there were $M<\infty$ such that $\frac{\|Z\|_{\psi_2|X,p_n}}{\|Zo_p(1)\|_{\psi_2|X,p_n}-\delta_n}\leq M$ for n sufficiently large, since $\frac{|Zo_p(1)|}{\|Z\|_{\psi_2|X,p_n}}\xrightarrow{p}0$, we have for sufficiently large n

$$
\mathbb{E}\left[\psi_2\left(\frac{|Zo_p(1)|}{\|Z\|_{\psi_2|X,p_n}}\cdot\frac{\|Z\|_{\psi_2|X,p_n}}{\|Zo_p(1)\|_{\psi_2|X,p_n}-\delta_n}\right)\middle| X,p_n\right]\leq\mathbb{E}\left[\psi_2\left(\frac{|Zo_p(1)|}{\|Z\|_{\psi_2|X,p_n}}\cdot M\right)\middle| X,p_n\right]\to 0
$$

by dominated convergence. Therefore, $\frac{\|Zo_p(1)\|_{\psi_2|X,p_n}-\delta_n}{\|Z\|_{\psi_2|X,p_n}}=o(1)$, so $\|Zo_p(1)\|_{\psi_2|X,p_n}=o\left(\|Z\|_{\psi_2|X,p_n}\right)$.

Let $\varphi^{\omega}_{n,ij,t}(\omega^*_{ni},\varepsilon_{ij})$ denote the t-th component of $\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})$, $t=1,\dots T$. Note that

$$\begin{aligned}\left\|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})\right\| &= \sqrt{\sum_{t=1}^{T}\left(\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij,t}(\omega^*_{ni},\varepsilon_{ij})\right)^2}\\ &\leq \sum_{t=1}^{T}\left|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij,t}(\omega^*_{ni},\varepsilon_{ij})\right|,\end{aligned}$$

so for any $x>0$,

$$\begin{aligned}&\Pr\left(\left\|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})\right\|>x\,\middle|\,X,p_n\right)\\ \leq&\Pr\left(\sum_{t=1}^{T}\left|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij,t}(\omega^*_{ni},\varepsilon_{ij})\right|>x\,\middle|\,X,p_n\right)\\ \leq&\sum_{t=1}^{T}\Pr\left(\left|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij,t}(\omega^*_{ni},\varepsilon_{ij})\right|>\frac{x}{T}\,\middle|\,X,p_n\right).\end{aligned}$$

It is clear that for any $t=1,\dots T$, and $i,j=1,\dots,n$,

$$\begin{aligned}&\left|\varphi^{\omega}_{n,ij,t}(\omega^*_{ni},\varepsilon_{ij})\right|<\left\|\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})\right\|\\ &\qquad\leq\left\|\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^{+}\right\|\left(\|\Lambda_{ni}\Phi'_{ni}Z_j\|+\|\Lambda_{ni}\|\,\|\omega^*_{ni}\|\right)\leq M_{n,ij}\leq M<\infty.\end{aligned}$$

By Hoeffding's inequality for bounded random variables (Boucheron, Lugosi, and Massart, 2013, Theorem 2.8) we have

$$\Pr\left(\left|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij,t}(\omega^*_{ni},\varepsilon_{ij})\right|>\frac{x}{T}\,\middle|\,X,p_n\right)\leq 2\exp\left(-\frac{(n-1)\,x^2}{2M^2T^2}\right),$$

so

$$\Pr\left(\left\|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})\right\|>x\,\middle|\,X,p_n\right)\leq 2T\exp\left(-\frac{(n-1)\,x^2}{2M^2T^2}\right).$$

Hence, by Lemma 2.2.1 in van der Vaart and Wellner (1996),

$$\left\|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})\right\|_{\psi_2|X,p_n}\leq\frac{\sqrt{2\,(2T+1)}TM}{\sqrt{n-1}}.$$

From (S.64) in the proof of Lemma S.9 we see that

$$\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}^*,\varepsilon_i\right)\right\|_{\psi|X,p_n}=o\left(1\right).$$

Following the proof for (S.64) and applying Theorems 2.14.5 and 2.14.1 in van der Vaart and Wellner (1996) for $p=2$ we can derive similarly

$$\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}^*,\varepsilon_i\right)\right\|_{\psi_2|X,p_n}=o\left(1\right),$$

so the second term on the right-hand side of (S.52) is $\frac{o(1)}{\sqrt{n-1}}$.

Combining the results yields

$$\begin{aligned}&\Pr\left(\max_{1\le i\le n}\sqrt{n}\left\|\omega_{ni}\left(\varepsilon_i\right)-\omega_{ni}^*\right\|^2>\delta\,\middle|\,X,p_n\right)\\ \le&\frac{K}{\delta}\sqrt{n}\ln\left(n+1\right)\left(\frac{\sqrt{2\left(2T+1\right)}TM}{\sqrt{n-1}}+\frac{o\left(1\right)}{\sqrt{n-1}}\right)^2\\ =&o\left(1\right).\end{aligned}$$

We conclude that $\max_{1\le i\le n}\left\|\omega_{ni}\left(\varepsilon_i\right)-\omega_{ni}^*\right\|^2=o_p\left(n^{-1/2}\right)$.

Part (ii): From (S.38), (S.40), (S.45), and (S.50) in Lemma S.7, the remainder $r_{ni}^{\omega}\left(\varepsilon_i\right)$ is given by

$$\begin{aligned}r_{ni}^{\omega}\left(\varepsilon_i\right)=&\nabla_{\omega'}\Gamma_{ni}^*\left(\omega_{ni}^*\right)^+\\ &\cdot\left(O_p\left(\left\|\omega_{ni}\left(\varepsilon_i\right)-\omega_{ni}^*\right\|^2\right)+\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}^*,\varepsilon_i\right)\right)\right).\end{aligned}$$

It is clear that $\max_{1\le i\le n}\left\|\nabla_{\omega'}\Gamma_{ni}^*\left(\omega_{ni}^*\right)^+\right\|\le M<\infty$. By Lemma S.9(ii)

$$\max_{1\le i\le n}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}^*,\varepsilon_i\right)\right\|=o_p\left(1\right),$$

so combining this with part (i) we obtain

$$
\begin{aligned}
&\max_{1\le i\le n}\left\|r^{\omega}_{ni}\left(\varepsilon_i\right)\right\| \\
\le &\max_{1\le i\le n}\left\|\nabla_{\omega'}\Gamma^*_{ni}\left(\omega^*_{ni}\right)^{+}\right\|\left(O_p\left(\max_{1\le i\le n}\left\|\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}\right\|^2\right)\right. \\
&\qquad\left.+\frac{1}{\sqrt{n-1}}\max_{1\le i\le n}\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega^*_{ni},\varepsilon_i\right)\right\|\right) \\
= &o_p\left(\frac{1}{\sqrt{n}}\right).
\end{aligned}
$$

The proof is complete. ■

Lemma S.9 (Stochastic equicontinuity) *Suppose that Assumptions 1-3 and 5 are satisfied. Then $\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega,\varepsilon_i\right)$ defined in (S.44) satisfies for any $\delta>0$,*

(i) if $\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}=o_p\left(1\right)$,

$$
\Pr\left(\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega^*_{ni},\varepsilon_i\right)\right\|>\delta\middle| X,p_n\right)\to 0 \tag{S.53}
$$

as $n\to\infty$, and

(ii) if $\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}=O_p\left(n^{-1/2}\right)$,

$$
\Pr\left(\max_{1\le i\le n}\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega^*_{ni},\varepsilon_i\right)\right\|>\delta\middle| X,p_n\right)\to 0 \tag{S.54}
$$

as $n\to\infty$.

Proof. Part (i): By consistency of $\omega_{ni}\left(\varepsilon_i\right)$, we can define h_{ni} by $\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}=r_n^{-1}h_{ni}$ for some $r_{ni}\to\infty$ at a rate slower than the rate at which ω_{ni} converges to ω^*_{ni} so that $h_{ni}\in\Omega$ if n is sufficiently large, because by Assumption 5 Ω contains a compact neighborhood of 0.

By Markov's inequality

$$
\begin{aligned}
&\Pr\left(\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega^*_{ni},\varepsilon_i\right)\right\|>\delta\middle| X,p_n\right) \\
\le &\Pr\left(\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega+\frac{h}{r_{ni}},\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega,\varepsilon_i\right)\right\|>\delta\middle| X,p_n\right) \\
\le &\frac{1}{\delta}\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega+\frac{h}{r_{ni}},\varepsilon_i\right)-\mathbb{G}_n\varphi^{\gamma}_{ni}\left(\omega,\varepsilon_i\right)\right\|\middle| X,p_n\right].
\end{aligned}
$$

We consider the empirical process

$$\mathbb{G}_n \varphi_{ni}^{\gamma}\left(\omega+\frac{h}{r_{ni}},\varepsilon_i\right)-\mathbb{G}_n \varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)$$

indexed by $\omega, h \in \Omega$. Recall that

$$\begin{aligned}
&\mathbb{G}_n \varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)-\mathbb{G}_n \varphi_{ni}^{\gamma}\left(\tilde{\omega},\varepsilon_i\right) \\
=&\frac{1}{\sqrt{n-1}}\sum_{j\neq i}\varphi_{n,ij}^{\gamma}\left(\omega,\varepsilon_{ij}\right)-\varphi_{n,ij}^{\gamma}\left(\tilde{\omega},\varepsilon_{ij}\right)-\left(\mathbb{E}\left[\varphi_{n,ij}^{\gamma}\left(\omega,\varepsilon_{ij}\right)-\varphi_{n,ij}^{\gamma}\left(\tilde{\omega},\varepsilon_{ij}\right)\middle| X,p_n\right]\right),
\end{aligned}$$

where $\varphi_{n,ij}^{\gamma}\left(\omega,\varepsilon_{ij}\right)$ is defined in (S.43) by

$$\varphi_{n,ij}^{\gamma}\left(\omega,\varepsilon_{ij}\right)=1\left\{U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega-\varepsilon_{ij}\geq 0\right\}\Lambda_{ni}\Phi_{ni}'Z_j.$$

Here we essentially need to show that this empirical process is stochastically equicontinuous. Notice that this empirical process is a triangular array with function $\varphi_{n,ij}^{\gamma}$ that varies across j (because we condition on X),[22] so most of the ready-to-use results for stochastic equicontinuity (e.g. Andrews, 1994) are not applicable. Instead, we apply maximal inequalities in van der Vaart and Wellner (1996) to directly prove the stochastic equicontinuity.

Observe that for any $\omega, \tilde{\omega} \in \Omega$ the function $\varphi_{n,ij}^{\gamma}\left(\omega,\varepsilon_{ij}\right)-\varphi_{n,ij}^{\gamma}\left(\tilde{\omega},\varepsilon_{ij}\right)$ can be bounded by

$$\begin{aligned}
&\left\|\varphi_{n,ij}^{\gamma}\left(\omega,\varepsilon_{ij}\right)-\varphi_{n,ij}^{\gamma}\left(\tilde{\omega},\varepsilon_{ij}\right)\right\| \\
\leq&\left\|\Lambda_{ni}\Phi_{ni}'Z_j\right\| \\
&\cdot\left|1\left\{U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega\geq\varepsilon_{ij}\right\}-1\left\{U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\tilde{\omega}\geq\varepsilon_{ij}\right\}\right| \\
\leq&\eta_{n,ij}\left(\omega,\tilde{\omega},\varepsilon_{ij}\right),
\end{aligned}$$

[22] Both $U_{n,ij}$ and Z_j vary in j.

with $\eta_{n,ij}(\omega, \tilde{\omega}, \varepsilon_{ij})$ given by

$$\eta_{n,ij}(\omega, \tilde{\omega}, \varepsilon_{ij}) = \begin{cases} \|\Lambda_{ni}\Phi'_{ni}Z_j\|, & \text{if } \varepsilon_{ij} \text{ lies between } U_{n,ij} + \frac{2(n-1)}{n-2}Z'_j\Phi_{ni}\Lambda_{ni}\omega \\ & \text{and } U_{n,ij} + \frac{2(n-1)}{n-2}Z'_j\Phi_{ni}\Lambda_{ni}\tilde{\omega}, \\ 0, & \text{otherwise.} \end{cases} \tag{S.55}$$

Next, we apply Theorem 2.14.1 in van der Vaart and Wellner (1996). This theorem gives a uniform upper bound to the absolute p-th moment of an empirical process that we take as

$$\mathbb{G}_n\left(\varphi_{ni}^{\gamma}\left(\omega + \frac{h}{r_{ni}}, \varepsilon_i\right) - \varphi_{ni}^{\gamma}(\omega, \varepsilon_i)\right) \tag{S.56}$$

indexed by $\omega, h \in \Omega$. We take the expectation conditional on X and p_n and set $p = 1$. The bound from Theorem 2.14.1 in van der Vaart and Wellner (1996) is[23]

$$\begin{aligned} &\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\left(\varphi_{ni}^{\gamma}\left(\omega + \frac{h}{r_{ni}}, \varepsilon_i\right) - \varphi_{ni}^{\gamma}(\omega, \varepsilon_i)\right)\right\| \middle| X, p_n\right] \\ &\leq K\mathbb{E}\left[J(1, \mathcal{F}_{ni}(\varepsilon_i)) \sup_{\omega,h\in\Omega}\left\|\eta_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)\right\|_n \middle| X, p_n\right], \end{aligned} \tag{S.57}$$

where $K > 0$ is a constant and

$$\left\|\eta_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)\right\|_n^2 = \frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_{ij}\right). \tag{S.58}$$

We first show that the uniform entropy integral $J(1, \mathcal{F}_{ni}(\varepsilon_i))$ in (S.57) is finite, where $\mathcal{F}_{ni}(\varepsilon_i)$ denotes the set of arrays

$$\mathcal{F}_{ni}(\varepsilon_i) = \left\{\left(\varphi_{n,ij}^{\gamma}\left(\omega + \frac{h}{r_{ni}}, \varepsilon_{ij}\right) - \varphi_{n,ij}^{\gamma}(\omega, \varepsilon_{ij}), j \neq i\right) : \omega, h \in \Omega\right\}, \tag{S.59}$$

[23] From the proof of Theorem 2.14.1 in van der Vaart and Wellner (1996) it follows that the empirical L_2 norm of an envelope of $\mathcal{F}_{ni}(\varepsilon_i)$ can be replaced by the sup of the empirical L_2 norm of the $n-1$ bounding functions in $\eta_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)$. Also the theorem holds for a triangular array with independent but non-identically distributed observations.

and $J(1, \mathcal{F}_{ni}(\varepsilon_i))$ is the uniform entropy integral of $\mathcal{F}_{ni}(\varepsilon_i)$

$$J(1, \mathcal{F}_{ni}(\varepsilon_i)) = \int_0^1 \sup_{\alpha \in \mathbb{R}_+^{n-1}} \sqrt{\ln D\left(\xi \left\|\alpha \odot \bar{\eta}_{ni}(\varepsilon_i)\right\|_n, \alpha \odot \mathcal{F}_{ni}(\varepsilon_i), \|\cdot\|_n\right)} d\xi. \quad \text{(S.60)}$$

In (S.60), $\bar{\eta}_{ni}(\varepsilon_i) = \sup_{\omega, h \in \Omega} \eta_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)$ is an $(n-1) \times 1$ vector of envelope functions of $\mathcal{F}_{ni}(\varepsilon_i)$, α is an $(n-1) \times 1$ vector of nonnegative constants, $\alpha \odot \bar{\eta}_{ni}(\varepsilon_i)$ is the Hadamard product of α and $\bar{\eta}_{ni}(\varepsilon_i)$, and $\alpha \odot \mathcal{F}_{ni}(\varepsilon_i)$ is the set of Hadamard products of α and the functions in $\mathcal{F}_{ni}(\varepsilon_i)$. Also $\|\cdot\|_n$ is the empirical L_2 norm defined in (S.58), and $D\left(\xi \left\|\alpha \odot \bar{\eta}_{ni}(\varepsilon_i)\right\|_n, \alpha \odot \mathcal{F}_{ni}(\varepsilon_i), \|\cdot\|_n\right)$ is the packing number, i.e., the maxinum number of points in the set $\alpha \odot \mathcal{F}_{ni}(\varepsilon_i)$ that are separated by the distance $\xi \left\|\alpha \odot \bar{\eta}_{ni}(\varepsilon_i)\right\|_n$ for the norm $\|\cdot\|_n$. The sup in (S.60) is taken over all $(n-1) \times 1$ vectors α of nonnegative constants.

