

# Identification and Estimation of a Partially Linear Regression Model using Network Data

Eric Auerbach\*

January 28, 2018

## Abstract

I study a regression model in which one covariate is an unknown function of a latent driver of link formation in a network. Rather than specify and fit a parametric network formation model, I introduce a new method based on matching pairs of agents with similar columns of the squared adjacency matrix, the  $ij$ th entry of which contains the number of other agents linked to both agents  $i$  and  $j$ . The intuition behind this approach is that for a large class of network formation models the columns of this matrix characterize all of the identifiable information about individual linking behavior. In the paper, I first describe the model and formalize this intuition. I then introduce estimators for the parameters of the regression model and characterize their large sample properties.

## 1 Introduction

Most economic outcomes are not determined in isolation. Rather agents are influenced by the behaviors and characteristics of other agents. For example, a high school student's academic performance might depend on the attitudes and expectations of that student's friends and family (see generally [Akerlof and Kranton 2002](#), [Austen-Smith and Fryer Jr 2005](#), [Marianne 2011](#), [Sacerdote 2011](#)).

Incorporating this social influence into the right-hand side of an economic model may be desirable when the researcher wants to understand its impact on the agents' outcomes

---

\*Department of Economics, Northwestern University. E-mail: eric.auerbach@northwestern.edu. I thank my advisors, James Powell and Bryan Graham for their advice and support. I also thank Ivan Canay, David Card, Christina Chung, Aluma Dembo, Joel Horowitz, Michael Jansson, Patrick Kline, Sheisha Kulkarni, Chuck Manski, Konrad Menzel, Carl Nadler, Stephen Nei, Aureo de Paula, Demian Pouzo, Mikkel Solvsten, Katalin Springel and participants at the UC Berkeley Econometrics Seminar for helpful feedback.

or when it confounds the impact of another explanatory variable such as the causal effect of some nonrandomized treatment. For instance in the above example, the researcher may want to learn the causal effect of a tutoring program on academic performance in which program enrollment and counterfactual academic performance are both partially determined by family expectations. However, in many cases the relevant social influence is not observed by the researcher and so it cannot simply be included as a covariate in the model. That is in the above example, the researcher does not have access to data on the family expectations that confound the causal effect of the tutoring program and thus cannot control for this variable using conventional methods.

One solution to this problem is to collect social network data and presume that the unknown social influence is revealed by agent linking behavior in the network (see generally [Jackson 2008; 2014](#), [Blume, Brock, Durlauf, and Ioannides 2010](#), [Boucher and Fortin 2015](#), [Chandrasekhar 2015](#), [Graham 2015](#), [de Paula 2016](#), [Kranton 2017](#)). For instance in the above example, the researcher might observe pairs of students who identify as friends and believe that students with similar reported friendships have similar family expectations. It is not immediately clear, however, how one might actually use network data to account for unobserved social influence in practice, since the total number of ways in which agents can be linked in a network is typically large relative to the sample size.

The main contribution of this paper is to demonstrate one way in which network data can be used as a substitute for this sort of unobserved heterogeneity in the context of a partially linear regression model. The paper consists of three steps: first, I specify a joint regression and network formation model; second, I establish sufficient conditions for the parameters of the regression model to be identified; third, I provide estimators for these parameters and characterize their large sample properties.

In the first step, described further in Section 2.1, I specify a model in which latent social characteristics determine both the social influence in the regression and links in the network. The model draws upon previous work by [Goldsmith-Pinkham and Imbens \(2013\)](#), [Chan \(2014\)](#), [Hsieh and Lee \(2014\)](#), [Johnsson and Moon \(2015\)](#), and [Arduini, Patacchini, and Rainone \(2015\)](#). However, these authors rely on relatively strong functional form assumptions on the network formation model that when wrong may lead to invalid inferences about the

parameters of the regression model. My method does not require such assumptions.

An assumption I do require is that the network links are conditionally independent given the agents' social characteristics. This assumption is not uncommon in the network formation literature (see [Bickel and Chen 2009](#), [Graham 2017](#), as well as the above literature). But when taken as a literal description of the agents' incentives to form links, the model does preclude behavior thought to characterize many economic networks (c.f. [Sheng 2012](#), [Chandrasekhar and Jackson 2014](#), [Leung 2015](#), [Menzel 2015](#), [Ridder and Sheng 2015](#), [Badev 2017](#), [Mele 2017](#)). In particular, under the random utility interpretation for this model outlined by [Candelaria \(2016\)](#), the utility two agents receive from forming a link cannot explicitly depend on the existence of links between other agents in the sample.

In Section 2.3, I propose an alternative interpretation of the model that does not imply such strong behavioral assumptions about the network formation process. This interpretation views the model not as a literal description of agent behavior, but as a reduced form characterization of the within-equilibrium distribution of network links generated by some economic game, not specified by the researcher. A similar representation argument underlies the estimation strategy of [Menzel \(2015\)](#). However, my interpretation also relies on the additional argument that, for the purposes of identifying and estimating the parameters of a regression model, conflating the true network formation process with its reduced form approximation is, in many cases, without loss of generality. A formal presentation of this argument, its limitations, and an example can be found in Section 2.3.

In the second step, described further in Sections 2.2 and 3.2, I provide sufficient conditions for the parameters of the regression model to be identified without strong functional form restrictions on the network formation model. The idea behind these conditions is that in a regression model in which the outcome depends on observed explanatory variables and an unobserved social influence term, the model is identified if agents with similar social characteristics have similar social influences but different explanatory variables. An innovation of this paper is the use of network distance, a novel measure of similarity between agents' social characteristics, to formalize these conditions make them straightforward to apply in practice.

To illustrate the use of these conditions, I study the identification of network peer effects

in a variation on the linear-in-menas model of [Bramoullé, Djebbari, and Fortin \(2009\)](#) and demonstrate that, in the setting of this paper, the network peer effects are not generally identified in the presence of unknown social influence. Similar results have been found in the related group peer effects literature (for instance, [Manski 1993](#), [Graham and Hahn 2005](#), [Graham 2008](#)) in which a group peer effect is not generally distinguishable from unobserved heterogeneity at the group level. More details about this example can be found at the end of Section 2.2.

In the third step, also described further in Section 2.2, I propose estimators for the parameters of the regression model based on matching pairs of agents with similar columns of the squared adjacency matrix. The adjacency matrix of a network is a matrix with the number of rows and columns equal to the number of agents. It contains a 1 in the  $ij$ th entry if agents  $i$  and  $j$  are linked and a 0 otherwise. The squared adjacency matrix refers to the matrix square of the adjacency matrix and agent  $i$ 's column of the squared adjacency matrix is the  $i$ th column of this matrix.

The rationale for this procedure follows from a new result in this paper that, under mild regularity conditions on the network formation model, agents with similar columns of the squared adjacency matrix necessarily have similar social characteristics, as measured by network distance. The logic is related to recent arguments from the link prediction literature (in particular [Bickel, Chen, and Levina 2011](#), [Rohe, Chatterjee, and Yu 2011](#), [Zhang, Levina, and Zhu 2015](#)), though to my knowledge the main result and its application to the identification and estimation of the parameters of a regression model are original. A formal statement of this result can be found in Section 3.3.1.

The estimators are simple to compute and, under certain regularity conditions, are consistent and asymptotically normal. In particular, the estimators can be approximated by ratios of nondegenerate U-statistics, so that their large sample distributions can be derived analytically using arguments from [Serfling \(2009\)](#) (see also [Powell, Stock, and Stoker 1989](#), [Ahn and Powell 1993](#)) and approximated using random sampling methods such as the bootstrap of [Efron \(1979\)](#) (see also [Bickel and Freedman 1981](#), [Bhattacharyya and Bickel 2015](#), [Menzel 2017](#)). Details about the large sample properties of these estimators can be found in Sections 3.3 and 3.4.

This analysis, however, is complicated by the fact that the matching variable is a vector of dimension equal to the sample size. As a result, constructions commonly used in the literature to characterize the rate of convergence and limiting distribution of the estimators, such as the density function of the matching variable, are not well defined in my setting. To resolve this problem, I appeal to arguments from the literature on functional nonparametrics (for instance, [Ferraty, Mas, and Vieu 2007](#), [Hong and Linton 2016](#)), in which this density function is replaced with a more general notion of a small ball probability. This construction can then be characterized using tools from the literature on dense graph limits (see generally [Lovász 2012](#)). The dimension of the matching variable also complicates correcting the bias of the slope parameter of the regression model, for which I propose a slight variation on the jackknife method of [Honoré and Powell \(1997\)](#). Details can be found in Section 3.3.3.

Section 4 contains simulation evidence from three Monte Carlo experiments and Section 5 concludes by discussing how the method of this paper might be extended to various nonlinear and nonparametric regression models, or to allow for weighted networks, directed networks, or networks with exogenous link covariates. I leave the formal study of these extensions to future work.

## 2 Model and Estimators

Section 2.1 provides an overview of the model. Section 2.2 provides an overview of the main identification conditions and proposed estimators. Section 2.3 contains a discussion about two behavioral interpretations of the model.

### 2.1 Model

Let  $i$  be an arbitrary agent from a large population. Associated with agent  $i$  is an outcome  $y_i \in \mathbb{R}$ , a vector of observed explanatory variables  $x_i \in \mathbb{R}^k$  for some positive integer  $k$ , and an unobserved index of social characteristics  $w_i \in [0, 1]$ . The three are related by the following regression model

$$y_i = x_i\beta + \lambda(w_i) + \varepsilon_i \tag{1}$$

in which  $\beta \in \mathbb{R}^k$  is an unknown slope parameter,  $\lambda$  is an unknown Lebesgue measurable function, and  $\varepsilon_i$  is an idiosyncratic error with  $E[\varepsilon_i|x_i, w_i] = 0$ . I emphasize that the semilinear structure of (1) is used to simplify the exposition of the paper; it is possible to extend the logic of this paper to various nonlinear and nonparametric regression models (see Section 5).

The parameters of interest are  $\beta$  and agent  $i$ 's social influence term  $\lambda(w_i)$ . In this paper, the social influence function  $\lambda : [0, 1] \rightarrow \mathbb{R}$  is not a parameter of interest because it is not separately identified from  $w_i$ . It is thus without loss to normalize the marginal distribution of  $w_i$  to be standard uniform. Apart from this normalization, the two main identification conditions given in Section 2.2, and various regularity conditions given in Section 3, the joint distribution of  $x_i$  and  $w_i$  is left unrestricted. In fact, under these conditions  $\beta$  and  $\lambda(w_i)$  may be identified even if one of the random variables is a deterministic function of the other (see the discussion after Theorem 1 in Section 3.2 for more details).

I assume the researcher draws a random sample of  $n$  agents from the population. The agents in this sample are described by the sequence of independent and identically distributed random variables  $\{y_i, x_i, w_i\}_{i=1}^n$ , although the researcher only observes  $\{y_i, x_i\}_{i=1}^n$  as data. In order to identify and estimate  $\beta$  and  $\lambda(w_i)$ , the researcher also observes  $D$ , an  $n \times n$  stochastic binary adjacency matrix corresponding to an unlabeled, unweighted, and undirected random network between the  $n$  agents. The existence of a link between agents  $i$  and  $j$  is determined by the following model

$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\} \mathbb{1}\{i \neq j\} \quad (2)$$

in which  $f$  is an unknown symmetric Lebesgue measurable function and  $\{\eta_{ij}\}_{i,j=1}^n$  is a symmetric matrix of unobserved scalar disturbances with independent upper diagonal entries that are mutually independent of  $\{x_i, w_i, \varepsilon_i\}_{i=1}^n$ . In this paper, the marginal distribution of  $\eta_{ij}$  is not separately identified from  $f$  and so is also normalized to be standard uniform.

Network formation is represented by  $\binom{n}{2}$  conditionally independent Bernoulli trials in which the probability that agents  $i$  and  $j$  link is proportional to  $f(w_i, w_j)$ . Examples of (2) in the network formation literature include [Holland and Leinhardt \(1981\)](#), [Duijn, Snijders, and Zijlstra \(2004\)](#), [Krivitsky, Handcock, Raftery, and Hoff \(2009\)](#), [Chatterjee, Diaconis, and Sly](#)

(2011), McCormick and Zheng (2012), Dzemeski (2014), Graham (2017), Candelaria (2016), Toth (2017) and Nadler (2016). Many of these authors also consider directed networks, weighted networks, or include exogenous link covariates in the right hand side of (2). Such extensions are also possible in my setting but not pursued in this paper.

One way to interpret (2) is as a literal description of how a researcher might use a subjective survey question to elicit information about a nonrandom relationship between two agents. For instance, the researcher might survey a random sample of agents about whether or not they identify as friends with  $D_{ij} = 1$  if agents  $i$  and  $j$  report a friendship,  $f(w_i, w_j)$  the frequency of positive social interactions between them, and  $\eta_{ij}$  an error that allows for heterogeneity in  $i$  and  $j$ 's subjective interpretation of whether or not the nature of their interactions qualifies as a friendship. This interpretation of the network formation model is not behavioral, in the sense that it takes the social interactions as deterministic and attributes all of the randomness in  $D$  to either sampling or measurement error. In Section 2.3, I provide two alternative interpretations of the model that are behavioral, in that they consider the randomness in  $D$  to be the result of agents' stochastic preferences over particular configurations of network links.

The following three examples illustrate applications of the model to the literature.

**Example 1 (Network Peer Effects):** Let  $y_i$  be student GPA,  $x_i$  be a vector of student covariates (age, grade, gender, etc.), and  $D_{ij} = 1$  if students  $i$  and  $j$  are friends and 0 otherwise. One extension of the Manski (1993) linear-in-means peer effects model to the network setting is

$$y_i = x_i\beta + E[x_j|D_{ij} = 1, w_i]\rho_1 + E[y_j|D_{ij} = 1, w_i]\rho_2 + \rho_3(w_i) + \varepsilon_i$$

$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\} \mathbb{1}\{i \neq j\}$$

in which  $w_i$  measures student  $i$ 's social ability,  $E[x_j|D_{ij} = 1, w_i]$  denotes the expected covariates of agent  $i$ 's friends given his social ability,  $E[y_j|D_{ij} = 1, w_i]$  denotes the expected GPA of agent  $i$ 's friends given his social ability, and  $\rho_3(w_i)$  is the direct effect of social ability on GPA (students with more social ability may have higher family expectations about GPA). Social influence corresponds to the inside three terms on the right hand side

of the regression model:  $\lambda(w_i) = E[x_j|D_{ij} = 1, w_i]\rho_1 + E[y_j|D_{ij} = 1, w_i]\rho_2 + \rho_3(w_i)$ .

Identification problems stemming from the fact that all three terms are functions of  $w_i$  is discussed in Section 2.2. [Bramoullé, Djebbari, and Fortin \(2009\)](#) consider a similar model with  $\rho_3(w_i) = 0$  and [Goldsmith-Pinkham and Imbens \(2013\)](#), [Chan \(2014\)](#), [Hsieh and Lee \(2014\)](#), [Johnsson and Moon \(2015\)](#), and [Arduini, Patacchini, and Rainone \(2015\)](#) consider related models with additional functional form restrictions on  $\rho_3$  or  $f$ .

In Example 1, the use of the expected peer outcomes  $E[y_j|D_{ij} = 1, w_i]$  instead of their sample analogs  $\sum_j y_j D_{ij} / \sum_j D_{ij}$  reflects a particular interpretation about the model and sampling procedure: the peer groups that determine agent behavior are not related to the random sample drawn by the researcher. In contrast, the literature generally assumes that the researcher has sampled all of the other agents whose outcomes and characteristics influence agent  $i$ 's outcome.

**Example 2 (Information Diffusion):** [Banerjee, Chandrasekhar, Duflo, and Jackson \(2013\)](#) model household participation in a microfinance program in which information about the program diffuses over a social network. The authors control for household-level heterogeneity in program information by specifying and simulating a joint model of information diffusion and program participation. To simplify the example, I suppose that the authors have access to a continuous measure of program participation (extending my method to nonlinear models is straightforward, but left to future work) and propose the alternative

$$y_i = x_i\beta + \lambda(w_i) + \varepsilon_i$$

$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}\mathbb{1}\{i \neq j\}$$

in which  $i$  indexes participating households,  $y_i$  is a continuous measure of participation (amount of money borrowed),  $x_i$  is a vector of observed household characteristics (caste, religion, wealth, etc.),  $D_{ij} = 1$  if households  $i$  and  $j$  report a social connection, and  $w_i$  are characteristics that influence social network formation (e.g. villager gregariousness). The social influence term,  $\lambda(w_i)$ , gives the direct effect of villager gregariousness on program



participation (more gregarious villagers might all else equal learn more about the program and thus may be willing to borrow more money).

**Example 3 (Research Productivity):** [Ductor, Fafchamps, Goyal, and van der Leij \(2014\)](#) study a model of research productivity in which a researcher’s current publication quality depends on past quality, researcher covariates, and a vector of network statistics derived from a coauthorship network (in which two researchers are linked if they have previously been coauthors) including agent degree, eigenvector centrality, etc. The authors experiment with several different models of productivity, including various combinations of network statistics. An alternative treats the unknown combination of network statistics as unobserved social influence

$$y_i = x_i\beta + \lambda(w_i) + \varepsilon_i$$

$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\} \mathbb{1}\{i \neq j\}$$

in which  $w_i$  indexes researcher  $i$ ’s participation in various academic communities (for instance, fields of study, physical locations, etc.) and the social influence term  $\lambda(w_i)$  is the direct effect of interacting with a particular collection of communities (as indexed by  $w_i$ ) on research productivity. The idea that a vector of network statistics can be represented by a function of the agent’s latent social characteristics is explained in more detail in Section 2.3.

## 2.2 Main Identification Conditions and Estimators

This section motivates the main identification conditions and estimators for  $\beta$  and  $\lambda(w_i)$ , deferring formal results to Section 3. I first focus on the identification and estimation of  $\beta$  and treat  $\lambda$  and  $f$  as nuisance functions. If the social characteristics were observed, (1) corresponds to the partially linear regression of [Engle, Granger, Rice, and Weiss \(1986\)](#) and the identification and estimation of  $\beta$  is well understood (see [Chamberlain 1986](#), [Powell 1987](#), [Newey 1988](#), [Robinson 1988](#), [Ritov and Bickel 1990](#)). If the social characteristics were unobserved but identified by the distribution of  $D$  (that is,  $w_i \neq w'_i$  implies  $\int_{\tau \in A} f(w_i, \tau) d\tau \neq \int_{\tau \in A} f(w'_i, \tau) d\tau$  for some  $A \subseteq [0, 1]$  with nonzero Lebesgue measure), one might extend these

methods by replacing the social characteristics with empirical analogs as in [Ahn and Powell \(1993\)](#), [Ahn \(1997\)](#), and [Hahn and Ridder \(2013\)](#). This is the approach taken by [Arduini, Patacchini, and Rainone \(2015\)](#) and [Johnsson and Moon \(2015\)](#).

However, in this paper I do not assume that the social characteristics are either observed or identified by the distribution of  $D$ . Instead, I propose to identify and estimate  $\beta$  by matching pairs of agents with similar network types (an object I define below). This idea is motivated by the following two observations.

The first observation is that  $\beta$  is identified if  $\lambda(w_i)$  depends on  $w_i$  only through the network type  $f(w_i, \cdot) : [0, 1] \rightarrow [0, 1]$  and if there is excess variation in the distribution of  $x_i$  not explained by  $f(w_i, \cdot)$ . I will explain what I mean by these conditions first, and then the logic behind them. The network type  $f(u, \cdot)$  gives the conditional probability that an agent with social characteristics  $u$  links with agents of every other social characteristic in  $[0, 1]$ . To compare two agents' network types, I use network distance, which is defined to be the following pseudometric on the space of social characteristics

$$d(u, v) = \|f(u, \cdot) - f(v, \cdot)\|_2 = \left( \int (f(u, \tau) - f(v, \tau))^2 d\tau \right)^{1/2}$$

In words,  $d(u, v)$  is the integrated squared difference in the network types of agents with social characteristics  $u$  and  $v$ . The main identification conditions are then that  $\beta$  is identified if  $E[(x_i - x_j)'(\lambda(w_i) - \lambda(w_j)) | d(w_i, w_j) = 0] = 0$  and  $E[(x_i - x_j)'(x_i - x_j) | d(w_i, w_j) = 0]$  is positive definite. A formal definition of these conditional expectations is provided at the end of Section 3.1.

The logic behind the first identification condition is that under (2),  $f(w_i, \cdot)$  describes the totality of information that the distribution of  $D$  contains about  $w_i$ . That is, if  $d(w_i, w_j) = 0$  then there is no feature of the network that can distinguish between  $w_i$  and  $w_j$ . Agents  $i$  and  $j$  will have the same probability of being connected in any particular configuration of links, and thus will have the same distribution of degrees, eigenvector centralities, average peer characteristics, and any other agent-specific statistic of  $D$  (see Theorem B in Section 2.3). If  $E[(x_i - x_j)'(\lambda(w_i) - \lambda(w_j)) | d(w_i, w_j) = 0] \neq 0$ , then matching agents with similar network types will not control for all of the unobserved heterogeneity in (1), but under (2)

there is no further information in the distribution of  $D$  that can identify it. Additionally, when  $w_i$  is identified by the distribution of  $D$ ,  $d(w_i, w_j) = 0$  implies  $|w_i - w_j| = 0$ , so that  $E[(x_i - x_j)'(\lambda(w_i) - \lambda(w_j)) | d(w_i, w_j) = 0] = 0$  holds by definition. As a consequence, this first identification condition is more general than that imposed by the literature cited in Section 2.1.

The logic behind the second identification condition is that if  $E[(x_i - x_j)'(x_i - x_j) | d(w_i, w_j) = 0]$  is not positive definite, then there is a dimension of the covariate space such that all of the variation in  $y_i$  can be explained by  $w_i$  regardless of  $\beta$ . Thus  $\beta$  is not identified, because any value of  $\beta$  along this dimension is consistent with the data. An example of a model that fails this condition is the network peer effects model of Example 1, which I discuss below.

