Social Interactions with Endogenous Group Formation*

Shuyang Sheng[†] Xiaoting Sun[‡]

March 30, 2023

Abstract

This paper investigates the identification and estimation of social interactions with endogenous group formation. We characterize group formation by a two-sided many-to-one matching model, where individuals choose among groups according to their preferences, and a group ranks individuals based on their qualifications and admits those with the highest qualifications until the capacity is reached. Following Azevedo and Leshno (2016), we show that an equilibrium in a finite market converges to a limit as the number of individuals in the market grows large. Based on this limiting approximation, we derive the selection bias as a group-specific nonparametric function of the group formation indices. Assuming the unobservables are exchangeable across groups, the peer effects can be identified by controlling for the selection bias as in a sample selection model. The excluded variables in group formation also provide instruments that can help resolve the reflection problem (Manski, 1993). We propose a multi-stage distribution-free semiparametric estimator based on our constructive identification results. The proposed estimator is $\sqrt{n}$ consistent and asymptotically normal and performs well in simulations.

*We are grateful to Xiaohong Chen, Aureo de Paula, Brian Krauth, Michael Leung, Arthur Lewbel, Rosa Matzkin, Krishna Pendakur and Geert Ridder for their helpful comments. We are also grateful to seminar and conference participants at Caltech, Jinan University (IESR), Rice, UCL, USC, Yale, 2021 Asia Meeting of the Econometric Society, 2021 Australasia Meeting of the Econometric Society, 2021 China Meeting of the Econometric Society, 2022 California Econometrics Conference, 37th Canadian Econometric Study Group Meetings, 55th Annual Conference of the Canadian Economics Association, ASSA/ES North American Winter Meeting 2021, Econometric Society World Congress 2020, and IAAE 2021 Annual Conference. Sun is grateful for financial support from the Social Sciences and Humanities Research Council of Canada through its Insight Development Grants Program. All errors are our own.

[†]Department of Economics, University of California at Los Angeles, Los Angeles, CA, USA, 90095, email: ssheng@ucla.edu

[‡]Department of Economics, Simon Fraser University, Vancouver, BC, Canada, V5A 1S6, email: xiaoting_sun@sfu.ca

1 Introduction

Social interaction models are useful for investigating the interdependence of individual outcomes in a wide range of contexts such as educational attainment, teenage smoking, and criminal activity. A salient feature of social interactions is that they often take place among individuals who belong to a certain social or economic group. For example, a student tends to interact with other students in her school or college (e.g., Kremer et al., 2009; Argys and Rees, 2008; Carrell et al., 2008; Gaviria and Raphael, 2001; Lavy and Schlosser, 2011; Sacerdote, 2011; Zimmerman, 2003). A resident tends to interact with other residents in the same neighborhood (e.g., Bayer et al., 2008; Damm and Dustmann, 2014; Åslund et al., 2011; Bobonis and Finan, 2009; Katz et al., 2001; Kling et al., 2005; Lewbel et al., 2021). It is evident that entering such a social or economic group, for example, by being admitted to a college or moving into a neighborhood, can be selective, thereby posing a challenge in quantifying the causal social effects among the group members (e.g., Ioannides and Zabel, 2008; Bayer et al., 2007; Epple and Romano, 1998).

Most of the literature on social interactions assumes that peer relationships, represented by an adjacency matrix, are exogenous (See Blume et al., 2011, for a survey). Several recent studies (e.g., Goldsmith-Pinkham and Imbens, 2013; Hsieh and Lee, 2016; Johnsson and Moon, 2021; Auerbach, 2022) relax this assumption by incorporating a network formation model and introducing unobserved individual heterogeneity that affects both the link formation and individual outcome, thus leading to endogeneity in the peer relationships. The endogeneity of this kind can be resolved once the individual heterogeneity is controlled for. Unlike these studies, we consider a more empirically motivated setting where the endogeneity in peer relationships stems from the selective entry into each group. For example, there is sizable empirical evidence documenting that students may be sorted into schools/classes and the potential sorting hinders the identification of the peer effects (Sacerdote, 2011; Epple and Romano, 2011; Friesen and Krauth, 2007). We develop a new framework to characterize the endogenous entry into groups and identify and estimate the causal social effects.

Specifically, we discover that group formation can be equivalently characterized by a two-sided many-to-one matching model, where the individuals who join a group can be regarded as being "matched with" the group. To form the groups, individuals choose among the groups based on their preferences and each group ranks the individuals based on their qualifications and admits those with the highest qualifications given the group capacity (Azevedo and Leshno, 2016; He et al., 2022). For example, in college admissions, each student chooses among the colleges, and only those who are admitted to a college can enter the college. The groups formed eventually are determined as an equilibrium outcome. This two-sided group formation framework covers one-sided group formation (e.g., neighborhood choice) as a special case, where each group has an unbounded capacity and joining a group reduces to a simple multinational discrete choice problem. To the best of our knowledge, this is the first paper that exploits a matching framework to characterize group formation.

Based on the group formation model, we show that the endogeneity in group formation leads to a selection bias in the individual outcome. To derive this selection bias, we follow the many-to-one matching literature and characterize an equilibrium in group formation (i.e., stable groups) by group *cutoffs*, where the cutoff of a group is the minimum qualification to join the group (Azevedo and Leshno, 2016; He et al., 2022). Through the equilibrium cutoffs, the groups formed in an equilibrium depend on the unobserved characteristics of all the individuals in the market, rendering it difficult to represent the selection bias. To overcome this difficulty, we approximate an equilibrium in a market with n individuals by the equilibrium in a limiting market as n goes to infinity. The limiting equilibrium has the feature that the group that an individual joins only depends on her own characteristics, thereby yielding a tractable expression of the selection bias. The idea of limiting approximation has been exploited in the matching literature (Menzel, 2015; Azevedo and Leshno, 2016; He et al., 2022). As far as we know, our paper is the first to apply the limiting approximation to derive the selection bias due to endogenous group formation. Under the limiting approximation, we can explicitly express the selection bias as a group-specific nonparametric function of the preference and qualification indices in group formation.

The group-specific selection bias gives rise to a further challenge in the identification of the social interaction effects, in both one-sided and two-sided group formation settings. In the case where the adjacency matrix is given by group averages, the group-level social effects can not be separately identified from the group-specific selection bias. To conquer the challenge, we impose an assumption that the unobservables in group formation are exchangeable across groups, so that the selection bias can be represented using a group-invariant selection function, provided that the cutoffs and group fixed effects are identified. We also provide constructive results on the identification of the cutoffs and group fixed effects. By partialling out this version of the selection bias, we can thus identify the exogenous social effect.

In a general setting with the presence of both exogenous and endogenous social interactions, it is well known that the social effects may not be separately identifiable, referred to as the reflection problem (Manski, 1993). The existing literature proposed to achieve identification using the exogenous variation in group size (Lee, 2007; Davezies et al., 2009; Graham, 2008) or the observed characteristics of friends of friends (Bramoullé et al., 2009). Brock and Durlauf (2001) pointed out that self-selection into groups may aid in the identification of the social effects. In this paper, we follow the insight of Brock and Durlauf (2001) and propose a new approach that exploits the exogenous variation in the observed characteristics in group formation to resolve the reflection problem. We discover that after the selection bias is partialled out, the excluded variables in group formation can be used as instruments to identify the social effects. Such instruments are valid regardless of whether the adjacency matrix is given by group averages as proposed in Manski (1993) or additional networks within each group. To our knowledge, our paper is the first to use excluded variables in group formation to resolve the reflection problem.

As for the estimation, we propose semiparametric methods to estimate the model parameters, where the distribution of the unobservables is assumed to be nonparametric. We first develop a distribution-free semiparametric estimator for the parameters in group formation. In particular, we propose a two-step kernel estimator for the cutoffs and group fixed effects based on our constructive identification results. To the best of our knowledge, this is the first paper that proposes a distribution-free estimator for the cutoffs in a two-sided many-to-one matching model. We then propose a semiparametric two-step GMM estimator for the parameters in social interactions, where we first partial out the selection bias by sieve estimators, and then estimate the social interaction parameters by GMM. The proposed estimators in both group formation and social interactions are $\sqrt{n}$ consistent and asymptotically normal.

The remainder of the paper is organized as follows. Section 2 introduces the model. Section 3 derives the nonparametric selection bias. Section 4 investigates the identification. Section 5 presents the estimation methods. Section 6 conducts a simulation study. Section 7 concludes the paper. Additional results are presented in the Online Appendix.

2 Model

2.1 Social Interactions

Consider a set of individuals $\mathcal{N} = \{1, 2, \ldots, n\}$ who interact following the standard linear-in-means social interaction model

$$y_i = \sum_{j=1}^{n} w_{ij} y_j \gamma_1 + \sum_{j=1}^{n} w_{ij} x_j' \gamma_2 + x_i' \gamma_3 + \epsilon_i. \tag{2.1}$$

In this specification, $y_i \in \mathbb{R}$ represents the outcome of interest (earnings, employment, or education), $x_i \in \mathbb{R}^{d_x}$ is a vector of observed characteristics, $\epsilon_i \in \mathbb{R}$ is an unobserved shock, and $w_{ij} \in \mathbb{R}_+$ denotes the weight of peer j on individual i. We assume that i's outcome y_i depends on $\sum_{j=1}^{n} w_{ij} y_j$ and $\sum_{j=1}^{n} w_{ij} x_j$, the weighted averages of outcomes and observed characteristics of i's peers. Following the terminology in Manski (1993), γ_1 captures the endogenous social effect, and γ_2 captures the exogenous/contextual social effect. The parameter of interest is $\gamma = (\gamma_1, \gamma_2', \gamma_3')' \in \mathbb{R}^{2d_x+1}$.

In this paper, we focus on a setting where the adjacency matrix $\boldsymbol{w} = (w_{ij}) \in \mathbb{R}_+^{n^2}$ presents a group structure. Suppose there is a set of groups $\mathcal{G} = \{1, \ldots, G\}$ that the individuals can join. We assume that the number of groups G is finite and each group $g \in \mathcal{G}$ has a predetermined capacity n_g that is proportional to n. The groups are non-overlapping (such as colleges and nursing homes), so one joins only one group. Let $g_i \in \mathcal{G}$ denote the group that i joins and $\boldsymbol{g} = (g_1, \ldots, g_n)'$ the $n \times 1$ vector that stacks g_i. We assume that $w_{ij} = 0$ if $g_i \neq g_j$ – an individual is influenced by her groupmates only. A typical example is given by group averages where we set $w_{ij} = \frac{1}{n_{g_i}}$ if $g_i = g_j$ and $w_{ij} = 0$ if $g_i \neq g_j$ (Manski, 1993).[1] We also allow w_{ij} to take a more general form so long as the interactions occur within a group. For example, group members may form additional friendships and only friends in the group can have an impact.

Let $\boldsymbol{y}$ denote the $n \times 1$ vector that stacks y_i, $\boldsymbol{x}$ the $n \times d_x$ matrix that stacks the row vectors x_i', and $\boldsymbol{\epsilon}$ the $n \times 1$ vector that stacks ϵ_i. Write (2.1) in a matrix form

$$\boldsymbol{y} = \boldsymbol{w}\boldsymbol{y}\gamma_1 + \boldsymbol{w}\boldsymbol{x}\gamma_2 + \boldsymbol{x}\gamma_3 + \boldsymbol{\epsilon}. \tag{2.2}$$

The literature on linear-in-means models typically assumes that the adjacency matrix

[1] This example assumes that all groups reach their full capacities.

$\boldsymbol{w}$ is independent of the unobservables $\boldsymbol{\epsilon}$. We relax this assumption to accommodate endogenous group formation. Specifically, we allow $\boldsymbol{w}$ to be endogenous, though the endogeneity only occurs at the group level, as detailed in Assumption 1 below. In the next section, we develop a model of group formation to account for the endogeneity of groups.

2.2 Group Formation

Suppose that, prior to social interactions, the groups are established through two-sided decisions. On one side, an individual chooses a group according to her preferences over groups. On the other side, a group ranks individuals based on their qualifications and admits those with the highest qualifications until its capacity is reached.[2] Two-sided group formation has various applications such as admissions to colleges or schools, medical residency programs, and nursing homes. The framework nests one-sided group formation as a special case, where each group has an infinite capacity, and thus an individual unilaterally determines which group to enter.[3]

Two-sided group formation can be equivalently characterized as two-sided many-to-one matching without transfers, where the individuals in a group are regarded as being "matched with" the group. Therefore, we specify group formation following the literature on two-sided many-to-one matching without transfers (Azevedo and Leshno, 2016; He et al., 2022).

Utility For individual $i \in \mathcal{N}$ and group $g \in \mathcal{G}$, let u_{ig} denote i's utility of joining group g

$$u_{ig} = z_i' \delta_g^u + \xi_{ig}, \tag{2.3}$$

and v_{gi} denote i's qualification for group g

$$v_{gi} = z_i' \delta_g^v + \eta_{gi}, \tag{2.4}$$

[2]For instance, in college admissions the qualifications of a student reflect colleges' preferences over the student; in school assignments, however, the qualifications may depend on (observed) priorities of the student such as the district that the student lives in and whether the student has siblings attending the school.

[3]Brock and Durlauf (2001; 2005) provided examples of one-sided group formation where individuals unilaterally select into a group. In Brock and Durlauf (2001), an individual decides whether to enter a potential group or not. In Brock and Durlauf (2005), an individual chooses one out of multiple groups.

where $z_i \in \mathbb{R}^{d_z}$ represents a vector of individual- or pair-specific observed characteristics that affect the preferences and qualifications in group formation.[4] The group-specific coefficients $\delta_g^u, \delta_g^v \in \mathbb{R}^{d_z}$ allow the effect of z_i to be heterogeneous across groups. Note that z_i may overlap with the observed characteristics in social interactions x_i.[5] $\xi_{ig} \in \mathbb{R}$ and $\eta_{gi} \in \mathbb{R}$ represent pair-specific unobserved utility and qualification shocks of individual i for group g. Let $\xi_i = (\xi_{i1}, \ldots, \xi_{iG})'$ and $\eta_i = (\eta_{1i}, \ldots, \eta_{Gi})'$. We assume that the joint distribution of ϵ_i, ξ_i, and η_i is nonparametric, which has the advantage of allowing ϵ_i to have flexible dependence with ξ_i and η_i.[6]

It is worthy to point out that our model does not allow for peer effects in group formation. In particular, the utility and qualification specified in equations (2.3) and (2.4) do not depend on prospective group members. Nevertheless, if we have additional information about the group members in the past (e.g., students enrolled in a school in previous years), we can approximate the peer effects by including in z_i the outcome/characteristics of previous group members, provided that these group-level measures do not vary over time. For example, assuming that the gender composition in a school remains stable over time, we can use the fraction of female in the previous year as a proxy for the gender peer effect.

Example 2.1 (College admissions)**.** College admissions provide an example of two-sided group formation, where students apply for colleges based on their utilities (u_{ig}), and colleges admit students based on qualifications (v_{gi}). In this example, ξ_i refers to student i's unobserved preferences for the colleges (family tradition), and η_i refers to student i's unobserved ability that affects her qualifications (extracurricular activities). z_i represents a set of characteristics that affect the preferences of the student or colleges, including student-specific characteristics (family income, parental education, and SAT scores) and pair-specific characteristics (distance to a college and the interaction between a student's minority status and the fraction of minorities in a college in the previous year).

[4] To illustrate that the specification in equations (2.3) and (2.4) covers both individual- and pair-specific characteristics, suppose that there are two groups and we specify $v_{gi} = s_i\delta_{g,s} + d_{ig}\delta_{g,d} + \eta_{gi}$, where s_i is an individual-specific variable and d_{ig} is a pair-specific variable (which can be an individual-specific variable interacted with a group-specific variable). This example can be represented by equation (2.4) with $z_i = (s_i, d_{i1}, d_{i2})'$, $\delta_1^v = (\delta_{1,s}, \delta_{1,d}, 0)'$, and $\delta_2^v = (\delta_{2,s}, 0, \delta_{2,d})'$.

[5] For the identification of γ, z_i has to have at least two components that are excluded from x_i. See Section 4 for more details.

[6] The nonparametric specification implies that z_i does not include a constant or group-specific variables, because these group-level heterogeneity cannot be separated from ξ_i and η_i.

Equilibrium Following the matching literature (Roth and Sotomayor, 1992), we assume that the group formation outcome is stable.[7] Azevedo and Leshno (2016) showed that a stable matching exists and can be characterized by group *cutoffs*. Let p_g denote the cutoff of group $g \in \mathcal{G}$. It is given by the lowest qualification among the group members if the capacity constraint is binding; otherwise, the cutoff is set to $-\infty$. Namely,

$$p_g = \begin{cases} \inf_{i:g_i=g} v_{gi}, & \text{if } \sum_{i\in\mathcal{N}} 1\{g_i = g\} = n_g, \\ -\infty, & \text{if } \sum_{i\in\mathcal{N}} 1\{g_i = g\} < n_g. \end{cases}$$

The cutoff of a group reflects how selective the group is.

Given the cutoffs $p = (p_1, \dots, p_G)'$, let $\mathcal{C}_i(p) = \{g \in \mathcal{G} : v_{gi} \geqslant p_g\} \subseteq \mathcal{G}$ denote individual $i's$ *choice set* (i.e., the subset of groups that individual i qualifies for) . Within $\mathcal{C}_i(p)$, i chooses the group that yields the highest utility

$$g_i = \arg\max_{g\in\mathcal{C}_i(p)} u_{ig}.$$

This is a multinomial discrete choice problem with the choice set $\mathcal{C}_i(p)$ determined endogenously by the cutoffs p. Individual i joins group g if (i) i qualifies for group g, and (ii) for any other group $h \neq g$, either i prefers group g to group h, or i does not qualify for group h.[8] That is,

$$\begin{aligned} & 1\{g_i = g\} \\ = \; & 1\{v_{gi} \geq p_g\} \cdot \prod_{h\neq g} 1\{u_{ih} < u_{ig} \text{ or } v_{hi} < p_h\} \\ = \; & 1\{\eta_{gi} \geq p_g - z_i'\delta_g^v\} \cdot \prod_{h\neq g} 1\{\xi_{ih} - \xi_{ig} < z_i'(\delta_g^u - \delta_h^u) \text{ or } \eta_{hi} < p_h - z_i'\delta_h^v\}. \end{aligned} \tag{2.5}$$

Equation (2.5) indicates that the group that individual i joins is a function of i's

[7] In the context of college admissions, stability can be achieved through various means. One way is for students to apply to all acceptable colleges and use a stable mechanism like the deferred acceptance algorithm to determine the matching Gale and Shapley (1962). Even if students choose not to apply to all acceptable colleges due to application costs Fack et al. (2019) or errors in the application process Artemov et al. (2020), stability can still be achieved theoretically as long as students know the criteria used by colleges to rank them.

[8] For simplicity of exposition, we assume that individuals always prefer to join a group. This can be relaxed by assuming that the utility of not joining any group is $u_{i0} = \xi_{i0}$. Such relaxation does not lead to substantive technical modifications of the results.

observed and unobserved characteristics z_i, ξ_i, η_i as well as the cutoffs p. We can write $g_i = g(z_i, \xi_i, \eta_i; p)$.

We remark that in one-sided group formation, the capacities are infinite and the cutoffs p_g are set to $-\infty$. The choice set $\mathcal{C}_i(p)$ is simply $\mathcal{G}$. The optimal decision in equation (2.5) reduces to $1\{g_i = g\} = \prod_{k \neq g} 1\{u_{ik} < u_{ig}\} = \prod_{k \neq g} 1\{\xi_{ik} - \xi_{ig} < z_i'(\delta_g^u - \delta_k^u)\}$, and we return to a standard multinomial discrete choice problem. The group that individual i joins is a function of z_i and ξ_i only, that is, $g_i = g(z_i, \xi_i)$.

In a stable matching, the cutoffs p clear the supply of and demand for each group.[9] Let $\boldsymbol{z}$ denote the $n \times d_z$ matrix that stacks z_i', $\boldsymbol{\xi}$ the $n \times G$ vector that stacks ξ_i, and $\boldsymbol{\eta}$ the $n \times G$ vector that stacks η_i. An equilibrium cutoff vector can be represented as $p(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta})$.[10] Given the equilibrium cutoffs, the groups are formed following equation (2.5), and the equilibrium groups can be written as $\boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}))$.

3 Selection Bias

In this section, we investigate the bias that results from selection into groups. Throughout the paper, we maintain the following assumptions.

Assumption 1. *The adjacency matrix $\boldsymbol{w}$ is independent of $\boldsymbol{\epsilon}$ conditional on $\boldsymbol{x}$, $\boldsymbol{z}$,* and *$\boldsymbol{g}$.*

Assumption 2. *(i) x_i, z_i, ϵ_i, ξ_i, and η_i are i.i.d. for $i = 1, \ldots, n$. (ii) The joint cdf of ϵ_i, ξ_i, and η_i is continuously differentiable. (iii) For all i, ϵ_i, ξ_i, and η_i are independent of x_i and z_i.*

Assumption 1 requires that the selection only occurs at the group level. For an adjacency matrix that represents group averages, this assumption is trivially satisfied. For an adjacency matrix of a more general form, the assumption requires that $\boldsymbol{w}$ is independent of $\boldsymbol{\epsilon}$ given the group structure $\boldsymbol{g}$ – for example, conditional on group memberships, how groupmates make friends is independent of $\boldsymbol{\epsilon}$. This assumption ensures that we can focus on the formation of groups to deal with the endogeneity of $\boldsymbol{w}$. Assumption 2(i) imposes an i.i.d. assumption that is typical in social interactions. Assumption 2(ii) imposes a smoothness assumption on the joint cdf of the

[9] The equilibrium cutoffs p satisfy the following market-clearing equations: $\sum_{i \in \mathcal{N}} 1\{g(z_i, \xi_i, \eta_i; p) = g\} \leq n_g$ and $\sum_{i \in \mathcal{N}} 1\{g(z_i, \xi_i, \eta_i; p) = g\} = n_g$ if $p_g > -\infty$, for all $g \in \mathcal{G}$.

[10] There may be multiple equilibrium cutoffs in a finite n economy.

unobservables. It is to ensure that the conditional probability that individual i joins group g is continuously differentiable. Assumption 2(iii) is a standard assumption that the observables are exogenous.

3.1 The Presence of Selection Bias

The social interaction model in equation (2.1) presents a selection bias if

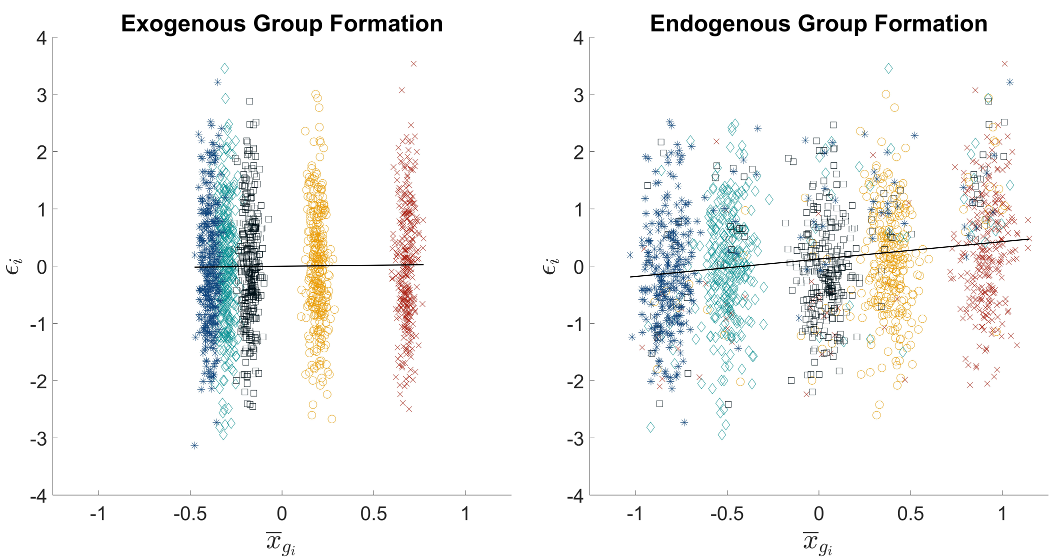
$$\mathbb{E}[\epsilon_i | \boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}))] \neq 0. \tag{3.1}$$

Under Assumption 2(i)(iii), this bias arises from the dependence between the outcome shock ϵ_i and the unobservables in group formation ξ_i and η_i. The dependence causes ϵ_i to be correlated with g_i because g_i is a function of ξ_i and η_i. Moreover, ϵ_i can be correlated with the entire group structure $\boldsymbol{g}$ through the equilibrium cutoffs $p(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta})$, as the latter depend on $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ which include ξ_i and η_i (the general equilibirum effects). Below we give an example to illustrate how ϵ_i is dependent of ξ_i and η_i and how the dependence leads to a selection bias.

Example 3.1 (Example 2.1 continued)**.** In the context of college admissions, ϵ_i represents unobserved ability (IQ and motivation) that affects y_i (labor market outcome). If this ability also affects an student's performance in high school, then ϵ_i is dependent of η_i. Moreover, ϵ_i is dependent of ξ_i if, for example, students of high ability prefer colleges of high (unobserved) quality. In the presence of the dependence, the admission process will sort higher-ability students into more selective colleges. By the same logic, if high SES increases a student's qualifications in college applications and stimulates preferences for colleges of high rankings, then students with higher SES will be sorted into more selective colleges. Therefore, we will observe a positive assortative matching between students and colleges in the sense that students who attend more selective colleges are also of higher ability and SES. The sorting yields a positive correlation between ϵ_i and the average characteristics/outcomes in a college. Without correcting for this sorting effect, we will overestimate the peer effects.

Simulation Evidence To further illustrate the selection bias, we provide simulation evidence using the design in Section 6. We consider both exogenous and endogenous group formation, depending on whether ϵ_i is correlated with η_i or not. Using the simulated data, we display in Figure 3.1 the correlation between the group average characteristic $\bar{x}_{g_i} = \sum_j w_{ij} x_j$ and the unobserved shock ϵ_i for both exogenous and endogenous groups.[11] Under the specification that x_i and z_i are correlated in the same direction as ϵ_i and η_i are correlated, we find that $\bar{x}_{g_i}$ is uncorrelated with ϵ_i in exogenous groups, but positively correlated with ϵ_i in endogenous groups. In the latter case, the OLS estimate of the coefficient of $\bar{x}_{g_i}$ will be upward biased.

Figure 3.1: Correlation Between Group Average Characteristic and Outcome Shock



Note: The figures plot the relationship between the residualized group average characteristic $\bar{x}_{g_i}$ and the residualized outcome shock ϵ_i, using one market in the simulated data in Section 6. In exogenous group formation (left figure), ϵ_i is independent of ξ_i and η_i. In endogenous group formation (right figure), ϵ_i is independent of ξ_i, but correlated with η_i. Other markets in the simulated data show a similar pattern.

3.2 Limiting Approximation

Because equilibrium cutoffs depend on the (observed and unobserved) characteristics of the n individuals in a market, the selection bias in equation (3.1) is a high-dimensional function that involves the observed characteristics of all the n individuals. To reduce its dimensionality, we propose a novel approach that exploits the limiting approximation of the market as n approaches infinity. We find that the correlation between ϵ_i and $\boldsymbol{g}$ through the equilibrium cutoffs becomes negligible as the market

[11]We control for the individual characteristic x_i by regressing $\bar{x}_{g_i}$ (resp. ϵ_i) on x_i and taking the residual of $\bar{x}_{g_i}$ (resp. ϵ_i).

grows large, thereby opening the door to reduce the dimensionality of the selection bias.

To this end, let $p_n = (p_{n,1}, \ldots, p_{n,G})'$ denote a vector of equilibrium cutoffs in a market with n individuals. Azevedo and Leshno (2016) showed that there is a unique stable matching in the limiting market as $n \to \infty$, which can be captured by a unique vector of equilibrium cutoffs in the limiting market, denoted by $p^* = (p_1^*, \ldots, p_G^*)'$. Unlike the finite-$n$ cutoffs p_n, the limiting cutoffs p^* are non-stochastic because they are determined by the distribution of the characteristics and each individual has a negligible impact.

In the following proposition, we follow Azevedo and Leshno (2016) and show that the equilibrium cutoffs in a finite-n market converge to the equilibrium cutoffs in the limiting market as $n \to \infty$. By continuous mapping, the selection bias in a finite-n market also converges to the selection bias in the limiting market.

Proposition 3.1 (Limiting approximation)**.** *Under Assumption 2(i)-(ii), we have*

$$\mathbb{E}[\epsilon_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p_n)] \xrightarrow{p} \mathbb{E}[\epsilon_i|x_i, z_i, g_i(z_i, \xi_i, \eta_i; p^*)]. \tag{3.2}$$

Proof. See Appendix A.2.1. □

The proposition indicates that the selection bias in a finite-n market can be approximated by the selection bias in the limiting market. Because the limiting cutoffs are non-stochastic, the selection bias of individual i in the limiting market depends on i's characteristics only. This reduces the dimensionality of the selection bias from $O(n)$ to a finite number.

In the subsequent analysis, we assume that the selection bias takes the limiting form. Accounting for the sampling error due to the limiting approximation is left to future research.

3.3 Nonparametric Form

Now we derive the selection bias. We start with a nonparametric form where the selection function is group-specific. This group-specific selection, however, poses a challenge in the identification of group-level peer effects. By imposing an additional exchangeability assumption, we can represent the selection bias in an alternative form where the selection function is group-invariant.

3.3.1 Group-specific selection function

Under nonparametric unobservables, the selection bias is a nonparametric function of the group formation indices. Specifically, let $\tau_i = (z_i'\delta_1^u, \ldots, z_i'\delta_G^u, z_i'\delta_1^v, \ldots, z_i'\delta_G^v)' \in \mathbb{R}^{2G}$ denote a vector of preference and qualification indices of individual i.[12] We can represent the selection bias as a group-specific nonparametric function of τ_i.

Proposition 3.2 (Selection bias)**.** *Under Assumption 2, for any $g \in \mathcal{G}$, there exists a function $\lambda_g : \mathbb{R}^{2G} \to \mathbb{R}$ such that $\mathbb{E}[\epsilon_i | x_i, z_i, g_i = g] = \lambda_g(\tau_i)$.*

Proof. See Appendix A.2.1. □

Below we illustrate the selection function $\lambda_g(\cdot)$ in the case of two groups.

Example 3.2. Suppose that there are two groups ($G = 2$) and the group formation indices are $\tau_i = (z_i'\delta_1^u, z_i'\delta_2^u, z_i'\delta_1^v, z_i'\delta_2^v)' \in \mathbb{R}^4$. Denote $\xi_i = (\xi_{i1}, \xi_{i2})'$ and $\eta_i = (\eta_{1i}, \eta_{2i})'$. Let $f(\epsilon_i, \xi_i, \eta_i)$ denote the joint pdf of ϵ_i, ξ_i, and η_i, and $f(\xi_i, \eta_i)$ the joint pdf of ξ_i and η_i. The selection bias of individual i if joining group 1 is

$$\begin{aligned}\mathbb{E}[\epsilon_i | x_i, z_i, g_i = 1] &= \mathbb{E}[\epsilon_i | \eta_{1i} \geq p_1 - z_i'\delta_1^v,\ \xi_{i2} - \xi_{i1} < z_i'(\delta_1^u - \delta_2^u) \text{ or } \eta_{2i} < p_2 - z_i'\delta_2^v] \\ &= \frac{\int_{R_1(\tau_i)} \epsilon_i f(\epsilon_i, \xi_i, \eta_i) \mathrm{d}\epsilon_i \mathrm{d}\xi_i \mathrm{d}\eta_i}{\int_{R_1(\tau_i)} f(\xi_i, \eta_i) \mathrm{d}\xi_i \mathrm{d}\eta_i} =: \lambda_1(\tau_i),\end{aligned}$$

where $R_1(\tau_i)$ denotes the conditioning event $R_1(\tau_i) = \{(\xi_i', \eta_i')' \in \mathbb{R}^4 : \eta_{1i} \geq p_1 - z_i'\delta_1^v \cap (\xi_{i2} - \xi_{i1} < z_i'(\delta_1^u - \delta_2^u) \cup \eta_{2i} < p_2 - z_i'\delta_2^v)\}$. Similarly, the selection bias of individual i if joining group 2 is

$$\begin{aligned}\mathbb{E}[\epsilon_i | x_i, z_i, g_i = 2] &= \mathbb{E}[\epsilon_i | \eta_{2i} \geq p_2 - z_i'\delta_2^v,\ \xi_{i1} - \xi_{i2} < z_i'(\delta_2^u - \delta_1^u) \text{ or } \eta_{1i} < p_1 - z_i'\delta_1^v] \\ &= \frac{\int_{R_2(\tau_i)} \epsilon_i f(\epsilon_i, \xi_i, \eta_i) \mathrm{d}\epsilon_i \mathrm{d}\xi_i \mathrm{d}\eta_i}{\int_{R_2(\tau_i)} f(\xi_i, \eta_i) \mathrm{d}\xi_i \mathrm{d}\eta_i} =: \lambda_2(\tau_i),\end{aligned}$$

where $R_2(\tau_i)$ denotes the conditioning event $R_2(\tau_i) = \{(\xi_i', \eta_i')' \in \mathbb{R}^4 : \eta_{2i} \geq p_2 - z_i'\delta_2^v \cap (\xi_{i1} - \xi_{i2} < z_i'(\delta_2^u - \delta_1^u) \cup \eta_{1i} < p_1 - z_i'\delta_1^v)\}$.

The result in Proposition 3.2 extends the standard sample selection models (Heckman, 1979; Das et al., 2003) to social interaction models with endogenous group

[12]Note that the qualification index of a group matters only if the capacity of the group is binding, If a group does not reach its capacity, then the qualification index of this group should be dropped from τ_i. Specifically, let $\overline{\mathcal{G}} \subseteq \mathcal{G}$ denote the subset of groups whose capacity is binding and $\overline{G}$ the cardinality of $\overline{\mathcal{G}}$. Then $\tau_i = (z_i'\delta_g^u, g \in \mathcal{G}; z_i'\delta_g^v, g \in \overline{\mathcal{G}}) \in \mathbb{R}^{G+\overline{G}}$.

formation. Das et al. (2003) explored a standard sample selection model where the model specification is fully nonparametric. They represented the selection bias as a nonparametric function of the propensity scores of a multivariate selection rule. In our setting, the propensity scores that correspond to the selection rules in equation (2.5) are not available because we do not observe individuals' rankings over the groups or whether they qualify for each group. Instead, we impose an index structure on the preferences and qualifications so that the selection bias can be represented as a function of these group formation indices. Brock and Durlauf (2001, Section 3.6) considered social interactions with endogenous one-sided group formation, where the unobservables follow a parametric distribution. We extend Brock and Durlauf (2001) to more general two-sided group formation with nonparametric unobservables.

Proposition 3.2 and Example 3.2 indicate that the selection function $\lambda_g(\cdot)$ is group-specific for three reasons. First, the cutoffs are group-specific and are absorbed into the selection function. Second, the distribution of the unobservables (ξ_{ig}, η_{gi}) may vary across groups. Third, the components of τ_i for group g and for the other groups $h \neq g$ play different roles in the selection function, through different selection rules as shown in (2.5). The group-specific feature poses a challenge in the identification of group-level peer effects. To see this, let w_i denote the ith row of $\boldsymbol{w}$. If both $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ are group averages that include i herself, they are invariant within a group.[13] The effects of these group-level variables cannot be separately identified from a group-specific nonparametric selection bias. This is similar to the case of panel data models, where the effects of time-invariant variables cannot be separately identified from an individual fixed effect.[14] In the next section, we propose a novel idea to resolve the problem of group-specific selection.

We remark that the selection bias is also individual-specific because it depends on τ_i. While the selection occurs at the entry into a group, individuals with different values of τ_i are subject to different selection biases. Therefore, we cannot correct for

[13] If $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ are group averages that exclude i herself, they converge to including-oneself group averages as the number of group members goes to infinity. Hence, the variation of $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ in a group vanishes to zero as the group size grows.

[14] If we have an additional network within each group, conditional on g_i there may be individual-level variation in $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ because individuals in a group can have different friends. In this case, we can partial out the group-specific selection bias by interacting the indices τ_i with group dummies and exploit the within-group variation in $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ to identify γ. This approach, however, requires further restrictions on network structure as it exploits within-group variation in $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ for identification.

the selection bias by simply introducing group fixed effects. An appropriate correction requires us to exploit the information in τ_i.

3.3.2 Group-invariant selection function

Our idea to tackle the group-specific selection function is motivated by the observation that if the cutoffs are known, by ordering the components of τ_i appropriately, the selection function becomes group-invariant so long as the distribution of the unobservables does not vary across groups. To this end, we introduce group fixed effects to account for group-level heterogeneity and assume that the remaining individual-varying unobservables are exchangeable across groups. If the group fixed effects are also known, we can establish an equivalent representation of the selection bias where the selection function is group-invariant.

