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

IDA JOHNSSON

psyML

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: January 30, 2019.

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 5.2 of the paper. We also appreciate the comments and discussions of the participants at the 2015 USC Dornsife INET Conference on Networks, the 2016 North American Summer Meeting of the Econometric Society, the 2016 California Econometrics Conference, the 2017 Asian Meeting of Econometric Society, the 2017 IAAE conference, the 2018 UCLA-USC Mini Conference, the econometrics seminars at University of British Columbia and Ohio State University. 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 Johnsson: ida@psyml.co. Hyungsik Roger Moon: Corresponding 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 application examples are found in recent review papers such as Blume et al. (2010), Manski (2000), Epple and Romano (2011), Brock and Durlauf (2001) and Graham (2011).

Many models considered in earlier literature assume that connections between individuals are independent of unobserved individual characteristics that influence outcomes. However, assuming exogeneity of the network or peer group is 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 unobserved variables that influence both friendship formation and individual grades. However, even if a study controls for observable individual characteristics such as gender, age, race and parents’ education, it is likely to omit factors that influence both students’ choice of friends and their GPA, for example parental expectations, psychological disorders or non-reported substance use. For more examples of endogenous peer groups see Brock and Durlauf (2001), Weinberg (2007), Shalizi (2012) and Hsieh and Lee (2016), among others.

In this paper we propose a method for estimating a linear-in-means model of peer effects, where the peer group is defined by a network that is endogenous in the outcome equation. Our model allows for 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, and 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. Second, 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 an estimator of the unobserved individual heterogeneity and the other one based on the node degrees of the network. Our estimation method is semiparametric because we do not restrict the functional form of the control function. Finally, 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. Other peer effects papers that considered endogenously formed peer groups and controlled the endogeneity via various control functions include Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016), Qu and Lee (2015), Arduini et al. (2015) and Auerbach (2016). We provide more detail on these paper in Section 2.3.

The remainder of the paper is organized as follows. In Section 2 we present a high level description of our approach and provide intuition as to its empirical applications. In Section 3 we formally present our model. In Section 4 we show how to identify peer effects using control functions. Estimation is discussed in Section 5 and in Section 6 we discuss the limiting distribution of the estimator and propose standard errors. In Section 7 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 1

We acknowledge that this approach is developed based on the idea provided by one of the referees. We thank the referee.
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. Main Idea

In this section we introduce a simple model in order to illustrate the main points of our approach. A more general model and detailed discussion of the model will follow later.

2.1. Simple Model. A simple peer effect model for the illustration of the main idea is

\[ y_i = \beta^0 \left( \frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}} \right) + v_i, \quad i = 1, ..., N, \]  

(2.1)

where \( x_i \) is a measure of observable characteristics of individual \( i \) and \( d_{ij} \) is an indicator of individual \( i \)'s peer, so \( d_{ij} = 1 \) if \( i \) and \( j \) are directly linked and 0 otherwise. In (2.1), the regressor of interest is the average of the characteristics of those individuals who are linked with \( i \), \( \frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}} \). For simplicity, we assume that \( x_i \) are exogenous with respect to all the unobserved components of the model, this will be relaxed later.

For the link formation, we consider the following dyadic network formation model,

\[ d_{ij} = \mathbb{I}(g(a_i, a_j) \geq u_{ij}) \mathbb{I}(i \neq j), \]  

(2.2)

where \( a_i \) and \( a_j \) are unobserved individual specific characteristics, \( u_{ij} \) is a link specific component, and \( g(\cdot, \cdot) \) is some function. It should be noted that this model of network formation does not allow for network effects in link formation, as a link between \( i \) and \( j \) only depends on the characteristics of \( i \) and \( j \). The unobserved individual characteristic \( a_i \) can be interpreted as social capital that increases the likelihood of forming a link. Depending on the context this could for example be trustworthiness, socioeconomic status or outspokenness. For example, [Weerdt and Fafchamps (2011)] measure the risk sharing links between households in Tanzania and the construct links between households based on the question whom individuals could “personally rely on for help”. [Fafchamps and Gubert (2007)] examine the
formation of risk-sharing networks using data from the rural Philippines. Banerjee et al. (2013) examine how participation in micro-finance diffuses through a social network which they measure using lending and trust. In these settings, we can think of $a_i$ as a measure of individual trustworthiness and integrity in financial matters.

Ductor et al. (2014) analyze whether knowledge of a researcher’s co-authorship network is helpful in predicting his or her productivity. In this setting $a_i$ can be interpreted as some unobserved productivity trait that induces the researcher to have more coauthors, and also to be more productive at writing papers.

2.2. Control Function and Its Implementation. The key feature of the peer effect model (2.1) and (2.2) is that individual $i$’s unobserved characteristic $a_i$, which impacts link formation, is correlated with $v_i$, $i$’s unobserved characteristic that affects the outcome $y_i$. For example, $a_i$ could be an unobserved component that affects the a researcher’s publication rate $y_i$, and also his or her co-authorship relationships, $d_{ij}$. Alternatively we can think of a situation where there are two types of agents: popular and unpopular agents. The popular agents are both more likely to be friends with other agents, and popular agents have better outcomes even in the absence of a peer effect. Then the peer formation $d_{ij}$ becomes correlated with the unobserved component $v_i$ of the outcome, and, as a consequence, the regressor of the peer effect, $\sum_{j \neq i} d_{ij}^{x_j} \sum_{j \neq i} d_{ij}^{x_i}$, becomes endogenous.

In this paper we use a control function method to handle the endogenous peer group problem. Let $D_N$ be the adjacency matrix that describes the network links $d_{ij}$. Suppose that the unobserved characteristics $(a_i, v_i)$ and $u_{ij}$ are randomly drawn over $i$ and $(i, j)$, respectively. Also assume that $u_{ij}$ is independent of $(a_i, v_i)$. Then, for any $i \neq j$, the link $d_{ij} = 1(g(a_i, a_j) \geq u_{ij})$ and $v_i$ are dependent only through $a_i$. Therefore, controlling for $a_i$, the network $D_N$ and $v_i$ become mean independent, that is,

$$\mathbb{E}(v_i \mid D_N, a_i) = \mathbb{E}(v_i \mid a_i) =: h(a_i).$$
Suppose that we observe $a_i$. Consider the outcome equation which controls for $a_i$ non-parametrically,

$$y_i = \beta_0 \left( \frac{\sum_{j \neq i} d_{ij} x_j}{\sum_{j \neq i} d_{ij}} \right) + h(a_i) + \varepsilon_i,$$

where $\varepsilon_i := v_i - h(a_i)$. Once we control the endogeneity of the network with $a_i$, then the regressor of the peer effect becomes exogenous, and we can estimate the peer effect coefficient $\beta_0$ using the conventional partially linear regression estimation method (e.g. Robinson (1988b)).

However, in most empirical applications, $a_i$ is not observed. Then the question becomes how to implement the control function approach. In this paper, as the main methodological contribution, we propose the following two procedures. Both procedures are implemented with a single snapshot of an observed network.

(i) First, suppose that $a_i$ can be consistently estimated. An example can be found in Graham (2017) with the specification $g(\alpha_i, \alpha_j) = \alpha_i + \alpha_j$. Then, we estimate $\beta_0$ by running the partially linear regression of $y_i$ on $\sum_{j \neq i} g_{ij} x_j$ and $h(\hat{a}_i)$ as in Robinson (1988a).

(ii) The second method is to use an observed control function that carries the same information as $a_i$. For this, first notice by the WLLN,

$$\text{deg}_i := \frac{1}{N} \sum_{j \neq i} d_{ij} = \frac{1}{N} \sum_{j \neq i} \mathbb{I}(g(a_i, a_j) \geq u_{ij}) \rightarrow_{p} \mathbb{P}(d_{ij} = 1 \mid a_i).$$

Suppose that the network formation probability conditional on $a_i$, $\mathbb{P}(d_{ij} = 1 \mid a_i)$, is a monotonic function of $a_i$. A sufficient condition for this is that $g(\cdot, a_j)$ is monotonic in the same direction for all $a_j$, for example

$$g(a_i, a_j) = a_i + a_j - \tau|a_i - a_j| \quad (2.3)$$

with $0 \leq \tau \leq 1$. In this case, the limit of the node degree carries the same information as the control function $a_i$, which justifies $\text{deg}_i$ as a proxy of the control function $a_i$, that is, $\mathbb{E}(v_i \mid a_i) \simeq \mathbb{E}(v_i \mid \text{deg}_i) =: h_*(\text{deg}_i)$. The peer effect coefficient $\beta_0$ can be
estimated by using \( \text{deg}_i \) as a control function. More specifically, we estimate \( \beta^0 \) by running the partially linear regression of \( y_i \) on \( \sum_{j \neq i} g_{ij} x_j \) and \( h_i(\text{deg}_i) \).

Intuitively, unobserved characteristics \( a_i \) drive heterogeneous degree sequences. We can therefore control for degree when estimating peer effects, ignoring the specific choice of structural model explaining heterogeneous degrees.

The use of degree as a control function requires much less restriction on the specification of the network. Intuitively, the unobserved node (or individual) fixed effects \( a_i \) control for heterogeneous degree sequences, so from an economic point of view what needs to be controlled is the agent’s degree. Hence the validity of the control function approach that uses \( \text{deg}_i \). This approach does not require a specification of the specific structural model explaining heterogeneous degree sequences. Consistent estimation of \( a_i \) usually requires a specific functional form. For example, Graham (2017) assumed an additive model and Chen et al. (2018) require an interactive form. However, there is a disadvantage in the degree approach. The degree approach cannot identify the coefficient of the observed exogenous regressor if the same regressor also impacts the network formation.

In Section 3, we generalize the simple model (2.1) by allowing for an additional peer effect, \( \sum_{j \neq i} d_{ij} y_j / \sum_{j \neq i} d_{ij} \), known as the endogenous peer effect, which measures the effects of the outcomes of the peer group on an individual outcome. In this case we have to deal with two kinds of endogeneity in the peer effect regressors, one from the endogenous regressors \( y_j \) and the other one from the endogenous peers \( d_{ij} \). In Section 3, we also generalize the dyadic network formation model by introducing a dyadic component based on observed individual characteristics. We provide application examples of the general model and discuss its features there. The identification of the peer effects in the general model will be discussed in Section 4. In Section 5, we shows how to implement the two aforementioned estimation methods in the general framework.
2.3. **Related Literature.** Closely related papers that adopt a control function approach include Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2016), Qu and Lee (2015), Arduini et al. (2015) and Auerbach (2016). Our paper adopts a frequentist approach based on a 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 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. Also, in Goldsmith-Pinkham and Imbens (2013), unobserved components account for homophily in link formation, whereas in our setup they mainly drive degree heterogeneity but are allowed to account for homophily as well, as in the example (2.3).

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 and the idiosyncratic network formation error, like the conventional sample selection model. This model can be interpreted as meeting opportunities being correlated with unobserved ability of the agent that affects the outcome. Arduini et al. (2015) consider control functions (both parametric and semiparametric) to deal with the selection bias problem and propose a semiparametric estimator that uses a power series to approximate selectivity bias terms. 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.
Among the aforementioned related papers, probably the one most closely related to ours is Auerbach (2016), and we want to discuss the differences between the two papers in more detail. The outcome model of Auerbach (2016) is a partially linear regression model where the nonparametric component is an unknown function of the unobserved network heterogeneity,

\[ y_i = \beta_0 x_i + h(a_i) + \varepsilon_i, \]

\[ d_{ij} = \mathbb{I}(g(a_i, a_j) \geq u_{ij}) \mathbb{I}(i \neq j). \]

In the simple peer effect example, the exogenous peer effect corresponds to the regressor \( x_i \) above. The network formation is the same as (2.2).

To compare the identification ideas, let’s assume that \( a_i \sim U[-1/2, 1/2] \) and \( u_{ij} \sim U[0, 1] \). In this case, \( d_i := (d_{i1}, ..., d_{in})' \) and the distribution of \( d_i \) of node \( i \), whose characteristic is \( a_i \), is fully characterized by the link formation probability profile \( g(a_i, \cdot) \).

The key condition of Auerbach (2016) is that \( h(a_i) \) and the the link formation distribution profile \( g_i(\cdot) := g(a_i, \cdot) \) be one-to-one a.s., that is, \( g(a, \cdot) \neq g(a^*, \cdot) \) a.s. if and only if \( h(a) \neq h(a^*) \). Then, for any distance measure between the two profiles \( g_i \) and \( g_j \), \( d(g_i, g_j) \), it follows that \( d(g_i, g_j) = 0 \) if and only if \( h(a_i) = h(a_j) \).

Based on this, Auerbach (2016) finds that one can control the network endogeneity by pair-wise differencing\(^2\) of the observations of the two individuals, \( i \) and \( j \), whose network formation distributions are the same, \( d(g_i, g_j) = 0 \). Based on this, Auerbach (2016) proposes a semiparametric estimator based on matching pairs of agents with similar columns of the squared adjacency matrix.

Notice that the identification condition of Auerbach (2016) is satisfied if \( g(a_i, \cdot) \) and \( a_i \) have a one-to-one relation. Compared to this, our second identification is based on the condition that \( a_i \) and the marginal network probability, \( \int g(a_i, \tau) d\tau \), have a one-to-one relation. We admit that this condition is more restrictive than the identification condition of Auerbach (2016), because our restriction is a special case his restriction. However, as mentioned in the introduction, our identification under the stronger condition allows for the omitted variable

\(^2\)This resembles Powell (1987), Heckman et al. (1998), and Abadie and Imbens (2006).
in the peer effects equation to be nonparametrically directly estimated, which results in the peer effect estimator having the parametric convergence rate $\sqrt{N}$. This feature is not necessarily guaranteed in the framework of Auerbach (2016)\textsuperscript{3}.

3. **General Model of Peer Effects with an Endogenous Network**

In this section, we introduce a general linear-in-means peer effect model that extends the simple illustrative outcome model with a peer effect in (2.1) and the simple dyadic network formation model in (2.2).

3.1. **General Linear-In-Means Peer Effects Model.** As in Section 2, $d_{ij,N}$ are the observed binary variables that measure undirected links among individuals $i \in \{1, 2, \ldots, N\}$. We assume that individual outcomes are given by the linear-in-means model of peer effects

$$y_i = \left(\sum_{j=1, j \neq i}^{N} g_{ij,N} y_j \right) \beta_{01}^0 + x_{1i}' \beta_{02}^0 + \left(\sum_{j=1, j \neq i}^{N} g_{ij,N} x_{1j} \right) ' \beta_{03}^0 + v_i,$$

(3.1)

where $x_{1i}$ are observed individual characteristics that affect the outcome $y_i$, $v_i$ are unobserved individual characteristics, and

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

is the weight of the peer effects. Using the terminology of Manski (1993), $\beta_{01}^0$ captures the endogenous social effect, and $\beta_{03}^0$ measures the exogenous social effect. We let $\beta^0 := (\beta_{01}^0, \beta_{02}^0, \beta_{03}^0)'$ and denote $\beta = (\beta_1, \beta_2', \beta_3')'$. We let $D_N$ 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 $G_N$ be the matrix whose $(i, j)^{th}$ element is $g_{ij,N}$, and $G_N$ is obtained by row-normalizing $D_N$. Denote $X_{1N} = (x_{11}', \ldots, x_{1N}')'$.

\textsuperscript{3}We thank one of the referees for suggesting the comparisons.
$y_N = (y_1, \ldots, y_N)'$ and $\upsilon_N = (\upsilon_1, \ldots, \upsilon_N)'$. Using this notation, we can express the linear-in-means peer effects model (3.1) as

$$y_N = G_N y_N \beta_1^0 + X_1 N \beta_2^0 + G_N X_1 N \beta_3^0 + \upsilon_N. \quad (3.2)$$

Throughout the paper, we assume that $|\beta_1^0| < 1$. It is known that when $G_N$ is row normalized (i.e., $\sum_{j \neq i} g_{ij,N} = 1$) and $|\beta_1^0| < 1$, the (equilibrium) solution of the peer effect model uniquely exists (e.g., see Lee (2004)) as

$$y_N = (I_N - \beta_1^0 G_N)^{-1} (X_1 N \beta_2^0 + G_N X_1 N \beta_3^0 + \upsilon_N) \quad (3.3)$$

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, $\mathbb{E}[\upsilon_i | X_1 N, G_N] = 0$. 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 $\upsilon_i$. For example, if $\upsilon_i \sim i.i.d. (0, \sigma^2)$, it is true that

$$\mathbb{E}[(G_N y_N)' \upsilon_N] = \mathbb{E}[(G_N (I_N - \beta_1^0 G_N)^{-1} (X_1 N \beta_2^0 + G_N X_1 N \beta_3^0 + \upsilon_N))' \upsilon_N] = \tau_0 tr(G_N (I_N - \beta_1^0 G_N)^{-1}) \neq 0. \quad (3.4)$$

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. In view of the expression of (3.3), when $\beta_1^0 \neq 0$, we can use $G_2^2 N X_1 N$ as the IV of the endogenous regressor $G_N y_N$ because $G_2^2 N X_1 N$ is uncorrelated with $\upsilon_N$ while it is correlated with the endogenous regressor $G_N y_N$ (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,

$$
\hat{\beta}^{2SLS}_N = \left( W_N^{'} Z_N (Z_N^{'} Z_N)^{-1} Z_N W_N \right)^{-1} W_N^{'} Z_N (Z_N^{'} Z_N)^{-1} Z_N^{'} y_N,
$$

(3.5)

where $W_N = [G_N y_N, X_{1N}, G_N X_{1N}]$ and $Z_N = [X_{1N}, G_N X_{1N}, G_N^2 X_{1N}]$ is the matrix of instruments. For the IVs $Z_N$ to be strong, we assume that $\beta_0^2 \neq 0$.

When the network matrix is endogenous, $E[G_N v_N] \neq 0$, 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 $Z_N = [X_{1N}, G_N X_{1N}, G_N^2 X_{1N}]$ is correlated with the error term $v_N$. Specifically, the validity of the 2SLS estimator depends on the orthogonality condition $E[v_N | Z_N] = 0$, which is implied if $E[v_N | X_{1N}, G_N] = 0$. However, it does not hold if the (row normalized) network $G_N$ is correlated with $v_N$, which is true if unobserved individual characteristics of $G_N$ directly influence both link formation and individual outcomes.

In this paper, we consider the case where it may be that $E[v_N | X_{1N}, G_N] \neq 0$, 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. Imagine kids from homes where parents help with homework only form friendships with kids from similar homes. If this unobserved characteristic of parental behavior is not taken into account, and if this is what really determines grades, this effect might falsely be classified as a peer effect. A more elaborate discussion of our framework and its empirical applications can be found in Section 2.

