2018-08-02

Estimating Equation Estimators for the Pair Correlation Function

Chong Zhao
University of Miami, zhaochongyct@gmail.com

UNIVERSITY OF MIAMI

ESTIMATING EQUATION ESTIMATORS FOR THE PAIR CORRELATION FUNCTION

By

Chong Zhao

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2018

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

ESTIMATING EQUATION ESTIMATORS FOR THE PAIR CORRELATION FUNCTION

Chong Zhao

Approved:

Yongtao Guan, Ph.D.
Professor of Management Science

Jingfei Zhang, Ph.D.
Assistant Professor of Management Science

Wei Sun, Ph.D.
Assistant Professor of Management Science

Guillermo Prado, Ph.D.
Dean of the Graduate School

Xiaodong Cai, Ph.D.
Professor of Electrical and Computer Engineering

CHONG, ZHAO (Ph.D., Management Science)
Estimating Equation Estimators for the Pair Correlation Function (August 2018)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Yongtao Guan.
No. of pages in text.(152)

The pair correlation function is an important and useful tool to explore spatial-temporal point pattern data. It determines the second order characteristics of spatial or temporal process and is often estimated by some nonparametric approach such as kernel smoothing. However, the estimating performance is highly dependent on the estimation of the first order intensity function and it is widely studied only in single point pattern scenario. An inappropriate estimated intensity function may lead to poor estimation for the pair correlation function. In this dissertation, we introduce two nonparametric estimators based on estimating equation technique in a replicated point patterns setting. The two estimators can be easily extended to semiparametric setting to incorporate the covariate(s) information. The two estimators have asymptotic consistency and normality according to our theoretical and simulation results. An empirical study where the points are users' tweets in a Twitter-type data illustrates the application of our estimators.

Acknowledgements

I would like to express the great appreciation to my advisor, the committee chair, Professor Yongtao Guan, who has guided me for five years with great patience, helpful suggestions, and consistent support. I am truly grateful for his constant encouragement and generous help in every aspect.

I am deeply grateful to Professor Jingfei Zhang and Professor Wei Sun for suggestions and help in my research and also job application. I appreciate all the tutoring and sharing the experience in research and writing papers. My sincere thanks also go to Professor Xiaodong Cai for joining my dissertation committee and providing comments and advice. In addition, I thank all other faculty members in the department for their great lectures and sincere help.

I also want to extend my gratitude to the help of other Ph.D. students, Ming Wang and Gang Xu, and my friends, Tianchi Zhang and Bohan Zhang, for discussion and inspiration on my research, and for all the fun we have had during the passing years.

Last but not at least, I would like to acknowledge my gratitude, the support, and love of my girlfriend, Yifan Sha. Special thanks to my family for their love, support, and sacrifices. I dedicate this thesis to my parents, Yunai Ma and Hengping Zhao, and to my sister, Hang Zhao.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

One major interest for analyzing point pattern is to determine whether the points being placed independently of each other, or whether they exhibit some kind of inter-point dependence. Actually, correlation, or generally covariance, is a standard statistical tool for measuring dependence. And it is classified as a second moment quantity. In fact, the second-moment quantities for point processes are intimately related to counting pairs of points or adding up contributions from each pair of points in the process. In this dissertation, we will focus on the correlation measure in point pattern and conduct novelty research in this area. More specifically this dissertation concentrates on proposing novelty estimating methods to describe correlation structure in replicated point patterns which are actually sets of point patterns instead of a single pattern.

If we look at the second order characteristics of point pattern more closely, we have Ripley's K-function (Ripley, 1976), Besag's L-function (Besag, 1977) and the pair correlation function $g(r)$ (Stoyan and Stoyan, 1994). Current literature has shown that the pair correlation function is a good way to statistical present the distribution and correlation no matter in single point pattern or replicated point patterns. This function has been regarded as the second-order fundamental functional summary on the dependence of the point pattern (Stoyan and Stoyan, 1994; Møller and Waagepetersen, 2003; Illian et al., 2008). While it contains the similar statistical information as the K and L function, it offers the information in a way which is easier to understand. The estimates of the pair correlation function are useful for assessing regularity or clustering of a spatial point pattern and can moreover be used for inferring parametric models for spatial point processes via minimum contrast estimation (Stoyan and Stoyan, 1994; Illian et al., 2008). In fact, modifications of the pair correlation function $g(r)$ have been widely used in some areas of science. Astronomers, for example, use what they call the 'correlation function', $\xi(r) = g(r)-1$ (Illian et al., 2008). Thus, it's very natural to consider estimate this function in these settings, actually, there is a lot of literature focusing on developing new methodologies to estimate the pair correlation function in a variety of settings.

While there are extensive methods for estimating the pair correlation function to analyze single point pattern datasets, little has been done in developing similar analysis methodologies for the replicated point patterns. More specially, replicated point patterns are datasets consisting of several point patterns which can be regarded as independent repetitions of the same experiment. Replication is vitally important in statistics because it enables us to observe variability and to separate different sources of variability (Baddeley et al., 2015). Also, replication enables us to analyze the difference in point pattern between the studying subjects and within each studying subject. In medical or materials science studies, several patterns are frequently considered at the same time instead of only a single sample of a point process is available with the aim of eventually analyzing them all together to extract information on the spatial or temporal behavior that is reflected in them (Illian et al., 2008). The data we are focusing on in this thesis is actually a typical replicated point patterns data, the SINA Weibo data. In this dataset, the random posts across timeline comprise of simple temporal point process for each user. Multiple users' temporal processes form a typical replicated point patterns. However, the limited literature in this field mainly focused on pooled summarized statistics and single point pattern analysis aggregation. Thus, the tools for analyzing replicated point patterns, especially with covariates, are still under development. Hence developing more suitable and efficient methodologies for this specific data structure are necessary and promising. In addition, techniques for analyzing replicated point patterns can also be pressed into service for analyzing a single point pattern (Baddeley et al., 2015).

Since studying replicated point pattern being vital and the pair correlation function being widely used, it is natural to extend this function to replicated point pattern settings. However, limited literature has been done to combine these together. In this dissertation, we proposed two nonparametric estimators for the pair correlation function based on the estimation equation technique in replicated point pattern setting. The estimators are easy to solve, enjoy good asymptotic properties and achieve extraordinary performance both in simulation and empirical application. The remainder of this thesis is organized as follows: In Chapter 2, a detailed and comprehensive literature review is being conducted in both replicated point pattern analysis and the pair correlation function estimation. In Chapter 3-4, we introduce two estimating equation estimators for the pair correlation function namely, the local linear estimator and the orthogonal estimator. A tuning parameter selection approach is illustrated in Chapter 5. In Chapter 6, a solid theoretical investigation is being conducted to access the asymptotic properties of the proposed estimators. We also proposed empirical variance estimators of the local constant estimator and orthogonal series estimator for inference purpose in this section. Then we conducted a simulation study comparing the performance of different methods in Chapter 7. In Chapter 8, we apply our proposed method to the Weibo dataset to validate the extraordinary performance of the proposed estimators. In Chapter 9, we extended the two proposed estimators to a semiparametric setting which make the estimators more general and applicable. In the last Chapter, we discussed the achieved properties for the proposed estimators. Further directions are being covered in this section.

Chapter 2

Literature Review

While there are extensive methods for analyzing single point pattern, little work has been done in developing analysis methodologies for replicated point pattern. The work reported here is motivated by a temporal point process dataset obtained from SINA Weibo (www.weibo.com) which is the largest Twitter-type social media service in China. This dataset has a lot of independent accounts with events comprise independent temporal point processes. In fact, more and more replicated point pattern data are collected and need to be analyzed in various areas, for example, in clinical neuroanatomy, in material science, molecular pharmacology, etc (Diggle et al., 2000; Bell and Grunwald, 2004).

When replicated point pattern data is available, nonparametric methods using the replication within the experimental design as the basis for inference can be used. Diggle et al. (1991) and Baddeley et al. (1993) introduced nonparametric approaches based on estimating pooled summary statistics and further applied these pooled summary statistics to test whether spatial patterns are different across different groups.

The method was being further applied in Diggle et al. (2000) and Schladitz et al. (2003). Diggle et al. (2000) compared the use of the K-function ANOVA Diggle et al. (1991) with the maximum pseudo-likelihood methods (Baddeley and Turner, 2000) and found that the nonparametric methods outperform the parametric ones under misspecification of the original model setting. In fact, each of the above-mentioned papers uses fixed-effect models or separate estimates from each pattern and incorporates the replication with somewhat ad hoc methods which result in poor utilization of the variation across different subjects. Bell and Grunwald (2004) proposed a mixed parametric model with random effects to capture these variations. More detailedly, they fit pairwise interaction point process models using pseudo-likelihood maximization and accommodate differences between replicates by treating model parameters as random effects. However, parametric modeling assumptions would be hard to justify and any misspecification can result in unreliable analysis result, thus a nonparametric approach is preferred which is consistent with the conclusion obtained by Diggle et al. (2000). In this dissertation, we propose novel nonparametric methods estimating the pair correlation function to analyze the data of replicated point pattern.

Current literature has shown that the pair correlation function is a good way to statistically present the distribution and correlation in point pattern (Møller and Waagepetersen, 2003; Illian et al., 2008). Describing the dependence information of the process, the pair correlation function can be used to assess regularity or clustering property of a spatial or temporal point pattern. Thus estimating the pair correlation function, usually through nonparametric approaches, has experienced a wide explosion of interest over the last 20 years. Stoyan and Stoyan (1994) brought up a kernel-smoothed version estimator of the pair correlation function which is computationally fast and works well except at small spatial lags. It was subsequently used by a number of authors including Møller et al. (1998) and was discovered that the kernel estimator has a serious bias problem when spatial lags are close to zero, the detailed discussion can be found in Stoyan and Stoyan (1994) and Møller and Waagepetersen (2003). In fact, the properties of the kernel estimator are highly dependent on the choice of bandwidth. Any inappropriate selection of bandwidth may result in an estimator with large bias or variance or both. To avoid the inappropriate choice of h, some general guidelines on the selection of h was being proposed in stationary case. For instance, $h = c\lambda^{-1/2}$ with Epanechnickove kernel was recommended in Stoyan and Stoyan (1994). An approximation for the variance of the pair correlation function based method was introduced in Stoyan and Stoyan (2000). For inhomogeneous processes, Guan (2007a,b) proposed least-squares cross-validation and composite likelihood cross-validation procedures to obtain the optimal bandwidth while Loh and Jang (2010) suggested a bootstrap approach together with a pilot estimate. The bias is actually the main drawback when one attempts to utilize nonparametric estimate to infer a parametric model as the behavior near zero is vital for determining the correctness of parametric model specification (Jalilian et al., 2013). To remedy the bias in the Kernel estimator, Jalilian et al. (2017) adapt the orthogonal series density estimators (see the reviews in Hall, 1987; Efromovich, 2010) to estimate the pair correlation function. They derived unbiased estimators of the coefficients in an orthogonal series expansion of the pair correlation function and the estimators were found to be less biased under clustered point patterns settings. Additionally, Yue and Loh (2010) proposed an alternative estimator based on Bayesian nonparametric regression approach by assigning a Gaussian smoothness prior to the function space. However, all the above methods are partially dependent on the estimation of first-order intensity function. In fact, there are numerous cases that the point pattern only contains a lot of shot processes or only have extremely sparse events dispersion in a certain time period for each process within studying time window. All of these will result in a poor estimation for first order intensity in the current nonparametric setting. Thus the subsequent estimation of second order characteristics for point pattern will inevitably corrupt no matter how pooled summary or aggregation strategy being used for replicated point pattern. Thus, how to extend the current research to replicated point patterns becomes vital and the robustness of methods for varied data structure also needs to be considered.

In this dissertation, we propose two nonparametric methods to extend the estimation of the pair correlation function from single point pattern to replicated point patterns. The pair correlation function in the kernel estimator takes a local linear form and is based on an orthogonal series expansion form in the latter mentioned approach. To my knowledge, our methods are the first methodologies in estimating pair correlation function where the within and between subjects variations in replicated point patterns are incorporated into the estimation procedure in nonparametric settings. In fact, our novelty estimators possess great advantages over existing methods in two main aspects. Firstly, our estimators bypass the first-order intensity function estimation successfully during the pair correlation function estimation. On the contrary, the existing smoothing nonparametric methods may result in poor estimation of the pair correlation function since the first-order intensity function can be approaching zero for a fairly long period during a day (people barely have tweeting activities during the night) in the data structures similar to what we analyzing. Our estimators are less sensitive to the structural characteristics of the replicated point patterns than the named methods above. Another advantage of our estimators is the ability to obtain estimates efficiently by utilizing the estimating equations procedure.

It makes the analysis process easy, neat and efficient. In fact, the untransformed estimators can even get explicit solutions without any optimization computation. To illustrate these, an elaborate introduction is being covered to describe the two estimators. Asymptotic behaviors of the estimators are being accessed to make sure the good properties of the estimators. Also the empirical variance estimators are being proposed to make inference for our pair correlation function estimates. Similar to other nonparametric estimators, we propose a tuning parameter criterion for the optimal parameters selection. The performance of the two estimators is also being carefully investigated in the simulation study. After that, an empirical study is being conducted to obtain more insight into the data and also validate the performance of the estimators in real data application.

Chapter 3

Motivation and Background

Let $\{X_i : i = 1, \ldots, m\}$ be m independent point processes defined on $\mathbb{R}$. Let $D \subset \mathbb{R}$ be an observation window over which X_i's are observed. For $B \subseteq \mathbb{R}$, let $N_i(B)$ denote the number of random events in $X_i \cap B$. We assume that X_i's have a common intensity function $\lambda(x)$ and a common second-order intensity function $\lambda^{(2)}(x, y)$ so that for bounded $A, B \subseteq \mathbb{R}$,

$$
\begin{aligned}
E[N_i(B)] &= \int_B \lambda(x)dx, \\
E[N_i(A)N_i(B)] &= \int_{A \bigcap B} \lambda(x)dx + \int_A \int_B \lambda^{(2)}(x, y)dxdy.
\end{aligned}
$$

Assuming that $\lambda(x)\lambda(y) > 0$, the pair correlation function $g(x, y)$ is defined as

$$
g(x, y) = \frac{\lambda^{(2)}(x, y)}{\lambda(x)\lambda(y)}
$$

otherwise we define $g(x, y) = 0$. In this paper, we assume that $g(x, y)$ is isotropic, i.e., $g(x, y)$ is a function of the lag distance $|x - y|$. Often a parametric model is assumed for $g(x,y)$ involving a vector of unknown parameters θ. With a slight abuse of notation, we denote such a model by $g(r;\theta)$ with r being the lag distance.

To estimate θ, we consider a weighted composite likelihood estimation approach following Guan (2006) and Waagepetersen (2007). Suppose that we know $\lambda(x), x \in D$, we can then estimate θ by maximizing $L(\theta) = \prod_{i=1}^{m} L_i(\theta)$, where

$$L_i(\theta) = \prod_{u,v \in X_i}^{\neq} [\lambda(u)\lambda(v)g(|u-v|;\theta)]^{w(u,v)} \exp\left[-\int_D \int_D w(x,y)\lambda(x)\lambda(y)g(|x-y|;\theta)dxdy\right].$$

$\prod^{\neq}$ denotes product over all distinct points, and $w(x,y)$ is some predefined weight function. The associated score function for the (log) weighted composite likelihood is

$$U(\theta) = \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} w(u,v) \frac{g^{(1)}(|u-v|;\theta)}{g(|u-v|;\theta)} - m \int_D \int_D w(x,y)\lambda(x)\lambda(y)g^{(1)}(|x-y|;\theta)dxdy, \tag{3.1}$$

where $\sum^{\neq}$ denotes summation over all distinct points. Maximizing $L(\theta)$ is equivalent to solving $U(\theta) = 0$. The validity of this estimation approach lies in the fact that $U(\theta)$ is an unbiased estimating function, i.e., $E[U(\theta_0)] = 0$ where θ_0 is the true value of θ.

Without knowing $\lambda(x)$, the double integrals in (3.1) cannot be calculated and thus must be estimated. Note that for any real function $f(x,y)$,

$$E\left[\sum_{u \in X_i} \sum_{v \in X_j} f(u,v)\right] = \int_D \int_D f(x,y)\lambda(x)\lambda(y)dxdy, \quad \text{if } i \neq j.$$

We therefore modify $U(\theta)$ as

$$\tilde{U}(\theta) = \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} w(u,v) \frac{g^{(1)}(u-v;\theta)}{g(u-v;\theta)} - \frac{1}{m-1} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} w(u,v) g^{(1)}(u-v;\theta) \quad (3.2)$$

It can be easily seen that $E[\tilde{U}(\theta_0)] = 0$, i.e., $\tilde{U}(\theta)$ is also an unbiased estimating function. We may therefore estimate θ by solving $\tilde{U}(\theta) = 0$.

Following this line of thought, we propose nonparametric estimators for the pair correlation function for replicated point patterns. To do so, we first approximate the pair correlation function by a function specified by some unknown parameters and then estimate these parameters using the aforementioned weighted composite likelihood estimation approach.

It should be noted that although we assume a common intensity function $\lambda(x)$, our proposed procedures below can be easily extended to more general settings. For example, suppose that

$$\lambda_i(x;\beta) = \lambda_0(x)\rho[Z_i(x)'\beta], \ i = 1, \ldots, m, \quad (3.3)$$

where $\lambda_0(x)$ is an unspecified baseline intensity function, $\rho(x)$ is a known function, e.g., $\exp(x)$, $Z_i(x)$ is a set of potentially time-varying covariates for the ith point process, and β is a vector of unknown parameters. The regression parameter β can be estimated using the method in Lawless and Nadeau (1995) and Lin et al. (2000) without having to specify $\lambda_0(x)$. Let $\hat{\beta}$ be the resulting estimator. Then by modifying

(3.2) as

$$\tilde{U}(\theta) = \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} w(u,v) \frac{g^{(1)}(u-v;\theta)}{g(u-v;\theta)\rho[Z_i(u)'\hat{\beta}]\rho[Z_i(v)'\hat{\beta}]} - \frac{1}{m-1} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} w(u,v) \frac{g^{(1)}(u-v;\theta)}{\rho[Z_i(u)'\hat{\beta}]\rho[Z_j(v)'\hat{\beta}]}, \tag{3.4}$$

our proposed estimators below can be easily extended this new setting. We give the detailed expressions in the Chapter 9.

Chapter 4

Estimating Equation Estimators

Based on the intuition and background, I will illustrate the two proposed nonparametric estimating equation estimators in this chapter. Transformed version of the original proposed estimators are also being covered in this chapter.

4.1 Local Linear Estimator

We first propose a local linear estimator for the pair correlation function. For any lag t in a small neighborhood around r, we assume that

$$g(t;\theta) \approx \theta_1 + \theta_2(t-r), \tag{4.1}$$

where $\theta = (\theta_1, \theta_2)$. As in standard local linear estimations, our main interest is to estimate θ_1 and the obtained estimator will then be used as an estimator for the pair correlation function at lag r, i.e., $g(r)$.

Let $K(x)$ be a kernel function and define $K_h(x) = K(x/h)/h$ where h is a bandwidth. Define the weight function $w(u,v)$ in (3.2) as

$$w(u,v) = K_h(|u-v|-r)g(|u-v|;\theta)/|D\cap(D-u+v)|,$$

where $|D\cap(D-x)|$ is the volume of the intersection between D and $D-x \equiv \{y : x+y\in D\}$. Let $G(x) = (1,x)^T$. Then $\tilde{U}(\theta)$ in (3.2) becomes

$$\begin{aligned}\tilde{U}(\theta) =& \sum_{i=1}^{m}\sum_{u,v\in X_i}^{\neq}\frac{K_h(|u-v|-r)G(|u-v|-r)}{|D\cap(D-u+v)|} - \\ & \frac{1}{m-1}\sum_{i\neq j}\sum_{u\in X_i}\sum_{v\in X_j}\frac{K_h(|u-v|-r)G(|u-v|-r)}{|D\cap(D-u+v)|}[\theta_1+\theta_2(|u-v|-r)].\end{aligned} \tag{4.2}$$

The resulting estimating equations $\tilde{U}(\theta) = 0$ can be solved by solving a system of linear equations $A\theta = B$, where

$$\begin{aligned}A &= \frac{1}{m(m-1)}\sum_{i\neq j}\sum_{u\in X_i}\sum_{v\in X_j}\frac{K_h(|u-v|-r)G(|u-v|-r)G(|u-v|-r)^T}{|D\cap(D-u+v)|}, \\ B &= \frac{1}{m}\sum_{i=1}^{m}\sum_{u,v\in X_i}^{\neq}\frac{K_h(|u-v|-r)G(|u-v|-r)}{|D\cap(D-u+v)|}.\end{aligned} \tag{4.3}$$

By definition, the pair correlation function has to be nonnegative. However, the proposed estimator cannot guarantee this. To tackle this problem, we may instead assume $g(t;\theta) \approx \exp[\theta_1 + \theta_2(t-r)]$ for t in a small neighborhood around r. Now consider a new weight function,

$$w(u,v) = K_h(|u-v|-r)/|D\cap(D-u+v)|.$$

Then $\tilde{U}(\theta)$ in (3.2) becomes

$$\tilde{U}(\theta) = \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{K_h(|u-v|-r)G(|u-v|-r)}{|D \cap (D-u+v)|} - \frac{1}{m-1} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{K_h(|u-v|-r)G(|u-v|-r)}{|D \cap (D-u+v)|} \exp[\theta_1 + \theta_2(|u-v|-r)]. \tag{4.4}$$

We may then solve $\tilde{U}(\theta) = 0$ to obtain an estimate $\hat{\theta}_1$ for θ_1 and use $\exp(\hat{\theta}_1)$ as an approximation for $g(r)$. This guarantee $\hat{g}(r) \geq 0$. Compared with (4.2), however, $\tilde{U}(\theta) = 0$ now cannot be solved explicitly.

4.2 Orthogonal Series Estimator

For some predefined $R > 0$, let $\{\phi_k\}_{k \geq 1}$ be a set of orthonormal basis functions defined on $[0, R]$. One example of such basis functions is the cosine basis functions, i.e., $\phi_1(r) = \frac{1}{\sqrt{R}}$ and $\phi_k(r) = \frac{\sqrt{2}}{\sqrt{R}} \cos((k-1)\pi r/R)$ for $k \geq 2$. For $r \in [0, R]$, we consider a truncated orthogonal series approximation of $g(r)$ as follows:

$$g(r;\theta) \approx \sum_{k=1}^{L} \theta_k \phi_k(r), \tag{4.5}$$

where $\theta = (\theta_1, \ldots, \theta_L)$ and L is a predefined positive integer. Our main interest is to estimate θ and the resulting estimators can then be plugged in (4.5) to obtain an estimator for $g(r)$.

Let $I(\cdot)$ be an indicator function. We now define the weight function $w(u, v)$ in (3.2) as

$$w(u,v) = I(|u-v| \leq R) g(|u-v|;\theta)/|D \cap (D-u+v)|,$$

Let $\Phi = (\phi_1, \ldots, \phi_L)^T$. Then $\tilde{U}(\theta)$ in (3.2) becomes

$$\begin{aligned}\tilde{U}(\theta) =& \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{I(|u-v| \leq R)\Phi(|u-v|)}{|D \cap (D-u+v)|} - \\ & \frac{1}{m-1} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{I(|u-v| \leq R)\Phi(|u-v|)}{|D \cap (D-u+v)|} \left[\sum_{k=1}^{L} \theta_k \phi_k(|u-v|) \right].\end{aligned} \tag{4.6}$$

The resulting estimating equations $\tilde{U}(\theta) = 0$ can again be solved by solving a system of linear equations $A\theta = B$, except that now

$$\begin{aligned} A &= \frac{1}{m(m-1)} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{I(|u-v| \leq R)\Phi(|u-v|)\Phi(|u-v|)^T}{|D \cap (D-u+v)|}, \\ B &= \frac{1}{m} \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{I(|u-v| \leq R)\Phi(|u-v|)}{|D \cap (D-u+v)|}. \end{aligned} \tag{4.7}$$

Similar to the local linear estimator case, we can consider a truncated log-orthogonal series approximation as follows:

$$g(r;\theta) = \exp\left[\sum_{k=1}^{L} \theta_k \phi_k(r) \right]. \tag{4.8}$$

Now consider a new weight function,

$$w(u,v) = I(|u-v| \leq R)/|D \cap (D-u+v)|.$$

Then $\tilde{U}(\theta)$ in (3.2) becomes

$$\begin{aligned}\tilde{U}(\theta) = &\sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{I(|u-v| \leq R)\Phi(|u-v|)}{|D \cap (D-u+v)|} - \\ &\frac{1}{m-1} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{I(|u-v| \leq R)\Phi(|u-v|)}{|D \cap (D-u+v)|} \exp\left[\sum_{k=0}^{L} \theta_k \phi_k(|u-v|)\right].\end{aligned} \tag{4.9}$$

We may then solve $\tilde{U}(\theta) = 0$ to obtain an estimate $\hat{\theta}$ for θ and plug $\hat{\theta}$ into (4.8) in order to obtain an estimate for $g(r)$. This guarantees $\hat{g}(r) \geq 0$. Compared with (4.6), however, $\tilde{U}(\theta) = 0$ now can no longer be solved explicitly and can be difficult to solve when L is large.

Chapter 5

Parameter Tuning

As in any nonparametric method, the statistical properties of our proposed estimators are highly dependent on the choice of the respective tuning parameters, i.e., the bandwidth h for the local linear estimator or the number of basis functions L for the orthogonal series estimator. An inappropriate choice of h or L values may result in an estimator with a large bias or variance or both. In this section we describe a data driven method for the selection of the smoothing parameters.

Similar to Guan (2007b), we propose to choose the tuning parameter by using a least-square cross-validation method. We will focus on the local linear estimator to illustrate our approach. Let R be the largest lag for which the pair correlation function is to be estimated. Our proposed criterion intends to minimize the following least-square discrepancy measure:

$$Q(h) = \int_D \int_D \tilde{w}(x, y)\lambda(x)\lambda(y)[\hat{g}_h(|x - y|) - g(|x - y|)]^2 dxdy, \tag{5.1}$$

where

$$\tilde{w}(x,y) = \frac{I(|x-y| \le R)}{|D \cap (D-x+y)||x-y|}.$$

In the special case of $\lambda(x) = \lambda$ for all $x \in D$, $M(h)$ is simply

$$Q(h) = \lambda^2 \int_0^R [\hat{g}_h(r) - g(r)]^2 dr.$$

To estimate (5.1), we need to estimate the following two terms:

$$Q_1(h) = \int_D \int_D \tilde{w}(x,y)\lambda(x)\lambda(y)[\hat{g}_h(|x-y|)]^2 dxdy,$$

and

$$Q_2(h) = \int_D \int_D \tilde{w}(x,y)\lambda(x)\lambda(y)\hat{g}_h(|x-y|)g(|x-y|)dxdy.$$

To do so, let $\hat{g}_h^{-(u,v)}$ be $\hat{g}_h$ obtained by removing the pair of events (u,v). We proposed the following two estimators for $Q_1(h)$ and $Q_2(h)$:

$$\hat{Q}_1(h) = \frac{1}{m(m-1)} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \tilde{w}(u,v) \left[\hat{g}_h^{-(u,v)}(u,v)\right]^2,$$

and

$$\hat{Q}_2(h) = \frac{1}{m} \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \tilde{w}(u,v)\hat{g}_h^{-(u,v)}(u,v).$$

Chapter 6

Asymptotic Properties

In this chapter, we develop asymptotic properties of the proposed local linear estimator based on (4.1) and the proposed orthogonal series estimator based on (4.5). We assume that there are m_n independent and identically distributed point processes being observed over a sequence of observation windows D_n. Let h_n be a sequence of bandwidths for the local linear estimator and L_n be a sequence of numbers of basis functions for the orthogonal series estimator. More detailed assumptions regarding these parameters will be discussed in the following sections. Generally, we let $\hat{\Theta}$ and $\hat{g}$ denote the estimators of Θ and g in the formulation of the asymptotic properties of the estimators.

In the results below we shall refer to higher order normalized joint intensities $g^{(k)}$ of X. Define the k'th order joint intensity by the identity

$$E \sum_{u_1 \cdots, u_k \in X}^{\neq} 1[u_1 \in A_1, \cdots, u_k \in A_k] = \int_{A_1 \times \cdots \times A_k} \lambda^{(k)}(x_1, \cdots, x_k) dx_1 \cdots dx_k$$

for bounded subsets $A_i \subset \mathbb{R}$, $i = 1, \cdots, k$, where the sum is over distinct $u_1 \cdots, u_k$. We then let $g^{(k)}(x_1, \cdots, x_k) = \lambda^{(k)}(x_1, \cdots, x_k)/(\lambda(x_1) \cdots \lambda(x_k))$ and assume that translation invariant for $k = 3, 4$, i.e. $g^{(k)}(x_1, \cdots, x_k) = g^{(k)}(x_2 - x_1, \cdots, x_k - x_1)$.

6.1 Consistency of the Local Linear Estimator

In this section, I will show the consistency of the original local linear estimator and also the consistency of the log local linear estimator.

6.1.1 Consistency of the Local Linear Estimator

In this subsection, the local linear estimator has a form as (4.1) with the estimating equation as (4.2) shows. For the sake of convenience, we define the expectation of matrix A and vector B as matrix M and vector V. Let $M^{(n)}$ and $V^{(n)}$ be the expected values of $A^{(n)}$ and $B^{(n)}$, where $A^{(n)}$ and $B^{(n)}$ are A and B in (4.3) or (4.7) that are calculated from the m_n point processes. Taking the local linear form as an example, we have

$$M_{l,k}^{(n)} = EA_{l,k}^{(n)} = \int_{(D_n)^2} \frac{f_{l,k}(|x-y|-r)}{\gamma_n(x-y)} \lambda(x)\lambda(y)dxdy, l,k = 1,2$$

and

$$V_l^{(n)} = EB_l^{(n)} = \int_{(D_n)^2} \frac{K_h(|x-y|-r)G_l(|x-y|-r)}{\gamma_n(u-v)} g(|x-y|)\lambda(x)\lambda(y)dxdy, l = 1,2$$

where $\gamma_n(x-y) = |D \cap D - x + y|$ and $f_{l,k}(r) = K_h(r)G_l(r)G_k(r)$. We assume that the following conditions hold in the derivation of asymptotic properties.

V1 For all $x \in \mathbb{R}$, $0 \leq \lambda(x) \leq \lambda_{\max} < \infty$.

V2 The constants $0 < C_I, C_g, C_{g^{(3)}}, C_{g^{(4)}} < \infty$ can be found such that

$$\int_0^\infty \Big|g(r) - 1\Big| dr < C_I$$

and for all $x_1, x_2, x_3, x_4 \in \mathbb{R}$, $g(|x_1 - x_2|) \leq C_g$, $g^{(3)}(x_1, x_2, x_3) \leq C_{g^{(3)}}$ and

$$G^{(4)}(x_1, x_2, x_3) = \int_{\mathbb{R}} \Big|g^{(4)}(x_1 + x_2, x_2, x_3 + y, y) - g(|x_1|)g(|x_3|)\Big| dy \leq C_{g^{(4)}}.$$

Note that Condition V1-V2 are mild conditions which are assumed true in most point pattern analysis.

Let

$$\gamma_n(s) = |D_n \cap (D_n - s)| = \int_{\mathbb{R}} \mathbb{I}[y \in D_n, y + s \in D_n] dy, \quad s \in \mathbb{R}.$$

Then for any fixed $s \in \mathbb{R}$, $\gamma_n(s)/|D_n| \to 1$ and hence a constant $0 < C_\gamma \leq 1$ can be found such that $\gamma_n(s) \geq C_\lambda |D_n|$ for all $-R \leq s \leq R$ and sufficiently large n.

Let $A^{(n)} = [A^{(n)}_{l,k}]$, $M^{(n)} = [M^{(n)}_{l,k}]$, $B^{(n)} = (B^{(n)}_l)$, $V^{(n)} = (V^{(n)}_l)$ and $\Theta_r = (\theta_1(r), \theta_2(r))$. Then $g(r) = (G(r - t))^\mathsf{T} \Theta_r$ and the above estimating equations are equivalent to

$$\sum_{m=1}^{2} A^{(n)}_{l,m} \hat{\theta}_m(r) = B^{(n)}_l, \quad l = 1, 2.$$

or $A^{(n)} \hat{\Theta}_r = B^{(n)}$, where $\hat{\Theta}_t = (\hat{\theta}_1(r), \hat{\theta}_2(r))$.

We consider the Euclidean norm $\|B^{(n)}\| = \left(\sum_{l=1}^{2} \left(B^{(n)}_l\right)^2\right)^{1/2}$ for vectors and the maximum absolute column sum norm (Isaacson and Keller, 1994, p. 9).

$$\|A^{(n)}\|_1 = \max_{1\le k\le 2} \sum_{l=1}^{2} |A^{(n)}_{l,k}|$$

and the Frobenius norm

$$\left\|A^{(n)}\right\|_{\mathrm{F}} = \left(\sum_{l=1}^{2} \sum_{k=1}^{2} (A^{(n)}_{l,k})^2 \right)^{1/2}$$

for matrices. It is known that $\|A^{(n)}\|_1 \le \sqrt{2}\|A^{(n)}\|_{\mathrm{F}}$ (See Golub and Van Loan, 1996, p.56).

6.1.1 Lemma. *Under conditions $V1$-$V2$, $\left\|A^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ and $\|B^{(n)} - V^{(n)}\| = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ and $\|V^{(n)} - M^{(n)}\Theta_r\| = O(h_n^2)$.*

Proof. For each $n \in \mathbb{N}$ and $l, k = 1, 2$,

$$\begin{aligned}
VarB_l^{(n)} &= \frac{1}{m_n} Var \sum_{i=1}^{m_n} \sum_{u,v\in X_i}^{\neq} \frac{K_h(|u-v|-r)G_l(|u-v|-r)}{\gamma_n(u-v)} \\
&= \frac{2}{m_n} \int_{(D_n)^2} \frac{K^2(|x-y|-r)}{h_n^2\gamma_n^2(x-y)} G_l^2(|x-y|-r) g(|x-y|)\lambda(x)\lambda(y)dxdy \\
&+ \frac{4}{m_n} \int_{(D_n)^3} \frac{K(|x-y|-r)K(|x-z|-r)}{h_n^2\gamma_n(x-y)\gamma_n(x-z)} G_l(|x-y|-r)G_l(|x-z|-r) \\
&\quad g^{(3)}(x,y,z)\lambda(x)\lambda(y)\lambda(z)dxdydz \\
&+ \frac{1}{m_n} \int_{(D_n)^4} \frac{K(|x_1-y_1|-r)K(|x_2-y_2|-r)}{h_n^2\gamma_n(x_1-y_1)\gamma_n(x_2-y_2)} G_l(|x_1-y_1|-r)G_l(|x_2-y_2|-r) \\
&\quad \left[g^{(4)}(x_1,y_1,x_2,y_2) - g(|x_1-y_1|)g(|x_2-y_2|)\right]\lambda(x_1)\lambda(y_1)\lambda(x_2)\lambda(y_2)dx_1dy_1dx_2dy_2
\end{aligned}$$

and

$$VarA_{l,k}^{(n)} = \frac{q_n^{(2)} + (m_n - 2)q_n^{(3)}}{m_n(m_n - 1)},$$

where

$$\begin{aligned} q_n^{(2)} &= 2\int_{(D_n)^2} \frac{f_{l,k}^2(|x-y|-r)}{\gamma_n^2(x-y)}\lambda(x)\lambda(y)dxdy \\ &+ 4\int_{(D_n)^3} \frac{f_{l,k}(|x_1-y|-r)f_{l,k}(|x_2-y|-r)}{\gamma_n(x_1-y)\gamma_n(x_2-y)} g(|x_1-x_2|)\lambda(x_1)\lambda(x_2)\lambda(y)dx_1dx_2dy \\ &+ \int_{(D_n)^4} \frac{f_{l,k}(|x_1-y_1|-r)f_{l,k}(|x_2-y_2|-r)}{\gamma_n(x_1-y_1)\gamma_n(x_2-y_2)} \\ &\qquad \Big[g(|x_1-x_2|)g(|y_1-y_2|) + g(|x_1-y_2|)g(|y_1-x_2|) - 2\Big] \\ &\qquad \lambda(x_1)\lambda(x_2)\lambda(y_1)\lambda(y_2)dx_1dy_1dx_2dy_2, \\ q_n^{(3)} &= 4\int_{(D_n)^3} \frac{f_{l,k}(|x_1-y|-r)f_{l,k}(|x_2-y|-r)}{\gamma_n(x_1-y)\gamma_n(x_2-y)}\lambda(x_1)\lambda(x_2)\lambda(y)dx_1dx_2dy \\ &+ 4\int_{(D_n)^4} \frac{f_{l,k}(|x_1-y_1|-r)f_{l,k}(|x_2-y_2|-r)}{\gamma_n(x_1-y_1)\gamma_n(x_2-y_2)}\big[g(|x_1-x_2|)-1\big] \\ &\qquad \lambda(x_1)\lambda(x_2)\lambda(y_1)\lambda(y_2)dx_1dy_1dx_2dy_2. \end{aligned}$$

Since for sufficiently large n, $\gamma_n(s) \geq C_\gamma|D_n|$, using conditions V1-V2, also $K(r) < 1$

we obtain

$$
\begin{aligned}
VarB_l^{(n)} \leq & \frac{2\lambda_{\max}^2 C_g}{m_n|D_n|C_\gamma} \int_{\mathbb{R}^2} \frac{K(|s|-r)\mathbb{I}[y \in D_n, y+s \in D_n]}{h_n^2 \gamma_n(s)} G_l^2(|s|-r) dy ds \\
& + \frac{4\lambda_{\max}^3 C_{g^{(3)}}}{m_n|D_n|C_\gamma} \int_{\mathbb{R}^3} \frac{K(|s_1|-r)K(|s_2|-r)\mathbb{I}[x \in D_n, x-s_1 \in D_n]}{h_n^2\gamma_n(s_1)} \\
& \qquad \big|G_l(|s_1|-r)\big|\big|G_l(|s_2|-r)\big| dx ds_1 ds_2 \\
& + \frac{\lambda_{\max}^4}{m_n|D_n|C_\gamma} \int_{\mathbb{R}^3} \frac{K(|s_1|-r)K(|s_2|-r)\mathbb{I}[y_1+s_1 \in D_n]}{h_n^2\gamma_n(s_1)} \\
& \qquad G^{(4)}(s_1, y_1, s_2)\big|G_l(|s_1|-r)\big|\big|G_l(|s_2|-r)\big| dy_1 ds_1 ds_2 \\
\leq & \frac{4\lambda_{\max}^2 C_g}{m_n|D_n|h_n^2 C_\gamma} \int_{r-h_n}^{r+h_n} G_l^2(|s|-r) ds \\
& + \frac{8\lambda_{\max}^3 C_{g^{(3)}} + 2\lambda_{\max}^4 C_{g^{(4)}}}{m_n|D_n|h_n^2 C_\gamma} \left(\int_{r-h_n}^{r+h_n} \big|G_l(|s|-r)\big| ds \right)^2
\end{aligned}
$$

and

$$
\begin{aligned}
q_n^{(2)} \leq & \frac{4\lambda_{\max}^2}{|D_n|h_n^2 C_\gamma} \int_{r-h_n}^{r+h_n} G_l^2(|s|-r) G_k^2(|s|-r) ds \\
& + \frac{8\lambda_{\max}^3}{|D_n|h_n^2 C_\gamma} \left(\int_{r-h_n}^{r+h_n} \big|G_l(|s|-r)\big|\big|G_k(|s|-r)\big| ds \right)^2 \\
& + \frac{2\lambda_{\max}^4}{|D_n|h_n^2 C_\gamma} \left(2(1+C_g) \int_{\mathbb{R}} \big|g(|s|)-1\big| ds \right) \left(\int_{r-h_n}^{r+h_n} \big|G_l(|s|-r)\big|\big|G_k(|s|-r)\big| ds \right)^2 \\
q_n^{(3)} \leq & \frac{8\lambda_{\max}^3}{|D_n|h_n^2 C_\gamma} \left(\int_{r-h_n}^{r+h_n} \big|G_l(|s|-r)\big|\big|G_k(|s|-r)\big| ds \right)^2 \\
& + \frac{8\lambda_{\max}^4}{|D_n|h_n^2 C_\gamma} \left(\int_{\mathbb{R}} \big|g(|s|)-1\big| ds \right) \left(\int_{r-h_n}^{r+h_n} \big|G_l(|s|-r)\big|\big|G_k(|s|-r)\big| ds \right)^2 .
\end{aligned}
$$

But $G_1(|s|-r)=1$ and $G_2(|s|-r)=|s|-r$ and

$$\begin{aligned}
&\int_{r-h_n}^{r+h_n}|G_1(|s|-r)|ds=\int_{r-h_n}^{r+h_n}\big|G_1(|s|-r)\big|^2ds=\int_{r-h_n}^{r+h_n}\big|G_1(|s|-r)\big|^4ds=2h_n,\\
&\int_{r-h_n}^{r+h_n}|G_2(|s|-r)|ds=\int_{r-h_n}^{r+h_n}\big||s|-r\big|ds\le\int_{r-h_n}^{r+h_n}|s-r|ds=h_n^2,\\
&\int_{r-h_n}^{r+h_n}\big|G_2(|s|-r)\big|^2ds=\int_{r-h_n}^{r+h_n}\big(|s|-r\big)^2ds\le\int_{r-h_n}^{r+h_n}(s-r)^2ds=\frac{2}{3}h_n^3,\\
&\int_{r-h_n}^{r+h_n}\big|G_2(|s|-r)\big|^4ds=\int_{r-h_n}^{r+h_n}\big(|s|-r\big)^4ds\le\int_{r-h_n}^{r+h_n}(s-r)^4ds=\frac{2}{5}h_n^5,\\
&\int_{r-h_n}^{r+h_n}\big|G_1(|s|-r)\big|\big|G_2(|s|-r)\big|ds=\int_{r-h_n}^{r+h_n}\big|G_2(|s|-r)\big|ds\le h_n^2\\
&\int_{r-h_n}^{r+h_n}\big|G_1(|s|-r)\big|^2\big|G_2(|s|-r)\big|^2ds=\int_{r-h_n}^{r+h_n}\big|G_2(|s|-r)\big|^2ds\le\frac{2}{3}h_n^3.
\end{aligned}$$

Thus, for sufficiently large n, $\max\{2h_n,h_n^2,2h_n^3/3,2h_n^5/5\}=2h_n$,

$$VarB_l^{(n)}\le\frac{8\lambda_{\max}^2C_gh_n+\big[8\lambda_{\max}^3C_{g^{(3)}}+2\lambda_{\max}^4C_{g^{(4)}}\big]4h_n^2}{m_n|D_n|h_n^2C_\gamma}$$

and

$$\begin{aligned}
VarA_{l,k}^{(n)}\le&\frac{8\lambda_{\max}^2h_n}{m_n(m_n-1)|D_n|h_n^2C_\gamma}+\frac{32\lambda_{\max}^3h_n^2+32\lambda_{\max}^4(1+C_g)C_Ih_n^2}{m_n(m_n-1)|D_n|h_n^2C_\gamma}\\
&+\frac{(m_n-2)\big(8\lambda_{\max}^3+16\lambda_{\max}^4C_I\big)4h_n^2}{m_n(m_n-1)|D_n|h_n^2C_\gamma}.
\end{aligned}$$

Therefore, $VarB_l^{(n)}\le C_{V_B}(m_n|D_n|h_n)^{-1}$ and $VarA_{l,k}^{(n)}\le C_{V_A}(m_n|D_n|h_n)^{-1}$ for sufficiently large n and some $0 < C_{V_B}, C_{V_A} < \infty$. Consequently,

$$\begin{aligned} E\big\|A^{(n)} - M^{(n)}\big\|_1^2 &\leq 2E\big\|A^{(n)} - M^{(n)}\big\|_{\mathrm{F}}^2 \\ &= 2\sum_{l=1}^{2}\sum_{k=1}^{2} E\big(A_{l,k}^{(n)} - M_{l,k}^{(n)}\big)^2 \\ &= 2\sum_{l=1}^{2}\sum_{k=1}^{2} Var A_{l,k}^{(n)} \leq \frac{8C_{V_A}}{m_n|D_n|h_n} \end{aligned}$$

and

$$E\|B^{(n)} - V^{(n)}\|^2 = \sum_{l=1}^{2} E\left(B_l^{(n)} - V_l^{(n)}\right)^2 = \sum_{l=1}^{2} Var B_l^{(n)} \leq \frac{2C_{V_B}}{m_n|D_n|h_n}.$$

By Markov's inequality, for any $n \in \mathbb{N}$ and $c > 0$

$$\begin{aligned} \mathbb{P}\big\{(m_n|D_n|h_n)^{1/2}\big\|A^{(n)} - M^{(n)}\big\|_1 > c\big\} &\leq \frac{E\big\|A^{(n)} - M^{(n)}\big\|_1^2}{c^2(m_n|D_n|h_n)^{-1}}, \\ \mathbb{P}\big\{(m_n|D_n|h_n)^{1/2}\|B^{(n)} - V^{(n)}\| > c\big\} &\leq \frac{E\|B^{(n)} - V^{(n)}\|^2}{c^2(m_n|D_n|h_n)^{-1}}. \end{aligned}$$

The right hand sides of the above identities can be made arbitrary small, for all n, by choosing sufficiently large c, which means that $\big\|A^{(n)} - M^{(n)}\big\|_1 = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ and $\|B^{(n)} - V^{(n)}\| = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ (See Van der Vaart, 2000, p. 20). Finally, by (4.1), for all n and $t - h_n \leq r \leq t + h_n$,

$$g(t) = \theta_1(r) + \theta_2(t-r) + O(h_n^2) = G_1(t-r)\theta_1(r) + G_2(t-r)\theta_2(r) + O(h_n^2)$$

and

$$
\begin{aligned}
V_l^{(n)} &= \int_{(D_n)^2} \frac{K(|x-y|-r)}{h_n\gamma_n(x-y)} G_l(|x-y|-r) g(|x-y|)\lambda(x)\lambda(y)dxdy \\
&= \int_{(D_n)^2} \frac{K(|x-y|-r)}{h_n\gamma_n(x-y)} G_l(|x-y|-r) \\
&\qquad \Big[G_1(|x-y|-r)\theta_1(r) + G_2(|x-y|-r)\theta_2(r) + O(h_n^2)\Big]\lambda(x)\lambda(y)dxdy \\
&= M_{l,1}^{(n)}\theta_1(r) + M_{l,2}^{(n)}\theta_2(r) + \frac{O(h_n^2)}{h_n}\int_{(D_n)^2} \frac{K(|x-y|-r)}{\gamma_n(x-y)} G_l(|x-y|-r)\lambda(x)\lambda(y)dx
\end{aligned}
$$

and hence

$$
\begin{aligned}
\left| V_l^{(n)} - \sum_{k=1}^{2} \theta_k(r) M_{l,k}^{(n)} \right| &\le \frac{\lambda_{\max}^2 O(h_n^2)}{h} \int_{(D_n)^2} \frac{K(|x-y|-r)}{\gamma_n(x-y)} |G_l(|x-y|-r)|dxdy \\
&\le \frac{\lambda_{\max}^2 O(h_n^2)}{h_n} \int_{r-h_n}^{r+h_n} |G_l(|s|-r)|ds \le 2\lambda_{\max}^2 O(h_n^2)
\end{aligned}
$$

Thus

$$
\|V^{(n)} - M^{(n)}\Theta_r\|^2 = \sum_{l=1}^{2}\left(V_l^{(n)} - \sum_{k=1}^{2}\theta_k(r)M_{l,k}^{(n)} \right)^2 \le 2\big(2\lambda_{\max}^2 O(h^2)\big)^2.
$$

Thus, we have $\|V^{(n)} - M^{(n)}\Theta_r\| = O(h_n^2)$. □

6.1.2 Lemma. *Assume* $h^{-5/2}(m_n|D_n|)^{-\frac{1}{2}} \to 0$. *Then* $\big\|(A^{(n)})^{-1}_{(1\cdot)}\big\| = O_{\mathbb{P}}(\big\|(M^{(n)})^{-1}_{(1\cdot)}\big\|)$, *where* $(A^{(n)})^{-1}_{(1\cdot)}$ *denotes the first row vector of matrix* $(A^{(n)})^{-1}$, *similar definition for* $(M^{(n)})^{-1}_{(1\cdot)}$.

