Identification and Estimation of Models with Endogenous Network Formation

Eric Auerbach*

November 11, 2016

JOB MARKET PAPER

Link to the current version here

Abstract

This paper studies a linear model in which the regressors and errors covary with drivers of link formation in a large network. Neither the endogenous relationship between the regressors and errors nor the distribution of network links are restricted parametrically. Instead, the model is identified by variation in the regressors unexplained by the distribution of network links. I first demonstrate that agents with similar columns of the squared adjacency matrix, the ijth entry of which contains the number of other agents linked to both agents i and j, necessarily have a similar distribution of network links. I then propose a semiparametric estimator based on matching pairs of agents with similar columns of the squared adjacency matrix. I find sufficient conditions for the estimator to be consistent and asymptotically normal, and provide a consistent estimator for its asymptotic variance. While this paper focuses on cases in which the network is represented by a binary, symmetric, and square adjacency matrix, I also discuss extensions to weighted, directed, bipartite, multiple, sampled, and higher-order networks.

Link to the online appendix here

*Department of Economics, UC Berkeley. E-mail: eric.auerbach@econ.berkeley.edu. I thank my advisors, James Powell and Bryan Graham for their advice and support. I also thank Aluma Dembo, Michael Jansson, Patrick Kline, Sheisha Kulkarni, Carl Nadler, Stephen Nei, Demian Pouzo, Mikkel Soelvsten and participants at the UC Berkeley Econometrics Seminar for helpful feedback.

1 Introduction

In many social networks, linked agents make similar decisions. One explanation for this phenomenon is peer effects, in which agents are influenced by or choose to imitate the behavior of their peers. Another is latent homophily, in which linked agents have underlying characteristics that generate correlated though otherwise unrelated behaviors. Distinguishing between peer effects and latent homophily matters because the former often suggests that a policy maker can efficiently influence mass behavior by manipulating only a small number of key agents or links.[1] However, recent work has questioned not only the existence of network peer effects, but the extent to which they can be identified in nonexperimental settings at all.[2]

This paper considers network peer effects as part of a broader study about the identification and estimation of models with endogenous network formation.[3] In this paper, I address two fundamental questions. First, when are models with endogenous networks identified? Second, how can data on network links be used to control for this sort of endogeneity in estimation?

I study these questions in the context of a linear model in which a correlation between the regressors and errors is caused by an omitted vector of unobserved social characteristics. I do not assume that the researcher has access to instrument or control variables for the endogenous regressors. Instead, relevant features of the social characteristics are to be inferred using variation in how agents link in a network. To do this, I consider a nonparametric model of link formation in which the probability that two agents link is some unknown function of their social characteristics. The model admits a basic random utility interpretation and is consistent with a number of network formation models from the literature, including Chandrasekhar and Jackson (2014), Graham (2014), Leung (2015), Ridder and Sheng (2015), and

[1]Recent examples include Ballester, Calvó-Armengol, and Zenou (2006), Christakis and Fowler (2007), Calvó-Armengol, Patacchini, and Zenou (2009), Banerjee, Chandrasekhar, Duflo, and Jackson (2013), and Elliott, Golub, and Jackson (2014)

[2]For instance, Shalizi and Thomas (2011), Carrell, Sacerdote, and West (2013), Angrist (2014), Jackson (2014), and Graham (2015)

[3]Endogeneity refers to models in which the regressors and errors are correlated. A network represents a collection of pairs of agents that are distinguished in some economically meaningful way (i.e, the pairs are "linked," "connected," "friends," etc.). Network endogeneity refers to models in which the correlation between the regressors and errors is explained by latent factors that influence link formation in a network.

Menzel (2015).

In recent work Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2014), Johnsson and Moon (2015), and Arduini, Patacchini, and Rainone (2015) all consider related models with endogenous networks. However, their models all impose parametric restrictions on the network formation model to identify and estimate the parameters of interest. As a result, the performance of their estimators generally depends on the accuracy of these assumptions which may potentially fail to capture the full heterogeneity in linking behavior underlying many real world networks.

The first contribution of this paper is to provide identification conditions that do not require parametric restrictions on the network model. The idea behind these conditions is familiar: the model is identified if conditional on the distribution of network links, the regressors and errors are uncorrelated and the distribution of the regressors is nondegenerate. A key feature of this paper is that it introduces new tools to formalize these conditions and make them straightforward to apply in practice.

For instance, I demonstrate that the linear peer effects model of Bramoullé, Djebbari, and Fortin (2009) is not generally identified when the network is endogenous. In particular, the nondegeneracy condition is violated because the explanatory variable of interest (an agent's expected peers' characteristics) is completely determined by the distribution of network links. Similar non-identification results are found in the related grouped peer effects literature (for instance, Manski 1993, Graham and Hahn 2005, Graham 2008), and I discuss how strategies from this literature might be used to restore identification in the network setting.

The second contribution of this paper is to propose a new matching procedure to estimate models with endogenous networks. Specifically, I propose matching pairs of agents with similar columns of the squared adjacency matrix, the ijth entry of which contains the number of other agents linked to both agents i and j.[4] The motivation for this procedure follows from a new result I derive in this setting that agents with similar columns of the squared adjacency matrix necessarily have a similar distribution of network links. The logic is related to recent

[4]Formally, the adjacency matrix of a network is a matrix with the number of rows and columns equal to the number of agents that contains a 1 in the ijth 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 ith column of this matrix.

arguments from the link prediction literature (for example, Bickel, Chen, and Levina 2011, Zhang, Levina, and Zhu 2015), though to my knowledge the results of this paper and its application to the study of network endogeneity are original.

The proposed estimator resembles other matching estimators from the literature (for instance, Powell 1987, Heckman, Ichimura, and Todd 1998, Abadie and Imbens 2006) and is similarly straightforward to implement and interpret. However, its large sample properties are nonstandard when compared to this literature for two reasons.

The first reason concerns the dimension of the matching variable. The above literature makes asymptotic approximations that require the density function of the matching variable to exist and be bounded away from zero. In this paper, the matching variable is a column vector of length equal to the sample size. Since the usual notion of a density function does not necessarily exist in this setting, these asymptotic approximations are generally inapplicable. I sidestep the issue by appealing to arguments from the functional nonparametrics literature (for example, Ferraty and Vieu 2006, Hong and Linton 2016) in which the density function is replaced by the more general notion of a small ball probability. I then adapt tools from the literature on dense graph limits (for instance, Lovász 2012) to characterize this probability and find sufficient conditions for consistency and asymptotic normality. As is common in the matching literature, the bias of my estimator is potentially large relative to its variance. Accurate inference requires a bias correction and I propose a variation on the jackknife technique proposed by Powell, Stock, and Stoker (1989).

The second reason this estimator is nonstandard is that even though the matching variable is generated in the sense that its entries are sample averages with variances on the order of the inverse of the sample size, this variation does not influence the asymptotic distribution of the proposed estimator. This result is unusual because it seemingly contradicts a developed literature on asymptotic variance formulas for semiparametric estimators (for instance, Newey 1994, Chen, Linton, and Van Keilegom 2003, Hahn and Ridder 2013). The intuition behind this result is that the average squared difference between two agents' matching variables estimates a particular measure of network distance between the agents. Evaluating the variance of my estimator does not require bounding the sampling variation of all of these estimated distances, but only those that correspond to pairs of matched agents.

Since the estimated distances between matched agents is small by construction, their means and variances must also be small, and under certain regularity conditions the total variation is small enough to be asymptotically negligible. As a result, the asymptotic variance of my estimator does not have the usual correction term for a first stage estimation error.

The matching logic extends to various nonlinear and nonparametric settings or to allow for weighted, directed, bipartite, multiple, sampled, or higher-order networks. I explore some of these extensions in an appendix to this paper, though formal results are left to future work. The method also has important limitations. The model and estimator generally require the network to be dense (the expected number of links is proportional to the square of the sample size) and that the network links are conditionally independent. Some sparsity can be accommodated by letting the link probabilities decrease with the sample size (as in Bickel and Chen 2009), and although the rate of convergence is likely to be affected, this may be unimportant if the total number of agents is large. The assumption of conditional link independence can also be weakened. For instance, it can be replaced with the conditional independence of some higher-order network event, such as the formation of cliques of a particular size, along the lines proposed by Chandrasekhar and Jackson (2014).

The structure of this paper is as follows. Section 2 introduces the model, identification conditions, and proposed estimator. Section 3 contains the main results of the paper. Section 3.2 provides the main identification results and section 3.3 the main asymptotic results: sufficient conditions for consistency and asymptotic normality. Section 4 provides simulation evidence and Section 5 concludes. Proofs of the various lemmas and theorems are collected in Appendix A and some extensions to the proposed model and estimator can be found in Appendix B. Appendices C and D contain additional context for the results. Appendix C illustrates the proposed matching strategy using three example parametric link distributions from the literature. Appendix D provides details about a behavioral interpretation for the model and estimator. Appendices B, C, and D have been collected in an online appendix, a link to which can be found on the title page of this paper.

2 Model and Estimator

2.1 Model

Let $\{y_i, x_i\}_{i=1}^n$ be an independent and identically distributed sequence of data for n agents with $y_i \in \mathbb{R}, x_i \in \mathbb{R}^k$ for some positive integer k, and D be an $n \times n$ stochastic binary adjacency matrix corresponding to an unlabelled, unweighted, and undirected random network between the n agents. The joint distribution of $\{y_i, x_i\}_{i=1}^n$ and D is determined by the following semiparametric model

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

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

in which $\{w_i\}_{i=1}^n$ is an independent and identically distributed sequence of unobserved social characteristics, λ and f are unknown Lebesgue measurable functions with the latter symmetric in its arguments, and $\{\eta_{ij}\}_{i,j=1}^n$ is a symmetric matrix of unobserved scalar disturbances with independent and identically distributed upper diagonal entries that are mutually independent of $\{x_i, w_i, \varepsilon_i\}_{i=1}^n$. I suppose for the sake of exposition that $E[\varepsilon_i|x_i, w_i] = 0$, although the main results of this paper will be derived under a weaker uncorrelatedness assumption. It is generally without loss to normalize the marginal distributions of w_i and η_{ij} to be standard uniform.

In this model, endogeneity takes the form of a dependence between x_i and the unobserved error $\lambda(w_i) + \varepsilon_i$ through w_i. 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 $f(w_i, w_j)$. Parametric examples of (2) in the network formation literature include Holland and Leinhardt (1981), Duijn, Snijders, and Zijlstra (2004), Krivitsky, Handcock, Raftery, and Hoff (2009), Dzemski (2014), Graham (2014) and Nadler (2016) (see section 3 of Graham 2015, for a review). Leung (2015), Ridder and Sheng (2015) and Menzel (2015) also consider network formation models with strategic interaction that imply equation (2) as a reduced form distribution of links. More details about a behavioral interpretation for this model can be found in Appendix D.

Example 1 (Network Peer Effects): Let y_i be student GPA, x_i be a vector of student characteristics (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 of student achievement to the network setting is

$$\begin{aligned} y_i &= x_i\beta + E[x_j|D_{ij} = 1, w_i]\rho_1 + E[y_j|D_{ij} = 1, w_i]\rho_2 + \lambda(w_i) + \varepsilon_i \\ D_{ij} &= \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}\mathbb{1}\{i \neq j\} \end{aligned}$$

in which $E[x_j|D_{ij} = 1, w_i]$ denotes the mean characteristics and $E[y_j|D_{ij} = 1, w_i]$ the mean GPA of agent i's friends, conditional on agent i's social characteristics w_i. Bramoullé, Djebbari, and Fortin (2009) consider a similar model in which the network is exogenous ($\lambda(w_i) = 0$) and Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2014), Johnsson and Moon (2015), and Arduini, Patacchini, and Rainone (2015) consider related models with additional parametric assumptions on λ or f.[5]

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. Ignoring for now that their outcome is binary,[6] I propose a semiparametric alternative

$$\begin{aligned} y_i &= x_i\beta + E[y_j|D_{ij} = 1, w_i]\rho + \lambda(w_i) + \varepsilon_i \\ D_{ij} &= \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}\mathbb{1}\{i \neq j\} \end{aligned}$$

In this linear example, $i = 1, ..., n$ indexes households with program participants, y_i is a measure of the intensity of participation (for example, the amount of money borrowed or

[5]The use of the expected peer outcomes $E[y_j|D_{ij} = 1, w_i]$ instead of their empirical counterparts $\sum_j y_j D_{ij} / \sum_j D_{ij}$ masks another endogeneity issue generated by having dependent variables on the right hand side of the outcome equation. Bramoullé, Djebbari, and Fortin (2009) resolve this issue by using functions of D and $\{x_i\}_{i=1}^n$ as instruments for $\sum_j y_j D_{ij} / \sum_j D_{ij}$. I ignore the complication here because the simultaneity issue is unrelated to the unobserved heterogeneity focus of this paper.

[6]In future work I plan to demonstrate how the results of this paper can be extended to certain nonlinear and nonparametric models along the lines of Manski (1987) and Honoré and Powell (1997).

the average time to repayment), x_i is a vector of observed household characteristics (caste, religion, wealth, etc.), $D_{ij} = 1$ if households i and j have a social connection, and w_i are characteristics that influence social network formation (for example, the physical location of the household). $\lambda(w_i)$, the probability that household i is informed about the program given their social characteristics, is a correction term for selection into the program due to heterogeneous information.

Example 3 (Job Mobility): Schmutte (2014) studies a bipartite labor market network[7] between workers and industry-occupations in which worker i and industry-occupation j are linked if worker i is observed working in industry-occupation j at some point in time. The author identifies several clusters of highly connected workers and industry-occupations in the labor market network and uses the clusters as proxy variables for unobserved worker and industry-occupation heterogeneity in a linear model of labor market earnings. I characterize the relationship between this unobserved heterogeneity and the observed network clusters using the network formation model of this paper and recast the model as a model with an endogenous network along the lines of

$$\log(y_{it}) = x_{it}\beta + \theta(\phi_1(w_i)) + \psi(\phi_2(w_{j(i,t)})) + \varepsilon_{it}$$
$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(\phi_1(w_i), \phi_2(w_j))\}$$

in which y_{it} is the earnings of worker i in time period t, x_{it} are worker characteristics (age, gender, race, education, etc.), $j(i,t)$ indexes the industry-occupation of worker i in period t, w_i and $w_{j(i,t)}$ denote unobserved worker and industry-occupation characteristics (for instance, ability or productivity), and ϕ_1 and ϕ_2 map worker and industry-occupation characteristics to the network clusters.