As the first step, rewrite the utility u_{ig} in equation (2.3) as

$$u_{ig} = \alpha_g + z_i'\delta_g^u + \xi_{ig}, \tag{3.3}$$

where α_g denotes a group fixed effect. We may also include group fixed effects in the qualification v_{gi}, but they cannot be distinguished from the group-specific cutoffs and thus are normalized to 0. Let $\alpha = (\alpha_1, \ldots, \alpha_G)'$ be a vector of group fixed effects. They can capture vertical preferences over the groups. For example, in college admissions, α may represent college quality, available resources, and reputation. We assume that the individual-varying unobservables are exchangeable across groups.

Assumption 3 (Exchangeability)**.** *The joint pdf of ϵ_i, ξ_i, and η_i satisfies*

$$f(\epsilon_i, \xi_{i1}, \ldots, \xi_{iG}, \eta_{1i}, \ldots, \eta_{Gi}) = f(\epsilon_i, \xi_{ik_1}, \ldots, \xi_{ik_G}, \eta_{k_1 i}, \ldots, \eta_{k_G i}),$$

for any permutation $(k_1, \ldots, k_G)$ of $(1, \ldots, G)$.

Assumption 3 assumes that the joint pdf of the unobservables is invariant under the permutations of the group labels. In other words, the joint distribution of the unobservables does not depend on the order of the groups. Exchangeability has been used in various contexts such as differentiated product markets (Berry et al., 1995; Gandhi and Houde, 2019), panel data (Altonji and Matzkin, 2005), matching (Fox et al., 2018), and network formation (Menzel, 2021, 2022). In our setting, we impose exchangeability so that the unobserved heterogeneity across groups is fully captured by group fixed effects. Note that Assumption 3 is less restrictive than an i.i.d. assumption as it permits the unobservables (ξ_{ig}, η_{gi}) to be dependent across groups, provided the dependence is symmetric. In particular, it allows for an individual effect in ξ_i and η_i.

Given the modified utility in equation (3.3), the selection bias in Proposition 3.2 can be rewritten as

$$\begin{aligned} & \mathbb{E}[\epsilon_i | x_i, z_i, g_i = g] \\ = \; & \mathbb{E}[\epsilon_i | \eta_{gi} \geq p_g - z_i'\delta_g^v, \text{ and } \forall h \neq g \\ & \qquad \xi_{ih} - \xi_{ig} < \alpha_g + z_i'\delta_g^u - \alpha_h - z_i'\delta_h^u, \text{ or } \eta_{hi} < p_h - z_i'\delta_h^v] \\ =: \; & \lambda^e(\tau_i^e(g)), \end{aligned} \tag{3.4}$$

where $\tau_i^e(g) = (\tau_{ig}^{e\prime}, \tau_{i,-g}^{e\prime})' \in \mathbb{R}^{2G}$ is a $2G \times 1$ vector, with the component $\tau_{ig}^e = (\alpha_g + z_i'\delta_g^u, p_g - z_i'\delta_g^v)' \in \mathbb{R}^2$ representing the *extended* indices for group g, and the component $\tau_{i,-g}^e = (\tau_{ih}^{e\prime}, \forall h \neq g)' \in \mathbb{R}^{2G-2}$ representing the *extended* indices for the other $G-1$ groups.[15] Because the cutoffs p and group fixed effects α are absorbed into $\tau_i^e(g)$ and the components for group g are properly separated, under exchangeability the selection function $\lambda^e(\cdot)$ is invariant across groups. In contrast to the expression in Proposition 3.2, the selection bias in equation (3.4) leverages the arguments $\tau_i^e(g)$ rather than the functional form $\lambda^e(\cdot)$ to capture the group-specific feature. This representation utilizes the structure of the selection bias and yields a nonparametric selection function that is group invariant.

As a remark, by exchangeability the order of the extended indices in $\tau_{i,-g}^e$ does not matter – the selection function is symmetric in the extended indices τ_{ih}^e and $\tau_{i\tilde{h}}^e$ for any distinct $h, \tilde{h} \neq g$. This symmetry can further reduce the number of nuisance parameters, which we will delve into further in Section 5.

4 Identification

Moving on to the identification of parameters, we start by discussing the identification of the group formation indices. Then we examine the identification of the peer effects

[15] More accurately, we can define $\tau_i^e(g)$ as a $(2G-1) \times 1$ vector, where we drop the extended utility index for group g and replace the extended utility index for group $h \neq g$ by the extended utility difference between h and g. In the estimation, we will use these utility differences to construct an estimator for the selection bias. See Section 5.2 for more discussions.

with the presence of the selection bias.

4.1 Identification of the Group Formation Indices

Let $\delta = (\delta_1^{u\prime}, \ldots, \delta_G^{u\prime}, \delta_1^{v\prime}, \ldots, \delta_G^{v\prime})'$ collect the slope parameters in group formation and $\theta = (\delta', p', \alpha')'$ all the group formation parameters. The identification of δ was established in He et al. (2022).[16] Hence, the raw indices $\tau_i = \tau(z_i, \delta)$ are identified. To identify the extended indices $\tau_i^e = \tau^e(z_i, g_i, \theta)$, we need to identify p and α in addition to δ, which we discuss below.

4.1.1 Identification of p and α

We start with the following assumption.

Assumption 4. *(i) $p_1 = 0$ and $\alpha_1 = 0$. (ii) The joint cdf of ξ_i and η_i is strictly increasing.*

Part (i) of the assumption is a location normalization because the joint distribution of ξ_i and η_i is fully nonparametric. Part (ii) guarantees a one-to-one relationship between the conditional probability of joining a group and an extended group formation index. It is satisfied for a wide range of common distributions such as normal distributions.

Proposition 4.1 (Identification of p and α)**.** *Suppose that δ is known. Under Assumptions 2-4, p and α are identified.*

Proof. See Appendix A.2.2. □

Proposition 4.1 exploits the idea that under exchangeability, the extended indices for two distinct groups 1 and g have the same impact on the conditional probability of joining a third group $h \neq g, 1$. Therefore, by monotonicity (Assumption 4(ii)) the extended indices that lead to the same conditional probability of joining group h must

[16] A main challenge in identifying δ is that an individual's choice set $\mathcal{C}_i(p)$ may be unobservable to the researchers. To achieve identification, He et al. (2022) utilized excluded variables that act as "demand shifters" and "choice-set shifters". These excluded variables trace out how utilities and qualifications affect the conditional probability of joining each group, respectively. By taking derivatives of the conditional probability of joining each group w.r.t. all the variables, excluded and non-excluded, they derived a system of linear equations that that link the effects of variations in demand and supply. This system of equations enables the nonparametric identification of the model by ensuring the existence of a unique solution.

be equal. We can then recover the difference between the cutoffs of groups 1 and g, and thus identify p_g under location normalization (Assumption 4(i)). The identification of α follows similarly. He et al. (2022) studied a similar setting without exchangeability. They relied on location normalization of the distribution of the unobservables and a large support assumption, so that p and α can be identified by the variation within groups. Our paper complements He et al. (2022) by leveraging the variation across groups under exchangeability to identify p and α.

4.2 Identification of γ

4.2.1 The reflection problem

Define $\nu_i = \epsilon_i - \lambda^e(\tau_i^e)$ to be the residual of ϵ_i after the selection bias is eliminated. We write the i^{th} equation in (2.2) as

$$y_i = w_i\boldsymbol{y}\gamma_1 + w_i\boldsymbol{x}\gamma_2 + x_i'\gamma_3 + \lambda^e(\tau_i^e) + \nu_i, \tag{4.1}$$

where w_i denotes the i^{th} row of the adjacency matrix $\boldsymbol{w}$. Equation (4.1) is a partially linear model (Robinson, 1988). Taking the expectation of equation (4.1) conditional on τ_i^e and subtracting it from equation (4.1) we obtain

$$\tilde{y}_i = \widetilde{w_i\boldsymbol{y}}\gamma_1 + \widetilde{w_i\boldsymbol{x}}\gamma_2 + \tilde{x}_i'\gamma_3 + \nu_i. \tag{4.2}$$

where $\tilde{y}_i = y_i - \mathbb{E}[y_i|\tau_i^e]$ and similarly for other variables. The identification of γ requires that the support of the regressors $\widetilde{w_i\boldsymbol{y}}$, $\widetilde{w_i\boldsymbol{x}}$, and $\tilde{x}_i$ is not contained in a proper linear subspace of $\mathbb{R}^{2d_x+1}$, which by Lemma 4.1 holds if and only if there is no linear combination of $w_i\boldsymbol{y}$, $w_i\boldsymbol{x}$, and x_i that is a function of τ_i^e almost surely. This rank condition prevents any element of $w_i\boldsymbol{y}$, $w_i\boldsymbol{x}$, and x_i being perfectly predictable by τ_i^e (e.g., x_i is a function of τ_i^e or contains a constant). Moreover, the rank condition fails when $w_i\boldsymbol{y}$, $w_i\boldsymbol{x}$, and x_i are linearly dependent, a scenario widely referred to as the reflection problem (Manski, 1993; Brock and Durlauf, 2001).

Lemma 4.1. *Suppose that X_i is a $d \times 1$ vector of variables. The support of $X_i - \mathbb{E}[X_i|\tau_i^e]$ is contained in a proper linear subspace of $\mathbb{R}^d$ if and only if there is a $d \times 1$ vector of constants $k \neq 0$ such that $k'X_i$ is a function of τ_i* with probability 1.

To investigate whether the reflection problem arises in our setting, we consider the social equilibrium in equation (4.1). Let $\boldsymbol{\lambda}^e := \boldsymbol{\lambda}^e(\boldsymbol{\tau}^e) = (\lambda^e(\tau_1^e), \dots, \lambda^e(\tau_n^e))'$ be the $n \times 1$ vector of selection biases, where $\boldsymbol{\tau}^e = (\tau_1^{e\prime}, \dots, \tau_n^{e\prime})'$, and $\boldsymbol{\nu} = (\nu_1, \dots, \nu_n)'$ be the $n \times 1$ vector of residuals after the selection biases are subtracted. Equation (4.1) can be written in a matrix form as

$$\boldsymbol{y} = \boldsymbol{w}\boldsymbol{y}\gamma_1 + \boldsymbol{w}\boldsymbol{x}\gamma_2 + \boldsymbol{x}\gamma_3 + \boldsymbol{\lambda}^e + \boldsymbol{\nu}.$$

Under the assumption that $|\gamma_1| < 1$ and $|||\boldsymbol{w}|||_\infty = \max_{i\in\mathcal{N}} \sum_{j=1}^n |w_{ij}| = 1$, the matrix $I - \gamma_1\boldsymbol{w}$ is invertible. The social equilibrium $\boldsymbol{w}\boldsymbol{y}$ is given by

$$\begin{aligned} \boldsymbol{w}\boldsymbol{y} &= (I - \gamma_1\boldsymbol{w})^{-1}(\boldsymbol{w}^2\boldsymbol{x}\gamma_2 + \boldsymbol{w}\boldsymbol{x}\gamma_3 + \boldsymbol{w}\boldsymbol{\lambda}^e + \boldsymbol{w}\boldsymbol{\nu}) \\ &= \boldsymbol{w}\boldsymbol{x}\gamma_3 + \sum_{k=0}^{\infty} \gamma_1^k \boldsymbol{w}^{k+2}\boldsymbol{x}(\gamma_1\gamma_3 + \gamma_2) + \sum_{k=0}^{\infty} \gamma_1^k \boldsymbol{w}^{k+1}\boldsymbol{\lambda}^e + \sum_{k=0}^{\infty} \gamma_1^k \boldsymbol{w}^{k+1}\boldsymbol{\nu}. \end{aligned} \tag{4.3}$$

The ith equation in (4.3) gives

$$w_i\boldsymbol{y} = w_i\boldsymbol{x}\gamma_3 + \sum_{k=0}^{\infty} \gamma_1^k w_i^{k+2}\boldsymbol{x}(\gamma_1\gamma_3 + \gamma_2) + \sum_{k=0}^{\infty} \gamma_1^k w_i^{k+1}\boldsymbol{\lambda}^e + \sum_{k=0}^{\infty} \gamma_1^k w_i^{k+1}\boldsymbol{\nu}, \tag{4.4}$$

where w_i^k denotes the i^{th} row of the matrix $\boldsymbol{w}^k$. From equation (4.4), $w_i\boldsymbol{y}$, $w_i\boldsymbol{x}$ and x_i are linearly independent if (i) $\gamma_1\gamma_3 + \gamma_2 \neq 0$ and the support of $(x_i, w_i\boldsymbol{x}, w_i^2\boldsymbol{x}, w_i^3\boldsymbol{x}, \dots)$ is not contained in a proper linear subspace of $\mathbb{R}^{2d_x+1}$, or (ii) the support of $(x_i, w_i\boldsymbol{x}, w_i\boldsymbol{\lambda}^e, w_i^2\boldsymbol{\lambda}^e, \dots)$ is not contained in a proper linear subspace of $\mathbb{R}^{2d_x+1}$.

Existing literature on social interactions established numerous conditions for case (i) to hold. For example, $w_i^2\boldsymbol{x}$, $w_i\boldsymbol{x}$, and x_i are linearly independent if there is a network within each group and each network contains an intransitive triad (Bramoullé et al., 2009). If $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ represent group averages that exclude i herself and there is variation in the group sizes, then $w_i^2\boldsymbol{x}$, $w_i\boldsymbol{x}$, and x_i are also linearly independent (Lee, 2007). This source of identification fails, however, if $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ represent group averages that include i herself because $\boldsymbol{w}^2 = \boldsymbol{w}$ (Manski, 1993; Bramoullé et al., 2009).

The presence of selection provides an alternative source of identification through case (ii). It is evident from equation (4.4) that the rank condition for identification holds if $w_i\boldsymbol{\lambda}^e$, $w_i\boldsymbol{x}$, and x_i are linearly independent and, in view of Lemma 4.1, $w_i\boldsymbol{\lambda}^e$ is not perfectly predictable by τ_i^e. This source of identification holds regardless of whether there are within-group networks and whether group averages include oneself.

The result is in accordance with the insight of Brock and Durlauf (2001, Section 3.6) who discovered that identification can be achieved through self-selection so long as the selection correction has within-neighborhood variation. Essentially, the selection bias serves as an individual variable whose average is excluded from the contextual effect.

Near multicollinearity Despite the theoretical prediction that $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ are linearly independent due to the presence of selection, in the case of group averages that include oneself, we find in the simulation study that $w_i\boldsymbol{y}$ and $w_i\boldsymbol{x}$ are nearly multicollinear, though not perfectly multicollinear, when the number of groups is small (e.g., $G = 5$). Note that because $\boldsymbol{w}^2 = \boldsymbol{w}$, the social equilibrium $w_i\boldsymbol{y}$ in this case reduces to

$$w_i\boldsymbol{y} \quad = \quad w_i\boldsymbol{x}\frac{\gamma_2 + \gamma_3}{1 - \gamma_1} + w_i\boldsymbol{\lambda}^e\frac{1}{1 - \gamma_1} + w_i\boldsymbol{\nu}\frac{1}{1 - \gamma_1}. \tag{4.5}$$

The near multicollinearity results from the fact that both $w_i\boldsymbol{\lambda}^e$ and $w_i\boldsymbol{x}$ are group-level averages, so they appear to be highly correlated in data if they are allowed to take only a small number of values. This near multicollinearity imposes an ill-conditioned problem in the estimation of γ, thereby leading to a biased estimate. In the next section, we propose an approach that may reduce the near multicollinearity even if the number of groups is small.

4.2.2 Instrumental variables

Our idea is to use the excluded variables in group formation z_i as an instrument for $w_i\boldsymbol{y}$. The basic intuition is that z_i affects the group that individual i joins and thus the average outcome of her neighbors. Because z_i is individual specific, it brings in variation that is linearly independent of $w_i\boldsymbol{x}$, which then helps alleviate the near multicollinearity problem.

Instrument validity To justify that z_i is a valid instrument, let $X_i = (w_i\boldsymbol{y}, w_i\boldsymbol{x}, x_i')'$ denote the vector of regressors and $Z_i = (z_i', w_i\boldsymbol{x}, x_i')'$ the vector of instruments. Rewrite equation (4.2) as $\tilde{y}_i = \tilde{X}_i'\gamma + \nu_i$. Because $\mathbb{E}[\nu_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{w}] = \mathbb{E}[\epsilon_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{w}] - \lambda^e(\tau_i^e) = 0$, Z_i satisfies the exclusion restriction $\mathbb{E}[Z_i\nu_i] = \mathbb{E}[Z_i\mathbb{E}[\nu_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{w}]] = 0$, that is, Z_i and ν_i are uncorrelated. The textbook literature on instrumental variables (e.g., Wooldridge, 2010) suggests that γ is identified if Assumption 5 is satisfied.

Assumption 5 (Rank). *(i) The matrix* $\mathbb{E}[\tilde{Z}_i\tilde{Z}_i']$ *has full rank. (ii) The matrix* $\mathbb{E}[\tilde{Z}_i\tilde{X}_i']$ *has full column rank.*

Proposition 4.2 (Identification of γ). *Suppose that δ is known. Under Assumptions 1-5, γ is identified.*

Proof. Under Assumption 5, there is a unique γ that satisfies the exclusion restriction $\mathbb{E}[Z_i(\tilde{y}_i - \tilde{X}_i'\gamma)] = 0$. □

Assumption 5(i) is the standard assumption that requires the instruments to be linearly independent. If some components of $\tilde{z}_i$ are linearly dependent (due to the fact that τ_i^e consists of indices of z_i), we can typically choose a subvector of z_i to make this assumption hold. Assumption 5(ii) is the standard rank condition for identification. An immediate implication of this condition is that the support of $\tilde{X}_i$ is not contained in a proper linear subspace of $\mathbb{R}^{2d_x+1}$, the same as discussed in Section 4.2.1. Moreover, the rank condition requires that the linear projection of $\widetilde{w_i\boldsymbol{y}}$ on $\tilde{Z}_i$ must have a nonzero coefficient for $\tilde{z}_i$. In other words, conditional on the other controls, the instrument $\tilde{z}_i$ must be correlated with $\widetilde{w_i\boldsymbol{y}}$, the usual relevance condition for an instrument to be valid. Note that because $\mathbb{E}[\mathbb{E}[Z_i|\tau_i^e]\tilde{X}_i'] = 0$, it is equivalent to using z_i and $\tilde{z}_i$ as the instrument.

It is evident that z_i is correlated with $w_i\boldsymbol{y}$ because z_i affects the group that i joins g_i and thus the average outcome in her neighborhood $w_i\boldsymbol{y}$. What is less obvious is that after controlling for the selection bias, there is still additional variation in z_i that can affect g_i. In fact, under exchangeability the selection bias can be controlled for without fixing the group that an individual joins. In Example 4.1, we provide an illustration that an individual may join different groups under different values of z_i, but the selection bias does not change with the group that she joins, because τ_i^e takes the same value regardless of which group she joins. Therefore, conditional on τ_i^e, there is additional variation in z_i that can affect the group that individual i joins and thus $w_i\boldsymbol{y}$.

Example 4.1. Consider the case of two groups ($G = 2$). Suppose that the utilities are $u_{i1} = \alpha_1 - z_{i1}^u + \xi_{i1}$ and $u_{i2} = \alpha_2 - z_{i2}^u + \xi_{i2}$ and the qualifications are $v_{1i} = z_{i1}^v + \eta_{1i}$ and $v_{2i} = z_{i2}^v + \eta_{2i}$. The group effects take the values $\alpha_1 = 4$ and $\alpha_2 = 2$, and the cutoffs are $p_1 = 2$ and $p_2 = 3$. Consider an individual i with the following values of the observables: $z = (z_1^u, z_2^u, z_1^v, z_2^v) = (2, 1, 0, 2)$, and $\bar{z} = (\bar{z}_1^u, \bar{z}_2^u, \bar{z}_1^v, \bar{z}_2^v) = (5, 2, 1, 1)$. If the unobservables $(\xi_{i1}, \xi_{i2}, \eta_{1i}, \eta_{2i})$ satisfy $-1 < \xi_{i2} - \xi_{i1} < 1$, and $\eta_{1i}, \eta_{2i} > 2$, then individual i joins group 1 when $z_i = z$ and joins group 2 when $z_i = \bar{z}$. In this case, individual i's group formation indices for the two groups take the same value ($\tau_i^e = \bar{\tau}_i^e$) because $\tau_i^e = (p_1 - z_1^v, p_2 - z_2^v, \alpha_1 - z_1^u - (\alpha_2 - z_2^u)) = (2, 1, 1)$ and $\bar{\tau}_i^e = (p_2 - \bar{z}_2^v, p_1 - \bar{z}_1^v, \alpha_2 - \bar{z}_2^u - (\alpha_1 - \bar{z}_1^u)) = (2, 1, 1)$. Conditional on τ_i^e, the additional variation in z_i can drive the individual to join different groups.

Example 4.1 demonstrates that there is variation in $\widetilde{w_i \boldsymbol{y}}$ that is driven by z_i. For the rank condition to hold, we further need that z_i operates not just through $\widetilde{w_i \boldsymbol{x}}$. To verify this, let us take a look at equation (4.5). Partialling out τ_i^e from equation (4.5) yields

$$\widetilde{w_i \boldsymbol{y}} = \widetilde{w_i \boldsymbol{x}} \frac{\gamma_2 + \gamma_3}{1 - \gamma_1} + \widetilde{w_i \boldsymbol{\lambda}^e} \frac{1}{1 - \gamma_1} + w_i \boldsymbol{\nu} \frac{1}{1 - \gamma_1}. \tag{4.6}$$

From the equation we can see that there are two channels for z_i to affect the average outcome $\widetilde{w_i \boldsymbol{y}}$: via affecting the group that i joins, z_i can affect both the average characteristics $\widetilde{w_i \boldsymbol{x}}$ and the average selection $\widetilde{w_i \boldsymbol{\lambda}^e}$ in i's group. Therefore, conditional on $\widetilde{w_i \boldsymbol{x}}$, z_i can influence the average outcome through $\widetilde{w_i \boldsymbol{\lambda}^e}$. Because there are different levels of selection in each group, z_i becomes a relevant instrument for $\widetilde{w_i \boldsymbol{y}}$ by varying it through this additional channel.

Reducing near multicollinearity So far we have focused on the validity of z_i being an instrument. Below we show that using z_i as an instrument brings individual-level variation to the predicted value of $w_i \boldsymbol{y}$, thereby reducing the collinearity with $w_i \boldsymbol{x}$.

To gain insight into this approach, consider the matrix in Assumption 5(ii). Denote $X_i = (w_i \boldsymbol{y}, X_{2i}')'$, where $X_{2i} = (w_i \boldsymbol{x}, x_i')'$. We can write the matrix as

$$\mathbb{E}[\tilde{Z}_i \tilde{X}_i'] = \left(\mathbb{E}[\tilde{Z}_i \widetilde{w_i \boldsymbol{y}}] \quad \mathbb{E}[\tilde{Z}_i \tilde{X}_{2i}'] \right) \tag{4.7}$$

where the first term on the right-hand side represents the first column of the matrix, and the second term represents the remaining $2d_x$ columns. Suppose that the linear projection of $\widetilde{w_i \boldsymbol{y}}$ on $\tilde{Z}_i$ takes the form $\tilde{z}_i' \beta_1 + \tilde{X}_{2i}' \beta_2$. By definition of a linear projection the first column of the matrix is given by

$$\mathbb{E}[\tilde{Z}_i \widetilde{w_i \boldsymbol{y}}] = \mathbb{E}[\tilde{Z}_i \tilde{z}_i']\beta_1 + \mathbb{E}[\tilde{Z}_i \tilde{X}_{2i}']\beta_2 \tag{4.8}$$

Note that the last term in equation (4.8) is a linear combination of the second term on the right-hand side of (4.7). The relevance of the instrument implies that $\beta_1 \neq 0$, so the first term on the right-hand side of (4.8) is present. This term involves individual-specific $\tilde{z}_i$, which is typically linearly independent of $\widetilde{w_i \boldsymbol{x}}$. Due to the presence of this term, the first column of the matrix (4.7) is therefore not likely to be collinear with the remaining columns.

Simulation evidence We provide further evidence on the effectiveness of IV in reducing multicollinearity using simulated data. We use the condition number of a matrix to measure the magnitude of multicollinearity.[17] Using our simulated samples, we calculate the condition numbers of the matrix in equation (4.7) and its counterpart for OLS $\mathbb{E}[\tilde{X}_i \tilde{X}_i']$.[18] Compared with the condition number of the matrix for OLS, which is on average 844, the average condition number of the matrix for IV reduces to 27. This result confirms that using z_i as an instrument can effectively reduce the near multicollinearity.

Remark 4.1. Motivated by equation (4.5), one may attempt to use the average excluded variables in group formation $w_i \boldsymbol{z}$ as an instrument for $\widetilde{w_i \boldsymbol{y}}$. In the case of group averages that include oneself, however, because the instrument $w_i \boldsymbol{z}$ is group-specific, the predicted value of $\widetilde{w_i \boldsymbol{y}}$ is still highly correlated with $\widetilde{w_i \boldsymbol{x}}$ when the number of groups is small. Using $w_i \boldsymbol{z}$ as an instrument does not resolve the near multicollinearity problem.

5 Estimation

In this section, we propose semiparametric methods to estimate the model parameters. We first develop distribution-free estimators for the parameters in group formation.

[17]The condition number of a matrix A is the ratio of the maximal and minimal singular values of A.

[18]To focus on the linear dependence between $w_i \boldsymbol{y}$ and $w_i \boldsymbol{x}$, we partial out x_i and a sieve basis of τ_i^e from $w_i \boldsymbol{y}$, $w_i \boldsymbol{x}$, and z_i, and calculate the condition number of a matrix constructed from their residuals.

Based on these estimators, we then propose a semiparametric two-step GMM estimator for the parameters in social interactions, where we first estimate the selection bias by sieve, and then estimate the social interaction parameters by GMM.

5.1 Estimating p and α in Group Formation

To construct the extended indices τ_i^e in the selection bias, we need to estimate the group formation parameter $\theta = (\delta', p', \alpha')'$. In this section, we discuss how to estimate the intercept parameters p and α. The estimation of the slope parameter δ follows He et al. (2022).

We construct a distribution-free estimator for p and α based on the identification results in Section 3.3.2. For simplicity, we discuss the estimation of p only. The estimation of α can be established similarly. Suppose that we want to estimate the cutoff p_g for group $g \neq 1$. Consider the conditional choice probability of individual i joining another group $h \neq 1, g$ given her qualification index for group g, that is, $\sigma_{h|g}(\tau_{ig}^v) = \mathbb{P}(g_i = h|\tau_{ig}^v)$, where recall that $\tau_{ig}^v = \tau_{ig}^v(\delta_g^v) = z_i'\delta_g^v$ denotes the qualification index for group g. Proposition 4.1 showed that the cutoff p_g can be obtained from the difference $\tau_{ig}^v - \tau_{j1}^v$ for any pair of i and j such that $\sigma_{h|g}(\tau_{ig}^v) = \sigma_{h|1}(\tau_{j1}^v)$ for all $h \neq 1, g$. Consequently, we propose to estimate p_g by a two-step kernel estimator.

Specifically, in the first step, we estimate $\sigma_{h|g}(\tau)$ for $\tau \in \mathbb{R}$ using a kernel estimator

$$\hat{\sigma}_{h|g}(\tau) = \frac{\sum_{i=1}^{n} 1\{g_i = h\} K_1(\frac{\tau - \hat{\tau}_{ig}^v}{\zeta_{1n}})}{\sum_{i=1}^{n} K_1(\frac{\tau - \hat{\tau}_{ig}^v}{\zeta_{1n}})},$$

where $\hat{\tau}_{ig}^v = z_i'\hat{\delta}_g^v$, $\hat{\delta}_g^v$ is an estimator of δ_g^v, $K_1(\cdot)$ is a kernel function, and ζ_{1n} is a bandwidth. Let $\hat{\boldsymbol{\sigma}}_{-\{1,g\}|g} = (\hat{\sigma}_{h|g}, h \neq 1, g)$ denote the vector of estimated $\sigma_{h|g}$ for all $h \neq 1, g$.

In the second step, we estimate p_g by a kernel estimator

$$\hat{p}_g = \frac{1}{n(n-1)(G-2)} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{\substack{h=2 \\ h \neq g}}^{G} \frac{1}{\zeta_{2n}} K_2(\frac{\hat{\sigma}_{h|g}(\hat{\tau}_{ig}^v) - \hat{\sigma}_{h|1}(\hat{\tau}_{j1}^v)}{\zeta_{2n}})(\hat{\tau}_{ig}^v - \hat{\tau}_{j1}^v),$$

where $K_2(\cdot)$ is a kernel function, and ζ_{2n} is a bandwidth. The kernel function is to approximate the criterion that a qualifying pair of i and j must satisfy $\sigma_{h|g}(\tau_{ig}^v) = \sigma_{h|1}(\tau_{j1}^v)$ for all $h \neq 1, g$. The estimator is similar in spirit to a propensity-score matching estimator.

To establish the asymptotic normality of $\hat{p}_g$, we impose the following regularity conditions:

Assumption 6 (First-step kernel). *(i)* $\int K_1(t)dt = 1$, *and for a positive integer* s_1 *and all* $j < s_1$, $\int K_1(t)t^j dt = 0$. *(ii)* $K_1(t)$ *is twice continuously differentiable. (iii)* $K_1(t)$ *is zero outside a bounded set.*

Assumption 7 (Second-step kernel). *(i)* $\int K_2(t)dt = 1$, *and for a positive integer* $s_2 \geq s_1$ *and all* $j < s_2$, $\int K_2(t)t^j dt = 0$. *(ii)* $K_2(t)$ *is twice continuously differentiable. (iii)* $K_2(t)$ *is zero outside a bounded set.*

Assumption 8. *For any* $g \in \mathcal{G}$, *(i) the pdf of* τ_{ig}^u (τ_{ig}^v) *is bounded away from zero on the support of* τ_{ig}^u (τ_{ig}^v). *(ii) the pdf of* τ_{ig}^u (τ_{ig}^v) *is* (s_2+1)*th continuously differentiable. (iii)* $\sigma_{h|g}(\tau)$ *is* (s_2+2)*th continuously differentiable.*

Assumption 9 (Bandwidth). *Define* $r_0 = (\ln n)^{1/2}(n\zeta_{1n})^{-1/2} + \zeta_{1n}^{s_1}$ *and* $r_1 = (\ln n)^{1/2}(n\zeta_{1n}^3)^{-1/2} + \zeta_{1n}^{s_1}$. *The bandwidths* $\zeta_{1n} \to 0$ *and* $\zeta_{2n} \to 0$ *satisfy (i)* $n^{1/2}\zeta_{2n}^{-3}r_0^2 \to 0$, *(ii)* $\zeta_{2n}^{-3}r_0 \to 0$ *and* $\zeta_{2n}^{-2}r_1 \to 0$, *(iii)* $n^{1/2}\zeta_{1n}\zeta_{2n}^2 \to \infty$, *(iv)* $n^{1/2}\zeta_{1n}^{s_1} \to 0$, *(v)* $n^{1/2}\zeta_{2n}^{s_2} \to 0$.

Assumption 10. *(i) The parameter* δ_0 *is an interior point in a compact space* Δ. *(ii)* $\hat{\delta} - \delta = n^{-1}\sum_{i=1}^n \psi^\delta(z_i) + o_p(n^{-1/2})$, where $\mathbb{E}[\psi^\delta(z_i)] = 0$.

Assumption 11 (Compactness). *(i) The variable* z_i *has a bounded support. (ii) The variable* x_i *has a bounded support.*

Assumption 6 and 7 impose regular conditions on the kernel functions. Assumption 9 guarantees that the bias introduced by the kernel estimators in both the first and second steps are of smaller order than $\sqrt{n}$.[19] Assumption 10 is satisfied when δ^v is parametrically or semiparametrically estimated.[20]

The following proposition shows the asymptotic distribution of $\hat{p}_g$.

[19] Assumption 9 holds when, for example, $s_1 = 3$, $s_2 = 5$, $\zeta_{1n} = O(n^{-\frac{11}{60}})$, and $\zeta_{2n} = O(n^{-\frac{5}{48}})$.

[20] See He et al. (2022) for discussions on a semiparametric GMM estimator based on average derivative estimators (Powell et al., 1989).

Proposition 5.1 (Asymptotic distribution of $\hat{p}_g$)**.** *Under Assumptions 2-4, 6-10, and 11(i), we have*

$$\sqrt{n}(\hat{p}_g - p_g) \xrightarrow{d} N(0, V_g),$$

where $V_g = Var[\psi_g(z_i, g_i)]$, where $\psi_g(z_i, g_i)$ is defined in proof.

Proof. See Appendix A.2.3. □

5.2 Estimating γ in Social Interactions

5.2.1 Symmetry of the selection correction

To estimate γ in equation (4.1), we need to partial out the selection bias $\lambda^e(\tau_i^e)$. In Section 3.3.2 we showed that the selection correction function $\lambda^e(\cdot)$ is symmetric in the index pairs in $\tau_{i,-g_i}^e$. This symmetry property allows us to reduce the number of nuisance parameters in the estimation and improve efficiency.

To account for the symmetry of the selection correction function, we consider the elementary symmetric functions of the extended indices in $\tau_{i,-g_i}^e$, denoted by $\tau_{i,-g_i}^s$. Following Altonji and Matzkin (2005), $\tau_{i,-g_i}^s$ is a tuple that consists of (i) the elementary symmetric functions of the extended utility indices $\tau_{ig_i}^{e,u} - \tau_{ih}^{e,u}$ for $h \neq g_i$ up to order $G-1$, that is, $\sum_{h_1 \neq g_i}(\tau_{ig_i}^{e,u} - \tau_{ih}^{e,u})$, $\sum_{(h_1,h_2)\neq g_i}(\tau_{ig_i}^{e,u} - \tau_{ih_1}^{e,u})(\tau_{ig_i}^{e,u} - \tau_{ih_2}^{e,u})$,$\ldots$,$\sum_{(h_1,\ldots,h_{G-1})\neq g_i}\prod_{k=1}^{G-1}(\tau_{ig_i}^{e,u} - \tau_{ih_k}^{e,u})$, and (ii) the elementary symmetric functions of the extended qualification indices $\tau_{ih}^{e,v}$ for $h \neq g_i$ up to order $G-1$, that is, $\sum_{h_1 \neq g_i}\tau_{ih_1}^{e,v}$, $\sum_{(h_1,h_2)\neq g_i}\tau_{ih_1}^{e,v}\tau_{ih_2}^{e,v}$,$\ldots$,$\sum_{(h_1,\ldots,h_{G-1})\neq g_i}\prod_{k=1}^{G-1}\tau_{ih_k}^{e,v}$, where $\sum_{(h_1,\ldots,h_k)\neq g_i}$ denotes the summation over all combinations of distinct $h_1, h_2, \ldots, h_k$ in $\mathcal{G}\backslash\{g_i\}$. By the fundamental theorem of symmetric functions and the Weierstrass approximation theorem, any symmetric function can be approximated arbitrarily closely by a polynomial function of the elementary symmetric functions (Altonji and Matzkin, 2005). Therefore, there is a function λ^s such that $\lambda^e(\tau_i^e) = \lambda^s(\tau_i^s)$, where $\tau_i^s = (\tau_{ig_i}^{e,v}, \tau_{i,-g_i}^s)$.

Using the symmetric representation of the selection bias can substantially reduce the number of nuisance parameters in a sieve approximation. For example, if we consider linear basis functions, the number of approximating functions in $\lambda^e(\tau_i^e)$ is $2G-1$, while the number of approximating functions in $\lambda^s(\tau_i^s)$ is only 3 – one functions for group g_i and the other two functions for the remaining $G-1$ groups combined. For basis functions of order two, the number of approximating functions in $\lambda^e(\tau_i^e)$ is

$(G+1)(2G-1)$ and 12 in $\lambda^s(\tau_i^s)$.[21]

5.2.2 Estimation of γ

Rewrite equation (4.1) using the symmetric representation of the selection bias $\lambda^s(\tau_i^s)$. Partialing out the selection bias yields $y_i - \mathbb{E}[y_i|\tau_i^s] = (X_i - \mathbb{E}[X_i|\tau_i^s])'\gamma + \nu_i$. From the independence of Z_i and ν_i, the true parameter γ_0 satisfies the moment condition

$$\begin{aligned} 0 &= \mathbb{E}[Z_i(y_i - \mathbb{E}[y_i|\tau_i^s] - (X_i - \mathbb{E}[X_i|\tau_i^s])'\gamma)] \\ &= \mathbb{E}[(Z_i - \mathbb{E}[Z_i|\tau_i^s])(y_i - X_i'\gamma)], \end{aligned} \tag{5.1}$$

where the last line follows from the law of iterated expectations.[22]

Based on the moment condition in equation (5.1), we propose a three-step semiparametric estimator for γ. In the first step, we estimate the group formation parameters $\theta_0 = (\delta_0', p^{*\prime}, \alpha_0')'$ by the estimator $\hat{\theta} = (\hat{\delta}', \hat{p}', \hat{\alpha}')'$ as discussed in Section 5.1. Note that τ_i^s can be represented as $\tau_i^s = \tau^s(z_i, g_i, \theta_0)$. Denote its estimator by $\hat{\tau}_i^s = \tau^s(z_i, g_i, \hat{\theta})$.