Proof. For matrix $M^{(n)}$, we have

$$
\begin{aligned}
M^{(n)} = EA_{l,k}^{(n)} &= \int_{(D_n)^2} \frac{f_{l,k}(|u-v|-r)}{\gamma_n(u-v)}\lambda(x)\lambda(y)dxdy \\
&= \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix}
\end{aligned}
$$

where $M_{l,k}$ has the following form:

$$
\begin{aligned}
|M_{11}| &= \int_{(D_n)^2} \frac{K(|x-y|-r)G_1(|x-y|-r)|G_1(|x-y|-r)|}{h_n\gamma_n(u-v)}\lambda(x)\lambda(y)dxdy \\
&\le \frac{\lambda_{max}^2}{h_n}\int_{\mathbb{R}^2} \frac{K(|s|-r)\mathbb{I}[y \in D_n, y+s \in D_n]}{\gamma_n(s)}G_1^2(|s|-r)dyds \\
&\le \frac{\lambda_{max}^2}{h_n}\int_{r-h_n}^{r+h_n} G_1^2(|s|-r)ds \\
&\le \lambda_{max}O(1)
\end{aligned}
$$

Similarly we can derive

$$
\begin{aligned}
|M_{21}| = |M_{12}| &= \int_{(D_n)^2} \frac{K(|x-y|-r)|G_1(|x-y|-r)G_2(|x-y|-r)|}{h_n\gamma_n(u-v)}\lambda(x)\lambda(y)dxdy \\
&\le \frac{\lambda_{max}^2}{h_n}\int_{\mathbb{R}^2} \frac{K(|s|-r)\mathbb{I}[y \in D_n, y+s \in D_n]}{\gamma_n(s)}|G_1(|s|-r)G_2(|s|-r)|dyds \\
&\le \frac{\lambda_{max}^2}{h_n}\int_{r-h_n}^{r+h_n} |G_1(|s|-r)G_2(|s|-r)|ds \\
&\le \lambda_{max}O(h_n)
\end{aligned}
$$

$$
\begin{aligned}
|M_{22}| &= \int_{(D_n)^2} \frac{K(|x-y|-r)|G_2^2(|x-y|-r)|}{h_n \gamma_n(u-v)} \lambda(x)\lambda(y) dx dy \\
&\le \frac{\lambda_{max}^2}{h_n} \int_{\mathbb{R}^2} \frac{K(|s|-r)\mathbb{I}[y \in D_n, y+s \in D_n]}{\gamma_n(s)} |G_2^2(|x-y|-r)| dy ds \\
&\le \frac{\lambda_{max}^2}{h_n} \int_{r-h_n}^{r+h_n} |G_2^2(|x-y|-r)| ds \\
&\le \lambda_{max} O(h_n^2)
\end{aligned}
$$

In fact, we can have the inverse of matrix $M^{(n)}$ as

$$
(M^{(n)})^{-1} = \frac{1}{M_{11}M_{22} - M_{12}M_{21}} \begin{bmatrix} M_{22} & -M_{12} \\ -M_{21} & M_{11} \end{bmatrix}
$$

Thus, we can have $\|(M^{(n)})^{-1}\|_1 = O(h_n^{-2})$. And if we only consider the first row vector of this matrix we have $\|(M^{(n)})^{-1}_{(1\cdot)}\|^2 = [(M^{(n)})^{-1}_{(11)}]^2 + [(M^{(n)})^{-1}_{(12)}]^2 = O(h^{-2})$ which means $\|(M^{(n)})^{-1}_{(1\cdot)}\| = O(h_n^{-1})$.

From the identity

$$
A^{(n)} = M^{(n)} + (A^{(n)} - M^{(n)}) = M^{(n)}\big[I_2 + (M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big], \tag{6.1}
$$

where I_2 is the 2×2 identity matrix, it follows that

$$
\big\|(A^{(n)})^{-1}\big\|_1 \le \big\|\big[I_2 + (M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big]^{-1}\big\|_1 \big\|(M^{(n)})^{-1}\big\|_1.
$$

But

$$
\big\|(M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big\|_1 \le \big\|(M^{(n)})^{-1}\big\|_1 \big\|A^{(n)} - M^{(n)}\big\|_1,
$$

which, from Lemma 6.1.1 and the above derivation, implies that $\big\|(M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big\|_1 = O_{\mathbb{P}}(h_n^{-2}(m_n|D_n|h_n)^{-1/2})$. Thus, if $h_n^{-5/2}(m_n|D_n|)^{-1/2} \to 0$, $\big\|(M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big\|_1 \xrightarrow{\mathbb{R}} 0$. In fact, we can write the (6.1) as:

$$A^{(n)}_{(1\cdot)} = M^{(n)}_{(1\cdot)}\big[I_2 + (M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big],$$

Let $E_n = \big\{\big\|(M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big\|_1 < 1\big\}$. Then $\lim_{n\to\infty} \mathbb{R}(\Omega\backslash E_n) = 0$ and for any $\omega \in E_n$, $\big\|(M^{(n)})^{-1}\big(A^{(n)}(\omega) - M_n\big)\big\|_1 < 1$ and hence (Isaacson and Keller 1994, p.16)

$$\big\|\big[I_2 + (M^{(n)})^{-1}\big(A^{(n)}(\omega) - M_n\big)\big)\big]^{-1}\big\|_1 \leq \frac{1}{1 - \big\|(M^{(n)})^{-1}\big(A^{(n)}(\omega) - M_n\big)\big\|_1},$$

which means that $\big\|\big[I_2 + (M^{(n)})^{-1}\big(A^{(n)}(\omega) - M_n\big)\big)\big]^{-1}\big\|_1 \leq C_M$ for some $1 \leq C_M < \infty$. Therefore, for any $\omega \in E_n$, $\big\|(A^{(n)}(\omega))^{-1}_{(1\cdot)}\big\| \leq C_M\big\|(M^{(n)})^{-1}_{(1\cdot)}\big\|$ and

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\big\{\big\|(A^{(n)})^{-1}_{(1\cdot)}\big\| \leq C_M\big\|(M^{(n)})^{-1}_{(1\cdot)}\big\|_1\big\} = \\
&\lim_{n\to\infty} \mathbb{P}\big(E_n \cap \big\{\big\|(A^{(n)})^{-1}_{(1\cdot)}\big\| \leq C_M\big\|(M^{(n)})^{-1}_{(1\cdot)}\big\|\big\}\big) \\
&+ \lim_{n\to\infty} \mathbb{P}\big((\Omega\backslash E_n) \cap \big\{\big\|(A^{(n)})^{-1}_{(1\cdot)}\big\| \leq C_M\big\|(M^{(n)})^{-1}_{(1\cdot)}\big\|\big\}\big) \\
&= \lim_{n\to\infty} \mathbb{P}(E_n) = 1;
\end{aligned}$$

i.e. for any given $\epsilon > 0$, there exists an $n_\epsilon \in \mathbb{N}$ such that

$$\sup_{n\geq n_\epsilon} \mathbb{P}\big\{\big\|(A^{(n)})^{-1}_{(1\cdot)}\big\| \leq C_M\big\|(M^{(n)})^{-1}_{(1\cdot)}\big\|\big\} \geq 1 - \epsilon.$$

which means that $\big\|(A^{(n)})^{-1}_{(1\cdot)}\big\| = O_{\mathbb{P}}(\big\|(M^{(n)})^{-1}_{(1\cdot)}\big\|)$. $\square$

6.1.3 Theorem. *Under assumptions $V1$-$V2$, $h_n^{-\frac{5}{2}}(m_n|D_n|)^{-1/2} \to 0$, then* $|\hat{\Theta}_1 -$

$\Theta_1| \xrightarrow{\mathbb{P}} 0$.

Proof. The estimating equation $A^{(n)}\hat{\Theta}_r = B^{(n)}$ can be written as

$$A^{(n)}(\hat{\Theta}_r - \Theta_r) = (B^{(n)} - M^{(n)}\Theta_r) + (M^{(n)} - A^{(n)})\Theta_r,$$

which implies that

$$\begin{aligned} |\hat{\Theta}_1 - \Theta_1| &= |(A^{(n)})^{-1}_{(1\cdot)}\big[(B^{(n)} - M^{(n)}\Theta_r) + (M^{(n)} - A^{(n)})\Theta_r\big]| \\ &\leq \big(\|(A^{(n)})^{-1})_{(1\cdot)}\|\big)\big(\|B^{(n)} - M^{(n)}\Theta_r\| + \big\|(M^{(n)} - A^{(n)})\Theta_r\big\|\big). \end{aligned}$$

From Lemma 6.1.1 and the triangle inequality

$$\|B^{(n)} - M^{(n)}\Theta_r\| \leq \|B^{(n)} - V^{(n)}\| + \|V^{(n)} - M^{(n)}\Theta_r\|$$

it follows that $\|B^{(n)} - M^{(n)}\Theta_r\| = O_{\mathbb{P}}(h_n^2 + (m_n|D_n|h_n)^{-1/2})$, because $O(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n + b_n)$. Moreover, from Lemma 6.1.1 $\big\|(M^{(n)} - A^{(n)})\Theta_r\big\| = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ because

$$\big\|(M^{(n)} - A^{(n)})\Theta_t\big\| \leq \big\|M^{(n)} - A^{(n)}\big\|_1 \|\Theta_r\|,$$

and

$$\|\Theta_r\|^2 = \sum_{k=1}^{2} \theta_k^2(r) < \infty.$$

Thus, $\|B^{(n)} - M^{(n)}\Theta_r\| + \big\|(M^{(n)} - A^{(n)})\Theta_r\big\| = O_{\mathbb{P}}(h_n^2 + (m_n|D_n|h_n)^{-1/2})$. It then follows from Lemma 6.1.2 and the fact that $O_{\mathbb{P}}(a_n)o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n b_n)$ that $|\hat{\Theta}_1 - \Theta_1| = O_{\mathbb{P}}(h_n + h_n^{-\frac{5}{2}}(m_n|D_n|)^{-1/2})$. Therefore $|\hat{\Theta}_1 - \Theta_1| \xrightarrow{\mathbb{P}} 0$ if $h_n^{-\frac{5}{2}}(m_n|D_n|)^{-1/2} \to 0$. □

6.1.2 Consistency of the Log Local Linear Estimator

In this subsection, I will focus on the log local linear estimator which can be written as a general local polynomial form of the pair correlation function $g(t)$ in a small neighborhood around r as follows

$$\tilde{g}_r(t;\boldsymbol{\theta}) = \exp\left[\theta_0 + \theta_1(|t| - r) + \cdots + \theta_p(|t| - r)^p\right], \tag{6.2}$$

where the parameter $\boldsymbol{\theta} = (\theta_0, \theta_1, \cdots, \theta_p)^T$ can be estimated by the following estimating equation

$$\begin{aligned} \tilde{\boldsymbol{U}}_{r,h}(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{u,v\in X_i}^{\neq}\sum w_{r,h}(|u-v|)\boldsymbol{G}_r(|u-v|) \\ - \frac{1}{m(m-1)}\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j}\sum w_{r,h}(|u-v|)\boldsymbol{G}_r(|u-v|)\tilde{g}_r(|u-v|;\boldsymbol{\theta}) = \mathbf{0}, \end{aligned} \tag{6.3}$$

where the vector-valued function $\boldsymbol{G}_r(t) = [1, t-r, \cdots, (t-r)^p]^T$ and the weight function $w_{r,h}(t)$ is defined as following

$$w_{r,h}(t) = \frac{K_h(t-r)}{\gamma_n(t)}, \quad \gamma_n(t) = \int_{\mathbb{R}} I\left(y \in D_n, y+t \in D_n\right) \mathrm{d}y \text{ for any } t \in \mathbb{R}, \tag{6.4}$$

with $K_h(t) = h^{-1}K(t/h)$ for some kernel function $K(\cdot)$. Denote the solution to estimating equations $\tilde{\boldsymbol{U}}_{r,h}(\boldsymbol{\theta}) = \mathbf{0}$ as $\widehat{\boldsymbol{\theta}} = (\hat{\theta}_0, \hat{\theta}_1, \cdots, \hat{\theta}_p)^T$, the final estimator of $g(r)$ is defined as

$$\hat{g}(r) = \exp(\hat{\theta}_0). \tag{6.5}$$

Define the $(p+1)\times(p+1)$ matrix valued functions for $k=1,2$ as follows

$$\mathbf{Q}_{n,h}^{(k)}(r) = (|D_n|h)^{k-1}\int_{D_n^2}\lambda(u)\lambda(v)w_{r,h}^k(|u-v|)\boldsymbol{A}_h(|u-v|-r)\boldsymbol{A}_h^T(|u-v|-r)\mathrm{d}u\mathrm{d}v, \quad (6.6)$$

where $\boldsymbol{A}_h(t) = [1, t/h, \cdots, t^p/h^p]^T$. Then some standard change of variable tricks yield that

$$\begin{aligned}\mathbf{Q}_{n,h}^{(k)}(r) = \frac{1}{|D_n|}\int_{D_n}\int_{-r/h}^{\infty}\Big[& I(u-hs-r\in D_n)\lambda(u-hs-r) \\ &+ I(u+hs+r\in D_n)\lambda(u+sh+r)\Big] \\ &\times \lambda(u)\left[\frac{|D_n|K(s)}{\gamma_n(hs+r)}\right]^k \boldsymbol{A}_1(s)\boldsymbol{A}_1^T(s)\mathrm{d}s\mathrm{d}u, \qquad k=1,2.\end{aligned}$$

Following conditions are sufficient for consistency

C1 There exists a constant C_λ such that the intensify function $0\le\lambda(u)\le C_\lambda$ for any $u\in D_n$.

C2 There exist positive constants c_g, C_g and C_f such that (a) $c_g\le g(t)\le C_g$; (b) $\max_{1\le j\le p+1}|f^{(j)}(t)|\le C_f$ for any $t\ge 0$ and that (c) $\int_0^\infty |g(s)-1|\mathrm{d}s < C_g$.

C3 There exist constants $C_{g^{(k)}}$'s such that (a) $\left|g^{(k)}(u_1,u_2,\cdots,u_k)\right|\le C_{g^{(k)}}$ for any $u_j\in D_n$, $j=1,\cdots,k$ and $k=3,4,5,6$; (b) $\int_{\mathbb{R}}|g_0^{(3)}(s,t)-g(|s|)|\mathrm{d}t\le C_{g^{(3)}}$; and (c) $\int_{\mathbb{R}}|g_0^{(4)}(s,t+w,w)-g(|s|)g(|t|)|\mathrm{d}w\le C_{g^{(4)}}$.

C4 The kernel function $K(x)$ has a bounded support, say $[-1,1]$, such that $\int_{-1}^1 K(x)dx = 1$.

C5 We only consider estimation of $g(r)$ with $r < R$, where $\gamma_n(R) \geq c_R|D_n|$ for some $c_R > 0$.

C6 Assume that $h \to 0$, $m|D_n|h \to \infty$, and there exists a constant $c_0 > 0$ such that

$$\eta_{\min}\left[\mathbf{Q}_{n,h}^{(k)}(r)\right] > c_0, \text{ for any } r \leq R \text{ and } k = 1, 2,$$

where $\eta_{\min}(\mathbf{Q})$ denotes the smallest eigenvalue of the matrix $\mathbf{Q}$.

Denote $f(t) = log[g(t)]$ and assume the jth derivative $f^{(j)}(t)$ exist for $j = 1, \cdots, p+1$. Suppose there exists a vector $\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*, \cdots, \theta_p^*)^T$ such that

$$\int_{D_n^2} \lambda(u)\lambda(v)w_{r,h}(|u-v|)\left[g(|u-v|) - \tilde{g}_r(|u-v|; \boldsymbol{\theta}^*)\right] \boldsymbol{G}_r(|u-v|)\mathrm{d}u\mathrm{d}v = \mathbf{0}, \quad (6.7)$$

where $\tilde{g}_r(\cdot;\cdot)$ is defined in (6.2).

6.1.4 Lemma. *Under conditions C1-C2 and C4-C6, we have that as* $h \to 0$,

$$h^j\left[\theta_j^* - f^{(j)}(r)/j!\right] = O(h^{p+1}), \quad j = 0, 1, \cdots, p, \quad (6.8)$$

$$|g(t) - \tilde{g}_r(t; \boldsymbol{\theta}^*)| = O(h^{p+1}), \quad \textit{for } t \in [r-h, r+h]. \quad (6.9)$$

Proof: see appendix A.1.

6.1.5 Lemma. *Under conditions C1-C6, we have that as* $m|D_n|h \to \infty$ *and* $h \to 0$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_h = O_p\left(\frac{1}{\sqrt{m|D_n|h}}\right), \tag{6.10}$$

where the norm $\|\boldsymbol{x}\|_h^2 = x_0^2 + (hx_1)^2 + \cdots + (h^p x_p)^2$ *for any* $\boldsymbol{x} = (x_0, x_1, \ldots, x_p)^T \in \mathbb{R}^{p+1}$ *and* $\boldsymbol{\theta}^*$ *is defined in equation* (A.1).

Proof: see appendix A.1.

6.2 Consistency of Orthogonal Series Estimator

In this section, I will show the consistency of the original orthogonal series estimator and also the consistency of the log orthogonal series estimator.

6.2.1 Consistency of Orthogonal Series Estimator

In this subsection, the orthogonal series estimator has a form as (4.5) with the estimating equation as (4.6) shows. We assume that the following conditions hold in the following derivation:

S1 The basis functions are uniformly bounded; i.e. $|\phi_l(r)| \leq C_\phi$ for all $l \in \mathbb{N}$, $0 \leq r \leq R$ and some $0 < C_\phi < \infty$.

S2 For all $u \in \mathbb{R}$, $0 \leq \lambda(u) \leq \lambda_{\max} < \infty$.

S3 The constants $0 < C_I, C_g, C_{g^{(3)}}, C_{g^{(4)}} < \infty$ can be found such that

$$\int_0^\infty \left|g(r) - 1\right| \mathrm{d}r < C_I$$

and for all $u_1, u_2, u_3, u_4 \in \mathbb{R}$, $g(|u_1 - u_2|) \le C_g$, $g^{(3)}(u_1, u_2, u_3) \le C_{g^{(3)}}$ and

$$G^{(4)}(u_1, u_2, u_3) = \int_{\mathbb{R}} \left| g^{(4)}(u_1 + u_2, u_2, u_3 + v, v) - g(|u_1|)g(|u_3|) \right| \mathrm{d}v \le C_{g^{(4)}}.$$

S4 For some $\eta > 0$, $\left\|(M^{(n)})^{-1}\right\|_1 = O(L_n^{\eta})$.

S5 The coefficients of the expansion of $g(r)$ satisfy $\theta_l = O(l^{-(1+\delta)})$ for some $\delta > \eta$.

Condition S1-S3 are also mild conditions that are usually assumed to be true. Detailed discussions regarding condition S4 and S5 can be found in the appendix $A.2$.

6.2.1 Lemma. *Under conditions $S1$-$S3$ and $S5$,* $\left\|A^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1/2})$, $\|B^{(n)} - V^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1/2})$ *and* $\|V^{(n)} - M^{(n)}\Theta_n\| = O(L_n^{-\delta})$.

Proof: see appendix $A.2$.

6.2.2 Lemma. *Assume condition $S4$ holds and $L_n^{3/2+\eta}(m_n|D_n|)^{-1} \to 0$. Then* $\left\|(A^{(n)})^{-1}\right\|_1 = O_{\mathbb{P}}(\left\|(M^{(n)})^{-1}\right\|_1)$.

Proof: see appendix $A.2$.

Based on the previous two lemma, we can get the consistency for parameter $\hat{\Theta}_n$:

6.2.3 Theorem. *Under assumptions $S2$-$S5$, if $L_n^{3+2\eta}(m_n|D_n|)^{-1} \to 0$ then* $\|\hat{\Theta}_n - \Theta_n\| \xrightarrow{\mathbb{P}} 0$.

The proof of Theorem 6.2.3 follows similar logic as shown in the proof of Theorem 6.1.3. Please check the appendix $A.2$ for rigorous proof. Then we show the consistency of the Orthogonal Series estimator.

For each $r \in [0, R]$, let

$$\hat{g}_n(r) = \sum_{l=1}^{L_n} \hat{\theta}_l \phi_l(r)$$

and define the functional norm

$$\|\hat{g}_n - g_n\|_2 = \left(\int_0^R \big(\hat{g}_n(r) - g_n(r)\big)^2 \mathrm{d}r \right)^{1/2}.$$

Then,

$$\int_0^R \big(\hat{g}_n(r) - g_n(r)\big)^2 \mathrm{d}r \le \|\hat{\Theta}_n - \Theta_n\|^2 \sum_{l=1}^{L_n} \int_0^R \phi_l^2(r)\mathrm{d}r = L_n \|\hat{\Theta}_n - \Theta_n\|^2.$$

and we can conclude that $\|\hat{g}_n - g_n\|_2 = O_{\mathbb{P}}(L_n^{1/2+\eta-\delta} + L_n^{2+\eta}(m_n|D_n|)^{-1/2})$ and state the following corollary.

6.2.4 Corollary. *Under conditions S2-S5, if* $\delta > \eta + 1/2$ *and* $L_n^{2+\eta}(m_n|D_n|)^{-1/2} \to 0$, *then* $\|\hat{g}_n - g_n\|_2 \xrightarrow{\mathbb{P}} 0$.

Corollary 6.2.4 implies that $\|\hat{g}_n - g\|_2 \xrightarrow{\mathbb{P}} 0$ under the same conditions, in the sense that $\hat{g}_n(r) - g(r) = \hat{g}_n(r) - g_n(r) - \zeta_n(r)$, $\|\zeta_n\|_2 \to 0$ where $\zeta_n(r) = g(r) - g_n(r) = \sum_{l=L_n+1}^{\infty} \theta_l \phi_l(r)$. Then for each $r \in (0, R)$, $|\zeta_n(r)| \to 0$. In addition (see Tolstov, 1962, p. 55),

$$\int_0^R \zeta_n^2(r)\mathrm{d}r = \sum_{l=L_n+1}^{\infty} \theta_l^2 \to 0.$$

and

$$\|\hat{g}_n - g\|_2 \leq \|\hat{g}_n - g_n\|_2 + \|\zeta_n\|_2.$$

6.2.2 Consistency of Log Orthogonal Series Estimator

Given a complete orthonormal basis of functions $\phi_l(r)$ on $[0, R]$, the orthogonal series expansion of the square-integrable log-pair correlation function $\log[g(r)]$ on $[0, R]$ is given by

$$\log[g(r)] = \sum_{l=1}^{\infty} \theta_{0,l}\phi_l(r), \quad \text{where} \quad \theta_{0,l} = \int_0^R \log[g(r)]\,\phi_l(r)w_o(r)\mathrm{d}r,$$

where basis functions $\phi_l(r)$'s are orthogonal with respect to some weight function $w_o(r) \geq 0$ such that $\int_0^R \phi_l(r)\phi_k(r)w_o(r)\mathrm{d}r = I(k = l)$.

By Parseval's identity (see Tolstov, 1962, p. 119),

$$\sum_{l=1}^{\infty} \theta_{0,l}^2 = \int_0^R w_o(r)\left\{\log[g(r)]\right\}^2 \mathrm{d}r,$$

and hence $\sum_{k=l}^{\infty} \theta_k^2 \to 0$ and $\theta_l \to 0$, as $l \to \infty$, if the right-hand side of the above equation is integrable. Consider the log orthogonal series estimator of the pair correlation function $g(t)$

$$\tilde{g}_L(r; \boldsymbol{\theta}) = \exp\left[\sum_{l=1}^{L} \theta_l\phi_l(r)\right], \qquad (6.11)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_L)^T$ and L is a predefined positive integer. Let

$$\tilde{\zeta}_L(r; \boldsymbol{\theta}_0) = \log[g(r)] - \log[\tilde{g}_L(r; \boldsymbol{\theta}_0)] = \sum_{l=L+1}^{\infty} \theta_{0,l}\phi_l(r). \tag{6.12}$$

Then for each $r \in (0, R)$, under appropriate conditions, we have $|\tilde{\zeta}_L(r; \boldsymbol{\theta}_0)| \to 0$ as $L \to \infty$. In addition (see Tolstov, 1962, p. 55),

$$\int_0^R w_o(r)\tilde{\zeta}_L^2(r; \boldsymbol{\theta}_0)\mathrm{d}r = \sum_{l=L+1}^{\infty} \theta_{0,l}^2 \to 0, \text{ as } L \to \infty.$$

The parameter vector $\boldsymbol{\theta}_0$ can then be obtained by solving the estimating equations (4.8) which can be written as

$$\begin{aligned} \tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}) = & \frac{1}{m}\sum_{i=1}^{m}\sum\sum_{u,v\in X_i}^{\neq} w(|u-v|)\boldsymbol{\phi}_L(|u-v|) \\ & - \frac{1}{m(m-1)}\sum\sum_{i\neq j}\sum_{u\in X_i, v\in X_j}\sum w(|u-v|)\boldsymbol{\phi}_L(|u-v|)\tilde{g}_L(|u-v|; \boldsymbol{\theta}) = \mathbf{0}, \end{aligned} \tag{6.13}$$

where the vector-valued function $\boldsymbol{\phi}_L(r) = [\phi_1(r), \phi_2(r), \cdots, \phi_L(r)]^T$ and the weight function is defined as following

$$w_R(r) = \frac{w_0(r)I(r < R)}{\gamma_n(r)}, \quad \gamma_n(r) = \int_{\mathbb{R}} I(y \in D_n, y + r \in D_n)\,\mathrm{d}y \text{ for any } r \in \mathbb{R} \tag{6.14}$$

After solving $\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}) = \mathbf{0}$, denote the solution as $\widehat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_L)^T$, the final estimator of $g(t)$ is defined as

$$\hat{g}_L(r) = \exp\left[\widehat{\boldsymbol{\theta}}^T \boldsymbol{\phi}(r)\right], \quad \text{for } 0 < r \leq R. \tag{6.15}$$

Suppose there exists a vector $\boldsymbol{\theta}^* = (\theta_1^*, \theta_2^*, \cdots, \theta_L^*)^T$ such that

$$\int_{D_n^2} \lambda(u)\lambda(v) w_R(|u-v|) \left[g(|u-v|) - \tilde{g}_L(|u-v|; \boldsymbol{\theta}^*)\right] \boldsymbol{\phi}_L(|u-v|) \mathrm{d}u\mathrm{d}v = \mathbf{0}. \quad (6.16)$$

Define the $L \times L$ matrix valued function as follows

$$\mathbf{Q}_L = \int_{D_n^2} \lambda(u)\lambda(v) w_R(|u-v|) g(|u-v|) \boldsymbol{\phi}_L(|u-v|) \boldsymbol{\phi}_L^T(|u-v|) \mathrm{d}u\mathrm{d}v. \quad (6.17)$$

Then some standard change of variable tricks yields that

$$\mathbf{Q}_L = \frac{1}{|D_n|} \int_{D_n} \int_0^R \left[I(u-s \in D_n)\lambda(u-s) + I(u+s \in D_n)\lambda(u+s)\right] \times \lambda(u) \left[\frac{|D_n|}{\gamma_n(s)}\right] \tilde{g}_L(s; \boldsymbol{\theta}^*) \boldsymbol{\phi}_L(s) \boldsymbol{\phi}_L^T(s) \mathrm{d}s\mathrm{d}u.$$

Following conditions are sufficient for consistency

E1 There exists a constant C_λ such that the intensify function $0 \le \lambda(u) \le C_\lambda$ for any $u \in D_n$.

E2 There exist positive constants c_g and C_g such that (a) $c_g \le g(t) \le C_g$; and that (b) $\int_0^\infty |g(s) - 1| \mathrm{d}s < C_g$.

E3 There exist constants $C_{g^{(k)}}$'s such that (a) $\left|g^{(k)}(u_1, u_2, \cdots, u_k)\right| \le C_{g^{(k)}}$ for any $u_j \in D_n$, $j = 1, \cdots, k$ and $k = 3, 4$; (b) $\int_{\mathbb{R}} |g_0^{(3)}(s,t) - g(|s|)| \mathrm{d}t \le C_{g^{(3)}}$; and (c) $\int_{\mathbb{R}} |g_0^{(4)}(s, t+w, w) - g(|s|)g(|t|)| \mathrm{d}w \le C_{g^{(4)}}$.

E4 There exists ν_1 such that the approximation error (6.12) satisfies (a) $\int_0^R w_o(r) \tilde{\zeta}_L^2(r; \boldsymbol{\theta}_0) \mathrm{d}r = \sum_{l=L+1}^\infty \theta_{0,l}^2 = O\left(L^{-2\nu_1}\right)$; (b) $\sup_{0<r\le R} |\tilde{\zeta}_L(r; \boldsymbol{\theta}_0)| = O\left(L^{-\nu_1+\tau_1}\right)$ for some $0 < \tau_1 < \nu_1$; and that (c) $\sup_{0<r\le R} \|\boldsymbol{\phi}_L(r)\| = O(L^{\nu_2})$ for some $0 \le \nu_2 < \nu_1$, where $\boldsymbol{\phi}_L(r) = [\phi_1(r), \phi_2(r), \cdots, \phi_L(r)]^T$; and (d) the weight function is uniformly bounded, i.e., $w_o(r) \leq C_w$ for any $0 < r \leq R$.

E5 We only consider estimation of $g(r)$ with $r < R$, where $\gamma_n(R) \geq c_R|D_n|$ for some $c_R > 0$.

E6 Assume that $L \to \infty$, there exists constant $c_0 > 0$ and $0 \leq 2\nu_0 < \nu_1 - \nu_2$ such that

$$\eta_{\min}\left[\mathbf{Q}_L^{(k)}\right] > c_0 L^{-\nu_0}, k = 1, 2,$$

where $\eta_{\min}(\mathbf{Q})$ denotes the smallest eigenvalue of the matrix $\mathbf{Q}$.

The following Lemma quantifies the distance between $g(r)$ and $\tilde{g}_L(r; \boldsymbol{\theta}^*)$.

6.2.5 Lemma. *Under conditions E1-E2 and E4-E6, we have that as $L \to \infty$,*

$$\begin{aligned}
\|\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*\| &= O(L^{\nu_0 - \nu_1}), && (6.18)\\
\sup_{0<r<R} |g(r) - \tilde{g}_L(r; \boldsymbol{\theta}^*)| &= O\left(L^{-\nu_1 + \max\{\tau_1, \nu_0 + \nu_2\}}\right) = o(1), && (6.19)\\
\sup_{0<r<R} |\tilde{g}_L(r; \boldsymbol{\theta}^*)| &= O(1), && (6.20)
\end{aligned}$$

where ν_0, ν_1, τ_1 and ν_2 are defined in conditions E4 and E6.

Proof: see appendix $A.3$.

6.2.6 Lemma. *Under conditions E1-E6, we have that as $L \to \infty$ and $L^{4\nu_0 + 2\nu_2}/m|D_n| \to 0$,*

$$\begin{aligned}
\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| &= O_p\left(\frac{L^{\nu_0}}{\sqrt{m|D_n|}}\right) && (6.21)\\
\sup_{0<r<R} \left|g(r) - \tilde{g}_L(r; \widehat{\boldsymbol{\theta}})\right| &= O\left(L^{-\nu_1 + \max\{\tau_1, \nu_0 + \nu_2\}}\right) + O_p\left(\frac{L^{\nu_0 + \nu_2}}{\sqrt{m|D_n|}}\right), && (6.22)
\end{aligned}$$

where $\boldsymbol{\theta}^*$ *is defined in equation* (6.16).

Proof: see appendix $A.3$.

6.3 Asymptotic Normality

In this section, I will establish asymptotic normality of estimators of the form (6.5) and (6.15) which corresponding to the two log-version estimating equation estimators. Detailed proof follows as below.

6.3.1 Asymptotic Normality of the Log Local Linear Estimator

To verify asymptotic normality of estimator (6.5), we have following conditions beyond of conditions C1-C6

N1 Either one of the following conditions are true (a) $m \to \infty$; or (b) the mixing coefficient satisfies $\alpha_X(s; (t+2R), \infty) = O(s^{-1-\varepsilon})$ for some $t, \varepsilon > 0$.

N2 There exist $\delta > 0$ and $L > 0$ such that $\left|g^{(k)}(u_1, u_2, \cdots, u_k)\right| \leq L$ for any $u_j \in D_n$, $j = 1, \cdots, k$ and $k = 2, \cdots, 2(2 + \lceil\delta\rceil)$, where $\lceil\delta\rceil$ is the smallest integer that is greater than δ.

Define two random vectors

$$\boldsymbol{Z}_1 = \frac{1}{m} \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} w_{r,h}(|u-v|) \boldsymbol{A}_h(|u-v|-r), \tag{6.23}$$

$$\boldsymbol{Z}_2(\boldsymbol{\theta}^*) = \frac{1}{m(m-1)} \sum\sum_{i \neq j} \sum\sum_{u \in X_i, v \in X_j} w_{r,h}(|u-v|)\tilde{g}_r(|u-v|;\boldsymbol{\theta}^*)\boldsymbol{A}_h(|u-v|-r) \quad (6.24)$$

By definition of $\boldsymbol{\theta}^*$ in (A.1), we have that

$$\mathbb{E}\boldsymbol{Z}_1 = \mathbb{E}\boldsymbol{Z}_2 = \int_{D_n^2} \lambda(u)\lambda(v)w_{r,h}(|u-v|)g(|u-v|)\boldsymbol{A}_h(|u-v|-r)\mathrm{d}u\mathrm{d}v. \quad (6.25)$$

6.3.1 Lemma. *Under conditions $C1$-$C6$, we have that, as $h \to 0$,*

$$\begin{aligned} (m|D_n|h)\mathrm{Var}\left(\boldsymbol{Z}_1\right) &= 2g(r)\mathbf{Q}_{n,h}^{(2)}(r) + O(h), & (6.26)\\ (m|D_n|h)\mathrm{Var}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] &= \frac{2}{m-1}g^2(r)\mathbf{Q}_{n,h}^{(2)}(r) + O(h), & (6.27)\\ (m|D_n|h)\mathrm{Cov}\left[\boldsymbol{Z}_1, \boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] &= O(h), & (6.28) \end{aligned}$$

where $\mathbf{Q}_{n,h}^{(2)}(r)$ is as defined in (6.6) and the convergence is entry-wise.

Note that by conditions C1-C6, it is trivial to see that eigenvalues of the matrix $\mathbf{Q}_{n,h}^{(2)}(r)$ are bounded from below and above at the same time. Proof: see appendix B.1.

6.3.2 Lemma. *Under conditions $C1$-$C6$, we have that, as $h \to 0$ and $m|D_n|h \to \infty$,*

$$\sqrt{m|D|_n h}\boldsymbol{\Sigma}_Z^{-1/2}(\boldsymbol{\theta}^*)\left[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] \to^d N\left(\mathbf{0}, \mathbf{I}\right), \quad (6.29)$$

where $\boldsymbol{\Sigma}_Z(\boldsymbol{\theta}^*) = 2(m-1+g(r))/(m-1)g(r)\mathbf{Q}_{n,h}^{(2)}(r)$ with $\mathbf{Q}_{n,h}^{(2)}(r)$ defined in (6.6)

Proof: see appendix B.1.

6.3.3 Lemma. *Denote* $\widehat{\boldsymbol{\theta}}$ *as the solution to estimating equations* (6.3)*, then under conditions* $C1$*-*$C6$*,* $N1$*-*$N2$*, we have that*

$$\begin{aligned}\boldsymbol{\Delta}_h(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) &\equiv \begin{bmatrix} (\hat{\theta}_0 - \theta_0^*) \\ h(\hat{\theta}_1 - \theta_1^*) \\ \vdots \\ h^p(\hat{\theta}_p - \theta_p^*) \end{bmatrix} \\ &= \left[g(r)\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1} \left[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\boldsymbol{\theta}^*) + o_p\left(\frac{1}{\sqrt{m|D|_n h}}\right)\right],\end{aligned} \tag{6.30}$$

where $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2(\boldsymbol{\theta}^*)$ *are defined in* (B.1) *and* (B.2)*, respectively.*

Proof: see appendix $B.1$.

6.3.4 Theorem. *Under conditions* $C1$*-*$C6$*,* $N1$*-*$N2$*, as* $h \to \infty$ *and* $m|D_n|h \to \infty$*, we have that*

$$\frac{\sqrt{m|D_n|h}\,[\hat{g}(r) - g(r) + b_{n,h}]}{\sigma_{m,n,h}} \to^D N(0,1),$$

where $b_{n,h} = O(h^{p+1})$ *and* $\sigma_{m,n,h}^2 = \frac{2g(r)[m-1+g(r)]}{m-1}\boldsymbol{e}^T \left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1} \mathbf{Q}_{n,h}^{(2)}(r) \left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1} \boldsymbol{e}$*, with* $\boldsymbol{e} = (1, 0, \cdots, 0)_{p+1}^T$ *and* $\mathbf{Q}_{n,h}^{(k)}(r)$*,* $k = 1, 2$*, are defined in* (6.6)*.*

Proof: see appendix $B.1$.

6.3.2 Asymptotic Normality of the Log Orthogonal Series Estimator

To verify asymptotic normality of estimator (6.15), we have following conditions beyond of conditions E1-E6

Q1 Either one of the following conditions are true (a) $m \to \infty$; or (b) the mixing coefficient satisfies $\alpha_X(s; (t+2R), \infty) = O(s^{-1-\varepsilon})$ for some $t, \varepsilon > 0$.

Q2 There exist $\delta > 0$ and $L > 0$ such that $\left| g^{(k)}(u_1, u_2, \cdots, u_k) \right| \leq L$ for any $u_j \in D_n$, $j = 1, \cdots, k$ and $k = 2, \cdots, 2(2 + \lceil \delta \rceil)$, where $\lceil \delta \rceil$ is the smallest integer that is greater than δ.

Q3 For $r \in [0, R]$, denote vector $\boldsymbol{\ell}(r) = (\mathbf{Q}_L)^{-1} \boldsymbol{\phi}_L^T(r)$ and its standardized version $\boldsymbol{\ell}_0(r) = \|\boldsymbol{\ell}(r)\|^{-1} \boldsymbol{\ell}(r)$. Assume that as $m|D_n| \to \infty$ and $L \to \infty$, (a) there exists some constant $c_u > 0$ such that $\boldsymbol{\ell}_0^T(r) \boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) \boldsymbol{\ell}_0(r) \geq c_u$ with $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|} \tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*) \right]$; and (b) the basis functions satisfy $\int_0^R \left[w_o(s) | \boldsymbol{\ell}_0^T(r) \boldsymbol{\phi}_L(s) | \right]^{2+\lceil \delta \rceil} \mathrm{d}s \leq C_\phi$, for some $C_\phi > 0$.

6.3.5 Lemma. *Let* $\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*) = \boldsymbol{\delta}_L^T \boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) \boldsymbol{\delta}_L^T$ *with* $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|} \tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*) \right]$. *If the vector* $\boldsymbol{\delta}_L$ *satisfies (a)* $\|\boldsymbol{\delta}_L\| = 1$; *(b)* $\int_0^R \left[w_o(s) | \boldsymbol{\delta}_L^T \boldsymbol{\phi}_L(s) | \right]^{2+\lceil \delta \rceil} \mathrm{d}s = O(1)$; *and (c)* $\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*) \geq c_u$ *for some constant* $c_u > 0$, *then under conditions E1-E6, Q1-Q2, we have that, as* $L \to \infty$ *and* $m|D_n| \to \infty$,

$$\frac{\sqrt{m|D_n|} \boldsymbol{\delta}_L^T \tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)}{\tilde{\sigma}_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*)} \to^d N(0, 1). \tag{6.31}$$

Proof: see appendix $B.2$.

6.3.6 Lemma. *Denote $\widehat{\boldsymbol{\theta}}$ as the solution to estimating equations (6.13), then under conditions E1-E6, we have that as $L \to \infty$ and $L^{4\nu_0+2\nu_2}/m|D_n| \to 0$, for any $0 < r \leq R$,*

$$\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) = \sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*) + o_p(1)\left\|\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\right\|, \quad (6.32)$$

where $\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*)$ and $\mathbf{Q}_L$ are defined in (6.13) and (6.17), respectively. Furthermore, under additional conditions Q1-Q3, we have that

$$\frac{\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)}{\sigma_L(r;\boldsymbol{\theta}^*)} \to^d N(0,1), \quad (6.33)$$

where $\sigma_L^2(r;\boldsymbol{\theta}^*) = \boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)$ and $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right]$.

Proof: see appendix $B.2$.

6.3.7 Theorem. *Under conditions E1-E6, Q1-Q3, as $L \to \infty$ and $m|D_n| \to \infty$, we have that*

$$\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}}) - g(r) + b_{n,L}\right]}{g(r)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}} \to^D N(0,1),$$

where $b_{n,L} = O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right)$, $\mathbf{Q}_L(r)$ is defined in (6.17) and the specific form of $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right]$ is given in the Appendix.

Proof: see appendix $B.2$.

6.4 Empiricical Variance Estimators

To make inference on $g(r)$, we need information on the distributional properties of $\hat{g}$. One standard approach is to derive its limiting distribution of $\hat{g}$ under an appropriate asymptotic frame work and then use it or find a approximation of it to be the distribution of $\hat{g}$ in a finte-sample setting.

In this section, we have derived the asymptotic variance form for the proposed estimators but it can not be calculated in practice. Thus, developing empirical estimators for variances become essential. Here we propose the empirical variance form of the local constant estimator and orthogonal series estimator in this section. And we will use the local constant estimator case as an example to illustrate. In Lemma 6.3.3 and 6.1.4, we have shown that

$$\mathrm{Var}[\hat{g}(r) - g(r)] = [\mathbf{Q}_{n,h}^{(1)}(r)]^{-1}\mathrm{Var}[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\theta^*)][\mathbf{Q}_{n,h}^{(1)}(r)]^{-1}. \tag{6.34}$$

The matrix $\mathbf{Q}_{n,h}^{(1)}(r)$ can simply be estimated as following

$$\mathbf{Q}_{n,h}^{(1)}(r) = \frac{1}{m(m-1)}\sum\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j} w_{r,h}(|u-v|)\boldsymbol{G}_r(|u-v|)\boldsymbol{G}_r^T(|u-v|) \tag{6.35}$$

The variance estimator $\mathrm{Var}[\hat{g}(r) - g(r)]$ differs for two case scenarios, Partition the interval $[0, D_n]$ into K equally spaced intervals, denoted as $\Delta_k = (d_{k-1}, d_k], k = 1, \cdots, K$.

