5-2023

Efficient estimation of generalized nonparametric model under additive structure

Ying XIA
Singapore Management University, ying.xia.2017@phdecons.smu.edu.sg

Efficient Estimation of Generalized Nonparametric Model

under Additive Structure

Ying Xia

A DISSERTATION

In

ECONOMICS

Presented to the Singapore Management University in Partial Fulfilment

of the Requirements for the Degree of PhD in Economics

2023

____

Supervisor of Dissertation

____

PhD in Economics, Programme Director

EFFICIENT ESTIMATION OF GENERALIZED NONPARAMETRIC MODEL UNDER ADDITIVE STRUCTURE

by

Ying Xia

Submitted to School of Economics in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Economics

Dissertation Committee:

Liangjun Su (Supervisor/Chair)
C.V. Starr Chair Professor of Economics
Tsinghua University

Yichong Zhang(Co-Supervisor/Chair)
Associate Professor of Economics
Singapore Management University

Peter C.B. PHILLIPS
Distinguished Term Professor of Economics
Singapore Management University

Xun Lu
Associate Professor of Economics
Chinese University of Hong Kong

Singapore Management University
2023

ABSTRACT

In this thesis, we develop novel nonparametric estimation techniques for two distinct classes of models: (1) Generalized Additive Models with Unknown Link Functions (GAMULF) and (2) Generalized Panel Data Transformation Models with Fixed Effects. Both models avoid parametric assumptions on their respective link or transformation functions, as well as the distribution of the idiosyncratic error terms.

The first chapter aims to provide an in-depth and systematic introduction to cross-sectional and panel-data nonparametric transformation models, encompassing practical applications, a diverse range of estimation techniques, and the study of asymptotic properties. We discuss the advantages and limitations of these models and estimation methods, delving into the latest advancements and innovations in the field. Furthermore, we propose a potential approach to mitigate the curse of dimensionality in the context of fully nonparametric transformation models with fixed effects in panel-data settings.

The second chapter proposes a three-stage nonparametric least squares (NPLS) estimation procedure for the additive functions in the GAMULF. In the first stage, we estimate conditional expectation by the local-linear kernel regression and then apply matching method to the splines series to obtain initial estimators. In the second stage, we use the local-polynomial kernel regression to estimate the link function. In the third stage, given the estimators in Stages 1 and 2, we apply the local-linear kernel regression to refine the initial estimator. The great advantage of such a procedure is that the estimators obtained at all stages have closed-form expressions, which overcomes the computational hurdle for existing estimators of the GAMULF model.

The third chapter proposes a multiple-stage Local Maximum Likelihood Estimator (LMLE)

for the structural functions in the generalized panel data transformation model with fixed effects. In the first stage, we apply the regularized logistic sieve method to estimate the sieve coefficients associated with the approximation of a composite function and then apply a matching method to obtain initial consistent estimators of the additive structural functions. In the second stage, we apply the local polynomial method to estimate certain composite function and its derivatives to be used later on. In the third stage we apply the local linear method to obtain the refined estimator of the additive structural functions based on the estimators obtained in Steps 1 and 2. The greatest advantage is that all minimization problems are convex and thus overcome the computational hurdle for existing approaches to the generalized panel data transformation model.

The final estimates of the additive terms in two models achieve the optimal one-dimensional convergence rate, asymptotic normality and oracle efficiency. The Monte Carlo simulations demonstrate that our new estimator performs well in finite samples.

The thesis demonstrates the effectiveness of the proposed nonparametric estimation techniques in addressing the complexities of generalized additive models with unknown link functions and panel data transformation models with fixed effects.

ACKNOWLEDGMENT

First and foremost, I wish to extend my heartfelt gratitude towards my supervisors, Professor Liangjun Su and Yichong Zhang. Their expert guidance and constructive feedback were instrumental throughout my Ph.D. journey. Their commitment of time, thoughtful insights, and diligent mentorship have undeniably enhanced the productivity and richness of my academic experience. Their fervor for avant-garde research has been both contagious and inspirational. It has indeed been a profound privilege and honor to study under their aegis.

My deepest appreciation is also owed to Associate Professor Anthony Tay and Nicolas L. Jacquet, the esteemed directors of the Ph.D. program, for their invaluable advice and unwavering support. My sincere thanks go to Qiu Ling Thor and Amelia Tan for their exemplary assistance with administrative tasks, ensuring the smooth progression of my studies.

I would like to extend special recognition to my committee members, Professor Peter C.B. Phillips, and Xun Lu. Their patience and assistance during the review of my dissertation have been indispensable. I am especially indebted to Professor Peter C.B. Phillips, who graciously supported my professional growth by writing a recommendation letter for my job applications.

My deepest gratitude extends to my beloved family and friends who have been my pillars of strength. I owe an immeasurable debt of thanks to my parents and grandparents for their unwavering understanding, motivation, companionship, and unconditional love. I am also grateful to my fellow Ph.D. peers at SMU: Shenxi Song, Robin Ng, and Yiren Wang for their camaraderie and support during our collective journey.

Last but certainly not least, my special thanks are reserved for my dearest friends in Singapore, Xichun Wang, and Jingyi Liao. Their ceaseless love and support have been my safe harbor. In their presence, I am reminded of my worth, and without them, I consider myself incomplete.

I dedicate my accomplishments to all these individuals, for their roles have been instrumental in my journey, both personally and academically.

DEDICATION

This dissertation is dedicated to my grandparents.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGMENT iv

DEDICATION vi

1 Exploring Nonparametric Transformation Models: A Comprehensive Review of Current Literature 1

1.1 Introduction 1

1.2 Cross-sectional Nonparametric Transformation Model 2

1.3 Panel-data Nonparametric Transformation Model 4

1.4 Conclusion 5

2 Efficient Nonparametric Estimation of the Generalized Additive Model with an Unknown Link Function 7

2.1 Introduction 7

2.2 Setup and Estimation Procedure 10

2.3 Asymptotic Propperties 13

2.3.1 Assumptions 13

2.4 Numerical Studies 20

2.4.1 Data Generating Processes 20

2.4.2 The Estimation Methods for Comparison 21

2.4.3 Choices of Tuning Parameters 23

2.4.4 Results 24

2.5 Application: Why Do Small and Medium Enterprises(SMEs) Demand Property Liability Insurance 25

2.6 Conclusion 30

3 Efficient Nonparametric Estimation of Generalized Panel Data Transformation Models with Fixed Effects 31

3.1 Introduction 31

3.2 Methodology 34

3.2.1 The Model 34

3.2.2 Estimation Procedure 35

3.3 Assumptions and Asymptotic Results 43

3.3.1 Assumptions 44

3.3.2 Asymptotic Properties 48

3.4 Numerical Studies 50

3.5 Application: the Effect of Income Shock on Job Creation 52

3.6 Conclusion 56

APPENDIX

A Appendix for Chapter 2 64

A.1 Proof of Theorem 2.3.1 64

A.2 Proof of Theorem 2.3.2 68

A.3 Proof of Theorem 2.3.3 71

A.4 Technical Lemmas 74

A.4.1 Lemmas for the proof of Theorem 2.3.1 74

A.4.2 Lemmas for the proof of Theorem 2.3.2 81

A.4.3 Lemmas for the proof of Theorem 2.3.3 93

B Appendix for Chapter 3 98

B.1 Proofs of the Main Results 98

B.1.1 Proof of Theorem 3.3.1 98

B.1.2 Proof of Theorem 3.3.2 103

B.1.3 Proof of Theorem 3.3.3 106

B.2 Technical Lemmas 114

Chapter One

Exploring Nonparametric Transformation Models: A Comprehensive Review of Current Literature

Authors: Ying Xia

1.1 Introduction

Asymptotic properties of different forms of transformation models have also attracted significant interest. Horowitz (1996) focuses on a transformation model with a nonparametric transformation function and a parametric structural function. Chiappori, Komunjer and Kristensen (2015) extend Horowitz's method to a transformation model with both nonparametric transformation and structural functions, considering endogeneity. Although fully nonparametric transformation models avoid misspecification, they suffer from the curse of dimensionality.

Inspired by the additive structure in generalized additive models with unknown link functions (Horowitz (2001), Horowitz and Mammen (2007), Horowitz and Mammen (2011), Lin, Pan, Lv and Zhang (2018)), Chen, Lu and Wang (2022) propose a fully nonparametric transformation model with fully additive structural functions. Additionally, they account for unobserved individual heterogeneity, specifically fixed effects in panel data. Prior to Chen et al. (2022), Horowitz and Lee (2004), Chen (2010), and Wang and Chen (2020) also analyzed panel transformation models, but all assumed that the structural function was parametric.

In summary, the transformation model has been widely studied and applied in various fields of economics, with researchers continually developing and extending its properties and applications. The exploration of fully nonparametric transformation models with additive structural functions, as well as accounting for unobserved individual heterogeneity, has expanded the model's capabilities, although challenges such as the curse of dimensionality still persist.

This chapter is organized as follows: Section 1 investigates the cross-sectional nonparametric and semiparametric transformation models, providing an overview of their estimations and asymptotic properties. Section 2 explores the panel-data nonparametric transformation models, delving into their advantages, challenges, and promising future topics in this area of research. Finally, Section 3 summarizes the key points discussed in the chapter.

1.2 Cross-sectional Nonparametric Transformation Model

Horowitz (1996) examines the subsequent model:

$$\Lambda(y) = x'\beta + u, \tag{1.2.1}$$

where $\Lambda(\cdot)$ represents an increasing and invertible function. To estimate β in Equation (1.2.1) without knowledge of $H(\cdot)$, consider the following relationship:

$$m(x) = E(Y|\,x) = E\left[\Lambda^{-1}(x'\beta + u)\right] = H(x'\beta). \qquad (1.2.2)$$

Here, Härdle and Stoker (1989) proposes the Average Derivatives Method for estimating β in Equation (1.2.2). Given the estimator of β, Horowitz (1996) proposes using the Cumulative Distribution Function (CDF) and Probability Density Function (PDF) of Y given $x'\beta$ to estimate $\Lambda(\cdot)$. The estimators of both β and $\Lambda(\cdot)$ achieve $\sqrt{n}$ convergence rates and mean-zero normal distributions asymptotically.

Chiappori et al. (2015) builds upon the method proposed by Horowitz (1996) to estimate $\Lambda(\cdot)$ in the absence of knowledge about $g(\cdot)$ in the fully nonparametric model with endogeneity:

$$\Lambda(Y) = g(x) + u, \qquad (1.2.3)$$

where $g(\cdot)$ is an unknown fully nonparametric function. For identification, Chiappori et al. (2015) assumes that there exist continuous exogenous independent variables x_I and endogenous independent variables x_{-I} such that $u \perp x_I|\,x_{-I}$. The estimator of $\Lambda(\cdot)$ in Chiappori et al. (2015) achieves $\sqrt{n}$ convergence rates and a mean-zero Gaussian Process.

Thus, regardless of whether the functional form is $x'\beta$ or $g(\cdot)$, and irrespective of the presence or absence of endogeneity, we can consistently estimate the nonparametric transformation function with convergence rate $\sqrt{n}$ using the approach outlined in Chiappori et al. (2015). Since $\sqrt{n}$ is much faster than the nonparametric convergence rate, we can replace $\Lambda(\cdot)$ with its estimator and treat it as known when estimating nonparametric function $g(\cdot)$. However, the estimation of $g(\cdot)$ still suffers from the curse of dimensionality. To avoid the curse of dimensionality, one can replace $g(x)$ with $G(x'\beta)$ and consider the model,

$$\Lambda(y) = G(x'\beta) + u. \qquad (1.2.4)$$

And then we have

$$m_1(x) = E(Y|\, x) = E\left\{\Lambda^{-1}\left[G(x'\beta) + u\right]\right\} = H_1(x'\beta), \tag{1.2.5}$$

which shares similarities with equation (1.2.2) and we can employ a comparable method in model (1.2.1) to estimate β in model (1.2.4).

An alternative approach to avoid the curse of dimensionality is to assume that the structure function $g(x)$ is additive,

$$g(x) = \sum_{l=1}^{d_x} g(x_l),$$

and then we have the model,

$$\Lambda(y) = \sum_{l=1}^{d_x} g_l(x_l) + u, \tag{1.2.6}$$

which is a fully nonparametric transformation model under additivity. To estimate model (1.2.6), since Chiappori et al. (2015) already provides the estimator of $\Lambda(\cdot)$ with convergence rate $\sqrt{n}$ without knowing $\{g_l(\cdot)\}_{l=1}^{d_x}$, we can employ the methods presented in Lin et al. (2018) and Horowitz and Mammen (2011) to estimate $\{g_l(\cdot)\}_{l=1}^{d_x}$.

In summary, oracle-efficient and optimal-convergence-rate estimators for the cross-sectional nonparametric and semiparametric transformation models are consistently available.

1.3 Panel-data Nonparametric Transformation Model

Recently, Chen et al. (2022) considered the fully nonparametric transformation model (1.3.1) under the panel data structure with individual fixed effects,

$$\Lambda(y_{it}) = g(x_{it}) + \alpha_i + u_{it}, \tag{1.3.1}$$

where α_i represents the individual fixed effect. To eliminate the unknown nonparametric transformation function and fixed effect, Chen et al. (2022) generated a new variable by comparing the dependent variable y of the same individual across different periods. The estimator of the structural function proposed by Chen et al. (2022) has a closed-form expression, which makes it easy to implement and study its asymptotic normality.

To be more specific, Chen et al. (2022) considered the following model,

$$\Lambda(y_{it}) = \sum_{l=1}^{d_x} g_l(x_{l,it}) + \alpha_i + u_{it}, \tag{1.3.2}$$

which assumes the additive structure of $g(\cdot)$. However, the estimation is done through matching with other covariates locally, and thus suffers substantially from the curse of dimensionality.

To overcome the curse of dimensionality, we propose a promising method to estimate $\{g_l(\cdot)\}_{l=1}^{d_x}$ in model (1.3.2). In the first step, we can employ the method in Chen et al. (2022) to eliminate the nonparametric transformation function and fixed effect. Subsequently, to achieve oracle efficiency, optimal convergence rate, and asymptotic normal distribution, we can extend the methods in Horowitz and Mammen (2011) and Lin et al. (2018) to estimate $\{g_l(\cdot)\}_{l=1}^{d_x}$ in model (1.3.2). This approach can serve as a future topic for discussion and exploration in the field of panel-data nonparametric transformation models.

1.4 Conclusion

This chapter provides a summary of the literature on both cross-sectional and panel-data nonparametric transformation models, highlighting their key developments, properties, and applications in various fields of economics. Additionally, the chapter presents a potential approach to overcome the curse of dimensionality when dealing with fully nonparametric transformation models that include fixed effects in a panel-data setting. By addressing this challenge, researchers can further enhance the capabilities and applicability of nonparametric transformation models in empirical work.

In chapter 2 and 3 in this thesis, we will develop novel nonparametric estimation techniques for two distinct classes of models: (1) Generalized Additive Models with Unknown Link Functions (GAMULF) and (2) Generalized Panel Data Transformation Models with Fixed Effects. Both models avoid parametric assumptions on their respective link or transformation functions, as well as the distribution of the idiosyncratic error terms.

The second chapter proposes a three-stage nonparametric least squares (NPLS) estimation procedure for the additive functions in the GAMULF. In the first stage, we estimate conditional expectation by the local-linear kernel regression and then apply matching method to the splines series to obtain initial estimators. In the second stage, we use the local-polynomial kernel regression to estimate the link function. In the third stage, given the estimators in Stages 1 and 2, we apply the local-linear kernel regression to refine the initial estimator. The great advantage of such a procedure is that the estimators obtained at all stages have closed-form expressions, which overcomes the computational hurdle for existing estimators of the GAMULF model.

The third chapter proposes a multiple-stage Local Maximum Likelihood Estimator (LMLE) for the structural functions in the generalized panel data transformation model with fixed effects. In the first stage, we apply the regularized logistic sieve method to estimate the sieve coefficients associated with the approximation of a composite function and then apply a matching method to obtain initial consistent estimators of the additive structural functions. In the second stage, we apply the local polynomial method to estimate certain composite function and its derivatives to be used later on. In the third stage we apply the local linear method to obtain the refined estimator of the additive structural functions based on the estimators obtained in Steps 1 and 2. The greatest advantage is that all minimization problems are convex and thus overcome the computational hurdle for existing approaches to the generalized panel data transformation model.

Chapter Two

Efficient Nonparametric Estimation of the Generalized Additive Model with an Unknown Link Function

Authors: Ying Xia, Yichong Zhang, Liangjun Su

2.1 Introduction

Economics theories usually do not specify a detailed parametric functional form for the conditional expectation. However, in practice, estimation of conditional expectation function with multiple covariates fully nonparametrically suffer from the curse of dimensionality. Instead, researchers often use the nonparametric additive model which allows each covariate to enter the conditional mean function in a nonparametric but additive manner. Due to the additive structure, the estimator of conditional mean function has a one-dimensional nonparametric convergence rate, and thus, circumvent the curse of dimensionality. However, it also rules out the interaction between distinct covariates. Generalized additive model (GAM) further introduces an link function outside the additive structure, which adheres the fast convergence rate, yet allows for the marginal effect of one regressor to depend on all other regressors. Due to these advantages, the nonparametric additive model and GAM have been widely studied and applied. See, for example, Breiman and Friedman (1985), Stone (1985), Stone (1986), Buja, Hastie and Tibshirani (1989), Linton and Nielsen (1995), Linton and Härdle (1996), Opsomer and Ruppert (1997), Fan, Härdle and Mammen (1998), Mammen, Linton and Nielsen (1999), Linton (2000), Opsomer (2000), Horowitz and Mammen (2004), Nielsen and Sperlich (2005), Mammen and Park (2006), Horowitz and Mammen (2007), Yu, Park and Mammen (2008), Horowitz and Mammen (2011), Liu, Yang and Härdle (2013), Hastie and Tibshirani (2017), Lin et al. (2018). Among them, Horowitz (2001) and Horowitz and Mammen (2007) consider the GAM with an unknown link function (GAMULF) and focus on the estimation and inference of the additive term within the link function up to location and scale normalizations.

Horowitz (2001) and Horowitz and Mammen (2007) propose estimators for the additive functions in GAMULF which do not achieve oracle efficiency. Here we say the estimator achieves Oracle efficiency if its asymptotic variance is the same as if other additive components and the link function are known. To achieve oracle efficiency, optimal convergence rate and asymptotic normality at the same time, Horowitz and Mammen (2011) and Lin et al. (2018) propose multi-stage estimation procedures for GAMULF. However, these multi-stage estimation procedures require non-linear and non-convex optimization or time-consuming iteration.

In this article, we propose a new multi-stage estimation procedure with close-form expression in each stage, which greatly reduces the computational burden in the estimation of GAMULF. In the first stage, we obtain the B-splines estimators of additive components. In the second stage, we obtain the kernel-based estimator of the link function. In the third stage, we estimate each of the additive component via a local linear regression with other unknown components and the link function replaced by their estimators obtained in the first and second stages, respectively. We then show the resulting estimator still enjoys the desired statistical properties as those for GAMULF proposed in the literature such as the rate-optimal convergence rate, asymptotic normality, and oracle efficiency.

The article is organized as follows. Section 2 describes our methodology. We present the asymptotic properties of our estimators in Section 3. Section 4 examines the finite sample performance of our estimators via Monte Carlo simulations. We apply our method to an empirical dataset in Section 5. Section 6 concludes. All the proofs of the main theorems are relegated to the appendix.

Notation. We denote $tr(\cdot)$ as the trace operator and $\equiv$ as "is defined as". For a real matrix A, let A', $\|A\| \left(\equiv [tr(AA')]^{1/2}\right)$, $\|A\|_{op} \left(\equiv \sqrt{\lambda_{max}(A'A)}\right)$, $\lambda_{max}(A)$, and $\lambda_{min}(A)$ be the transpose of A, the Frobenius norm, the operator norm, and the biggest and smallest eigenvalues of matrix A, respectively. For any function $f(\cdot)$ defined on the real line, let $\partial^k f(\cdot)$ be the k th-order derivatives, for $k = 1, 2, \cdots$. Let $\xrightarrow{D}$ and $\xrightarrow{P}$ be convergences in distribution and probability, respectively. A kernel function $H(\omega)$ is said to be of order a, if the following conditions are satisfied, (i) $\int H(\omega)\, d\omega = 1$, (ii) $\int H(\omega)\, \omega^s d\omega = 0$, for $s = 1, \cdots, a-1$ and (iii)$\int H(\omega)\, \omega^a d\omega \in (0, +\infty)$. For a vector $a = (a_1, \cdots, a_d)$, let (a) be a diagonal matrix with entries $a_1, \cdots, a_d$. For any scalar bandwidth h, denote $H_h(\cdot) = \frac{1}{h} H\left(\frac{\cdot}{h}\right)$ and $\partial^k H_{h_2}(x) = \frac{1}{h} \partial^k H(y)\Big|_{y=\frac{x}{h}}$, for $x \in \mathcal{X}$ and $k = 1, 2, \cdots$. To implement multi-dimensional nonparametric regression, for $i \neq j$, let $H_h(i,j) = \prod_{l=1}^{d_x} T_{h_l}(X_{l,i} - X_{l,j})$, where $h = (h_1, \cdots h_{d_x})'$ is a vector of bandwidths and $\{X_i\}_{i=1}^n$ are the covariates. For a one-dimensional kernel function $T(\cdot)$, we define $T_h(\cdot)$ and $T_h(i,j)$ in the same manner as $H_h(\cdot)$ and $H_h(i,j)$, respectively. Let $\{p_k(\cdot), k = 1, 2, \cdots\}$ be a sequence of one-dimensional B-splines basis functions. Let $K = K(n)$ be some integer such that $K(n) \to \infty$ as $n \to \infty$.Let $P^K(\cdot) \equiv [p_1(\cdot), \cdots, p_K(\cdot)]'$ be one-dimensional B-splines and then, for $x = (x_1, \cdots, x_{d_x})'$, let $P^K(x) \equiv \left[P^K(x_1)', \cdots, P^K(x_{d_x})'\right]'$. For $i \neq j$, let $\triangle X_{i,j} = X_i - X_j$, $\triangle P_{i,j}^K = P^K(X_i) - P^K(X_j)$, $\triangle P_{i,j}^{(1)} = p_1(X_{1,i}) - p_1(X_{1,j})$ and

$$\triangle P_{i,j}^{(-1)} = \Big(p_2(X_{1,i}) - p_2(X_{1,j}), \cdots, p_K(X_{1,i}) - p_K(X_{1,j}), \left(P^K(X_{2,i}) - P^K(X_{2,j})\right)',$$

$$\cdots, \left(P^K\left(X_{d_x,i}\right) - P^K\left(X_{d_x,j}\right)\right)'\Big)'.$$

2.2 Setup and Estimation Procedure

In this section, we consider GAMULF:

$$Y = F\left[g_1(X_1) + \cdots + g_{d_x}(X_{d_x})\right] + V, \tag{2.2.1}$$

where $X = (X_1, \ldots, X_{d_x})'$ is a $d_x \times 1$ vector of exogenous regressors, $F(\cdot)$ and $\{g_l(\cdot)\}_{l=1}^{d_x}$ are unknown and smooth functions of interest and V is an unobserved random variable satisfying $E\left(V|X\right) = 0$. Following Horowitz and Mammen's (2011) lead, we replace additive terms with B-splines and implement the nonparametric sieve estimation. B-splines have low multicollinearity and recursive formula for calculation,[1] which provide computational advantages, and thus, are widely used in practice. For ease of notation, we assume all coordinates of X_i are continuously distributed. If in an application some elements of X are discrete, the dimension d_x is interpreted as the dimension of the continuous covariates. All results in this section can then be extended in a conceptually straight-forward manner by using the continuous covariates only within samples that are homogeneous in discrete covariates. For $x \in \mathcal{X}$, let $g(x) = \sum_{l=1}^{d_x} g_l(x_l)$. For $i \neq j \in \{1, \cdots, n\}$, let $\triangle g_{i,j} = g\left(X_j\right) - g\left(X_i\right)$. For $i \in \{1, \cdots, n\}$, let $F_i = F[g(X_i)]$. Denote $f(\cdot)$ as the density function of X_i.