Example 4 (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 characteristics, and a vector of network

[7] A bipartite network is a network in which the agents can be sorted into two groups such that two agents in the same group never form a link. In Appendix B, I describe how one might extend the methods of this paper to the bipartite setting.

statistics derived from a coauthorship network (in which two researchers are linked if they have previously been coauthors) including agent degree, eigenvector centrality, betweeness centrality, etc. The authors experiment with several different models of productivity, including various combinations of network statistics. An alternative approach treats the unknown combination of network statistics as unobserved network heterogeneity

$$y_i = x_i\beta + \lambda(w_i) + \varepsilon_i$$
$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}\mathbb{1}\{i \neq j\}$$

in which x_i is a vector of researcher characteristics, w_i characterizes the academic community of researcher i (for instance, a field of study) and $\lambda(w_i)$ represents heterogeneity in research productivity due to this community. A key feature of this model is that the estimation of β (which measures the impact of researcher characteristics on publication quality) does not require the researcher to correctly identify the relevant features of the network that make up $\lambda(w_i)$.

In many cases, the function λ (or the functions ϕ and ψ in Example 3) are not nuisance parameters, but also objects of interest in the analysis. In future work I plan to demonstrate how the tools of this paper can be extended to estimate and conduct inference about features of these parameters as well.

2.2 Estimator

Estimation is complicated by the fact that the social characteristics $\{w_i\}_{i=1}^n$ are unobserved. If the social characteristics were observed, (1) corresponds to the partially linear regression of Engle, Granger, Rice, and Weiss (1986), and many tools exist to estimate β (for example, Chamberlain 1986, Powell 1987, Newey 1988, Robinson 1988). If the social characteristics were unobserved but identified by the distribution of D, one can 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 particular approach is taken by Arduini, Patacchini, and Rainone (2015) and Johnsson and Moon (2015).

However, in many empirical applications the social characteristics are neither observed nor identified by the distribution of D. This paper demonstrates that identifying, estimating, and conducting inference about β is still possible without imposing parametric restrictions on either f or λ by matching pairs of agents with similar link distributions. The result is motivated by two key insights.

One insight concerns the identification of β, which holds if two conditions are satisfied. The first condition is that $\lambda(w_i)$ depends on w_i only through the schedule of linking probabilities $f(w_i, \cdot) : [0,1] \to [0,1]$. The second is that there is excess variation in the distribution of x_i that is not explained by $f(w_i, \cdot)$. Formally, consider the pseudometric on the space of social characteristics defined by

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

The linking function $f(u, \cdot)$ gives the collection of probabilities that an agent with social characteristics u links with the other agents in the network as indexed by their social characteristics in $[0, 1]$. The pseudometric $d(u, v)$ is then the integrated squared difference in the linking functions of agents with social characteristics u and v. The identification conditions are then that β 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. These conditions are similar to the usual identification conditions for linear models with unobserved heterogeneity in the panel data setting (see for example Wooldridge 2010, Chapter 10): it is the notion of the network distance measure d used to partial out the endogenous variation that is different.

The logic behind the first identification condition is that d 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 the social characteristics of agents i and j. They 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-level statistic of D. 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 link distributions 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 trivially. As a consequence, this first identification condition is more general than those imposed by Goldsmith-Pinkham and Imbens (2013), Hsieh and Lee (2014), Johnsson and Moon (2015), and Arduini, Patacchini, and Rainone (2015).

A sufficient condition for $E[(x_i - x_j)'(\lambda(w_i) - \lambda(w_j))|d(w_i, w_j) = 0] = 0$ is for $\lambda(w_i)$ to be continuous in d (i.e, if $\{w_t\}_{t=1}^{\infty}$ is such that $d(w_i, w_t) \to 0$ then $|\lambda(w_i) - \lambda(w_t)| \to 0$). An advantage of the more general condition is that in some cases there is variation in $\lambda(w_i)$ that is not continuous in d but is uncorrelated with x_i. For instance, suppose the omitted function is an indicator for whether or not an agent is linked to agent 1, or $\lambda(w_i) = D_{i1}$. Then $\lambda(w_i)$ is not continuous with respect to d, but $D_{i1} = E[D_{i1}|w_i] + (D_{i1} - E[D_{i1}|w_i])$ in which the first summand is continuous with respect to d and the second is uncorrelated with x_i.

The logic behind the second identification condition is that matching agents with similar link distributions only identifies β if there is excess variation in the distribution of x_i not explained by the linking function $f(w_i, \cdot)$. Otherwise 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 the magnitude of β. One example of this is when x_i contains agent-level statistics of the adjacency matrix. Another is the case of a linear-in-means network peer effects model. I discuss these cases in more detail below.

The second insight is that the average squared difference in the ith and jth 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 δ on $[0,1]^2$ 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 ith and jth columns of the squared adjacency matrix

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

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 such network moments and the distribution of links have also been exploited in arguments by Lovász and Szegedy (2007; 2010), Bickel, Chen, and Levina (2011), Lovász (2012), and Zhang, Levina, and Zhu (2015).

The two insights indicate that when the ith and jth columns of the squared adjacency matrix are similar and the identification conditions for β hold then $(y_i - y_j)$ and $(x_i - x_j)\beta + (\varepsilon_i - \varepsilon_j)$ are approximately equal. This result is limited in the sense that it is insufficient to estimate λ by a series approximation as in Newey (1988) and Ai and Chen (2003) because w_i is not necessarily identified. However, one can recover β by matching pairs of agents with d-similar social characteristics. This paper demonstrates that under certain regularity conditions β is consistently estimated by a pairwise difference estimator

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

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

The estimator has a form similar to established pairwise difference estimators from the literature (in particular Ahn and Powell 1993). However, the large sample properties of $\hat{\beta}$ are not typical of this literature. For example, unless the researcher is willing to put substantial structure on the unknown linking function f, the distribution of $\hat{\delta}_{ij}$ can be difficult to characterize near 0, complicating the usual balancing of large sample bias and variance. The problem is related to the small ball problem in the functional nonparametrics literature (see for instance Masry 2005, Ferraty and Vieu 2006, Hong and Linton 2016) and can severely amplify the usual curse of dimensionality. Of particular concern is the possibility that the quantity of matches shrinks to zero quicker than the averages in (4) converge, though in the proofs of this paper I demonstrate how the structure of the network model sufficiently mitigates this problem such that under certain regularity conditions the proposed estimator is consistent and asymptotically normal.

The following translates the identification conditions for the examples posed above

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

$$y_i = x_i\beta + E[x_j|D_{ij} = 1, w_i]\rho_1 + E[y_j|D_{ij} = 1, w_i]\rho_2 + \lambda(w_i) + \varepsilon_i$$
$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}\mathbb{1}\{i \neq j\}$$

the parameter β is identified if there is variation in x_i that is unrelated to the distribution of network links $f(w_i, \cdot)$. However, the parameters ρ_1 and ρ_2 are not identified since $E[x_j|D_{ij} = 1, w_i] = E[x_j D_{ij}|w_i]/E[D_{ij}|w_i]$ is a fixed function of w_i that is indistinguishable from $\lambda(w_i)$. In particular, the model violates the nondegeneracy identification condition since

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

and $d(w_i, w_{i'}) = ||f(w_i, \cdot) - f(w_{i'}, \cdot)||_2 = 0$ implies

$$E[(E[x_j|D_{ij} = 1, w_i] - E[x_j|D_{i'j} = 1, w_{i'}])^2 \, |d(w_i, w_{i'}) = 0] = 0$$

The same logic applies for the variable $E[y_j|D_{ij} = 1, w_i]$.

It is helpful to contrast the nonidentification result with the setting of Goldsmith-Pinkham and Imbens (2013). They study a model along the lines of

$$y_i = x_i\beta + E[x_j|D_{ij} = 1, w_i, Z_{ij}]\rho_1 + E[y_j|D_{ij} = 1, w_i, Z_{ij}]\rho_2 + w_i\rho_3 + \varepsilon_i$$
$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq |w_i - w_j|\gamma_1 + Z_{ij}\gamma_2\}\mathbb{1}\{i \neq j\}$$

Their model is identified by two restrictions. The first is the functional form restriction on the network heterogeneity $\lambda(w_i) = w_i\rho_3$. The second is the introduction of exogenous link covariates Z_{ij}, assumed to be independent of w_i and w_j.[8]

[8] It is also possible to incorporate link covariates into the framework of this paper by replacing equation (2) with $D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j, Z_{ij})\}$. In the appendix, I demonstrate how the estimator of this paper can be extended to models with link covariates by matching on conditional codegree vectors, although a formal study of the large sample properties of such an estimator is left to future work.

Example 2 (Information Diffusion) In the microfinance program participation model

$$y_i = x_i\beta + E[y_j|D_{ij} = 1, w_i]\rho + \lambda(w_i) + \varepsilon_i$$
$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}\mathbb{1}\{i \neq j\}$$

the parameter ρ is not identified following previous arguments. The parameter β is identified if two households with the same distribution of links have the same probability of being informed about the program and a household's covariates are not completely determined by their distribution of links. The first condition is satisfied in the information diffusion model of Banerjee, Chandrasekhar, Duflo, and Jackson (2013). The second condition may be violated if households only link to other households of the same religion or caste, which does not seem to be the case in this setting (see Jackson 2014, for a discussion). A key feature of the model and estimator proposed in this paper is that they do not require many-networks asymptotics.

Example 3 (Job Mobility): In the labor market earnings model

$$\log(y_{it}) = x_{it}\beta + \theta(\phi_1(w_i)) + \psi(\phi_2(w_{j(i,t)})) + \varepsilon_{it}$$
$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(\phi_1(w_i), \phi_2(w_j))\}$$

β is identified if agents in different network clusters have a different distribution of network links and there is excess variation in the worker and industry-occupation covariates that are not explained by the network clusters. The first is satisfied by construction since Schmutte (2014) defines the clusters as collections workers and industry-occupations with few links between clusters. The second is satisfied if the covariates have overlapping support across clusters, which is the case in this particular setting.

Example 4 (Research Productivity): In the research productivity model

$$y_i = x_i\beta + \lambda(w_i) + \varepsilon_i$$
$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\}\mathbb{1}\{i \neq j\}$$

β is identified if there is excess variation in the covariates that is not explained by the network links. This may not be satisfied if researchers only coauthor with other researchers with similar publication histories. This does not seem to be the case empirically.

3 Main Results

3.1 Terminology and Notation

This section details additional constructions required for the lemmas, theorems, and proofs. I define agent i's network type to be the projection of the link function f onto his social characteristics: $f_{w_i}(\cdot) := f(w_i, \cdot) : [0,1] \to [0,1]$. 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. As suggested by the notation of the previous section, $d(w_i, w_j) = ||f_{w_i} - f_{w_j}||_2$ is the L^2 metric on the space of network types.

I require two network theoretic constructions: (average) agent degrees and (average) agent-pair codegrees, as well as their population analogs. The degree of agent i is the fraction of other agents linked to agent i in D, or $(n-1)^{-1}\sum_{t\neq i} D_{it}$. Under (2), that $(n-1)^{-1}\sum_{t\neq i} D_{it} \to_{a.s.} \int f_{w_i}(\tau)d\tau$ follows from the usual strong law of large numbers. I refer to $\int f_{w_i}(\tau)d\tau$ as agent i's population degree.

Similarly, for $i \neq j$ the codegree of agent pair (i,j) is the fraction of other agents linked to both agent i and agent j, or $(n-2)^{-1}\sum_{t\neq i,j} D_{it}D_{jt}$. Again, under (2), $(n-2)^{-1}\sum_{t\neq i,j} D_{it}D_{jt} \to_{a.s.} \int f_{w_i}(\tau)f_{w_j}(\tau)d\tau = \langle f_{w_i}, f_{w_j}\rangle_{L^2}$. I define $\hat{p}_{ij} := (n-2)^{-1}\sum_{t\neq i,j} D_{it}D_{jt}$ and $p(w_i, w_j) := \int f_{w_i}(\tau)f_{w_j}(\tau)d\tau$ and refer to $p(w_i, w_j)$ as the population codegree of agents i and j. I emphasize that $p(w_i, w_i)$ refers to the population codegree of two distinct agents with social characteristics equal to w_i and *not* to the limiting degree of agent i. That is $p(w_i, w_i) := \int f_{w_i}(\tau)^2 d\tau = ||f_{w_i}||_2^2 \neq \int f_{w_i}(\tau)d\tau$.

Notice that p also defines a link function, in which $p(w_i, w_j)$ gives the probability that agents i and j are both linked to a third agent, 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 $p_{w_i}(\cdot) := p(w_i, \cdot) : [0,1] \to [0,1]$ as agent i's codegree type. I also take codegree types to be elements of $L^2([0,1])$. I refer to the pseudometric on $[0,1]$ induced by L^2-differences in codegree types with δ, so that

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

for any pair of social characteristics u and v. Under (2), my Lemma 1 demonstrates that the root average squared difference in the ith and jth columns of the squared adjacency matrix (given by (3)) provides a uniformly consistent estimator for $\delta(w_i, w_j)$ over $[0,1]^2$.

I 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 $E[Z_{ij}|\, ||f_{w_i} - f_{w_j}||_2 = x]$ refers to the conditional expectation

$$\lim_{h \to 0} E[Z_{ij}|\, (w_i, w_j) \in \{(u,v) \in [0,1]^2 : x \leq ||f_u - f_v||_2 \leq x + h\}]$$

and $E\left[Z_i | f_{w_i} = f\right]$ refers to the conditional expectation

$$\lim_{h \to 0} E\left[Z_i |\, w_i \in \{w \in [0,1] : ||f_w - f||_2 \leq h\}\right]$$

Though f_{w_i} is a random function, these conditional expectations implicitly refer to the measure induced by the random variable w_i. Conditional means with respect to the agent codegree differences or types are defined in an analogous way.

Let $u_i = \lambda(w_i) + \varepsilon_i$. I use the functional $\lambda(f)$ to denote $E[u_i | f_{w_i} = f]$ and ν_i for the associated residual $u_i - \lambda(f_{w_i})$. This allows me to rewrite the model (equations (1) and (2)) in a way that emphasizes the identification and estimation strategy described in the previous section.

$$y_i = x_i \beta + \lambda(f_{w_i}) + \nu_i \tag{5}$$

$$D_{ij} = \mathbb{1}\{\eta_{ij} \leq f(w_i, w_j)\} \tag{6}$$

3.2 Model Identification

This section provides conditions for agents with similar network types but different regressors to identify β.