1. **When m is large.** In this case, we can make use of the U-statistic approxi-

mation to $Z_2(\boldsymbol{\theta}^*)$ as follows

$$\tilde{Z}_2(\boldsymbol{\theta}^*) = \frac{2}{m}\sum_{i=1}^{m}\sum_{u\in X_i} q(u) - \mathbb{E}Z_2(\boldsymbol{\theta}^*), \tag{6.36}$$

where $q(u) = \int_{D_n} \lambda_0(v) w_{r,h}(|u-v|)\tilde{g}_r(|u-v|;\boldsymbol{\theta}^*)\mathbf{A}_h(|u-v|-r)\mathrm{d}v$. Then we can decompose $Z_1 - Z_2(\boldsymbol{\theta}^*)$ as

$$\begin{aligned}
Z_1 - Z_2(\boldsymbol{\theta}^*) &\approx \frac{1}{m}\sum_{i=1}^{m}\left[\sum\sum_{u,v\in X_i} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r) - 2\sum_{u\in X_i} q(u)\right] + \mathbb{E}Z_2(\boldsymbol{\theta}^*)\\
&\approx \frac{1}{m}\sum_{i=1}^{m}\Bigg[\sum\sum_{u,v\in X_i} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r)\\
&- 2\sum_{u\in X_i}\frac{1}{m-1}\sum_{j\neq i}\sum_{v\in X_j} w_{r,h}(|u-v|)\hat{g}_r(|u-v|)\boldsymbol{A}_h(|u-v|-r)\Bigg] + \mathbb{E}Z_2(\boldsymbol{\theta}^*)\\
&\approx \frac{1}{m}\sum_{k=1}^{K}\sum_{i=1}^{m}(\mathbf{Y}_{ik} - \mathbb{E}\mathbf{Y}_{ik}),
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{Y}_{ik} = &\sum_{u\in\Delta_k\cap X_i}\sum_{v\in X_i} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r)\\
&- 2\sum_{u\in\Delta_k\cap X_i}\frac{1}{m-1}\sum_{j\neq i}\sum_{v\in X_j} w_{r,h}(|u-v|)\hat{g}_r(|u-v|)\boldsymbol{A}_h(|u-v|-r).
\end{aligned}$$

Then an unbiased estimator for $\mathrm{Var}\left[Z_1 - Z_2(\boldsymbol{\theta}^*)\right]$ becomes

$$\widehat{\mathrm{Var}}\left[Z_1 - Z_2(\boldsymbol{\theta}^*)\right] = \frac{1}{m}\sum_{k=1}^{K}\frac{1}{m-1}\sum_{i=1}^{m}(\mathbf{Y}_{ik} - \overline{\mathbf{Y}}_k)(\mathbf{Y}_{ik} - \overline{\mathbf{Y}}_k)^T, \text{ where } \overline{\mathbf{Y}}_k = \frac{1}{m}\sum_{i=1}^{m}\mathbf{Y}_{ik}.$$

2. **When m is small and D_n is large.**

$$\begin{aligned} Z_1 - Z_2(\boldsymbol{\theta}^*) &= \frac{1}{m}\sum_{i=1}^{m}\Bigg[\sum\sum_{u,v\in X_i} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r) \\ &\quad - \sum_{u\in X_i}\frac{1}{m-1}\sum_{j\neq i}\sum_{v\in X_j} w_{r,h}(|u-v|)\hat{g}_r(|u-v|)\boldsymbol{A}_h(|u-v|-r)\Bigg] \\ &\approx \sum_{k=1}^{K}\frac{1}{m}\sum_{i=1}^{m}\tilde{\mathbf{Y}}_{ik} \end{aligned}$$

where

$$\begin{aligned} \tilde{\mathbf{Y}}_{ik} = &\sum_{u\in\Delta_k\cap X_i}\sum_{v\in X_i} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r) \\ &- \sum_{u\in\Delta_k\cap X_i}\frac{1}{m-1}\sum_{j\neq i}\sum_{v\in X_j} w_{r,h}(|u-v|)\hat{g}_r(|u-v|)\boldsymbol{A}_h(|u-v|-r). \end{aligned}$$

Then an unbiased estimator for $\mathrm{Var}\,[Z_1 - Z_2(\boldsymbol{\theta}^*)]$ becomes

$$\widehat{\mathrm{Var}}\,[Z_1 - Z_2(\boldsymbol{\theta}^*)] = \sum_{k=1}^{K}\sum_{i=1}^{m}(\bar{\tilde{\mathbf{Y}}}_k - \bar{\tilde{\mathbf{Y}}}_{..})(\bar{\tilde{\mathbf{Y}}}_k - \bar{\tilde{\mathbf{Y}}}_{..})^T$$

, where

$$\bar{\tilde{\mathbf{Y}}}_k = \frac{1}{m}\sum_{i=1}^{m}\tilde{\mathbf{Y}}_{ik} \text{ and } \bar{\tilde{\mathbf{Y}}}_{..} = \frac{1}{K}\bar{\tilde{\mathbf{Y}}}_k.$$

Using above estimators, we can finally obtain the estimator for the variance as

$$\widehat{\mathrm{Var}}[\hat{g}(r) - g(r)] = [\widehat{\mathbf{Q}}^{(1)}_{n,h}(r)]^{-1}\widehat{\mathrm{Var}}[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\theta^*)][\widehat{\mathbf{Q}}^{(1)}_{n,h}(r)]^{-1}$$

Similar variance estimator can be obtained for the orthogonal series estimator as well. Now we can make inference for the pair correlation function estimates that we obtained from estimating equations accurately by incorporating with these empirical variance estimators no matter in simulation or in real data application. In the following chapter, a simulation study will be conducted to validate the asymptotic variance form and also the empirical variance estimators derived in this section.

Chapter 7

Simulation Studies

We conducted a simulation study to evaluate the finite sample performance of the proposed pair correlation function estimators. We generate m replicated realizations of inhomogeneous Thomas processes (Waagepetersen, 2007) on $[0, T]$, where $m = 100, 200$ and $T = 30, 60$. Let ρ, μ and σ denote the parent intensity of the process, the mean number of offspring generated per parent, and the standard deviation of an offspring's position relative to its parent, respectively. In this simulation study, we consider two different scenarios for μ which correspond to the two following settings respectively . Also an increasing sample setting are being conducted to supplement our theoretical findings. In addition, a detailed discussion about the asymptotic and empirical variance form of the pair correlation function estimates are being covered in this section.

7.1 Fixed Intensity Setting

In the first scenario, we set $\mu = 6$, i.e., each parent generates on average six offspring events. Each simulated offspring event is retained randomly with a probability equal to $p(x) = 0.28[\sin(2\pi x) + \sin(4\pi x) + 1.811256]$; this probability function is used such that the average number of events per realization when $T = 30$ is similar to that in our real application. The resulting intensity function is hence $\lambda(x) = \rho\mu p(x)$, and the pair correlation function is

$$g(t) = 1 + \frac{\exp(-\frac{t^2}{4\sigma^2})}{2\sqrt{\pi}\rho\sigma}. \tag{7.1}$$

Note that a smaller ρ and a smaller σ both lead to larger $g(t)$ values at small lags. We set $\rho = 0.8, 1.0, 1.2$ and $\sigma = 0.025, 0.030, 0.035$ to reflect the different strengths of clustering at such lags.

For each simulation, we estimate the pair correlation function using the proposed two local linear estimators, the two orthogonal series estimators using cosine basis function and one orthogonal series estimator using spline basis. For the local linear (orthogonal series based on cosine basis) estimators, we will refer to the estimator with the log transformation as the log local linear (orthogonal series estimator) estimator and the other estimator as the original local linear (orthogonal series) estimator. And we will refer to the orthogonal series estimator based on spline basis as the spline estimator. For comparison, we also include the local constant estimator considered by Guan (2011).

Table 7.1 shows the averages from 100 simulations of the integrated squared errors (ISEs), i.e., $\int_0^R [\hat{g}(t) - g(t)]^2 dt$ for some prespecified R value in the first scenario being considered, where $\hat{g}(t)$ is a generic pair correlation function estimator. Three R values, i.e., $R = 0.06, 0.12, 0.18$, are used to show the performance at the different scales. For a fair comparison, the average ISE values being presented are the smallest such values based on a wide range of tuning parameters for each of the estimators. For all combinations of (ρ, σ, R) values, the two orthogonal series estimators always yield the smallest average ISE values, followed by the spline estimator, the local constant estimator and then the two local linear estimators. There is not much difference between the original and the log local linear (orthogonal series) estimators.

7.2 Random and Semiparametric Intensity Setting

In the second scenario, we generate μ_i, $i = 1, \ldots, m$, from a $\Gamma(k, \theta)$ distribution, where $(k, \theta) = (150, 0.04)$, $(120, 0.05)$ or $(100, 0.06)$. Note that the mean and variance of a Gamma random variable are $k\theta$ and $k\theta^2$, respectively. Thus, the expected value of μ_i is 6, which is the same as in the first scenario. For $\theta = 0.06$, there is a larger variability in the simulated μ_i values than when $\theta = 0.05$. We use the same probability function $p(x)$ for retaining a simulated offspring event. Given μ_i, the resulting first- and second-order intensity functions for the ith process are $\lambda_i(x) = \rho\mu_i p(x)$ and $\lambda_i^{(2)}(x, x+t) = \rho^2\mu_i^2 p(x)p(x+t)g(t)$ and consequently the pair correlation function remains the same as in the first scenario. However, without knowing μ_i, the marginal pair correlation function is

$$g(t) = \left[1 + \frac{\exp(-\frac{t^2}{4\sigma^2})}{2\sqrt{\pi}\rho\sigma}\right] \frac{E(\mu_i^2)}{[E(\mu_i)]^2}. \tag{7.2}$$

Note that all these estimators estimate the marginal pair correlation functions, as given in (7.1) and (7.2) for the two different scenarios.

In the second scenario, the intensity function for the ith process can be rewritten as $\lambda_i(x) = \rho \exp(Z_i\beta)p(x)$ with $Z_i = \log\mu_i$ and $\beta = 1$. If we treat Z_i as an observed covariate and β as an unknown regression coefficient, this becomes a special case of the general intensity function given in (3.3). We may then estimate β by using the procedure in Lawless and Nadeau (1995) and subsequently estimate the pair correlation function using our proposed estimators that are generalized for the setting of (3.3). Detailed introduction about the estimating procedure will be discussed in Chapter 9. The new pair correlation function estimators now estimate (7.1) instead of (7.2). Of course, in a real-life setting, we won't observe Z_i in the given form for a Thomas process. The purpose of this additional analysis is merely to illustrate how the pair correlation estimates will change by incorporating the available covariate(s).

Table 7.2 shows the average ISEs in the second scenario. As described before, we consider two: the first estimat the random intensity functions estimated through the model $\lambda_i(x) = \rho \exp(Z_i\beta)p(x)$ and without estimating them. For both settings, the relative performances of the various estimators are the same as in the first scenario. When the random intensity functions are estimated, the average ISEs are slightly smaller than those obtained without estimating the random intensity functions.

7.3 Increasing Sample Size Setting

To supplement our theoretical result in Chapter 6, we consider increasing the number of subjects within each replicated point patterns (m) and the number of days (T) in the simulation. For ease of computation, we only choose one fixed intensity structure

Table 7.1: MISE in fixed intensity setting (MISE in 10^{-2})

σ	ρ	Time Range	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$
		$\hat{g}_c$	0.9569	1.0863	1.1572
		$\hat{g}_l$	0.9959	1.1228	1.1937
	$\rho = 0.8$	$\hat{g}_{l_e}$	0.9855	1.0980	1.1692
		$\hat{g}_o$	0.8791	0.9871	1.0557
		$\hat{g}_{o_e}$	0.9005	1.0002	1.0649
		$\hat{g}_s$	0.9415	1.0641	1.1452
		$\hat{g}_c$	0.4311	0.4982	0.5584
		$\hat{g}_l$	0.4435	0.5104	0.5678
$\sigma = 0.025$	$\rho = 1.0$	$\hat{g}_{l_e}$	0.4557	0.5164	0.5739
		$\hat{g}_o$	0.3586	0.4168	0.4733
		$\hat{g}_{o_e}$	0.3629	0.4170	0.4701
		$\hat{g}_s$	0.3977	0.4662	0.5299
		$\hat{g}_c$	0.2812	0.3344	0.3774
		$\hat{g}_l$	0.2857	0.3405	0.3816
	$\rho = 1.2$	$\hat{g}_{l_e}$	0.2918	0.3401	0.3811
		$\hat{g}_o$	0.2460	0.2908	0.3309
		$\hat{g}_{o_e}$	0.2510	0.2925	0.3301
		$\hat{g}_s$	0.2711	0.3186	0.3609
		$\hat{g}_c$	0.7209	0.8587	0.9171
		$\hat{g}_l$	0.7696	0.9033	0.9616
	$\rho = 0.8$	$\hat{g}_{l_e}$	0.7685	0.8864	0.9449
		$\hat{g}_o$	0.6706	0.7814	0.8373
		$\hat{g}_{o_e}$	0.6758	0.7804	0.8303
		$\hat{g}_s$	0.7068	0.8231	0.8855
		$\hat{g}_c$	0.3479	0.4172	0.4682
		$\hat{g}_l$	0.3602	0.4302	0.4790
$\sigma = 0.030$	$\rho = 1.0$	$\hat{g}_{l_e}$	0.3743	0.4369	0.4858
		$\hat{g}_o$	0.2923	0.3501	0.3981
		$\hat{g}_{o_e}$	0.2966	0.3508	0.3954
		$\hat{g}_s$	0.3143	0.3758	0.4289
		$\hat{g}_c$	0.2128	0.2648	0.3017
		$\hat{g}_l$	0.2157	0.2695	0.3052
	$\rho = 1.2$	$\hat{g}_{l_e}$	0.2237	0.2680	0.3035
		$\hat{g}_o$	0.1864	0.2266	0.2596
		$\hat{g}_{o_e}$	0.1900	0.2293	0.2609
		$\hat{g}_s$	0.1952	0.2358	0.2722
		$\hat{g}_c$	0.5802	0.7235	0.7768
		$\hat{g}_l$	0.6080	0.7458	0.7989
	$\rho = 0.8$	$\hat{g}_{l_e}$	0.6102	0.7365	0.7896
		$\hat{g}_o$	0.5389	0.6543	0.7062
		$\hat{g}_{o_e}$	0.5593	0.6749	0.7220
		$\hat{g}_s$	0.5875	0.7095	0.7735
		$\hat{g}_c$	0.2854	0.3558	0.4007
		$\hat{g}_l$	0.2963	0.3659	0.4094
$\sigma = 0.035$	$\rho = 1.0$	$\hat{g}_{l_e}$	0.3092	0.3728	0.4164
		$\hat{g}_o$	0.2406	0.3020	0.3439
		$\hat{g}_{o_e}$	0.2523	0.3106	0.3499
		$\hat{g}_s$	0.2716	0.3363	0.3869
		$\hat{g}_c$	0.1696	0.2183	0.2531
		$\hat{g}_l$	0.1735	0.2205	0.2551
	$\rho = 1.2$	$\hat{g}_{l_e}$	0.1772	0.2186	0.2529
		$\hat{g}_o$	0.1477	0.1823	0.2137
		$\hat{g}_{o_e}$	0.1521	0.1878	0.2163
		$\hat{g}_s$	0.1597	0.1992	0.2355

Table 7.2: MISE in random & Semiparametric intensity (MISE in 10^{-2})

		random intensity			semiparametric intensity		
case	grid	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$
	$\tilde{g}_c$	0.4268	0.5108	0.5574	0.4328	0.5139	0.5595
	$\tilde{g}_l$	0.4391	0.5242	0.5687	0.4474	0.5292	0.5727
$\Gamma(150, 0.04)$	$\tilde{g}_{l_e}$	0.4672	0.5420	0.5848	0.4585	0.5347	0.5802
	$\tilde{g}_o$	0.3656	0.4366	0.4798	0.3682	0.4374	0.4791
	$\tilde{g}_{o_e}$	0.3671	0.4361	0.4756	0.3699	0.4361	0.4745
	$\tilde{g}_s$	0.3872	0.4611	0.5079	0.3911	0.4636	0.5094
	$\tilde{g}_c$	0.3261	0.3892	0.4400	0.3162	0.3778	0.4301
	$\tilde{g}_l$	0.3517	0.4130	0.4636	0.3453	0.4065	0.4567
$\Gamma(120, 0.05)$	$\tilde{g}_{l_e}$	0.3740	0.4296	0.4791	0.3524	0.4094	0.4595
	$\tilde{g}_o$	0.2894	0.3377	0.3897	0.2824	0.3304	0.3823
	$\tilde{g}_{o_e}$	0.2908	0.3386	0.3856	0.2835	0.3312	0.3779
	$\tilde{g}_s$	0.3123	0.3635	0.4185	0.3065	0.3576	0.4126
	$\tilde{g}_c$	0.4115	0.4870	0.5380	0.4008	0.4691	0.5185
	$\tilde{g}_l$	0.4124	0.4923	0.5420	0.4114	0.4802	0.5275
$\Gamma(100, 0.06)$	$\tilde{g}_{l_e}$	0.4297	0.4931	0.5425	0.4183	0.4792	0.5263
	$\tilde{g}_o$	0.3656	0.4266	0.4737	0.3601	0.4185	0.4619
	$\tilde{g}_{o_e}$	0.3669	0.4262	0.4700	0.3610	0.4178	0.4581
	$\tilde{g}_s$	0.3850	0.4457	0.4976	0.3817	0.4390	0.4875

and one random intensity structure to implement this study. For the fixed intensity function case, we fix ρ to 1.0 and σ to 0.030. For the random intensity case, a gamma distribution with parameter $k = 120$ and $\theta = 0.05$ is being used to generate μ.

One may observe that the ISEs of all methods decreases dramatically as we increase m and T no matter which intensity structure we use (Table 7.3 and 7.4). In addition, we may also notice that the sine series estimator still outperforms all other estimators as the previous two settings. Also, the other two estimators have the similar performance we as observed in above two settings.

Table 7.3: MISE in increasing sample size setting: Fix Intensity (MISE in 10^{-2})

fixed intensity function				
case	Time Range	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$
m=100 & T=30	$\hat{g}_c$	0.3479	0.4172	0.4682
	$\hat{g}_l$	0.3602	0.4302	0.4790
	$\hat{g}_{l_e}$	0.3743	0.4369	0.4858
	$\hat{g}_o$	0.2923	0.3501	0.3981
	$\hat{g}_{o_e}$	0.2966	0.3508	0.3954
	$\hat{g}_s$	0.3143	0.3758	0.4289
m=200 & T=30	$\hat{g}_c$	0.1507	0.1886	0.2113
	$\hat{g}_l$	0.1545	0.1912	0.2139
	$\hat{g}_{l_e}$	0.1564	0.1889	0.2116
	$\hat{g}_o$	0.1275	0.1557	0.1770
	$\hat{g}_{o_e}$	0.1296	0.1568	0.1775
	$\hat{g}_s$	0.1339	0.1640	0.1888
m=100 & T=60	$\hat{g}_c$	0.1993	0.2538	0.2799
	$\hat{g}_l$	0.2046	0.2573	0.2833
	$\hat{g}_{l_e}$	0.1991	0.2419	0.2669
	$\hat{g}_o$	0.1823	0.2208	0.2449
	$\hat{g}_{o_e}$	0.1828	0.2220	0.2450
	$\hat{g}_s$	0.1899	0.2314	0.2594

Table 7.4: MISE in increasing sample size setting: Random Intensity (MISE in 10^{-2})

		random intensity			semiparametric intensity		
case	grid	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$	$[0, 0.06]$	$[0, 0.12]$	$[0, 0.18]$
m=100 & T=30	$\tilde{g}_c$	0.3261	0.3892	0.4400	0.3162	0.3778	0.4301
	$\tilde{g}_l$	0.3517	0.4130	0.4636	0.3453	0.4065	0.4567
	$\tilde{g}_{l_e}$	0.3740	0.4296	0.4791	0.3524	0.4094	0.4595
	$\tilde{g}_o$	0.2894	0.3377	0.3897	0.2824	0.3304	0.3823
	$\tilde{g}_{o_e}$	0.2908	0.3386	0.3856	0.2835	0.3312	0.3779
	$\tilde{g}_s$	0.3123	0.3635	0.4185	0.3065	0.3576	0.4126
m=200 & T=30	$\tilde{g}_c$	0.1743	0.2116	0.2389	0.1673	0.2039	0.2307
	$\tilde{g}_l$	0.1833	0.2194	0.2467	0.1766	0.2121	0.2389
	$\tilde{g}_{l_e}$	0.1944	0.2253	0.2516	0.1788	0.2112	0.2380
	$\tilde{g}_o$	0.1541	0.1805	0.2063	0.1476	0.1741	0.1995
	$\tilde{g}_{o_e}$	0.1572	0.1832	0.2084	0.1506	0.1765	0.2013
	$\tilde{g}_s$	0.1630	0.1912	0.2211	0.1571	0.1852	0.2147
m=100 & T=60	$\tilde{g}_c$	0.2357	0.2920	0.3201	0.2295	0.2840	0.3118
	$\tilde{g}_l$	0.2411	0.3009	0.3282	0.2350	0.2878	0.3156
	$\tilde{g}_{l_e}$	0.2654	0.3122	0.3387	0.2374	0.2833	0.3111
	$\tilde{g}_o$	0.2048	0.2471	0.2736	0.1987	0.2391	0.2654
	$\tilde{g}_{o_e}$	0.2066	0.2494	0.2742	0.2002	0.2413	0.2659
	$\tilde{g}_s$	0.2206	0.2651	0.2945	0.2145	0.2572	0.2864

7.4 Empirical Variance Estimators

To evaluate the performance of the empirical variance estimators proposed in section 6.4, a simulation study was conducted. And this simulation is used to verify the correctness of the asymptotic variance form of the estimator which derived in last section. In this scenario, we fixed $\mu = 6$, $\rho = 0.6$ and $\sigma = 0.016$ for ease of computation. Also we consider increasing the number of subjects within each replicated point pattern(m) and the number of days(T) in the simulation to supplement our theoretical result of the asymptotic estimators' variance.

For the local constant case, we denote the two empirical variance estimator which derived in section 6.4 as $\widehat{\mathrm{Var}}_1(\hat{g}_c)$ and $\widehat{\mathrm{Var}}_2(\hat{g}_c)$. The first one suits the case that when m is large while the second one suits for the case when T is large. For comparison purpose, I include standard error of $\hat{g}_c$ (denotes as $[\mathrm{Var}(\hat{g}_c)]^{1/2}$), the asymptotic standard error in theory($[\mathrm{Var}_{asy}(\hat{g}_c)]^{1/2}$) and the Monte Carlo mean and standard error(denote S) of the two empirical variance estimators. For the orthogonal series case, I include standard error of $\hat{g}_o$ (denotes as $[\mathrm{Var}(\hat{g}_o)]^{1/2}$) and the Monte Carlo mean and standard error(denote S) of the two empirical variance estimator which denote as $\widehat{\mathrm{Var}}_1(\hat{g}_o)$ and $\widehat{\mathrm{Var}}_2(\hat{g}_o)$ respectively .

One may observe that no matter the variance of the pair correlation function estimate, the variance estimates or the theoretical variance all keep decreasing as we increase the number of subjects(m) within the replicated point pattern or the number of days(T). Also the empirical variance becomes more and more close to the asymptotic variance by doing so. If we look at the two empirical variance estimators we may observe that $\mathbb{E}[\widehat{\mathrm{Var}}_1(\hat{g}_c)]^{1/2}$ is more close to $[\mathrm{Var}(\hat{g}_c)]^{1/2}$ than $\mathbb{E}[\widehat{\mathrm{Var}}_2(\hat{g}_c)]^{1/2}$ when $m = 200$ and $T = 30$. And $\mathbb{E}[\widehat{\mathrm{Var}}_2(\hat{g}_c)]^{1/2}$ is more close to $[\mathrm{Var}(\hat{g}_c)]^{1/2}$ than

Table 7.5: Empirical Variance Estimator Result for Local Constant Case

	r	$[\text{Var}(\hat{g}_c)]^{1/2}$	$[\text{Var}_{asy}(\hat{g}_c)]^{1/2}$	$\mathbb{E}[\hat{\text{Var}}_1(\hat{g}_c)]^{1/2}$	$\mathbb{E}[\widehat{\text{Var}}_2(\hat{g}_c)]^{1/2}$	$S[\widehat{\text{Var}}_1(\hat{g}_c)]^{1/2}$	$S[\widehat{\text{Var}}_1(\hat{g}_c)]^{1/2}$
	0.001	1.277	1.140	1.613	1.206	0.109	0.389
	0.028	0.616	0.668	0.713	0.555	0.047	0.181
m=100	0.055	0.271	0.300	0.249	0.216	0.036	0.077
&	0.082	0.221	0.199	0.193	0.160	0.042	0.060
T=30	0.100	0.210	0.198	0.193	0.172	0.040	0.068
	0.135	0.216	0.206	0.202	0.185	0.044	0.068
	0.180	0.213	0.218	0.206	0.178	0.040	0.062
	0.001	0.958	0.570	1.136	0.933	0.059	0.182
	0.028	0.512	0.334	0.500	0.432	0.024	0.088
m=100	0.055	0.173	0.150	0.176	0.166	0.016	0.037
&	0.082	0.154	0.099	0.140	0.136	0.021	0.037
T=60	0.100	0.137	0.099	0.140	0.134	0.021	0.034
	0.135	0.138	0.103	0.142	0.132	0.021	0.031
	0.180	0.157	0.109	0.141	0.125	0.024	0.029
	0.001	0.935	0.891	0.990	0.773	0.047	0.279
	0.028	0.482	0.493	0.454	0.380	0.021	0.145
m=200	0.055	0.159	0.213	0.172	0.144	0.016	0.049
&	0.082	0.154	0.141	0.141	0.117	0.024	0.041
T=30	0.100	0.159	0.141	0.140	0.117	0.023	0.041
	0.135	0.131	0.146	0.143	0.120	0.020	0.044
	0.180	0.144	0.155	0.140	0.121	0.017	0.041

$\mathbb{E}[\widehat{\text{Var}}_1(\hat{g}_c)]^{1/2}$ when $m = 100$ and $T = 60$ which are consistent with what we expect in section 6.4. Also we may find that the Monte Carlo standard error of the two empirical variance estimators are acceptable comparing with the value of the Monte Carlo mean. For the orthogonal series case, we can find the two empirical variance estimators' cases perform similarly as what we have in the local constant case, which they both have a better performance case when being compared with each other.

Table 7.6: Empirical Variance Estimator Result for Orthogonal Series Case

	r	$[\text{Var}(\hat{g}_o)]^{1/2}$	$\mathbb{E}[\widehat{\text{Var}}_1(\hat{g}_o)]^{1/2}$	$\mathbb{E}[\widehat{\text{Var}}_2(\hat{g}_o)]^{1/2}$	$S[\widehat{\text{Var}}_1(\hat{g}_o)]^{1/2}$	$S[\widehat{\text{Var}}_1(\hat{g}_o)]^{1/2}$
	0.001	1.348	1.572	1.169	0.107	0.374
	0.028	0.586	0.676	0.515	0.048	0.169
m=100	0.055	0.266	0.236	0.205	0.035	0.072
&	0.082	0.171	0.157	0.134	0.031	0.052
T=30	0.100	0.203	0.184	0.163	0.038	0.062
	0.135	0.212	0.190	0.176	0.044	0.071
	0.180	0.248	0.219	0.195	0.052	0.079
	0.001	0.935	1.105	0.915	0.055	0.186
	0.028	0.465	0.473	0.411	0.023	0.086
m=100	0.055	0.168	0.169	0.162	0.018	0.041
&	0.082	0.123	0.114	0.111	0.017	0.028
T=60	0.100	0.135	0.134	0.129	0.020	0.034
	0.135	0.125	0.134	0.125	0.018	0.029
	0.180	0.161	0.149	0.134	0.025	0.037
	0.001	0.925	0.960	0.748	0.044	0.264
	0.028	0.422	0.429	0.360	0.020	0.138
m=200	0.055	0.150	0.164	0.135	0.017	0.045
&	0.082	0.122	0.113	0.093	0.019	0.034
T=30	0.100	0.154	0.135	0.112	0.022	0.044
	0.135	0.123	0.135	0.114	0.019	0.040
	0.180	0.169	0.150	0.134	0.021	0.043

Chapter 8

Empirical Study

As our empirical study, we present here a real social-network example about SINA Weibo (www.weibo.com), which can be viewed as a largest Twitter-type social media in the Chinese community. For illustration purpose, the dataset, specifically, contains $m = 1,695$ active followers of an official MBA program. Their tweeting timestamps and related covariates are recorded for a total of $T = 30$ days. For each user, gender, the number of followers (which indicates the accounts that are following the user) and the number of followees (which indicates the accounts that the user are following) information are being recorded in the data.

8.1 Pair Correlation Function Analysis

The objective of this study is to describe the posting pattern of the SINA Weibo users in terms of their posting activity. For illustration purpose, we start with the Weibo accounts incorporating all the associated covariates. As we mentioned before, the first-order intensity function can be approaching zero for a fairly long period during a day since people barely have tweeting activities during the night in this data structure. Thus, the existing nonparametric smoothing methods may result in poor estimation of the pair correlation function. Taken aforementioned covariates into consideration, three group segmentation settings are being introduced when applying the two estimators with empirical variance estimators to estimate the pair correlation function. We conducted a posting pattern analysis with the two well-tuned estimators for the replicated point patterns within each group segmentation setting.

In the first group segmentation scenario, all users are being classified into two main groups naturally according to their gender information. As a result, 1221 male accounts and 474 female accounts are being split into two replicated point patterns. The second scenario is being built with the covariate followers information which has an extremely right-skewed distribution. I split all users by the third quantile of this distribution into two groups: the group with more followers will be referred as 'celebrity' users (424 accounts) and the other one will be referred as 'nonCelebirty' users (1271 accounts). For the followees information which has a bi-modes distribution, I split the data by the number of 2000 since only paid-up members can have more than 2000 followees (This number is also the second mode position within its distribution). I will refer to the users who have more followees as 'member' users (107 accounts) and the other group as 'nonMemeber' Users (1588 accounts) which forms the last scenario.

In fact, we can still use Kernel smoothing (Diggle, 1985) to estimate the intensity function under the three scenarios. Limited posting points during the night would lead to a poor estimation of the intensity function, but the overall shapes are still worthy referring to. If we look at the figure 8.1 more closely, two replicated point

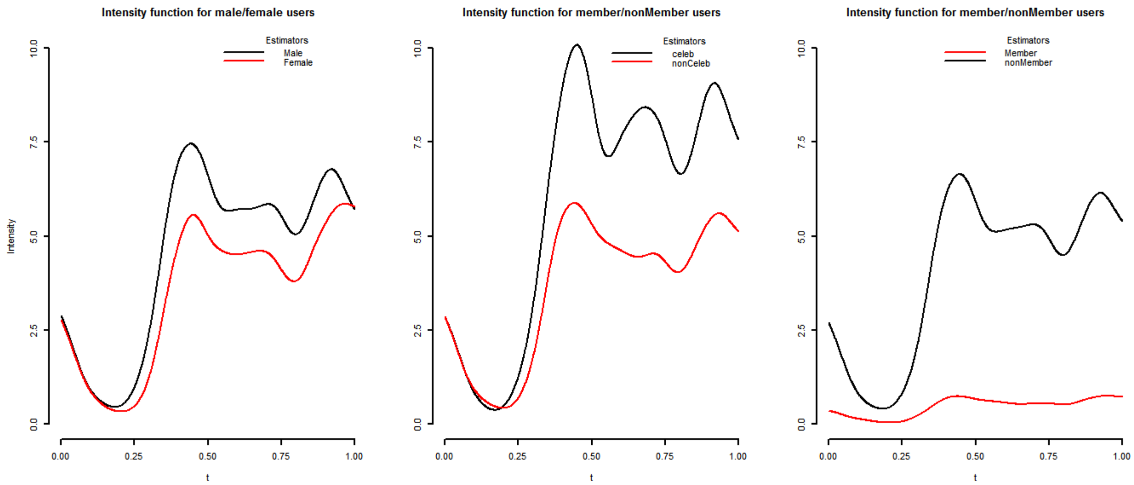

Figure 8.1: Intensity function under three settings

patterns within each scenario all have distinct overall shapes which indicate the varied magnitude of how likely user tends to post across a day. The first plot shows that the male users tend to have a larger intensity than the female users with a similar shape. The second plot shows celebrity users have a larger intensity than the non-Celebrity users during the most of the day. The last plot shows intensity functions for member users and nonMember users. The estimates of the member users' intensity function seems corrupt. And the nonMember users have a larger intensity function than the member users whose intensity function doesn't fluctuate very much. Existing nonparametric methods for the pair correlation function based on this estimation are foreseeably ill-estimated. The local constant and the orthogonal series estimators (Both estimators are being applied with the log-transformed version instead of the original one, we will refer to these two estimators as local constant and orthogonal series estimator in this section) are being used to estimate the pair correlation function within the above three settings in a time lag range $[0, 0.12]$. The associated parameters for the two methods are being well-tuned follow the procedure introduced in

Chapter 5. Besides the pair correlation function estimates, we also include the 95% confidence interval incorporating with the empirical variance estimator discussed in Chapter 6.

Figure 8.2 shows that the two estimates are quite similar no matter for the male users or the female users. Overall speaking, the female users display slightly larger correlation than the male users in short time lag. Thus, male users tend to post more weibo in a day and these weibos have a more clustered pattern than the female users. The overall dependence range for the two groups of users is similar. Also we can observe that no matter the 95% confidence interval of the local constant estimate or the 95% confidence interval of the orthogonal series estimate contains the other estimate most of the time.

Figure 8.3 shows that the two estimates are still quite similar no matter for the celebrity users or the nonCelebrity users. The two 95% confidence interval contain the two estimates nearly across all time lag from 0 to R. For short time lag, the non-Celebrity users have markedly larger correlation comparing with the celebrity users. In fact, the celebrity users have notably larger intensity function than nonCelebrity users. Thus, it turns out that users who have more followers tend to post more frequently with a less clustered tweeting pattern than users who has fewer followers.

In figure 8.4, the two estimates for two replicated point patterns are nearly the same. The clustered posting pattern for nonMember users doesn't have a big difference comparing with the Member users even though the nonMember users have prominently larger intensity. So do their dependence range.

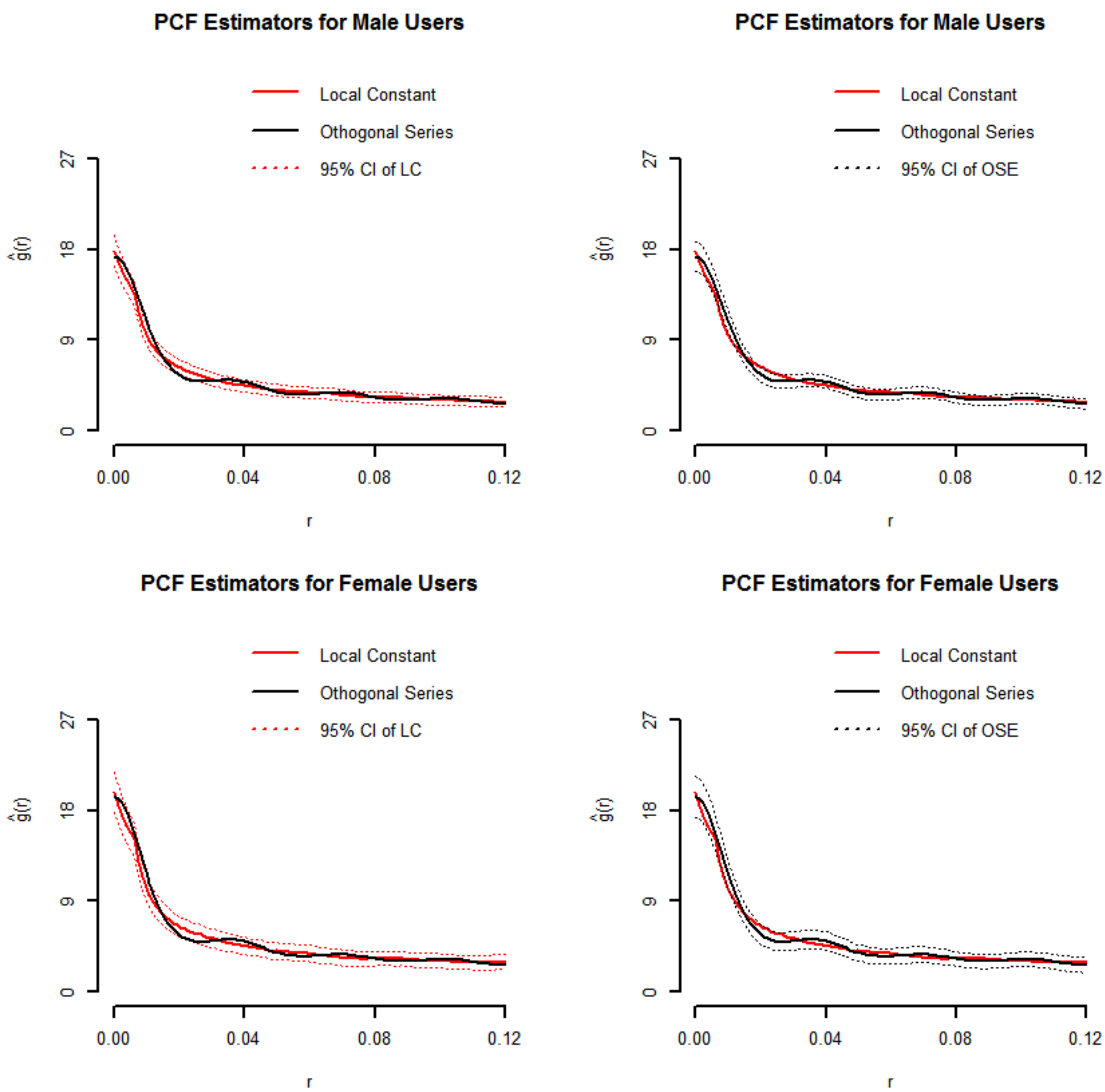

Figure 8.2: Estimated pair correlation functions for Male/Female Users

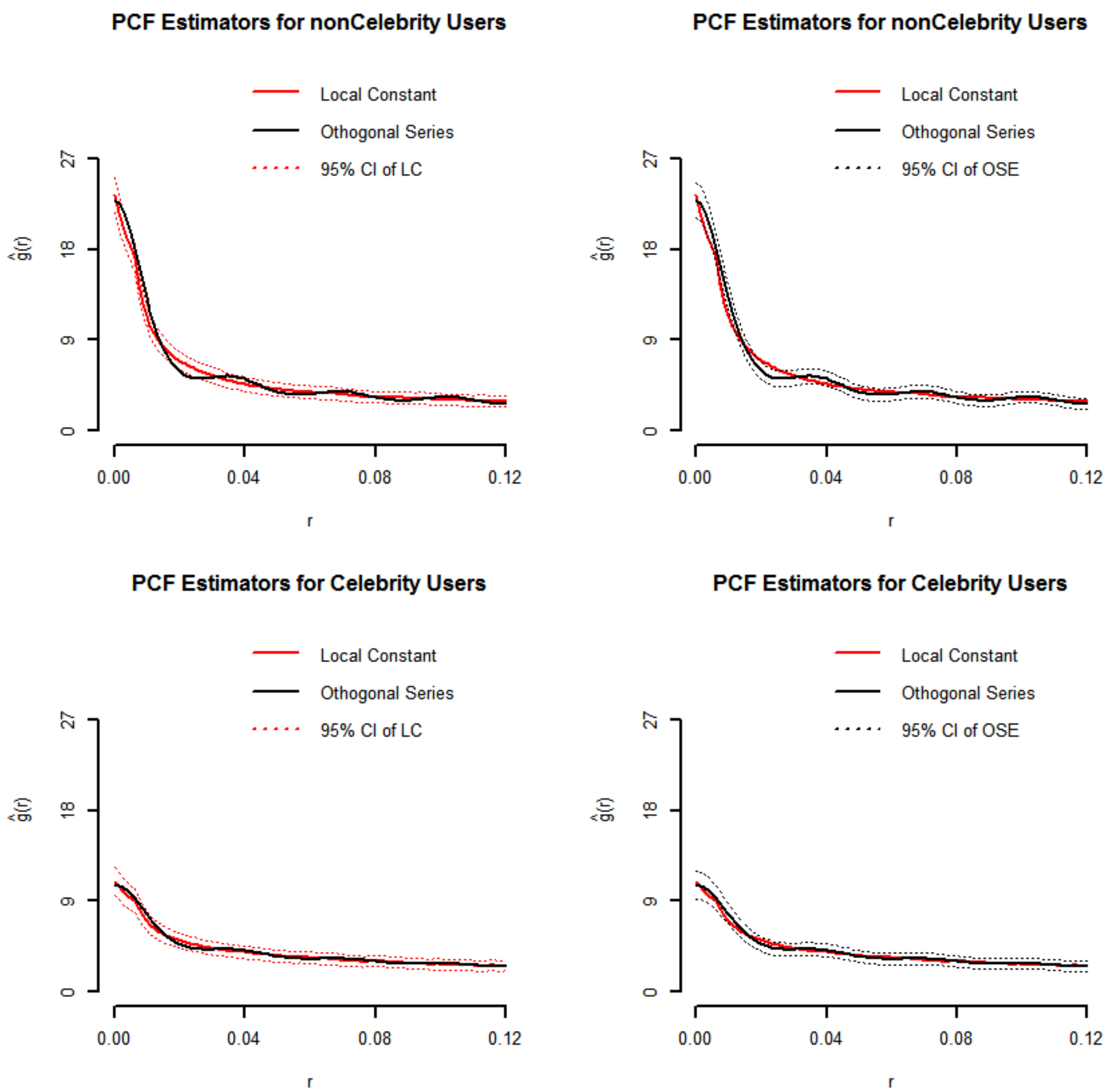

Figure 8.3: Estimated pair correlation functions for Celebrity/nonCelebrity Users

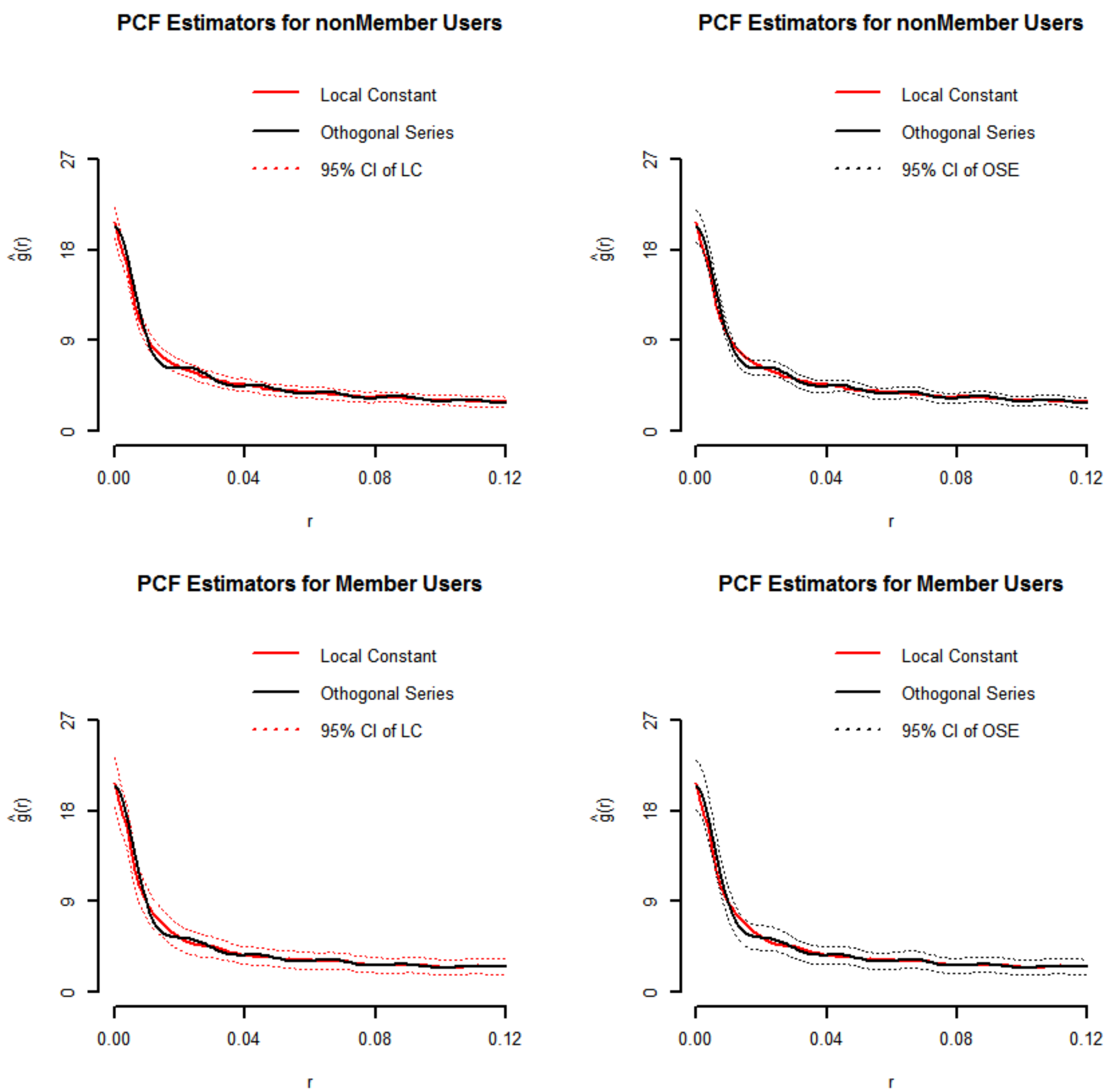

Figure 8.4: Estimated pair correlation functions for Member/nonMember Users

8.2 Pair Correlation Surface Analysis

Through figure 8.3, we find out that the clustered posting pattern within two groups of users has marked difference. To study this posting pattern more closely incorporating this number of followers covariate, we propose a surface analysis of the pair correlation function across the distribution of the number of followers for Weibo users. For the extremely right-skewed distribution of covariate number of followers, we have data range from 24 to $82,437,362$ with nearly 90% users being within $[0, 440, 000]$. Thus, we pick equal-distance covariate grids within this range and extract 250 users from a moving window for each grid to get the estimated pair correlation function. It means we have 125 users with covariate followers value greater(less) than each covariate grid and we have 250 users forming a replicated point patterns within each moving window (The number of users within the moving window may vary due to the edge effect). Through this formulation, we can obtain a surface estimation for users with a different number of followers within time lag range $[0, 0.12]$.