3.2. Model of Network Formation. Let $x_{2i}$ be a vector of observable characteristics of individual $i$, and let $x_i = x_{1i} \cup x_{2i}$. Define $X_{2N}$ analogously to $X_{1N}$ and let $X_N = X_{1N} \cup X_{2N}$.

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 $x_i$. We denote the vector of individual unobserved characteristics by $a_N = (a_1, a_2, \ldots, a_N)^\prime$. Individuals are connected by an undirected network $D_N$, with the $(i, j)^{th}$ element $d_{ij,N} = 1$ if $i$ and $j$ are directly

---

4 If $\beta_0^2 = 0$, $y_N$ does not depend on $X_{1N}$ and $G_N^2 X_{1N}$ is not a relevant instrument for $G_N y_N$.
connected and 0 otherwise. We assume the network to be undirected \( d_{ij,N} = d_{ji,N} \), and assume \( d_{ii,N} = 0 \) for all \( i \), following the convention. In this case, there are \( n = \binom{N}{2} \) dyads. Let \( t_{ij} \) denote an \( l_T \times 1 \) vector of dyad-specific characteristics of dyad \( ij \), and we assume that \( t_{ij} = t(x_{2i}, x_{2j}) \). Agents form links according to

\[
d_{ij,N} = \mathbb{I}(g(t(x_{2i}, x_{2j}), a_i, a_j) - u_{ij} \geq 0),
\]

where \( \mathbb{I}(\cdot) \) is an indicator function. In this setup, link surplus is transferable across directly linked agents and consists of three components: \( t_{ij} := t(x_{2i}, x_{2j}) \) 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 \( t_{ij} \) 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 (3.6) are used, (see for example Jackson (2005), Graham (2017)). An important example of a parametric specification is the one in Graham (2017),

\[
d_{ij,N} = \mathbb{I}(t(x_{2i}, x_{2j})'\lambda + a_i + a_j - u_{ij} > 0).
\]

For the purpose of the paper, particularly in constructing the estimators that we introduce in Section 5, we do not need a parametric specification.

The restrictions on the network formation (3.6) are made in Assumption 11 (iii) - (vi). There are two main implications implied by the restrictions in Assumption 11. First, Assumption 11 (vi) in the Appendix implies that the link formation probability of individual \( i \) with characteristics \((x_{2i}, a_i)\) is monotonic with respect to the unobserved characteristic \( a_i \), that is, for all \( x_{2i} \),

\[
a_i \neq a_i^* \text{ if and only if } \mathbb{P}(d_{ij,N} = 1 \mid x_{2i}, a_i) \neq \mathbb{P}(d_{ij,N} = 1 \mid x_{2i}, a_i^*).
\]

\(^5\)Our analysis can be extended to the directed network case, but we do not pursue it in this paper.
Obviously, this condition is satisfied in the parametric model (3.7). This monotonic condition justifies the use of the average node degree in implementing the control function in Section 5.2.

The second implication of the restrictions is that the network formed by (3.6) 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 \( g(t_{ij}, a_i, a_j) \) is bounded (see Assumption 11 (iii), (iv), and (v) in the Appendix). In this case, the probability of any 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. Notice that the dense network assumption and the share restriction on the net surplus function \( g \) are necessary for implementing the control function in Section 5 and establishing the asymptotic theory of the control function based estimators in Section 6. If \( a_i \) is observed, we can identify and estimate the peer effects without these assumptions (see Section 4).

Regarding the network formation model (3.6), it is important to notice that the network formation model (3.6) rules out interdependent link preferences, and it assumes that links are formed independently conditional on observed individual characteristics and unobserved fixed effects. 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 (3.6) 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 (2017) 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.
4. Identification of peer effects using a control function approach

In this section we provide an identification argument for the peer effect equation based on a control function when the network is endogenous.

4.1. 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 (3.6). First we introduce a basic assumption that we will maintain throughout the paper.

Assumption 1. (i) \((x_i, a_i, u_i)\) are i.i.d. for all \(i\), \(i = 1, \ldots, N\), (ii) \(\{u_{ij}\}_{i,j=1,\ldots,N}\) are independent of \((X_N, a_N, u_N)\) and i.i.d. across \((i, j)\) with cdf \(\Phi(\cdot)\), and (iii) \(E(u_i|x_i, a_i) = E(u_i|a_i)\).

Assumption (i) implies that the observables \(x_i\) and the unobservable characteristics, \((a_i, u_i)\), are randomly drawn. This is a standard assumption in the peer effects literature. Assumption (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 \(u_i\). This means that the unobserved error \(u_i\) 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 \(x_i\) of the outcome equation and the network formation to be correlated with the unobserved components \((u_i, a_i)\), so that the regressor \(x_{1i}\) can be endogenous in the outcome equation, and the network formation observables \(x_{2i}\) can be arbitrarily correlated with the unobserved individual characteristic \(a_i\). In Assumption (iii), we assume that the dependence between \(x_i\) and \(v_i\) exists only through \(a_i\). That is, \(a_i\) is the fixed effect of individual \(i\) and controls the endogeneity of \(x_i\) with respect to \(v_i\).

Notice that the network \(D_N\) defined in (3.6) and the (row normalized) network \(G_N\) are measurable functions of \((x_{2i}, x_{2,-i}, a_i, a_{-i}, \{u_{ij}\}_{i,j=1,\ldots,N})\), where \(x_{2,-i} = (x_{2,1}, \ldots, x_{2,i-1}, x_{2,i+1}, \ldots, x_{2,N})\).
and \( a_{-i} \) is defined analogously. Under Assumption 1 we have

\[
E[u_i|X_N, G_N, a_i] = E[u_i|x_{-i}, G_N(x_{2,-i}, a_{-i}, \{ u_{ij} \}_{i,j=1,...,N}, x_{2i}, a_i), x_i, a_i]
\]

\[
= E[u_i|x_i, a_i] = E[u_i|a_i],
\]

where the second equality holds because \((x_{-i}, a_{-i}, \{ u_{ij} \}_{i,j=1,...,N})\) and \((x_i, a_i, u_i)\) are independent under Assumptions 1 (i) and (ii). This shows \( u_i \) and \((x_{-i}, G_N(x_{2,-i}, a_{-i}, \{ u_{ij} \}_{i,j=1,...,N}, x_{2i}, a_i))\) are mean-independent conditioning on \((x_i, a_i)\). The last line follows by the fixed effect assumption, Assumption 1 (iii).

Result 4.1 shows that conditional on the unobserved heterogeneity \( a_i \) in the network formation (and any subcomponent of \( x_i \)), the unobserved characteristic \( u_i \) that affects the outcome \( y_i \) becomes uncorrelated with the (row normalized) network \( G_N \) (and the observables \( X_N \)). This implies that the network endogeneity can be controlled by \( a_i \) (or together with any subcomponent of \( x_i \)). We summarize the discussion above in the following lemma:

Lemma 1 (Control Function of Peer Group Endogeneity). Suppose that Assumption 1 holds. Then, \( E[u_i|X_N, G_N, a_i] = E[u_i|x_{s,i}, a_i] \).

4.2. Identification of Peer Effects with \( a_i \) as Control Function. In this section we show how to identify the peer effects in the outcome question when the endogenous network is formed by (3.6). We provide two identification methods depending on whether we control the network (peer group) endogeneity with \( a_i \) or \( a_i \) together with \( x_{2i} \), in the case when \( x_{2i} \) and \( x_{1i} \) do not overlap.

First notice that regardless of the possible endogeneity of the (row normalized) network \( G_N \), we need to control for the endogeneity of the term \( \sum_{j \neq i} g_{ij,N}y_j \) that represents the so-called endogenous peer effects. When the peer group \( G_N \) is exogenous and uncorrelated with \( u_N \), \( G_N^2 X_{1N} \) is often used as an IV for the endogenous peer effects term \( G_N y_N \) (See, for example, Kelejian and Prucha (1998), Lee (2003), Bramoullé et al. (2009)).

Let \( Z_N = [X_{1N}, G_N X_{1N}, G_N^2 X_{1N}] \) be the usual IV matrix used in 2SLS estimation of the peer effects equation. Note that \( Z_N \) is not a valid IV matrix anymore in our framework.
because the peer group defined by the network $G_N$ is correlated with $v_N$ due to potential correlation between the unobserved $v_i$ and $a_i$. Let $W_N = [G_N Y_N, X_{1N}, G_N X_{1N}]$. Further, denote the transpose of the $i$th row of $Z_N$ and $W_N$ by $z_i$ and $w_i$, respectively.

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

$$
\mathbb{E} [ (z_i - \mathbb{E}[z_i|a_i]) (v_i - \mathbb{E}(v_i|a_i)) | a_i ] = \mathbb{E}[z_i v_i | a_i] - \mathbb{E}[z_i | a_i] \mathbb{E}[v_i | a_i]
$$

$$
= \mathbb{E} [ \mathbb{E}[z_i | a_i, X_{1N}, G_N] | a_i ] - \mathbb{E}[z_i | a_i] \mathbb{E}[v_i | a_i]
$$

$$
= \mathbb{E} [ z_i \mathbb{E}[v_i | a_i, X_{1N}, G_N] | a_i ] - \mathbb{E}[z_i | a_i] \mathbb{E}[v_i | a_i]
$$

$$
\overset{(1)}{=} \mathbb{E} [ z_i \mathbb{E}[v_i | a_i] | a_i ] - \mathbb{E}[z_i | a_i] \mathbb{E}[v_i | a_i]
$$

$$
= 0,
$$

(4.1)

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

Furthermore, if $\mathbb{E} [ (z_i - \mathbb{E}[z_i|a_i]) (w_i - \mathbb{E}[w_i|a_i])' ]$ has full rank, then we can identify the peer effect coefficients $\beta^0$ as

$$
0 = \mathbb{E} [ (z_i - \mathbb{E}[z_i|a_i]) (y_i - w_i' \beta - \mathbb{E}[y_i - w_i' \beta|a_i]) ]
$$

$$
= \mathbb{E} [ (z_i - \mathbb{E}[z_i|a_i]) (w_i - \mathbb{E}[w_i|a_i])' ] (\beta - \beta^0) + \mathbb{E} [ (z_i - \mathbb{E}[z_i|a_i]) (v_i - \mathbb{E}[v_i|a_i]) ]
$$

$$
\overset{(1)}{=} \mathbb{E} [ (z_i - \mathbb{E}[z_i|a_i]) (w_i - \mathbb{E}[w_i|a_i])' ] (\beta - \beta^0)
$$

$$
\overset{(2)}{=} \beta = \beta^0,
$$

where equality (1) follows by the orthogonality result in (4.1) and equality (2) follows from the full rank condition.

**Assumption 2** (Rank condition). $\mathbb{E} [ (z_i - \mathbb{E}[z_i|a_i]) (w_i - \mathbb{E}[w_i|a_i])' ]$ has full rank.

For the full rank condition in Assumption 2 it is necessary that the IVs $z_i$ and the regressors $w_i$ 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 $z_i$ and $w_i$ become close
to functions that depend only on \((x_i, a_i)\). In this case, for the full rank condition to be satisfied, it is necessary that there be additional random components in \(x_i\) that are different from \(a_i\), so that the limits of \(z_i\) and \(w_i\) are not linearly dependent. As a summary, we have the following first identification theorem.

**Theorem 4.1** (Identification). Under Assumptions 1 and 2, the parameter \(\beta^0\) is identified by the moment condition \(E[(z_i - E(z_i|a_i))(y_i - E(y_i|a_i) - (w_i - E(w_i|a_i))' \beta^0)] = 0\):

\[
E[(z_i - E(z_i|a_i))(y_i - E(y_i|a_i) - (w_i - E(w_i|a_i))' \beta)] = 0 \iff \beta = \beta^0.
\]

Theorem 4.1 shows that we can identify the parameter \(\beta^0\) by controlling the unobserved network heterogeneity \(a_i\) in the outcome equation and taking the residuals \(y_i - E(y_i|a_i) - (w_i - E(w_i|a_i))' \beta\) and using the instrumental variables \(z_i - E[z_i|a_i]\).

**4.3. Identification of Peer Effects using \((x_{2i}, a_i)\) as Control Function.** In view of the derivation of the control function in (4.1) under Assumption 1, it is possible to use any regressors in \(x_i\) in addition to the unobserved heterogeneity \(a_i\). In this section, we discuss identification of the peer effects using \((x_{2i}, a_i)\) as control function. The reason to consider this particular control function is that we can implement it in the absence of a consistent estimator of \(a_i\), which will be discussed in detail in Section 5.

First, suppose that there is no overlap between the regressors in the outcome equation, \(x_{1i}\), and the regressors in the network formation equation, \(x_{2i}\), and assume the conditions in Assumption 1\(^6\).

**Assumption 3.** Assume that the conditions (i),(ii), and (iii) of Assumption 1 hold. Also, assume that (iv) the explanatory variables in \(x_{1i}\) and \(x_{2i}\) do not overlap (i.e., \(x_{1i} \cap x_{2i} = \emptyset\)).

\(^6\)Later in this section, we will discuss a more general case where \(x_{1i}\) and \(x_{2i}\) intersect.
Then, under Assumption 1 and (4.1), it follows that
\[
E[\nu_i | X_N, G_N, a_i] = E[\nu_i | a_i] = E[\nu_i | x_{2i}, a_i],
\]
(4.2)
where the last line holds by Assumption 1(iii). Then, similar to (4.1), we can show that
\[
E \left[ (z_i - E[z_i | x_{2i}, a_i]) (v_i - E(v_i | x_{2i}, a_i)) | x_{2i}, a_i \right] = 0.
\]
(4.3)
Furthermore, suppose that the following full rank assumption is satisfied:

**Assumption 4** (Rank condition). \( E \left[ \left( z_i - E[z_i | x_{2i}, a_i] \right) \left( w_i - E[w_i | x_{2i}, a_i] \right)' \right] \) has full rank.

Notice that if \( x_{1i} \) and \( x_{2i} \) intersect, then the full rank condition in Assumption 4 does not hold.

Using similar arguments that lead to Theorem 4.1, we can identify the peer effect coefficients \( \beta^0 \) as
\[
0 = E \left[ (z_i - E[z_i | x_{2i}, a_i]) \left( y_i - w_i' \beta - E[y_i - w_i' \beta | x_{2i}, a_i] \right) \right] \iff \beta = \beta^0,
\]
(4.4)
This is summarized in the following theorem.

**Theorem 4.2** (Alternative Identification). Under Assumptions 1, 3, and 4, the parameter \( \beta^0 \) is identified by the moment condition
\[
E[(z_i - E(z_i | x_{2i}, a_i)) \left( (y_i - E(y_i | x_{2i}, a_i) - (w_i' - E(w_i | x_{2i}, a_i))' \beta \right)] = 0 \iff \beta = \beta^0.
\]
So far, we have considered the case where the regressors \( x_{1i} \) and \( x_{2i} \) do not intersect. A more general case is when the regressors \( x_{1i} \) consist of two components, where one component is different from the observed control function \( x_{2i} \) and the other is part of \( x_{2i} \). That is, 
\( x_{1i} = (x_{11i}, x_{12i}) \), where \( x_{11i} \) does not share any elements with \( x_{2i} \) and \( x_{11i} \) is nonempty, and \( x_{12i} \subset x_{2i} \). Let \( \beta_2^0 = (\beta_{21}^0, \beta_{22}^0) \), \( \beta_3^0 = (\beta_{31}^0, \beta_{32}^0) \) conformable to the dimensions of \((x_{11i}, x_{12i})\). Similarly let \( \beta_2 = (\beta_{21}, \beta_{22}), \beta_3 = (\beta_{31}, \beta_{32}) \).
In this case, with a properly modified rank condition of $z_{(2),i}$ and $w_{(2),i}$ that excludes the variables associated with $x_{12,i}$ and $\sum_{j=1,\neq i}^{N}g_{ij,N}x_{12,j}$, we can identify the coefficients $\beta^{0}_{(2)} := (\beta^{0}_1, \beta^{0}_{21}, \beta^{0}_{31})$ using the same argument that leads to the identification in (4.4). However, we cannot identify the coefficients that correspond to the variable $x_{12,i}$ and $\sum_{j=1,\neq i}^{N}g_{ij,N}x_{12,j}$. The reason is that controlling the network endogeneity with the control variable $(x_{2i}, a_{i})$ wipes out the information in $(x_{12,i}, \sum_{j=1,\neq i}^{N}g_{ij,N}x_{12,j})$:

$$x_{12,i} - E[x_{12,i}|x_{2i}, a_{i}] = 0$$

$$\sum_{j=1,\neq i}^{N}g_{ij,N}x_{12,j} - E\left[\sum_{j=1,\neq i}^{N}g_{ij,N}x_{12,j} | x_{2i}, a_{i}\right] \rightarrow_p 0,$$

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

Throughout the rest of the paper, when we consider $(x_{2i}, a_{i})$ as control function, we will without loss of generality apply the restriction in Assumption 3 that $x_{1i}$ and $x_{2i}$ do not overlap.

5. Estimation

In this section we present two estimation methods. In subsections 5.1 and 5.2 we discuss estimation using $a_{i}$ and $(x_{2i}, a_{i})$ as control functions, respectively.

5.1. With $a_{i}$ as Control Function. The identification scheme of Theorem 4.1 identifies the parameter of interest $\beta^0$ with the two step procedure: (i) control $a_{i}$ in the outcome equation and yield $y_{i} - E(y_{i}|a_{i}) = (w_{i} - E(w_{i}|a_{i}))'\beta_{0} + v_{i} - E(v_{i})$, and then (ii) use $z_{i} - E(z_{i}|a_{i})$ as IVs for $w_{i} - E(w_{i}|a_{i})$. If we observe $a_{i}$ and know the conditional mean functions $h(a_{i}) = (h_y(a_{i}), h_w(a_{i}), h_z(a_{i})) := (E[y_{i}|a_{i}], E[w_{i}|a_{i}], E[z_{i}|a_{i}])$, then $\beta^0$ can be estimated using 2SLS
In this paper we consider a (linear) sieve estimation method. Estimation of \( h \) \( \hat{h} \).

See Section A.2.1 in the Appendix for more details on the estimator \( h \).

However, since the individual heterogeneity \( a_i \) is not observed and the conditional mean functions \( h(a_i) = \langle E(y_i|a_i), E(w_i|a_i), E(z_i|a_i) \rangle \) are not known either, the estimator \( \hat{h}_{inf}^{2SLS} \) is not feasible.

