Density-Weighted Average Derivatives for Dyadic Data

Fengshi Niu *

November 15, 2021

Latest version available here

Abstract

This paper studies the density-weighted average derivatives, a direct estimator of index coefficients in a semiparametric single index regression model, for dyadic data, in which variables are naturally defined for pairs of agents. Dyadic data attracts increasing applied interest and exhibits dyadic dependence: variables of dyads sharing no agents are independent and variables of dyads sharing one or two agents are allowed to be dependent. Borrowing the technical tools for small-bandwidth asymptotics, I show that even though the presence/absence of dyadic dependence and the choice of small/large bandwidth sequence may affect the convergence rate of this estimator, its asymptotical normality holds robustly under mild conditions. This result presents an interesting contrast between kernel-based semiparametric estimators and the sample mean of dyadic data, the latter of which exhibits asymptotic non-normality when dyadic dependence is absent and whose uniform nonconservative inference procedure don't exist.

1 Introduction

Many social and economic variables have a natural underlying network or clustering structure. One set of notable examples falling into the category of *dyadic data* are trade flow among countries (Tinbergen, 1962), research collaboration among scientist (Newman, 2001; Ductor et al., 2014), sale and purchase of inputs between firms (Atalay et al., 2011), risk-sharing relationship across households (Fafchamps and Lund, 2003; Fafchamps and Gubert, 2007), and friendship relationship between individuals (Christakis et al., 2020). In these examples, the primary interests are variables, either continuous or discrete, between *every* pair of agents. Another set of examples falling into the category of *two-way clustered data* are wages between matched employer-employee (Abowd et al., 1999), test scores among matched students and teachers (Hanushek, 1971), and two-way panel data. In these examples, the primary interests are variables on a bipartite network between *two* dimensions, e.g. individual and time, or two groups of agents.

*Department of Economics, University of California - Berkeley, e-mail: fniu@berkeley.edu

*I thank seminar audiences at University of California - Berkeley for helpful feedback. I also thank Bryan Graham, Michael Jansson and James Powell for useful comments and discussion. All the usual disclaimers apply.

Both dyadic data and two-way clustered data exhibit *dyadic dependence*, i.e. variables of pairs sharing one or two indices are allowed to be dependent and those who don't share any indices are independent. Sources of dyadic dependence could often be *unobserved heterogeneity* in microeconometric data or demand/supply shocks to agents for trade data. In a matched data example, the test score of the same student with different teachers may be positively correlated because the student's unobserved ability affects these test scores in the same direction. In an international trade example, the import volume from various countries to the same country could be positively correlated because of the destination country's aggregate demand shock. In a panel data example, outcomes of the same individual or in the same time period may be positively correlated due to individual or time effects.

While semi/nonparametric estimators have been extensively studied for cross-sectional data, time-series data, and certain panel data, rigorous analysis of their performance and inference procedure for data subject to dyadic dependence are now only emerging. Recent efforts in this direction include Graham (2020a,b), which studies the nonparametric kernel density and regression estimators for dyadic data, and Chiang and Tan (2020), which extends double machine learning method to multiway clustering data and proposes robust standard errors.

In this paper, I extend and study the classical semiparametric estimator, kernel-based density-weighted average derivative in Powell et al. (1989) henceforth PSS, to data subject to dyadic dependence. The target parameters here are coefficients of a single index regression model, which covers many parametric models like linear models or censored regression for continuous outcomes and probit or logit models for binary outcomes. This estimator is a canonical example of semiparametric estimators employing nonparametric kernel estimators of unknown nuisance functions and has been studied extensively in the cross-sectional setting. PSS proposes this estimator establishes its asymptotic normality. Nishiyama and Robinson (1999, 2000, 2005) studied the rate of convergence to normality and establishes asymptotic refinement of bootstrap procedures in terms of estimating the estimator's sampling distribution. Cattaneo et al. (2014b,a) henceforth CCJ employs an asymptotic quadratic approximation to establish asymptotic normality and investigates various resampling inference procedures under considerably weaker assumptions on the bandwidth and the kernel.