Assumption 1: The random sequence $\{x_i, \nu_i, w_i\}_{i=1}^n$ is independent and identically distributed with entries mutually independent of $\{\eta_{ij}\}_{j>i=1}^n$, a symmetric random array with independent and identically distributed entries above the diagonal. The variables w_i and η_{ij} have standard uniform marginals. The conditional distributions of $\{y_i\}_{i=1}^n$ and D are given by equations (5) and (6) respectively. The functions $\lambda : [0,1] \to \mathbb{R}$ and $f : [0,1]^2 \to [0,1]$ are Lebesgue-measurable with the latter symmetric in its arguments.

Assumption 1 is a restatement of the discussed model and is included primarily as a reference. Since the marginal distributions of w_i and η_{ij} are not separately identified from f, the assumption of standard uniform marginals is without loss of generality (see Bickel and Chen 2009, Orbanz and Roy 2015, for a discussion).

Assumption 2: The variables x_i and u_i both have finite sixth moments with $E[(x_i - x_j)'(u_i - u_j)| \, ||f_{w_i} - f_{w_j}||_2 = 0] = 0$.

The second part of Assumption 2 is satisfied if x_i and u_i are uncorrelated conditional on f_{w_i}.

Assumption 3: The conditional covariance matrix $\Gamma_0 = E\left[(x_i - x_j)'\,(x_i - x_j)\,|\,||f_{w_i} - f_{w_j}||_2 = 0\right]$ is positive definite.

Assumption 3 states that there is some independent variation in each of the regressors that is not explained by the network types. Section 2 explores cases when it may not be satisfied, for example when the regressors include functions of the adjacency matrix. The assumption can be weakened in cases when the researcher has some additional information about the network formation process (for example, exogenous link covariates) or structure on the endogenous covariation in equation (5).

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

Theorem 1 demonstrates that β is identified from the joint distribution of (y_i, x_i, f_{w_i}). The fact that the network types f_{w_i} are in turn identified by the distribution of the adjacency matrix D is shown in the following section.

3.3 Model Estimation

This section characterizes the large sample properties of $\hat{\beta}$. The first part provides sufficient conditions for consistency. The second part provides sufficient conditions for the limiting distribution to be normal. Accurate inference may require a bias correction and the third part demonstrates how a variation on the jackknife method proposed by Powell, Stock, and Stoker (1989) can be used for this purpose. The fourth part provides a consistent estimator for the asymptotic variance.

3.3.1 Consistency

Consistency of $\hat{\beta}$ requires an additional continuity condition on the conditional expectation functions from Assumptions 2 and 3, and restrictions on the bandwidth sequence and kernel density function.

Assumption 4: The conditional expectation functions satisfy
$\lim_{h\to 0} E[(x_i - x_j)'(u_i - u_j)| \, ||f_{w_i} - f_{w_j}||_2 = h] = 0$ and
$\lim_{h\to 0} E[(x_i - x_j)'(x_i - x_j)| \, ||f_{w_i} - f_{w_j}||_2 = h] = \Gamma_0$.

Assumption 4 is satisfied if Assumptions 2 and 3 hold and the conditional expectation functionals $E[x_i' u_i | f_{w_i}]$ and $E[x_i' x_i | f_{w_i}]$ as defined in Section 3.1 are continuous with respect to f_{w_i} in the L^2-sense. This condition might not be satisfied if the network is sparse, because f_{w_i} may be uniformly close to zero so that small variations in f_{w_i} correspond to large variations in x_i and u_i.[9] In the appendix I discuss a number of ways in which the model and estimator can be altered to mitigate this problem, for example, by including observable link covariates.

[9]The problem is not unique to the sparse case. If the network is very dense so that f is uniformly close to 1, a similar problem occurs. Thus it is not sparsity per se that is a problem for the model of this paper, but situations in which the relative amount of information that the network types contain about the covariation between x_i and u_i is small.

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

The first two restrictions on the bandwidth sequence are standard. The third condition, that $nr_n \to \infty$ is less so. This condition is required to ensure that the number of matches used to estimate $\hat{\beta}$ is increasing 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 average number of agent pairs with similar codegree types would then be on the order of nh_n^d, which increases with n if the second bandwidth condition were changed to $n^{1-\gamma}h_n^d \to \infty$. However, 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 this third bandwidth condition is required.

The conditions on the kernel density function K are satisfied by a type-II kernel density function (examples include the Epanechnikov, Biweight, and Bartlett kernels). It is possible to extend this proof to include type-I kernel density functions (for example, the uniform kernel), although kernels supported on all of $\mathbb{R}$ (for example the Gaussian kernel) may potentially cause problems in this setting (see Hong and Linton 2016, for a discussion).

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 some 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 the codegree differences $\{||p_{w_i} - p_{w_j}||_2\}_{i\neq j}$. Second, agent-pairs with small codegree differences have small network differences. These results are stated in Lemmas 1 and 2 respectively.

Lemma 1: Suppose Assumptions 1 and 5 hold. Then

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

in which γ refers to the exponent from Assumption 5.

Lemma 1 demonstrates that the collection of $\binom{n}{2}$ empirical codegree differences observed by the researcher converges uniformly to their population analogs at a rate slightly slower than $n^{-1/2}$ (since h_n can be taken to be arbitrarily close to $n^{-1/2}$ by taking γ close to 0). 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 matching strategy of this paper. 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,[10] which suggests that differences in codegree types provide a coarser notion of network distance than differences in network types. Nevertheless, pairs of agents with similar codegree types have similar network types with high probability.

The lemma is related to Theorem 13.27 of Lovász (2012), which demonstrates that $||p_{w_i} - p_{w_j}||_2 = 0$ implies $||f_{w_i} - f_{w_j}||_2 = 0$ when f is continuous. The logic of his result is sketched below.

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

Essentially, the result follows from the fact 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 τ. In particular, the difference is

[10] To see this, note $||p_{w_i} - p_{w_j}||_2^2 = \int \left(\int f(t,s) \left(f(w_i, s) - f(w_j, s) \right) ds \right)^2 dt \leq \int \left(\int \left(f(t,s) \left(f(w_i, s) - f(w_j, s) \right) \right)^2 ds \right) dt \leq \int \left(f(w_i, s) - f(w_j, s) \right)^2 ds = ||f_{w_i} - f_{w_j}||_2^2$, where the first inequality is due to Jensen and the second due to the fact that f is bounded between 0 and 1.

uncorrelated with both f_{w_i} and f_{w_j}, the network types of agents i and j. However, this can only be the case if the network types of i and 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 $\int f(\tau, s)\left(f(w_i, s) - f(w_j, s)\right) ds = 0$ for exactly all τ, which is not guaranteed by the condition $||p_{w_i} - p_{w_j}||_2^2 \leq \epsilon$ for any $\epsilon > 0$. Despite this, the proof of Lemma 2 demonstrates that the two notions of distance are similar in enough places on $[0, 1]^2$ that matching agents with similar codegree types is sufficient to partial out $\lambda(f_{w_i})$ in equation (5) and consistently estimate β.

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

Theorem 2 is almost a direct consequence of Lemmas 1 and 2, several applications of the continuous mapping theorem, and Lemma 3.1 from Powell, Stock, and Stoker (1989).

3.3.2 Asymptotic Normality

I provide two asymptotic normality results. The first result concerns the case when the support of the network and codegree types is finite, so 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 $P(0 < ||f_{w_i} - f_{w_j}||_2 < \epsilon) = P(0 < ||p_{w_i} - p_{w_j}||_2 < \epsilon) = 0$.

Theorem 3: Suppose Assumptions 1-5 hold. Further suppose the support of f_{w_i} is finite. Then

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

where $V_{3,n} = \Gamma_0^{-1}\Omega_0\Gamma_0^{-1} \times s/n$, Γ_0 is as defined in Assumption 3, I_k is the $k \times k$ identity matrix, and

$$s = P(||p_{w_i} - p_{w_j}||_2 = 0, ||p_{w_i} - p_{w_k}||_2 = 0)/P(||p_{w_i} - p_{w_j}||_2 = 0)^2$$
$$\Omega_0 = E\left[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k)|\ ||p_{w_i} - p_{w_j}||_2 = 0, ||p_{w_i} - p_{w_k}||_2 = 0\right]$$

When the support of network and codegree types is finite, pairs of agents with similar codegree types have identical network types with high probability, and so the proof of Theorem 3 follows from Assumptions 1-5, Lemmas 1 and 2, and standard arguments. I include this theorem for three reasons. First, it adds to a literature noting that the adverse effects of unobserved heterogeneity can be mild when the support of this variation is finite (for example Hahn and Moon 2010, Bonhomme and Manresa 2015). Second, the assumption of discrete heterogeneity is not uncommon in empirical work (for instance Schmutte 2014, Bonhomme, Lamadon, and Manresa 2015). Third, it provides an easy to interpret condition such that $\hat{\beta}$ is consistent and asymptotically normal at the $\sqrt{n}$-rate.

The second result concerns the more general case when the support of f_{w_i} is not necessarily finite. In this case, the proof of asymptotic normality requires additional structure on f and the conditional expectations from Assumption 4, which is given in the following Assumptions 6 and 7. Assumption 8 modifies the bandwidth sequence accordingly.

Assumption 6: There exists an integer K and a partition of $[0, 1)$ into K equally spaced, adjacent, and non-intersecting intervals $\cup_{t=1}^{K}[x_t^1, x_t^2)$ with $x_1^1 = 0$ and $x_K^2 = 1$ such that for any $t \in \{1, ..., K\}$ and almost every $x, y \in [x_t^1, x_t^2)$ and $s \in [0, 1]$, $|f(x, s) - f(y, s)|$ $\leq C_6|x - y|^\alpha$, for some $C_6 \geq 0$ and $\alpha > 0$.

Assumption 6 imposes 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 some order. The partition allows for discrete jumps of the link function as to include discrete models such as the stochastic blockmodel (see Appendix C for a definition and discussion) as a special case. The restriction that the partition is uniformly sized is without loss, and the results can also be extended to let $K_n \to 0$ slowly with n. This corresponds to a stochastic blockmodel with a growing number of blocks as in Wolfe and Olhede (2013). A similar condition is used by Zhang, Levina, and Zhu (2015).

Assumption 7: The conditional expectation
$E[(x_i - x_j)'(u_i - u_j)| \, ||f_{w_i} - f_{w_j}||_2 = h] \leq C_7 h^{\zeta}$ for some C_7, $\zeta > 0$ and all h in a neighborhood to the right of 0.

Assumption 7 strengthens the first condition of Assumption 4 so that the slope of the conditional expectaton $E[(x_i - x_j)'(u_i - u_j)| \, ||f_{w_i} - f_{w_j}||_2]$ is bounded by a fractional polynomial to the right of 0.

Assumption 8: The bandwidth sequence $h_n = C_8 \times n^{-\rho}$ for $\rho \in \left(\frac{\alpha}{4+8\alpha}, \frac{\alpha}{2+4\alpha}\right)$ and some $C_8 > 0$. $K(\sqrt{u})$ is supported, bounded, and twice differentiable on $[0, 1]$, and strictly positive on $[0, 1)$.

The rate of convergence of the bandwidth sequence depends on the exponent from Assumption 6. When $\alpha = 1$ this bandwidth choice is approximately on the order of magnitude considered by Ahn and Powell (1993). The proof of Theorem 4 is simplified by requiring $K(\sqrt{u})$ to be twice differentiable at 0 and all of the kernel density functions in the discussion of Assumption 5 satisfy this additional condition.

The second asymptotic normality proof uses Assumption 6 to strengthen Lemma 2 in the following way.

Lemma 3: Suppose Assumptions 1 and 6 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_6 and α are the constants from Assumption 6.

Theorem 4: Suppose Assumptions 1-4 and 6-8 hold. Further suppose $\alpha \times \zeta > 1/2$. Then

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

where $V_{4,n} = \Gamma_0^{-1} \Omega_n \Gamma_0^{-1}/n$, Γ_0 is as defined in Assumption 3, r_n is as defined in Assumption

5, and I_k is the $k \times k$ identity matrix, and

$$\beta_{h_n} = \beta + (\Gamma_0)^{-1} E\left[(x_i - x_j)'(u_i - u_j) K\left(\frac{||p_i - p_j||_2}{h_n}\right)\right] / (2r_n)$$

$$\Omega_n = E\left[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k) K\left(\frac{||p_{w_i} - p_{w_j}||_2}{h_n}\right) K\left(\frac{||p_{w_i} - p_{w_k}||_2}{h_n}\right)\right] / \left(r_n^2\right)$$

The statement of Theorem 4 warrants three 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 $r_n^{-2} E\left[K\left(\frac{||p_{w_i}-p_{w_j}||_2}{h_n}\right) K\left(\frac{||p_{w_i}-p_{w_k}||_2}{h_n}\right)\right]$ can potentially diverge with n. When this variance converges to a limit, then $\left(\hat{\beta} - \beta_{h_n}\right)$ is asymptotically normal with variance $\Gamma_0 \Omega_0 \Gamma_0 \times \sigma / n$ where $\sigma = \lim_{n\to\infty} r_n^{-2} E\left[K\left(\frac{||p_{w_i}-p_{w_j}||_2}{h_n}\right) K\left(\frac{||p_{w_i}-p_{w_k}||_2}{h_n}\right)\right]$ and Ω_0 is as defined in Theorem 3. Even when this variance diverges, Assumptions 6-8 and Lemma 3 ensure that the rate of convergence for $V_{4,n}$ is on the order of at least $n^{-1/2}$ and is close to n^{-1} when α is close to 1. In the appendix, I propose an adaptive bandwidth procedure that requires each agent to belong to the same number of matches, which normalizes $r_n^{-2} E\left[K\left(\frac{||p_{w_i}-p_{w_j}||_2}{h_n}\right) K\left(\frac{||p_{w_i}-p_{w_k}||_2}{h_n}\right)\right] = 1$. Though this choice of bandwidth potentially inflates the bias of the estimator relative to $\hat{\beta}$, simulation evidence suggests that this inflation is often small relative to the reduction in variance.

