Estimation of Large Network Formation Games*

Geert Ridder† Shuyang Sheng‡

September 20, 2019

Abstract

This paper develops estimation methods for network formation models using observed data from a single large network. We characterize network formation as a simultaneous-move game with incomplete information, where we allow for utility externalities from indirect friends such as friends of friends and friends in common. As a consequence the expected utility can be nonlinear in the link choices of an agent. In a network with n members each individual faces a discrete choice problem with 2^{n-1} overlapping alternatives, which is difficult to solve without simplification. 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 allows us to derive a closed-form expression for the conditional choice probability (CCP). The closed-form CCP 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 choices is captured by a 'sufficient statistic' for this dependence. Using this CCP we propose a two-step estimation procedure that requires few assumptions on equilibrium selection, is simple to compute, and provides asymptotically valid estimators for the parameters, accounting for the dependence between the choices. Monte Carlo results show that the estimation procedure performs well, even in moderately large networks.

*We are grateful to 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, UBC, Ohio State, TAMU, Econometric Society World Congress (2015), Econometric Society Winter Meeting and Summer Meeting (2016), INET conferences on econometrics of networks at University of Cambridge (2015) and USC (2015), Berkeley/CeMMAP conference on networks at Berkeley (2016), NYU conference (2017), Cowles foundataion 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.

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

JEL Codes: C13, C31, C57, D85

1 Introduction

This paper contributes to the growing literature on the estimation of game-theoretic models of network formation.[1] The purpose of the empirical analysis is to recover the preferences of the members of the network, in particular the preferences that determine whether a member of the network 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 simplification is in a number of steps. First we assume that agents have incomplete information on unobservables when making their link choices. We assume that agents know the unobserved (by the econometrician) link utility shocks for their own potential links, but not the unobserved link utility shocks in the preference for potential links of other agents. Alternative assumptions are complete information under which agents know not just the unobserved link characteristics of their own potential links, but also those of the links for all other agents in the network, and a stronger type of incomplete information under which agents only know the unobserved utility shock for the link under consideration. The complete information models are the hardest to estimate and there is set and not point identification of the parameters of the utility function (Miyauchi (2013) and Sheng (2017)). Estimation under the stronger type of incomplete information is easier (Leung (2015)). We show that his method requires that the utility function is additively separable, so that the two types

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

of incomplete information imply the same link choices. In that case the optimal strategy is a sequence 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 link utility function depends on the choice of potential partners in a non-separable 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 the number of links in common. If links-in-common has 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 links choice is still equivalent to a sequence of (correlated) binary link choices. To obatin this equivalence we use the Legendre transform to linearize the expected utility function. This linearization introduces an auxiliary variable that depends on the unobserved link characteristics of the agent's links. This auxiliary variable is itself the solution to a (non-differentiable) optimization problem. Because of the linearization 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 (2017) 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 the auxiliary variable converges to a constant that does not depend on the unobserved link characteristics. It turns out that the dependence vanishes at the rate $\frac{1}{n}$ which implies that our two-step estimator that is based on n^2 observations on links is consistent even with this dependence, but that the dependence has to be accounted for in the calculation of 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 link-formation probability that is computationally tractable. 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 parameter $\theta_\varepsilon \in \Theta_\varepsilon \subset \mathbb{R}^{d_\varepsilon}$. The distribution 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 parameter $\theta_u = (\beta', \gamma')'$ in a compact set $\Theta_u \subset \mathbb{R}^{d_u}$.

In (2.1), $u_i(G_j, X; \theta_u)$ 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

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

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 reciprocal 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 $v_i(G_j, G_k, X; \theta_u)$. An important example is the utility derived from friends in common

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

where the first term captures the utility of friends in common that are directly connected[2] 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 $G_i(X, \varepsilon_i)$ denote individual i's link decisions, which is a mapping from i's information (X, ε_i) to a vector of links $G_i \in \mathcal{G}_i = \{0,1\}^{n-1}$. Write $G = (G_i, G_{-i})$, where $G_{-i} = (G_j)_{j \neq i}$ 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 decisions by maximizing her expected utility $\mathbb{E}[U_i(g_i, G_{-i}, X, \varepsilon_i) | X, \varepsilon_i]$ over $g_i \in \mathcal{G}_i$, where the expectation is taken with respect to the link decisions of other individuals G_{-i}. Since G_{-i} is a function of X and $\varepsilon_{-i} = \left(\varepsilon_j'\right)_{j \neq i}'$, 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 $\sigma_i(g_i | X) = \Pr(G_i(X, \varepsilon_i) = g_i | X)$ 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 $\sigma_{-i}(g_{-i} | X) = \prod_{j \neq i} \sigma_j(g_j | X)$. Let $\sigma(X) = \{\sigma_i(g_i | X), \forall g_i \in \mathcal{G}_i, \forall i\}$ denote the belief profile. For a given belief profile σ, the expected utility of individual i is given by

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

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

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

$$\mathbb{E}\left[u_i\left(G_j,X\right)\middle|X,\sigma\right]=\beta_0+X_i'\beta_1+\left|X_i-X_j\right|'\beta_2+\mathbb{E}\left[G_{ji}\middle|X,\sigma\right]\beta_3 \\ +\frac{1}{n-2}\sum_{k\neq i,j}\mathbb{E}\left[G_{jk}\middle|X,\sigma\right]\beta_4\left(X_i,X_j,X_k\right) \tag{2.5}$$

and

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

with e.g.

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

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

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

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

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

for all link decisions $g_i\in\mathcal{G}_i$ and all $i=1,\ldots,n$.

In this paper, we focus on equilibria that are symmetric in individuals' 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 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. It can be shown that there exists a symmetric equilibrium. We assume that the observed equilibrium is symmetric.

Proposition 2.1 *For any $X\in\mathcal{X}^n$, there exists a symmetric equilibrium $\sigma(X)$.*

[3] Note that i and j have the same choice probabilities, but with probability 1 different expected utilities due to different random utility shocks.

Proof. See the appendix. ■

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, 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 $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\middle| X, \sigma\right]$ in (2.6), is symmetric in j and k.[4] Moreover, 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 estimation, 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.[5]

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 type s and t, respectively,

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

The notation emphasizes that $V_{i,st}$ depends on the publicly known characteristics of all agents and on the equilibrium σ. Clearly $V_{i,st}\left(X, \sigma\right) = V_{i,ts}\left(X, \sigma\right)$. We arrange the type-specific

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

[5] In Section 5, we will discuss how to extend our approach to continuous X_i.

expected utilities of friends in common in a $T \times T$ symmetric matrix

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

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

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

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

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

and

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

The term $Z_j' V_i(X,\sigma) Z_k$ represents the additional expected utility that individual i receives if she links to both j and k and this addition 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 $V_i(X,\sigma)$ is real and symmetric, it has a real spectral decomposition

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

where $\Lambda_i(X,\sigma) = diag(\lambda_{i1}(X,\sigma), \ldots, \lambda_{iT}(X,\sigma))$ is the $T \times T$ diagonal matrix of eigenvalues $\lambda_{it}(X,\sigma) \in \mathbb{R}$, $t = 1, \ldots, T$, and $\Phi_i(X,\sigma) = (\phi_{i1}(X,\sigma), \ldots, \phi_{iT}(X,\sigma))$ is the $T \times T$ orthogonal matrix of eigenvectors $\phi_{it}(X,\sigma) \in \mathbb{R}^T$, $t = 1, \ldots, T$. 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., $\frac{1}{n-1}\sum_{j \neq i} G_{ij} Z_j' \phi_{it}(X,\sigma)$, $t = 1, \ldots, T$.

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

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

where the maximization is over ω. By choosing Y as the linear function $\frac{1}{n-1}\sum_{j\neq i} G_{ij}Z_j'\phi_{it}(X,\sigma)$, 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*

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

Proof. See the appendix. ■

To derive the optimal decision, recall that it is a link vector $G_i=(G_{ij}, j\neq i)'$ that maximizes the expected utility. Given the transformation in (3.6), if the eigenvalues $\lambda_{it}(X,\sigma)$, $t=1,\ldots,T$, are nonnegative (Assumption 3), we can move the auxiliary maximization over each $\omega_t\in\mathbb{R}$ to the beginning of the expected utility. Hence, by interchanging these auxillary maximizations with the maximization over G_i, we can derive the optimal G_i from a simple maximization with an objective function that is linear in the components of G_i. Evaluating the optimal G_i at the optimal $\omega=(\omega_1,\ldots,\omega_T)'$ solved from an auxiliary maximization, we obtain the optimal decision that maximizes the expected utility.

Theorem 3.2 establishes the validity of our approach and gives the optimal decision. In Section 5.1 we show that our approach, with some modifications, remains valid if Assumption 3 is relaxed.

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 $V_i(X,\sigma)$ is nonnegative.*

Although this assumption is not necessary for the simplification of the link decisions in Theorem 3.2, we note that the assumption will hold if link preferences have a large degree of homophily. If we define the type-specific link probability

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

and assume $\gamma_1(X_i, X_j, X_k) \equiv \gamma_1 > 0$, i.e., friends in common have positive utility, and $\gamma_2 \equiv 0$, then by (2.6)

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

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

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

We now present the solution for the link decisions.

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

$$G_{ij}(\varepsilon_i, X, \sigma) = 1\left\{U_{ij}(X,\sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_i(X,\sigma)\, \Lambda_i(X,\sigma)\, \omega_i(\varepsilon_i, X, \sigma) - \varepsilon_{ij} \geq 0\right\}, \ \forall j \neq i \tag{3.7}$$

where the $T \times 1$ *vector* $\omega_i(\varepsilon_i, X, \sigma) = (\omega_{i1}(\varepsilon_i, X, \sigma), \ldots, \omega_{iT}(\varepsilon_i, X, \sigma))'$ *is a solution to the maximization problem*

$$\begin{aligned} &\max_{\omega} \Pi_i(\omega, \varepsilon_i, X, \sigma) \\ &= \max_{\omega} \sum_{j \neq i} \left[U_{ij}(X,\sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_i(X,\sigma)\, \Lambda_i(X,\sigma)\, \omega - \varepsilon_{ij}\right]_+ - \frac{(n-1)^2}{n-2} \omega' \Lambda_i(X,\sigma)\, \omega \end{aligned} \tag{3.8}$$

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

Proof. See the appendix. ■

To understand the role and interpretation of $\omega_i(\varepsilon_i, X, \sigma)$ we consider the first-order condition of (3.8) derived in Lemma 8.1

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

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

$$\frac{2(n-1)}{n-2} Z_j' \Phi_i(X,\sigma) \Lambda_i(X,\sigma) \omega_i(\varepsilon_i, X, \sigma) = \frac{2}{n-2} \sum_{k \neq i} G_{ik}(\varepsilon_i, X, \sigma) Z_j' V_i(X,\sigma) Z_k \quad (3.9)$$

Note that the left-hand side is the component of the choice index in (3.7) associated with friends in common. The right-hand side is the normalized 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. By including this expected marginal utility into the latent utility of each link, we can account for the interaction effects among the links and each optimal link can be obtained from a binary decision as in (3.7).

From Theorem 3.2, we can further derive a reduced-form representation of the optimal decision $G_i(\varepsilon_i, X, \sigma)$ that resembles the pure-stategy Nash equilibria in entry games (Tamer (2003), Ciliberto and Tamer (2009)).

Corollary 3.3 *Suppose Assumption 1-3 are satisfied. For each i, the optimal decision $G_i(\varepsilon_i, X, \sigma) = (G_{ij}(\varepsilon_i, X, \sigma), j \neq i)' \in \{0,1\}^{n-1}$ satisfies*

$$G_{ij}(\varepsilon_i, X, \sigma) = 1\left\{U_{ij}(X,\sigma) + \frac{2}{n-2} \sum_{k \neq i} G_{ik}(\varepsilon_i, X, \sigma) Z_j' V_i(X,\sigma) Z_k \geq \varepsilon_{ij}\right\}, \ \forall j \neq i \quad (3.10)$$

and

$$\begin{aligned} &\sum_{j \neq i} G_{ij}(\varepsilon_i, X, \sigma)\left(U_{ij}(X,\sigma) + \frac{1}{n-2} \sum_{k \neq i} G_{ik}(\varepsilon_i, X, \sigma) Z_j' V_i(X,\sigma) Z_k - \varepsilon_{ij}\right) \\ \geq &\max_{\substack{g_i^l \in \{0,1\}^{n-1} \text{ s.t.} \\ g_i^l \text{ satisfies (3.10) a.s.}}} \sum_{j \neq i} g_{ij}^l \left(U_{ij}(X,\sigma) + \frac{1}{n-2} \sum_{k \neq i} g_{ik}^l Z_j' V_i(X,\sigma) Z_k - \varepsilon_{ij}\right) \end{aligned} \quad (3.11)$$

with probability 1*. Moreover, for each $g_i = (g_{ij}, j \neq i)' \in \{0,1\}^{n-1}$, define the set*

$$\mathcal{E}(g_i, X, \sigma) = \left\{\varepsilon_i \in \mathbb{R}^{n-1} : g_i \text{ satisfies both (3.10) and (3.11)}\right\}. \quad (3.12)$$

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

Proof. See the appendix. ■

From the corollary, the optimal decision $G_{ij}(\varepsilon_i, X, \sigma)$ is a solution to the simultanous discrete choice model in (3.10), where each optimal link is determined by a binary choice problem with an augmented latent utility that includes the interaction effect from friends in common as in (3.9). Inspired by the similarity between the model in (3.10) and an entry game (Tamer (2003), Ciliberto and Tamer (2009)), we can view the decision of links G_{ij}, $j \neq i$, as an "entry game" played by the $n-1$ potential links of individual i, with strategic interactions resulted from the externality of friends in common.

Like an entry game which typically has multiple equilibria, the system in (3.10) could have multiple solutions. Because individual i chooses links that maximize her expected utility, we have a natural selection mechanism. That is, among the solutions to system (3.10), i chooses the optimal $G_i(\varepsilon_i, X, \sigma)$ that gives the highest expected utility, as characterized by (3.11). The set $\mathcal{E}(g_i, X, \sigma)$ defined in (3.12) gives the collection of $\varepsilon_i \in \mathbb{R}^{n-1}$ that support g_i to be the optimal solution, i.e., $G_i(\varepsilon_i, X, \sigma) = g_i$. From Theorem 3.2 there is a unique optimal $G_i(\varepsilon_i, X, \sigma)$ that satisfies both (3.10) and (3.11) with probability 1. Therefore, the sets $\mathcal{E}(g_i, X, \sigma)$ for all $g_i \in \{0,1\}^{n-1}$ form a partition of the space of ε_i, with each set corresponding to a unique optimal decision with probability 1. These results are similar to those established in entry games (Tamer (2003), Ciliberto and Tamer (2009)) and they are useful to analyze the properties of conditional choice probabilities in Section 4.

The auxillary variable $\omega_i(\varepsilon_i, X, \sigma)$ provides an explicit expression for the correlation of link choices of an agent. Note that $\omega_i(\varepsilon_i, X, \sigma)$ is an optimal solution to the problem in (3.8), whose objective function depends on all ε_{ij}, $j \neq i$, so $\omega_i(\varepsilon_i, X, \sigma)$ is a function of all ε_{ij}, $j \neq i$. Under Assumption 1, two optimal link choices G_{ij} and G_{ik} are correlated because (1) they both depend on $\omega_i(\varepsilon_i, X, \sigma)$, as shown in (3.7), and (2) $\omega_i(\varepsilon_i, X, \sigma)$ 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 correlation 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 (3.7), the indeterminate $\omega_{it}(\varepsilon_i, X, \sigma)$ are irrelevant for the optimal link decisions. For that reason we can arbitrarily set $\omega_{it}(\varepsilon_i, X, \sigma) = 0$ if $\lambda_{it}(X, \sigma_i, \sigma) = 0$. This ensures that $\omega_i(\varepsilon_i, X, \sigma)$ is unique as stated in Theorem 3.2.

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 decisions in (3.7) reduce to

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

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

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 correlated 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.[6] 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 $X = (X_1', \ldots X_n')'$ 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 perferences $\varepsilon_i = (\varepsilon_{ij}, j \neq i)$ 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 Bayesian Nash equilibria, and among the equilibria 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 select this particular equilibrium.

Observe that the expected utility in (2.4) and the optimal link choice in (3.7) depend on the equilibrium choice probabilities σ_n through the link probabilities of each pair only. Hence it suffices to consider the link probabilities

$$p_{n,ij} = \Pr\left(G_{n,ij} = 1 \middle| X, \sigma_n\right), \; i, j = 1, \ldots, n, \; i \neq j.$$

Let $p_n = (p_{n,ij}, i, j = 1, \ldots, n, i \neq j)$ be the link probability profile. The optimal link choice

[6] 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.

$G_{n,ij}\left(\varepsilon_i, X, \theta, p_n\right)$ in (3.7) implies the structural choice probability

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

The Bayesian Nash equilibrium in (2.8) implies that for all $i \neq j$

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

Because of the symmetric equilibrium and the discrete X, each $p_{n,ij}$ depends on individuals i and j only through their types. It suffices to estimate the type-specific link probabilities

$$p_{n,st} = \Pr\left(G_{n,ij} = 1 \middle| X_i = x_s, X_j = x_t, X, p_n \right), \;\; s, t = 1, \ldots, T.$$

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 frequency that pairs with characteristics x_s and x_t form a link

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

Let $\hat{p}_n = (\hat{p}_{n,st}, s, t = 1, \ldots, T)$ be the frequency estimator of p_n. In the second step, we estimate θ based on the moment conditions implied by (4.2), with p_n replaced by $\hat{p}_n$. From now on we omit X in $P_{n,ij}\left(X, \theta, p_n\right)$. Define the sample unconditional moment function

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

where $\hat{W}_{n,ij} \in \mathbb{R}^{d_\theta}$, $i, j = 1, \ldots, n$, is a $d_\theta \times 1$ vector of instruments/weights 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

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

where $W_{n,ij} \in \mathbb{R}^{d_\theta}$, $i, j = 1, \ldots, n$, is a $d_\theta \times 1$ vector of population instruments/weights 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 $\Psi_n(\theta_0, p_n) = 0$.

Remark 4.1 (Multiple equilibria) *Two-step estimation is appropriate even with multiple equilibria, and our model in general has multiple equilibria. The reason is that the parameters of the utility function and therefore the parameters of the structural link choice in (3.7) are the same in all equilibria. Only p_n varies between equilibria. On the other hand, if we wanted to estimate θ and p_n simultaneously under the equilibrium condition in (2.8) as in one-step estimation, we would have to specify an equilibrium selection mechanism. Because we consider the scenario of a single network and all observed links are generated from the same equilibrium, there is no need for a restriction that is typically imposed in the two-step approach if we have many networks. The restriction is that all networks have the same equilibrium. For all n the first-step estimator estimates the equilibrium link probabilities, whatever that equilibrium is.*[7]

Under Assumptions 1-4, we show that $\hat{\theta}_n$ 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/weights $\hat{W}_{n,ij}$ and $W_{n,ij}$, $i, j = 1, \ldots, n$, satisfy $\max_{i,j=1,\ldots,n} \left\| \hat{W}_{n,ij} - W_{n,ij} \right\| \xrightarrow{p} 0$ and $\max_{i,j=1,\ldots,n} \|W_{n,ij}\| < \infty$. (iv) For $s, t = 1, \ldots, T$, $\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} 1\{X_i = x_s, X_j = x_t\} \xrightarrow{p} \pi_{st} > 0$ as $n \to \infty$.*

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

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

as $n \to \infty$.

Proof. See the appendix. ■

Assumptions 4(i) and 4(iii) are regularity conditions. Assumption 4(iv) imposes a mild assumption on the data generating process of X. The restriction is that the fraction of pairs of all types x_s and x_t is positive if $n \to \infty$, 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.

[7] There is a mild restriction we need to impose on the equilibrium selection: equilibrium selection does not depend on the unobservable ε_n so p_n does not depend on ε_n.

Assumption 4(ii) is an identification condition. We can think of this as a local identification condition, because it is an identification condition for θ for a given equilibrium p_n.[8] The identification condition holds with θ_0 invariant to the the particular equilibrium selected p_n for all n. Given this identification condition, we establish the consistency of $\hat{\theta}_n$ by first showing that $\hat{p}_n$ is consistent for p_n (Lemma 8.2 in the appendix) and then showing that local identification is enough to establish the consistency of $\hat{\theta}_n$.

For consistency, in addition to the identification condition, we need a uniform LLN for the sample moment $\hat{\Psi}_n(\theta, p_n)$ (Lemma 8.3 in the appendix). Note that links formed by different individuals in a single network are conditionally independent given X and p_n. This conditional independence across individuals is crucial for a LLN to hold. Moreover, because link choices depend on the number of nodes we consider the data on the links in a network with n nodes as non-identically distributed. Therefore, we establish 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 $\omega_{ni}(\varepsilon_i, \theta_0, p_n) \in \mathbb{R}^T$ 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 ω_{ni}, and then investigate how ω_{ni} 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

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

[8] The reason that we impose local identification rather than a stronger global identification condition that $\Psi_n(\theta, p) = 0$ has a unique solution (θ_0, p_n) is to allow for multiple equilibria.

We assume that $\Pi_{ni}^{*}(\omega, \theta_0, p_n)$ has a unique maximizer $\omega_{ni}^{*}(\theta_0, p_n)$

$$\omega_{ni}^{*}(\theta_0, p_n) = \arg\max_{\omega \in \mathbb{R}^T} \Pi_{ni}^{*}(\omega, \theta_0, p_n). \tag{4.9}$$

Note that $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$ depends on ε_i whereas $\omega_{ni}^{*}(\theta_0, p_n)$ is a deterministic vector given X and p_n. We can view $\omega_{ni}^{*}(\theta_0, p_n)$ as the population counterpart of $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$.

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 a Jacobian matrix $\nabla_{\omega'}\Gamma_{ni}^{*}(\omega, \theta_0, p_n)$*[9]

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

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

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

is nonsingular at $\omega_{ni}^{*}(\theta_0, p_n)$ and its inverse is bounded.