Consider the function

$$g_{n,ij}(\omega, \varepsilon_{ij}) = 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \geq 0\right\}.$$

It is an indicator with the argument being a linear function of ω. We can show that the set $\{(g_{n,ij}(\omega, \varepsilon_{ij}), j \neq i) : \omega \in \Omega\}$ has a pseudo-dimension of at most T,[24] so it is Euclidean (Pollard, 1990, Corollary 4.10). Note that $\varphi_{n,ij}^{\gamma}(\omega, \varepsilon_{ij}) = g_{n,ij}(\omega, \varepsilon_{ij}) \Lambda_{ni} \Phi_{ni}' Z_j$,

[24] To see this, by the definition of pseudo-dimension, it suffices to show that for each index set $I = \{j_1, \ldots, j_{T+1}\} \in \{1, \ldots, n\} \setminus \{i\}$ and each point $c \in \mathbb{R}^{T+1}$, there is a subset $J \subseteq I$ such that no $\omega \in \Omega$ can satisfy the inequalities

$$1\left\{U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \geq 0\right\} \begin{cases} > c_j & \text{for } j \in J \\ < c_j & \text{for } j \in I \backslash J \end{cases}$$

If c has a component c_j that lies outside of $(0,1)$, we can choose J such that $j \in J$ if $c_j \geq 1$ and $j \in I \backslash J$ if $c_j \leq 0$ so no ω can satisfy the inequalities above. It thus suffices to consider c with all the components in $(0,1)$ and for such c the inequalities reduce to

$$U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \begin{cases} \geq 0 & \text{for } j \in J \\ < 0 & \text{for } j \in I \backslash J \end{cases}$$

Since $Z_j' \Phi_{ni} \Lambda_{ni} \in \mathbb{R}^T$ for all j, there exists a non-zero vector $\tau = (\tau_1, \ldots, \tau_{T+1}) \in \mathbb{R}^{T+1}$ such that $\sum_{t=1}^{T+1} \tau_t Z_{j_t}' \Phi_{ni} \Lambda_{ni} = 0$, so $\sum_{t=1}^{T+1} \tau_t \frac{2(n-1)}{n-2} Z_{j_t}' \Phi_{ni} \Lambda_{ni} \omega = 0$ for all $\omega \in \Omega$. We may assume that $\tau_t > 0$ for at least one t. If $\sum_{t=1}^{T+1} \tau_t (U_{n,ij_t} - \varepsilon_{n,ij_t}) \geq 0$, it is impossible to find a $\omega \in \Omega$ satisfying these inequalities for the choice $J = \{j_t \in I : \tau_t \leq 0\}$, because this would lead to the contradiction $\sum_{t=1}^{T+1} \tau_t (U_{n,ij_t} - \varepsilon_{n,ij_t}) = \sum_{t=1}^{T+1} \tau_t (U_{n,ij_t} - \varepsilon_{n,ij_t}) + \sum_{t=1}^{T+1} \tau_t \frac{2(n-1)}{n-2} Z_{j_t}' \Phi_{ni} \Lambda_{ni} \omega = \sum_{t=1}^{T+1} \tau_t \left(U_{n,ij_t} + \frac{2(n-1)}{n-2} Z_{j_t}' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{n,ij_t}\right) < 0$. If $\sum_{t=1}^{T+1} \tau_t (U_{n,ij_t} - \varepsilon_{n,ij_t}) < 0$, we could choose $J = \{j_t \in I : \tau_t \geq 0\}$ to reach a similar contradiction.

and $\Lambda_{ni}\Phi'_{ni}Z_j$ is a $T \times 1$ vector that does not depend on ω. From the stability results in Pollard (1990, Section 5), each component of the doubly indexed process $\{(\varphi^{\gamma}_{ni}\left(\omega + \frac{h}{r_{ni}}, \varepsilon_i\right) - \varphi^{\gamma}_{ni}(\omega, \varepsilon_i)), j \neq i) : \omega, h \in \Omega\}$ is Euclidean. Therefore, the set $\mathcal{F}_{ni}(\varepsilon_i)$ has a finite uniform entropy integral, i.e.,

$$J\left(1, \mathcal{F}_{ni}(\varepsilon_i)\right) \leq \bar{J} \tag{S.61}$$

uniformly in ε_i and n for some $\bar{J} < \infty$.

Next we consider the empirical L_2 norm in the bound. By Jensen's inequality

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left\|\eta_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)\right\|_n \middle| X, p_n\right] \\
=&\mathbb{E}\left[\left(\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\eta^2_{n,ij}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_{ij}\right)\right)^{1/2}\middle| X, p_n\right] \\
\leq&\left(\mathbb{E}\left[\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\eta^2_{n,ij}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_{ij}\right)\middle| X, p_n\right]\right)^{1/2}.
\end{aligned} \tag{S.62}$$

To derive an upper bound on the last term in (S.62), we consider the empirical process

$$\begin{aligned}
&\mathbb{G}_n\eta^2_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right) \\
=&\frac{1}{\sqrt{n-1}}\sum_{j\neq i}\left(\eta^2_{n,ij}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_{ij}\right) - \mathbb{E}\left[\eta^2_{n,ij}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_{ij}\right)\middle| X, p_n\right]\right)
\end{aligned}$$

indexed by $\omega, h \in \Omega$. Note that each $\eta^2_{n,ij}$ is bounded by $\|\Lambda_{ni}\Phi'_{ni}Z_j\|^2 \leq \max_{t=1,\cdots,T}\lambda^2_{ni,t}T$ $\leq \bar{\eta}^2 < \infty$. Similarly to (S.57), we apply Theorem 2.14.1 in van der Vaart and Wellner (1996) to this empirical process with $p = 1$ and get an upper bound

$$\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left|\mathbb{G}_n\eta^2_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)\right|\middle| X, p_n\right] \leq K^{\eta}\mathbb{E}\left[J\left(1, \mathcal{F}^{\eta}_{ni}(\varepsilon_i)\right)\left\|\bar{\eta}^2\right\|_n \middle| X, p_n\right],$$

with $K^{\eta} > 0$ a constant,

$$\left\|\bar{\eta}^2\right\|_n = \sqrt{\frac{1}{n-1}\sum_{j\neq i}\bar{\eta}^4} = \bar{\eta}^2,$$

and $J\left(1,\mathcal{F}_{ni}^{\eta}\left(\varepsilon_i\right)\right)$ the uniform entropy integral of the set

$$\mathcal{F}_{ni}^{\eta}\left(\varepsilon_i\right)=\left\{\left(\eta_{n,ij}^2\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right),j\neq i\right):\omega,h\in\Omega\right\}.$$

Similarly to the argument for the set $\mathcal{F}_{ni}\left(\varepsilon_i\right)$ in (S.59), we can show that the set $\mathcal{F}_{ni}^{\eta}\left(\varepsilon_i\right)$ has a finite uniform entropy integral

$$J\left(1,\mathcal{F}_{ni}^{\eta}\left(\varepsilon_i\right)\right)\leq\bar{J}^{\eta}<\infty.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\right.\\
&\qquad\left.-\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\mathbb{E}\left[\eta_{n,ij}^2\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\middle|X,p_n\right]\middle|X,p_n\right]\\
\leq&\frac{1}{\sqrt{n-1}}\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left|\mathbb{G}_n\eta_{ni}^2\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_i\right)\right|\middle|X,p_n\right]\\
\leq&\frac{K^{\eta}\bar{J}^{\eta}\bar{\eta}^2}{\sqrt{n-1}}\equiv\frac{M^{\eta}}{\sqrt{n-1}}.
\end{aligned}$$

For any $\omega,h\in\Omega$ and any $j\neq i$, by the mean-value theorem, we have

$$\begin{aligned}
&\mathbb{E}\left[\eta_{n,ij}^2\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\middle|X,p_n\right]\\
=&\left|F_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\left(\omega+\frac{h}{r_{ni}}\right)\right)-F_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega\right)\right|\\
&\cdot\left\|\Lambda_{ni}\Phi_{ni}'Z_j\right\|^2\\
=&f_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\left(\omega+t_{n,ij}\frac{h}{r_{ni}}\right)\right)\frac{2\left(n-1\right)}{n-2}\left|Z_j'\Phi_{ni}\Lambda_{ni}\frac{h}{r_{ni}}\right|\left\|\Lambda_{ni}\Phi_{ni}'Z_j\right\|^2\\
\leq&\frac{1}{r_{ni}}f_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\left(\omega+t_{n,ij}\frac{h}{r_{ni}}\right)\right)\frac{2\left(n-1\right)}{n-2}\left\|\Lambda_{ni}\Phi_{ni}'Z_j\right\|^3\sup_{h\in\Omega}\|h\|
\end{aligned}\tag{S.63}$$

for some $t_{n,ij}\in[0,1]$. By Assumption 1, the density f_{ε} is bounded. There is also a finite bound on the eigenvalues in Λ_{ni} that does not depend on i. We conclude that there is a finite M with

$$\mathbb{E}\left[\eta_{n,ij}^{2}\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\middle| X,p_n\right]\leq\frac{M}{r_{ni}}$$

for all $\omega,h\in\Omega$ and all j. Hence

$$\mathbb{E}\left[\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^{2}\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\middle| X,p_n\right]\leq\frac{M^{\eta}}{\sqrt{n-1}}+\frac{M}{r_{ni}}.$$

Combining the results we obtain the upper bound

$$\Pr\left(\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}^{*},\varepsilon_i\right)\right\|>\delta\middle| X,p_n\right)\leq\frac{K\bar{J}}{\delta}\sqrt{\frac{M^{\eta}}{\sqrt{n-1}}+\frac{M}{r_{ni}}},$$

which for all $\delta>0$ can be made arbitrarily small by making n sufficiently large. Part (i) is proved.

Part (ii): Because $\omega_{ni}\left(\varepsilon_i\right)-\omega_{ni}^{*}=O_p\left(n^{-1/2}\right)$, we can define h_{ni} by $\omega_{ni}\left(\varepsilon_i\right)-\omega_{ni}^{*}=n^{-\kappa}h_{ni}$ for $0<\kappa<1/2$, and $h_{ni}\in\Omega$ if n is sufficiently large. By Markov's inequality

$$\begin{aligned}
&\Pr\left(\max_{1\leq i\leq n}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}^{*},\varepsilon_i\right)\right\|>\delta\middle| X,p_n\right)\\
\leq&\Pr\left(\max_{1\leq i\leq n}\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)\right\|>\delta\middle| X,p_n\right)\\
\leq&\frac{1}{\delta}\mathbb{E}\left[\max_{1\leq i\leq n}\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)\right\|\middle| X,p_n\right].
\end{aligned}$$

Because for any random variable $\mathbb{E}\left[\left|Z\right|\middle| X,p_n\right]\leq\left\|Z\right\|_{\psi|X,p_n}$ with $\left\|\cdot\right\|_{\psi|X,p_n}$ the conditional Orlicz norm given X and p_n and $\psi\left(x\right)=e^{x}-1$, by the maximal inequality in Lemma 2.2.2 in van der Vaart and Wellner (1996)

$$\begin{aligned}
&\mathbb{E}\left[\max_{1\leq i\leq n}\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)\right\|\middle| X,p_n\right]\\
\leq&\left\|\max_{1\leq i\leq n}\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)\right\|\right\|_{\psi|X,p_n}\\
\leq&K\ln\left(n+1\right)\max_{1\leq i\leq n}\left\|\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)\right\|\right\|_{\psi|X,p_n},
\end{aligned}$$

where $K > 0$ is a constant.

By Theorem 2.14.5 (with $p = 1$) and Lemma 2.2.2 in van der Vaart and Wellner (1996), we have

$$\begin{aligned}
&\left\| \sup_{\omega,h\in\Omega} \left\| \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega + \frac{h}{n^{\kappa}}, \varepsilon_i \right) - \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega, \varepsilon_i \right) \right\| \right\|_{\psi|X,p_n} \\
\leq & K_1 \left(\mathbb{E}\left[\sup_{\omega,h\in\Omega} \left\| \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega + \frac{h}{n^{\kappa}}, \varepsilon_i \right) - \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega, \varepsilon_i \right) \right\| \middle| X, p_n \right] \right. \\
& \left. + \frac{\ln n}{\sqrt{n-1}} \max_{j\neq i} \left\| \sup_{\omega,h\in\Omega} \left\| \eta_{n,ij} \left(\omega + \frac{h}{n^{\kappa}}, \omega, \varepsilon_{ij} \right) \right\| \right\|_{\psi|X,p_n} \right).
\end{aligned}$$

In Part (i) we derived for the first term in the upper bound

$$\mathbb{E}\left[\sup_{\omega,h\in\Omega} \left\| \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega + \frac{h}{n^{\kappa}}, \varepsilon_i \right) - \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega, \varepsilon_i \right) \right\| \middle| X, p_n \right] \leq K \bar{J} \sqrt{\frac{M^{\eta}}{\sqrt{n-1}} + \frac{M}{n^{\kappa}}}.$$

For the second term by the definition of $\eta_{n,ij}$,

$$\begin{aligned}
\max_{j\neq i} \left\| \sup_{\omega,h\in\Omega} \left\| \eta_{n,ij} \left(\omega + \frac{h}{n^{\kappa}}, \omega, \varepsilon_{ij} \right) \right\| \right\|_{\psi|X,p_n} &\leq \max_{j\neq i} \left\| \Lambda_{ni} \Phi_{ni}' Z_j \right\| \\
&\leq \max_{t=1,,T} \lambda_{ni,t} \sqrt{T} \leq \bar{\eta} < \infty.
\end{aligned}$$

Therefore, for $1 \leq i \leq n$

$$\begin{aligned}
&\left\| \sup_{\omega,h\in\Omega} \left\| \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega + \frac{h}{n^{\kappa}}, \varepsilon_i \right) - \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega, \varepsilon_i \right) \right\| \right\|_{\psi|X,p_n} \\
\leq & K_1 \left(K \bar{J} \sqrt{\frac{M^{\eta}}{\sqrt{n-1}} + \frac{M}{n^{\kappa}}} + \frac{\bar{\eta} \ln n}{\sqrt{n-1}} \right).
\end{aligned} \tag{S.64}$$

Combining the results yields the upper bound

$$\begin{aligned}
&\Pr\left(\max_{1\leq i\leq n} \left\| \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega_{ni} \left(\varepsilon_i \right), \varepsilon_i \right) - \mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega_{ni}^*, \varepsilon_i \right) \right\| > \delta \middle| X, p_n \right) \\
\leq & \frac{K K_1 \ln (n+1)}{\delta} \left(K \bar{J} \sqrt{\frac{M^{\eta}}{\sqrt{n-1}} + \frac{M}{n^{\kappa}}} + \frac{\bar{\eta} \ln n}{\sqrt{n-1}} \right),
\end{aligned}$$

which for all $\delta > 0$ can be made arbitrarily small by making n sufficiently large. The proof is complete. ■

Asymptotic Distribution of $\hat{\theta}_n$

Proof of Theorem 4.2. The GMM estimator of θ_0 satisfies the sample unconditional moment condition

$$\hat{\Psi}_n(\hat{\theta}_n, \hat{p}_n) = \frac{1}{n\,(n-1)} \sum_i \sum_{j \neq i} \hat{W}_{n,ij} \left(G_{n,ij} - P_{n,ij}(\hat{\theta}_n, \hat{p}_n) \right) = o_p\left(\frac{1}{n}\right)$$

with $\hat{p}_n$ the $T \times T$ matrix of empirical link frequencies between the types. We arrange the link frequencies in a vector and with abuse of notation we use $\hat{p}_n$ for $\text{vec}(\hat{p}_n')$.