We approximate $g_l(\cdot)$ by $p^K(\cdot)'\beta^{x_l}$ for $l = 1, \cdots, d_x$. Further denote $\beta = (\beta^{x_1\prime}, \cdots, \beta^{x_{d_x}\prime})'$ as a $d_xK \times 1$ vector of unknown parameters to be estimated and $\{\beta^{x_l} = (\beta_1^{x_l}, \cdots, \beta_K^{x_l})'\}_{l=1}^{d_x}$ as $K \times 1$ vectors.

Given observations $\{Y_i, X_i\}_{i=1}^n$, we estimate $\{g_l(\cdot)\}_{l=1}^{d_x}$ and $F(\cdot)$ in (2.2.1) via the following algorithm.

[1] See Chapter 19 of Powell (1981) and Chapter 4 of Schumaker (2007) for more detail.

Algorithm 2.2.1 (Three-stage Estimation Procedure)

1. Undersmoothed estimation $\{\bar{g}_l(\cdot)\}_{l=1}^{d_x}$ of $\{g_l(\cdot)\}_{l=1}^{d_x}$ in model (2.2.1).

   (a) Initial estimation $\bar{E}(Y|X)$ of $E(Y|X)$ by the local-polynomial regression of order R.

   For $j \in \{1, \cdots, n\}$,

$$\begin{aligned}&\left(\bar{E}(Y|X_j), \partial\bar{E}(Y|X_j)', \cdots, (R!)^{-1}\partial^R\bar{E}(Y|X_j)'\right)\\&= \underset{\left(a,\{b_s\}_{s=1}^R\right)'}{\arg\min}\frac{1}{n-1}\sum_{i=1,\neq j}^{n}\left\{Y_i - a - \sum_{s=1}^{R} b_s \cdot (\triangle X_{i,j})^{\otimes s}\right\}^2 T_{h_1}(i,j),\end{aligned} \tag{2.2.2}$$

   where $h_1 = (h_{1,1}, h_{1,2}, \cdots, h_{1,d_x})$.

   (b) Initial estimation $\bar{\beta} = \left(1, \bar{\theta}'\right)'$ of β.

$$\bar{\theta} = \arg\min_{\theta}\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left\{\triangle P_{i,j}^{(1)} + \theta'\triangle P_{i,j}^{(-1)}\right\}^2 H_{h_2}\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i)\right]. \tag{2.2.3}$$

   which implies

$$\begin{aligned}\bar{\theta} =& -\left\{\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}H_{h_2}\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i)\right]\right\}^{-1}\\&\times\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\triangle P_{i,j}^{(1)}\triangle P_{i,j}^{(-1)}H_{h_2}\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i)\right].\end{aligned} \tag{2.2.4}$$

   Then, define $\bar{g}_l(x_l) = P^K(x_l)'\bar{\beta}^{x_l}$, for $l = 1, \cdots, d_x$ and $\bar{g}(x) = \sum_{l=1}^{d_x}\bar{g}_l(x_l)$.

2. Oversmoothed estimation $\left\{\left(\widehat{F}_i, \widehat{\partial F}_i\right)\right\}_{i=1}^{n}$ of $\{(F_i, \partial F_i)\}_{i=1}^{n}$ by the local-polynomial regression of order S.

$$\begin{aligned}&\left(\widehat{F}_i, \widehat{\partial F}_i, \cdots, (S!)^{-1}\widehat{\partial^{h_3}F}_i\right)\\&= \underset{\left(a,b_1,\cdots,b_{h_3}\right)'}{\arg\min}\frac{1}{n}\sum_{j=1,\neq i}^{n}\left\{Y_j - a - \sum_{l=1}^{h_3}\left[\triangle\bar{g}_{i,j}\right]^l b_l\right\}^2 T_{h_3}(\triangle\bar{g}_{i,j}).\end{aligned}$$

The corresponding closed-form solution is

$$\left(\widehat{F}_i, \widehat{\partial F}_i, \cdots, (S!)^{-1}\widehat{\partial^{h_3} F}_i\right) = \left\{\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix} 1 \\ \triangle\bar{g}_{i,j} \\ \vdots \\ [\triangle\bar{g}_{i,j}]^{h_3}\end{pmatrix}\left(1, \triangle\bar{g}_{i,j}, \cdots, [\triangle\bar{g}_{i,j}]^{h_3}\right)\right\}^{-1}$$

$$\times\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)Y_j\begin{pmatrix} 1 \\ \triangle\bar{g}_{i,j} \\ \vdots \\ [\triangle\bar{g}_{i,j}]^{h_3}\end{pmatrix}. \quad (2.2.5)$$

3. Rate-optimal kernel estimation $\left\{\left(\widehat{g}_l(\cdot), \widehat{\partial g}_l(\cdot)\right)\right\}_{l=1}^{d_x}$ of $\{(g_l(\cdot), \partial g_l(\cdot))\}_{l=1}^{d_x}$.

   For $l \in \{1, \cdots, d_x\}$,

$$\left(\widehat{g}_l(x_l), \widehat{\partial g}_l(x_l)\right)' = \arg\min_{(a,b)'}\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i - \widehat{F}_i + \widehat{\partial F}_i\bar{g}_l\left(X_{l,i}\right) - \widehat{\partial F}_i\left[a + b\left(X_{l,i} - x_l\right)\right]\right\}^2 T_{h_{4,l}}\left(X_{l,i} - x_l\right).$$

   The corresponding closed-form solution is

$$\begin{pmatrix}\widehat{g}_l(x_l) \\ \widehat{\partial g}_l(x_l)\end{pmatrix} = \left\{\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ X_{l,i} - x_l\end{pmatrix}\left(1, X_{l,i} - x_l\right)\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i} - x_l\right)\right\}^{-1}$$

$$\times\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ X_{l,i} - x_l\end{pmatrix}\widehat{\partial F}_i T_{h_{4,l}}\left(X_{l,i} - x_l\right)\left(Y_i - \widehat{F}_i + \widehat{\partial F}_i\bar{g}_l\left(X_{l,i}\right)\right). \quad (2.2.6)$$

Stages 1 and 2 in Algorithm (2.2.1) are standard nonparametric estimations in the literature. The third stage is designed based on a Taylor's expansion of $g_l\left(X_{l,i}\right)$:

$$g_l\left(X_{l,i}\right) \approx g_l\left(x_l\right) + \partial g_l\left(x_l\right)\left(X_{l,i} - x_l\right),$$

which implies

$$\sum_{i=1}^{n}\left\{Y_i - F\left[\sum_{j=1,\neq l}^{d_x} g(X_j)\left(X_{j,i}\right) + g_l\left(x_l\right) + \partial g_l\left(x_l\right)\left(X_{l,i}-x_l\right)\right]\right\}^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)$$
$$\approx \sum_{i=1}^{n}\left\{Y_i - F_i + \partial F_i g_l\left(X_{l,i}\right) - \partial F_i\left[g_l\left(x_l\right) + \partial g_l\left(x_l\right)\left(X_{l,i}-x_l\right)\right]\right\}^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right).$$

This motivates our third-stage regression. We also note that all these three stages can be implemented without any forms of numerical optimization.

2.3 Asymptotic Propperties

In this section, we first present the assumptions and then study the asymptotic properties of the estimators of the structural functions.

2.3.1 Assumptions

To proceed, we introduce some notation. A real-valued m-times continuously differentiable function $q(u)$ on $\mathcal{U} \subset \mathbb{R}$ is said to be a γ-smooth function if, for some $r = \gamma - m \in (0, 1]$, $\exists c_q$, $|\partial^m q(u) - \partial^m q(u^*)| \leq c_q \left|u - u^*\right|^r$ holds for all $u, u^* \in \mathcal{U}$. It is well known that γ-smooth functions can be approximated well by various linear B-splines (e.g., Chen (2007)). So we will assume that $\{g_l(\cdot)\}_{l=1}^{d}$ are γ-smooth functions below.

We will use $\mathcal{X} = \otimes_{l=1}^{d} \mathcal{X}_l$ to denote the support of $X_{it} = (X_{1,i}, ...X_{d_x,i})'$. We make the following Assumptions.

Assumption 1 *(1) $\{Y_i, X_i\}_{i=1}^{n}$ are i.i.d.*

(2) The support $\mathcal{X}$ of X is compact.

(3) $\forall i \in \{1, \cdots, n\}$, $E(e_i|X_i) = 0$.

(4) There exist positive and finite constants $\underline{c}_e, \overline{c}_e$ and c_e such that $\underline{c}_e \leq E\left(e^2 \left| X = x\right.\right) \leq \overline{c}_e$ for all $x \in \mathcal{X}$ and $E\left|e\right|^j \leq c_e^{j-2} j! E\left(e^2\right) < \infty$ for all $j \geq 2$.

Assumption 1 imposes specific conditions on $\{Y_i, X_i, e_i\}$. Assumption 1(1) enforces that the observations are independent and identically distributed (i.i.d.); Assumptions 1(2) and (3) require the exogenous independent variables to have compact supports; while Assumption 1(4) establishes certain moment conditions on the error terms, thus simplifying the derivation.

The subsequent assumption pertains to the properties of the additive functions $g_l(\cdot)_{l=1}^{d_x}$.

Assumption 2 *(1) The function $g_l(\cdot)$ is bounded and γ-smooth function with $\gamma \geq 2$ within its support for $l \in \{1, \cdots, d_x\}$.*[2]

(2) There exist constant $C_\beta \in [0, \infty)$ and vectors $\beta_0 = (1, \theta_0) \in interior\left(B^K\right)$, such that, for $l \in \{1, \cdots, d_x\}$, $\sup_{x_l \in \mathcal{X}_l} \left|g_l(x_l) - P^K(x_l)'\beta_{l,0}\right| = O(K^{-\gamma})$.

(3) The set of basis functions, $\left\{p^k(\cdot)\right\}_{k=1}^{\infty}$, are twice continuously differentiable everywhere on $\mathcal{X}$.

(4) $\max_{0 \leq s \leq r} \sup_{\omega \in \mathcal{X}} \left\|\partial^{h_3} P^K(\omega)\right\| \leq \varsigma_{rK}$.

Remark 1 *Assumption 2(1) postulates that all one-dimensional nonparametric functions exhibit sufficient smoothness. Assumptions 2(2)-(4) specify the approximation error for γ-smooth functions, with polynomials, splines, wavelets, and certain other basis functions satisfying these conditions. For the sake of scale normalization, the first element of β_0 is set to* 1.

The next assumption concerns the properties of the unknown link function $F(\cdot)$.

Assumption 3 *(1) Let $\Omega = \left\{\sum_{l=1}^{d_x} g_l(x_l) : \text{ for } l \in \{1, \cdots, d_x\}, x_l \in \mathcal{X}_l\right\}$. Then, $\exists \underline{c}_2$, $\overline{c}_2$, $\underline{c}_3$ and $\overline{c}_3$, such that $\forall \omega \in \Omega$,*

$$-\infty < \underline{c}_2 \leq F(\omega) \leq \overline{c}_2 < \infty \quad \text{and} \quad 0 < \underline{c}_3 \leq \partial F(\omega) \leq \overline{c}_3 < \infty.$$

[2] A real-valued and m-times continuously differentiable function $q(o)$ on $\mathcal{O} \subset R$ is said to be a γ-smooth function if, for $r = \gamma - m \in (0, 1]$, $\exists c_q$, $|\partial^m q(o) - \partial^m q(\widetilde{o})| \leq c_q |o - \widetilde{o}|^r$ holds $\forall o$ and $\widetilde{o} \in \mathcal{O}$.

(2) $F(\cdot)$ *is* γ_F*-smooth functions with* $\gamma_F \geq 2$.

Next, we state the assumptions on the kernel function.

Assumption 4 *(1) The kernel function* $T(\cdot)$ *is symmetric PDFs and its support is compact and within its supports, it is bounded and* γ*-smooth function with* $\gamma \geq 2$.

(2) $H(\cdot)$ *is a kernel function of order* a_H *and* a_H *is even.*

(3) The orders of polynomial regression, R *and* S*, are odd.*

Assumption 5 *(1) The probability density function (PDF) of* X_l, $f_{X_l}(\cdot)$, *is bounded and bounded away from zero within its support,* $\forall\ l \in \{1, \cdots, d_x\}$.

(2) the probability density function (PDF) of $g(X) = \sum_{l=1}^{d_x} g(X_l)$, $f_{g(X)}(\cdot)$, *and its derivatives ,* $\partial f_{g(X)}(\cdot)$ *and* $\partial^2 f_{g(X)}(\cdot)$, *are bounded and bounded away from zero within its support.*

(3) the probability density function (PDF) of $E(Y|X_j) - E(Y|X_i)$, $f_{E(Y|X_j)-E(Y|X_i)}(\cdot)$, *and its derivatives,* $\partial F f_{E(Y|X_j)-E(Y|X_i)}(\cdot)$ *and* $\partial^2 f_{E(Y|X_j)-E(Y|X_i)}(\cdot)$ *are bounded and bounded away from zero within support,* $\forall\ i \neq j \in \{1, \cdots, n\}$.

Remark 2 *Assumption 5 are standard in the literature of B-splines estimation and kernel estimation.*

Assumption 6 *(1) there exist positive constants* C_1 *and* C_2 *such that for* $i \neq j \in \{1, \cdots, n\}$,

$$
\begin{aligned}
0 < C_1 &\leq \lambda_{\min}\left(E\left[P^K(X_i)P^K(X_i)' \middle| g(X_i)\right]\right) \\
&\leq \lambda_{\max}\left(E\left[P^K(X_i)P^K(X_i)' \middle| g(X_i)\right]\right) \leq C_2 < \infty \\
0 < C_1 &\leq \lambda_{\min}\left(E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E(Y|X_j) = E(Y|X_i)\right]\right) \\
&\leq \lambda_{\max}\left(E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E(Y|X_j) = E(Y|X_i)\right]\right) \leq C_2 < \infty,
\end{aligned}
$$

$$
\begin{aligned}
0 < C_1 &\le \lambda_{\min}\left(\partial E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right]\right) \\
&\le \lambda_{\max}\left(\partial E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right]\right) \le C_2 < \infty, \\
0 < C_1 &\le \lambda_{\min}\left(\partial^2 E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right]\right) \\
&\le \lambda_{\max}\left(\partial^2 E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right]\right) \le C_2 < \infty, \\
0 < C_1 &\le \lambda_{\min}\left(E\left\{\left|\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right|\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right\}\right) \\
&\le \lambda_{\max}\left(E\left\{\left|\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right|\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right\}\right) \le C_2 < \infty,
\end{aligned}
$$

and

$$
\begin{aligned}
&\max_{i=1,\cdots,n}\lambda_{\max}\left(\partial E\left[P^K(X_i)P^K(X_i)'\right| g\left(X_i\right)\right]\right) \le C_2 < \infty, \\
&\max_{i=1,\cdots,n}\lambda_{\max}\left(\partial^2 E\left[P^K(X_i)P^K(X_i)'\right| g\left(X_i\right)\right]\right) \le C_2 < \infty,
\end{aligned}
$$

hold uniformly in $K = 1, 2, \cdots,$ *where*

$$
\begin{aligned}
&\partial^l E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right] \\
&= \left.\frac{\partial^l E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\right| E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right]}{\partial\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right]^l}\right|_{E(Y|X_j)-E(Y|X_i)=0},
\end{aligned}
$$

for l *= 1, 2 and the definitions of other terms are similar.*

(2) As $n \to \infty$, $K \to \infty$, $\{h_{1,l}\}_{l=1}^{d_x} \to 0$, $h_2 \to 0$, $h_3 \to 0$ *and* $\{h_{4,l}\}_{t=1}^{d_x} \to 0$.

(3) $\left(h_2^{-1}\delta_{h,n}\right)^2 = o_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H}\right)$, $\sqrt{K}h_2^{-1}\delta_{h,n} = o_p(1)$,

$\sum_{l=1}^{d_x} h_{1,l}^{R+1} = o_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H}\right)$,

and $Kh^4 = o(1)$, *where* $\delta_{h,n} = O_p\left(\sum_{l=1}^{d_x} h_{1,l}^{R+1} + \sqrt{\frac{\log n}{n\prod_{l=1}^{d_x} h_{1,l}}}\right)$.

(4) $h_3^{-1}\sqrt{K}\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H}\right) = o(1)$ *and* $\left(h_3^{S+1} + \sqrt{\frac{\log n}{nh_3}}\right)\sqrt{K} \le O(1)$.

(5) $K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}+h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}=o_p\left(\sqrt{\frac{\log n}{nh_{4,l}}}+h_{4,l}^2\right)$.

Remark 3 *Assumption 6 (1) ensures the existence and nonsingularity of the asymptotic covariance matrices for B-splines estimations, which is a standard condition in the B-splines estimation literature. Assumptions 6(2)-(5) guarantee that the asymptotic biases and variances of the first-stage and second-stage estimators are sufficiently small for achieving oracle efficiency in the third stage. Under standard bandwidth selection methods for local polynomial regression, Assumptions 6 (3), (4), and the condition* $K^{-\alpha}+h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}=o_p\left(\sqrt{\frac{\log n}{nh_{4,l}}}+h_{4,l}^2\right)$ *in (5) hold when* $R=3$, $S=3$, *and* α *is sufficiently large. Consequently, Assumption 6 (5) simplifies to* $\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}=o_p\left(\sqrt{\frac{\log n}{nh_{4,l}}}+h_{4,l}^2\right)$.

Theorem 2.3.1 *Suppose that Assumptions 1 - 4, 5 (1) and (3) and 6 (1) - (2) hold. Then*

1.

$$\begin{aligned}
&\bar{\theta}-\theta_0\\
&=-\left\{\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}H_{h_2}\left[E\left(Y\,|X_j\right)-E\left(Y\,|X_i\right)\right]\right\}^{-1}\\
&\times\left\{\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\triangle g_{i,j}\triangle P_{i,j}^{(-1)}H_{h_2}\left[E\left(Y\,|X_j\right)-E\left(Y\,|X_i\right)\right]\right.\\
&\left.-\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left(\triangle P_{i,j}^{K\prime}\beta_0^{x,z}-\triangle g_{i,j}\right)\triangle P_{i,j}^{(-1)}H_{h_2}\left[E\left(Y\,|X_j\right)-E\left(Y\,|X_i\right)\right]\right\}\\
&+R_{n,5},
\end{aligned} \tag{2.3.1}$$

where $\|R_{n,5}\|=\left(h_2^{-1}\delta_{h,n}\right)^2+O_p\left(\sum_{l=1}^{d_x}h_{1,l}^{R+1}+\sqrt{\frac{1}{n}}\right)$.

2.

$$\left\|\bar{\theta}-\theta_0\right\|=O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right),$$

$$\left\|\bar{\beta}-\beta_0\right\|=O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right).$$

3.

$$\max_{1\le l\le d_x}\sup_{x_l\in\mathcal{X}_l}\left|\bar{g}_l\left(x_l\right)-g_l\left(x_l\right)\right|=O_p\left[\sqrt{K}\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right)\right]. \tag{2.3.2}$$

Remark 4 *Given Assumption 6 (3), Theorem 2.3.1 (1) signifies the asymptotically negligible dominant effect of the estimation of $E(Y|X)$.*

Theorem 2.3.2 *Suppose that Assumptions 1 - 5, 6 (1), (2) and (3) hold. Then*

1. *There exists a positive constant C_1 such that*

$$\left|\widehat{F}_i-F_i\right|=C_1\left|\bar{g}(X_i)-g(X_i)\right|+O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}+h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}\right), \tag{2.3.3}$$

hold uniformly over $i\in\{1,\cdots,n\}$.

2. *There exist a positive constant C_2 such that*

$$\left|\widehat{\partial F}_i-\partial F_i\right|=h_3^{-1}C_2\left|\bar{g}(X_i)-g(X_i)\right|+h_3^{-1}O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}+h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}\right), \tag{2.3.4}$$

hold uniformly over $i\in\{1,\cdots,n\}$.

Withe Theorems (2.3.1) and (2.3.2), it is straigntforward to show the asymptotic joint distribution of our three-stage estimators of $\{(g_l(\cdot),\partial g_l(\cdot))\}_{l=1}^{d_x}$.

Theorem 2.3.3 *Suppose that Assumptions Assumptions 1 - 6 hold. Let $\kappa_{ab}=\int u^a\left[T(u)\right]^b du$ for $a,b=0,1,2$. Then for $l=1,\cdots,d$,*

1.

$$\begin{pmatrix}1 & 0\\ 0 & h_{4,l}\end{pmatrix}\left[\begin{pmatrix}\widehat{g}_l(x_l)\\ \widehat{\partial g}_l(x_l)\end{pmatrix}-\begin{pmatrix}g_l(x_l)\\ g_l(x_l)\end{pmatrix}\right]$$

$$= \left\{ \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \left(1, \frac{X_{l,i} - x_l}{h_{4,l}}\right) (\partial F_i)^2 T_{h_{4,l}} (X_{l,i} - x_l) \right\}^{-1}$$

$$\times \left\{ \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \partial F_i T_{h_{4,l}} (X_{l,i} - x_l) e_i + \frac{1}{2} \partial^2 g_l (x_l) \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \partial F_i^2 T_{h_{4,l}} (X_{l,i} - x_l) (X_{l,i} - x_l)^2 \right\}$$

$$+ R_{n,3} (x_l), \tag{2.3.5}$$

where $R_{n,3} (x_l) = o_p \left(\sqrt{\frac{1}{nh_{4,l}}} + h_{4,l}^2 \right) + O_p \left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H} + h_3^{S+1} + \sqrt{\frac{\log n}{nh_3}} \right)$ *holds for each* $x_l \in \mathcal{X}_l$ *and* $R_{n,3} (x_l) = o_p \left(\sqrt{\frac{\log n}{nh_{4,l}}} + h_{4,l}^2 \right) + O_p \left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H} + h_3^{S+1} + \sqrt{\frac{\log n}{nh_3}} \right)$ *holds uniformly over* $x_l \in \mathcal{X}_l$.

