Estimation of Large Network Formation Games*

Geert Ridder† Shuyang Sheng‡

December 6, 2017

(Preliminary and Incomplete)

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. The main issue is to show that the asymptotic distribution of the estimator is not dominated by the dependence

*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.

between the link choices of the agent. Monte Carlo results show that the estimation procedure performs well, even in moderately large networks.

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. Besides on observable variables, 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 is large and grows without bounds, this seems an intractable discrete choice problem. In addition the link choices of the members of the network have to be consistent. This is achieved by assuming that the realized network is a Bayesian Nash equilibrium in which agents maximize the expected utility of their link choices. The first simplification of the link choice model is due to the assumption of incomplete information under which agents base their link choices on the unobserved characteristics of the their potential links, but agents do not know the unobserved link characteristics of the potential links of the other agents and these unobserved link characteristics are independent across agents. The alternative assumption is that of complete information under which agents know not just the unobserved link characteristics of their own potential links, but also those of the links by all other agents in the network. The complete information models are the hardest to estimate and they achieve set and not point identification of the parameters of the utility function (Miyauchi (2013) and Sheng (2017)).

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

A further simplification is obtained if the utility function for links depends on the position of a potential partner in a restrictive way. Leung (2015) shows that if the utility function depends on the choice of potential partners in a separable way, e.g. through the (weighted) number of links of the potential partner, then the link choice that maximizes the expected utility is myopic, because it is equivalent to a sequence of myopic choices in which an agent chooses to form a link with another member if the utility of that link is greater than the utility of not forming that link. This equivalence to a sequence of myopic choices does not hold, if the utility function depends on the choice of potential partners in a non-separable way. An important example is that the utility function depends on the number of links in common which means that the agent considers linking to two members simultaneously. Allowing for the utility to depend on links-in-common is important if networks exhibit clustering.

The main contribution of this paper is that 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 myopic link choices. Using the Legendre transform we linearize the expected utility function. This linearization introduces an auxiliary variable in the expected utility function 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. Thanks to 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 is due to 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 means that the dependence cannot be ignored in the calculation of the asymptotic variance of the two-step estimator that is based on n^2 observations on links. However the dependence can be accounted for in an obvious way.

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 considers a number of extensions of the model and estimator that will be addressed in future research. Section 6 reports the results of a simulation study.

2 Model

Suppose that n individuals can play to form links. The links form a network, which we denote by $G \in \mathcal{G}$. This is an $n \times n$ binary matrix. The (i, j) element $G_{ij} = 1$ if i,j are linked and 0 otherwise. The diagonal elements G_{ii} are set to be 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 5.1.

Each individual i has a vector of observed characteristics $X_i \in \mathcal{X}$ and a vector of unobserved preference 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 preference for link ij. We assume that the characteristic profile $X = (X_1, \ldots, X_n) \in \mathcal{X}^n$ is public information of all the individuals, but the shock vector ε_i is the private information of i. We also assume that the private 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 $F_\varepsilon(\theta_\varepsilon)$ supported over $\mathbb{R}$ that is absolutely continuous with respect to the Lebesgue measure. $F_\varepsilon(\theta_\varepsilon)$ is known up to parameter $\theta_\varepsilon \in \Theta_\varepsilon \subset \mathbb{R}^{d_\varepsilon}$. (ii) $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ and X are independent.*

Utility Given the network G, characteristic profile X, and private shocks ε_i, individual i has the utility

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

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 the specification in (2.1), $u_i(G_j, X; \beta)$ represents the incremental utility from a link that is separable in i's links G_i. A typical example of $u_i(G_j, X; \beta)$ is

$$u_i(G_j, X; \beta) = \beta_0 + X_i'\beta_1 + |X_i - X_j|'\beta_2 + G_{ji}\beta_3 + \frac{1}{n-2} \sum_{k \neq i,j} G_{jk}\beta_4(X_i, X_j, X_k) \tag{2.2}$$

The first four terms in (2.2) capture the direct utility from the link with j, which consists of the homophily effect (β_2) and the reciprocal effect (β_3). The last term in (2.2) captures the indirect utility from j's friends, which may vary in the characteristics of the individuals involved. This specification is similar to that in Leung (2015).

The main difference between our setting and Leung (2015) is that in addition to the utility that is separable in one's own links, we also allow for the utility from indirect friends that are nonseparable in one's own links, represented by $v_i(G_j, G_k, X; \gamma)$. For example, if we want to capture the utility from friends in common, we may specify

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

where the first term captures the effect of friends in common that are directly connected and the second term captures the effect of friends in common that are indirectly connected. Allowing for such nonseparable externalities is crucial if we want to model networks with a feature of clustering, that is, two individuals with friends in common are more likely to become friends (Jackson, 2008).

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

Equilibrium Let $G_i\left(X, \varepsilon_i\right)$ denote individual i's link 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 $G_{-i} = \left(G_j\right)_{j\neq i}$ 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 $\mathbb{E}\left[U_i(g_i, G_{-i}, X, \varepsilon_i)\right| X, \varepsilon_i]$ 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 $\sigma_i\left(g_i\right| X) = \Pr\left(G_i\left(X, \varepsilon_i\right) = g_i\right| X)$ be the conditional probability that individual i chooses g_i given X. The independence of the private shocks also implies that the link decisions G_i are independent across i given X, so individual i's belief about the link decisions of others can be represented as $\sigma_{-i}\left(g_{-i}\right| X) = \prod_{j\neq i}\sigma_j\left(g_j\right| X)$. Let $\sigma\left(X\right) = \left\{\sigma_i\left(g_i\right| X), \forall g_i \in \mathcal{G}_i.\forall i\right\}$ denote the belief profile. For a given belief profile σ, the expected utility of individual i is given by

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

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

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

and

$$\mathbb{E}\left[v_i\left(G_j,G_k,X\right)\middle|\,X,\sigma\right]=\mathbb{E}\left[G_{jk}\middle|\,X,\sigma\right]\mathbb{E}\left[G_{kj}\middle|\,X,\sigma\right]G_{jk}\gamma_1\left(X_i,X_j,X_k\right) \\ +\frac{1}{n-3}\sum_{l\neq i,j,k}\mathbb{E}\left[G_{jl}\middle|\,X,\sigma\right]\mathbb{E}\left[G_{kl}\middle|\,X,\sigma\right]\gamma_2\left(X_i,X_j,X_k\right) \tag{2.5}$$

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

$$\Pr\left(G_i=g_i\middle|\,X,\sigma\right) \\ =\Pr\left(\mathbb{E}\left[U_i(g_i,G_{-i},X,\varepsilon_i)\middle|\,X,\varepsilon_i,\sigma\right]\geq\max_{\tilde{g}_i\in\mathcal{G}_i}\mathbb{E}\left[U_i(\tilde{g}_i,G_{-i},X,\varepsilon_i)\middle|\,X,\varepsilon_i,\sigma\right]\middle|\,X,\sigma\right). \tag{2.6}$$

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

$$\sigma_i^*\left(g_i\middle|\,X\right)=\Pr\left(G_i=g_i\middle|\,X,\sigma^*\left(X\right)\right) \tag{2.7}$$

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' observed characteristics. We say that an equilibrium $\sigma\left(X\right)$ is *symmetric* if for i and j with $X_i=X_j$, we have $\sigma_i\left(X\right)=\sigma_j\left(X\right)$, where $\sigma_i\left(X\right)=\left\{\sigma_i\left(g_i\middle|\,X\right),\forall g_i\in\mathcal{G}_i\right\}$ denotes the conditional choice probability profile of individual i. In social networks, individuals typically have no identities and are labelled arbitrarily. It is thus reasonable to assume that a realized equilibrium is symmetric because otherwise the conditional choice probability profile of an individual may depend on how we label the individuals. It can be shown that there exists a symmetric equilibrium. We assume that the equilibrium realized in data is symmetric.

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

Proof. See the appendix. ■

3 Representation of the Optimal Link Decisions

The difficulty in estimating the formation model in the previous section results from the fact that the expected utility in (2.4) is nonlinear in the link decisions of an individual, so it is not clear how to represent the optimal link decisions and quantify their correlation. In this section, we propose a novel approach to overcome this difficulty. The main idea is to linearize the expected utility through certain auxiliary variables so that the optimal link decisions can be derived in closed form. In the next section, we show how to use this closed-form representation to characterize the dependence among the links of an individual and derive asymptotic properties for estimators as the network size n goes to infinity.

To facilitate the presentation, we assume that X_i has a finite support so that the optimal link decisions can be derived in matrix notation.

Assumption 2 (Discrete X) *The support of X_i has a finite number of distinct values, $\mathcal{X} = \{x_1, \ldots, x_T\}$.*

To proceed, observe that the expected utility of friends in common $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\right| X, \sigma]$ in (2.5) is symmetric in subscripts j and k. Moreover, in a symmetric equilibrium $\sigma\left(X\right)$ if two individuals j and k have the same characteristics (i.e., $X_j = X_k$) we have $\sigma_j\left(X\right) = \sigma_k\left(X\right)$. This implies that $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\right| X, \sigma]$ depends on j and k only through the values of X_j and X_k. Therefore, given $X_{-jk} = \left(X_l\right)_{l \neq j,k}$, we can view $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\right| X, \sigma]$ as a symmetric function of X_j and X_k. For any $s, t = 1, \ldots, T$, let $V_{i,st}\left(X, \sigma\right)$ denote the value of $\mathbb{E}\left[v_i\left(G_j, G_k, X\right)\right| X, \sigma]$ when $X_j = x_s$ and $X_k = x_t$,

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

Arrange all such values in a $T \times T$ matrix

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

We take two steps to linearize the expected utility. First, since $V_i\left(X, \sigma\right)$ is a real symmetric matrix, it has a real spectral decomposition

$$V_i\left(X, \sigma\right) = \Phi_i\left(X, \sigma\right) \Lambda_i\left(X, \sigma\right) \Phi_i\left(X, \sigma\right)' \tag{3.1}$$

where $\Lambda_i\left(X, \sigma\right) = diag\left(\lambda_{i1}\left(X, \sigma\right), \ldots, \lambda_{iT}\left(X, \sigma\right)\right)$ is the $T \times T$ diagonal matrix of the eigenvalues $\lambda_{it}\left(X, \sigma\right) \in \mathbb{R}$, $t = 1, \ldots, T$, and $\Phi_i\left(X, \sigma\right) = \left(\phi_{i1}\left(X, \sigma\right), \ldots, \phi_{iT}\left(X, \sigma\right)\right)$ is the $T \times T$

orthogonal matrix of the eigenvectors $\phi_{it}(X,\sigma) \in \mathbb{R}^T$, $t = 1, \ldots, T$. Using this spectral decomposition, we can transform the double summation in the expected utility in (2.4) to a canonical form that involves only squares of linear functions of G_i.

Second, once we obtain the canonical form of the double summation, we can represent the squares of linear functions of G_i using a special case of Legendre transform (Rockafellar, 1970), namely,

$$Y^2 = \max_{\omega \in \mathbb{R}} \left\{ 2Y\omega - \omega^2 \right\} \tag{3.2}$$

Note that the maximand on the right hand side of (3.2) is linear in Y. Replacing each square function by the right hand side of (3.2) with an auxiliary variable ω, we can represent the expected utility in a form that is linear in G_i. The details are given in Proposition 3.1.

Proposition 3.1 *Under Assumptions 1-2, the expected utility in (2.4) is equal to*

$$\begin{aligned}
&\mathbb{E}\left[U_i(G_i, G_{-i}, X, \varepsilon_i) \middle| X, \varepsilon_i, \sigma \right] \\
&= \sum_{j \neq i} G_{ij} \left(U_{ij}(X,\sigma) - \varepsilon_{ij} \right) + \frac{(n-1)^2}{n-2} \sum_t \lambda_{it}(X,\sigma) \left(\frac{1}{n-1} \sum_{j \neq i} G_{ij} D_j' \phi_{it}(X,\sigma) \right)^2 \\
&= \sum_{j \neq i} G_{ij} \left(U_{ij}(X,\sigma) - \varepsilon_{ij} \right) \\
&\quad + \frac{(n-1)^2}{n-2} \sum_t \lambda_{it}(X,\sigma) \max_{\tilde{\omega}_t \in \mathbb{R}} \left\{ 2 \left(\frac{1}{n-1} \sum_{j \neq i} G_{ij} D_j' \phi_{it}(X,\sigma) \right) \tilde{\omega}_t - \tilde{\omega}_t^2 \right\}
\end{aligned} \tag{3.3}$$

where

$$U_{ij}(X,\sigma) = \mathbb{E}\left[u_i(G_j, X) \middle| X, \sigma \right] - \frac{1}{n-2} D_j' diag\left(V_i(X,\sigma) \right) D_j \tag{3.4}$$

and

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

Proof. See the appendix. ■

The optimal link decisions are solved by maximizing the expected utility over G_i. Observe that under the transformation in (3.3) the expected utility becomes linear in G_i. If we can move the maximization over $\tilde{\omega} = (\tilde{\omega}_1, \ldots, \tilde{\omega}_T)' \in \mathbb{R}^T$ in (3.3) to the beginning of the expected utility function and interchange this maximization with the maximization over G_i, the optimal link decisions can be solved easily. Moving the maximization over $\tilde{\omega}$ to the beginning of the expected function is valid if the eigenvalues $\lambda_{it}(X,\sigma)$ are assumed to be nonnegative (Assumption 3). Moreover, it is guaranteed by Lemma 3.2 below that the swap of the two maximizations does not change the optimal solution. Therefore, we can obtain the optimal link decisions from the linearized expected utility in closed form.

To get around the eigenvalues and eigenvectors when representing the optimal link decisions, we further transform the maximization over $\tilde{\omega}$ to a maximization over another vector ω, defined by $\omega = (\omega_1, \ldots, \omega_T)' = V_i \Phi_i \tilde{\omega} \in \mathbb{R}^T$. Let V_i^+ denote the Moore-Penrose generalized inverse of V_i. By the spectral decomposition in (3.1) and the orthogonality of Φ_i, we have $\omega = \Phi_i \Lambda_i \tilde{\omega}$ and $\omega' V_i^+ \omega = \tilde{\omega}' \Lambda_i \tilde{\omega}$. It can be shown that the maximization over $\tilde{\omega}$ can be represented equivalently as a maximization over ω. Formal results about the optimal link decisions are established in Theorem 3.3.

Assumption 3 (Positive semi-definite V_i) *For any X and σ the eigenvalues of $V_i(X,\sigma)$ are nonnegative, i.e., $\lambda_{it}(X,\sigma) \geq 0$ for all i and all t.*

Lemma 3.2 *For any function $f(x,y) : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ with $\sup_{x,y} f(x,y) < \infty$, we have*

$$\max_y \max_x f(x,y) = \max_x \max_y f(x,y) \tag{3.5}$$

Therefore, if there is a unique (x^*, y^*) such that $f(x^*, y^*) = \max_y \max_x f(x,y)$, then (x^*, y^*) is also the unique solution to $\max_x \max_y f(x,y)$.

Theorem 3.3 *Under Assumptions 1-3, the optimal link decisions $G_i(X, \varepsilon_i, \sigma) = (G_{ij}(X, \varepsilon_i, \sigma))_{j \neq i}$ are given by*

$$G_{ij}(X, \varepsilon_i, \sigma) = 1\left\{U_{ij}(X,\sigma) + \frac{n-1}{n-2} 2 D_j' \omega_i(\varepsilon_i, X, \sigma) - \varepsilon_{ij} \geq 0\right\}, \ \forall j \neq i \tag{3.6}$$

where $\omega_i(\varepsilon_i, X, \sigma)$ is a maximizer of the problem

$$\max_\omega \Pi_i(\omega, \varepsilon_i, X, \sigma) = \sum_{j \neq i} \left[U_{ij}(X,\sigma) + \frac{n-1}{n-2} 2 D_j' \omega - \varepsilon_{ij}\right]_+ - \frac{(n-1)^2}{n-2} \omega' V_i^+(X,\sigma) \omega \tag{3.7}$$

with $[x]_+ = \max\{x, 0\}$. Moreover, the optimal $G_{ij}(X, \varepsilon_i, \sigma)$ is unique almost surely.

Proof. See the appendix. ■

4 Estimation

In this section, we discuss how to estimate the model parameter θ. We propose a two-step estimation procedure, where in the first step we estimate link formation probabilities directly from data and in the second step we estimate the parameter θ using the first-step estimates. Unlike most econometric literature on games of incomplete information which typically assumes a large number of independent games, we use links from a single network and consider the asymptotic properties when the network size n increases to infinity.[2] The model in Section 2 implies that conditional on an equilibrium links are independent across individuals due to the independence of private information, while links within an individual are correlated, as indicated by (3.6). The main challenge in deriving the asymptotic properties of our estimator is then to deal with the correlation among the links of an individual using the representation of the links in (3.6) and show how such correlation affects the rate of convergence and asymptotic distribution of the proposed estimator.

To facilitate the asymptotic analysis, we add subscript n to G, X, and ε and denote them as G_n, X_n, and ε_n, where $G_n = (G_{n,ij})_{\forall i \neq j}$ and $\varepsilon_n = (\varepsilon_{n,ij})_{\forall i \neq j}$. We also use $G_{ni} = (G_{n,ij})_{j \neq i}$ and $\varepsilon_{ni} = (\varepsilon_{n,ij})_{j \neq i}$ to denote the ith row of G_n and ε_n. For convenience, we write $X_n = (X_{n,ij})_{\forall i \neq j}$, where $X_{n,ij} = (X_{ni}, X_{nj})$ represents the characteristics of pair i and j. Let $\sigma_n(X_n)$ denote the equilibrium observed in data.

We start with the first step. Given the utility specification in (2.2) and (2.3), the expected utility in (2.4) depends on the equilibrium $\sigma_n(X_n)$ only through the link formation probabilities $p_{n,ij}(X_n)$, $\forall i \neq j$, where

$$p_{n,ij}(X_n) = \Pr\left(G_{n,ij} = 1 \middle| X_n, \sigma_n(X_n)\right)$$

is the probability that individuals i and j form a link. It thus suffices to consider such link formation probabilities in the first step.

Let $p_n(X_n) = (p_{n,ij}(X_n))_{\forall i \neq j}$ be the link probability profile observed in data. Define the conditional link probability as a function of the link probability profile p

$$P_{n,ij}(X_n, p) = \Pr\left(G_{n,ij}(X_n, \varepsilon_{ni}, p) = 1 \middle| X_n, p\right) \tag{4.1}$$

where $G_{n,ij}(X_n, \varepsilon_{ni}, p)$ is given by (3.6). The equilibrium condition implies that $p_n(X_n)$ satisfies

$$p_{n,ij}(X_n) = P_{n,ij}(X_n, p_n(X_n)) \tag{4.2}$$

for all $i \neq j$.

Since the observed equilibrium $\sigma_n(X_n)$ is assumed to be symmetric, the value of $p_{n,ij}(X_n)$ depends on i and j only through $X_{n,ij}$. For any $s, t = 1, \ldots, T$, with abuse of notation let

[2] If we observe more than one network, we can implement the first-step estimation network by network, i.e., we estimate the link formation probabilities in each network separately. In the second step, we pool the likelihoods or moments from each network to estimate the parameter θ.

$p_{n,st}(X_n)$ denote the value of $p_{n,ij}(X_n)$ when $X_{n,ij} = x_{st} = (x_s, x_t) \in \mathcal{X}^2$,

$$p_{n,st}(X_n) = \Pr\left(G_{n,ij} = 1 \middle| X_{n,ij} = x_{st}, X_{n,-ij}, \sigma_n\right)$$

where $X_{n,-ij} = (X_{n,kl})_{\forall(k,l)\neq(i,j)}$, and let $p_n(X_n) = (p_{n,st}(X_n))_{\forall s,t}$. For simplicity, we abbreviate $p_n(X_n)$ as p_n.

The symmetry assumption reduces the dimensionality of p_n from $n(n-1)$ to T^2. We estimate each $p_{n,st}$ by the empirical frequency that all the pairs with characteristics x_{st} form a link

$$\hat{p}_{n,st} = \frac{\sum_i \sum_{j\neq i} G_{n,ij} 1\{X_{n,ij} = x_{st}\}}{\sum_i \sum_{j\neq i} 1\{X_{n,ij} = x_{st}\}} \tag{4.3}$$

Let $\hat{p}_n = (\hat{p}_{n,st})_{\forall s,t}$ denote the estimator of p_n.

With the estimates $\hat{p}_n$, we obtain an estimator $\hat{\theta}_n$ of the parameter θ in the second step by solving the system of equations

$$\hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right) = 0 \tag{4.4}$$

where $\hat{\Psi}_n(\theta, p)$ is a sample moment function defined by

$$\hat{\Psi}_n(\theta, p) = \frac{1}{n(n-1)} \sum_i \sum_{j\neq i} W_{n,ij}(\theta, p)\left(G_{n,ij} - P_{n,ij}(\theta, p)\right) \tag{4.5}$$

with $P_{n,ij}(\theta, p) = P_{n,ij}(X_n, p, \theta)$ being the conditional link probability in (4.1) and $W_{n,ij}(\theta, p)$ a $\dim(\theta) \times 1$ vector representing certain weights. Depending on whether we estimate θ by Quasi-MLE (QMLE) or GMM, the weights could be different.

Example 1 (QMLE) *We can estimate θ by QMLE. Construct the (quasi) log likelihood function*

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

Let $\hat{\theta}_n$ be the maximizer of the log likelihood with p replaced by $\hat{p}_n$

$$\max_\theta \mathcal{L}_n(\theta, \hat{p}_n)$$

It satisfies the first-order condition

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

where $\nabla_\theta P_{n,ij}(\theta,p)$ is the gradient of $P_{n,ij}(\theta,p)$ with respect to θ.[3] *It is clear that by choosing the weight*

$$W_{n,ij}(\theta,p) = \frac{\nabla_\theta P_{n,ij}(\theta,p)}{P_{n,ij}(\theta,p)(1-P_{n,ij}(\theta,p))} \tag{4.6}$$

we get the moment function in (4.5).

Example 2 (GMM) *An alternative way is to estimate θ by GMM. The equilibrium condition in (4.2) implies the moment restrictions*

$$\mathbb{E}\left[G_{n,ij} - P_{n,ij}(\theta,p) \middle| X_n, p_n\right] = 0$$

Since X_{ni} is discrete, these moment restrictions are equivalent to

$$\mathbb{E}\left[(G_{n,ij} - P_{n,ij}(\theta,p))\, 1\{X_{n,ij} = x_{st}\} \middle| X_n, p_n\right] = 0$$

for $s,t = 1,\ldots,T$. Define $m_{n,st}(\theta,p)$ to be the sample analogue of the above moment restriction

$$m_{n,st}(\theta,p) = \frac{1}{n(n-1)} \sum_i \sum_{j\neq i} (G_{n,ij} - P_{n,ij}(\theta,p))\, 1\{X_{n,ij} = x_{st}\}$$

and let $m_n(\theta,p) = (m_{n,st}(\theta,p))_{\forall s,t}$ be the $T^2 \times 1$ vector of sample moments. A GMM estimator $\hat{\theta}_n$ is the optimal solution to the problem

$$\min_\theta m_n(\theta,\hat{p}_n)' W_n(\theta,\hat{p}_n)' W_n(\theta,\hat{p}_n)\, m_n(\theta,\hat{p}_n)$$

where $W_n(\theta,p)$ is a $\dim(\theta)\times T^2$ weighting matrix. By choosing the weighting matrix $W_n(\theta,p)$ to be

$$W_n(\theta,p) = (W_{n,ij}(\theta,p)\, 1\{X_{n,ij} = x_{11}\}, \ldots, W_{n,ij}(\theta,p)\, 1\{X_{n,ij} = x_{TT}\})$$

with $W_{n,ij}(\theta,p)$ being the weight given in (4.5), we obtain the moment function in (4.5). In particular, if we choose the weight $W_{n,ij}(\theta,p)$ as in (4.6), the GMM estimator is equivalent to the QMLE estimator.

Now we examine the asymptotic properties of the estimator $\hat{\theta}_n$ defined by (4.4). Throughout the section, we treat X_n and p_n as fixed and derive the asymptotic theory conditional on X_n and p_n. The main issue in the asymptotic analysis is that links made by the same individual are correlated. The representation in (3.6) indicates that such correlation results

[3] We show in Lemma 8.3 that $P_{n,ij}(\theta,p)$ is differentiable in (θ,p).

from the presence of $\omega_{ni}(\varepsilon_{ni}, X_n, p_n)$. Recall that $\omega_{ni}(\varepsilon_{ni}, X_n, p_n)$ is a solution to the problem

$$\omega_{ni}(\varepsilon_{ni}, X_n, p_n) = \arg\max_{\omega} \Pi_{ni}(\omega, \varepsilon_{ni}, X_n, p_n) \tag{4.7}$$

where

$$\Pi_{ni}(\omega, \varepsilon_{ni}, X_n, p_n) = \sum_{j \neq i} \left[U_{n,ij}(X_n, p_n) + \frac{n-1}{n-2} 2D_j' \omega - \varepsilon_{n,ij} \right]_+ - \frac{(n-1)^2}{n-2} \omega' V_i^+(X_n, p_n) \omega \tag{4.8}$$

It is clear that $\omega_{ni}(\varepsilon_{ni}, X_n, p_n)$ depends on the entire ε_{ni}, so is a random vector. Hence two links $G_{n,ij}$ and $G_{n,ik}$ are correlated either through the randomness in $\omega_{ni}(\varepsilon_{ni}, X_n, p_n)$ or through the correlation between $\omega_{ni}(\varepsilon_{ni}, X_n, p_n)$ and $\varepsilon_{n,ij}$ or $\varepsilon_{n,ik}$. Nevertheless, by law of large numbers we expect that $\Pi_{ni}(\omega, \varepsilon_{ni}, X_n, p_n)$ converges to its conditional expectation $\mathbb{E}[\Pi_{ni}(\omega, \varepsilon_{ni}, X_n, p_n) | X_n, p_n]$ (under the normalization of $n-1$), so $\omega_{ni}(\varepsilon_{ni}, X_n, p_n)$ tends to converge to the maximizer of the conditional expectation

$$\omega_i^*(X_n, p_n) = \arg\max_{\omega} \mathbb{E}[\Pi_{ni}(\omega, \varepsilon_{ni}, X_n, p_n) | X_n, p_n] \tag{4.9}$$

Since $\omega_i^*(X_n, p_n)$ is a deterministic vector, we expect that the correlation between two links of an individual vanishes as n approaches infinity and in the limit the links become independent. Proposition 4.1 provides a formal proof that the link correlation indeed vanishes to 0 as n increases at the rate of n^{-1}. Obtaining the rate of link correlation is crucial in deriving the asymptotic distribution of the estimator. We show later that link correlation vanishing at such a rate will not slow down the rate of convergence of the estimator, but will increase the asymptotic variance.

Proposition 4.1 (Rate of link correlation) *Suppose that Assumptions 1-3 are satisfied. Given X_n and p_n, for any distinct i, j and k, we have*

$$\mathbb{E}\left((G_{n,ij} - P_{n,ij}(\theta_0, p_n))(G_{n,ik} - P_{n,ik}(\theta_0, p_n)) | X_n, p_n \right) = O\left(\frac{1}{n}\right).$$

Proof. See the appendix. ■

To establish the consistency and asymptotic distribution of the estimator $\hat{\theta}_n$, define $\Psi_n(\theta, p)$ to be the conditional expectation of $\hat{\Psi}_n(\theta, p)$ given X_n and p_n

$$\begin{aligned} \Psi_n(\theta, p) &= \mathbb{E}\left[\hat{\Psi}_n(\theta, p) \middle| X_n, p_n \right] \\ &= \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} W_{n,ij}(\theta, p) \left(\mathbb{E}[G_{n,ij} | X_n, p_n] - P_{n,ij}(\theta, p) \right) \end{aligned} \tag{4.10}$$

Let θ_0 denote the true value of θ. It is clear that $\Psi_n(\theta_0, p_n) = 0$ because $\mathbb{E}[G_{n,ij}| X_n, p_n] = P_{n,ij}(\theta_0, p_n)$. Theorem 4.2 establishes that $(\hat{\theta}_n, \hat{p}_n)$ is consistent.

Theorem 4.2 (Consistency) *Suppose that Assumptions 1-3 are satisfied. Given X_n and p_n,*

$$(\hat{\theta}_n, \hat{p}_n) - (\theta_0, p_n) \xrightarrow{p} 0$$

as $n \to \infty$.

Proof. See the appendix. ■

Next we derive the asymptotic distribution of $\hat{\theta}_n$. To account for the impact of the first-stage estimate $\hat{p}_n$, recall that $\hat{p}_n - p_n$ can be represented as

$$\hat{p}_n - p_n = \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} Q_{n,ij} \left(G_{n,ij} - P_{n,ij}(\theta_0, p_n)\right) \tag{4.11}$$

where $Q_{n,ij} = \left(Q^{st}_{n,ij}\right)_{\forall s,t}$ is a $T^2 \times 1$ vector with the components

$$Q^{st}_{n,ij} = \frac{1\{X_{n,ij} = x_{st}\}}{(n(n-1))^{-1} \sum_i \sum_{j \neq i} 1\{X_{n,ij} = x_{st}\}}$$

Define an augmented sample moment at (θ_0, p_n)

$$\begin{aligned}\tilde{\Psi}_n(\theta_0, p_n) &= \hat{\Psi}_n(\theta_0, p_n) + \nabla_p \Psi_n(\theta_0, p_n)(\hat{p}_n - p_n) \\ &= \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} \tilde{W}_{n,ij}(\theta_0, p_n)\left(G_{n,ij} - P_{n,ij}(\theta_0, p_n)\right)\end{aligned} \tag{4.12}$$

where $\nabla_p \Psi_n(\theta_0, p_n)$ is the gradient of $\Psi_n(\theta, p)$ with respect to p at (θ_0, p_n), and $\tilde{W}_{n,ij}(\theta_0, p_n)$ represents the augmented weight

$$\tilde{W}_{n,ij}(\theta_0, p_n) = W_{n,ij}(\theta_0, p_n) + \nabla_p \Psi_n(\theta_0, p_n) Q_{n,ij} \tag{4.13}$$

Define $\tilde{\Omega}_n(X_n, p_n)$ to be the conditional variance of $\sqrt{n(n-1)}\tilde{\Psi}_n(\theta_0, p_n)$

$$\tilde{\Omega}_n(X_n, p_n) = Var\left(\sqrt{n(n-1)}\tilde{\Psi}_n(\theta_0, p_n)\middle| X_n, p_n\right) \tag{4.14}$$

Theorem 4.3 shows that $\hat{\theta}_n - \theta_0$ converges at the rate of n to a normal distribution.