Under Assumption 5, we show that $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$ is consistent for $\omega_{ni}^{*}(\theta_0, p_n)$ and has an asymptotically linear representation as $n \to \infty$ (Lemmas 8.5-8.7 in the appendix)

$$\omega_{ni}(\varepsilon_i, \theta_0, p_n) - \omega_{ni}^{*}(\theta_0, p_n) = \frac{1}{n-1} \sum_{j \neq i} \varphi_{n,ij}^{\omega}(\omega_{ni}^{*}(\theta_0, p_n), \varepsilon_{ij}, \theta_0, p_n) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{4.10}$$

In this representation $\varphi_{n,ij}^{\omega}(\omega, \varepsilon_{ij}, \theta_0, p_n) \in \mathbb{R}^T$ is the mean 0 influence function

$$\varphi_{n,ij}^{\omega}(\omega, \varepsilon_{ij}, \theta_0, p_n) = \nabla_{\omega'}\Gamma_{ni}^{*}(\omega, \theta_0, p_n)^{+} \varphi_{n,ij}^{\pi}(\omega, \varepsilon_{ij}, \theta_0, p_n)$$

where the function $\varphi_{n,ij}^{\pi}(\omega, \varepsilon_{ij}, \theta_0, p_n) \in \mathbb{R}^T$ is defined by

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

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

Since $\omega_i^*(\theta_0, p_n)$ is deterministic, the convergence of $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$ to $\omega_{ni}^*(\theta_0, p_n)$ indicates that the correlation between links $G_{n,ij}$ and $G_{n,ik}$ vanishes as n approaches infinity. Moreover, the asymptotically linear representation in (4.10) implies that $\omega_{ni}(\varepsilon_i, \theta_0, p_n)$ converges to $\omega_{ni}^*(\theta_0, p_n)$ at the rate of $n^{-1/2}$. This rate is crucial in deriving the asymptotic distribution of the estimator. We show below that the vanishing link correlation does not affect the rate of convergence of the estimator, but it increases the asymptotic variance.

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

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), with the derivative at (θ_0, p_n) given by*

$$\nabla_{(\theta,p)} P_{n,ij}(\theta_0, p_n) = (\nabla_{\theta'} P_{n,ij}(\theta_0, p_n), \nabla_{p'} P_{n,ij}(\theta_0, p_n))'$$

(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.

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

$$\begin{aligned}&\Sigma_n(\theta_0, p_n) \\ &= \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} J_n^\theta(\theta_0, p_n)^{-1} \mathbb{E}\left[\varphi_{n,ij}^m(\varepsilon_{ij}, \theta_0, p_n) \varphi_{n,ij}^m(\varepsilon_{ij}, \theta_0, p_n)' \middle| X, p_n\right] \left(J_n^\theta(\theta_0, p_n)^{-1}\right)'\end{aligned} \tag{4.11}$$

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

$$\begin{aligned}\varphi_{n,ij}^m(\varepsilon_{ij}, \theta_0, p_n) =& \tilde{W}_{n,ij}(\theta_0, p_n)\left(g_{n,ij}(\omega_{ni}^*(\theta_0, p_n), \varepsilon_{ij}, \theta_0, p_n) - P_{n,ij}^*(\omega_{ni}^*(\theta_0, p_n), \theta_0, p_n)\right) \\ &+ \tilde{J}_{ni}^\omega(\omega_{ni}^*(\theta_0, p_n), \theta_0, p_n) \varphi_{n,ij}^\omega(\omega_{ni}^*(\theta_0, p_n), \varepsilon_{ij}, \theta_0, p_n)\end{aligned} \tag{4.12}$$

for all $i \neq j$, where $\omega_{ni}^*(\theta_0, p_n) \in \mathbb{R}^T$ is the optimal solution in (4.9). 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/weights that include the *contribution of the first-step estimates*

$$\tilde{W}_{n,ij}(\theta_0, p_n) = W_{n,ij} - \left(\frac{1}{n(n-1)} \sum_k \sum_{l \neq k} W_{n,kl} \nabla_{p'} P_{n,kl}(\theta_0, p_n) \right) Q_{n,ij} \tag{4.13}$$

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 (4.10). 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 appendix. ■

The asymptotic variance of $\hat{\theta}_n$ reflects the variability in the link choices that affect both the first-step estimate of p_n and the second-step estimate. The influence function $\varphi^m_{n,ij}$ has two components where the first captures the mentioned variability in the link choices which induces variation in the estimate of the utility function parameters, through variation in both the first and the second stage of the estimation procedure. Note that $\varphi^m_{n,ij}$, $i, j = 1, \ldots, n$, are uncorrelated. The correlation between the link choices is through the auxiliary variable ω_{ni} and this is reflected in the influence function $\varphi^\omega_{n,ij}$ that is multiplied by a Jacobian to transform to variability in the moment function, yielding the second component of $\varphi^m_{n,ij}$. The derivation of the influence function of ω_{ni} is challenging because ω_{ni} enters link choices through a non-differentiable indicator function as shown in (3.7). We deal with this non-differentiability issue by empirical process methods (see Lemma 8.10 in the appendix).

In practice, we need to choose the instrument/weight $\hat{W}_{n,ij}$ for the second step. One option is to use the weight derived from 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 $\hat{p}_n$

$$\mathcal{L}_n(\theta, \hat{p}_n) = \sum_i \sum_{j \neq i} G_{n,ij} \ln P_{n,ij}(\theta, \hat{p}_n) + (1 - G_{n,ij}) \ln(1 - P_{n,ij}(\theta, \hat{p}_n)). \tag{4.14}$$

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). So the log likelihood in (4.14) is a limited-information likelihoood.

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) we get the instrument/weight

$$\hat{W}_{n,ij}(\theta) = \frac{\nabla_\theta P_{n,ij}(\theta, \hat{p}_n)}{P_{n,ij}(\theta, \hat{p}_n)(1 - P_{n,ij}(\theta, \hat{p}_n))}, \; i, j = 1, \ldots, n, \; i \neq j. \tag{4.15}$$

This instrument/weight depends on θ. In practice, we use the instrument/weight $\hat{W}_{n,ij}\left(\tilde{\theta}\right)$ evaluated at an initial estimator $\tilde{\theta}$ that is obtained from the second step based on some initial instrument/weight, i.e., $\hat{W}_{n,ij}(\theta)$ is the $d_\theta \times 1$ vector of powers and interactions of X_i and X_j. We could also use continuous updating as in Hansen, Heaton and Yaron (1996) and treat the instrument/weight as part of the moment function when estimating θ.

Note that the instrument/weight in (4.15) depends on the derivative $\nabla_\theta P_{n,ij}(\theta, \hat{p}_n)$. In Lemma 8.4 in the appendix, we show that $P_{n,ij}(\theta, p)$ is continuous in θ and p. To establish continuity, we apply the reduced-form optimal decision in Corollary 3.3 and partition the space of ε_i into regions that correspond to each value that the vector of link choices $G_{ni}(\varepsilon_i, \theta, p)$ can take, and these regions have boundaries that are continuous in θ and p. The proof shows that the boundaries of the regions can have kinks due to the intersection with the additional inequalities in (3.11) that ensure a 1-1 mapping from the link choices of i and the sets in the partition. Therefore, $P_{n,ij}(\theta, p_n)$ can be non-differentiable in θ at certain values. We assume that $P_{n,ij}(\theta, p)$ is continuously differentiable in a neighborhood of the true parameter value (θ_0, p_n) (Assumption 6(i)) to avoid complications in deriving the asymptotic distribution of $\hat{\theta}$. In Section 5.3 we show that the link choice probability $P_{n,ij}(\theta, p)$ converges to a limiting link choice probability as $n \to \infty$, and the limiting link choice probability is differentiable in θ. This suggests that we can use the instrument/weight derived from the limiting link choice probability as a differentiable approximation for the finite-n instrument/weight in (4.15). We will come back to this after we discuss the limiting game in Section 5.3.

Remark 4.2 (Simulation) *The link choice probability $P_{n,ij}(\theta, p)$ in general has no closed form and needs to be calculated by simulation. To simulate $P_{n,ij}(\theta, p)$, we draw ε_i independently R times, and for each simulated $\varepsilon_{i,r}$, $r = 1, \ldots, R$, we compute the auxiliary variable $\omega_{ni}(\varepsilon_{i,r}, \theta, p)$ and the vector of link choices $G_{ni}(\varepsilon_{i,r}, \theta, p)$ in (3.7). Then for each $j \neq i$ we average over the R simulated $G_{n,ij}(\varepsilon_{i,r}, \theta, p)$, $r = 1, \ldots, R$, to get a simulated value of $P_{n,ij}(\theta, p)$. This simulation procedure does not affect the consistency, rate of convergence, and asymptotic normality of the estimator. Because it maintains the same correlation structure between the link choices $G_{n,ij}$ and $G_{n,ik}$ as in the data, it affects the asymptotic variance only by scaling it up by the scalar constant $1 + R^{-1}$ (Pakes and Pollard (1989)).*

Remark 4.3 (Sampled networks) *Our estimation procedure may be applicable even if a network is not fully observed. Suppose that we observe all the n individuals in a network, but due to data collecting limitations for each individual we only observe a random sample of $d < n - 1$ potential links, yielding nd total links in the data. In this sampled link scenario, we can apply the proposed two-step estimation by estimating p_n in the first step and θ in the second step using the nd observed links. Notice that the link choice probability $P_{n,ij}(\theta, p)$ needs to be calculated using the entire X, which is available because we observe all the nodes. In this case, the asymptotic results still hold, except that the rate of convergence becomes $\sqrt{nd}$ for an increasing d and $\sqrt{n}$ for a fixed d.*

A more challenging scenario is when we only observe a random sample of the n individuals in a network. Suppose that we only observe $n_1 < n$ randomly selected individuals in a network, yielding a sample of $n_1(n_1 - 1)$ links. In this sampled node scenario, we can still apply the two-step estimation procedure using the observed nodes and links, but the complication is that $P_{n,ij}(\theta, p)$ has to be approximated by the link choice probability $P_{n_1,ij}(\theta, p)$ calculated based on the n_1 observed nodes. Under the additional assumptions of i.i.d. X_i and $\frac{n_1}{n} \to \rho > 0$, we expect that the utility terms $U_{n_1,ij}$ and $V_{n_1 i}$ and the auxiliary variable $\omega_{n_1 i}$ based on the n_1 nodes are close enough to $U_{n,ij}$, V_{ni}, and ω_{ni} so that $P_{n_1,ij}(\theta, p)$ is a good estimate of $P_{n,ij}(\theta, p)$ when n is large. A in-depth asymptotic analysis with sampled nodes *is left to future research.*

5 Extensions

In this section, we discuss a number of extensions of our method that will be explored in future research.

5.1 General V_i

We first examine how to relax Assumption 3 that requires that the matrix $V_i(X, \sigma)$ is positive semi-definite. This imposes restrictions on the utility attached to friends in common and implicitly on the degree of homophily in the network. In fact, Assumption 3 is not crucial to our approach. Without this assumption, however, the auxiliary variable is solved from a more complicated 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*

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

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

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

with $\mathcal{T}_+ = \{t : \lambda_{it}(X, \sigma) > 0\}$ *and* $\mathcal{T}_- = \{t : \lambda_{it}(X, \sigma) < 0\}$. *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 appendix. ■

Theorem 5.1 extends Theorem 3.2 and shows that for a general $V_i(X, \sigma)$ we can still represent an optimal link choice as binary choices, if an auxiliary variable $\omega_i(\varepsilon_i, X, \sigma)$ is included. This $\omega_i(\varepsilon_i, X, \sigma)$ is solved from an optimization problem, with the same objective function as in (3.8). The only difference is that we solve for $\omega_i(\varepsilon_i, X, \sigma)$ from a maximim problem over $\omega \in \mathbb{R}^T$, where we maximize over components of ω that correspond to the positive eigenvalues of $V_i(X, \sigma)$, while we minimize over components of ω that correspond to the negative eigenvalues of $V_i(X,\sigma)$. If an eigenvalue is 0 the objective function does not depend on the corresponding component of ω and we set the component to 0.

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 has as components the expected utility of friends in common which depend on the types of the agents that i links to. In the case considered here (see Section 3)

$$V_i = \begin{bmatrix} V_{i,11} & V_{i,12} \\ V_{i,12} & V_{i,22} \end{bmatrix} = \gamma_1 \begin{bmatrix} p_{i,11}^2 & p_{i,12}p_{i,21} \\ p_{i,12}p_{i,21} & p_{i,22}^2 \end{bmatrix}$$

Suppose $V_{i,12}>0$, so being friends with two individuals of different types that are friends yields a positive expected utility. Let λ_{i1} and λ_{i2} be the eigenvalues of V_i with $\lambda_{i1}\geq\lambda_{i2}$, and 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]

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

$$\Lambda_i\omega_i = \Lambda_i\Phi_i'\frac{1}{n-1}\sum_{j\neq i}G_{ij}Z_j, \text{ a.s.}$$

If both λ_{i1} and λ_{i2} are nonzero, each component of ω_i is given by

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

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}. Since $\phi_{i,11}\phi_{i,12}>0$, ω_{i1} is large (in absolute value) if i has many friends, irrespective of type. On the other hand, because $\phi_{i,21}\phi_{i,22}<0$, ω_{i2} is large (in absolute value) if i has friends that are of one type. Intuitively, ω_{i1} captures the preference for overall friends irrespective of type, and ω_{i2} captures the preference for a circle of friends of the same type.

[10] 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 $\lambda_{i1}>\max\{V_{i,11},V_{i,22}\}$ and $\lambda_{i2}<\min\{V_{i,11},V_{i,22}\}$. 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 $\phi_{i,21}\phi_{i,22}<0$.

By $V_i = \Phi_i \Lambda_i \Phi_i'$ and the first-order condition in (5.3), the expected utility in (3.2) can be expressed as

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

From the last expression, we can see how the expected utility is related to the value of ω_i. If an eigenvalue is positive, to maximize the expected utility we should choose the largest (absolute) value for the corresponding component of ω_i. If an eigenvalue is negative, the expected utility is maximized if the corresponding component of ω_i is equal to 0.

In our case, because $V_{i,11} > 0$, the first eigenvalue $\lambda_{i1} > 0$, so in problem (5.2) we maximize over ω_{i1}. If friends in common have a positive contribution to the expected utility ($\gamma_1 > 0$), the maximization over ω_{i1} reflects and individual's preference for more friends irrespective of type, which also leads to more friends in common.

If $V_{i,11}V_{i,22} > V_{i,12}^2$, the second eigenvalue is also positive $\lambda_{i2} > 0$, so in (5.2) we maximize over ω_{i2} as well. As argued above ω_{i2} is large if friends of the same type are preferred. If $V_{i,11}V_{i,22} < V_{i,12}^2$, the eigenvalue $\lambda_{i2} < 0$ is negative, so that in (5.2) we minimize over ω_{i2}. In this case, agents prefer a balanced circle of friends.

In a special case where $V_{i,11} = V_{i,22} = 0$, 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,

$$\begin{aligned}\omega_{i1} &= \frac{1}{\sqrt{2}\,(n-1)}\sum_{j\neq i} G_{ij}\left(Z_{j1} + Z_{j2}\right) \\ \omega_{i2} &= \frac{1}{\sqrt{2}\,(n-1)}\sum_{j\neq i} G_{ij}\left(Z_{j1} - Z_{j2}\right)\end{aligned}$$

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 and chooses her friends in each type to be same.

5.2 Undirected Networks

In this section we examine the case of undirected networks. Let G_{ij} now denote an undirected link between i and j and G the adjacency matrix of an 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 identification of parameters is less straightforward.

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 common 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 other agents, S_{-i}, when choosing S_i. Given X, let $\sigma_i\left(s_i\middle|X\right)$ be the conditional probability that agent i proposes s_i given X and let $\sigma(X) = \left\{\sigma_i\left(s_i\middle|X\right), s_i\in\{0,1\}^{n-1}, i = 1,\ldots,n\right\}$ be the belief profile. For a belief profile σ, the expected utility of agent i is given by

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

where

$$\mathbb{E}\left[S_{ji}u_i\left(G_j,X\right)\middle|X,\sigma\right]=\mathbb{E}\left[S_{ji}\middle|X,\sigma\right]\left(\beta_0+X_i'\beta_1+\left|X_i-X_j\right|'\beta_2\right)$$
$$+\frac{1}{n-2}\sum_{k\neq i,j}\mathbb{E}\left[S_{ji}S_{jk}\middle|X,\sigma\right]\mathbb{E}\left[S_{kj}\middle|X,\sigma\right]\beta_3\left(X_i,X_j,X_k\right) \tag{5.5}$$

and

$$\mathbb{E}\left[S_{ji}S_{ki}v_i\left(G_j,G_k,X\right)\middle|X,\sigma\right]$$
$$=\mathbb{E}\left[S_{ji}S_{jk}\middle|X,\sigma\right]\mathbb{E}\left[S_{ki}S_{kj}\middle|X,\sigma\right]\gamma_1\left(X_i,X_j,X_k\right)$$
$$+\frac{1}{n-3}\sum_{l\neq i,j,k}\mathbb{E}\left[S_{ji}S_{jl}\middle|X,\sigma\right]\mathbb{E}\left[S_{ki}S_{kl}\middle|X,\sigma\right]\mathbb{E}\left[S_{lj}S_{lk}\middle|X,\sigma\right]\gamma_2\left(X_i,X_j,X_k\right) \tag{5.6}$$

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. It is easily checked that (5.6) is symmetric in j and k if $\gamma_1\left(X_i,X_j,X_k\right)$ and $\gamma_2\left(X_i,X_j,X_k\right)$ are symmetric in X_j and X_k. The expected utility in (5.4) is the sum of a linear and quadratic functions in S_i, just as in (2.4)-(2.6). Note that the coefficients in these functions can depend on the probability that an agent proposes a link to two other agents simultaneously, as in $\mathbb{E}\left[S_{ji}S_{jk}\middle|X,\sigma\right]$.

As in a directed network, a Bayesian Nash equilibrium $\sigma^*\left(X\right)$ is a fixed point such that $\sigma_i^*\left(s_i\middle|X\right)=\Pr\left(S_i=s_i\middle|X,\sigma^*\left(X\right)\right)$, $s_i\in\{0,1\}^{n-1}$, $i=1,\ldots,n$. One could argue that the Bayesian Nash equilibrium is not an appropriate equilibrium concept for undirected networks. This concern is understandable because if there is complete information, i.e., agents know the idiosyncratic link preferences ε_{ij} not just for themselves, but also for all other agents, then the usual equilibrium concept is pairwise stability (Jackson and Wolinsky (1996)), because agents in such an environment coordinate their link proposals (see e.g., de Paula et al. (2017), Sheng (2018), Menzel (2017)). However, under the incomplete information assumption made in the paper agents are unable to coordinate their link proposals because they cannot anticipate what the other agents will propose. They can at most coordinate on the choice probabilities of their link proposals. In fact, there could be a trivial Bayesian Nash equilibrium that yields an empty network; If an agent believes that no other agent will propose a link, it is (weakly) optimal for the agent not to propose a link either, so proposing no link is a Bayesian Nash equilibrium. We assume that agents can coordinate so that this equilibrium is not selected.

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 again make Assumption 2, so that X takes T values x_t,

$t = 1, \ldots, T$. For $s, t = 1, \ldots, T$, we define $V_{i,st}^{u}(X, \sigma)$ as the coefficient of $S_{ij}S_{ik}$ in (5.4) if j is of type s and k is of type t

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

$V_{i,st}^{u}(X, \sigma)$ is the expected utility of friends in common (j and k) 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, $V_{i,st}^{u}(X, \sigma)$ is symmetric in s and t. Let $V_i^u(X, \sigma)$ denote the symmetric $T \times T$ matrix with components $V_{i,st}^{u}(X, \sigma)$

$$V_i^u(X,\sigma) = \begin{bmatrix} V_{i,11}^{u}(X,\sigma) & \cdots & V_{i,1T}^{u}(X,\sigma) \\ \vdots & & \vdots \\ V_{i,T1}^{u}(X,\sigma) & \cdots & V_{i,TT}^{u}(X,\sigma) \end{bmatrix}$$

Let $\lambda_{it}^{u}(X, \sigma)$, $t = 1, \ldots, T$, be the eigenvalues of the matrix $V_i^u(X, \sigma)$ and $\phi_{it}^{u}(X, \sigma)$, $t = 1, \ldots, T$, the corresponding eigenvectors. Further, $\Lambda_i^u(X, \sigma) = diag\left(\lambda_{i1}^{u}(X, \sigma), \ldots, \lambda_{iT}^{u}(X, \sigma)\right)$ and $\Phi_i^u(X, \sigma) = \left(\phi_{i1}^{u}(X, \sigma), \ldots, \phi_{iT}^{u}(X, \sigma)\right)$. The following corollary follows immediately from Theorem 5.1 and gives the optimal proposal decision 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*

$$S_{ij}(\varepsilon_i, X, \sigma) = 1\left\{U_{ij}^{u}(X,\sigma) + \frac{2(n-1)}{n-2}Z_j'\Phi_i^u(X,\sigma)\Lambda_i^u(X,\sigma)\omega_i^u(\varepsilon_i, X, \sigma) - \sigma_{ji}\varepsilon_{ij} \geq 0\right\}, \ \forall j \neq i \tag{5.7}$$

almost surely, where the $T \times 1$ *vector* $\omega_i^u(\varepsilon_i, X, \sigma) = (\omega_{it}^{u}(\varepsilon_i, X, \sigma), t = 1, \ldots, T)'$ *is a solution to the maximin problem*

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

with

$$U_{ij}^{u}(X,\sigma) = \mathbb{E}\left[S_{ji}u_i(G_j, X)\middle| X, \sigma\right] - \frac{1}{n-2}Z_j'V_i^u(X,\sigma)Z_j,$$

$\sigma_{ji} = \mathbb{E}\left[S_{ji}\middle| X, \sigma\right]$, $\mathcal{T}_+ = \{t : \lambda_{it}^{u}(X, \sigma) > 0\}$, *and* $\mathcal{T}_- = \{t : \lambda_{it}^{u}(X, \sigma) < 0\}$. *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.*

Corollary 5.2 shows that an optimal link proposal is a binary choice if we include an auxiliary variable $\omega_i^u(\varepsilon_i, X, \sigma)$, which plays the same role as $\omega_i(\varepsilon_i, X, \sigma)$ in a directed network.

Now we discuss how to extend the two-step procedure in Section 4 to estimate the parameters using data from an undirected network. Since the expected utility in (5.4)-(5.6) depends on the conditional probability that an agent i proposes a link with j as well as on the conditional probability that i proposes simultaneously links with j and k, these conditional probabilities have to be estimated in the first step. In the sequel agent i is of type r, i.e., $X_i = x_r$, agent j is of type s and agent k is of type t. As in a directed network the proposals S and equilibrium σ depend on the network size n so we add the subscript n hereafter.

In the first step we need to estimate

$$\begin{aligned} p_{n,rs} &= \Pr\left(S_{n,ij} = 1 \,| X_i = x_r, X_j = x_s, X, \sigma_n\right) \\ q_{n,rst} &= \Pr\left(S_{n,ij} = 1, S_{n,ik} = 1 \,| X_i = x_r, X_j = x_s, X_k = x_t, X, \sigma_n\right). \end{aligned}$$

Because $\Pr\left(S_{n,ij} = 1, S_{n,ik} = 1 \,| X, \sigma_n\right)$ is symmetric in j and k, $q_{n,rst}$ is symmetric in s and t. These conditional probabilities are not directly estimable because we do not observe the proposals. Nevertheless, we can treat them as parameters and estimate them together with the utility function parameters using observed links G_{ij}. The conditional link probabilities $p_{n,st}$ are organized in the $T \times T$ matrix p_n and the $q_{n,rst}$ in the 3-dimensional array q_n.