If we look at figure 8.5, we can find that the estimated pair correlation function estimates within small time lags seems to have an overall decreasing trend as their followers number going up which is consistent with the analysis result we obtained from figure 8.3. Users with more followers tend to have less clustered posting pattern. More specially speaking, users who have a small number of followers (which corresponds to the first one-third part of the surface across the 'n' axis) tends to have a large and greatly-varied correlation at small time lag. The users who have a large number of followers tend to have smaller correlation within small time lag but doesn't vary too much comparing with the other users. The stability of these users may be caused by the user sparsity across this followers number range. And it would result in

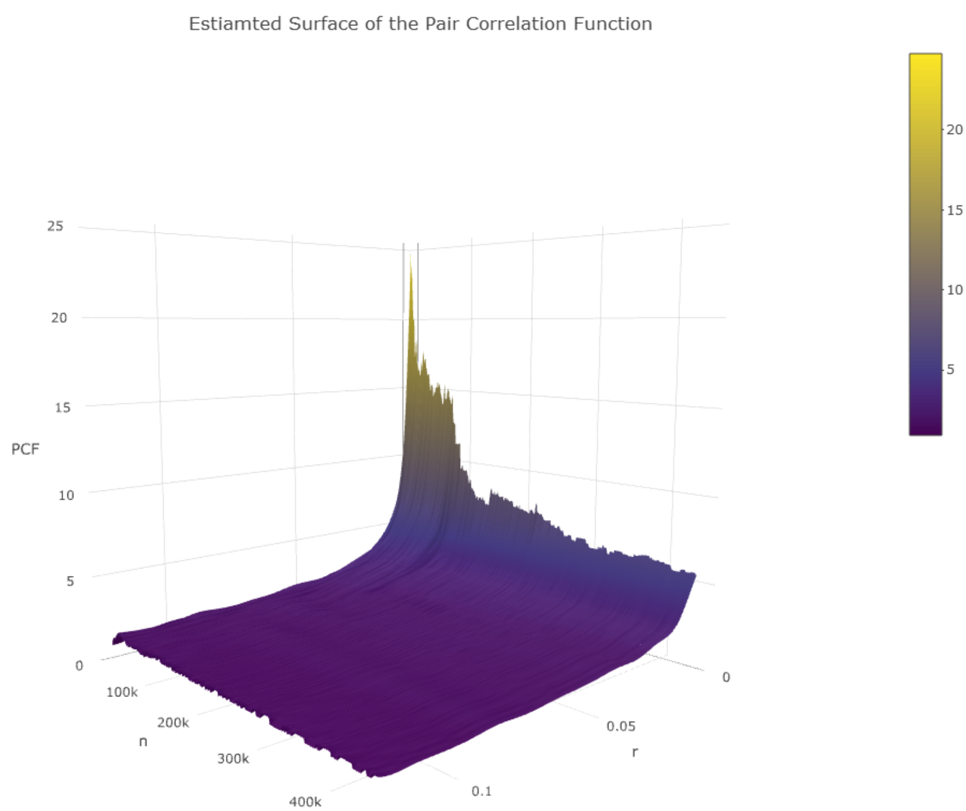

Figure 8.5: Estimated pair correlation function surface

the replicated point patterns of each grid within this followers number range doesn't change too much.

Chapter 9

Semiparametric Estimating Equation Estimators

In chapter 3, we assume that all subjects share a same intensity function without specifying the accurate form. However it might be not reasonable to assume so in reality even though we don't need to know the accurate form in our proposed method. Actually, different subjects may have different point pattern and subjects with similar pattern may tend to belong to same group. In fact, this assumption can be relaxed with small modification of our methods.

9.1 Semiparametric Estimators

Now let's re-assume the intensity function of each subject takes following semi-parametric form:

$$\lambda_i(x) = \lambda_0(x)\psi(Z_i;\beta),\ i = 1, \ldots, m \tag{9.1}$$

where $\lambda_0(x)$ is an unspecified baseline intensity function and $\rho(Z_i, \beta)$ is a semi-parametric function depending on both some predictors $Z_i = (Z_{i1}, \dots, Z_{ip})^\intercal$ and unknown parameters β. We assume that Z_i are time-invariants covariates associated with point process X_i for ease of theoretical derivation. However, our methods can be easily generalized to the case when Z_i is time varying like we mentioned in (3.4).

The regression parameter β can be estimated using the method in Lawless and Nadeau (1995) and Lin et al. (2000) without having to specify $\lambda_0(x)$. More specifically, β is estimated by solving the estimating equation $u_n(\beta) = 0$, where

$$u_n(\beta) = \frac{1}{m_n} \sum_{i=1}^{m_n} u_{n,i}(\beta),$$

and

$$u_{n,i}(\beta) = \sum_{u \in X_i \cap D_n} \left(\frac{\rho^{(1)}(Z_i; \beta)}{\rho(Z_i; \beta)} - \frac{\sum_{j=1}^{m_n} \rho^{(1)}(Z_j; \beta)}{\sum_{j=1}^{m_n} \rho(Z_j; \beta)} \right) \tag{9.2}$$

Through (9.2) we can obtain the consistent estimator of β denotes as $\hat{\beta}$.

9.1.1 Semiparametric Local Linear Estimator

Based on (4.2) and (9.2), the $\hat{U}(\theta)$ in (3.4)

$$\begin{aligned}\tilde{U}(\theta) = &\sum_{i=1}^{m} \sum_{(u,v) \in X_i} \frac{K_h(|u-v|-t)G(|u-v|-t)}{|D \cap (D-u+v)|\rho^2(Z_i, \hat{\beta})} \\ &- \frac{1}{m-1} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{K_h(|u-v|-t)G(|u-v|-t)}{|D \cap (D-u+v)|\rho(Z_i, \hat{\beta})\rho(Z_j, \hat{\beta})} [\theta_1(t) + \theta_2(|u-v|-t)]\end{aligned} \tag{9.3}$$

where $K_h(.)$ and $G(.)$ have the same definition as in (4.2), $\hat{\beta}$ is obtained from last step (9.2). Similar solving method can be applied to this modified estimator and the bandwidth h could still follow a data-driven cross-validation parameter tuning guideline similar to equation (5.1) in Chapter 5.

9.1.2 Semiparametric Orthogonal Series Estimator

Using the same intuition, we can obtain new proposed semiparametric orthogonal series estimator based on (4.6) and (9.2):

$$\begin{aligned}\tilde{U}(\theta) = &\sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{I(|u-v| \leq R)\Phi(|u-v|)}{|D \cap (D-u+v)|\rho^2(Z_i,\hat{\beta})} - \\ &\frac{1}{m-1} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{I(|u-v| \leq R)\Phi(|u-v|)}{|D \cap (D-u+v)|\rho(Z_i,\hat{\beta})\rho(Z_j,\hat{\beta})} \left[\sum_{k=1}^{L} \theta_k \phi_k(|u-v|) \right].\end{aligned} \tag{9.4}$$

for $l = 1, \cdots, L$. Similar to the semiparametric local linear estimator in (9.3), here we could truncate the expansion using L orthogonal basis by the help of data-driven cross validation parameter tuning guideline.

9.2 Estimation Solutions and Extension

To solve these estimating equations under the two modified settings above, we can obtain the matrix A and vector B similar to (4.3) and (4.7) as:

$$
\begin{aligned}
A &= \frac{1}{m(m-1)} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{K_h(|u-v|-r) G(|u-v|-r) G(|u-v|-r)^T}{|D \cap (D-u+v)| \rho(Z_i, \hat{\beta}) \rho(Z_j, \hat{\beta})}, \\
B &= \frac{1}{m} \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{K_h(|u-v|-r) G(|u-v|-r)}{|D \cap (D-u+v)| \rho^2(Z_i, \hat{\beta})}. \qquad (9.5)
\end{aligned}
$$

and

$$
\begin{aligned}
A &= \frac{1}{m(m-1)} \sum_{i \neq j} \sum_{u \in X_i} \sum_{v \in X_j} \frac{I(|u-v| \leq R) \Phi(|u-v|) \Phi(|u-v|)^T}{|D \cap (D-u+v)| \rho(Z_i, \hat{\beta}) \rho(Z_j, \hat{\beta})}, \\
B &= \frac{1}{m} \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \frac{I(|u-v| \leq R) \Phi(|u-v|)}{|D \cap (D-u+v)| \rho^2(Z_i, \hat{\beta})}. \qquad (9.6)
\end{aligned}
$$

In fact, the exponential-transformed estimators we discussed in (4.4) and (4.9) can be easily extend to the semiparametric setting. With this transformation for the estimators, we can guarantee $\hat{g}(r) > 0$. Similarly, the $\tilde{U}(\theta) = 0$ cannot be solved explicitly.

9.3 Parameters Tuning

To choose the optimal bandwidth (dimension) of the modified estimators, we can follow the similar criterion as we introduced in (5.1). The only difference is the estimator of the pair correlation function in this two formula need to be calculated incorporating with the modification in (9.3) and (9.4).

9.4 Asymptotic Properties

In this semiparametric setting, the two estimators actually follow a two-step procedure. The overall asymptotic property of the two estimators are also dependent on the estimation of the regression parameters in the intensity function. Here we also proof the consistency of the two estimators. Rigorous proof can be found in appendix $A.4$ and $A.5$.

Chapter 10

Discussion

In this dissertation, we proposed two nonparametric estimators for the pair correlation function in replicated point patterns. The two estimators are being built by utilizing estimating equation technique. More specially, the estimating equations are constructed by the score function of the weighted composite likelihood function in Guan (2006) and Waagepetersen (2007). Tailored weight functions are being applied to the score function to obtain the estimators respectively. The simple recipe can be extended to similar likelihood based methods in other settings. In fact, we bypass the first order intensity function estimation successfully in this procedure through this estimating equation technique. Moreover, the two estimators can be transformed to exponential version to guarantee the non-negativeness of the pair correlation function estimates with similar performance. One may observe that the two estimators are actually not quite sensitive to the data structure in replicated point patterns and the two can also be generalized to semiparametric setting with ease. In addition, empirical variance estimators are being proposed to make inference for the pair correlation function estimates. A simulation study and theoretical derivation are conducted to investigate its finite and asymptotic properties and to access their performance.

Some improvements can be made for future development on this work. First, we can combine our likelihood based motivation with the penalized model to propose new estimating procedure to complete the model estimation and the parameter tuning simultaneously. Second, the data example we studied in this dissertation need incorporate more traditional analysis in replicated point patterns to get more idea about the new proposed estimators and also the data.

Bibliography

Baddeley, A., Rubak, E., and Turner, R. (2015). *Spatial point patterns: methodology and applications with R.* CRC Press.

Baddeley, A. and Turner, R. (2000). Practical maximum pseudolikelihood for spatial point patterns: (with discussion). *Australian & New Zealand Journal of Statistics*, 42(3):283–322.

Baddeley, A. J., Moyeed, R. A., Howard, C. V., and Boyde, A. (1993). Analysis of a Three-Dimensional Point Pattern with Replication. *Applied Statistics*, 42(4):641–668.

Bell, M. L. and Grunwald, G. K. (2004). Mixed models for the analysis of replicated spatial point patterns. *Biostatistics*, 5:633–648.

Besag, J. (1977). Contribution to the discussion on dr. ripley's paper. *JR Stat. Soc. Ser*, 39:193–195.

Bhatia, R. (1997). *Matrix analysis.* Springer-Verlag, New York.

Diggle, P. (1985). A kernel method for smoothing point process data. *Applied statistics*, pages 138–147.

Diggle, P. J., Lange, N., and Benes, F. M. (1991). Analysis of Variance for Replicated Spatial Point Patterns in Clinical Neuroanatomy. *Journal of the American Statistical Association*, 86:618–625.

Diggle, P. J., Mateu, J., and Clough, H. E. (2000). A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Advances in Applied Probability*, 32(2):331–343.

Efromovich, S. (2008). *Nonparametric curve estimation: methods, theory, and applications.* Springer Science & Business Media.

Efromovich, S. (2010). Orthogonal series density estimation. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):467–476.

Golub, G. H. and Van Loan, C. F. (1996). *Matrix computations. third*. The John Hopkins University Press.

Guan, Y. (2006). A composite likelihood approach in fitting spatial point process models. *Journal of the American Statistical Association*, 101(476):1502–1512.

Guan, Y. (2007a). A composite likelihood cross-validation approach in selecting bandwidth for the estimation of the pair correlation function. *Scandinavian journal of statistics*, 34(2):336–346.

Guan, Y. (2007b). A least-squares cross-validation bandwidth selection approach in pair correlation function estimations. *Statistics & Probability Letters*, 77(18):1722–1729.

Guan, Y. (2011). Second-order analysis of semiparametric recurrent event processes. *Biometrics*, 67(3):730–739.

Hall, P. (1987). Cross-validation and the smoothing of orthogonal series density estimators. *Journal of Multivariate Analysis*, 21(2):189–206.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*, volume 76. Wiley.

Isaacson, E. and Keller, H. B. (1994). *Analysis of numerical methods*. Courier Corporation.

Jalilian, A., Guan, Y., and Waagepetersen, R. (2013). Decomposition of variance for spatial cox processes. *Scandinavian Journal of Statistics*, 40(1):119–137.

Jalilian, A., Guan, Y., and Waagepetersen, R. (2017). Orthogonal series estimation of the pair correlation function of a spatial point process.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168.

Lin, D., Wei, L., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730.

Loader, C. R. et al. (1996). Local likelihood density estimation. *The Annals of Statistics*, 24(4):1602–1618.

Loh, J. M. and Jang, W. (2010). Estimating a cosmological mass bias parameter with bootstrap bandwidth selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(5):761–779.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998). Log gaussian cox processes. *Scandinavian journal of statistics*, 25(3):451–482.

Møller, J. and Waagepetersen, R. P. (2003). *Statistical Inference and Simulation for Spatial Point Processes*. Chapman and Hall/CRC,Boca Raton.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of applied probability*, 13(2):255–266.

Schladitz, K., Särkkä, A., Pavenstädt, I., Haferkamp, O., and Mattfeldt, T. (2003). Statistical analysis of intramembranous particles using freeze fracture specimens. *Journal of microscopy*, 211(2):137–153.

Stoyan, D. and Stoyan, H. (1994). *Fractals, random shapes, and point fields: methods of geometrical statistics*, volume 302. John Wiley & Sons Inc.

Stoyan, D. and Stoyan, H. (2000). Improving ratio estimators of second order point process characteristics. *Scandinavian Journal of Statistics*, 27(4):641–656.

Tolstov, G. P. (1962). *Fourier Series: Translated from the Russian by Richard A. Silverman*. Prentice-Hall.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Waagepetersen, R. and Guan, Y. (2009). Two-step estimation for inhomogeneous spatial point processes. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 71:685–702.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous Neyman-Scott processes. *Biometrics*, 63(1).

Yue, Y. R. and Loh, J. M. (2010). Bayesian nonparametric estimation of pair correlation function for inhomogeneous spatial point processes. *Journal of Nonparametric Statistics*, 25(2):463–474.

Appendices

Appendix A

Asymptotic Consistency

A.1 Consistency of the Log Local linear Estimator

Suppose there exists a vector $\boldsymbol{\theta}^* = (\theta_0^*, \theta_1^*, \cdots, \theta_p^*)^T$ such that

$$\int_{D_n^2} \lambda(u)\lambda(v)w_{r,h}(|u-v|)\left[g(|u-v|) - \tilde{g}_r(|u-v|;\boldsymbol{\theta}^*)\right]\boldsymbol{G}_r(|u-v|)\mathrm{d}u\mathrm{d}v = \mathbf{0}, \quad \text{(A.1)}$$

where $\tilde{g}_r(\cdot;\cdot)$ is defined in (6.2).

The following Lemma quantifies the distance between $g(t)$ and $\tilde{g}_r(t;\boldsymbol{\theta}^*)$.

A.1.1 Lemma. *Under conditions $C1$-$C2$ and $C4$-$C6$, we have that as $h \to 0$,*

$$\begin{aligned} h^j\left[\theta_j^* - f^{(j)}(r)/j!\right] &= O(h^{p+1}), \quad j = 0, 1, \cdots, p, && \text{(A.2)} \\ |g(t) - \tilde{g}_r(t;\boldsymbol{\theta}^*)| &= O(h^{p+1}), \quad \textit{for } t \in [r-h, r+h]. && \text{(A.3)} \end{aligned}$$

Proof. Define function

$$g_{r,0}(t) = \exp\left\{f(r) + f^{(1)}(r)/1!(t-r) + \cdots + f^{(p)}(r)/p!(t-r)^p\right\}.$$

Then it is straightforward to show that

$$\begin{aligned}
&\int_{D_n^2} w_{r,h}(|u-v|)\lambda(u)\lambda(v)\left[g(|u-v|)-g_{r,0}(|u-v|)\right]\boldsymbol{A}_h(|u-v|-r)\mathrm{d}u\mathrm{d}v \\
&=\int_{D_n^2} w_{r,h}(|u-v|)\lambda(u)\lambda(v)g(|u-v|)\left\{1-\exp\left[\log g_{r,0}(|u-v|)-f(|u-v|)\right]\right\}\boldsymbol{A}_h(|u-v|-r)\mathrm{d}u\mathrm{d}v \\
&=\int_{D_n^2} w_{r,h}(|u-v|)\lambda(u)\lambda(v)g(|u-v|)\left\{1-\exp\left[\frac{f^{(p+1)}(r^*_{|u-v|})(|u-v|-r)^{p+1}}{(p+1)!}\right]\right\}\boldsymbol{A}_h(|u-v|-r)\mathrm{d}u\mathrm{d}v \qquad \text{(A.4)} \\
&=O(1)|D_n|\int_{\mathbb{R}} \frac{K_h(|s|-r)}{\gamma_n(|s|)}g(|s|)\left|1-\exp\left[\frac{f^{(p+1)}(r^*_{|s|})(|s|-r)^{p+1}}{(p+1)!}\right]\right|\boldsymbol{A}_h(||s|-r|)\mathrm{d}s \\
&=O(1)\int_{\mathbb{R}} K_h(|s|-r)g(|s|)\left|1-\exp\left[\frac{f^{(p+1)}(r^*_{|s|})(|s|-r)^{p+1}}{(p+1)!}\right]\right|\boldsymbol{A}_h(||s|-r|)\mathrm{d}s,
\end{aligned}$$

where the second last equality follows from condition C1 and the last equality follows from condition C4-C5. Note that by condition C4, the kernel function $K(s)$ has the bounded support in $[-1,1]$, hence $\gamma_n(|s|)$ in the integral only takes value for $|s| \in [r-h, r+h]$. Together with condition C 2(a)-(b), C5 and the fact that $|1-e^x| \le |x|e^{|x|}$, we have that as $h \to 0$,

$$\begin{aligned}
&\int_{D_n^2} w_{r,h}(|u-v|)\lambda(u)\lambda(v)\left[g(|u-v|)-g_{r,0}(|u-v|)\right]\boldsymbol{A}_h(|u-v|-r)\mathrm{d}u\mathrm{d}v \\
&=O(h^{p+1})\int_0^\infty K_h(s-r)\boldsymbol{A}_h(|s-r|)\mathrm{d}s \\
&=O(h^{p+1})\int_{-r/h}^\infty K(s)|\boldsymbol{A}_1(|s|)\mathrm{d}s \\
&=O(h^{p+1}).
\end{aligned}$$

By the definition of $\boldsymbol{\theta}^*$ in (A.1), using the above equation, it is straightforward to see that

$$\int_{D_n^2} w_{r,h}(|u-v|)\lambda(u)\lambda(v)\left[\tilde{g}(|u-v|;\boldsymbol{\theta}^*)-g_{r,0}(|u-v|)\right]\boldsymbol{A}_h(|u-v|-r)\mathrm{d}u\mathrm{d}v = O(h^{p+1}). \qquad \text{(A.5)}$$

Let $\boldsymbol{a} = (a_0, a_1, \cdots, a_p)^T$, define a new multivariate function as follows

$$\boldsymbol{F}(\boldsymbol{a}) = \int_{D_n^2} w_{r,h}(|u-v|)\lambda(u)\lambda(v)q_{\boldsymbol{a},h}(|u-v|-r)\boldsymbol{A}_h(|u-v|-r)\mathrm{d}u\mathrm{d}v,$$

where $q_{\boldsymbol{a},h}(t) = \exp\left(a_0 + a_1 t/h + \cdots + a_p t^p/h^p\right)$. Then equation (A.5) immediately yields that

$$\boldsymbol{F}(\theta_0^*, \theta_1^* h, \cdots, \theta_p^* h^p) - \boldsymbol{F}(f(r), hf^{(1)}(r)/1!, \cdots, h^p f^{(p)}(r)/p!) = O(h^{p+1}). \quad \text{(A.6)}$$

Some straightforward change of variables tricks give that

$$\boldsymbol{F}(\boldsymbol{a}) = \int_{D_n} \int_{-r/h}^{\infty} [I(u - hs - r \in D_n)\lambda(u - hs - r) + I(u + hs + r \in D_n)\lambda(u + sh + r)] \\ \times \lambda(u) \frac{K(s)}{\gamma_n(hs + r)} q_{\boldsymbol{a},1}(s) \boldsymbol{A}_1(s) \mathrm{d}s\mathrm{d}u.$$

The partial derivatives of $\boldsymbol{F}(\boldsymbol{a})$ then becomes

$$\frac{\partial \boldsymbol{F}(\boldsymbol{a})}{\partial \boldsymbol{a}^T} = \int_{D_n} \int_{-r/h}^{\infty} [I(u - hs - r \in D_n)\lambda(u - hs - r) + I(u + hs + r \in D_n)\lambda(u + sh + r)] \\ \times \lambda(u) \frac{K(s)}{\gamma_n(hs + r)} q_{\boldsymbol{a},1}(s) \boldsymbol{A}_1(s) \boldsymbol{A}_1^T(s) \mathrm{d}s\mathrm{d}u.$$

Plugging in $\boldsymbol{a}_0 = (f(r), 0, \cdots, 0)^T$ back to the above equation, we have that

$$\frac{\partial \boldsymbol{F}(\boldsymbol{a})}{\partial \boldsymbol{a}^T}|_{\boldsymbol{a}=\boldsymbol{a}_0} \equiv g(r) \mathbf{Q}_{n,h}^{(1)}(r),$$

where $\mathbf{Q}_{n,h}^{(1)}(r)$ is as defined in condition C6. Using condition C2(a) and C6, we have that $\frac{\partial \boldsymbol{F}(\boldsymbol{a})}{\partial \boldsymbol{a}^T}$ is strictly positive definite around $\boldsymbol{a}_0 = (f(r), 0, \cdots, 0)^T$. Based on equation (A.6) and a simple application of the inverse function theorem (**?**, page 223) to function $\boldsymbol{F}(\boldsymbol{a})$ in the neighborhood of $\boldsymbol{a}_0 = (f(r), 0, \cdots, 0)^T$ gives that

$$h^j \left[\theta_j^* - f^{(j)}(r)/j!\right] = O(h^{p+1}), \quad j = 0, 1, \cdots, p.$$

Similar argument has been used in, e.g., Loader et al. (1996). When $j = 0$, we have that

$$|\exp(\theta_0^*) - g(r)| = g(r)|\exp(\theta_0^* - f(r)) - 1| \leq g(r)|\theta_0^* - f(r)| \exp(|\theta_0^* - f(r)|) = O(h^{p+1}).$$

Finally, we have that for $|t-r| \le h$,

$$
\begin{aligned}
&|g(t) - \tilde{g}_r(t;\boldsymbol{\theta}^*)| = g(t)|1 - \exp\left[\theta_0^* + \theta_1^*(t-r) + \cdots + \theta_p^*(t-r)^p - f(t)\right]| \\
&= g(t)\Bigg|1 - \exp\Bigg[\theta_0^* - f(r) + h(\theta_1^* - f^{(1)}(r)/1!)\frac{(t-r)}{h} + \cdots \\
&+ h^p(\theta_p^* - f^{(p)}(r))\frac{(t-r)^p/p!}{h^p} + O(h^{p+1})\Bigg]\Bigg| \\
&= O(h^{p+1}),
\end{aligned}
$$

which concludes the proof. □

A.1.2 Lemma. *Under conditions $C1$-$C6$, we have that as $m|D_n|h \to \infty$ and $h \to 0$,*

$$
\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_h = O_p\left(\frac{1}{\sqrt{m|D_n|h}}\right), \tag{A.7}
$$

where the norm $\|\boldsymbol{x}\|_h^2 = x_0^2 + (hx_1)^2 + \cdots + (h^p x_p)^2$ for any $\boldsymbol{x} = (x_0, x_1, \ldots, x_p)^T \in \mathbb{R}^{p+1}$ and $\boldsymbol{\theta}^*$ is defined in equation (A.1).

Proof. It is straightforward to see that solving estimating equation (6.3) is equivalent to maximizing the following loss function with respect to $\boldsymbol{\theta}$

$$
\begin{aligned}
\tilde{L}_{r,h}(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{u,v\in X_i}\sum^{\neq} w_{r,h}(|u-v|)\log\left[\tilde{g}_r(|u-v|;\boldsymbol{\theta})\right] \\
- \frac{1}{m(m-1)}\sum_{i\neq j}\sum\ \sum_{u\in X_i, v\in X_j}\sum w_{r,h}(|u-v|)\tilde{g}_r(|u-v|;\boldsymbol{\theta}),
\end{aligned}
\tag{A.8}
$$

whose Hessian matrix is negative definitive as follows

$$
\frac{\partial^2 \tilde{L}_{r,h}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = -\sum_{i\neq j}\sum\ \sum_{u\in X_i, v\in X_j}\sum \frac{w_{r,h}(|u-v|)}{m(m-1)}\tilde{g}_r(|u-v|;\boldsymbol{\theta})\boldsymbol{G}_r(|u-v|)\boldsymbol{G}_r^T(|u-v|),
$$

which implies that $\tilde{L}_{r,h}(\boldsymbol{\theta})$ is a concave function of $\boldsymbol{\theta}$.

We shall first show that for any given ε there exists a large constant C such that, for large m and/or large n,

$$
P\left\{\sup_{\|\boldsymbol{\delta}_{m,n}\|_h=1} \tilde{L}_{r,h}(\boldsymbol{\theta}^* + CJ_{m,n}^{-1/2}\boldsymbol{\delta}_{m,n}) < \tilde{L}_{r,h}(\boldsymbol{\theta}^*)\right\} \ge 1-\varepsilon. \tag{A.9}
$$

Inequality (A.9) implies that with probability tending to 1 there is a local maximum, denoted as $\widehat{\boldsymbol{\theta}}$, in the ellipsoid $\{\boldsymbol{\theta}^* + J_{m,n}^{-1/2} C \boldsymbol{\delta}_{m,n} : \|\boldsymbol{\delta}_{m,n}\|_h = 1\}$ such that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_h^2 = O_p(J_{m,n}^{-1})$.

To show (A.9), define a function of z as

$$H_{m,n}(z) = -\tilde{L}_{r,h}(\boldsymbol{\theta}^* + z J_{m,n}^{-1/2} \boldsymbol{\delta}_{m,n}). \tag{A.10}$$

Consider the Taylor expansion of the loss function $\tilde{L}_{r,h}(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$

$$\begin{aligned} \tilde{L}_{r,h}(\boldsymbol{\theta}^* + C J_{m,n}^{-1/2} \boldsymbol{\delta}_{m,n}) - \tilde{L}_{r,h}(\boldsymbol{\theta}^*) &= C J_{m,n}^{-1/2} \left[\tilde{\boldsymbol{U}}_{r,h}(\boldsymbol{\theta}^*)\right]^T \boldsymbol{\delta}_{m,n} \\ &\quad + \frac{1}{2} C^2 J_{m,n}^{-1} \boldsymbol{\delta}_{m,n}^T \left[\frac{\partial^2 \tilde{L}_{r,h}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} |_{\boldsymbol{\theta} = \tilde{\boldsymbol{\theta}}_{m,n}^*}\right] \boldsymbol{\delta}_{m,n} \\ &= -C \left[H'_{m,n}(0) + \frac{C}{2} H''_{m,n}(z_0)\right], \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}_{m,n}^* = \boldsymbol{\theta}^* + z_0 J_{m,n}^{-1/2} \boldsymbol{\delta}_{m,n}$ for some $0 < z_0 < C$. By definition, $H_{m,n}(z)$ is a convex function of z since $H''_{m,n}(z) \geq 0$ for any constant z. Therefore, to show (A.9), it suffices to show that

$$H'_{m,n}(0) = O_p\left[H''_{m,n}(z_0)\right], \text{ for any finite } z_0. \tag{A.11}$$

By the definition of $H_{m,n}(\cdot)$ in (A.10), it is straightforward to show that

$$\begin{aligned} H'_{m,n}(0) = &-\frac{J_{m,n}^{-1/2}}{m} \sum_{i=1}^{m} \sum\sum_{u,v \in X_i}^{\neq} w_{r,h}(|u-v|) \boldsymbol{G}_r(|u-v|)^T \boldsymbol{\delta}_{m,n} \\ &+ \frac{J_{m,n}^{-1/2}}{m} \sum\sum_{i \neq j} \sum\sum_{u \in X_i, v \in X_j} \frac{w_{r,h}(|u-v|)}{m-1} \tilde{g}_r(|u-v|; \boldsymbol{\theta}^*) \boldsymbol{G}_r(|u-v|)^T \boldsymbol{\delta}_{m,n}, \end{aligned} \tag{A.12}$$

$$H''_{m,n}(z_0) = \frac{J_{m,n}^{-1}}{m} \sum\sum_{i \neq j} \sum\sum_{u \in X_i, v \in X_j} \frac{w_{r,h}(|u-v|)}{m-1} \tilde{g}_r(|u-v|; \tilde{\boldsymbol{\theta}}_{m,n}^*) \left[\boldsymbol{G}_r(|u-v|)^T \boldsymbol{\delta}_{m,n}\right]^2, \tag{A.13}$$

where $\tilde{\boldsymbol{\theta}}_{m,n}^* = \boldsymbol{\theta}^* + z_0 J_{m,n}^{-1/2} \boldsymbol{\delta}_{m,n}$.

We first investigate $H'_{m,n}(0)$. By the definition of $\boldsymbol{\theta}^*$ in (A.1), we have that

$\mathbb{E}\left[H'_{m,n}(0)\right] = 0$. Furthermore, the variance can be shown as

$$
\begin{aligned}
mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right] = &\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)g^{(4)}(u_1,v_1,u_2,v_2) \\
&\times \left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ 4\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)g^{(3)}(u_1,v_1,u_2) \\
&\times \left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ 2\int_{D_n^2} \lambda(u_1)\lambda(v_1)\left[w_{r,h}(|u_1-v_1|)\right]^2 g(|u_1-v_1|)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}v_1 \\
&+ (m-1)\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)g(|u_1-v_1|) \\
&\times g(|u_2-v_2|)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&- 4\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)g^{(3)}(u_1,v_1,u_2) \\
&\times \tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta})\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&- 8\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)g(|u_1-v_1|)\tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*) \\
&\times \left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&- 2(m-2)\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)g(|u_1-v_1|) \\
&\times \tilde{g}_r(u_2,v_2;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ \frac{2}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*) \\
&\times \tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)g(|u_1-u_2|)g(|v_1-v_2|)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right] \\
&\qquad \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ \frac{4}{m-1}\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*) \\
&\times \tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*)g(|v_1-u_2|)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ \frac{2}{m-1}\int_{D_n^2} \lambda(u_1)\lambda(v_1)w_{r,h}^2(|u_1-v_1|)\tilde{g}_r^2(|u_1-v_1|;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}v_1 \\
&+ \frac{4(m-2)}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)g(|u_1-u_2|) \\
&\times \tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right] \\
&\qquad \mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v_1\mathrm{d}v_2 \\
&+ \frac{4(m-2)}{m-1}\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*) \\
&\times \left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ \frac{(m-2)(m-3)}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*) \\
&\times \tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2
\end{aligned}
$$

Using the definition of $\boldsymbol{\theta}^*$ in (A.1), we can simply the Var $\left[H'_{m,n}(0)\right]$ as following

$$
\begin{aligned}
mJ_{m,n}\text{Var}\left[H'_{m,n}(0)\right] = &\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)\big[g^{(4)}(u_1,v_1,u_2,v_2)\\
&- g(|u_1-v_1|)g(|u_2-v_2|)\big] \times \left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2\\
&- 4\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)[g^{(3)}(u_1,v_1,u_2) - g(|u_1-v_1|)]\\
&\times \tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 +\\
&\frac{2}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)\\
&\times [g(|u_1-u_2|)g(|v_1-v_2|) - 1]\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2\\
&+ \frac{4(m-2)}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)[g(|u_1-u_2|) - 1]\\
&\times \tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]du_1du_2dv_1dv_2\\
&+ 4\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)g^{(3)}(u_1,v_1,u_2)\\
&\times \left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\\
&- 4\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\left[2g(|u_1-v_1|) - \tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\right]\\
&\times \tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\\
&+ \frac{4}{m-1}\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\\
&\times \tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*)[g(|v_1-u_2|) - 1]\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\\
&+ 2\int_{D_n^2} \lambda(u_1)\lambda(v_1)\left[w_{r,h}(|u_1-v_1|)\right]^2 g(|u_1-v_1|)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}v_1\\
&+ \frac{2}{m-1}\int_{D_n^2} \lambda(u_1)\lambda(v_1)w_{r,h}^2(|u_1-v_1|)\tilde{g}_r^2(|u_1-v_1|;\boldsymbol{\theta}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}v_1
\end{aligned}
$$

Note that by definition, $\boldsymbol{G}_r(t)^T\boldsymbol{\delta}_{m,n} = \boldsymbol{A}_h(t-r)\boldsymbol{\eta}_{m,n}$ for some $\|\boldsymbol{\eta}_{m,n}\|^2 = 1$, which implies that $|\boldsymbol{G}_r(t)^T\boldsymbol{\delta}_{m,n}|^2 \le \|\boldsymbol{A}_h(t-r)\|^2 \le p+1$ for any $r-h \le t \le r+h$. Therefore, under conditions C1, C2(a)-(b), C4 and equation (A.3), we can further simplify $mJ_{m,n}\text{Var}\left[H'_{m,n}(0)\right]$

as follows

$$
\begin{aligned}
mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right] = O(1)\int_{D_n^4} & w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)\big|g^{(4)}(u_1,v_1,u_2,v_2) \\
& \qquad - g(|u_1-v_1|)g(|u_2-v_2|)\big|\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
& + O(1)\int_{D_n^4} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)|g^{(3)}(u_1,v_1,u_2)-g(|u_1-v_1|)|\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
& + \frac{1}{m}O(1)\int_{D_n^4} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)|g(|u_1-u_2|)g(|v_1-v_2|)-1|\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
& + O(1)\int_{D_n^4} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)|g(|u_1-u_2|)-1|du_1du_2dv_1dv_2 \\
& + O(1)\int_{D_n^3} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
& + \frac{1}{m}O(1)\int_{D_n^3} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)|g(|v_1-u_2|)-1|\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
& + O(1)\int_{D_n^2}\left[w_{r,h}(|u_1-v_1|)\right]^2\mathrm{d}u_1\mathrm{d}v_1 \\
= O(1)|D_n|\int_{\mathbb{R}^3} & \frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\big|g^{(4)}(s,t+w,w)-g(|s|)g(|t|)\big|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
& + O(1)|D_n|\int_{\mathbb{R}^3}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g^{(3)}(s,w)-g(|s|)|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
& + \frac{1}{m}O(1)|D_n|\int_{\mathbb{R}^3}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g(|w|)g(|t-s+w|)-1|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
& + O(1)|D_n|\int_{\mathbb{R}^3}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g(|w|)-1|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
& + O(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t \\
& + \frac{1}{m}O(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g(|s-t|)-1|\mathrm{d}s\mathrm{d}t \\
& + O(1)|D_n|\int_{\mathbb{R}}\left[\frac{K_h(|s|-r)}{\gamma_n(|s|)}\right]^2\mathrm{d}s \\
= O(1)|D_n|\int_{\mathbb{R}^2} & \frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t + O(1)|D_n|\int_{\mathbb{R}}\left[\frac{K_h(|s|-r)}{\gamma_n(|s|)}\right]^2\mathrm{d}s,
\end{aligned}
$$

where the last equality follows from condition C3.

Finally, using condition C5, we have that

$$
\begin{aligned}
mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right] &= O(1)|D_n|^{-1}\int_{\mathbb{R}^2}K(|s|-r/h)K(|t|-r/h)\mathrm{d}s\mathrm{d}t \\
&\quad + O(1)|D_n|^{-1}h^{-1}\int_{\mathbb{R}}\left[K(|s|-r/h)\right]^2\mathrm{d}s \\
&= O\left(\frac{1}{|D_n|}\right) + O\left(\frac{1}{|D_n|h}\right) \\
&= O\left(\frac{1}{|D_n|h}\right).
\end{aligned}
$$

Combing with the fact that $\mathbb{E}\left[H'_{m,n}(0)\right]=0$, we have that

$$H'_{m,n}(0)=O_p\left(\frac{1}{\sqrt{m|D_n|hJ_{m,n}}}\right). \tag{A.14}$$

Now we proceed to study $H''_{m,n}(z_0)$. Some tedious algebra gives that

$$\begin{aligned}
J_{m,n}^2 m(m-1)\mathrm{Var}\left[H''_{m,n}(z_0)\right] &= 2\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)\tilde{g}_r(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)\\
&\times \tilde{g}_r(|u_2-v_2|;\tilde{\boldsymbol{\theta}}^*)[g(|u_1-u_2|)g(|v_1-v_2|)-1]\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]^2\\
&\qquad \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2\\
&+4\int_{D_n^3}\lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\tilde{g}_r(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)\\
&\times \tilde{g}_r(|u_1-u_2|;\tilde{\boldsymbol{\theta}}^*)g(|v_1-u_2|)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\\
&+2\int_{D_n^2}\lambda(u_1)\lambda(v_1)w_{r,h}^2(|u_1-v_1|)\tilde{g}_r^2(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)g(|u_1-v_1|)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^4\mathrm{d}u_1\mathrm{d}v_1\\
&+4(m-2)\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)[g(|u_1-u_2|)-1]\\
&\times \tilde{g}_r(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)\tilde{g}_r(|u_2-v_2|;\tilde{\boldsymbol{\theta}}^*)\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\left[\boldsymbol{G}_r(|u_2-v_2|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v_1\mathrm{d}v_2\\
&+4(m-2)\int_{D_n^3}\lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\tilde{g}_r(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)\tilde{g}_r(|u_1-u_2|;\tilde{\boldsymbol{\theta}}^*)\\
&\times\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\left[\boldsymbol{G}_r(|u_1-u_2|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2.
\end{aligned}$$

Recall that we have shown that $|\boldsymbol{G}_r(t)^T\boldsymbol{\delta}_{m,n}|^2\le p+1$ for any $r-h\le t\le r+h$. Then under conditions C1, C2(a)-(b), C4 and equation (A.3), we can further simplify $J_{m,n}^2 m(m-$

1)Var $\left[H''_{m,n}(z_0)\right]$ as follows

$$
\begin{aligned}
J^2_{m,n}m(m-1)\mathrm{Var}\left[H''_{m,n}(z_0)\right] &= O(1)\int_{D_n^4} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)|g(|u_1-u_2|)g(|v_1-v_2|)-1| \\
&\qquad \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ O(1)\int_{D_n^3} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 + O(1)\int_{D_n^2} w^2_{r,h}(|u_1-v_1|)\mathrm{d}u_1\mathrm{d}v_1 \\
&+ mO(1)\int_{D_n^4} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)|g(|u_1-u_2|)-1|du_1du_2dv_1dv_2 \\
&+ mO(1)\int_{D_n^3} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&= O(1)|D_n|\int_{\mathbb{R}^3}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g(|w|)g(|t-s+w|)-1|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ O(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t + O(1)|D_n|\int_{\mathbb{R}}\left[\frac{K_h(|s|-r)}{\gamma_n(|s|)}\right]^2\mathrm{d}s \\
&+ mO(1)|D_n|\int_{\mathbb{R}^3}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g(|w|)-1|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ mO(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t \\
&= O(1)|D_n|\int_{\mathbb{R}}\left[\frac{K_h(|s|-r)}{\gamma_n(|s|)}\right]^2\mathrm{d}s + mO(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t
\end{aligned}
$$

Then, by condition C5, we finally have that

$$J^2_{m,n}m(m-1)\mathrm{Var}\left[H''_{m,n}(z_0)\right] = \frac{1}{|D_n|h}O(1) + \frac{m}{|D_n|}O(1),$$

which implies that

$$\mathrm{Var}\left[H''_{m,n}(z_0)\right] = O\left(\frac{1}{J^2_{m,n}m^2|D_n|h} + \frac{1}{J^2_{m,n}m|D_n|}\right). \tag{A.15}$$

On the other hand, we have that

$$\mathbb{E}\left[H''_m(z_0)\right] = J^{-1}_{m,n}\int_{D_n^2}\lambda(u_1)\lambda(v_1)w_{r,h}(|u_1-v_1|)\tilde{g}_r(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*_{m,n})\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2\mathrm{d}u_1\mathrm{d}v_1.$$

Observe that by definition $\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n} = \boldsymbol{A}^T_h(|u_1-v_1|-r)\boldsymbol{\eta}_{m,n}$ with $\|\boldsymbol{\eta}_{m,n}\|^2 = 1$ and the fact that the smallest eigenvalue satisfy the condition

$$\eta_{\min}\left[\mathbf{Q}^{(1)}_{n,h}(r)\right] = \inf_{\|\boldsymbol{\eta}\|^2=1}\boldsymbol{\eta}^T\mathbf{Q}^{(1)}_{n,h}(r)\boldsymbol{\eta}.$$

Using definition of $\mathbf{Q}_{n,h}^{(1)}(r)$ in condition C6, we have that

$$\begin{aligned}
\mathbb{E}\left[J_{m,n}H_m''(z_0)\right] - g(r)\eta_{\min}\left[\mathbf{Q}_{n,h}^{(1)}(r)\right] &\geq \int_{D_n^2} \lambda(u_1)\lambda(v_1)w_{r,h}(|u_1-v_1|) \\
&\times \left[\tilde{g}_r(|u_1-v_1|;\tilde{\boldsymbol{\theta}}_{m,n}^*) - g(r)\right]\left[\boldsymbol{G}_r(|u_1-v_1|)^T\boldsymbol{\delta}_{m,n}\right]^2 \mathrm{d}u_1\mathrm{d}v_1 \\
= O(1)\int_{D_n^2} w_{r,h}(|u_1-v_1|)\left|\tilde{g}_r(|u_1-v_1|;\tilde{\boldsymbol{\theta}}_{m,n}^*) - g(r)\right| \mathrm{d}u_1\mathrm{d}v_1 \\
= O(1)|D_n|\int_{\mathbb{R}} w_{r,h}(|s|)\left|\tilde{g}_r(|s|;\tilde{\boldsymbol{\theta}}_{m,n}^*) - g(r)\right| \mathrm{d}s
\end{aligned}$$

Note that by the definition of $\tilde{\boldsymbol{\theta}}_{m,n}^* = \boldsymbol{\theta}^* + z_0J_{m,n}^{-1/2}\boldsymbol{\delta}_{m,n}$ and equation (A.2) of Lemma A.1.1, it is straightforward to show that

$$\left|\tilde{g}_r(t;\tilde{\boldsymbol{\theta}}_{m,n}^*) - g(r)\right| = g(r)O\left(h + J_{m,n}^{-1/2}\right), \text{ for any } r-h \leq t \leq r+h.$$

Therefore, we have that, under conditions C4-C5,

$$\begin{aligned}
\mathbb{E}\left[J_{m,n}H_m''(z_0)\right] - g(r)\eta_{\min}\left[\mathbf{Q}_{n,h}^{(1)}(r)\right] &\geq O(J_{m,n}^{-1/2}+h)|D_n|\int_{\mathbb{R}} \frac{K_h(|s|-r)}{\gamma_n(|s|)}\mathrm{d}s \\
&= O(J_{m,n}^{-1/2}+h).
\end{aligned}$$

By condition C2(a) and C6, the above equation gives that

$$\mathbb{E}\left[J_{m,n}H_m''(z_0)\right] \geq c_gc_0 + O(J_{m,n}^{-1/2}+h). \tag{A.16}$$

Hence for the constant $c = c_0c_g$, we have that

$$\begin{aligned}
&P\left(J_{m,n}H_m''(z_0) < c/2\right) = P\left\{J_{m,n}H_m''(z_0) - \mathbb{E}\left[J_{m,n}H_m''(z_0)\right] < c/2 - \mathbb{E}\left[J_{m,n}H_m''(z_0)\right]\right\} \\
&\leq P\Big\{|J_{m,n}H_m''(z_0) - \mathbb{E}\left[J_{m,n}H_m''(z_0)\right]| \\
&> |c/2 - \mathbb{E}\left[J_{m,n}H_m''(z_0)\right]|\Big\} I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right] > c/2\right\} + I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right] \leq c/2\right\} \\
&\leq \frac{\mathrm{Var}\left[J_{m,n}H_{m,n}''(z_0)\right]}{|c/2 - \mathbb{E}\left[J_{m,n}H_m''(z_0)\right]|^2} I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right] > c/2\right\} + I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right] \leq c/2\right\} \\
&= O\left(\frac{1}{m^2|D_n|h} + \frac{1}{m|D_n|}\right) + o(1),
\end{aligned}$$

where the last equality follows from equations (A.15) and (A.16) as $J_{m,n} \to \infty$ and $h \to 0$. Therefore, as long as $m|D_n|h \to \infty$, $J_{m,n} \to \infty$ and $h \to 0$, we have that

$$P\left(J_{m,n}H_m''(z_0) \geq c_0c_g/2\right) \to 1, \tag{A.17}$$

where c_g, c_0 are constants defined in conditions C2(a) and C6, respectively.