Theorem 4.3 (Asymptotic Distribution) *Suppose that Assumptions 1-3 are satisfied*

$$\sqrt{n(n-1)}\tilde{\Omega}_n^{-1/2}(X_n, p_n) \nabla_\theta \Psi_n(\theta_0, p_n)\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{d} N(0, I)$$

as $n \to \infty$. *Moreover,* $\tilde{\Omega}_n(X_n, p_n) = O(1)$, *so* $\hat{\theta}_n - \theta_0 = O_p(n^{-1})$.

Proof. See the appendix. ■

5 Extensions

5.1 Undirected Networks

In this section we consider the case of undirected networks. We show that the methodology in the previous sections also works for undirected networks under mild modifications. To proceed, let G_{ij} denote an undirected link between i and j. Clearly $G_{ij} = G_{ji}$. It is useful to denote by S_{ij} a directed link from i to j. Under the link announcement framework, S_{ij} represents whether i proposes to form a link with j. The link is formed if both i and j propose to form it, so $G_{ij} = S_{ij}S_{ji}$. We may write $G(S)$ to indicate that G is the network induced by proposals S.

We consider the same utility specification as in (2.1), with G_{ij} representing an undirected link. With abuse of notation we use G_{-i} to denote the submatrix of G_{-i} with the ith row and column deleted. We maintain the same information assumption. The strategy of individual i, denoted by $S_i(X, \varepsilon_i) = (S_{ij}(X, \varepsilon_i))_{j \neq i}$ is a mapping from her information (X, ε_i) to a vector of link proposals in $\mathcal{S}_i = \{0,1\}^{n-1}$. The optimal strategy maximizes her expected utility $\mathbb{E}[U_i(G(S_i, S_{-i}), X, \varepsilon_i) | X, \varepsilon_i, \sigma]$ over $s_i \in \mathcal{S}_i$, where the expectation is taken with respect to others' proposals $S_{-i} = (S_j)_{j \neq i}$. The choice probability $\sigma_i(s_i | X)$ and belief profile $\sigma(X)$ are defined similarly as before, with links G_i replaced by proposals S_i.

Given belief profile σ, individual i's expected utility is now given by

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

For the specifications in (2.2) and (2.3) we have

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

and

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

Let $\Pr\left(S_i=s_i\middle|X,\sigma\right)$ be the probability that i proposes s_i given X and σ. A Bayesian Nash equilibrium $\sigma^*\left(X\right)$ is a fixed point that solves $\sigma_i^*\left(s_i\middle|X\right)=\Pr\left(S_i=s_i\middle|X,\sigma^*\left(X\right)\right)$.

Remark 1 *A potential concern with Nash is that in undirected networks players may coordinate. This is reasonable under complete information, where pairwise stability (Jackson and Wolinsky (1996)) and Nash equilibrium are nonnested and neither of them implies the other. However, under incomplete information players won't be able to coordinate even in undirected networks; because i does not observe ε_{ji}, he cannot predict what j proposes and coordinate on that (unless in a trivial equilibrium where $\sigma\left(X\right)\equiv 0$). In fact, if we define a Bayesian version of the pairwise stability, that is, a network G is Bayesian pairwise stable if*

$$G_{ij}=1\left\{\Delta_{ij}\mathbb{E}\left[U_i(G\left(S_i,S_{-i}\right),X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right]\geq 0\ \&\ \Delta_{ji}\mathbb{E}\left[U_j(G\left(S_j,S_{-j}\right),X,\varepsilon_j)\middle|X,\varepsilon_j,\sigma\right]\geq 0\right\}$$

for any $i\neq j$, where $\Delta_{ij}\mathbb{E}\left[U_i(G\left(S_i,S_{-i}\right),X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right]$ is the expected marginal utility if i proposes the link with j, i.e.,

$$\mathbb{E}\left[U_i(G\left(S_i:S_{ij}=1,S_{-i}\right),X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right]-\mathbb{E}\left[U_i(G\left(S_i:S_{ij}=0,S_{-i}\right),X,\varepsilon_i)\middle|X,\varepsilon_i,\sigma\right]$$

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

Like in the directed case, define

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

and the matrix

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

Following the same idea in the previous sections, we can derive the optimal proposals in closed form.

Corollary 1 *Under Assumptions 1-3, the optimal link proposals* $S_i(X,\varepsilon_i,\sigma) = (S_{ij}(X,\varepsilon_i,\sigma))_{j\neq i}$ *are given by*

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

where $\omega_i(\varepsilon_i,X,\sigma)$ *is a maximizer of the problem*

$$\max_{\omega}\sum_{j\neq i}\left[U_{ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'\omega - \sigma_{ji}\varepsilon_{ij}\right]_+ - \frac{(n-1)^2}{n-2}\omega' V_i(X,\sigma)\,\omega$$

with

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

and $\sigma_{ji} = \mathbb{E}\left[S_{ji}\middle| X,\sigma\right]$. *Moreover, the optimal* $S_{ij}(X,\varepsilon_i,\sigma)$ *is unique almost surely.*

In a n-player network, given X_n and σ_n, the probability that individual i proposes to link to j and the probability that i proposes to link to both j and k are given by

$$\begin{aligned}P_{n,ij}(X_n,\sigma_n) &= \Pr\left(S_{n,ij}(X_n,\varepsilon_{ni},\sigma_n) = 1|X_n,\sigma_n\right)\\ &= \Pr\left(U_{n,ij}(X_n,\sigma_n) + \frac{n-1}{n-2}2D_j'\omega_{ni}(\varepsilon_{ni},X_n,\sigma_n) - \sigma_{n,ji}\varepsilon_{n,ij} \geq 0|X_n,\sigma_n\right)\end{aligned}$$

and

$$\begin{aligned}Q_{n,ijk}(X_n,\sigma_n) &= \Pr\left(S_{n,ij}(X_n,\varepsilon_{ni},\sigma_n) = 1, S_{n,ik}(X_n,\varepsilon_{ni},\sigma_n) = 1|X_n,\sigma_n\right)\\ &= \Pr\Bigg(U_{n,ij}(X_n,\sigma_n) + \frac{n-1}{n-2}2D_j'\omega_{ni}(\varepsilon_{ni},X_n,\sigma_n) - \sigma_{n,ji}\varepsilon_{n,ij} \geq 0,\\ &\qquad U_{n,ik}(X_n,\sigma_n) + \frac{n-1}{n-2}2D_k'\omega_{ni}(\varepsilon_{ni},X_n,\sigma_n) - \sigma_{n,ki}\varepsilon_{n,ik} \geq 0\,\Bigg|\, X_n,\sigma_n\Bigg)\end{aligned}$$

The latter is relevant because the expected utility depends on the beliefs about two proposals $S_{n,ji}S_{n,jk}$.

5.2 Limiting Game Approach

In this section, we explore the asymptotic behavior of the n-player game when n approaches infinity. Applying the representation in (3.6)-(3.7), we can show that the optimal link decisions of an individual and the probability she forms a link converge to some limiting strategy and link formation probability as n goes to infinity. The optimal link decisions in the limit has a simple form that conditional on some sufficient statistics that control for the spillover effects between one's own links, the links of an individual are formed independently. These asymptotic features are crucial for deriving simple enough while asymptotically valid estimators of the model parameters.

From (3.6) the probability that individual i forms a link to j conditional on characteristic profile X and belief profile σ is

$$\begin{aligned} P_{n,ij}\left(X_{n,ij}; X_n, \sigma_n\right) &= \Pr\left(G_{n,ij}\left(X_n, \varepsilon_{ni}, \sigma_n\right) = 1 \middle| X_n, \sigma_n\right) \\ &= \Pr\left(U_{n,ij}\left(X_n, \sigma_n\right) + \frac{n-1}{n-2} 2D_j'\omega_{ni}\left(\varepsilon_{ni}, X_n, \sigma_n\right) - \varepsilon_{n,ij} \geq 0 \middle| X_n, \sigma_n\right) \end{aligned} \tag{5.1}$$

If the utility specification ensures that $U_{n,ij}(X_n, \sigma)$ converges to some limit $U_{ij}(\sigma)$ as $n \to \infty$, as stated in Assumption 4, we expect that the link formation probability $P_{n,ij}(X_{ij}; X, \sigma)$ converges to a limit given by

$$P_{ij}\left(X_{ij}; \sigma\right) = \Pr\left(U_{ij}(\sigma) + 2D_j'\omega_i(\sigma) - \varepsilon_{ij} \geq 0 \middle| X_{ij}, \sigma\right) \tag{5.2}$$

as $n \to \infty$, where $\omega_i(\sigma)$ is a maximizer of the problem

$$\max_{\omega} \Pi\left(\omega, X_i, \sigma\right) = \mathbb{E}\left(\left[U_{ij}(\sigma) + 2D_j'\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right) - \omega' V_i^+(\sigma)\,\omega \tag{5.3}$$

The expectation in (5.3) is taken with respect to X_j and ε_{ij}. The equilibrium belief profile σ^* is a fixed point solved from

$$\sigma_{ij}^*\left(X_{ij}\right) = P\left(X_{ij}; \sigma^*\right)$$

where $\sigma_{ij}(X_{ij}) = \Pr\left(G_{ij} = 1 | X_{ij}\right)$.

Assumption 4 *For $U_{n,ij}(X_n, \sigma)$ defined in (3.4), there is $U_{ij}(\sigma)$ such that*

(a) Given any X_{ij} $\sup_\sigma \left|U_{n,ij}(X_n, \sigma) - U_{ij}(\sigma)\right| \xrightarrow{p} 0$, and

(b) Given any X_i, $\sup_\sigma \frac{1}{n-1}\sum_{j\neq i}\left|U_{n,ij}(X_n, \sigma) - U_{ij}(\sigma)\right| \xrightarrow{p} 0$.

Example 3 *Consider the separable utility specification in (2.2). For any X_n and σ, we have*

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

We verify in the appendix that $U_{ij}(\sigma)$ is given by

$$U_{ij}(\sigma) = \beta_0 + X_i'\beta_1 + |X_i - X_j|'\beta_2 + \sigma_{ji}(X)\beta_3 + \mathbb{E}\left[\left.\sum_{k \neq i,j} \sigma_{jk}(X)\beta_4(X_i, X_j, X_k)\right| X_{ij}\right]$$

satisfies Assumption 4.

We refer to $P(X_{ij}; \sigma)$ defined in (5.2) the limiting choice probability. It can be understood as the choice probability derived from a "limiting game" with infinite number of players, where each individual i forms a link with j following a limiting strategy given by

$$G_{ij} = 1\left\{U_{ij}(\sigma) + 2D_j'\omega_i(\sigma) - \varepsilon_{ij} \geq 0\right\}$$

In this limiting game, conditional on sufficient statistic $\omega_i(\sigma)$ which summarizes the spillover effects from one's other links due to nonseparable utility, each individual makes link decisions myopically and considers each link as a binary choice independent of her other links. Moreover, while statistic $\omega_{ni}(\varepsilon_{ni}, X_n, \sigma)$ in the n-player game may be dependent of $\varepsilon_{n,ij}$, the dependence vanishes as $n \to \infty$ and in the limit statistic $\omega_i(\sigma)$ is independent of ε_{ij}.

Under Assumption 5 that ensures the maximizer of $\Pi(\omega, X_i, \sigma)$ is identified, we can show that the statistic $\omega_{n,i}$ and choice probability $P_{n,ij}(X_{n,ij}; X_n, \sigma_n)$ in the finite game converge to the proposed limits given in (5.3) and (5.2) as $n \to \infty$.

Assumption 5 *For any X_i and symmetric σ, $\Pi(\omega, X_i, \sigma)$ defined in (5.3) has a unique maximizer $\omega_i^*(\sigma)$.*

Proposition 5.1 *Under Assumptions 1-5, given any X_i,*

$$\sup_{\sigma} |\omega_{ni}(\varepsilon_{ni}, X_n, \sigma) - \omega_i^*(\sigma)| \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{5.4}$$

Moreover, given any $X_{n,ij} = X_{ij}$, we have

$$\sup_{\sigma} |P_{n,ij}(X_{n,ij}; X_n, \sigma_n) - P(X_{ij}; \sigma)| \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{5.5}$$

Proof. See the appendix. ∎

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

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

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 (**??**), with $P_n(X_{ij}; X_n, p_n)$ replaced by the limiting choice probability $P(X_{ij}; p_n)$ given in (5.2). Such approximation has a substantial advantage in computation because $P(X_{ij}; p_n)$ 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-game choice probability $P_n(X_{ij}; X_n, p_n)$ does not have a closed form because $\omega_{n,i}(X, \varepsilon_i)$ depends on ε_i. It needs to be computed by simulation. In practice, we simulate $P_n(X_{ij}; X_n, p_n)$ by a frequency simulator, where in each simulation we generate a vector ε_i and for this ε_i we solve for the optimal $G_{n,ij}(X, \varepsilon_i)$ numerically. We obtain $G_{n,ij}(X, \varepsilon_i)$ by solving the integer programming problem directly when $n \leq 100$ or applying the equivalent binary representation as in (3.6) 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.

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.

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 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.

8 Appendix

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 (2.7) 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 statement. Denote $D_{it} = 1\{X_i = x_t\}$. Consider the quadratic term in the expected utility

$$\begin{aligned}
&\sum_{j\neq i}\sum_{k\neq i,j} G_{ij}G_{ik}\mathbb{E}\left[v_i(G_j, G_k, X) | X, \sigma\right] \\
&= \sum_{j\neq i}\sum_{k\neq i} G_{ij}G_{ik}\sum_s\sum_t D_{js}D_{kt}V_{i,st}(X,\sigma) - \sum_{j\neq i} G_{ij}\sum_t D_{jt}V_{i,tt}(X,\sigma) \\
&= \sum_s\sum_t V_{i,st}(X,\sigma)\sum_{j\neq i} G_{ij}D_{js}\sum_{k\neq i} G_{ik}D_{kt} - \sum_{j\neq i} G_{ij}\sum_t D_{jt}V_{i,tt}(X,\sigma) \\
&= \left(\sum_{j\neq i} G_{ij}D_j'\right) V_i(X,\sigma)\left(\sum_{j\neq i} G_{ij}D_j\right) - \sum_{j\neq i} G_{ij}D_j' diag\left(V_i(X,\sigma)\right) D_j \qquad (8.1)
\end{aligned}$$

By the real spectral decomposition of $V_i(X,\sigma)$,

$$\begin{aligned}
&\left(\sum_{j\neq i} G_{ij}D_j'\right) V_i(X,\sigma)\left(\sum_{j\neq i} G_{ij}D_j\right)\\
&=\left(\sum_{j\neq i} G_{ij}D_j'\right)\Phi_i(X,\sigma)\Lambda_i(X,\sigma)\Phi_i(X,\sigma)'\left(\sum_{j\neq i} G_{ij}D_j\right)\\
&=\left(\sum_{j\neq i} G_{ij}D_j'\Phi_i(X,\sigma)\right)\Lambda_i(X,\sigma)\left(\sum_{j\neq i} G_{ij}\Phi_i(X,\sigma)'D_j\right)\\
&=\sum_t \lambda_{it}(X,\sigma)\left(\sum_{j\neq i} G_{ij}D_j'\phi_{it}(X,\sigma)\right)^2 \qquad (8.2)
\end{aligned}$$

The desired statement then follows from (8.1), (8.2), and some simple algebra. ■

Proof of Lemma 3.2. Note that

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

Similarly,

$$\max_x\max_y f(x,y)\geq \max_y\max_x f(x,y)$$

Hence (3.5) is proved. If (x^*,y^*) is the unique solution to the LHS of (3.5), then by $f(x^*,y^*)=\max_y\max_x f(x,y)=\max_x\max_y f(x,y)$ it is also the unique solution to the RHS of (3.5). ■

Proof of Theorem 3.3. From Proposition 3.1 and Lemma 3.2

$$\begin{aligned}
&\max_{G_i}\mathbb{E}\left[U_i(G_i,G_{-i},X,\varepsilon_i)\middle| X,\varepsilon_i,\sigma\right]\\
&\Leftrightarrow \max_{G_i}\max_{\tilde\omega}\sum_{j\neq i}G_{ij}\left(U_{ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'\sum_t\phi_{it}(X,\sigma)\lambda_{it}(X,\sigma)\tilde\omega_t-\varepsilon_{ij}\right)-\frac{(n-1)^2}{n-2}\sum_t\lambda_{it}(X,\sigma)\tilde\omega_t^2\\
&\Leftrightarrow \max_{\tilde\omega}\max_{G_i}\sum_{j\neq i}G_{ij}\left(U_{ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'\Phi_i(X,\sigma)\Lambda_i(X,\sigma)\tilde\omega-\varepsilon_{ij}\right)-\frac{(n-1)^2}{n-2}\tilde\omega'\Lambda_i(X,\sigma)\tilde\omega\\
&\Leftrightarrow \max_{\omega}\max_{G_i}\sum_{j\neq i}G_{ij}\left(U_{ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'\omega-\varepsilon_{ij}\right)-\frac{(n-1)^2}{n-2}\omega'V_i^+(X,\sigma)\omega \qquad (8.3)
\end{aligned}$$

The optimal $G_i(X, \varepsilon_i, \sigma)$ follows immediately from (8.3), so does $\omega_i(\varepsilon_i, X, \sigma)$.

As for the the uniqueness, under the assumption that ε_i has a continuous distribution, the optimal $G_i(X, \varepsilon_i, \sigma)$ solved from the original problem in (2.6) is unique almost surely, so by Lemma 3.2 the optimal $G_i(X, \varepsilon_i, \sigma)$ given in (3.6) is unique almost surely. ■

8.3 Proofs in Section 4

Before we prove Proposition 4.1, we first prove two lemmas.

Lemma 8.1 *Given X_n and p_n, $\omega_{ni}(\varepsilon_{ni})$ has an asymptotic linear form*

$$\omega_{ni}(\varepsilon_{ni}) - \omega_{ni}^* = -\frac{1}{n-1}\sum_{j\neq i} H_{ni}^{-1}(\omega_{ni}^*)\,\psi_{n,ij}(\omega_{ni}^*, \varepsilon_{n,ij}) + r_n(\varepsilon_{ni}), \tag{8.4}$$

where $\psi_{n,ij}(\omega, \varepsilon_{n,ij})$ is given by

$$\psi_{n,ij}(\omega, \varepsilon_{n,ij}) = 1\left\{U_{n,ij} + \frac{n-1}{n-2}2D_j'\omega - \varepsilon_{n,ij} > 0\right\} D_j - V_i^+\omega, \tag{8.5}$$

$H_{ni}(\omega_{ni}^*)$ is the Jacobian matrix of $\frac{1}{n-1}\sum_{j\neq i}\mathbb{E}\left[\psi_{n,ij}(\omega, \varepsilon_{n,ij})\middle| X_n, p_n\right]$ at ω_{ni}^*,

$$H_{ni}(\omega_{ni}^*) = \frac{1}{n-1}\sum_{j\neq i}\frac{\partial\mathbb{E}\left[\psi_{n,ij}(\omega_{ni}^*, \varepsilon_{n,ij})\middle| X_n, p_n\right]}{\partial\omega'}, \tag{8.6}$$

and the remainder in (8.4) $r_n(\varepsilon_{ni})$ satisfies

$$r_n(\varepsilon_{ni}) = o_p\left(\frac{1}{\sqrt{n}}\right).$$

Lemma 8.2 *Given X_n and p_n, suppose that ε_{ni}^1 and $\varepsilon_{ni}^2 \in \mathbb{R}^{n-1}$ differ in at most two components, i.e., there are j and k such that $\varepsilon_{n,il}^1 = \varepsilon_{n,il}^2$ for all $l \neq i, j, k$. Then*

$$\begin{aligned}\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2) = &-\frac{1}{n-1}H_{ni}^{-1}(\omega_{ni}^*)\left(\psi_{n,ij}(\omega_{ni}^*, \varepsilon_{n,ij}^1) - \psi_{n,ij}(\omega_{ni}^*, \varepsilon_{n,ij}^2)\right.\\ &\left.+\,\psi_{n,ik}(\omega_{ni}^*, \varepsilon_{n,ik}^1) - \psi_{n,ik}(\omega_{ni}^*, \varepsilon_{n,ik}^2)\right) + r_n(\varepsilon_{ni}^1) - r_n(\varepsilon_{ni}^2)\end{aligned}$$

where $\psi_{n,ij}(\omega_{ni}^*, \varepsilon_{n,ij})$, $H_{ni}(\omega_{ni}^*)$, and $r_n(\varepsilon_{ni})$ are defined as in Lemma 8.1, and $r_n(\varepsilon_{ni}^1) - r_n(\varepsilon_{ni}^2)$ satisfies

$$r_n(\varepsilon_{ni}^1) - r_n(\varepsilon_{ni}^2) = O_p\left(\frac{1}{n}\right), \tag{8.7}$$

and

$$\mathbb{E}\left[\sup_{\varepsilon^1_{n,ij},\varepsilon^1_{n,ik}} \left\|r_n\left(\varepsilon^1_{ni}\right)-r_n\left(\varepsilon^2_{ni}\right)\right\| \middle| X_n,p_n\right] \leq O\left(\frac{1}{n}\right). \tag{8.8}$$

Proof of Lemma 8.1. Since $\Pi_{ni}(\omega,\varepsilon_{ni})$ is sub-differentiable at all ω,[4] by optimality of $\omega_{ni}(\varepsilon_{ni})$, $\Pi_{ni}(\omega,\varepsilon_{ni})$ has subgradient 0 at $\omega_{ni}(\varepsilon_{ni})$, that is, $\omega_{ni}(\varepsilon_{ni})$ satisfies the first-order condition

$$\begin{aligned}&\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega-\varepsilon_{n,ij}>0\right\}D_j - V_i^{+}\omega \\ &= -\frac{1}{n-1}\sum_{j\neq i} 1\left\{U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega-\varepsilon_{n,ij}=0\right\}diag\left(\tau\right)D_j,\end{aligned} \tag{8.9}$$

for some $\tau=(\tau_1,\ldots,\tau_T)\in[0,1]^T$. Define the right-hand side of (8.9) as $\Delta_n(\omega,\varepsilon_{ni})$. For any ω,

$$\begin{aligned}&\Pr\left(\left\|\Delta_n(\omega,\varepsilon_{ni})\right\|>0\middle| X_n,p_n\right)\\ &\leq \Pr\left(\exists j, U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega=\varepsilon_{n,ij}\middle| X_n,p_n\right)\\ &\leq \sum_{j\neq i}\Pr\left(U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega=\varepsilon_{n,ij}\middle| X_n,p_n\right)=0,\end{aligned}$$

because $\varepsilon_{n,ij}$ has a continuous distribution. Hence the first-order condition (8.9) holds with $\Delta_n(\omega,\varepsilon_{ni})$ replaced by 0 with probability 1.

Define the left-hand side of (8.9) as $\hat{\Gamma}_{ni}(\omega,\varepsilon_{ni})$

$$\hat{\Gamma}_{ni}(\omega,\varepsilon_{ni})=\frac{1}{n-1}\sum_{j\neq i}\psi_{n,ij}(\omega,\varepsilon_{n,ij}), \tag{8.10}$$

with $\psi_{n,ij}(\omega,\varepsilon_{n,ij})$ defined in (8.5)

$$\psi_{n,ij}(\omega,\varepsilon_{n,ij})=1\left\{U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega-\varepsilon_{n,ij}>0\right\}D_j-V_i^{+}\omega.$$

Let $\Gamma_{ni}(\omega)$ be the conditional expectation of $\hat{\Gamma}_{ni}(\omega,\varepsilon_{ni})$ given X_n and p_n

$$\Gamma_{ni}(\omega)=\mathbb{E}\left[\hat{\Gamma}_{ni}(\omega,\varepsilon_{ni})\middle| X_n,p_n\right]=\frac{1}{n-1}\sum_{j\neq i}\mathbb{E}\left[\psi_{n,ij}(\omega,\varepsilon_{n,ij})\middle| X_n,p_n\right], \tag{8.11}$$

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

where

$$\mathbb{E}\left[\psi_{n,ij}\left(\omega,\varepsilon_{n,ij}\right)\middle| X_n, p_n\right] = F_\varepsilon\left(U_{n,ij} + \frac{n-1}{n-2}2D_j'\omega\right)D_j - V_i^+\omega.$$

As we have shown $\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right),\varepsilon_{ni}\right) = 0$ with probability 1. Moreover, by the definition of ω_{ni}^* we have $\Gamma_{ni}\left(\omega_{ni}^*\right) = 0$. Therefore,

$$\begin{aligned}
&\sqrt{n-1}\left(\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right)\right) - \Gamma_{ni}\left(\omega_{ni}^*\right)\right)\\
&= \sqrt{n-1}\left(\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right)\right) - \hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right),\varepsilon_{ni}\right)\right)\\
&= -\sqrt{n-1}\hat{\Gamma}_{ni}\left(\omega_{ni}^*,\varepsilon_{ni}\right) - \sqrt{n-1}\left(\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right),\varepsilon_{ni}\right) - \Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right)\right) - \left(\hat{\Gamma}_{ni}\left(\omega_{ni}^*,\varepsilon_{ni}\right) - \Gamma_{ni}\left(\omega_{ni}^*\right)\right)\right).
\end{aligned} \tag{8.12}$$

The Jacobian matrix of $\Gamma_{ni}\left(\omega\right)$ is equal to $H_{ni}\left(\omega\right)$

$$H_{ni}\left(\omega\right) = \frac{\partial\Gamma_{ni}\left(\omega\right)}{\partial\omega} = \frac{2}{n-2}\sum_{j\neq i} f_\varepsilon\left(U_{n,ij} + \frac{n-1}{n-2}2D_j'\omega\right)D_jD_j' - V_i^+.$$

Clearly $H_{ni}\left(\omega\right)$ is continuous in ω. By Assumption 5, $H_{ni}\left(\omega\right)$ is continously invertible at ω_{ni}^*. Hence there exists a constant $c > 0$ such that $\left\|H_{ni}\left(\omega_{ni}^*\right)\left(\omega - \omega_{ni}^*\right)\right\| \geq c\left\|\omega - \omega_{ni}^*\right\|$ for every ω. Combining this with the differentiability of Γ_{ni} we obtain

$$\left\|\Gamma_{ni}\left(\omega\right) - \Gamma_{ni}\left(\omega_{ni}^*\right)\right\| \geq c\left\|\omega - \omega_{ni}^*\right\| + o\left(\left\|\omega - \omega_{ni}^*\right\|\right). \tag{8.13}$$

We show that $\omega_{ni}\left(\varepsilon_{ni}\right)$ is consistent for ω_{ni}^*. Note that inequality (8.13) provides an identification condition for ω_{ni}^*: for any $\xi > 0$, if $\left\|\omega - \omega_{ni}^*\right\| > \xi$, then $\left\|\Gamma_{ni}\left(\omega\right) - \Gamma_{ni}\left(\omega_{ni}^*\right)\right\| > \left(c + o\left(1\right)\right)\xi$. Let $\eta > 0$ satisfy $\eta \leq \left(c + o\left(1\right)\right)\xi$. Thus

$$\begin{aligned}
&\Pr\left(\left\|\omega_{ni}\left(\varepsilon_{ni}\right) - \omega_{ni}^*\right\| > \xi\middle| X_n, p_n\right)\\
&\leq \Pr\left(\left\|\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right)\right) - \Gamma_{ni}\left(\omega_{ni}^*\right)\right\| > \eta\middle| X_n, p_n\right)\\
&= \Pr\left(\left\|\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right)\right) - \hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right),\varepsilon_{ni}\right)\right\| > \eta\middle| X_n, p_n\right)\\
&\leq \Pr\left(\sup_\omega\left\|\hat{\Gamma}_{ni}\left(\omega,\varepsilon_{ni}\right) - \Gamma_{ni}\left(\omega\right)\right\| > \eta\middle| X_n, p_n\right),
\end{aligned} \tag{8.14}$$

where the equality follows because $\Gamma_{ni}\left(\omega_{ni}^*\right) = 0$ and $\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}\right),\varepsilon_{ni}\right) = 0$ with probability 1. We can show the consistency of $\omega_{ni}\left(\varepsilon_{ni}\right)$ by applying a uniform law of large numbers to $\hat{\Gamma}_{ni}\left(\omega,\varepsilon_{ni}\right) - \Gamma_{ni}\left(\omega\right)$.

