Essays on Econometrics of Dyadic Data

by

Fengshi Niu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Bryan S. Graham, Chair
Professor James L. Powell
Professor Michael Jansson
Professor Adityanand Guntuboyina

Summer 2021

Essays on Econometrics of Dyadic Data

Abstract

Essays on Econometrics of Dyadic Data

by

Fengshi Niu

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Bryan S. Graham, Chair

Many important social and economic variables are naturally defined for pairs of agents (or dyads). Examples include trade between pairs of countries (e.g., Tinbergen, 1962), input purchases and sales between pairs of firms (e.g., Atalay et al., 2011), research and development (R&D) partnerships across firms (e.g., König et al., 2019) and friendships between individuals (e.g., Christakis et al., 2020). Dyadic data arises frequently in the analysis of social and economic issues. See Graham (2020a) for many other examples and references. While the statistical analysis of network data began almost a century ago, rigorously justified methods of inference for dyadic or network statistics are only now emerging (cf., Goldenberg et al., 2010).

This dissertation studies statistical inference problems of dyadic data. Throughout I focus on target parameters of fundamental theoretical and applied interest. These include *density functions*, *regression functions*, *density-weighted average derivatives*, and *coefficients in linear regressions*. Dyadic data exhibits a distinct kind of local dependence property: i.e., any random variables of dyads that share one or two indices/agents may be dependent. The four chapters of this dissertation develop a broad set of theoretical results for estimation and inference of nonparametric, parametric, and semiparametric models for dyadic data and present generic and in some cases surprising implications of the local dependence.

In Chapter 1 I study nonparametric estimation of density functions for undirected dyadic random variables (i.e., random variables defined for all $n \overset{def}{\equiv} \binom{N}{2}$ unordered pairs of agents/nodes in a weighted network of order N). In this setting, I show that density functions may be estimated by an application of the kernel estimation method of Rosenblatt et al. (1956) and Parzen (1962). I suggest an estimate of their asymptotic variances inspired by a combination of (i) Newey's (1994) method of variance estimation for kernel estimators in the "monadic" setting and (ii) a variance estimator for the (estimated) density of a simple network first suggested by Holland and Leinhardt (1976). More unusual are the rates of convergence and

asymptotic (normal) distributions of these dyadic density estimates. Specifically, I show that they converge at the same rate as the (unconditional) dyadic sample mean: the square root of the number, N, of nodes. This differs from the results for nonparametric estimation of densities and regression functions for monadic data, which generally have a slower rate of convergence than their corresponding sample mean. Then I study the robustness of the normality-based and the bootstrap-based inference procedures. Since the distribution of this kernel density estimator depends on both the unknown presence/absence of dyadic dependence and the bandwidth choice, successfully approximating its distribution under a wide range of scenarios is both nonstandard and especially desirable. Toward this goal, I first establish the robustness of the normality-based inference by showing that the consistency of variance estimator and asymptotic normal approximation are valid under both dependence regimes with both commonly used and small-bandwidth asymptotics (Cattaneo et al., 2014b). Then, I establish asymptotic inconsistency of a wide class of generalized bootstrap (tailored toward U-statistics) in this setting. Finally, I propose a simple modification of the bootstrap procedure and show its consistency holds robustly. The chapter ends with a semiparametric efficiency bound calculation for density estimation and shows that the kernel density estimator achieves optimal asymptotic variance. Section 1.1, 1.2, 1.3 of this chapter are joint work with Bryan Graham and James Powell.

In Chapter 2 I study nonparametric estimation of regression functions for directed dyadic data. Let $i = 1, \ldots, N$ index a simple random sample of units drawn from some large population. For each unit, researchers observe the vector of regressors X_i and, for each of the $N(N-1)$ ordered pairs of units, an outcome Y_{ij}. The outcomes Y_{ij} and Y_{kl} are independent if their indices are disjoint, but dependent otherwise (i.e., "dyadically dependent"). Let $W_{ij} = \left(X_i', X_j'\right)'$; using the sampled data I seek to construct a nonparametric estimate of the mean regression function $g(W_{ij}) \overset{def}{\equiv} \mathbb{E}\left[Y_{ij} \middle| X_i, X_j\right]$. I present two sets of results. First, I calculate lower bounds on the minimax risk for estimating the regression function at (i) a point and (ii) under the infinity norm. Second, I calculate (i) pointwise and (ii) uniform convergence rates for the dyadic analog of the familiar Nadaraya-Watson (NW) kernel regression estimator. I show that the NW kernel regression estimator achieves the optimal rates suggested by the risk bounds when an appropriate bandwidth sequence is chosen. This optimal rate differs from the one available under iid data: the effective sample size is smaller and $d_W = \dim(W_{ij})$ influences the rate differently. This chapter is joint work with Bryan Graham and James Powell.

In Chapter 3 I study estimation of the density-weighted average derivative for directed dyadic data. This parameter is of substantial practical interest as it is proportional to the coefficients in single index models (Powell et al., 1989), which encompasses various models of limited dependent variables. Besides carefully setting up the directed dyadic single index regression model with both monadic and dyadic explainable variables, the main contributions of this chapter are extending the kernel-based estimator of the density-weighted average derivatives from the "monadic" iid setup (e.g. Stoker, 1986; Powell et al., 1989; Newey and Stoker, 1993) to directed dyadic data and proving its robust asymptotic normality (asymptotic normality holds under both nondegeneracy and degeneracy and across a wide range of bandwidth sequences) using asymptotic quadratic approximation. This robust asymptotic normality result presents an interesting contrast between this kernel-based semiparametric estimator and the sample mean of dyadic data, which exhibits asymptotic non-normality when dyadic dependence is absent and whose uniform nonconservative inference procedure does not exist (Menzel, 2021). This chapter marks the start of my analysis of estimation of semiparametric models for dyadic data, which is continued in the next chapter.

In Chapter 4 I study error components models of dyadic data, of which a major motivation is separating the monadic and dyadic components of variation. The development parallels that of error components with panel data: I progressively enrich the random effect model by going from being without covariates to being with covariates and from homoskedasticity to multiplicative heteroskedasticity. Throughout enriching the models, I focus on estimating the coefficients in a linear regression, which includes both monadic and dyadic explanatory variables. To understand the nature of the estimation problem under different error components models, I study the performance of intuitive OLS estimators, propose more efficient estimators, calculate the asymptotic efficiency bounds (Cramér-Rao lower bound, CRLB), and compare the efficiency bounds to variances of the estimators. Under homoskedasticity, I prove the sample mean, which converges at rate $O(N^{-1/2})$, and least square estimator with double-differencing operation, which converges at rate $O\left(\binom{N}{2}^{-1/2}\right)$, achieve the CRLB and are asymptotically efficient for estimating the marginal expectation and the coefficients of dyadic variables in a linear regression respectively. Under unknown multiplicative heteroskedasticity, I show that the intuitive two-step semiparametric generalized score estimator for estimating the linear regression coefficients, which is a natural extension of the classical feasible generalized least square estimator (FGLS) for linear regression with heteroskedasticity for the "monadic" iid data, is not adaptive to the unknown heteroskedasticity. Its convergence rate is faster than that of the OLS estimator, $O(N^{-1/2})$, but slower than the rate suggested by CRLB, $O\left(\binom{N}{2}^{-1/2}\right)$. This result makes a distinction from a familiar result in the monadic iid setting, i.e. a two-step semiparametric generalized score estimator often indeed achieves adaptivity and CRLB in iid setting. The gap between the performance of the best available estimator and the CRLB suggests that for this estimation problem with dyadic data either there exists a better estimator that is adaptive and achieves the CRLB, or there is a tighter efficiency bound. I point this gap out for further research.

To my parents, Xiuqin Jia and Xiangchen Niu

Contents

Contents **ii**

1 Estimating the Density of A Dyadic Random Variable **1**
- 1.1 Introduction and Summary . . . 1
- 1.2 Setup and Assumptions . . . 3
- 1.3 Kernel Density Estimator: Rate of Convergence, Consistent Variance Estimation, Asymptotic Normality . . . 5
- 1.4 Bootstrap Failure and Robust Generalized Bootstrap . . . 11
- 1.5 Efficient Influence Function and Optimal Asymptotic Variance . . . 17
- 1.6 Appendix: Proofs . . . 19

2 Nonparametric Dyadic Regression **37**
- 2.1 Introduction and Summary . . . 37
- 2.2 Lower Bounds on the Minimax Risk . . . 38
- 2.3 Kernel Estimator of Dyadic Regression . . . 41
- 2.4 Appendix: Proofs . . . 44

3 Density-Weighted Average Derivatives **62**
- 3.1 Introduction and Summary . . . 62
- 3.2 Population and Sampling Framework . . . 63
- 3.3 Estimand, Estimator, and Hoeffding Decomposition . . . 64
- 3.4 Assumptions . . . 67
- 3.5 Asymptotic Quadratic Approximation and Asymptotic Normality . . . 68
- 3.6 Appendix: Proofs . . . 75

4 Error Components Models for Dyadic Data **88**
- 4.1 Introduction and Summary . . . 88
- 4.2 A Variance Decomposition . . . 89
- 4.3 Error Components Regression . . . 90
- 4.4 Cramér-Rao Lower Bound . . . 97
- 4.5 Appendix: Proofs . . . 100

Bibliography **103**

Acknowledgments

First, I want to thank my advisors, Bryan Graham and James Powell, for their invaluable advice and unreserved support. With a full plate of other commitments, Bryan and Jim have never been short on time with advising and providing timely and insightful feedback on my research. Both of them encouraged me and shared many of their experience and insights with me. I have fond memories of many triad meetings we had at either Bryan's office, Jim's office, or the cloud meeting room. Our collaboration has also been very fruitful. The joint research projects are the basis for part of my dissertation's first and second chapter. I have also greatly benefited from conversations with Michael Jansson, who is also on my dissertation committee. It has been a repeated experience that bouncing ideas off Michael leads to new directions and new ideas.

I also want to thank my fellow PhD colleagues and my friends. This includes Eric Auerbach, Pablo Muñoz Henríquez, Junyi Hou, Weijia Li, Maximilian Müller, Maxime Sauzet, Zenan Wang, and Jerry Yunhao Zhang. The friendship and occasionally mutual mentorship with this wonderful group of people are invaluable throughout my PhD journey. I am indebted to Alexander White at Tsinghua University, who is my undergraduate thesis advisor and coauthor. Alex is resourceful and has always been supportive since my undergraduate years.

Finally, I would like to express my gratitude to my parents, Xiuqin Jia and Xiangchen Niu, who have always loved me, trusted me, and encouraged me unconditionally. This dissertation would not have come to life without their support.

Chapter 1

Estimating the Density of A Dyadic Random Variable [1]

1.1 Introduction and Summary

Many important social and economic variables are naturally defined for pairs of agents (or dyads). Examples include trade between pairs of countries (e.g., Tinbergen, 1962), input purchases and sales between pairs of firms (e.g., Atalay et al., 2011), research and development (R&D) partnerships across firms (e.g., König et al., 2019) and friendships between individuals (e.g., Christakis et al., 2020). Dyadic data arises frequently in the analysis of social and economic networks. In economics, such analyses are predominant in, for example, the analysis of international trade flows. See Graham (2020a) for many other examples and references.

While the statistical analysis of network data began almost a century ago, rigorously justified methods of inference for network statistics are only now emerging (cf., Goldenberg et al., 2010). In this chapter, we study nonparametric estimation of the density function of a (continuously-valued) dyadic random variable. Examples included the density of migration across states, trade across nations, liabilities across banks, or minutes of telephone conversation among individuals. While nonparametric density estimation using independent and identically distributed random samples, henceforth "monadic" data, is well-understood, its dyadic counterpart has, to our knowledge, not yet been studied.

Holland and Leinhardt (1976) derived the sampling variance of the link frequency in a simple network (and of other low order subgraph counts). A general asymptotic distribution theory for subgraph counts, exploiting recent ideas from the probability literature on dense graph limits (e.g., Diaconis and Janson, 2008; Lovász, 2012), was presented in Bickel et al. (2011).[2] Menzel (2021) presents bootstrap procedures for inference on the mean of a dyadic random variable. Our focus on nonparametric density estimation appears to be novel.

[1] Section 1.1, 1.2, 1.3 of this chapter are joint work with Bryan Graham and James Powell.

[2] See Nowicki (1991) for a summary of earlier research in this area.

Density estimation is, of course, a topic of intrinsic interest to econometricians and statisticians, but it also provides a relatively simple and canonical starting point for understanding nonparametric estimation more generally.

We show that an (obvious) adaptation of the Rosenblatt et al. (1956) and Parzen (1962) kernel density estimator applies to dyadic data. While our dyadic density estimator is straightforward to define, its rate of convergence and asymptotic sampling properties, depart significantly from its monadic counterpart. Let N be the number of sampled agents and $n = \binom{N}{2}$ the corresponding number of dyads. Estimation is based upon the n dyadic outcomes. Due to dependence across dyads sharing an agent in common, the rate of convergence of our density estimate is (generally) much *slower* than it would be with n i.i.d. outcomes. This rate of convergence is also invariant across a wide range of bandwidth sequences. This property is familiar from the econometric literature on semiparametric estimation (e.g., Powell, 1994). Indeed, from a certain perspective, our nonparametric dyadic density estimate can be viewed as a semiparametric estimator (in the sense that it can be thought of as an average of nonparametrically estimated densities).

We also explore the impact of "degeneracy" – which arises either when dependence across dyads vanishes or when the bandwidth sequence is small – on our sampling theory; such degeneracy features prominently in Menzel's (2021) innovative analysis of inference on dyadic means. To address the concern of degeneracy, we incorporate both the traditional and the small-bandwidth asymptotics (Cattaneo et al., 2014b) and show that variance estimator and inference based on asymptotic normality are consistent both under nondegeneracy and degeneracy.

To further explore the impact of degeneracy on the inference problem, we study the behavior of a class of generalized bootstrap procedures tailored toward U-statistics and show bootstrap failure in our setting. More specifically, bootstrap procedures consistent in the nondegenerate case when the bandwidth is large are strictly conservative in either the degenerate case or cases with small bandwidth. The conservativeness is severe in the sense that the confidence interval based on the bootstrap approximation is at least 40% wider than the percentile interval based on the true distribution of the estimator. This vanilla version of bootstrap-based inference is hence much less robust compared to inference based on normal approximation. We then propose a simple algorithmic fix to get a robust bootstrap and prove its consistency in all scenarios. The robust bootstrap uses a multiple of the original bandwidth to generate bootstrap samples. The multiplier factor depends on a specific weight used in the generalized bootstrap and can be calculated analytically. We think bootstrapping the kernel density estimator provides a relatively simple and canonical starting point for understanding bootstrapping nonparametric estimators for dyadic data more generally.

We then calculate the efficient influence function and the optimal asymptotic variance for estimating the density under degeneracy and verify that the kernel density estimator is asymptotically efficient. The calculation is based on reducing the problem to the traditional iid setting by inspecting a particular submodel and then leverage existing theory on efficient influence function.

In what follows, section 1.2 presents the setup and assumptions. Section 1.3 shows the rate of convergence, consistent variance estimation, and asymptotic normality results for the kernel density estimator. Section 1.4 shows bootstrap failure and presents a robust generalized bootstrap procedure that works. Section 1.5 calculate the efficient influence function and the optimal asymptotic variance. Section 1.6 contains proofs of results in the main text.

1.2 Setup and Assumptions

We use a nonparametric model of dyadic variables. Let $i = 1, \ldots, N$ index a random sample of N individuals. An undirected pair of individuals is called a dyad. $\{A_1, \ldots, A_N\}$ is an individual level random sample and $\{V_{ij}, 1 \le i < j \le N\}$ is a dyad level random sample. We assume these two are independent. $n = \binom{N}{2}$ total number of dyads have their corresponding scalar outcomes $\{W_{ij}, 1 \le i < j \le N\}$, where

$$W_{ij} = W(A_i, A_j, V_{ij}). \tag{1.1}$$

To keep a coherent undirected setup, the function W is assumed to be symmetric w.r.t. its first two arguments, i.e. $W(a_1, a_2, v) = W(a_2, a_1, v)$ for any a_1, a_2, v, and $V_{ji} \equiv V_{ij}$ so that $W_{ji} = W(A_j, A_i, V_{ij}) = W(A_i, A_j, V_{ji})$. The statistician observes $\{W_{ij}, 1 \le i < j \le N\}$ only.

In what follows we directly maintain (1.1), however, it also a consequence of assuming that the infinite graph sampled from is jointly exchangeable (Aldous, 1981; Hoover, 1979). Joint exchangeability of the sampled graph $\mathbf{W} = [W_{ij}]$ implies that

$$[W_{ij}] \stackrel{D}{=} \left[W_{\pi(i)\pi(j)}\right] \tag{1.2}$$

for every $\pi \in \Pi$ where $\pi : \{1, \ldots, N\} \to \{1, \ldots, N\}$ is a permutation of the node indices. Put differently, when node labels have no meaning we have that the "likelihood" of any simultaneous row and column permutation of $\mathbf{W}$ is the same as that of $\mathbf{W}$ itself.[3] See Menzel (2021) for a related discussion.

The object of interest is the marginal density f_W of W_{ij} at point w_0, i.e.

$$f_W(w_0).$$

This marginal density $f_W(w_0)$ of W_{ij} can be estimated using an immediate extension of the kernel density estimator for monadic data first proposed by Rosenblatt et al. (1956) and Parzen (1962):

$$\hat{f}_{W,N}(w) = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} K_{N,h_N,ij}(w_0),$$

[3]For $\mathbf{W} = [W_{ij}]$ the $N \times N$ weighted adjacency matrix and $\mathbf{P}$ any conformable permutation matrix

$$\Pr(\mathbf{W} \le \mathbf{w}) = \Pr(\mathbf{PWP} \le \mathbf{w})$$

for all $\mathbf{w} \in \mathbb{W} = \mathbb{R}^{\binom{N}{2}}$.

where $K_{N,h_N,ij}(w_0) \equiv h_N^{-1} K\left(\frac{w_0 - W_{ij}}{h_N}\right)$. h_N is the vanishing bandwidth and K is a fixed kernel function. h_N is the vanishing bandwidth and K is a fixed kernel function. A discussion of the motivation for the kernel estimator $\hat{f}_W(w)$ and its statistical properties under random sampling (of monadic variables) can be found in Silverman (1986, Chapers 2 & 3).

In the following we will omit subscripts W, N, h_N to save notation when there is no confusion. For example, we will often use K_{ij} to denote $K_{N,h_N,ij}(w_0)$. We will use the following assumptions of the model. They are meant to define the scope of our study.

The kernel density estimator of $f_W(w_0)$ is

$$\hat{f}_{W,N}(w) = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} K_{N,h_N,ij}(w_0),$$

where $K_{N,h_N,ij}(w_0) \equiv h_N^{-1} K\left(\frac{w_0 - W_{ij}}{h_N}\right)$. h_N is the vanishing bandwidth and K is a fixed kernel function. In the following, we will omit subscripts W, N, h_N to save notation when there is no confusion. For example, we will often use K_{ij} to denote $K_{N,h_N,ij}(w_0)$. We will use the following assumptions of the model. They are meant to define the scope of our study.

Assumption 1.2.1 (**Model**). (a) **(Marginal Density)** W_{12} is absolutely continuous with density f_W. $f_W(w_0) > 0$. f_W is continuously differentiable with derivative f'_W. f'_W is L-Lipschitz, i.e. $|f'_W(w_1) - f'_W(w_2)| \leq L|w_1 - w_2|$ for any $w_1,\ w_2 \in \mathbb{R}$.

(b) **(Conditional Density)** Let $f_{W|A}(w|a)$ be the conditional density of W_{ij} given $A_i = a$. Assume the conditional density exists. For all a, $f_{W|A}(w|a)$ is continuously differentiable in w and its first derivative is L-Lipschitz. $f_{W|A}$ is bounded by a constant M, i.e. $\sup_{w,a} f_{W|A}(w|a) \leq M < \infty$.

Note that this model nests two special cases. Special case 1: $W_{ij} = W(A_i, A_j)$ is a deterministic function of individual-level random variables. It does not depend on the dyad level idiosyncratic terms V_{ij}. Special case 2: $W_{ij} = W(V_{ij})$ is a function of the dyad level idiosyncratic terms V_{ij} only. It does not depend on the individual-level terms A_i, A_j. The following analysis will work well for all cases. The key dichotomy of the problem will be

$$\text{Var}\left(f_{W|A}\left(w_0|A_1\right)\right) > 0 \text{ or } \text{Var}\left(f_{W|A}\left(w_0|A_1\right)\right) = 0.$$

We will refer to the case in which $\text{Var}\left(f_{W|A}\left(w_0|A_1\right)\right) > 0$ as the **nondegenerate case** and the case in which $\text{Var}\left(f_{W|A}\left(w_0|A_1\right)\right) = 0$ as the **degenerate case**. To save notation, we will use Ω_1 to denote $\text{Var}\left(f_{W|A}\left(w_0|A_1\right)\right)$ from here on

$$\Omega_1 = \text{Var}\left(f_{W|A}\left(w_0|A_1\right)\right).$$

Here is the reason why this variance drives the nature of the estimation problem. Ω_1 pins down the magnitude of the Hajék projection. If it is zero, the remainder terms of the Hajék projection will be the leading term and the U-type statistic is degenerate. Special case 2 is a representative example of the degenerate case. Since dyadic dependence is absent (in first order) under degeneracy, variance of the estimator and convergence rate could have qualitatively different magnitude. Our analysis will take care of this degeneracy.

We will use the following standard assumptions of the kernel throughout.

Assumption 1.2.2 (Kernel). K integrates to one, i.e. $\int K(u)du = 1$. K is even, i.e. $K(u) = K(-u)$ for any $u \in \mathbb{R}$. K is bounded, i.e. $\sup_u |K(u)| \leq M$. K has bounded support, i.e. $K(u) = 0$ for any $|u| > M$.

Note that symmetry of the kernel implies the kernel is orthogonal to the first-order polynomial, i.e. $\int uK(u)du = 0$. We do not require the use of a higher-order kernel. Given the assumption of boundedness of conditional density, the assumptions of boundedness and bounded support conditions of the kernel ensure the integrability of all generic moments that matter in later analysis. These assumptions are not essential as they can be replaced by directly imposing relevant integrability conditions. For example, Gaussian kernel, which has rapidly vanishing tails, would ensure integrability in many cases. Even though it does not strictly satisfy our assumptions, our results will still hold for it with some rare exceptions.

The purpose of the following assumption is to ensure consistency of the estimator.

Assumption 1.2.3 (Consistency). The sequence of bandwidth h_N satisfies $N^{-2} \ll h_N \ll 1$.

Here $a_N \ll b_N$ means $\frac{a_N}{b_N} \to 0$ as $N \to \infty$ for the two sequences $\{a_N\}$ and $\{b_N\}$. Note this assumption is both necessary and sufficient for consistency. In the following, we will explicitly impose additional bandwidth conditions for relevant results.

1.3 Kernel Density Estimator: Rate of Convergence, Consistent Variance Estimation, Asymptotic Normality

We develop results on the rate of convergence, consistent variance estimation, and asymptotic normality in this section. These results also serve as a warm-up for the discussion of the bootstrap-based inference. Our results incorporate both degenerate case and small-bandwidth range.

Bias, Variance, and Consistency

Examining the bias and especially variance of the estimator is the most instructive way to understand its behavior and appreciate the concern of degeneracy and small bandwidth range. We will present the results first and follow them with the key calculation steps.

Theorem 1.3.1. *If assumptions 1.2.1, 1.2.2, and 1.2.3 hold and* *(a)* (**Bias**) $\mathrm{Bias}\left(\hat{f}_W(w_0)\right) \equiv \mathbb{E}\left[\hat{f}_W(w_0)\right] - f_W(w_0) = O\left(h_N^2\right).$

(b) (**Variance**)

$$\sigma_N^2 \equiv \mathrm{Var}\left(\hat{f}_W(w_0)\right) = \begin{cases} N^{-1}\left(1 - \frac{1}{N-1}\right) 4\Omega_1 + \binom{N}{2}^{-1} h_N^{-1}\Omega_2 + O\left(N^{-2} + N^{-1}h_N^2\right) & \text{if } \Omega_1 > 0 \\ \binom{N}{2}^{-1} h_N^{-1}\Omega_2 + O\left(N^{-2} + N^{-1}h_N^4\right) & \text{if } \Omega_1 = 0 \end{cases}, \tag{1.3}$$

where $\Omega_1 \equiv \mathrm{Var}\left(f^2_{W|A}(w_0|A_1)\right)$ *and* $\Omega_2 \equiv f_W(w_0)\int K^2(u)du.$

(c) (**Consistency**)

$$\mathrm{MSE}\left(\hat{f}_W(w_0)\right) = \begin{cases} O\left(h_N^4 + N^{-1} + N^{-2}h_N^{-1}\right) & \text{if } \Omega_1 > 0 \\ O\left(h_N^4 + N^{-1}h_N^4 + N^{-2}h_N^{-1}\right) & \text{if } \Omega_1 = 0 \end{cases}. \tag{1.4}$$

And the estimator is consistent $\hat{f}_W(w_0) \xrightarrow{\mathrm{P}} f_W(w_0)$.

Proofs can be found in the appendix. Here is the sketch of the variance calculation, which resembles variance calculation of a U-statistic.

$$\begin{aligned} \mathrm{Var}\left(\hat{f}_W(w_0)\right) &= \mathrm{Var}\left(\binom{N}{2}^{-1}\sum_{i<j} K_{ij}\right) \\ &= \binom{N}{2}^{-2}\sum_{i_1<j_1}\sum_{i_2<j_2}\mathrm{Cov}\left(K_{i_1j_1}, K_{i_2j_2}\right) \\ &= \binom{N}{2}^{-2}\left[\binom{N}{2}\mathrm{Var}\left(K_{12}\right) + N(N-1)(N-2)\,\mathrm{Cov}\left(K_{12}, K_{13}\right)\right] \\ &= N^{-1}(N-1)^{-1}\left(N-2\right)4\ \Omega_{1,N} + N^{-1}(N-1)^{-1}h_N^{-1}2\Omega_{2,N}, \end{aligned} \tag{1.5}$$

where $\Omega_{2,N} \equiv h_N \mathrm{Var}\left(K_{12}\right)$ converges to Ω_2 and $\Omega_{1,N} \equiv \mathrm{Cov}\left(K_{12}, K_{13}\right)$ converges to Ω_1. The third equality follows elementary counting and the fact that $K_{i_1j_1}$ and $K_{i_2j_2}$ sharing zero index are independent.

To further appreciate the results in the nondegenerate case, provided that $\Omega_1(w_0) \neq 0$ (i.e. the nondegenerate case) and the bandwidth sequence h_N is chosen such that

$$Nh \to \infty, \qquad Nh^4 \to 0 \tag{1.6}$$

as $N \to \infty$, we get that

$$\begin{aligned} \mathrm{MSE}\left(\hat{f}_W(w_0)\right) &= o\left(\frac{1}{N}\right) + O\left(\frac{1}{N}\right) + o\left(\frac{1}{N}\right) \\ &= O\left(\frac{1}{N}\right), \end{aligned}$$

and hence that

$$\sqrt{N}(\hat{f}_W(w_0) - f_W(w_0)) = O_p(1).$$

In fact, the rate of convergence of $\hat{f}_W(w_0)$ to $f_W(w_0)$ will be $\sqrt{N}$ as long as $Nh^4 \leq C \leq Nh$ for some $C > 0$ as $N \to \infty$, although the mean-squared error will include an additional bias or variance term of $O(N^{-1})$ if either Nh or $(Nh^4)^{-1}$ does not diverge to infinity.

To derive the MSE-optimal bandwidth sequence, we minimize (1.4) with respect to its first and third terms, this yields an optimal bandwidth sequence of

$$h_N^* = O\left(N^{-\frac{2}{5}}\right).$$

This sequence satisfies condition (1.6) above.

Interestingly, the rate of convergence of $\hat{f}_W(w_0)$ to $f_W(w_0)$ under condition (1.6) is the same as the rate of convergence of the sample mean

$$\bar{W} \overset{def}{\equiv} \frac{1}{n}\sum_{i<j} W_{ij} \tag{1.7}$$

to its expectation $\mu_W \overset{def}{\equiv} \mathbb{E}[W_{ij}]$ when $\mathbb{E}[W_{ij}^2] < \infty$. Similar variance calculations to those for $\hat{f}_w(w_0)$ yield (see also Holland and Leinhardt (1976) and Menzel (2021))

$$\begin{aligned}\mathrm{Var}(\bar{W}) &= O\left(\frac{\mathrm{Var}(W_{ij})}{n}\right) + O\left(\frac{4\,\mathrm{Var}(\mathbb{E}[W_{ij}|A_i])}{N}\right)\\ &= O\left(\frac{1}{N}\right),\end{aligned}$$

provided $\mathbb{E}[W_{ij}|A_i]$ is non-degenerate, yielding

$$\sqrt{N}(\bar{W} - \mu) = O_p(1).$$

Thus, in contrast to the case of i.i.d monadic data, there is no convergence-rate “cost” associated with nonparametric estimation of $f_W(w_0)$. The presence of dyadic dependence, due to its impact on estimation variance, does slow down the feasible rate of convergence substantially. With iid data, the relevant rate for density estimation would be $n^{2/5}$ when the MSE-optimal bandwidth sequence is used. Recalling that $n = O(N^2)$, the $\sqrt{N}$ rate we find here corresponds to an $n^{1/4}$ rate. The slowdown from $n^{2/5}$ to $n^{1/4}$ captures the rate of convergence costs of dyadic dependence on the variance of our density estimate.

The lack of dependence of the convergence rate of $\hat{f}_W(w_0)$ to $f_W(w_0)$ on the precise bandwidth sequence chosen is analogous to that for semiparametric estimators defined as averages over nonparametrically-estimated components (e.g., Newey, 1994). Defining $K_{ji} \overset{def}{\equiv} K_{ij}$, the estimator $\hat{f}_W(w_0)$ can be expressed as

$$\hat{f}_W(w_0) = \frac{1}{N}\sum_{i=1}^{N} \hat{f}_{W|A}(w_0|A_i),$$

where

$$\hat{f}_{W|A}(w_0|A_i) \overset{def}{\equiv} \frac{1}{N-1}\sum_{j\neq i, j=1}^{N} K_{ij}.$$

Holding i fixed, the estimator $\hat{f}_{W|A}(w_0|A_i)$ can be shown to converge to $f_{W|A}(w_0|A_i)$ at the nonparametric rate $\sqrt{Nh}$, but the average of this nonparametric estimator over A_i converges at the faster (parametric) rate $\sqrt{N}$. In comparison, while

$$\bar{W} = \frac{1}{N}\sum_{i=1}^{N} \hat{\mathbb{E}}\left[W_{ij}|\, A_i\right],$$

for

$$\hat{\mathbb{E}}\left[W_{ij}|\, A_i\right] \overset{def}{\equiv} \frac{1}{N-1}\sum_{j\neq i, j=1}^{N} W_{ij},$$

the latter converges at the parametric rate $\sqrt{N}$, and the additional averaging to obtain $\bar{W}$ does not improve upon that rate.

In the degenerate case (i.e. $\Omega_1 = 0$), the convergence rate of the estimator could be faster than $\sqrt{N}$ because the covariance of terms sharing a single index $\Omega_{1,N} = \mathrm{Cov}(K_{ij}, K_{il})$ vanishes to zero as N goes to infinity. In the extreme case where $W_{ij} = W(A_i, A_j, V_{ij}) = W^*(V_{ij})$, W_{ij}s are independent from each other and the density estimation problem of "dyadic" variables reduces to the standard density estimation problem of n iid variables. In this case, the problem is truely "nonparametric" in the sense that we rediscover the optimal nonparametric rate of convergence $\mathrm{MSE}(\hat{f}_W(w_0)) = O(n^{-\frac{4}{5}})$ by setting the optimal bandwidth $h_N^* = O\left(N^{-2/5}\right)$

Variance Estimation

To quantify the uncertainty of the estimator $\hat{f}_W(w_0)$, we need a consistent estimator of its variance. Motivated by the variance expression $\sigma_N^2 = \binom{N}{2}^{-2}\sum_{i_1<j_1}\sum_{i_2<j_2} d(i_1,j_1,i_2,j_2)\,\mathrm{Cov}\left(K_{i_1j_1}, K_{i_2j_2}\right)$, where $d(i_1,j_1,i_2,j_2) = \mathbb{1}\left(i_1 = i_2 \text{ or } i_1 = j_2 \text{ or } j_1 = i_2 \text{ or } j_1 = j_2\right)$, we propose the following analog estimator

$$\hat{\sigma}_N^2 = \binom{N}{2}^{-2}\sum_{i_1<j_1}\sum_{i_2<j_2} d(i_1,j_1,i_2,j_2)\left(K_{i_1j_1} - \hat{f}_W(w_0)\right)\left(K_{i_2j_2} - \hat{f}_W(w_0)\right).$$

The following theorem gives us a consistency result that holds under degeneracy and small bandwidth range.

Theorem 1.3.2 (Consistent Variance Estimation). *If assumptions 1.2.1, 1.2.2, 1.2.3 hold, and $h_N \ll N^{-1/4}$, then*

$$\frac{\hat{\sigma}_N^2}{\sigma_N^2} = 1 + o_P(1).$$

Notice this theorem holds under a broad range of bandwidth sequences $N^{-2} \ll h_N \ll N^{-1/4}$ for both degenerate and non-degenerate case. The proof starts by rewriting the estimator as

$$\begin{aligned}\hat{\sigma}_N^2 &= \binom{N}{2}^{-2} \sum_{i_1<j_i} \sum_{i_2<j_2} d(i_1, j_1, i_2, j_2) \left(K_{i_1 j_1} - \hat{f}_W(w_0)\right) \left(K_{i_2 j_2} - \hat{f}_W(w_0)\right) \\ &= N^{-1}(N-1)^{-1}(N-2)4\, \hat{\Omega}_{1,N} + N^{-1}(N-1)^{-1} h_N^{-1} 2\, \hat{\Omega}_{2,N}, \end{aligned} \tag{1.8}$$

where

$$\begin{aligned}\hat{\Omega}_{1,N} &\equiv \binom{N}{3}^{-1} \sum_{i<j<k} \left(\frac{K_{ij}K_{ik} + K_{ij}K_{jk} + K_{ik}K_{jk}}{3} \right) - \hat{f}_W^2(w_0) \\ \hat{\Omega}_{2,N} &\equiv \binom{N}{2}^{-1} \sum_{i<j} h_N K_{ij}^2 - h_N \hat{f}_W^2(w_0).\end{aligned}$$

This expression mirrors the variance expression (1.5). It proceeds by showing the discrepancy $\hat{\Omega}_{1,N} - \Omega_1$ and $\hat{\Omega}_{2,N} - \Omega_2$ goes to zero fast enough under the assumptions on the bandwidth sequence. See the detail of the proof in the appendix.

Asymptotic Normality

The following theorem ensures asymptotic normality of $\hat{f}_W(w_0)$ for both nondegenerate and degenerate cases and for both standard and small bandwidth range.

Theorem 1.3.3 (Asymptotic Normality). *If assumptions 1.2.1, 1.2.2, 1.2.3 hold, and $h_N \ll N^{-2/5}$, then*

$$\sigma_N^{-1} \left(\hat{f}_W(w_0) - f_W(w_0) \right) \rightsquigarrow N(0, 1).$$

The proof of theorem 1.3.3 uses martingale CLT. The martingale structure is constructed by arranging terms in Hoeffding decomposition carefully.

$$
\begin{aligned}
\hat{f}_W(w_0) - \mathbb{E}\hat{f}_W(w_0) &= \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (K_{ij} - \mathbb{E}K_{ij}) \\
&= N^{-1} \sum_{i} 2\left(\mathbb{E}\left[K_{ij}|A_i\right] - \mathbb{E}K_{ij}\right) \\
&\quad + \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left(K_{ij} - \mathbb{E}\left[K_{ij}|A_i\right] - \mathbb{E}\left[K_{ij}|A_j\right] + \mathbb{E}K_{ij}\right) \\
&= \sum_{i=1}^{N} \left[N^{-1} 2D_i + \binom{N}{2}^{-1} \sum_{j=1}^{i-1} E_{ji} \right] \\
&= \sum_{i=1}^{N} Y_{N,i} \ ,
\end{aligned} \tag{1.9}
$$

where $D_i \equiv \mathbb{E}\left[K_{ij}|A_i\right] - \mathbb{E}K_{ij}$, $E_{ij} \equiv K_{ij} - \mathbb{E}\left[K_{ij}|A_i\right] - \mathbb{E}\left[K_{ij}|A_j\right] + \mathbb{E}K_{ij}$, and $Y_{N,i} \equiv N^{-1}2D_i + \binom{N}{2}^{-1} \sum_{j=1}^{i-1} E_{ji}$. The first equality is by definition. The second equality is the Hoeffding decomposition of U-statistics. The third and fourth equality reveals the martingale structure. More precisely, let $\mathcal{F}_i \equiv \sigma\left(A_1, \ldots, A_i, V_{kl}, 1 \leq k < l \leq i\right)$ be a filtration. Then $\mathbb{E}\left[D_i|\mathcal{F}_{i-1}\right] = 0$, $\mathbb{E}\left[E_{ji}|\mathcal{F}_{i-1}\right] = 0, \forall j < i$, hence $\mathbb{E}\left[Y_{N,i}|\mathcal{F}_{i-1}\right] = 0$. The sequence $(Y_{N,i}, i = 1, \ldots, N)$ therefore is a martingale adapted to $(\mathcal{F}_i, i = 1, \ldots, N)$. Applying the martingale CLT in Hall and Heyde (1980) gives the desired normality result for the centered statistic $\hat{f}_W(w_0) - \mathbb{E}\hat{f}_W(w_0)$. The under-smoothing bandwidth assumption $h_N \ll N^{-2/5}$ ensures the bias $\mathbb{E}\hat{f}_W(w_0) - f_W(w_0)$ is of smaller order. Together these implied the asymptotic normality result. See the proof in the appendix.

To better appreciate the result, remember Menzel (2021) shows that under degeneracy, the limit distribution of the sample mean, $\bar{W}$, equation (1.7), may be non-Gaussian. This occurs because the second-order terms in the Hoeffding decomposition of $\bar{W}$ dominate the variation under degeneracy and could be asymptotically non-Gaussian (as is familiar from the theory of U-Statistics, e.g., Chapter 12 of Van der Vaart (2000)).

The situation is both more complicated and simpler here. In the case of the estimated density $\hat{f}_W(w_0)$, due to the presence of the vanishing bandwidth, Liaponuv condition, an important condition for the martingale CLT

$$
\sigma_N^{-4} \sum_{i=1}^{N} \mathbb{E}\left[Y_{N,i}^4\right] \xrightarrow{\text{P}} 0,
$$

continues to hold under degeneracy. It follows then $\frac{1}{\sigma_N}\left(\hat{f}_W(w_0) - f_W(w_0)\right)$ continues to be normal in the limit. This result ensures the normality based inference methods is robust. Namely, the following Wald-based confidence interval

$$
\text{CI}(f_W(w_0)) := [\hat{f}_W(w_0) - \Phi^{-1}(1-\alpha/2) \cdot \hat{\sigma}_N, \hat{f}_W(w_0) + \Phi^{-1}(1-\alpha/2) \cdot \hat{\sigma}_N],
$$

where Φ is the cdf of standard normal, is asymptotically consistent with coverage probability $1-\alpha$ under the assumptions of theorem 1.3.3.

1.4 Bootstrap Failure and Robust Generalized Bootstrap

In this section, we review a general class of weighted bootstrap for U-statistic, show its inconsistency for estimating the distribution of our estimator by a detailed variance calculation, and propose a simple fix. We then show the bootstrap consistency of the modified bootstrap procedure. The corresponding bootstrap statistic is preferable both analytically and computationally.

Generalized Bootstrap

Bootstrapping U-statistics has a well-developed line of literature (see e.g. Bickel and Freedman (1981), Arcones and Gine (1992), Janssen (1994), Mccullagh (2000), Owen (2007), Menzel (2021), Bose and Chatterjee (2018), and Levin and Levina (2019)). A unifying class of weighted bootstrap statistics proposed by Bose and Chatterjee (2018) encompasses most state-of-the-art specific proposals. To be concrete, let's consider a generic second-order U-statistic $U_N = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} H_N(X_i, X_j)$. Its corresponding weighted bootstrap statistic is then

$$U_N^{\Delta} = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \Delta_{N,ij} H_N(X_i, X_j),$$

where the weights satisfy the following assumption.

Assumption 1.4.1 (Weights). $(\Delta_{N,ij}, 1 \leq i < j \leq N) \in \mathbb{R}^{\binom{N}{2}}$ is a vector of random weights independent from the observed data. It satisfies

(a) Exchangeability in the sense that $(\Delta_{N,\pi(1)\pi(2)}, \ldots, \Delta_{N,\pi(1)\pi(N)}, \Delta_{N,\pi(2)\pi(3)}, \ldots, \Delta_{N,\pi(2)\pi(N)}, \ldots, \Delta_{\pi(N-1)\pi(N)})$ has the same distribution for all permutation π of $\{1, \ldots, N\}$

(b) $\mathbb{E}\Delta_{N,12} = \mu_N \to 1$ as $N \to \infty$

(c) $\mathrm{Var}(\Delta_{N,12}) = \upsilon_{N,2} \to \upsilon_2$, $\mathrm{Cov}(\Delta_{N,12}, \Delta_{N,13}) = \upsilon_{N,1} \to \upsilon_1$, $\mathrm{Cov}(\Delta_{N,12}, \Delta_{N,34}) = \upsilon_{N,0} \to 0$, where $\upsilon_2, \upsilon_1 > 0$ are positive constants.

Multiple ways of extending bootstrap statistic of sample mean in iid setting to U-statstic setting are special cases lying in this class. Product weights of the form $\Delta_{N,ij} = \Delta_{N,i}\Delta_{N,j}$ with $(\Delta_{N,1}, \ldots, \Delta_{N,N}) \sim \mathrm{Mult}(N, \frac{1}{N}, \ldots, \frac{1}{N})$ is an analog of Efron's nonparametric bootstrap. An analog of Bayesian bootstrap is $(\Delta_{N,1}, \ldots, \Delta_{N,N}) \sim N \cdot \mathrm{Dirichlet}(N, 1, \ldots, 1)$. Both cases satisfy assumption 4 with $\upsilon_2 = 3, \upsilon_1 = 1$. Additive weights of the form $\Delta_{N,ij} = \Delta_{N,i} + \Delta_{N,j} - 1$

with $(\Delta_{N,1}, \ldots, \Delta_{N,N}) \sim \mathrm{Mult}(N, \frac{1}{N}, \ldots, \frac{1}{N})$ corresponds to pigeon-hole bootstrap. This case satisfies assumption 4 with $\upsilon_2 = 2, \upsilon_1 = 1$.

Inconsistency of Generalized Bootstrap

We will write down the generalized bootstrap statistic of the centered kernel density estimator and show its inconsistency by comparing its conditional variance to the variance of the kernel density estimator. More specifically, bootstrap consistent in the nondegenerate case when the bandwidth is large is strictly conservative in either the degenerate case or cases with small bandwidth. The conservativeness is severe in the sense that the confidence intervals based on these bootstrap approximations are at least 40% wider than the percentile interval based on the true distribution of the estimator.

As notation we use $P, \mathbb{E}, \mathrm{Var}$ to denote probability, expectation, variance under $\mathbb{P}$ and use $P^*, \mathbb{E}^*, \mathrm{Var}^*$ to denote the conditional probability, expectation, variance under the conditional probability given $\mathcal{F}_N = \sigma\left(A_i, V_{jk}, 1 \leq i \leq N,\ 1 \leq j < k \leq N\right)$.

Consider the following exchangeably weighted bootstrap statistics of the centered statistic $\hat{f}_N - \mathbb{E}\hat{f}_N$

$$\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right)^{\Delta} = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \Delta_{N,ij}\left(K_{N,h_N,ij} - \bar{K}_{N,h_N,\cdot\cdot}\right),$$

We use $\bar{K}_{N,h_N,\cdot\cdot}$ to denote $\hat{f}_N$ in this context in order to emphasize it is a sample average.

The most instructive way to understand the behavior of the weighted bootstrap statistic is to inspect its conditional variance.

Lemma 1.4.1. *If assumptions 1.2.1, 1.2.2, 1.2.3, 1.4.1 hold and $h_N \ll N^{-1/4}$, then*

$$\mathrm{Var}^*\left(\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right)^{\Delta}\right) = \left(N^{-1}\left(1 - \frac{1}{N-1}\right)4\Omega_1 \cdot \upsilon_1 + \binom{N}{2}^{-1} h_N^{-1}\Omega_2 \cdot \upsilon_2\right) \cdot (1 + o_p(1)). \tag{1.10}$$

The proof of the result is in the appendix. The proof uses basic variance expansion and convergence results used in the proof of theorem 1.3.2, consistency of the variance estimator.

Armed with this lemma and theorem 1.3.3, we are ready to state the the bootstrap inconsistency result.