In the second step, we estimate the nuisance parameter $\mu_0^Z(\tau_i^s) = \mathbb{E}[Z_i|\tau_i^s]$ by a sieve estimator $\hat{\mu}^Z(\hat{\tau}_i^s)$. Specifically, define $b^K(\tau_i^s) = (b_{1K}(\tau_i^s), \ldots, b_{KK}(\tau_i^s))'$ to be a $K \times 1$ vector of approximating functions, and let $B_K(\boldsymbol{\tau}^s) = (b^K(\tau_1^s), \ldots, b^K(\tau_n^s))'$ be the $n \times K$ matrix of approximating functions for all i. We estimate $\mu_0^Z(\tau_i^s)$ by the predicted value from the regression of $\boldsymbol{Z} = (Z_1, \ldots, Z_n)'$ on the estimated approximating functions $\hat{B}_K = (b^K(\hat{\tau}_1^s), \ldots, b^K(\hat{\tau}_n^s))'$. That is,

$$\hat{\mu}^Z(\hat{\tau}_i^s) = \boldsymbol{Z}'\hat{B}_K(\hat{B}_K'\hat{B}_K)^{-1}b^K(\hat{\tau}_i^s).$$

In the third step, we estimate the parameter γ by GMM. Let $\omega_i = (y_i, X_i, Z_i)$ and

[21] For $\lambda^e(\tau_i^e)$, there are $G(2G-1)$ functions of order two: $2G-1$ squared indices and $(2G-1)(G-1)$ interactions between indices. For $\lambda^s(\tau_i^s)$, we have 9 functions of order two: three terms that involve $\tau_{ig}^{e,v}$: $(\tau_{ig}^{e,v})^2$, $\tau_{ig}^{e,v}\sum_{h\neq g}(\tau_{ih}^{e,u} - \tau_{ig}^{e,u})$, and $\tau_{ig}^{e,v}\sum_{h\neq g}\tau_{ih}^{e,v}$, three monomial symmetric functions of the form $m(2,0)$: $\sum_{h\neq g}(\tau_{ih}^{e,u} - \tau_{ig}^{e,u})^2$, $\sum_{h\neq g}(\tau_{ih}^{e,v})^2$, and $\sum_{h\neq g}(\tau_{ih}^{e,u} - \tau_{ig}^{e,u})\tau_{ih}^{e,v}$, and three monomial symmetric functions of the form $m(1,1)$: $\sum_{(h,h')\neq g}(\tau_{ih}^{e,u} - \tau_{ig}^{e,u})(\tau_{ih'}^{e,u} - \tau_{ig}^{e,u})$, $\sum_{(h,h')\neq g}\tau_{ih}^{e,v}\tau_{ih'}^{e,v}$, and $\sum_{(h,h')\neq g}(\tau_{ih}^{e,u} - \tau_{ig}^{e,u})\tau_{ih'}^{e,v}$.

[22] Note that $\mathbb{E}[Z_i(y_i - \mathbb{E}[y_i|\tau_i^s] - (X_i - \mathbb{E}[X_i|\tau_i^s])'\gamma] = \mathbb{E}[Z_i(y_i - X_i'\gamma)] - \mathbb{E}[Z_i\mathbb{E}[y_i - X_i'\gamma|\tau_i^s]] = \mathbb{E}[Z_i(y_i - X_i'\gamma)] - \mathbb{E}[\mathbb{E}[Z_i|\tau_i^s]\mathbb{E}[y_i - X_i'\gamma|\tau_i^s]] = \mathbb{E}[(Z_i - \mathbb{E}[Z_i|\tau_i^s])(y_i - X_i'\gamma)]$.

$m(\omega_i, \gamma, \mu_0^Z(\tau_i^s))$ denote the (infeasible) moment function

$$m(\omega_i, \gamma, \mu_0^Z(\tau_i^s)) = (Z_i - \mu_0^Z(\tau_i^s))(y_i - X_i'\gamma).$$

Define $m_0(\gamma, \mu^Z) = \mathbb{E}[m(\omega_i, \gamma, \mu^Z(\tau_i^s))]$ and $\hat{m}_n(\gamma, \hat{\mu}^Z) = \frac{1}{n}\sum_{i=1}^n m(\omega_i, \gamma, \hat{\mu}^Z(\hat{\tau}_i^s))$. Let W be a $d_z \times d_z$ weighting matrix and $\hat{W}$ a consistent estimator of W. We obtain a GMM estimator $\hat{\gamma}$ by solving

$$\min_{\gamma\in\Gamma} \hat{m}_n(\gamma, \hat{\mu}^Z)'\hat{W}\hat{m}_n(\gamma, \hat{\mu}^Z). \tag{5.2}$$

To establish the asymptotic properties of the GMM estimator, we impose the following assumptions:

Assumption 12. *(i) The parameter θ_0 is an interior point in a compact space Θ. (ii) $\hat{\theta} - \theta_0 = n^{-1}\sum_{i=1}^n \phi_\theta(z_i, \theta_0) + o_p(n^{-1/2})$, where $\mathbb{E}[\phi_\theta(z_i, \theta_0)] = 0$ and $\mathbb{E}[\phi_\theta(z_i, \theta_0)^2] < \infty$.*[23]

Assumption 13 (Sieve)**.** *Let $K \to \infty$ and $K/n \to 0$. The basis functions $b^K(\tau) \in \mathbb{R}^K$ satisfy the following conditions. (i) $\mathbb{E}[b^K(\tau)b^K(\tau)'] = I_K$.*[24] *(ii) There exist β^Z and a constant $a > 0$ such that $\sup_\tau|\mu_0^Z(\tau) - b^K(\tau)'\beta^Z| = O(K^{-a})$. (iii) $\sup_\tau\|b^K(\tau)\| \leq \varrho_0(K)$ for a sequence of constants $\varrho_0(K)$ such that $\varrho_0(K)^2K/n \to 0$. (iv) $\sup_\tau \|\partial b^K(\tau)/\partial\tau'\| \leq \varrho_1(K)$ for a sequence of constants $\varrho_1(K)$ such that $\varrho_1(K)/\sqrt{n} \to 0$.*

Assumption 14 (Adjacency matrix)**.** *The adjacency matrix $\boldsymbol{w} = (w_{ij}) \in \mathbb{R}_+^{n^2}$ satisfies the following conditions. (i) $\|\|\boldsymbol{w}\|\|_\infty = \max_{i\in\mathcal{N}}\sum_{j=1}^n |w_{ij}| = 1$. (ii) $\mathbb{E}[\|\boldsymbol{w}\|_\infty^4] = O(n^{-4})$. (iii) There exist i.i.d. ς_i, $i \in \mathcal{N}$, such that (a) $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g})$ is a function of $\boldsymbol{\varsigma} = (\varsigma_i, i \in \mathcal{N})$, (b) $\mathbb{E}[\nu_i|\boldsymbol{\varsigma}] = 0$, and (c) $\boldsymbol{w}$ is independent of $\boldsymbol{\nu}$ conditional on $\boldsymbol{\varsigma}$. (iv) For $h_{ij} = h(\varsigma_i, \varsigma_j) \in \mathbb{R}$ such that $\max_{i,j\in\mathcal{N}}|h_{ij}| < \infty$ and $t = 0, 1$,*

[23] This assumption is satisfied by Assumption 10 and Proposition 5.1.

[24] Assumption 13(i) is a normalization. In fact, we can assume alternatively that the smallest eigenvalue of $\mathbb{E}[b^K(\tau)b^K(\tau)']$ is bounded away from zero uniformly in K. Under this assumption, denote $Q_0 = \mathbb{E}[b^K(\tau)b^K(\tau)']$ and let $Q_0^{-1/2}$ be the symmetric square root of Q_0^{-1}. Then $\tilde{b}^K(\tau) = Q_0^{-1/2}b^K(\tau)$ is a nonsingular transformation of $b^K(\tau)$ that satisfies $\mathbb{E}[\tilde{b}^K(\tau)\tilde{b}^K(\tau)'] = I_K$. Note that nonparametric series estimators are invariant w.r.t. nonsingular transformations of $b^K(\tau)$ – let $\tilde{\beta}^Z = Q_0^{1/2}\beta^Z$ then $\tilde{b}^K(\tau)'\tilde{\beta}^Z = b^K(\tau)'\beta^Z$. Further, $\tilde{b}^K(\tau)$ satisfies Assumption 13(iii)(iv) if and only if $b^K(\tau)$ does. Thus, all parts of Assumption 13 hold with $b^K(\tau)$ replaced by $\tilde{b}^K(\tau)$ (Li and Racine, 2007, p.480).

$\max_{r,\tilde{r}\geq 1} \max_{i,j,k.l\in\mathcal{N}:\{i,j\}\cap\{k,l\}=\emptyset} |Cov(h_{ij}((\boldsymbol{w}^t)'\boldsymbol{w}^r)_{ij}, h_{kl}((\boldsymbol{w}^t)'\boldsymbol{w}^{\tilde{r}})_{kl})| = o(n^{-2})$. *(v) For K that satisfies Assumption 13,* $\max_{i,j,k,l\in\mathcal{N}:\{i,j\}\cap\{k,l\}=\emptyset} \mathbb{E}[Cov(w_{ij}, w_{kl}|\boldsymbol{\varsigma})^2] = o(n^{-4}/K)$ *and* $\max_{i,j\in\mathcal{N}} \mathbb{E}[(\mathbb{E}[w_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[w_{ij}|\varsigma_i, \varsigma_j])^4] = o(n^{-4}/K^2)$. *(vi) For $\{i,j\}$ and $\{k,l\}$ that overlap in one element,* $\max_{i,j,k,l\in\mathcal{N}:|\{i,j\}\cap\{k,l\}|=1} \mathbb{E}[Cov(w_{ij}, w_{kl}|\boldsymbol{\varsigma})^2] = o(n^{-4})$.

Assumption 15 (GMM)**.** *(i) $\hat{W} \xrightarrow{p} W$. W is positive semi-definite and bounded, and $Wm_0(\gamma, \mu_0^Z) \neq 0$ for all $\gamma \neq \gamma_0$. (ii) The parameter γ_0 is an interior point in a compact space Γ. (iii) $M_n'WM_n$ is nonsingular for $M_n = \mathbb{E}[\frac{\partial \hat{m}_n(\gamma,\mu_0^Z)}{\partial \gamma'}]$.*[25]

Assumption 16. *(i) The unobservable ϵ_i has finite fourth moment. (ii) For any $\theta \in \Theta$, $\mathbb{E}[Z_i|\tau^s(z_i, g_i, \theta)]$ and $\mathbb{E}[\epsilon_i|\tau^s(z_i, g_i, \theta)]$ are continuously differentiable in $\tau^s(z_i, g_i, \theta)$.*

We establish the asymptotics of $\hat{\gamma}$ based on Newey (1994a) and Hahn and Ridder (2013). The main challenge in deriving the asymptotics involves dependency of $w_i\boldsymbol{x}$ and $w_i\boldsymbol{y}$ due to the adjacency matrix $\boldsymbol{w}$. Different from the literature that study deterministic $\boldsymbol{w}$, we consider $\boldsymbol{w}$ to be random and correlated with the observables, with its randomness originating from the randomness of the group links and within-group networks. Assumption 14 imposes a set of sufficient conditions on a dense network that ensure the correlation within the network vanishes as n grows large. Specifically, part (iv) of the assumption addresses the dependency of $w_i\boldsymbol{y}$ by restricting the covariance of higher-order adjacency matrices. Part (v) is necessary to establish the consistency of the sieve estimator, while part (vi) is used for deriving the asymptotic distribution. Our proof of asymptotics generalize Lee (1990, Secion 3.7.5) for weighted U-statistics to accommodate random weights.

Theorem 5.1 (Consistency of $\hat{\gamma}$)**.** *Under Assumptions 1–5, 11–16, we have $\hat{\gamma} - \gamma_0 = o_p(1)$.*

Proof. See Appendix A.2.3. □

Theorem 5.2 (Asymptotic distribution of $\hat{\gamma}$)**.** *Under Assumptions 1–5, 11–16, we have $\sqrt{n}\Sigma_n^{-1/2}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, I_{d_\gamma})$, where the variance Σ_n is defined in the proof.*

Proof. See Appendix A.2.3. □

[25]Note that $M_n = -n^{-1}\sum_{i=1}^n \mathbb{E}[(Z_i - \mu_0^Z(\tau_i))X_i']$ does not depends on γ.

6 Simulations

6.1 Setup

In this section, we conduct a simulation study to evaluate the performance of the estimators. We generate a market with $n = 2000$ individuals. They form $G = 5$ groups following the model in Section 2, where the capacities of the groups are $\{340, 320, 340, 320, 340\}$, implying a total of $1,660$ seats. After the groups are formed, they interact according to the social interaction model

$$y_i = \gamma_1 w_i \boldsymbol{y} + w_i \boldsymbol{x} \gamma_2 + x_i \gamma_3 + \epsilon_i, \tag{6.1}$$

where ϵ_i is i.i.d. $N(0,1)$, x_i is i.i.d. $N(5,25)$, and the social interaction parameters are $\gamma = (\gamma_1, \gamma_2, \gamma_3) = (0.5, 1, 1)$. We consider two scenarios of the adjacency matrix w: (i) pure groups; and (ii) additional networks within groups. In the case of pure groups, $w_i \boldsymbol{y}$ and $w_i \boldsymbol{x}$ are group means. When there is an additional network within each group, $w_i \boldsymbol{y}$ and $w_i \boldsymbol{x}$ represent the averages among friends. We generate a network as follows: for individuals i and j in a group, draw ζ_{ij} from $U[0,1]$, and the link ij is formed if and only if $\zeta_{ij} \geq 0.5$. The weights w_{ij} in w_i are given by $w_{ij} = \frac{d_{ij}}{d_i}$, where $d_{ij} = 1\{\zeta_{ij} \geq 1/2\} 1\{g_i = g_j\}$, and $d_i = \sum_{j \neq i} d_{ij}$.

In group formation, we specify the utility and qualification of an individual i for group g as follows

$$\begin{aligned} u_{ig} &= \alpha_g + \delta_1^u z_{1,ig}^u + \delta_2^u z_{2,i} + \xi_{ig} \\ v_{gi} &= \delta_1^v z_{1,ig}^v + \delta_2^v z_{2,i} + \eta_{gi}, \end{aligned}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (9, 6, 4, 2, 0)$ are the group fixed effects, and the group formation parameters are $\delta = (\delta_1^u, \delta_2^u, \delta_1^v, \delta_2^v) = (-1, 1, 1, 1)$. Both $z_{1,ig}^u$ and $z_{1,ig}^v$ are individual-and-group-specific characteristics that are i.i.d. across individuals and groups with the distribution $N(0, 9)$. $z_{2,i}$ denotes an individual-specific characteristic that we allow to be correlated with x_i. In particular, we assume that $z_{2,i} = \log(x_i + q_i + 20)$, where q_i follows i.i.d. $N(0, 4)$. Let $z_i = (z_{1,ig}^u, z_{1,ig}^v, z_{2,i})$. If an individual i does not join any group, her utility is ξ_{i0}. The utility shocks ξ_{ig} is i.i.d. across i and $g \in \mathcal{G} \cup \{0\}$ with the type I extreme value distribution . The preference shocks η_{gi} is i.i.d. with the distribution $N(\epsilon_i, 1)$, and thus is correlated with ϵ_i.

We simulate 200 markets independently. For each market, we calculate the group formation outcomes using the individual-proposing Deferred-Acceptance algorithm. The capacity constraint is binding for all groups. The estimation of group formation parameters are discussed in Appendix O.A.

6.2 Results

Panel A of Table 6.1 presents the estimation results for the case of pure groups in the absence of endogenous social interactions (i.e., $\gamma_1 = 0$). The estimates of γ_2 by OLS (Column 1) and OLS with school fixed effects (Column 2) are upward biased, though the latter is less biased. Controlling for school fixed effects tends to eliminate some selection bias, but not all. In Column 3, we use a polynomial series of the extended group formation indices to correct for the selection bias. To account for the symmetry of the selection bias, we combine the indices for groups other than the group an individual joins using the elementary symmetric functions. The sieve OLS estimates of γ tend to be unbiased. This demonstrates that the control function we construct performs well in approximating the selection bias.

Panel A of Table 6.2 presents the estimation results for the case of pure groups when endogenous social interactions are present (i.e., $\gamma_1 \neq 0$). The OLS estimates of γ_1 and γ_2 (Columns 1 and 2) are heavily biased due to the endogeneity in groups and the reflection problem. In Column 3, we correct for the selection bias using the polynomial series . The sieve OLS estimates are still biased as a result of the reflection problem. In Columns 4 and 5, in addition to correcting for the selection bias as in Column 3, we also instrument for $\widetilde{w_i \boldsymbol{y}}$ and estimate the parameters by sieve 2SLS. Column 4 shows that using $\widetilde{w_i \boldsymbol{z}}$ (average z of group members, partialling out τ_i^s) as instruments does not resolve the reflection problem – the estimates of γ_2 and γ_3 are similar to those by sieve OLS. F test of $\widetilde{w_i \boldsymbol{z}}$ implies perfect multicolinearity between $\widetilde{w_i \boldsymbol{y}}$ and $\widetilde{w_i \boldsymbol{z}}$. Therefore, $\widetilde{w_i \boldsymbol{z}}$ contains exactly the same information as $\widetilde{w_i \boldsymbol{y}}$, leading to identical results between sieve OLS and sieve 2SLS using $\widetilde{w_i \boldsymbol{z}}$ as instruments. In contrast, Column 5 shows that using $\tilde{z}_i$ (own z, partialling out τ_i^s) as instruments performs well: the sieve 2SLS estimates of γ seem unbiased. This suggests that using one's own excluded variables in group formation as instruments can resolve the reflection problem. Additionally, F test suggests that $\tilde{z}_i$ is strongly correlated with

Table 6.1: Estimation Results: *Without* Endogenous Interactions ($\gamma_1 = 0$)

		OLS	OLS	Sieve OLS
Panel A: Pure groups				
γ_2	bias	0.363	0.127	-0.004
	sd	0.397	1.067	0.122
	rmse	0.537	1.072	0.122
γ_3	bias	-0.005	-0.005	-0.001
	sd	0.005	0.005	0.005
	rmse	0.007	0.007	0.005
Panel B: Network				
γ_2	bias	0.166	0.046	-0.002
	sd	0.140	0.106	0.058
	rmse	0.217	0.115	0.058
γ_3	bias	-0.004	-0.005	-0.001
	sd	0.005	0.005	0.005
	rmse	0.007	0.007	0.005
Selection control?		No	Group FE	Sieve

Note: This table presents estimates for the coefficients of the social interactions model without endogenous social interactions. The coefficients are estimated using 200 MC samples, where each sample contains 5 groups and 1,660 individuals who join one of the groups. The basis functions include a polynomial series (up to order 2) of the extended qualification indices and differences in the extended preference indices. We combine the indices for groups other than the group an individual joins using the elementary symmetric functions.

Table 6.2: Estimation Results: *With* Endogenous Interactions ($\gamma_1 \neq 0$)

		OLS	OLS	Sieve OLS	Sieve 2SLS	Sieve 2SLS
Panel A: Pure groups						
γ_1	bias	0.500	-0.358	0.059	0.059	-0.022
	sd	0.000	2.799	0.118	0.118	0.236
	rmse	0.500	2.815	0.131	0.131	0.236
γ_2	bias	-1.995	0.749	-0.267	-0.267	0.098
	sd	0.005	18.300	0.560	0.560	1.115
	rmse	1.995	18.270	0.619	0.619	1.116
γ_3	bias	-0.005	-0.005	-0.001	-0.001	-0.001
	sd	0.005	0.005	0.005	0.005	0.005
	rmse	0.007	0.007	0.005	0.005	0.005
F-stat of instruments					Inf[a]	74.973
p-value					0.000	0.000
Over-identification stat					2.353	9.572
p-value					0.959	0.521
Panel B: Network						
γ_1	bias	0.168	-0.812	0.003	0.010	-0.018
	sd	0.060	0.416	0.041	0.041	0.084
	rmse	0.178	0.912	0.041	0.042	0.085
γ_2	bias	-0.271	0.792	-0.013	-0.028	0.037
	sd	0.131	0.414	0.106	0.106	0.194
	rmse	0.301	0.894	0.106	0.109	0.196
γ_3	bias	-0.005	-0.005	-0.001	-0.001	-0.001
	sd	0.005	0.005	0.005	0.005	0.005
	rmse	0.008	0.007	0.005	0.005	0.005
F-stat of instruments					4835.600	49.192
p-value					0.000	0.000
Over-identification stat					10.424	8.721
p-value					0.468	0.580
Selection control?		No	Group FE	Sieve	Sieve	Sieve
Instruments?		No	No	No	$\widetilde{w_i \boldsymbol{z}}$	$\widetilde{z}_i$

Note: This table presents estimates for the coefficients of the social interactions model with endogenous social effects. The coefficients are estimated using 200 MC samples, where each sample contains 5 groups and 1,660 individuals who join one of the groups. The basis functions include a polynomial series (up to order 2) of the extended qualification indices and differences in the extended preference indices. We combine the indices for groups other than the group an individual joins using the elementary symmetric functions.

a. F statistic is infinity in 105 (out of 200) samples, while the mean of F statistic in the remaining samples is 8×10^{13}. This is due to perfect multicolinearity between $\widetilde{w_i \boldsymbol{y}}$ and $\widetilde{w_i \boldsymbol{z}}$. In this case, the sieve 2sls using $\widetilde{w_i \boldsymbol{z}}$ reduces to sieve ols.

$\widetilde{w_i \boldsymbol{y}}$, and the over-identification test indicates that we can not reject the exogeneity of $\tilde{z}_i$.

As a comparison, we repeat the exercises above for the case where there is an additional network within each group. The estimation results are presented in Panel B of Tables 6.1 and 6.2. Because the network within each group generates additional individual-level variation in $w_i \boldsymbol{y}$ and $w_i \boldsymbol{x}$, the model does not suffer severely from the reflection problem. In particular, sieve OLS, sieve 2SLS using $\widetilde{w_i \boldsymbol{z}}$ as instruments, and sieve 2SLS using $\tilde{z}_i$ as instruments yield similar results. Moreover, F test and over-identification indicate validity of both $\widetilde{w_i \boldsymbol{z}}$ and $\tilde{z}_i$ as instruments.

7 Conclusion

This paper considers social interaction models with endogenous non-overlapping group formation. The endogeneity is due to the correlation between the unobservables in the group formation and the unobservable in the outcome equation. In this paper, we derive a tractable expression of the correction term by exploiting a many-to-one matching framework to characterize group formation. We characterize the selection bias as a nonparametric function of group formation indices. We show identification results of the model and propose a multi-stage estimation strategy.

A Appendix

A.1 Adjacency Matrix: Examples

In this section, we verify Assumption 14 for several adjacency matrices that are widely used in the literature.

Example A.1 (Group averages including oneself)**.** Suppose that $\boldsymbol{w}$ represents group averages that include oneself and the group capacities are binding. We can write $w_{ij} = \sum_{g=1}^{G} \frac{1}{n_g} 1\{g_i = g\} 1\{g_j = g\}$. By construction, $|||\boldsymbol{w}|||_\infty = \max_{i \in \mathcal{N}} \sum_{j=1}^{n} |w_{ij}| = 1$ and $\|\boldsymbol{w}\|_\infty = \max_{i,j \in \mathcal{N}} |w_{ij}| \leq \frac{1}{\min_{g \in \mathcal{G}} n_g} = \frac{1}{n \min_{g \in \mathcal{G}} r_g}$, where $r_g = \frac{n_g}{n} > 0$ for $g \in \mathcal{G}$. Hence, $\mathbb{E}[\|\boldsymbol{w}\|_\infty^4] \leq \frac{1}{n^4 \min_{g \in \mathcal{G}} r_g^4} = O(n^{-4})$ and Assumption 14(i) and (ii) are satisfied. Because n_g is a constant, w_{ij} is a function of g_i and g_j – once we know the groups that i and j join, we know w_{ij}. In this case, $\boldsymbol{w}$ is a function of $\boldsymbol{\varsigma} = (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g})$ and

w_{ij} depends on $\boldsymbol{\varsigma}$ only through ς_i and ς_j. Therefore, Assumption 14(iii), (v), and (vi) are satisfied. Further, for group averages that include oneself we have $\boldsymbol{w}' = \boldsymbol{w}$ and $\boldsymbol{w}^2 = \boldsymbol{w}$.[26] Because $h_{ij}w_{ij}$ and $h_{kl}w_{kl}$ are independent for disjoint $\{i,j\}$ and $\{k,l\}$, Assumption 14(iv) is satisfied.

Example A.2 (Group averages excluding oneself)**.** Suppose that $\boldsymbol{w}$ represents group averages that exclude oneself and the group capacities are binding. We have $w_{ij} = \sum_{g=1}^{G} \frac{1}{n_g - 1} 1\{g_i = g\}1\{g_j = g\}$ for $i \neq j$ and $w_{ii} = 0$. By construction, $|||\boldsymbol{w}|||_\infty = 1$ and $\|\boldsymbol{w}\|_\infty \leq \frac{1}{n \min_{g \in \mathcal{G}} r_g - 1}$. Hence, $\mathbb{E}[\|\boldsymbol{w}\|_\infty^4] \leq \frac{1}{(n \min_{g \in \mathcal{G}} r_g - 1)^4} = O(n^{-4})$ and Assumption 14(i) and (ii) are satisfied. Similarly as in Example A.1, w_{ij} is a function of g_i and g_j and Assumption 14(iii), (v) and (vi) are satisfied for $\boldsymbol{\varsigma} = (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g})$. For group averages that exclude oneself, we have $\boldsymbol{w}' = \boldsymbol{w}$ and for $r \geq 1$, the (i,j) element of $\boldsymbol{w}^r$ takes the form $(\boldsymbol{w}^r)_{ij} = \sum_{g=1}^{G} c_{ij,g}(r) 1\{g_i = g\}1\{g_j = g\}$, where $c_{ij,g}(r)$ is a constant that only depends on r and n_g. For example, $c_{ij,g}(2) = \frac{n_g - 2}{(n_g - 1)^2}$ for $i \neq j$ and $c_{ii,g}(2) = \frac{1}{n_g - 1}$. This structure suggests that $(\boldsymbol{w}^r)_{ij}$ is a function of g_i and g_j. Hence, $h_{ij}(\boldsymbol{w}^r)_{ij}$ and $h_{kl}(\boldsymbol{w}^{\tilde{r}})_{kl}$ are independent for disjoint $\{i,j\}$ and $\{k,l\}$ and Assumption 14(iv) is satisfied.

Example A.3 (Dyadic networks)**.** Suppose that individuals in a group form additional connections, for example, schoolmates make friends. Let $d_{ij,g}$ denote an indicator for whether individuals i and j are connected in group g and $d_{i,g} = \sum_{j=1}^{n} d_{ij,g} 1\{g_j = g\}$ the number of connections that i has in group g. Suppose that no individual is isolated so $d_{i,g_i} > 0$ for all $i \in \mathcal{N}$. Typically, w_{ij} is specified as $w_{ij} = \sum_{g=1}^{G} \frac{d_{ij,g}}{d_{i,g}} 1\{g_i = g\}1\{g_j = g\}$ – if both i and j join group g, then the weight of j has on i depends on whether j is connected to i, normalized by the number of connections that i has in the group.

Following the literature on dyadic network formation with fixed effects (Graham, 2017; Johnsson and Moon, 2021), we specify $d_{ij,g}$ as

$$d_{ij,g} = 1\{f_g(x_i, x_j, a_i, a_j) \geq \psi_{ij}\}, \ \forall i \neq j, \tag{A.1}$$

and $d_{ii,g} = 0$, where $a_i \in \mathbb{R}$ and $\psi_{ij} \in \mathbb{R}$ represent individual- and pair-specific unobserved heterogeneity. Without loss of generality we normalize $\psi_{ij} \sim U[0,1]$ and

[26] For any $i,j \in \mathcal{N}$, $(\boldsymbol{w}^2)_{ij} = \sum_{k=1}^{n} w_{ik}w_{kj} = \sum_{k=1}^{n} (\sum_{g=1}^{G} \frac{1}{n_g} 1\{g_i = g\}1\{g_k = g\})(\sum_{g=1}^{G} \frac{1}{n_g} 1\{g_k = g\}1\{g_j = g\}) = \sum_{k=1}^{n} \sum_{g=1}^{G} \frac{1}{n_g^2} 1\{g_i = g\}1\{g_j = g\}1\{g_k = g\} = \sum_{g=1}^{G} \frac{1}{n_g} 1\{g_i = g\}1\{g_j = g\} = w_{ij}$, where we have used $n_g = \sum_{k=1}^{n} 1\{g_k = g\}$.

assume $0 \leq f_g \leq 1$. Denote $\boldsymbol{a} = (a_i, i \in \mathcal{N})$ and $\boldsymbol{\psi} = (\psi_{ij}, i, j \in \mathcal{N})$. Assume that (a) both a_i and ψ_{ij} are i.i.d., (b) $\boldsymbol{\psi}$ is independent of $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{a})$, and (c) $(\boldsymbol{a}, \boldsymbol{\psi})$ is independent of $\boldsymbol{\epsilon}$ conditional on $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g})$. The last is consistent with Assumption 1 – conditional on $(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g})$, $\boldsymbol{w}$ is a function of $(\boldsymbol{a}, \boldsymbol{\psi})$ and is thus independent of $\boldsymbol{\epsilon}$. Let $\boldsymbol{\varsigma} = (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{a})$, where $\varsigma_i = (x_i, z_i, g_i, a_i)$ is i.i.d. across i. Note that $\frac{1}{n-1}\mathbb{E}[d_{i,g}|\varsigma_i] = \frac{1}{n-1}\sum_{j\neq i}\mathbb{E}[d_{ij,g}1\{g_j = g\}|\varsigma_i] = \mathbb{E}[f_g(x_i, x_j, a_i, a_j)1\{g_j = g\}|\varsigma_i]$. We assume that $\min_{i\in\mathcal{N}}\min_{g\in\mathcal{G}}\mathbb{E}[f_g(x_i, x_j, a_i, a_j)1\{g_j = g\}|\varsigma_i] \geq c > 0$.

Lemma A.1. *For $\boldsymbol{w}$ in Example A.3, Assumption 14 is satisfied with $\boldsymbol{\varsigma} = (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{a})$.*

Proof. See Appendix A.2.4. □

Example A.4 (Group averages, continued)**.** Examples A.1 and A.2 assume that the group capacities are binding. If the groups have infinite capacities as in one-sided group formation or there is a group that does not reach its full capacity, then the number of members in a group is determined endogenously. This setting can be regarded as a special case of Example A.3, where we set $d_{ij,g} = 1$ for all $i, j \in \mathcal{N}$ (if oneself is included in an average) or $d_{ij,g} = 1$ for all $i \neq j$ and $d_{ii,g} = 0$ (if oneself is excluded in an average). Following the proof of Lemma A.1, we can show that $\boldsymbol{w}$ in this setting satisfies Assumption 14 with $\boldsymbol{\varsigma} = (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g})$.

Example A.5 (Strategic networks)**.** Follow the setup in Example A.3. To account for strategic network formation, we replace equation (A.1) with

$$d_{ij,g} = 1\{f_g(x_i, x_j, a_i; \boldsymbol{x}) \geq \psi_{ij}\}, \ \forall i \neq j, \tag{A.2}$$

and $d_{ii,g} = 0$, where a_i and ψ_{ij} are specified as in Example A.3. This specification is motivated by strategic network formation under incomplete information (Leung, 2015; Ridder and Sheng, 2022), where we assume that $\boldsymbol{x}$ is publicly observed by all the individuals, and a_i and $\psi_i = (\psi_{ij}, j \neq i)$ are privately observed by individual i. The presence of $\boldsymbol{x}$ is to capture the equilibrium effect that results from strategic interactions. We impose the same assumptions on $\boldsymbol{a}$ and $\boldsymbol{\psi}$ as in Example A.3 and set $\boldsymbol{\varsigma} = (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{a})$. Note that $\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] = \sum_{j\neq i} f_g(x_i, x_j, a_i; \boldsymbol{x})1\{g_j = g\}$. We assume that $\min_{i,j\in\mathcal{N}}\min_{g\in\mathcal{G}} f_g(x_i, x_j, a_i; \boldsymbol{x}) \geq c > 0$.

Suppose that the equilibrium effect has a limiting approximation in the sense that for each $g \in \mathcal{G}$,

$$\max_{i,j\in\mathcal{N}} \mathbb{E}[(f_g(x_i, x_j, a_i; \boldsymbol{x}) - f_g^*(x_i, x_j, a_i))^4] = O(n^{-2}), \tag{A.3}$$

for some function $0 \leq f_g^*(x_i, x_j, a_i) \leq 1$ as $n \to \infty$, and $\min_{i,j\in\mathcal{N}} \min_{g\in\mathcal{G}} f_g^*(x_i, x_j, a_i) \geq c^* > 0$.[27]

Lemma A.2. *For $\boldsymbol{w}$ in Example A.5, Assumption 14 is satisfied with $\boldsymbol{\varsigma} = (\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}, \boldsymbol{a})$.*

Proof. See Appendix A.2.4. □

A.2 Proofs

Notation We use $\|\cdot\|$ to denote the Euclidean norm. For an $n \times 1$ vector $x \in \mathbb{R}^n$ and an $n \times n$ matrix $A \in \mathbb{R}^{n^2}$, we have $\|x\| = (\sum_{i=1}^n x_i^2)^{1/2}$ and $\|A\| = (\text{tr}(AA'))^{1/2} = (\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2)^{1/2}$. For a function $b(z) \in \mathbb{R}$ and a nonnegative integer s, $\|b\|_s$ denotes the Sobolev norm of order s. Specifically, let $\frac{\partial^j b(z)}{\partial z^j}$ denote the vector that consists of all distinct jth-order partial derivatives of all elements of $b(z)$, and let $\mathcal{Z}$ denote a set that is contained in the support of z . We have $\|b\|_s = \max_{j\leq s} \sup_{z\in\mathcal{Z}} \|\frac{\partial^j b(z)}{\partial z^j}\|$.

A.2.1 Proofs in Section 3

Proof of Proposition 3.1. Following Azevedo and Leshno (2016), we can show that the cutoffs converge, that is, $p_n \xrightarrow{p} p^*$ as $n \to \infty$, and the limiting cutoffs p^* are non-stochastic.