Second, the estimator has an oracle property in the sense that the estimation error of $\hat{\delta}_{ij}$ around $\delta(w_i, w_j)$ is asymptotically negligible, so that the researcher may conduct inference as though the codegree differences between agents were known. The intuition in this case is that conditional on (w_i, w_j), the asymptotic variance of $\sqrt{n}(\hat{\delta}_{ij} - \delta(w_i, w_j))$ is bounded from above by $d(w_i, w_j)$. If $\hat{\delta}_{ij}$ is close to zero and the sample size is large then Lemmas 2 and 3 imply that $d(w_i, w_j)$ is close to zero so that the variance of $\sqrt{n}(\hat{\delta}_{ij} - \delta(w_i, w_j))$ is also close to zero. When $\alpha \times \zeta > 1/2$ this variance is sufficiently small as to not influence the asymptotic distribution of $\hat{\beta}$. This is distinct from the results of Ahn and Powell (1993). Their approach would roughly correspond to matching agents based on $\delta(\hat{w}_i, \hat{w}_j)$, where $\hat{w}_i$ is a consistent estimator for w_i and δ is known. In their case, the variation of $\hat{w}_i$ around w_i and $\hat{w}_j$ around w_j is unrelated to $\delta(w_i, w_j)$, so that the variance of $\sqrt{n}(\delta(\hat{w}_i, \hat{w}_j) - \delta(w_i, w_j))$ is not small. As a result, this variation does inflate the asymptotic variance of their estimator.

Third, the asymptotic distribution $\hat{\beta}$ is not centered at β, but at the pseudo-truth β_{h_n}. Though Theorem 2 implies that β_{h_n} converges to β, the rate of convergence can be slow depending on the size of α and ζ. This problem is common with matching estimators, although it is exacerbated here by the relatively weak relationship between the codegree and network distances as described by Lemma 3. In particular, Assumptions 6-8 and Lemma 3 only imply that $|\beta_{h_n} - \beta| = O_p\left(n^{\frac{-\zeta\alpha^2}{2(1+2\alpha)^2}}\right)$ which can imply a large worst-case scenario bias on the order of $n^{-1/36}$.

3.3.3 Bias Correction

Inferences about β based on the asymptotic distribution provided by Theorem 4 will only be valid if $V_{4,n}^{-1/2}\left(\beta_{h_n} - \beta\right) = o_p(1)$. Otherwise, accurate inference requires a bias correction. The technique proposed in this paper requires an additional smoothness condition.

Assumption 9: The pseudo-truth function β_h satisfies $\beta_h = \sum_{l=1}^{L} C_l h^{l/\theta} + O\left(h^{(L+1)/\theta}\right)$ for some positive integer $L > ((1+2\alpha)\theta - \alpha)/\alpha$, k-dimensional constants $C_1, C_2, ..., C_L$, $\theta > 0$, and h in a fixed open neighborhood to the right of 0.

Assumption 9 essentially requires that the asymptotic bias from Theorem 4 is sufficiently smooth with respect to the bandwidth choice.

I propose the following jackknife bias corrected estimator $\bar{\beta}_L$. For an arbitrary sequence of distinct positive numbers $\{c_1, c_2, ..., c_L\}$ 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$\{a_1, a_2, ...a_L\}$ satisfies

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

Theorem 5: Suppose Assumptions 1-4 and 6-9 hold, and $L > ((1+2\alpha)\theta - \alpha)/\alpha$. Then

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

where $V_{5,n} = \sum_{l_1=1}^{L}\sum_{l_2=1}^{L} a_{l_1}a_{l_2}\Gamma_0^{-1}\Omega_{n,l_1l_2}\Gamma_0^{-1}/n$, Γ_0 is as defined in Assumption 3, $r_{n,l} = E\left[K\left(\frac{||p_{w_i}-p_{w_j}||_2}{c_l h_n}\right)\right]$, I_k is the $k \times k$ identity matrix, and

$$\Omega_{n,l_1l_2} = E\left[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k)K\left(\frac{||p_{w_i} - p_{w_j}||_2}{c_{l_1}h_n}\right)K\left(\frac{||p_{w_i} - p_{w_k}||_2}{c_{l_2}h_n}\right)\right] / \left(r_{n,l_1}r_{n,l_2}\right)$$

3.3.4 Variance Estimation

The asymptotic variances from Theorems 3-5 can be consistently estimated using the sample analogs of Γ_0 and Ω_{n,l_1l_2}. That is for $\hat{u}_i = y_i - \hat{\beta}x_i$,

$$\hat{\Gamma}_h = \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left(x_i - x_j\right)'\left(x_i - x_j\right)K\left(\frac{||\hat{p}_i - \hat{p}_j||_2}{h}\right)$$

and $\hat{\Omega}_{h_1,h_2} =$

$$\binom{n}{3}^{-2}\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1}\sum_{k=j+1}^{n}\left(x_i - x_j\right)'\left(x_i - x_k\right)\left(\hat{u}_i - \hat{u}_j\right)\left(\hat{u}_i - \hat{u}_k\right)K\left(\frac{||\hat{p}_i - \hat{p}_j||_2}{h_1}\right)K\left(\frac{||\hat{p}_i - \hat{p}_k||_2}{h_2}\right)$$

then

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

A corollary to Theorem 6 is that $\hat{\Gamma}_{h_n}^{-1}\hat{\Omega}_{h_n,h_n}\hat{\Gamma}_{h_n}^{-1}$ also consistently estimates $nV_{3,n}$ under the hypothesis of Theorem 3. These statistics can be used to build confidence intervals or test hypotheses about β under the relevant assumptions in the usual way. Asymptotic theory has little to say about the actual choices of bandwidths and constants used in the construction of the estimators in this section. The setting potentially allows for choices based on cross validation which I leave to future work.

4 Simulations

This section presents simulation evidence for three types of network formation models: a stochastic blockmodel, a beta model, and a homophily model. To simplify the exposition, a detailed explanation of the models is deferred to Appendix C. 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 covariance given by the identity matrix. I also draw a random symmetric 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 $D = \mathbb{1}\{\eta_{ij} \leq f\left(\Phi(\omega_i), \Phi(\omega_j)\right)\}$ where Φ 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 a beta model and homophily model respectively 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 refer to 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 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,[11] and average peer covariates $\sum_{j=1}^n D_{ij}x_j / \sum_{j=1}^n D_{ij}$. Z_2 denotes the stacked vector $\{x_i, c_i\}_{i=1}^n$.

[11] Agent i's eigenvector centrality statistics refers to the ith entry of the eigenvector of D associated with the largest eigenvalue.

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 (7), and $\hat{\beta}_6$ is the pairwise difference estimator with an adaptive bandwidth but without bias correction (see Appendix A for more information). The pairwise difference estimators all use the Epanechnikov kernel $K(u) = 3(1-u^2)\mathbb{1}\{u^2 < 1\}/4$ and the bandwidth sequence $n^{-1/9}/10$. Since $n^{1/9}$ is roughly equal to 2 for the sample sizes considered, the results are close to a constant bandwidth choice of $h_n = 1/20$. The adaptive bandwidth estimator uses $L = 2$ with $(c_1, c_2) = (1, 2)$.

Tables 1-3 demonstrate results for $R = 1000$, $\beta = \gamma = 1$ and for n equal to 50, 100, 200, 500, and 800. For each model, estimator and sample size, the first row (bias) gives the mean minus $\beta = 1$, the second row (MAE) gives the mean absolute error around $\beta = 1$, and the third row (rMAE) gives the mean absolute error around $\beta = 1$ divided by that of the infeasible benchmark $\hat{\beta}_1$. The fourth row (size) is the proportion of draws that are rejected by rule $|\hat{\beta}_k - 1|/\sqrt{\hat{V}_k} > 1.96$ for $k = 1, ..., 6$. For the first three estimators $\hat{V}_1 = \left[(Z_1'Z_1)^{-1}Z_1'(y - Z_1\hat{\beta}_1)(y - Z_1\hat{\beta}_1)'Z_1(Z_1'Z_1)^{-1}\right]_{1,1}$, $\hat{V}_2 = (x'x)^{-1}x'(y - \hat{\beta}_2 x)(y - \hat{\beta}_2 x)'x(x'x)^{-1}$, and $\hat{V}_3 = \left[(Z_2'Z_2)^{-1}Z_2'(y - Z_2\hat{\beta}_3)(y - Z_2\hat{\beta}_3)'(Z_2'Z_2)^{-1}\right]_{1,1}$. For the last three estimators, $\hat{V}_4 = \hat{V}_{4,n}$, $\hat{V}_5 = \hat{V}_{5,n}$, and $\hat{V}_6 = \hat{V}_{6,n}$ (see Appendix A for more details) respectively.

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 $\hat{\beta}_3$ also has a large and persistent bias that decreases with n but at a very slow rate.

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 the pairwise difference estimator with an adaptive bandwidth $\hat{\beta}_6$ behave similar to the infeasible $\hat{\beta}_2$, for n greater than 50. 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

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	-0.000	0.831	0.444	0.075	0.038	0.065
	MAE	0.118	0.831	0.444	0.194	0.201	0.155
	rMAE	1.000	7.042	3.763	1.644	1.703	1.313
	size	0.056	1.000	0.880	0.130	0.127	0.163
100							
	bias	-0.000	0.826	0.387	0.008	-0.036	0.005
	MAE	0.082	0.826	0.387	0.099	0.110	0.090
	rMAE	1.000	10.073	4.720	1.207	1.342	1.098
	size	0.044	1.000	0.958	0.052	0.080	0.085
200							
	bias	0.001	0.825	0.322	0.002	-0.042	0.002
	MAE	0.056	0.825	0.322	0.060	0.073	0.060
	rMAE	1.000	14.732	5.750	1.071	1.304	1.071
	size	0.055	1.000	0.958	0.059	0.110	0.065
500							
	bias	0.002	0.825	0.237	0.002	-0.043	0.002
	MAE	0.036	0.825	0.237	0.036	0.051	0.037
	rMAE	1.000	22.917	6.583	1.000	1.417	1.028
	size	0.042	1.000	0.959	0.042	0.168	0.045
800							
	bias	-0.001	0.824	0.187	-0.000	-0.045	-0.001
	MAE	0.029	0.824	0.187	0.029	0.049	0.029
	rMAE	1.000	28.414	6.448	1.000	1.690	1.000
	size	0.056	1.000	0.928	0.056	0.265	0.065

Table 1: This table contains simulation results for 1000 replications and a sample size of $n = 50, 100, 200, 500, 800$ and $\beta = 1$. Bias gives the simulation mean minus β. MAE gives the mean absolute error around β. rMAE gives the mean absolute error divided by that of the benchmark $\hat{\beta}_1$. Size gives the proportion of draws that fall outside the asymptotic 0.95 confidence interval based on a normal distribution with mean β and variances given in the text.

performance of the estimator slowly deteriorates as n increases (though the bias and variance of this estimator are 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 (a similar point is made in Section 5 of Johnsson and Moon 2015). As per the discussion in Section 3, this model demonstrates complications due to network sparsity. In Appendix C I demonstrate that the social characteristics are 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: $|u - v| \leq 40 \times d(u, v)$.

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 $\hat{\beta}_3$, even without a bias correction. When $n = 100$, $\hat{\beta}_5$ has approximately half the bias and mean absolute error of $\hat{\beta}_2$ while $\hat{\beta}_3$ offers a reduction of less than ten percent. When $n = 800$ the difference between the estimators is even more dramatic.

Table 3 contains results for the homophily model. As in the case of the beta model, I demonstrate in Appendix C that the social characteristics are also identified in the homophily model. Unlike the beta model, there is a relatively large amount of information about the network positions in the link 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 adaptive bandwidth and bias corrected estimators outperform both estimators over all of the sample sizes considered.

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	-0.002	0.499	0.470	0.380	0.331	0.366
	MAE	0.120	0.499	0.470	0.382	0.343	0.372
	rMAE	1.000	4.158	3.917	3.183	2.858	3.100
	size	0.066	0.979	0.902	0.644	0.394	0.654
100							
	bias	0.005	0.500	0.458	0.320	0.240	0.296
	MAE	0.080	0.500	0.458	0.321	0.243	0.297
	rMAE	1.000	6.250	5.725	4.013	3.038	3.713
	size	0.048	1.000	0.999	0.856	0.558	0.791
200							
	bias	0.003	0.501	0.447	0.260	0.146	0.227
	MAE	0.054	0.501	0.447	0.260	0.148	0.227
	rMAE	1.000	9.278	8.278	4.815	2.741	4.204
	size	0.040	1.000	1.000	0.943	0.484	0.870
500							
	bias	-0.000	0.500	0.406	0.193	0.055	0.146
	MAE	0.035	0.500	0.406	0.193	0.062	0.146
	rMAE	1.000	14.286	11.600	5.514	1.771	4.171
	size	0.046	1.000	1.000	0.992	0.249	0.848
800							
	bias	0.001	0.501	0.378	0.170	0.029	0.121
	MAE	0.029	0.501	0.378	0.170	0.040	0.121
	rMAE	1.000	17.276	13.035	5.862	1.379	4.172
	size	0.054	1.000	1.000	1.000	0.145	0.880

Table 2: This table contains simulation results for 1000 replications and a sample size of $n = 50, 100, 200, 500, 800$ and $\beta = 1$. Bias gives the simulation mean minus β. MAE gives the mean absolute error around β. rMAE gives the mean absolute error divided by that of the benchmark $\hat{\beta}_1$. Size gives the proportion of draws that fall outside the asymptotic 0.95 confidence interval based on a normal distribution with mean β and variances given in the text.

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	-0.003	0.497	0.224	0.160	0.105	0.126
	MAE	0.118	0.497	0.232	0.193	0.176	0.205
	rMAE	1.000	4.212	1.996	1.636	1.492	1.737
	size	0.064	0.978	0.356	0.175	0.122	0.279
100							
	bias	0.003	0.497	0.138	0.127	0.067	0.071
	MAE	0.078	0.497	0.147	0.141	0.112	0.118
	rMAE	1.000	6.372	1.885	1.8077	1.436	1.513
	size	0.051	1.000	0.274	0.180	0.098	0.174
200							
	bias	-0.000	0.502	0.083	0.096	0.030	0.049
	MAE	0.056	0.502	0.091	0.106	0.075	0.078
	rMAE	1.000	8.964	1.625	1.893	1.339	3.393
	size	0.049	1.000	0.219	0.215	0.075	0.150
500							
	bias	-0.001	0.498	0.044	0.074	0.006	0.025
	MAE	0.036	0.498	0.052	0.078	0.045	0.045
	rMAE	1.000	13.833	1.444	2.167	1.250	1.250
	size	0.048	1.000	0.170	0.312	0.061	0.114
800							
	bias	-0.000	0.502	0.037	0.063	-0.007	0.021
	MAE	0.028	0.502	0.042	0.065	0.035	0.035
	rMAE	1.000	17.929	1.500	2.321	1.250	1.250
	size	0.050	1.000	0.184	0.338	0.054	0.115

Table 3: This table contains simulation results for 1000 replications and a sample size of $n = 50, 100, 200, 500, 800$ and $\beta = 1$. Bias gives the simulation mean minus β. MAE gives the mean absolute error around β. rMAE gives the mean absolute error divided by that of the benchmark $\hat{\beta}_1$. Size gives the proportion of draws that fall outside the asymptotic 0.95 confidence interval based on a normal distribution with mean β and variance given in the text.