By a Taylor-series expansion of $P_{n,ij}(\hat{\theta}_n, \hat{p}_n)$ around (θ_0, p_n)

$$\begin{aligned} P_{n,ij}(\hat{\theta}_n, \hat{p}_n) =& P_{n,ij}(\theta_0, p_n) + \nabla_{\theta'} P_{n,ij}(\theta_0, p_n)(\hat{\theta}_n - \theta_0) \\ &+ \nabla_{p'} P_{n,ij}(\theta_0, p_n)(\hat{p}_n - p_n) + o_p\left(\left\| (\hat{\theta}_n, \hat{p}_n) - (\theta_0, p_n) \right\| \right) \end{aligned}$$

and upon rearranging the terms of the expansion we have

$$\begin{aligned} &\frac{1}{n\,(n-1)} \sum_i \sum_{j \neq i} \hat{W}_{n,ij} \nabla_{\theta'} P_{n,ij}(\theta_0, p_n)(\hat{\theta}_n - \theta_0) \\ =& \frac{1}{n\,(n-1)} \sum_i \sum_{j \neq i} \hat{W}_{n,ij} \left(G_{n,ij} - P_{n,ij}(\theta_0, p_n) \right) \\ &- \frac{1}{n\,(n-1)} \sum_i \sum_{j \neq i} \hat{W}_{n,ij} \nabla_{p'} P_{n,ij}(\theta_0, p_n)(\hat{p}_n - p_n) \\ &- \frac{1}{n\,(n-1)} \sum_i \sum_{j \neq i} \hat{W}_{n,ij} o_p\left(\left\| (\hat{\theta}_n, \hat{p}_n) - (\theta_0, p_n) \right\| \right) + o_p\left(\frac{1}{n}\right), \end{aligned}$$

where we assume that n is sufficiently large, so that $(\hat{\theta}_n, \hat{p}_n)$ is in a neighborhood of (θ_0, p_n) where $P_{n,ij}(\theta, p)$ is continuously differentiable.

The instruments $\hat{W}_{n,ij}$ are estimated, but we have $\max_{i,j=1,\ldots,n} \left\| \hat{W}_{n,ij} - W_{n,ij} \right\| = o_p(1)$ by Assumption 4(iii), so the sampling variation in the instruments has no effect on the asymptotic distribution of $\hat{\theta}_n$. The GMM estimator thus satisfies

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}W_{n,ij}\nabla_{\theta'}P_{n,ij}(\theta_0,p_n)\left(\hat{\theta}_n-\theta_0\right)\\
=&\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}W_{n,ij}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right)\\
&-\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}W_{n,ij}\nabla_{p'}P_{n,ij}(\theta_0,p_n)\left(\hat{p}_n-p_n\right)\\
&+o_p\left(\left\|\hat{\theta}_n-\theta_0\right\|\right)+o_p\left(\|\hat{p}_n-p_n\|\right)+o_p\left(\frac{1}{n}\right).
\end{aligned}\tag{S.65}$$

where we have used

$$\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}W_{n,ij}o_p\left(\left\|(\hat{\theta}_n,\hat{p}_n)-(\theta_0,p_n)\right\|\right)\leq o_p\left(\left\|\hat{\theta}_n-\theta_0\right\|\right)+o_p\left(\|\hat{p}_n-p_n\|\right)$$

because $\left\|(\hat{\theta}_n,\hat{p}_n)-(\theta_0,p_n)\right\|\leq\left\|\hat{\theta}_n-\theta_0\right\|+\|\hat{p}_n-p_n\|$ and $\max_{i,j=1,\ldots,n}\|W_{n,ij}\|<\infty$ (Assumption 4(iii)).

Let us examine the first two terms on the right-hand side of (S.65), with the first being the main term while the second gives the contribution of the first-stage estimation of the link probabilities. Recall that $\hat{p}_n$ is the vector of empirical fractions of pairs of type s, t that have a link and p_n is the vector of link probabilities of pairs of type s, t so

$$\hat{p}_n-p_n=\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}Q_{n,ij}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right),$$

where $Q_{n,ij}=(Q_{n,ij,11},\ldots,Q_{n,ij,1T},\ldots,Q_{n,ij,T1},\ldots,Q_{n,ij,TT})'\in\mathbb{R}^{T^2}$ with

$$Q_{n,ij,st}=\frac{1\left\{X_i=x_s,X_j=x_t\right\}}{\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}1\left\{X_i=x_s,X_j=x_t\right\}},\quad s,t=1,\ldots,T.$$

Hence, the first two terms on the right-hand side of (S.65) can be combined as

$$\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\tilde{W}_{n,ij}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right),$$

where $\tilde{W}_{n,ij}$ is the augmented instrument that combines the instruments for the first-stage and second-stage estimation

$$\tilde{W}_{n,ij} = W_{n,ij} - \left(\frac{1}{n(n-1)} \sum_{k} \sum_{l \neq k} W_{n,kl} \nabla_{p'} P_{n,kl}(\theta_0, p_n) \right) Q_{n,ij}.$$

Applying Lemma S.10 twice for the instrument vectors $\tilde{W}_{n,ij}$ and $Q_{n,ij}$ we get

$$\frac{1}{n(n-1)} \sum_{i} \sum_{j \neq i} \tilde{W}_{n,ij} \left(G_{n,ij} - P_{n,ij}(\theta_0, p_n) \right) = O_p\left(\frac{1}{n}\right), \tag{S.66}$$

and

$$\hat{p}_n - p_n = O_p\left(\frac{1}{n}\right).$$

so (S.65) becomes

$$J_n^{\theta}(\theta_0, p_n)\left(\hat{\theta}_n - \theta_0\right) = O_p\left(\frac{1}{n}\right) + o_p\left(\left\|\hat{\theta}_n - \theta_0\right\|\right) + o_p\left(\frac{1}{n}\right),$$

with $J_n^{\theta}(\theta_0, p_n)$ being the Jacobian matrix

$$J_n^{\theta}(\theta_0, p_n) = \frac{1}{n(n-1)} \sum_{i} \sum_{j \neq i} W_{n,ij} \nabla_{\theta'} P_{n,ij}(\theta_0, p_n).$$

By Assumption 6(ii) $J_n^{\theta}(\theta_0, p_n)$ is nonsingular, so

$$\left\| J_n^{\theta}(\theta_0, p_n)(\theta - \theta_0) \right\| \geq c \left\| \theta - \theta_0 \right\|$$

with $c = \lambda_{\min}\left(J_n^{\theta}(\theta_0, p_n)' J_n^{\theta}(\theta_0, p_n) \right) > 0$. Therefore,

$$\left\| \hat{\theta}_n - \theta_0 \right\| (c + o_p(1)) \leq O_p\left(\frac{1}{n}\right) + o_p\left(\frac{1}{n}\right).$$

This implies that

$$\hat{\theta}_n - \theta_0 = O_p\left(\frac{1}{n}\right).$$

i.e., $\hat{\theta}_n$ is n-consistent for θ_0.

To derive the asymptotic distribution of $\hat{\theta}_n$, we rewrite (S.65) as

$$\begin{aligned}
&\sqrt{n(n-1)}\left(\hat{\theta}_n-\theta_0\right) \\
=&\frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}J_n^{\theta}(\theta_0,p_n)^{-1}\tilde{W}_{n,ij}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right)+o_p(1).
\end{aligned}$$

Note that for each i the $G_{n,ij}$ are correlated over j, so that a CLT for independent random variables cannot be used. We need the result for dependent random variables in Lemma S.10. The link chocies of i in the $n-1$ vector G_{ni} are correlated through their dependence on $\omega_{ni}(\varepsilon_i)$. The correlation goes to 0 as $n\to\infty$, so the sample moments have an asymptotic normal distribution with a finite variance that accounts for the variation in $\omega_{ni}(\varepsilon_i)$.

We apply Lemma S.10 for the instrument vector $J_n^{\theta}(\theta_0,p_n)^{-1}\tilde{W}_{n,ij}$. Define the $d_\theta\times d_\theta$ matrix

$$\Sigma_n(\theta_0,p_n)=\frac{1}{n(n-1)}\sum_i\sum_{j\neq i}\Sigma_{n,ij}(\theta_0,p_n)$$

with

$$\begin{aligned}
&\Sigma_{n,ij}(\theta_0,p_n) \\
=&J_n^{\theta}(\theta_0,p_n)^{-1}\mathbb{E}\left[\left(\tilde{W}_{n,ij}\left(g_{n,ij}(\omega_{ni}^*,\varepsilon_{ij})-P_{n,ij}^*(\omega_{ni}^*)\right)+\tilde{J}_{ni}^{\omega}(\omega_{ni}^*)\varphi_{n,ij}^{\omega}(\omega_{ni}^*,\varepsilon_{ij})\right)\right. \\
&\cdot\left.\left(\tilde{W}_{n,ij}\left(g_{n,ij}(\omega_{ni}^*,\varepsilon_{ij})-P_{n,ij}^*(\omega_{ni}^*)\right)+\tilde{J}_{ni}^{\omega}(\omega_{ni}^*)\varphi_{n,ij}^{\omega}(\omega_{ni}^*,\varepsilon_{ij})\right)'\right|X,p_n\right]\left(J_n^{\theta}(\theta_0,p_n)^{-1}\right)'
\end{aligned}\tag{S.67}$$

where $\omega_{ni}^*\in\mathbb{R}^T$ maximizes $\Pi_{ni}^*(\omega)$ in (4.8). The indicator function $g_{n,ij}(\omega,\varepsilon_{ij})$ and the corresponding probability $P_{n,ij}^*(\omega)$ are defined in Lemma S.10. The $d_\theta\times T$ matrix $\tilde{J}_{ni}^{\omega}(\omega)$ is defined by

$$\tilde{J}_{ni}^{\omega}(\omega)=\frac{1}{n-1}\sum_{j\neq i}\tilde{W}_{n,ij}\nabla_{\omega'}P_{n,ij}^*(\omega)$$

The function $\varphi_{n,ij}^{\omega}(\omega_{ni}^*,\varepsilon_{ij})\in\mathbb{R}^T$ is the j-th term of the influence function of ω_{ni} defined in (S.32) in Lemma S.7. By Lemma S.10,

$$\sqrt{n(n-1)}\Sigma_n^{-1/2}(\theta_0,p_n)\left(\hat{\theta}_n-\theta_0\right)\xrightarrow{d}N\left(0,I_{d_\theta}\right)$$

as $n \to \infty$. The proof is complete. ■

Lemma S.10 (Asymptotic normality of the sample moment) *Suppose that Assumption 1-3 and 5 are satisfied. Define*

$$Y_n = \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} W_{n,ij} \left(G_{n,ij} - P_{n,ij}\left(\theta_0, p_n\right)\right).$$

where $W_{n,ij}$ is a $d_\theta \times 1$ instrument vector.[25] *Let Σ_n be the $d_\theta \times d_\theta$ positive-definite matrix*

$$\Sigma_n = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \Sigma_{n,ij} \tag{S.68}$$

with

$$\begin{aligned} \Sigma_{n,ij} =& \mathbb{E}\Big[\left(W_{n,ij}\left(g_{n,ij}\left(\omega_{ni}^*, \varepsilon_{ij}\right) - P_{n,ij}^*\left(\omega_{ni}^*\right)\right) + J_{ni}^\omega\left(\omega_{ni}^*\right) \varphi_{n,ij}^\omega\left(\omega_{ni}^*, \varepsilon_{ij}\right)\right) \\ & \cdot \left(W_{n,ij}\left(g_{n,ij}\left(\omega_{ni}^*, \varepsilon_{ij}\right) - P_{n,ij}^*\left(\omega_{ni}^*\right)\right) + J_{ni}^\omega\left(\omega_{ni}^*\right) \varphi_{n,ij}^\omega\left(\omega_{ni}^*, \varepsilon_{ij}\right)\right)' \Big| X, p_n \Big] \end{aligned} \tag{S.69}$$

where $\omega_{ni}^* \in \mathbb{R}^T$ maximizes the function

$$\Pi_{ni}^*(\omega) = \sum_{j \neq i} \mathbb{E}\left[\left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \right]_+ \middle| X, p_n \right] - \frac{(n-1)^2}{n-2} \omega' \Lambda_{ni} \omega.$$

The indicator functions $g_{n,ij}(\omega, \varepsilon_{ij})$, the corresponding probabilities $P_{n,ij}^*(\omega)$, and the $d_\theta \times T$ matrix $J_{ni}^\omega(\omega)$ are defined by

$$\begin{aligned} g_{n,ij}\left(\omega, \varepsilon_{ij}\right) &= 1\left\{ U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \geq 0 \right\} \\ P_{n,ij}^*(\omega) &= F_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega \right) \\ J_{ni}^\omega(\omega) &= \frac{1}{n-1} \sum_{j \neq i} W_{n,ij} \nabla_{\omega'} P_{n,ij}^*(\omega), \end{aligned}$$

and $\varphi_{n,ij}^\omega\left(\omega_{ni}^*, \varepsilon_{ij}\right) \in \mathbb{R}^T$ is the j-th term of the influence function of $\omega_{ni}(\varepsilon_i)$ defined

[25] $W_{n,ij}$ is a generic valid instrument vector. It can be the augmented instrument vector $\tilde{W}_{n,ij}$ or the vector of first-stage instruments $Q_{n,ij}$. We discuss the choice of instrument in Section 4.

in (S.32) in Lemma S.7. Then

$$\Sigma_n^{-1/2} Y_n \xrightarrow{d} N\left(0, I_{d_\theta}\right)$$

as $n \to \infty$, where I_θ is the $d_\theta \times d_\theta$ identity matrix.