The second observation is that the average squared difference in the  $i$ th and  $j$ th columns of the squared adjacency matrix ( $D \times D$ ) can be used to bound  $d(w_i, w_j)$ . The logic has two steps. First, there exists another pseudometric  $\delta$  on the space of social characteristics such that  $d(w_i, w_j)$  can be bounded in terms of  $\delta(w_i, w_j)$ . Second,  $\delta(w_i, w_j)$  can be consistently estimated by the root average squared difference in the  $i$ th and  $j$ th columns of the squared adjacency matrix

$$\hat{\delta}_{ij} = \left( \frac{1}{n} \sum_{t=1}^n \left( \frac{1}{n} \sum_{s=1}^n D_{ts}(D_{is} - D_{js}) \right)^2 \right)^{1/2} \quad (3)$$

Here, the codegree  $\sum_{s=1}^n D_{ts}D_{is}$  gives the number of other agents that are linked to both agents  $i$  and  $t$ ,  $\{\sum_{s=1}^n D_{ts}D_{is}\}_{t=1}^n$  is the collection of codegrees between agent  $i$  and the other agents in the sample, and  $\hat{\delta}_{ij}$  gives the root average squared difference in  $i$ 's and  $j$ 's collection of codegrees. Similar relationships between configurations of network moments and the distribution of links are also used by [Bickel, Chen, and Levina \(2011\)](#), [Lovász and Szegedy \(2010\)](#), [Rohe, Chatterjee, and Yu \(2011\)](#), and [Zhang, Levina, and Zhu \(2015\)](#), but to different ends.

The two observations indicate that when the  $i$ th and  $j$ th columns of the squared adjacency matrix are similar and the identification conditions for  $\beta$  hold then  $(y_i - y_j)$  and  $(x_i - x_j)\beta + (\varepsilon_i - \varepsilon_j)$  are approximately equal. Under additional regularity conditions,  $\beta$  is then

consistently estimated by the pairwise difference estimator

$$\hat{\beta} = \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)'(x_i - x_j) K \left( \frac{\hat{\delta}_{ij}^2}{h_n} \right) \right)^{-1} \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)'(y_i - y_j) K \left( \frac{\hat{\delta}_{ij}^2}{h_n} \right) \right) \quad (4)$$

in which  $K$  is a kernel density function and  $h_n$  a bandwidth parameter depending on the sample size.

This estimator for  $\beta$  can be used to construct an estimator for  $\lambda(w_i)$  when the first main identification condition is strengthened to  $E[(\lambda(w_i) - \lambda(w_j))^2 | d(w_i, w_j) = 0] = 0$ . That is, if two agents have the same distribution of network links they have the same unobserved social influences. Under this assumption, consistency of  $\hat{\beta}$ , and additional regularity conditions,  $\lambda(w_i)$  is consistently estimated by the following nonparametric regression of the residuals  $(y_i - x_i \hat{\beta})$  on  $w_i$  using differences in the columns of the squared adjacency matrix  $\hat{\delta}_{ij}$ .

$$\widehat{\lambda}(w_i) = \left( \sum_{t=1}^n K \left( \frac{\hat{\delta}_{it}^2}{h_n} \right) \right)^{-1} \left( \sum_{t=1}^n (y_t - x_t \hat{\beta}) K \left( \frac{\hat{\delta}_{it}^2}{h_n} \right) \right) \quad (5)$$

This logic might also be used to estimate other network effects. For instance, another extension of the [Manski \(1993\)](#) linear-in-means peer effects model to the network setting is  $y_i = x_i \beta + E[x_i | w_i] \rho_1 + E[y_i | w_i] \rho_2 + \rho_3(w_i) + \varepsilon_i$ , in which  $E[x_i | w_i]$  is the expected covariates of agent  $i$  given his social characteristics. This model differs from that in Example 1 in that agents react to their expected characteristics rather than the expected characteristics of their friends. While the two models are identical in Manski's setting, when linking behavior is heterophilic, they have different implications. When  $\rho_3(\cdot) = 0$ , one might use  $\{\hat{\delta}_{ij}\}_{i \neq j}$  to estimate  $E[x_i | w_i]$  and  $E[y_i | w_i]$ , and then estimate  $\rho_1$  and  $\rho_2$  by regressing  $\widehat{\lambda}(w_i)$  on  $\widehat{E}[x_i | w_i]$  and  $\widehat{E}[y_i | w_i]$ .

I now discuss the identification of  $\beta$  and  $\lambda(w_i)$  in the context of Example 1.

**Example 1 (Network Peer Effects):** In the network peer effects model

$$y_i = x_i \beta + E[x_j | D_{ij} = 1, w_i] \rho_1 + E[y_j | D_{ij} = 1, w_i] \rho_2 + \rho_3(w_i) + \varepsilon_i$$

$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\} \mathbb{1}\{i \neq j\}$$

the parameter  $\beta$  is identified if the two main identification conditions hold. For example,  $\beta$  is identified if a student’s expected number of network links ( $\int f(w_i, \tau)d\tau$ ) is a monotonic function of social ability, but students with some fixed social ability do not all have the same covariates. The parameters  $\rho_1$  and  $\rho_2$  are not separately identified from the function  $\rho_3$ , since  $E[x_j|D_{ij} = 1, w_i] = E[x_j D_{ij}|w_i]/E[D_{ij}|w_i]$  is a function of  $w_i$ . In fact, the model violates the second identification condition because

$$E[x_j|D_{ij} = 1, w_i] = \int E[x_j|w_j = w]f(w_i, w)dw / \int f(w_i, w)dw$$

is a smooth functional of  $f(w_i, \cdot)$  (so long as  $\inf_{u \in [0,1]} \int f(u, w)dw > 0$ ). Identifying  $\rho_1$  and  $\rho_2$  in this model requires additional assumptions.

## 2.3 Discussion

This section discusses two motivations for the proposed model and main identification conditions. The first motivation is due to [Goldsmith-Pinkham and Imbens \(2013\)](#) and [Jackson \(2014\)](#), who view  $w_i$  as literally corresponding to an exogenous social attribute, such as socioeconomic status or social ability. Agent  $i$ ’s incentive to form links and his social influence in the regression model are both determined by  $w_i$ . For instance, a student’s socioeconomic status might determine both his friendships and his parent’s expectations about academic performance. Under this interpretation, the network is relevant because it allows the researcher to identify the agents’ social characteristics (up to the equivalence class defined by  $d$ ) and incorporate them into the regression model.

That the unobserved social characteristics can be learned from the columns of the squared adjacency matrix supposes that the true network formation model is of the form given in (2). A key implication of this model is that  $D_{ij}$  and  $D_{kl}$  are independent conditional on  $\{w_i, w_j, w_k, w_l\}$ . This assumption is also made by the literature cited in Section 2.1. However, when  $f(w_i, w_j) - \eta_{ij}$  is interpreted as the utility agents  $i$  and  $j$  receive from forming a link, the assumption is often thought to be unrealistic because it does not allow for endogenous link formation, a phenomena in which the utility agents  $i$  and  $j$  receive from forming a link depends on the existence of other links in the sample [Sheng \(2012\)](#), [Leung \(2014; 2015\)](#),

Ridder and Sheng (2015), Menzel (2015), Mele (2017) and Mele and Zhu (2017) all consider network formation models in which some linking behavior is endogenous.

If the network formation model exhibits endogenous link formation, then under this first motivation the columns of the squared adjacency matrix do not necessarily reveal any meaningful information about the underlying social characteristics, and the methodology of this paper is potentially invalid. Recent work by Hsieh and Lee (2014), Griffith (2016) and Badev (2017) consider parametric models of social interaction and network formation that, among other things, explicitly account for endogenous link formation. The extent to which the parametric structure in their models can be relaxed as in (2) is to my knowledge an open question.

In the second motivation, the social characteristics do not literally correspond to exogenous agent attributes. Instead, (2) is a reduced form description of the equilibrium distribution of network links implied by some unspecified structural network formation game on a population that is large relative to the sample size. Linking behavior in this game may be endogenous. This interpretation relies on two main results that I summarize first and then explain below.

The first result is that many network formation models from the economics literature generate a distribution of network links that can be described by (2), for some choice of linking function  $f$  and collection of social characteristics  $\{w_i\}_{i=1}^n$ . The second result (original to this paper) is that many network statistics from the economics literature can be approximated by functionals of the network types implied by this reduced form approximation. When the network formation game occurs on a population that is large relative to the size of the sample observed by the researcher, any regression model in which the outcome depends on a vector of observed covariates and an unknown combination of population network statistics (satisfying certain regularity conditions) is closely approximated by (1), for some social influence function  $\lambda$  satisfying the first main identification condition of Section 2.2. I first explain the results below, describe the interpretation in more detail, and then illustrate the argument using the Example 3 from Section 2.1.

The first result is that if the network formation model is jointly exchangeable then the equilibrium distribution of network links is described by (2) for some choice of  $f$  and  $\{w_i\}_{i=1}^n$ .

In this paper, a network formation model on  $n$  agents is characterized by the joint distribution of the elements of the associated random  $n \times n$  adjacency matrix  $\{D_{ij}\}_{i \neq j}$ , and is jointly exchangeable if for any automorphism  $\pi$  on  $\{1, \dots, n\}$ ,  $\{D_{ij}\}_{i \neq j}$  and  $\{D_{\pi(i)\pi(j)}\}_{i \neq j}$  are equal in distribution. Intuitively, joint exchangeability imposes that the distribution of network links does not depend on how the agents are indexed. Almost all of the network formation models cited in this paper are jointly exchangeable.

The following theorem is generally attributed to Hoover (1979), Aldous (1981) and Kallenberg (1989), although this particular version is Corollary III.6 to Theorem III.2 of Orbanz and Roy (2015)

**Theorem A:** The network formation model characterized by  $\{D_{ij}\}_{i \neq j}$  is jointly exchangeable if and only if there exist iid uniform random variables  $w$ ,  $\{w_i\}_{i=1}^n$  and  $\{\eta_{ij}\}_{i \neq j}$  that are mutually independent of each other and a measurable function  $f$  such that

$$D_{ij} =_d \mathbb{1}\{\eta_{ij} < f(w, w_i, w_j)\} \quad (6)$$

The theorem is similar in spirit to arguments made by Leung (2015), Menzel (2015), Ridder and Sheng (2015), Mele and Zhu (2017), who broadly view (6) as either a limiting game associated with a model with endogenous link formation or as a reduced form description of the within-equilibrium distribution of links between agents (the random variable  $w$ , which indexes variation at the population level, captures variation due to equilibrium selection). Conditional on  $w$ , (6) is equivalent to (2), and the arguments of this paper can be applied. In what follows, I refer to the function  $f(w, \cdot, \cdot)$  as the reduced form linking function and  $f(w, w_i, \cdot)$  as agent  $i$ 's reduced form network type.

The second result is that even if (6) does not literally describe the structural network formation game that generated the observed network links, the main identification conditions of this paper may still be satisfied if the social influence in the regression model can be described by a combination of agent-specific network statistics as formalized below. A similar class of regression models is studied by Chandrasekhar and Lewis (2011) (and includes the three main examples from Section 2.1).

Let  $\lambda(D, i)$  denote an arbitrary agent-specific network statistic, that is a real-valued func-

tion of an adjacency matrix  $D$  and an agent index  $i$  from some network on  $m$  agents, satisfying two assumptions. The first assumption is about symmetry: for any automorphism  $\pi$  on  $\{1, \dots, m\}$  such that  $\pi(i) = i$ ,  $\lambda(D, i) = \lambda(\{D_{\pi(s), \pi(t)}\}_{s \neq t}, i)$ . In words, the assumption says that the network statistic for agent  $i$  does not depend on how the other agents are indexed. The second assumption is about bounded deviations: for any two  $m \times m$  adjacency matrixes  $D$  and  $D'$ ,  $|\lambda(D, i) - \lambda(D', j)| = O_p\left(\frac{1}{m} \sum_{t \neq i, j} |D_{it} - D'_{jt}| + \frac{1}{m^2} \sum_{s, t \neq i, j} |D_{st} - D'_{st}|\right)$ . This assumption states that altering one of agent  $i$ 's links in the network only changes the statistic for agent  $i$  by a factor of  $O\left(\frac{1}{m}\right)$  while altering any other link only changes it by  $O\left(\frac{1}{m^2}\right)$ . Many network statistics from the economics literature satisfy these conditions, including average degree, eigenvector centrality, and average peer characteristics.

**Theorem B:** Suppose  $\lambda(D, i)$  satisfies the above symmetry and bounded deviations assumptions. Further suppose (6) holds. Then

$$|\lambda(D, i) - \lambda(D, j)| \leq C \|f(w, w_i, \cdot) - f(w, w_j, \cdot)\|_2 + O_p(m^{-1/2})$$

for some  $C$  depending on  $f$  and  $w$ .

The proof of Theorem B can be found in the Appendix. The choice of  $1/m$  in the bounded deviations condition is arbitrary, affecting only the rate of convergence on the right hand side of the conclusion.

Theorems A and B motivate a class of models for which the conditional independence assumption does not restrict agent behavior in the network formation process. The model is defined on a large population of size  $m$ . On this population, agents make linking decisions according to some jointly exchangeable network formation game, represented by (6). Agent outcomes in the regression model are a linear function of the observed covariates and a collection of unknown population network statistics defined on an equilibrium of the game and satisfying the hypothesis of Theorem B. The researcher then draws a random sample of agents of size  $n$ , where  $n$  is small relative to  $m$ . The relationship between the outcome, covariates, and network statistics is then described by (1), for some choice of social influence function  $\lambda$ , up to a negligible error. It also satisfies the first main identification assumption of Section 2.2 (in which network distance is now defined using the reduced form network



types  $d(w, w_i, w_j) := \|f(w, w_i, \cdot) - f(w, w_j, \cdot)\|_2$ .

I now apply this second interpretation to Example 3.

**Example 3 (Research Productivity):** Consider a variation on Example 3 from Section 2.1 in which  $y_i$ , the productivity of researcher  $i$ , is explained by the model

$$y_i = x_i\beta + \lambda(D^*, i) + \varepsilon_i$$

in which researcher  $i$  belongs to a large population of size  $m$ ,  $D^*$  is an  $m \times m$  binary adjacency matrix corresponding to a random network on that population (where two researchers are linked if they interact professionally), and  $\lambda(D^*, i)$  is the direct effect of the collection of professional interactions on agent  $i$ 's research quality. Suppose  $D^*$  and  $\lambda(D^*, i)$  satisfy the hypothesis of Theorems A and B and the two main identification assumptions. For example, the professional interactions matrix  $D^*$  might correspond to an equilibrium of some unspecified economic game and  $\lambda(D^*, i)$  is the equilibrium quantity of connections researcher  $i$  has with key researchers in the  $m$ -sized population. The data then consists of, for a random sample of size  $n$  drawn from this population, the outcomes, covariates, and professional links connecting these researchers. Following the above arguments, the joint distribution of the data can be described by the model

$$y_i = x_i\beta + \lambda(w, w_i) + O_p(m^{-1/2}) + \varepsilon_i$$

$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w, w_i, w_j)\} \mathbb{1}\{i \neq j\}$$

in which  $\lambda(w, w_i)$  is a Lipschitz continuous functional of  $f(w, w_i, \cdot)$  as per Theorem B and  $f$  is the reduced form linking function implied by Theorem A. When  $m$  is large relative to  $n$  (that is  $m/n \rightarrow \infty$ ), the approximation error in the regression model does not affect the identification and estimation of  $\beta$  or  $\lambda(w_i)$ , and the methodology of this paper can be applied.

### 3 Identification and Large Sample Results

This section formalizes the discussion about identification and estimation from Section 2.2. Section 3.1 introduces notation, Section 3.2 discusses the identification of  $\beta$  and  $\lambda(w_i)$ , and Sections 3.3 and 3.4 characterize the large sample properties of  $\hat{\beta}$  and  $\widehat{\lambda}(w_i)$  respectively.

#### 3.1 Terminology and Notation

I define agent  $i$ 's network type to be the projection of the link function  $f$  onto his or her social characteristics:  $f_{w_i}(\cdot) := f(w_i, \cdot) : [0, 1] \rightarrow [0, 1]$ . In words, it is the collection of probabilities that agent  $i$  links to agents with each social characteristic in  $[0, 1]$ . I consider network types to be elements of  $L^2([0, 1])$ , the usual inner product space of square integrable functions on the unit interval. The previously defined pseudometric  $d(w_i, w_j) = \|f_{w_i} - f_{w_j}\|_2$  is the usual  $L^2$  metric on the space of network types.

I also define two constructions from network theory: (average) agent degrees and (average) agent-pair codegrees. The degree of agent  $i$  is  $(n - 1)^{-1} \sum_{t \neq i} D_{it}$ , the fraction of other agents linked to agent  $i$ . Under (2), that  $(n - 1)^{-1} \sum_{t \neq i} D_{it} \rightarrow_{a.s.} \int f_{w_i}(\tau) d\tau$  follows from the usual strong law of large numbers. Similarly, for  $i \neq j$  the codegree of agent pair  $(i, j)$  is  $(n - 2)^{-1} \sum_{t \neq i, j} D_{it} D_{jt}$ , the fraction of other agents linked to both agent  $i$  and agent  $j$ . Again, under (2),  $(n - 2)^{-1} \sum_{t \neq i, j} D_{it} D_{jt} \rightarrow_{a.s.} \int f_{w_i}(\tau) f_{w_j}(\tau) d\tau = \langle f_{w_i}, f_{w_j} \rangle_{L^2}$ . For reference, I denote this codegree by  $\hat{p}_{ij}$  and its almost sure limit with  $p(w_i, w_j)$  or  $p_{ij}$ . I emphasize that  $p(w_i, w_i)$  refers to the limiting codegree of two distinct agents with social characteristics equal to  $w_i$  and *not* to the limiting degree of agent  $i$ . That is,  $p(w_i, w_i) := \int f_{w_i}(\tau)^2 d\tau = \|f_{w_i}\|_2^2 \neq \int f_{w_i}(\tau) d\tau$ .

The function  $p$  also defines a link function in which  $p(w_i, w_j)$  gives the probability that agents  $i$  and  $j$  have a link in common, as opposed to  $f(w_i, w_j)$ , which gives the probability that they are directly linked themselves. To distinguish  $p$  from  $f$  I refer to it as the codegree link function (associated with  $f$ ), and the function  $p_{w_i}(\cdot) := p(w_i, \cdot) : [0, 1] \rightarrow [0, 1]$  as agent  $i$ 's codegree type, also taken to be an element of  $L^2([0, 1])$ . I refer to the pseudometric on

$[0, 1]$  induced by  $L^2$ -differences in codegree types with  $\delta$ , so that

$$\delta(u, v) = \|p(u, \cdot) - p(v, \cdot)\|_2 = \left( \int \left( \int f(\tau, s) (f(u, s) - f(v, s)) ds \right)^2 d\tau \right)^{1/2}$$

for any pair of social characteristics  $u$  and  $v$ .

I also use two different conditional expectations defined over events on the network types. Let  $Z_i$  and  $Z_{ij}$  be arbitrary random matrices indexed at the agent and agent-pair level respectively. Then for any positive real  $x$ ,  $E[Z_{ij} | \|f_{w_i} - f_{w_j}\|_2 = x]$  refers to

$$\lim_{h \rightarrow 0} E[Z_{ij} | (w_i, w_j) \in \{(u, v) : x \leq \|f_u - f_v\|_2 \leq x + h\}]$$

and for any  $f$  in  $L^2([0, 1])$ ,  $E[Z_i | f_{w_i} = f]$  refers to

$$\lim_{h \rightarrow 0} E[Z_i | w_i \in \{w : \|f_w - f\|_2 \leq h\}]$$

Though  $f_{w_i}$  is a random function, these conditional expectations implicitly refer to the measure induced by the random variable  $w_i$ . Conditional expectations with respect to the codegree types are defined in the same way.

## 3.2 Identification

This section restates the two main identification conditions from Section 2.2 that are sufficient for  $\beta$  and  $\lambda(w_i)$  to be identified by  $L^2$  differences in the agent network types. That the network types are identified by the distribution of  $D$  follows from Lemmas 1 and 2 in Section 3.3.

**Assumption 1:** The random sequence  $\{x_i, \varepsilon_i, w_i\}_{i=1}^n$  is independent and identically distributed with entries mutually independent of  $\{\eta_{ij}\}_{i,j=1}^n$ , a symmetric random array with independent and identically distributed entries above the diagonal. The outcomes  $\{y_i\}_{i=1}^n$  and  $D$  are given by equations (1) and (2) respectively. The variables  $x_i$  and  $\varepsilon_i$  have finite eighth moments,  $w_i$  and  $\eta_{ij}$  have standard uniform marginals, and  $E[\varepsilon_i | x_i, w_i] = 0$ .

Assumption 1 restates the model discussed in Section 2.1 and is included as a reference. Since the marginal distributions of  $w_i$  and  $\eta_{ij}$  are not separately identified from  $f$ , the assumption of standard uniform marginals is without loss (see [Bickel and Chen 2009](#)).

**Assumption 2:** The covariance matrix  $\Gamma_0 := E [(x_i - x_j)'(x_i - x_j) | \|f_{w_i} - f_{w_j}\|_2 = 0]$  is positive definite.

Assumption 2 is a full rank condition that states that there is independent variation in each of the regressors not explained by the network types. This assumption may be unrealistic when the regressors include agent-specific network statistics as discussed in Example 1 of Section 2.2 (see also Theorem B of Section 2.3).

**Assumption 3:** The social influence function  $\lambda$  satisfies  $E [(\lambda(w_i) - \lambda(w_j))^2 | \|f_{w_i} - f_{w_j}\|_2 = 0] = 0$ .

Assumption 3 states that agents with similar network types have similar social influences, as motivated in Section 2.2. The parameter  $\beta$  is also identified under the weaker orthogonality condition  $E [(x_i - x_j)'(\lambda(w_i) - \lambda(w_j)) | \|f_{w_i} - f_{w_j}\|_2 = 0] = 0$ .

**Theorem 1:** Suppose Assumptions 1-3 hold. Then  $\beta$  is the unique minimizer of  $E [((y_i - y_j) - (x_i - x_j)b)^2 | \|f_{w_i} - f_{w_j}\|_2 = 0]$  over  $b \in \mathbb{R}^k$  and  $\lambda(w_i) = E [(y_i - x_i\beta) | f_{w_i}]$ .

The proof of Theorem 1 follows from standard arguments. Assumptions 1-3 do not rule out cases where either  $x_i$  or  $w_i$  is a deterministic function of the other variable. For example, if  $x_i = w_i$ , the assumptions may still be satisfied if  $|\{w \in [0, 1] : d(w, w_i) = 0\}| > 0$  for almost every  $w_i$ . That is, more than one value of social characteristics is associated with any particular network type. The implication is that it is generally acceptable to include observed drivers of link formation in the right hand side of the regression model, so long as these variables cannot be perfectly predicted (in the mean-squared sense) by the agents' network types.

### 3.3 Large Sample Properties of $\hat{\beta}$

Section 3.3.1 provides sufficient conditions for  $\hat{\beta}$  to be consistent for  $\beta$ . Section 3.3.2 provides sufficient conditions for its limiting distribution to be normal. Accurate inference may require a bias correction and Section 3.3.3 provides sufficient conditions such that a variation on the jackknife method of [Honoré and Powell \(1997\)](#) can be used for this purpose. Section 3.3.4 provides two consistent estimators for the asymptotic variance.