5 Directions for Future Work

I highlight two directions for future work. The first is to consider models in which the parameter of interest depends on the distribution of network links. For example, one might be interested in the functions $\beta(w_i)$ and $\lambda(w_i)$ in the model $y_i = x_i\beta(w_i) + \lambda(w_i) + \varepsilon_i$. To see why, suppose that x_i is an indicator for the adoption of some treatment. Then the function β describes how the treatment effect varies over the network, which intuitively might be nonconstant if the impact of treatment for a particular agent depends on the fraction of that agent's social connections that have been similarly treated. Estimating $\beta(w_i)$ potentially allows the researcher to determine which positions in the network are associated with, for example, the largest or smallest treatment effects. I plan to demonstrate how the tools of this paper might be used to estimate these and other features of both $\beta(w_i)$ and $\lambda(w_i)$ in future work.

The second direction for future work concerns a behavioral motivation for the model and estimator of this paper. In Appendix D, I provide a basic random utility interpretation for the network model along the lines of Graham (2014). However, the discussion is otherwise largely divorced from a developed literature on network formation models with strategic interaction. In future work, I hope to explore more connections between the setting of this paper and that literature.

One connection is potentially provided by the literature on network formation games with private information. Recent work in this literature employs a similar network formation model as a within-equilibrium reduced form characterization of linking behavior (see for example Leung 2015, Ridder and Sheng 2015, Menzel 2015). Here the social characteristics constitute public information about individual agents and the linking probabilities are conditionally independent given these characteristics and some equilibrium selection process. In this setting the link errors $\{\eta_{ij}\}_{i \neq j}$ constitute private information about the quality of individual links.

Understanding the mapping between structural models of network formation and this reduced-form representation might be mutually beneficial for both the network formation and network endogeneity literatures. For instance, the tools of this paper could be used to fit models of network formation in which not all of the public information that informs linking decisions is observed by the researcher. At the same time, a deeper understanding of network formation is important to help researchers fitting models with endogenous network formation identify and control for the many types of unobserved heterogeneity potentially lurking in the model's errors.

References

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

Ahn, H. (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.

Ai, C. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*(6), 1795–1843.

Angrist, J. D. (2014). The perils of peer effects. *Labour Economics 30*, 98–108.

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).

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006). Who's who in networks. wanted: the key player. *Econometrica 74*(5), 1403–1417.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson (2013). The diffusion of microfinance. *Science 341*(6144), 1236498.

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.

Bonhomme, S., T. Lamadon, and E. Manresa (2015). A distributional framework for matched employer employee data. Technical report, Technical Report.

Bonhomme, S. and E. Manresa (2015). Grouped patterns of heterogeneity in panel data. *Econometrica 83*(3), 1147–1184.

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

Calvó-Armengol, A., E. Patacchini, and Y. Zenou (2009). Peer effects and social networks in education. *The Review of Economic Studies 76*(4), 1239–1267.

Carrell, S. E., B. I. Sacerdote, and J. E. West (2013). From natural variation to optimal policy? the importance of endogenous peer group formation. *Econometrica 81*(3), 855–882.

Chamberlain, G. (1986). Notes on semiparametric regression. *manuscript, Department of Economics, University of Wisconsin-Madison*.

Chandrasekhar, A. G. and M. O. Jackson (2014). Tractable and consistent random graph models.

Chen, X., O. Linton, and I. Van Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica 71*(5), 1591–1608.

Christakis, N. A. and J. H. Fowler (2007). The spread of obesity in a large social network over 32 years. *New England journal of medicine 357*(4), 370–379.

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.

Dzemski, A. (2014). An empirical model of dyadic link formation in a network with unobserved heterogeneity. Technical report, University of Manheim Working Paper.

Elliott, M., B. Golub, and M. O. Jackson (2014). Financial networks and contagion. *The American economic review 104*(10), 3115–3153.

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. and P. Vieu (2006). *Nonparametric functional data analysis: theory and practice.* Springer Science & Business Media.

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. (2014). An empirical model of network formation: detecting homophily when agents are heterogenous.

Graham, B. S. (2015). Methods of identification in social networks. *Annu. Rev. Econ. 7*(1), 465–485.

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.

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.

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

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, 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.

Hsieh, C.-S. and L. F. Lee (2014). A social interactions model with endogenous friendship formation and selectivity. *Journal of Applied Econometrics*.

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.

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. 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 (2007). Szemerédis lemma for the analyst. *GAFA Geometric And Functional Analysis 17*(1), 252–270.

Lovász, L. and B. Szegedy (2010). Regularity partitions and the topology of graphons. In *An irregular mind*, pp. 415–446. Springer.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data. *Econometrica: Journal of the Econometric Society*, 357–362.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies 60*(3), 531–542.

Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Processes and their Applications 115*(1), 155–177.

Menzel, K. (2015). Strategic networkformation with many agents.

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.

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

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.

Robinson, P. M. (1988). Root-n-consistent semiparametric regression. *Econometrica 56*(4), 931–54.

Schmutte, I. M. (2014). Free to move? a network analytic approach for learning the limits to job mobility. *Labour Economics 29*, 49–61.

Shalizi, C. R. and A. C. Thomas (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research 40*(2), 211–239.

Wolfe, P. J. and S. C. Olhede (2013). Nonparametric graphon estimation. *arXiv preprint arXiv:1309.5936*.

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

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 that are not formally stated in the paper are labelled Lemma A1, Lemma A2, etc.

A.1 Lemmas and Theorems in Section 3.2

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

Proof of Theorem 1:

$$
\begin{aligned}
E\left[((y_i - y_j) - (x_i - x_j)b)^2 \,|\, ||f_{w_i} - f_{w_j}||_2 = 0\right] &= E\left[((x_i - x_j)(\beta - b) + (u_i - u_j))^2 \,|\, ||f_{w_i} - f_{w_j}||_2 = 0\right] \\
= (\beta - b)' E\left[(x_i - x_j)'(x_i - x_j)| \; ||f_{w_i} - f_{w_j}||_2 = 0\right](\beta - b) &+ E[(u_i - u_j)^2| \; ||f_{w_i} - f_{w_j}||_2 = 0] \\
- 2(\beta - b)' E[(x_i - x_j)'(u_i - u_j)| \; ||f_{w_i} - f_{w_j}||_2 = 0]
\end{aligned}
$$

The first summand is unique minimized at $b = \beta$ by Assumption 3. The second summand does not depend on b. The third summand is equal to 0 by Assumption 2. Notice Assumptions 2 and 3 are also necessary: if either assumption fails the sum of the first and third terms may be minimized at a b that is not equal β. □

A.2 Lemmas and Theorems in Section 3.3.1

Lemma 1: Suppose Assumptions 1 and 5 hold. Then

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

Proof of Lemma 1: The lemma is proved in four steps. Set $h'_n = n^{-\gamma/4}h_n$ and recall $h'_n n^{(1-\gamma)/2} \to \infty$, $p_{w_iw_j} = \int f_{w_i}(\tau)f_{w_j}(\tau)d\tau$ and $\hat{p}_{ij} = \frac{1}{n-2}\sum_{t\neq i,j} D_{it}D_{jt}$. Let $||\hat{p}_{w_i} - \hat{p}_{w_j}||_{2,n} := \hat{\delta}_{ij}$. I first show that $\max_{(i\neq j)} h_n'^{-1}|\hat{p}_{w_iw_j} - p_{w_iw_j}| \to_{a.s.} 0$. By Bernstein's Inequality, for any $\epsilon > 0$

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

and so by the union bound

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

Since $(n-2)^{1-\gamma/2}h_n'^2 \to \infty$ by the assumed choice of bandwidth sequence and $\sum_{n=3}^{\infty} n(n-1)\exp\left(\frac{-(n-2)(h'_n\epsilon)^2}{1+2h'_n\epsilon/3}\right) < \infty$ by the ratio test, $P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}|\hat{p}_{w_iw_j} - p_{w_iw_j}| > \epsilon\right) = 0$ follows from the first Borel-Cantelli Lemma. To see the summability claim, note that $(n-2)h_n'^2 > (n-2)^{\gamma}$ and $2h'_n\epsilon < 1$ eventually, so that $\sum_{n=3}^{\infty} n(n-1)\exp\left(\frac{-(n-2)(h'_n\epsilon)^2}{1+2h'_n\epsilon/3}\right)$ is finite if $\sum_{n=3}^{\infty} n(n-1)\exp\left(\frac{-(n-2)^{\gamma}\epsilon^2}{2}\right)$ is. Letting $m(n) = (n-2)^{1/\gamma}$, the latter sum is eventually less than
$\sum_{m=1}^{\infty} 2m^{2/\gamma}\exp\left(\frac{-m\epsilon^2}{2}\right) \times |\{n \in \{\mathbb{N}+2\} : n^{\gamma/2} \in (m-1, m]\}| \leq \sum_{m=1}^{\infty} 2m^{4/\gamma}\exp\left(\frac{-m\epsilon^2}{2}\right)$.
This final sum is absolutely convergent by the ratio test, for any $\gamma > 0$.

Second, let $||\hat{p}_{w_i} - p_{w_i}||_{2,n,j} = \left((n-2)^{-1}\sum_{s\neq i,j}\left(\hat{p}_{w_iw_s} - p_{w_iw_s}\right)^2\right)^{1/2}$. Then

$\max_{(i\neq j)} h_n'^{-1}|\hat{p}_{w_iw_j} - p_{w_iw_j}| \rightarrow_{a.s.} 0$ implies $\max_{(i\neq j)} h_n'^{-1}||\hat{p}_{w_i} - p_{w_i}||_{2,n,j} \rightarrow_{a.s.} 0$, since

$$\begin{aligned}
&P\left(\limsup_{n\to\infty} \max_{(i\neq j)} h_n'^{-1}||\hat{p}_{w_i} - p_{w_i}||_{2,n,j} > \epsilon\right) \\
&= P\left(\limsup_{n\to\infty} \max_{(i\neq j)} h_n'^{-1}\left((n-2)^{-1}\sum_{s\neq i,j}\left(\hat{p}_{w_iw_s} - p_{w_iw_s}\right)^2\right)^{1/2} > \epsilon\right) \\
&\leq P\left(\limsup_{n\to\infty} \max_{(i\neq j)} h_n'^{-1}\left|\hat{p}_{w_iw_j} - p_{w_iw_j}\right| > \epsilon\right)
\end{aligned}$$

because $h_n'^{-1}\left((n-2)^{-1}\sum_{s\neq i,j}\left(\hat{p}_{w_iw_s} - p_{w_iw_s}\right)^2\right)^{1/2} > \epsilon$ only if $h_n'^{-1}\left|\hat{p}_{w_iw_t} - p_{w_iw_t}\right| > \epsilon$ for some $t \neq i, j$.

Third, for $||p_{w_i} - p_{w_j}||_{2,n} = \left((n-2)^{-1}\sum_{s\neq i,j}\left(p_{w_iw_s} - p_{w_iw_s}\right)^2\right)^{1/2}$,
$\max_{(i\neq j)} h_n'^{-1}\left|||p_{w_i} - p_{w_j}||_{2,n} - ||p_{w_i} - p_{w_j}||_2\right| \rightarrow_{a.s.} 0$ since

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

with the last line again by Bernstein. So

$$P\left(\max_{(i\neq j)} h_n'^{-1}\left|||p_{w_i} - p_{w_j}||_{2,n} - ||p_{w_i} - p_{w_j}||_2\right| > \epsilon\right) \leq 2(n-2)^2\exp\left(\frac{-(n-2)h_n'\epsilon}{2 + 2\sqrt{h_n'\epsilon}/3}\right)$$

which is again absolutely summable for the assumed choice of h_n', since it is eventually bounded above by the summable sequence considered in the first part of this proof.

Finally, the second and third parts of this proof and a few applications of the triangle inequality yield

$$
\begin{aligned}
&P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}\left|||\hat{p}_{w_i}-\hat{p}_{w_j}||_{2,n}-||p_{w_i}-p_{w_j}||_2\right|>\epsilon\right)\\
&=P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}\left|||\hat{p}_{w_i}-\hat{p}_{w_j}||_{2,n}-||p_{w_i}-p_{w_j}||_{2,n}+||p_{w_i}-p_{w_j}||_{2,n}-||p_{w_i}-p_{w_j}||_2\right|>\epsilon\right)\\
&\leq P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}\left|||\hat{p}_{w_i}-\hat{p}_{w_j}||_{2,n}-||p_{w_i}-p_{w_j}||_{2,n}\right|>\epsilon/2\right)\\
&\qquad\qquad+P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}\left|||p_{w_i}-p_{w_j}||_{2,n}-||p_{w_i}-p_{w_j}||_2\right|>\epsilon/2\right)\\
&=P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}\left|||\hat{p}_{w_i}-\hat{p}_{w_j}||_{2,n}-||p_{w_i}-p_{w_j}||_{2,n}\right|>\epsilon/2\right)\\
&\leq P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}||(\hat{p}_{w_i}-\hat{p}_{w_j})-(p_{w_i}-p_{w_j})||_{2,n}>\epsilon/2\right)\\
&\leq 2P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}||(\hat{p}_{w_i}-p_{w_i})||_{2,n,j}>\epsilon/4\right)=0
\end{aligned}
$$

where $P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}\left|||p_{w_i}-p_{w_j}||_{2,n}-||p_{w_i}-p_{w_j}||_2\right|>\epsilon/2\right)$ in the second equality follows from the third part of the proof, and $P\left(\limsup_{n\to\infty}\max_{(i\neq j)} h_n'^{-1}||(\hat{p}_{w_i}-p_{w_i})||_{2,n,j}>\epsilon/4\right)$ in the final inequality from the second part of the proof. Since $h_n' = n^{-\gamma/4}h_n$, this completes the argument. □

Lemma 2: Suppose Assumption 1 holds. Then for every (w_i, w_j) pair

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

Furthermore, 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: To see the first part, observe that for every (w_i, w_j) pair

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

where the first inequality is due to Jensen and the second is due to the fact that $|f(\tau, s)| \leq 1$ for every $(\tau, s) \in [0, 1]^2$.

