

# Density-Weighted Average Derivatives for Dyadic Data

Fengshi Niu

November 14, 2021

[Preliminary and incomplete comments welcome but please don't cite]

## Abstract

We study the density-weighted average derivatives, a direct estimator of index coefficients in a semiparametric single index regression model proposed by Powell, Stock, Stoker (1988), for dyadic data. This type of data attracts increasing applied interest and exhibits dyadic dependency: variables of dyads with non-overlap monads are independent from each other and variables of dyads with nonempty overlap are allowed to be dependent. Borrowing the technical tools developed by Cattaneo, Crump, and Jansson (2014) for small-bandwidth asymptotics, we show that even though the presence/absence of dyadic dependency and the choice of small/large bandwidth sequence may affect the convergence rate of this estimator, its asymptotical normality holds robustly under mild condition across these scenarios. This result presents an interesting contrast between kernel-based semiparametric estimator and the sample mean of dyadic data, which exhibits asymptotic non-normality when dyadic dependency is absent and whose uniform nonconservative inference procedure don't exist (Menzel (2017)).

## 1 Introduction

Many social and economic variables have a natural underlying network structure. One set of notable examples falling into the category of *dyadic data* are trade flow among countries (e.g., 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 (cf., Fafchamps & Lund, 2003; Fafchamps & Gubert, 2007), and friendship relationship between individuals (e.g., Christakis et al., 2010). In these examples, the primary interest 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 score among matched students and teachers (Hanushek 1971), and two-way panel data. In these examples, the primary interest are variables on a bipartite network between *two* dimensions, e.g. individual and time, or two groups of agents.

Both dyadic data and two-way clustered data exhibit *dyadic dependency*, i.e. variables of pairs sharing one or two indices are allowed to be dependent and those don't share any indices are independent. Sources of dyadic dependency could often be *unobserved heterogeneity* in microeconomic data and 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 a same country could be positively correlated because of 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 dependency are now only emerging. Recent efforts in this direction include Chiang et al. (2020), which extends double machine learning method to multiway clustering data and proposes robust standard errors, and Graham et al. (2019), which studies the nonparametric kernel density estimator for dyadic data.

In this paper, I extend and study the classical semiparametric estimator, kernel-based density weighted average derivative in Powell, Stock, Stoker (1989) henceforth PSS, to data subject to dyadic dependency. 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 (1989) proposes this estimator establishes its asymptotic normality. Robinson (1995) and Nishiyama and Robinson (2000, 2001, 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, Crump, and Jansson (2014a, b) 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 dependency 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 dependency being present, and is the much faster  $\sqrt{\binom{N}{2}}$  under degeneracy, i.e. the dyadic dependency being absent. Intuitively, the presence and absence of dyadic dependency pivots the effective sample size between the number of agent 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 2017), and the kernel density estimators (GNP 2019). Because the limit distribution is very sensitive to the level of dependency, inference method valid uniformly is particularly worth pursuing. However, Menzel (2017) 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 tuning parameter, i.e. bandwidth. It is well-known that under context specific assumptions, asymptotic linearity of kernel-based semiparametric estimators often leads to a limit Gaussian 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. Seminal paper CCJ establishes 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 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 rate 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 first-step two-way leave-out estimator of the nuisance function. This resonates the two-way cross fitting proposed in Chiang et al. (2020). I then give conditions for an asymptotic quadratic representation of the estimator

and shows 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 mean 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 (2019) first shows similar result for kernel-based density estimator. Hall (1984) first points out degenerate U-statistics with variable kernel may exhibits asymptotic normality. These results are interesting as they make robust inference for these kernel-based estimators much more easier than that of the sample mean and statistics asymptotically equivalent to it. Moreover, it is reasonable to conjecture that inference method of this estimator based on normal approximation may be uniformly valid and asymptotically nonconservative.

The next section 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. Section 6 concludes.