Dyadic dependence poses a distinct challenge for inference of this estimator and many other statistics of dyadic data. In particular, the convergence rate of this estimator is $\sqrt{N}$ under nondegeneracy, i.e. the dyadic dependence being present, and is the much faster $\sqrt{\binom{N}{2}}$ under degeneracy, i.e. the dyadic dependence being absent. Intuitively, the presence and absence of dyadic dependence pivot the effective sample size between the number of agents and the number of dyads. This phenomenon also shows up for a variety of other estimators, including the sample mean, OLS estimator for linear regression (Menzel, 2021), and the kernel density estimators (GNP). Because the limit distribution is very sensitive to the level of dependence, inference method valid uniformly is particularly worth pursuing. However, Menzel (2021) points out the asymptotic distribution of the sample mean could be gaussian under nondegeneracy and nongaussian under degeneracy. In addition, he shows that there is no uniformly valid and nonconservative inference procedure adaptive to degeneracy for the sample mean.

Another challenge for inference of the estimator is the presence of the tuning parameter, i.e. bandwidth. It is well-known that under context-specific assumptions, asymptotic linearity of kernel-based semiparametric estimators often leads to an asymptotic normal distribution with variance invariant to the choice of bandwidth and kernel. This suggests these asymptotic approximations are too coarse to capture the finite-sample effect of the specificity of the bandwidth and kernel. CCJ establishes an asymptotic normal approximation of density-weighted average derivative estimator without relying on asymptotic linearity. In particular, they show a more refined asymptotic quadratic approximation is valid under a broader range of bandwidth sequences and establish asymptotic normality of the quadratic objects with martingale central limit theorem.

The main result of this paper shows the asymptotic normality of this estimator under both nondegeneracy and degeneracy and across a broad range of bandwidth sequences. In other words, both challenges mentioned previously are met. Toward this end, I first set up the population and sampling framework for dyadic data. Compared to previous literature, this setup makes a clear distinction between monadic regressors and dyadic regressors, which have different convergence rates under some bandwidth sequences. Then, I extend the density-weighted average derivative estimator to the dyadic setting. Compared to the estimator in the cross-sectional setting, this extension features a two-way leave-out estimator of the nuisance function. This resonates with the two-way cross fitting proposed in Chiang et al. (2019). I then give conditions for an asymptotic quadratic representation of the estimator and show its asymptotic normality. The technical tool behind this result is first developed by CCJ and used for small-bandwidth asymptotics in the cross-sectional setting. In the dyadic setting, this approach is especially appealing because it automatically handles the degenerate case. Intuitively, even though degeneracy means the asymptotically linear part of this estimator is zero, the more refined asymptotic quadratic representation can still capture the leading variation of the estimator.

The robust asymptotic normality result is novel and captures a key distinction of kernel-based semi/nonparametric estimators from the sample mean for dyadic data. While the sample means may converge asymptotically to gaussian chaos under degeneracy, the density-weighted average derivatives estimator is asymptotically normal under both nondegeneracy and degeneracy. For dyadic data, GNP first shows a similar result for kernel-based density estimator. Hall (1984) first points out degenerate U-statistics with variable kernel may exhibit asymptotic normality. These results are interesting as they make robust inference for these kernel-based estimators much easier than that of the sample mean and statistics asymptotically equivalent to it. Moreover, it is reasonable to conjecture that the inference method of this estimator based on normal approximation may be uniformly valid and asymptotically nonconservative.

Section 2 lays out the population and sampling framework. Section 3 introduces the target parameter, index coefficient, and specifies the estimator, density-weighted average derivative. Section 4 explains the assumptions. Section 5 presents the main theoretical results, asymptotic normality of the estimator.

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,

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

for every permutation $\pi \in \Pi$ where $\pi : \{1, 2, \dots, N\} \to \{1, 2, \dots, N\}$ 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 $[X_{\sigma_X(i)\sigma_X(j)}] \stackrel{d}{=} [X_{ij}]$:

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

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 $(V_{ij}, V_{ji}) \stackrel{d}{=} (V_{ji}, V_{ij})$ and $(\epsilon_{ij}, \epsilon_{ji}) \stackrel{d}{=} (\epsilon_{ji}, \epsilon_{ij})$. 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 (2020b), 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 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:

$$\mathbb{E}\left[Y_{ij} | B_i, C_i, B_j, C_j, D_{ij}, D_{ji}\right] = \mathbb{E}[Y_{ij} | \underbrace{B_i, C_j, D_{ij}}_{X_{ij}}] = g\left(X_{ij}\right). \tag{1}$$

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 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

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

[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.