Since

$$\hat{\Gamma}_{ni}(\omega, \varepsilon_{ni}) - \Gamma_{ni}(\omega) = \frac{1}{n-1} \sum_{j \neq i} \left(\psi_{n,ij}(\omega, \varepsilon_{n,ij}) - \mathbb{E}\left[\psi_{n,ij}(\omega, \varepsilon_{n,ij}) \right| X_n, p_n] \right),$$

and $\|\psi_{n,ij}(\omega, \varepsilon_{n,ij}) - \mathbb{E}[\psi_{n,ij}(\omega, \varepsilon_{n,ij})| X_n, p_n]\| \leq 2$, to apply the uniform law of large numbers for triangular arrays in Theorem 8.3 of Pollard (1990), we need to show that the set of arrays $\left\{ (\psi_{n,ij}(\omega, \varepsilon_{n,ij}))_{j \neq i} : \omega \in \mathbb{R}^T \right\}$ is manageable, in the sense of Definition 7.9 in Pollard (1990). Note that the set $\left\{ V_i^+ \omega : \omega \in \mathbb{R}^T \right\}$ has a pseudodimension of at most T so is manageable by Theorem 4.8 in Pollard (1990). It suffices to show that the set of arrays

$$\left\{ \left(1 \left\{ U_{n,ij} + \frac{n-1}{n-2} 2D_j' \omega - \varepsilon_{n,ij} > 0 \right\} \right)_{j \neq i} : \omega \in \mathbb{R}^T \right\} \tag{8.15}$$

also has a pseudodimension of at most T and thus is manageable as well. To this end, by the definition of pseudodimension (Definition 4.3 in Pollard (1990)), we need to determine, for any index set $I = \{j_1, \ldots, j_K\} \in \{1, \ldots, n\} \setminus \{i\}$ with $K \geq T+1$, and any point $c \in \mathbb{R}^K$, whether it is possible to find for each subset $J \subseteq I$ a $\omega \in \mathbb{R}^T$ for which

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

Since the indicator function lies between 0 and 1, choosing J to be the empty set requires that $c_j > 0$ for all j, and choosing $J = I$ requires that $c_j < 1$ for all j. The problem is equivalent to determining whether for each subset $J \subseteq I$ there is a $\omega \in \mathbb{R}^T$ satisfying

$$U_{n,ij} + \frac{n-1}{n-2} 2D_j' \omega - \varepsilon_{n,ij} \begin{cases} > 0 & \text{for } j \in J \\ < 0 & \text{for } j \in I \backslash J \end{cases}$$

Since $D_j \in \mathbb{R}^T$ for all j, there exists a non-zero vector $\tau = (\tau_1, \ldots, \tau_K) \in \mathbb{R}^K$ such that $\sum_{k=1}^K \tau_k D_{j_k} = 0$, so $\sum_{k=1}^K \tau_k \frac{n-1}{n-2} 2D_{j_k}' \omega = 0$ for all $\omega \in \mathbb{R}^T$. If $\sum_{k=1}^K \tau_k (U_{n,ij_k} - \varepsilon_{n,ij_k}) \leq 0$, it is impossible to find a $\omega \in \mathbb{R}^T$ satisfying these inequalities for the choice $J = \{j_k \in I : \tau_k > 0\}$ because this would lead to the contradiction $\sum_{k=1}^K \tau_k (U_{n,ij_k} - \varepsilon_{n,ij_k}) = \sum_{k=1}^K \tau_k (U_{n,ij_k} - \varepsilon_{n,ij_k}) + \sum_{k=1}^K \tau_k \frac{n-1}{n-2} 2D_{j_k}' \omega = \sum_{k=1}^K \tau_k \left(U_{n,ij_k} + \frac{n-1}{n-2} 2D_{j_k}' \omega - \varepsilon_{n,ij_k} \right) > 0$. If $\sum_{k=1}^K \tau_k (U_{n,ij_k} - \varepsilon_{n,ij_k}) > 0$, we would choose $J = \{j_k \in I : \tau_k \leq 0\}$ to reach a similar contradiction. We conclude that the set of arrays in (8.15) has a pseudodimension of at most T. Hence, by the uniform law of large numbers in Theorem 8.3 of Pollard (1990),

$$\Pr\left(\sup_{\omega}\left\|\hat{\Gamma}_{ni}(\omega,\varepsilon_{ni})-\Gamma_{ni}(\omega)\right\|>\eta\middle|X_n,p_n\right)\to 0,$$

as $n\to\infty$. Combining this with (8.14) shows that $\omega_{ni}(\varepsilon_{ni})$ is consistent for ω^*_{ni}

$$\omega_{ni}(\varepsilon_{ni})-\omega^*_{ni}=o_p(1).$$

For consistent $\omega_{ni}(\varepsilon_{ni})$, (8.13) implies that

$$\|\Gamma_{ni}(\omega_{ni}(\varepsilon_{ni}))-\Gamma_{ni}(\omega^*_{ni})\|\geq\|\omega_{ni}(\varepsilon_{ni})-\omega^*_{ni}\|(c+o_p(1)), \tag{8.16}$$

by the continuous mapping theorem.

Next we derive the asymptotic linear representation of $\omega_{ni}(\varepsilon_{ni})$ in (8.4). Consider the last two terms in (8.12). By the Lindeberg-Feller Central Limit Theorem the second last term in (8.12) is

$$\sqrt{n-1}\hat{\Gamma}_{ni}(\omega^*_{ni},\varepsilon_{ni})=O_p(1). \tag{8.17}$$

Define the stochastic process

$$\mathbb{G}_n(\omega,\varepsilon_{ni})=\sqrt{n-1}\left(\hat{\Gamma}_{ni}(\omega,\varepsilon_{ni})-\Gamma_{ni}(\omega)\right) \tag{8.18}$$

indexed by ω, so the the last term in (8.12) is given by

$$\begin{aligned}&\mathbb{G}_n(\omega_{ni}(\varepsilon_{ni}),\varepsilon_{ni})-\mathbb{G}_n(\omega^*_{ni},\varepsilon_{ni})\\&=\frac{1}{\sqrt{n-1}}\sum_{j\neq i}\phi_{n,ij}(\omega_{ni}(\varepsilon_{ni}),\varepsilon_{n,ij})-\phi_{n,ij}(\omega^*_{ni},\varepsilon_{n,ij})\\&\qquad-\left(\mathbb{E}\left[\phi_{n,ij}(\omega_{ni}(\varepsilon_{ni}),\varepsilon_{n,ij})-\phi_{n,ij}(\omega^*_{ni},\varepsilon_{n,ij})\right|X_n,p_n]\right),\end{aligned}$$

with $\phi_{n,ij}(\omega,\varepsilon_{n,ij})$ defined by

$$\phi_{n,ij}(\omega,\varepsilon_{n,ij})=1\left\{U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega-\varepsilon_{n,ij}>0\right\}D_j.$$

Note that $\mathbb{G}_n(\omega,\varepsilon_{ni})-\mathbb{G}_n(\omega^*_{ni},\varepsilon_{ni})$ is an empirical process defined over the triangular array $(\phi_{n,ij}(\omega,\varepsilon_{n,ij})-\phi_{n,ij}(\omega^*_{ni},\varepsilon_{n,ij}))_{j\neq i}$, $n=2,3,\ldots$. We want to show that

$$\mathbb{G}_n(\omega_n(\varepsilon_{ni}),\varepsilon_{ni})-\mathbb{G}_n(\omega^*_{ni},\varepsilon_{ni})=O_p\left(\|\omega_n(\varepsilon_{ni})-\omega^*_{ni}\|^{1/2}\right), \tag{8.19}$$

so the last term in (8.12) is negligible. The asymptotic linear representation in (8.4) can then be obtained from (8.12) and (8.16).

To show (8.19), for $M > 0$,

$$\begin{aligned}
&\Pr\left(\left\|\frac{\mathbb{G}_n\left(\omega_n\left(\varepsilon_{ni}\right),\varepsilon_{ni}\right)-\mathbb{G}_n\left(\omega_{ni}^*,\varepsilon_{ni}\right)}{\left\|\omega_n\left(\varepsilon_{ni}\right)-\omega_{ni}^*\right\|^{1/2}}\right\|>M\,\middle|\,X_n,p_n\right)\\
&\le \Pr\left(\sup_{\left\|\omega-\omega_{ni}^*\right\|>0}\left\|\frac{\mathbb{G}_n\left(\omega,\varepsilon_{ni}\right)-\mathbb{G}_n\left(\omega_{ni}^*,\varepsilon_{ni}\right)}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}\right\|>M\,\middle|\,X_n,p_n\right)\\
&\le \frac{1}{M}\mathbb{E}\left[\sup_{\left\|\omega-\omega_{ni}^*\right\|>0}\left\|\frac{\mathbb{G}_n\left(\omega,\varepsilon_{ni}\right)-\mathbb{G}_n\left(\omega_{ni}^*,\varepsilon_{ni}\right)}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}\right\|\,\middle|\,X_n,p_n\right].
\end{aligned}$$

by Markov's inequality. It suffices to show

$$\mathbb{E}\left[\sup_{\left\|\omega-\omega_{ni}^*\right\|>0}\left\|\frac{\mathbb{G}_n\left(\omega,\varepsilon_{ni}\right)-\mathbb{G}_n\left(\omega_{ni}^*,\varepsilon_{ni}\right)}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}\right\|\,\middle|\,X_n,p_n\right]<\infty. \tag{8.20}$$

Consider the process

$$\begin{aligned}
&\frac{\mathbb{G}_n\left(\omega,\varepsilon_{ni}\right)-\mathbb{G}_n\left(\omega_{ni}^*,\varepsilon_{ni}\right)}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}\\
&=\frac{1}{\sqrt{n-1}}\sum_{j\neq i}\frac{\phi_{n,ij}\left(\omega,\varepsilon_{n,ij}\right)-\phi_{n,ij}\left(\omega_{ni}^*,\varepsilon_{n,ij}\right)}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}-\frac{\mathbb{E}\left[\phi_{n,ij}\left(\omega,\varepsilon_{n,ij}\right)-\phi_{n,ij}\left(\omega_{ni}^*,\varepsilon_{n,ij}\right)\middle|\,X_n,p_n\right]}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}
\end{aligned}$$

The function $\frac{\phi_{n,ij}(\omega,\varepsilon_{n,ij})-\phi_{n,ij}\left(\omega_{ni}^*,\varepsilon_{n,ij}\right)}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}$ can be bounded by

$$\begin{aligned}
&\frac{\left\|\phi_{n,ij}\left(\omega,\varepsilon_{n,ij}\right)-\phi_{n,ij}\left(\omega_{ni}^*,\varepsilon_{n,ij}\right)\right\|}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}\\
&=\frac{\left|1\left\{U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega>\varepsilon_{n,ij}\right\}-1\left\{U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega_{ni}^*>\varepsilon_{n,ij}\right\}\right|}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}}\\
&\le \eta_{n,ij}\left(\omega,\omega_{ni}^*,\varepsilon_{n,ij}\right),
\end{aligned}$$

where the bound function $\eta_{n,ij}\left(\omega,\omega_{ni}^*,\varepsilon_{n,ij}\right)$ is

$$\eta_{n,ij}\left(\omega,\omega_{ni}^*,\varepsilon_{n,ij}\right)=\begin{cases}\frac{1}{\left\|\omega-\omega_{ni}^*\right\|^{1/2}} & \text{if } \varepsilon_{n,ij} \text{ lies between } U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega \text{ and } U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega_{ni}^*\\ 0 & \text{otherwise,}\end{cases} \tag{8.21}$$

Hence, we can combine the proof for Theorem 2.14.1 of Van der Vaart and Wellner (1996) (setting $p = 1$ and taking expectation conditional on X_n and p_n)[5] with the random entropy in the definition of manageability (Definition 7.9) in Pollard (1990) for triangular arrays, and derive the maximal inequality

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\|\omega-\omega_{ni}^*\|>0}\left\|\frac{\mathbb{G}_n(\omega,\varepsilon_{ni})-\mathbb{G}_n(\omega_{ni}^*,\varepsilon_{ni})}{\|\omega-\omega_{ni}^*\|^{1/2}}\right\|\Bigg| X_n, p_n\right] \\
&\leq K\mathbb{E}\left[J\left(1,\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni})\right)\sup_{\|\omega-\omega_{ni}^*\|>0}\|\eta_{ni}(\omega,\omega_{ni}^*,\varepsilon_{ni})\|_n \Bigg| X_n, p_n\right], \qquad (8.22)
\end{aligned}$$

where $K > 0$ is a universal constant, $\|\eta_{ni}(\omega,\omega_{ni}^*,\varepsilon_{ni})\|_n$ is the $L_2(\mathbb{P}_n)$ norm of the bound array $\eta_{ni}(\omega,\omega_{ni}^*,\varepsilon_{ni}) = (\eta_{n,ij}(\omega,\omega_{ni}^*,\varepsilon_{n,ij}))_{j\neq i}$

$$\|\eta_{ni}(\omega,\omega_{ni}^*,\varepsilon_{ni})\|_n = \left(\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2(\omega,\omega_{ni}^*,\varepsilon_{n,ij})\right)^{1/2}, \qquad (8.23)$$

$\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni})$ denotes the set of arrays

$$\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni}) = \left\{\left(\frac{\phi_{n,ij}(\omega,\varepsilon_{n,ij})-\phi_{n,ij}(\omega_{ni}^*,\varepsilon_{n,ij})}{\|\omega-\omega_{ni}^*\|^{1/2}}\right)_{j\neq i} : \omega\in\mathbb{R}^T, \|\omega-\omega_{ni}^*\|>0\right\}, \qquad (8.24)$$

and $J(1,\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni}))$ represents the uniform entropy integral of $\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni})$

$$J(1,\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni})) = \int_0^1 \sup_\alpha \sqrt{\log D\left(\xi\|\alpha\odot\bar{\eta}_{ni}(\omega_{ni}^*,\varepsilon_{ni})\|_n, \alpha\odot\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni}), \|\cdot\|_n\right)}d\xi. \qquad (8.25)$$

In the expression in (8.25), $\alpha\in\mathbb{R}^{n-1}$ is a $(n-1)\times 1$ vector of nonnegative constants, $\bar{\eta}_{ni}(\omega_{ni}^*,\varepsilon_{ni}) = \sup_{\|\omega-\omega_{ni}^*\|>0}\eta_{ni}(\omega,\omega_{ni}^*,\varepsilon_{ni})$, $\alpha\odot\bar{\eta}_{ni}(\omega_{ni}^*,\varepsilon_{ni})$ is the pointwise product of α and $\bar{\eta}_{ni}(\omega_{ni}^*,\varepsilon_{ni})$, $\alpha\odot\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni})$ is the set $\{\alpha\odot f : f\in\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni})\}$, $\|\cdot\|_n$ is the $L_2(\mathbb{P}_n)$ norm defined in (8.23), and $D(\xi\|\alpha\odot\bar{\eta}_{ni}(\omega_{ni}^*,\varepsilon_{ni})\|_n, \alpha\odot\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni}), \|\cdot\|_n)$ is the packing number of the set $\alpha\odot\mathcal{F}_n(\omega_{ni}^*,\varepsilon_{ni})$ at distance $\xi\|\alpha\odot\bar{\eta}_{ni}(\omega_{ni}^*,\varepsilon_{ni})\|_n$ under the norm $\|\cdot\|_n$. The supremum in (8.25) is taken over all vectors $\alpha\in\mathbb{R}^{n-1}$ with nonnegative constants. Note that the set $\left\{\|\omega-\omega_{ni}^*\|^{-1/2} : \omega\in\mathbb{R}^T, \|\omega-\omega_{ni}^*\|>0\right\}$ has a pseudodimension of at most 1

[5] In the proof for Theorem 2.14.1 of 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 stochastic process in (8.18) and the envelope functions in (8.21).

because $\|\omega-\omega_{ni}^{*}\|^{-1/2}\in\mathbb{R}$. Hence, by the manageability of the set in (8.15), Corollary 4.10, Definition 7.9, and the stability results of packing numbers under addition and multiplication in Section 5 of Pollard (1990) we show that the set $\mathcal{F}_n(\omega_{ni}^{*},\varepsilon_{ni})$ is manageable and thus

$$J\left(1,\mathcal{F}_n\left(\omega_{ni}^{*},\varepsilon_{ni}\right)\right)<\infty \tag{8.26}$$

uniformly in ε_{ni} and n. Next we consider the conditional expectation of $\sup_{\|\omega-\omega_{ni}^{*}\|>0}\|\eta_{ni}(\omega,\omega_{ni}^{*},\varepsilon_{ni})\|_n$. By Jensen's inequality

$$\begin{aligned}
&\mathbb{E}\left[\left.\sup_{\|\omega-\omega_{ni}^{*}\|>0}\|\eta_{ni}(\omega,\omega_{ni}^{*},\varepsilon_{ni})\|_n\right|X_n,p_n\right]\\
&=\mathbb{E}\left[\left.\left(\sup_{\|\omega-\omega_{ni}^{*}\|>0}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^{2}(\omega,\omega_{ni}^{*},\varepsilon_{n,ij})\right)^{1/2}\right|X_n,p_n\right]\\
&\leq\left(\mathbb{E}\left[\left.\sup_{\|\omega-\omega_{ni}^{*}\|>0}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^{2}(\omega,\omega_{ni}^{*},\varepsilon_{n,ij})\right|X_n,p_n\right]\right)^{1/2}
\end{aligned}$$

Following the argument for (8.15) it is similar to show that the set of arrays

$$\left\{\left(1\left\{\varepsilon_{n,ij}\text{ lies between }U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega\text{ and }U_{n,ij}+\frac{n-1}{n-2}2D_j'\omega_{ni}^{*}\right\}\right)_{j\neq i}:\omega\in\mathbb{R}^T\right\}$$

is manageable and so is the set of arrays

$$\left\{\left(\eta_{n,ij}^{2}(\omega,\omega_{ni}^{*},\varepsilon_{n,ij})\right)_{j\neq i}:\omega\in\mathbb{R}^T,\|\omega-\omega_{ni}^{*}\|>0\right\}.$$

Thus by the uniform law of large numbers in Theorem 8.3 of Pollard (1990)

$$\begin{aligned}
&\left|\sup_{\|\omega-\omega_{ni}^{*}\|>0}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^{2}(\omega,\omega_{ni}^{*},\varepsilon_{n,ij})-\sup_{\|\omega-\omega_{ni}^{*}\|>0}\frac{1}{n-1}\sum_{j\neq i}\mathbb{E}\left[\left.\eta_{n,ij}^{2}(\omega,\omega_{ni}^{*},\varepsilon_{n,ij})\right|X_n,p_n\right]\right|\\
&\leq\sup_{\|\omega-\omega_{ni}^{*}\|>0}\left|\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^{2}(\omega,\omega_{ni}^{*},\varepsilon_{n,ij})-\mathbb{E}\left[\left.\eta_{n,ij}^{2}(\omega,\omega_{ni}^{*},\varepsilon_{n,ij})\right|X_n,p_n\right]\right|\to 0,\text{ a.s.}
\end{aligned}$$

as $n \to \infty$. Moreover, by the definition of $\eta_{n,ij}(\omega, \omega_{ni}^*, \varepsilon_{n,ij})$ and mean-value theorem

$$\begin{aligned}
&\mathbb{E}\left[\eta_{n,ij}^2(\omega, \omega_{ni}^*, \varepsilon_{n,ij}) \middle| X_n, p_n\right] \\
=&\frac{\left|F_\varepsilon\left(U_{n,ij} + \frac{n-1}{n-2}2D_j'\omega\right) - F_\varepsilon\left(U_{n,ij} + \frac{n-1}{n-2}2D_j'\omega_{ni}^*\right)\right|}{\|\omega - \omega_{ni}^*\|} \\
=&f_\varepsilon\left(U_{n,ij} + \frac{n-1}{n-2}2D_j'\left(t_{n,ij}\omega + (1-t_{n,ij})\omega_{ni}^*\right)\right)\frac{2(n-1)}{n-2}
\end{aligned} \tag{8.27}$$

for some $t_{n,ij} \in [0,1]$. Hence,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\|\omega-\omega_{ni}^*\|>0} \|\eta_{ni}(\omega, \omega_{ni}^*, \varepsilon_{ni})\|_n \middle| X_n, p_n\right] \\
\leq &\left(\sup_{\|\omega-\omega_{ni}^*\|>0} \frac{1}{n-1}\sum_{j\neq i}\mathbb{E}\left[\eta_{n,ij}^2(\omega, \omega_{ni}^*, \varepsilon_{n,ij}) \middle| X_n, p_n\right]\right)^{1/2} + o(1) \\
= &\left(\sup_{\|\omega-\omega_{ni}^*\|>0} \frac{1}{n-1}\sum_{j\neq i} f_\varepsilon\left(U_{n,ij} + \frac{n-1}{n-2}2D_j'\left(t_{n,ij}\omega + (1-t_{n,ij})\omega_{ni}^*\right)\right)\frac{2(n-1)}{n-2}\right)^{1/2} + o(1) \\
< &\infty
\end{aligned} \tag{8.28}$$

by the dominated convergence theorem. Combining (8.22), (8.26), (8.28) proves (8.20) and thus (8.19). Therefore, the last term in (8.12) is

$$\sqrt{n-1}\left(\hat{\Gamma}_{ni}(\omega_{ni}(\varepsilon_{ni}), \varepsilon_{ni}) - \Gamma_{ni}(\omega_{ni}(\varepsilon_{ni})) - (\hat{\Gamma}_{ni}(\omega_{ni}^*, \varepsilon_{ni}) - \Gamma_{ni}(\omega_{ni}^*))\right) = O_p\left(\|\omega_{ni}(\varepsilon_{ni}) - \omega_{ni}^*\|^{1/2}\right) \tag{8.29}$$

Applying (8.16), (8.17) and (8.29) to (8.12) we obtain

$$\sqrt{n-1}\,\|\omega_{ni}(\varepsilon_{ni}) - \omega_{ni}^*\|\,(c + o_p(1)) \leq O_p(1) + O_p\left(\|\omega_{ni}(\varepsilon_{ni}) - \omega_{ni}^*\|^{1/2}\right)$$

By consistency $\omega_{ni}(\varepsilon_{ni}) - \omega_{ni}^* = o_p(1)$, the second term on the right-hand side is negligible compared with the first term, so the right-hand side is $O_p(1)$. This implies that

$$\omega_{ni}(\varepsilon_{ni}) - \omega_{ni}^* = O_p\left(\frac{1}{\sqrt{n}}\right)$$

For $\omega_{ni}(\varepsilon_{ni})$ converging at this rate, (8.29) becomes

$$\sqrt{n-1}\left(\hat{\Gamma}_{ni}(\omega_{ni}(\varepsilon_{ni}), \varepsilon_{ni}) - \Gamma_{ni}(\omega_{ni}(\varepsilon_{ni})) - (\hat{\Gamma}_{ni}(\omega_{ni}^*, \varepsilon_{ni}) - \Gamma_{ni}(\omega_{ni}^*))\right) = O_p\left(\frac{1}{n^{1/4}}\right)$$

Applying this and the differentiability of Γ_{ni} to (8.12) again yields

$$H_{ni}\left(\omega_{ni}^{*}\right)\left(\omega_{ni}\left(\varepsilon_{ni}\right)-\omega_{ni}^{*}\right)+o_{p}\left(\left\|\omega_{ni}\left(\varepsilon_{ni}\right)-\omega_{ni}^{*}\right\|\right)=-\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}\right)+O_{p}\left(\frac{1}{n^{3/4}}\right)$$

By the invertibility of $H_{ni}\left(\omega_{ni}^{*}\right)$, we have

$$\omega_{ni}\left(\varepsilon_{ni}\right)-\omega_{ni}^{*}=-H_{ni}^{-1}\left(\omega_{ni}^{*}\right)\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}\right)+o_{p}\left(\frac{1}{\sqrt{n}}\right)$$

The proof is complete. ■

Proof of Lemma 8.2. First we prove (8.7). From the definition of ω_{ni}^{*}, $\omega_{ni}\left(\varepsilon_{ni}^{1}\right)$, and $\omega_{ni}\left(\varepsilon_{ni}^{2}\right)$, we have $\Gamma_{ni}\left(\omega_{ni}^{*}\right)=0$ and $\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right),\varepsilon_{ni}^{1}\right)=\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{2}\right)=0$ with probability 1. Hence, with probability 1 we can decompose $\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right)\right)-\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right)\right)$ as

$$\begin{aligned}
&\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right)\right)-\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right)\right)\\
=&\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right)\right)-\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right),\varepsilon_{ni}^{1}\right)-\left(\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right)\right)-\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{2}\right)\right)\\
=&\left(\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}^{2}\right)-\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}^{1}\right)\right)-\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_{n}\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right),\varepsilon_{ni}^{1}\right)-\mathbb{G}_{n}\left(\omega_{ni}^{*},\varepsilon_{ni}^{1}\right)\right)\\
&+\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_{n}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{2}\right)-\mathbb{G}_{n}\left(\omega_{ni}^{*},\varepsilon_{ni}^{2}\right)\right)\\
=&\left(\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}^{2}\right)-\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}^{1}\right)\right)-\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_{n}\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right),\varepsilon_{ni}^{1}\right)-\mathbb{G}_{n}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{1}\right)\right)\\
&+\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_{n}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{2}\right)-\mathbb{G}_{n}\left(\omega_{ni}^{*},\varepsilon_{ni}^{2}\right)-\left(\mathbb{G}_{n}\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{1}\right)-\mathbb{G}_{n}\left(\omega_{ni}^{*},\varepsilon_{ni}^{1}\right)\right)\right)
\end{aligned} \tag{8.30}$$

where $\hat{\Gamma}_{ni}$, Γ_{ni}, and $\mathbb{G}_{n}$ are defined in (8.10), (8.11) and (8.18). By the definition of $\hat{\Gamma}_{ni}$ and $\mathbb{G}_{n}$, the first and last terms on the right-hand side of (8.30) are

$$\begin{aligned}
&\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}^{2}\right)-\hat{\Gamma}_{ni}\left(\omega_{ni}^{*},\varepsilon_{ni}^{1}\right)\\
=&\frac{1}{n-1}\left(\psi_{n,ij}\left(\omega_{ni}^{*},\varepsilon_{n,ij}^{2}\right)-\psi_{n,ij}\left(\omega_{ni}^{*},\varepsilon_{n,ij}^{1}\right)+\psi_{n,ik}\left(\omega_{ni}^{*},\varepsilon_{n,ik}^{2}\right)-\psi_{n,ik}\left(\omega_{ni}^{*},\varepsilon_{n,ik}^{1}\right)\right)\\
=&O_{p}\left(\frac{1}{n}\right)
\end{aligned} \tag{8.31}$$

and

$$
\begin{aligned}
&\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^2\right)-\mathbb{G}_n\left(\omega_{ni}^*,\varepsilon_{ni}^2\right)-\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^1\right)-\mathbb{G}_n\left(\omega_{ni}^*,\varepsilon_{ni}^1\right)\right)\right)\\
=&\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^2\right)-\hat{\Gamma}_{ni}\left(\omega_{ni}^*,\varepsilon_{ni}^2\right)-\left(\hat{\Gamma}_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^1\right)-\hat{\Gamma}_{ni}\left(\omega_{ni}^*,\varepsilon_{ni}^1\right)\right)\\
=&\frac{1}{n-1}\left(\psi_{n,ij}(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,ij}^2)-\psi_{n,ij}(\omega_{ni}^*,\varepsilon_{n,ij}^2)-\left(\psi_{n,ij}(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,ij}^1)-\psi_{n,ij}(\omega_{ni}^*,\varepsilon_{n.ij}^1)\right)\right.\\
&\quad\left.+\,\psi_{n,ik}\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,ik}^2\right)-\psi_{n,ik}\left(\omega_{ni}^*,\varepsilon_{n,ik}^2\right)-\left(\psi_{n,ik}\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,ik}^1\right)-\psi_{n,ik}\left(\omega_{ni}^*,\varepsilon_{n,ik}^1\right)\right)\right)\\
=&O_p\left(\frac{1}{n}\right).
\end{aligned}
\tag{8.32}
$$

Next, we want to show

$$
\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\varepsilon_{ni}^1\right)-\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^1\right)=O_p\left(\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right)-\omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2}\right) \tag{8.33}
$$

so the rate of the second last term on the right-hand side of (8.30) can be controlled by $n^{-1/2}\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right)-\omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2}$. Similar to the argument for (8.20) in Lemma 8.1, it suffices to show

$$
\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\frac{\mathbb{G}_n\left(\omega,\varepsilon_{ni}\right)-\mathbb{G}_n\left(\tilde{\omega},\varepsilon_{ni}\right)}{\|\omega-\tilde{\omega}\|^{1/2}}\right\|\middle| X_n,p_n\right]<\infty. \tag{8.34}
$$

Note that (8.34) is slightly different from the result in (8.20) in Lemma 8.1 because (8.20) holds for a fixed ω_{ni}^*, while here we need the supremum to be over both ω and $\tilde{\omega}$. Modifying the proof for Theorem 2.14.1 in Van der Vaart and Wellner (1996) and the argument for (8.22) in Lemma 8.1, we can derive a version of (8.22) that is uniform over both ω and $\tilde{\omega}$,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\frac{\mathbb{G}_n\left(\omega,\varepsilon_{ni}\right)-\mathbb{G}_n\left(\tilde{\omega},\varepsilon_{ni}\right)}{\|\omega-\tilde{\omega}\|^{1/2}}\right\|\middle| X_n,p_n\right]\\
&\leq K\mathbb{E}\left[J\left(1,\mathcal{F}_n\left(\varepsilon_{ni}\right)\right)\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\eta_{ni}\left(\omega,\tilde{\omega},\varepsilon_{ni}\right)\right\|_n\middle| X_n,p_n\right],
\end{aligned}
\tag{8.35}
$$

where $K>0$ is a universal constant, $\eta_{ni}\left(\omega,\tilde{\omega},\varepsilon_{ni}\right)$ is the bound array defined in (8.21) with ω_{ni}^* replaced by $\tilde{\omega}$, $\mathcal{F}_n\left(\varepsilon_{ni}\right)$ is the set of arrays

$$
\mathcal{F}_n\left(\varepsilon_{ni}\right)=\left\{\left(\frac{\phi_{n,ij}\left(\omega,\varepsilon_{n,ij}\right)-\phi_{n,ij}\left(\tilde{\omega},\varepsilon_{n,ij}\right)}{\|\omega-\tilde{\omega}\|^{1/2}}\right)_{j\neq i}:\omega,\tilde{\omega}\in\mathbb{R}^T,\|\omega-\tilde{\omega}\|>0\right\}, \tag{8.36}
$$

and $J\left(1,\mathcal{F}_n\left(\varepsilon_{ni}\right)\right)$ is the uniform entropy integral as defined in (8.25) for the set $\mathcal{F}_n\left(\varepsilon_{ni}\right)$

$$J\left(1,\mathcal{F}_n\left(\varepsilon_{ni}\right)\right)=\int_0^1 \sup_{\alpha}\sqrt{\log D\left(\xi\left\|\alpha\odot\sup_{\|\omega-\tilde{\omega}\|>0}\eta_{ni}\left(\omega,\tilde{\omega},\varepsilon_{ni}\right)\right\|_n,\alpha\odot\mathcal{F}_n\left(\varepsilon_{ni}\right),\|\cdot\|_n\right)}d\xi. \tag{8.37}$$

where all the ingredients are the same as in (8.25) except that $\bar{\eta}_{ni}\left(\omega_{ni}^*,\varepsilon_{ni}\right)$ is replaced with $\sup_{\|\omega-\tilde{\omega}\|>0}\eta_{ni}\left(\omega,\tilde{\omega},\varepsilon_{ni}\right)$. We have shown in Lemma 8.1 that the set of arrays in (8.15) is manageable, so by the stability of packing numbers under addition and multiplication in Section 5 of Pollard (1990) the set $\mathcal{F}_n\left(\varepsilon_{ni}\right)$ is also manageable. The definition of manageability then yields

$$J\left(1,\mathcal{F}_n\left(\varepsilon_{ni}\right)\right)<\infty \tag{8.38}$$

uniformly in ε_{ni} and n. Furthermore, by Jensen's inequality

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\eta_{ni}\left(\omega,\tilde{\omega},\varepsilon_{ni}\right)\right\|_n\middle| X_n,p_n\right]\\
&=\mathbb{E}\left[\left(\sup_{\|\omega-\tilde{\omega}\|>0}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega,\tilde{\omega},\varepsilon_{n,ij}\right)\right)^{1/2}\middle| X_n,p_n\right]\\
&\leq\left(\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega,\tilde{\omega},\varepsilon_{n,ij}\right)\middle| X_n,p_n\right]\right)^{1/2}
\end{aligned}$$