## 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  $\pi : \{1, 2, \dots, N\} \rightarrow \{1, 2, \dots, N\}$  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  $\sigma_X : \mathbb{N} \rightarrow \mathbb{N}$  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 doesn't depend on the node labelling. Third, outcomes of dyads sharing zero index are independent from each other. Outcomes of dyads sharing one index, like  $Y_{ij}$  and  $Y_{il}$ , are correlated with each other even conditional on the observed covariates. This dyadic dependence is crucial for the statistical analysis. For many statistical procedures, the dyadic dependence renders the effective sample size to be the number of nodes instead 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

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

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

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

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 (2020) and Menzel (2017).

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 serves as elements of the exchangeable array representation and are not necessarily substantive on their own. Interestingly, we will see later on the monadic and dyadic regressors may exhibit different rates of convergence.

### 3 Estimand and Estimator

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

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  $X_{ij} := (B_i, C_j, D_{ij}) \in \mathbb{R}^{d_X}$ ,<sup>1</sup>  $d_X = d_B + d_C + d_D$ .  $B_i$  captures the ego-relevant effect,  $C_j$  captures the alter-relevant effect,  $D_{ij}$  captures the dyadic effect:

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

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:  $g(x) = G(x'\beta_0)$ . 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  $\beta_0$ . 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 PSS

---

<sup>1</sup>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.

the following representation holds

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

This representation motivates the density weighted average derivative estimator

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

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

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

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 resonate with the multiway cross fitting procedure in Chiang et.al (2020), and facilitates setting up the estimator as a fourth-order ‘‘U’’-statistics<sup>2</sup>.

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  $i, j, l, m$ :

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

Write the estimator as a ‘‘U’’-statistics:

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

We will study its Hoeffding decomposition to understand its asymptotic behavior. Toward

---

<sup>2</sup>The quotation marks around the U indicates that the statistics is not strictly speaking a U-statistics in the traditional sense as an average of symmetric functions of combinations of iid random variables because of the presence of dyadic primitive variables  $\epsilon$  and  $V$  in this statistics. However, this won't make the analysis of this statistics any different from a U-statistics as will be revealed by the Hoeffding decomposition.

this end, define the information sets for one, two, and three indices as

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

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

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

with

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

Namely  $U_{N,0}(h) = \mathbb{E}(p_{ijlm}(h))$  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  $\theta_0$  is well-defined and the estimator  $\hat{\theta}_N$  is well-behaved.

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

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

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

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

The  $\Omega$  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

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

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

**Assumption 2** (Kernel). (a)  $K$  follows a product form: for  $x = (b, c, d)$  where  $b \in \mathbb{R}^{d_B}$ ,  $c \in \mathbb{R}^{d_C}$ ,  $d \in \mathbb{R}^{d_D}$ ,  $K(x) = K_B(b)K_C(c)K_D(d)$ .  $\int_{\mathbb{R}^{d_B}} K_B(b)db = \int_{\mathbb{R}^{d_C}} K_C(c)dc = \int_{\mathbb{R}^{d_D}} K_D(d)dd = 1$

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

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

**Assumption 3** (Bias).  $Nh_N^{Q + \frac{d_B}{2}} \rightarrow 0$ .

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).  $Nh_N^{\frac{d_B + d_D + 2}{2}} \rightarrow \infty$

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 that for asymptotic linearity. To see this note  $\frac{d_B + d_D + 2}{2} = \frac{1}{4}(d_B + 2) + \frac{3}{4}\left(\frac{d_B + 2d_D + 2}{3}\right)$  is a convex combination of  $d_B + 2$  and  $\frac{d_B + 2d_D + 2}{3}$  and is weakly smaller than their max. Also, this condition ensures the second-order term always dominates the fourth-order term both under degeneracy and nondegeneracy. Under nondegeneracy, this condition can be weakened to  $Nh_N^{\min\{\frac{d_B + d_D + 2}{2}, \frac{d_B + 2d_D + 2}{3}\}} \rightarrow \infty$ .