The proof of the second part is more complicated. 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 $\eta' > 0$, f is uniformly continuous when restricted to a closed subset A of $[0, 1]^2$ with measure at least $1 - \eta'$.

It follows that for any $\eta > 0$ there must also exist 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 $A' = \{(b, c) \in [0, 1]^2 : b \in B, c \in C(b)\}$.

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 $\eta \leq \epsilon'/16$.

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 τ 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

$$\int \left(\int f_\tau(s)(f_{w_i}(s) - f_{w_j}(s))ds \right)^2 d\tau > (\epsilon'/2)^2 \times 2\omega(\epsilon', \eta)$$

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 $\left|\int f_{w_i}(s)(f_{w_i}(s) - f_{w_j}(s))ds\right| < \epsilon'$ and $\left|\int f_{w_j}(s)(f_{w_i}(s) - f_{w_j}(s))ds\right| < \epsilon'$ also imply that $\left|\int (f_{w_i}(s) - f_{w_j}(s))(f_{w_i}(s) - f_{w_j}(s))ds\right| < 2\epsilon'$ 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 $\delta(\epsilon, \eta) = (\epsilon^2/4) \times (2\omega(\epsilon^2/2, \eta))^{1/2}$. $\square$

Notice ϵ depends on η for a given δ through the choice of $\omega(\epsilon, \eta)$, so that η cannot be chosen to be arbitrarily small for a fixed δ. Doing so requires a decoupling of the link function approximation error (due to the fact that f might not be smooth off of the set A') from the codegree approximation error (due to the fact that p induces a strictly coarser topology on $[0, 1]$ than f). Lemma 3 accomplishes this by replacing the measurability of f with a stronger continuity assumption, which essentially implies that the former error does not exist.

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

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

Proof of Lemma A1: As in the proof of the second part of Lemma 2, I begin with an appeal to Lusin's theorem (Dudley (2002), Theorem 7.5.2): for any $\eta > 0$ there must exist 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 $||f_u - f_v||_2 \leq \left(\epsilon'^2(1 - 2\eta) + 2\eta\right)^{1/2} \leq \epsilon' + \sqrt{2\eta}$. Since w_i, w_j are independent with standard uniform marginals, this means that $||f_{w_i} - f_{w_j}||_2 \leq \epsilon' + \sqrt{2\eta}$ with probability at least $(1 - 2\eta)\,\omega(\epsilon', \eta)$. Now just choose $\epsilon' < \epsilon/2$ and $\eta < \epsilon'^2/2$ to get $P\left(||f_{w_i} - f_{w_j}||_2 \leq \epsilon\right) \geq (1 - \epsilon^2/8)\,\omega(\epsilon/2, \epsilon^2/8) > 0$. $\square$

A direct implication of the first part of Lemma 2 and Lemma A1 is that for any $\epsilon > 0$, $P\left(||p_{w_i} - p_{w_j}||_2 \leq \epsilon\right) > 0$.

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

Proof of Theorem 2: Write

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

I show $\left(\binom{n}{2}r_n\right)^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i-x_j)'(x_i-x_j)K\left(\frac{\hat{\delta}_{ij}}{h_n}\right)\to_p 2\Gamma_0$, which is positive definite under Assumption 3. Similar arguments yield
$\left(\binom{n}{2}r_n\right)^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i-x_j)'(u_i-u_j)K\left(\frac{\hat{\delta}_{ij}}{h_n}\right)\to_p 0$, so that the claim follows from Slutsky and the continuous mapping theorem. Since $r_n>0$ with high probability from Lemma A1, both statistics are eventually well-defined.

Let $D_n=\left(\binom{n}{2}E\left[K\left(\frac{\delta_{ij}}{h_n}\right)\right]\right)^{-1}\sum_i\sum_j(x_i-x_j)'(x_i-x_j)K\left(\frac{\hat{\delta}_{ij}}{h_n}\right)$ then by the mean value theorem $D_n=\left(\binom{n}{2}E\left[K\left(\frac{\delta_{ij}}{h_n}\right)\right]\right)^{-1}\sum_i\sum_j(x_i-x_j)'(x_i-x_j)\left[K\left(\frac{\delta_{ij}}{h_n}\right)+K'\left(\frac{\iota_{ij}}{h_n}\right)\left(\frac{\hat{\delta}_{ij}-\delta_{ij}}{h_n}\right)\right]$ where $\{\iota_{ij}\}_{i\neq j}$ is the collection of intermediate values implied by that theorem. By Lemma 1 $\max_{i\neq j}\frac{\hat{\delta}_{ij}-\delta_{ij}}{h_n}=o_p\left(n^{-\gamma/4}\right)$ and by Markov's inequality $K'\left(\frac{\iota_{ij}}{h_n}\right)=o_p(r_nn^{\gamma/2})$, since $P\left(K'\left(\frac{\iota_{ij}}{h_n}\right)\geq r_nn^{\gamma/2}\right)\leq\frac{E\left[K'\left(\frac{\iota_{ij}}{h_n}\right)\right]}{r_nn^{\gamma/4}}=o(1)$ by choice of kernel density function in Assumption 5. It follows that

$$
\begin{aligned}
D_n&=\left(\binom{n}{2}E\left[K\left(\frac{\delta_{ij}}{h_n}\right)\right]\right)^{-1}\sum_i\sum_j(x_i-x_j)'(x_i-x_j)K\left(\frac{\delta_{ij}}{h_n}\right)+o_p(1)\\
&=D_n'+o_p(1)
\end{aligned}
$$

since x_i has finite second moments and $K'(u)$ is bounded.

Recall that $\delta_{ij}=\delta(w_i,w_j)$ so that D_n' 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

$$
D_n'=\left(E\left[K\left(\frac{\delta_{ij}}{h_n}\right)\right]\right)^{-1}E\left[(x_i-x_j)'(x_i-x_j)K\left(\frac{\delta_{ij}}{h_n}\right)\right]+o_p(1)
$$

since $nr_n\to\infty$. Additionally, measurability of f and Assumption 4 imply

$$
\begin{aligned}
&E\left[(x_i-x_j)'(x_i-x_j)K\left(\frac{\delta_{ij}}{h_n}\right)\right]=\int E\left[(x_i-x_j)'(x_i-x_j)|\delta_{ij}=u\right]K\left(\frac{u}{h_n}\right)dP(\delta_{ij}=u)\\
&=\int\left(\Gamma_0+o_p(1)\right)K\left(\frac{u}{h_n}\right)dP(\delta_{ij}=u)=\Gamma_0r_n+o_p\left(r_n\right)
\end{aligned}
$$

with the second equality is due to $E\left[(x_i-x_j)'(x_i-x_j)|\delta_{ij}\leq u\right]=\Gamma_0+o_p(1)$ by Lemma 2 and Assumptions 3 and 4. So $D_n = \Gamma_0 + o_p(1)$

A nearly identical argument gives

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

since $E\left[(x_i - x_j)'(u_i - u_j) | d(w_i, w_j) = h_n\right] = o_p(1)$ by Assumptions 2 and 4. $D_n^{-1} U_n = o_p(1)$ then follows from Slutsky and the continuous mapping theorem. □

A.3 Lemmas and Theorems in Section 3.3.2

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

Lemma A2: Suppose Assumption 5 holds and f_{w_i} has finite support. Then

$$\max_{(i \neq j)} ||\hat{p}_{w_i} - \hat{p}_{w_j}||_{2,n} \times \mathbb{1}\{||\hat{p}_{w_i} - \hat{p}_{w_j}||_{2,n} \leq \epsilon/2\} = o_{a.s.}(n^{-1/2} h_n)$$

Proof of Lemma A2: The assumption that f_{w_i} has finite support implies $\delta_{ij} 1\{\delta_{ij} \leq \epsilon\} = 0$ and $m_{ijt} 1\{\delta_{ij} \leq \epsilon/2\} := \left(p_{w_i w_t} - p_{w_j w_t}\right) \times 1\{\delta_{ij} \leq \epsilon/2\} = 0$ both with probability one. Consider the decomposition of $\hat{\delta}_{ij} 1\{\hat{\delta}_{ij} \leq \epsilon/2\}$ into

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

I first show $\max_{i \neq j} \sqrt{n} h_n^{-1} \hat{\delta}_{ij} 1\{\delta_{ij} \leq \epsilon/2\} = o_{a.s.}(1)$. As in the proof of Lemma 1, Bernstein's inequality gives

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

so that by the union bound

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

Averaging over t implies

$$P\left(\max_{i,j}\sqrt{n}h_n^{-1}\hat{\delta}_{ij}1\{\delta_{ij}\leq\epsilon/2\}\geq\eta\right)\leq 16(n-3)^3\exp\left(\frac{-(n-3)\eta h_n}{3\sqrt{n}}\right)$$

so long as $n\geq 6$. Since the right hand side is absolutely summable by arguments made in the proof of Lemma 1, $\max_{i\neq j}\sqrt{n}h_n^{-1}\hat{\delta}_{ij}1\{\delta_{ij}\leq\epsilon/2\}=o_{a.s.}(1)$.

I now show $\max_{i\neq j}\sqrt{n}h_n^{-1}\hat{\delta}_{ij}\left(1\{\hat{\delta}_{ij}\leq\epsilon/2\}-1\{\delta_{ij}\leq\epsilon/2\}\right)=o_{a.s.}(1)$. First,

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

Since $\delta_{ij}1\{\delta_{ij}\leq\epsilon\}=0$ with probability one, $\delta_{ij}\in(\epsilon/4,3\epsilon/4)$ is a probability zero event, and so it is sufficient to show

$$\max_{i\neq j}\sqrt{n}h_n^{-1}1\{|\hat{\delta}_{ij}-\delta_{ij}|>\epsilon/4\}=o_{a.s.}(1)$$

Using the inequality from before, the left hand side is nonzero on a set of probability at most $16(n-3)^3\exp\left(\frac{-(n-3)\epsilon h_n}{12\sqrt{n}}\right)$. Since this is again absolutely summable, $\sup_{i\neq j}\sqrt{n}h_n^{-1}\hat{\delta}_{ij}\left(1\{\hat{\delta}_{ij}\leq\epsilon/2\}-1\{\delta_{ij}\leq\epsilon/2\}\right)=o_{a.s.}(1)$ follows.

Taken together, the two arguments demonstrate that $\max_{i\neq j}\sqrt{n}h_n^{-1}\hat{\delta}_{ij}1\{\hat{\delta}_{ij}\leq\epsilon\}=o_{a.s.}(1)$, as claimed. □

Theorem 3: Suppose Assumptions 1-5 hold and the support of f_{w_i} is finite. Then

$$V_3^{-1/2}\left(\hat{\beta}-\beta\right)\to_d\mathcal{N}\left(0,I_k\right)$$

where $V_3 = \Gamma_0^{-1}\Omega_0\Gamma_0^{-1} \times s/n$, Γ_0 is as defined in Assumption 3, I_k is the $k \times k$ identity matrix, and

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

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

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

where $m = n(n-1)/2$. Since the support of f_{w_i} is finite, $E\left[K\left(\frac{\delta_{ij}}{h_n}\right)\right]$ $= K(0)P(||f_{w_i} - f_{w_j}||_2 = 0) > 0$ eventually (for $h_n \leq \epsilon$) since $P(\delta_{ij} = 0) > 0$.

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

$$\begin{aligned} U_n &= \frac{1}{m}\sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j) K\left(\frac{\hat{\delta}_{ij}}{h_n}\right)\right) \\ &= \frac{1}{m}\sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j) \left[K\left(\frac{\delta_{ij}}{h_n}\right) + K'\left(\frac{\iota_{ij}}{h_n}\right)\left(\frac{\hat{\delta}_{ij} - \delta_{ij}}{h_n}\right) 1\{\hat{\delta}_{ij} \leq h_n\}\right]\right) \end{aligned}$$

where ι_{ij} is a mean value between δ_{ij} and $\hat{\delta}_{ij}$. I first show $\frac{1}{m}\sum_i \sum_{j>i} \left((x_{il} - x_{jl}) (u_i - u_j) K'\left(\frac{\iota_{ij}}{h_n}\right)\left(\frac{\hat{\delta}_{ij} - \delta_{ij}}{h_n}\right) 1\{\hat{\delta}_{ij} \leq h_n\}\right) = o_p(n^{-1/2})$ for any positive integer $l \leq k$. By Cauchy-Schwartz

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

where $\bar{K}' = \sup_{u\in[0,1]} K'(u) < \infty$, $\sum_i \sum_{j>i} \left((x_{il} - x_{jl}) (u_i - u_j)\right)^2 = O_p(m)$ since x_i and u_i

have finite fourth moments, and $\max_{i \neq j} \left(\frac{\hat{\delta}_{ij} - \delta_{ij}}{h_n}\right) 1\{\hat{\delta}_{ij} \leq h_n\} = o_{a.s.}\left(n^{-1/2}\right)$ by Lemma A2. It follows that

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

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)

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

where $U = E\left[(x_i - x_j))' (u_i - u_j) K \left(\frac{\delta_{ij}}{h_n}\right)\right]$ and for $Z_i = (x_i, \nu_i, w_i)$

$$V = \lim_{h \to 0} 4E \left[E \left[(x_i - x_j)' (u_i - u_j) K \left(\frac{\delta_{ij}}{h} \right) \mid Z_i \right] E \left[(x_i - x_j) (u_i - u_j) K \left(\frac{\delta_{ij}}{h} \right) \mid Z_i \right] \right]$$
$$= \lim_{h \to 0} 4E \left[(x_i - x_j)' (x_i - x_k) (u_i - u_j)(u_i - u_k) K \left(\frac{\delta_{ij}}{h} \right) K \left(\frac{\delta_{ik}}{h} \right) \right]$$

Since f_{w_i} has finite support, $E[\delta_{ij} | \delta_{ij} \leq \epsilon] = 0$ for some $\epsilon > 0$, and so $U = E\left[(x_i - x_j))' (u_i - u_j) K(0) 1\{\delta_{ij} = 0\}\right]$ for n sufficiently large such that $h_n \leq \epsilon$. By Lemma 2, $1\{\delta_{ij} = 0\} = 1\{d_{ij} = 0\}$ with probability one, so Assumption 5 implies that $U = 0$ for any choice of $h_n \leq \epsilon$ (i.e. $U = 0$ eventually). Similarly $V = 4\Omega_0 K(0)^2 P\left(||f_{w_i} - f_{w_j}||_2 = 0, ||f_{w_i} - f_k||_2 = 0\right)$ so long as $h_n \leq \epsilon$. So by Slutsky's Theorem,

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

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

Lemma 3: Suppose Assumptions 1 and 6 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 $||p_{w_i} - p_{w_j}||_2 < \sqrt{8C_6}K^{-\alpha}$.