Following the argument in Lemma 8.1, it is similar to show that the set of array

$$\left\{\left(\eta_{n,ij}^2\left(\omega,\tilde{\omega},\varepsilon_{n,ij}\right)\right)_{j\neq i}:\omega,\tilde{\omega}\in\mathbb{R}^T,\|\omega-\tilde{\omega}\|>0\right\}$$

is manageable, so by the uniform law of large numbers in Theorem 8.3 of Pollard (1990)

$$\begin{aligned}
&\left|\sup_{\|\omega-\tilde{\omega}\|>0}\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega,\tilde{\omega},\varepsilon_{n,ij}\right)-\sup_{\|\omega-\tilde{\omega}\|>0}\frac{1}{n-1}\sum_{j\neq i}\mathbb{E}\left[\eta_{n,ij}^2\left(\omega,\tilde{\omega},\varepsilon_{n,ij}\right)\middle| X_n,p_n\right]\right|\\
&\leq\sup_{\|\omega-\tilde{\omega}\|>0}\left|\frac{1}{n-1}\sum_{j\neq i}\eta_{n,ij}^2\left(\omega,\tilde{\omega},\varepsilon_{n,ij}\right)-\mathbb{E}\left[\eta_{n,ij}^2\left(\omega,\tilde{\omega},\varepsilon_{n,ij}\right)\middle| X_n,p_n\right]\right|\to 0,\text{ a.s.}
\end{aligned}$$

as $n \to \infty$. Hence, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0} \|\eta_{ni}(\omega,\tilde{\omega},\varepsilon_{ni})\|_n \middle| X_n, p_n\right] \\
&\leq \left(\sup_{\|\omega-\tilde{\omega}\|>0} \frac{1}{n-1}\sum_{j\neq i} \mathbb{E}\left[\eta_{n,ij}^2(\omega,\tilde{\omega},\varepsilon_{n,ij})\middle| X_n, p_n\right]\right)^{1/2} + o(1) \\
&= \left(\sup_{\|\omega-\tilde{\omega}\|>0} \frac{1}{n-1}\sum_{j\neq i} f_\varepsilon\left(U_{n,ij} + \frac{n-1}{n-2}2D_j'(t_{n,ij}\omega + (1-t_{n,ij})\tilde{\omega})\right)\frac{2(n-1)}{n-2}\right)^{1/2} + o(1) \\
&< \infty \qquad (8.39)
\end{aligned}$$

by the dominated convergence theorem and (8.27). Combining (8.35), (8.38), (8.39) proves (8.34) and (8.33). Thus the second last term on the right-hand side of (8.30) satisfies

$$\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\varepsilon_{ni}^1\right) - \mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^1\right)\right) = O_p\left(\frac{1}{\sqrt{n}}\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2}\right) \qquad (8.40)$$

Now we go back to the left-hand side of (8.30). By the differentiability of Γ_{ni} and $\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2) = o_p(1)$ derived from Lemma 8.1, the left-hand side of (8.30) is

$$\begin{aligned}
&\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^1\right)\right) - \Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^2\right)\right) \\
&= H_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^2\right)\right)\left(\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right) + o_p\left(\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|\right) \\
&= H_{ni}\left(\omega_{ni}^*\right)\left(\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right) + o_p\left(\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|\right) \qquad (8.41)
\end{aligned}$$

where the last equality comes from the fact that $H_{ni}(\omega_{ni}(\varepsilon_{ni}^2)) - H_{ni}(\omega_{ni}^*) = o_p(1)$ as a result of the continuity of H_{ni} and the consistency of $\omega_{ni}(\varepsilon_{ni}^2)$. Since $H_{ni}(\omega_{ni}^*)$ is invertible, there is a constant $c > 0$ such that $\|H_{ni}(\omega_{ni}^*)(\omega^1 - \omega^2)\| \geq c\|\omega^1 - \omega^2\|$ for every ω^1 and ω^2. Applying this to (8.41) yields

$$\left\|\Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^1\right)\right) - \Gamma_{ni}\left(\omega_{ni}\left(\varepsilon_{ni}^2\right)\right)\right\| \geq \left\|\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|(c + o_p(1)) \qquad (8.42)$$

Combining (8.30), (8.31), (8.32), (8.40), and (8.42) we obtain

$$\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|(c + o_p(1)) \leq O_p\left(\frac{1}{n}\right) + O_p\left(\frac{1}{\sqrt{n}}\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2}\right). \qquad (8.43)$$

From Lemma 8.1, $\|\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2)\| \leq \|\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}^*\| + \|\omega_{ni}(\varepsilon_{ni}^2) - \omega_{ni}^*\|$ is at most $O_p\left(n^{-1/2}\right)$. If $\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2)$ converges slower than $O_p(n^{-1})$, the second term on the right-hand side of (8.43) would dominate the first term, implying $\|\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2)\| \leq O_p\left(n^{-1/2}\|\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2)\|^{1/2}\right)$. This can be true only if $\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2)$ converges faster than $O_p(n^{-1})$, leading to a contradiction. Therefore, we must have

$$\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right) = O_p\left(\frac{1}{n}\right).$$

For $\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2)$ converging at this rate, (8.40) becomes

$$\frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right), \varepsilon_{ni}^1\right) - \mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^1\right)\right) = O_p\left(\frac{1}{n}\right) \tag{8.44}$$

Combining (8.41) with (8.30) yields

$$\begin{aligned}
&H_{ni}\left(\omega_{ni}^*\right)\left(\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right)\right) \\
=&\hat{\Gamma}_{ni}\left(\omega_{ni}^*, \varepsilon_{ni}^2\right) - \hat{\Gamma}_{ni}\left(\omega_{ni}^*, \varepsilon_{ni}^1\right) \\
&- \frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right), \varepsilon_{ni}^1\right) - \mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^1\right)\right) \\
&+ \frac{1}{\sqrt{n-1}}\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^2\right) - \mathbb{G}_n\left(\omega_{ni}^*, \varepsilon_{ni}^2\right) - \left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^1\right) - \mathbb{G}_n\left(\omega_{ni}^*, \varepsilon_{ni}^1\right)\right)\right) + o_p\left(\frac{1}{n}\right)
\end{aligned}$$

By the invertiability of $H_{ni}(\omega_{ni}^*)$ and

$$\begin{aligned}
&\hat{\Gamma}_{ni}\left(\omega_{ni}^*, \varepsilon_{ni}^2\right) - \hat{\Gamma}_{ni}\left(\omega_{ni}^*, \varepsilon_{ni}^1\right) \\
=&\frac{1}{n-1}\left(\psi_{n,ij}\left(\omega_{ni}^*, \varepsilon_{n,ij}^2\right) - \psi_{n,ij}\left(\omega_{ni}^*, \varepsilon_{n,ij}^1\right) + \psi_{n,ik}\left(\omega_{ni}^*, \varepsilon_{n,ik}^2\right) - \psi_{n,ik}\left(\omega_{ni}^*, \varepsilon_{n,ik}^1\right)\right)
\end{aligned}$$

we can derive $\omega_{ni}(\varepsilon_{ni}^1) - \omega_{ni}(\varepsilon_{ni}^2)$ as

$$\begin{aligned}
&\omega_{ni}\left(\varepsilon_{ni}^1\right) - \omega_{ni}\left(\varepsilon_{ni}^2\right) \\
=&-\frac{1}{n-1}H_{ni}^{-1}\left(\omega_{ni}^*\right)\left(\psi_{n,ij}\left(\omega_{ni}^*, \varepsilon_{n,ij}^1\right) - \psi_{n,ij}\left(\omega_{ni}^*, \varepsilon_{n,ij}^2\right) + \psi_{n,ik}\left(\omega_{ni}^*, \varepsilon_{n,ik}^1\right) - \psi_{n,ik}\left(\omega_{ni}^*, \varepsilon_{n,ik}^2\right)\right) \\
&+ r_n\left(\varepsilon_{ni}^1\right) - r_n\left(\varepsilon_{ni}^2\right)
\end{aligned}$$

where $r_n(\varepsilon_{ni}^1) - r_n(\varepsilon_{ni}^2)$ should be equal to

$$
\begin{aligned}
& r_n\left(\varepsilon_{ni}^1\right) - r_n\left(\varepsilon_{ni}^2\right) \\
= & \frac{1}{\sqrt{n-1}} H_{ni}^{-1}\left(\omega_{ni}^*\right)\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^2\right) - \mathbb{G}_n\left(\omega_{ni}^*, \varepsilon_{ni}^2\right) - \left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^1\right) - \mathbb{G}_n\left(\omega_{ni}^*, \varepsilon_{ni}^1\right)\right)\right) \\
& - \frac{1}{\sqrt{n-1}} H_{ni}^{-1}\left(\omega_{ni}^*\right)\left(\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right), \varepsilon_{ni}^1\right) - \mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^1\right)\right) + o_p\left(\frac{1}{n}\right) \qquad (8.45)
\end{aligned}
$$

The rates in (8.32) and (8.44) imply that

$$
r_n\left(\varepsilon_{ni}^1\right) - r_n\left(\varepsilon_{ni}^2\right) = O_p\left(\frac{1}{n}\right).
$$

This proves (8.7).

We now turn to the proof for (8.8). Consider $r_n(\varepsilon_{ni}^1) - r_n(\varepsilon_{ni}^2)$ given in (8.45). In the last expression in (8.32), $\psi_{n,ij}$ and $\psi_{n,ik}$ are bounded for all $\varepsilon_{n,ij}^1$ and $\varepsilon_{n,ik}^1$, so the supremium of the first term on the right-hand side of (8.45) over $\varepsilon_{n,ij}^1$ and $\varepsilon_{n,ik}^1$ has a conditional expectation that is $O(n^{-1})$. It suffices to consider the empirical process in the last line of (8.45). In particular, it suffices to show

$$
\mathbb{E}\left[\sup_{\varepsilon_{n,ij}^1, \varepsilon_{n,ik}^1} \left\| \mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right), \varepsilon_{ni}^1\right) - \mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right), \varepsilon_{ni}^1\right)\right\| \middle| X_n, p_n\right] \leq O\left(\frac{1}{\sqrt{n}}\right) \qquad (8.46)
$$

By the definition of $\mathbb{G}_n$ in (8.18), for any ω and $\tilde{\omega}$

$$
\mathbb{G}_n(\omega, \varepsilon_{ni}) - \mathbb{G}_n(\tilde{\omega}, \varepsilon_{ni}) = \frac{1}{\sqrt{n-1}} \sum_{l \neq i} \tilde{\phi}_{n,il}(\omega, \tilde{\omega}, \varepsilon_{n,il})
$$

where

$$
\tilde{\phi}_{n,il}(\omega, \tilde{\omega}, \varepsilon_{n,il}) = \phi_{n,il}(\omega, \varepsilon_{n,il}) - \phi_{n,il}(\tilde{\omega}, \varepsilon_{n,il}) - \left(\mathbb{E}\left[\phi_{n,il}(\omega, \varepsilon_{n,il}) - \phi_{n,il}(\tilde{\omega}, \varepsilon_{n,il})\right| X_n, p_n\right]\right)
$$

We separate the jth and kth components in $\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\varepsilon_{ni}^1\right)-\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^1\right)$ from the rest and obtain

$$\begin{aligned}
&\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\left\|\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\varepsilon_{ni}^1\right)-\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{ni}^1\right)\right\| \\
\leq&\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\left\|\frac{1}{\sqrt{n-1}}\sum_{l\neq i,j,k}\tilde{\phi}_{n,il}\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,il}^1\right)\right\| \\
&+\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\frac{1}{\sqrt{n-1}}\left\|\tilde{\phi}_{n,ij}\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,ij}^1\right)+\tilde{\phi}_{n,ik}\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,ik}^1\right)\right\|
\end{aligned} \tag{8.47}$$

Since $\tilde{\phi}_{n,ij}$ and $\tilde{\phi}_{n,ik}$ are bounded uniformly in $\varepsilon_{n,ij}^1$ and $\varepsilon_{n,ik}^1$, the last term in (8.47) has a conditional expectation that is at most $O\left(n^{-1/2}\right)$. As for the first term on the right-hand side of (8.47), by the property of supremium we have

$$\begin{aligned}
&\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\left\|\frac{1}{\sqrt{n-1}}\sum_{l\neq i,j,k}\tilde{\phi}_{n,il}\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,il}^1\right)\right\| \\
\leq&\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\left\|\frac{1}{\sqrt{n-1}}\sum_{l\neq i,j,k}\frac{\tilde{\phi}_{n,il}\left(\omega_{ni}\left(\varepsilon_{ni}^1\right),\omega_{ni}\left(\varepsilon_{ni}^2\right),\varepsilon_{n,il}^1\right)}{\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right)-\omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2}}\right\|\cdot\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right)-\omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2} \\
\leq&\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\frac{1}{\sqrt{n-1}}\sum_{l\neq i,j,k}\frac{\tilde{\phi}_{n,il}\left(\omega,\tilde{\omega},\varepsilon_{n,il}^1\right)}{\left\|\omega-\tilde{\omega}\right\|^{1/2}}\right\|\cdot\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right)-\omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2} \\
=&\frac{\sqrt{n-3}}{\sqrt{n-1}}\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\frac{\mathbb{G}_{n,-jk}\left(\omega,\varepsilon_{ni,-jk}^1\right)-\mathbb{G}_{n,-jk}\left(\tilde{\omega},\varepsilon_{ni,-jk}^1\right)}{\left\|\omega-\tilde{\omega}\right\|^{1/2}}\right\|\cdot\sup_{\varepsilon_{n,ij}^1,\varepsilon_{n,ik}^1}\left\|\omega_{ni}\left(\varepsilon_{ni}^1\right)-\omega_{ni}\left(\varepsilon_{ni}^2\right)\right\|^{1/2}
\end{aligned}$$

where in the second last sup term $\varepsilon_{ni,-kj}^1=\left(\varepsilon_{n,il}^1\right)_{l\neq i,j,k}$ and

$$\mathbb{G}_{n,-jk}\left(\omega,\varepsilon_{ni,-jk}^1\right)-\mathbb{G}_{n,-jk}\left(\tilde{\omega},\varepsilon_{ni,-jk}^1\right)=\frac{1}{\sqrt{n-3}}\sum_{l\neq i,j,k}\tilde{\phi}_{n,il}\left(\omega,\tilde{\omega},\varepsilon_{n,il}^1\right)$$

represents the empirical process obtained from $\mathbb{G}_n\left(\omega,\varepsilon_{ni}^1\right)-\mathbb{G}_n\left(\tilde{\omega},\varepsilon_{ni}^1\right)$ by deleting the jth and kth components and normalizing by $n-3$. Taking the conditional expectation then yields

$$\mathbb{E}\left[\sup_{\varepsilon^1_{n,ij},\varepsilon^1_{n,ik}}\left\|\frac{1}{\sqrt{n-1}}\sum_{l\neq i,j,k}\tilde{\phi}_{n,il}\left(\omega_{ni}\left(\varepsilon^1_{ni}\right),\omega_{ni}\left(\varepsilon^2_{ni}\right),\varepsilon^1_{n,il}\right)\right\|\middle| X_n,p_n\right]$$
$$\leq\frac{\sqrt{n-3}}{\sqrt{n-1}}\left(\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\frac{\mathbb{G}_{n,-jk}\left(\omega,\varepsilon^1_{ni,-jk}\right)-\mathbb{G}_{n,-jk}\left(\tilde{\omega},\varepsilon^1_{ni,-jk}\right)}{\|\omega-\tilde{\omega}\|^{1/2}}\right\|^2\middle| X_n,p_n\right]\right)^{1/2}$$
$$\cdot\left(\mathbb{E}\left[\sup_{\varepsilon^1_{n,ij},\varepsilon^1_{n,ik}}\left\|\omega_{ni}\left(\varepsilon^1_{ni}\right)-\omega_{ni}\left(\varepsilon^2_{ni}\right)\right\|\middle| X_n,p_n\right]\right)^{1/2} \tag{8.48}$$

by Cauchy-Schwarz inequality. Consider the second moment in the second line of (8.48). Similar to the argument for (8.34), we modify the proof for Theorem 2.14.1 in Van der Vaart and Wellner (1996) in line with (8.35) except that we apply the theorem with the L_2 norm (setting $p=2$) instead of the L_1 norm (setting $p=1$). We then derive a second-moment version of (8.35) for the empirical process in the second line of (8.48)

$$\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\frac{\mathbb{G}_{n,-jk}\left(\omega,\varepsilon_{ni,-jk}\right)-\mathbb{G}_{n,-jk}\left(\tilde{\omega},\varepsilon_{ni,-jk}\right)}{\|\omega-\tilde{\omega}\|^{1/2}}\right\|^2\middle| X_n,p_n\right]$$
$$\leq K\mathbb{E}\left[J\left(1,\mathcal{F}_{n,-jk}\left(\varepsilon_{ni,-jk}\right)\right)^2\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\eta_{ni,-jk}\left(\omega,\tilde{\omega},\varepsilon_{ni,-jk}\right)\right\|^2_{n-2}\middle| X_n,p_n\right], \tag{8.49}$$

where $K>0$ is a universal constant, $\left\|\eta_{ni,-jk}\left(\omega,\tilde{\omega},\varepsilon_{ni,-jk}\right)\right\|_{n-2}$ is the $L_2\left(\mathbb{P}_{n-2}\right)$ norm for the deleting-two bound array $\eta_{ni,-jk}\left(\omega,\tilde{\omega},\varepsilon_{ni,-jk}\right)=\left(\eta_{n,il}\left(\omega,\tilde{\omega},\varepsilon_{n,il}\right)\right)_{l\neq i,j,k}$

$$\left\|\eta_{ni,-jk}\left(\omega,\tilde{\omega},\varepsilon_{ni,-jk}\right)\right\|_{n-2}=\left(\frac{1}{n-3}\sum_{l\neq i,j,k}\eta^2_{n,il}\left(\omega,\tilde{\omega},\varepsilon_{n,il}\right)\right)^{1/2},$$

$\mathcal{F}_{n,-jk}\left(\varepsilon_{ni,-jk}\right)$ is the set of deleting-two arrays

$$\mathcal{F}_{n,-jk}\left(\varepsilon_{ni,-jk}\right)=\left\{\left(\frac{\phi_{n,il}\left(\omega,\varepsilon_{n,il}\right)-\phi_{n,il}\left(\tilde{\omega},\varepsilon_{n,il}\right)}{\|\omega-\tilde{\omega}\|^{1/2}}\right)_{l\neq i,j,k}:\omega,\tilde{\omega}\in\mathbb{R}^T,\|\omega-\tilde{\omega}\|>0\right\},$$

and $J\left(1,\mathcal{F}_{n,-jk}\left(\varepsilon_{ni,-jk}\right)\right)$ is the uniform entropy integral as in (8.37) for the set $\mathcal{F}_{n,-jk}\left(\varepsilon_{ni,-jk}\right)$. We have shown that the uniform entropy integral $J\left(1,\mathcal{F}_n\left(\varepsilon_{ni}\right)\right)$ is bounded uniformly in ε_{ni} and n. This implies trivially

$$J\left(1,\mathcal{F}_{n,-jk}\left(\varepsilon_{ni,-jk}\right)\right)<\infty \tag{8.50}$$

uniformly in $\varepsilon_{ni,-jk}$ and n. Moreover, following the proof for (8.39) we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\eta_{ni,-jk}\left(\omega,\tilde{\omega},\varepsilon_{ni,-jk}\right)\right\|_{n-2}^{2}\middle| X_n,p_n\right]\\
&=\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\frac{1}{n-3}\sum_{l\neq i,j,k}\eta_{n,il}^{2}\left(\omega,\tilde{\omega},\varepsilon_{n,il}\right)\middle| X_n,p_n\right]\\
&\leq\sup_{\|\omega-\tilde{\omega}\|>0}\frac{1}{n-3}\sum_{l\neq i,j,k}\mathbb{E}\left[\eta_{n,il}^{2}\left(\omega,\tilde{\omega},\varepsilon_{n,il}\right)\middle| X_n,p_n\right]+o\left(1\right)\\
&=\sup_{\|\omega-\tilde{\omega}\|>0}\frac{1}{n-3}\sum_{l\neq i,j,k}f_{\varepsilon}\left(U_{n,il}+\frac{n-1}{n-2}2D_l'\left(t_{n,il}\omega+\left(1-t_{n,il}\right)\tilde{\omega}\right)\right)\frac{2\left(n-1\right)}{n-2}+o\left(1\right)\\
&<\infty
\end{aligned}\tag{8.51}$$

by the uniform law of large numbers, dominated convergence theorem and (8.27). Combining (8.49), (8.50), and (8.51) yields

$$\mathbb{E}\left[\sup_{\|\omega-\tilde{\omega}\|>0}\left\|\frac{\mathbb{G}_{n,-jk}\left(\omega,\varepsilon_{ni,-jk}\right)-\mathbb{G}_{n,-jk}\left(\tilde{\omega},\varepsilon_{ni,-jk}\right)}{\left\|\omega-\tilde{\omega}\right\|^{1/2}}\right\|^{2}\middle| X_n,p_n\right]<\infty$$

This together with (8.47) and (8.48) implies

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\varepsilon_{n,ij}^{1},\varepsilon_{n,ik}^{1}}\left\|\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right),\varepsilon_{ni}^{1}\right)-\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{1}\right)\right\|\middle| X_n,p_n\right]\\
&\leq O\left(\frac{1}{\sqrt{n}}\right)+O\left(\mathbb{E}\left[\sup_{\varepsilon_{n,ij}^{1},\varepsilon_{n,ik}^{1}}\left\|\omega_{ni}\left(\varepsilon_{ni}^{1}\right)-\omega_{ni}\left(\varepsilon_{ni}^{2}\right)\right\|\middle| X_n,p_n\right]^{1/2}\right)
\end{aligned}\tag{8.52}$$

Combining (8.30), (8.31), (8.32), (8.42), and (8.52), taking supremium over $\varepsilon_{n,ij}^{1}$ and $\varepsilon_{n,ik}^{1}$, and taking conditional expectation with respect to the rest of ε_{ni}^{1} and ε_{ni}^{2}, we obtain

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\varepsilon_{n,ij}^{1},\varepsilon_{n,ik}^{1}}\left\|\omega_{ni}\left(\varepsilon_{ni}^{1}\right)-\omega_{ni}\left(\varepsilon_{ni}^{2}\right)\right\|\middle| X_n,p_n\right]\left(c+o\left(1\right)\right)\\
&\leq O\left(\frac{1}{n}\right)+O\left(\frac{1}{\sqrt{n}}\mathbb{E}\left[\sup_{\varepsilon_{n,ij}^{1},\varepsilon_{n,ik}^{1}}\left\|\omega_{ni}\left(\varepsilon_{ni}^{1}\right)-\omega_{ni}\left(\varepsilon_{ni}^{2}\right)\right\|\middle| X_n,p_n\right]^{1/2}\right)
\end{aligned}\tag{8.53}$$

The $O\left(n^{-1}\right)$ term in (8.53) reflects the fact that (8.31), (8.32) and the last term in (8.47) (multiplied by the additional $\left(n-1\right)^{-1/2}$ scaling in front of $\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^{1}\right),\varepsilon_{ni}^{1}\right)-\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon_{ni}^{2}\right),\varepsilon_{ni}^{1}\right)$ in (8.30)) all have the common feature of being equal to n^{-1} times some bound functions, so the rate n^{-1} could pass to the outside of their expectations. Complete the square to see that (8.53) implies

$$(c+o(1))\left(\sqrt{n}\mathbb{E}\left[\sup_{\varepsilon^1_{n,ij},\varepsilon^1_{n,ik}}\left\|\omega_{ni}\left(\varepsilon^1_{ni}\right)-\omega_{ni}\left(\varepsilon^2_{ni}\right)\right\|\middle| X_n,p_n\right]^{1/2}-O(1)\right)^2\leq O(1)$$

This can be true only if

$$\mathbb{E}\left[\sup_{\varepsilon^1_{n,ij},\varepsilon^1_{n,ik}}\left\|\omega_{ni}\left(\varepsilon^1_{ni}\right)-\omega_{ni}\left(\varepsilon^2_{ni}\right)\right\|\middle| X_n,p_n\right]\leq O\left(\frac{1}{n}\right) \tag{8.54}$$

Applying (8.54) to (8.52) we derive the desired rate in (8.46)

$$\mathbb{E}\left[\sup_{\varepsilon^1_{n,ij},\varepsilon^1_{n,ik}}\left\|\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon^1_{ni}\right),\varepsilon^1_{ni}\right)-\mathbb{G}_n\left(\omega_{ni}\left(\varepsilon^2_{ni}\right),\varepsilon^1_{ni}\right)\right\|\middle| X_n,p_n\right]\leq O\left(\frac{1}{\sqrt{n}}\right)$$

and (8.8) follows. The proof is complete. ■

Now we go back to the proof of Proposition 4.1.

Proof of Proposition 4.1. Write the conditional covariance of $G_{n,ij}$ and $G_{n,ik}$ as

$$\begin{aligned}&\mathbb{E}\left(\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right)\left(G_{n,ik}-P_{n,ik}(\theta_0,p_n)\right)\middle| X_n,p_n\right)\\&=\mathbb{E}\left(G_{n,ij}G_{n,ik}\middle| X_n,p_n\right)-\mathbb{E}\left(G_{n,ij}\middle| X_n,p_n\right)\mathbb{E}\left(G_{n,ik}\middle| X_n,p_n\right)\end{aligned} \tag{8.55}$$

Recall that

$$G_{n,ij}=1\left\{U_{n,ij}(\theta_0,p_n)+\frac{n-1}{n-2}2D_j'V_i(\theta_0,p_n)\,\omega_{ni}(\varepsilon_{ni};\theta_0,p_n)-\varepsilon_{n,ij}\geq 0\right\}$$

so

$$\begin{aligned}\mathbb{E}\left(G_{n,ij}G_{n,ik}\middle| X_n,p_n\right)=\Pr\Bigg(&\varepsilon_{n,ij}\leq U_{n,ij}(\theta_0,p_n)+\frac{n-1}{n-2}2D_j'\omega_{ni}(\varepsilon_{ni};\theta_0,p_n),\\&\varepsilon_{n,ik}\leq U_{n,ik}(\theta_0,p_n)+\frac{n-1}{n-2}2D_k'\omega_{ni}(\varepsilon_{ni};\theta_0,p_n)\Bigg| X_n,p_n\Bigg)\end{aligned} \tag{8.56}$$

and

$$\mathbb{E}\left(G_{n,ij}\middle| X_n,p_n\right)=\Pr\left(\varepsilon_{n,ij}\leq U_{n,ij}(\theta_0,p_n)+\frac{n-1}{n-2}2D_j'\omega_{ni}(\varepsilon_{ni};\theta_0,p_n)\middle| X_n,p_n\right)$$

From (8.56) we see that $G_{n,ij}$ and $G_{n,ik}$ are correlated for two reasons: (1) both $G_{n,ij}$ and $G_{n,ik}$ contain $\omega_{ni}(\varepsilon_{ni})$ which is random, and (2) $\omega_{ni}(\varepsilon_{ni})$ depends on ε_{ni}, so $\omega_{ni}(\varepsilon_{ni})$ in $G_{n,ij}$ can be correlated with $\varepsilon_{n,ik}$ in $G_{n,ik}$. Such correlation can be complicated because $\omega_{ni}(\varepsilon_{ni})$ and $\varepsilon_{n,ij}$ enter $G_{n,ij}$ through an indicator function. We deal with the correlation by dividing the space of $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ into subsets within which the indicator functions are fixed. We will show that both types of correlation vanish at the rate of n^{-1}, so the conditional covariance in (8.55) has the rate of n^{-1}.

To proceed, suppress (θ_0, p_n) hereafter for notational simplicity. Define the sets

$$\mathcal{E}_{n,ij} = \left\{\varepsilon_{n,ij} \in \mathbb{R} : \varepsilon_{n,ij} \leq U_{n,ij} + \frac{n-1}{n-2} 2D_j' \omega_{ni}^*\right\}$$
$$\mathcal{E}_{n,ik} = \left\{\varepsilon_{n,ik} \in \mathbb{R} : \varepsilon_{n,ik} \leq U_{n,ik} + \frac{n-1}{n-2} 2D_k' \omega_{ni}^*\right\}$$

We divide the space of $(\varepsilon_{n,ij}, \varepsilon_{n,ik}) \in \mathbb{R}^2$ into four subsets depending on whether $\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}$ and whether $\varepsilon_{n,ik} \in \mathcal{E}_{n,ik}$. The cross-product expectation and individual expectation product in (8.55) can thus be expressed as a sum of four parts, corresponding to the four subsets,

$$\begin{aligned}
\mathbb{E}\left(G_{n,ij}G_{n,ik} \middle| X_n, p_n\right) = & \mathbb{E}\left(G_{n,ij}G_{n,ik} 1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \\
& + \mathbb{E}\left(G_{n,ij}G_{n,ik} 1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \\
& + \mathbb{E}\left(G_{n,ij}G_{n,ik} 1\left\{\varepsilon_{n,ij} \notin \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \\
& + \mathbb{E}\left(G_{n,ij}G_{n,ik} 1\left\{\varepsilon_{n,ij} \notin \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \qquad (8.57)
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\left(G_{n,ij} \middle| X_n, p_n\right) \mathbb{E}\left(G_{n,ik} \middle| X_n, p_n\right) = & \mathbb{E}\left(G_{n,ij} 1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}\right\} \middle| X_n, p_n\right) \mathbb{E}\left(G_{n,ik} 1\left\{\varepsilon_{n,ik} \in \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \\
& + \mathbb{E}\left(G_{n,ij} 1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}\right\} \middle| X_n, p_n\right) \mathbb{E}\left(G_{n,ik} 1\left\{\varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \\
& + \mathbb{E}\left(G_{n,ij} 1\left\{\varepsilon_{n,ij} \notin \mathcal{E}_{n,ij}\right\} \middle| X_n, p_n\right) \mathbb{E}\left(G_{n,ik} 1\left\{\varepsilon_{n,ik} \in \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \\
& + \mathbb{E}\left(G_{n,ij} 1\left\{\varepsilon_{n,ij} \notin \mathcal{E}_{n,ij}\right\} \middle| X_n, p_n\right) \mathbb{E}\left(G_{n,ik} 1\left\{\varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}\right\} \middle| X_n, p_n\right) \\
& \qquad (8.58)
\end{aligned}$$

We will show that each term in (8.57) subtracting its corresponding term in (8.58) will yield a difference that is $O(n^{-1})$, so summing up the four differences the conditional covariance in (8.55) is also $O(n^{-1})$.