A natural implementation of the infeasible estimator \( \hat{h}_{inf}^{2SLS} \) is to replace the conditional mean function \( h(a_i) \) with its estimate. Suppose that \( \hat{a}_i \) is an estimator of \( a_i \) and \( \hat{h}(\hat{a}_i) \) is a nonparametric estimator of \( h(a_i) \). Then we can implement the infeasible estimator \( \hat{h}_{inf}^{2SLS} \) with

\[
\hat{h}_{2SLS}^i := \left[ \sum_{i=1}^{N} (w_i - \hat{h}^w(\hat{a}_i))(z_i - \hat{h}^z(\hat{a}_i))' \left( \sum_{i=1}^{N} (z_i - \hat{h}^z(\hat{a}_i))(z_i - \hat{h}^z(\hat{a}_i))' \right)^{-1} \sum_{i=1}^{N} (z_i - \hat{h}^z(\hat{a}_i))(w_i - \hat{h}^w(\hat{a}_i))' \right]^{-1}
\]

\[
\times \left[ \sum_{i=1}^{N} (w_i - \hat{h}^w(\hat{a}_i))(z_i - \hat{h}^z(\hat{a}_i))' \left( \sum_{i=1}^{N} (z_i - \hat{h}^z(\hat{a}_i))(z_i - \hat{h}^z(\hat{a}_i))' \right)^{-1} \sum_{i=1}^{N} (z_i - \hat{h}^z(\hat{a}_i))(y_i - \hat{h}^y(\hat{a}_i))' \right].
\]

(5.3)

See Section A.2.1 in the Appendix for more details on the estimator \( \hat{h}_{2SLS} \).

**Estimation of \( h(\cdot) \):** We can estimate \( h(\cdot) \) using various standard nonparametric methods. In this paper we consider a (linear) sieve estimation method.\(^7\) Suppose that \( h'(a) \) is the \( l^{th} \) element in \( h(a) \) for \( l = 1, \ldots, L \), where \( L \) is the dimension of \( (y_i, w_i', z_i')' \). The sieve

\(^7\)In principle we can use other nonparametric estimation methods such as kernel smoothing or local polynomial methods.
estimation method assumes that each function $h^l(a), l = 1, ..., L$ is well approximated by a linear combination of base functions $(q_1(a), ..., q_{K_N}(a))$: \[
 h^l(a) \approx \sum_{k=1}^{K_N} q_k(a)\alpha_k^l, \tag{5.4}
\]
as the truncation parameter $K_N \to \infty$. A linear sieve (or series) estimator of a function, for example $\hat{h}^y(\hat{a}_i)$, is the OLS projection of $y_i$ on the sieve basis $q^K(\cdot) = (q_1(\cdot), ..., q_K(\cdot))'$ with $\hat{a}_i$ plugged in,

$\hat{h}^y(\hat{a}_i) := q^K(\hat{a}_i)' \left( \sum_{i=1}^{N} q^K(\hat{a}_i)q^K(\hat{a}_i)' \right)^{-1} \sum_{i=1}^{N} q^K(\hat{a}_i)y_i.$

For the regularity conditions of the sieve basis $q^K(a_i)$, we impose standard conditions such as those proposed by Newey (1997) and Li and Racine (2007). These assumptions ensure that $\sum_{i=1}^{N} q^K(a_i)q^K(a_i)'$ 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 $\hat{\beta}_{2SLS}$ (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

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

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. 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.
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.

**Estimation of $a_i$:** 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 $\max_i |\hat{a}_i - a_i| = O_p (\zeta_a(N)^{-1})$, where $\zeta_a(N) \to \infty$ as $N \to \infty$, satisfying Assumption 8 in the Appendix.

Here $\zeta_a(N)$ is the order of magnitude that measures the Lipschitz smoothness of the sieve basis. The assumption puts restrictions on the uniform bound of the convergence rate of $\hat{a}_i$, and we need a more accurate estimator of $a_i$ when the average curvature of the sieve basis is larger.

For the purpose of our paper, any estimation method that yields an estimator $\hat{a}_i$ satisfying the restriction in Assumption 5 can be adopted. For example, assuming the parametric specification as in (3.7),

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

with regularity conditions of Assumption 6 in the Appendix, including the error $u_{ij}$ following a logistic distribution, Graham (2017) showed that the joint maximum likelihood estimator that solves

$$(\hat{a}_1, ..., \hat{a}_N)$$

satisfies

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

(5.6)
with probability $1 - O(N^{-2})$. In this case we have $\zeta(N) = \sqrt{\frac{N}{\ln N}}$. Notice that the requirement that the network formation in (5.5) be dense is necessary for $\hat{a}_i$ to satisfy the desired uniform convergence rate in (5.6). Examples of other estimation methods include Fernández-Val and Weidner (2013), Jochmans (2016), Dzemski (2017), and Jochmans (2017).

5.2. With $(x_{2i}, a_i)$ as Control Function. As we assume in Section 4.3, we consider the case where $x_{1i}$ and $x_{2i}$ do not overlap. When $a_i$ is observed and the conditional expectations $h^*(x_{2i}, a_i) := (E(y_i|x_{2i}, a_i), E(w_i|a_i), E(z_i|a_i))$ are known, we can estimate $\beta^0$ by the 2SLS similar to $\hat{\beta}^\inf_{2SLS}$ in (5.1),

$$\hat{\beta}^\inf_{2SLS} = \left[ \sum_{i=1}^{N} (w_i - h^w_*(x_{2i}, a_i))(z_i - h^z_*(x_{2i}, a_i))' \left( \sum_{i=1}^{N} (z_i - h^z_*(x_{2i}, a_i))(z_i - h^z_*(x_{2i}, a_i))' \right)^{-1} \times \sum_{i=1}^{N} (z_i - h^z_*(x_{2i}, a_i))(w_i - h^w_*(x_{2i}, a_i))' \right]^{-1} \times \left[ \sum_{i=1}^{N} (w_i - h^w_*(x_{2i}, a_i))(z_i - h^z_*(x_{2i}, a_i))' \left( \sum_{i=1}^{N} (z_i - h^z_*(x_{2i}, a_i))(z_i - h^z_*(x_{2i}, a_i))' \right)^{-1} \times \sum_{i=1}^{N} (z_i - h^z_*(x_{2i}, a_i))(y_i - h^y_*(x_{2i}, a_i))' \right]^{-1}. \quad (5.7)$$

When $a_i$ is unknown and $x_{2i}$ is also used in the control function, under the monotonicity condition of the link formation as in (3.8), we can implement the infeasible estimator using the average node degree without estimating $a_i$. To be more specific, first we denote

$$\mathbb{P}(d_{ij,N} = 1|x_{2i}, a_i) =: \deg(x_{2i}, a_i) =: \deg_i.$$

Under the monotonicity condition in (3.8), $(x_{2i}, a_i)$ and $(x_{2i}, \deg_i)$ are one-to-one. This implies that for any $b_i \in \{y_i, w_i, z_i\}$,

$$h^b_*(x_{2i}, a_i) = \mathbb{E}(b_i|x_{2i}, a_i) = \mathbb{E}(b_i|x_{2i}, \deg_i) =: h^b_{**}(x_{2i}, \deg_i).$$
Notice that the natural estimator of \( \deg_i \) is the node degree of \( i \), the number of connections with node (individual) \( i \) in the network scaled by the network size:

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

Recall that the link \( d_{ij,N} \) is formed by

\[
d_{ij,N} = \mathbb{I}(g(t(x_{2i}, x_{2j}), a_i, a_j) - u_{ij} \geq 0).
\]

Also 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 \( \pi(x_2, a) \) be the joint density function of \( (x_{2i}, a_i) \). Then, for each \( (x_{2i}, a_i) \), by the WLLN conditioning on \( (x_{2i}, a_i) \), we have

\[
\hat{\deg}_i := \frac{1}{N - 1} \sum_{j=1,\neq i}^N \mathbb{I}(g(t(x_{2i}, x_{2j}), a_i, a_j) - u_{ij} \geq 0) \rightarrow_p \int \Phi(g(t(x_{2i}, x_{2j}), a_i, a_j)) \pi(x_2, a) d\mathbf{x}_2 da
\]

\[
= \mathbb{P}(d_{ij,N} = 1|\mathbf{x}_{2i}, a_i)
\]

\[=: \deg_i > 0 \quad (5.8)\]

as the network size \( N \) grows to infinity. Here the limit of the average network \( \deg_i > 0 \) follows since we assume the network is dense.

This shows that \( \hat{\deg}_i \) can be used as an estimator of \( \deg_i \). In fact, we can show that under the regularity conditions in Assumption \([11]\) in the Appendix, \( \sup_i \mathbb{E}[(\sqrt{N}(\hat{\deg}_i - \deg_i))^{2B}] < \infty \) for any finite integer \( B \geq 2 \), from which we can deduce that

\[
\max_{1 \leq i \leq N} |\hat{\deg}_i - \deg_i| = O_p \left( \zeta_{\deg}(N)^{-1} \right), \quad (5.9)
\]

where

\[
\zeta_{\deg}(N) := o(1)N^{B-1 \over 2B}.
\]

This corresponds to the regularity condition in Assumption \([5]\).
Suppose that $r^K(x_{2i}, \deg_{i}) = (r_1(x_{2i}, \deg_{i}), \ldots, r_K(x_{2i}, \deg_{i}))'$ is a sieve basis of the unknown function $h_*(x_{2i}, a_i)$. For each $b_i \in \{y_i, w_i, z_i\}$, a sieve estimator of $h_{*n}^b(x_{2i}, \deg_{i}) = \mathbb{E}(b_i|x_{2i}, a_i)$ is the OLS projection of $b_i$ on $r^K(x_{2i}, \deg_{i})$. For example,

$$
\hat{h}_{*n}^y(x_{2i}, a_i) = \hat{h}_{*n}^y(x_{2i}, \deg_{i}) = \mathbf{r}^K(x_{2i}, \deg_{i})' \left( \sum_{i=1}^N \mathbf{r}^K(x_{2i}, \deg_{i}) \mathbf{r}^K(x_{2i}, \deg_{i})' \right)^{-1} \sum_{i=1}^N \mathbf{r}^K(x_{2i}, \deg_{i})y_i.
$$

Then, we have

$$
\bar{\beta}_{2SLS} = \left[ \sum_{i=1}^N (w_i - \hat{h}_{*n}^w(x_{2i}, a_i))(z_i - \hat{h}_{*n}^z(x_{2i}, a_i))' \left( \sum_{i=1}^N (z_i - \hat{h}_{*n}^z(x_{2i}, a_i))(z_i - \hat{h}_{*n}^z(x_{2i}, a_i))' \right)^{-1} \sum_{i=1}^N (z_i - \hat{h}_{*n}^z(x_{2i}, a_i)) (w_i - \hat{h}_{*n}^w(x_{2i}, a_i))' \right]^{-1}
\times \left[ \sum_{i=1}^N (w_i - \hat{h}_{*n}^w(x_{2i}, a_i))(z_i - \hat{h}_{*n}^z(x_{2i}, a_i))' \left( \sum_{i=1}^N (z_i - \hat{h}_{*n}^z(x_{2i}, a_i))(z_i - \hat{h}_{*n}^z(x_{2i}, a_i))' \right)^{-1} \sum_{i=1}^N (z_i - \hat{h}_{*n}^z(x_{2i}, a_i)) (y_i - \hat{h}_{*n}^y(x_{2i}, a_i))' \right]^{-1}.
$$

(5.10)

For more details see Section A.2.2 in the Appendix.

The two different estimators $\bar{\beta}_{2SLS}$ and $\bar{\beta}_{2SLS}$ are implemented using different control functions, and these two approaches have their own pros and cons. For $\bar{\beta}_{2SLS}$, a good estimator of $a_i$ is required, which imposes restrictions on the network formation model (3.6) in the form of (3.7). Compared to this, the estimator $\bar{\beta}_{2SLS}$ that uses $(x_{2i}, \deg_{i})$ as control functions does not require a restriction like (3.7). It requires only the monotonicity of the net surplus function as in (3.8) of Section 3.2. However, there is a disadvantage of $\bar{\beta}_{2SLS}$. Because it uses $x_{2i}$ as a part of the control function, as discussed in Section 4.3, this approach cannot identify and estimate the coefficients of the regressor $x_{2i}$ if $x_{2i}$ is a relevant regressor of the outcome.
Later in Section 7 of Monte Carlo simulations, we compare the finite sample properties of \( \hat{\beta}_{2SLS} \) and \( \bar{\beta}_{2SLS} \) in both dense and sparse setups. There we observe that the finite sample properties of \( \bar{\beta}_{2SLS} \) are better than \( \hat{\beta}_{2SLS} \) particularly in the sparse setup. The reason for this is that the implementation of the control function with the node degree is less sensitive with respect to the denseness of the network formation than with \( \hat{a}_i \).

6. 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.

6.1. Limiting Distribution and Standard Error of \( \hat{\beta}_{2SLS} \). Recall the definitions \( h^y(a_i) := \mathbb{E}[y_i|a_i], \ h^v(a_i) := \mathbb{E}[v_i|a_i], \ h^w(a_i) := \mathbb{E}(w_i|a_i), \ h^z(a_i) := \mathbb{E}(z_i|a_i) \). Define \( \eta^y_i := y_i - h^y(a_i), \ \eta^v_i := v_i - h^v(a_i), \ \eta^w_i := w_i - h^w(a_i), \ \eta^z_i := z_i - h^z(a_i) \). Let \( \eta^v_N = (\eta^v_1, ..., \eta^v_N)' \) and \( \eta^w_N = (\eta^w_1, ..., \eta^w_N)' \). Let \( \eta^z_N = (\eta^z_1, ..., \eta^z_N)' \).

In the appendix, we derive the asymptotic distribution of \( \hat{\beta}_{2SLS} \) in three steps. First, we show that the sampling error caused by the use of \( \hat{a}_i \) instead of \( a_i \) 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 \( l \in \{v, w, z\} \). In Lemma 7 of the 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

\[
\sqrt{N}(\hat{\beta}_{2SLS} - \beta_{2SLS}) = o_p(1).
\]
The last step is to derive the limiting distribution of the infeasible estimator $\sqrt{N}(\hat{\beta}_{2SLS} - \beta^0)$. In the Supplementary Appendix S.1.3 we show the following:

$$\frac{1}{N} \sum_{i=1}^{N} (w_i - h^w(a_i))(z_i - h^z(a_i))' \xrightarrow{P} S_{wz}$$

$$\frac{1}{N} \sum_{i=1}^{N} (z_i - h^z(a_i))(z_i - h^z(a_i))' \xrightarrow{P} S_{zz}$$

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - h^z(a_i))\eta_i \Rightarrow N(0, S_{zz})$$

where the closed forms of the limits $S_{wz}$ and $S_{zz}$ are found in Lemma 11 and $S_{zz}$ in Lemma 12 of Supplementary Appendix.

Notice that the derivation of the limiting distribution in (6.3) allows $\eta_i = \eta_i - E(\eta_i\mid a_i)$ to be conditionally heteroskedastic, and so $\sigma^2(x_i, a_i) := E[(v_i - E[v_i\mid a_i])^2\mid x_i, a_i]$ is allowed to depend on $(x_i, a_i)$.

Combining all the limit results leads to the following theorem.

**Theorem 6.1 (Limiting Distribution).** Suppose that Assumptions 1, 2, 5, 7, 8, and 11(i)-(v) in the Appendix hold. Then, we have

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

where

$$\Omega = (S_{wz} (S_{zz})^{-1} S_{wz})'^{-1} (S_{wz} (S_{zz})^{-1} S_{zz})^{-1} (S_{wz})' \left(S_{wz} (S_{zz})^{-1} (S_{wz})'\right)^{-1}.\quad (6.4)$$

The theorem requires several regularity conditions. These conditions are presented in Appendix A.1. Assumption 1 requires that $(y_i, x_i, a_i)$ 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 $h(a_i)$ and $\hat{h}(\hat{a}_i)$ converges to zero at a suitable rate. Assumption 11 imposes further restrictions on the outcome model (3.1) and the network formation model (3.6). It requires that $|\beta_i|$ 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 the observables \((y_i, x_i)\) and \(t_{ij}\) 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 (6.1), (6.2), and (6.3), which involves some uniformity in the limit.

The asymptotic variance can be consistently estimated by

\[
\hat{\Omega} = \left( \hat{S}_{wz} \left( \hat{S}_{zz} \right)^{-1} (\hat{S}_{wz})' \right)^{-1} \left( \hat{S}_{wz} \left( \hat{S}_{zz} \right)^{-1} \hat{S}_{zz2} \left( \hat{S}_{zz} \right)^{-1} (\hat{S}_{wz})' \right) \left( \hat{S}_{wz} \left( \hat{S}_{zz} \right)^{-1} (\hat{S}_{wz})' \right)^{-1},
\]

where

\[
\hat{S}_{wz} = \frac{1}{N} \sum_{i=1}^{N} \left( w_i - \hat{h}_w(\hat{a}_i) \right) \left( z_i - \hat{h}_z(\hat{a}_i) \right)'
\]

\[
\hat{S}_{zz} = \frac{1}{N} \sum_{i=1}^{N} \left( z_i - \hat{h}_z(\hat{a}_i) \right) \left( z_i - \hat{h}_z(\hat{a}_i) \right)'
\]

\[
\hat{S}_{zz2} = \frac{1}{N} \sum_{i=1}^{N} \left( z_i - \hat{h}_z(\hat{a}_i) \right) \left( z_i - \hat{h}_z(\hat{a}_i) \right)' \left( \hat{\eta}_i' \right)^2,
\]

and \( \hat{\eta}_i = y_i - \hat{h}_y(\hat{a}_i) - (w_i - \hat{h}_w(\hat{a}_i))' \hat{\beta}_{2SLS} \).

6.2. Limiting Distribution and Standard Error of \( \hat{\beta}_{2SLS} \). The process is analogous to the one presented in the previous section. Again, let \( b_i^l \) be the \( l^{th} \) element in \((y_i, w_i', z_i')'\).

Recall the definition that

\[
h_i^l(x_{2i}; a_i) = \mathbb{E}[b_i| x_{2i}, a_i] = \mathbb{E}[b_i| x_{2i}, \deg_i] =: h_{i*}(x_{2i}, \deg_i).
\]

Further, let \( \eta_{si} = b_i^l - h_i^l(x_{2i}, a_i) = b_i^l - h_i^l(x_{2i}, \deg_i) \), and let \( \hat{h}_i^l(x_{2i}, \deg_i) \) denote a sieve estimator of \( h_i^l(x_{2i}, \deg_i) \).