Theorem 1.4.2 (**Bootstrap Inconsistency**)**.** *If assumptions 1.2.1, 1.2.2, 1.2.3, 1.4.1 hold and $h_N \ll N^{-1/4}$, and if the bootstrap is consistent, i.e. $\sigma_N^{-1}\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right)^{\Delta} | \mathcal{F}_N \rightsquigarrow_p N(0,1)$, in the nondegenerate case when $h_N \gg N^{-1}$, then it is* ***in****consistent both in the nondegenerate case when $h \leq\leq N^{-1}$ and in the degenerate case.*

Proof. A necessary condition for bootstrap consistency is the ratio of the conditional variance $\mathrm{Var}^*\left(\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right)^{\Delta}\right)$ in equation (1.10) to the variance $\sigma_N^2 = \mathrm{Var}\left(\hat{f}_{N,h_N}(w_0)\right)$ in equation (1.3)

$$\left(N^{-1}\left(1-\frac{1}{N-1}\right)4\Omega_1\cdot\upsilon_1+\binom{N}{2}^{-1}h_N^{-1}\Omega_2\cdot\upsilon_2\right)\Big/\left(N^{-1}\left(1-\frac{1}{N-1}\right)4\Omega_1+\binom{N}{2}^{-1}h_N^{-1}\Omega_2\right)$$

should converge in probability to 1. Both equations hold under the assumptions. We can check for what kind of weight regimes the limit of the ratio is 1 under different scenarios.

In the nondegenerate case when $h_N \gg N^{-1}$, the limit of the ratio being 1 would imply $\upsilon_1 = 1$. In either the nondegenerate case when $h_N \leq\leq N^{-1}$ or the degenerage case, the limit of the ratio being 1 would imply $\upsilon_2 = 1$, However, $\upsilon_1 = 1$ and $\upsilon_2 = 1$ can never happen together. To see this, notice by the definition of υ_1 and υ_2, for any constant integer k,

$$\begin{pmatrix}\upsilon_2 & \upsilon_1 & & \\ \upsilon_1 & \upsilon_2 & \ddots & \\ & \ddots & \ddots & \upsilon_1 \\ & & \upsilon_1 & \upsilon_2\end{pmatrix}_{k\times k} = \lim_{N\to\infty}\begin{pmatrix}\upsilon_{N,2} & \upsilon_{N,1} & \upsilon_{N,0} & \cdots & \cdots & \upsilon_{N,0} \\ \upsilon_{N,1} & \upsilon_{N,2} & \upsilon_{N,1} & \ddots & \ddots & \upsilon_{N,0} \\ \upsilon_{N,0} & \upsilon_{N,1} & \upsilon_{N,2} & \ddots & \ddots & \upsilon_{N,0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \upsilon_{N,1} \\ \upsilon_{N,0} & \upsilon_{N,0} & \upsilon_{N,0} & \cdots & \upsilon_{N,1} & \upsilon_{N,2}\end{pmatrix}_{k\times k} = \lim_{N\to\infty}\mathrm{Var}\begin{pmatrix}\Delta_{N,12} \\ \Delta_{N,23} \\ \vdots \\ \Delta_{N,kk+1}\end{pmatrix}$$

is positive semidefinite. Specificaly, $k\upsilon_2 - 2(k-1)\upsilon_1 = \lim_{N\to\infty}\mathrm{Var}\left(\sum_{i=1}^{k}(-1)^i\Delta_{ii+1}\right) \geq 0$. This means $\upsilon_2 \geq (2-\frac{2}{k})\upsilon_1$ for any positive integer k. Taking the limit gives us $\upsilon_2 \geq 2\upsilon_1$. This implies for example, if $\upsilon_1 = 1$, then $\upsilon_2 \geq 2$ and $\upsilon_2 \neq 2$. (This bound is sharp since it is attained by for example additive multinomial weights.) This means there exists no weight regimes satisfying bootstrap consistency simultaneously in both cases. □

Here is the intuition behind this bootstrap failure. The variance $\mathrm{Var}\left(\hat{f}_{N,h_N}(w_0)\right)$ is a summation of dyad specific variance and the covariance due to node sharing as shown in equation (1.5). In the nondegenerate case when $h_N \gg N^{-1}$, the covariance due to node sharing is dominating. Bootstrap consistency in this case requires the bootstrap weights to exhibit a similar dependence structure with covariance due to node sharing. As a side effect, this class of dependent weights will necessarily capture too much dyad specific variance. This is not a problem in this case per se because the contribution of dyad specific variance is infinitesimal in the limit. However, this is a serious problem both in the nondegenerate case when $h_N \leq\leq N^{-1}$ and in the degenerate case, where the contribution of dyad specific variance is first-order. A serious upward bias will show up. In this situation, the "best" weights would be those achieving the lower bound $\upsilon_2 = 2\upsilon_1$, e.g. additive weights, so that this upward bias is the smallest.

Roughly speaking, bootstrap consistency in the first case requires weights to exhibits nondegenerate U-type dependence in the limit; bootstrap consistency in the second case requires weights to exhibits independence (or equivalently degenerate U-type dependence) in the limit. The set of weights satisfying both is empty.

Robustness in the sense of consistency across all cases is not achievable by any weight regimes. However, it is achievable by altering the bootstrapped terms through altering the bandwidth used to construct them. This idea first shows up in a remark in Cattaneo et al. (2014a).

Modification: Bootstrapping with a Constant Multiple Bandwidth

We propose a simple algorithmic fix to get a robust bootstrap consistent in all scenarios. This robust bootstrap uses a multiple of the original bandwidth to generate bootstrap samples. The multiplier factor depends on a specific weight used in the generalized bootstrap and can be calculated analytically.

Consider the following modified bootstrap statistic

$$\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta} = \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \Delta_{N,ij} \left(K_{N,h'_N,ij} - \bar{K}_{N,h'_N,\cdot\cdot}\right).$$

This is exactly the previously discussed bootstrap statistic with a different bandwidth, h'_N.

$$h'_N \equiv \frac{\upsilon_{2,N}}{\upsilon_{1,N}} h_N$$

is a multiple of h_N with a known weight-specific multiplier $\frac{\upsilon_{2,N}}{\upsilon_{1,N}}$. Plugging this bandwidth into equation (1.10) gives us the conditional variance

$$\begin{aligned}\mathrm{Var}^*\left(\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta}\right) &= \left(N^{-1}\left(1 - \frac{1}{N-1}\right) 4\Omega_1 \cdot \upsilon_1 + \binom{N}{2}^{-1} h_N^{-1} \Omega_2 \cdot \upsilon_2 \cdot \frac{\upsilon_{1,N}}{\upsilon_{2,N}}\right) \\ &\quad \cdot (1 + o_p(1)) \\ &= \mathrm{Var}\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right) \cdot \upsilon_1.\end{aligned}$$

This means this bootstrap statistic has consistent conditional variance simultaneously across all cases for weights with $\upsilon_1 = 1$. Enlarging the bandwidth by a ratio of $\frac{\upsilon_{2,N}}{\upsilon_{1,N}}$ reduces the dyad specific variance while retains the covariance contribution. This precisely cancels out the previously discussed upward bias. We summarize this result in the following lemma.

Lemma 1.4.3. *If assumptions 1.2.1, 1.2.2, 1.2.3, 1.4.1 hold and* $h_N \ll N^{-1/4}$,*, and* $\upsilon_1 = 1$, *then* $\mathrm{Var}^*\left(\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta}\right) = \mathrm{Var}\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right) \cdot (1 + o_p(1))$.

Robust Bootstrap Consistency

We will show consistency of the modified bootstrap statistic with the following additive weights.

Assumption 1.4.2 (Additive Multinomial Weights). The weights $(\Delta_{N,ij}, 1 \leq i < j \leq N)$ is of the additive form $\Delta_{N,ij} = \Delta_{N,i} + \Delta_{N,j} - 1$ with $(\Delta_{N,1}, \ldots, \Delta_{N,N}) \sim \text{Mult}(N, \frac{1}{N}, \ldots, \frac{1}{N})$.

Bootstrap with additive multinomial weights enjoys both analytical convenience and computational efficiency due to the following algebraic relation.

$$\begin{aligned}
\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta} &= \binom{N}{2}^{-1} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} (\Delta_{N,i} + \Delta_{N,j} - 1) \left(K_{N,h'_N,ij} - \bar{K}_{N,h'_N,\cdot\cdot}\right) \\
&= N^{-1} \sum_{i=1}^{N} \Delta_{N,i} 2 \left(\bar{K}_{N,h'_N,i\cdot} - \bar{K}_{N,h'_N,\cdot\cdot}\right), \qquad (1.11)
\end{aligned}$$

where $\bar{K}_{N,h'_N,i\cdot} \equiv (N-1)^{-1} \sum_{j \neq i} K_{N,h'_N,ij}$.

Expression (1.11) reveals the same probabilistic structure as the well-studied nonparametric bootstrap in iid setting. The bootstrapped object here is the empirical Hájek projection in the sense that $2\left(\bar{K}_{N,h'_N,i\cdot} - \bar{K}_{N,h'_N,\cdot\cdot}\right)$ in the expression is an empirical analog of terms $2\left(\mathbb{E}\left(K_{ij}|A_i\right) - \mathbb{E}\left(K_{ij}\right)\right)$ in the actual Hájek projection (1.9). This interpretation is intuitively appealing in its own right. Moreover, this expression enables directly applying probability results of the iid setting, which simplifies the asymptotic analysis.

In terms of computation, expression (1.11) implies that the time complexity and space complexity of sampling this bootstrap statistic B times are $O\left(N(B+1) + N^2\right)$ and $O(N)$ as shown in Bose and Chatterjee (2018). Those of bootstrap with general weights are $O\left(N^2(B+1)\right)$ and $O\left(N^2\right)$, both of which are significantly improved under additivity.

In what follows we will prove consistency of using the distribution of $\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta}$ to estimate the distribution of the original estimation error $\hat{f}_{N,h_N} - f$. The key technical tool behind this proof is what we call conditional Lindeberg-Feller central limit theorem.

Theorem 1.4.4 (Conditional Lindeberg-Feller CLT). *Let $\mathcal{F}_N$ be a sequence of $\sigma-$algebra. If*

1. *$X_{N,1}, \ldots, X_{N,k_N}$ are random variables conditionally independent given $\mathcal{F}_N$ with $\mathbb{E}\left[X_{N,m}|\mathcal{F}_N\right] = 0,\ m = 1, 2, \ldots, k_N$*

2. *$\sum_{i=1}^{k_N} \mathbb{E}\left[X_{N,i}^2|\mathcal{F}_N\right] \rightsquigarrow_p 1$*

3. *For all $\epsilon > 0$, $\sum_{i=1}^{k_N} \mathbb{E}\left[|X_{N,i}|^2 \mathbb{1}\left(|X_{N,i}| > \epsilon\right)|\mathcal{F}_N\right] \rightsquigarrow_p 0$,*

then $S_n = \sum_{i=1}^{k_n} X_{n,i}$ *has* $S_n|\mathcal{F}_n \rightsquigarrow N(0,1)$ *in the sense that for any* $t \in \mathbb{R}$, $P\left(S_n \le t|\mathcal{F}_n\right) \rightsquigarrow_p \Phi\left(t\right)$.

This theorem says if the triangular array $X_{n,1},\dots,X_{n,k_n}$ satisfies the conditions of the classic Lindeberg-Feller CLT conditional on some information set $\mathcal{F}_N$, then the sum $S_n = \sum_{i=1}^{k_n} X_{n,i}$ given $\mathcal{F}_N$ weakly converge to the standard normal distribution in probability. The proof of the theorem relies on a subsequence argument.

Proof. Denote $Q_N(t) \equiv P\left(S_N \le t|\mathcal{F}_N\right)$, $V_N = \sum_{i=1}^{k_N} \mathbb{E}\left[X_{N,k_N}^2|\mathcal{F}_N\right]$, and $T_N(\epsilon) = \sum_{i=1}^{k_N} \mathbb{E}\left[|X_{N,i}|^2 \mathbb{1}\left(|X_{N,i}| > \epsilon\right)|\mathcal{F}_N\right]$. We will prove that for any $t \in \mathbb{R}$, $Q_N(t) \rightsquigarrow_p \Phi\left(t\right)$ by a subsequence argument. Since for any fixed t, $Q_N(t) \rightsquigarrow_p \Phi\left(t\right)$ if and only if for every subsequence $Q_{N(m)}(t)$ there is a further subsequence $Q_{N(m_k)}(t)$ that converges almost surely to $\Phi\left(t\right)$, we can focus on proving the latter in order to prove the former.

For any fixed sequence of indices $N(m)$, condition 2 and 3 imply that there is a subsequence $N(m_k)$ and E with $P(E) = 1$ s.t. $V_{N(m_k)} \to 1$ and for all $\epsilon > 0$, $T_{N(m_k)}(\epsilon) \to 0$ on E. Lindeberg-Feller CLT for $Q_{N(m_k)}$ on E implies $Q_{N(m_k)}(t) \to \Phi\left(t\right)$ on E. In other word, $Q_{N(m_k)}(t) \xrightarrow{a.s.} \Phi\left(t\right)$. We've proved the latter. Hence, for any $t \in \mathbb{R}$, $Q_N(t) \rightsquigarrow_p \Phi\left(t\right)$. □

To use the conditional Lindeberg-Feller CLT, the first thing we do is writing the statistic $\sigma_N^{-1}\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta}$ as a sum of conditionally independent random variables given realization of the data.

$$\sigma_N^{-1}\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta} = \sum_{i=1}^{N} \sigma_N^{-1} 2\left(\bar{K}^*_{N,h'_N,i\cdot} - \bar{K}_{N,h'_N,\cdot\cdot}\right), \tag{1.12}$$

where $\bar{K}^*_{i\cdot}$s are independent uniform draws with replacement from $\bar{K}_{1\cdot}, \bar{K}_{2\cdot}, \dots, \bar{K}_{N\cdot}$. This rewriting uses the special structure of multinomial weights. The RHS term has similar stochastic nature as a generic statistic from nonparametric bootstrap in the iid setting.

Theorem 1.4.5 (Robust Bootstrap Consistency). *If assumptions 1.2.1, 1.2.2, 1.2.3, 1.4.1, 1.4.2 hold and* $h_N \ll N^{-2/5}$, *then*

$$\sigma_N^{-1}\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta}\Big|\mathcal{F}_N \rightsquigarrow N(0,1), \tag{1.13}$$

which together with the asymptotic normality of the original estimator implies bootstrap consistency $\sup_{t\in\mathbb{R}}\left|P^*\left(\sigma_N^{-1}\left(\hat{f}_{N,h'_N} - \mathbb{E}\hat{f}_{N,h'_N}\right)^{\Delta} < t\right) - P\left(\sigma_N^{-1}\left(\hat{f}_{N,h_N} - f\right) < t\right)\right| \xrightarrow{P} 0.$

Proof. We will verify the three conditions of theorem 1.4.4 (conditional Lindeberg-Feller CLT) to prove the weak convergence of (1.12). The conditioning information here is $\mathcal{F}_N = \sigma\left(A_i, V_{jk}, 1 \le i \le N,\ 1 \le j < k \le N\right)$. The triangular array is $X_{N,i} = \sigma_N^{-1} 2\left(\bar{K}^*_{N,h'_N,i\cdot} - \bar{K}_{N,h'_N,\cdot\cdot}\right)$

The first condition, conditional independence and mean zero, holds by the definition of $\bar{K}^*_{N,h'_N,i\cdot}$s and $\mathbb{E}\left(\bar{K}^*_{N,h'_N,i\cdot} - \bar{K}_{N,h'_N,\cdot\cdot}|\mathcal{F}_N\right) = \sum_{i=1}^N \frac{1}{N}\left(\bar{K}^*_{N,h'_N,i\cdot} - \bar{K}_{N,h'_N,\cdot\cdot}\right) = 0$. The second condition, convergence of conditional variance in probability, is established in lemma 1.4.3. The third condition, conditional Lindeberg's condition in probability, is verified by moment calculations in lemma 1.6.4 in the appendix. Since all three conditions are satisfied, theorem 1.4.4 implies the weak convergence in probability (1.13). $\square$

1.5 Efficient Influence Function and Optimal Asymptotic Variance

Now that we have developed an in-depth understanding of a specific estimator, the kernel density estimator, we will take one step further and find out the efficiency bound of this density estimation problem across all possible estimators. The following theorem rigorously derives the efficient influence function under nondegeneracy, $\Omega_1 \neq 0$, and shows that the kernel estimator is asymptotically efficient.

Theorem 1.5.1 (Efficient Influence Function and Optimal Asymptotic Variance)**.** *If assumption 1.2.1 holds and the problem is nondegenerate, $\Omega_1 \neq 0$, then the efficient influence function for estimating $f_W(w_0)$ is*

$$\tilde{\psi}(a) = 2 \cdot [f_{W|A}(w_0|A = a) - f_W(w_0)]$$

and the optimal asymptotic variance is $4\operatorname{Var}\left(f_{W|A}(w_0|A = a)\right)$. If in addition assumptions 1.2.2, 1.2.2 holds and $N^{-1} \ll h_N \ll N^{-2/5}$, the kernel density estimator is asymptotically efficient.

Proof. Imagine the situation where we observe $(A_1, \ldots, A_N)$ in addition to $(W_{ij}, 1 \le i < j \le N)$. Since the set of estimators as a function of the infeasible hypothetical data $(A_1, \ldots, A_N, W_{ij}, 1 \le i < j \le N)$ is larger than the set of estimators as a function of the actual data $(W_{ij}, 1 \le i < j \le N)$, the asymptotic efficiency bound of the former serves as a lower bound of that of the latter. We will work with the former set of estimators to get the efficiency bound and then show that the bound obtained are achievable by sequences of estimators using only the actual data $(W_{ij}, 1 \le i < j \le N)$, which concludes the proof.

To start the discussion, let's index the full nonparametric model by the pair $\left(P_A, P_{W|AA}\right)$. P_A denotes the distribution of A. $P_{W|AA}$ denotes the conditional distribution of $W_{ij}|A_i, A_j$. We denote the full model by

$$\mathcal{P} = \left\{P_{\left(P_A, P_{W|AA}\right)} : \text{Assumption 1.2.1 is satisfied}\right\}.$$

The first step of calculating the efficiency bound is reducing the problem to the familiar iid setting by a sufficiency argument. Consider the following submodel with a fixed and known primitive $P_{W|AA} = P_{W|AA,0}$

$$\mathcal{P}(P_{W|AA,0}) = \left\{ P_{\left(P_A, P_{W|AA}\right)} \in \mathcal{P} : P_{W|AA} = P_{W|AA,0} \right\},$$

where the conditional distribution of W_{12} given A_1, A_2 is fixed and known. This submodel is indexed by P_A, which is the only unknown. Since for every $(a_1, \ldots, a_N)$ and P_A, the conditional distribution of $(A_1, \ldots, A_N, W_{ij}, 1 \leq i < j \leq N)$ under $P_{\left(P_A, P_{W|AA,0}\right)}$ given $(A_1, \ldots, A_N) = (a_1, \ldots, a_N)$ does not depend on P_A, $(A_1, \ldots, A_N)$ is a sufficient statistic for $\mathcal{P}(P_{W|AA,0})$ (or for $(A_1, \ldots, A_N, W_{ij}, 1 \leq i < j \leq N)$). As a result, for any estimator as a function of $(A_1, \ldots, A_N, W_{ij}, 1 \leq i < j \leq N)$, there exists a (random) estimator as a function of $(A_1, \ldots, A_N)$ only with the same distribution. Hence, for the purpose of getting distributional properties of estimators, working with estimators as a function of $(A_1, \ldots, A_N)$ only is completely general in this submodel. Notice this means we are back in the iid setting in the sense we are looking at the set of estimators which are functions of iid random variables $(A_1, \ldots, A_N)$ and the only unknown is P_A. We can use all the machinery developed for iid data to get an efficiency bound among estimators which are functions of $(A_1, \ldots, A_N)$ for estimating $f_W(w_0)$ in this submodel.

In the submodel $\mathcal{P}\left(P_{W|AA,0}\right)$ with the data $(A_1, \ldots, A_N)$, write the estimand $f_W(w_0)$ as a function of the unknown primitive P_A

$$f_W(w_0) = \nu\left(P_A\right) = \frac{d}{dw}\Big|_{w=w_0} \int\int P_{W|AA,0}(w|a_1, a_2) dP_A(a_1) dP_A(a_2).$$

Take a directional derivative of the parameter ν with respect to $h - P_A$

$$\begin{aligned}
\dot{\nu}\left(P_A\right)\left(h - P_A\right) &= \frac{d}{dt}\Big|_{t=0} \nu\left(P_A + t(h - P_A)\right) \\
&= \frac{d}{dt}\Big|_{t=0} \frac{d}{dw}\Big|_{w=w_0} \Bigg(\int\int P_{W|AA,0}(w|a_1, a_2) \\
&\qquad\qquad d\left(P_A + t(h - P_A)\right)(a_1) d\left(P_A + t(h - F_A)\right)(a_2) \Bigg) \\
&= 2 \int \frac{d}{dw}\Big|_{w=w_0} \int P_{W|AA,0}(w|a_1, a_2) dP_A(a_1) d\left(h - P_A\right)(a_2) \\
&= \int 2 f_{W|A}(w_0|A = a_2) d\left(h - P_A\right)(a_2) \\
&= P_{h - P_A} \psi,
\end{aligned}$$

where $\psi(a) \equiv 2 f_{W|A}(w_0|A = a)$ is identified as the influence function. Since the model is nonparametric in the sense P_A can be any probability law on the sample space, the tangent set at P_A consists of all measurable functions g satisfying $\int g dP_A = 0$. The efficient influence function, which is the projection of influence function $\psi(a)$ onto the tangent set, is

$$\tilde{\psi}(a) = \psi(a) - \mathbb{E}\psi(A) = 2 \cdot [f_{W|A}(w_0|A = a) - f_W(w_0)].$$

An estimator is efficient if and only if it has the asymptotic linear expansion $\hat{\nu}(P_A) = \nu(P_A) + N^{-1/2}\sum_{i=1}^{N}\tilde{\psi}_{P_A}(A_i) + o_P\left(N^{-1/2}\right)$. Under assumption 1.2.2, 1.2.3, and $N^{-1} \ll h_N \ll N^{-2/5}$, the kernel density estimator has exactly this asymptotic linear expansion. This implies the kernel density estimator is asymptotically efficient in any submodel $\mathcal{P}(P_{W|AA,0})$ and it is asymptotically efficient in the full nonparametric model $\mathcal{P}$. □

1.6 Appendix: Proofs

This Appendix contains proofs of results in the main text of Chapter 1.

Proof of Theorem 1.3.1

Proof. (a)

$$\begin{aligned}
\mathbb{E}\left[\hat{f}_W(w_0)\right] &= \mathbb{E}K_{12} \\
&= \mathbb{E}\left[h_N^{-1}K\left(\frac{w_0 - W_{12}}{h_N}\right)\right] \\
&= \int h_N^{-1}K\left(\frac{w_0 - s}{h_N}\right) f_W(s)ds \\
&= \int K(u)\, f_W(w_0 - h_N u)du \\
&= \int K(u)\left[f_W(w_0) - f_W'(w_0 - h_N\bar{u})hu\right]du \\
&= \int K(u)\left[f_W(w_0) - f_W'(w_0)h_N u + O\left(h_N^2 u^2\right)\right]du \\
&= f_W(w_0) + O\left(h_N^2\right).
\end{aligned}$$

The third equality uses change of variable $u \equiv \frac{w_0 - s}{h_N}$. The fourth equality follows mean value theorem guaranteed by continuous differentiability of f_W. The fifth equality follows Lipschitz condition of f_W'. The last equality follows the fact that $\int K(u)u\, du = 0$ because $K(u) = K(-u)$. Integrability of all terms are guaranteed by boundedness of K and f_W and the bounded support assumption on the kernel K.

(b)

$$\begin{aligned}
\mathbb{E}\left[K_{ij}^2\right] &= \int h_N^{-2}\left[K\left(\frac{w_0-s}{h_N}\right)\right]^2 f_W(s)ds \\
&= h_N^{-1}\int [K(u)]^2 f_W(w_0-h_Nu)du \\
&= h_N^{-1}\int [K(u)]^2 \left[f_W(w_0)-f_W'(w_0-h_N\bar{u})hu)\right] du \\
&= h_N^{-1}\int [K(u)]^2 \left[f_W(w_0)-f_W'(w_0)h_Nu+O(h_N^2u^2))\right] du \\
&= h_N^{-1}f_W(w_0)\int [K(u)]^2 du - f_W'(w_0)\int [K(u)]^2 u\ du + O(h_N) \\
&= h_N^{-1}f_W(w_0)\int K^2(u)du + O(1).
\end{aligned}$$

Plug it into the definition of $\Omega_{2,N}$.

$$\begin{aligned}
\Omega_{2,N} &= h_N\mathbb{E}\left[K_{12}^2\right] - h_N\mathbb{E}\left[\hat{f}_W(w_0)\right]^2 \\
&= h_N\cdot\left[h_N^{-1}f_W(w_0)\int [K(u)]^2 du + O(1)\right] - h_N\left[f_W(w_0)+O\left(h_N^2\right)\right]^2 \\
&= \Omega_2 + O(h_N).
\end{aligned}$$

$$\begin{aligned}
\Omega_{1,N} &= \mathbb{E}\left[\left(\mathbb{E}\left[K_{12}|A_1\right]-\mathbb{E}K_{12}\right)^2\right] \\
&= \mathbb{E}\left\{\left[f_{W|A}(w_0|A_1)-f_W(w_0)+O\left(h_N^2\right)\right]^2\right\} \\
&= \begin{cases} \Omega_1 + O\left(h_N^2\right) & \text{if } \Omega_1>0 \\ O\left(h_N^4\right) & \text{if } \Omega_1 = 0. \end{cases}.
\end{aligned}$$

The second equality follows the observation that

$$\mathbb{E}\left(K_{12}|A_1\right) = f_{W|A}(w_0|A_1) + O\left(h_N^2\right) \tag{1.14}$$

and $\mathbb{E}K_{12} = f_W(w_0) + O\left(h_N^2\right)$. In the special case $\mathrm{Var}\left(f_{W|A}^2(w_0|A_1)\right) = 0$, we have $f_{W|A}(w_0|A_1) - f_W(w_0) \overset{\text{a.s.}}{=} 0$ and $\mathbb{E}\left\{\left[f_{W|A}(w_0|A_1)-f_W(w_0)+O\left(h_N^2\right)\right]^2\right\} = \mathbb{E}\left\{\left[0+O\left(h_N^2\right)\right]^2\right\} = O\left(h_N^4\right)$. Plug these back into variance formula (1.5).

$$\sigma_N^2 \equiv \mathrm{Var}\left(\hat{f}_W(w_0)\right) = \begin{cases} N^{-1}\left(1-\frac{1}{N-1}\right)4\Omega_1 + \binom{N}{2}^{-1}h_N^{-1}\Omega_2 + O\left(N^{-2}+N^{-1}h_N^2\right) & \text{if } \Omega_1>0 \\ \binom{N}{2}^{-1}h_N^{-1}\Omega_2 + O\left(N^{-2}+N^{-1}h_N^4\right) & \text{if } \Omega_1 = 0 \end{cases},$$

(c) This part is implied by (a) and (b).

□

Proof of Theorem 1.3.2

Proof. The variance estimator is

$$\begin{aligned}\hat{\sigma}_N^2 &= \binom{N}{2}^{-2} \sum_{i_1<j_i} \sum_{i_2<j_2} d(i_1, j_1, i_2, j_2) \left(K_{i_1j_1} - \hat{f}_W(w_0)\right) \left(K_{i_2j_2} - \hat{f}_W(w_0)\right) \\ &= N^{-1}(N-1)^{-1}(N-2)4\ \hat{\Omega}_{1,N} + N^{-1}(N-1)^{-1}h_N^{-1}2\ \hat{\Omega}_{2,N},\end{aligned}$$

where

$$\begin{aligned}\hat{\Omega}_{1,N} &\equiv \binom{N}{3}^{-1} \sum_{i<j<k} \left(\frac{K_{ij}K_{ik} + K_{ij}K_{jk} + K_{ik}K_{jk}}{3}\right) - \hat{f}_W^2(w_0) \\ \hat{\Omega}_{2,N} &\equiv \binom{N}{2}^{-1} \sum_{i<j} h_N K_{ij}^2 - h_N \hat{f}_W^2(w_0).\end{aligned}$$

The following lemma is useful for proving theorem 1.3.2.

Lemma 1.6.1. *Under assumption 1.2.1, 1.2.2, 1.2.3,*

(a) $\mathbb{E}\hat{\Omega}_{1,N} = \Omega_{1,N} - \sigma_N^2,\ \mathbb{E}\hat{\Omega}_{2,N} = \Omega_{2,N} - h_N\sigma_N^2.$

(b)

$$\operatorname{Var}\left(\hat{\Omega}_{1,N}\right) \le \begin{cases} O\left(N^{-3}h_N^{-2} + N^{-2}h_N^{-1} + N^{-1}\right) & \text{if } \Omega_1 > 0 \\ O\left(N^{-3}h_N^{-2} + N^{-2}h_N^{-1} + N^{-1}h_N^2\right) & \text{if } \Omega_1 = 0 \end{cases}$$

(c) $\operatorname{Var}\left(\hat{\Omega}_{2,N}\right) \le O\left(N^{-2}h_N^{-1} + N^{-1}\right).$

The proof of this lemma can be found at the end of this subsection. Armed with this lemma, we are ready to prove theorem 1.3.2.

By the definition of $\sigma_N^2,\ \hat{\sigma}_N^2$ in equation (1.3), (1.8),

$$\begin{aligned}\frac{\hat{\sigma}_N^2}{\sigma_N^2} &= \frac{N^{-1}(N-1)^{-1}h_N^{-1}2\ \hat{\Omega}_{2,N} + N^{-1}(N-1)^{-1}(N-2)4\ \hat{\Omega}_{1,N}}{N^{-1}(N-1)^{-1}h_N^{-1}2\ \Omega_{2,N} + N^{-1}(N-1)^{-1}(N-2)4\ \Omega_{1,N}} \\ &= \frac{\hat{\Omega}_{2,N} + (N-2)h_N 2\ \hat{\Omega}_{1,N}}{\Omega_{2,N} + (N-2)h_N 2\ \Omega_{1,N}}.\end{aligned}$$

By lemma (1.6.1) (a),

$$\mathbb{E}\left[\frac{\hat{\sigma}_N^2}{\sigma_N^2}\right] = \frac{\mathbb{E}\hat{\Omega}_{2,N} + (N-2)h_N 2\ \mathbb{E}\hat{\Omega}_{1,N}}{\Omega_{2,N} + (N-2)h_N 2\ \Omega_{1,N}} = 1 - \frac{1 + 2(N-2)h_N^2}{\Omega_{2,N} + 2(N-2)h_N\ \Omega_{1,N}}\sigma_N^2 = 1 + o(1).$$

This result doesn't depend on assumption of Ω_1. The variance is

$$\begin{aligned}\operatorname{Var}\left(\frac{\hat{\sigma}_N^2}{\sigma_N^2}\right) &= \operatorname{Var}\left(\frac{\hat{\Omega}_{2,N} + (N-2)h2\ \hat{\Omega}_{1,N}}{\Omega_{2,N} + (N-2)h2\ \Omega_{1,N}}\right) \\ &\leq [\Omega_{2,N} + (N-2)h2\ \Omega_{1,N}]^{-2}\left[2\operatorname{Var}\left(\hat{\Omega}_{2,N}\right) + 2\operatorname{Var}\left((N-2)h2\ \hat{\Omega}_{1,N}\right)\right].\end{aligned}$$

Based on bounds in lemma 1.6.1 and bound in the proof of theorem 1.3.1(b), we get
Case 1: $\Omega_1 > 0$

$$\begin{aligned}\operatorname{Var}\left(\frac{\hat{\sigma}_N^2}{\sigma_N^2}\right) &= O\left((1+Nh_N)^{-2}\left[N^{-2}h_N^{-1} + N^{-1} + N^2h_N^2(N^{-3}h_N^{-2} + N^{-2}h_N^{-1} + N^{-1})\right]\right) \\ &= O\left((1+Nh_N)^{-2}\left[N^{-2}h_N^{-1} + N^{-1} + h_N + Nh_N^2\right]\right) \\ &= O\left(N^{-2}h_N^{-1} + N^{-1} + h_N + N^{-1}\right) \\ &= o(1).\end{aligned}$$

Case 2: $\Omega_1 = 0$

$$\begin{aligned}\operatorname{Var}\left(\frac{\hat{\sigma}_N^2}{\sigma_N^2}\right) &= O\left(1^{-2}\left[N^{-2}h_N^{-1} + N^{-1} + N^2h_N^2(N^{-3}h_N^{-2} + N^{-2}h_N^{-1} + N^{-1}h_N^2)\right]\right) \\ &= O\left(N^{-2}h_N^{-1} + N^{-1} + N^{-1} + h_N + Nh_N^4\right) \\ &= o(1).\end{aligned}$$

The second step and the last step uses the fact that $h_N \ll N^{-1/4}$

To summarize, we've shown

$$\mathbb{E}\left[\frac{\hat{\sigma}_N^2}{\sigma_N^2}\right] = 1 + o(1),\ \operatorname{Var}\left(\frac{\hat{\sigma}_N^2}{\sigma_N^2}\right) = o(1).$$

Hence, $\frac{\hat{\sigma}_N^2}{\sigma_N^2} = 1 + o_p(1)$. $\square$

Proof of Lemma 1.6.2

The following technical lemma and the calculation in its proof is useful for proving lemma 1.6.1. We will prove it first.

Lemma 1.6.2 (Moments)**.** *Under assumption 1.2.1, 1.2.2, 1.2.3,*

(a) For any $c \in \mathbb{N}$, $\mathbb{E}\left(|K_{12}|^c\right) = O\left(h_N^{-(c-1)_+}\right)$.

(b) *For any* $k \in \{1, 2, \ldots, N\}$, $c_{ij} \in \mathbb{N}$, $\mathbb{E}\left[\prod\prod_{i<j} |K_{ij}|^{c_{ij}}\right] \le O\left(h_N^{-\left[\left(\sum_{i \neq k} c_{ik}\right)-1\right]_+}\right) \cdot$
$\mathbb{E}\left[\prod\prod_{\substack{i<j \\ i,j\neq k}} |K_{ij}|^{c_{ij}}\right]$.

(c) *Let* $m \equiv |\operatorname{unique}(i_1, j_1, i_2, j_2, i_3, j_3, i_4, j_4)|$ *be the number of distinct nodes of the four dyads* $(i_1, j_1), (i_2, j_2), (i_3, j_3), (i_4, j_4)$. *Then*

$$\operatorname{Cov}\left(K_{i_1 j_1} K_{i_2 j_2}, K_{i_3 j_3} K_{i_4 j_4}\right) = \begin{cases} 0 & \text{if } m = 8 \\ O\left(h_N^{-\lfloor 4-m/2 \rfloor}\right) & \text{if } 2 \le m \le 7 \end{cases}$$

Moreover, if $\Omega_1 = 0$, *then* $\operatorname{Cov}\left(K_{i_1 j_1} K_{i_2 j_2}, K_{i_3 j_3} K_{i_4 j_4}\right) = O(h_N^2)$ *in case* $m = 1$.

Proof. (a) If $c = 0$, then $(c-1)_+ = 0$, $h_N^0 = 1$ and $\mathbb{E}\left(|K_{12}|^0\right) = 1 = O(1)$ is trivial. If $c \geq 1$, then

$$\begin{aligned}
\mathbb{E}\left(|K_{12}|^c\right) &= h_N^{-c} \int \left|K\left(\frac{w_0 - s}{h_N}\right)\right|^c f_W(s) ds \\
&= h_N^{-(c-1)} \int |K(u)|^c f_W(w_0 - h_N u) ds \\
&\le h_N^{-(c-1)} \left(f_W(w_0) \int |K(u)|^c ds + O(h_N)\right) \\
&= O\left(h_N^{-(c-1)}\right).
\end{aligned}$$

(b) By the iterated expectation formula,

$$\begin{aligned}
\mathbb{E}\left[\prod\prod_{i<j} |K_{ij}|^{c_{ij}}\right] &= \mathbb{E}\left\{\mathbb{E}\left[\prod\prod_{i<j} |K_{ij}|^{c_{ij}} \middle| A_1, \ldots, A_{k-1}, A_{k+1}, \ldots, A_N\right]\right\} \\
&= \mathbb{E}\left\{\mathbb{E}\left[\prod\prod_{\substack{i<j \\ i,j\neq k}} |K_{ij}|^{c_{ij}} \middle| A_1, \ldots, A_{k-1}, A_{k+1}, \ldots, A_N\right]\right. \\
&\qquad \left. \cdot\, \mathbb{E}\left[\prod_{i \neq k} |K_{ik}|^{c_{ik}} \middle| A_1, \ldots, A_{k-1}, A_{k+1}, \ldots, A_N\right]\right\} \qquad (1.15)
\end{aligned}$$

By Jensen's inequality,

$$\prod_{i\neq k} |K_{ik}|^{c_{ik}} \le \sum_{i \neq k} \frac{c_{ik}}{\sum_{l\neq k} c_{lk}} \left|K_{ik}^{\sum_{l \neq k} c_{lk}}\right|.$$

The second term in (1.15) is hence upper bounded by

$$
\begin{aligned}
&\mathbb{E}\left[\prod_{i\neq k}|K_{ik}|^{c_{ik}}\middle|A_1,\ldots,A_{k-1},A_{k+1},\ldots,A_N\right]\\
&\le \mathbb{E}\left[\sum_{i\neq k}\frac{c_{ik}}{\sum_{l\neq k}c_{lk}}\left|K_{ik}^{\sum_{l\neq k}c_{lk}}\right|\middle|A_1,\ldots,A_{k-1},A_{k+1},\ldots,A_N\right]\\
&=\sum_{i\neq k}\frac{c_{ik}}{\sum_{l\neq k}c_{lk}}\mathbb{E}\left[\left|K_{ik}^{\sum_{l\neq k}c_{lk}}\right|\middle|A_i\right]\\
&=\sum_{i\neq k}\frac{c_{ik}}{\sum_{l\neq k}c_{lk}}O\left(h^{-\left[\left(\sum_{l\neq k}c_{lk}\right)-1\right]_+}\right)\\
&=O\left(h_N^{-\left[\left(\sum_{i\neq k}c_{ik}\right)-1\right]_+}\right).
\end{aligned}
$$

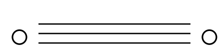
Plug this back in (1.15). We get

$$
\begin{aligned}
\mathbb{E}\left[\prod\prod_{i<j}|K_{ij}|^{c_{ij}}\right]&=\mathbb{E}\left\{\mathbb{E}\left[\prod\prod_{i<j}|K_{ij}|^{c_{ij}}\middle|A_1,\ldots,A_{k-1},A_{k+1},\ldots,A_N\right]\right\}\\
&\le O\left(h_N^{-\left[\left(\sum_{i\neq k}c_{ik}\right)-1\right]_+}\right)\cdot\mathbb{E}\left\{\mathbb{E}\left[\prod\prod_{\substack{i<j\\ i,j\neq k}}|K_{ij}|^{c_{ij}}\middle|A_1,\ldots,A_{k-1},A_{k+1},\ldots,A_N\right]\right\}\\
&=O\left(h_N^{-\left[\left(\sum_{i\neq k}c_{ik}\right)-1\right]_+}\right)\cdot\mathbb{E}\left[\prod\prod_{\substack{i<j\\ i,j\neq k}}|K_{ij}|^{c_{ij}}\right].
\end{aligned}
$$

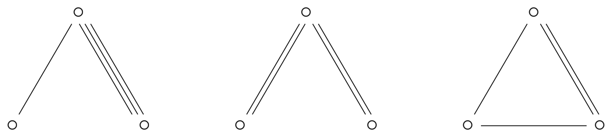
(c) If $m=8$, $K_{i_1j_1}K_{i_2j_2}$, $K_{i_3j_3}K_{i_4j_4}$ are independent from each other and their covariance is zero.
If $m\le 7$, we will use the bound from part (b) to iteratively remove nodes and bound the moments $\mathbb{E}\,|K_{i_1j_1}K_{i_2j_2}K_{i_3j_3}K_{i_4j_4}|$. The exact bound will depend on the graph

$$G=(V=\text{unique}(i_1,j_1,i_2,j_2,i_3,j_3,i_4,j_4),E=\{(i_1,j_1),(i_2,j_2),(i_3,j_3),(i_4,j_4)\})\,.$$

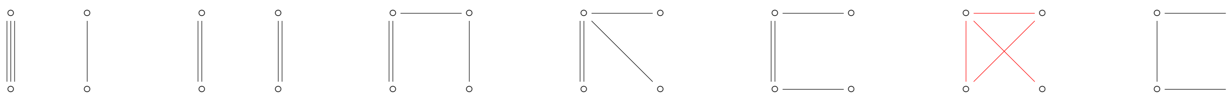
If $m=2$, there is only a single possible graph



If $m = 3$, all possible graphs are

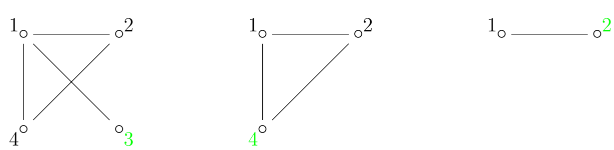


If $m = 4$, then all possible graphs are



It's also feasible to list the possible graph isomorphisms for $m = 5, 6, 7$.

We will give an example of how to use the bound in part (a),(b) for a typical graph like the red one above when $m = 4$. We start from the first graph below on the left. To apply bound of part (b), we typically pick the node k as the node with the smallest number of dyads in the graph. Here in step one, we pick node $k = 3$, which is highlighted in green. Since node 3 is in a single dyad, its contribution is of order $O\left(h_N^{(1-1)_+}\right) = O(1)$. After applying the bound, we can then delete node 3 and reduce the graph to the second graph. In the second step, we apply the bound in part (b) with node $k = 4$. Since node 4 is in two dyad in this graph, its contribution is of order $O\left(h_N^{(2-1)_+}\right) = O(h)$. We can then delete node 4 and reduce the graph to the third graph. In the third step, we apply the bound in part (b). Since there is a single dyad left in the graph, its contribution is of order $O\left(h_N^{(1-1)_+}\right) = O(1)$. The total magnitude of $\mathbb{E}\left|K_{i_1j_1}K_{i_2j_2}K_{i_3j_3}K_{i_4j_4}\right|$ is therefore bounded by the sum of contributions in all three steps, which is $O(1) \cdot O(h_N) \cdot O(1) = O(h_N)$.



Apply this bound iteratively gives us the result of part (c).

In the special case where $\Omega_1 = 0$, we have $f_W(w_0|A_1) \overset{\text{a.s.}}{=} f_W(w_0)$ and hence $\mathbb{E}\left(K_{ij}|A_i\right) = f_W(w_0|A_1) + O\left(h_N^2\right) \overset{\text{a.s.}}{=} f_W(w_0) + O\left(h_N^2\right)$. Calculation based on this fact shows

$$\begin{aligned}
\mathbb{E}\left(K_{12}K_{13}K_{14}K_{15}\right) &= f_W^4(w_0) + O(h_N^2) \\
\mathbb{E}\left(K_{12}K_{13}K_{14}K_{45}\right) &= f_W^4(w_0) + O(h_N^2) \\
\mathbb{E}\left(K_{12}K_{23}K_{14}K_{45}\right) &= f_W^4(w_0) + O(h_N^2) \\
(\mathbb{E}K_{12}K_{13})^2 &= f_W^4(w_0) + O(h_N^2),
\end{aligned}$$

which implies if $\Omega_1 = 0$, then $\operatorname{Cov}\left(K_{i_1j_1}K_{i_2j_2}, K_{i_3j_3}K_{i_4j_4}\right) = O(h^2)$ in case $m = 1$.