To this end, we first introduce some notation. Recall that by Lemma 8.1 $\omega_{ni}(\varepsilon_{ni})$ has an asymptotic linear representation

$$\omega_{ni}(\varepsilon_{ni}) = \omega_{ni}^* + \frac{1}{n-1}\sum_{l\neq i} H_{ni}^{-1}(\omega_{ni}^*)\,\psi_{n,il}(\omega_{ni}^*, \varepsilon_{n,il}) + r_n(\varepsilon_{ni})$$

where

$$\psi_{n,il}(\omega_{ni}^*, \varepsilon_{n,il}) = 1\left\{U_{n,il} + \frac{n-1}{n-2}2D_l'\omega_{ni}^* - \varepsilon_{n,il} > 0\right\} D_j - V_i^+\omega_{ni}^*.$$

Since $\varepsilon_{n,il}$ enters $\psi_{n,il}(\omega_{ni}^*, \varepsilon_{n,il})$ through an indicator function defined by set $\mathcal{E}_{n,il}$, once we restrict $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ to either one of the four subsets, the functions $\psi_{n,ij}(\omega_{ni}^*, \varepsilon_{n,ij})$ and $\psi_{n,ik}(\omega_{ni}^*, \varepsilon_{n,ik})$ no longer depend on $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$. Depending on which subset $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ lies in, we denote the value of $\omega_{ni}(\varepsilon_{ni})$ as

$$\begin{aligned}
\omega_{ni}(\varepsilon_{ni}) &= \begin{cases} \omega_{ni}^{1\cdot}(\varepsilon_{ni,-j}), & \text{if } \varepsilon_{n,ij} \in \mathcal{E}_{n,ij}, \\ \omega_{ni}^{0\cdot}(\varepsilon_{ni,-j}), & \text{if } \varepsilon_{n,ij} \notin \mathcal{E}_{n,ij}, \end{cases} + r_n(\varepsilon_{ni}) \\
&= \begin{cases} \omega_{ni}^{\cdot 1}(\varepsilon_{ni,-k}), & \text{if } \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}, \\ \omega_{ni}^{\cdot 0}(\varepsilon_{ni,-k}), & \text{if } \varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}, \end{cases} + r_n(\varepsilon_{ni}) \\
&= \begin{cases} \omega_{ni}^{11}(\varepsilon_{ni,-jk}), & \text{if } \varepsilon_{n,ij} \in \mathcal{E}_{n,ij},\ \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}, \\ \omega_{ni}^{10}(\varepsilon_{ni,-jk}), & \text{if } \varepsilon_{n,ij} \in \mathcal{E}_{n,ij},\ \varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}, \\ \omega_{ni}^{01}(\varepsilon_{ni,-jk}), & \text{if } \varepsilon_{n,ij} \notin \mathcal{E}_{n,ij},\ \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}, \\ \omega_{ni}^{00}(\varepsilon_{ni,-jk}), & \text{if } \varepsilon_{n,ij} \notin \mathcal{E}_{n,ij},\ \varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}, \end{cases} + r_n(\varepsilon_{ni})
\end{aligned}$$

where

$$\omega_{ni}^{s\cdot}(\varepsilon_{ni,-j}) = \omega_{ni}^* + \frac{1}{n-1}\sum_{l\neq i,j} H_{ni}^{-1}(\omega_{ni}^*)\,\psi_{n,il}(\omega_{ni}^*, \varepsilon_{n,il}) + \frac{1}{n-1}H_{ni}^{-1}(\omega_{ni}^*)\cdot\begin{cases} D_j - V_i^+\omega_{ni}^*, & \text{if } s=1 \\ -V_i^+\omega_{ni}^*, & \text{if } s=0 \end{cases}$$

$$\omega_{ni}^{\cdot t}(\varepsilon_{ni,-k}) = \omega_{ni}^* + \frac{1}{n-1}\sum_{l\neq i,k} H_{ni}^{-1}(\omega_{ni}^*)\,\psi_{n,il}(\omega_{ni}^*, \varepsilon_{n,il}) + \frac{1}{n-1}H_{ni}^{-1}(\omega_{ni}^*)\cdot\begin{cases} D_k - V_i^+\omega_{ni}^*, & \text{if } t=1 \\ -V_i^+\omega_{ni}^*, & \text{if } t=0 \end{cases}$$

and

$$\begin{aligned}
\omega_{ni}^{st}(\varepsilon_{ni,-jk}) = \omega_{ni}^* &+ \frac{1}{n-1}\sum_{l\neq i,j,k} H_{ni}^{-1}(\omega_{ni}^*)\,\psi_{n,il}(\omega_{ni}^*, \varepsilon_{n,il}) \\
&+ \frac{1}{n-1}H_{ni}^{-1}(\omega_{ni}^*)\cdot\begin{cases} D_j + D_k - 2V_i^+\omega_{ni}^*, & \text{if } s=1, t=1 \\ D_j - 2V_i^+\omega_{ni}^*, & \text{if } s=1, t=0 \\ D_k - 2V_i^+\omega_{ni}^*, & \text{if } s=0, t=1 \\ -2V_i^+\omega_{ni}^*, & \text{if } s=0, t=0 \end{cases}
\end{aligned}$$

Note that $\omega_{ni}^{s\cdot}$, $\omega_{ni}^{\cdot t}$, and ω_{ni}^{st} depend on ε_{ni} only through $\varepsilon_{ni,-j} = (\varepsilon_{n,il})_{l\neq i,j}$, $\varepsilon_{ni,-k} = (\varepsilon_{n,il})_{l\neq i,k}$ and $\varepsilon_{ni,-jk} = (\varepsilon_{n,il})_{l\neq i,j,k}$, respectively, for $s,t \in \{0,1\}$. For such variants of ω_{ni}, we further define

$$
\begin{aligned}
C_{n,il}^{*} &= U_{n,il} + \frac{n-1}{n-2} 2D_l' \omega_{ni}^{*} \\
C_{n,il}(\varepsilon_{ni}) &= U_{n,il} + \frac{n-1}{n-2} 2D_l' \omega_{ni}(\varepsilon_{ni}) \\
C_{n,il}^{s\cdot}(\varepsilon_{ni,-j}) &= U_{n,il} + \frac{n-1}{n-2} 2D_l' \omega_{ni}^{s\cdot}(\varepsilon_{ni,-j}), \quad s \in \{0,1\} \\
C_{n,il}^{\cdot t}(\varepsilon_{ni,-k}) &= U_{n,il} + \frac{n-1}{n-2} 2D_l' \omega_{ni}^{\cdot t}(\varepsilon_{ni,-k}), \quad t \in \{0,1\} \\
C_{n,il}^{st}(\varepsilon_{ni,-jk}) &= U_{n,il} + \frac{n-1}{n-2} 2D_l' \omega_{ni}^{st}(\varepsilon_{ni,-jk}), \quad s,t \in \{0,1\}
\end{aligned}
$$

The last three terms only cover the leading term in ω_{ni}. To deal with the remainder $r_n(\varepsilon_{ni})$ in ω_{ni}, we choose $\left(\tilde{\varepsilon}_{n,ij}^1, \tilde{\varepsilon}_{n,ik}^1\right)$ and $\left(\tilde{\varepsilon}_{n,ij}^2, \tilde{\varepsilon}_{n,ik}^2\right)$ to be independent copies of $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$, i.e., these three pairs are independent of each other and $\left(\tilde{\varepsilon}_{n,ij}^1, \tilde{\varepsilon}_{n,ik}^1\right)$ and $\left(\tilde{\varepsilon}_{n,ij}^2, \tilde{\varepsilon}_{n,ik}^2\right)$ follow the same distribution as $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$. Replacing $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ in ε_{ni} by $\left(\tilde{\varepsilon}_{n,ij}^1, \tilde{\varepsilon}_{n,ik}^1\right)$ and $\left(\tilde{\varepsilon}_{n,ij}^2, \tilde{\varepsilon}_{n,ik}^2\right)$, we construct two vectors $\tilde{\varepsilon}_{ni}^1$ and $\tilde{\varepsilon}_{ni}^2$

$$
\tilde{\varepsilon}_{ni}^1 = \left(\tilde{\varepsilon}_{n,ij}^1, \tilde{\varepsilon}_{n,ik}^1, \varepsilon_{ni,-jk}\right), \quad \tilde{\varepsilon}_{ni}^2 = \left(\tilde{\varepsilon}_{n,ij}^2, \tilde{\varepsilon}_{n,ik}^2, \varepsilon_{ni,-jk}\right)
$$

By construction, ε_{ni}, $\tilde{\varepsilon}_{ni}^1$ and $\tilde{\varepsilon}_{ni}^2$ are independent conditional on $\varepsilon_{ni,-jk}$. We then define

$$
\begin{aligned}
\Delta_{n,il}(\varepsilon_{ni}) &= \frac{n-1}{n-2} 2D_l' r_n(\varepsilon_{ni}) \\
\tilde{\Delta}_{n,il}^1\left(\tilde{\varepsilon}_{ni}^1\right) &= \frac{n-1}{n-2} 2D_l' r_n\left(\tilde{\varepsilon}_{ni}^1\right) \\
\tilde{\Delta}_{n,il}^2\left(\tilde{\varepsilon}_{ni}^2\right) &= \frac{n-1}{n-2} 2D_l' r_n\left(\tilde{\varepsilon}_{ni}^2\right)
\end{aligned}
$$

$\Delta_{n,il}$, $\tilde{\Delta}_{n,il}^1$, and $\tilde{\Delta}_{n,il}^2$ are also independent conditional on $\varepsilon_{ni,-jk}$.

Now we examine the difference between (8.57) and (8.58). We divide the task into four parts, depending on which of the four subsets $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ lies in.

Part I: $\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}$.

We start with the first terms in (8.57) and (8.58). By definition

$$
\begin{aligned}
&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij}\in\mathcal{E}_{n,ij},\varepsilon_{n,ik}\in\mathcal{E}_{n,ik}\right\}\middle|\, X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\},\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\Delta_{n,ik},C^{*}_{n,ik}\right\}\middle|\, X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\},\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\middle|\, X_n,p_n\right)\\
&+\mathbb{E}\Bigg(1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\\
&\qquad\cdot\left(1\left\{\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\Delta_{n,ik},C^{*}_{n,ik}\right\}\right\}-1\left\{\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\right\}\right)\Bigg|\, X_n,p_n\Bigg)\\
&+\mathbb{E}\Bigg(1\left\{\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\right\}\\
&\qquad\cdot\left(1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\right)\Bigg|\, X_n,p_n\Bigg)\\
&+\mathbb{E}\Bigg(\left(1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\right)\\
&\qquad\cdot\left(1\left\{\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\Delta_{n,ik},C^{*}_{n,ik}\right\}\right\}-1\left\{\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\right\}\right)\Bigg|\, X_n,p_n\Bigg)
\end{aligned}
\tag{8.59}
$$

Moreover,

$$
\begin{aligned}
&\mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij}\in\mathcal{E}_{n,ij}\right\}\middle|\, X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{1\cdot}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\middle|\, X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\middle|\, X_n,p_n\right)\\
&+\mathbb{E}\left(1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{1\cdot}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\middle|\, X_n,p_n\right)
\end{aligned}
$$

and similar derivation holds for $\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik}\in\mathcal{E}_{n,ik}\right\}\mid X_n,p_n\right)$, with j replaced by k and $C^{1\cdot}_{n,ij}$ and $\tilde{\Delta}^1_{n,ij}$ replaced by $C^{\cdot 1}_{n,ik}$ and $\tilde{\Delta}^2_{n,ik}$, so

$$
\begin{aligned}
&\mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij}\in\mathcal{E}_{n,ij}\right\}\mid X_n,p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik}\in\mathcal{E}_{n,ik}\right\}\mid X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\middle|X_n,p_n\right)\Pr\left(\varepsilon_{n,ik}\le\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\middle|X_n,p_n\right)\\
&+\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\middle|X_n,p_n\right)\\
&\quad\cdot\mathbb{E}\left(1\left\{\varepsilon_{n,ik}\le\min\left\{C^{\cdot 1}_{n,ik}+\Delta_{n,ik},C^*_{n,ik}\right\}\right\}-1\left\{\varepsilon_{n,ik}\le\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\right\}\middle|X_n,p_n\right)\\
&+\Pr\left(\varepsilon_{n,ik}\le\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\middle|X_n,p_n\right)\\
&\quad\cdot\mathbb{E}\left(1\left\{\varepsilon_{n,ij}\le\min\left\{C^{1\cdot}_{n,ij}+\Delta_{n,ij},C^*_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\right\}\middle|X_n,p_n\right)\\
&+\mathbb{E}\left(1\left\{\varepsilon_{n,ij}\le\min\left\{C^{1\cdot}_{n,ij}+\Delta_{n,ij},C^*_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\right\}\middle|X_n,p_n\right)\\
&\quad\cdot\mathbb{E}\left(1\left\{\varepsilon_{n,ik}\le\min\left\{C^{\cdot 1}_{n,ik}+\Delta_{n,ik},C^*_{n,ik}\right\}\right\}-1\left\{\varepsilon_{n,ik}\le\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\right\}\middle|X_n,p_n\right)
\end{aligned}
\tag{8.60}
$$

We then compare the terms in (8.59) with those in (8.60). We do this in four steps.

Step 1.1: Compare the first terms in (8.59) and (8.60). Because $C^{11}_{n,ij}$ is a function of $\varepsilon_{ni,-jk}$, $\tilde{\Delta}^1_{n,ij}$ and $\tilde{\Delta}^2_{n,ij}$ are independent conditional on $\varepsilon_{ni,-jk}$, and $\varepsilon_{n,ij}$ are i.i.d. by Assumption 1, the first term in (8.59) is given by

$$
\begin{aligned}
&\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\},\varepsilon_{n,ik}\le\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\middle|X_n,p_n\right)\\
&=\mathbb{E}\left(\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\},\varepsilon_{n,ik}\le\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\middle|\varepsilon_{ni,-jk},X_n,p_n\right)\middle|X_n,p_n\right)\\
&=\mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\right)F_\varepsilon\left(\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\right)\middle|X_n,p_n\right)
\end{aligned}
$$

where F_ε is the CDF of $\varepsilon_{n,ij}$. Moreover,

$$
\begin{aligned}
&\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\middle|X_n,p_n\right)\\
&=\mathbb{E}\left(\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\middle|\varepsilon_{ni,-jk},X_n,p_n\right)\middle|X_n,p_n\right)\\
&=\Pr\left(F_\varepsilon\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\right)\middle|X_n,p_n\right)
\end{aligned}
$$

Therefore, if we subtract the first term in (8.60) from the first term in (8.59), the difference is given by

$$\begin{aligned}
&\Pr\left(\varepsilon_{n,ij} \leq \min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1}, C_{n,ij}^{*}\right\}, \varepsilon_{n,ik} \leq \min\left\{C_{n,ik}^{11} + \tilde{\Delta}_{n,ik}^{2}, C_{n,ik}^{*}\right\} \middle| X_n, p_n\right) \\
&\quad - \Pr\left(\varepsilon_{n,ij} \leq \min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1}, C_{n,ij}^{*}\right\} \middle| X_n, p_n\right) \Pr\left(\varepsilon_{n,ik} \leq \min\left\{C_{n,ik}^{11} + \tilde{\Delta}_{n,ik}^{2}, C_{n,ik}^{*}\right\} \middle| X_n, p_n\right) \\
=&\mathbb{E}\left(F_\varepsilon\left(\min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1}, C_{n,ij}^{*}\right\}\right) F_\varepsilon\left(\min\left\{C_{n,ik}^{11} + \tilde{\Delta}_{n,ik}^{2}, C_{n,ik}^{*}\right\}\right) \middle| X_n, p_n\right) \\
&\quad - \mathbb{E}\left(F_\varepsilon\left(\min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1}, C_{n,ij}^{*}\right\}\right) \middle| X_n, p_n\right) \mathbb{E}\left(F_\varepsilon\left(\min\left\{C_{n,ik}^{11} + \tilde{\Delta}_{n,ik}^{2}, C_{n,ik}^{*}\right\}\right) \middle| X_n, p_n\right) \\
=&Cov\left(F_\varepsilon\left(\min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1}, C_{n,ij}^{*}\right\}\right), F_\varepsilon\left(\min\left\{C_{n,ik}^{11} + \tilde{\Delta}_{n,ik}^{2}, C_{n,ik}^{*}\right\}\right) \middle| X_n, p_n\right) \\
=&Cov\left(F_\varepsilon\left(\min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1}, C_{n,ij}^{*}\right\}\right) - F_\varepsilon\left(C_{n,ij}^{*}\right), F_\varepsilon\left(\min\left\{C_{n,ik}^{11} + \tilde{\Delta}_{n,ik}^{2}, C_{n,ik}^{*}\right\}\right) - F_\varepsilon\left(C_{n,ik}^{*}\right) \middle| X_n, p_n\right)
\end{aligned} \tag{8.61}$$

We will show this conditional covariance is $O\left(n^{-1}\right)$. By Taylor expansion,

$$\begin{aligned}
&F_\varepsilon\left(\min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1}, C_{n,ij}^{*}\right\}\right) - F_\varepsilon\left(C_{n,ij}^{*}\right) \\
&= f_\varepsilon\left(C_{n,ij}^{*}\right) \min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1} - C_{n,ij}^{*}, 0\right\} + o\left(\left|\min\left\{C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1} - C_{n,ij}^{*}, 0\right\}\right|\right)
\end{aligned} \tag{8.62}$$

where f_ε is the density of F_ε. From the definition of $C_{n,ij}^{11}$, $\tilde{\Delta}_{n,ij}^{1}$ and $C_{n,ij}^{*}$,

$$\begin{aligned}
&C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1} - C_{n,ij}^{*} \\
&= \frac{n-1}{n-2} 2D_j'\left(\frac{1}{n-1}\sum_{l \neq i,j,k} H_{ni}^{-1}\left(\omega_{ni}^{*}\right)\psi_{n,il}\left(\omega_{ni}^{*}, \varepsilon_{n,il}\right) + \frac{1}{n-1} H_{ni}^{-1}\left(\omega_{ni}^{*}\right)\left(D_j + D_k - 2V_i^{+}\omega_{ni}^{*}\right) + r_n\left(\tilde{\varepsilon}_{ni}^{1}\right)\right)
\end{aligned}$$

In the parentheses above, the second term is $O\left(n^{-1}\right)$ and by Lemma 8.1 the remainder $r_n\left(\tilde{\varepsilon}_{ni}^{1}\right) = o_p\left(n^{-1/2}\right)$, so they are dominated by the first term, which is $O_p\left(n^{-1/2}\right)$ by CLT. Thus

$$C_{n,ij}^{11} + \tilde{\Delta}_{n,ij}^{1} - C_{n,ij}^{*} = 2D_j' \frac{1}{n-2}\sum_{l \neq i,j,k} H_{ni}^{-1}\left(\omega_{ni}^{*}\right)\psi_{n,il}\left(\omega_{ni}^{*}, \varepsilon_{n,il}\right) + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{8.63}$$

Similarly, we can derive

$$C_{n,ik}^{11} + \tilde{\Delta}_{n,ik}^{2} - C_{n,ik}^{*} = 2D_k' \frac{1}{n-2}\sum_{l \neq i,j,k} H_{ni}^{-1}\left(\omega_{ni}^{*}\right)\psi_{n,il}\left(\omega_{ni}^{*}, \varepsilon_{n,il}\right) + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{8.64}$$

Hence,

$$
\begin{aligned}
&\left|Cov\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij}-C^{*}_{n,ij},0\right\},\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-C^{*}_{n,ik},0\right\}\middle|\,X_n,p_n\right)\right| \\
&\leq Var\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij}-C^{*}_{n,ij},0\right\}\middle|\,X_n,p_n\right)^{\frac{1}{2}} Var\left(\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-C^{*}_{n,ik},0\right\}\middle|\,X_n,p_n\right)^{\frac{1}{2}} \\
&\leq Var\left(C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij}-C^{*}_{n,ij}\middle|\,X_n,p_n\right)^{\frac{1}{2}} Var\left(C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-C^{*}_{n,ik}\middle|\,X_n,p_n\right)^{\frac{1}{2}} \\
&= O\left(\frac{1}{n}\right) \qquad (8.65)
\end{aligned}
$$

where the second inequality holds because for any random variable Z we can write $Z = [Z]_- + [Z]_+$, where $[Z]_- = \min\{Z,0\}$ and $[Z]_+ = \max\{Z,0\}$, and note that

$$Cov\left([Z]_-,[Z]_+\right)=\mathbb{E}\left([Z]_-\cdot[Z]_+\right)-\mathbb{E}\left([Z]_-\right)\mathbb{E}\left([Z]_+\right)=-\mathbb{E}\left([Z]_-\right)\mathbb{E}\left([Z]_+\right)\geq 0$$

so

$$Var\left(Z\right)=Var\left([Z]_-\right)+Var\left([Z]_+\right)+2Cov\left([Z]_-,[Z]_+\right)\geq Var\left([Z]_-\right) \qquad (8.66)$$

The third equality in (8.65) follows because from (8.63) we obtain

$$
\begin{aligned}
&Var\left(C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij}-C^{*}_{n,ij}\middle|\,X_n,p_n\right) \\
&\leq 4D_j' Var\left(\frac{1}{n-1}\sum_{l\neq i,j,k}H^{-1}_{ni}\left(\omega^{*}_{ni}\right)\psi_{n,il}\left(\omega^{*}_{ni},\varepsilon_{n,il}\right)\middle|\,X_n,p_n\right)D_j+o\left(\frac{1}{n}\right) \\
&= O\left(\frac{1}{n}\right)
\end{aligned}
$$

and similarly $Var\left(C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}-C^*_{n,ik}\middle| X_n,p_n\right)=O\left(n^{-1}\right)$. Therefore applying the Taylor expansion in (8.62) we derive the conditional covariance in (8.61) as

$$\begin{aligned}
&Cov\left(F_\varepsilon\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}-F_\varepsilon\left(C^*_{n,ij}\right)\right),F_\varepsilon\left(\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}-F_\varepsilon\left(C^*_{n,ik}\right)\right)\middle| X_n,p_n\right)\\
=&f_\varepsilon\left(C^*_{n,ij}\right)f_\varepsilon\left(C^*_{n,ik}\right)Cov\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij}-C^*_{n,ij},0\right\},\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}-C^*_{n,ik},0\right\}\middle| X_n,p_n\right)\\
&+f_\varepsilon\left(C^*_{n,ij}\right)Cov\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij}-C^*_{n,ij},0\right\},o\left(\left|\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}-C^*_{n,ik},0\right\}\right|\right)\middle| X_n,p_n\right)\\
&+f_\varepsilon\left(C^*_{n,ik}\right)Cov\left(\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}-C^*_{n,ik},0\right\},o\left(\left|\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij}-C^*_{n,ij},0\right\}\right|\right)\middle| X_n,p_n\right)\\
&+Cov\left(o\left(\left|\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij}-C^*_{n,ij},0\right\}\right|\right),o\left(\left|\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}-C^*_{n,ik},0\right\}\right|\right)\middle| X_n,p_n\right)\\
\leq&f_\varepsilon\left(C^*_{n,ij}\right)f_\varepsilon\left(C^*_{n,ik}\right)Cov\left(\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij}-C^*_{n,ij},0\right\},\min\left\{C^{11}_{n,ik}\left(\varepsilon_{ni,-jk}\right)-C^*_{n,ik},0\right\}\middle| X_n,p_n\right)+o\left(\frac{1}{n}\right)\\
=&O\left(\frac{1}{n}\right) \qquad (8.67)
\end{aligned}$$

Combining (8.61) and (8.67) proves

$$\begin{aligned}
&\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\},\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\middle| X_n,p_n\right)\\
&-\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}\middle| X_n,p_n\right)\Pr\left(\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\middle| X_n,p_n\right)\\
&\leq O\left(\frac{1}{n}\right)
\end{aligned}$$

While the events $\varepsilon_{n,ij}\leq\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij},C^*_{n,ij}\right\}$ and $\varepsilon_{n,ik}\leq\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}$ are correlated due to the correlation of $C^{11}_{n,ij}+\tilde{\Delta}^1_{n,ij}$ and $C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}$ through $\varepsilon_{ni,-jk}$, because $C^{11}_{n,ij}$ and $C^{11}_{n,ik}$ converge to $C^*_{n,ij}$ and $C^*_{n,ik}$ respectively at the rate of $n^{-1/2}$, such correlation vanishes to 0 at the rate of n^{-1}.

Step 1.2: Consider the second terms in (8.59) and (8.60). Since

$$\begin{aligned}
&\left|\min\left\{C^{\cdot1}_{n,ik}+\Delta_{n,ik},C^*_{n,ik}\right\}-\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik},C^*_{n,ik}\right\}\right|\\
=&\left|\min\left\{C^{\cdot1}_{n,ik}+\Delta_{n,ik}-C^*_{n,ik},0\right\}-\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}-C^*_{n,ik},0\right\}\right|\\
\leq&\left|C^{\cdot1}_{n,ik}+\Delta_{n,ik}-C^*_{n,ik}-\left(C^{11}_{n,ik}+\tilde{\Delta}^2_{n,ik}-C^*_{n,ik}\right)\right|\\
=&\left|C^{\cdot1}_{n,ik}-C^{11}_{n,ik}+\Delta_{n,ik}-\tilde{\Delta}^2_{n,ik}\right|
\end{aligned}$$

where the inequality follows because $|\min\{x,0\}-\min\{y,0\}| \leq |x-y|$, and by definition and Lemma 8.2,

$$
\begin{aligned}
& C^{\cdot 1}_{n,ik} - C^{11}_{n,ik} + \Delta_{n,ik} - \tilde{\Delta}^2_{n,ik} \\
&= \frac{n-1}{n-2} 2D'_k \left(\frac{1}{n-1} H^{-1}_{ni}(\omega^*_{ni}) \left(\psi_{n,ij}(\omega^*_{ni}, \varepsilon_{n,ij}) - \left(D_j - V_i^+ \omega^*_{ni}\right)\right) + r_n(\varepsilon_{ni}) - r_n\left(\tilde{\varepsilon}^2_{ni}\right)\right)
\end{aligned}
$$

we define d_n

$$
d_n = \frac{2}{n-2} \left| D'_k H^{-1}_{ni}(\omega^*_{ni}) \left(\psi_{n,ij}(\omega^*_{ni}, \varepsilon_{n,ij}) - \left(D_j - V_i^+ \omega^*_{ni}\right)\right)\right| + \frac{2(n-1)}{n-2} \sup_{\varepsilon_{n,ik}} \left\| r_n(\varepsilon_{ni}) - r_n\left(\tilde{\varepsilon}^2_{ni}\right)\right\| \tag{8.68}
$$

Then d_n provides an upper bound

$$
\left| \min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\} - \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right| \leq d_n
$$

that is independent of $\varepsilon_{n,ik}$. Hence a rate-determining part in both second terms in (8.59) and (8.60) can be bounded as

$$
\begin{aligned}
& \left| \mathbb{E}\left(1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right\} \middle| X_n, p_n\right)\right| \\
& \leq \mathbb{E}\left(\left| 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right\}\right| \middle| X_n, p_n\right) \\
& \leq \mathbb{E}\left(1\left\{\varepsilon_{n,ik} \text{ lies between } \min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\} \text{ and } \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right\} \middle| X_n, p_n\right) \\
& \leq \mathbb{E}\left(1\left\{\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\} - d_n \leq \varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\} + d_n\right\} \middle| X_n, p_n\right) \\
& = \mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\} + d_n\right) - F_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\} - d_n\right) \middle| X_n, p_n\right),
\end{aligned}
$$

where the second inequality is from the fact that if events $\varepsilon_{n,ik} \leq \min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}$ and $\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}$ do not occur at the same time, $\varepsilon_{n,ik}$ must lie between $\min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}$ and $\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}$. The last equality follows because $\varepsilon_{n,ik}$ is independent of $C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}$ and d_n. By Taylor expansion

$$
\begin{aligned}
& \mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\} + d_n\right) - F_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\} - d_n\right) \middle| X_n, p_n\right) \\
& = \mathbb{E}\left(2f_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right) d_n + o(d_n) \middle| X_n, p_n\right) \\
& = O\left(\frac{1}{n}\right).
\end{aligned}
$$

where the last equality follows because $\mathbb{E}\left(d_n \middle| X_n, p_n\right) = O\left(n^{-1}\right)$ by (8.8) in Lemma 8.2. Therefore, the second terms in (8.59) and (8.60) satisfy

$$
\begin{aligned}
&\left|\mathbb{E}\left(1\left\{\varepsilon_{n,ij} \leq \min\left\{C^{11}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right\}\right.\right. \\
&\left.\left.\cdot\left(1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right\}\right)\right| X_n, p_n\right)\right| \\
&\leq \mathbb{E}\left(\left|1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right\}\right| \middle| X_n, p_n\right) \\
&\leq O\left(\frac{1}{n}\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\left|\Pr\left(\varepsilon_{n,ij} \leq \min\left\{C^{11}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\} \middle| X_n, p_n\right)\right. \\
&\left.\cdot\mathbb{E}\left(1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right\} \middle| X_n, p_n\right)\right| \\
&\leq \mathbb{E}\left(\left|1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{\cdot 1}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right\}\right| \middle| X_n, p_n\right) \\
&\leq O\left(\frac{1}{n}\right).
\end{aligned}
$$

We conclude that both second terms in (8.59) and (8.60) are $O\left(n^{-1}\right)$.

Step 1.3: Consider the third terms in (8.59) and (8.60). Following the same proof as in Step 1.2, with j and k swapped and $C^{\cdot 1}_{n,ik}$, $\tilde{\Delta}^2_{n,ik}$ and $\tilde{\varepsilon}^2_{ni}$ replaced by $C^{1\cdot}_{n,ij}$, $\tilde{\Delta}^1_{n,ij}$ and $\tilde{\varepsilon}^1_{ni}$, we can show that both terms are $O\left(n^{-1}\right)$.