For any p, the selection bias $\mathbb{E}[\epsilon_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p)]$ is continuous in p because the cdf of the unobservables is continuous under Assumption 2(ii). Therefore, by the continuous mapping theorem, we have

$$\mathbb{E}[\epsilon_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p_n)] \xrightarrow{p} \mathbb{E}[\epsilon_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p^*)]. \tag{A.4}$$

Let $\boldsymbol{x}_{-i} = (x_j, j \neq i)$ and define $\boldsymbol{z}_{-i}$, $\boldsymbol{g}_{-i}$, $\boldsymbol{\xi}_{-i}$, $\boldsymbol{\eta}_{-i}$ analogously. The selection bias evaluated at the limiting cutoffs satisfies

$$\begin{aligned} \mathbb{E}[\epsilon_i|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{g}(\boldsymbol{z}, \boldsymbol{\xi}, \boldsymbol{\eta}; p^*)] &= \mathbb{E}[\epsilon_i|x_i, \boldsymbol{x}_{-i}, z_i, \boldsymbol{z}_{-i}, g_i(z_i, \xi_i, \eta_i; p^*), \boldsymbol{g}_{-i}(\boldsymbol{z}_{-i}, \boldsymbol{\xi}_{-i}, \boldsymbol{\eta}_{-i}; p^*)] \\ &= \mathbb{E}[\epsilon_i|x_i, z_i, g_i(z_i, \xi_i, \eta_i; p^*)], \end{aligned} \tag{A.5}$$

[27] Ridder and Sheng (2022, Section 5.2) demonstrated the existence of a limiting approximation.

where the last equality holds because given the non-stochastic cutoffs, each g_j in $\boldsymbol{g}_{-i}$ depends on individual $j's$ characteristics (z_j, ξ_j, η_j) only and for all $j \neq i$, $(x_j, z_j, \xi_j, \eta_j)$ are independent of ϵ_i under Assumption 2(i). Combining equations (A.4) and (A.5) completes the proof. □

Proof of Proposition 3.2. The event that individual i joins group g can be represented by the selection rules in equation (2.5). Hence, individual i's selection bias from joining group g is given by

$$\mathbb{E}[\epsilon_i | x_i, z_i, g_i = g] = \mathbb{E}[\epsilon_i | z_i, g_i = g] = \lambda_g(\tau_i^o), \tag{A.6}$$

where $\lambda_g(\tau_i^o) := \mathbb{E}[\epsilon_i | \eta_{gi} \geq p_g - z_i'\delta_g^v$, and for all $h \neq g$, $\xi_{ih} - \xi_{ig} < z_i'(\delta_g^u - \delta_h^u)$ or $\eta_{hi} < p_h - z_i'\delta_h^v]$. The first equality follows from the exogeneity of x_i (Assumption 2(iii)) and the fact that g_i is not a function of x_i. The second equality follows from equation (2.5) and the exogeneity of z_i (Assumption 2(iii)). □

A.2.2 Proofs in Section 4

Proof of Proposition 4.1. We focus on p in the proof and the identification of α can be established similarly. Suppose that our goal is to identify the cutoff p_g of group $g \neq 1$. Take another group $h \neq g, 1$. Let $\sigma_{h|g}^e(p_g - \tau_{ig}^v) := \mathbb{P}(g_i = h | p_g - \tau_{ig}^v) = \mathbb{P}(g_i = h | \tau_{ig}^v)$ denote the conditional probability that individual i joins group h given her qualification index for group g. The last equality holds because p_g is a constant, so conditioning on $p_g - \tau_{ig}^v$ is the same as conditioning on τ_{ig}^v. Similarly, let $\sigma_{h|1}^e(p_1 - \tau_{j1}^v) := \mathbb{P}(g_j = h | p_1 - \tau_{j1}^v) = \mathbb{P}(g_j = h | \tau_{j1}^v)$. By equation (2.5) and Assumption 3, $\sigma_{h|g}^e(\cdot)$ and $\sigma_{h|1}^e(\cdot)$ have the same functional form, that is, $\sigma_{h|g}^e(\cdot) = \sigma_{h|1}^e(\cdot) =: \sigma_h^e(\cdot)$. Moreover, because the unobservables have a strictly increasing cdf (Assumption 4(ii)), $\sigma_h^e(\cdot)$ is strictly monotone. Therefore, for any i and j such that $0 < \mathbb{P}(g_i = h | \tau_{ig}^v) = \mathbb{P}(g_j = h | \tau_{j1}^v) < 1$, because $\mathbb{P}(g_i = h | \tau_{ig}^v) = \sigma_h^e(p_g - \tau_{ig}^v)$ and $\mathbb{P}(g_j = h | \tau_{j1}^v) = \sigma_h^e(p_1 - \tau_{j1}^v)$, we obtain $p_g - \tau_{ig}^v = p_1 - \tau_{j1}^v$. This, together with Assumption 4(i), implies that $p_g = \tau_{ig}^v - \tau_{j1}^v$ is identified. □

Proof of Lemma 4.1. Suppose that the support of $X_i - \mathbb{E}[X_i | \tau_i^e]$ is contained in a proper linear subspace of $\mathbb{R}^d$. There is a $d \times 1$ vector of constants $k \neq 0$ such that $k'(X_i - \mathbb{E}[X_i | \tau_i^e]) = k'X_i - \mathbb{E}[k'X_i | \tau_i^e] = 0$ with probability 1. Because $\mathbb{E}[k'X_i | \tau_i^e]$ is a function of τ_i^e, $k'X_i$ is a function of τ_i^e with probability 1.

To show the reverse, let $k \neq 0$ be the $d \times 1$ vector of constants such that $k'X_i$ is a function of τ_i^e with probability 1. This implies that with probability 1 we have $\mathbb{E}[k'X_i|\tau_i^e] = k'X_i$ and thus $k'(X_i - \mathbb{E}[X_i|\tau_i^e]) = 0$. Therefore, the support of $X_i - \mathbb{E}[X_i|\tau_i^e]$ is contained in a proper linear subspace of $\mathbb{R}^d$. $\square$

A.2.3 Proofs in Section 5

Proof of Proposition 5.1. Observe that $\hat{p}_g = (G-2)^{-1} \sum_{h=2,h\neq g}^{G} \hat{p}_{g,h}$, where

$$\hat{p}_{g,h} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} \frac{1}{\zeta_{2n}} K_2\Big(\frac{\hat{\sigma}_{h|g}(\hat{\tau}_{ig}^v) - \hat{\sigma}_{h|1}(\hat{\tau}_{j1}^v)}{\zeta_{2n}}\Big)(\hat{\tau}_{ig}^v - \hat{\tau}_{j1}^v)$$

for $h \neq 1, g$. To derive the asymptotic distribution of $\hat{p}_g$, we derive the influence function of each $\hat{p}_{g,h}$, and take their average to obtain the influence function of $\hat{p}_g$. Define $\hat{\tau}_i^v = (\hat{\tau}_{ig}^v, g \in \mathcal{G})$, $\hat{\sigma}_h = (\hat{\sigma}_{h|g}, g \in \mathcal{G})$, and the function

$$h_{gn}(\hat{\tau}_i^v, \hat{\tau}_j^v; \hat{\sigma}_h) = \frac{1}{\zeta_{2n}} K_2\Big(\frac{\hat{\sigma}_{h|g}(\hat{\tau}_{ig}^v) - \hat{\sigma}_{h|1}(\hat{\tau}_{j1}^v)}{\zeta_{2n}}\Big)(\hat{\tau}_{ig}^v - \hat{\tau}_{j1}^v). \tag{A.7}$$

We can view $\hat{p}_{g,h}$ as a V-statistic with the kernel h_{gn} which is asymmetric in i and j.

We decompose

$$\begin{aligned}
\hat{p}_{g,h} - p_g &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} (h_{gn}(\hat{\tau}_i^v, \hat{\tau}_j^v; \hat{\sigma}_h) - h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)) \\
&\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} (h_{gn}(\tau_i^v, \tau_j^v; \sigma_h) - \mathbb{E}[h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)]) \\
&\quad + \mathbb{E}[h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)] - p_g \\
&\equiv T_{1n} + T_{2n} + T_{3n},
\end{aligned} \tag{A.8}$$

where $\tau_i^v = (\tau_{ig}^v, g \in \mathcal{G})$ and $\sigma_h = (\sigma_{h|g}, g \in \mathcal{G})$. The first term T_{1n} captures the estimation error due to the first-step estimators $\hat{\delta}^v$ and $\hat{\sigma}_h$, the second term T_{2n} captures the estimation error in the second step if properly centered, and the third term T_{3n} captures the bias due to kernel smoothing.

Lemmas O.B.1 and O.B.2 show that both T_{1n} and T_{2n} have an asymptotically linear form, $T_{1n} = n^{-1} \sum_{i=1}^{n} \psi_{g,h,1}(z_i, g_i) + o_p(n^{-1/2})$ and $T_{2n} = n^{-1} \sum_{i=1}^{n} \psi_{g,h,2}(z_i) + o_p(n^{-1/2})$. Lemma O.B.3 shows that $T_{3n} = o(n^{-1/2})$ and is thus negligible. Averaging over $h \neq 1, g$, we derive $\sqrt{n}(\hat{p}_g - p_g) = n^{-1/2} \sum_{i=1}^{n} \psi_g(z_i, g_i) + o_p(1)$, where $\psi_g(z_i, g_i) =$

$(G-2)^{-1}\sum_{h\neq 1,g}(\psi_{g,h,1}(z_i,g_i)+\psi_{g,h,2}(z_i))$ and $\mathbb{E}[\psi_g(z_i,g_i)]=0$. By the Central Limit Theorem, we have $\sqrt{n}(\hat{p}_g-p_g)\xrightarrow{d}N(0,V_g)$, where $V_g=\text{Var}(\psi_g(z_i,g_i))$. □

Proof of Theorem 5.1. Denote $Q(\gamma,\mu^Z)=m_0(\gamma,\mu^Z)'Wm_0(\gamma,\mu^Z)$ and $\hat{Q}_n(\gamma,\mu^Z)=\hat{m}_n(\gamma,\mu^Z)'\hat{W}\hat{m}_n(\gamma,\mu^Z)$. Fix $\delta>0$. Let $\mathcal{B}_\delta(\gamma_0)=\{\gamma\in\Gamma:\|\gamma-\gamma_0\|<\delta\}$ be an open δ-ball centered at γ_0. If $Q(\hat{\gamma},\hat{\mu}^Z)<\inf_{\gamma\in\Gamma\setminus\mathcal{B}_\delta(\gamma_0)}Q(\gamma,\hat{\mu}^Z)$, then $\hat{\gamma}\in\mathcal{B}_\delta(\gamma_0)$. Therefore,

$$\Pr(\|\gamma-\gamma_0\|<\delta)\geq\Pr(Q(\hat{\gamma},\hat{\mu}^Z)<\inf_{\gamma\in\Gamma\setminus\mathcal{B}_\delta(\gamma_0)}Q(\gamma,\hat{\mu}^Z)). \tag{A.9}$$

From the triangle inequality and the optimality of $\hat{\gamma}$, we obtain

$$Q(\hat{\gamma},\mu_0^Z)\leq\hat{Q}_n(\hat{\gamma},\hat{\mu}^Z)+|\hat{Q}_n(\hat{\gamma},\hat{\mu}^Z)-Q(\hat{\gamma},\mu_0^Z)|\leq\sup_{\gamma\in\Gamma}|\hat{Q}_n(\gamma,\hat{\mu}^Z)-Q(\gamma,\mu_0^Z)|+o_p(1).$$

The uniform convergence of the moment in Lemma O.C.1 together with $\hat{W}-W=o_p(1)$ and the boundedness of W and $m_0(\gamma,\mu^Z)$ (Assumptions 11, 14(i), 15(i)) implies that $\sup_{\gamma\in\Gamma}|\hat{Q}_n(\gamma,\hat{\mu}^Z)-Q(\gamma,\mu_0^Z)|=o_p(1)$, and hence $Q(\hat{\gamma},\mu_0^Z)=o_p(1)$.

By Assumption 15(i), $Wm_0(\gamma,\mu^Z)=0$ if and only if $\gamma=\gamma_0$ and therefore $Q(\gamma,\mu_0^Z)$ has a unique minimizer at $\gamma=\gamma_0$. Hence, by the compactness of $\Gamma\setminus\mathcal{B}_\delta(\gamma_0)$ and the continuity of $Q(\gamma,\mu_0^Z)$, we have $\inf_{\gamma\in\Gamma\setminus\mathcal{B}_\delta(\gamma_0)}Q(\gamma,\mu_0^Z)=Q(\bar{\gamma},\mu_0^Z)>Q(\gamma_0,\mu_0^Z)=0$ for some $\bar{\gamma}\in\Gamma\setminus\mathcal{B}_\delta(\gamma_0)$.

Combining the results we can see that the right-hand side of equation (A.9) goes to 1, and the consistency of $\hat{\gamma}$ is proved. □

Proof of Theorem 5.2. For simplicity, we write τ_i^s as τ_i. $\hat{\gamma}$ satisfies the first-order condition

$$\frac{\partial\hat{m}_n(\hat{\gamma},\hat{\mu}^Z)'}{\partial\gamma}\hat{W}\hat{m}_n(\hat{\gamma},\hat{\mu}^Z)=0. \tag{A.10}$$

Expanding equation (A.10) around γ_0 and solving for $\sqrt{n}(\hat{\gamma}-\gamma_0)$ gives

$$\sqrt{n}(\hat{\gamma}-\gamma_0)=-\left(\frac{\partial\hat{m}_n(\hat{\gamma},\hat{\mu}^Z)'}{\partial\gamma}\hat{W}\frac{\partial\hat{m}_n(\bar{\gamma},\hat{\mu}^Z)}{\partial\gamma'}\right)^{-1}\frac{\partial\hat{m}_n(\hat{\gamma},\hat{\mu}^Z)'}{\partial\gamma}\hat{W}\sqrt{n}\hat{m}_n(\gamma_0,\hat{\mu}^Z), \tag{A.11}$$

where $\bar{\gamma}$ is a mean value that lies between $\hat{\gamma}$ and γ_0.

Consider the derivatives in equation (A.11). We have

$$\frac{\partial\hat{m}_n(\gamma,\hat{\mu}^Z)}{\partial\gamma'} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial m(\omega_i,\gamma,\hat{\mu}^Z(\hat{\tau}_i))}{\partial\gamma'}$$

$$
\begin{aligned}
&= -\frac{1}{n}\sum_{i=1}^{n}(Z_i - \hat{\mu}^Z(\hat{\tau}_i))X_i' \\
&= -\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mu_0^Z(\tau_i))X_i' + \frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))X_i' \\
&= -\frac{1}{n}\sum_{i=1}^{n}(Z_i - \mu_0^Z(\tau_i))X_i' + o_p(1),
\end{aligned}
$$

where the last equality follows from Lemma O.C.2. By definition $M_n = \mathbb{E}[\frac{\partial \hat{m}_n(\gamma,\mu_0^Z)}{\partial \gamma'}] = -n^{-1}\sum_{i=1}^{n}\mathbb{E}[(Z_i - \mu_0^Z(\tau_i))X_i']$. Lemma O.C.3 shows that $n^{-1}\sum_{i=1}^{n}((Z_i - \mu_0^Z(\tau_i))X_i' - \mathbb{E}[(Z_i - \mu_0^Z(\tau_i))X_i']) = o_p(1)$. Therefore, $\frac{\partial \hat{m}_n(\gamma,\hat{\mu}^Z)}{\partial \gamma'} - M_n = o_p(1)$.

By Lemmas O.C.11 and O.C.12, the last term in equation (A.11) has the asymptotic distribution $\sqrt{n}\Omega_n^{-1/2}\hat{m}_n(\gamma_0,\hat{\mu}^Z) \xrightarrow{d} N(0, I_{d_Z})$, where Ω_n is defined in Lemma O.C.12. By equation (A.11) and Slutsky's theorem, we obtain $\sqrt{n}\Sigma_n^{-1/2}(\hat{\gamma} - \gamma_0) \xrightarrow{d} N(0, I_{d_\gamma})$, where $\Sigma_n = (M_n'WM_n)^{-1}M_n'W\Omega_n WM_n(M_n'WM_n)^{-1}$. □

A.2.4 Proofs in Appendix A.1

Proof of Lemma A.1. By construction, $|||\boldsymbol{w}|||_\infty = 1$ and $\|\boldsymbol{w}\|_\infty \leq \frac{1}{\min_{i\in\mathcal{N}}\min_{g\in\mathcal{G}} d_{i,g}}$. Observe that $\frac{1}{n-1}(d_{i,g} - \mathbb{E}[d_{i,g}|\varsigma_i]) = o_p(1)$ by the law of large numbers. Because $\frac{1}{n-1}\mathbb{E}[d_{i,g}|\varsigma_i] > c$ for all $i \in \mathcal{N}$ and $g \in \mathcal{G}$, we have $\mathbb{E}[\frac{1}{\min_{i\in\mathcal{N}}\min_{g\in\mathcal{G}}(\frac{1}{n-1}d_{i,g})^4}] \to \mathbb{E}[\frac{1}{\min_{i\in\mathcal{N}}\min_{g\in\mathcal{G}}(\frac{1}{n-1}\mathbb{E}[d_{i,g}|\varsigma_i])^4}] < \infty$ by dominated convergence and thus $\mathbb{E}[\|\boldsymbol{w}\|_\infty^4] = O(n^{-4})$. Conditional on $(\boldsymbol{x},\boldsymbol{z},\boldsymbol{g})$, $\boldsymbol{a}$ is independent of $\boldsymbol{\nu}$, so we have $\mathbb{E}[\nu_i|\boldsymbol{\varsigma}] = \mathbb{E}[\nu_i|\boldsymbol{x},\boldsymbol{z},\boldsymbol{g}] = 0$. Moreover, conditional on $\boldsymbol{\varsigma}$, $\boldsymbol{w}$ is a function of $\boldsymbol{\psi}$ and is thus independent of $\boldsymbol{\nu}$.[28] Hence, Assumptions 14(i)–(iii) are satisfied.

To verify Assumptions 14(iv)–(vi), define $w_{ij,g} = \frac{d_{ij,g}}{d_{i,g}}$, $\bar{w}_{ij,g} = \frac{d_{ij,g}}{\mathbb{E}[d_{i,g}|\varsigma_i]}$, and $e_{ij,g} = w_{ij,g} - \bar{w}_{ij,g}$. By Taylor expansion,

$$
e_{ij,g} = -\frac{d_{ij,g}}{\mathbb{E}[d_{i,g}|\varsigma_i]^2}(d_{i,g} - \mathbb{E}[d_{i,g}|\varsigma_i]) + \frac{d_{ij,g}}{\mathbb{E}[d_{i,g}|\varsigma_i]^3}(d_{i,g} - \mathbb{E}[d_{i,g}|\varsigma_i])^2 - \cdots \tag{A.12}
$$

It suffices to consider the leading term in $e_{ij,g}$. Recall that $d_{i,g} - \mathbb{E}[d_{i,g}|\varsigma_i] = \sum_{j\neq i} r_{ij,g}$, where $r_{ij,g} = d_{ij,g}1\{g_j = g\} - \mathbb{E}[d_{ij,g}1\{g_j = g\}|\varsigma_i]$. Note that $|r_{ij,g}| \leq 1$ and $\mathbb{E}[r_{ij,g}|\varsigma_i] = 0$. For any $j \neq k$, conditional on ς_i, $r_{ij,g}$ is a function of $(x_j, a_j, \psi_{ij}, g_j)$ and $r_{ik,g}$ is a

[28] Because $(\boldsymbol{a},\boldsymbol{\psi})$ is independent of $\boldsymbol{\nu}$ conditional on $(\boldsymbol{x},\boldsymbol{z},\boldsymbol{g})$, we can show that $\boldsymbol{\psi}$ is independent of $\boldsymbol{\nu}$ conditional on $\boldsymbol{\varsigma}$.

function of $(x_k, a_k, \psi_{ik}, g_k)$, so $r_{ij,g}$ and $r_{ik,g}$ are independent. Therefore,

$$\begin{aligned}\mathbb{E}[(d_{i,g} - \mathbb{E}[d_{i,g}|\varsigma_i])^4|\varsigma_i] &= \sum_{j,k,l,m\neq i} \mathbb{E}[r_{ij,g}r_{ik,g}r_{il,g}r_{im,g}|\varsigma_i] \\ &= \sum_{j\neq i} \mathbb{E}[r_{ij,g}^4|\varsigma_i] + \sum_{j\neq k, j,k\neq i} \mathbb{E}[r_{ij,g}^2 r_{ik,g}^2|\varsigma_i] \leq O(n^2). \quad \text{(A.13)}\end{aligned}$$

Combining equations (A.12) and (A.13) yields $\max_{i,j\in\mathcal{N}} \max_{g\in\mathcal{G}} \mathbb{E}[e_{ij,g}^4] = O(n^{-6})$. Further, summing over the groups we define $\bar{w}_{ij} = \sum_{g=1}^{G} \bar{w}_{ij,g} 1\{g_i = g\}1\{g_j = g\}$ and $e_{ij} = w_{ij} - \bar{w}_{ij}$. From the previous results we obtain $\max_{i,j\in\mathcal{N}} |\bar{w}_{ij}| \leq \frac{1}{\min_{i\in\mathcal{N}} \min_{g\in\mathcal{G}} \mathbb{E}[d_{i,g}|\varsigma_i]} \leq O(n^{-1})$ and $\max_{i,j\in\mathcal{N}} \mathbb{E}[e_{ij}^4] \leq G \max_{i,j\in\mathcal{N}} \max_{g\in\mathcal{G}} \mathbb{E}[e_{ij,g}^4] = O(n^{-6})$.

Assumption 14(v). For any disjoint $\{i,j\}$ and $\{k,l\}$, we have $\text{Cov}(w_{ij}, w_{kl}|\boldsymbol{\varsigma}) = \text{Cov}(w_{ij}, e_{kl}|\boldsymbol{\varsigma}) = \text{Cov}(e_{ij}, e_{kl}|\boldsymbol{\varsigma})$. The first equality holds because conditional on $\boldsymbol{\varsigma}$, w_{ij} is a function of $(\psi_{i\tilde{j}}, \tilde{j} \in \mathcal{N}_{g_i}\backslash\{i\})$, where $\mathcal{N}_{g_i} = \{j : g_j = g_i\}$, and $\bar{w}_{kl}$ is a function of ψ_{kl}, so $\text{Cov}(w_{ij}, \bar{w}_{kl}|\boldsymbol{\varsigma}) = 0$. The second equality follows similarly from $\text{Cov}(\bar{w}_{ij}, e_{kl}|\boldsymbol{\varsigma}) = 0$. By Cauchy-Schwarz inequality and Jensen's inequality, $\mathbb{E}[\text{Cov}(e_{ij}, e_{kl}|\boldsymbol{\varsigma})^2] \leq C\mathbb{E}[e_{ij}^4]$. Therefore, $\max_{i,j,k,l\in\mathcal{N}:\{i,j\}\cap\{k,l\}=\emptyset} \mathbb{E}[\text{Cov}(w_{ij}, w_{kl}|\boldsymbol{\varsigma})^2] = O(n^{-6}) = o(n^{-4}/K)$ because $K/n^2 \to 0$.

Moreover, for any $i,j \in \mathcal{N}$, $\mathbb{E}[w_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[w_{ij}|\varsigma_i, \varsigma_j] = \mathbb{E}[e_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[e_{ij}|\varsigma_i, \varsigma_j]$ because $\bar{w}_{ij}$ depends on $\boldsymbol{\varsigma}$ only through ς_i and ς_j and thus $\mathbb{E}[\bar{w}_{ij}|\boldsymbol{\varsigma}] = \mathbb{E}[\bar{w}_{ij}|\varsigma_i, \varsigma_j]$. We can bound $\mathbb{E}[(\mathbb{E}[e_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[e_{ij}|\varsigma_i, \varsigma_j])^4] \leq C\mathbb{E}[e_{ij}^4]$. Hence, $\max_{i,j\in\mathcal{N}} \mathbb{E}[(\mathbb{E}[w_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[w_{ij}|\varsigma_i, \varsigma_j])^4] = O(n^{-6}) = o(n^{-4}/K^2)$ because $K/n \to 0$. Assumption 14(v) is satisfied.

Assumption 14(vi). For $\{i,j\}$ and $\{k,l\}$ that overlap in one element, $\bar{w}_{ij}$ and $\bar{w}_{kl}$ are independent conditional on $\boldsymbol{\varsigma}$. Therefore, $\text{Cov}(w_{ij}, w_{kl}|\boldsymbol{\varsigma}) = \text{Cov}(e_{ij}, \bar{w}_{kl}|\boldsymbol{\varsigma}) + \text{Cov}(\bar{w}_{ij}, e_{kl}|\boldsymbol{\varsigma}) + \text{Cov}(e_{ij}, e_{kl}|\boldsymbol{\varsigma})$. By Cauchy-Schwarz inequality and Jensen's inequality, we can bound $\mathbb{E}[\text{Cov}(e_{ij}, \bar{w}_{kl}|\boldsymbol{\varsigma})^2] \leq C(\mathbb{E}[e_{ij}^4]\mathbb{E}[\bar{w}_{kl}^4])^{1/2} = O(n^{-5}) = o(n^{-4})$ uniformly. Assumption 14(vi) thus holds.

Assumption 14(iv). For any $r \geq 1$, $(\boldsymbol{w}^r)_{ij} = \sum_{(t_0,\ldots,t_r):(t_0,t_r)=(i,j)} \dot{w}_{t_0,\ldots,t_r}$, where $\dot{w}_{t_0,\ldots,t_r} = \prod_{s=0}^{r-1} w_{t_s t_{s+1}}$ and the sum is over tuples $(t_0, \ldots, t_r)$ with $t_0 = i$ and $t_r = j$. For any disjoint $\{i,j\}$ and $\{k,l\}$, we can write

$$\begin{aligned}&\text{Cov}(h_{ij}(\boldsymbol{w}^r)_{ij}, h_{kl}(\boldsymbol{w}^{\tilde{r}})_{kl}) \\ = &\sum_{\substack{(t_0,\ldots,t_r,\tilde{t}_0\ldots,\tilde{t}_{\tilde{r}}):(t_0,t_r,\tilde{t}_0,\tilde{t}_{\tilde{r}})=(i,j,k,l),\\ \{t_0,\ldots,t_r\}\cap\{\tilde{t}_0\ldots,\tilde{t}_{\tilde{r}}\}\neq\emptyset}} \text{Cov}(h_{ij}\dot{w}_{t_0,\ldots,t_r}, h_{kl}\dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}})\end{aligned}$$

$$+\sum_{\substack{(t_0,\ldots,t_r,\tilde{t}_0\ldots,\tilde{t}_{\tilde{r}}):(t_0,t_r,\tilde{t}_0,\tilde{t}_{\tilde{r}})=(i,j,k,l),\\ \{t_0,\ldots,t_r\}\cap\{\tilde{t}_0\ldots,\tilde{t}_{\tilde{r}}\}=\emptyset}} \mathrm{Cov}(h_{ij}\dot{w}_{t_0,\ldots,t_r}, h_{kl}\dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}). \qquad \text{(A.14)}$$

The first sum in (A.14) consists of $O(n^{(r+\tilde{r})-3})$ terms, and each term can be bounded uniformly by $C(\mathbb{E}[\|\boldsymbol{w}\|_\infty^{r+\tilde{r}}]+\mathbb{E}[\|\boldsymbol{w}\|_\infty^r]\mathbb{E}[\|\boldsymbol{w}\|_\infty^{\tilde{r}}]) = O(n^{-(r+\tilde{r})})$. Hence, the first sum in (A.14) is $O(n^{(r+\tilde{r})-3})\cdot O(n^{-(r+\tilde{r})}) = o(n^{-2})$, uniformly in i, j, k, l, r, and $\tilde{r}$. The second sum in (A.14) consists of $O(n^{(r+\tilde{r})-2})$ terms. To derive a uniform bound on each term, for disjoint $\{t_0,\ldots,t_r\}$ and $\{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}\}$ with $(t_0,t_r,\tilde{t}_0,\tilde{t}_{\tilde{r}}) = (i,j,k,l)$, we write

$$\begin{aligned}
&\mathrm{Cov}(h_{ij}\dot{w}_{t_0,\ldots,t_r}, h_{kl}\dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_r})\\
=\ &\mathbb{E}[h_{ij}h_{kl}\mathrm{Cov}(\dot{w}_{t_0,\ldots,t_r}, \dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_r}|\boldsymbol{\varsigma})] + \mathrm{Cov}(h_{ij}\mathbb{E}[\dot{w}_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}], h_{kl}\mathbb{E}[\dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_r}|\boldsymbol{\varsigma}]).
\end{aligned}$$

Define $\dot{\bar{w}}_{t_0,\ldots,t_r} = \prod_{s=0}^{r-1}\bar{w}_{t_s t_{s+1}}$ and $e_{t_0,\ldots,t_r} = \dot{w}_{t_0,\ldots,t_r} - \dot{\bar{w}}_{t_0,\ldots,t_r}$. For any $t_0,\ldots,t_r$, by continuous mapping and Slutsky's theorem we can bound $e_{t_0,\ldots,t_r}$ uniformly by $o_p(n^{-r})$. Observe that $\mathrm{Cov}(\dot{w}_{t_0,\ldots,t_r}, \dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\boldsymbol{\varsigma}) = \mathrm{Cov}(\dot{w}_{t_0,\ldots,t_r}, e_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\boldsymbol{\varsigma}) = \mathrm{Cov}(e_{t_0,\ldots,t_r}, e_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\boldsymbol{\varsigma})$. The first equality holds because conditional on $\boldsymbol{\varsigma}$, $\dot{w}_{t_0,\ldots,t_r}$ is a function of $(\psi_{t_s j}, j\in\mathcal{N}_{g_{t_s}}\backslash\{t_s\}, s=0,\ldots,r-1)$ and $\dot{\bar{w}}_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}$ is a function of $(\psi_{\tilde{t}_s\tilde{t}_{s+1}}, s=0,\ldots,\tilde{r}-1)$, and thus $\mathrm{Cov}(\dot{w}_{t_0,\ldots,t_r}, \dot{\bar{w}}_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\boldsymbol{\varsigma}) = 0$. The second equality follows similarly from $\mathrm{Cov}(\dot{\bar{w}}_{t_0,\ldots,t_r}, e_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\boldsymbol{\varsigma}) = 0$. Therefore, from the boundedness of h_{ij} and dominated convergence, we obtain that $\mathbb{E}[h_{ij}h_{kl}\mathrm{Cov}(\dot{w}_{t_0,\ldots,t_r}, \dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_r}|\boldsymbol{\varsigma})]$ has a uniform bound that is of the rate of $o(n^{-(r+\tilde{r})})$.

Moreover, because $\dot{\bar{w}}_{t_0,\ldots,t_r}$ depends on $\boldsymbol{\varsigma}$ only through $\varsigma_{t_0},\ldots,\varsigma_{t_r}$, $\mathbb{E}[\dot{\bar{w}}_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}] = \mathbb{E}[\dot{\bar{w}}_{t_0,\ldots,t_r}|\varsigma_{t_0},\ldots,\varsigma_{t_r}]$. For disjoint $\{t_0,\ldots,t_r\}$ and $\{\tilde{t}_0\ldots,\tilde{t}_{\tilde{r}}\}$ with $(t_0,t_r,\tilde{t}_0,\tilde{t}_{\tilde{r}}) = (i,j,k,l)$, $h_{ij}\mathbb{E}[\dot{\bar{w}}_{t_0,\ldots,t_r}|\varsigma_{t_0},\ldots,\varsigma_{t_r}]$ and $h_{kl}\mathbb{E}[\dot{\bar{w}}_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\varsigma_{\tilde{t}_0},\ldots,\varsigma_{\tilde{t}_{\tilde{r}}}]$ are independent. Hence, $\mathrm{Cov}(h_{ij}\mathbb{E}[\dot{w}_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}], h_{kl}\mathbb{E}[\dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_r}|\boldsymbol{\varsigma}]) = \mathrm{Cov}(h_{ij}\mathbb{E}[\dot{\bar{w}}_{t_0,\ldots,t_r}|\varsigma_{t_0},\ldots,\varsigma_{t_r}]),\ h_{kl}\mathbb{E}[e_{\tilde{t}_0,\ldots,\tilde{t}_r}|\boldsymbol{\varsigma}])+ \mathrm{Cov}(h_{ij}\mathbb{E}[e_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}], h_{kl}\mathbb{E}[\dot{\bar{w}}_{\tilde{t}_0,\ldots,\tilde{t}_r}|\varsigma_{\tilde{t}_0},\ldots,\varsigma_{\tilde{t}_{\tilde{r}}}])+\mathrm{Cov}(h_{ij}\mathbb{E}[e_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}], h_{kl}\mathbb{E}[e_{\tilde{t}_0,\ldots,\tilde{t}_r}|\boldsymbol{\varsigma}])$. Similarly as before, we can derive that $\mathrm{Cov}(h_{ij}\mathbb{E}[\dot{w}_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}], h_{kl}\mathbb{E}[\dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_r}|\boldsymbol{\varsigma}])$ has a uniform bound that is $o(n^{-(r+\tilde{r})})$. Combining the results yields Assumption 14(iv). □

Proof of Lemma A.2. For any $j\neq k$, conditional on $\boldsymbol{\varsigma}$, $d_{ij,g}$ is a function of ψ_{ij} and $d_{ik,g}$ is a function of ψ_{ik}, so $d_{ij,g}$ and $d_{ik,g}$ are independent. Hence, we obtain $\mathbb{E}[(d_{i,g}-\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}])^2] = O(n)$ and thus $\frac{1}{n-1}(d_{i,g}-\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}]) = o_p(1)$. Note that $c < \frac{1}{n-1}\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] \leq 1$ for all $i\in\mathcal{N}$ and $g\in\mathcal{G}$. Following the argument in Lemma A.1, we can show that Assumptions 14(i)–(iii) are satisfied.

To verify Assumptions 14(iv)–(vi), modify $\bar{w}_{ij,g}$ and $e_{ij,g}$ in Lemma A.1 with $\bar{w}_{ij,g} = \frac{d_{ij,g}}{\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}]}$ and $e_{ij,g} = w_{ij,g} - \bar{w}_{ij,g}$. Write $d_{i,g} - \mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] = \sum_{j \neq i} r_{ij,g}$, where $r_{ij,g} = d_{ij,g} 1\{g_j = g\} - \mathbb{E}[d_{ij,g} 1\{g_j = g\}|\boldsymbol{\varsigma}]$. Observe that $|r_{ij,g}| \leq 1$ and $\mathbb{E}[r_{ij,g}|\boldsymbol{\varsigma}] = 0$. For any $j \neq k$, $r_{ij,g}$ and $r_{ik,g}$ are independent conditional on $\boldsymbol{\varsigma}$. Therefore, we have

$$\begin{aligned} \mathbb{E}[(d_{i,g} - \mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}])^4|\boldsymbol{\varsigma}] &= \sum_{j,k,l,m \neq i} \mathbb{E}[r_{ij,g} r_{ik,g} r_{il,g} r_{im,g}|\boldsymbol{\varsigma}] \\ &= \sum_{j \neq i} \mathbb{E}[r_{ij,g}^4|\boldsymbol{\varsigma}] + \sum_{j \neq k, j,k \neq i} \mathbb{E}[r_{ij,g}^2 r_{ik,g}^2|\boldsymbol{\varsigma}] \leq O(n^2). \end{aligned} \quad \text{(A.15)}$$

Combining a modification of equation (A.12) where we replace $\mathbb{E}[d_{i,g}|\varsigma_i]$ with $\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}]$ and equation (A.15) yields $\max_{i,j \in \mathcal{N}} \max_{g \in \mathcal{G}} \mathbb{E}[e_{ij,g}^4] = O(n^{-6})$. Further, summing over the groups we define $\bar{w}_{ij} = \sum_{g=1}^{G} \bar{w}_{ij,g} 1\{g_i = g\} 1\{g_j = g\}$ and $e_{ij} = w_{ij} - \bar{w}_{ij}$. From the previous results we obtain $\max_{i,j \in \mathcal{N}} |\bar{w}_{ij}| \leq \frac{1}{\min_{i \in \mathcal{N}} \min_{g \in \mathcal{G}} \mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}]} \leq O(n^{-1})$ and $\max_{i,j \in \mathcal{N}} \mathbb{E}[e_{ij}^4] \leq G \max_{i,j \in \mathcal{N}} \max_{g \in \mathcal{G}} \mathbb{E}[e_{ij,g}^4] = O(n^{-6})$. Conditional on $\boldsymbol{\varsigma}$, $\bar{w}_{ij}$ is a function of ψ_{ij}.

With the modified $\bar{w}_{ij}$, the rest of the proof in Lemma A.1 still holds, except that $\bar{w}_{ij}$ also depends on ς_k, $k \neq i, j$, due to the presence of strategic interactions. In fact, $\mathbb{E}[\bar{w}_{ij}|\boldsymbol{\varsigma}] = \sum_{g=1}^{G} \mathbb{E}[\bar{w}_{ij,g}|\boldsymbol{\varsigma}] 1\{g_i = g\} 1\{g_j = g\}$ and $\mathbb{E}[\bar{w}_{ij,g}|\boldsymbol{\varsigma}] = \frac{\mathbb{E}[d_{ij,g}|\boldsymbol{\varsigma}]}{\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}]} = \frac{f_{ij,g}}{\sum_{k \neq i} f_{ik,g} 1\{g_k = g\}}$, where $f_{ij,g}$ is shorthand for $f_g(x_i, x_j, a_i; \boldsymbol{x})$. To overcome this problem, we exploit the limiting approximation in equation (A.3).