#### 3.3.1 Consistency

I suppose the bandwidth sequence and kernel density function used in (4) satisfy the following conditions.

**Assumption 4:** The bandwidth sequence satisfies  $h_n \rightarrow 0$ ,  $n^{1-\gamma}h_n^2 \rightarrow \infty$  for some  $\gamma > 0$ , and  $nr_n \rightarrow \infty$  for  $r_n = E \left[ K \left( \frac{\|p_{w_i} - p_{w_j}\|_2}{h_n} \right) \right]$  as  $n \rightarrow \infty$ .  $K$  is supported, bounded, and differentiable on  $[0, 1]$ , strictly positive and smooth on  $[0, 1)$ , and bounded away from 0 on  $[0, .5]$ .

The restrictions on the kernel density function are satisfied by a type-II kernel density function (examples include the Epanechnikov, Biweight, and Bartlett kernels). The first two restrictions on the bandwidth sequence are also standard. The third, that  $nr_n \rightarrow \infty$ , ensures that the number of matches used to estimate  $\hat{\beta}$  increases with  $n$ . If  $p_{w_i}$  was a  $d$ -dimensional random vector with compact support and a strictly positive density function,  $P(\|p_{w_i} - p_{w_j}\|_2 \leq h_n)$  would be on the order of  $h_n^d$ . The number of agent-pairs with similar codegree types would then be on the order of  $nh_n^d$ , which increases with  $n$  if  $n^{1-\gamma}h_n^d \rightarrow \infty$ . Since  $p_{w_i}$  is infinite dimensional,  $P(\|p_{w_i} - p_{w_j}\|_2 \leq h_n)$  cannot necessarily be approximated by a polynomial of  $h_n$  of known order and so the third condition is required. One can verify it in practice (in the same sense that one can choose a sequence of bandwidths that satisfies the first two conditions) by computing the empirical analog of  $r_n$  and choosing  $h_n$  such that this statistic is large relative to  $1/n$ . The framework of this paper also allows for  $h_n$  to be chosen in a data-dependent way, for example by cross-validation in the sense of [Hall \(1984\)](#), [Stone \(1984\)](#) and [Hardle and Marron \(1985\)](#) (see [Nolan and Pollard 1987; 1988](#)), however I leave the formal study of such an estimator to future work.

If the collection of network differences between agents  $\{\|f_{w_i} - f_{w_j}\|_2\}_{i \neq j}$  were observed and used to construct the matches in  $\hat{\beta}$ , the arguments for consistency would be similar to those of [Ahn and Powell \(1993\)](#), though with alterations to accommodate the dimensionality of  $f_{w_i}$ . That the estimator is still consistent when  $\|f_{w_i} - f_{w_j}\|_2$  is replaced by  $\hat{\delta}_{ij}$  follows from two arguments. First,  $\{\hat{\delta}_{ij}\}_{i \neq j}$  converges uniformly to  $\{\|p_{w_i} - p_{w_j}\|_2\}_{i \neq j}$  over all agent-pairs. Second, agent-pairs similar with respect to the codegree distance are also similar with respect to the network distance. These results are the following Lemmas 1 and 2 respectively.

**Lemma 1:** Suppose Assumptions 1 and 4 hold. Then

$$\max_{i \neq j} \left| \hat{\delta}_{ij} - \|p_{w_i} - p_{w_j}\|_2 \right| = o_p(n^{-\gamma/4} h_n)$$

in which  $\gamma$  refers to the exponent from Assumption 4.

Lemma 1 demonstrates that the collection of  $\binom{n}{2}$  empirical codegree distances converges uniformly to their population analogs at a rate slightly slower than  $n^{-1/2}$ . The proof involves repeated applications of Bernstein's Inequality and the union bound over the  $\binom{n}{2}$  distinct empirical codegrees that make up  $\{\hat{\delta}_{ij}\}_{i \neq j}$ .

**Lemma 2:** Suppose Assumption 1 holds. Then for every  $\epsilon > 0$  there exists a  $\delta > 0$  such that with probability at least  $1 - \epsilon^2/4$

$$\|p_{w_i} - p_{w_j}\|_2 \leq \delta \implies \|f_{w_i} - f_{w_j}\|_2 \leq \epsilon$$

Lemma 2 is the main justification for the codegree matching procedure. The result is somewhat unexpected since  $\|p_{w_i} - p_{w_j}\|_2 \leq \|f_{w_i} - f_{w_j}\|_2$  is almost an immediate consequence of Jensen's inequality. That is,

$$\begin{aligned} \|p_{w_i} - p_{w_j}\|_2^2 &= \int \left( \int f(t, s) (f(w_i, s) - f(w_j, s)) ds \right)^2 dt \\ &\leq \int \left( \int (f(t, s) (f(w_i, s) - f(w_j, s)))^2 ds \right) dt \\ &\leq \int (f(w_i, s) - f(w_j, s))^2 ds = \|f_{w_i} - f_{w_j}\|_2^2 \end{aligned}$$

where the first inequality is due to Jensen and the second due to the fact that  $f$  is bounded between 0 and 1. Lemma 2 is related to Theorem 13.27 of Lovász (2012), the logic of which demonstrates that  $\|p_{w_i} - p_{w_j}\|_2 = 0$  implies  $\|f_{w_i} - f_{w_j}\|_2 = 0$  when  $f$  is continuous. Its proof is sketched below.

$$\begin{aligned}
\|p_{w_i} - p_{w_j}\|_2^2 = 0 &\implies \int \left( \int f(\tau, s) (f(w_i, s) - f(w_j, s)) ds \right)^2 d\tau = 0 \\
&\implies \int f(\tau, s) (f(w_i, s) - f(w_j, s)) ds = 0 \text{ for every } \tau \\
&\implies \int f(w_i, s) (f(w_i, s) - f(w_j, s)) ds = 0 \text{ and } \int f(w_j, s) (f(w_i, s) - f(w_j, s)) ds = 0 \\
&\implies \int (f(w_i, s) - f(w_j, s))^2 ds = 0 \implies \|f_{w_i} - f_{w_j}\|_2^2 = 0
\end{aligned}$$

The intuition is that if agents  $i$  and  $j$  have identical codegree types, then the difference in their network types ( $f_{w_i} - f_{w_j}$ ) must be uncorrelated with each other network type in the population, as indexed by  $\tau$ . In particular, the difference is uncorrelated with  $f_{w_i}$  and  $f_{w_j}$ , the network types of agents  $i$  and  $j$ . However, this can only be the case if  $f_{w_i}$  and  $f_{w_j}$  are perfectly correlated.

Lovász's theorem demonstrates that agent-pairs with identical codegree types also have identical network types. However, consistency of  $\hat{\beta}$  requires a stronger result, that agent-pairs with similar but not necessarily equivalent codegree types have similar network types. This is the statement of Lemma 2. Unfortunately the above proof cannot simply be extended by replacing each occurrence of 0 with some function of a small  $\epsilon > 0$ , because the third implication relies on  $\int f(\tau, s) (f(w_i, s) - f(w_j, s)) ds = 0$  for exactly all  $\tau$ , which is not guaranteed by the condition  $\|p_{w_i} - p_{w_j}\|_2^2 \leq \epsilon$  for any  $\epsilon > 0$ . Still, the proof of Lemma 2 demonstrates that the two notions of distance are similar in enough places that matching agents with similar codegree types is sufficient to partial out  $\lambda(w_i)$  in the regression model (1) and consistently estimate  $\beta$  under Assumptions 1-4.

**Theorem 2:** Suppose Assumptions 1-4 hold. Then  $(\hat{\beta} - \beta) \rightarrow_p 0$  as  $n \rightarrow \infty$ .

The proof of Theorem 2 is a direct consequence of Lemmas 1 and 2 and the continuous mapping theorem.

### 3.3.2 Asymptotic Normality

I provide two asymptotic normality results. The first concerns the case when the distribution of  $f_{w_i}$  has finite support in that  $P(\|f_{w_i} - f_{w_j}\|_2 = 0) = P(\|p_{w_i} - p_{w_j}\|_2 = 0) > 0$  and there exists an  $\epsilon > 0$  such that  $P(0 < \|f_{w_i} - f_{w_j}\|_2 < \epsilon) = P(0 < \|p_{w_i} - p_{w_j}\|_2 < \epsilon) = 0$ . This assumption is satisfied by the stochastic blockmodel of [Holland, Laskey, and Leinhardt \(1983\)](#) (see also [Bickel, Choi, Chang, and Zhang 2013](#)).

**Theorem 3:** Suppose Assumptions 1-4 hold and  $f_{w_i}$  has finite support. Then as  $n \rightarrow \infty$

$$V_{3,n}^{-1/2} \left( \hat{\beta} - \beta \right) \rightarrow_d \mathcal{N}(0, I_k)$$

where  $V_{3,n} = \Gamma_0^{-1} \Omega_0 \Gamma_0^{-1} \times s/n$ ,  $\Gamma_0$  is as defined in Assumption 3,  $I_k$  is the  $k \times k$  identity matrix, and

$$s = P(\|p_i - p_j\|_2 = 0, \|p_i - p_k\|_2 = 0) / P(\|p_i - p_j\|_2 = 0)^2$$

$$\Omega_0 = E[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k) \mid \|p_i - p_j\|_2 = 0, \|p_i - p_k\|_2 = 0]$$

with  $u_i = \lambda(w_i) + \varepsilon_i$ .

Theorem 3 is included in this paper for three reasons. First, it adds to a literature noting that some of the adverse effects of unobserved heterogeneity might be mitigated when the support of this variation is finite (see [Hahn and Moon 2010](#), [Bonhomme and Manresa 2015](#)). Second, the assumption of discrete heterogeneity is not uncommon in empirical work (for instance [Schmutte 2014](#)). Third, it provides an easy to interpret condition such that  $\hat{\beta}$  converges to  $\beta$  at the  $\sqrt{n}$ -rate.

The second result concerns the more general case when the support of  $f_{w_i}$  is not necessarily finite. It requires additional structure on the linking function  $f$  and the bandwidth sequence  $h_n$  given in Assumptions 5 and 6 respectively.

**Assumption 5:** There exists an integer  $K$  and a partition of  $[0, 1)$  into  $K$  equally spaced, adjacent, and non-intersecting intervals  $\cup_{t=1}^K [x_t^1, x_t^2)$  such that for any  $t \in \{1, \dots, K\}$  and



almost every  $x, y \in [x_t^1, x_t^2)$  and  $s \in [0, 1]$ ,  $|f(x, s) - f(y, s)| \leq C_5|x - y|^\alpha$ , for some  $C_5 \geq 0$  and  $\alpha > 0$ .

Assumption 5 supposes that the space of social characteristics can be partitioned into  $K$  segments such that on each partition segment the link function  $f$  is almost everywhere Hölder continuous of order  $\alpha$ . The partition allows for discrete jumps of the link function and so includes the discrete heterogeneity models from Theorem 3 as a special case. The restriction that the partition is uniformly sized is without loss.

**Assumption 6:** The bandwidth sequence  $h_n = C_7 \times n^{-\rho}$  for  $\rho \in \left(\frac{\alpha}{4+8\alpha}, \frac{\alpha}{2+4\alpha}\right)$  and some  $C_7 > 0$ .  $K$  is supported, bounded, and differentiable on  $[0, 1]$ , strictly positive and smooth on  $[0, 1)$ , and bounded away from 0 on  $[0, .5]$ .

The assumptions on the kernel density function in Assumption 6 are the same as in Assumption 4. However, the rate of convergence of the bandwidth sequence now depends on the exponent from Assumption 5. When  $\alpha = 1$  this bandwidth choice is approximately on the order of magnitude used by [Ahn and Powell \(1993\)](#). In this paper,  $\alpha$  is a parameter to be chosen by the researcher. All of the network formation models cited in Section 2.1 essentially assume  $\alpha = 1$ .

The second asymptotic normality proof uses Assumption 5 to strengthen Lemma 2 in the following way.

**Lemma 3:** Suppose Assumptions 1 and 5 hold. Then for almost every  $(w_i, w_j)$  pair

$$\|p_{w_i} - p_{w_j}\|_2 \leq \|f_{w_i} - f_{w_j}\|_2 \leq 32 C_6^{\frac{1}{2+4\alpha}} \left(\|p_{w_i} - p_{w_j}\|_2\right)^{\frac{\alpha}{1+2\alpha}}$$

so long as  $\|p_{w_i} - p_{w_j}\|_2 < \sqrt{8C_6}K^{-\alpha}$ , where  $C_5$  and  $\alpha$  are the constants from Assumption 5.

**Theorem 4:** Suppose Assumptions 1-3 and 5-6 hold. Then as  $n \rightarrow \infty$

$$V_{4,n}^{-1/2} \left( \hat{\beta} - \beta_{h_n} \right) \rightarrow_d \mathcal{N}(0, I_k)$$

where  $V_{4,n} = \Gamma_0^{-1} \Omega_n \Gamma_0^{-1} / n$ ,  $\Gamma_0$  is as defined in Assumption 3,  $r_n$  is as defined in Assumption 4,  $I_k$  is the  $k \times k$  identity matrix, and

$$\begin{aligned} \beta_{h_n} &= \beta + (\Gamma_0)^{-1} E \left[ (x_i - x_j)' (\lambda(w_i) - \lambda(w_j)) K \left( \frac{\|p_i - p_j\|_2}{h_n} \right) \right] / (2r_n) \\ \Omega_n &= \frac{4}{r_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_1 j_2} K \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) \right] \\ &\quad + \frac{1}{r_n^2 h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_1 s_{21} s_{22}} - \delta_{i_2 j_2}^2) \right] \\ &\quad + \frac{4}{r_n^2 h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_2 s_{11} s_{22}} - \delta_{i_2 j_2}^2) \right] \end{aligned}$$

with  $\Delta_{ij} = (x_i - x_j)'(u_i - u_j)$ ,  $u_i = \lambda(w_i) + \varepsilon_i$ ,  $\delta_{ij} = \delta(w_i, w_j) = \|p_{w_i} - p_{w_j}\|_2$ , and  $F_{ij t s_1 s_2} = f(w_t, w_{s_1}) f(w_t, w_{s_2}) (f(w_i, w_{s_1}) - f(w_j, w_{s_1})) (f(w_i, w_{s_2}) - f(w_j, w_{s_2}))$ .

The statement of Theorem 4 warrants two remarks. First, the variance is not necessarily on the order of the inverse of the sample size. This is because the variance of the kernel  $r_n^{-2} E \left[ K \left( \frac{\|p_i - p_j\|_2}{h_n} \right) K \left( \frac{\|p_i - p_k\|_2}{h_n} \right) \right]$  can potentially diverge with  $n$ . Even when this variance diverges, Assumptions 5, 6, and Lemma 3 ensure that the rate of convergence for  $V_{4,n}$  is on the order of at least  $n^{-1/2}$ . One can remove this term from the variance (that is, set  $K \left( \frac{\|p_i - p_j\|_2}{h_n} \right) = r_n$ ) by choosing a variable bandwidth in which each agent belongs to the same number of matches, though the strategy also generally inflates the bias of the estimator relative to  $\hat{\beta}$ .

The variance is also inflated relative to the infeasible pairwise difference regression using the unknown codegree distances  $\{\delta(w_i, w_j)\}_{i \neq j}$ , due to the variability of the estimated codegree differences  $\{\hat{\delta}_{ij}\}_{i \neq j}$  around their probability limits. A previous version of this paper gave conditions such that these components of the variance are asymptotically small. However, the current statement of Theorem 4 is more general and potentially allows for more accurate inferences. In Section 3.3.4, I provide two consistent estimators for  $V_{4,n}$ .

The second remark is that the asymptotic distribution of  $\hat{\beta}$  is not centered at  $\beta$ , but at the pseudo-truth  $\beta_{h_n}$ . Though  $\beta_{h_n}$  converges to  $\beta$ , the rate of convergence can be slow depending on the size of  $\alpha$  and the conditional expectation function  $E[(x_i - x_j)'(\lambda(w_i) - \lambda(w_j)) | \|f_{w_i} - f_{w_j}\|_2 = h_n]$ . This problem is common with matching estimators (see also

Abadie and Imbens 2006; 2012), although the problem is exacerbated here by the relatively weak relationship between the codegree and network distances given by Lemma 3. Accurate inferences about  $\beta$  using Theorem 4 will generally require a bias correction.

### 3.3.3 Bias Correction

I propose a variation on the jackknife technique of Honoré and Powell (1997), which relies on the following smoothness condition.

**Assumption 7:** The pseudo-truth function  $\beta_h$  satisfies  $\beta_h = \sum_{l=1}^L C_l h^{l/\theta} + O(h^{(L+1)/\theta})$  for some positive integer  $L > 2\theta(1 + 2\alpha)/\alpha - 1$ ,  $k$ -dimensional constants  $C_1, C_2, \dots, C_L$ ,  $\theta > 0$ , and  $h$  in a fixed open neighborhood to the right of 0.

Assumption 7 assumes that the pseudo-truth  $\beta_{h_n}$  can be well approximated by a series of fractional polynomials. The assumption holds, for example, with  $\theta = 1$  if  $f$  and  $\lambda$  are smooth functions and  $\delta^{-1}(w_i, h) = \{w \in [0, 1] : \delta(w_i, w) \leq h\}$  is equal to the union of a finite (uniformly over  $i = 1, \dots, n$ ) number of disjoint intervals for almost every  $w_i$  and all  $h$  in an open interval to the right of 0. In other words, the measure of other social characteristics that are  $\delta$ -similar to  $w_i$  (i.e. have codegree distance  $\leq h$ ) does not change drastically for  $h$  near zero. Most of the network formation models cited in Section 2.1 satisfy these conditions for  $\alpha = \theta = 1$ . In this case  $L$  may be chosen (depending on the choice of bandwidth sequence) to be between 2 and 5.

The method produces a bias-corrected estimator  $\bar{\beta}_L$ . For an arbitrary sequence of distinct positive numbers  $\{c_1, c_2, \dots, c_L\}$  with  $c_1 = 1$ ,  $\bar{\beta}_L$  is defined to be

$$\bar{\beta}_L = \sum_{l=1}^L a_l \hat{\beta}_{c_l h_n} \tag{7}$$

in which  $\hat{\beta}_{c_l h_n}$  refers to the pairwise difference estimator (4) with the choice of bandwidth

$c_l \times h_n$  and the sequence  $\{a_1, a_2, \dots, a_L\}$  solves

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & c_2^{2/\theta} & \dots & c_L^{2/\theta} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & c_2^{L/\theta} & \dots & c_L^{L/\theta} \end{pmatrix} \times \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_L \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

**Theorem 5:** Suppose Assumptions 1-3 and 5-7 hold. Then as  $n \rightarrow \infty$

$$V_{5,n}^{-1/2} (\bar{\beta}_L - \beta) \rightarrow_d \mathcal{N}(0, I_k)$$

where  $V_{5,n} = \sum_{l_1=1}^L \sum_{l_2=1}^L a_{l_1} a_{l_2} \Gamma_0^{-1} \Omega_{n,l_1 l_2} \Gamma_0^{-1} / n$ ,  $\Gamma_0$  is as defined in Assumption 3,  $r_{nl} = E \left[ K \left( \frac{\delta_{ij}^2}{c_l h_n} \right) \right]$ ,  $I_k$  is the  $k \times k$  identity matrix, and

$$\begin{aligned} \Omega_{n,l_1 l_2} &= \frac{4}{r_{nl_1} r_{nl_2}} E \left[ \Delta_{i_1 j_1} \Delta'_{i_1 j_2} K \left( \frac{\delta_{i_1 j_1}^2}{c_{l_1} h_n} \right) K \left( \frac{\delta_{i_2 j_2}^2}{c_{l_2} h_n} \right) \right] \\ &+ \frac{1}{r_{nl_1} c_{l_1} r_{nl_2} c_{l_2} h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{c_{l_1} h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{c_{l_2} h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_1 s_{21} s_{22}} - \delta_{i_2 j_2}^2) \right] \\ &+ \frac{4}{r_{nl_1} c_{l_1} r_{nl_2} c_{l_2} h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{c_{l_1} h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{c_{l_2} h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_2 s_{11} s_{22}} - \delta_{i_2 j_2}^2) \right] \end{aligned}$$

with  $\Delta_{ij} = (x_i - x_j)'(u_i - u_j)$ ,  $u_i = \lambda(w_i) + \varepsilon_i$ ,  $\delta_{ij} = \delta(w_i, w_j) = \|p_{w_i} - p_{w_j}\|_2$ , and

$$F_{ijts_1 s_2} = f(w_t, w_{s_1}) f(w_t, w_{s_2}) (f(w_i, w_{s_1}) - f(w_j, w_{s_1})) (f(w_i, w_{s_2}) - f(w_j, w_{s_2})).$$

### 3.3.4 Variance Estimation

This section demonstrates that, under the various sets of assumptions provided above, the asymptotic variances of  $\hat{\beta} - \beta_{h_n}$  and  $\bar{\beta}_L - \beta$  can be consistently estimated at least two ways.