Step 1.4: Consider the last terms in (8.59) and (8.60). From Steps 1.2-1.3, the last term in (8.60) is clearly $O\left(n^{-2}\right)$. As for the last term in (8.59), we follow the idea in Steps 1.2-1.3. Observe that

$$
\begin{aligned}
&\left|\min\left\{C^{11}_{n,ij} + \Delta_{n,ij}, C^*_{n,ij}\right\} - \min\left\{C^{11}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right| \\
&\leq \left|\Delta_{n,ij} - \tilde{\Delta}^1_{n,ij}\right| \\
&= \left|\frac{n-1}{n-2} 2D'_j\left(r_n\left(\varepsilon_{ni}\right) - r_n\left(\tilde{\varepsilon}^1_{ni}\right)\right)\right|
\end{aligned}
$$

and

$$\begin{aligned}
&\left|\min\left\{C^{11}_{n,ik}+\Delta_{n,ik},C^{*}_{n,ik}\right\}-\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\right| \\
&\leq\left|\Delta_{n,ik}-\tilde{\Delta}^{2}_{n,ik}\right| \\
&=\left|\frac{n-1}{n-2}2D'_{k}\left(r_{n}\left(\varepsilon_{ni}\right)-r_{n}\left(\tilde{\varepsilon}^{2}_{ni}\right)\right)\right|
\end{aligned}$$

Define d^1_n and d^2_n by

$$\begin{aligned}
d^1_n &= \frac{2\left(n-1\right)}{n-2}\sup_{\varepsilon_{n,ij},\varepsilon_{n,ik}}\left\|r_n\left(\varepsilon_{ni}\right)-r_n\left(\tilde{\varepsilon}^{1}_{ni}\right)\right\| \\
d^2_n &= \frac{2\left(n-1\right)}{n-2}\sup_{\varepsilon_{n,ij},\varepsilon_{n,ik}}\left\|r_n\left(\varepsilon_{ni}\right)-r_n\left(\tilde{\varepsilon}^{2}_{ni}\right)\right\| \qquad (8.69)
\end{aligned}$$

Then d^1_n and d^2_n provide upper bounds

$$\begin{aligned}
\left|\min\left\{C^{11}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}-\min\left\{C^{11}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right| &\leq d^1_n \\
\left|\min\left\{C^{11}_{n,ik}+\Delta_{n,ik},C^{*}_{n,ik}\right\}-\min\left\{C^{11}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\right| &\leq d^2_n
\end{aligned}$$

that are independent of $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ and are independent of each other conditional on $\varepsilon_{ni,-jk}$. Hence, the last term in (8.59) satisfies

$$
\begin{aligned}
&\left|\mathbb{E}\left(\left(1\left\{\varepsilon_{n,ij} \leq \min\left\{C^{11}_{n,ij} + \Delta_{n,ij}, C^{*}_{n,ij}\right\}\right\} - 1\left\{\varepsilon_{n,ij} \leq \min\left\{C^{11}_{n,ij} + \tilde{\Delta}^{1}_{n,ij}, C^{*}_{n,ij}\right\}\right\}\right)\right.\right.\\
&\quad \left.\left.\cdot\left(1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \Delta_{n,ik}, C^{*}_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^{2}_{n,ik}, C^{*}_{n,ik}\right\}\right\}\right)\right| X_n, p_n\right)\right|\\
\leq&\mathbb{E}\left(\left|1\left\{\varepsilon_{n,ij} \leq \min\left\{C^{11}_{n,ij} + \Delta_{n,ij}, C^{*}_{n,ij}\right\}\right\} - 1\left\{\varepsilon_{n,ij} \leq \min\left\{C^{11}_{n,ij} + \tilde{\Delta}^{1}_{n,ij}, C^{*}_{n,ij}\right\}\right\}\right|\right.\\
&\quad \left.\cdot\left|1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \Delta_{n,ik}, C^{*}_{n,ik}\right\}\right\} - 1\left\{\varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^{2}_{n,ik}, C^{*}_{n,ik}\right\}\right\}\right|\right| X_n, p_n\right)\\
\leq&\mathbb{E}\left(1\left\{\min\left\{C^{11}_{n,ij} + \tilde{\Delta}^{1}_{n,ij}, C^{*}_{n,ij}\right\} - d^1_n \leq \varepsilon_{n,ij} \leq \min\left\{C^{11}_{n,ij} + \tilde{\Delta}^{1}_{n,ij}, C^{*}_{n,ij}\right\} + d^1_n\right\}\right.\\
&\quad \left.\cdot 1\left\{\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^{2}_{n,ik}, C^{*}_{n,ik}\right\} - d^2_n \leq \varepsilon_{n,ik} \leq \min\left\{C^{11}_{n,ik} + \tilde{\Delta}^{2}_{n,ik}, C^{*}_{n,ik}\right\} + d^2_n\right\}\right| X_n, p_n\right)\\
=&\mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{11}_{n,ij} + \tilde{\Delta}^{1}_{n,ij}, C^{*}_{n,ij}\right\} + d^1_n\right) - F_\varepsilon\left(\min\left\{C^{11}_{n,ij} + \tilde{\Delta}^{1}_{n,ij}, C^{*}_{n,ij}\right\} - d^1_n\right)\right| X_n, p_n\right)\\
&\quad \cdot\mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^{2}_{n,ik}, C^{*}_{n,ik}\right\} + d^2_n\right) - F_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^{2}_{n,ik}, C^{*}_{n,ik}\right\} - d^2_n\right)\right| X_n, p_n\right)\\
=&\mathbb{E}\left(2 f_\varepsilon\left(\min\left\{C^{11}_{n,ij} + \tilde{\Delta}^{1}_{n,ij}, C^{*}_{n,ij}\right\}\right) d^1_n + o\left(d^1_n\right)\right| X_n, p_n\right)\\
&\quad \cdot\mathbb{E}\left(2 f_\varepsilon\left(\min\left\{C^{11}_{n,ik} + \tilde{\Delta}^{2}_{n,ik}, C^{*}_{n,ik}\right\}\right) d^2_n + o\left(d^2_n\right)\right| X_n, p_n\right)\\
\leq& O\left(\frac{1}{n^2}\right)
\end{aligned}
$$

where the first equality results from the assumptions that (i) $\varepsilon_{n,ij}$ and $\varepsilon_{n,ik}$ are independent, (ii) $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ is independent of $\left(\tilde{\Delta}^1_{n,ij}, \tilde{\Delta}^2_{n,ik}\right)$ and (d^1_n, d^2_n), (iii) $\tilde{\Delta}^1_{n,ij}$ and $\tilde{\Delta}^2_{n,ik}$ are independent conditional on $\varepsilon_{ni,-jk}$, and (iv) d^1_n and d^2_n are independent conditional on $\varepsilon_{ni,-jk}$. The last equality follows because $\mathbb{E}\left(d^1_n\middle| X_n, p_n\right) \leq O\left(n^{-1}\right)$ and $\mathbb{E}\left(d^2_n\middle| X_n, p_n\right) \leq O\left(n^{-1}\right)$ by (8.8) in Lemma 8.2. We conclude that the last terms in (8.59) and (8.60) are $O\left(n^{-2}\right)$ and thus negligible.

Combining Steps 1.1-1.4, we have proved that

$$
\begin{aligned}
&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}\right\}\middle| X_n, p_n\right)\\
&- \mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}\right\}\middle| X_n, p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik} \in \mathcal{E}_{n,ik}\right\}\middle| X_n, p_n\right)\\
\leq& O\left(\frac{1}{n}\right)
\end{aligned}
$$

Part II: $\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}$.

We turn to the second terms in (8.57) and (8.58). They can be represented as

$$
\begin{aligned}
&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij}\in\mathcal{E}_{n,ij},\varepsilon_{n,ik}\notin\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\},C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\Delta_{n,ik}\middle|\,X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\},C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\middle|\,X_n,p_n\right)\\
&+\mathbb{E}\Bigg(1\left\{\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\\
&\quad\cdot\left(1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\right)\Bigg|\,X_n,p_n\Bigg)\\
&+\mathbb{E}\Bigg(1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\\
&\quad\cdot\left(1\left\{\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\right)\Bigg|\,X_n,p_n\Bigg)\\
&+\mathbb{E}\Bigg(\left(1\left\{\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\right)\\
&\quad\cdot\left(1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\right)\Bigg|\,X_n,p_n\Bigg)
\end{aligned}
\tag{8.70}
$$

and

$$
\begin{aligned}
&\mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij}\in\mathcal{E}_{n,ij}\right\}\middle|\,X_n,p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik}\notin\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{1\cdot}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\middle|\,X_n,p_n\right)\Pr\left(C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\middle|\,X_n,p_n\right)\\
=&\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\middle|\,X_n,p_n\right)\Pr\left(C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\middle|\,X_n,p_n\right)\\
&+\Pr\left(\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\middle|\,X_n,p_n\right)\\
&\quad\cdot\mathbb{E}\left(1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\
&+\Pr\left(C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\middle|\,X_n,p_n\right)\\
&\quad\cdot\mathbb{E}\left(1\left\{\varepsilon_{n,ij}\le\min\left\{C^{1\cdot}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\middle|\,X_n,p_n\right)\\
&+\mathbb{E}\left(1\left\{\varepsilon_{n,ij}\le\min\left\{C^{1\cdot}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}\right\}-1\left\{\varepsilon_{n,ij}\le\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\middle|\,X_n,p_n\right)\\
&\quad\cdot\mathbb{E}\left(1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\middle|\,X_n,p_n\right)
\end{aligned}
\tag{8.71}
$$

The representations in (8.70) and (8.71) are similar to those in (8.59) and (8.60). The main difference is that $\varepsilon_{n,ik}\le\min\left\{C^{11}_{n,ik}+\Delta_{n,ik},C^{*}_{n,ik}\right\}$ is replaced by $C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{10}_{n,ik}+\Delta_{n,ik}$. Following the proof in Part I, we show in four steps that the difference between (8.70) and (8.71) is also $O\left(n^{-1}\right)$.

Step 2.1: Compare the first terms in (8.70) and (8.71). Since

$$
\begin{aligned}
&\Pr\left(\varepsilon_{n,ij} \leq \min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}, C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right| X_n, p_n\Big) \\
=&\mathbb{E}\left(\Pr\left(\varepsilon_{n,ij} \leq \min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}, C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right| \varepsilon_{ni,-jk}, X_n, p_n\Big)\Big| X_n, p_n\right) \\
=&\mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right)\left(F_\varepsilon\left(\max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right) - F_\varepsilon\left(C^*_{n,ik}\right)\right)\Big| X_n, p_n\right)
\end{aligned}
$$

and

$$
\begin{aligned}
&\Pr\left(C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right| X_n, p_n\Big) \\
=&\mathbb{E}\left(\Pr\left(C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right| \varepsilon_{ni,-jk}, X_n, p_n\Big)\Big| X_n, p_n\right) \\
=&\mathbb{E}\left(F_\varepsilon\left(\max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right) - F_\varepsilon\left(C^*_{n,ik}\right)\Big| X_n, p_n\right)
\end{aligned}
$$

subtracting the first term in (8.71) from the first term in (8.70) yields

$$
\begin{aligned}
&\Pr\left(\varepsilon_{n,ij} \leq \min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}, C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right| X_n, p_n\Big) \\
&- \Pr\left(\varepsilon_{n,ij} \leq \min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\Big| X_n, p_n\right) \Pr\left(C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right| X_n, p_n\Big) \\
=&\mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right)\left(F_\varepsilon\left(\max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right) - F_\varepsilon\left(C^*_{n,ik}\right)\right)\Big| X_n, p_n\right) \\
&- \mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right)\Big| X_n, p_n\right)\mathbb{E}\left(F_\varepsilon\left(\max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right) - F_\varepsilon\left(C^*_{n,ik}\right)\Big| X_n, p_n\right) \\
=&Cov\left(F_\varepsilon\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right), F_\varepsilon\left(\max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right) - F_\varepsilon\left(C^*_{n,ik}\right)\Big| X_n, p_n\right) \\
=&Cov\left(F_\varepsilon\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right) - F_\varepsilon\left(C^*_{n,ik}\right), F_\varepsilon\left(\max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}, C^*_{n,ik}\right\}\right) - F_\varepsilon\left(C^*_{n,ik}\right)\Big| X_n, p_n\right)
\end{aligned}
\tag{8.72}
$$

Similar to the proof in Step 1.1, we have

$$
\begin{aligned}
&\left|Cov\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij} - C^*_{n,ij}, 0\right\}, \max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik} - C^*_{n,ik}, 0\right\}\Big| X_n, p_n\right)\right| \\
&\leq Var\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij} - C^*_{n,ij}, 0\right\}\Big| X_n, p_n\right)^{\frac{1}{2}} Var\left(\max\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik} - C^*_{n,ik}, 0\right\}\Big| X_n, p_n\right)^{\frac{1}{2}} \\
&\leq Var\left(C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij} - C^*_{n,ij}\Big| X_n, p_n\right)^{\frac{1}{2}} Var\left(C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik} - C^*_{n,ik}\Big| X_n, p_n\right)^{\frac{1}{2}} \\
&= O\left(\frac{1}{n}\right)
\end{aligned}
$$

where the second inequality follows because (8.66) implies that $Var\left(\max\left\{Z,0\right\}\right) \leq Var\left(Z\right)$. Therefore, applying the Taylor expansion

$$\begin{aligned}&F_{\varepsilon}\left(\max\left\{C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\right)-F_{\varepsilon}\left(C^{*}_{n,ik}\right)\\&=f_{\varepsilon}\left(C^{*}_{n,ik}\right)\max\left\{C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-C^{*}_{n,ik},0\right\}o\left(\left|\max\left\{C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-C^{*}_{n,ik},0\right\}\right|\right)\end{aligned}$$

we bound the covariance in (8.72) as

$$\begin{aligned}&Cov\left(F_{\varepsilon}\left(\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}-F_{\varepsilon}\left(C^{*}_{n,ij}\right)\right),F_{\varepsilon}\left(\max\left\{C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik},C^{*}_{n,ik}\right\}\right)-F_{\varepsilon}\left(C^{*}_{n,ik}\right)\middle|X_{n},p_{n}\right)\\&\leq f_{\varepsilon}\left(C^{*}_{n,ij}\right)f_{\varepsilon}\left(C^{*}_{n,ik}\right)Cov\left(\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij}-C^{*}_{n,ij},0\right\},\max\left\{C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-C^{*}_{n,ik},0\right\}\middle|X_{n},p_{n}\right)+o\left(\frac{1}{n}\right)\\&=O\left(\frac{1}{n}\right)\end{aligned}$$

This proves that the first terms in (8.70) and (8.71) have a difference that is $O\left(n^{-1}\right)$.

$$\begin{aligned}&\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\},C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\middle|X_{n},p_{n}\right)\\&-\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\middle|X_{n},p_{n}\right)\Pr\left(C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\middle|X_{n},p_{n}\right)\\&\leq O\left(\frac{1}{n}\right)\end{aligned}$$

Step 2.2: Consider the second terms in (8.70) and (8.71). Observe that

$$\begin{aligned}&\left|1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\right|\\&\leq 1\left\{\varepsilon_{n,ik}\text{ lies between }C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\text{ and }C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\end{aligned}$$

Moreover, by definition and Lemma 8.2,

$$\begin{aligned}&C^{\cdot 0}_{n,ik}+\Delta_{n,ik}-\left(C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right)\\&=\frac{n-1}{n-2}2D'_{k}\left(\frac{1}{n-1}H^{-1}_{ni}\left(\omega^{*}_{ni}\right)\left(\psi_{n,ij}\left(\omega^{*}_{ni},\varepsilon_{n,ij}\right)-\left(D_{j}-V^{+}_{i}\omega^{*}_{ni}\right)\right)+r_{n}\left(\varepsilon_{ni}\right)-r_{n}\left(\tilde{\varepsilon}^{2}_{ni}\right)\right)\end{aligned}$$

so the d_n defined in (8.68) provides an upper bound

$$\left|C^{\cdot 0}_{n,ik}+\Delta_{n,ik}-\left(C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right)\right|\leq d_{n}$$

that is independent of $\varepsilon_{n,ik}$. Then

$$\begin{aligned}
&\mathbb{E}\left(\left|1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\right|\,\middle|\, X_n,p_n\right)\\
\leq&\mathbb{E}\left(1\left\{\varepsilon_{n,ik}\text{ lies between }C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\text{ and }C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\,\middle|\, X_n,p_n\right)\\
\leq&\mathbb{E}\left(1\left\{C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-d_n\leq\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}+d_n\right\}\,\middle|\, X_n,p_n\right)\\
=&\mathbb{E}\left(F_{\varepsilon}\left(C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}+d_n\right)-F_{\varepsilon}\left(C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}-d_n\right)\,\middle|\, X_n,p_n\right)\\
=&\mathbb{E}\left(2f_{\varepsilon}\left(C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right)d_n+o\left(d_n\right)\,\middle|\, X_n,p_n\right)\\
=&O\left(\frac{1}{n}\right)
\end{aligned}$$

Therefore, both second terms in (8.70) and (8.71) are $O\left(n^{-1}\right)$ since

$$\begin{aligned}
&\left|\mathbb{E}\left(1\left\{\varepsilon_{n,ij}\leq\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right\}\right.\right.\\
&\left.\left.\cdot\left(1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\right)\,\middle|\, X_n,p_n\right)\right|\\
\leq&\mathbb{E}\left(\left|1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\right|\,\middle|\, X_n,p_n\right)\\
\leq&O\left(\frac{1}{n}\right)
\end{aligned}$$

and

$$\begin{aligned}
&\left|\Pr\left(\varepsilon_{n,ij}\leq\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\,\middle|\, X_n,p_n\right)\right.\\
&\left.\cdot\mathbb{E}\left(1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\,\middle|\, X_n,p_n\right)\right|\\
\leq&\mathbb{E}\left(\left|1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\right\}-1\left\{C^{*}_{n,ik}<\varepsilon_{n,ik}\leq C^{10}_{n,ik}+\tilde{\Delta}^{2}_{n,ik}\right\}\right|\,\middle|\, X_n,p_n\right)\\
\leq&O\left(\frac{1}{n}\right)
\end{aligned}$$

Step 2.3: Observe that the third terms in (8.70) and (8.71) are obtained from the second terms with j and k swapped. Using the same proof in Step 2.2 we can show that the third terms in (8.70) and (8.71) are also $O\left(n^{-1}\right)$.

Step 2.4: Consider the last terms in (8.70) and (8.71). From Steps 1.2 and 2.2, it is clear that the last term in (8.71) is $O\left(n^{-2}\right)$. Moreover, since

$$\left|\min\left\{C^{10}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ij}\right\}-\min\left\{C^{10}_{n,ij}+\tilde{\Delta}^{1}_{n,ij},C^{*}_{n,ij}\right\}\right|\leq\left|\Delta_{n,ij}-\tilde{\Delta}^{1}_{n,ij}\right|$$

and

$$\left|C^{10}_{n,ik} + \Delta_{n,ik} - \left(C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right)\right| \leq \left|\Delta_{n,ik} - \tilde{\Delta}^2_{n,ik}\right|$$

the d^1_n and d^2_n defined in (8.69) provide upper bounds

$$\begin{aligned}\left|\min\left\{C^{10}_{n,ij} + \Delta_{n,ij}, C^*_{n,ij}\right\} - \min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right| &\leq d^1_n \\ \left|C^{10}_{n,ik} + \Delta_{n,ik} - \left(C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right)\right| &\leq d^2_n\end{aligned}$$

that are independent of $(\varepsilon_{n,ij}, \varepsilon_{n,ik})$ and are independent of each other conditional on $\varepsilon_{ni,-jk}$. Then the last term in (8.59) satisfies

$$\begin{aligned}&\Big|\mathbb{E}\Big(\Big(1\left\{\varepsilon_{n,ij} \leq \min\left\{C^{10}_{n,ij} + \Delta_{n,ij}, C^*_{n,ij}\right\}\right\} - 1\left\{\varepsilon_{n,ij} \leq \min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right\}\Big) \\ &\quad \cdot \Big(1\left\{C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \Delta_{n,ik}\right\} - 1\left\{C^*_{n,ik} < \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right\}\Big)\Big| X_n, p_n\Big)\Big| \\ \leq&\mathbb{E}\Big(1\left\{\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\} - d^1_n \leq \varepsilon_{n,ij} \leq \min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\} + d^1_n\right\} \\ &\quad \cdot 1\left\{C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik} - d^2_n \leq \varepsilon_{n,ik} \leq C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik} + d^2_n\right\}\Big| X_n, p_n\Big) \\ =&\mathbb{E}\left(F_\varepsilon\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\} + d^1_n\right) - F_\varepsilon\left(\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\} - d^1_n\right)\Big| X_n, p_n\right) \\ &\quad \cdot \mathbb{E}\left(F_\varepsilon\left(C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik} + d^2_n\right) - F_\varepsilon\left(C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik} - d^2_n\right)\Big| X_n, p_n\right) \\ =&\mathbb{E}\left(2f_\varepsilon\left(\min\left\{C^{10}_{n,ij} + \tilde{\Delta}^1_{n,ij}, C^*_{n,ij}\right\}\right) d^1_n + o\left(d^1_n\right)\Big| X_n, p_n\right)\mathbb{E}\left(2f_\varepsilon\left(C^{10}_{n,ik} + \tilde{\Delta}^2_{n,ik}\right) d^2_n + o\left(d^2_n\right)\Big| X_n, p_n\right) \\ \leq&O\left(\frac{1}{n^2}\right)\end{aligned}$$

Like in Part I, the last terms in (8.70) and (8.71) are also $O\left(n^{-2}\right)$ and thus negligible.

Combining Steps 2.1-2.4, we have proved that

$$\begin{aligned}&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}\right\}\middle| X_n, p_n\right) \\ &- \mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij} \in \mathcal{E}_{n,ij}\right\}\middle| X_n, p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik} \notin \mathcal{E}_{n,ik}\right\}\middle| X_n, p_n\right) \\ &\leq O\left(\frac{1}{n}\right)\end{aligned}$$

Part III: $\varepsilon_{n,ij} \notin E_{n,ij}, \varepsilon_{n,ik} \in E_{n,ik}$.

The third terms in (8.57) and (8.58) are given by

$$\begin{aligned}&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij} \notin \mathcal{E}_{n,ij}, \varepsilon_{n,ik} \in \mathcal{E}_{n,ik}\right\}\middle| X_n, p_n\right) \\ &= \Pr\left(C^*_{n,ij} < \varepsilon_{n,ij} \leq C^{01}_{n,ij} + \Delta_{n,ij}, \varepsilon_{n,ik} \leq \min\left\{C^{01}_{n,ik} + \Delta_{n,ik}, C^*_{n,ik}\right\}\middle| X_n, p_n\right)\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij}\notin\mathcal{E}_{n,ij}\right\}\middle|\,X_n,p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik}\in\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\=&\Pr\left(C^{*}_{n,ij}<\varepsilon_{n,ij}\le C^{0\cdot}_{n,ij}+\Delta_{n,ij}\middle|\,X_n,p_n\right)\Pr\left(\varepsilon_{n,ik}\le\min\left\{C^{\cdot 1}_{n,ik}+\Delta_{n,ik},C^{*}_{n,ik}\right\}\middle|\,X_n,p_n\right)\end{aligned}$$

They are obtained from the terms for $\{\varepsilon_{n,ij}\in\mathcal{E}_{n,ij},\varepsilon_{n,ik}\notin\mathcal{E}_{n,ik}\}$ in Part II with j and k swapped, so from the proof in Part II the difference between them is also $O\left(n^{-1}\right)$.

$$\begin{aligned}&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij}\notin\mathcal{E}_{n,ij},\varepsilon_{n,ik}\in\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\&-\mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij}\notin\mathcal{E}_{n,ij}\right\}\middle|\,X_n,p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik}\in\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\&\le O\left(\frac{1}{n}\right)\end{aligned}$$

Part IV: $\varepsilon_{n,ij}\notin E_{n,ij},\varepsilon_{n,ik}\notin E_{n,ik}$.

By definition,

$$\begin{aligned}&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij}\notin\mathcal{E}_{n,ij},\varepsilon_{n,ik}\notin\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\=&\Pr\left(C^{*}_{n,ij}<\varepsilon_{n,ij}\le C^{00}_{n,ij}+\Delta_{n,ij},C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{00}_{n,ik}+\Delta_{n,ik}\middle|\,X_n,p_n\right)\end{aligned}$$

and

$$\begin{aligned}&\mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij}\notin\mathcal{E}_{n,ij}\right\}\middle|\,X_n,p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik}\notin\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\=&\Pr\left(C^{*}_{n,ij}<\varepsilon_{n,ij}\le C^{0\cdot}_{n,ij}+\Delta_{n,ij}\middle|\,X_n,p_n\right)\Pr\left(C^{*}_{n,ik}<\varepsilon_{n,ik}\le C^{\cdot 0}_{n,ik}+\Delta_{n,ik}\middle|\,X_n,p_n\right)\end{aligned}$$

With a slight modification to the proof in Part II, we can show

$$\begin{aligned}&\mathbb{E}\left(G_{n,ij}G_{n,ik}1\left\{\varepsilon_{n,ij}\notin\mathcal{E}_{n,ij},\varepsilon_{n,ik}\notin\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\&-\mathbb{E}\left(G_{n,ij}1\left\{\varepsilon_{n,ij}\notin\mathcal{E}_{n,ij}\right\}\middle|\,X_n,p_n\right)\mathbb{E}\left(G_{n,ik}1\left\{\varepsilon_{n,ik}\notin\mathcal{E}_{n,ik}\right\}\middle|\,X_n,p_n\right)\\&\le O\left(\frac{1}{n}\right)\end{aligned}$$

Combining Parts I-IV, we conclude that the conditional covariance of $G_{n,ij}$ and $G_{n,ik}$ given (X_n,p_n) is $O\left(n^{-1}\right)$

$$\mathbb{E}\left(G_{n,ij}G_{n,ik}\middle|\,X_n,p_n\right)-\mathbb{E}\left(G_{n,ij}\middle|\,X_n,p_n\right)\mathbb{E}\left(G_{n,ik}\middle|\,X_n,p_n\right)=O\left(\frac{1}{n}\right)$$

The proof is complete. ■

To prove Theorem 4.2 and Theorem 4.3, we first prove in the following lemma that the population objective function is differentiable.

Lemma 8.3 (Differentiability) *Suppose that Assumptions 1-3 are satisfied. Given* X_n *and* p_n, $\Psi_n(\theta, p)$ *defined in (4.10) is differentiable at* (θ_0, p_n) *with a derivative* $\nabla_{(\theta,p)}\Psi_n(\theta_0, p_n) = [\nabla_\theta \Psi_n(\theta_0, p_n), \nabla_p \Psi_n(\theta_0, p_n)]$ *that satisfies*

$$\left\|\Psi_n(\theta, p) - \Psi_n(\theta_0, p_n) - \nabla_{(\theta,p)}\Psi_n(\theta_0, p_n)((\theta, p) - (\theta_0, p_n))\right\| = o(\|(\theta, p) - (\theta_0, p_n)\|)$$

Proof. By the definition of Ψ_n

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

Since $W_{n,ij}(\theta, p)$ is a function of $P_{n,ij}(\theta, p)$, we start by proving the differentiability of $P_{n,ij}(\theta, p)$.

By definition

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

where $f_{\varepsilon_{n,i}}$ represents the density of $\varepsilon_{n,i}$. From the definition of $U_{n,ij}$ in (3.4), $U_{n,ij}(\theta, p)$ is differentiable in (θ, p). To establish the differentiability of $P_{n,ij}(\theta, p)$, we need to investigate how $\omega_{ni}(\varepsilon_{n,i}, \theta, p)$ would change in (θ, p).

From the first-order condition for ω in (8.9) in the proof for Lemma 8.1, $\omega_{ni}(\varepsilon_{n,i}, \theta, p)$ satisfies

$$\frac{1}{n-1} \sum_{j \neq i} 1\left\{U_{n,ij}(\theta, p) + \frac{n-1}{n-2} 2D_j' \omega_{ni}(\varepsilon_{n,i}, \theta, p) - \varepsilon_{n,ij} > 0\right\} D_j = V_i^+(\theta, p)\, \omega_{ni}(\varepsilon_{n,i}, \theta, p) \tag{8.73}$$

almost surely. The left-hand side of (8.73) as a function of (θ, p) has a derivative at (θ_0, p_n) that equals 0, except for the set

$$\left\{\varepsilon_{n,i} : \exists j, U_{n,ij}(\theta_0, p_n) + \frac{n-1}{n-2} 2D_j' \omega_{ni}(\varepsilon_{n,i}, \theta_0, p_n) = \varepsilon_{n,ij}\right\},$$

which occurs with probability 0 since

$$\Pr\left(\exists j, U_{n,ij}(\theta_0, p_n) + \frac{n-1}{n-2} 2D_j'\omega_{ni}(\varepsilon_{n,i}, \theta_0, p_n) = \varepsilon_{n,ij} \middle| X_n, p_n\right)$$
$$\leq \sum_{j\neq i} \Pr\left(U_{n,ij}(\theta_0, p_n) + \frac{n-1}{n-2} 2D_j'\omega_{ni}(\varepsilon_{n,i}, \theta_0, p_n) = \varepsilon_{n,ij} \middle| X_n, p_n\right) = 0.$$

The last equality follows because the right-hand side of (8.73) is a linear function of ω, and the left-hand side is a step function, so when $\varepsilon_{n,ij}$ changes the optimal ω_{ni} changes at most finite number of times which occurs with probability 0. Therefore, with probability 1 the right-hand side of (8.73) as a function of (θ, p) also has a derivative at (θ_0, p_n) that is equal to 0, i.e.,

$$\left|V_i^+(\theta, p)\,\omega_{ni}(\varepsilon_{n,i}, \theta, p) - V_i^+(\theta_0, p_n)\,\omega_{ni}(\varepsilon_{n,i}, \theta_0, p_n)\right| = o\left(\|(\theta, p) - (\theta_0, p_n)\|\right)$$

almost surely. Moreover, from the definition of ω in Section 3, we have $\omega = V_i\Phi_i\tilde{\omega}$, where Φ_i is the matrix of eigenvectors and $\tilde{\omega}$ is given in (3.3), so $\omega = V_iV_i^+V_i\Phi_i\tilde{\omega} = V_iV_i^+\omega$. Hence,

$$\begin{aligned}
\omega_{ni}(\varepsilon_{n,i}, \theta, p) &= V_i(\theta, p)\,V_i^+(\theta, p)\,\omega_{ni}(\varepsilon_{n,i}, \theta, p) \\
&= V_i(\theta, p)\,V_i^+(\theta_0, p_n)\,\omega_{ni}(\varepsilon_{n,i}, \theta_0, p_n) + o\left(\|(\theta, p) - (\theta_0, p_n)\|\right) \\
&= \omega_{ni}(\varepsilon_{n,i}, \theta_0, p_n) + \left(V_i(\theta, p) - V_i(\theta_0, p_n)\right)V_i^+(\theta_0, p_n)\,\omega_{ni}(\varepsilon_{n,i}, \theta_0, p_n) \\
&\quad + o\left(\|(\theta, p) - (\theta_0, p_n)\|\right) \qquad (8.74)
\end{aligned}$$

almost surely.