We have already shown in equation (A.14) that

$$H'_{m,n}(0) = O_p\left(\frac{1}{\sqrt{J_{m,n} m|D_n|h}}\right),$$

hence as long as $\frac{J_{m,n}}{m|D_n|h} = O(1)$, we have that $H'_{m,n}(0) = O_p(H''_m(z_0))$. In other words, by taking $J_{m,n} = m|D_n|h$, (A.9) holds, which completes the proof of equation (A.7). □

A.2 Consistency of the Orthogonal Series Estimator

Let $X_1, \ldots, X_{m_n}$ be independent point processes on $\mathbb{R}$ with the same intensity function $\lambda(x)$ and pair correlation function $g(|x-y|)$. We assume that

A1 $D_n \subset \mathbb{R}$ is an arbitrary sequence of observation windows,

A2 m_n is and increasing sequence of natural numbers such that $m_n \to \infty$, and

A3 L_n is an increasing sequence of natural numbers such that $L_n \to \infty$.

Given a complete orthonormal basis of functions $\phi_l(r)$ on $[0, R]$, the orthogonal series expansion of the square-integrable pair correlation function $g(r)$ on $[0, R]$ is given by

$$g(r) = \sum_{l=1}^{\infty} \theta_l \phi_l(r), \quad \text{where} \quad \theta_l = \int_0^R g(r) \mathrm{d}r.$$

By Parseval's identity (see Tolstov, 1962, p. 119),

$$\sum_{l=1}^{\infty} \theta_l^2 = \int_0^R g^2(r) \mathrm{d}r < \infty,$$

and hence $\sum_{k=l}^{\infty} \theta_k^2 \to 0$ and $\theta_l \to 0$, as $l \to \infty$. Let

$$g_n(r) = \sum_{l=1}^{L_n} \theta_l \phi_l(r) \quad \text{and} \quad \zeta_n(r) = g(r) - g_n(r) = \sum_{l=L_n+1}^{\infty} \theta_l \phi_l(r).$$

Then for each $r \in (0, R)$, $|\zeta_n(r)| \to 0$. In addition (see Tolstov, 1962, p. 55),

$$\int_0^R \zeta_n^2(r)\mathrm{d}r = \sum_{l=L_n+1}^{\infty} \theta_l^2 \to 0.$$

Let

$$\gamma_n(h) = |D_n \cap (D_n - h)| = \int_{\mathbb{R}} \mathbb{I}[y \in D_n, y + h \in D_n]\mathrm{d}y, \quad h \in \mathbb{R}.$$

Then for any fixed $h \in \mathbb{R}$, $\gamma_n(h)/|D_n| \to 1$ and hence a constant $0 < C_\gamma \leq 1$ can be found such that $\gamma_n(h) \geq C_\lambda |D_n|$ for all $-R \leq h \leq R$ and sufficiently large n. Define

$$A_{l,k}^{(n)} = \frac{1}{m_n(m_n - 1)} \sum_{i \neq j = 1}^{m_n} \sum_{x \in X_i \cap D_n} \sum_{y \in X_j \cap D_n} \frac{f_{l,k}(|x - y|)}{\gamma_n(x - y)}, \quad l, k = 1, \ldots, L_n,$$

and

$$B_l^{(n)} = \frac{1}{m_n} \sum_{i=1}^{m_n} \sum_{x, y \in X_i \cap D_n} \frac{\mathbb{I}[0 < |x - y| \leq R]}{\gamma_n(x - y)} \phi_l(|x - y|), \quad l = 1, \ldots, L_n,$$

where $f_{l,k}(r) = \mathbb{I}[0 < r \leq R]\phi_l(r)\phi_k(r)$ and $\mathbb{I}[\,\cdot\,]$ denotes the indicator function. Then,

$$M_{l,k}^{(n)} = \mathbb{E}\, A_{l,k}^{(n)} = \int_{(D_n)^2} \frac{f_{l,k}(|x - y|)}{\gamma_n(x - y)} \lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y$$

and

$$V_l^{(n)} = \mathbb{E}\, B_l^{(n)} = \int_{(D_n)^2} \frac{\mathbb{I}[0 < |x - y| \leq R]}{\gamma_n(x - y)} \phi_l(|x - y|) g(|x - y|)\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y.$$

Let $A^{(n)} = [A_{l,k}^{(n)}]$, $M^{(n)} = [M_{l,k}^{(n)}]$, $B^{(n)} = (B_l^{(n)})$, $V^{(n)} = (V_l^{(n)})$ and $\Theta_n = (\theta_1, \ldots, \theta_{L_n})$. The estimating equations (1.3) are equivalent to

$$\sum_{m=1}^{L_n} A_{l,m}^{(n)} \hat{\theta}_m = B_l^{(n)}, \quad l = 1, \ldots, L_n,$$

or $A^{(n)}\hat{\Theta}_n = B^{(n)}$.

We consider the Euclidean norm $\|B^{(n)}\| = \left(\sum_{l=1}^{L_n} \left(B_l^{(n)}\right)^2\right)^{1/2}$ for vectors and the maximum absolute column sum norm (Isaacson and Keller, 1994, p. 9)

$$\|A^{(n)}\|_1 = \max_{1\le k\le L_n} \sum_{l=1}^{L_n} |A_{l,k}^{(n)}|$$

and the Frobenius norm

$$\left\|A^{(n)}\right\|_{\mathrm{F}} = \left(\sum_{l=1}^{L_n}\sum_{k=1}^{L_n}(A_{l,k}^{(n)})^2\right)^{1/2}$$

for matrices. It is known that $\|A^{(n)}\|_1 \le \sqrt{L_n}\|A^{(n)}\|_{\mathrm{F}}$ (see Golub and Van Loan, 1996, p. 56)

The condition S1 is satisfied for the cosine basis

$$\phi_l(r) = \begin{cases} 1/\sqrt{R} & l=1 \\ (\sqrt{2}/\sqrt{R})\cos\big((l-1)\pi r/R\big) & l\ge 2 \end{cases}$$

with $C_\phi = \sqrt{2}/R$. The condition S4 holds in the homogeneous case where $\lambda(x) \equiv \lambda_0$, because

$$\lim_{n\to\infty}\int_{(D_n)^2} \frac{f_{l,k}(|x-y|)}{\gamma_n(x-y)}\mathrm{d}x\mathrm{d}y = 2\int_0^R \phi_l(r)\phi_k(r)\mathrm{d}r = 2\mathbb{I}[l=k].$$

Regarding condition S5, the rate of convergence of θ_l to zero depends on the smoothness of the pair correlation function. In fact, if $g(r)$ is differentiable on $(0, R)$ and $\int_0^R |g'(r)|\mathrm{d}r < \infty$, then $\theta_l = O(l^{-1})$ and if $g(r)$ is twice differentiable and $\int_0^R |g''(r)|\mathrm{d}r < \infty$ then $\theta_l = O(l^{-2})$ (Efromovich, 2008, p. 32). Thus S5 holds if $g(r)$ is twice differentiable and $\int_0^R |g''(r)|\mathrm{d}r < \infty$.

A.2.1 Lemma. *Under conditions S1-S3 and S5,* $\left\|A^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1/2})$, $\|B^{(n)}-V^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1/2})$ *and* $\|V^{(n)}-M^{(n)}\Theta_n\| = O(L_n^{-\delta})$.

Proof. For each $n \in \mathbb{N}$ and $l, k = 1, \ldots, L_n$,

$$\begin{aligned}\mathrm{Var} B_l^{(n)} &= \frac{2}{m_n} \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \leq R]}{\gamma_n^2(x-y)} \phi_l^2(|x-y|) g(|x-y|) \lambda(x)\lambda(y) \mathrm{d}x \mathrm{d}y \\ &+ \frac{4}{m_n} \int_{(D_n)^3} \frac{\mathbb{I}[|x-y| \leq R, |x-z| \leq R]}{\gamma_n(x-y)\gamma_n(x-z)} \phi_l(|x-y|)\phi_l(|x-z|) \\ &\qquad g^{(3)}(x,y,z)\lambda(x)\lambda(y)\lambda(z)\mathrm{d}x\mathrm{d}y\mathrm{d}z \\ &+ \frac{1}{m_n} \int_{(D_n)^4} \frac{\mathbb{I}[|x_1-y_1| \leq R, |x_2-y_2| \leq R]}{\gamma_n(x_1-y_1)\gamma_n(x_2-y_2)} \phi_l(|x_1-y_1|)\phi_l(|x_2-y_2|) \\ &\qquad \left[g^{(4)}(x_1,y_1,x_2,y_2) - g(|x_1-y_1|)g(|x_2-y_2|)\right] \\ &\qquad \lambda(x_1)\lambda(y_1)\lambda(x_2)\lambda(y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2\end{aligned}$$

and

$$\mathrm{Var} A_{l,k}^{(n)} = \frac{q_n^{(2)} + (m_n - 2) q_n^{(3)}}{m_n(m_n-1)},$$

where

$$\begin{aligned}q_n^{(2)} &= 2\int_{(D_n)^2} \frac{f_{l,k}^2(|x-y|)}{\gamma_n^2(x-y)} \lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y \\ &+ 4\int_{(D_n)^3} \frac{f_{l,k}(|x_1-y|) f_{l,k}(|x_2-y|)}{\gamma_n(x_1-y)\gamma_n(x_2-y)} g(|x_1-x_2|)\lambda(x_1)\lambda(x_2)\lambda(y)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y \\ &+ \int_{(D_n)^4} \frac{f_{l,k}(|x_1-y_1|) f_{l,k}(|x_2-y_2|)}{\gamma_n(x_1-y_1)\gamma_n(x_2-y_2)} \\ &\qquad \Big[g(|x_1-x_2|)g(|y_1-y_2|) + g(|x_1-y_2|)g(|y_1-x_2|) - 2\Big] \\ &\qquad \lambda(x_1)\lambda(x_2)\lambda(y_1)\lambda(y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2, \\ q_n^{(3)} &= 4\int_{(D_n)^3} \frac{f_{l,k}(|x_1-y|) f_{l,k}(|x_2-y|)}{\gamma_n(x_1-y)\gamma_n(x_2-y)} \lambda(x_1)\lambda(x_2)\lambda(y)\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}y \\ &+ 4\int_{(D_n)^4} \frac{f_{l,k}(|x_1-y_1|) f_{l,k}(|x_2-y_2|)}{\gamma_n(x_1-y_1)\gamma_n(x_2-y_2)} \left[g(|x_1-x_2|) - 1\right] \\ &\qquad \lambda(x_1)\lambda(x_2)\lambda(y_1)\lambda(y_2)\mathrm{d}x_1\mathrm{d}y_1\mathrm{d}x_2\mathrm{d}y_2.\end{aligned}$$

Since for sufficiently large n, $\gamma_n(h) \geq C_\gamma |D_n|$ and

$$\int_0^R |\phi_l(r)|\mathrm{d}r \leq \left(\int_0^R \phi_l^2(r)\mathrm{d}r\right)^{1/2} \left(\int_0^R \mathrm{d}r\right)^{1/2} = \sqrt{R},$$

using conditions S2 and S3 we obtain

$$
\begin{aligned}
\mathrm{Var} B_l^{(n)} \le\ & \frac{2\lambda_{\max}^2 C_g}{m_n |D_n| C_\gamma} \int_{\mathbb{R}^2} \frac{\mathbb{I}[|h| \le R, y \in D_n, y+h \in D_n]}{\gamma_n(h)} \phi_l^2(|h|) \mathrm{d}y \mathrm{d}h \\
& + \frac{4\lambda_{\max}^3 C_{g^{(3)}}}{m_n |D_n| C_\gamma} \int_{\mathbb{R}^3} \frac{\mathbb{I}[|h_1| \le R, |h_2| \le R, x \in D_n, x - h_1 \in D_n]}{\gamma_n(h_1)} \\
& \qquad \big|\phi_l(|h_1|)\big| \big|\phi_l(|h_2|)\big| \mathrm{d}x \mathrm{d}h_1 \mathrm{d}h_2 \\
& + \frac{\lambda_{\max}^4}{m_n |D_n| C_\gamma} \int_{\mathbb{R}^3} \frac{\mathbb{I}[|h_1| \le R, |h_2| \le R, y_1 \in D_n, y_1 + h_1 \in D_n]}{\gamma_n(h_1)} \\
& \qquad G^{(4)}(h_1, y_1, h_2) \big|\phi_l(|h_1|)\big| \big|\phi_l(|h_2|)\big| \mathrm{d}y_1 \mathrm{d}h_1 \mathrm{d}h_2 \\
\le\ & \frac{4\lambda_{\max}^2 C_g}{m_n |D_n| C_\gamma} \int_0^R \phi_l^2(r) \mathrm{d}r + \frac{8\lambda_{\max}^3 C_{g^{(3)}} + 2\lambda_{\max}^4 C_{g^{(4)}}}{m_n |D_n| C_\gamma} \left(\int_0^R \big|\phi_l(r)\big| \mathrm{d}r \right)^2 \\
=\ & \frac{4\lambda_{\max}^2 C_g + 8\lambda_{\max}^3 C_{g^{(3)}} R + 2\lambda_{\max}^4 C_{g^{(4)}} R}{m_n |D_n| C_\gamma}.
\end{aligned}
$$

Similarly, using conditions S1-S3,

$$
\begin{aligned}
q_n^{(2)} \le\ & \frac{4\lambda_{\max}^2}{|D_n| C_\gamma} \int_0^R \phi_l^2(r) \phi_k^2(r) \mathrm{d}r + \frac{8\lambda_{\max}^3}{|D_n| C_\gamma} \left(\int_0^R \big|\phi_l(r)\big| \big|\phi_k(r)\big| \mathrm{d}r \right)^2 \\
& + \frac{2\lambda_{\max}^4}{|D_n| C_\gamma} \left(2(1 + C_g) \int_{\mathbb{R}} \big|g(|h|) - 1\big| \mathrm{d}h \right) \left(\int_0^R \big|\phi_l(r)\big| \big|\phi_k(r)\big| \mathrm{d}r \right)^2 \\
q_n^{(3)} \le\ & \frac{8\lambda_{\max}^3}{|D_n| C_\gamma} \left(\int_0^R \big|\phi_l(r)\big| \big|\phi_k(r)\big| \mathrm{d}r \right)^2 \\
& + \frac{8\lambda_{\max}^4}{|D_n| C_\gamma} \left(\int_{\mathbb{R}} \big|g(|h|) - 1\big| \mathrm{d}h \right) \left(\int_0^R \big|\phi_l(r)\big| \big|\phi_k(r)\big| \mathrm{d}r \right)^2.
\end{aligned}
$$

and hence

$$
\begin{aligned}
\mathrm{Var} A_{l,k}^{(n)} \le\ & \frac{4\lambda_{\max}^2 C_\phi^2 + 8\lambda_{\max}^3 + 8\lambda_{\max}^4 (1 + C_g) C_I}{m_n (m_n - 1) |D_n| C_\gamma} \\
& + \frac{(m_n - 2)\big(8\lambda_{\max}^3 + 16\lambda_{\max}^4 C_I\big)}{m_n (m_n - 1) |D_n| C_\gamma}.
\end{aligned}
$$

Therefore, $\mathrm{Var} B_l^{(n)} \le C_{V_B} (m_n |D_n|)^{-1}$ and $\mathrm{Var} A_{l,k}^{(n)} \le C_{V_A} (m_n |D_n|)^{-1}$ for sufficiently large n and some $0 < C_{V_B}, C_{V_A} < \infty$. Consequently,

$$\begin{aligned}\mathbb{E}\left\|A^{(n)} - M^{(n)}\right\|_1^2 &\leq L_n\,\mathbb{E}\left\|A^{(n)} - M^{(n)}\right\|_{\mathrm{F}}^2 \\ &= L_n \sum_{l=1}^{L_n}\sum_{k=1}^{L_n} \mathbb{E}\left(A_{l,k}^{(n)} - M_{l,k}^{(n)}\right)^2 \\ &= L_n \sum_{l=1}^{L_n}\sum_{k=1}^{L_n} \mathrm{Var} A_{l,k}^{(n)} \leq C_{V_A}\frac{L_n^3}{m_n|D_n|}\end{aligned}$$

and

$$\mathbb{E}\,\|B^{(n)} - V^{(n)}\|^2 = \sum_{l=1}^{L_n} \mathbb{E}\left(B_l^{(n)} - V_l^{(n)}\right)^2 = \sum_{l=1}^{L_n} \mathrm{Var} B_l^{(n)} \leq C_{V_B}\frac{L_n}{m_n|D_n|}.$$

By Markov's inequality, for any $n \in \mathbb{N}$ and $c > 0$

$$\begin{aligned}\mathbb{P}\left\{L_n^{-3/2}(m_n|D_n|)^{1/2}\left\|A^{(n)} - M^{(n)}\right\|_1 > c\right\} &\leq \frac{\mathbb{E}\left\|A^{(n)} - M^{(n)}\right\|_1^2}{c^2 L_n^3 (m_n|D_n|)^{-1}}, \\ \mathbb{P}\left\{L_n^{-1/2}(m_n|D_n|)^{1/2}\|B^{(n)} - V^{(n)}\| > c\right\} &\leq \frac{\mathbb{E}\,\|B^{(n)} - V^{(n)}\|^2}{c^2 L_n (m_n|D_n|)^{-1}}.\end{aligned}$$

The right hand sides of the above identities can be made arbitrary small, for all n, by choosing sufficiently large c, which means that $\left\|A^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1/2})$ and $\|B^{(n)} - V^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1/2})$ (see Van der Vaart, 2000, p. 10).

Finally,

$$\begin{aligned}V_l^{(n)} &= \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \leq R]}{\gamma_n(x-y)}\phi_l(|x-y|)g(|x-y|)\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \leq R]}{\gamma_n(x-y)}\phi_l(|x-y|)\left(\sum_{k=1}^{\infty}\theta_k\phi_k(|x-y|)\right)\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y \\ &= \sum_{k=1}^{L_n}\theta_k M_{l,k}^{(n)} + \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \leq R]}{\gamma_n(x-y)}\phi_l(|x-y|)\zeta_n(|x-y|)\lambda(x)\lambda(y)\mathrm{d}x\mathrm{d}y,\end{aligned}$$

and hence

$$\left|V_l^{(n)} - \sum_{k=1}^{L_n} \theta_k M_{l,k}^{(n)}\right| \le \lambda_{\max}^2 \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \le R]}{\gamma_n(x-y)} \big|\phi_l(|x-y|)\big|\big|\zeta_n(|x-y|)\big| \mathrm{d}x\mathrm{d}y$$
$$= 2\lambda_{\max}^2 \int_0^R \big|\phi_l(r)\big|\big|\zeta_n(r)\big| \mathrm{d}r$$
$$\le 2\lambda_{\max}^2 \left(\int_0^R \phi_l^2(r)\mathrm{d}r\right)^{1/2} \left(\int_0^R \zeta_n^2(r)\mathrm{d}r\right)^{1/2}$$
$$= 2\lambda_{\max}^2 \left(\sum_{k=L_n+1}^{\infty} \theta_k^2\right)^{1/2}.$$

Thus

$$\|V^{(n)} - M^{(n)}\Theta_n\|^2 = \sum_{l=1}^{L_n} \left(V_l^{(n)} - \sum_{k=1}^{L_n} \theta_k M_{l,k}^{(n)}\right)^2$$
$$\le 4\lambda_{\max}^4 L_n \sum_{l=L_n+1}^{\infty} \theta_l^2.$$

By condition S5,

$$\sum_{l=L_n+1}^{\infty} \theta_l^2 = O(L_n^{-(1+2\delta)}),$$

which guarantees that $L_n \sum_{l=L_n+1}^{\infty} \theta_l^2 = O(L_n^{-2\delta})$ and hence $\|V^{(n)} - M^{(n)}\Theta_n\| = O(L_n^{-\delta})$. □

A.2.2 Lemma. *Assume condition S4 holds and* $L_n^{3/2+\eta}(m_n|D_n|)^{-1} \to 0$. *Then* $\big\|(A^{(n)})^{-1}\big\|_1 = O_{\mathbb{P}}(\big\|(M^{(n)})^{-1}\big\|_1)$.

Proof. From the identity

$$A^{(n)} = M^{(n)} + (A^{(n)} - M^{(n)}) = M^{(n)}\big[I_{L_n} + (M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big],$$

it follows that

$$\big\|(A^{(n)})^{-1}\big\|_1 \le \big\|\big[I_{L_n} + (M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big]^{-1}\big\|_1 \big\|(M^{(n)})^{-1}\big\|_1.$$

But

$$\big\|(M^{(n)})^{-1}\big(A^{(n)} - M_n\big)\big\|_1 \le \big\|(M^{(n)})^{-1}\big\|_1 \big\|A^{(n)} - M^{(n)}\big\|_1,$$

which, from Lemma6.2.1 and condition S4, implies that $\|(M^{(n)})^{-1}(A^{(n)} - M_n)\|_1 = O_{\mathbb{P}}(L_n^{3/2+\eta}(m_n|D_n|)^{-1/2})$. Thus, if $L_n^{3/2+\eta}(m_n|D_n|)^{-1/2} \to 0$, $\|(M^{(n)})^{-1}(A^{(n)} - M_n)\|_1 \xrightarrow{\mathbb{P}} 0$.

Let $E_n = \{\|(M^{(n)})^{-1}(A^{(n)} - M_n)\|_1 < 1\}$. Then $\lim_{n\to\infty} \mathbb{P}(\Omega \backslash E_n) = 0$ and for any $\omega \in E_n$, $\|(M^{(n)})^{-1}(A^{(n)}(\omega) - M_n)\|_1 < 1$ and hence (Isaacson and Keller, 1994, p. 16)

$$\|[I_{L_n} + (M^{(n)})^{-1}(A^{(n)}(\omega) - M_n))]^{-1}\|_1 \le \frac{1}{1 - \|(M^{(n)})^{-1}(A^{(n)}(\omega) - M_n)\|_1},$$

which means that $\|[I_{L_n} + (M^{(n)})^{-1}(A^{(n)}(\omega) - M_n))]^{-1}\|_1 \le C_M$ for some $1 \le C_M < \infty$. Therefore, for any $\omega \in E_n$, $\|(A^{(n)}(\omega))^{-1}\|_1 \le C_M \|(M^{(n)})^{-1}\|_1$ and

$$\begin{aligned}
&\lim_{n\to\infty} \mathbb{P}\{\|(A^{(n)})^{-1}\|_1 \le C_M \|(M^{(n)})^{-1}\|_1\} = \lim_{n\to\infty} \mathbb{P}\left(E_n \cap \{\|(A^{(n)})^{-1}\|_1 \le C_M \|(M^{(n)})^{-1}\|_1\}\right) \\
&+ \lim_{n\to\infty} \mathbb{P}\left((\Omega\backslash E_n) \cap \{\|(A^{(n)})^{-1}\|_1 \le C_M \|(M^{(n)})^{-1}\|_1\}\right) \\
&= \lim_{n\to\infty} \mathbb{P}(E_n) = 1;
\end{aligned}$$

i.e. for any given $\epsilon > 0$, there exists an $n_\epsilon \in \mathbb{N}$ such that

$$\sup_{n \ge n_\epsilon} \mathbb{P}\{\|(A^{(n)})^{-1}\|_1 \le C_M \|(M^{(n)})^{-1}\|_1\} \ge 1 - \epsilon.$$

which means that $\|(A^{(n)})^{-1}\|_1 = O_{\mathbb{P}}(\|(M^{(n)})^{-1}\|_1)$. □

A.2.3 Theorem. *Under assumptions S2-S5, if $L_n^{3+2\eta}(m_n|D_n|)^{-1} \to 0$ then $\|\hat{\Theta}_n - \Theta_n\| \xrightarrow{\mathbb{P}} 0$.*

Proof. The estimating equation $A^{(n)}\hat{\Theta}_n = B^{(n)}$ can be written as

$$A^{(n)}(\hat{\Theta}_n - \Theta_n) = (B^{(n)} - M^{(n)}\Theta_n) + (M^{(n)} - A^{(n)})\Theta_n,$$

which implies that

$$\begin{aligned}
\|\hat{\Theta}_n - \Theta_n\| &= \|(A^{(n)})^{-1}[(B^{(n)} - M^{(n)}\Theta_n) + (M^{(n)} - A^{(n)})\Theta_n]\| \\
&\le \|(A^{(n)})^{-1}\|_1 (\|B^{(n)} - M^{(n)}\Theta_n\| + \|(M^{(n)} - A^{(n)})\Theta_n\|).
\end{aligned}$$

From Lemma 6.2.1 and the triangle inequality

$$\|B^{(n)} - M^{(n)}\Theta_n\| \le \|B^{(n)} - V^{(n)}\| + \|V^{(n)} - M^{(n)}\Theta_n\|$$

it follows that $\|B^{(n)} - M^{(n)}\Theta_n\| = O_{\mathbb{P}}(L_n^{-\delta} + L_n^{1/2}(m_n|D_n|)^{-1/2})$, because $O(a_n) +$

$O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n + b_n)$. Moreover, from Lemma 6.2.1 $\left\|(M^{(n)} - A^{(n)})\Theta_n\right\| = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1/2})$ because

$$\left\|(M^{(n)} - A^{(n)})\Theta_n\right\| \leq \left\|M^{(n)} - A^{(n)}\right\|_1 \|\Theta_n\|,$$

and

$$\|\Theta_n\|^2 = \sum_{m=1}^{L_n} \theta_m^2 \leq \sum_{m=1}^{\infty} \theta_m^2 < \infty.$$

Thus, $\|B^{(n)} - M^{(n)}\Theta_n\| + \left\|(M^{(n)} - A^{(n)})\Theta_n\right\| = O_{\mathbb{P}}(L_n^{-\delta} + L_n^{3/2}(m_n|D_n|)^{-1/2})$, because $O_{\mathbb{P}}(a_n) + O_{\mathbb{P}}(b_n) = O_{\mathbb{P}}(a_n)$, if $a_n \geq b_n$. It then follows from Lemma A.2.2 and the fact that $O_{\mathbb{P}}(a_n)o_{\mathbb{P}}(b_n) = o_{\mathbb{P}}(a_n b_n)$ that $\|\hat{\Theta}_n - \Theta_n\| = O_{\mathbb{P}}(L_n^{\eta-\delta} + L_n^{3/2+\eta}(m_n|D_n|)^{-1/2})$. Therefore $\|\hat{\Theta}_n - \Theta_n\| \xrightarrow{\mathbb{P}} 0$ if $L_n^{3/2+\eta}(m_n|D_n|)^{-1/2} \to 0$. □

For each $r \in [0, R]$, let

$$\hat{g}_n(r) = \sum_{l=1}^{L_n} \hat{\theta}_l \phi_l(r).$$

and define the functional norm

$$\|\hat{g}_n - g_n\|_2 = \left(\int_0^R \big(\hat{g}_n(r) - g_n(r)\big)^2 \mathrm{d}r\right)^{1/2}.$$

Then,

$$\big(\hat{g}_n(r) - g_n(r)\big)^2 = \left(\sum_{l=1}^{L_n} (\hat{\theta}_l - \theta_l)\phi_l(r)\right)^2 \leq \left(\sum_{l=1}^{L_n} (\hat{\theta}_l - \theta_l)^2\right)\left(\sum_{l=1}^{L_n} \phi_l^2(r)\right)$$

and

$$\int_0^R \big(\hat{g}_n(r) - g_n(r)\big)^2 \mathrm{d}r \leq \|\hat{\Theta}_n - \Theta_n\|^2 \sum_{l=1}^{L_n} \int_0^R \phi_l^2(r)\mathrm{d}r = L_n \|\hat{\Theta}_n - \Theta_n\|^2.$$

and we can conclude that $\|\hat{g}_n - g_n\|_2 = O_{\mathbb{P}}(L_n^{1/2+\eta-\delta} + L_n^{2+\eta}(m_n|D_n|)^{-1/2})$ and state the following corollary.

A.2.4 Corollary. *Under conditions S2-S5, if $\delta > \eta + 1/2$ and $L_n^{2+\eta}(m_n|D_n|)^{-1/2} \to 0$, then $\|\hat{g}_n - g_n\|_2 \xrightarrow{\mathbb{P}} 0$.*

The above corollary implies that $\|\hat{g}_n - g\|_2 \xrightarrow{\mathbb{P}} 0$ under the same conditions, because $\hat{g}_n(r) - g(r) = \hat{g}_n(r) - g_n(r) - \zeta_n(r)$, $\|\zeta_n\|_2 \to 0$ and

$$\|\hat{g}_n - g\|_2 \leq \|\hat{g}_n - g_n\|_2 + \|\zeta_n\|_2.$$

A.3 Consistency of the Log Orthogonal Series Estimator

The following Lemma quantifies the distance between $g(r)$ and $\tilde{g}_L(r;\boldsymbol{\theta}^*)$.

A.3.1 Lemma. *Under conditions E1-E2 and E4-E6, we have that as $L \to \infty$,*

$$\begin{aligned} \|\boldsymbol{\theta}_0 - \boldsymbol{\theta}^*\| &= O(L^{\nu_0-\nu_1}), && \text{(A.18)} \\ \sup_{0<r<R} |g(r) - \tilde{g}_L(r;\boldsymbol{\theta}^*)| &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) = o(1), && \text{(A.19)} \\ \sup_{0<r<R} |\tilde{g}_L(r;\boldsymbol{\theta}^*)| &= O(1), && \text{(A.20)} \end{aligned}$$

where ν_0, ν_1, τ_1 and ν_2 are defined in conditions E4 and E6.

Proof. It is straightforward to see that by definition, $\boldsymbol{\theta}^*$ is the solution to (6.16), which also maximizes the following loss function

$$\ell(\boldsymbol{\theta}) = \int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)\left\{g(|u-v|)\boldsymbol{\theta}^T\boldsymbol{\phi}_L(|u-v|) - \exp\left[\boldsymbol{\theta}^T\boldsymbol{\phi}_L(|u-v|)\right]\right\}\mathrm{d}u\mathrm{d}v.$$

Define functions $\Delta_n(r) = \boldsymbol{\nu}_n^T\boldsymbol{\phi}_L(r)$ such that $\|\boldsymbol{\nu}_n\| = L^{-\nu_1+\nu_0}$. Let $f_0(r) = \boldsymbol{\theta}_0^T\boldsymbol{\phi}_L(r)$, $0 < r < R$, and define the function of a scalar z as follows

$$\begin{aligned} h(z) = &\int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|) \\ &\times \left\{g(|u-v|)\left[f_0(|u-v|) + z\Delta_n(|u-v|)\right] - \exp\left[f_0(|u-v|) + z\Delta_n(|u-v|)\right]\right\}\mathrm{d}u\mathrm{d}v. \end{aligned}$$

We shall show that for any $z_0 > 0$, $h'(z_0) < 0$ and $h'(-z_0) > 0$. This implies that the maximizer of $\ell(\boldsymbol{\theta})$ satisfies that $\boldsymbol{\theta}^{*T}\boldsymbol{\phi}_L(r)$ is between $f_0(r) \pm z_0\Delta_n(r)$, using the fact that $\ell(\boldsymbol{\theta})$ is a concave function of $\boldsymbol{\theta}$. Some straightforward calculus gives that

$$
\begin{aligned}
h'(z) &= \int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\Delta_n(|u-v|)\left\{1-\exp\left[-\tilde{\zeta}_L(|u-v|;\boldsymbol{\theta}_0)+z\Delta_n(|u-v|)\right]\right\}\mathrm{d}u\mathrm{d}v \\
&= \int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\Delta_n(|u-v|)\left\{1-\exp\left[-\tilde{\zeta}_L(|u-v|;\boldsymbol{\theta}_0)\right]\right\}\mathrm{d}u\mathrm{d}v \\
&+ \int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\Delta_n(|u-v|)\exp\left[-\tilde{\zeta}_L(|u-v|;\boldsymbol{\theta}_0)\right]\left\{1-\exp\left[z\Delta_n(|u-v|)\right]\right\}\mathrm{d}u\mathrm{d}v \\
&= \int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\Delta_n(|u-v|)\tilde{\zeta}_L(|u-v|;\boldsymbol{\theta}_0)\left[1+o(1)\right]\mathrm{d}u\mathrm{d}v \\
&+ \int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\Delta_n(|u-v|)\left[-z\Delta_n(|u-v|)\right]\left[1+o(1)\right]\mathrm{d}u\mathrm{d}v,
\end{aligned}
$$

where $\tilde{\zeta}_L(r;\boldsymbol{\theta}_0)$ is the approximation error defined in (6.12). The last equation follows from the Taylor expansion $1-e^x = -x + e^{x^*}x^{*2}$, $|x^*| < |x|$ and conditions E4, E6, which ensure that $\sup_{0<r<R}|\Delta_n(r)| \le \|\boldsymbol{\nu}_n\| \sup_{0<r<R}\|\boldsymbol{\phi}_L(r)\| = O(L^{\nu_0+\nu_2-\nu_1}) = o(1)$ as $L \to \infty$.

When $z > 0$, using conditions E1,E2,E4, E5 and the Hölder's inequality, we can derive that

$$
\begin{aligned}
h'(z) &\le O(1)\int_0^R w_o(s)g(s)\Delta_n(s)\tilde{\zeta}_L(s;\boldsymbol{\theta}_0)\mathrm{d}s \\
&\quad - z\left[1+o(1)\right]\boldsymbol{\nu}_n^T \underbrace{\left\{\int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\boldsymbol{\phi}_L(|u-v|)\boldsymbol{\phi}_L^T(|u-v|)\mathrm{d}u\mathrm{d}v\right\}}_{Q_L \text{ defined in (6.17)}}\boldsymbol{\nu}_n \\
&\le O(1)\sqrt{\int_0^R w_o(s)\Delta_n^2(s)\mathrm{d}s}\sqrt{\int_0^R w_o(s)\tilde{\zeta}_L^2(s;\boldsymbol{\theta}_0)\mathrm{d}s} - z\left[1+o(1)\right]\|\boldsymbol{\nu}_n\|^2 \times \eta_{\min}\left[Q_L\right] \\
&= O\left(L^{-\nu_1}\right)\|\boldsymbol{\nu}_n\| - z\left[1+o(1)\right]\|\boldsymbol{\nu}_n\|^2 \times \eta_{\min}\left[Q_L\right].
\end{aligned}
$$

Finally, under condition E6, $\|\boldsymbol{\nu}_n\| = L^{-\nu_1+\nu_0}$ is sufficient to ensure that there exists a $z_0 > 0$ such that $h(z_0) < 0$.

Similarly, $\|\boldsymbol{\nu}_n\| = L^{-\nu_1+\nu_0}$ is sufficient to ensure that

$$
\begin{aligned}
h'(-z_0) &\ge -O(1)\int_0^R w_o(s)g(s)|\Delta_n(s)||\tilde{\zeta}_L(s;\boldsymbol{\theta}_0)|\mathrm{d}s \\
&\quad + z_0\left[1+o(1)\right]\boldsymbol{\nu}_n^T \underbrace{\left\{\int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\boldsymbol{\phi}_L(|u-v|)\boldsymbol{\phi}_L^T(|u-v|)\mathrm{d}u\mathrm{d}v\right\}}_{Q_L}\boldsymbol{\nu}_n \\
&\ge -O(1)\sqrt{\int_0^R w_o(s)\Delta_n^2(s)\mathrm{d}s}\sqrt{\int_0^R w_o(s)\tilde{\zeta}_L^2(s;\boldsymbol{\theta}_0)\mathrm{d}s} + z_0\left[1+o(1)\right]\|\boldsymbol{\nu}_n\|^2 \times \eta_{\min}\left[Q_L\right] \\
&= -O\left(L^{-\nu_1}\right)\|\boldsymbol{\nu}_n\| + z_0\left[1+o(1)\right]\|\boldsymbol{\nu}_n\|^2 \times \eta_{\min}\left[Q_L\right] > 0.
\end{aligned}
$$

Therefore, we have shown that $\boldsymbol{\theta}^{*T}\boldsymbol{\phi}_L(r)$ is between $\boldsymbol{\theta}_0^T\boldsymbol{\phi}_L(s) \pm z_0\boldsymbol{\nu}_n^T\boldsymbol{\phi}_L(s)$, and hence

$$\|\boldsymbol{\theta}^* - \boldsymbol{\theta}_0\|^2 = \int_0^R w_o(s)\left[\boldsymbol{\theta}^{*T}\boldsymbol{\phi}_L(s) - \boldsymbol{\theta}_0^T\boldsymbol{\phi}_L(s)\right]^2 \mathrm{d}s \le z_0^2 \int_0^R w_o(s)\left[\boldsymbol{\nu}_n^T\boldsymbol{\phi}_L(s)\right]^2 \mathrm{d}s = z_0^2\|\boldsymbol{\nu}_n\|^2,$$

which completes the proof of equation (A.18).

Furthermore, to show (A.19), note that, under condition E6(b),

$$|g(r) - \tilde{g}_L(r;\boldsymbol{\theta}_0)| = g(r)\left|1 - \exp\left[-\sum_{l=L+1}^{\infty}\theta_{0,l}\phi_l(r)\right]\right| = g(r)O(L^{-\nu_1+\tau_1}) = O(L^{-\nu_1+\tau_1}).$$

Under condition E2(a), the above result also implies that $\sup_{0<r<R}\tilde{g}_L(r;\boldsymbol{\theta}_0) = O(1)$. Then, we have that

$$\begin{aligned}
|g(r) - \tilde{g}_L(r;\boldsymbol{\theta}^*)| &\le |g(r) - \tilde{g}_L(r;\boldsymbol{\theta}_0)| + |\tilde{g}_L(r;\boldsymbol{\theta}_0) - \tilde{g}_L(r;\boldsymbol{\theta}^*)| \\
&= O(L^{-\nu_1+\tau_1}) + \tilde{g}_L(r;\boldsymbol{\theta}_0)\left|1 - \exp\left[\sum_{l=1}^{L}(\theta_l^* - \theta_{0,l})\phi_l(r)\right]\right| \\
&= O(L^{-\nu_1+\tau_1}) + \tilde{g}_L(r;\boldsymbol{\theta}_0)O\left(\sup_{0<r\le R}\|\boldsymbol{\phi}_L(r)\|\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|\right) \\
&= O(L^{-\nu_1+\tau_1}) + O(L^{\nu_0-\nu_1+\nu_2}) \\
&= o(1),
\end{aligned}$$

where the last equality follows from condition E6, where we have assumed that $0 \le 2\nu_0 < \nu_1 - \nu_2$. Equation (A.19) immediately follows by noting that all the upper bounds do not depend on r. Equation (A.20) is trivial by combining equation (A.19) and condition E2(a). □

A.3.2 Lemma. *Under conditions E1-E6, we have that as* $L \to \infty$ *and* $L^{4\nu_0+2\nu_2}/m|D_n| \to 0$,

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = O_p\left(\frac{L^{\nu_0}}{\sqrt{m|D_n|}}\right) \quad \text{(A.21)}$$

$$\sup_{0<r<R}\left|g(r) - \tilde{g}_L(r;\widehat{\boldsymbol{\theta}})\right| = O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) + O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right), \quad \text{(A.22)}$$

where $\boldsymbol{\theta}^*$ *is defined in equation* (6.16).

Proof. It is straightforward to see that solving estimating equation (6.13) is equivalent to maximizing the following loss function with respect to $\boldsymbol{\theta}$

$$\tilde{L}(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{u,v\in X_i}^{\neq}\sum w_R(|u-v|)\log\left[\tilde{g}_L(|u-v|;\boldsymbol{\theta})\right] - \frac{1}{m(m-1)}\sum_{i\neq j}\sum \sum_{u\in X_i, v\in X_j}\sum w_R(|u-v|)\tilde{g}_L(|u-v|;\boldsymbol{\theta}), \tag{A.23}$$

whose Hessian matrix is negative definitive as follows

$$\frac{\partial^2\tilde{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T} = -\sum_{i\neq j}\sum \sum_{u\in X_i, v\in X_j}\sum \frac{w_R(|u-v|)}{m(m-1)}\tilde{g}_L(|u-v|;\boldsymbol{\theta})\boldsymbol{\phi}_L(|u-v|)\boldsymbol{\phi}_L^T(|u-v|),$$

which implies that $\tilde{L}(\boldsymbol{\theta})$ is a concave function of $\boldsymbol{\theta}$.

We shall first show that for any given ε there exists a large constant C such that, for large m or/and n,

$$P\left\{\sup_{\|\boldsymbol{\delta}_L\|=1}\tilde{L}(\boldsymbol{\theta}^* + CJ_{m,n}^{-1/2}\boldsymbol{\delta}_L) < \tilde{L}(\boldsymbol{\theta}^*)\right\} \geq 1-\varepsilon. \tag{A.24}$$

Inequality (A.24) implies that with probability tending to 1 there is a local maximum, denoted as $\widehat{\boldsymbol{\theta}}$, in the ball $\{\boldsymbol{\theta}^* + J_{m,n}^{-1/2}C\boldsymbol{\delta}_L : \|\boldsymbol{\delta}_L\| = 1\}$ such that $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|^2 = O_p(J_{m,n}^{-1})$.

To show (A.24), define a function of z as

$$H_{m,n}(z) = -\tilde{L}(\boldsymbol{\theta}^* + zJ_{m,n}^{-1/2}\boldsymbol{\delta}_L). \tag{A.25}$$

Consider the Taylor expansion of the loss function $\tilde{L}(\boldsymbol{\theta})$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^*$

$$\begin{aligned}\tilde{L}(\boldsymbol{\theta}^* + CJ_{m,n}^{-1/2}\boldsymbol{\delta}_L) - \tilde{L}(\boldsymbol{\theta}^*) &= CJ_{m,n}^{-1/2}\left[\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*)\right]^T\boldsymbol{\delta}_L + \frac{1}{2}C^2J_{m,n}^{-1}\boldsymbol{\delta}_L^T\left[\frac{\partial^2\tilde{L}(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}^*}\right]\boldsymbol{\delta}_L \\ &= -C\left[H'_{m,n}(0) + \frac{C}{2}H''_{m,n}(z_0)\right],\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}_m^* = \boldsymbol{\theta}^* + z_0J_{m,n}^{-1/2}\boldsymbol{\delta}_L$ for some $0 < z_0 < C$. By definition, $H_{m,n}(z)$ is a convex function of z since $H''_{m,n}(z) \geq 0$ for any constant z. Therefore, to find a large enough C so that (A.24) holds, it suffices to show that

$$H'_{m,n}(0) = O_p\left[H''_{m,n}(z_0)\right], \text{ for any finite } z_0. \tag{A.26}$$

By the definition of $H_{m,n}(\cdot)$ in (A.25), it is straightforward to show that

$$H'_{m,n}(0) = -\frac{J_{m,n}^{-1/2}}{m} \sum_{i=1}^{m} \sum\sum_{u,v \in X_i}^{\neq} w_R(|u-v|) \boldsymbol{\phi}_L(|u-v|)^T \boldsymbol{\delta}_L \\ + \frac{J_{m,n}^{-1/2}}{m} \sum\sum_{i \neq j} \sum\sum_{u \in X_i, v \in X_j} \frac{w_R(|u-v|)}{m-1} \tilde{g}_L(|u-v|; \boldsymbol{\theta}^*) \boldsymbol{\phi}_L(|u-v|)^T \boldsymbol{\delta}_L, \tag{A.27}$$

$$H''_{m,n}(z_0) = \frac{J_{m,n}^{-1}}{m} \sum\sum_{i \neq j} \sum\sum_{u \in X_i, v \in X_j} \frac{w_R(|u-v|)}{m-1} \tilde{g}_L(|u-v|; \tilde{\boldsymbol{\theta}}_m^*) \left[\boldsymbol{\phi}_L(|u-v|)^T \boldsymbol{\delta}_L\right]^2 \tag{A.28}$$

where $\tilde{\boldsymbol{\theta}}_m^* = \boldsymbol{\theta}^* + z_0 J_{m,n}^{-1/2} \boldsymbol{\delta}_L$.

