ESTIMATION OF PEER EFFECTS IN ENDOGENOUS SOCIAL NETWORKS: CONTROL FUNCTION APPROACH

IDA JOHNSSON

Department of Economics, University of Southern California

HYUNGSIK ROGER MOON

Department of Economics, University of Southern California,

and School of Economics, Yonsei University

Abstract. We propose a method of estimating the linear-in-means model of peer effects in which the peer group, defined by a social network, is endogenous in the outcome equation for peer effects. Endogeneity is due to unobservable individual characteristics that influence both link formation in the network and the outcome of interest. We propose two estimators of the peer effect equation that control for the endogeneity of the social connections using a control function approach. We leave the functional form of the control function unspecified and treat it as unknown. To estimate the model, we use a sieve semiparametric approach, and we establish asymptotics of the semiparametric estimator.

Keywords: peer effects, endogenous network, sieve estimation, control function

JEL Classification: C14, C21

Date: September 2017.

We thank Bryan Graham and three referees for their helpful and valuable comments and suggestions. We are particularly grateful to one of the referees for suggesting the idea that is presented in Section 4.2 of the paper. We also appreciate the comments and discussions of the participants at the 2015 USC Dornsife INET Conference on Networks, University of Southern California, 2016 North American Summer Meeting of the Econometric Society, 2016 California Econometrics Conference, the 2017 Asian Meeting of Econometric Society, the 2017 IAAE conference. The first draft of the paper was written while Johnsson was a graduate fellow of USC Dornsife INET and Moon was the associate director of USC Dornsife INET. Moon acknowledges that this work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A2A01023679).

Ida Johnsoon: johnsson@usc.edu. Hyungsik Roger Moon: Corrpesponding author. moonr@usc.edu.

1. Introduction

The ways in which interconnected individuals influence each other are usually referred to as peer effects. One of the first to formally model peer effects is Manski (1993). He proposes the linear-in-means model, in which an individual's action depends on the average action of other individuals and possibly also on their average characteristics. Manski (1993) assumes that all individuals within a given group are connected. Later literature allows for more complex patterns of connections, in which an individual might be directly influenced by a subset of the group. Examples are Bramoullé et al. (2009), Lee et al. (2010), Lee (2007b) among others. Models of peer effects have been applied in various areas, such as education, health and development, and various application examples are found in recent review papers such as Blume et al. (2010), Manski (2000), Epple and Romano (2011), and Graham (2011).

Many models considered in earlier literature assume that connections between individuals are independent of unobserved individual characteristics that influence outcomes. However, the exogeneity of the network or peer group might be restrictive in many applications. For example, consider the following, widely studied, empirical application of peer effects: peer influence on scholarly achievement. The assumption that friendships are exogenous in the outcome equation for scholarly achievement means that there are no unobservables that influence both friendship formation and individual grades. However, even if the researchers control for observable individual characteristics such as gender, age, race and parents' education, it is likely they leave out factors that influence both students' choice of friends and their GPA, for example parents' expectations, psychological disorders or substance use. For more examples of endogenous peer groups and networks, see Weinberg (2007), Shalizi (2012) and Hsieh and Lee (2016), among others.

In this paper we investigate estimating a linear-in-means model of peer effects, where the peer group is defined by a network that is endogenous to the outcome equation. Our model allows correlation between the unobserved individual heterogeneity that impacts network formation and the unobserved characteristics of the outcome. For this, we use a dyadic network formation model that allows the unobserved individual attributes of two different agents to influence link formation, in which links are pairwise independent conditional on the observed and unobserved individual attributes. The network formation we consider in the paper is dense and nonparametric.

The main contributions of the paper are methodological. First, given the endogenous peer group formation, we show that we can identify the peer effects by controlling the unobserved individual heterogeneity of the network formation equation. Secondly, we propose an empirically tractable implementation of the control function whose functional form is not parametrically specified. For this, we propose two approaches, one based on the estimator of the unobserved individual heterogeneity and the other one based on the node degrees of the network.[1] Our estimation method is semiparametric because we do not restrict the functional form of the control function. Thirdly, we derive the limiting distributions of the estimators within a large single network. The main challenge of the asymptotics is to handle the strong dependence of observables caused by the dense network.

Closely related papers that adopt a control function approach include Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016), Qu and Lee (2015), and Arduini et al. (2015). Our paper adopts a frequentist approach based on nonparametric specification of the network formation, while Goldsmith-Pinkham and Imbens (2013) and Hsieh and Lee (2016) use the Bayesian method based on a full parametric specification of the network formation and the outcome equation. Like our paper, Qu and Lee (2015) assume the network (spatial weights in their model) to be endogenous through the unobserved individual heterogeneity. However, our paper is different from Qu and Lee (2015) in many aspects. They consider sparse network formation models while we consider a dense network. They restrict the functional form of the control function to be linear, while we impose no restriction on the functional form. The two papers propose different implementations of the control function. Our paper is different from Arduini et al. (2015) regarding the main source of the endogeneity of the network and the form of the control function. Arduini et al. (2015) assume that the endogeneity of the network is allowed through dependence between the outcome equation error

[1]We acknowledge that this approach is developed based on the idea provided by one of the referees. We thank the referee.

and the idiosyncratic network formation error, like the conventional sample selection model. Arduini et al. (2015) consider a control function (both parametric and semiparametric) to deal with the selection bias problem. Regarding asymptotics, in both Qu and Lee (2015) and Arduini et al. (2015) the asymptotics are derived using near-epoch dependence and are based on the assumption that the number of connections does not increase at the same rate as the square of the network size. After the completion of the first draft of our paper we were made aware of the paper by Auerbach (2016). His network formation model is the same as ours. To identify and estimate peer effects, he proposes a pairwise matching method which resembles Powell (1987), Heckman et al. (1998), and Abadie and Imbens (2006).

The remainder of the paper is organized as follows. In Section 2 we formally present our model. In Section 3, we show how to identify the peer effects using control functions. Estimation is discussed in Section 4, and in Section 5 we discuss the limiting distribution of the estimator and propose standard errors. In Section 6 we present results of Monte Carlo simulations. There we compare the finite sample performance of our two semiparametric estimators against an estimator that assumes unobserved characteristics enter in a linear way, as well as an instrumental variable (IV) estimator that does not control for network endogeneity. We investigate both high degree and low degree networks. Section 7 concludes.

A word on notation. In what follows we denote scalars by lowercase letters, vectors by lowercase bold letters, and matrices by uppercase bold letters.

2. Model of Peer Effects with an Endogenous Network

Suppose that $d_{ij,N}$ are the observed links among individuals $i \in \{1, ..., N\}$, such that $d_{ij,N} = 1$ if i and j are directly connected and 0 otherwise. We assume that individual outcomes, y_i, are given by the linear-in-means model of peer effects

$$y_i = \beta_1^0 \sum_{\substack{j=1 \\ j \neq i}}^{N} g_{ij,N} y_j + \mathbf{x}_{1i}' \beta_2^0 + \left(\sum_{\substack{j=1 \\ j \neq i}}^{N} g_{ij,N} \mathbf{x}_{1j} \right)' \beta_3^0 + \upsilon_i, \tag{2.1}$$

where $\mathbf{x}_{1i}$ are observed individual characteristics that affect the outcome y_i, υ_i are unobserved individual characteristics, and

$$g_{ij,N} = \begin{cases} 0 & \text{if} \quad i = j \\ \frac{d_{ij,N}}{\sum_{j \neq i} d_{ij,N}} & \text{otherwise.} \end{cases}$$

is the weight of the peer effects. Using the terminology of Manski (1993), β_1^0 captures the endogenous social effect, and β_3^0 measures the exogenous social effect. We let $\beta^0 := (\beta_1^0, \beta_2^{0\prime}, \beta_3^{0\prime})'$ and denote $\beta = (\beta_1, \beta_2', \beta_3')'$. Throughout the paper, we assume that $|\beta_1^0| < 1$, and so the peer effect model has a unique solution.

We let $\mathbf{D}_N$ be the $(N \times N)$ adjacency matrix of the network whose $(i,j)^{th}$ element is $d_{ij,N}$. We let $d_{ii,N} = 0$ for all i, following the convention. Let $\mathbf{G}_N$ be the matrix whose $(i,j)^{th}$ element is $g_{ij,N}$, and $\mathbf{G}_N$ is obtained by row-normalizing $\mathbf{D}_N$. Denote $\mathbf{X}_{1N} = (\mathbf{x}_{11}', \ldots, \mathbf{x}_{1N}')'$, $\mathbf{y}_N = (y_1, \ldots, y_N)'$ and $\boldsymbol{\upsilon}_N = (\upsilon_1, \ldots, \upsilon_N)'$.

In the standard linear-in-means model of peer effects, the main focus has been identification and estimation of peer effects, assuming that the peer group (or the network) is exogenous, that is, $\mathbb{E}[\upsilon_i | \mathbf{X}_{1N}, \mathbf{G}_N] = 0$. For example, see Manski (1993) and Bramoullé et al. (2009), Lee (2007b), and Blume et al. (2015). To identify and estimate the linear-in-means model of peer effects when the peer group is exogenous it is necessary to take into account the fact that the regressor $\sum_{i=1}^{N} g_{ij,N} y_j$ is correlated with the error term υ_i. For example, if $\upsilon_i \sim \ i.i.d.(0, \sigma^2)$, it is true that

$$\begin{aligned} \mathbb{E}[(\mathbf{G}_N \mathbf{y}_N)' \boldsymbol{\upsilon}_N] &= [(\mathbf{G}_N (\mathbf{I}_N - \beta_1 \mathbf{G}_N)^{-1} (\mathbf{X}_{1N} \beta_2 + \mathbf{G}_N \mathbf{X}_{1N} \beta_3 + \boldsymbol{\upsilon}_N))' \boldsymbol{\upsilon}_N] \\ &= \mathbb{E}[(\mathbf{G}_N (\mathbf{I}_N - \beta_1 \mathbf{G}_N)^{-1} \boldsymbol{\upsilon}_N)' \boldsymbol{\upsilon}_N] = \sigma_0 tr(\mathbf{G}_N (\mathbf{I}_N - \beta_1 \mathbf{G}_N)^{-1}) \neq 0. \end{aligned} \tag{2.2}$$

To solve this endogeneity problem different estimators have been proposed in the literature, see for example Kelejian and Prucha (1998), Lee (2003) and Lee (2007a). One of the widely used estimation methods is the Instrumental Variable (IV) approach based on using $\mathbf{G}_N^2 \mathbf{X}_{1N}$ as the IV of the endogenous regressor $\mathbf{G}_N \mathbf{y}_N$ (see for example Kelejian and Prucha (1998), Lee (2003), and Bramoullé et al. (2009)). Then, the natural estimator is the Two-Stage

Least Squares (2SLS) estimator

$$\hat{\beta}_N^{2SLS} = (\mathbf{W}_N'\mathbf{Z}_N(\mathbf{Z}_N'\mathbf{Z}_N)^{-1}\mathbf{Z}_N\mathbf{W}_N)^{-1}\mathbf{W}_N'\mathbf{Z}_N(\mathbf{Z}_N'\mathbf{Z}_N)^{-1}\mathbf{Z}_N'\mathbf{y}_N, \tag{2.3}$$

where $\mathbf{W}_N = [\mathbf{G}_N\mathbf{y}_N,\ \mathbf{X}_{1N},\ \mathbf{G}_N\mathbf{X}_{1N}]$ and $\mathbf{Z}_N = [\mathbf{X}_{1N},\ \mathbf{G}_N\mathbf{X}_{1N},\ \mathbf{G}_N^2\mathbf{X}_{1N}]$ is the matrix of instruments. For the IVs $\mathbf{Z}_N$ to be strong, we assume that $\beta_2^0 \neq 0$.

When the network matrix is endogenous, $\mathbb{E}[\mathbf{G}_N\boldsymbol{v}_N] \neq 0$, and the procedure used by Kelejian and Prucha (1998), Lee (2003), Bramoullé et al. (2009) and others is no longer valid since the IV matrix $\mathbf{Z}_N = [\mathbf{X}_{1N},\ \mathbf{G}_N\mathbf{X}_{1N},\ \mathbf{G}_N^2\mathbf{X}_{1N}]$ is correlated with the error term $\boldsymbol{v}_N$. Specifically, the validity of the 2SLS estimator depends on the orthogonality condition $\mathbb{E}[\boldsymbol{v}_N|\mathbf{Z}_N] = 0$, which is implied if $\mathbb{E}[\boldsymbol{v}_N|\mathbf{X}_{1N}, \mathbf{D}_N] = 0$. However, it does not hold if the network $\mathbf{D}_N$ (or equivalently, the network $\mathbf{G}_N$) is correlated with $\boldsymbol{v}_N$, which is true if unobserved individual characteristics of the network $\mathbf{D}_N$ (or $\mathbf{G}_N$) directly influence both link formation and individual outcomes.

In this paper, we consider the case where it may be that $\mathbb{E}[\boldsymbol{v}_N|\mathbf{X}_{1N}, \mathbf{D}_N] \neq 0$, so that unobserved characteristics that influence link formation can also have a direct effect on individual outcomes. This is an important consideration in many common applications, for example the impact of school friendships on scholarly achievement or substance use.

3. Endogenous Network Formation and Identification of peer effects using a control function approach

In this section, we introduce a network formation model for $d_{ij,N}$ and discuss the assumptions and implication of the model. Then we provide an identification argument for the peer effect equation when the network is endogenous.

3.1. **Model of Network Formation.** Let $\mathbf{x}_{2i}$ be a vector of observable characteristics of individual i, and let $\mathbf{x}_i = \mathbf{x}_{1i} \cup \mathbf{x}_{2i}$. Define $\mathbf{X}_{2N}$ analogously to $\mathbf{X}_{1N}$ and let $\mathbf{X}_N = \mathbf{X}_{1N} \cup \mathbf{X}_{2N}$. We introduce a_i, a scalar unobserved characteristic of individual i, which is treated as an individual fixed-effect, and hence, might be correlated with $\mathbf{x}_i$. We denote the vector of individual unobserved characteristics by $\mathbf{a}_N = (a_1, a_2, \ldots, a_N)'$. Individuals are connected by an undirected network $\mathbf{D}_N$, with the $(i,j)^{th}$ element $d_{ij,N} = 1$ if i and j are directly connected and 0 otherwise. We assume the network to be undirected[2], $d_{ij,N} = d_{ji,N}$, and assume $d_{ii,N} = 0$ for all i, following the convention. In this case, there are $n = \binom{N}{2}$ dyads. Let $\mathbf{t}_{ij}$ denote an $l_T \times 1$ vector of dyad-specific characteristics of dyad ij, and we assume that $\mathbf{t}_{ij} = t(\mathbf{x}_{2i}, \mathbf{x}_{2j})$.

Agents form links according to

$$d_{ij,N} = \mathbb{I}(g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) - u_{ij} \geq 0), \tag{3.1}$$

where $\mathbb{I}(\bullet)$ is an indicator function. In this setup, link surplus is transferable across directly linked agents and consists of three components: $\mathbf{t}_{ij} := t(\mathbf{x}_{2i}, \mathbf{x}_{2j})$ is a systematic component that varies with observed dyad attributes and accounts for homophily, a_i and a_j account for unobserved dyad attributes (degree heterogeneity), and u_{ij} is an idiosyncratic shock that is i.i.d. across dyads and independent of $\mathbf{t}_{ij}$ and a_i for all i, j. Since links are undirected, the surplus of link $d_{ij,N}$ must be the same for individual i and j. Hence, we assume that the function t_{ij} is symmetric in i and j, and the function g is symmetric in a_i and a_j.

In the literature, various parametric versions of the network formation in (3.1) are used (see Graham (2015) and the references therein.). An important example of a parametric specification is the one in Graham (2017),

$$d_{ij,N} = \mathbb{I}(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})'\lambda + a_i + a_j - u_{ij} > 0). \tag{3.2}$$

For the purpose of the paper, particularly in constructing the estimators that we introduce in Section 4, we do not need a parametric specification. Instead, we need the following two restrictions.

First, we assume that the network formed by (3.1) is dense in the sense that the expected number of connections is proportional to the square of the network size. For this, we assume that the error u_{ij} is drawn randomly from a distribution with full support, while $g(\mathbf{t}_{ij}, a_i, a_j)$ is bounded (see Assumption 11 in the Appendix). In this case, the probability of any

[2]Our analysis can be extended to the directed network case, but we do not pursue it in this paper.

two individuals forming a link is bounded away from zero. The dense network model is appropriate for scenarios where any two individuals can plausibly form a link.

Secondly, we assume that the net link surplus is a monotonic function of a_i (and a_j), that is, $g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j)$ is a strictly monotonic function of a_i (and a_j) (see Assumption 11(vi) in the Appendix.). Obviously, this condition is satisfied in the parametric model (3.2). This monotonic condition is crutial in constructing the estimator in Section 4.2.

Regarding the network formation model (3.1), it is important to notice that the network formation model (3.1) rules out interdependent link preferences, and it assumes that links are formed independently conditional on observed and unobserved characteristics. As discussed in Graham (2017), this assumption is appropriate for settings where link formation is driven predominantly by bilateral concerns, such as certain types of friendship networks, trade networks and some models of conflict between nation-states. The model in (3.1) is not a good choice when important strategic aspects influence link formation, for example when the identity of the nodes to which j is linked influences i's return from forming a link with j. A discussion of networks with interdependent links can be found in Graham (2015) and De Paula (2016). Also, when network externalities are present, the additional complication of multiple equilibria has to be considered, see for example Sheng (2012) for more details.

3.2. **Control Function of Network Endogeneity.** In this subsection we discuss how to control the endogeneity of the peer group defined by the network formed in equation (3.1). First we introduce a basic assumption that we will maintain throughout the paper.

Assumption 1. *(i)* $(\mathbf{x}_i, a_i, \upsilon_i)$ *are i.i.d. for all* i, $i = 1, \ldots, N$.
(ii) $\{u_{ij}\}_{i,j=1,\ldots,N}$ *are independent of* $(\mathbf{X}_N, \mathbf{a}_N, \boldsymbol{\upsilon}_N)$ *and i.i.d. across* (i, j) *with cdf* $\Phi(\cdot)$.
(iii) $\mathbb{E}(\upsilon_i|\mathbf{x}_i, a_i) = \mathbb{E}(\upsilon_i|a_i)$.

Assumption 1(i) implies that the observables $\mathbf{x}_i$ and the unobservable characteristics, (a_i, υ_i), are randomly drawn. This is a standard assumption in the peer effects literature. Assumption 1(ii) assumes that the link formation error u_{ij} is orthogonal to all other observables and unobservables in the model. This means that the dyad-specific unobservable shock u_{ij} from the link formation process does not influence outcomes $y_1, \ldots, y_N$. However, we allow for endogeneity of the social interaction group through dependence between the two unobserved components a_i and υ_i. This means that the unobserved error υ_i in the outcome equation can be correlated with unobserved individual characteristics a_i that are determinants of link formation. We also allow that the observed characteristics, $\mathbf{x}_i$, of the outcome equation and the network formation be correlated with the unobserved components, (υ_i, a_i), so that the regressor $\mathbf{x}_{1i}$ can be endogenous in the outcome equation, and the network formation observables $\mathbf{x}_{2i}$ can be arbitrarily correlated with the unobserved individual characteristic a_i. In Assumption 1(iii), we assume that the dependence between $\mathbf{x}_i$ and υ_i exists only through a_i. That is, a_i is the fixed effect of individual i and controls the endogeneity of $\mathbf{x}_i$ with respect to υ_i.

Notice that the network $\mathbf{D}_N$ defined in (3.1) and the (row normalized) network $\mathbf{G}_N$ are measurable functions of

$$(\mathbf{x}_{2i}, \mathbf{x}_{2,-i}, a_i, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\ldots,N}),$$

where $\mathbf{x}_{2,-i} = (\mathbf{x}_{2,1}, \ldots, \mathbf{x}_{2,i-1}, \mathbf{x}_{2,i+1}, \ldots, \mathbf{x}_{2N})$ and $\mathbf{a}_{-i}$ is defined analogously. Under Assumption 1 we have

$$\begin{aligned} \mathbb{E}[\upsilon_i|\mathbf{X}_N, \mathbf{D}_N, a_i] &= \mathbb{E}[\upsilon_i|\mathbf{x}_{-i}, \mathbf{D}_N(\mathbf{x}_{2,-i}, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\ldots,N}, \mathbf{x}_{2i}, a_i), \mathbf{x}_i, a_i] \\ &= \mathbb{E}[\upsilon_i|\mathbf{x}_i, a_i] \\ &= \mathbb{E}[\upsilon_i|a_i], \end{aligned} \tag{3.3}$$

where the second equality holds because $(\mathbf{x}_{-i}, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\ldots,N})$ and $(\mathbf{x}_i, a_i, \upsilon_i)$ are independent under Assumptions 1 (i) and (ii). This shows υ_i and $(\mathbf{x}_{-i}, \mathbf{D}_N(\mathbf{x}_{2,-i}, \mathbf{a}_{-i}, \{u_{ij}\}_{i,j=1,\ldots,N}, \mathbf{x}_{2i}, a_i))$ are mean-independent conditioning on $(\mathbf{x}_i, a_i)$. The last line follows by the fixed effect assumption, Assumption 1 (iii).

The result (3.3) shows conditional on the unobserved heterogeneity a_i in the network formation, the unobserved characteristic υ_i that affects the outcome y_i becomes uncorrelated with the network $\mathbf{D}_N$ (and the observables $\mathbf{X}_N$). This implies that the network endogeneity can be controlled by a_i. We summarize the discussion above in the following lemma:

Lemma 1 (Control Function of Peer Group Endogeneity)**.** *Suppose that Assumption 1 holds. Then,*

$$\mathbb{E}[\upsilon_i|\mathbf{X}_N, \mathbf{D}_N, a_i] = \mathbb{E}[\upsilon_i|a_i].$$

3.3. **Identification of Peer Effects.** In this section we show how to identify the peer effects in the outcome question when the endogenous network is formed by (3.1). We provide two identification methods depending on whether we control the network (peer group) endogeneity with a_i or together with $\mathbf{x}_{2i}$, in the case when $\mathbf{x}_{2i}$ and $\mathbf{x}_{1i}$ do not overlap.

First notice that regardless of the possible endogeneity of the network $\mathbf{G}_N$, we need to control for the endogeneity of the term $\sum_{\substack{j=1\\j\neq i}}^N g_{ij,N} y_j$ that represents so-called the endogenous peer effects. When the peer group $\mathbf{G}_N$ is exogenous and uncorrelated with υ_N, $\mathbf{G}_N^2\mathbf{X}_{1N}$ is often used as IVs for the endogenous peer effects term $\mathbf{G}_N\mathbf{y}_N$ (See, for example, Kelejian and Prucha (1998), Lee (2003), Bramoullé et al. (2009).).

Let $\mathbf{Z}_N = [\mathbf{X}_{1N}, \mathbf{G}_N\mathbf{X}_{1N}, \mathbf{G}_N^2\mathbf{X}_{1N}]$ be the usual IV matrix used in 2SLS estimation of the peer effects equation. Note that $\mathbf{Z}_N$ is not a valid IV matrix anymore in our framework because the peer group defined by the network $\mathbf{G}_N$ is correlated with υ_N due to potential correlation between two unobserved characteristics υ_i and a_i, Let $\mathbf{W}_N = [\mathbf{G}_N\mathbf{y}_N, \mathbf{X}_{1N}, \mathbf{G}_N\mathbf{X}_{1N}]$. Further, denote the transpose of the ith row of $\mathbf{Z}_N$ and $\mathbf{W}_N$ by $\mathbf{z}_i$ and $\mathbf{w}_i$, respectively.

Suppose that Assumption 1 holds and so a_i controls the network endogeneity. Then,

$$\begin{aligned}
\mathbb{E}\left[\,(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i])\,(\upsilon_i - \mathbb{E}(\upsilon_i|a_i))\mid a_i\right] &= \mathbb{E}[\mathbf{z}_i\upsilon_i \mid a_i] - \mathbb{E}[\mathbf{z}_i \mid a_i]\mathbb{E}[\upsilon_i \mid a_i] \\
&= \mathbb{E}\left[\mathbb{E}[\mathbf{z}_i\upsilon_i \mid a_i, \mathbf{X}_{1N}, \mathbf{G}_N] \mid a_i\right] - \mathbb{E}[\mathbf{z}_i|a_i]\mathbb{E}[\upsilon_i \mid a_i] \\
&= \mathbb{E}\left[\mathbf{z}_i\mathbb{E}[\upsilon_i \mid a_i, \mathbf{X}_{1N}, \mathbf{G}_N] \mid a_i\right] - \mathbb{E}[\mathbf{z}_i \mid a_i]\mathbb{E}[\upsilon_i \mid a_i] \\
&\overset{(1)}{=} \mathbb{E}\left[\mathbf{z}_i\mathbb{E}[\upsilon_i \mid a_i] \mid a_i\right] - \mathbb{E}[\mathbf{z}_i \mid a_i]\mathbb{E}[\upsilon_i \mid a_i] \\
&= 0, \qquad (3.4)
\end{aligned}$$

where equality (1) holds by Lemma 1(a). This shows that the instrumental variables $\mathbf{z}_i$ or $\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i]$ become orthogonal to $\upsilon_i - \mathbb{E}[\upsilon_i|a_i]$, the residual of υ_i after projecting out a_i.

Furthermore, if $\mathbb{E}\left[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i|a_i])'\right]$ has full rank, then we can identify the peer effect coefficients β^0 as

$$\begin{aligned} 0 &= \mathbb{E}\left[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i])(y_i - \mathbf{w}_i'\beta - \mathbb{E}[y_i - \mathbf{w}_i'\beta|a_i])\right] \\ &= \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i|a_i])'](\beta - \beta^0) + \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i])(\upsilon_i - \mathbb{E}[\upsilon_i|a_i])] \\ &\overset{(1)}{=} \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i|a_i])'](\beta - \beta^0) \\ &\overset{(2)}{\Leftrightarrow} \beta = \beta^0, \end{aligned}$$

where equality (1) follows by the orthogonality result in (3.4) and equality (2) follows from the full rank condition.

Assumption 2 (Rank condition). $\mathbb{E}\left[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i|a_i])'\right]$ *has full rank.*

For the full rank condition in Assumption 2, it is necessary that the IVs $\mathbf{z}_i$ and the regressors $\mathbf{w}_i$ have additional variation after projecting out the control function a_i. As shown in the Supplementary Appendix S.1.3, when N is large, both $\mathbf{z}_i$ and $\mathbf{w}_i$ become close to functions that depend only on $(\mathbf{x}_i, a_i)$. In this case, for the full rank condition to be satisfied, it is necessary that there are additional random components in $\mathbf{x}_i$ that are different from a_i, so that the limits of $\mathbf{z}_i$ and $\mathbf{w}_i$ are not linearly dependent.

As a summary, we have the following first identification theorem.

Theorem 3.1 (Identification). *Under Assumptions 1 and 2, the parameter β^0 is identified by the moment condition $\mathbb{E}[(\mathbf{z}_i - \mathbb{E}(\mathbf{z}_i|a_i))(y_i - \mathbb{E}(y_i|a_i) - (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i|a_i))'\beta^0)] = 0$:*

$$\mathbb{E}[(\mathbf{z}_i - \mathbb{E}(\mathbf{z}_i|a_i))(y_i - \mathbb{E}(y_i|a_i) - (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i|a_i))'\beta)] = 0 \iff \beta = \beta^0.$$

Theorem 3.1 shows that we can identify parameter β^0 by controlling the unobserved network heterogeneity a_i in the outcome equation and taking the residuals $y_i - \mathbb{E}(y_i|a_i) - (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i|a_i))'\beta$, and then by using the instrumental variables $\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i]$.

3.3.1. *Alternative Identification.* In view of the derivation of the control function in (3.3) under Assumption 1, it is possible to use any regressors in $\mathbf{x}_i$ in addition to the unobserved heterogeneity a_i. In this section, we discuss identification of the peer effects when we use $(\mathbf{x}_{2i}, a_i)$ as control function. The reason to consider this particular control function is that we can implement this control function differently from the control function that uses a_i only, which will be discussed in detail in Section 4.

