A Semiparametric Network Formation Model with Unobserved Linear Heterogeneity

Luis E. Candelaria*

February 1, 2024

Abstract

This paper analyzes a semiparametric model of network formation in the presence of unobserved agent-specific heterogeneity. Its objective is to identify and estimate the preference parameters associated with observed homophily when the distribution of the unobserved factors is not parametrically specified. This paper offers two main contributions to the literature on network formation. First, it establishes a new point identification result for the vector of parameters that relies on the existence of a special regressor. The identification proof is constructive and characterizes a closed form for the parameter of interest. Second, it introduces a two-step semiparametric estimator with a first-step kernel estimator. This estimator is consistent and has a limiting normal distribution under sparse network asymptotics. Monte Carlo experiments demonstrate that the estimator performs well in finite samples. Finally, the methodology is implemented to estimate the homophily parameters in a friendship network using the Add Health dataset.

JEL classification: C14; C25; C31; D85.

Keywords: Network formation; Special regressor; Unobserved heterogeneity; Semiparametric estimation.

*Department of Economics, University of Warwick, U.K. Email: l.candelaria@warwick.ac.uk.

I am deeply grateful to Federico Bugni, Shakeeb Khan, Matthew Masten, and Arnaud Maurel for their excellent guidance, constant encouragement, and helpful discussions. I also thank Áureo de Paula, Andreas Dzemski, Koen Jochmans, Bryan Graham, Bo Honoré, Arthur Lewbel, Thierry Magnac, James Powell, Adam Rosen, Martin Weidner, and seminar participants at different institutions for their comments. This research uses data from Add Health. See references for full citation.

1 Introduction

Individuals tend to connect with other individuals with whom they share similar observed attributes. This observation is known as homophily and is one of the main objects of study in the literature on social networks (McPherson et al. 2001). However, few have investigated the role of homophily when individuals also have preferences for unobserved attributes. Proper policy evaluation requires us to distinguish between the contributions of observed and unobserved attributes since they have different policy implications. For example, high-school students might form friendships based on similarities in their observed socioeconomic attributes just as well as on their preferences for unobserved social skills, such as extraversion. While socioeconomic attributes can be influenced by a given policy intervention, preferences for unobserved social skills are harder to change via targeted policies. In this paper, I study the identification and estimation of the preference parameters associated with the observed attributes in a model of network formation that accounts for valuations on unobserved agent-specific heterogeneity. In particular, I develop identification and estimation strategies that do not depend on distributional assumptions of the unobserved random components.

I consider a semiparametric model of network formation with unobserved agent-specific heterogeneity. Specifically, two distinct agents, i and j, form an undirected link according to the following network formation equation

$$D_{ij} = \mathbf{1}\left[v_{ij} + X_{ij}^{\top}\theta_0 + A_i + A_j - U_{ij} \geq 0\right], \tag{1}$$

where $\mathbf{1}[\cdot]$ is the indicator function, D_{ij} is a binary outcome variable that takes a value of 1 if agents i and j form a link and 0 otherwise, v_{ij} and X_{ij} are pair-specific observed attributes, $\theta_0 \in \Re^{d_\theta}$ is a vector of unknown parameters, A_i and A_j are agent-specific unobserved random variables, and U_{ij} is a link-specific unobserved disturbance term.

Intuitively, Eq. (1) says that an undirected link between two agents is formed if the net benefit of the link between agents i and j is non-negative. The components in Eq. (1) can be classified into three different categories. The first class, given by the linear index $v_{ij} + X_{ij}^{\top}\theta_0$, captures agents' preferences for establishing a link based on observed characteristics. For instance, this component is known as homophily on observed attributes when it captures assortative matching based on sharing similar traits. The second class, formed by the factors A_i and A_j, captures preferences for establishing connections based on agent-specific unobserved attributes. These factors account for unobserved heterogeneity across the individuals' decisions and challenge the identification of θ_0 because of their correlation with the observed attributes. Finally, the third category comprises a link-specific disturbance term U_{ij} that captures the exogenous factors influencing the decision to form a specific link. The components in the last two categories are known to the agents but unobserved to the researcher.

This paper offers two main contributions to the literature on network formation. First, it proposes a new point identification strategy to recover the vector of coefficients in a semiparametric network formation model with unobserved agent-specific factors. The point identification result relies on the presence of a special regressor. A special regressor is an observed covariate that satisfies two key properties: (i) it is conditionally independent of the unobserved attributes, and (ii) it is a continuous variable with a large support. The methodology of the special regressor has been employed in a wide variety of micro-econometric models, but to the best of my knowledge, this paper represents the first generalization of the special regressor to the analysis of a network formation model (Lewbel 1998, and Lewbel 2014 for a survey on the special regressor). Notably, the identification strategy developed in this paper is constructive and yields a closed form for the parameters of interest. In Section 3, I describe in detail the defining properties of the special regressor and provide sufficient conditions to point identify the vector of coefficients.

As a second contribution, this paper introduces a two-step semiparametric estimator for the vector of coefficients θ_0. This estimator has a least-squares closed form and is computationally tractable even in large networks. The estimator uses a first-step kernel estimator to inversely weight the linking decisions D_{ij} by the conditional density of the special regressor. The second step is a linear regression of the pairwise variation of the inversely weighted linking decisions on the pairwise variation of the observed attributes across all the distinct groups of four agents in the network, also known as tetrads. In Section 4, I provide sufficient conditions that ensure the estimator is consistent and has a limiting normal distribution under sparse network asymptotics. I perform inference in a setting where only one large network is observed in the data. In Section 5, I use Monte Carlo simulations to show that the method performs well in finite samples and under different degrees of sparsity.

Finally, in Section 6, I illustrate the performance of this methodology in an empirical application. I use the National Longitudinal Study of Adolescent Health (Add Health) to study the factors driving the formation of a friendship network among high-school students (Harris et al. 2009). As a special regressor, I consider students' birth weight, in deviation from the sample mean. Based on the established relationship between birth weight and individuals' personality traits (Almond et al. 2018), I discuss in detail why, conditional on a rich set of covariates, birth weight is likely to represent a valid special regressor.[1] I estimate the vector of preference parameters and find evidence for homophily on several socio-demographic characteristics and risky behaviors. The results are intuitive and expand on the previous findings described by the literature on the formation of friendship networks by estimating a dyadic network model with agent-specific heterogeneity (cf. Goldsmith-Pinkham and Imbens 2013; Miyauchi 2016; Christakis et al. 2020).

In the rest of this section, I relate my results to the existing literature. This paper is closely related to the literature that studies dyadic network formation models with unobserved heterogeneity (see, e.g., Graham 2017, and Graham 2020 for an additional survey).[2] Within this literature, the

[1] Specifically, the estimation method controls for a wide range of socio-demographic, economic, health, and educational factors of both the students and their parents.

[2] Another branch of the literature on network formation specifies the network as the equilibrium outcome of a

papers by Charbonneau (2017); Jochmans (2017, 2018); Dzemski (2019); Yan et al. (2019) have analyzed the formation of a directed network. Their methodologies differ substantially from the one proposed here since they follow a parametric conditional likelihood approach. In contrast, I study the formation of an undirected network and follow a semiparametric approach.[3]

This paper builds on the seminal work of Graham (2017), which aims to detect preferences for homophily in an undirected network model with agent heterogeneity. Graham (2017) introduced the Tetrad Logit estimator with identification and asymptotic properties that rely on U_{ij} following a logistic distribution. The point identification and estimation results presented here relax this requirement and can be applied to settings where the distribution of U_{ij} is not parametrically specified. Hence, this paper provides a feasible alternative to the Tetrad Logit when the assumption on the logistic distribution is unlikely to hold.

Since the initial draft of this paper was circulated, recent studies have appeared analyzing semiparametric or nonparametric variations of a dyadic network formation model with unobserved heterogeneity; these include the studies by Toth (2017); Gao (2020); and Zeleneev (2020). Similarly to this paper, Toth (2017) studies the identification of a dyadic network formation model in which the distribution of U_{ij} is unknown. However, he uses a different strategy than the one developed in this paper. In particular, he implements an identification strategy similar to the maximum rank by Han (1987). His methodology assumes that each component in the vector of observed attributes X_i and A_i are continuously distributed. This restriction is not required by the method developed in this paper. Relaxing this restriction is empirically relevant as it is common to control for continuous and discrete attributes when studying social networks.

Gao (2020) studies the identification of a dyadic network model with an unknown and strictly increasing cumulative distribution for U_{ij}.[4] He introduces an in-fill and out-expansion that allows him to identify the homophily function. His identification strategy relies on normalizing the location of two different quantiles of the distribution of U_{ij} and using the strictly increasing property of this distribution. In contrast, the methodology presented in this paper relies on an alternative normalization that sets the conditional mean of U_{ij} equal to zero and does not restrict this distribution to be strictly increasing.

Finally, Zeleneev (2020) studies the identification and estimation of a dyadic network formation model with a nonparametric structure of the unobserved heterogeneity. His identification strategy generalizes Auerbach (2022) approach to introduce a pseudo-distance between a pair of agents i and j, which is instrumental in recovering groups of agents with the same levels of agent-specific unobserved heterogeneity. After conditioning on the matched agents with similar unobserved het-

strategic game of link formation (de Paula 2020). The identification and estimation methods used in that literature differ substantially from the ones proposed here.

[3] Related to this literature, Chernozhukov et al. (2020) estimate quantile treatment effects in a two-way fixed effects model when the distribution of U_{ij} is known. Ma et al. (2020) uses a multi-step estimation to detect the latent communities in a dyadic network formation model with logistic error terms.

[4] Gao (2020) considers extensions on the functional form of the unobserved heterogeneity (e.g., Gao 2020 and Zeleneev 2020). Those extensions are beyond the scope of this paper and left for future research.

erogeneity, the residual variation in the observed characteristics is used to identify the vector of coefficients. In this paper, a direct approach is used instead, where the vector of coefficients is identified without relying on recovering first the groups of agents with the same levels of agent-specific heterogeneity.

In contrast to these three studies, the identification strategy proposed here is based on the presence of a special regressor. The vector of parameters is point identified from the information in the pairwise difference of the linking decisions, inversely weighted by the conditional density of the special regressor given the observed attributes. This transformation is not nested in any existing work. Consequently, this result represents a novel contribution to the literature on the formation of networks. The main advantage of the special regressor methodology is that, after inversely weighting the linking decisions, the factors capturing the agent-specific heterogeneity in the network formation equation are partialled out by a pairwise difference strategy. As a result, θ_0 is recovered in closed form without the need to sort the individual types at an initial stage. Moreover, this unique identification strategy yields a two-step semiparametric estimator for θ_0 with a least-squares analytic form and known limiting distribution under sparse network asymptotics. In contrast, Toth (2017) and Zeleneev (2020) recover the vector of parameters at a non-parametric rate due to the initial sorting of the individual indices embedded in their methodologies.[5]

Finally, this paper also contributes to a broad econometrics literature that uses the special regressor method to identify micro-econometric models. Some applications include binary and multinomial choice models, nonlinear panel data models with fixed effects, valuation models, and static games with incomplete information (see, e.g., Lewbel 2014, and references therein). This paper represents the first generalization of the special regressor to network data. The network formation model in Eq. (1) is a nonlinear model with multiple unobserved heterogeneity and dyadic dependence. This setting is not contained in any model of the existing literature. Thus, the methodology proposed in this paper represents a novel contribution to the literature that has used the special regressor in micro-econometric models.

The rest of the paper is organized as follows. Section 2 introduces the setup of the network formation model. Section 3 provides the identification result for the vector of parameters. Section 4 introduces the semiparametric estimator and proves the main asymptotic results. Section 5 reports simulation evidence. Section 6 describes an empirical application, and Section 7 concludes. The appendices collect the proofs, simulation tables, and information on the Add Health dataset.

2 Network Formation Model

A network is an ordered pair $(\mathbf{N}_n, \mathbf{D}_n)$ formed by a set of n agents denoted by $\mathbf{N}_n = \{1, \cdots, n\}$ and an $n \times n$ adjacency matrix $\mathbf{D}_n$, which represents the links between the agents in $\mathbf{N}_n$. Let D_{ij} denote the (i, j)th entry of the matrix $\mathbf{D}_n$. I assume the network is undirected and unweighted. A network

[5] No inference method is provided for the identification results in Gao (2020).

is undirected if the adjacency matrix is symmetric, i.e., $D_{ij} = D_{ji}$. A network is unweighted if any (i, j)th entry of the adjacency matrix takes one of two values, where these values are normalized to be 0 and 1. In other words, $D_{ij} \in \{0, 1\}$, where $D_{ij} = 1$ if the agents i and j share a link and $D_{ij} = 0$ otherwise. Moreover, I normalize the value of self-ties to zero, i.e., $D_{ii} = 0$ for any agent i.

Each agent $i \in \mathbf{N}_n$ is endowed with a $d_\theta + 1$-dimensional vector of observed attributes $(v_i, X_i^\top)^\top$ and an unobserved scalar component term A_i. The unobserved term U_{ij} captures exogenous stochastic factors that influence the (i, j)th pair-specific decision to establish a link between these agents.

For any distinct agents i and j, let $X_{ij} = g_x(X_{ij})$ be a $d_\theta \times 1$ random vector of pair-specific attributes, where $g_x : \Re^{d_\theta} \times \Re^{d_\theta} \mapsto \Re^{d_\theta}$ is a vector-valued function that is known, nonlinear, and symmetric. Similarly, let $v_{ij} = g_v(v_i, v_j)$ with $g_v : \Re \times \Re \mapsto \Re$. The functions g_x and g_v are assumed to be symmetric on their terms due to the undirected nature of the network. The specification of g_x and g_v are chosen by the researcher and vary according to the empirical application. For instance, these functions can be specified to capture preferences for assortative matching. Suppose that X_i represents agent i's gender, then $X_{ij} = \mathbf{1}[X_i = X_j]$ accounts for homophily on gender.

Using this notation, the formation of an undirected link between two distinct agents i and j in $\mathbf{N}_n$ is represented by Eq. (1). The coefficient associated with v_{ij} has been normalized to 1. Scale normalizations are standard in the binary choice literature and are necessary for point identification when the distribution of the error term is not parametrized (Powell 1994). The main parameter of interest is θ_0.[6]

2.1 Notation

For any pair of agents $i, j \in \mathbf{N}_n$, let $Z_i = \{v_i, X_i, A_i\}$ and $Z_{ij} = \{v_i, v_j, X_i, X_j, A_i, A_j\}$. The sequence of pair-specific observed attributes is denoted by $\mathbf{X}_n = \{X_{ij} : i, j \in \mathbf{N}_n\}$. Similarly, let $\mathbf{Z}_n = \{Z_{ij} : i, j \in \mathbf{N}_n\}$ denote the sequence of observed and unobserved attributes for all individuals in the network. Additionally, let $\mathbf{X}_{-ij} = \{X_{kl} : k, l \neq i, j\}$ and $\mathbf{Z}_{-ij} = \{Z_{kl} : k, l \neq i, j\}$.

The identification and estimation strategies introduced in Sections 3 and 4 use the information contained in subnetworks formed by the 4-tuples $\{i, j, k, l\}$ of individuals, known as tetrads. The following notation describes variables at the tetrad level. Given a network of size n, there is a total of $m_n = \binom{n}{4}$ tetrads with distinct indices $i, j, k, l \in \mathbf{N}_n$. Let σ be a function that maps these tetrads to the index set $\mathbf{N}_{m_n} = \{1, \cdots, m_n\}$. Thus, each tetrad with distinct indices $\{i, j, k, l\}$ corresponds to a unique value $\sigma(\{i, j, k, l\}) \in \mathbf{N}_{m_n}$. I will denote the subnetwork formed by the tetrad $\{i, j, k, l\}$ by $\sigma(\{i, j, k, l\}) \in \mathbf{N}_{m_n}$ (cf. Jochmans 2018).

[6] In Appendix C, I discuss how this network formation model can be derived as a stable outcome from a static game with transferable utilities.

3 Identification

This section introduces the identification result for the semiparametric network formation model with unobserved agent-specific factors. This result relies on the presence of a special regressor. The following assumptions specify the underlying framework that is used to show point identification of θ_0 in the network formation model.

Assumption 3.1 (Sampling)**.** *(i) $\{Z_i\}_{i=1}^n$ is i.i.d. across $i \in \mathbf{N}_n$. (ii) $\{U_{ij} \mid \mathbf{Z}_n\}_{i \neq j}$ is i.i.d. across $i, j \in \mathbf{N}_n$ (iii) For any $i, j \in \mathbf{N}_n$, $U_{ij} \perp\!\!\!\perp \mathbf{Z}_{-ij} \mid Z_{ij}$. (iv) $v_{ij} \perp \mathbf{X}_{-ij} \mid X_{ij}$.*

Assumption 3.1 describes the sampling process. Condition (i) specifies that individuals are drawn independently from an identical distribution. This condition is widely used to describe network data (see, e.g., Graham 2017; Jochmans 2018; and Auerbach 2022). Condition (ii) states that, conditional on $\mathbf{Z}_n$, the link-specific disturbance terms $\{U_{ij}\}_{i \neq j}$ are independent across dyads $\{i, j\}$ and drawn from the same distribution. Furthermore, Condition (iii) requires that conditional on the dyad-specific attributes Z_{ij}, the link-specific disturbance term U_{ij} is independent of any observed or unobserved feature in $\mathbf{Z}_{-ij}$. Condition (iv) states that the dyad-level component X_{ij} controls for any dependence between v_{ij} and the vector of observed attributes $\mathbf{X}_n$. This condition is design-specific and holds under different empirically interesting designs, for example, when both $v_{ij} = g_v(v_i, v_j)$ and $X_{ij} = g_x(X_i, X_j)$ account for assortative matching. Assumption 3.1 ensures that each of the linking decisions is conditionally independent across dyads.

Notice that Assumption 3.1 allows for heteroskedasticity of a general form in the distribution of U_{ij}. Moreover, it allows for flexible dependence between the unobserved agent-specific factors and the observed attributes. In other words, Assumption 3.1 does not restrict the joint distribution of Z_{ij}. Assumption 3.1 is commonly used in semiparametric nonlinear panel data models, for example in Arellano and Honoré (2001). In network formation models, full stochastic independence $U_{ij} \perp \mathbf{Z}_n, \mathbf{A}_n$ is usually imposed as in Leung (2015); Graham (2017); Toth (2017); and Gao (2020). Arbitrary heteroskedasticity is also considered in Zeleneev (2020).

Assumption 3.2 (Exclusion)**.** *For any distinct $i, j \in \mathbf{N}_n$, let $e_{ij} = A_i + A_j - U_{ij}$. Denote by $F_{e|X}(e_{ij} \mid \mathbf{X}_n)$ and $\mathbb{S}_{e|X}(e_{ij} \mid \mathbf{X}_n)$ the conditional CDF and support of e_{ij} given $\mathbf{X}_n$. (i) $\mathbb{E}[e_{ij} \mid \mathbf{X}_n] < \infty$. (ii) $e_{ij} \perp v_{ij} \mid X_{ij}$.*

Assumption 3.2 represents an exclusion restriction, and it entails that the regressor v_{ij} is conditionally independent of e_{ij} given the observed attributes X_{ij}. In other words, after controlling for X_{ij}, the regressor v_{ij} is statistically independent from the unobserved attributes in e_{ij}. This is one of the main defining properties of the special regressor in the sense of Lewbel (1998, 2000).

Assumption 3.3 (Large Support)**.** *For any distinct $i, j \in \mathbf{N}_n$, the following holds: (i) The conditional distribution of v_{ij} given X_{ij} is absolutely continuous with PDF given by $f_{v|X}(v_{ij} \mid X_{ij})$ and support given by $\mathbb{S}_{v|X}(X_{ij}) = [\underline{v}(X_{ij}), \overline{v}(X_{ij})]$, with $-\infty \leq \underline{v}(X_{ij}) \leq 0 \leq \overline{v}(X_{ij}) \leq \infty$. (ii)*

The density $f_{v|x}(v_{ij} \mid X_{ij})$ *is bounded away from zero. (iii) Conditional on* X_{ij}*, the support of* $-\left(X_{ij}^{\top}\theta_0 + e_{ij}\right)$ *is a subset of* $\mathbb{S}_{v|X}(X_{ij})$*.*

Assumption 3.3 is a large support condition, and it ensures that, conditional on X_{ij}, v_{ij} has a density function $f_{v|X}(v_{ij} \mid X_{ij})$ on $\mathbb{S}_{v|X}(X_{ij})$. Furthermore, it requires that, given X_{ij}, the support of $-(X_{ij}'\theta_0 + e_{ij})$ is contained in $\mathbb{S}_{v|X}(X_{ij})$. Notice that Assumption 3.3 does not require $v_{ij} \mid X_{ij}$ to have full support on the real line. Hence, the point identification result introduced in this section is general enough to include both: the full support case and the bounded support case, as long as Condition (ii) holds. Moreover, observe that Assumption 3.3 leaves unrestricted the distribution of the observed attributes X_{ij}. Hence, it is possible to control for discrete and continuous covariates in X_{ij}, which is desirable in an empirical application. Together, Assumptions 3.2 and 3.3 imply that v_{ij} is a special regressor.

The network formation model specified by Eq. (1) and Assumptions 3.1-3.3 represent, to the best of my knowledge, the first generalization of the special regressor to analyze network data.

The following theorem formalizes the point identification result for θ_0. For any $i, j \in \mathbf{N}_n$, consider the next transformation of the pair-specific linking decision D_{ij}:

$$D_{ij}^{*} \equiv \left(\frac{D_{ij} - \mathbf{1}\left[v_{ij} > 0\right]}{f_{v|x}(v_{ij} \mid X_{ij})} \right). \tag{2}$$

This transformation is instrumental in shifting the scale and location of the conditional expectation $\mathbb{E}\left[D_{ij}^{*} \mid \mathbf{X}_n\right]$ as described in Lemma 1. Also, for any tetrad $\sigma(\{i, j, k, l\}) \in \mathbf{N}_{m_n}$, let

$$\begin{aligned} G_{\sigma}^{*} &\equiv (D_{ik}^{*} - D_{il}^{*}) - (D_{jk}^{*} - D_{jl}^{*}) \\ W_{\sigma} &\equiv (X_{ik} - X_{il}) - (X_{jk}^{*} - D_{jl}^{*}). \end{aligned}$$

The variable G_{σ}^{*} denotes the pairwise variation of $\left\{D_{ik}^{*}, D_{il}^{*}, D_{jk}^{*}, D_{jl}^{*}\right\}$, which represent the inversely-weighted linking decisions. While, W_{σ} denotes the pairwise variation of $\{X_{ik}, X_{il}, X_{jk}, X_{jl}\}$.

The identification and estimation of θ_0 will make use of those contributing tetrads $\sigma \in \mathbf{N}_{m_n}$ for which $G_{\sigma}^{*} \neq 0$. To this end, for any $\sigma(\{i, j, k, l\}) \in \mathbf{N}_{m_n}$, let $\omega_{\sigma} \equiv \mathbf{1}\left[G_{\sigma}^{*} \neq 0\right]$. The total number of contributing tetrads is given by

$$m_n^{*} \equiv \sum_{\sigma \in \mathbf{N}_{m_n}} \omega_{\sigma},$$

and the expected share of contributing tetrads is

$$\rho_n \equiv \frac{1}{m_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \Pr\left[\omega_{\sigma} = 1\right]. \tag{3}$$

Assumption 3.4. *For any distinct* $i, j \in \mathbf{N}_n$*, the following holds: (i)* $\mathbb{E}\left[U_{ij} \mid \mathbf{X}_n\right] = 0$*. (ii) The* *matrix*

$$\Gamma_0 \equiv \lim_{n\to\infty} \frac{1}{\mathbb{E}\left[m_n^*\right]} \sum_{\sigma\in\mathbf{N}_{m_n}} \mathbb{E}\left[W_\sigma W_\sigma^\top\right]$$

is finite and nonsingular.