Define $d_{ij,g}^* = 1\{f_{ij,g}^* \geq \psi_{ij}\}$ for $i \neq j$, $d_{ii,g}^* = 0$, and $d_{i,g}^* = \sum_{j=1}^{n} d_{ij,g}^* 1\{g_j = g\}$, where $f_{ij,g}^*$ is shorthand for $f_g^*(x_i, x_j, a_i)$. Define $\bar{w}_{ij,g}^* = \frac{d_{ij,g}^*}{\mathbb{E}[d_{i,g}^*|\varsigma_i]}$, $e_{ij,g}^* = \bar{w}_{ij,g} - \bar{w}_{ij,g}^*$, $\bar{w}_{ij}^* = \sum_{g=1}^{G} \bar{w}_{ij,g}^* 1\{g_i = g\} 1\{g_j = g\}$, and $e_{ij}^* = \bar{w}_{ij} - \bar{w}_{ij}^*$. Note that $w_{ij} = e_{ij} + e_{ij}^* + \bar{w}_{ij}^*$. Because $\mathbb{E}[\bar{w}_{ij}^*|\boldsymbol{\varsigma}] = \mathbb{E}[\bar{w}_{ij}^*|\varsigma_i, \varsigma_j]$, we can write $\mathbb{E}[w_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[w_{ij}|\varsigma_i, \varsigma_j] = \mathbb{E}[e_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[e_{ij}|\varsigma_i, \varsigma_j] + \mathbb{E}[e_{ij}^*|\boldsymbol{\varsigma}] - \mathbb{E}[e_{ij}^*|\varsigma_i, \varsigma_j]$. The proof in Lemma A.1 provides a bound for $\mathbb{E}[e_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[e_{ij}|\varsigma_i, \varsigma_j]$. Here we derive a similar bound for $\mathbb{E}[e_{ij}^*|\boldsymbol{\varsigma}] - \mathbb{E}[e_{ij}^*|\varsigma_i, \varsigma_j]$. By Taylor expansion,

$$\begin{aligned} e_{ij,g}^* &= \frac{1}{\mathbb{E}[d_{i,g}^*|\varsigma_i]}(d_{ij,g} - d_{ij,g}^*) - \frac{d_{ij,g}^*}{\mathbb{E}[d_{i,g}^*|\varsigma_i]^2}(\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] - \mathbb{E}[d_{i,g}^*|\varsigma_i]) \\ &\quad - \frac{1}{\mathbb{E}[d_{i,g}^*|\varsigma_i]^2}(d_{ij,g} - d_{ij,g}^*)(\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] - \mathbb{E}[d_{i,g}^*|\varsigma_i]) \\ &\quad + \frac{d_{ij,g}^*}{\mathbb{E}[d_{i,g}^*|\varsigma_i]^3}(\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] - \mathbb{E}[d_{i,g}^*|\varsigma_i])^2 - \cdots \end{aligned} \quad \text{(A.16)}$$

It suffices to consider the first two leading terms. By equation (A.3), we derive $\mathbb{E}[\mathbb{E}[d_{ij,g} - d^*_{ij,g}|\boldsymbol{\varsigma}]^4] = \mathbb{E}[(f_{ij,g} - f^*_{ij,g})^4] = O(n^{-2})$ uniformly. Moreover, observe that $\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] = \sum_{j\neq i} f_{ij,g}1\{g_j = g\}$, $\mathbb{E}[d^*_{i,g}|\boldsymbol{\varsigma}] = \sum_{j\neq i} f^*_{ij,g}1\{g_j = g\}$, and $\mathbb{E}[d^*_{i,g}|\varsigma_i] = \sum_{j\neq i} \mathbb{E}[f^*_{ij,g}1\{g_j = g\}|\varsigma_i]$. By Cauchy-Schwarz inequality and equation (A.3), we can bound $\mathbb{E}[(\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] - \mathbb{E}[d^*_{i,g}|\boldsymbol{\varsigma}])^4] \leq (n-1)^4 \max_{i,j\in\mathcal{N}} \max_{g\in\mathcal{G}} \mathbb{E}[(f_{ij,g} - f^*_{ij,g})^4] = O(n^2)$. Next, note that for any $j \neq k$, $f^*_{ij,g}1\{g_j = g\}$ and $f^*_{ik,g}1\{g_k = g\}$ are independent conditional on ς_i. Similarly as in equation (A.13), we can thus derive $\mathbb{E}[(\mathbb{E}[d^*_{i,g}|\boldsymbol{\varsigma}] - \mathbb{E}[d^*_{i,g}|\varsigma_i])^4] = O(n^2)$ uniformly. Therefore, by $(a+b)^4 \leq 8(a^4 + b^4)$, we obtain $\mathbb{E}[(\mathbb{E}[d_{i,g}|\boldsymbol{\varsigma}] - \mathbb{E}[d^*_{i,g}|\varsigma_i])^4] = O(n^2)$ uniformly. Combining these results with equation (A.16) yields $\max_{i,j\in\mathcal{N}} \max_{g\in\mathcal{G}} \mathbb{E}[\mathbb{E}[e^*_{ij,g}|\boldsymbol{\varsigma}]^4] = O(n^{-6})$ and hence $\max_{i,j\in\mathcal{N}} \mathbb{E}[\mathbb{E}[e^*_{ij}|\boldsymbol{\varsigma}]^4] \leq G \max_{i,j\in\mathcal{N}} \max_{g\in\mathcal{G}} \mathbb{E}[\mathbb{E}[e^*_{ij,g}|\boldsymbol{\varsigma}]^4] = O(n^{-6})$. By iterated expectations and Jensen's inequality $\mathbb{E}[\mathbb{E}[e^*_{ij}|\varsigma_i,\varsigma_j]^4] \leq \mathbb{E}[\mathbb{E}[e^*_{ij}|\boldsymbol{\varsigma}]^4]$, so we obtain the uniform bound $\max_{i,j\in\mathcal{N}} \mathbb{E}[(\mathbb{E}[e^*_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[e^*_{ij}|\varsigma_i,\varsigma_j])^4] \leq C \max_{i,j\in\mathcal{N}} \mathbb{E}[\mathbb{E}[e^*_{ij}|\boldsymbol{\varsigma}]^4] = O(n^{-6})$. This proves that Assumption 14(vi) is satisfied.

Assumption 14(v) can be justified by the same proof as in Lemma A.1. As for Assumption 14(iv), the proof in Lemma A.1 remains valid except the last paragraph. Define $\dot{\bar{w}}^*_{t_0,\ldots,t_r} = \prod_{s=0}^{r-1} \bar{w}^*_{t_s t_{s+1}}$. Because $\dot{\bar{w}}^*_{t_0,\ldots,t_r}$ depends on $\boldsymbol{\varsigma}$ only through $\varsigma_{t_0},\ldots,\varsigma_{t_r}$, $\mathbb{E}[\dot{\bar{w}}^*_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}] = \mathbb{E}[\dot{\bar{w}}^*_{t_0,\ldots,t_r}|\varsigma_{t_0},\ldots,\varsigma_{t_r}]$. For disjoint $\{t_0,\ldots,t_r\}$ and $\{\tilde{t}_0\ldots,\tilde{t}_{\tilde{r}}\}$ with $(t_0,t_r,\tilde{t}_0,\tilde{t}_{\tilde{r}}) = (i,j,k,l)$, $h_{ij}\mathbb{E}[\dot{\bar{w}}^*_{t_0,\ldots,t_r}|\varsigma_{t_0},\ldots,\varsigma_{t_r}]$ and $h_{kl}\mathbb{E}[\dot{\bar{w}}^*_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\varsigma_{\tilde{t}_0},\ldots,\varsigma_{\tilde{t}_{\tilde{r}}}]$ are independent. Using a similar argument as in Lemma A.1 with $\dot{\bar{w}}^*_{t_0,\ldots,t_r}$ in place of $\dot{\bar{w}}_{t_0,\ldots,t_r}$, we derive that $\mathrm{Cov}(h_{ij}\mathbb{E}[\dot{w}_{t_0,\ldots,t_r}|\boldsymbol{\varsigma}], h_{kl}\mathbb{E}[\dot{w}_{\tilde{t}_0,\ldots,\tilde{t}_{\tilde{r}}}|\boldsymbol{\varsigma}])$ has a uniform bound that is $o(n^{-(r+\tilde{r})})$. Assumption 14(iv) is thus satisfied. □

References

Altonji, J. G. and R. L. Matzkin (2005): "Cross section and panel data estimators for nonseparable models with endogenous regressors," *Econometrica*, 73, 1053–1102.

Andrews, D. W. (1994): "Chapter 37 Empirical process methods in econometrics," in *Handbook of Econometrics*, Elsevier, 2247–2294.

Argys, L. M. and D. I. Rees (2008): "Searching for peer group effects: A test of the contagion hypothesis," *The Review of Economics and Statistics*, 90, 442–458.

Artemov, G., Y.-K. Che, and Y. He (2020): "Strategic 'Mistakes': Implications for Market Design Research," Manuscript.

Åslund, O., P.-A. Edin, P. Fredriksson, and H. Grönqvist (2011): "Peers, neighborhoods, and immigrant student achievement: Evidence from a placement policy," *American Economic Journal: Applied Economics*, 3, 67–95.

Auerbach, E. (2022): "Identification and Estimation of a Partially Linear Regression Model Using Network Data," *Econometrica*, 90, 347–365.

Azevedo, E. M. and J. D. Leshno (2016): "A supply and demand framework for two-sided matching markets," *Journal of Political Economy*, 124, 1235–1268.

Bayer, P., F. Ferreira, and R. McMillan (2007): "A unified framework for measuring preferences for schools and neighborhoods," *Journal of political economy*, 115, 588–638.

Bayer, P., S. L. Ross, and G. Topa (2008): "Place of work and place of residence: Informal hiring networks and labor market outcomes," *Journal of political Economy*, 116, 1150–1196.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 841–890.

Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2011): "Identification of social interactions," in *Handbook of social economics*, Elsevier, vol. 1, 853–964.

Bobonis, G. J. and F. Finan (2009): "Neighborhood peer effects in secondary school enrollment decisions," *The Review of Economics and Statistics*, 91, 695–716.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009): "Identification of peer effects through social networks," *Journal of econometrics*, 150, 41–55.

Brock, W. A. and S. N. Durlauf (2001): "Interactions-based models," in *Handbook of econometrics*, Elsevier, vol. 5, 3297–3380.

——— (2005): "Multinomial Choice with Social Interactions," *The Economy As an Evolving Complex System, III: Current Perspectives and Future Directions*, 3, 175.

Carrell, S. E., F. V. Malmstrom, and J. E. West (2008): "Peer effects in academic cheating," *Journal of human resources*, 43, 173–207.

Damm, A. P. and C. Dustmann (2014): "Does growing up in a high crime neighborhood affect youth criminal behavior?" *American Economic Review*, 104, 1806–32.

Das, M., W. K. Newey, and F. Vella (2003): "Nonparametric estimation of sample selection models," *The Review of Economic Studies*, 70, 33–58.

Davezies, L., X. D'Haultfoeuille, and D. Fougère (2009): "Identification of peer effects using group size variation," *The Econometrics Journal*, 12, 397–413.

Epple, D. and R. E. Romano (1998): "Competition between private and public schools, vouchers, and peer-group effects," *American Economic Review*, 33–62.

——— (2011): "Peer Effects in Education," in *Handbook of Social Economics*, Elsevier, 1053–1163.

Fack, G., J. Grenet, and Y. He (2019): "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economic Review*, 109, 1486–1529.

Fox, J. T., C. Yang, and D. H. Hsu (2018): "Unobserved Heterogeneity in Matching Games," *Journal of Political Economy.*

Friesen, J. and B. Krauth (2007): "Sorting and inequality in Canadian schools," *Journal of public Economics*, 91, 2185–2212.

Gale, D. and L. Shapley (1962): "College Admissions and the Stability of Marriage," *The American Mathematical Monthly*, 69, 9–15.

Gandhi, A. and J.-F. Houde (2019): "Measuring Substitution Patterns in Differentiated-Products Industries," Tech. rep., National Bureau of Economic Research.

Gaviria, A. and S. Raphael (2001): "School-based peer effects and juvenile behavior," *Review of Economics and Statistics*, 83, 257–268.

Goldsmith-Pinkham, P. and G. W. Imbens (2013): "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 31, 253–264.

Graham, B. S. (2008): "Identifying social interactions through conditional variance restrictions," *Econometrica*, 76, 643–660.

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

Hahn, J. and G. Ridder (2013): "Asymptotic Variance of Semiparametric Estimators With Generated Regressors," *Econometrica*, 81, 315–340.

He, Y., S. Sinha, and X. Sun (2022): "Identification and Estimation in Many-to-One Two-Sided Matching without Transfers," *Working Paper, https://arxiv.org/abs/2104.02009*.

Heckman, J. J. (1979): "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, 153–161.

Hoeffding, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution," *The Annals of Mathematical Statistics*, 293–325.

Hsieh, C.-S. and L. F. Lee (2016): "A social interactions model with endogenous friendship formation and selectivity," *Journal of Applied Econometrics*, 31, 301–319.

Ioannides, Y. M. and J. E. Zabel (2008): "Interactions, neighborhood selection and housing demand," *Journal of urban economics*, 63, 229–252.

Johnsson, I. and H. R. Moon (2021): "Estimation of peer effects in endogenous social networks: Control function approach," *Review of Economics and Statistics*, 1–51.

Katz, L. F., J. R. Kling, and J. B. Liebman (2001): "Moving to opportunity in Boston: Early results of a randomized mobility experiment," *The Quarterly Journal of Economics*, 116, 607–654.

Kling, J. R., J. Ludwig, and L. F. Katz (2005): "Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment," *The Quarterly Journal of Economics*, 120, 87–130.

Kremer, M., E. Miguel, and R. Thornton (2009): "Incentives to learn," *The Review of Economics and Statistics*, 91, 437–456.

Lavy, V. and A. Schlosser (2011): "Mechanisms and impacts of gender peer effects at school," *American Economic Journal: Applied Economics*, 3, 1–33.

Lee, A. J. (1990): *U-Statistics: Theory and Practice*, 270 Madison Ave, New York, NY 10016: Marcel Dekker, Inc.

Lee, L.-F. (2002): "Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Augotregressive Models," *Econometric Theory*, 18, 252–277.

——— (2007): "Identification and estimation of econometric models with group interactions, contextual factors and fixed effects," *Journal of Econometrics*, 140, 333–374.

Leung, M. P. (2015): "Two-step estimation of network-formation models with incomplete information," *Journal of Econometrics*, 188, 182–195.

Lewbel, A., S. Norris, K. Pendakur, X. Qu, et al. (2021): "Consumption peer effects and utility needs in India," *Working Paper*.

Li, Q. and J. S. Racine (2007): *Nonparametric econometrics: theory and practice*, Princeton University Press.

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

McFadden, D. (1989): "A method of simulated moments for estimation of discrete response models without numerical integration," *Econometrica: Journal of the Econometric Society*, 995–1026.

Menzel, K. (2015): "Large Matching Markets as Two-Sided Demand Systems," *Econometrica*, 83, 897–941.

——— (2021): "Central Limit Theory for Models of Strategic Network Formation," *arXiv preprint arXiv:2111.01678*.

——— (2022): "Strategic network formation with many agents," *Working Paper, New York University.*

Newey, W. K. (1994a): "The asymptotic variance of semiparametric estimators," *Econometrica: Journal of the Econometric Society*, 1349–1382.

——— (1994b): "Kernel Estimation of Partial Means and a General Variance Estimator," *Econometric Theory*, 10, 233–253.

——— (1997): "Convergence rates and asymptotic normality for series estimators," 79, 147–168.

Newey, W. K. and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, Elsevier, vol. 4, 2111–2245.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): "Semiparametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Ridder, G. and S. Sheng (2022): "Two-Step Estimation of A Strategic Network Formation Model with Clustering," *arXiv preprint arXiv:2001.03838.*

Robinson, P. M. (1988): "Root-N-Consistent Semiparametric Regression," *Econometrica*, 56, 931.

Roth, A. E. and M. Sotomayor (1992): "Two-sided matching," *Handbook of game theory with economic applications*, 1, 485–541.

Sacerdote, B. (2011): "Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far?" in *Handbook of the Economics of Education*, Elsevier, 249–277.

Serfling, R. J. (1980): *Approximation theorems of mathematical statistics.*

Sun, X. (2019): "Identification and estimation of many-to-one matching with an application to the U.S. college admissions," *Working paper, UCLA*.

Train, K. E. (2009): *Discrete choice methods with simulation*, Cambridge university press.

Wooldridge, J. M. (2010): *Econometric analysis of cross section and panel data*, MIT press.

Zimmerman, D. J. (2003): "Peer effects in academic outcomes: Evidence from a natural experiment," *Review of Economics and statistics*, 85, 9–23.

Online Appendix to

Social Interactions with Endogenous Group Formation

Shuyang Sheng Xiaoting Sun

O.A Simulations: Group Formation Parameters

The parameters to be estimated include the group formation parameters δ, the group fixed effects α, and the cutoffs p, a total of 14 parameters. We use maximum simulated likelihood (MSL) for estimation. Let $\eta_i^{(1)}, \ldots, \eta^{(R)}$ be R draws from the multivariate standard normal distribution. The simulated conditional probability for each $g \in \mathcal{G}$ is

$$\check{\sigma}_g(z_i) = \frac{1}{R}\sum_{r=1}^{R} \frac{\exp\left(\alpha_g + \delta_1^u z_{1,ig}^u + \delta_2^u z_{2,i}\right) \cdot 1\left(\delta_1^v z_{1,ig}^v + \delta_2^v z_{2,i} + \eta_{gi}^{(r)} \geq p_g\right)}{\sum_{k=1}^{G} \exp\left(\alpha_k + \delta_1^u z_{1,ik}^u + \delta_2^u z_{2,i}\right) \cdot 1\left(\delta_1^v z_{1,ik}^v + \delta_2^v z_{2,i} + \eta_{ki}^{(r)} \geq p_k\right) + 1},$$

and for those who do not join any group,

$$\check{\sigma}_0(z_i) = \frac{1}{R}\sum_{r=1}^{R} \frac{1}{\sum_{k=1}^{G} \exp\left(\alpha_k + \delta_1^u z_{1,ik}^u + \delta_2^u z_{2,i}\right) \cdot 1\left(\delta_1^v z_{1,ik}^v + \delta_2^v z_{2,i} + \eta_{ki}^{(r)} \geq p_k\right) + 1}.$$

The MSL estimator $(\hat{\delta}, \hat{\alpha}, \hat{p})$ maximizes

$$L(\delta, \alpha, p) = \frac{1}{n}\sum_{i=1}^{n}\sum_{g=0}^{G} 1\left(g_i = g\right) \log \check{\sigma}_g(z_i).$$

To mitigate the numerical difficulties caused by the nonsmoothness of the indicator functions, we replace $1\left(\delta_1^v z_{1,ig}^v + \delta_2^v z_{2,i} + \eta_{gi}^{(r)} \geq p_g\right)$ with a smoothed A-R simulator(McFadden, 1989; Train, 2009) $1/(1 + \exp(p_g - (\delta_1^v z_{1,ig}^v + \delta_2^v z_{2,i} + \eta_{gi}^{(r)}))/\kappa)$, where

$\kappa > 0$ is a scale factor which we specify as 0.05.[29] The estimation results are presented in Table O.A.1.

Table O.A.1: Estimation Results for Group Formation

Group formation parameters			Group fixed effects			Cutoffs		
	bias	sd		bias	sd		bias	sd
δ_1^u	0.026	0.073	α_1	0.101	0.652	p_1	0.048	0.142
δ_1^v	0.025	0.196	α_2	0.049	0.472	p_2	0.014	0.163
δ_2^u	-0.047	0.039	α_3	-0.031	0.448	p_3	0.002	0.163
δ_2^v	-0.069	0.044	α_4	-0.044	0.467	p_4	-0.027	0.177
			α_5	-0.056	0.519	p_5	0.004	0.226

Note: This table presents estimates for the coefficients of the group formation model. The coefficients are estimated using 200 MC samples, where each sample contains 5 groups and 2,000 individuals.

O.B Lemmas in the Asymptotic Analysis of $\hat{p}$ and $\hat{\alpha}$

Notation Define $\tau_i^v(\delta^v) = (\tau_{ig}^v(\delta_g^v), g \in \mathcal{G})$. Then $\tau_i^v = \tau_i^v(\delta^v)$ and $\hat{\tau}_i^v = \tau_i^v(\hat{\delta}^v)$. Similarly, for $h \neq 1, g$, define $\sigma_h(\delta^v) = (\sigma_{h|g}(\tau_{ig}^v(\delta_g^v)), \sigma_{h|1}(\tau_{i1}^v(\delta_1^v)))$ and $\hat{\sigma}_h(\delta^v) = (\hat{\sigma}_{h|g}(\tau_{ig}^v(\delta_g^v)), \hat{\sigma}_{h|1}(\tau_{i1}^v(\delta_1^v)))$. We have $\sigma_h = \sigma_h(\delta^v)$ and $\hat{\sigma}_h = \hat{\sigma}_h(\hat{\delta}^v)$. For any $g \neq h$, define $\pi_{h|g,i} = \sigma_{h|g}(\tau_{ig}^v)$ and $\hat{\pi}_{h|g,i} = \hat{\sigma}_{h|g}(\tau_{ig}^v)$. Under Assumption 4(ii), the inverse of $\sigma_{h|g}(\cdot)$ exists and hence $\tau_{ig}^v = \sigma_{h|g}^{-1}(\pi_{h|g,i})$.[30] Let $f_{\pi_{h|g}}$ denote the pdf of $\pi_{h|g,i}$. Let $0 < C < \infty$ denote a universal constant.

Lemma O.B.1. *The term T_{1n} in equation (A.8) satisfies $T_{1n} = n^{-1}\sum_{i=1}^n \psi_{g,h,1}(z_i, g_i) + o_p(n^{-1/2})$, where $\psi_{g,h,1}(z_i, g_i)$ is defined in equation (O.B.10).*

Proof. We decompose T_{1n} as

$$\begin{aligned} T_{1n} \quad = \quad & \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j\neq i}^n (h_{gn}(\hat{\tau}_i^v, \hat{\tau}_j^v; \hat{\sigma}_h(\hat{\delta}^v)) - h_{gn}(\tau_i^v, \tau_j^v; \hat{\sigma}_h(\delta^v))) \\ & + \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1, j\neq i}^n (h_{gn}(\tau_i^v, \tau_j^v; \hat{\sigma}_h(\delta^v)) - h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)) \end{aligned}$$

[29] The smaller κ is, the better the simulator approximates the indicator function.

[30] Under Assumption 4(ii), we can show that $\sigma_{h|g}(\tau)$ is strictly monotone for all τ and thus its inverse function $\sigma_{h|g}^{-1}(\tau)$ exists.

$$\equiv \quad T^{\delta}_{1n} + T^{\sigma}_{1n}, \tag{O.B.1}$$

where T^{δ}_{1n} captures the estimation error due to $\hat{\delta}^v$, and T^{σ}_{1n} captures the estimation error due to $\hat{\sigma}_h$.

Step 1: we start with T^{σ}_{1n} in equation (O.B.1). Define the remainder term

$$\begin{aligned} R_{n,ij} &= \frac{1}{\zeta_{2n}} K_2(\frac{\hat{\pi}_{h|g,i} - \hat{\pi}_{h|1,j}}{\zeta_{2n}}) - \frac{1}{\zeta_{2n}} K_2(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}}) \\ &\quad - \frac{1}{\zeta^2_{2n}} K'_2(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}})((\hat{\pi}_{h|g,i} - \pi_{h|g,i}) - (\hat{\pi}_{h|1,j} - \pi_{h|1,j})) \\ &= \frac{1}{2}\zeta^{-3}_{2n} K''_2(\triangle_{n,ij})((\hat{\pi}_{h|g,i} - \pi_{h|g,i}) - (\hat{\pi}_{h|1,j} - \pi_{h|1,j}))^2, \end{aligned}$$

where the last equality follows by Taylor expansion, with $\triangle_{n,ij}$ an intermediate value between $\frac{\hat{\pi}_{h|g,i} - \hat{\pi}_{h|1,j}}{\zeta_{2n}}$ and $\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}}$. We can write

$$\begin{aligned} T^{\sigma}_{1n} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{1}{\zeta^2_{2n}} K'_2(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}})(\tau^v_{ig} - \tau^v_{j1}) \\ &\quad \cdot((\hat{\pi}_{h|g,i} - \pi_{h|g,i}) - (\hat{\pi}_{h|1,j} - \pi_{h|1,j})) \\ &\quad + \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} R_{n,ij}(\tau^v_{ig} - \tau^v_{j1}). \end{aligned}$$

For any $g \neq h$, write $\hat{\pi}_{h|g,i} = \hat{b}_{h|g}(\tau^v_{ig})/\hat{b}_g(\tau^v_{ig})$, where $\hat{b}_{h|g}(\tau^v_{ig}) = \frac{1}{n\zeta_{1n}} \sum_{k=1}^{n} 1\{g_k = h\} K_1(\frac{\tau^v_{ig} - \tau^v_{kg}}{\zeta_{1n}})$ and $\hat{b}_g(\tau^v_{ig}) = \frac{1}{n\zeta_{1n}} \sum_{k=1}^{n} K_1(\frac{\tau^v_{ig} - \tau^v_{kg}}{\zeta_{1n}})$. Moreover, let $b_g(\tau^v_{ig})$ denote the pdf of τ^v_{ig}, and define $b_{h|g}(\tau^v_{ig}) = \pi_{h|g,i} b_g(\tau^v_{ig})$. We express $\hat{\pi}_{h|g,i} - \pi_{h|g,i}$ as a linear functional of the kernel estimators $\hat{b}_{h|g}$ and $\hat{b}_g$. Specifically, we follow Newey and McFadden (1994, p.2204) and Newey (1994b, Lemma B.3)[31] and derive

$$\begin{aligned} &\max_i |\hat{\pi}_{h|g,i} - \pi_{h|g,i} - 1/b_g(\tau^v_{ig})(\hat{b}_{h|g}(\tau^v_{ig}) - \hat{b}_g(\tau^v_{ig})\pi_{h|g,i})| \\ \leq\ & \max_i 1/(|\hat{b}_g(\tau^v_{ig})| b_g(\tau^v_{ig}))(1 + \pi_{h|g,i})((\hat{b}_{h|g}(\tau^v_{ig}) - b_{h|g}(\tau^v_{ig}))^2 + (\hat{b}_g(\tau^v_{ig}) - b_g(\tau^v_{ig}))^2) \\ \leq\ & O_p(1) \sup_\tau ((\hat{b}_{h|g}(\tau) - b_{h|g}(\tau))^2 + (\hat{b}_g(\tau) - b_g(\tau))^2) \\ =\ & O_p(((\ln n)^{1/2}(n\zeta_{1n})^{-1/2} + \zeta^{s_1}_{1n})^2), \end{aligned} \tag{O.B.2}$$

[31] Lemma B.3 in Newey (1994b) holds by Assumptions 6, 8(ii)(iii), 10(i), 11(i), and $n^{1/2}\zeta_{1n}/\ln n \to \infty$. The last condition is implied by Assumption 9(ii). To see this, note that the second condition of Assumption 9(ii) implies that $n\zeta^3_{1n} \to \infty$, or $\zeta_{1n} = c_n n^{-1/3}$ with $c_n \to \infty$. Therefore, $n^{1/2}\zeta_{1n}/\ln n = c_n n^{1/2}/\ln n \to \infty$, and Lemma B.3 in Newey (1994b) holds under the assumptions we impose here.

where the second inequality holds because $b_g(\cdot)$ is bounded away from zero under Assumption 8(i) and $\hat{b}_g(\cdot)$ is uniformly close to $b_g(\cdot)$. Under Assumption 9(i), we can bound the linearization error as $\zeta_{2n}^{-2}\max_i|\hat{\pi}_{h|g,i}-\pi_{h|g,i}-1/b_g(\tau_{ig}^v)(\hat{b}_{h|g}(\tau_{ig}^v)-\hat{b}_g(\tau_{ig}^v)\pi_{h|g,i})|=o_p(n^{-1/2})$ and similarly $\zeta_{2n}^{-2}\max_{i,j}|\hat{\pi}_{h|1,i}-\pi_{h|1,i}-1/b_1(\tau_{i1}^v)(\hat{b}_{h|1}(\tau_{i1}^v)-\hat{b}_1(\tau_{i1}^v)\pi_{h|1,i})|=o_p(n^{-1/2})$. Applying the boundedness of $K_2'(\cdot)$ and τ_{ig}^v (Assumptions 7(ii)(iii), 10(i), and 11(i)) we can see that the overall linearization error is $o_p(n^{-1/2})$.

Further, observe that $\max_i|1/b_g(\tau_{ig}^v)(\hat{b}_{h|g}(\tau_{ig}^v)-\hat{b}_g(\tau_{ig}^v)\pi_{h|g,i})|\leq C\sup_\tau(|\hat{b}_{h|g}(\tau)-b_{h|g}(\tau)|+|\hat{b}_g(\tau)-b_g(\tau)|)=O_p((\ln n)^{1/2}(n\zeta_{1n})^{-1/2}+\zeta_{1n}^{s_1})$ by Newey (1994b, Lemma B.3). Combining this with equation (O.B.2) yields $\max_i|\hat{\pi}_{h|g,i}-\pi_{h|g,i}|=O_p((\ln n)^{1/2}(n\zeta_{1n})^{-1/2}+\zeta_{1n}^{s_1})$ and similarly for $\hat{\pi}_{h|1,j}-\pi_{h|1,j}$. Hence, by Assumptions 7(ii)(iii), Assumption 8(i), 9(i), 10(i), and 11(i), the remainder term is negligible, that is, $\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n|R_{n,ij}(\tau_{ig}^v-\tau_{j1}^v)|=\zeta_{2n}^{-3}O_p(((\ln n)^{1/2}(n\zeta_{1n})^{-1/2}+\zeta_{1n}^{s_1})^2)=o_p(n^{-1/2})$.

Overall, we obtain

$$
\begin{aligned}
T_{1n}^\sigma &= \frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\frac{1}{\zeta_{2n}^2}K_2'(\frac{\pi_{h|g,i}-\pi_{h|1,j}}{\zeta_{2n}})(\tau_{ig}^v-\tau_{j1}^v)\\
&\quad\cdot\left(\frac{\hat{b}_{h|g}(\tau_{ig}^v)-\hat{b}_g(\tau_{ig}^v)\pi_{h|g,i}}{b_g(\tau_{ig}^v)}-\frac{\hat{b}_{h|1}(\tau_{j1}^v)-\hat{b}_1(\tau_{j1}^v)\pi_{h|1,j}}{b_1(\tau_{j1}^v)}\right)+o_p(n^{-1/2}) \qquad \text{(O.B.3)}
\end{aligned}
$$

Let $\omega_i=(\tau_i^v,g_i)$. Plugging in the expressions of the kernel estimators, we can further write

$$
\begin{aligned}
T_{1n}^\sigma &= \frac{1}{n^2(n-1)}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n 1\{i\neq j\}q_n(\omega_i,\omega_j,\omega_k)+o_p(n^{-1/2})\\
&= \frac{1}{n(n-1)(n-2)}\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n 1\{i\neq j\neq k\}q_n(\omega_i,\omega_j,\omega_k)+o_p(n^{-1/2})\\
&\equiv Q_n+o_p(n^{-1/2}), \qquad \text{(O.B.4)}
\end{aligned}
$$

where

$$
\begin{aligned}
& q_n(\omega_i,\omega_j,\omega_k)\\
=\ & \frac{1}{\zeta_{1n}\zeta_{2n}^2}K_2'(\frac{\pi_{h|g,i}-\pi_{h|1,j}}{\zeta_{2n}})(\tau_{ig}^v-\tau_{j1}^v)\\
& \cdot\left(K_1(\frac{\tau_{ig}^v-\tau_{kg}^v}{\zeta_{1n}})\frac{1\{g_k=h\}-\pi_{h|g,i}}{b_g(\tau_{ig}^v)}-K_1(\frac{\tau_{j1}^v-\tau_{k1}^v}{\zeta_{1n}})\frac{1\{g_k=h\}-\pi_{h|1,j}}{b_1(\tau_{j1}^v)}\right).
\end{aligned}
$$

The second equality in (O.B.4) follows because the terms with $i = k$ or $j = k$ are negligible.[32] Q_n is a V-statistic of degree 3 with an asymmetric kernel function.

We derive an asymptotically linear representation of Q_n using the idea of Hoeffding projection (Hoeffding, 1948). Let $I = (i_1, i_2, i_3)$ with $i_1 \neq i_2 \neq i_3$ and $\omega_I = (\omega_i : i \in I)$. We can write $Q_n = \frac{1}{n(n-1)(n-2)} \sum_I q_n(\omega_I)$. Define $q_n^*(\omega_i) = \frac{1}{(n-1)(n-2)} \sum_{I:i\in I} \mathbb{E}[q_n(\omega_I)|\omega_i]$ and $Q_n^* = \frac{1}{n}\sum_{i=1}^n q_n^*(\omega_i) - 2\mathbb{E}[q_n(\omega_I)]$. Observe that $\mathbb{E}[q_n^*(\omega_i)] = 3\mathbb{E}[q_n(\omega_I)]$ and thus $\mathbb{E}[Q_n - Q_n^*] = 0$.[33] By construction, we obtain $\text{Cov}(Q_n, Q_n^*) = n^{-1}\sum_{i=1}^n \text{Cov}(Q_n, q_n^*(\omega_i))$ and for each i,

$$\begin{aligned}
\text{Cov}(Q_n, q_n^*(\omega_i)) &= \frac{1}{n(n-1)(n-2)} \sum_{I:i\in I} \text{Cov}(q_n(\omega_I), q_n^*(\omega_i)) \\
&= \frac{1}{n(n-1)(n-2)} \sum_{I:i\in I} \text{Cov}(\mathbb{E}[q_n(\omega_I)|\omega_i], q_n^*(\omega_i)) \\
&= \frac{1}{n}\text{Var}(q_n^*(\omega_i)),
\end{aligned}$$

where the first equality holds because for $i \notin I$ we have $\text{Cov}(q_n(\omega_I), q_n^*(\omega_i)) = 0$, and the second equality holds by iterated expectations. It then follows that $\text{Cov}(Q_n, Q_n^*) = n^{-2}\sum_{i=1}^n \text{Var}(q_n^*(\omega_i)) = \text{Var}(Q_n^*)$ and hence $\mathbb{E}[(Q_n - Q_n^*)^2] = \text{Var}(Q_n) - \text{Var}(Q_n^*)$. By Markov inequality, we obtain $Q_n = Q_n^* + o_p(n^{-1/2})$ if $\text{Var}(Q_n) - \text{Var}(Q_n^*) = o(n^{-1})$.

To show the last result, note that $q_n(\omega_I)$ and $q_n(\omega_J)$ are independent for disjoint I and J. Therefore,

$$\begin{aligned}
\text{Var}(Q_n) &= \frac{1}{(n(n-1)(n-2))^2} \sum_{(I,J):|I\cap J|=1} \text{Cov}(q_n(\omega_I), q_n(\omega_J)) \\
&\quad + \frac{1}{(n(n-1)(n-2))^2} \sum_{(I,J):|I\cap J|>1} \text{Cov}(q_n(\omega_I), q_n(\omega_J)). \quad \text{(O.B.5)}
\end{aligned}$$

For comparison, because $\text{Var}(Q_n^*) = n^{-2}\sum_{i=1}^n \text{Var}(q_n^*(\omega_i))$ we can write

$$\begin{aligned}
&\text{Var}(Q_n^*) \\
&= \frac{1}{(n(n-1)(n-2))^2} \sum_{i=1}^n \sum_{(I,J):\{i\}=I\cap J} \text{Cov}(\mathbb{E}[q_n(\omega_I)|\omega_i], \mathbb{E}[q_n(\omega_J)|\omega_i])
\end{aligned}$$

[32] The terms with $i = k$ are negligible because $\frac{1}{n^2(n-1)}\sum_{i=1}^n\sum_{j=1}^n 1\{i \neq j\} q_n(\omega_i, \omega_j, \omega_i) \leq (n\zeta_{1n}\zeta_{2n}^2)^{-1} \sup_t |K_1(t)| \sup_t |K_2'(t)| \frac{1}{n(n-1)} \sum_{i=1}^n\sum_{j=1}^n 1\{i \neq j\}|\tau_{ig}^v - \tau_{j1}^v|(1/b_g(\tau_{ig}^v) + 1/b_1(\tau_{j1}^v)) = O_p((n\zeta_{1n}\zeta_{2n}^2)^{-1}) = o_p(n^{-1/2})$ by Assumptions 6(ii)(iii), 7(ii)(iii), 8(i), 9(iii), 10(i), and 11(i). A similar argument shows that the terms with $j = k$ are negligible.

[33] The sum over I with $i \in I$ consists of $3(n-1)(n-2)$ terms.

$$+\frac{1}{(n(n-1)(n-2))^2}\sum_{i=1}^{n}\sum_{(I,J):\{i\}\subsetneq I\cap J}\text{Cov}(\mathbb{E}[q_n(\omega_I)|\omega_i],\mathbb{E}[q_n(\omega_J)|\omega_i]).\quad\text{(O.B.6)}$$

For I and J such that $\{i\}=I\cap J$, $q_n(\omega_I)$ and $q_n(\omega_J)$ are independent conditional on ω_i and thus $\text{Cov}(q_n(\omega_I),q_n(\omega_J))=\text{Cov}(\mathbb{E}[q_n(\omega_I)|\omega_i],\mathbb{E}[q_n(\omega_J)|\omega_i])$. This implies that the first sum in $\text{Var}(Q_n)$ is equal to the first sum in $\text{Var}(Q_n^*)$. Moreover, the second sums in both $\text{Var}(Q_n)$ and $\text{Var}(Q_n^*)$ consist of $O(n^4)$ terms. For any I and J, by Assumptions 6(ii)(iii), 7(ii)(iii), 8(i), 10(i), and 11(i), we can bound both $\text{Cov}(q_n(\omega_I),q_n(\omega_J))$ and $\text{Cov}(\mathbb{E}[q_n(\omega_I)|\omega_i],\mathbb{E}[q_n(\omega_J)|\omega_i])$ uniformly by $O(\zeta_{1n}^{-2}\zeta_{2n}^{-4})$. Therefore, the second sums in $\text{Var}(Q_n)$ and $\text{Var}(Q_n^*)$ are both $(n(n-1)(n-2))^{-2}\cdot O(n^4)\cdot O(\zeta_{1n}^{-2}\zeta_{2n}^{-4})=O(n^{-2}\zeta_{1n}^{-2}\zeta_{2n}^{-4})=o(n^{-1})$ by Assumption 9(iii). Combining the results yields $\text{Var}(Q_n)-\text{Var}(Q_n^*)=o(n^{-1})$ and thus $Q_n=Q_n^*+o_p(n^{-1/2})$.