First, suppose that there is no overlap between the regressors in the outcome equation, $\mathbf{x}_{1i}$, and the network formation equation, $\mathbf{x}_{2i}$ and assume the conditions in Assumption 1.

Assumption 3. *Assume that the conditions (i),(ii), and (iii) of Assumption 1 hold. Also, assume that (iv)* $\mathbf{x}_{1i} \cap \mathbf{x}_{2i} = \emptyset$.

Later in this section, we will discuss a more general case where $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ intersect. Then, under Assumption 1 and (3.3), it follows that

$$\mathbb{E}[\upsilon_i|\mathbf{X}_N, \mathbf{D}_N, a_i] = \mathbb{E}[\upsilon_i|a_i] = \mathbb{E}[\upsilon_i|\mathbf{x}_{2i}, a_i], \tag{3.5}$$

where the last line holds by Assumption 1(iii). Then, similar to (3.4), we can show that

$$\begin{aligned}
& \mathbb{E}\left[\, (\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])\,(\upsilon_i - \mathbb{E}(\upsilon_i|\mathbf{x}_{2i}, a_i)) \mid \mathbf{x}_{2i}, a_i\right] \\
= \;& \mathbb{E}[\mathbf{z}_i\upsilon_i|\mathbf{x}_{2i}, a_i] - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i]\mathbb{E}[\upsilon_i|\mathbf{x}_{2i}, a_i] \\
= \;& \mathbb{E}\left[\mathbf{z}_i\mathbb{E}[\upsilon_i|a_i, \mathbf{X}_{1N}, \mathbf{G}_N]|\mathbf{x}_{2i}, a_i\right] - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i]\mathbb{E}[\upsilon_i|\mathbf{x}_{2i}, a_i] \\
\overset{(1)}{=} \;& \mathbb{E}\left[\mathbf{z}_i\mathbb{E}[\upsilon_i|\mathbf{x}_{2i}, a_i]|\mathbf{x}_{2i}, a_i\right] - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i]\mathbb{E}[\upsilon_i|\mathbf{x}_{2i}, a_i] \\
= \;& 0,
\end{aligned} \tag{3.6}$$

where equality (1) holds by (3.5). Furthermore, suppose that the following full rank assumption is satisfied:

Assumption 4 (Rank condition). $\mathbb{E}\left[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i|\mathbf{x}_{2i}, a_i])'\right]$ *has full rank.*

Notice that if $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ are overlapped, then the full rank condition in Assumption 4 does not hold.

Using similar arguments that lead to Theorem 3.1, we can identify the peer effect coefficients β^0 as

$$\begin{aligned}
0 &= \mathbb{E}\left[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])(y_i - \mathbf{w}_i'\beta - \mathbb{E}[y_i - \mathbf{w}_i'\beta|\mathbf{x}_{2i}, a_i])\right] \\
&= \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i|\mathbf{x}_{2i}, a_i])'](\beta - \beta^0) + \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])(v_i - \mathbb{E}[v_i|\mathbf{x}_{2i}, a_i])] \\
&\overset{(1)}{=} \mathbb{E}[(\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, a_i])(\mathbf{w}_i - \mathbb{E}[\mathbf{w}_i|\mathbf{x}_{2i}, a_i])'](\beta - \beta^0) \\
&\overset{(2)}{\Leftrightarrow} \beta = \beta^0,
\end{aligned} \tag{3.7}$$

where equality (1) follows by the orthogonality result in (3.6) and equality (2) follows from the full rank condition in Assumption 4. This is summarized in the following theorem.

Theorem 3.2 (Alternative Identification). *Under Assumptions 1, 3, and 4, the parameter β^0 is identified by the moment condition*
$\mathbb{E}[(\mathbf{z}_i - \mathbb{E}(\mathbf{z}_i|\mathbf{x}_{2i}, a_i))((y_i - \mathbb{E}(y_i|\mathbf{x}_{2i}, a_i) - (\mathbf{w}_i' - \mathbb{E}(\mathbf{w}_i|\mathbf{x}_{2i}, a_i))'\beta^0] = 0$*:*

$$\mathbb{E}[(\mathbf{z}_i - \mathbb{E}(\mathbf{z}_i|\mathbf{x}_{2i}, a_i))((y_i - \mathbb{E}(y_i|\mathbf{x}_{2i}, a_i) - (\mathbf{w}_i' - \mathbb{E}(\mathbf{w}_i|\mathbf{x}_{2i}, a_i))'\beta] = 0 \quad \Longleftrightarrow \quad \beta = \beta^0.$$

So far, we have considered the case where the regressors $\mathbf{x}_{i1}$ and $\mathbf{x}_{2i}$ do not intersect. A more general case is when the regressors $\mathbf{x}_{1i}$ consist of two components, where one component is different from the observed control function $\mathbf{x}_{2i}$ and the other is a part of $\mathbf{x}_{2i}$. That is, $\mathbf{x}_{1i} = (\mathbf{x}_{11i}, \mathbf{x}_{12i})$, where $\mathbf{x}_{11i}$ does not share any elements with $\mathbf{x}_{2i}$ and $\mathbf{x}_{11i}$ is nonempty, and $\mathbf{x}_{12i} \subset \mathbf{x}_{2i}$. Let $\beta_2^0 = (\beta_{21}^0, \beta_{22}^0), \beta_3^0 = (\beta_{31}^0, \beta_{32}^0)$ conformable to the dimensions of $(\mathbf{x}_{11i}, \mathbf{x}_{12i})$. Similarly let $\beta_2 = (\beta_{21}, \beta_{22}), \beta_3 = (\beta_{31}, \beta_{32})$.

In this case, with a properly modified rank condition of $\mathbf{z}_{(2),i}$ and $\mathbf{w}_{(2),i}$ that excludes the variables associated with $\mathbf{x}_{12,i}$ and $\sum_{j=1,\neq i}^{N} g_{ij,N}\mathbf{x}_{12,j}$, we can identify the coefficients $\beta_{(2)}^0 :=$

$(\beta_1^0, \beta_{21}^0, \beta_{31}^0)$ using the same argument that leads to the identification in (3.7). However, we cannot identify the coefficients that correspond to the variable $\mathbf{x}_{12,i}$ and $\sum_{j=1,\neq i}^N g_{ij,N}\mathbf{x}_{12,j}$. The reason is that the controlling the network endogeneity with the control variable $(\mathbf{x}_{2i}, a_i)$ wipes out the information in $(\mathbf{x}_{12,i}, \sum_{j=1,\neq i}^N g_{ij,N}\mathbf{x}_{12,j})$:

$$\mathbf{x}_{12,i} - \mathbb{E}[\mathbf{x}_{12,i}|\mathbf{x}_{2i}, a_i] = 0$$

$$\sum_{j=1,\neq i}^N g_{ij,N}\mathbf{x}_{12,j} - \mathbb{E}\left[\sum_{j=1,\neq i}^N g_{ij,N}\mathbf{x}_{12,j}\,|\mathbf{x}_{2i}, a_i\right] \to_p 0,$$

where the second convergence holds because $\sum_{j=1,\neq i}^N g_{ij,N}\mathbf{x}_{12,j}$ converges to a function that depends only on $\mathbf{x}_{2i}, a_i$ (see Section S.1.3 in the Supplementary Appendix.).

Through the rest of the paper, when we consider $(\mathbf{x}_{2i}, a_i)$ as control function, we will without loss of generality apply the restriction in Assumption 3 that $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ do not overlap.

4. Estimation

In this section we present two estimation methods. In subsections 4.1 and 4.2 we discuss estimation using a_i and $(\mathbf{x}_{2i}, a_i)$ as control function, respectively.

4.1. **With a_i as Control Function.**

The identification scheme of Theorem 3.1 identifies the parameter of interest β^0 with the two step procedure: (i) control a_i in the outcome equation and yield $y_i - \mathbb{E}(y_i|a_i) = (\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i|a_i))'\beta^0 + \upsilon_i - \mathbb{E}(\upsilon_i)$, and then (ii) use $\mathbf{z}_i - \mathbb{E}[\mathbf{z}_i|a_i]$ as IVs for $\mathbf{w}_i - \mathbb{E}(\mathbf{w}_i|a_i)$. Let $\mathbf{h}(a_i) = (h^y(a_i), \mathbf{h}^w(a_i), \mathbf{h}^z(a_i)) := (\mathbb{E}[y_i|a_i], \mathbb{E}[\mathbf{w}_i|a_i], \mathbb{E}[\mathbf{z}_i|a_i])$.

Let $\widetilde{\mathbf{W}}_N = (\mathbf{w}_1 - \mathbf{h}^w(a_1), ..., \mathbf{w}_N - \mathbf{h}^w(a_N))'$. Similarly we define $\widetilde{\mathbf{Z}}_N, \tilde{\mathbf{y}}_N$. Suppose that we observe $\mathbf{h}(a_i)$. In view of the identification scheme of Theorem 3.1, we can estimate β^0 by

$$\hat{\beta}_{2SLS}^{\text{inf}} = \left(\widetilde{\mathbf{W}}_N'\widetilde{\mathbf{Z}}_N\left(\widetilde{\mathbf{Z}}_N'\widetilde{\mathbf{Z}}_N\right)^{-1}\widetilde{\mathbf{Z}}_N'\widetilde{\mathbf{W}}_N\right)^{-1}\widetilde{\mathbf{W}}_N'\widetilde{\mathbf{Z}}_N\left(\widetilde{\mathbf{Z}}_N'\widetilde{\mathbf{Z}}_N\right)^{-1}\widetilde{\mathbf{Z}}_N'\tilde{\mathbf{y}}_N.$$

However, since the individual heterogeneity a_i is, not observed and the functions $\mathbf{h}(\cdot)$ are not known, the estimator $\hat{\beta}_{2SLS}^{\text{inf}}$ is not feasible. A natural implementation of the infeasible estimator $\hat{\beta}_{2SLS}^{\text{inf}}$ is to replace $\mathbf{h}(a_i)$ in $\widetilde{\mathbf{W}}_N, \widetilde{\mathbf{Z}}_N$, and $\tilde{\mathbf{y}}_N$ with its estimate, say $\hat{\mathbf{h}}(\hat{a}_i)$.

Notice that if a_i is observed, we can estimate $\mathbf{h}(\cdot)$ using various nonparametric methods. In this paper we consider a sieve estimation method.[3] Suppose that $h^l(a)$ is the l^{th} element in $\mathbf{h}(a)$ for $l = 1, ..., L$, where L is the dimension of $(y_i, \mathbf{w}_i', \mathbf{z}_i')'$. The sieve estimation method assumes that each function $h^l(a)$, $l = 1, ..., L$ is well approximated by a linear combination of base functions $(q_1(a), ..., q_{K_N}(a))$:

$$h^l(a) \cong \sum_{k=1}^{K_N} q_k(a)\alpha_k^l, \tag{4.1}$$

as the truncation parameter $K_N \to \infty$.

Let $\mathbf{q}^K(a) = (q_1(a), \ldots, q_K(a))'$, $\mathbf{Q}_N := \mathbf{Q}_N(\mathbf{a}_N) = (q^K(a_1), \ldots, q^K(a_N))'$, $\mathbf{h}^l(\mathbf{a}_N) = (h^l(a_1), \ldots, h^l(a_N))'$, and $\boldsymbol{\alpha}_N^l = (\alpha_1^l, \ldots, \alpha_{K_N}^l)'$. Let b_i^l be the l^{th} element in $(y_i, \mathbf{w}_i', \mathbf{z}_i')'$ and denote $\mathbf{b}_N^l = (b_1^l, ..., b_N^l)$.

If $\mathbf{a}_N = (a_1, ..., a_N)'$ is observed, in view of (4.1), we can estimate the unknown function $\mathbf{h}^l(\mathbf{a}_N)$ by the OLS of b_i^l on $\mathbf{q}^K(a_i)$: for $l = 1, ..., L$,

$$\hat{\mathbf{h}}^l(\mathbf{a}_N) = \mathbf{P}_{\mathbf{Q}_N}\mathbf{b}_N^l, \tag{4.2}$$

where $\mathbf{P}_{\mathbf{Q}_N} = \mathbf{Q}_N(\mathbf{Q}_N'\mathbf{Q}_N)^{-}\mathbf{Q}_N'$. Here $^-$ denotes any symmetric generalized inverse.

Given this, we suggest to estimate $\mathbf{h}^l(\mathbf{a}_N)$ as follows: (i) first, we estimate the unobserved individual heterogeneity and then (ii) plug the estimate in $\hat{\mathbf{h}}^l(\mathbf{a}_N)$ of (4.2). To be more specific, suppose $\hat{\mathbf{a}}_N = (\hat{a}_1, ..., \hat{a}_N)'$ is an estimator of $\mathbf{a}_N = (a_1, ..., a_N)'$. Denote $\hat{\mathbf{Q}}_N := \mathbf{Q}_N(\hat{\mathbf{a}}_N) = (\mathbf{q}^{K_N}(\hat{a}_1), \ldots, \mathbf{q}^{K_N}(\hat{a}_N))'$. Then the first estimator of $\mathbf{h}^l(\mathbf{a}_N)$ is defined by

$$\hat{\mathbf{h}}^l := \hat{\mathbf{h}}^l(\hat{\mathbf{a}}_N) = \mathbf{P}_{\hat{\mathbf{Q}}_N}\mathbf{b}_N^l \tag{4.3}$$

for $l = 1, ..., L$, and this leads the following estimator of β^0:

$$\begin{aligned} \hat{\beta}_{2SLS} &= \left(\mathbf{W}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\left(\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{W}_N\right)^{-1} \\ &\quad \times\mathbf{W}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\left(\mathbf{Z}'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{y}_N, \end{aligned} \tag{4.4}$$

[3] In principle we can use other nonparametric estimation methods such as kernel smoothing or local polynomial methods.

where $\mathbf{M}_{\widehat{\mathbf{Q}}_N} = I_N - \mathbf{P}_{\widehat{\mathbf{Q}}_N}$.

A desired estimator of a_i should satisfy the following high level condition.

Assumption 5 (Estimation of a_i)**.** *We assume that we can estimate a_i with $\hat{a}_i$ such that $\max_i |\hat{a}_i - a_i| = O_p\left(\zeta_a(N)^{-1}\right)$, where $\zeta_a(N) \to \infty$ as $N \to \infty$, satisfying Assumption 8 in the Appendix.*

For the purpose of our paper, any estimation method that yields the estimator $\hat{a}_i$ satisfying the restriction in Assumption 5 can be adopted. For example, assuming the parametric specification (3.2),

$$d_{ij,N} = \mathbb{I}(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})'\lambda + a_i + a_j \geq u_{ij}),$$

with regularity conditions of Assumption 6 in the Appendix, Graham (2017) shows that the joint maximum likelihood estimator that solves

$$(\hat{a}_1, ..., \hat{a}_N) := \underset{\lambda, \mathbf{a}_N}{\operatorname{argmax}} \left(\sum_{i=1}^{N} \sum_{j<i} d_{ij,N} \exp\left(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})'\lambda + a_i + a_j\right) - \ln\left[1 + \exp(t(\mathbf{x}_{2i}, \mathbf{x}_{2j})'\lambda + a_i + a_j)\right] \right) \tag{4.5}$$

satisfies

$$\sup_{1 \leq i \leq N} |\hat{a}_i - a_i| < O\left(\sqrt{\frac{\ln N}{N}}\right).$$

with probability $1 - O(N^{-2})$. In this case we have $\zeta_a(N) = \sqrt{\frac{N}{\ln N}}$. Examples of other estimation methods include Fernández-Val and Weidner (2013), Jochmans (2016), Dzemski (2017), and Jochmans (2017).

4.2. **With $(\mathbf{x}_{2i}, a_i)$ as Control Function.** Suppose that $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ do not intersect as assumed in Assumption 3. The idea of the second approach is to implement a control function with the node degree of the network together with the regressors $\mathbf{x}_{2i}$.

The degree of node (or individual) i is the number of connections with node (individual) i in the network. Let $\widehat{\deg}_i$ be the degree of node i scaled by the network size:

$$\widehat{\deg}_i := \frac{1}{N-1} \sum_{j=1, \neq i}^{N} d_{ij,N}.$$

Recall that the link $d_{ij,N}$ is formed by

$$d_{ij,N} = \mathbb{I}(g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) - u_{ij} \geq 0).$$

Recall that the unobserved link-specific error terms u_{ij} are assumed to be independent of all the other variables and randomly drawn. Let $\Phi(\cdot)$ be the cdf of u_{ij}. Also let $\pi(\mathbf{x}_2, a)$ be the joint density funciton of $(\mathbf{x}_{2i}, a_i)$. Then, for each $(\mathbf{x}_{2i}, a_i)$, by the WLLN conditioning on $(\mathbf{x}_{2i}, a_i)$, we have

$$\begin{aligned} \widehat{\deg}_i &:= \frac{1}{N-1} \sum_{j=1, \neq i}^{N} \mathbb{I}(g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) - u_{ij} \geq 0) \\ &\rightarrow_p \int \Phi\left(g(t(\mathbf{x}_{2i}, \mathbf{x}_2), a_i, a)\right) \pi(\mathbf{x}_2, a) d\mathbf{x}_2 da \\ &=: \deg(\mathbf{x}_{2i}, a_i) \\ &=: \deg_i \end{aligned} \tag{4.6}$$

as the network size N grows to infinity. Furthermore, we can show that under the regularity conditions in Assumption 11 in the Appendix, $\sup_i \mathbb{E}[(\sqrt{N}(\widehat{\deg}_i - \deg_i))^{2B}] < \infty$ for any finite integer $B \geq 2$, from which we can deduce that

$$\max_{1 \leq i \leq N} |\widehat{\deg}_i - \deg_i| = O_p\left(\zeta_{deg}(N)^{-1}\right), \tag{4.7}$$

where

$$\zeta_{deg}(N) := o(1) N^{\frac{B-1}{2B}}.$$

This corresponds to the condition in Assumption 5.

Suppose that u_{ij} is a continuous random variable so that $\Phi(\cdot)$ is a strictly increasing function. Then, since the function g is a strictly monotonic function of a_i, the function $\deg(\mathbf{x}_{2i}, a_i)$ is a strictly monotonic function of a_i, too, for each $\mathbf{x}_{2i}$.

Let b_i^l be the l^{th} element in $(y_i, \mathbf{w}_i', \mathbf{z}_i')'$. The strict monotonicity of $\deg_i = \deg(\mathbf{x}_{2i}, a_i)$ in a_i for each $\mathbf{x}_{2i}$ implies that $(\mathbf{x}_{2i}, a_i)$ and $(\mathbf{x}_{2i}, \deg_i)$ have an one-to-one relation, and therefore we have

$$h^l_*(\mathbf{x}_{2i}, a_i) := \mathbb{E}[b^l_i|\mathbf{x}_{2i}, a_i] = \mathbb{E}[b^l_i|\mathbf{x}_{2i}, \deg_i] =: h^l_{**}(\mathbf{x}_{2i}, \deg_i).$$

Suppose that the function $h^l_*(\mathbf{x}_{2i}, \deg_i), l = 1, ..., L$ is well approximated by a linear combination of base functions $(r_1(\mathbf{x}_2, \deg_i), ..., r_K(\mathbf{x}_2, \deg_i))$:

$$h^l_{**}(\mathbf{x}_{2i}, \deg_i) \cong \sum_{k=1}^{K_N} r_k(\mathbf{x}_2, \deg_i)\gamma^l_k$$

as the truncation parameter $K_N \to \infty$.

Let $\mathbf{Deg}_N = (\deg_1, ..., \deg_N)'$. Let $\mathbf{r}^K(\mathbf{x}_{2i}, \deg_i) = (r_1(\mathbf{x}_{2i}, \deg_i), \ldots, r_K(\mathbf{x}_{2i}, \deg_i))'$, $\mathbf{R}_N := \mathbf{R}_N(\mathbf{X}_{2N}, \mathbf{Deg}_N) = (\mathbf{r}^K(\mathbf{x}_{21}, \deg_1), \ldots, \mathbf{r}^K(\mathbf{x}_{2N}, \deg_N))'$, and $\boldsymbol{\gamma}^l = (\gamma^l_1, \ldots, \gamma^l_{K_N})'$. Let $\mathbf{b}^l_N = (b^l_1, ..., b^l_N)$. In the case where $(\mathbf{x}_{2i}, \deg_i)$ are observed, we can estimate $\mathbf{h}^l_{**}(\mathbf{X}_{2N}, \mathbf{Deg}_N) = (h^l_{**}(\mathbf{x}_{2,1}, \deg_1), ..., h^l_{**}(\mathbf{x}_{2,N}, \deg_N)$ for $l = 1, .., L$ with

$$\widehat{\mathbf{h}}^l_{**}(\mathbf{X}_{2N}, \mathbf{Deg}_N) := \mathbf{P}_{\mathbf{R}_N}\mathbf{b}^l_N, \tag{4.8}$$

where $\mathbf{P}_{\mathbf{R}_N} = \mathbf{R}_N(\mathbf{R}'_N\mathbf{R}_N)^-\mathbf{R}'_N$. Here $^-$ denotes any symmetric generalized inverse.

In view of (4.6), the natural estimator of $\deg_i$ is $\widehat{\deg}_i$. This suggests that we estimate $\hat{h}^l_{**}(\mathbf{x}_{2i}, \deg_i)$ by using $\widehat{\deg}_i$ in place of $\deg_i$. To be more specific, suppose that $\widehat{\mathbf{Deg}}_N = (\widehat{\deg}_1, ..., \widehat{\deg}_N)$. Denote $\widehat{\mathbf{R}}_N := \mathbf{R}_N(\mathbf{X}_{2N}, \widehat{\mathbf{Deg}}_N) = (\mathbf{r}^K(\mathbf{x}_{21}, \widehat{\deg}_1), ..., \mathbf{r}^K(\mathbf{x}_{2N}, \widehat{\deg}_N))'$. The estimator of $h^l_*(\mathbf{x}_{2i}, a_i) = h^l_{**}(\mathbf{x}_{2i}, \deg_i)$ is defined by the i^{th} element of

$$\widehat{\mathbf{h}}^l_*(\mathbf{X}_{2N}, \mathbf{a}_N) := \widehat{\mathbf{h}}^l_{**}(\mathbf{X}_{2N}, \widehat{\mathbf{Deg}}_N) = \mathbf{P}_{\widehat{\mathbf{R}}_N}\mathbf{b}^l_N.$$

Then, it leads the following second estimator of β^0:

$$\begin{aligned} \bar{\beta}_{2SLS} \;\; := \;\; & \left(\mathbf{W}'_N\mathbf{M}_{\widehat{\mathbf{R}}_N}\mathbf{Z}_N\left(\mathbf{Z}'_N\mathbf{M}_{\widehat{\mathbf{R}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}'_N\mathbf{M}_{\widehat{\mathbf{R}}_N}\mathbf{W}_N\right)^{-1} \\ & \times\mathbf{W}'_N\mathbf{M}_{\widehat{\mathbf{R}}_N}\mathbf{Z}_N\left(\mathbf{Z}'\mathbf{M}_{\widehat{\mathbf{R}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}'_N\mathbf{M}_{\widehat{\mathbf{R}}_N}\mathbf{y}_N, \end{aligned} \tag{4.9}$$

where $\mathbf{M}_{\widehat{\mathbf{R}}_N} = I_N - \mathbf{P}_{\widehat{\mathbf{R}}_N}$.

4.3. **Sieve Estimation.** In this subsection we discuss the sieve estimators used in estimating $\hat{\beta}_{2SLS}$ and $\bar{\beta}_{2SLS}$. For the regularity conditions of the sieve basis $\mathbf{Q}_N$ and $\mathbf{R}_N$, we impose standard conditions such as those proposed by Newey (1997) and Li and Racine (2007). These assumptions ensure that $\mathbf{Q}_N'\mathbf{Q}_N$ is asymptotically non-singular and control the rate of approximation of the sieve estimator. These assumptions are formally stated in Assumptions 7 and 9 of the Appendix.

Additionally, we require that the sieve basis satisfy a Lipschitz condition, which allows to control for the error introduced by the estimation of a_i with $\hat{a}_i$ in the estimation of $\hat{\beta}_{2SLS}$, and the estimation of deg_i with $\widehat{deg_i}$ in the estimation of $\bar{\beta}_{2SLS}$[4] (see Assumptions 8 and 10). As an example, define the polynomial sieve as follows. Let $Pol(K_N)$ denote the space of polynomials on $[-1,1]$ of degree K_N,

$$Pol(K_N) = \left\{ \nu_0 + \sum_{k=1}^{K_N} \nu_k a^k,\ a \in [-1,1], \nu_k \in \mathbb{R} \right\}.$$

For any k we have

$$\left| a_1^k - a_2^k \right| = k|\tilde{a}^k||a_1 - a_2| \leq Mk|a_1 - a_2|,$$

where $\tilde{a} \in [-1,1]$ and M is a finite constant.

In sieve estimations an important issue is how to choose the truncation parameter K_N. Well-known procedures for selecting K_N are Mallows' C_L, generalized cross-validation and leave-one-out cross-validation. For more on these methods see Chapter 15.2 in Li and Racine (2007), Li (1987), Wahba (1985), Andrews (1991) and Hansen (2014). However, these methods are applicable mostly when the observations are cross-sectionally independent, which is not true in our case, especially when the network is dense, as we assume. Developing a data-driven choice of K_N is beyond the scope of this paper and we leave it for future work.

[4]This issue is similar to the two step series estimation problem in Newey (2009). Other papers that investigated the problem of nonparametric or semiparametric analysis with generated regressors include Ahn and Powell (1993), Mammen et al. (2012), Hahn and Ridder (2013), and Escanciano et al. (2014), for example.

5. Limit Distribution and Standard Error

In this section we present the asymptotic distributions of the two 2SLS estimators $\hat{\beta}_{2SLS}$ and $\bar{\beta}_{2SLS}$, and show how to estimate standard errors. We also discuss key technical issues in deriving the limits. All details of the technical derivations and proofs can be found in the Appendix.

5.1. Limiting Distribution and Standard Error of $\hat{\beta}_{2SLS}$.

Recall the definitions $h^y(a_i) := \mathbb{E}[y_i|a_i]$, $h^\upsilon(a_i) := \mathbb{E}[\upsilon_i|a_i]$, $\mathbf{h}^{\mathbf{w}}(a_i) := \mathbb{E}(\mathbf{w}_i|a_i)$, $\mathbf{h}^{\mathbf{z}}(a_i) := \mathbb{E}(\mathbf{z}_i|a_i)$. Define $\eta_i^y := y_i - h^y(a_i)$, $\eta_i^\upsilon := \upsilon_i - h^\upsilon(a_i)$, $\eta_i^{\mathbf{w}} = \mathbf{w}_i - \mathbf{h}^{\mathbf{w}}(a_i)$, $\eta_i^{\mathbf{z}} = \mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i)$. Let $\boldsymbol{\eta}_N^\upsilon = (\eta_1^\upsilon, ..., \eta_N^\upsilon)'$ and $\mathbf{H}_N^\upsilon(\mathbf{a}_N) = (h^\upsilon(a_1), ..., h^\upsilon(a_N))'$. Let $\hat{h}^\upsilon(a_i)$, $\hat{\mathbf{h}}^{\mathbf{w}}(a_i)$, and $\hat{\mathbf{h}}^{\mathbf{z}}(a_i)$ denote the sieve estimators of $h^\upsilon(a_i)$, $h^{\mathbf{w}}(a_i)$ and $h^{\mathbf{z}}(a_i)$, respectively.