Condition (i) of Assumption 3.4 normalizes the conditional mean of the link-specific disturbance term U_{ij} to be equal zero given $\mathbf{X}_n$. Condition (ii) of Assumption 3.4 is a full rank condition on the pairwise variation of the observed attributes $\mathbb{E}\left[W_\sigma W_\sigma^\top\right]$, which is scaled by the expected share of contributing tetrads. This condition ensures that θ_0 is point identified.

Lemma 1. *If Assumptions 3.1-3.4 hold in Eq.* (1)*, then for any distinct $i, j \in \mathbf{N}_n$*

$$\mathbb{E}\left[D_{ij}^* \mid \mathbf{X}_n\right] = X_{ij}^\top\theta_0 + \mathbb{E}\left[A_i + A_j \mid \mathbf{X}_n\right].$$

Proof. See Appendix A.1. □

Lemma 1 conveys two main insights. First, it shows that the conditional expectation of the inversely-weighted linking decision D_{ij}^* given $\mathbf{X}_n$ is a linear function of the unobserved term $\mathbb{E}[A_i + A_j \mid \mathbf{X}_n]$ and the linear index term $X_{ij}^\top\theta_0$. Second, the within-individual i difference of the transformed linking decisions follows a partially linear structure (cf. Robinson 1988). That is, $\mathbb{E}[D_{ik}^* - D_{il}^* \mid \mathbf{X}_n] = (X_{ik} - X_{il})^\top\theta_0 + \mathbb{E}\left[A_k - A_l \mid \mathbf{X}_n\right]$ for any $i \in \mathbf{N}_n$ with $i \neq k, l$. The component $\mathbb{E}\left[A_k - A_l \mid \mathbf{X}_n\right]$ represents a nuisance parameter that is a common element across the within-individual i variations with $i \neq k, l$. Hence, point identification of θ_0 will follow from the pairwise variation of the weighted linking decisions G_σ^*, which will differentiate out the nuisance parameter.

Theorem 3.1. *If Assumptions 3.1-3.4 hold in Eq.* (1)*, then for any $\sigma \in \mathbf{N}_{m_n}$*

$$\mathbb{E}\left[W_\sigma G_\sigma^*\right] = \mathbb{E}\left[W_\sigma W_\sigma^\top\right]\theta_0,$$

and thus

$$\theta_0 = \Gamma_0^{-1} \times \Psi_0, \tag{4}$$

where

$$\Psi_0 \equiv \lim_{n\to\infty} \frac{1}{\mathbb{E}\left[m_n^*\right]} \sum_{\sigma\in\mathbf{N}_{m_n}} \mathbb{E}\left[W_\sigma G_\sigma^*\right].$$

Proof. See Appendix A.1. □

Theorem 3.1 demonstrates that θ_0 is point identified using the information contained in the joint distribution of the tetrads $\{G^*_\sigma, W_\sigma\}_{\sigma \in \mathbf{N}_{m_n}}$. Moreover, the identification of θ_0 is constructive and yields a least-squares analytic form for the parameter of interest. This least-squares structure will be the foundation for the semiparametric estimator introduced in Section 4.

Remark 1 (Partially-linear regression)**.** *Although the within-individual i variation follows a partially linear structure, note that the identification of θ_0 does not require additional exclusion restrictions, as it is often imposed in partially linear models (cf. Robinson 1988). In contrast, the full rank condition in Assumption 3.4 is sufficient for the identification result.*

Remark 2 (Individual-specific attributes)**.** *Unlike the case in panel data models with fixed effects, notice that this strategy can account for individual-specific attributes in X_i. Thus, this methodology overcomes the impossibility of estimating time-invariant characteristics in panel models with fixed effects (cf. Honoré and Lewbel 2002). This result is a consequence of defining the dyad-specific attributes X_{ij} as a nonlinear function of (X_i, X_j) and represents an empirically relevant property since when studying social networks it is common to control for individual-specific attributes, such as gender and ethnicity.*

Finally, given the results in Lemma 1 and Theorem 3.1, notice that the average contribution of the unobserved agent-specific factors to forming a link is identified.

Corollary 1. *If Assumptions 3.1-3.4 hold in Eq. (1), then $\mathbb{E}\left[A_i + A_j\right] = \mathbb{E}\left[D^*_{ij}\right] - \mathbb{E}\left[X_{ij}\right]^\top \theta_0$ for any i and j in $\mathbf{N}_n$.*

4 Inference

This section introduces a semiparametric estimator for θ_0 based on the point identification result. The estimator for θ_0, denoted by $\widehat{\theta}_n$, is a two-step estimator with a nonparametric estimate of the conditional distribution $f_{v|x}(v_{ij} \mid X_{ij})$. Below, I discuss sufficient conditions to derive the large sample properties of $\widehat{\theta}_n$. Theorem 4.1 proves that $\widehat{\theta}_n$ is a consistent estimator for θ_0. Theorem 4.2 shows that the limiting distribution of $\widehat{\theta}_n$ is normal.

The estimator for θ_0 is defined as the sample analogue of Eq. (4) and represents the regression coefficient in the regression of G^*_σ on W_σ across all distinct tetrads $\sigma \in \mathbf{N}_{m_n}$. Given that the inverse of $f_{v|x}(v_{ij} \mid X_{ij})$ is used as a weight in the definition of Ψ_0, a trimming sequence is used to avoid boundary effects due to the first-step estimation of $f_{v|x}(v_{ij} \mid X_{ij})$. Let $I_\tau(v_{ij} \mid X_{ij})$ denote this trimming sequence with trimming parameter given by τ_n.

Recall that G^*_σ is defined as the pairwise variation across the inversely-weighted linking decisions for a given tetrad $\sigma \in \mathbf{N}_{m_n}$. I extend this notation to define the pairwise variation of the trimmed network links given the parameter τ_n as follows:

$$\begin{aligned} G^*_{\sigma,\tau} &\equiv \left(D^*_{ik,\tau} - D^*_{il,\tau}\right) - \left(D^*_{jk,\tau} - D^*_{jl,\tau}\right) \\ \widehat{G}^*_{\sigma,\tau} &\equiv \left(\widehat{D}^*_{ik,\tau} - \widehat{D}^*_{il,\tau}\right) - \left(\widehat{D}^*_{jk,\tau} - \widehat{D}^*_{jl,\tau}\right), \end{aligned}$$

where, for any distinct i and j in $\mathbf{N}_n$

$$\begin{aligned} D^*_{ij,\tau} &\equiv f_{v|x}(v_{ij} \mid X_{ij})^{-1} \left(D_{ij} - \mathbf{1}\left[v_{ij} > 0\right]\right) I_\tau(v_{ij} \mid X_{ij}) \\ \widehat{D}^*_{ij,\tau} &\equiv \widehat{f}_{v|x}(v_{ij} \mid X_{ij})^{-1} \left(D_{ij} - \mathbf{1}\left[v_{ij} > 0\right]\right) I_\tau(v_{ij} \mid X_{ij}). \end{aligned}$$

Here, $\widehat{f}_{v|x}(v_{ij} \mid X_{ij})$ denotes the kernel estimator of the true conditional density function of v_{ij} given X_{ij}, denoted by $f_{v|x}(v_{ij} \mid X_{ij})$. Thus, $G^*_{\sigma,\tau}$ denotes the pairwise variation of the trimmed network links, assuming that the conditional distribution of the special regressor given the observed attributes is known. Conversely, $\widehat{G}^*_{\sigma,\tau}$ denotes the pairwise variation of the trimmed network links when $f_{v|x}(v_{ij} \mid X_{ij})$ is replaced by a first-stage kernel estimator $\widehat{f}_{v|x}(v_{ij} \mid X_{ij})$.

The trimming sequence $I_\tau(v_{ij} \mid X_{ij})$ is a function of the conditional distribution of v_{ij} given X_{ij}, and it converges to 1 as the trimming parameter $\tau_n \to 0$ when $n \to \infty$. Assumption 4.4 describes the conditions imposed on τ_n (see e.g., Honoré and Lewbel 2002 and Khan and Tamer 2010).

To ease the exposition, I will denote $I_{\tau,ij} = I_\tau(v_{ij} \mid X_{ij})$, $f_{vx,ij} = f_{vx}(v_{ij}, X_{ij})$, and $f_{x,ij} = f_x(X_{ij})$. Using this notation, the semiparametric estimator for θ_0 is defined as

$$\widehat{\theta}_n \equiv \widehat{\Gamma}_n^{-1} \times \widehat{\Psi}_{n,\tau}, \tag{5}$$

where

$$\begin{aligned} \widehat{\Gamma}_n &\equiv \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma W_\sigma^\top \\ \widehat{\Psi}_{n,\tau} &\equiv \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma \widehat{G}^*_{\sigma,\tau}. \end{aligned}$$

The first-stage kernel estimator $\widehat{f}_{v|x}(v_{ij} \mid X_{ij})$ is defined as the ratio of the kernel estimators $\widehat{f}_{vx,ij}$ and $\widehat{f}_{x,ij}$ with

$$\widehat{f}_{vx}(v_{12}, x_{12}) = \frac{1}{n(n-1)} \sum_{l_1=1}^{n} \sum_{l_2 \neq l_1} K_{vx,h}\left[v_{l_1 l_2} - v_{12}, X_{l_1 l_2} - x_{12}\right] \tag{6}$$

$$\widehat{f}_{x}(x_{12}) = \frac{1}{n(n-1)} \sum_{l_1=1}^{n} \sum_{l_2 \neq l_1} K_{x,h}\left[X_{l_1 l_2} - x_{12}\right]. \tag{7}$$

The kernels $K_{vx,h}$ and $K_{x,h}$ are defined as

$$K_{vx,h}\left[v_{l_1l_2}-v_{12},X_{l_1l_2}-x_{12}\right]\equiv h_n^{d_\theta+1}\kappa_{vx}\left[\frac{v_{l_1l_2}-v_{12}}{h_n},\frac{X_{l_1l_2}-x_{12}}{h_n}\right]$$
$$K_{x,h}\left[X_{l_1l_2}-x_{12}\right]\equiv h_n^{d_\theta+1}\kappa_{x}\left[\frac{X_{l_1l_2}-x_{12}}{h_n}\right].$$

where h_n denotes a bandwidth parameter. Assumption 4.3 below describes the conditions imposed on the kernel functions κ_{vx} and κ_x, and bandwidth parameter h_n.

The estimator $\widehat{\theta}_n$ in Eq. (5) offers a novel contribution to the literature of dyadic network formation models with agent heterogeneity. In particular, it constitutes the first semiparametric methodology to conduct valid inference on the vector of homophily parameters when a special regressor is available and under different degrees of sparsity in the network. As an appealing property, the estimator $\widehat{\theta}_n$ is computationally simple to calculate as it has a least-squares analytical form with an initial kernel density estimator.

The structure of the estimator is related to the work of Honoré and Lewbel (2002) and Graham (2017); however, the asymptotic theory presented here expands the existing literature. In particular, the theory in Honoré and Lewbel (2002) is intended for short panels, and thus, it is not suitable for network data with dyadic dependence. In contrast, the asymptotic results developed in this paper use higher order U-statistics, Hájek projections and CLTs that are tailored to unpack the conditional independence structure in dyadic data that is described by Assumption 3.1. Conversely, the estimator in Graham (2017) is a conditional maximum likelihood estimator of a logistic dyadic regression, which does not involve the non-parametric estimation of a density function. In contrast, $\widehat{\theta}_n$ is instead a two-step semiparametric estimator with a least-squares analytic form. The asymptotic theory presented here is developed to account for the effect that the non-parametric estimation of $f_{v|x}(v_{ij}\mid X_{ij})$ has on the influence function of $\widehat{\theta}_n$. Moreover, as the density $f_{v|x}(v_{ij}\mid X_{ij})$ is used as an inverse weight, a trimming approach is implemented to derive the asymptotic results.[7]

The following technical conditions are needed to prove Theorems 4.1 and 4.2. For simplicity, the theorems are stated assuming that all of the elements of X_{ij} are continuously distributed. However, the results can be extended to include discretely distributed variables by applying the density estimator separately to each discrete data cell.

Assumption 4.1 (Compact Support). *(i) $\theta_0\in int(\Theta)$, with Θ a compact subset of $\Re^{d_\theta}$. (ii) For any $\sigma\in\mathbf{N}_{m_n}$, the support of W_σ is $\mathbb{W}$, a compact subset of $\Re^{d_\theta}$, and $\mathbb{E}\mid(W_{r,\sigma}W_{s,\sigma})^2\mid<\infty$ for any $r,s=1,\cdots,d_\theta$, where $W_{r,\sigma}$ denotes the rth entry of W_σ.*

Condition (i) of Assumption 4.1 is standard in the context of semiparametric estimation (Powell 1994). Assumption 4.1 ensures that $W_\sigma^\top\theta_0$ has bounded contribution to the pairwise regression of $G^*_{\sigma,\tau}$ on W_σ. This condition is not essential for the main results but simplifies the proof of

[7] The nonparametric estimation of density functions with dyadic data is an important topic that has attracted recent attention, for example, in Graham et al. (2019, 2021).

the asymptotic distribution of the semiparametric estimator. Alternatively, Condition (ii) can be relaxed by requiring that a sufficiently high-order moment of W_σ is finite. Assumption 4.1 has been used in Graham (2017) (cf. Jochmans 2018).

Assumption 4.2 (Density)**.** *For any $i, j \in \mathbf{N}_n$, denote the probability density functions of the covariates X_{ij} and (v_{ij}, X_{ij}) by $f_x(X_{ij})$ and $f_{vx}(v_{ij}, X_{ij})$, with supports given by $\mathbb{S}_x$ and $\mathbb{S}_{vx}$. Suppose the following holds.*

(i) $0 < \underline{B}_x \leq \inf_{x_{12}\in\mathbb{S}_x} f_x(x_{12}) < \sup_{x_{12}\in\mathbb{S}_x} f_x(x_{12}) \leq \overline{B}_x < \infty$ *and*

$$0 < \underline{B}_{vx} \leq \inf_{(v_{12},x_{12})\in\mathbb{S}_{vx}} f_{vx}(v_{12}, x_{12}) < \sup_{(v_{12},x_{12})\in\mathbb{S}_{vx}} f_{vx}(v_{12}, x_{12}) \leq \overline{B}_{vx} < \infty.$$

(ii) For some $\delta > d_\theta + 1$, $f_x(x_{12})$ is δ-times differentiable with differential operator

$$\nabla^\lambda = \frac{\partial^{|\lambda|}}{\partial^{\lambda_1} x_{12,1} \cdots \partial^{\lambda_{d_\theta}} x_{12,d_\theta}}$$

where $\lambda = (\lambda_1, \cdots, \lambda_{d_\theta}) \in \Re^{d_\theta}$ and $\mid \lambda \mid = \lambda_1 + \cdots + \lambda_{d_\theta}$. The derivative $\nabla^\lambda f_x(x_{12})$ is bounded,

$$\max_{0\leq|\lambda|\leq\delta} \sup_{x_{12}\in\mathbb{S}_x} \mid \nabla^\lambda f_x(x_{12}) \mid \leq \overline{D}_x.$$

The same conditions hold for $f_{vx}(v_{ij}, x_{ij})$ with differential operator $\nabla^{\tilde{\lambda}}$ with $\tilde{\lambda} \in \Re^{d_\theta+1}$ and derivatives bounded by $\overline{D}_{vx}$.

Assumption 4.2 ensures that the densities $f_{x,ij}$ and $f_{vx,ij}$ are continuous and δ-times differentiable. This assumption is standard in the literature of density estimation, for example, in Ahn and Powell (1993), Hansen (2008), and Graham et al. (2021).

Assumption 4.3 (Kernel)**.** *The kernel function $\kappa_x(x) : \Re^{d_\theta} \mapsto \Re$ and bandwidth parameter h_n satisfy the following conditions.*

(i) The kernel is symmetric around zero, $\kappa_x(x) = \kappa_x(-x)$.

(ii) $\kappa_x(x) = 0$ for all x outside a convex bounded subset of $\mathbb{S}_x$. This subset has a nonempty interior with 0 as an interior point.

(iii) The kernel is bounded, $\sup_{x\in\mathbb{S}_x} \mid \kappa_x(x) \mid \leq \overline{\kappa}_x < \infty$.

(iv) $\int \kappa_x(x)dx = 1$, *and* $\int \kappa_x(x)^2 dx = \overline{Q}_x < \infty$.

(v) The kernel $\kappa_x(x)$ is bias reducing of order $\delta > d_\theta + 1$, where for $\lambda \in \Re^{d_\theta}$

$$\int x_1^{\lambda_1} \cdots x_{d_\theta}^{\lambda_{d_\theta}} \kappa_x(x)dx = \begin{cases} 0 & \text{if } \mid \lambda \mid < \delta \\ \overline{T}_x & \text{if } \mid \lambda \mid = \delta \end{cases}$$

with $\overline{T}_x < \infty$.

(vi) The kernel $\kappa_x(x)$ is differentiable of order δ with bounded derivatives.

$$\max_{0\leq|\lambda|\leq\delta} \sup_{x\in\mathbb{S}_x} \mid \nabla^{\lambda}\kappa_x(x) \mid\leq \overline{L}_x.$$

(vii) The bandwidth rate satisfies $nh_n^{d_\theta+1} \to \infty$, $nh_n^{2\delta} \to 0$, and $\log n/nh_n^{d_\theta+1} \to 0$ as $n \to \infty$.

The kernel function $\kappa_{v,x}(v,x)$ satisfies all the same properties, replacing the constant values $\overline{\kappa}_{vx}$ in point (iii), $\overline{Q}_{vx}$ in (iv), $\overline{T}_{vx}$ in (v), and $\overline{L}_{vx}$ in (vi).

Conditions (i)-(iv) are standard in kernel estimation (see e.g., Honoré and Lewbel 2002; Hansen 2008; Graham et al. 2019). Conditions (v) and (vi) ensure that κ_x and κ_v are higher-order kernels that are selected to control the bias induced by using the inverse of $f_{v|x}(v_{ij} \mid X_{ij})$ as a weighting function. Condition (vii) imposes rate conditions on the bandwidth parameter h_n, which ensures the consistent estimation of $f_x(X_{ij})$ and $f_{vx}(v_{ij}, X_{ij})$ (Hansen 2008; Graham et al. 2019).

Assumption 4.4 (Trimming Sequence)**.** *(i) For any $i, j \in \mathbf{N}_n$ and $\tau_n \in [0,1]$, the trimming function $I_\tau(v_{ij} \mid X_{ij})$ is equal to zero if v_{ij} is within a distance τ_n of the boundary of the support of $\mathbb{S}_{v|x}(X_{ij})$, and $I_\tau(v_{ij} \mid X_{ij})$ equals one, otherwise. In particular, $\Pr\left[I_\tau(v_{ij} \mid X_{ij}) = 1\right] = 1 - \tau_n$. (ii) The trimming parameter τ_n satisfies the rate condition $\tau_n \to 0$ as $n \to \infty$.*

Due to the inverse weighting used in the definition of $\widehat{D}^*_{ij}$, boundary effects could arise from the density estimation step when computing $\widehat{\Psi}_{n,\tau}$. Assumptions 4.2 and 4.4 deal with this technicality by assuming that $f_{vx,ij}$ is bounded away from zero and by introducing a trimming sequence $I_{\tau,ij}$ that sets to zero the terms in $\widehat{\Psi}_{n,\tau}$ with data within a τ_n distance of the boundary of $\mathbb{S}_{vx}$ (see e.g., Lewbel 1997, 2000; Honoré and Lewbel 2002; and Khan and Tamer 2010). The trimming parameter vanishes asymptotically as $n \to \infty$.

Assumption 4.5 (Bounded Moments)**.** *For any $\sigma(\{i,j,k,l\}) \in \mathbf{N}_{m_n}$*

$$\sup_{x_{i_1i_2}\in\mathbb{S}_x} \mathbb{E}\left[\left(D_{\sigma_{i_1i_2},\tau}\right)^2 \mid x_{i_1i_2}\right] f_x(x_{i_1i_2}) \leq \overline{E}_x < \infty$$

$$\sup_{(v_{i_1i_2},x_{i_1i_2})\in\mathbb{S}_{v,x}} \mathbb{E}\left[\left(D^*_{i_1i_2,\tau}\right)^2 \mid v_{i_1i_2}, x_{i_1i_2}\right] f_{vx}(v_{i_1i_2}, x_{i_1i_2}) \leq \overline{E}_{vx} < \infty$$

where $\sigma_{i_1i_2} \in \{(i,k),(i,l),(j,k),(j,l)\}$.

Assumption 4.5 ensures the existence and boundedness of the conditional expectations defined above. Similar conditions have been used in Ahn and Powell (1993); Aradillas-Lopez (2012); Hansen (2008), and Graham et al. (2021).

4.1 Consistency

Using the assumptions above, the next theorem shows that $\widehat{\theta}_n$ is a consistent estimator of θ_0.

Theorem 4.1. *Let Assumptions 3.1-4.3 hold and $n\rho_n \to \infty$ as $n \to \infty$. Then $(\widehat{\theta}_n - \theta_0) \xrightarrow{\text{p}} \mathbf{0}$ as $n \to \infty$.*

Proof. See Appendix A.3. □

The rate condition $n\rho_n \to \infty$ states that the number of identifying tetrads grows as the sample size grows. In other words, the term $n\rho_n$ represents the effective sample size. The proof of Theorem 4.1 consists of showing that $\widehat{\Gamma}_n \xrightarrow{\text{p}} \Gamma_0$ and $\widehat{\Psi}_{n,\tau} \xrightarrow{\text{p}} \Psi_0$, followed by invoking a Continuous Mapping Theorem and Slutsky's Theorem. The convergence in probability of $\widehat{\Gamma}_n$ follows from a variance calculation. Meanwhile, the convergence in probability of $\widehat{\Psi}_{n,\tau}$ deserves additional attention as it accounts for the trimming effect, as well as the non-parametric first-stage estimation.