□

Proof of Lemma 1.6.1

Proof. (a)

$$\begin{aligned}
\mathbb{E}\hat{\Omega}_{1,N} &= \mathbb{E}\left[K_{12}K_{13}\right] - \mathbb{E}\left[\hat{f}_W^2(w_0)\right] = \Omega_{1,N} - \sigma_N^2 \\
\mathbb{E}\hat{\Omega}_{2,N} &= h\mathbb{E}\left[K_{12}^2\right] - h\mathbb{E}\left[\hat{f}_W^2(w_0)\right] = \Omega_{2,N} - h_N\sigma_N^2.
\end{aligned}$$

(b)

$$\begin{aligned}
\operatorname{Var}\left(\hat{\Omega}_{1,N}\right) &= \operatorname{Var}\left(\binom{N}{3}^{-1}\sum_{i<j<k}\left(\frac{K_{ij}K_{ik}+K_{ij}K_{jk}+K_{ik}K_{jk}}{3}\right) - \hat{f}_W^2(w_0)\right) \\
&\leq 2\operatorname{Var}\left(\binom{N}{3}^{-1}\sum_{i<j<k}\left(\frac{K_{ij}K_{ik}+K_{ij}K_{jk}+K_{ik}K_{jk}}{3}\right)\right) + 2\operatorname{Var}\left(\hat{f}_W^2(w_0)\right).
\end{aligned} \tag{1.16}$$

Calculation of the first term is similar to the previous variance calculation.

$$\begin{aligned}
&\operatorname{Var}\left(\binom{N}{3}^{-1}\sum_{i<j<k}\left(\frac{K_{ij}K_{ik}+K_{ij}K_{jk}+K_{ik}K_{jk}}{3}\right)\right) \\
&\qquad = O\left(N^{-3}\right)\cdot\gamma_{3,N} + O\left(N^{-2}\right)\cdot\gamma_{2,N} + O\left(N^{-1}\right)\cdot\gamma_{1,N}
\end{aligned} \tag{1.17}$$

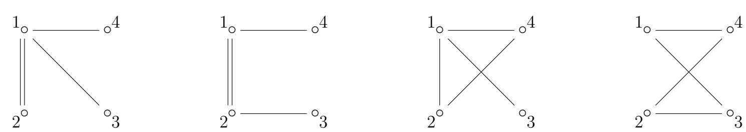
where

$$\begin{aligned}
\gamma_{3,N} &\equiv \operatorname{Var}\left(s_{123}\right) \\
\gamma_{2,N} &\equiv \operatorname{Cov}\left(s_{123}, s_{124}\right) \\
\gamma_{1,N} &\equiv \operatorname{Cov}\left(s_{123}, s_{145}\right) \\
s_{ijk} &\equiv \frac{K_{ij}K_{ik}+K_{ij}K_{jk}+K_{ik}K_{jk}}{3}.
\end{aligned}$$

We will get general bounds of $\gamma_{3,N}, \gamma_{2,N}, \gamma_{1,N}$ using lemma 1.6.2.

$$
\begin{aligned}
\gamma_{3,N} &= \operatorname{Var}\left(s_{123}\right) = \operatorname{Var}\left(\frac{K_{12}K_{13}+K_{12}K_{23}+K_{13}K_{23}}{3}\right) \leq \operatorname{Var}\left(K_{12}K_{13}\right) \leq \mathbb{E}\left[K_{12}^2K_{13}^2\right] \\
&\leq O\left(h_N^{-2}\right) \\
|\gamma_{2,N}| &= |\operatorname{Cov}\left(s_{123}, s_{124}\right)| \\
&= 9^{-1}\left|\operatorname{Cov}\left(K_{12}K_{13}+K_{12}K_{23}+K_{13}K_{23}, K_{12}K_{14}+K_{12}K_{24}+K_{14}K_{24}\right)\right| \\
&= 9^{-1}\Big\{\mathbb{E}\left|K_{12}K_{13}K_{12}K_{14}\right| + \mathbb{E}\left|K_{12}K_{13}K_{12}K_{24}\right| + \mathbb{E}\left|K_{12}K_{13}K_{14}K_{24}\right| \\
&\qquad + \mathbb{E}\left|K_{12}K_{23}K_{12}K_{14}\right| + \mathbb{E}\left|K_{12}K_{23}K_{12}K_{24}\right| + \mathbb{E}\left|K_{12}K_{23}K_{14}K_{24}\right| \\
&\qquad + \mathbb{E}\left|K_{13}K_{23}K_{12}K_{14}\right| + \mathbb{E}\left|K_{13}K_{23}K_{12}K_{24}\right| + \mathbb{E}\left|K_{13}K_{23}K_{14}K_{24}\right|\Big\} + O(1) \\
&= O\left(h_N^{-1}\right).
\end{aligned}
$$

The last equality follows the same argument as the proof in lemma 1.6.2(c). The four different relevant graph isomorphisms in the nine terms are



Lastly, we bound $|\gamma_{1,N}|$ by

$$
\begin{aligned}
&|\gamma_{1,N}| \\
&= |\operatorname{Cov}\left(s_{123}, s_{145}\right)| \\
&= 9^{-1}\left|\operatorname{Cov}\left(K_{12}K_{13}+K_{12}K_{23}+K_{13}K_{23}, K_{14}K_{15}+K_{14}K_{45}+K_{15}K_{45}\right)\right| \\
&= 9^{-1}\Big|\operatorname{Cov}\left(K_{12}K_{13}, K_{14}K_{15}\right) + \operatorname{Cov}\left(K_{12}K_{13}, K_{14}K_{45}\right) + \operatorname{Cov}\left(K_{12}K_{13}, K_{15}K_{45}\right) \\
&\qquad + \operatorname{Cov}\left(K_{12}K_{23}, K_{14}K_{15}\right) + \operatorname{Cov}\left(K_{12}K_{23}, K_{14}K_{45}\right) + \operatorname{Cov}\left(K_{12}K_{23}, K_{15}K_{45}\right) \\
&\qquad + \operatorname{Cov}\left(K_{13}K_{23}, K_{14}K_{15}\right) + \operatorname{Cov}\left(K_{13}K_{23}, K_{14}K_{45}\right) + \operatorname{Cov}\left(K_{13}K_{23}, K_{15}K_{45}\right)\Big| \\
&= 9^{-1}\left|\operatorname{Cov}\left(K_{12}K_{13}, K_{14}K_{15}\right) + 4\operatorname{Cov}\left(K_{12}K_{13}, K_{14}K_{45}\right) + 4\operatorname{Cov}\left(K_{12}K_{23}, K_{14}K_{45}\right)\right| \\
&= 9^{-1}\left|\mathbb{E}\left(K_{12}K_{13}K_{14}K_{15}\right) + 4\mathbb{E}\left(K_{12}K_{13}K_{14}K_{45}\right) + 4\mathbb{E}\left(K_{12}K_{23}K_{14}K_{45}\right) - 9\left(\mathbb{E}K_{12}K_{13}\right)^2\right| \\
&\leq O\left(1\right).
\end{aligned}
$$

The last inequality follows from applying lemma (1.6.2)(a)(b). In the special case when $\Omega_1 = 0$, lemma (1.6.2)(c) implies $\gamma_{1,N} = O(h_N^2)$.

Use these bounds in equation (1.17). We get in general

$$\begin{aligned}
&\operatorname{Var}\left(\binom{N}{3}^{-1}\sum_{i<j<k}\left(\frac{K_{ij}K_{ik}+K_{ij}K_{jk}+K_{ik}K_{jk}}{3}\right)\right)\\
=&O\left(N^{-3}\right)\ \gamma_{3,N}+O\left(N^{-2}\right)\ \gamma_{2,N}+O\left(N^{-1}\right)\ \gamma_{1,N}\\
\leq&O\left(N^{-3}h_N^{-2}+N^{-2}h_N^{-1}+N^{-1}\right).
\end{aligned}$$

In the special case in which $\Omega_1=0$,

$$\operatorname{Var}\left(\binom{N}{3}^{-1}\sum_{i<j<k}\left(\frac{K_{ij}K_{ik}+K_{ij}K_{jk}+K_{ik}K_{jk}}{3}\right)\right)\leq\ O\left(N^{-3}h_N^{-2}+N^{-2}h_N^{-1}+N^{-1}h_N^{2}\right).$$

To calculate the second term in (1.16), notice

$$\begin{aligned}
\operatorname{Var}\left(\hat{f}_W^2(w_0)\right)&=\operatorname{Var}\left(\binom{N}{2}^{-2}\sum_{i_1<j_1}\sum_{i_2<j_2}K_{i_1j_1}K_{i_2j_2}\right)\\
&=\binom{N}{2}^{-4}\sum_{i_1<j_1}\sum_{i_2<j_2}\sum_{i_3<j_3}\sum_{i_4<j_4}\operatorname{Cov}\left(K_{i_1j_1}K_{i_2j_2},K_{i_3j_3}K_{i_4j_4}\right)\\
&=\binom{N}{2}^{-4}\sum_{m=2}^{7}O\left(N^m\right)O\left(h_N^{-\lfloor 4-m/2\rfloor}\right)\\
&=O\left(N^{-8}\right)\cdot O\left(N^7\cdot 1+N^6\cdot h_N^{-1}+N^5h_N^{-1}+N^4h_N^{-2}+N^3h_N^{-2}+N^2h_N^{-3}\right)\\
&=O\left(N^{-2}h_N^{-1}+N^{-1}\right).
\end{aligned}$$

The third equality follow the fact that there are $O\left(\binom{N}{m}\right)=O\left(N^m\right)$ number of terms with $|\,\text{unique}(i_1,j_1,i_2,j_2,i_3,j_3,i_4,j_4)|=m$ number of distinct associated nodes and the bound from lemma (1.6.2). In the special case $\Omega_1=0$, we have a tighter bound for $m=1$ and $\operatorname{Var}\left(\hat{f}_W^2(w_0)\right)=O(N^{-2}h_N^{-1}+N^{-1}h_N^2)$

Going back to (1.16), we have proved

$$\operatorname{Var}\left(\hat{\Omega}_{1,N}\right)\leq\begin{cases}O\left(N^{-3}h_N^{-2}+N^{-2}h_N^{-1}+N^{-1}\right) & \text{if } \Omega_1>0\\ O\left(N^{-3}h_N^{-2}+N^{-2}h_N^{-1}+N^{-1}h_N^2\right) & \text{if } \Omega_1=0\end{cases}$$

(c)

$$\begin{aligned}
\operatorname{Var}\left(\hat{\Omega}_{2,N}\right)&=\operatorname{Var}\left(\binom{N}{2}^{-1}\sum_{i<j}h_NK_{ij}^2-h_N\hat{f}_W^2(w_0)\right)\\
&\leq 2\operatorname{Var}\left(\binom{N}{2}^{-1}\sum_{i<j}h_NK_{ij}^2\right)+2\operatorname{Var}\left(h_N\hat{f}_W^2(w_0)\right). \qquad (1.18)
\end{aligned}$$

Calculation of the first term in (1.18) is similar to the variance calculation done previously.

$$\mathrm{Var}\left(\binom{N}{2}^{-1}\sum_{i<j} h_N K_{ij}^2\right) = O\left(N^{-2}\right)\eta_{2,N} + O\left(N^{-1}\right)\eta_{1,N},$$

where

$$\begin{aligned}
\eta_{2,N} &\equiv \mathrm{Var}\left(h_N K_{12}^2\right) = h_N^2 \mathbb{E}\left[K_{12}^4\right] - h_N^2 \mathbb{E}\left[K_{12}^2\right]^2 = h_N^2 \cdot O\left(h_N^{-3}\right) - h_N^2 O\left(h_N^{-1}\right)^2 = O\left(h_N^{-1}\right) \\
\eta_{1,N} &\equiv \mathrm{Cov}\left(h_N K_{12}^2, h_N K_{13}^2\right) = h_N^2 \mathbb{E}\left[K_{12}^2 K_{13}^2\right] - h_N^2 \mathbb{E}\left[K_{12}^2\right]^2 \\
&= h_N^2 \mathbb{E}\left[\mathbb{E}\left(K_{12}^2 | A_1\right)\mathbb{E}\left(K_{13}^2 | A_1\right)\right] - h_N^2 \mathbb{E}\left[K_{12}^2\right]^2 \\
&= h_N^2 \cdot O\left(h_N^{-1}\right)^2 - h_N^2 \cdot O\left(h_N^{-1}\right)^2 \\
&= O(1).
\end{aligned}$$

Plug these back into the variance expression.

$$\mathrm{Var}\left(\binom{N}{2}^{-1}\sum_{i<j} h_N K_{ij}^2\right) \le O\left(N^{-2}h_N^{-1} + N^{-1}\right).$$

The second term in (1.18) has been bounded already in part (b). Combine these together. We get $\mathrm{Var}\left(\hat{\Omega}_{2,N}\right) = O\left(N^{-2}h_N^{-1} + N^{-1}\right)$.

□

Proof of Theorem 1.3.3

Proof. We will apply Brown's martingale central limit theorem, as rephrased in in [Hall and Heyde] Corollary 3.1, p. 58. This requires us check two conditions:

1. Conditional Lindeberg condition

$$\text{for any } \epsilon > 0,\ \sum_{i=1}^{N} \mathbb{E}\left[\frac{Y_{N,i}^2}{\sigma_N^2}\mathbb{1}\left(\frac{Y_{N,i}}{\sigma_N} > \epsilon\right)\middle|\mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 0. \tag{1.19}$$

2. Stability condition

$$\sum_{i=1}^{N} \mathbb{E}\left[\frac{Y_{N,i}^2}{\sigma_N^2}\middle|\mathcal{F}_{i-1}\right] \xrightarrow{\mathrm{P}} 1. \tag{1.20}$$

In order to verify these, we will investigate their sufficient conditions which are easier to work with, namely

1'. Lyapunov condition

$$\sigma_N^{-4}\sum_{i=1}^N \mathbb{E}\left[Y_{N,i}^4\right] \xrightarrow{\mathrm{P}} 0, \tag{1.21}$$

2'. Stronger stability condition

$$\sigma_N^{-2}\sum_{i=1}^N Y_{N,i}^2 \xrightarrow{\mathrm{P}} 1. \tag{1.22}$$

(1.21) implies (1.19). (1.21) (1.22) together implies (1.20). The following two facts are particularly useful for verifying (1.21) and (1.22).

$$|D_i| = \left|\mathbb{E}\left(K_{ij}|A_i\right) - \mathbb{E}K_{ij}\right| = \left|f_{W|A}(w_0|A_i) - f_W(w_0) + O\left(h_N^2\right)\right| = \left\{\begin{array}{ll} O(1) & \text{if } \Omega_1 > 0 \\ O(h_N^2) & \text{if } \Omega_1 = 0 \end{array}\right. \tag{1.23}$$

$$\mathbb{E}\left(|K_{ij}|\middle|A_i\right) = O(1).$$

First, we verify Lyapunov condition (1.21). Since

$$(a+b)^4 \le 8a^4 + 8b^4,$$

to show

$$\sigma_N^{-4}\sum_{i=1}^N \mathbb{E}\left[Y_{N,i}^4\right] = \sigma_N^{-4}\sum_{i=1}^N \mathbb{E}\left[\left(N^{-1}2D_i + \binom{N}{2}^{-1}\sum_{j=1}^{i-1} E_{ji}\right)^4\right] \to 0,$$

we only need to show

$$\sigma_N^{-4}\sum_{i=1}^N \mathbb{E}\left[\left(N^{-1}2D_i\right)^4\right] \to 0, \text{ and } \sigma_N^{-4}\sum_{i=1}^N \mathbb{E}\left[\left(\binom{N}{2}^{-1}\sum_{j=1}^{i-1} E_{ji}\right)^4\right] \to 0.$$

The first term is

$$\sigma_N^{-4}\sum_{i=1}^N \mathbb{E}\left[N^{-4}D_i^4\right] = \left\{\begin{array}{ll} O\left(\left(N^{-2}h_N^{-1} + N^{-1}\right)^{-2} N^{-3}\right) = O\left(N^{-1}\right) & \text{if } \Omega_1 > 0 \\ O\left(\left(N^{-2}h_N^{-1}\right)^{-2} N^{-3}h_N^8\right) = O\left(Nh_N^{10}\right) & \text{if } \Omega_1 = 0 \end{array}\right. ,$$

which converges to zero as long as $h_N \ll N^{-1/10}$. To simplify the second term

$$\mathbb{E}\left[\left(\sum_{j=1}^{i-1} E_{ji}\right)^4\right] = \mathbb{E}\left[\sum_{j=1}^{i-1} E_{ji}^4 + 2\sum_{j_1=2}^{i-1}\sum_{j_2=1}^{j_1-1} E_{j_1 i}^2 E_{j_2 i}^2\right],$$

where we use the fact that $\mathbb{E}\left[E_{j_1 i}^3 E_{j_2 i}\right] = \mathbb{E}\left[E_{j_1 i} E_{j_2 i} E_{j_3 i}^2\right] = \mathbb{E}\left[E_{j_1 i} E_{j_2 i} E_{j_3 i} E_{j_4 i}\right] = 0, \forall j_1 \neq j_2 \neq j_3 \neq j_4, 1 \leq j_1, j_2, j_3, j_4 \leq i-1$. Hence

$$\begin{aligned}
\sum_{i=1}^{N} \mathbb{E}\left[\left(\sum_{j=1}^{i-1} E_{ji}\right)^4\right] &= \sum_{i=1}^{N}\sum_{j=1}^{i-1} \mathbb{E}\left[E_{ji}^4\right] + 2\sum_{i=1}^{N}\sum_{j_1=2}^{i-1}\sum_{j_2=1}^{j_1-1} \mathbb{E}\left[E_{j_1 i}^2 E_{j_2 i}^2\right] \\
&= O\left(N^2 h_N^{-3} + N^3 h_N^{-2}\right).
\end{aligned}$$

Back to the original term, in all cases

$$\begin{aligned}
\sigma_N^{-4} \sum_{i=1}^{N} \mathbb{E}\left[\left(\binom{N}{2}^{-1} \sum_{j=1}^{i-1} E_{ji}\right)^4\right] &= O\left(\left(N^{-2} h_N^{-1}\right)^{-2} N^{-8}\left(N^2 h_N^{-3} + N^3 h_N^{-2}\right)\right) \\
&= O\left(N^{-2} h_N^{-1} + N^{-1}\right) \\
&= o(1).
\end{aligned}$$

By now, we've verified Lyapunov condition (1.21). Second, let's verify the stronger stability condition (1.22).

$$\begin{aligned}
\sum_{i=1}^{N} Y_{N,i}^2 &= \sum_{i=1}^{N}\left(N^{-1} 2 D_i + \binom{N}{2}^{-1} \sum_{j=1}^{i-1} E_{ji}\right)^2 \\
&= N^{-2} 4 \sum_{i=1}^{N} D_i^2 + N^{-1}\binom{N}{2}^{-1} 4 \sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ji} D_i + \binom{N}{2}^{-2} \sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ji}^2 \\
&\quad + \binom{N}{2}^{-2} 2 \sum_{i=3}^{N}\sum_{j_1=2}^{i-1}\sum_{j_2=1}^{j_1-1} E_{j_1 i} E_{j_2 i}. \qquad (1.24)
\end{aligned}$$

We will bound the variance of these four terms. To bound the variance of the first term in (1.24), note that

$$\operatorname{Var}\left(\sum_{i=1}^{N} D_i^2\right) = N \operatorname{Var}\left(D_i^2\right) = \begin{cases} O(N) & \text{if } \Omega_1 > 0 \\ O\left(N h_N^4\right) & \text{if } \Omega_1 = 0 \end{cases}.$$

To bound the variance of the second term in (1.24), note that

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ji} D_i\right) &= \sum_{j<i}\sum_{l<k} \operatorname{Cov}\left(E_{ji} D_i, E_{lk} D_k\right) \\
&= \sum_{j<i} \operatorname{Var}\left(E_{ji} D_i\right) + 2 \sum_{j<i<k} \operatorname{Cov}\left(E_{ji} D_i, E_{jk} D_k\right) \\
&= \begin{cases} O\left(N^2 h_N^{-1} + N^3\right) & \text{if } \Omega_1 > 0 \\ O\left(N^2 h_N^3 + N^3 h_N^4\right) & \text{if } \Omega_1 = 0 \end{cases}.
\end{aligned}$$

The second equality follows from the fact that $\operatorname{Cov}\left(E_{ji}D_i, E_{lk}D_k\right) = 0$ when $j \neq l$. The third inequality follows from repeatedly applying the bound (1.23). To bound the variance of the third term in (1.24), note that

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ji}^2\right) &= \sum_{j<i}\sum_{l<k} \operatorname{Cov}\left(E_{ji}^2, E_{lk}^2\right) \\
&= \sum_{j<i} \operatorname{Var}\left(E_{ji}^2\right) + 2\sum_{j<i<k}\left[\operatorname{Cov}\left(E_{ji}^2, E_{jk}^2\right) + \operatorname{Cov}\left(E_{ji}^2, E_{ik}^2\right) + \operatorname{Cov}\left(E_{jk}^2, E_{ik}^2\right)\right] \\
&= O\left(N^2 h_N^{-3} + N^3 h_N^{-2}\right).
\end{aligned}$$

The last equality follows

$$\begin{aligned}
\operatorname{Var}\left(E_{ji}^2\right) &\leq \mathbb{E}\left(E_{ji}^4\right) = O\left(\mathbb{E}\left(K_{ji}^4\right)\right) = O\left(h_N^{-3}\right) \\
\operatorname{Cov}\left(E_{ji}^2, E_{jk}^2\right) &\leq \mathbb{E}\left(E_{ji}^2 E_{jk}^2\right) = \mathbb{E}\left(\mathbb{E}\left(E_{ji}^2 | A_j\right)^2\right) = O\left(\mathbb{E}\left(\mathbb{E}\left(K_{ji}^2 | A_j\right)^2\right)\right) = O\left(h_N^{-2}\right).
\end{aligned}$$

To bound the variance of the fourth term in (1.24), note that

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=3}^{N}\sum_{j_1=2}^{i-1}\sum_{j_2=1}^{j_1-1} E_{j_1 i}E_{j_2 i}\right) &= \sum_{j_2<j_1<i} \operatorname{Var}\left(E_{j_1 i}E_{j_2 i}\right) + 2\sum_{j_2<j_1<i_2<i_1} \operatorname{Cov}\left(E_{j_1 i_1}E_{j_2 i_1}, E_{j_1 i_2}E_{j_2 i_2}\right) \\
&= O\left(N^3 h_N^{-2} + N^4 h_N^{-1}\right).
\end{aligned}$$

The first equality is due to $\operatorname{Cov}\left(E_{j_1 i_1}E_{j_2 i_1}, E_{k_1 i_2}E_{k_2 i_2}\right) \neq 0$ only if ($j_1 = k_1$ and $j_2 = k_2$) or ($j_1 = k_2$ and $j_2 = k_1$). The second equality follows

$$\begin{aligned}
\operatorname{Var}\left(E_{j_1 i}E_{j_2 i}\right) &\leq \mathbb{E}\left(E_{j_1 i}^2 E_{j_2 i}^2\right) = \mathbb{E}\left(\mathbb{E}\left(E_{j_1 i}^2 | A_i\right)^2\right) = O\left(\mathbb{E}\left(\mathbb{E}\left(K_{j_1 i}^2 | A_i\right)^2\right)\right) = O\left(h_N^{-2}\right) \\
\operatorname{Cov}\left(E_{j_1 i_1}E_{j_2 i_1}, E_{j_1 i_2}E_{j_2 i_2}\right) &= O\left(\mathbb{E}\left(K_{j_1 i_1}K_{j_2 i_1}K_{j_1 i_2}K_{j_2 i_2}\right)\right) = O\left(h_N^{-1}\right).
\end{aligned}$$

Now go back to equation (1.24).

$$\begin{aligned}
\operatorname{Var}\left(\sum_{i=1}^{N} Y_{N,i}^2\right) &\leq 4\operatorname{Var}\left(N^{-2}4\sum_{i=1}^{N} D_i^2\right) + 4\operatorname{Var}\left(N^{-1}\binom{N}{2}^{-1}4\sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ji}D_i\right) + 4\operatorname{Var}\left(\binom{N}{2}^{-2}\sum_{i=2}^{N}\sum_{j=1}^{i-1} E_{ji}^2\right) \\
&\quad + 4\operatorname{Var}\left(\binom{N}{2}^{-2} 2\sum_{i=3}^{N}\sum_{j_1=2}^{i-1}\sum_{j_2=1}^{j_1-1} E_{j_1 i}E_{j_2 i}\right) \\
&\leq \begin{cases} O\left(N^{-4}\cdot N + N^{-6}\cdot\left(N^2 h_N^{-1} + N^3\right) + N^{-8}\cdot\left(N^2 h_N^{-3} + N^3 h_N^{-2}\right) + N^{-8}\cdot\left(N^3 h_N^{-2} + N^4 h_N^{-1}\right)\right) & \text{if } \Omega_1 > 0 \\ O\left(N^{-4}\cdot N h^4 + N^{-6}\cdot\left(N^2 h_N^3 + N^3 h_N^4\right) + N^{-8}\cdot\left(N^2 h_N^{-3} + N^3 h_N^{-2}\right) + N^{-8}\cdot\left(N^3 h_N^{-2} + N^4 h_N^{-1}\right)\right) & \text{if } \Omega_1 = 0 \end{cases} \\
&= \begin{cases} O\left(N^{-3} + N^{-4}h_N^{-1} + N^{-5}h_N^{-2} + N^{-6}h_N^{-3}\right) & \text{if } \Omega_1 > 0 \\ O\left(N^{-3}h_N^4 + N^{-4}h_N^3 + N^{-4}h_N^{-1} + N^{-5}h_N^{-2} + N^{-6}h_N^{-3}\right) & \text{if } \Omega_1 = 0 \end{cases}
\end{aligned}$$

This suggestsd

$$\begin{aligned}
\sigma_N^{-4}\operatorname{Var}\left(\sum_{i=1}^{N} Y_{N,i}^2\right) &= \begin{cases} O\left(\left(N^{-2}h_N^{-1} + N^{-1}\right)^{-2}\left(N^{-3} + N^{-4}h_N^{-1} + N^{-5}h_N^{-2} + N^{-6}h_N^{-3}\right)\right) & \text{if } \Omega_1 > 0 \\ O\left(\left(N^{-2}h_N^{-1}\right)^{-2}\left(N^{-3}h_N^4 + N^{-4}h_N^3 + N^{-4}h_N^{-1} + N^{-5}h_N^{-2} + N^{-6}h_N^{-3}\right)\right) & \text{if } \Omega_1 = 0 \end{cases} \\
&= \begin{cases} O\left(N^{-1} + N^{-2}h_N^{-1}\right) & \text{if } \Omega_1 > 0 \\ O\left(N h_N^6 + h_N + N^{-1} + N^{-2}h_N^{-1}\right) & \text{if } \Omega_1 = 0 \end{cases}
\end{aligned}$$

If $N^{-2} \ll h \ll N^{-1/6}$, then this converge to zero and the stability condition holds. Now that we've verified both conditions, the CLT implies

$$\sigma_N^{-1}\left(\hat{f}_W(w_0) - \mathbb{E}\hat{f}_W(w_0)\right) \rightsquigarrow \mathrm{N}(0,1).$$

Under the under-smoothing bandwidth condition $h_N \ll N^{-2/5}$, the bias is of smaller order $\mathrm{Bias}^2(\hat{f}_W(w_0)) = o_p(\hat{f}_W(w_0) - \mathbb{E}\hat{f}_W(w_0))$. Together these implies

$$\sigma_N^{-1}\left(\hat{f}_W(w_0) - f_W(w_0)\right) \rightsquigarrow \mathrm{N}(0,1).$$

□

Proof of Lemma 1.4.1

Proof.

$$\begin{aligned}
&\mathrm{Var}^*\left(\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right)^{\Delta}\right) \\
&= \binom{N}{2}^{-2} \sum_{i_1<j_1}\sum_{i_2<j_2} \mathrm{Cov}\left(\Delta_{N,i_1j_1}, \Delta_{N,i_2j_2}\right)\left(K_{N,h_N,i_1j_1} - \bar{K}_{N,h_N,\cdot\cdot}\right)\left(K_{N,h_N,i_2j_2} - \bar{K}_{N,h_N,\cdot\cdot}\right) \\
&= \binom{N}{2}^{-2}\Bigg\{\upsilon_{N,2}\left(\sum_{i<j} K^2_{N,h_N,ij} - \binom{N}{2}\bar{K}^2_{N,h_N,\cdot\cdot}\right) + 2\upsilon_{N,1}\Bigg[\sum_{i<j<k}\Big(K_{N,h_N,ij}K_{N,h_N,ik} \\
&\qquad + K_{N,h_N,ij}K_{N,h_N,jk} + K_{N,h_N,ik}K_{N,h_N,jk}\Big) - \frac{1}{2}N(N-1)(N-2)\bar{K}^2_{N,h_N,\cdot\cdot}\Bigg] \\
&\qquad + 2\upsilon_{N,0}\Bigg[\sum_{i<j<k<l}\Big(K_{N,h_N,ij}K_{N,h_N,kl} + K_{N,h_N,ik}K_{N,h_N,jl} \\
&\qquad + K_{N,h_N,il}K_{N,h_N,jk}\Big) - \frac{1}{8}N(N-1)(N-2)(N-3)\bar{K}^2_{N,h_N,\cdot\cdot}\Bigg]\Bigg\} \\
&= a_N S_2 + b_N S_3 - c_N S_4
\end{aligned}$$

where

$$\begin{aligned}
S_2 &= \frac{1}{\binom{N}{2}}\sum_{i<j} K_{ij}^2 \\
S_3 &= \frac{1}{\binom{N}{3}}\sum_{i<j<k}\frac{1}{3}\left(K_{ij}K_{ik} + K_{ij}K_{jk} + K_{ik}K_{jk}\right) \\
S_4 &= \frac{1}{\binom{N}{4}}\sum_{i<j<k<l}\frac{1}{3}\left(K_{ij}K_{kl} + K_{ik}K_{jl} + K_{il}K_{jk}\right)
\end{aligned}$$

and

$$\begin{aligned}
a_N &= \upsilon_{N,2}\binom{N}{2}^{-1} - \left(\upsilon_{N,2}\binom{N}{2}^{-1} + \upsilon_{N,1}\frac{4(N-2)}{N(N-1)} + \upsilon_{N,0}\frac{(N-2)(N-3)}{N(N-1)}\right)\frac{2}{N(N-1)} \\
&= \upsilon_{N,2}\binom{N}{2}^{-1}(1+o(1)) \\
b_N &= \upsilon_{N,1}\frac{4(N-2)}{N(N-1)} - \left(\upsilon_{N,2}\binom{N}{2}^{-1} + \upsilon_{N,1}\frac{4(N-2)}{N(N-1)} + \upsilon_{N,0}\frac{(N-2)(N-3)}{N(N-1)}\right)\frac{4(N-2)}{N(N-1)} \\
&= \upsilon_{N,1}\frac{4(N-2)}{N(N-1)}(1+o(1)) \\
c_N &= \left[\upsilon_{N,0} - \left(\upsilon_{N,2}\binom{N}{2}^{-1} + \upsilon_{N,1}\frac{4(N-2)}{N(N-1)} + \upsilon_{N,0}\frac{(N-2)(N-3)}{N(N-1)}\right)\right]\frac{(N-2)(N-3)}{N(N-1)} \\
&= \left(\upsilon_{N,2}\binom{N}{2}^{-1} + \upsilon_{N,1}\frac{4(N-2)}{N(N-1)}\right)(1+o(1)).
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}^*\left(\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right)^{\Delta}\right) &= a_N S_2 + b_N S_3 - c_N S_4 \\
&= \left[\upsilon_{N,2}\binom{N}{2}^{-1}(S_2 - S_4) + \upsilon_{N,1}\frac{4(N-2)}{N(N-1)}(S_1 - S_4)\right](1+o_p(1))
\end{aligned} \tag{1.25}$$

Now use the following lemma and the assumption 1.4.1. This lemma is a result of lemma 1.6.2 and 1.6.1.

Lemma 1.6.3. *If assumptions 1.2.1, 1.2.2, 1.2.3 hold, and $h_N \ll N^{-1/4}$, then*

(a) $h_N S_2 \rightsquigarrow_p \Omega_2$

(b) $S_3 \rightsquigarrow_p \mathbb{E}\left[f^2(w|A_i)\right] = \Omega_1 + f_W^2(w)$

(c) $S_4 \rightsquigarrow_p f_W^2(w)$.

Equation (1.25) becomes

$$\mathrm{Var}^*\left(\left(\hat{f}_{N,h_N} - \mathbb{E}\hat{f}_{N,h_N}\right)^{\Delta}\right) = \left[N^{-1}\left(1 - \frac{1}{N-1}\right)4\Omega_1 \cdot \upsilon_1 + \binom{N}{2}^{-1} h_N^{-1}\Omega_2 \cdot \upsilon_2\right](1+o_p(1)).$$

□

Proof of Conditional Lindeberg's Condition In Probability

Lemma 1.6.4 (Conditional Lindeberg's Condition In Probability). *If assumptions 1.2.1, 1.2.2, 1.2.3, 1.4.1 hold and $h_N \ll N^{-1/4}$, for all $\epsilon > 0$,*

$$T_N(\epsilon) = \sum_{i=1}^{N} \mathbb{E}\left[\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot}^{*} - \bar{K}_{\cdot\cdot}\right)\right|^2 \mathbb{1}\left(\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot}^{*} - \bar{K}_{\cdot\cdot}\right)\right| > \epsilon\right) \middle| \mathcal{F}_N\right] \rightsquigarrow_p 0,$$

where $\mathcal{F}_N \equiv \sigma\left(W_{ij}, 1 \leq i < j \leq N\right)$.

Proof.

$$\begin{aligned}
T_N(\epsilon) &= \sum_{i=1}^{N} \mathbb{E}\left[\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot}^{*} - \bar{K}_{\cdot\cdot}\right)\right|^2 \mathbb{1}\left(\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot}^{*} - \bar{K}_{\cdot\cdot}\right)\right| > \epsilon\right) \middle| \mathcal{F}_N\right] \\
&= \sum_{i=1}^{N} \left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot} - \bar{K}_{\cdot\cdot}\right)\right|^2 \mathbb{1}\left(\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot} - \bar{K}_{\cdot\cdot}\right)\right| > \epsilon\right)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}T_N(\epsilon) &= N \cdot \mathbb{E}\left[\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot} - \bar{K}_{\cdot\cdot}\right)\right|^2 \mathbb{1}\left(\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot} - \bar{K}_{\cdot\cdot}\right)\right| > \epsilon\right)\right] \\
&\leq N \cdot \frac{1}{\epsilon^2}\mathbb{E}\left[\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot} - \bar{K}_{\cdot\cdot}\right)\right|^4\right] \\
&= \frac{8}{\epsilon^2 N^3 \sigma_N^4}\mathbb{E}\left[\left|\bar{K}_{i\cdot} - f_W(w_0)\right|^4 + \left|\bar{K}_{\cdot\cdot} - f_W(w_0)\right|^4\right] \\
&\leq \frac{16}{\epsilon^2 N^3 \sigma_N^4}\mathbb{E}\left[\left(\bar{K}_{i\cdot} - f_W(w_0)\right)^4\right] \\
&= \begin{cases} O\left(\frac{1}{N^3\left(N^{-1}+N^{-2}h_N^{-1}\right)^2}\right) \cdot O\left(1 + N^{-1}h_N^{-1} + N^{-2}h_N^{-2} + N^{-3}h_N^{-3}\right) & \text{if } \Omega_1 > 0 \\ O\left(\frac{1}{N^3\left(N^{-2}h_N^{-1}\right)^2}\right) \cdot O\left(h^8 + N^{-1}h_N^{3} + N^{-2}h_N^{-2} + N^{-3}h_N^{-3}\right) & \text{if } \Omega_1 = 0 \end{cases} \\
&= \begin{cases} O\left(N^{-1} + N^{-2}h_N^{-1}\right) & \text{if } \Omega_1 > 0 \\ O\left(Nh_N^{10} + h_N^{5} + N^{-1} + N^{-2}h_N^{-1}\right) & \text{if } \Omega_1 = 0 \end{cases} \\
&= o(1).
\end{aligned}$$

The second line is due to $\mathbb{1}\left(\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot} - \bar{K}_{\cdot\cdot}\right)\right| > \epsilon\right) \leq \frac{1}{\epsilon^2}\left|\frac{1}{N\sigma_N}\left(\bar{K}_{i\cdot} - \bar{K}_{\cdot\cdot}\right)\right|^2$. The third line is due to $|x+y|^4 \leq 8(|x|^4+|y|^4)$. The fourth line is due to $\bar{K}_{\cdot\cdot} - f_W(w) = N^{-1}\sum_i(\bar{K}_{i\cdot} - f_W(w_0))$ and $\mathbb{E}(\bar{K}_{\cdot\cdot} - f_W(w_0))^4 \leq \mathbb{E}[N^{-1}\sum_i(\bar{K}_{i\cdot} - f_W(w_0))^4] = \mathbb{E}[(\bar{K}_{i\cdot} - f_W(w_0))^4]$. The fifth line is due to 1.3.1(b) and the fact that

$$
\begin{aligned}
&\mathbb{E}\left[\left(\bar{K}_{i\cdot} - f_W(w_0)\right)^4\right] \\
&= O\Bigg(\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)\left(K_{13} - f_W(w_0)\right)\left(K_{14} - f_W(w_0)\right)\left(K_{15} - f_W(w_0)\right)\right] \\
&\qquad + N^{-1}\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)^2\left(K_{13} - f_W(w_0)\right)\left(K_{14} - f_W(w_0)\right)\right] \\
&\qquad + N^{-2}\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)^3\left(K_{13} - f_W(w_0)\right) + \left(K_{12} - f_W(w_0)\right)^2\left(K_{12} - f_W(w_0)\right)^2\right] \\
&\qquad + N^{-3}\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)^4\right]\Bigg) \\
&= \begin{cases} O(1 + N^{-1}h_N^{-1} + N^{-2}h_N^{-2} + N^{-3}h_N^{-3}) & \text{if } \Omega_1 > 0 \\ O(h_N^8 + N^{-1}h_N^3 + N^{-2}h_N^{-2} + N^{-3}h_N^{-3}) & \text{if } \Omega_1 = 0 \end{cases},
\end{aligned}
$$

the second line of which is based on implications of lemma 1.6.2 and equation (1.23)

$$
\begin{aligned}
&\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)\left(K_{13} - f_W(w_0)\right)\left(K_{14} - f_W(w_0)\right)\left(K_{15} - f_W(w_0)\right)\right] = \begin{cases} O(1) & \text{if } \Omega_1 > 0 \\ O(h_N^8) & \text{if } \Omega_1 = 0 \end{cases} \\
&\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)^2\left(K_{13} - f_W(w_0)\right)\left(K_{14} - f_W(w_0)\right)\right] = \begin{cases} O(h_N^{-1}) & \text{if } \Omega_1 > 0 \\ O(h_N^3) & \text{if } \Omega_1 = 0 \end{cases} \\
&\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)^3\left(K_{13} - f_W(w_0)\right) + \left(K_{12} - f_W(w_0)\right)^2\left(K_{12} - f_W(w_0)\right)^2\right] = O(h_N^{-2}) \\
&\mathbb{E}\left[\left(K_{12} - f_W(w_0)\right)^4\right] = O(h_N^{-3}).
\end{aligned}
$$

$T_N(\epsilon) > 0$ and $\mathbb{E}T_N(\epsilon) \to 0$ as $N \to \infty$ implies $T_N(\epsilon) \xrightarrow{\text{P}} 0$. $\square$

Chapter 2

Nonparametric Dyadic Regression [1]

2.1 Introduction and Summary

Let $i = 1, \ldots, N$ index a simple random sample of units drawn from some large population. For each unit we observe the vector of regressors X_i and, for each of the $N(N-1)$ ordered pairs of units, or *directed dyads*, we observe the "dyadic" outcome Y_{ij} (e.g., total exports from country i to country j). The outcomes Y_{ij} and Y_{kl} are independent if their indices are disjoint, but dependent otherwise (e.g., exports from Japan to Korea may covary with those from Japan to Vietnam).

Let $W_{ij} = \left(X_i', X_j'\right)'$; using the sampled data we seek to construct a nonparametric estimate of the mean regression function

$$g(W_{ij}) \stackrel{def}{\equiv} \mathbb{E}\left[\left. Y_{ij} \right| X_i, X_j\right]. \tag{2.1}$$

We present two sets of results. First, we calculate lower bounds on the minimax risk for estimating the regression function at (i) a point and (ii) under the infinity norm. Second, we calculate (i) pointwise and (ii) uniform convergence rates for the dyadic analog of the familiar Nadaraya-Watson (NW) kernel regression estimator. We show that the NW kernel regression estimator achieves the optimal rates suggested by our risk bounds when an appropriate bandwidth sequence is chosen.

Analogous results are widely available in the i.i.d. setting. For nonparametric regression risk bounds see, for example, Stone (1980, 1982) and Ibragimov and Has' Minskii (1982, 1984). Tsybakov (2008) provides a masterful synthesis of these results, from which we draw in formulating our own proofs.

Uniform convergence of kernel averages with i.i.d. data, as well as stationary strong mixing data, have been studied by, for example, Newey (1994) and Hansen (2008) respectively. The latter paper includes additional references to the extensive literature in this area. Our uniform convergence proofs build upon those of Hansen (2008). Nonparametric density esti-

[1] This chapter is joint work with Bryan Graham and James Powell.

mation with dyadic data was first considered by Graham et al. (2019); Chiang et al. (2019) present uniform convergence results for dyadic density estimators.[2]

Our results provide insight in the structure of dyadic nonparametric estimation problems. Our minimax risk bounds suggest that, N, the number of units, *not* $n \overset{def}{\equiv} N \times (N-1)$, the number of dyadic outcomes, is the relevant "sample size" for dyadic estimation problems. This is consistent with the long standing intuition among empirical researchers that dyadic dependence makes inference less precise (see Aronow et al. (2017) and the references cited therein), as well as with a small, but growing, number of more formal rates-of-convergence results (cf., Graham, 2020a).

More surprisingly, we find that the relevant dimension of our estimation problem is just $d_X = \dim(X_i)$, not $d_W = 2d_X$. We provide two intuitions for this fact. The first, described below, stems from the thought experiment underlying our minimax risk bound calculations. The second, arises from the fact that the Hájek projection of the NW estimator has a "partial-mean-like" structure. As is well known, averaging over the marginal distribution of some regressors, while holding the remaining ones fixed, improves rates-of-convergence (e.g., Newey, 1994; Linton and Nielsen, 1995).

Graham (2020a) surveys empirical studies in economics utilizing dyadic data. Interest in, as well as the availability of, such data are growing in economics, other academic fields, and in enterprise settings. This paper provides an initial set of results for nonparametric regression with dyadic data. These results are, of course, of direct interest. They should, as has been true with their i.i.d. predecessors, also be useful for proving consistency of two-step semiparametric M-estimators under dyadic dependence (see Chiang et al. (2019) for some results on double machine learning with dyadic data).

2.2 Lower Bounds on the Minimax Risk

Let $i = 1, \ldots, N$ index a simple random sample of units drawn from some large population. The econometrician observes the vector of regressors, X_i, for each sampled unit as well as the scalar outcome, Y_{ij}, for each directed pair of sampled units (i.e., each directed dyad). Let $\mathbf{Z}_N = (X_1, \ldots, X_N, Y_{ij}, 1 \leq i \neq j \leq N)$ be the observable data when N units are sampled. The regression function of interest is (2.1) above. The goal is to construct a nonparametric estimate of $g : \mathbb{R}^{d_W} \to \mathbb{R}$ where $d_W = 2d_x$.

We assume that Y_{ij} is generated according to the following conditionally independent dyad (CID) model (cf., Graham, 2020a, Section 3.3).

$$Y_{ij} = h(X_i, X_j, U_i, U_j, V_{ij}). \tag{2.2}$$

Random sampling ensures that (X_i, U_i) is independent and identically distributed for $i = 1, \ldots, N$. We further assume that $\{(V_{ij}, V_{ji})\}_{1 \leq i < j \leq N}$ are i.i.d. and indepenent of

[2] It is possible that the methods of inference presented in Chiang et al. (2019) could be adapted to our setting.

$\mathbf{X} = (X_1, \ldots, X_N)'$ and $\mathbf{U} = (U_1, \ldots, U_N)$. Here h is an unknown function, often called the *graphon*. This set-up, which can also be derived as an implication of more primitive exchangeability assumptions, has the following implications (see Graham (2020a,b) for additional discussion):

1. The Y_{ij} are relatively exchangeable given the W_{ij}. Namely, the conditional distribution of $\mathbf{Y}$ is invariant across permutations of the indices $\sigma : \mathbb{N} \to \mathbb{N}$ satisfying the restriction $[W_{\sigma(i)\sigma(j)}] \overset{d}{=} [W_{ij}]$:
$$[Y_{ij}] \overset{d}{=} [Y_{\sigma(i)\sigma(j)}].$$

2. Y_{ij} and Y_{kl} are independent if their indices are disjoint.

3. Y_{ij} and Y_{kl} are dependent (unconditionally or conditionally given $X_1, \ldots, X_N$) if they share at least one index in common.

The statistical problem is to estimate the regression function g when the only prior restriction on it is that it belongs to the Hölder class of functions.

Definition 2.2.1. (Hölder Class) Given a vector $s = (s_1, \ldots, s_d)$, define $|s| = s_1 + \cdots + s_d$ and
$$D^s = \frac{\partial^{s_1+\cdots+s_d}}{\partial^{s_1} w_1 \cdots \partial^{s_d} w_d}.$$
Let β and L be two positive numbers. The *Hölder class* $\Sigma(\beta, L)$ on $\mathbb{R}^d$ is defined as the set of $l = \lfloor \beta \rfloor$ times differentiable functions $g : \mathbb{R}^d \to \mathbb{R}$ whose partial derivative $D^s g$ satisfies
$$|D^s g(w) - D^s g(w')| \leq L||w - w'||_\infty^{\beta - l}, \quad \forall w, w' \in \mathbb{R}^d$$
for all s such that $|s| = \lfloor \beta \rfloor$. $\lfloor \beta \rfloor$ denotes the greatest integer strictly less than the real number β.

An estimator $\hat{g}_N$ is a function $w \mapsto \hat{g}_N(w) = \hat{g}_N(w, \mathbf{Z}_N)$ measurable with respect to $\mathbf{Z}$. Our first result establishes a lower bound on the minimax risk for estimating the regression function at a single point and under the infinity norm. We state this result under a Gaussian error assumption, which simplifies the proof.

Theorem 2.2.1. (Minimax Risk Lower Bound) *Suppose that $\beta > 0$ and $L > 0$; X_i is continuously distributed on $\mathbb{R}^{d_X}$ with density f and $\sup_x f(x) \leq B_3 < \infty$; and Y_{ij} is generated according to the following nonparametric regression model:*
$$Y_{ij} = g(W_{ij}) + e_{ij}, \quad i \neq j,$$
with $e_{ij} = U_i + U_j + V_{ij}$, $U_i \overset{\text{iid}}{\sim} N(0, 1)$, and $V_{ij} \overset{\text{iid}}{\sim} N(0, 1)$, then

(i) *For all* $w \in \mathbb{R}^{d_W}$,

$$\liminf_{N\to\infty} \inf_{\hat{g}_N} \sup_{g\in\Sigma(\beta,L)} \mathbb{E}_g \left[N^{\frac{2\beta}{2\beta+d_X}} \left(\hat{g}_N(w) - g(w)\right)^2 \right] \geq c_1,$$

where $c_1 > 0$ *depends only on* β *and* L.