## 5 Results

This section presents the main theoretical results. The final result shows asymptotic normality of the estimator robust to nondegeneracy and small bandwidth sequences.

Both lemma 1, 2 below are imported from PSS. 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  $h_N \rightarrow 0$ ,*

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

Lemma 3 below involves calculation specific for the dyadic data. 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  $\mathbb{E}[p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}]$ ,  $\mathbb{E}[p_{ijlm}(h_N)|\mathcal{F}_{\{i,j\}}]$ ,  $\mathbb{E}[p_{ijlm}(h_N)|\mathcal{F}_{\{i,j,l\}}]$ , 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  $h_N \rightarrow 0$ , the order of magnitude of terms in the Hoeffding decomposition are*

$$\text{Var}(\sqrt{N}U_{N,1}) = \frac{1}{16}\Sigma + O(h_N^Q) \quad (6)$$

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

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

$$\text{Var}(U_{N,4}) = N^{-4}O(h_N^{-(d_X+2)}). \quad (9)$$

Proof of this lemma is in the appendix.

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 an order of  $N$ . The relative magnitude of the fourth-order term vs 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 vanishes. The concern of degeneracy motivates us to explore a more robust asymptotic approximation using a quadratic form, which incorporate the second-order term in addition to the first-order term. 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 (CCJ (2014)).

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

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

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

To prepare, notice lemma 6 shows that  $\mathbb{E}[p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}] = \frac{1}{4}\eta_i + O(h_N^Q)$ , in which  $\eta_i = \mathbb{E} \left[ -\frac{\partial}{\partial X_{ij}} g f(X_{ij}) - \frac{\partial}{\partial X_{ji}} g f(X_{ji}) + Y_{ij} \frac{\partial}{\partial X_{ij}} f(X_{ij}) + Y_{ji} \frac{\partial}{\partial X_{ji}} f(X_{ji}) \middle| \mathcal{F}_{\{i\}} \right]$ .  $\eta_i$  doesn't depend on  $N$ . This implies

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

because  $\sqrt{N}4U_{N,1}(h_N) - \frac{1}{\sqrt{N}} \sum_{i=1}^N (\eta_i - \mathbb{E}\eta_i) = \frac{1}{\sqrt{N}} \sum_{i=1}^N [4\mathbb{E}[p_{ijlm}(h_N)|\mathcal{F}_{\{i\}}] - \eta_i - (4\mathbb{E}[p_{ijlm}(h_N)] - \mathbb{E}\eta_i)]$  whose variance is bounded above by  $O(h_N^{2Q})$ .

The multivariate CLT (10) holds once

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

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

$$\begin{aligned} u_{N,1} &= \frac{1}{N} \sum_{i=1}^N l_i, \quad l_i = \lambda_1'(\eta_i - \mathbb{E}\eta_i), \quad \sigma^2 = \lambda_1' \Sigma \lambda_1, \\ u_{N,2} &= \binom{N}{2}^{-1} \sum_{i < j} w_{N,ij}, \quad w_{N,ij} = \lambda_2' \begin{pmatrix} h_N^{\frac{d_B+2}{2}} & 0 & 0 \\ 0 & h_N^{\frac{d_B+2}{2}} & 0 \\ 0 & 0 & h_N^{\frac{d_B}{2}} \end{pmatrix} q_{2,ij}(h_N), \quad \omega^2 = \lambda_2' \Omega \lambda_2. \end{aligned}$$

Assuming  $\lambda_1$  and  $\lambda_2$  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

$$\binom{N}{2}^{-1} \max_{1 \leq j \leq N} \sum_{i=1}^N \mathbb{E} [w_{N,ij}^2] \rightarrow 0 \quad (12)$$

$$\binom{N}{2}^2 \mathbb{E} u_{N,2}^4 \rightarrow 3\omega^4 \quad (13)$$

$$N^{-2} \sum_{i=1}^N \mathbb{E} [l_i^4] \rightarrow 0 \quad (14)$$

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

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

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

by lemma (8).