Proof. Define the link choice indicator at $\omega \in \Omega$

$$g_{n,ij}\left(\omega, \varepsilon_{ij}\right) = 1\left\{U_{n,ij} + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \geq 0\right\},\ j \neq i.$$

By Theorem 3.2, the observed link choice $G_{n,ij}$ is given by $g_{n,ij}\left(\omega, \varepsilon_{ij}\right)$ evaluated at $\omega_{ni}\left(\varepsilon_i\right)$, i.e.,

$$G_{n,ij} = g_{n,ij}\left(\omega_{ni}\left(\varepsilon_i\right), \varepsilon_{ij}\right),\ j \neq i, \tag{S.70}$$

where $\omega_{ni}\left(\varepsilon_i\right)$ maximizes the function

$$\Pi_{ni}\left(\omega\right) = \sum_{j \neq i}\left[U_{n,ij} + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij}\right]_+ - \frac{\left(n-1\right)^2}{n-2} \omega' \Lambda_{ni} \omega,\ i = 1, \ldots, n.$$

The conditional choice probability $P_{n,ij}\left(\theta_0, p_n\right)$ is thus the conditional expectation of $g_{n,ij}\left(\omega_{ni}\left(\varepsilon_i\right), \varepsilon_{ij}\right)$

$$P_{n,ij} = \mathbb{E}\left[g_{n,ij}\left(\omega_{ni}\left(\varepsilon_i\right), \varepsilon_{ij}\right) \middle| X, p_n\right],\ j \neq i. \tag{S.71}$$

The challenge in deriving the asymptotic distribution of the normalized sample moment Y_n lies in the fact that the link choices of an individual i, i.e., $G_{n,ij}$ and $G_{n,ik}$, are correlated through $\omega_{ni}\left(\varepsilon_i\right)$. As shown in Lemma S.5 $\omega_{ni}\left(\varepsilon_i\right)$ converges in probability to ω_{ni}^* that does not depend on ε_i. Let $\Pi_{ni}^*\left(\omega\right)$ be the expectation of $\Pi_{ni}\left(\omega\right)$

$$\Pi_{ni}^*\left(\omega\right) = \sum_{j \neq i} \mathbb{E}\left[\left[U_{n,ij} + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij}\right]_+ \middle| X, p_n\right] - \frac{\left(n-1\right)^2}{n-2} \omega' \Lambda_{ni} \omega,$$

$i = 1, \ldots, n$. By Assumption 5 $\Pi_{ni}^*\left(\omega\right)$ has a unique maximizer ω_{ni}^* that does not depend on ε_i.

Define the function $P^*_{n,ij}(\omega)$

$$\begin{aligned} P^*_{n,ij}(\omega) &= \mathbb{E}\left[g_{n,ij}(\omega, \varepsilon_{ij}) \middle| X, p_n \right] \\ &= F_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega \right), \ j \neq i. \end{aligned} \tag{S.72}$$

Here we treat ω as a parameter and take the expectation with respect to ε_{ij} only.

The normalized sample moment Y_n is equal to

$$\begin{aligned} Y_n =& \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} W_{n,ij}\left(g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) - \mathbb{E}\left[g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) \middle| X, p_n \right] \right) \\ =& T_{1n} + T_{2n} + T_{3n} + T_{4n}, \end{aligned} \tag{S.73}$$

where

$$\begin{aligned} T_{1n} =& \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} W_{n,ij}\left(g_{n,ij}(\omega^*_{ni}, \varepsilon_{ij}) - P^*_{n,ij}(\omega^*_{ni}) \right) \\ T_{2n} =& \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} W_{n,ij}\left(g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) - g_{n,ij}(\omega^*_{ni}, \varepsilon_{ij}) \right. \\ & \left. - \left(P^*_{n,ij}(\omega_{ni}(\varepsilon_i)) - P^*_{n,ij}(\omega^*_{ni}) \right) \right) \\ T_{3n} =& \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} W_{n,ij}\left(P^*_{n,ij}(\omega_{ni}(\varepsilon_i)) - P^*_{n,ij}(\omega^*_{ni}) \right) \\ T_{4n} =& -\frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} W_{n,ij}\left(\mathbb{E}\left[g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) \middle| X, p_n \right] - P^*_{n,ij}(\omega^*_{ni}) \right). \end{aligned} \tag{S.74}$$

The terms in the decomposition have an interpretation. T_{1n} is the sample moment condition if we replace $\omega_{ni}(\varepsilon_i)$ by its limit ω^*_{ni}. This substitution removes the correlation between the link choices of an individual. The term T_{2n} contains the difference between the dependent (through $\omega_{ni}(\varepsilon_i)$) sample moment function and the independent one in T_{1n}. The fact that this term is shown to be negligible shows that the correlation between the link choices vanishes if n is large. The sampling variation in $\omega_{ni}(\varepsilon_i)$ is captured by T_{3n} which contributes to the asymptotic variance of the moment function. Finally, the linear approximation of T_{3n} has non-negligible approximation errors (from both a Taylor series expansion remainder and a remainder in the asymptotically linear approximation of $\omega_{ni}(\varepsilon_i)$) that are $o_p(1)$ if we add T_{4n}.

Let us now examine the four terms in (S.74).

Step 1: T_{1n}.

The term T_{1n} is a normalized sum of link indicators that are evaluated at ω^*_{ni} rather than $\omega_{ni}(\varepsilon_i)$ and thus are independent. This is the main term in Y_n with an asymptotically normal distribution, because the CLT applies. It captures the sampling in the link choices, due to sampling variation in ε_{ij}.

Step 2: T_{2n}.

We show that T_{2n} in (S.74) is $o_p(1)$. Define for each i the empirical process

$$\mathbb{G}_n W_{ni} g_{ni}(\omega, \varepsilon_i) = \frac{1}{\sqrt{n-1}} \sum_{j \neq i} W_{n,ij} \left(g_{n,ij}(\omega, \varepsilon_{ij}) - P^*_{n,ij}(\omega) \right), \ \omega \in \Omega,$$

so that

$$T_{2n} = \frac{1}{\sqrt{n}} \sum_i \mathbb{G}_n W_{ni} \left(g_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - g_{ni}(\omega^*_{ni}, \varepsilon_i) \right).$$

Since each $\mathbb{G}_n W_{ni} \left(g_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - g_{ni}(\omega^*_{ni}, \varepsilon_i) \right)$ only involves ε_i, conditional on X and p_n, they are independent. By Lemma S.7 $\omega_{ni}(\varepsilon_i) - \omega^*_{ni} = O_p\left(n^{-1/2}\right)$, so if we define h_{ni} by $\omega_{ni}(\varepsilon_i) - \omega^*_{ni} = n^{-\kappa} h_{ni}$ for $0 < \kappa < 1/2$, then $h_{ni} \in \Omega$ if n is sufficiently large, because by Assumption 5(i) Ω contains a compact neighborhood of 0. Note that T_{2n} is a normalized average of terms that are each $o_p(1)$ by establishing stochastic equicontinuity. Hence we cannot directly invoke a stochastic equicontinuity argument to show that their sum T_{2n} is $o_p(1)$.

By Chebyshev's inequality, for any $\delta > 0$

$$\begin{aligned}
&\Pr\left(\left\| \frac{1}{\sqrt{n}} \sum_i \mathbb{G}_n W_{ni} \left(g_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - g_{ni}(\omega^*_{ni}, \varepsilon_i) \right) \right\| > \delta \middle| X, p_n \right) \\
\leq & \frac{1}{\delta^2 n} \sum_i \mathbb{E}\left[\left\| \mathbb{G}_n W_{ni} \left(g_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - g_{ni}(\omega^*_{ni}, \varepsilon_i) \right) \right\|^2 \middle| X, p_n \right] \\
\leq & \frac{1}{\delta^2 n} \sum_i \mathbb{E}\left[\sup_{\omega, h \in \Omega} \left\| \mathbb{G}_n W_{ni} \left(g_{ni}\left(\omega + \frac{h}{n^\kappa}, \varepsilon_i \right) - g_{ni}(\omega, \varepsilon_i) \right) \right\|^2 \middle| X, p_n \right] \qquad \text{(S.75)}
\end{aligned}$$

Observe that for any $\omega, \tilde{\omega} \in \Omega$, the function $W_{n,ij}\left(g_{n,ij}\left(\omega, \varepsilon_{ij}\right) - g_{n,ij}\left(\tilde{\omega}, \varepsilon_{ij}\right)\right)$ can be bounded by

$$
\begin{aligned}
& \left\| W_{n,ij}\left(g_{n,ij}\left(\omega, \varepsilon_{ij}\right) - g_{n,ij}\left(\tilde{\omega}, \varepsilon_{ij}\right)\right) \right\| \\
\leq & \left\| W_{n,ij} \right\| \left| 1\left\{ U_{n,ij} + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \geq 0 \right\} \right. \\
& \qquad \left. -1\left\{ U_{n,ij} + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \tilde{\omega} - \varepsilon_{ij} \geq 0 \right\} \right| \\
\leq & \eta_{n,ij}\left(\omega, \tilde{\omega}, \varepsilon_{ij}\right),
\end{aligned}
$$

with $\eta_{n,ij}\left(\omega, \tilde{\omega}, \varepsilon_{ij}\right)$ given by

$$
\begin{aligned}
& \eta_{n,ij}\left(\omega, \tilde{\omega}, \varepsilon_{ij}\right) \\
= & \begin{cases} \left\| W_{n,ij} \right\|, & \text{if } \varepsilon_{ij} \text{ lies between } U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega \\ & \text{and } U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \tilde{\omega}, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{S.76}
$$

Next, we apply Theorem 2.14.1 in van der Vaart and Wellner (1996) (with $p = 2$). This theorem gives a uniform upper bound to the absolute p-th moment of an empirical process that we take as

$$
\mathbb{G}_n W_{ni}\left(g_{ni}\left(\omega + \frac{h}{n^\kappa}, \varepsilon_i\right) - g_{ni}\left(\omega, \varepsilon_i\right)\right) \tag{S.77}
$$

indexed by $\omega, h \in \Omega$. The bound from Theorem 2.14.1 in van der Vaart and Wellner (1996) is[26]

$$
\begin{aligned}
& \mathbb{E}\left[\sup_{\omega, h \in \Omega} \left\| \mathbb{G}_n W_{ni}\left(g_{ni}\left(\omega + \frac{h}{n^\kappa}, \varepsilon_i\right) - g_{ni}\left(\omega, \varepsilon_i\right)\right) \right\|^2 \middle| X, p_n \right] \\
\leq & K \mathbb{E}\left[J\left(1, \mathcal{F}_{ni}\left(\varepsilon_i\right)\right)^2 \sup_{\omega, h \in \Omega} \left\| \eta_{ni}\left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_i\right) \right\|_n^2 \middle| X, p_n \right],
\end{aligned}
$$

[26]Similarly to the proof in Lemma S.9, from the proof of Theorem 2.14.1 in van der Vaart and Wellner (1996) it follows that the empirical L_2 norm of an envelope of $\mathcal{F}_{ni}\left(\varepsilon_i\right)$ can be replaced by the sup of the empirical L_2 norm of the $n-1$ bounding functions in $\eta_{ni}\left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_i\right)$. Also the theorem holds for a triangular array with independent but non-identically distributed observations.

where $K > 0$ is a constant and

$$\left\| \eta_{ni} \left(\omega + \frac{h}{n^{\kappa}}, \omega, \varepsilon_i \right) \right\|_n^2 = \frac{1}{n-1} \sum_{j \neq i} \eta_{n,ij}^2 \left(\omega + \frac{h}{n^{\kappa}}, \omega, \varepsilon_{ij} \right).$$

We now show that the uniform entropy integral $J\left(1, \mathcal{F}_{ni}\left(\varepsilon_i\right)\right)$ (defined as in S.60) is finite, where $\mathcal{F}_{ni}\left(\varepsilon_i\right)$ denotes the set of functions

$$\mathcal{F}_{ni}\left(\varepsilon_i\right) = \left\{ \left(W_{n,ij} \left(g_{n,ij} \left(\omega + \frac{h}{n^{\kappa}}, \varepsilon_{ij} \right) - g_{n,ij}\left(\omega, \varepsilon_{ij}\right) \right), j \neq i \right) : \omega, h \in \Omega \right\}.$$

Consider

$$g_{n,ij}\left(\omega, \varepsilon_{ij}\right) = 1 \left\{ U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \geq 0 \right\}.$$

It is an indicator with the argument being a linear function of ω. Following the argument in Lemma S.9, we can show that the set $\left\{\left(g_{n,ij}\left(\omega, \varepsilon_{ij}\right), j \neq i\right) : \omega \in \Omega\right\}$ has a pseudo-dimension of at most T, so it is Euclidean (Pollard, 1990, Corollary 4.10). Note that $W_{n,ij}$ is a $d_\theta \times 1$ vector that does not depend on ω. From the stability results in Pollard (1990, Section 5), each component of the doubly indexed process $\{(W_{n,ij}(g_{n,ij}\left(\omega + n^{-\kappa} h, \varepsilon_{ij}\right) - g_{n,ij}\left(\omega, \varepsilon_{ij}\right)), j \neq i) : \omega, h \in \Omega\}$ is Euclidean. Therefore, the set $\mathcal{F}_{ni}\left(\varepsilon_i\right)$ has a finite uniform entropy integral bounded by some $\bar{J} < \infty$.

Observe that

$$\begin{aligned} &\mathbb{E}\left[\sup_{\omega, h \in \Omega} \left\| \eta_{ni} \left(\omega + \frac{h}{n^{\kappa}}, \omega, \varepsilon_i \right) \right\|_n^2 \middle| X, p_n \right] \\ =&\mathbb{E}\left[\sup_{\omega, h \in \Omega} \frac{1}{n-1} \sum_{j \neq i} \eta_{n,ij}^2 \left(\omega + \frac{h}{n^{\kappa}}, \omega, \varepsilon_{ij} \right) \middle| X, p_n \right]. \end{aligned}$$

Note that $\eta_{n,ij}^2$ is bounded by $\left\| W_{n,ij} \right\|^2 \leq \max_{i,j=1,\ldots,n} \left\| W_{n,ij} \right\|^2 \equiv \bar{W}^2 < \infty$ (Assumption 4(iii)). Similar to the argument in Lemma S.9 we can show that the set of functions

$$\left\{ \left(\eta_{n,ij}^2 \left(\omega + \frac{h}{n^{\kappa}}, \omega, \varepsilon_{ij} \right), j \neq i \right) : \omega, h \in \Omega \right\}$$

has a finite uniform entropy integral bounded by some $\bar{J}^{\eta} < \infty$. Hence, we can apply Theorem 2.14.1 in van der Vaart and Wellner (1996) and derive an upper bound on the expectation of the empirical process

$$\begin{aligned}
&\mathbb{G}_n \eta_{ni}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_i\right) \\
&= \frac{1}{\sqrt{n-1}} \sum_{j \neq i} \left(\eta_{n,ij}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) - \mathbb{E}\left[\eta_{n,ij}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) \middle| X, p_n \right] \right)
\end{aligned}$$

indexed by $\omega, h \in \Omega$. The bound is

$$\mathbb{E}\left[\sup_{\omega, h \in \Omega} \left| \mathbb{G}_n \eta_{ni}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) \right| \middle| X, p_n \right] \leq K^\eta \bar{J}^\eta \bar{W}^2.$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\omega, h \in \Omega} \frac{1}{n-1} \sum_{j \neq i} \eta_{n,ij}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) \right. \\
&\qquad \left. - \sup_{\omega, h \in \Omega} \frac{1}{n-1} \sum_{j \neq i} \mathbb{E}\left[\eta_{n,ij}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) \middle| X, p_n \right] \middle| X, p_n \right] \\
\leq& \frac{1}{\sqrt{n-1}} \mathbb{E}\left[\sup_{\omega, h \in \Omega} \left| \mathbb{G}_n \eta_{ni}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) \right| \middle| X, p_n \right] \\
\leq& \frac{K^\eta \bar{J}^\eta \bar{W}^2}{\sqrt{n-1}} \equiv \frac{M^\eta}{\sqrt{n-1}}. \qquad \text{(S.78)}
\end{aligned}$$