(ii)

$$\liminf_{N\to\infty} \inf_{\hat{g}_N} \sup_{g\in\Sigma(\beta,L)} \mathbb{E}_g \left[\left(\frac{N}{\ln N}\right)^{\frac{2\beta}{2\beta+d_X}} ||\hat{g}_N - g||_\infty^2 \right] \geq c_2,$$

where $c_2 > 0$ *also depends only on* β *and* L.

Our proof follows the general recipe outlined in Chapter 2 of Tsybakov (2008). The lower bound at a point is based on Le Cam's method of two hypotheses. The lower bound under the infinity norm is based on Fano's method of multiple hypotheses.

The key, and novel, step in our proof involves constructing hypotheses close enough to one other in terms of Kullback-Leibler (KL) divergence while being at the same time different enough in terms of the target regression function.

An essential feature of our construction is additive separability of the regression functions. In the hypotheses we consider, $Y_{ij} = k(X_i) + k(X_j) + U_i + U_j + V_{ij}$. Next suppose we also observe $T_i \stackrel{def}{\equiv} k(X_i) + U_i$. Observe that $(X_i, T_i, i = 1, \ldots, N)$ is sufficient with respect to $(X_i, T_i, i = 1, \ldots, N, Y_{kl}, 1 \leq k \neq l \leq N)$ for the parameter k.

It is well-known that the optimal rates of convergence for estimating k using iid data $(X_i, T_i, i = 1, \ldots, N)$ are $N^{-\frac{\beta}{2\beta+d_X}}$ pointwise and $\left(\frac{N}{\ln N}\right)^{-\frac{\beta}{2\beta+d_X}}$ for the infinity norm. We expect the rates for estimating g to be no faster than these. The proof of Theorem 2.2.1 makes this intuition rigorous.

Relative to its iid counterpart, there are two distinctive features of Theorem 2.2.1. First, the relevant sample size is not the number of observed dyadic outcomes $n = N \times (N-1)$, but instead the number of sampled units, N. Dependence across outcomes sharing indices in common is strong enough to slow down the feasible rate of convergence. Second, although the regression function has $d_W = 2d_X$ arguments, the relevant dimension reflected in the rate of convergence result is just d_X (i.e., just half of what might naively be expected).

The form of our constructed hypotheses provides one intuition for this second finding: clearly the relevant dimension of the problem of estimating $k(x)$ is just d_X. Relatedly this finding is consistent with those of Linton and Nielsen (1995) in their analysis of additively separable, but otherwise nonparametric, regression functions (see also Newey (1994)).

The pairwise structure of dyadic data results in apparent data abundance (sample N agents, but observe $O(N^2)$ outcomes!). This abundance is both illusory, in the sense that the effective sample size is indeed just N, and real, in the sense that availability of the pairwise outcome data allows for an effective reduction in the dimensionality of the problem via partial mean like average (as in Newey (1994) and Linton and Nielsen (1995) in a different context).

2.3 Kernel Estimator of Dyadic Regression

In this section we study the properties of a specific nonparametric regression estimator. Namely, the dyadic analog of the well-known Nadaraya-Watson (NW) kernel regression estimator. While our results are specific to this estimator, they could, for example, be extended to apply to local linear regression (e.g., Hansen, 2008).

The dyadic NW kernel regression estimator is

$$\hat{g}_N(w) := \frac{\sum_{1\leq i\neq j\leq N} K_{ij,N}(w)Y_{ij}}{\sum_{1\leq i\neq j\leq N} K_{ij,N}(w)}, \tag{2.3}$$

where

$$K_{ij,N}(w) := \frac{1}{h_N^{d_W}} K\left(\frac{W_{ij}-w}{h_N}\right),$$

K is a fixed multivariate kernel function, and h_N is a vanishing bandwidth sequence.

We first develop a sequence of results useful for bounding the variance of kernel objects of the form

$$\hat{\Psi}_N(w) := \frac{1}{N(N-1)} \sum_{1\leq i\neq j\leq N} Y_{ij}K_{ij,N}(w) \tag{2.4}$$

and then apply these results to the NW regression estimator. We then bound the NW estimator's bias and combine the two sets of results to formulate a risk bound.

Variance Bound and Uniform Convergence

Here we are interested in bounding the deviation of $\hat{\Psi}_N(w)$ from its mean. We begin with a presentation of our maintained assumptions.

Assumption 2.3.1 (Model)**.** The data generating process is as described in Section 2.2 with

(i) X_i continuously distributed with marginal density $f(x)$ s.t. $\sup_{x\in\mathbb{R}^{d_X}} f(x) \leq B_3 < \infty$;

(ii) $\sup_{x_1,x_2\in\mathbb{R}^{d_X}} \mathbb{E}\left[|Y_{12}|^2 \middle| (X_1,X_2)=(x_1,x_2)\right] \cdot f(x_1)f(x_2) \leq B_4 < \infty$,
$\sup_{x_1,x_2,x_3\in\mathbb{R}^{d_X}} \mathbb{E}\left[|Y_{12}Y_{13}| \middle| (X_1,X_2,X_3)=(x_1,x_2,x_3)\right] \cdot f(x_1)f(x_2)f(x_3) \leq B_5 < \infty$.

Condition (i) is a standard condition in the context of kernel estimation, while (ii) ensures that various second moments appearing in our variance calculations are finite.

Assumption 2.3.2 (Kernel, Part A)**.** $\sup_{w\in\mathbb{R}^{d_W}} |K(w)| \leq K_{\max} < \infty$, $\int_{w\in\mathbb{R}^{d_W}} |K(w)|\mathrm{d}w \leq B_1 < \infty$, and $\sup_{x\in\mathbb{R}^{d_X}} \int |K(x,x')|\mathrm{d}x' \leq B_2 < \infty$.

Assumption 2.3.2 is satisfied by many widely-used multivariate kernel functions. Our first result holds under Assumptions 2.3.1 and 2.3.2.

Theorem 2.3.1 (Variance Bound). *Under Assumptions 2.3.1 and 2.3.2, and the bandwidth condition $Nh_N^{d_X} \to \infty$ as $N \to \infty$, there exists a constant $M_0 < \infty$ such that for N sufficiently large*

$$\mathrm{Var}\left(\hat{\Psi}_N(w)\right) \leq \frac{M_0}{Nh_N^{d_X}}$$

for all $w \in \mathbb{R}^{d_W}$.

A proof is available in the appendix. Mirroring our risk bound results, two features of Theorem 2.3.1 merit comment. First, N not $n = N \times (N-1)$ appears in the denominator. This is due to the effects of dependence across dyads sharing units in common. Second, the relevant dimension of the problem is d_X, not $d_W = 2d_X$, this reflects the U-statistic like structure of kernel weighted averages and the partial mean like averaging this structure induces.

To establish uniform convergence, we need additional moment conditions on Y_{ij} as well as some smoothness conditions on the kernel K. As in Hansen (2008), we require the kernel to either have bounded support and be Lipschitz or have bounded derivatives and an integrable tail. See Hansen (2008) for additional discussion about these conditions. As with Assumption 2.3.2 above, most commonly used kernels satisfy these conditions.

Assumption 2.3.3 (Regularity Condition). (i) For some $s > 2$, $\mathbb{E}|Y_{12}|^s < \infty$ and $\sup_{x_1,x_2 \in \mathbb{R}^{d_X}} \mathbb{E}\left[|Y_{12}|^s \middle| (X_1, X_2) = (x_1, x_2)\right] \cdot f(x_1, x_2) \leq B_{4,s} < \infty$;

(ii) For some $\Lambda_1 < \infty$ and $L < \infty$, either (a) or (b) holds

(a) $K(w) = 0$ for $||w|| > L$, and $|K(w) - K(w')| \leq \Lambda_1 ||w - w'||$ for all $w, w' \in \mathbb{R}^{2d}$

(b) $K(w)$ is differentiable, $\left|\left|\frac{\partial}{\partial w}K(w)\right|\right| \leq \Lambda_1$, where $\left|\left|\frac{\partial}{\partial w}K(w)\right|\right| = \left|\left|\left(\frac{\partial}{\partial w_1}K(w), \ldots \frac{\partial}{\partial w_{2d}}K(w)\right)\right|\right|_\infty$, and for some $\nu > 1$, $\left|\left|\frac{\partial}{\partial w}K(w)\right|\right| \leq \Lambda_1 ||w||^{-\nu}$ for $||w|| > L$.

Part (ii) coincides with Assumption 3 in Hansen (2008). This assumption implies that for all $||w_1 - w_2|| \leq \delta \leq L$,

$$|K(w_2) - K(w_1)| \leq \delta K^*(w_1),$$

where $K^*(u)$ satisfies Assumption 2.3.1. If case (a) holds, then $K^*(u) = 2d\Lambda_1 \mathbb{1}(||u|| \leq 2L)$. If case (b) holds, then, $K^*(u) = 2d[\Lambda_1 \mathbb{1}(||u|| \leq 2L) + (||u|| - L)^{-\nu} \mathbb{1}(||u|| > 2L)]$. In both cases K^* is bounded and integrable and therefore satisfies Assumption 2.3.1.

Define

$$a_N := \left(\frac{\ln N}{Nh_N^{d_X}}\right)^{1/2}.$$

Theorem 2.3.2 (Weak Uniform Convergence). *Under Assumptions 2.3.1, 2.3.2, 2.3.3, and the bandwidth conditions*

$\max\left\{\min\left\{(a_N h_N^{2d_X})^{-\frac{1}{s-1}}, [N^2(\ln(\ln N))^2 \ln N]^{\frac{1}{s}}\right\}, a_N^{-\frac{1}{s-1}}\right\} \ll \min\left\{a_N^{-1}, \frac{N}{\ln N} h_N^{\frac{3}{2}d_X}\right\}$ and $\frac{N}{\ln N} h_N^{d_X} \to \infty$, we have for any $q > 0$, $c_N = N^q$,

$$\sup_{||w|| \leq c_N} \left|\hat{\Psi}_N(w) - \mathbb{E}\hat{\Psi}_N(w)\right| = O_P(a_N).$$

This theorem establishes uniform convergence of $\hat{\Psi}_N(w)$ to its mean in probability over an expanding set with radius growing at a polynomial rate.

In the proof, we decompose $\hat{\Psi}_N(w)$ into two parts

$$\hat{\Psi}_N(w) = \tilde{\Psi}_N(w) + R_N(w),$$

in which $\tilde{\Psi}_N(w) = \frac{1}{N(N-1)} \sum_{1 \leq i \neq j \leq N} Y_{ij} \cdot \mathbb{1}\left(|Y_{ij}| < \tau_N\right) K_{ij,N}$ is a truncated version of $\hat{\Psi}_N(w)$ with a carefully chosen threshold parameter τ_N and $R_N(w)$ is a residual. The boundedness induced by this truncation is technically convenient as it facilitates the application of various concentration inequalities. To establish concentration of $\tilde{\Psi}_N$, we apply Bernstein inequality to its Hájek Projection (i.e., to the first-order terms in the Hoeffding decomposition) and apply Arcones and Gine (1993)'s concentration inequalities for degenerate U-statistics to the second-order terms in the Hoeffding decompositon. Both these bounds requires the truncation threshold to be small enough. To bound the magnitude of the residual R_N, we can either apply a triangular inequality to bound the sup of its first moment or use the Borel-Cantelli Lemma to bound its probability of being nonzero. Both these bounds requires the truncation threshold to be large.

A proper truncation threshold satisfying both requirements exists only if the bandwidth sequence satisfies the condition

$$\max\left\{\min\left\{(a_N h_N^{2d_X})^{-\frac{1}{s-1}}, [N^2(\ln(\ln N))^2 \ln N]^{\frac{1}{s}}\right\}, a_N^{-\frac{1}{s-1}}\right\} \ll \min\left\{a_N^{-1}, \frac{N}{\ln N} h_N^{\frac{3}{2}d_X}\right\}.$$

The complicated form of this condition is technical in nature. When all (conditional) moments of Y_{12} are bounded, such that $s = \infty$ (of Assumption 2.3.3 above), this condition simplifies to $\frac{N}{\ln N} h_N^{\frac{3}{2}d_X} \gg 1$.

In order to state the weak uniform convergence result for the kernel regression estimator $\hat{g}_N$, we need additional smoothness assumptions on the kernel. As in other applications of kernel estimation, these assumptions are employed for bias reduction purpose.

Assumption 2.3.4 (Kernel, Part B)**.**

$$\int_{\mathbb{R}^{d_W}} w_1^{l_1} \cdots w_{d_W}^{l_{d_W}} K(w) dw = \begin{cases} 1, & \text{if } l_1 = \cdots = l_{d_W} = 0 \\ 0, & \text{if } (l_1, \ldots, l_{d_W})' \in \mathbb{Z}_+^{d_W} \text{ and } l_1 + \cdots + l_{d_X} < \beta \end{cases}$$

We can now give a uniform convergence result for the NW regression estimator under dyadic dependence over a sequence of expanding sets.

Theorem 2.3.3. *Suppose $f_W, g \in \Sigma(\beta, L)$ and $\delta_N = \inf_{||w|| \leq C_N} f_W(w) > 0$, $\delta_N^{-1} a_N^* \to 0$ where $a_N^* := \left(\frac{\ln N}{N h_N^{d_X}}\right)^{1/2} + h_N^{\beta}$. Under the Assumptions of Theorem, 2.3.2 and Assumption 2.3.4*

$$\sup_{||w|| \leq C_N} |\hat{g}_N(w) - g(w)| = O_p(\delta_N^{-1} a_N^*).$$

The optimal convergence rate is

$$\sup_{||w|| \leq C_N} |\hat{g}_N(w) - g(w)| = O_p\left(\delta_N^{-1}\left(\frac{\ln N}{N}\right)^{\frac{\beta}{2\beta + d_X}}\right).$$

As in the iid case, the KW estimator achieves the optimal rate suggested by Theorem 2.2.1 for a compact set with $C_N = C$. If we look at a sequence of expanding sets approaching the entire space $\mathbb{R}^{d_W}$, then there is an additional penalty term δ_N due to the presence of the denominator $f_W(w)$.

2.4 Appendix: Proofs

This Appendix contains proofs of results in the main text of Chapter 2.

All notation is as established in the main text unless noted otherwise. Equation numbering continues in sequence with that of the main text.

Proof of Theorem 2.2.1

Our method of proof follows the general approach outlined in Chapter 2 of Tsybakov (2008). To prove part (i) we use Le Cam's two-point method to find a lower risk bound for estimation of the regression function at a point. To prove statement (ii), which involves the infinity-norm metric, we use Fano's method.

Proof of statement (i)

Our proof of statement (i) essentially involves checking the conditions, as specially formulated for our dyadic regression problem, of Theorem 2.3 of (Tsybakov, 2008).

For $k = 0, 1$, let P_{kN} be a probability measure for the observed data $\{(X_i', Y_{ij})\}_{1 \leq i \neq j \leq N}$ with regression function g_{kN}. The general reduction scheme outlined in Section 2.2 of Tsybakov (2008), as well as his Theorems 2.1 and 2.2, imply that our Theorem 2.2.1 will hold if we can construct two sequences of hypotheses g_{0N}, g_{1N} such that

(a) the regression functions g_{0N}, g_{1N} are in the Hölder class $\Sigma(\beta, L)$;

(b) $d(\theta_1, \theta_0) = |g_{1N}(w) - g_{0N}(w)| \geq 2A\psi_N$ with $\psi_N = N^{-\frac{\beta}{2\beta + d_X}}$ and $\theta_0 = g_{0N}(w)$ and $\theta_1 = g_{1N}(w)$ for some fixed $w \in \mathbb{X} \times \mathbb{X}$;

(c) the Kullback-Leibler divergence of P_{0N} from P_{1N} is bounded: $\mathrm{KL}(P_{0N}, P_{1N}) \leq \alpha < \infty$.

The "trick" of the proof is choosing these two sequences of hypotheses appropriately. Letting $w = (x_{10}, x_{20})$ we choose the sequences:

$$g_{0N}(x_1, x_2) \equiv 0$$
$$g_{1N}(x_1, x_2) = \frac{Lh_N^\beta}{2}\left[K\left(\frac{x_1 - x_{10}}{h_N}\right) + K\left(\frac{x_1 - x_{20}}{h_N}\right) + K\left(\frac{x_2 - x_{10}}{h_N}\right) + K\left(\frac{x_2 - x_{20}}{h_N}\right)\right]$$

where $h_N = c_0 N^{-\frac{1}{2\beta+d_X}}$ and the function $K : \mathbb{R}^{d_W} \to [0, \infty)$ satisfies

$$K \in \Sigma\left(\beta, 1/2\right) \cap C^\infty(\mathbb{R}^{d_X}) \text{ and } K(x) > 0 \iff ||x||_\infty \in (-1/2, 1/2). \tag{2.5}$$

There exist functions K satisfying this condition. For example, for a sufficiently small $a > 0$, we can take

$$K(x) = \Pi_{i=1}^{d_X} \lambda(x_i), \quad \text{where } \lambda(u) = a\eta(2u) \text{ and } \eta(u) = \exp\left(-\frac{1}{1-u^2}\right)\mathbb{1}(|u| \leq 1).$$

See also Equation (2.34) in Tsybakov (2008).

We verify conditions (a), (b) and (c) in sequence.

Verification of (a) $g_{0N}, g_{1N} \in \Sigma(\beta, L)$

For $s = (\underbrace{s_1, \ldots, s_{d_X}}_{\mathcal{S}_1}, \underbrace{s_{d_X+1}, \ldots, s_{2d_X}}_{\mathcal{S}_2})$ with $|s| = \lfloor \beta \rfloor$, $w = (x_1, x_2)$ and $w' = (x_1', x_2')$, the s^{th} order derivative of g_{1N} is

$$\begin{aligned}
&D^s g_{1N}(w)\\
&= Lh_N^\beta\left[D^s K\left(\frac{x_1 - x_{10}}{h_N}\right) + D^s K\left(\frac{x_1 - x_{20}}{h_N}\right) + D^s K\left(\frac{x_2 - x_{10}}{h_N}\right) + D^s K\left(\frac{x_2 - x_{20}}{h_N}\right)\right]\\
&= \begin{cases} 0 & \text{if } |\mathcal{S}_1| \notin \{0, |s|\} \\ \frac{Lh_N^{\beta-\lfloor\beta\rfloor}}{2}\left[D^{\mathcal{S}_1}K\left(\frac{x_1-x_{10}}{h_N}\right) + D^{\mathcal{S}_1}K\left(\frac{x_1-x_{20}}{h_N}\right)\right] & \text{if } |\mathcal{S}_1| = |s| \\ \frac{Lh_N^{\beta-\lfloor\beta\rfloor}}{2}\left[D^{\mathcal{S}_2}K\left(\frac{x_2-x_{10}}{h_N}\right) + D^{\mathcal{S}_2}K\left(\frac{x_2-x_{20}}{h_N}\right)\right] & \text{if } |\mathcal{S}_1| = 0 \end{cases}.
\end{aligned}$$

Therefore, if $|\mathcal{S}_1| \notin \{0, |s|\}$, then $|D^s g_{1N}(w) - D^s g_{1N}(w')| = 0$; if $|\mathcal{S}_1| = |s|$, then

$$\begin{aligned}
&|D^s g_{1N}(w) - D^s g_{1N}(w')|\\
&= \frac{Lh_N^{\beta-\lfloor\beta\rfloor}}{2}\left[\left|D^{\mathcal{S}_1}K\left(\frac{x_1 - x_{10}}{h_N}\right) - D^{\mathcal{S}_1}K\left(\frac{x_1' - x_{10}}{h_N}\right)\right|\right.\\
&\qquad \left.+ \left|D^{\mathcal{S}_1}K\left(\frac{x_1 - x_{20}}{h_N}\right) - D^{\mathcal{S}_1}K\left(\frac{x_1' - x_{20}}{h_N}\right)\right|\right]\\
&\leq L||x_1 - x_1'||_\infty^{\beta-\lfloor\beta\rfloor}\\
&\leq L||w - w'||_\infty^{\beta-\lfloor\beta\rfloor};
\end{aligned}$$

and, finally, if $|\mathcal{S}_1| = 0$, then

$$\begin{aligned}
&|D^s g_{1N}(w) - D^s g_{1N}(w')| \\
&= \frac{L h_N^{\beta-\lfloor\beta\rfloor}}{2}\left[\left|D^{\mathcal{S}_2} K\left(\frac{x_2 - x_{10}}{h_N}\right) - D^{\mathcal{S}_2} K\left(\frac{x_2' - x_{10}}{h_N}\right)\right|\right. \\
&\qquad \left. + \left|D^{\mathcal{S}_2} K\left(\frac{x_2 - x_{20}}{h_N}\right) - D^{\mathcal{S}_2} K\left(\frac{x_2' - x_{20}}{h_N}\right)\right|\right] \\
&\leq L||x_2 - x_2'||_\infty^{\beta-\lfloor\beta\rfloor} \\
&\leq L||w - w'||_\infty^{\beta-\lfloor\beta\rfloor}.
\end{aligned}$$

Hence $g_{1N} \in \Sigma(\beta, L)$. We also have that $g_{0N} \in \Sigma(\beta, L)$ by inspection.

Verification of (b): $d\left(\theta\left(P_{0N}\right), \theta\left(P_{1N}\right)\right) = |g_{1N}(w) - g_{0N}(w)| \geq 2A\psi_N$ **with** $\psi_N = N^{-\frac{\beta}{2\beta+d}}$

Here we check that our hypotheses are $2s$-separated. We have that

$$\begin{aligned}
|g_{1N}(w) - g_{0N}(w)| &= \frac{L h_N^\beta}{2}\left[2K(0) + K\left(\frac{x_{10} - x_{20}}{h_N}\right) + K\left(\frac{x_{20} - x_{10}}{h_N}\right)\right] \geq 2L h_N^\beta K(0) \\
&= LK(0)\, c_0^\beta \psi_N,
\end{aligned}$$

and hence condition (b) holds with $A = \frac{LK(0)c_0^\beta}{2}$.

Verification of (c): $\mathrm{KL}(P_{0N}, P_{1N}) \leq \alpha < \infty$

This condition allows for the application of part (iii) of Theorem 2.2 in Tsybakov (2008). We begin by establishing some helpful notation. Let $\mathbf{Y} = [Y_{ij}]_{1\leq i,j\leq N}$ be the $N \times N$ adjacency matrix; $\mathbf{G}_k = [g_{kN}(W_{ij})]_{1\leq i,j\leq N}$ for $k = 0, 1$ the associated matrices of regression functions for the two sequences of hypotheses; and $\mathbf{V} = [V_{ij}]_{1\leq i,j\leq N}$ the corresponding matrix of dyadic-specific disturbances. Note the diagonals of each of these matrices consist of "structural" zeros. Further let $\mathbf{U} = [U_i]_{1\leq i\leq N}$ be the $N \times 1$ vector of agent-specific disturbances. Finally let $\mathbf{K}$ be the $N \times 1$ vector with i^{th} element $\frac{L h_N^\beta}{2}\left[K\left(\frac{X_i - x_{10}}{h_N}\right) + K\left(\frac{X_i - x_{20}}{h_N}\right)\right]$.

Let ι_J denote a $J \times 1$ vector of ones, $\underline{0}_{K,J}$ a $K \times J$ matrix of zeros, and I_J the $J \times J$ identity matrix. We also define the following selection matrices:

$$\mathcal{T}_1 = \begin{pmatrix} \iota_{N-1} & 0 & 0 & \cdots & 0 & 0 \\ \underline{0} & \iota_{N-2} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{\binom{N}{2}\times N}, \quad \mathcal{T}_2 = \begin{pmatrix} \underline{0}_{N-1,1} & I_{N-1} \\ \hdashline \underline{0}_{N-2,2} & I_{N-2} \\ \vdots & \vdots \\ \hdashline 0 & 1 \end{pmatrix}_{\binom{N}{2}\times N},$$

from which we form $\mathcal{T} = \mathcal{T}_1 + \mathcal{T}_2$ and, finally, $\mathbf{T} = \iota_2 \otimes \mathcal{T}$. Next let $\mathbf{y} = (\mathrm{vech}(\mathbf{Y}')', \mathrm{vech}(\mathbf{Y})')'$ be the $N(N-1) \times 1$ vectorization of the dyadic outcomes. Similarly let $\mathbf{g}_k$ for $k = 0, 1$ and $\mathbf{v}$ be the corresponding vectorizations of, respectively, $\mathbf{G}_k$ and $\mathbf{V}$.

Using this notation we can write the $N(N-1)\times 1$ vector of composite regression errors $e_{ij} = U_i + U_j + V_{ij}$ as $\mathbf{e} = \mathbf{TU} + \mathbf{v}$ and its variance covariance matrix as

$$\Omega = \mathrm{Var}\left(\mathbf{e}\right) = \mathbf{I}_{N(N-1)\times N(N-1)} + \mathbf{TT}^{T}.$$

Under P_{0N} we have that

$$\mathbf{g}_0 = 0,\ \mathbf{y} = \mathbf{e},\ \mathbf{y}|\mathbf{X} \sim \mathrm{N}\left(0, \Omega\right).$$

While under P_{1N} we instead have that

$$\mathbf{g}_1 = \mathbf{TK},\ \mathbf{y} = \mathbf{TK} + \mathbf{e},\ \mathbf{y}|\mathbf{X} \sim \mathrm{N}\left(\mathbf{TK}, \Omega\right).$$

Let $K_{\max} = \max_u K(u)$ and recall that $h_N = c_0 N^{-\frac{1}{2\beta+d_X}}$. We can now evaluate the KL divergence as follows:

$$\begin{aligned}
\mathrm{KL}\left(P_{0N}, P_{1N}\right) &= \int \log \frac{\mathrm{d}P_{0N}}{\mathrm{d}P_{1N}} \mathrm{d}P_{0N} \\
&= \int \log \frac{p_{0N}(\mathbf{y}|\mathbf{X})}{p_{1N}(\mathbf{y}|\mathbf{X})} \mathrm{d}P_{0N} \\
&= -\frac{1}{2} \int \mathbf{y}^\top \Omega^{-1} \mathbf{y} - (\mathbf{y} - \mathbf{g}_1)^\top \Omega^{-1} (\mathbf{y} - \mathbf{g}_1) \mathrm{d}P_{0N} \\
&= \frac{1}{2} \int \mathbf{g}_1^\top \Omega^{-1} \mathbf{g}_1 \mathrm{d}P_{0N} \\
&= \frac{1}{2} \mathbb{E}_{P_{0N}} \left[\mathbf{K}^\top \mathbf{T}^\top (\mathbf{I} + \mathbf{TT}^\top)^{-1} \mathbf{TK}\right] \\
&\leq \frac{1}{2} \mathbb{E}_{P_{0N}} \left[\mathbf{K}^\top \mathbf{K}\right] \\
&\leq \frac{1}{2} L^2 K_{\max}^2 B_3 h_N^{2\beta+d_X} N \\
&= \frac{1}{2} L^2 K_{\max}^2 B_3 c_0^{2\beta+d_X},
\end{aligned} \tag{2.6}$$

for N large enough such that $Nh_N^{d_X} \geq 1$ and $LK_{\max} h_N^{2\beta}$ bounded above.

In the derivation above, the third equality follows from the form of the multivariate normal density. The weak inequality in line six holds because

$$\begin{aligned}
\mathbf{K}^\top \mathbf{K} - \mathbf{K}^\top \mathbf{T}^\top (\mathbf{I} + \mathbf{TT}^\top)^{-1} \mathbf{TK} &= \mathbf{K}^\top \left[\mathbf{I}_N - \mathbf{T}^\top (\mathbf{I} + \mathbf{TT}^\top)^{-1} \mathbf{T}\right] \mathbf{K} \\
&= \mathbf{K}^\top \left[\mathbf{I}_N + \mathbf{T}^\top \mathbf{T}\right]^{-1} \mathbf{K} \\
&\geq 0.
\end{aligned}$$

Finally, the weak inequality in line seven holds because, using condition (2.5) above,

$$
\begin{aligned}
&\mathbb{E}\left[\left(K\left(\frac{X_i - x_{10}}{h_N}\right) + K\left(\frac{X_i - x_{20}}{h_N}\right)\right)^2\right] \\
&\leq 2\mathbb{E}\left[\left(K\left(\frac{X_i - x_{10}}{h_N}\right)\right)^2 + \left(K\left(\frac{X_i - x_{20}}{h_N}\right)\right)^2\right] \\
&= 2\int \left(K\left(\frac{x - x_{10}}{h_N}\right)\right)^2 + \left(K\left(\frac{x - x_{20}}{h_N}\right)\right)^2 \mathrm{d}F(x) \\
&\leq 2K_{\max}^2 \int \mathbb{1}\left(\left|\frac{x - x_{10}}{h_N}\right| \leq \frac{1}{2}\right) + \mathbb{1}\left(\left|\frac{x - x_{20}}{h_N}\right| \leq \frac{1}{2}\right) \mathrm{d}F(x) \\
&= 2K_{\max}^2 h_N^{d_X} \left[\int \mathbb{1}\left(|u| \leq \frac{1}{2}\right) [f(x_{10} + h_N u) + f(x_{20} + h_N u)] \mathrm{d}u\right] \\
&\leq 4h_N^{d_X} B_3 K_{\max}^2,
\end{aligned}
$$

and where it is also helpful to remind oneself of the definition of $\mathbf{K}$ given earlier.

If we take $c_0 = \left(\frac{2\alpha}{L^2 K_{max}^2 B_3}\right)^{\frac{1}{2\beta + d_X}}$, then we obtain $\mathrm{KL}\left(P_{0N}, P_{1N}\right) \leq \alpha$. This result, and condition (b) above, gives – invoking equations (2.7) and (2.9) on p. 29 of Tsybakov (2008) as well as part (iii) of his Theorem 2.2:

$$
\inf_{\hat{g}_N} \sup_{g \in \Sigma(\beta, L)} \mathbb{E}_g \left[1\left(|g_{1N}(w) - g_{0N}(w)| \geq A\psi_N\right)\right] \geq \max\left(\frac{1}{4}\exp\left(-\alpha\right), \frac{1 - \sqrt{\frac{\alpha}{2}}}{2}\right)
$$

for N large enough. Some rearrangement and the Markov Inequality then yield

$$
\inf_{\hat{g}_N} \sup_{g \in \Sigma(\beta, L)} \mathbb{E}_g \left[N^{\frac{2\beta}{2\beta + d_X}} \left(g_{1N}(w) - g_{0N}(w)\right)^2\right] \geq A^2 \max\left(\frac{1}{4}\exp\left(-\alpha\right), \frac{1 - \sqrt{\frac{\alpha}{2}}}{2}\right).
$$

Since the constant to the right of the inequality only depends on β and L part (i) of the Theorem follows after taking the limit inferior of the expression above as $N \to \infty$.

Proof of statement (ii)

Again let P_{kN} be the probability measure of the observed data $(X_i, Y_{ij}, 1 \leq i \neq j \leq N)$ with the regression function g_{kN}. Theorem 2.5 of Tsybakov (2008) implies that part (ii) will hold if we can construct sequences of hypotheses $P_{0N}, P_{1N}, \ldots, P_{M_N N}$ such that

(a) $g_{0N}, g_{kN} \in \Sigma(\beta, L)$, $k = 1, \ldots, M_N$;

(b) $d\left(\theta_k, \theta_l\right) = ||g_{kN} - g_{lN}||_\infty \geq 2A\psi_N$, $\psi_N = \left(\frac{N}{\ln N}\right)^{-\frac{\beta}{2\beta + d}}$ and $\theta_k = g_{kN}$ and $\theta_l = g_{lN}$ for $k \neq l$ and $k, l = 1, \ldots, M_N$;

(c) $\frac{1}{M_N}\sum_{k=1}^{M_N} KL(P_{kN}, P_{0N}) \leq \alpha \ln M_N$.

Define the hypotheses:

$$\begin{aligned} g_{0N} :&(x_1, x_2) \to 0 \\ g_{kN} :&(x_1, x_2) \to Lh_N^\beta \left[K\left(\frac{x_1 - x_{kN}}{h_N}\right) + K\left(\frac{x_2 - x_{kN}}{h_N}\right)\right] \end{aligned}$$

where $k \in \mathcal{I}_N = \{1, 2, \ldots, m_N\}^{d_X}$, $h_N = c_0\left(\frac{N}{\ln N}\right)^{-\frac{1}{2\beta+d_X}}$, $m_N = \lceil h_N^{-1} \rceil$, $M_N = |\mathcal{I}_N| = m_N^{d_X}$, and for $k = (k_1, k_2, \ldots, k_d)$, $x_{kN} = \left(\frac{k_1-1/2}{m_N}, \frac{k_2-1/2}{m_N}, \ldots, \frac{k_d-1/2}{m_N}\right)$, the function $K : \mathbb{R}^{d_X} \to [0, \infty)$ satisfies (2.5). Notice the supports of these functions for the same N are disjoint. The results follows by verifying conditions (a), (b) and (c). We have already shown that condition (a) holds in the proof of part (i). The condition (b) holds with $A = LK(0)c_0^\beta$ because

$$||g_{kN} - g_{lN}||_\infty \geq |g_{kN}(x_{kN}, x_{kN}) - g_{lN}(x_{kN}, x_{kN})| = 2Lh_N^\beta K(0) = 2LK(0)c_0^\beta \psi_N.$$

To verify condition (c) we evaluate the KL-divergence:

$$\begin{aligned} \frac{1}{M_N}\sum_{k\in\mathcal{I}_N} \mathrm{KL}(P_{kN}, P_{0N}) &\leq \frac{1}{M_N}\sum_{k\in\mathcal{I}_N} \frac{1}{2}\mathbb{E}_{P_{0N}}\left[\mathbf{K}_k^\top \mathbf{K}_k\right] \\ &\leq \frac{1}{M_N}\sum_{k\in\mathcal{I}_N} 2L^2 h_N^{2\beta} K_{\max}^2 \sum_{i=1}^N \int \mathbb{1}\left(\left|\frac{x_i - x_{kN}}{h_N}\right| \leq \frac{1}{2}\right) \mathrm{d}F(x_i) \\ &= \frac{1}{M_N} 2L^2 h_N^{2\beta} K_{\max}^2 \sum_{i=1}^N \int \sum_{k\in\mathcal{I}_N} \mathbb{1}\left(\left|\frac{x_i - x_{kN}}{h_N}\right| \leq \frac{1}{2}\right) \mathrm{d}F(x_i) \\ &\leq 2L^2 h_N^{2\beta+d_X} K_{\max}^2 N \\ &= 2L^2 K_{\max}^2 c_0^{2\beta+d_X} \ln N. \end{aligned}$$

The first and second line are proved in part (i). The fourth line use the fact that the support of functions $g_{kN}, k \in \mathcal{I}_N$ are disjoint and $\sum_{k\in\mathcal{I}_N} \mathbb{1}\left(\left|\frac{x_i - x_{kN}}{h_N}\right| \leq \frac{1}{2}\right) \leq 1$. We have $\ln M_N = \ln(m_N^{d_X}) \geq \frac{d_X}{2\beta+d_X}\ln\left(\frac{N}{\ln N}\right) - d_X \ln c_0 \geq \frac{d_X}{2\beta+d_X+1}\ln N$ for sufficiently large N. The condition is thus satisfied with sufficiently large c_0. The result follows from Theorem 2.5 of Tsybakov (2008).

Proof of Theorem 2.3.1

Applying the variance operator to $\hat{\Psi}(w)$ yields

$$\mathbb{V}\left(\hat{\Psi}(w)\right) = \frac{4}{N}\frac{N-2}{N-1}V_{N,1} + \binom{N}{2}^{-1} V_{N,2}$$

where, starting with the second term,

$$
\begin{aligned}
V_{N,2} &= \mathbb{V}\left(\frac{1}{2}\left[Y_{12}K_{12}+Y_{21}K_{21}\right]\right) \leq \mathbb{V}\left(Y_{12}K_{12}\right) \leq \mathbb{E}\left(Y_{12}^2K_{12}^2\right) \\
&= h_N^{-4d_X}\int \mathbb{E}\left[Y_{12}^2|(X_1,X_2)=(x_1,x_2)\right]K^2\left(\frac{x-x_1}{h_N},\frac{x-x_2}{h_N}\right)f(x_1)f(x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= h_N^{-2d_X}\int \mathbb{E}\left[Y_{12}^2|(X_1,X_2)=(x-h_Ns_1,x'-h_Ns_2)\right] \\
&\qquad\qquad \cdot f(x-h_Ns_1)f(x'-h_Ns_2)K^2\left(s_1,s_2\right)\mathrm{d}s_1\mathrm{d}s_2 \\
&\leq h_N^{-2d_X}B_4K_{\max}B_1.
\end{aligned}
$$

Next, consider the first term. We get that

$$
\begin{aligned}
V_{N,1} &= \mathbb{C}\left(\frac{1}{2}\left(Y_{12}K_{12}+Y_{21}K_{21}\right),\frac{1}{2}\left(Y_{13}K_{13}+Y_{31}K_{31}\right)\right) \\
&= \mathbb{V}\left(\mathbb{E}\left[\frac{1}{2}\left(Y_{12}K_{12}+Y_{21}K_{21}\right)\middle|X_1,U_1\right]\right) \\
&\leq \frac{1}{2}\operatorname{Var}\left(\mathbb{E}\left(Y_{12}K_{12}\middle|X_1,U_1\right)\right)+\frac{1}{2}\operatorname{Var}\left(\mathbb{E}\left(Y_{21}K_{21}\middle|X_1,U_1\right)\right) \\
&\leq \frac{1}{2}\mathbb{E}\left(Y_{12}K_{12}Y_{13}K_{13}\right)+\frac{1}{2}\mathbb{E}\left(Y_{21}K_{21}Y_{31}K_{31}\right) \\
&= \frac{1}{2}h_N^{-4d_X}\int \mathbb{E}\left(Y_{12}Y_{13}|(X_1,X_2,X_3)=(x_1,x_2,x_3)\right) \\
&\qquad\qquad \cdot K\left(\frac{x-x_1}{h_N},\frac{x'-x_2}{h_N}\right)K\left(\frac{x-x_1}{h_N},\frac{x'-x_3}{h_N}\right)f(x_1)f(x_2)f(x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \\
&\quad+\frac{1}{2}h_N^{-4d_X}\int \mathbb{E}\left(Y_{21}Y_{31}|(X_1,X_2,X_3)=(x_1,x_2,x_3)\right) \\
&\qquad\qquad \cdot K\left(\frac{x-x_2}{h_N},\frac{x'-x_1}{h_N}\right)K\left(\frac{x-x_3}{h_N},\frac{x'-x_1}{h_N}\right)f(x_1)f(x_2)f(x_3)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3 \\
&= h_N^{-d_X}\frac{1}{2}\int \mathbb{E}\left(Y_{12}Y_{13}|(X_1,X_2,X_3)=(x-h_Ns_1,x'-h_Ns_2,x'-h_Ns_3)\right) \\
&\qquad\qquad \cdot f(x-h_Ns_1)f(x'-h_Ns_2)f(x'-h_Ns_3)K\left(s_1,s_2\right)K\left(s_1,s_3\right)\mathrm{d}s_1\mathrm{d}s_2\mathrm{d}s_3 \\
&\quad+h_N^{-d_X}\frac{1}{2}\int \mathbb{E}\left(Y_{21}Y_{31}|(X_1,X_2,X_3)=(x'-h_Ns_1,x-h_Ns_2,x-h_Ns_3)\right) \\
&\qquad\qquad \cdot f(x'-h_Ns_1)f(x-h_Ns_2)f(x-h_Ns_3)K\left(s_1,s_2\right)K\left(s_1,s_3\right)\mathrm{d}s_1\mathrm{d}s_2\mathrm{d}s_3 \\
&\leq h_N^{-d_X}B_5\int \left|K\left(s_1,s_2\right)\right|\left|K\left(s_1,s_3\right)\right|\mathrm{d}s_1\mathrm{d}s_2\mathrm{d}s_3 \\
&\leq h_N^{-d_X}B_5B_2B_1. \qquad (2.7)
\end{aligned}
$$

These two bounds imply the variance bound

$$\begin{aligned}\mathbb{V}\left(\hat{\Psi}(w)\right) &\leq \binom{N}{2}^{-1} h_N^{-2d_X} B_4 K_{\max} B_1 + \frac{4(N-2)}{N(N-1)} h_N^{-d_X} B_5 B_2 B_1 \\ &= N^{-1} h_N^{-d_X}\left[\frac{N-2}{N-1} 4 B_5 B_2 B_1 + N^{-1} h_N^{-d_X} \frac{4N}{N-1} B_4 K_{\max} B_1\right],\end{aligned}$$

which, in turn, implies that for $M_0 = 4B_5B_2B_1 + 1$ and sufficiently large N, $\mathbb{V}\left(\hat{\Psi}(w)\right) \leq \frac{M_0}{Nh_N^{d_X}}$ for all $w \in \mathbb{R}^{d_W}$ as claimed.

Proof of Theorem 2.3.2

For τ_N a sequence of positive truncation parameters we consider the sum

$$\begin{aligned}\tilde{\Psi}_N(w) = \frac{1}{\binom{N}{2}} \sum_{1\leq i<j\leq N} \frac{1}{2}\Bigg[& Y_{ij} \cdot \mathbb{1}\left(|Y_{ij}| < \tau_N\right) \frac{1}{h_N^{d_W}} K\left(\frac{w - W_{ij}}{h_N}\right) \\ &+ Y_{ji} \cdot \mathbb{1}\left(|Y_{ji}| < \tau_N\right) \frac{1}{h_N^{d_W}} K\left(\frac{w - W_{ji}}{h_N}\right)\Bigg].\end{aligned}$$

We will use $\tilde{Z}_{N,ij}$ to denote the summands in the above expression in what follows. The Hoeffding decomposition of this U-like statistic is

$$\tilde{\Psi}(w) = \mathbb{E}\tilde{\Psi}(w) + \underbrace{\frac{2}{N}\sum_{i=1}^{N} \bar{Z}_{N,i}}_{T_{N,1}(w)} + \underbrace{\frac{1}{\binom{N}{2}} \sum_{1\leq i<j\leq N} \breve{Z}_{N,ij}}_{T_{N,2}(w)},$$

where

$$\begin{aligned}\bar{Z}_{N,i} &= \mathbb{E}\left[\tilde{Z}_{N,ij}\middle| X_i, U_i\right] - \mathbb{E}\tilde{Z}_{N,ij} \\ \breve{Z}_{N,ij} &= \tilde{Z}_{N,ij} - \mathbb{E}\left[\tilde{Z}_{N,ij}\middle| X_i, U_i\right] - \mathbb{E}\left[\tilde{Z}_{N,ij}\middle| X_j, U_j\right] + \mathbb{E}\tilde{Z}_{N,ij}.\end{aligned}$$

Notice that $T_{N,1}(w)$ is an average of N iid mean-zero random variables while $T_{N,2}(w)$ is a degenerate second-order U-like statistic.

To proceed further we require the following Lemma.

Lemma 2.4.1. *Under Assumptions 2.3.1 and 2.3.2, for any $\alpha > 0$, there exists constant M_α such that*

(i) if $\tau_N \ll a_N^{-1}$, then $\sup_{w\in\mathbb{R}^{d_W}} P\left(|T_{N,1}(w)| > M_\alpha a_N\right) = O\left(N^{-\alpha}\right)$;

(ii) if $\tau_N \ll Nh^{\frac{3}{2}d_X}/\ln N$ and $a_N = o(1)$, then $\sup_{w\in\mathbb{R}^{d_W}} P\left(|T_{N,2}(w)| > M_\alpha a_N\right) = O\left(N^{-\alpha}\right)$;

(iii) *if for some* $s > 1$, $\sup_{x_1,x_2\in\mathbb{R}^{d_X}} \mathbb{E}\left[|Y_{12}|^s \middle| (X_1,X_2) = (x_1,x_2)\right]\cdot f(x_1,x_2) \le B_{4,s} < \infty$ *and* $\tau_N \gg a_N^{-\frac{1}{s-1}}$, *then* $\sup_{w\in\mathbb{R}^{d_W}} \left|\mathbb{E}\left(\hat{\Phi}(w) - \tilde{\Phi}(w)\right)\right| = o\left(a_N\right)$*;*

(iv) *if for some* $s > 1$, $\mathbb{E}\left|Y_{12}\right|^s \le B_{6,s}$ *and* $\tau_N \gg (a_N h_N^{2d_X})^{-\frac{1}{s-1}}$, *then* $\sup_{w\in\mathbb{R}^{d_W}} \left|\hat{\Phi}_N(w) - \tilde{\Phi}_N(w)\right| = o_P\left(a_N\right)$*;*

(v) *if for some* $s > 2$, $\tau_N = (N^2\phi_N)^{\frac{1}{s}}$ *where* $\phi_N = (\ln(\ln N))^2 \ln N$, *and* $\mathbb{E}\left|Y_{12}\right|^s \le B_{6,s}$, *then* $P(\hat{\Phi}_N = \tilde{\Phi}_N) = P\left(\hat{\Phi}_N(w) = \tilde{\Phi}_N(w), \forall w \in \mathbb{R}^{2d_X}\right) \to 1$ *as* $N \to \infty$.