2.

$$\sqrt{nh_{4,l}} \begin{pmatrix} 1 & 0 \\ 0 & h_{4,l} \end{pmatrix} \left(\begin{pmatrix} \widehat{g}_l(x_l) \\ \widehat{\partial g}_l(x_l) \end{pmatrix} - \begin{pmatrix} g_l(x_l) \\ g_l(x_l) \end{pmatrix} - \begin{pmatrix} \frac{1}{2} h_{4,l}^2 \kappa_{21} \partial^2 g_l(x_l) \\ 0 \end{pmatrix} \right)$$

$$\xrightarrow{d} N \left(0, \frac{E\left\{ E\left(e^2 \middle| X\right) \partial F \left[g\left(X\right)\right]^2 \middle| X_l = x_l \right\}}{\left(E\left\{ \left[\partial F \left(g\left(X\right)\right)\right]^2 \middle| X_l = x_l \right\} \right)^2} \begin{pmatrix} \kappa_{02} & 0 \\ 0 & \frac{\kappa_{22}}{\kappa_{21}^2} \end{pmatrix} \right).$$

3.

$$\sup_{x_l \in \mathcal{X}_l} \|\widehat{g}_l(x_l) - g_l(x_l)\| = O_p \left(h_{4,l}^2 + \sqrt{\frac{\log n}{nh_{4,l}}} \right).$$

Remark 5 *The rate $O(n^{-1/5})$ is widely recognized as the asymptotically optimal bandwidth for one-dimensional kernel mean regression, specifically when the conditional mean function possesses continuous second derivatives.*

Remark 6 *Theorem 2.3.3 (1) furnishes linear representations of the nonparametric estimators $\left\{ \left(\widehat{g}_l(\cdot), \widehat{\partial g}_l(\cdot) \right) \right\} l = 1^{d_x}$, uniformly controlling the remainder terms. This theorem serves as a foundational component for both pointwise and uniform inference. According to Chernozhukov, Chetverikov and Kato (2014, Corollary 3.1), uniform inference based on the multiplier bootstrap is feasible.*

Remark 7 *Theorem 2.3.3 (2) elucidates the asymptotic properties of our three-stage estimators for* $\{(g_l(\cdot), \partial g_l(\cdot))\}\, l = 1^{d_x}$*. Notably, the asymptotic distribution of the local-linear estimator remains unaffected by random sampling errors emanating from the first two-stage estimators. Essentially, the three-stage estimator of* $(g_l(\cdot), \partial g_l(\cdot))$ *retains the same asymptotic distribution that would be expected if other additive components* $\{(g_{l_1}(\cdot), \partial g_{l_1}(\cdot))\}\, l_1 = 1, \neq l^{d_x}$ *and the link function* $F(\cdot)$ *were known, for* $\forall l \in \{1, \cdots, d_x\}$*.*

Remark 8 *Theorem 2.3.3 (3) establishes the uniform convergence rate for* $\{g_l(\cdot)\}_{l=1}^{d_x}$*. Furthermore, following standard practices in the nonparametric kernel literature, we can also affirm that the estimators of* $(g_{l_1}(\cdot), \partial g_{l_1}(\cdot))$ *and* $(g_{l_2}(\cdot), \partial g_{l_2}(\cdot))$ *for* $\forall l_1 \neq l_2 \in \{1, \cdots, d_x\}$ *are asymptotically independently distributed.*

2.4 Numerical Studies

2.4.1 Data Generating Processes

We use three data generating processes (DGPs) to study the finite sample performance of estimation methods proposed in this paper and the existing literature.

We consider the following settings.

Example 1 (*Continuous Case*) In model (2.2.1), we set the dimension as 2, and

$$g_1(x) = \sin(x), \quad g_2(x) = 2\Phi(x - 0.5), \quad F(x) = \Phi(x),$$

where $\Phi(\cdot)$ is the CDF of normal distribution. We generate $X_{i,1}$ and $X_{i,2}$, $i = 1, \cdots, n$, from the uniform distribution $U[-1, 1]$. Y_i , given $X_{i,1}$ and X_{i2}, is generated through $Y_i = F\left[g_1\left(X_{i,1}\right) + g_2\left(X_{i,2}\right)\right] + V_i$, where V_i is generated from the normal distribution $N(0, 0.02^2)$ and independent of $X_{i,1}$ and X_{i2}. The simulated results are in Table (3.1).

Example 2 (*Continuous Case with Four Components*) In model (2.2.1), we set the dimension as 4, and

$$g_1(x) = \sin(x), \quad g_2(x) = \Phi(x), \quad g_3(x) = \frac{1}{4}(x-5)^2, \quad g_4(x) = \cos\left(\frac{1}{2}x - \frac{3}{2}\right), \quad F(x) = \Phi(x),$$

where $\Phi(\cdot)$ is the CDF of normal distribution. We generate $X_{i,1}$, $X_{i,2}$, $X_{i,3}$ and $X_{i,4}$, $i = 1, \cdots, n$, from the uniform distribution $U[-1, 1]$. Y_i , given $X_{i,1}$, $X_{i,2}$, $X_{i,3}$ and $X_{i,4}$, is generated through $Y_i = F\left[g_1\left(X_{i,1}\right) + g_2\left(X_{i,2}\right) + g_3\left(X_{i,1}\right) + g_4\left(X_{i,2}\right)\right] + V_i$, where V_i is generated from the normal distribution $N(0, 3^2)$ and independent of $X_{i,1}$, $X_{i,2}$, $X_{i,3}$ and $X_{i,4}$. The simulated results are in Table (3.2).

Example 3 (*Binary Case*) In model (2.2.1), we set the dimension as 2, and

$$g_1(x) = \sin(x), \quad g_2(x) = \frac{1}{8}\left(x-5\right)^2 - 3, \quad F(x) = \frac{\exp(x)}{1+\exp(x)}.$$

We generate $X_{i,1}$ and $X_{i,2}$, $i = 1, \cdots, n$, from the uniform distribution $U[-1, 1]$. Y_i , given $X_{i,1}$ and X_{i2}, has the Bernoulli distribution $B(1, p_i)$ with

$$p_i = E\left(Y_i \middle| X_{i,1}, X_{i2}\right) = F\left[g_1(X_{i,1}) + g_2(X_{i,2})\right].$$

The link function $F(\cdot)$ is the commonly used logistic function for binary response. The simulated results are in Table (3.3).

2.4.2 The Estimation Methods for Comparison

In this section, we use simulated examples to demonstrate how well the proposed estimation procedure works. For continuous case, we compare the proposed estimation procedure with the method in Horowitz and Mammen (2011), denoted by HM. For binary case, we compare the proposed estimation procedure with the method in Horowitz and Mammen (2011) and Klein and Spady (1993), denoted by HM and KS.

We define the bias, standard deviation (SD), and root mean integrated squared error (RMISE) of an estimator $\hat{f}(\cdot)$ of $f(\cdot)$ as

$$\text{bias} = \int \left| \mathrm{E}\left[\hat{f}(v)\right] - f(v) \right| dv,$$

$$\text{SD} = \int sd\left[\hat{f}(v)\right] dv$$

and

$$RMISE = (\text{bias}^2 + \text{SD}^2)^{1/2},$$

respectively, and use them to assess the accuracy of the estimator $\hat{f}(\cdot)$.

In the continuous case, we implement both the proposed estimation procedure and the method from Horowitz and Mammen (2011) on the simulated data. Both methods utilize B-splines to acquire initial estimators, which are subsequently refined by one-dimensional kernels to enhance efficiency. The key distinction between the two techniques lies in the initial B-splines estimator. The initial estimator as per Horowitz and Mammen (2011) encounters challenges with non-convex optimization and tends to get trapped at local optima, which, despite its perfect theoretical properties, hampers its performance with simulated data. For the binary case, we compare our method with both methods presented in Horowitz and Mammen (2011) and Klein and Spady (1993). Klein and Spady (1993) estimates a single-index model with a binary response using the Maximum Likelihood Estimation (MLE) method. Given that the model $Y = F\left[P^K(X)'\beta\right] + \widetilde{e}$, where $\widetilde{e} = F\left[g_1(X_1) + \cdots + g_{d_x}(X_{d_x})\right] - F\left[P^K(X)'\beta\right] + V$, resembles the single-index model, and both the proposed method and the one in Klein and Spady (1993) employ kernel estimation to estimate the link function, we also present the simulation results using the method in Klein and Spady (1993) for the binary case. To maintain a fair comparison, we extend and refine the method in Klein and Spady (1993). Since Klein and Spady (1993) employs the MLE method and Horowitz and Mammen (2007) uses a method akin to weighted OLS, Klein and Spady (1993) should provide superior performance in the binary case.

2.4.3 Choices of Tuning Parameters

All the hyperparameters are tunned by cross validation based on grid search. The kernel function used in the proposed estimation procedure is the Epanechnikov kernel for all simulated examples in this section. For each simulated example, we assess the accuracy of the proposed estimation procedure for sample size $n = 400$ and for each case, we compute the bias, SD and RMISE of an obtained estimator based on 1000 simulations.

For Example 1 with proposed method, the bandwidths for multi-dimensional kernel estimation in stage 1(a) are $h_{1,1} = 2$ and $h_{1,2} = 2$, 5-interior-knot cubic B-splines are used for the first additive term and 3-interior-knot cubic B-splines are used for the second additive term in stage 1(b), the bandwidth for matching estimator in stage 1(b) is $h_2 = 0.001$, the bandwidth for link function in stage 2 is $h_3 = 2$ and the bandwidths for additive terms in stage 3 are $h_{4,1} = 0.2$ and $h_{4,2} = 0.15$. For Example 1 with method in Horowitz and Mammen (2011), the bandwidth for the initial estimator is optimized by cross-validation method in each replicate. The bandwidth for link function in stage 2 is $h_3 = 3$ and the bandwidths for additive terms in stage 3 are $h_{4,1} = 0.2$ and $h_{4,2} = 0.1$.

For Example 2 with proposed method, the bandwidths for multi-dimensional kernel estimation in stage 1(a) are $h_{1,1} = 10$, $h_{1,2} = 10$, $h_{1,3} = 2$ and $h_{1,4} = 10$. 3-interior-knot cubic B-splines are used for each additive term in stage 1(b), the bandwidth for matching estimator in stage 1(b) is $h_2 = 0.3$, the bandwidth for link function in stage 2 is $h_3 = 2$ and the bandwidths for additive terms in stage 3 are $h_{4,1} = 0.12$, $h_{4,2} = 0.14$, $h_{4,3} = 0.12$ and $h_{4,4} = 0.18$. For Example 2 with method in Horowitz and Mammen (2011), the bandwidth for the initial estimator is optimized by cross-validation method in each replicate. The bandwidth for link function in stage 2 is $h_3 = 2$ and the bandwidths for additive terms in stage 3 are $h_{4,1} = 0.1$, $h_{4,2} = 0.1$, $h_{4,3} = 0.1$ and $h_{4,4} = 0.1$.

For Example 3 with proposed method, the bandwidths for multi-dimensional kernel estimation in stage 1(a) are $h_{1,1} = 2$ and $h_{1,2} = 2$, 5-interior-knot cubic B-splines are used for the first additive term and 3-interior-knot cubic B-splines are used for the second additive term in stage 1(b), the bandwidth for matching estimator in stage 1(b) is $h_2 = 0.001$, the bandwidth for link function in stage 2 is $h_3 = 3$ and the bandwidths for additive terms in stage 3 are $h_{4,1} = 2$ and $h_{4,2} = 1$. For Example 3 with method in Horowitz and Mammen (2011), the bandwidth for the initial estimator is optimized by cross-validation method in each replicate. The bandwidth for link function in stage 2 is $h_3 = 0.6$ and the bandwidths for additive terms in stage 3 are $h_{4,1} = 2.5$ and $h_{4,2} = 1.5$. For Example 3 with method in Klein and Spady (1993), the bandwidth for the initial estimator is optimized by cross-validation method in each replicate. The bandwidth for link function in stage 2 is $h_3 = 3$ and the bandwidths for additive terms in stage 3 are $h_{4,1} = 2$ and $h_{4,2} = 1$.

2.4.4 Results

Table 2.1 The simulation results for Example 1 with continuous response

		$\hat{g}_1$	$\hat{g}_2$
Prop.	RMSE	0.0067	0.0081
	Bias	0.0018	0.0018
	SD	0.0064	0.0079
HM	RMSE	0.0668	0.0462
	Bias	0.0147	0.0093
	SD	0.0652	0.0453

In the methodologies proposed by Horowitz and Mammen (2011) as well as Klein and Spady (1993), optimizing the bandwidth for the initial estimator in each simulation proves to be time-consuming and impractical, especially when dealing with four additive components. Furthermore, the computation of the initial estimator involves non-convex and non-linear optimization, leading to results that may be local optima and vary significantly across

Table 2.2 The simulation results for Example 2 with four components

		$\hat{g}_1$	$\hat{g}_2$	$\hat{g}_3$	$\hat{g}_4$
Prop.	RMSE	0.0071	0.0057	0.0065	0.0055
	Bias	0.0022	0.0010	0.0012	0.0010
	SD	0.0067	0.0056	0.0064	0.0054
HM	RMSE	0.0456	0.0177	0.1375	0.0367
	Bias	0.0101	0.0050	0.0254	0.0060
	SD	0.0444	0.0170	0.1351	0.0362

simulations. Consequently, the computation of the initial estimator in the methods of Horowitz and Mammen (2011) and Klein and Spady (1993) is not only time-consuming but also unstable.

The biases, standard deviations (SDs), and root mean squared errors (RMISEs) of the function estimators $g_1(\cdot)$, $g_2(\cdot)$, $g_3(\cdot)$, $g_4(\cdot)$, and $F(\cdot)$, based on 1000 replicates, are presented in Tables (1) to (3).

A comparison presented in Tables (1) to (3) shows that the proposed estimator significantly outperforms those of Horowitz and Mammen (2011) and Klein and Spady (1993) in terms of Bias, SD, and RMISE across all three examples. Thus, it can be concluded that the proposed estimation procedure exhibits superior performance compared to the methods introduced in Horowitz and Mammen (2011) and Klein and Spady (1993).

2.5 Application: Why Do Small and Medium Enterprises(SMEs) Demand Property Liability Insurance

The purchase of insurance is a common practice for businesses of all sizes, industries, and locations. Property liability insurance, which protects against financial losses from property

Table 2.3 The simulation results for Example 3 with binary response

		$\hat{g}_1$	$\hat{g}_2$
Prop.	RMSE	0.0991	0.0804
	Bias	0.027	0.0359
	SD	0.0953	0.072
HM	RMSE	0.2474	0.1247
	Bias	0.1384	0.054
	SD	0.2051	0.1124
KS	RMSE	0.1497	0.0971
	Bias	0.0798	0.0666
	SD	0.1267	0.0707

damage or bodily injury caused by one's actions or negligence, is vital for both businesses and individuals. Studies have shown that a firm's ln(Asset), Credit Score, and the number of banks SME transacts with (Banks) are significantly related to insurance demand. While most research focuses on listed firms, few examine small and medium enterprises (SMEs) due to data limitations.

Yoshihiro (2019) conducted a survey to analyze the insurance demand of SMEs in Japan. The study's findings can be concisely summarized as follows: Firstly, SMEs with a higher risk of bankruptcy often demand less insurance. Teikoku Data Bank's financial statements provide credit scores, which indicate the likelihood of bankruptcy for SMEs. For instance, lower credit scores suggest a higher bankruptcy probability. Credit scores thus facilitate the analysis of bankruptcy risk on insurance demand. One might expect that SMEs with low credit scores and high bankruptcy risks would demand more insurance to mitigate bankruptcy. However, such SMEs may struggle to afford adequate insurance coverage. As a result, they may choose not to purchase insurance because they have less to lose in case of bankruptcy. In contrast, SMEs with high credit scores and low bankruptcy risks may demand more insurance, as they have more at stake in the event of bankruptcy. This suggests that higher credit scores should correlate with increased insurance demand. Secondly, SMEs that have weaker connections with their primary banks are more likely to seek additional insurance coverage. When these businesses are unable to secure adequate funding from their main bank, they turn to other financial institutions. Consequently, relying on multiple banks for financing can signal increased financial limitations. As a result, SMEs that engage with a higher number of banks are more inclined to obtain increased insurance coverage to mitigate potential property liability losses, knowing that securing loans may prove challenging.

To investigate the effects of ln(Asset), Credit Score and Banks on the demand of property liability insurance, we are interested in estimating regressions of the form:

$$ln\left(\text{Insurance Demand}_i\right) = F\left[g_1\left(ln\left(\text{Asset}_i\right)\right) + g_2\left(\text{Credit Score}_i\right) + g_3\left(\text{Banks}_i\right)\right] + \epsilon_i, \tag{2.5.1}$$

where Banks stands for the number of banks SME transacts with and it is normalized between 0 and 1. We utilize the same dataset as the one employed in Asai (2019). For further details on data preparation, please refer to Asai (2019). This study uses The Management Survey of Corporate Insurance Issues in Japan, which was conducted in January and February of 2014. The survey asked SMEs about their characteristics, insurance purchases, bank relationships, and the Great East Japan Earthquake. To have a basic idea about what the data is like, Table (3.4) reports summary statistics. Similarly, Asai (2019) estimated the linear version of the regression:

$$ln\left(\text{Insurance Demand}_i\right) = \beta_0 + \beta_1 ln\left(\text{Asset}_i\right) + \beta_2 \text{Credit Score}_i + \beta_3 \text{Banks}_i + \epsilon_i,$$

and the study shows that both the parameters of Credit Score and Banks (the number of banks SME transacts with) are significantly positive and the parameters of $ln\left(\text{Asset}_i\right)$ is significantly negative.

Table 2.4 Summary Statistics

	Number of Sample	Mean	Standard Deviation	Median	Min	Max
Dpendent Variables						
ln*(Insurance Demand)*	758	0.782	1.525	0.878	-10.034	6.073
Independent Variables						
ln*(Assets)*	758	6.145	0.397	6.134	4.834	7.270
Credit Score	758	0.585	0.152	0.581	0.000	1.000
Banks	758	0.438	0.245	0.444	0.000	1.000

Figure 1 provides a visual representation of the relationship among ln (Insurance Demand), ln (Asset), Credit Score and Banks in our study. This effects of Credit Score and Banks are found to be both overall positive and statistically significant, corroborating the results of the Ordinary Least Squares (OLS) analysis conducted by Asai (2019). Our research approach, however, offers a more detailed and nuanced perspective compared to the OLS analysis alone.

We observe that when Credit Score is lower than 0.52 or higher than 0.65, the effect on ln (Insurance Demand) is relatively stable. The phenomenon suggests that when Credit Scores are either relatively low or high, they don't have a substantial influence on the demand for Property Liability Insurance. This might be due to the presence of other factors or risk management strategies that play a more significant role in determining the insurance demand. Businesses with either very low or very high Credit Scores might already have risk management strategies in place that don't rely solely on insurance, reducing the impact of the Credit Score on insurance demand.

We also observe that when the number of banks it transacts with is between 0.30 and 0.45, the effect on ln (Insurance Demand) is relatively stable. This suggests that within this range, the number of banks a business works with does not significantly impact the demand for Property Liability Insurance. The possible explanation is that businesses with either fewer or more bank relationships have different risk management strategies compared to those with a moderate number of bank relationships. This difference in risk management approaches could lead to varying demand for Property Liability Insurance.

According to Figure 1, when ln (Asset) is lower than 6.0, its effects is found to be negative and statistically significant, corroborating the results of the Ordinary Least Squares (OLS) analysis conducted by Asai (2019). Our research approach provides a more detailed and nuanced perspective on the relationship between a company's assets and the demand for Property Liability Insurance. When ln(Asset) exceeds 6.1, the effect is observed to be positive and statistically significant, while when ln(Asset) surpasses 6.3, the effect becomes stable, deviating from the outcomes of the Ordinary Least Squares (OLS) analysis conducted by Asai (2019). There could be several reasons for this observation. When a company has a low asset level, it may not have sufficient resources to self-insure or absorb potential losses. However, as the asset level increases, the company might be in a better position to bear potential liabilities without relying on Property Liability Insurance. As a result, the demand for this insurance might decrease as the company becomes more capable of handling its own risks. As assets increase beyond a certain level, companies might experience a higher potential for liabilities due to factors such as business expansion or increased complexity, leading to a higher demand for Property Liability Insurance. When assets reach a significantly high level, the company may already have optimized its risk management strategies, and the potential liabilities may not increase proportionally with the assets, resulting in a minimal effect on Property Liability Insurance Demand.

In summary, our approach showcases the importance of using a more detailed and nuanced research approach to better understand the complex effect of different factors on Property Liability Insurance Demand

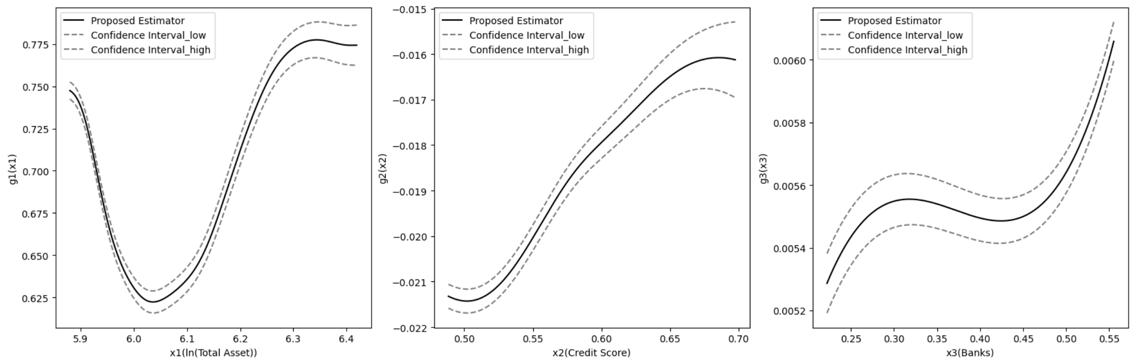

Figure 2.1 Estimation of Structure Functions and Their 95% Confidence Intervals

2.6 Conclusion

In this paper, we create oracle-efficient estimators for a generalized additive model with an unknown link function (GAMULF), with the assumption that the structural functions are additive. Our estimators for the conditional mean and gradient exhibit consistency and asymptotic normality. To estimate the component functions, we suggest a multi-stage algorithm with a refinement stage that employs a one-dimensional kernel, thus bypassing the curse of dimensionality. Furthermore, the multi-stage algorithm either provides closed-form solutions or involves convex optimizations, significantly reducing computational load. Through simulation studies and real data analysis, we demonstrate that our estimator outperforms existing methods in terms of efficiency and robustness.

Chapter Three

Efficient Nonparametric Estimation of Generalized Panel Data Transformation Models with Fixed Effects

Authors: Ying Xia, Peter C. B. Phillips, Liangjun Su

3.1 Introduction

Since the pioneering work of Box and Cox (1964), transformation models have been widely studied. They include many popular models, such as the accelerated failure time model, the Weibull hazard model, the proportional hazard model and the mixed proportional hazard model. Due to their popularity, transformation models have been widely applied to empirical work in various areas of economics to study issues that include the length of unemployment spell, the time between purchases of a particular good, the time intervals between two child births, and the insurance claim durations, among others. See Van den Berg (2001) for a survey on the applications of duration models. Meanwhile, the asymptotic properties of different forms of transformation models have received a great deal of interest. For example, Horowitz (1996) focus on a transformation model with a nonparametric transformation function and a parametric structural function. Chiappori et al. (2015) extend the method in Horowitz (1996) to a transformation model with both nonparametric transformation functions and nonparametric structural functions under endogeneity.

Even though a fully-nonparametric transformation model avoids various misspecification issues, it suffers from the curse of dimensionality. For this reason, there has developed a large literature that applies the additive structure in generalized additive models with an unknown link function; see Horowitz (2001), Horowitz and Mammen (2007), Horowitz and Mammen (2011) and Lin et al. (2018), among others. Recently, Chen et al. (2022) have considered a fully nonparametric transformation model with additive structural functions in a panel data model with fixed effects. In contrast with the early works such as Horowitz and Lee (2004), Chen (2010) and Wang and Chen (2020) who also analyze panel transformation models but assume parametric structural functions, Chen et al. (2022) is the only paper that considers a generalized transformation model with fixed effects under additivity and avoids imposing any parametric assumption. The estimator of the structural function Chen et al. (2022) has a closed-form expression, which makes it is easy to implement and to study the asymptotically normality. Nevertheless, the estimation is done through the matching with other covariates locally and thus suffers from the curse of dimensionality substantially.