Notice that given a trimming parameter $\tau_n > 0$, $\mathbb{E}\left[W_\sigma G^*_{\sigma,\tau}\right]$ is different from $\mathbb{E}\left[W_\sigma G^*_\sigma\right]$. In Appendix A.3, I show that $\widehat{\Psi}_{n,\tau} - \Psi_0 \xrightarrow{\text{p}} 0$ follows from proving that each component of the following decomposition converges in probability to zero,

$$\left\{\frac{1}{m^*_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma \widehat{G}^*_{\sigma,\tau} - \frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau}\right]\right\} + \frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} \left\{\mathbb{E}\left[W_\sigma G^*_{\sigma,\tau}\right] - \mathbb{E}\left[W_\sigma G^*_\sigma\right]\right\}.$$

The first sum accounts for the fact that given the nonzero trimming sequence, the estimator $\widehat{\Psi}_{n,\tau}$ is centered around the trimmed parameter $(m_n\rho_n)^{-1}\sum_{\sigma\in\mathbf{N}_{m_n}}\mathbb{E}\left[W_\sigma G^*_{\sigma,\tau}\right]$ rather than on Ψ_0. Meanwhile, the second sum isolates the effect of trimming on $\mathbb{E}\left[W_\sigma G^*_\sigma\right]$. In Lemma 6 of Appendix A.2, I show that the effect of trimming is negligible, that is

$$\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} \left\{\mathbb{E}\left[W_\sigma G^*_{\sigma,\tau}\right] - \mathbb{E}\left[W_\sigma G^*_\sigma\right]\right\} = o_p(1).$$

Regarding the first sum, in Appendix A.2, I also show that the effect of the first-stage nonparametric estimation into the estimator $\widehat{\Psi}_{n,\tau}$ can be decomposed as follows

$$\widehat{\Psi}_{n,\tau} = \Psi_{n,\tau} + \tilde{\Upsilon}_{1,n\tau} - \tilde{\Upsilon}_{2,n\tau} + o_p(1) \tag{8}$$

with

$$\Psi_{n,\tau} = \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma G^*_{\sigma,\tau}$$

$$\begin{aligned}\tilde{\Upsilon}_{1,n\tau} = \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma \Bigg\{ & \left(D^*_{\sigma_{13},\tau} \frac{\widehat{f}_{x,\sigma_{13}} - f_{x,\sigma_{13}}}{f_{x,\sigma_{13}}} - D^*_{\sigma_{14},\tau} \frac{\widehat{f}_{x,\sigma_{14}} - f_{x,\sigma_{14}}}{f_{x,\sigma_{14}}} \right) \\ & - \left(D^*_{\sigma_{23},\tau} \frac{\widehat{f}_{x,\sigma_{23}} - f_{x,\sigma_{23}}}{f_{x,\sigma_{23}}} - D^*_{\sigma_{24},\tau} \frac{\widehat{f}_{x,\sigma_{24}} - f_{x,\sigma_{24}}}{f_{x,\sigma_{24}}} \right) \Bigg\}\end{aligned}$$

$$\begin{aligned}\tilde{\Upsilon}_{2,n\tau} = \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma \Bigg\{ & \left(D^*_{\sigma_{13},\tau} \frac{\widehat{f}_{vx,\sigma_{13}} - f_{vx,\sigma_{13}}}{f_{vx,\sigma_{13}}} - D^*_{\sigma_{14},\tau} \frac{\widehat{f}_{vx,\sigma_{14}} - f_{vx,\sigma_{14}}}{f_{vx,\sigma_{14}}} \right) \\ & - \left(D^*_{\sigma_{23},\tau} \frac{\widehat{f}_{vx,\sigma_{23}} - f_{vx,\sigma_{23}}}{f_{vx,\sigma_{23}}} - D^*_{\sigma_{24},\tau} \frac{\widehat{f}_{vx,\sigma_{24}} - f_{vx,\sigma_{24}}}{f_{vx,\sigma_{24}}} \right) \Bigg\}\end{aligned}$$

and $\sigma_{i_1 i_2} \in \{(i,k),(i,l),(j,k),(j,l)\}$ for any fixed $\sigma(\{i,j,k,l\}) \in \mathbf{N}_{m_n}$. The term $\tilde{\Upsilon}_{1,n\tau}$ accounts for the effect of estimating the marginal density $f_{x,ij}$ at a first-stage, while $\tilde{\Upsilon}_{2,n\tau}$ accounts for the effect of estimating the joint density $f_{vx,ij}$. The proof is completed by showing that

$$\widehat{\Psi}_{n,\tau} - \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau} \right] \xrightarrow{\text{p}} 0,$$

and $\tilde{\Upsilon}_{1,n\tau}, \tilde{\Upsilon}_{2,n\tau} \xrightarrow{\text{p}} 0$. The convergence in probability of the latter components follows from estimating consistently the probability densities using the kernel estimators in Eq. (6) and (7), i.e.

$$\begin{aligned}\sup_{(v_{12},x_{12})} \mid \widehat{f}_{vx}(v_{12},x_{12}) - f_{vx}(v_{12},x_{12}) \mid &= o_p(1) \\ \sup_{x_{12}} \mid \widehat{f}_x(x_{12}) - f_x(x_{12}) \mid &= o_p(1).\end{aligned}$$

4.2 Asymptotic Distribution

In this section, I derive the asymptotic distribution of $\widehat{\theta}_n$. In Appendix A.4, I show that semiparametric estimator $\widehat{\theta}_n$ has the following asymptotic linear representation

$$\left(\widehat{\theta}_n - \theta_0 \right) = \Gamma_0^{-1} \times S_{n,\tau} + o_p(1)$$

where $S_{n,\tau}$ is a sixth-order U-statistic given by

$$S_{n,\tau} = \binom{n}{6}^{-1} \sum_{\phi \in \mathbf{N}_{M_n}} \left\{ \psi_{\phi n\tau} - \mathbb{E}\left[\psi_{\phi n\tau} \mid \mathbf{Z}_n \right] \right\}$$

with kernel function $\psi_{\phi n\tau} = \psi_{0,\phi n\tau} + \psi_{x,\phi n\tau} - \psi_{vx,\phi n\tau}$. The term $S_{n,\tau}$ is a U-statistic of order six as the initial order of $\widehat{\Psi}_{n,\tau}$ is augmented by the kernel estimator $\widehat{f}_{v|x,ij}$, which can be represented as a second-order U-statistic.[8] The functions $\psi_{0,\phi n\tau}$, $\psi_{x,\phi n\tau}$, and $\psi_{vx,\phi n\tau}$ capture the influence functions of $\Psi_{n,\tau}$, $\tilde{\Upsilon}_{1,n\tau}$, and $\tilde{\Upsilon}_{2,n\tau}$ in the decomposition (8) and are defined in Appendix A.2.

A key step in characterizing the asymptotic distribution of $\widehat{\theta}_n$ consists in deriving the Hájek Projection of $S_{n,\tau}$ onto the space of arbitrary functions $(v_{ij}, X_{ij}, A_i, A_j)$, which drives the asymptotic distribution of the two-step semiparametric estimator. In particular, in Appendix A.4, I show that the Hájek Projection of $S_{n,\tau}$ yields the following equivalant representation

$$\left(\widehat{\theta}_n - \theta_0\right) = 15 \times \Gamma_0^{-1} \times \left\{ \binom{n}{2}^{-1} \sum_{i<j} \zeta_{ij} \right\} + o_p(1)$$

where

$$S^*_{n,\tau} \equiv \binom{n}{2}^{-1} \sum_{i<j} \zeta_{ij} \tag{9}$$

with $\zeta_{ij} \equiv \mathbb{E}\left[\widetilde{\psi}_{\phi n\tau} \mid Z_{ij}, U_{ij}\right]$ and $\widetilde{\psi}_{\phi n\tau} = \psi_{\phi n\tau} - \mathbb{E}\left[\psi_{\phi n\tau} \mid \mathbf{Z}_n\right]$. Moreover, the asymptotic variance $\mathbb{V}\left(S^*_{n,\tau}\right)$ satisfies the order condition $\tilde{\Omega} \equiv n(n-1)\rho_n \mathbb{V}\left(S^*_{n,\tau}\right) = O\left(1\right)$.

Assumptions 3.1 ensures that the elements of the Hájek Projection in Eq. (9) are conditionally independent given $\mathbf{Z}_n$, with conditional mean equal zero. This conditional independence structure is used to establish the following asymptotic distribution

$$\widehat{\Omega}_n^{-1/2} S^*_{n,\tau} \rightsquigarrow N\left(0, I\right)$$

where I represents the $d_\theta \times d_\theta$ identity matrix and $\widehat{\Omega}_n$ denotes a consistent estimator $\mathbb{V}\left(S^*_{n,\tau}\right)$ and is defined as

$$\widehat{\Omega}_n \equiv \binom{n}{2}^{-1} \sum_{i<j} \zeta_{ij} \zeta_{ij}^\top.$$

The next theorem formalizes the limiting distribution of $\widehat{\theta}_n$.

Theorem 4.2. *Suppose Assumptions 3.1-4.3 hold and $n\rho_n \to \infty$ as $n \to \infty$. Then*

$$\mathcal{V}_n^{-1/2} \left(\widehat{\theta}_n - \theta_0\right) \rightsquigarrow \mathcal{N}\left(0, 15^2 I\right)$$

[8] Alike to the notation used for $\sigma\left(\{i,k,k,l\}\right)$, which maps each tetrads in a network of size n into the set $\mathbf{N}_{m_n}$, the function ϕ is defined as a function that maps each unique 6-tuple $\{i,j,k,l,s,p\}$ in a network of size n to the index set $\mathbf{N}_{M_n} = \{1, \cdots, M_n\}$, where $M_n = \binom{n}{6}$ denotes the total number of 6-tuples with distinct indices $\{i,j,k,l,s,p\} \in \mathbf{N}_n$.

where

$$\mathcal{V}_n \equiv \widehat{\Gamma}_n^{-1} \times \widehat{\Omega}_n \times \widehat{\Gamma}_n^{-1}. \tag{10}$$

Proof. See Appendix A.4. □

Theorem 4.2 describes the asymptotic distribution of $\widehat{\theta}_n$. The semiparametric estimator $\widehat{\theta}_n$ converges to θ_0 at a rate $\sqrt{n(n-1)\rho_n}$, which is ensured by the order condition of $\tilde{\Omega}$. This rate is the square root of the effective sample given by the number of dyads $n(n-1)$ and the expected number of contributing tetrads ρ_n. Under sparse network asymptotic, the share of contributing tetrads ρ_n is allowed to convergence to zero as the network grows, but a lower rate than that at which $n \to \infty$, i.e, $\rho_n \to 0$ and $n\rho_n \to \infty$ as $n \to \infty$. Whereas if the network is dense, then $\rho_n \to \rho_0 > 0$ and $\widehat{\theta}_n$ converges at a parametric rate given by $\sqrt{n(n-1)}$.

5 Simulations

This section presents simulation evidence for the finite sample performance of the semiparametric estimator introduced in Section 4. I compare the performance of this estimator with the Tetrad Logit estimator introduced in Graham (2017), which relies on the assumption that the link-specific disturbance terms are logistically distributed. I consider a wide array of DGP designs that are meant to capture differences in distributional assumptions, sample size, and degree of connections in the network.

The undirected network is simulated according to the network model in Eq. (1). I consider a single observed attribute in X_i, which is drawn as $X_i \sim \text{Beta}(2,2) - \frac{1}{2}$. The pair-specific covariate X_{ij} is constructed to account for complementarities on the observed attributes and is defined as $X_{ij} = X_i X_j$. The agent-specific unobserved factor A_i is generated such that it is correlated with X_i and depends on the sample size n. This last feature offers a useful approach to control the degree of links formed in the network. In particular, I set $A_i = \lambda X_i - (1-\lambda) C_n \times \text{Beta}(0.5, 0.5)$, where the Beta random variable is independent of X_i and concentrates mass at the boundary of the unit interval. This implies that conditional on X_i, the individuals cluster at small or high types of unobserved attributes. The parameter $\lambda \in (0,1)$ controls the degree of correlation between the agent-specific heterogeneity and the observed covariate X_i, which is set to $\lambda = \frac{3}{4}$. The constant C_n depends on the size of the network and takes the values $C_n \in \left\{\log(\log(n)), \log(n)^{1/2}, \log(n), n^{1/3}\right\}$. Under this design, the choice of C_n regulates the degree of link formation. For instance, if C_n takes large values, fewer links will be formed in the network, thus generating a more sparse network. Alternatively, smaller values of C_n will produce denser networks.

Regarding the simulation of the special regressor and link-specific disturbance term, I consider two main DGP specifications. In the first DGP, I simulate the special regressor as $v_{ij} \sim N(0, 1.5)$ for $i < j$, and the link-specific disturbance term is generated as $U_{ij} \sim \text{Beta}(2,2) - \frac{1}{2}$ for $i < j$.

Note that under this specification, the independence and support conditions in Assumptions 3.2 and 3.3 are satisfied. This DGP is intended to illustrate the case in which v_{ij} has a large support and the support of U_{ij} is bounded. Moreover, under this design, the Tetrad Logit estimator studied in Graham (2017) is misspecified.

In the second DGP, I simulate the special regressor as $v_{ij} \sim \text{Logistic}\,(0, 1.5)$ for $i < j$, and the link-specific disturbance term is generated as $U_{ij} \sim \text{Logistic}(0, 1)$ for $i < j$. This DGP is intended to illustrate the case in which both v_{ij} and U_{ij} follow unbounded distributions, but the distribution of v_{ij} has relatively heavier tails than the remaining components in the network formation model. Under this DGP, the Tetrad Logit estimator in Graham (2017) is correctly specified. This specification also satisfies the independence and support conditions in Assumptions 3.2 and 3.3.

The true DGP design is completed by setting the parameter value $\theta_0 = 1.5$ and considering two different network sizes $n \in \{100, 200\}$. The choices of the network size are intended to be representative of the real-world network studied in Section 6.

I compute the semiparametric two-step estimator $\widehat{\theta}_n$ as defined in Eq. (5). The implementation of the semiparametric estimator for θ_0 requires estimating the conditional density of v_{ij} in a nonparametric first stage. Although Assumption 4.5 instructs the use of higher-order kernels to eliminate the asymptotic bias, I compute $\widehat{\theta}_n$ using a standard second-order kernel. I do this because semiparametric estimators computed using high-order kernels tend to have inferior finite sample properties compared to those obtained using standard kernels. Furthermore, this choice is common in many semiparametric applications (e.g., Rothe 2009). I use the standard-normal density as the kernel function. The bandwidth parameter h is set to be equal to 0.025, but I also consider different values for the bandwidth parameter, obtaining qualitatively similar results. These results are summarized in Appendix B. I also consider a fixed trimming design given by $I_{\tau,ij} = 1\,[| \; v_{ij} \; |< \tau]$ with $\tau = 2std(v_{ij})$.

Table 1 summarizes the results of computing the semiparametric two-step estimator $\widehat{\theta}_n$ and the Tetrad Logit estimator under the first DGP and over 1000 Monte Carlo replications. In particular, I report the mean, median, standard deviation (std), and mean squared error (MSE) of the two estimators over the total number of simulations. The final column of Table 1 reports the average degree of the network across the total number of simulations. I will use this information to describe the degree of link formation across the different designs.

The top panel in Table 1 shows the results of computing both estimators for θ_0 in a network with a size of $n = 100$. Both the mean and median show that the semiparametric estimator $\widehat{\theta}_n$ approximates well the true value of $\theta_0 = 1.5$ independently of the network degree. Furthermore, these results suggest that the estimator $\widehat{\theta}_n$ presents the smallest dispersion in dense network designs, e.g., the MSE is 0.190 when $C_n = \log(\log(n))$ and the average network degree is 39%. As fewer links are present in the network and the value of C_n increases, the dispersion of this estimator increases. Relative to the Tetrad Logit estimator, the semiparametric estimator $\widehat{\theta}_n$ has a smaller bias but a larger MSE across all the designs. This behavior is expected as, under this DGP $U_{ij} \sim \text{Beta}(2,2) - \frac{1}{2}$, and thus, the Tetrad Logit is not correctly specified. The larger dispersion of $\widehat{\theta}_n$ is explained as its std has to account for the approximation error induced by the nonparametric first-stage estimation.

In the bottom panel of Table 1, I show the results of estimating θ_0 in a large network with $n = 200$. The evidence in this scenario reinforces the previous findings and suggests that the performance of the estimator $\widehat{\theta}_n$ improves across all the designs. For example, in the densest network scenario when $C_n = \log(\log(n))$, the bias shrinks from 0.06 to 0.03, the std decreases by a factor of 2, and the MSE by a factor of 4. A similar pattern is observed in the sparsest network case when $C_n = n^{1/3}$ and only 19% of the links are formed. Moreover, notice that although the MSE of the Tetrad Logit estimator decreases in the larger network, the bias fails to disappear. This result highlights the consequence of using the Tetrad Logit estimator to recover θ_0 when the distribution of U_{ij} is not logistic.

Table 2 summarizes the results of computing the semiparametric two-step estimator $\widehat{\theta}_n$ and the Tetrad Logit estimator under the second DGP, which assumes the error term U_{ij} follows a logistic distribution. The results are qualitatively similar, which suggests that $\widehat{\theta}_n$ estimates well the true value of θ_0 when both v_{ij} and U_{ij} follow an unbounded distribution. Overall, these numerical experiments convey two main insights. First, the semiparametric estimator $\widehat{\theta}_n$ yields a reliable inference for the parameter θ_0 across different degrees of network formation. Second, in contrast to the Tetrad Logit, the estimator $\widehat{\theta}_n$ represents a viable alternative when the distribution of U_{ij} is unknown.

6 Empirical Application

In this section, I study a friendship network of high-school students. The objective is to estimate the preference parameters associated with socioeconomic, demographic, health, and educational factors that could drive the formation of this network. I use the self-reported friendship connections from the Add Health dataset to construct an undirected network of high-school friends (Harris et al. 2009) and compute the semiparametric estimator $\widehat{\theta}_n$ introduced in Section 4.

This is the first paper that uses the Add Health dataset to estimate a dyadic network model with unobserved heterogeneity using a special regressor. Thus, the results presented below provide new and richer evidence on the predictive factors that drive the formation of a friendship network.[9]

In this analysis, I use a representative sample of four saturated high schools included in Wave 1 of the In-Home survey.[10] In Appendix D, I describe in detail the Add Health dataset and the

[9] The Add Health dataset has been employed to study the relationship between social interactions and economic and behavioral outcomes (see Masten 2018 and references therein). Fewer studies have utilized this dataset to estimate the formation of a friendship network, but none has studied a dyadic specification with unobserved agent-specific heterogeneity (cf. Goldsmith-Pinkham and Imbens 2013 and Christakis et al. 2020).

[10] Saturated schools are those where all the students were selected for In-Home interviews regardless of whether

construction of the sample for this study. I also provide evidence of the representativeness of this sample. The final sample includes 273 students.

An undirected friendship link D_{ij} for any students i and j is recorded to be equal to 1 if either i or j name each other as friends, regardless of the order in which they do it. Of the 273 students in the network, 34 did not form any connection and thus remained isolated. On average, each student named up to 3 friends, and the maximum number of friendship connections formed by a student is 10. Figure 1 represents the empirical distribution of the network degree.

As a special regressor, I consider students' birth weight. There is abundant empirical evidence documenting the relationship between birth weight and cognitive and non-cognitive skills, health-related and socioeconomic outcomes (see, e.g., Almond et al. 2018 for a survey of the literature on early childhood conditions). Based on the established relationship between birth weight and individuals' personality traits, I use birth weight as a contributing factor to the formation of a friendship network among high-school students.

To examine the effect of birth weight on the formation of a friendship network, I specify the birth weight of student i in deviation from the average birth weight in my sample. In other words, the birth weight of student i is defined as $v_i - \bar{v}_n$, where $\bar{v}_n = n^{-1}\sum_{i=1}^{n} v_i$ is approximately 3400 grams.[11] In terms of the dyad-specific variable v_{ij} in the network model of Eq. (1), v_{ij} captures assortative matching among those students with a birth weight value above or below the sample average, i.e., $v_{ij} = (v_i - \bar{v}_n)\times(v_j - \bar{v}_n)$.

Importantly, the special regressor is required to satisfy the conditional independence and support conditions described in Assumptions 3.2 and 3.3. Here, I discuss why birth weight is likely to satisfy both assumptions.

Conditional Independence: Assumption 3.2 implies that the students' birth weight v_{ij} must be conditionally independent from $e_{ij} = A_i + A_j - U_{ij}$ given X_{ij}. While birth weight is not randomly assigned, the richness of the Add Health dataset allows me to control for a wide range of socio-demographic, economic, health, and educational factors of both the students and their parents, which have been shown to be related to both students' birth weight and their unobserved individual-specific attributes. In particular, I include covariates that aim to capture assortative matching across demographic, academic, economic, and physical attributes, as well as on early childhood conditions. In Appendix D, I describe in detail the observed characteristics used in this analysis, their relationship with students' birth weight and individual traits, and their potential effect on establishing friendship links.[12]

While potentially there could be other unobserved factors that are correlated with birth weight

they had completed an In-School questionnaire. I focus on saturated schools since their design allows me to recover the complete friendship network for all the students enrolled in these schools.

[11] While there is no consensus on how to specify birth weight in a regression (e.g., levels, logs, and deviations from the mean), this specification has also been used in Black et al. (2007) and Maruyama and Heinesen (2020).

[12] Appendix Table D1 summarizes their descriptive statistics.

and affect the formation of a friendship network, controlling for this rich set of attributes should mitigate concerns in this respect. In turn, any residual variation in birth weight is likely to be as good as idiosyncratic.[13] In other words, the specification of the network formation model that I propose here should contribute to ensuring that birth weight satisfies Assumption 3.2.

Large Support: the special regressor is also required to satisfy the support condition described in Assumption 3.3. Birth weight is recorded as a continuous variable, and after centering it around zero, the demeaned variable $v_i - \bar{v}_n$ has a support that ranges from -1567 to 1975 grams. Figure 2 presents the empirical distribution of birth weight, which, jointly with the descriptive statistics reported in Table D1, indicates that the special regressor is a continuous variable that has larger variance and support than the remaining covariates in the network formation model in Eq. (1).[14] Therefore, Assumption 3.3 holds in this context. In summary, birth weight is likely to satisfy both Assumptions 3.2 and 3.3, and thus, represent a valid special regressor.

Table 3 summarizes the results of estimating the network formation model in Eq. (1) using the two-step semiparametric estimator $\widehat{\theta}_n$ over a sample of 273 students taking $37,128$ unique linking decisions. This table also includes the results from two parametric designs that will be used to assess the performance of the semiparametric approach: (i) a model with a logistic error term with fixed effects and (ii) a model with a logistic error term without fixed effects.

The semiparametric estimates in column (1) indicate positive homophily effects on the covariates of age, gender, repeated grade, smoking and drinking habits, religion, divorced parents, and breastfed at birth. These findings suggest that students tend to form friendship connections with other students of similar age and gender, but also with those who have lived similar experiences, such as repeating an academic grade or experiencing their parents' divorce. Homophily among students who have been breastfed might represent a proxy for different mechanisms, such as establishing friendship connections among students nurtured by their mothers. Additionally, the homophily effects on smoking and drinking habits and religion suggest that sharing similar habits or social norms increases the probability of establishing a friendship connection. Overall, the results of this analysis seem plausible. Moreover, this analysis documents the importance of additional factors beyond those that have been identified by the literature on friendship network formation, such as age and gender (Miyauchi 2016; Christakis et al. 2020).[15]

Next, I compare the results of the semiparametric estimator with those of the parametric specifications. First, the semiparametric methodology estimates similar homophily coefficients to the

[13] If the researcher was to collect the network data by conducting a survey or an experiment, the special regressor could be constructed as a covariate that is independently distributed of the remaining individual attributes as in the experiment on willingness to pay for protecting wetland habitats and wildlife in California discussed in Lewbel et al. (2011).

[14] Appendix Table D2 summarizes the descriptive statistics at a dyadic level, which reinforces this evidence.

[15] Although imprecisely estimated, the signs of the remaining estimates are also empirically interesting. Notably, the results suggest positive homophily effects on overall GPA, clubs, number of siblings, attractiveness, and mother's participation in the labour market. Moreover, these results indicate a negative homophily effect on covariates such as depression, order of siblings, mother's age at birth, and mother's health risk factors.

logistic model with fixed effects, though, as expected, it generates larger standard errors. When the two methodologies differ, the results of the semiparametric estimator seem more intuitive. For instance, the parametric model predicts a negative effect on attractiveness, disability, and mother's participation in the labour market, and positive homophily effects on depression, which seem at odds with the existing evidence on their relationship with individuals' personality traits (Brunello and Schlotter 2011). As for the second parametric specification, when the fixed effects are not included, the estimated coefficients on gender, ethnicity, and overall GPA are incompatible with those documented in the literature on social networks (Miyauchi 2016; Christakis et al. 2020).