The proof of the above Lemma may be found below. The bandwidth conditions stated in the hypotheses of Theorem 2.3.2 ensure that we can pick truncation thresholds τ_N which satisfy the following conditions

1. $\tau_N \ll a_N^{-1}$;
2. $\tau_N \ll \frac{N}{\ln N} h_N^{\frac{3}{2}d_X}$;
3. $\tau_N \gg a_N^{-\frac{1}{s-1}}$;
4. $\tau_N \gg (N^2\phi_N)^{\frac{1}{s}}$ or $\tau_N \gg (a_N h_N^{2d_X})^{-\frac{1}{s-1}}$.

These conditions allow for the application of Lemma 2.4.1. Denote $R_N(w) := \hat{\Psi}_N(w) - \tilde{\Psi}_N(w)$. For any set $\mathcal{C}_N \subset \mathbb{R}^{2d}$,

$$\begin{aligned}
&P\left(\sup_{w\in\mathcal{C}_N}\left|\hat{\Psi}_N(w) - \mathbb{E}\hat{\Psi}_N(w)\right| > 8Ma_N\right)\\
&= P\left(\sup_{w\in\mathcal{C}_N}\left|\tilde{\Psi}_N(w) - \mathbb{E}\tilde{\Psi}_N(w) + R_N(w) - \mathbb{E}R_N(w)\right| > 8Ma_N\right)\\
&\le P\left(\sup_{w\in\mathcal{C}_N}\left|\tilde{\Psi}_N(w) - \mathbb{E}\tilde{\Psi}_N(w)\right| > 6Ma_N\right) + P\left(\sup_{w\in\mathcal{C}_N}\left|R_N(w) - \mathbb{E}R_N(w)\right| > 2Ma_N\right).
\end{aligned} \tag{2.8}$$

The second term in inequality (2.8) converges to zero because

$$\begin{aligned}
&P\left(\sup_{w\in\mathcal{C}_N}\left|R_N(w) - \mathbb{E}R_N(w)\right| > 2Ma_N\right)\\
&\le P\left(\sup_{w\in\mathbb{R}^{d_W}}\left|R_N(w) - \mathbb{E}R_N(w)\right| > 2Ma_N\right)\\
&\le P\left(\sup_{w\in\mathbb{R}^{d_W}}\left|R_N(w)\right| > Ma_N\right) + \mathbb{1}\left(\sup_{w\in\mathbb{R}^{d_W}}\left|\mathbb{E}R_N(w)\right| > Ma_N\right) \qquad (2.9)\\
&= o(1).
\end{aligned}$$

The last line holds because

$$\mathbb{1}\left(\sup_{w\in\mathbb{R}^{d_W}} |\mathbb{E}R_N(w)| > Ma_N\right) = 0 \qquad \text{for large } N \tag{2.10}$$

$$P\left(\sup_{w\in\mathbb{R}^{d_W}} |R_N(w)| > Ma_N\right) = o_P(1). \tag{2.11}$$

To see (2.10), notice part (iii) of Lemma 2.4.1 implies that $\sup_{w\in\mathbb{R}^{d_W}} |\mathbb{E}R_N(w)| = o(a_N)$. Hence $\mathbb{1}\left(\sup_{w\in\mathbb{R}^{2d}} |\mathbb{E}R_N(w)| > Ma_N\right) = 0$ for large N. To see (2.11), notice the inequality

$$P\left(\sup_{w\in\mathbb{R}^{d_W}} |R_N(w)| > Ma_N\right) \le \min\left\{1 - P(\hat{\Phi}_N = \tilde{\Phi}_N), \frac{\mathbb{E}\sup_{w\in\mathbb{R}^{d_W}} |R_N(w)|}{Ma_N}\right\},$$

suggests we can bound either term on the right-hand side to bound the term on the left-hand side. The threshold we pick meets the conditions of both parts (iv) and (v) of Lemma 2.4.1, which ensures either $1 - P(\hat{\Phi}_N = \tilde{\Phi}_N) = o(1)$ or $\frac{\mathbb{E}\sup_{w\in\mathbb{R}^{d_W}} |R_N(w)|}{Ma_N} = o(1)$. This implies (2.11).

To show the first term in inequality (2.8) converges to zero, we will use a covering argument to reduce finding the supremum over an infinite number points to finding the maximum over a finite number of points. We then invoke point-wise concentration bounds. This part closely follows the argument in Hansen (2008). Cover any compact region $\mathcal{C}_N \subset \mathbb{R}^{d_W}$ by finite number of balls of radius $a_N h_N$ centered at grid points in the set $L_N = \{w_{N,1}, w_{N,2}, \ldots, w_{N,L_N}\}$ (Here we abuse the notation a bit: L_N is used to refer to both the set and its cardinality). Denote the ball $A_{N,j} = \{w \in \mathbb{R}^{d_W} : ||w - w_{N,j}|| \le a_N h_N\}$. For N large enough such that $a_N < L$ (L is the constant appearing in Assumption 2.3.3), for any point $w \in A_{N,j}$ within the ball, assumption 2.3.3 (ii) implies

$$\left|K\left(\frac{w - W_{ij}}{h}\right) - K\left(\frac{w_{N,j} - W_{ij}}{h}\right)\right| \le a_N K^*\left(\frac{w_{N,j} - W_{ij}}{h}\right). \tag{2.12}$$

Define

$$\breve{\Phi}_N(w) := \frac{1}{N(N-1)} \sum_{1\le i\neq j\le N} Y_{ij} \cdot \mathbb{1}\left(|Y_{ij}| < \tau_N\right) \frac{1}{h^{d_W}} K^*\left(\frac{w - W_{ij}}{h}\right),$$

which is a version of $\tilde{\Phi}(w)$ with K replaced by K^*. The bound (2.12) implies

$$\left|\tilde{\Psi}_N(w) - \tilde{\Psi}_N(w_{N,j})\right| \le a_N \breve{\Phi}_N(w_{N,j}),$$

with $|\mathbb{E}\breve{\Phi}_N(w_{N,j})| \leq B_4^{1/2}B_3^{1/2}\int|K^*(w)|\mathrm{d}w < \infty$. Next bound the sup within the ball by a value at the center and the sup discrepancy

$$
\begin{aligned}
&\sup_{w\in A_{N,j}}\left|\tilde{\Psi}_N(w)-\mathbb{E}\tilde{\Psi}_N(w)\right| \\
&\leq \left|\tilde{\Psi}_N(w_{N,j})-\mathbb{E}\tilde{\Psi}_N(w_{N,j})\right| + \sup_{w\in A_{N,j}}\left|\tilde{\Psi}_N(w)-\tilde{\Psi}_N(w_{N,j})\right| + \sup_{w\in A_{N,j}}\left|\mathbb{E}\left(\tilde{\Psi}_N(w)-\tilde{\Psi}_N(w_{N,j})\right)\right| \\
&\leq \left|\tilde{\Psi}_N(w_{N,j})-\mathbb{E}\tilde{\Psi}_N(w_{N,j})\right| + a_N\left[\breve{\Phi}_N(w_{N,j})+\mathbb{E}\breve{\Phi}_N(w_{N,j})\right] \\
&\leq \left|\tilde{\Psi}_N(w_{N,j})-\mathbb{E}\tilde{\Psi}_N(w_{N,j})\right| + a_N\left|\breve{\Phi}_N(w_{N,j})-\mathbb{E}\breve{\Phi}_N(w_{N,j})\right| + 2a_N\mathbb{E}\breve{\Phi}_N(w_{N,j}) \\
&\leq \left|\tilde{\Psi}_N(w_{N,j})-\mathbb{E}\tilde{\Psi}_N(w_{N,j})\right| + \left|\breve{\Phi}_N(w_{N,j})-\mathbb{E}\breve{\Phi}_N(w_{N,j})\right| + 2a_N\mathbb{E}\breve{\Phi}_N(w_{N,j}).
\end{aligned}
$$

The last inequality follow because $a_N \leq 1$ for N large enough. For any constant $M \geq B_4^{1/2}B_3^{1/2}\int|K^*(w)|\mathrm{d}w \geq \mathbb{E}\breve{\Phi}_N(w_{N,j})$,

$$
\begin{aligned}
&P\left(\sup_{w\in A_{N,j}}\left|\tilde{\Psi}_N(w)-\mathbb{E}\tilde{\Psi}_N(w)\right| > 6Ma_N\right) \\
&\leq P\left(\left|\tilde{\Psi}_N(w_{N,j})-\mathbb{E}\tilde{\Psi}_N(w_{N,j})\right| + \left|\breve{\Phi}_N(w)-\mathbb{E}\breve{\Phi}_N(w)\right| + 2a_N\mathbb{E}\breve{\Phi}_N(w) > 6Ma_N\right) \\
&\leq P\left(\left|\tilde{\Psi}_N(w_{N,j})-\mathbb{E}\tilde{\Psi}_N(w_{N,j})\right| > 2Ma_N\right) + P\left(\left|\breve{\Phi}_N(w)-\mathbb{E}\breve{\Phi}_N(w)\right| > 2Ma_N\right),
\end{aligned}
$$

as well as

$$
\begin{aligned}
&P\left(\sup_{w\in\mathcal{C}_N}\left|\tilde{\Psi}_N(w)-\mathbb{E}\tilde{\Psi}_N(w)\right| > 6Ma_N\right) \\
&\leq \sum_{j=1}^{L_N} P\left(\sup_{w\in A_{N,j}}\left|\tilde{\Psi}_N(w)-\mathbb{E}\tilde{\Psi}_N(w)\right| > 6Ma_N\right) \\
&\leq L_N \max_{j\in\{1,2,\ldots,L_N\}} P\left(\sup_{w\in A_{N,j}}\left|\tilde{\Psi}_N(w)-\mathbb{E}\tilde{\Psi}_N(w)\right| > 6Ma_N\right) \\
&\leq L_N \max_{j\in\{1,2,\ldots,L_N\}} P\left(\left|\tilde{\Psi}_N(w_{N,j})-\mathbb{E}\tilde{\Psi}_N(w_{N,j})\right| > 2Ma_N\right) \\
&\qquad + L_N \max_{j\in\{1,2,\ldots,L_N\}} P\left(\left|\breve{\Phi}_N(w)-\mathbb{E}\breve{\Phi}_N(w)\right| > 2Ma_N\right). \qquad (2.13)
\end{aligned}
$$

We now bound the two terms in (2.13) using the same argument, as both K and K^* satisfy Assumption 2.3.1, and this is the only property of the function K or K^* we will use. For any $\alpha > 0$ and M_α as in Lemma 2.4.1, for any $w \in \mathbb{R}^{d_W}$

$$\begin{aligned}\sup_{w\in\mathbb{R}^{d_W}} P\left(\left|\tilde{\Psi}_N(w) - \mathbb{E}\tilde{\Psi}_N(w)\right| > 2M_\alpha a_N\right) &= \sup_{w\in\mathbb{R}^{d_W}} P\left(|T_{N,1}(w) + T_{N,2}(w)| > 2M_\alpha a_N\right) \\ &\leq \sup_{w\in\mathbb{R}^{d_W}} P\left(|T_{N,1}(w)| > M_\alpha a_N\right) \\ &\quad + \sup_{w\in\mathbb{R}^{d_W}} P\left(|T_{N,2}(w)| > M_\alpha a_N\right) \\ &= O\left(N^{-\alpha}\right).\end{aligned}$$

Hence

$$P\left(\sup_{w\in\mathcal{C}_N}\left|\tilde{\Psi}_N(w) - \mathbb{E}\tilde{\Psi}_N(w)\right| > 6Ma_N\right) \leq O\left(L_N N^{-\alpha}\right).$$

If we take $\mathcal{C}_N = \{w \in \mathbb{R}^{d_W} : ||w|| < c_N\}$ where $c_N = N^q$, then $\mathcal{C}_N$ can be covered by $L_N = 2\left(\frac{c_N}{a_N h_N}\right)^{d_W}$ number of balls with radius $a_N h_N$. Hence we can take α large enough, e.g. $\alpha = (q + \frac{1}{2})d_W + 3$, so that $O\left(L_N N^{-\alpha}\right) = O\left(\left(\frac{c_N}{a_N h_N}\right)^{d_W} N^{-\alpha}\right) = O\left(N^{(q+\frac{1}{2})d_W + 2 - \alpha}\right) = O\left(N^{-1}\right) = o(1)$. We have therefore shown that

$$P\left(\sup_{w\in\mathcal{C}_N}\left|\tilde{\Psi}_N(w) - \mathbb{E}\tilde{\Psi}_N(w)\right| > 6Ma_N\right) = o(1). \tag{2.14}$$

Together the two bounds (2.9) and (2.14) imply that the right-hand side of equality (2.8) is $o(1)$. This is saying for sufficiently large $M < \infty$, we have $P\left(\sup_{w\in\mathcal{C}_N}\left|\hat{\Psi}_N(w) - \mathbb{E}\hat{\Psi}_N(w)\right| > 8Ma_N\right) = o(1)$, which is sufficient for

$$\sup_{||w||\leq c_N}\left|\hat{\Psi}_N(w) - \mathbb{E}\hat{\Psi}_N(w)\right| = O(a_N).$$

as required.

Proof of Lemma 2.4.1

Proof of claim (i)

To prove the first claim of the Lemma we will apply the classic Bernstein's inequality (see equation 2.10 on p. 36 of the textbook Boucheron et al. (2013)).

> Let $Q_1, \ldots, Q_N$ be independent random variables with finite variance such that $Q_i \leq b$ for some $b > 0$ almost surely for all $i < N$. Let $S = \sum_{i=1}^N (Q_i - \mathbb{E}Q_i)$ and $v = \sum_{i=1}^N \mathbb{E}\left[Q_i^2\right]$. Then for any $t > 0$,
>
> $$P(S \geq t) \leq \exp\left(-\frac{t^2}{2(v + bt/3)}\right).$$

In order to invoke the inequality, we first show that $Q_i(w) := \tau_N^{-1} h_N^{d_X} \bar{Z}_{N,i}(w)$ is bounded. In the following we will use the abbreviated notation Q_i for $Q_i(w)$. Remember $\bar{Z}_{N,i}$ is the mean-normalized version of $\mathbb{E}\left[\tilde{Z}_{N,ij}\middle|X_i, U_i\right]$. Since

$$\begin{aligned}
&\tau_N^{-1} h^{d_X} \left|\mathbb{E}\left[\tilde{Z}_{N,ij}\middle|X_i, U_i\right]\right| \\
&= \tau_N^{-1} h^{d_X} \left|\mathbb{E}\left[\frac{1}{2}\left[Y_{ij} \cdot \mathbb{1}\left(|Y_{ij}| < \tau_N\right) K_{ij}\right.\right.\right. \\
&\left.\left.\left. + Y_{ji} \cdot \mathbb{1}\left(|Y_{ji}| < \tau_N\right) K_{ji}\right] \middle| X_i, U_i\right]\right| \\
&\leq h_N^{d_X} \frac{1}{2} \mathbb{E}\left[|K_{ij}| + |K_{ji}| \middle| X_i, U_i\right] \\
&= h_N^{-d_X} \frac{1}{2} \mathbb{E}\left[\left|K\left(\frac{w - W_{ij}}{h_N}\right)\right| + \left|K\left(\frac{w - W_{ji}}{h_N}\right)\right| \middle| X_i\right] \\
&= h_N^{-d_X} \frac{1}{2} \int \left[\left|K\left(\frac{x - x_i}{h_N}, \frac{x' - x_j}{h_N}\right)\right| + \left|K\left(\frac{x - x_j}{h_N}, \frac{x' - x_i}{h_N}\right)\right|\right] f(x_j) \mathrm{d}x_j \\
&= \frac{1}{2} \int \left|K\left(\frac{x - x_i}{h_N}, s\right)\right| f(x' - h_N s) + \left|K\left(s, \frac{x' - x_i}{h_N}\right)\right| f(x - h_N s) \mathrm{d}s \\
&\leq B_2 B_3,
\end{aligned}$$

we have $|Q_i| = |\tau_N^{-1} h_N^{d_X} \bar{Z}_{N,i}| < 2B_2 B_3$. Write $P\left(T_{N,1}(w) > M a_N\right)$ in the form suitable for applying Bernstein's inequality

$$\begin{aligned}
P\left(T_{N,1}(w) > M a_N\right) &= P\left(\frac{2}{N} \sum_{i=1}^{N} \bar{Z}_{N,i} > M a_N\right) \\
&= P\left(\sum_{i=1}^{N} \tau_N^{-1} h_N^{d_X} \bar{Z}_{N,i} > \frac{M}{2} N h_N^{d_X} a_N \tau_N^{-1}\right) \\
&= P(S \geq t),
\end{aligned}$$

in which $S = \sum_{i=1}^{N} Q_i$ and $t = \frac{M}{2} N h_N^{d_X} a_N \tau_N^{-1}$. Applying Bernstein's inequality gives us $P(S \geq t) \leq \exp\left(-\frac{t^2}{2(v + bt/3)}\right)$ where $v := \sum_{i=1}^{N} \mathbb{E}\left[Q_i^2\right]$ and $b = 2B_2 B_3$. Since the function $\exp\left(-\frac{t^2}{2(v+bt/3)}\right)$ is increasing in v, we have for any $v' > v$

$$P(S \geq t) \leq \exp\left(-\frac{t^2}{2(v' + bt/3)}\right). \tag{2.15}$$

The upper bound v' we are going to use is the following one

$$v = \sum_{i=1}^{N} \mathbb{E}\left[Q_i^2\right] = \sum_{i=1}^{N} \mathbb{E}\left[\left(\tau_N^{-1} h_N^{d_X} \bar{Z}_{N,i}\right)^2\right] = \tau_N^{-2} h_N^{2d_X} N V_{N,1} \leq \tau_N^{-2} N h_N^{d_X} B_5 B_2 B_1 := v',$$

in which the inequality is an implication of (2.7). Plugging the expression of v', t, b, and a_N into the RHS of (2.15) gives us

$$\exp\left(-\frac{t^2}{2(v'+bt/3)}\right) = \exp\left(-\frac{M^2}{8B_5B_2B_1 + 8B_2B_3Ma_N\tau_N/3}\ln N\right).$$

By assumption $a_N\tau_N \to 0$ as $N \to \infty$, we can pick N_0 such that $8B_2B_3a_N\tau_N/3 \leq 1$ for any $N > N_0$. For any $\alpha > 0$, we can pick M large enough so that $\frac{M^2}{8B_5B_2B_1+M} \geq \alpha$ and $\exp\left(-\frac{M^2}{8B_5B_2B_1+M}\ln N\right) < N^{\alpha}$. In particular, $M_\alpha = \frac{\alpha+\sqrt{\alpha^2+32B_5B_2B_1\alpha}}{2}$ will work. This means we have proved

$$P\left(T_{N,1}(w) > M_\alpha a_N\right) = O\left(N^{-\alpha}\right).$$

We get the two-sided bound by applying the same argument twice for $T_{N,1}(w)$ and $-T_{N,1}(w)$. Moreover, because the derivation of the bound and the value of M_α doesn't depend on the specific point w, we have also proved our desired result

$$\sup_{w\in\mathbb{R}^{d_W}} P\left(|T_{N,1}(w)| > M_\alpha a_N\right) = O\left(N^{-\alpha}\right).$$

Proof of claim (ii)

We will use Propsition 2.3(c), a concentration inequality, from Arcones and Gine (1993)[3] to prove the second claim.

> Let $\{X_i, i\in\mathbb{N}\}$ and $\{V_{i_1,\ldots,i_m}, (i_1,\ldots,i_m)\in I_m^{\mathbb{N}}\}$ be independent random samples; $||f||_\infty \leq c$, $\mathbb{E}\,]f(X_1,\ldots,X_m,V_{1,\ldots,m})] = 0$, $\sigma^2 = \mathbb{E}\left[f^2(X_1,\ldots,X_m,V_{1,\ldots,m})\right]$; f is P-canonical, then there are constants c_i depending only on m such that for any $t > 0$,
>
> $$P\left(\left|N^{-m/2}\sum_{(i_1,\ldots,i_m)\in I_m^N} f(X_{i_1},\ldots,X_{i_m},V_{i_1,\ldots,i_m})\right| > t\right) \leq c_1\exp\left(-\frac{c_2t^{2/m}}{\sigma^{2/m}+\left(ct^{1/m}N^{-1/2}\right)^{2/(m+1)}}\right).$$

In order to apply the inequality, we first show that $\tau_N^{-1}h_N^{2d_X}\breve{Z}_{N,ij}$ is bounded. Decompose

$$\breve{Z}_{N,ij} = \tilde{Z}_{N,ij} - \mathbb{E}\left[\tilde{Z}_{N,ij}\middle|X_i,U_i\right] - \mathbb{E}\left[\tilde{Z}_{N,ij}\middle|X_j,U_j\right] + \mathbb{E}\tilde{Z}_{N,ij}.$$

[3] There is a small modification compared to the original proposition. Since our statistic is not exactly a U-statistic as there are the iid V_{ij} variables in our setup, we include this additional term in the statement of inequality. The proof of the inequality in our setup could follow the same steps of the original Arcones and Gine (1993) one. The reason this works is that the V_{ij} terms are iid and won't affect the randomization inequality, decoupling inequality, and the hypercontractivity inequality used in the proof.

The last three terms on the right-hand side are bounded because $\tau_N^{-1}\mathbb{E}\tilde{Z}_{N,ij} = O(1)$ and $\tau_N^{-1}\mathbb{E}\left[\tilde{Z}_{N,ij}\middle|X_i,U_i\right] = O(h_N^{-d_X})$. Moreover, $|\tau_N^{-1}h_N^{2d_X}\tilde{Z}_{N,ij}| = \frac{1}{2}|\tau_N^{-1}Y_{ij}\cdot \mathbb{1}\left(|Y_{ij}|<\tau_N\right)K\left(\frac{w-W_{ij}}{h_N}\right)|+\frac{1}{2}|\tau_N^{-1}Y_{ji}\cdot\mathbb{1}\left(|Y_{ji}|<\tau_N\right)K\left(\frac{w-W_{ji}}{h_N}\right)| \le K_{\max}$. Hence, there exists constant $c>0$ s.t. $|\tau_N^{-1}h_N^{2d_X}\breve{Z}_{N,ij}|<c$. Applying the concentration inequality to $T_{N,2}(w)$ then gives us

$$\begin{aligned}
P\left(|T_{N,2}(w)|>Ma_N\right) &= P\left(\left|\frac{1}{\binom{N}{2}}\sum_{1\le i<j\le N}\breve{Z}_{N,ij}\right|>Ma_N\right)\\
&= P\left(\left|N^{-1}\sum_{1\le i<j\le N}\tau_N^{-1}h_N^{2d_X}\breve{Z}_{N,ij}\right|>M\frac{N-1}{2}h_N^{2d_X}a_N\tau_N^{-1}\right)\\
&\le c_1\exp\left(-\frac{c_2t}{\sigma+\left(ct^{1/2}N^{-1/2}\right)^{2/3}}\right)\\
&= c_1\exp\left(-\frac{c_2t}{\sigma\ln N+\left(ct^{1/2}N^{-1/2}\right)^{2/3}\ln N}\cdot\ln N\right)
\end{aligned}$$

where $t = M\frac{N-1}{2}h_N^{2d_X}a_N\tau_N^{-1}$ and $\sigma^2 = \operatorname{Var}\left(\tau_N^{-1}h_N^{2d_X}\breve{Z}_{N,ij}\right)$. We will show that $\frac{c_2t}{\sigma\ln N+\left(ct^{1/2}N^{-1/2}\right)^{2/3}\ln N}\to\infty$ as $N\to\infty$ by showing both $\frac{t}{\sigma\ln N}\to\infty$ and $\frac{t}{\left(ct^{1/2}N^{-1/2}\right)^{2/3}\ln N}\to\infty$ as $N\to\infty$.

Beginning with the former claim:

$$\begin{aligned}
\frac{t}{\sigma\ln N} &= \frac{M\frac{N-1}{2}h_N^{2d_X}a_N\tau_N^{-1}}{\tau_N^{-1}h_N^{2d_X}\operatorname{Var}\left(\breve{Z}_{N,ij}\right)^{1/2}\ln N} = \frac{M(N-1)a_N}{2\operatorname{Var}\left(\breve{Z}_{N,ij}\right)^{1/2}\ln N} \ge \frac{M(N-1)a_N}{2V_{N,2}^{1/2}\ln N}\\
&\ge \frac{MNa_N}{4\left(h_N^{-2d_X}B_4K_{max}B_1\right)^{1/2}\ln N} = \frac{M}{4\left(B_4K_{max}B_1\right)^{1/2}}a_N\left(\frac{\ln N}{Nh_N^{d_X}}\right)^{-1}\\
&= \frac{M}{4\left(B_4K_{max}B_1\right)^{1/2}}a_N^{-1}\\
&\to\infty,\ \text{ as } N\to\infty.
\end{aligned}$$

The latter claim follows because:

$$\begin{aligned}
\frac{t}{\left(ct^{1/2}N^{-1/2}\right)^{2/3}\ln N} &= \left(\frac{t^2 N}{c^2(\ln N)^3}\right)^{1/3} = \left(\frac{(M\frac{N-1}{2}h_N^{2d_X}a_N\tau_N^{-1})^2 N}{c^2(\ln N)^3}\right)^{1/3} \\
&\geq \left(\frac{M^2}{16c^2}N^3(\ln N)^{-3}h_N^{4d_X}a_N^2\tau_N^{-2}\right)^{1/3} \\
&= \left(\frac{M^2}{16c^2}\right)^{1/3}\left(Nh_N^{\frac{3}{2}d_X}(\ln N)^{-1}\tau_N^{-1}\right)^{2/3} \\
&\to \infty, \text{ as } N \to \infty.
\end{aligned}$$

The last line above is an implication of the condition $\tau_N \ll Nh_N^{\frac{3}{2}d_X}/\ln N$. Combining these two limit results gives us $\frac{c_2 t}{\sigma \ln N + \left(ct^{1/2}N^{-1/2}\right)^{2/3}\ln N} \to \infty$ as $N \to \infty$. Notice the bound again doesn't depend on w and the inequality still holds when we take the sup over $w \in \mathbb{R}^{d_W}$ on the left-hand side. Hence for any $M > 0$ and any $\alpha > 0$, $\sup_{w\in\mathbb{R}^{d_W}} P\left(|T_{N,2}(w)| > Ma_N\right) = O\left(N^{-\alpha}\right)$.

Proof of claim (iii)

Direct evaluation yields

$$\begin{aligned}
\left|\mathbb{E}\left(\hat{\Phi}(w) - \tilde{\Phi}(w)\right)\right| &= \left|\mathbb{E}\left[Y_{ij}\mathbb{1}\left(|Y_{ij}| > \tau_N\right)\frac{1}{h_N^{2d_X}}K\left(\frac{w - W_{ij}}{h}\right)\right]\right| \\
&\leq \mathbb{E}\left[|Y_{ij}|\left|\tau_N^{-1}Y_{ij}\right|^{s-1}\mathbb{1}\left(|Y_{ij}| > \tau_N\right)\frac{1}{h_N^{2d_X}}\left|K\left(\frac{w - W_{ij}}{h_N}\right)\right|\right] \\
&\leq \tau_N^{-(s-1)}\mathbb{E}\left[|Y_{ij}|^s\frac{1}{h_N^{2d_X}}\left|K\left(\frac{w - W_{ij}}{h}\right)\right|\right] \\
&= \tau_N^{-(s-1)}\int \mathbb{E}\left[|Y_{12}|^s\,|(X_1, X_2) = (x_1, x_2)\right]\frac{1}{h_N^{2d_X}} \\
&\quad \left|K\left(\frac{x - x_1}{h_N}, \frac{x - x_2}{h_N}\right)\right| f(x_1, x_2)\mathrm{d}x_1\mathrm{d}x_2 \\
&= \tau_N^{-(s-1)}\int \mathbb{E}\left[|Y_{12}|^s\,|(X_1, X_2) = (x - h_N s_1, x' - h_N s_2)\right] \\
&\quad \times f(x - h_N s_1, x' - h_N s_2)\left|K\left(s_1, s_2\right)\right|\mathrm{d}s_1\mathrm{d}s_2 \\
&\leq \tau_N^{-(s-1)}B_{4,s}B_1.
\end{aligned}$$

Since the last expression doesn't depend on w, we have $\sup_{w\in\mathbb{R}^{d_W}}\left|\mathbb{E}\left(\hat{\Phi}(w) - \tilde{\Phi}(w)\right)\right| = o(a_N)$.

Proof of claim (iv)

First, we eliminate the sup by upper bounding the terms involving K by $K_{\max}$.

$$\begin{aligned}\sup_{w\in\mathbb{R}^{d_W}}\left|\hat{\Phi}_N(w)-\tilde{\Phi}_N(w)\right| &= \sup_{w\in\mathbb{R}^{d_W}}\left|\frac{1}{N(N-1)}\sum_{1\le i\neq j\le N}Y_{ij}\mathbb{1}\left(|Y_{ij}|>\tau_N\right)\frac{1}{h_N^{d_W}}K\left(\frac{w-W_{ij}}{h_N}\right)\right| \\ &\le \frac{1}{N(N-1)}\sum_{1\le i\neq j\le N}|Y_{ij}|\,\mathbb{1}\left(|Y_{ij}|>\tau_N\right)\frac{1}{h_N^{2d_X}}\sup_{w\in\mathbb{R}^{d_W}}\left|K\left(\frac{w-W_{ij}}{h_N}\right)\right| \\ &\le K_{\max}h_N^{-2d_X}\tau_N^{-(s-1)}\frac{1}{N(N-1)}\sum_{1\le i\neq j\le N}|Y_{ij}|^s.\end{aligned}$$

Then, taking expectation on both sides yields

$$\mathbb{E}\left(\sup_{w\in\mathbb{R}^{d_W}}\left|\hat{\Phi}_N(w)-\tilde{\Phi}_N(w)\right|\right)\le K_{\max}h_N^{-2d_X}\tau_N^{-(s-1)}\mathbb{E}\left(|Y_{ij}|^s\right)\le K_{\max}B_{6,s}h_N^{-2d_X}\tau_N^{-(s-1)}=o(a_N).$$

Proof of claim (v)

If all the $|Y_{ij}|, 1\le i\neq j\le N$ are smaller than the truncation threshold τ_N, then $\hat{\Phi}_N=\tilde{\Phi}_N$,

$$P\left(\hat{\Phi}_N=\tilde{\Phi}_N\right)\ge P\left(\max_{1\le i<j\le N}|Y_{ij}|\le\tau_N\right).$$

We now show that the RHS converges to 1. Observe

$$\begin{aligned}\sum_{N=2}^{\infty}\sum_{i=1}^{N-1}\left[P\left(|Y_{iN}|>\tau_N\right)+P\left(|Y_{Ni}|>\tau_N\right)\right] &\le \sum_{N=2}^{\infty}\sum_{i=1}^{N-1}\left[\mathbb{E}\left(|Y_{iN}|^s\tau_N^{-s}\right)+\mathbb{E}\left(|Y_{Ni}|^s\tau_N^{-s}\right)\right] \\ &= \mathbb{E}\left(|Y_{iN}|^s\right)\sum_{N=2}^{\infty}\sum_{i=1}^{N-1}2N^{-2}\phi_N^{-1} \\ &\le \mathbb{E}\left(|Y_{iN}|^s\right)\sum_{N=2}^{\infty}\frac{2}{N(\ln\ln N)^2\ln N} \\ &< \infty,\end{aligned}$$

The Borel-Cantelli lemma implies $P(A_{ij}, i\neq j, i.o.)=0$ where the set $A_{ij}=\{\omega: Y_{ij}(\omega)>\tau_{\max\{i,j\}}\}$. This means, except for a null set $\mathcal{N}$, for any $\omega\in\mathcal{N}^c$, there exists a $N(\omega)$ s.t. for all $N\ge N(\omega)$, $Y_{iN}(\omega)\le\tau_N$. Since $\tau_N\uparrow\infty$ as $N\to\infty$, we can take $N^*(\omega)\ge N(\omega)$ such that $\tau_{N^*(w)}>\max_{i,j\le N(\omega)}|Y_{ij}(\omega)|$. Then for any $N\ge N^*(\omega)$, we have $\max_{1\le i<j\le N}|Y_{ij}(\omega)|\le\tau_N$ and hence $\hat{\Phi}_N=\tilde{\Phi}_N$. Define the set $E_N:=\{\omega: N^*(\omega)\le N\}\subset\{\omega:\hat{\Phi}_N=\tilde{\Phi}_N\}$. Since $E_N\uparrow\mathcal{N}^c$ and $P(\mathcal{N}^c)=1$, we have $P(\hat{\Phi}_N=\tilde{\Phi}_N)\ge P(E_N)\to 1$ as $N\to\infty$.

Proof of Theorem 2.3.3

The proof follows the general approach used in Hansen (2008). Denote $\hat{f}_{W,N}(w) = \frac{1}{N(N-1)} \sum_{1\leq i\neq j\leq N} K_{ij,N}(w)$. We can write

$$\hat{g}_N(w) = \frac{\hat{\Psi}_N(w)}{\hat{f}_{W,N}(w)}.$$

We examine the numerator and denominator separately. An application of Theorem 2.3.2 yields

$$\sup_{||w||\leq C_N} |\hat{\Psi}_N(w) - \mathbb{E}\hat{\Psi}_N(w)| = O_p(a_N)$$
$$\sup_{||w||\leq C_N} |\hat{f}_{W,N}(w) - \mathbb{E}\hat{f}_{W,N}(w)| = O_p(a_N).$$

Standard bias calculations give

$$\sup_{||w||\leq C_N} |\mathbb{E}\hat{\Psi}_N(w) - \Psi(w)| = O(h_N^{\beta})$$
$$\sup_{||w||\leq C_N} |\mathbb{E}\hat{f}_{W,N}(w) - f_W(w)| = O(h_N^{\beta}).$$

Combining these results we get

$$\sup_{||w||\leq C_N} |\hat{\Psi}_N(w) - \Psi(w)| = O_p(a_N) + O(h_N^{\beta}) = O(a_N^*)$$
$$\sup_{||w||\leq C_N} |\hat{f}_{W,N}(w) - f_W(w)| = O_p(a_N) + O(h_N^{\beta}) = O(a_N^*).$$

Uniformly over $||w|| \leq C_N$ we have

$$\begin{aligned}\frac{\hat{\Psi}_N(w)}{\hat{f}_{W,N}(w)} &= \frac{\hat{\Psi}_N(w)/f_W(w)}{\hat{f}_{W,N}(w)/f_W(w)} = \frac{g(w) + (\hat{\Psi}_N(w) - \Psi(w))/f_W(w)}{1 + (\hat{f}_{W,N}(w) - f_W(w))/f_W(w)} = \frac{g(w) + O_p(\delta_N^{-1}a_N^*)}{1 + O_p(\delta_N^{-1}a_N^*)} \\ &= g(w) + O_p(\delta_N^{-1}a_N^*)\end{aligned}$$

as claimed. The optimal rate is obtained by setting $h_N \asymp \left(\frac{\ln N}{N}\right)^{\frac{1}{2\beta+d_X}}$.

Chapter 3

Density-Weighted Average Derivatives for Dyadic Data

3.1 Introduction and Summary

In this chapter, we study estimation of the density-weighted average derivative for directed dyadic data. This parameter is of substantial practical interest as it is proportional to the coefficients in single index models (Powell et al., 1989), which encompasses various models of limited dependent variables. The main contributions of this chapter are extending the classical kernel-based estimator of the density-weighted average derivatives from the "monadic" iid setup (e.g. Stoker, 1986; Powell et al., 1989; Newey and Stoker, 1993) to directed dyadic data, and proving its robust asymptotic normality (in the sense asymptotic normality holds under both nondegeneracy and degeneracy and across a wide range of bandwidth sequences) using asymptotic quadratic approximation. This robust asymptotic normality result presents an interesting contrast between this kernel-based semiparametric estimator and the sample mean of dyadic data, which exhibits asymptotic non-normality when dyadic dependence is absent and whose uniform nonconservative inference procedure doesn't exist (Menzel, 2021).

Compared to the previous two chapters, this chapter studies a richer setup for directed dyadic data in the following sense. Instead of only considering monadic regressors as in chapter 2, we consider both monadic and dyadic regressors in this chapter. To appreciate this, consider a model of export/import among countries. Natural explainable variables include both GDP of each country, a monadic variable, and export/import tax between pairs of countries, a dyadic variable. As we will see in the results, the rate of convergence of estimators may differ depending on whether we are estimating the coefficient of a monadic variable or a dyadic variable. A similar phenomenon also shows up in Chapter 4 (for different reasons though).

3.2 Population and Sampling Framework

We consider an empirical problem where $i \in \mathbb{N}$ indices agents in an infinite population of interest. Agents are also referred as nodes, individuals, or monads. A (directed) pair of agent (i, j) constitutes a *dyad*, in which we refer to agent i as "ego" and agent j as "alter".

The data generating process specifies the joint distribution of the dyadic outcome Y_{ij} for dyad (i, j) together with observed monadic ego regressors B_i, monadic alter regressors C_j, and dyadic regressors D_{ij}.

In a toy specification of the gravity equation in international trade, Y_{ij} is the logarithm of export from country i to country j. B_i, C_j are the logarithm of the values of exporter's output and the value of importer's expenditure, respectively. D_{ij} is the logarithm of the distance from country i to j.

We are going to use a latent space model satisfying the following key modeling assumptions. First, the regressors $X_{ij} := (B_i, C_j, D_{ij})$ are jointly exchangeable. Namely,

$$[X_{ij}] \overset{d}{=} [X_{\sigma(i)\sigma(j)}]$$

for every permutation $\pi \in \Pi$ where $\pi : \{1, 2, \ldots, N\} \to \{1, 2, \ldots, N\}$ is a permutation of the nodes indices. In other words, the joint likelihood of regressors is invariant to the node labeling. Second, the outcome are relatively exchangeable given the regressors. Namely, the conditional distribution of $\mathbf{Y}$ is invariant against permutation of indices $\sigma_X : \mathbb{N} \to \mathbb{N}$ satisfying the restriction $[X_{\sigma_X(i)\sigma_X(j)}] \overset{d}{=} [X_{ij}]$:

$$[Y_{ij}] \overset{d}{=} [Y_{\sigma_X(i)\sigma_X(j)}].$$

Put differently, the conditional likelihood of outcomes given the regressor values does not depend on the node labeling. Third, outcomes of dyads sharing zero index are independent of each other. Outcomes of dyads sharing one index, like Y_{ij} and Y_{il}, are allowed to be correlated with each other even conditional on the observed covariates. This dyadic dependence is crucial for statistical analysis. For many statistical procedures, the dyadic dependence renders the effective sample size to be the number of nodes instead of the much larger number of dyads.

We state the sampling assumptions of all relevant latent variables first. Then we construct the observed variables using the latent variables. Let $\{(A_i, B_i, C_i)\}_{i\geq 1}$, $\{(U_i)\}_{i\geq 1}$, $\{(V_{ij}, V_{ji})\}_{i,j\geq 1, i<j}$, $\{(\epsilon_{ij}, \epsilon_{ji})\}_{i,j\geq 1, i<j}$ be sequences of i.i.d. random variables additionally independent of one another. (V_{ij}, V_{ji}) and $(\epsilon_{ij}, \epsilon_{ji})$ are both symmetric, namely $(V_{ij}, V_{ji}) \overset{d}{=} (V_{ji}, V_{ij})$ and $(\epsilon_{ij}, \epsilon_{ji}) \overset{d}{=} (\epsilon_{ji}, \epsilon_{ij})$. We explain specific roles of these variables in the following.

For each agent i, we observe the monadic random variables (B_i, C_i) in which $B_i \in \mathbb{B} \subset \mathbb{R}^{d_B}$ is ego-relevant and $C_i \in \mathbb{C} \subset \mathbb{R}^{d_C}$ is alter-relevant. For each dyad (i, j), we observe dyadic explanatory variables $D_{ij} \in \mathbb{D} \subset \mathbb{R}^{d_D}$ generated according to

$$D_{ij} = D\left(A_i, A_j, \epsilon_{ij}\right),$$

where A_i is an agent-specific random vector of attributes (of arbitrary dimension, not necessarily observable) and ϵ_{ij} is an unobservable random vector. The scalar dyadic outcome variable $Y_{ij} \in \mathbb{Y} \subset \mathbb{R}$ is generated according to

$$Y_{ij} = Y\left(B_i, C_j, D_{ij}, U_i, U_j, V_{ij}\right),$$

where U_i is an unobservable scalar random variable uniformly distributed on $[0, 1]$ and V_{ij} is an unobservable scalar random variable.

This latent space model could be motivated by the Aldous-Hoover representation of jointly exchangeable array (Aldous, 1981; Hoover, 1979) and its extension to relative exchangeable arrays by Crane and Towsner (2018). More extensive discussion can be found in Graham (2020a), Menzel (2021), and Davezies et al. (2021).

Compared to existing setups in the literature, the model here highlights the distinct roles of monadic and dyadic regressors. The latent variables A_i, U_i for example serve as elements of the exchangeable array representation and do not necessarily have an explicit structural interpretation. Interestingly, we will see later on the monadic and dyadic regressors may exhibit different rates of convergence. One crucial assumption is that the dimension of monadic regressors is greater or equal to one. We will give more comments on this after we present the rate of convergence result.

3.3 Estimand, Estimator, and Hoeffding Decomposition

We introduce the target parameter of our interest here and follow it with a specification and discussion of the estimator, density weighted average derivatives.

The object of our interest is the regression function of the outcome Y_{ij} given all the observed variables involving nodes i and j, $(B_i, C_i, B_j, C_j, D_{ij}, D_{ji})$. This regression function is assumed to be a function of $X_{ij} := (B_i, C_j, D_{ij}) \in \mathbb{R}^{d_X}$, [1] $d_X = d_B + d_C + d_D$. B_i captures the ego-relevant effect, C_j captures the alter-relevant effect, D_{ij} captures the dyadic effect:

$$\mathbb{E}\left[Y_{ij} | B_i, C_i, B_j, C_j, D_{ij}, D_{ji}\right] = \mathbb{E}[Y_{ij} | \underbrace{B_i, C_j, D_{ij}}_{X_{ij}}] = g\left(X_{ij}\right). \tag{3.1}$$

To save notation, we will assume $d_B = d_C$, which is the most common case in a typical setting. Besides its innate conceptual meaning, the ego vs alter distinction is technically relevant as it facilitates the statement of support conditions, which is crucial for nonparametric estimation.

The main assumption on this regression function is the single index restriction: $g(x) = G(x'\beta_0)$. This assumption says the regression function depends on the projection of x

[1] Notice there is no hidden exclusion restriction from a representation perspective, because we can always pinpoint the ego-relevant and alter-relevant variables after inspecting the conditional expectation given full information. However, real exclusion restrictions may be present when any specific specification is being used.

onto a single dimension identified by β_0. Under this assumption, $\theta_0 := \mathbb{E}\left(f(X)\frac{\partial g(X)}{\partial X}\right) = \mathbb{E}\left(f(X)\dot{G}(X'\beta_0)\right)\beta_0$ is proportional to the index coefficient β_0. If in addition the function $g(x)f^2(x)$ vanishes at the boundary, which facilitates the application of integration by parts, then as shown by Powell et al. (1989) (PSS thereafter) the following representation holds

$$\theta_0 = -2\mathbb{E}\left[Y\frac{\partial f(X)}{\partial X}\right]. \tag{3.2}$$

This representation motivates the density weighted average derivative estimator

$$\hat{\theta}_N = -2\frac{1}{N(N-1)}\sum_{i\neq j} Y_{ij}\frac{\partial}{\partial X_{ij}}\hat{f}_{N,ij}\left(X_{ij}\right), \tag{3.3}$$

where the two-way leave-out kernel density estimator is

$$\hat{f}_{N,ij}(x) = \frac{1}{(N-2)(N-3)}\sum_{\substack{(l,m)\\ l\neq m\neq i\neq j}}\frac{1}{h_N^d}K\left(\frac{x-X_{lm}}{h_N}\right). \tag{3.4}$$

Notice we leave dyads involving i and j out when we are estimating the density at point X_{ij}. This trick is a generalization of PSS's leave-out trick with the cross-sectional data to the dyadic data. It makes the bias much simpler to analyze, which resonates with the multiway cross fitting procedure in Chiang and Tan (2020). It also facilitates setting up the estimator as a fourth-order "U"-statistic, [2]

Define

$$\delta_{ijlm}(h) := Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h^d}K\left(\frac{X_{ij}-X_{lm}}{h}\right)$$

and its symmetrized version

$$p_{ijlm}(h) := \frac{1}{24}\sum_{(i',j',l',m')\in\mathcal{P}_{ijlm}}\delta_{i'j'l'm'}(h).$$

Each term $p_{ijlm}(h)$ is a symmetric function of primitive node variables and dyad variables involving i, j, l, m:

$$\begin{aligned} p_{ijlm}(h) = p\big(h;\ & A_i, B_i, C_i, U_i, A_j, B_j, C_j, U_j, A_l, B_l, C_l, U_l, A_m, B_m, C_m, U_m, \\ & \epsilon_{ij}, V_{ij}, \epsilon_{ji}, V_{ji}, \epsilon_{il}, V_{il}, \epsilon_{li}, V_{li}, \epsilon_{im}, V_{im}, \epsilon_{mi}, V_{mi}, \epsilon_{jl}, V_{jl}, \epsilon_{lj}, V_{lj}, \epsilon_{lm}, V_{lm}, \epsilon_{ml}, V_{ml}\big). \end{aligned}$$

[2] The quotation marks around the U indicates that the statistic is not strictly speaking a U-statistic in the traditional sense as an average of symmetric functions of combinations of iid random variables because of the presence of dyadic primitive variables ϵ and V in this statistics. However, this will not make the analysis of this statistic any different from a U-statistic, as will be revealed by the Hoeffding decomposition.