Define the conditional probability that agent i of type r links with agent j of type s by $\pi_{n,rs}$. Also define the conditional probability that i links simultaneously with j and k by $\psi_{n,rst}$

$$\begin{aligned} \pi_{n,rs} &= \Pr\left(G_{n,ij} = 1 \,| X_i = x_r, X_j = x_s, X, \sigma_n\right) \\ \psi_{n,rst} &= \Pr\left(G_{n,ij} = 1, G_{n,ik} = 1 \,| X_i = x_r, X_j = x_s, X_k = x_t, X, \sigma_n\right) \end{aligned}$$

Because $G_{ij} = S_{ij} S_{ji}$, and S_i, $i = 1, \ldots, n$, are independent given X, p_n, and q_n, we have

$$\begin{aligned} \pi_{n,rs} &= p_{n,rs} p_{n,sr}, & r \le s = 1, \ldots, T \\ \psi_{n,rst} &= q_{n,rst} p_{n,sr} p_{n,tr}, & r, s \le t = 1, \ldots, T \end{aligned} \tag{5.8}$$

The conditional probabilities on the left-hand side can be estimated by relative frequencies. Because $\pi_{n,rs}$ is symmetric in r and s and $\psi_{n,rst}$ is symmetric in s and t, we have $\frac{1}{2}T(T+1) + \frac{1}{2}T^2(T+1) = \frac{1}{2}T(T+1)^2$ equations in the $T^2 + \frac{1}{2}T^2(T+1)$ unknowns $p_n = (p_{n,rs}, r \le s = 1, \ldots, T)$ and $q_n = (q_{n,rst}, r, s \le t = 1, \ldots, T)$, so for $T \ge 2$ additional restrictions are needed.

We obtain these restrictions from the optimal link proposal decision in (5.7) that is a function of p_n and q_n and depends on σ_n only through these probabilities. The optimal proposal of a link from an agent i to an agent j is therefore $S_{n,ij}(\varepsilon_i, X, p_n, q_n, \theta)$. Define the conditional choice probabilities

$$P_{n,rs}(X, p_n, q_n, \theta) = \Pr(S_{n,ij} = 1 \,|X_i = x_r, X_j = x_s, X, p_n, q_n)$$
$$Q_{n,rst}(X, p_n, q_n, \theta) = \Pr(S_{n,ij} = 1, S_{n,ik} = 1 \,|X_i = x_r, X_j = x_s, X_k = x_t, X, p_n, q_n)$$

They satisfy the equations

$$\begin{aligned} \pi_{n,rs} &= P_{n,rs}(X, p_n, q_n, \theta) P_{n,sr}(X, p_n, q_n, \theta), & r \le s = 1, \ldots, T \\ \psi_{n,rst} &= Q_{n,rst}(X, p_n, q_n, \theta) P_{n,sr}(X, p_n, q_n, \theta) P_{n,tr}(X, p_n, q_n, \theta), & r, s \le t = 1, \ldots, T \end{aligned} \quad (5.9)$$

If we combine the equations in (5.8) and (5.9) we have $T(T+1)^2$ equations in $T^2 + \frac{1}{2}T^2(T+1) + \dim\theta$ unknowns. For instance, if $T = 2$ we have 18 equations in $10 + \dim\theta$ unknowns. For $T = 3$ the number of equations is 48 with $27 + \dim\theta$ unknowns. Therefore, it is possible to identify the parameters θ, if the number of types T is large enough (i.e., $\frac{1}{2}T(T^2 + T + 2) > \dim\theta$). An in-depth analysis of this procedure is left to future research.

5.3 Limiting Game

In this section, we investigate the asymptotic properties of our 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 game provides a useful asymptotic approximation for the finite-n game. It could be used to deal with complicated issues in the estimation and inference based on the finite-n game, such as non-differentiability and computational complexity.

Recall that the probability that individual i forms a link to j conditional on characteristic profile X and equilibrium σ is

$$\begin{aligned} & P_{n,ij}(X, \sigma) \\ &= \Pr(G_{n,ij}(\varepsilon_i, X, \sigma) = 1 |\, X, \sigma) \\ &= \Pr\left(U_{n,ij}(X, \sigma) + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni}(X, \sigma) \Lambda_{ni}(X, \sigma) \omega_{ni}(\varepsilon_i, X, \sigma) - \varepsilon_{ij} \ge 0 \,\middle|\, X, \sigma \right). \end{aligned} \quad (5.10)$$

For any equilibrium profile σ, if we assume that X_i, $i = 1, \ldots, n$, are i.i.d. and the utility specification satisfies that $U_{n,ij}(X, \sigma)$ and $V_{ni}(X, \sigma)$ converges to some limits $U_{ij}(X_i, X_j, \sigma)$ and $V_i(X_i, \sigma)$ as $n \to \infty$, as formally stated in Assumption 7, we expect that the link formation probability $P_{n,ij}(X,\sigma)$ has a limit given by

$$P_{ij}(X_i,X_j,\sigma)=\Pr\left(U_{ij}(X_i,X_j,\sigma)+2Z_j'\Phi_i(X_i,\sigma)\Lambda_i(X_i,\sigma)\omega_i(X_i,\sigma)-\varepsilon_{ij}\geq 0\middle| X_i,X_j,\sigma\right) \tag{5.11}$$

as $n\to\infty$, where $\Lambda_i(X_i,\sigma)$ and $\Phi_i(X_i,\sigma)$ are the eigenvalue and eigenvector matrices of $V_i(X_i,\sigma)$, $\omega_i(X_i,\sigma)$ is a maximizer of the problem

$$\begin{aligned}&\max_{\omega}\Pi_i(\omega,X_i,\sigma)\\ =&\max_{\omega}\mathbb{E}\left(\left[U_{ij}(X_i,X_j,\sigma)+2Z_j'\Phi_i(X_i,\sigma)\Lambda_i(X_i,\sigma)\omega-\varepsilon_{ij}\right]_+\middle| X_i,\sigma\right)-\omega'\Lambda_i(X_i,\sigma)\omega.\end{aligned} \tag{5.12}$$

The expectation in (5.12) 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 $U_{n,ij}(X,\sigma)$ and $V_{ni}(X,\sigma)$ defined in (3.3) and (), there exist $U_{ij}(X_i,X_j,\sigma)$ and $V_i(X_i,\sigma)$ such that for any i and any X_i and X_j, $\max_{j\neq i}|U_{n,ij}(X,\sigma)-U_{ij}(X_i,X_j,\sigma)|=o_p(1)$ and $V_{ni}(X,\sigma)-V_i(X_i,\sigma)=o_p(1)$. (iii) For any X_i and σ, $\Pi_i(\omega,X_i,\sigma)$ defined in (5.12) has a unique maximizer $\omega_i(X_i,\sigma)$.*

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

$$P_{n,ij}(X,\sigma)-P_{ij}(X_i,X_j,\sigma)=o_p(1) \tag{5.13}$$

as $n\to\infty$.

Proof. See the appendix. ■

We refer to $P_{ij}(X_i,X_j,\sigma)$ defined in (5.11) as the limiting choice probability. It is understood as the choice probability derived from a "limiting game" with a continuum of players, where each individual i forms a link with j following a "limiting strategy" given by

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

Similar to $\omega_{ni}(\varepsilon_i,X,\sigma)$ in the finite-n game, $\omega_i(X_i,\sigma)$ in the limiting strategy represents a sufficient statistic that captures the interaction effects due to friends in common. With the inclusion of $\omega_i(X_i,\sigma)$ individual i myopically chooses to form a link as in a binary choice problem. As shown in the proof of Theorem 5.11, $\omega_i(X_i,\sigma)$ is the limit of $\omega_{ni}(\varepsilon_i,X,\sigma)$ as $n\to\infty$ under i.i.d. X_i, and it no longer depends on the entire X, just X_i.

Assumption 7(i) is crucial in achieving the convergence of the CCP. Unlike the asymptotic analysis in Section 4 where we treat X as fixed and consider the limit conditional on X, here we treat X_i, $i=1,\ldots,n$, as random so that in the limit $X_{-ij}=(X_k, k\neq i,j)$ are integrated out and the limiting CCP depends only on X_i and X_j.[11] Assumption 7(ii) is a high-level assumption we impose on the utility specification. In the following example, we verify that under i.i.d. X_i, Assumption 7(ii) is satisfied for our utility specification in Section 2. Assumption 7(iii) is a standard identification condition.

Example 5.1 *Consider the expected utility in (2.5)-(2.6). For any X and σ, we have*

$$\begin{aligned} U_{n,ij}(X,\sigma)=&\beta_0+X_i'\beta_1+|X_i-X_j|'\beta_2+\sigma_{ji}(X_j,X_i)\beta_3\\ &+\frac{1}{n-2}\sum_{k\neq i,j}\sigma_{jk}(X_j,X_k)\beta_4(X_i,X_j,X_k)-\frac{1}{n-2}Z_j'V_{ni}(X,\sigma)Z_j \end{aligned}$$

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

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

In the appendix we verify that under i.i.d. X_i (Assumption 7(i)) the $U_{n,ij}(X,\sigma)$ and $V_{ni}(X,\sigma)$ satisfy Assumption 7(ii), with the limits $U_{ij}(X_i,X_j,\sigma)$ and $V_i(X_i,\sigma)$ given by

$$\begin{aligned} U_{ij}(X_i,X_j,\sigma)=&\beta_0+X_i'\beta_1+|X_i-X_j|'\beta_2+\sigma_{ji}(X_j,X_i)\beta_3\\ &+\mathbb{E}\left[\sigma_{jk}(X_j,X_k)\beta_4(X_i,X_j,X_k)\right|X_i,X_j,\sigma] \end{aligned}$$

and $V_i(X_i,\sigma)=(V_{i,st}(X_i,\sigma), s,t=1,\ldots,T)$, where

$$\begin{aligned} V_{i,st}(X_i,\sigma)=&\sigma_{jk}(x_s,x_t)\sigma_{kj}(x_t,x_s)\gamma_1(X_i,x_s,x_t)\\ &+\mathbb{E}\left[\sigma_{jl}(x_s,X_l)\sigma_{kl}(x_t,X_l)\gamma_2(X_i,x_s,x_t)\right|X_i,\sigma]. \end{aligned}$$

Recall that in Section 4 we propose a two-step GMM estimator where the instrument/weight may depend on the derivative of the finite-n CCP $P_{n,ij}(X,\sigma)$, while $P_{n,ij}(X,\sigma)$ can be non-differentiable in θ at certain values. Observe that the limiting CCP $P_{ij}(X_i,X_j,\sigma)$ depends on $\omega_i(X_i,\sigma)$, which, unlike $\omega_{ni}(\varepsilon_i,X,\sigma)$, no longer involves ε_i. Under our utility specification it is easy to show that $P_{ij}(X_i,X_j,\sigma)$ is differentiable in θ. Therefore, we can get around the non-differentiability issue by replacing the finite-n CCP with the limiting CCP in the instrument/weight. Given that the finite-n CCP converges to the limiting CCP, we

[11] We could relax Assumption 7(i) to allow for random but non-i.i.d. X_i, but for simplicity we do not pursue this general setting in the paper.

expect that the instrument/weight constructed based on the latter would be asymptotically as efficient as the instrument/weight based on the former.

One may further attempt to approximate the moment condition in (4.4) with the limiting CCP in replace of the finite-n CCP. This approximation can have a big advantage in computation, especially for large n, because the limiting CCP can be calculated analytically without simulation, though it leads to a misspecified moment condition. Because the finite-n CCP converges to the limiting CCP, we expect that the misspecification vanishes as n approaches infinity. However, to establish asymptotic properties based on the approximated moment condition, we have to further investigate the asymptotic behaviors of the finite-n CCP as it converges to the limiting CCP. An additional complication is the presence of multiple equilibria. Unlike estimation based on the true finite-n CCP in Section 4 which requires no assumption on equilibrum selection, for estimation based on the limiting CCP to be consistent, we need to assume that the sequence of equilibrium selection mechanisms along n yields a sequence of equilibria that converges to a certain equilibrium in the limiting game, similar in spirit to the convergence of equilibria in Menzel (2016). In particular, this assumption rules out cases where equilibrium selection oscillates between equilibria as n increases. A theoretical asymptotic analysis based on the limiting approximation is beyond the scope of the paper and is left to future research, but we provide some simulation evidence on the asymptotic properties in our simulation study.

6 Simulation

In this section, we implement the proposed methods in a simulation study. We focus on directed networks and assume the following utility specification

$$
\begin{aligned}
&U_i(G, X, \varepsilon_i; \theta) \\
&= \textstyle\sum_{j\neq i} G_{ij} \left(\beta_0 + X_i\beta_1 + |X_i - X_j|\,\beta_2 + \dfrac{1}{n-2}\textstyle\sum_{k\neq i,j} G_{jk}\beta_3 + \dfrac{1}{n-2}\textstyle\sum_{k\neq i,j} G_{ik}G_{jk}G_{kj}\gamma - \varepsilon_{ij} \right)
\end{aligned}
$$

where X_i is a binary random variable taking values in $\{0, 1\}$ with equal probability and ε_{ij} is standard normal $N(0, 1)$. The true values of the parameters are $(\beta_0, \beta_1, \beta_2, \beta_3, \gamma) = (-1, 1, -2, 1, 1)$. The networks are generated according to the n-player incomplete information game described in Section 2, with n taking different values of $10, 25, 50, 100, 250$, and 500. For each value of n, we generate a single network and use it to estimate the parameters by two-step MLE or GMM. Each experiment is repeated 100 times. We report the means and standard errors of the estimated parameters.

Since the limiting game approximates the finite game asymptotically, we first use the limiting game to estimate the parameters and check how well such estimates could perform. In particular, we construct the likelihood as in (4.14), with the finite n CCP replaced by the limiting CCP given in (5.11). Such approximation has a substantial advantage in computation because the limiting CCP has a probit-type closed form. The estimates are reported in Table 1. It is not surprising that the estimates in small networks perform poorly, as the limiting game is only an asymptotic approximation of the finite game. Nevertheless, when the network size gets large (e.g. $n \geq 100$), the estimates become close to the truth. This indicates that the limiting game is a valid approximation of the finite game asymptotically and estimation based on this approximation may yield good estimates for the parameters if networks are sufficiently large.

Next we estimate the parameters by the finite game, and compare the estimates with those from the limiting game. Note that the finite n CCP does not have a closed form because $\omega_{ni}(\varepsilon_i)$ depends on ε_i. It needs to be computed by simulation. In practice, we simulate the finite n CCP by a frequency simulator, where in each simulation we generate a vector ε_i and for this ε_i we solve for the optimal decision numerically. We obtain the optimal decision by solving the integer programming problem directly when $n \leq 100$ or applying the equivalent binary representation as in (3.7) and solving for $\omega_{n,i}(X, \varepsilon_i)$ when $n > 100$. Table 2 reports the MLE estimates and Tables 3 and 4 report the GMM estimates. By correctly specifying the choice probability, we improve the estimates in small networks. The estimation precision is also improved. For example, the 95% quantiles of γ from the finite game are closer to the truth, especially in small networks. These results indicate that we should use the finite game rather than the limiting game for estimation for n relatively small. We want to point out that the small network performance of the finite-game estimates is still unsatisfactory. To deliver satisfactory estimates, we need to have large networks or conduct some bias correction for the estimates.

7 Conclusion

In this paper, we provide estimation methods for network formation using observed data from a single large network. We model network formation as a simultaneous-move game with private information and extend Leung (2015) by allowing for nonseparable utility such as the effect of friends in common. The main innovation is to provide an approach to explicitly represent the pure strategy of an individual in the game. This closed form representation enables us to analyze the asymptotic features of the game as the number of players approaches infinity and thus construct asymptotically valid estimators for the parameters. We propose a two-step estimation procedure which makes little assumption about equilibrium selection

Table 1: MLE Estimates Using the Limiting Game

n	β_0	β_1	β_2	β_3	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.

Table 2: MLE Estimates Using the Finite Game

n	β_0	β_1	β_2	β_3	γ
10	−1.042	1.143	−2.312	0.901	0.891
	(0.674)	(0.563)	(0.920)	(0.656)	(1.297)
25	−1.005	1.059	−2.123	0.906	0.969
	(0.122)	(0.269)	(0.324)	(0.379)	(0.290)
50	−1.009	0.995	−1.997	1.024	1.011
	(0.063)	(0.154)	(0.148)	(0.186)	(0.221)
100	−0.990	0.992	−2.013	1.006	0.990
	(0.028)	(0.064)	(0.060)	(0.105)	(0.095)
250	−0.995	1.002	−2.004	1.017	0.985
	(0.010)	(0.024)	(0.023)	(0.046)	(0.035)
500	−0.999	1.014	−2.004	0.994	0.982
	(0.006)	(0.014)	(0.014)	(0.027)	(0.022)
DGP	−1	1	−2	1	1

Note: Mean estimates and standard errors from 100 repeated samples using the finite game, where the CCPs are computed from 500 simulations by either solving integer programming (for $n \leq 100$) or applying Legendre transform (for $n > 100$).

Table 3: GMM Estimates Using the Finite Game (Weights from the Finite Game)

n	β_0	β_1	β_2	β_3	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 game, with the GMM weights simulated independently from the finite game. The CCPs are computed from 500 simulations by either solving integer programming (for $n \leq 100$) or applying Legendre transform (for $n > 100$).

Table 4: GMM Estimates Using the Finite Game (Weights from the Limiting Game)

n	β_0	β_1	β_2	β_3	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 game, with the GMM weights simulated independently from the limiting game. The CCPs are computed from 500 simulations by either solving integer programming (for $n \leq 100$) or applying Legendre transform (for $n > 100$).

and is computationally simple. Our approach can apply to both directed and undirected networks. We focus on discrete observables in this paper, but expect our approach could be extended to continuous observables.

Future work: relax the assumption of i.i.d. unobservables. Suppose there is an individual effect α_i. Suppose it is a fixed effect, so can be correlated with X. Treat α_i as parameters. We can estimate α_i and θ jointly. Bias-correction is needed. We leave this extension to the paper by Ridder and Sheng (2017) where the focus is to deal with the endogeneity in social interactions.

8 Appendix

In the proofs $\|.\|$ is the Euclidean norm. If the argument is a matrix A the norm is the matrix Euclidean norm $\|A\| = \sqrt{tr\,(A'A)}$.

8.1 Proofs in Section 2

Proof of Proposition 2.1. Define the set of symmetric $\sigma(X)$

$$\Sigma^s(X) = \left\{\sigma(X) \in [0,1]^{n2^{n-1}} : \sigma_i(X) = \sigma_j(X) \text{ if } X_i = X_j\right\}$$

It is clear that $\Sigma^s(X)$ is a convex and compact subset of $[0,1]^{n2^{n-1}}$. Equations in (2.8) forms a mapping from $\Sigma^s(X)$ to $\Sigma^s(X)$, because if $\sigma \in \Sigma^s(X)$, then $\Pr(G_i = g_i | X, \sigma(X)) = \Pr(G_j = g_j | X, \sigma(X))$ for $X_i = X_j$ and $g_i = g_j$ with (g_{ii}, g_{ij}) swapped with (g_{jj}, g_{ji}), so $\Pr(G_i = g_i | X, \sigma(X))$ is also symmetric. The mapping is continuous in σ because the expected utilities are continuous in $\sigma(X)$ and ε_i has a continuous distribution under Assumption 1. By Brouwer's fixed point theorem there is a fixed point. ■

8.2 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 doublesum 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

$$\begin{aligned}
&\mathbb{E}\left[U_i(G_i,G_{-i},X,\varepsilon_i)\right| X,\varepsilon_i,\sigma] \\
=&\sum_{j\neq i} G_{ij}\left(U_{ij}(X,\sigma)-\varepsilon_{ij}\right) \\
&+\sum_{t=1}^{T}\lambda_{it}(X,\sigma)\max_{\omega_t\in\mathbb{R}}\left\{\frac{2(n-1)}{n-2}\sum_{j\neq i} G_{ij}Z_j'\phi_{it}(X,\sigma)\omega_t-\frac{(n-1)^2}{n-2}\omega_t^2\right\} \\
=&\max_{\omega\in\mathbb{R}^T}\sum_{j\neq i} G_{ij}\left(U_{ij}(X,\sigma)+\frac{2(n-1)}{n-2}Z_j'\sum_{t=1}^{T}\phi_{it}(X,\sigma)\lambda_{it}(X,\sigma)\omega_t-\varepsilon_{ij}\right) \\
&-\frac{(n-1)^2}{n-2}\sum_{t=1}^{T}\lambda_{it}(X,\sigma)\omega_t^2 \\
=&\max_{\omega\in\mathbb{R}^T}\sum_{j\neq i} G_{ij}\left(U_{ij}(X,\sigma)+\frac{2(n-1)}{n-2}Z_j'\Phi_i(X,\sigma)\Lambda_i(X,\sigma)\omega-\varepsilon_{ij}\right) \\
&-\frac{(n-1)^2}{n-2}\omega'\Lambda_i(X,\sigma)\omega \qquad (8.1)
\end{aligned}$$

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

Denote by $\tilde{\Pi}_i(G_i, \omega, \varepsilon_i, X, \sigma)$ the objective function of the maximization problem in (8.1)

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

From (8.1), the maximized expected utility can be derived from

$$\begin{aligned} &\max_{G_i} \mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right] \\ =&\max_{G_i} \max_{\omega} \tilde{\Pi}_i(G_i, \omega, \varepsilon_i, X, \sigma) \\ =&\max_{\omega} \max_{G_i} \tilde{\Pi}_i(G_i, \omega, \varepsilon_i, X, \sigma) \\ =&\max_{\omega} \Pi_i(\omega, \varepsilon_i, X, \sigma) \end{aligned} \tag{8.2}$$

where $\Pi_i(\omega, \varepsilon_i, X, \sigma)$ is defined in (3.8). The second equality in (8.2) 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, i.e., $\max_{G_i} \max_\omega \tilde{\Pi}_i(G_i, \omega) \geq \max_\omega \max_{G_i} \tilde{\Pi}_i(G_i, \omega)$. The last equality follows from the definition of $\Pi_i(\omega, \varepsilon_i, X, \sigma)$. The result in (8.2) shows that the maximum expected utility can be obtained by solving the last maximization problem in (8.2) or equivalently (3.8).