Overall, these results document new factors of homophilic preferences in the formation of friendship networks. They also provide empirical support for the validity of the methodology proposed here and its strength relative to other methods.

7 Conclusion

This paper studies a network formation model with unobserved agent-specific heterogeneity and offers two main contributions to the literature on network formation. First, it proposes a new identification strategy that recovers the preference parameters associated with homophily on observed attributes. The identification result relies on the existence of a special regressor, and to the best of my knowledge, this paper represents the first generalization of the special regressor methodology to analyze network models.

The second contribution of this paper is to introduce a two-step semiparametric estimator for the parameter of interest. The estimator has a least-squares analytic form and is computationally tractable even in large networks. In Monte Carlo simulations, I show that it performs well in finite samples and networks with different degrees of sparsity.

Finally, in an empirical application, I use the methodology developed in this paper to study the factors that drive the formation of a friendship network among high-school students in the Add Health dataset. As a special regressor, I consider students' birth weight and control for a wide range of observed attributes of the students and their parents. Consistent with the results of the literature, I find evidence for homophily on age and gender and document additional effects on school achievements, habits, social norms, and parental strategies. A comparison of these results with those obtained under parametric specifications provides empirical support for the validity of the methodology proposed here and its strength relative to other methods.

References

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

Aizer, A. and J. Currie (2014). The intergenerational transmission of inequality: maternal disadvantage and health at birth. *Science 344*(6186), 856–861.

Almond, D. and J. Currie (2011). Human capital development before age five. *Handbook of Labor Economics 4*, 1315–1486.

Almond, D., J. Currie, and V. Duque (2018). Childhood circumstances and adult outcomes: Act ii. *Journal of Economic Literature 56*(4), 1360–1446.

Aradillas-Lopez, A. (2012). Pairwise-difference estimation of incomplete information games. *Journal of Econometrics 168*(1), 120–140.

Arellano, M. and B. Honoré (2001). Panel data models: some recent developments. *Handbook of Econometrics 5*, 3229–3296.

Auerbach, E. (2022). Identification and estimation of a partially linear regression model using network data. *Econometrica 90*(1), 347–365.

Black, S. E., P. J. Devereux, and K. G. Salvanes (2007). From the cradle to the labor market? the effect of birth weight on adult outcomes. *Quarterly Journal of Economics 122*(1), 409–439.

Brunello, G. and M. Schlotter (2011). Non-cognitive skills and personality traits: Labour market relevance and their development in education & training systems.

Charbonneau, K. B. (2017). Multiple fixed effects in binary response panel data models. *The Econometrics Journal 20*(3), S1–S13.

Chernozhukov, V., I. Fernandez-Val, and M. Weidner (2020). Network and panel quantile effects via distribution regression. *Journal of Econometrics*.

Christakis, N., J. Fowler, G. W. Imbens, and K. Kalyanaraman (2020). An empirical model for strategic network formation. *The Econometric Analysis of Network Data*, 123–148.

de Paula, Á. (2020). Econometric models of network formation. *Annual Review of Economics 12*, 775–799.

Del Bono, E., J. Ermisch, and M. Francesconi (2012). Intrafamily resource allocations: a dynamic structural model of birth weight. *Journal of Labor Economics 30*(3), 657–706.

Dzemski, A. (2019). An empirical model of dyadic link formation in a network with unobserved heterogeneity. *Review of Economics and Statistics 101*(5), 763–776.

Figlio, D., J. Guryan, K. Karbownik, and J. Roth (2014). The effects of poor neonatal health on children's cognitive development. *American Economic Review 104*(12), 3921–55.

Fitzsimons, E. and M. Vera-Hernández (2015). Breastfeeding and child development. *University College London and Institute for Fiscal Studies*.

Gao, W. Y. (2020). Nonparametric identification in index models of link formation. *Journal of Econometrics 215*(2), 399–413.

Goldsmith-Pinkham, P. and G. W. Imbens (2013). Social networks and the identification of peer effects. *Journal of Business & Economic Statistics 31*(3), 253–264.

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

Graham, B. S. (2020). Network data. In *Handbook of Econometrics*, Volume 7, pp. 111–218. Elsevier.

Graham, B. S., F. Niu, and J. L. Powell (2019). Kernel density estimation for undirected dyadic data. *arXiv preprint arXiv:1907.13630*.

Graham, B. S., F. Niu, and J. L. Powell (2021). Minimax risk and uniform convergence rates for nonparametric dyadic regression. Technical report, National Bureau of Economic Research.

Han, A. K. (1987). Non-parametric analysis of a generalized regression model: the maximum rank correlation estimator. *Journal of Econometrics 35*(2), 303–316.

Hansen, B. E. (2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory 24*(3), 726–748.

Harris, K. M., C. T. Halpern, E. Whitsel, J. Hussey, J. Tabor, P. Entzel, and J. R. Udry (2009). The national longitudinal study of adolescent health: Research design. *WWW document*.

Honoré, B. E. and A. Lewbel (2002). Semiparametric binary choice panel data models without strictly exogeneous regressors. *Econometrica 70*(5), 2053–2063.

Jackson, M. O. and A. Wolinsky (1996). A strategic model of social and economic networks. *Journal of Economic Theory 71*(1), 44–74.

Jochmans, K. (2017). Two-way models for gravity. *Review of Economics and Statistics 99*(3), 478–485.

Jochmans, K. (2018). Semiparametric analysis of network formation. *Journal of Business & Economic Statistics 36*(4), 705–713.

Khan, S. and E. Tamer (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica 78*(6), 2021–2042.

Leung, M. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics 188*(1), 182–195.

Lewbel, A. (1997). Semiparametric estimation of location and other discrete choice moments. *Econometric Theory 13*(01), 32–51.

Lewbel, A. (1998). Semiparametric latent variable model estimation with endogenous or mismeasured regressors. *Econometrica*, 105–121.

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics 97*(1), 145–177.

Lewbel, A. (2014). An overview of the special regressor method. In *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometrics and Statistics*, pp. 38–62. Oxford University Press.

Lewbel, A., D. McFadden, and O. Linton (2011). Estimating features of a distribution from binomial data. *Journal of Econometrics 162*(2), 170–188.

Ma, S., L. Su, and Y. Zhang (2020). Detecting latent communities in network formation models. *arXiv preprint arXiv:2005.03226*.

Maruyama, S. and E. Heinesen (2020). Another look at returns to birthweight. *Journal of Health Economics 70*, 102269.

Masten, M. A. (2018). Random coefficients on endogenous variables in simultaneous equations models. *The Review of Economic Studies 85*(2), 1193–1250.

McPherson, M., L. Smith-Lovin, and J. M. Cook (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 415–444.

Miyauchi, Y. (2016). Structural estimation of pairwise stable networks with nonnegative externality. *Journal of Econometrics 195*(2), 224–235.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. *Handbook of Econometrics 4*, 2111–2245.

Powell, J. L. (1994). Estimation of semiparametric models. *Handbook of Econometrics 4*, 2443–2521.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica*, 931–954.

Rothe, C. (2009). Semiparametric estimation of binary response models with endogenous regressors. *Journal of Econometrics 153*(1), 51–64.

Serfling, R. J. (2009). *Approximation Theorems of Mathematical Statistics*, Volume 162. John Wiley & Sons.

Toth, P. (2017). Semiparametric estimation in networks with homophily and degree heterogeneity. Technical report, Working paper, University of Nevada.

Yan, T., B. Jiang, S. E. Fienberg, and C. Leng (2019). Statistical inference in a directed network model with covariates. *Journal of the American Statistical Association 114*(526), 857–868.

Zeleneev, A. (2020). Identification and estimation of network models with nonparametric unobserved heterogeneity. Working Paper.

8 Tables

Table 1: Simulation Results

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
					$n = 100$				
$\log(\log(n))$	1.5610	1.5546	0.4327	0.1909	1.6528	1.6526	0.2839	0.1040	0.3990
$\log(n)^{1/2}$	1.5529	1.5534	0.4838	0.2368	1.6421	1.6415	0.3038	0.1125	0.3526
$\log(n)$	1.5584	1.5662	0.6267	0.3962	1.6437	1.6373	0.3585	0.1492	0.2386
$n^{1/3}$	1.5546	1.5513	0.6110	0.3763	1.6321	1.6325	0.3658	0.1513	0.2368
					$n = 200$				
$\log(\log(n))$	1.5307	1.5233	0.2133	0.0465	1.6413	1.6429	0.1402	0.0396	0.3916
$\log(n)^{1/2}$	1.5280	1.5174	0.2318	0.0545	1.6379	1.6427	0.1458	0.0403	0.3348
$\log(n)$	1.5276	1.5334	0.3263	0.1072	1.6373	1.6347	0.1839	0.0527	0.2135
$n^{1/3}$	1.5150	1.5165	0.3517	0.1239	1.6433	1.6445	0.1770	0.0519	0.1953

[1] DGP: $v_{ij} \sim N\,(0, 1.5)$ and $U_{ij} \sim Beta(2,2) - \frac{1}{2}$.

[2] Total number of Monte Carlo simulations = 1000.

[3] Bandwidth parameter $h = 0.025$.

Table 2: Simulation Results

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
					$n = 100$				
$\log(\log(n))$	1.5201	1.5232	1.1119	1.2368	1.3777	1.3778	0.8961	0.8179	0.4459
$\log(n)^{1/2}$	1.5595	1.5660	1.1017	1.2173	1.3960	1.3587	0.9006	0.8219	0.4211
$\log(n)$	1.5015	1.5166	1.2400	1.5377	1.4122	1.3956	0.9287	0.8702	0.3518
$n^{1/3}$	1.3850	1.4209	1.1783	1.4017	1.3493	1.3620	0.9269	0.8818	0.3512
					$n = 200$				
$\log(\log(n))$	1.4746	1.4622	0.5207	0.2718	1.3698	1.3499	0.4180	0.1917	0.4437
$\log(n)^{1/2}$	1.4882	1.4802	0.5245	0.2752	1.3821	1.3892	0.4308	0.1995	0.4125
$\log(n)$	1.4746	1.4396	0.5667	0.3218	1.3915	1.3892	0.4438	0.2087	0.3346
$n^{1/3}$	1.4201	1.4168	0.6100	0.3785	1.3941	1.3927	0.4726	0.2346	0.3194

[1] DGP: $v_{ij} \sim Logistic\,(0, 1.5)$ and $U_{ij} \sim Logistic\,(0, 1)$

[1] Total number of Monte Carlo simulations = 1000.

[2] Bandwidth parameter $h = 0.025$.

Figure 1: Network degree empirical distribution

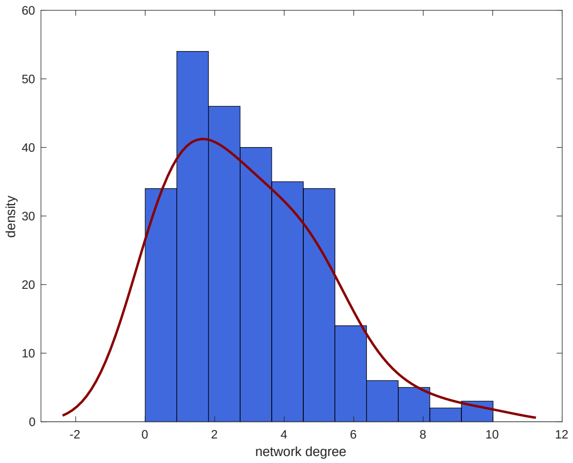

Figure 2: Birth weight empirical distribution

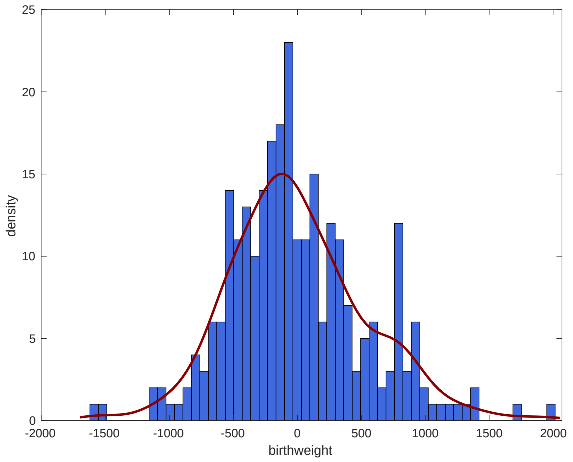

Table 3: Point Estimates

	Semiparametric		Logistic with FE		Logistic without FE	
	(1)		(2)		(3)	
	estimate	s.e.	estimate	s.e.	estimate	s.e.
age	1.045**	(0.405)	1.769***	(0.070)	0.584***	(0.113)
gender	0.397*	(0.230)	0.139**	(0.056)	−0.204**	(0.096)
white	0.007	(3.469)	4.911***	(0.203)	−0.764***	(0.010)
grade	−0.766	(0.234)	2.985***	(0.074)	1.704***	(0.108)
overall GPA	0.763	(0.540)	0.365***	(0.052)	−0.111***	(0.016)
clubs	0.030	(0.028)	0.146***	(0.010)	0.073***	(0.010)
repeated grade	1.713**	(0.675)	1.103***	(0.222)	−0.767***	(0.297)
depressed	−0.122	(0.244)	0.240***	(0.072)	0.076***	(0.088)
number of siblings	0.467	(0.754)	0.029	(0.029)	0.069***	(0.017)
order of siblings	−0.594	(0.486)	0.194***	(0.047)	−0.138***	(0.038)
mother highly educated	−1.173	(1.252)	−0.590***	(0.141)	0.697***	(0.132)
mother works	1.348	(1.384)	−2.456***	(0.233)	−0.215**	(0.102)
S&D habits	0.743*	(0.443)	1.035***	(0.110)	0.121	(0.108)
friends' S&D habits	0.096	(0.169)	0.272***	(0.028)	0.090***	(0.025)
good neighborhood	−0.011	(0.703)	−0.605***	(0.125)	−0.327***	(0.101)
religion	1.232*	(0.737)	1.075***	(0.089)	1.004***	(0.110)
college expectations	−0.134	(0.221)	0.130***	(0.037)	−0.116***	(0.007)
attractiveness	0.023	(0.148)	−0.206***	(0.054)	0.172***	(0.061)
mother's health	0.030	(0.273)	0.110***	(0.036)	−0.132***	(0.009)
divorced parents	2.091**	(0.484)	1.021***	(0.130)	0.069	(0.178)
breastfed	1.669**	(0.658)	0.018	(0.117)	0.669***	(0.149)
disability	0.648	(2.588)	−2.670***	(0.501)	−2.469**	(1.056)
mother's age at birth	−0.812	(0.529)	0.133	(0.111)	−0.156	(0.225)
mother's health risk factors	−0.894	(0.635)	−0.004	(0.121)	−0.370**	(0.168)
log household income	−0.497	(0.667)	1.081***	(0.074)	0.205**	(0.102)

Sample size $n = 273$. Significance levels: *10%, **5%, and ***1%.

A SEMIPARAMETRIC NETWORK FORMATION MODEL WITH UNOBSERVED LINEAR HETEROGENEITY

Appendices

A Mathematical Proofs

A.1 Identification

Proof of Lemma 1. Let $Q(x,e) = -x^\top \theta_0 - e$. For any i and j consider consider

$$\begin{aligned}
\mathbb{E}\left[D^*_{ik} \mid X_{ij}, \boldsymbol{X}_{-ij}\right] &= \mathbb{E}\left[\mathbb{E}\left[D^*_{ij} \mid v_{ij}, \boldsymbol{X}_n\right] \mid X_{ij}, \boldsymbol{X}_{-ij}\right] \\
&= \mathbb{E}\left[\frac{\mathbb{E}\left[D^*_{ij} \mid v_{ij}, \boldsymbol{X}_n\right] - \mathbf{1}\left[v_{ij} > 0\right]}{f_{v|x}(v_{ij} \mid X_{ij})} \mid X_{ij}, \boldsymbol{X}_{-ij}\right] \\
&= \int_{\underline{v}(X_{ij})}^{\overline{v}(X_{ij})} \left\{\frac{\mathbb{E}\left[D^*_{ij} \mid v_{ij}, \boldsymbol{X}_n\right] - \mathbf{1}\left[v_{ij} > 0\right]}{f_{v|x}(v_{ij} \mid X_{ij})}\right\} f_{v|x}(v_{ij} \mid X_{ij}) dv_{ij} \\
&= \int_{\underline{v}(X_{ij})}^{\overline{v}(X_{ij})} \left\{\int_{\mathbb{S}_{e|X}} \left\{\mathbf{1}\left[v_{ij} \geq Q(X_{ij}, e_{ij})\right]\right\} dF_{e|x}(e_{ij} \mid \mathbf{X}_n) - \mathbf{1}\left[v_{ij} > 0\right]\right\} dv_{ij} \\
&= \int_{\mathbb{S}_{e|X}} \int_{\underline{v}(X_{ij})}^{\overline{v}(X_{ij})} \left\{\mathbf{1}\left[v_{ij} \geq Q(X_{ij}, e_{ij})\right] - \mathbf{1}\left[v_{ij} > 0\right]\right\} dv_{ij} dF_{e|x}(e_{ij} \mid \mathbf{X}_n) \\
&= \int_{\mathbb{S}_{e|X}} -Q(X_{ij}, e_{ij}) dF_{e|x}(e_{ij} \mid \mathbf{X}_n) \\
&= X_{ij}^\top \theta_0 + \mathbb{E}\left[A_i + A_j \mid \boldsymbol{X}_n\right]
\end{aligned}$$

where the third equality follows from Assumption 3.1 which states that v_{ij} is conditionally independent from $\mathbf{X}_{-ij}$ given X_{ij}. The second to last equality is the result of

$$\int_{\underline{v}(X_{ij})}^{\overline{v}(X_{ij})} \left\{\mathbf{1}\left[v_{ij} \geq Q(X_{ij}, e_{ij})\right] - \mathbf{1}\left[v_{ij} > 0\right]\right\} dv_{ij} \quad = \quad -Q(X_{ij}, e_{ij}).$$

□

Proof of Theorem 3.1. Fix any tetrad $\sigma\left(\{i,j,k,l\}\right) \in \mathbf{N}_{M_n}$, notice that

$$\mathbb{E}\left[D^*_{ik} - D^*_{il} \mid \boldsymbol{X}_n\right] = \left(X_{ik} - X_{il}\right)^\top \theta_0 + \mathbb{E}\left[A_k - A_l \mid \boldsymbol{X}_n\right],$$

and, consequently

$$\mathbb{E}\left[G^*_\sigma \mid \boldsymbol{X}_n\right] = W_\sigma^\top \theta_0 \tag{A.1}$$

and

$$\mathbb{E}\left[W_\sigma G^*_\sigma \mid \boldsymbol{X}_n\right] = W_\sigma W_\sigma^\top \theta_0.$$

The result follows from a Law of Iterated Expectations, aggregating all the information across all the contributing tetrads $\sigma \in \mathbf{N}_{m_n}$, and invoking Assumption 3.4, which ensures that Γ_0 is nonsingular. □

A.2 Nonparametric Kernel Estimator

The nonparametric kernel estimators are defined as

$$\widehat{f}_{vx}(v_{12}, x_{12}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K_{vxh,ij}\left[v_{12}, x_{12}\right]$$
$$\widehat{f}_{x,ij}(x_{12}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} K_{xh,ij}\left[x_{12}\right]$$

where the following notation is used for the kernels

$$K_{vxh,ij}\left[v_{12}, x_{12}\right] \equiv \frac{1}{h_n^{d_\theta+1}} K_{vx,ij}\left[v_{12}, x_{12}\right]$$
$$K_{xh,ij}\left[x_{12}\right] \equiv \frac{1}{h_n^{d_\theta}} K_{x,ij}\left[x_{12}\right].$$

The bandwidth parameter is denoted by h_n and

$$K_{vx,ij}\left[v_{12}, x_{12}\right] = K_{vx}\left[v_{ij} - v_{12}, X_{ij} - x_{12}\right] = \kappa_{vx}\left[\frac{v_{ij} - v_{12}}{h_n}, \frac{X_{ij} - x_{12}}{h_n}\right]$$
$$K_{x,ij}\left[x_{12}\right] = K_x\left[X_{ij} - x_{12}\right] = \kappa_x\left[\frac{X_{ij} - x_{12}}{h_n}\right].$$

Lemma 2. *Suppose that the assumptions in Theorem 4.1 hold. Then*

$$\sup_{(v_{12}, x_{12})} \mid \widehat{f}_{vx}(v_{12}, x_{12}) - f_{vx}(v_{12}, x_{12}) \mid = O_p\left(\alpha_{1n}\right)$$
$$\sup_{x_{12}} \mid \widehat{f}_x(x_{12}) - f_x(x_{12}) \mid = O_p\left(\alpha_{2n}\right),$$

with

$$\alpha_{1n} \equiv \left(\frac{\log n}{n h_n^{d_\theta+1}}\right)^{1/2}$$
$$\alpha_{2n} \equiv \left(\frac{\log n}{n h_n^{d_\theta}}\right)^{1/2}$$

Proof. This result follows from Theorem 3.2 in Graham et al. (2021) or Theorem 6 in Hansen (2008). □

Lemma 3. *Under assumptions 3.1, 4.2, and 4.3, the kernel estimator has the following representation*

$$\widehat{f}_x(x_{12}) - f_x(x_{12}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left\{K_{xh,ij}\left(x_{12}\right) - \mathbb{E}\left[K_{xh,ij}\left(x_{12}\right)\right]\right\} + o_p(1/\sqrt{n})$$

$$\widehat{f}_{vx}(v_{12}, x_{12}) - f_{vx}(v_{12}, x_{12}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\left\{K_{vxh,ij}\left(v_{12}, x_{12}\right) - \mathbb{E}\left[K_{vxh,ij}\left(v_{12}, x_{12}\right)\right]\right\} + o_p(1/\sqrt{n}).$$

Proof. Consider the result for $\widehat{f}_{x,12}$. Notice that

$$\begin{aligned}
\mathbb{E}\left[\widehat{f}_{x,12}\right] &= \mathbb{E}\left[K_{xh,ij}\left(x_{12}\right)\right] \\
&= \frac{1}{h_n^{d_\theta}}\mathbb{E}\left[K_{x,ij}\left(x_{12}\right)\right] \\
&= \frac{1}{h_n^{d_\theta}}\int K_x\left(X_{ij} - x_{12}\right) f_{x,ij} dX_{ij} \\
&= \int f_x(x_{12} + \nu h)\kappa_x\left(\nu\right) d\nu \\
&= \int \left[f_{x,12} + \sum_{|\lambda|=1}^{\delta}\frac{h_n^{|\lambda|}}{\mid\lambda\mid}\nabla^{|\lambda|} f_x(x_{12} + h_n\tilde{v})\left(\nu_1^{\lambda_1}\cdots\nu_{d_\theta}^{\lambda_{d_\theta}}\right)^{|\lambda|}\right]\kappa_x\left(\nu\right) d\nu \\
&= f_{x,12} + o(h_n^{\delta}),
\end{aligned}$$

where the fourth equality follows from the changes of variables $x_{12} = X_{ij} + \nu h_n$. The fifth equality follows from Assumption 4.2, which ensures that $f_x(\cdot)$ is δ-times differentiable. The last equality follows from Assumption 4.3, which ensures that $\kappa_x(\cdot)$ is a bias-reducing kernel of order δ. The result follows from Assumptions 4.3 which requires that $nh_n^{2\delta} \to 0$ as $n \to \infty$. □