Write the estimator as a "U"-statistics:

$$\hat{\theta}_N(h) = -2\binom{N}{4}^{-1} \sum_{i<j<l<m} p_{ijlm}(h). \tag{3.5}$$

We will study its Hoeffding decomposition to understand its asymptotic behavior. Toward this end, define the information sets (σ-algebra) for one, two, and three indices as

$$\begin{aligned}
\mathcal{F}_{\{i\}} &= \sigma\left(A_i, B_i, C_i, U_i\right) \\
\mathcal{F}_{\{i,j\}} &= \sigma\left(A_i, B_i, C_i, U_i, A_j, B_j, C_j, U_j, \epsilon_{ij}, V_{ij}, \epsilon_{ji}, V_{ji}\right) \\
\mathcal{F}_{\{i,j,l\}} &= \sigma\left(A_i, B_i, C_i, U_i, A_j, B_j, C_j, U_j, A_l, B_l, C_l, U_l, \epsilon_{ij}, V_{ij}, \epsilon_{ji}, V_{ji}, \epsilon_{il}, V_{il}, \epsilon_{li}, V_{li}, \epsilon_{jl}, V_{jl}, \epsilon_{lj}, V_{lj}\right).
\end{aligned}$$

Define the Hoeffding decomposition of the "U"-statistics $U_N(h) := \binom{N}{4}^{-1} \sum_{i<j<l<m} p_{ijlm}(h)$ by

$$U_N(h) = \sum_{c=0}^{4} \binom{4}{c} U_{N,c}(h),$$

in which each $U_{N,c}(h)$ is defined by

$$U_{N,c}(h) = \frac{1}{\binom{N}{c}} \sum_{1 \le i_1 < \cdots < i_c \le N} q_{c,i_1 \ldots i_c}(h)$$

with

$$\begin{aligned}
q_0(h) &= \mathbb{E}\left(p_{ijlm}(h)\right) \\
q_{1,i}(h) &= \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i\}}\right) - \mathbb{E}\left(p_{ijlm}(h)\right) \\
q_{2,ij}(h) &= \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,j\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{j\}}\right) + \mathbb{E}\left(p_{ijlm}(h)\right) \\
q_{3,ijl}(h) &= \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,j,l\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,j\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,l\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{j,l\}}\right) \\
&\quad + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i\}}\right) + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{j\}}\right) + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{l\}}\right) - \mathbb{E}\left(p_{ijlm}(h)\right) \\
q_{4,ijlm}(h) &= p_{ijlm}(h) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,j,l\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,j,m\}}\right) \\
&\quad - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,l,m\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{j,m,l\}}\right) \\
&\quad + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,j\}}\right) + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,l\}}\right) + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i,m\}}\right) + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{j,l\}}\right) \\
&\quad + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{j,m\}}\right) + \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{l,m\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{i\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{j\}}\right) \\
&\quad - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{l\}}\right) - \mathbb{E}\left(p_{ijlm}(h)|\mathcal{F}_{\{m\}}\right) + \mathbb{E}\left(p_{ijlm}(h)\right).
\end{aligned}$$

Namely $U_{N,0}(h) = \mathbb{E}\left(p_{ijlm}(h)\right)$ is the expectation. $U_{N,1}(h), U_{N,2}(h), U_{N,3}(h), U_{N,4}(h)$ are "U"-statistics of order 1, 2, 3, 4. $U_{N,1}(h)$, the first-order terms, is often referred as the Hajek projection. The Hajek projection gives a best approximation by a sum of functions of one node information $\mathcal{F}_{\{i\}}$ at a time. The Hoeffding decomposition gives improved approximation by using sums of functions of two, three, or four nodes information.

3.4 Assumptions

This section presents assumptions on the model, kernel, and bandwidth sequences. These assumptions ensure that the estimand θ_0 is well-defined and the estimator $\hat{\theta}_N$ is well-behaved.

Assumption 3.4.1 (Model). (a) $\mathbb{E}Y_{ij}^4 < \infty$.

(b) Ω as defined in lemma 3.6.3 is positive definite.

(c) The density function f and the function gf is $(Q+1)$ times differentiable, and f, gf and their first $(Q+1)$ derivatives are bounded, for some $Q \geq 2$.

(d) $\lim_{||x|| \to \infty}[f(x) + |g(x)f(x)|] = 0$ where $|| \cdot ||$ is the Euclidean norm.

(e) (B_i, D_{ij}) conditional on $\mathcal{F}_{\{j\}}$ has density $f_{B_i, D_{ij}|\mathcal{F}_{\{j\}}}\left(b, d|\mathcal{F}_{\{j\}}\right)$, which is bounded together with its first two derivatives. $\mathbb{E}\left[Y_{ij}|B_i = b, D_{ij} = d, \mathcal{F}_{\{j\}}\right] f_{B_i, D_{ij}|\mathcal{F}_{\{j\}}}\left(b, d|\mathcal{F}_{\{j\}}\right)$ and its first two derivatives are bounded.

(f) (C_j, D_{ij}) conditional on $\mathcal{F}_{\{i\}}$ has density $f_{C_j, D_{ij}|\mathcal{F}_i}\left(c, d|\mathcal{F}_{\{i\}}\right)$, which is bounded together with its first two derivatives. $\mathbb{E}\left[Y_{ij}|C_j = c, D_{ij} = d, \mathcal{F}_{\{i\}}\right] f_{C_j, D_{ij}|\mathcal{F}_{\{i\}}}\left(c, d|\mathcal{F}_{\{j\}}\right)$ and its first two derivatives are bounded.

The Ω in assumption 3.4.1 Part (b) is the asymptotic variance of the second-order terms in the Hoeffding decomposition. This assumption says this second-order term is nondegenerate. We are not imposing assumptions on the first-order term. The asymptotic variance of the first-order term

$$\Sigma := \operatorname{Var}\left(\mathbb{E}\left[\frac{\partial}{\partial X_{ij}} gf(X_{ij}) + \frac{\partial}{\partial X_{ji}} gf(X_{ji}) - Y_{ij}\frac{\partial}{\partial X_{ij}} f\left(X_{ij}\right) - Y_{ji}\frac{\partial}{\partial X_{ji}} f\left(X_{ji}\right) \middle| \mathcal{F}_{\{i\}}\right]\right)$$

is allowed to be not strictly positive definite. In this sense, we are allowing for first-order degeneracy.

The other assumptions are mostly technical. They ensures relevant smoothness and boundedness.

Assumption 3.4.2 (Kernel). (a) K follows a product form: for $x = (b, c, d)$ where $b \in \mathbb{R}^{d_B}$, $c \in \mathbb{R}^{d_C}$, $d \in \mathbb{R}^{d_D}$, $K(x) = K_B(b)K_C(c)K_D(d)$. $\int_{\mathbb{R}^{d_B}} K_B(b)db = \int_{\mathbb{R}^{d_C}} K_C(c)dc = \int_{\mathbb{R}^{d_D}} K_D(\mathbf{d})d\mathbf{d} = 1$

(b) K_B, K_C, K_D are even and differentiable. K and its first derivative are bounded.

(c) $\int \dot{K}(u)\dot{K}(u)'du$ is positive definite where $\dot{K}(u) = \frac{\partial}{\partial u}K(u)$.

(d) For some $Q > 2$, $\int_{\mathbb{R}^d} |K(u)|\left(1 + ||u||^Q\right)du + \int ||\dot{K}(u)||\left(1 + ||u||^2\right)du < \infty$, and

$$\int_{\mathbb{R}^{d_X}} u_1^{l_1} \cdots u_{d_X}^{l_{d_X}} K(u)du = \begin{cases} 1, & \text{if } l_1 = \cdots = l_{d_X} = 0 \\ 0, & \text{if } (l_1, \ldots, l_d)' \in \mathbb{Z}_+^{d_X} \text{ and } l_1 + \cdots + l_{d_X} < Q \end{cases}$$

The assumptions on the kernel are standard. We use multiplicative kernels, $K(x) = K_B(b)K_C(c)K_D(d)$, to simplify the derivation.

Assumption 3.4.3 (Bias). $Nh_N^{Q+\frac{d_B}{2}} \to 0$.

This assumption is equivalent to $N^{-2}h_N^{-d_B} \gg h_N^{2Q}$, which ensures the square bias of the estimator is always asymptotically smaller than the second-order term $U_{N,2}$ in the Hoeffding decomposition.

Assumption 3.4.4 (Asymptotically Quadratic). $Nh_N^{\frac{d_B+d_D+2}{2}} \to \infty$

This condition ensures the estimator $\hat{\theta}_N$ is asymptotically equivalent to a second-order U-statistic, the sum of its first-order and second-order terms in the Hoeffding decomposition. This condition is weaker than the more frequently seen asymptotic linearity condition, which would imply the estimator $\hat{\theta}_N$ is asymptotically equivalent to a sample mean, a first-order U-statistic. To see this more closely, note the condition for asymptotic linearity is $Nh_N^{\max\left\{d_B+2, \frac{d_B+2d_D+2}{3}\right\}} \to \infty$. Because $\frac{d_B+d_D+2}{2} = \frac{1}{4}(d_B+2) + \frac{3}{4}\left(\frac{d_B+2d_D+2}{3}\right)$ is a convex combination of d_B+2 and $\frac{d_B+2d_D+2}{3}$, assumption 3.4.4 is a weakly weaker condition. Also, this condition ensures the second-order term always dominates the fourth-order term both. Under nondegeneracy, this condition can be weaken to $Nh_N^{\min\left\{\frac{d_B+d_D+2}{2}, \frac{d_B+2d_D+2}{3}\right\}} \to \infty$.

3.5 Asymptotic Quadratic Approximation and Asymptotic Normality

This section presents the main theoretical results. The final goals are to show the asymptotic normality of the estimator and [to present a consistent variance estimator].

Both lemma 3.5.1, 3.5.2 below are imported from PSS. Lemma 3.5.1 gives us the representation motivating the estimator. Lemma 3.5.2 bounds the order of magnitude of the bias. Their proofs are in the appendix.

Lemma 3.5.1 (Representation). *Given Assumptions 3.4.1,*

$$\theta_0 := \mathbb{E}\left(f(X)\frac{\partial g(X)}{\partial X}\right) = -2\mathbb{E}\left(Y\frac{\partial f(X)}{\partial X}\right)$$

Lemma 3.5.2 (Bias). *Given Assumption 3.4.1, 3.4.2 and $h_N \to 0$,*

$$\mathbb{E}\hat{\theta}_N - \theta = O\left(h_N^Q\right)$$

Lemma 3.5.3 below involves calculation specific for the dyadic data setup. The detail of the calculation is in the appendix. The key task in the calculation is figuring out the order of magnitude of the second moments of $\mathbb{E}\left[p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}\right]$, $\mathbb{E}\left[p_{ijlm}(h_N)|\mathcal{F}_{\{i,j\}}\right]$, $\mathbb{E}\left[p_{ijlm}(h_N)|\mathcal{F}_{\{i,j,l\}}\right]$, and $p_{ijlm}(h_N)$. Taking conditional expectation of the kernel object $p_{ijlm}(h_N)$ given some information set means smoothing (integrating) over variables conditionally continuously distributed while keeping variables fully pinned down by these information fixed. Intuition tells us conditional expectation given less information has smaller order of magnitude because more conditionally continuously distributed variables are getting smoothed over. The support conditions specified in Assumption 3.4.1 on the regressor $X_{ij} = (B_i, C_j, D_{ij})$ solidify this intuition by making sure all the variables not fully pinned down to be conditionally continuously distributed.

Lemma 3.5.3 (Order of Magnitude of the Hoeffding Decomposition)**.** *Given Assumption 3.4.1, 3.4.2, and $h_N \to 0$, the order of magnitude of terms in the Hoeffding decomposition are*

$$\operatorname{Var}\left(\sqrt{N}U_{N,1}\right) = \frac{1}{16}\Sigma + O\left(h_N^Q\right) \tag{3.6}$$

$$\operatorname{Var}\left(\sqrt{\binom{N}{2}}\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,2}(h_N)\right) = \frac{1}{36}\Omega + O\left(h_N\right) \tag{3.7}$$

$$\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,3}\right) = O\left(N^{-3}\right) \tag{3.8}$$

$$\operatorname{Var}\left(U_{N,4}\right) = N^{-4}O\left(h_N^{-(d_X+2)}\right). \tag{3.9}$$

Proof of this lemma is in the appendix. Notice that these results crucially depend on the presence of monadic regressors. Without the presence of monadic regressors, the convergence rate of the dyadic regressor will be parametric (either $\sqrt{N}$ or $\sqrt{\binom{N}{2}}$ depending on (non)degeneracy).

Notice the variance of the Hajek projection $4U_{N,1}$ is of order $O(N^{-1})$ under nondegeneracy. The variance of the second-order term $U_{N,2}$ is always larger than that of the third-order term $U_{N,3}$ by order of N. The relative magnitude of the fourth-order term versus the second-order term depends on the bandwidth: only when the bandwidth is very small the fourth-order term is larger. When the bandwidth is large, the Hajek projection captures the leading variance. This together with Lindberg-Levy CLT leads to the result of first-order asymptotic normality. This result will technically break if the U-statistic is degenerate, in which case the first-order terms vanish. The concern of degeneracy motivates us to explore a more robust asymptotic approximation using a quadratic form, which incorporates the second-order terms in addition to the first-order terms. Verifying conditions of a martingale CLT in Eubank and Wang (1999) gives us asymptotic normality of the quadratic approximation, which holds under both degeneracy and nondegeneracy and across a broader range of bandwidth sequences (Cattaneo et al., 2014b).

Lemma 3.5.4 (CLT). *Given Assumption 3.4.1, 3.4.2, 3.4.3, 3.4.4,*

$$\begin{pmatrix} \sqrt{N}4U_{N,1}(h_N) \\ \sqrt{\binom{N}{2}} \begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} 6U_{N,2}(h_N) \end{pmatrix} \rightsquigarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Omega \end{pmatrix}\right). \tag{3.10}$$

Proof. Our proof strategy follows that of CCJ (2014).

To prepare, notice lemma 3.6.2 shows that $\mathbb{E}[p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}] = \frac{1}{4}\eta_i + O\left(h_N^Q\right)$, in which $\eta_i = \mathbb{E}\left[-\frac{\partial}{\partial X_{ij}}gf(X_{ij}) - \frac{\partial}{\partial X_{ji}}gf(X_{ji}) + Y_{ij}\frac{\partial}{\partial X_{ij}}f\left(X_{ij}\right) + Y_{ji}\frac{\partial}{\partial X_{ji}}f\left(X_{ji}\right)\middle|\mathcal{F}_{\{i\}}\right]$. η_i doesn't depend on N. This implies

$$\sqrt{N}4U_{N,1}(h_N) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\eta_i - \mathbb{E}\eta_i) + O_P\left(h_N^Q\right),$$

because

$$\begin{aligned} &\sqrt{N}4U_{N,1}(h_N) - \frac{1}{\sqrt{N}}\sum_{i=1}^{N}(\eta_i - \mathbb{E}\eta_i) \\ &= \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\left[4\mathbb{E}\left[p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}\right] - \eta_i - (4\mathbb{E}[p_{ijlm}(h_N)] - \mathbb{E}\eta_i)\right] \end{aligned}$$

whose variance is bounded above by $O\left(h_N^{2Q}\right)$.

The multivariate CLT (3.10) holds once

$$\sqrt{N}4u_{N,1} + \sqrt{\binom{N}{2}}6u_{N,2}(h_N) \rightsquigarrow \mathrm{N}\left(0, \sigma^2 + \omega^2\right) \tag{3.11}$$

for any $\lambda_1 \in \mathbb{R}^{d_X}$ and $\lambda_2 \in \mathbb{R}^{d_X}$, where

$$u_{N,1} = \frac{1}{N}\sum_{i=1}^{N} l_i, \qquad l_i = \lambda_1'(\eta_i - \mathbb{E}\eta_i), \qquad \sigma^2 = \lambda_1'\Sigma\lambda_1,$$

$$u_{N,2} = \binom{N}{2}^{-1}\sum_{i<j} w_{N,ij}, \qquad w_{N,ij} = \lambda_2' \begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N), \qquad \omega^2 = \lambda_2'\Omega\lambda_2.$$

Assuming λ_1 and λ_2 are both none zero, we establish (3.11) by invoking the theorem of Eubank and Wang (1999). In our notation, conditions (1.3)-(1.6) of Eubank and Wang (1999) are

$$\binom{N}{2}^{-1} \max_{1\leq j\leq N} \sum_{i=1}^{N} \mathbb{E}\left[w_{N,ij}^2\right] \to 0 \tag{3.12}$$

$$\binom{N}{2}^{2} \mathbb{E}u_{N,2}^4 \to 3\omega^4 \tag{3.13}$$

$$N^{-2}\sum_{i=1}^{N} \mathbb{E}\left[l_i^4\right] \to 0 \tag{3.14}$$

$$\binom{N}{2}^{-1} N^{-1}\mathbb{E}\left[\left(\sum_{j=2}^{N}\sum_{i=1}^{j-1}\mathbb{E}\left(w_{N,ij}l_j|\mathcal{F}_{\{1,\ldots,j-1\}}\right)\right)^2\right] \to 0. \tag{3.15}$$

Because of exchangeability, (3.12) is equivalent to $N^{-1}\mathbb{E}\left[w_{N,ij}^2\right] \to 0$, which is satisfied because of

$$\mathbb{E}\left[w_{N,ij}^2\right] = \lambda_2' \operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N)\right)\lambda_2 \to \lambda_2'\Omega\lambda_2 = \omega^2 \tag{3.16}$$

by lemma (3.6.4).

By de Jong (1987) Prop. 3.1, condition (3.13) is satisfied if

$$N^{-2}\mathbb{E}[w_{N,ij}^4] \to 0 \tag{3.17}$$

$$N^{-1}\mathbb{E}[w_{N,ij}^2 w_{N,il}^2] \to 0 \tag{3.18}$$

$$\mathbb{E}[w_{N,ij}w_{N,il}w_{N,jm}w_{N,lm}] \to 0 \tag{3.19}$$

$$\mathbb{E}[w_{N,ij}^2] \to \omega^2 \tag{3.20}$$

We verify each of the condition one by one.
(3.17): $\mathbb{E}[w_{N,ij}^4] = h_N^{-d_B}$ by change of variable and (3.17) holds if $N^2 h_N^{d_B} \to \infty$, which is ensured by assumption 3.4.4.
(3.18): $\mathbb{E}[w_{N,ij}^2 w_{N,il}^2] = \mathbb{E}[\mathbb{E}[w_{N,ij}^2|\mathcal{F}_{\{i\}}]^2] = O(1)$ and (3.18) holds because $N^{-1}O(1) = o(1)$.
(3.19): $\mathbb{E}[w_{N,ij}w_{N,il}w_{N,jm}w_{N,lm}] = \mathbb{E}[\mathbb{E}[w_{N,ij}w_{N,il}|\mathcal{F}_{\{j,l\}}]^2] = h_N^d \to 0$. This can be shown using change of variable.
(3.20): This is the same as (3.16).
Condition (3.14) is ensured by $\mathbb{E}[l_i^4] = O(1)$ and $N^{-2}\mathbb{E}[l_i^4] = N^{-2}O(1) = o(1)$.
Condition (3.15) is equivalent to $\mathbb{E}\left[\left(\mathbb{E}\left(w_{N,ij}l_j|\mathcal{F}_{\{i\}}\right)\right)^2\right] \to 0$. Since $\mathbb{E}\left(w_{N,ij}l_j|\mathcal{F}_{\{i\}}\right) = h_N^{\frac{d_B}{2}}$ by integration by parts and bounding arguments, (3.15) is satisfied. □

Now we are ready to prove the asymptotic normality result based on asymptotic quadratic approximation.

Theorem 3.5.5 (Asymptotic Quadratic Approximation)**.** *Given Assumptions 3.4.1, 3.4.2, 3.4.3, 3.4.4, the second-order approximation of U_N dominates the approximation error,*

$$U_N - U_{N,0}\left(h_N\right) = \left[4U_{N,1}\left(h_N\right) + 6U_{N,2}\left(h_N\right)\right]\left(1 + o_P(1)\right).$$

And $\hat{\theta}_N$ is asymptotically normal

$$\operatorname{Var}\left(\hat{\theta}_N\right)^{-1/2}\left(\hat{\theta}_N - \theta_0\right) \rightsquigarrow N(0, I_{d_X}),$$

in which the variance

$$\operatorname{Var}\left(\hat{\theta}_N\right) = 4\left[N^{-1}\Sigma + \binom{N}{2}^{-1} h_N^{-d_B}\begin{pmatrix} h_N^{-1} & 0 & 0 \\ 0 & h_N^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}\Omega\begin{pmatrix} h_N^{-1} & 0 & 0 \\ 0 & h_N^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix}\right](1 + o(1))$$

Proof. Remember by definition the estimator $\hat{\theta}_N = -2U_N$. The Hoeffding decompositon of U_N is

$$U_N(h_N) - U_{N,0}(h_N) = 4U_{N,1}(h_N) + 6U_{N,2}(h_N) + 4U_{N,3}(h_N) + U_{N,4}(h_N).$$

$U_{N,1}(h_N)$, $U_{N,2}(h_N)$, $U_{N,3}(h_N)$, $U_{N,4}(h_N)$ are all mean zero. They are uncorrelated with each other and their variance are calculated in lemma 3.5.3.

First, we will show the bias is asymptotically negligible. Notice

$$\begin{aligned} \hat{\theta}_N - \theta_0 &= -2U_N - \theta_0 \\ &= -2(U_N - U_{N,0}) + (-2U_{N,0} - \theta_0), \end{aligned}$$

in which the first part is the deviation from the mean and the second part is the bias. Lemma 3.5.2 tells us the bias square is bounded above by $O\left(h_N^{2Q}\right)$. Lemma 3.5.3 together with assumption 3.4.1 (b) tells us the variance of the second-order term $\operatorname{Var}\left(U_{N,2}\right)$ is of order at least $N^{-2}h_N^{d_B}$. These together with the bandwidth condition assumption 3.4.3 ensures $\left(-2U_{N,0} - \theta_0\right)^2 = o\left(\operatorname{Var}\left(U_{N,2}\right)\right)$ and $\left(-2U_{N,0} - \theta_0\right)^2 = o\left(\operatorname{Var}\left(U_N\right)\right)$. As a result, bias plays an asymptotically negligible role.

$$\begin{aligned} \operatorname{Var}\left(\hat{\theta}_N\right)^{-1/2}\left(\hat{\theta}_N - \theta_0\right) &= -\operatorname{Var}\left(U_N\right)^{-1/2}\left(U_N - U_{N,0}\right) - \operatorname{Var}\left(U_N\right)^{-1/2}\left(U_{N,0} + \frac{\theta_0}{2}\right) \\ &= -\operatorname{Var}\left(U_N\right)^{-1/2}\left(U_N - U_{N,0}\right) + o(1). \end{aligned}$$

Second, we will show CLT holds for the centered statistic $U_N - U_{N,0}$.

Under Assumption 3.4.1 (b) about second order nondegeneracy and assumption 3.4.4, lemma 3.5.3 ensures the second-order term dominates the third-order term and the fourth-order term $\operatorname{Var}\left(U_{N,3}(h_N)\right) = o\left(\operatorname{Var}\left(U_{N,2}(h_N)\right)\right)$ and $\operatorname{Var}\left(U_{N,4}(h_N)\right) = o\left(\operatorname{Var}\left(U_{N,2}(h_N)\right)\right)$.

Consequently, the sum of the first- and second-order term dominates in the Hoeffding decomposition $\mathrm{Var}\left(4U_{N,3}+U_{N,4}\right)=o\left(\mathrm{Var}\left(4U_{N,1}+6U_{N,2}\right)\right)$. Hence,

$$
\begin{aligned}
&\mathrm{Var}\left(U_N\right)^{-1/2}\left(U_N-U_{N,0}\right)\\
&=\mathrm{Var}\left(4U_{N,1}+6U_{N,2}+4U_{N,3}+U_{N,4}\right)^{-1/2}\left(4U_{N,1}+6U_{N,2}+4U_{N,3}+U_{N,4}\right)\\
&=\left[\mathrm{Var}\left(4U_{N,1}+6U_{N,2}\right)\right]^{-1/2}\left(4U_{N,1}+6U_{N,2}\right)+o(1).
\end{aligned}
$$

CLT in lemma 3.5.4 together with the fact that

$$
\begin{aligned}
4U_{N,1}+6U_{N,2}&=N^{-1/2}4\sqrt{N}U_{N,1}\\
&+\binom{N}{2}^{-1/2}\begin{pmatrix} h_N^{-\frac{d_B+2}{2}} & 0 & 0\\ 0 & h_N^{-\frac{d_B+2}{2}} & 0\\ 0 & 0 & h_N^{-\frac{d_B}{2}}\end{pmatrix}\sqrt{\binom{N}{2}}\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0\\ 0 & h_N^{\frac{d_B+2}{2}} & 0\\ 0 & 0 & h_N^{\frac{d_B}{2}}\end{pmatrix}6U_{N,2}
\end{aligned}
$$

implies the CLT holds for $4U_{N,1}+6U_{N,2}$: $\mathrm{Var}\left(U_N\right)^{-1/2}\left(U_N-U_{N,0}\right)\rightsquigarrow \mathrm{N}(0,I_{d_X})$. Hence,

$$
\mathrm{Var}\left(\hat{\theta}_N\right)^{-1/2}\left(\hat{\theta}_N-\theta_0\right)\rightsquigarrow \mathrm{N}(0,I_{d_X}).
$$

□

This robust asymptotic normality result based on asymptotic quadratic approximation suggests the possibility of constructing a normality-based robust confidence interval. The only ingredient left to be filled is a consistent variance estimator. We propose a variance estimator in the following. Though we do not have a formal consistency result for this estimator, we conjecture that this estimator will be consistent under similar to that in theorem 3.5.5.

Variance Estimation

Motivated by the variance expression

$$
\frac{1}{4}\mathrm{Var}\left(\hat{\theta}_N\right)=\binom{N}{4}^{-2}\sum_{i_1<j_1<l_1<m_1}\sum_{i_2<j_2<l_2<m_2}\mathrm{Cov}\left(p_{i_1j_1l_1m_1}(h_N),p_{i_2j_2l_2m_2}(h_N)\right),
$$

we propose the following analog estimator

$$
\begin{aligned}
\frac{1}{4}\hat{V}\left(\hat{\theta}_N\right)=\binom{N}{4}^{-2}\sum_{i_1<j_1<l_1<m_1}\sum_{i_2<j_2<l_2<m_2}&\Big\{d\left(i_1,j_1,l_1,m_1,i_2,j_2,l_2,m_2\right)\\
&\cdot\left[p_{i_1j_1l_1m_1}(h_N)-U_N\right]\left[p_{i_2j_2l_2m_2}(h_N)-U_N\right]'\Big\}.
\end{aligned}
$$

Consistency result of this estimator is left for further research.

To prove consistency, we will decompose this variance estimator into four parts each corresponding to one term in the Hoeffding decomposition.

$$
\begin{aligned}
d\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) = {} & d_1\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) + d_2\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) \\
& + d_3\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) + d_4\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right),
\end{aligned}
$$

where

$$
\begin{aligned}
d_1\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) &= \mathbb{1}\left(\left|\left\{i_1, j_1, l_1, m_1\right\} \cap \left\{i_2, j_2, l_2, m_2\right\}\right| = 1\right) \\
d_2\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) &= \mathbb{1}\left(\left|\left\{i_1, j_1, l_1, m_1\right\} \cap \left\{i_2, j_2, l_2, m_2\right\}\right| = 2\right) \\
d_3\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) &= \mathbb{1}\left(\left|\left\{i_1, j_1, l_1, m_1\right\} \cap \left\{i_2, j_2, l_2, m_2\right\}\right| = 3\right) \\
d_4\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) &= \mathbb{1}\left(\left|\left\{i_1, j_1, l_1, m_1\right\} \cap \left\{i_2, j_2, l_2, m_2\right\}\right| = 4\right).
\end{aligned}
$$

Write $\hat{V}\left(U_N\right)$ into four parts.

$$
\begin{aligned}
\hat{V}\left(U_N\right) &= \hat{V}_1 + \hat{V}_2 + \hat{V}_3 + \hat{V}_4 \\
\hat{V}_1 &= \binom{N}{4}^{-2} \sum_{i_1<j_1<l_1<m_1} \sum_{i_2<j_2<l_2<m_2} \Big\{ d_1\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) \\
&\qquad \cdot \left[p_{i_1 j_1 l_1 m_1}(h_N) - U_N\right]\left[p_{i_2 j_2 l_2 m_2}(h_N) - U_N\right]' \Big\} \\
\hat{V}_2 &= \binom{N}{4}^{-2} \sum_{i_1<j_1<l_1<m_1} \sum_{i_2<j_2<l_2<m_2} \Big\{ d_2\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) \\
&\qquad \cdot \left[p_{i_1 j_1 l_1 m_1}(h_N) - U_N\right]\left[p_{i_2 j_2 l_2 m_2}(h_N) - U_N\right]' \Big\} \\
\hat{V}_3 &= \binom{N}{4}^{-2} \sum_{i_1<j_1<l_1<m_1} \sum_{i_2<j_2<l_2<m_2} \Big\{ d_3\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) \\
&\qquad \cdot \left[p_{i_1 j_1 l_1 m_1}(h_N) - U_N\right]\left[p_{i_2 j_2 l_2 m_2}(h_N) - U_N\right]' \Big\} \\
\hat{V}_4 &= \binom{N}{4}^{-2} \sum_{i_1<j_1<l_1<m_1} \sum_{i_2<j_2<l_2<m_2} \Big\{ d_4\left(i_1, j_1, l_1, m_1, i_2, j_2, l_2, m_2\right) \\
&\qquad \cdot \left[p_{i_1 j_1 l_1 m_1}(h_N) - U_N\right]\left[p_{i_2 j_2 l_2 m_2}(h_N) - U_N\right]' \Big\}
\end{aligned}
$$

We'd like to show

$$N\hat{V}_1 \xrightarrow{\text{P}} \Sigma \tag{3.21}$$

$$\binom{N}{2} h_N^{d_B} \begin{pmatrix} h_N & 0 & 0 \\ 0 & h_N & 0 \\ 0 & 0 & 1 \end{pmatrix} \hat{V}_2 \begin{pmatrix} h_N & 0 & 0 \\ 0 & h_N & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{\text{P}} \Omega \tag{3.22}$$

$$\hat{V}_3 + \hat{V}_4 = o_P \left(N^{-1}\Sigma + \binom{N}{2}^{-1} h_N^{-d_B} \begin{pmatrix} h_N^{-1} & 0 & 0 \\ 0 & h_N^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \Omega \begin{pmatrix} h_N^{-1} & 0 & 0 \\ 0 & h_N^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \tag{3.23}$$

Completion of this proof outline is left for further research.

3.6 Appendix: Proofs

This Appendix contains proofs of results in the main text of Chapter 3.

Proof of Lemma 3.5.1

Proof. Partition x into its first coordinate vs the rest, $x = (x_1, x_{-1}')'$,

$$\begin{aligned} \mathbb{E}\left(f(X) \frac{\partial g(X)}{\partial X} \right) &= \int f^2(x) \frac{\partial g(x)}{\partial x_1} dx \\ &= \int dx_0 \int f^2(x_1, x_{-1}) \frac{\partial g(x_1, x_{-1})}{\partial x_1} dx_1 \\ &= -2 \int dx_0 \int f(x_1, x_{-1}) \frac{\partial f(x_1, x_{-1})}{\partial x_1} g(x_1, x_{-1}) dx_1 \\ &= -2\mathbb{E}\left(Y \frac{\partial f(X)}{\partial X} \right), \end{aligned}$$

where the third equality follows by integration by part and the limit condition

$$\begin{aligned} \int f^2(x_1, x_{-1}) \frac{\partial g(x_1, x_{-1})}{\partial x_1} dx_1 &= g(x_1, x_{-1}) f^2(x_1, x_{-1})|_{-\infty}^{+\infty} \\ &\quad - \int 2f(x_1, x_{-1}) \frac{\partial f(x_1, x_{-1})}{\partial x_1} g(x_1, x_{-1}) dx_1 \\ &= -\int 2f(x_1, x_{-1}) \frac{\partial f(x_1, x_{-1})}{\partial x_1} g(x_1, x_{-1}) dx_1. \end{aligned}$$

□

Proof of Lemma 3.5.2

Proof. Use the following notation to distinguish the first coordinate from the rest in $X_{ij} = (X_{ij,1}, X'_{ij,-1})'$.

$$\begin{aligned}\mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij,1}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\right] &= \mathbb{E}\left[g\left(X_{ij}\right)\mathbb{E}\left[\frac{\partial}{\partial X_{ij,1}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)|X_{ij}\right]\right] \\ &= \mathbb{E}\left[g\left(X_{ij}\right)\left(\frac{\partial}{\partial X_{ij,1}}f(X_{ij}) + O\left(h^Q\right)\right)\right] \\ &= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij,1}}f(X_{ij})\right] + O\left(h^Q\right),\end{aligned}$$

which implies $\mathbb{E}\hat{\theta}_N = \theta + O\left(h^Q\right)$. The second equality follows by

$$\begin{aligned}&\mathbb{E}\left[\frac{\partial}{\partial X_{ij,1}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)|X_{ij}\right] \\ &= \int\frac{\partial}{\partial X_{ij,1}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-x}{h_N}\right)f(x)dx \\ &= \int\int\frac{1}{h_N}\frac{\partial}{\partial u_1}K\left(u_1,u_{-1}\right)f(X_{ij,1}-h_Nu_1, X_{ij,-1}-h_Nu_{-1})du_1du_{-1} \\ &= \int K\left(u\right)\frac{\partial}{\partial(X_{ij,1}-h_Nu)}f(X_{ij}-h_Nu)du \\ &= \frac{\partial}{\partial X_{ij,1}}f(X_{ij}) + O\left(h^Q\right),\end{aligned}$$

in which the third equality follows by integration by part:

$$\begin{aligned}&\int\frac{1}{h_N}\frac{\partial}{\partial u_1}K\left(u_1,u_{-1}\right)f(X_{ij,1}-h_Nu_1, X_{ij,-1}-h_Nu_{-1})du_1 \\ &= \frac{1}{h_N}K(u_1,u_{-1})f(X_{ij,1}-h_Nu_1, X_{ij,-1}-h_Nu_{-1})|_{-\infty}^{+\infty} \\ &\qquad + \int K(u_1,u_{-1})\frac{\partial}{\partial(X_{ij,1}-h_Nu)}f(X_{ij,1}-h_Nu_1, X_{ij,-1}-h_Nu_{-1})du_1 \\ &= \int K(u_1,u_{-1})\frac{\partial}{\partial(X_{ij,1}-h_Nu)}f(X_{ij,1}-h_Nu_1, X_{ij,-1}-h_Nu_{-1})du_1\end{aligned}$$

□

Proof of Lemma 3.5.3

To prove lemma 3.5.3, we will need to calculate $\mathbb{E}\left[p_{ijlm}(h)|\mathcal{F}_{\{i\}}\right]$, $\mathbb{E}\left[p_{ijlm}(h)|\mathcal{F}_{\{i,j\}}\right]$, $\mathbb{E}\left[p_{ijlm}(h)|\mathcal{F}_{\{i,j,l\}}\right]$. Toward this end, we will first calculate the conditional expectations of

$\delta_{ijlm}(h)$ in lemma 3.6.1. Then we use these result to calculate and bound corresponding conditional second moments of $p_{ijlm}(h)$ in lemma 3.6.3 and variance of $q_{N,1,i}, q_{N,2,ij}, q_{N,3,ijl}, q_{N,4,ijlm}$ in lemma 3.6.4. Lemma 3.5.3 is a natural result of lemma 3.6.4.

In the following, for a vector $V = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$, we use V^2 to denote element-wise square $V^2 = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}^2 = \begin{pmatrix} v_1^2 \\ v_2^2 \end{pmatrix}$.

Lemma 3.6.1. *Under Assumption 3.4.1, 3.4.2, and* $h_N \to 0$,

(a)

$$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j,l,m\}}\right] = \begin{pmatrix} \frac{1}{h_N^{d_C}} K_C\left(\frac{C_j - C_m}{h_N}\right)\left[-\frac{\partial}{\partial B_l}\left\{\mathbb{E}\left[Y_{ij}|B_i = B_l, D_{ij} = D_{lm}, \mathcal{F}_{\{j\}}\right] f_{B_i, D_{ij}|\mathcal{F}_{\{j\}}}\left(B_l, D_{lm}|\mathcal{F}_{\{j\}}\right)\right\} + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_C+1}} \dot{K}_C\left(\frac{C_j - C_m}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|B_i = B_l, D_{ij} = D_{lm}, \mathcal{F}_{\{j\}}\right] f_{B_i, D_{ij}|\mathcal{F}_{\{j\}}}\left(B_l, D_{lm}|\mathcal{F}_{\{j\}}\right) + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_C}} K_C\left(\frac{C_j - C_m}{h_N}\right)\left[-\frac{\partial}{\partial D_{lm}}\left\{\mathbb{E}\left[Y_{ij}|B_i = B_l, D_{ij} = D_{lm}, \mathcal{F}_{\{j\}}\right] f_{B_i, D_{ij}|\mathcal{F}_{\{j\}}}\left(B_l, D_{lm}|\mathcal{F}_{\{j\}}\right)\right\} + O\left(h_N\right)\right]. \end{pmatrix},$$

$$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,l,m\}}\right] = \begin{pmatrix} \frac{1}{h_N^{d_B+1}} \dot{K}_B\left(\frac{B_i - B_l}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|C_j = C_m, D_{ij} = D_{lm}, \mathcal{F}_{\{i\}}\right] f_{C_j, D_{ij}|\mathcal{F}_{\{i\}}}\left(C_m, D_{lm}|\mathcal{F}_{\{i\}}\right) + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_B}} K_B\left(\frac{B_i - B_l}{h_N}\right)\left[-\frac{\partial}{\partial C_m}\left\{\mathbb{E}\left[Y_{ij}|C_j = C_m, D_{ij} = D_{lm}, \mathcal{F}_{\{i\}}\right] f_{C_j, D_{ij}|\mathcal{F}_{\{i\}}}\left(C_m, D_{lm}|\mathcal{F}_{\{i\}}\right)\right\} + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_B}} K_B\left(\frac{B_i - B_l}{h_N}\right)\left[-\frac{\partial}{\partial D_{lm}}\left\{\mathbb{E}\left[Y_{ij}|C_j = C_m, D_{ij} = D_{lm}, \mathcal{F}_{\{i\}}\right] f_{C_j, D_{ij}|\mathcal{F}_{\{i\}}}\left(C_m, D_{lm}|\mathcal{F}_{\{i\}}\right)\right\} + O\left(h_N\right)\right]. \end{pmatrix},$$

$$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,j,m\}}\right] = \begin{pmatrix} \frac{1}{h_N^{d_C}} K_C\left(\frac{C_j - C_m}{h}\right) Y_{ij}\left[\frac{\partial}{\partial B_i} f_{D_{lm}, B_l|\mathcal{F}_{\{m\}}}\left(D_{ij}, B_i|\mathcal{F}_{\{m\}}\right) + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_C+1}} \dot{K}_C\left(\frac{C_j - C_m}{h}\right) Y_{ij}\left[f_{D_{lm}, B_l|\mathcal{F}_{\{m\}}}\left(D_{ij}, B_i|\mathcal{F}_{\{m\}}\right) + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_C}} K_C\left(\frac{C_j - C_m}{h}\right) Y_{ij}\left[\frac{\partial}{\partial D_{ij}} f_{D_{lm}, B_l|\mathcal{F}_{\{m\}}}\left(D_{ij}, B_i|\mathcal{F}_{\{m\}}\right) + O\left(h_N\right)\right] \end{pmatrix},$$

$$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,j,l\}}\right] = \begin{pmatrix} \frac{1}{h_N^{d_B+1}} \dot{K}_B\left(\frac{B_i - B_l}{h}\right) Y_{ij}\left[f_{D_{lm}, C_m|\mathcal{F}_{\{l\}}}\left(D_{ij}, C_j|\mathcal{F}_{\{l\}}\right) + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_B}} K_B\left(\frac{B_i - B_l}{h}\right) Y_{ij}\left[\frac{\partial}{\partial C_m} f_{D_{lm}, C_m|\mathcal{F}_{\{l\}}}\left(D_{ij}, C_j|\mathcal{F}_{\{l\}}\right) + O\left(h_N\right)\right] \\ \frac{1}{h_N^{d_B}} K_B\left(\frac{B_i - B_l}{h}\right) Y_{ij}\left[\frac{\partial}{\partial D_{lm}} f_{D_{lm}, C_m|\mathcal{F}_{\{l\}}}\left(D_{ij}, C_j|\mathcal{F}_{\{l\}}\right) + O\left(h_N\right)\right] \end{pmatrix}.$$

(b) $\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,j\}}\right] = Y_{ij}\frac{\partial}{\partial X_{ij}} f(X_{ij}) + O\left(h_N^Q\right)$,

$$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,l\}}\right] = \begin{pmatrix} \frac{1}{h_N^{d_B+1}} \dot{K}_B\left(\frac{B_i - B_l}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|C_j - C_m = 0, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{i,l\}}\right] f_{C_j - C_m, D_{ij} - D_{lm}|\mathcal{F}_{il}}(0, 0|\mathcal{F}_{\{i,l\}}) + O(h_N)\right] \\ \frac{1}{h_N^{d_B}} K_B\left(\frac{B_i - B_l}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|C_j - C_m = t, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{i,l\}}\right] f_{C_j - C_m, D_{ij} - D_{lm}|\mathcal{F}_{\{i,l\}}}(t, 0|\mathcal{F}_{\{i,l\}}) + O(h_N)\right] \\ \frac{1}{h_N^{d_B}} K_B\left(\frac{B_i - B_l}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|C_j - C_m = 0, D_{ij} - D_{lm} = t, \mathcal{F}_{\{i,l\}}\right] f_{C_j - C_m, D_{\{i,l\}} - D_{lm}|\mathcal{F}_{\{i,l\}}}(0, t|\mathcal{F}_{\{i,l\}}) + O(h_N)\right] \end{pmatrix},$$

$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,m\}}\right] = -\frac{\partial}{\partial t}\Big|_{t=0}\left\{\mathbb{E}\left[Y_{ij}|X_{ij} - X_{lm} = t, \mathcal{F}_{\{i,m\}}\right] f_{X_{ij} - X_{lm}|\mathcal{F}_{\{i,m\}}}\left(t|\mathcal{F}_{\{i,m\}}\right)\right\} + O\left(h_N\right)$,

$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j,l\}}\right] = -\frac{\partial}{\partial t}\Big|_{t=0}\left\{\mathbb{E}\left[Y_{ij}|X_{ij} - X_{lm} = t, \mathcal{F}_{\{j,l\}}\right] f_{X_{ij} - X_{lm}|\mathcal{F}_{\{j,l\}}}\left(t|\mathcal{F}_{\{j,l\}}\right)\right\} + O\left(h_N\right).$,

$$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j,m\}}\right] = \begin{pmatrix} \frac{1}{h_N^{d_C}} K_C\left(\frac{C_j - C_m}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|B_i - B_l = t, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{j,m\}}\right] f_{B_i - B_l, D_{ij} - D_{lm}|\mathcal{F}_{\{j,m\}}}(t, 0|\mathcal{F}_{\{j,m\}}) + O(h_N)\right] \\ \frac{1}{h_N^{d_C+1}} \dot{K}_C\left(\frac{C_j - C_m}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|B_i - B_l = 0, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{j,m\}}\right] f_{B_i - B_l, D_{ij} - D_{lm}|\mathcal{F}_{\{j,m\}}}(0, 0|\mathcal{F}_{\{j,m\}}) + O(h_N)\right] \\ \frac{1}{h_N^{d_C}} K_C\left(\frac{C_j - C_m}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|B_i - B_l = 0, D_{ij} - D_{lm} = t, \mathcal{F}_{\{j,m\}}\right] f_{B_i - B_l, D_{ij} - D_{lm}|\mathcal{F}_{\{j,m\}}}(0, t|\mathcal{F}_{\{j,m\}}) + O(h_N)\right] \end{pmatrix},$$

$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{l,m\}}\right] = -\frac{\partial}{\partial X_{lm}} g f(X_{lm}) + O\left(h_N^Q\right)$.

(c) $\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i\}}\right] = \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}f(X_{ij})\middle|\mathcal{F}_{\{i\}}\right] + O\left(h_N^Q\right)$,
$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j\}}\right] = \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}f(X_{ij})\middle|\mathcal{F}_{\{j\}}\right] + O\left(h_N^Q\right)$,
$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{l\}}\right] = \mathbb{E}\left[-\frac{\partial}{\partial X_{lm}}gf\left(X_{lm}\right)\middle|\mathcal{F}_{\{l\}}\right] + O\left(h_N^Q\right)$,
$\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{m\}}\right] = \mathbb{E}\left[-\frac{\partial}{\partial X_{lm}}gf\left(X_{lm}\right)\middle|\mathcal{F}_{\{m\}}\right] + O\left(h_N^Q\right)$.