The derivation of the asymptotic distribution of $\hat{\beta}_{2SLS}$ consists of three steps. First, we show that the sampling error caused by the use of $\hat{\mathbf{a}}_N$ instead of $\mathbf{a}_N$ is asymptotically negligible (see Lemma 2 of the Supplementary Appendix S.1.1.). Next, we control the error introduced by the non-parametric estimation of $h^l(a_i)$, where $l \in \{\upsilon, \mathbf{w}, \mathbf{z}\}$. In Lemma 7 of Supplementary Appendix S.1.2 we show that under the regularity conditions, the estimation error in $\hat{h}^l(a_i)$ vanishes at a suitable rate. Combining these two, we deduce

$$\sqrt{N}(\hat{\beta}_{2SLS} - \hat{\beta}_{2SLS}^{\inf}) = o_p(1).$$

The last step is to derive the limiting distribution of the infeasible estimator $\sqrt{N}(\hat{\beta}_{2SLS}^{\inf} - \beta^0)$. In the Supplementary Appendix S.1.3 we show the following:

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}_i - \mathbf{h}^{\mathbf{w}}(a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))' \xrightarrow{p} \mathbf{S}^{\mathbf{wz}} \tag{5.1}$$

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))' \xrightarrow{p} \mathbf{S}^{\mathbf{zz}} \tag{5.2}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))\eta_i^\upsilon \Rightarrow \mathcal{N}(0, \mathbf{S}^{\mathbf{zz}\sigma}), \tag{5.3}$$

where the closed forms of the limits $\mathbf{S}^{\mathbf{wz}}$ and $\mathbf{S}^{\mathbf{zz}}$ are found in Lemma 11 and $\mathbf{S}^{\mathbf{zz}\sigma}$ in Lemma 12 of Supplementary Appendix.

Notice that the derivation of the limiting distribution in (5.3) allows $\eta_i^{\upsilon} = \upsilon_i - \mathbb{E}(\upsilon_i|a_i)$ to be conditionally heteroskedastic, and so $\sigma^2(\mathbf{x}_i, a_i) := \mathbb{E}[(\upsilon_i - \mathbb{E}[\upsilon_i|a_i])^2|\mathbf{x}_i, a_i]$ is allowed to depend on $(\mathbf{x}_i, a_i)$.

Combining all the limit results deduce the following theorem.

Theorem 5.1 (Limiting Distribution)**.** *Suppose that Assumptions 1, 2, 5, 7, 8, and 11(i)-(v) in the Appendix hold. Then, we have*

$$\sqrt{N}(\hat{\beta}_{2SLS} - \beta^0) \Rightarrow \mathcal{N}\left(0, \Omega\right),$$

where

$$\Omega = \left(\mathbf{S}^{\mathbf{wz}}\left(\mathbf{S}^{\mathbf{zz}}\right)^{-1}\left(\mathbf{S}^{\mathbf{wz}}\right)'\right)^{-1}\left(\mathbf{S}^{\mathbf{wz}}\left(\mathbf{S}^{\mathbf{zz}}\right)^{-1}\mathbf{S}^{\mathbf{zz}\sigma}\left(\mathbf{S}^{\mathbf{zz}}\right)^{-1}\left(\mathbf{S}^{\mathbf{wz}}\right)'\right)\left(\mathbf{S}^{\mathbf{wz}}\left(\mathbf{S}^{\mathbf{zz}}\right)^{-1}\left(\mathbf{S}^{\mathbf{wz}}\right)'\right)^{-1}. \quad (5.4)$$

The theorem requires several regularity conditions. Assumption 1 requires that $(y_i, \mathbf{x}_i, a_i)$ be randomly drawn and Assumption 2 is a full rank condition. Assumptions 5, 7 and 8 ensure that a_i can be consistently estimated, and that the error between $\mathbf{h}(a_i)$ and $\hat{\mathbf{h}}^l(\hat{a}_i)$ converges to zero at a suitable rate. Assumption 11 imposes further restrictions on the outcome model (2.1) and the network formation model (3.1). It requires that $|\beta_1^0|$ be bounded below 1 so that the spillover effect has a unique solution, and that $\|\beta_2^0\|$ be bounded above 0 so that the IVs are strong. It also assumes that observables $(y_i, \mathbf{x}_i)$ and $\mathbf{t}_{ij}$ are bounded, and a_i has a compact support in $[-1, 1]$. These boundedness conditions are required as a technical regularity condition in deriving the limits in (5.1), (5.2), and (5.3), which involves some uniformity in the limit.

The asymptotic variance can be consistently estimated by

$$\widehat{\Omega} = \left(\widehat{\mathbf{S}}^{\mathbf{wz}}\left(\widehat{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})'\right)^{-1}\left(\widehat{\mathbf{S}}^{\mathbf{wz}}\left(\widehat{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}\widehat{\mathbf{S}}^{\mathbf{zz}\sigma}\left(\widehat{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})'\right)\left(\widehat{\mathbf{S}}^{\mathbf{wz}}\left(\widehat{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}(\widehat{\mathbf{S}}^{\mathbf{wz}})'\right)^{-1}, \quad (5.5)$$

where

$$\hat{\mathbf{S}}^{\mathbf{wz}} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{w}_i - \hat{\mathbf{h}}^{\mathbf{w}}(\hat{a}_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(\hat{a}_i)\right)'$$

$$\hat{\mathbf{S}}^{\mathbf{zz}} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(\hat{a}_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(\hat{a}_i)\right)'$$

$$\hat{\mathbf{S}}^{ZZ\sigma^2} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(\hat{a}_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(\hat{a}_i)\right)'(\hat{\eta}_i^v)^2,$$

and $\hat{\eta}_i^v = y_i - \hat{h}^y(\hat{a}_i) - (\mathbf{w}_i - \hat{\mathbf{h}}^{\mathbf{w}}(\hat{a}_i))'\hat{\beta}_{2SLS}$.

5.2. **Limiting Distribution and Standard Error of $\bar{\beta}_{2SLS}$.** The process is analogous to the one presented in the previous section. Again, let b_i^l be the l^{th} element in $(y_i, \mathbf{w}_i', \mathbf{z}_i')'$. Recall the definition that

$$h_*^l(\mathbf{x}_{2i}, a_i) = \mathbb{E}[b_i^l|\mathbf{x}_{2i}, a_i] = \mathbb{E}[b_i^l|\mathbf{x}_{2i}, \deg_i] =: h_{**}^l(\mathbf{x}_{2i}, \deg_i).$$

Further, let $\eta_{*i}^l = b_i^l - h_*^l(\mathbf{x}_{2i}, a_i) = b^l - h_{**}^l(\mathbf{x}_{2i}, \deg_i)$, and let $\hat{h}_{**}^l(\mathbf{x}_{2i}, \deg_i)$ denote a sieve estimator of $h_{**}^l(\mathbf{x}_{2i}, \deg_i)$.

As in the previous section, we derive the asymptotic distribution of $\bar{\beta}_{2SLS}$ in three steps. First, we show that the error that stems from the use of the estimate $\widehat{\deg}_i$ for $\deg_i$, $\hat{h}_{**}^l(\mathbf{x}_{2i}, \widehat{\deg}_i) - \hat{h}_{**}^l(\mathbf{x}_{2i}, \deg_i)$, is asymptotically negligible. In the second step, we control the error introduced by the non-parametric estimation of $h_{**}^l(\mathbf{x}_{2i}, \deg_i)$, $\hat{h}_{**}^l(\mathbf{x}_{2i}, \deg_i) - h_{**}^l(\mathbf{x}_{2i}, \deg_i)$. This implies

$$\sqrt{N}(\bar{\beta}_{2SLS} - \bar{\beta}_{2SLS}^{\inf}) = o_p(1),$$

where

$$\bar{\beta}_{2SLS}^{\inf} = \left(\widetilde{\mathbf{W}}_{*N}'\widetilde{\mathbf{Z}}_{*N}\left(\widetilde{\mathbf{Z}}_{*N}'\widetilde{\mathbf{Z}}_{*N}\right)^{-1}\widetilde{\mathbf{Z}}_{*N}'\widetilde{\mathbf{W}}_{*N}\right)^{-1}\widetilde{\mathbf{W}}_{*N}'\widetilde{\mathbf{Z}}_{*N}\left(\widetilde{\mathbf{Z}}_{*N}'\widetilde{\mathbf{Z}}_{*N}\right)^{-1}\widetilde{\mathbf{Z}}_{*N}'\tilde{\mathbf{y}}_{*N}$$

and $\widetilde{\mathbf{W}}_{*N} = (\mathbf{w}_1 - \mathbf{h}_*^w(\mathbf{x}_{21}, a_1), ..., \mathbf{w}_N - \mathbf{h}_*^w(\mathbf{x}_{2N}, a_N))'$ and $\widetilde{\mathbf{Z}}_{*N}, \tilde{\mathbf{y}}_{*N}$ are defined analogously.

The last step is to derive the limiting distribution of the infeasible estimator $\sqrt{N}(\bar{\beta}^{\inf}_{2SLS} - \beta^0)$ by showing

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}_i - \mathbf{h}^{\mathbf{w}}_*(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_*(\mathbf{x}_{2i}, a_i))' \xrightarrow{p} \bar{\mathbf{S}}^{\mathbf{wz}}$$

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_*(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_*(\mathbf{x}_{2i}, a_i)' \xrightarrow{p} \bar{\mathbf{S}}^{\mathbf{zz}}$$

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_*(\mathbf{x}_{2i}, a_i))\eta^{\upsilon}_{*i} \Rightarrow \mathcal{N}(0, \bar{\mathbf{S}}^{\mathbf{zz}\sigma}),$$

Combining all the limit results we have the following theorem.

Theorem 5.2 (Limiting Distribution)**.** *Suppose that Assumptions 1, 3, 4, 9, 10, and 11 hold. Then, we have*

$$\sqrt{N}(\bar{\beta}_{2SLS} - \beta^0) \Rightarrow \mathcal{N}\left(0, \bar{\Omega}\right),$$

where

$$\bar{\Omega} = \left(\bar{\mathbf{S}}^{\mathbf{wz}}\left(\bar{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}(\bar{\mathbf{S}}^{\mathbf{wz}})'\right)^{-1}\left(\bar{\mathbf{S}}^{\mathbf{wz}}\left(\bar{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}\bar{\mathbf{S}}^{\mathbf{zz}\sigma}\left(\bar{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}(\bar{\mathbf{S}}^{\mathbf{wz}})'\right) \times \left(\bar{\mathbf{S}}^{\mathbf{wz}}\left(\bar{\mathbf{S}}^{\mathbf{zz}}\right)^{-1}(\bar{\mathbf{S}}^{\mathbf{wz}})'\right)^{-1}.$$

Assumption 3 assumes that the regressors in the outcome equation, $\mathbf{x}_{1i}$ and the observables in the network formation $\mathbf{x}_{2i}$ do not overlap. Assumption 4 is a full rank condition for $\bar{\beta}_{2SLS}$. Assumptions 9 and 10 assume the sieve used in constructing the estimator $\bar{\beta}_{2SLS}$. Comparing the assumptions assumed in Theorem 5.1, Theorem 5.2 does not require the high level condition of Assumption 5 because we do not use an estimator of a_i. Instead it requires an additional restriction that the net surplus of link formation be strictly monotonic in a_i, as explained in Section 4.2.

Like in the case of $\hat{\beta}_{2SLS}$, we allow $\eta^{\upsilon}_{*i} = \upsilon_i - \mathbb{E}(\upsilon_i|\mathbf{x}_{2i}, a_i)$ to be conditionally heteroskedastic, and $\sigma^2_*(\mathbf{x}_i, a_i) := \mathbb{E}[(\upsilon_i - \mathbb{E}[\upsilon_i|\mathbf{x}_{2i}, a_i])^2|\mathbf{x}_i, a_i]$ is allowed to depend on $(\mathbf{x}_i, a_i)$.

The asymptotic variance can be consistently estimated by

$$\hat{\hat{\Omega}} = \left(\hat{\hat{\mathbf{S}}}^{\mathbf{wz}}\left(\hat{\hat{\mathbf{S}}}^{\mathbf{zz}}\right)^{-1}(\hat{\hat{\mathbf{S}}}^{\mathbf{wz}})')\right)^{-1}\left(\hat{\hat{\mathbf{S}}}^{\mathbf{wz}}\left(\hat{\hat{\mathbf{S}}}^{\mathbf{zz}}\right)^{-1}\hat{\hat{\mathbf{S}}}^{\mathbf{zz}\sigma}\left(\hat{\hat{\mathbf{S}}}^{\mathbf{zz}}\right)^{-1}(\hat{\hat{\mathbf{S}}}^{\mathbf{wz}})')\right)$$
$$\times\left(\hat{\hat{\mathbf{S}}}^{\mathbf{wz}}\left(\hat{\hat{\mathbf{S}}}^{\mathbf{zz}}\right)^{-1}(\hat{\hat{\mathbf{S}}}^{\mathbf{wz}})')\right)^{-1}, \tag{5.6}$$

where

$$\hat{\hat{S}}^{\mathbf{wz}} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{w}_i - \hat{\mathbf{h}}^{\mathbf{w}}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i)\right)'$$

$$\hat{\hat{S}}^{\mathbf{zz}} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i)\right)'$$

$$\hat{\hat{S}}^{\mathbf{zz}\sigma^2} = \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i)\right)'(\hat{\eta}^{v}_{**i})^2,$$

and $\hat{\eta}^{v}_{**i} = y_i - \hat{h}^{y}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i) - (\mathbf{w}_i - \hat{\mathbf{h}}^{\mathbf{w}}_{**}(\mathbf{x}_{2i}, \widehat{\deg}_i))'\bar{\beta}_{2SLS}$.

6. Monte Carlo

The Monte Carlo design of the network formation process follows Graham (2017). Links are formed according to

$$d_{ij,N} = \mathbb{I}\left\{x_{2i}x_{2j}\lambda + a_i + a_j - u_{ij} \geq 0\right\},$$

where $x_{2i} \in \{-1, 1\}$, $\lambda = 1$ and u_{ij} follows a logistic distribution. This link rule implies that agents have a strong taste for homophilic matching since $x_{2i}x_{2j}\lambda = 1$ when $x_{2i} = x_{2j}$ and $x_{2i}x_{2j}\lambda = -1$ when $x_{2i} \neq x_{2j}$. Individual-level degree heterogeneity is generated according to

$$a_i = \varphi(\alpha_L \mathbb{I}\left\{x_{2i} = -1\right\} + \alpha_H \mathbb{I}\left\{x_{2i} = 1\right\} + \xi_i),$$

with $\alpha_L \leq \alpha_H$ and ξ_i a centered Beta random variable $\xi_i|x_{2i} \sim \left\{Beta(\mu_0, \mu_1) - \frac{\mu_0}{\mu_0+\mu_1}\right\}$ so that $a_i \in \left(\varphi\left[\alpha_L - \frac{\mu_0}{\mu_0+\mu_1}, \alpha_H + \frac{\mu_1}{\mu_0+\mu_1}\right]\right)$. φ is a scaling factor that assures that $|a_i| \leq 1$ in the designs that so require.

We set the parameter values $\alpha_L = -3/2$, $\alpha_H = 1$, $\mu_0 = 1/4$ and $\mu_1 = 3/4$. This design involves degree heterogeneity distributions that are correlated with x_{2i} and right skewed, which mimics distributions observed in real world networks. We have explored other specifications of the network formation parameters and the results of the Monte Carlo simulations are not notably different. These results are available upon request.

Individual outcomes are generated according to

$$y_i = \beta_1 \sum_{\substack{j=1 \\ j\neq i}}^{N} g_{ij,N} y_j + \beta_2 x_{1i} + \beta_3 \sum_{\substack{j=1 \\ j\neq i}}^{N} g_{ij,N} x_{1j} + h(a_i) + \varepsilon_i.$$

In the simulations, we set $\beta_2 = \beta_3 = 1$, $\beta_1 \in \{0.2,\ 0.5,\ 0.8\}$, $x_{1i} = 3q_1 + \cos(q_2)/0.8 + \epsilon_i$, where $q_1, q_2 \sim \mathcal{N}(x_{2i}, 1)$, and $\varepsilon_i, \epsilon_i \sim \mathcal{N}(0, 1)$. For $h(a_i)$ we use the following functional forms: $h(a_i) = \exp(\kappa a_i)$, and $h(a_i) = \sin(\kappa a_i)$ with $\kappa = 3$.

We estimate the outcome equation coefficients $(\beta_1, \beta_2, \beta_3)$ using the standard 2SLS estimator for peer effects without controlling for the network endogeneity, using a control function linear in $\hat{a}_i$, $\hat{h}(\hat{a}_i)$, $\hat{h}(\widehat{deg}_i, x_{2i})$[5], and $h(a_i)$ with polynomial and Hermite polynomial sieve bases.

We perform simulations with network size $N = 100, 250$ with a dense and sparse network design. The average number of links for the dense design is 24 for $N = 100$ and 60 for $N = 250$. The corresponding numbers for the sparse design are 2 and 5, respectively.

In the paper, we present Monte Carlo results with the Hermite polynomial sieve and $h(a_i) = sin(3a_i)$ for selected values of K_N. Specifically, Tables 1 and 2 include results for the dense network specification for $K_N = 4$ and $K_N = 6$, respectively. Results for the sparse network for $K_N = 3$ are presented in Table 3. Results for the other specifications are not notably different and are included in the Supplementary Appendix. For the sparse design, we present results with the control function $\hat{h}(\widehat{deg}_i, x_{2i})$.

[5] Note that since x_{2i} is a discrete with a finite support $x_{2i} \in \{x_1, ..., x_M\}$ we have $r(x_{2i}, deg_i) = \sum_{m=1}^{M} r(x_m, deg_i)\mathbb{I}\{x_{2i} = x_m\}$. We can then approximate $r(x_{2i}, deg_i) \simeq \sum_{k=1}^{K_N} \left\{ \sum_{m=1}^{M} \alpha_{m,k} q_k^d(deg)\mathbb{I}\{x_{2i} = x_m\} \right\}$.

We also perform the conventional leave-one-out cross validation to find data-dependent K_N (chosen as the K_N that minimizes the Root Mean Square Error (RMSE) of the prediction based on the leave-one-out estimator, see for example Andrews (1991), Hansen (2014)). We report the data-dependent K_N, $K^*_{N,\beta}$, for each design in the footnotes of the tables in the case of control function $\hat{h}(\widehat{deg}_i, x_{2i})$. The differences in RMSE are very small between the different values of K_N.

Analyzing Monte Carlo results we can see that the order of K_N does not have a significant impact on the estimates for K_N between 3 and 8. The conventional data-driven method for the choice of K_N gives us K^*_N in this range. In almost all specifications, as expected from our asymptotic theories, the control functions $\hat{h}(\hat{a}_i)$ and $\hat{h}(\widehat{deg}_i, x_{2i})$ perform better than the estimator with a linear control function, as well as the estimator that does not control for the endogeneity of the network, both in terms of mean bias and size. Also, as expected, the true control function $h(a_i)$ performs best of all specifications. An interesting observation is that our estimator continues to show good finite sample performances even in the specification with the sparse network presented in Table 3.

Table 1. **Hermite Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \sin(3a_i)$ and $K_N = 4$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.200$	-1.681	-0.398	-0.101	0.007	-0.010	-0.445	-0.630	-0.091	-0.023	-0.019	*mean bias*
	(26.702)	(0.325)	(0.243)	(0.349)	(0.246)	(160.512)	(0.352)	(0.232)	(0.372)	(0.297)	*std*
	0.597	0.301	0.078	0.069	0.018	0.668	0.563	0.078	0.063	0.014	*size*
$\beta_2 = 1$	0.128	0.022	0.002	0.001	0.001	-0.043	0.019	0.001	0.000	0.001	*mean bias*
	(2.379)	(0.041)	(0.036)	(0.037)	(0.035)	(6.620)	(0.024)	(0.021)	(0.021)	(0.022)	*std*
	0.372	0.109	0.076	0.086	0.016	0.557	0.137	0.057	0.061	0.018	*size*
$\beta_3 = 1$	2.674	0.520	0.094	-0.002	0.015	0.627	0.937	0.097	0.029	0.033	*mean bias*
	(41.110)	(0.530)	(0.411)	(0.473)	(0.425)	(277.957)	(0.605)	(0.407)	(0.499)	(0.532)	*std*
	0.544	0.203	0.062	0.062	0.017	0.668	0.430	0.069	0.058	0.013	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.5$	-1.095	-0.211	-0.062	0.003	-0.005	-0.462	-0.321	-0.056	-0.015	-0.009	*mean bias*
	(14.839)	(0.170)	(0.151)	(0.219)	(0.140)	(25.186)	(0.171)	(0.142)	(0.230)	(0.147)	*std*
	0.771	0.313	0.078	0.069	0.023	0.787	0.576	0.079	0.063	0.015	*size*
$\beta_2 = 1$	0.148	0.015	0.001	0.001	0.000	0.023	0.013	0.001	0.000	0.000	*mean bias*
	(2.399)	(0.037)	(0.035)	(0.036)	(0.033)	(1.468)	(0.023)	(0.021)	(0.021)	(0.021)	*std*
	0.316	0.100	0.075	0.083	0.021	0.529	0.096	0.057	0.063	0.020	*size*
$\beta_3 = 1$	2.154	0.317	0.060	0.002	0.011	0.910	0.571	0.063	0.022	0.022	*mean bias*
	(30.259)	(0.364)	(0.329)	(0.363)	(0.310)	(55.722)	(0.383)	(0.321)	(0.369)	(0.336)	*std*
	0.663	0.174	0.061	0.064	0.018	0.786	0.378	0.066	0.055	0.015	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.8$	-0.168	-0.093	-0.036	0.002	-0.002	-0.320	-0.136	-0.033	-0.009	-0.004	*mean bias*
	(0.285)	(0.079)	(0.088)	(0.130)	(0.070)	(5.928)	(0.075)	(0.083)	(0.136)	(0.065)	*std*
	0.802	0.321	0.078	0.069	0.025	0.897	0.592	0.079	0.064	0.022	*size*
$\beta_2 = 1$	0.026	0.008	0.000	0.001	-0.000	0.028	0.008	0.001	0.000	0.000	*mean bias*
	(0.072)	(0.034)	(0.035)	(0.035)	(0.031)	(0.470)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.135	0.084	0.077	0.082	0.021	0.227	0.072	0.057	0.065	0.020	*size*
$\beta_3 = 1$	0.401	0.162	0.037	0.004	0.007	0.879	0.301	0.041	0.017	0.013	*mean bias*
	(0.817)	(0.251)	(0.274)	(0.292)	(0.219)	(18.665)	(0.244)	(0.265)	(0.288)	(0.203)	*std*
	0.479	0.127	0.062	0.068	0.024	0.849	0.279	0.064	0.053	0.020	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg}_i, x_{2i})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 7$, $K^*_{100,0.5} = 3$, $K^*_{100,0.8} = 3$
$K^*_{250,0.2} = 8$, $K^*_{250,0.5} = 6$, $K^*_{250,0.8} = 4$

TABLE 2. **Hermite Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \sin(3a_i)$ and $K_N = 6$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.200$	-1.681	-0.398	-0.102	0.005	-0.010	-0.445	-0.630	-0.090	-0.023	-0.019	*mean bias*
	(26.702)	(0.325)	(0.245)	(0.350)	(0.246)	(160.512)	(0.352)	(0.233)	(0.372)	(0.297)	*std*
	0.597	0.301	0.082	0.065	0.018	0.668	0.563	0.078	0.065	0.014	*size*
$\beta_2 = 1$	0.128	0.022	0.002	0.001	0.001	-0.043	0.019	0.001	0.000	0.001	*mean bias*
	(2.379)	(0.041)	(0.037)	(0.037)	(0.035)	(6.620)	(0.024)	(0.021)	(0.021)	(0.022)	*std*
	0.372	0.109	0.077	0.085	0.016	0.557	0.137	0.057	0.063	0.018	*size*
$\beta_3 = 1$	2.674	0.520	0.096	0.000	0.015	0.627	0.937	0.095	0.029	0.033	*mean bias*
	(41.110)	(0.530)	(0.415)	(0.475)	(0.425)	(277.957)	(0.605)	(0.409)	(0.499)	(0.532)	*std*
	0.544	0.203	0.062	0.061	0.017	0.668	0.430	0.073	0.056	0.013	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.5$	-1.095	-0.211	-0.063	0.002	-0.005	-0.462	-0.321	-0.055	-0.015	-0.009	*mean bias*
	(14.839)	(0.170)	(0.152)	(0.220)	(0.140)	(25.186)	(0.171)	(0.143)	(0.230)	(0.147)	*std*
	0.771	0.313	0.080	0.066	0.023	0.787	0.576	0.078	0.066	0.015	*size*
$\beta_2 = 1$	0.148	0.015	0.001	0.001	0.000	0.023	0.013	0.001	0.000	0.000	*mean bias*
	(2.399)	(0.037)	(0.036)	(0.036)	(0.033)	(1.468)	(0.023)	(0.021)	(0.021)	(0.021)	*std*
	0.316	0.100	0.078	0.089	0.021	0.529	0.096	0.055	0.064	0.020	*size*
$\beta_3 = 1$	2.154	0.317	0.061	0.004	0.011	0.910	0.571	0.062	0.022	0.022	*mean bias*
	(30.259)	(0.364)	(0.332)	(0.364)	(0.310)	(55.722)	(0.383)	(0.323)	(0.370)	(0.336)	*std*
	0.663	0.174	0.063	0.060	0.018	0.786	0.378	0.070	0.056	0.015	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.8$	-0.168	-0.093	-0.037	0.001	-0.002	-0.320	-0.136	-0.032	-0.009	-0.004	*mean bias*
	(0.285)	(0.079)	(0.089)	(0.130)	(0.070)	(5.928)	(0.075)	(0.084)	(0.136)	(0.065)	*std*
	0.802	0.321	0.080	0.065	0.025	0.897	0.592	0.078	0.066	0.022	*size*
$\beta_2 = 1$	0.026	0.008	0.000	0.001	-0.000	0.028	0.008	0.001	0.000	0.000	*mean bias*
	(0.072)	(0.034)	(0.035)	(0.036)	(0.031)	(0.470)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.135	0.084	0.078	0.090	0.021	0.227	0.072	0.057	0.064	0.020	*size*
$\beta_3 = 1$	0.401	0.162	0.038	0.006	0.007	0.879	0.301	0.040	0.017	0.013	*mean bias*
	(0.817)	(0.251)	(0.277)	(0.293)	(0.219)	(18.665)	(0.244)	(0.267)	(0.289)	(0.203)	*std*
	0.479	0.127	0.061	0.069	0.024	0.849	0.279	0.068	0.054	0.020	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg}_i, x_{2i})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 7$, $K^*_{100,0.5} = 3$, $K^*_{100,0.8} = 3$
$K^*_{250,0.2} = 8$, $K^*_{250,0.5} = 6$, $K^*_{250,0.8} = 4$

TABLE 3. **Hermite Polynomial Sieve**: Parameter values across 1000 MC replications with $h(a) = \sin(3a_i)$ and $K_N = 3$

N	100			250			
CF	(0)	(1)	(2)	(0)	(1)	(2)	
$\beta_1 = 0.2$	0.011 (0.048) 0.329	-0.003 (0.055) 0.075	-0.002 (0.049) 0.050	0.034 (0.042) 0.366	-0.001 (0.051) 0.052	-0.001 (0.045) 0.048	*mean bias* *std* *size*
$\beta_2 = 1$	0.003 (0.035) 0.320	0.001 (0.040) 0.066	0.001 (0.035) 0.054	-0.005 (0.022) 0.290	0.001 (0.024) 0.054	0.001 (0.022) 0.053	*mean bias* *std* *size*
$\beta_3 = 1$	-0.003 (0.078) 0.306	0.004 (0.081) 0.065	0.003 (0.077) 0.044	-0.034 (0.069) 0.324	0.000 (0.074) 0.048	0.001 (0.071) 0.042	*mean bias* *std* *size*
N	100			250			
CF	(0)	(1)	(2)	(0)	(1)	(2)	
$\beta_1 = 0.5$	0.007 (0.029) 0.313	-0.002 (0.033) 0.071	-0.001 (0.030) 0.054	0.019 (0.025) 0.377	-0.001 (0.032) 0.042	-0.001 (0.027) 0.038	*mean bias* *std* *size*
$\beta_2 = 1$	0.003 (0.034) 0.305	-0.001 (0.038) 0.057	-0.000 (0.034) 0.047	-0.002 (0.021) 0.293	0.001 (0.022) 0.067	0.001 (0.021) 0.059	*mean bias* *std* *size*
$\beta_3 = 1$	-0.003 (0.071) 0.313	0.001 (0.074) 0.067	0.002 (0.071) 0.052	-0.021 (0.058) 0.321	0.003 (0.063) 0.049	0.002 (0.059) 0.039	*mean bias* *std* *size*
N	100			250			
CF	(0)	(1)	(2)	(0)	(1)	(2)	
$\beta_1 = 0.8$	0.003 (0.012) 0.295	-0.000 (0.014) 0.050	-0.000 (0.012) 0.036	0.008 (0.011) 0.384	-0.001 (0.017) 0.051	-0.001 (0.014) 0.035	*mean bias* *std* *size*
$\beta_2 = 1$	0.004 (0.031) 0.288	-0.001 (0.037) 0.057	-0.000 (0.031) 0.049	-0.001 (0.019) 0.329	-0.000 (0.022) 0.061	-0.000 (0.019) 0.048	*mean bias* *std* *size*
$\beta_3 = 1$	-0.001 (0.063) 0.289	0.000 (0.065) 0.065	0.001 (0.063) 0.043	-0.011 (0.048) 0.313	0.002 (0.056) 0.049	0.002 (0.051) 0.038	*mean bias* *std* *size*

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\widehat{deg}_i, x_{2i})$.
Average number of links for $N = 100$ is 1.9, for $N = 250$ it is 4.9.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$.
$K^*_{100,0.2} = 8$, $K^*_{100,0.5} = 5$, $K^*_{100,0.8} = 6$
$K^*_{250,0.2} = 4$, $K^*_{250,0.5} = 6$, $K^*_{250,0.8} = 5$

7. Conclusions

In this paper we show that, whenever the network is likely endogenous, it is important to control for this endogeneity when estimating peer effects. Failing to control for the endogeneity of the connections matrix in general leads to biased estimates of peer effects. We show that under specific assumptions, we can use the control function approach to deal with the endogeneity problem. We assume that unobserved individual characteristics directly affect link formation and individual outcomes. We leave the functional form through which unobserved individual characteristics enter the outcome equation unspecified and estimate it using a non-parametric approach. The estimators we propose is easy to use in applied work, and Monte Carlo results show that they preform well compared to a linear control function estimator. Erroneously assuming that unobserved characteristics enter the outcome equation in a linear fashion can lead to a serious bias in the estimated parameters.