As in the previous section, we derive the asymptotic distribution of \( \hat{\beta}_{2SLS} \) in three steps. First, we show that the error that stems from the use of the estimate \( \hat{\deg_i} \) for \( \deg_i \), \( \hat{h}_i^l(x_{2i}, \hat{\deg}_i) - \hat{h}_i^l(x_{2i}, \deg_i) \), is asymptotically negligible. In the second step, we control the error introduced by the non-parametric estimation of \( h_i^l(x_{2i}, \deg_i) \), \( \hat{h}_i^l(x_{2i}, \deg_i) - h_i^l(x_{2i}, \deg_i) \). This
implies
\[
\sqrt{N}(\hat{\beta}_{2SLS} - \beta^0) = o_p(1),
\]
where
\[
\hat{\beta}^{infty}_{2SLS} = \left( \tilde{W}_{*N}' \tilde{Z}_{*N} \left( \tilde{Z}'_{*N} \tilde{Z}_{*N} \right)^{-1} \tilde{Z}'_{*N} \tilde{W}_{*N} \right)^{-1} \left( \tilde{W}_{*N}' \tilde{Z}_{*N} \left( \tilde{Z}'_{*N} \tilde{Z}_{*N} \right)^{-1} \tilde{Z}'_{*N} \tilde{y}_{*N} \right)
\]
and \( \tilde{W}_{*N} = (w_1 - h_w(x_{21}, a_1), ..., w_N - h_w(x_{2N}, a_N))' \) and \( \tilde{Z}_{*N}, \tilde{y}_{*N} \) are defined analogously.

The last step is to derive the limiting distribution of the infeasible estimator \( \sqrt{N}(\hat{\beta}^{infty}_{2SLS} - \beta^0) \) by showing
\[
\frac{1}{N} \sum_{i=1}^{N} (w_i - h_w(x_{2i}, a_i))(z_i - h_z(x_{2i}, a_i))' \xrightarrow{p} \bar{S}_{wz}
\]
\[
\frac{1}{N} \sum_{i=1}^{N} (z_i - h_z(x_{2i}, a_i))(z_i - h_z(x_{2i}, a_i))' \xrightarrow{p} \bar{S}_{zz}
\]
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - h_z(x_{2i}, a_i))\eta_{zi}^{w} \Rightarrow \mathcal{N}(0, \bar{S}_{zz} \sigma^{2})
\]

Combining all the limit results we have the following theorem.

**Theorem 6.2 (Limiting Distribution).** Suppose that Assumptions 1, 3, 4, 9, 10, and 11 hold. Then, we have
\[
\sqrt{N}(\hat{\beta}_{2SLS} - \beta^0) \Rightarrow \mathcal{N}(0, \bar{\Omega}),
\]
where
\[
\bar{\Omega} = \left( \bar{S}_{wz} \left( \bar{S}_{zz} \right)^{-1} \left( \bar{S}_{wz} \right)' \right)^{-1} \left( \bar{S}_{wz} \left( \bar{S}_{zz} \right)^{-1} \bar{S}_{zz} \sigma \left( \bar{S}_{zz} \right)^{-1} \left( \bar{S}_{wz} \right)' \right) \left( \bar{S}_{wz} \left( \bar{S}_{zz} \right)^{-1} \left( \bar{S}_{wz} \right)' \right)^{-1}.
\]

The asymptotic result in Theorem 6.2 requires the following regularity conditions which are formally presented in the appendix. First, Assumption 3 assumes that the regressors in the outcome equation, \( x_{1i} \) and the observables in the network formation \( x_{2i} \) do not overlap. Assumption 4 is a full rank condition for \( \hat{\beta}_{2SLS} \). Assumptions 9 and 10 assume the sieve used in constructing the estimator \( \hat{\beta}_{2SLS} \). Comparing the assumptions assumed in Theorem
Theorem 6.2 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 function in the link formation is strictly monotonic in $a_i$ conditional on $(x_{2i}, x_{2j}, a_j)$, which implies the required monotonicity condition in (3.8).

Like in the case of $\hat{\beta}_{2SLS}$, we allow $\eta_{\epsilon_i} = v_i - \mathbb{E}(v_i | x_{2i}, a_i)$ to be conditionally heteroskedastic, and $\sigma^2(x_i, a_i) := \mathbb{E}[(v_i - \mathbb{E}(v_i | x_{2i}, a_i))^2 | x_i, a_i]$ is allowed to depend on $(x_i, a_i)$.

The asymptotic variance can be consistently estimated by

$$\hat{\Omega} = \left( \hat{\bar{S}}^{wz} \left( \hat{\bar{S}}^{-1} \hat{\bar{S}}^{wz} \right)' \right)^{-1} \left( \hat{\bar{S}}^{wz} \left( \hat{\bar{S}}^{-1} \hat{\bar{S}}^{zz} \hat{\bar{S}}^{-1} \hat{\bar{S}}^{wz} \right)' \right)^{-1} \left( \hat{\bar{S}}^{wz} \left( \hat{\bar{S}}^{-1} \hat{\bar{S}}^{wz} \right)' \right)^{-1} \left( \hat{\bar{S}}^{wz} \left( \hat{\bar{S}} - \hat{\bar{y}}_{\beta_{2SLS}} \right)' \right),$$

(6.6)

where

$$\hat{\bar{S}}^{wz} = \frac{1}{N} \sum_{i=1}^{N} (w_i - \hat{h}_{w}^w(x_{2i}, \hat{\text{deg}}_i)) \left( z_i - \hat{h}_{zz}^z(x_{2i}, \hat{\text{deg}}_i) \right)'$$

$$\hat{\bar{S}}^{zz} = \frac{1}{N} \sum_{i=1}^{N} (z_i - \hat{h}_{zz}^z(x_{2i}, \hat{\text{deg}}_i)) \left( z_i - \hat{h}_{zz}^z(x_{2i}, \hat{\text{deg}}_i) \right)'$$

$$\hat{\bar{S}}^{zz^2} = \frac{1}{N} \sum_{i=1}^{N} (z_i - \hat{h}_{zz}^z(x_{2i}, \hat{\text{deg}}_i)) \left( z_i - \hat{h}_{zz}^z(x_{2i}, \hat{\text{deg}}_i) \right)' (\hat{\eta}_{\epsilon_i})^2,$$

and $\hat{\eta}_{\epsilon_i} = y_i - \hat{h}_{zz}^z(x_{2i}, \hat{\text{deg}}_i) - (w_i - \hat{h}_{w}^w(x_{2i}, \hat{\text{deg}}_i))'\hat{\beta}_{2SLS}$.

7. Monte Carlo

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

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

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
$a_i = \varphi(\alpha_L \mathbb{1}\{x_{2i} = -1\} + \alpha_H \mathbb{1}\{x_{2i} = 1\} + \xi_i),$ \\
with $\alpha_L \leq \alpha_H$ and $\xi_i$ a centered Beta random variable $\xi_i | x_{2i} \sim \{Beta(\mu_0, \mu_1) - \frac{\mu_0}{\mu_0 + \mu_1}\}$ so that $a_i \in (\varphi \left[ \alpha_L - \frac{\mu_0}{\mu_0 + \mu_1}, \alpha_H + \frac{\mu_1}{\mu_0 + \mu_1} \right]),$ $\varphi$ 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

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

In the simulations, we set $\beta_1 = 0.8,$ $\beta_2 = \beta_3 = 5,$ $x_{1i} = 3q_1 + \cos(q_2)/0.8 + \varepsilon_i,$ where $q_1, q_2 \sim \mathcal{N}(x_{2i}, 1),$ and $\varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, 1).$ For $h(a_i)$ we use the following functional forms: $h(a_i) = \exp(\kappa a_i),$ $h(a_i) = \cos(\kappa a_i)$ $h(a_i) = \sin(\kappa a_i)$ with $\kappa = 3.$ A plot of $h(a_i)$ for these functional forms is presented in Figure [1]. We can see that the exponential function gives a strongly increasing impact on the individual outcome, with the cosine functions the returns are increasing up to a certain point and then decreasing, whereas the sine function gives a more irregular pattern.

We estimate the outcome equation coefficients ($\beta_1, \beta_2, \beta_3$) using the standard 2SLS estimator for peer effects and the Hermite polynomial sieve. We consider both a dense and a sparse network case. For the dense network case, we estimate $a_i$ using $\widehat{a}_i$ and implement the

---

9We have also performed simulations using the polynomial sieve, the results are very similar and available upon request.
following control functions: using a control function linear in $\hat{a}_i$, $\hat{h}(a_i)$, $\hat{h}(deg_i, x_{2i})$ and $h(a_i)$. For the sparse network case the estimator of $a_i$ is not reliable, and we implement the following control functions: linear in $a_i$, $\hat{h}(a_i)$, $\hat{h}(deg_i, x_{2i})$ and $h(a_i)$. In both the dense and sparse setup we also implement a benchmark model with no control for the endogeneity of the network.

We perform simulations with network size $N = 100, 250$. 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 $K_N = 4$. Specifically, Tables 1 and 2 include results for the dense and sparse network specifications, respectively. Results for the other orders of $K_N$ are not notably different and are available upon request.

**Figure 1.** $h(a_i)$ for selected functional forms of $h(a_i)$

![Figure 1](image)

We also perform 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 statistics on the cross-validation in Table 3. The differences in RMSE are very small between the different values of $K_N$.

Note that since $x_{2i}$ is a discrete with a finite support $x_{2i} \in \{x_1, ..., x_M\}$ we have $r(x_{2i}, deg_i) = \sum_{m=1}^{M} r(x_m, deg_i)I\{x_{2i} = x_m\}$. We can then approximate $r(x_{2i}, deg_i) \approx \sum_{k=1}^{K_N} \left\{ \sum_{m=1}^{M} \alpha_{m,k} q_k^{x_m}(deg)I\{x_{2i} = x_m\} \right\}$. 
Analyzing the Monte Carlo results for the dense network specification in Tables 1, we can see that, as expected from our asymptotic theories, the control functions \( \hat{h}(\hat{a}_i) \) and \( \hat{h}(\text{deg}_i, x_{2i}) \) 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 in terms of mean bias. This difference is more pronounced in the case when \( h(a_i) \) is the sine or cosine function. The control for degree approach seems more robust - it yields a lower bias than the \( \hat{h}(\hat{a}_i) \) approach in almost all cases. Also, the former approach has the correct size on all coefficients in all cases, whereas the latter has the incorrect size in some cases. In the simulations we also implemented the control function \( \hat{h}(a_i) \), that is, using the true \( a_i \) instead of \( \hat{a}_i \), in which case the results were much more precise than for \( \hat{h}(\hat{a}_i) \). This suggest that the approach of using \( \hat{h}(\hat{a}_i) \) as a control function works very well when a highly precise estimator of \( a_i \) is available, for example when the network size \( N \) is large.

Looking at Table 2 and the results for the sparse design, we can see that the control for degree approach performs very well across all functional forms of \( h(a_i) \). In the sparse setup, the bias of all estimates, including those that do not control for the endogeneity of the network, is small. However, the size of the no control and linear control estimates are not correct. If a precise estimator of \( a_i \) is available, the control function \( \hat{h}(\hat{a}_i) \) also performs well with low bias and correct size in all cases.

Table 3 shows that the performance of the estimators does not vary notably for different values of \( K_N \). As for the choice of \( K_N \) we present in the tables, we have run simulations for a range of values of \( K_N \) and the results did not differ significantly. As deriving a theory for a data driven choice of \( K_N \) is beyond the scope of this paper, for applied researchers we suggest estimating the model over a range of \( K_N \) and seeing whether the results vary significantly. As shown in our Monte Carlo simulations, the control function approach yields results robust to the choice of \( K_N \) for different non-linear functions. Finally, unless \( a_i \) can be estimated with a high degree of precision, we have shown that the control for degree
## Table 1. Dense Network: Parameter values across 1000 Monte Carlo replications with $K_N = 4$ and Hermite Polynomial sieve

<table>
<thead>
<tr>
<th>$N$</th>
<th>$h(a_i) = \exp(3a_i)$</th>
<th>$h(a_i) = \sin(3a_i)$</th>
<th>$h(a_i) = \cos(3a_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0) (1) (2) (3) (4) (5)</td>
<td>(0) (1) (2) (3) (4) (5)</td>
<td>(0) (1) (2) (3) (4) (5)</td>
</tr>
<tr>
<td>CF</td>
<td>$\beta_1 = 0.8$</td>
<td>$\beta_2 = 5$</td>
<td>$\beta_3 = 5$</td>
</tr>
<tr>
<td></td>
<td>0.094 0.007 0.001 0.001 0.001 0.001</td>
<td>0.008 0.003 0.004 0.001 0.001 0.001</td>
<td>0.020 0.036 0.008 0.005 0.003 0.001</td>
</tr>
<tr>
<td></td>
<td>(0.014 ) (0.020 ) (0.061 ) (0.061 ) (0.062 ) (0.062 )</td>
<td>(0.024 ) (0.027 ) (0.033 ) (0.036 ) (0.040 ) (0.040 )</td>
<td>(0.151 ) (0.169 ) (0.345 ) (0.349 ) (0.352 ) (0.352 )</td>
</tr>
<tr>
<td></td>
<td>0.178 0.165 0.060 0.068 0.061 0.050</td>
<td>0.081 0.057 0.058 0.058 0.071 0.047</td>
<td>0.090 0.071 0.056 0.061 0.058 0.066</td>
</tr>
<tr>
<td></td>
<td>(0.063 ) (0.049 ) (0.060 ) (0.063 ) (0.061 ) (0.066 )</td>
<td>(0.015 ) (0.011 ) (0.017 ) (0.015 ) (0.017 ) (0.017 )</td>
<td>(0.155 ) (0.141 ) (0.141 ) (0.141 ) (0.141 )</td>
</tr>
<tr>
<td></td>
<td>0.058 0.052 0.058 0.058 0.071 0.047</td>
<td>0.102 0.059 0.053 0.053 0.052 0.060</td>
<td>0.103 0.064 0.058 0.062 0.064</td>
</tr>
<tr>
<td></td>
<td>(0.025 ) (0.022 ) (0.025 ) (0.025 ) (0.025 ) (0.025 )</td>
<td>(0.021 ) (0.022 ) (0.022 ) (0.022 ) (0.025 ) (0.025 )</td>
<td>(0.523 ) (0.518 ) (0.518 ) (0.518 ) (0.518 )</td>
</tr>
<tr>
<td></td>
<td>0.058 0.058 0.058 0.058 0.058 0.058</td>
<td>0.058 0.058 0.058 0.058 0.058 0.058</td>
<td>0.064 0.064 0.064 0.064 0.064</td>
</tr>
</tbody>
</table>

CF - control function. (0) - none, (1) - $\tilde{a}_i$, (2) - $\tilde{h}(\tilde{a}_i)$, (3) - $\tilde{h}(\tilde{a}_i)$, (4) - $\tilde{h}(\tilde{deg}_i, x_{2i})$, (5) - $h(a_i)$.

Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2.

size is the empirical size of t-test against the truth.

The bias of $\hat{a}_i$ is calculated as $a_i - \hat{a}_i$. For $N = 100$, $\hat{a}$ - mean bias=0.021, median bias=0.008, std=0.269.

For $N = 250$, $\hat{a}$ - mean bias=0.008, median bias=0.003 , std=0.165.

### 8. 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 approach is more robust and also easier to implement as it does not require an estimate of $a_i$. 


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.
Table 3. **Cross-Validation results**: Parameter values across 1000 Monte Carlo replications

<table>
<thead>
<tr>
<th>$N$</th>
<th>100</th>
<th>250</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h(a_i)$</td>
<td>$\hat{h}(a_i)$</td>
<td>$\hat{h}(a_i)$</td>
</tr>
<tr>
<td>$\exp(a_i)$</td>
<td>$\exp(a_i)$</td>
<td>$\exp(a_i)$</td>
</tr>
<tr>
<td>mean</td>
<td>1.432</td>
<td>1.427</td>
</tr>
<tr>
<td>median</td>
<td>0.617</td>
<td>0.616</td>
</tr>
<tr>
<td>std</td>
<td>2.172</td>
<td>2.198</td>
</tr>
<tr>
<td>iqr</td>
<td>1.608</td>
<td>1.673</td>
</tr>
</tbody>
</table>

| $\cos(a_i)$ | $\cos(a_i)$ | $\cos(a_i)$ |
| mean | 1.840 | 1.841 | 1.836 | 1.848 | 1.717 | 1.661 | 1.693 | 1.718 |
| median | 0.854 | 0.841 | 0.844 | 0.828 | 0.798 | 0.779 | 0.791 | 0.798 |
| std | 2.585 | 2.624 | 2.629 | 2.601 | 2.400 | 2.286 | 2.360 | 2.377 |
| iqr | 2.262 | 2.239 | 2.237 | 2.261 | 2.122 | 2.063 | 2.085 | 2.128 |

| $\sin(a_i)$ | $\sin(a_i)$ | $\sin(a_i)$ |
| mean | 1.852 | 1.815 | 1.876 | 1.870 | 2.329 | 2.282 | 2.185 | 2.179 |
| median | 0.857 | 0.846 | 0.866 | 0.856 | 0.816 | 0.800 | 0.826 | 0.822 |
| std | 2.597 | 2.554 | 2.703 | 2.663 | 4.028 | 4.072 | 3.719 | 3.854 |
| iqr | 2.263 | 2.232 | 2.281 | 2.270 | 2.530 | 2.456 | 2.401 | 2.433 |

| Control function: $\hat{h}(\hat{deg}_i, x_{2i})$ |
|-----|-----|
| $\exp(a_i)$ | $\exp(a_i)$ |
| mean | 2.329 | 2.282 | 2.185 | 2.438 | 2.220 | 2.016 | 2.125 | 2.018 | 2.006 |
| median | 0.838 | 0.862 | 0.845 | 0.893 | 0.849 | 0.791 | 0.838 | 0.826 | 0.809 |
| iqr | 2.530 | 2.456 | 2.401 | 2.550 | 2.431 | 2.220 | 2.361 | 2.289 | 2.253 |

| $\cos(a_i)$ | $\cos(a_i)$ |
| mean | 2.708 | 2.686 | 2.847 | 2.636 | 2.273 | 2.347 | 2.273 | 2.512 | 2.379 |
| median | 1.154 | 1.147 | 1.176 | 1.149 | 1.149 | 1.053 | 1.037 | 1.108 | 1.079 |
| iqr | 3.123 | 3.109 | 3.180 | 3.120 | 3.165 | 2.806 | 2.756 | 2.972 | 2.866 |

| $\sin(a_i)$ | $\sin(a_i)$ |
| mean | 2.633 | 2.556 | 2.736 | 2.687 | 2.648 | 2.371 | 2.338 | 2.342 | 2.405 |
| median | 1.061 | 1.078 | 1.085 | 1.072 | 1.070 | 0.984 | 0.999 | 1.005 | 1.013 |
| iqr | 2.953 | 2.959 | 2.983 | 2.966 | 2.986 | 2.721 | 2.763 | 2.769 | 2.789 |

Average number of links for $N = 100$ is 24.1, for $N = 250$ it is 60.2. The statistics are based on conventional leave one out cross-validation.

using a non-parametric approach. The estimators we propose are 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**


generalized cross-validation in regression with heteroskedastic errors. *Journal of Economet-


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 interac-
tions. *working paper*.