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

$$N^{-2} \mathbb{E} [w_{N,ij}^4] \rightarrow 0 \quad (17)$$

$$N^{-1} \mathbb{E} [w_{N,ij}^2 w_{N,il}^2] \rightarrow 0 \quad (18)$$

$$\mathbb{E} [w_{N,ij} w_{N,il} w_{N,jm} w_{N,lm}] \rightarrow 0 \quad (19)$$

$$\mathbb{E} [w_{N,ij}^2] \rightarrow \omega^2 \quad (20)$$

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  $N^2 h_N^{d_B} \rightarrow \infty$ , 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):  $\mathbb{E} [w_{N,ij} w_{N,il} w_{N,jm} w_{N,lm}] = \mathbb{E} [\mathbb{E} [w_{N,ij} w_{N,il} | \mathcal{F}_{\{j,l\}}]^2] = h_N^d \rightarrow 0$  [Deserves a detailed calculation]

(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  $\mathbb{E} \left[ \left( \mathbb{E} (w_{N,ij} l_j | \mathcal{F}_{\{i\}}) \right)^2 \right] \rightarrow 0$ . Since  $\mathbb{E} (w_{N,ij} l_j | \mathcal{F}_{\{i\}}) = h_N^{\frac{d_B}{2}}$

by integration by parts and bounding arguments, (15) is satisfied.  $\square$

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

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

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

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

in which the variance

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

*Proof.* Remember by definition the estimator  $\hat{\theta}_N = -2U_N$ . The Hoeffding decomposition 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  $O(h_N^2 Q)$ . Lemma 3 together with assumption 1 (b) tells us the variance of the second-order term  $\text{Var}(U_{N,2})$  is of order at least  $N^{-2}h_N^{d_B}$ . These together with assumption 3 ensures  $(-2U_{N,0} - \theta_0)^2 = o(\text{Var}(U_{N,2}))$  and  $(-2U_{N,0} - \theta_0)^2 = o(\text{Var}(U_N))$ . As a result, bias plays a asymptotically negligible role.

$$\text{Var}(\hat{\theta}_N)^{-1/2}(\hat{\theta}_N - \theta_0) = \text{Var}(U_N)^{-1/2} \frac{1}{2}(\hat{\theta}_N - \theta_0) = -\text{Var}(U_N)^{-1/2}(U_N - U_{N,0}) + o(1).$$

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

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

CLT in lemma 4 together with the fact that

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

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

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

□

## 6 Conclusion

Dyadic data and two-way clustering data has attracted lots of applied interest. Semiparametric estimation and its corresponding robust inference for these data lacks a theoretical investigation. In this paper, we have extended a canonical kernel-based semiparametric estimator, the density-weighted average derivative of PSS, to dyadic data, and proved its asymptotic normality under both nondegeneracy and degeneracy and across a wide range of bandwidth sequences. This result suggests we may expect inference procedures of this estimator based on asymptotic normality to be robust. This result presents an interesting contrast between kernel-based semiparametric estimator and the sample mean of dyadic data, which exhibits asymptotic non-normality when dyadic dependency is absent and whose uniform nonconservative inference procedure don't exist.