The first way is direct computation. Let  $\hat{u}_i = y_i - x_i \hat{\beta}$ ,

$$\hat{\Gamma}_h = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)' (x_i - x_j) K \left( \frac{\hat{\delta}_{ij}^2}{h} \right)$$

and

$$\begin{aligned}\hat{\Omega}_{n,h_1h_2} &= \frac{1}{n^3} \sum_{i=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \hat{\Delta}_{ij_1} \hat{\Delta}'_{ij_2} K \left( \frac{\hat{\delta}_{ij_1}^2}{h_1} \right) K \left( \frac{\hat{\delta}_{ij_2}^2}{h_2} \right) \\ &+ \frac{1}{n^5 h_1 h_2} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \sum_{t=1}^n \hat{\Delta}_{i_1 j_1} \hat{\Delta}_{i_2 j_2} K' \left( \frac{\hat{\delta}_{i_1 j_1}^2}{h_1} \right) K' \left( \frac{\hat{\delta}_{i_2 j_2}^2}{h_2} \right) \left( \hat{F}_{i_1 j_1 t} - \hat{\delta}_{i_1 j_1}^2 \right) \left( \hat{F}_{i_2 j_2 t} - \hat{\delta}_{i_2 j_2}^2 \right) \\ &+ \frac{4}{n^5 h_1 h_2} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \sum_{t=1}^n \hat{\Delta}_{i_1 j_1} \hat{\Delta}_{i_2 j_2} K' \left( \frac{\hat{\delta}_{i_1 j_1}^2}{h_1} \right) K' \left( \frac{\hat{\delta}_{i_2 j_2}^2}{h_2} \right) \left( \hat{F}'_{i_1 j_1 t} - \hat{\delta}_{i_1 j_1}^2 \right) \left( \hat{F}'_{i_2 j_2 t} - \hat{\delta}_{i_2 j_2}^2 \right)\end{aligned}$$

where  $h_l = c_l h_n$ ,  $\hat{\Delta}_{ij} = (x_i - x_j)'(\hat{u}_i - \hat{u}_j)$ ,  $\hat{F}_{ijt} = \frac{1}{n^2} \sum_{s_1=1}^n \sum_{s_2=1}^n D_{ts_1} D_{ts_2} (D_{is_1} - D_{js_1}) (D_{is_2} - D_{js_2})$ , and  $\hat{F}'_{ijs_1} = \frac{1}{n^2} \sum_{t=1}^n \sum_{s_2=1}^n D_{ts_1} D_{ts_2} (D_{is_1} - D_{js_1}) (D_{is_2} - D_{js_2})$

**Theorem 6:** Suppose Assumptions 1-4 hold. Then  $\left( \hat{\Gamma}_{h_n}^{-1} \hat{\Omega}_{n,h_n,h_n} \hat{\Gamma}_{h_n}^{-1} - nV_{4,n} \right) \rightarrow_p 0$  and  $\left( \sum_{l_1=1}^L \sum_{l_2=1}^L a_{l_1} a_{l_2} \hat{\Gamma}_{c_{l_1} h_n}^{-1} \hat{\Omega}_{n,c_{l_1} h_n, c_{l_2} h_n} \hat{\Gamma}_{c_{l_2} h_n}^{-1} - nV_{5,n} \right) \rightarrow_p 0$  as  $n \rightarrow \infty$ .

A corollary to Theorem 6 is that  $\hat{\Gamma}_{h_n}^{-1} \hat{\Omega}_{n,h_n,h_n} \hat{\Gamma}_{h_n}^{-1}$  also consistently estimates  $nV_{3,n}$  under the hypothesis of Theorem 3, although one can omit the last two summands when computing  $\hat{\Omega}_{n,h_n,h_n}$ .

Another way to estimate the asymptotic variances uses the bootstrap. Let  $\{\{i_{tr}\}_{t=1}^n\}_{r=1}^R$  denote a sequence of  $R$  independent samples of agents of size  $n$  drawn from  $\{1, \dots, n\}$  with replacement. With this notation,  $i_{tr}$  denotes the original index of the agent in the  $t$ th index of the  $r$ th sample. Let  $(y_{tr}, x_{tr}, w_{tr})$  denote the outcome, covariates, and social characteristics of agent  $i_{tr}$  and  $D_r = \{D_{str}\}_{s \neq t}$  be the  $n \times n$  adjacency matrix induced by the agents in the  $r$ th sample in which  $D_{str} = D_{i_s r i_{tr}}$ . Let  $\{\hat{\beta}_r\}_{r=1}^R$  and  $\{\bar{\beta}_{Lr}\}_{r=1}^R$  denote the estimators from (4) and (7) constructed using  $\{y_{tr}, x_{tr}\}_{t=1}^n$  and  $D_r$ .

**Theorem 7:** Suppose Assumptions 1-3 and 5-6 hold. Then

$$\frac{1}{R} \sum_{r=1}^R \left( \hat{\beta}_r - \hat{\beta} \right) \left( \hat{\beta}_r - \hat{\beta} \right)' \rightarrow_p V_{4,n} \text{ and } \frac{1}{R} \sum_{r=1}^R \left( \bar{\beta}_{Lr} - \bar{\beta}_L \right) \left( \bar{\beta}_{Lr} - \bar{\beta}_L \right)' \rightarrow_p V_{5,n} \text{ as } n, R \rightarrow \infty.$$

Consistency of the bootstrap variance estimators follows from the fact that under Assumptions 1-3 and 5-6,  $\hat{\beta}$  and  $\bar{\beta}_L$  (and thus  $\hat{\beta}_r$  and  $\bar{\beta}_{Lr}$ ) are asymptotically averages of functions of the iid sequence  $\{y_i, x_i, w_i\}_{i=1}^n$ . Theorem 7 is then a consequence of Theorem 2.2 of [Bickel and Freedman \(1981\)](#).

### 3.4 Large Sample Properties of $\widehat{\lambda}(w_i)$

This section provides two results about the estimators for the social influence term  $\lambda(w_i)$ : consistency and asymptotic normality. The two results mirror those Section 3.3.1 and Section 3.3.2 respectively and so only a limited discussion is provided here.

The first result is that Assumptions 1-4 are sufficient for  $\{\widehat{\lambda}(w_i)\}_{i=1}^n$ , the collection of estimators for the sampled agents to be consistent for their population analogs in the mean-squared sense. This result is stated as Theorem 8.

**Theorem 8:** Suppose Assumptions 1-4 hold. Then  $E \left[ \left( \widehat{\lambda}(w_i) - \lambda(w_i) \right)^2 \right] \rightarrow_p 0$  as  $n \rightarrow \infty$ , where the expectation is taken with respect to  $w_i$ .

Theorem 8 follows almost immediately from Theorem 2 and Lemmas 1 and 2. It can be strengthened to convergence in the sup norm sense under an analogous strengthening of Assumption 3.

The second result is that, under an additional restriction on the choice of bandwidth sequence, these conditions are also sufficient for  $\{\lambda(w_i)\}_{i \in S}$ , the collection of estimators corresponding to a finite (i.e. fixed in  $n$ ) set of agents  $S \subset \{1, \dots, n\}$ , to be asymptotically normal. This additional restriction is given by Assumption 8 and the result is stated as the following Theorem 9.

**Assumption 8:** The bandwidth sequence  $h_n$  satisfies  $nr_{n,i} \rightarrow \infty$  and  $b_{n,i}n/r_{n,i} \rightarrow 0$  where  $r_{n,i} = E \left[ K \left( \frac{\|p_{w_i} - p_{w_j}\|_2}{h_n} \right) | w_i \right]$ ,  $r'_{n,i} = E \left[ \lambda(w_j) K \left( \frac{\|p_{w_i} - p_{w_j}\|_2}{h_n} \right) | w_i \right]$ , and  $b_{n,i} = \left( \lambda(w_i)r_{n,i} - r'_{n,i} \right)^2$  for all  $i \in S$ .

The first condition  $nr_{n,i} \rightarrow \infty$  states that the number of matches to agent  $i$  grows with the sample size, and is analogous to the third bandwidth condition in Assumption 4. The second condition  $b_{n,i}n/r_{n,i} \rightarrow 0$  is an undersmoothing condition that assumes that the bandwidth is chosen to be small enough so that the estimators are asymptotically unbiased. These conditions can be approximated in practice using the empirical analogs of  $r_{n,i}$ ,  $r'_{n,i}$ , and  $\lambda(w_i)$  (see also the discussion of Assumption 4 in Section 3.3.1). The setting of this paper also potentially allows for  $h_n$  to be chosen using a data dependent method such as cross-validation, although I leave the formal study of such an estimator to future work.

**Theorem 9:** Suppose Assumptions 1-4 and 8 hold. Let  $\hat{\lambda}_S = \{\widehat{\lambda(w_i)}\}_{i \in S}$  for some finite collection of agents  $S$ . Then as  $n \rightarrow \infty$

$$V_{8,n}^{-1/2} (\hat{\lambda}_S - \lambda_S) \rightarrow_d \mathcal{N}(0, I_{|S|})$$

where  $\lambda_S = \{\lambda(w_i)\}_{i \in S}$ ,  $I_{|S|}$  is the  $|S| \times |S|$  identity matrix,  $V_{8,n} = \text{diag}(\{V_{8,n,i}\}_{i \in S})$ , and

$$V_{8,n,i} = \sum_{t=1}^n \left( \left( u_t K \left( \frac{\delta_{it}}{h_n} \right) - r'_{n,i} \right) - \frac{r'_{n,i}}{r_{n,i}} \left( K \left( \frac{\delta_{it}}{h_n} \right) - r_{n,i} \right) \right)^2 / (nr_{n,i}^2)$$

One can estimate  $V_{8,n,i}$  directly as in the first part of Section 3.3.4 using the empirical analogs of  $u_t$ ,  $\delta_{it}$ ,  $r_{n,i}$  and  $r'_{n,i}$ , along the lines of Theorem 6, or by using the bootstrap, along the lines of Theorem 7. Consistency of the resulting variance estimators follows from identical arguments, and so is not demonstrated here. One can potentially extend the conclusion of Theorem 9 to allow  $|S|$  to increase with  $n$  using arguments from [Horowitz and Lee \(2016\)](#) or [Chernozhukov, Chetverikov, and Kato \(2017\)](#), although such an extension is not considered in this paper.

## 4 Simulation Evidence

This section presents simulation evidence for three types of network formation models: a stochastic blockmodel, a beta model, and a homophily model. For each of  $R$  simulations, I draw a random sample of  $n$  observations  $\{\xi_i, \varepsilon_i, \omega_i\}_{i=1}^n$  from a trivariate normal distribution with mean 0 and variance-covariance given by the identity matrix, and a random symmetric  $n \times n$  matrix  $\{\eta_{ij}\}_{i,j=1}^n$  with independent and identically distributed upper diagonal entries with standard uniform marginals. For each of the following link functions  $f$ , the adjacency matrix  $D$  is formed by  $D = \mathbb{1}\{\eta_{ij} \leq f(\Phi(\omega_i), \Phi(\omega_j))\}$  where  $\Phi$  is the cumulative distribution function for the standard univariate normal distribution.

The first design draws  $D$  from a stochastic blockmodel where

$$f_1(u, v) = \begin{cases} 1/3 & \text{if } u \leq 1/3 \text{ and } v > 1/3 \\ 1/3 & \text{if } 1/3 < u \leq 2/3 \text{ and } v \leq 2/3 \\ 1/3 & \text{if } u > 2/3 \text{ and } (v > 2/3 \text{ or } v \leq 1/3) \\ 0 & \text{otherwise} \end{cases}$$

The linking function  $f_1$  generates network types with finite support as in the hypothesis of Theorem 3. For this model, I take  $\lambda(\omega_i) = \lceil 3\Phi(\omega_i) \rceil$ ,  $x_i = \xi_i + \lambda(\omega_i)$ , and  $y_i = \beta x_i + \gamma \lambda(\omega_i) + \varepsilon_i$ . The second and third designs draw  $D$  from the beta model and homophily model where

$$f_2(u, v) = \frac{\exp(u + v)}{1 + \exp(u + v)} \text{ and } f_3(u, v) = 1 - (u - v)^2$$

For these models,  $\lambda(\omega_i) = \omega_i$ ,  $x_i = \xi_i + \lambda(\omega_i)$  and  $y_i = \beta x_i + \gamma \lambda(\omega_i) + \varepsilon_i$ .

I use  $x$  and  $y$  to denote the stacked  $n$ -dimensional vector of observations  $\{x_i\}_{i=1}^n$  and  $\{y_i\}_{i=1}^n$ , and  $Z_1$  for the  $(n \times 2)$  matrix  $\{x_i, \lambda(\omega_i)\}_{i=1}^n$ . I also use  $c_i$  to denote a vector of network statistics for agent  $i$  based on  $D$  containing agent degree  $n^{-1} \sum_{j=1}^n D_{ij}$ , eigenvector centrality (the  $i$ th entry of the eigenvector of  $D$  associated with the largest eigenvalue in absolute value), and average peer covariates  $\sum_{j=1}^n D_{ij} x_j / \sum_{j=1}^n D_{ij}$ .  $Z_2$  denotes the stacked vector  $\{x_i, c_i\}_{i=1}^n$ .

For each design, I evaluate the performance of six estimators. The benchmark is  $\hat{\beta}_1 = (Z_1' Z_1)^{-1} (Z_1' y)$ , the infeasible OLS regression of  $y$  on  $x$  and  $\lambda(\omega_i)$ .  $\hat{\beta}_2 = (x' x)^{-1} (x' y)$  is the naïve OLS regression of  $y$  on  $x$ .  $\hat{\beta}_3 = (Z_2' Z_2)^{-1} (Z_2' y)$  is the OLS regression of  $y$  on  $x$  and the vector of network controls  $c$ .  $\hat{\beta}_4$  is the proposed pairwise difference estimator given in (4) without bias correction,  $\hat{\beta}_5$  is the bias corrected estimator (using  $L = 3$ ), and  $\hat{\beta}_6$  is the pairwise difference estimator with an adaptive bandwidth but without bias correction (specifically, the bandwidth depends on  $i$  and is chosen such that each agent is matched to exactly  $n \times h_n$  other agents). The pairwise difference estimators all use the Epanechnikov kernel  $K(u) = 3(1 - u^2) \mathbb{1}\{u^2 < 1\}/4$ . Estimators  $\hat{\beta}_4$  and  $\hat{\beta}_5$  use the bandwidth sequence  $n^{-1/9}/10$  and the estimator  $\hat{\beta}_6$  uses the bandwidth sequence  $n^{-1/9}/5$ . Since  $n^{1/9}$  is roughly equal to 2 for the sample sizes considered in this section, the results are close to a constant



bandwidth choice of  $h_n = .05$  and  $.1$  respectively.

Tables 1-3 demonstrates the results for  $R = 1000$ ,  $\beta = \gamma = 1$  and for each  $n$  in  $\{50, 100, 200, 500, 800\}$ . For each model, estimator and sample size, the first row gives the mean, the second gives the mean absolute error of the simulated estimators around  $\beta$ , the third gives the mean absolute error divided by that of  $\hat{\beta}_1$ , and the fourth gives the proportion of the simulation draws that fall outside of a 0.95 confidence interval based on the asymptotic distributions derived Section 3. The relevant asymptotic variances are approximated directly using Theorem 6.

Table 1 contains results for the stochastic blockmodel. The naïve estimator  $\hat{\beta}_2$  has a large and stable positive bias that is not reduced as  $n$  is increased. The OLS estimator with network controls  $\hat{\beta}_3$  is not asymptotically well defined in this example because the network statistics converge to constants. The results in Table 1 instead demonstrate a common “fix” in the literature, which is to instead calculate  $(Z_2'Z_2)^+(Z_2'y)$  where  $+$  refers to the Moore-Penrose pseudo-inverse. The results for this estimator indicate that adding network controls mitigates some of the bias in  $\beta_1$  (due to sampling variation in the number of agents in each block), however the estimator is otherwise poorly behaved. Notice this bias returns when the block sizes stabilize (in particular when  $n = 800$ ).

The results for the pairwise difference estimators illustrate the content of Theorem 3, that when the unobserved heterogeneity is discrete, the proposed estimator identifies pairs of agents of the same type with high probability. As a result, the pairwise difference estimators  $\hat{\beta}_4$  and  $\hat{\beta}_6$  behave similarly to the infeasible  $\hat{\beta}_2$ . For the stochastic blockmodel, Assumption 9 is not valid, and so the jackknife bias correction actually inflates both the bias and variance of  $\hat{\beta}_4$ . Looking at the relative mean absolute error for this estimator, it is clear that the relative performance of the error is deteriorating as  $n$  increases (though the bias and variance of this estimator is still on the order of  $1/\sqrt{n}$ ).

Table 2 contains results for the beta model. Relative to the stochastic blockmodel, all of the estimators for the beta model (except infeasible OLS) have large biases. This is because the link function  $f_2$  is very flat, so that the variation in linking probabilities that identifies the network positions is relatively small (Johnsson and Moon 2015, describe a similar phenomenon in their simulations). One can show that the social characteristics are

Table 1: Simulation Results, Stochastic Blockmodel

n		Infeasible OLS $\hat{\beta}_1$	Naïve OLS $\hat{\beta}_2$	OLS with Controls $\hat{\beta}_3$	Pairwise Difference $\hat{\beta}_4$	Bias Corrected $\hat{\beta}_5$	Adaptive Bandwidth $\hat{\beta}_6$
50	bias	0.004	0.829	0.268	0.060	0.022	0.106
	MAE	0.116	0.829	0.274	0.224	0.240	0.150
	rMAE	1.000	7.147	2.362	1.931	2.069	1.293
	size	0.057	0.063	0.072	0.115	0.123	0.067
100	bias	0.003	0.829	0.226	0.021	-0.022	0.019
	MAE	0.083	0.829	0.229	0.089	0.094	0.084
	rMAE	1.000	9.988	2.759	1.072	1.133	1.012
	size	0.064	0.053	0.108	0.053	0.058	0.056
200	bias	0.001	0.823	0.180	0.004	-0.040	0.002
	MAE	0.056	0.823	0.183	0.058	0.069	0.058
	rMAE	1.000	14.696	3.268	1.036	1.232	1.036
	size	0.049	0.044	0.215	0.045	0.064	0.058
500	bias	0.000	0.824	0.172	0.006	0.038	0.001
	MAE	0.035	0.824	0.174	0.035	0.048	0.035
	rMAE	1.000	23.543	4.971	1.000	1.371	1.000
	size	0.033	0.061	0.777	0.037	0.047	0.044
800	bias	0.001	0.823	0.314	0.008	-0.036	0.000
	MAE	0.029	0.823	0.314	0.029	0.043	0.029
	rMAE	1.000	28.379	10.828	1.000	1.483	1.000
	size	0.057	0.038	0.127	0.054	0.068	0.062

Table 1: This table contains simulation results for 1000 replications and a sample size of  $n = 50, 100, 200, 500, 800$ . Bias gives the mean estimator minus 1. MAE gives the mean absolute error of the estimator around 1. rMAE gives the mean absolute error relative to the benchmark  $\hat{\beta}_1$ . Size gives the proportion of draws that fall outside the asymptotic 0.95 confidence interval.

Table 2: Simulation Results, Beta Model

n		Infeasible OLS $\hat{\beta}_1$	Naïve OLS $\hat{\beta}_2$	OLS with Controls $\hat{\beta}_3$	Pairwise Difference $\hat{\beta}_4$	Bias Corrected $\hat{\beta}_5$	Adaptive Bandwidth $\hat{\beta}_6$
50	bias	0.000	0.496	0.462	0.379	0.335	0.365
	MAE	0.119	0.496	0.463	0.381	0.341	0.366
	rMAE	1.000	4.168	3.891	3.202	2.866	3.076
	size	0.064	0.063	0.075	0.049	0.066	0.070
100	bias	0.006	0.501	0.462	0.336	0.269	0.298
	MAE	0.082	0.501	0.462	0.336	0.270	0.299
	rMAE	1.000	6.110	5.634	4.098	3.293	3.646
	size	0.055	0.053	0.055	0.039	0.062	0.081
200	bias	0.002	0.501	0.444	0.290	0.200	0.231
	MAE	0.058	0.501	0.444	0.290	0.200	0.231
	rMAE	1.000	8.638	7.655	5.000	3.448	3.983
	size	0.050	0.041	0.036	0.033	0.054	0.070
500	bias	0.003	0.499	0.403	0.246	0.136	0.151
	MAE	0.036	0.499	0.403	0.246	0.136	0.151
	rMAE	1.000	13.861	11.194	6.833	3.778	4.194
	size	0.049	0.042	0.054	0.022	0.033	0.076
800	bias	0.000	0.500	0.385	0.237	0.122	0.122
	MAE	0.028	0.500	0.385	0.237	0.122	0.122
	rMAE	1.000	17.857	13.750	8.464	4.357	4.357
	size	0.050	0.054	0.078	0.037	0.050	0.062

Table 2: This table contains simulation results for 1000 replications and a sample size of  $n = 50, 100, 200, 500, 800$ . Bias gives the mean estimator minus 1. MAE gives the mean absolute error of the estimator around 1. rMAE gives the mean absolute error relative to the benchmark  $\hat{\beta}_1$ . Size gives the proportion of draws that fall outside the asymptotic 0.95 confidence interval.

identified by the distribution of  $D$  (they are consistently estimated by the order statistics of the degree distribution), but the bound on the deviation of the social characteristics given by the network metric is large:  $|u - v| \leq 20 \times d(u, v)$ .

Still, the proposed pairwise difference estimator offers a substantial improvement in performance relative to both the naïve estimator  $\hat{\beta}_2$  and the estimator with network controls (which includes agent degree)  $\hat{\beta}_3$ . For example, when  $n = 100$ ,  $\hat{\beta}_5$  has approximately half the bias and mean absolute error of  $\hat{\beta}_1$  while  $\hat{\beta}_3$  offers a reduction of less than ten percent. When  $n = 800$  the reduction in bias is over three times as large (75% relative to 23%).

Table 3 contains results for the homophily model. As in the case of the beta model, one can show that the social characteristics are also identified in the homophily model (the social characteristics are uniquely identified by the agents' codegrees with any two randomly drawn agents from the sample). Unlike the beta model, there is a relatively large amount of information about the network positions in the linking probabilities so that all of the estimators in Table 3 are much better behaved. In fact, for this model  $|u - v| \leq d(u, v)$ .

In this example, the OLS estimator with network controls actually performs comparably to the uncorrected pairwise difference estimator  $\hat{\beta}_4$ . This is because the peer characteristics variable  $\sum_{j=1}^n D_{ij}x_j / \sum_{j=1}^n D_{ij}$  is a good approximation of  $w_i$  when  $n$  is large. However, the bias corrected estimator  $\hat{\beta}_5$  outperforms both estimators over all of the sample sizes considered.