Blume, L. E., W. A. Brock, S. N. Durlauf, and R. Jayaraman (2015). Linear social interac-


and panel data.

Microdata Methods and Practice*.

Ductor, L., M. Fafchamps, S. Goyal, and M. J. van der Leij (2014). Social networks and

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

evidence* (1 ed.), Volume 1. Elsevier B.V.


In this section we introduce the assumptions that are required for the two asymptotic results, Theorem 6.1 for $\hat{\beta}_{2SLS}$ and Theorem 6.2 for $\bar{\beta}_{2SLS}$. We also outline the proof of Theorem 6.1. Detailed proofs are available in the Supplementary Appendix which is available at http://www-bcf.usc.edu/~moonr/. Since the proof of Theorem 6.2 is similar to that of Theorem 6.1, we provide only a sketch of the proof of Theorem 6.2 in the Supplementary Appendix.
In this section we introduce the assumptions used in the proof of Theorem 6.1. 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) $t_{ij} = t_{ji}$. (ii) $u_{ij} \sim i.i.d.$ for all $ij$ a logistic distribution. (iii) The supports of $\lambda$, $t_{ij}$, $a_i$ 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 $\hat{\beta}_{2SLS}$.

**Assumption 7** (Sieve). For every $K_N$ there is a non-singular matrix of constants $B$ such that for $\hat{q}^{KN}(a) = Bq^{KN}(a)$, we assume the following. (i) The smallest eigenvalue of $E[q^{KN}(a_i)\hat{q}^{KN}(a_i)f]$ is bounded away from zero uniformly in $K_N$. (ii) There exists a sequence of constants $\zeta_0(K_N)$ that satisfy the condition $\sup_{a \in A} \|q^{KN}(a)\| \leq \zeta_0(K_N)$, where $K_N$ satisfies $\zeta_0(K_N)^2K_N/N \to 0$ as $N \to \infty$. (iii) For $f(a)$ being an element of $h(a) = (E[y_i|a_i = a], E[z_i|a_i = a], E[w_i|a_i = a])$, there exists a sequence of $\alpha^{f}_{K_N}$ and a number $\kappa > 0$ such that

$$\sup_{a \in A} \|f(a) - q^{KN}(a)\alpha^{f}_{K_N}\| = O(K_N^{-\kappa})$$

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

$$\|q_k(a) - q_k(a')\| \leq \zeta_1(k)\|a - a'\| \quad \forall k = 1, \ldots, K_N$$

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 the polynomial 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 $\zeta_a(N)^{-2}O(K_N^3) = o(1)$.

The next two assumptions are for the sieves used in $\hat{\beta}_{2SLS}$. These assumptions modify Assumption 7 and Assumption 8.
For every $K_N$ there is a non-singular matrix of constants $B$ such that for $\tilde{r}^{K_N}(x_{2i}, \text{deg}_i) = B(x_{2i}, \text{deg}_i)$. We assume the following. (i) The smallest eigenvalue of $E[\tilde{r}^{K_N}(x_{2i}, \text{deg}_i)\tilde{r}^{K_N}(x_{2i}, \text{deg}_i)']$ is bounded away from zero uniformly in $K_N$. (ii) There exists a sequence of constants $\zeta_{0**}(K_N)$ that satisfy the condition $\sup \| \tilde{r}^{K_N}(x_{2i}, \text{deg}_i) \| \leq \zeta_{0**}(K_N)$, where $K_N$ satisfies $\zeta_{0**}(K_N)^2 K_N/N \to 0$ as $N \to \infty$, and $S$ is the domain of $(x_{2i}, \text{deg}_i)$. (iii) For $f(x_{2i}, \text{deg}_i)$ being an element of $h_{**}(x_{2i}, \text{deg}_i) = (E[y_i|x_{2i}, \text{deg}_i], E[z_i|x_{2i}, \text{deg}_i], E[w_i|x_{2i}, \text{deg}_i])$, there exists a sequence of $\gamma_{K_N}^f$ and a number $\kappa > 0$ such that

$$\sup_{(x_{2i}, \text{deg}_i) \in S} \| f - r^{K_N} \gamma_{K_N}^f \| = O(K_N^{-\kappa})$$

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

Recall from (1.7) that $\sup_i \tilde{\text{deg}}_i - \text{deg}_i = O(\zeta_{deg}(N)^{-1})$ with $\zeta_{deg}(N) = o(1) N^{2/3}$ 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

$$\| r_k(x_{2i}, \text{deg}_i) - r_k(x_{2i}, \text{deg}_i') \| \leq \zeta_{1**}(k) \| \text{deg}_i - \text{deg}_i' \| \quad \forall \ k = 1, \ldots, K_N$$

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

The next assumptions restrict the models of the outcome in (3.1) and the network formation of (3.6). We need Assumption 11 to derive the limiting distribution of $\hat{\beta}_{2SLS}$ in Theorem 6.1.

Assumption 11. We assume the following: (i) The true coefficients satisfies $|\beta_0| \leq 1 - \epsilon$ and $\| \beta_2 \| \geq \epsilon$ for some small $\epsilon$. (ii) The parameter set $B$ for $\beta$ is bounded. (iii) The observables $(y_i, x_i)$ 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 $g(t_{ij}, a_i, a_j)$ is bounded by a finite constant, where $t_{ij} := t(x_{2i}, x_{2j})$. (vi) The net surplus of the network $g(t_{ij}, a_i, a_j)$ is a strictly monotonic function of $a_i$ for fixed $(x_{2i}, x_{2j})$ and $a_j$.

Condition (i) is standard in the linear-in-means peer effect literature. As discussed in the main text, the condition $|\beta_0| \leq 1 - \epsilon$ is required for a unique solution of the spillover effect. We need the restriction $\| \beta_2 \| > \epsilon$ 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 $\mathbb{E} [d_{ij,N} = 1] \geq \xi > 0$.

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

Appendix A.2. Estimators

A.2.1. $\hat{\beta}_{2SLS}$. Let $h(a_i) = (h^y(a_i), h^w(a_i), h^z(a_i)) := (E[y_i|a_i], E[w_i|a_i], E[z_i|a_i])$.

To present the estimator $\hat{\beta}_{2SLS}$ in matrix notation, we let $\hat{W}_N = (w_1 - h^y(a_1), \ldots, w_N - h^y(a_N))'$. Similarly we define $\hat{Z}_N, \hat{y}_N$. Suppose that we observe $h(a_i)$. In view of the identification scheme
of Theorem 4.1, we can estimate $\beta^0$ by

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

Let $q^K(a) = (q_1(a), \ldots, q_K(a))'$, $Q_N := Q_N(a_N) = (q^K(a_1), \ldots, q^K(a_N))'$, $h'(a_N) = (h'(a_1), \ldots, h'(a_N))'$, and $\alpha_N^l = (\alpha^l_1, \ldots, \alpha^l_K)$'. Let $b^l_i$ be the $l$th element in $(y_i, w_i', z_i')'$ and denote $b_N^l = (b^l_1, \ldots, b^l_N)$.

If $a_N = (a_1, \ldots, a_N)'$ is observed, in view of (5.4), we can estimate the unknown function $h'(a_N)$ by the OLS of $b^l_i$ on $q^K(a_i)$: for $l = 1, \ldots, L$,

$$\hat{h}'(a_N) = P_{Q_N} b_N^l,$$  \hspace{1cm} (A.2.1.1)

where $P_{Q_N} = Q_N(Q_N' Q_N)^{-} Q_N'$. Here $^-$ denotes any symmetric generalized inverse.

Given this, we suggest to estimate $h'(a_N)$ as follows: (i) first, we estimate the unobserved individual heterogeneity and then (ii) plug the estimate in $\hat{h}'(a_N)$ of (A.2.1.1). To be more specific, suppose that $\hat{a}_N = (\hat{a}_1, \ldots, \hat{a}_N)'$ is an estimator of $a_N = (a_1, \ldots, a_N)'$. Denote $\hat{Q}_N := Q_N(\hat{a}_N) = (q^K(\hat{a}_1), \ldots, q^K(\hat{a}_N))'$. Then the first estimator of $h'(a_N)$ is defined by

$$\hat{h}' := \hat{h}'(\hat{a}_N) = P_{\hat{Q}_N} b_N^l$$  \hspace{1cm} (A.2.1.2)

for $l = 1, \ldots, L$, and this leads the following estimator of $\beta^0$:

$$\hat{\beta}_{2SLS} = \left( W_N'M_{\hat{Q}_N} Z_N \left( Z_N'M_{\hat{Q}_N} Z_N \right)^{-1} Z_N'M_{\hat{Q}_N} W_N \right)^{-1}$$

$$\times W_N'M_{\hat{Q}_N} Z_N \left( Z'M_{\hat{Q}_N} Z_N \right)^{-1} Z_N'M_{\hat{Q}_N} y_N,$$  \hspace{1cm} (A.2.1.3)

where $M_{\hat{Q}_N} = I_N - P_{\hat{Q}_N}$.

### A.2. $\hat{\beta}_{2SLS}$

Suppose that the function $h'(x_{2i}, \text{deg}_i), l = 1, \ldots, L$ is well approximated by a linear combination of base functions $(r_1(x_{2i}, \text{deg}_i), \ldots, r_K(x_{2i}, \text{deg}_i))$:

$$h_{**}(x_{2i}, \text{deg}_i) \equiv \sum_{k=1}^{K_N} r_k(x_{2i}, \text{deg}_i) \gamma_k^l$$

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

Let $\text{Deg}_N = (\text{deg}_1, \ldots, \text{deg}_N)'$. Let $r^K(x_{2i}, \text{deg}_i) = (r_1(x_{2i}, \text{deg}_i), \ldots, r_K(x_{2i}, \text{deg}_i))'$, $R_N := R_N(X_{2N}, \text{Deg}_N) = (r^K(x_{21}, \text{deg}_1), \ldots, r^K(x_{2N}, \text{deg}_N))'$, and $\gamma^l = (\gamma^l_1, \ldots, \gamma^l_K)'$. Let $b_N^l = (b^l_1, \ldots, b^l_N)$. In the case where $(x_{2i}, \text{deg}_i)$ are observed, we can estimate

$$h_{**}(x_{2N}, \text{Deg}_N) = (h_{**}(x_{21}, \text{deg}_1), \ldots, h_{**}(x_{2N}, \text{deg}_N))$$

for $l = 1, \ldots, L$ with

$$\hat{h}_{**}(x_{2N}, \text{Deg}_N) := P_{R_N} b_N^l,$$  \hspace{1cm} (A.2.2.1)

where $P_{R_N} = R_N(R_N' R_N)^{-} R_N'$. Here $^-$ denotes any symmetric generalized inverse.

In view of (5.8), the natural estimator of $\text{deg}_i$ is $\hat{\text{deg}}_i$. This suggests that we estimate $\hat{h}_{**}(x_{2i}, \text{deg}_i)$ by using $\hat{\text{deg}}_i$ in place of $\text{deg}_i$. To be more specific, suppose that $\hat{\text{Deg}}_N = (\hat{\text{deg}}_1, \ldots, \hat{\text{deg}}_N)$. Denote $\hat{R}_N := R_N(X_{2N}, \hat{\text{Deg}}_N) = (r^K(x_{21}, \hat{\text{deg}}_1), \ldots, r^K(x_{2N}, \hat{\text{deg}}_N))'$. The estimator of $h_{**}(x_{2i}, a_i) = \hat{h}_{**}(x_{2i}, a_i)$. 


Then, it leads to the following second estimator of \( \beta_0 \):

\[
\tilde{\beta}_{2SLS} := \left( W_N'M_{\hat{R}_N}Z_N \left( Z'_N M_{\hat{R}_N} Z_N \right)^{-1} Z'_N M_{\hat{R}_N} W_N \right)^{-1} \times W_N'M_{\hat{R}_N}Z_N \left( Z'M_{\hat{R}_N} Z_N \right)^{-1} Z'_N M_{\hat{R}_N} y_N, \tag{A.2.2.2}
\]

where \( M_{\hat{R}_N} = I_N - P_{\hat{R}_N} \).

### Appendix A.3. Outline of the proof of Theorem 6.1

By definition, we have

\[
\tilde{\beta}_{2SLS} - \beta_0 = \left( \frac{1}{N} W_N'M_{\hat{Q}_N}Z_N \left( \frac{1}{N} Z'_N M_{\hat{Q}_N} Z_N \right)^{-1} \frac{1}{N} Z'_N M_{\hat{Q}_N} W_N \right)^{-1} \times \frac{1}{N} W_N'M_{\hat{Q}_N}Z_N \left( \frac{1}{N} Z'_N M_{\hat{Q}_N} Z_N \right)^{-1} \frac{1}{N} Z'_N M_{\hat{Q}_N} \eta_N^\nu - h^\nu(a_N) - \hat{Q}_N \alpha_{KN}^\nu \right).
\]

The derivation of the asymptotic distribution of \( \tilde{\beta}_{2SLS} \) consists of three steps.

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

\[
\sqrt{N} \left( \tilde{\beta}_{2SLS} - \beta_0 \right) = \left( \frac{1}{N} W_N'M_{\hat{Q}_N}Z_N \left( \frac{1}{N} Z'_N M_{\hat{Q}_N} Z_N \right)^{-1} \frac{1}{N} Z'_N M_{\hat{Q}_N} W_N \right)^{-1} \times \frac{1}{N} W_N'M_{\hat{Q}_N}Z_N \left( \frac{1}{N} Z'_N M_{\hat{Q}_N} Z_N \right)^{-1} \frac{1}{N} Z'_N M_{\hat{Q}_N} \eta_N^\nu + o_p(1).
\]

(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 \( h^w(a_i) = \mathbb{E}(w_i|a_i), h^z(a_i) = \mathbb{E}(z_i|a_i) \) and \( \eta_i^z = z_i - h^z(a_i) \). Let \( \hat{h}^w(a_i) \) and \( \hat{h}^z(a_i) \) denote the series approximation of \( h^w(a_i) \) and \( h^z(a_i) \), 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

\[ \frac{1}{N} W_N' M_Q Z_N = \frac{1}{N} \sum_{i=1}^{N} \left( w_i - \tilde{h}^w(a_i) \right) \left( z_i - \tilde{h}^z(a_i) \right)' \]

\[ = \frac{1}{N} \sum_{i=1}^{N} (w_i - h^w(a_i)) (z_i - h^z(a_i))' + o_p(1) \]

\[ \frac{1}{N} Z_N' M_Q Z_N = \frac{1}{N} \sum_{i=1}^{N} (z_i - \tilde{h}^z(a_i)) (z_i - \tilde{h}^z(a_i))' \]

\[ = \frac{1}{N} \sum_{i=1}^{N} (z_i - h^z(a_i)) (z_i - h^z(a_i))' + o_p(1), \]

\[ \frac{1}{\sqrt{N}} Z_N' M_Q \eta_N^v = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - \tilde{h}^z(a_i)) \eta_i^v = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - h^z(a_i)) \eta_i^v + o_p(1). \]

Step 3. The consequence of these two approximations is that \( \sqrt{N}(\hat{\beta}_{2SLS} - \beta_0^{2SLS}) = o_p(1) \). Finally in Step 3, we derive the limiting distribution of the infeasible estimator \( \sqrt{N}(\hat{\beta}_{2SLS}^{inf} - \beta_0) \).
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 $A$, we define matrix norms as follows:

$$
\|A\| = \left(\sum_{i,j} |a_{ij}|^2\right)^{1/2} \quad \text{denotes the Frobenius norm,} 
\|A\|_o \quad \text{denotes the operator norm of matrix } A, 
\text{that is, } \|A\|_o = \lambda_{\max}(A'A)^{1/2}, 
\lambda_{\min}(A) \quad \text{denotes the minimum eigenvalue of } A. 
$$

Notice that

$$
\|A\|_o \leq \|A\| \leq \|A\|_o \text{rank}(A). 
$$

Further, for matrix $A$, $[a]_i$ denotes the $i$’th row of $A$. Denote $[GX]_i$ by $X_{1,G,i}$, $[G^2X]_i$ by $X_{1,G^2,i}$, $[GY]_i$ by $Y_{G,i}$. The $i$th row of the instrument matrix $Z_N$ is given by $z_i' = [X_{2,i}, X_{1,G,i}, X_{1,G^2,i}]$. $z_i$ is $(3l_x) \times 1$. Similarly, $w_i' = [Y_{G,i}, X_{1,i}, X_{1,G,i}]$. We denote matrices by uppercase bold letters and vectors by lowercase bold letters, $Z_N = (Z_1', \ldots, Z_N')$, $W_N = (W_1', \ldots, W_N')$ and $a_N = (a_1, \ldots, a_N)'$.