Lemma 4. *Under assumptions 3.1, 4.2, and 4.3, the variance of the kernels are the following*

$$\mathbb{V}\left(\widehat{f}_x(x_{12})\right) = O\left(\frac{1}{n(n-1)h_n^{d_\theta}}\Omega_2(x_{12})\right) + O\left(\frac{1}{n}\Omega_1(x_{12})\right)$$

$$\mathbb{V}\left(\widehat{f}_{vx}(v_{12}, x_{12})\right) = O\left(\frac{1}{n(n-1)h_n^{d_\theta+1}}\Omega_2(v_{12}, x_{12})\right) + O\left(\frac{1}{n}\Omega_1(v_{12}, x_{12})\right)$$

where

$$\Omega_1(x_{12}) = \mathbb{E}\left[f_x\left(x_{12} \mid X_1\right)^2\right] \qquad \Omega_2(x_{12}) = f_{x,12}\int \kappa_x\left(v\right)^2 dv$$

$$\Omega_1(v_{12}, x_{12}) = \mathbb{E}\left[f_{vx}\left(v_{12}, x_{12} \mid v_1, X_1\right)^2\right] \qquad \Omega_2(v_{12}, x_{12}) = f_{vx,12}\int \kappa_{vx}\left(v\right)^2 dv.$$

Proof. A proof of this Lemma can be found in Graham et al. (2019). We show the result for $\widehat{f}_x(x_{12})$

for completeness.

$$\begin{aligned}
&Var\left(\widehat{f}_{x,12}\right)\\
&=Cov\left(\frac{1}{n(n-1)}\sum_{i\neq j}K_{xh,ij}\left(x_{12}\right),\frac{1}{n(n-1)}\sum_{k\neq l}K_{xh,kl}\left(x_{12}\right)\right)\\
&=\left(\frac{1}{n(n-1)}\right)^2\sum_{i\neq j}\sum_{k\neq l}Cov\left(K_{xh,ij}\left(x_{12}\right),K_{xh,kl}\left(x_{12}\right)\right)\\
&=\left(\frac{1}{n(n-1)}\right)^2\left\{\sum_{i\neq j}Var\left(K_{xh,ij}\left(x_{12}\right)\right)+\sum_{i=1}^{n}\sum_{j\neq i}\sum_{k\neq i,j}Cov\left(K_{xh,ij}\left(x_{12}\right),K_{xh,kl}\left(x_{12}\right)\right)\right\}\\
&=\left(\frac{1}{n(n-1)}\right)^2\left\{n(n-1)Var\left(K_{xh,ij}\left(x_{12}\right)\right)+n(n-1)(n-2)Cov\left(K_{xh,ij}\left(x_{12}\right),K_{xh,ik}\left(x_{12}\right)\right)\right\}\\
&=\left\{\frac{1}{n(n-1)}Var\left(K_{xh,ij}\left(x_{12}\right)\right)+\frac{(n-2)}{n(n-1)}Cov\left(K_{xh,ij}\left(x_{12}\right),K_{xh,il}\left(x_{12}\right)\right)\right\}
\end{aligned}$$

that is

$$Var\left(\widehat{f}_{x,12}\right)=O\left(\frac{Var\left(K_{xh,ij}\left(x_{12}\right)\right)}{n(n-1)}\right)+O\left(\frac{Cov\left(K_{xh,ij}\left(x_{12}\right),K_{xh,ik}\left(x_{12}\right)\right)}{n}\right). \tag{A.2}$$

Moreover,

$$\begin{aligned}
&\mathbb{E}\left[K_{xh,ij}\left(x_{12}\right)^2\right]\\
=&\frac{1}{h_n^{2d_\theta}}\int K_{x,ij}\left(X_{ij}-x_{12}\right)^2 f_{x,ij}dX_{ij}\\
=&\frac{1}{h_n^{d_\theta}}\int f_x(x_{12}+\nu h_n)\kappa_x\left(\nu\right)^2 d\nu\\
=&\frac{1}{h_n^{d_\theta}}\int\left\{f_{x,12}+\sum_{|\lambda|=1}^{\delta}\frac{h_n^{|\lambda|}}{|\,\lambda\,|}\nabla^{|\lambda|}f_x(x_{12}+h_n\tilde{v})\left(\nu_1^{\lambda_1}\cdots\nu_{d_\theta}^{\lambda_{d_\theta}}\right)^{|\lambda|}\right\}\kappa_x\left(\nu\right)^2 d\nu\\
=&\frac{1}{h_n^{d_\theta}}\left\{f_{x,12}\int\kappa_x\left(\nu\right)^2 d\nu\right\}\\
&+\frac{1}{h_n^{d_\theta}}\int\left[\sum_{|\lambda|=1}^{\delta}\frac{h_n^{|\lambda|}}{|\,\lambda\,|}\nabla^{|\lambda|}f_x(x_{12}+h_n\tilde{v})\left(\nu_1^{\lambda_1}\cdots\nu_{d_\theta}^{\lambda_{d_\theta}}\right)^{|\lambda|}\right]\kappa_x\left(\nu\right)^2 d\nu\\
&=\frac{1}{h_n^{d_\theta}}\Omega_2(x_{12})+O(1)
\end{aligned}$$

where the second equality follows from the changes of variables $X_{ij}=X_{12}+\nu h_n$. The third equality follows from Assumption 4.2, which ensures that $f_x(\cdot)$ is δ-times differentiable. The fourth equality follows from Assumption 4.3, which ensures that $\kappa_x(\cdot)$ has finite second moments. The last equality uses

$$\Omega_2(x_{12}) = f_{x,12} \int \kappa_x(\nu)^2 \, d\nu.$$

Thus

$$\begin{aligned} Var\left(K_{xh,ij}(x_{12})\right) &= \mathbb{E}\left[K_{xh,ij}(x_{12})^2\right] - \left(\mathbb{E}\left[K_{xh,ij}(x_{12})\right]\right) \\ &= \frac{1}{h_n^{d_\theta}} \Omega_2(x_{12}) - f_{x,12}^2 + o\left(h_n^{2\delta}\right). \end{aligned}$$

Similarly,

$$\begin{aligned} &\mathbb{E}\left[K_{xh,ij}\left(x_{12}\right) K_{xh,ik}\left(x_{12}\right)\right] \\ &= \frac{1}{h_n^{2d_\theta}} \mathbb{E}\left[\mathbb{E}\left[K_{x,ij}\left(x_{12}\right) K_{x,ik}\left(x_{12}\right) \mid X_i\right]\right] \\ &= \frac{1}{h_n^{2d_\theta}} \mathbb{E} \int\int K_x\left(X_{ij} - x_{12}\right) K_x\left(X_{ik} - X_{12}\right) f_x(X_{ij} \mid X_i) f_x(X_{ik} \mid X_i) dX_{ij} dX_{ik} \\ &= \frac{1}{h_n^{2d_\theta}} \mathbb{E} \int K_x\left(X_{ij} - x_{12}\right) f_x(X_{ij} \mid X_i) dX_{ij} \int K_x\left(X_{ik} - X_{12}\right) f_x(X_{ik} \mid X_i) dX_{ik} \\ &= \mathbb{E} \int \kappa_x\left(\nu_1\right) f_x(x_{12} + \nu_1 h_n \mid X_i) d\nu_1 \int \kappa_x\left(\nu_2\right) f_x(x_{12} + \nu_2 h_n \mid X_i) d\nu_2 \\ &= \Omega_1(x_{12}) + o(h^\delta), \end{aligned}$$

where

$$\Omega_1(x_{12}) = \mathbb{E}\left[f_x(x_{12} \mid X_i)^2\right]. \tag{A.3}$$

The third equality follows from Assumption 3.1. The fourth equality follows from the change of variables $X_{ij} = x_{12} + \nu_1 h_n$ and $X_{ik} = x_{12} + \nu_2 h_n$. It follows from the previous results that

$$\begin{aligned} &Var\left(K_{xh,ij}\left(x_{12}\right)\right) = \mathbb{E}\left[K_{xh,ij}\left(x_{12}\right)^2\right] - \mathbb{E}\left[K_{xh,ij}\left(x_{12}\right)\right]^2 = \frac{1}{h_n^{d_\theta}} \Omega_2(x_{12}) - f_{x,12}^2 \\ &Cov\left(K_{xh,ij}\left(x_{12}\right), K_{xh,ik}\left(x_{12}\right)\right) = \Omega_1(x_{12}) - f_{x,12}^2. \end{aligned}$$

Consequently

$$\begin{aligned} Var\left(\widehat{f}_{x,12}\right) &= O\left(\frac{Var\left(K_{xh,ij}\left(x_{12}\right)\right)}{n(n-1)}\right) + O\left(\frac{Cov\left(K_{xh,ij}\left(x_{12}\right), K_{xh,ik}\left(x_{12}\right)\right)}{n}\right) \\ &= O\left(\frac{1}{n(n-1)h_n^{d_\theta}}\right) \Omega_2(x_{12}) + O\left(\frac{1}{n}\right) \Omega_1(x_{12}). \end{aligned}$$

□

Lemma 5 (Consistency)**.** *Suppose the assumptions of Theorem 4.1 hold. For any* $\sigma(\{i,k,j,l\}) \in$

$\mathbf{N}_{m_n}$, *let* $\sigma_{i_1 i_2} \in \{(i,k),(i,l),(j,k),(j,l)\}$,

$$\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{ W_\sigma G^*_{\sigma,\tau} - \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau} \right] \right\} = o_p(1)$$

Proof. For any $\epsilon > 0$, consider

$$\begin{aligned}
&\Pr\left[\left| \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{ W_\sigma G^*_{\sigma,\tau} - \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau} \right] \right\} \right| > \epsilon \right] \\
&\leq \frac{1}{\epsilon^2 m_n^2 \rho_n^2} \mathbb{E}\left[\left\{ \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{ W_\sigma G^*_{\sigma,\tau} - \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau} \right] \right\} \right\}^2 \right] \\
&\leq \frac{1}{\epsilon^2 n m_n \rho_n^2} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{ \mathbb{E}\left[\left(W_\sigma G^*_{\sigma,\tau} - \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau} \right] \right)^2 \mid \omega_\sigma = 1 \right] \Pr\left[\omega_\sigma = 1 \right] \right\} \\
&= O_p\left(\frac{1}{n \rho_n} \right).
\end{aligned}$$

The first inequality follows from Chebyshev's inequality. The second inequality follows from Cauchy-Schwarz inequality, and the fact that for two tetrads σ_1 and σ_2 with zero overlapping indices $\mathbb{E}\left[\left(W_{\sigma_1} G^*_{\sigma_1,\tau} - \mathbb{E}\left[W_{\sigma_1} G^*_{\sigma_1,\tau} \right] \right) \left(W_{\sigma_2} G^*_{\sigma_2,\tau} - \mathbb{E}\left[W_{\sigma_2} G^*_{\sigma_2,\tau} \right] \right)^\top \right] = 0$. The last relationship follows from Assumption 4.1 which ensures that $\mathbb{W}$ is a compact subsets of $\Re^{d_\theta}$. Hence

$$\frac{1}{m_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{ \mathbb{E}\left[\left(W_\sigma G^*_{\sigma,\tau} - \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau} \right] \right)^2 \mid \omega_\sigma = 1 \right] \Pr\left[\omega_\sigma = 1 \right] \right\} = O_p\left(\rho_n \right)$$

The converge in probability to zero follows from the rate condition $n\rho_n \to \infty$ as $n \to \infty$. □

Lemma 6 (Trimming)**.** *Suppose the assumptions of Theorem 4.1 hold. For any* $\sigma(\{i,k,j,l\}) \in \mathbf{N}_{m_n}$, *let* $\sigma_{i_1 i_2} \in \{(i,k),(i,l),(j,k),(j,l)\}$,

$$\Psi_{0,\tau} - \Psi_0 = \frac{1}{\mathbb{E}\left[m_n^* \right]} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{ \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau} \right] - \mathbb{E}\left[W_\sigma G^*_\sigma \right] \right\} = o_p(1) \tag{A.4}$$

Proof. For any $\sigma(\{i,j,k,l\}) \in \mathbf{N}_{M_n}$ and $\sigma_{i_1 i_2} \in \{(i,k),(i,l),(j,k),(j,l)\}$, consider

$$\begin{aligned}
&\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{M_n}} W_\sigma \left\{ G^*_\sigma - G^*_{\sigma,\tau} \right\} \\
=&\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{M_n}} W_\sigma \left\{ \left(D^*_{\sigma_{13}} I^c_{\sigma_{13},\tau} - D^*_{\sigma_{14}} I^c_{\sigma_{14},\tau} \right) - \left(D^*_{\sigma_{23}} I^c_{\sigma_{23},\tau} - D^*_{\sigma_{24}} I^c_{\sigma_{24},\tau} \right) \right\}.
\end{aligned}$$

The last step follows from noticing that for any $\sigma_{i_1 i_2} \in \{(i,k),(i,l),(j,k),(j,l)\}$,

$$D^*_{\sigma_{i_1 i_2}} - D^*_{\sigma_{i_1 i_2},\tau} = D^*_{\sigma_{i_1 i_2}} \left\{ 1 - I_{\sigma_{i_1 i_2},\tau} \right\} = D^*_{\sigma_{i_1 i_2}} I^c_{\sigma_{i_1 i_2},\tau} \tag{A.5}$$

where $I^c_{\sigma_{i_1 i_2},\tau} = 1 - I_{\sigma_{i_1 i_2},\tau}$. For any $\epsilon > 0$, consider

$$
\begin{aligned}
&\Pr\left[\left|\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{M_n}} W_\sigma\left\{G^*_\sigma - G^*_{\sigma,\tau}\right\}\right| > \epsilon\right] \\
\leq&\epsilon^2\left(\frac{1}{n\rho_n^2}\right)\frac{1}{m_n}\sum_{\sigma\in\mathbf{N}_{M_n}}\mathbb{E}\left[\left(W_\sigma\left\{\left(D^*_{\sigma_{13}}I^c_{\sigma_{13},\tau} - D^*_{\sigma_{14}}I^c_{\sigma_{14},\tau}\right) - \left(D^*_{\sigma_{23}}I^c_{\sigma_{23},\tau} - D^*_{\sigma_{24}}I^c_{\sigma_{24},\tau}\right)\right\}\right)^2\right] \\
=&O\left(\frac{\tau_n}{n\rho_n}\right)
\end{aligned}
$$

where the inequality follows from Chebyshev's inequality and subsequently Cauchy-Schwarz inequality. The equality follows from

$$
\begin{aligned}
&\frac{1}{m_n}\sum_{\sigma\in\mathbf{N}_{M_n}}\mathbb{E}\left[\left(W_\sigma\left\{\left(D^*_{\sigma_{13}}I^c_{\sigma_{13},\tau} - D^*_{\sigma_{14}}I^c_{\sigma_{14},\tau}\right) - \left(D^*_{\sigma_{23}}I^c_{\sigma_{23},\tau} - D^*_{\sigma_{24}}I^c_{\sigma_{24},\tau}\right)\right\}\right)^2\right] \\
=&O\left(\frac{1}{m_n}\sum_{\sigma\in\mathbf{N}_{M_n}}\mathbb{E}\left[\mathbb{E}\left[W_\sigma W_\sigma^\top\left\{(G^*_\sigma)\right\}^2 \mid I^c_{\sigma_{13},\tau} = 1, I^c_{\sigma_{14},\tau} = 1, I^c_{\sigma_{24},\tau} = 1, I^c_{\sigma_{24},\tau} = 1\right]\right]\tau_n\right) \\
&= O\left(\rho_n\tau_n\right)
\end{aligned}
$$

where the first equality follows Assumption 4.4, which ensures that $\Pr\left[I^c_{ij,\tau} = 1\right] = \tau_n$. The second equality follows Assumption 4.1, and the fact that the sample average of the inner expectation across all tetrads is of order ρ_n. Therefore, the rate conditions $n\rho_n \to \infty$ and $\tau_n \to 0$ as $n \to \infty$ ensure that

$$
\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{M_n}} W_\sigma\left\{G^*_\sigma - G^*_{\sigma,\tau}\right\} = o_p(1).
$$

and consequently

$$
\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{\mathbb{E}\left[W_\sigma G^*_{\sigma,\tau}\right] - \mathbb{E}\left[W_\sigma G^*_\sigma\right]\right\} \quad = \quad o(1).
$$

□

Lemma 7 (Ψ_n Expansion). *Let Assumptions 3.1-4.3 hold. Then*

$$
\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma\widehat{G}^*_{\sigma,\tau} = \Psi_{n,\tau} + \Upsilon_{1,n\tau} - \Upsilon_{2,n\tau} + o_p(1)
$$

where

$$\Psi_{n,\tau} = \frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma G^*_{\sigma,\tau}$$

$$\Upsilon_{1,n\tau} = \frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma \left\{\left(D^*_{\sigma_{13},\tau}\frac{\widehat{f}_{x,\sigma_{13}}-f_{x,\sigma_{13}}}{f_{x,\sigma_{13}}} - D^*_{\sigma_{14},\tau}\frac{\widehat{f}_{x,\sigma_{14}}-f_{x,\sigma_{14}}}{f_{x,\sigma_{14}}}\right) - \left(D^*_{\sigma_{23},\tau}\frac{\widehat{f}_{x,\sigma_{23}}-f_{x,\sigma_{23}}}{f_{x,\sigma_{23}}} - D^*_{\sigma_{24},\tau}\frac{\widehat{f}_{x,\sigma_{24}}-f_{x,\sigma_{24}}}{f_{x,\sigma_{24}}}\right)\right\}$$

$$\Upsilon_{2,n\tau} = \frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma \left\{\left(D^*_{\sigma_{13},\tau}\frac{\widehat{f}_{vx,\sigma_{13}}-f_{vx,\sigma_{13}}}{f_{vx,\sigma_{13}}} - D^*_{\sigma_{14},\tau}\frac{\widehat{f}_{vx,\sigma_{14}}-f_{vx,\sigma_{14}}}{f_{vx,\sigma_{14}}}\right) - \left(D^*_{\sigma_{23},\tau}\frac{\widehat{f}_{vx,\sigma_{23}}-f_{vx,\sigma_{23}}}{f_{vx,\sigma_{23}}} - D^*_{\sigma_{24},\tau}\frac{\widehat{f}_{vx,\sigma_{24}}-f_{vx,\sigma_{24}}}{f_{vx,\sigma_{24}}}\right)\right\}.$$

Proof. Notice that

$$\begin{aligned}\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma \widehat{G}^*_{\sigma,\tau} &= \frac{1}{m_n\rho_n}\sum_{\sigma(\{i,j,k,l\})\in\mathbf{N}_{m_n}} W_\sigma\left\{\left(\widehat{D}_{\sigma_{13},\tau}-\widehat{D}_{\sigma_{14},\tau}\right)-\left(\widehat{D}_{\sigma_{23},\tau}-\widehat{D}_{\sigma_{24},\tau}\right)\right\}\\ &= (\widehat{\eta}_{13,\tau}-\widehat{\eta}_{14,\tau})-(\widehat{\eta}_{23,\tau}-\widehat{\eta}_{24,\tau})\end{aligned}$$

where for any $\sigma(\{i,k,j,l\})\in\mathbf{N}_{m_n}$ and $\sigma_{i_1i_2}\in\{(i,k),(i,l),(j,k),(j,l)\}$

$$\widehat{\eta}_{i_1i_2,\tau} \equiv \frac{1}{m_n\rho_n}\sum_{\sigma(\{i,j,k,l\})\in\mathbf{N}_{m_n}} W_\sigma\widehat{D}_{\sigma_{i_1i_2},\tau}.$$

Given the definition of $\widehat{D}_{\sigma_{i_1i_2},\tau}$, consider a second-order Taylor expansion of $\widehat{f}_{x,\sigma_{i_1i_2}}/\widehat{f}_{vx,\sigma_{i_1i_2}}$ around $f_{x,\sigma_{i_1i_2}}/f_{vx,\sigma_{i_1i_2}}$. The quadratic terms in the expansion involve 2nd order derivatives of $f_{x,\sigma_{i_1i_2}}/f_{vx,\sigma_{i_1i_2}}$ evaluated at $\bar{f}_{x,\sigma_{i_1i_2}}$ and $\bar{f}_{vx,\sigma_{i_1i_2}}$, where $\bar{f}_{x,\sigma_{i_1i_2}}$ lies between $\widehat{f}_{x,\sigma_{i_1i_2}}$ and $f_{x,\sigma_{i_1i_2}}$, and similarly, $\bar{f}_{vx,\sigma_{i_1i_2}}$ lies between $\widehat{f}_{vx,\sigma_{i_1i_2}}$ and $f_{vx,\sigma_{i_1i_2}}$. It follows from substituting a second-order Taylor expansion of $\widehat{f}_{x,\sigma_{i_1i_2}}/\widehat{f}_{vx,\sigma_{i_1i_2}}$ around $f_{x,\sigma_{i_1i_2}}/f_{vx,\sigma_{i_1i_2}}$ into $\widehat{\eta}_{i_1i_2,\tau}$

$$\widehat{\eta}_{i_1i_2,\tau} = \frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma D^*_{\sigma_{i_1i_2},\tau}\left\{1+\frac{\widehat{f}_{x,\sigma_{i_1i_2}}-f_{x,\sigma_{i_1i_2}}}{f_{x,\sigma_{i_1i_2}}}-\frac{\widehat{f}_{vx,\sigma_{i_1i_2}}-f_{vx,\sigma_{i_1i_2}}}{f_{vx,\sigma_{i_1i_2}}}\right\}+R_{i_1i_2,n},$$

where $R_{i_1i_2,n}$ denotes the reminder term. It follows after aggregating $\widehat{\eta}_{13,\tau}$, $\widehat{\eta}_{14,\tau}$, $\widehat{\eta}_{23,\tau}$, and $\widehat{\eta}_{24,\tau}$ that

$$\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma\widehat{G}^*_{\sigma,\tau} = \Psi_{n,\tau}+\Upsilon_{1,n\tau}-\Upsilon_{1,n\tau}+R_n \tag{A.6}$$

where $R_n = R_{13,n} - R_{14,n} - R_{23,n} + R_{24,n}$. The proof is complete if we show that $R_n = o_p(1)$. Let

$$\begin{aligned}\widetilde{f}_{x,\sigma_{i_1 i_2}} &= \widehat{f}_{x,\sigma_{i_1 i_2}} - f_{x,\sigma_{i_1 i_2}}\\ \widetilde{f}_{vx,\sigma_{i_1 i_2}} &= \widehat{f}_{vx,\sigma_{i_1 i_2}} - f_{vx,\sigma_{i_1 i_2}}.\end{aligned}$$