For any $\omega, h \in \Omega$ and any $j \neq i$, by the mean-value theorem, we have

$$\begin{aligned}
&\mathbb{E}\left[\eta_{n,ij}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) \middle| X, p_n \right] \\
=& \left| \left| F_\varepsilon \left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \left(\omega + \frac{h}{n^\kappa}\right) \right) - F_\varepsilon \left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega \right) \right| \right. \\
&\cdot \|W_{n,ij}\|^2 \\
=& f_\varepsilon \left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \left(\omega + t_{n,ij} \frac{h}{n^\kappa}\right) \right) \frac{2(n-1)}{n-2} \left| Z_j' \Phi_{ni} \Lambda_{ni} \frac{h}{n^\kappa} \right| \|W_{n,ij}\|^2 \\
\leq& \frac{1}{n^\kappa} f_\varepsilon \left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \left(\omega + t_{n,ij} \frac{h}{n^\kappa}\right) \right) \frac{2(n-1)}{n-2} \|\Lambda_{ni} \Phi_{ni}' Z_j\| \|W_{n,ij}\|^2 \sup_{h \in \Omega} \|h\|
\end{aligned}$$

for some $t_{n,ij} \in [0, 1]$. By Assumption 1, the density f_ε is bounded. There is also a finite bound on the eigenvalues in Λ_{ni} and on the instruments $W_{n,ij}$ that does not depend on i and j. We conclude that there is a finite M with

$$\mathbb{E}\left[\eta_{n,ij}^{2}\left(\omega+\frac{h}{n^{\kappa}},\omega,\varepsilon_{ij}\right)\middle|X,p_{n}\right]\leq\frac{M}{n^{\kappa}}$$

for all $\omega, h\in\Omega$ and all i, j so that

$$\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\mathbb{E}\left[\eta_{n,ij}^{2}\left(\omega+\frac{h}{n^{\kappa}},\omega,\varepsilon_{ij}\right)\middle|X,p_{n}\right]\leq\frac{M}{n^{\kappa}}.$$

By (S.78)

$$\mathbb{E}\left[\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^{2}\left(\omega+\frac{h}{n^{\kappa}},\omega,\varepsilon_{ij}\right)\middle|X,p_{n}\right]\leq\frac{M^{\eta}}{\sqrt{n-1}}+\frac{M}{n^{\kappa}}.$$

Combining the results we obtain the upper bound

$$\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_{n}W_{ni}\left(g_{ni}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_{i}\right)-g_{ni}\left(\omega,\varepsilon_{i}\right)\right)\right\|^{2}\middle|X,p_{n}\right]\leq K\bar{J}\left(\frac{M^{\eta}}{\sqrt{n-1}}+\frac{M}{n^{\kappa}}\right). \tag{S.79}$$

Hence the upper bound in (S.75) is

$$\frac{K\bar{J}}{\delta^{2}}\left(\frac{M^{\eta}}{\sqrt{n-1}}+\frac{M}{n^{\kappa}}\right)$$

which for all $\delta>0$ can be made arbitrarily small by making n sufficiently large. We conclude that $T_{2n}=o_{p}(1)$.

Step 3: T_{3n}.

We use the delta method to derive an asymptotically linear representation of T_{3n}, from which we can see how T_{3n} contributes to the asymptotic distribution of Y_n.

By (S.72) the probability $P_{n,ij}^{*}(\omega)$ is differentiable in ω with the derivative

$$\nabla_{\omega}P_{n,ij}^{*}(\omega)=\frac{2(n-1)}{n-2}f_{\varepsilon}\left(U_{n,ij}+\frac{2(n-1)}{n-2}Z_{j}^{\prime}\Phi_{ni}\Lambda_{ni}\omega\right)\Lambda_{ni}\Phi_{ni}^{\prime}Z_{j},$$

By a Taylor series expansion of $P_{n,ij}^{*}(\omega_{ni}(\varepsilon_{i}))$ around ω_{ni}^{*}, T_{3n} can be written as

$$T_{3n}=\frac{1}{\sqrt{n(n-1)}}\sum_{i}\sum_{j\neq i}W_{n,ij}\left(\nabla_{\omega^{\prime}}P_{n,ij}^{*}\left(\omega_{ni}^{*}\right)\left(\omega_{ni}\left(\varepsilon_{i}\right)-\omega_{ni}^{*}\right)+O_{p}\left(\left\|\omega_{ni}\left(\varepsilon_{i}\right)-\omega_{ni}^{*}\right\|^{2}\right)\right)$$

By Lemma S.7, for any i, $\omega_{ni}(\varepsilon_i) - \omega_{ni}^* = O_p\left(n^{-1/2}\right)$ and $\omega_{ni}(\varepsilon_i)$ has the asymptotically linear approximation

$$\omega_{ni}(\varepsilon_i) - \omega_{ni}^* = \frac{1}{n-1}\sum_{j\neq i}\varphi_{n,ij}^{\omega}(\omega_{ni}^*, \varepsilon_{ij}) + r_{ni}^{\omega},$$

with the influence function $\varphi_{n,ij}^{\omega}$ defined in Lemma S.7. The remainder r_{ni}^{ω} satisfies $\max_{1\leq i\leq n}\|r_{ni}^{\omega}\| = o_p\left(n^{-1/2}\right)$ (Lemma S.8(ii)). Denote by $J_{ni}^{\omega}(\omega_{ni}^*)$ the $d_\theta \times T$ Jacobian matrix

$$J_{ni}^{\omega}(\omega_{ni}^*) = \frac{1}{n-1}\sum_{j\neq i} W_{n,ij}\nabla_{\omega'}P_{n,ij}^*(\omega_{ni}^*), \; i = 1, \ldots, n.$$

and by $\bar{W}_{ni}$ the average instrument vector

$$\bar{W}_{ni} = \frac{1}{n-1}\sum_{j\neq i} W_{n,ij}, \; i = 1, \ldots, n.$$

Note that both $J_{ni}^{\omega}(\omega_{ni}^*)$ and $\bar{W}_{ni}$ are bounded uniformly over i. Replacing $\omega_{ni}(\varepsilon_i) - \omega_{ni}^*$ with its asymptotically linear approximation we derive

$$\begin{aligned}
T_{3n} =& \frac{1}{\sqrt{n}}\sum_i\left(\frac{1}{n-1}\sum_{j\neq i}W_{n,ij}\nabla_{\omega'}P_{n,ij}^*(\omega_{ni}^*)\right)\left(\frac{1}{\sqrt{n-1}}\sum_{j\neq i}\varphi_{n,ij}^{\omega}(\omega_{ni}^*, \varepsilon_{ij}) + \sqrt{n-1}r_{ni}^{\omega}\right) \\
&+ \frac{1}{\sqrt{n}}\sum_i\left(\frac{1}{n-1}\sum_{j\neq i}W_{n,ij}\right)O_p\left(\sqrt{n-1}\,\|\omega_{ni} - \omega_{ni}^*\|^2\right) \\
=& T_{3n}^l + r_{1n} + r_{2n}, \qquad \text{(S.80)}
\end{aligned}$$

where

$$T_{3n}^l = \frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}J_{ni}^{\omega}(\omega_{ni}^*)\,\varphi_{n,ij}^{\omega}(\omega_{ni}^*, \varepsilon_{ij}),$$

and

$$\begin{aligned}
r_{1n} &= \sqrt{\frac{n-1}{n}}\sum_i J_{ni}^{\omega}(\omega_{ni}^*)\,r_{ni}^{\omega}, \\
r_{2n} &= \sqrt{\frac{n-1}{n}}\sum_i \bar{W}_{ni}O_p\left(\|\omega_{ni} - \omega_{ni}^*\|^2\right).
\end{aligned}$$

The term T_{3n}^{l} in (S.80) contributes to the asymptotic distribution and has an asymptotically normal distribution. It captures the random variation in $\omega_{ni}(\varepsilon_i)$. We will combine it with T_{1n} to derive the asymptotic distribution of Y_n.

The two remainder terms r_{1n} and r_{2n} are not asymptotically negligible. The sum of these terms and the fourth term T_{4n} in (S.74) is however $o_p(1)$.

Step 4: T_{4n}.

Observe that $\mathbb{E}[Y_n | X, p_n] = 0$, so

$$\begin{aligned} 0 &= \mathbb{E}[T_{1n} + T_{2n} + T_{3n} + T_{4n} | X, p_n] \\ &= \mathbb{E}\left[T_{1n} + T_{2n} + T_{3n}^{l} + r_{1n} + r_{2n} + T_{4n} \middle| X, p_n\right] \end{aligned}$$

It is clear that $\mathbb{E}[T_{1n} | X, p_n] = \mathbb{E}\left[T_{3n}^{l} \middle| X, p_n\right] = 0$. We have shown in Step 2 that $\mathbb{E}[T_{2n}^2 | X, p_n] = o(1)$, so $\mathbb{E}[T_{2n} | X, p_n] = o(1)$. This implies that

$$\mathbb{E}[r_{1n} + r_{2n} + T_{4n} | X, p_n] = \mathbb{E}[r_{1n} | X, p_n] + \mathbb{E}[r_{2n} | X, p_n] + T_{4n} = o(1).$$

Hence,

$$r_{1n} + r_{2n} + T_{4n} = r_{1n} + r_{2n} - \mathbb{E}[r_{1n} | X, p_n] - \mathbb{E}[r_{2n} | X, p_n] + o(1).$$

Note that

$$\begin{aligned} \mathbb{E}[r_{1n} | X, p_n] &= \sqrt{\frac{n-1}{n}} \sum_i J_{ni}^{\omega}(\omega_{ni}^*) \mathbb{E}[r_{ni}^{\omega} | X, p_n], \\ \mathbb{E}[r_{2n} | X, p_n] &= \sqrt{\frac{n-1}{n}} \sum_i \bar{W}_{ni} \mathbb{E}\left[O_p\left(\|\omega_{ni} - \omega_{ni}^*\|^2\right) \middle| X, p_n\right]. \end{aligned}$$

Below we show that the two centered remainders $r_{1n} - \mathbb{E}[r_{1n} | X, p_n]$ and $r_{2n} - \mathbb{E}[r_{2n} | X, p_n]$ are both $o_p(1)$.

We show $r_{1n} - \mathbb{E}[r_{1n}| X, p_n] = o_p(1)$ and the proof for $r_{2n} - \mathbb{E}[r_{2n}| X, p_n]$ is similar. By Chebyshev's inequality, for any $\delta > 0$,

$$
\begin{aligned}
&\Pr\left(\|r_{1n} - \mathbb{E}[r_{1n}| X, p_n]\| > \delta | X, p_n \right) \\
\leq &\frac{1}{\delta^2} \mathbb{E}\left(\|r_{1n} - \mathbb{E}[r_{1n}| X, p_n]\|^2 \Big| X, p_n \right) \\
= &\frac{n-1}{n\delta^2} \sum_i \mathbb{E}\left((r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n])' J_{ni}^\omega(\omega_{ni}^*)' J_{ni}^\omega(\omega_{ni}^*) (r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n]) \Big| X, p_n \right) \\
\leq &\frac{n-1}{\delta^2} \max_i \mathbb{E}\left((r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n])' J_{ni}^\omega(\omega_{ni}^*)' J_{ni}^\omega(\omega_{ni}^*) (r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n]) \Big| X, p_n \right),
\end{aligned}
$$

where the equality follows because conditional on X and p_n, r_{ni}^ω depends on ε_i only and therefore they are independent over i by Assumption 1. From Lemma S.8(ii), $\max_{1\leq i\leq n} \|r_{ni}^\omega\| = o_p\left(n^{-1/2}\right)$, by the dominated convergence theorem we have $\mathbb{E}[\max_{1\leq i\leq n} \|r_{ni}^\omega\| | X, p_n] = o\left(n^{-1/2}\right)$, so $\max_{1\leq i\leq n} \|r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n]\| \leq \max_{1\leq i\leq n} \|r_{ni}^\omega\| + \mathbb{E}[\max_{1\leq i\leq n} \|r_{ni}^\omega\| | X, p_n] = o_p\left(n^{-1/2}\right)$. This together with the boundedness of $J_{ni}(\omega_{ni}^*)$ implies that the term $(r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n])' J_{ni}(\omega_{ni}^*)' J_{ni}(\omega_{ni}^*) (r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n])$ is $o_p(n^{-1})$ uniformly over i. By the dominated convergence theorem again, we obtain

$$
\mathbb{E}\left(\max_{1\leq i\leq n} (r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n])' J_{ni}(\omega_{ni}^*)' J_{ni}(\omega_{ni}^*) (r_{ni}^\omega - \mathbb{E}[r_{ni}^\omega| X, p_n]) \Bigg| X, p_n \right) = o\left(\frac{1}{n}\right),
$$

This shows that

$$
r_{1n} - \mathbb{E}[r_{1n}| X, p_n] = o_p(1).
$$

Similarly, with $O_p\left(\|\omega_{ni} - \omega_{ni}^*\|^2\right)$ in place of r_{ni}^ω and $\bar{W}_{ni}$ in place of $J_{ni}(\omega_{ni}^*)$, and by $\max_{1\leq i\leq n} \|\omega_{ni}(\varepsilon_i) - \omega_{ni}^*\|^2 = o_p\left(n^{-1/2}\right)$ (Lemma S.8(i)), we can derive that

$$
r_{2n} - \mathbb{E}[r_{2n}| X, p_n] = o_p(1).
$$

Combining the results yields

$$
r_{1n} + r_{2n} + T_{4n} = r_{1n} + r_{2n} - \mathbb{E}[r_{1n}| X, p_n] - \mathbb{E}[r_{2n}| X, p_n] + o(1) = o_p(1).
$$

Now we return to the two main terms T_{1n} in (S.74) and T_{3n}^l in (S.80). Both are normalized averages of independent random variables. If we define the $d_\theta \times 1$ random vector

$$Y_{ni}^{l}=\frac{1}{\sqrt{n\left(n-1\right)}}\sum_{j\neq i}W_{n,ij}\left(g_{n,ij}\left(\omega_{ni}^{*},\varepsilon_{ij}\right)-P_{n,ij}^{*}\left(\omega_{ni}^{*}\right)\right)+J_{ni}^{\omega}\left(\omega_{ni}^{*}\right)\varphi_{n,ij}^{\omega}\left(\omega_{ni}^{*},\varepsilon_{ij}\right)$$

then

$$T_{1n}+T_{3n}^{l}=\sum_{i}Y_{ni}^{l}.$$

Note that conditional on X and p_n, Y_{ni}^{l}, $i=1,\ldots,n$, is an independent triangular array becuase each Y_{ni}^{l} depends on ε_i only, and ε_i, $i=1,\ldots,n$, are i.i.d. by Assumption 1. Conditional on X and p_n the Y_{ni}^{l} are not identically distributed so we have to use the Lindeberg-Feller central limit theorem (CLT) for triangular arrays to derive the asymptotic distribution of $\sum_i Y_{ni}^{l}$.