References

Abadie, A. and G. W. Imbens (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica 74*(1), 235–267.

Ahn, H. and J. L. Powell (1993). Semiparametric estimation of censored selection models with a nonparametric selection mechanism. *Journal of Econometrics 58*(1-2), 3–29.

Andrews, D. W. (1991). Asymptotic optimality of generalized cl, cross-validation, and generalized cross-validation in regression with heteroskedastic errors. *Journal of Econometrics 47*(2-3), 359–377.

Arduini, T., E. Patacchini, and E. Rainone (2015). Parametric and Semiparametric IV Estimation of Network Models with Selectivity. *Working Paper*.

Auerbach, E. (2016). Identification and estimation of models with endogenous network formation. *working paper*.

Blume, L., W. Brock, S. Durlauf, and Y. Ioannides (2010). Identification of social interactions. *working paper*.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear social interactions models. *Journal of Political Economy 123*(2), 444–496.

Bramoullé, Y., H. Djebbari, and B. Fortin (2009, May). Identification of peer effects through social networks. *Journal of Econometrics 150*(1), 41–55.

De Paula, A. (2016). Econometrics of network models. *cemmap working paper, Centre for Microdata Methods and Practice*.

Dzemski, A. (2017). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *working paper*.

Epple, D. and R. E. Romano (2011). *Peer effects in education: A survey of the theory and evidence* (1 ed.), Volume 1. Elsevier B.V.

Escanciano, J. C., D. T. Jacho-Chávez, and A. Lewbel (2014). Uniform convergence of weighted sums of non and semiparametric residuals for estimation and testing. *Journal of Econometrics 178*, 426–443.

Fernández-Val, I. and M. Weidner (2013). Individual and time effects in nonlinear panel models with large N, T. *arXiv preprint arXiv:1311.7065.*

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social Networks and the Identification of Peer Effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

Graham, B. (2011). Econometric Methods for the Analysis of Assignment Problems in the Presence of Complimentarity and Social Spillovers. *Handbook of Social Economics 1B*, 965–1062.

Graham, B. (2015). Methods of identification in social networks. *Annual Review of Economics 7*, 1–60.

Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica 85*(4), 1033–1063.

Hahn, J. and G. Ridder (2013). Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica 81*(1), 315–340.

Hall, P. and C. C. Heyde (2014). *Martingale limit theory and its application.* Academic press.

Hansen, B. E. (2014). Nonparametric sieve regression: Least squares, averaging least squares, and cross-validation. *Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics, forthcoming*.

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The review of economic studies 65*(2), 261–294.

Hsieh, C.-S. and L. F. Lee (2016). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics 31*(2), 301–319.

Jochmans, K. (2016). Modified-likelihood estimation of the b-model. Technical report, Sciences Po Departement of Economics.

Jochmans, K. (2017). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics*, 1–9.

Kelejian, H. and I. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics 17*, 99–121.

Lee, L. (2003, January). Best Spatial TwoStage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews 22*(4), 307–335.

Lee, L.-f. (2007a, April). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics 137*(2), 489–514.

Lee, L.-f. (2007b, October). Identification and estimation of econometric models with group interactions, contextual factors and fixed effects. *Journal of Econometrics 140*(2), 333–374.

Lee, L.-f., X. Liu, and X. Lin (2010). Specification and estimation of social interaction models with network structures. *Econometrics Journal 13*, 145–176.

Li, K.-C. (1987). Asymptotic optimality for cp, cl, cross-validation and generalized cross-validation: discrete index set. *The Annals of Statistics*, 958–975.

Li, Q. and S. J. Racine (2007). *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Mammen, E., C. Rothe, M. Schienle, et al. (2012). Nonparametric regression with nonparametrically generated covariates. *The Annals of Statistics 40*(2), 1132–1170.

Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*.

Manski, C. F. (2000). Economic analysis of social interactions. Technical report, NBER.

Newey, W. K. (1997, July). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Newey, W. K. (2009). Two-step series estimation of sample selection models. *The Econometrics Journal 12*(s1).

Powell, J. L. (1987). Semiparametric estimation of bivariate latent variable models.

Qu, X. and L.-f. Lee (2015, February). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics 184*(2), 209–232.

Shalizi, C. R. (2012, January). Comment on "Why and When 'Flawed' Social Network Analyses Still Yield Valid Tests of no Contagion". *Statistics, Politics, and Policy 3*(1).

Sheng, S. (2012). Identification and estimation of network formation games. *Unpublished manuscript*.

Wahba, G. (1985). A comparison of gcv and gml for choosing the smoothing parameter in the generalized spline smoothing problem. *The Annals of Statistics*, 1378–1402.

Weinberg, B. (2007). Social interactions with endogenous associations.

Appendix

In this section we introduce the assumptions that are required for the two asymptotic results, Theorem 5.1 for $\hat{\beta}_{2SLS}$ and Theorem 5.2 for $\bar{\beta}_{2SLS}$. We also outline the proof of Theorem 5.1. The proof of Theorem 5.2 is similar to that of Theorem 5.1.

Detailed proofs are available in the Supplementary Appendix which is available at http://www-bcf.usc.edu

Appendix A.1. Assumptions

In this section we introduce the assumptions used in the proof of Theorem 5.1.

First, we introduce a set of sufficient conditions under which we can estimate a_i satisfying the conditions in Assumption 5. This assumption corresponds to Assumptions 1, 2, 3 and 5 of Graham (2017).

Assumption 6 (Sufficient Conditions for Assumption 5)**.** *(i)* $\mathbf{t}_{ij} = \mathbf{t}_{ji}$*. (ii)* $u_{ij} \sim$ *i.i.d. for all* ij *a logistic distribution. (iii) The supports of* λ, $\mathbf{t}_{ij}$, a_i *are compact.*

The next four assumptions are about the sieves used in the semiparametric estimators. The first two are for $\hat{\beta}_{2SLS}$ and the next two are for $\bar{\beta}_{2SLS}$.

Assumption 7 (Sieve)**.** *For every* K_N *there is a non-singular matrix of constants* $\mathbf{B}$ *such that for* $\tilde{\mathbf{q}}^{K_N}(a) = \mathbf{B}\mathbf{q}^{K_N}(a)$*,*

(i) The smallest eigenvalue of $\mathbb{E}[\tilde{\mathbf{q}}^{K_N}(a_i)\tilde{\mathbf{q}}^{K_N}(a_i)']$ *is bounded away from zero uniformly in* K_N*.*

(ii) There exists a sequence of constants $\zeta_0(K_N)$ *that satisfy the condition*

$$\sup_{a\in\mathcal{A}} \|\tilde{\mathbf{q}}^{K_N}(a)\| \leq \zeta_0(K_N),$$

where K_N *satisfies* $\zeta_0(K_N)^2 K_N/N \to 0$ *as* $N \to \infty$*.*

(iii) For $f(a)$ *being an element of* $\mathbf{h}(a) = (E[y_i|a_i = a], \mathbb{E}[\mathbf{z}_i|a_i = a], \mathbb{E}[\mathbf{w}_i|a_i = a])$*, there exists a sequence of* $\boldsymbol{\alpha}^f_{K_N}$ *and a number* $\kappa > 0$ *such that*

$$\sup_{a\in\mathcal{A}} \|f(a) - \mathbf{q}^{K_N}(a)'\boldsymbol{\alpha}^f_{K_N}\| = O(K_N^{-\kappa})$$

as $K_N \to \infty$.

(iv) As $N \to \infty$, $K_N \to \infty$ *with* $\sqrt{N}K_N^{-\kappa} \to 0$ *and* $K_N/N \to 0$.

Assumption 8 (Lipschitz condition)**.** *The sieve basis satisfies the following condition: there exists a positive number* $\zeta_1(k)$ *such that*

$$\|\mathbf{q}_k(a) - \mathbf{q}_k(a')\| \leq \zeta_1(k)\|a - a'\| \;\forall\; k = 1, \ldots, K_N$$

with

$$\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1^2(k) = o(1)$$

and

$$\zeta_0(K_N)^6\left(\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1^2(k)\right) = o(1).$$

In our paper, we use the following sieves for the Monte Carlo simulations.

(i) Polynomial: For $|a| \leq 1$, define

$$Pol(K_N) = \left\{\nu_0 + \sum_{k=1}^{K_N}\nu_k a^k,\; a \in [-1,1]\, \nu_k \in \mathbb{R}\right\}$$

(ii) The Hermite Polynomial sieve:

$$HPol(K_N) = \left\{\sum_{k=1}^{K_N+1}\nu_k H_k(a)\exp\left(\frac{-a^2}{2}\right),\; a \in [-1,1], \nu_k \in \mathbb{R}\right\},$$

where $H_k(a) = (-1)^k e^{a^2}\frac{d^k}{da^k}e^{-a^2}$.

For polynormial sieve, it is known that $\zeta_0 = O(K_N)$ (e.g., Newey (1997)). Then, since $\zeta_1(k) = O(k)$, $\sum_{k=1}^{K_N}\zeta_1^2(k) = O(K_N^3)$. Hence, the conditions that must be satisfied for the polynomial sieve are $K_N^3/N \to 0$ and $\sqrt{N}K_N^{-\kappa} \to 0$. Further, when $\zeta_a(N)^2 = \frac{N}{\ln N}$, we need $\zeta_a(N)^{-2}O(K_N^9) = o(1)$.

The next two assumptions are for the sieves used in $\bar{\beta}_{2SLS}$. These assumptions modify Assumption 7 and Assumption 8.

Assumption 9 (Sieve)**.** *For every K_N there is a non-singular matrix of constants $\mathbf{B}$ such that for $\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, deg_i) = \mathbf{B}\mathbf{r}^{K_N}(\mathbf{x}_{2i}, deg_i)$,*

(i) The smallest eigenvalue of $\mathbb{E}[\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, deg_i)\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, deg_i)']$ is bounded away from zero uniformly in K_N.

(ii) There exists a sequence of constants $\zeta_{0**}(K_N)$ that satisfy the condition

$$\sup_{(\mathbf{x}_{2i}, deg_i)\in\mathcal{S}} \|\tilde{\mathbf{r}}^{K_N}(\mathbf{x}_{2i}, deg_i)\| \leq \zeta_{0**}(K_N),$$

where K_N satisfies $\zeta_{0**}(K_N)^2 K_N/N \rightarrow 0$ as $N \rightarrow \infty$, and $\mathcal{S}$ is the domain of $(\mathbf{x}_{2i}, deg_i)$.

(iii) For $f(\mathbf{x}_{2i}, deg_i)$ being an element of $\mathbf{h}_{**}(\mathbf{x}_{2i}, deg_i) = (\mathbb{E}[y_i|\mathbf{x}_{2i}, deg_i], \mathbb{E}[\mathbf{z}_i|\mathbf{x}_{2i}, deg_i], \mathbb{E}[\mathbf{w}_i|\mathbf{x}_{2i}, deg_i])$, there exists a sequence of $\boldsymbol{\gamma}^f_{K_N}$ and a number $\kappa > 0$ such that

$$\sup_{(\mathbf{x}_{2i}, deg_i)\in\mathcal{S}} \|f - \mathbf{r}^{K_N\prime}\boldsymbol{\gamma}^f_{K_N}\| = O(K_N^{-\kappa})$$

as $K_N \rightarrow \infty$.

(iv) As $N \rightarrow \infty$, $K_N \rightarrow \infty$ with $\sqrt{N}K_N^{-\kappa} \rightarrow 0$ and $K_N/N \rightarrow 0$.

Recall from (17) that $\sup_i |\widehat{deg}_i - deg_i| = O(\zeta_{deg}(N)^{-1})$ with $\zeta_{deg}(N) = o(1)N^{\frac{B-1}{2B}}$ for some integer $B \geq 2$.

Assumption 10 (Lipschitz)**.** *For $\zeta_{0**}(K_N)$ being the constant from Assumption 10, there exists a positive number $\zeta_{1**}(k)$ such that*

$$\|\mathbf{r}_k(\mathbf{x}_{2i}, deg_i) - \mathbf{r}_k(\mathbf{x}_{2i}, deg_i')\| \leq \zeta_{1**}(k)\|deg_i - deg_i'\| \;\forall\; k = 1, \ldots, K_N$$

with

$$\zeta_{deg}(N)^{-2}\sum_{k=1}^{K_N}\zeta^2_{1**}(k) = o(1)$$

and

$$\zeta_{0**}(K_N)^6\left(\zeta_{deg}(N)^{-2}\sum_{k=1}^{K_N}\zeta^2_{1**}(k)\right) = o(1).$$

The next assumptions restrict the models of the outcome in (2.1) and the network formation of (3.1). We need Assumption 11 to derive the limiting distribution of $\hat{\beta}_{2SLS}$ in Theorem 5.1.

Assumption 11. *We assume the following:*

(i) *The true coefficients satisfies* $|\beta_1^0| \leq 1-\epsilon$ *and* $\|\beta_2^0\| > \epsilon$ *for some small* ϵ.
(ii) *The parameter set* $\mathbb{B}$ *for* β *is bounded.*
(iii) *The observables* $(y_i, \mathbf{x}_i)$ *are bounded. The unobserved characteristic* a_i *has a compact support in* $[-1,1]$.
(iv) *The network formation error* u_{ij} *has an unbounded full support* $\mathbb{R}$.
(v) *The net surplus of the network* $g(\mathbf{t}_{ij}, a_i, a_j)$ *is bounded by a finite constant.*
(vi) *The net surplus function* $g(\mathbf{t}_{ij}, a_i, a_j)$ *is strictly monotonic in* a_i *and* a_j *for all* $\mathbf{t}_{ij}$.

Condition (i) is standard in the linear-in-means peer effect literature. The condition $|\beta_1^0| \leq 1-\epsilon$ is required for the unique solution of the spillover effect. We need the restriction $\|\beta_2^0\| > \epsilon$ for the IVs to be strong. The boundedness conditions in (ii) and (iii) are important technical assumptions for asymptotics which require some uniform convergence. Also, these conditions imply key regularity conditions for the CLT. Conditions (vi) and (v) assume that the network is dense and $\mathbb{E}[d_{ij,N}=1] \geq \underline{\kappa} > 0$.

Finally, notice that Assumption 11 allows $\upsilon_i - \mathbb{E}(\upsilon_i|a_i)$ to be conditionally heteroskedastic, and so $\sigma^2(\mathbf{x}_i, a_i) := \mathbb{E}[(\upsilon_i - \mathbb{E}[\upsilon_i|a_i])^2|\mathbf{x}_i, a_i]$ depends on $(\mathbf{x}_i, a_i)$. This is also true for $\upsilon_i - \mathbb{E}(\upsilon_i|a_i)$

Appendix A.2. Outline of the proof of Theorem 5.1

By definition, we have

$$\begin{aligned}\hat{\beta}_{2SLS} - \beta^0 &= \left(\mathbf{W}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\left(\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{W}_N\right)^{-1} \\ &\quad \times\mathbf{W}_N'M_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\left(\mathbf{Z}'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\left(\boldsymbol{\eta}_N^{\upsilon} - \mathbf{h}^{\upsilon}(\mathbf{a}_N) - \widehat{\mathbf{Q}}_N\boldsymbol{\alpha}_{K_N}^{\upsilon}\right).\end{aligned}$$

The derivation of the asymptotic distribution of $\hat{\beta}_{2SLS}$ consists of three steps.

Step 1. First, we control the sampling error coming from the fact that we do not observe $\mathbf{a}_N$ and approximate it with $\hat{\mathbf{a}}_N$. Under suitable assumptions (see Supplementary Appendix S.1.1), we show that the error that stems from the estimation of $\mathbf{a}_N$ by $\hat{\mathbf{a}}_N$ is asymptotically negligible:

$$
\begin{aligned}
&\sqrt{N}\left(\hat{\beta}_{2SLS}-\beta^0\right)\\
=\;&\left(\frac{1}{N}\mathbf{W}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\left(\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\right)^{-1}\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{W}_N\right)^{-1}\\
&\times\frac{1}{N}\mathbf{W}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\left(\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\right)^{-1}\frac{1}{\sqrt{N}}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\boldsymbol{\eta}_N^{v}+o_p(1).
\end{aligned}
$$

(See Lemma 2 in Supplementary Appendix S.1.1)

Step 2. Next, we consider the error introduced by the non-parametric estimation of $h(a_i)$. Let $\mathbf{h}^{\mathbf{w}}(a_i)=\mathbb{E}(\mathbf{w}_i|a_i), \eta_i^{\mathbf{w}}=\mathbf{w}_i-\mathbf{h}^{\mathbf{w}}(a_i)$, $\mathbf{h}^{\mathbf{z}}(a_i)=\mathbb{E}(\mathbf{z}_i|a_i)$ and $\eta_i^{\mathbf{z}}=\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)$. Let $\hat{\mathbf{h}}^{\mathbf{w}}(a_i)$ and $\hat{\mathbf{h}}^{\mathbf{z}}(a_i)$ denote the series approximation of $\mathbf{h}^{\mathbf{w}}(a_i)$ and $\mathbf{h}^{\mathbf{z}}(a_i)$, respectively. In Lemma 7 in Supplementary Appendix S.1.2 we show that under the regularity conditions (see Supplementary Appendix S.1.2), the error from estimating $h(a_i)$ with $\hat{h}(a_i)$ converges to zero at a suitable rate and we have

$$
\begin{aligned}
\frac{1}{N}\mathbf{W}_N'\mathbf{M}_{\mathbf{Q}}\mathbf{Z}_N &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{w}_i-\hat{\mathbf{h}}^{\mathbf{w}}(a_i)\right)\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)'\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{w}_i-\mathbf{h}^{\mathbf{w}}(a_i)\right)\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)'+o_p(1)\\
\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}}\mathbf{Z}_N &= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)'\\
&= \frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)'+o_p(1),\\
\frac{1}{\sqrt{N}}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}}\boldsymbol{\eta}_N^{\boldsymbol{v}} &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)\eta_i^{v}=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)\eta_i^{v}+o_p(1).
\end{aligned}
$$

Step 3. The consequence of these two approximations is that $\sqrt{N}(\hat{\beta}_{2SLS} - \hat{\beta}^{\text{inf}}_{2SLS}) = o_p(1)$. Finally in Step 3, we derive the limiting distribution of the infeasible estimator $\sqrt{N}(\hat{\beta}^{\text{inf}}_{2SLS} - \beta^0)$.

Supplementary Appendix

to the paper

Estimation of Peer Effects in Endogenous Social Networks: Control Function Approach (2017)

Ida Johnsson* and Roger Moon†

We use the following notation. M denotes a finite generic constant and $a \perp b$ means that a and b are orthogonal to each other. For an $N \times N$ matrix $\mathbf{A}$, we define matrix norms as follows: $\|\mathbf{A}\| = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2}$ denotes the Frobenius norm, $\|\mathbf{A}\|_o$ denotes the operator norm of matrix $\mathbf{A}$, that is, $\|\mathbf{A}\|_o = \lambda_{\max}(\mathbf{A}'\mathbf{A})^{1/2}$, $\lambda_{min}(\mathbf{A})$ denotes the minimum eigenvalue of $\mathbf{A}$. Notice that

$$\|\mathbf{A}\|_o \leq \|\mathbf{A}\| \leq \|\mathbf{A}\|_o \operatorname{rank}(\mathbf{A}). \tag{S.0.0.1}$$

Further, for matrix $\mathbf{A}$, $[\mathbf{a}]_i$ denotes the i'th row of $\mathbf{A}$. Denote $[\mathbf{G}\mathbf{X}_1]_i$ by $\mathbf{X}_{1,G,i}$, $[\mathbf{G}^2\mathbf{X}_1]_i$ by $\mathbf{X}_{1,G^2,i}$, $[\mathbf{G}\mathbf{y}]^i$ by $\mathbf{Y}_{G,i}$. The ith row of the instrument matrix $\mathbf{Z}_N$ is given by $\mathbf{z}_i' = [\mathbf{X}_{2,i}', \mathbf{X}_{1,G,i}, \mathbf{X}_{1,G^2,i}]$, $\mathbf{z}_i$ is $(3l_x) \times 1$. Similarly, $\mathbf{w}_i' = [\mathbf{Y}_{G,i}, \mathbf{X}_{1,i}', \mathbf{X}_{1,G,i}]$. We denote matrices by uppercase bold letters and vectors by lowercase bold letters, $\mathbf{Z}_N = (\mathbf{Z}_1', \ldots, \mathbf{Z}_N')'$, $\mathbf{W}_N = (\mathbf{W}_1', \ldots, \mathbf{W}_N')'$ and $\mathbf{a}_N = (a_1, \ldots, a_N)'$.

Appendix S.1. For $\hat{\beta}_{2SLS}$

Outline of the proof of Theorem 5.1: By definition, we have

$$\begin{aligned} \hat{\beta}_{2SLS} - \beta^0 &= \left(\mathbf{W}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\left(\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{W}_N\right)^{-1} \\ &\quad \times \mathbf{W}_N' M_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\left(\mathbf{Z}'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\left(\boldsymbol{\eta}_N^{v} - \mathbf{h}^{v}(\mathbf{a}_N) - \widehat{\mathbf{Q}}_N\boldsymbol{\alpha}_{K_N}^{v}\right). \end{aligned}$$

The derivation of the asymptotic distribution of $\hat{\beta}_{2SLS}$ consists of three steps.

Step 1. First, we control the sampling error coming from the fact that we do not observe $\mathbf{a}_N$ and approximate it with $\hat{\mathbf{a}}_N$. Under suitable assumptions (see Appendix S.1.1), we

*University of Southern California, johnsson@usc.edu

†University of Southern California and Yonsei University, moonr@usc.edu (corresponding author)

show that the error that stems from the estimation of $\mathbf{a}_N$ by $\hat{\mathbf{a}}_N$ is asymptotically negligible:

$$
\begin{aligned}
&\sqrt{N}\left(\hat{\beta}_{2SLS}-\beta^0\right)\\
=\;&\left(\frac{1}{N}\mathbf{W}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\left(\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\right)^{-1}\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{W}_N\right)^{-1}\\
&\times\frac{1}{N}\mathbf{W}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\left(\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\mathbf{Z}_N\right)^{-1}\frac{1}{\sqrt{N}}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\boldsymbol{\eta}_N^v+o_p(1).
\end{aligned}
$$

(See Lemma 2 in Appendix S.1.1)

Step 2. Next, we consider the error introduced by the non-parametric estimation of $h(a_i)$. Let $\mathbf{h}^{\mathbf{w}}(a_i)=\mathbb{E}(\mathbf{w}_i|a_i)$, $\eta_i^{\mathbf{w}}=\mathbf{w}_i-\mathbf{h}^{\mathbf{w}}(a_i)$, $\mathbf{h}^{\mathbf{z}}(a_i)=\mathbb{E}(\mathbf{z}_i|a_i)$ and $\eta_i^{\mathbf{z}}=\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)$. Let $\hat{\mathbf{h}}^{\mathbf{w}}(a_i)$ and $\hat{\mathbf{h}}^{\mathbf{z}}(a_i)$ denote the series approximation of $\mathbf{h}^{\mathbf{w}}(a_i)$ and $\mathbf{h}^{\mathbf{z}}(a_i)$, respectively. In Lemma 7 in Appendix S.1.2 we show that under the regularity conditions (see Appendix S.1.2), the error from estimating $h(a_i)$ with $\hat{h}(a_i)$ converges to zero at a suitable rate and we have

$$
\begin{aligned}
\frac{1}{N}\mathbf{W}_N'\mathbf{M}_{\mathbf{Q}}\mathbf{Z}_N &= \frac{1}{N}\sum_{i=1}^N\left(\mathbf{w}_i-\hat{\mathbf{h}}^{\mathbf{w}}(a_i)\right)\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)'\\
&= \frac{1}{N}\sum_{i=1}^N\left(\mathbf{w}_i-\mathbf{h}^{\mathbf{w}}(a_i)\right)\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)'+o_p(1)\\
\frac{1}{N}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}}\mathbf{Z}_N &= \frac{1}{N}\sum_{i=1}^N\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)'\\
&= \frac{1}{N}\sum_{i=1}^N\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)'+o_p(1),\\
\frac{1}{\sqrt{N}}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}}\boldsymbol{\eta}_N^{\boldsymbol{v}} &= \frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\mathbf{z}_i-\hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)\eta_i^v=\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i)\right)\eta_i^v+o_p(1).
\end{aligned}
$$

Step 3. The consequence of these two approximation is that $\sqrt{N}(\hat{\beta}_{2SLS}-\hat{\beta}_{2SLS}^{\inf})=o_p(1)$. Finally in Step 3, we derive the limiting distribution of the infeasible estimator $\sqrt{N}(\hat{\beta}_{2SLS}^{\inf}-\beta^0)$.

S.1.1. **Controlling the Sampling Error $\hat{a}_i - a_i$ in Sieve Estimation.** In this section, we show that the error coming from the estimation of a_i by $\hat{a}_i$ is of order $o_p(1)$. All supporting Lemmas can be found in Appendix S.1.1.1.

Lemma 2. *Assume Assumptions 1 2, 7, 8, and 11. Then the following hold.*

(a) $\frac{1}{N}(\mathbf{Z}'_N\mathbf{P}_{\widehat{\mathbf{Q}}_N}\mathbf{W}_N - \mathbf{Z}'_N\mathbf{P}_{\mathbf{Q}_N}\mathbf{W}_N) = o_p(1)$.
(b) $\frac{1}{N}(\mathbf{Z}'_N\mathbf{P}_{\widehat{\mathbf{Q}}_N}\mathbf{Z}_N - \mathbf{Z}'_N\mathbf{P}_{\mathbf{Q}_N}\mathbf{Z}_N) = o_p(1)$.
(c) $\frac{1}{\sqrt{N}}(\mathbf{Z}'_N\mathbf{P}_{\widehat{\mathbf{Q}}_N}\boldsymbol{\eta}^v_N - \mathbf{Z}'_N\mathbf{P}_{\mathbf{Q}_N}\boldsymbol{\eta}^v_N) = o_p(1)$.
(d) $\frac{1}{\sqrt{N}}(\mathbf{Z}'\mathbf{M}_{\widehat{\mathbf{Q}}_N}(\mathbf{h}^v(\mathbf{a}_N) - \widehat{\mathbf{Q}}_N\boldsymbol{\alpha}^v_{K_N})) = o_p(1)$.