To combat the curse of dimensionality, in this paper we propose a three-stage estimation procedure for the generalized transformation model with fixed effects and additive structures. We assume that the nonparametric structural function $g(\cdot)$ exhibits an additive structure: $g(x) = \sum_{l=1}^{d} g_l(x_l)$. Inspired by Horowitz and Mammen (2004, 2011) and Ozabaci, Henderson and Su (2014), we aim to obtain estimators of the additive structural functions that enjoy the orcale efficiency in the sense that they can be estimated as asymptotically efficiently as the oracle estimator obtained when the other additive components are observed. In the first stage, we first consider a regularized sieve method to estimate the logit sieve coefficients associated with the approximation of a composite function of the inverse

$L^{-1}(\cdot)$ of logit-CDF $L(\cdot)$, the CDF $F(\cdot)$ of the error difference, and the structural function $g(\cdot)$, and then generalize the "pairwise differencing" or "matching" method of Blundell and Powell (2004) to obtain initial consistent estimators $\bar{g}_l(\cdot)$ of the structural functions $g_l(\cdot)$. In the second stage, we consider the local polynomial estimation of $LF(\cdot) \equiv L^{-1}(F(\cdot))$ and its first order derivative based on the preliminary consistent estimates $\bar{g}_l(\cdot)$. In the third stage, we apply the local linear method to estimate the one-dimensional object $g_l(\cdot)$ based on the consistent estimates $\{\bar{g}_l(\cdot)\}$ of $\{g_l(\cdot)\}$ and those of $LF(\cdot)$ and its first order derivative. Since only one-dimensional nonparametric objects are estimated in the second and third stage and the additive structure of $g(\cdot)$ is imposed in the whole procedure, the whole estimation procedure does not have the curse of dimensionality issue.

Interestingly, all the minimization problems in our three-stage approach are convex problems. This overcomes the computational hurdle in some existing procedure for transformation models. Furthermore, our estimator achieve optimal convergence rate, asymptotic normality and oracle efficiency.

The article is organized as follows. Section 2 describes our methodology. We present the asymptotic properties of our estimators in Section 3. Section 4 examines the finite sample performance of our estimators via Monte Carlo simulations. We apply our method to an empirical dataset in Section 5. Section 6 concludes. All the proofs of the main theorems are relegated to the appendix.

Notation. For a real matrix A, let A' denote its transpose, and let $\|A\|$ and $\|A\|_{op}$ to denote its Frobenius norm and operator norm, respectively: $\|A\| \equiv [tr(AA')]^{1/2}$ and $\|A\|_{op} \equiv \sqrt{\lambda_{max}(A'A)}$, where $\equiv$ signifies a definitional relationship, $tr(\cdot)$ is a trace operator, and $\lambda_{max}(\cdot)$ denotes the maximum eigenvalue of a real symmetric matrix. Similarly, we use $\lambda_{min}(\cdot)$ to denote the minimum eigenvalue of a real symmetric matrix. For any function $f(\cdot)$ defined on the real line, let $\dot{f}(\cdot)$, $\ddot{f}(\cdot)$, and $\dddot{f}(\cdot)$ be its first, second, and third order derivatives and let $\partial^a f(\cdot)$ be the ath order partial derivative of $f(\cdot)$. Let $\xrightarrow{D}$ and $\xrightarrow{P}$ be convergence in distribution and convergence in probability. Let $\mathbf{1}\{A\}$ denote the usual indicator function which takes one 1 if A holds true and 0 otherwise. For any positive integer c, we write $[c] = \{1, 2, ..., c\}$. For a vector v, $|v|_0$ denotes the number of nonzero elements in v.

3.2 Methodology

In this section we first present the panel data transformation model and then propose a multi-step procedure to estimate it.

3.2.1 The Model

We consider the following transformation model:

$$\Lambda\left(Y_{it}\right) = g\left(X_{it}\right) + \alpha_i + \epsilon_{it} = \sum_{l=1}^{d} g_l\left(X_{l,it}\right) + \alpha_i + \epsilon_{it}, \tag{3.2.1}$$

where $i = 1, \cdots, n,\ t = 1, \cdots, T$, Y_{it} is the observed dependent/response variable, $(X_{1,it}, \cdots, X_{d,it})'$ is a $d \times 1$ vector of observed covariates, $g\left(X_{it}\right) = \sum_{l=1}^{d} g_l\left(X_{l,it}\right)$, α_i is the individual fixed effect that captures the unobserved individual heterogeneity, ϵ_{it} is the idiosyncratic error term, and $\Lambda\left(\cdot\right)$ is an unknown transformation function that is strictly increasing. Note that the model in (3.2.1) specifies a structural relationship between the response variable Y_{it} and the covariates in X_{it}. We address the important issue of "curse of dimensionality" by imposing additive structures on the covariates. Also, for simplicity and clarity we assume that $g_l(\cdot)$, $l = 1, \cdots, d$ are all unknown smooth functions defined on the real line so that each $X_{l,it}$ is a scalar random variable. Even though $g_l(\cdot)$'s are only components of the structural relationship, they are often parameters of interest in empirical applications and we shall refer to them as the structural functions in this paper. In addition, the derivatives, $\dot{g}_1(\cdot), \cdots, \dot{g}_d(\cdot)$, which measure the marginal effects, are also of interest in practice. For example, $\dot{g}_l\left(X_{l,it}\right)$ can be interpreted as the marginal effect of $X_{l,it}$ on $\Lambda\left(Y_{it}\right)$. The main goal of this paper is to estimate $(g_1(\cdot), \cdots, g_d(\cdot))$ and their derivatives $(\dot{g}_1(\cdot), \cdots, \dot{g}_d(\cdot))$. Let $g(x) = \sum_{l=1}^{d} g_l(x_l)$ where $x = (x_1, ..., x_l)'$.

Throughout the paper we focus on a short panel with T being fixed but allow the individual effect α_i to be correlated with the covariates in arbitrarily unknown form. To deal with the fixed effects α_i, we rewrite the model in (3.2.1) as follows:

$$Y_{it} = \Lambda^{-1}\left(g\left(X_{it}\right) + \alpha_i + \epsilon_{it}\right) = \Lambda^{-1}\left(\sum_{l=1}^{d} g_l\left(X_{l,it}\right) + \alpha_i + \epsilon_{it}\right), \tag{3.2.2}$$

where $\Lambda^{-1}(\cdot)$ is the inverse function of $\Lambda(\cdot)$. Clearly, the above expression indicates that the model (3.2.1) is different from the classical panel data model of the following form:

$$Y_{it} = \Lambda^{-1}\left(g\left(X_{it}\right)\right) + \alpha_i + \epsilon_{it} = \Lambda^{-1}\left(\sum_{l=1}^{d} g_l\left(X_{l,it}\right)\right) + \alpha_i + \epsilon_{it}. \tag{3.2.3}$$

For the model in (3.2.3), we can eliminate the fixed effects through various transformations such as first-differing and within-group transformation. Nevertheless, for the model in (3.2.1) or (3.2.2), we cannot apply such transformations to remove α_i due to the presence of the nonlinear function Λ^{-1}. Fortunately, Chen et al. (2022) find that the distribution of $D_{i,ts} \equiv \mathbf{1}\left\{Y_{it} > Y_{is}\right\}$ is free of α_i. This motivates the estimation of the structural functions based on such a non-smooth transformation of the dependent variables.

3.2.2 Estimation Procedure

For clarity, we focus on the case where $T = 2$ and then remark on the general case with $T > 2$ later on. To avoid complication that arises from the presence of discrete covariates, we assume that all covariates are continuous variables.

Following the lead of Chen et al. (2022), we compare Y_{i2} with Y_{i1} by defining $D_i \equiv \mathbf{1}\left\{Y_{i2} > Y_{i1}\right\}$. Since $\Lambda(\cdot)$ is strictly increasing, we have

$$D_i = \mathbf{1}\left\{\Lambda\left(Y_{i2}\right) > \Lambda\left(Y_{i1}\right)\right\}$$

$$
\begin{aligned}
&= \mathbf{1}\left\{g\left(X_{i2}\right)+\epsilon_{i2}>g\left(X_{i1}\right)+\epsilon_{i1}\right\} \\
&= \mathbf{1}\left\{g\left(X_{i2}\right)-g\left(X_{i1}\right)>\Delta_i\right\} \\
&= \mathbf{1}\left\{\sum_{l=1}^{d} g_l\left(X_{l,i2}\right)-\sum_{l=1}^{d} g_l\left(X_{l,i1}\right)>\Delta_i\right\},
\end{aligned} \tag{3.2.4}
$$

where $\Delta_i = \epsilon_{i1} - \epsilon_{i2}$. Obviously, the fixed effect α_i has been removed via the above nonlinear transformation so that the distribution of D_i is free of α_i. Let $X_i = (X_{i1}, X_{i2})$. Let $f(\cdot)$ and $F(\cdot)$ denote the probability density function (PDF) and cumulative distribution function (CDF) of Δ_i, respectively. Then

$$
E\left(D_i|X_i\right) = \Pr\left(g\left(X_{i2}\right)-g\left(X_{i1}\right)>\Delta_i\right) = F\left(g\left(X_{i2}\right)-g\left(X_{i1}\right)\right). \tag{3.2.5}
$$

Inspired by Horowitz and Mammen (2004, 2011) and Ozabaci et al. (2014), we propose a three-step procedure to estimate the structural functions and their derivatives below. In the first stage, we first consider a regularized sieve method to estimate the sieve coefficients associated with the approximation of a composite function of the inverse $L^{-1}(\cdot)$ of logit-CDF $L(\cdot)$, the CDF $F(\cdot)$ of Δ_i, and the structural equation $g(\cdot)$, and then generalize the "pairwise differencing" or "matching" method of Blundell and Powell (2004) to obtain initial consistent estimators $\bar{g}_l(\cdot)$ of the structural functions $g_l(\cdot)$. In the second stage, we consider the local polynomial estimation of $LF(\cdot) \equiv L^{-1}(F(\cdot))$ and its first order derivative based on the preliminary consistent estimates $\bar{g}_l(\cdot)$. Note that $LF(\cdot)$ is a one-dimensional smooth function and its estimation does not have the curse of dimensionality issue. In the third stage, we apply the local linear method to estimate the one-dimensional object $g_l(\cdot)$ based on early consistent estimates $\{\bar{g}_l(\cdot)\}$ of $\{g_l(\cdot)\}$ and those of $LF(\cdot)$ and its first order derivative. Again, here there is no curse of dimensionality involved here.

First-stage estimation of $\{g_l(\cdot)\}_{l=1}^{d}$

In the first stage, we consider initial consistent estimation of the structural functions $\{g_l(\cdot)\}_{l=1}^{d}$ in model (3.2.1), which is done through two sub-steps.

In principle, we can estimate $\{g_l(\cdot)\}_{l=1}^d$ via least squares based on the model for the response variable D_i by using sieve approximation for the structural functions in (3.2.4). Nevertheless, the least squares estimates do not perform well as it cannot ensure the resulting probability estimates to lie between 0 and 1. To ensure the probability estimates to always lie between 0 and 1 during the computation, we follow the lead of Hirano, Imbens and Ridder (2003) and consider the method of logit sieve.

To proceed, we introduce some notations. Let $\{p_l(\cdot), l = 1, 2, \cdots\}$ denote a sequence of B-spline basis functions. Let $K = K(n)$ be some integer such that $K(n) \to \infty$ as $n \to \infty$.

Let $P^K(x_{it}) = \left[p^K(x_{1,it})', \cdots, p^K(x_{d,it})'\right]'$ where $p^K(x_l) \equiv [p_1(x_l), \cdots, p_K(x_l)]'$ for $l = 1, ..., d$. Then under suitable smooth conditions, we can approximate $g_l(\cdot)$ by $p^K(\cdot)'\beta^{x_l}$ where $\beta^{x_l} = (\beta_1^{x_l}, \cdots, \beta_K^{x_l})'$ is a $K \times 1$ vector of parameters. Let $\beta = (\beta^{x_1\prime}, \cdots, \beta^{x_d\prime})'$. In the sequel, we propose to use B-spline estimation as it has faster uniform convergence rate than the estimation based on the power splines. In addition, it is well known that B-splines have low multicollinearity and recursive formula for calculation, which provides great computational advantages in practice. See Chapter 19 of Powell (1981) and Chapter 4 of Schumaker (2007) for more details on B-splines.

Let $\Delta g(X_i) = g(X_{i2}) - g(X_{i1})$, and $LF(\cdot) = L^{-1}(F(\cdot))$. In the first substep, we try to approximate the composite function $LF(\Delta g(\cdot))$. Even though the additive structure in $g(\cdot)$ implies that that of $\Delta g : \Delta g(X_i) = \sum_{l=1}^d [g_l(X_{l,i2}) - g_l(X_{l,i1})]$, $LF(\Delta g(X_i))$ can not be written as additive functions of $(X_{1,i1}, ..., X_{d,i1}, X_{1,i2}, ..., X_{d,i2})$. This implies that if one uses $\{p^K(x_{l,it}), l \in [d], t \in [2]\}$ to approximate this composite function, one has to use their $2d$-dimensional tensor product to form the basis functions, resulting in the "curse of dimensionality". Fortunately, noting that $\Delta g(\cdot)$ is additive and $LF(\cdot)$ is a one-dimensional function, we can avoid the "curse of dimensionality" via two sieve approximations to the composite function. First, we approximate $\Delta g(X_i)$ as follows:

$$\Delta g(X_i) = \left(P^K(X_{i2}) - P^K(X_{i1})\right)' \beta_0 + [r_1(X_{i2}) - r_1(X_{i1})]$$

$$\equiv \quad \Delta P^K(X_i)'\beta_0 + \Delta r_1(X_i) \qquad (3.2.6)$$

where $r_1(X_i)$ is the approximation error in the sieve approximation of $\Delta g(X_i)$. Then under certain smooth conditions on $F(\cdot)$, we can approximate $LF(\Delta g(X_i))$ as follows

$$\begin{aligned} LF(\Delta g(X_i)) &= LF\left(\Delta P^K(X_i)'\beta_0 + \Delta r_1(X_i)\right) \\ &= LF\left(\Delta P^K(X_i)'\beta_0\right) + LF(\Delta g_i^*)\,\Delta r_1(X_i) \\ &= \sum_{\ell=0}^{R} \alpha_{\ell,0}\left(\Delta P^K(X_i)'\beta_0\right)^\ell + \left[r_2(X_i) + LF(\Delta g_i^*)\,\Delta r_1(X_i)\right] \\ &\equiv \sum_{\ell=0}^{R} \alpha_{\ell,0}\left(\Delta P^K(X_i)'\beta_0\right)^\ell + r(X_i), \end{aligned} \qquad (3.2.7)$$

where Δg_i^* lies between $\Delta g_i(X_i)$ and $\Delta P^K(X_i)'\beta_0$, $r_2(X_i)$ can be regarded as the remainder term in the Rth order Taylor expansion of $LF(\cdot)$, and $r(X_i) = [r_2(X_i) + LF(\Delta g_i^*)\,\Delta r_1(X_i)]$. Intuitively, as long as both $g_l(\cdot)$'s and $F(\cdot)$ are sufficiently smooth, and both K and R diverge to infinity, we can control the overall approximation error $r(X_i)$ uniformly well. In practice, we propose to use the following functions as the vector of base functions to approximate $LF(\Delta g(X_i))$:

$$1,\ \Delta P^K(X_i),\ \text{the tensor product of } \Delta P^K(X_i) \text{ up to order } R. \qquad (3.2.8)$$

For notational simplicity, we denote the above vector of base functions simply as $R(X_i) \equiv R^{K_R}(X_i)$ where K_R signifies the dimension of the vector $R(X_i)$. Clearly, K_R is a deterministic function of K and R. Then we have

$$LF(\Delta g(X_i)) \approx R(X_i)'\,\pi_0 \text{ for some } \pi_0 \in \mathbb{R}^{K_R}.$$

Note that the true values of the elements of π_0 depend on the coefficients $\alpha_{\ell,0}$'s and β_0 nonlinearly, but it is hard to incorporate such restrictions in the following estimation procedure. Instead, we will consider a regularized procedure to estimate π_0. Specifically, we propose to estimate π_0 by the regularized logit sieve (RLS) method:

$$\bar{\pi} = \arg\min_{\pi} -\frac{1}{n}\sum_{i=1}^{N}\left[D_i \cdot \ln\left(L\left(R(X_i)'\pi\right)\right) + (1-D_i)\cdot \ln\left(1 - L\left(R(X_i)'\pi\right)\right)\right] + \lambda\left\|\pi\right\|_1, \tag{3.2.9}$$

where $L(\cdot)$ is the Logit CDF: $L(x) = \frac{\exp(x)}{1+\exp(x)}$, $\|\cdot\|_1$ is the L_1 norm, and $\lambda = \lambda(n)$ is a tuning parameter that shrinks to zero as $n \to \infty$. In comparison with the standard logit sieve estimation, we use regularization in the above minimization problem. Following Belloni, Chernozhukov, Fernández-Val and Hansen (2017), we can set

$$\lambda = cn^{-1/2}\Phi^{-1}\left(1 - c_{\lambda n}/\left\{2K_R\right\}\right) \tag{3.2.10}$$

where $c > 1$ is slack constant (e.g., 1.1), $c_{\lambda n} = 0.1/\log(n)$ and $\Phi^{-1}(\cdot)$ is the inverse function of the standard norm CDF Φ. Let $\bar{m}_i \equiv \bar{E}\left(D_i|X_i\right) = L\left(R(X_i)'\bar{\pi}\right)$, which serves as an initial consistent estimator for $m_i \equiv E\left(D_i|X_i\right)$. Note that even though the true link function $F(\cdot)$ is not a Logistic function, we can use Logistic function inside the function $\ln(\cdot)$ in (3.2.9). Following Hirano et al. (2003) and Belloni et al. (2017), we can establish the convergence rate for the above regularized logit sieve estimator under some suitable conditions.

In the second substep, we consider the use of a matching method to estimate the structural functions. To see how the idea of "matching" works, note that

$$m_i = E\left(D_i|X_i\right) = F\left(\Delta g\left(X_i\right)\right).$$

By the strict monotonicity property of the CDF function $F(\cdot)$,

$$m_i \approx m_j \text{ if and only if } \Delta g\left(X_i\right) \approx \Delta g\left(X_j\right).$$

So in principle, one can consider minimizing the average squared distance between $\Delta g\left(X_i\right)$ and $\Delta g\left(X_j\right)$ when we control m_i to lie close to m_j. In practice, both $\Delta g\left(X_i\right)$ and m_i's are not observed, we need to use sieve approximation to obtain the former one and replace the latter one by its preliminary consistent estimate. Note that

$$m_i = F\left(g\left(X_{i2}\right) - g\left(X_{i1}\right)\right) \approx F\left(\left[P^K(X_{i2}) - P^K(X_{i1})\right]'\beta^0\right).$$

For $i \in \{1, \cdots, n\}$, let

$$\begin{aligned} \triangle P_i^K &= P^K(X_{i2}) - P^K(X_{i1}), \\ q_{k,i} &= p_k(X_{1,i2}) - p_k(X_{1,i1}) \text{ for } k = 1, ..., K, \\ Q_{l,i} &= p^K(X_{l,i2}) - p^K(X_{l,i1}) \text{ for } l = 1, ..., d. \end{aligned}$$

For $i \neq j \in \{1, \cdots, n\}$, let

$$\begin{aligned} \triangle P_{i,j}^K &= \triangle P_i^K - \triangle P_j^K, \ \triangle P_{i,j}^{1,K} = q_{1,i} - q_{1,j}, \\ \triangle P_i^{K-1,K} &= \left(q_{2,i}, ..., q_{K,i}, Q_{2,i}', ..., Q_{d,i}'\right)' \text{ and } \triangle P_{i,j}^{K-1,K} = \triangle P_i^{K-1,K} - \triangle P_j^{K-1,K}. \end{aligned}$$

Note that $\triangle P_{i,j}^K = \left(\triangle P_{i,j}^{1,K}, \left(\triangle P_{i,j}^{K-1,K}\right)'\right)'$. To estimate β^0, we normalize its first element to be 1 and rewrite it as $\beta^0 = (1, \theta^{0\prime})'$. The matching estimator of θ^0 is obtained as follows:

$$\bar{\theta} = \arg\min_{\theta} \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \left[\triangle P_{i,j}^{1,K} + \theta' \triangle P_{i,j}^{K-1,K}\right]^2 H_{1h_1}(\bar{m}_j - \bar{m}_i) \tag{3.2.11}$$

$$= -\left\{\frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} \bar{H}_{1h_1,ji}\right\}^{-1} \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \triangle P_{i,j}^{1,K} \triangle P_{i,j}^{K-1,K} \bar{H}_{1h_1,ji}, \tag{3.2.12}$$

where $\bar{H}_{1h_1,ji} = H_{1h_1}(\bar{m}_j - \bar{m}_i)$, $H_{1h_1}(\cdot) \equiv h_1^{-1} H_1(\cdot/h_1)$, $H_1(\cdot)$ is a one-dimensional kernel function, and h_1 is a bandwidth. Let $\bar{\beta} = (1, \bar{\theta}') = \left(\bar{\beta}^{x_1\prime}, ..., \bar{\beta}^{x_d\prime}\right)'$, where $\bar{\beta}^{x_l}$ serves as an estimator of β^{x_l} for $l = 1, ..., d$. Then we obtain the estimate of $g_l(x_l)$ by $\bar{g}_l(x_l) = p^K(x_l)'\bar{\beta}^{x_l}$ for $l = 1, \cdots, d$ and that of $g(x)$ by $\bar{g}(x) \equiv \sum_{l=1}^d \bar{g}_l(x_l)$, where $x = (x_1, ..., x_d)'$.

Second-stage estimation

To motivate the second-stage estimation, we add some notation. Let

$$\triangle g_i = g(X_{i2}) - g(X_{i1}) \text{ and } \triangle g_{i,j} = \triangle g_i - \triangle g_j.$$

Let $LF(\cdot) = L^{-1}(F(\cdot))$, $LF_i = LF(\Delta g_i)$ and $\dot{LF}_i = \dot{LF}(\Delta g_i)$. Note that

$$\sum_{i=1}^N \{D_i \ln[F(\Delta g_i)] + (1 - D_i) \ln[1 - F(\Delta g_i)]\}$$

$$= \sum_{i=1}^{N} \{D_i \ln [L (LF (\Delta g_i))] + (1 - D_i) \ln [1 - L (LF (\Delta g_i))]\} . \quad (3.2.13)$$

By Taylor expansions, for any $i \neq j \in \{1, \cdots, n\}$,

$$LF(\Delta g_i) = LF(\Delta g_j + (\Delta g_i - \Delta g_j)) \approx LF(\Delta g_j) + \sum_{l=1}^{a_2} \partial^{a_2} LF (\Delta g_j) \frac{1}{l!} (\triangle g_{i,j})^l ,$$

where $\triangle g_{i,j}$ is close to zero and $LF (\cdot)$ is a_2-order continuously differentiable.