We first investigate $H'_{m,n}(0)$. By the definition of $\boldsymbol{\theta}^*$ in (6.16), we have that

$\mathbb{E}\left[H'_{m,n}(0)\right] = 0$. Furthermore, the variance can be shown as

$$
\begin{aligned}
mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right] = &\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)\big[g^{(4)}(u_1,v_1,u_2,v_2) \\
&\quad - g(|u_1-v_1|)g(|u_2-v_2|)\big] \times \left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right] \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&- 4\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)[g^{(3)}(u_1,v_1,u_2) - g(|u_1-v_1|)] \\
&\quad \times \tilde{g}_L(|u_2-v_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right] \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ \frac{2}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)\tilde{g}_L(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_L(|u_2-v_2|;\boldsymbol{\theta}^*) \\
&\quad \times [g(|u_1-u_2|)g(|v_1-v_2|)-1]\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right] \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ \frac{4(m-2)}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)[g(|u_1-u_2|)-1] \\
&\quad \times \tilde{g}_L(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_L(|u_2-v_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right] du_1du_2dv_1dv_2 \\
&+ 4\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_R(|u_1-v_1|)w_R(|u_1-u_2|)g^{(3)}(u_1,v_1,u_2) \\
&\quad \times \left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right] \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&- 4\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_R(|u_1-v_1|)w_R(|u_1-u_2|)[2g(|u_1-v_1|) - \tilde{g}_L(|u_1-v_1|;\boldsymbol{\theta}^*)] \\
&\quad \times \tilde{g}_L(|u_1-u_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right] \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ \frac{4}{m-1}\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_R(|u_1-v_1|)w_R(|u_1-u_2|)\tilde{g}_L(|u_1-v_1|;\boldsymbol{\theta}^*) \\
&\quad \times \tilde{g}_L(|u_1-u_2|;\boldsymbol{\theta}^*)[g(|v_1-u_2|)-1]\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right] \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ 2\int_{D_n^2} \lambda(u_1)\lambda(v_1)\left[w_R(|u_1-v_1|)\right]^2 g(|u_1-v_1|)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1\mathrm{d}v_1 \\
&+ \frac{2}{m-1}\int_{D_n^2} \lambda(u_1)\lambda(v_1)w_R^2(|u_1-v_1|)\tilde{g}_L^2(|u_1-v_1|;\boldsymbol{\theta}^*)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1\mathrm{d}v_1
\end{aligned}
$$

Therefore, under conditions E1, E2(a)-(b), E4 and equations (A.19)-(A.20), we can further simplify $mL^2J_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right]$ as follows

$$
\begin{aligned}
mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right] &= O(1)\int_{D_n^4} w_R(|u_1-v_1|)w_R(|u_2-v_2|)\big[g^{(4)}(u_1,v_1,u_2,v_2) \\
&\quad - g(|u_1-v_1|)g(|u_2-v_2|)\big] \times \left|\boldsymbol{\phi}_L^T(|u_1-v_1|)\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L^T(|u_2-v_2|)\boldsymbol{\delta}_L\right| \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ O(1)\int_{D_n^4} w_R(|u_1-v_1|)w_R(|u_2-v_2|)[g^{(3)}(u_1,v_1,u_2) - g(|u_1-v_1|)] \\
&\quad \times \left|\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right| \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ O(m^{-1})\int_{D_n^4} w_R(|u_1-v_1|)w_R(|u_2-v_2|)[g(|u_1-u_2|)g(|v_1-v_2|) - 1] \\
&\quad \times \left|\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right| \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ O(1)\int_{D_n^4} w_R(|u_1-v_1|)w_R(|u_2-v_2|)[g(|u_1-u_2|) - 1] \\
&\quad \times \left|\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right| du_1du_2dv_1dv_2 \\
&+ O(1)\int_{D_n^3} w_R(|u_1-v_1|)w_R(|u_1-u_2|) \left|\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right| \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ O(1)\int_{D_n^2} [w_R(|u_1-v_1|)]^2 \left|\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right|^2 \mathrm{d}u_1\mathrm{d}v_1 \\
&= O(1)|D_n|\int_{\mathbb{R}^3} w_R(|s|)w_R(|t|)\left|g^{(4)}(s,t+w,w) - g(|s|)g(|t|)\right| \left|\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right| \mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ O(1)|D_n|\int_{\mathbb{R}^3} w_R(|s|)w_R(|t|)|g^{(3)}(s,w) - g(|s|)| \left|\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right| \mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ \frac{1}{m}O(1)|D_n|\int_{\mathbb{R}^3} w_R(|s|)w_R(|t|)|g(|w|)g(|t-s+w|) - 1| \left|\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right| \mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ O(1)|D_n|\int_{\mathbb{R}^3} w_R(|s|)w_R(|t|)|g(|w|) - 1| \left|\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right| \mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ O(1)|D_n|\int_{\mathbb{R}^2} w_R(|s|)w_R(|t|) \left|\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right| \mathrm{d}s\mathrm{d}t \\
&+ O(1)|D_n|\int_{\mathbb{R}} [w_R(|s|)]^2 \left|\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right|^2 \mathrm{d}s \\
&= O(1)|D_n|\int_{\mathbb{R}^2} w_R(|s|)w_R(|t|) \left|\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right| \left|\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right| \mathrm{d}s\mathrm{d}t + O(1)|D_n|\int_{\mathbb{R}} \left[w_R(|s|)\boldsymbol{\phi}_L^T(|s|)\right]^2 \mathrm{d}s,
\end{aligned}
$$

where the last equality follows from conditions E2-E3 and equations (A.19)-(A.20). Recall that by definition of orthogonal basis, we have that $\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s = \boldsymbol{\delta}_L^T\boldsymbol{\delta}_L = 1$.

Finally, by condition E5, we have that

$$
\begin{aligned}
mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right] &= O(1)|D_n|^{-1}\left\{\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]\mathrm{d}s\right\}^2 \\
&\quad + O(1)|D_n|^{-1}\int_0^R w_o^2(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^2\mathrm{d}s \\
&\le O(1)|D_n|^{-1}\left\{\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^2\mathrm{d}s\right\}\int_0^R w_o(s)\mathrm{d}s \\
&\quad + O(1)|D_n|^{-1}\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^2\mathrm{d}s = O\left(\frac{1}{|D_n|}\right),
\end{aligned}
$$

where the second lat inequality follows from the Hölder's inequality.

Combing with the fact that $\mathbb{E}\left[H'_{m,n}(0)\right]=0$, we have that

$$
H'_{m,n}(0) = O_p\left(\frac{1}{\sqrt{m|D_n|J_{m,n}}}\right). \tag{A.29}
$$

Now we proceed to study $H''_{m,n}(z_0)$. Some tedious algebra gives that

$$
\begin{aligned}
J_{m,n}^2 m(m-1)\mathrm{Var}\left[H''_{m,n}(z_0)\right] &= 2\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)\tilde{g}_L(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*) \\
&\quad\times\tilde{g}_L(|u_2-v_2|;\tilde{\boldsymbol{\theta}}^*)[g(|u_1-u_2|)g(|v_1-v_2|)-1]\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right]^2 \\
&\qquad\qquad \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ 4\int_{D_n^3}\lambda(u_1)\lambda(v_1)\lambda(u_2)w_R(|u_1-v_1|)w_R(|u_1-u_2|)\tilde{g}_L(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*) \\
&\quad\times\tilde{g}_L(|u_1-u_2|;\tilde{\boldsymbol{\theta}}^*)g(|v_1-u_2|)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right]^2\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ 2\int_{D_n^2}\lambda(u_1)\lambda(v_1)w_R^2(|u_1-v_1|)\tilde{g}_L^2(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)g(|u_1-v_1|)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^4\mathrm{d}u_1\mathrm{d}v_1 \\
&+ 4(m-2)\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)[g(|u_1-u_2|)-1] \\
&\quad\times\tilde{g}_L(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)\tilde{g}_L(|u_2-v_2|;\tilde{\boldsymbol{\theta}}^*)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right]^2\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v_1\mathrm{d}v_2 \\
&+ 4(m-2)\int_{D_n^3}\lambda(u_1)\lambda(v_1)\lambda(u_2)w_R(|u_1-v_1|)w_R(|u_1-u_2|)\tilde{g}_L(|u_1-v_1|;\tilde{\boldsymbol{\theta}}^*)\tilde{g}_L(|u_1-u_2|;\tilde{\boldsymbol{\theta}}^*) \\
&\quad\times\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right]^2\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2.
\end{aligned}
$$

Using equations (A.19)-(A.20), under conditions E1, E2(a)-(b), E4, we can further simplify $J_{m,n}^2 m(m-1)\mathrm{Var}\left[H''_{m,n}(z_0)\right]$ as follows

$$
\begin{aligned}
J_{m,n}^2 m(m-1)\mathrm{Var}\left[H''_{m,n}(z_0)\right] &= O(1)\int_{D_n^4} w_R(|u_1-v_1|)w_R(|u_2-v_2|)|g(|u_1-u_2|)g(|v_1-v_2|)-1| \\
&\qquad \times \left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ O(1)\int_{D_n^3} w_R(|u_1-v_1|)w_R(|u_1-u_2|)\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ O(1)\int_{D_n^2} w_R^2(|u_1-v_1|)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^4 \mathrm{d}u_1\mathrm{d}v_1 \\
&+ mO(1)\int_{D_n^4} w_R(|u_1-v_1|)w_R(|u_2-v_2|)|g(|u_1-u_2|)-1| \\
&\qquad \times \left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right]^2 du_1du_2dv_1dv_2 \\
&+ mO(1)\int_{D_n^3} w_R(|u_1-v_1|)w_R(|u_1-u_2|)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&= O(1)|D_n|\int_{\mathbb{R}^3} w_R(|s|)w_R(|t|)|g(|w|)g(|t-s+w|)-1|\left[\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ O(1)|D_n|\int_{\mathbb{R}^2} w_R(|s|)w_R(|t|)\left[\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s\mathrm{d}t \\
&+ O(1)|D_n|\int_{\mathbb{R}} [w_R(|s|)]^2\left[\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right]^4 \mathrm{d}s \\
&+ mO(1)|D_n|\int_{\mathbb{R}^3} w_R(|s|)w_R(|t|)|g(|w|)-1|\left[\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ mO(1)|D_n|\int_{\mathbb{R}^2} w_R(|s|)w_R(|t|)\left[\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s\mathrm{d}t \\
&= O(1)|D_n|\int_{\mathbb{R}} [w_R(|s|)]^2\left[\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right]^4 \mathrm{d}s \\
&+ mO(1)|D_n|\int_{\mathbb{R}^2} w_R(|s|)w_R(|t|)\left[\boldsymbol{\phi}_L^T(|s|)\boldsymbol{\delta}_L\right]^2\left[\boldsymbol{\phi}_L^T(|t|)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s\mathrm{d}t
\end{aligned}
$$

Recall that by definition of orthogonal basis, we have that $\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s = \boldsymbol{\delta}_L^T\boldsymbol{\delta}_L = 1$. Then, by the condition E5, we have that

$$
\begin{aligned}
J_{m,n}^2 m(m-1)\mathrm{Var}\left[H''_{m,n}(z_0)\right] &= O(1)|D_n|^{-1}\int_0^R w_o^2(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^4 \mathrm{d}s \\
&+ O(1)m|D_n|^{-1}\left\{\int_{\mathbb{R}^2} w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s\right\}^2 \\
&\le O(1)|D_n|^{-1}\sup_{0<r\le R}\|\boldsymbol{\phi}_L(r)\|^2 \times \int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}_L\right]^2 \mathrm{d}s + O(1)m|D_n|^{-1} \\
&= O(L^{2\nu_2}|D_n|^{-1}) + O(m|D_n|^{-1}),
\end{aligned}
$$

which immediately implies that

$$\text{Var}\left[H''_{m,n}(z_0)\right] = O\left(\frac{L^{2\nu_2}}{J^2_{m,n} m^2 |D_n|}\right) + O\left(\frac{1}{J^2_{m,n} m |D_n|}\right). \tag{A.30}$$

On the other hand, we have that

$$\mathbb{E}\left[H''_m(z_0)\right] = J^{-1}_{m,n} \int_{D_n^2} \lambda(u_1)\lambda(v_1) w_R(|u_1 - v_1|) \tilde{g}_L(|u_1 - v_1|; \tilde{\boldsymbol{\theta}}^*_m) \left[\boldsymbol{\phi}_L(|u_1 - v_1|)^T \boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1 \mathrm{d}v_1.$$

Recall that the smallest eigenvalue satisfy the condition

$$\eta_{\min}\left[\mathbf{Q}_L\right] = \inf_{\|\boldsymbol{\eta}\|^2 = 1} \boldsymbol{\eta}^T \mathbf{Q}_L \boldsymbol{\eta}.$$

Using definition of $\mathbf{Q}_L$ in condition E6, we have that

$$\begin{aligned}
\mathbb{E}\left[J_{m,n} H''_m(z_0)\right] - \eta_{\min}\left[\mathbf{Q}_{n,h}\right] &\geq \int_{D_n^2} \lambda(u_1)\lambda(v_1) w_R(|u_1 - v_1|) \\
&\quad \times \left[\tilde{g}_L(|u_1 - v_1|; \tilde{\boldsymbol{\theta}}^*_m) - g(|u_1 - v_1|)\right] \left[\boldsymbol{\phi}_L(|u_1 - v_1|)^T \boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1 \mathrm{d}v_1 \\
&= O(1)|D_n| \int_{\mathbb{R}} w_R(|s|) \left|\tilde{g}_L(|s|; \tilde{\boldsymbol{\theta}}^*_m) - g(|s|)\right| \left[\boldsymbol{\phi}_L(|s|)^T \boldsymbol{\delta}_L\right]^2 \mathrm{d}s \\
&= O(1) \int_0^R w_o(s) \left|\tilde{g}_L(s; \tilde{\boldsymbol{\theta}}^*_m) - g(s)\right| \left[\boldsymbol{\phi}_L(s)^T \boldsymbol{\delta}_L\right]^2 \mathrm{d}s \\
&= O(1) \sup_{0 < r \leq R} \|\boldsymbol{\phi}_L(r)\| \int_0^R w_o(s) \left|\tilde{g}_L(s; \tilde{\boldsymbol{\theta}}^*_m) - g(s)\right| \left|\boldsymbol{\phi}_L(s)^T \boldsymbol{\delta}_L\right| \mathrm{d}s.
\end{aligned} \tag{A.31}$$

Note that by the definition of $\tilde{\boldsymbol{\theta}}^*_m = \boldsymbol{\theta}^* + z_0 J^{-1/2}_{m,n} \boldsymbol{\delta}_L$, equation (A.18) of Lemma A.3.1, it is straightforward to show that, under conditions E4 and provided the $\sup_{0<r\leq R} \left|J^{-1/2}_{m,n} \boldsymbol{\delta}^T_L \boldsymbol{\phi}_L(r)\right| = O(1)$,

$$\begin{aligned}
\left|\tilde{g}_L(r; \tilde{\boldsymbol{\theta}}^*_m) - g(r)\right| &= g(r) \left|1 - \exp\left[(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)^T \boldsymbol{\phi}_L(r) + z_0 J^{-1/2}_{m,n} \boldsymbol{\delta}^T_L \boldsymbol{\phi}_L(r) - \tilde{\zeta}_L(r; \boldsymbol{\theta}_0)\right]\right| \\
&= O(1) \left\{\left|(\boldsymbol{\theta}^* - \boldsymbol{\theta}_0)^T \boldsymbol{\phi}_L(r)\right| + \left|z_0 J^{-1/2}_{m,n} \boldsymbol{\delta}^T_L \boldsymbol{\phi}_L(r)\right| + \left|\tilde{\zeta}_L(r; \boldsymbol{\theta}_0)\right|\right\},
\end{aligned}$$

which further gives that, under condition E4 and using (A.18) in Lemma A.3.1,

$$\begin{aligned}
\int_0^R w_o(s)\left|\tilde{g}_L(s;\tilde{\boldsymbol{\theta}}_m^*)-g(s)\right|\left|\boldsymbol{\phi}_L(|s|)^T\boldsymbol{\delta}_L\right|\mathrm{d}s &= O(1)\int_0^R w_o(s)\left|(\boldsymbol{\theta}^*-\boldsymbol{\theta}_0)^T\boldsymbol{\phi}_L(r)\right|\left|\boldsymbol{\phi}_L(|s|)^T\boldsymbol{\delta}_L\right|\mathrm{d}s\\
&+O(1)\int_0^R w_o(s)\left|z_0J_{m,n}^{-1/2}\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(r)\right|\left|\boldsymbol{\phi}_L(|s|)^T\boldsymbol{\delta}_L\right|\mathrm{d}s\\
&+O(1)\int_0^R w_o(s)\left|\tilde{\zeta}_L(r;\boldsymbol{\theta}_0)\right|\left|\boldsymbol{\phi}_L(|s|)^T\boldsymbol{\delta}_L\right|\mathrm{d}s\\
&=O(1)\left[\int_0^R w_o(s)\left|(\boldsymbol{\theta}^*-\boldsymbol{\theta}_0)^T\boldsymbol{\phi}_L(r)\right|^2\mathrm{d}s\right]^{1/2}+O(1)z_0J_{m,n}^{-1/2}+O(1)\left[\int_0^R w_o(s)\tilde{\zeta}_L^2(r;\boldsymbol{\theta}_0)\mathrm{d}s\right]^{1/2}\\
&=O(\|\boldsymbol{\theta}^*-\boldsymbol{\theta}_0\|+J_{m,n}^{-1/2}+L^{-\nu_1})\\
&=O(L^{\nu_0-\nu_1}+J_{m,n}^{-1/2}+L^{-\nu_1})=O(L^{\nu_0-\nu_1}+J_{m,n}^{-1/2}).
\end{aligned}$$

Combining the above result, equation (A.31) and condition E2, we have that if $J_{m,n}^{-1/2}L^{\nu_2}=O(1)$,

$$\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]-\eta_{\min}\left[\mathbf{Q}_{n,h}\right]=O(L^{\nu_2+\nu_0-\nu_1}+J_{m,n}^{-1/2}L^{\nu_2})$$

By condition E2(a) and E6, the above equation gives that

$$L^{\nu_0}\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]\geq c_0+O(L^{\nu_2+2\nu_0-\nu_1}+J_{m,n}^{-1/2}L^{\nu_0+\nu_2}). \tag{A.32}$$

Hence for the constant $c=c_0$, we have that

$$\begin{aligned}
P\left(L^{\nu_0}J_{m,n}H_m''(z_0)<c/2\right)&=P\left\{J_{m,n}H_m''(z_0)-\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]<c/2L^{-\nu_0}-\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]\right\}\\
&\leq P\left\{|J_{m,n}H_m''(z_0)-\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]|>|c/2L^{-\nu_0}-\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]|\right\}\\
&\quad\times I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]>c/2L^{-\nu_0}\right\}+I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]\leq c/2L^{-\nu_0}\right\}\\
&\leq\frac{\mathrm{Var}\left[J_{m,n}H_{m,n}''(z_0)\right]}{|c/2L^{-\nu_0}-\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]|^2}I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]>c/2L^{-\nu_0}\right\}+I\left\{\mathbb{E}\left[J_{m,n}H_m''(z_0)\right]\leq c/2L^{-\nu_0}\right\}\\
&=O\left(\frac{L^{2\nu_2+2\nu_0}}{m^2|D_n|}\right)+O\left(\frac{L^{2\nu_0}}{m|D_n|}\right)+o(1),
\end{aligned}$$

where the last equality follows from equations (A.30) and (A.32) when $J_{m,n}\to\infty$ and $L^{2\nu_0+2\nu_2}/J_{m,n}\to0$. Therefore, as long as $\frac{L^{2\nu_0+2\nu_2}}{m^2|D_n|}+\frac{L^{2\nu_0}}{m|D_n|}\to0$, $J_{m,n}\to\infty$ and $L^{2\nu_0+2\nu_2}/J_{m,n}\to0$, we have that

$$P\left(J_{m,n}H_m''(z_0)\geq c_0L^{-\nu_0}/2\right)\to1, \tag{A.33}$$

where c_0 is the constant defined in condition E6.

We have already shown in equation (A.29) that

$$H_{m,n}'(0)=O_p\left(\frac{1}{\sqrt{m|D_n|J_{m,n}}}\right).$$

hence as long as $\frac{J_{m,n}L^{2\nu_0}}{m|D_n|}=O(1)$, we have that $H_{m,n}'(0)=O_p(H_m''(z_0))$. In other words, for any $\frac{J_{m,n}L^{2\nu_0}}{m|D_n|} \to 0$ and $L^{2\nu_0+2\nu_2}/J_{m,n} \to 0$, (A.24) holds, which completes the proof of equation (A.21).

To show (A.22), using equation (A.19)-(A.20), we have that

$$\begin{aligned}\left|g(r) - \tilde{g}_L(r;\widehat{\boldsymbol{\theta}})\right| &\le |g(r) - \tilde{g}_L(r;\boldsymbol{\theta}^*)| + \left|\tilde{g}_L(r;\boldsymbol{\theta}^*) - \tilde{g}_L(r;\widehat{\boldsymbol{\theta}})\right| \\ &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) + \tilde{g}_L(r;\boldsymbol{\theta}^*)\left|1 - \exp\left[\sum_{l=1}^{L}(\hat{\theta}_l - \theta_l^*)\phi_l(r)\right]\right| \\ &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) + \tilde{g}_L(r;\boldsymbol{\theta}^*)O\left(\sup_{0<r\le R}\|\boldsymbol{\phi}_L(r)\|\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|\right) \\ &= O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right) + O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right) \\ &= o(1),\end{aligned}$$

where the upper bounds does not depend on r, which completes the proof. □

A.4 Consistency of the Semiparametric Local Linear Estimator

Let $X_1, \ldots, X_{m_n}$ be independent point processes on $\mathbb{R}$ with the intensity functions $\lambda_i(x)$, $i = 1, \ldots, m_n$, and the same pair correlation function $g(|u - v|)$. For each single intensity function, we assume $\lambda_i(x) = \lambda_0(x)\rho(Z_i; \beta^*)$, $i = 1, \cdots, m_n$, where $\lambda_0(x)$ is the baseline intensity function, $Z_i = (Z_{i1}, \ldots, Z_{ip})^\mathsf{T}$ are time invariant covariates associated with the point process X_i, $\beta^* = (\beta_1^*, \ldots, \beta_p^*)^\mathsf{T}$ is the vector of true coefficients corresponding to each covariate and ρ is some known nonnetative function. We assume that

B1 $D_n \subset \mathbb{R}$ is an arbitrary sequence of observation windows,

B2 m_n is an increasing sequence of natural numbers such that $m_n \to \infty$, and

B3 h_n is a decreasing sequence of positive real numbers such that $h_n \to 0$.

Let

$$\gamma_n(s) = |D_n \cap (D_n - s)| = \int_{\mathbb{R}} \mathbb{I}[y \in D_n, y + s \in D_n]\mathrm{d}y, \quad s \in \mathbb{R}.$$

Then for any fixed $s \in \mathbb{R}$, $\gamma_n(s)/|D_n| \to 1$ and hence a constant $0 < C_\gamma \le 1$ can be found such that $\gamma_n(s) \ge C_\lambda|D_n|$ for all $-R_0 \le s \le R_0$ and sufficiently large n.

For given $t \geq 0$ and $h_n > 0$, we further assume that

$$g(t) = \theta_1(r) + \theta_2(r)(t-r) + O(h_n^2), \quad r - h_n \leq t \leq r + h_n, \tag{A.34}$$

where $\theta_1(t) = g(t)$ and $\theta_2(t) = g'(t)$.

Similar to the proof of the consistency of local linear estimator we have matrix M_n and vector B_n:

$$M_{l,k}^{(n)} = \mathbb{E}\, A_{l,k}^{(n)} = \int_{(D_n)^2} \frac{f_{l,k}(|x-y|-r)}{\gamma_n(x-y)\rho(Z_i;\beta^*)\rho(Z_j;\beta^*)}\lambda_0(x)\lambda_0(y)\mathrm{d}x\mathrm{d}y, \qquad l,k = 1,2$$

and

$$V_l^{(n)} = EB_l^{(n)} = \int_{(D_n)^2} \frac{K_h(|x-y|-r)G_l(|x-y|-r)}{\gamma_n(x-y)\rho^2(Z_i;\beta^*)} g(|x-y|)\lambda(x)\lambda(y)dxdy, l = 1,2$$

where $\gamma_n(x-y) = |D \cap D - x + y|$ and $f_{l,k}(r) = K_h(r)G_l(r)G_k(r)$.

The regression parameter β can be estimated using the method in Lawless and Nadeau (1995) and Lin et al. (2000) without having to specify $\lambda_0(t)$. More specifically, β is estimated by solving the estimating equation $u_n(\beta) = 0$, where

$$u_n(\beta) = \frac{1}{m_n}\sum_{i=1}^{m_n} u_{n,i}(\beta),$$

and

$$u_{n,i}(\beta) = \sum_{x \in X_i \cap D_n} \left(\frac{\rho^{(1)}(Z_i;\beta)}{\rho(Z_i;\beta)} - \frac{\sum_{j=1}^{m_n} \rho^{(1)}(Z_j;\beta)}{\sum_{j=1}^{m_n} \rho(Z_j;\beta)} \right)$$

This equation does not dependent on time events, but relies on the number of events in each single pattern.

Then we can substitute the true value β^* with $\hat{\beta}_n$ to get the practical versions of $\hat{A}_{l,k}^{(n)}$ and $\hat{B}_l^{(n)}$

Let $A^{(n)} = [A_{l,k}^{(n)}]$, $M^{(n)} = [M_{l,k}^{(n)}]$, $\hat{A}^{(n)} = [\hat{A}_{l,k}^{(n)}]$, $B^{(n)} = (B_l^{(n)})$, $\hat{B}^{(n)} = (\hat{B}_l^{(n)})$, $V^{(n)} = (V_l^{(n)})$ and $\Theta_t = (\theta_1(r), \theta_2(r))$. Then $g(t) = (G(r-t))^{\mathsf{T}}\Theta_r$ and the above estimating equations are equivalent to

$$\sum_{m=1}^{2} \hat{A}_{l,m}^{(n)}\hat{\theta}_m(t) = \hat{B}_l^{(n)}, \quad l = 1,2.$$

or $\hat{A}^{(n)}\hat{\Theta}_t = \hat{B}^{(n)}$, where $\hat{\Theta}_t = (\hat{\theta}_1(t), \hat{\theta}_2(t))$.

We consider the Euclidean norm $\|B^{(n)}\| = \left(\sum_{l=1}^{2}\left(B_l^{(n)}\right)^2\right)^{1/2}$ for vectors and the maximum absolute column sum norm (Isaacson and Keller, 1994, p. 9)

$$\|A^{(n)}\|_1 = \max_{1\le k\le 2}\sum_{l=1}^{2}|A_{l,k}^{(n)}|$$

and the Frobenius norm

$$\left\|A^{(n)}\right\|_{\mathrm{F}} = \left(\sum_{l=1}^{2}\sum_{k=1}^{2}(A_{l,k}^{(n)})^2\right)^{1/2}$$

for matrices. It is known that $\|A^{(n)}\|_1 \le \sqrt{2}\|A^{(n)}\|_{\mathrm{F}}$ (see Golub and Van Loan, 1996, p. 56)

We assume that the following conditions hold:

F1 For all $x \in \mathbb{R}$, $0 \le \lambda_0(x) \le \lambda_{\max} < \infty$.

F2 For any $z, \beta \in \mathbb{R}^P$, $0 < \rho_{\min} < \rho(z;\beta) < \rho_{\max} < \infty$.

F3 For any fixed z, $\rho(z;\beta)$ is twice continuously differentiable as a function of β and $\|\rho^{(1)}(z;\beta)\| < C_\rho\|\beta - \beta^*\|$ if $\|\beta - \beta^*\| < \varepsilon_\rho$ for some $0 < \varepsilon_\rho, C_\rho < \infty$, where $\rho^{(1)}(z;\beta) = \mathrm{d}\rho(z;\beta)/\mathrm{d}\beta$.

F4 The constants $0 < C_I, C_g, C_{g^{(3)}}, C_{g^{(4)}} < \infty$ can be found such that

$$\int_0^\infty \left|g(r) - 1\right|\mathrm{d}r < C_I$$

and for all $x_1, x_2, x_3, x_4 \in \mathbb{R}$, $g(|x_1 - x_2|) \le C_g$, $g^{(3)}(x_1, x_2, x_3) \le C_{g^{(3)}}$ and

$$G^{(4)}(x_1, x_2, x_3) = \int_{\mathbb{R}} \left|g^{(4)}(x_1 + x_2, x_2, x_3 + z, z) - g(|x_1|)g(|x_3|)\right|\mathrm{d}z \le C_{g^{(4)}}.$$

A.4.1 Lemma. *Let* $\breve{\Sigma}_n = \mathrm{Var} u_{n,1}(\beta^*)$ *and assume that* $\breve{\Sigma}_n \to \breve{\Sigma}$*, where* $\breve{\Sigma}$ *is a positive definite matrix. Under conditions* $F1$*-*$F4$*, there exists a sequence* $\{\hat{\beta}_n\}$ *where* $\|\hat{\beta}_n - \beta^*\| = O_{\mathbb{P}}((m_n|D_n|)^{-1/2})$ *and* $u_n(\hat{\beta}_n) = 0$ *with a probability tending to 1 as* $n \to \infty$.

Proof. Define $\tilde{\Sigma}_n = \mathrm{Var} u_n(\beta^*) = m_n^{-1}\breve{\Sigma}_n$ and $\tilde{V}^{(n)} = (m_n|D_n|)^{1/2}\breve{\Sigma}_n^{1/2}$, where $\breve{\Sigma}_n^{1/2}$ is the unique positive semidefinite square root of $\breve{\Sigma}_n$. Since

$$\breve{\Sigma}_n = \breve{\Sigma} + (\breve{\Sigma}_n - \breve{\Sigma}) = \breve{\Sigma}\left[I_p + \breve{\Sigma}^{-1}(\breve{\Sigma}_n - \breve{\Sigma})\right],$$

we have

$$\breve{\Sigma}_n^{-1} = \big[I_p + \breve{\Sigma}^{-1}(\breve{\Sigma}_n - \breve{\Sigma})\big]^{-1}\breve{\Sigma}^{-1}.$$

The assumption $\breve{\Sigma}_n \to \breve{\Sigma}$ implies that for any sufficiently large n,

$$\big\|\breve{\Sigma}^{-1}(\breve{\Sigma}_n - \breve{\Sigma})\big\|_1 \le \big\|\breve{\Sigma}^{-1}\big\|_1\big\|\breve{\Sigma}_n - \breve{\Sigma}\big\|_1 < 1$$

and (Isaacson and Keller, 1994, p. 16)

$$\big\|\big[I_p + \breve{\Sigma}^{-1}(\breve{\Sigma}_n - \breve{\Sigma})\big]^{-1}\big\|_1 \le \frac{1}{1 - \big\|\breve{\Sigma}^{-1}(\breve{\Sigma}_n - \breve{\Sigma})\big\|_1},$$

which means that $\big\|\big[I_p + \breve{\Sigma}^{-1}(\breve{\Sigma}_n - \breve{\Sigma})\big]^{-1}\big\|_1$ and hence $\|\breve{\Sigma}_n^{-1}\|_1$ are bounded for all n. The Frobenius norm is a unitarily invariant matrix norm and hence the inequality (see Bhatia, 1997, Theorem IX.4.5)

$$\|\breve{\Sigma}_n^{-1/2}\|_{\mathrm{F}} \le \frac{1}{2}\big\|I_p + \breve{\Sigma}_n^{-1}\big\|_{\mathrm{F}}$$

holds. The boundedness of $\|\breve{\Sigma}_n^{-1/2}\|_{\mathrm{F}}$ follows by $\|I_p + \breve{\Sigma}_n^{-1}\|_{\mathrm{F}} \le \sqrt{p} + \|\breve{\Sigma}_n^{-1}\|_{\mathrm{F}}$ and the equivalence of matrix norms (Isaacson and Keller, 1994, Theorem 2′, p. 11). Thus

$$\big\|(\tilde{V}^{(n)})^{-1}\big\|_{\max} = (m_n|D_n|)^{-1/2}\big\|\breve{\Sigma}_n^{-1/2}\big\|_{\max} \le (m_n|D_n|)^{-1/2}\big\|\breve{\Sigma}_n^{-1/2}\big\|_{\mathrm{F}},$$

which implies that $\|(\tilde{V}^{(n)})^{-1}\|_{\max} \to 0$ as $n \to \infty$.

Let

$$\begin{aligned} J_n(\beta) &= -\frac{\mathrm{d}}{\mathrm{d}\beta^{\mathsf{T}}} u_n(\beta) \\ &= -\frac{1}{m_n}\sum_{i=1}^{m_n}\sum_{x \in X_i \cap D_n}\left\{\frac{\rho(Z_i;\beta)\rho^{(2)}(Z_i;\beta) - \rho^{(1)}(Z_i;\beta)\big[\rho^{(1)}(Z_i;\beta)\big]^{\mathsf{T}}}{\rho^2(Z_i;\beta)}\right. \\ &\quad \left. - \frac{\Big(\sum_{j=1}^{m_n}\rho(Z_j;\beta)\Big)\Big(\sum_{j=1}^{m_n}\rho^{(2)}(Z_j;\beta)\Big) - \Big(\sum_{j=1}^{m_n}\rho^{(1)}(Z_j;\beta)\Big)\Big(\sum_{j=1}^{m_n}\rho^{(1)}(Z_j;\beta)\Big)^{\mathsf{T}}}{\Big(\sum_{j=1}^{m_n}\rho(Z_j;\beta)\Big)^2}\right\} \end{aligned} \tag{A.35}$$

and

$$\tilde{I}_n = \frac{1}{|D_n|}\,\mathbb{E}\,J_n(\beta^*) = \frac{1}{m_n|D_n|}\int_{D_n}\lambda_0(x)\mathrm{d}x\Bigg\{\sum_{i=1}^{m_n}\frac{\rho^{(1)}(Z_i;\beta^*)\big[\rho^{(1)}(Z_i;\beta^*)\big]^{\mathsf{T}}}{\rho(Z_i;\beta^*)} - \frac{\left(\sum_{j=1}^{m_n}\rho^{(1)}(Z_j;\beta^*)\right)\left(\sum_{j=1}^{m_n}\rho^{(1)}(Z_j;\beta^*)\right)^{\mathsf{T}}}{\sum_{j=1}^{m_n}\rho(Z_j;\beta^*)}\Bigg\}.$$

For any non-zero $v \in \mathbb{R}^p$, let $a_i = \left[\rho^{(1)}(Z_i;\beta^*)\right]^{\mathsf{T}} v = v^{\mathsf{T}}\rho^{(1)}(Z_i;\beta^*)$. Then by the Cauchy-Schwarz inequality

$$\left(\sum_{j=1}^{m_n} a_j\right)^2 = \left(\sum_{j=1}^{m_n}\frac{a_j}{\sqrt{\rho(Z_i;\beta^*)}}\sqrt{\rho(Z_i;\beta^*)}\right)^2 \le \left(\sum_{i=1}^{m_n}\frac{a_i^2}{\rho(Z_i;\beta^*)}\right)\left(\sum_{j=1}^{m_n}\rho(Z_j;\beta^*)\right)$$

and the equality holds if and only if $a_i = b\rho(Z_i;\beta^*)$ for $i = 1,\ldots,m_n$ and some $b \in \mathbb{R}$. Thus for all $v \in \mathbb{R}^p \backslash \mathcal{N}$, where $\mathcal{N} = \{v \in \mathbb{R}^p : \exists b \in \mathbb{R}, v^{\mathsf{T}}\rho^{(1)}(Z_i;\beta^*) = b\rho(Z_i;\beta^*), i = 1,\ldots,m_n\}$ is a Lebesgue null set,

$$v^{\mathsf{T}}\tilde{I}_n v = \frac{1}{m_n|D_n|}\int_{D_n}\lambda_0(x)\mathrm{d}x\left\{\sum_{i=1}^{m_n}\frac{a_i^2}{\rho(Z_i;\beta^*)} - \frac{\left(\sum_{j=1}^{m_n} a_j\right)^2}{\sum_{j=1}^{m_n}\rho(Z_j;\beta^*)}\right\} > 0,$$

which means that $\tilde{I}_n$ is a positive-definite $p \times p$ matrix and hence all its eigenvalues are positive (see equation (4) in Waagepetersen and Guan, 2009). Then by Lemma 3-4 in Waagepetersen and Guan (2009),

$$\liminf_{n\to\infty}\left(\inf\left\{\beta^{\mathsf{T}}\tilde{\Sigma}_n^{-1/2}\tilde{I}_n\tilde{\Sigma}_n^{-1/2}\beta : \beta \in \mathbb{R}^p, \|\beta\| = 1\right\}\right) > 0$$

and hence there exists a $0 < l < \infty$ such that $\mathbb{P}\{l_n < l\} \to 0$, where

$$l_n = \inf\left\{\beta^{\mathsf{T}}(\tilde{V}^{(n)})^{-1}J_n(\beta^*)(\tilde{V}^{(n)})^{-1}\beta : \beta \in \mathbb{R}^p, \|\beta\| = 1\right\},$$

and

$$\sup\left\{\left\|(\tilde{V}^{(n)})^{-1}\big[J_n(\hat{\beta}_n) - J_n(\beta^*)\big](\tilde{V}^{(n)})^{-1}\right\|_{\mathrm{M}} : \beta \in \mathbb{R}^p, \|\tilde{V}^{(n)}(\beta - \beta^*)\| \le q\right\} \xrightarrow{\mathbb{P}} 0,$$

for any $q > 0$. Finally, Lemma 5 in Waagepetersen and Guan (2009) ensures that $\left\|(\tilde{V}^{(n)})^{-1}u_n(\beta^*)\right\| = O_{\mathbb{P}}(1)$. Therefore, conditions (a)-(d) of Theorem 2 in Waagepetersen and Guan (2009) guarantees that there exists a sequence $\{\hat{\beta}_n\}$ where $\|\hat{\beta}_n - \beta^*\| = O_{\mathbb{P}}((m_n|D_n|)^{-1/2})$ and $u_n(\hat{\beta}_n) = 0$ with a probability tending to 1 as

$n \to \infty$ (see Waagepetersen and Guan, 2009, Appendix A). $\square$

A.4.2 Lemma. *Under conditions F1-F4,* $\left\|\hat{A}^{(n)} - A^{(n)}\right\|_1 = O_{\mathbb{P}}((m_n|D_n|)^{-1/2})$, $\left\|\hat{B}^{(n)} - B^{(n)}\right\| = O_{\mathbb{P}}((m_n|D_n|)^{-1/2})$, $\left\|\hat{A}^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ *and* $\|\hat{B}^{(n)} - V^{(n)}\| = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ *and* $\|V^{(n)} - M^{(n)}\Theta_n\| = O(h_n^2)$.

Proof. Similar to the proof of Lemma 6.1.1 in Chapter 6, we can easily derive that $\left\|A^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ and $\|B^{(n)} - V^{(n)}\| = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ (see Van der Vaart, 2000, p. 10).

Let's consider the difference between $\hat{B}^{(n)}$ and $B^{(n)}$ and $\hat{A}^{(n)}_{l,k}$ and $A^{(n)}_{l,k}$. For each $n \in \mathbb{N}$ and $l, k = 1, 2$, we have

$$\hat{B}_l^{(n)} - B_l^{(n)} = \frac{1}{m_n}\sum_{i=1}^{m_n}\sum_{u,v\in X_i\cap D_n}\frac{K(|u-v|-r)}{h_n\gamma_n(u-v)\rho^2(Z_i;\hat{\beta}_n)}G_l(|u-v|-r)$$
$$-\frac{1}{m_n}\sum_{i=1}^{m_n}\sum_{u,v\in X_i\cap D_n}\frac{K(|u-v|-r)}{h_n\gamma_n(u-v)\rho^2(Z_i;\beta^*)}G_l(|u-v|-r)$$
$$=\frac{1}{m_n h_n}\sum_{i=1}^{m_n}\left\{\left[\frac{1}{\rho^2(Z_i;\hat{\beta}_n)} - \frac{1}{\rho^2(Z_i;\beta^*)}\right]\sum_{u,v\in X_i\cap D_n}\frac{K(|u-v|-r)}{\gamma_n(u-v)}G_l(|u-v|-r)\right\}.$$

and

$$\hat{A}_{l,k}^{(n)} - A_{l,k}^{(n)} = \frac{1}{m_n(m_n-1)h_n}\sum_{i\neq j=1}^{m_n}\left\{\left[\frac{1}{\rho(Z_i;\hat{\beta}_n)\rho(Z_j;\hat{\beta}_n)} - \frac{1}{\rho(Z_i;\beta^*)\rho(Z_j;\beta^*)}\right]\right.$$
$$\left.\sum_{x\in X_i\cap D_n}\sum_{y\in X_j\cap D_n}\frac{K(|u-v|-r)G_l(|u-v|)G_k(|u-v|)}{\gamma_n(u-v)}\right\}.$$

Using the Taylor series expansions

$$\frac{1}{\rho^2(Z_i;\beta)} = \frac{1}{\rho^2(Z_i;\beta^*)} - \frac{2}{\rho^3(Z_i;\beta^*)}\left[\rho^{(1)}(Z_i;\beta^*)\right]^{\mathsf{T}}(\beta-\beta^*) + O(\|\beta-\beta^*\|^2),$$
$$\frac{1}{\rho(Z_i;\beta)\rho(Z_j;\beta)} = \frac{1}{\rho(Z_i;\beta^*)\rho(Z_i;\beta^*)}$$
$$-\frac{1}{\rho^2(Z_i;\beta^*)\rho^2(Z_i;\beta^*)}\left[\rho(Z_j;\beta^*)\rho^{(1)}(Z_i;\beta^*) + \rho(Z_i;\beta^*)\rho^{(1)}(Z_j;\beta^*)\right]^{\mathsf{T}}(\beta-\beta^*)$$
$$+\,O(\|\beta-\beta^*\|^2)$$

combined with conditions F2 and F3, we have

$$\left|\frac{1}{\rho^2(Z_i;\hat{\beta}_n)} - \frac{1}{\rho^2(Z_i;\beta^*)}\right| \le \frac{2}{\rho^3_{\min}}\left\|\rho^{(1)}(Z_i;\beta^*)\right\|\|\hat{\beta}_n-\beta^*\| + O(\|\hat{\beta}_n-\beta^*\|^2)$$
$$\le \tilde{C}_\rho\|\hat{\beta}_n-\beta^*\|^2$$

and

$$\left|\frac{1}{\rho(Z_i;\hat{\beta}_n)\rho(Z_j\hat{\beta}_n)} - \frac{1}{\rho(Z_i;\beta^*)\rho(Z_j;\beta^*)}\right|$$
$$\le \frac{1}{\rho^3_{\min}}\left(\left\|\rho^{(1)}(Z_i;\beta^*)\right\| + \left\|\rho^{(1)}(Z_j;\beta^*)\right\|\right)\|\hat{\beta}_n-\beta^*\| + O(\|\hat{\beta}_n-\beta^*\|^2)$$
$$\le \bar{C}_\rho\|\hat{\beta}_n-\beta^*\|^2$$

for sufficiently large n and some $0 < \tilde{C}_\rho, \bar{C}_\rho < \infty$. Thus

$$|\hat{B}^{(n)}_l - B^{(n)}_l| \le \tilde{C}_\rho\|\hat{\beta}_n-\beta^*\|^2\tilde{B}^{(n)}_l,$$
$$\left|\hat{A}^{(n)}_{l,k} - A^{(n)}_{l,k}\right| \le \bar{C}_\rho\|\hat{\beta}_n-\beta^*\|^2\tilde{A}^{(n)}_{l,k},$$

where

$$\tilde{B}^{(n)}_l = \frac{1}{m_n h_n}\sum_{i=1}^{m_n}\sum_{u,v\in X_i\cap D_n}\frac{K(|u-v|-r)}{\gamma_n(u-v)}G_l(|u-v|-r),$$
$$\tilde{A}^{(n)}_{l,k} = \frac{1}{m_n(m_n-1)h_n}\sum_{i\neq j=1}^{m_n}\sum_{x\in X_i\cap D_n}\sum_{y\in X_j\cap D_n}\frac{K(|u-v|-r)G_l(|u-v|)G_k(|u-v|)}{\gamma_n(u-v)}$$

Similar to $B^{(n)}_l$, we have $\mathrm{Var}\tilde{B}^{(n)}_l \le C_{V_{\tilde{B}}}(m_n|D_n|h_n)^{-1}$ and $\left(\mathbb{E}\,\tilde{B}^{(n)}_l\right)^2 < C^2_{E_{\tilde{B}}}$, for some $0 < C^2_{E_{\tilde{B}}}, C_{V_{\tilde{B}}} < \infty$. Similarly, it can be seen that $\mathrm{Var}\tilde{A}^{(n)}_{l,k} \le C_{V_{\tilde{A}}}(m_n|D_n|h_n)^{-1}$ and

$\big(\mathbb{E}\,\tilde{A}_{l,k}^{(n)}\big)^2 < C_{E_{\tilde{A}}}^2$, for some $0 < C_{E_{\tilde{A}}}^2, C_{V_{\tilde{A}}} < \infty$. Thus,

$$\begin{aligned}
\mathbb{E}\,\|\tilde{B}^{(n)}\|^2 = \sum_{l=1}^{2}\mathbb{E}(\tilde{B}_l^{(n)})^2 &= \sum_{l=1}^{2}[\mathrm{Var}\tilde{B}_l^{(n)} + (\mathbb{E}\,\tilde{B}_l^{(n)})^2] \\
&\le 2\big(C_{V_{\tilde{B}}}(m_n|D_n|h_n)^{-1} + C_{E_{\tilde{B}}}^2\big), \\
\mathbb{E}\,\|\tilde{A}^{(n)}\|_1^2 \le 2\,\mathbb{E}\,\|\tilde{A}^{(n)}\|_{\mathrm{F}}^2 &= 2\sum_{l=1}^{2}\sum_{k=1}^{2}[\mathrm{Var}\tilde{A}_{l,k}^{(n)} + (\mathbb{E}\,\tilde{A}_{l,k}^{(n)})^2] \\
&\le 8\big(C_{V_{\tilde{A}}}(m_n|D_n|h_n)^{-1} + C_{E_{\tilde{A}}}^2\big),
\end{aligned}$$

which mean that $\mathbb{E}\,\|\tilde{B}^{(n)}\|^2 = O(1)$ and $\mathbb{E}\,\|\tilde{A}^{(n)}\|_1^2 = O(1)$. Since

$$\begin{aligned}
\big\|\hat{B}^{(n)} - B^{(n)}\big\| &\le \tilde{C}_\rho\|\hat{\beta}_n - \beta^*\|^2\|\tilde{B}^{(n)}\|, \\
\big\|\hat{A}_{l,k}^{(n)} - A_{l,k}^{(n)}\big\| &\le \bar{C}_\rho\|\hat{\beta}_n - \beta^*\|^2\|\tilde{A}_{l,k}^{(n)}\|,
\end{aligned}$$

Lemma A.4.1 results in $\big\|\hat{A}^{(n)} - A^{(n)}\big\|_1 = O_{\mathbb{P}}((m_n|D_n|)^{-1})$ and $\|\hat{B}^{(n)} - B^{(n)}\| = O_{\mathbb{P}}((m_n|D_n|)^{-1})$. Thus we have $\big\|\hat{A}^{(n)} - M^{(n)}\big\|_1 = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ and $\|\hat{B}^{(n)} - V^{(n)}\| = O_{\mathbb{P}}((m_n|D_n|h_n)^{-1/2})$ by applying the triangle inequality.

Finally, similar to 6.1.1 in chapter 6, we can derive that $\|V^{(n)} - M^{(n)}\Theta_t\| = O(h_n^2)$. □

A.4.3 Lemma. *Assume $h^{-5/2}(m_n|D_n|)^{-\frac{1}{2}} \to 0$. Then $\big\|(\hat{A}^{(n)})_{(1\cdot)}^{-1}\big\| = O_{\mathbb{P}}(\big\|(M^{(n)})_{(1\cdot)}^{-1}\big\|)$, where $(\hat{A}^{(n)})_{(1\cdot)}^{-1}$ denotes the first row vector of matrix $(\hat{A}^{(n)})^{-1}$, similar definition for $(M^{(n)})_{(1\cdot)}^{-1}$.*

Proof. The proof here is similar to what we have derived in Lemma 6.1.2 in Chapter 6. □

A.4.4 Theorem. *Under assumptions $F1$-$F4$, $h_n^{-\frac{5}{2}}(m_n|D_n|)^{-1/2} \to 0$, then $|\hat{\Theta}_1 - \Theta_1| \xrightarrow{\mathbb{P}} 0$.*

Proof of this theorem can be derived similar to the proof in Theorem 6.1.3.