## Reference

- Abowd, J. M., Kramarz, F., & 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., Hortaçsu, A., Roberts, J., & Syverson, C. (2011). Network structure of production. *Proceedings of the National Academy of Sciences*, 108(13), 5199 – 5202.
- Cattaneo, M. D., Crump, R. K., & Jansson, M. (2014). Small bandwidth asymptotics for density-weighted average derivatives. *Econometric Theory*, 176-200.
- Cattaneo, M. D., Crump, R. K., & Jansson, M. (2014). Bootstrapping density-weighted average derivatives. *Econometric Theory*, 1135-1164.
- Chiang, H. D., Kato, K., Ma, Y., & Sasaki, Y. (2019). Multiway cluster robust double/debiased machine learning. *arXiv preprint arXiv:1909.03489*.
- Christakis, N., Fowler, J., Imbens, G. W., & Kalyanaraman, K. (2020). An empirical model for strategic network formation. In *The Econometric Analysis of Network Data* (pp. 123-148). Academic Press.
- Crane, H., & Towsner, H. (2018). Relatively exchangeable structures. *The Journal of Symbolic Logic*, 83(2), 416-442.
- Ductor, L., Fafchamps, M., & Goyal, S. (2014). Social networks and research output. *Review of Economics and Statistics*, 96(5), 936 – 948.
- Eubank, R. L., & Wang, S. (1999). A central limit theorem for the sum of generalized linear and quadratic forms. *Statistics: A Journal of Theoretical and Applied Statistics*, 33(1), 85-91.
- Fafchamps, M. & Gubert, F. (2007). The formation of risk sharing networks. *Journal of Development Economics*, 83(2), 326 – 350.
- Fafchamps, M. & Lund, S. (2003). Risk sharing networks in rural philippines. *Journal of Development Economics*, 71(2), 261 – 287.
- Graham, B. S. (2019). Network Data. *arXiv preprint arXiv:1912.06346*.



- Graham, B. S., Niu, F., & Powell, J. L. (2019). Kernel density estimation for undirected dyadic data. arXiv preprint arXiv:1907.13630.
- 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. (2017). Bootstrap with clustering in two or more dimensions. arXiv preprint arXiv:1703.03043.
- Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2), 404 – 409.
- Nishiyama, Y. & P.M. Robinson (2000) Edgeworth expansions for semiparametric averaged derivatives. *Econometrica* 68, 931–979.
- Nishiyama, Y. & P.M. Robinson (2001) Studentization in edgeworth expansions for estimates of semiparametric index models. In C. Hsiao, K. Morimune, & J.L. Powell (eds.), *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*, pp. 197–240. Cambridge University Press.
- Nishiyama, Y. & P.M. Robinson (2005) The bootstrap and the Edgeworth correction for semiparametric averaged derivatives. *Econometrica* 73, 903–948.
- Powell, J. L., Stock, J. H., & Stoker, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica: Journal of the Econometric Society*, 1403-1430.
- Tinbergen, J. (1962). *Shaping the world economy; suggestions for an international economic policy*. New York: Twentieth Century Fund.

## Appendix: Proofs

### 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

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

□

### Proof of Lemma 2

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

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

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

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

in which the third equality follows by integration by part:

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

□

### Proof of Lemma 3

To prove lemma 3, we will need to calculate  $\mathbb{E}[p_{ijlm}(h)|\mathcal{F}_{\{i\}}]$ ,  $\mathbb{E}[p_{ijlm}(h)|\mathcal{F}_{\{i,j\}}]$ ,  $\mathbb{E}[p_{ijlm}(h)|\mathcal{F}_{\{i,j,l\}}]$ . 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  $h_N \rightarrow 0$ ,*

$$(a) \quad \mathbb{E} \left[ \delta_{ijlm}(h_N) \middle| \mathcal{F}_{\{j,l,m\}} \right]$$