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

$$\begin{aligned} &\max_{\omega} \Pi_i(\omega, \varepsilon_i, X, \sigma) \\ =&\tilde{\Pi}_i(G_i(X, \varepsilon_i, \sigma), \omega_i(X, \varepsilon_i, \sigma); X, \varepsilon_i, \sigma) \\ \leq&\max_{\omega} \tilde{\Pi}_i(G_i(X, \varepsilon_i, \sigma), \omega, \varepsilon_i, X, \sigma) \\ =&\mathbb{E}\left[U_i(G_i(X, \varepsilon_i, \sigma), G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right] \end{aligned} \tag{8.3}$$

where the last equality comes from (8.1). Combining (8.2) and (8.3) we see that the inequality in (8.3) becomes an equality. Therefore, $\mathbb{E}\left[U_i(G_i(X, \varepsilon_i, \sigma), G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right] = \max_{G_i} \mathbb{E}[U_i(G_i, G_{-i}, X, \varepsilon_i) | X, \varepsilon_i, \sigma]$, and $G_i(X, \varepsilon_i, \sigma)$ 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 (8.3) implies that $\tilde{\Pi}_i(G_i(X, \varepsilon_i, \sigma), \omega_i(X, \varepsilon_i, \sigma); X, \varepsilon_i, \sigma) = \max_\omega \tilde{\Pi}_i(G_i(X, \varepsilon_i, \sigma), \omega, \varepsilon_i, X, \sigma)$, i.e., $\omega_i(X, \varepsilon_i, \sigma)$ 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

$$\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma) = \frac{1}{n-1}\Lambda_i(X,\sigma)\,\Phi_i'(X,\sigma)\sum_{j\neq i} G_{ij}(X,\varepsilon_i,\sigma)\,Z_j \tag{8.4}$$

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 8.1 *Suppose Assumption 1-3 are satisfied. An $\omega_i(\varepsilon_i, X, \sigma)$ that solves the maximization problem in (3.8) 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,[12] 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

$$\begin{aligned}&\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega - \varepsilon_{ij} > 0\right\}\Lambda_i\Phi_i'Z_j - \Lambda_i\omega \\ &= -\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega - \varepsilon_{ij} = 0\right\} diag(\tau)\,\Lambda_i\Phi_i'Z_j,\end{aligned} \tag{8.5}$$

for some $\tau = (\tau_1, \ldots, \tau_T) \in [0,1]^T$. Define the right-hand side of (8.5) as $\Delta_n(\omega, \varepsilon_i)$. For any ω,

$$\begin{aligned}&\Pr\left(\left\|\Delta_n(\omega,\varepsilon_i)\right\| > 0 \middle| X,\sigma\right) \\ &\leq \Pr\left(\exists j\neq i, U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega = \varepsilon_{ij}\middle| X,\sigma\right) \\ &\leq \sum_{j\neq i}\Pr\left(U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega = \varepsilon_{ij}\middle| X,\sigma\right) = 0,\end{aligned} \tag{8.6}$$

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

because ε_{ij} has a continuous distribution. Hence the first-order condition (8.5) holds with $\Delta_n(\omega, \varepsilon_i)$ replaced by 0 with probability 1. By (8.6) 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 8.1, $\omega_i(\varepsilon_i)$ is a solution to the first-order condition

$$\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega \geq \varepsilon_{ij}\right\}\Lambda_i\Phi_i'Z_j = \Lambda_i\omega, \text{ a.s..} \tag{8.7}$$

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 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

$$g_{ij}(\omega; \varepsilon_{ij}) = 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega \geq \varepsilon_{ij}\right\}, \ \forall j \neq i. \tag{8.8}$$

The first-order condition (8.7) defines a system of equations over ω

$$\Lambda_i\omega = \frac{1}{n-1}\sum_{j\neq i} g_{ij}(\omega; \varepsilon_{ij})\Lambda_i\Phi_i'Z_j, \text{ a.s..} \tag{8.9}$$

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

$$\omega_i(g_i) = \frac{1}{n-1}\sum_{j\neq i} g_{ij}\Phi_i'Z_j. \tag{8.10}$$

We derive a system of equations over g_i

$$g_{ij} = 1\left\{U_{ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_i\Lambda_i\omega_i(g_i) \geq \varepsilon_{ij}\right\}, \ \forall j \neq i. \tag{8.11}$$

We show that with probability 1 there is a one-to-one mapping between the solutions to (8.9) and the solutions to (8.11).

First, for any local solution $\omega_i^l(\varepsilon_i)$ that solves system (8.9), 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 (8.11) with probability 1. To see this, note that by the first-order condition in (8.9) and the definition of $\omega_i(g_i)$ in (8.10)

$\omega_i^l(\varepsilon_i)$ satisfies

$$\begin{aligned}\Lambda_i\omega_i^l(\varepsilon_i) &= \frac{1}{n-1}\sum_{j\neq i} g_{ij}\left(\omega_i^l(\varepsilon_i);\varepsilon_{ij}\right)\Lambda_i\Phi_i' Z_j, \text{ a.s.} \\ &= \Lambda_i\omega_i\left(g_i\left(\omega_i^l(\varepsilon_i);\varepsilon_i\right)\right). \end{aligned} \tag{8.12}$$

Then by the definition of $g_i(\omega;\varepsilon_i)$ in (8.8), 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 (8.11) with probability 1. Second, for two distinct $\omega_i^{l_1}(\varepsilon_i)\neq\omega_i^{l_2}(\varepsilon_i)$, with probability 1 we must have $g_i\left(\omega_i^{l_1}(\varepsilon_i);\varepsilon_i\right)\neq g_i\left(\omega_i^{l_2}(\varepsilon_i);\varepsilon_i\right)$, otherwise by (8.12) we derive the contradiction that $\omega_i^{l_1}(\varepsilon_i)=\omega_i^{l_2}(\varepsilon_i)$. Therefore, there is a one-to-one mapping between the solutions to (8.9) and (8.11) with probability 1.

The equivalence between systems (8.9) and (8.11) motivates us to analyze the relationship between $\omega_i(\varepsilon_i)$ and ε_i through the relationship between the solutions to (8.11) and ε_i. By the definition of $\omega_i(g_i)$ in (8.10), write system (8.11) explicitly as

$$g_{ij}=1\left\{U_{ij}+\frac{2}{n-2}\sum_{k\neq i} g_{ik}Z_j'V_iZ_k\geq\varepsilon_{ij}\right\},\ \forall j\neq i,\ \text{a.s.} \tag{8.13}$$

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

$$\mathcal{E}_i^l(g_i)=\left\{\varepsilon_i\in\mathbb{R}^{n-1}: g_i \text{ satisfies } (8.13)\right\}. \tag{8.14}$$

This set can be regarded as the collection of ε_i that support g_i as a solution to (8.13). 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 (8.13) may have multiple solutions, resembling the presence of multiple equilibria in entry games (Tamer (2003), Ciliberto and Tamer (2009)). In particular, it is possible that the sets in (8.14) for two different g_i and g_i' overlap and in the overlapping area both g_i and g_i' satisfy (8.13). For example, assume that all elements in V_i are positive so link choices are strategic complementarity. 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 (8.13).

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 (8.8) 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 (8.13). For a given $\varepsilon_i \in \mathbb{R}^{n-1}$, system (8.13) 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 (8.9), 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

$$\begin{aligned}
&\Pi_i\left(\omega^l, \varepsilon_i\right) \\
=& \sum_{j \neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right) \left(U_{ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_i \Lambda_i \omega^l - \varepsilon_{ij} \right) \\
&- \frac{(n-1)^2}{n-2} \left(\frac{1}{n-1} \sum_{j \neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right) \Lambda_i \Phi_i' Z_j \right)' \omega^l, \text{ a.s.} \\
=& \sum_{j \neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right) \left(U_{ij} + \frac{n-1}{n-2} Z_j' \Phi_i \Lambda_i \omega^l - \varepsilon_{ij} \right) \\
=& \sum_{j \neq i} g_{ij}\left(\omega^l; \varepsilon_{ij}\right) \left(U_{ij} + \frac{1}{n-2} \sum_{k \neq i} g_{ik}\left(\omega^l; \varepsilon_{ik}, \theta, p\right) Z_j' V_i Z_k - \varepsilon_{ij} \right), \text{ a.s.} \qquad (8.15)
\end{aligned}$$

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 $\omega_i(\varepsilon_i)$ we have $\Pi_i(\omega_i(\varepsilon_i),\varepsilon_i) \geq \Pi_i\left(\omega^l,\varepsilon_i\right)$ for all ω^l that solves the first-order condition (8.9). By the represetation of $\Pi_i\left(\omega^l,\varepsilon_i\right)$ in (8.15) and the equivalence between (8.9) (thus (8.7)) and (8.13), if $G_i(\varepsilon_i)$ takes a value $g_i \in \{0,1\}^{n-1}$, then g_i must satisfy both (8.13) and

$$\sum_{j\neq i} g_{ij}\left(U_{ij}+\frac{1}{n-2}\sum_{k\neq i} g_{ik}Z_j'V_iZ_k-\varepsilon_{ij}\right) \geq \sum_{j\neq i} g_{ij}^l\left(U_{ij}+\frac{1}{n-2}\sum_{k\neq i} g_{ik}^l Z_j'V_iZ_k-\varepsilon_{ij}\right) \tag{8.16}$$

almost surely, for all g_i^l that solve (8.13). Note that with probability 1 there is a unique $G_i(\varepsilon_i)$ because the system of equations obtained from (8.16) with "$\leq$" in replace of "=" holds with probability 0 (i.e., two distinct link decisions yield the same Π_i with probability 0). We can view (8.16) as a selection criterion that determines which solution to (8.13) is selected.

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

$$\mathcal{E}_i(g_i) = \left\{\varepsilon_i \in \mathbb{R}^{n-1} : g_i \text{ satisfies both (8.13) and (8.16)}\right\}. \tag{8.17}$$

This set is the collection of ε_i that support g_i as the unique optimal decision, i.e., for any $\varepsilon_i \in \mathcal{E}_i(g_i)$, we have $G_i(\varepsilon_i) = g_i$. The uniqueness implies that if $g_i \neq g_i'$, the sets $\mathcal{E}_i(g_i)$ and $\mathcal{E}_i(g_i')$ are disjoint. The collection of such sets $\mathcal{E}_i(g_i)$ 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 (Tamer (2003), Ciliberto and Tamer (2009)). The proof is complete. ■

8.3 Proofs in Section 4

8.3.1 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 $\hat{\theta}_n \notin \Theta \backslash \mathcal{B}_\delta(\theta_0)$, or equivalently, $\hat{\theta}_n \in \mathcal{B}_\delta(\theta_0)$. Therefore,

$$\begin{aligned}
&\Pr\left(\left\|\hat{\theta}_n - \theta_0\right\| < \delta \middle| X, p_n\right) \\
\geq &\Pr\left(\left\|\Psi_n\left(\hat{\theta}_n, p_n\right)\right\| < \inf_{\theta \in \Theta \backslash \mathcal{B}_\delta(\theta_0)} \left\|\Psi_n(\theta, p_n)\right\| \middle| X, p_n\right).
\end{aligned} \tag{8.18}$$

Because by Assumption 4(i)-(ii) and Lemma 8.4

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

the right-hand side in (8.18) goes to 1, if

$$\left\|\Psi_n\left(\hat{\theta}_n, p_n\right)\right\| = o_p(1). \tag{8.19}$$

Now by the triangle inequality

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

By the uniform LLN in Lemma 8.3 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

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

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

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

where $P_{n,(st)}(\theta, p)$ represents the value of $P_{n,ij}(\theta, p)$ if $X_i = x_s$ and $X_j = x_t$. By Lemma 8.4 $P_{n,(st)}(\theta, p)$ 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 $\sup_{\theta\in\Theta} \left\|\hat{\Psi}_n(\theta, p) - \hat{\Psi}_n(\theta, p_n)\right\|$ is continuous in p at p_n.

By the continuous mapping theorem and the consistency of $\hat{p}_n$ in Lemma 8.2,

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

as $n \to \infty$, so (8.19) holds and weak consistency is proven. ■

Lemma 8.2 (Consistency of $\hat{p}_n$) *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

$$\mathbb{E}\left[(G_{n,ij} - p_{n,st}) \mid X, p_n\right] 1\{X_i = x_s, X_j = x_t\} = 0 \tag{8.20}$$

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

$$\begin{aligned} \mathbb{E}\left[\|\hat{p}_n - p_n\|^2 \middle| X, p_n \right] &= \mathbb{E}\left[\sum_s \sum_t \left(\hat{p}_{n,st} - p_{n,st} \right)^2 \middle| X, p_n \right] \\ &= \sum_s \sum_t \mathbb{E}\left[\left(\hat{p}_{n,st} - p_{n,st} \right)^2 \middle| X, p_n \right]. \end{aligned} \tag{8.21}$$

We can write

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

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

$$\begin{aligned} &\frac{1}{n^2 (n-1)^2} \sum_i \sum_{j \neq i} \mathbb{E}\left[\left(G_{n,ij} - p_{n,st} \right)^2 \middle| X, p_n \right] 1\left\{ X_i = x_s, X_j = x_t \right\} \\ &+ \frac{1}{n^2 (n-1)^2} \sum_i \sum_{j \neq i} \sum_{k \neq i,j} \mathbb{E}\left[\left(G_{n,ij} - p_{n,st} \right) \left(G_{n,ik} - p_{n,st} \right) \middle| X, p_n \right] 1\left\{ X_i = x_s, X_j = x_t, X_k = x_t \right\} \\ &+ \frac{1}{n^2 (n-1)^2} \sum_i \sum_{j \neq i} \sum_{k \neq i} \sum_{l \neq k} \mathbb{E}\left[\left(G_{n,ij} - p_{n,st} \right) \left(G_{n,kl} - p_{n,st} \right) \middle| X, p_n \right] \\ &\qquad \cdot 1\left\{ X_i = x_s, X_j = x_t, X_k = x_s, X_l = x_t \right\} \end{aligned} \tag{8.22}$$

Because the link choices are independent between individuals, the last term in (8.22) is 0 by (8.20). Further,

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

and

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

so the numerator in (8.22) is bounded by

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

Because $\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} 1\{X_i = x_s, X_j = x_t\}$ converges to a strictly positive limit q_{st} by Assumption 4(iv), the denominator of the conditional variance of $\hat{p}_{n,st} - p_{n,st}$ converges to q_{st}^2. Therefore, the conditional variance of $\hat{p}_{n,st} - p_{n,st}$ is $o(1)$ 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 8.3 (Uniform LLN for Sample Moments) *For any $\delta > 0$,*

$$\Pr\left(\sup_{\theta \in \Theta}\left\|\hat{\Psi}_n(\theta, p_n) - \Psi_n(\theta, p_n)\right\| > \delta \middle| X, p_n\right) \to 0 \tag{8.23}$$

as $n \to \infty$.

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

$$\begin{aligned}
&\hat{\Psi}_n(\theta, p_n) - \Psi_n(\theta, p_n) \\
=&\frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \hat{W}_{n,ij}\left(G_{n,ij} - P_{n,ij}(\theta, p_n)\right) - W_{n,ij}\left(\mathbb{E}\left[G_{n,ij} \middle| X, p_n\right] - P_{n,ij}(\theta, p_n)\right) \\
=&\frac{1}{n(n-1)} \sum_i \sum_{j \neq i}\left(\hat{W}_{n,ij} - W_{n,ij}\right)\left(G_{n,ij} - P_{n,ij}(\theta, p_n)\right) \\
&+ \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} W_{n,ij}\left(G_{n,ij} - \mathbb{E}\left[G_{n,ij} \middle| X, p_n\right]\right)
\end{aligned} \tag{8.24}$$

The first term in the last expression in (8.24) is $o_p(1)$ uniformly over $\theta \in \Theta$ because

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

by Assumption 4(iii). Write the last term in (8.24) 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

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

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

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

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 8.4 (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

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

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 $\omega_{ni}(\varepsilon_i,\theta,p)$ would vary in 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, so $P_{n,ij}(\theta,p)$ can be expressed as

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

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 (8.17)

$$\mathcal{E}_i(g_{ni},\theta,p) = \left\{\varepsilon_i \in \mathbb{R}^{n-1} : g_{ni} \text{ satisfies both (8.13) and (8.16)}\right\}.$$

For any $g_{ni} \in \{0,1\}^{n-1}$, the equations in (8.13) define an orthant in $\mathbb{R}^{n-1}$

$$\varepsilon_{ij}\begin{cases} < U_{n,ij}(\theta,p) + \frac{2}{n-2}\sum_{k\neq i} g_{n,ik} Z_j' V_{ni}(\theta,p) Z_k & \text{if } g_{n,ij}=1 \\ \geq U_{n,ij}(\theta,p) + \frac{2}{n-2}\sum_{k\neq i} g_{n,ik} Z_j' V_{ni}(\theta,p) Z_k & \text{if } g_{n,ij}=0 \end{cases}, \ \forall j\neq i \tag{8.25}$$

Since both $U_{n,ij}(\theta,p)$ and $V_{ni}(\theta,p)$ are continuous in θ and p, the boundary of this orthant is continuous in θ and p. Moreover, the inequalities in (8.16) define half-spaces in $\mathbb{R}^{n-1}$ given by the hyperplanes

$$\sum_{j\neq i}\left(g_{n,ij} - g_{n,ij}^l\right)\varepsilon_{ij} \leq \sum_{j\neq i}\left(g_{n,ij} - g_{n,ij}^l\right) U_{n,ij}(\theta,p) + \frac{1}{n-2}\sum_{k\neq i}\left(g_{n.ik}^l - g_{n,ik}^l\right) Z_j' V_{ni}(\theta,p) Z_k, \tag{8.26}$$

for all g_{ni}^l that solve (8.13) with probability 1. Because the right-hand side of (8.26) 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 (8.25) and the half-spaces defined by (8.26). 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

$$
\begin{aligned}
P_{n,ij}(\theta, p) &= \sum_{g_{ni} \in \{0,1\}^{n-1}} \int_{\mathcal{E}_i(g_{ni},\theta,p)} 1\left\{ U_{n,ij}(\theta,p) + \frac{2}{n-2}\sum_{k\neq i} g_{n,ik} Z_j' V_{ni}(\theta,p) Z_k \geq \varepsilon_{ij} \right\} f_{\varepsilon_i}(\varepsilon_i;\theta)\, d\varepsilon_i \\
&= \sum_{\substack{g_{ni} \in \{0,1\}^{n-1} \\ g_{n,ij}=1}} \int_{\mathcal{E}_i(g_{ni},\theta,p)} f_{\varepsilon_i}(\varepsilon_i;\theta)\, d\varepsilon_i \qquad (8.27)
\end{aligned}
$$

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 (8.27) is continuous in θ and p. Summing up the integrals, we conclude that $P_{n,ij}(\theta, p)$ is continuous in θ and p. The proof is complete. ■

8.3.2 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 8.5). Moreover, $\omega_{ni}(\varepsilon_i)$ has an asymptotically linear representation (Lemma 8.7) and satisfies certain uniformity properties (Lemma 8.8). Additional results that are needed to prove these lemmas are in Lemma 8.6 and 8.9.

Remark 8.1 *By Lemma 8.1 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

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

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 8.5 (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 $\delta > 0$*

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

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 $\Pi^*_{ni}(\omega_{ni}(\varepsilon_i)) > \sup_{\omega \in \Omega \setminus \mathcal{B}_\delta(\omega^*_{ni})} \Pi^*_{ni}(\omega)$,

$\omega_{ni}(\varepsilon_i) \notin \Omega \backslash \mathcal{B}_\delta(\omega_{ni}^*)$, or equivalently, $\omega_{ni}(\varepsilon_i) \in \mathcal{B}_\delta(\omega_{ni}^*)$. Therefore,

$$\begin{aligned}
&\Pr\left(\|\omega_{ni}(\varepsilon_i) - \omega_{ni}^*\| < \delta \,|\, X, p_n\right) \\
&\geq \Pr\left(\Pi_{ni}^*(\omega_{ni}(\varepsilon_i)) > \sup_{\omega\in\Omega\backslash\mathcal{B}_\delta(\omega_{ni}^*)} \Pi_{ni}^*(\omega) \,\middle|\, X, p_n\right) \\
&= \Pr\left(\Pi_{ni}^*(\omega_{ni}^*) - \Pi_{ni}^*(\omega_{ni}(\varepsilon_i)) < \Pi_{ni}^*(\omega_{ni}^*) - \sup_{\omega\in\Omega\backslash\mathcal{B}_\delta(\omega_{ni}^*)} \Pi_{ni}^*(\omega) \,\middle|\, X, p_n\right). \qquad (8.28)
\end{aligned}$$

By Assumption 5(i)-(ii)

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

so the right-hand size of (8.28) goes to 1 if

$$\frac{1}{n-1}\left(\Pi_{ni}^*(\omega_{ni}^*) - \Pi_{ni}^*(\omega_{ni}(\varepsilon_i))\right) \leq o_p(1). \qquad (8.29)$$

By the optimality of $\omega_{ni}(\varepsilon_i)$ we have

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

By the uniform LLN for $\Pi_{ni}(\omega, \varepsilon_i)$ in Lemma 8.6,

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

so (8.29) holds and the consistency is proved. ■

Lemma 8.6 (Uniform LLN for Π_{ni}) *Suppose that Assumptions 1-3 and 5 are satisfied. Then for any $\delta > 0$,*

$$\Pr\left(\sup_{\omega\in\Omega} \frac{1}{n-1}\left|\left(\Pi_{ni}(\omega, \varepsilon_i) - \Pi_{ni}^*(\omega)\right)\right| > \delta \,\middle|\, X, p_n\right) \to 0 \qquad (8.30)$$

as $n \to \infty$.