This representation motivates the density weighted average derivative estimator

$$\hat{\theta}_N = -2\frac{1}{N(N-1)}\sum_{i\neq j} Y_{ij}\frac{\partial}{\partial X_{ij}}\hat{f}_{N,ij}\left(X_{ij}\right), \tag{3}$$

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

$$\hat{f}_{N,ij}(x) = \frac{1}{(N-2)(N-3)}\sum_{\substack{(l,m)\\ l\neq m\neq i\neq j}}\frac{1}{h_N^d}K\left(\frac{x-X_{lm}}{h_N}\right). \tag{4}$$

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:

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

Write the estimator as a "U"-statistics:

$$\hat{\theta}_N(h) = -2\binom{N}{4}^{-1}\sum_{i<j<l<m} p_{ijlm}(h). \tag{5}$$

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

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

[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.

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.

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

(b) Ω *as defined in lemma 7 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 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

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

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 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 (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 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 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 $Nh_N^{\min\left\{\frac{d_B+d_D+2}{2},\frac{d_B+2d_D+2}{3}\right\}} \to \infty$.

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 1, 2 below are imported from PSS. Lemma 1 gives us the representation motivating the estimator. Lemma 2 bounds the order of magnitude of the bias. Their proofs are in the appendix.

Lemma 1 (Representation)**.** *Given Assumptions 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 2 (Bias)**.** *Given Assumption 1, 2 and $h_N \to 0$,*

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

Lemma 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 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 (Order of Magnitude of the Hoeffding Decomposition)**.** *Given Assumption 1, 2,* *and $h_N \to 0$, the order of magnitude of terms in the Hoeffding decomposition are*

$$\mathrm{Var}\left(\sqrt{N}U_{N,1}\right) = \frac{1}{16}\Sigma + O\left(h_N^Q\right) \tag{6}$$

$$\mathrm{Var}\left(\sqrt{\binom{N}{2}}\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,2}(h_N)\right) = \frac{1}{36}\Omega + O\left(h_N\right) \tag{7}$$

$$\mathrm{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,3}\right) = O\left(N^{-3}\right) \tag{8}$$

$$\mathrm{Var}\left(U_{N,4}\right) = N^{-4}O\left(h_N^{-(d_X+2)}\right). \tag{9}$$

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 4 (CLT)**.** *Given Assumption 1, 2, 3, 4,*

$$\begin{pmatrix} \sqrt{N}4U_{N,1}(h_N) \\ \sqrt{\binom{N}{2}}\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} 6U_{N,2}(h_N) \end{pmatrix} \rightsquigarrow N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & 0 \\ 0 & \Omega \end{pmatrix}\right). \tag{10}$$

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

To prepare, notice lemma 6 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 (10) holds once

$$\sqrt{N}4u_{N,1} + \sqrt{\binom{N}{2}}6u_{N,2}(h_N) \rightsquigarrow \mathrm{N}\left(0, \sigma^2 + \omega^2\right) \tag{11}$$

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 (11) by invoking the theorem of Eubank and Wang (1999). In our notation, conditions (1.3)-(1.6) of Eubank and Wang (1999) are

$$\binom{N}{2}^{-1}\max_{1\leq j\leq N}\sum_{i=1}^{N}\mathbb{E}\left[w_{N,ij}^2\right] \to 0 \tag{12}$$

$$\binom{N}{2}^{2}\mathbb{E}u_{N,2}^4 \to 3\omega^4 \tag{13}$$

$$N^{-2}\sum_{i=1}^{N}\mathbb{E}\left[l_i^4\right] \to 0 \tag{14}$$

$$\binom{N}{2}^{-1}N^{-1}\mathbb{E}\left[\left(\sum_{j=2}^{N}\sum_{i=1}^{j-1}\mathbb{E}\left(w_{N,ij}l_j|\mathcal{F}_{\{1,\ldots,j-1\}}\right)\right)^2\right] \to 0. \tag{15}$$

Because of exchangeability, (12) is equivalent to $N^{-1}\mathbb{E}\left[w_{N,ij}^2\right] \to 0$, which is satisfied because of

$$\mathbb{E}\left[w_{N,ij}^2\right] = \lambda_2' \operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N)\right)\lambda_2 \to \lambda_2'\Omega\lambda_2 = \omega^2 \tag{16}$$

by lemma (8).