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

$$\begin{aligned}
(b) \quad &\mathbb{E} [\delta_{ijlm}(h_N) | \mathcal{F}_{\{i, j\}}] = Y_{ij} \frac{\partial}{\partial X_{ij}} f(X_{ij}) + O(h_N^Q), \\
&\mathbb{E} [\delta_{ijlm}(h_N) | \mathcal{F}_{\{i, l\}}] = \left( \frac{1}{h_N^{d_B+1}} \dot{K}_B \left( \frac{B_i - B_l}{h_N} \right) [\mathbb{E} [Y_{ij} | C_j - C_m = 0, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{i, l\}}] f_{C_j - C_m, D_{ij} - D_{lm} | \mathcal{F}_{\{i, l\}}} (0, 0 | \mathcal{F}_{\{i, l\}}) + O(h_N)] \right. \\
&\quad \left. \frac{1}{h_N^{d_B}} K_B \left( \frac{B_i - B_l}{h_N} \right) \left[ -\frac{\partial}{\partial t} \Big|_{t=0} \mathbb{E} [Y_{ij} | C_j - C_m = t, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{i, l\}}] f_{C_j - C_m, D_{ij} - D_{lm} | \mathcal{F}_{\{i, l\}}} (t, 0 | \mathcal{F}_{\{i, l\}}) + O(h_N) \right] \right. \\
&\quad \left. \frac{1}{h_N^{d_B}} K_B \left( \frac{B_i - B_l}{h_N} \right) \left[ -\frac{\partial}{\partial t} \Big|_{t=0} \mathbb{E} [Y_{ij} | C_j - C_m = 0, D_{ij} - D_{lm} = t, \mathcal{F}_{\{i, l\}}] f_{C_j - C_m, D_{ij} - D_{lm} | \mathcal{F}_{\{i, l\}}} (0, t | \mathcal{F}_{\{i, l\}}) + O(h_N) \right] \right) \\
&\mathbb{E} [\delta_{ijlm}(h_N) | \mathcal{F}_{\{i, m\}}] = -\frac{\partial}{\partial t} \Big|_{t=0} \left\{ \mathbb{E} [Y_{ij} | X_{ij} - X_{lm} = t, \mathcal{F}_{\{i, m\}}] f_{X_{ij} - X_{lm} | \mathcal{F}_{\{i, m\}}} (t | \mathcal{F}_{\{i, m\}}) \right\} + \\
&O(h_N), \\
&\mathbb{E} [\delta_{ijlm}(h_N) | \mathcal{F}_{\{j, l\}}] = -\frac{\partial}{\partial t} \Big|_{t=0} \left\{ \mathbb{E} [Y_{ij} | X_{ij} - X_{lm} = t, \mathcal{F}_{\{j, l\}}] f_{X_{ij} - X_{lm} | \mathcal{F}_{\{j, l\}}} (t | \mathcal{F}_{\{j, l\}}) \right\} + \\
&O(h_N), \\
&\mathbb{E} [\delta_{ijlm}(h_N) | \mathcal{F}_{\{j, m\}}] = \left( \frac{1}{h_N^{d_C}} K_C \left( \frac{C_j - C_m}{h_N} \right) \left[ -\frac{\partial}{\partial t} \Big|_{t=0} \mathbb{E} [Y_{ij} | B_i - B_l = t, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{j, m\}}] f_{B_i - B_l, D_{ij} - D_{lm} | \mathcal{F}_{\{j, m\}}} (t, 0 | \mathcal{F}_{\{j, m\}}) + O(h_N) \right] \right. \\
&\quad \left. \frac{1}{h_N^{d_C+1}} \dot{K}_C \left( \frac{C_j - C_m}{h_N} \right) [\mathbb{E} [Y_{ij} | B_i - B_l = 0, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{j, m\}}] f_{B_i - B_l, D_{ij} - D_{lm} | \mathcal{F}_{\{j, m\}}} (0, 0 | \mathcal{F}_{\{j, m\}}) + O(h_N)] \right. \\
&\quad \left. \frac{1}{h_N^{d_C}} K_C \left( \frac{C_j - C_m}{h_N} \right) \left[ -\frac{\partial}{\partial t} \Big|_{t=0} \mathbb{E} [Y_{ij} | B_i - B_l = 0, D_{ij} - D_{lm} = t, \mathcal{F}_{\{j, m\}}] f_{B_i - B_l, D_{ij} - D_{lm} | \mathcal{F}_{\{j, m\}}} (0, t | \mathcal{F}_{\{j, m\}}) + O(h_N) \right] \right) \\
&\mathbb{E} [\delta_{ijlm}(h_N) | \mathcal{F}_{\{l, m\}}] = -\frac{\partial}{\partial X_{lm}} g f(X_{lm}) + O(h_N^Q).
\end{aligned}$$