## 5 Directions for Future Work

I highlight two directions for future work. The first direction is to relax the partially linear structure of the regression model (1). With a little work, I suspect that the main ideas of this paper also apply to nonlinear regression models in which other applications of the pairwise differencing logic in (4) has been effective (see [Honoré and Powell 1994; 1997](#), [Hong and Shum 2010](#), [Aradillas-Lopez 2012](#)), so long as the unobserved heterogeneity in the regression model is continuous with respect to network distance in the sense of Assumption 2. For example, one might consider the partially linear Logit model  $y_i = \mathbb{1}\{x\beta + \lambda(w_i) - \varepsilon_i > 0\}$  where the errors  $\{\varepsilon_i\}_{i=1}^n$  are iid logistically distributed random variables and, along the lines

Table 3: Simulation Results, Homophily Model

n		Infeasible OLS $\hat{\beta}_1$	Naïve OLS $\hat{\beta}_2$	OLS with Controls $\hat{\beta}_3$	Pairwise Difference $\hat{\beta}_4$	Bias Corrected $\hat{\beta}_5$	Adaptive Bandwidth $\hat{\beta}_6$
50	bias	0.007	0.505	0.269	0.128	0.087	0.140
	MAE	0.120	0.505	0.274	0.108	0.121	0.211
	rMAE	1.000	4.208	2.283	0.900	1.008	1.758
	size	0.068	0.051	0.063	0.062	0.068	0.132
100	bias	0.005	0.502	0.162	0.100	0.057	0.089
	MAE	0.081	0.502	0.167	0.124	0.108	0.116
	rMAE	1.000	6.198	2.062	1.531	1.333	1.432
	size	0.049	0.059	0.061	0.053	0.066	0.083
200	bias	0.001	0.503	0.095	0.085	0.039	0.055
	MAE	0.057	0.503	0.100	0.097	0.075	0.077
	rMAE	1.000	8.825	1.754	1.702	1.316	1.351
	size	0.054	0.059	0.054	0.050	0.057	0.069
500	bias	0.000	0.501	0.047	0.074	0.028	0.035
	MAE	0.035	0.501	0.053	0.077	0.048	0.046
	rMAE	1.000	14.314	1.514	2.200	1.371	1.314
	size	0.043	0.059	0.039	0.045	0.058	0.051
800	bias	0.000	0.501	0.034	0.070	0.023	0.030
	MAE	0.028	0.501	0.086	0.072	0.039	0.038
	rMAE	1.000	17.893	3.071	2.571	1.392	1.357
	size	0.039	0.040	0.041	0.038	0.050	0.047

Table 3: This table contains simulation results for 1000 replications and a sample size of  $n = 50, 100, 200, 500, 800$ . Bias gives the mean estimator minus 1. MAE gives the mean absolute error of the estimator around 1. rMAE gives the mean absolute error relative to the benchmark  $\hat{\beta}_1$ . Size gives the proportion of draws that fall outside the asymptotic 0.95 confidence interval.

of [Honoré and Powell \(1997\)](#), estimate  $\beta$  by maximizing

$$-\sum_{i \neq j} K\left(\frac{\hat{\delta}_{ij}}{h_n}\right) (y_i \ln(1 + \exp((x_j - x_i)b)) + y_j \ln(1 + \exp((x_i - x_j)b)))$$

over  $b \in \mathbb{R}^k$ . This model could be used to study binary outcomes as in the program participation application in Example 2 of Section 2.1.

One might also consider the nonparametric regression model  $y_i = m(x_i, w_i) + \varepsilon_i$ , and estimate features of  $m$  along the lines of Theorems 8 and 9, by local averaging in the sense of [Nadaraya \(1965\)](#) and [Watson \(1964\)](#), using the empirical codegree distance. Nonparametric predictions of  $y_i$  may be useful to the literature on contagion, in which one object of interest is the conditional probability that agent  $i$  becomes infected conditional on that agent's observed characteristics and his social or economic connections to other agents.

The second direction for future work is to extend the network formation model (2) as to allow for more general network structures. With a little work, I suspect that the main ideas of this paper also apply to directed and weighted networks, where  $E[D_{ij}|w_i, w_j] = f(w_i, w_j)$ , so long as  $f$  is square integrable. If linking behavior is still conditionally independent, the main arguments in Section 2.3 are still valid and analogs of Lemmas 1 and 2 still hold.

One can also incorporate exogenous covariates into the right-hand side of the network formation model, by redefining the agent network types. That is, one can consider the model  $D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j, z_{ij})\}$ , where  $\{z_{ij}\}_{i \neq j}$  is observed data. If  $z_{ij}$  has finite support, then one can assign each agent a network type for each element of that support, i.e.  $\{f(w_i, \cdot, z)\}_{z \in \text{supp}(z_{ij})}$ . Extending this logic to the case where  $z_{ij}$  has continuous support is, to my knowledge, nontrivial, and a topic I plan to explore in more detail in future work. As mentioned in the discussion of Example 1 in Section 2.2, one potential use of exogenous link covariates is to identify network peer effects in the presence of unobserved social influence.

## References

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

- Abadie, A. and G. W. Imbens (2012). Bias-corrected matching estimators for average treatment effects. *Journal of Business & Economic Statistics*.
- Ahn, H. (1997). Semiparametric estimation of a single-index model with nonparametrically generated regressors. *Econometric Theory* 13(01), 3–31.
- 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.
- Akerlof, G. A. and R. E. Kranton (2002). Identity and schooling: Some lessons for the economics of education. *Journal of Economic Literature* 40(4), 1167–1201.
- Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis* 11(4), 581–598.
- Aradillas-Lopez, A. (2012). Pairwise-difference estimation of incomplete information games. *Journal of econometrics* 168(1), 120–140.
- Arduini, T., E. Patacchini, and E. Rainone (2015). Parametric and semiparametric iv estimation of network models with selectivity. Technical report, Einaudi Institute for Economics and Finance (EIEF).
- Austen-Smith, D. and R. G. Fryer Jr (2005). An economic analysis of acting white. *The Quarterly Journal of Economics* 120(2), 551–583.
- Badev, A. (2017). Discrete games in endogenous networks: Equilibria and policy. *arXiv preprint arXiv:1705.03137*.
- Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science* 341(6144), 1236498.
- Bhattacharyya, S. and P. J. Bickel (2015). Subsampling bootstrap of count features of networks. *The Annals of Statistics* 43(6), 2384–2411.
- Bickel, P., D. Choi, X. Chang, and H. Zhang (2013). Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels. *The Annals of Statistics* 41(4), 1922–1943.

- Bickel, P. J. and A. Chen (2009). A nonparametric view of network models and newman–girvan and other modularities. *Proceedings of the National Academy of Sciences* 106(50), 21068–21073.
- Bickel, P. J., A. Chen, and E. Levina (2011). The method of moments and degree distributions for network models. *The Annals of Statistics* 39(5), 2280–2301.
- Bickel, P. J. and D. A. Freedman (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics*, 1196–1217.
- Blume, L. E., W. A. Brock, S. N. Durlauf, and Y. M. Ioannides (2010). Identification of social interactions. *Available at SSRN 1660002*.
- Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica* 83(3), 1147–1184.
- Boucher, V. and B. Fortin (2015). Some challenges in the empirics of the effects of networks. *Cahier de recherche/Working Paper 15*, 04.
- Bramoullé, Y., H. Djebbari, and B. Fortin (2009). Identification of peer effects through social networks. *Journal of econometrics* 150(1), 41–55.
- Breza, E., A. G. Chandrasekhar, T. H. McCormick, and M. Pan (2017). Using aggregated relational data to feasibly identify network structure without network data. Technical report, National Bureau of Economic Research.
- Candelaria, L. E. (2016). A semiparametric network formation model with multiple linear fixed effects. *Duke University*.
- Chamberlain, G. (1986). Notes on semiparametric regression. *manuscript, Department of Economics, University of Wisconsin-Madison*.
- Chan, J. (2014). An additive model with endogenous network formation.
- Chandrasekhar, A. (2015). Econometrics of network formation. Technical report, working paper, 2014.



- Chandrasekhar, A. and R. Lewis (2011). Econometrics of sampled networks. *Unpublished manuscript, MIT.[422]*.
- Chandrasekhar, A. G. and M. O. Jackson (2014). Tractable and consistent random graph models.
- Chatterjee, S., P. Diaconis, and A. Sly (2011). Random graphs with a given degree sequence. *The Annals of Applied Probability*, 1400–1435.
- Chernozhukov, V., D. Chetverikov, and K. Kato (2017). Central limit theorems and bootstrap in high dimensions. *The Annals of Probability* 45(4), 2309–2352.
- de Paula, A. (2016). Econometrics of network models. Technical report, cemmap working paper, Centre for Microdata Methods and Practice.
- Ductor, L., M. Fafchamps, S. Goyal, and M. J. van der Leij (2014). Social networks and research output. *Review of Economics and Statistics* 96(5), 936–948.
- Dudley, R. M. (2002). *Real analysis and probability*, Volume 74. Cambridge University Press.
- Duijn, M. A., T. A. Snijders, and B. J. Zijlstra (2004). p2: a random effects model with covariates for directed graphs. *Statistica Neerlandica* 58(2), 234–254.
- Dzemeski, A. (2014). An empirical model of dyadic link formation in a network with unobserved heterogeneity. Technical report, University of Manheim Working Paper.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 7(1), 1–26.
- Engle, R. F., C. W. Granger, J. Rice, and A. Weiss (1986). Semiparametric estimates of the relation between weather and electricity sales. *Journal of the American statistical Association* 81(394), 310–320.
- Ferraty, F., A. Mas, and P. Vieu (2007). Nonparametric regression on functional data: inference and practical aspects. *Australian & New Zealand Journal of Statistics* 49(3), 267–286.

- 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. (2008). Identifying social interactions through conditional variance restrictions. *Econometrica* 76(3), 643–660.
- Graham, B. S. (2015). Methods of identification in social networks. *Annu. Rev. Econ.* 7(1), 465–485.
- Graham, B. S. (2017). An econometric model of network formation with degree heterogeneity. *Econometrica* 85(4), 1033–1063.
- Graham, B. S. and J. Hahn (2005). Identification and estimation of the linear-in-means model of social interactions. *Economics Letters* 88(1), 1–6.
- Griffith, A. (2016). Random assignment with non-random peers: A structural approach to counterfactual treatment assessment.
- Hahn, J. and H. R. Moon (2010). Panel data models with finite number of multiple equilibria. *Econometric Theory* 26(03), 863–881.
- Hahn, J. and G. Ridder (2013). Asymptotic variance of semiparametric estimators with generated regressors. *Econometrica* 81(1), 315–340.
- Hall, P. (1984). Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function. *Probability Theory and Related Fields* 67(2), 175–196.
- Hardle, W. and J. S. Marron (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, 1465–1481.
- Holland, P. W., K. B. Laskey, and S. Leinhardt (1983). Stochastic blockmodels: First steps. *Social networks* 5(2), 109–137.
- Holland, P. W. and S. Leinhardt (1981). An exponential family of probability distributions for directed graphs. *Journal of the American Statistical Association* 76(373), 33–50.

- Hong, H. and M. Shum (2010). Pairwise-difference estimation of a dynamic optimization model. *The Review of Economic Studies* 77(1), 273–304.
- Hong, S. Y. and O. B. Linton (2016). Asymptotic properties of a nadaraya-watson type estimator for regression functions of infinite order. *Available at SSRN 2766822*.
- Honoré, B. E. and J. Powell (1997). *Pairwise difference estimators for nonlinear models*. na.
- Honoré, B. E. and J. L. Powell (1994). Pairwise difference estimators of censored and truncated regression models. *Journal of Econometrics* 64(1-2), 241–278.
- Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Preprint, Institute for Advanced Study, Princeton, NJ 2*.
- Horowitz, J. L. and S. Lee (2016). Nonparametric estimation and inference under shape restrictions. Technical report, cemmap working paper, Centre for Microdata Methods and Practice.
- Hsieh, C.-S. and L. F. Lee (2014). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*.
- Jackson, M. (2008). *Social and economic networks*, Volume 3. Princeton university press Princeton.
- Jackson, M. O. (2014). Networks in the understanding of economic behaviors. *The Journal of Economic Perspectives* 28(4), 3–22.
- Johnsson, I. and H. R. Moon (2015). Estimation of peer effects in endogenous social networks: Control function approach.
- Kallenberg, O. (1989). On the representation theorem for exchangeable arrays. *Journal of Multivariate Analysis* 30(1), 137–154.
- Kranton, R. E. (2017). Networks in economics: Remarks. In *Advances in Economics and Econometrics: Volume 1: Eleventh World Congress*, Volume 58, pp. 324. Cambridge University Press.

- Krivitsky, P. N., M. S. Handcock, A. E. Raftery, and P. D. Hoff (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks* 31(3), 204–213.
- Leung, M. (2014). A random-field approach to inference in large models of network formation. *Available at SSRN*.
- Leung, M. P. (2015). Two-step estimation of network-formation models with incomplete information. *Journal of Econometrics* 188(1), 182–195.
- Lovász, L. (2012). *Large networks and graph limits*, Volume 60. American Mathematical Soc.
- Lovász, L. and B. Szegedy (2010). Regularity partitions and the topology of graphons. In *An irregular mind*, pp. 415–446. Springer.
- Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies* 60(3), 531–542.
- Marianne, B. (2011). New perspectives on gender. *Handbook of labor economics* 4, 1543–1590.
- McCormick, T. H. and T. Zheng (2012). Latent demographic profile estimation in hard-to-reach groups. *The annals of applied statistics* 6(4), 1795.
- Mele, A. (2017). A structural model of dense network formation. *Econometrica* 85(3), 825–850.
- Mele, A. and L. Zhu (2017). Approximate variational estimation for a model of network formation.
- Menzel, K. (2015). Strategic network formation with many agents.
- Menzel, K. (2017). Bootstrap with clustering in two or more dimensions. *arXiv preprint arXiv:1703.03043*.

- Nadaraya, E. (1965). On non-parametric estimates of density functions and regression curves. *Theory of Probability & Its Applications* 10(1), 186–190.
- Nadler, C. (2016). Networked inequality: Evidence from freelancers.
- Newey, W. K. (1988). Adaptive estimation of regression models via moment restrictions. *Journal of Econometrics* 38(3), 301–339.
- Nolan, D. and D. Pollard (1987). U-processes: rates of convergence. *The Annals of Statistics*, 780–799.
- Nolan, D. and D. Pollard (1988). Functional limit theorems for u-processes. *The Annals of Probability*, 1291–1298.
- Orbanz, P. and D. M. Roy (2015). Bayesian models of graphs, arrays and other exchangeable random structures. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 37(2), 437–461.
- Powell, J. L. (1987). *Semiparametric estimation of bivariate latent variable models*. University of Wisconsin–Madison, Social Systems Research Institute.
- Powell, J. L., J. H. Stock, and T. M. Stoker (1989). Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, 1403–1430.
- Ridder, G. and S. Sheng (2015). Estimation of large network formation games.
- Ritov, Y. and P. J. Bickel (1990). Achieving information bounds in non and semiparametric models. *The Annals of Statistics*, 925–938.
- Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica* 56(4), 931–54.
- Rohe, K., S. Chatterjee, and B. Yu (2011). Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics* 39(4), 1878–1915.
- Sacerdote, B. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far. *Handbook of the Economics of Education* 3(3), 249–277.

- Schmutte, I. M. (2014). Free to move? a network analytic approach for learning the limits to job mobility. *Labour Economics* 29, 49–61.
- Serfling, R. J. (2009). *Approximation theorems of mathematical statistics*, Volume 162. John Wiley & Sons.
- Sheng, S. (2012). Identification and estimation of network formation games.
- Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *The Annals of Statistics*, 1285–1297.
- Toth, P. (2017). Semiparametric estimation in networks with homophily and degree heterogeneity.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 359–372.
- Zhang, Y., E. Levina, and J. Zhu (2015). Estimating network edge probabilities by neighborhood smoothing. *arXiv preprint arXiv:1509.08588*.

## A Proofs of Lemmas and Theorems

This section contains proofs of the various Lemmas and Theorems from Section 3. Auxiliary lemmas not formally stated in the paper are labelled Lemma A1, Lemma A2, et cetera.

### A.1 Lemmas and Theorems from Section 2.3

**Theorem B:** Suppose  $\lambda(D, i)$  satisfies the symmetry and bounded deviations assumptions, and that (6) holds. Then

$$|\lambda(D, i) - \lambda(D, j)| \leq C \|f(w, w_i, \cdot) - f(w, w_j, \cdot)\|_2 + O_p(m^{-1/2})$$

for some  $C$  depending on  $f$  and  $w$ .

**Proof of Theorem B** First write

$$\begin{aligned}\lambda(D, i) - \lambda(D, j) &= (\lambda(D, i) - \lambda(D, i - j)) + (\lambda(D, i - j) - \lambda(D, j - i)) \\ &\quad + (\lambda(D, j - i) - \lambda(D, j))\end{aligned}$$

where  $\lambda(D, i - j) = \lambda(D'', i)$  for  $D''_{st} = D_{st} \mathbb{1}\{s, t \neq j\}$ . Intuitively,  $\lambda(D, i - j)$  is the network statistic based on network  $D$  and agent  $i$  with all of agent  $j$ 's links removed.

The bounded deviations condition implies

$$(\lambda(D, i) - \lambda(D, i - j)) + (\lambda(D, j - i) - \lambda(D, j)) = O_p(m^{-1})$$

since agent  $j$  has at most one connection to agent  $i$  and  $m - 2$  connections to other agents.

It remains to be shown that

$$(\lambda(D, i - j) - \lambda(D, j - i)) \leq C \|f_{w_i} - f_{w_j}\|_2 + O_p(m^{-1/2})$$

where  $f_{w_i}$  is shorthand for the reduced form network type  $f(w, w_i, \cdot)$ .

To see this, write

$$\begin{aligned}(\lambda(D, i - j) - \lambda(D, j - i)) &= \sum_{\tau \neq i, j} (\lambda(D, i - j, \tau, 1) - \lambda(D, i - j, \tau, 0)) (D_{i\tau} - D_{j\tau}) \\ &= \sum_{\tau \neq i, j} (\lambda(D, i - j, \tau, 1) - \lambda(D, i - j, \tau, 0)) [(D_{i\tau} - f_{i\tau}) + (f_{j\tau} - D_{j\tau}) + (f_{i\tau} - f_{j\tau})]\end{aligned}$$

where  $f_{i\tau} = f(w, w_i, w_\tau)$ ,  $\lambda(D, i - j, \tau, p) = \lambda(D'''(p), i)$  and

$$\begin{aligned}D'''_{st}(p) &= D_{st} \mathbb{1}\{s, t \neq i, j\} + D_{it} \mathbb{1}\{i = s \text{ and } t > \tau\} \\ &\quad + D_{jt} \mathbb{1}\{i = s \text{ and } t < \tau\} + p \mathbb{1}\{i = s \text{ and } t = \tau\}\end{aligned}$$

Intuitively,  $\lambda(D, i - j, \tau, p)$  is the network statistic on network  $D$  for agent  $i$  with the link

between  $i$  and  $\tau$  replaced with  $p$  and if  $t < \tau$  the link between  $i$  and  $t$  is replaced with  $D_{jt}$ .

I consider the three summands in the square brackets of the second line separately. The second summand

$$\sum_{\tau \neq i, j} (\lambda(D, i - j, \tau, 1) - \lambda(D, i - j, \tau, 0)) (f_{j\tau} - D_{j\tau})$$

is a martingale with respect to the filtration

$\mathcal{F}_\tau = \sigma(w, \{D_{st} : s, t \neq j \text{ or } s = j, t \leq \tau\} \cup \{w_t : t \in \mathbb{N}\})$ , and so the summand is

$O_p(m^{-1/2})$  by Azuma's inequality since

$|(\lambda(D, i - j, \tau, 1) - \lambda(D, i - j, \tau, 0)) (f_{j\tau} - D_{j\tau})| = O_p(m^{-1})$  by the bounded deviations condition.

The first summand can be rewritten as

$$\sum_{\tau' \neq i', j'} (\lambda(D, j' - i', \tau', 0) - \lambda(D, j' - i', \tau', 1)) (D_{i'\tau'} - f_{i'\tau'})$$

where  $\tau' = m + 1 - \tau$ ,  $i' = m + 1 - i$ , and  $j' = m + 1 - j$ . This sum is also a martingale with respect to the filtration  $\mathcal{F}_{\tau'} = \sigma(w, \{D_{st} : s, t \neq i' \text{ or } s = i', t \leq \tau'\} \cup \{w_t : t \in \mathbb{N}\})$  and so is also  $O_p(m^{-1/2})$  by Azuma's inequality.

Altogether,

$$\begin{aligned} \lambda(D, i) - \lambda(D, j) &= \sum_{\tau \neq i, j} (\lambda(D, i - j, \tau, 1) - \lambda(D, i - j, \tau, 0)) (f_{i\tau} - f_{j\tau}) + O_p(m^{-1/2}) \\ &\leq \sqrt{\sum_{\tau \neq i, j} (\lambda(D, i - j, \tau, 1) - \lambda(D, i - j, \tau, 0))^2 \times \sum_{\tau \neq i, j} (f_{i\tau} - f_{j\tau})^2} + O_p(m^{-1/2}) \\ &\leq C \times \sqrt{\frac{1}{m} \sum_{\tau \neq i, j} (f_{i\tau} - f_{j\tau})^2} + O_p(m^{-1/2}) \end{aligned}$$

where  $C = 2 \times \text{plim}_{m \rightarrow \infty} \sqrt{m \sum_{\tau \neq i, j} (\lambda(D, i - j, \tau, 1) - \lambda(D, i - j, \tau, 0))^2} < \infty$ , the first



inequality due to Cauchy-Schwarz, and the second due to the bounded deviations condition. The final term is equal to  $C\|f_{w_i} - f_{w_j}\|_2 + O_p(m^{-1/2})$  by Markov's inequality, which completes the proof.  $\square$

## A.2 Lemmas and Theorems in Section 3.2

**Theorem 1:** Suppose Assumptions 1-3 hold. Then  $\beta$  is the unique minimizer of  $E [((y_i - y_j) - (x_i - x_j)b)^2 | \|f_{w_i} - f_{w_j}\|_2 = 0]$  over  $b \in \mathbb{R}^k$  and  $\lambda(w_i) = E [(y_i - x_i\beta) | f_{w_i}]$ .

**Proof of Theorem 1:** Let  $d_{ij}$  shorthand  $\|f_{w_i} - f_{w_j}\|_2$  and  $u_i = y_i - x_i\beta$ . The second claim follows from  $E [\varepsilon | x_i, w_i] = 0$  and Assumption 3 since

$$E [u_i | f_{w_i}] = E [\lambda(w_i) | f_{w_i}] + E [E [\varepsilon_i | x_i, w_i] | f_{w_i}] = \lambda(w_i)$$

The first claim follows from

$$\begin{aligned} E [((y_i - y_j) - (x_i - x_j)b)^2 | d_{ij} = 0] &= E [((x_i - x_j)(\beta - b) + (u_i - u_j))^2 | d_{ij} = 0] \\ &= (\beta - b)' E [(x_i - x_j)'(x_i - x_j) | d_{ij} = 0] (\beta - b) + E [(u_i - u_j)^2 | d_{ij} = 0] \\ &\quad - 2(\beta - b)' E [(x_i - x_j)'(u_i - u_j) | d_{ij} = 0] \end{aligned}$$

in which first summand is uniquely minimized at  $b = \beta$  by Assumption 2, the second summand does not depend on  $b$ , and the third summand is equal to 0 by Assumption 3 (since  $E [\varepsilon | x_i, w_i] = 0$ ).