The first component of R_n is

$$\begin{aligned}&\left|\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma\left\{\frac{D^*_{\sigma_{13},\tau}}{\bar{f}^2_{vx,\sigma_{13}}}\left(\widetilde{f}_{vx,\sigma_{13}}\right)^2 - \frac{D^*_{\sigma_{14},\tau}}{\bar{f}^2_{vx,\sigma_{14}}}\left(\widetilde{f}_{vx,\sigma_{14}}\right)^2 - \frac{D^*_{\sigma_{23},\tau}}{\bar{f}^2_{vx,\sigma_{23}}}\left(\widetilde{f}_{vx,\sigma_{23}}\right)^2 + \frac{D^*_{\sigma_{24},\tau}}{\bar{f}^2_{vx,\sigma_{24}}}\left(\widetilde{f}_{vx,\sigma_{24}}\right)^2\right\}\right|\\ \leq&\underline{B}_{vx}^{-2}\left(\sup_{(v_{12},x_{12})\in\mathbb{S}_{vx}}|\widehat{f}_{vx,12} - f_{vx,12}|\right)^2\left(\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma G^*_{\sigma,\tau}\right)\\ =&O_p(1)\left(\sup_{(v_{12},x_{12})\in\mathbb{S}_{vx}}|\widehat{f}_{vx,12} - f_{vx,12}|\right)^2\\ =&O_p\left(\left(\frac{\log n}{nh^{L+1}}\right)\right).\end{aligned}$$

where the first inequality follows from Assumptions 4.2. The first equality follows from Lemma 5. The last equality follows from the from the uniform rate of convergence of the kernel estimator in Lemma 2. The remaining component of R_n is

$$\begin{aligned}&\left|\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma G^*_{\sigma,\tau}\left\{\frac{D^*_{\sigma_{13},\tau}}{\bar{f}^2_{vx,\sigma_{13}}}\widetilde{f}_{vx,\sigma_{13}}\widetilde{f}_{x,\sigma_{13}} - \frac{D^*_{\sigma_{14},\tau}}{\bar{f}^2_{vx,\sigma_{14}}}\widetilde{f}_{vx,\sigma_{14}}\widetilde{f}_{x,\sigma_{14}} - \frac{D^*_{\sigma_{23},\tau}}{\bar{f}^2_{vx,\sigma_{23}}}\widetilde{f}_{vx,\sigma_{23}}\widetilde{f}_{x,\sigma_{23}} + \frac{D^*_{\sigma_{24},\tau}}{\bar{f}^2_{vx,\sigma_{24}}}\widetilde{f}_{vx,\sigma_{24}}\widetilde{f}_{x,\sigma_{24}}\right\}\right|\\ \leq&\underline{B}_{vx}^{-2}\sup_{(v_{12},x_{12})\in\mathbb{S}_{vx}}\mid\widehat{f}_{vx,12} - f_{vx,12}\mid\sup_{x_{12}\in\mathbb{S}_x}\mid\widehat{f}_{x,12} - f_{x,12}\mid\left[\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma G^*_{\sigma,\tau}\right]\\ =&O_p(1)\left(\sup_{(v_{12},x_{12})\in\mathbb{S}_{vx}}\mid\widehat{f}_{vx,12} - f_{vx,12}\mid\right)\left(\sup_{x_{12}\in\mathbb{S}_x}\mid\widehat{f}_{x,12} - f_{x,12}\mid\right).\\ =&O_p\left(\alpha_{1,n}\alpha_{2,n}\right).\end{aligned}$$

The result follows from the uniform rates of convergence in Lemma 2. □

Lemma 8 (First-stage Estimator)**.** *Let $\phi = \phi(\{i,j,k,l,s,p\})$ denotes the 6-tuples in $\mathcal{N}_n$. The set of all 6-tuples is denoted by $\mathbf{N}_{M_n}$ where $M_n = \binom{n}{6}^{-1}$. The following representation as U-statistics* *of order 6 holds.*

$$
\begin{aligned}
\Upsilon_{1,n\tau} &= \frac{1}{M_n\rho_n}\sum_{\phi\in\mathbf{N}_{m_n}}\psi_{x,\phi n\tau} + o_p(1)\\
\Upsilon_{2,n\tau} &= \frac{1}{M_n\rho_n}\sum_{\phi\in\mathbf{N}_{m_n}}\psi_{vx,\phi n\tau} + o_p(1)
\end{aligned}
$$

where

$$
\begin{aligned}
\psi_{x,\phi n\tau} = W_\phi \Bigg\{ &\left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}}\overline{K}_{xh,\phi_{56}}(x_{\phi_{13}}) - \frac{D^*_{\phi_{14},\tau}}{f_{x,\phi_{14}}}\overline{K}_{xh,\phi_{56}}(x_{\phi_{14}})\right)\\
&\qquad - \left(\frac{D^*_{\phi_{23},\tau}}{f_{x,\phi_{23}}}\overline{K}_{xh,\phi_{56}}(x_{\phi_{23}}) - \frac{D^*_{\phi_{24},\tau}}{f_{x,\phi_{24}}}\overline{K}_{xh,\phi_{56}}(x_{\phi_{24}})\right)\Bigg\}
\end{aligned}
$$

and

$$
\begin{aligned}
\psi_{vx,\phi n\tau} = W_\phi \Bigg\{ &\left(\frac{D^*_{\phi_{13},\tau}}{f_{vx,\phi_{13}}}\overline{K}_{vh,\phi_{56}}(v_{\phi_{13}}, x_{\phi_{13}}) - \frac{D^*_{\phi_{14},\tau}}{f_{vx,\phi_{14}}}\overline{K}_{vxh,\phi_{56}}(v_{\phi_{14}}, x_{\phi_{14}})\right)\\
&\qquad - \left(\frac{D^*_{\phi_{23},\tau}}{f_{vx,\phi_{23}}}\overline{K}_{vxh,\phi_{56}}(v_{\phi_{23}}, x_{\phi_{23}}) - \frac{D^*_{\phi_{24},\tau}}{f_{vx,\phi_{24}}}\overline{K}_{vxh,\phi_{56}}(v_{\phi_{24}}, x_{\phi_{24}})\right)\Bigg\}.
\end{aligned}
$$

The following notation is used for the kernel

$$
\begin{aligned}
\overline{K}_{xh,\phi_{i_1j_2}}(x_{\phi_{i_1i_2}}) &= K_{xh,\phi_{i_1j_2}}(x_{\phi_{i_1i_2}}) - \mathbb{E}\left[K_{xh,\phi_{i_1j_2}}(x_{\phi_{i_1i_2}})\right]\\
\overline{K}_{vxh,\phi_{i_1j_2}}(v_{\phi_{i_1i_2}}, x_{\phi_{i_1i_2}}) &= K_{vxh,\phi_{i_1j_2}}(x_{\phi_{i_1i_2}}, x_{\phi_{i_1i_2}}) - \mathbb{E}\left[K_{vxh,\phi_{i_1j_2}}(v_{\phi_{i_1i_2}}, x_{\phi_{i_1i_2}})\right].
\end{aligned}
$$

Proof. We show the result for $\Upsilon_{1,n\tau}$. Given $\sigma(\{i,j,k,l\}) \in \mathcal{N}$, let $\sigma_{i_1i_2} \in \{(i,k),(i,l),(j,k),(j,l)\}$, we consider instead

$$
\widehat{f}_x(x_{\sigma_{i_1i_2}}) - f_x(x_{\sigma_{i_1i_2}}) = \frac{1}{(n-5)(n-6)}\sum_{s,p\neq\sigma}\left(K_{xh,sp}(x_{\sigma_{i_1i_2}}) - \mathbb{E}\left[K_{xh,sp}(x_{\sigma_{i_1i_2}})\right]\right) + o(n^\delta)
$$

As in the proof of Lemma 7, notice that $\Upsilon_{1,n\tau} = \left(\tilde{H}_{x,\sigma_{13}} - \tilde{H}_{x,\sigma_{14}}\right) - \left(\tilde{H}_{x,\sigma_{23}} - \tilde{H}_{x,\sigma_{24}}\right)$, with

$$
\begin{aligned}
\tilde{H}_{x,\sigma_{i_1i_2}} &= \frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}} W_\sigma \frac{D^*_{\sigma_{i_1i_2},\tau}}{f_{x,\sigma_{i_1i_2}}}\left(\widehat{f}_{x,\sigma_{i_1i_2}} - f_{x,\sigma_{i_1i_2}}\right)\\
&= \frac{1}{M_n\rho_n}\sum_{\phi\in\mathbf{N}_{M_n}} W_\phi \frac{D^*_{\phi_{i_1i_2},\tau}}{f_{x,\phi_{i_1i_2}}}\left\{K_{xh,\phi_{i_5i_6}}(x_{\phi_{i_1i_2}}) - \mathbb{E}\left[K_{xh,\phi_{i_5i_6}}(x_{\phi_{i_1i_2}})\right]\right\} + o_p(n^\delta)
\end{aligned}
$$

since

$$
\frac{1}{m_n(n-5)(n-5)\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\sum_{s,p\neq\sigma} W_\sigma \frac{D^*_{\sigma_{i_1i_2},\tau}}{f_{x,\sigma_{i_1i_2}}}\left\{\mathbb{E}\left[K_{xh,sp}(x_{\sigma_{i_1i_2}})\right] - f_{x,\sigma_{i_1i_2}}\right\} = o_p(n^\delta)
$$

and the fact that the difference between a $U-$statistics and V-statistics is asymptotically negligible, see e.g., Newey and McFadden (1994) Chapter 8.2 and Serfling (2009) Chapter 5.7.3. Hence

$$\Upsilon_{1,n\tau} = \frac{1}{M_n \rho_n} \sum_{\phi \in \mathbf{N}_{M_n}} W_\phi \Bigg\{ \bigg(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{13}}) - \frac{D^*_{\phi_{14},\tau}}{f_{x,\phi_{14}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{14}}) \bigg) - \bigg(\frac{D^*_{\phi_{23},\tau}}{f_{x,\phi_{23}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{23}}) - \frac{D^*_{\phi_{24},\tau}}{f_{x,\phi_{24}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{24}}) \bigg) \Bigg\} + o_p(n^\delta).$$

where ϕ is defined as a function that maps the 6-tuples to the index set $\mathbf{N}_{M_n} = \{1, \cdots, M_n\}$ where $M_n = \binom{n}{6}$ denotes the total number of 6-tuples with distinct indices $i, j, k, l, s, p \in \mathbf{N}_n$ in a network of size n.

The result for $\Upsilon_{2,n\tau}$ follows from using

$$\widehat{f}_{vx}(v_{\sigma_{i_1 i_2}}, x_{\sigma_{i_1 i_2}}) - f_{vx}(v_{\sigma_{i_1 i_2}}, x_{\sigma_{i_1 i_2}}) = \frac{1}{(n-5)(n-6)} \sum_{s,p \neq \sigma} \overline{K}_{vxh,sp}(v_{\sigma_{i_1 i_2}}, x_{\sigma_{i_1 i_2}}) + o_p(n^\delta).$$

□

Lemma 9 (U-statistic)**.** *The estimator*

$$\frac{1}{\rho_n} \widehat{\Psi}_{n,\tau}$$

can be represented as a sixth-order U-statistic. That is

$$\frac{1}{\rho_n} \widehat{\Psi}_{n,\tau} = \frac{1}{M_n \rho_n} \sum_{\phi \in \mathbf{N}_{M_n}} \psi_{\phi n\tau} + o_p(1),$$

where $\psi_{\phi n\tau} = \psi_{0,\phi n\tau} + \psi_{x,\phi n\tau} - \psi_{vx,\phi n\tau}$, $\psi_{0,\phi n\tau} = W_\phi G^*_{\phi,\tau}$, and $M_n = \binom{n}{6}$.

Proof. The proof resembles the one of Lemma 8. □

A.3 Consistency

Proof of Theorem 4.1

Proof. The estimator is defined as $\widehat{\theta}_n = \widehat{\Gamma}_n^{-1} \times \widehat{\Psi}_{n,\tau}$ with

$$\begin{aligned}\widehat{\Gamma}_n &= \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma W_\sigma^\top \\ \widehat{\Psi}_{n,\tau} &= \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma \widehat{G}_{\sigma,\tau}^*.\end{aligned}$$

First, we will show that $\widehat{\Gamma}_n \xrightarrow{p} \Gamma_0$ and $\widehat{\Psi}_{n,\tau} \xrightarrow{p} \Psi_0$. The result will follow from Assumption 3.4, the use of the Continuous Mapping Theorem and Slutsky's Theorem.

Part 1.

$$\begin{aligned}\widehat{\Gamma}_{n,\tau} - \Gamma_0 &= \frac{1}{m_n^*} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma W_\sigma^\top - \frac{1}{\mathbb{E}\left[m_n^*\right]} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma W_\sigma^\top \\ &= \left\{\frac{\mathbb{E}\left[m_n^*\right]}{m_n^*} - 1\right\} \frac{1}{\mathbb{E}\left[m_n^*\right]} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma W_\sigma^\top + \frac{1}{\mathbb{E}\left[m_n^*\right]} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{W_\sigma W_\sigma^\top - \mathbb{E}\left[W_\sigma W_\sigma^\top\right]\right\}.\end{aligned}$$

We will show that each component converges in probability to zero. Notice that for any $\epsilon > 0$,

$$\begin{aligned}&\Pr\left(\left|\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{W_\sigma W_\sigma^\top - \mathbb{E}\left[W_\sigma W_\sigma^\top\right]\right\}\right| > \epsilon\right) \\ &\leq \frac{1}{\epsilon^2 \left(m_n \rho_n\right)^2} \mathbb{E}\left[\left(\sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma W_\sigma^\top - \mathbb{E}\left[W_\sigma W_\sigma^\top\right]\right)^2\right] \\ &\leq \frac{n^3}{\epsilon^2 \left(m_n \rho_n\right)^2} \sum_{\sigma \in \mathbf{N}_{m_n}} \mathbb{E}\left[\left(W_\sigma W_\sigma^\top - \mathbb{E}\left[W_\sigma W_\sigma^\top\right]\right)^2\right] \\ &= O\left(\frac{1}{n \rho_n^2 m_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \Pr\left[W_\sigma \neq \mathbf{0}\right]\right) \\ &= O_p\left(\frac{1}{n \rho_n}\right)\end{aligned}$$

The first inequality follows from Chebyshev's inequality. The second inequality follows from Cauchy-Schwarz inequality and the fact that when σ_1 and σ_2 have zero overlapping indices we have that $\mathbb{E}\left[\left(W_{\sigma_1} W_{\sigma_1}^\top - \mathbb{E}\left[W_{\sigma_1} W_{\sigma_1}^\top\right]\right)\left(W_{\sigma_2} W_{\sigma_2}^\top - \mathbb{E}\left[W_{\sigma_2} W_{\sigma_2}^\top\right]\right)\right] = 0$. The equality results from the identification condition in Eq. (A.1), which ensures that

$$\frac{1}{m_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\mathbb{E}\left[\left(W_\sigma W_\sigma^\top-\mathbb{E}\left[W_\sigma W_\sigma^\top\right]\right)^2\right]=O_p\left(\rho_n\right).$$

To prove the second result. Notice that $\frac{1}{\mathbb{E}[m_n^*]}\sum_{\sigma\in\mathbf{N}_{m_n}}W_\sigma W_\sigma^\top=O_p(1)$ and $\rho_n-\frac{m_n^*}{m_n}\to 0$, and thus $\frac{\mathbb{E}[m_n^*]}{m_n^*}-1\to 0$ as $n\to\infty$.

Part 2: Consider

$$\begin{aligned}\widehat{\Psi}_{n,\tau}-\Psi_0&=\frac{1}{m_n^*}\sum_{\sigma\in\mathbf{N}_{m_n}}W_\sigma\widehat{G}_{\sigma,\tau}^*-\frac{1}{\mathbb{E}\left[m_n^*\right]}\sum_{\sigma\in\mathbf{N}_{m_n}}\mathbb{E}\left[W_\sigma G_\sigma^*\right]\\&=\left\{\frac{\mathbb{E}\left[m_n^*\right]}{m_n^*}-1\right\}\frac{1}{\mathbb{E}\left[m_n^*\right]}\sum_{\sigma\in\mathbf{N}_{m_n}}W_\sigma\widehat{G}_{\sigma,\tau}^*+\frac{1}{\mathbb{E}\left[m_n^*\right]}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{W_\sigma\widehat{G}_{\sigma,\tau}^*-\mathbb{E}\left[W_\sigma G_\sigma^*\right]\right\}.\end{aligned}$$

We will show that each term convergences in probability to zero. To convergence of the first term follows from similar steps as those in Part 1. Consider the second term in the right-hand side

$$\begin{aligned}&\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{W_\sigma\widehat{G}_{\sigma,\tau}^*-\mathbb{E}\left[W_\sigma G_\sigma^*\right]\right\}\\&\qquad=\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{W_\sigma\widehat{G}_{\sigma,\tau}^*-\mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\}&\text{(A.7)}\\&\qquad\quad+\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{\mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]-\mathbb{E}\left[W_\sigma G_\sigma^*\right]\right\}&\text{(A.8)}\end{aligned}$$

We need to show that (A.7) and (A.8) converge in probability to zero. Lemma 6 ensures that (A.8) converge in probability to zero. We proceed to show the result for (A.7).

Part 2.1 Lemma 7 ensures that the term given by (A.7) can be written as

$$\begin{aligned}&\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{W_\sigma\widehat{G}_{\sigma,\tau}^*-\mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\}\\=&\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{W_\sigma G_{\sigma,\tau}^*-\mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\}+\Upsilon_{1,n\tau}-\Upsilon_{2,n\tau}+o_p(1).\end{aligned}$$

Moreover, Lemma 5 shows that

$$\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{W_\sigma G_{\sigma,\tau}^*-\mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\}=o_p(1).$$

The result follows from proving that $\Upsilon_{1,n\tau}-\Upsilon_{2,n\tau}=o_p(1)$. We will show that $\Upsilon_{1,n\tau}=o_p(1)$, the result for $\Upsilon_{2,n\tau}$ follows from analogous steps. Notice that $\Upsilon_{1,n\tau}$ is a U-statistic of second-order that depend on the initial kernel estimators $\widehat{f}_{vx,i_1 i_2}$. It follows from Lemma 8 that after plugging-in the kernel estimator $\widehat{f}_{vx,i_1 i_2}$, $\Upsilon_{1,n\tau}$ can be written as the following sixth-order U-statistic

$$\Upsilon_{1,n\tau} = \frac{1}{M_n \rho_n} \sum_{\phi \in \mathbf{N}_{M_n}} W_\phi \left\{ \left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{13}}) - \frac{D^*_{\phi_{14},\tau}}{f_{x,\phi_{14}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{14}}) \right) \right.$$
$$\left. - \left(\frac{D^*_{\phi_{23},\tau}}{f_{x,\phi_{23}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{23}}) - \frac{D^*_{\phi_{24},\tau}}{f_{x,\phi_{24}}} \overline{K}_{xh,\phi_{56}}(x_{\phi_{24}}) \right) \right\} + o_p(1).$$

Notice that each element in $\Upsilon_{1,n\tau}$ has conditional mean zero as the kernel $\overline{K}_{xh,\phi_{56}}(x_{\phi_{13}}) \equiv K_{xh,\phi_{56}}(x_{\phi_{13}}) - \mathbb{E}\left[K_{xh,\phi_{56}}(x_{\phi_{13}})\right]$. For any $\epsilon > 0$, consider

$$\begin{aligned}
&\Pr\left[\left| \frac{1}{M_n \rho_n} \sum_{\phi \in \mathbf{N}_{M_n}} W_\phi \left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}} \right) \overline{K}_{xh,\phi_{56}}(x_{\phi_{13}}) \right| > \epsilon \right] \\
&\leq \frac{1}{\epsilon^2 M_n^2 \rho_n^2} \mathbb{E}\left[\left\{ \sum_{\phi \in \mathbf{N}_{M_n}} W_\phi \left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}} \right) \overline{K}_{xh,\phi_{56}}(x_{\phi_{13}}) \right\}^2 \right] \\
&\leq \frac{1}{\epsilon^2 n \rho_n M_n} \sum_{\phi \in \mathbf{N}_{M_n}} \mathbb{E}\left[W_\phi W_\phi^\top \left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}} \right)^2 \overline{K}_{xh,\phi_{56}}(x_{\phi_{13}})^2 \right] \\
&= O\left(\frac{1}{n \rho_n} \sum_{\phi \in \mathbf{N}_{M_n}} \mathbb{E}\left[W_\phi W_\phi^\top \left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}} \right)^2 K_{xh,\phi_{56}}(x_{\phi_{13}})^2 \right] \right) \\
&\leq O\left(\frac{\overline{\kappa}_x \overline{E}_x}{n h_n^{d_\theta}} \right)
\end{aligned}$$

the first two inequalities follow from similar steps as those in Part 1. The last inequality follows from noticing that

$$
\begin{aligned}
&\mathbb{E}\left[\left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}}\right)^2 K_{xh,\phi_{56}}(x_{\phi_{13}})^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}}\right)^2 \mid x_{\phi_{13}}, x_{\phi_{56}}\right] K_{xh,\phi_{56}}(x_{\phi_{13}})^2\right] \\
&\leq \overline{\kappa}_x \left(\frac{1}{h_n^{d_\theta}}\right)^2 \int \mathbb{E}\left[\left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}}\right)^2 \mid x_{\phi_{13}}\right] K_{xh,\phi_{56}}(x_{\phi_{13}}) f_{x,\phi_{56}} f_{x,\phi_{13}} dX_{\phi_{56}} dX_{\phi_{13}} \\
&= \overline{\kappa}_x \left(\frac{1}{h_n^{d_\theta}}\right)^2 \int \left\{\mathbb{E}\left[\left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}}\right)^2 \mid x_{\phi_{13}}\right] \times \int K_{xh,\phi_{56}}(x_{\phi_{13}}) f_{x,\phi_{56}} dX_{\phi_{56}}\right\} f_{x,\phi_{13}} dX_{\phi_{13}} \\
&= \overline{\kappa}_x \left(\frac{1}{h_n^{d_\theta}}\right) \int \left\{\mathbb{E}\left[\left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}}\right)^2 \mid x_{\phi_{13}}\right] \times f_{x,\phi_{13}}\right\} f_{x,\phi_{13}} dX_{\phi_{13}} + o\left(n^{\delta}\right) \\
&\leq \overline{\kappa}_x \overline{E}_x \left(\frac{1}{h_n^{d_\theta}}\right)
\end{aligned}
$$

where the first inequality uses Assumption 3.1 and the fact that the kernel is bounded by Assumptions 4.3. The last equality follows from Assumption 3.1 and the fact that the kernel is a bias-reducing kernel by by Assumptions 4.3. Therefore

$$
\frac{1}{M_n \rho_n} \sum_{\phi \in \mathbf{N}_{M_n}} W_\phi \left(\frac{D^*_{\phi_{13},\tau}}{f_{x,\phi_{13}}}\right) \left\{K_{xh,\phi_{56}}(x_{\phi_{13}}) - \mathbb{E}\left[K_{xh,\phi_{56}}(x_{\phi_{13}})\right]\right\} = o_p(1). \tag{A.9}
$$

Consequently, $\Upsilon_{1,n\tau} = o_p(1)$. It follows from similar steps that $\Upsilon_{2,n\tau} = o_p(1)$. □

A.4 Asymptotic Distribution

Lemma 10. *Suppose the assumptions of Theorem 4.2. Let*

$$
\begin{aligned}
\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n) &\equiv \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \mathbb{E}\left[W_\sigma \widehat{G}^*_{\sigma,\tau} \mid \mathbf{Z}_n\right] \\
\Psi_{0,\tau} &\equiv \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \mathbb{E}\left[W_\sigma G^*_{\sigma,\tau}\right]
\end{aligned}
$$

Then

$$
\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n) - \Psi_{0,\tau} = o_p(1)
$$

$$
\mathbb{E}\left[\widehat{\Psi}_{n,\tau} \mid \boldsymbol{Z}_n\right] - \Psi_{0\tau} = \frac{1}{\rho_n}\binom{n}{6}^{-1} \sum_{\phi \in \mathbf{N}_{M_n}} \left\{\mathbb{E}\left[\psi_{\phi n\tau} \mid \boldsymbol{Z}_n\right] - \mathbb{E}\left[\psi_{0,\phi\tau}\right]\right\} = o_p(1)
$$