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

**Outline of the proof of Theorem 6.1** By definition, we have

$$
\hat{\beta}_{2SLS} - \beta^0 = \left( W_N'M_NQ_N Z_N \left( Z_N'M_NQ_N Z_N \right)^{-1} Z_N'M_N Q_N W_N \right)^{-1} 
\times W_N'M_NQ_N Z_N \left( Z_N'M_N Q_N Z_N \right)^{-1} \left( \eta_N^∗ - h^*(a_N) - \hat{Q}_N \alpha_{K_N}^\nu \right). 
$$

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 $a_N$ and approximate it with $\hat{a}_N$. Under suitable assumptions (see Appendix S.1.1), we show that the error that stems from the estimation of $a_N$ by $\hat{a}_N$ is asymptotically negligible:

$$
\sqrt{N} \left( \hat{\beta}_{2SLS} - \beta^0 \right) 
= \left( \frac{1}{N} W_N'M_NQ_N Z_N \left( \frac{1}{N} Z_N'M_NQ_N Z_N \right)^{-1} \frac{1}{N} Z_N'M_N Q_N W_N \right)^{-1} 
\times \frac{1}{N} W_N'M_NQ_N Z_N \left( \frac{1}{N} Z_N'M_N Q_N Z_N \right)^{-1} \frac{1}{\sqrt{N}} Z_N'M_N Q_N \eta_N^∗ \tilde{+} o_p(1).
$$

(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 $h^w(a_i) = \mathbb{E}(w_i|a_i)$, $\eta^w_i = w_i - h^w(a_i)$, $h^z(a_i) = \mathbb{E}(z_i|a_i)$ and $\eta_i^z = z_i - h^z(a_i)$. Let $\tilde{h}^w(a_i)$

\*psyML, ida@psyml.co
\*University of Southern California and Yonsei University, moonr@usc.edu (corresponding author)
and \( \hat{h}^z(a_i) \) denote the series approximation of \( h^w(a_i) \) and \( h^z(a_i) \), 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

\[
\begin{align*}
\frac{1}{N} W_N' M_Q Z_N &= \frac{1}{N} \sum_{i=1}^{N} \left( w_i - \hat{h}^w(a_i) \right) \left( z_i - \hat{h}^z(a_i) \right)' \\
&= \frac{1}{N} \sum_{i=1}^{N} \left( w_i - h^w(a_i) \right) \left( z_i - h^z(a_i) \right)' + o_p(1) \\
\frac{1}{N} Z_N' M_Q Z_N &= \frac{1}{N} \sum_{i=1}^{N} \left( z_i - \hat{h}^z(a_i) \right) \left( z_i - \hat{h}^z(a_i) \right)' \\
&= \frac{1}{N} \sum_{i=1}^{N} \left( z_i - h^z(a_i) \right) \left( z_i - h^z(a_i) \right)' + o_p(1), \\
\frac{1}{\sqrt{N}} Z_N' M_Q \eta_N &= \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( z_i - \hat{h}^z(a_i) \right) \eta_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( z_i - h^z(a_i) \right) \eta_N + o_p(1).
\end{align*}
\]

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

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 3, 4, 8 and 11. Then the following hold.

\[
\begin{align*}
(a) \quad &\frac{1}{N} (Z_N' P_{Q_N} W_N - Z_N' P_{Q_N} W_N) = o_p(1). \\
(b) \quad &\frac{1}{N} (Z_N' P_{Q_N} Z_N - Z_N' P_{Q_N} Z_N) = o_p(1). \\
(c) \quad &\frac{1}{\sqrt{N}} (Z_N' P_{Q_N} \eta_N - Z_N' P_{Q_N} \eta_N) = o_p(1). \\
(d) \quad &\frac{1}{\sqrt{N}} (Z' M_{Q_N} (h^v(a_N) - Q_N \alpha_{K_N}^v)) = o_p(1).
\end{align*}
\]
Proof. Part (a).

\[
\frac{1}{N}(Z_N'P_{Q_N}W_N - Z_N'P_{Q_N}W_N) \\
= Z_N'\left(\hat{Q}_N^T - Q_N^T\right)N^{-1}\hat{Q}_N^T W_N - Z_N'Q_N N^{-1}Q_N^{-1} - Z_N'\left(\hat{Q}_N^T - Q_N^T\right)N^{-1}\hat{Q}_N^T W_N \\
+ Z_N'Q_N N^{-1}\left(\hat{Q}_N^T - Q_N^T\right)N^{-1}\hat{Q}_N^T W_N \\
= Z_N'\left(\hat{Q}_N^T - Q_N^T\right)N^{-1}\hat{Q}_N^T W_N + Z_N'\left(\hat{Q}_N^T - Q_N^T\right)N^{-1}\hat{Q}_N^T W_N \\
- Z_N'Q_N N^{-1}Q_N^{-1} - Z_N'Q_N N^{-1}\hat{Q}_N^T W_N \\
= I_1 + I_2 - I_3 + I_4, \text{ say.}
\]

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

\[
\left\| \frac{1}{N}(Z_N'P_{Q_N}W_N - Z_N'P_{Q_N}W_N) \right\|_o = o_p(1),
\]

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 $I_1$,

\[
\|I_1\|_o \leq \frac{Z_N}{\sqrt{N}} \left\| \left(\hat{Q}_N^T - Q_N^T\right)N^{-1}\hat{Q}_N^T W_N \right\|_o = O_p\left(\frac{1}{\zeta_a(N)^2} \sum_{k=1}^{K_N} \zeta_1(k)^2\right) O_P(1)O(1) = o_p(1),
\]

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

For term $I_2$,

\[
\|I_2\|_o \leq \frac{Z_N}{\sqrt{N}} \left\| \left(\hat{Q}_N^T - Q_N^T\right)N^{-1}\hat{Q}_N^T W_N \right\|_o = O_p\left(\frac{1}{\zeta_a(N)^2} \sum_{k=1}^{K_N} \zeta_1(k)^2\right)^{1/2} O_P(1)\zeta_0(K_N)O(1) = o_p(1),
\]

where the last line holds by (S.1.1.1), Lemmas 4 and 6 and by Assumption 8.
For term $I_3$, write

\[ I_3 = \frac{Z_N'Q_N}{N} \left( \frac{\hat{Q}_N'\hat{Q}_N}{N} \right)^{-1} \left\{ \left( \frac{\hat{Q}_N'\hat{Q}_N}{N} \right) - \left( \frac{Q_N'Q_N}{N} \right) \right\} \left( \frac{Q_N'Q_N}{N} \right)^{-1} \frac{Q_N'W_N}{N} \]

\[ = \frac{Z_N'Q_N}{N} \left( \frac{\hat{Q}_N'\hat{Q}_N}{N} \right)^{-1} \left( \frac{\hat{Q}_N'\hat{Q}_N - Q_N'Q_N}{N} \right) \left( \frac{Q_N'Q_N}{N} \right)^{-1} \frac{Q_N'W_N}{N} \]

\[ + \frac{Z_N'Q_N}{N} \left( \frac{Q_N'Q_N}{N} \right)^{-1} \left( \frac{\hat{Q}_N - Q_N}{N} \right) \left( \frac{Q_N'Q_N}{N} \right)^{-1} \frac{Q_N'W_N}{N} \]

Then,

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

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).

\[ \frac{1}{\sqrt{N}}(Z_N'P\hat{Q}_N\eta_N^v - Z_N'PQ_N\eta_N^v) \]

\[ = \frac{Z_N'\left( \hat{Q}_N - Q_N \right)}{N} \left( \frac{\hat{Q}_N'\hat{Q}_N}{\sqrt{N}} \right)^{-1} \left( \frac{\hat{Q}_N'\hat{Q}_N - Q_N'Q_N}{\sqrt{N}} \right) \frac{Q_N'\eta_N^v}{\sqrt{N}} + \frac{Z_N'\left( \hat{Q}_N - Q_N \right)}{N} \left( \frac{\hat{Q}_N'\hat{Q}_N}{\sqrt{N}} \right)^{-1} \frac{Q_N'\eta_N^v}{\sqrt{N}} \]

\[ - \frac{Z_N'Q_N}{N} \left\{ \left( \frac{Q_N'Q_N}{N} \right)^{-1} - \left( \frac{\hat{Q}_N'\hat{Q}_N}{\sqrt{N}} \right)^{-1} \right\} \frac{Q_N'\eta_N^v}{\sqrt{N}} + \frac{Z_N'Q_N}{N} \left( \frac{\hat{Q}_N'\hat{Q}_N}{\sqrt{N}} \right)^{-1} \left( \frac{\hat{Q}_N - Q_N}{N} \right)' \eta_N^v \]

\[ = III_1 + III_2 + III_3 + III_4, \text{say}, \]

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

\[ \|III_j\| = o_p(1). \]

First, for term $III_1$, we have

\[ \|III_1\| \leq \left\| \frac{Z_N}{\sqrt{N}} \right\| \left\| \frac{\hat{Q}_N - Q_N}{\sqrt{N}} \right\| \left\| \left( \frac{\hat{Q}_N'\hat{Q}_N}{N} \right)^{-1} \right\| \left\| \left( \frac{\hat{Q}_N - Q_N}'\eta_N^v}{\sqrt{N}} \right) \right\| \]

\[ = O_p(1) \left( \frac{1}{\zeta_o(N)^2} \sum_{k=1}^{K_N} \zeta_1(k)^2 \right)^{1/2} O_p(1) \left\| \frac{\hat{Q}_N - Q_N}'\eta_N^v}{\sqrt{N}} \right\| , \]
where the last line holds by (S.1.1.1), Lemmas 4 and 6. Under Assumption we can show that

$$
\mathbb{E} \left[ \left\| \frac{\hat{Q}_N - Q_N}{\sqrt{N}} \right\|^2 \right| X_{1N}, G_N, a_N = \frac{1}{N} \left\| \hat{Q}_N - Q_N \right\|^2.
$$

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

$$
\frac{1}{\sqrt{N}}(Z_N' M_{\hat{Q}_N} (h^v(a_N) - \hat{Q}_N \alpha_{K_N}^v))
$$

$$
= \frac{1}{\sqrt{N}} Z_N' M_{\hat{Q}_N} h^v(a_N)
$$

$$
= \frac{1}{\sqrt{N}} Z_N' (M_{\hat{Q}_N} - M_{Q_N}) h^v(a_N) + \frac{1}{\sqrt{N}} Z_N' M_{Q_N} (h^v(a_N) - Q_N \alpha_{K_N}^v)
$$

$$
= IV_1 + IV_2, \text{ say.}
$$

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

$$
\left\| IV_2 \right\| = \left\| IV_2 \right\| = \left\| \frac{1}{\sqrt{N}} Z_N \right\| \left\| M_{Q_N} \right\| \left\| h^v(a_N) - Q_N \alpha_{K_N}^v \right\| = \frac{1}{\sqrt{N}} Z_N \left\| h^v(a_N) - Q_N \alpha_{K_N}^v \right\|
$$

$$
= O_p(1) \sqrt{N} O(K_N^{-\kappa}) = o_p(1)
$$

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

$$
\frac{1}{N} \| Z_N \|^2 = O_p(1), \frac{1}{N} \| W_N \|^2 = O_p(1).
$$

(S.1.1.1)

**Lemma 3.** Under Assumption 7, we have

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

**Proof.**

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

by Assumption 7 (ii).

**Lemma 4.** Under Assumptions 7, 3, 4, and 5, we have

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

\[
\frac{1}{N} \| \hat{Q}_N - Q_N \|^2 = \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K_N} \| q_k(\hat{a}_i) - q_k(a_i) \|^2 \leq \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K_N} \zeta_1(k)^2 \| \hat{a}_i - a_i \|^2 
\]

\[
\leq \frac{1}{N} \sum_{i=1}^{N} \sum_{k=1}^{K_N} \zeta_1(k)^2 \frac{1}{\zeta_a(N)^2} = \frac{1}{\zeta_a(N)^2} \sum_{k=1}^{K_N} \zeta_1(k)^2, 
\]

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

\[\square\]

Lemma 5. For symmetric matrices \( A \) and \( B \) it is true that

\[
|\lambda_{\min}(A) - \lambda_{\min}(B)| \leq \| A - B \|
\]

Proof. Let \( x_A \) be the eigenvector associated with the minimum eigenvalue of \( A \). Define \( x_B \) analogously. First we show \( |\lambda_{\min}(A) - \lambda_{\min}(B)| \leq \| A - B \| \).

\[
\lambda_{\min}(A) - \lambda_{\min}(B) = x_A^t A x_A - x_B^t B x_B 
\]

\[
\leq x_B^t (A - B) x_B 
\]

\[
\leq |x_B^t (A - B) x_B| \leq \| A - B \|.
\]

Also, we can prove the other direction. Notice that

\[
\lambda_{\min}(A) - \lambda_{\min}(B) = x_A^t A x_A - x_B^t B x_B 
\]

\[
\geq x_A^t (A - B) x_A 
\]

\[
\geq - |x_A^t (A - B) x_A| \geq - \| A - B \|.
\]

Then, we have the required result.

\[\square\]

Lemma 6. Under 7, 8, W.p.a.1, there exists a positive constant \( C > 0 \) such that

\[
\frac{1}{C} \leq \lambda_{\min} \left( \frac{Q_N' Q_N}{N} \right), \lambda_{\min} \left( \frac{\hat{Q}_N' \hat{Q}_N}{N} \right).
\]

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

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

For this, by Lemma 5 we have

\[
\left| \lambda_{\min} \left( \frac{Q_N' Q_N}{N} \right) - \mathbb{E}[q^{K_N}(a_i)q^{K_N}(a_i)'] \right| \leq \left\| \frac{Q_N' Q_N}{N} - \mathbb{E}[q^{K_N}(a_i)q^{K_N}(a_i)'] \right\|
\]

\[
= \left\| \frac{1}{N} \sum_{i=1}^{N} (q^{K_N}(a_i)q^{K_N}(a_i)') - \mathbb{E}[q^{K_N}(a_i)q^{K_N}(a_i)'] \right\|.
\]
Then, by Assumption 7(ii), we have

$$\mathbb{E}\left|\frac{1}{N} \sum_{i=1}^{N} (q_i^{KN}(a_i)q_i^{KN}\eta_i + 2 \mathbb{E}[q_i^{KN}(a_i)q_i^{KN}\eta_i])\right|^2$$

$$= \sum_{k=1}^{KN} \sum_{l=1}^{KN} \mathbb{E}\left[ \frac{1}{N} \sum_{i=1}^{N} (q_k(a_i)q_l(a_i) - \mathbb{E}[q_k(a_i)q_l(a_i)]) \right]^2$$

$$\leq \frac{1}{N} \sum_{k=1}^{KN} \sum_{l=1}^{KN} \mathbb{E}[q_k(a_i)q_l(a_i)]^2 \leq \frac{1}{N} \sup_a \left( \sum_{k=1}^{KN} q_k(a)^2 \right)^2$$

$$\leq \frac{\zeta_0(KN)^4}{N} = o(1),$$

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

$$|\lambda_{min}(Q_N) - \lambda_{min}(\hat{Q}_N)| = o_p(1).$$

Notice by Lemma 3 for symmetric matrices $A$ and $B$, we have

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

Then,

$$|\lambda_{min}(\hat{Q}_N) - \lambda_{min}(Q_N)| \leq \|\hat{Q}_N - Q_N\|$$

$$\leq \left\| \frac{Q_N - \hat{Q}_N}{\sqrt{N}} \right\| + \left\| \frac{Q_N - \hat{Q}_N}{\sqrt{N}} \right\|$$

$$\leq \frac{\zeta_0(KN)^4}{N} = o(1),$$

as desired. 

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

**Lemma 7** (Series Approximation). Assume the assumptions in Lemma 3. Then, we have

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

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

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

**Proof.** Lemma 7 follows if we show
(i) \( \frac{1}{N} \sum_{i=1}^{N} (\hat{h}_w(a_i) - h_w(a_i))(\hat{h}_w(a_i) - h_w(a_i))' = o_p(1) \).

(ii) \( \frac{1}{N} \sum_{i=1}^{N} (\hat{h}_z(a_i) - h_z(a_i))(\hat{h}_z(a_i) - h_z(a_i))' = o_p(1) \).

(iii) \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (\hat{h}_z(a_i) - h_z(a_i)) \eta_i = o_p(1) \).

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

Following Newey (1997), we assume \( B = I \) in Assumption 7 hence, \( \hat{q}^K(a) = q^K(a) \).

Also, we assume \( P = \mathbb{E}[q^K(a_i)(q^K(a_i))'] = I \).

Lemma 8. Assume Assumption 7. Then, \( \mathbb{E}[\|\hat{P} - I\|^2] = O(\zeta_0(K_N)^2 K_N/N) \), where \( \hat{P} = (Q_N' Q_N)/N \).

**Proof.** For proof see Li and Racine (2007) page 481. \( \square \)

Note that Lemmas 5 and 8 imply that

\[ |\lambda_{\min}(\hat{P}) - 1| \leq \|\hat{P} - I\| = O_p(\zeta_0(K_N)\sqrt{K_N/N}) = o_p(1). \]

That is, the smallest eigenvalue of \( \hat{P} \) converges to one in probability. Letting \( 1_N \) be the indicator function for the smallest eigenvalue of \( \hat{P} \) being greater than 1/2, we have \( \Pr(1_N = 1) \to 1 \).

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

**Proof.**

\[
\begin{align*}
1_N\|\hat{a}^f - \alpha^f\| &= 1_N\|Q_N'(Q_N^{-1}Q_N' f - Q_N\alpha^f)\|
= 1_N\{(f - Q_N\alpha^f)'Q_N(Q_N^{-1}Q_N' f - Q_N\alpha^f)/N\}^{1/2}
= 1_NO_p(1)\{(f - Q_N\alpha^f)'Q_N(Q_N^{-1}Q_N' f - Q_N\alpha^f)/N\}^{1/2}
\leq O_p(1)\{(f - Q_N\alpha^f)'(f - Q_N\alpha^f)/N\}^{1/2} = O_p(K_N^{-\kappa})
\end{align*}
\]

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

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

**Proof.** The required result for the lemma follows because

\[
\begin{align*}
\frac{1}{N}\|f - \tilde{f}\|^2 &\leq \frac{1}{N}\{\|f - Q_N\alpha_N^f\|^2 + \|Q_N(\alpha_N^f - \hat{a}_N^f)\|^2\}
= O(K_N^{-2\kappa}) + (\alpha_N^f - \hat{a}_N^f)'(Q_N' Q_N/N)(\alpha_N^f - \hat{a}_N^f)
= O(K_N^{-2\kappa}) + O_p(1)\|\alpha_N^f - \hat{a}_N^f\|^2 = O_p(K_N^{-2\kappa})
\end{align*}
\]

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

\( \text{The Lemmas in this section follow Section 15.6 in Li and Racine (2007).} \)
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 $s_0(x_i, a_i)$ be a function of $(x_i, a_i)$ such that $s_0(\cdot, \cdot)$ is bounded over the support of $(x_i, a_i)$. We denote an $N$ vector-valued function that stacks $s_0(x_i, a_i)$ over $i = 1, \ldots, N$ as $S_{0,N} = (s_0(x_1, a_1), \ldots, s_0(x_N, a_N))'$. Define

$$s_{0,N,i} := s_0(x_i, a_i).$$  \hspace{1cm} (S.1.3.1)

Next, for $m = 1, 2, \ldots$, we define recursively

$$s_{m,N,i} := \sum_{j=1, j \neq i}^{N} g_{ij,N} s_{m-1,N,i} = [G_N S_{m-1,N}]_i,$$  \hspace{1cm} (S.1.3.2)

where

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

For $m = 0, 1, 2, \ldots$, we define $s_{x, m,N,i}^1$ and $S_{x, m,N}$ with initial function $s_{0,N,i} = s_0(x_i, a_i) = x_{1i}$, and define $s_{a, m,N,i}^a$ and $S_{a, m,N}$ with initial function $s_{0,N,i} = s_0(x_i, a_i) = h^v(a_i)$.