A.5 Consistency of the Semiparametric Orthogonal Series Estimator

Let $X_1, \dots, X_{m_n}$ be independent point processes on $\mathbb{R}$ with the intensity functions $\lambda_i(x)$ and the same pair correlation function $g(|x-y|)$. For each single intensity function, we assume $\lambda_i(x) = \lambda_0(x)\rho(Z_i;\beta^*)$, $i = 1, \cdots, m_n$, where $\lambda_0(x)$ is the baseline intensity function, $Z_i = (Z_{i1}, \dots, Z_{ip})^{\mathsf{T}}$ are time invariant covariates associated with point process X_i, $\beta^* = (\beta_1^*, \dots, \beta_p^*)^{\mathsf{T}}$ is the vector of true coefficients corresponding to each covariate and ρ is some known nonnetative function. We assume that A1-A3 similar to what we assume in section A.2.

Given a complete orthonormal basis of functions $\phi_l(r)$ on $[0, R]$, the orthogonal series expansion of the square-integrable pair correlation function $g(r)$ on $[0, R]$ is given by

$$g(r) = \sum_{l=1}^{\infty} \theta_l \phi_l(r), \quad \text{where} \quad \theta_l = \int_0^R \phi_l \cdot g(r) \mathrm{d}r.$$

By Parseval's identity (see Tolstov, 1962, p. 119),

$$\sum_{l=1}^{\infty} \theta_l^2 = \int_0^R g^2(r) \mathrm{d}r < \infty,$$

and hence $\sum_{k=l}^{\infty} \theta_k^2 \to 0$ and $\theta_l \to 0$, as $l \to \infty$. Let

$$g_n(r) = \sum_{l=1}^{L_n} \theta_l \phi_l(r) \quad \text{and} \quad \zeta_n(r) = g(r) - g_n(r) = \sum_{l=L_n+1}^{\infty} \theta_l \phi_l(r).$$

Then for each $r \in (0, R)$, $|\zeta_n(r)| \to 0$. In addition (see Tolstov, 1962, p. 55),

$$\int_0^R \zeta_n^2(r) \mathrm{d}r = \sum_{l=L_n+1}^{\infty} \theta_l^2 \to 0.$$

Let

$$\gamma_n(h) = |D_n \cap (D_n - h)| = \int_{\mathbb{R}} \mathbb{I}[y \in D_n, y + h \in D_n] \mathrm{d}y, \quad h \in \mathbb{R}.$$

Then for any fixed $h \in \mathbb{R}$, $\gamma_n(h)/|D_n| \to 1$ and hence a constant $0 < C_\gamma \leq 1$ can be found such that $\gamma_n(h) \geq C_\lambda |D_n|$ for all $-R \leq h \leq R$ and sufficiently large n. Define

$$A_{l,k}^{(n)} = \frac{1}{m_n(m_n-1)} \sum_{i \neq j=1}^{m_n} A_{l,k}^{(n,i,j)}, \quad l,k = 1,\ldots,L_n,$$

and

$$B_l^{(n)} = \frac{1}{m_n} \sum_{i=1}^{m_n} B_l^{(n,i)}, \quad l = 1,\ldots,L_n,$$

where

$$A_{l,k}^{(n,i,j)} = \sum_{x \in X_i \cap D_n} \sum_{y \in X_j \cap D_n} \frac{f_{l,k}(|x-y|)}{\gamma_n(x-y)\rho(Z_i;\beta^*)\rho(Z_j;\beta^*)},$$
$$B_l^{(n,i)} = \sum_{x,y \in X_i \cap D_n} \frac{\mathbb{I}[0 < |x-y| \leq R]}{\gamma_n(x-y)\rho^2(Z_i;\beta^*)} \phi_l(|x-y|),$$

$f_{l,k}(r) = \mathbb{I}[0 < r \leq R]\phi_l(r)\phi_k(r)$ and $\mathbb{I}[\,\cdot\,]$ denotes the indicator function. Then,

$$M_{l,k}^{(n)} = \mathbb{E}\, A_{l,k}^{(n)} = \mathbb{E}\, A_{l,k}^{(n,1,2)} = \int_{(D_n)^2} \frac{f_{l,k}(|x-y|)}{\gamma_n(x-y)} \lambda_0(x)\lambda_0(y)\mathrm{d}x\mathrm{d}y$$

and

$$\begin{aligned} V_l^{(n)} = \mathbb{E}\, B_l^{(n)} = \mathbb{E}\, B_l^{(n,1)} \\ = \int_{(D_n)^2} \frac{\mathbb{I}[0 < |x-y| \leq R]}{\gamma_n(x-y)} \phi_l(|x-y|) g(|x-y|) \lambda_0(x)\lambda_0(y)\mathrm{d}x\mathrm{d}y. \end{aligned}$$

The regression parameter β can be estimated using the method in Lawless and Nadeau (1995) and Lin et al. (2000) without having to specify $\lambda_0(t)$. More specifically, β is estimated by solving the estimating equation $u_n(\beta) = 0$, where

$$u_n(\beta) = \frac{1}{m_n} \sum_{i=1}^{m_n} u_{n,i}(\beta),$$

and

$$u_{n,i}(\beta) = \sum_{x \in X_i \cap D_n} \left(\frac{\rho^{(1)}(Z_i;\beta)}{\rho(Z_i;\beta)} - \frac{\sum_{j=1}^{m_n} \rho^{(1)}(Z_j;\beta)}{\sum_{j=1}^{m_n} \rho(Z_j;\beta)} \right)$$

This equation does not dependent on time events, but relies on the number of events in each single pattern.

Then we can substitute the true value β^* with $\hat{\beta}_n$ in previous two functions to get the practical versions

$$\hat{A}_{l,k}^{(n)} = \frac{1}{m_n(m_n-1)} \sum_{i \neq j=1}^{m_n} \hat{A}_{l,k}^{(n,i,j)}, \quad l,k = 1,\dots,L_n,$$

and

$$\hat{B}_l^{(n)} = \frac{1}{m_n} \sum_{i=1}^{m_n} \hat{B}_l^{(n,i)}, \quad l = 1,\dots,L_n,$$

where

$$\hat{A}_{l,k}^{(n,i,j)} = \sum_{x \in X_i \cap D_n} \sum_{y \in X_j \cap D_n} \frac{f_{l,k}(|x-y|)}{\gamma_n(x-y)\rho(Z_i;\hat{\beta})\rho(Z_j;\hat{\beta})},$$

$$\hat{B}_l^{(n,i)} = \sum_{x,y \in X_i \cap D_n} \frac{\mathbb{I}[0 < |x-y| \leq R]}{\gamma_n(x-y)\rho^2(Z_i;\hat{\beta})} \phi_l(|x-y|).$$

Let $A^{(n)} = [A_{l,k}^{(n)}]$, $\hat{A}^{(n)} = [\hat{A}_{l,k}^{(n)}]$, $M^{(n)} = [M_{l,k}^{(n)}]$, $B^{(n)} = (B_l^{(n)})$, $\hat{B}^{(n)} = (\hat{B}_l^{(n)})$ $V^{(n)} = (V_l^{(n)})$ and $\Theta_n = (\theta_1,\dots,\theta_{L_n})$. Then Θ_n is estimated by solving the estimating equations

$$\sum_{m=1}^{L_n} \hat{A}_{l,m}^{(n)} \hat{\theta}_m - \hat{B}_l^{(n)} = 0, \quad l = 1,\dots,L_n,$$

or $\hat{A}^{(n)}\hat{\Theta}_n - \hat{B}^{(n)} = 0$.

We consider the Euclidean norm $\|B^{(n)}\| = \left(\sum_{l=1}^{L_n} \left(B_l^{(n)}\right)^2\right)^{1/2}$ for vectors and the maximum absolute column sum norm (Isaacson and Keller, 1994, p. 9)

$$\|A^{(n)}\|_1 = \max_{1 \leq k \leq L_n} \sum_{l=1}^{L_n} |A_{l,k}^{(n)}|,$$

the Frobenius norm

$$\left\|A^{(n)}\right\|_{\mathrm{F}} = \left(\sum_{l=1}^{L_n} \sum_{k=1}^{L_n} (A_{l,k}^{(n)})^2\right)^{1/2}$$

and the elementwise max norm

$$\left\|A^{(n)}\right\|_{\max} = \max_{1 \leq l,k \leq L_n} |A_{l,k}^{(n)}|$$

for matrices. It is known that $\|A^{(n)}\|_1 \leq \sqrt{L_n}\|A^{(n)}\|_{\mathrm{F}}$ (see Golub and Van Loan, 1996, p. 56)

We assume that the following conditions hold:

W1 The basis functions are uniformly bounded; i.e. $|\phi_l(r)| \leq C_\phi$ for all $l \in \mathbb{N}$, $0 \leq r \leq R$ and some $0 < C_\phi < \infty$.

W2 For all $x \in \mathbb{R}$, $0 \leq \lambda_0(x) \leq \lambda_{\max} < \infty$.

W3 For any $z, \beta \in \mathbb{R}^P$, $0 < \rho_{\min} < \rho(z;\beta) < \rho_{\max} < \infty$.

W4 For any fixed z, $\rho(z;\beta)$ is twice continuously differentiable as a function of β and $\|\rho^{(1)}(z;\beta)\| < C_\rho \|\beta - \beta^*\|$ if $\|\beta - \beta^*\| < \varepsilon_\rho$ for some $0 < \varepsilon_\rho, C_\rho < \infty$, where $\rho^{(1)}(z;\beta) = \mathrm{d}\rho(z;\beta)/\mathrm{d}\beta$.

W5 The constants $0 < C_I, C_g, C_{g^{(3)}}, C_{g^{(4)}} < \infty$ can be found such that

$$\int_0^\infty \Big|g(r) - 1\Big| \mathrm{d}r < C_I$$

and for all $x_1, x_2, x_3, x_4 \in \mathbb{R}$, $g(|x_1 - x_2|) \leq C_g$, $g^{(3)}(x_1, x_2, x_3) \leq C_{g^{(3)}}$ and

$$G^{(4)}(x_1, x_2, x_3) = \int_{\mathbb{R}} \Big| g^{(4)}(x_1 + x_2, x_2, x_3 + z, z) - g(|x_1|) g(|x_3|) \Big| \mathrm{d}z \leq C_{g^{(4)}}.$$

W6 For some $\eta > 0$, $\big\|(M^{(n)})^{-1}\big\|_1 = O(L_n^\eta)$.

W7 The coefficients of the expansion of $g(r)$ satisfy $\theta_l = O(l^{-(1+\delta)})$ for some $\delta > \eta$.

The condition W1 is satisfied for the cosine basis

$$\phi_l(r) = \begin{cases} 1/\sqrt{R} & l = 1 \\ (\sqrt{2}/\sqrt{R}) \cos\big((l-1)\pi r/R\big) & l \geq 2 \end{cases}$$

with $C_\phi = \sqrt{2}/R$. The condition W1-condition W5 are mild conditions. The condition W6 holds in the homogeneous case where $\lambda(x) \equiv \lambda_0$, because

$$\lim_{n\to\infty} \int_{(D_n)^2} \frac{f_{l,k}(|x-y|)}{\gamma_n(x-y)} \mathrm{d}x\mathrm{d}y = 2\int_0^R \phi_l(r)\phi_k(r)\mathrm{d}r = 2\mathbb{I}[l = k].$$

Regarding condition W7, the rate of convergence of θ_l to zero depends on the smoothness of the pair correlation function. In fact, if $g(r)$ is differentiable on $(0, R)$ and $\int_0^R |g'(r)|\mathrm{d}r < \infty$, then $\theta_l = O(l^{-1})$ and if $g(r)$ is twice differentiable and $\int_0^R |g''(r)|\mathrm{d}r < \infty$ then $\theta_l = O(l^{-2})$ (Efromovich, 2008, p. 32). Thus W7 holds if $g(r)$ is twice differentiable and $\int_0^R |g''(r)|\mathrm{d}r < \infty$.

A.5.1 Lemma. *Let $\breve{\Sigma}_n = \mathrm{Var} u_{n,1}(\beta^*)$ and assume that $\breve{\Sigma}_n \to \breve{\Sigma}$, where $\breve{\Sigma}$ is a positive definite matrix. Under conditions $W1$-$W5$ and $W7$, there exists a sequence $\{\hat{\beta}_n\}$ where $\|\hat{\beta}_n - \beta^*\| = O_{\mathbb{P}}((m_n|D_n|)^{-1/2})$ and $u_n(\hat{\beta}_n) = 0$ with a probability tending to 1 as $n \to \infty$.*

Proof. The proof can be found in Lemma A.4.1 □

A.5.2 Lemma. *Under conditions $W1$-$W5$ and $W7$, $\left\|\hat{A}^{(n)} - A^{(n)}\right\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1})$, $\|\hat{B}^{(n)} - B^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1})$, $\left\|\hat{A}^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1/2})$, $\|\hat{B}^{(n)} - V^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1/2})$ and $\|V^{(n)} - M^{(n)}\Theta_n\| = O(L_n^{-\delta})$.*

Proof. Similar to the proof of Lemma A.2.1, we can easily derive that $\left\|A^{(n)} - M^{(n)}\right\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1/2})$ and $\|B^{(n)} - V^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1/2})$ (see Van der Vaart, 2000, p. 10).

Let's consider the difference between $\hat{B}^{(n)}$ and $B^{(n)}$ and $\hat{A}^{(n)}_{l,k}$ and $A^{(n)}_{l,k}$. For each $n \in \mathbb{N}$ and $l, k = 1, \cdots, L_n$, we have

$$\begin{aligned}
\hat{B}^{(n)}_l - B^{(n)}_l =& \frac{1}{m_n}\sum_{i=1}^{m_n}\sum_{x,y\in X_i\cap D_n}\frac{\mathbb{I}[0<|x-y|\le R]}{\gamma_n(x-y)\rho^2(Z_i;\hat{\beta}_n)}\phi_l(|x-y|) \\
&- \frac{1}{m_n}\sum_{i=1}^{m_n}\sum_{x,y\in X_i\cap D_n}\frac{\mathbb{I}[0<|x-y|\le R]}{\gamma_n(x-y)\rho^2(Z_i;\beta^*)}\phi_l(|x-y|) \\
=& \frac{1}{m_n}\sum_{i=1}^{m_n}\left\{\left[\frac{1}{\rho^2(Z_i;\hat{\beta}_n)} - \frac{1}{\rho^2(Z_i;\beta^*)}\right]\sum_{x,y\in X_i\cap D_n}\frac{\mathbb{I}[0<|x-y|\le R]}{\gamma_n(x-y)}\phi_l(|x-y|)\right\}.
\end{aligned}$$

and

$$\begin{aligned}
\hat{A}^{(n)}_{l,k} - A^{(n)}_{l,k} = \frac{1}{m_n(m_n-1)}\sum_{i\neq j=1}^{m_n}&\left\{\left[\frac{1}{\rho(Z_i;\hat{\beta}_n)\rho(Z_j;\hat{\beta}_n)} - \frac{1}{\rho(Z_i;\beta^*)\rho(Z_j;\beta^*)}\right]\right. \\
&\left.\sum_{x\in X_i\cap D_n}\sum_{y\in X_j\cap D_n}\frac{f_{l,k}(|x-y|)}{\gamma_n(x-y)}\right\}.
\end{aligned}$$

Using the Taylor series expansions

$$\frac{1}{\rho^2(Z_i;\beta)} = \frac{1}{\rho^2(Z_i;\beta^*)} - \frac{2}{\rho^3(Z_i;\beta^*)}\left[\rho^{(1)}(Z_i;\beta^*)\right]^{\mathsf{T}}(\beta-\beta^*) + O(\|\beta-\beta^*\|^2),$$
$$\frac{1}{\rho(Z_i;\beta)\rho(Z_j;\beta)} = \frac{1}{\rho(Z_i;\beta^*)\rho(Z_i;\beta^*)}$$
$$- \frac{1}{\rho^2(Z_i;\beta^*)\rho^2(Z_i;\beta^*)}\left[\rho(Z_j;\beta^*)\rho^{(1)}(Z_i;\beta^*) + \rho(Z_i;\beta^*)\rho^{(1)}(Z_j;\beta^*)\right]^{\mathsf{T}}(\beta-\beta^*)$$
$$+ O(\|\beta-\beta^*\|^2)$$

combined with conditions W3 and W4, we have

$$\left|\frac{1}{\rho^2(Z_i;\hat\beta_n)} - \frac{1}{\rho^2(Z_i;\beta^*)}\right| \le \frac{2}{\rho^3_{\min}}\left\|\rho^{(1)}(Z_i;\beta^*)\right\|\|\hat\beta_n-\beta^*\| + O(\|\hat\beta_n-\beta^*\|^2)$$
$$\le \tilde C_\rho\|\hat\beta_n-\beta^*\|^2$$

and

$$\left|\frac{1}{\rho(Z_i;\hat\beta_n)\rho(Z_j\hat\beta_n)} - \frac{1}{\rho(Z_i;\beta^*)\rho(Z_j;\beta^*)}\right|$$
$$\le \frac{1}{\rho^3_{\min}}\left(\left\|\rho^{(1)}(Z_i;\beta^*)\right\| + \left\|\rho^{(1)}(Z_j;\beta^*)\right\|\right)\|\hat\beta_n-\beta^*\| + O(\|\hat\beta_n-\beta^*\|^2)$$
$$\le \bar C_\rho\|\hat\beta_n-\beta^*\|^2$$

for sufficiently large n and some $0 < \tilde C_\rho, \bar C_\rho < \infty$. Thus

$$|\hat B_l^{(n)} - B_l^{(n)}| \le \tilde C_\rho\|\hat\beta_n-\beta^*\|^2\tilde B_l^{(n)},$$
$$\left|\hat A_{l,k}^{(n)} - A_{l,k}^{(n)}\right| \le \bar C_\rho\|\hat\beta_n-\beta^*\|^2\tilde A_{l,k}^{(n)},$$

where

$$\tilde B_l^{(n)} = \frac{1}{m_n}\sum_{i=1}^{m_n}\sum_{x,y\in X_i\cap D_n}\frac{\mathbb{I}[0<|x-y|\le R]}{\gamma_n(x-y)}\phi_l(|x-y|),$$
$$\tilde A_{l,k}^{(n)} = \frac{1}{m_n(m_n-1)}\sum_{i\ne j=1}^{m_n}\sum_{x\in X_i\cap D_n}\sum_{y\in X_j\cap D_n}\frac{f_{l,k}(|x-y|)}{\gamma_n(x-y)}.$$

Similar to $B_l^{(n)}$, we have $\mathrm{Var}\tilde B_l^{(n)} \le C_{V_{\tilde B}}(m_n|D_n|)^{-1}$ and $\left(\mathbb{E}\,\tilde B_l^{(n)}\right)^2 < C^2_{E_{\tilde B}}$, for some $0 < C^2_{E_{\tilde B}}, C_{V_{\tilde B}} < \infty$. Similarly, it can be seen that $\mathrm{Var}\tilde A_{l,k}^{(n)} \le C_{V_{\tilde A}}(m_n|D_n|)^{-1}$ and

$\big(\mathbb{E}\,\tilde{A}_{l,k}^{(n)}\big)^2 < C_{E_{\tilde{A}}}^2$, for some $0 < C_{E_{\tilde{A}}}^2, C_{V_{\tilde{A}}} < \infty$. Thus,

$$\begin{aligned}
\mathbb{E}\,\|\tilde{B}^{(n)}\|^2 = \sum_{l=1}^{L_n} \mathbb{E}(\tilde{B}_l^{(n)})^2 &= \sum_{l=1}^{L_n} [\mathrm{Var}\tilde{B}_l^{(n)} + (\mathbb{E}\,\tilde{B}_l^{(n)})^2] \\
&\le L_n\big(C_{V_{\tilde{B}}}(m_n|D_n|)^{-1} + C_{E_{\tilde{B}}}^2\big), \\
\mathbb{E}\,\|\tilde{A}^{(n)}\|_1^2 \le L_n\,\mathbb{E}\,\|\tilde{A}^{(n)}\|_{\mathrm{F}}^2 &= L_n \sum_{l=1}^{L_n}\sum_{k=1}^{L_n} [\mathrm{Var}\tilde{A}_{l,k}^{(n)} + (\mathbb{E}\,\tilde{A}_{l,k}^{(n)})^2] \\
&\le L_n^3\big(C_{V_{\tilde{A}}}(m_n|D_n|)^{-1} + C_{E_{\tilde{A}}}^2\big),
\end{aligned}$$

which mean that $\mathbb{E}\,\|\tilde{B}^{(n)}\|^2 = O(L_n)$ and $\mathbb{E}\,\|\tilde{A}^{(n)}\|_1^2 = O(L_n^3)$. By Markov's inequality, for any $n \in \mathbb{N}$ and $c > 0$,

$$\begin{aligned}
\mathbb{P}\left\{L_n^{-3/2}\big\|\tilde{A}^{(n)}\big\|_1 > c\right\} &\le \frac{\mathbb{E}\,\big\|\tilde{A}^{(n)}\big\|_1^2}{c^2 L_n^3}, \\
\mathbb{P}\left\{L_n^{-1/2}\|\tilde{B}^{(n)}\| > c\right\} &\le \frac{\mathbb{E}\,\|\tilde{B}^{(n)}\|^2}{c^2 L_n}.
\end{aligned}$$

The right hand sides of the above inequalities can be made arbitrary small, for all n, by choosing sufficiently large c and hence $\big\|\tilde{A}^{(n)}\big\| = O_{\mathbb{P}}(L_n^{3/2})$ and $\|\tilde{B}^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2})$.

Since

$$\begin{aligned}
\big\|\hat{B}^{(n)} - B^{(n)}\big\| &\le \tilde{C}_\rho\|\hat{\beta}_n - \beta^*\|^2\|\tilde{B}^{(n)}\|, \\
\big\|\hat{A}_{l,k}^{(n)} - A_{l,k}^{(n)}\big\| &\le \bar{C}_\rho\|\hat{\beta}_n - \beta^*\|^2\|\tilde{A}_{l,k}^{(n)}\|,
\end{aligned}$$

Lemma A.5.1 results in $\big\|\hat{A}^{(n)} - A^{(n)}\big\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1})$ and $\|\hat{B}^{(n)} - B^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1})$. Thus we can have $\big\|\hat{A}^{(n)} - M^{(n)}\big\|_1 = O_{\mathbb{P}}(L_n^{3/2}(m_n|D_n|)^{-1/2})$ and $\|\hat{B}^{(n)} - V^{(n)}\| = O_{\mathbb{P}}(L_n^{1/2}(m_n|D_n|)^{-1/2})$ by applying triangle inequality.

Finally,

$$\begin{aligned}
V_l^{(n)} &= \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \le R]}{\gamma_n(x-y)}\phi_l(|x-y|)g(|x-y|)\lambda_0(x)\lambda_0(y)\mathrm{d}x\mathrm{d}y \\
&= \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \le R]}{\gamma_n(x-y)}\phi_l(|x-y|)\left(\sum_{k=1}^{\infty}\theta_k\phi_k(|x-y|)\right)\lambda_0(x)\lambda_0(y)\mathrm{d}x\mathrm{d}y \\
&= \sum_{k=1}^{L_n}\theta_k M_{l,k}^{(n)} + \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \le R]}{\gamma_n(x-y)}\phi_l(|x-y|)\zeta_n(|x-y|)\lambda_0(x)\lambda_0(y)\mathrm{d}x\mathrm{d}y,
\end{aligned}$$

and hence

$$\left|V_l^{(n)} - \sum_{k=1}^{L_n} \theta_k M_{l,k}^{(n)}\right| \leq \lambda_{\max}^2 \int_{(D_n)^2} \frac{\mathbb{I}[|x-y| \leq R]}{\gamma_n(x-y)} \big|\phi_l(|x-y|)\big|\big|\zeta_n(|x-y|)\big| \mathrm{d}x\mathrm{d}y$$
$$= 2\lambda_{\max}^2 \int_0^R \big|\phi_l(r)\big|\big|\zeta_n(r)\big| \mathrm{d}r$$
$$\leq 2\lambda_{\max}^2 \left(\int_0^R \phi_l^2(r)\mathrm{d}r\right)^{1/2} \left(\int_0^R \zeta_n^2(r)\mathrm{d}r\right)^{1/2}$$
$$= 2\lambda_{\max}^2 \left(\sum_{k=L_n+1}^{\infty} \theta_k^2\right)^{1/2}.$$

Thus

$$\|V^{(n)} - M^{(n)}\Theta_n\|^2 = \sum_{l=1}^{L_n} \left(V_l^{(n)} - \sum_{k=1}^{L_n} \theta_k M_{l,k}^{(n)}\right)^2$$
$$\leq 4\lambda_{\max}^4 L_n \sum_{l=L_n+1}^{\infty} \theta_l^2.$$

By condition W7,

$$\sum_{l=L_n+1}^{\infty} \theta_l^2 = O(L_n^{-(1+2\delta)}),$$

which guarantees that $L_n \sum_{l=L_n+1}^{\infty} \theta_l^2 = O(L_n^{-2\delta})$ and hence $\|V^{(n)} - M^{(n)}\Theta_n\| = O(L_n^{-\delta})$. □

A.5.3 Lemma. *Assume condition W6 holds and* $L_n^{3/2+\eta}(m_n|D_n|)^{-1} \to 0$. *Then* $\big\|(\hat{A}^{(n)})^{-1}\big\|_1 = \big\|(A^{(n)})^{-1}\big\|_1 = O_{\mathbb{P}}(\big\|(M^{(n)})^{-1}\big\|_1)$.

Proof. The proof here is similar to what we have derived in Lemma A.2.2 in appendix A. □

A.5.4 Theorem. *Under assumptions W2-W7, if* $L_n^{3+2\eta}(m_n|D_n|)^{-1} \to 0$ *then* $\|\hat{\Theta}_n - \Theta_n\| \xrightarrow{\mathbb{P}} 0$.

Proof. Proof of this theorem can be derived similar to the proof in Theorem A.2.3. □

For each $r \in [0, R]$, let

$$\hat{g}_n(r) = \sum_{l=1}^{L_n} \hat{\theta}_l \phi_l(r).$$

and define the functional norm

$$\|\hat{g}_n - g_n\|_2 = \left(\int_0^R \big(\hat{g}_n(r) - g_n(r)\big)^2 \mathrm{d}r \right)^{1/2}.$$

Then,

$$\big(\hat{g}_n(r) - g_n(r)\big)^2 = \left(\sum_{l=1}^{L_n} (\hat{\theta}_l - \theta_l)\phi_l(r) \right)^2 \leq \left(\sum_{l=1}^{L_n} (\hat{\theta}_l - \theta_l)^2 \right) \left(\sum_{l=1}^{L_n} \phi_l^2(r) \right)$$

and

$$\int_0^R \big(\hat{g}_n(r) - g_n(r)\big)^2 \mathrm{d}r \leq \|\hat{\Theta}_n - \Theta_n\|^2 \sum_{l=1}^{L_n} \int_0^R \phi_l^2(r)\mathrm{d}r = L_n \|\hat{\Theta}_n - \Theta_n\|^2.$$

and we can conclude that $\|\hat{g}_n - g_n\|_2 = O_{\mathbb{P}}(L_n^{1/2+\eta-\delta} + L_n^{2+\eta}(m_n|D_n|)^{-1/2})$ and state the following corollary.

A.5.5 Corollary. *Under conditions W2-W7, if $\delta > \eta + 1/2$ and $L_n^{2+\eta}(m_n|D_n|)^{-1/2} \to 0$, then $\|\hat{g}_n - g_n\|_2 \xrightarrow{\mathbb{P}} 0$.*

The above corollary implies that $\|\hat{g}_n - g\|_2 \xrightarrow{\mathbb{P}} 0$ under the same conditions, because $\hat{g}_n(r) - g(r) = \hat{g}_n(r) - g_n(r) - \zeta_n(r)$, $\|\zeta_n\|_2 \to 0$ and

$$\|\hat{g}_n - g\|_2 \leq \|\hat{g}_n - g_n\|_2 + \|\zeta_n\|_2.$$

Appendix B

Asymptotic Normality

B.1 Asymptotic Normality of the Log Local Linear Estimator

Define two random vectors

$$\boldsymbol{Z}_1 = \frac{1}{m} \sum_{i=1}^{m} \sum_{u,v \in X_i}^{\neq} \sum w_{r,h}(|u-v|) \boldsymbol{A}_h(|u-v|-r), \tag{B.1}$$

$$\boldsymbol{Z}_2(\boldsymbol{\theta}^*) = \frac{1}{m(m-1)} \sum_{i \neq j} \sum \sum_{u \in X_i, v \in X_j} \sum w_{r,h}(|u-v|) \tilde{g}_r(|u-v|; \boldsymbol{\theta}^*) \boldsymbol{A}_h(|u-v|-r) \tag{B.2}$$

By definition of $\boldsymbol{\theta}^*$ in (A.1), we have that

$$\mathbb{E}\boldsymbol{Z}_1 = \mathbb{E}\boldsymbol{Z}_2 = \int_{D_n^2} \lambda(u)\lambda(v) w_{r,h}(|u-v|) g(|u-v|) \boldsymbol{A}_h(|u-v|-r) \mathrm{d}u\mathrm{d}v. \tag{B.3}$$

B.1.1 Lemma. *Under conditions $C1$-$C6$, we have that, as $h \to 0$,*

$$
\begin{aligned}
(m|D_n|h)\mathrm{Var}\,(\boldsymbol{Z}_1) &= 2g(r)\mathbf{Q}_{n,h}^{(2)}(r) + O(h), & \text{(B.4)}\\
(m|D_n|h)\mathrm{Var}\,[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)] &= \frac{2}{m-1}g^2(r)\mathbf{Q}_{n,h}^{(2)}(r) + O(h), & \text{(B.5)}\\
(m|D_n|h)\mathrm{Cov}\,[\boldsymbol{Z}_1, \boldsymbol{Z}_2(\boldsymbol{\theta}^*)] &= O(h), & \text{(B.6)}
\end{aligned}
$$

where $\mathbf{Q}_{n,h}^{(2)}(r)$ is as defined in (6.6) *and the convergence is entry-wise.*

Note that by conditions C1-C6, it is trivial to see that eigenvalues of the matrix $\mathbf{Q}_{n,h}^{(2)}(r)$ are bounded from below and above at the same time.

Proof. Under conditions C1-C5, using similar arguments as those in the proof of Lemma A.1.2, we can immediately show that

$$
\begin{aligned}
m\mathrm{Var}\,(\boldsymbol{Z}_1) = &\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)[g^{(4)}(u_1,v_1,u_2,v_2)\\
&\quad - g(|u_1-v_1|)g(|u_2-v_2|)]\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_2-v_2|-r)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2\\
&+ 4\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)g^{(3)}(u_1,v_1,u_2)\\
&\quad \times \boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h(|u_1-u_2|-r)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\\
&+ 2\int_{D_n^2} \lambda(u_1)\lambda(v_1)\left[w_{r,h}(|u_1-v_1|)\right]^2 g(|u_1-v_1|)\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-v_1|-r)\mathrm{d}u_1\mathrm{d}v_1\\
= &\ 2\int_{D_n^2} \lambda(u_1)\lambda(v_1)\left[w_{r,h}(|u_1-v_1|)\right]^2 g(|u_1-v_1|)\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-v_1|-r)\mathrm{d}u_1\mathrm{d}v_1 + O(|D_n|^{-1})\\
= &\ 2g(r)\left\{1+O(h)\right\}\int_{D_n^2} \lambda(u_1)\lambda(v_1)\left[w_{r,h}(|u_1-v_1|)\right]^2 \boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-v_1|-r)\mathrm{d}u_1\mathrm{d}v_1\\
&\quad + O(|D_n|^{-1}),
\end{aligned}
$$

where the last equality follows from the fact that $|g(t)-g(r)| = O(h)$ for any $r-h \le t \le r+h$ as $h \to 0$. Similarly, we can show that under conditions C1-C5, we have that

$$
\begin{aligned}
m\mathrm{Var}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] &= \frac{2}{m-1}\int_{D_n^2}\lambda(u_1)\lambda(v_1)w_{r,h}^2(|u_1-v_1|)\tilde{g}_r^2(|u_1-v_1|;\boldsymbol{\theta}^*) \\
&\quad\times \boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-v_1|-r)\mathrm{d}u_1\mathrm{d}v_1 \\
&+ \frac{4}{m-1}\int_{D_n^3}\lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*) \\
&\quad\times \tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*)g(|v_1-u_2|)\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-u_2|-r)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ \frac{4(m-2)}{m-1}\int_{D_n^3}\lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*) \\
&\quad\times \boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-u_2|-r)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ \frac{2}{m-1}\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)[g(|u_1-u_2|)g(|v_1-v_2|)-1] \\
&\quad\times \tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_2-v_2|-r)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ \frac{4(m-2)}{m-1}\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)[g(|u_1-u_2|)-1] \\
&\quad\times \tilde{g}_r(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*)\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_2-v_2|-r)\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v_1\mathrm{d}v_2 \\
&= \int_{D_n^2}\frac{2\lambda(u_1)\lambda(v_1)}{m-1}w_{r,h}^2(|u_1-v_1|)\tilde{g}_r^2(|u_1-v_1|;\boldsymbol{\theta}^*)\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-v_1|-r)\mathrm{d}u_1\mathrm{d}v_1 + O(|D_n|^{-1}) \\
&= g^2(r)\left\{2+O(h)\right\}\int_{D_n^2}\frac{\lambda(u_1)\lambda(v_1)}{m-1}w_{r,h}^2(|u_1-v_1|)\boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_1-v_1|-r)\mathrm{d}u_1\mathrm{d}v_1 + O(|D_n|^{-1}).
\end{aligned}
$$

and that

$$
\begin{aligned}
m\mathrm{Cov}\left[\boldsymbol{Z}_1,\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] &= 4\int_{D_n^3}\lambda(u_1)\lambda(v_1)\lambda(u_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)g(|u_1-v_1|)\tilde{g}_r(|u_1-u_2|;\boldsymbol{\theta}^*) \\
&\quad\times \boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h(|u_1-U_2|-r)^T\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&+ 2\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)g^{(3)}(u_1,v_1,u_2)\tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*) \\
&\quad\times \boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_2-v_2|-r)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&- 2\int_{D_n^4}\lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)g(|u_1-v_1|)\tilde{g}_r(|u_2-v_2|;\boldsymbol{\theta}^*) \\
&\quad\times \boldsymbol{A}_h(|u_1-v_1|-r)\boldsymbol{A}_h^T(|u_2-v_2|-r)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&= O(|D_n|^{-1}).
\end{aligned}
$$

Combining above three equalities, we can conclude that as $h \to 0$

$$
\begin{aligned}
(m|D_n|h)\mathrm{Var}\left(\boldsymbol{Z}_1\right) &= 2g(r)\mathbf{Q}_{n,h}^{(2)}(r) + O(h), \\
(m|D_n|h)\mathrm{Var}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] &= \frac{2}{m-1}g^2(r)\mathbf{Q}_{n,h}^{(2)}(r) + O(h), \\
(m|D_n|h)\mathrm{Cov}\left[\boldsymbol{Z}_1,\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] &= O(h),
\end{aligned}
$$

where $\mathbf{Q}_{n,h}^{(2)}(r)$ is as defined in (6.6) and the convergence is entry-wise. Note that by conditions C1-C6, it is trivial to see that eigenvalues of the matrix $\mathbf{Q}_{n,h}^{(2)}(r)$ are bounded from below and above at the same time. □

B.1.2 Lemma. *Under conditions C1-C6, we have that, as $h \to 0$ and $m|D_n|h \to \infty$,*

$$\sqrt{m|D|_n h}\boldsymbol{\Sigma}_Z^{-1/2}(\boldsymbol{\theta}^*)\left[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] \to^d N(\mathbf{0}, \mathbf{I}), \tag{B.7}$$

where $\boldsymbol{\Sigma}_Z(\boldsymbol{\theta}^*) = 2(m-1+g(r))/(m-1)g(r)\mathbf{Q}_{n,h}^{(2)}(r)$ with $\mathbf{Q}_{n,h}^{(2)}(r)$ defined in (6.6)

Proof. By equation (B.6) of Lemma B.1.1, we can see that $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2(\boldsymbol{\theta}^*)$ are asymptotically independent as $h \to 0$. Hence, it suffices to consider asymptotic normality of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2(\boldsymbol{\theta}^*)$ separately. We divide our discussions into two case scenarios: (1) $m \to \infty$ and (2) m is fixed.

Case I: when $m \to \infty$. In this case, from equations (B.4)-(B.5), we can see that as $m \to \infty$,

$$\sqrt{m|D_n|h}\left\{\boldsymbol{Z}_2(\boldsymbol{\theta}^*) - \mathbb{E}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right]\right\} = o_p(1),$$

which implies that

$$\sqrt{m|D|_n h}\left[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] = \sqrt{m|D|_n h}\left(\boldsymbol{Z}_1 - \mathbb{E}\boldsymbol{Z}_1\right) + o_p(1),$$

since $\mathbb{E}\boldsymbol{Z}_1 = \mathbb{E}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right]$. Let $\boldsymbol{Y}_i = \sum\sum_{u,v\in X_i}^{\neq} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r)$ and then $\boldsymbol{Z}_1 = \frac{1}{m}\sum_{i=1}^m \boldsymbol{Y}_i$. By definition, $\boldsymbol{Y}_i$'s are identically distributed, thus it immediately follows from the standard multivariate central limit theorem that as $m \to \infty$,

$$\left[\mathrm{Var}(\boldsymbol{Z}_1)\right]^{-1/2}\left(\boldsymbol{Z}_1 - \mathbb{E}\boldsymbol{Z}_1\right) \to^d N(\mathbf{0}, \mathbf{I}),$$

which coincides with (B.7) after plugging (B.4) back to the above equation.

Case II: when m is fixed. In this case, condition $m|D_n|h \to \infty$ requires that $|D_n| \to \infty$. In other words, we need to consider the case where the observation window of the point processes is expanding. Define a partition of $\mathbb{R} = \cup_{t\in\mathbb{Z}}\Delta(t)$, where $\Delta_h(t) = (h^{-1}(t-1/2), h^{-1}(t+1/2)]$. Note that by this definition, $\Delta_h(t_1)\cap\Delta_h(t_2) = \emptyset$ if $t_1 \neq t_2 \in \mathbb{Z}$. Define random vectors

$$\boldsymbol{Y}_{1,n}(t) = \frac{|D_n|h}{m}\sum_{i=1}^m \sum_{u\in X_i\cap\Delta_h(t), v\in X_i}\sum^{\neq} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r),$$

$$\boldsymbol{Y}_{2,n}(t) = \frac{|D_n|h}{m(m-1)}\sum\sum_{i\neq j}\sum\sum_{u\in X_i\cap\Delta(t), v\in X_j} w_{r,h}(|u-v|)\tilde{g}_r(|u-v|;\boldsymbol{\theta}^*)\boldsymbol{A}_h(|u-v|-r).$$

Then by definition, we have that

$$\boldsymbol{Z}_1 = \frac{1}{|D_n|h} \sum_{t \in \mathcal{T}_n} \boldsymbol{Y}_{1,n}(t), \quad \boldsymbol{Z}_2(\boldsymbol{\theta}^*) = \frac{1}{|D_n|h} \sum_{t \in \mathcal{T}_n} \boldsymbol{Y}_{2,n}(t),$$

where $\mathcal{T}_n = \{t \in \mathbb{Z} : \Delta_h(t) \cap D_n \neq \emptyset\}$.

Under conditions C4-C5, it is straightforward to see that there exists a constant C_1 such that

$$|D_n| h w_{r,h}(|u-v|) |\boldsymbol{A}_h(|u-v|-r)| \leq C_1 I(||u-v|-r| < h).$$

A simple application of the Jensen's inequality gives that $(m^{-1}\sum_{i=1}^m |x_i|)^{2+\lceil\delta\rceil} \leq m^{-1}\sum_{i=1}^m |x_i|^{2+\lceil\delta\rceil}$ (note that $f(x) = x^{2+\lceil\delta\rceil}$ is convex for $x > 0$)

$$\begin{aligned}
\mathbb{E}\left|\boldsymbol{Y}_{1,n}(t)\right|^{2+\lceil\delta\rceil} &\leq \frac{1}{m}\sum_{i=1}^m \mathbb{E}\left|\sum_{u \in X_i \cap \Delta_h(t), v \in X_i}^{\neq}\sum |D_n| h w_{r,h}(|u-v|) \boldsymbol{A}_h(|u-v|-r)\right|^{2+\lceil\delta\rceil} \\
&= \mathbb{E}\left|\sum_{u \in X_1 \cap \Delta_h(t), v \in X_1}^{\neq}\sum |D_n| h w_{r,h}(|u-v|) \boldsymbol{A}_h(|u-v|-r)\right|^{2+\lceil\delta\rceil} \\
&\leq \mathbb{E}\left\{\sum_{u \in X_1 \cap \Delta_h(t), v \in X_1}^{\neq}\sum |D_n| h w_{r,h}(|u-v|) |\boldsymbol{A}_h(|u-v|-r)|\right\}^{2+\lceil\delta\rceil} \\
&\leq C_1^{2+\lceil\delta\rceil} \mathbb{E}\left\{\sum_{u \in X_1 \cap \Delta_h(t), v \in X_1}^{\neq}\sum I\left(||u-v|-r| < h\right)\right\}^{2+\lceil\delta\rceil},
\end{aligned}$$

where the last expectation is essentially bounded by sums of integrals involving $\lambda(u)$, $g(s)$, $g^{(k)}(u_1, \cdots, u_k)$, $k = 3, \cdots, 6$. All terms are bounded under conditions C1-C3 and condition N2, hence these integrals are bounded uniformly in t and n. Recall that δ is defined in condition N2. Therefore, we have that

$$\sup_{n \geq 1} \sup_{t \in \mathcal{T}_n} \mathbb{E}\left|\boldsymbol{Y}_{1,n}(t)\right|^{2+\delta} \leq \sup_{n \geq 1} \sup_{t \in \mathcal{T}_n} \mathbb{E}\left|\boldsymbol{Y}_{1,n}(t)\right|^{2+\lceil\delta\rceil} < \infty. \tag{B.8}$$

Similar, using equation (A.3) in Lemma A.1.1 and condition C2a, we have that $\tilde{g}_r(t; \boldsymbol{\theta}^*)$ is also uniformly bounded and following similar arguments as above, we can show that

$$\sup_{n \geq 1} \sup_{t \in \mathcal{T}_n} \mathbb{E}\left|\boldsymbol{Y}_{2,n}(t)\right|^{2+\delta} \leq \sup_{n \geq 1} \sup_{t \in \mathcal{T}_n} \mathbb{E}\left|\boldsymbol{Y}_{2,n}(t)\right|^{2+\lceil\delta\rceil} < \infty. \tag{B.9}$$

Note that the total number of disjoint partitions in $\mathbb{R} = \cup_{t\in\mathbb{Z}}\Delta(t)$ is of the order $|D_n|h$, hence we can check that, using equations (B.4)-(B.5),

$$(|D_n|h)^{-1}\text{Var}\left[\sum_{t\in\mathcal{T}_n}\boldsymbol{Y}_{1,n}(t)\right] = (|D_n|h)^{-1}\text{Var}\left(|D_n|h\boldsymbol{Z}_1\right) = |D_n|h\text{Var}\left(\boldsymbol{Z}_1\right),$$

$$(|D_n|h)^{-1}\text{Var}\left[\sum_{t\in\mathcal{T}_n}\boldsymbol{Y}_{2,n}(t)\right] = (|D_n|h)^{-1}\text{Var}\left[|D_n|h\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] = |D_n|h\text{Var}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right],$$

both of above matrices have strictly positive eigenvalues under condition C6. Therefore, using conditions $N1(b)$ and $N2$, together with inequalities (B.22)-(B.9), it follows from Theorem 3.1 of Biscio and Waagepetersen (2016) that as $|D_n|h \to \infty$,

$$\left\{\text{Var}\left[\sum_{t\in\mathcal{T}_n}\boldsymbol{Y}_{k,n}(t)\right]\right\}^{-1/2}\sum_{t\in\mathcal{T}_n}\left[\boldsymbol{Y}_{k,n}(t) - \mathbb{E}\boldsymbol{Y}_{k,n}(t)\right] \to^d N(\mathbf{0},\mathbf{I}), \quad k = 1, 2,$$

which is equivalent to stating that

$$\left[\text{Var}(\boldsymbol{Z}_1)\right]^{-1/2}\left(\boldsymbol{Z}_1 - \mathbb{E}\boldsymbol{Z}_1\right) \to^d N(\mathbf{0},\mathbf{I}),$$

and

$$\left\{\text{Var}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right]\right\}^{-1/2}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*) - \mathbb{E}\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] \to^d N(\mathbf{0},\mathbf{I}).$$

Recall that by Lemma B.1.1, $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2(\boldsymbol{\theta}^*)$ are asymptotically independent as $h \to 0$, and that $\mathbb{E}\boldsymbol{Z}_1 = \mathbb{E}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right]$ by definition, we can conclude that

$$\left\{\text{Var}(\boldsymbol{Z}_1) + \text{Var}\left[\boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right]\right\}^{-1/2}\left[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\boldsymbol{\theta}^*)\right] \to^d N(\mathbf{0},\mathbf{I}),$$

which coincides with (B.7) after plugging (B.4)-(B.5) back to the above equation. The proof is complete. □

B.1.3 Lemma. *Denote* $\widehat{\boldsymbol{\theta}}$ *as the solution to estimating equations* (6.3)*, then under conditions* $C1$*-*$C6$*,* $N1$*-*$N2$*, we have that*

$$\boldsymbol{\Delta}_h(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \equiv \begin{bmatrix} (\hat{\theta}_0 - \theta_0^*) \\ h(\hat{\theta}_1 - \theta_1^*) \\ \vdots \\ h^p(\hat{\theta}_p - \theta_p^*) \end{bmatrix} = \left[g(r)\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\left[\boldsymbol{Z}_1 - \boldsymbol{Z}_2(\boldsymbol{\theta}^*) + o_p\left(\frac{1}{\sqrt{m|D|_n h}}\right)\right], \tag{B.10}$$

where $\boldsymbol{Z}_1$ *and* $\boldsymbol{Z}_2(\boldsymbol{\theta}^*)$ *are defined in* (B.1) *and* (B.2)*, respectively.*

Proof. By the definition of $\tilde{\boldsymbol{U}}_{r,h}(\boldsymbol{\theta})$ in (6.3), solving $\tilde{\boldsymbol{U}}_{r,h}(\boldsymbol{\theta}) = \mathbf{0}$ is equivalent to solving $\tilde{\boldsymbol{V}}_{r,h}(\boldsymbol{\theta}) = \mathbf{0}$ for $\boldsymbol{\theta}$, where

$$\tilde{\boldsymbol{V}}_{r,h}(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\sum_{u,v\in X_i}\sum^{\neq} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r) - \frac{1}{m(m-1)}\sum\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j} w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r)\tilde{g}_r(|u-v|;\boldsymbol{\theta}).$$

Using the first order Taylor expansion, we can show that

$$\tilde{\boldsymbol{V}}_{r,h}(\widehat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{V}}_{r,h}(\boldsymbol{\theta}^*) = -\tilde{\mathbf{H}}_{h,r}(\tilde{\boldsymbol{\theta}}^*)\boldsymbol{\Delta}_h(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \tag{B.11}$$

where $\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_h$ with $\|\cdot\|_h$ as defined in Lemma A.1.2 and

$$\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}) = \sum\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j}\frac{w_{r,h}(|u-v|)}{m(m-1)}\boldsymbol{A}_h(|u-v|-r)\boldsymbol{A}_h^T(|u-v|-r)\tilde{g}_r(|u-v|;\boldsymbol{\theta}). \tag{B.12}$$

By definition, we have that for any $r-h \leq t \leq r+h$,

$$\begin{aligned}|\tilde{g}_r(t;\boldsymbol{\theta}^*) - \tilde{g}_r(t;\tilde{\boldsymbol{\theta}}^*)| &= \tilde{g}_r(t;\boldsymbol{\theta}^*)\left|1 - \exp\left[\theta_0^* - \tilde{\theta}_0^* + h(\theta_1^* - \tilde{\theta}_1^*)\frac{t-r}{h}\cdots + h^p(\theta_p^* - \tilde{\theta}_p^*)\frac{(t-r)^p}{h^p}\right]\right| \\ &\leq \tilde{g}_r(t;\boldsymbol{\theta}^*)\sqrt{p+1}\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h \exp(\sqrt{p+1}\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h),\end{aligned}$$

where the last inequality follows form the fact that $|1-e^x| \leq |x|e^{|x|}$ and Cauchy-Schwartz inequality. Since $\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|_h = O_p\left(1/\sqrt{m|D_n|h}\right)$ by Lemma A.1.2, we have that

$$\begin{aligned}\eta_{\max}\left[\tilde{\mathbf{H}}_{h,r}(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right] &= \sup_{\|\boldsymbol{\delta}\|=1}\boldsymbol{\delta}^T\left[\tilde{\mathbf{H}}_{h,r}(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right]\boldsymbol{\delta} \\ &\leq \sup_{\|\boldsymbol{\delta}\|=1}\sum\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j}\frac{w_{r,h}(|u-v|)}{m(m-1)}[\boldsymbol{\delta}^T\boldsymbol{A}_h(|u-v|-r)]^2|\tilde{g}_r(|u-v|;\tilde{\boldsymbol{\theta}}^*) - \tilde{g}_r(|u-v|;\boldsymbol{\theta}^*)| \\ &\leq \sup_{\|\boldsymbol{\delta}\|=1}\sum\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j}\frac{w_{r,h}(|u-v|)}{m(m-1)}[\boldsymbol{\delta}^T\boldsymbol{A}_h(|u-v|-r)]^2\tilde{g}_r(|u-v|;\boldsymbol{\theta}^*) \\ &\qquad\times\sqrt{p+1}\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h\exp(\sqrt{p+1}\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h) \\ &= \eta_{\max}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right]\times\sqrt{p+1}\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h\exp(\sqrt{p+1}\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|_h) \\ &= \eta_{\max}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right]O_p\left(1/\sqrt{m|D_n|h}\right).\end{aligned}$$

Following exactly the same steps, we can also show that

$$-\eta_{\min}\left[\tilde{\mathbf{H}}_{h,r}(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right] = \eta_{\max}\left[-\tilde{\mathbf{H}}_{h,r}(\tilde{\boldsymbol{\theta}}^*) + \tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right] = \eta_{\max}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right]O_p\left(1/\sqrt{m|D_n|h}\right),$$

which implies that

$$\tilde{\mathbf{H}}_{h,r}(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*) = \eta_{\max}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right] O_p\left(1/\sqrt{m|D_n|h}\right), \tag{B.13}$$

where the convergence is entry-wise.