Proof. (a)

$$
\begin{aligned}
&\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,j,m\}}\right] \\
&= Y_{ij}\mathbb{E}\left[\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,j,m\}}\right] \\
&= Y_{ij}\mathbb{E}\left[\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K_B\left(\frac{B_i-B_l}{h}\right)K_C\left(\frac{C_j-C_m}{h}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h}\right)\middle|\mathcal{F}_{\{i,j,m\}}\right] \\
&= Y_{ij}\frac{\partial}{\partial X_{ij}}\Bigg\{\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h}\right) \\
&\qquad\qquad \cdot\mathbb{E}\left[\frac{1}{h_N^{d_B+d_D}}K_B\left(\frac{B_i-B_l}{h}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h}\right)\middle|\mathcal{F}_{\{i,j,m\}}\right]\Bigg\} \\
&= Y_{ij}\frac{\partial}{\partial X_{ij}}\left\{\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h}\right)\left[f_{D_{lm},B_l|\mathcal{F}_m}\left(D_{ij},B_i|\mathcal{F}_{\{m\}}\right)+O\left(h_N\right)\right]\right\} \\
&= \begin{pmatrix}
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h}\right)Y_{ij}\left[\frac{\partial}{\partial B_i}f_{D_{lm},B_l|\mathcal{F}_m}\left(D_{ij},B_i|\mathcal{F}_{\{m\}}\right)+O\left(h_N\right)\right] \\
\frac{1}{h_N^{d_C+1}}\dot{K}_C\left(\frac{C_j-C_m}{h}\right)Y_{ij}\left[f_{D_{lm},B_l|\mathcal{F}_m}\left(D_{ij},B_i|\mathcal{F}_{\{m\}}\right)+O\left(h_N\right)\right] \\
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h}\right)Y_{ij}\left[\frac{\partial}{\partial D_{ij}}f_{D_{lm},B_l|\mathcal{F}_m}\left(D_{ij},B_i|\mathcal{F}_{\{m\}}\right)+O\left(h_N\right)\right]
\end{pmatrix}.
\end{aligned}
$$

Similarly,

$$
\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,j,l\}}\right] = \begin{pmatrix}
\frac{1}{h_N^{d_B+1}}\dot{K}_B\left(\frac{B_i-B_l}{h}\right)Y_{ij}\left[f_{D_{lm},C_m|\mathcal{F}_l}\left(D_{ij},C_j|\mathcal{F}_{\{l\}}\right)+O\left(h_N\right)\right] \\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h}\right)Y_{ij}\left[\frac{\partial}{\partial C_m}f_{D_{lm},C_m|\mathcal{F}_l}\left(D_{ij},C_j|\mathcal{F}_{\{l\}}\right)+O\left(h_N\right)\right] \\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h}\right)Y_{ij}\left[\frac{\partial}{\partial D_{lm}}f_{D_{lm},C_m|\mathcal{F}_l}\left(D_{ij},C_j|\mathcal{F}_{\{l\}}\right)+O\left(h_N\right)\right]
\end{pmatrix}.
$$

$$
\begin{aligned}
&\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j,l,m\}}\right]\\
&=\mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,l,m\}}\right]\\
&=\begin{pmatrix}
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_B+d_D}}\frac{\partial}{\partial B_i}K_B\left(\frac{B_i-B_l}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,l,m\}}\right]\\
\frac{1}{h_N^{d_C+1}}\dot{K}_C\left(\frac{C_j-C_m}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_B+d_D}}K_B\left(\frac{B_i-B_l}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,l,m\}}\right]\\
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_B+d_D}}K_B\left(\frac{B_i-B_l}{h_N}\right)\frac{\partial}{\partial D_{ij}}K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,l,m\}}\right]
\end{pmatrix}\\
&=\begin{pmatrix}
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h_N}\right)\left[-\frac{\partial}{\partial B_l}\left\{\mathbb{E}\left[Y_{ij}|B_i=B_l,D_{ij}=D_{lm},\mathcal{F}_{\{j\}}\right]f_{B_i,D_{ij}|\mathcal{F}_{\{j\}}}\left(B_l,D_{lm}|\mathcal{F}_{\{j\}}\right)\right\}+O\left(h_N\right)\right]\\
\frac{1}{h_N^{d_C+1}}\dot{K}_C\left(\frac{C_j-C_m}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|B_i=B_l,D_{ij}=D_{lm},\mathcal{F}_{\{j\}}\right]f_{B_i,D_{ij}|\mathcal{F}_{\{j\}}}\left(B_l,D_{lm}|\mathcal{F}_{\{j\}}\right)+O\left(h_N\right)\right]\\
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h_N}\right)\left[-\frac{\partial}{\partial D_{lm}}\left\{\mathbb{E}\left[Y_{ij}|B_i=B_l,D_{ij}=D_{lm},\mathcal{F}_{\{j\}}\right]f_{B_i,D_{ij}|\mathcal{F}_{\{j\}}}\left(B_l,D_{lm}|\mathcal{F}_{\{j\}}\right)\right\}+O\left(h_N\right)\right]
\end{pmatrix}.
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,l,m\}}\right]\\
&=\begin{pmatrix}
\frac{1}{h_N^{d_B+1}}\dot{K}_B\left(\frac{B_i-B_l}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|C_j=C_m,D_{ij}=D_{lm},\mathcal{F}_{\{i\}}\right]f_{C_j,D_{ij}|\mathcal{F}_{\{i\}}}\left(C_m,D_{lm}|\mathcal{F}_{\{i\}}\right)+O\left(h_N\right)\right]\\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h_N}\right)\left[-\frac{\partial}{\partial C_m}\left\{\mathbb{E}\left[Y_{ij}|C_j=C_m,D_{ij}=D_{lm},\mathcal{F}_{\{i\}}\right]f_{C_j,D_{ij}|\mathcal{F}_{\{i\}}}\left(C_m,D_{lm}|\mathcal{F}_{\{i\}}\right)\right\}+O\left(h_N\right)\right]\\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h_N}\right)\left[-\frac{\partial}{\partial D_{lm}}\left\{\mathbb{E}\left[Y_{ij}|C_j=C_m,D_{ij}=D_{lm},\mathcal{F}_{\{i\}}\right]f_{C_j,D_{ij}|\mathcal{F}_{\{i\}}}\left(C_m,D_{lm}|\mathcal{F}_{\{i\}}\right)\right\}+O\left(h_N\right)\right]
\end{pmatrix}.
\end{aligned}
$$

(b)

$$
\begin{aligned}
\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,j\}}\right]&=\mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,j\}}\right]\\
&=Y_{ij}\mathbb{E}\left[\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,j\}}\right]\\
&=Y_{ij}\frac{\partial}{\partial X_{ij}}f(X_{ij})+O\left(h_N^{Q}\right).
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,l\}}\right] \\
&= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,l\}}\right] \\
&= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K_B\left(\frac{B_i-B_l}{h_N}\right)K_C\left(\frac{C_j-C_m}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,l\}}\right] \\
&= \begin{pmatrix}
\frac{1}{h_N^{d_B+1}}\dot{K}_B\left(\frac{B_i-B_l}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_C+d_D}}K_C\left(\frac{C_j-C_m}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,l\}}\right] \\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_C+d_D}}\frac{\partial}{\partial C_j}K_C\left(\frac{C_j-C_m}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,l\}}\right] \\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_C+d_D}}K_C\left(\frac{C_j-C_m}{h_N}\right)\frac{\partial}{\partial D_{ij}}K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,l\}}\right]
\end{pmatrix} \\
&= \begin{pmatrix}
\frac{1}{h_N^{d_B+1}}\dot{K}_B\left(\frac{B_i-B_l}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|C_j-C_m=0,D_{ij}-D_{lm}=0,\mathcal{F}_{\{i,l\}}\right]f_{C_j-C_m,D_{ij}-D_{lm}|\mathcal{F}_{il}}(0,0|\mathcal{F}_{\{i,l\}})+O(h_N)\right] \\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|C_j-C_m=t,D_{ij}-D_{lm}=0,\mathcal{F}_{\{i,l\}}\right]f_{C_j-C_m,D_{ij}-D_{lm}|\mathcal{F}_{\{i,l\}}}(t,0|\mathcal{F}_{\{i,l\}})+O(h_N)\right] \\
\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_l}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|C_j-C_m=0,D_{ij}-D_{lm}=t,\mathcal{F}_{\{i,l\}}\right]f_{C_j-C_m,D_{\{i,l\}}-D_{lm}|\mathcal{F}_{\{i,l\}}}(0,t|\mathcal{F}_{\{i,l\}})+O(h_N)\right]
\end{pmatrix}
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j,m\}}\right] \\
&= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,m\}}\right] \\
&= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_x}}K_B\left(\frac{B_i-B_l}{h_N}\right)K_C\left(\frac{C_j-C_m}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,m\}}\right] \\
&= \begin{pmatrix}
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_B+d_D}}\frac{\partial}{\partial B_i}K_B\left(\frac{B_i-B_l}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,m\}}\right] \\
\frac{1}{h_N^{d_C+1}}\dot{K}_C\left(\frac{C_j-C_m}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_B+d_D}}K_B\left(\frac{B_i-B_l}{h_N}\right)K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,m\}}\right] \\
\frac{1}{h_N^{d_C}}K_B\left(\frac{C_j-C_m}{h_N}\right)\mathbb{E}\left[Y_{ij}\frac{1}{h_N^{d_B+d_D}}K_B\left(\frac{B_i-B_l}{h_N}\right)\frac{\partial}{\partial D_{ij}}K_D\left(\frac{D_{ij}-D_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,m\}}\right]
\end{pmatrix} \\
&= \begin{pmatrix}
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|B_i-B_l=t,D_{ij}-D_{lm}=0,\mathcal{F}_{\{j,m\}}\right]f_{B_i-B_l,D_{ij}-D_{lm}|\mathcal{F}_{\{j,m\}}}(t,0|\mathcal{F}_{\{j,m\}})+O(h_N)\right] \\
\frac{1}{h_N^{d_C+1}}\dot{K}_C\left(\frac{C_j-C_m}{h_N}\right)\left[\mathbb{E}\left[Y_{ij}|B_i-B_l=0,D_{ij}-D_{lm}=0,\mathcal{F}_{\{j,m\}}\right]f_{B_i-B_l,D_{ij}-D_{lm}|\mathcal{F}_{\{j,m\}}}(0,0|\mathcal{F}_{\{j,m\}})+O(h_N)\right] \\
\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_m}{h_N}\right)\left[-\frac{\partial}{\partial t}|_{t=0}\mathbb{E}\left[Y_{ij}|B_i-B_l=0,D_{ij}-D_{lm}=t,\mathcal{F}_{\{j,m\}}\right]f_{B_i-B_l,D_{ij}-D_{lm}|\mathcal{F}_{\{j,m\}}}(0,t|\mathcal{F}_{\{j,m\}})+O(h_N)\right]
\end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i,m\}}\right] &= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,m\}}\right] \\
&= -\frac{\partial}{\partial t}\bigg|_{t=0}\left\{\mathbb{E}\left[Y_{ij}|X_{ij}-X_{lm}=t,\mathcal{F}_{\{i,m\}}\right]f_{X_{ij}-X_{lm}|\mathcal{F}_{\{i,m\}}}\left(t|\mathcal{F}_{\{i,m\}}\right)\right\}+O\left(h_N\right).
\end{aligned}
$$

$$
\begin{aligned}
&\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j,l\}}\right] \\
&= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{j,l\}}\right] \\
&= -\frac{\partial}{\partial t}\bigg|_{t=0}\left\{\mathbb{E}\left[Y_{ij}|X_{ij}-X_{lm}=t,\mathcal{F}_{\{j,l\}}\right]f_{X_{ij}-X_{lm}|\mathcal{F}_{\{j,l\}}}\left(t|\mathcal{F}_{\{j,l\}}\right)\right\}+O\left(h_N\right).
\end{aligned}
$$

$$\begin{aligned}
\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{l,m\}}\right] &= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{l,m\}}\right] \\
&= \int \frac{1}{h_N^{d_X}}\frac{\partial}{\partial x}K\left(\frac{x-X_{lm}}{h_N}\right)gf(x)dx \\
&= -\frac{\partial}{\partial X_{lm}}gf(X_{lm}) + O\left(h_N^Q\right).
\end{aligned}$$

(c)

$$\begin{aligned}
\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{i\}}\right] &= \mathbb{E}\left[\mathbb{E}\left(Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{i,j\}}\right)\middle|\mathcal{F}_{\{i\}}\right] \\
&= \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}f(X_{ij})\middle|\mathcal{F}_{\{i\}}\right] + O\left(h_N^Q\right).
\end{aligned}$$

Similarly, $\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{j\}}\right] = \mathbb{E}\left[Y_{ij}\frac{\partial}{\partial X_{ij}}f(X_{ij})\middle|\mathcal{F}_{\{j\}}\right] + O\left(h_N^Q\right)$.

$$\begin{aligned}
\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{l\}}\right] &= \mathbb{E}\left[\mathbb{E}\left(Y_{ij}\frac{\partial}{\partial X_{ij}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)\middle|\mathcal{F}_{\{l,m\}}\right)\middle|\mathcal{F}_{\{l\}}\right] \\
&= \mathbb{E}\left[-\frac{\partial}{\partial X_{lm}}gf\left(X_{lm}\right)\middle|\mathcal{F}_{\{l\}}\right] + O\left(h_N^Q\right).
\end{aligned}$$

Similarly, $\mathbb{E}\left[\delta_{ijlm}(h_N)\middle|\mathcal{F}_{\{m\}}\right] = \mathbb{E}\left[-\frac{\partial}{\partial X_{lm}}gf\left(X_{lm}\right)\middle|\mathcal{F}_{\{m\}}\right] + O\left(h_N^Q\right)$.

□

Lemma (3.6.1) directly implies the following lemma (3.6.2).

Lemma 3.6.2. *Under Assumption 3.4.1, 3.4.2, and* $h_N \to 0$,

(a) $\mathbb{E}\left(p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}\right) = \frac{1}{4}\eta_i + O\left(h_N^Q\right)$, *in which*

$\eta_i = \mathbb{E}\left[-\frac{\partial}{\partial X_{ij}}gf(X_{ij}) - \frac{\partial}{\partial X_{ji}}gf(X_{ji}) + Y_{ij}\frac{\partial}{\partial X_{ij}}f\left(X_{ij}\right) + Y_{ji}\frac{\partial}{\partial X_{ji}}f\left(X_{ji}\right)\middle|\mathcal{F}_{\{i\}}\right]$

(b)

$$
\begin{aligned}
&\mathbb{E}\left[12p_{ijlm}\left(h_N\right)|\mathcal{F}_{\{i,j\}}\right] \\
=&Y_{ij}\frac{\partial}{\partial X_{ij}}f(X_{ij})+Y_{ji}\frac{\partial}{\partial X_{ji}}f(X_{ji})-\frac{\partial}{\partial X_{ij}}gf(X_{ij})-\frac{\partial}{\partial X_{ji}}gf(X_{ji})+O(h_N) \\
&-\left.\frac{\partial}{\partial t}\right|_{t=0}\left\{\mathbb{E}\left[Y_{il}|X_{il}-X_{mj}=t,\mathcal{F}_{\{i,j\}}\right]f_{X_{il}-X_{mj}|\mathcal{F}_{\{i,j\}}}\left(t|\mathcal{F}_{\{i,j\}}\right)\right\} \\
&-\left.\frac{\partial}{\partial t}\right|_{t=0}\left\{\mathbb{E}\left[Y_{jl}|X_{jl}-X_{mi}=t,\mathcal{F}_{\{i,j\}}\right]f_{X_{jl}-X_{mi}|\mathcal{F}_{\{i,j\}}}\left(t|\mathcal{F}_{\{i,j\}}\right)\right\} \\
&-\left.\frac{\partial}{\partial t}\right|_{t=0}\left\{\mathbb{E}\left[Y_{lj}|X_{lj}-X_{im}=t,\mathcal{F}_{\{i,j\}}\right]f_{X_{lj}-X_{im}|\mathcal{F}_{\{i,j\}}}\left(t|\mathcal{F}_{\{i,j\}}\right)\right\} \\
&-\left.\frac{\partial}{\partial t}\right|_{t=0}\left\{\mathbb{E}\left[Y_{li}|X_{li}-X_{jm}=t,\mathcal{F}_{\{i,j\}}\right]f_{X_{li}-X_{jm}|\mathcal{F}_{\{i,j\}}}\left(t|\mathcal{F}_{\{i,j\}}\right)\right\} \\
&+\begin{pmatrix}
\frac{1}{h_N^{d_B+1}}\dot{K}_B\left(\frac{B_i-B_j}{h_N}\right)\left\{\mathbb{E}\left[Y_{il}|C_l-C_m=0,D_{il}-D_{jm}=0,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{il}-D_{jm}|\mathcal{F}_{\{i,j\}}}\left(0,0|\mathcal{F}_{\{i,j\}}\right)+O(h_N)\right\} \\
-\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_j}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{il}|C_l-C_m=t,D_{il}-D_{jm}=0,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{il}-D_{jm}|\mathcal{F}_{\{i,j\}}}\left(t,0|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right] \\
-\frac{1}{h_N^{d_B}}K_B\left(\frac{B_i-B_j}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{il}|C_l-C_m=0,D_{il}-D_{jm}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{il}-D_{jm}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right]
\end{pmatrix} \\
&+\begin{pmatrix}
\frac{1}{h_N^{d_B+1}}\dot{K}_B\left(\frac{B_j-B_i}{h_N}\right)\left\{\mathbb{E}\left[Y_{jl}|C_l-C_m=0,D_{jl}-D_{im}=0,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{jl}-D_{im}|\mathcal{F}_{\{i,j\}}}\left(0,0|\mathcal{F}_{\{i,j\}}\right)+O(h_N)\right\} \\
-\frac{1}{h_N^{d_B}}K_B\left(\frac{B_j-B_i}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{jl}|C_l-C_m=t,D_{jl}-D_{im}=0,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{jl}-D_{im}|\mathcal{F}_{\{i,j\}}}\left(t,0|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right] \\
-\frac{1}{h_N^{d_B}}K_B\left(\frac{B_j-B_i}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{jl}|C_l-C_m=0,D_{jl}-D_{im}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{jl}-D_{im}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right]
\end{pmatrix} \\
&+\begin{pmatrix}
-\frac{1}{h_N^{d_C}}K_C\left(\frac{C_i-C_j}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{li}|B_l-B_m=t,D_{li}-D_{mj}=0,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{li}-D_{mj}|\mathcal{F}_{\{i,j\}}}\left(t,0|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right] \\
\frac{1}{h_N^{d_C+1}}\dot{K}_C\left(\frac{C_i-C_j}{h_N}\right)\left\{\mathbb{E}\left[Y_{li}|B_l-B_m=0,D_{li}-D_{mj}=0,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{li}-D_{mj}|\mathcal{F}_{\{i,j\}}}\left(0,0|\mathcal{F}_{\{i,j\}}\right)+O(h_N)\right\} \\
-\frac{1}{h_N^{d_C}}K_C\left(\frac{C_i-C_j}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{li}|B_l-B_m=0,D_{li}-D_{mj}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{li}-D_{mj}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right]
\end{pmatrix} \\
&+\begin{pmatrix}
-\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_i}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{lj}|B_l-B_m=t,D_{lj}-D_{mi}=0,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{lj}-D_{mi}|\mathcal{F}_{\{i,j\}}}\left(t,0|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right] \\
\frac{1}{h_N^{d_C+1}}\dot{K}_C\left(\frac{C_j-C_i}{h_N}\right)\left\{\mathbb{E}\left[Y_{lj}|B_l-B_m=0,D_{lj}-D_{mi}=0,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{lj}-D_{mi}|\mathcal{F}_{\{i,j\}}}\left(0,0|\mathcal{F}_{\{i,j\}}\right)+O(h_N)\right\} \\
-\frac{1}{h_N^{d_C}}K_C\left(\frac{C_j-C_i}{h_N}\right)\left[\left.\frac{\partial}{\partial t}\right|_0\left\{\mathbb{E}\left[Y_{lj}|B_l-B_m=t,D_{lj}-D_{mi}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{lj}-D_{mi}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}+O(h_N)\right]
\end{pmatrix}.
\end{aligned}
$$

Proof. These are direct results of lemma (3.6.1) and the definition of $p_{ijlm}(h_N)$ □

Lemma 3.6.3. *Under Assumption 3.4.1, 3.4.2, and $h_N \to 0$,*

(a) $\mathbb{E}\left[\mathbb{E}\left(p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}\right)^2\right]=\frac{1}{16}\Sigma+O\left(h_N^Q\right)$, *in which*

$$\Sigma=\mathrm{Var}\left(\mathbb{E}\left[-\frac{\partial}{\partial X_{ij}}gf(X_{ij})-\frac{\partial}{\partial X_{ji}}gf(X_{ji})+Y_{ij}\frac{\partial}{\partial X_{ij}}f\left(X_{ij}\right)+Y_{ji}\frac{\partial}{\partial X_{ji}}f\left(X_{ji}\right)\middle|\mathcal{F}_{\{i\}}\right]\right)$$

(b) *For $d_B=d_C$,*

$$\mathrm{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix}\mathbb{E}\left[p_{ijlm}(h_N)|\mathcal{F}_{\{i,j\}}\right]\right)=\frac{1}{36}\Omega+O(h_N).$$

in which $\Omega = \begin{pmatrix} \Omega_{BB} & \Omega_{BC} & \Omega_{BD} \\ \Omega_{BC}^T & \Omega_{CC} & \Omega_{CD} \\ \Omega_{BD}^T & \Omega_{CD}^T & \Omega_{DD} \end{pmatrix}$.

$$\Omega = \frac{1}{4} \lim_{N\to\infty} \frac{1}{h_N^{d_B}} \operatorname{Var}\left(\begin{pmatrix} R_B \\ R_C \\ R_D \end{pmatrix} \right),$$

in which

$$\begin{aligned}
R_B = \dot{K}_B \left(\frac{B_i - B_j}{h_N} \right) \Big\{ & \mathbb{E}\left[Y_{il} | C_l - C_m = 0, D_{il} - D_{jm} = 0, \mathcal{F}_{\{i,j\}} \right] f_{C_l - C_m, D_{il} - D_{jm} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) \\
& - \mathbb{E}\left[Y_{jl} | C_l - C_m = 0, D_{jl} - D_{im} = 0, \mathcal{F}_{\{i,j\}} \right] f_{C_l - C_m, D_{jl} - D_{im} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) + O(h_N) \Big\} \\
R_C = \dot{K}_C \left(\frac{C_i - C_j}{h_N} \right) \Big\{ & \mathbb{E}\left[Y_{li} | B_l - B_m = 0, D_{li} - D_{mj} = 0, \mathcal{F}_{\{i,j\}} \right] f_{B_l - B_m, D_{li} - D_{mj} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) \\
& - \mathbb{E}\left[Y_{lj} | B_l - B_m = 0, D_{lj} - D_{mi} = 0, \mathcal{F}_{\{i,j\}} \right] f_{B_l - B_m, D_{lj} - D_{mi} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) + O(h_N) \Big\} \\
R_D = -K_B \left(\frac{B_i - B_j}{h_N} \right) & \left[\frac{\partial}{\partial t}\Big|_0 \Big\{ \mathbb{E}\left[Y_{il} | C_l - C_m = 0, D_{il} - D_{jm} = t, \mathcal{F}_{\{i,j\}} \right] f_{C_l - C_m, D_{il} - D_{jm} | \mathcal{F}_{\{i,j\}}} \left(0, t | \mathcal{F}_{\{i,j\}} \right) \Big\} \right. \\
& \left. + \frac{\partial}{\partial t}\Big|_0 \Big\{ \mathbb{E}\left[Y_{jl} | C_l - C_m = 0, D_{jl} - D_{im} = t, \mathcal{F}_{\{i,j\}} \right] f_{C_l - C_m, D_{jl} - D_{im} | \mathcal{F}_{\{i,j\}}} \left(0, t | \mathcal{F}_{\{i,j\}} \right) \Big\} + O(h_N) \right] \\
- K_C \left(\frac{C_i - C_j}{h_N} \right) & \left[\frac{\partial}{\partial t}\Big|_0 \Big\{ \mathbb{E}\left[Y_{li} | B_l - B_m = 0, D_{li} - D_{mj} = t, \mathcal{F}_{\{i,j\}} \right] f_{B_l - B_m, D_{li} - D_{mj} | \mathcal{F}_{\{i,j\}}} \left(0, t | \mathcal{F}_{\{i,j\}} \right) \Big\} \right. \\
& \left. + \frac{\partial}{\partial t}\Big|_0 \Big\{ \mathbb{E}\left[Y_{lj} | B_l - B_m = t, D_{lj} - D_{mi} = t, \mathcal{F}_{\{i,j\}} \right] f_{B_l - B_m, D_{lj} - D_{mi} | \mathcal{F}_{\{i,j\}}} \left(0, t | \mathcal{F}_{\{i,j\}} \right) \Big\} + O(h_N) \right]
\end{aligned}$$

$$\begin{aligned}
\Omega_{BB} = \frac{1}{4} & \int \dot{K}_B(b) \dot{K}_B^\top(b)\, db \\
& \cdot \mathbb{E} \Big\{ \Big[\mathbb{E}\left[Y_{il} | C_l - C_m = 0, D_{il} - D_{jm} = 0, \mathcal{F}_{\{i,j\}} \right] f_{C_l - C_m, D_{il} - D_{jm} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) \\
& \qquad - \mathbb{E}\left[Y_{jl} | C_l - C_m = 0, D_{jl} - D_{im} = 0, \mathcal{F}_{\{i,j\}} \right] f_{C_l - C_m, D_{jl} - D_{im} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) \Big]^2 | B_i - B_j = 0 \Big\} \\
& \cdot f_{B_i - B_j}(0) + O(h_N)
\end{aligned}$$

$$\begin{aligned}
\Omega_{CC} = \frac{1}{4} & \int \dot{K}_C(c) \dot{K}_C^\top(c)\, dc \\
& \cdot \mathbb{E} \Big\{ \Big[\mathbb{E}\left[Y_{li} | B_l - B_m = 0, D_{li} - D_{mj} = 0, \mathcal{F}_{\{i,j\}} \right] f_{B_l - B_m, D_{li} - D_{mj} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) \\
& \qquad - \mathbb{E}\left[Y_{lj} | B_l - B_m = 0, D_{lj} - D_{mi} = 0, \mathcal{F}_{\{i,j\}} \right] f_{B_l - B_m, D_{lj} - D_{mi} | \mathcal{F}_{\{i,j\}}} \left(0, 0 | \mathcal{F}_{\{i,j\}} \right) \Big]^2 | C_i - C_j = 0 \Big\} \\
& \cdot f_{C_i - C_j}(0) + O(h_N)
\end{aligned}$$

$$
\begin{aligned}
&\Omega_{DD}\\
&=\lim_{N\to\infty}\frac{1}{4}\frac{1}{h_N^{d_B}}\operatorname{Var}\Bigg(-K_B\left(\frac{B_i-B_j}{h_N}\right)\left[\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{il}|C_l-C_m=0,D_{il}-D_{jm}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{il}-D_{jm}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right.\\
&\qquad\left.+\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{jl}|C_l-C_m=0,D_{jl}-D_{im}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{jl}-D_{im}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right]\\
&\qquad-K_C\left(\frac{C_i-C_j}{h_N}\right)\left[\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{li}|B_l-B_m=0,D_{li}-D_{mj}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{li}-D_{mj}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right.\\
&\qquad\left.+\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{lj}|B_l-B_m=t,D_{lj}-D_{mi}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{lj}-D_{mi}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right]\Bigg)\\
&=\frac{1}{4}\int K_B^2(b)db\,\mathbb{E}\Bigg(\left[\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{il}|C_l-C_m=0,D_{il}-D_{jm}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{il}-D_{jm}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right.\\
&\qquad\left.+\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{jl}|C_l-C_m=0,D_{jl}-D_{im}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{jl}-D_{im}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right]^2\Bigg|B_i-B_j=0\Bigg)\\
&\quad+\frac{1}{4}\int K_C^2(c)dc\,\mathbb{E}\Bigg(\left[\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{li}|B_l-B_m=0,D_{li}-D_{mj}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{li}-D_{mj}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right.\\
&\qquad\left.+\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{lj}|B_l-B_m=t,D_{lj}-D_{mi}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{lj}-D_{mi}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right]^2\Bigg|C_i-C_j=0\Bigg)\\
&\quad+\frac{1}{4}\lim_{N\to\infty}\frac{2}{h_N^{d_B}}\operatorname{Cov}\Bigg(-K_B\left(\frac{B_i-B_j}{h_N}\right)\left[\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{il}|C_l-C_m=0,D_{il}-D_{jm}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{il}-D_{jm}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right.\\
&\qquad\left.+\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{jl}|C_l-C_m=0,D_{jl}-D_{im}=t,\mathcal{F}_{\{i,j\}}\right]f_{C_l-C_m,D_{jl}-D_{im}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right],\\
&\qquad-K_C\left(\frac{C_i-C_j}{h_N}\right)\left[\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{li}|B_l-B_m=0,D_{li}-D_{mj}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{li}-D_{mj}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right.\\
&\qquad\left.+\frac{\partial}{\partial t}\bigg|_0\left\{\mathbb{E}\left[Y_{lj}|B_l-B_m=t,D_{lj}-D_{mi}=t,\mathcal{F}_{\{i,j\}}\right]f_{B_l-B_m,D_{lj}-D_{mi}|\mathcal{F}_{\{i,j\}}}\left(0,t|\mathcal{F}_{\{i,j\}}\right)\right\}\right]\Bigg).
\end{aligned}
$$

(c) $\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0\\ 0 & h_N^{\frac{d_B+2}{2}} & 0\\ 0 & 0 & h_N^{\frac{d_B}{2}}\end{pmatrix}\mathbb{E}\left[p_{N,ijlm}|\mathcal{F}_{\{i,j,l\}}\right]\right)=O\left(1\right),$

(d) $\operatorname{Var}\left(p_{ijlm}(h_N)\right)=O\left(h_N^{-(d_X+2)}\right).$

Proof. (a) This is a direct result of lemma (3.6.2)(a).

(b) The scaling is exactly making the variance asymptotically converging to a finite PSD matrix. To see this, focus on the first d_B components:

$$\frac{1}{h_N^{d_B}} \operatorname{Var}\left(\dot{K}_B\left(\frac{B_i - B_j}{h_N}\right)\left\{\mathbb{E}\left[Y_{il}|C_l - C_m = 0, D_{il} - D_{jm} = 0, \mathcal{F}_{\{i,j\}}\right] f_{C_l - C_m, D_{il} - D_{jm}|\mathcal{F}_{\{i,j\}}}\left(0, 0|\mathcal{F}_{\{i,j\}}\right)\right.\right.$$
$$\left.\left. - \mathbb{E}\left[Y_{jl}|C_l - C_m = 0, D_{jl} - D_{im} = 0, \mathcal{F}_{\{i,j\}}\right] f_{C_l - C_m, D_{jl} - D_{im}|\mathcal{F}_{\{i,j\}}}\left(0, 0|\mathcal{F}_{\{i,j\}}\right) + O(h_N)\right\}\right)$$
$$= \frac{1}{h_N^{d_B}}\mathbb{E}\Bigg(\dot{K}_B\left(\frac{B_i - B_j}{h_N}\right)\dot{K}_B^\top\left(\frac{B_i - B_j}{h_N}\right)$$
$$\left\{\mathbb{E}\left[Y_{il}|C_l - C_m = 0, D_{il} - D_{jm} = 0, \mathcal{F}_{\{i,j\}}\right] f_{C_l - C_m, D_{il} - D_{jm}|\mathcal{F}_{\{i,j\}}}\left(0, 0|\mathcal{F}_{\{i,j\}}\right)\right.$$
$$\left. - \mathbb{E}\left[Y_{jl}|C_l - C_m = 0, D_{jl} - D_{im} = 0, \mathcal{F}_{\{i,j\}}\right] f_{C_l - C_m, D_{jl} - D_{im}|\mathcal{F}_{\{i,j\}}}\left(0, 0|\mathcal{F}_{\{i,j\}}\right) + O(h_N)\right\}^2\Bigg)$$
$$+ O\left(h_N^{d_B+2}\right)$$
$$= \Omega_{BB} + o(h_N).$$

Calculation for the middle d_C components and the last d_D components are conceptually similar. The covariance term in the definition of Ω_{DD} is written as a limit because we know the covariance term is not exploding but pinning down its limit requires additional specification of the joint support conditions of $B_i - B_l, C_i - C_l$, which we don't need to get into.

(c) Integration by parts, change of variable, and bounding arguments give us the result.

(d) Integration by part together with change of variable give us the result.

$$\begin{aligned}
\mathbb{E}\left[\delta_{ijlm}(h_N)^2\right] &= \mathbb{E}\left[Y_{ij}^2 \frac{1}{h_N^{2d_X}}\left(\frac{\partial}{\partial X_{ij}} K\left(\frac{X_{ij} - X_{lm}}{h_N}\right)\right)^2\right] \\
&= \frac{1}{h_N^{2d_X}}\mathbb{E}\left[\mathbb{E}\left(Y_{ij}^2|X_{ij}\right)\mathbb{E}\left[\left(\frac{\partial}{\partial X_{ij}} K\left(\frac{X_{ij} - X_{lm}}{h_N}\right)\right)^2 \middle| X_{ij}\right]\right] \\
&= \frac{1}{h_N^{d_X+2}}\mathbb{E}\left[\mathbb{E}\left(Y_{ij}^2|X_{ij}\right)\int\left(\frac{\partial}{\partial u} K(u)\right)^2 f\left(X_{ij} - h_N u\right) du\right] \\
&= O\left(h_N^{-(d_X+2)}\right)
\end{aligned}$$
$$\mathbb{E}\left[p_{ijlm}(h_N)^2\right] \leq \mathbb{E}\left[\delta_{ijlm}(h_N)^2\right] = O\left(h_N^{-(d_X+2)}\right).$$

□

Lemma 3.6.4. *Under Assumption 3.4.1, 3.4.2, and $h_N \to 0$,*

(a) $\operatorname{Var}\left(q_{1,i}(h_N)^2\right) = \frac{1}{16}\Sigma + O\left(h_N^Q\right)$,

(b) $\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N)\right) = \frac{1}{36}\Omega + O\left(h_N\right),$

(c) $\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{3,ijl}(h_N)\right) = O(1),$

(d) $\operatorname{Var}\left(q_{4,ijlm}(h_N)\right) = O\left(h_N^{-(d_X+2)}\right).$

Proof. (a) $\operatorname{Var}\left(q_{1,i}(h_N)\right) = \operatorname{Var}\left(\mathbb{E}\left(p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}\right)\right) = \frac{1}{16}\Sigma + O\left(h_N^Q\right).$

(b)
$$\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N)\right) = \operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} \mathbb{E}\left(p_{ijlm}(h_N)|\mathcal{F}_{\{i,j\}}\right)\right) - 2\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} \mathbb{E}\left(p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}\right)\right)$$
$$= \frac{1}{36}\Omega + O(h_N).$$

(c) $\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{3,ijl}(h_N)\right) \leq \operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} \mathbb{E}\left(p_{ijlm}(h_N)|\mathcal{F}_{\{i,j,l\}}\right)\right) = O(1).$

(d) Similarly, $\operatorname{Var}\left(q_{4,ijlm}(h_N)\right) \leq \operatorname{Var}\left(p_{ijlm}(h_N)\right) = O\left(h_N^{-(d_X+2)}\right).$

□

Proof of lemma 3.5.3

Proof. These are results of lemma 3.6.4 together with the facts

$$\operatorname{Var}\left(U_{N,1}(h_N)\right) = N^{-1}\operatorname{Var}\left(q_{N,1,i}^2\right)$$

$$\operatorname{Var}\left(\sqrt{\binom{N}{2}}\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,2}(h_N)\right) = \operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N)\right)$$

$$\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,3}(h_N)\right) = \binom{N}{3}^{-1}\operatorname{Var}\left(q_{3,ijl}(h_N)\right)$$

$$\operatorname{Var}\left(U_{N,4}(h_N)\right) = \binom{N}{4}^{-1}\mathbb{E}\left(q_{N,4,ijlm}^2\right).$$

$\square$

Chapter 4

Error Components Models for Dyadic Data

4.1 Introduction and Summary

In this chapter, we study error components models of dyadic data, of which a major motivation is separating the monadic and dyadic components of variation. Our development parallels that of error components with panel data: we progressively enrich the random effect model by going from being without covariates to being with covariates and from homoskedasticity to multiplicative heteroskedasticity. Throughout enriching the models, we focus on estimation of the coefficients in a linear regression, which includes both monadic and dyadic explanatory variables. To understand the nature of the estimation problems under different error components models, we study the performance of intuitive OLS estimators, propose more efficient estimators, calculate the asymptotic efficiency bounds (Cramér-Rao lower bound, CRLB), and compare the efficiency bounds to the estimators we propose. A central theme of this dissertation is understanding estimation problems of non/semiparametric models for dyadic data. The linear model with homo/heteroskedasticity is an ideal playground for this themed exploration.

Under homoskedasticity, we prove the sample mean, which converges at rate $O(N^{-1/2})$, and least square estimator with double-differencing operation, which converges at rate $O\left(\binom{N}{2}^{-1/2}\right)$, achieve the CRLB and are asymptotically efficient for estimating the marginal expectation and the coefficients of dyadic variables in a linear regression respectively. Under unknown multiplicative heteroskedasticity, we show that the intuitive two-step semiparametric generalized score estimator for estimating the linear regression coefficients, which is a natural extension of the classical feasible generalized least square estimator (FGLS) for linear regression with heteroskedasticity for the "monadic" iid data, is not adaptive to the unknown heteroskedasticity. Its convergence rate is faster than that of the OLS estimator, $O(N^{-1/2})$, but slower than the rate suggested by CRLB, $O\left(\binom{N}{2}^{-1/2}\right)$. This result makes a distinction from a familiar result in the monadic iid setting, i.e. a two-step semiparametric generalized score estimator often indeed achieves adaptivity and CRLB in iid setting. The gap between the performance of the best available estimator and the CRLB suggests that for this estimation problem with dyadic data either there exists a better estimator that is adaptive and achieves the CRLB, or there is a tighter efficiency bound. We point this gap out for further research.

Compared to previous literature, one of our main contributions is extending the discussion of estimating parameters of linear regression from the homoskedasticity model to the multiplicative heteroskedasticity model. From a modeling perspective, we are moving the discussion from a parametric model to a semiparametric model. Another main contribution is calculating the efficiency bound under both homoskedasticity and heteroskedasticity models. Even though we do not have a formal theory for efficiency bound for dyadic data yet, the CRLB calculation is intuitive and valuable in its own right. A third contribution is that we propose and characterize the asymptotic behavior of tetrad-difference estimators, "fixed-effect" regression estimators, and a feasible two-step semiparametric generalized score estimator for the unknown heteroskedasticity model. A final contribution is pointing out the gap between CRLB and the asymptotic behavior of these proposed estimators.

4.2 A Variance Decomposition

We start from a simple variance-components model

$$\begin{aligned}
\text{Model: } & Y_{ij} = \mu + U_i + U_j + V_{ij}, 1 \le i < j \le N \\
& U_1, \ldots, U_N \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma_U^2\right), \quad V_{ij} \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma_V^2\right), \\
& \left(U_1, \ldots, U_N, V_{ij}, 1 \le i < j \le N\right) \text{ are independent} \\
& \mu,\ \sigma_A^2, \sigma_V^2 \text{ are unknown} \\
\text{Data: } & \left(Y_{ij}, 1 \le i < j \le N\right) \text{ only, } U_i \text{ are unobserved} \\
\text{Estimand: } & \mu, \sigma_U^2, \sigma_V^2.
\end{aligned}$$

Write the variables in matrix form

$$\mathbf{Y}_{\binom{N}{2}\times 1} = \mu \boldsymbol{\iota}_{\binom{N}{2}} + \mathbf{T}_{\binom{N}{2}\times N} \mathbf{U}_{N\times 1} + \mathbf{V}_{\binom{N}{2}\times 1},$$

in which the matrix collecting all the dummies is

$$\mathbf{T} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & 1 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 1 \end{pmatrix}_{\binom{N}{2}\times N} = (a_1, \ldots, a_N).$$

Conditionally

$$\mathbf{Y}|\mathbf{U} \sim \mathrm{N}\left(\mu\boldsymbol{\iota}_{\binom{N}{2}} + \mathbf{T}_{\binom{N}{2}\times N}\mathbf{U}_{N\times 1},\ \sigma_V^2 I_{\binom{N}{2}}\right).$$

Marginally

$$\mathbf{Y} \sim \mathrm{N}\left(\mu\boldsymbol{\iota}_{\binom{N}{2}},\ \Sigma = \sigma_V^2 I_{\binom{N}{2}} + \sigma_U^2 \mathbf{T}\mathbf{T}^\top\right).$$

The (unconditional) likelihood is

$$l\left(\mu, \sigma_U^2, \sigma_V^2; \mathbf{Y}\right) = -\frac{1}{2}\left(\mathbf{Y} - \mu\boldsymbol{\iota}_{\binom{N}{2}}\right)^\top \Sigma^{-1}\left(\mathbf{Y} - \mu\boldsymbol{\iota}_{\binom{N}{2}}\right) - \frac{1}{2}\ln(\det(\Sigma)) - \binom{N}{2}\frac{1}{2}\ln(2\pi).$$

The MLE is

$$\hat{\mu} = \bar{Y} = \binom{N}{2}^{-1}\sum_{i<j} Y_{ij}.$$

Notice that the $\hat{\mu}$ here coincides with the GLS with known Σ. Gauss-Markov theorem implies $\hat{\mu}$ is BLUE. It is also asymptotic efficient. $\sqrt{N}\left(\hat{\mu} - \mu\right) \rightsquigarrow \mathrm{N}\left(0, 4\sigma_U^2\right)$. The MLE for the variances $\hat{\sigma}_U^2,\ \hat{\sigma}_V^2$ seems reasonable but they are not our main focus.

4.3 Error Components Regression

Homoskedasticity

The regression version of the previous model with conditioning variables based on individual covariates is

$$\begin{aligned} \text{Model: } Y_{ij} &= Y\left(X_i, U_i, X_j, U_j, V_{ij}\right) = W_{ij}^\top \beta + \left(f_i + f_j\right)^\top \gamma + e_{ij}, 1 \le i < j \le N \\ e_{ij} &= U_i + U_j + V_{ij}, \end{aligned}$$

where the regressor $W_{ij} = W(X_i, X_j)$ for a known function $W : \mathbb{R}^{2d} \to \mathbb{R}^{p_W}$, $f_i = f(X_i)$ for a known function $f : \mathbb{R}^d \to \mathbb{R}^{p_f}$ and

$$\begin{aligned}
\text{Sampling assumption: } & X_1, \dots, X_N \overset{\text{iid}}{\sim} P_X \text{ known, } U_1, \dots, U_N \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma_U^2\right), \; V_{ij} \overset{\text{iid}}{\sim} \mathrm{N}\left(0, \sigma_V^2\right) \\
& \left(X_1, \dots, X_N, U_1, \dots, U_N, V_{ij}, 1 \le i < j \le N\right) \text{ are independent.} \\
\text{Primitives: } & \beta \in \mathbb{R}^{p_W}, \gamma \in \mathbb{R}^{p_f}, \sigma_U^2, \sigma_V^2 \in \mathbb{R}_+ \text{ are unknown} \\
& W, f, P_X \text{ are known} \\
\text{Data : } & \left(Y_{ij}, 1 \le i < j \le N, X_1, \dots, X_N\right) \text{ only, } U_i \text{ are unobserved} \\
\text{Estimand: } & \beta, \gamma, \sigma_U^2, \sigma_V^2.
\end{aligned}$$

Compared to the error-components model in the previous section, this regression model specifies a richer form of error components allowing for a potential correlation between the monadic effect and the dyadic effect.

Conditionally

$$\mathbf{Y}|\mathbf{X}, \mathbf{U} \sim \mathrm{N}\left(\mathbf{W}_{\binom{N}{2} \times p_W} \beta + \mathbf{T}_{\binom{N}{2} \times N} \mathbf{f}_{N \times p_f} \gamma + \mathbf{T}_{\binom{N}{2} \times N} \mathbf{U}_{N \times 1}, \; \sigma_V^2 I_{\binom{N}{2}}\right).$$

Marginally

$$\mathbf{Y}|\mathbf{X} \sim \mathrm{N}\left(\mathbf{W}_{\binom{N}{2} \times p_W} \beta + \mathbf{T}_{\binom{N}{2} \times N} \mathbf{f}_{N \times p_f} \gamma, \; \Sigma = \sigma_V^2 I_{\binom{N}{2}} + \sigma_U^2 \mathbf{T}\mathbf{T}^\top\right).$$

The (unconditional) likelihood is

$$\begin{aligned}
l\left(\beta, \gamma, \sigma_U^2, \sigma_V^2; \mathbf{Y}, \mathbf{X}\right) = & -\frac{1}{2}\left(\mathbf{Y} - \mathbf{W}\beta - \mathbf{T}\mathbf{f}\gamma\right)^\top \Sigma^{-1}\left(\mathbf{Y} - \mathbf{W}\beta - \mathbf{T}\mathbf{f}\gamma\right) \\
& - \frac{1}{2}\ln(\det(\Sigma)) - \binom{N}{2}\frac{1}{2}\ln(2\pi).
\end{aligned}$$

MLE

Here we report the MLE of the linear coefficients (β, γ). We are using $\mathbf{S}$ and $\mathbf{T}$ interchangeably here ($\mathbf{S} = \mathbf{T}$).

$$\begin{aligned}
\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} &= \left[\left(\mathbf{W} \vdots \mathbf{S}\mathbf{f}\right)^\top \hat{\Sigma}^{-1} \left(\mathbf{W} \vdots \mathbf{S}\mathbf{f}\right)\right]^{-1} \left(\mathbf{W} \vdots \mathbf{S}\mathbf{f}\right)^\top \hat{\Sigma}^{-1} Y \\
\hat{\Sigma} &= \hat{\sigma}_V^2 I_{\binom{N}{2}} + \hat{\sigma}_U^2 \mathbf{T}\mathbf{T}^\top.
\end{aligned}$$

The MLE in this type of model is not guaranteed to have good performance. It seems to perform reasonably in line with the intuition. It could be tedious but insightful to understand its performance. We are not writing down full explicit analytical expressions here.