Proof of Lemma 3: The first inequality follows from the first part of Lemma 2 holding exactly for every (w_i, w_j) pair. The proof of the second inequality essentially mirrors the second part of Lemma 2, and so only a quick sketch is provded here. I first demonstrate that $||p_{w_i} - p_{w_j}||_2 \leq \left(4(4C_6)^{1/\alpha}\right)^{-1} \epsilon'^{\frac{4\alpha+2}{\alpha}}$ and $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}} < K^{-1}$ imply that $||f_{w_i} - f_{w_j}||_2 \leq \sqrt{2\epsilon'}$ with probability one.

Suppose $\left|\int f_{w_i}(s)\left(f_{w_i}(s) - f_{w_j}(s)\right) ds\right| > \epsilon'$. Then $\left|\int f_{\tau}(s)\left(f_{w_i}(s) - f_{w_j}(s)\right) ds\right| > \epsilon'/2$ for $\tau \in [0,1]$ so long as τ 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 τ in $[0,1]$ that satisfty these conditions is at least $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}}$. It follows that so long as $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}} < K^{-1}$

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

with probability one.

Following the logic of Lemma 2, I conclude that $||p_i - p_j||_2 \leq \left(4(4C_6)^{1/\alpha}\right)^{-1} \epsilon'^{\frac{4\alpha+2}{\alpha}}$ implies that $||f_{w_i} - f_{w_j}||_2 \leq \sqrt{2\epsilon'}$ with probability one so long as $\left(\frac{\epsilon'}{4C_6}\right)^{\frac{1}{\alpha}} < K^{-1}$. Replacing ϵ' with $\epsilon^2/2$ yields

$$2^{\frac{2\alpha}{4\alpha+2}} 4^{\frac{4}{4\alpha+2}} (4C_6)^{\frac{1}{4\alpha+2}} ||p_i - p_j||_2^{\frac{2\alpha}{4\alpha+2}} \leq \epsilon \text{ implies that } ||f_{w_i} - f_{w_j}||_2 \leq \epsilon$$

with probability one if $\left(\frac{\epsilon^2}{8C_6}\right)^{\frac{1}{\alpha}} < K^{-1}$.

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 statement of the lemma 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 6 hold. Then $P\left(||f_{w_i} - f_{w_j}||_2 \leq \epsilon\right) > C_6^{-1/\alpha}\epsilon^{1/\alpha}$, so long as $\epsilon \leq C_6 K^{-\alpha}$

Proof of Lemma A3: The proof of Lemma A3 essentially mirrors that of Lemma A1, except Assumption 6 allows for the replacement of $\omega(\epsilon, \eta)$ with $\left(\frac{\epsilon}{C_6}\right)^{1/\alpha}$. Notice that that 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 $\left(\frac{\epsilon}{C_6}\right)^{1/\alpha}$. So $P\left(||f_{w_i} - f_{w_j}||_2 \leq \epsilon\right) > \frac{1}{C_6^{1/\alpha}}\epsilon^{1/\alpha}$ as claimed. $\square$

Theorem 4: Suppose Assumptions 1-4 and 6-8 hold and $\alpha \times \zeta > 1/2$. Then

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

where $V_{4,n} = \Gamma_0^{-1}\Omega_n\Gamma_0^{-1}/n$, Γ_0 is as defined in Assumption 3, r_n is as defined in Assumption 5, and I_k is the $k \times k$ identity matrix, and

$$\begin{aligned}\beta_{h_n} &= \beta + \left(\Gamma_0\right)^{-1} E\left[(x_i - x_j)'(u_i - u_j)K\left(\frac{||p_i - p_j||_2}{h_n}\right)\right] / \left(2r_n\right)\\ \Omega_n &= E\left[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k)K\left(\frac{||p_i - p_j||_2}{h_n}\right) K\left(\frac{||p_i - p_k||_2}{h_n}\right)\right] / \left(r_n^2\right)\end{aligned}$$

Proof of Theorem 4: The proof is simplified by squaring the empirical codegree differences so that

$$\begin{aligned}\left(\hat{\beta} - \beta\right) = &\left(\frac{1}{\binom{n}{2}r_n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i - x_j)'(x_i - x_j)K_{1/2}\left(\frac{\hat{\delta}_{ij}^2}{h_n^2}\right)\right)^{-1}\\ &\frac{1}{\binom{n}{2}r_n}\left(\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}(x_i - x_j)'(u_i - u_j)K_{1/2}\left(\frac{\hat{\delta}_{ij}^2}{h_n^2}\right)\right)\end{aligned}$$

where $r_n = E\left[K_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right)\right]$ and $K_{1/2}(u) = K(\sqrt{u})$ is supported, positive, and twice differentiable on $[0,1)$ by Assumption 8. Recall $r_n > 0$ by Lemma A1.

The proof of Theorem 2 demonstrates that Assumptions 1-5 are sufficient for the denominator to converge in probability to $2\Gamma_0$, which is eventually invertible by Assumption 3. As for the numerator,

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

where ι_{ij} is the intermediate value between $\hat{\delta}_{ij}^2$ and δ_{ij}^2 suggested by the mean value theorem and Taylor's theorem. I consider each of the summands individually. I first show that

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

Let $s_n = n^{-1/2} h_n^4 r_n$. Since $\delta_{ij} \le C|w_i - w_j|^\alpha$ by the first part of Lemma 2 and Assumption 6, $r_n \ge K C^{-1/\alpha} h_n^{1/\alpha}$ for $K = \liminf_{h\to 0} E\left[K\left(\frac{\delta_{ij}}{h}\right)|\delta_{ij} \le h\right] > 0$ by Lemma A2. Since $n^{1/2-\gamma} h_n^{4+1/\alpha} \to \infty$ for some $\gamma > 0$ by Assumption 9, $n^{1-\gamma} s_n \to \infty$, and so Lemma 1 implies that $\sup_{i\neq j}\left(\frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{\sqrt{s}}\right)^2 = o_{a.s.}(1)$ or $\sup_{i\neq j}\left(\frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n^2 \sqrt{r_n}}\right)^2 = o_{a.s.}\left(n^{-1/2}\right)$. It follows that

$$\begin{aligned}
&\frac{1}{\binom{n}{2} r_n} \sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j) K''_{1/2}\left(\frac{\iota_{ij}}{h_n}\right)\left(\frac{\hat{\delta}_{ij}^2 - \delta_{ij}^2}{h_n^2}\right)^2 \right) \\
&\qquad \le \frac{\bar{K}''_{1/2}}{\binom{n}{2}} \sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j)\right) \times o_{a.s.}\left(n^{-1/2}\right)
\end{aligned}$$

where $\bar{K}''_{1/2} = \sup_{u\in[0,1]} K''_{1/2}(u)$ and the last line is $o_p\left(n^{-1/2}\right)$ because x_i and u_i are assumed to have finite fourth moments by Assumption 2. Thus

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

Now let

$$\begin{aligned} \tilde{\delta}_{ij} = \tilde{\delta}(w_i, w_j)^2 = \frac{1}{n}\sum_{t=1}^{n} & \left(\frac{1}{n} \sum_{s_1=1}^{n} f(w_t, w_{s_1}) \left(f(w_i, w_{s_1}) - f(w_j, w_{s_1}) \right) \right) \\ & \times \left(\frac{1}{n} \sum_{s_2=1}^{n} f(w_t, w_{s_2}) \left(f(w_i, w_{s_2}) - f(w_j, w_{s_2}) \right) \right) \end{aligned}$$

and rewrite the numerator as

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

In the remainder of this proof, I show that the second summand is $o_p(n^{-1/2})$, while the first part is a fifth-order U-statistic. First,

$$\frac{1}{\binom{n}{2} r_n} \sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j) \left[K_{1/2}'\left(\frac{\delta_{ij}^2}{h_n^2}\right)\left(\frac{\hat{\delta}_{ij}^2 - \tilde{\delta}_{ij}^2}{h_n^2}\right) \right] \right) = o_p\left(n^{-1/2}\right)$$

by Chebyshev's inequality, since $E\left[\left(\frac{\hat{\delta}_{ij}^2 - \tilde{\delta}_{ij}^2}{h^2}\right) | x_i, x_j, u_i, u_j, w_i, w_j\right] = 0$ implies

$\frac{1}{\binom{n}{2} r_n} \sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j) \left[K'_{1/2} \left(\frac{\delta_{ij}^2}{h_n^2} \right) \left(\frac{\hat{\delta}_{ij}^2 - \tilde{\delta}_{ij}^2}{h_n^2} \right) \right] \right)$ is mean zero and

$$E\left[\left(\frac{1}{\binom{n}{2} r_n} \sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j) \left[K'_{1/2} \left(\frac{\delta_{ij}^2}{h_n^2} \right) \left(\frac{\hat{\delta}_{ij}^2 - \tilde{\delta}_{ij}^2}{h_n^2} \right) \right] \right) \right)^2 \right]$$

$$= \frac{1}{\binom{n}{2}^2 n^6 r_n^2 h_n^4} E\Bigg[\sum_{i_1} \sum_{i_2} \sum_{j_1} \sum_{j_2} \sum_{t_1} \sum_{t_2} \sum_{s_{11}} \sum_{s_{12}} \sum_{s_{21}} \sum_{s_{22}}$$
$$(x_{i_1} - x_{j_1})'(x_{i_2} - x_{j_2})(u_{i_1} - u_{j_1})(u_{i_2} - u_{j_2}) K'_{1/2} \left(\frac{\delta_{i_1 j_1}^2}{h_n^2} \right) K'_{1/2} \left(\frac{\delta_{i_2 j_2}^2}{h_n^2} \right)$$
$$\times [D_{t_1 s_{11}} D_{t_1 s_{12}} (D_{i_1 s_{11}} - D_{j_1 s_{11}})(D_{i_1 s_{12}} - D_{j_1 s_{12}}) - f_{t_1 s_{11}} f_{t_1 s_{12}} (f_{i_1 s_{11}} - f_{j_1 s_{11}})(f_{i_1 s_{12}} - f_{j_1 s_{12}})]$$
$$\times [D_{t_2 s_{21}} D_{t_2 s_{22}} (D_{i_2 s_{21}} - D_{j_2 s_{21}})(D_{i_2 s_{22}} - D_{j_2 s_{22}}) - f_{t_2 s_{21}} f_{t_2 s_{22}} (f_{i_2 s_{21}} - f_{j_2 s_{21}})(f_{i_2 s_{22}} - f_{j_2 s_{22}})] \Bigg]$$

is $o(n^{-1})$. To see this, 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

$$E\Bigg[[D_{t_1 s_{11}} D_{t_1 s_{12}} (D_{i_1 s_{11}} - D_{j_1 s_{11}})(D_{i_1 s_{12}} - D_{j_1 s_{12}}) - f_{t_1 s_{11}} f_{t_1 s_{12}} (f_{i_1 s_{11}} - f_{j_1 s_{11}})(f_{i_1 s_{12}} - f_{j_1 s_{12}})]$$
$$\times [D_{t_2 s_{21}} D_{t_2 s_{22}} (D_{i_2 s_{21}} - D_{j_2 s_{21}})(D_{i_2 s_{22}} - D_{j_2 s_{22}}) - f_{t_2 s_{21}} f_{t_2 s_{22}} (f_{i_2 s_{21}} - f_{j_2 s_{21}})(f_{i_2 s_{22}} - f_{j_2 s_{22}})]$$
$$| Z_{i_1}, Z_{i_2}, Z_{j_1}, Z_{j_2}, Z_{t_1}, Z_{t_2}, Z_{s_{11}}, Z_{s_{12}}, Z_{s_{21}}, Z_{s_{22}} \Bigg] = 0$$

where $Z_i = \{x_i, w_i, \nu_i\}$. Since $K'_{1/2}\left(\frac{\delta_{i_1 j_1}^2}{h^2}\right)$ is $O_p(r_n)$ by Assumption 8 (see the proof of Theorem 2 for the formal argument), the desired term is $o(n^{-1})$ since $nh_n^4 \to \infty$.

It follows that

$$U_n = \frac{1}{\binom{n}{2} r_n} \sum_i \sum_{j>i} \left((x_i - x_j)' (u_i - u_j) \left[K_{1/2}\left(\frac{\delta_{ij}^2}{h^2}\right) + K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \left(\frac{\tilde{\delta}_{ij}^2 - \delta_{ij}^2}{h_n^2}\right) \right] \right) + o_p\left(n^{-1/2}\right)$$

$$= \frac{1}{\binom{n}{5}^2 r_n} \sum_i \sum_{j>i} \sum_{t>j} \sum_{s_1>t} \sum_{s_2>s_1} (x_i - x_j)' (u_i - u_j) \left[K_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \right.$$
$$\left. + h_n^{-2} K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \left(f_{ts_1} f_{ts_2} (f_{is_1} - f_{js_1})(f_{is_2} - f_{js_2}) - \delta_{ij}^2 \right) \right] + o_p\left(n^{-1/2}\right)$$

so that U_n is equivalent to a 5th order U-statistic up to a $o_p(1/\sqrt{n})$ error. As in Theorem 3, I apply Lemma 3.2 from Powell et al. (1989) to rewrite this statistic as the sum of first order projections.

$$U_n = E[U_n] + \frac{2}{n r_n} \sum_{\tau=1}^n \left(E\left[(x_\tau - x_j)'(u_\tau - u_j) K_{1/2}\left(\frac{\delta_{\tau j}^2}{h_n^2}\right) |Z_\tau \right] - E[U_n] \right)$$
$$+ \frac{1}{n r_n h_n^2} \sum_{\tau=1}^n E\left[(x_i - x_j)' (u_i - u_j) K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \left(f_{\tau s_1} f_{\tau s_2} (f_{is_1} - f_{js_1})(f_{is_2} - f_{js_2}) - \delta_{ij}^2 \right) |Z_\tau \right]$$
$$+ \frac{2}{n r_n h_n^2} \sum_{\tau=1}^n E\left[(x_i - x_j)' (u_i - u_j) K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \left(f_{t\tau} f_{ts_2} (f_{i\tau} - f_{j\tau})(f_{is_2} - f_{js_2}) - \delta_{ij}^2 \right) |Z_\tau \right]$$
$$+ o_p\left(n^{-1/2}\right)$$

where $E[U_n] = E\left[(x_i - x_j)'(u_i - u_j) K_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \right]$ and $Z_\tau = \{x_\tau, w_\tau, \nu_\tau\}$.