Now we construct the derivative of $P_{n,ij}(\theta, p)$ at (θ_0, p_n). Observe that $V_i(\theta, p)$ is differentiable in (θ, p). Define the augmented parameter $\tilde{\theta} = (\theta, p)$ and $\tilde{\theta}_0 = (\theta_0, p_n)$. Note that the dimension of $\tilde{\theta}$ is $\dim(\theta) + T^2$. For $k = 1, \ldots, \dim(\theta) + T^2$, let $\tilde{\theta}_k$ be the kth component of $\tilde{\theta}$ and define

$$\nabla_{\tilde{\theta}_k} V_i(\theta_0, p_n) = \begin{bmatrix} \frac{\partial V_{i,11}(\theta_0, p_n)}{\partial \tilde{\theta}_k} & \cdots & \frac{\partial V_{i,1T}(\theta_0, p_n)}{\partial \tilde{\theta}_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial V_{i,T1}(\theta_0, p_n)}{\partial \tilde{\theta}_k} & \cdots & \frac{\partial V_{i,TT}(\theta_0, p_n)}{\partial \tilde{\theta}_k} \end{bmatrix}$$

Furthermore, define

$$\nabla_{(\theta,p)} U_{n,ij}(\theta_0, p_n) = \left(\frac{\partial U_{n,ij}(\theta_0, p_n)}{\partial \theta'}, \frac{\partial U_{n,ij}(\theta_0, p_n)}{\partial p'}\right).$$

We define the vector $\nabla_{(\theta,p)}P_{n,ij}(\theta_0,p_n)$ by

$$\begin{aligned}
&\nabla_{(\theta,p)}P_{n,ij}(\theta_0,p_n)((\theta,p)-(\theta_0,p_n)) \\
&=\int\Bigg(\nabla_{(\theta,p)}U_{n,ij}(\theta_0,p_n)((\theta,p)-(\theta_0,p_n)) \\
&\qquad+\frac{n-1}{n-1}2{\textstyle\sum_k}D_j'\nabla_{\tilde{\theta}_k}V_i(\theta_0,p_n)V_i^{+}(\theta_0,p_n)\,\omega_{ni}(\varepsilon_{n,i},\theta_0,p_n)\left(\tilde{\theta}_k-\tilde{\theta}_{0k}\right)\Bigg) \\
&\quad\cdot f_{\varepsilon}\left(\varepsilon_{n,ij}:\varepsilon_{n,ij}=U_{n,ij}(\theta_0,p_n)+\frac{n-1}{n-1}2D_j'\omega_{ni}(\varepsilon_{n,i},\theta_0,p_n)\right)f_{\varepsilon_{n,-ij}}(\varepsilon_{n,-ij})\,d\varepsilon_{n,-ij},
\end{aligned} \tag{8.75}$$

where $f_{\varepsilon_{n,-ij}}$ represents the density of $\varepsilon_{n,-ij}=(\varepsilon_{n,ik})_{k\neq i,j}$ and the f_ε term in the last line is the density of $\varepsilon_{n,ij}$ evaluated at the value such that $\varepsilon_{n,ij}=U_{n,ij}(\theta_0,p_n)+\frac{n-1}{n-1}2D_j'\omega_{ni}(\varepsilon_{n,i},\theta_0,p_n)$ for a given $\varepsilon_{n,-ij}$. We show that the $\nabla_{(\theta,p)}P_{n,ij}(\theta_0,p_n)$ defined in (8.75) gives the derivative of $P_{n,ij}(\theta,p)$ at (θ_0,p_n) in the sense that

$$\left|P_{n,ij}(\theta,p)-P_{n,ij}(\theta_0,p_n)-\nabla_{(\theta,p)}P_{n,ij}(\theta_0,p_n)((\theta,p)-(\theta_0,p_n))\right|=o(\|(\theta,p)-(\theta_0,p_n)\|).$$

To see this, consider the difference between $P_{n,ij}(\theta,p)$ and $P_{n,ij}(\theta_0,p_n)$

$$\begin{aligned}
&P_{n,ij}(\theta,p)-P_{n,ij}(\theta_0,p_n) \\
&=\int\Bigg(1\left\{U_{n,ij}(\theta,p)+\frac{n-1}{n-1}2D_j'\omega_{ni}(\varepsilon_{n,i},\theta,p)\geq\varepsilon_{n,ij}\right\} \\
&\qquad-1\left\{U_{n,ij}(\theta_0,p_n)+\frac{n-1}{n-1}2D_j'\omega_{ni}(\varepsilon_{n,i},\theta_0,p_n)\geq\varepsilon_{n,ij}\right\}\Bigg)f_{\varepsilon_{n,i}}(\varepsilon_{n,i})\,d\varepsilon_{n,i} \\
&=\int\Delta_{n,ij}(\varepsilon_{n,i},\theta,p,\theta_0,p_n)\,\Delta\tilde{U}_{n,ij}(\varepsilon_{n,i},\theta,p)\,f_{\varepsilon_{n,i}}(\varepsilon_{n,i})\,d\varepsilon_{n,i}
\end{aligned}$$

where we define

$$\begin{aligned}
\tilde{U}_{n,ij}(\varepsilon_{n,i},\theta,p)&=U_{n,ij}(\theta,p)+\frac{n-1}{n-1}2D_j'\omega_{ni}(\varepsilon_{n,i},\theta,p) \\
\Delta\tilde{U}_{n,ij}(\varepsilon_{n,i},\theta,p)&=\tilde{U}_{n,ij}(\varepsilon_{n,i},\theta,p)-\tilde{U}_{n,ij}(\varepsilon_{n,i},\theta_0,p_n)
\end{aligned}$$

and

$$
\begin{aligned}
&\Delta_{n,ij}\left(\varepsilon_{n,i},\theta,p,\theta_0,p_n\right)\\
&=\frac{1}{\Delta\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta,p\right)}\left(1\left\{U_{n,ij}\left(\theta,p\right)+\frac{n-1}{n-1}2D_j'\omega_{ni}\left(\varepsilon_{n,i},\theta,p\right)\geq\varepsilon_{n,ij}\right\}\right.\\
&\qquad\qquad\left.-1\left\{U_{n,ij}\left(\theta_0,p_n\right)+\frac{n-1}{n-1}2D_j'\omega_{ni}\left(\varepsilon_{n,i},\theta_0,p_n\right)\geq\varepsilon_{n,ij}\right\}\right).
\end{aligned}
$$

By (8.74) and the differentiability of $U_{n,ij}\left(\theta,p\right)$ and $V_i\left(\theta,p\right)$, $\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta,p\right)$ has a derivative $\nabla_{(\theta,p)}\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)$ at $\left(\theta_0,p_n\right)$ given by

$$
\begin{aligned}
&\nabla_{(\theta,p)}\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\left(\left(\theta,p\right)-\left(\theta_0,p_n\right)\right)\\
=&\nabla_{(\theta,p)}U_{n,ij}\left(\theta_0,p_n\right)\left(\left(\theta,p\right)-\left(\theta_0,p_n\right)\right)\\
&+\frac{n-1}{n-1}2{\textstyle\sum_k}D_j'\nabla_{\tilde{\theta}_k}V_i\left(\theta_0,p_n\right)V_i^{+}\left(\theta_0,p_n\right)\omega_{ni}\left(\varepsilon_{n,i},\theta_0,p_n\right)\left(\tilde{\theta}_k-\tilde{\theta}_{0k}\right)
\end{aligned}
$$

in the sense that

$$
\left|\Delta\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta,p\right)-\nabla_{(\theta,p)}\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\left(\left(\theta,p\right)-\left(\theta_0,p_n\right)\right)\right|=o\left(\left\|\left(\theta,p\right)-\left(\theta_0,p_n\right)\right\|\right)\quad(8.76)
$$

almost surely. Hence, for the $\nabla_{(\theta,p)}P_{n,ij}\left(\theta_0,p_n\right)$ defined in (8.75) we have

$$
\begin{aligned}
&P_{n,ij}\left(\theta,p\right)-P_{n,ij}\left(\theta_0,p_n\right)-\nabla_{(\theta,p)}P_{n,ij}\left(\theta_0,p_n\right)\left(\left(\theta,p\right)-\left(\theta_0,p_n\right)\right)\\
=&\int\Delta_{n,ij}\left(\varepsilon_{n,i},\theta,p,\theta_0,p_n\right)\Delta\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta,p\right)f_{\varepsilon_{n,i}}\left(\varepsilon_{n,i}\right)d\varepsilon_{n,i}\\
&-\int\nabla_{(\theta,p)}\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\left(\left(\theta,p\right)-\left(\theta_0,p_n\right)\right)\cdot f_{\varepsilon}\left(\varepsilon_{n,ij}:\varepsilon_{n,ij}=\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\right)f_{\varepsilon_{n,-ij}}\left(\varepsilon_{n,-ij}\right)d\varepsilon_{n,i}\\
=&\int\Delta_{n,ij}\left(\varepsilon_{n,i},\theta,p,\theta_0,p_n\right)\left(\Delta\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta,p\right)-\nabla_{(\theta,p)}\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\left(\left(\theta,p\right)-\left(\theta_0,p_n\right)\right)\right)f_{\varepsilon_{n,i}}\left(\varepsilon_{n,i}\right)d\varepsilon_{n,i}\\
&+\int\left(\Delta_{n,ij}\left(\varepsilon_{n,i},\theta,p,\theta_0,p_n\right)f_{\varepsilon}\left(\varepsilon_{n,ij}\right)-f_{\varepsilon}\left(\varepsilon_{n,ij}:\varepsilon_{n,ij}=\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\right)\right)\\
&\qquad\cdot\nabla_{(\theta,p)}\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\left(\left(\theta,p\right)-\left(\theta_0,p_n\right)\right)f_{\varepsilon_{n,-ij}}\left(\varepsilon_{n,-ij}\right)d\varepsilon_{n,i}\qquad(8.77)
\end{aligned}
$$

where the last equality follows by adding and subtracting the same term. The first term on the last right-hand side of (8.77) is $o\left(\left\|\left(\theta,p\right)-\left(\theta_0,p_n\right)\right\|\right)$ by (8.76) and dominated convergence. It suffices to consider the last term in (8.77). For a given $\varepsilon_{n,-ij}$,

$$
\int\Delta_{n,ij}\left(\varepsilon_{n,i},\theta,p,\theta_0,p_n\right)f_{\varepsilon}\left(\varepsilon_{n,ij}\right)d\varepsilon_{n,ij}-f_{\varepsilon}\left(\varepsilon_{n,ij}:\varepsilon_{n,ij}=\tilde{U}_{n,ij}\left(\varepsilon_{n,i},\theta_0,p_n\right)\right)\to 0\qquad(8.78)
$$

as $(\theta, p) \to (\theta_0, p_n)$ and thus $\Delta\tilde{U}_{n,ij}(\varepsilon_{n,i}, \theta, p) \to 0$ by the Lebesgue Differentiation Theorem and the fact that $\omega_{ni}(\varepsilon_{n,i}, \theta, p)$ does not change in $\varepsilon_{n,ij}$ except for a finite number of points. Moreover, the derivative $\nabla_{(\theta,p)}\tilde{U}_{n,ij}(\varepsilon_{n,i}, \theta_0, p_n)$ is Lipschitz continuous in (θ, p), i.e., there is a constant $C > 0$,

$$\left|\nabla_{(\theta,p)}\tilde{U}_{n,ij}(\varepsilon_{n,i}, \theta_0, p_n)((\theta, p) - (\theta_0, p_n))\right| \leq C \|(\theta, p) - (\theta_0, p_n)\|. \tag{8.79}$$

Hence by (8.78), (8.79) and dominated convergence we derive that the last term in (8.77) is also $o(\|(\theta, p) - (\theta_0, p_n)\|)$.

We have proved that

$$\left|P_{n,ij}(\theta, p) - P_{n,ij}(\theta_0, p_n) - \nabla_{(\theta,p)}P_{n,ij}(\theta_0, p_n)((\theta, p) - (\theta_0, p_n))\right| = o(\|(\theta, p) - (\theta_0, p_n)\|)$$

so the $\nabla_{(\theta,p)}P_{n,ij}(\theta_0, p_n)$ defined in (8.75) is the desired derivative for $P_{n,ij}(\theta, p)$ at (θ_0, p_n). ■

Proof of Theorem 4.2. By Lemma 8.3 and Assumption, Ψ_n has a continuously invertible derivative at (θ_0, p_n), so there exists a constant $c > 0$ such that

$$\|\Psi_n(\theta, p) - \Psi_n(\theta_0, p_n)\| > c\|(\theta, p) - (\theta_0, p_n)\| + o(\|(\theta, p) - (\theta_0, p_n)\|) \tag{8.80}$$

for every (θ, p). Similar to the consistency proof for ω_{ni} in Lemma 8.1, inequality (8.80) provides an identification condition for (θ_0, p_n): for any $\xi > 0$, if $\|(\theta, p) - (\theta_0, p_n)\| > \xi$, then $\|\Psi_n(\theta, p) - \Psi_n(\theta_0, p_n)\| > (c + o(1))\xi$. Let $\eta > 0$ satisfy $\eta \leq (c + o(1))\xi$. Then,

$$\begin{aligned}
&\Pr\left(\left\|\left(\hat{\theta}_n, \hat{p}_n\right) - (\theta_0, p_n)\right\| > \xi \middle| X_n, p_n\right) \\
&\leq \Pr\left(\left\|\Psi_n\left(\hat{\theta}_n, \hat{p}_n\right) - \Psi_n(\theta_0, p_n)\right\| > \eta \middle| X_n, p_n\right) \\
&= \Pr\left(\left\|\Psi_n\left(\hat{\theta}_n, \hat{p}_n\right) - \hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right)\right\| > \eta \middle| X_n, p_n\right) \\
&\leq \Pr\left(\sup_{(\theta,p)}\left\|\hat{\Psi}_n(\theta, p) - \Psi_n(\theta, p)\right\| > \eta \middle| X_n, p_n\right)
\end{aligned} \tag{8.81}$$

where the equality follows because $\Psi_n(\theta_0, p_n) = 0$ and $\hat{\Psi}_n\left(\hat{\theta}_n, \hat{p}_n\right) = 0$. By the definition of $\hat{\Psi}_n$ and Ψ_n

$$\begin{aligned}
\hat{\Psi}_n(\theta, p) - \Psi_n(\theta, p) &= \frac{1}{n(n-1)} \sum_i \sum_{j \neq i} W_{n,ij}(\theta, p)\left(G_{n,ij} - \mathbb{E}\left(G_{n,ij} \middle| X_n, p_n\right)\right) \\
&= \frac{1}{n} \sum_i Y_{ni}(\theta, p) - \mathbb{E}\left(Y_{ni}(\theta, p) \middle| X_n, p_n\right)
\end{aligned}$$

where

$$Y_{ni}(\theta, p) = \frac{1}{n-1} \sum_{j \neq i} W_{n,ij}(\theta, p)\, G_{n,ij}$$

By assumption the weight $W_{n,ij}(\theta, p)$ is bounded by an envelope $\|W_{n,ij}(\theta, p)\| \leq M_n < \infty$ for all i and j and all (θ, p), so $\|Y_{ni}(\theta, p)\| \leq M_n$. Moreover, by the contruction of $W_{n,ij}(\theta, p)$ and Lemma 8.3 $W_{n,ij}(\theta, p)$ is differentiable with an derivative $\nabla_{(\theta,p)} W_{n,ij}(\theta, p)$ which is bounded by $C_n = \max_{i,j} \sup_{(\theta,p)} \left\|\nabla_{(\theta,p)} W_{n,ij}(\theta, p)\right\| < \infty$, so $Y_{ni}(\theta, p)$ is Lipschitz continuous in (θ, p) with a Lipschitz constant C_n

$$\left\|Y_{ni}(\theta, p) - Y_{ni}\left(\tilde{\theta}, \tilde{p}\right)\right\| \leq C_n \left\|(\theta, p) - \left(\tilde{\theta}, \tilde{p}\right)\right\| \tag{8.82}$$

Define the array $Y_n(\theta, p) = \left(Y_{ni}(\theta, p)\right)_{1 \leq i \leq n}$ and the set $\mathcal{F} = \left\{Y_n(\theta, p) : (\theta, p) \in \mathbb{R}^{d_\theta + T^2}\right\}$. Let $\alpha \in \mathbb{R}^n$ be an arbitrary $n \times 1$ vector of nonnegative constants. Consider the set of arrays

$$\alpha \odot \mathcal{F} = \left\{\alpha \odot Y_n(\theta, p) : (\theta, p) \in \mathbb{R}^{d_\theta + T^2}\right\}$$

where $\alpha \odot Y_{ni}$ is the pointwise product of α and Y_n. The arrays in $\alpha \odot \mathcal{F}$ have the envelope $\|\alpha\| M_n$. Further, the Lipschitz condition in (8.82) implies that for any (θ, p) and $\left(\tilde{\theta}, \tilde{p}\right)$

$$\begin{aligned}
\left\|\alpha \odot Y_n(\theta, p) - \alpha \odot Y_n\left(\tilde{\theta}, \tilde{p}\right)\right\| &= \left(\sum_{i=1}^n \alpha_i^2 \left\|Y_{ni}(\theta, p) - Y_{ni}\left(\tilde{\theta}, \tilde{p}\right)\right\|^2\right)^{1/2} \\
&\leq \left(\sum_{i=1}^n \alpha_i^2 C_n^2 \left\|(\theta, p) - \left(\tilde{\theta}, \tilde{p}\right)\right\|^2\right)^{1/2} \\
&= C_n \|\alpha\| \left\|(\theta, p) - \left(\tilde{\theta}, \tilde{p}\right)\right\|
\end{aligned}$$

Therefore, the packing numbers of the sets $\alpha \odot \mathcal{F}$ and $\Theta \times \mathcal{P} = \left\{(\theta, p) \in \mathbb{R}^{d_\theta + T^2}\right\}$ satisfy the relationship

$$D\left(\zeta \|\alpha\| M_n, \alpha \odot \mathcal{F}, \|\cdot\|\right) \leq D\left(\zeta \frac{M_n}{C_n}, \Theta \times \mathcal{P}, \|\cdot\|\right)$$

for $\zeta > 0$. It is clear that the set $\Theta \times \mathcal{P}$ has a pseudodimension of at most $d_\theta + T^2$, so by Corollary 4.10 and Definition 7.9 in Pollard (1990) the set $\mathcal{F}$ is manageable. Applying the uniform law of large number in Theorem 8.3 of Pollard (1990) yields

$$\Pr\left(\sup_{(\theta,p)} \left\|\hat{\Psi}_n(\theta,p) - \Psi_n(\theta,p)\right\| > \eta \middle| X_n, p_n\right) \to 0$$

as $n \to \infty$. Combining this with (8.81) proves

$$\left(\hat{\theta}_n, \hat{p}_n\right) - (\theta_0, p_n) = o_p(1)$$

The proof is complete. ■

To prove Theorem 4.3, we also need the following lemma.

Lemma 8.4 (CLT for Sample Moments) *Suppose that Assumptions 1-3 are satisfied. For the augmented sample moment function $\tilde{\Psi}_n(\theta_0, p_n)$ defined in (4.12) and the conditional variance $\tilde{\Omega}_n(X_n, p_n)$ defined in (4.14). we have*

$$\sqrt{n(n-1)}\tilde{\Omega}_n^{-1/2}(X_n, p_n)\tilde{\Psi}_n(\theta_0, p_n) \xrightarrow{d} N(0, I) \tag{8.83}$$

as $n \to \infty$. Moreover, $\tilde{\Omega}_n(X_n, p_n) = O(1)$, so $\tilde{\Psi}_n(\theta_0, p_n)$ converges at the order of n^{-1}.

Proof. Define

$$\sum_i Y_{ni} = \sqrt{n(n-1)}\tilde{\Omega}_n^{-1/2}(X_n, p_n)\tilde{\Psi}_n(\theta_0, p_n)$$

where Y_{ni} is given by

$$Y_{ni} = \frac{1}{\sqrt{n(n-1)}}\tilde{\Omega}_n^{-1/2}(X_n, p_n)\sum_{j\neq i}\tilde{W}_{n,ij}(\theta_0, p_n)(G_{n,ij} - P_{n,ij}(\theta_0, p_n))$$

with the augmented weights $\tilde{W}_{n,ij}(\theta_0, p_n)$ defined in (4.13). Note that conditional on (X_n, p_n), $Y_{n1}, \ldots, Y_{nn}$ form an independent triangular array because each Y_{ni} involves $\varepsilon_{n,i}$ only (though $G_{n,i}$), and $\varepsilon_{n,i}$ is i.i.d. by Assumption 1. We derive the asymptotic distribution of $\sum_i Y_{ni}$ by applying the Lindeberg-Feller Central Limit Theorem (CLT). The rest of the the proof is to check whether the conditions in the Lindeberg-Feller CLT hold.

Since Y_{ni} is a $d_\theta \times 1$ vector, where d_θ is the dimension of θ, we use the Cramer-Wold method. Take any vector $a \in \mathbb{R}^{d_\theta}$. We verify the conditions in the Lindeberg-Feller CLT for

$a'Y_{ni}$. First, given X_n and p_n the conditional mean of $a'Y_{ni}$ is 0

$$\mathbb{E}\left(a'Y_{ni} \middle| X_n, p_n \right) = 0$$

Second, the conditional variance of $\sum_i a'Y_{ni}$ given X_n and p_n is

$$\begin{aligned} & Var\left(\sum_i a'Y_{ni} \middle| X_n, p_n \right) \\ & = a'\tilde{\Omega}_n^{-1/2}\left(X_n, p_n\right) Var\left(\sqrt{n\left(n-1\right)}\tilde{\Psi}_n\left(\theta_0, p_n\right) \middle| X_n, p_n \right) \tilde{\Omega}_n^{-1/2}\left(X_n, p_n\right) a \\ & = a'Ia = \left\|a\right\|^2 \end{aligned}$$

Next we verify the Lindeberg condition. It requires that for every $\xi > 0$

$$\lim_{n\to\infty} \frac{1}{\left\|a\right\|^2} \sum_i \mathbb{E}\left(\left(a'Y_{ni}\right)^2 1\left\{\left|a'Y_{ni}\right| \geq \xi \left\|a\right\|\right\} \middle| X_n, p_n \right) = 0 \tag{8.84}$$

Since

$$\begin{aligned} & \frac{1}{\left\|a\right\|^2} \sum_i \mathbb{E}\left(\left(a'Y_{ni}\right)^2 1\left\{\left|a'Y_{ni}\right| \geq \xi \left\|a\right\|\right\} \middle| X_n, p_n \right) \\ & = \frac{1}{\left\|a\right\|^2}\mathbb{E}\left(\sum_i a'Y_{ni}Y_{ni}'a1\left\{\left|a'Y_{ni}\right| \geq \xi \left\|a\right\|\right\} \middle| X_n, p_n \right) \\ & \leq \frac{1}{\left\|a\right\|^2}\mathbb{E}\left(\sum_i a'Y_{ni}Y_{ni}'a1\left\{ \frac{\max_{1\leq i\leq n}\left|a'Y_{ni}\right|}{\left\|a\right\|} \geq \xi \right\} \middle| X_n, p_n \right) \end{aligned}$$

it suffices to show

$$\frac{\max_{1\leq i\leq n}\left|a'Y_{ni}\right|}{\left\|a\right\|} \xrightarrow{p} 0 \tag{8.85}$$

as $n \to \infty$, because

$$\frac{1}{\left\|a\right\|^2}\mathbb{E}\left(\sum_i a'Y_{ni}Y_{ni}'a1\left\{ \frac{\max_{1\leq i\leq n}\left|a'Y_{ni}\right|}{\left\|a\right\|} \geq \xi \right\} \middle| X_n, p_n \right) \leq \frac{1}{\left\|a\right\|^2}\mathbb{E}\left(\sum_i a'Y_{ni}Y_{ni}'a \middle| X_n, p_n \right) = 1$$

so the Lindeberg condition in (8.84) follows by the dominated convergence theorem.

To show (8.85), by Chebyshev's inequality

$$\Pr\left(\frac{\max_{1\leq i\leq n}\left|a'Y_{ni}\right|}{\left\|a\right\|} \geq \xi \middle| X_n, p_n \right) \leq \frac{1}{\xi^2\left\|a\right\|^2}\mathbb{E}\left(\max_{1\leq i\leq n} a'Y_{ni}Y_{ni}'a \middle| X_n, p_n \right) \tag{8.86}$$

Define the function $\psi(z) = e^z - 1$, and $\|Z\|_\psi$ to be the Orlicz norm of random variable Z conditional on (X_n, p_n), i.e., $\|Z\|_\psi = \inf\left\{C > 0 : \mathbb{E}\left(\psi\left(\frac{|Z|}{C}\right)\middle| X_n, p_n\right) \leq 1\right\}$. Since $z \leq e^z - 1$, it is easy to show $\mathbb{E}\left(|Z|\middle| X_n, p_n\right) \leq \|Z\|_\psi$, so

$$\mathbb{E}\left(\max_{1\leq i\leq n} a'Y_{ni}Y_{ni}'a\middle| X_n, p_n\right) \leq \left\|\max_{1\leq i\leq n} a'Y_{ni}Y_{ni}'a\right\|_\psi \tag{8.87}$$

The right-hand side of (8.87) can be further bounded using the maximal inequality in Lemma 2.2.2 of Van der Vaart and Wellner (1996),

$$\left\|\max_{1\leq i\leq n} a'Y_{ni}Y_{ni}'a\right\|_\psi \leq K\log(1+n)\max_{1\leq i\leq n}\|a'Y_{ni}Y_{ni}'a\|_\psi \tag{8.88}$$

where K is a constant depending only on ψ. We will show that $\max_{1\leq i\leq n}\|a'Y_{ni}Y_{ni}'a\|_\psi = O(n^{-1})$. By the definition of Y_{ni},

$$\begin{aligned}
&a'Y_{ni}Y_{ni}'a\\
=&\frac{1}{n(n-1)}a'\tilde{\Omega}_n^{-1/2}(\theta_0, p_n)\left(\sum_{j\neq i}\tilde{W}_{n,ij}(\theta_0, p_n)\tilde{W}_{n,ij}'(\theta_0, p_n)(G_{n,ij} - P_{n,ij}(\theta_0, p_n))^2\right.\\
&\left.+\sum_{j\neq i}\sum_{k\neq i,j}\tilde{W}_{n,ij}(\theta_0, p_n)\tilde{W}_{n,ik}'(\theta_0, p_n)(G_{n,ij} - P_{n,ij}(\theta_0, p_n))(G_{n,ik} - P_{n,ik}(\theta_0, p_n))\right)\tilde{\Omega}_n^{-1/2}(X_n, p_n)a
\end{aligned}$$

We will show later in this proof that $\left\|\tilde{W}_{n,ij}(\theta_0, p_n)\right\| < \infty$ and $\tilde{\Omega}_n(X_n, p_n) = O(1)$, so for any i, $a'Y_{ni}Y_{ni}'a$ is bounded. For any bounded random variable Z, we can show that $\|Z\|_\psi \lesssim \mathbb{E}\left(|Z|\middle| X_n, p_n\right)$.[6] Moreover $\mathbb{E}\left(a'Y_{ni}Y_{ni}'a\middle| X_n, p_n\right) = O(n^{-1})$ because $\left\|\tilde{W}_{n,ij}(\theta_0, p_n)\right\| < \infty$, $\tilde{\Omega}_n(X_n, p_n) = O(1)$ and $\left|\mathbb{E}\left((G_{n,ij} - P_{n,ij}(\theta_0, p_n))(G_{n,ik} - P_{n,ik}(\theta_0, p_n))\middle| X_n, p_n\right)\right| = O(n^{-1})$ for $j \neq k$ due to Proposition 4.1. Therefore,

$$\|a'Y_{ni}Y_{ni}'a\|_\psi \lesssim \mathbb{E}\left(a'Y_{ni}Y_{ni}'a\middle| X_n, p_n\right) = O\left(\frac{1}{n}\right)$$

[6] Suppose random variable $|Z| \leq M$ for constant $M < \infty$. Since $\psi(z) = e^z - 1$ is convex in z, by Jensen's inequality $e^{|Z|} - 1 \leq \frac{|Z|}{M}\left(e^M - 1\right)$. Therefore

$$\mathbb{E}\left(e^{|Z|/C} - 1\middle| X_n, p_n\right) \leq \frac{\left(e^M - 1\right)}{M}\frac{\mathbb{E}\left(|Z|\middle| X_n, p_n\right)}{C}$$

Choose $C = \frac{\left(e^M-1\right)}{M}\mathbb{E}\left(|Z|\middle| X_n, p_n\right)$ so the right-hand side is 1. By the definition of the Orlicz norm, $\|Z\|_\psi \leq C = \frac{\left(e^M-1\right)}{M}\mathbb{E}\left(|Z|\middle| X_n, p_n\right)$.

for all i. This implies that

$$\max_{1\leq i\leq n} \left\| a'Y_{ni}Y_{ni}'a \right\|_{\psi} = O\left(\frac{1}{n}\right) \tag{8.89}$$

Combining (8.86)-(8.89) we obtain

$$\begin{aligned}
\Pr\left(\frac{\max_{1\leq i\leq n} |a'Y_{ni}|}{\|a\|} \geq \xi \middle| X_n, p_n \right) &\leq \frac{K \log(1+n)}{\xi^2 \|a\|^2} \max_{1\leq i\leq n} \left\| a'Y_{ni}Y_{ni}'a \right\|_{\psi} \\
&= \frac{K \log(1+n)}{\xi^2 \|a\|^2} O\left(\frac{1}{n}\right) \\
&\to 0
\end{aligned}$$

as $n \to \infty$. Condition (8.85) and thus the Lindeberg condition (8.84) are proved.

We have verified that the conditions for the Lindeberg-Feller CLT are satisfied, so applying the theorem we obtain $\frac{1}{\|a\|^2} \sum_i a'Y_{ni} \xrightarrow{d} N(0,1)$, or equivalently,

$$a' \sum_i Y_{ni} \xrightarrow{d} N(0, a'Ia)$$

as $n \to \infty$. This implies that

$$\sum_i Y_{ni} \xrightarrow{d} N(0, I)$$

as $n \to \infty$, thereby proving (8.83).