Proof. It follows from Lemma 9 that

$$\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n) - \Psi_{0,\tau} = \frac{1}{\rho_n}\binom{n}{6}^{-1} \sum_{\phi \in \mathbf{N}_{M_n}} h_{\phi,\tau}$$

where $h_{\phi,\tau} = \mathbb{E}\left[\psi_{\phi n\tau} \mid \boldsymbol{Z}_n\right] - \mathbb{E}\left[\psi_{0,\phi\tau}\right]$. Let

$$\begin{aligned}
\tilde{h}_{1,\tau}(Z_{i_1}) &= \mathbb{E}\left[\psi_{\phi n\tau} \mid Z_{i_1}\right] - \mathbb{E}\left[\psi_{0,\phi\tau}\right] \\
\tilde{h}_{2,\tau}(Z_{i_1}, Z_{i_2}) &= \mathbb{E}\left[\psi_{\phi n\tau} \mid Z_{i_1}, Z_{i_2}\right] - \mathbb{E}\left[\psi_{0,\phi\tau}\right] \\
&\vdots \\
\tilde{h}_{6,\tau}(Z_{i_1}, \cdots, Z_{i_6}) &= \mathbb{E}\left[\psi_{\phi n\tau} \mid Z_{i_1}, \cdots, Z_{i_6}\right] - \mathbb{E}\left[\psi_{0,\phi\tau}\right]
\end{aligned}$$

and

$$\begin{aligned}
g_{1,\tau}(Z_{i_1}) =& \tilde{h}_{1,\tau}(Z_{i_1}) \\
g_{2,\tau}(Z_{i_1}, Z_{i_2}) =& \tilde{h}_{2,\tau}(Z_{i_1}, Z_{i_2}) - \sum_{1 \le k \le 2} g_{1,\tau}(Z_{i_k}) \\
g_{3,\tau}(Z_{i_1}, Z_{i_2}, Z_{i_3}) =& \tilde{h}_{3,\tau}(Z_{i_1}, Z_{i_2}, Z_{i_3}) - \sum_{1 \le k_1 < k_2 \le 3} g_{2,\tau}(Z_{i_{k_1}}, Z_{i_{k_2}}) - \sum_{1 \le k \le 3} g_{1,\tau}(Z_{i_k}) \\
g_{6,\tau}(Z_{i_1}, \cdots, Z_{i_6}) =& \tilde{h}_{6,\tau}(Z_{i_1}, \cdots, Z_{i_6}) - \sum_{1 \le k_1 < \cdots < k_5 \le 6} g_{5,\tau}(Z_{i_{k_1}}, \cdots, Z_{i_{k_5}}) - \cdots - \\
& - \sum_{1 \le k_1 < k_2 \le 6} g_{2,\tau}(Z_{i_{k_1}}, Z_{i_{k_2}}) - \sum_{1 \le k \le 6} g_{1,\tau}(Z_{i_k})
\end{aligned}$$

It follows from a Hoeffding decomposition, that $\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n) - \Psi_{0,\tau}$ can be written as

$$\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n) - \Psi_{0,\tau} = \frac{1}{\rho_n} \sum_{c=1}^{6} \binom{6}{c} \binom{n}{c}^{-1} S_{cn}$$

with

$$S_{cn,\tau} = \sum_{1 \le i_1 < \cdots < i_c \le n} g_c\left(Z_{i_1}, \cdots, Z_{i_c}\right)$$

where for each $c = 1, \cdots, 6$, $S_{cn,\tau}$ are uncorrelated. Thus,

$$\mathbb{V}\left(\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n)\right) = \frac{1}{\rho_n^2} \sum_{c=1}^{6} \binom{6}{c}^2 \binom{n}{c}^{-2} \mathbb{V}\left(S_{cn,\tau}\right)$$

with $\mathbb{V}\left(S_{cn,\tau}\right)=\tilde{\Sigma}_{cn}$. Consequently,

$$\begin{aligned}
&\mathbb{V}\left(\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n)\right)\\
=&\frac{1}{\rho_n^2}\left\{6^2\binom{n}{1}^{-1}\tilde{\Sigma}_{1n}+15^2\binom{n}{2}^{-1}\tilde{\Sigma}_{2n}+20^2\binom{n}{3}^{-1}\tilde{\Sigma}_{3n}+15^2\binom{n}{4}^{-1}\tilde{\Sigma}_{4n}+6^2\binom{n}{5}^{-1}\tilde{\Sigma}_{5n}+\binom{n}{6}^{-1}\tilde{\Sigma}_{6n}\right\}\\
=&\frac{1}{\rho_n^2}\left\{6^2\binom{n}{1}^{-1}\tilde{\Sigma}_{1n}+15^2\binom{n}{2}^{-1}\tilde{\Sigma}_{2n}+20^2\binom{n}{3}^{-1}\tilde{\Sigma}_{3n}+15^2\binom{n}{4}^{-1}\tilde{\Sigma}_{4n}+6^2\binom{n}{5}^{-1}\tilde{\Sigma}_{5n}+\binom{n}{6}^{-1}\tilde{\Sigma}_{6n}\right\}\\
=&O\left(\frac{1}{n\rho_n}\right)+o\left(\frac{1}{n\rho_n}\right)
\end{aligned}$$

which uses the fact that $\tilde{\Sigma}_{1n}=O\left(\rho_n\right)$.

It follows then that for any $\epsilon>0$

$$\Pr\left[\left|\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n)-\Psi_{0\tau}\right|>\epsilon\right]\leq\frac{1}{\epsilon^2}\,\mathbb{V}\left(\widehat{\Psi}_{n,\tau}(\boldsymbol{Z}_n)\right)=O\left(\frac{1}{n\rho_n}\right)$$

which converges in probability to zero as $n\to\infty$. □

Lemma 11 (Hájek Projection)**.** *Suppose the assumptions of Theorem 4.2. Let*

$$\mathcal{S}_{n,\tau}\equiv\frac{1}{m_n\rho_n}\sum_{\sigma\in\mathbf{N}_{m_n}}\left\{W_\sigma\widehat{G}^*_{\sigma,\tau}-\mathbb{E}\left[W_\sigma\widehat{G}^*_{\sigma,\tau}\mid\mathbf{Z}_n\right]\right\}.$$

The, Hájek Projection of $\mathcal{S}_{n,\tau}$ into arbitrary functions (Z_{ij},U_{ij}) is given by

$$\mathcal{S}_{n,\tau}=\frac{15}{\rho_n}\binom{n}{2}^{-1}\sum_{i<j}\zeta_{ij,\tau}+o_p(1)$$

where

$$\zeta_{ij,\tau}\equiv\mathbb{E}\left[\widetilde{\psi}_{\phi n\tau}\mid Z_{ij},U_{ij}\right]$$

and $\widetilde{\psi}_{\phi n\tau}=\psi_{\phi n\tau}-\mathbb{E}\left[\psi_{\phi n\tau}\mid\boldsymbol{Z}_n\right]$.

Proof. It follows from Lemma 9 that $S_{n,\tau}$ can be written as a U-statistic. In particular,

$$S_{n,\tau}=\frac{1}{\rho_n}\left\{\widehat{\Psi}_{n,\tau}-\mathbb{E}\left[\widehat{\Psi}_{n,\tau}\mid\mathbf{Z}_n\right]\right\}=\frac{1}{\rho_n}\binom{n}{6}^{-1}\sum_{\phi\in\mathbf{N}_{M_n}}\widetilde{\psi}_{\phi n\tau}$$

where

$$\widetilde{\psi}_{\phi n\tau}=\widetilde{\psi}_{0,\phi n\tau}+\widetilde{\psi}_{x,\phi n\tau}-\widetilde{\psi}_{vx,\phi n\tau}$$

and $\widetilde{\psi}_{\phi n\tau} = \psi_{\phi n\tau} - \mathbb{E}\left[\psi_{\phi n\tau} \mid \boldsymbol{Z}_n\right]$, $\widetilde{\psi}_{0,\phi n\tau} = \psi_{0,\phi n\tau} - \mathbb{E}\left[\psi_{0,\phi n\tau} \mid \boldsymbol{Z}_n\right]$, $\widetilde{\psi}_{x,\phi n\tau} = \psi_{x,\phi n\tau} - \mathbb{E}\left[\psi_{x,\phi n\tau} \mid \boldsymbol{Z}_n\right]$, $\widetilde{\psi}_{vx,\phi n\tau} = \psi_{vx,\phi n\tau} - \mathbb{E}\left[\psi_{vx,\phi n\tau} \mid \boldsymbol{Z}_n\right]$.

Part 1: Variance of $\mathcal{S}_{n,\tau}$. For any two indices ϕ_1 and ϕ_2, let

$$\Sigma_{cn} \equiv \mathbb{C}\left(\widetilde{\psi}_{\phi_1 n\tau}, \widetilde{\psi}_{\phi_2 n\tau}\right)$$

denote the covariance of $\widetilde{\psi}_{\phi_1 n\tau}$ and $\widetilde{\psi}_{\phi_2 n\tau}$ when the 6-tuples ϕ_1 and ϕ_2 have $c = 0, 1, \cdots, 6$ indices in common. Notice that $\mathcal{S}_{n,\tau}$ has degeneracy of order 1, which is a consequence of Assumption 3.1, and the conditional mean zero, $\mathbb{E}\left[\mathcal{S}_{n,\tau} \mid \boldsymbol{Z}_n\right] = 0$. In particular, notice that for any ϕ_1 and ϕ_2 with $c = 0, 1$ indices in common

$$\begin{aligned}
\mathbb{C}\left(\widetilde{\psi}_{0,\phi_1 n\tau}, \widetilde{\psi}_{0,\phi_2 n\tau}\right) &= 0 \\
\mathbb{C}\left(\widetilde{\psi}_{x,\phi_1 n\tau}, \widetilde{\psi}_{x,\phi_2 n\tau}\right) &= 0 \\
\mathbb{C}\left(\widetilde{\psi}_{vx,\phi_1 n\tau}, \widetilde{\psi}_{vx,\phi_2 n\tau}\right) &= 0 \\
\mathbb{C}\left(\widetilde{\psi}_{0,\phi_1 n\tau}, \widetilde{\psi}_{x,\phi_2 n\tau}\right) &= 0 \\
\mathbb{C}\left(\widetilde{\psi}_{0,\phi_1 n\tau}, \widetilde{\psi}_{vx,\phi_2 n\tau}\right) &= 0 \\
\mathbb{C}\left(\widetilde{\psi}_{x,\phi_1 n\tau}, \widetilde{\psi}_{vx,\phi_2 n\tau}\right) &= 0.
\end{aligned}$$

Which ensures that, for any ϕ_1 and ϕ_2 the covariances $\Sigma_{0n} = 0$ and $\Sigma_{1n} = 0$. It follows then that

$$\begin{aligned}
\mathbb{V}\left(\mathcal{S}_{n,\tau}\right) &= \frac{1}{\rho_n^2}\binom{n}{6}^{-2} \sum_{\phi_1 \in \mathbf{N}_{M_n}} \sum_{\phi_2 \in \mathbf{N}_{M_n}} \mathbb{C}\left(\widetilde{\psi}_{\phi_1 n\tau}, \widetilde{\psi}_{\phi_2 n\tau}\right) \\
&= \frac{1}{\rho_n^2}\binom{n}{6}^{-1} \sum_{c=2}^{6} \binom{6}{c}\binom{n-6}{6-c}\Sigma_{cn} \\
&\simeq \frac{1}{\rho_n^2}\left\{15^2\binom{n}{2}^{-1}\Sigma_{2n} + 20^2\binom{n}{3}^{-1}\Sigma_{3n} + 15^2\binom{n}{4}^{-1}\Sigma_{4n} + 6^2\binom{n}{5}^{-1}\Sigma_{5n} + \Sigma_{6n}\right\}.
\end{aligned}$$

Part 2: Hájek Projection. Consider the projection of $\mathcal{S}_{n,\tau}$ into an arbitrary set functions of (Z_{ij}, U_{ij}). Let

$$\zeta_{ij} = \mathbb{E}\left[\widetilde{\psi}_{\phi n\tau} \mid Z_{ij}, U_{ij}\right].$$

Hence

$$\begin{aligned}
\mathbb{E}\left[\mathcal{S}_{n,\tau} \mid Z_{ij}, U_{ij}\right] &= \frac{1}{\rho_n}\binom{n}{6}^{-1} \sum_{\phi \in \mathbf{N}_{M_n}} \mathbb{E}\left[\widetilde{\psi}_{\phi n\tau} \mid Z_{ij}, U_{ij}\right] \\
&= \frac{1}{\rho_n}\binom{n}{6}^{-1}\binom{n-2}{4} \mathbb{E}\left[\widetilde{\psi}_{\phi n\tau} \mid Z_{ij}, U_{ij}\right] \\
&= \frac{15}{\rho_n}\binom{n}{2}^{-1} \zeta_{ij}.
\end{aligned}$$

The Hájek Projection of $\mathcal{S}_{n,\tau}$ is given by

$$\mathcal{S}^*_{n,\tau} = \frac{15}{\rho_n}\binom{n}{2}^{-1} \sum_{i<j} \zeta_{ij}.$$

Part 3: Variance of $\mathcal{S}^*_{n,\tau}$*.* The variance of the Hájek Projection is

$$\begin{aligned}
\Omega_n \equiv \mathbb{V}\left(\mathcal{S}^*_{n,\tau}\right) &= \left(\frac{15}{\rho_n}\right)^2 \binom{n}{2}^{-2} \sum_{i<j} \mathbb{V}\left(\zeta_{ij}\right) \\
&= \left(\frac{15}{\rho_n}\right)^2 \binom{n}{2}^{-1} \mathbb{V}\left(\zeta_{ij}\right)
\end{aligned}$$

where $\mathbb{V}\left(\zeta_{ij}\right) = \mathbb{E}\left[\zeta_{ij}\zeta_{ij}^\top\right]$.

Part 4: Asymptotic Equivalence. The asymptotic equivalence between $\mathcal{S}_{n,\tau}$ and $\mathcal{S}^*_{n,\tau}$ follows from showing that

$$\Omega_n^{-1/2}\, \mathbb{E}\left[\left(\mathcal{S}_{n,\tau} - \mathcal{S}^*_{n,\tau}\right)^2\right] \Omega_n^{-1/2} = o_p(1).$$

Part 1 of the proof shows that the leading term of the variance of $S_{n,\tau}$ is of order $O\left(\frac{\Sigma_{2n}}{(n\rho_n)^2}\right)$. The number of terms with more than one of dyad in common is $o\left((n\rho_n)^2\right)$. That is

$$\mathbb{V}\left(\mathcal{S}_{n,\tau}\right) \simeq \left\{\left(\frac{15}{\rho_n}\right)^2 \binom{n}{2}^{-1} \Sigma_{2n}\right\} + o\left((n\rho_n)^2\right)$$

where

$$\begin{aligned}
\Sigma_{2n} &= \mathbb{C}\left(\widetilde{\psi}_{\phi_1 n\tau}, \widetilde{\psi}_{\phi_2 n\tau}\right) \\
&= \mathbb{E}\left[\mathbb{E}\left[\widetilde{\psi}_{\phi_1 n\tau}\left(\widetilde{\psi}_{\phi_2 n\tau}\right)^{\top} \mid Z_{ij}, U_{ij}\right]\right] \\
&= \mathbb{E}\left[\zeta_{ij}\zeta_{ij}^{\top}\right] \\
&= \mathbb{V}\left(\zeta_{ij}\right)
\end{aligned}$$

where the third equality follows from the fact that ϕ_1 and ϕ_2 with 2 indices in common, denoted by i, j, the arrays ϕ_1 and ϕ_2 are conditionally independent by Assumption 3.1. Next, notice that

$$\begin{aligned}
\mathbb{C}\left(\mathcal{S}_{n,\tau}, \mathcal{S}_{n,\tau}^{*}\right) &= \mathbb{E}\left[\mathcal{S}_{n,\tau}\left(\mathcal{S}_{n,\tau}^{*}\right)^{\top}\right] \\
&= \mathbb{E}\left[\left\{\mathcal{S}_{n,\tau} - \mathcal{S}_{n,\tau}^{*}\right\}\left(\mathcal{S}_{n,\tau}^{*}\right)^{\top}\right] + \mathbb{E}\left[\mathcal{S}_{n,\tau}^{*}\left(\mathcal{S}_{n,\tau}^{*}\right)^{\top}\right] \\
&= \mathbb{V}\left(\mathcal{S}_{n,\tau}^{*}\right)
\end{aligned}$$

where $\mathbb{E}\left[\left\{\mathcal{S}_{n,\tau} - \mathcal{S}_{n,\tau}^{*}\right\}\left(\mathcal{S}_{n,\tau}^{*}\right)^{\top}\right] = 0$ by definition of a projection. It follows then that

$$\begin{aligned}
\Omega_n^{-1/2}\,\mathbb{E}\left[\left(\mathcal{S}_{n,\tau} - \mathcal{S}_{n,\tau}^{*}\right)^{2}\right]\Omega_n^{-1/2} &= \Omega_n^{-1/2}\left\{\mathbb{V}\left(\mathcal{S}_{n,\tau}\right) + \mathbb{V}\left(\mathcal{S}_{n,\tau}^{*}\right) - 2\,\mathbb{C}\left(\mathcal{S}_{n,\tau}, \mathcal{S}_{n,\tau}^{*}\right)\right\}\Omega_n^{-1/2} \\
&= \Omega_n^{-1/2}\left\{\mathbb{V}\left(\mathcal{S}_{n,\tau}\right) - \mathbb{V}\left(\mathcal{S}_{n,\tau}^{*}\right)\right\}\Omega_n^{-1/2} \\
&\to 0
\end{aligned}$$

as $n \to \infty$. □

Proof of Theorem 4.2

Proof. Consider

$$\begin{aligned}
\widehat{\theta}_n - \theta_0 &= \widehat{\Gamma}_n^{-1} \times \widehat{\Psi}_{n,\tau} - \Gamma_0^{-1} \times \Psi_0 \\
&= \widehat{\Gamma}_n^{-1} \times \left(\widehat{\Psi}_{n,\tau} - \Psi_0\right) + \left(\widehat{\Gamma}_n^{-1} - \Gamma_0^{-1}\right)\Psi_0 \\
&= \Gamma_0^{-1} \times \left(\widehat{\Psi}_{n,\tau} - \Psi_0\right) + o_p(1)
\end{aligned}$$

where $\widehat{\Gamma}_n^{-1} - \Gamma_0^{-1} = o_p(1)$ was established in the proof of Theorem 4.1. The proof follows from establishing the asymptotic distribution of $\widehat{\Psi}_{n,\tau} - \Psi_0$. Consider, the decomposition

$$\begin{aligned}\widehat{\Psi}_{n,\tau} - \Psi_0 =& \left\{\frac{\mathbb{E}\left[m_n^*\right]}{m_n^*} - 1\right\} \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} W_\sigma \widehat{G}_{\sigma,\tau}^* + \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{W_\sigma \widehat{G}_{\sigma,\tau}^* - \mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\} \\ & - \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{\mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right] - \mathbb{E}\left[W_\sigma G_\sigma^*\right]\right\}\end{aligned}$$

It follows from Theorem 4.1 that first term in the right-hand side of the equality convergences in probability to zero. Moreover, it follows from Lemma 6 that the trimming effect is asymptotically negligible, and thus the third term in the right-hand side of the equality also convergences in probability to zero. That is

$$\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{\mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right] - \mathbb{E}\left[W_\sigma G_\sigma^*\right]\right\} = o_p(1).$$

Notice that the second term can be decomposed as

$$\begin{aligned}&\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{W_\sigma \widehat{G}_{\sigma,\tau}^* - \mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\} \\ =& \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{W_\sigma \widehat{G}_{\sigma,\tau}^* - \mathbb{E}\left[W_\sigma \widehat{G}_{\sigma,\tau}^* \mid \mathbf{Z}_n\right]\right\} + \frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{\mathbb{E}\left[W_\sigma \widehat{G}_{\sigma,\tau}^* \mid \mathbf{Z}_n\right] - \mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\}.\end{aligned}$$

Lemma 10 ensures that

$$\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{\mathbb{E}\left[W_\sigma \widehat{G}_{\sigma,\tau}^* \mid \mathbf{Z}_n\right] - \mathbb{E}\left[W_\sigma G_{\sigma,\tau}^*\right]\right\} = o_p(1),$$

while Lemma 11 shows that Hájek projection of the first term on the right-hand side is

$$\frac{1}{m_n \rho_n} \sum_{\sigma \in \mathbf{N}_{m_n}} \left\{W_\sigma \widehat{G}_{\sigma,\tau}^* - \mathbb{E}\left[W_\sigma \widehat{G}_{\sigma,\tau}^* \mid \mathbf{Z}_n\right]\right\} = \frac{15}{\rho_n} \binom{n}{2}^{-1} \sum_{i<j} \zeta_{ij,\tau} + o_p\left(1/\sqrt{n(n-1)\rho_n}\right)$$

with the score given by

$$\zeta_{ij} \equiv \mathbb{E}\left[\widetilde{\psi}_{\phi n \tau} \mid Z_{ij}, U_{ij}\right].$$

Notice that conditional on $\boldsymbol{Z}_n$, the components of the Hájek Projection are conditionally independent of each other with mean zero. Moreover, it follows from the proof of Theorem that

$$\mathbb{E}\left[\zeta_{ij} \zeta_{ij}^\top\right] = O\left(\rho_n\right)$$

and thus

$$\mathbb{V}\left(\mathcal{S}_{n,\tau}^{*}\right)=\left(\frac{15}{\rho_n}\right)^2\binom{n}{2}^{-1}\mathbb{V}\left(\zeta_{ij}\right)=O\left(\frac{1}{n(n-1)\rho_n}\right).$$

which ensures that $\tilde{\Omega}=n(n-1)\rho_n\,\mathbb{V}\left(\mathcal{S}_{n,\tau}^{*}\right)=O\left(1\right)$. Consequently,

$$\widehat{\Psi}_{n,\tau}-\Psi_0=\frac{15}{\rho_n}\binom{n}{2}^{-1}\sum_{i<j}\zeta_{ij,\tau}+o_p\left(1/\sqrt{n(n-1)\rho_n}\right)$$

As established in Graham (2017), the conditional independence structure of the terms in the Hájek Projection ensures that the following convergence in distribution holds

$$\widehat{\Omega}_n^{-1/2}\left\{\binom{n}{2}^{-1}\sum_{i<j}\zeta_{ij,\tau}\right\}\rightsquigarrow\mathcal{N}\left(0,I\right)$$

with

$$\widehat{\Omega}_n=\binom{n}{2}^{-1}\sum_{i<j}\zeta_{ij,\tau}\zeta_{ij,\tau}^{\top}.$$

Let $\mathcal{V}_n=\widehat{\Gamma}_n^{-1}\widehat{\Omega}_n\widehat{\Gamma}_n^{-1}$, it follows from Slutsky's Theorem that

$$\mathcal{V}_n^{-1/2}\left(\widehat{\theta}_n-\theta_0\right)\rightsquigarrow\mathcal{N}\left(0,15^2I\right).$$

□

B Additional Monte Carlo Simulations

DGP 1: $v_{ij} \sim N(0, 1.5)$ **and** $U_{ij} \sim Beta(2,2) - \frac{1}{2}$

Table B1: Simulation results for the semiparametric estimator $\widehat{\theta}_n$ and the Tetrad Logit

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
				$n = 100$					
$\log(\log(n))$	1.5182	1.5159	0.4277	0.1832	1.6244	1.6377	0.2933	0.1015	0.3982
$\log(n)^{1/2}$	1.5379	1.5102	0.4653	0.2179	1.6388	1.6459	0.3017	0.1103	0.3511
$\log(n)$	1.5440	1.5481	0.6179	0.3837	1.6580	1.6595	0.3507	0.1479	0.2391
$n^{1/3}$	1.5135	1.5256	0.6459	0.4174	1.6341	1.6383	0.3516	0.1416	0.2377
				$n = 200$					
$\log(\log(n))$	1.5307	1.5233	0.2133	0.0465	1.6413	1.6429	0.1402	0.0396	0.3916
$\log(n)^{1/2}$	1.5280	1.5174	0.2318	0.0545	1.6379	1.6427	0.1458	0.0403	0.3348
$\log(n)$	1.5276	1.5334	0.3263	0.1072	1.6373	1.6347	0.1839	0.0527	0.2135
$n^{1/3}$	1.5150	1.5165	0.3517	0.1239	1.6433	1.6445	0.1770	0.0519	0.1953

[1] Total number of Monte Carlo simulations = 1000.
[2] Bandwidth parameter $h = 0.05$.