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

*Proof.* (a)

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

Similarly,

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

$$\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] \\
&= \left( \frac{\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]} \right. \\
&\quad \left. \frac{\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]}{\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]} \right) \\
&= \left( \frac{\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} [Y_{ij} | B_i = B_l, D_{ij} = D_{lm}, \mathcal{F}_{\{j\}}] f_{B_i, D_{ij} | \mathcal{F}_{\{j\}}}(B_l, D_{lm} | \mathcal{F}_{\{j\}}) \right\} + 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} [Y_{ij} | B_i = B_l, D_{ij} = D_{lm}, \mathcal{F}_{\{j\}}] f_{B_i, D_{ij} | \mathcal{F}_{\{j\}}}(B_l, D_{lm} | \mathcal{F}_{\{j\}}) + O(h_N) \right]} \right. \\
&\quad \left. \frac{\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} [Y_{ij} | B_i = B_l, D_{ij} = D_{lm}, \mathcal{F}_{\{j\}}] f_{B_i, D_{ij} | \mathcal{F}_{\{j\}}}(B_l, D_{lm} | \mathcal{F}_{\{j\}}) \right\} + 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 D_{lm}} \left\{ \mathbb{E} [Y_{ij} | B_i = B_l, D_{ij} = D_{lm}, \mathcal{F}_{\{j\}}] f_{B_i, D_{ij} | \mathcal{F}_{\{j\}}}(B_l, D_{lm} | \mathcal{F}_{\{j\}}) \right\} + O(h_N) \right]} \right)
\end{aligned}$$

Similarly,

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

(b)

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

$$\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] \\ &= \left( \begin{aligned} & \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{aligned} \right) \\ &= \left( \begin{aligned} & \frac{1}{h_N^{d_B+1}} \dot{K}_B \left( \frac{B_i - B_l}{h_N} \right) \left[ \mathbb{E} [Y_{ij} | C_j - C_m = 0, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{i,l\}}] f_{C_j - C_m, D_{ij} - D_{lm} | \mathcal{F}_{\{i,l\}}} (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} \Big|_{t=0} \mathbb{E} [Y_{ij} | C_j - C_m = t, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{i,l\}}] 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} \Big|_{t=0} \mathbb{E} [Y_{ij} | C_j - C_m = 0, D_{ij} - D_{lm} = t, \mathcal{F}_{\{i,l\}}] f_{C_j - C_m, D_{ij} - D_{lm} | \mathcal{F}_{\{i,l\}}} (0, t | \mathcal{F}_{\{i,l\}}) + O(h_N) \right] \end{aligned} \right) \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] \\
&= \left( \frac{\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]} \right. \\
&\quad \left. \frac{\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]}{\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) \frac{\partial}{\partial D_{ij}} K_D \left( \frac{D_{ij} - D_{lm}}{h_N} \right) \middle| \mathcal{F}_{\{j,m\}} \right]} \right) \\
&= \left( \frac{\frac{1}{h_N^{d_C}} K_C \left( \frac{C_j - C_m}{h_N} \right) \left[ -\frac{\partial}{\partial t} \middle|_{t=0} \mathbb{E} [Y_{ij} | B_i - B_l = t, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{j,m\}}] 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} [Y_{ij} | B_i - B_l = 0, D_{ij} - D_{lm} = 0, \mathcal{F}_{\{j,m\}}] f_{B_i - B_l, D_{ij} - D_{lm} | \mathcal{F}_{\{j,m\}}} (0, 0 | \mathcal{F}_{\{j,m\}}) + O(h_N)] \right]} \right) \\
&\quad \left( \frac{\frac{1}{h_N^{d_C}} K_C \left( \frac{C_j - C_m}{h_N} \right) \left[ -\frac{\partial}{\partial t} \middle|_{t=0} \mathbb{E} [Y_{ij} | B_i - B_l = 0, D_{ij} - D_{lm} = t, \mathcal{F}_{\{j,m\}}] f_{B_i - B_l, D_{ij} - D_{lm} | \mathcal{F}_{\{j,m\}}} (0, t | \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[ -\frac{\partial}{\partial t} \middle|_{t=0} \mathbb{E} [Y_{ij} | B_i - B_l = 0, D_{ij} - D_{lm} = t, \mathcal{F}_{\{j,m\}}] f_{B_i - B_l, D_{ij} - D_{lm} | \mathcal{F}_{\{j,m\}}} (0, t | \mathcal{F}_{\{j,m\}}) + O(h_N)] \right]} \right)
\end{aligned}$$