Let $\triangle \bar{g}_i = \bar{g}(X_{i2}) - \bar{g} (X_{i1})$ and $\triangle \bar{g}_{i,j} = \triangle \bar{g}_i - \triangle \bar{g}_j$. Define

$$\begin{aligned} & Q_n (\triangle \bar{g}_j, \{b_l\}_{l=0}^{a_2}) \\ = \; & \frac{-1}{n} \sum_{i=1}^{N} H_{2h_2} (\triangle \bar{g}_{i,j}) \\ & \times \left\{ D_i \ln \left[L \left(b_0 + \sum_{l=1}^{a_2} \frac{1}{h_2^l l!} (\triangle \bar{g}_{i,j})^l b_l \right) \right] + (1 - D_i) \ln \left[1 - L \left(b_0 + \sum_{l=1}^{a_2} \frac{1}{h_2^l l!} (\triangle \bar{g}_{i,j})^l b_l \right) \right] \right\} . \end{aligned}$$

where $H_{2h_2}(\cdot) \equiv h_2^{-1} H_2 (\cdot/h_2)$, $H_2(\cdot)$ is a one-dimensional kernel function, and h_2 is a bandwidth. Obviously, $b_0 + \sum_{l=1}^{a_2} \frac{1}{h_2^l l!} (\triangle \bar{g}_{i,j})^l b_l$ serves as an a_2-order Taylor series approximation of $LF (\Delta \bar{g}_i)$ in the neighborhood of $\triangle \bar{g}_j$. Then we can estimate $\left(LF_j, h_2 \dot{L}F_j \right)$ by the minimizing $Q_n (\triangle \bar{g}_j, \{b_l\}_{l=0}^{a_2})$ with respect to $\{b_l\}_{l=0}^{a_2}$:

$$\left(\widehat{LF}_j, h_2 \widehat{\partial LF}_j, \cdots, h_2^{a_2} \widehat{\partial^{a_2} LF}_j \right) = \arg \min_{\{b_l\}_{l=0}^{a_2}} Q_n \left(\triangle \bar{g}_j, \{b_l\}_{l=0}^{a_{h_2}} \right) .$$

Let $\widehat{\dot{L}F}_j = \widehat{\partial LF}_j$.

Third-stage estimation

In this stage, we refine the early estimates of the structural functions. Our objective is to obtain an estimator of $g_l (\cdot)$ that is as asymptotically efficient as that obtained when the other $(d - 1)$ the structural functions $\{g_{l^*} (\cdot)\}_{l^*=1, l^* \neq l}^{d}$ were known.

Note that $\Delta g_i = \sum_{j=1}^{d} [g_j (X_{j,i2}) - g_j (X_{j,i1})]$ enters the Logit sieve objective function. For the moment, suppose that $\{g_{l^*} (\cdot)\}_{l^*=1, l^* \neq l}^{d}$ is known, we aim at estimating $g_l (\cdot)$ alone by the local linear method. Noting that $g_l(\cdot)$ appears twice in Δg_i, one may be tempted to conduct the local linear approximation of $g_l(X_{l,i2})$ and $g_l(X_{l,i1})$ simultaneously around a point x_l. But to control the approximation well, one would need to ensure both $X_{l,i2}$ and $X_{l,i1}$ are around x_l. This will yield a local linear estimator with a slower convergence rate than the usual one-dimensional local linear estimate. To avoid such slow convergence, we consider Taylor expansion of $g_l(X_{l,i2})$ and $g_l(X_{l,i1})$ separately around a point x_l below.

First, by the Taylor expansion of $g_l(X_{l,i1})$ around x_l, we have $g_l(X_{l,i1}) \approx g_l(x_l) + \dot{g}_l(x_l)(X_{l,i1} - x_l)$. It follows that

$$\begin{aligned}&\sum_{j=1,\neq l}^{d}[g_j(X_{j,i2}) - g_j(X_{j,i1})] + g_l(X_{l,i2}) - g_l(x_l) - \dot{g}_l(x_l)(X_{l,i1} - x_l)\\&= \Delta g_i + g_l(X_{l,i1}) - g_l(x_l) - \dot{g}_l(x_l)(X_{l,i1} - x_l) \equiv G_{l1,i},\end{aligned}$$

and

$$LF(G_{l1,i}) \approx LF(\Delta g_i) - \dot{LF}(\Delta g_i)[g_l(x_l) + \dot{g}_l(x_l)(X_{l,i1} - x_l) - g_l(X_{l,i1})] \equiv LF_{i1}(x_l). \tag{3.2.14}$$

Similarly, using $g_l(X_{l,i2}) \approx g_l(x_l) + \dot{g}_l(x_l)(X_{l,i2} - x_l)$ by Taylor expansion of $g_l(X_{l,i2})$ around x_l, we have

$$\begin{aligned}&\sum_{j=1,j\neq l}^{d}[g_j(X_{j,i2}) - g_j(X_{j,i1})] + g_l(x_l) + \dot{g}_l(x_l)(X_{l,i2} - x_l) - g_l(X_{l,i1})\\&= \Delta g_i + g_l(x_l) + \dot{g}_l(x_l)(X_{l,i2} - x_l) - g_l(X_{l,i2}) \equiv G_{l2,i},\end{aligned}$$

and

$$LF(G_{l2,i}) \approx LF(\Delta g_i) + \dot{LF}(\Delta g_i)[g_l(x_l) + \dot{g}_l(x_l)(X_{l,i2} - x_l) - g_l(X_{l,i2})] \equiv LF_{i2}(x_l). \tag{3.2.15}$$

Obviously, $G_{l1,i}$ is an approximation version of Δg_i in which only $g_l(X_{l,i1})$ is replaced by its first order Taylor expansion at x_l, and $G_{l2,i}$ is that of Δg_i in which only $g_l(X_{l,i12})$ is replaced by its first order Taylor expansion at x_l. Then we may consider the following local likelihood function to estimate $(g_l(x_l), \dot{g}_l(x_l))$

$$\sum_{t=1}^{2}\sum_{i=1}^{N} H_{3h_3}(X_{l,it}-x_l)\{D_i \ln[L(LF(G_{lt,i}))]+(1-D_i)\ln[1-L(LF(G_{lt,i}))]\}$$
$$\approx \sum_{t=1}^{2}\sum_{i=1}^{N} T_{h_3}(X_{l,it}-x_l)\{D_i \ln[L(LF_{it}(x_l))]+(1-D_i)\ln[1-L(LF_{it}(x_l))]\}, \quad (3.2.16)$$

where $H_{3h_3}(\cdot) \equiv h_3^{-1}H_3(\cdot/h_3)$, $H_3(\cdot)$ is a one-dimensional kernel function, and h_3 is a bandwidth.

Of course, we cannot minimize the negative of (3.2.16) with respect to $(g_l(x_l), \dot{g}_l(x_l))$ given the unknown nature of $LF(\Delta g_i)$ and $\dot{LF}(\Delta g_i)$ in the definitions of $LF_{i1}(x_l)$ and $LF_{i2}(x_l)$. A feasible objective function is given by

$$W_{n,x_l}(c) \equiv -\sum_{t=1}^{T}\sum_{i=1}^{N} H_{3h_3}(X_{l,it}-x_l)\left[D_i \ln\left(L\left(\widehat{LF}_{it,x_l}(c)\right)\right)+(1-D_i)\ln\left(1-L\left(\widehat{LF}_{it,x_l}(c)\right)\right)\right], \quad (3.2.17)$$

where $c \equiv (c_0, c_1)'$,

$$\widehat{LF}_{i1,x_l}(c) = \widehat{LF}_i - \widehat{\dot{LF}}_i \cdot \left[c_0 + c_1\frac{1}{h_3}(X_{l,i1}-x_l) - \bar{g}_l(X_{l,i1})\right], \text{ and}$$
$$\widehat{LF}_{i2,x_l}(c) = \widehat{LF}_i + \widehat{\dot{LF}}_i \cdot \left[c_0 + c_1\frac{1}{h_3}(X_{l,i2}-x_l) - \bar{g}_l(X_{l,i2})\right].$$

By minimizing the objective function in (3.2.17) with respect to (c_0, c_1) yields the following estimates

$$\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right) = \arg\min_{(c_0,c_1)} W_{n,x_l}(c_0, c_1).$$

In the next section we will show that the estimators $\widehat{g}_l(x_l)$ and $\widehat{\dot{g}}_l(x_l)$ are oracle efficient.

3.3 Assumptions and Asymptotic Results

In this section, we first present the assumptions and then study the asymptotic properties of the estimators of the structural functions.

3.3.1 Assumptions

To proceed, we introduce some notation. A real-valued m-times continuously differentiable function $q(u)$ on $\mathcal{U} \subset \mathbb{R}$ is said to be a γ-smooth function if, for some $r = \gamma - m \in (0, 1]$, $\exists c_q$, $|\partial^m q(u) - \partial^m q(u^*)| \leq c_q |u - u^*|^r$ holds for all $u, u^* \in \mathcal{U}$. It is well known that γ-smooth functions can be approximated well by various linear B-splines (e.g., Chen (2007)). So we will assume that $\{g_l(\cdot)\}_{l=1}^d$ are γ-smooth functions below.

We will use $\mathcal{X} = \otimes_{l=1}^d \mathcal{X}_l$ to denote the support of $X_{it} = (X_{1,it}, ...X_{d,it})'$. Let $\mathcal{X}^{\otimes 2} = \mathcal{X} \times \mathcal{X}$ denote the support of (X_{i1}, X_{i2}). We make the following assumptions.

Assumption 7 *1. $\{Y_i, X_i\}_{i=1}^n$ are i.i.d.;*

2. The support $\mathcal{X} = \otimes_{l=1}^d \mathcal{X}_l$ of $X_{it} = (X_{1,it}, ...X_{d,it})'$ is compact;

3. ϵ_{it} is strictly stationary over time.

4. $(\epsilon_{i1}, \epsilon_{i2})$ is independent of (X_{i1}, X_{i2});

5. There exist positive constants $\underline{c}_\epsilon$, $\overline{c}_\epsilon$ and c_ϵ such that $\underline{c}_\epsilon \leq E(\epsilon_{it}^2) \leq \overline{c}_\epsilon$ and $E|\epsilon_{it}|^j \leq c_\epsilon^{j-2} j! E(\epsilon_{it}^2) < \infty$ for all $j \geq 2$.

Assumption 7 imposes some conditions on $\{Y_i, X_i, \epsilon_{it}\}$. Assumption 7(1) assumes the observations are i.i.d.; Assumption 7(2) assumes the exogenous independent variables have compact supports. Assumptions 7(3) is made to simplify the notation. Assumption 7(4) is commonly assumed in the nonparametric transformation models to avoid the estimation of certain conditional distributions. Assumption 7(5) imposes some moment conditions on the error terms to simplify the derivation.

Assumption 8 *1. The link function $\Lambda(\cdot)$ is strictly increasing;*

2. $\beta_0 = (1, \theta_0')'$.

Assumption 2 is an identification condition. Note that we impose a strictly monotone condition on the link function in Assumption 2(1) and normalize the first element of β_0 to be 1 in Assumption 2(2). Without the scale normalization, the structural functions $\{g_l(\cdot)\}_{l=1}^d$ cannot be separately identified from $\Lambda(\cdot)$.

Assumption 9 *1. The CDF $F(\cdot)$ of $\Delta_i = \epsilon_{i1} - \epsilon_{i2}$ is strictly monotone and $(R+1)th$ order continuously differentiable.*

2. *There exists a small positive constant c such that* $0 < c < \inf_{x=(x_1,x_2)\in\mathcal{X}^{\otimes 2}} E(D_i|X_i = x) \leqslant \inf_{x=(x_1,x_2)\in\mathcal{X}^{\otimes 2}} E(D_i|X_i = x) \leqslant 1 - c$.

3. *The set of basis functions $\{p_k(\cdot)\}_{k=1}^{\infty}$ are twice continuously differentiable on their supports;* $\max_{0\leq h_2\leq r}\max_{1\leq l\leq d}\sup_{x_l\in\mathcal{X}_l}\left\|\partial^{h_2}p^K(x_l)\right\| \leq C\zeta_{rK}$ *for $r = 0, 1, 2$ for some large constant C.*

4. *The functions $\{g_l(\cdot)\}_{l=1}^d$ are bounded and γ-smooth function with $\gamma \geq 2$ on their supports; there exist a vector $\beta_0 = \left(\beta_0^{x_1\prime}, ..., \beta_0^{x_d\prime}\right)'$ such that $\beta_0^{x_l} \in interior(\mathcal{B})$ for some compact set $\mathcal{B}$ in $\mathbb{R}^K$ and all $l = 1, ..., d$, and* $\max_{1\leq l\leq d}\sup_{x_l\in\mathcal{X}_l}\left|g_l(x_l) - p^K(x_l)'\beta_0^{x_l}\right| = O(K^{-\gamma})$ *for some $\gamma > 2$.*

5. *There exist a vector $\pi^0 \in interior(\Pi)$ for some compact set Π in $\mathbb{R}^{K_R}$ such that* $\sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}}|LF(\Delta g(x)) - R(x)'\pi^0| = O(K^{-\gamma} + R^{-(R+1)})$*; we can decompose $R(x) = (R_1(x)', R(x)')'$ and $\pi^0 = (\pi_1^{0\prime}, \pi_2^{0\prime})'$ accordingly such that $s_{\pi_1}^2\log^2\left(K^R \vee n\right) \leq K^{-\gamma}n$, and* $\sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}}|R_2(x)'\pi^0| = O(K^{-\gamma})$ *where $s_{\pi_1} \equiv |\pi_1^0|_0$.*

Assumption 9(1) imposes some smooth conditions on $F(\cdot)$ to ensure the second sieve approximation considered in the first stage estimation. Assumption 9(2) ensures the desirable asymptotic properties of the sieve logit estimator in the first stage. Assumption 9(3)-(4) quantify the properties of the base functions $\{p_k(\cdot)\}_{k=1}^{\infty}$ and the approximation error for one-dimensional γ-smooth functions. Note that many basis functions such as polynomials, splines and wavelets satisfy these conditions with various controls on ζ_{rK}. For splines, it is well known that $\zeta_{rK} = K^{1/2+r}$; see Newey (1997). Assumption 9(5) reflects the error in the approximation of $LF(\Delta g(x))$ by $R(x)'\pi^0$ is uniformly well controlled where the term $K^{-\gamma}$ is carried over from the approximation of the additive function $\Delta g(\cdot)$ by $\Delta P^K(\cdot)'\beta_0$ and the term $R^{-(R+1)}$ signifies the error in the approximation of the $(R+1)$th-order continuously derivative function $LF(\cdot)$ by power series. Clearly, $R^{-(R+1)} \ll K^{-\gamma}$ provided $R \geq \underline{c}\log(K)$ for some $\underline{c} > 0$. This indicates to suffices to consider R to be proportioal to $\log(K)$. Our simulations indicates that a choice of R like 3 or 4 works sufficiently well in general. In addition, Assumption 9(5) indicates that π^0 should be approximately sparse to facilitate the asymptotic analysis.

Assumption 10 *For every K and R that is sufficiently large,*

1. *There exist positive constants C_1 and C_2 such that*

$$0 < C_1 \leq \lambda_{\min}\left(E\left[R(X_i)R(X_i)'\right]\right) \leq \lambda_{\max}\left(E\left[R(X_i)R(X_i)'\right]\right) \leq C_2 < \infty.$$

2. *Let $\eta(m_i) \equiv E\left[\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K}{}'|m_i\right]$ where $j \neq i$. Let $f_m(\cdot)$ denote the density of m_i. All eigenvalues of $E[\eta(m_i) f_m(m_i)]$ are bounded and bounded away from zero: $0 < C_{1L} \leq \lambda_{\min}\left(E\left[\eta(m_i) f_m(m_i)\right]\right) \leq \lambda_{\max}\left(E\left[\eta(m_i) f_m(m_i)\right]\right) \leq C_{2L} < \infty$.*

Assumption 10(1) impose some standard conditions to ensure the logit sieve estimator to be well behaved. Assumption 10(2) ensures the matching estimator in the second substep of the first stage estimation is well behaved.

Assumption 11 1. *The probability density function (PDF) $f_{X_l}(\cdot)$ of $X_{l,it}$, is bounded and bounded away from zero within its support $\mathcal{X}_l$, for $l \in [d]$.*

Assumption 11 imposes some standard conditions on the density of the regressors.

Assumption 12 *1. The kernel function $H_1(\cdot)$ is an a_1-order symmetric kernel function with compact support where $a_1 \geq 2$ is even; it is third order continuously differentiable.*

2. Both $H_2(\cdot)$ and $H_3(\cdot)$ are second order symmetric kernel functions with compact support.

Assumption 12(i) imposes some conditions on the kernel function $H_1(\cdot)$ used in the first stage estimation. To eliminate the effect of the first stage estimation, we typically resort to a higher order kernel with $a_1 \geq 4$. Assumption 12(ii) indicates that we can use the usual second order kernel function in the second stage local polynomial regression and the third stage local linear estimation. Note that we cannot use higher order kernel in local linear or polynomial regressions to avoid asymptotic singularity, but it is fine to set $H_3(\cdot) = H_2(\cdot)$.

Assumption 13 *1. As $n \to \infty$, $K \to \infty$, $R \to \infty$, $h_\ell \to 0$ $\forall \ell \in [3]$, and $R^{-(R+1)} = O\left(K^{-\gamma}\right)$;*

2. $h_2^{a_2+1} + \sqrt{\log(n)/(nh_2)} + \sqrt{K\log(n)/n} + \sqrt{K}h_1^{a_1} + K^{-\gamma+1/2} = o(h_3^2 + (nh_3)^{-1/2})$

3. $K^3\log(n)/n = o(1)$ and $\sqrt{K}(\sqrt{K}h_1^{a_1} + h_1^{-1}(\sqrt{s_{\pi_1}\log(R^R \vee n)/n} + K^{-\gamma})) = o\left(1\right)$.

Assumption 13 imposes some conditions on the bandwidths h_ℓ's, the sieve approximating terms K and R, the order of the kernel used in the first stage estimation, and the order of the local polynomial used in the second stage estimation. Assumption 13(i) is standard and minimal except the last part, which ensures that the second sieve approximation error is no bigger than the first sieve approximation studied in Step 1. Assumption 13(ii) ensures that the asymptotic biases and variances of the first-stage and second-stage estimators are sufficiently small to achieve the oracle efficiency in the third stage. To ensure the last stage local linear estimator of $g_l\left(\cdot\right)$ to enjoy the optimal rate of convergence, we need to choose h_3 to be proportional to $n^{-1/5}$. To be specific, we consider the case where $a_1 = 4$, $a_2 = 3$

and $h_3 \propto n^{-1/5}$. Assumption 13(ii) requires that

$$\begin{aligned}
K &\propto n^{c_K} \text{ for } c_K \in \left(\frac{2}{5(\gamma-1/2)}, \frac{1}{3}\right) \\
h_1 &\propto n^{-c_1} \text{ for some } c_1 \in \left(\frac{1}{10}+\frac{c_K}{8}, 1\right) \\
h_2 &\propto n^{-c_2} \text{ for some } c_2 \in \left(\frac{1}{10}, \frac{1}{5}\right).
\end{aligned}$$

For example, if $\gamma > 2.5$, we can simplify choose $K = n^{1/5}$.

3.3.2 Asymptotic Properties

In this subsection we study the asymptotic properties of our three-step estimators.

The following theorem establishes the asymptotic properties of the first-stage estimator $\bar{\theta}$.

Theorem 3.3.1 *Suppose that Assumptions 7–10, 12(1) and 13(i) and (iii) hold. Let* $\eta_{1Kn} = \sqrt{s_{\pi_1}\log(R^R \vee n)/n} + K^{-\gamma}$ *and* $\eta_{2Kn} = \eta_{1Kn} + \sqrt{K}h_1^{a_1} + K^{-\gamma+1/2}$. *Let* $H_{1h_1,ji} \equiv H_{1h_1}(m_j - m_i)$. *Then*

(i)

$$\begin{aligned}
\bar{\theta} - \theta_0 = &-\left\{\frac{1}{n^2}\sum_{1\le i\neq j\le n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1,ji}\right\}^{-1} \left\{\frac{1}{n^2}\sum_{1\le i\neq j\le n} \triangle g_{i,j} \triangle P_{i,j}^{K-1,K} H_{1h_1,ji}\right. \\
&\left. + \frac{1}{n^2}\sum_{1\le i\neq j\le n} \left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right) \triangle P_{i,j}^{K-1,K} H_{1h_1,ji}\right\} + R_{1n},
\end{aligned}$$

(ii) $\left\|\bar{\theta} - \theta_0\right\| = O_p(\eta_{2Kn})$;

(iii) $\frac{1}{n}\sum_{i=1}^{n}\left[\bar{g}_l(X_{l,i}) - g_l(X_{l,i})\right]^2 = O_p(\eta_{2Kn})$ *for* $l = 1, \cdots, d$;

(iv) $\sup_{x_l\in\mathcal{X}_l}\left|\bar{g}_l(x_l) - g_l(x_l)\right| = O_p(\sqrt{K}\eta_{2Kn})$ *for* $l = 1, \cdots, d$;

where $\|R_{1n}\| = O_p(\eta_{1Kn})$.

Theorem 3.3.1(i) establishes a Bahadur-type representation for the first-stage estimator $\bar{\theta}$. Theorem 3.3.1(ii) establishes the Euclidean norm for $\bar{\theta}$. Theorem 3.3.1(iii)-(iv) establishes the mean square convergence and uniform convergence of $\bar{g}_l(\cdot)$, respectively.

The following theorem establishes the asymptotic properties of the second-stage estimators.

Theorem 3.3.2 *Suppose that Assumptions 7–10, 12 and 13(i) and (iii) hold. Let* $\eta_{3Kn} = \eta_{2Kn} + h_2^{a_2+1} + \sqrt{\ln(n)/(nh_2)}$. *Then*

(i) There exists a positive constant c_F *such that*

$$\left\| \left(\widehat{LF}(\triangle g_j)), h_2 \widehat{\partial LF}(\triangle g_j) \right) - (LF(\triangle g_j)), h_2 \partial LF(\triangle g_j)) \right\| \leq c_F \left| \triangle g_j - \triangle \bar{g}_j \right| + O_p(\eta_{3Kn})$$

uniformly over $j \in \{1, \cdots, n\}$;

(ii) $\frac{1}{n}\sum_{j=1}^{n} \left[\widehat{LF}(\triangle g_j) - LF(\triangle g_j) \right]^2 = O_p(\eta_{3Kn}^2)$, *and* $\frac{1}{n}\sum_{j=1}^{n} \left[h_2 \widehat{\partial LF}(\triangle g_j) - h_2 \partial LF(\triangle g_j) \right]^2 = O_p(\eta_{3Kn}^2)$.

Theorem 3.3.2(i) establishes the asymptotic expansions for $\widehat{LF}(\triangle g_j))$ and $h_2 \widehat{\partial LF}(\triangle g_j)$; Theorem 3.3.2(ii) establishes the mean square error convergence rate for the estimators of $LF(\triangle g_j)$ and $h_2 \partial LF(\triangle g_j)$, respectively.

With Theorems 3.3.1 and 3.3.2, we can establish the asymptotic properties of the third stage estimator of $\{(g_l(\cdot), \dot{g}_l(\cdot))\}_{l=1}^{d}$.