Table B2: Simulation results for the semiparametric estimator $\widehat{\theta}_n$ and the Tetrad Logit

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
				$n = 100$					
$\log(\log(n))$	1.5191	1.5185	0.4183	0.1754	1.6334	1.6453	0.2890	0.1013	0.3979
$\log(n)^{1/2}$	1.5342	1.5322	0.4484	0.2022	1.6231	1.6276	0.2950	0.1022	0.3525
$\log(n)$	1.5015	1.5099	0.6342	0.4022	1.6241	1.6350	0.3517	0.1391	0.2395
$n^{1/3}$	1.5226	1.4816	0.6347	0.4034	1.6100	1.5966	0.3513	0.1355	0.2386
				$n = 200$					
$\log(\log(n))$	1.5307	1.5233	0.2133	0.0465	1.6413	1.6429	0.1402	0.0396	0.3916
$\log(n)^{1/2}$	1.5280	1.5174	0.2318	0.0545	1.6379	1.6427	0.1458	0.0403	0.3348
$\log(n)$	1.5276	1.5334	0.3263	0.1072	1.6373	1.6347	0.1839	0.0527	0.2135
$n^{1/3}$	1.5150	1.5165	0.3517	0.1239	1.6433	1.6445	0.1770	0.0519	0.1953

[1] Total number of Monte Carlo simulations = 1000.
[2] Bandwidth parameter $h = 0.1$.

Table B3: Simulation results for the semiparametric estimator $\widehat{\theta}_n$ and the Tetrad Logit

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
				$n = 100$					
$\log(\log(n))$	1.5564	1.5642	0.4289	0.1871	1.6288	1.6232	0.2954	0.1039	0.3986
$\log(n)^{1/2}$	1.5544	1.5502	0.4592	0.2139	1.6272	1.6308	0.3181	0.1174	0.3524
$\log(n)$	1.5468	1.5495	0.6127	0.3777	1.6407	1.6357	0.3410	0.1361	0.2392
$n^{1/3}$	1.5532	1.5695	0.6255	0.3940	1.6371	1.6416	0.3586	0.1474	0.2367
				$n = 200$					
$\log(\log(n))$	1.5324	1.5347	0.2152	0.0474	1.6337	1.6336	0.1384	0.0370	0.3923
$\log(n)^{1/2}$	1.5371	1.5462	0.2437	0.0608	1.6346	1.6348	0.1479	0.0400	0.3354
$\log(n)$	1.5412	1.5490	0.3325	0.1122	1.6398	1.6362	0.1753	0.0503	0.2141
$n^{1/3}$	1.4892	1.5025	0.3580	0.1283	1.6392	1.6321	0.1848	0.0535	0.1952

[1] Total number of Monte Carlo simulations = 1000.
[2] Bandwidth parameter $h = 0.2$.

DGP 2: $v_{ij} \sim Logistic(0, 1.5)$ **and** $U_{ij} \sim Logistic(0, 1)$

Table B4: Simulation results for the semiparametric estimator $\widehat{\theta}_n$ and the Tetrad Logit

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
				$n = 100$					
$\log(\log(n))$	1.4742	1.4654	1.1009	1.2126	1.3753	1.3579	0.8840	0.7970	0.4457
$\log(n)^{1/2}$	1.4500	1.4454	1.1031	1.2194	1.3410	1.3640	0.9142	0.8611	0.4218
$\log(n)$	1.4624	1.4202	1.2157	1.4793	1.3664	1.3542	0.9406	0.9026	0.3523
$n^{1/3}$	1.4649	1.4396	1.1548	1.3348	1.3311	1.2840	0.8938	0.8274	0.3512
				$n = 200$					
$\log(\log(n))$	1.5018	1.4898	0.5265	0.2772	1.3916	1.3840	0.4226	0.1904	0.4435
$\log(n)^{1/2}$	1.4961	1.5007	0.5208	0.2712	1.3849	1.3756	0.4175	0.1875	0.4125
$\log(n)$	1.4827	1.4826	0.5685	0.3235	1.3956	1.4010	0.4393	0.2039	0.3352
$n^{1/3}$	1.4559	1.4441	0.5951	0.3561	1.3996	1.4107	0.4657	0.2270	0.3196

[1] Total number of Monte Carlo simulations = 1000.
[2] Bandwidth parameter $h = 0.05$.

Table B5: Simulation results for the semiparametric estimator $\widehat{\theta}_n$ and the Tetrad Logit

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
$n = 100$									
$\log(\log(n))$	1.4860	1.4265	1.0722	1.1498	1.4018	1.4011	0.8405	0.7160	0.4457
$\log(n)^{1/2}$	1.4752	1.5015	1.1167	1.2476	1.3495	1.3304	0.9070	0.8452	0.4211
$\log(n)$	1.4880	1.4876	1.1675	1.3631	1.3718	1.4099	0.9220	0.8665	0.3521
$n^{1/3}$	1.5122	1.5241	1.1744	1.3794	1.3996	1.4208	0.9477	0.9083	0.3514
$n = 200$									
$\log(\log(n))$	1.4739	1.4934	0.5116	0.2624	1.3705	1.3710	0.4132	0.1875	0.4435
$\log(n)^{1/2}$	1.4976	1.4802	0.5365	0.2878	1.4006	1.4015	0.4349	0.1991	0.4126
$\log(n)$	1.4764	1.4516	0.5635	0.3181	1.3895	1.3936	0.4495	0.2143	0.3348
$n^{1/3}$	1.4775	1.4853	0.6088	0.3711	1.4052	1.4109	0.4795	0.2389	0.3195

[1] Total number of Monte Carlo simulations = 1000.

[2] Bandwidth parameter $h = 0.1$.

Table B6: Simulation results for the semiparametric estimator $\widehat{\theta}_n$ and the Tetrad Logit

	$\widehat{\theta}_n$				Tetrad Logit				
	mean	median	std	MSE	mean	median	std	MSE	Degree
$n = 100$									
$\log(\log(n))$	1.4585	1.4583	1.0656	1.1372	1.3725	1.3539	0.8818	0.7939	0.4457
$\log(n)^{1/2}$	1.5320	1.5240	1.0826	1.1731	1.4074	1.3894	0.8846	0.7911	0.4217
$\log(n)$	1.5087	1.5273	1.1567	1.3381	1.3775	1.3477	0.9560	0.9290	0.3519
$n^{1/3}$	1.4617	1.4946	1.1640	1.3563	1.3747	1.3705	0.9125	0.8483	0.3511
$n = 200$									
$\log(\log(n))$	1.4987	1.4924	0.5117	0.2618	1.3860	1.3914	0.4129	0.1835	0.4436
$\log(n)^{1/2}$	1.5044	1.5113	0.5216	0.2720	1.4089	1.4358	0.4167	0.1820	0.4123
$\log(n)$	1.4709	1.4455	0.5713	0.3272	1.3861	1.3803	0.4582	0.2229	0.3350
$n^{1/3}$	1.4530	1.4549	0.5906	0.3510	1.3999	1.3890	0.4539	0.2160	0.3197

[1] Total number of Monte Carlo simulations = 1000.

[2] Bandwidth parameter $h = 0.2$.

C Random Utility Model with Transferable Utilities

The network formation model described in Equation 1 can be obtained as a stable outcome of a random utility model with transferable utilities. For instance, let $\bar{u}_{ij}(Z_{ij}, A_j, U_{ij})$ denote individual i's latent valuation of establishing a link with j given their shared observed attributes Z_{ij}, agent $j's$ unobserved type A_j, and their common unobserved factor U_{ij}. It follows that the joint net benefit of adding the link $\{i, j\}$ to the network $\mathbf{D}_n$ is

$$\bar{u}_{ij}(Z_{ij}, A_j, U_{ij}) + \bar{u}_{ji}(Z_{ij}, A_i, U_{ij}) = Z_{ij}'\beta_0 + A_i + A_j - U_{ij}. \tag{C.1}$$

Notice that the joint net benefit accounts for the preferences based on the observed attributes Z_{ij}, as well as on the agent-specific characteristics $A_i + A_j$, and the exogenous factors affecting the decision to establish a link U_{ij}. Moreover, Equation C.1 implies that two distinct individuals i and j in $\mathbf{N}_n$ only have utility valuations for their own observed and unobserved attributes. In other words, the (i, j)th linking decision does not depend on the attributes of other individuals $k \in \mathbf{N}_n$ with $k \neq i, j$, as well as on the structure of the network $\mathbf{D}_n$, which would give rise to network externalities.

Next, I introduce the definition of stability.

Definition 1 (Stability). *A network $\mathbf{D}_n$ is stable with transfers if for any distinct $i, j \in \mathbf{N}_n$:*

1. *$D_{ij} = 1$ only if $\bar{u}_{ij}(Z_{ij}, A_j, U_{ij}) + \bar{u}_{ji}(Z_{ij}, A_i, U_{ij}) \geq 0$; and*
2. *$D_{ij} = 0$ only if $\bar{u}_{ij}(Z_{ij}, A_j, U_{ij}) + \bar{u}_{ji}(Z_{ij}, A_i, U_{ij}) < 0$.*

Notice that this definition adapts the pairwise stability in Jackson and Wolinsky (1996) to allow for transferable utilities. Intuitively, this condition states that a link within dyad $\{i, j\}$ is established if the net benefit of that connection is nonnegative.

D Data Description

D.1 Data References

This research uses data from Add Health, a program project directed by Kathleen Mullan Harris and designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD31921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Special acknowledgment is due Ronald R. Rindfuss and Barbara Entwisle for assistance in the original design. Information on how to obtain the Add Health data files is available on the Add Health website (www.cpc.unc.edu/projects/addhealth). No direct support was received from grant P01-HD31921 for this analysis.

D.2 Add Health sample

The Add Health dataset is a nationally representative survey of adolescents in grades 7-12 in the United States during the 1994-1995 school year. It has been designed to study the impact of the social environment, such as participants' schools, neighborhoods, friends and families, on adolescents' behavior. It is is a longitudinal survey collected in five waves of in-school and in-home interviews, and linked to school administrative data.[D.1]

I use data from the Wave 1 in-home survey, which contains information on a total of 20,745 students, including their self-reported friendship connections. The in-home survey is more suitable for this analysis than the in-school survey as it also includes information on students' health in the early years of life, their parents and households, such as parents' health-related behavior, marriage relationships, and household income. Below, I discuss in detail the attributes used for this analysis.[D.2]

In 16 out of the total 145 high schools in the sample, known as saturated schools, all the students were selected for in-home interviews regardless of whether they had completed an in-school questionnaire.[D.3] In these interviews, the students were asked to name up to five male and five female students.[D.4] The saturated high schools include two large schools and 14 smaller schools with a total of 3,702 enrolled students. The two larger schools were selected deliberately to represent student populations with opposite ethnic compositions. One is predominantly white

[D.1] The Add Health website describes the data in detail, www.cpc.unc.edu/projects/addhealth.

[D.2] I matched the in-home survey data with the school administrative data to input for missing observations, which yields a sample of 20,369 matched participants.

[D.3] The code books *Adolescent In-Home Questionnaire* and *School Information* provide more information on the definition of saturated schools.

[D.4] In all the remaining schools, most of the students were allowed to nominate only one male and one female friend. In fact, the average and median share of students across high schools that were allowed to list up to five male and five female friends in non-saturated schools are 22% and 13%, respectively. Meanwhile, in saturated schools, these figures reached 84% and 94%, respectively.

and is located in a small town; the other is characterized by a large ethnic heterogeneity and is located in a major metropolitan area. The remaining 14 schools are drawn to represent both rural and urban areas and are similar in terms of the number of enrolled students.

In my analysis, I focus on saturated schools since their design allows me to recover the complete friendship network for all the students enrolled in these schools. Moreover, to avoid inducing any bias due to oversampled ethnic groups, I select a representative sample of four high schools among the 14 smaller saturated schools.[D.5] These four high schools are the largest in each stratification region (West, Midwest, South, and Northeast). Table D3 provides evidence suggesting that, for the majority of the covariates used in the analysis, the selected sample is representative of the population of students enrolled in the remaining saturated high schools.[D.6] The final sample includes 273 students, after dropping missing observations.

D.3 Observed Attributes

In this section, I describe the covariates included in the estimation of the network formation model given by Equation 1. The covariates are classified in the following five categories:

1. Socio-demographic attributes:

- `Age` is a discrete variable that indicates the student's age.
- `Gender` is a binary variable that takes the value of 1 if the student is a female and 0 otherwise.
- `White` ethnicity is a binary variable takes the value of 1 if the student is of white ethnicity and 0 otherwise.
- `Religion` is a variable that describes the student's religion.

In this sample, the average age of the students is 15 years old and the shares of female and white students are 57% and 85%, respectively. From the 28 different religious beliefs recorded in the Add Health dataset, 21% of students reported to be Baptist, 15% Methodist, 14% Catholic, 14% Disciples of Christ, and 14% Atheist. The remaining students follow other protestant denominations.

2. Educational factors:

- `Grade` is a variable that indicates the current academic grade of the student and ranges from 7th to 12th grade.
- `Overall GPA` represents the student's average GPA across English, History, Mathematics, and Science courses.

[D.5] Notice that this setting is consistent with the sampling framework described in Section 4.

[D.6] Naturally, the empirical analysis can be conducted by aggregating the information across the 14 saturated high schools without modifying the theoretical framework, but at a higher computational cost.

- `Clubs` is an index that records the number of artistic, educational, language or sports clubs that the student attends. This index takes a value from 0 to 4, where 0 to 3 correspond to the actual number of clubs that the student attends, and 4 indicates that this number is 4 or more.

- `Repeated Grade` is a dummy variable that takes the value of 1 if the student has ever repeated an academic grade and 0 otherwise.

- `College Expectations` is a categorical variable indicating on a scale of 1 to 5, where 1 is low and 5 is high, to what extent students want to go to college.

On average, students in this network have a 2.9 GPA, are members of approximately 2 clubs, and have strong college aspirations. The share of students that has repeated an academic grade is 18%.

3. Economic factors:

- `Mother Highly Educated` is a dummy variable that indicates whether or not the student's mother has attended at least some years of college.

- `Mother Works` is a binary variable that takes the value of 1 if the student's mother participates in the labour market.

- `Log Household income` is the logarithm value of the total household income before taxes in 1994, expressed as deviations from the median value.

- `Divorced Parents` is a dummy variable that takes the value of 1 if the students' parents are divorced or separated.

- `Number of Siblings` is a categorical variable that indicates the student's number of siblings. This variable takes a value from 0 to 5, where 5 corresponds to five siblings or more.

- `Birth Order` is a variable that records the order in which the student was born in her family. This covariate takes values from 1 to 5, where 5 denotes the fifth children or lower in the ranking.

- `Good neighborhood` is a binary variable that indicates whether or not the student's family signalled their current neighborhood as having lower crime and drug use relative to other neighborhoods, as well as better schools.

On average, 41% of the students' mothers in this sample have attended college and 78% do paid work. Moreover, the students have on average two siblings and they are the second child.

4. Physical and health-related factors:

- `Depressed` is a categorical variable that indicates how often a student felt depressed during the previous week. It takes a value from 0 to 3, where 0 represents never or rarely, and 3 represents all the time.

- `Attractiveness` is variable that scores the student's level of attractiveness according to the interviewer. It takes values on a scale from -2 to 2, where -2 is very unattractive and 2 is very attractive.

- `S&D habits` is a dummy variable that takes the value of 1 if the student either smokes or drinks regularly.[D.7]

- `Friends' S&D habits` records the number of friends out of the student's three best friends that drink or smoke regularly.

- `Mother's Health` is a categorical variable that indicates the mother's general health on a scale from 0 to 5, where 0 is poor, and 5 is excellent.

5. Early childhood conditions

- `Breastfed` is a dummy variable that takes the value of 1 if the student was breastfed at birth for a period of at least three months, and 0 otherwise.

- `Intellectual or Physical Disability` is a binary variable that captures whether or not the student is intellectually or physically disabled since birth, for example as having a learning disability.

- `Mother's Age at Birth` is variable that records the age at which the mother gave birth to the student.

- `Mother Health Risk Factors` is a binary variable that takes the value of 1 if the mother suffers from alcoholism, diabetes, or overweight, and 0 otherwise.

The descriptive statistics of the observed covariates are summarized in Table D1 at individual level and in Table D2 at a dyad level.

Now, I explain the motivation for using this set of attributes. The rationale for including socio-demographic attributes is to account for homophily on age, gender, race, and religion. Meanwhile, the educational factors considered are likely to be affected by students' birth weight and capture homophily patterns in friendship connections across extracurricular activities, academic performance, and aspirations (see e.g., Almond and Currie 2011; Figlio et al. 2014). Similarly, the economic factors are likely to affect students' birth weight and individual personality traits (see e.g., Aizer and Currie 2014; Almond et al. 2018). Also, they capture assortative matching based on economic factors. For example, high-skilled mothers might help to raise more gregarious students, while disruptions in the household might damper the student's social skills.

The physical and health-related factors are likely to be related to the students' birth weight but also contribute to shaping their individual characteristics and affect the probability of establishing

[D.7] Smoking and drinking regularly is defined as smoking at least 1 cigarette ever day for the past 30 days and drinking alcohol at least 3 days a week during the last 12 months.

friendship connections (see e.g., Del Bono et al. 2012; Brunello and Schlotter 2011). For example, the tendency to be depressed may hamper students' ability to engage in social interactions. In contrast, common social interests, such as individual and friends' smoking and drinking habits, may increase it. Finally, the attributes accounting for early childhood conditions may contribute to building up a stronger immune system, as well as influence the actual realization of students' birth weight and affect the development of students' individual traits (see e.g., Aizer and Currie 2014; Fitzsimons and Vera-Hernández 2015; Maruyama and Heinesen 2020).

Table D1: Descriptive statistics

	mean	median	std	min	max
network degree	2.894	3.000	2.188	0.000	10.000
age	15.582	16.000	1.708	13.000	20.000
gender	0.531	1.000	0.499	0.000	1.000
white	0.853	1.000	0.354	0.000	1.000
grade	9.206	9.000	1.658	7.000	12.000
overall GPA	2.922	3.000	0.811	0.000	4.000
clubs	1.982	2.000	1.609	0.000	4.000
repeated grade	0.187	0.000	0.390	0.000	1.000
depressed	0.421	0.000	0.686	0.000	3.000
number of siblings	2.359	2.000	1.053	0.000	5.000
order of siblings	1.593	1.000	0.833	1.000	4.000
mother highly educated	0.414	0.000	0.493	0.000	1.000
mother works	0.788	1.000	0.409	0.000	1.000
S&D habits	0.209	0.000	0.406	0.000	1.000
friends' S&D habits	0.978	1.000	1.135	0.000	3.000
good neighborhood	0.648	1.000	0.477	0.000	1.000
religion	10.205	5.000	8.118	0.000	28.000
college expectations	4.491	5.000	0.926	1.000	5.000
attractiveness	0.531	0.000	0.770	-2.000	2.000
mother's health	3.681	4.000	0.993	0.000	5.000
divorced parents	0.289	0.000	0.453	0.000	1.000
breastfed	0.352	0.000	0.477	0.000	1.000
disability	0.103	0.000	0.303	0.000	1.000
mother's age at birth	25.029	24.000	5.166	14.000	46.000
mother's health risks factors	0.308	0.000	0.462	0.000	1.000
log household income	3.597	3.638	0.649	1.099	5.451
birth weight	0.000	-64.881	530.349	-1566.901	1975.599
Sample size $n = 273$					

Table D2: Descriptive statistics at a dyad level

	mean	median	std	min	max
network degree	2.894	3.000	2.188	0.000	10.000
age	0.161	0.000	6.076	0.000	1.000
gender	0.502	1.000	8.261	0.000	1.000
white	0.728	1.000	7.349	0.000	1.000
grade	0.171	0.000	6.220	0.000	1.000
overall GPA	8.539	8.250	56.449	0.000	16.000
clubs	3.927	0.000	85.922	0.000	16.000
repeated grade	0.035	0.000	3.032	0.000	1.000
depressed	0.177	0.000	10.303	0.000	9.000
number of siblings	5.565	4.000	60.878	0.000	25.000
order of siblings	2.539	2.000	33.109	1.000	16.000
mother highly educated	0.171	0.000	6.226	0.000	1.000
mother works	0.620	1.000	8.019	0.000	1.000
S&D habits	0.670	1.000	7.772	0.000	1.000
friends' S&D habits	0.957	0.000	33.566	0.000	9.000
good neighborhood	0.420	0.000	8.156	0.000	1.000
religion	0.140	0.000	5.728	0.000	1.000
college expectations	20.168	20.000	98.227	1.000	25.000
attractiveness	0.282	0.000	13.693	-4.000	4.000
mother's health	13.552	12.000	86.972	0.000	25.000
divorced parents	0.084	0.000	4.577	0.000	1.000
breastfed	0.124	0.000	5.439	0.000	1.000
disability	0.011	0.000	1.686	0.000	1.000
mother's age at birth	0.059	0.000	3.896	0.000	1.000
mother's health risks factors	0.095	0.000	4.837	0.000	1.000
log household income	0.002	0.000	6.996	-4.604	6.446
birth weight	826282.7	664176.9	12185755.3	-1897778.3	8320917.1
Sample size $n = 273$					

Table D3: Two-sample t-test for equal means

	Mean selected schools	Mean other saturated schools	Differences	P-values
age	15.58	15.09	-3.81	0.00
female	0.53	0.51	-0.46	0.64
white	0.85	0.84	-0.49	0.62
grade	9.21	8.69	-4.20	0.00
overall GPA	2.92	2.85	-1.07	0.28
clubs	1.98	1.74	-2.05	0.04
repeated grade	0.19	0.15	-1.40	0.16
depressed	0.42	0.35	-1.40	0.16
number of siblings	2.36	2.42	0.77	0.44
order of siblings	1.37	1.63	2.96	0.00
mother highly educated	0.41	0.40	-0.51	0.61
mother works	0.79	0.79	0.01	0.99
S&D habits	0.21	0.18	-1.01	0.31
friends' S&D habits	0.98	0.85	-1.52	0.13
good neighborhood	0.65	0.77	3.67	0.00
religion	10.21	11.08	1.37	0.17
college expectations	4.49	4.38	-1.27	0.20
attractiveness	0.53	0.69	2.68	0.01
mother's health	3.68	3.76	1.02	0.31
divorced parents	0.29	0.29	-0.07	0.94
breastfed	0.35	0.27	-2.35	0.02
disability	0.10	0.16	2.27	0.02
mother's age at birth	25.03	27.57	0.95	0.34
mother's health risk factors	0.31	0.19	-3.88	0.00
log household income	3.60	3.57	-0.47	0.64
birth weight	3,409.00	3,386.25	-0.52	0.60