Proof. Recall that

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

and

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

Define

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

Hence,

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

By Assumption 5(i)

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

Therefore for all $\omega\in\Omega$

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

with

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

Also $\pi_{n,ij}\left(\omega,\varepsilon_{ij}\right)$ is continuous in ω on a compact set Ω. Therefore the conditions of the uniform LLN for triangular arrays are satisfied (Jennrich (1969)) and (8.30) follows. ■

Lemma 8.7 (Asymptotically linear representation of $\omega_{ni}\left(\varepsilon_i\right)$) *Suppose that Assumptions 1-3 and 5 are satisfied. For each $i=1,\ldots,n$, $\omega_{ni}\left(\varepsilon_i\right)$ has an asymptotically linear representation*

$$\omega_{ni}\left(\varepsilon_i\right)-\omega_{ni}^*=\frac{1}{n-1}\sum_{j\neq i}\varphi_{n,ij}^{\omega}\left(\omega_{ni}^*,\varepsilon_{ij}\right)+r_{ni}^{\omega}\left(\varepsilon_i\right)\tag{8.31}$$

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

$$\varphi_{n,ij}^{\omega}\left(\omega_{ni}^*,\varepsilon_{ij}\right)=-\nabla_{\omega'}\Gamma_{ni}^*\left(\omega_{ni}^*\right)^+\varphi_{n,ij}^{\pi}\left(\omega_{ni}^*,\varepsilon_{ij}\right),\tag{8.32}$$

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

$$\varphi^{\pi}_{n,ij}(\omega, \varepsilon_{ij}) = 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega - \varepsilon_{ij} \geq 0\right\} \Lambda_{ni} \Phi_{ni}' Z_j - \Lambda_{ni} \omega, \tag{8.33}$$

and

$$\nabla_{\omega'} \Gamma^*_{ni}(\omega^*_{ni}) = \frac{2}{n-2} \sum_{j \neq i} f_{\varepsilon}\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega^*_{ni}\right) \Lambda_{ni} \Phi_{ni}' Z_j Z_j' \Phi_{ni} \Lambda_{ni} - \Lambda_{ni}. \tag{8.34}$$

with by Assumption 5(iii) the generalized inverse

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

Moreover, the remainder $r^{\omega}_{ni}(\varepsilon_i)$ in (8.31) satisfies

$$r^{\omega}_{ni}(\varepsilon_i) = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{8.35}$$

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

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

where $\varphi^{\pi}_{n,ij}(\omega, \varepsilon_{ij}) \in \mathbb{R}^T$ is defined in (8.33). By Lemma 8.1 $\omega_{ni}(\varepsilon_i)$ satisfies the first-order condition

$$\Gamma_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) = 0, \text{ a.s.} \tag{8.36}$$

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

$$\Gamma^*_{ni}\left(\omega^*_{ni}\right) = 0. \tag{8.37}$$

By a Taylor expansion of $\Gamma^*_{ni}(\omega)$ at ω^*_{ni} and the consistency of $\omega_{ni}(\varepsilon_i)$ in Lemma 8.5, we have

$$\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right) = \nabla_{\omega'}\Gamma^*_{ni}\left(\omega^*_{ni}\right)\left(\omega_{ni}\left(\varepsilon_i\right) - \omega^*_{ni}\right) + O_p\left(\left\|\omega_{ni}\left(\varepsilon_i\right) - \omega^*_{ni}\right\|^2\right), \tag{8.38}$$

where $\nabla_{\omega'}\Gamma^*_{ni}\left(\omega^*_{ni}\right)$ is the Jacobian matrix of $\Gamma^*_{ni}(\omega)$ at ω^*_{ni} defined in (8.34) that we rewrite as

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

with

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

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

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

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

$$\left\|\Gamma^*_{ni}\left(\omega_{ni}\left(\varepsilon_i\right)\right)\right\| \geq \left\|\omega_{ni}\left(\varepsilon_i\right) - \omega^*_{ni}\right\|\left(c + o_p(1)\right). \tag{8.39}$$

By (8.36) and (8.37), we can write $\Gamma^*_{ni}(\omega_{ni}(\varepsilon_i))$ as

$$\begin{aligned}&\Gamma^*_{ni}(\omega_{ni}(\varepsilon_i))\\&= -\Gamma_{ni}(\omega^*_{ni},\varepsilon_i) - (\Gamma_{ni}(\omega_{ni}(\varepsilon_i),\varepsilon_i) - \Gamma^*_{ni}(\omega_{ni}(\varepsilon_i)) - (\Gamma_{ni}(\omega^*_{ni},\varepsilon_i) - \Gamma^*_{ni}(\omega^*_{ni}))), \text{ a.s.}\end{aligned} \tag{8.40}$$

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

$$\Gamma_{ni}(\omega^*_{ni},\varepsilon_i) = O_p\left(\frac{1}{\sqrt{n}}\right). \tag{8.41}$$

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

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

so that

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

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 $\xi > 0$

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

with

$$\begin{aligned}\Sigma^\gamma_{ni} &= \sum_{j\neq i} Var\left(Y^\gamma_{n,ij}\middle| X, p_n\right)\\ &= \frac{1}{n-1}\sum_{j\neq i} F_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z'_j \Phi_{ni} \Lambda_{ni} \omega^*_{ni}\right)\\ &\qquad \cdot\left(1 - F_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z'_j \Phi_{ni} \Lambda_{ni} \omega^*_{ni}\right)\right) \Lambda_{ni} \Phi'_{ni} Z_j Z'_j \Phi_{ni} \Lambda_{ni}.\end{aligned}$$

We have

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

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

$$\frac{\max_{j\neq i}\left|a'Y^{\gamma}_{n,ij}\right|}{\sqrt{a'\Sigma^{\gamma}_{ni}a}}=o_p(1), \tag{8.42}$$

then

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

Finally, this random variable is bounded by $\sum_{j\neq i}\left(a'Y^{\gamma}_{n,ij}\right)^2$ that has a finite expectation. We conclude that by dominated convergence the Lindeberg condition is satisfied if (8.42) holds.

By Chebyshev's inequality,

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

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

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

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 $\psi(z)=e^z-1$.[13] Then $\mathbb{E}\left[|Z|\,|\,X,p_n\right]\leq\|Z\|_{\psi|X,p_n}$[14] so that

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

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

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

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

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

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

[14] 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 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

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

so the Lindeberg condition holds.

As for the second term on the right-hand side of (8.40), note that

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

with $\varphi_{n,ij}^{\gamma}(\omega,\varepsilon_{ij})$ defined by

$$\varphi_{n,ij}^{\gamma}(\omega,\varepsilon_{ij}) = 1\left\{U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega - \varepsilon_{ij} \geq 0\right\}\Lambda_{ni}\Phi_{ni}'Z_j. \tag{8.43}$$

Define the empirical process

$$\begin{aligned}\mathbb{G}_n\varphi_{ni}^{\gamma}(\omega,\varepsilon_i) &= \sqrt{n-1}\left(\Gamma_{ni}(\omega,\varepsilon_i) - \Gamma_{ni}^*(\omega)\right)\\ &= \frac{1}{\sqrt{n-1}}\sum_{j\neq i}\left(\varphi_{n,ij}^{\gamma}(\omega,\varepsilon_{ij}) - \mathbb{E}\left[\varphi_{n,ij}^{\gamma}(\omega,\varepsilon_{ij})\middle| X, p_n\right]\right),\ \omega\in\Omega\end{aligned} \tag{8.44}$$

so the second term on the right-hand side of (8.40) can be written as

$$\begin{aligned}&\Gamma_{ni}(\omega_{ni}(\varepsilon_i),\varepsilon_i) - \Gamma_{ni}^*(\omega_{ni}(\varepsilon_i)) - \left(\Gamma_{ni}(\omega_{ni}^*,\varepsilon_i) - \Gamma_{ni}^*(\omega_{ni}^*)\right)\\ =&\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\varphi_{ni}^{\gamma}(\omega_{ni}(\varepsilon_i),\varepsilon_i) - \mathbb{G}_n\varphi_{ni}^{\gamma}(\omega_{ni}^*,\varepsilon_i)\right).\end{aligned} \tag{8.45}$$

In Lemma 8.9(i) we show that

$$\mathbb{G}_n\varphi_{ni}^{\gamma}(\omega_{ni}(\varepsilon_i),\varepsilon_i) - \mathbb{G}_n\varphi_{ni}^{\gamma}(\omega_{ni}^*,\varepsilon_i) = o_p(1). \tag{8.46}$$

Hence, the second term on the right-hand side of (8.40) is $o_p\left(n^{-1/2}\right)$

$$\Gamma_{ni}(\omega_{ni}(\varepsilon_i),\varepsilon_i) - \Gamma_{ni}^*(\omega_{ni}(\varepsilon_i)) - \left(\Gamma_{ni}(\omega_{ni}^*,\varepsilon_i) - \Gamma_{ni}^*(\omega_{ni}^*)\right) = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{8.47}$$

Applying (8.39), (8.41) and (8.47) to (8.40) we obtain

$$\|\omega_{ni}(\varepsilon_i) - \omega_{ni}^*\| (c + o_p(1)) \leq O_p\left(\frac{1}{\sqrt{n}}\right) + o_p\left(\frac{1}{\sqrt{n}}\right). \tag{8.48}$$

This implies that

$$\omega_{ni}(\varepsilon_i) - \omega_{ni}^* = O_p\left(\frac{1}{\sqrt{n}}\right), \tag{8.49}$$

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

Combining (8.38), (8.40), and (8.47) 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

$$\begin{aligned}\omega_{ni}(\varepsilon_i) - \omega_{ni}^* &= -\nabla_{\omega'}\Gamma_{ni}^*(\omega_{ni}^*)^+ \Gamma_{ni}(\omega_{ni}^*, \varepsilon_i) + r_{ni}^{\omega}(\varepsilon_i) \\ &= -\frac{1}{n-1}\sum_{j\neq i}\nabla_{\omega'}\Gamma_{ni}^*(\omega_{ni}^*)^+ \varphi_{n,ij}^{\pi}(\omega_{ni}^*, \varepsilon_{ij}) + r_{ni}^{\omega}(\varepsilon_i)\end{aligned} \tag{8.50}$$

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

Lemma 8.8 (Uniform Properties of $\omega_{ni}(\varepsilon_i)$) *Suppose that Assumptions 1-3 and 5 are satisfied. Then (i) $\omega_{ni}(\varepsilon_i)$ satisfies*

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

(ii) The remainder $r_{ni}^{\omega}(\varepsilon_i)$ defined in Lemma 8.7 satisfies

$$\max_{1\leq i\leq n}\|r_{ni}^{\omega}(\varepsilon_i)\| = o_p\left(\frac{1}{\sqrt{n}}\right).$$

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

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

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 $\mathbb{E}[|Z||X, p_n] \leq \|Z\|_{\psi|X,p_n}$ for any random variable Z and the maximal inequality in Lemma 2.2.2 in Van der Vaart and Wellner (1996) we derive

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

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 $\|Z^2\|_{\psi|X,p_n}=\|Z\|_{\psi_2|X,p_n}^2$. Hence,

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

From (8.38), (8.40), (8.45), and (8.50) in Lemma 8.7, the remainder $r_{ni}^{\omega}(\varepsilon_i)$ is

$$r_{ni}^{\omega}(\varepsilon_i)=\nabla_{\omega'}\Gamma_{ni}^*(\omega_{ni}^*)^+\left(O_p\left(\|\omega_{ni}(\varepsilon_i)-\omega_{ni}^*\|^2\right)+\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\varphi_{ni}^{\gamma}(\omega_{ni}(\varepsilon_i),\varepsilon_i)-\mathbb{G}_n\varphi_{ni}^{\gamma}(\omega_{ni}^*,\varepsilon_i)\right)\right). \tag{8.51}$$

Therefore, from (8.50) and (8.51), we obtain

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

so by the triangle inequality (that holds for the Orlicz norm)[15] and the boundedness of the inverse Jacobian $\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+$

$$\begin{aligned}
&\left\|(\omega_{ni}(\varepsilon_i)-\omega^*_{ni})(1+o_p(1))\right\|_{\psi_2|X,p_n} \\
\leq &\left\|\frac{1}{n-1}\sum_{j\neq i}\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij})\right\|_{\psi_2|X,p_n} \\
&+\frac{1}{\sqrt{n-1}}\left\|\nabla_{\omega'}\Gamma^*_{ni}(\omega^*_{ni})^+\right\|\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}(\omega_{ni}(\varepsilon_i),\varepsilon_i)-\mathbb{G}_n\varphi^{\gamma}_{ni}(\omega^*_{ni},\varepsilon_i)\right\|_{\psi_2|X,p_n}. \qquad (8.52)
\end{aligned}$$

Notice that $\|(\omega_{ni}(\varepsilon_i)-\omega^*_{ni})(1+o_p(1))\|_{\psi_2|X,p_n}=\|(\omega_{ni}(\varepsilon_i)-\omega^*_{ni})\|_{\psi_2|X,p_n}(1+o(1))$.[16]

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

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

[15] 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 $\|X+Y\|_{\psi_2|X,p_n}\leq\|X\|_{\psi_2|X,p_n}+\|Y\|_{\psi_2|X,p_n}$.

[16] 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 $\varphi^{\omega}_{n,ij,t}\left(\omega^*_{ni},\varepsilon_{ij}\right)$ denote the t-th component of $\varphi^{\omega}_{n,ij}\left(\omega^*_{ni},\varepsilon_{ij}\right)$, $t=1,\ldots T$. Note that

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

so for any $x>0$,

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

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

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

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

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

so

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

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

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

From (8.65) in the proof of Lemma 8.9 we see that

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

Following the proof for (8.65) and applying Theorem 2.14.5 and Theorem 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 (8.52) is $\frac{o(1)}{\sqrt{n-1}}$.

Combining the results yields

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

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

Part (ii): From (8.38), (8.40), (8.45), and (8.50) in Lemma 8.7, the remainder $r_{ni}^{\omega} \left(\varepsilon_i \right)$ is given by

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

It is clear that $\max_{1 \leq i \leq n} \left\| \nabla_{\omega'} \Gamma_{ni}^{*} \left(\omega_{ni}^{*} \right)^{+} \right\| \leq M < \infty$. By Lemma 8.9(ii)

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

so combining this with part (i) we obtain

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

The proof is complete. ■

Lemma 8.9 (Stochastic equicontinuity) *Suppose that Assumptions 1-3 and 5 are satisfied. The* $\mathbb{G}_n \varphi_{ni}^{\gamma} \left(\omega, \varepsilon_i \right)$ *defined in (8.44) satisfies that for any* $\delta > 0$*,* *(i) for* $\omega_{ni}(\varepsilon_i) - \omega^*_{ni} = o_p(1)$

$$\Pr\left(\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - \mathbb{G}_n\varphi^{\gamma}_{ni}(\omega^*_{ni}, \varepsilon_i)\right\| > \delta \middle| X, p_n\right) \to 0 \tag{8.53}$$

as $n \to \infty$*, and*

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

$$\Pr\left(\max_{1\leq i\leq n}\left\|\mathbb{G}_n\varphi^{\gamma}_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - \mathbb{G}_n\varphi^{\gamma}_{ni}(\omega^*_{ni}, \varepsilon_i)\right\| > \delta \middle| X, p_n\right) \to 0 \tag{8.54}$$

as $n \to \infty$*.*

Proof. Part (i): By consistency of $\omega_{ni}(\varepsilon_i)$, we can define h_{ni} by $\omega_{ni}(\varepsilon_i) - \omega^*_{ni} = r_n^{-1}h_{ni}$ for some $r_{ni} \to \infty$, and $h_{ni} \in \Omega$ if n is sufficiently large, because by Assumption 5 Ω contains a compact neighborhood of 0.

By Markov's inequality

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

We consider the empirical process

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

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

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

where $\varphi^{\gamma}_{n,ij}(\omega, \varepsilon_{ij})$ is defined in (8.43) by

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

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

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

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

$$
\begin{aligned}
&\eta_{n,ij}(\omega, \tilde{\omega}, \varepsilon_{ij}) \\
= &\begin{cases} \left\|\Lambda_{ni}\Phi'_{ni}Z_j\right\|, & \text{if } \varepsilon_{ij} \text{ lies between } U_{n,ij} + \frac{2(n-1)}{n-2} Z'_j \Phi_{ni}\Lambda_{ni}\omega \text{ and } U_{n,ij} + \frac{2(n-1)}{n-2} Z'_j \Phi_{ni}\Lambda_{ni}\tilde{\omega}, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
\tag{8.55}
$$

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

$$
\mathbb{G}_n\left(\varphi^{\gamma}_{ni}\left(\omega + \frac{h}{r_{ni}}, \varepsilon_i\right) - \varphi^{\gamma}_{ni}(\omega, \varepsilon_i)\right) \tag{8.56}
$$

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[17]

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\left(\varphi^{\gamma}_{ni}\left(\omega + \frac{h}{r_{ni}}, \varepsilon_i\right) - \varphi^{\gamma}_{ni}(\omega, \varepsilon_i)\right)\right\| \middle| X, p_n\right] \\
&\leq K\mathbb{E}\left[J(1, \mathcal{F}_{ni}(\varepsilon_i)) \sup_{\omega,h\in\Omega}\left\|\eta_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)\right\|_n \middle| X, p_n\right],
\end{aligned}
\tag{8.57}
$$

where $K > 0$ is a constant and

$$
\left\|\eta_{ni}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_i\right)\right\|_n^2 = \frac{1}{n-1}\sum_{j\neq i}\eta^2_{n,ij}\left(\omega + \frac{h}{r_{ni}}, \omega, \varepsilon_{ij}\right). \tag{8.58}
$$

[17] To apply Theorem 2.14.1 in Van der Vaart and Wellner, we need to reprove the symmetrization lemma in Lemma 2.3.1 and reapply the Hoeffding's inequality in Lemma 2.2.7 and the maximal inequality in Corollary 2.2.5 (all in Van der Vaart and Wellner) for the empirical process in (8.56) and the bound function in (8.55).

We now show that the uniform entropy integral $J(1,\mathcal{F}_{ni}(\varepsilon_i))$ in (8.57) is finite, where $\mathcal{F}_{ni}(\varepsilon_i)$ denotes the set of arrays

$$\mathcal{F}_{ni}(\varepsilon_i)=\left\{\left(\varphi^{\gamma}_{n,ij}\left(\omega+\frac{h}{r_{ni}},\varepsilon_{ij}\right)-\varphi^{\gamma}_{n,ij}(\omega,\varepsilon_{ij}),j\neq i\right):\omega,h\in\Omega\right\}, \tag{8.59}$$

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

$$J(1,\mathcal{F}_{ni}(\varepsilon_i))=\int_0^1\sup_{\alpha\in\mathbb{R}^{n-1}_+}\sqrt{\ln D\left(\xi\left\|\alpha\odot\bar{\eta}_{ni}(\varepsilon_i)\right\|_n,\alpha\odot\mathcal{F}_{ni}(\varepsilon_i),\|\cdot\|_n\right)}d\xi. \tag{8.60}$$

In the expression in (8.60), $\alpha\in\mathbb{R}^{n-1}_+$ is a $(n-1)\times 1$ vector of nonnegative constants, $\bar{\eta}_{ni}(\varepsilon_i)=\sup_{\omega,h\in\Omega}\eta_{ni}\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_i\right)$, $\alpha\odot\bar{\eta}_{ni}(\varepsilon_i)$ is the pointwise product of α and $\bar{\eta}_{ni}(\varepsilon_i)$, $\alpha\odot\mathcal{F}_{ni}(\varepsilon_i)$ is the set $\{\alpha\odot f:f\in\mathcal{F}_{ni}(\varepsilon_i)\}$, $\|\cdot\|_n$ is the $L_2(\mathbb{P}_n)$ norm defined in (8.58), 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 of the set $\alpha\odot\mathcal{F}_{ni}(\varepsilon_i)$ at distance $\xi\left\|\alpha\odot\bar{\eta}_{ni}(\varepsilon_i)\right\|_n$ under the norm $\|\cdot\|_n$. The sup in (8.60) is taken over all vectors $\alpha\in\mathbb{R}^{n-1}_+$ with nonnegative constants.

Consider the function

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

It is an indicator with the argument being a linear function of ω. We can show that the set $\left\{\left(\bar{\varphi}^{\gamma}_{n,ij}(\omega,\varepsilon_{ij}),j\neq i\right):\omega\in\Omega\right\}$ has a pseudo-dimension of at most T,[18] so it is manageable (Corollary 4.10 in Pollard (1990)). Note that $\varphi^{\gamma}_{n,ij}(\omega,\varepsilon_{ij})=\bar{\varphi}^{\gamma}_{n,ij}(\omega,\varepsilon_{ij})\Lambda_{ni}\Phi_{ni}'Z_j$, and $\Lambda_{ni}\Phi_{ni}'Z_j$ is a $T\times 1$ vector that does not depend on ω. From the stability results in Section in Pollard (1990), 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 manageable. Therefore, the set $\mathcal{F}_{ni}(\varepsilon_i)$ is manageable and

[18] 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

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

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 $\omega\in\Omega$. If $\sum_{t=1}^{T+1}\tau_t(U_{n,ij_t}-\varepsilon_{n,ij_t}-c_{j_t})\leq 0$, it is impossible to find a $\omega\in\Omega$ satisfying these inequalities for the choice $J=\{j_t\in I:\tau_t>0\}$, because this would lead to the contradiction $\sum_{t=1}^{T+1}\tau_t(U_{n,ij_t}-\varepsilon_{n,ij_t}-c_{j_t})=\sum_{t=1}^{T+1}\tau_t(U_{n,ij_t}-\varepsilon_{n,ij_t}-c_{j_t})+\sum_{t=1}^{T+1}\tau_t\frac{2(n-1)}{n-2}Z_{j_t}'\Phi_{ni}\Lambda_{ni}\omega=\sum_{t=1}^{T+1}\tau_t\left(U_{n,ij_t}+\frac{2(n-1)}{n-2}Z_{j_t}'\Phi_{ni}\Lambda_{ni}\omega-\varepsilon_{n,ij_t}-c_{j_t}\right)>0$. If $\sum_{t=1}^{T+1}\tau_t(U_{n,ij_t}-\varepsilon_{n,ij_t}-c_{j_t})>0$, we would choose $J=\{j_t\in I:\tau_t\leq 0\}$ to reach a similar contradiction.

has a finite uniform entropy integral, i.e.,

$$J\left(1, \mathcal{F}_{ni}\left(\varepsilon_i\right)\right) \leq \bar{J} \tag{8.61}$$

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

By Jensen's inequality

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\omega,h\in\Omega}\left\|\eta_{ni}\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_i\right)\right\|_n \middle| X, p_n\right] \\
=&\mathbb{E}\left[\left(\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\right)^{1/2}\middle| X, p_n\right] \\
\leq&\left(\mathbb{E}\left[\sup_{\omega,h\in\Omega}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\middle| X, p_n\right]\right)^{1/2}.
\end{aligned} \tag{8.62}$$

To derive an upper bound on the last term in (8.62), we consider the empirical process

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

indexed by $\omega, h \in \Omega$. Note that each $\eta_{n,ij}^2$ is bounded by $\left\|\Lambda_{ni}\Phi_{ni}' Z_j\right\|^2 \leq \max_{t=1,\cdots,T}\lambda_{ni,t}^2 T \leq \bar{\eta}^2 < \infty$. Similarly to (8.57), 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

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

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 (8.59), we can show that the set $\mathcal{F}_{ni}^\eta\left(\varepsilon_i\right)$ is manageable so it has a finite uniform entropy integral

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

Therefore,

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

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

$$
\begin{aligned}
&\mathbb{E}\left[\eta_{n,ij}^{2}\left(\omega+\frac{h}{r_{ni}},\omega,\varepsilon_{ij}\right)\middle| X,p_n\right]\\
=&\left|F_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\left(\omega+\frac{h}{r_{ni}}\right)\right)-F_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\omega\right)\right|\left\|\Lambda_{ni}\Phi_{ni}'Z_j\right\|^{2}\\
=&f_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\left(\omega+t_{n,ij}\frac{h}{r_{ni}}\right)\right)\frac{2\left(n-1\right)}{n-2}\left|Z_j'\Phi_{ni}\Lambda_{ni}\frac{h}{r_{ni}}\right|\left\|\Lambda_{ni}\Phi_{ni}'Z_j\right\|^{2}\\
\leq&\frac{1}{r_{ni}}f_{\varepsilon}\left(U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_{ni}\Lambda_{ni}\left(\omega+t_{n,ij}\frac{h}{r_{ni}}\right)\right)\frac{2\left(n-1\right)}{n-2}\left\|\Lambda_{ni}\Phi_{ni}'Z_j\right\|^{3}\sup_{h\in\Omega}\|h\| \qquad (8.63)
\end{aligned}
$$

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} that does not depend on i and j. We conclude that there is a finite M with

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

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

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

Because for any random variable $\mathbb{E}\left[|Z|\,|\,X,p_n\right]\le\|Z\|_{\psi|X,p_n}$, by the maximal inequality in Lemma 2.2.2 in Van der Vaart and Wellner (1996)

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

where $K>0$ is a constant.