Proof. Part (a).

$$
\begin{aligned}
&\frac{1}{N}(\mathbf{Z}'_N\mathbf{P}_{\widehat{\mathbf{Q}}_N}\mathbf{W}_N - \mathbf{Z}'_N\mathbf{P}_{Q_N}\mathbf{W}_N)\\
= \;& \frac{\mathbf{Z}'_N\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)}{N}\left(\frac{\widehat{\mathbf{Q}}'_N\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{\widehat{\mathbf{Q}}'_N\mathbf{W}_N}{N} - \frac{\mathbf{Z}'_N\mathbf{Q}_N}{N}\left\{\left(\frac{\mathbf{Q}'_N\mathbf{Q}_N}{N}\right)^{-1} - \left(\frac{\widehat{\mathbf{Q}}'_N\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\right\}\frac{\mathbf{Q}'_N\mathbf{W}_N}{N}\\
&+\frac{\mathbf{Z}'_N\mathbf{Q}_N}{N}\left(\frac{\widehat{\mathbf{Q}}'_N\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)'\mathbf{W}_N}{N}\\
= \;& \frac{\mathbf{Z}'_N\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)}{N}\left(\frac{\widehat{\mathbf{Q}}'_N\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N)'\mathbf{W}_N}{N} + \frac{\mathbf{Z}'_N\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)}{N}\left(\frac{\widehat{\mathbf{Q}}'_N\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{\mathbf{Q}'_N\mathbf{W}_N}{N}\\
&-\frac{\mathbf{Z}'_N\mathbf{Q}_N}{N}\left\{\left(\frac{\mathbf{Q}'_N\mathbf{Q}_N}{N}\right)^{-1} - \left(\frac{\widehat{\mathbf{Q}}'_N\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\right\}\frac{\mathbf{Q}'_N\mathbf{W}_N}{N} + \frac{\mathbf{Z}'_N\mathbf{Q}_N}{N}\left(\frac{\widehat{\mathbf{Q}}'_N\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)'\mathbf{W}_N}{N}\\
= \;& I_1 + I_2 - I_3 + I_4, \textit{say}.
\end{aligned}
$$

For the desired result, by (S.0.0.1) we show that

$$\left\|\frac{1}{N}(\mathbf{Z}'_N\mathbf{P}_{\widehat{\mathbf{Q}}_N}\mathbf{W}_N - \mathbf{Z}'_N\mathbf{P}_{\mathbf{Q}_N}\mathbf{W}_N)\right\|_o = o_p(1),$$

which follows by triangular inequality if we show

$$\|I_1\|_o, \|I_2\|_o, \|I_3\|_o, \|I_4\|_o = o_p(1).$$

For term I_1,

$$\begin{aligned}\|I_1\|_o &\leq \left\|\frac{\mathbf{Z}_N}{\sqrt{N}}\right\| \left\|\frac{\widehat{\mathbf{Q}}_N - \mathbf{Q}_N}{\sqrt{N}}\right\|^2 \left\|\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\right\|_o \left\|\frac{\mathbf{W}_N}{\sqrt{N}}\right\| \\ &= O_p(1)\left(\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1(k)^2\right) O_P(1)O(1) = o_p(1),\end{aligned}$$

where the last line holds by (S.1.1.1), Lemmas 4 and 6, and by Assumption 8.

For term I_2,

$$\begin{aligned}\|I_2\|_o &\leq \left\|\frac{\mathbf{Z}_N}{\sqrt{N}}\right\| \left\|\frac{\widehat{\mathbf{Q}}_N - \mathbf{Q}_N}{\sqrt{N}}\right\| \left\|\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\right\|_o \left\|\frac{\mathbf{Q}_N}{\sqrt{N}}\right\| \left\|\frac{\mathbf{W}_N}{\sqrt{N}}\right\| \\ &= O_p(1)\left(\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1(k)^2\right)^{1/2} O_P(1)\zeta_0(K_N)O(1) = o_p(1),\end{aligned}$$

where the last line holds by (S.1.1.1), Lemmas 4 and 6, and by Assumption 8.

For term I_3, write

$$\begin{aligned}I_3 &= \frac{\mathbf{Z}_N'\mathbf{Q}_N}{N}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\left\{\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right) - \left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)\right\}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)^{-1}\frac{\mathbf{Q}_N'\mathbf{W}_N}{N} \\ &= \frac{\mathbf{Z}_N'\mathbf{Q}_N}{N}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\left(\frac{\widehat{\mathbf{Q}}_N'(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N)}{N}\right)\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)^{-1}\frac{\mathbf{Q}_N'\mathbf{W}_N}{N} \\ &\quad + \frac{\mathbf{Z}_N'\mathbf{Q}_N}{N}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\left(\frac{(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N)'\mathbf{Q}_N}{N}\right)\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)^{-1}\frac{\mathbf{Q}_N'\mathbf{W}_N}{N}.\end{aligned}$$

Then,

$$\|I_3\|_o \leq O_p(1)\zeta_0(K_N)O_p(1)\zeta_0(K_N)\left(\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1(k)^2\right)^{1/2} O_p(1)\zeta_0(K_N)O_p(1) = o_p(1),$$

where the last equality follows by Assumption 8.

The desired result of term I_4 follows by similar argument used for term I_2.

Part (b) can be shown in a similar way as Part (a).

Part (c).

$$\begin{aligned}
&\frac{1}{\sqrt{N}}(\mathbf{Z}_N'\mathbf{P}_{\widehat{\mathbf{Q}}_N}\boldsymbol{\eta}_N^{\upsilon} - \mathbf{Z}_N'\mathbf{P}_{Q_N}\boldsymbol{\eta}_N^{\upsilon}) \\
= \;& \frac{\mathbf{Z}_N'\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)}{N}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N)'\boldsymbol{\eta}_N^{\upsilon}}{\sqrt{N}} + \frac{\mathbf{Z}_N'\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)}{N}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{\mathbf{Q}_N'\boldsymbol{\eta}_N^{\upsilon}}{\sqrt{N}} \\
& -\frac{\mathbf{Z}_N'\mathbf{Q}_N}{N}\left\{\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)^{-1} - \left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\right\}\frac{\mathbf{Q}_N'\boldsymbol{\eta}_N^{\upsilon}}{\sqrt{N}} + \frac{\mathbf{Z}_N'\mathbf{Q}_N}{N}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\frac{\left(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right)'\boldsymbol{\eta}_N^{\upsilon}}{\sqrt{N}} \\
= \;& III_1 + III_2 - III_3 + III_4, say,
\end{aligned}$$

and the desired result of Part (c) follows if we show that for $j = 1, ..., 4$,

$$\|III_j\| = o_p(1).$$

First, for term III_1, we have

$$\begin{aligned}
\|III_1\| \;&\leq\; \left\|\frac{\mathbf{Z}_N}{\sqrt{N}}\right\|\left\|\frac{\widehat{\mathbf{Q}}_N - \mathbf{Q}_N}{\sqrt{N}}\right\|\left\|\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right)^{-1}\right\|\left\|\frac{(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N)'\boldsymbol{\eta}_N^{\upsilon}}{\sqrt{N}}\right\| \\
&=\; O_p(1)\left(\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1(k)^2\right)^{1/2} O_p(1)\left\|\frac{(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N)'\boldsymbol{\eta}_N^{\upsilon}}{\sqrt{N}}\right\|,
\end{aligned}$$

where the last line holds by (S.1.1.1), Lemmas 4 and 6. Under Assumption we can show that

$$\mathbb{E}\left[\left\|\frac{(\widehat{\mathbf{Q}}_N - \mathbf{Q}_N)'\boldsymbol{\eta}_N^{\upsilon}}{\sqrt{N}}\right\|^2 \,\middle|\, \mathbf{X}_{1N}, \mathbf{G}_N, \mathbf{a}_N\right] = \frac{1}{N}\left\|\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\right\|^2.$$

Then, by Lemma 4 and Assumption 8, we have the required result for term III_1.

The rest of the required results follow by similar fashion and we omit the proof.

Part (d).

Notice that

$$
\begin{aligned}
& \frac{1}{\sqrt{N}}(\mathbf{Z}_N'\mathbf{M}_{\widehat{Q}_N}(\mathbf{h}^{\upsilon}(\mathbf{a}_N) - \widehat{\mathbf{Q}}_N\boldsymbol{\alpha}_{K_N}^{\upsilon})) \\
= \ & \frac{1}{\sqrt{N}}\mathbf{Z}_N'\mathbf{M}_{\widehat{\mathbf{Q}}_N}\mathbf{h}^{\upsilon}(\mathbf{a}_N) \\
= \ & \frac{1}{\sqrt{N}}\mathbf{Z}_N'\left(\mathbf{M}_{\widehat{\mathbf{Q}}_N} - \mathbf{M}_{\mathbf{Q}_N}\right)\mathbf{h}^{\upsilon}(\mathbf{a}_N) + \frac{1}{\sqrt{N}}\mathbf{Z}_N'\mathbf{M}_{\mathbf{Q}_N}\left(\mathbf{h}^{\upsilon}(\mathbf{a}_N) - \mathbf{Q}_N\boldsymbol{\alpha}_{K_N}^{\upsilon}\right) \\
= \ & IV_1 + IV_2, say.
\end{aligned}
$$

We can show $IV_1 = o_p(1)$ by applying similar arguments used in the proof of Part (a). For term IV_2, notice that

$$
\begin{aligned}
\|IV_2\| &= \|IV_2\|_o \\
&\leq \left\|\frac{1}{\sqrt{N}}\mathbf{Z}_N\right\|_o \|\mathbf{M}_{\mathbf{Q}_N}\|_o \left\|\mathbf{h}^{\upsilon}(\mathbf{a}_N) - \mathbf{Q}_N\boldsymbol{\alpha}_{K_N}^{\upsilon}\right\|_o \\
&= \left\|\frac{1}{\sqrt{N}}\mathbf{Z}_N\right\| \left\|\mathbf{h}^{\upsilon}(\mathbf{a}_N) - \mathbf{Q}_N\boldsymbol{\alpha}_{K_N}^{\upsilon}\right\| \\
&= O_p(1)\sqrt{N}O(K_N^{-\kappa}) = o_p(1)
\end{aligned}
$$

by Assumption 7 (iii) and (iv).

□

S.1.1.1. *Supporting Lemmas.* First notice that by the boundedness condition (ii) and (iii) in Assumption 11, we have

$$
\frac{1}{N}\|\mathbf{Z}_N\|^2 = O_p(1),\ \frac{1}{N}\|\mathbf{W}_N\|^2 = O_p(1). \tag{S.1.1.1}
$$

Lemma 3. *Under Assumption 7, we have*

$$
\frac{1}{N}\|\mathbf{Q}_N\|^2 \leq M\zeta_0^2(K_N).
$$

Proof.

$$
\frac{1}{N}\|\mathbf{Q}_N\|^2 = \frac{1}{N}\sum_{i=1}^{N}\|\mathbf{q}^K(a_i)\|^2 \leq \sup_i \|\mathbf{q}^K(a_i)\|^2 = \zeta_0^2(K_N)
$$

by Assumption 7 (ii). □

Lemma 4. *Under Assumptions 1, 5, 7, and 8, we have*

$$\frac{1}{N}\|\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\|^2 = M\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1(k)^2.$$

Proof.

$$\begin{aligned}\frac{1}{N}\|\widehat{\mathbf{Q}}_N - \mathbf{Q}_N\|^2 &= \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_N}\|q_k(\hat{a}_i) - q_k(a_i)\|^2 \leq \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_N}\zeta_1(k)^2\|\hat{a}_i - a_i\|^2 \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K_N}\zeta_1(k)^2\frac{1}{\zeta_a(N)^2} = \frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1(k)^2,\end{aligned}$$

where the first inequality follows from Assumption 8 and the second inequality follows from Assumption 5. □

Lemma 5. *For symmetric matrices* $\mathbf{A}$ *and* $\mathbf{B}$ *it is true that*

$$|\lambda_{min}(\mathbf{A}) - \lambda_{min}(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|$$

Proof. Let $\underline{\mathbf{x}}_A$ be the eigenvector associated with the minimum eigenvalue of $\mathbf{A}$. Define $\underline{x\mathbf{x}}_B$ analogously. First we show $|\lambda_{min}(\mathbf{A}) - \lambda_{min}(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|$.

$$\begin{aligned}\lambda_{min}(\mathbf{A}) - \lambda_{min}(\mathbf{B}) &= \underline{\mathbf{x}}_A'\mathbf{A}\underline{\mathbf{x}}_A - \underline{x}_B'\mathbf{B}\underline{\mathbf{x}}_B \\ &\leq \underline{\mathbf{x}}_B'(\mathbf{A} - \mathbf{B})\underline{\mathbf{x}}_B \\ &\leq |\underline{\mathbf{x}}_B'(\mathbf{A} - \mathbf{B})\underline{\mathbf{x}}_B| \leq \|\mathbf{A} - \mathbf{B}\|.\end{aligned}$$

Also, we can prove the other direction. Notice that

$$\begin{aligned}\lambda_{min}(\mathbf{A}) - \lambda_{min}(B) &= \underline{\mathbf{x}}_A'\mathbf{A}\underline{x}_A - \underline{\mathbf{x}}_B'\mathbf{B}\underline{\mathbf{x}}_B \\ &\geq \underline{\mathbf{x}}_A'(\mathbf{A} - \mathbf{B})\underline{\mathbf{x}}_A \\ &\geq -|\underline{\mathbf{x}}_B'(\mathbf{A} - \mathbf{B})\underline{\mathbf{x}}_B| \geq -\|\mathbf{A} - \mathbf{B}\|.\end{aligned}$$

Then, we have the required result.

$\square$

Lemma 6. *Under 1, 5, 7, and 8, W.p.a.1, there exists a positive constant $C > 0$ such that*

$$\frac{1}{C} \leq \lambda_{min}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right),\ \lambda_{min}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right).$$

Proof. First we show that there exists a positive constant C such that $\frac{1}{C} \leq \lambda_{min}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)$, which follows by Assumption 7(i) if we show

$$\left|\lambda_{min}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right) - \mathbb{E}[\mathbf{q}^{K_N}(a_i)\mathbf{q}^{K_N}(a_i)']\right| = o_p(1).$$

For this, by Lemma 5, we have

$$\begin{aligned}
&\left|\lambda_{min}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right) - \mathbb{E}[\mathbf{q}^{K_N}(a_i)\mathbf{q}^{K_N}(a_i)']\right| \leq \left\|\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N} - \mathbb{E}[\mathbf{q}^{K_N}(a_i)\mathbf{q}^{K_N}(a_i)']\right\| \\
= &\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{q}^{K_N}(a_i)\mathbf{q}^{K_N}(a_i)' - \mathbb{E}[\mathbf{q}^{K_N}(a_i)\mathbf{q}^{K_N}(a_i)']\right)\right\|.
\end{aligned}$$

Then, by Assumption 7(ii), we have

$$\begin{aligned}
&\mathbb{E}\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{q}^{K_N}(a_i)\mathbf{q}^{K_N}(a_i)' - \mathbb{E}[\mathbf{q}^{K_N}(a_i)\mathbf{q}^{K_N}(a_i)']\right)\right\|^2 \\
= &\sum_{k=1}^{K_N}\sum_{l=1}^{K_N}\mathbb{E}\left(\frac{1}{N}\sum_{i=1}^{N}\left(\mathbf{q}_k(a_i)\mathbf{q}_l(a_i) - \mathbb{E}[\mathbf{q}_k(a_i)\mathbf{q}_l(a_i)]\right)\right)^2 \\
\leq &\frac{1}{N}\sum_{k=1}^{K_N}\sum_{l=1}^{K_N}\mathbb{E}[\mathbf{q}_k(a_i)\mathbf{q}_l(a_i)]^2 \leq \frac{1}{N}\sup_a\left(\sum_{k=1}^{K_N}\mathbf{q}_k(a)^2\right)^2 \\
\leq &\frac{\zeta_0(K_N)^4}{N} = o(1),
\end{aligned}$$

where the last line holds by Assumptions 7(ii) and 8.

Next, given the first part of the lemma, the second claim of the lemma follows if we show

$$\left|\lambda_{min}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right) - \lambda_{min}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)\right| = o_p(1).$$

Notice by Lemma 5, for symmetric matrices $\sqrt{A}$ and $\mathbf{B}$, we have

$$\|\lambda_{min}(\mathbf{A}) - \lambda_{min}(\mathbf{B})\| \leq \|\mathbf{A} - \mathbf{B}\|.$$

Then,

$$\begin{aligned}
\left|\lambda_{min}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right) - \lambda_{min}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)\right| &\leq \left\|\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N} - \frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right\| \\
&\leq \left\|\frac{(\hat{\mathbf{Q}}_N - \mathbf{Q}_N)'}{\sqrt{N}}\frac{\mathbf{Q}_N}{\sqrt{N}}\right\| + \left\|\frac{\mathbf{Q}_N'}{\sqrt{N}}\frac{(\hat{\mathbf{Q}}_N - \mathbf{Q}_N)}{\sqrt{N}}\right\| \\
&+ \left\|\frac{(\hat{\mathbf{Q}}_N - \mathbf{Q}_N)'}{\sqrt{N}}\frac{(\hat{\mathbf{Q}}_N - \mathbf{Q}_N)}{\sqrt{N}}\right\|.
\end{aligned}$$

Then, by lemmas 3 and 4 and by Assumption 8, we have

$$\left|\lambda_{min}\left(\frac{\widehat{\mathbf{Q}}_N'\widehat{\mathbf{Q}}_N}{N}\right) - \lambda_{min}\left(\frac{\mathbf{Q}_N'\mathbf{Q}_N}{N}\right)\right| \leq M\left(\zeta_0(K_N)\sqrt{\frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1^2(k)} + \frac{1}{\zeta_a(N)^2}\sum_{k=1}^{K_N}\zeta_1^2(k)\right) = o_p(1),$$

as desired. □

S.1.2. **Controlling the Series Approximation Error.**

Lemma 7 (Series Approximation)**.** *Assume the assumptions in Lemma 2. Then, we have*

(a) $\frac{1}{N}\sum_{i=1}^N\left(\mathbf{w}_i - \hat{\mathbf{h}}^{\mathbf{w}}(a_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)' = \frac{1}{N}\sum_{i=1}^N\left(\mathbf{w}_i - \mathbf{h}^{\mathbf{w}}(a_i)\right)\left(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i)\right)' + o_p(1),$

(b) $\frac{1}{N}\sum_{i=1}^N\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)' = \frac{1}{N}\sum_{i=1}^N\left(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i)\right)\left(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i)\right)' + o_p(1),$

(c) $\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}(a_i)\right)\eta_i^v = \frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i)\right)\eta_i^v + o_p(1).$

Proof. Lemma 7 follows if we show

(i) $\frac{1}{N}\sum_{i=1}^N\left(\hat{\mathbf{h}}^{\mathbf{w}}(a_i) - \mathbf{h}^{\mathbf{w}}(a_i)\right)\left(\hat{\mathbf{h}}^{\mathbf{w}}(a_i) - \mathbf{h}^{\mathbf{w}}(a_i)\right)' = o_p(1).$

(ii) $\frac{1}{N}\sum_{i=1}^N\left(\hat{\mathbf{h}}^{\mathbf{z}}(a_i) - \mathbf{h}^{\mathbf{z}}(a_i)\right)\left(\hat{\mathbf{h}}^{\mathbf{z}}(a_i) - \mathbf{h}^{\mathbf{z}}(a_i)\right)' = o_p(1).$

(iii) $\frac{1}{\sqrt{N}}\sum_{i=1}^N\left(\hat{\mathbf{h}}^{\mathbf{z}}(a_i) - \mathbf{h}^{\mathbf{z}}(a_i)\right)\eta_i^v = o_p(1).$

Lemma 7 (i) and (ii) is true by Lemma 10 and Lemma 7 (iii) follows from (ii). See the remainder of this section. □

Following Newey (1997), we assume $\mathbf{B} = \mathbf{I}$ in Assumption 7, hence, $\tilde{q}^K(a) = q^K(a)$. Also, we assume $P = \mathbb{E}[\mathbf{q}^K(a_i)(\mathbf{q}^K(a_i))'] = I$.[6]

Lemma 8. *Assume Assumption 7. Then,* $\mathbb{E}[\|\tilde{\mathbf{P}} - \mathbf{I}\|^2] = O(\zeta_0(K_N)^2 K_N/N)$, *where* $\tilde{\mathbf{P}} = (\mathbf{Q}_N'\mathbf{Q}_N)/N$.

Proof. For proof see Li and Racine (2007) page 481. □

Note that this Lemma implies that $\|\tilde{\mathbf{P}} - \mathbf{I}\| = O_p(\zeta_0(K_N)\sqrt{K_N/N} = o_p(1)$. Also, since the smallest eigenvalue of $\tilde{\mathbf{P}} - \mathbf{I}$ is bounded by $\|\tilde{\mathbf{P}} - \mathbf{I}\|$, this implies that the smallest eigenvalue of $\tilde{\mathbf{P}}$ converges to one in probability. Letting $\mathbf{1}_N$ be the indicator function for the smallest eigenvalue of $\tilde{\mathbf{P}}$ being greater than 1/2, we have $\Pr(\mathbf{1}_N = 1) \to 1$.

Lemma 9. *Assume Assumption 7. Then,* $\|\tilde{\alpha}^f - \alpha^f\| = O_p(K_N^{-\kappa})$, *where* $\tilde{\alpha}^f = (\mathbf{Q}_N'\mathbf{Q}_N)^{-1}\mathbf{Q}_N' f$, *where* $\alpha^{(f)}$ *satisfies Assumption 7 and* $f(a) \in \{h^y(a), h^{\mathbf{z}}(a), h^{\mathbf{w}}(a)\}$.

Proof.

$$\begin{aligned}
\mathbf{1}_N\|\tilde{\alpha}^{(f)} - \alpha^{(f)}\| &= \mathbf{1}_N\|(\mathbf{Q}_N'\mathbf{Q}_N)^{-1}\mathbf{Q}_N'(f - \mathbf{Q}_N\alpha^f)\| \\
&= \mathbf{1}_N\{(f - \mathbf{Q}_N\alpha^f)'\mathbf{Q}_N(\mathbf{Q}_N'\mathbf{Q}_N)^{-1}(\mathbf{Q}_N'\mathbf{Q}_N/N)^{-1}\mathbf{Q}_N'(f - \mathbf{Q}_N\alpha^f)/N\}^{1/2} \\
&= \mathbf{1}_N O_P(1)\{(f - \mathbf{Q}_N\alpha^{(f)})'\mathbf{Q}_N(\mathbf{Q}_N'\mathbf{Q}_N)^{-1}\mathbf{Q}_N'(f - \mathbf{Q}_N\alpha^f)/N\}^{1/2} \\
&\leq O_p(1)\{(f - \mathbf{Q}_N\alpha^f)'(f - \mathbf{Q}_N\alpha^f)/N\}^{1/2} = O_p(K_N^{-\kappa})
\end{aligned}$$

by Lemma 8, Assumption 7(iii), the fact that $\mathbf{Q}_N(\mathbf{Q}_N'\mathbf{Q}_N)^{-1}\mathbf{Q}_N'$ is idempotent and $\Pr(\mathbf{1}_N = 1) \to 1$. □

Lemma 10. *Assume Assumption 7. Let* $f(a) \in (h^y(a), \mathbf{h}^z\mathbf{z}(a), \mathbf{h}^{\mathbf{w}}(a))$ *and* $\tilde{f} = \mathbf{Q}_N\tilde{\boldsymbol{\alpha}}_N^f$. *Then,* $\frac{1}{N}\|f - \tilde{f}\|^2 = O_p(K_N^{-2\kappa}) = o_p(N^{-1/2})$.

[6] The Lemmas in this section follow Section 15.6 in Li and Racine (2007).

Proof. The required result for the lemma follows because

$$\begin{aligned}
\frac{1}{N}\|f-\tilde{f}\|^2 &\leq \frac{1}{N}\{\|f-\mathbf{Q}_N\boldsymbol{\alpha}_N^f\|^2+\|\mathbf{Q}_N(\boldsymbol{\alpha}_N^{(f)}-\tilde{\boldsymbol{\alpha}}_N^f)\|^2\}\\
&= O(K_N^{-2\kappa})+(\boldsymbol{\alpha}_N^f-\tilde{\boldsymbol{\alpha}}_N^f)'(\mathbf{Q}_N'\mathbf{Q}_N/N)(\boldsymbol{\alpha}_N^f-\tilde{\boldsymbol{\alpha}}_N^f)\\
&= O(K_N^{-2\kappa})+O_p(1)\|\boldsymbol{\alpha}_N^f-\tilde{\boldsymbol{\alpha}}_N^f\|^2=O_p(K_N^{-2\kappa})
\end{aligned}$$

by Assumption 7(iii), Lemma 8 and Lemma 9. □

S.1.3. **Limiting Distribution of $\hat{\beta}_{2SLS}$.** In this section we derive the distribution of the infeasible estimator $\hat{\beta}_{2SLS}^{inf}$. All supporting lemmas can be found in Section S.1.4.

We introduce the following notation. Let $s_0(\mathbf{x}_i,a_i)$ be a function of $(\mathbf{x}_i,a_i)$ such that $s_0(\cdot,\cdot)$ is bounded over the support of $(\mathbf{x}_i,a_i)$. We denote an N vector-valued function that stacks $s_0(\mathbf{x}_i,a_i)$ over $i=1,...,N$ as $\mathbf{S}_{0,N}=(s_0(\mathbf{x}_1,a_1),\ldots,s_0(\mathbf{x}_N,a_N))'$. Define

$$s_{0,N,i}:=s_0(\mathbf{x}_i,a_i). \tag{S.1.3.1}$$

Next, for $m=1,2,...$, we define recursively

$$s_{m,N,i}:=\sum_{j=1,\neq i}^{N}g_{ij,N}s_{m-1,N,i}=[\mathbf{G}_N\mathbf{S}_{m-1,N}]_i, \tag{S.1.3.2}$$

where

$$\mathbf{S}_{m-1,N}:=(s_{m-1,N,1},\ldots,s_{m-1,N,N})'.$$

For $m=0,1,2,...$, we define $s_{m,N,i}^{\mathbf{x}_1}$ and $\mathbf{S}_{m,N}^{\mathbf{x}_1}$ with initial function $s_{0,N,i}=s_0(\mathbf{x}_i,a_i)=\mathbf{x}_{1i}$, and define $s_{m,N,i}^{a}$ and $\mathbf{S}_{m,N}^{a}$ with initial function $s_{0,N,i}=s_0(\mathbf{x}_i,a_i)=h^{\upsilon}(a_i)$.