By de Jong (1987) Prop. 3.1, condition (13) is satisfied if

$$N^{-2}\mathbb{E}[w_{N,ij}^4] \to 0 \tag{17}$$
$$N^{-1}\mathbb{E}[w_{N,ij}^2 w_{N,il}^2] \to 0 \tag{18}$$
$$\mathbb{E}[w_{N,ij}w_{N,il}w_{N,jm}w_{N,lm}] \to 0 \tag{19}$$
$$\mathbb{E}[w_{N,ij}^2] \to \omega^2 \tag{20}$$

We verify each of the condition one by one.
(17): $\mathbb{E}[w_{N,ij}^4] = h_N^{-d_B}$ by change of variable and (17) holds if $N^2 h_N^{d_B} \to \infty$, which is ensured by assumption 4.
(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 (18) holds because $N^{-1}O(1) = o(1)$.
(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.
(20): This is the same as (16).
Condition (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 (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, (15) is satisfied. □

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

Theorem 1 (Asymptotic Quadratic Approximation)**.** *Given Assumptions 1, 2, 3, 4, the second-order approximation of U_N dominates the approximation error,*

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

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.

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 2 tells us the bias square is bounded above by $O\left(h_N^{2Q}\right)$. Lemma 3 together with assumption 1 (b) tells us the variance of the second-order term $\mathrm{Var}\left(U_{N,2}\right)$ is of order at least $N^{-2}h_N^{d_B}$. These together with the bandwidth condition assumption 3 ensures $\left(-2U_{N,0} - \theta_0\right)^2 = o\left(\mathrm{Var}\left(U_{N,2}\right)\right)$ and $\left(-2U_{N,0} - \theta_0\right)^2 = o\left(\mathrm{Var}\left(U_N\right)\right)$. As a result, bias plays an asymptotically negligible role.

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

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

Under Assumption 1 (b) about second order nondegeneracy and assumption 4, lemma 3 ensures the second-order term dominates the third-order term and the fourth-order term $\mathrm{Var}\left(U_{N,3}(h_N)\right) = o\left(\mathrm{Var}\left(U_{N,2}(h_N)\right)\right)$ and $\mathrm{Var}\left(U_{N,4}(h_N)\right) = o\left(\mathrm{Var}\left(U_{N,2}(h_N)\right)\right)$. 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 4 together with the fact that

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

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,

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

$\square$

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 1.

Variance Estimation

Motivated by the variance expression

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

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

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

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

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

We'd like to show

$$N\hat{V}_1 \xrightarrow{\text{P}} \Sigma \tag{21}$$

$$\binom{N}{2} h_N^{d_B} \begin{pmatrix} h_N & 0 & 0 \\ 0 & h_N & 0 \\ 0 & 0 & 1 \end{pmatrix} \hat{V}_2 \begin{pmatrix} h_N & 0 & 0 \\ 0 & h_N & 0 \\ 0 & 0 & 1 \end{pmatrix} \xrightarrow{\text{P}} \Omega \tag{22}$$

$$\hat{V}_3 + \hat{V}_4 = o_P\left(N^{-1}\Sigma + \binom{N}{2}^{-1} h_N^{-d_B} \begin{pmatrix} h_N^{-1} & 0 & 0 \\ 0 & h_N^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \Omega \begin{pmatrix} h_N^{-1} & 0 & 0 \\ 0 & h_N^{-1} & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \tag{23}$$

Completion of this proof outline is left for further research.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Aldous, D. J. (1981). Representations for partially exchangeable arrays of random variables. *Journal of Multivariate Analysis*, 11(4):581–598.

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.

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.

Ductor, L., Fafchamps, M., Goyal, S., and Van der Leij, M. J. (2014). Social networks and research output. *Review of Economics and Statistics*, 96(5):936–948.

Fafchamps, M. and Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2):326–350.

Fafchamps, M. and Lund, S. (2003). Risk-sharing networks in rural philippines. *Journal of Development Economics*, 71(2):261–287.

Graham, B. S. (2020a). Dyadic regression. In *The Econometric Analysis of Network Data*, pages 23–40. Elsevier.

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

Hall, P. (1984). Integrated Square Error Properties of Kernel Estimators of Regression Functions. *The Annals of Statistics*, 12(1):241 – 260.

Hanushek, E. (1971). Teacher characteristics and gains in student achievement: Estimation using micro data. *The American Economic Review*, 61(2):280–288.

Hoover, D. N. (1979). Relations on probability spaces and arrays of random variables. *Technical Report, Institute for Advanced Study, Princeton, NJ.*

Menzel, K. (2021). Bootstrap with cluster-dependence in two or more dimensions. *Econometrica*.

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409.

Nishiyama, Y. and Robinson, P. M. (1999). Studentization in edgworth expansions for estimates of semiparametric index models. In *Nonlinear Statistical Modeling (Festschrift for Takeshi Amemiya)*, pages 197–240. Cambridge University Press.

Nishiyama, Y. and Robinson, P. M. (2000). Edgeworth expansions for semiparametric averaged derivatives. *Econometrica*, 68(4):931–979.