We follow the proof of Theorem 2.14.5 in Van der Vaart and Wellner (1996) and derive an upper bound on the Orlicz norm of the sup norm of the empirical process

$$\mathbb{G}_n\left(\varphi_{ni}^{\gamma}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_i\right)-\varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)\right)\tag{8.64}$$

indexed by $\omega,h\in\Omega$. By Theorem 2.14.5 (with $p=1$) and Lemma 2.2.2 in Van der Vaart and Wellner (1996), the upper bound is

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

In Part (i) we have derived that

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

Moreover, by the definition of $\eta_{n,ij}$,

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

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

$$\left\|\sup_{\omega,h\in\Omega}\left\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega+\frac{h}{n^{\kappa}},\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega,\varepsilon_i\right)\right\|\right\|_{\psi|X,p_n}\leq K_1\left(K\bar{J}\sqrt{\frac{M^{\eta}}{\sqrt{n-1}}+\frac{M}{n^{\kappa}}}+\frac{\bar{\eta}\ln n}{\sqrt{n-1}}\right). \tag{8.65}$$

Combining the results yields the upper bound

$$\begin{aligned}&\Pr\left(\max_{1\leq i\leq n}\|\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}\left(\varepsilon_i\right),\varepsilon_i\right)-\mathbb{G}_n\varphi_{ni}^{\gamma}\left(\omega_{ni}^{*},\varepsilon_i\right)\|>\delta\middle| X,p_n\right)\\&\leq\frac{KK_1\ln\left(n+1\right)}{\delta}\left(K\bar{J}\sqrt{\frac{M^{\eta}}{\sqrt{n-1}}+\frac{M}{n^{\kappa}}}+\frac{\bar{\eta}\ln n}{\sqrt{n-1}}\right),\end{aligned}$$

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

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

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

$$
\begin{aligned}
&\frac{1}{n\,(n-1)}\sum_{i}\sum_{j\neq i}W_{n,ij}\nabla_{\theta'}P_{n,ij}\left(\theta_0,p_n\right)\left(\hat{\theta}_n-\theta_0\right)\\
=&\frac{1}{n\,(n-1)}\sum_{i}\sum_{j\neq i}W_{n,ij}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right)\\
&-\frac{1}{n\,(n-1)}\sum_{i}\sum_{j\neq i}W_{n,ij}\nabla_{p'}P_{n,ij}\left(\theta_0,p_n\right)\left(\hat{p}_n-p_n\right)\\
&+o_p\left(\left\|\hat{\theta}_n-\theta_0\right\|\right)+o_p\left(\frac{1}{n}\right).
\end{aligned}
\tag{8.66}
$$

where we have used

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

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

Let us examine the first two terms on the right-hand side of (8.66), with the first being the main term while the second gives the contribution of the 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

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

Hence, the first two terms on the right-hand side of (8.66) 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 weight/instrument

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

By Lemma 8.10,

$$\begin{aligned} & \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \tilde{W}_{n,ij} \left(G_{n,ij} - P_{n,ij}(\theta_0, p_n) \right) \\ = & \frac{1}{\sqrt{n(n-1)}} \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} \tilde{W}_{n,ij} \left(G_{n,ij} - P_{n,ij}(\theta_0, p_n) \right) \\ = & O_p\left(\frac{1}{n}\right), \end{aligned} \tag{8.67}$$

so (8.66) 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 (8.66) as

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

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 8.10. 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 8.10 for the weight function $J_n^\theta(\theta_0, p_n)^{-1} \tilde{W}_{n,ij}$. 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

$$\begin{aligned}
&\Sigma_{n,ij}(\theta_0, p_n) \\
=&J_n^\theta(\theta_0, p_n)^{-1} \mathbb{E}\left[\left(\tilde{W}_{n,ij}\left(g_{n,ij}(\omega_{ni}^*, \varepsilon_{ij}) - P_{n,ij}^*(\omega_{ni}^*) \right) + \tilde{J}_{ni}^\omega(\omega_{ni}^*) \varphi_{n,ij}^\omega(\omega_{ni}^*, \varepsilon_{ij}) \right) \right. \\
&\left. \cdot \left(\tilde{W}_{n,ij}\left(g_{n,ij}(\omega_{ni}^*, \varepsilon_{ij}) - P_{n,ij}^*(\omega_{ni}^*) \right) + \tilde{J}_{ni}^\omega(\omega_{ni}^*) \varphi_{n,ij}^\omega(\omega_{ni}^*, \varepsilon_{ij}) \right)' \right| X, p_n \right] \left(J_n^\theta(\theta_0, p_n)^{-1} \right)'
\end{aligned} \tag{8.68}$$

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 8.10. 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 (8.32) in Lemma 8.7. By Lemma 8.10,

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

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

Lemma 8.10 (Asymptotic normality of the sample moment function) *Suppose that Assumption 1-3 and 5 are satisfied. Define*

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

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

$$\Sigma_n = \frac{1}{n(n-1)}\sum_{i}\sum_{j\neq i}\Sigma_{n,ij} \tag{8.69}$$

with

$$\begin{aligned}\Sigma_{n,ij} =& \mathbb{E}\Big[\left(W_{n,ij}\left(g_{n,ij}(\omega_{ni}^{*},\varepsilon_{ij}) - P_{n,ij}^{*}(\omega_{ni}^{*})\right) + J_{ni}^{\omega}(\omega_{ni}^{*})\varphi_{n,ij}^{\omega}(\omega_{ni}^{*},\varepsilon_{ij})\right) \\ &\cdot \left(W_{n,ij}\left(g_{n,ij}(\omega_{ni}^{*},\varepsilon_{ij}) - P_{n,ij}^{*}(\omega_{ni}^{*})\right) + J_{ni}^{\omega}(\omega_{ni}^{*})\varphi_{n,ij}^{\omega}(\omega_{ni}^{*},\varepsilon_{ij})\right)'\Big| X, p_n\Big]\end{aligned} \tag{8.70}$$

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

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

the indicator functions $g_{n,ij}(\omega,\varepsilon_{ij})$, the corresponding probabilities $P_{n,ij}^{*}(\omega)$, and the $d_\theta \times T$

matrix $J^{\omega}_{ni}(\omega)$ *are defined by*

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

and $\varphi^{\omega}_{n,ij}(\omega^*_{ni},\varepsilon_{ij}) \in \mathbb{R}^T$ *is the* j*-th term of the influence function of* $\omega_{ni}(\varepsilon_i)$ *defined in (8.32) in Lemma 8.7. In the expressions above* $U_{n,ij}$*,* Φ_{ni} *and* V_{ni} *are all evaluated at* (θ_0, p_n)*. Under these assumptions*

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

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

Proof. Define the link choice indicator at the population parameters θ_0, p_n (suppressed in the notation) and at $\omega \in \Omega$

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

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

$$G_{n,ij} = g_{n,ij}(\omega_{ni}(\varepsilon_i),\varepsilon_{ij}),\ j\neq i, \tag{8.71}$$

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

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

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

$$P_{n,ij} = \mathbb{E}\left[g_{n,ij}(\omega_{ni}(\varepsilon_i),\varepsilon_{ij})\,|\,X,p_n\right],\ j\neq i. \tag{8.72}$$

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 $\omega_{ni}(\varepsilon_i)$. As shown in Lemma 8.5 $\omega_{ni}(\varepsilon_i)$ converges in probability to ω^*_{ni} that does not depend on ε_i. Let $\Pi^*_{ni}(\omega)$ be the expectation of $\Pi_{ni}(\omega)$

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

By Assumption 5 $\Pi^*_{ni}(\omega)$ has a unique maximizer ω^*_{ni} that does not depend on ε_i.

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

$$\begin{aligned} P^*_{n,ij}(\omega) &= \mathbb{E}\left[g_{n,ij}(\omega, \varepsilon_{ij}) \middle| X, p_n\right] \\ &= F_\varepsilon\left(U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega\right), \ j \neq i. \end{aligned} \tag{8.73}$$

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

$$\begin{aligned} Y_n =& \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j\neq i} W_{n,ij} \left(g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) - \mathbb{E}\left[g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) \middle| X, p_n\right]\right) \\ =& T_{1n} + T_{2n} + T_{3n} + T_{4n}, \end{aligned} \tag{8.74}$$

where

$$\begin{aligned} T_{1n} &= \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j\neq i} W_{n,ij} \left(g_{n,ij}(\omega^*_{ni}, \varepsilon_{ij}) - P^*_{n,ij}(\omega^*_{ni})\right) \\ T_{2n} &= \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j\neq i} W_{n,ij} \left(g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) - g_{n,ij}(\omega^*_{ni}, \varepsilon_{ij}) - \left(P^*_{n,ij}(\omega_{ni}(\varepsilon_i)) - P^*_{n,ij}(\omega^*_{ni})\right)\right) \\ T_{3n} &= \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j\neq i} W_{n,ij} \left(P^*_{n,ij}(\omega_{ni}(\varepsilon_i)) - P^*_{n,ij}(\omega^*_{ni})\right) \\ T_{4n} &= -\frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j\neq i} W_{n,ij} \left(\mathbb{E}\left[g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) \middle| X, p_n\right] - P^*_{n,ij}(\omega^*_{ni})\right). \end{aligned} \tag{8.75}$$

The terms in the decomposition have an interpretation. T_{1n} captures the sampling variation of the link choices after removing the correlation between the link choices of an individual. The term T_{2n} contains the approximation error in the main term due to evaluaton at ω^*_{ni} which removes the correlation between the link choices. The fact that this term is negligible shows that the correlation between the link choices vanishes if n is large. Instead of a smooth approximation of a non-smooth moment condition, this approximation is to a moment condition with independent link choices. The sampling variation in $\omega_{ni}(\varepsilon_i)$ is captured in T_{3n}. Finally, the linear approximation of T_{3n} has non-negligible approximation errors that cancel out if we add T_{4n}.

Let us now examine the four terms in (8.75).

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 error in the link choices.

Step 2: T_{2n}.

We show that T_{2n} in (8.75) 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

$$\begin{aligned} T_{2n} &= \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j \neq i} W_{n,ij} \left(g_{n,ij}(\omega_{ni}(\varepsilon_i), \varepsilon_{ij}) - g_{n,ij}(\omega_{ni}^*, \varepsilon_{ij}) - \left(P_{n,ij}^*(\omega_{ni}(\varepsilon_i)) - P_{n,ij}^*(\omega_{ni}^*) \right) \right) \\ &= \frac{1}{\sqrt{n}} \sum_i \mathbb{G}_n W_{ni} \left(g_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - g_{ni}(\omega_{ni}^*, \varepsilon_i) \right). \end{aligned}$$

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 8.7 $\omega_{ni}(\varepsilon_i) - \omega_{ni}^* = O_p\left(n^{-1/2}\right)$, 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 compacct neighborhood of 0. Note that T_{2n} is a normalized average of terms that are $o_p(1)$ by establishing stochastic equicontinuity. Hence we cannot directly invoke a stochastic equicontinuity argument to show that T_{2n} is $o_p(1)$.

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

$$\begin{aligned} &\Pr\left(\left\| \frac{1}{\sqrt{n}} \sum_i \mathbb{G}_n W_{ni} \left(g_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - g_{ni}(\omega_{ni}^*, \varepsilon_i) \right) \right\| > \delta \,\middle|\, X, p_n \right) \\ \leq & \frac{1}{\delta^2 n} \sum_i \mathbb{E}\left[\left\| \mathbb{G}_n W_{ni} \left(g_{ni}(\omega_{ni}(\varepsilon_i), \varepsilon_i) - g_{ni}(\omega_{ni}^*, \varepsilon_i) \right) \right\|^2 \middle| X, p_n \right] \\ \leq & \frac{1}{\delta^2 n} \sum_i \mathbb{E}\left[\sup_{\omega, h \in \Omega} \left\| \mathbb{G}_n W_{ni} \left(g_{ni}\left(\omega + \frac{h}{n^\kappa}, \varepsilon_i \right) - g_{ni}(\omega, \varepsilon_i) \right) \right\|^2 \middle| X, p_n \right] \end{aligned} \tag{8.76}$$

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

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

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

$$\begin{aligned}
&\eta_{n,ij}(\omega, \tilde{\omega}, \varepsilon_{ij}) \\
&= \begin{cases} \|W_{n,ij}\|, & \text{if } \varepsilon_{ij} \text{ lies between } U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \omega \text{ and } U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' \Phi_{ni} \Lambda_{ni} \tilde{\omega}, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{8.77}$$

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

$$\mathbb{G}_n W_{ni} \left(g_{ni} \left(\omega + \frac{h}{n^\kappa}, \varepsilon_i \right) - g_{ni}(\omega, \varepsilon_i) \right) \tag{8.78}$$

indexed by $\omega, h \in \Omega$. The bound from Theorem 2.14.1 in Van der Vaart and Wellner (1996) is[19]

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

where $K > 0$ is a constant and

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

[19] Similar to the proof in Lemma 8.9, to apply Theorem 2.14.1 in Van der Vaart and Wellner, we need to reprove the symmetrization lemma in Lemma 2.3.1 and reapply the Hoeffding's inequality in Lemma 2.2.7 and the maximal inequality in Corollary 2.2.5 (all in Van der Vaart and Wellner) for the empirical process in (8.78) and the bound function in (8.77).

We now show that the uniform entropy integral $J\left(1, \mathcal{F}_{ni}\left(\varepsilon_{i}\right)\right)$ (defined as in 8.60) 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

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

It is an indicator with the argument being a linear function of ω. We can show that the set $\left\{\left(g_{n,ij}\left(\omega, \varepsilon_{ij}\right), j \neq i\right): \omega \in \Omega\right\}$ has a pseudo-dimension of at most T,[20] so it is manageable (Corollary 4.10 in Pollard (1990)). Note that $W_{n,ij}$ is a $d_{\theta} \times 1$ vector that does not depend on ω. From the stability results in Section 5 in Pollard (1990), 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 manageable. Therefore, the set $\mathcal{F}_{ni}\left(\varepsilon_{i}\right)$ is manageable and has a finite uniform entropy integral bounded by some $\bar{J}<\infty$.

Observe that

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

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 8.9 we can show that the set of functions

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

[20] To see this, by the definition of pseudo-dimension, it suffices to show that for each index set $I=\left\{j_{1}, \ldots, j_{T+1}\right\} \in\{1, \ldots, n\} \backslash\{i\}$ 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

$$U_{n,ij}+\frac{2(n-1)}{n-2} Z_{j}^{\prime} \Phi_{ni} \Lambda_{ni} \omega-\varepsilon_{ij}\left\{\begin{array}{ll}
>c_{j} & \text{for } j \in J \\
<c_{j} & \text{for } j \in I \backslash J
\end{array}\right.$$

Since $Z_{j}^{\prime} \Phi_{ni} \Lambda_{ni} \in \mathbb{R}^{T}$ for all j, there exists a non-zero vector $\tau=\left(\tau_{1}, \ldots, \tau_{T+1}\right) \in \mathbb{R}^{T+1}$ such that $\sum_{t=1}^{T+1} \tau_{t} Z_{j_{t}}^{\prime} \Phi_{ni} \Lambda_{ni}=0$, so $\sum_{t=1}^{T+1} \tau_{t} \frac{2(n-1)}{n-2} Z_{j_{t}}^{\prime} \Phi_{ni} \Lambda_{ni} \omega=0$ for all $\omega \in \Omega$. If $\sum_{t=1}^{T+1} \tau_{t}\left(U_{n,ij_{t}}-\varepsilon_{ij_{t}}-c_{j_{t}}\right) \leq 0$, it is impossible to find a $\omega \in \Omega$ satisfying these inequalities for the choice $J=\left\{j_{t} \in I: \tau_{t}>0\right\}$, because this would lead to the contradiction $\sum_{t=1}^{T+1} \tau_{t}\left(U_{n,ij_{t}}-\varepsilon_{ij_{t}}-c_{j_{t}}\right)=\sum_{t=1}^{T+1} \tau_{t}\left(U_{n,ij_{t}}-\varepsilon_{ij_{t}}-c_{j_{t}}\right)+\sum_{t=1}^{T+1} \tau_{t} \frac{2(n-1)}{n-2} Z_{j_{t}}^{\prime} \Phi_{ni} \Lambda_{ni} \omega=\sum_{t=1}^{T+1} \tau_{t}\left(U_{n,ij_{t}}+\frac{2(n-1)}{n-2} Z_{j_{t}}^{\prime} \Phi_{ni} \Lambda_{ni} \omega-\varepsilon_{ij_{t}}-c_{j_{t}}\right)>0$. If $\sum_{t=1}^{T+1} \tau_{t}\left(U_{n,ij_{t}}-\varepsilon_{ij_{t}}-c_{j_{t}}\right)>0$, we would choose $J=\left\{j_{t} \in I: \tau_{t} \leq 0\right\}$ to reach a similar contradiction.

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

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

indexed by $\omega, h \in \Omega$, i.e.,

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

Therefore,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\omega, h \in \Omega} \frac{1}{n-1} \sum_{j \neq i} \eta_{n,ij}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right) - \sup_{\omega, h \in \Omega} \frac{1}{n-1} \sum_{j \neq i} \mathbb{E}\left[\eta_{n,ij}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right)\middle| X, p_n\right]\middle| X, p_n\right] \\
\leq& \frac{1}{\sqrt{n-1}} \mathbb{E}\left[\sup_{\omega, h \in \Omega} \left|\mathbb{G}_n \eta_{ni}^2 \left(\omega + \frac{h}{n^\kappa}, \omega, \varepsilon_{ij}\right)\right| \middle| X, p_n\right] \\
\leq& \frac{K^\eta \bar{J}^\eta \bar{W}^2}{\sqrt{n-1}} \equiv \frac{M^\eta}{\sqrt{n-1}}. \qquad (8.79)
\end{aligned}$$

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

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

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 (8.79)

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

Combining the results we obtain the upper bound

$$\mathbb{E}\left[\sup_{\omega,h\in\Omega} \left\| \mathbb{G}_n W_{ni} \left(g_{ni} \left(\omega + \frac{h}{n^\kappa}, \varepsilon_i \right) - g_{ni}\left(\omega, \varepsilon_i\right) \right) \right\|^2 \middle| X, p_n \right] \leq K\bar{J} \left(\frac{M^\eta}{\sqrt{n-1}} + \frac{M}{n^\kappa} \right). \tag{8.80}$$

Hence the upper bound in (8.76) 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 (8.73) the probability $P_{n,ij}^*(\omega)$ is differentiable in ω with the derivative

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

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

$$T_{3n} = \frac{1}{\sqrt{n(n-1)}} \sum_i \sum_{j\neq i} W_{n,ij} \left(\nabla_{\omega'} P_{n,ij}^*(\omega_{ni}^*)(\omega_{ni} - \omega_{ni}^*) + O_p\left(\|\omega_{ni} - \omega_{ni}^*\|^2 \right) \right)$$

By Lemma 8.7, 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 $\varphi^{\omega}_{n,ij}$ defined in Lemma 8.7 and the remainder r^{ω}_{ni} satisfies $\max_{1\leq i\leq n}\|r^{\omega}_{ni}\| = o_p\left(n^{-1/2}\right)$ (Lemma 8.8(ii)). Denote by $J^{\omega}_{ni}\left(\omega^*_{ni}\right)$ the $d_\theta \times T$ Jacobian matrix

$$J^{\omega}_{ni}\left(\omega^*_{ni}\right) = \frac{1}{n-1}\sum_{j\neq i} W_{n,ij}\nabla_{\omega'}P^*_{n,ij}\left(\omega^*_{ni}\right),\ i = 1,\ldots,n.$$

and by $\bar{W}_{ni}$ the average weight/instrument

$$\bar{W}_{ni} = \frac{1}{n-1}\sum_{j\neq i} W_{n,ij},\ i = 1,\ldots,n.$$

Note that both $J^{\omega}_{ni}\left(\omega^*_{ni}\right)$ and $\bar{W}_{ni}$ are bounded uniformly over i. Replacing $\omega_{ni}\left(\varepsilon_i\right) - \omega^*_{ni}$ with its asymptotically linear approximation we derive

$$\begin{aligned} T_{3n} =& \frac{1}{\sqrt{n}}\sum_i\left(\frac{1}{n-1}\sum_{j\neq i} W_{n,ij}\nabla_{\omega'}P^*_{n,ij}\left(\omega^*_{ni}\right)\right)\left(\frac{1}{\sqrt{n-1}}\sum_{j\neq i}\varphi^{\omega}_{n,ij}\left(\omega^*_{ni},\varepsilon_{ij}\right) + \sqrt{n-1}r^{\omega}_{ni}\right) \\ &+ \frac{1}{\sqrt{n}}\sum_i\left(\frac{1}{n-1}\sum_{j\neq i} W_{n,ij}\right)O_p\left(\sqrt{n-1}\left\|\omega_{ni} - \omega^*_{ni}\right\|^2\right) \\ =& T^l_{3n} + r_{1n} + r_{2n}, \end{aligned} \tag{8.81}$$

where

$$T^l_{3n} = \frac{1}{\sqrt{n\left(n-1\right)}}\sum_i\sum_{j\neq i} J^{\omega}_{ni}\left(\omega^*_{ni}\right)\varphi^{\omega}_{n,ij}\left(\omega^*_{ni},\varepsilon_{ij}\right),$$

and

$$\begin{aligned} r_{1n} &= \sqrt{\frac{n-1}{n}}\sum_i J^{\omega}_{ni}\left(\omega^*_{ni}\right)r^{\omega}_{ni}, \\ r_{2n} &= \sqrt{\frac{n-1}{n}}\sum_i \bar{W}_{ni}O_p\left(\left\|\omega_{ni} - \omega^*_{ni}\right\|^2\right). \end{aligned}$$

The term T^l_{3n} in (8.81) contributes to the asymptotic distribution and has an asymptotically normal distribution. It captures the random variation in $\omega_{ni}\left(\varepsilon_i\right)$. 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 (8.75) is however $o_p\left(1\right)$.

Step 4: T_{4n}.