Theorem 3.3.3 *Suppose that Assumptions 7–13 hold. Let* $\kappa_{ab} = \int u^a [H_3(u)]^b du$ *for* $a, b = 0, 1, 2$. *Then for* $l = 1, \cdots, d$,

(i)

$$\begin{pmatrix} \sqrt{nh_3} & 0 \\ 0 & \sqrt{nh_3^3} \end{pmatrix} \left[\begin{pmatrix} \widehat{g}_l(x_l) \\ \widehat{\dot{g}}_l(x_l) \end{pmatrix} - \begin{pmatrix} g_l(x_l) \\ \dot{g}_l(x_l) \end{pmatrix} - \frac{1}{2}\ddot{g}_l(x_l) \begin{pmatrix} h_3^2 \kappa_{21} \\ 0 \end{pmatrix} \right]$$
$$\xrightarrow{d} N\left(0, \left\{ E\left[\frac{\dot{F}^2(\triangle g(X_i))}{F(\triangle g(X_i))\left[1 - F(\triangle g(X_i))\right]} \middle| X_{l,i1} = x_l \right] + E\left[\frac{\dot{F}^2(\triangle g(X_i))}{F(\triangle g(X_i))\left[1 - F(\triangle g(X_i))\right]} \middle| X_{l,i2} = x_l \right] \right\}^{-1} \times \begin{pmatrix} \kappa_{02} & 0 \\ 0 & \frac{\kappa_{22}}{\kappa_{21}^2} \end{pmatrix} \right).$$

(ii) $\sup_{x_l \in \mathcal{X}_l} \left\| \widehat{g}_l(x_l) - g_l(x_l) \right\| = O_p\left(h_3^2 + \sqrt{\ln(n)/(nh_3)} \right)$.

Theorem 3.3.3 reports the asymptotic properties of the third step local linear estimator of $\{(g_l(\cdot), \dot{g}_l(\cdot))\}_{l=1}^{d}$. Theorem 3.3.3(i) indicates that the asymptotic distribution of the local linear estimator of is not affected by random sampling errors in the first two stage estimation. In fact, our local linear estimator of $(g_l(\cdot), \dot{g}_l(\cdot))$ has the same asymptotic distribution that we would have if the other additive components $\{(g_j(\cdot), \dot{g}_j(\cdot))\}_{j=1,\neq l}^{d}$ and link function $F(\cdot)$ were known. This indicates the oracle efficiency of the estimator. Theorem 3.3.3(ii) gives the uniform convergence rate for $g_l(\cdot)$. Following the standard exercise in the nonparametric kernel literature, we can also demonstrate that these estimators of $(g_{l_1}(\cdot), \dot{g}_{l_1}(\cdot))$ and $(g_{l_2}(\cdot), \dot{g}_{l_2}(\cdot))$, $\forall l_1 \neq l_2 \in \{1, \cdots, n\}$ are asymptotically independent.

In the proof of Theorem 3.3.3, we give the linear representations of the nonparametric estimators $\left\{\left(\widehat{g}_l(\cdot), \widehat{\dot{g}}_l(\cdot)\right)\right\}_{l=1}^{d}$ with uniform control of the reminder terms. It serves as a building block for both pointwise and uniform inference. For example, one can consider uniform inference based on the multiplier bootstrap as in Chernozhukov et al. (2014, Corollary 3.1). For brevity, we skip the details.

3.4 Numerical Studies

In this section, we are going to use simulated examples to demonstrate how well the proposed estimation procedure works. We use the same DGPs in Chen et al. (2022) to compare their estimator with the proposed estimator. To save space, we only report the detailed results for the estimator of $g_1(\cdot)$. We consider four data generating processes (DGPs).

DGP I: $\Lambda(Y_{it}) = X_{1,it}^2 + X_{2,it}^2 + \alpha_i + \epsilon_{it}$, where $\epsilon_{it} \sim U(0,1)$.

DGP II: $\Lambda(Y_{it}) = X_{1,it}^2 + X_{2,it}^2 + \alpha_i + \epsilon_{it}$, where $(a\epsilon_{it} + b) \sim \mathcal{X}^2(2)$ with $a = \frac{1}{2}\left(\frac{9}{8}\right)^3$ and $b = \frac{1}{2}\exp\left(-\frac{1}{2a}\right)$.

DGP III: $\Lambda(Y_{it}) = X_{1,it}^3 + 0.5X_{1,it}^2 + X_{2,it}^2 + \alpha_i + \epsilon_{it}$, where $\epsilon_{it} \sim U(0,1)$.

All DGPs take the Box-Cox transformation of Bickel and Doksum (1981) with $\Lambda(y) =$

$\frac{|y|^\lambda sgn(y)-1}{\lambda}$ for $\lambda = 0.8$. Both $X_{1,it}$ and $X_{2,it}$ follow $U(-1,1)$ and their correlation coefficient is 0.2. $\alpha_i = 0.5\left(X_{1,it} + X_{2,it}\right) + 0.5\eta$, where η_i is a $N(0,1)$ random variable. The error term either follows symmetric normal distribution or asymmetric Chi-square distribution of freedom 2.

We define the bias, standard deviation (SD), and root mean integrated squared error (RMISE) of an estimator $\widehat{f}(\cdot)$ of $f(\cdot)$ as

$$\text{bias} = \int \left| \mathrm{E}\left[\widehat{f}(v)\right] - f(v) \right| dv,$$

$$\text{SD} = \int sd\left[\widehat{f}(v)\right] dv$$

and

$$RMISE = (\text{bias}^2 + \text{SD}^2)^{1/2},$$

respectively, and use them to assess the accuracy of the estimator $\widehat{f}(\cdot)$.

The kernel function used in the proposed estimation procedure is the standard Gaussian kernel for all simulated examples in this section. For each simulated example, we assess the accuracy of the proposed estimation procedure for sample size $n = 500$ and for each case, we compute the bias, SD and RMISE of an obtained estimator based on 1000 simulations. Method in Chen et al. (2022) chooses bandwidth by minimizing the leave-one-out cross-validation (CV) function. The proposed method chooses bandwidth by grid search to minimize CV function.

Table 3.1 - 3.3 report bias (Bias), standard deviation (SD) and root mean square error (RMSE) of $g_1(x_1)$ for DGPs I-III, respectively. When the error terms follow normal distribution in DGP I and III, the proposed method works better than the method in Chen et al. (2022), especially at boundary points. When the error term follows Chi-square distribution in DGP II, the proposed method defeats the method in Chen et al. (2022) at boundary points, e.g. $x_1 = -0.8, -0.6, -0.4, 0.4, 0.6, 0.8$, and does not function well at center points, e.g. $x_1 = -0.2, 0.2$. As expected, we usually observe a relatively larger RMSE when the

Table 3.1 Estimation results for DGP I

x_1	-0.8	-0.6	-0.4	-0.2	0.2	0.4	0.6	0.8
$g_1(x_1)$	0.64	0.36	0.16	0.04	0.04	0.16	0.36	0.64
				Chen et al. (2022)				
RMSE	0.2459	0.1489	0.0946	0.0571	0.0576	0.0928	0.143	0.2358
Bias	-0.1942	-0.091	-0.03	-0.0025	0.004	-0.0232	-0.0828	-0.1826
SD	0.151	0.118	0.0898	0.0571	0.0575	0.09	0.1167	0.1492
				the proposed estimator				
RMSE	0.108	0.0804	0.0614	0.0423	0.0403	0.0609	0.0834	0.1116
Bias	-0.0074	-0.0016	0.0128	0.0179	0.0132	0.0039	-0.0129	-0.0176
SD	0.1077	0.0804	0.0601	0.0383	0.0381	0.0608	0.0824	0.1102

evaluation point is close to the boundary and it is much more obvious in Chen et al. (2022). The dimension of variables does not influence the simulation performance of the proposed method, however, the method in Chen et al. (2022) suffers from the curse of dimensionality in implementation.

3.5 Application: the Effect of Income Shock on Job Creation

In our empirical study, we investigate the impact of fluctuations in regional income on employment generation within the nontradable sector for firms. The nontradable sector encompasses goods and services that cannot be easily traded or transported across regions, such as hairdressing, restaurants, and local retail stores. These types of businesses predominantly rely on the demand from consumers within their respective regions. Mian and Sufi (2012) argue that firms operating in the nontradable sector are heavily influenced by local

Table 3.2 Estimation results for DGP II

x_1	-0.8	-0.6	-0.4	-0.2	0.2	0.4	0.6	0.8
$g_1(x_1)$	0.64	0.36	0.16	0.04	0.04	0.16	0.36	0.64
				Chen et al. (2022)				
RMSE	0.289	0.1668	0.0962	0.0566	0.0594	0.1013	0.1714	0.2962
Bias	-0.2476	-0.1213	-0.0422	-0.0044	-0.0035	-0.0435	-0.1243	-0.2534
SD	0.149	0.1145	0.0865	0.0564	0.0593	0.0915	0.1181	0.1535
				the proposed estimator				
RMSE	0.1858	0.1389	0.1073	0.0821	0.0829	0.105	0.1335	0.186
Bias	-0.0428	-0.0073	0.024	0.039	0.0396	0.0259	-0.0029	-0.0332
SD	0.1808	0.1387	0.1046	0.0723	0.0728	0.1018	0.1335	0.183

demand, which in turn is impacted by the income levels of the local population. As a result, when there is a positive shock to income in a region, the purchasing power of the residents typically increases. This surge in consumer spending leads to a higher demand for goods and services in the nontradable sector.

To illustrate this relationship, let's consider a hypothetical example. Suppose a region experiences an economic boom, resulting in increased income levels for the majority of its residents. As these individuals now have more disposable income, they are more likely to spend on goods and services such as dining out, visiting local attractions, or utilizing personal services like hair salons. This increase in local demand, driven by higher income levels, would then create more opportunities for businesses in the nontradable sector, potentially leading to the establishment of new businesses or the expansion of existing ones. Consequently, this growth would translate into the creation of additional employment opportunities within the nontradable sector. To test this hypothesis, we are interested in

Table 3.3 Estimation results for DGP III

x_1	-0.8	-0.6	-0.4	-0.2	0.2	0.4	0.6	0.8
$g_1(x_1)$	0.64	0.36	0.16	0.04	0.04	0.16	0.36	0.64
Chen et al. (2022)								
RMSE	0.1471	0.1354	0.1163	0.0759	0.0715	0.0935	0.1215	0.2705
Bias	-0.0238	-0.0792	-0.075	-0.0456	0.0431	0.0344	-0.0405	-0.2282
SD	0.1452	0.1099	0.0889	0.0607	0.0571	0.087	0.1146	0.1453
the proposed estimator								
RMSE	0.0904	0.0934	0.0896	0.0471	0.0909	0.0838	0.1006	0.0407
Bias	-0.0312	-0.0755	-0.0404	-0.0179	0.0213	0.0322	0.0912	0.0186
SD	0.0848	0.055	0.08	0.0436	0.0884	0.0774	0.0425	0.0362

estimating regressions of the form:

$$\Lambda\left(\text{Job Creation}_{it}\right) = g_1\left(\text{Income Growth Rate}_{it}\right) + g_2\left(ln\left(\text{Total Wage}_{it}\right)\right) + \alpha_i + \epsilon_{it}, \quad (3.5.1)$$

where Job Creation$_{it}$ is the net employment creation in firms in the nontradable sector in each age category t - startups (0-1 year old), 2-3 year old, 4-5 year old, and firms 6 year old or older. We scale all employment numbers by total nontradable sector employment as of 2000. Income growth rate is the two-year growth in total wages and salaries in the county level. We utilize the same dataset as the one employed in Adelino, Ma and Robinson (2017). For further details on data preparation, please refer to Adelino et al. (2017). The net employment creation by firm age is computed from the publicly available data from the U.S. Census Quarterly Workforce Indicators (QWI). Income data at the county level come from the Internal Revenue Srvice(IRS) Statistics of Income and is measured in calendar years (i.e., January to December of each year). To have a basic idea about what the data is like, Table 3.4 reports summary statistics. Similarly, Adelino et al. (2017) estimated the linear version of the regression:

$$\Lambda\,(\text{Job Creation}_{it}) = \beta_0 + \beta_1 \text{Income Growth Rate}_{it} + \beta_2 ln\,(\text{Total Wage}_{it}) + \beta_3 ln(\text{Labor Force Population}_{it}) + \beta_4 ln(\text{Percentage of High School Degree or Above}_{it}) + \epsilon_{it},$$

where Labor Force Population$_{it}$ and Percentage of High School Degree or Above$_{it}$ are control variables. Since model (3.5.1) considers individual fixed effects and these two control variables are absorbed into fixed effects. Adelino et al. (2017) shows that both the parameters of Income Growth Rate and $ln\,(\text{Total Wage}_{it})$ are significantly positive.

Table 3.4 Summary Statistics

	Employment Creation	Income Growth	ln(Total Wages)
count	2005	2005	2005
mean	0.0082	0.0267	0.9844
std	0.0483	0.0439	1.4064
min	-0.3423	-0.0914	-2.4323
max	0.2158	0.2505	5.7083

Figure 1 provides a visual representation of the relationship among net employment creation, income growth, and ln(total wages) in our study. This relationship is found to be both positive and statistically significant, corroborating the results of the Ordinary Least Squares (OLS) analysis conducted by Adelino et al. (2017). Our research approach, however, offers a more detailed and nuanced perspective compared to the OLS analysis alone.

We observe that when income growth is close to zero, the effect on employment creation is relatively minimal. This indicates that a substantial increase in income is necessary for it to significantly influence job creation. Furthermore, our analysis reveals that when total wages reach exorbitantly high levels, they actually hinder the creation of new employment opportunities. This suggests that there is an optimal range for total wages to encourage job growth.

In summary, Figure 1 demonstrates that net employment creation is positively and significantly influenced by income growth and ln(total wages) in our study. Our approach offers additional insights, highlighting the importance of income growth and the potential negative effects of excessively high total wages on job creation. A noteworthy finding from our study is the strong impact of increased income on non-tradable employment. This implies that as income rises, there is an increased demand for goods and services in the nontradable sector, such as local retail and hospitality industries. This heightened demand, in turn, stimulates the creation of more job opportunities within the sector, ultimately contributing to overall higher total employment.

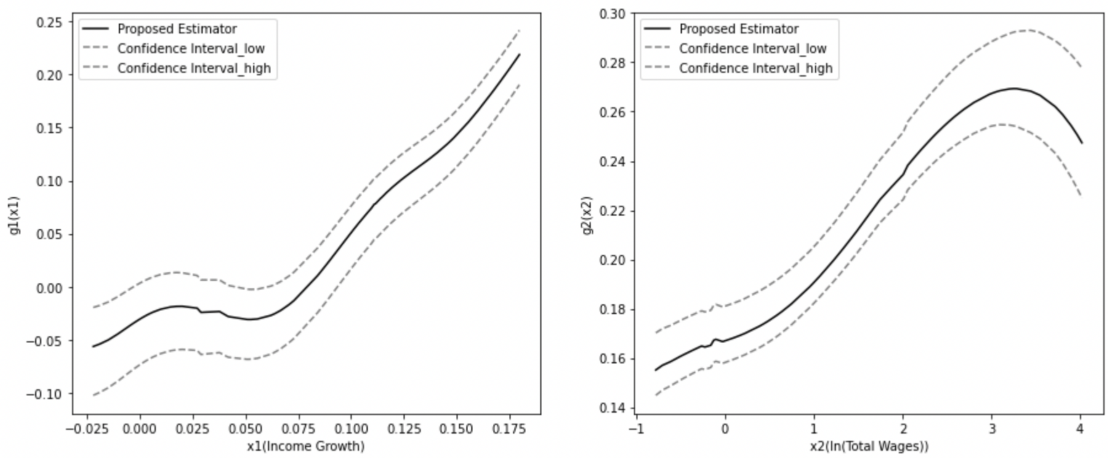

Figure 3.1 Estimation of Structure Functions and Their 95% Confidence Intervals

3.6 Conclusion

In this paper, we create oracle-efficient estimators for a broadened panel data transformation model that includes fixed effects, with the assumption that the structural functions are additive. Our estimators for the conditional mean and gradient exhibit consistency and asymptotic normality. To estimate the component functions, we suggest a multi-stage algorithm with a refinement stage that employs a one-dimensional kernel, thus bypassing the curse of dimensionality. Furthermore, the multi-stage algorithm either provides closed-form solutions or involves convex optimizations, significantly reducing computational load. Through simulation studies and real data analysis, we demonstrate that our estimator outperforms existing methods in terms of efficiency and robustness.

REFERENCES

Adelino, M., Ma, S., Robinson, D., 2017. Firm age, investment opportunities, and job creation. The Journal of Finance 72, 999–1038.

Asai, Y., 2019. Why do small and medium enterprises (smes) demand property liability insurance? Journal of Banking & Finance 106, 298–304.

Belloni, A., Chernozhukov, V., Fernández-Val, I., Hansen, C., 2017. Program evaluation and causal inference with high-dimensional data. Econometrica 85, 233–298.

Van den Berg, G.J., 2001. Duration models: specification, identification and multiple durations, in: Handbook of econometrics. Elsevier. volume 5, pp. 3381–3460.

Bickel, P.J., Doksum, K.A., 1981. An analysis of transformations revisited. Journal of the american statistical association 76, 296–311.

Blundell, R.W., Powell, J.L., 2004. Endogeneity in semiparametric binary response models. The Review of Economic Studies 71, 655–679.

Box, G.E., Cox, D.R., 1964. An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological) 26, 211–243.

Breiman, L., Friedman, J.H., 1985. Estimating optimal transformations for multiple regression and correlation. Journal of the American statistical Association 80, 580–598.

Buja, A., Hastie, T., Tibshirani, R., 1989. Linear smoothers and additive models. The Annals of Statistics , 453–510.

Chen, S., 2010. Root-n-consistent estimation of fixed-effect panel data transformation models with censoring. Journal of Econometrics 159, 222–234.

Chen, S., Lu, X., Wang, X., 2022. Nonparametric estimation of generalized transformation models with fixed effects. Econometric Theory , 1–32.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. Handbook of econometrics 6, 5549–5632.

Chernozhukov, V., Chetverikov, D., Kato, K., 2014. Anti-concentration and honest, adaptive confidence bands. The Annals of Statistics 42, 1787–1818.

Chiappori, P.A., Komunjer, I., Kristensen, D., 2015. Nonparametric identification and estimation of transformation models. Journal of Econometrics 188, 22–39.

Fan, J., Härdle, W., Mammen, E., 1998. Direct estimation of low-dimensional components in additive models. The Annals of Statistics 26, 943–971.

Härdle, W., Stoker, T.M., 1989. Investigating smooth multiple regression by the method of average derivatives. Journal of the American statistical Association 84, 986–995.

Hastie, T.J., Tibshirani, R.J., 2017. Generalized additive models. Routledge.

Hirano, K., Imbens, G.W., Ridder, G., 2003. Efficient estimation of average treatment effects using the estimated propensity score. Econometrica 71, 1161–1189.

Horowitz, J.L., 1996. Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. Econometrica 64, 103.

Horowitz, J.L., 2001. Nonparametric estimation of a generalized additive model with an unknown link function. Econometrica 69, 499–513.

Horowitz, J.L., Lee, S., 2004. Semiparametric estimation of a panel data proportional hazards model with fixed effects. Journal of Econometrics 119, 155–198.

Horowitz, J.L., Mammen, E., 2004. Nonparametric estimation of an additive model with a link function. The Annals of Statistics 32, 2412–2443.

Horowitz, J.L., Mammen, E., 2007. Rate-optimal estimation for a general class of nonparametric regression models with unknown link functions. The Annals of Statistics 35, 2589–2619.

Horowitz, J.L., Mammen, E., 2011. Oracle-efficient nonparametric estimation of an additive model with an unknown link function. Econometric Theory 27, 582–608.

Klein, R.W., Spady, R.H., 1993. An efficient semiparametric estimator for binary response models. Econometrica: Journal of the Econometric Society , 387–421.

Lin, H., Pan, L., Lv, S., Zhang, W., 2018. Efficient estimation and computation for the generalised additive models with unknown link function. Journal of Econometrics 202, 230–244.

Linton, O., Härdle, W., 1996. Estimation of additive regression models with known links. Biometrika 83, 529–540.

Linton, O., Nielsen, J.P., 1995. A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika , 93–100.

Linton, O.B., 2000. Efficient estimation of generalized additive nonparametric regression models. Econometric Theory 16, 502–523.

Liu, R., Yang, L., Härdle, W.K., 2013. Oracally efficient two-step estimation of generalized additive model. Journal of the American Statistical Association 108, 619–631.

Mammen, E., Linton, O., Nielsen, J., 1999. The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. The Annals of Statistics 27, 1443–1490.

Mammen, E., Park, B.U., 2006. A simple smooth backfitting method for additive models. The Annals of Statistics 34, 2252–2271.

Nielsen, J.P., Sperlich, S., 2005. Smooth backfitting in practice. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 43–61.

Opsomer, J.D., 2000. Asymptotic properties of backfitting estimators. Journal of Multivariate Analysis 73, 166–179.

Opsomer, J.D., Ruppert, D., 1997. Fitting a bivariate additive model by local polynomial regression. The Annals of Statistics 25, 186–211.

Ozabaci, D., Henderson, D.J., Su, L., 2014. Additive nonparametric regression in the presence of endogenous regressors. Journal of Business & Economic Statistics 32, 555–575.

Powell, J.L., Stock, J.H., Stoker, T.M., 1989. Semiparametric estimation of index coefficients. Econometrica: Journal of the Econometric Society , 1403–1430.

Powell, M.J.D., 1981. Approximation theory and methods. Cambridge university press.

Schumaker, L., 2007. Spline functions: basic theory. Cambridge University Press.

Stone, C.J., 1985. Additive regression and other nonparametric models. The annals of Statistics , 689–705.

Stone, C.J., 1986. The dimensionality reduction principle for generalized additive models. The Annals of Statistics , 590–606.

Vershynin, R., 2018. High-dimensional probability: An introduction with applications in data science. volume 47. Cambridge university press.

Wang, X., Chen, S., 2020. Semiparametric estimation of generalized transformation panel data models with nonstationary error. The Econometrics Journal 23, 386–402.

Yu, K., Park, B.U., Mammen, E., 2008. Smooth backfitting in generalized additive models. The Annals of Statistics 36, 228–260.