When $\alpha \times \zeta > 1/2$ the second and third terms are both $o_p\left(n^{-1/2}\right)$. For the second term, I show this by fixing some $\epsilon > 0$ and writing

$$P\left(E\left[(x_i - x_j)' (u_i - u_j) K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \left(f_{\tau s_1} f_{\tau s_2} (f_{is_1} - f_{js_1})(f_{is_2} - f_{js_2}) - \delta_{ij}^2 \right) |Z_\tau \right] \geq r_n h_n^2 \epsilon \right)$$
$$= P\left(E\left[(x_i - x_j)' (u_i - u_j) K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) \left(E[f_{\tau s}(f_{is} - f_{js})|Z_i, Z_j, Z_\tau]^2 - \delta_{ij}^2 \right) |Z_\tau \right] \geq r_n h_n^2 \epsilon \right)$$
$$\leq E\left[\left| E\left[(x_i - x_j)' (u_i - u_j) K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) |Z_\tau \right] \right| \times \left(E[f_{\tau s}(f_{is} - f_{js})|Z_i, Z_j, Z_\tau]^2 + \delta_{ij}^2 \right) |Z_\tau \right] / r_n h_n^2 \epsilon$$

with the last line by Markov's inequality and the triangle inequality. Since $\left| E\left[(x_i - x_j)' (u_i - u_j) K'_{1/2}\left(\frac{\delta_{ij}^2}{h_n^2}\right) |Z_\tau \right] \right| = o_p(r_n)$ and both $E[f_{\tau s}(f_{is} - f_{js})|Z_i, Z_j, Z_\tau]^2$ and

δ_{ij}^2 are $O_p(h_n^2)$, the term is $o_p(1)$. So the second summand

$$\frac{1}{nr_nh_n^2}\sum_{\tau=1}^{n}E\left[(x_i-x_j)'(u_i-u_j)K_{1/2}'\left(\frac{\delta_{ij}^2}{h_n^2}\right)\left(f_{\tau s_1}f_{\tau s_2}(f_{is_1}-f_{js_1})(f_{is_2}-f_{js_2})-\delta_{ij}^2\right)|Z_\tau\right]$$

is an average of n independent random variables with finite third moments (since x_i and u_i have finite sixth moments) that are each $o_p(1)$, and so must be $o_p\left(n^{-1/2}\right)$ by the Lindeberg-Levy central limit theorem.

Bounding the third term is a bit more complicated. Again fix some $\epsilon > 0$ and write

$$P\left(E\left[(x_i-x_j)'(u_i-u_j)K_{1/2}'\left(\frac{\delta_{ij}^2}{h_n^2}\right)\left(f_{t\tau}f_{ts_2}(f_{i\tau}-f_{j\tau})(f_{is_2}-f_{js_2})-\delta_{ij}^2\right)|Z_\tau\right]\geq r_nh_n^2\epsilon\right)$$

However, this time Markov's inequality only provides the upper bound

$$E\left[\left|E\left[(x_i-x_j)'(u_i-u_j)K_{1/2}'\left(\frac{\delta_{ij}^2}{h_n^2}\right)|Z_\tau\right]\right|\right.$$
$$\left.\times\left(E[f_{t\tau}(f_{i\tau}-f_{j\tau})|Z_i,Z_j,Z_\tau]E[f_{ts}(f_{is}-f_{js})|Z_i,Z_j]+\delta_{ij}^2\right)|Z_\tau\right]/r_nh_n^2\epsilon$$

Here δ_{ij}^2 is $O_p(h_n^2)$ and $E[f_{ts}(f_{is}-f_{js})|Z_i,Z_j]$ is $O_p(h_n)$ by Jensen's inequality, but it is only possible to demonstrate that $E[f_{t\tau}(f_{i\tau}-f_{j\tau})|Z_i,Z_j,Z_\tau]\leq ||f_{w_i}-f_{w_j}||_2=O_p(h_n^{2\alpha/(1+2\alpha)})$ by Lemma 3. This is where I use the $\zeta\times\alpha>1/2$ condition so that $\left|E\left[(x_i-x_j)'(u_i-u_j)K_{1/2}'\left(\frac{\delta_{ij}^2}{h_n^2}\right)|Z_\tau\right]\right|$ is not just $o_p(r_n)$ but $o_p(h_n^{2\alpha\zeta/(1+2\alpha)}r_n)$. Together, these rates imply that the term is $o_p(1)$, and that the third summand

$$\frac{1}{nr_nh_n^2}\sum_{\tau=1}^{n}E\left[(x_i-x_j)'(u_i-u_j)K_{1/2}'\left(\frac{\delta_{ij}^2}{h_n^2}\right)\left(f_{t\tau}f_{ts_2}(f_{i\tau}-f_{j\tau})(f_{is_2}-f_{js_2})-\delta_{ij}^2\right)|Z_\tau\right]$$

is $o_p\left(n^{-1/2}\right)$ by previous arguments.

It follows from these two arguments that

$$U_n=E[U_n]+\frac{2}{nr_n}\sum_{\tau=1}^{n}\left(E\left[(x_\tau-x_j)'(u_\tau-u_j)K_{1/2}\left(\frac{\delta_{\tau j}^2}{h_n^2}\right)|Z_\tau\right]-E[U_n]\right)+o_p\left(n^{-1/2}\right)$$

U_n is simply an iid sum of random variables with bounded third moments, so by the Lindeberg-Levy central limit theorem

$$V_n^{\prime\prime -1/2}\left(U_n - E[U_n]\right) \to_d \mathcal{N}\left(0, I_k\right)$$

where

$$\begin{aligned} V_n^{\prime\prime} &= E\left[\left(\frac{4}{r_n^2}\left(E\left[(x_\tau - x_j)'(u_\tau - u_j)K_{1/2}\left(\frac{\delta_{\tau j}^2}{h_n^2}\right)|Z_\tau\right] - E[U_n]\right)\right)\right. \\ &\qquad \left.\times\left(E\left[(x_\tau - x_j)(u_\tau - u_j)K_{1/2}\left(\frac{\delta_{\tau j}^2}{h_n^2}\right)|Z_\tau\right] - E[U_n]\right)\right] \\ &= \frac{4}{r_n^2}E\left[(x_\tau - x_j)'(x_\tau - x_k)(u_\tau - u_j)(u_\tau - u_k)K_{1/2}\left(\frac{\delta_{\tau j}^2}{h_n^2}\right)K_{1/2}\left(\frac{\delta_{\tau k}^2}{h_n^2}\right)\right] \end{aligned}$$

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

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

where $E[U_n] = r_n^{-1}E\left[(x_i - x_j)'(u_i - u_j)K\left(\frac{\delta(w_i,w_j)}{h_n}\right)\right]$ as claimed. □

A.4 Theorems in Sections 3.3.3 and 3.3.4

Theorem 5: Suppose Assumptions 1-4 and 6-9 hold, and $L > ((1+2\alpha)\theta - \alpha)/\alpha$. Then

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

where $V_{5,n} = \sum_{l_1=1}^{L}\sum_{l_2=1}^{L} a_{l_1}a_{l_2}\Gamma_0^{-1}\Omega_{n,l_1 l_2}\Gamma_0^{-1}/n$, Γ_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

$$\Omega_{n,l_1 l_2} = E\left[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k)K\left(\frac{||p_i - p_j||_2}{h_n}\right)K\left(\frac{||p_i - p_k||_2}{h_n}\right)\right] / \left(r_n^2\right)$$

Proof of Theorem 5: Since $\bar{\beta}_L = \sum_{l=1}^{L} a_l \hat{\beta}_{C_l h_n}$, the logic of Theorem 4 and the continuous mapping theorem imply

$$\sqrt{n}\left(\bar{\beta}_L - \bar{\beta}_{L,h_n}\right) = \sum_{l=1}^{L} a_l \sqrt{n}\left(\hat{\beta}_{C_l h_n} - \beta_{C_l h_n}\right) \to_d \mathcal{N}\left(0, \sum_{l_1=1}^{n}\sum_{l_2=1}^{n} \Gamma_0^{-1}\Omega_{l_1 l_2, h_n}\Gamma_0^{-1}\sigma_{l_1,l_2,h_n}\right)$$

where $\bar{\beta}_{L,h} = \sum_{l=1}^{L} a_l \beta_{C_l h}$ and
$\Omega_{n,l_1 l_2} = E\left[(x_i - x_j)'(x_i - x_k)(u_i - u_j)(u_i - u_k)K\left(\frac{||p_i - p_j||_2}{h_n}\right)K\left(\frac{||p_i - p_k||_2}{h_n}\right)\right] / (r_n^2)$. By Assumption 9 and the definition of $\{a_1, ..., a_L\}$, $\bar{\beta}_{L,h}$ can be written as

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

since $\sum_{l_2} a_{l_2} = 1$ by choice of $\{a_1, ..., a_L\}$. Furthermore, $\{a_1, ..., a_L\}$ also satisfies $\left[\sum_{l_1} a_{l_1} c_{l_1}^{l_2/\theta}\right] = 0$ for all $l_2 \in \{1, ..., L\}$, so the second summand is 0 and $\bar{\beta}_{L,h} = \beta + o_p\left(n^{-1/2}\right)$. The claim follows. $\square$

Theorem 6: Suppose Assumptions 1-5 hold. Then $\hat{\Gamma}_{h_n}^{-1}\hat{\Omega}_{h_n,h_n}\hat{\Gamma}_{h}^{-1}/\sqrt{n} \to_p V_{4,n}$ and $\sum_{l_1=1}^{L}\sum_{l_2=1}^{L}\hat{\Gamma}_{c_{l_1}h_n}^{-1}\hat{\Omega}_{c_{l_1}h_n, c_{l_2}h_n}\hat{\Gamma}_{c_{l_2}h_n}^{-1}/\sqrt{n} \to_p V_{5,n}$

Proof of Theorem 6 It is sufficient to prove the second result, which nests the first as a special case. In the proof of Theorem 2 I demonstrate that Assumptions 1-5 are sufficient for $\left(E\left[K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)\right]\right)^{-1}\hat{\Gamma}_{ch_n} = 2\Gamma_0 + o_p(1)$ for any constant $c > 0$. It remains to be shown that $\left(E\left[K\left(\frac{\delta(w_i,w_j)}{c_1 h_n}\right)\right]\right)^{-1}\left(E\left[K\left(\frac{\delta(w_i,w_j)}{c_2 h_n}\right)\right]\right)^{-1}\hat{\Omega}_{c_1 h_n, c_2 h_n}$ converges to $\Omega_{nc_1c_2}$.
I first fix agent i and $Z_i = \{x_i, w_i, \nu_i\}$ and study the average $\left(E\left[K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)\right]\right)^{-1}(n-2)^{-1}\sum_{j>i}(x_i - x_j)'(\hat{u}_i - \hat{u}_j)K\left(\frac{\hat{\delta}_{ij}}{ch_n}\right)$ for some fixed $c > 0$. Since

$\hat{u}_i = u_i + x_i(\hat{\beta} - \beta)$ this average can be rewritten

$$\left(E\left[K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)\right]\right)^{-1}(n-2)^{-1}\sum_{j>i}(x_i-x_j)'\left[(u_i-u_j)-(x_i-x_j)(\hat{\beta}-\beta)\right]K\left(\frac{\hat{\delta}_{ij}}{ch_n}\right)$$
$$=\left(E\left[K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)\right]\right)^{-1}(n-2)^{-1}\sum_{j>i}(x_i-x_j)'(u_i-u_j)K\left(\frac{\hat{\delta}_{ij}}{ch_n}\right)$$
$$-\left(E\left[K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)\right]\right)^{-1}(n-2)^{-1}\sum_{j>i}(x_i-x_j)'(x_i-x_j)K\left(\frac{\hat{\delta}_{ij}}{ch_n}\right)(\hat{\beta}-\beta)$$

The first summand converges to $\left(E\left[K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)\right]\right)^{-1}E\left[(x_i-x_j)'(u_i-u_j)K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)|Z_i\right]$ following from arguments made in Theorem 3. The first part of the second summand $\left(E\left[K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)\right]\right)^{-1}(n-2)^{-1}\sum_{j>i}(x_i-x_j)'(x_i-x_j)K\left(\frac{\delta(w_i,w_j)}{ch_n}\right)$ is bounded following arguments made in Theorem 2, and so the second summand converges to 0 in probability since $(\hat{\beta}-\beta)=o_p(1)$ by Theorem 2. As a result, $\left(E\left[K\left(\frac{\delta(w_i,w_j)}{c_1h_n}\right)\right]\right)^{-1}\left(E\left[K\left(\frac{\delta(w_i,w_j)}{c_2h_n}\right)\right]\right)^{-1}\hat{\Omega}_{c_1h_n,c_2h_n}$ can be written as

$$(n-2)^{-1}4\sum_{i=1}^{n-1}E\left[(x_i-x_j)'(u_i-u_j)K\left(\frac{\delta(w_i,w_j)}{c_1h_n}\right)|Z_i\right]E\left[(x_i-x_j)(u_i-u_j)K\left(\frac{\delta(w_i,w_j)}{c_2h_n}\right)|Z_i\right]$$
$$\times\left(E\left[K\left(\frac{\delta(w_i,w_j)}{c_1h_n}\right)\right]E\left[K\left(\frac{\delta(w_i,w_j)}{c_2h_n}\right)\right]\right)^{-1}$$
$$=(n-2)^{-1}4\sum_{i=1}^{n-1}E\left[(x_i-x_j)'(x_i-x_k)(u_i-u_j)(u_i-u_k)K\left(\frac{\delta(w_i,w_j)}{c_1h_n}\right)K\left(\frac{\delta(w_i,w_k)}{c_2h_n}\right)|Z_i\right]$$
$$\times\left(E\left[K\left(\frac{\delta(w_i,w_j)}{c_1h_n}\right)\right]E\left[K\left(\frac{\delta(w_i,w_j)}{c_2h_n}\right)\right]\right)^{-1}$$

Together, the two results imply that $\hat{\Gamma}^{-1}_{c_{l_1}h_n}\hat{\Omega}_{c_{l_1}h_n,c_{l_2}h_n}\hat{\Gamma}^{-1}_{c_{l_2}h_n}\to_p\Gamma_0^{-1}\Omega_{c_1h_n,c_2h_n}\Gamma_0^{-1}$, and the claim follows from the continuous mapping theorem. □