Now we calculate the influence function of Q_n^*. Because $\omega_i=(\tau_i^v,g_i)$ are i.i.d., to calculate $q_n^*(\omega_i)$ it is sufficient to consider $\mathbb{E}[q_n(\omega_I)|\omega_i]$ for the three cases where i appears as the first, second, or third element of I. Fix $i\neq j\neq k$. We start with the case where i appears as the third element (i.e., $\mathbb{E}[q_n(\omega_j,\omega_k,\omega_i)|\omega_i]$). Write $q_n(\omega_j,\omega_k,\omega_i)=q_{gn}(\omega_j,\omega_k,\omega_i)-q_{1n}(\omega_j,\omega_k,\omega_i)$. Note that conditional on ω_i, τ_{ig}^v and $\pi_{h|g,i}=\sigma_{h|g}(\tau_{ig}^v)$ are constants. By the change of variables $t=\zeta_{2n}^{-1}(\pi_{h|g,j}-\pi_{h|1,k})$ and using $\tau_{k1}^v=\sigma_{h|1}^{-1}(\pi_{h|1,k})$, we obtain

$$\begin{aligned}\mathbb{E}[q_{gn}(\omega_j,\omega_k,\omega_i)|\omega_i,\omega_j] &= \frac{1}{\zeta_{1n}}K_1\left(\frac{\tau_{jg}^v-\tau_{ig}^v}{\zeta_{1n}}\right)\frac{1\{g_i=h\}-\pi_{h|g,j}}{b_g(\tau_{jg}^v)}\\ &\quad\cdot\int\frac{1}{\zeta_{2n}}K_2'(t)(\sigma_{h|1}^{-1}(\pi_{h|g,j}-t\zeta_{2n})-\tau_{jg}^v)f_{\pi_{h|1}}(\pi_{h|g,j}-t\zeta_{2n})\mathrm{d}t.\end{aligned}$$

Moreover, by Assumptions 7(i)(iii), 8(ii)(iii), 10(i), and 11(i),[34] we obtain

$$\begin{aligned}&\int\frac{1}{\zeta_{2n}}K_2'(t)(\sigma_{h|1}^{-1}(\pi_{h|g,j}-t\zeta_{2n})-\tau_{jg}^v)f_{\pi_{h|1}}(\pi_{h|g,j}-t\zeta_{2n})\mathrm{d}t\\ =&\int K_2(t)\frac{\partial((\tau_{jg}^v-\sigma_{h|1}^{-1}(\pi_{h|g,j}-t\zeta_{2n}))f_{\pi_{h|1}}(\pi_{h|g,j}-t\zeta_{2n}))}{\partial\pi_{h|g,j}}\mathrm{d}t\end{aligned}$$

[34]For any $g\in\mathcal{G}$, by definition $\pi_{h|g,i}=\sigma_{h|g}(\tau_{ig}^v)$ and thus $f_{\pi_{h|g}}(\pi_{h|g,i})=b_g(\sigma_{h|g}^{-1}(\pi_{h|g,i}))|(\sigma_{h|g}^{-1})'(\pi_{h|g,i})|=b_g(\tau_{ig}^v)|(\sigma_{h|g}^{-1})'(\pi_{h|g,i})|$. By the chain rule, the (s_2+2)th order continuous differentiability of $\sigma_{h|g}(\tau)$ (Assumption 8(iii)) implies that $\sigma_{h|g}^{-1}(\tau)$ is (s_2+2)th continuously differentiable. Because b_g is (s_2+1)th continuously differentiable (Assumption 8(ii)), we derive that $f_{\pi_{h|g}}$ is (s_2+1)th continuously differentiable. The boundedness of τ_{ig}^v (Assumptions 10(i) and 11(i)) then implies that the (s_2+1)th derivative of $f_{\pi_{h|g}}$ is bounded.

$$= \frac{\partial((\tau^v_{jg} - \sigma^{-1}_{h|1}(\pi_{h|g,j}))f_{\pi_{h|1}}(\pi_{h|g,j}))}{\partial \pi_{h|g,j}} + r_{gn}(\tau^v_{jg}),$$

with $\max_{i,j} |r_{gn}(\tau^v_{jg})| \leq C\zeta^{s_2}_{2n}$ and hence

$$\begin{aligned}
& \mathbb{E}[q_{gn}(\omega_j, \omega_k, \omega_i)|\omega_i] \\
= \; & \mathbb{E}[\mathbb{E}[q_{gn}(\omega_j, \omega_k, \omega_i)|\omega_i, \omega_j]|\omega_i] \\
= \; & \int K_1(t)(1\{g_i = h\} - \sigma_{h|g}(\tau^v_{ig} + t\zeta_{1n})) \\
& \cdot \left(\frac{\partial(\tau^v_{ig} + t\zeta_{1n} - \sigma^{-1}_{h|1}(\sigma_{h|g}(\tau^v_{ig} + t\zeta_{1n})))f_{\pi_{h|1}}(\sigma_{h|g}(\tau^v_{ig} + t\zeta_{1n}))}{\partial \sigma_{h|g}((\tau^v_{ig} + t\zeta_{1n}))} + r_{gn}(\tau^v_{ig} + t\zeta_{1n}) \right) \mathrm{d}t \\
= \; & (1\{g_i = h\} - \pi_{h|g,i}) \frac{\partial(\tau^v_{ig} - \sigma^{-1}_{h|1}(\pi_{h|g,i}))f_{\pi_{h|1}}(\pi_{h|g,i})}{\partial \pi_{h|g,i}} + r_{gn,i},
\end{aligned}$$

with $\max_i |r_{gn,i}| \leq C(\zeta^{s_1}_{1n} + \zeta^{s_2}_{2n})$. The second equality follows from the change of variables $t = \zeta^{-1}_{1n}(\tau^v_{jg} - \tau^v_{ig})$,[35] and the last equality follows from Taylor expansion and Assumptions 6(i), 8(ii)(iii), 10(i), and 11(i). The form of $\mathbb{E}[q_{1n}(\omega_j, \omega_k, \omega_i)|\omega_i]$ can be derived similarly. Overall, we obtain $\mathbb{E}[q_n(\omega_j, \omega_k, \omega_i)|\omega_i] = \psi^\sigma_{g,h}(\omega_i) + r_{n,i}$, where

$$\begin{aligned}
\psi^\sigma_{g,h}(\omega_i) \;=\; & (1\{g_i = h\} - \pi_{h|g,i}) \frac{\partial(\tau^v_{ig} - \sigma^{-1}_{h|1}(\pi_{h|g,i}))f_{\pi_{h|1}}(\pi_{h|g,i})}{\partial \pi_{h|g,i}} \\
& -(1\{g_i = h\} - \pi_{h|1,i}) \frac{\partial(\tau^v_{i1} - \sigma^{-1}_{h|g}(\pi_{h|1,i}))f_{\pi_{h|g}}(\pi_{h|1,i})}{\partial \pi_{h|g,i}}, \qquad \text{(O.B.7)}
\end{aligned}$$

and $r_{n,i}$ satisfies $\max_i |r_{n,i}| \leq C(\zeta^{s_1}_{1n} + \zeta^{s_2}_{2n})$. Using similar arguments, we can show that both $\max_i |\mathbb{E}[q_n(\omega_i, \omega_j, \omega_k)|\omega_i]|$ and $\max_i |\mathbb{E}[q_n(\omega_j, \omega_i, \omega_k)|\omega_i]|$ are bounded by $C\zeta^{s_1}_{1n}$. We obtain $q^*_n(\omega_i) = \mathbb{E}[q_n(\omega_j, \omega_k, \omega_i)|\omega_i] + \mathbb{E}[q_n(\omega_i, \omega_j, \omega_k)|\omega_i] + \mathbb{E}[q_n(\omega_j, \omega_i, \omega_k)|\omega_i] = \psi^\sigma_{g,h}(\omega_i) + r^*_{n,i}$, where $\max_i |r^*_{n,i}| \leq C(\zeta^{s_1}_{1n} + \zeta^{s_2}_{2n})$.

Note that $\mathbb{E}[1\{g_i = h\} - \pi_{h|g,i}|\tau^v_{ig}] = 0$, so $\mathbb{E}[\psi^\sigma_{g,h}(\omega_i)] = 0$. This together with $\max_i |r^*_{n,i}| \leq C(\zeta^{s_1}_{1n} + \zeta^{s_2}_{2n})$ and Assumption 9(iv)(v) yields $\mathbb{E}[q^*_n(\omega_i)] = o(n^{-1/2})$. It follows that $T^\sigma_{1n} = Q^*_n + o_p(n^{-1/2}) = \frac{1}{n}\sum^n_{i=1} \psi^\sigma_{g,h}(\omega_i) + o_p(n^{-1/2})$.

Step 2: now we examine T^δ_{1n} in equation (O.B.1). Under Assumption 6(ii) $\hat{\sigma}_h$ is twice differentiable in δ, so by applying the Taylor expansion we obtain $h_{gn}(\hat{\tau}^v_i, \hat{\tau}^v_j; \hat{\sigma}_h(\hat{\delta}^v)) - h_{gn}(\tau^v_i, \tau^v_j; \hat{\sigma}_h(\delta^v)) = \frac{\partial h_{gn}(\tau^v_i, \tau^v_j; \hat{\sigma}_h(\delta^v))}{\partial \delta^{v\prime}}(\hat{\delta}^v - \delta^v) + O_p(\|\hat{\delta}^v - \delta^v\|^2)$. Because $\hat{\delta}^v - \delta^v = n^{-1}\sum^n_{i=1} \psi^{\delta^v}(z_i) + o_p(n^{-1/2})$, where $\psi^{\delta^v}(z_i)$ is the influence function of $\hat{\delta}^v$ with $\mathbb{E}[\psi^{\delta^v}(z_i)] =$

[35] Recall that $b_g(\tau^v_{jg})$ is the density of τ^v_{jg}, so the two cancel out.

0 (Assumption 10(ii)), we have

$$T_{1n}^{\delta} = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\frac{\partial h_{gn}(\tau_i^v,\tau_j^v;\hat{\sigma}_h(\delta^v))}{\partial\delta^{v\prime}}\right)\psi^{\delta^v}(z_k)+o_p(n^{-1/2}).$$

The double sum term in parentheses is a V-statistic with a nested kernel estimator $\hat{\sigma}_h(\delta^v)$.

Observe that $\frac{\partial h_{gn}(\tau_i^v,\tau_j^v;\hat{\sigma}_h(\delta^v))}{\partial\delta^{v\prime}} = (\frac{\partial h_{gn}(\tau_i^v,\tau_j^v;\hat{\sigma}_h(\delta^v))}{\partial\delta_1^{v\prime}},\ldots,\frac{\partial h_{gn}(\tau_i^v,\tau_j^v;\hat{\sigma}_h(\delta^v))}{\partial\delta_G^{v\prime}})$, where

$$\begin{aligned}
\frac{\partial h_{gn}(\tau_i^v,\tau_j^v;\hat{\sigma}_h(\delta^v))}{\partial\delta_1^{v\prime}} &= -\frac{1}{\zeta_{2n}^2}K_2'(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})\hat{\sigma}_{h|1}'(\tau_{j1}^v)(\tau_{ig}^v-\tau_{j1}^v)z_j' \\
&\quad -\frac{1}{\zeta_{2n}}K_2(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})z_j' \\
\frac{\partial h_{gn}(\tau_i^v,\tau_j^v;\hat{\sigma}_h(\delta^v))}{\partial\delta_g^{v\prime}} &= \frac{1}{\zeta_{2n}^2}K_2'(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})\hat{\sigma}_{h|g}'(\tau_{ig}^v)(\tau_{ig}^v-\tau_{j1}^v)z_i' \\
&\quad +\frac{1}{\zeta_{2n}}K_2(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})z_i',
\end{aligned}$$

and the remaining $G-2$ subvectors equal to zero. We focus on the gth subvector, and the first subvector can be analyzed similarly. Applying the mean-value theorem (with $\triangle_{n,ij}$ an intermediate value), we obtain the bound

$$\begin{aligned}
& \zeta_{2n}^{-1}\max_{i,j}\left|K_2(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})-K_2(\frac{\pi_{h|g,i}-\pi_{h|1,j}}{\zeta_{2n}})\right| \\
\leq\ & \zeta_{2n}^{-2}\max_{i,j}|K_2'(\Delta_{n,ij})||\hat{\pi}_{h|g,i}-\pi_{h|g,i}-(\hat{\pi}_{h|1,j}-\pi_{h|1,j})| \\
=\ & \zeta_{2n}^{-2}O_p((\ln n)^{1/2}(n\zeta_{1n})^{-1/2}+\zeta_{1n}^{s_1})=o_p(1)
\end{aligned}$$

by Assumptions 7(ii)(iii), 9(ii), and Newey (1994b, Lemma B.3). Similarly, we have

$$\begin{aligned}
& \zeta_{2n}^{-2}\max_{i,j}\left|K_2'(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})\hat{\sigma}_{h|g}'(\tau_{ig}^v)-K_2'(\frac{\pi_{h|g,i}-\pi_{h|1,j}}{\zeta_{2n}})\sigma_{h|g}'(\tau_{ig}^v)\right| \\
\leq\ & \zeta_{2n}^{-2}\max_{i,j}\left|K_2'(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})\right|\sup_{\tau}|\hat{\sigma}_{h|g}'(\tau_{ig}^v)-\sigma_{h|g}'(\tau_{ig}^v)| \\
& +\zeta_{2n}^{-2}\max_{i,j}\left|K_2'(\frac{\hat{\pi}_{h|g,i}-\hat{\pi}_{h|1,j}}{\zeta_{2n}})-K_2'(\frac{\pi_{h|g,i}-\pi_{h|1,j}}{\zeta_{2n}})\right|\sup_{\tau}|\sigma_{h|g}'(\tau_{ig}^v)| \\
=\ & \zeta_{2n}^{-2}O_p((\ln n)^{1/2}(n\zeta_{1n}^3)^{-1/2}+\zeta_{1n}^{s_1})+\zeta_{2n}^{-3}O_p((\ln n)^{1/2}(n\zeta_{1n})^{-1/2}+\zeta_{1n}^{s_1})=o_p(1).
\end{aligned}$$

by Assumptions 7(ii)(iii), 8(i)(ii), 9(ii), 10(i), 11(i), Sun (2019, Lemma 3) and Newey (1994b, Lemma B.3).Therefore, by the boundedness of τ_{ig}^v and z_i (Assumptions 10(i) and 11(i)), we have the approximation

$$\begin{aligned} &\left\| \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \left(\frac{\partial h_{gn}(\tau_i^v, \tau_j^v; \hat{\sigma}_h(\delta^v))}{\partial \delta_g^{v\prime}} - \frac{\partial h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)}{\partial \delta_g^{v\prime}}\right) \right\| \\ \leq \;& o_p(1) \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (\|(\tau_{ig}^v - \tau_{j1}^v) z_i\| + \|z_i\|) = o_p(1). \end{aligned}$$

Following a standard argument for the Law of Large Number for U-statistics Serfling (1980, p. 190), we can show

$$\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \frac{\partial h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)}{\partial \delta_g^{v\prime}} = \mathbb{E}\left[\frac{\partial h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)}{\partial \delta_g^{v\prime}}\right] + o_p(1).$$

Now we derive the form of $\mathbb{E}[\partial h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)/\partial \delta_g^{v\prime}]$. Recall that

$$\frac{\partial h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)}{\partial \delta_g^{v\prime}} = \frac{1}{\zeta_{2n}^2} K_2'\left(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}}\right) \sigma_{h|g}'(\tau_{ig}^v)(\tau_{ig}^v - \tau_{j1}^v) z_i' + \frac{1}{\zeta_{2n}^1} K_2\left(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}}\right) z_i'.$$

By the change of variables $t = \zeta_{2n}^{-1}(\pi_{h|g,i} - \pi_{h|1,j})$, integration by parts, Taylor expansion, and Assumption 2(i), 7(i)(iii), and 8(ii)(iii), 10(i), and 11(i), we have

$$\begin{aligned} \left|\int \zeta_{2n}^{-2} K_2'\left(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}}\right) f_{\pi_{h|1}}(\pi_{h|1,j}) \mathrm{d}\pi_{h|1,j} + f_{\pi_{h|1}}'(\pi_{h|g,i})\right| &\leq C\zeta_{2n}^{s_2} \\ \left|\int \zeta_{2n}^{-2} K_2'\left(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}}\right) \tau_{j1}^v f_{\pi_{h|1}}(\pi_{h|1,j}) \mathrm{d}\pi_{h|1,j} + (\sigma_{h|1}^{-1}(\pi_{h|g,i}) f_{\pi_{h|1}}(\pi_{h|g,i}))'\right| &\leq C\zeta_{2n}^{s_2} \\ \left|\int \zeta_{2n}^{-1} K_2\left(\frac{\pi_{h|g,i} - \pi_{h|1,j}}{\zeta_{2n}}\right) f_{\pi_{h|1}}(\pi_{h|1,j}) \mathrm{d}\pi_{h|1,j} + f_{\pi_{h|1}}(\pi_{h|g,i})\right| &\leq C\zeta_{2n}^{s_2}. \end{aligned}$$

Therefore, from Assumption 9(v) we derive $\mathbb{E}[\partial h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)/\partial \delta_g^{v\prime}] = D_{g,h}^{\delta_g^v} + o(n^{-1/2})$, where

$$D_{g,h}^{\delta_g^v} = \mathbb{E}[(((\sigma_{h|1}^{-1}(\pi_{h|g,i}) f_{\pi_{h|1}}(\pi_{h|g,i}))' - f_{\pi_{h|1}}'(\pi_{h|g,i}) \tau_{ig}^v) \sigma_{h|g}'(\tau_{ig}^v) + f_{\pi_{h|1}}(\pi_{h|g,i})) z_i']. \tag{O.B.8}$$

Similarly we can derive $\mathbb{E}[\partial h_{gn}(\tau_i^v,\tau_j^v;\sigma_h)/\partial\delta_1^{v\prime}] = D_{g,h}^{\delta_1^v} + o(n^{-1/2})$, where

$$D_{g,h}^{\delta_1^v} = \mathbb{E}[(((\sigma_{h|g}^{-1}(\pi_{h|1,i})f_{\pi_{h|g}}(\pi_{h|1,i}))' - f'_{\pi_{h|g}}(\pi_{h|1,i})\tau_{i1}^v)\sigma'_{h|1}(\tau_{i1}^v) + f_{\pi_{h|g}}(\pi_{h|1,i}))z_i']. \tag{O.B.9}$$

Combining the results yields $T_{1n}^{\delta} = \frac{1}{n}\sum_{i=1}^n (D_{g,h}^{\delta_g^v}\psi^{\delta_g^v}(z_i) - D_{g,h}^{\delta_1^v}\psi^{\delta_1^v}(z_i)) + o_p(n^{-1/2})$, where $\psi^{\delta_1^v}(z_i)$ and $\psi^{\delta_g^v}(z_i)$ are the influence functions for $\hat\delta_1^v$ and $\hat\delta_g^v$, respectively. Define

$$\psi_{g,h,1}(z_i,g_i) = D_{g,h}^{\delta_g^v}\psi^{\delta_g^v}(z_i) - D_{g,h}^{\delta_1^v}\psi^{\delta_1^v}(z_i) + \psi_{g,h}^{\sigma}(\omega_i). \tag{O.B.10}$$

We thus have $T_{1n} = \frac{1}{n}\sum_{i=1}^n \psi_{g,h,1}(z_i,g_i) + o_p(n^{-1/2})$. □

Lemma O.B.2. The term T_{2n} in equation (A.8) satisfies $T_{2n} = n^{-1}\sum_{i=1}^n \psi_{g,h,2}(z_i) + o_p(n^{-1/2})$, where $\psi_{g,h,2}(z_i)$ is defined in equation (O.B.12).

Proof. Recall that T_{2n} is a V-statistic of degree 2 with an asymmetric kernel function and $\mathbb{E}[T_{2n}] = 0$. Similarly as in Lemma O.B.1, we use the idea of Hoeffding projection and derive an asymptotically linear representation of T_{2n}. Let $I = (i_1, i_2)$ with $i_1 \neq i_2$ and $\tau_I^v = (\tau_i^v : i \in I)$. We can write $T_{2n} = \frac{1}{n(n-1)}\sum_I (h_{gn}(\tau_I^v) - \mathbb{E}[h_{gn}(\tau_I^v)])$. Define $h_{gn}^*(\tau_i^v) = \frac{1}{n-1}\sum_{I:i\in I}(\mathbb{E}[h_{gn}(\tau_I^v)|\tau_i^v] - \mathbb{E}[h_{gn}(\tau_I^v)])$ and $T_{2n}^* = \frac{1}{n}\sum_{i=1}^n h_{gn}^*(\tau_i^v)$. Observe that $\mathbb{E}[T_{2n}^*] = 0$. Following Lemma O.B.1, we can show that $\mathrm{Cov}(T_{2n}, T_{2n}^*) = \mathrm{Var}(T_{2n}^*)$[36] and hence $\mathbb{E}[(T_{2n} - T_{2n}^*)^2] = \mathrm{Var}(T_{2n}) - \mathrm{Var}(T_{2n}^*)$. By Markov inequality, we obtain $T_{2n} = T_{2n}^* + o_p(n^{-1/2})$ if $\mathrm{Var}(Q_n) - \mathrm{Var}(Q_n^*) = o(n^{-1})$.

To show the last result, we express $\mathrm{Var}(T_{2n})$ and $\mathrm{Var}(T_{2n}^*)$ similarly as in equations (O.B.5) and (O.B.6). Follow the argument that compares the two equations and note that for any I and J, by Assumptions 7(ii)(iii), 10(i), and 11(i), we can bound both $\mathrm{Cov}(h_{gn}(\tau_I^v), h_{gn}(\tau_J^v))$ and $\mathrm{Cov}(\mathbb{E}[h_{gn}(\tau_I^v)|\tau_i^v], \mathbb{E}[h_{gn}(\tau_J^v)|\tau_i^v])$ uniformly by $O(\zeta_{2n}^{-2})$. Therefore, $\mathrm{Var}(Q_n) - \mathrm{Var}(Q_n^*) = (n(n-1))^{-2}\cdot O(n^2)\cdot O(\zeta_{2n}^{-2}) = O(n^{-2}\zeta_{2n}^{-2}) = o(n^{-1})$ by Assumption 9(iii)[37] and then $T_{2n} = T_{2n}^* + o_p(n^{-1/2})$ follows.

Next we calculate the influence function of T_{2n}^*. Conditional on τ_i^v, both τ_{ig}^v and $\pi_{h|g,i} = \sigma_{h|g}(\tau_{ig}^v)$ are constants. Under Assumption 2(i), we obtain

$$\begin{aligned}\mathbb{E}[h_{gn}(\tau_i^v,\tau_j^v;\sigma_h)|\tau_i^v] &= \int \frac{1}{\zeta_{2n}}K_2\Big(\frac{\pi_{h|g,i}-\pi_{h|1,j}}{\zeta_{2n}}\Big)(\tau_{ig}^v - \sigma_{h|1}^{-1}(\pi_{h|1,j}))f_{\pi_{h|1}}(\pi_{h|1,j})\mathrm{d}\pi_{h|1,j} \\ &= \int K_2(t)(\tau_{ig}^v - \sigma_{h|1}^{-1}(\pi_{h|g,i} - t\zeta_{2n}))f_{\pi_{h|1}}(\pi_{h|g,i} - t\zeta_{2n})\mathrm{d}t\end{aligned}$$

[36] This is because $\mathrm{Cov}(T_{2n}, T_{2n}^*) = \frac{1}{n}\sum_{i=1}^n \mathrm{Cov}(T_{2n}, h_{gn}^*(\tau_i^v))$ and for each i, $\mathrm{Cov}(T_{2n}, h_{gn}^*(\tau_i^v)) = \frac{1}{n(n-1)}\sum_{I:i\in I}\mathrm{Cov}(h_{gn}(\tau_I^v), h_{gn}^*(\tau_i^v)) = \frac{1}{n}\mathrm{Var}(h_{gn}^*(\tau_i^v))$.

[37] Assumption 9(iii) (i.e., $n^{1/2}\zeta_{1n}\zeta_{2n}^2 \to \infty$) implies $n^{1/2}\zeta_{2n} \to \infty$.

$$= \quad (\tau_{ig}^v - \sigma_{h|1}^{-1}(\pi_{h|g,i}))f_{\pi_{h|1}}(\pi_{h|g,i}) + \tilde{r}_{n,i}, \tag{O.B.11}$$

with $\max_i |\tilde{r}_{n,i}| \leq C\zeta_{2n}^{s_2}$. The second equality follows from the change of variables $t = \zeta_{2n}^{-1}(\pi_{h|g,i} - \pi_{h|1,j})$, and the third equality holds by Assumptions 7(i) and 8. Similarly, we can derive $|\mathbb{E}[h_{gn}(\tau_j^v, \tau_i^v; \sigma_h)|\tau_i^v] - (\sigma_{h|g}^{-1}(\pi_{h|1,i}) - \tau_{i1}^v)f_{\pi_{h|g}}(\pi_{h|1,i})| \leq C\zeta_{2n}^{s_2}$. By Assumption 9(v), $O(\zeta_{2n}^{s_2}) = o(n^{-1/2})$. All together, we obtain $T_{2n}^* = \frac{1}{n}\sum_{i=1}^n \psi_{g,h,2}(z_i) + o_p(n^{-1/2})$, where

$$\begin{aligned}\psi_{g,h,2}(z_i) =&(\tau_{ig}^v - \sigma_{h|1}^{-1}(\pi_{h|g,i}))f_{\pi_{h|1}}(\pi_{h|g,i}) + (\sigma_{h|g}^{-1}(\pi_{h|1,i}) - \tau_{i1}^v)f_{\pi_{h|g}}(\pi_{h|1,i}) \\ &- \mathbb{E}[(\tau_{ig}^v - \sigma_{h|1}^{-1}(\pi_{h|g,i}))f_{\pi_{h|1}}(\pi_{h|g,i}) + (\sigma_{h|g}^{-1}(\pi_{h|1,i}) - \tau_{i1}^v)f_{\pi_{h|g}}(\pi_{h|1,i})].\end{aligned} \tag{O.B.12}$$

□

Lemma O.B.3. *The third term T_{3n} in equation (A.8) satisfies $T_{3n} = o(n^{-1/2})$.*

Proof. Equation (O.B.11) implies

$$\mathbb{E}[h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)] = \int (\sigma_{h|g}^{-1}(\pi_{h|g,i}) - \sigma_{h|1}^{-1}(\pi_{h|g,i}))f_{\pi_{h|1}}(\pi_{h|g,i})f_{\pi_{h|g}}(\pi_{h|g,i})\mathrm{d}\pi_{h|g,i} + O(\zeta_{2n}^{s_2}).$$

From the identification result in Section 4, we can represent p_g as

$$\begin{aligned} p_g \quad &= \quad \mathbb{E}[1\{\pi_{h|g,i} = \pi_{h|1,j}\}(\tau_{ig}^v - \tau_{j1}^v)] \\ &= \quad \int [\sigma_{h|g}^{-1}(\pi_{h|g,i}) - \sigma_{h|1}^{-1}(\pi_{h|g,i})]f_{\pi_{h|g}}(\pi_{h|g,i})f_{\pi_{h|1}}(\pi_{h|g,i})\mathrm{d}\pi_{h|g,i}.\end{aligned}$$

Hence, $T_{3n} = \mathbb{E}[h_{gn}(\tau_i^v, \tau_j^v; \sigma_h)] - p_g = O(\zeta_{2n}^{s_2}) = o(n^{-1/2})$ by Assumption 9(v). □

O.C Lemmas in the Asymptotic Analysis of $\hat{\gamma}$

Notation Let $A = (a_{ij}) \in \mathbb{R}^{n^2}$ denote an $n \times n$ matrix and $x = (x_1, \dots, x_n)' \in \mathbb{R}^n$ denote an $n \times 1$ vector. Denote by $|||A|||_\infty = \max_{1\leq i\leq n}\sum_{j=1}^n |a_{ij}|$ the maximum row sum norm and $|||A|||_1 = \max_{1\leq j\leq n}\sum_{i=1}^n |a_{ij}|$ the maximum column sum norm. Note that $|||A'|||_\infty = |||A|||_1$ and $|||A'|||_1 = |||A|||_\infty$. Denote by $\|\cdot\|_\infty$ and $\|\cdot\|_1$ the l_∞ and l_1 norms. That is, $\|A\|_\infty = \max_{1\leq i,j\leq n}|a_{ij}|$, $\|A\|_1 = \sum_{i,j=1}^n |a_{ij}|$, $\|x\|_\infty = \max_{1\leq i\leq n}|x_i|$, and $\|x\|_1 = \sum_{i=1}^n |x_i|$. For $n \times n$ matrices A and B and $n \times 1$ vectors x and y, we can show $\|AB\|_\infty \leq |||A|||_\infty \|B\|_\infty$, $\|AB\|_1 \leq |||A|||_1 \|B\|_1$, $\|Ax\|_\infty \leq |||A|||_\infty \|x\|_\infty$, and

$\|Ax\|_1 \leq |||A|||_1 \|x\|_1$. Moreover, $|x'Ay| \leq |||A|||_\infty \|x\|_\infty \|y\|_1 \leq n\, |||A|||_\infty \|x\|_\infty \|y\|_\infty$.[38] Let $0 < C < \infty$ denote a universal constant. For simplicity, we write τ_i^s as τ_i throughout this section.

O.C.1 Consistency of $\hat{\gamma}$

Lemma O.C.1 (ULLN of the moment)**.** $\sup_{\gamma\in\Gamma} \|\hat{m}_n(\gamma, \hat{\mu}^Z) - m_0(\gamma, \mu_0^Z)\| = o_p(1)$.

Proof. Because $y_i - X_i'\gamma = y_i - X_i'\gamma_0 - X_i'(\gamma - \gamma_0) = \epsilon_i - X_i'(\gamma - \gamma_0)$, we have

$$\begin{aligned}
&\hat{m}_n(\gamma, \hat{\mu}^Z) - m_0(\gamma, \mu_0^Z) \\
= \;& \frac{1}{n}\sum_{i=1}^n (Z_i - \hat{\mu}^Z(\hat{\tau}_i))(y_i - X_i'\gamma) - \mathbb{E}[(Z_i - \mu_0^Z(\tau_i))(y_i - X_i'\gamma)] \\
= \;& -\frac{1}{n}\sum_{i=1}^n (\hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))\epsilon_i + \frac{1}{n}\sum_{i=1}^n (\hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))X_i'(\gamma - \gamma_0) \\
& + \frac{1}{n}\sum_{i=1}^n ((Z_i - \mu_0^Z(\tau_i))\epsilon_i - \mathbb{E}[(Z_i - \mu_0^Z(\tau_i))\epsilon_i]) \\
& - \frac{1}{n}\sum_{i=1}^n ((Z_i - \mu_0^Z(\tau_i))X_i' - \mathbb{E}[(Z_i - \mu_0^Z(\tau_i))X_i'])(\gamma - \gamma_0).
\end{aligned}$$

In the last expression, the first two average terms are $o_p(1)$ by Lemma O.C.2, and the last two average terms are $o_p(1)$ by Lemma O.C.3. By the compactness of Γ (Assumption 15(ii)) we have $\sup_{\gamma\in\Gamma} \|\hat{m}_n(\gamma, \hat{\mu}^Z) - m_0(\gamma, \mu_0^Z)\| = o_p(1)$. □

Lemma O.C.2. *For $t_i = (X_i', \epsilon_i)'$, we have*

$$\frac{1}{n}\sum_{i=1}^n (\hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))t_i' = o_p(1). \tag{O.C.1}$$

Proof. Recall that $X_i = (w_i\boldsymbol{y}, w_i\boldsymbol{x}, x_i')'$. Because $t_i \in \mathbb{R}^{2d_x+2}$ is finite dimensional, we can prove equation (O.C.1) for each component of t_i separately. For $r = 1, \ldots, 2d_x+2$, let t_{ir} denote the rth component of t_i, and $\boldsymbol{t}_r = (t_{1r}, \ldots, t_{nr})'$. By construction, we have $\hat{\mu}^Z(\hat{\tau}_i) = \hat{\beta}^Z(\hat{\boldsymbol{\tau}})'b^K(\hat{\tau}_i) \in \mathbb{R}^{d_Z}$ and $\hat{\mu}^Z(\tau_i) = \hat{\beta}^Z(\boldsymbol{\tau})'b^K(\tau_i) \in \mathbb{R}^{d_Z}$, where $\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) = (\hat{B}_K'\hat{B}_K)^{-1}\hat{B}_K'\boldsymbol{Z}$ and $\hat{\beta}^Z(\boldsymbol{\tau}) = (B_K'B_K)^{-1}B_K'\boldsymbol{Z}$, with $\hat{B}_K = B_K(\hat{\boldsymbol{\tau}})$ and $B_K = B_K(\boldsymbol{\tau})$. Denote $\boldsymbol{\mu}_0^Z = (\mu_0^Z(\tau_1), \ldots, \mu_0^Z(\tau_n))'$. The left-hand side of equation

[38] These results can be found in Horn and Johnson (1985, Section 5.6) or proved similarly.

(O.C.1) for component t_{ir} satisfies

$$\begin{aligned} \|\frac{1}{n}\sum_{i=1}^{n}(\hat{\mu}^Z(\hat{\tau}_i)-\mu_0^Z(\tau_i))t_{ir}\|^2 &= n^{-2}\boldsymbol{t}_r'(\hat{B}_K\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\boldsymbol{\mu}_0^Z)(\hat{B}_K\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\boldsymbol{\mu}_0^Z)'\boldsymbol{t}_r \\ &\leq n^{-2}\|\hat{B}_K\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\boldsymbol{\mu}_0^Z\|^2\boldsymbol{t}_r'\boldsymbol{t}_r, \end{aligned}$$

where the inequality follows because we can bound the largest eigenvalue of the matrix $(\hat{B}_K\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\boldsymbol{\mu}_0^Z)(\hat{B}_K\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\boldsymbol{\mu}_0^Z)'$ by $\|\hat{B}_K\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\boldsymbol{\mu}_0^Z\|^2$.

For t_{ir} that represents a component of $w_i\boldsymbol{x}$ or x_i, we have $\max_i|t_{ir}|<\infty$ (Assumptions 11(ii), 14(i)). Therefore, $n^{-1}\boldsymbol{t}_r'\boldsymbol{t}_r = n^{-1}\sum_{i=1}^{n}t_{ir}^2 \leq \max_i t_{ir}^2 < \infty$. For $t_{ir}=\epsilon_i$, because ϵ_i is i.i.d., by the law of large numbers and Assumption 16(i) $n^{-1}\boldsymbol{t}_r'\boldsymbol{t}_r = n^{-1}\sum_{i=1}^{n}\epsilon_i^2 = \mathbb{E}[\epsilon_i^2]+o_p(1)=O_p(1)$. For $t_{ir}=w_i\boldsymbol{y}$, we have $n^{-1}(\boldsymbol{wy})'\boldsymbol{wy}=O_p(1)$ by Lemma O.C.4. We conclude that $n^{-1}\boldsymbol{t}_r'\boldsymbol{t}_r=O_p(1)$.

By the triangle inequality and $(a+b+c)^2\leq 3(a^2+b^2+c^2)$,

$$\begin{aligned} & n^{-1}\|\hat{B}_K\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\boldsymbol{\mu}_0^Z\|^2 \\ \leq\ & n^{-1}(\|(\hat{B}_K-B_K)\hat{\beta}^Z(\hat{\boldsymbol{\tau}})\|+\|B_K(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)\|+\|B_K\beta^Z-\boldsymbol{\mu}_0^Z\|)^2 \\ \leq\ & 3n^{-1}(\|\hat{B}_K-B_K\|^2\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})\|^2+\|B_K(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)\|^2+\|B_K\beta^Z-\boldsymbol{\mu}_0^Z\|^2). \end{aligned}$$

It suffices to show that the last three terms are $o_p(1)$.

By equation (O.C.5), $n^{-1}\|\hat{B}_K-B_K\|^2=O_p(\varrho_1(K)^2/n)$. Moreover,

$$\begin{aligned} \|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})\|^2 &= \mathrm{tr}(\boldsymbol{Z}'\hat{B}_K(\hat{B}_K'\hat{B}_K)^{-2}\hat{B}'_K\boldsymbol{Z}) \\ &\leq O_p(n^{-1})\mathrm{tr}(\boldsymbol{Z}'\hat{B}_K(\hat{B}_K'\hat{B}_K)^{-1}\hat{B}'_K\boldsymbol{Z}) \\ &\leq O_p(n^{-1})\mathrm{tr}(\boldsymbol{Z}'\boldsymbol{Z})=O_p(1). \end{aligned} \tag{O.C.2}$$

The first inequality follows from Lemmas O.C.5 and O.C.6.[39] The second inequality follows because $\hat{B}_K(\hat{B}_K'\hat{B}_K)^{-1}\hat{B}'_K$ is idempotent and thus $\hat{B}_K(\hat{B}_K'\hat{B}_K)^{-1}\hat{B}'_K\leq I_K$. The last equality holds because $n^{-1}\mathrm{tr}(\boldsymbol{Z}'\boldsymbol{Z})=n^{-1}\sum_{i=1}^{n}\|Z_i\|^2<\infty$ by the boundedness of Z_i. We conclude that $n^{-1}\|\hat{B}_K-B_K\|^2\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})\|^2=o_p(1)$.