Conditional on X and p_n, Y_{ni}^{l} has mean 0. By independence of Y_{ni}^{l}, the variance of $\sum_i Y_{ni}^{l}$ is given by

$$\sum_{i}Var\left(Y_{ni}^{l}\right|X,p_{n}\Big)=\sum_{i}\mathbb{E}\left[Y_{ni}^{l}\left(Y_{ni}^{l}\right)'\right|X,p_{n}\Big]=\frac{1}{n\left(n-1\right)}\sum_{i}\sum_{j\neq i}\Sigma_{n,ij}=\Sigma_{n}$$

where $\Sigma_{n,ij}$ is defined in (S.69).

Since Y_{ni}^{l} is a vector, we verify that the conditions in the Lindeberg-Feller CLT hold for $a'\sum_i Y_{ni}^{l}$ for any vector of constants $a\in\mathbb{R}^{d_\theta}$ so that $\left(a'\Sigma_n a\right)^{-1/2}a'\sum_i Y_{ni}^{l}$ converges in distribution to $N\left(0,1\right)$. By the Cramér–Wold theorem, this implies that $\Sigma_n^{-1/2}\sum_i Y_{ni}^{l}$ converges in distribution to $N\left(0,I_{d_\theta}\right)$.

Observe that given X and p_n, $a'\sum_i Y_{ni}^{l}$ has mean 0 and variance $a'\Sigma_n a$. For the Lindeberg condition, we need to show that for any $\xi>0$,

$$\lim_{n\to\infty}\frac{1}{a'\Sigma_{n}a}\sum_{i}\mathbb{E}\left[\left(a'Y_{ni}^{l}\right)^{2}1\left\{\left|a'Y_{ni}^{l}\right|\geq\xi\sqrt{a'\Sigma_{n}a}\right\}\right|X,p_{n}\Big]=0.\qquad\text{(S.81)}$$

We have

$$\begin{aligned}&\sum_{i}\mathbb{E}\left[\left(a'Y_{ni}^{l}\right)^{2}1\left\{\left|a'Y_{ni}^{l}\right|\geq\xi\sqrt{a'\Sigma_{n}a}\right\}\right|X,p_{n}\Big]\\\leq&\mathbb{E}\left[\sum_{i}\left(a'Y_{ni}^{l}\right)^{2}1\left\{\frac{\max_{1\leq i\leq n}\left|a'Y_{ni}^{l}\right|}{\sqrt{a'\Sigma_{n}a}}\geq\xi\right\}\right|X,p_{n}\Bigg],\end{aligned}$$

Note that $\sum_i \left(a' Y_{ni}^l\right)^2$ has a finite expectation and is therefore $O_p(1)$. Hence if

$$\frac{\max_{1\leq i\leq n}\left|a' Y_{ni}^l\right|}{\sqrt{a'\Sigma_n a}} = o_p(1) \tag{S.82}$$

then

$$\sum_i \left(a' Y_{ni}^l\right)^2 1\left\{\frac{\max_{1\leq i\leq n}\left|a' Y_{ni}^l\right|}{\sqrt{a'\Sigma_n a}} \geq \xi\right\} = O_p(1)\, o_p(1) = o_p(1)$$

Finally, this random variable is bounded by $\sum_i \left(a' Y_{ni}^l\right)^2$ that has a finite expectation. We conclude that by dominated convergence the Lindeberg condition is satisfied if (S.82) holds.

By Chebyshev's inequality

$$\Pr\left(\frac{\max_{1\leq i\leq n}\left|a' Y_{ni}^l\right|}{\sqrt{a'\Sigma_n a}} \geq \xi \,\middle|\, X, p_n\right) \leq \frac{1}{\xi^2 a'\Sigma_n a}\mathbb{E}\left[\max_{1\leq i\leq n}\left(a' Y_{ni}^l\right)^2 \,\middle|\, X, p_n\right].$$

By the maximal inequality in Lemma 2.2.2 in van der Vaart and Wellner (1996),

$$\mathbb{E}\left[\max_{1\leq i\leq n}\left(a' Y_{ni}^l\right)^2 \,\middle|\, X, p_n\right] \leq K \ln(n+1) \max_{1\leq i\leq n}\left\|\left(a' Y_{ni}^l\right)^2\right\|_{\psi|X,p_n},$$

where K is a constant depending only on ψ and $\|Z\|_{\psi|X,p_n}$ is the conditional Orlicz norm of a random variable Z given X and p_n for the convex function $\psi(z) = e^z - 1$. By convexity of ψ, $\mathbb{E}(|Z|| X, p_n) \leq \|Z\|_{\psi|X,p_n}$.

Next we derive a bound on $\max_{1\leq i\leq n}\left\|\left(a' Y_{ni}^l\right)^2\right\|_{\psi|X,p_n}$. Recall that

$$a' Y_{ni}^l = \frac{1}{\sqrt{n(n-1)}}\sum_{j\neq i} a'\left(W_{n,ij}\left(g_{n,ij}\left(\omega_{ni}^*, \varepsilon_{ij}\right) - P_{n,ij}^*\left(\omega_{ni}^*\right)\right) + J_{ni}^\omega\left(\omega_{ni}^*\right)\varphi_{n,ij}^\omega\left(\omega_{ni}^*, \varepsilon_{ij}\right)\right).$$

Each term in the average is bounded by

$$\begin{aligned}
&\left|a'\left(W_{n,ij}\left(g_{n,ij}\left(\omega_{ni}^*, \varepsilon_{ij}\right) - P_{n,ij}^*\left(\omega_{ni}^*\right)\right) + J_{ni}^\omega\left(\omega_{ni}^*\right)\varphi_{n,ij}^\omega\left(\omega_{ni}^*, \varepsilon_{ij}\right)\right)\right| \\
&\leq \|a\|\left(2\left\|W_{n,ij}\right\| + \left\|J_{ni}^\omega\left(\omega_{ni}^*\right)\right\|\left\|\varphi_{n,ij}^\omega\left(\omega_{ni}^*, \varepsilon_{ij}\right)\right\|\right) \equiv M_{n,ij} \leq M_n < \infty,
\end{aligned}$$

so each normalized term in the sum has a support that is contained in

$$\left[-\frac{M_{n,ij}}{\sqrt{n\left(n-1\right)}},\frac{M_{n,ij}}{\sqrt{n\left(n-1\right)}}\right].$$

By Hoeffding's inequality for bounded random variables (Boucheron, Lugosi, and Massart, 2013, Theorem 2.8), we have for any $t>0$,

$$\Pr\left(\left|a'Y_{ni}^{l}\right|>t\right|X,p_{n}\right)\leq 2\exp\left(-\frac{n\left(n-1\right)t^{2}}{2\sum_{j\neq i}M_{n,ij}^{2}}\right).$$

Therefore, by Lemma 2.2.1 in van der Vaart and Wellner (1996),

$$\left\|\left(a'Y_{ni}^{l}\right)^{2}\right\|_{\psi|X,p_{n}}\leq\frac{6\sum_{j\neq i}M_{n,ij}^{2}}{n\left(n-1\right)},$$

Combining the results we obtain

$$\mathbb{E}\left[\max_{1\leq i\leq n}\left(a'Y_{ni}^{l}\right)^{2}\right|X,p_{n}\right]\leq\frac{6K\ln\left(n+1\right)\sum_{j\neq i}M_{n,ij}^{2}}{n\left(n-1\right)}\leq\frac{6KM_{n}\ln\left(n+1\right)}{n}\to 0,$$

so (S.82) holds and and the Lindeberg condition (S.81) is proved.

By the Lindeberg-Feller CLT

$$\frac{a'\sum_{i}Y_{ni}^{l}}{\sqrt{a'\Sigma_{n}a}}\xrightarrow{d}N\left(0,1\right).$$

Because Σ_n is a symmetric and positive-definite matrix, there is a nonsingular symmetric matrix $\Sigma_n^{1/2}$ such that $\Sigma_n^{1/2}\Sigma_n^{1/2}=\Sigma_n$. Let $\tilde{a}=\Sigma_n^{1/2}a$, then $a'\sum_i Y_{ni}=\tilde{a}'\Sigma_n^{-1/2}\sum_i Y_{ni}^l$ and $a'\Sigma_n a=\tilde{a}'\Sigma_n^{-1/2}\Sigma_n\Sigma_n^{-1/2}\tilde{a}=\tilde{a}'\tilde{a}$. Note that Σ_n is nonsingular, so $\tilde{a}$ is also an arbitrary vector in $\mathbb{R}^{d_\theta}$. The previous result then implies that

$$\tilde{a}'\Sigma_{n}^{-1/2}\sum_{i}Y_{ni}^{l}\xrightarrow{d}N\left(0,\tilde{a}'\tilde{a}\right).$$

By the Cramer-Wold device,

$$\Sigma_{n}^{-1/2}\sum_{i}Y_{ni}^{l}\xrightarrow{d}N\left(0,I_{d_{\theta}}\right).$$

where I_{d_θ} is the $d_\theta \times d_\theta$ identity matrix.

Because

$$Y_n = \sum_i Y_{ni}^l + o_p(1),$$

we conclude that by Slutsky's theorem Y_n has the asymptotic distribution

$$\Sigma_n^{-1/2} Y_n \xrightarrow{d} N(0, I_{d_\theta}).$$

■

S.3 Proofs in Section 5

Proof of Theorem 5.1. We prove the theorem in the case where both $\mathcal{T}_{i+}$ and $\mathcal{T}_{i-}$ are nonempty. The proof also holds for special cases where $\mathcal{T}_{i+}$ is empty (i.e., all the eigenvalues of $V_i(X,\sigma)$ are nonpositive) or $\mathcal{T}_{i-}$ is empty (i.e., all the eigenvalues of $V_i(X,\sigma)$ are nonnegative) without modification. Note that the latter special case has been proved in Theorem 3.2.

From Proposition 3.1, the expected utility satisfies

$$\begin{aligned}
&\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\
=&\sum_{j\neq i} G_{ij}\left(U_{ij}(X,\sigma) - \varepsilon_{ij}\right) \\
&+ \sum_t \lambda_{it}(X,\sigma) \max_{\omega_t\in\mathbb{R}} \left\{ \frac{2(n-1)}{n-2} \sum_{j\neq i} G_{ij} Z_j' \phi_{it}(X,\sigma)\omega_t - \frac{(n-1)^2}{n-2}\omega_t^2 \right\} \\
=&\max_{(\omega_t, t\in\mathcal{T}_{i+})} \min_{(\omega_t, t\in\mathcal{T}_{i-})} \sum_{j\neq i} G_{ij}\left(U_{ij}(X,\sigma) + \frac{2(n-1)}{n-2} Z_j' \sum_t \phi_{it}(X,\sigma)\lambda_{it}(X,\sigma)\omega_t - \varepsilon_{ij}\right) \\
&- \frac{(n-1)^2}{n-2} \sum_t \lambda_{it}(X,\sigma)\omega_t^2 \qquad \text{(S.83)} \\
=&\max_{(\omega_t, t\in\mathcal{T}_{i+})} \min_{(\omega_t, t\in\mathcal{T}_{i-})} \sum_{j\neq i} G_{ij}\left(U_{ij}(X,\sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_i(X,\sigma)\Lambda_i(X,\sigma)\omega - \varepsilon_{ij}\right) \\
&- \frac{(n-1)^2}{n-2}\omega'\Lambda_i(X,\sigma)\omega. \qquad \text{(S.84)}
\end{aligned}$$

The second equality in (S.83) follows because if we move an eigenvalue λ_{it} inside a maximization, it remains a maximzation if $\lambda_{it} \geq 0$ and becomes a minimization if

$\lambda_{it} < 0$. Note that the transformed expected utility is separable in each maximization, so the order of the maximizations and minimizations in (S.83) and (S.84) does not matter.

Denote by $\tilde{\Pi}\left(G_i, \omega, \varepsilon_i, X, \sigma\right)$ the objective function of the maximin problem in (S.84)

$$\begin{aligned}\tilde{\Pi}_i\left(G_i, \omega, \varepsilon_i, X, \sigma\right) = &\sum_{j \neq i} G_{ij}\left(U_{ij}\left(X, \sigma\right) + \frac{2\left(n-1\right)}{n-2} Z_j' \Phi_i\left(X, \sigma\right) \Lambda_i\left(X, \sigma\right) \omega - \varepsilon_{ij}\right) \\ &- \frac{\left(n-1\right)^2}{n-2} \omega' \Lambda_i\left(X, \sigma\right) \omega.\end{aligned}$$

We have

$$\begin{aligned}&\max_{G_i} \mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\ =&\max_{G_i} \max_{(\omega_t, t \in \mathcal{T}_{i+})} \min_{(\omega_t, t \in \mathcal{T}_{i-})} \tilde{\Pi}_i\left(G_i, \omega, \varepsilon_i, X, \sigma\right) \\ \leq& \max_{(\omega_t, t \in \mathcal{T}_{i+})} \min_{(\omega_t, t \in \mathcal{T}_{i-})} \max_{G_i} \tilde{\Pi}_i\left(G_i, \omega, \varepsilon_i, X, \sigma\right) \\ =& \max_{(\omega_t, t \in \mathcal{T}_{i+})} \min_{(\omega_t, t \in \mathcal{T}_{i-})} \Pi_i\left(\omega, \varepsilon_i, X, \sigma\right),\end{aligned} \tag{S.85}$$

where $\Pi_i\left(\omega, \varepsilon_i, X, \sigma\right)$ is defined in (5.2). The inequality follows because $\tilde{\Pi}_i\left(G_i, \omega\right) \leq \max_{G_i} \tilde{\Pi}_i\left(G_i, \omega\right)$ for all G_i and ω, so we have

$$\max_{(\omega_t, t \in \mathcal{T}_{i+})} \min_{(\omega_t, t \in \mathcal{T}_{i-})} \tilde{\Pi}_i\left(G_i, \omega\right) \leq \max_{(\omega_t, t \in \mathcal{T}_{i+})} \min_{(\omega_t, t \in \mathcal{T}_{i-})} \max_{G_i} \tilde{\Pi}_i\left(G_i, \omega\right)$$

for all G_i, and thus the maximum of the left-hand side over G_i is bounded above by the right-hand side. The last equality in (S.85) holds because for any ω, $\tilde{\Pi}_i\left(G_i, \omega\right)$ is separable in each G_{ij} so the optimal G_{ij} is given by (5.1) with $\omega_i\left(X, \varepsilon_i, \sigma\right)$ replaced by ω and $\max_{G_i} \tilde{\Pi}_i\left(G_i, \omega\right) = \Pi_i\left(\omega\right)$.