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

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

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

$$s_{1,N,i} = [G_N S_{0,N}]_i$$

$$= \left( \frac{1}{N} \sum_{j \neq i} d_{ij,N} \right)^{-1} \left( \frac{1}{N} \sum_{j \neq i} d_{ij,N} s_0(x_j, a_j) \right)$$

$$= \left( \frac{1}{N} \sum_{j \neq i} \mathbb{I} \{g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij} \} \right)^{-1} \times \frac{1}{N} \sum_{j \neq i} \mathbb{I} \{g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij} \} s_0(x_j, a_j)$$

$$\overset{p}{\rightarrow} \int \int \int \int p(g(t(x_{2i}, x_{2j}), a_i, a) \in [x, a]) \pi(x, a) dx da$$

$$\int \int p(g(t(x_{2i}, x_{2j}), a_i, a) \in [x, a]) \pi(x, a) dx da$$

$$= \frac{\mathbb{E}[d_{ij,N} s_0(x_j, a_j)|x_i, a_i]}{\mathbb{E}[d_{ij,N}|x_i, a_i]} =: \tilde{s}_1(x_i, a_i),$$  \hspace{1cm} (S.1.3.3)

where $\pi(x, a)$ with $x = (x_1, x_2)$ is the joint density of $x_i = (x_{1i}, x_{2i})$ and $a_i$, and $\pi(x_2, a)$ is the joint density of $(x_{2i}, a_i)$. Here note that the limit $\tilde{s}_1(x_i, a_i)$ depends only on $(x_i, a_i)$, not on $(x_{-i}, a_{-i})$, while $s_{1,N,i}$ depends on both $(x_i, a_i)$ and $(x_{-i}, a_{-i})$. 
We define the following recursively for \( m = 2, 3, \ldots \) as follows:

\[
\hat{s}_m(x_i, a_i) := \frac{E[d_{ij,N}\hat{s}_{m-1}(x_j, a_j)|x_i, a_i]}{E[d_{ij,N}|x_i, a_i]}
\]

\[
= \frac{\int \int p(g(t(x_2, x_2), a_i, a)\hat{s}_{m-1}(x, a)\pi(x, a)dx} \int \int p(t(x_2, x_2), a_i, a)\pi(x_2, a)dx_2da
\]

\[
= \text{plim}_{N \to \infty} \left( \frac{1}{N} \sum_{j \neq i} d_{ij,N} \right)^{-1} \frac{1}{N} \sum_{j \neq i} d_{ij,N}\hat{s}_{m-1}(x_j, a_j)
\]

\[
= \text{plim}_{N \to \infty} \left[ G_N \hat{S}_{m-1} \right]_i,
\]

where \( \hat{S}_m = (\hat{s}_m(x_1, a_1), \ldots, \hat{s}_m(x_N, a_N)) \).

Using this general definitions of [S.1.3.3] and [S.1.3.4], with \( \hat{s}_0^x(x_i, a_i) = \hat{s}_0^xa(x_i, a_i) = x_{1i} \) and \( \hat{s}_0^a(x_i, a_i) = s_0^a(x_i, a_i) = h(a_i) \), we define \( \hat{s}_m^x(x_i, a_i) \) and \( \hat{s}_m^a(x_i, a_i) \), respectively, for \( m = 1, 2, \ldots \).

Let \( \hat{S}_m^x = (\hat{s}_m^x(x_1, a_1), \ldots, \hat{s}_m^x(x_N, a_N))' \), and \( \hat{S}_m^a = (\hat{s}_m^a(x_1, a_1), \ldots, \hat{s}_m^a(x_N, a_N))' \).

Next, with the initial function \( s_{0,N,i}^v = \eta_i^v \) and \( S_{0,N}^v := (s_{0,N,1}, \ldots, s_{0,N,N})' \), we define recursively

\[
s_{m,N,i}^v := [G_N S_{m-1,N}]_i = \sum_{j=1}^{N} g_{ij,N} s_{m-1,N,i}^v, \quad (S.1.3.5)
\]

and \( S_{m,N}^v := (s_{m,N,1}, \ldots, s_{m,N,N})' \) for \( m = 1, 2, \ldots \).

**Lemma 11.** Under Assumptions \([7]\) and \([11]\) as \( N \to \infty \), we have

\( a) \quad \frac{1}{N} \sum_{i=1}^{N} (w_i - h^w(a_i))(z_i - h^z(a_i))' \)

\[=: \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right)' \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right) \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right) \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right) \]

\[\overset{P}{\to} \left( S_{G1}^z, S_{G1,G1}^z \right) =: S_{wz}, \]

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

\[=: \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right)' \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right) \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right) \left( \frac{1}{N} \sum_{i=1}^{N} \eta_i^G(x_i) \eta_i^x \right) \]

\[\overset{P}{\to} \left( S_{G1}^z, S_{G1,G1}^z \right) =: S_{zz}, \]
where

\[ S_{GY,G^*x_i} = \mathbb{E} \left[ \left( \sum_{m=0}^{\infty} \beta_0^m \tilde{s}_m(x_i, a_i) + \beta_3^0 \tilde{s}_m(x_i, a_i) + \tilde{s}_m(x_i, a_i) \right) \left( \tilde{s}_r(x_i, a_i) \right) \right], \quad r = 0, 1, 2 \]

\[ S_{G^*x_i,G^*x_i} = \mathbb{E} \left[ \tilde{s}_r(x_i, a_i) \left( \tilde{s}_s(x_i, a_i) \right) \right], \quad r, s = 0, 1, 2 \]

\[ \tilde{s}_m(x_i, a_i) = \tilde{s}_m(x_i, a_i) - \mathbb{E}[s_m(x_i, a_i)| a_i] \quad \text{with} \quad \tilde{s}_0(x_i, a_i) = x_{1i} \]

\[ \tilde{s}_m(x_i, a_i) = \tilde{s}_m(x_i, a_i) - \mathbb{E}[s_m(x_i, a_i)| a_i] \quad \text{with} \quad \tilde{s}_0(x_i, a_i) = h^v(a_i). \]

and \( \tilde{s}_m(x_i, a_i) \) and \( \tilde{s}_m(x_i, a_i) \) are defined recursively as in \([S.1.3.4]\).

**Proof**

We take the element \( \frac{1}{N} \sum_{i=1}^{N} \eta_i^{GY}(\eta_i^{G^2x_1})' \) as an example. The proofs of the rest are similar and we omit them.

When \( |\beta|^2 < 1 \),

\[ G_N y_N = \sum_{m=0}^{\infty} (\beta^0_1)^m G_N^m(X_{1N}, a^0_2) + G_N X_{1N}, a^0_2 + h^v(a_N) + \eta_N^v, \]

and

\[ [G_N y_N]_i \]

\[ = \beta_2^0 \left[ \sum_{m=0}^{\infty} (\beta_1^0)^m G_N^m X_{1N} \right]_i + \beta_3^0 \left[ \sum_{m=0}^{\infty} (\beta_1^0)^m G_N^{m+1} X_{1N} \right]_i \]

\[ + \left[ \sum_{m=0}^{\infty} (\beta_1^0)^m G_N^m h(a_N) \right]_i + \left[ \sum_{m=0}^{\infty} (\beta_1^0)^m G_N^m \eta_N^v \right]_i. \]

Set \( s^x_{N}(x_i, a_i) = \tilde{s}^x_{N}(x_i, a_i) = x_{1i} \). We have

\[ \frac{1}{N} \sum_{i=1}^{N} \eta_i^{GY}(\eta_i^{G^2x_1})' = 1 \]

\[ \frac{1}{N} \sum_{i=1}^{N} \left( [G_N y_N]_i - \mathbb{E}[G_N y_N]|a_i) \right) \left( [G_N^2 X_{1N}]_i - \mathbb{E}[G_N^2 X_{1N}]|a_i) \right)' \]

\[ = \frac{1}{N} \sum_{i=1}^{N} \left( \beta_2^0 \sum_{m=0}^{\infty} (\beta_1^0)^m \left( s^x_{m,N,i} - \mathbb{E}[s^x_{m,N,i}|a_i) \right) \right) \left( s^x_{2,N,i} - \mathbb{E}[s^x_{2,N,i}|a_i) \right)' \]

\[ + \frac{1}{N} \sum_{i=1}^{N} \left( \beta_3^0 \sum_{m=0}^{\infty} (\beta_1^0)^m \left( s^x_{m+1,N,i} - \mathbb{E}[s^x_{m+1,N,i}|a_i) \right) \right) \left( s^x_{2,N,i} - \mathbb{E}[s^x_{2,N,i}|a_i) \right)' \]

\[ + \frac{1}{N} \sum_{i=1}^{N} \left( \beta_2^0 \sum_{m=0}^{\infty} (\beta_1^0)^m \left( s^a_{m,N,i} - \mathbb{E}[s^a_{m,N,i}|a_i) \right) \right) \left( s^a_{2,N,i} - \mathbb{E}[s^a_{2,N,i}|a_i) \right)' \]

\[ + \frac{1}{N} \sum_{i=1}^{N} \left( \beta_2^0 \sum_{m=0}^{\infty} (\beta_1^0)^m \left( s^v_{m,N,i} - \mathbb{E}[s^v_{m,N,i}|a_i) \right) \right) \left( s^v_{2,N,i} - \mathbb{E}[s^v_{2,N,i}|a_i) \right)' \]

\[ = I + II + III + IV, \quad \text{say.} \]
Consider term $I$,

$$\frac{1}{N} \sum_{i=1}^{N} \left( \beta_2^{\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i} - \mathbb{E}[s_{m,N,i}^{x_1}] \right\} \right) \left( s_{2,N,i}^{x_1} - \mathbb{E}[s_{2,N,i}^{x_1}] \right)^{\prime}.$$

Denote

$$A_{1i} := \beta_2^{\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{x_1} - \mathbb{E}[s_{m,N,i}^{x_1}] \right\}$$

$$A_{2i} := s_{2,N,i}^{x_1} - \mathbb{E}[s_{2,N,i}^{x_1}]$$

$$A_{3i} := \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ s_{m,N,i}^{x_1} - \mathbb{E}[s_{m,N,i}^{x_1}] \right\}$$

$$B_{1i} := \beta_2^{\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ \tilde{s}_{m}^{x_1}(x_i, a_i) - \mathbb{E}[^{x_1} s_{m}^{x_1}(x_i, a_i)] \right\}$$

$$B_{2i} := \tilde{s}_{2}^{x_1}(x_i, a_i) - \mathbb{E}[\tilde{s}_{2}^{x_1}(x_i, a_i) | a_i]$$

$$B_{3i} := \eta_{i}^{\prime} = v_i - \mathbb{E}[v_i | a_i].$$

First, notice that

$$\left\| \frac{1}{N} \sum_{i=1}^{N} A_{1i} A_{2i} - \frac{1}{N} \sum_{i=1}^{N} B_{1i} B_{2i} \right\| \leq \left\| \frac{1}{N} \sum_{i=1}^{N} (A_{1i} - B_{1i}) A_{2i}^{\prime} + \frac{1}{N} \sum_{i=1}^{N} B_{1i}(A_{2i} - B_{2i})^{\prime} \right\|$$

$$\leq \sup_i \left\| A_{1i} - B_{1i} \right\| \sup_i \left\| A_{2i} \right\| + \sup_i \left\| B_{1i} \right\| \sup_i \left\| A_{2i} - B_{2i} \right\|$$

(S.1.3.6)

According to Lemma 10 and Lemma 14 we have

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

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

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

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

$$\frac{1}{N} \sum_{i=1}^{N} B_{1i} B_{2i}^{\prime} \xrightarrow{p} \mathbb{E} \left[ B_{1i} B_{2i} \right]$$

$$= \mathbb{E} \left[ \left( \beta_2^{\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \left\{ \tilde{s}_{m}^{x_1}(x_i, a_i) - \mathbb{E}[^{x_1} \tilde{s}_{m}^{x_1}(x_i, a_i)] \right\} \right) \left( \tilde{s}_{2}^{x_1}(x_i, a_i) - \mathbb{E}[\tilde{s}_{2}^{x_1}(x_i, a_i) | a_i] \right) \right]$$

$$= \mathbb{E} \left[ \left( \beta_2^{\prime} \sum_{m=0}^{\infty} (\beta_1^0)^m \tilde{s}_{m}^{x_1}(x_i, a_i) \right) \tilde{s}_{2}^{x_1}(x_i, a_i) \right]$$
We can derive the probability limits of terms II and III by similar fashion.

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

\[
E[s_{m,N,i}^u|a_i] = E\left( [G_N^m \eta_N^v]_i|a_i \right) \\
= E \{ E([G_N^m \eta_N^v]_i|X_N, D_N, a_i)|a_i \} \\
= E \{ G_N^m E(\eta_N^v|X_N, D_N, a_i)|a_i \} = 0,
\]

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

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

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

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

\[
\frac{1}{N} \sum_{i=1}^{N} B_{3i} B_{2i}' \xrightarrow{P} E[B_{3i} B_{2i}'] \\
= E[\eta_i^u (s_2^X i(x_i, a_i) - E[s_2^X i(x_i, a_i)|a_i])] \\
= E [(v_i - E[v_i|a_i]) s_2^X i(x_i, a_i)] \\
= E \{ E(v_i - E[v_i|a_i]|x_i, a_i) s_2^X i(x_i, a_i) \} \\
= 0.
\]

Let $\sigma^2(x_i, a_i) := E[(\eta_i^u)^2|x_i, a_i] = E[(v_i - E[v_i|a_i])^2|x_i, a_i]$.

**Lemma 12.** Under Assumptions 7 and 11 as $N \rightarrow \infty$, we have

\[
\frac{1}{N} \sum_{i=1}^{N} (z_i - h^z(a_i))(z_i - h^z(a_i))' \sigma^2(x_i, a_i) \xrightarrow{P} S_{zz}^\sigma,
\]

where the limit variance $S_{zz}^\sigma$ 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 7 and 11 as $N \rightarrow \infty$, we have

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - h^z(a_i)) \eta_i^v \Rightarrow \mathcal{N}(0, S_{zz}^\sigma),
\]
According to Lemma 12, for some finite constant $M$ and so

$$\tilde{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{N} \left[ \ell' (z_i - h^2(a_i)) (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i \right] = \frac{1}{N} \sum_{i=1}^{n} \left[ \ell' (z_i - h^2(a_i)) (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i \right] = \frac{1}{N} \sum_{i=1}^{n} \left[ \ell' (z_i - h^2(a_i)) (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i \right]$$

Let $\ell$ be a nonzero vector whose dimension is the same as the IVs $z_i$. Then,

$$\mathbb{E}[\ell' (\eta_i^v)^2 | F_i] = (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i] = (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i]$$

Let

$$s^2_N := 1 - \sum_{i=1}^{N} \frac{\mathbb{E}[\ell' (z_i - h^2(a_i)) (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i]}{\mathbb{E}[\ell' (z_i - h^2(a_i)) (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i]} = \frac{1}{N} \sum_{i=1}^{n} \frac{\mathbb{E}[\ell' (z_i - h^2(a_i)) (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i]}{\mathbb{E}[\ell' (z_i - h^2(a_i)) (z_i - h^2(a_i))^\ell (\eta_i^v)^2 | F_i]}$$

According to Lemma 12

$$\frac{s^2_N}{p} \rightarrow S^{z\sigma}.$$
Also, since \(\ell'(z_i - h^2(a_i))\eta_i^\nu = \ell'(z_i - h^2(a_i))(v_i - E[v_i|a_i])\) is bounded by a constant, under Assumption [11] the Lindeberg-Feller condition is satisfied, that is, for any \(\epsilon > 0\),

\[
\frac{1}{N} \sum_{i=1}^{N} E \left[ \left( \ell'(z_i - h^2(a_i))(z_i - h^2(a_i))'\ell(\eta_i^\nu) \right)^2 I\left\{ |\ell'(z_i - h^2(a_i))\eta_i^\nu| > \epsilon \sqrt{N} \right\} |F_i \right]
\]

\[
\leq \sum_{i=1}^{N} \frac{1}{\epsilon^2 N^2} E \left[ |\ell'(z_i - h^2(a_i))(z_i - h^2(a_i))'-\ell(\eta_i^\nu)|^4 |F_i \right]
\]

\[
\leq \frac{M}{\epsilon N} \rightarrow 0
\]

as \(N \rightarrow \infty\).

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

\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - h^2(a_i))\eta_i^\nu \Rightarrow N(0, Szz\sigma).
\]

\[\square\]

**Proof of Theorem 6.1.**

Theorem 6.1 follows from Lemma 2, Lemma 7, Lemma 11, and Lemma 13. \[\square\]

### 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 \(s_0(x_i, a_i)\) is a bounded function of \(x_i\) and \(a_i\). Suppose that we define \(s_{m,N,i}\) as in \(\text{(S.1.3.2)}\) and consider its probability limit \(\tilde{s}_m(x_i, a_i)\) in equation \(\text{(S.1.3.4)}\) for each \(i\). Then, for each \(m = 0, 1, 2, \cdots\)

\[
(a) \sup_{1 \leq i \leq N} |s_{m,N,i} - \tilde{s}_m(x_i, a_i)| = o_p(1)
\]

\[
(b) \sup_{1 \leq i \leq N} |E[s_{m,N,i}|a_i] - E[\tilde{s}_m(x_i, a_i)|a_i]| = o_p(1).
\]

**Proof**

**Part (a).**

For \(m = 0\).

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

Next we show the required result for \(m = 1\) and then use mathematical induction for the rest \(m = 2, 3, \ldots\)

For \(m = 1\).