Next, we define recursively the probability limit of $s_{m,N,i}$ defined with the initial function $s_{0,N,i}=s_0(\mathbf{x}_i,a_i)$ for each i as $N\to\infty$. For this, let

$$\tilde{s}_0(\mathbf{x}_i,a_i)=s_0(\mathbf{x}_i,a_i)=s_{0,N,i}.$$

Note that for fixed i, $s_{1,N,i}$ has the following limit as $N \to \infty$:

$$
\begin{aligned}
s_{1,N,i} &= [\mathbf{G}_N \mathbf{S}_{0,N}]_i \\
&= \left(\frac{1}{N} \sum_{j \neq i} d_{ij,N} \right)^{-1} \frac{1}{N} \sum_{j \neq i} d_{ij,N} s_0(\mathbf{x}_j, a_j) \\
&= \left(\frac{1}{N} \sum_{j \neq i} \mathbb{I}\left\{ g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) \geq u_{ij} \right\} \right)^{-1} \\
&\quad \times \frac{1}{N} \sum_{j \neq i} \mathbb{I}\left\{ g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) \geq u_{ij} \right\} s_0(\mathbf{x}_j, a_j) \\
&\xrightarrow{p} \frac{\int \int \int p(g(t(\mathbf{x}_{2i}, \mathbf{x}_2), a_i, a) s_0(\mathbf{x}, a) \pi(\mathbf{x}, a) d\mathbf{x} da}{\int \int p(g(t(\mathbf{x}_{2i}, \mathbf{x}_2), a_i, a) \pi(\mathbf{x}_2, a) d\mathbf{x}_2 da} \\
&= \frac{\mathbb{E}[d_{ij,N} s_0(\mathbf{x}_j, a_j) | \mathbf{x}_i, a_i]}{\mathbb{E}[d_{ij,N} | \mathbf{x}_i, a_i]} =: \tilde{s}_1(\mathbf{x}_i, a_i),
\end{aligned}
\tag{S.1.3.3}
$$

where $\pi(\mathbf{x}, a)$ with $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ is the joint density of $\mathbf{x}_i = (\mathbf{x}_{1i}, \mathbf{x}_{2i})$ and a_i, and $\pi(\mathbf{x}_2, a)$ is the joint density of $(\mathbf{x}_{2i}, a_i)$. Here note that the limit $\tilde{s}_1(\mathbf{x}_i, a_i)$ depends only on $(\mathbf{x}_i, a_i)$, not on $(\mathbf{x}_{-i}, a_{-i})$, while $s_{1,N,i}$ depends on both $(\mathbf{x}_i, a_i)$ and $(\mathbf{x}_{-i}, a_{-i})$.

We define the following recursively for $m = 2, 3, \cdots$ as follows:

$$
\begin{aligned}
\tilde{s}_m(\mathbf{x}_i, a_i) &:= \frac{\mathbb{E}[d_{ij,N} \tilde{s}_{m-1}(\mathbf{x}_j, a_j) | \mathbf{x}_i, a_i]}{\mathbb{E}[d_{ij,N} | \mathbf{x}_i, a_i]} \\
&= \frac{\int \int p(g(t(\mathbf{x}_{2i}, \mathbf{x}_2), a_i, a) \tilde{s}_{m-1}(\mathbf{x}, a) \pi(\mathbf{x}, a) d\mathbf{x} da}{\int \int p(t(\mathbf{x}_{2i}, \mathbf{x}_2), a_i, a) \pi(\mathbf{x}_2, a) d\mathbf{x}_2 da} \\
&= \operatorname*{plim}_{N \to \infty} \left(\frac{1}{N} \sum_{j \neq i} d_{ij,N} \right)^{-1} \frac{1}{N} \sum_{j \neq i} d_{ij,N} \tilde{s}_{m-1}(\mathbf{x}_j, a_j) \\
&= \operatorname*{plim}_{N \to \infty} [\mathbf{G}_N \tilde{\mathbf{S}}_{m-1}]_i,
\end{aligned}
\tag{S.1.3.4}
$$

where $\tilde{\mathbf{S}}_m = (\tilde{s}_m(\mathbf{x}_1, a_1), ..., \tilde{s}_m(\mathbf{x}_N, a_N))$.

Using this general definitions of (S.1.3.3) and (S.1.3.4), with $\tilde{s}_0^{\mathbf{x}_1}(\mathbf{x}_i, a_i) = s_0^{\mathbf{x}_1}(\mathbf{x}_i, a_i) = \mathbf{x}_{1i}$ and $\tilde{s}_0^{a}(\mathbf{x}_i, a_i) = s_0^{a}(\mathbf{x}_i, a_i) = h(a_i)$, we define $\tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i, a_i)$ and $\tilde{s}_m^{a}(\mathbf{x}_i, a_i)$, respectively, for $m = 1, 2, ...$ Let $\tilde{S}_m^{\mathbf{x}_1} = (\tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_1, a_1), \ldots \tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_N, a_N))'$. and $\tilde{S}_m^{a} = (\tilde{s}_m^{a}(\mathbf{x}_1, a_1), \ldots \tilde{s}_m^{a}(\mathbf{x}_N, a_N))'$.

Next, with the initial function $s^{\upsilon}_{0,N,i} = \eta^{\upsilon}_i$ and $\mathbf{S}^{\upsilon}_{0,N} := (s^{\eta}_{0,N,1}, \ldots, s^{\eta}_{0,N,N})'$, we define recursively

$$s^{\upsilon}_{m,N,i} := [\mathbf{G}_N \mathbf{S}^{\upsilon}_{m-1,N}]_i = \sum_{j=1,\neq i}^{N} g_{ij,N} s^{\upsilon}_{m-1,N,i}, \tag{S.1.3.5}$$

and $\mathbf{S}^{\upsilon}_{m,N} := (s^{\upsilon}_{m,N,1}, \ldots, s^{\upsilon}_{m,N,N})'$ for $m = 1, 2, ...$

Lemma 11. *Under Assumptions 1 and 11, as $N \to \infty$, we have*

$$(a) \frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}_i - \mathbf{h}^{\mathbf{w}}(a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))'$$

$$=: \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{GY}_i(\boldsymbol{\eta}^{\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{GY}_i(\boldsymbol{\eta}^{G\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{GY}_i(\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i)' \\ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{\mathbf{x}_1}_i(\boldsymbol{\eta}^{\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{\mathbf{x}_1}_i(\boldsymbol{\eta}^{G\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{\mathbf{x}_1}_i(\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i)' \\ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G\mathbf{x}_1}_i(\boldsymbol{\eta}^{\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G\mathbf{x}_1}_i(\boldsymbol{\eta}^{G\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G\mathbf{x}_1}_i(\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i)' \end{pmatrix}$$

$$\xrightarrow{p} \begin{pmatrix} S^{GY,\mathbf{x}_1} & S^{GY,G\mathbf{x}_1} & S^{GY,G^2\mathbf{x}_1} \\ S^{\mathbf{x}_1,\mathbf{x}_1} & S^{\mathbf{x}_1,G\mathbf{x}_1} & S^{\mathbf{x}_1,G^2\mathbf{x}_1} \\ S^{G\mathbf{x}_1,\mathbf{x}_1} & S^{G\mathbf{x}_1,G\mathbf{x}_1} & S^{G\mathbf{x}_1,G^2\mathbf{x}_1} \end{pmatrix} =: S^{\mathbf{wz}},$$

$$(b) \quad \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))'$$

$$=: \begin{pmatrix} \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{\mathbf{x}_1}_i(\boldsymbol{\eta}^{\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{\mathbf{x}_1}_i(\boldsymbol{\eta}^{G\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{\mathbf{x}_1}_i(\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i)' \\ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G\mathbf{x}_1}_i(\boldsymbol{\eta}^{\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G\mathbf{x}_1}_i(\boldsymbol{\eta}^{\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G\mathbf{x}_1}_i(\boldsymbol{\eta}^{G\mathbf{x}_1}_i)' \\ \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i(\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i(\boldsymbol{\eta}^{\mathbf{x}_1}_i)' & \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i(\boldsymbol{\eta}^{G^2\mathbf{x}_1}_i)' \end{pmatrix}$$

$$\xrightarrow{p} \begin{pmatrix} S^{\mathbf{x}_1,\mathbf{x}_1} & S^{\mathbf{x}_1,G\mathbf{x}_1} & S^{\mathbf{x}_1,G^2\mathbf{x}_1} \\ S^{G\mathbf{x}_1,\mathbf{x}_1} & S^{G\mathbf{x}_1,G\mathbf{x}_1} & S^{G\mathbf{x}_1 G^2\mathbf{x}_1} \\ S^{G^2\mathbf{x}_1,\mathbf{x}_1} & S^{G^2\mathbf{x}_1,G\mathbf{x}_1} & S^{G^2\mathbf{x}_1,G^2\mathbf{x}_1} \end{pmatrix} =: S^{\mathbf{zz}},$$

where

$$S^{GY,G^r\mathbf{x}_1} = \mathbb{E}\left[\left(\sum_{m=0}^{\infty}\beta_2^{0\prime}\tilde{\tilde{s}}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i)+\beta_3^{0\prime}\tilde{\tilde{s}}_{m+1}^{\mathbf{x}_1}(\mathbf{x}_i,a_i)+\tilde{\tilde{s}}_m^{a}(\mathbf{x}_i,a_i)\right)\left(\tilde{\tilde{s}}_r^{\mathbf{x}_1}(\mathbf{x}_i,a_i)\right)'\right],\ r=0,1,2$$

$$S^{G^r\mathbf{x}_1,G^s\mathbf{x}_1} = \mathbb{E}\left[\tilde{\tilde{s}}_r^{\mathbf{x}_1}(\mathbf{x}_i,a_i))\left(\tilde{\tilde{s}}_s^{\mathbf{x}_1}(\mathbf{x}_i,a_i)\right)'\right],\ r,s=0,1,2$$

$$\tilde{\tilde{s}}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i) = \tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i)-\mathbb{E}[\tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i)|a_i])\quad \textit{with}\quad \tilde{s}_0^{\mathbf{x}_1}(\mathbf{x}_i,a_i)=\mathbf{x}_{1i}$$

$$\tilde{\tilde{s}}_m^{a}(\mathbf{x}_i,a_i) = \tilde{s}_m^{a}(\mathbf{x}_i,a_i)-\mathbb{E}[\tilde{s}_m^{a}(\mathbf{x}_i,a_i)|a_i])\quad \textit{with}\quad \tilde{s}_0^{a}(\mathbf{x}_i,a_i)=h^v(a_i).$$

and $\tilde{\tilde{s}}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i)$ *and* $\tilde{\tilde{s}}_m^{a}(\mathbf{x}_i,a_i)$ *are defined recursively as in (S.1.3.4).*

Proof

We take the element $\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{\eta}_i^{GY}(\boldsymbol{\eta}_i^{G^2\mathbf{x}_1})'$ as an example. The proofs of the rest are similar and we omit them.

When $|\beta_1^0|<1$,

$$\mathbf{G}_N\mathbf{y}_N=\sum_{m=0}^{\infty}(\beta_1^0)^m\mathbf{G}_N^m(\mathbf{X}_{1N}\beta_2^0+\mathbf{G}_N\mathbf{X}_{1N}\beta_3^0+\mathbf{h}^v(\mathbf{a}_N)+\boldsymbol{\eta}_N^v),$$

and

$$\begin{aligned}&[\mathbf{G}_N\mathbf{y}_N]_i\\&=\beta_2^{0\prime}\left[\sum_{m=0}^{\infty}(\beta_1^0)^m\mathbf{G}_N^m\mathbf{X}_{1N}\right]_i+\beta_3^{0\prime}\left[\sum_{m=0}^{\infty}(\beta_1^0)^m\mathbf{G}_N^{m+1}\mathbf{X}_{1N}\right]_i\\&\quad+\left[\sum_{m=0}^{\infty}(\beta_1^0)^m\mathbf{G}_N^m\mathbf{h}(\mathbf{a}_N)\right]_i+\left[\sum_{m=0}^{\infty}(\beta_1^0)^m\mathbf{G}_N^m\boldsymbol{\eta}_N^v\right]_i.\end{aligned}$$

Set $s_0^{\mathbf{x}_1}(\mathbf{x}_i, a_i) = \tilde{s}_0^{\mathbf{x}_1}(\mathbf{x}_i, a_i) = \mathbf{x}_{1i}$. We have

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N} \boldsymbol{\eta}_i^{GY} (\boldsymbol{\eta}_i^{G^2\mathbf{x}_1})' \\
&= \frac{1}{N}\sum_{i=1}^{N} \left([\mathbf{G}_N \mathbf{y}_N]_i - \mathbb{E}\{[\mathbf{G}_N \mathbf{y}_N]_i | a_i\}\right) \left([\mathbf{G}_N^2 \mathbf{X}_{1N}]_i - \mathbb{E}\{[\mathbf{G}_N^2 \mathbf{X}_{1N}]_i | a_i\}\right)' \\
&= \frac{1}{N}\sum_{i=1}^{N} \left(\beta_2^{0\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{m,N,i}^{\mathbf{x}_1} | a_i] \right\}\right) \left(s_{2,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{2,N,i}^{\mathbf{x}_1} | a_i]\right)' \\
&+ \frac{1}{N}\sum_{i=1}^{N} \left(\beta_3^{0\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m+1,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{m+1,N,i}^{\mathbf{x}_1} | a_i] \right\}\right) \left(s_{2,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{2,N,i}^{\mathbf{x}_1} | a_i]\right)' \\
&+ \frac{1}{N}\sum_{i=1}^{N} \left(\beta_2^{0\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{a} - \mathbb{E}[s_{m,N,i}^{a} | a_i] \right\}\right) \left(s_{2,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{2,N,i}^{\mathbf{x}_1} | a_i]\right)' \\
&+ \frac{1}{N}\sum_{i=1}^{N} \left(\beta_2^{0\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{\upsilon} - \mathbb{E}[s_{m,N,i}^{\upsilon} | a_i] \right\}\right) \left(s_{2,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{2,N,i}^{\mathbf{x}_1} | a_i]\right)' \\
&= I + II + III + IV, \quad say.
\end{aligned}
$$

Consider term I,

$$
\frac{1}{N}\sum_{i=1}^{N} \left(\beta_2^{0\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{m,N,i}^{\mathbf{x}_1} | a_i] \right\}\right) \left(s_{2,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{2,N,i}^{\mathbf{x}_1} | a_i]\right)'.
$$

Denote

$$
\begin{aligned}
A_{1i} &:= \beta_2^{0\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{m,N,i}^{\mathbf{x}_1} | a_i] \right\} \\
A_{2i} &:= s_{2,N,i}^{\mathbf{x}_1} - \mathbb{E}[s_{2,N,i}^{\mathbf{x}_1} | a_i] \\
A_{3i} &:= \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{\upsilon} - \mathbb{E}[s_{m,N,i}^{\upsilon} | a_i] \right\} \\
B_{1i} &:= \beta_2^{0\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ \tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i, a_i) - \mathbb{E}[\tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i, a_i) | a_i] \right\} \\
B_{2i} &:= \tilde{s}_2^{\mathbf{x}_1}(\mathbf{x}_i, a_i) - \mathbb{E}[\tilde{s}_2^{\mathbf{x}_1}(\mathbf{x}_i, a_i) | a_i] \\
B_{3i} &:= \eta_i^{\upsilon} = \upsilon_i - \mathbb{E}[\upsilon | a_i].
\end{aligned}
$$

First, notice that

$$\begin{aligned}\left\|\frac{1}{N}\sum_{i=1}^{N}A_{1i}A_{2i}'-\frac{1}{N}\sum_{i=1}^{N}B_{1i}B_{2i}'\right\| &= \left\|\frac{1}{N}\sum_{i=1}^{N}(A_{1i}-B_{1i})A_{2i}'+\frac{1}{N}\sum_{i=1}^{N}B_{1i}(A_{2i}-B_{2i})'\right\| \\ &\leq \left\|\frac{1}{N}\sum_{i=1}^{N}(A_{1i}-B_{1i})A_{2i}'\right\|+\left\|\frac{1}{N}\sum_{i=1}^{N}B_{1i}(A_{2i}-B_{2i})'\right\| \\ &\leq \sup_i\|A_{1i}-B_{1i}\|\sup_i\|A_{2i}\|+\sup_i\|B_{1i}\|\sup_i\|A_{2i}-B_{2i}\|\end{aligned} \tag{S.1.3.6}$$

According to Lemma 16 and Lemma 14, we have

$$\sup_i\|A_{1i}-B_{1i}\|=o_p(1),\quad \sup_i\|A_{2i}-B_{2i}\|=o_p(1).$$

Also, under Assumption 11, $\sup_i\|A_{2i}\|$ and $\sup_i\|B_{1i}\|$ are bounded by a finite constant. Therefore, we deduce that

$$I=\frac{1}{N}\sum_{i=1}^{N}B_{1i}B_{2i}'+o_p(1).$$

Then, we apply the WLLN to $\frac{1}{N}\sum_{i=1}^{N}B_{1i}B_{2i}'$ and deduce

$$\begin{aligned}\frac{1}{N}\sum_{i=1}^{N}B_{1i}B_{2i}' &\xrightarrow{p}\mathbb{E}\left[B_{1i}B_{2i}'\right] \\ &=\mathbb{E}\left[\left(\beta_2^{0\prime}\sum_{m=0}^{\infty}(\beta_1^0)^m\left\{\tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i)-\mathbb{E}[\tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i)|a_i]\right\}\right)(\tilde{s}_2^{\mathbf{x}_1}(\mathbf{x}_i,a_i)-\mathbb{E}[s_2^{\mathbf{x}_1}(\mathbf{x}_i,a_i)|a_i])\right] \\ &=\mathbb{E}\left[\left(\beta_2^{0\prime}\sum_{m=0}^{\infty}(\beta_1^0)^m\tilde{\tilde{s}}_m^{\mathbf{x}_1}(\mathbf{x}_i,a_i)\right)\tilde{\tilde{s}}_2^{\mathbf{x}_1}(\mathbf{x}_i,a_i)\right]\end{aligned}$$

We can derive the probability limits of terms II and III by similar fashion.

For term IV, first notice that for each $m=0,1,2,...,$

$$\begin{aligned}\mathbb{E}[s_{m,N,i}^{v}|a_i] &= \mathbb{E}\left([\mathbf{G}_N^m\boldsymbol{\eta}_N^v]_i|a_i\right) \\ &= \mathbb{E}\left\{\mathbb{E}\left([\mathbf{G}_N^m\boldsymbol{\eta}_N^v]_i|\mathbf{X}_N,\mathbf{D}_N,a_i\right)|a_i\right\} \\ &= \mathbb{E}\left\{[\mathbf{G}_N^m\mathbb{E}(\boldsymbol{\eta}_N^v|\mathbf{X}_N,\mathbf{D}_N,a_i)]_i|a_i\right\}=0,\end{aligned}$$

where the last equality holds by Lemma 1. Then, $A_{3i} := \sum_{m=0}^{\infty}(\beta_1^0)^m s_{m,N,i}^{\upsilon}$.

Similar to the bound in (S.1.3.6), notice that

$$\left\| \frac{1}{N}\sum_{i=1}^{N} A_{3i}A_{2i}' - \frac{1}{N}\sum_{i=1}^{N} B_{3i}B_{2i}' \right\| \leq \sup_i \|A_{3i} - B_{3i}\| \sup_i \|A_{2i}\| + \sup_i \|B_{3i}\| \sup_i \|A_{2i} - B_{2i}\|.$$

According to Lemma 16 and Lemma 14,

$$\sup_i \|A_{3i} - B_{3i}\| = o_p(1), \quad \sup_i \|A_{2i} - B_{2i}\| = o_p(1).$$

Also, under Assumption 11, $\sup_i \|A_{2i}\|$ and $\sup_i \|B_{3i}\|$ are bounded by a finite constant. Therefore, we deduce that

$$IV = \frac{1}{N}\sum_{i=1}^{N} B_{3i}B_{2i}' + o_p(1).$$

Then, we apply the WLLN to $\frac{1}{N}\sum_{i=1}^{N} B_{3i}B_{2i}'$ and deduce

$$\begin{aligned}
\frac{1}{N}\sum_{i=1}^{N} B_{3i}B_{2i}' &\xrightarrow{p} \mathbb{E}\left[B_{3i}B_{2i}'\right] \\
&= \mathbb{E}\left[\eta_i^a\left(\tilde{s}_2^{\mathbf{x}_1}(\mathbf{x}_i, a_i) - \mathbb{E}[s_2^{\mathbf{x}_1}(\mathbf{x}_i, a_i)|a_i]\right)\right] \\
&= \mathbb{E}\left[(\upsilon_i - \mathbb{E}[\upsilon_i|a_i])\tilde{\tilde{s}}_2^{\mathbf{x}_1}(\mathbf{x}_i, a_i)\right] \\
&= \mathbb{E}\left\{\mathbb{E}\left(\upsilon_i - \mathbb{E}[\upsilon_i|a_i]|\mathbf{x}_i, a_i\right)\tilde{\tilde{s}}_2^{\mathbf{x}_1}(\mathbf{x}_i, a_i)\right\} \\
&= 0.
\end{aligned}$$

□

Let $\sigma^2(\mathbf{x}_i, a_i) := \mathbb{E}[(\eta_i^{\upsilon})^2|\mathbf{x}_i, a_i] = \mathbb{E}[(\upsilon_i - \mathbb{E}[\upsilon_i|a_i])^2|\mathbf{x}_i, a_i]$.

Lemma 12. *Under Assumptions 1 and 11, as $N \to \infty$, we have*

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))'\sigma^2(\mathbf{x}_i, a_i) \xrightarrow{p} \mathbf{S}^{\mathbf{zz}\sigma},$$

where the limit variance $\mathbf{S}^{\mathbf{zz}\sigma}$ is defined in Lemma 13.

Proof

The proof is similar to that of the results in Lemma 11 and we omit it. □

Lemma 13. *Under Assumptions 1 and 11, as $N \to \infty$, we have*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i)) \eta_i^{\upsilon} \Rightarrow \mathcal{N}(0, \mathbf{S}^{\mathbf{zz}\sigma}),$$

where

$$\mathbf{S}^{\mathbf{zz}\sigma} = \begin{pmatrix} S^{\mathbf{x}_1\mathbf{x}_1\sigma} & S^{\mathbf{x}_1 G\mathbf{x}_1\sigma} & S^{\mathbf{x}_1 G^2\mathbf{x}_1\sigma} \\ S^{G\mathbf{x}_1\mathbf{x}_1\sigma} & S^{G\mathbf{x}_1 G\mathbf{x}_1\sigma} & S^{G\mathbf{x}_1 G^2\mathbf{x}_1\sigma} \\ S^{G^2\mathbf{x}_1\mathbf{x}_1\sigma} & S^{G^2\mathbf{x}_1 G\mathbf{x}_1\sigma} & S^{G^2\mathbf{x}_1 G^2\mathbf{x}_1\sigma} \end{pmatrix}$$

and

$$S^{G^r\mathbf{x}_1 G^s\mathbf{x}_1\sigma} = \mathbb{E}\left[\tilde{\tilde{s}}_r^{\mathbf{x}_1}(\mathbf{x}_i, a_i)) \left(\tilde{\tilde{s}}_s^{\mathbf{x}_1}(\mathbf{x}_i, a_i)\right)' \sigma^2(\mathbf{x}_i, a_i)\right], \ r, s = 0, 1, 2$$

$$\tilde{\tilde{s}}_m^{\mathbf{x}_1}(\mathbf{x}_i, a_i) = \tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i, a_i) - \mathbb{E}[\tilde{s}_m^{\mathbf{x}_1}(\mathbf{x}_i, a_i)|a_i]) \quad \textit{with} \quad \tilde{s}_0^{\mathbf{x}_1}(\mathbf{x}_i, a_i) = \mathbf{x}_{1i}$$

$$\sigma^2(\mathbf{x}_i, a_i) := \mathbb{E}[(\eta_i^{\upsilon})^2|\mathbf{x}_i, a_i] = \mathbb{E}[(\upsilon_i - \mathbb{E}[\upsilon_i|a_i])^2|\mathbf{x}_i, a_i],$$

where $\tilde{\tilde{s}}_m^{\mathbf{x}_1}(\mathbf{x}_i, a_i)$ is defined recursively as in (S.1.3.4).

Proof

Let $\mathcal{F}_i = (\mathbf{X}_{1N}, \mathbf{D}_N, a_i, \eta_1^{\upsilon}, \ldots, \eta_{i-1}^{\upsilon})$. Conditional on $(\mathbf{X}_{1N}, \mathbf{D}_N, a_i)$,

$$\mathbb{E}[(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))\eta_i^{\upsilon}|\mathcal{F}_i] = (\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))\mathbb{E}[\eta_i^{\upsilon}|\mathcal{F}_i] = 0,$$

and so $\{(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}(a_i))\eta_i^{\upsilon}, \mathcal{F}_i\}$ is a martingale difference sequence.

Since $\eta_i^{\upsilon} = \upsilon_i - \mathbb{E}[\upsilon_i|a_i]$ is bounded by a constant under Assumption 11,

$$\mathbb{E}[(\eta_i^{\upsilon})^4|F_{i-1}] < M \tag{S.1.3.7}$$

for some finite constant M.

Also notice under Assumptions 1, we have

$$\begin{aligned}\mathbb{E}[(\eta_i^{\upsilon})^2|\mathcal{F}_i] &= \mathbb{E}[(\upsilon_i-\mathbb{E}(\upsilon|a_i))^2|\mathbf{x}_i,a_i,\mathbf{x}_{-i},\mathbf{a}_{-i},\mathbf{D}_N(\mathbf{x}_{-i},\mathbf{a}_{-i},\{u_{ij}\}_{i,j=1,\dots,N},\mathbf{x}_i,a_i),\{\eta_j^{\upsilon}\}_{j<i}]\\ &= \mathbb{E}[(\upsilon_i-\mathbb{E}(\upsilon|a_i))^2|\mathbf{x}_i,a_i]\\ &=: \sigma^2(\mathbf{x}_i,a_i).\end{aligned}$$

Let ℓ be a nonzero vector whose dimension is the same as the IVs $\mathbf{z}_i$. Then,

$$\begin{aligned}\mathbb{E}[\ell'\boldsymbol{\eta}_i^Z(\boldsymbol{\eta}_i^Z)'\ell(\eta_i^{\upsilon})^2|\mathcal{F}_i] &= [\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))'\ell]\mathbb{E}[(\eta_i^{\upsilon})^2|\mathcal{F}_i]\\ &= [\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))'\ell]\sigma^2(\mathbf{x}_i,a_i).\end{aligned}$$

Let

$$s_N^2 := \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}[\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))'\ell(\eta_i^{\upsilon})^2|\mathcal{F}_i] = \frac{1}{N}\sum_{i=1}^{n}[\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))'\ell]\sigma^2(\mathbf{x}_i,a_i).$$

According to Lemma 12,

$$s_N^2 \xrightarrow{p} \mathbf{S}^{\mathbf{zz}\sigma}.$$

Also, since $\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))\eta_i^{\upsilon} = \ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))(\upsilon_i-\mathbb{E}[\upsilon_i|a_i])$ is bounded by a constant, under Assumption 11 the Lindeberg-Feller condition is satisfied, that is, for any $\epsilon>0$,

$$\begin{aligned}&\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left[[\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))'\ell](\eta_i^{\upsilon})^2\mathbb{I}\left\{|\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))\eta_i^{\upsilon}|>\epsilon\sqrt{N}\right\}|\mathcal{F}_i\right]\\ &\leq \sum_{i=1}^{N}\frac{1}{\epsilon^2N^2}\mathbb{E}\left[[\ell'(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))'\ell]^2(\eta_i^{\upsilon})^4|\mathcal{F}_i\right]\\ &\leq \frac{M}{\epsilon N}\to 0\end{aligned}$$

as $N\to\infty$.

Then, by the Martingale Central Limit Theorem (e.g., see Corollary 3.1 Hall and Heyde (2014)), we have the desired result for theorem:

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i-\mathbf{h}^{\mathbf{z}}(a_i))\eta_i^{\upsilon} \Rightarrow \mathcal{N}(0,\mathbf{S}^{\mathbf{zz}\sigma}).$$

□

Proof of Theorem 5.1.

Theorem 5.1 follows from Lemma 2, Lemma 7, Lemma 11, and Lemma 13. □

S.1.4. **Further Supporting Lemmas.**

Lemma 14 (Uniform Convergence of $s_{m,N,i}$ in i)**.** *Assume Assumptions 1, 5, 7, 8 and 11. Suppose that $s_0(\mathbf{x}_i, a_i)$ is a bounded function of $\mathbf{x}_i$ and a_i. Suppose that we define $s_{m,N,i}$ as in (S.1.3.2) and consider its probability limit $\tilde{s}_m(\mathbf{x}_i, a_i)$ in equation (S.1.3.4) for each i. Then, for each $m = 0, 1, 2, \cdots$*

$$(a) \sup_{1 \le i \le N} |s_{m,N,i} - \tilde{s}_m(\mathbf{x}_i, a_i)| = o_p(1)$$

$$(b) \sup_{1 \le i \le N} |\mathbb{E}[s_{m,N,i}|a_i] - \mathbb{E}[\tilde{s}_m(\mathbf{x}_i, a_i)|a_i]| = o_p(1).$$

Proof

Part (a).

For $m = 0$.