OLS in Levels

Here we report the OLS (in levels) estimator of (β, γ) and outline its asymptotic properties.

$$\begin{aligned}
\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} &= \left[\left(\mathbf{W} \vdots \mathbf{Sf} \right)^\top \left(\mathbf{W} \vdots \mathbf{Sf} \right) \right]^{-1} \left(\mathbf{W} \vdots \mathbf{Sf} \right)^\top \mathbf{Y} \\
\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} &= \begin{pmatrix} \mathbf{W}^\top \mathbf{W} & \mathbf{W}^\top \mathbf{Sf} \\ (\mathbf{Sf})^\top \mathbf{W} & (\mathbf{Sf})^\top \mathbf{Sf} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{W}^\top \\ (\mathbf{Sf})^\top \end{pmatrix} \mathbf{e} \\
&= \left[\sum_{i<j} \begin{pmatrix} W_{ij} W_{ij}^\top & W_{ij}(f_i + f_j)^\top \\ (f_i + f_j) W_{ij}^\top & (f_i + f_j)(f_i + f_j)^\top \end{pmatrix} \right]^{-1} \sum_{i<j} \begin{pmatrix} W_{ij} \\ (f_i + f_j) \end{pmatrix} e_{ij}
\end{aligned}$$

In the special case where we know γ, the infeasible OLS of a scalar β has

$$\begin{aligned}
&\sqrt{N} \left(\hat{\beta} - \beta \right) \\
&= \left[\binom{N}{2}^{-1} \sum_{i<j} W_{ij}^2 \right]^{-1} \sqrt{N} \binom{N}{2}^{-1} \sum_{i<j} W_{ij} e_{ij} \rightsquigarrow \mathrm{N} \left(0, (\mathbb{E} W_{12}^2)^{-1} 4 \operatorname{Var} \left(P_{\{1\}} W_{12} \right) \sigma_U^2 \right).
\end{aligned}$$

Here $P_{\{1\}} W_{12} := \mathbb{E}[W_{12} | X_1] - \mathbb{E} W_{12}$ is the Hajek projection of W_{12} on index 1. In the special case where we know β, the infeasible OLS of a scalar γ has

$$\begin{aligned}
\sqrt{N} \left(\hat{\gamma} - \gamma \right) &= \left[\binom{N}{2}^{-1} \sum_{i<j} (f_i + f_j)^2 \right]^{-1} \sqrt{N} \binom{N}{2}^{-1} \sum_{i<j} (f_i + f_j) e_{ij} \\
&\rightsquigarrow \mathrm{N} \left(0, (\mathbb{E}(f_1 + f_2)^2)^{-1} 4 \operatorname{Var} \left(f_1 \right) \sigma_U^2 \right).
\end{aligned}$$

The OLS estimator $\hat{\gamma}$ is asymptotically efficient, as we will see in the discussion of efficiency. The OLS estimator $\hat{\beta}$ is rate inefficient. Since β could be identified by dyadic variation, we naturally expect a root-$\binom{N}{2}$ convergence rate. This could be achieved by several estimators, including "fixed-effect" estimator and tetrad difference estimator.

Tetrad Estimator

Consider the following estimator of β.

$$\begin{aligned}
Q_N(b) &\equiv \frac{1}{N(N-1)(N-2)(N-3)} \sum_{1 \le i \ne j \ne k \ne l \le N} \frac{1}{2} \left[\Delta_{ijkl} Y - \Delta_{ijkl} W^\top b \right]^2 \\
\hat{\beta} &= \operatorname*{arg\,min}_{b \in \mathbb{R}^{p_W}} Q_N(b),
\end{aligned}$$

where we use the notation $\Delta_{ijkl} Z \equiv Z_{ij} - Z_{ik} - Z_{jl} + Z_{kl}$. Tetrad differencing removes the the monadic effects and keeps only the dyadic effects left

$$\begin{aligned}
\Delta_{ijkl} Y &= \Delta_{ijkl} W^\top \beta + \Delta_{ijkl} (f + f) + \Delta_{ijkl} (U + U + V) \\
&= \Delta_{ijkl} W^\top \beta + \Delta_{ijkl} V.
\end{aligned}$$

We can write $\hat{\beta}$ explicitly as

$$\hat{\beta} - \beta = \left(\frac{1}{N(N-1)(N-2)(N-3)} \sum_{1 \leq i \neq j \neq k \neq l \leq N} \Delta_{ijkl} W \Delta_{ijkl} W^\top \right)^{-1} \cdot \left(\frac{1}{N(N-1)(N-2)(N-3)} \sum_{1 \leq i \neq j \neq k \neq l \leq N} \Delta_{ijkl} W \Delta_{ijkl} V \right). \tag{4.1}$$

Moreover we write the second term in (4.1) as a U-statistic

$$\frac{1}{N(N-1)(N-2)(N-3)} \sum_{1 \leq i \neq j \neq k \neq l \leq N} \Delta_{ijkl} W \Delta_{ijkl} V = \binom{N}{4}^{-1} \sum_{i<j<k<l} \xi_{ijkl}$$

$$\xi_{ijkl} = \frac{1}{4!} \sum_{(a,b,c,d) \in \mathbf{p}(\{i,j,k,l\})} \Delta_{abcd} W \Delta_{abcd} V,$$

where ξ_{ijkl} is a symmetrized kernel. Since

$$\mathbb{E}\left[\xi_{1234} | X_1, U_i\right] = 0,$$

this U-statistic is degenerate in the sense that its Hajek projection is zero, which implies the variance of this second term is of order $O(N^{-2})$. Together with the fact the first converges to a constant in the limit, it implies $\hat{\beta}$ converges at rate root-$\binom{N}{2}$.

"Fixed-Effect" Estimator

The fixed-effect estimator achieves similar goal as the tetrad difference estimator.

$$\hat{\beta} = \left[\mathbf{W}^\top \left(I - P_{\mathbf{T}}\right) \mathbf{W}\right]^{-1} \mathbf{W}^\top \left(I - P_{\mathbf{T}}\right) \mathbf{Y},$$

where $P_{\mathbf{T}} = T(T^\top T)^{-1} T^\top$ is the projection matrix of column space of $\mathbf{T}$. This estimator uses $I - P_{\mathbf{T}}$ to extract pure dyadic variation and regress the projection residual of Y onto that of W. This estimator is intuitively again converging at rate root-$\binom{N}{2}$.

Infeasible OLS with Second-Order Terms in Hoeffding Decomposition

For this particular model, Hoeffding decomposition decomposes monadic effect and dyadic effect in a clear way.[1] For any random variable Z, denote $PZ \equiv \mathbb{E}Z$, $P_{\{i\}}Z \equiv \mathbb{E}[Z|X_i, U_i] - \mathbb{E}Z$, $P_{\{i,j\}}Z \equiv Z - \mathbb{E}[Z|X_i, U_i] - \mathbb{E}[Z|X_j, U_j] + \mathbb{E}Z$. Let's write down the decomposition explicitly

$$\begin{aligned}
Y_{ij} &= PY_{ij} + (P_{\{i\}}Y_{ij} + P_{\{j\}}Y_{ij}) + P_{\{i,j\}}Y_{ij} \\
PY_{ij} &= PW^\top \beta + 2Pf^\top \gamma \\
(P_{\{i\}}Y_{ij} + P_{\{j\}}Y_{ij}) &= (P_{\{i\}}W_{ij} + P_{\{j\}}W_{ij})^\top \beta + (f_i + f_j - 2Pf)^\top \gamma + U_i + U_j \\
P_{\{i,j\}}Y_{ij} &= P_{\{i,j\}}W_{ij}^\top \beta + V_{ij}
\end{aligned}$$

[1] We will see later on that there may still be subtlety unrevealed by Hoeffding decomposition

Both the tetrad estimator and the "fixed-effect" estimator are essentially asymptotically equivalent to OLS of

$$P_{\{i,j\}}Y_{ij} = P_{\{i,j\}}W_{ij}^\top\beta + V_{ij}.$$

Estimators based on the second-order terms in the Hoeffding decomposition may exhibit rate root-$\binom{N}{2}$ convergence. This is because the "score" term $\binom{N}{2}^{-1}\sum_{i<j} P_{\{i,j\}}W_{ij}\cdot V_{ij}$ is degenerate with zero Hajek projection. Namely, $\mathbb{E}\left[\left(P_{\{i,j\}}W_{ij}\right)\cdot V_{ij}|X_i,U_i\right]=0$ or equivalently

$$\left(P_{\{i,j\}}W_{ij}\right)\cdot V_{ij} \in H_{\{i,j\}} \equiv \left\{P_{\{i,j\}}Z(X_i,U_i,X_j,U_j,V_{ij}):\ \forall Z:\mathbb{R}^{2d+3}\to\mathbb{R}^p, p=1,2,\ldots\right\}.$$

Heteroskedasticity

Here we extend the homoskedastic model in the previous section by introducing a specific form of multiplicative heteroskedasticity.

$$\begin{aligned}\text{Model: } Y_{ij} &= W_{ij}^\top\beta + (f_i+f_j)^\top\gamma + r_{ij}, 1\le i<j\le N\\ r_{ij} &= \sigma(X_i,X_j)\cdot e_{ij} = \sigma(X_i,X_j)\cdot(U_i+U_j+V_{ij}),\end{aligned}$$

where $\sigma:\mathbb{R}^{2d}\to\mathbb{R}_+$ specifies the form of heteroskedasticity. The sampling assumptions stay the same as the homoskedastic model. One may suspect the tetrad estimator or the "fixed-effect" estimator has the same fast rate of convergence in this enriched model. However, this is not the case. We will see tetrad/fixed-effect estimator will lose their fast root-$\binom{N}{2}$ rate in general. The best way to see this is by inspecting the second-order terms in the Hoeffding decomposition.

$$\begin{aligned}P_{\{i,j\}}Y_{ij} &= P_{\{i,j\}}W_{ij}^\top\beta + P_{\{i,j\}}r_{ij}\\ &= P_{\{i,j\}}W_{ij}^\top\beta\\ &\quad + [\sigma(X_i,X_j) - P\sigma - P_{\{i\}}\sigma(X_i,X_j)]U_i + [\sigma(X_i,X_j) - P\sigma - P_{\{j\}}\sigma(X_i,X_j)]U_j\\ &\quad + \sigma(X_i,X_j)V_{ij}.\end{aligned}$$

The key observation for understanding the (infeasible) OLS based on this equation is the following. The "score" term $\binom{N}{2}^{-1}\sum_{i<j}[P_{i,j}W_{ij}\cdot P_{i,j}r_{ij}]$ is a nondegenerate U-statistic in general. To see this, we need to show that its Hajek projection is not zero. Let's compute the projection of $P_{\{i,j\}}W_{ij}\cdot P_{\{i,j\}}r_{ij}$ onto the space of (X_i,U_i).

$$\begin{aligned}P_{\{i\}}\left(P_{\{i,j\}}W_{ij}\cdot P_{\{i,j\}}r_{ij}\right) &= \mathbb{E}\left(P_{\{i,j\}}W_{ij}\cdot P_{\{i,j\}}r_{ij}\middle|X_i,U_i\right)\\ &= \mathbb{E}\left(P_{\{i,j\}}W_{ij}\cdot[\sigma(X_i,X_j)-P\sigma-P_{\{i\}}\sigma(X_i,X_j)]U_i\middle|X_i,U_i\right)\\ &= U_i\mathbb{E}\left(P_{\{i,j\}}W_{ij}\cdot[\sigma(X_i,X_j)-P\sigma-P_{\{i\}}\sigma(X_i,X_j)]\middle|X_i\right),\end{aligned}$$

which is not zero in general. [2] To interpret this nondegeneracy, notice

$$P_{\{i,j\}}r_{ij} = [\sigma(X_i,X_j)-P\sigma-P_{\{i\}}\sigma(X_i,X_j)]U_i+[\sigma(X_i,X_j)-P\sigma-P_{\{i\}}\sigma(X_i,X_j)]U_i+\sigma(X_i,X_j)V_{ij},$$

[2] For example, if $W_{ij}=X_iX_j$ and $\sigma(X_i,X_j)=1+X_i^2X_j^2$, then the last term is $U_iX_i^2(X_i-\mathbb{E}X)\operatorname{Cov}\left(X_j,X_j^2\right)$.

which involves both monadic and dyadic variation. $P_{\{i,j\}}r_{ij}$ itself being degenerate doesn't guarantee its product with $P_{\{i,j\}}W_{ij}$ to be degenerate. In fact, the product is degenerate only in special cases, e.g. homoskedasticity. When $\sigma(X_i, X_j) = \sigma$ is a constant, the product is degenerate. To make this more explicit, $\sigma(X_i, X_j) = \sigma$ implies $P_{\{i,j\}}r_{ij} = \sigma P_{\{i,j\}}e_{ij} = \sigma V_{ij}$, which is totally independent of the monadic variation. Unfortunately, this degeneracy is special rather than general.

The nondegeneracy of $P_{\{i,j\}}W_{ij} \cdot P_{\{i,j\}}r_{ij}$ implies that the infeasible OLS converges at rate root-N instead of root-$\binom{N}{2}$. The remaining question is "Can we construct root-$\binom{N}{2}$ estimator?"

Before trying to answer this question, it is instructive to investigate a local to homoskedasticity asymptotics.

Local to Homoskedasticity

We are looking at a sequence of model with moving primitives. In particular, $\sigma_N(X_i, X_j) = \sigma + N^{-\alpha}\tilde{\sigma}(X_i, X_j)$ where σ is a constant and $\tilde{\sigma}(X_i, X_j)$ is a mean-zero function. $\alpha \geq 0$ controls the speed of slipping into homoskedasticity. In this sequence of model, the projection term

$$P_{\{i\}}\left(P_{\{i,j\}}W_{ij} \cdot P_{\{i,j\}}r_{ij}\right) = N^{-\alpha}U_i\mathbb{E}\left(P_{\{i,j\}}W_{ij} \cdot [\tilde{\sigma}(X_i, X_j) - P_{\{i\}}\tilde{\sigma}(X_i, X_j)]\big|X_i, U_i\right),$$

is shrinking to zero at rate $N^{-\alpha}$. This implies the variance of the score term $\binom{N}{2}^{-1}\sum_{i<j}\left(P_{\{i,j\}}W_{ij} \cdot P_{\{i,j\}}r_{ij}\right)$ is of order $O\left(N^{-1}\operatorname{Var}\left(P_{\{i\}}\left(P_{\{i,j\}}W_{ij} \cdot P_{\{i,j\}}r_{ij}\right)\right) + N^{-2}\right) = O\left(N^{-1-2\alpha} + N^{-2}\right)$. The convergence rate of the infeasible OLS is root-$N^{-\min\{1+2\alpha,2\}}$. This asymptotic analysis interpolates the "homo-" and the "hetero-" worlds

Infeasible Weighted OLS with Infeasible Weight

A simple (infeasible) reweighting, namely

$$\begin{aligned}
\tilde{Y}_{ij} &\equiv \sigma(X_i, X_j)^{-1}Y_{ij} \\
&= \sigma(X_i, X_j)^{-1}W_{ij}^\top\beta + \sigma(X_i, X_j)^{-1}(f_i + f_j)^\top\gamma + \sigma(X_i, X_j)^{-1}r_{ij} \\
&= \tilde{W}_{ij}^\top\beta + \tilde{f}_{ij}^{\top}\gamma + e_{ij},
\end{aligned}$$

gets us back into the nice situation under homoskedasticity. Here we list some comments about this result. First, the infeasible OLS with infeasible weight is expected to converge at rate root-$\binom{N}{2}$. Second, it seems to open up the possibility to estimate γ at a faster rate than root-N. The main takeaway is that we can construct a more efficient estimator if we have additional knowledge about the heteroskedasticity structure. We will see in the following that even if we do not know this heteroskedasticity form σ, we can still estimate it and use the estimated σ to construct a more efficient estimator.

Feasible Weighted Tetrad Estimator

Suppose we do not know the exact form of the heteroskedasticity function $\sigma(X_i, X_j)$. We would like to construct a feasible version of the previous infeasible OLS with infeasible weight. This task is similar to constructing a feasible GLS in the classical linear regression with heteroskedasticity model. Here is the proposed algorithm.

Algorithm 1 Feasible Weighted Tetrad Estimator

Input: Monadic variables $\{X_i, f_i, i = 1, \ldots, N\}$, dyadic outcome $\{Y_{ij}, 1 \le i < j \le N\}$, function W.

Output: estimate $(\hat{\beta}, \hat{\gamma})$

Fit a simple OLS (in level) to get an initial estimate of β and γ

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\gamma} \end{pmatrix} = \left[\left(\mathbf{W} \vdots \mathbf{Sf} \right)^\top \left(\mathbf{W} \vdots \mathbf{Sf} \right) \right]^{-1} \left(\mathbf{W} \vdots \mathbf{Sf} \right)^\top \mathbf{Y}$$

Compute the regression residual $\hat{r}_{ij} = Y_{ij} - W_{ij}^\top \tilde{\beta} - (f_i + f_j)^\top \tilde{\gamma}$ [3]

Estimate σ^2 function by running a (non)parametric regression of $\hat{r}_{ij}^2$ on regressor (X_i, X_j). Use $\hat{\sigma}^2$ to denote the estimated function

Construct transformed data using $\hat{\sigma}^2$. $\tilde{Y}_{ij} = \hat{\sigma}(X_i, X_j)^{-1} Y_{ij}, \tilde{W}_{ij} = \hat{\sigma}(X_i, X_j)^{-1} W_{ij}, \tilde{f}_{ij} = \hat{\sigma}(X_i, X_j)^{-1}(f_i + f_j)$.

Generate the final estimate by an OLS on tetrad-difference terms

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \underset{b, c \in \mathbb{R}^{p_W + p_f}}{\arg\min} \; Q_N(b, c),$$

$$Q_N(b, c) \equiv \frac{1}{N(N-1)(N-2)(N-3)} \sum_{1 \le i \ne j \ne k \ne l \le N} \frac{1}{2} \left[\Delta_{ijkl} \tilde{Y} - \Delta_{ijkl} \tilde{W}^\top b - \Delta_{ijkl} \tilde{f}^\top c \right]^2.$$

In the following, we will outline an informal analysis of this estimator. The first step OLS estimator has an estimation error of order root-N as we have shown before. As a result, the discrepancy between the residual $\hat{r}$ and the true regression error r is also of order $O(N^{-1/2})$. The nonparametric regression for estimating σ^2 has an estimation error at rate $O\left(N^{-\frac{\beta}{2\beta + d_x}}\right)$. This estimation error together with the already existing discrepancy $\hat{r}^2 = r^2 + O(N^{-1/2})$ implies the $\hat{\sigma} - \sigma = O\left(N^{-\frac{\beta}{2\beta + d_x}}\right)$. The residual in the transformed data

[3] A more principled way to generate these residuals involves sample splitting (or graph splitting) and constructing residuals from out-of-sample predictions. This trick help us bypass the in-sample overfitting problem and simplify the bias analysis.

$\tilde{Y}_{ij} = \tilde{W}_{ij}^\top \beta + \tilde{f}_{ij}^\top \gamma + \tilde{r}_{ij}$ is

$$\tilde{r}_{ij} = \frac{r_{ij}}{\hat{\sigma}(X_i, X_j)} = \left(1 - \frac{\hat{\sigma}(X_i, X_j) - \sigma(X_i, X_j)}{\hat{\sigma}(X_i, X_j)}\right) \cdot e_{ij}.$$

Since $\frac{\hat{\sigma}(X_i,X_j)-\sigma(X_i,X_j)}{\hat{\sigma}(X_i,X_j)} = O\left(N^{-\frac{\beta}{2\beta+d_x}}\right)$, the transformed data is local to homoskedasticity. The result from local to homoskedasticity analysis suggests the infeasible OLS will converge at rate

$$\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} = O\left(N^{-\left(\frac{1}{2} + \frac{\beta}{2\beta+d_x}\right)}\right).$$

We expect that replacing the infeasible OLS with the tetrad estimator in the last step will not change this result. Notice that using this nonparametrically estimated weight is not asymptotically equivalent to using the infeasible weight. Weighting by the estimated weight pushes the estimator in the heteroskedastic model to be equivalent to the unweighted estimator in the local to homoskedasticity model (but not equivalent to the unweighted estimator in the homoskedastic model).

One way to think about the problem here is that the estimated weight is converging slower than (or equal to) the usual cutoff rate, $n^{-1/4} = \binom{N}{2}^{-1/4} = N^{-1/2}$, in a typical analysis of two-step semiparametric estimator. The analogy to an unweighted estimator under local to homoskedasticity gives a more precise intuition.

Suppose we are willing to specify a parametric model for σ^2 and estimate it using parametric dyadic regression instead of nonparametric dyadic regression. In this case, the weight can be estimated at a faster rate $\hat{\sigma} - \sigma = O(N^{-1/2})$. This is a knife-edge case in the following sense. The estimation error is exactly equal to the $n^{-1/4}$ cutoff rate for two-step semiparametric estimation. Local to homoskedasticity analysis will tell us $\begin{pmatrix} \hat{\beta} - \beta \\ \hat{\gamma} - \gamma \end{pmatrix} = O\left(N^{-1}\right)$, which is the same rate as the infeasible weighted infeasible OLS. This is the best convergence rate we can achieve. We conjecture the knife-edge nature implies a non-vanishing asymptotic bias or an additional variance term when we go from an infeasible estimator to a feasible estimator.

4.4 Cramér-Rao Lower Bound

In previous sections, we see that parameters can be estimated at different rates ranging between root-N and root-$\binom{N}{2}$ and the convergence rate may be different depending on what estimator we use. In this section, we want to understand the nature of the estimation problems in a different way by calculating the Cramér-Rao lower bound of parameters in the linear dyadic regression. The exercise delivers not only the optimal rate of convergence but also the optimal asymptotic variance. Even though we do not have a theorem formalizing this, we conjecture the CRLB gives us an efficiency bound in dyadic setting. We work out the calculation of CRLB in the following.

We are working with a slightly more general nonlinear regression model with heteroskedasticity here.

$$\begin{aligned}\text{Model: } & Y_{ij} = \mu_\theta(X_i, X_j) + r_{ij}, 1 \le i < j \le N \\ & r_{ij} = \sigma(X_i, X_j) \cdot e_{ij} = \sigma(X_i, X_j) \cdot (U_i + U_j + V_{ij}), U_i \overset{\text{iid}}{\sim} \mathrm{N}(0,1), V_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0,1). \end{aligned} \tag{4.2}$$

Suppose the density of X denoted by λ is also known. The only unknown parameters in this model are θ and σ.

The log likelihood and the score of observing the hypothetical data $(\mathbf{X}, \mathbf{U}, \mathbf{Y}) = (X_1, \ldots, X_N, U_1, \ldots, U_N, Y_{ij}, i < j)$ is

$$\begin{aligned} l\left(\theta, \sigma; \mathbf{X}, \mathbf{U}, \mathbf{Y}\right) &= \sum_{i=1}^{N} [\log \lambda\left(X_i\right) + \log \phi\left(U_i\right)] \\ &\quad + \sum_{i<j} \left[\log \phi \left(\frac{Y_{ij} - \mu_\theta\left(X_i, X_j\right)}{\sigma(X_i, X_j)} - U_i - U_j\right) - \log \sigma(X_i, X_j)\right] \\ \dot{l}_{\theta,\sigma}\left(\theta, \sigma; \mathbf{X}, \mathbf{U}, \mathbf{Y}\right) &:= \frac{\partial}{\partial \theta} l\left(\theta, \sigma; \mathbf{X}, \mathbf{U}, \mathbf{Y}\right) = \sum_{i<j} V_{ij} \cdot \frac{\dot{\mu}_{ij}(\theta)}{\sigma_{ij}} \\ \frac{\partial}{\partial t}\Big|_{t=0} l\left(\theta, \sigma(1+tg)^{-1}; \mathbf{X}, \mathbf{U}, \mathbf{Y}\right) &= \sum_{i<j} (-V_{ij} e_{ij} + 1) g_{ij}. \end{aligned}$$

ϕ here denotes the density of the standard normal distribution. The last line is calculating the directional derivative of the log likelihood w.r.t. the σ in a direction defined by function g. To calculate the score of the actual data $(\mathbf{X}, \mathbf{Y})$, notice

$$\begin{aligned} \dot{l}\left(\theta, \sigma; \mathbf{X}, \mathbf{Y}\right) &:= \frac{\partial}{\partial \theta} l\left(\theta, \sigma; \mathbf{X}, \mathbf{Y}\right) \\ &= \frac{\frac{\partial}{\partial \theta} p_{\theta,\sigma}\left(\mathbf{X}, \mathbf{Y}\right)}{p_{\theta,\sigma}\left(\mathbf{X}, \mathbf{Y}\right)} = \frac{\frac{\partial}{\partial \theta}\left(\int p_{\theta,\sigma}\left(\mathbf{X}, \mathbf{u}, \mathbf{Y}\right) d\mathbf{u}\right)}{p_{\theta,\sigma}\left(\mathbf{X}, \mathbf{Y}\right)} = \frac{\int \frac{\frac{\partial}{\partial \theta} p_{\theta,\sigma}(\mathbf{X}, \mathbf{u}, \mathbf{Y})}{p_{\theta,\sigma}(\mathbf{X}, \mathbf{u}, \mathbf{Y})} \cdot p_{\theta,\sigma}\left(\mathbf{X}, \mathbf{u}, \mathbf{Y}\right) d\mathbf{u}}{p_{\theta,\sigma}\left(\mathbf{X}, \mathbf{Y}\right)} \\ &= \int \dot{l}\left(\theta, \sigma; \mathbf{X}, \mathbf{u}, \mathbf{Y}\right) p_{\theta,\sigma}\left(\mathbf{u} \,\middle|\, \mathbf{X}, \mathbf{Y}\right) d\mathbf{u} \\ &= \mathbb{E}_{\theta,\sigma}\left[\dot{l}\left(\theta, \sigma; \mathbf{X}, \mathbf{U}, \mathbf{Y}\right) \,\middle|\, \mathbf{X}, \mathbf{Y}\right]. \end{aligned}$$

Applying this to our problem gives

$$\begin{aligned} \dot{l}_{\theta,\sigma}\left(\theta, \sigma; \mathbf{X}, \mathbf{Y}\right) &= \sum_{i<j} \mathbb{E}\left(V_{ij} \,\middle|\, \mathbf{e}\right) \cdot \frac{\dot{\mu}_{ij}(\theta)}{\sigma_{ij}} \\ \frac{\partial}{\partial t}\Big|_{t=0} l\left(\theta, \sigma(1+tg)^{-1}; \mathbf{X}, \mathbf{Y}\right) &= \sum_{i<j} \left[-\mathbb{E}\left(V_{ij} \,\middle|\, \mathbf{e}\right) e_{ij} + 1\right] g_{ij}, \end{aligned}$$

where $\mathbf{e} = (e_{ij}, i < j)$. The overall "efficient score" for θ is

$$\begin{aligned}\tilde{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y}) &:= \dot{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y}) - \Pi\ \dot{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y}) \\ &= \dot{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y}) - 0 \\ &= \dot{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y}), \end{aligned} \tag{4.3}$$

in which $\Pi\ \dot{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y})$ denotes the L_2 projection of $\dot{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y})$ onto the tangent space spanned by the overall score of the nuisance parameter $\left\{\frac{\partial}{\partial t}\Big|_{t=0} l\left(\theta, \sigma(1+tg)^{-1}; \mathbf{X}, \mathbf{Y}\right) :\ g : \mathbb{R}^{2d_x} \to \mathbb{R}, ||g||_\infty < \infty\right\}$. The projection is zero because for any direction g

$$\begin{aligned}&\mathrm{Cov}\left(\dot{l}_\theta(\theta, \sigma; \mathbf{X}, \mathbf{Y}), \frac{\partial}{\partial t}\Big|_{t=0} l\left(\theta, \sigma(1+tg)^{-1}; \mathbf{X}, \mathbf{Y}\right)\right) \\ &= \mathrm{Cov}\left(\sum_{i<j} \mathbb{E}\left(V_{ij}\Big|\mathbf{e}\right)\frac{\dot{\mu}_{ij}(\theta)}{\sigma_{ij}}, \sum_{i<j}\left[-\mathbb{E}\left(V_{ij}\Big|\mathbf{e}\right)e_{ij} + 1\right]g_{ij}\right) \\ &= \sum_{i_1<j_1}\sum_{i_2<j_2}\mathbb{E}\left\{\mathbb{E}\left(V_{i_1j_1}\Big|\mathbf{e}\right)\cdot\left[-\mathbb{E}\left(V_{i_2j_2}\Big|\mathbf{e}\right)e_{i_2j_2} + 1\right]\right\}\cdot\mathbb{E}\left[\frac{\dot{\mu}_{i_1j_1}(\theta)}{\sigma_{i_1j_1}}g_{i_2j_2}\right] \\ &= \sum_{i_1<j_1}\sum_{i_2<j_2} 0\cdot\mathbb{E}\left[\frac{\dot{\mu}_{i_1j_1}(\theta)}{\sigma_{i_1j_1}}g_{i_2j_2}\right] \\ &= 0.\end{aligned}$$

Here the second equality holds because $\mathbb{E}\left(V_{i_1j_1}\Big|\mathbf{e}\right)\cdot\mathbb{E}\left(V_{i_2j_2}\Big|\mathbf{e}\right)\cdot e_{i_2j_2}$ can be expressed as third-order (see equation (4.8) in the appendix) polynomials of Us and Vs and all of these third-order polynomials have zero expectation. (4.3) implies the "efficient score function" $\tilde{l}_{\theta,\sigma}$ doesn't depend on whether the nuisance parameter σ is known or not. Remarkably, the efficiency bound remains the same even if we do not know σ. Remember a similar result holds for linear regression with unknown heteroskedasticity in the cross-sectional iid setting.

We summarize the previous analysis into the following theorem.

Theorem 4.4.1 (CRLB of Dyadic Regression with Heteroskedasticity)**.** *For the dyadic regression model with heteroskedasticity in* (4.2)*, the overall efficient score for θ is*

$$\tilde{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y}) := \dot{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y}) = \sum_{i<j}\mathbb{E}\left(V_{ij}\Big|\mathbf{e}\right)\cdot\frac{\dot{\mu}_{ij}(\theta)}{\sigma_{ij}}. \tag{4.4}$$

The Fisher information of θ is $\mathcal{I}(\theta) = \mathrm{Var}\left(\tilde{l}_{\theta,\sigma}(\theta, \sigma; \mathbf{X}, \mathbf{Y})\right)$ and the CRLB is $\mathcal{I}(\theta)^{-1}$.

To appreciate this result, it's useful to calculate the Fisher information explicitly in special cases. Suppose the true model is indeed homoskedastic with $\sigma(X_i, X_j) = 1$. If $\mu_\theta(x_1, x_2) =$

$(x_1 + x_2) \cdot \theta$, then $\tilde{l}\,(\theta; \mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{N} X_i U_i + o_{L^2}\left(\sqrt{N}\right)$, $I(\theta) = N\mathbb{E}\left[X_1^2\right](1 + o(1))$. If $\mu_\theta(x_1, x_2) = x_1 \cdot x_2 \cdot \theta$, then $\tilde{l}\,(\theta; \mathbf{X}, \mathbf{Y}) = \left[\sum_{i<j} V_{ij}\left(X_i - \mu_X\right)\left(X_j - \mu_X\right)\right](1 + o_P(1))$, $I(\beta) = \binom{N}{2} \operatorname{Var}\left(X_1\right)^2 (1 + o_P(1))$. The results are in line with our intuition. The coefficient of the monadic regressor will at best converge at rate $O\left(N^{-1/2}\right)$. The coefficient of the dyadic regressor will at best converge at rate $O\left(\binom{N}{2}^{-1/2}\right)$. Careful inspection reveals that OLS achieves the optimal asymptotic variance for the former model and the tetrad estimator achieves the optimal asymptotic variance for the later model.

Let's go back to the model in 4.3 now. Because the CRLB stays the same either we know σ or we don't know σ, we expect an efficient estimator will deliver $O\left(\binom{N}{2}^{-1/2}\right)$ rate of convergence for the parameter β, γ. The best estimator we proposed, the feasible weighted tetrad estimator in 4.3, only converges at a slower rate $O\left(N^{-\left(\frac{1}{2}+\frac{\beta}{2\beta+d_x}\right)}\right)$. This gap between the estimator and the efficiency bound is quite interesting in its own right. Remember there is not any such gap for linear regression with unknown heteroskedasticity with iid data, where a similar two-step estimator achieves the asymptotic efficiency bound. For dyadic regression, either there exists a better estimator that is adaptive and achieves the CRLB, or there is a tighter efficiency bound. We are not sure which case is happening. We point this out as an open question for further research.

4.5 Appendix: Proofs

This appendix gives some detailed calculations of score functions. Here we give an analytical expression of the term $\mathbb{E}\left(V_{ij}|\mathbf{e}\right)$, which shows up repeatedly in the derivation and in the expression of the efficient score (4.4). Toward this goal, we first calculate $\mathbb{E}\left(U_i|\mathbf{e}\right)$. A useful expression is $\mathbf{e} = \mathbf{TU} + \mathbf{V}$. Because all the variables Us, Vs, and es are jointly normal, calculating conditional expectation reduces to calculating the best linear approximation.

The normality assumption comes in very handy in this case.

$$
\begin{aligned}
\mathbb{E}\left(U_1\middle|\mathbf{e}\right) &= \mathrm{Cov}\left(U_1, \mathbf{e}\right)\mathrm{Var}\left(\mathbf{e}\right)^{-1}\mathbf{e} \\
&= (\underbrace{1,\dots,1}_{N-1}, 0, \dots 0)_{1\times\binom{N}{2}}\left(I_{\binom{N}{2}} + \mathbf{T}\mathbf{T}^\top\right)^{-1}\mathbf{e} \\
&= (\underbrace{1,\dots,1}_{N-1}, 0, \dots 0)_{1\times\binom{N}{2}}\left[I_{\binom{N}{2}} - (N-1)^{-1}\mathbf{T}\mathbf{T}^\top + 4(N-1)^{-1}(2N-1)^{-1}1_{\binom{N}{2}}1^\top_{\binom{N}{2}}\right]\mathbf{e} \\
&&& (4.5) \\
&= \left[\frac{1}{N-1}(1^\top_{N-1}, 0^\top_{\binom{N}{2}-N-1}) - \frac{2}{(N-1)(2N-1)}1^\top_{\binom{N}{2}}\right]\mathbf{e} && (4.6) \\
&= e_1 - \frac{N}{2N-1}e, && (4.7)
\end{aligned}
$$

where $e_1 := \frac{1}{N-1}\sum_{j\neq 1} e_{1j}$ and $e := \frac{1}{\binom{N}{2}}\sum_{i<j} e_{ij}$. The third equality (4.5) uses the following explicit expression of the inverse matrix.

$$
\begin{aligned}
\left(I_{\binom{N}{2}} + \mathbf{T}\mathbf{T}^\top\right)^{-1} &= I_{\binom{N}{2}} - \mathbf{T}\left(I_N + \mathbf{T}^\top\mathbf{T}\right)^{-1}\mathbf{T}^\top \\
&= I_{\binom{N}{2}} - \mathbf{T}(N-1)^{-1}\left(I_N - (2N-1)^{-1}1_N1_N^\top\right)\mathbf{T}^\top \\
&= I_{\binom{N}{2}} - (N-1)^{-1}\mathbf{T}\mathbf{T}^\top + 4(N-1)^{-1}(2N-1)^{-1}1_{\binom{N}{2}}1^\top_{\binom{N}{2}}
\end{aligned}
$$

The fourth equality (4.6) is due to

$$
\begin{aligned}
&(1^\top_{N-1}, 0^\top_{\binom{N}{2}-N-1})\left[I_{\binom{N}{2}} - (N-1)^{-1}\mathbf{T}\mathbf{T}^\top + 4(N-1)^{-1}(2N-1)^{-1}1_{\binom{N}{2}}1^\top_{\binom{N}{2}}\right] \\
=\ &(1^\top_{N-1}, 0^\top_{\binom{N}{2}-N-1}) - (N-1)^{-1}(N-1, 1, \dots, 1)_{N\times 1}\mathbf{T}^\top + 4(N-1)^{-1}(2N-1)^{-1}(N-1)1^\top_{\binom{N}{2}} \\
=\ &(1^\top_{N-1}, 0^\top_{\binom{N}{2}-N-1}) - (N-1)^{-1}\left[(N-2)(1^\top_{N-1}, 0^\top_{\binom{N}{2}-N-1}) + 21^\top_{\binom{N}{2}}\right] + \frac{4}{2N-1}1^\top_{\binom{N}{2}} \\
=\ &\frac{1}{N-1}(1^\top_{N-1}, 0^\top_{\binom{N}{2}-N-1}) - \frac{2}{(N-1)(2N-1)}1^\top_{\binom{N}{2}}.
\end{aligned}
$$

We are now ready to calculate $\mathbb{E}\left(V_{12}|\mathbf{e}\right)$.

$$
\begin{aligned}
\mathbb{E}\left(V_{12}\middle|\mathbf{e}\right) &= \mathbb{E}\left(e_{12} - U_1 - U_2\middle|\mathbf{e}\right) \\
&= e_{12} - \mathbb{E}\left(U_1\middle|\mathbf{e}\right) - \mathbb{E}\left(U_2\middle|\mathbf{e}\right) \\
&= e_{12} - e_1 - e_2 + \frac{2N}{2N-1}e && (4.8) \\
&= V_{12} - V_1 - V_2 + \left(1 + \frac{1}{2N-1}\right)V + \frac{1}{N-1}\left(U_1 + U_2\right) - \frac{2N}{(N-1)(2N-1)}U,
\end{aligned}
$$

where $V_1 := \frac{1}{N-1}\sum_{j\neq 1} V_{1j}$, $V := \frac{1}{\binom{N}{2}}\sum_{i<j} V_{ij}$, and $U = \frac{1}{N}\sum_i U_i$. In the main text we made the following claim: if $\mu_\theta(x_1, x_2) = (x_1 + x_2)\cdot\theta$, then $\tilde{l}\left(\theta; \mathbf{X}, \mathbf{Y}\right) = \sum_{i=1}^N X_i U_i + o_{L^2}\left(\sqrt{N}\right)$, $I(\theta) = N\mathbb{E}\left[X_1^2\right](1 + o(1))$. We give its derivation here. Theorem 4.4.1 implies

$$\tilde{l}_{\theta,\sigma}\left(\theta, \sigma; \mathbf{X}, \mathbf{Y}\right) = \sum_{i<j}\mathbb{E}\left(V_{ij}\middle|\mathbf{e}\right)\cdot\left(X_i + X_j\right).$$

Plug in the expression of $\mathbb{E}\left(V_{12}\middle|\mathbf{e}\right)$.

$$\begin{aligned}
&\sum_{i<j}\mathbb{E}\left(V_{ij}\middle|\mathbf{e}\right)\cdot\left(X_i + X_j\right)\\
&= \sum_{i=1}^N\left[X_i\cdot\sum_{j\neq i}\mathbb{E}\left(V_{ij}\middle|\mathbf{e}\right)\right]\\
&= \sum_{i=1}^N X_i\cdot\left(V_i - \frac{N}{2N-1}V + U_i\left(1 - \frac{1}{N-1}\right) + \frac{N}{(N-1)(2N-1)}U\right)\\
&= \left(1 - \frac{1}{N-1}\right)\sum_{i=1}^N X_i U_i + \sum_{i=1}^N X_i V_i + \sum_{i=1}^N X_i\left(-\frac{N}{2N-1}V + \frac{N}{(N-1)(2N-1)}U\right)\\
&= \sum_{i=1}^N X_i U_i + o_{L^2}\left(\sqrt{N}\right).
\end{aligned}$$

Bibliography

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598.

Arcones, M. A. and Gine, E. (1992). On the Bootstrap of U and V Statistics. *The Annals of Statistics*, 20(2):655 – 674.

Arcones, M. A. and Gine, E. (1993). Limit Theorems for U-Processes. *The Annals of Probability*, 21(3):1494 – 1542.

Aronow, P. M., Samii, C., and Assenova, V. A. (2017). Cluster–robust variance estimation for dyadic data. *Political Analysis*, 23(4):564–577.

Atalay, E., Hortacsu, A., Roberts, J., and Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13):5199–5202.

Bickel, P. J., Chen, A., and Levina, E. (2011). The method of moments and degree distributions for network models. *The Annals of Statistics*, 39(5):2280–2301.

Bickel, P. J. and Freedman, D. A. (1981). Some Asymptotic Theory for the Bootstrap. *The Annals of Statistics*, 9(6):1196 – 1217.

Bose, A. and Chatterjee, S. (2018). *U-statistics, Mm-estimators and Resampling*. Springer.

Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press.

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2014a). Bootstrapping density-weighted average derivatives. *Econometric Theory*, 30(6):1135–1164.

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2014b). Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory*, 30(1):176–200.

Chiang, H. D., Kato, K., Ma, Y., and Sasaki, Y. (2019). Multiway cluster robust double/debiased machine learning. *arXiv preprint arXiv:1909.03489*.

Chiang, H. D. and Tan, B. Y. (2020). Empirical likelihood and uniform convergence rates for dyadic kernel density estimation. *arXiv preprint arXiv:2010.08838*.

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

Crane, H. and Towsner, H. (2018). Relatively exchangeable structures. *The Journal of Symbolic Logic*, 83(2):416–442.

Davezies, L., D'Haultfœuille, X., and Guyonvarch, Y. (2021). Empirical process results for exchangeable arrays. *The Annals of Statistics*, 49(2):845–862.

de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probability Theory and Related Fields*, 75(2):261–277.

Diaconis, P. and Janson, S. (2008). Graph limits and exchangeable random graphs. *Rendiconti di Matematica e delle sue Applicazioni. Serie VII*, pages 33–61.

Goldenberg, A., Zheng, A. X., Fienberg, S. E., and Airoldi, E. M. (2010). A survey of statistical network models. *Foundations and Trends® in Machine Learning*, 2(2):129–233.

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

Graham, B. S. (2020b). Sparse network asymptotics for logistic regression. *arXiv preprint arXiv:2010.04703*.

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

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

Holland, P. W. and Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7:1–45.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Technical Report, Institute for Advanced Study, Princeton, NJ.*

Ibragimov, I. A. and Has' Minskii, R. Z. (1982). Bounds for the risks of nonparametric regression estimates. *Theory of Probability and Its Applications*, 16:84–99.

Ibragimov, I. A. and Has' Minskii, R. Z. (1984). Asymptotic bounds on the quality of the nonparametric regression estimation in l_o. *Journal of Soviet Mathematics*, 25:540–550.

Janssen, P. (1994). Weighted bootstrapping of u-statistics. *Journal of Statistical Planning and Inference*, 38(1):31–41.

König, M. D., Liu, X., and Zenou, Y. (2019). R&d networks: Theory, empirics, and policy implications. *Review of Economics and Statistics*, 101(3):476–491.

Levin, K. and Levina, E. (2019). Bootstrapping networks with latent space structure. *arXiv preprint arXiv:1907.10821*.

Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika*, 82(1):93–100.

Lovász, L. (2012). *Large Networks and Graph Limits*. American Mathematical Society.

Mccullagh, P. (2000). Resampling and exchangeable arrays. *Bernoulli*, 6(2):285 – 301.

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica*.

Newey, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory*, 10(2):1–21.

Newey, W. K. and Stoker, T. M. (1993). Efficiency of weighted average derivative estimators and index models. *Econometrica*, 61(5):1199–1223.

Nowicki, K. (1991). Asymptotic distributions in random graphs with applications to social networks. *Statistica Neerlandica*, 45(3):295–325.

Owen, A. B. (2007). The pigeonhole bootstrap. *The Annals of Applied Statistics*, 1(2):386–411.

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076.

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

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430.

Rosenblatt, M. et al. (1956). Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837.

Silverman, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall.

Stoker, T. M. (1986). Consistent estimation of scaled coefficients. *Econometrica*, 54(6):1461–1481.

Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics*, 8(6):1348–1360.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics*, 10(4):1040–1053.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. Twentieth Century Fund.

Tsybakov, A. B. (2008). *Introduction to Nonparametric Estimation*. Springer.

Van der Vaart, A. W. (2000). *Asymptotic Statistics*. Cambridge University Press.