We now show that the inequality in (S.85) is an equality. Since $\omega_i\left(X, \varepsilon_i, \sigma\right)$ is a solution to the maximin problem in the last line of (S.85), similarly as in Lemma S.1

it satisfies the first-order condition

$$\begin{aligned}&\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)\\=&\frac{1}{n-1}\sum_{j\neq i}1\left\{U_{ij}(X,\sigma)+\frac{2(n-1)}{n-2}Z_j'\Phi_i(X,\sigma)\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)-\varepsilon_{ij}\geq 0\right\}\\&\qquad\cdot\Lambda_i(X,\sigma)\,\Phi_i'(X,\sigma)\,Z_j,\ \text{a.s..}\end{aligned}$$

Pre-multiplication by $\Phi_i(X,\sigma)$ gives

$$\Phi_i(X,\sigma)\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)=\frac{1}{n-1}V_i(X,\sigma)\sum_{j\neq i}G_{ij}(X,\varepsilon_i,\sigma)\,Z_j,\ \text{a.s.,}\tag{S.86}$$

where $G_{ij}(X,\varepsilon_i,\sigma)$ is given in (5.1). By the definition of $G_i(X,\varepsilon_i,\sigma)$ and $\omega_i(X,\varepsilon_i,\sigma)$, the maximin value of $\Pi(\omega,\varepsilon_i,X,\sigma)$ is given by

$$\begin{aligned}&\max_{(\omega_t,t\in\mathcal{T}_{i+})}\min_{(\omega_t,t\in\mathcal{T}_{i-})}\Pi_i(\omega,\varepsilon_i,X,\sigma)\\=&\tilde{\Pi}\left(G_i(X,\varepsilon_i,\sigma),\omega_i(X,\varepsilon_i,\sigma);X,\varepsilon_i,\sigma\right)\\=&\sum_{j\neq i}G_{ij}(X,\varepsilon_i,\sigma)\left(U_{ij}(X,\sigma)-\varepsilon_{ij}\right)\\&+\frac{2(n-1)}{n-2}\sum_{j\neq i}G_{ij}(X,\varepsilon_i,\sigma)\,Z_j'\Phi_i(X,\sigma)\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)\\&-\frac{(n-1)^2}{n-2}\omega_i(X,\varepsilon_i,\sigma)'\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma).\end{aligned}$$

Let $V_i^+(X,\sigma)$ and $\Lambda_i^+(X,\sigma)$ be the Moore-Penrose generalized inverse of $V_i(X,\sigma)$ and $\Lambda_i(X,\sigma)$, respectively. Clearly $V_i^+(X,\sigma)=\Phi_i(X,\sigma)\Lambda_i^+(X,\sigma)\Phi_i(X,\sigma)'$. The quadratic term in the last display satisfies

$$\begin{aligned}&\omega_i(X,\varepsilon_i,\sigma)'\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)\\=&\omega_i(X,\varepsilon_i,\sigma)'\Lambda_i(X,\sigma)\Lambda_i^+(X,\sigma)\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)\\=&\omega_i(X,\varepsilon_i,\sigma)'\Lambda_i(X,\sigma)\Phi_i(X,\sigma)'\Phi_i(X,\sigma)\Lambda_i^+(X,\sigma)\Phi_i(X,\sigma)'\Phi_i(X,\sigma)\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)\\=&\frac{1}{(n-1)^2}\left(V_i(X,\sigma)\sum_{j\neq i}G_{ij}(X,\varepsilon_i,\sigma)\,Z_j\right)'V_i^+(X,\sigma)\left(V_i(X,\sigma)\sum_{j\neq i}G_{ij}(X,\varepsilon_i,\sigma)\,Z_j\right),\ \text{a.s.,}\end{aligned}$$

where we have used (S.86) to derive the last equality. Therefore,

$$
\begin{aligned}
&\max_{(\omega_t, t\in\mathcal{T}_{i+})}\min_{(\omega_t, t\in\mathcal{T}_{i-})}\Pi_i\left(\omega,\varepsilon_i,X,\sigma\right)\\
=&\sum_{j\neq i}G_{ij}\left(X,\varepsilon_i,\sigma\right)\left(U_{ij}\left(X,\sigma\right)-\varepsilon_{ij}\right)\\
&+\frac{2}{n-2}\left(\sum_{j\neq i}G_{ij}\left(X,\varepsilon_i,\sigma\right)Z_j\right)'V_i\left(X,\sigma\right)\left(\sum_{j\neq i}G_{ij}\left(X,\varepsilon_i,\sigma\right)Z_j\right)\\
&-\frac{1}{n-2}\left(V_i\left(X,\sigma\right)\sum_{j\neq i}G_{ij}\left(X,\varepsilon_i,\sigma\right)Z_j\right)'V_i^{+}\left(X,\sigma\right)\left(V_i\left(X,\sigma\right)\sum_{j\neq i}G_{ij}\left(X,\varepsilon_i,\sigma\right)Z_j\right)\\
=&\sum_{j\neq i}G_{ij}\left(X,\varepsilon_i,\sigma\right)\left(U_{ij}\left(X,\sigma\right)-\varepsilon_{ij}\right)\\
&+\frac{1}{n-2}\sum_{j\neq i}\sum_{k\neq i}G_{ij}\left(X,\varepsilon_i,\sigma\right)G_{ik}\left(X,\varepsilon_i,\sigma\right)Z_j'V_i\left(X,\sigma\right)Z_k,\text{ a.s.}\\
=&\mathbb{E}\left[U_i(G_i\left(X,\varepsilon_i,\sigma\right),G_{-i},X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right],\text{ a.s..}
\end{aligned}
\tag{S.87}
$$

Combining (S.85) and (S.87) yields

$$
\begin{aligned}
&\max_{G_i}\mathbb{E}\left[U_i(G_i,G_{-i},X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right]\\
\leq&\max_{(\omega_t, t\in\mathcal{T}_{i+})}\min_{(\omega_t, t\in\mathcal{T}_{i-})}\Pi_i\left(\omega,\varepsilon_i,X,\sigma\right)\\
=&\mathbb{E}\left[U_i(G_i\left(X,\varepsilon_i,\sigma\right),G_{-i},X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right],\text{ a.s..}
\end{aligned}
$$

Because $\max_{G_i}\mathbb{E}\left[U_i(G_i,G_{-i},X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right]\geq\mathbb{E}[U_i(G_i(X,\varepsilon_i,\sigma),G_{-i},X,\varepsilon_i)\,|X,\varepsilon_i,\sigma]$, the inequality becomes an equality, and all the terms are equal. Hence, $G_i\left(X,\varepsilon_i,\sigma\right)$ is an optimal solution almost surely.

As for the uniqueness, $G_i\left(X,\varepsilon_i,\sigma\right)$ is unique almost surely because ε_i has a continuous distribution, so two link decisions achieve the same utility with probability zero. The uniqueness of $\Lambda_i\left(X,\sigma\right)\omega_i\left(X,\varepsilon_i,\sigma\right)$ follows from the uniqueness of $G_i\left(X,\varepsilon_i,\sigma\right)$, (S.86) and the invertibility of $\Phi_i\left(X,\sigma\right)$. The proof is complete. ■

Proof of Theorem 5.3. For simplicity, we omit the arguments (X,σ) (or (X_i,σ)) whenever possible. Define $\tilde{\omega}_{ni}\left(\varepsilon_i\right)=\Phi_{ni}\omega_{ni}\left(\varepsilon_i\right)$ and $\tilde{\omega}_i=\Phi_i\omega_i$. The finite-$n$ and limiting conditional choice probabilities depend on $\tilde{\omega}_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i$, respectively, i.e.,

$$P_{n,ij}(X,\sigma) = \Pr\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega}_{ni}(\varepsilon_i) - \varepsilon_{ij} \geq 0 \middle| X, \sigma \right)$$

and

$$P_{ij}(X_i, X_j, \sigma) = \Pr\left(U_{ij} + 2 Z_j' V_i \tilde{\omega}_i - \varepsilon_{ij} \geq 0 \middle| X_i, X_j, \sigma \right).$$

Notice that

$$\begin{aligned} \omega' \Lambda_{ni} \omega &= (\Phi_{ni}\omega)' V_{ni} (\Phi_{ni}\omega) \\ \omega' \Lambda_i \omega &= (\Phi_i \omega)' V_i (\Phi_i \omega). \end{aligned}$$

Since Φ_{ni} and Φ_i are nonsingular, there are one-to-one mappings between ω and $\Phi_{ni}\omega$ and between ω and $\Phi_i\omega$. Therefore, $\tilde{\omega}_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i$ are the solutions to

$$\max_{\tilde{\omega}} \tilde{\Pi}_{ni}(\tilde{\omega}, \varepsilon_i, X, \sigma) = \max_{\tilde{\omega}} \frac{1}{n-1} \sum_{j \neq i} \left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega} - \varepsilon_{ij} \right]_+ - \frac{n-1}{n-2} \tilde{\omega}' V_{ni} \tilde{\omega}$$

and

$$\max_{\tilde{\omega}} \tilde{\Pi}_i(\tilde{\omega}, X_i, \sigma) = \max_{\tilde{\omega}} \mathbb{E}\left[\left[U_{ij} + 2 Z_j' V_i \tilde{\omega} - \varepsilon_{ij} \right]_+ \middle| X_i, \sigma \right] - \tilde{\omega}' V_i \tilde{\omega},$$

respectively. The advantage of the change of variables is that we get rid of the eigenvalues and eigenvectors in the expressions so that the conditional choice probabilities and the objective functions $\tilde{\Pi}_{ni}$ and $\tilde{\Pi}_i$ only involve V_{ni} and V_i.

By the definition of $P_{n,ij}$ and P_{ij},

$$\begin{aligned} & \left| P_{n,ij}(X,\sigma) - P_{ij}(X_i, X_j, \sigma) \right| \\ \leq & \mathbb{E}\left[\left| 1\left\{ U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega}_{ni}(\varepsilon_i) \geq \varepsilon_{ij} \right\} - 1\left\{ U_{ij} + 2 Z_j' V_i \tilde{\omega}_i \geq \varepsilon_{ij} \right\} \right| \middle| X, \sigma \right] \\ \leq & \mathbb{E}\left[\left| 1\left\{ U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega}_{ni}(\varepsilon_i) \geq \varepsilon_{ij} \right\} - 1\left\{ U_{n,ij} + 2 Z_j' V_{ni} \tilde{\omega}_i \geq \varepsilon_{ij} \right\} \right| \middle| X, \sigma \right] \\ & + \mathbb{E}\left[\left| 1\left\{ U_{n,ij} + 2 Z_j' V_{ni} \tilde{\omega}_i \geq \varepsilon_{ij} \right\} - 1\left\{ U_{ij} + 2 Z_j' V_i \tilde{\omega}_i \geq \varepsilon_{ij} \right\} \right| \middle| X, \sigma \right]. \end{aligned} \tag{S.88}$$

Observe that

$$\begin{aligned}
&U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega}_{ni}(\varepsilon_i) - \left(U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i\right) \\
=&2Z_j' V_{ni} \left(\tilde{\omega}_{ni}(\varepsilon_i) - \tilde{\omega}_i\right) + \frac{2}{n-2} Z_j' V_{ni} \tilde{\omega}_{ni}(\varepsilon_i) \\
\equiv&\Delta_{ni}(\varepsilon_i)
\end{aligned}$$

so the first term in the last expression in (S.88) can be bounded by

$$\begin{aligned}
&\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X, \sigma\right) \\
&+ \Pr\left(U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i - \delta_n \leq \varepsilon_{ij} \leq U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i + \delta_n \middle| X, \sigma\right)
\end{aligned} \tag{S.89}$$

for $\delta_n > 0$. This is because if ε_{ij} lies between $U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega}_{ni}(\varepsilon_i)$ and $U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i$, and if their difference $\left|\Delta_{ni}(\varepsilon_i)\right|$ is at most δ_n, then ε_{ij} must lie between $U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i - \delta_n$ and $U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i + \delta_n$.

Given X_i and σ, $\tilde{\omega}_{ni}(\varepsilon_i) - \tilde{\omega}_i = o_p(1)$ by Lemma S.11. Since V_{ni} and $\tilde{\omega}_{ni}(\varepsilon_i) = \Phi_{ni} \omega_{ni}(\varepsilon_i)$ are bounded, we have for any $\delta_n > 0$

$$\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, \sigma\right) \to 0$$

as $n \to \infty$. By the law of iterated expectation

$$\begin{aligned}
\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, \sigma\right) &= \mathbb{E}\left[\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, X_j, \sigma\right) \middle| X_i, \sigma\right] \\
&= \sum_{t=1}^{T} \Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, X_j = x_t, \sigma\right) \Pr\left(X_j = x_t\right).
\end{aligned}$$

In the expression the expectation is taken with respect to X_j. If there is $t \in \{1, \ldots, T\}$ such that $\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, X_j = x_t, \sigma\right)$ does not converge to 0, then $\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, \sigma\right)$ cannot converge to 0. This implies that given X_i, X_j, and σ we have

$$\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, X_j, \sigma\right) \to 0.$$

Note that $X = (X_i, X_j, X_{-ij})$, where $X_{-ij} = (X_k, k \neq i, j)$. By the law of iterated expectation again

$$\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, X_j, \sigma\right) = \mathbb{E}\left[\Pr\left(\left|\Delta_{ni}(\varepsilon_i)\right| > \delta_n \middle| X_i, X_j, X_{-ij}, \sigma\right) \middle| X_i, X_j, \sigma\right],$$

where the expectation is taken respect to X_{-ij}. Because convergence in mean implies convergence in probability (by Markov's inequality), given X_i, X_j, and σ we must have

$$\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right| > \delta_n \middle| X_i, X_j, X_{-ij}, \sigma\right) = o_p\left(1\right). \tag{S.90}$$

For the second term in the bound in (S.89), by the mean-value theorem

$$\begin{aligned}
&\Pr\left(U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i - \delta_n \leq \varepsilon_{ij} \leq U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i + \delta_n \middle| X_i, X_j, X_{-ij}, \sigma\right) \\
=&F_\varepsilon\left(U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i + \delta_n\right) - F_\varepsilon\left(U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i - \delta_n\right) \\
=&2f_\varepsilon\left(U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i + t_{n,ij}\delta_n\right)\delta_n
\end{aligned}$$

for some $t_{n,ij} \in [-1, 1]$. Since the density f_ε is bounded, $f_\varepsilon\left(U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i + t_{n,ij}\delta_n\right)$ is bounded as well. We choose $\delta_n > 0$ with $\delta_n \downarrow 0$ as $n \to \infty$, so

$$\Pr\left(U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i - \delta_n \leq \varepsilon_{ij} \leq U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i + \delta_n \middle| X_i, X_j, X_{-ij}, \sigma\right) = o_p\left(1\right). \tag{S.91}$$

Combining (S.90) and (S.91), given X_i, X_j, and σ the first term in the last expression in (S.88) is $o_p\left(1\right)$

$$\begin{aligned}
\mathbb{E}\Bigg[1\left\{U_{n,ij} + \frac{2\left(n-1\right)}{n-2} Z_j' V_{ni}\tilde{\omega}_{ni}\left(\varepsilon_i\right) \geq \varepsilon_{ij}\right\} \\
- 1\left\{U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i \geq \varepsilon_{ij}\right\}\Bigg| X_i, X_j, X_{-ij}, \sigma\Bigg] = o_p\left(1\right).
\end{aligned}$$

The last term in (S.88) satisfies

$$\begin{aligned}
&\mathbb{E}\left[\left|1\left\{U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i \geq \varepsilon_{ij}\right\} - 1\left\{U_{ij} + 2Z_j' V_i\tilde{\omega}_i \geq \varepsilon_{ij}\right\}\right| X_i, X_j, X_{-ij}, \sigma\right] \\
=&\left|F_\varepsilon\left(U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i\right) - F_\varepsilon\left(U_{ij} + 2Z_j' V_i\tilde{\omega}_i\right)\right| \\
=&2f_\varepsilon\left(\tilde{t}_{n,ij}\right)\left|U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i - \left(U_{ij} + 2Z_j' V_i\tilde{\omega}_i\right)\right| \\
\leq&2f_\varepsilon\left(\tilde{t}_{n,ij}\right)\left(\left|U_{n,ij} - U_{ij}\right| + 2\left\|V_{ni} - V_i\right\|\left\|\tilde{\omega}_i\right\|\right)
\end{aligned}$$

for some $\tilde{t}_{n,ij}$ between $U_{n,ij} + 2Z_j' V_{ni}\tilde{\omega}_i$ and $U_{ij} + 2Z_j' V_i\tilde{\omega}_i$, where the second equality follows from the mean-value theorem. Since the density f_ε is bounded, and given X_i, X_j, and σ, $U_{n,ij} - U_{ij} = o_p\left(1\right)$ and $V_{ni} - V_i = o_p\left(1\right)$ by Assumption 7, the last term in (S.88) is also $o_p(1)$

$$\mathbb{E}\left[\left|1\left\{U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i\geq\varepsilon_{ij}\right\}-1\left\{U_{ij}+2Z_j'V_i\tilde{\omega}_i\geq\varepsilon_{ij}\right\}\right|\,X_i,X_j,X_{-ij},\sigma\right]=o_p(1).$$

Combining the results we conclude that given X_i, X_j, and σ

$$P_{n,ij}(X_i,X_j,X_{-ij},\sigma)-P_{ij}(X_i,X_j,\sigma)=o_p(1).$$

The proof is complete. ■

Lemma S.11 (Consistency of $\tilde{\omega}_{ni}(\varepsilon_i)$ for $\tilde{\omega}_i$) *Suppose that Assumptions 1-3 and 7 are satisfied. Given X_i and σ, $\tilde{\omega}_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i$ defined in the proof of Theorem 5.3 satisfy $\tilde{\omega}_{ni}(\varepsilon_i)-\tilde{\omega}_i=o_p(1)$, i.e., for any $\delta>0$,*

$$\Pr\left(\left\|\tilde{\omega}_{ni}(\varepsilon_i)-\tilde{\omega}_i\right\|>\delta\right|X_i,\sigma)\to 0 \tag{S.92}$$

as $n\to\infty$.