The required result for the lemma holds trivially because of the definition that $s_{0,N,i} = \tilde{s}_0(\mathbf{x}_i, a_i)$.

Next we show the required result for $m = 1$ and then use mathematical induction for the rest $m = 2, 3, ...$

For $m = 1$.

The claim for the case $m = 1$ is proved in three steps.

Step 1.

Notice that

$$\begin{aligned} s_{1,N,i} &= \left(\frac{1}{N}\sum_{j\neq i} d_{ij,N}\right)^{-1} \frac{1}{N}\sum_{j\neq i} d_{ij,N} s_{1,N,j} \\ &= \left(\frac{1}{N}\sum_{j\neq i} \mathbb{I}\left\{g(t(\mathbf{x}_{2i},\mathbf{x}_{2j}),a_i,a_j) \geq u_{ij}\right\}\right)^{-1} \\ &\quad \times \frac{1}{N}\sum_{j\neq i} \mathbb{I}\left\{g(t(\mathbf{x}_{2i},\mathbf{x}_{2j}),a_i,a_j) \geq u_{ij}\right\} s_0(\mathbf{x}_j,a_j). \end{aligned}$$

Then, by the WLLN, for each i,

$$\begin{aligned} \frac{1}{N}\sum_{j\neq i} \mathbb{I}\left\{g(t(\mathbf{x}_{2i},\mathbf{x}_{2j}),a_i,a_j) \geq u_{ij}\right\} &\xrightarrow{p} \int\int \Phi((t(\mathbf{x}_{2i},\mathbf{x}_2),a_i,a)\pi(\mathbf{x}_2,a)d\mathbf{x}_2 da \\ &= \mathbb{E}[d_{ij,N}|\mathbf{x}_i,a_i] \qquad \text{(S.1.4.1)} \\ \frac{1}{N}\sum_{j\neq i} \mathbb{I}\left\{g(t(\mathbf{x}_{2i},\mathbf{x}_{2j}),a_i,a_j) \geq u_{ij}\right\} s_0(\mathbf{x}_j,a_j) &\xrightarrow{p} \int\int \Phi(t(\mathbf{x}_{2i},\mathbf{x}_2)'\lambda^0 + a_i + a)s_0(\mathbf{x},a)\pi(\mathbf{x},a)d\mathbf{x}da \\ &= \mathbb{E}[d_{ij,N}s_0(\mathbf{x}_j,a_j)|\mathbf{x}_i,a_i]. \qquad \text{(S.1.4.2)} \end{aligned}$$

Since $\mathbb{E}[d_{ij,N}|\mathbf{x}_i,a_i] > 0$ uniformly in i,j under Assumption 11 (vi),(v), and (vi) for each i as $N\to\infty$, we have

$$s_{1,N,i} \to_p \tilde{s}_1(\mathbf{x}_i,a_i) = \frac{\int\int \Phi(g(t(\mathbf{x}_{2i},\mathbf{x}_2),a_i,a)s_0(\mathbf{x},a)\pi(\mathbf{x},a)d\mathbf{x}da}{\int\int \Phi(g(t(\mathbf{x}_{2i},\mathbf{x}_2),a_i,a)\pi(\mathbf{x}_2,a)d\mathbf{x}_2 da}$$

Step 2.

In this step, we show that the convergences in (S.1.4.1) and (S.1.4.2) hold uniformly in i. For this, we introduce the following notation. Let

$$\zeta_{i,N,1} = \frac{1}{N}\sum_{j=1,\neq i}^{N} \left(d_{ij,N} - \mathbb{E}[d_{ij,N}|\mathbf{x}_i,a_i]\right)$$

and

$$\zeta_{i,N,2} = \frac{1}{N}\sum_{j=1,\neq i}^{N} \left(d_{ij,N}s_0(\mathbf{x}_j,a_j) - \mathbb{E}[d_{ij,N}s_0(\mathbf{x}_j,a_j)|\mathbf{x}_i,a_i]\right).$$

Notice that conditional on $(\mathbf{x}_i, a_i)$, $d_{ij,N}$ and $d_{ij,N}s_0(\mathbf{x}_j, a_j)$ are iid with conditional mean zero and bounded by a constant across $j = 1, ..., N, \neq i$. Then, there exists a finite constant M_1 such that

$$\sup_i \mathbb{E}\left(\|\sqrt{N}\zeta_{i,N,k}\|^4|\mathbf{x}_i, a_i\right) \leq M_1,$$

and we can deduce the desired result

$$\sup_i \|\zeta_{i,N,k}\| = O_p(N^{-1/4}) = o_p(1)$$

because for any $\epsilon >$, we choose $M_2 = \frac{\epsilon}{M_1}$ and then

$$\begin{aligned}
\mathbb{P}\{\sup_i \|\zeta_{i,N,k}\| \geq N^{-1/4}M_2^{1/4}|\mathbf{x}_i, a_i\} &= \mathbb{P}\{\sup_i N^{-1/4}\|\sqrt{N}\zeta_{i,N,k}\| \geq M_2^{1/4}|\mathbf{x}_i, a_i\} \\
&= \mathbb{P}\{\sup_i N^{-1}\|\sqrt{N}\zeta_{i,N,k}\|^4 \geq M_2|\mathbf{x}_i, a_i\} \\
&\leq \mathbb{P}\left\{\frac{1}{N}\sum_{i=1}^{N}\|\sqrt{N}\zeta_{i,N,k}\|^4 \geq M_2|\mathbf{x}_i, a_i\right\} \\
&\leq \frac{1}{M_2}\frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\left(\|\sqrt{N}\zeta_{i,N,k}\|^4|\mathbf{x}_i, a_i\right) \\
&\leq \frac{M_1}{M_2} = \epsilon.
\end{aligned}$$

Step 3.

Now we prove the desired result for the case $m = 1$. Define $\Psi_{i,N,1} = \frac{1}{N}\sum_{j\neq i} d_{ij,N}$ and $\Psi_{i,N,2} = \frac{1}{N}\sum_{j\neq i} d_{ij,N}s_0(\mathbf{x}_j, a_j)$. Then,

$$s_{1,N,i} = \frac{\Psi_{i,N,1}}{\Psi_{i,N,2}}.$$

Let $\phi_{i,1} = \frac{1}{N}\sum_{j=1,\neq i}^{N} \mathbb{E}[d_{ij,N}|\mathbf{x}_i, a_i]$ and $\psi_{i,2} = \frac{1}{N}\sum_{j=1,\neq i}^{N} \mathbb{E}[d_{ij,N} s_0(\mathbf{x}_j, a_j)|\mathbf{x}_i, a_i]$. Notice that

$$\begin{aligned}
\sup_i \|s_{1,N,i}\| &= \sup_i \left\| \frac{\Psi_{i,N,2}}{\Psi_{i,N,1}} - \frac{\Psi_{i,2}}{\Psi_{i,1}} \right\| \\
&\leq \sup_i \left\| \frac{\Psi_{i,N,2} - \Psi_{i,2}}{\Psi_{i,N,1}} \right\| + \sup_i \left\| \frac{\Psi_{i,2}(\Psi_{i,N,1} - \Psi_{i,1})}{\Psi_{i,N,1}\Psi_{i,1}} \right\| \qquad = o_p(1),
\end{aligned}$$

where the last line holds because $\|\Psi_{i,N,k} - \Psi_{i,k}\| = o_p(1)$ by Step 2, and $\Psi_{i,1} > 0$ and $\|\Psi_{i,2}\|$ is bounded by a constant. This shows the required result

$$\sup_i \|s_{1,N,i} - \tilde{s}_1(\mathbf{x}_i, a_i)\| = o_p(1).$$

For $m \geq 2$.

Given that we show the required result of the lemma with $m = 1$, we show the rest by mathematical induction. For this, suppose that

$$\sup_{1\leq i\leq N} \|s_{m,N,i} - \tilde{s}_m(\mathbf{x}_i, a_i)\| = o_p(1).$$

Then, we have

$$\begin{aligned}
&\sup_{1\leq i\leq N} \|s_{m+1,N,i} - \tilde{s}_{m+1}(\mathbf{x}_i, a_i)\| \\
&= \sup_{1\leq i\leq N} \left\| \frac{\frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N} s_{m,N,i}}{\frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N}} - \frac{\mathbb{E}[d_{ij,N}\tilde{s}_m(\mathbf{x}_j, a_j)|\mathbf{x}_i, a_i]}{\mathbb{E}[d_{ij,N}|\mathbf{x}_i, a_i]} \right\| \\
&\leq \sup_{1\leq i\leq N} \frac{\left\| \frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N}\left(s_{m,N,i} - \mathbb{E}[d_{ij,N}\tilde{s}_m(\mathbf{x}_j, a_j)|\mathbf{x}_i, a_i]\right) \right\|}{\frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N}} \\
&\quad + \sup_{1\leq i\leq N} \|\mathbb{E}[d_{ij,N}\tilde{s}_m(\mathbf{x}_j, a_j)|\mathbf{x}_i, a_i]\| \sup_{1\leq i\leq N} \left| \frac{1}{\frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N}} - \frac{1}{\mathbb{E}[d_{ij,N}|\mathbf{x}_i, a_i]} \right|.
\end{aligned}$$

For the first term, we have by the definition of $g_{ij,N} = \frac{d_{ij,N}}{\sum_{j=1,\neq i}^{N} d_{ij,N}}$ and since $\sum_{j=1,\neq i} g_{ij,N} = 1$, we have

$$\begin{aligned}
&\sup_{1\leq i\leq N} \frac{\left\| \frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N}\left(s_{m,N,i} - \mathbb{E}[d_{ij,N}\tilde{s}_m(\mathbf{x}_j,a_j)|\mathbf{x}_i,a_i]\right)\right\|}{\frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N}} \\
&= \sup_{1\leq i\leq N} \left\| \frac{1}{N}\sum_{j=1,\neq i}^{N} g_{ij,N}\left(s_{m,N,i} - \mathbb{E}[d_{ij,N}\tilde{s}_m(\mathbf{x}_j,a_j)|\mathbf{x}_i,a_i]\right)\right\| \\
&\leq \sup_{1\leq i\leq N} \left\| s_{m,N,i} - \mathbb{E}[d_{ij,N}\tilde{s}_m(\mathbf{x}_j,a_j)|\mathbf{x}_i,a_i]\right\| \\
&= o_p(1),
\end{aligned}$$

where the last line holds by the assumption of mathematical induction. We can show the second term

$$\sup_{1\leq i\leq N} \left\|\mathbb{E}[d_{ij,N}\tilde{s}_m(\mathbf{x}_j,a_j)|\mathbf{x}_i,a_i]\right\| \sup_{1\leq i\leq N}\left| \frac{1}{\frac{1}{N}\sum_{j=1,\neq i}^{N} d_{ij,N}} - \frac{1}{\mathbb{E}[d_{ij,N}|\mathbf{x}_i,a_i]}\right| = o_p(1)$$

by using similar argument used in the proof of Step 3 of the case $m = 1$. □

Part (b).

Notice that under Assumption 11, $\mathbb{E}[s_{m,N,i}|a_i]$ and $\mathbb{E}[\tilde{s}_m(\mathbf{x}_i,a_i)|a_i]$ are bounded by a finite constant. The required argument follows by similar arguments used in the proof of Part (a).
□

Lemma 15 (Uniform Convergence of $s^{\upsilon}_{m,N,i}$ in i)**.** *Assume Assumptions Assumptions 1, 5, 7, 8 and 11. Suppose that we define $s^{\upsilon}_{m,N,i}$ as in (S.1.3.5). Then, for each $m = 1, 2, \cdots$*

$$\sup_{1\leq i\leq N} |s^{\upsilon}_{m,N,i}| = o_p(1).$$

Proof

The proof is similar to that of Lemma 14. First, we show that for each i and $m = 1, 2, ...$ the probability limit of $s^{\upsilon}_{m,N,i}$ defined with $s^{\upsilon}_{0,i} = \eta^{\upsilon}_i = \upsilon_i - \mathbb{E}[\upsilon_i|a_i]$ recursively as (S.1.3.5) is zero as $N \to \infty$. To verify this, let

$$\tilde{s}^{\upsilon}_{0,i} = \eta^{\upsilon}_i = \upsilon_i - \mathbb{E}[\upsilon_i|a_i].$$

For $m = 1$,

$$s^{\upsilon}_{1,N,i} = \left(\frac{1}{N}\sum_{j\neq i} d_{ij,N}\right)^{-1} \frac{1}{N}\sum_{j\neq i} d_{ij,N} s^{\upsilon}_{0,j}.$$

Consider the numerator. Notice by definition that

$$d_{ij,N} s^{\upsilon}_{0,j} = \mathbb{I}\left\{g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}, a_i, a_j) \geq u_{ij}\right\}(\upsilon_j - \mathbb{E}[\upsilon_j|a_j])$$

are i.i.d. across j conditioning on $(\mathbf{x}_{2i}, a_i)$ and bounded by a finite constant under Assumption 11. Then, by the WLLN conditioning on $(\mathbf{x}_{2i}, a_i)$, we have

$$\begin{aligned}
\frac{1}{N}\sum_{j\neq i} d_{ij,N} s^{\upsilon}_{0,j} &\xrightarrow{p} \mathbb{E}\left[d_{ij,N}(\upsilon_j - \mathbb{E}[\upsilon_j|a_j])|\mathbf{x}_{2i}, a_i\right] \\
&= \mathbb{E}\left[d_{ij,N}\mathbb{E}\left(\upsilon_j - \mathbb{E}[\upsilon_j|a_j]|\mathbf{X}_N, \mathbf{D}_N, a_i\right)|\mathbf{x}_{2i}, a_i\right] \\
&= 0,
\end{aligned}$$

where the last equality holds by Lemma 1. The denominator converges to

$$\frac{1}{N}\sum_{j\neq i} \mathbb{I}\left\{g(t(\mathbf{x}_{2i}, \mathbf{x}_{2j}), a_i, a_j) \geq u_{ij}\right\} \to_p \int\int \Phi(g(t(\mathbf{x}_{2i}, \mathbf{x}_2), a_i, a)\pi(\mathbf{x}_2, a)d\mathbf{x}_2 da > 0,$$

where the last inequality holds under Assumption 11.

This shows that as $N \to \infty$

$$\frac{1}{N}\sum_{j\neq i} g_{ij,N} s^{\upsilon}_{0,j} \xrightarrow{p} 0 =: \tilde{s}^{\upsilon}_{1,i}$$

for each i.

Then, using similar argument in Step 2 of the proof of Lemma 14, we deduce

$$\sup_{1\leq i\leq N}\left|\frac{1}{N}\sum_{j\neq i} g_{ij,N} s^{\upsilon}_{0,j}\right| = o_p(1).$$

Also, for $m = 2, ...$, we follow the same mathematical induction argument in Steps 3 and 4 of the proof of Lemma 14 and deduce that

$$\sup_{1 \leq i \leq N} \left| \frac{1}{N} \sum_{j \neq i} g_{ij,N} s^{\upsilon}_{m,N,j} \right| = o_p(1).$$

□

Lemma 16. *Assume Assumptions 1, 5, 7, 8 and 11. Suppose that $s_0(\mathbf{x}_i, a_i)$ is a bounded function of $\mathbf{x}_i$ and a_i. Suppose that we define $s_{m,N,i}$ as in equation (S.1.3.2) and consider its probability limit $\tilde{s}_m(\mathbf{x}_i, a_i)$ in equation (S.1.3.4) for each i. Then,*

$$(a) \sup_{1 \leq i \leq N} \left| \sum_{m=0}^{\infty} (\beta_1^0)^m \left(s_{m,N,i} - \tilde{s}_m(\mathbf{x}_i, a_i) \right) \right| = o_p(1)$$

$$(b) \sup_{1 \leq i \leq N} \left| \sum_{m=0}^{\infty} (\beta_1^0)^m \left(\mathbb{E}[s_{m,N,i}|a_i] - \mathbb{E}[\tilde{s}_m(\mathbf{x}_i, a_i)|a_i] \right) \right| = o_p(1).$$

Also, suppose that we define $s^{\eta}_{m,N,i}$ as in equation (S.1.3.5). Let $\tilde{s}^{\eta}_{0,i} = \eta^a_i$ and $\tilde{s}^{\eta}_{m,i} = 0$ for $m = 1, 2, ...$ Then,

$$(c) \quad \sup_{1 \leq i \leq N} \left| \sum_{m=0}^{\infty} (\beta_1^0)^m \left(s^{\eta}_{m,N,i} - \tilde{s}^{\eta}_m \right) \right| = o_p(1).$$

Proof

Part (a).

Notice from Assumption 11 that $|\beta_1^0| < 1$ and $s_{m,N,i}, \tilde{s}_m(\mathbf{x}_i, a_i), \mathbb{E}[s_{m,N,i}|a_i], \mathbb{E}[\tilde{s}_m(\mathbf{x}_i, a_i)|a_i]$ are bounded by a finite constant, say, M. For given $\epsilon > 0$, we choose m^* such that $2M \sum_{m=m^*+1}^{\infty} (\beta_1^0)^m \leq \epsilon$. Then, by definition, we have

$$\sup_{1 \leq i \leq N} \left| \sum_{m=m^*+1}^{\infty} (\beta_1^0)^m \left(s_{m,N,i} - \tilde{s}_m(\mathbf{x}_i, a_i) \right) \right| \leq 2M \sum_{m=m^*+1}^{\infty} (\beta_1^0)^m \leq \epsilon.$$

Notice that

$$\begin{aligned}
\sup_{1\le i\le N}\left|\sum_{m=0}^{\infty}(\beta_1^0)^m\left(s_{m,N,i}-\tilde{s}_m(\mathbf{x}_i,a_i)\right)\right| &\le \sup_{1\le i\le N}\left|\sum_{m=0}^{m^*}(\beta_1^0)^m\left(s_{m,N,i}-\tilde{s}_m(\mathbf{x}_i,a_i)\right)\right|+\epsilon\\
&\le m^*\sup_{1\le i\le N}\left|s_{m,N,i}-\tilde{s}_m(\mathbf{x}_i,a_i)\right|+\epsilon\\
&= o_p(1)+\epsilon,
\end{aligned}$$

where the last inequality holds since m^* is finite and by Lemma 16. Since ϵ is arbitrary, we have the desired result for Part (a). $\square$

Parts (b) and (c).

Under Assumption 11, $\mathbb{E}[s_{m,N,i}|a_i],\mathbb{E}[\tilde{s}_m(\mathbf{x}_i,a_i)|a_i]$, and $\eta_i^\upsilon=\upsilon_i-h^\upsilon(a_i)$ are bounded by a constant. Apply the same argument used in the proof of Part (a), then we deduce the required result of Parts (b) and (c). $\square$

Appendix S.2. For $\bar{\beta}_{2SLS}$

S.2.1. **Limiting distribution of $\bar{\beta}_{2SLS}$.** Recall the definition that for any variable b_i^l being an element of $(y_i,\mathbf{w}_i,\mathbf{w}_i)$ and υ_i,

$$\eta_{*i}^l:=b_i^l-h_*^l(\mathbf{x}_{2i},a_i)=b_i^l-h_{**}^l(\mathbf{x}_{2i},\deg_i),\quad \eta_{*i}^\upsilon:=\upsilon_i-h_*^\upsilon(\mathbf{x}_{2i},a_i)\upsilon_i-h_{**}^\upsilon(\mathbf{x}_{2i},\deg_i).$$

Let $\boldsymbol{\eta}_{*N}^\upsilon=(\eta_{*1}^\upsilon,...,\eta_{*N}^\upsilon)'$.

Outline:

Step 1 Show that

$$\begin{aligned}
&\sqrt{N}(\bar{\beta}_{2SLS}-\beta^0)\\
=\;&\left(\mathbf{W}_N'\mathbf{M}_{\mathbf{R}_N}\mathbf{Z}_N\left(\mathbf{Z}_N'\mathbf{M}_{\mathbf{R}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\mathbf{R}_N}\mathbf{W}_N\right)^{-1}\\
&\times\mathbf{W}_N'\mathbf{M}_{\mathbf{R}_N}\mathbf{Z}_N\left(\mathbf{Z}'\mathbf{M}_{\mathbf{R}_N}\mathbf{Z}_N\right)^{-1}\mathbf{Z}_N'\mathbf{M}_{\mathbf{R}_N}\boldsymbol{\eta}_{*N}^\upsilon+o_p(1). \qquad \text{(S.2.1.1)}
\end{aligned}$$

Step 2 Show

$$\frac{1}{N}\sum_{i=1}^{N}\left(b_i^l - \hat{h}_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)\left(b_i^l - \hat{h}_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)' \frac{1}{N}\sum_{i=1}^{N}\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)' + o_p(1)$$

and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(b_i^l - \hat{h}_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)\eta_{*i}^v = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)\eta_{*i}^v + o_p(1).$$

Step 3 Derive the limits of

$$\frac{1}{N}\sum_{i=1}^{N}\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)' = \frac{1}{N}\sum_{i=1}^{N}\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, a_i)\right)\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, a_i)\right)'$$

and

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(b_i^l - h_{**}^l(\mathbf{x}_{2i}, \deg_i)\right)\eta_{*i}^v = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left(b_i^l - h_{*}^l(\mathbf{x}_{2i}, a_i)\right)\eta_{*i}^v$$

S.2.2. **Controlling the Sampling Error $\widehat{deg}_i - deg_i$ in Sieve Estimation.** Equation (S.2.1.1) holds if the following Lemma is true.

Lemma 17. *Assume Assumptions Assumptions 1, 3, 4, 9, 10 and 11. Then the following holds.*

(a) $\frac{1}{N}(\mathbf{Z}_N'\mathbf{P}_{\widehat{R}_N}\mathbf{W}_N - \mathbf{Z}_N'\mathbf{P}_{R_N}\mathbf{W}_N) = o_p(1)$.

(b) $\frac{1}{N}(\mathbf{Z}_N'\mathbf{P}_{\widehat{R}_N}\mathbf{Z}_N - \mathbf{Z}_N'\mathbf{P}_{R_N}\mathbf{Z}_N) = o_p(1)$.

(c) $\frac{1}{\sqrt{N}}(\mathbf{Z}_N'\mathbf{P}_{\widehat{R}_N}\boldsymbol{\eta}_{*N}^{\boldsymbol{v}} - \mathbf{Z}_N'\mathbf{P}_{R_N}\boldsymbol{\eta}_{*N}^{\boldsymbol{v}}) = o_p(1)$.

(d) $\frac{1}{\sqrt{N}}(\mathbf{Z}'\mathbf{M}_{\widehat{R}_N}(H(\mathbf{a}_N) - \widehat{\boldsymbol{R}}_N\gamma)) = o_p(1)$.

Proof. We can apply a similar argument as in Lemma 2 and derive the desired result. □

S.2.3. **Controlling the Series Approximation Error for $\mathbf{r}^K(\mathbf{x}_{2i}, \mathbf{deg}_i)$.**

Lemma 18 (Series Approximation)**.** *Assume the assumptions in Lemma 17. Then, we have*

(a) $\frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}_i - \hat{\mathbf{h}}_{**}^{\mathbf{w}}(\mathbf{x}_{2i}, deg_i))(\mathbf{z}_i - \hat{\mathbf{h}}_{**}^{\mathbf{z}}(\mathbf{x}_{2i}, deg_i))' = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}_i - \mathbf{h}_{**}^{\mathbf{w}}(\mathbf{x}_{2i}, deg_i))(\mathbf{z}_i - \mathbf{h}_{**}^{\mathbf{z}}(\mathbf{x}_{2i}, deg_i))' + o_p(1)$,

(b) $\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, deg_i))(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, deg_i))' = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, deg_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, deg_i))' + o_p(1)$,

(c) $\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i - \hat{\mathbf{h}}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, deg_i))\eta^{v}_{*i} = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{**}(\mathbf{x}_{2i}, deg_i))\eta^{v}_{*i} + o_p(1)$.

Then the proofs are analogous to the proofs presented in Section S.1.2 and we omit them.

S.2.4. **Limiting distribution of $\bar{\beta}_{2SLS}$.** Note that $h^{l}_{**}(\mathbf{x}_{2i}, deg_i) = h^{l}_{*}(\mathbf{x}_{2i}, a_i)$. Using this relationship we can state the following Lemmas.