The claim for the case \(m = 1\) is proved in three steps.
Step 1.
Notice that
\[ s_{1,N,i} = \left( \frac{1}{N} \sum_{j \neq i} d_{ij,N} \right)^{-1} \frac{1}{N} \sum_{j \neq i} d_{ij,N} s_{1,N,j} \]
\[ = \left( \frac{1}{N} \sum_{j \neq i} \mathbb{I} \{ g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij} \} \right)^{-1} \]
\[ \times \frac{1}{N} \sum_{j \neq i} \mathbb{I} \{ g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij} \} s_0(x_j, a_j). \]

Then, by the WLLN, for each \( i \),
\[ \frac{1}{N} \sum_{j \neq i} \mathbb{I} \{ g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij} \} \xrightarrow{p} \int \int \Phi((t(x_{2i}, x_2), a_i, a) \pi(x_2, a) dx_2 da \]
\[ = \mathbb{E}[d_{ij,N}|x_i, a_i] \quad (S.1.4.1) \]
\[ \frac{1}{N} \sum_{j \neq i} \mathbb{I} \{ g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij} \} s_0(x_j, a_j) \xrightarrow{p} \int \int \Phi(t(x_{2i}, x_2)' \lambda^0 + a_i + a)s_0(x, a) \pi(x, a) dx da \]
\[ = \mathbb{E}[d_{ij,N}s_0(x_j, a_j)|x_i, a_i]. \quad (S.1.4.2) \]

Since \( \mathbb{E}[d_{ij,N}|x_i, a_i] > 0 \) uniformly in \( i, j \) under Assumption (vi),(v), and (vi) for each \( i \) as \( N \to \infty \), we have
\[ s_{1,N,i} \xrightarrow{p} \tilde{s}_1(x_i, a_i) = \frac{\int \int \Phi(g(t(x_{2i}, x_2), a_i, a)s_0(x, a) \pi(x, a) dx da}{\int \int \Phi(g(t(x_{2i}, x_2), a_i, a) \pi(x, a) dx da}. \]

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
\[ \zeta_{i,N,1} = \frac{1}{N} \sum_{j=1, \neq i}^N (d_{ij,N} - \mathbb{E}[d_{ij,N}|x_i, a_i]) \]
and
\[ \zeta_{i,N,2} = \frac{1}{N} \sum_{j=1, \neq i}^N (d_{ij,N}s_0(x_j, a_j) - \mathbb{E}[d_{ij,N}s_0(x_j, a_j)|x_i, a_i]). \]

Notice that conditional on \( (x_i, a_i) \), \( d_{ij,N} \) and \( d_{ij,N}s_0(x_j, a_j) \) are iid with conditional mean zero and bounded by a constant across \( j = 1, \ldots, N, \neq i \). Then, there exists a finite constant \( M_1 \) such that
\[ \sup_i \mathbb{E} \left( \|N\zeta_{i,N,k}\|^4|x_i, a_i\right) \leq M_1, \]
and we can deduce the desired result
\[ \sup_i \|\zeta_{i,N,k}\| = O_p(N^{-1/4}) = o_p(1) \]
because for any $\epsilon > 0$, we choose $M_2 = \frac{\epsilon}{M_1^4}$ and then

$$
P\{\sup_i \|\zeta_{i,N,k}\| \geq N^{-1/4}M_2^{1/4} | x_i, a_i\} = P\{\sup_i N^{-1/4}\sqrt{N} \zeta_{i,N,k} \| \geq M_2^{1/4} | x_i, a_i\}
$$

$$= P\{\sup_i N^{-1/4}\sqrt{N} \zeta_{i,N,k} \|^4 \geq M_2^4 | x_i, a_i\}
$$

$$\leq P\left\{ \frac{1}{N} \sum_{i=1}^N \|\sqrt{N} \zeta_{i,N,k}\|^4 \geq M_2^4 | x_i, a_i\right\}
$$

$$\leq \frac{1}{M_2^2} \frac{1}{N} \sum_{i=1}^N E \left( \|\sqrt{N} \zeta_{i,N,k}\|^4 | x_i, a_i\right)
$$

$$\leq \frac{M_1}{M_2} \epsilon.
$$

**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 $\Psi_{i,N,2} = \frac{1}{N} \sum_{j \neq i} d_{ij,N}s_0(x_j, a_j)$. Then,

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

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

$$\sup_i \|s_{1,N,i}\| = \sup_i \left\| \frac{\Psi_{i,N,2}}{\Psi_{i,N,1}} - \frac{\Psi_{i,2}}{\Psi_{i,1}} \right\|
$$

$$\leq \sup_i \left\| \frac{\Psi_{i,N,2} - \Psi_{i,2}}{\Psi_{i,N,1}} \right\| + \sup_i \left\| \frac{\Psi_{i,2}(\Psi_{i,N,1} - \Psi_{i,1})}{\Psi_{i,N,1}\Psi_{i,1}} \right\| = o_p(1),
$$

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

$$\sup_i \|s_{1,N,i} - \hat{s}_1(x_i, a_i)\| = o_p(1).
$$

**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

$$\sup_{1 \leq i \leq N} \|s_{m,N,i} - \hat{s}_m(x_i, a_i)\| = o_p(1).$$
Then, we have

\[
\sup_{1 \leq i \leq N} \left\| s_{m+1,N,i} - \tilde{s}_{m+1}(x_i, a_i) \right\| = \sup_{1 \leq i \leq N} \left\| \frac{1}{N} \sum_{j=1}^{N} d_{ij,N} s_{m,N,i} - \mathbb{E}[d_{ij,N} \tilde{s}_m(x_j, a_j)|x_i, a_i] - \mathbb{E}[d_{ij,N}|x_i, a_i] \right\|
\]

\[
\leq \sup_{1 \leq i \leq N} \left\| \frac{1}{N} \sum_{j=1, \neq i}^{N} d_{ij,N} (s_{m,N,i} - \mathbb{E}[d_{ij,N} \tilde{s}_m(x_j, a_j)|x_i, a_i]) \right\|
\]

\[
+ \sup_{1 \leq i \leq N} \left\| \mathbb{E}[d_{ij,N} \tilde{s}_m(x_j, a_j)|x_i, a_i] \sup_{1 \leq i \leq N} \left| \frac{1}{N} \sum_{j=1, \neq i}^{N} d_{ij,N} - \frac{1}{N} \sum_{j=1, \neq i}^{N} d_{ij,N} \right| \right. \]

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 \( \sum_{j=1, \neq i}^{N} g_{ij,N} = 1 \), we have

\[
\sup_{1 \leq i \leq N} \left\| \frac{1}{N} \sum_{j=1, \neq i}^{N} d_{ij,N} (s_{m,N,i} - \mathbb{E}[d_{ij,N} \tilde{s}_m(x_j, a_j)|x_i, a_i]) \right\| = \sup_{1 \leq i \leq N} \frac{1}{N} \sum_{j=1, \neq i}^{N} g_{ij,N} (s_{m,N,i} - \mathbb{E}[d_{ij,N} \tilde{s}_m(x_j, a_j)|x_i, a_i])
\]

\[
\leq \sup_{1 \leq i \leq N} \left\| s_{m,N,i} - \mathbb{E}[d_{ij,N} \tilde{s}_m(x_j, a_j)|x_i, a_i] \right\|
\]

\[
= o_p(1),
\]

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

\[
\sup_{1 \leq i \leq N} \left\| \mathbb{E}[d_{ij,N} \tilde{s}_m(x_j, a_j)|x_i, a_i] \sup_{1 \leq i \leq N} \left| \frac{1}{N} \sum_{j=1, \neq i}^{N} d_{ij,N} - \frac{1}{N} \sum_{j=1, \neq i}^{N} d_{ij,N} \right| \right. \]

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 \( \mathbb{E}[\tilde{s}_m(x_i, a_i)|a_i] \) 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 \( s_{m,N,i}^u \) in \( i \)). Assume Assumptions 7, 8, 9, 10 and 11. Suppose that we define \( s_{m,N,i}^u \) as in (S.1.3.5). Then, for each \( m = 1, 2, \ldots \)

\[
\sup_{1 \leq i \leq N} |s_{m,N,i}^u| = o_p(1).
\]

**Proof**

The proof is similar to that of Lemma 14. First, we show that for each \( i \) and \( m = 1, 2, \ldots \) the probability limit of \( s_{m,N,i}^u \) defined with \( s_{0,i}^u = n_i^u = v_i - \mathbb{E}[v_i|a_i] \) recursively as (S.1.3.5) is zero as
To verify this, let
\[ \tilde{s}_{0,i}^v = \eta_i^v = v_i - \mathbb{E}[v_i|a_i]. \]

For \( m = 1 \),
\[ s_{1,N,i}^v = \left( \frac{1}{N} \sum_{j \neq i} d_{ij,N} \right)^{-1} \left( \frac{1}{N} \sum_{j \neq i} d_{ij,N} s_{0,j}^v \right). \]

Consider the numerator. Notice by definition that
\[ d_{ij,N} s_{0,j}^v = \mathbb{I}\{g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij}\} (v_j - \mathbb{E}[v_j|a_j]) \]
are i.i.d. across \( j \) conditioning on \((x_{2i}, a_i)\) and bounded by a finite constant under Assumption 11. Then, by the WLLN conditioning on \((x_{2i}, a_i)\), we have
\[
\frac{1}{N} \sum_{j \neq i} d_{ij,N} s_{0,j}^v p \to \mathbb{E}[d_{ij,N}(v_j - \mathbb{E}[v_j|a_j])|x_{2i}, a_i]
= \mathbb{E}[d_{ij,N} \mathbb{E}(v_j - \mathbb{E}[v_j|a_j]|X_N, D_N, a_i)|x_{2i}, a_i]
= 0,
\]
where the last equality holds by Lemma 1. The denominator converges to
\[
\frac{1}{N} \sum_{j \neq i} \mathbb{I}\{g(t(x_{2i}, x_{2j}), a_i, a_j) \geq u_{ij}\} \to_p \int \int \Phi(g(t(x_{2i}, x_2), a_i, a)) \pi(x_2, a) dx_2 da > 0,
\]
where the last inequality holds under Assumption 11.

This shows that as \( N \to \infty \)
\[
\mathbb{E} \left[ \frac{1}{N} \sum_{j \neq i} g_{ij,N} s_{0,j}^v \right] p \to 0 =: \tilde{s}_{1,i}^v
\]
for each \( i \).

Then, using similar argument in Step 2 of the proof of Lemma 14 we deduce
\[
\sup_{1 \leq i \leq N} \left| \frac{1}{N} \sum_{j \neq i} g_{ij,N} s_{0,j}^v \right| = o_p(1).
\]

Also, for \( m = 2, \ldots \), we follow the same mathematical induction argument in Steps 3 and 4 of the proof of Lemma 14 and deduce that
\[
\sup_{1 \leq i \leq N} \left| \frac{1}{N} \sum_{j \neq i} g_{ij,N} s_{m,N,j}^v \right| = o_p(1).
\]

\[ \square \]

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

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

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

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

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

**Proof**

**Part (a).**

Notice from Assumption 11 that \( |\beta_1^0| < 1 \) and \( s_{m,N,i}, \tilde{s}_m(x_i, a_i), \mathbb{E}[s_{m,N,i}|a_i], \mathbb{E}[\tilde{s}_m(x_i, a_i)|a_i] \) 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

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

Notice that

\[
\sup_{1 \leq i \leq N} \left| \sum_{m=0}^{\infty} (\beta_1^0)^m (s_{m,N,i} - \tilde{s}_m(x_i, a_i)) \right| \leq \sup_{1 \leq i \leq N} \left| \sum_{m=0}^{m^*} (\beta_1^0)^m (s_{m,N,i} - \tilde{s}_m(x_i, a_i)) \right| + \epsilon
\]

\[= m^* \sup_{1 \leq i \leq N} |s_{m,N,i} - \tilde{s}_m(x_i, a_i)| + \epsilon = o_p(1) + \epsilon,
\]

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

**Parts (b) and (c).**

Under Assumption 11 \( \mathbb{E}[s_{m,N,i}|a_i], \mathbb{E}[\tilde{s}_m(x_i, a_i)|a_i] \), and \( \eta_i^v = v_i - h^v(a_i) \) 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 \( \tilde{\beta}_{2SLS} \)**

**S.2.1. Limiting distribution of \( \tilde{\beta}_{2SLS} \).** Recall the definition that for any variable \( b_i^l \) being an element of \( (y_i, w_i, s_i) \) and \( v_i, \)

\[
\eta_{si}^l := b_i^l - h_{s_i}(x_{2i}, a_i) = b_i^l - h_{s_i}(x_{2i}, \text{deg}_i), \quad \eta_{si}^v := v_i - h_{s_i}(x_{2i}, a_i) v_i - h_{s_i}(x_{2i}, \text{deg}_i).
\]

Let \( \eta_{sN}^v = (\eta_{s1}^v, \ldots, \eta_{sN}^v)' \).

Outline:
Step 1 Show that
\[
\sqrt{N} (\hat{\beta}_{2SLS} - \beta^0) = \left( W_N^t M_{RN} Z_N (Z_N^t M_{RN} Z_N)^{-1} Z_N^t M_{RN} W_N \right)^{-1} \times W_N^t M_{RN} Z_N (Z_N^t M_{RN} Z_N)^{-1} Z_N^t M_{RN} \eta^v_{iN} + o_p(1). \] (S.2.1.1)

Step 2 Show
\[
\frac{1}{N} \sum_{i=1}^{N} \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right) \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right)' + o_p(1)
\]
and
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right) \eta^v_{i1} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right) \eta^v_{i1} + o_p(1).
\]

Step 3 Derive the limits of
\[
\frac{1}{N} \sum_{i=1}^{N} \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right) \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right)' = \frac{1}{N} \sum_{i=1}^{N} \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right) \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right)'
\]
and
\[
\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right) \eta^v_{i1} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \left( b_i^t - h_{ss}^t(x_{2i}, \text{deg}_i) \right) \eta^v_{i1}
\]

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


(a) \( \frac{1}{N} (Z_N^t P_{RN} W_N - Z_N^t P_{RN} W_N) = o_p(1) \).
(b) \( \frac{1}{N} (Z_N^t P_{RN} Z_N - Z_N^t P_{RN} Z_N) = o_p(1) \).
(c) \( \frac{1}{\sqrt{N}} (Z_N^t P_{RN} \eta^v_{iN} - Z_N^t P_{RN} \eta^v_{iN}) = o_p(1) \).
(d) \( \frac{1}{\sqrt{N}} (Z_N^t M_{RN} (H(a_N - R_N^5)) = o_p(1) \).

*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** \( r^K(x_{2i}, \text{deg}_i) \).

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

(a) \( \frac{1}{N} \sum_{i=1}^{N} (w_i - \hat{h}_{ss}^w(x_{2i}, \text{deg}_i))(z_i - \hat{h}_{ss}^z(x_{2i}, \text{deg}_i))'(z_i - \hat{h}_{ss}^z(x_{2i}, \text{deg}_i))(z_i - \hat{h}_{ss}^z(x_{2i}, \text{deg}_i))' + o_p(1), \)
(b) \( \frac{1}{N} \sum_{i=1}^{N} (z_i - \hat{h}_{ss}^z(x_{2i}, \text{deg}_i))(z_i - \hat{h}_{ss}^z(x_{2i}, \text{deg}_i))' = \frac{1}{N} \sum_{i=1}^{N} (z_i - h_{ss}^z(x_{2i}, \text{deg}_i))(z_i - h_{ss}^z(x_{2i}, \text{deg}_i))' + o_p(1), \)
(c) \( \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - \hat{h}_{ss}^z(x_{2i}, \text{deg}_i))\eta^v_{i1} = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} (z_i - h_{ss}^z(x_{2i}, \text{deg}_i))\eta^v_{i1} + o_p(1). \)

Then the proofs are analogous to the proofs presented in Section S.1.2 and we omit them.
Lemma 21. Under Assumption 1, 3, and 11, where the limit variance \( \bar{S} \)

\[
\bar{S} = \begin{pmatrix}
S_{G_y,1} & S_{G_y,G_1} & S_{G_y,G_2,1} \\
S_{G_1,1} & S_{G_1,G_1} & S_{G_1,G_2,1} \\
S_{G_2,1} & S_{G_2,G_1} & S_{G_2,G_2,1}
\end{pmatrix}
\]

we have

\[
1 \over N \sum_{i=1}^{N} (w_i - h_w^*(x_{2i}, a_i))(z_i - h_w^*(x_{2i}, a_i))' \xrightarrow{p} \bar{S}^{wz},
\]

and

\[
1 \over N \sum_{i=1}^{N} (z_i - h_w^*(x_{2i}, a_i))(z_i - h_w^*(x_{2i}, a_i))' \xrightarrow{p} \bar{S}^{zz},
\]

where

\[
\begin{align*}
\bar{S}_{G_y,G_1} & = E \left[ \sum_{m=0}^{\infty} \beta_2^m \bar{s}_{x_{2m}}(x_i, a_i) + \beta_3^m \bar{s}_{x_{2m+1}}(x_i, a_i) + \bar{s}_{x_{2m}}(x_i, a_i) \right] \left( \bar{s}_{x_{2m}}(x_i, a_i) \right)', \\
\bar{S}_{G_1,G_1} & = E \left[ \bar{s}_{x_{2m}}(x_i, a_i) \right] \left( \bar{s}_{x_{2m}}(x_i, a_i) \right)', \\
\bar{S}_{G_2,G_2,G_1} & = E \left[ \bar{s}_{x_{2m}}(x_i, a_i) \right] \left( \bar{s}_{x_{2m}}(x_i, a_i) \right)', \\
\end{align*}
\]

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

\[
1 \over N \sum_{i=1}^{N} (w_i - h_w^*(x_{2i}, a_i))(z_i - h_w^*(x_{2i}, a_i))' \xrightarrow{p} \bar{S}^{wz},
\]

and

\[
1 \over N \sum_{i=1}^{N} (z_i - h_w^*(x_{2i}, a_i))(z_i - h_w^*(x_{2i}, a_i))' \xrightarrow{p} \bar{S}^{zz},
\]

where \( \bar{s}_{x_{2m}}(x_i, a_i) \) and \( \bar{s}_{x_{2m}}(x_i, a_i) \) are defined recursively as in (S.1.4).

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

\[
1 \over N \sum_{i=1}^{N} (z_i - h_w^*(x_{2i}, a_i))(z_i - h_w^*(x_{2i}, a_i))' \sigma^2(x_i, a_i) \xrightarrow{p} \bar{S}^{zz},
\]

where the limit variance \( \bar{S}^{zz} \) is defined in Lemma 21.

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

\[
1 \over \sqrt{N} \sum_{i=1}^{N} (z_i - h_w^*(x_{2i}, a_i)) \eta_i^v \Rightarrow N(0, \bar{S}^{zz}),
\]

where

\[
\bar{S}^{zz} = \begin{pmatrix}
\bar{S}_{x_1,x_1} & \bar{S}_{x_1,G_1} & \bar{S}_{x_1,G_2,1} \\
\bar{S}_{G_1,x_1} & \bar{S}_{G_1,G_1} & \bar{S}_{G_1,G_2,1} \\
\bar{S}_{G_2,x_1} & \bar{S}_{G_2,G_1} & \bar{S}_{G_2,G_2,1}
\end{pmatrix}
\]

and

\[
\bar{S}_{G_2,x_1} = E \left[ \bar{s}_{x_{2m}}(x_i, a_i) \right] \left( \bar{s}_{x_{2m}}(x_i, a_i) \right)',
\]

where

\[
\begin{align*}
\bar{s}_{x_{2m}}(x_i, a_i) & = \bar{s}_{x_{2m}}(x_i, a_i) - E[\bar{s}_{x_{2m}}(x_i, a_i)|x_{2i}, a_i] \quad \text{with} \quad \bar{s}_{x_{2m}}(x_i, a_i) = x_{1i} \\
\sigma^2(x_i, a_i) & = E[\eta_i^v]^2|x_i, a_i| = E[(v_i - E[v_i|x_{2i}, a_i])^2|x_i, a_i].
\end{align*}
\]
where $\tilde{s}_{s_m}^{x_i}(x_i, a_i)$ is defined recursively as in (S.1.3.4).