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

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

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

(c)

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

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

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

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

□

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

**Lemma 6.** *Under Assumption 1, 2, and  $h_N \rightarrow 0$ ,*

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

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

(b)



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

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

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

(b) For  $d_B = d_C$ ,

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

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

$$\Omega = \frac{1}{4} \lim_{N \rightarrow \infty} \frac{1}{h_N^{d_B}} \text{Var} \left( \begin{pmatrix} K_B \left( \frac{B_i - B_j}{h_N} \right) \left\{ \mathbb{E} [Y_{il} | C_l - C_m = 0, D_{il} - D_{jm} = 0, \mathcal{F}_{\{i,j\}}] f_{C_l - C_m, D_{il} - D_{jm} | \mathcal{F}_{\{i,j\}}} (0, 0 | \mathcal{F}_{\{i,j\}}) - \mathbb{E} [Y_{il} | C_l - C_m = 0, D_{il} - D_{jm} = 0, \mathcal{F}_{\{i,j\}}] f_{C_l - C_m, D_{il} - D_{jm} | \mathcal{F}_{\{i,j\}}} (0, 0 | \mathcal{F}_{\{i,j\}}) + O(h_N) \right\} \\ K_C \left( \frac{C_i - C_j}{h_N} \right) \left\{ \mathbb{E} [Y_{il} | B_l - B_m = 0, D_{li} - D_{mj} = 0, \mathcal{F}_{\{i,j\}}] f_{B_l - B_m, D_{li} - D_{mj} | \mathcal{F}_{\{i,j\}}} (0, 0 | \mathcal{F}_{\{i,j\}}) - \mathbb{E} [Y_{il} | B_l - B_m = 0, D_{li} - D_{mj} = 0, \mathcal{F}_{\{i,j\}}] f_{B_l - B_m, D_{li} - D_{mj} | \mathcal{F}_{\{i,j\}}} (0, 0 | \mathcal{F}_{\{i,j\}}) + O(h_N) \right\} \\ - K_B \left( \frac{B_i - B_j}{h_N} \right) \left\{ \mathbb{E} [Y_{il} | C_l - C_m = 0, D_{il} - D_{jm} = t, \mathcal{F}_{\{i,j\}}] f_{C_l - C_m, D_{il} - D_{jm} | \mathcal{F}_{\{i,j\}}} (0, t | \mathcal{F}_{\{i,j\}}) \right\} + K_C \left( \frac{C_i - C_j}{h_N} \right) \left\{ \mathbb{E} [Y_{il} | B_l - B_m = 0, D_{li} - D_{mj} = t, \mathcal{F}_{\{i,j\}}] f_{B_l - B_m, D_{li} - D_{mj} | \mathcal{F}_{\{i,j\}}} (0, t | \mathcal{F}_{\{i,j\}}) \right\} + O(h_N) \end{pmatrix} \right)$$

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

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

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

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

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

*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:

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

Calculation for the middle  $d_C$  components and the last  $d_D$  components are conceptually similar. The covariance term in the definition of  $\Omega_{DD}$  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.

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

□

**Lemma 8.** Under Assumption 1, 2, and  $h_N \rightarrow 0$ ,

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

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

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

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

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

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

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

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

□

### Proof of lemma 3

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

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

□