Observe that $n^{-1}\|B_K(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)\|^2=n^{-1}\|B_K(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)\|^2=n^{-1}\mathrm{tr}((\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)'B_K'B_K(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z))\leq O_p(1)\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z\|^2$, where the inequality holds because by Lemma O.C.5 $B_K'B_K/n\leq CI_K$ with probability approaching one. By the triangle

[39] By Lemmas O.C.5 and O.C.6, the smallest eigenvalue of $\hat{Q}_K=\hat{B}_K'\hat{B}_K/n$ converges to one in probability and hence $(\hat{B}_K'\hat{B}_K/n)^{-1}\leq CI_K$ with probability approaching one.

inequality, $\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) - \beta^Z\| \leq \|\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) - \hat{\beta}^Z(\boldsymbol{\tau})\| + \|\hat{\beta}^Z(\boldsymbol{\tau}) - \beta^Z\|$. Lemma O.C.7 shows that $\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) - \hat{\beta}^Z(\boldsymbol{\tau})\| = O_p(\varrho_1(K)/\sqrt{n})$. Moreover, by Lemma 15.3 in Li and Racine (2007) for x_i and z_i in Z_i and Lemma O.C.8 for $w_i\boldsymbol{x}$ in Z_i, we have $\|\hat{\beta}^Z(\boldsymbol{\tau}) - \beta^Z\| = o_p(1)$. Combining these results yields $n^{-1}\|B_K(\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) - \beta^Z)\|^2 = o_p(1)$.

Finally, $n^{-1}\|B_K\beta^Z - \boldsymbol{\mu}_0^Z\|^2 = n^{-1}\sum_{i=1}^n \|\beta^{Z\prime}b^K(\tau_i) - \mu_0^Z(\tau_i)\|^2 \leq \sup_\tau \|\beta^{Z\prime}b^K(\tau) - \mu_0^Z(\tau)\|^2 = O(K^{-2a})$ by Assumption 13(ii). □

Lemma O.C.3. *For $t_i = (X_i', \epsilon_i)'$, we have*

$$\frac{1}{n}\sum_{i=1}^n((Z_i - \mu_0^Z(\tau_i))t_i' - \mathbb{E}[(Z_i - \mu_0^Z(\tau_i))t_i']) = o_p(1). \tag{O.C.3}$$

Proof. Recall that $Z_i = (w_i\boldsymbol{x}, x_i', z_i')'$ and $X_i = (w_i\boldsymbol{y}, w_i\boldsymbol{x}, x_i')'$. Because both Z_i and t_i are finite dimensional, we can prove equation (O.C.3) component by component. For simplicity, we assume that both z_i and x_i are scalars. Depending on which components of Z_i and t_i under consideration, we divide the proof into six cases (Table O.C.1).

Table O.C.1: The Six Cases in Lemma O.C.3

		Component of t_i		
		x_i, ϵ_i	$w_i\boldsymbol{x}$	$w_i\boldsymbol{y}$
Component of Z_i	x_i, z_i	Case (a)	Case (c)	Case (e)
	$w_i\boldsymbol{x}$	Case (b)	Case (d)	Case (f)

Case (a): Because (x_i, z_i, ϵ_i) is i.i.d., the result follows by the law of large numbers.

Case (b): Take x_i in t_i as an example and ϵ_i in t_i can be proved similarly.

$$\begin{aligned} & n^{-1}\sum_{i=1}^n((w_i\boldsymbol{x} - \mu_0^{w_i\boldsymbol{x}}(\tau_i))x_i - \mathbb{E}[(w_i\boldsymbol{x} - \mu_0^{w_i\boldsymbol{x}}(\tau_i))x_i]) \\ = \; & n^{-1}\sum_{i=1}^n(w_i\boldsymbol{x}x_i - \mathbb{E}[w_i\boldsymbol{x}x_i]) - n^{-1}\sum_{i=1}^n(\mu_0^{w_i\boldsymbol{x}}(\tau_i)x_i - \mathbb{E}[\mu_0^{w_i\boldsymbol{x}}(\tau_i)x_i]). \end{aligned}$$

Because $\mu_0^{w_i\boldsymbol{x}}(\tau_i)x_i$ is independent across i, the second term on the right-hand side is $o_p(1)$ by the law of large numbers. Write the first term on the right-hand side as $n^{-1}(\boldsymbol{x}'\boldsymbol{w}\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}'\boldsymbol{w}\boldsymbol{x}])$. Applying Lemma O.C.9 with $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{x}$ and $\boldsymbol{q} = \boldsymbol{w}$ yields $n^{-1}(\boldsymbol{x}'\boldsymbol{w}\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}'\boldsymbol{w}\boldsymbol{x}]) = o_p(1)$. Equation (O.C.3) thus holds for case (b).

Case (c): Without loss of generality we take z_i in Z_i as an example. Denote $\tilde{z}_i = z_i - \mu_0^z(\tau_i)$ and $\tilde{\boldsymbol{z}} = (\tilde{z}_1, \ldots, \tilde{z}_n)'$. Write $n^{-1}\sum_{i=1}^n((z_i - \mu_0^z(\tau_i))w_i\boldsymbol{x} - \mathbb{E}[(z_i -$

$\mu_0^z(\tau_i))w_i\boldsymbol{x}]) = n^{-1}(\tilde{\boldsymbol{z}}'\boldsymbol{w}\boldsymbol{x} - \mathbb{E}[\tilde{\boldsymbol{z}}'\boldsymbol{w}\boldsymbol{x}])$. Applying Lemma O.C.9 with $\boldsymbol{a} = \tilde{\boldsymbol{z}}$, $\boldsymbol{b} = \boldsymbol{x}$, and $\boldsymbol{q} = \boldsymbol{w}$, we prove equation (O.C.3) for case (c).

Case (d): Write

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}((w_i\boldsymbol{x} - \mu_0^{w_i\boldsymbol{x}}(\tau_i))w_i\boldsymbol{x} - \mathbb{E}[(w_i\boldsymbol{x} - \mu_0^{w_i\boldsymbol{x}}(\tau_i))w_i\boldsymbol{x}]) \\
= \; & n^{-1}\sum_{i=1}^{n}(w_i\boldsymbol{x}w_i\boldsymbol{x} - \mathbb{E}[w_i\boldsymbol{x}w_i\boldsymbol{x}]) - n^{-1}\sum_{i=1}^{n}(\mu_0^{w_i\boldsymbol{x}}(\tau_i)w_i\boldsymbol{x} - \mathbb{E}[\mu_0^{w_i\boldsymbol{x}}(\tau_i)w_i\boldsymbol{x}]) \\
= \; & n^{-1}(\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{w}\boldsymbol{x} - \mathbb{E}[\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{w}\boldsymbol{x}]) - n^{-1}(\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{w}\boldsymbol{x} - \mathbb{E}[\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{w}\boldsymbol{x}]),
\end{aligned}
$$

where $\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau}) = (\mu_0^{w_1\boldsymbol{x}}(\tau_1), \ldots, \mu_0^{w_n\boldsymbol{x}}(\tau_n))'$. Applying Lemma O.C.9 twice, one with $\boldsymbol{a} = \boldsymbol{b} = \boldsymbol{x}$ and $\boldsymbol{q} = \boldsymbol{w}'\boldsymbol{w}$, and the other with $\boldsymbol{a} = \boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})$, $\boldsymbol{b} = \boldsymbol{x}$ and $\boldsymbol{q} = \boldsymbol{w}$, we obtain that the last two terms are both $o_p(1)$. Equation (O.C.3) holds for case (d).

Case (e): Recall that $\boldsymbol{y} = \boldsymbol{s}(\boldsymbol{w}\boldsymbol{x}\gamma_2 + \boldsymbol{x}\gamma_3 + \boldsymbol{\epsilon})$, where $\boldsymbol{s} = (I_n - \gamma_1\boldsymbol{w})^{-1}$, and $\boldsymbol{\lambda} = \boldsymbol{\lambda}(\boldsymbol{\tau})$ is short for $\boldsymbol{\lambda}^s(\boldsymbol{\tau}^s)$. Take z_i in Z_i as an example. Write

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}((z_i - \mu_0^z(\tau_i))w_i\boldsymbol{y} - \mathbb{E}[(z_i - \mu_0^z(\tau_i))w_i\boldsymbol{y}]) \\
= \; & n^{-1}(\tilde{\boldsymbol{z}}'\boldsymbol{wswx} - \mathbb{E}[\tilde{\boldsymbol{z}}'\boldsymbol{wswx}])\gamma_2 + n^{-1}(\tilde{\boldsymbol{z}}'\boldsymbol{wsx} - \mathbb{E}[\tilde{\boldsymbol{z}}'\boldsymbol{wsx}])\gamma_3 \\
& + n^{-1}(\tilde{\boldsymbol{z}}'\boldsymbol{ws\epsilon} - \mathbb{E}[\tilde{\boldsymbol{z}}'\boldsymbol{ws\epsilon}]).
\end{aligned}
$$

Applying Lemma O.C.9 to each term in the last line proves equation (O.C.3) for case (e).

Case (f): Write

$$
\begin{aligned}
& n^{-1}\sum_{i=1}^{n}((w_i\boldsymbol{x} - \mu_0^{w_i\boldsymbol{x}}(\tau_i))w_i\boldsymbol{y} - \mathbb{E}[(w_i\boldsymbol{x} - \mu_0^{w_i\boldsymbol{x}}(\tau_i))w_i\boldsymbol{y}]) \\
= \; & n^{-1}\sum_{i=1}^{n}(w_i\boldsymbol{x}w_i\boldsymbol{y} - \mathbb{E}[w_i\boldsymbol{x}w_i\boldsymbol{y}]) - n^{-1}\sum_{i=1}^{n}(\mu_0^{w_i\boldsymbol{x}}(\tau_i)w_i\boldsymbol{y} - \mathbb{E}[\mu_0^{w_i\boldsymbol{x}}(\tau_i)w_i\boldsymbol{y}]) \\
= \; & n^{-1}(\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{wswx} - \mathbb{E}[\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{wswx}])\gamma_2 + n^{-1}(\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{wsx} - \mathbb{E}[\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{wsx}])\gamma_3 \\
& + n^{-1}(\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{ws\epsilon} - \mathbb{E}[\boldsymbol{x}'\boldsymbol{w}'\boldsymbol{ws\epsilon}]) - n^{-1}(\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{wswx} - \mathbb{E}[\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{wswx}])\gamma_2 \\
& - n^{-1}(\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{wsx} - \mathbb{E}[\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{wsx}])\gamma_3 - n^{-1}(\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{ws\epsilon} - \mathbb{E}[\boldsymbol{\mu}_0^{\boldsymbol{wx}}(\boldsymbol{\tau})'\boldsymbol{ws\epsilon}]).
\end{aligned}
$$

Applying Lemma O.C.9 to each term in the last line proves equation (O.C.3). □

Lemma O.C.4 (Boundness of $\boldsymbol{wy}$)**.** $n^{-1}(\boldsymbol{wy})'\boldsymbol{wy} = O_p(1)$.

Proof. Let $T = \boldsymbol{w}^2\boldsymbol{x}\gamma_2 + \boldsymbol{w}\boldsymbol{x}\gamma_3$ be an $n \times 1$ vector, and recall that $\boldsymbol{w}\boldsymbol{y} = \boldsymbol{s}(\boldsymbol{w}^2\boldsymbol{x}\gamma_2 + \boldsymbol{w}\boldsymbol{x}\gamma_3 + \boldsymbol{w}\boldsymbol{\epsilon}) = \boldsymbol{s}(T + \boldsymbol{w}\boldsymbol{\epsilon})$. We can write

$$\begin{aligned} n^{-1}(\boldsymbol{w}\boldsymbol{y})'\boldsymbol{w}\boldsymbol{y} &= n^{-1}(T + \boldsymbol{w}\boldsymbol{\epsilon})'\boldsymbol{s}'\boldsymbol{s}(T + \boldsymbol{w}\boldsymbol{\epsilon}) \\ &= n^{-1}T'\boldsymbol{s}'\boldsymbol{s}T + 2n^{-1}T'\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}\boldsymbol{\epsilon} + n^{-1}\boldsymbol{\epsilon}'\boldsymbol{w}'\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}\boldsymbol{\epsilon}. \end{aligned} \quad \text{(O.C.4)}$$

We show that $\boldsymbol{w}$ and $\boldsymbol{s}$ are uniformly bounded in both row and column sums.[40] In fact, under Assumption 14(i), we have $|||\boldsymbol{w}|||_\infty = 1$ and thus $|||\boldsymbol{s}|||_\infty \leq \sum_{r=0}^{\infty} |\gamma_1|^r\ |||\boldsymbol{w}|||_\infty^r < \infty$. For any $r \geq 1$, we can bound $\|\boldsymbol{w}^r\|_\infty \leq |||\boldsymbol{w}\ |||_\infty^{r-1}\ \|\boldsymbol{w}\|_\infty = \|\boldsymbol{w}\|_\infty$. Therefore, $|||\boldsymbol{w}^r|||_1 = \max_{j\in\mathcal{N}} \sum_{i=1}^{n} |(\boldsymbol{w}^r)_{ij}| \leq n\|\boldsymbol{w}^r\|_\infty \leq n\|\boldsymbol{w}\|_\infty$ and thus $|||\boldsymbol{s}|||_1 \leq \sum_{r=0}^{\infty} |\gamma_1|^r\ |||\ \boldsymbol{w}^r|||_1 \leq \sum_{r=0}^{\infty} |\gamma_1|^r n\|\boldsymbol{w}\|_\infty = O_p(1)$.

By the boundedness of x_i and γ, we can bound $\|T\|_\infty \leq |||\boldsymbol{w}\ |||_\infty^2\ \|\boldsymbol{x}\gamma_2\|_\infty + |||\boldsymbol{w}\ |||_\infty\ \|\boldsymbol{x}\gamma_3\|_\infty < \infty$. Therefore, the first term in the last line of (O.C.4) is $n^{-1}T'\boldsymbol{s}'\boldsymbol{s}T \leq |||\boldsymbol{s}'\boldsymbol{s}\ |||_\infty\ \|T\|_\infty^2 \leq |||\boldsymbol{s}\ |||_1\ |||\boldsymbol{s}\ |||_\infty\ \|T\|_\infty^2 = O_p(1)$. The second to last term in equation (O.C.4) satisfies $n^{-1}|T'\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}\boldsymbol{\epsilon}| \leq |||\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}\ |||_\infty\ \|T\|_\infty \|\boldsymbol{\epsilon}/n\|_1 = O_p(1)$, because $\|\boldsymbol{\epsilon}/n\|_1 = n^{-1}\sum|\epsilon_i| = \mathbb{E}[|\epsilon_i|] + o_p(1) = O_p(1)$ by the law of large numbers and Assumption 16(i) and $|||\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}|||_\infty \leq |||\boldsymbol{s}\ |||_1\ |||\boldsymbol{s}\ |||_\infty\ |||\boldsymbol{w}|||_\infty = O_p(1)$. Finally, the last term in (O.C.4) satisfies $n^{-1}\boldsymbol{\epsilon}'\boldsymbol{w}'\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}\boldsymbol{\epsilon} \leq n^{-1}\lambda_{\max}(\boldsymbol{w}'\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w})\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = O_p(1)$, because $n^{-1}\boldsymbol{\epsilon}'\boldsymbol{\epsilon} = n^{-1}\sum_{i=1}^{n}\epsilon_i^2 = \mathbb{E}[\epsilon_i^2] + o_p(1) = O_p(1)$ and $\lambda_{\max}(\boldsymbol{w}'\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}) \leq |||\boldsymbol{w}'\boldsymbol{s}'\boldsymbol{s}\boldsymbol{w}|||_\infty \leq |||\boldsymbol{w}\ |||_1\ |||\boldsymbol{s}\ |||_1\ |||\boldsymbol{s}\ |||_\infty\ |||\boldsymbol{w}|||_\infty = O_p(1)$. Combining the three terms, we complete the proof. □

Lemma O.C.5. Let $Q_K = B_K'B_K/n$. Then $\|Q_K - I_K\| = O_p(\varrho_0(K)\sqrt{K/n})$.

Proof. The result follows from Lemma 15.2 in Li and Racine (2007, p.481). □

Lemma O.C.6. *Let $\hat{Q}_K = \hat{B}_K'\hat{B}_K/n$. Then $\|\hat{Q}_K - Q_K\| = O_p(\varrho_1(K)/\sqrt{n})$.*

Proof. Because $\hat{B}_K'\hat{B}_K - B_K'B_K = (\hat{B}_K - B_K)^2 + B_K'(\hat{B}_K - B_K) + (\hat{B}_K - B_K)'B_K$, we have $\|\hat{Q}_K - Q_K\| = \|\hat{B}_K'\hat{B}_K - B_K'B_K\|/n \leq \|\hat{B}_K - B_K\|^2/n + 2\|(\hat{B}_K - B_K)'B_K\|/n$. The $\sqrt{n}$-consistency of $\hat{\theta}$ and boundedness of z (Assumptions 11(i) and 12(ii)) imply that $\max_i \|\hat{\tau}_i - \tau_i\| = O_p(n^{-1/2})$. Therefore,

$$\|\hat{B}_K - B_K\| = (\sum_{i=1}^{n} \|b^K(\hat{\tau}_i) - b^K(\tau_i)\|^2)^{1/2} \leq n^{1/2}\varrho_1(K)\max_i \|\hat{\tau}_i - \tau_i\| = O_p(\varrho_1(K)), \quad \text{(O.C.5)}$$

[40] See Lee (2002, Lemma 1) for similar results.

by the mean-value theorem and Assumption 13(iv). Moreover,

$$\begin{aligned}\|(\hat{B}_K - B_K)' B_K\|/n &= \operatorname{tr}((\hat{B}_K - B_K)' B_K B_K' (\hat{B}_K - B_K))^{1/2}/n \\ &\leq O_p(1)\operatorname{tr}((\hat{B}_K - B_K)' B_K (B_K' B_K)^{-1} B_K' (\hat{B}_K - B_K))^{1/2}/\sqrt{n} \\ &\leq O_p(1)\|\hat{B}_K - B_K\|/\sqrt{n} = O_p(\varrho_1(K)/\sqrt{n}).\end{aligned}$$

The first inequality above holds because by Lemma O.C.5 $I_K \leq C(B_K' B_K/n)^{-1}$ with probability approaching one. The second inequality follows by $B_K(B_K'B_K)^{-1}B_K'$ idempotent. The last equality follows from equation (O.C.5). We conclude that $\|\hat{Q}_K - Q_K\| \leq O_p(\varrho_1(K)^2/n) + O_p(\varrho_1(K)/\sqrt{n}) = O_p(\varrho_1(K)/\sqrt{n})$. □

Lemma O.C.7. $\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) - \hat{\beta}^Z(\boldsymbol{\tau})\| = O_p(\varrho_1(K)/\sqrt{n})$.

Proof. Recall that $\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) = \hat{Q}_K^{-1}\hat{B}_K'\boldsymbol{Z}/n$ and $\hat{\beta}^Z(\boldsymbol{\tau}) = Q_K^{-1}B'_K\boldsymbol{Z}/n$. We have

$$\begin{aligned}\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}}) - \hat{\beta}^Z(\boldsymbol{\tau})\| &= \operatorname{tr}(\boldsymbol{Z}'(\hat{Q}_K^{-1}\hat{B}_K' - Q_K^{-1}B'_K)'(\hat{Q}_K^{-1}\hat{B}_K' - Q_K^{-1}B'_K)\boldsymbol{Z}/n^2)^{1/2} \\ &\leq \|(\hat{Q}_K^{-1}\hat{B}_K' - Q_K^{-1}B'_K)/\sqrt{n}\|\operatorname{tr}(\boldsymbol{Z}'\boldsymbol{Z}/n)^{1/2}.\end{aligned}$$

By the boundedness of Z_i, $\operatorname{tr}(\boldsymbol{Z}'\boldsymbol{Z}/n) = n^{-1}\sum_{i=1}^n \|Z_i\|^2 < \infty$. Moreover, $\|(\hat{Q}_K^{-1}\hat{B}_K' - Q_K^{-1}B'_K)/\sqrt{n}\| \leq \|(\hat{Q}_K^{-1} - Q_K^{-1})\hat{B}_K'/\sqrt{n}\| + \|Q_K^{-1}(\hat{B}_K - B_K)'/\sqrt{n}\|$. Observe

$$\begin{aligned}\|(\hat{Q}_K^{-1} - Q_K^{-1})\hat{B}_K'/\sqrt{n}\| &= \operatorname{tr}((\hat{Q}_K^{-1} - Q_K^{-1})\hat{B}_K'\hat{B}_K(\hat{Q}_K^{-1} - Q_K^{-1})/n)^{1/2} \\ &= \operatorname{tr}(Q_K^{-1}(Q_K - \hat{Q}_K)\hat{Q}_K^{-1}(Q_K - \hat{Q}_K)Q_K^{-1})^{1/2} \\ &\leq O_p(1)\operatorname{tr}((Q_K - \hat{Q}_K)Q_K^{-2}(Q_K - \hat{Q}_K))^{1/2} \\ &\leq O_p(1)\|Q_K - \hat{Q}_K\| = O_p(\varrho_1(K)/\sqrt{n}),\end{aligned}$$

where the inequalities follow from Lemmas O.C.5 and O.C.6.[41] The last equality follows from Lemma O.C.6. As for the second term, we have

$$\begin{aligned}\|Q_K^{-1}(\hat{B}_K - B_K)'/\sqrt{n}\| &= \operatorname{tr}((\hat{B}_K - B_K)Q_K^{-2}(\hat{B}_K - B_K)'/n)^{1/2} \\ &\leq O_p(1)\|(\hat{B}_K - B_K)/\sqrt{n}\| = O_p(\varrho_1(K)/\sqrt{n}).\end{aligned}$$

where the last equality holds by equation (O.C.5). □

[41] By Lemmas O.C.5 and O.C.6, the smallest eigenvalue of $\hat{Q}_K$ converges to one in probability and hence the largest eigenvalue of $\hat{Q}_K^{-1}$ is bounded with probability approaching one. Similarly, by Lemma O.C.5, the largest eigenvalue of Q_K^{-2} is bounded with probability approaching one.

Lemma O.C.8. $\|\hat{\beta}^{\boldsymbol{wx}}(\boldsymbol{\tau}) - \beta^{\boldsymbol{wx}}\| = o_p(1)$.

Proof. Recall that $\hat{\beta}^{\boldsymbol{wx}}(\boldsymbol{\tau}) = Q_K^{-1} B_K' \boldsymbol{wx}/n$. We can write $\hat{\beta}^{\boldsymbol{wx}}(\boldsymbol{\tau}) - \beta^{\boldsymbol{wx}} = Q_K^{-1} B_K'(\boldsymbol{wx} - B_K \beta^{\boldsymbol{wx}})/n$. Observe that

$$\begin{aligned} \|\hat{\beta}^{\boldsymbol{wx}}(\boldsymbol{\tau}) - \beta^{\boldsymbol{wx}}\| &= \operatorname{tr}((\boldsymbol{wx} - B_K \beta^{\boldsymbol{wx}})' B_K Q_K^{-2} B_K' (\boldsymbol{wx} - B_K \beta^{\boldsymbol{wx}})/n^2)^{1/2} \\ &\leq O_p(1) \|B_K'(\boldsymbol{wx} - B_K \beta^{\boldsymbol{wx}})/n\|, \end{aligned}$$

where we used that the largest eigenvalue of Q_K^{-2} is bounded with probability approaching one. Write $\boldsymbol{wx} - B_K \beta^{\boldsymbol{wx}} = (\boldsymbol{wx} - \boldsymbol{\mu}_0^{\boldsymbol{wx}}) + (\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})$. We derive

$$\begin{aligned} & \|B_K'(\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})/n\| \\ = \; & \operatorname{tr}((\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})' B_K B_K' (\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})/n^2)^{1/2} \\ \leq \; & O_p(1) \operatorname{tr}((\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})' B_K (B_K' B_K)^{-1} B_K' (\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})/n)^{1/2} \\ \leq \; & O_p(1) \|(\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})/\sqrt{n}\| = O_p(K^{-a}), \end{aligned}$$

where the first inequality holds by Lemma O.C.5,[42] the second inequality follows by $B_K(B_K' B_K)^{-1} B_K'$ idempotent, and the last equality follows by Assumption 13(ii).[43] If we can show $\|B_K'(\boldsymbol{wx} - \boldsymbol{\mu}_0^{\boldsymbol{wx}})/n\| = o_p(1)$, then combining the results completes the proof. Because x_i is finite dimensional, we can prove the equation for each component of x_i separately. For simplicity, assume that x_i is a scalar.

Write $B_K'(\boldsymbol{wx} - \boldsymbol{\mu}_0^{\boldsymbol{wx}})/n = n^{-1} \sum_i \sum_j b^K(\tau_i)(w_{ij} x_j - \mathbb{E}[w_{ij} x_j | \tau_i]) = n^{-1} \sum_i \sum_j r_{ij}$, where $r_{ij} = b^K(\tau_i)(w_{ij} x_j - \mathbb{E}[w_{ij} x_j | \tau_i])$. Because $\mathbb{E}[r_{ij} | \tau_i] = 0$, we have $\mathbb{E}[r_{ij}] = 0$. By construction,

$$\begin{aligned} \mathbb{E}\|B_K'(\boldsymbol{wx} - \boldsymbol{\mu}_0^{\boldsymbol{wx}})/n\|^2 &= n^{-2} \sum_{(i,j)} \sum_{(k,l):\{i,j\} \cap \{k,l\} \neq \emptyset} \mathbb{E}[r_{ij}' r_{kl}] \\ &\quad + n^{-2} \sum_{(i,j)} \sum_{(k,l):\{i,j\} \cap \{k,l\} = \emptyset} \mathbb{E}[r_{ij}' r_{kl}]. \qquad \text{(O.C.6)} \end{aligned}$$

For any $i, j, k, l \in \mathcal{N}$, we have $|\mathbb{E}[r_{ij}' r_{kl}]| \leq \mathbb{E}|b^K(\tau_i)' b^K(\tau_k)(w_{ij} x_j - \mathbb{E}[w_{ij} x_j | \tau_i])(w_{kl} x_l - \mathbb{E}[w_{kl} x_l | \tau_k])| \leq O(n^{-2})(\mathbb{E}[(b^K(\tau_i)' b^K(\tau_k))^2])^{1/2} = O(n^{-2}\sqrt{K})$. The second inequality

[42] By Lemma O.C.5, the largest eigenvalue of $Q_K = B_K' B_K/n$ converges to one in probability and hence $CI_K \leq (B_K' B_K/n)^{-1}$ with probability approaching one.

[43] Under Assumption 13(ii), we have $\|(\boldsymbol{\mu}_0^{\boldsymbol{wx}} - B_K \beta^{\boldsymbol{wx}})/\sqrt{n}\| = (n^{-1} \sum_{i=1}^n \|\mu_0^{\boldsymbol{wx}}(\tau_i) - \beta^{\boldsymbol{wx}\prime} b^K(\tau_i)\|^2)^{1/2} \leq \sup_\tau \|\mu_0^{\boldsymbol{wx}}(\tau) - \beta^{\boldsymbol{wx}\prime} b^K(\tau)\| = O(K^{-a})$.

follows from Assumptions 11(ii) and 14(ii),[44] and the last equality follows because by Assumptions 2(i) and 13(i), $\mathbb{E}[(b^K(\tau_i)'b^K(\tau_k))^2] = \mathbb{E}[b^K(\tau_i)'b^K(\tau_k)b^K(\tau_k)'b^K(\tau_i)] = \mathbb{E}[\text{tr}(b^K(\tau_i)b^K(\tau_i)'b^K(\tau_k)b^K(\tau_k)')] = \text{tr}(\mathbb{E}[b^K(\tau_i)b^K(\tau_i)']\mathbb{E}[b^K(\tau_k)b^K(\tau_k)']) = \text{tr}(I_K) = K$. The sum over overlapping $\{i,j\}$ and $\{k,l\}$ contains $O(n^3)$ terms. Therefore, the first term in equation (O.C.6) is $n^{-2}\cdot O(n^3)\cdot O(n^{-2}\sqrt{K}) = O(\sqrt{K}/n)$.

Further, for disjoint $\{i,j\}$ and $\{k,l\}$, we have

$$\begin{aligned}
\mathbb{E}[r'_{ij}r_{kl}|\boldsymbol{\varsigma}] &= b^K(\tau_i)'b^K(\tau_k)\mathbb{E}[(w_{ij}x_j - \mathbb{E}[w_{ij}x_j|\tau_i])(w_{kl}x_l - \mathbb{E}[w_{kl}x_l|\tau_k])|\boldsymbol{\varsigma}] \\
&= b^K(\tau_i)'b^K(\tau_k)(\text{Cov}(w_{ij}, w_{kl}|\boldsymbol{\varsigma})x_jx_l \\
&\quad +(\mathbb{E}[w_{ij}|\boldsymbol{\varsigma}]x_j - \mathbb{E}[w_{ij}x_j|\tau_i])(\mathbb{E}[w_{kl}|\boldsymbol{\varsigma}]x_l - \mathbb{E}[w_{kl}x_l|\tau_k])) \\
&= b^K(\tau_i)'b^K(\tau_k)(\mathbb{E}[w_{ij}|\varsigma_i,\varsigma_j]x_j - \mathbb{E}[w_{ij}x_j|\tau_i])(\mathbb{E}[w_{kl}|\varsigma_k,\varsigma_l]x_l - \mathbb{E}[w_{kl}x_l|\tau_k]) \\
&\quad +b^K(\tau_i)'b^K(\tau_k)e_{ij,kl},
\end{aligned}$$

where $e_{ij,kl} = \text{Cov}(w_{ij}, w_{kl}|\boldsymbol{\varsigma})x_jx_l + (\mathbb{E}[w_{ij}|\boldsymbol{\varsigma}] - \mathbb{E}[w_{ij}|\varsigma_i,\varsigma_j])x_j(\mathbb{E}[w_{kl}|\boldsymbol{\varsigma}]x_l - \mathbb{E}[w_{kl}x_l|\tau_k]) + (\mathbb{E}[w_{kl}|\boldsymbol{\varsigma}] - \mathbb{E}[w_{kl}|\varsigma_k,\varsigma_l])x_l(\mathbb{E}[w_{ij}|\varsigma_i,\varsigma_j]x_j - \mathbb{E}[w_{ij}x_j|\tau_i])$. By Assumptions 11(ii), 14(ii)(v), and Cauchy-Schwarz inequality, we derive $\mathbb{E}[(e_{ij,kl})^2] \leq o(n^{-4}/K)$. Observe that the terms $b^K(\tau_i)(\mathbb{E}[w_{ij}|\varsigma_i,\varsigma_j]x_j - \mathbb{E}[w_{ij}x_j|\tau_i])$ and $b^K(\tau_k)(\mathbb{E}[w_{kl}|\varsigma_k,\varsigma_l]x_l - \mathbb{E}[w_{kl}x_l|\tau_k])$ are independent, both with mean zero. Therefore, we have the uniform bound $|\mathbb{E}[r'_{ij}r_{kl}]| \leq (\mathbb{E}[(b^K(\tau_i)'b^K(\tau_k))^2])^{1/2}(\mathbb{E}[(e_{ij,kl})^2])^{1/2} = \sqrt{K}\cdot o(n^{-2}/\sqrt{K}) = o(n^{-2})$. The sum over disjoint $\{i,j\}$ and $\{k,l\}$ contains $O(n^4)$ terms. Hence, the second term in (O.C.6) can be bounded by $n^{-2}\cdot O(n^4)\cdot o(n^{-2}) = o(1)$. Combining the results we prove $\mathbb{E}\|B'_K(\boldsymbol{wx} - \boldsymbol{\mu}_0^{\boldsymbol{wx}})/n\|^2 = o(1)$ and thus $\|B'_K(\boldsymbol{wx} - \boldsymbol{\mu}_0^{\boldsymbol{wx}})/n\| = o_p(1)$. □

Lemma O.C.9. *Suppose that $\boldsymbol{a} = (a_1,\ldots,a_n)'$ and $\boldsymbol{b} = (b_1,\ldots,b_n)'$ are $n\times 1$ vectors in $\mathbb{R}^n$ such that (i) (a_i,b_i) is independent across i; (ii) $\boldsymbol{a}$ is a function of $\boldsymbol{\varsigma}$ with $\max_{i\in\mathcal{N}}|a_i| < \infty$; (iii) $\boldsymbol{b}$ is either a function of $\boldsymbol{\varsigma}$ with $\max_{i\in\mathcal{N}}|b_i| < \infty$ or independent of $\boldsymbol{w}$ conditional on $\boldsymbol{\varsigma}$ with $\max_{i\in\mathcal{N}}|\mathbb{E}[b_i|\varsigma_i]| < \infty$. Let $\boldsymbol{q}$ be a matrix that takes the form of (a) $\boldsymbol{w}$, (b) $\boldsymbol{w}'\boldsymbol{w}$, (c) $\boldsymbol{wsw}^t$, $t = 0,1$, or (d) $\boldsymbol{w}'\boldsymbol{wsw}^t$, $t = 0,1$, where $\boldsymbol{s} = (I_n - \gamma_1\boldsymbol{w})^{-1}$. Then $n^{-1}(\boldsymbol{a}'\boldsymbol{qb} - \mathbb{E}[\boldsymbol{a}'\boldsymbol{qb}]) = o_p(1)$.*

Proof. By Markov's inequality, it suffices to show that the second moment of $n^{-1}(\boldsymbol{a}'\boldsymbol{qb}-$

[44] By Cauchy-Schwarz inequality, $(a+b)^4 \leq 8(a^4+b^4)$, Jensen's inequality, and iterated expectations, we have $(\mathbb{E}[(w_{ij}x_j - \mathbb{E}[w_{ij}x_j|\tau_i])^2(w_{kl}x_l - \mathbb{E}[w_{kl}x_l|\tau_k])^2])^{1/2} \leq (\mathbb{E}[(w_{ij}x_j - \mathbb{E}[w_{ij}x_j|\tau_i])^4])^{1/4}(\mathbb{E}[(w_{kl}x_l - \mathbb{E}[w_{kl}x_l|\tau_k])^4])^{1/4} \leq 4(\mathbb{E}[(w_{ij}x_j)^4])^{1/4}(\mathbb{E}[(w_{kl}x_l)^4])^{1/4} \leq C\mathbb{E}[\|\boldsymbol{w}\|_\infty^4]^{1/2} = O(n^{-2})$.

$\mathbb{E}[\boldsymbol{a}'\boldsymbol{q}\boldsymbol{b}]$) is $o(1)$. The second moment is

$$\begin{aligned} n^{-2}\mathbb{E}(\boldsymbol{a}'\boldsymbol{q}\boldsymbol{b} - \mathbb{E}[\boldsymbol{a}'\boldsymbol{q}\boldsymbol{b}])^2 &= n^{-2}\mathrm{Cov}\left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j q_{ij}, \sum_{i=1}^{n}\sum_{j=1}^{n} a_i b_j q_{ij}\right) \\ &= n^{-2}\sum_{(i,j,k,l):\{i,j\}\cap\{k,l\}\neq\emptyset} \mathrm{Cov}(a_i b_j q_{ij}, a_k b_l q_{kl}) \\ &\quad + n^{-2}\sum_{(i,j,k,l):\{i,j\}\cap\{k,l\}=\emptyset} \mathrm{Cov}(a_i b_j q_{ij}, a_k b_l q_{kl}) \quad \text{(O.C.7)} \end{aligned}$$

In the last expression, the first term sums over all the indices i, j, k, and l such that $\{i, j\}$ and $\{k, l\}$ have at least one common element, and the second term sums over all the indices i, j, k, and l such that $\{i, j\}$ and $\{k, l\}$ do not overlap.