Proof. Recall that $\tilde{\omega}_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i$ are solutions to the transformed maximization problems

$$\max_{\tilde{\omega}}\tilde{\Pi}_{ni}(\tilde{\omega},\varepsilon_i,X,\sigma)=\max_{\tilde{\omega}}\frac{1}{n-1}\sum_{j\neq i}\left[U_{n,ij}+\frac{2(n-1)}{n-2}Z_j'V_{ni}\tilde{\omega}-\varepsilon_{ij}\right]_+-\frac{n-1}{n-2}\tilde{\omega}'V_{ni}\tilde{\omega}$$

and

$$\max_{\tilde{\omega}}\tilde{\Pi}_i(\tilde{\omega},X_i,\sigma)=\max_{\tilde{\omega}}\mathbb{E}\left[\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+\right|X_i,\sigma\right]-\tilde{\omega}'V_i\tilde{\omega}. \tag{S.93}$$

Because the original maximization problem in (5.8) has a unique solution ω_i by Assumption 7, the one-to-one relationship between ω_i and $\tilde{\omega}_i$ implies that $\tilde{\omega}_i$ is the unique solution to the transformed maximization problem.

Observe that

$$\frac{\partial}{\partial c}\mathbb{E}[c-\varepsilon]_+=\frac{\partial}{\partial c}\int_{-\infty}^{c}(c-\varepsilon)f_\varepsilon(\varepsilon)\,d\varepsilon=F_\varepsilon(c).$$

The first-order condition of (S.93) is given by

$$\nabla_{\tilde{\omega}}\tilde{\Pi}_i(\tilde{\omega},X_i,\sigma)=2V_i\mathbb{E}\left[Z_jF_\varepsilon\left(U_{ij}+2Z_j'V_i\tilde{\omega}\right)\right|X_i,\sigma\right]-2V_i\tilde{\omega}=0. \tag{S.94}$$

It is easy to see that any $\tilde{\omega}$ that satisfies the first-order condition must be bounded. Without loss of generality we can assume that $\tilde{\omega}_i$ is in a compact set $\tilde{\Omega}$. Since $\tilde{\Pi}_i(\tilde{\omega}, X_i, \sigma)$ is continuous in $\tilde{\omega}$, if we can further establish a uniform LLN for the objective functions, i.e.,

$$\sup_{\tilde{\omega}} \left| \tilde{\Pi}_{ni}(\tilde{\omega}, \varepsilon_i, X, \sigma) - \tilde{\Pi}_i(\tilde{\omega}, X_i, \sigma) \right| = o_p(1) \tag{S.95}$$

as $n \to \infty$, following a standard consistency proof (Newey and McFadden, 1994) we can prove (S.92).

By the triangle inequality

$$\begin{aligned}
&\sup_{\tilde{\omega}} \left| \tilde{\Pi}_{ni}(\tilde{\omega}, \varepsilon_i, X, \sigma) - \tilde{\Pi}_i(\tilde{\omega}, X_i, \sigma) \right| \\
\leq &\sup_{\tilde{\omega}} \frac{1}{n-1} \sum_{j \neq i} \left| \left[U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega} - \varepsilon_{ij} \right]_+ - \left[U_{ij} + 2Z_j' V_i \tilde{\omega} - \varepsilon_{ij} \right]_+ \right| \\
&+ \sup_{\tilde{\omega}} \left| \frac{1}{n-1} \sum_{j \neq i} \left[U_{ij} + 2Z_j' V_i \tilde{\omega} - \varepsilon_{ij} \right]_+ - \mathbb{E}\left[\left[U_{ij} + 2Z_j' V_i \tilde{\omega} - \varepsilon_{ij} \right]_+ \middle| X_i, \sigma \right] \right| \\
&+ \sup_{\tilde{\omega}} \left| \frac{n-1}{n-2} \tilde{\omega}' V_{ni} \tilde{\omega} - \tilde{\omega}' V_i \tilde{\omega} \right|.
\end{aligned} \tag{S.96}$$

Because $\left| [x]_+ - [y]_+ \right| \leq |x - y|$, the first term on the right-hand side can be bounded by

$$\begin{aligned}
&\sup_{\tilde{\omega}} \frac{1}{n-1} \sum_{j \neq i} \left| (U_{n,ij} - U_{ij}) + 2Z_j' (V_{ni} - V_i) \tilde{\omega} + \frac{2}{n-2} Z_j' V_{ni} \tilde{\omega} \right| \\
\leq &\max_{j \neq i} |U_{n,ij} - U_{ij}| + 2 \|V_{ni} - V_i\| \sup_{\tilde{\omega}} \|\tilde{\omega}\| + \frac{2}{n-2} \|V_{ni}\| \sup_{\tilde{\omega}} \|\tilde{\omega}\| \\
= &o_p(1),
\end{aligned}$$

where the last equality follows because $\max_{j \neq i} |U_{n,ij} - U_{ij}| = o_p(1)$ and $V_{ni} - V_i = o_p(1)$ by Assumption 7, and $\|V_{ni}\|$ and $\sup_{\tilde{\omega}} \|\tilde{\omega}\|$ are bounded. Similarly, we can bound the last term on the right-hand side of (S.96) by

$$\begin{aligned}
&\sup_{\tilde{\omega}} \left|\tilde{\omega}'\left(V_{ni}-V_i\right)\tilde{\omega}\right| + \sup_{\tilde{\omega}} \frac{1}{n-2}\left|\tilde{\omega}'V_{ni}\tilde{\omega}\right| \\
\leq &\left(\left\|V_{ni}-V_i\right\| + \frac{1}{n-2}\left\|V_{ni}\right\|\right)\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|^2 \\
= &o_p(1).
\end{aligned}$$

For the second term on the right-hand side of (S.96), observe that given X_i and σ, the functions $\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+$ are i.i.d. across j. These functions have an envelope

$$\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+ \leq \left[U_{ij}+2\left\|V_i\right\|\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|-\varepsilon_{ij}\right]_+, \ \forall\tilde{\omega}\in\tilde{\Omega},$$

that is integrable since

$$\begin{aligned}
&\mathbb{E}\left[\left[U_{ij}+2\left\|V_i\right\|\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|-\varepsilon_{ij}\right]_+\middle| X_i,\sigma\right] \\
\leq &\mathbb{E}\left[\left|U_{ij}+2\left\|V_i\right\|\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|-\varepsilon_{ij}\right|\middle| X_i,\sigma\right] \\
\leq &\left(\mathbb{E}\left[\left(U_{ij}+2\left\|V_i\right\|\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|-\varepsilon_{ij}\right)^2\middle| X_i,\sigma\right]\right)^{1/2} < \infty.
\end{aligned}$$

Note that $U_{ij}+2Z_j'V_i\tilde{\omega}$ is linear in ω and the function $[x]_+$ is Lipschitz in x because $\left|[x]_+-[y]_+\right| \leq |x-y|$. The function $\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+$ is therefore Lipschitz in $\tilde{\omega}$, i.e., for any $\tilde{\omega}^1,\tilde{\omega}^2\in\tilde{\Omega}$

$$\left|\left[U_{ij}+2Z_j'V_i\tilde{\omega}^1-\varepsilon_{ij}\right]_+ - \left[U_{ij}+2Z_j'V_i\tilde{\omega}^2-\varepsilon_{ij}\right]_+\right| \leq 2\left\|V_i\right\|\left\|\tilde{\omega}^1-\tilde{\omega}^2\right\|,$$

so the class of functions $\left\{\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+, \tilde{\omega}\in\tilde{\Omega}\right\}$ is a type II class as defined in Andrews (1994). It thus satisfies Pollard's entropy condition (Andrews, 1994, Theorem 2), and the uniform LLN follows

$$\sup_{\tilde{\omega}}\left|\frac{1}{n-1}\sum_{j\neq i}\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+ - \mathbb{E}\left[\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+\middle| X_i,\sigma\right]\right| = o_p(1).$$

Hence (S.95) is proved. ■

Proof of Example 5.1. We verify that under Assumption 7(i), $U_{n,ij}(X,\sigma)$ and $V_{ni}(X,\sigma)$ given in Example 5.1 satisfy Assumption 7(ii). Recall that

$$
\begin{aligned}
&U_{n,ij}(X,\sigma)-U_{ij}(X_i,X_j,\sigma)\\
=&\frac{1}{n-2}\sum_{k\neq i,j}\sigma(X_j,X_k)\beta_5(X_i,X_j,X_k)-\mathbb{E}\left[\sigma(X_j,X_k)\beta_5(X_i,X_j,X_k)\right|X_i,X_j,\sigma]\\
&-\frac{1}{n-2}Z_j'V_{ni}(X,\sigma)Z_j,
\end{aligned}
$$

and for $s,t=1,\ldots,T$,

$$
\begin{aligned}
&V_{ni,st}(X,\sigma)-V_{i,st}(X_i,\sigma)\\
=&\frac{1}{n-3}\sum_{l\neq i,j,k}\left(\sigma(x_s,X_l)\sigma(x_t,X_l)\gamma_2(X_i,x_s,x_t)\right.\\
&\qquad\left.-\mathbb{E}\left[\sigma(x_s,X_l)\sigma(x_t,X_l)\gamma_2(X_i,x_s,x_t)\right|X_i,\sigma]\right).
\end{aligned}
$$

Denote

$$
\begin{aligned}
&\Delta^U(X_i,X_j,X_k,\sigma)\\
=&\sigma(X_j,X_k)\beta_5(X_i,X_j,X_k)-\mathbb{E}\left[\sigma(X_j,X_k)\beta_5(X_i,X_j,X_k)\right|X_i,X_j,\sigma],
\end{aligned}
$$

and for $s,t=1,\ldots,T$,

$$
\begin{aligned}
&\Delta^V_{st}(X_i,x_s,x_t,X_l,\sigma)\\
=&\sigma(x_s,X_l)\sigma(x_t,X_l)\gamma_2(X_i,x_s,x_t)-\mathbb{E}\left[\sigma(x_s,X_l)\sigma(x_t,X_l)\gamma_2(X_i,x_s,x_t)\right|X_i,\sigma].
\end{aligned}
$$

We first look at $U_{n,ij}(X,\sigma) - U_{ij}(X_i, X_j, \sigma)$. Given X_i, X_j and σ, for any $\delta > 0$, by Chebyshev's inequality

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)\right| > \delta \,\middle|\, X_i,X_j,\sigma\right) \\
\leq &\frac{1}{\delta^2}\mathbb{E}\left[\left(\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)\right)^2 \,\middle|\, X_i,X_j,\sigma\right] \\
= &\frac{1}{\delta^2(n-2)^2}\sum_{k\neq i,j}\mathbb{E}\left[\left(\Delta^U(X_i,X_j,X_k,\sigma)\right)^2 \,\middle|\, X_i,X_j,\sigma\right] \to 0
\end{aligned}$$

as $n \to \infty$, where the equality follows because X_k, $k \neq i,j$, are i.i.d. (Assumption 7(i)). This proves

$$\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma) = o_p(1)$$

for any $j \neq i$. Because $\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)$ depends on j only through X_j,[27] and X_j takes only T values, we have

$$\begin{aligned}
\max_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)\right| &= \max_{t=1,\ldots,T}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,x_t,X_k,\sigma)\right| \\
&\leq \sum_{t=1}^{T}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,x_t,X_k,\sigma)\right| = o_p(1).
\end{aligned}$$

The second term in $U_{n,ij}(X,\sigma) - U_{ij}(X_i,X_j,\sigma)$ satisfies

$$\max_{j\neq i}\frac{1}{n-2}\left|Z_j' V_{ni}(X,\sigma) Z_j\right| \leq \frac{1}{n-2}\left\|V_{ni}(X,\sigma)\right\| = o_p(1),$$

because $V_{ni}(X,\sigma)$ is bounded. Therefore,

$$\begin{aligned}
&\max_{j\neq i}\left|U_{n,ij}(X,\sigma) - U_{ij}(X_i,X_j,\sigma)\right| \\
\leq &\max_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)\right| + \max_{j\neq i}\frac{1}{n-2}\left|Z_j' V_{ni}(X,\sigma) Z_j\right| = o_p(1).
\end{aligned}$$

[27] Note that $\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma) = \frac{1}{n-2}\sum_{k\neq i,j'}\Delta^U(X_i,X_{j'},X_k,\sigma)$ for any $j \neq j'$ with $X_j = X_{j'}$.

As for $V_{ni}(X,\sigma) - V_i(X_i,\sigma)$, by Chebyshev's inequality and i.i.d. X_l

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{n-3}\sum_{l\neq i,j,k}\Delta^V_{st}\left(X_i,x_s,x_t,X_l,\sigma\right)\right|>\delta\right|X_i,\sigma\Bigg) \\
\leq&\frac{1}{\delta^2}\mathbb{E}\left[\left(\frac{1}{n-3}\sum_{l\neq i,j,k}\Delta^V_{st}\left(X_i,x_s,x_t,X_l,\sigma\right)\right)^2\right|X_i,\sigma\Bigg] \\
=&\frac{1}{\delta^2\left(n-3\right)^2}\sum_{l\neq i,j,k}\mathbb{E}\left[\left(\Delta^V_{st}\left(X_i,x_s,x_t,X_l,\sigma\right)\right)^2\right|X_i,\sigma\Big]\to 0
\end{aligned}$$

as $n\to\infty$. Hence, for $s,t=1,\ldots,T$,

$$V_{ni,st}\left(X,\sigma\right)-V_{i,st}\left(X_i,\sigma\right)=o_p\left(1\right).$$

We conclude that

$$\left\|V_{ni}\left(X,\sigma\right)-V_i\left(X_i,\sigma\right)\right\|\leq\max_{s,t=1,\ldots,T}\left|V_{ni,st}\left(X,\sigma\right)-V_{i,st}\left(X_i,\sigma\right)\right|=o_p\left(1\right).$$

The proof is complete. ■