Finally, we show $\left\| \tilde{W}_{n,ij}(\theta_0, p_n) \right\| < \infty$ and $\tilde{\Omega}_n(X_n, p_n) = O(1)$. Recall that the augmented weights $\tilde{W}_{n,ij}(\theta_0, p_n)$ are defined as

$$\tilde{W}_{n,ij}(\theta, p) = W_{n,ij}(\theta_0, p_n) + \nabla_p \Psi_n(\theta_0, p_n) Q_{n,ij}$$

We choose weights $W_{n,ij}(\theta_0, p_n)$ in the second step to be bounded, i.e., $\|W_{n,ij}(\theta_0, p_n)\| < \infty$. Moreover, the estimation of p_n in the first step lead to extra weights given by

$$
\begin{aligned}
&\nabla_p \Psi_n(\theta_0, p_n) Q_{n,ij} \\
=&\frac{1}{n(n-1)} \sum_k \sum_{l \neq k} \left. \frac{\partial W_{n,kl}(\theta, p)\left(\mathbb{E}\left[G_{n,kl} \mid X_n, p_n\right] - P_{n,kl}(\theta, p)\right)}{\partial p'} \right|_{(\theta,p)=(\theta_0,p_n)} Q_{n,ij} \\
=&-\frac{1}{n(n-1)} \sum_k \sum_{l \neq k} W_{n,kl}(\theta_0, p_n) \frac{\partial P_{n,kl}(\theta_0, p_n)}{\partial p'} Q_{n,ij} \\
=&-\frac{1}{n(n-1)} \sum_k \sum_{l \neq k} W_{n,kl}(\theta_0, p_n) \sum_{s,t} \frac{\partial P_{n,kl}(\theta_0, p_n)}{\partial p^{st}} \frac{1\left\{X_{n,ij} = x^{st}\right\}}{(n(n-1))^{-1} \sum_{k'} \sum_{l' \neq k'} 1\left\{X_{n,k'l'} = x^{st}\right\}} \\
=&-\sum_{s,t} \frac{\sum_k \sum_{l \neq k} W_{n,kl}(\theta_0, p_n) \frac{\partial P_{n,kl}(\theta_0,p_n)}{\partial p^{st}}}{\sum_k \sum_{l \neq k} 1\left\{X_{n,kl} = x^{st}\right\}} 1\left\{X_{n,ij} = x^{st}\right\}
\end{aligned}
$$

Under the assumption, for any $s, t \leq T$

$$
\frac{1}{n(n-1)} \sum_k \sum_{l \neq k} 1\left\{X_{n,kl} = x^{st}\right\} \to q^{st}
$$

for $q^{st} > 0$ as $n \to \infty$. Hence

$$
\left\| \frac{\sum_k \sum_{l \neq k} W_{n,kl}(\theta_0, p_n) \frac{\partial P_{n,kl}(\theta_0,p_n)}{\partial p^{st}}}{\sum_k \sum_{l \neq k} 1\left\{X_{n,kl} = x^{st}\right\}} \right\| < \infty
$$

and thus

$$
\left\| \nabla_p \Psi_n(\theta_0, p_n) Q_{n,ij} \right\| < \infty
$$

This proves

$$
\left\| \tilde{W}_{n,ij}(\theta, p)(\theta_0, p_n) \right\| < \infty \tag{8.90}
$$

Note that both the first-step and second-step weights have the same order $O(1)$, so both of them will contribute to the conditional variance $\tilde{\Omega}_n(X_n, p_n)$.

To show the rate of $\tilde{\Omega}_n(X_n, p_n)$, recall that

$$
\begin{aligned}
&\tilde{\Omega}_n(X_n, p_n) \\
=&Var\left(\sqrt{n(n-1)}\tilde{\Psi}_n(\theta_0, p_n) \middle| X_n, p_n \right) \\
=&\frac{1}{n(n-1)}\left(\sum_i \sum_{j\neq i} \tilde{W}_{n,ij}(\theta_0, p_n) \tilde{W}'_{n,ij}(\theta_0, p_n) \mathbb{E}\left((G_{n,ij} - P_{n,ij}(\theta_0, p_n))^2 | X_n, p_n \right) \right. \\
&\left. + \sum_i \sum_{j\neq i} \sum_{k\neq i,j} \tilde{W}_{n,ij}(\theta_0, p_n) \tilde{W}'_{n,ik}(\theta_0, p_n) \mathbb{E}\left((G_{n,ij} - P_{n,ij}(\theta_0, p_n)) (G_{n,ik} - P_{n,ik}(\theta_0, p_n)) \middle| X_n, p_n \right) \right)
\end{aligned}
\tag{8.91}
$$

The double summation in (8.91) consists of the conditional variances of each link. It satisfies

$$
\begin{aligned}
&\frac{1}{n(n-1)} \sum_i \sum_{j\neq i} \tilde{W}_{n,ij}(\theta_0, p_n) \tilde{W}'_{n,ij}(\theta_0, p_n) \mathbb{E}\left((G_{n,ij} - P_{n,ij}(\theta_0, p_n))^2 | X_n, p_n \right) \\
=&\frac{1}{n(n-1)} \sum_i \sum_{j\neq i} \tilde{W}_{n,ij}(\theta_0, p_n) \tilde{W}'_{n,ij}(\theta_0, p_n) P_{n,ij}(\theta_0, p_n) (1 - P_{n,ij}(\theta_0, p_n)) \\
\leq& \max_{i,j} W_{n,ij}(\theta_0, p_n) W'_{n,ij}(\theta_0, p_n) = O(1)
\end{aligned}
\tag{8.92}
$$

where the last inequality follows because $\left\| \tilde{W}_{n,ij}(\theta_0, p_n) \right\| < \infty$. Next we look at the triple summation in (8.91). Since $\left| \mathbb{E}\left((G_{n,ij} - P_{n,ij}(\theta_0, p_n)) (G_{n,ik} - P_{n,ik}(\theta_0, p_n)) \middle| X_n, p_n \right) \right| \leq O(n^{-1})$ for $j \neq k$ by Proposition 4.1, the triple summation satisfies

$$
\begin{aligned}
&\frac{1}{n(n-1)} \sum_i \sum_{j\neq i} \sum_{k\neq i,j} \tilde{W}_{n,ij}(\theta_0, p_n) \tilde{W}'_{n,ik}(\theta_0, p_n) \mathbb{E}\left((G_{n,ij} - P_{n,ij}) (G_{n,ik} - P_{n,ik}) \middle| X_n, p_n \right) \\
\leq& \frac{n(n-1)(n-2)}{n(n-1)} O\left(\frac{1}{n}\right) \max_{i,j,k} \tilde{W}_{n,ij}(\theta_0, p_n) \tilde{W}'_{n,ik}(\theta_0, p_n) = O(1)
\end{aligned}
\tag{8.93}
$$

Combining (8.92)-(8.93) yields

$$
\tilde{\Omega}_n(X_n, p_n) = O(1) \tag{8.94}
$$

The proof is complete. ■

Proof of Theorem 4.3. By the definition of $\hat{\theta}_n$ and $\Psi_n(\theta_0, p_n) = 0$

$$\begin{aligned}
&\sqrt{n(n-1)}\left(\Psi_n(\hat{\theta}_n, \hat{p}_n) - \Psi_n(\theta_0, p_n)\right) \\
&= \sqrt{n(n-1)}\left(\Psi_n(\hat{\theta}_n, \hat{p}_n) - \hat{\Psi}_n(\hat{\theta}_n, \hat{p}_n)\right) \\
&= -\sqrt{n(n-1)}\hat{\Psi}_n(\theta_0, p_n) - \sqrt{n(n-1)}\left((\hat{\Psi}_n - \Psi_n)(\hat{\theta}_n, \hat{p}_n) - (\hat{\Psi}_n - \Psi_n)(\theta_0, p_n)\right) \quad (8.95)
\end{aligned}$$

Since $\nabla_{(\theta,p)}\Psi_n(\theta_0, p_n)$ is continuously invertible by Lemma 8.3 and Assumption, there exists a constant $c > 0$ such that $\left\|\nabla_{(\theta,p)}\Psi_n(\theta_0, p_n)\left((\theta, p) - (\theta_0, p_n)\right)\right\| \geq c\left\|(\theta, p) - (\theta_0, p_n)\right\|$ for every (θ, p). Combining this with the differentiability of Ψ_n yields

$$\left\|\Psi_n(\theta, p) - \Psi_n(\theta_0, p_n)\right\| \geq c\left\|(\theta, p) - (\theta_0, p_n)\right\| + o\left(\left\|(\theta, p) - (\theta_0, p_n)\right\|\right) \quad (8.96)$$

Next we consider the last two terms in (8.95). Using the same proof in Lemma 8.4, with $\tilde{W}_{n,ij}(\theta_0, p_n)$ replaced by $W_{n,ij}(\theta_0, p_n)$ and $\tilde{\Omega}_n(X_n, p_n)$ replaced by

$$\Omega_n(X_n, p_n) = Var\left(\sqrt{n(n-1)}\hat{\Psi}_n(\theta_0, p_n)\Big| X_n, p_n\right)$$

we can show that

$$\sqrt{n(n-1)}\Omega_n^{-1/2}(X_n, p_n)\hat{\Psi}_n(\theta_0, p_n) \xrightarrow{d} N(0, I)$$

as $n \to \infty$ and $\Omega_n(X_n, p_n) = O(1)$, so the second last term in (8.95) satisfies

$$\sqrt{n(n-1)}\hat{\Psi}_n(\theta_0, p_n) = O_p(1) \quad (8.97)$$

As for the last term in (8.95), let $\mathbb{G}_n(\theta, p)$ be the empirical process

$$\mathbb{G}_n(\theta, p) = \sqrt{n(n-1)}(\hat{\Psi}_n - \Psi_n)(\theta, p)$$

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

$$\begin{aligned}
\mathbb{G}_n\left(\tilde{\theta}, \tilde{p}\right) - \mathbb{G}_n(\theta, p) &= \frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}\left(W_{n,ij}\left(\tilde{\theta}, \tilde{p}\right) - W_{n,ij}(\theta, p)\right)\left(G_{n,ij} - P_{n,ij}(\theta_0, p_n)\right) \\
&= \frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}\frac{\partial W_{n,ij}(\theta, p)}{\partial(\theta, p)'}\left(G_{n,ij} - P_{n,ij}(\theta_0, p_n)\right)\left(\left(\tilde{\theta}, \tilde{p}\right) - (\theta, p)\right) \\
&+ \frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}\left(G_{n,ij} - P_{n,ij}(\theta_0, p_n)\right) o\left(\left\|\left(\tilde{\theta}, \tilde{p}\right) - (\theta, p)\right\|\right)
\end{aligned}$$

(8.98)

where the second equality follows from the differentiability of $W_{n,ij}(\theta,p)$ due to the definition of $W_{n,ij}(\theta,p)$ and Lemma 8.3. Using the proof in Lemma 8.4 again twice, once for

$$\frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}\frac{\partial W_{n,ij}(\theta,p)}{\partial(\theta,p)'}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right) \tag{8.99}$$

with $\tilde{W}_{n,ij}(\theta_0,p_n)$ replaced by $\frac{\partial W_{n,ij}(\theta,p)}{\partial(\theta,p)'}$ and $\tilde{\Omega}_n(X_n,p_n)$ replaced by the conditional variance of (8.99) given X_n and p_n

$$Var\left(\frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}\frac{\partial W_{n,ij}(\theta,p)}{\partial(\theta,p)'}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right)\middle| X_n,p_n\right)$$

and once for

$$\frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right) \tag{8.100}$$

with $\tilde{W}_{n,ij}(\theta_0,p_n)$ replaced by 1 and $\tilde{\Omega}_n(X_n,p_n)$ replaced by the conditional variance of (8.100) given X_n and p_n

$$Var\left(\frac{1}{\sqrt{n(n-1)}}\sum_i\sum_{j\neq i}\left(G_{n,ij}-P_{n,ij}(\theta_0,p_n)\right)\middle| X_n,p_n\right)$$

we can show that both (8.99) and (8.100) are $O_p(1)$. Applying this to (8.98) yields

$$\mathbb{G}_n\left(\tilde{\theta},\tilde{p}\right)-\mathbb{G}_n(\theta,p)=O_p\left(\left\|\left(\tilde{\theta},\tilde{p}\right)-(\theta,p)\right\|\right)$$

for every $\left(\tilde{\theta},\tilde{p}\right)$ and (θ,p). Therefore, the last term in (8.95) satisfies

$$\sqrt{n(n-1)}\left((\hat{\Psi}_n-\Psi_n)(\hat{\theta}_n,\hat{p}_n)-(\hat{\Psi}_n-\Psi_n)(\theta_0,p_n)\right)=o_p\left(\sqrt{n(n-1)}\left\|(\hat{\theta}_n,\hat{p}_n)-(\theta_0,p_n)\right\|\right)$$

Combining (8.96)-(8.98) and the consistency of $(\hat{\theta}_n,\hat{p}_n)$ by Theorem 4.2 to (8.95) we obtain

$$\sqrt{n(n-1)}\left\|(\hat{\theta}_n,\hat{p}_n)-(\theta_0,p_n)\right\|(c+o_p(1))\leq O_p(1)+o_p\left(\sqrt{n(n-1)}\left\|(\hat{\theta}_n,\hat{p}_n)-(\theta_0,p_n)\right\|\right)$$

This implies that

$$\sqrt{n(n-1)}\left\|(\hat{\theta}_n,\hat{p}_n)-(\theta_0,p_n)\right\|\leq O_p(1)$$

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

To show the asymptotic distribution of $(\hat{\theta}_n, \hat{p}_n)$, by the differentiability of Ψ_n, the first line of (8.95) can be replaced by

$$\sqrt{n(n-1)}\left(\nabla_\theta \Psi_n(\theta_0, p_n)(\hat{\theta}_n - \theta_0) + \nabla_p \Psi_n(\theta_0, p_n)(\hat{p}_n - p_n)\right) + o_p\left(\sqrt{n(n-1)}\left\|(\hat{\theta}_n, \hat{p}_n) - (\theta_0, p_n)\right\|\right)$$

where the last term is $o_p(1)$ as also is the right-hand side of (8.98). Combining this with (8.95) yields

$$\begin{aligned}\sqrt{n(n-1)}\nabla_\theta \Psi_n(\theta_0, p_n)(\hat{\theta}_n - \theta_0) &= -\sqrt{n(n-1)}\left(\hat{\Psi}_n(\theta_0, p_n) + \nabla_p \Psi_n(\theta_0, p_n)(\hat{p}_n - p_n)\right) + o_p(1) \\ &= -\sqrt{n(n-1)}\tilde{\Psi}_n(\theta_0, p_n) + o_p(1)\end{aligned}$$

Lemma 8.4 shows that

$$\sqrt{n(n-1)}\tilde{\Omega}_n^{-1/2}(X_n, p_n)\tilde{\Psi}_n(\theta_0, p_n) \xrightarrow{d} N(0, I)$$

as $n \to \infty$. This implies that

$$\sqrt{n(n-1)}\tilde{\Omega}_n^{-1/2}(X_n, p_n)\nabla_\theta \Psi_n(\theta_0, p_n)(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I)$$

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

8.4 Proofs in Section 5

Proof of Example 3. We verify that $U_{n,ij}(X, \sigma)$ and $U_{ij}(\sigma)$ given in Example 3 satisfy Assumption 4. Note that $\sup_\sigma \left|\frac{1}{n-2} D_j' diag(V_i(\sigma)) D_j\right| = o_p(1)$ (because $V_i(\sigma)$ is uniformly bounded) and that $\sup_\sigma \frac{1}{n-1}\sum_{j\neq i}\left|\frac{1}{n-2}D_j' diag(V_i(\sigma)) D_j\right| = \frac{1}{n-2}\sup_\sigma \frac{1}{n-1}\sum_{j\neq i}\left|D_j' diag(V_i(\sigma)) D_j\right| = o_p(1)$ by applying uniform law of large numbers to the second sup term, which is appropriate because the space of symmetric σ is compact, $V_i(\sigma)$ is continuous in σ and is uniformly bounded. It suffices to show that for any θ,

$$\sup_\sigma \left|\frac{1}{n-2}\sum_{k\neq i,j}\sigma_{jk}(X_j, X_k)\, w(X_i, X_k)\,\beta_4 - \mathbb{E}\left[\sigma_{jk}(X_j, X_k)\, w(X_i, X_k)\right| X_i, X_j\right]\beta_4\right| = o_p(1)$$

and

$$\sup_\sigma \frac{1}{(n-1)}\sum_{j\neq i}\left|\frac{1}{n-2}\sum_{k\neq i,j}\sigma_{jk}(X_j, X_k)\, w(X_i, X_k)\,\beta_4 - \mathbb{E}\left[\sigma_{jk}(X_j, X_k)\, w(X_i, X_k)\right| X_i, X_j\right]\beta_4\right| = o_p(1)$$

The former is satisfied by applying uniform law of large numbers again. As for the latter, write $\Delta_i(X_j, X_k; \sigma) = \sigma_{jk}(X_j, X_k) w(X_i, X_k) \beta_4 - \mathbb{E}[\sigma_{jk}(X_j, X_k) w(X_i, X_k) | X_i, X_j] \beta_4$, which has zero conditional mean $\mathbb{E}[\Delta_i(X_j, X_k; \sigma) | X_i] = 0$. By Cauchy-Schwarz inequality, we have $\left(\frac{1}{n}\sum_i |y_i|\right)^2 \leq \frac{1}{n}\sum_i y_i^2$, so

$$\begin{aligned}
&\left(\frac{1}{(n-1)(n-2)} \sum_{j \neq i} \left| \sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma) \right| \right)^2 \\
&\leq \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \left(\sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma) \right)^2 \\
&= \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma)^2 \\
&+ \frac{1}{(n-1)(n-2)^2} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \Delta_i(X_j, X_k; \sigma) \Delta_i(X_j, X_l; \sigma)
\end{aligned}$$

The last two terms are U-processes. It remains to show that they are $o_p(1)$ uniformly in σ. For the first term, applying Corollary 7 in Sherman (1994) yields

$$\sup_{\sigma} \left| \frac{1}{(n-1)(n-2)} \sum_{j \neq i} \sum_{k \neq i,j} \Delta_i(X_j, X_k; \sigma)^2 - \mathbb{E}\left[\Delta_i(X_j, X_k; \sigma)^2 \middle| X_i \right] \right| = O_p\left(\frac{1}{\sqrt{n}} \right)$$

so the first terms is $O_p\left(\frac{1}{n}\right)$ uniformly in σ. As for the second term, note that the product $\Delta_i(X_j, X_k; \sigma) \Delta_i(X_j, X_l; \sigma)$ has zero mean conditional on X_i, so by Corollary 7 in Sherman (1994) again we obtain

$$\sup_{\sigma} \left| \frac{1}{(n-1)(n-2)(n-3)} \sum_{j \neq i} \sum_{k \neq i,j} \sum_{l \neq i,j,k} \Delta_i(X_j, X_k; \sigma) \Delta_i(X_j, X_l; \sigma) \right| = O_p\left(\frac{1}{\sqrt{n}} \right)$$

which proves that the second term is $o_p(1)$ uniformly in σ. The proof is complete. ■

Proof of Proposition 5.1. Step 1: We first prove (5.4). Because $\frac{\partial}{\partial c}\mathbb{E}[c - \varepsilon]_+ = \frac{\partial}{\partial c}\int_{-\infty}^{c} (c - \varepsilon) f_\varepsilon(\varepsilon) d\varepsilon = F_\varepsilon(c)$, the first order condition of the problem (5.3) is

$$\nabla_\omega \Pi(\omega, X_i, \sigma) = 2V_i(\sigma) \mathbb{E}\left(D_j F_\varepsilon\left(U_{ij}(\sigma) + 2D_j' V_i(\sigma) \omega \right) \middle| X_i \right) - 2V_i(\sigma)\omega = 0 \qquad (8.101)$$

It is easy to see that any $\omega_i(\sigma)$ that satisfies the first order condition must be bounded. Hence without loss of generality we can assume that $\omega_i(\sigma)$ is in a compact set Ω. Since $\Pi(\omega, X_i, \sigma)$ is continuous in σ, the compactness of Ω implies that the unique maximizer

$V_i(\sigma)\omega_i^*(\sigma)$ (by Assumption 5) is also well separated. If we can further show that

$$\sup_{\omega,\sigma}\left|\Pi_n(\omega, X, \varepsilon_i, \sigma) - \Pi(\omega, X_i, \sigma)\right| = o_p(1) \tag{8.102}$$

then following a standard proof for uniform consistency we can prove (5.4). Specifically, from (8.102) we have $\sup_\sigma \Pi_n(\omega_{n,i}, X, \varepsilon_i, \sigma) \geq \sup_\sigma \Pi_n(\omega_i^*, X, \varepsilon_i, \sigma) \geq \sup_\sigma \Pi(\omega_i^*, X_i, \sigma) - o_p(1)$, whence,

$$\begin{aligned}\sup_\sigma \Pi(\omega_i^*, X_i, \sigma) - \sup_\sigma \Pi(\omega_{n,i}, X_i, \sigma) &\leq \sup_\sigma \Pi_n(\omega_{n,i}, X_i, \varepsilon_i, \sigma) - \sup_\sigma \Pi(\omega_{n,i}, X_i, \sigma) + o_p(1) \\ &\leq \sup_{\omega,\sigma}\left|\Pi_n(\omega, X_i, \varepsilon_i, \sigma) - \Pi(\omega, X_i, \sigma)\right| + o_p(1) = o_p(1).\end{aligned}$$

Well-separateness of $V_i(\sigma)\omega_i^*(\sigma)$ implies that for any $\varepsilon > 0$, there is $\eta > 0$ such that, for any symmetric σ, $\Pi(\omega, X_i, \sigma) < \Pi(\omega_i^*, X_i, \sigma) - \eta$ for every ω with $\|V_i(\sigma)\omega - V_i(\sigma)\omega_i^*(\sigma)\| \geq \varepsilon$. Therefore,

$$\begin{aligned}\Pr\left(\sup_\sigma \|V_i(\sigma)\omega_{n,i}(\sigma) - V_i(\sigma)\omega_i^*(\sigma)\| \geq \varepsilon\right) &\leq \Pr\left(\sup_\sigma \left[\Pi(\omega_{n,i}, X_i, \sigma) - \Pi(\omega_i^*, X_i, \sigma)\right] < -\eta\right) \\ &\leq \Pr\left(\sup_\sigma \Pi(\omega_{n,i}, X_i, \sigma) - \sup_\sigma \Pi(\omega^*, X_i, \sigma) < -\eta\right) \\ &\to 0\end{aligned}$$

in view of the preceding display (5.4) is proved.

Now we prove (8.102). The left hand side of (8.102) equals

$$\begin{aligned}&\sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[U_{n,ij}(X,\sigma) + \frac{n-1}{n-2}2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+ - \mathbb{E}\left(\left[U_{ij}(\sigma) + 2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right)\right| \\ &\leq \sup_{\omega,\sigma}\frac{1}{n-1}\sum_{j\neq i}\left[\frac{1}{n-2}2D_j'V_i(\sigma)\omega\right]_+ \\ &+ \sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[U_{n,ij}(X,\sigma) + 2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+ - \left[U_{ij}(\sigma) + 2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+\right| \\ &+ \sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[U_{ij}(\sigma) + 2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+ - \mathbb{E}\left(\left[U_{ij}(\sigma) + 2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right)\right|\end{aligned} \tag{8.103}$$

For the third term in (8.103), because $\left[U_{ij}(\sigma) + 2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+$ are i.i.d. conditional on X_i with conditional mean $\mathbb{E}\left(\left[U_{ij}(\sigma) + 2D_j'V_i(\sigma)\omega - \varepsilon_{ij}\right]_+ \middle| X_i\right)$, are continuous in ω and σ, and are bounded by

$$\left[U_{ij}(\sigma)+2D_j'V_i(\sigma)\omega-\varepsilon_{ij}\right]_+ \leq \left[\sup_{\omega,\sigma}\left(U_{ij}(\sigma)+2D_j'V_i(\sigma)\omega\right)-\varepsilon_{ij}\right]_+$$

which is absolute integrable because of the continuity of $U_{ij}(\sigma)$ and $V_i(\sigma)$ and compactness of the spaces of ω and σ, uniform law of large numbers holds, so

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

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

$$\begin{aligned}&\sup_{\omega,\sigma}\left|\frac{1}{n-1}\sum_{j\neq i}\left[U_{n,ij}(\sigma)+2D_j'V_i(\sigma)\omega-\varepsilon_{ij}\right]_+ - \left[U_{ij}(\sigma)+2D_j'V_i(\sigma)\omega-\varepsilon_{ij}\right]_+\right| \\ &\leq \sup_{\sigma}\frac{1}{n-1}\sum_{j\neq i}\left|U_{n,ij}(\sigma)-U_{ij}(\sigma)\right| = o_p(1)\end{aligned}$$

by Assumption 4(ii). Finally, the first term in (8.103) is $o_p(1)$, again by uniform law of large numbers. Hence (8.102) is proved.

Step 2: Next we prove (5.5). By the definition of P_n and P,

$$\begin{aligned}&\left|P_n(X_i,X_j;X,\sigma)-P(X_i,X_j;\sigma)\right| \\ &\leq \int\left|1\left\{U_{n,ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'V_i(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)\geq\varepsilon_{ij}\right\}-1\left\{U_{ij}(\sigma)+2D_j'V_i(\sigma)\omega_i^*(\sigma)\geq\varepsilon_{ij}\right\}\right|dF_{\varepsilon_i}(\varepsilon_i) \\ &\leq \Pr\left(U_{n,ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'V_i(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)\geq\varepsilon_{ij}>U_{ij}(\sigma)+2D_j'V_i(\sigma)\omega_i^*(\sigma)\middle| X\right) \\ &\Pr\left(U_{n,ij}(X,\sigma)+2D_j'V_i(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)<\varepsilon_{ij}\leq U_{ij}(\sigma)+2D_j'V_i(\sigma)\omega_i^*(\sigma)\middle| X\right)\end{aligned}$$

Since the last two terms are similar, without loss of generality it suffices to show that the second last term is $o_p(1)$ uniformly in σ. Define

$$\begin{aligned}M_n(X,\varepsilon_i,\sigma) &= U_{n,ij}(X,\sigma)+\frac{n-1}{n-2}2D_j'V_i(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma) \\ M(\sigma) &= U_{ij}(\sigma)+2D_j'V_i(\sigma)\omega_i^*(\sigma)\end{aligned}$$

Fix $\eta>0$. Because $\varepsilon_{ij}\in(M(\sigma),M_n(X,\sigma)]$ implies that $\varepsilon_{ij}\in(M(\sigma),M(\sigma)+\eta]$ or

$M_n(X,\sigma) > M(\sigma)+\eta$, we can bound the second last term as follows

$$\begin{aligned}
&\Pr\left(M_n(X,\varepsilon_i,\sigma) \geq \varepsilon_{ij} > M(\sigma)\middle| X\right)\\
&\leq \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma)\middle| X\right) + \Pr\left(M_n(X,\varepsilon_i,\sigma) > M(\sigma)+\eta\middle| X\right)\\
&\leq \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma)\middle| X\right) + \Pr\left(2D_j'V_i(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma)-\omega_i^*(\sigma)\right) > \frac{\eta}{2}\middle| X\right)\\
&+1\left\{U_{n,ij}(X,\sigma)-U_{ij}(\sigma) > \frac{\eta}{2}\right\}
\end{aligned}$$

where the last inequality follows because $M_n(X,\varepsilon_i,\sigma)-M(\sigma) > \eta$ implies that $U_{n,ij}(X,\sigma)-U_{ij}(\sigma) > \frac{\eta}{2}$ or $2D_j'V_i(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma)-\omega_i^*(\sigma)\right) > \frac{\eta}{2}$. It suffices to show that the last three terms in the display are $o_p(1)$ uniformly in σ. For any $\delta > 0$, the last term satisfies

$$\begin{aligned}
&\Pr\left(\sup_\sigma 1\left\{U_{n,ij}(X,\sigma)-U_{ij}(\sigma) > \frac{\eta}{2}\right\} > \delta\middle| X_i,X_j\right)\\
&\leq \Pr\left(\sup_\sigma \left(U_{n,ij}(X,\sigma)-U_{ij}(\sigma)\right) > \frac{\eta}{2}\middle| X_i,X_j\right) \to 0
\end{aligned}$$

as $n\to\infty$ by Assumption 4(i), so it is $o_p(1)$ uniformly in σ. For the second term, we have

$$\begin{aligned}
&\Pr\left(\sup_\sigma \Pr\left(2D_j'V_i(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma)-\omega_i^*(\sigma)\right) > \frac{\eta}{2}\middle| X\right) > \delta\middle| X_i,X_j\right)\\
&\leq \frac{1}{\delta}\mathbb{E}\left(\sup_\sigma \Pr\left(2D_j'V_i(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma)-\omega_i^*(\sigma)\right) > \frac{\eta}{2}\middle| X\right)\middle| X_i,X_j\right)\\
&\leq \frac{1}{\delta}\mathbb{E}\left(\Pr\left(\sup_\sigma 2D_j'V_i(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma)-\omega_i^*(\sigma)\right) > \frac{\eta}{2}\middle| X\right)\middle| X_i,X_j\right)\\
&= \frac{1}{\delta}\Pr\left(\sup_\sigma 2D_j'V_i(\sigma)\left(\frac{n-1}{n-2}\omega_{n,i}(X,\varepsilon_i,\sigma)-\omega_i^*(\sigma)\right) > \frac{\eta}{2}\middle| X_i,X_j\right) \to 0
\end{aligned}$$

as $n\to\infty$ by Markov inequality, law of iterated expectation, and uniform consistency of $V_i(\sigma)\omega_{n,i}(X,\varepsilon_i,\sigma)$ proved earlier, so the second term is also $o_p(1)$ uniformly in σ. As for the first term,

$$\begin{aligned}
&\sup_\sigma \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma)\middle| X\right)\\
&= \sup_\sigma \left(F_{\varepsilon_{ij}}(M(\sigma)+\eta) - F_{\varepsilon_{ij}}(M(\sigma))\right)\\
&= \sup_\sigma f_{\varepsilon_{ij}}(M(\sigma)+\tilde{\eta})\,\eta
\end{aligned}$$

for some $\tilde{\eta}\in[0,\eta]$ where the last equality is from mean value theorem. Since η is arbitrary, choosing η to be $o(1)$, we can get $\sup_\sigma \Pr\left(M(\sigma)+\eta \geq \varepsilon_{ij} > M(\sigma)\middle| X,\sigma\right) = o_p(1)$. 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. Hong, H., Krainer, J., & Nekipelov, D. (2010) Computing equilibria in static games of incomplete information using the all-solution homotopy, mimeo, University of Minnesota.

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

[4] De Paula, Richards-Shubik and Tamer (2017) Identification of Preferences in Network Formation Games, *Econometrica* (forthcoming).

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

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

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

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

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

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

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

[12] Sheng, S. (2017) A Structural Econometric Analysis of Network Formation Games, UCLA, working paper.

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

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

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