Because $\boldsymbol{a}$ is a function of $\boldsymbol{\varsigma}$ and $\boldsymbol{q}$ is a function of $\boldsymbol{w}$, if $\boldsymbol{b}$ is independent of $\boldsymbol{w}$ conditional on $\boldsymbol{\varsigma}$, we can write the covariance as

$$\begin{aligned} & \mathrm{Cov}(a_i b_j q_{ij}, a_k b_l q_{kl}) \\ = \;& \mathbb{E}[a_i a_k \mathbb{E}[b_j|\varsigma_j]\mathbb{E}[b_l|\varsigma_l]\mathbb{E}[q_{ij}q_{kl}|\boldsymbol{\varsigma}]] - \mathbb{E}[a_i\mathbb{E}[b_j|\varsigma_j]\mathbb{E}[q_{ij}|\boldsymbol{\varsigma}]]\mathbb{E}[a_k\mathbb{E}[b_l|\varsigma_l]\mathbb{E}[q_{kl}|\boldsymbol{\varsigma}]] \\ = \;& \mathbb{E}[a_i a_k \mathbb{E}[b_j|\varsigma_j]\mathbb{E}[b_l|\varsigma_l] q_{ij} q_{kl}] - \mathbb{E}[a_i \mathbb{E}[b_j|\varsigma_j] q_{ij}]\mathbb{E}[a_k\mathbb{E}[b_l|\varsigma_l] q_{kl}] \\ = \;& \mathrm{Cov}(a_i\mathbb{E}[b_j|\varsigma_j]q_{ij}, a_k\mathbb{E}[b_l|\varsigma_l]q_{kl}), \end{aligned}$$

where in the first equality we used $\mathbb{E}[b_j b_l|\boldsymbol{\varsigma}] = \mathbb{E}[b_j|\varsigma_j]\mathbb{E}[b_l|\varsigma_l]$ and $\mathbb{E}[b_j|\boldsymbol{\varsigma}] = \mathbb{E}[b_j|\varsigma_j]$ because (b_i, ς_i) is i.i.d.. Let $h_{ij} = a_i b_j$ (if $\boldsymbol{b}$ is a function of $\boldsymbol{\varsigma}$) or $h_{ij} = a_i\mathbb{E}[b_j|\varsigma_j]$ (if $\boldsymbol{b}$ is independent of $\boldsymbol{w}$ conditional on $\boldsymbol{\varsigma}$). The boundedness assumptions in conditions (ii) and (iii) then imply that $\max_{i,j\in\mathcal{N}} |h_{ij}| < \infty$.

The first sum in (O.C.7) consists of $O(n^3)$ terms. By Lemma O.C.10(i), each covariance term can be bounded by $O(n^{-2})$ uniformly in i, j, k, and l. Hence, the first sum is $n^{-2}\cdot O(n^3)\cdot O(n^{-2}) = o(1)$. The second sum in (O.C.7) consists of $O(n^4)$ terms. Applying Lemma O.C.10(ii) yields $\max_{i,j,k,l\in\mathcal{N}:\{i,j\}\cap\{k,l\}=\emptyset} |\mathrm{Cov}(a_i b_j q_{ij}, a_k b_l q_{kl})| = o(n^{-2})$. Hence, the last sum in equation (O.C.7) is $n^{-2}\cdot O(n^4)\cdot o(n^{-2}) = o(1)$. □

Lemma O.C.10. *Let $\boldsymbol{q}$ be a matrix that takes the form of (a) $\boldsymbol{w}$, (b) $\boldsymbol{w}'\boldsymbol{w}$, (c) $\boldsymbol{w}\boldsymbol{s}\boldsymbol{w}^t$, $t = 0, 1$, or (d) $\boldsymbol{w}'\boldsymbol{w}\boldsymbol{s}\boldsymbol{w}^t$, $r = 0, 1$, where $\boldsymbol{s} = (I_n - \gamma_1\boldsymbol{w})^{-1}$. For $h_{ij} = h(\varsigma_i, \varsigma_j) \in \mathbb{R}$ such that $\max_{i,j\in\mathcal{N}} |h_{ij}| < \infty$, $\boldsymbol{q}$ satisfies (i) $\max_{i,j,k,l\in\mathcal{N}} |\mathrm{Cov}(h_{ij}q_{ij}, h_{kl}q_{kl})| = O(n^{-2})$ and (ii) $\max_{i,j,k,l\in\mathcal{N}:\{i,j\}\cap\{k,l\}=\emptyset} |Cov(h_{ij}q_{ij}, h_{kl}q_{kl})| = o(n^{-2})$.*

Proof. Part (i). Write $\mathrm{Cov}(h_{ij}q_{ij}, h_{kl}q_{kl}) = \mathbb{E}[h_{ij}h_{kl}q_{ij}q_{kl}] - \mathbb{E}[h_{ij}q_{ij}]\mathbb{E}[h_{kl}q_{kl}]$. By the boundedness of h_{ij} and Cauchy-Schwarz inequality, it is sufficient to show that

$\mathbb{E}[\|\boldsymbol{q}\|_\infty^2] = O(n^{-2})$. For case (a) with $\boldsymbol{q} = \boldsymbol{w}$, the result follows immediately from Assumption 14(ii). For case (b) with $\boldsymbol{q} = \boldsymbol{w}'\boldsymbol{w}$, we can bound $\|\boldsymbol{q}\|_\infty \leq \|\|\boldsymbol{w}\|\|_1 \|\boldsymbol{w}\|_\infty \leq n\|\boldsymbol{w}\|_\infty^2$. Hence, $\mathbb{E}[\|\boldsymbol{q}\|_\infty^2] \leq n^2\mathbb{E}[\|\boldsymbol{w}\|_\infty^4] = O(n^{-2})$ by Assumption 14(ii). For case (c), because $\boldsymbol{s} = (I_n - \gamma_1\boldsymbol{w})^{-1} = \sum_{r=0}^\infty \gamma_1^r \boldsymbol{w}^r$, we have $\boldsymbol{q} = \boldsymbol{w}\boldsymbol{s}\boldsymbol{w}^t = \sum_{r=t+1}^\infty \gamma_1^{r-(t+1)}\boldsymbol{w}^r$ and $(q_{ij})^2 = \sum_{r=t+1}^\infty \sum_{\tilde{r}=t+1}^\infty \gamma_1^{r+\tilde{r}-2(t+1)}(\boldsymbol{w}^r)_{ij}(\boldsymbol{w}^{\tilde{r}})_{ij}$, $t = 0, 1$. For any $r \geq 1$, we have $\|\boldsymbol{w}^r\|_\infty \leq \|\|\boldsymbol{w}\|\|_\infty \|\boldsymbol{w}^{r-1}\|_\infty \leq \cdots \leq \|\|\boldsymbol{w}\|\|_\infty^{r-1} \|\boldsymbol{w}\|_\infty = \|\boldsymbol{w}\|_\infty$. Therefore, $\mathbb{E}[\|\boldsymbol{q}\|_\infty^2] \leq \sum_{r=t+1}^\infty \sum_{\tilde{r}=t+1}^\infty \gamma_1^{r+\tilde{r}-2(t+1)}\mathbb{E}[\|\boldsymbol{w}\|_\infty^2] = O(n^{-2})$ by Assumption 14(ii). Similarly as in cases (b)(c), we can show that the result holds for case (d).

Part (ii). For case (a) with $\boldsymbol{q} = \boldsymbol{w}$ and case (b) with $\boldsymbol{q} = \boldsymbol{w}'\boldsymbol{w}$, the statement follows immediately from Assumption 14(iv). For case (c), consider $i, j, k, l \in \mathcal{N}$ such that $\{i, j\} \cap \{k, l\} = \emptyset$. We have

$$\mathrm{Cov}(h_{ij}q_{ij}, h_{kl}q_{kl}) = \sum_{r=t+1}^\infty \sum_{\tilde{r}=t+1}^\infty \gamma_1^{r+\tilde{r}-2(t+1)}\mathrm{Cov}(h_{ij}(\boldsymbol{w}^r)_{ij}, h_{kl}(\boldsymbol{w}^{\tilde{r}})_{kl}).$$

By Assumption 14(iv), each term in the sum has an uniform bound $o(n^{-2})$ that does not depend on i, j, k, l, r, and $\tilde{r}$. The statement is thus satisfied for case (c). Case (d) can be proved similarly. □

O.C.2 Asymptotic Distribution of $\hat{\gamma}$

Lemma O.C.11 (Asymptotically linear representation of the moment)**.** *We have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n m(\omega_i, \gamma_0, \hat{\mu}^Z(\hat{\tau}_i)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n ((Z_i - \mu_0^Z(\tau_i))\nu_i + M_\theta\phi_\theta(z_i, \theta_0)) + o_p(1), \quad \text{(O.C.8)}$$

where $M_\theta = -\mathbb{E}[(\mathbb{E}[Z_i|z_i] - \mu_0^Z(\tau_i))\frac{\partial\lambda_0(\tau_i)}{\partial\tau}\frac{\partial\tau(z_i,g_i,\theta_0)}{\partial\theta}]$.

Proof. Consider the decomposition

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n m(\omega_i, \gamma_0, \hat{\mu}^Z(\hat{\tau}_i)) \\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n m(\omega_i, \gamma_0, \mu_0^Z(\tau_i)) + \sqrt{n}\int D(\epsilon_i, \hat{\mu}^Z(\hat{\tau}_i) - \mu^Z(\hat{\tau}_i))dF(z_i, g_i, \epsilon_i) \\
&+ \sqrt{n}\int D(\epsilon_i, \mu^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))dF(z_i, g_i, \epsilon_i) \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n (D(\epsilon_i, \hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i)) - \int D(\epsilon_i, \hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))dF(z_i, g_i, \epsilon_i)), \quad \text{(O.C.9)}
\end{aligned}$$

where $D(\epsilon_i, \mu) = -\mu\epsilon_i$ for any $\mu \in \mathbb{R}^{d_Z}$, $\mu^Z(\hat{\tau}_i) = \mathbb{E}[Z_i|\tau(z_i, g_i, \hat{\theta})]$, and $F(z_i, g_i, \epsilon_i)$ denotes the cdf of (z_i, g_i, ϵ_i). The first term is a leading term. The second term is to adjust for the estimation of μ_0^Z, and the third term is to adjust for the estimation of θ_0 (Hahn and Ridder, 2013), Both terms contribute to the asymptotic distribution of $\hat{\gamma}$. The last term is $o_p(1)$ by Lemma O.C.14.

The second term in (O.C.9) can be analyzed following Newey (1994a). Observe that for an arbitrary mean square integrable function $\mu(\tau(z_i, g_i, \theta)) \in \mathbb{R}^{d_Z}$ that is continuously differentiable in τ, by iterated expectations we have $\mathbb{E}[D(\epsilon_i, \mu(\tau(z_i, g_i, \theta))] = -\mathbb{E}[\mu(\tau(z_i, g_i, \theta))\mu^\epsilon(\tau(z_i, g_i, \theta))]$, where $\mu^\epsilon(\tau(z_i, g_i, \theta)) = \mathbb{E}[\epsilon_i|\tau(z_i, g_i, \theta)]$. Hence, the correction term in Newey (1994a, Proposition 4) takes the form $\alpha^Z(\omega_i, \tau(z_i, g_i, \theta)) = -(Z_i - \mu^Z(\tau(z_i, g_i, \theta)))\mu^\epsilon(\tau(z_i, g_i, \theta))$ and thus $\sqrt{n}\int D(\epsilon_i, \hat{\mu}^Z(\hat{\tau}_i) - \mu^Z(\hat{\tau}_i))dF(z_i, g_i, \epsilon_i) = n^{-1/2}\sum_{i=1}^n \alpha_0^Z(\omega_i, \hat{\tau}_i)$. Moreover, recall that $\hat{\tau}_i = \tau(z_i, g_i, \hat{\theta})$ and $\tau_i = \tau(z_i, g_i, \theta_0)$. Define $\alpha_0^Z(\omega_i, \tau_i) = -(Z_i - \mu_0^Z(\tau_i))\lambda_0(\tau_i)$. Under Assumption 16(ii), expanding $\alpha^Z(\omega_i, \hat{\tau}_i)$ around θ_0 yields $\alpha^Z(\omega_i, \hat{\tau}_i) = \alpha_0^Z(\omega_i, \tau_i) + \frac{\partial \alpha^Z(\omega_i, \tau_i)}{\partial \theta'}(\hat{\theta} - \theta_0) + o_p(\|\hat{\theta} - \theta_0\|)$. By Lemma O.C.15 and Assumption 12(ii), $n^{-1}\sum_{i=1}^n \frac{\partial \alpha^Z(\omega_i, \tau_i)}{\partial \theta'} = o_p(1)$ and $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$. Therefore, we can represent $\sqrt{n}\int D(\epsilon_i, \hat{\mu}^Z(\hat{\tau}_i) - \mu^Z(\hat{\tau}_i))dF(z_i, g_i, \epsilon_i) = n^{-1/2}\sum_{i=1}^n \alpha_0^Z(\omega_i, \tau_i) + o_p(1)$.

The third term in (O.C.9) can be analyzed following Hahn and Ridder (2013). Observe that $\frac{\partial D(\epsilon_i, \mu_0^Z(\tau_i))}{\partial \mu^Z} = -\epsilon_i$ and $\mathbb{E}[\frac{\partial D(\epsilon_i, \mu_0^Z(\tau_i))}{\partial \mu^Z}|\tau_i = \tau] = -\lambda_0(\tau)$. The first term in Hahn and Ridder (2013, Theorem 4) takes the form $-\mathbb{E}[(\epsilon_i - \lambda_0(\tau_i))\frac{\partial \mu_0^Z(\tau_i)}{\partial \tau}\frac{\partial \tau(z_i, g_i, \theta_0)}{\partial \theta}] = 0$,where we used $\epsilon_i - \lambda_0(\tau_i) = \nu_i$ and $\mathbb{E}[\nu_i|z_i, g_i] = 0$. Therefore, by Hahn and Ridder (2013, Theorem 4),

$$\begin{aligned} & \sqrt{n}\int D(\epsilon_i, \mu^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))dF(z_i, g_i, \epsilon_i) \\ = \; & -\mathbb{E}\left[(\mathbb{E}[Z_i|z_i] - \mu_0^Z(\tau_i))\frac{\partial \lambda_0(\tau_i)}{\partial \tau}\frac{\partial \tau(z_i, g_i, \theta_0)}{\partial \theta}\right]\sqrt{n}(\hat{\theta} - \theta_0) = M_\theta\sqrt{n}(\hat{\theta} - \theta_0). \end{aligned}$$

Because $\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \phi_\theta(z_i, \theta_0) + o_p(1)$, we can represent $\sqrt{n}\int D(\epsilon_i, \mu^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))dF(z_i, g_i, \epsilon_i) = n^{-1/2}\sum_{i=1}^n M_\theta\phi_\theta(z_i, \theta_0) + o_p(1)$. Note that $m(\omega_i, \gamma_0, \mu_0^Z(\tau_i)) + \alpha^Z(\omega_i, \tau_i) = (Z_i - \mu_0^Z(\tau_i))\nu_i$. Combining the results we obtain (O.C.8). □

Lemma O.C.12 (CLT of the moment)**.** *Let* $\Phi_n = n^{-1/2}\sum_{i=1}^n((Z_i - \mu_0^Z(\tau_i))\nu_i + M_\theta\phi_\theta(z_i, \theta_0))$. *Then* $\Omega_n^{-1/2}\Phi_n \xrightarrow{d} N(0, I_{d_Z})$, *where* $\varphi_n(x_i, z_i, \nu_i) \in \mathbb{R}^{d_Z}$ *is defined in equation (O.C.10),* $\Omega_n = n^{-1}\sum_{i=1}^n \mathbb{E}[\varphi_n(x_i, z_i, \nu_i)\varphi_n(x_i, z_i, \nu_i)']$, *and* I_{d_Z} *is the* $d_Z \times d_Z$ *identity matrix.*

Proof. Recall that $Z_i = (w_i\boldsymbol{x}, x_i', z_i')'$. While x_i and z_i are i.i.d., $w_i\boldsymbol{x}$ are correlated across i. Lemma O.C.13 establishes the Hoeffding projection $n^{-1/2}\sum_{i=1}^n (w_i\boldsymbol{x})'\nu_i = n^{-1/2}\sum_{i=1}^n h_n^*(x_i,\nu_i) + o_p(1)$, where $h_n^*(x_i,\nu_i) = \sum_j(\mathbb{E}[w_{ij}x_j\nu_i|x_i,\nu_i] + \mathbb{E}[w_{ji}\nu_j x_i|x_i,\nu_i]) \in \mathbb{R}^{d_x}$. Define the function $\varphi_n(x_i,z_i,\nu_i) \in \mathbb{R}^{d_Z}$ by

$$\begin{aligned}\varphi_n(x_i,z_i,\nu_i) \quad &= \quad n^{-1/2}((h_n^*(x_i,\nu_i)' - \mathbb{E}[(w_i\boldsymbol{x})'|\tau_i]\nu_i, (x_i' - \mathbb{E}[x_i'|\tau_i])\nu_i, (z_i' - \mathbb{E}[z_i'|\tau_i])\nu_i)' \\ &\quad + M_\theta\phi_\theta(z_i,\theta_0)). \qquad \text{(O.C.10)}\end{aligned}$$

Then $\Phi_n = \sum_{i=1}^n \varphi_n(x_i,z_i,\nu_i) + o_p(1)$. Because $\mathbb{E}[h_n^*(x_i,\nu_i)] = 0$, $\mathbb{E}[\nu_i|\boldsymbol{\varsigma}] = 0$, and $\mathbb{E}[\phi_\theta(z_i,\theta_0)] = 0$, we obtain $\mathbb{E}[\varphi_n(x_i,z_i,\nu_i)] = 0$.

Write $\varphi_{ni} = \varphi_n(x_i,z_i,\nu_i)$. Observe that $\{\varphi_{ni}, i = 1,\ldots,n\}$ forms a triangular array. We apply the Lindeberg-Feller CLT to derive the asymptotic distribution of $\sum_{i=1}^n \varphi_{ni}$. By the Cramer-Wold device it suffices to show that $a'\sum_{i=1}^n \varphi_{ni}$ satisfies the Lindeberg condition for any $d_Z \times 1$ vector of constants $a \in \mathbb{R}^{d_Z}$. The Lindeberg condition is that for any $\kappa > 0$, $\lim_{n\to\infty}\sum_{i=1}^n \mathbb{E}[\frac{(a'\varphi_{ni})^2}{a'\Omega_n a}1\{|a'\varphi_{ni}| \geq \kappa\sqrt{a'\Omega_n a}\}] = 0$. The sum is bounded by $\mathbb{E}[\sum_i \frac{(a'\varphi_{ni})^2}{a'\Omega_n a}1\{\max_i |a'\varphi_{ni}| \geq \kappa\sqrt{a'\Omega_n a}\}]$, where the random variable $\sum_i \frac{(a'\varphi_{ni})^2}{a'\Omega_n a}$ has a finite expectation and is therefore $O_p(1)$. Moreover, we can derive $\max_i |a'\varphi_{ni}| = o_p(1)$,[45] and therefore $\sum_i \frac{(a'\varphi_{ni})^2}{a'\Omega_n a}1\{\max_i |a'\varphi_{ni}| \geq \kappa\sqrt{a'\Omega_n a}\} = O_p(1)o_p(1) = o_p(1)$. This random variable is bounded by $\sum_i \frac{(a'\varphi_{ni})^2}{a'\Omega_n a}$ which has a finite expectation. We conclude that by dominated convergence the Lindeberg condition is satisfied. By Lindeberg-Feller CLT, $\Omega_n^{-1/2}\Phi_n = \Omega_n^{-1/2}\sum_{i=1}^n \varphi_n(x_i,z_i,\nu_i) + o_p(1) \xrightarrow{d} N(0, I_{d_Z})$. □

Lemma O.C.13 (Hoeffding projection)**.** *Let $W_n = n^{-1/2}\sum_i\sum_j w_{ij}x_j\nu_i$. Define $W_n^* = n^{-1/2}\sum_i h_n^*(x_i,\nu_i)$, where $h_n^*(x_i,\nu_i) = \sum_j(\mathbb{E}[w_{ij}x_j\nu_i|x_i,\nu_i] + \mathbb{E}[w_{ji}x_i\nu_j|x_i,\nu_i])$. Then $\|W_n - W_n^*\| = o_p(1)$.*

Proof. Our proof is based on Lee (1990, Secion 3.7.5) for weighted U-statistics. The idea is to generalize the Hoeffding projection to allow for weights. What differs from Lee (1990, Secion 3.7.5) is that Lee assumed a constant weight w_{ij}, while we allow w_{ij} to be a random variable that is correlated with $\boldsymbol{x}$.

Let $I = \{i_1, i_2\}$ be an ordered 2-subset of $\mathcal{N}$ and $t_i = (x_i,\nu_i)$. Define $w_I = w_{i_1 i_2}$ and $h(t_I) = h(t_{i_1}, t_{i_2}) = x_{i_2}\nu_{i_1}$. We can write $W_n = n^{-1/2}\sum_I w_I h(t_I)$ and $h_n^*(x_i,\nu_i) = h_n^*(t_i) = \sum_{I:i\in I}\mathbb{E}[w_I h(t_I)|t_i]$. Because $\boldsymbol{w}$ is independent of $\boldsymbol{\nu}$ conditional on $\boldsymbol{\varsigma}$ and

[45] Because x_i and z_i are bounded, $\mathbb{E}[\|\boldsymbol{w}\|_\infty^4] = O(n^{-4})$, and $\mathbb{E}[\nu_i^4] < \infty$, we can bound $\mathbb{E}[\max_i(a'\varphi_{ni})^2] \leq \|a\|^2\mathbb{E}[\max_i\|\varphi_{ni}\|^2] \leq O(n^{-1}) = o(1)$.

$\mathbb{E}[\nu_i|\boldsymbol{\varsigma}] = 0$ (Assumption 14(iii)), we have $\mathbb{E}[w_I h(t_I)] = \mathbb{E}[\mathbb{E}[w_I|\boldsymbol{\varsigma}]\mathbb{E}[h(t_I)|\boldsymbol{\varsigma}]] = 0$ and $\mathbb{E}[h_n^*(t_i)] = \sum_{I:i\in I}\mathbb{E}[w_I h(t_I)] = 0$. By Markov's inequality, it suffices to show $\mathbb{E}\|W_n - W_n^*\|^2 = o(1)$.

By definition, $\mathbb{E}[W_n' W_n^*] = n^{-1/2}\sum_i \mathbb{E}[W_n' h_n^*(t_i)]$, and for each i, $\mathbb{E}[W_n' h_n^*(t_i)] = n^{-1/2}\sum_{I:i\in I}\mathbb{E}[w_I h(t_I)' h_n^*(t_i)] = n^{-1/2}\mathbb{E}[h_n^*(t_i)' h_n^*(t_i)]$, where the first equality holds because for $i \notin I$, $\mathbb{E}[w_I h(t_I)' h_n^*(t_i)] = \mathbb{E}[\mathbb{E}[w_I|\boldsymbol{\varsigma}]\mathbb{E}[h(t_I)' h_n^*(t_i)|\boldsymbol{\varsigma}]] = 0$ under Assumption 14(iii), and the second equality follows by iterated expectations. It then follows that $\mathbb{E}[W_n' W_n^*] = n^{-1}\sum_i \mathbb{E}[h_n^*(t_i)' h_n^*(t_i)] = \mathbb{E}\|W_n^*\|^2$ and thus $\mathbb{E}\|W_n - W_n^*\|^2 = \mathbb{E}\|W_n\|^2 - \mathbb{E}\|W_n^*\|^2$. It remains to show that $\mathbb{E}\|W_n\|^2 - \mathbb{E}\|W_n^*\|^2 = o(1)$.

To show the last result, note that for disjoint I and J, we have $\mathbb{E}[w_I w_J h(t_I)' h(t_J)] = \mathbb{E}[\mathbb{E}[w_I w_J|\boldsymbol{\varsigma}]\mathbb{E}[h(t_I)' h(t_J)|\boldsymbol{\varsigma}]] = 0$, where the first equality holds by Assumption 14(iii), and the last equality follows from $\mathbb{E}[h(t_I)' h(t_J)|\boldsymbol{\varsigma}] = 0$ because (ν_i, ς_i) is i.i.d.. Hence,

$$\begin{aligned}\mathbb{E}\|W_n\|^2 &= n^{-1}\sum_{(I,J):|I\cap J|=1}\mathbb{E}[w_I w_J h(t_I)' h(t_J)] \\ &\quad + n^{-1}\sum_{(I,J):|I\cap J|=2}\mathbb{E}[w_I w_J h(t_I)' h(t_J)].\end{aligned}$$

For comparison, because $\mathbb{E}\|W_n^*\|^2 = n^{-1}\sum_{i=1}^n \mathbb{E}\|h_n^*(t_i)\|^2$ we can write

$$\begin{aligned}\mathbb{E}\|W_n^*\|^2 &= n^{-1}\sum_{i=1}^n \sum_{(I,J):\{i\}=I\cap J}\mathbb{E}[\mathbb{E}[w_I h(t_I)'|t_i]\mathbb{E}[w_J h(t_J)|t_i]] \\ &\quad + n^{-1}\sum_{i=1}^n \sum_{(I,J):\{i\}\subsetneq I\cap J}\mathbb{E}[\mathbb{E}[w_I h(t_I)'|t_i]\mathbb{E}[w_J h(t_J)|t_i]].\end{aligned}$$

The first sums in $\mathbb{E}\|W_n\|^2$ and $\mathbb{E}\|W_n^*\|^2$ consist of the same number of terms. Consider I and J such that $|I\cap J| = 1$. Because $\boldsymbol{w}$ and $\boldsymbol{\nu}$ are independent conditional on $\boldsymbol{\varsigma}$,

$$\begin{aligned}\mathbb{E}[w_I w_J h(t_I)' h(t_J)] &= \mathbb{E}[\mathbb{E}[w_I w_J|\boldsymbol{\varsigma}, \boldsymbol{\nu}] h(t_I)' h(t_J)] \\ &= \mathbb{E}[\mathbb{E}[w_I w_J|\boldsymbol{\varsigma}] h(t_I)' h(t_J)] \\ &= \mathbb{E}[(\mathrm{Cov}(w_I, w_J|\boldsymbol{\varsigma}) + \mathbb{E}[w_I|\boldsymbol{\varsigma}]\mathbb{E}[w_J|\boldsymbol{\varsigma}]) h(t_I)' h(t_J)] \\ &= \mathbb{E}[\mathbb{E}[w_I|\varsigma_I]\mathbb{E}[w_J|\varsigma_J] h(t_I)' h(t_J)] + o(n^{-2}), \qquad \text{(O.C.11)}\end{aligned}$$

where the $o(n^{-2})$ term does not depend on I and J. To see the last equality, note that $h(t_I)' h(t_J)$ is square integrable by the boundedness of x and $\mathbb{E}[\nu_i^4] < \infty$ under Assumption 11(i) and 16(i). The last equality then follows from Assumption

14(ii)(v)(vi).[46] Similarly, we can derive

$$\begin{aligned}\mathbb{E}[\mathbb{E}[w_I h(t_I)'|t_i]\mathbb{E}[w_J h(t_J)|t_i]] &= \mathbb{E}(\mathbb{E}[\mathbb{E}[w_I|\boldsymbol{\varsigma},\boldsymbol{\nu}]h(t_I)'|t_i]\mathbb{E}[\mathbb{E}[w_J|\boldsymbol{\varsigma},\boldsymbol{\nu}]h(t_J)|t_i]] \\ &= \mathbb{E}(\mathbb{E}[\mathbb{E}[w_I|\boldsymbol{\varsigma}]h(t_I)'|t_i]\mathbb{E}[\mathbb{E}[w_J|\boldsymbol{\varsigma}]h(t_J)|t_i]] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{E}[w_I|\varsigma_I]h(t_I)'|t_i]\mathbb{E}[\mathbb{E}[w_J|\varsigma_J]h(t_J)|t_i]] + o(n^{-2}) \\ &= \mathbb{E}[\mathbb{E}[w_I|\varsigma_I]\mathbb{E}[w_J|\varsigma_J]h(t_I)'h(t_J)] + o(n^{-2}), \quad \text{(O.C.12)}\end{aligned}$$

where the last equality follows because for I and J with $\{i\} = I \cap J$, $\mathbb{E}[w_I|\varsigma_I]h(t_I)$ and $\mathbb{E}[w_J|\varsigma_J]h(t_J)$ are independent conditional on t_i. Comparing (O.C.11) with (O.C.12) we can see that the two covariances differ by $o(n^{-2})$ uniformly in I and J. Because the first sums in $\mathbb{E}\|W_n\|^2$ and $\mathbb{E}\|W_n^*\|^2$ consist of $O(n^3)$ terms, they differ by $n^{-1} \cdot O(n^3) \cdot o(n^{-2}) = o(1)$.

The second sums in $\mathbb{E}\|W_n\|^2$ and $\mathbb{E}\|W_n^*\|^2$ consist of $O(n^2)$ terms. For any I and J, both $\mathbb{E}[w_I w_J h(t_I)'h(t_J)]$ and $\mathbb{E}[\mathbb{E}[w_I h(t_I)'|t_i]\mathbb{E}[w_J h(t_J)|t_i]]$ can be uniformly bounded by $O(n^{-2})$ (Assumption 14(ii)). Therefore, the second sums in $\mathbb{E}\|W_n\|^2$ and $\mathbb{E}\|W_n^*\|^2$ are both $n^{-1}\cdot O(n^2)\cdot O(n^{-2}) = o(1)$. We conclude that $\mathbb{E}\|W_n\|^2 - \mathbb{E}\|W_n^*\|^2 = o(1)$. □

Lemma O.C.14.

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(D(\epsilon_i, \hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i)) - \int D(\epsilon_i, \hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))dF(z_i, g_i, \epsilon_i)) = o_p(1). \quad \text{(O.C.13)}$$

Proof. Let $\mu = \mu(\tau(z_i, g_i, \theta)) \in \mathbb{R}^{d_Z}$ be a function of $\tau(z_i, g_i, \theta)$. Define the empirical process $\mathbb{G}_n(\mu) = \frac{1}{\sqrt{n}}\sum_i(D(\epsilon_i, \mu) - \mathbb{E}[D(\epsilon_i, \mu)])$ indexed by μ. We can represent the left-hand side of equation (O.C.13) as $\mathbb{G}_n(\hat{\mu}^Z(\hat{\boldsymbol{\tau}})) - \mathbb{G}_n(\mu_0^Z(\boldsymbol{\tau}))$.

Observe that $D(\epsilon_i, \mu) = -\mu\epsilon_i$ is linear in μ. This together with the boundedness of Z_i and $\mathbb{E}[\epsilon_i^2] < \infty$ (Assumptions 11, 14(i), and 16(i)) implies that the empirical process $\mathbb{G}_n(\mu)$ is stochastically equicontinuous under L_2 norm (Andrews, 1994, Theorems 1-2). It remains to show that $\int \|\hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))\|^2 dF(z_i, g_i) = o_p(1)$, where $F(z_i, g_i)$ denotes the cdf of (z_i, g_i). We prove it following Newey (1997, Theorem 1).

By the triangle inequality and $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$, we derive

$$\int \|\hat{\mu}^Z(\hat{\tau}_i) - \mu_0^Z(\tau_i))\|^2 dF(z_i, g_i)$$

[46] Observe that Assumption 14(v) implies that $\max_{I\subseteq\mathcal{N}} \mathbb{E}[(\mathbb{E}[w_I|\boldsymbol{\varsigma}] - \mathbb{E}[w_I|\varsigma_I])^4] = o(n^{-4}/K^2) \leq o(n^{-4})$.

$$
\begin{aligned}
&\leq \; 3\int(\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})'(b^K(\hat{\tau}_i)-b^K(\tau_i))\|^2+\|(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)'b^K(\tau_i)\|^2\\
&\quad \|\beta^{Z\prime}b^K(\tau_i)-\mu_0^Z(\tau_i)\|^2)dF(z_i,g_i). \qquad \text{(O.C.14)}
\end{aligned}
$$

Consider the three terms in the last equation. The first term satisfies

$$
\begin{aligned}
&\int\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})'(b^K(\hat{\tau}_i)-b^K(\tau_i))\|^2dF(z_i,g_i)\\
\leq\; & O_p(\varrho_1(K)^2)\int\max_{1\leq i\leq n}\|\hat{\tau}_i-\tau_i\|^2dF(z_i,g_i)=O_p(\varrho_1(K)^2/n),
\end{aligned}
$$

where the inequality holds by equation (O.C.2), the mean-value theorem and Assumption 13(iv), and the equality holds because the $\sqrt{n}$-consistency of $\hat{\theta}$ and boundedness of z imply that $\max_{1\leq i\leq n}\|\hat{\tau}_i-\tau_i\|=O_p(n^{-1/2})$. As for the second term in (O.C.14), by $\mathbb{E}[b^K(\tau_i)b^{K\prime}(\tau_i)]=I_K$ we obtain

$$
\begin{aligned}
&\int\|(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)'b^K(\tau_i)\|^2dF(z_i,g_i)\\
=\; & \mathrm{tr}((\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z)'\int b^K(\tau_i)b^{K\prime}(\tau_i)dF(z_i,g_i)(\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z))\\
=\; & \|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z\|^2=O_p(\varrho_1(K)^2/n)+o_p(1),
\end{aligned}
$$

where the last equality follows from $\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\beta^Z\|^2\leq 2(\|\hat{\beta}^Z(\hat{\boldsymbol{\tau}})-\hat{\beta}^Z(\boldsymbol{\tau})\|^2+\|\hat{\beta}^Z(\boldsymbol{\tau})-\beta^Z\|^2)$, Lemmas O.C.7 and O.C.8, and Li and Racine (2007, Lemma 15.3). The third term in (O.C.14) has the bound $\int\|\beta^{Z\prime}b^K(\tau_i)-\mu_0^Z(\tau_i)\|^2dF(z_i,g_i)\leq\sup_\tau\|\beta^{Z\prime}b^K(\tau)-\mu_0^Z(\tau)\|=O(K^{-2a})$ by Assumption 13(ii). Combining the results yields $\int\|\hat{\mu}^Z(\hat{\tau}_i)-\mu_0^Z(\tau_i))\|^2dF(z_i,g_i)=o_p(1)$ and $\mathbb{G}_n(\hat{\mu}^Z(\hat{\boldsymbol{\tau}}))-\mathbb{G}_n(\mu_0^Z(\boldsymbol{\tau}))=o_p(1)$.

□

Lemma O.C.15. $\frac{1}{n}\sum_{i=1}^n\frac{\partial\alpha^Z(\omega_i,\tau_i)}{\partial\theta'}=o_p(1)$.

Proof. Recall that $\alpha^Z(\omega_i,\tau(z_i,g_i,\theta))=-(Z_i-\mu^Z(\tau(z_i,g_i,\theta)))\mu^\epsilon(\tau(z_i,g_i,\theta))$, where $\mu^Z(\tau(z_i,g_i,\theta))=\mathbb{E}[Z_i|\tau(z_i,g_i,\theta)]$ and $\mu^\epsilon(\tau(z_i,g_i,\theta))=\mathbb{E}[\epsilon_i|\tau(z_i,g_i,\theta)]$. By the law of iterated expectations we have $\mathbb{E}[\alpha^Z(\omega_i,\tau(z_i,g_i,\theta))]=0$, so $\mathbb{E}[\partial\alpha^Z(\omega_i,\tau_i)/\partial\theta']=\partial\mathbb{E}[\alpha^Z(\omega_i,\tau(z_i,g_i,\theta))]/\partial\theta'=0$.

Differentiating $\alpha^Z(\omega_i,\tau(z_i,g_i,\theta))$ with respect to θ at θ_0 yields

$$
\frac{\partial\alpha^Z(\omega_i,\tau_i)}{\partial\theta'}=\left(\frac{\partial\mu^Z(\tau_i)}{\partial\tau_i}\mu^\epsilon(\tau_i)-(Z_i-\mu_0^Z(\tau_i))\frac{\partial\mu^\epsilon(\tau_i)}{\partial\tau_i}\right)\frac{\partial\tau(z_i,g_i,\theta_0)}{\partial\theta'}.
$$

Because $\tau_i=\tau(z_i,g_i,\theta_0)$ is bounded and $\mu^Z(\tau_i)$ and $\mu^\epsilon(\tau_i)$ are continuously differentiable in τ_i (Assumptions 11(i), 12(i), and 16(ii)), $\mu^Z(\tau_i)$, $\mu^\epsilon(\tau_i)$, $\frac{\partial \mu^Z(\tau_i)}{\partial \tau_i}$, and $\frac{\partial \mu^\epsilon(\tau_i)}{\partial \tau_i}$ are bounded. Observe that (z_i, τ_i) is i.i.d.. By the law of large numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\partial \mu^Z(\tau_i)}{\partial \tau_i}\mu^\epsilon(\tau_i)\frac{\partial \tau(z_i, g_i, \theta_0)}{\partial \theta'} - \mathbb{E}\left[\frac{\partial \mu^Z(\tau_i)}{\partial \tau_i}\mu^\epsilon(\tau_i)\frac{\partial \tau(z_i, g_i, \theta_0)}{\partial \theta'}\right]\right) = o_p(1). \tag{O.C.15}$$

Moreover, following Lemma O.C.3 we can show that

$$\frac{1}{n}\sum_{i=1}^{n}\left((Z_i - \mu_0^Z(\tau_i))\frac{\partial \mu^\epsilon(\tau_i)}{\partial \tau_i}\frac{\partial \tau(z_i, g_i, \theta_0)}{\partial \theta'} - \mathbb{E}\left[(Z_i - \mu_0^Z(\tau_i))\frac{\partial \mu^\epsilon(\tau_i)}{\partial \tau_i}\frac{\partial \tau(z_i, g_i, \theta_0)}{\partial \theta'}\right]\right) = o_p(1). \tag{O.C.16}$$

Combining (O.C.15) and (O.C.16) proves the lemma. □