APPENDIX

Appendix A

Appendix for Chapter 2

A.1 Proof of Theorem 2.3.1

Substitute the following equality

$$\triangle P_{i,j}^{(1)} = \triangle g_{i,j} - \theta_0{}' \triangle P_{i,j}^{(-1)} + \left(\triangle P_{i,j}^{K\prime} \beta_0^{x,z} - \triangle g_{i,j}\right) \tag{A.1.1}$$

into the equation ((3.2.12)) and then we have

$$\begin{aligned}
&\bar{\theta} - \theta_0 \\
=& - \left\{ \frac{1}{n^2} \sum_{j=2}^{n} \sum_{i<j}^{n} \triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} H_{h_2} \left[\bar{E}\left(Y \left| X_j \right.\right) - \bar{E}\left(Y \left| X_i \right.\right) \right] \right\}^{-1} \\
&\times \frac{1}{n^2} \sum_{j=2}^{n} \sum_{i<j}^{n} \triangle g_{i,j} \triangle P_{i,j}^{(-1)} H_{h_2} \left[\bar{E}\left(Y \left| X_j \right.\right) - \bar{E}\left(Y \left| X_i \right.\right) \right] \\
&- \left\{ \frac{1}{n^2} \sum_{j=2}^{n} \sum_{i<j}^{n} \triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} H_{h_2} \left[\bar{E}\left(Y \left| X_j \right.\right) - \bar{E}\left(Y \left| X_i \right.\right) \right] \right\}^{-1} \\
&\times \frac{1}{n^2} \sum_{j=2}^{n} \sum_{i<j}^{n} \left(\triangle P_{i,j}^{K\prime} \beta_0^{x,z} - \triangle g_{i,j}\right) \triangle P_{i,j}^{(-1)} H_{h_2} \left[\bar{E}\left(Y \left| X_j \right.\right) - \bar{E}\left(Y \left| X_i \right.\right) \right] \\
=& - L_{0,n}^{-1} L_{1,n} - L_{0,n}^{-1} L_{2,n}.
\end{aligned} \tag{A.1.2}$$

By $\triangle P_{i,j}^{(-1)} = -\triangle P_{j,i}^{K-1,K}$,

$$\begin{aligned}
&L_{1,n} \\
=& \frac{1}{n^2} \sum_{j=2}^{n} \sum_{i<j}^{n} g\left(X_j\right) \triangle P_{i,j}^{(-1)} H_{h_2} \left[\bar{E}\left(Y \left| X_j \right.\right) - \bar{E}\left(Y \left| X_i \right.\right) \right]
\end{aligned}$$

$$
\begin{aligned}
&-\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n} g(X_i) \triangle P_{i,j}^{(-1)} H_{h_2}\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i)\right] \\
=&\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n} g(X_j) \triangle P_{i,j}^{(-1)} H_{h_2}\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i)\right]. \qquad \text{(A.1.3)}
\end{aligned}
$$

We apply Taylor Expansion to $H_{h_2}\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i)\right]$ at $E(Y|X_j) - E(Y|X_i)$, for $i \neq j \in \{1, \cdots, n\}$

$$
\begin{aligned}
&H_{h_2}\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i)\right] - H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right] \\
=&h_2^{-1}\partial H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i) - E(Y|X_j) + E(Y|X_i)\right] \\
&+h_2^{-2}\partial^2 H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i) - E(Y|X_j) + E(Y|X_i)\right]^2 \\
&+h_2^{-3}\partial^3 H_{h_2}\left[\bar{\bar{E}}(Y|X_j) - \bar{\bar{E}}(Y|X_i)\right]\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i) - E(Y|X_j) + E(Y|X_i)\right]^3 \\
=&h_2^{-1}\partial H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_j) - E(Y|X_j)\right] \\
&-h_2^{-1}\partial H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_i) - E(Y|X_i)\right] \\
&-2h_2^{-2}\partial^2 H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_i) - E(Y|X_i)\right]\left[\bar{E}(Y|X_j) - E(Y|X_j)\right] \\
&+h_2^{-2}\partial^2 H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_j) - E(Y|X_j)\right]^2 \\
&+h_2^{-2}\partial^2 H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_i) - E(Y|X_i)\right]^2 \\
&+h_2^{-3}\partial^3 H_{h_2}\left[\bar{\bar{E}}(Y|X_j) - \bar{\bar{E}}(Y|X_i)\right]\left[\bar{E}(Y|X_j) - \bar{E}(Y|X_i) - E(Y|X_j) + E(Y|X_i)\right]^3, \\
&\qquad \text{(A.1.4)}
\end{aligned}
$$

where $\bar{\bar{E}}(Y|X_j) - \bar{\bar{E}}(Y|X_i)$ is between $\bar{E}(Y|X_j) - \bar{E}(Y|X_i)$ and $E(Y|X_j) - E(Y|X_i)$. By (A.4.3) in Lemma (A.4.1),

$$
\begin{aligned}
&h_2^{-1}\partial H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_j) - E(Y|X_j)\right] \\
&-h_2^{-1}\partial H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\left[\bar{E}(Y|X_i) - E(Y|X_i)\right] \\
=&h_2^{-1}\delta_{h,n}\left|\partial H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\right| \qquad \text{(A.1.5)}
\end{aligned}
$$

holds uniformly over $i \neq j \in \{1, \cdots, n\}$.

Substitute ((A.1.4)) into $L_{1,n}$ ((B.1.9)) and we have

$$
\begin{aligned}
&L_{1,n}\\
=&\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n} g\left(X_j\right)\triangle P_{i,j}^{(-1)} H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\\
&+h^{-1}\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n} g\left(X_j\right)\triangle P_{i,j}^{(-1)} \partial H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\left[\bar{E}\left(Y\left|X_j\right.\right)-E\left(Y\left|X_j\right.\right)\right]\\
&-h^{-1}\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n} g\left(X_j\right)\triangle P_{i,j}^{(-1)} \partial H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\left[\bar{E}\left(Y\left|X_i\right.\right)-E\left(Y\left|X_i\right.\right)\right]\\
&+\left(h_2^{-1}\delta_{h,n}\right)^2\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left|g\left(X_j\right)\triangle P_{i,j}^{(-1)}\partial^2 H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right|\\
=&L_{1,n,1}+L_{1,n,2}-L_{1,n,3}+L_{1,n,4}. \qquad \text{(A.1.6)}
\end{aligned}
$$

Similarly, we also have

$$
\begin{aligned}
&L_{0,n}\\
=&\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime} H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\\
&+h_2^{-1}\delta_{h,n}\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left|\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\partial H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right|\\
&+\left(h_2^{-1}\delta_{h,n}\right)^2\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left|\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\partial^2 H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right|\\
=&L_{0,n,1}+L_{0,n,2}+L_{0,n,3}; \qquad \text{(A.1.7)}
\end{aligned}
$$

and

$$
\begin{aligned}
&L_{2,n}\\
=&\frac{2}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left(\triangle P_{i,j}^{K\prime}\beta_0^{x,z}-\triangle g_{i,j}\right)\triangle P_{i,j}^{(-1)} H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\\
&+h_2^{-1}\delta_{h,n}\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left|\left(\triangle P_{i,j}^{K\prime}\beta_0^{x,z}-\triangle g_{i,j}\right)\triangle P_{i,j}^{(-1)}\partial H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right|
\end{aligned}
$$

$$+\left(h_2^{-1}\delta_{h,n}\right)^2\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left|\left(\triangle P_{i,j}^{K\prime}\beta_0^{x,z}-\triangle g_{i,j}\right)\triangle P_{i,j}^{(-1)}\partial^2 H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right|$$

$$=L_{2,n,1}+L_{2,n,2}+L_{2,n,3}. \tag{A.1.8}$$

Since the local - polynomial regression is used to reduce bias, the kernel function $H(\cdot)$ could be the second-order and $H_{h_2}(\cdot)=\sqrt{H_{h_2}(\cdot)}\sqrt{H_{h_2}(\cdot)}$. To express terms $L_{1,n}$ and $L_{2,n}$ in a matrix form, let

$$vector_0=\left(\frac{1}{n}g\left(X_j\right)\sqrt{H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]}\right)_{i<j},$$

$$vectoh_{1,1}=\left(\frac{1}{n}\left\{\triangle P_{i,j}^{K\prime}\beta_0^{x,z}-\triangle g_{i,j}\right\}\sqrt{H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]}\right)_{i<j},$$

and

$$vector_2=\left(\frac{1}{n}\triangle P_{i,j}^{(-1)}\sqrt{H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]}\right)_{i<j}.$$

Thus, $L_{0,n,1}=vector_2{}'vector_2$, $L_{1,n,1}=vector_0{}'vector_2$ and $L_{2,n,1}=vector_1{}'vector_2$.

Given (B.1.8), (B.1.12), (A.1.7) and (B.1.13),

$$\begin{aligned}
&\bar{\theta}-\theta_0\\
=&-\left\{\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right\}^{-1}\\
&\times\left\{\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\triangle g_{i,j}\triangle P_{i,j}^{(-1)}H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right.\\
&\left.-\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left(\triangle P_{i,j}^{K\prime}\beta_0^{x,z}-\triangle g_{i,j}\right)\triangle P_{i,j}^{(-1)}H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right\}\\
&+R_{n,5},
\end{aligned} \tag{A.1.9}$$

where $\left\|R_{n,5}\right\|=\left(h_2^{-1}\delta_{h,n}\right)^2+O_p\left(\sum_{l=1}^{d_x}h_{1,l}^{R+1}+\sqrt{\frac{1}{n}}\right)+o_p\left(K^{-\alpha}\right)$.

Given the Assumption 6 (3),

$$\left\|\bar{\beta}-\beta_0\right\|=O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right),$$

and

$$\max_{1\le l\le d_x}\sup_{x_l\in\mathcal{X}_l}\left|\bar{g}_l\left(x_l\right)-g_l\left(x_l\right)\right|=\sqrt{K}O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right),$$

where $\sqrt{\frac{K}{n}}$ is the convergence rate and $K^{-\alpha}+\sqrt{K}h_2^{a_H}$ are the order of bias.

A.2 Proof of Theorem 2.3.2

By Taylor Expansion, given the Assumption 3(1) ,

$$\begin{aligned}
&F_j\\
=&F\left[g(X_j)\right]\\
=&\sum_{d=0}^{S+2}\frac{1}{d!}\partial^d F_i\left[g(X_j)-g(X_i)\right]^d+O\left(\partial^{S+3}F_i\left[g(X_j)-g(X_i)\right]^{S+3}\right)\\
=&F_i+\partial F_i\left[\bar{g}(X_j)-\bar{g}(X_i)\right]-\partial F_i\left[\bar{g}(X_j)-g(X_j)\right]+\partial F_i\left[\bar{g}(X_i)-g(X_i)\right]\\
+&\frac{1}{2!}\partial^2 F_i\left[\bar{g}(X_j)-\bar{g}(X_i)\right]^2+\frac{1}{2!}\partial^2 F_i\left[\bar{g}(X_j)-g(X_j)\right]^2+\frac{1}{2!}\partial^2 F_i\left[\bar{g}(X_i)-g(X_i)\right]^2\\
+&\frac{1}{3!}\partial^3 F_i\left[\bar{g}(X_j)-\bar{g}(X_i)\right]^3+\frac{1}{S!}\partial^{h_3} F_i\left[\bar{g}(X_j)-\bar{g}(X_i)\right]^{h_3}+\frac{1}{(S+1)!}\partial^{S+1} F_i\left[g(X_j)-g(X_i)\right]^{S+1}\\
+&\frac{1}{(S+2)!}\partial^{S+2} F_i\left[g(X_j)-g(X_i)\right]^{S+2}+R_{n,2,i,j}\\
=&\sum_{d=0}^{S}\partial^d F_i\left[\bar{g}(X_j)-\bar{g}(X_i)\right]^d+\frac{1}{(S+1)!}\partial^{S+1} F_i\left[g(X_j)-g(X_i)\right]^{S+1}+\frac{1}{(S+2)!}\partial^{S+2} F_i\left[g(X_j)-g(X_i)\right]^{S+2}\\
+&R_{n,2,i,j}-\partial F_i\left[\bar{g}(X_j)-g(X_j)\right]+\partial F_i\left[\bar{g}(X_i)-g(X_i)\right]+\frac{1}{2!}\partial^2 F_i\left[\bar{g}(X_j)-g(X_j)\right]^2+\frac{1}{2!}\partial^2 F_i\left[\bar{g}(X_i)-g(X_i)\right]^2,
\end{aligned}\tag{A.2.1}$$

where $R_{n,2,i,j}=o_p\left(\left[g(X_j)-g(X_i)\right]^{S+3}+\left[\bar{g}(X_j)-g(X_j)\right]^2+\left[\bar{g}(X_i)-g(X_i)\right]^2\right)$ and then

$$\begin{pmatrix}1&0&\cdots&0\\0&s&\cdots&0\\0&0&\cdots&s^R\end{pmatrix}\begin{pmatrix}\widehat{F}_i-F_i\\ \partial\widehat{F}_i-\partial F_i\\ \cdots\\ (R!)^{-1}\left(\partial^R\widehat{F}_i-\partial^R F_i\right)\end{pmatrix}$$

$$
=\left\{\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1 & \frac{\triangle\bar{g}_{i,j}}{h_3} & \cdots & \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\\ \frac{\triangle\bar{g}_{i,j}}{h_3} & \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{2} & \cdots, & \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{S+1}\\ \vdots & \vdots & & \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3} & \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{S+1} & \cdots & \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{2S}\end{pmatrix}\right\}^{-1}
$$

$$
\times\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}\left\{-\partial F_i\left[\bar{g}(X_j)-g(X_j)\right]+\partial F_i\left[\bar{g}(X_i)-g(X_i)\right]+\frac{1}{2}\partial^2F_i\left[\bar{g}(X_j)-g(X_j)\right]^2\right.
$$

$$
\left.+\frac{1}{2!}\partial^2F_i\left[\bar{g}(X_i)-g(X_i)\right]^2+\frac{1}{(S+1)!}\partial^{S+1}F_i\left[g(X_j)-g(X_i)\right]^{S+1}+\frac{1}{(S+1)!}\partial^{S+2}F_i\left[g(X_j)-g(X_i)\right]^{S+2}+e_i+R_{n,2,i,j}\right\}
$$

$$
=A_1^{-1}B_1. \tag{A.2.2}
$$

Before we derive the asymptotic properties of B_1, we first show the asymptotic properties of some important components in B_1. Let

$$
B_2=\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\left[\bar{g}(X_j)-g(X_j)\right]^2,
$$

and

$$
B_3=\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\left[\bar{g}(X_j)-g(X_j)\right]^4.
$$

Similar with ((A.1.4)), by Taylor Expansion,

$$
\begin{aligned}
&T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\\
=&T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)+h_3^{-1}\partial T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\bar{g}\left(X_j\right)-g\left(X_j\right)\right]-h_3^{-1}\left[\bar{g}\left(X_i\right)-g\left(X_i\right)\right]\partial T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\\
&+h_3^{-1}\partial^2T_{h_3}\left[g\left(X_j\right)-g\left(X_i\right)\right]\left[\bar{g}\left(X_j\right)-g\left(X_j\right)\right]^2+h_3^{-2}\left[\bar{g}\left(X_i\right)-g\left(X_i\right)\right]^2\partial^2T_{h_3}\left[g\left(X_j\right)-g\left(X_i\right)\right]+R_{n,1,i,j}
\end{aligned} \tag{A.2.3}
$$

holds uniformly over $i\neq j\in\{1,\cdots,n\}$, where

$$
R_{n,1,i,j}=h_3^{-2}o_p\left(\partial^2T_{h_3}\left[g\left(X_j\right)-\bar{\bar{g}}\left(X_i\right)\right]\left[\bar{g}\left(X_j\right)-g\left(X_j\right)\right]^2+\left[\bar{g}\left(X_i\right)-g\left(X_i\right)\right]^2\partial^2T_{h_3}\left[g\left(X_j\right)-g\left(X_i\right)\right]\right).
$$

$$
\begin{aligned}
&B_2\\
=&\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\left[\bar{g}(X_j)-g(X_j)\right]^2
\end{aligned}
$$

$$
\begin{aligned}
=&\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\bar{g}(X_j)-g(X_j)\right]^2+h_3^{-1}\frac{1}{n}\sum_{j=1,\neq i}^{n}\partial T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\bar{g}\left(X_j\right)-g\left(X_j\right)\right]^3\\
&-h_3^{-1}\left[\bar{g}\left(X_i\right)-g\left(X_i\right)\right]\frac{1}{n}\sum_{j=1,\neq i}^{n}\partial T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\bar{g}(X_j)-g(X_j)\right]^2+h_3^{-2}\frac{1}{n}\sum_{j=1,\neq i}^{n}\partial^2 T_{h_3}\left[g\left(X_j\right)-g\left(X_i\right)\right]\left[\bar{g}\left(X_j\right)-g\left(X_j\right)\right]\\
&+h_3^{-2}\left[\bar{g}\left(X_i\right)-g\left(X_i\right)\right]^2\frac{1}{n}\sum_{j=1,\neq i}^{n}\partial^2 T_{h_3}\left[g\left(X_j\right)-g\left(X_i\right)\right]\left[\bar{g}(X_j)-g(X_j)\right]^2+\frac{1}{n}\sum_{j=1,\neq i}^{n} R_{n,1,i,j}\left[\bar{g}(X_j)-g(X_j)\right]^2\\
=&B_{2,1}+B_{2,2}-B_{2,3}+B_{2,4}+B_{2,5}+B_{2,6}. \qquad \text{(A.2.4)}
\end{aligned}
$$

We decompose B_1 as the following,

$$
\begin{aligned}
&B_1\\
&=-\partial F_i\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}\left[\bar{g}(X_j)-g(X_j)\right]\\
&+\partial F_i\left[\bar{g}(X_i)-g(X_i)\right]\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}\\
&+\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}\left[\bar{g}(X_j)-g(X_j)\right]^2\\
&+\frac{1}{2!}\partial^2 F_i\left[\bar{g}(X_i)-g(X_i)\right]^2\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}
\end{aligned}
$$

$$+\frac{1}{(S+1)!}\partial^{S+1}F_i\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}[g(X_j)-g(X_i)]^{S+1}$$

$$+\frac{1}{(S+2)!}\partial^{S+2}F_i\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}[g(X_j)-g(X_i)]^{S+2}$$

$$+\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}e_i+\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots\\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}R_{n,2,i,j}$$

$$=B_{1,1}+B_{1,2}+B_{1,3}+B_{1,4}+B_{1,5}+B_{1,6}+B_{1,7}+B_{1,8}. \tag{A.2.5}$$

By (A.2.2), Assumption 3, Lemma (A.4.5) and (A.4.6), there exists positive constant C_1 such that

$$\widehat{F}_i-F_i=C_1\left[\bar{g}(X_i)-g(X_i)\right]+O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}+h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}\right),$$

hold uniformly over $i\in\{1,\cdots,n\}$; and there exists positive constant C_2 such that

$$\partial\widehat{F}_i-\partial F_i=h_3^{-1}C_2\left[\bar{g}(X_i)-g(X_i)\right]+h_3^{-1}O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}+h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}\right),$$

hold uniformly over $i\in\{1,\cdots,n\}$.

Thus, (2.3.3) and (2.3.4) in Theorem 2.3.2 are proved.

A.3 Proof of Theorem 2.3.3

Given the Taylor Expansion that, for $i\in\{1,\cdots,n\}$,

$$\begin{aligned}\bar{g}_l\left(X_{l,i}\right)=&g_l\left(X_{l,i}\right)+\bar{g}_l\left(X_{l,i}\right)-g_l\left(X_{l,i}\right)\\ =&\bar{g}_l\left(X_{l,i}\right)-g_l\left(X_{l,i}\right)+g_l\left(x_l\right)+\partial g_l\left(x_l\right)\left(X_{l,i}-x_l\right)+\frac{1}{2}\partial^2 g_l\left(x_l\right)\left(X_{l,i}-x_l\right)^2+\frac{1}{6}\partial^2 g_l\left(\bar{X}_{l,i}\right)\left(X_{l,i}-x_l\right)^3,\end{aligned}$$

where $\bar{X}_{l,i}$ is between $X_{l,i}$ and x_l, and then

$$Y_i-\widehat{F}_i+\widehat{\partial F}_i\bar{g}_l\left(X_{l,i}\right)=Y_i-\widehat{F}_i+\widehat{\partial F}_i\left[\bar{g}_l\left(X_{l,i}\right)-g_l\left(X_{l,i}\right)\right]+\widehat{\partial F}_i\left[g_l\left(x_l\right)+\partial g_l\left(x_l\right)\left(X_{l,i}-x_l\right)\right]+\widehat{\partial F}_i\frac{1}{2}\partial^2 g_l\left(x_l\right)\left(X_{l,i}-x_l\right)^2$$

$$+\widehat{\partial F}_i \frac{1}{6}\partial^2 g_l\left(\bar{X}_{l,i}\right)\left(X_{l,i}-x_l\right)^3$$

$$=\widehat{\partial F}_i\left[g_l\left(x_l\right)+\partial g_l\left(x_l\right)\left(X_{l,i}-x_l\right)\right]+e_i+\widehat{\partial F}_i\left[\bar{g}_l\left(X_{l,i}\right)-g_l\left(X_{l,i}\right)\right]+\left(F_i-\widehat{F}_i\right)$$

$$+\widehat{\partial F}_i \frac{1}{2}\partial^2 g_l\left(x_l\right)\left(X_{l,i}-x_l\right)^2+\widehat{\partial F}_i \frac{1}{6}\partial^2 g_l\left(\bar{X}_{l,i}\right)\left(X_{l,i}-x_l\right)^3. \tag{A.3.1}$$