Observe that $\mathbb{E}\left[Y_n|\, X, p_n\right] = 0$, so

$$\begin{aligned} 0 &= \mathbb{E}\left[T_{1n} + T_{2n} + T_{3n} + T_{4n}|\, X, p_n\right] \\ &= \mathbb{E}\left[T_{1n} + T_{2n} + T_{3n}^l + r_{1n} + r_{2n} + T_{4n}\middle|\, X, p_n\right] \end{aligned}$$

Clearly $\mathbb{E}\left[T_{1n}|\, X, p_n\right] = \mathbb{E}\left[T_{3n}^l\middle|\, X, p_n\right] = 0$. We have shown in Step 2 that $\mathbb{E}\left[T_{2n}^2|\, X, p_n\right] = o\left(1\right)$, so $\mathbb{E}\left[T_{2n}|\, X, p_n\right] = o\left(1\right)$. This implies that

$$\mathbb{E}\left[r_{1n} + r_{2n} + T_{4n}|\, X, p_n\right] = \mathbb{E}\left[r_{1n}|\, X, p_n\right] + \mathbb{E}\left[r_{2n}|\, X, p_n\right] + T_{4n} = o\left(1\right).$$

Hence,

$$r_{1n} + r_{2n} + T_{4n} = r_{1n} + r_{2n} - \mathbb{E}\left[r_{1n}|\, X, p_n\right] - \mathbb{E}\left[r_{2n}|\, X, p_n\right] + o\left(1\right).$$

Note that

$$\begin{aligned} \mathbb{E}\left[r_{1n}|\, X, p_n\right] &= \sqrt{\frac{n-1}{n}} \sum_i J_{ni}^{\omega}\left(\omega_{ni}^*\right) \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right], \\ \mathbb{E}\left[r_{2n}|\, X, p_n\right] &= \sqrt{\frac{n-1}{n}} \sum_i \bar{W}_{ni} \mathbb{E}\left[O_p\left(\|\omega_{ni} - \omega_{ni}^*\|^2\right)\middle|\, X, p_n\right]. \end{aligned}$$

Below we show that the two centered remainders $r_{1n} - \mathbb{E}\left[r_{1n}|\, X, p_n\right]$ and $r_{2n} - \mathbb{E}\left[r_{2n}|\, X, p_n\right]$ are both $o_p\left(1\right)$.

We show $r_{1n} - \mathbb{E}\left[r_{1n}|\, X, p_n\right] = o_p\left(1\right)$ and the proof for $r_{2n} - \mathbb{E}\left[r_{2n}|\, X, p_n\right]$ is similar. By Chebyshev's inequality, for any $\delta > 0$,

$$\begin{aligned} &\Pr\left(\|r_{1n} - \mathbb{E}\left[r_{1n}|\, X, p_n\right]\| > \delta|\, X, p_n\right) \\ \leq &\frac{1}{\delta^2} \mathbb{E}\left(\|r_{1n} - \mathbb{E}\left[r_{1n}|\, X, p_n\right]\|^2\middle|\, X, p_n\right) \\ = &\frac{n-1}{n\delta^2} \sum_i \mathbb{E}\left(\left(r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]\right)' J_{ni}^{\omega}\left(\omega_{ni}^*\right)' J_{ni}^{\omega}\left(\omega_{ni}^*\right) \left(r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]\right)\middle|\, X, p_n\right) \\ \leq &\frac{n-1}{\delta^2} \max_i \mathbb{E}\left(\left(r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]\right)' J_{ni}^{\omega}\left(\omega_{ni}^*\right)' J_{ni}^{\omega}\left(\omega_{ni}^*\right) \left(r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]\right)\middle|\, X, p_n\right), \end{aligned}$$

where the equality follows because (1) conditional on X and p_n, each r_{ni}^{ω} depends on ε_i only, (2) ε_{ij} are i.i.d. by Assumption 1, and (3) $r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]$ is mean zero. From Lemma 8.8(ii), $\max_{1\leq i\leq n} \|r_{ni}^{\omega}\| = o_p\left(n^{-1/2}\right)$, by the dominated convergence theorem we have $\mathbb{E}\left[\max_{1\leq i\leq n} \|r_{ni}^{\omega}\||\, X, p_n\right] = o\left(n^{-1/2}\right)$, so $\max_{1\leq i\leq n} \|r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]\| \leq \max_{1\leq i\leq n} \|r_{ni}^{\omega}\| + \mathbb{E}\left[\max_{1\leq i\leq n} \|r_{ni}^{\omega}\||\, X, p_n\right] = o_p\left(n^{-1/2}\right)$. This together with the boundedness of $J_{ni}\left(\omega_{ni}^*\right)$ implies that the quadratic form $\left(r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]\right)' J_{ni}\left(\omega_{ni}^*\right)' J_{ni}\left(\omega_{ni}^*\right) \left(r_{ni}^{\omega} - \mathbb{E}\left[r_{ni}^{\omega}|\, X, p_n\right]\right)$ is

$o_p\left(n^{-1}\right)$ uniformly over i. By the dominated convergence theorem again, we obtain

$$\mathbb{E}\left(\max_{1\leq i\leq n}\left(r^{\omega}_{ni}-\mathbb{E}\left[r^{\omega}_{ni}\right|X,p_n]\right)'J_{ni}\left(\omega^*_{ni}\right)'J_{ni}\left(\omega^*_{ni}\right)\left(r^{\omega}_{ni}-\mathbb{E}\left[r^{\omega}_{ni}\right|X,p_n]\right)\Bigg|X,p_n\right)=o\left(\frac{1}{n}\right),$$

This shows that

$$r_{1n}-\mathbb{E}\left[r_{1n}\right|X,p_n]=o_p\left(1\right).$$

Similarly, with $O_p\left(\left\|\omega_{ni}-\omega^*_{ni}\right\|^2\right)$ in place of r^{ω}_{ni} and $\bar{W}_{ni}$ in place of $J_{ni}\left(\omega^*_{ni}\right)$, and by $\max_{1\leq i\leq n}\left\|\omega_{ni}\left(\varepsilon_i\right)-\omega^*_{ni}\right\|^2=o_p\left(n^{-1/2}\right)$ (Lemma 8.8(i)), we can derive that

$$r_{2n}-\mathbb{E}\left[r_{2n}\right|X,p_n]=o_p\left(1\right).$$

Combining the results yields

$$r_{1n}+r_{2n}+T_{4n}=r_{1n}+r_{2n}-\mathbb{E}\left[r_{1n}\right|X,p_n]-\mathbb{E}\left[r_{2n}\right|X,p_n]+o\left(1\right)=o_p\left(1\right).$$

Now we return to the two main terms T_{1n} in (8.75) and T^l_{3n} in (8.81). Both are normalized averages of independent random variables. If we define the $d_\theta\times 1$ random vector

$$Y^l_{ni}=\frac{1}{\sqrt{n\left(n-1\right)}}\sum_{j\neq i}W_{n,ij}\left(g_{n,ij}\left(\omega^*_{ni},\varepsilon_{ij}\right)-P^*_{n,ij}\left(\omega^*_{ni}\right)\right)+J^{\omega}_{ni}\left(\omega^*_{ni}\right)\varphi^{\omega}_{n,ij}\left(\omega^*_{ni},\varepsilon_{ij}\right)$$

then

$$T_{1n}+T^l_{3n}=\sum_i Y^l_{ni}.$$

Note that conditional on X and p_n, Y^l_{ni}, $i=1,\ldots,n$, is an independent triangular array becuase each Y^l_{ni} 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^l_{ni} 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 $\sum_i Y^l_{ni}$.

Conditional on X and p_n, Y^l_{ni} has mean 0. By independence of Y^l_{ni}, the variance of $\sum_i Y^l_{ni}$ is given by

$$\sum_i Var\left(Y^l_{ni}\right|X,p_n)=\sum_i\mathbb{E}\left[Y^l_{ni}\left(Y^l_{ni}\right)'\Big|X,p_n\right]=\frac{1}{n\left(n-1\right)}\sum_i\sum_{j\neq i}\Sigma_{n,ij}=\Sigma_n$$

where $\Sigma_{n,ij}$ is defined in (8.70).

Since Y^l_{ni} is a vector, we verify that the conditions in the Lindeberg-Feller CLT hold for $a'\sum_i Y^l_{ni}$ for any vector of constants $a\in\mathbb{R}^{d_\theta}$ so that $\left(a'\Sigma_n a\right)^{-1/2}a'\sum_i Y^l_{ni}$ converges in distribution to $N(0,1)$. By the Cramér–Wold theorem, this implies that $\Sigma_n^{-1/2}\sum_i Y_{ni}^l$ converges in distribution to $N(0, I_{d_\theta})$.

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 $\xi > 0$,

$$\lim_{n\to\infty}\frac{1}{a'\Sigma_n a}\sum_i \mathbb{E}\left[\left(a'Y_{ni}^l\right)^2 1\left\{\left|a'Y_{ni}^l\right| \geq \xi\sqrt{a'\Sigma_n a}\right\}\right| X, p_n\bigg] = 0. \tag{8.82}$$

We have

$$\begin{aligned}&\sum_i \mathbb{E}\left[\left(a'Y_{ni}^l\right)^2 1\left\{\left|a'Y_{ni}^l\right| \geq \xi\sqrt{a'\Sigma_n a}\right\}\right| X, p_n\bigg] \\ \leq &\mathbb{E}\left[\sum_i \left(a'Y_{ni}^l\right)^2 1\left\{\frac{\max_{1\leq i\leq n}\left|a'Y_{ni}^l\right|}{\sqrt{a'\Sigma_n a}} \geq \xi\right\}\right| X, p_n\Bigg],\end{aligned}$$

Note that $\sum_i \left(a'Y_{ni}^l\right)^2$ has a finite expectation and is therefore $O_p(1)$. Hence if

$$\frac{\max_{1\leq i\leq n}\left|a'Y_{ni}^l\right|}{\sqrt{a'\Sigma_n a}} = o_p(1) \tag{8.83}$$

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 (8.83) 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 of 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

$$a'Y_{ni}^{l}=\frac{1}{\sqrt{n\left(n-1\right)}}\sum_{j\neq i}a'\left(W_{n,ij}\left(g_{n,ij}\left(\omega_{ni}^{*},\varepsilon_{ij}\right)-P_{n,ij}^{*}\left(\omega_{ni}^{*}\right)\right)+J_{ni}^{\omega}\left(\omega_{ni}^{*}\right)\varphi_{n,ij}^{\omega}\left(\omega_{ni}^{*},\varepsilon_{ij}\right)\right).$$

Each term in the average is bounded by

$$\begin{aligned}&\left|a'\left(W_{n,ij}\left(g_{n,ij}\left(\omega_{ni}^{*},\varepsilon_{ij}\right)-P_{n,ij}^{*}\left(\omega_{ni}^{*}\right)\right)+J_{ni}^{\omega}\left(\omega_{ni}^{*}\right)\varphi_{n,ij}^{\omega}\left(\omega_{ni}^{*},\varepsilon_{ij}\right)\right)\right|\\&\leq\left\|a\right\|\left(2\left\|W_{n,ij}\right\|+\left\|J_{ni}^{\omega}\left(\omega_{ni}^{*}\right)\right\|\left\|\varphi_{n,ij}^{\omega}\left(\omega_{ni}^{*},\varepsilon_{ij}\right)\right\|\right)\equiv M_{n,ij}\leq M_n<\infty,\end{aligned}$$

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 (e.g., Theorem 2.8 in Boucheron, Lugosi, Massart (2013)), we have for any $t>0$,

$$\Pr\left(\left|a'Y_{ni}^{l}\right|>t\right|X,p_n\left.\right)\leq2\exp\left(-\frac{n\left(n-1\right)t^{2}}{2\sum_{j\neq i}M_{n,ij}^{2}}\right).$$

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

$$\mathbb{E}\left[\max_{1\leq i\leq n}\left(a'Y_{ni}^{l}\right)^{2}\right|X,p_n\left.\right]\leq\frac{6K\ln\left(n+1\right)\sum_{j\neq i}M_{n,ij}^{2}}{n\left(n-1\right)}\leq\frac{6KM_n\ln\left(n+1\right)}{n}\rightarrow0,$$

so (8.83) holds and and the Lindeberg condition (8.82) 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

$$Y_n = \sum_i Y_{ni}^l + o_p\left(1\right),$$

we conclude that by Slutsky's theorem Y_n has the asymptotic distribution

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

■

8.4 Proofs in Section 5

Proof of Theorem 5.1. We prove the theorem in the general case where both $\mathcal{T}_+$ and $\mathcal{T}_-$ are nonempty. The proof also holds for special cases where $\mathcal{T}_+$ is empty (i.e., all the eigenvalues of $V_i\left(X,\sigma\right)$ are nonpositive) or $\mathcal{T}_-$ is empty (i.e., all the eigenvalues of $V_i\left(X,\sigma\right)$ are nonnegative) without modification. Note that the latter special case has been proved in Theorem 3.2 in a slightly stronger sense.

From Proposition 3.1, the expected utility satisfies

$$\begin{aligned}
&\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\
=&\sum_{j\neq i} G_{ij}\left(U_{ij}\left(X,\sigma\right)-\varepsilon_{ij}\right) \\
&+\sum_t \lambda_{it}\left(X,\sigma\right)\max_{\omega_t\in\mathbb{R}}\left\{\frac{2\left(n-1\right)}{n-2}\sum_{j\neq i} G_{ij}Z_j'\phi_{it}\left(X,\sigma\right)\omega_t-\frac{\left(n-1\right)^2}{n-2}\omega_t^2\right\} \\
=&\max_{(\omega_t)_{t\in\mathcal{T}_+}}\min_{(\omega_t)_{t\in\mathcal{T}_-}}\sum_{j\neq i} G_{ij}\left(U_{ij}\left(X,\sigma\right)+\frac{2\left(n-1\right)}{n-2}Z_j'\sum_t \phi_{it}\left(X,\sigma\right)\lambda_{it}\left(X,\sigma\right)\omega_t-\varepsilon_{ij}\right) \\
&-\frac{\left(n-1\right)^2}{n-2}\sum_t \lambda_{it}\left(X,\sigma\right)\omega_t^2 \qquad (8.84) \\
=&\max_{(\omega_t)_{t\in\mathcal{T}_+}}\min_{(\omega_t)_{t\in\mathcal{T}_-}}\sum_{j\neq i} G_{ij}\left(U_{ij}\left(X,\sigma\right)+\frac{2\left(n-1\right)}{n-2}Z_j'\Phi_i\left(X,\sigma\right)\Lambda_i\left(X,\sigma\right)\omega-\varepsilon_{ij}\right) \\
&-\frac{\left(n-1\right)^2}{n-2}\omega'\Lambda_i\left(X,\sigma\right)\omega \qquad (8.85)
\end{aligned}$$

The second equality in (8.84) 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 (8.84) and (8.85) does not matter.

Denote by $\tilde{\Pi}\left(G_i,\omega,\varepsilon_i,X,\sigma\right)$ the objective function of the maximin problem in (8.85)

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

We have

$$\begin{aligned}
&\max_{G_i}\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma\right] \\
=&\max_{G_i}\max_{(\omega_t)_{t\in\mathcal{T}_+}}\min_{(\omega_t)_{t\in\mathcal{T}_-}}\tilde{\Pi}_i\left(G_i,\omega,\varepsilon_i,X,\sigma\right) \\
\leq&\max_{(\omega_t)_{t\in\mathcal{T}_+}}\min_{(\omega_t)_{t\in\mathcal{T}_-}}\max_{G_i}\tilde{\Pi}_i\left(G_i,\omega,\varepsilon_i,X,\sigma\right) \\
=&\max_{(\omega_t)_{t\in\mathcal{T}_+}}\min_{(\omega_t)_{t\in\mathcal{T}_-}}\Pi_i\left(\omega,\varepsilon_i,X,\sigma\right) \qquad (8.86)
\end{aligned}$$

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 $\max_{(\omega_t)_{t\in\mathcal{T}_+}} \min_{(\omega_t)_{t\in\mathcal{T}_-}} \tilde{\Pi}_i(G_i,\omega) \le \max_{(\omega_t)_{t\in\mathcal{T}_+}} \min_{(\omega_t)_{t\in\mathcal{T}_-}} \max_{G_i} \tilde{\Pi}_i(G_i,\omega)$ 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 (8.86) holds because for any ω, $\tilde{\Pi}_i(G_i,\omega)$ is separable in each G_{ij} so the optimal G_{ij} is given by (5.1) with $\omega_i(X,\varepsilon_i,\sigma)$ replaced by ω and $\max_{G_i} \tilde{\Pi}_i(G_i,\omega) = \Pi_i(\omega)$.

Since $\omega_i(X,\varepsilon_i,\sigma)$ is a solution to the maximin problem in the last line of (8.86), similarly as in Lemma 8.1 it satisfies the first-order condition

$$
\begin{aligned}
&\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)\\
=&\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{ij}(X,\sigma)+\frac{2(n-1)}{n-2}Z_j'\Phi_i(X,\sigma)\,\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)-\varepsilon_{ij}\ge 0\right\}\Lambda_i(X,\sigma)\,\Phi_i'(X,\sigma)\,Z_j
\end{aligned}
$$

a.s., which implies

$$
\Phi_i(X,\sigma)\,\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)=\frac{1}{n-1}V_i(X,\sigma)\sum_{j\neq i}G_{ij}(X,\varepsilon_i,\sigma)\,Z_j,\quad \text{a.s.} \tag{8.87}
$$

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 maxmin value of $\Pi(\omega,\varepsilon_i,X,\sigma)$ is given by

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

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

$$
\begin{aligned}
&\omega_i (X, \varepsilon_i, \sigma)' \Lambda_i (X, \sigma) \omega_i (X, \varepsilon_i, \sigma) \\
=&\omega_i (X, \varepsilon_i, \sigma)' \Lambda_i (X, \sigma) \Lambda_i^+ (X, \sigma) \Lambda_i (X, \sigma) \omega_i (X, \varepsilon_i, \sigma) \\
=&\omega_i (X, \varepsilon_i, \sigma)' \Lambda_i (X, \sigma) \Phi_i (X, \sigma)' \Phi_i (X, \sigma) \Lambda_i^+ (X, \sigma) \Phi_i (X, \sigma)' \Phi_i (X, \sigma) \Lambda_i (X, \sigma) \omega_i (X, \varepsilon_i, \sigma) \\
=&\frac{1}{(n-1)^2} \left(V_i (X, \sigma) \sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) Z_j \right)' V_i^+ (X, \sigma) \left(V_i (X, \sigma) \sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) Z_j \right), \quad \text{a.s.}
\end{aligned}
$$

where we have used (8.87) to derive the last equality. Therefore,

$$
\begin{aligned}
&\max_{(\omega_t)_{t \in \mathcal{T}_+}} \min_{(\omega_t)_{t \in \mathcal{T}_-}} \Pi_i (\omega, \varepsilon_i, X, \sigma) \\
=&\sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) (U_{ij} (X, \sigma) - \varepsilon_{ij}) \\
&+ \frac{2}{n-2} \left(\sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) Z_j \right)' V_i (X, \sigma) \left(\sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) Z_j \right) \\
&- \frac{1}{n-2} \left(V_i (X, \sigma) \sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) Z_j \right)' V_i^+ (X, \sigma) \left(V_i (X, \sigma) \sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) Z_j \right) \\
=&\sum_{j \neq i} G_{ij} (X, \varepsilon_i, \sigma) (U_{ij} (X, \sigma) - \varepsilon_{ij}) \\
&+ \frac{1}{n-2} \sum_{j \neq i} \sum_{k \neq i} G_{ij} (X, \varepsilon_i, \sigma) G_{ik} (X, \varepsilon_i, \sigma) Z_j' V_i (X, \sigma) Z_k, \quad a.s. \\
=&\mathbb{E} \left[U_i(G_i (X, \varepsilon_i, \sigma), G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right], \quad a.s. \qquad (8.88)
\end{aligned}
$$

Combining (8.86) and (8.88) yields

$$
\begin{aligned}
&\max_{G_i} \mathbb{E} \left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right] \\
\leq &\max_{(\omega_t)_{t \in \mathcal{T}_+}} \min_{(\omega_t)_{t \in \mathcal{T}_-}} \Pi_i (\omega, \varepsilon_i, X, \sigma) \\
=&\mathbb{E} \left[U_i(G_i (X, \varepsilon_i, \sigma), G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right], \quad a.s.
\end{aligned}
$$

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, $G_i (X, \varepsilon_i, \sigma)$ is an optimal solution almost surely.

As for the uniqueness, $G_i (X, \varepsilon_i, \sigma)$ 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 $\Lambda_i(X,\sigma)\,\omega_i(X,\varepsilon_i,\sigma)$ follows from the uniqueness of $G_i(X,\varepsilon_i,\sigma)$, (8.87) and the invertibility of $\Phi_i(X,\sigma)$. The proof is complete. ■

Proof of Theorem 5.3. For simplicity, we omit the arguments (X,σ) (or (X_i,σ)) whenever possible. Define $\tilde{\omega}_{ni}(\varepsilon_i) = \Phi_{ni}\omega_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i = \Phi_i\omega_i$. We can represent the finite n and limiting conditional choice probabilities in terms of $\tilde{\omega}_{ni}(\varepsilon_i)$ 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

$$P_{ij}(X_i,X_j,\sigma) = \Pr\left(U_{ij} + 2Z_j'V_i\tilde{\omega}_i - \varepsilon_{ij} \geq 0 \middle| X_i,X_j,\sigma\right).$$

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 is a one-to-one mapping between ω and $\Phi_{ni}\omega$ and between ω and $\Phi_i\omega$. Therefore, $\tilde{\omega}_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i$ can be equivalently solved from the transformed maximization problems

$$\max_{\tilde{\omega}}\tilde{\Pi}_{ni}(\tilde{\omega},\varepsilon_i,X,\sigma) = \max_{\tilde{\omega}}\frac{1}{n-1}\sum_{j\neq i}\left[U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'\tilde{\omega} - \varepsilon_{ij}\right]_+ - \frac{n-1}{n-2}\tilde{\omega}'V_{ni}\tilde{\omega}$$

and

$$\max_{\tilde{\omega}}\tilde{\Pi}_i(\tilde{\omega},X_i,\sigma) = \max_{\tilde{\omega}}\mathbb{E}\left[\left[U_{ij} + 2Z_j'V_i\tilde{\omega} - \varepsilon_{ij}\right]_+ \middle| X_i,\sigma\right] - \tilde{\omega}'V_i\tilde{\omega}.$$

It is sufficient to consider $\tilde{\omega}_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i$. The advantage of the change of variables is that we can 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},

$$\begin{aligned}&|P_{n,ij}(X,\sigma) - P_{ij}(X_i,X_j,\sigma)|\\ \leq&\mathbb{E}\left[\left|1\left\{U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'V_{ni}\tilde{\omega}_{ni}(\varepsilon_i) \geq \varepsilon_{ij}\right\} - 1\left\{U_{ij} + 2Z_j'V_i\tilde{\omega}_i \geq \varepsilon_{ij}\right\}\right| X,\sigma\right]\\ \leq&\mathbb{E}\left[\left|1\left\{U_{n,ij} + \frac{2(n-1)}{n-2}Z_j'V_{ni}\tilde{\omega}_{ni}(\varepsilon_i) \geq \varepsilon_{ij}\right\} - 1\left\{U_{n,ij} + 2Z_j'V_{ni}\tilde{\omega}_i \geq \varepsilon_{ij}\right\}\right| X,\sigma\right]\\ &+\mathbb{E}\left[\left|1\left\{U_{n,ij} + 2Z_j'V_{ni}\tilde{\omega}_i \geq \varepsilon_{ij}\right\} - 1\left\{U_{ij} + 2Z_j'V_i\tilde{\omega}_i \geq \varepsilon_{ij}\right\}\right| X,\sigma\right].\end{aligned} \tag{8.89}$$

Observe that

$$\begin{aligned}
&U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_j'V_{ni}\tilde{\omega}_{ni}\left(\varepsilon_i\right)-\left(U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i\right)\\
=&2Z_j'V_{ni}\left(\tilde{\omega}_{ni}\left(\varepsilon_i\right)-\tilde{\omega}_i\right)+\frac{2}{n-2}Z_j'V_{ni}\tilde{\omega}_{ni}\left(\varepsilon_i\right)\\
\equiv&\Delta_{ni}\left(\varepsilon_i\right)
\end{aligned}$$

so the first term in the last expression in (8.89) can be bounded by

$$\begin{aligned}
&\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right|>\delta_n\right|X,\sigma)\\
&+\Pr\left(U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i-\delta_n\leq\varepsilon_{ij}\leq U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i+\delta_n\right|X,\sigma\Big)
\end{aligned}\tag{8.90}$$

for an arbitrary $\delta_n>0$. This is because if ε_{ij} lies between $U_{n,ij}+\frac{2(n-1)}{n-2}Z_j'V_{ni}\tilde{\omega}_{ni}\left(\varepsilon_i\right)$ and $U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i$, and if their difference $\Delta_{ni}\left(\varepsilon_i\right)$ is at most δ_n, then ε_{ij} must lie between $U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i-\delta_n$ and $U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i+\delta_n$.