## A.3 Lemmas and Theorems in Section 3.3.1

**Lemma 1:** Suppose Assumptions 1 and 4 hold. Then

$$\max_{(i \neq j)} \left| \hat{\delta}_{ij}^2 - \|p_{w_i} - p_{w_j}\|_2^2 \right| = o_p(n^{-\gamma/4} h_n)$$

**Proof of Lemma 1:** Let  $h'_n = n^{-\gamma/4} h_n$ ,  $p_{w_i w_j} = \int f_{w_i}(\tau) f_{w_j}(\tau) d\tau$ ,  $\hat{p}_{w_i w_j} = (n-2)^{-1} \sum_{t \neq i, j} D_{it} D_{jt}$ ,  $\|\hat{p}_{w_i} - p_{w_i}\|_{2, n, j}^2 = (n-2)^{-1} \sum_{s \neq i, j} (\hat{p}_{w_i w_s} - p_{w_i w_s})^2$ , and

$\|p_{w_i} - p_{w_j}\|_{2,n}^2 = (n-2)^{-1} \sum_{s \neq i,j} (p_{w_i w_s} - p_{w_j w_s})^2$ . Then for any fixed  $\epsilon > 0$ ,

$$\begin{aligned}
& P \left( \max_{i \neq j} h'_n{}^{-1} \left| \hat{\delta}_{ij}^2 - \|p_{w_i} - p_{w_j}\|_2^2 \right| > \epsilon \right) \\
&= P \left( \max_{i \neq j} h'_n{}^{-1} \left| \hat{\delta}_{ij}^2 - \|p_{w_i} - p_{w_j}\|_{2,n}^2 + \|p_{w_i} - p_{w_j}\|_{2,n}^2 - \|p_{w_i} - p_{w_j}\|_2^2 \right| > \epsilon \right) \\
&\leq P \left( \max_{i \neq j} h'_n{}^{-1} \left| \hat{\delta}_{ij}^2 - \|p_{w_i} - p_{w_j}\|_{2,n}^2 \right| > \epsilon/2 \right) \\
&\quad + P \left( \max_{i \neq j} h'_n{}^{-1} \left| \|p_{w_i} - p_{w_j}\|_{2,n}^2 - \|p_{w_i} - p_{w_j}\|_2^2 \right| > \epsilon/2 \right) \\
&= P \left( \max_{i \neq j} h'_n{}^{-1} \left| \hat{\delta}_{ij} - \|p_{w_i} - p_{w_j}\|_{2,n} \right| > \epsilon/2 \right) \\
&\leq P \left( \max_{i \neq j} h'_n{}^{-1} \left| (n-2)^{-1} \sum_{s \neq i,j} ((\hat{p}_{w_i w_s} - \hat{p}_{w_j w_s}) - (p_{w_i w_s} - p_{w_j w_s})) \right| > \epsilon/8 \right) \\
&\leq 2P \left( \max_{i \neq j} h'_n{}^{-1} (n-2)^{-1} \sum_{s \neq i,j} |\hat{p}_{w_i w_s} - p_{w_i w_s}| > \epsilon/16 \right) = o(1)
\end{aligned}$$

in which  $P \left( \max_{i \neq j} h'_n{}^{-1} \left| \|p_{w_i} - p_{w_j}\|_{2,n}^2 - \|p_{w_i} - p_{w_j}\|_2^2 \right| > \epsilon/2 \right) = o(1)$  in the second equality and  $P \left( \max_{i \neq j} h'_n{}^{-1} (n-2)^{-1} \sum_{s \neq i,j} |\hat{p}_{w_i w_s} - p_{w_i w_s}| > \epsilon/16 \right) = o(1)$  in the final equality are demonstrated below, the first inequality is due to the triangle inequality, the second inequality is due to the fact that  $|p_{w_i} + \hat{p}_{w_i}| \leq 2$  for every  $w_i \in [0, 1]$ , and the final inequality is due to the triangle and Jensen's inequality.

The second result, that  $P \left( \max_{i \neq j} h'_n{}^{-1} (n-2)^{-1} \sum_{s \neq i,j} |\hat{p}_{w_i w_s} - p_{w_i w_s}| > \epsilon/16 \right) = o(1)$  follows from the fact that  $\max_{i \neq j} h'_n{}^{-1} |\hat{p}_{w_i w_j} - p_{w_i w_j}| \rightarrow_p 0$  by Bernstein's inequality and the union bound. Specifically, the former implies that for any  $\epsilon > 0$

$$\begin{aligned}
P \left( h'_n{}^{-1} |\hat{p}_{w_i w_j} - p_{w_i w_j}| > \epsilon \right) &= P \left( h'_n{}^{-1} \left| (n-2)^{-1} \sum_{t \neq i,j} (D_{it} D_{jt} - p_{w_i w_j}) \right| > \epsilon \right) \\
&\leq 2 \exp \left( \frac{-(n-2)(h'_n \epsilon)^2}{2 + 2h'_n \epsilon/3} \right)
\end{aligned}$$

and the latter gives

$$P\left(\max_{i \neq j} h_n'^{-1} |\hat{p}_{w_i w_j} - p_{w_i w_j}| > \epsilon\right) \leq 2n(n-1) \exp\left(\frac{-(n-2)(h_n' \epsilon)^2}{2 + 2h_n' \epsilon/3}\right)$$

which is  $o(1)$  since  $n^{1-\gamma} h_n'^2 \rightarrow \infty$  for some  $\gamma > 0$ .

The first result, that  $P(\max_{i \neq j} h_n'^{-1} \left| \|p_{w_i} - p_{w_j}\|_{2,n}^2 - \|p_{w_i} - p_{w_j}\|_2^2 \right| > \epsilon/2) = o(1)$ , also follows from Bernstein's inequality since

$$\begin{aligned} & P(h_n'^{-1} \left| \|p_{w_i} - p_{w_j}\|_{2,n}^2 - \|p_{w_i} - p_{w_j}\|_2^2 \right| > \epsilon) \\ &= P\left(h_n'^{-1} \left| (n-2)^{-1} \sum_{s \neq i,j} (p_{w_i w_s} - p_{w_j w_s})^2 - \int (p_{w_i}(s) - p_{w_j}(s))^2 ds \right| > \epsilon\right) \\ &\leq 2 \exp\left(\frac{-(n-2)h_n' \epsilon}{2 + 2\sqrt{h_n' \epsilon}/3}\right) \end{aligned}$$

which is  $o(1)$  since  $nh_n' \rightarrow \infty$ . This completes the proof.  $\square$

**Lemma 2:** Suppose Assumption 1 holds. Then for every  $\epsilon > 0$  there exists a  $\delta > 0$  such that with probability at least  $1 - \epsilon^2/4$

$$\|p_{w_i} - p_{w_j}\|_2 \leq \delta \implies \|f_{w_i} - f_{w_j}\|_2 \leq \epsilon$$

**Proof of Lemma 2:** I first note that since  $f$  is Lebesgue measurable, Lusin's theorem (Dudley (2002), Theorem 7.5.2) implies that it is almost everywhere equivalent to a uniformly continuous function. That is, for any  $\eta' > 0$ , there exists a closed subset  $A$  of  $[0, 1]^2$  with measure at least  $1 - \eta'$  such that  $f$  is uniformly continuous when restricted to  $A$ .

It follows that for any  $\eta > 0$  there must also exist, a closed subset  $B$  of  $[0, 1]$  with measure of at least  $1 - \eta$  such that for any  $b \in B$ , there exists another closed subset  $C(b)$  of  $[0, 1]$  with measure of at least  $1 - \eta$ , such that for any  $c \in C(b)$ ,  $f$  is uniformly continuous when restricted to the set  $A' = \{(b, c) \in [0, 1]^2 : b \in B, c \in C(b)\}$ .

Second, I show that for all  $\epsilon' > 0$  there exists a  $\delta(\epsilon', \eta) > 0$  such that  $\|p_{w_i} - p_{w_j}\|_2 \leq \delta(\epsilon', \eta)$  implies  $|\int f_{w_i}(s)(f_{w_i}(s) - f_{w_j}(s))ds| < \epsilon'$  with probability at least  $1 - \epsilon'/4$ , so long as  $\eta \leq \epsilon'/16$ .

Specifically, I prove the contrapositive. Suppose  $|\int f_{w_i}(s)(f_{w_i}(s) - f_{w_j}(s))ds| \geq \epsilon'$ . Then by the negative triangle inequality  $|\int f_\tau(s)(f_{w_i}(s) - f_{w_j}(s))ds| > \epsilon'/2$  for any  $\tau \in [0, 1]$  chosen such that  $|\int (f_\tau(s) - f_{w_i}(s))(f_{w_i}(s) - f_{w_j}(s))ds| < \epsilon'/4$ . Since  $\|f_{w_i} - f_{w_j}\|_2 \leq 1$  for every  $(w_i, w_j)$  pair, it follows by Cauchy-Schwartz that  $\|f_{w_i} - f_\tau\|_2 \leq \epsilon'/4$  implies  $|\int f_\tau(s)(f_{w_i}(s) - f_{w_j}(s))ds| > \epsilon'/2$ .

Since  $f$  is uniformly continuous when restricted to  $A'$ , there exists a universal  $\omega(\epsilon', \eta) > 0$  such that  $|\tau - w_i| < \omega(\epsilon', \eta)$  implies that  $\|f_\tau - f_{w_i}\|_2 < \epsilon'/8 + 2\eta$  so long as  $w_i, \tau \in B$ . Taking  $\eta \leq \epsilon'/16$  gives  $|\tau - w_i| < \omega(\epsilon', \eta)$  implies that  $\|f_\tau - f_{w_i}\|_2 < \epsilon'/4$  so long as  $w_i, \tau \in B$ . It follows that choosing  $\tau$  such that  $|\tau - w_i| < \omega(\epsilon', \eta)$  implies  $|\int f_\tau(s)(f_{w_i}(s) - f_{w_j}(s))ds| > \epsilon'/2$

It is without loss to further restrict  $\omega(\epsilon', \eta) < \epsilon'/16$ . Since  $w_i$  is uniformly distributed on  $[0, 1]$ , the probability that  $w_i$  is in the  $\epsilon'/16$  interior of  $B$  (that is, the interval  $(w_i - \epsilon'/16, w_i + \epsilon'/16)$  is contained in  $B$ ) is greater than  $1 - \eta - 2\omega(\epsilon', \eta) \geq 1 - \epsilon'/4$ . This implies that  $|\int f_\tau(s)(f_{w_i}(s) - f_{w_j}(s))ds| > \epsilon'/2$  on a subset of  $[0, 1]$  of measure at least  $2\omega(\epsilon', \eta)$  with probability at least  $1 - \epsilon'/4$ .

Thus  $|\int f_{w_i}(s)(f_{w_i}(s) - f_{w_j}(s))ds| \geq \epsilon'$  implies  $\int (\int f_\tau(s)(f_{w_i}(s) - f_{w_j}(s))ds)^2 d\tau > (\epsilon'/2)^2 \times 2\omega(\epsilon', \eta)$  with probability at least  $1 - \epsilon'/4$ . Since the left hand side is just  $\|p_i - p_j\|_2^2$ , it follows that  $\|p_i - p_j\|_2 > (\epsilon'/2) \times (2\omega(\epsilon', \eta))^{1/2}$  with probability at least  $1 - \epsilon'/4$ , which proves this second part. Taking the contrapositive yields  $\|p_i - p_j\|_2 \leq \delta(\epsilon', \eta)$  implies that  $|\int f_{w_i}(s)(f_{w_i}(s) - f_{w_j}(s))ds| < \epsilon'$  with probability at least  $1 - \epsilon'/4$ , where  $\delta(\epsilon', \eta) = (\epsilon'/2) \times (2\omega(\epsilon', \eta))^{1/2}$ .

To finish the proof, note that  $|\int f_{w_i}(s)(f_{w_i}(s) - f_{w_j}(s))ds| < \epsilon'$  and  $|\int f_{w_j}(s)(f_{w_i}(s) - f_{w_j}(s))ds| < \epsilon'$  also imply that  $|\int (f_{w_i}(s) - f_{w_j}(s))(f_{w_i}(s) - f_{w_j}(s))ds| < 2\epsilon'$  by the triangle inequality, so that  $\|p_i - p_j\|_2 \leq (\epsilon'/2) \times (2\omega(\epsilon', \eta))^{1/2}$  implies  $\|f_{w_i} - f_{w_j}\|_2 < \sqrt{2\epsilon'}$  with probability at least  $1 - \epsilon'/2$ . Thus  $\|p_i - p_j\|_2 \leq \delta(\epsilon, \eta)$  implies  $\|f_{w_i} - f_{w_j}\|_2 < \epsilon$  with probability at least  $1 - \epsilon^2/4$  as claimed, where  $\delta(\epsilon, \eta) = (\epsilon^2/4) \times (2\omega(\epsilon^2/2, \eta))^{1/2}$ .  $\square$

The proof of Theorem 2 also relies on the auxiliary Lemma A1.

**Lemma A1:** Suppose Assumption 1 holds. Then for any  $\epsilon > 0$ ,  $P(\|f_{w_i} - f_{w_j}\|_2 \leq \epsilon) > 0$ .

**Proof of Lemma A1:** Following the first part of the proof of Lemma 2, Lusin's theorem implies that for any  $\eta > 0$  there exists  $B$ , a closed subset of  $[0, 1]$  with measure of at least  $1 - \eta$  such that for any  $b \in B$ , there exists another closed subset  $C(b)$  of  $[0, 1]$  with measure of at least  $1 - \eta$ , such that for any  $c \in C(b)$ ,  $f$  is uniformly continuous when restricted to the set  $A' = \{(b, c) \in [0, 1]^2 : b \in B, c \in C(b)\}$ . That is, for all  $\epsilon' > 0$  and  $u, v \in B$  there exists a  $\omega(\epsilon', \eta) > 0$  such that  $|u - v| \leq \omega(\epsilon', \eta)$  implies that  $|f(u, t) - f(v, t)| \leq \epsilon'$  for  $t \in C(u) \cap C(v)$ , a set with Lebesgue measure at least  $1 - 2\eta$ .

So  $|u - v| \leq \omega(\epsilon', \eta)$  and  $u, v \in B$  imply that  $\|f_u - f_v\|_2 \leq (\epsilon'^2(1 - 2\eta) + 2\eta)^{1/2} \leq \epsilon' + \sqrt{2\eta}$ . Since  $w_i, w_j$  are independent with standard uniform marginals,  $\|f_{w_i} - f_{w_j}\|_2 \leq \epsilon' + \sqrt{2\eta}$  with probability at least  $(1 - 2\eta)\omega(\epsilon', \eta)$ . Now just choose  $\epsilon' < \epsilon/2$  and  $\eta < \epsilon'^2/2$  to get  $P(\|f_{w_i} - f_{w_j}\|_2 \leq \epsilon) \geq (1 - \epsilon^2/8)\omega(\epsilon/2, \epsilon^2/8) > 0$ .  $\square$

Lemma 2 and Lemma A1 imply that for any  $\epsilon > 0$ ,  $P(\|p_{w_i} - p_{w_j}\|_2 \leq \epsilon) > 0$ .

**Theorem 2:** Suppose Assumptions 1-5 hold. Then  $\hat{\beta} \rightarrow_p \beta$ .

**Proof of Theorem 2:** Let  $u_i = y_i - x_i\beta$ ,  $\delta_{ij} = \delta(w_i, w_j)$ ,  $r_n = E\left[K\left(\frac{\delta_{ij}^2}{h_n}\right)\right]$ , and write

$$\hat{\beta} = \beta + \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)'(x_i - x_j) K\left(\frac{\hat{\delta}_{ij}^2}{h_n}\right) \right)^{-1} \left( \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)'(u_i - u_j) K\left(\frac{\hat{\delta}_{ij}^2}{h_n}\right) \right)$$

I show  $\binom{(n)}{(2)}r_n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)'(x_i - x_j)K\left(\frac{\hat{\delta}_{ij}^2}{h_n}\right) \rightarrow_p 2\Gamma_0$ . Similar arguments yield  $\binom{(n)}{(2)}r_n^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)'(u_i - u_j)K\left(\frac{\hat{\delta}_{ij}^2}{h_n}\right) \rightarrow_p 0$ , so that the claim follows from Slutsky and the continuous mapping theorem. Since  $r_n > 0$  with probability one from Lemma A1, both statistics are well-defined.

Let  $D_n = \binom{(n)}{(2)}r_n^{-1} \sum_i \sum_j (x_i - x_j)'(x_i - x_j)K\left(\frac{\hat{\delta}_{ij}^2}{h_n}\right)$ . By the mean value theorem  $D_n = \binom{(n)}{(2)}r_n^{-1} \sum_i \sum_j (x_i - x_j)'(x_i - x_j) \left[ K\left(\frac{\delta_{ij}^2}{h_n}\right) + K'\left(\frac{\iota_{ij}}{h_n}\right) \left(\frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n}\right) \right]$  where  $\{\iota_{ij}\}_{i \neq j}$  is the collection of intermediate values implied by that theorem. By Lemma 1  $\max_{i \neq j} \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} = o_p(n^{-\gamma/4})$  and by Markov's inequality  $K'\left(\frac{\iota_{ij}}{h_n}\right) = o_p(r_n n^{\gamma/4})$ , since  $P\left(K'\left(\frac{\iota_{ij}}{h_n}\right) \geq r_n n^{\gamma/4}\right) \leq \frac{E[|K'(\frac{\iota_{ij}}{h_n})|]}{r_n n^{\gamma/4}} = O(n^{-\gamma/4})$  by choice of kernel density function. It follows that  $D_n = \binom{(n)}{(2)}r_n^{-1} \sum_i \sum_j (x_i - x_j)'(x_i - x_j)K\left(\frac{\delta_{ij}^2}{h_n}\right) + o_p(1)$  since  $x_i$  has finite second moments and  $K'(u)$  is bounded.

$D'_n := \binom{(n)}{(2)}r_n^{-1} \sum_i \sum_j (x_i - x_j)'(x_i - x_j)K\left(\frac{\delta_{ij}^2}{h_n}\right)$  is a second order U-statistic with kernel depending on  $n$ , in the sense of [Ahn and Powell \(1993\)](#). In particular, their Lemma A.3 implies  $D'_n = r_n^{-1} E \left[ (x_i - x_j)'(x_i - x_j)K\left(\frac{\delta_{ij}}{h_n}\right) \right] + o_p(1)$  since  $nr_n \rightarrow \infty$ . Measurability of  $f$  and Assumption 4 further give

$$\begin{aligned} E \left[ (x_i - x_j)'(x_i - x_j)K\left(\frac{\delta_{ij}}{h_n}\right) \right] &= \int E[(x_i - x_j)'(x_i - x_j) | \delta_{ij} = u] K\left(\frac{u}{h_n}\right) dP(\delta_{ij} = u) du \\ &= \int (\Gamma_0 + o_p(1)) K\left(\frac{u}{h_n}\right) dP(\delta_{ij} = u) du = \Gamma_0 r_n + o_p(r_n) \end{aligned}$$

in which  $dP(\delta_{ij} = u)$  denotes the Radon-Nikodym derivative of the measure  $P(\delta_{ij} \leq u)$  with respect to the Lebesgue measure on  $[0, 1]^2$ ,<sup>1</sup> the second equality is due to

$E[(x_i - x_j)'(x_i - x_j) | \delta_{ij} \leq u] = \Gamma_0 + o_p(1)$  by Lemma 2 and Assumptions 2 and 4. So  $D_n = \Gamma_0 + o_p(1)$ .

---

<sup>1</sup> That is,  $dP(\delta_{ij} = u)$  satisfies  $\int_{u \in A} dP(\delta_{ij} = u) du = P(\delta_{ij} \in A)$  for any Lebesgue measurable subset of  $[0, 1]^2$ . Existence of this derivative follows from measurability of  $f$ , see [Dudley \(2002\)](#), Theorem 5.5.4.

A nearly identical argument yields

$$U_n = \left( \binom{n}{2} r_n \right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_i - x_j)' (u_i - u_j) K \left( \frac{\hat{\delta}_{ij}}{h_n} \right) = o_p(1)$$

since  $E[(x_i - x_j)'(u_i - u_j) | d(w_i, w_j) = h_n] = o_p(1)$  by Assumptions 3 and 4.

$(\hat{\beta} - \beta) = D_n^{-1} U_n = o_p(1)$  follows from Slutsky's theorem.  $\square$

#### A.4 Lemmas and Theorems in Section 3.3.2

The proof of Theorem 3 uses the discreteness of the network types to strengthen Lemma 1 to auxiliary Lemma A2.

**Lemma A2:** Suppose Assumption 4 holds and  $f_{w_i}$  has finite support. Then there exists an  $\epsilon > 0$  such that  $\max_{i \neq j} \hat{\delta}_{ij}^2 \times \mathbb{1}\{\hat{\delta}_{ij}^2 \leq \epsilon/2\} = o_p(n^{-1/2} h_n)$

**Proof of Lemma A2:** The assumption that  $f_{w_i}$  has finite support implies there exists an  $\epsilon > 0$  such that  $\delta_{ij}^2 \mathbb{1}\{\delta_{ij}^2 \leq \epsilon\} = 0$  and  $(p_{w_i w_t} - p_{w_j w_t}) \times \mathbb{1}\{\delta_{ij}^2 \leq \epsilon/2\} = 0$  both with probability one. For such an  $\epsilon$ , write

$$\hat{\delta}_{ij}^2 \mathbb{1}\{\hat{\delta}_{ij}^2 \leq \epsilon/2\} = \hat{\delta}_{ij}^2 \left( \mathbb{1}\{\hat{\delta}_{ij}^2 \leq \epsilon/2\} - \mathbb{1}\{\delta_{ij}^2 \leq \epsilon/2\} \right) + \hat{\delta}_{ij}^2 \mathbb{1}\{\delta_{ij}^2 \leq \epsilon/2\}$$

First,  $\max_{i \neq j} \sqrt{n} h_n^{-1} \hat{\delta}_{ij}^2 \mathbb{1}\{\delta_{ij}^2 \leq \epsilon/2\} = o_p(1)$  because  $(p_{w_i w_t} - p_{w_j w_t}) \times \mathbb{1}\{\delta_{ij}^2 \leq \epsilon/2\} = 0$ , Bernstein's inequality, and the union bound imply

$$P \left( \max_{i,j,t} \left[ (n-3)^{-1} \sum_{s \neq i,j,t} D_{ts} (D_{is} - D_{js}) \right]^2 \mathbb{1}\{\delta_{ij}^2 \leq \epsilon/2\} \geq \eta \right) \leq 2n^3 \exp \left( \frac{-(n-3)\eta}{3} \right)$$

and averaging over  $t \neq i, j$  gives

$$P \left( \max_{i,j} \sqrt{n} h_n^{-1} \hat{\delta}_{ij}^2 \mathbb{1}\{\delta_{ij}^2 \leq \epsilon/2\} \geq \eta \right) \leq 2n^3 \exp \left( \frac{-(n-3)\eta h_n}{3\sqrt{n}} \right) = o(1)$$

Second, since  $\delta_{ij}^2 \in (\epsilon/4, 3\epsilon/4)$  is a probability zero event,

$$\begin{aligned} \sqrt{n}h_n^{-1}|\hat{\delta}_{ij}^2 \left(1\{\hat{\delta}_{ij}^2 \leq \epsilon/2\} - 1\{\delta_{ij}^2 \leq \epsilon/2\}\right)| &\leq 2\sqrt{n}h_n^{-1} \times 1\{|\hat{\delta}_{ij}^2 - \delta_{ij}^2| > |\epsilon/2 - \delta_{ij}^2|\} \\ &\leq 2\sqrt{n}h_n^{-1}1\{|\hat{\delta}_{ij}^2 - \delta_{ij}^2| > \epsilon/4\} \end{aligned}$$

and so  $\max_{i \neq j} \sqrt{n}h_n^{-1}|\hat{\delta}_{ij}^2 \left(1\{\hat{\delta}_{ij}^2 \leq \epsilon/2\} - 1\{\delta_{ij}^2 \leq \epsilon/2\}\right)| = o_p(1)$  by previous arguments.  $\square$

**Theorem 3:** Suppose Assumptions 1-4 hold and  $f_{w_i}$  has finite support. Then

$$V_{3,n}^{-1/2} \left( \hat{\beta} - \beta \right) \rightarrow_d \mathcal{N}(0, I_k)$$

where  $V_3 = \Gamma_0^{-1} \Omega_0 \Gamma_0^{-1} \times s/n$ ,  $\Gamma_0$  is as defined in Assumption 3,  $I_k$  is the  $k \times k$  identity matrix, and

$$\begin{aligned} s &= P(\|p_i - p_j\|_2 = 0, \|p_i - p_k\|_2 = 0) / P(\|p_i - p_j\|_2 = 0)^2 \\ \Omega_0 &= E[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k) \mid \|p_i - p_j\|_2 = 0, \|p_i - p_k\|_2 = 0] \end{aligned}$$

**Proof of Theorem 3:** In the proof of Theorem 2, I demonstrate that Assumptions 1-5 are sufficient for

$$\frac{1}{m} \sum_i \sum_{j>i} (x_i - x_j)'(x_i - x_j) K \left( \frac{\hat{\delta}_{ij}^2}{h_n} \right) \rightarrow_p 2\Gamma_0 E \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) \right]$$

where  $m = n(n-1)/2$  and  $\delta_{ij} = \delta(w_i, w_j)$ . Since the support of  $f_{w_i}$  is finite,  $E \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) \right] = K(0)P(\delta_{ij} = 0) > 0$  eventually.