Lemma 19. *Under Assumption 1, 3, and 11, we have*

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{w}_i - \mathbf{h}^{\mathbf{w}}_{*}(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{*}(\mathbf{x}_{2i}, a_i))' \xrightarrow{p} \begin{pmatrix} \bar{S}^{GY,\mathbf{x}_1} & \bar{S}^{GY,G\mathbf{x}_1} & \bar{S}^{GY,G^2\mathbf{x}_1} \\ \bar{S}^{\mathbf{x}_1,\mathbf{x}_1} & \bar{S}^{\mathbf{x}_1,G\mathbf{x}_1} & \bar{S}^{\mathbf{x}_1,G^2\mathbf{x}_1} \\ \bar{S}^{G\mathbf{x}_1,\mathbf{x}_1} & \bar{S}^{G\mathbf{x}_1,G\mathbf{x}_1} & \bar{S}^{G\mathbf{x}_1,G^2\mathbf{x}_1} \end{pmatrix} =: \bar{S}^{\mathbf{wz}},$$

and

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{*}(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{*}(\mathbf{x}_{2i}, a_i))' \xrightarrow{p} \begin{pmatrix} \bar{S}^{\mathbf{x}_1,\mathbf{x}_1} & \bar{S}^{\mathbf{x}_1,G\mathbf{x}_1} & \bar{S}^{\mathbf{x}_1,G^2\mathbf{x}_1} \\ \bar{S}^{G\mathbf{x}_1,\mathbf{x}_1} & \bar{S}^{G\mathbf{x}_1,G\mathbf{x}_1} & \bar{S}^{G\mathbf{x}_1 G^2\mathbf{x}_1} \\ \bar{S}^{G^2\mathbf{x}_1,\mathbf{x}_1} & \bar{S}^{G^2\mathbf{x}_1,G\mathbf{x}_1} & \bar{S}^{G^2\mathbf{x}_1,G^2\mathbf{x}_1} \end{pmatrix} =: \bar{S}^{\mathbf{zz}},$$

where

$$\bar{S}^{GY,G^r\mathbf{x}_1} = \mathbb{E}\left[\left(\sum_{m=0}^{\infty}\beta_2^{0\prime}\tilde{\tilde{s}}^{\mathbf{x}_1}_{*m}(\mathbf{x}_i, a_i) + \beta_3^{0\prime}\tilde{\tilde{s}}^{\mathbf{x}_1}_{*,m+1}(\mathbf{x}_i, a_i) + \tilde{\tilde{s}}^{a}_{*m}(\mathbf{x}_i, a_i)\right)\left(\tilde{\tilde{s}}^{\mathbf{x}_1}_{*r}(\mathbf{x}_i, a_i)\right)'\right],\ r = 0, 1, 2$$

$$\bar{S}^{G^r\mathbf{x}_1,G^s\mathbf{x}_1} = \mathbb{E}\left[\tilde{\tilde{s}}^{\mathbf{x}_1}_{*r}(\mathbf{x}_i, a_i))\left(\tilde{\tilde{s}}^{\mathbf{x}_1}_{*s}(\mathbf{x}_i, a_i)\right)'\right],\ r, s = 0, 1, 2$$

$$\tilde{\tilde{s}}^{\mathbf{x}_1}_{*m}(\mathbf{x}_i, a_i) = \tilde{s}^{\mathbf{x}_1}_{*m}(\mathbf{x}_i, a_i) - \mathbb{E}[\tilde{s}^{\mathbf{x}_1}_{*m}(\mathbf{x}_i, a_i)|\mathbf{x}_{2i}, a_i]) \quad \textit{with} \quad \tilde{s}^{*\mathbf{x}_1}_{0}(\mathbf{x}_i, a_i) = \mathbf{x}_{1i}$$

$$\tilde{\tilde{s}}^{a}_{*m}(\mathbf{x}_i, a_i) = \tilde{s}^{a}_{*m}(\mathbf{x}_i, a_i) - \mathbb{E}[\tilde{s}^{a}_{*m}(\mathbf{x}_i, a_i)|\mathbf{x}_{2i}, a_i]) \quad \textit{with} \quad \tilde{s}^{a}_{*0}(\mathbf{x}_i, a_i) = h^{v}_{*}(\mathbf{x}_{2i}, a_i),$$

where $\tilde{\tilde{s}}^{\mathbf{x}_1}_{*m}(\mathbf{x}_i, a_i)$ *and* $\tilde{\tilde{s}}^{a}_{*m}(\mathbf{x}_i, a_i)$ *are defined recursively as in (S.1.3.4).*

Lemma 20. *Under Assumption 1, 3, and 11,*

$$\frac{1}{N}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{*}(\mathbf{x}_{2i}, a_i))(\mathbf{z}_i - \mathbf{h}^{\mathbf{z}}_{*}(\mathbf{x}_{2i}, a_i))'\sigma^2_{*}(\mathbf{x}_i, a_i) \to_p \bar{\mathbf{S}}^{\mathbf{zz}\sigma},$$

where the limit variance $\bar{\mathbf{S}}^{\mathbf{zz}\sigma}$ *is defined in Lemma 21.*

□

Lemma 21. *Under Assumption 1, 3, and 11,*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\mathbf{z}_i - \mathbf{h}_*^{\mathbf{z}}(\mathbf{x}_{2i}, a_i))\eta_{*i}^{\upsilon} \Rightarrow \mathcal{N}(0, \bar{\mathbf{S}}^{\mathbf{zz}\sigma}),$$

where

$$\bar{\mathbf{S}}^{\mathbf{zz}\sigma} = \begin{pmatrix} \bar{S}^{\mathbf{x}_1\mathbf{x}_1\sigma} & \bar{S}^{\mathbf{x}_1 G\mathbf{x}_1\sigma} & \bar{S}^{\mathbf{x}_1 G^2\mathbf{x}_1\sigma} \\ \bar{S}^{G\mathbf{x}_1\mathbf{x}_1\sigma} & \bar{S}^{G\mathbf{x}_1 G\mathbf{x}_1\sigma} & \bar{S}^{G\mathbf{x}_1 G^2\mathbf{x}_1\sigma} \\ \bar{S}^{G^2\mathbf{x}_1\mathbf{x}_1\sigma} & \bar{S}^{G^2\mathbf{x}_1 G\mathbf{x}_1\sigma} & \bar{S}^{G^2\mathbf{x}_1 G^2\mathbf{x}_1\sigma} \end{pmatrix}$$

and

$$\begin{aligned} \bar{S}^{G^r\mathbf{x}_1 G^s\mathbf{x}_1\sigma} &= \mathbb{E}\left[\tilde{\tilde{s}}_{*r}^{\mathbf{x}_1}(\mathbf{x}_i, a_i))\left(\tilde{\tilde{s}}_{*s}^{\mathbf{x}_1}(\mathbf{x}_i, a_i)\right)'\sigma_*^2(\mathbf{x}_i, a_i)\right],\ r, s = 0, 1, 2 \\ \tilde{\tilde{s}}_{*m}^{\mathbf{x}_1}(\mathbf{x}_i, a_i) &= \tilde{s}_{*m}^{\mathbf{x}_1}(\mathbf{x}_i, a_i) - \mathbb{E}[\tilde{s}_{*m}^{\mathbf{x}_1}(\mathbf{x}_i, a_i)|\mathbf{x}_{2i}, a_i]) \quad \textit{with} \quad \tilde{s}_{*0}^{\mathbf{x}_1}(\mathbf{x}_i, a_i) = \mathbf{x}_{1i} \\ \sigma_*^2(\mathbf{x}_i, a_i) &:= \mathbb{E}[(\eta_{*i}^{\upsilon})^2|\mathbf{x}_i, a_i] = \mathbb{E}[(\upsilon_i - \mathbb{E}[\upsilon_i|\mathbf{x}_{2i}, a_i])^2|\mathbf{x}_i, a_i], \end{aligned}$$

where $\tilde{\tilde{s}}_{*m}^{\mathbf{x}_1}(\mathbf{x}_i, a_i)$ *is defined recursively as in (S.1.3.4).*

Appendix S.3. Supplementary Monte Carlo results

TABLE 4. **Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \exp(3a_i)$ and $K_N = 4$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0$	0.081	0.074	0.011	0.007	-0.007	0.126	0.115	0.002	-0.018	-0.008	*mean bias*
	(0.140)	(0.178)	(0.214)	(0.308)	(0.165)	(0.109)	(0.163)	(0.205)	(0.331)	(0.135)	*std*
	0.383	0.089	0.057	0.061	0.048	0.537	0.109	0.060	0.068	0.044	*size*
$\beta_2 = 1$	-0.005	-0.004	0.000	0.001	0.000	-0.004	-0.003	0.000	0.000	0.000	*mean bias*
	(0.031)	(0.033)	(0.035)	(0.036)	(0.031)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.312	0.061	0.069	0.077	0.047	0.334	0.072	0.061	0.063	0.056	*size*
$\beta_3 = 1$	-0.113	-0.100	-0.006	-0.003	0.012	-0.201	-0.181	0.007	0.024	0.016	*mean bias*
	(0.268)	(0.335)	(0.384)	(0.437)	(0.298)	(0.214)	(0.315)	(0.379)	(0.458)	(0.247)	*std*
	0.348	0.076	0.057	0.064	0.051	0.465	0.092	0.057	0.057	0.050	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.5$	0.042	0.042	0.007	0.004	-0.004	0.062	0.064	0.001	-0.012	-0.004	*mean bias*
	(0.076)	(0.104)	(0.135)	(0.197)	(0.095)	(0.059)	(0.093)	(0.129)	(0.209)	(0.073)	*std*
	0.385	0.091	0.059	0.062	0.046	0.540	0.110	0.060	0.068	0.043	*size*
$\beta_2 = 1$	-0.003	-0.003	0.000	0.001	0.000	-0.003	-0.003	0.000	0.000	0.000	*mean bias*
	(0.031)	(0.032)	(0.034)	(0.035)	(0.031)	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)	*std*
	0.315	0.060	0.070	0.077	0.049	0.335	0.071	0.064	0.064	0.058	*size*
$\beta_3 = 1$	-0.067	-0.065	-0.001	0.002	0.008	-0.116	-0.122	0.008	0.019	0.011	*mean bias*
	(0.196)	(0.259)	(0.313)	(0.343)	(0.223)	(0.153)	(0.237)	(0.306)	(0.348)	(0.172)	*std*
	0.338	0.078	0.057	0.070	0.051	0.425	0.083	0.058	0.062	0.046	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.8$	0.015	0.018	0.004	0.002	-0.001	0.019	0.027	0.001	-0.007	-0.001	*mean bias*
	(0.033)	(0.049)	(0.080)	(0.118)	(0.045)	(0.025)	(0.042)	(0.076)	(0.125)	(0.031)	*std*
	0.388	0.090	0.058	0.062	0.041	0.539	0.110	0.060	0.068	0.042	*size*
$\beta_2 = 1$	-0.001	-0.002	0.000	0.001	-0.000	-0.001	-0.001	0.000	0.000	-0.000	*mean bias*
	(0.030)	(0.032)	(0.034)	(0.035)	(0.030)	(0.020)	(0.020)	(0.020)	(0.021)	(0.020)	*std*
	0.307	0.060	0.071	0.077	0.046	0.331	0.066	0.063	0.063	0.054	*size*
$\beta_3 = 1$	-0.021	-0.032	0.001	0.004	0.004	-0.035	-0.065	0.009	0.015	0.006	*mean bias*
	(0.135)	(0.192)	(0.264)	(0.281)	(0.154)	(0.099)	(0.164)	(0.256)	(0.276)	(0.105)	*std*
	0.314	0.072	0.058	0.065	0.041	0.361	0.070	0.058	0.060	0.041	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg_i, x_{2i}})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 6$, $K^*_{100,0.5} = 3$, $K^*_{100,0.8} = 4$
$K^*_{250,0.2} = 8$, $K^*_{250,0.5} = 3$, $K^*_{250,0.8} = 7$

TABLE 5. **Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \exp(3a_i)$ and $K_N = 6$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0$	0.081	0.074	0.010	0.006	-0.007	0.126	0.115	0.003	-0.018	-0.008	*mean bias*
	(0.140)	(0.178)	(0.216)	(0.314)	(0.165)	(0.109)	(0.163)	(0.206)	(0.332)	(0.135)	*std*
	0.383	0.089	0.059	0.067	0.048	0.537	0.109	0.060	0.069	0.044	*size*
$\beta_2 = 1$	-0.005	-0.004	0.001	0.001	0.000	-0.004	-0.003	0.000	0.000	0.000	*mean bias*
	(0.031)	(0.033)	(0.035)	(0.036)	(0.031)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.312	0.061	0.072	0.078	0.047	0.334	0.072	0.065	0.065	0.056	*size*
$\beta_3 = 1$	-0.113	-0.100	-0.005	-0.000	0.012	-0.201	-0.181	0.005	0.023	0.016	*mean bias*
	(0.268)	(0.335)	(0.389)	(0.443)	(0.298)	(0.214)	(0.315)	(0.381)	(0.460)	(0.247)	*std*
	0.348	0.076	0.061	0.067	0.051	0.465	0.092	0.058	0.058	0.050	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.5$	0.042	0.042	0.006	0.003	-0.004	0.062	0.064	0.002	-0.012	-0.004	*mean bias*
	(0.076)	(0.104)	(0.137)	(0.200)	(0.095)	(0.059)	(0.093)	(0.129)	(0.210)	(0.073)	*std*
	0.385	0.091	0.060	0.067	0.046	0.540	0.110	0.060	0.069	0.043	*size*
$\beta_2 = 1$	-0.003	-0.003	0.000	0.001	0.000	-0.003	-0.003	0.000	0.000	0.000	*mean bias*
	(0.031)	(0.032)	(0.035)	(0.036)	(0.031)	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)	*std*
	0.315	0.060	0.072	0.072	0.049	0.335	0.071	0.065	0.069	0.058	*size*
$\beta_3 = 1$	-0.067	-0.065	-0.000	0.004	0.008	-0.116	-0.122	0.006	0.018	0.011	*mean bias*
	(0.196)	(0.259)	(0.317)	(0.348)	(0.223)	(0.153)	(0.237)	(0.307)	(0.350)	(0.172)	*std*
	0.338	0.078	0.059	0.067	0.051	0.425	0.083	0.059	0.066	0.046	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.8$	0.015	0.018	0.004	0.002	-0.001	0.019	0.027	0.001	-0.007	-0.001	*mean bias*
	(0.033)	(0.049)	(0.081)	(0.121)	(0.045)	(0.025)	(0.042)	(0.077)	(0.126)	(0.031)	*std*
	0.388	0.090	0.060	0.068	0.041	0.539	0.110	0.060	0.069	0.042	*size*
$\beta_2 = 1$	-0.001	-0.002	0.000	0.000	-0.000	-0.001	-0.001	0.000	0.000	-0.000	*mean bias*
	(0.030)	(0.032)	(0.034)	(0.035)	(0.030)	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)	*std*
	0.307	0.060	0.072	0.070	0.046	0.331	0.066	0.068	0.068	0.054	*size*
$\beta_3 = 1$	-0.021	-0.032	0.002	0.006	0.004	-0.035	-0.065	0.007	0.014	0.006	*mean bias*
	(0.135)	(0.192)	(0.268)	(0.285)	(0.154)	(0.099)	(0.164)	(0.257)	(0.278)	(0.105)	*std*
	0.314	0.072	0.063	0.069	0.041	0.361	0.070	0.057	0.065	0.041	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg_i}, x_{2i})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 6$, $K^*_{100,0.5} = 3$, $K^*_{100,0.8} = 4$
$K^*_{250,0.2} = 8$, $K^*_{250,0.5} = 3$, $K^*_{250,0.8} = 7$

TABLE 6. **Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \exp(3a_i)$ and $K_N = 8$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0$	0.081	0.074	0.009	0.007	-0.007	0.126	0.115	0.003	-0.017	-0.008	*mean bias*
	(0.140)	(0.178)	(0.219)	(0.314)	(0.165)	(0.109)	(0.163)	(0.207)	(0.332)	(0.135)	*std*
	0.383	0.089	0.058	0.069	0.048	0.537	0.109	0.060	0.066	0.044	*size*
$\beta_2 = 1$	-0.005	-0.004	0.000	0.001	0.000	-0.004	-0.003	0.001	0.000	0.000	*mean bias*
	(0.031)	(0.033)	(0.036)	(0.036)	(0.031)	(0.020)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.312	0.061	0.074	0.074	0.047	0.334	0.072	0.063	0.067	0.056	*size*
$\beta_3 = 1$	-0.113	-0.100	-0.003	-0.001	0.012	-0.201	-0.181	0.004	0.023	0.016	*mean bias*
	(0.268)	(0.335)	(0.394)	(0.443)	(0.298)	(0.214)	(0.315)	(0.382)	(0.461)	(0.247)	*std*
	0.348	0.076	0.058	0.064	0.051	0.465	0.092	0.060	0.059	0.050	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.5$	0.042	0.042	0.005	0.003	-0.004	0.062	0.064	0.002	-0.011	-0.004	*mean bias*
	(0.076)	(0.104)	(0.138)	(0.201)	(0.095)	(0.059)	(0.093)	(0.130)	(0.210)	(0.073)	*std*
	0.385	0.091	0.059	0.071	0.046	0.540	0.110	0.060	0.066	0.043	*size*
$\beta_2 = 1$	-0.003	-0.003	0.000	0.000	0.000	-0.003	-0.003	0.000	0.000	0.000	*mean bias*
	(0.031)	(0.032)	(0.035)	(0.035)	(0.031)	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)	*std*
	0.315	0.060	0.074	0.067	0.049	0.335	0.071	0.065	0.071	0.058	*size*
$\beta_3 = 1$	-0.067	-0.065	0.001	0.003	0.008	-0.116	-0.122	0.006	0.018	0.011	*mean bias*
	(0.196)	(0.259)	(0.321)	(0.348)	(0.223)	(0.153)	(0.237)	(0.308)	(0.350)	(0.172)	*std*
	0.338	0.078	0.062	0.069	0.051	0.425	0.083	0.063	0.065	0.046	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.8$	0.015	0.018	0.003	0.002	-0.001	0.019	0.027	0.001	-0.007	-0.001	*mean bias*
	(0.033)	(0.049)	(0.082)	(0.121)	(0.045)	(0.025)	(0.042)	(0.077)	(0.126)	(0.031)	*std*
	0.388	0.090	0.060	0.071	0.041	0.539	0.110	0.060	0.066	0.042	*size*
$\beta_2 = 1$	-0.001	-0.002	-0.000	0.000	-0.000	-0.001	-0.001	0.000	0.000	-0.000	*mean bias*
	(0.030)	(0.032)	(0.034)	(0.035)	(0.030)	(0.020)	(0.020)	(0.021)	(0.021)	(0.020)	*std*
	0.307	0.060	0.074	0.067	0.046	0.331	0.066	0.067	0.070	0.054	*size*
$\beta_3 = 1$	-0.021	-0.032	0.004	0.006	0.004	-0.035	-0.065	0.007	0.014	0.006	*mean bias*
	(0.135)	(0.192)	(0.272)	(0.285)	(0.154)	(0.099)	(0.164)	(0.258)	(0.278)	(0.105)	*std*
	0.314	0.072	0.067	0.072	0.041	0.361	0.070	0.063	0.064	0.041	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg_i, x_{2i}})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 6$, $K^*_{100,0.5} = 3$, $K^*_{100,0.8} = 4$
$K^*_{250,0.2} = 8$, $K^*_{250,0.5} = 3$, $K^*_{250,0.8} = 7$

TABLE 7. **Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \sin(3a_i)$ and $K_N = 4$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.200$	-1.681	-0.398	-0.101	0.006	-0.010	-0.445	-0.630	-0.091	-0.024	-0.019	*mean bias*
	(26.702)	(0.325)	(0.243)	(0.349)	(0.246)	(160.512)	(0.352)	(0.232)	(0.372)	(0.297)	*std*
	0.638	0.301	0.075	0.066	0.018	0.698	0.563	0.080	0.064	0.014	*size*
$\beta_2 = 1$	0.128	0.022	0.002	0.001	0.001	-0.043	0.019	0.001	0.000	0.001	*mean bias*
	(2.379)	(0.041)	(0.036)	(0.037)	(0.035)	(6.620)	(0.024)	(0.021)	(0.021)	(0.022)	*std*
	0.511	0.109	0.073	0.088	0.016	0.616	0.137	0.056	0.062	0.018	*size*
$\beta_3 = 1$	2.674	0.520	0.094	-0.002	0.015	0.627	0.937	0.097	0.030	0.033	*mean bias*
	(41.110)	(0.530)	(0.411)	(0.473)	(0.425)	(277.957)	(0.605)	(0.407)	(0.499)	(0.532)	*std*
	0.597	0.203	0.062	0.062	0.017	0.705	0.430	0.071	0.054	0.013	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.5$	-1.095	-0.211	-0.062	0.003	-0.005	-0.462	-0.321	-0.056	-0.015	-0.009	*mean bias*
	(14.839)	(0.170)	(0.151)	(0.219)	(0.140)	(25.186)	(0.171)	(0.142)	(0.230)	(0.147)	*std*
	0.779	0.313	0.075	0.066	0.023	0.810	0.576	0.080	0.064	0.015	*size*
$\beta_2 = 1$	0.148	0.015	0.001	0.001	0.000	0.023	0.013	0.001	0.000	0.000	*mean bias*
	(2.399)	(0.037)	(0.035)	(0.036)	(0.033)	(1.468)	(0.023)	(0.021)	(0.021)	(0.021)	*std*
	0.497	0.100	0.074	0.085	0.021	0.599	0.096	0.058	0.065	0.020	*size*
$\beta_3 = 1$	2.154	0.317	0.060	0.003	0.011	0.910	0.571	0.063	0.023	0.022	*mean bias*
	(30.259)	(0.364)	(0.329)	(0.363)	(0.310)	(55.722)	(0.383)	(0.321)	(0.369)	(0.336)	*std*
	0.701	0.174	0.059	0.065	0.018	0.804	0.378	0.067	0.056	0.015	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.8$	-0.168	-0.093	-0.036	0.001	-0.002	-0.320	-0.136	-0.033	-0.009	-0.004	*mean bias*
	(0.285)	(0.079)	(0.088)	(0.130)	(0.070)	(5.928)	(0.075)	(0.083)	(0.136)	(0.065)	*std*
	0.813	0.321	0.075	0.067	0.025	0.906	0.592	0.080	0.066	0.022	*size*
$\beta_2 = 1$	0.026	0.008	0.000	0.001	-0.000	0.028	0.008	0.001	0.000	0.000	*mean bias*
	(0.072)	(0.034)	(0.035)	(0.035)	(0.031)	(0.470)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.364	0.084	0.075	0.084	0.021	0.461	0.072	0.058	0.066	0.020	*size*
$\beta_3 = 1$	0.401	0.162	0.037	0.005	0.007	0.879	0.301	0.041	0.018	0.013	*mean bias*
	(0.817)	(0.251)	(0.274)	(0.293)	(0.219)	(18.665)	(0.244)	(0.265)	(0.288)	(0.203)	*std*
	0.585	0.127	0.060	0.069	0.024	0.846	0.279	0.064	0.054	0.020	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg}_i, x_{2i})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 8$, $K^*_{100,0.5} = 7$, $K^*_{100,0.8} = 6$
$K^*_{250,0.2} = 7$, $K^*_{250,0.5} = 5$, $K^*_{250,0.8} = 7$

TABLE 8. **Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \sin(3a_i)$ and $K_N = 6$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.200$	-1.681	-0.398	-0.102	0.005	-0.010	-0.445	-0.630	-0.089	-0.024	-0.019	*mean bias*
	(26.702)	(0.325)	(0.246)	(0.354)	(0.246)	(160.512)	(0.352)	(0.233)	(0.373)	(0.297)	*std*
	0.638	0.301	0.081	0.071	0.018	0.698	0.563	0.076	0.071	0.014	*size*
$\beta_2 = 1$	0.128	0.022	0.002	0.001	0.001	-0.043	0.019	0.001	0.000	0.001	*mean bias*
	(2.379)	(0.041)	(0.037)	(0.037)	(0.035)	(6.620)	(0.024)	(0.021)	(0.021)	(0.022)	*std*
	0.511	0.109	0.079	0.081	0.016	0.616	0.137	0.058	0.065	0.018	*size*
$\beta_3 = 1$	2.674	0.520	0.096	0.001	0.015	0.627	0.937	0.094	0.029	0.033	*mean bias*
	(41.110)	(0.530)	(0.416)	(0.479)	(0.425)	(277.957)	(0.605)	(0.409)	(0.501)	(0.532)	*std*
	0.597	0.203	0.060	0.062	0.017	0.705	0.430	0.072	0.059	0.013	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.5$	-1.095	-0.211	-0.063	0.002	-0.005	-0.462	-0.321	-0.055	-0.015	-0.009	*mean bias*
	(14.839)	(0.170)	(0.152)	(0.222)	(0.140)	(25.186)	(0.171)	(0.143)	(0.231)	(0.147)	*std*
	0.779	0.313	0.079	0.073	0.023	0.810	0.576	0.076	0.072	0.015	*size*
$\beta_2 = 1$	0.148	0.015	0.001	0.001	0.000	0.023	0.013	0.001	0.000	0.000	*mean bias*
	(2.399)	(0.037)	(0.036)	(0.036)	(0.033)	(1.468)	(0.023)	(0.021)	(0.021)	(0.021)	*std*
	0.497	0.100	0.079	0.080	0.021	0.599	0.096	0.057	0.066	0.020	*size*
$\beta_3 = 1$	2.154	0.317	0.062	0.005	0.011	0.910	0.571	0.061	0.022	0.022	*mean bias*
	(30.259)	(0.364)	(0.333)	(0.368)	(0.310)	(55.722)	(0.383)	(0.323)	(0.371)	(0.336)	*std*
	0.701	0.174	0.059	0.061	0.018	0.804	0.378	0.069	0.060	0.015	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
$\beta_1 = 0.8$	-0.168	-0.093	-0.037	0.001	-0.002	-0.320	-0.136	-0.032	-0.009	-0.004	*mean bias*
	(0.285)	(0.079)	(0.089)	(0.132)	(0.070)	(5.928)	(0.075)	(0.084)	(0.136)	(0.065)	*std*
	0.813	0.321	0.079	0.073	0.025	0.906	0.592	0.077	0.072	0.022	*size*
$\beta_2 = 1$	0.026	0.008	0.000	0.001	-0.000	0.028	0.008	0.001	-0.000	0.000	*mean bias*
	(0.072)	(0.034)	(0.035)	(0.036)	(0.031)	(0.470)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.364	0.084	0.079	0.084	0.021	0.461	0.072	0.057	0.067	0.020	*size*
$\beta_3 = 1$	0.401	0.162	0.038	0.007	0.007	0.879	0.301	0.040	0.017	0.013	*mean bias*
	(0.817)	(0.251)	(0.278)	(0.296)	(0.219)	(18.665)	(0.244)	(0.266)	(0.289)	(0.203)	*std*
	0.585	0.127	0.058	0.072	0.024	0.846	0.279	0.066	0.059	0.020	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg}_i, x_{2i})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 8$, $K^*_{100,0.5} = 7$, $K^*_{100,0.8} = 6$
$K^*_{250,0.2} = 7$, $K^*_{250,0.5} = 5$, $K^*_{250,0.8} = 7$

TABLE 9. **Polynomial Sieve**: Parameter values across 1000 Monte Carlo replications with $h(a) = \sin(3a_i)$ and $K_N = 8$

N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
	-1.681	-0.398	-0.103	0.005	-0.010	-0.445	-0.630	-0.089	-0.023	-0.019	*mean bias*
$\beta_1 = 0.200$	(26.702)	(0.325)	(0.249)	(0.354)	(0.246)	(160.512)	(0.352)	(0.234)	(0.374)	(0.297)	*std*
	0.638	0.301	0.081	0.073	0.018	0.698	0.563	0.078	0.068	0.014	*size*
	0.128	0.022	0.002	0.001	0.001	-0.043	0.019	0.001	0.000	0.001	*mean bias*
$\beta_2 = 1$	(2.379)	(0.041)	(0.037)	(0.037)	(0.035)	(6.620)	(0.024)	(0.021)	(0.021)	(0.022)	*std*
	0.511	0.109	0.081	0.078	0.016	0.616	0.137	0.055	0.065	0.018	*size*
	2.674	0.520	0.099	0.000	0.015	0.627	0.937	0.093	0.029	0.033	*mean bias*
$\beta_3 = 1$	(41.110)	(0.530)	(0.422)	(0.480)	(0.425)	(277.957)	(0.605)	(0.411)	(0.501)	(0.532)	*std*
	0.597	0.203	0.062	0.065	0.017	0.705	0.430	0.073	0.060	0.013	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
	-1.095	-0.211	-0.064	0.002	-0.005	-0.462	-0.321	-0.055	-0.015	-0.009	*mean bias*
$\beta_1 = 0.5$	(14.839)	(0.170)	(0.154)	(0.222)	(0.140)	(25.186)	(0.171)	(0.144)	(0.231)	(0.147)	*std*
	0.779	0.313	0.081	0.074	0.023	0.810	0.576	0.078	0.070	0.015	*size*
	0.148	0.015	0.001	0.001	0.000	0.023	0.013	0.001	0.000	0.000	*mean bias*
$\beta_2 = 1$	(2.399)	(0.037)	(0.036)	(0.036)	(0.033)	(1.468)	(0.023)	(0.021)	(0.021)	(0.021)	*std*
	0.497	0.100	0.076	0.078	0.021	0.599	0.096	0.058	0.064	0.020	*size*
	2.154	0.317	0.064	0.004	0.011	0.910	0.571	0.061	0.022	0.022	*mean bias*
$\beta_3 = 1$	(30.259)	(0.364)	(0.338)	(0.368)	(0.310)	(55.722)	(0.383)	(0.324)	(0.371)	(0.336)	*std*
	0.701	0.174	0.061	0.061	0.018	0.804	0.378	0.068	0.061	0.015	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					
N	100					250					
CF	(0)	(1)	(2)	(3)	(4)	(0)	(1)	(2)	(3)	(4)	
	-0.168	-0.093	-0.037	0.001	-0.002	-0.320	-0.136	-0.032	-0.009	-0.004	*mean bias*
$\beta_1 = 0.8$	(0.285)	(0.079)	(0.090)	(0.132)	(0.070)	(5.928)	(0.075)	(0.084)	(0.136)	(0.065)	*std*
	0.813	0.321	0.081	0.074	0.025	0.906	0.592	0.078	0.071	0.022	*size*
	0.026	0.008	-0.000	0.000	-0.000	0.028	0.008	0.001	-0.000	0.000	*mean bias*
$\beta_2 = 1$	(0.072)	(0.034)	(0.035)	(0.036)	(0.031)	(0.470)	(0.021)	(0.021)	(0.021)	(0.020)	*std*
	0.364	0.084	0.075	0.085	0.021	0.461	0.072	0.057	0.065	0.020	*size*
	0.401	0.162	0.040	0.006	0.007	0.879	0.301	0.039	0.016	0.013	*mean bias*
$\beta_3 = 1$	(0.817)	(0.251)	(0.281)	(0.296)	(0.219)	(18.665)	(0.244)	(0.268)	(0.290)	(0.203)	*std*
	0.585	0.127	0.057	0.073	0.024	0.846	0.279	0.062	0.058	0.020	*size*
	$\hat{a}$ - mean bias=0.016, median bias=0.007					$\hat{a}$ - mean bias=0.006, median bias=0.003					

CF - control function. (0) - none, (1) - $\hat{a}_i$, (2) - $\hat{h}(\hat{a}_i)$, (3) - $\hat{h}(\widehat{deg_i}, x_{2i})$, (4) - $h(a_i)$.
Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.
The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$
$K^*_{100,0.2} = 8$, $K^*_{100,0.5} = 7$, $K^*_{100,0.8} = 6$
$K^*_{250,0.2} = 7$, $K^*_{250,0.5} = 5$, $K^*_{250,0.8} = 7$