The next step is to quantify the variabilities of elements in $\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)$, denoted as H_{ij}'s. Following steps as those in the proof of Lemma A.1.2 about Var $\left[H''_{m,n}(z_0)\right]$, under conditions C1, C2(a)-(b), C4 and equation (A.3), some tedious algebra gives that

$$\begin{aligned}
m(m-1)\mathrm{Var}\left(H_{ij}\right) &= O(1)\int_{D_n^4} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)|g(|u_1-u_2|)g(|v_1-v_2|)-1| \\
&\qquad \mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2 \\
&+ O(1)\int_{D_n^3} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 + O(1)\int_{D_n^2} w_{r,h}^2(|u_1-v_1|)\mathrm{d}u_1\mathrm{d}v_1 \\
&+ mO(1)\int_{D_n^4} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_2-v_2|)|g(|u_1-u_2|)-1|du_1du_2dv_1dv_2 \\
&+ mO(1)\int_{D_n^3} w_{r,h}(|u_1-v_1|)w_{r,h}(|u_1-u_2|)\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2 \\
&= O(1)|D_n|\int_{\mathbb{R}^3}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g(|w|)g(|t-s+w|)-1|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ O(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t + O(1)|D_n|\int_{\mathbb{R}}\left[\frac{K_h(|s|-r)}{\gamma_n(|s|)}\right]^2\mathrm{d}s \\
&+ mO(1)|D_n|\int_{\mathbb{R}^3}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}|g(|w|)-1|\mathrm{d}s\mathrm{d}t\mathrm{d}w \\
&+ mO(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t \\
&= O(1)|D_n|\int_{\mathbb{R}}\left[\frac{K_h(|s|-r)}{\gamma_n(|s|)}\right]^2\mathrm{d}s + mO(1)|D_n|\int_{\mathbb{R}^2}\frac{K_h(|s|-r)K_h(|t|-r)}{\gamma_n(|s|)\gamma_n(|t|)}\mathrm{d}s\mathrm{d}t
\end{aligned}$$

Then, by condition C5, we finally have that

$$\mathrm{Var}\left(H_{ij}\right) = \frac{1}{m^2|D_n|h}O(1) + \frac{1}{m|D_n|}O(1) \to 0, \text{ as } m|D_n|h \to \infty,$$

which gives that as $m|D_n|h \to \infty$,

$$\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*) = \mathbb{E}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right] + o_p(1). \tag{B.14}$$

Next, we study $\mathbb{E}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right]$. By definition

$$\mathbb{E}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right] = \int_{D_n^2} \lambda(u)\lambda(v)w_{r,h}(|u-v|)\boldsymbol{A}_h(|u-v|-r)\boldsymbol{A}_h^T(|u-v|-r)\tilde{g}_r(|u-v|;\boldsymbol{\theta}^*)\mathrm{d}u\mathrm{d}v.$$

Recall the definition of $\mathbf{Q}_{n,h}^{(1)}(r)$ in (6.6) and the fact that for any $r-h \leq t \leq r+h$, $|\tilde{g}_r(|t|;\boldsymbol{\theta}^*) - g(r)| = g(r)O(h)$ by Lemma A.1.1, following the similar proof as that of equation (B.13), we have that

$$\mathbb{E}\left[\tilde{\mathbf{H}}_{h,r}(\boldsymbol{\theta}^*)\right] - g(r)\mathbf{Q}_{n,h}^{(1)}(r) = g(r)\eta_{\max}\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]O(h). \tag{B.15}$$

Combining equations (B.11)-(B.15), we have that

$$\tilde{\boldsymbol{V}}_{r,h}(\boldsymbol{\theta}^*) = \left\{g(r)\mathbf{Q}_{n,h}^{(1)}(r) + g(r)\eta_{\max}\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]\left[O(h) + o_p(1/\sqrt{m|D_n|h})\right] + o_p(1)\right\}\boldsymbol{\Delta}_h(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*),$$

which further gives that, under conditions C2(a) and Lemma A.1.2,

$$\boldsymbol{\Delta}_h(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) = \left[g(r)\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\left[\tilde{\boldsymbol{V}}_{r,h}(\boldsymbol{\theta}^*) + o_p\left(\frac{1}{\sqrt{m|D|_n h}}\right)\right]. \tag{B.16}$$

The proof is completed by observing that the definition of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2(\boldsymbol{\theta}^*)$ gives that $\tilde{\boldsymbol{V}}_{r,h}(\boldsymbol{\theta}^*) = \boldsymbol{Z}_1 - \boldsymbol{Z}_2(\boldsymbol{\theta}^*)$. □

B.1.4 Theorem. *Under conditions C1-C6, N1-N2, as $h \to \infty$ and $m|D_n|h \to \infty$, we have that*

$$\frac{\sqrt{m|D_n|h}\,[\hat{g}(r) - g(r) + b_{n,h}]}{\sigma_{m,n,h}} \to^D N(0,1),$$

where $b_{n,h} = O(h^{p+1})$ and $\sigma_{m,n,h}^2 = \frac{2g(r)[m-1+g(r)]}{m-1}\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}$, with $\boldsymbol{e} = (1,0,\cdots,0)_{p+1}^T$ and $\mathbf{Q}_{n,h}^{(k)}(r)$, $k=1,2$, are defined in (6.6).

Proof. By applying delta method to $\hat{g}(r) = \exp(\hat{\theta}_0)$ with Lemmas B.1.2 and B.1.3, we have that

$$\frac{\sqrt{m|D_n|h}\left[\hat{g}(r)-\exp(\theta_0^*)\right]}{\exp(\theta_0^*)\sqrt{2(m-1+g(r))/(m-1)[g(r)]^{-1}\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}}} \to N(0,1).$$

By Lemma A.1.1, we have that $\theta_0^* - \log[g(r)] = O(h^{p+1})$, which gives that

$$\exp(\theta_0^*) - g(r) = O(h^{p+1}).$$

Therefore, it readily follows that

$$\frac{\sqrt{m|D_n|h}\,[\hat{g}(r)-g(r)]}{\exp(\theta_0^*)\sqrt{2(m-1+g(r))/(m-1)[g(r)]^{-1}\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}}}$$
$$=\frac{\sqrt{m|D_n|h}\,[\hat{g}(r)-\exp(\theta_0^*)+\exp(\theta_0^*)-g(r)]}{\sqrt{2(m-1+g(r))/(m-1)g(r)\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}}}+o_p(1)$$
$$=\frac{\sqrt{m|D_n|h}\,[\hat{g}(r)-\exp(\theta_0^*)+O(h^{p+1})]}{\sqrt{2(m-1+g(r))/(m-1)g(r)\boldsymbol{e}^T\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\mathbf{Q}_{n,h}^{(2)}(r)\left[\mathbf{Q}_{n,h}^{(1)}(r)\right]^{-1}\boldsymbol{e}}}+o_p(1),$$

which completes the proof. □

B.2 Asymptotic Normality of the Log Orthogonal Series Estimator

B.2.1 Lemma. *Let* $\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*)=\boldsymbol{\delta}_L^T\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\delta}_L^T$ *with* $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)=\mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right]$*. If the vector* $\boldsymbol{\delta}_L$ *satisfies (a)* $\|\boldsymbol{\delta}_L\|=1$*; (b)* $\int_0^R\left[w_o(s)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(s)|\right]^{2+\lceil\delta\rceil}\mathrm{d}s=O(1)$*; and (c)* $\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*)\ge c_u$ *for some constant* $c_u>0$*, then under conditions C1-C6, N1-N2, we have that, as* $L\to\infty$ *and* $m|D_n|\to\infty$*,*

$$\frac{\sqrt{m|D_n|}\boldsymbol{\delta}_L^T\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)}{\tilde{\sigma}_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*)}\to^d N(0,1). \tag{B.17}$$

Proof. Following exact the same arguments in the proof of finding $mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right]$ in Lemma A.3.2, we have shown that

$$\mathrm{Var}\left[\boldsymbol{\delta}_L^T\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right]=O(m^{-1}|D_n|^{-1})\Rightarrow\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*)=O(1).$$

To study the asymptotic normality of $\boldsymbol{\delta}_L^T\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)$, we define two random vectors such that $\boldsymbol{\delta}_L^T\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)=Z_1-Z_2(\boldsymbol{\theta}^*)$ as follows

$$Z_1=\frac{1}{m}\sum_{i=1}^m\sum_{u,v\in X_i}\sum^{\neq}w_R(|u-v|)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|), \tag{B.18}$$

$$Z_2(\boldsymbol{\theta}^*) = \frac{1}{m(m-1)} \sum\sum_{i\neq j} \sum_{u\in X_i, v\in X_j}\sum w_R(|u-v|)\tilde{g}_L(|u-v|;\boldsymbol{\theta}^*)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|) \quad \text{(B.19)}$$

By definition of $\boldsymbol{\theta}^*$ in (6.16), we have that

$$\mathbb{E}Z_1 = \mathbb{E}Z_2(\boldsymbol{\theta}^*) = \int_{D_n^2} \lambda(u)\lambda(v)w_R(|u-v|)g(|u-v|)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|)\mathrm{d}u\mathrm{d}v. \quad \text{(B.20)}$$

We shall divide our discussions into two case scenarios: (1) $m \to \infty$ and (2) m is fixed.

Case I: when $m \to \infty$. In this case, the normality of Z_1 is easy to show since it is an average of independent random variables. The normality of $Z_2(\boldsymbol{\theta}^*)$ is less straightforward since it has a structure similar to a U-Statistics. To resolve issue, we define the following approximation

$$\tilde{Z}_2(\boldsymbol{\theta}^*) = \frac{2}{m}\sum_{i=1}^{m}\sum_{u\in X_i} q(u) - \mathbb{E}Z_2(\boldsymbol{\theta}^*), \quad \text{(B.21)}$$

where $q(u) = \int_{D_n} \lambda_0(v)w_R(|u-v|)\tilde{g}_L(|u-v|;\boldsymbol{\theta}^*)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|)\mathrm{d}v$. It is trivial to see that $\mathbb{E}Z_2(\boldsymbol{\theta}^*) = \mathbb{E}\tilde{Z}_2(\boldsymbol{\theta}^*)$ by definition. Following similar arguments as those in the proof of finding $mJ_{m,n}\mathrm{Var}\left[H'_{m,n}(0)\right]$ in Lemma A.3.2, some tedious algebra gives that

$$\begin{aligned}
&m\mathrm{Var}\left[\tilde{Z}_2(\boldsymbol{\theta}^*) - Z_2(\boldsymbol{\theta}^*)\right] = \frac{2}{m-1}\int_{D_n^2} \lambda(u_1)\lambda(v_1)w_R^2(|u_1-v_1|)\tilde{g}_L^2(|u_1-v_1|;\boldsymbol{\theta}^*)g(|u_1-v_1|)\\
&\times \left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]^2 \mathrm{d}u_1\mathrm{d}v_1\\
&\frac{4}{m-1}\int_{D_n^3} \lambda(u_1)\lambda(v_1)\lambda(u_2)w_R(|u_1-v_1|)w_R(|u_1-u_2|)\tilde{g}_L(|u_1-v_1|;\boldsymbol{\theta}^*)\\
&\times \tilde{g}_L(|u_1-u_2|;\boldsymbol{\theta}^*)\left[g(|v_1-u_2|)-1\right]\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_1-u_2|)^T\boldsymbol{\delta}_L\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\\
&+\frac{2}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)\tilde{g}_L(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_L(|u_2-v_2|;\boldsymbol{\theta}^*)\\
&\times g(|u_1-u_2|)\left[g(|v_1-v_2|)-1\right]\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right]\mathrm{d}u_1\mathrm{d}v_1\mathrm{d}u_2\mathrm{d}v_2\\
&-\frac{2}{m-1}\int_{D_n^4} \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2)w_R(|u_1-v_1|)w_R(|u_2-v_2|)\left[g(|u_1-u_2|)-1\right]\\
&\times \tilde{g}_L(|u_1-v_1|;\boldsymbol{\theta}^*)\tilde{g}_L(|u_2-v_2|;\boldsymbol{\theta}^*)\left[\boldsymbol{\phi}_L(|u_1-v_1|)^T\boldsymbol{\delta}_L\right]\left[\boldsymbol{\phi}_L(|u_2-v_2|)^T\boldsymbol{\delta}_L\right]\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}v_1\mathrm{d}v_2\\
&= O(m^{-1}|D_n|^{-1})\int_0^R w_o^2(s)\left[\boldsymbol{\phi}_L(s)^T\boldsymbol{\delta}_L\right]^2\mathrm{d}s + O(m^{-1}|D_n|^{-1})\left\{\int_0^R w_o(s)\left[\boldsymbol{\phi}_L(s)^T\boldsymbol{\delta}_L\right]\mathrm{d}s\right\}^2\\
&= O(m^{-1}|D_n|^{-1}).
\end{aligned}$$

Therefore, as $m \to \infty$, we have that

$$\sqrt{m|D_n|}\left[\tilde{Z}_2(\boldsymbol{\theta}^*) - Z_2(\boldsymbol{\theta}^*)\right] = O_p(m^{-1}) = o_p(1),$$

and hence

$$\sqrt{m|D_n|}\boldsymbol{\delta}_L^T\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*) = \sqrt{m|D_n|}\left[\tilde{Z}_2(\boldsymbol{\theta}^*) - Z_2(\boldsymbol{\theta}^*)\right] + o_p(1).$$

Since $m|D_n|\mathrm{Var}\left[\boldsymbol{\delta}_L^T\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right] \geq c_u$ for some constant $c_u > 0$, it suffices to show the asymptotic normality of

$$\sqrt{m|D_n|}\left[Z_1 - \tilde{Z}_2(\boldsymbol{\theta}^*)\right] = \frac{\sqrt{m|D_n|}}{m}\sum_{i=1}^m Y_i,$$

where Y_i's are i.i.d. random variables of the form as follows

$$Y_i = \sum_{u,v\in X_i}\sum^{\neq} w_R(|u-v|)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|) - 2\sum_{u\in X_i} q(u) + \mathbb{E}Z_2(\boldsymbol{\theta}^*).$$

Note that $\sqrt{|D_n|}Y_i$'s are i.i.d random variables with a bounded variance (straightforward to show), (B.17) immediately follows from the standard central limit theorem as $m \to \infty$.

Case II: when m is fixed. In this case, condition $m|D_n| \to \infty$ requires that $|D_n| \to \infty$. In other words, we need to consider the case where the observation window of the point processes is expanding. Define a partition of $\mathbb{R} = \cup_{t\in\mathbb{Z}}\Delta(t)$, where $\Delta(t) = (s(t-1/2), s(t+1/2)]$ with $s > 0$ as the length of the interval. Note that by this definition, $\Delta(t_1) \cap \Delta(t_2) = \emptyset$ if $t_1 \neq t_2 \in \mathbb{Z}$. Define random variables

$$Y_{1,n}(t) = \frac{|D_n|}{m}\sum_{i=1}^m \sum_{u\in X_i\cap\Delta(t),v\in X_i}\sum^{\neq} w_R(|u-v|)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|),$$

$$Y_{2,n}(t) = \frac{|D_n|}{m(m-1)}\sum_{i\neq j}\sum \sum_{u\in X_i\cap\Delta(t),v\in X_j}\sum w_R(|u-v|)\tilde{g}_L(|u-v|;\boldsymbol{\theta}^*)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|).$$

Then by definition, we have that

$$Z_1 = \frac{1}{|D_n|}\sum_{t\in\mathcal{T}_n}\boldsymbol{Y}_{1,n}(t), \quad Z_2(\boldsymbol{\theta}^*) = \frac{1}{|D_n|}\sum_{t\in\mathcal{T}_n}\boldsymbol{Y}_{2,n}(t),$$

where $\mathcal{T}_n = \{t \in \mathbb{Z} : \Delta(t) \cap D_n \neq \emptyset\}$.

A simple application of the Jensen's inequality gives that $(m^{-1}\sum_{i=1}^m |x_i|)^{2+\lceil\delta\rceil} \leq$

$m^{-1}\sum_{i=1}^m |x_i|^{2+\lceil\delta\rceil}$ (note that $f(x) = |x|^{2+\lceil\delta\rceil}$ is convex)

$$\begin{aligned}
\mathbb{E}\,|Y_{1,n}(t)|^{2+\lceil\delta\rceil} &\le \frac{1}{m}\sum_{i=1}^m \mathbb{E}\left|\sum_{u\in X_i\cap\Delta(t),v\in X_i}\sum^{\neq} |D_n| w_R(|u-v|)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|)\right|^{2+\lceil\delta\rceil} \\
&= \mathbb{E}\left|\sum_{u\in X_1\cap\Delta(t),v\in X_1}\sum^{\neq} |D_n| w_R(|u-v|)\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|)\right|^{2+\lceil\delta\rceil} \\
&\le \mathbb{E}\left\{\sum_{u\in X_1\cap\Delta(t),v\in X_1}\sum^{\neq} |D_n| w_R(|u-v|)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|)|\right\}^{2+\lceil\delta\rceil} \\
&= O(1)\mathbb{E}\left\{\sum_{u\in X_1\cap\Delta(t),v\in X_1}\sum^{\neq} I(|u-v|<R) w_o(|u-v|)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(|u-v|)|\right\}^{2+\lceil\delta\rceil},
\end{aligned}$$

where the last expectation is essentially bounded by sums of integrals involving $w_o^k(s)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(s)|^k,\ k\le 2+\lceil\delta\rceil,\ \lambda(u),\ g(s),\ g^{(k)}(u_1,\cdots,u_k),\ k=3,\cdots,6$. All terms are bounded under conditions E1-E3 and condition Q2 except the first batch, hence we only need to consider upper bounds of integrals of the form

$$\int_0^R w_o^k(s)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(s)|^k \mathrm{d}s, \quad k\le 2+\lceil\delta\rceil.$$

For any $k<2+\lceil\delta\rceil$, by the Höder's inequality with $p=(2+\lceil\delta\rceil)/k, q=1/[1-k/(2+\lceil\delta\rceil)]$ such that $1/p+1/q=1$, we have that

$$\begin{aligned}
\int_0^R w_o^k(s)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(s)|^k \mathrm{d}s &= \int_0^R [w_o(s)]^{k-1+1/q}\left\{\left[w_o(s)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(s)|\right]^{2+\lceil\delta\rceil}\right\}^{1/p}\mathrm{d}s \\
&\le \left\{\int_0^R \left[w_o(s)|\boldsymbol{\delta}_L^T\boldsymbol{\phi}_L(s)|\right]^{2+\lceil\delta\rceil}\mathrm{d}s\right\}^{1/p}\left\{\int_0^R w_o^{q(k-1)+1}(s)\mathrm{d}s\right\}^{1/q} \\
&= O(1),
\end{aligned}$$

where the last equality follows from the condition for $\boldsymbol{\delta}_L$. Therefore, we have that there exists a constant C_1 such that

$$\mathbb{E}\,|Y_{1,n}(t)|^{2+\lceil\delta\rceil} < C_1.$$

Similar arguments also yield that for some constant $C_2>0$

$$\mathbb{E}\,|-Y_{2,n}(t)|^{2+\lceil\delta\rceil} < C_2.$$

Then by the Minkowski inequality, we have that

$$\begin{aligned}\mathbb{E}\left|Y_{1,n}(t)-Y_{2,n}(t)\right|^{2+\lceil\delta\rceil} &\leq \left\{\left[\mathbb{E}\left|Y_{1,n}(t)\right|^{2+\lceil\delta\rceil}\right]^{1/(2+\lceil\delta\rceil)}+\left[\mathbb{E}\left|Y_{2,n}(t)\right|^{2+\lceil\delta\rceil}\right]^{1/(2+\lceil\delta\rceil)}\right\}^{2+\lceil\delta\rceil}\\ &< 2^{2+\lceil\delta\rceil}\max\{C_1,C_2\},\end{aligned}$$

which further gives that

$$\sup_{n\geq 1}\sup_{t\in\mathcal{T}_n}\mathbb{E}\left|Y_{1,n}(t)-Y_{2,n}(t)\right|^{2+\delta} \leq \sup_{n\geq 1}\sup_{t\in\mathcal{T}_n}\mathbb{E}\left|Y_{1,n}(t)-Y_{2,n}(t)\right|^{2+\lceil\delta\rceil}<\infty. \tag{B.22}$$

Note that the total number of disjoint partitions in $\mathbb{R}=\cup_{t\in\mathbb{Z}}\Delta(t)$ is of the order $|D_n|$, hence we can check that,

$$(|D_n|)^{-1}\mathrm{Var}\left\{\sum_{t\in\mathcal{T}_n}\left[Y_{1,n}(t)-Y_{2,n}(t)\right]\right\}=(|D_n|)^{-1}\mathrm{Var}\left(|D_n|\boldsymbol{\delta}_L^T\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right)=m^{-1}\tilde{\sigma}^2_{\boldsymbol{\delta}}(\boldsymbol{\theta}^*)\geq c_u/m.$$

Therefore, using conditions $N1(b)$ and $N2$, together with inequality (B.22), it follows from Theorem 3.1 of Biscio and Waagepetersen (2016) that as $|D_n|\to\infty$,

$$\left\{\mathrm{Var}\left[\sum_{t\in\mathcal{T}_n}Y_{1,n}(t)-\sum_{t\in\mathcal{T}_n}Y_{2,n}(t)\right]\right\}^{-1/2}\sum_{t\in\mathcal{T}_n}\left[Y_{1,n}(t)-Y_{2,n}(t)\right]\to^d N(0,1),$$

which coincides with (B.17) by definition of $Y_{k,n}$'s, $k=1,2$. □

B.2.2 Lemma. *Denote $\widehat{\boldsymbol{\theta}}$ as the solution to estimating equations* (6.13)*, then under conditions E*1*-E*6*, we have that as $L\to\infty$ and $L^{4\nu_0+2\nu_2}/m|D_n|\to 0$, for any $0<r\leq R$,*

$$\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)=\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*)+o_p(1)\left\|\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\right\|, \tag{B.23}$$

where $\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*)$ and $\mathbf{Q}_L$ are defined in (6.13) *and* (6.17)*, respectively. Furthermore, under additional conditions Q*1*-Q*3*, we have that*

$$\frac{\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)}{\sigma_L(r;\boldsymbol{\theta}^*)}\to^d N(0,1), \tag{B.24}$$

where $\sigma_L^2(r;\boldsymbol{\theta}^*) = \boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)$ *and* $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right]$.

Proof. Recall the definition

$$\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}) = \frac{1}{m}\sum_{i=1}^{m}\sum\sum_{u,v\in X_i}^{\neq} w_R(|u-v|)\boldsymbol{\phi}_L(|u-v|) \\ - \frac{1}{m(m-1)}\sum\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j} w_R(|u-v|)\boldsymbol{\phi}_L(|u-v|)\tilde{g}_L(|u-v|;\boldsymbol{\theta}).$$

Using the first order Taylor expansion, we can show that

$$\tilde{\boldsymbol{U}}_L(\widehat{\boldsymbol{\theta}}) - \tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*) = -\tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \tag{B.25}$$

where $\|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\| \leq \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|$ and

$$\tilde{\mathbf{H}}_L(\boldsymbol{\theta}) = \sum\sum_{i\neq j}\sum\sum_{u\in X_i, v\in X_j} \frac{w_R(|u-v|)}{m(m-1)}\boldsymbol{\phi}_L(|u-v|)\boldsymbol{\phi}_L^T(|u-v|)\tilde{g}_L(|u-v|;\boldsymbol{\theta}). \tag{B.26}$$

Observe that $\tilde{\boldsymbol{U}}_L(\widehat{\boldsymbol{\theta}}) = \mathbf{0}$, we can re-write expansion (B.25) as follows

$$\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*) = \left(\mathbf{Q}_L + \mathbf{Q}^{\Delta}\right)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*), \tag{B.27}$$

where $\mathbf{Q}_L$ is defined in (6.17) and $\mathbf{Q}^{\Delta} = \tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) - \mathbf{Q}_L$. From the above new expansion, we have that

$$\begin{aligned} \boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) &= \boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\left[\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*) - \mathbf{Q}^{\Delta}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)\right] \\ &\leq \boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*) + \left\|\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\right\|\sigma_{\max}\left[\mathbf{Q}^{\Delta}\right]\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|, \end{aligned} \tag{B.28}$$

where $\sigma_{\max}(\boldsymbol{A})$ stands for the largest singular value of the matrix $\boldsymbol{A}$. We have shown the order of $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|$ in Lemma (A.3.2), so it remains to quantify $\sigma_{\max}\left[\mathbf{Q}^{\Delta}\right]$. Note that we can further decompose $\mathbf{Q}^{\Delta}$ as follows

$$\mathbf{Q}^{\Delta} = \tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*) + \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*) - \mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] + \mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] - \mathbf{Q}_L.$$

By the property of the singular value, we readily have that

$$\begin{aligned} \sigma_{\max}\left[\mathbf{Q}^{\Delta}\right] \leq \sigma_{\max}\left[\tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] + \sigma_{\max}\left\{\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*) - \mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]\right\} \\ + \sigma_{\max}\left\{\mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] - \mathbf{Q}_L\right\}, \end{aligned} \tag{B.29}$$

which will be studied one by one.

By definition of $\tilde{\boldsymbol{\theta}}^*$ and Lemma A.3.2, $\sup_{0<r\le R}\left|(\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^T\boldsymbol{\phi}_L(r)\right| \le \|\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*\|\sup_{0<r\le R}\|\boldsymbol{\phi}_L(r)\| = \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\|O(L^{\nu_2}) = O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right) = o_p(1)$. Then, it is straightforward to see that

$$\begin{aligned}|\tilde{g}_L(r;\boldsymbol{\theta}^*) - \tilde{g}_L(r;\tilde{\boldsymbol{\theta}}^*)| &= \tilde{g}_L(r;\boldsymbol{\theta}^*)\left|1 - \exp\left[(\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^T\boldsymbol{\phi}_L(r)\right]\right| \\ &= \tilde{g}_L(r;\boldsymbol{\theta}^*)O(1)\left|(\tilde{\boldsymbol{\theta}}^* - \boldsymbol{\theta}^*)^T\boldsymbol{\phi}_L(r)\right| = \tilde{g}_L(r;\boldsymbol{\theta}^*)O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right),\end{aligned}$$

which further implies that

$$\begin{aligned}&\eta_{\max}\left[\tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] = \sup_{\|\boldsymbol{\delta}\|=1}\boldsymbol{\delta}^T\left[\tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]\boldsymbol{\delta} \\ &\le \sup_{\|\boldsymbol{\delta}\|=1}\sum\sum_{i\ne j}\sum\sum_{u\in X_i, v\in X_j}\frac{w_R(|u-v|)}{m(m-1)}[\boldsymbol{\delta}^T\boldsymbol{\phi}_L(|u-v|)]^2|\tilde{g}_L(|u-v|;\tilde{\boldsymbol{\theta}}^*) - \tilde{g}_L(|u-v|;\boldsymbol{\theta}^*)| \\ &= O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right)\sup_{\|\boldsymbol{\delta}\|=1}\sum\sum_{i\ne j}\sum\sum_{u\in X_i, v\in X_j}\frac{w_R(|u-v|)}{m(m-1)}[\boldsymbol{\delta}^T\boldsymbol{\phi}_L(|u-v|)]^2\tilde{g}_L(|u-v|;\boldsymbol{\theta}^*) \\ &= O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right)\eta_{\max}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right].\end{aligned}$$

Following exactly the same steps, we can also show that

$$-\eta_{\min}\left[\tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] = \eta_{\max}\left[-\tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) + \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] = \eta_{\max}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right),$$

which implies that

$$\sigma_{\max}\left[\tilde{\mathbf{H}}_L(\tilde{\boldsymbol{\theta}}^*) - \tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right] = \eta_{\max}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}\right), \tag{B.30}$$

where the convergence is entry-wise.

The next step is to quantify the magnitude of $\sigma_{\max}\left\{\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*) - \mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]\right\}$. Using the standard random matrix theory, it suffice to consider the variability of $\boldsymbol{\delta}^T\left\{\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*) - \mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]\right\}\boldsymbol{\delta}$ for any $\|\boldsymbol{\delta}\| = 1$. Following steps as those in the proof of Lemma A.3.2 about Var $\left[H''_{m,n}(z_0)\right]$, we immediately have that

$$\sup_{\|\boldsymbol{\delta}\|=1}\left|\boldsymbol{\delta}^T\left\{\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)-\mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]\right\}\boldsymbol{\delta}\right| = O_p\left(\frac{L^{2\nu_2}}{m^2|D_n|}\right)+O\left(\frac{1}{m|D_n|}\right),$$

hence that

$$\sigma_{\max}\left\{\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)-\mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]\right\} = O_p\left(\frac{L^{2\nu_2}}{m^2|D_n|}\right)+O\left(\frac{1}{m|D_n|}\right). \tag{B.31}$$

Next, we proceed to bound the largest singular value of $\mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]-\mathbf{Q}_L$. For any $\|\boldsymbol{\delta}\|=1$,

$$\begin{aligned}
&\boldsymbol{\delta}^T\left\{\mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]-\mathbf{Q}_L\right\}\boldsymbol{\delta}\\
&=\int_{D_n^2}\lambda(u)\lambda(v)w_R(|u-v|)\left[\boldsymbol{\phi}_L^T(|u-v|)\boldsymbol{\delta}\right]^2\left[\tilde{g}_L(|u-v|;\boldsymbol{\theta}^*)-g(|u-v|)\right]\mathrm{d}u\mathrm{d}v\\
&=\int_{D_n^2}\lambda(u)\lambda(v)w_R(|u-v|)\left[\boldsymbol{\phi}_L^T(|u-v|)\boldsymbol{\delta}\right]^2 g(|u-v|)\\
&\qquad\qquad\times\left\{1-\exp\left[(\boldsymbol{\theta}^*-\boldsymbol{\theta}_0)^T\boldsymbol{\phi}(|u-v|)-\tilde{\zeta}_L(|u-v|;\boldsymbol{\theta}_0)\right]\right\}\mathrm{d}u\mathrm{d}v\\
&=O(1)\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}\right]^2\left[|(\boldsymbol{\theta}^*-\boldsymbol{\theta}_0)^T\boldsymbol{\phi}(s)|+|\tilde{\zeta}_L(s;\boldsymbol{\theta}_0)|\right]\mathrm{d}s\\
&=O(1)\|\boldsymbol{\theta}^*-\boldsymbol{\theta}_0\|\sup_{0<r\le R}\|\boldsymbol{\phi}(r)\|\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}\right]^2\mathrm{d}s\\
&\quad+O(1)\sqrt{\int_0^R w_o(s)\tilde{\zeta}_L^2(|u-v|;\boldsymbol{\theta}_0)\mathrm{d}s}\times\sqrt{\int_0^R w_o(s)\left[\boldsymbol{\phi}_L^T(s)\boldsymbol{\delta}\right]^2\mathrm{d}s}\\
&=O(L^{\nu_0+\nu_2-\nu_1})
\end{aligned}$$

where the last equality follows from condition E4 and Lemma A.3.1. This further gives that

$$\sigma_{\max}\left\{\mathbb{E}\left[\tilde{\mathbf{H}}_L(\boldsymbol{\theta}^*)\right]-\mathbf{Q}_L\right\} = O(L^{\nu_0+\nu_2-\nu_1}). \tag{B.32}$$

Combining equations (B.29)-(B.31), we have that where $\sigma_{\max}\left(\mathbf{Q}^{\Delta}\right) = O_p\left(\frac{L^{\nu_0+\nu_2}}{\sqrt{m|D_n|}}+L^{\nu_0+\nu_2-\nu_1}\right)$. In addition, we have shown in Lemma A.3.2 that $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\| = O_p\left(\frac{L^{\nu_0}}{\sqrt{m|D_n|}}\right)$. Plugging these two equations back to (B.28), we have that

$$\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*) = \boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*) + O_p\left(\frac{L^{2\nu_0+\nu_2}}{m|D_n|} + \frac{L^{\nu_0+\nu_2-\nu_1}}{\sqrt{m|D_n|}}\right)\left\|\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\right\|,$$

which gives (B.23), recall that $\nu_2 + 2\nu_0 < \nu_1$ in condition E6 and the condition $L^{4\nu_0+2\nu_2}/m|D_n| \to 0$.

To show (B.24), define vector $\boldsymbol{\ell}(r) = \left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L^T(r)$ and its standardized version $\boldsymbol{\ell}_0(r) = \|\boldsymbol{\ell}(r)\|^{-1}\boldsymbol{\ell}(r)$ as in condition N3. Then applying Lemma B.2.1 to $\boldsymbol{\ell}_0^T(r)\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)$, under condition N3, we have that

$$\frac{\sqrt{m|D_n|}\boldsymbol{\ell}_0^T(r)\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)}{\sqrt{\boldsymbol{\ell}_0^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}_0(r)}} = \frac{\sqrt{m|D_n|}\boldsymbol{\ell}^T(r)\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)}{\sqrt{\boldsymbol{\ell}^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}(r)}} \to^d N(0,1),$$

where $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right]$. Then using (B.23), we have that

$$\begin{aligned}\frac{\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*)}{\sqrt{\boldsymbol{\ell}^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}(r)}} &= \frac{\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*)}{\sqrt{\boldsymbol{\ell}^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}(r)}} + o_p(1)\frac{\|\boldsymbol{\ell}(r)\|}{\sqrt{\boldsymbol{\ell}^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}(r)}} \\ &= \frac{\sqrt{m|D_n|}\boldsymbol{\ell}^T(r)\tilde{\boldsymbol{U}}_L(\boldsymbol{\theta}^*)}{\sqrt{\boldsymbol{\ell}^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}(r)}} + o_p(1)\frac{1}{\sqrt{\boldsymbol{\ell}_0^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}_0(r)}} \to^d N(0,1),\end{aligned}$$

where the last equality follows from condition Q3(a), which requires that $\sqrt{\boldsymbol{\ell}_0^T(r)\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\boldsymbol{\ell}_0(r)} \geq c_u$. The proof is complete. □

B.2.3 Theorem. *Under conditions E1-E6, Q1-Q3, as $L \to \infty$ and $m|D_n| \to \infty$, we have that*

$$\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}}) - g(r) + b_{n,L}\right]}{g(r)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}} \to^D N(0,1),$$

where $b_{n,L} = O\left(L^{-\nu_1+\max\{\tau_1,\nu_0+\nu_2\}}\right)$, $\mathbf{Q}_L(r)$ is defined in (6.17) and the specific form of $\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*) = \mathrm{Var}\left[\sqrt{m|D_n|}\tilde{\boldsymbol{U}}(\boldsymbol{\theta}^*)\right]$ is given in the following paragraph.

Proof. Recall the definition $\tilde{g}_L(r;\boldsymbol{\theta}) = \exp\left[\boldsymbol{\theta}^T\boldsymbol{\phi}_L(r)\right]$, then applying delta method to the asymptotic distribution of $\sqrt{m|D_n|}\boldsymbol{\phi}_L^T(r)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*)$ from B.2.2, we have that

$$\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}}) - \tilde{g}_L(r;\boldsymbol{\theta}^*)\right]}{\tilde{g}_L(r;\boldsymbol{\theta}^*)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}} \to N(0,1).$$

By equation (A.19) in Lemma A.3.1, we have that $\sup_{0<r<R}|g(r) - \tilde{g}_L(r;\boldsymbol{\theta}^*)| = O(L^{-\nu_1+\tau_1} + L^{\nu_0-\nu_1+\nu_2}) = o(1)$, it readily follows that

$$\begin{aligned}\frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}}) - \tilde{g}_L(r;\boldsymbol{\theta}^*)\right]}{\tilde{g}_L(r;\boldsymbol{\theta}^*)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}} &= \frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}}) - g(r) + g(r) - \tilde{g}_L(r;\boldsymbol{\theta}^*)\right]}{\tilde{g}_L(r;\boldsymbol{\theta}^*)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}} \\ &= \frac{\sqrt{m|D_n|}\left[\tilde{g}_L(r;\widehat{\boldsymbol{\theta}}) - g(r) + O(L^{-\nu_1+\tau_1} + L^{\nu_0-\nu_1+\nu_2})\right]}{g(r)\sqrt{\boldsymbol{\phi}_L^T(r)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\Sigma}_U(\boldsymbol{\theta}^*)\left(\mathbf{Q}_L\right)^{-1}\boldsymbol{\phi}_L(r)}} + o_p(1) \to N(0,1),\end{aligned}$$

which completes the proof. And we have

$$
\begin{aligned}
\mathbf{\Sigma}_U(\boldsymbol{\theta}^*) = |D_n| \int_{D_n^4} & \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2) w_R(|u_1 - v_1|) w_R(|u_2 - v_2|) \big[g^{(4)}(u_1, v_1, u_2, v_2) \\
& - g(|u_1 - v_1|) g(|u_2 - v_2|) \big] \times \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_2 - v_2|) \mathrm{d}u_1 \mathrm{d}v_1 \mathrm{d}u_2 \mathrm{d}v_2 \\
- 4|D_n| \int_{D_n^4} & \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2) w_R(|u_1 - v_1|) w_R(|u_2 - v_2|) [g^{(3)}(u_1, v_1, u_2) - g(|u_1 - v_1|)] \\
& \times \tilde{g}_L(|u_2 - v_2|; \boldsymbol{\theta}^*) \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_2 - v_2|) \mathrm{d}u_1 \mathrm{d}v_1 \mathrm{d}u_2 \mathrm{d}v_2 \\
+ \frac{2|D_n|}{m-1} \int_{D_n^4} & \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2) w_R(|u_1 - v_1|) w_R(|u_2 - v_2|) \tilde{g}_L(|u_1 - v_1|; \boldsymbol{\theta}^*) \tilde{g}_L(|u_2 - v_2|; \boldsymbol{\theta}^*) \\
& \times [g(|u_1 - u_2|) g(|v_1 - v_2|) - 1] \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_2 - v_2|) \mathrm{d}u_1 \mathrm{d}v_1 \mathrm{d}u_2 \mathrm{d}v_2 \\
+ \frac{4(m-2)|D_n|}{m-1} \int_{D_n^4} & \lambda(u_1)\lambda(v_1)\lambda(u_2)\lambda(v_2) w_R(|u_1 - v_1|) w_R(|u_2 - v_2|) [g(|u_1 - u_2|) - 1] \\
& \times \tilde{g}_L(|u_1 - v_1|; \boldsymbol{\theta}^*) \tilde{g}_L(|u_2 - v_2|; \boldsymbol{\theta}^*) \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_2 - v_2|) du_1 du_2 dv_1 dv_2 \\
+ 4|D_n| \int_{D_n^3} & \lambda(u_1)\lambda(v_1)\lambda(u_2) w_R(|u_1 - v_1|) w_R(|u_1 - u_2|) g^{(3)}(u_1, v_1, u_2) \\
& \times \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_1 - u_2|) \mathrm{d}u_1 \mathrm{d}v_1 \mathrm{d}u_2 \\
- 4|D_n| \int_{D_n^3} & \lambda(u_1)\lambda(v_1)\lambda(u_2) w_R(|u_1 - v_1|) w_R(|u_1 - u_2|) [2g(|u_1 - v_1|) - \tilde{g}_L(|u_1 - v_1|; \boldsymbol{\theta}^*)] \\
& \times \tilde{g}_L(|u_1 - u_2|; \boldsymbol{\theta}^*) \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_1 - u_2|) \mathrm{d}u_1 \mathrm{d}v_1 \mathrm{d}u_2 \\
+ \frac{4|D_n|}{m-1} \int_{D_n^3} & \lambda(u_1)\lambda(v_1)\lambda(u_2) w_R(|u_1 - v_1|) w_R(|u_1 - u_2|) \tilde{g}_L(|u_1 - v_1|; \boldsymbol{\theta}^*) \\
& \times \tilde{g}_L(|u_1 - u_2|; \boldsymbol{\theta}^*) [g(|v_1 - u_2|) - 1] \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_1 - u_2|) \mathrm{d}u_1 \mathrm{d}v_1 \mathrm{d}u_2 \\
+ 2|D_n| \int_{D_n^2} & \lambda(u_1)\lambda(v_1) \left[w_R(|u_1 - v_1|) \right]^2 g(|u_1 - v_1|) \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_1 - v_1|) \mathrm{d}u_1 \mathrm{d}v_1 \\
+ \frac{2|D_n|}{m-1} \int_{D_n^2} & \lambda(u_1)\lambda(v_1) w_R^2(|u_1 - v_1|) \tilde{g}_L^2(|u_1 - v_1|; \boldsymbol{\theta}^*) \boldsymbol{\phi}_L(|u_1 - v_1|) \boldsymbol{\phi}_L^T(|u_1 - v_1|) \mathrm{d}u_1 \mathrm{d}v_1
\end{aligned}
$$

It has been shown that $\eta_{\max} [\mathbf{\Sigma}_U(\boldsymbol{\theta}^*)] = O(1)$. □