As for the numerator, I follow the proof of Theorem 2 to write

$$U_n = \frac{1}{m} \sum_i \sum_{j>i} \Delta_{ij} K \left( \frac{\hat{\delta}_{ij}^2}{h_n} \right) = \frac{1}{m} \sum_i \sum_{j>i} \Delta_{ij} \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) + K' \left( \frac{\iota_{ij}}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right) 1\{\hat{\delta}_{ij}^2 \leq h_n\} \right]$$

where  $\iota_{ij}$  is a mean value between  $\delta_{ij}^2$  and  $\hat{\delta}_{ij}^2$  and  $\Delta_{ij} = (x_i - x_j)'(u_i - u_j)$ . I first show



$\frac{1}{m} \sum_i \sum_{j>i} \Delta_{ijl} K' \left( \frac{\iota_{ij}}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right) 1_{\{\hat{\delta}_{ij}^2 \leq h_n\}} = o_p(n^{-1/2})$  where  $\Delta_{ijl}$  is the  $l$ th component of  $\Delta_{ij}$ . By Cauchy-Schwartz

$$\begin{aligned} & \frac{1}{m} \left| \sum_i \sum_{j>i} \left( \Delta_{ijl} K' \left( \frac{\iota_{ij}}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right) \right) \right| \\ & \leq \frac{\bar{K}'}{m} \left( \sum_i \sum_{j>i} \Delta_{ijl}^2 \right)^{1/2} \times \left( \sum_i \sum_{j>i} \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right)^2 1_{\{\hat{\delta}_{ij}^2 \leq h_n\}} \right)^{1/2} \end{aligned}$$

where  $\bar{K}' = \sup_{u \in [0,1]} K'(u)$ ,  $\sum_i \sum_{j>i} \Delta_{ijl}^2 = O_p(m)$  since  $x_i$  and  $u_i$  have finite fourth moments, and  $\max_{i \neq j} \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right) 1_{\{\hat{\delta}_{ij}^2 \leq h_n\}} = o_p(n^{-1/2})$  by Lemma A2.

It follows from this result that

$$U_n = \frac{1}{m} \sum_i \sum_{j>i} \Delta_{ij} K \left( \frac{\delta_{ij}^2}{h_n} \right) + o_p(n^{-1/2})$$

The first summand is a second order U-statistic with symmetric  $L^2$ -integrable kernel, so by Lemma A.3 of [Ahn and Powell \(1993\)](#)

$$\sqrt{n}(U_n - U) \rightarrow \mathcal{N}(0, V)$$

where  $U = E \left[ \Delta_{ij} K \left( \frac{\delta_{ij}^2}{h_n} \right) \right]$  and for  $Z_i = (x_i, \varepsilon_i, w_i)$

$$V = \lim_{h \rightarrow 0} 4E \left[ E \left[ \Delta_{ij} K \left( \frac{\delta_{ij}^2}{h} \right) \mid Z_i \right] E \left[ \Delta'_{ij} K \left( \frac{\delta_{ij}^2}{h} \right) \mid Z_i \right] \right] = \lim_{h \rightarrow 0} 4E \left[ \Delta_{ij} \Delta'_{ik} K \left( \frac{\delta_{ij}^2}{h} \right) K \left( \frac{\delta_{ik}^2}{h} \right) \right]$$

Since  $f_{w_i}$  has finite support,  $E[\delta_{ij}^2 | \delta_{ij}^2 \leq \epsilon] = 0$  for some  $\epsilon > 0$ , and so

$U = E[\Delta_{ij} K(0) 1_{\{\delta_{ij} = 0\}}] = 0$  for  $n$  sufficiently large such that  $h_n \leq \epsilon$ . Similarly

$V = 4\Omega_0 K(0)^2 P(\delta_{ij} = 0, \delta_{ik} = 0)$ . So by Slutsky's Theorem,

$$\sqrt{n}(\hat{\beta} - \beta) \rightarrow_d \mathcal{N}(0, V_3)$$

where  $V_3 = \Gamma_0^{-1} \Omega_0 \Gamma_0^{-1} \times s$  as claimed.  $\square$

**Lemma 3:** Suppose Assumptions 1 and 5 hold. Then for almost every  $(w_i, w_j)$  pair

$$\|p_{w_i} - p_{w_j}\|_2 \leq \|f_{w_i} - f_{w_j}\|_2 \leq 32 C_6^{\frac{1}{2+4\alpha}} (\|p_{w_i} - p_{w_j}\|_2)^{\frac{\alpha}{1+2\alpha}}$$

so long as  $\|p_{w_i} - p_{w_j}\|_2 < \sqrt{8C_6}K^{-\alpha}$ .

**Proof of Lemma 3:** The first inequality is proven in the text and holds exactly for any measurable  $f$  with  $\|f\|_\infty \leq 1$  and every  $(w_i, w_j)$  pair. The proof of the second inequality essentially mirrors the proof of Lemma 2, and so only a sketch is provided here. I first demonstrate  $\|p_{w_i} - p_{w_j}\|_2 \leq (4(4C_6)^{1/\alpha})^{-1} \epsilon'^{\frac{4\alpha+2}{\alpha}}$  and  $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}} < K^{-1}$  imply  $\|f_{w_i} - f_{w_j}\|_2 \leq \sqrt{2\epsilon'}$  with probability one.

Suppose  $|\int f_{w_i}(s) (f_{w_i}(s) - f_{w_j}(s)) ds| > \epsilon'$ . Then  $|\int f_\tau(s) (f_{w_i}(s) - f_{w_j}(s)) ds| > \epsilon'/2$  for  $\tau \in [0, 1]$  so long as  $\tau$  and  $w_i$  are in the same block of the partition of  $[0, 1]$  and  $C_6|w_i - \tau|^\alpha < \epsilon'/4$ . If  $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}} < K^{-1}$ , then the measure of  $\tau$  in  $[0, 1]$  that satisfies these conditions is at least  $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}}$ . It follows that so long as  $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}} < K^{-1}$

$$\int \left( \int f_\tau(s) (f_{w_i}(s) - f_{w_j}(s)) ds \right)^2 d\tau > \left(\frac{\epsilon'}{2}\right)^2 \left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}}$$

with probability one. The claim then follows from the last step in the proof of Lemma 2.

Now set  $\epsilon = \sqrt{2\epsilon'}$ . It follows that for almost every  $w_i$  and  $w_j$ ,  $2^{\frac{2\alpha+10}{4\alpha+2}} C_6^{\frac{1}{4\alpha+2}} \|p_i - p_j\|_2^{\frac{2\alpha}{4\alpha+2}} = \epsilon$  implies that  $\|f_{w_i} - f_{w_j}\|_2 \leq \epsilon$ , so long as  $\epsilon < \sqrt{8C_6}K^{-\alpha/2}$ . The claim then follows by noting that  $2^{\frac{2\alpha+10}{4\alpha+2}}$  is bounded below 32 when  $\alpha > 0$ .  $\square$

The proof of Theorem 4 relies on the following strengthening of auxiliary Lemma A1 to auxiliary Lemma A3.

**Lemma A3:** Suppose Assumptions 1 and 5 hold. Then  $P(\|f_{w_i} - f_{w_j}\|_2 \leq \epsilon) > C_6^{-1/\alpha} \epsilon^{1/\alpha}$ , so long as  $\epsilon \leq C_6 K^{-\alpha}$

**Proof of Lemma A3:** The proof of Lemma A3 mirrors that of Lemma A1 except Assumption 6 allows us to replace  $\omega(\epsilon, \eta)$  with  $\left(\frac{\epsilon}{C_6}\right)^{1/\alpha}$ . So long as  $K \leq \left(\frac{\epsilon}{C_6}\right)^{-\frac{1}{\alpha}}$  the probability that  $w_i$  and  $w_j$  are in the same partition of  $[0, 1]$  and that  $|w_i - w_j| \leq \left(\frac{\epsilon}{C_6}\right)^{1/\alpha}$  is bounded from below by  $\left(\frac{\epsilon}{C_6}\right)^{1/\alpha}$  and the claim follows.  $\square$

**Theorem 4:** Suppose Assumptions 1-3 and 5-6 hold. Then

$$V_{4,n}^{-1/2} \left( \hat{\beta} - \beta_{h_n} \right) \rightarrow_d \mathcal{N} \left( 0, I_k \right)$$

where  $V_{4,n} = \Gamma_0^{-1} \Omega_n \Gamma_0^{-1} / n$ ,  $\Gamma_0$  is as defined in Assumption 3,  $r_n$  is as defined in Assumption 5,  $I_k$  is the  $k \times k$  identity matrix, and

$$\begin{aligned} \beta_{h_n} &= \beta + (\Gamma_0)^{-1} E \left[ (x_i - x_j)' (\lambda(w_i) - \lambda(w_j)) K \left( \frac{\|p_i - p_j\|_2}{h_n} \right) \right] / (2r_n) \\ \Omega_n &= \frac{4}{r_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_1 j_2} K \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) \right] \\ &\quad + \frac{1}{r_n^2 h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_1 s_{21} s_{22}} - \delta_{i_2 j_2}^2) \right] \\ &\quad + \frac{4}{r_n^2 h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_2 s_{11} s_{22}} - \delta_{i_2 j_2}^2) \right] \end{aligned}$$

with  $\Delta_{ij} = (x_i - x_j)'(u_i - u_j)$ ,  $u_i = \lambda(w_i) + \varepsilon_i$ ,  $\delta_{ij} = \delta(w_i, w_j)$ , and

$$F_{i_j t s_1 s_2} = f(w_t, w_{s_1}) f(w_t, w_{s_2}) (f(w_i, w_{s_1}) - f(w_j, w_{s_1})) (f(w_i, w_{s_2}) - f(w_j, w_{s_2})).$$

**Proof of Theorem 4:** The proof of Theorem 2 demonstrates that Assumptions 1-3 and 6 are sufficient for the denominator to converge in probability to  $2\Gamma_0$ . As for the numerator,

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{2} r_n} \sum_i \sum_{j>i} \Delta_{ij} K \left( \frac{\hat{\delta}_{ij}^2}{h_n} \right) \\ &= \frac{1}{\binom{n}{2} r_n} \sum_i \sum_{j>i} \Delta_{ij} \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) + K' \left( \frac{\delta_{ij}^2}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n^2} \right) + K'' \left( \frac{\iota_{ij}}{h_n^2} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n^2} \right)^2 \right] \end{aligned}$$

where  $\iota_{ij}$  is the intermediate value between  $\hat{\delta}_{ij}^2$  and  $\delta_{ij}^2$ , suggested by Taylor and the mean

value theorem. First, I show that

$$\frac{1}{\binom{n}{2}r_n} \sum_i \sum_{j>i} \Delta_{ij} K'' \left( \frac{\iota_{ij}}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right)^2 = o_p(n^{-1/2})$$

Let  $s_n = n^{-1/2}h_n^4 r_n$ . Since  $\delta_{ij} \leq C|w_i - w_j|^\alpha$  by the first part of Lemma 2 and Assumption 5,  $r_n \geq KC^{-1/\alpha}h_n^{1/\alpha}$  for  $K = \liminf_{h \rightarrow 0} E \left[ K \left( \frac{\delta_{ij}}{h} \right) \mid \delta_{ij} \leq h \right] > 0$  by Lemma A2. Since  $n^{1/2-\gamma}h_n^{4+1/\alpha} \rightarrow \infty$  for some  $\gamma > 0$  by Assumption 9,  $n^{1-\gamma}s_n \rightarrow \infty$ , and so Lemma 1 implies that  $\sup_{i \neq j} \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{\sqrt{s}} \right)^2 = o_p(1)$  or  $\sup_{i \neq j} \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n^2 \sqrt{r_n}} \right)^2 = o_p(n^{-1/2})$ . It follows that

$$\frac{1}{\binom{n}{2}r_n} \sum_i \sum_{j>i} \Delta_{ij} K'' \left( \frac{\iota_{ij}}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right)^2 \leq \frac{\bar{K}''}{\binom{n}{2}} \sum_i \sum_{j>i} \Delta_{ij} \times o_p(n^{-1/2})$$

where  $\bar{K}'' = \sup_{u \in [0,1]} K''(u)$  and the last line is  $o_p(n^{-1/2})$  because  $x_i$  and  $u_i$  have finite fourth moments. It follows from this first step that

$$U_n = \frac{1}{\binom{n}{2}r_n} \sum_i \sum_{j>i} \Delta_{ij} \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) + K' \left( \frac{\delta_{ij}^2}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right) \right] + o_p(n^{-1/2})$$

Second, I show that

$$U_n = \frac{1}{\binom{n}{5}^2 r_n} \sum_i \sum_{j>i} \sum_{t>j} \sum_{s_1>t} \sum_{s_2>s_1} \Delta_{ij} \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) + \frac{1}{h_n} K' \left( \frac{\delta_{ij}^2}{h_n} \right) (F_{ijts_1s_2} - \delta_{ij}^2) \right] + o_p(n^{-1/2})$$

where  $F_{ijts_1s_2} = f_{ts_1} f_{ts_2} (f_{is_1} - f_{js_1})(f_{is_2} - f_{js_2})$ . Let

$\tilde{\delta}_{ij}^2 = \binom{n-j}{3}^{-1} \sum_{t>j} \sum_{s_1>t} \sum_{s_2>s_1} F_{ijts_1s_2}$ . Then

$$\begin{aligned} U_n &= \frac{1}{\binom{n}{2}r_n} \sum_i \sum_{j>i} \Delta_{ij} \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) + K' \left( \frac{\delta_{ij}^2}{h_n} \right) \left( \frac{\tilde{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right) \right] \\ &\quad + \frac{1}{\binom{n}{2}r_n} \sum_i \sum_{j>i} \Delta_{ij} \left[ K' \left( \frac{\delta_{ij}^2}{h_n} \right) \left( \frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n} \right) \right] + o_p(n^{-1/2}) \end{aligned}$$

and the second summand is  $o_p(n^{-1/2})$  by Chebyshev's inequality, since it has mean zero

and variance

$$\frac{1}{\binom{n}{2}^2 n^6 r_n^2 h_n^2} E \left[ \sum_{i_1} \sum_{i_2} \sum_{j_1} \sum_{j_2} \sum_{t_1} \sum_{t_2} \sum_{s_{11}} \sum_{s_{12}} \sum_{s_{21}} \sum_{s_{22}} \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) \right. \\ \left. \times (D_{i_1 j_1 t_1 s_{11} s_{12}} - F_{i_1 j_1 t_1 s_{11} s_{12}}) \times (D_{i_2 j_2 t_2 s_{21} s_{22}} - F_{i_2 j_2 t_2 s_{21} s_{22}}) \right]$$

where  $D_{ijts_1s_2} = D_{ts_1} D_{ts_2} (D_{is_1} - D_{js_1})(D_{is_2} - D_{js_2})$ . To see that this variance is  $o(n^{-1})$ , note that unless two elements from the set  $\{i_1, j_1, t_1, s_{11}, s_{12}\}$  equal two in  $\{i_2, j_2, t_2, s_{21}, s_{22}\}$ ,  $\{\eta_{t_1 s_{11}}, \eta_{t_1 s_{12}}, \eta_{i_1 s_{11}}, \eta_{j_1 s_{11}}, \eta_{i_1 s_{12}}, \eta_{j_1 s_{12}}\}$  is independent of  $\{\eta_{t_2 s_{21}}, \eta_{t_2 s_{22}}, \eta_{i_2 s_{21}}, \eta_{j_2 s_{21}}, \eta_{i_2 s_{22}}, \eta_{j_2 s_{22}}\}$  and so

$$E \left[ [D_{i_1 j_1 t_1 s_{11} s_{12}} - F_{i_1 j_1 t_1 s_{11} s_{12}}] \times [D_{i_2 j_2 t_2 s_{21} s_{22}} - F_{i_2 j_2 t_2 s_{21} s_{22}}] | Z_{i_1 j_1 t_1 s_{11} s_{12}}, Z_{i_2 j_2 t_2 s_{21} s_{22}} \right] = 0$$

where  $Z_i = \{x_i, w_i, \nu_i\}$  and  $Z_{ijts_1s_2} = \{Z_i, Z_j, Z_t, Z_{s_1}, Z_{s_2}\}$ . Since  $K' \left( \frac{\delta_{i_1 j_1}^2}{h} \right)$  is  $O_p(r_n)$ ,  $nh_n^4 \rightarrow \infty$  implies that this variance is  $o(n^{-1})$  and so the second summand is  $o_p(n^{-1/2})$ .

Let

$$U'_n = \frac{1}{\binom{n}{5}^2 r_n} \sum_i \sum_{j>i} \sum_{t>j} \sum_{s_1>t} \sum_{s_2>s_1} \Delta_{ij} \left[ K \left( \frac{\delta_{ij}^2}{h_n} \right) + \frac{1}{h_n} K' \left( \frac{\delta_{ij}^2}{h_n} \right) (F_{ijts_1s_2} - \delta_{ij}^2) \right] \\ = U_n + o_p(n^{-1/2})$$

$U'_n$  is a 5th order U-statistic which can be represented by the following iid sum (see for instance Lemma 3.2 of [Powell, Stock, and Stoker 1989](#))

$$U_n = E[U_n] + \frac{2}{nr_n} \sum_{\tau=1}^n \left( E \left[ \Delta_{\tau j} K \left( \frac{\delta_{\tau j}^2}{h_n} \right) | Z_\tau \right] - E[U_n] \right) \\ + \frac{1}{nr_n h_n} \sum_{\tau=1}^n E \left[ \Delta_{ij} K' \left( \frac{\delta_{ij}^2}{h_n} \right) (F_{ij\tau s_1 s_2} - \delta_{ij}^2) | Z_\tau \right] \\ + \frac{2}{nr_n h_n} \sum_{\tau=1}^n E \left[ \Delta_{ij} K' \left( \frac{\delta_{ij}^2}{h_n} \right) (F_{ij\tau s_2} - \delta_{ij}^2) | Z_\tau \right] + o_p(n^{-1/2})$$

where  $E[U_n] = r_n^{-1} E \left[ \Delta_{ij} K \left( \frac{\delta_{ij}^2}{h_n} \right) \right]$  and  $Z_\tau = \{x_\tau, w_\tau, \nu_\tau\}$ . In particular,  $U_n$  can be represented asymptotically by an iid sum of random variables, so by the Lindeberg Central Limit Theorem

$$V_n''^{-1/2} (U_n - E[U_n]) \rightarrow_d \mathcal{N}(0, I_k)$$

where for a collection of ten distinct agents  $\{i_1, i_2, j_1, j_2, t_1, t_2, s_{11}, s_{12}, s_{21}, s_{22}\}$

$$\begin{aligned} V_n'' &= \frac{4}{r_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_1 j_2} K \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) \right] \\ &+ \frac{1}{r_n^2 h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_2 s_{21} s_{22}} - \delta_{i_2 j_2}^2) \right] \\ &+ \frac{4}{r_n^2 h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_2 s_{11} s_{22}} - \delta_{i_2 j_2}^2) \right] \end{aligned}$$

since  $E[U_n] \rightarrow_p 0$  by Theorem 2. It follows from Slutsky's Theorem that

$$V_{4,n}^{-1/2} \left( \hat{\beta} - \beta - (2\Gamma_0)^{-1} E[U_n] \right) \rightarrow_d \mathcal{N}(0, I_k)$$

where  $E[U_n] = E \left[ \Delta_{ij} K \left( \frac{\delta_{ij}^2}{h_n} \right) \right]$  as claimed.  $\square$