Nishiyama, Y. and Robinson, P. M. (2005). The bootstrap and the edgeworth correction for semiparametric averaged derivatives. *Econometrica*, 73(3):903–948.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, 57(6):1403–1430.

Tinbergen, J. (1962). *Shaping the World Economy: Suggestions for an International Economic Policy*. Twentieth Century Fund.

Appendix: Proofs

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

Proof of Lemma 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 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

$$
\begin{aligned}
&\mathbb{E}\left[\frac{\partial}{\partial X_{ij,1}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-X_{lm}}{h_N}\right)|X_{ij}\right] \\
&= \int \frac{\partial}{\partial X_{ij,1}}\frac{1}{h_N^{d_X}}K\left(\frac{X_{ij}-x}{h_N}\right)f(x)dx \\
&= \int\int \frac{1}{h_N}\frac{\partial}{\partial u_1}K\left(u_1,u_{-1}\right)f(X_{ij,1}-h_N u_1, X_{ij,-1}-h_N u_{-1})du_1 du_{-1} \\
&= \int K\left(u\right)\frac{\partial}{\partial (X_{ij,1}-h_N u)}f(X_{ij}-h_N u)du \\
&= \frac{\partial}{\partial X_{ij,1}}f(X_{ij}) + O\left(h^Q\right),
\end{aligned}
$$

in which the third equality follows by integration by part:

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

□

Proof of Lemma 3

To prove lemma 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 5. Then we use these result to calculate and bound corresponding conditional second moments of $p_{ijlm}(h)$ in lemma 7 and variance of $q_{N,1,i}, q_{N,2,ij}, q_{N,3,ijl}, q_{N,4,ijlm}$ in lemma 8. Lemma 3 is a natural result of lemma 8.

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 5. *Under Assumption 1, 2, and $h_N \to 0$,*

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

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

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

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

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

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

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

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

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

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

(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)

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

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 (5) directly implies the following lemma (6).

Lemma 6. *Under Assumption 1, 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 (5) and the definition of $p_{ijlm}(h_N)$ $\square$

Lemma 7. *Under Assumption 1, 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 $\Omega = \begin{pmatrix} \Omega_{BB} & \Omega_{BC} & \Omega_{BD} \\ \Omega_{BC}^T & \Omega_{CC} & \Omega_{CD} \\ \Omega_{BD}^T & \Omega_{CD}^T & \Omega_{DD} \end{pmatrix}$.

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

in which

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

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

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

$$
\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) \Bigg[\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\} \\
&\qquad + \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\} \Bigg] \\
&\qquad - K_C\left(\frac{C_i - C_j}{h_N}\right) \Bigg[\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\} \\
&\qquad + \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\} \Bigg] \Bigg) \\
&= \frac{1}{4} \int K_B^2(b) db \mathbb{E}\Bigg(\Bigg[\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\} \\
&\qquad + \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\} \Bigg]^2 \Bigg| B_i - B_j = 0 \Bigg) \\
&\quad + \frac{1}{4} \int K_C^2(c) dc \mathbb{E}\Bigg(\Bigg[\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\} \\
&\qquad + \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\} \Bigg]^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) \Bigg[\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\} \\
&\qquad + \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\} \Bigg], \\
&\qquad - K_C\left(\frac{C_i - C_j}{h_N}\right) \Bigg[\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\} \\
&\qquad + \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\} \Bigg] \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 (6)(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:

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

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.

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

□

Lemma 8. *Under Assumption 1, 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)
$$\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{3,ijl}(h_N)\right) \le \operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} \mathbb{E}\left(p_{ijlm}(h_N)|\mathcal{F}_{\{i,j,l\}}\right)\right) = O(1).$$

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

□

Proof of lemma 3

Proof. These are results of lemma 8 together with the facts

$$\operatorname{Var}\left(U_{N,1}(h_N)\right) = N^{-1}\operatorname{Var}\left(q_{N,1,i}^2\right)$$
$$\operatorname{Var}\left(\sqrt{\binom{N}{2}}\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,2}(h_N)\right) = \operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N)\right)$$
$$\operatorname{Var}\left(\begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} U_{N,3}(h_N)\right) = \binom{N}{3}^{-1}\operatorname{Var}\left(q_{3,ijl}(h_N)\right)$$
$$\operatorname{Var}\left(U_{N,4}(h_N)\right) = \binom{N}{4}^{-1}\mathbb{E}\left(q_{N,4,ijlm}^2\right).$$

□