From now on we fix X_i and X_j and let $X_{-ij}=(X_k,k\neq i,j)$ be random. Note that $U_{n,ij}$, V_{ni}, $\tilde{\omega}_{ni}\left(\varepsilon_i\right)$, and $\Delta_{ni}\left(\varepsilon_i\right)$ depend on $X=(X_i,X_j,X_{-ij})$, so given X_i and X_j they are random because of X_{-ij}.

Given X_i and σ, $\tilde{\omega}_{ni}\left(\varepsilon_i\right)-\tilde{\omega}_i=o_p\left(1\right)$ by Lemma 8.11. Since V_{ni} and $\tilde{\omega}_{ni}\left(\varepsilon_i\right)=\Phi_{ni}\omega_{ni}\left(\varepsilon_i\right)$ are bounded, we have $\Delta_{ni}\left(\varepsilon_i\right)=o_p\left(1\right)$. Hence,

$$\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right|>\delta_n\right|X_i,\sigma)\rightarrow 0$$

as $n\rightarrow\infty$. From this result, we can derive that given X_i, X_j, and σ, the random variable

$$P_{ni}^{\Delta}\left(X_i,X_j,X_{-ij},\sigma,\delta_n\right)\equiv\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right|>\delta_n\right|X_i,X_j,X_{-ij},\sigma)$$

must be $o_p\left(1\right)$. If it is not true, there is a $\kappa>0$ such that

$$\Pr\left(P_{ni}^{\Delta}\left(X_i,X_j,X_{-ij},\sigma,\delta_n\right)>\kappa\right|X_i,X_j,\sigma\Big)\nrightarrow 0$$

as $n \to \infty$, then by iterated expectations

$$\begin{aligned}
&\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right| > \delta_n \right| X_i, X_j, \sigma) \\
=&\mathbb{E}\left[P_{ni}^{\Delta}\left(X_i, X_j, X_{-ij}, \sigma, \delta_n\right)\right| X_i, X_j, \sigma\Big] \\
\geq&\kappa \Pr\left(P_{ni}^{\Delta}\left(X_i, X_j, X_{-ij}, \sigma, \delta_n\right) > \kappa \right| X_i, X_j, \sigma\Big) \\
&+ \mathbb{E}\left[P_{ni}^{\Delta}\left(X_i, X_j, X_{-ij}, \sigma, \delta_n\right) 1\left\{ P_{ni}^{\Delta}\left(X_i, X_j, X_{-ij}, \sigma, \delta_n\right) \leq \kappa \right\}\right| X_i, X_j, \sigma\Big] \\
\geq&\kappa \Pr\left(P_{ni}^{\Delta}\left(X_i, X_j, X_{-ij}, \sigma, \delta_n\right) > \kappa \right| X_i, X_j, \sigma\Big) \nrightarrow 0
\end{aligned}$$

and thus

$$\begin{aligned}
\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right| > \delta_n \right| X_i, \sigma) &= \mathbb{E}\left[\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right| > \delta_n \right| X_i, X_j, \sigma)\right| X_i, \sigma\Big] \\
&= \sum_{t=1}^{T} \Pr\left(X_j = x_t\right) \Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right| > \delta_n \right| X_i, X_j = x_t, \sigma) \nrightarrow 0.
\end{aligned}$$

Therefore, given X_i, X_j, and σ we must have

$$\Pr\left(\left|\Delta_{ni}\left(\varepsilon_i\right)\right| > \delta_n \right| X_i, X_j, X_{-ij}, \sigma) = o_p(1). \tag{8.91}$$

For the second term in the bound in (8.90), by the mean-value theorem

$$\begin{aligned}
&\Pr\left(U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i - \delta_n \leq \varepsilon_{ij} \leq U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i + \delta_n \right| X_i, X_j, X_{-ij}, \sigma\Big) \\
=&F_{\varepsilon}\left(U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i + \delta_n\right) - F_{\varepsilon}\left(U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i - \delta_n\right) \\
=&2 f_{\varepsilon}\left(U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i + t_{n,ij}\delta_n\right) \delta_n
\end{aligned}$$

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 $O_p(1)$ given X_i, X_j, and σ. We choose $\delta_n \to 0$ as $n \to \infty$, so

$$\Pr\left(U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i - \delta_n \leq \varepsilon_{ij} \leq U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i + \delta_n \right| X_i, X_j, X_{-ij}, \sigma\Big) = o_p(1). \tag{8.92}$$

Combining (8.91) and (8.92), given X_i, X_j, and σ the first term in the last expression in (8.89) is $o_p(1)$

$$\mathbb{E}\left[\left| 1\left\{ U_{n,ij} + \frac{2(n-1)}{n-2} Z_j' V_{ni} \tilde{\omega}_{ni}\left(\varepsilon_i\right) \geq \varepsilon_{ij} \right\} - 1\left\{U_{n,ij} + 2Z_j' V_{ni} \tilde{\omega}_i \geq \varepsilon_{ij}\right\} \right| \Bigg| X_i, X_j, X_{-ij}, \sigma \right] = o_p(1).$$

The last term in (8.89) satisfies

$$\begin{aligned}
&\mathbb{E}\left[\left|1\left\{U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i \geq \varepsilon_{ij}\right\}-1\left\{U_{ij}+2Z_j'V_i\tilde{\omega}_i \geq \varepsilon_{ij}\right\}\right|\middle| X_i, X_j, X_{-ij}, \sigma\right] \\
=&\left|F_\varepsilon\left(U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i\right)-F_\varepsilon\left(U_{ij}+2Z_j'V_i\tilde{\omega}_i\right)\right| \\
=&2f_\varepsilon\left(\tilde{t}_{n,ij}\right)\left|U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i-\left(U_{ij}+2Z_j'V_i\tilde{\omega}_i\right)\right| \\
\leq&2f_\varepsilon\left(\tilde{t}_{n,ij}\right)\left(\left|U_{n,ij}-U_{ij}\right|+2\left\|V_{ni}-V_i\right\|\left\|\tilde{\omega}_i\right\|\right)
\end{aligned}$$

for some $\tilde{t}_{n,ij}$ that lies 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, $U_{n,ij}-U_{ij}=o_p(1)$ and $V_{ni}-V_i=o_p(1)$ by Assumption 7, the last term in (8.89) is also $o_p(1)$

$$\mathbb{E}\left[\left|1\left\{U_{n,ij}+2Z_j'V_{ni}\tilde{\omega}_i \geq \varepsilon_{ij}\right\}-1\left\{U_{ij}+2Z_j'V_i\tilde{\omega}_i \geq \varepsilon_{ij}\right\}\right|\middle| X_i, X_j, X_{-ij}, \sigma\right]=o_p(1).$$

Combining the results we conclude that given X_i, X_j, and σ

$$P_{n,ij}\left(X_i, X_j, X_{-ij}, \sigma\right)-P_{ij}\left(X_i, X_j, \sigma\right)=o_p(1).$$

The proof is complete. ■

Lemma 8.11 (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 $\delta>0$,*

$$\Pr\left(\left\|\tilde{\omega}_{ni}(\varepsilon_i)-\tilde{\omega}_i\right\|>\delta \middle| X_i, \sigma\right) \rightarrow 0 \tag{8.93}$$

as $n \rightarrow \infty$.

Proof. Recall that $\tilde{\omega}_{ni}(\varepsilon_i)$ and $\tilde{\omega}_i$ are solutions to the transformed maximization problems

$$\max_{\tilde{\omega}} \tilde{\Pi}_{ni}\left(\tilde{\omega}, \varepsilon_i, X, \sigma\right)=\max_{\tilde{\omega}} \frac{1}{n-1} \sum_{j \neq i}\left[U_{n,ij}+\frac{2(n-1)}{n-2} Z_j'\tilde{\omega}-\varepsilon_{ij}\right]_{+}-\frac{n-1}{n-2} \tilde{\omega}' V_{ni} \tilde{\omega}$$

and

$$\max_{\tilde{\omega}} \tilde{\Pi}_i\left(\tilde{\omega}, X_i, \sigma\right)=\max_{\tilde{\omega}} \mathbb{E}\left[\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_{+}\middle| X_i, \sigma\right]-\tilde{\omega}' V_i \tilde{\omega}. \tag{8.94}$$

Because the original maximization problem in (5.12) 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
$$\frac{\partial}{\partial c}\mathbb{E}\left[c-\varepsilon\right]_{+}=\frac{\partial}{\partial c}\int_{-\infty}^{c}\left(c-\varepsilon\right)f_{\varepsilon}\left(\varepsilon\right)d\varepsilon=F_{\varepsilon}\left(c\right).$$
The first-order condition of (8.94) is given by

$$\nabla_{\tilde{\omega}}\tilde{\Pi}_{i}\left(\tilde{\omega},X_{i},\sigma\right)=2V_{i}\mathbb{E}\left[\left.Z_{j}F_{\varepsilon}\left(U_{ij}+2Z_{j}^{\prime}V_{i}\tilde{\omega}\right)\right|X_{i},\sigma\right]-2V_{i}\tilde{\omega}=0. \tag{8.95}$$

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 $\tilde{\omega}$, the compactness of $\tilde{\Omega}$ implies that the unique maximizer $\tilde{\omega}_i$ is also well separated. If we can further establish a uniform LLN for the objective functions, i.e.,

$$\sup_{\tilde{\omega}}\left|\tilde{\Pi}_{ni}\left(\tilde{\omega},\varepsilon_{i},X,\sigma\right)-\tilde{\Pi}_{i}\left(\tilde{\omega},X_{i},\sigma\right)\right|=o_{p}(1) \tag{8.96}$$

as $n\to\infty$, following a standard consistency proof (e.g. Newey and McFadden (1994)) we can prove (8.93).

By triangle inequality

$$\begin{aligned}
&\sup_{\tilde{\omega}}\left|\tilde{\Pi}_{ni}\left(\tilde{\omega},\varepsilon_{i},X,\sigma\right)-\tilde{\Pi}_{i}\left(\tilde{\omega},X_{i},\sigma\right)\right|\\
\leq&\sup_{\tilde{\omega}}\frac{1}{n-1}\sum_{j\neq i}\left|\left[U_{n,ij}+\frac{2\left(n-1\right)}{n-2}Z_{j}^{\prime}V_{ni}\tilde{\omega}-\varepsilon_{ij}\right]_{+}-\left[U_{ij}+2Z_{j}^{\prime}V_{i}\tilde{\omega}-\varepsilon_{ij}\right]_{+}\right|\\
&+\sup_{\tilde{\omega}}\left|\frac{1}{n-1}\sum_{j\neq i}\left[U_{ij}+2Z_{j}^{\prime}V_{i}\tilde{\omega}-\varepsilon_{ij}\right]_{+}-\mathbb{E}\left[\left.\left[U_{ij}+2Z_{j}^{\prime}V_{i}\tilde{\omega}-\varepsilon_{ij}\right]_{+}\right|X_{i},\sigma\right]\right|\\
&+\sup_{\tilde{\omega}}\left|\frac{n-1}{n-2}\tilde{\omega}^{\prime}V_{ni}\tilde{\omega}-\tilde{\omega}^{\prime}V_{i}\tilde{\omega}\right|.
\end{aligned} \tag{8.97}$$

Because $\left|[x]_{+}-[y]_{+}\right|\leq|x-y|$, the first term on the right-hand side can be bounded by

$$\begin{aligned}
&\sup_{\tilde{\omega}}\frac{1}{n-1}\sum_{j\neq i}\left|\left(U_{n,ij}-U_{ij}\right)+2Z_{j}^{\prime}\left(V_{ni}-V_{i}\right)\tilde{\omega}+\frac{2}{n-2}Z_{j}^{\prime}V_{ni}\tilde{\omega}\right|\\
\leq&\max_{j\neq i}\left|U_{n,ij}-U_{ij}\right|+2\left\Vert V_{ni}-V_{i}\right\Vert\sup_{\tilde{\omega}}\left\Vert\tilde{\omega}\right\Vert+\frac{2}{n-2}\left\Vert V_{ni}\right\Vert\sup_{\tilde{\omega}}\left\Vert\tilde{\omega}\right\Vert\\
=&o_{p}\left(1\right),
\end{aligned}$$

where the last equality follows because $\max_{j\neq i}\left|U_{n,ij}-U_{ij}\right|=o_p\left(1\right)$ and $V_{ni}-V_i=o_p\left(1\right)$ by Assumption 7, and $\left\Vert V_{ni}\right\Vert$ and $\sup_{\tilde{\omega}}\left\Vert\tilde{\omega}\right\Vert$ are bounded. Similarly, we can bound the last term on the right-hand side of (8.97) by

$$\begin{aligned}
&\sup_{\tilde{\omega}} \left|\tilde{\omega}'\left(V_{ni}-V_i\right)\tilde{\omega}\right| + \sup_{\tilde{\omega}} \frac{1}{n-2}\left|\tilde{\omega}' V_{ni}\tilde{\omega}\right| \\
\leq &\left(\left\|V_{ni}-V_i\right\| + \frac{1}{n-2}\left\|V_{ni}\right\|\right)\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|^2 \\
= &o_p\left(1\right).
\end{aligned}$$

For the second term on the right-hand side of (8.97), 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 with mean $\mathbb{E}\left(\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+\Big| X_i,\sigma\right)$. 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 absolutely integrable since

$$\begin{aligned}
\mathbb{E}\left[\left[U_{ij}+2\left\|V_i\right\|\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|-\varepsilon_{ij}\right]_+\Bigg| X_i,\sigma\right] &\leq \mathbb{E}\left[\left|U_{ij}+2\left\|V_i\right\|\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|-\varepsilon_{ij}\right|\Bigg| X_i,\sigma\right] \\
&\leq \left(\mathbb{E}\left[\left(U_{ij}+2\left\|V_i\right\|\sup_{\tilde{\omega}}\left\|\tilde{\omega}\right\|-\varepsilon_{ij}\right)^2\Bigg| X_i,\sigma\right]\right)^{1/2} \\
&< \infty.
\end{aligned}$$

Note that $U_{ij}+2Z_j'V_i\tilde{\omega}$ is linear in ω and the max function $[x]_+$ is Lipschitz in x. We can show that the function $\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+$ is 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

$$\sup_{\tilde{\omega}}\left|\frac{1}{n-1}\sum_{j\neq i}\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+ - \mathbb{E}\left[\left[U_{ij}+2Z_j'V_i\tilde{\omega}-\varepsilon_{ij}\right]_+\Big| X_i,\sigma\right]\right| = o_p\left(1\right).$$

Hence (8.96) 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

$$\begin{aligned}
&U_{n,ij}(X,\sigma) - U_{ij}(X_i, X_j, \sigma)\\
=&\frac{1}{n-2}\sum_{k\neq i,j}\sigma_{jk}(X_j,X_k)\beta_4(X_i,X_j,X_k) - \mathbb{E}\left[\sigma_{jk}(X_j,X_k)\beta_4(X_i,X_j,X_k)\middle| X_i,X_j,\sigma\right]\\
&-\frac{1}{n-2}Z_j'V_{ni}(X,\sigma)Z_j,
\end{aligned}$$

and for $s,t=1,\ldots,T$,

$$\begin{aligned}
&V_{ni,st}(X,\sigma) - V_{i,st}(X_i,\sigma)\\
=&\frac{1}{n-3}\sum_{l\neq i,j,k}\sigma_{jl}(x_s,X_l)\sigma_{kl}(x_t,X_l)\gamma_2(X_i,x_s,x_t) - \mathbb{E}\left[\sigma_{jl}(x_s,X_l)\sigma_{kl}(x_t,X_l)\gamma_2(X_i,x_s,x_t)\middle| X_i,\sigma\right].
\end{aligned}$$

Denote

$$\Delta^U(X_i,X_j,X_k,\sigma) = \sigma_{jk}(X_j,X_k)\beta_4(X_i,X_j,X_k) - \mathbb{E}\left[\sigma_{jk}(X_j,X_k)\beta_4(X_i,X_j,X_k)\middle| X_i,X_j,\sigma\right],$$

and

$$\begin{aligned}
&\Delta^V_{st}(X_i,x_s,x_t,X_l,\sigma)\\
&=\sigma_{jl}(x_s,X_l)\sigma_{kl}(x_t,X_l)\gamma_2(X_i,x_s,x_t) - \mathbb{E}\left[\sigma_{jl}(x_s,X_l)\sigma_{kl}(x_t,X_l)\gamma_2(X_i,x_s,x_t)\middle| X_i,\sigma\right].
\end{aligned}$$

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

$$\begin{aligned}
&\Pr\left(\max_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)\right| > \delta\middle| X_i,X_j,\sigma\right)\\
&\leq \frac{1}{\delta^2}\mathbb{E}\left[\max_{j\neq i}\left(\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)\right)^2\middle| X_i,X_j,\sigma\right].
\end{aligned}$$

Because X_k, $k\neq i,j$, are i.i.d. (Assumption 7(i)), we have

$$\begin{aligned}
\max_{j\neq i}\left(\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i,X_j,X_k,\sigma)\right)^2 &= \max_{j\neq i}\frac{1}{(n-2)^2}\sum_{k\neq i,j}\left(\Delta^U(X_i,X_j,X_k,\sigma)\right)^2\\
&\leq \frac{1}{n-2}\max_{j\neq i}\max_{k\neq i,j}\left(\Delta^U(X_i,X_j,X_k,\sigma)\right)^2.
\end{aligned}$$

Since $\Delta^U(X_i, X_j, X_k, \sigma)$ is bounded uniformly in j and k, the expectation of the last term can be bounded as

$$\mathbb{E}\left[\max_{j\neq i}\max_{k\neq i,j}\left(\Delta^U(X_i, X_j, X_k, \sigma)\right)^2\middle| X_i, X_j, \sigma\right] \leq M$$

for some $M < \infty$. Hence,

$$\Pr\left(\max_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i, X_j, X_k, \sigma)\right| > \delta\middle| X_i, X_j, \sigma\right) \leq \frac{M}{\delta^2(n-2)} \to 0$$

as $n \to \infty$. This proves

$$\max_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i, X_j, X_k, \sigma)\right| = o_p(1).$$

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,

$$\begin{aligned}
&\max_{j\neq i}\left|U_{n,ij}(X, \sigma) - U_{ij}(X_i, X_j, \sigma)\right| \\
\leq&\max_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\Delta^U(X_i, X_j, X_k, \sigma)\right| + \max_{j\neq i}\frac{1}{n-2}\left|Z_j' V_{ni}(X, \sigma) Z_j\right| \\
=&o_p(1).
\end{aligned}$$

As for $V_{ni}(X, \sigma) - V_i(X_i, \sigma)$, by Chebyshev's inequality and i.i.d. X_l

$$\begin{aligned}
&\Pr\left(\left|\frac{1}{n-3}\sum_{l\neq i,j,k}\Delta_{st}^V(X_i, x_s, x_t, X_l, \sigma)\right| > \delta\middle| X_i, \sigma\right) \\
&\leq \frac{1}{\delta^2}\mathbb{E}\left[\left(\frac{1}{n-3}\sum_{l\neq i,j,k}\Delta_{st}^V(X_i, x_s, x_t, X_l, \sigma)\right)^2\middle| X_i, \sigma\right] \\
&= \frac{1}{\delta^2(n-3)^2}\sum_{l\neq i,j,k}\mathbb{E}\left[\left(\Delta_{st}^V(X_i, x_s, x_t, X_l, \sigma)\right)^2\middle| X_i, \sigma\right] \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Hence, for $s, t = 1, \ldots, T$,

$$V_{ni,st}(X, \sigma) - V_{i,st}(X_i, \sigma) = o_p(1).$$

We conclude that

$$\|V_{ni}(X, \sigma) - V_i(X_i, \sigma)\| \leq \max_{s,t=1,\ldots,T} |V_{ni,st}(X, \sigma) - V_{i,st}(X_i, \sigma)| = o_p(1).$$

The proof is complete. ∎

References

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

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

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

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

[5] Ciliberto, F. and E. Tamer (2009). Market Structure and Multiple Equilibria in Airline Markets, *Econometrica*, 77(6), 1791-1828.

[6] De Paula, A., S. Richards-Shubik and E. Tamer. (2018) Identifying Preferences in Networks with Bounded Degree, *Econometrica*, 86(1): 263-288.

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

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

[9] Leung, M. (2015) Two-Step Estimation of Network-Formation Models with Incomplete Information, *Journal of Econometrics*, 188: 182-195.

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

[11] Mele, A. (2017): A Structural Model of Dense Network Formation, *Econometrica*, 85(3): 825–850.

[12] Menzel, K. (2016) Inference for Games with Many Players, *Review of Economic Studies,* 83(1): 306-337.

[13] Menzel, K. (2017) Strategic Network Formation with Many Players, NYU, working paper.

[14] Pollard, D. (1990) *Empirical Processes: Theory and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Volumn 2, Institute of Mathematical Statistics.

[15] Sheng, S. (2018) A Structural Econometric Analysis of Network Formation Games, *Econometrica* (accepted).

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

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

[18] Tamer, E. (2003) Incomplete Simultaneous Discrete Response Model with Multiple Equilibria, *Review of Economic Studies*, 70, 147-165.

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