## A.5 Theorems in Sections 3.3.3 and 3.3.4

**Theorem 5:** Suppose Assumptions 1-3 and 5-7 hold, and  $L > (2\theta(1 + 2\alpha))/\alpha - 1$ . Then

$$V_{5,n}^{-1/2} (\bar{\beta}_L - \beta) \rightarrow_d \mathcal{N}(0, I_k)$$

where  $V_{5,n} = \sum_{l_1=1}^L \sum_{l_2=1}^L a_{l_1} a_{l_2} \Gamma_0^{-1} \Omega_{n,l_1 l_2} \Gamma_0^{-1} / n$ ,  $\Gamma_0$  is as defined in Assumption 3,  $r_{nl} = E \left[ K \left( \frac{\delta_{ij}^2}{c_l h_n} \right) \right]$ ,  $I_k$  is the  $k \times k$  identity matrix, and

$$\begin{aligned} \Omega_{n,l_1 l_2} &= \frac{4}{r_{nl_1} r_{nl_2}} E \left[ \Delta_{i_1 j_1} \Delta'_{i_1 j_2} K \left( \frac{\delta_{i_1 j_1}^2}{c_{l_1} h_n} \right) K \left( \frac{\delta_{i_2 j_2}^2}{c_{l_2} h_n} \right) \right] \\ &+ \frac{1}{r_{nl_1} c_{l_1} r_{nl_2} c_{l_2} h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{c_{l_1} h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{c_{l_2} h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_1 s_{21} s_{22}} - \delta_{i_2 j_2}^2) \right] \\ &+ \frac{4}{r_{nl_1} c_{l_1} r_{nl_2} c_{l_2} h_n^2} E \left[ \Delta_{i_1 j_1} \Delta'_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}^2}{c_{l_1} h_n} \right) K' \left( \frac{\delta_{i_2 j_2}^2}{c_{l_2} h_n} \right) (F_{i_1 j_1 t_1 s_{11} s_{12}} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t_2 s_{11} s_{22}} - \delta_{i_2 j_2}^2) \right] \end{aligned}$$

**Proof of Theorem 5:** Since  $\bar{\beta}_L = \sum_{l=1}^L a_l \hat{\beta}_{C_l h_n}$ , a straightforward extension of the proof of Theorem 4 (and the continuous mapping theorem) implies that

$$V_{5,n}^{-1/2} (\bar{\beta}_L - \bar{\beta}_{L,h_n}) = V_{5,n}^{-1/2} \sum_{l=1}^L a_l \left( \hat{\beta}_{C_l h_n} - \beta_{C_l h_n} \right) \rightarrow_d \mathcal{N}(0, I_k)$$

where  $\bar{\beta}_{L,h} = \sum_{l=1}^L a_l \beta_{C_l h}$  is the pseudo-truth associated with  $\beta_L$ , which can also be written

$$\begin{aligned} \bar{\beta}_{L,h} &= \beta + \sum_{l_1=1}^L \sum_{l_2=1}^L a_{l_1} (2\Gamma_0)^{-1} C_{l_2} (c_{l_1} h)^{l_2/\theta} + o_p(n^{-1/2}) \\ &= \beta + (2\Gamma_0)^{-1} \sum_{l_2} C_{l_2} \left[ \sum_{l_1} a_{l_1} c_{l_1}^{l_2/\theta} \right] h^{l_2/\theta} + o_p(n^{-1/2}) \end{aligned}$$

since  $\sum_{l_2} a_{l_2} = 1$  by choice of  $\{a_1, \dots, a_L\}$ . The second summand is 0 because,  $\{a_1, \dots, a_L\}$  also satisfies  $\left[ \sum_{l_1} a_{l_1} c_{l_1}^{l_2/\theta} \right] = 0$  for all  $l_2 \in \{1, \dots, L\}$  and the claim follows.  $\square$

**Theorem 6:** Suppose Assumptions 1-4 hold. Then  $\left( \hat{\Gamma}_{h_n}^{-1} \hat{\Omega}_{n,h_n,h_n} \hat{\Gamma}_{h_n}^{-1} - nV_{4,n} \right) \rightarrow_p 0$  and  $\left( \sum_{l_1=1}^L \sum_{l_2=1}^L a_{l_1} a_{l_2} \hat{\Gamma}_{c_{l_1} h_n}^{-1} \hat{\Omega}_{n,c_{l_1} h_n, c_{l_2} h_n} \hat{\Gamma}_{c_{l_2} h_n}^{-1} - nV_{5,n} \right) \rightarrow_p 0$

**Proof of Theorem 6:** I prove the second claim, which includes the first as a special case.

The proof of Theorem 2 demonstrates that Assumptions 1-4 are sufficient for

$r_{n,c}^{-1} \hat{\Gamma}_{c h_n} = 2\Gamma_0 + o_p(1)$  for any  $c > 0$  where  $\delta_{ij} = \delta(w_i, w_j)$  and  $r_{n,c} = \left( E \left[ K \left( \frac{\delta_{ij}}{c h_n} \right) \right] \right)$ . It

remains to be shown that  $(r_{n,c_1} r_{n,c_2})^{-1} \hat{\Omega}_{c_1 h_n, c_2 h_n}$  converges to  $\Omega_{n,c_1 c_2}$  for any  $c_1, c_2 > 0$ . I

consider the three terms that make up  $\hat{\Omega}_{c_1 h_n, c_2 h_n}$  separately.

The first term is  $\frac{1}{n^3 r_{n,c_1} r_{n,c_2}} \sum_{i=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \hat{\Delta}_{ij_1} \hat{\Delta}'_{ij_2} K\left(\frac{\hat{\delta}_{ij_1}}{h_1}\right) K\left(\frac{\hat{\delta}_{ij_2}}{h_2}\right)$ , where  $\hat{\Delta}_{ij} = (x_i - x_j)'(\hat{u}_i - \hat{u}_j)$  and  $\hat{u}_i = y_i - x_i \hat{\beta}$ . Lemma 2 and Theorem 2 imply that  $\hat{\delta}_{ij} = \delta_{ij} + o(1)$  and  $\hat{u}_i = u_i + o(1)$  where  $u_i = y_i - x_i \beta$ , and so the term converges to  $\frac{1}{n^3 r_{n,c_1} r_{n,c_2}} \sum_{i=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \Delta_{ij_1} \Delta'_{ij_2} K\left(\frac{\delta_{ij_1}}{h_1}\right) K\left(\frac{\delta_{ij_2}}{h_2}\right)$  by the continuous mapping theorem, which is a third order V-statistic in the sense of [Ahn and Powell \(1993\)](#), and thus converges in probability to  $\frac{1}{r_{n,c_1} r_{n,c_2}} E\left[\Delta_{ij_1} \Delta'_{ij_2} K\left(\frac{\delta_{ij_1}}{h_1}\right) K\left(\frac{\delta_{ij_2}}{h_2}\right)\right]$ .

The second term is

$$\frac{1}{n^5 c_1 h_1 r_{n,c_1} c_2 h_2 r_{n,c_2}} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \sum_{t=1}^n \hat{\Delta}_{i_1 j_1} \hat{\Delta}_{i_2 j_2} K'\left(\frac{\hat{\delta}_{i_1 j_1}^2}{c_1 h_1}\right) K'\left(\frac{\hat{\delta}_{i_2 j_2}^2}{c_2 h_2}\right) (\hat{F}_{i_1 j_1 t} - \hat{\delta}_{i_1 j_1}^2) (\hat{F}_{i_2 j_2 t} - \hat{\delta}_{i_2 j_2}^2)$$

where  $\hat{F}_{ijt} = \frac{1}{n^2} \sum_{s_1=1}^n \sum_{s_2=1}^n D_{ts_1} D_{ts_2} (D_{is_1} - D_{js_1}) (D_{is_2} - D_{js_2})$ . By previous arguments this converges to the fifth-order V-statistic

$$\frac{1}{n^5 c_1 h_1 r_{n,c_1} c_2 h_2 r_{n,c_2}} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \sum_{t=1}^n \Delta_{i_1 j_1} \Delta_{i_2 j_2} K'\left(\frac{\delta_{i_1 j_1}^2}{c_1 h_1}\right) K'\left(\frac{\delta_{i_2 j_2}^2}{c_2 h_2}\right) (F_{i_1 j_1 t} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t} - \delta_{i_2 j_2}^2)$$

where  $F_{ijt} = E[D_{ts_1} D_{ts_2} (D_{is_1} - D_{js_1}) (D_{is_2} - D_{js_2}) | w_i, w_j, w_t]$  is the probability limit of  $\hat{F}_{ijt}$ . The second term thus converges to

$$\frac{1}{c_1 h_1 r_{n,c_1} c_2 h_2 r_{n,c_2}} E\left[\Delta_{i_1 j_1} \Delta_{i_2 j_2} K'\left(\frac{\delta_{i_1 j_1}^2}{c_1 h_1}\right) K'\left(\frac{\delta_{i_2 j_2}^2}{c_2 h_2}\right) (F_{i_1 j_1 t} - \delta_{i_1 j_1}^2) (F_{i_2 j_2 t} - \delta_{i_2 j_2}^2)\right]$$

The third term is

$$\frac{4}{n^5 h_1 h_2} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \sum_{t=1}^n \hat{\Delta}_{i_1 j_1} \hat{\Delta}_{i_2 j_2} K'\left(\frac{\hat{\delta}_{i_1 j_1}}{h_1}\right) K'\left(\frac{\hat{\delta}_{i_2 j_2}}{h_2}\right) (\hat{F}'_{i_1 j_1 t} - \hat{\delta}_{i_1 j_1}^2) (\hat{F}'_{i_2 j_2 t} - \hat{\delta}_{i_2 j_2}^2)$$

where  $\hat{F}'_{ijs_1} = \frac{1}{n^2} \sum_{t=1}^n \sum_{s_2=1}^n D_{ts_1} D_{ts_2} (D_{is_1} - D_{js_1}) (D_{is_2} - D_{js_2})$ . By previous arguments



this converges to the fifth order V-statistic

$$\frac{4}{n^5 h_1 h_2} \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{j_1=1}^n \sum_{j_2=1}^n \sum_{t=1}^n \Delta_{i_1 j_1} \Delta_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}}{h_1} \right) K' \left( \frac{\delta_{i_2 j_2}}{h_2} \right) (F'_{i_1 j_1 t} - \delta_{i_1 j_1}^2) (F'_{i_2 j_2 t} - \delta_{i_2 j_2}^2)$$

where  $F'_{ij_{s_1}} = E[D_{ts_1} D_{ts_2} (D_{is_1} - D_{js_1}) (D_{is_2} - D_{js_2}) | w_i, w_j, w_{s_1}]$  is the probability limit of  $\hat{F}'_{ij_{s_1}}$ . The third term thus converges to

$$\frac{1}{c_1 h_n r_{n, c_1} c_2 h_2 r_{n, c_2}} E \left[ \Delta_{i_1 j_1} \Delta_{i_2 j_2} K' \left( \frac{\delta_{i_1 j_1}}{h_1} \right) K' \left( \frac{\delta_{i_2 j_2}}{h_2} \right) (F'_{i_1 j_1 t} - \delta_{i_1 j_1}^2) (F'_{i_2 j_2 t} - \delta_{i_2 j_2}^2) \right]$$

The claim then follows from the continuous mapping theorem.  $\square$

**Theorem 7:** Suppose Assumptions 1-3 and 5-6 hold. Then

$$\frac{1}{R} \sum_{r=1}^R (\hat{\beta}_r - \hat{\beta}) (\hat{\beta}_r - \hat{\beta})' \rightarrow_p V_{4,n} \text{ and } \frac{1}{R} \sum_{r=1}^R (\bar{\beta}_{Lr} - \bar{\beta}_L) (\bar{\beta}_{Lr} - \bar{\beta}_L)' \rightarrow_p V_{5,n} \text{ as } n, R \rightarrow \infty.$$

**Proof of Theorem 7:** The first claim essentially follows from the asymptotically linear representation for  $\hat{\beta}$  given in the proof of Theorem 4 and by Theorem 2.2 of [Bickel and Freedman \(1981\)](#). The second follows by identical arguments.

The proof of Theorem 4 indicates that under Assumptions 1-3 and 5-6

$$\hat{\beta} - \beta_{h_n} = \frac{1}{n} \sum_{\tau=1}^n (2\Gamma_0)^{-1} g_n(Z_\tau) + o_p(n^{-1/2})$$

where  $Z_\tau = \{X_\tau, w_\tau, \varepsilon_\tau\}$  and

$$g_n(Z_\tau) = 2 \left( E \left[ \Delta_{\tau j} K \left( \frac{\delta_{\tau j}^2}{h_n} \right) | Z_\tau \right] - E[U_n] \right) + E \left[ \Delta_{ij} K' \left( \frac{\delta_{ij}^2}{h_n} \right) (F_{ij\tau s_1 s_2} - \delta_{ij}^2) | Z_\tau \right] + 2E \left[ \Delta_{ij} K' \left( \frac{\delta_{ij}^2}{h_n} \right) (F_{ijt\tau s_2} - \delta_{ij}^2) | Z_\tau \right]$$

By definition of  $\hat{\beta}_r$

$$\hat{\beta}_r - \hat{\beta} = (2\Gamma_0)^{-1} \left( \frac{1}{n} \sum_{\tau'=1}^n g_n(Z_{\tau'r}) - \frac{1}{n} \sum_{\tau=1}^n g_n(Z_\tau) \right) + o_p(n^{-1/2})$$

in which  $Z_{\tau'r} = Z_{i_{\tau'r}}$ . By Theorem 2.2 (a) of [Bickel and Freedman \(1981\)](#),  $\{\sqrt{n}(\frac{1}{n} \sum_{\tau'=1}^n g_n(Z_{\tau'r}) - \frac{1}{n} \sum_{\tau=1}^n g_n(Z_{\tau}))\}_{r=1}^R$  is a conditionally independent (given  $\{Z_{\tau}\}_{\tau=1}^n$ ) sequence with entries weakly convergent (as  $n \rightarrow \infty$ ) to a  $k$ -dimensional normal distribution with mean 0 and variance  $E[g_n(Z_1)g_n(Z_1)']$ . The sufficient condition for this Theorem to hold is for  $E[||g_n(Z_1)||^2]$  to be finite, which follows from Assumption 1 and the choice of kernel density function in Assumption 4. Since  $\sigma(\{Z_{\tau}\}_{\tau=1}^n)$  is degenerate in the  $n \rightarrow \infty$  limit, the usual strong law of large numbers gives that  $R \rightarrow \infty$  implies  $\frac{1}{R} \sum_{r=1}^R (\hat{\beta}_r - \hat{\beta})(\hat{\beta}_r - \hat{\beta})'$  converges in probability to  $E[g(Z_1)g(Z_1)'] / n = V_{4,n}$  so long as  $E[||g_n(Z_1)||^4] < \infty$ . This last condition also follows from Assumptions 1 and 4, which completes the proof.  $\square$

## A.6 Theorems in Section 3.4

**Theorem 7:** Suppose Assumptions 1-4 hold. Then  $E \left[ \left( \widehat{\lambda}(w_i) - \lambda(w_i) \right)^2 \right] \rightarrow_p 0$ , where the expectation is taken with respect to  $w_i$ .

**Proof of Theorem 7** Let  $\lambda_i$ ,  $\hat{\lambda}_i$ , and  $\delta_{it}$  shorthand  $\lambda(w_i)$ ,  $\widehat{\lambda}(w_i)$ , and  $\delta(w_i, w_t)$  respectively. Recall that  $\hat{\lambda}_i = \sum_{t=1}^n (y_t - x_t \hat{\beta}) K \left( \frac{\delta_{it}^2}{h_n} \right) / \sum_{t=1}^n K \left( \frac{\delta_{it}^2}{h_n} \right)$ . First consider the denominator. Along the lines of the proof of Theorem 2, Lemma 1 and continuous differentiability of  $K$  implies

$$\max_{i=1, \dots, n} \left| \frac{1}{n} \sum_{t=1}^n K \left( \frac{\hat{\delta}_{it}^2}{h_n} \right) - \frac{1}{n} \sum_{t=1}^n K \left( \frac{\delta_{it}^2}{h_n} \right) \right| = o_p(n^{-\gamma/4} h_n)$$

while Hoeffding and Boole's inequality gives

$$\max_{i=1, \dots, n} \left| \frac{1}{n} \sum_{t=1}^n K \left( \frac{\delta_{it}^2}{h_n} \right) - E \left[ K \left( \frac{\delta_{it}^2}{h_n} \right) | w_i \right] \right| = o_p \left( \sqrt{\frac{\log n}{n}} \right)$$

so by the triangle inequality and choice of bandwidth sequence

$$\max_{i=1, \dots, n} \left| \frac{1}{n} \sum_{t=1}^n K \left( \frac{\hat{\delta}_{it}^2}{h_n} \right) - E \left[ K \left( \frac{\delta_{it}^2}{h_n} \right) | w_i \right] \right| = o_p(n^{-\gamma/4} h_n)$$

Applying the same logic to the numerator yields

$$\max_{i=1,\dots,n} \left| \frac{1}{n} \sum_{t=1}^n (y_t - x_t \hat{\beta}) K \left( \frac{\hat{\delta}_{it}^2}{h_n} \right) - E \left[ (y_t - x_t \hat{\beta}) K \left( \frac{\delta_{it}^2}{h_n} \right) | w_i \right] \right| = o_p(n^{-\gamma/4} h_n)$$

Let  $\underline{k} = \inf_{u \in [0,5]} K(u)$  with  $\underline{k} > 0$  by choice of kernel in Assumption 4. Then

$E \left[ K \left( \frac{\delta_{it}^2}{h_n} \right) | w_i \right] > \underline{k} P(\delta_{it}^2 \leq h_n/2 | w_i) > 0$  with probability one by Lemma A1, so the continuous mapping theorem implies

$$\max_{i=1,\dots,n} \left| \hat{\lambda}_i - \frac{E \left[ \lambda_t K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right]}{E \left[ K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right]} + \frac{E \left[ x_t K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right] (\hat{\beta} - \beta)}{E \left[ K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right]} \right| = o_p(n^{-\gamma/4} h_n)$$

Since  $x_i$  has finite second moments,  $\frac{E[x_t K(\frac{\delta_{it}}{h_n}) | w_i]}{E[K(\frac{\delta_{it}}{h_n}) | w_i]}$  is uniformly bounded with probability one, and so  $\max_{i=1,\dots,n} \frac{E[x_t K(\frac{\delta_{it}}{h_n}) | w_i] (\hat{\beta} - \beta)}{E[K(\frac{\delta_{it}}{h_n}) | w_i]} = o_p(1)$  by Theorem 2. It follows that

$$\begin{aligned} E \left[ \left( \hat{\lambda}_i - \lambda_i \right)^2 \right] &= E \left[ \left( \frac{E \left[ (\lambda_t - \lambda_i) K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right]}{E \left[ K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right]} \right)^2 \right] + o_p(n^{-\gamma/4} h_n) \\ &= E \left[ \left( \frac{\int E \left[ (\lambda_i - \lambda_t) | \delta_{it} = u, w_i \right] K \left( \frac{u}{h_n} \right) dP(\delta_{it} = u | w_i)}{\int K \left( \frac{u}{h_n} \right) dP(\delta_{it} = u | w_i)} \right)^2 \right] + o_p(n^{-\gamma/4} h_n) \end{aligned}$$

in which  $dP(\delta_{it} = u | w_i)$  refers to the Radon-Nikodym derivative of the measure

$P(\delta_{ij} \leq u | w_i)$  with respect to the Lebesgue measure on  $[0, 1]$  (see proof of Theorem 2 for more details). The first term in the last line is  $o_p(1)$  by Assumption 3 and Lemma 2.  $\square$

**Theorem 8:** Suppose Assumptions 1-4 and 8 hold. Let  $\hat{\lambda}_S = \{\widehat{\lambda}(w_i)\}_{i \in S}$  for some finite collection of agents  $S$ . Then

$$V_{8,n}^{-1/2} \left( \hat{\lambda}_S - \lambda_S \right) \rightarrow_d \mathcal{N}(0, I_{|S|})$$

where  $\lambda_S = \{\lambda(w_i)\}_{i \in S}$ ,  $V_{8,n} = \text{diag}(\{V_{8,n,i}\}_{i \in S})$ , and

$$V_{8,n,i} = \sum_{t=1}^n \left( \left( u_t K \left( \frac{\delta_{it}}{h_n} \right) - r'_{n,i} \right) - \frac{r'_{n,i}}{r_{n,i}} \left( K \left( \frac{\delta_{it}}{h_n} \right) - r_{n,i} \right) \right)^2 / (nr_{n,i}^2)$$

**Proof of Theorem 8** The proof of Theorem 8 closely follows that of Theorem 4, and so only a sketch is provided here. Let  $\lambda_i$ ,  $\hat{\lambda}_i$ , and  $\delta_{it}$  shorthand  $\lambda(w_i)$ ,  $\widehat{\lambda}(w_i)$ , and  $\delta(w_i, w_t)$  respectively. Then

$$\begin{aligned} (\hat{\lambda}_i r_{n,i} - r'_{n,i}) &= \frac{1}{n} \sum_{t=1}^n \left( \hat{u}_t K \left( \frac{\hat{\delta}_{it}}{h_n} \right) - r'_{n,i} \right) - \frac{r'_{n,i}}{r_{n,i}} \left( K \left( \frac{\hat{\delta}_{it}}{h_n} \right) - r_{n,i} \right) + rem_{n,i} \\ &= \frac{1}{n} \sum_{t=1}^n \left( u_t K \left( \frac{\delta_{it}}{h_n} \right) - r'_{n,i} \right) - \frac{r'_{n,i}}{r_{n,i}} \left( K \left( \frac{\delta_{it}}{h_n} \right) - r_{n,i} \right) + o_p \left( \sqrt{\frac{r_{n,i}}{n}} \right) + rem_{n,i} \end{aligned}$$

where  $r_{n,i} = E \left[ K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right]$ ,  $r'_{n,i} = E \left[ u_t K \left( \frac{\delta_{it}}{h_n} \right) | w_i \right]$ , and  $rem_{n,i}$  is an error that is stochastically small (i.e. the remainder from a first order Taylor approximation). See the proof of Theorem 3 for more details. The second line follows from the fact that  $|\hat{\beta} - \beta|$  and  $\max_{i \neq j} |\hat{\delta}_{ij} - \delta_{ij}|$  are both  $o_p \left( (nr_{n,i})^{-1/2} \right)$  and the smoothness conditions on  $K$  given in Assumption 8.

It follows from the Lindeberg Central Limit Theorem that  $\frac{\hat{\lambda}_i r_{n,i} - r'_{n,i}}{\sqrt{n V_{8,n,i}}} \rightarrow_d \mathcal{N}(0, 1)$  where

$$V_{8,n,i} = n^{-1} \sum_{t=1}^n \left( \left( u_t K \left( \frac{\delta_{it}}{h_n} \right) - r'_{n,i} \right) - \frac{r'_{n,i}}{r_{n,i}} \left( K \left( \frac{\delta_{it}}{h_n} \right) - r_{n,i} \right) \right)^2$$

and since  $b_{n,i} n / r_{n,i} \rightarrow_p 0$  for all  $i \in S$ ,  $\frac{\hat{\lambda}_i - \lambda_i}{\sqrt{n V_{8,n,i} / r_{n,i}^2}} \rightarrow_d \mathcal{N}(0, 1)$ . The claim then follows from the fact that the entries of  $\hat{\lambda}_S$  are all asymptotically independent.  $\square$