We substitute (A.3.1) into (2.2.6) in Algorithm (2.2.1) and the numerator of $\left(\widehat{g}_l(x_l,\widehat{\partial g}_l(x_l)\right)'-(g_l(x_l),\partial g_l(x_l))'$ can be expressed as

$$\begin{pmatrix}1 & 0\\ 0 & h_{4,l}\end{pmatrix}^{-1}\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ X_{l,i}-x_l\end{pmatrix}\widehat{\partial F}_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left\{Y_i-\widehat{F}_i+\widehat{\partial F}_i\left[\bar{g}_l\left(X_{l,i}\right)+g_l\left(x_l\right)+\partial g_l\left(x_l\right)\left(X_{l,i}-x_l\right)\right]\right\}$$

$$=\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\widehat{\partial F}_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)e_i+\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\widehat{\partial F}_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(F_i-\widehat{F}_i\right)$$

$$+\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left[\bar{g}_l\left(X_{l,i}\right)-g_l\left(X_{l,i}\right)\right]$$

$$+\frac{1}{2}\partial^2 g_l\left(x_l\right)\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(X_{l,i}-x_l\right)^2$$

$$+\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\frac{1}{6}\partial^2 g_l\left(\bar{X}_{l,i}\right)\left(X_{l,i}-x_l\right)^3$$

$$=B_{4,1}+B_{4,2}+B_{4,3}+B_{4,4}+B_{4,5}.$$

Therefore,

$$\begin{pmatrix}1 & 0\\ 0 & h_{4,l}\end{pmatrix}\left[\begin{pmatrix}\widehat{g}_l(x_l)\\ \widehat{\partial g}_l(x_l)\end{pmatrix}-\begin{pmatrix}g_l(x_l)\\ g_l(x_l)\end{pmatrix}\right]$$

$$=\left\{\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left(1,\frac{X_{l,i}-x_l}{h_{4,l}}\right)\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\right\}^{-1}$$

$$\times\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\widehat{\partial F}_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(Y_i-\widehat{F}_i+\widehat{\partial F}_i\bar{g}_l\left(X_{l,i}\right)\right)$$

$$=A_4^{-1}\left(B_{4,1}+B_{4,2}+B_{4,3}+B_{4,4}+B_{4,5}\right). \tag{A.3.2}$$

Thus, by ((A.4.59)), ((A.4.60)) and ((A.4.61)), ((A.4.62)), ((A.4.63)) in Lemma (A.4.7),

$$\begin{pmatrix}1 & 0\\ 0 & h_{4,l}\end{pmatrix}\left[\begin{pmatrix}\widehat{g}_l(x_l)\\ \widehat{\partial g}_l(x_l)\end{pmatrix}-\begin{pmatrix}g_l(x_l)\\ g_l(x_l)\end{pmatrix}\right]$$

$$= \left\{ \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \left(1, \frac{X_{l,i} - x_l}{h_{4,l}}\right) (\partial F_i)^2 T_{h_{4,l}} (X_{l,i} - x_l) \right\}^{-1}$$

$$\times \left\{ \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \partial F_i T_{h_{4,l}} (X_{l,i} - x_l) e_i + \frac{1}{2} \partial^2 g_l (x_l) \frac{1}{n} \sum_{i=1}^{n} \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \partial F_i^2 T_{h_{4,l}} (X_{l,i} - x_l) (X_{l,i} - x_l)^2 \right\}$$

$$+ R_{n,3} (x_l), \tag{A.3.3}$$

where $R_{n,3} (x_l) = o_p \left(\sqrt{\frac{1}{nh_{4,l}}} + h_{4,l}^2 \right) + O_p \left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H} + h_3^{S+1} + \sqrt{\frac{\log n}{nh_3}} \right)$ holds for each $x_l \in \mathcal{X}_l$ and $R_{n,3} (x_l) = o_p \left(\sqrt{\frac{\log n}{nh_{4,l}}} + h_{4,l}^2 \right) + O_p \left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H} + h_3^{S+1} + \sqrt{\frac{\log n}{nh_3}} \right)$ holds uniformly over $x_l \in \mathcal{X}_l$;

given the Assumption 6 (4) , and then,

$$\sqrt{nh_{4,l}} \left(\begin{pmatrix} 1 & 0 \\ 0 & h_{4,l} \end{pmatrix} \left[\begin{pmatrix} \widehat{g}_l(x_l) \\ \widehat{\partial g_l}(x_l) \end{pmatrix} - \begin{pmatrix} g_l(x_l) \\ g_l(x_l) \end{pmatrix} \right] - \left(E \left\{ \partial F [g(X)]^2 \middle| X_l = x_l \right\} \begin{pmatrix} \int T(\omega) d\omega & 0 \\ 0 & \int \omega^2 T(\omega) d\omega \end{pmatrix} \right) \right)^{-1}$$

$$\times \left(\begin{matrix} h_{4,l}^2 \frac{1}{2} \partial^2 g_l (x_l) E \left\{ \partial F [g(X)]^2 \middle| X_l = x_l \right\} \kappa_{21} \\ 0 \end{matrix} \right) + o_p(h_{4,l}^2) \Bigg)$$

$$\xrightarrow{d} \left(E \left\{ [\partial F (g(X))]^2 \middle| X_l = x_l \right\} \begin{pmatrix} \int T(\omega) d\omega & 0 \\ 0 & \int \omega^2 T(\omega) d\omega \end{pmatrix} \right)^{-1}$$

$$\times N \left(0, E \left\{ E \left(e^2 \middle| X \right) \partial F [g(X)]^2 \middle| X_l = x_l \right\} \begin{pmatrix} \kappa_{02} & 0 \\ 0 & \int \omega^2 T(\omega)^2 d\omega \end{pmatrix} \right);$$

and thus,

$$\sqrt{nh_{4,l}} \begin{pmatrix} 1 & 0 \\ 0 & h_{4,l} \end{pmatrix} \left(\begin{pmatrix} \widehat{g}_l(x_l) \\ \widehat{\partial g_l}(x_l) \end{pmatrix} - \begin{pmatrix} g_l(x_l) \\ g_l(x_l) \end{pmatrix} - \begin{pmatrix} \frac{1}{2} h_{4,l}^2 \kappa_{21} \partial^2 g_l(x_l) \\ 0 \end{pmatrix} \right)$$

$$\xrightarrow{d} N \left(0, \frac{E \left\{ E \left(e^2 \middle| X \right) \partial F [g(X)]^2 \middle| X_l = x_l \right\}}{\left(E \left\{ [\partial F (g(X))]^2 \middle| X_l = x_l \right\} \right)^2} \begin{pmatrix} \kappa_{02} & 0 \\ 0 & \frac{\kappa_{22}}{\kappa_{21}^2} \end{pmatrix} \right).$$

A.4 Technical Lemmas

A.4.1 Lemmas for the proof of Theorem 2.3.1

Lemma A.4.1 *Suppose that Assumptions 2, 3 and 4 (2)hold. Then,*

$$\max_{j\in\{1,\cdots,n\}}\left|\bar{E}\left(Y\left|X_j\right.\right)-E\left(Y\left|X_j\right.\right)\right|=O_p\left(\sum_{l=1}^{d_x}h_{1,l}^{R+1}+\sqrt{\frac{\log n}{n\prod_{l=1}^{d_x}h_{1,l}}}\right). \tag{A.4.1}$$

Proof. Given Assumptions 2 (1) and 3, we have the Taylor Expansion that , for $i\in\{1,\cdots,n\}$,

$$\begin{aligned}Y_i=&E\left(Y|\,X_i\right)+e_i=E\left(Y|\,X_j\right)+\partial E\left(Y|\,X_j\right)'\triangle X_{ij}+\cdots+\frac{1}{R!}\partial^R E\left(Y|\,X_j\right)\cdot\left(\triangle X_{ij}\right)^{\otimes R}+e_i\\&+\frac{1}{(R+1)!}\partial^{R+1}E\left(Y|\,X_j\right)\cdot\left(\triangle X_{ij}\right)^{\otimes R+1}+\frac{1}{(R+2)!}\partial^{R+2}E\left(Y|\,X_j\right)\cdot\left(\triangle X_{ij}\right)^{\otimes R+2}\\&+\frac{1}{(R+3)!}\partial^{R+3}E\left(Y|\,\bar{X}_{i,j}\right)\cdot\left(\triangle X_{ij}\right)^{\otimes R+3},\end{aligned}$$

where $\bar{X}_{i,j}$ is between X_i and X_j.

And then, by equation ((2.2.2)), for $i\in\{1,\cdots,n\}$,

$$\begin{aligned}&\left(\bar{E}\left(Y\left|X_j\right.\right)-E\left(Y\left|X_j\right.\right),diag(r)\left[\partial\bar{E}\left(Y\left|X_j\right.\right)'-\partial E\left(Y\left|X_j\right.\right)'\right],\cdots,\frac{diag(r^{\otimes R})}{R!}\left[\partial^R\bar{E}\left(Y\left|X_j\right.\right)'-\partial^R E\left(Y\left|X_j\right.\right)'\right]\right)\\
=&\left[\frac{\sum_{i=1,\neq j}^n T_{h_1}(i,j)}{n-1}\begin{pmatrix}1 & \frac{X_{1,i}-X_{1,j}}{h_{1,l}} & \cdots & \left(\frac{X_{d_x,i}-X_{d_x,j}}{h_{1,d_x}}\right)^R\\ \frac{X_{1,i}-X_{1,j}}{h_{1,1}} & \left(\frac{X_{1,i}-X_{1,j}}{h_{1,1}}\right)^2 & \cdots & \left(\frac{X_{1,i}-X_{1,j}}{h_{1,1}}\right)\left(\frac{X_{d_x,i}-X_{d_x,j}}{h_{1,d_x}}\right)^R\\ \vdots & \vdots & & \vdots\\ \left(\frac{X_{d_x,i}-X_{d_x,j}}{h_{1,d_x}}\right)^R & \left(\frac{X_{d_x,i}-X_{d_x,j}}{h_{1,d_x}}\right)^R\left(\frac{X_{1,i}-X_{1,j}}{h_{1,1}}\right) & \cdots & \left(\frac{X_{d_x,i}-X_{d_x,j}}{h_{1,d_x}}\right)^{2R}\end{pmatrix}\right]^{-1}\\
&\times\frac{\sum_{i=1,\neq j}^n T_{h_1}(i,j)}{n-1}\left(1\quad\frac{X_{1,i}-X_{1,j}}{h_{1,1}}\quad\cdots\quad\frac{X_{d_x,i}-X_{d_x,j}}{h_{1,d_x}}\quad\cdots\quad\left(\frac{X_{1,i}-X_{1,j}}{h_{1,1}}\right)^R\quad\cdots\quad\left(\frac{X_{d_x,i}-X_{d_x,j}}{h_{1,d_x}}\right)^R\right)'\\
&\times\left[e_i+\frac{\partial^{R+1}E\left(Y|\,X_j\right)\cdot\left(\triangle X_{ij}\right)^{\otimes R+1}}{(R+1)!}+\frac{\partial^{R+2}E\left(Y|\,X_j\right)\cdot\left(\triangle X_{ij}\right)^{\otimes R+2}}{(R+2)!}+\frac{\partial^{R+3}E\left(Y|\,\bar{X}_{i,j}\right)\cdot\left(\triangle X_{ij}\right)^{\otimes R+3}}{(R+3)!}\right],\end{aligned} \tag{A.4.2}$$

Given ((A.4.2)) and the Assumption 4, it is straightforward to have

$$\max_{j\in\{1,\cdots,n\}}\left|\bar{E}\left(Y\left|X_j\right.\right)-E\left(Y\left|X_j\right.\right)\right|=O\left(\sum_{l=1}^{d_x}h_{1,l}^{R+1}+\sqrt{\frac{\log n}{n\prod_{l=1}^{d_x}h_{1,l}}}\right). \tag{A.4.3}$$

To simplify the notation, let

$$\delta_{h,n}=O_p\left(\sum_{l=1}^{d_x}h_{1,l}^{R+1}+\sqrt{\frac{\log n}{n\prod_{l=1}^{d_x}h_{1,l}}}\right).$$

■

Lemma A.4.2 *Suppose that Assumptions 2, 3, 4 (2)and 5 (3), 6 (1)-(3) hold. Then*

(i)

$$\left\| L_{0,n,1} - E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right) = 0 \right] f_{E(Y|X_j)-E(Y|X_i)}(0) \right\| = O_p\left(\sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H} \right). \tag{A.4.4}$$

(ii)

$$\left\| L_{0,n,2} \right\| = \sqrt{K} h_2^{-1} \delta_{h,n}, \tag{A.4.5}$$

and

$$\left\| L_{0,n,3} \right\| = \sqrt{K} \left(h_2^{-1} \delta_{h,n} \right)^2. \tag{A.4.6}$$

(iii)

$$\left\| L_{0,n} - E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right) = 0 \right] f_{E(Y|X_j)-E(Y|X_i)}(0) \right\| = o_p(1). \tag{A.4.7}$$

and

$$\lambda_{\min}\left(L_{0,n}\right) = O_p(1). \tag{A.4.8}$$

Proof. (i) It is trivial to prove that

$$\left\| L_{0,n,1} - E\left\{ \triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right) \right] \right\} \right\| = O_p\left(\sqrt{\frac{K}{n}} \right). \tag{A.4.9}$$

By Taylor Expansion,

$$\begin{aligned}
&E\left\{ \triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right) \right] \right\} \\
=&E\left\{ E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right) \right] H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right) \right] \right\} \\
=&E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right) \right] f_{E(Y|X_i)}(E\left(Y\left|X_j\right.\right)) \\
+&\sum_{d=1}^{a_H} \frac{d!\,(a_H-d)!}{a_H} \partial^l E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right) \right] \partial^{a_H-d} f_{E(Y|X_i)}(E\left(Y\left|X_j\right.\right)) O(h_2^{a_H}),
\end{aligned}$$

and then

$$\begin{aligned}
&\Bigg\| E\left\{ \triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right) \right] \right\} \\
&- E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right) \right] f_{E(Y|X_i)}(E\left(Y\left|X_j\right.\right)) \Bigg\| \\
\leq&O(h_2^{a_H}) \left\| \sum_{d=1}^{a_H} \frac{d!\,(a_H-d)!}{a_H} \partial^l E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \middle| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right) \right] \partial^{a_H-d} f_{E(Y|X_i)}(E\left(Y\left|X_j\right.\right)) \right\|
\end{aligned}$$

$$(A.4.10)$$

holds uniformly over $i \neq j \in \{1, \cdots, n\}$, where $f_{E(Y|X_j)-E(Y|X_i)}(\cdot)$ is the density function of $E(Y|X_j) - E(Y|X_i)$, given the Assumption 5 (3). Since $f_{E(Y|X_i)}(\cdot)$ is bounded, there exists a positive constant c such that

$$\begin{aligned}
&\left\|\partial^{a_H} E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E(D|X_j) = E(Y|X_i)\right] f_{E(Y|X_i)}(E(D|X_j))\Big\|^2 \\
\leq & c\left\|\partial^{a_H} E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E(D|X_j) = E(Y|X_i)\right]\Big\|^2 \\
\leq & O(K)\,\lambda_{\max}^2 \left\{\partial^{a_H} E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E(D|X_j) = E(Y|X_i)\right]\right\}
\end{aligned}$$

Given the Assumptions 5 (3) and 6 (2) and (3) ,

$$\begin{aligned}
&\left\|\partial^{d} E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E(Y|X_j) - E(Y|X_i) = 0\right] \partial^{a_H - d} f_{E(Y|X_j)-E(Y|X_i)}(0)\Big\|\, O\left(h_2^{a_H}\right) \\
= & O\left(\sqrt{K} h_2^{a_H}\right), \qquad (A.4.11)
\end{aligned}$$

holds uniformly over $d = 1, \cdots, a_H$. Given the Assumption 6 (1) and (A.4.11)

$$\begin{aligned}
&\Big\|E\left\{\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\right\} \\
-& E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E(Y|X_j) - E(Y|X_i) = 0\right] f_{E(Y|X_j)-E(Y|X_i)}(0)\Big\| \\
= & O\left(\sqrt{K} h_2^{a_H}\right) = o(1). \qquad (A.4.12)
\end{aligned}$$

Thus, by ((A.4.9)) and ((A.4.12)),

$$\left\|L_{0,n,1} - E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E(Y|X_j) - E(Y|X_i) = 0\right] f_{E(Y|X_j)-E(Y|X_i)}(0)\Big\| = O_p\left(\sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H}\right),$$

and then (A.4.4) holds.

(ii) Given the Assumption 6 (3),

$$\begin{aligned}
&\|L_{0,n,2}\| \\
\leq & h_2^{-1}\delta_{h,n}\left\|\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}\left|\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} \partial H_{h_2}\left[E(Y|X_j) - E(Y|X_i)\right]\right|\right\| \\
= & \sqrt{K} h_2^{-1}\delta_{h,n}, \qquad (A.4.13)
\end{aligned}$$

and then (A.4.5) holds. It is trivial to prove (A.4.6).

(iii) Therefore, given the Assumption 6 (2) , $\|L_{0,n,2}\| = o_p(1)$, $\|L_{0,n,3}\| = o_p(1)$ and then (A.4.7) holds. Furthermore, we are given the Assumption 6 (3),

$$\lambda_{\min}\left(E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E(Y|X_j) - E(Y|X_i) = 0\right] f_{E(Y|X_j)-E(Y|X_i)}(0)\Big) = O(1). \qquad (A.4.14)$$

and then (A.4.8) holds. ■

Lemma A.4.3 *Suppose that Assumptions 2, 3, 4 (2)and 5 (3), 6 (1)-(3) hold. Then*

(i)

$$\left\|L_{0,n,1}^{-1}L_{1,n,1}\right\| = O_p\left(\sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H}\right). \tag{A.4.15}$$

(ii)

$$\|L_{1,n,2}\| = O_p\left(\sum_{l=1}^{d_x} h_{1,l}^{R+1} + \sqrt{\frac{1}{n}}\right). \tag{A.4.16}$$

(iii)

$$\|L_{1,n,3}\| = O_p\left(\sum_{l=1}^{d_x} h_{1,l}^{R+1} + \sqrt{\frac{1}{n}}\right), \tag{A.4.17}$$

and

$$\|L_{1,n,4}\| = \left(h_2^{-1}\delta_{h,n}\right)^2. \tag{A.4.18}$$

(iv)

$$\|L_{1,n}\| = O_p\left(\sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H} + \left(h_2^{-1}\delta_{h,n}\right)^2 + \sum_{l=1}^{d_x} h_{1,l}^{R+1} + \sqrt{\frac{1}{n}}\right). \tag{A.4.19}$$

(v)

$$\|L_{2,n}\| = O_p\left(K^{-\alpha}\right). \tag{A.4.20}$$

Proof. (i) Given the fact that $tr(AB) \leq \lambda_{\max}(B)tr(A)$ for any symmetric matrix B and positive semidefinite matrix A,

$$\left\|L_{0,n,1}^{-1}L_{1,n,1}\right\|^2 = tr\left\{L_{0,n,1}^{-1}L_{1,n,1}L_{1,n,1}{}'L_{0,n,1}^{-1}\right\} = O\left(\lambda_{\min}^{-2}\left(L_{0,n,1}\right)\right)tr\left\{L_{1,n,1}L_{1,n,1}{}'\right\} = O(1)\left\|L_{1,n,1}\right\|^2, \tag{A.4.21}$$

the last equality is by ((A.4.8)). Given the Assumption 3 (2), the link function $F_{\triangle}(\cdot)$ is monotonically increasing and thus, $g(X_i)$ is also identified given $E\left(Y\left|X_i\right.\right) = F_{\triangle}\left(g(X_i)\right)$. $L_{1,n,1} = \frac{1}{n^2}\sum_{i\neq j} g(X_j)\triangle P_{i,j}^{(-1)}H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\right.$ is a vector of U-statistics and then we apply Lemma 3.1 in Powell, Stock and Stoker (1989) to $L_{1,n,1}$. Let

$$q_n(X_i, X_j) = g(X_j)\triangle P_{i,j}^{(-1)}H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right],$$

$$U_n = \binom{n}{2}^{n-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} q_n(X_i, X_j),$$

and

$$L_{1,n,1} = \frac{\binom{n}{2}}{n^2} U_n. \tag{A.4.22}$$

Also define

$$r_n(X_j) = E\left[q_n(X_i, X_j)\right| X_j\right],$$

$$\theta_n = E\left[q_n(X_j)\right],$$

and

$$\widehat{U}_n = \theta_n + \frac{2}{n}\sum_{i=1}^{n}\left[r_n(X_j) - \theta_n\right].$$

Before we apply Lemma 3.1 in Powell et al. (1989) to $\widehat{U}_n$, we need to verify the condition $E\left\|q_n(X_i, X_j)\right\|^2 = o(n)$. By Assumption 6 (3), we have

$$\begin{aligned}
& E\left\|q_n(X_i, X_j)\right\|^2 \\
=& E\ tr\left\{g(X_j)^2 \triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime} H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right]^2\right\} \\
=& O(K) = o(n),
\end{aligned}$$

and by the proof of Lemma 3.1 in Powell et al. (1989),we have $E\left\|\widehat{U}_n - U_n\right\|^2 = o(n^{-1})$. By Markov Inequality, we have

$$\left\|\widehat{U}_n - U_n\right\| = o_p(n^{-1/2}). \tag{A.4.23}$$

To derive the convergence rate of $\widehat{U}_n$, we need to calculate $E\left\|r_n\left(X_j\right) - \theta_n\right\|^2$. We take conditional expectation of $q_n\left(X_i, X_j\right)$,

$$\begin{aligned}
& r_n\left(X_j\right) \\
=& E\left\{g(X_j) \triangle P_{i,j}^{(-1)} H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right]\right| X_j\right\} \\
=& E\left\{g(X_j)\left[P_j^{(-1)} - P_i^{(-1)}\right] H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right]\right| X_j\right\} \\
=& g(X_j) P_j^{(-1)} E\left\{H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right]\right| X_j\right\} \\
& - g(X_j) E\left\{P_i^{K-1,K} H_{h_2}\left[E\left(Y\left|X_j\right.\right) - E\left(Y\left|X_i\right.\right)\right]\right| X_j\right\} \\
=& g(X_j) P_j^{(-1)}\left\{f_{E(Y|X_i)}\left(E\left(Y\left|X_j\right.\right)\right) + \frac{h_2^{a_H}}{a_H!}\partial^{a_H} f_{E(Y|X_i)}\left(E\left(Y\left|X_j\right.\right)\right) + o(h_2^{a_H})\right\} \\
& - g(X_j)\left\{E\left[P_i^{(-1)}\right| E\left(Y\left|X_i\right.\right) = E\left(Y\left|X_j\right.\right)\right] f_{E(Y|X_i)}(E\left(Y\left|X_j\right.\right)) \\
& + h_2^{a_H}\sum_{l=1}^{a_H}\frac{l!\left(a_H - l\right)!}{a_H}\partial^l E\left[\triangle P_{i,j}^{(-1)} \triangle P_{i,j}^{(-1)\prime}\right| E\left(Y\left|X_j\right.\right) = E\left(Y\left|X_i\right.\right)\right]\partial^{a_H - l} f_{E(Y|X_i)}(E\left(Y\left|X_j\right.\right)) + o(h_2^{a_H})\right\}
\end{aligned}$$

$$
\begin{aligned}
&=g(X_j)\left\{P_j^{(-1)}-E\left[P_i^{(-1)}\middle| E(Y|X_i)=E(Y|X_j)\right]\right\}f_{E(Y|X_i)}(E(Y|X_j))\\
&+g(X_j)P_j^{(-1)}\left\{\frac{h_2^{a_H}}{a_H!}\partial^{a_H}f_{E(Y|X_i)}(E(Y|X_j))+o(h_2^{a_H})\right\}\\
&-g(X_j)\left\{h_2^{a_H}\sum_{l=1}^{a_H}\frac{l!\,(a_H-l)!}{a_H}\partial^l E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\middle| E(Y|X_j)=E(Y|X_i)\right]\partial^{a_H-l}f_{E(Y|X_i)}(E(Y|X_j))+o(h_2^{a_H})\right\},
\end{aligned}
$$

and we take expectation again,

$$
\begin{aligned}
&\theta_n\\
&=E\left\{g(X_j)\triangle P_{i,j}^{(-1)}H_{h_2}\left[E(Y|X_j)-E(Y|X_i)\right]\right\}\\
&=E\left(g(X_j)E\left[P_i^{(-1)}\middle| E(Y|X_i)=E(Y|X_j)\right]\left\{\frac{h_2^{a_H}}{a_H!}\partial^{a_H}f_{E(Y|X_i)}(E(Y|X_j))+o(h_2^{a_H})\right\}\right.\\
&-g(X_j)\left\{h_2^{a_H}\sum_{l=1}^{a_H}\frac{l!\,(a_H-l)!}{a_H}\partial^l E\left[\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}\middle| E(Y|X_j)=E(Y|X_i)\right]\partial^{a_H-l}f_{E(Y|X_i)}(E(Y|X_j))+o(h_2^{a_H})\right\},
\end{aligned}
$$

and then we get,

$$
\|\theta_n\|=O\left(\sqrt{K}h_2^{a_H}\right). \tag{A.4.24}
$$

Before we apply ChebyshevÂ’s inequality, we calculate

$$
\begin{aligned}
&E\left\|r_n(X_j)\right\|^2\\
&=E\left\|E\left\{g(X_j)\triangle P_{i,j}^{(-1)}H_{h_2}\left[E(Y|X_j)-E(Y|X_i)\right]\middle| X_j\right\}\right\|^2\\
&=E\left\|g(X_j)\left\{P^{K-1,K}(X_j)-E\left[P^{K-1,K}(X_i)\middle| E(Y|X_i)=E(Y|X_j)\right]\right\}f_{E(Y|X_i)}(E(Y|X_j))+O\left(h_2^{a_H}\right)\right\|^2\\
&=O\left[tr\left(E\left\{g(X_j)^2\left\{P_j^{(-1)}-E\left[P_i^{(-1)}\middle| E(Y|X_i)=E(Y|X_j)\right]\right\}\right.\right.\right.\\
&\times\left.\left.\left.\left\{P_j^{(-1)}-E\left[P_i^{(-1)}\middle| E(Y|X_i)=E(Y|X_j)\right]\right\}'f_{E(Y|X_i)}(E(Y|X_j))^2\right\}\right)\right]\\
&=O(K).
\end{aligned}
\tag{A.4.25}
$$

Given ((A.4.24)) and ((A.4.25)),

$$
E\left\|\widehat{U}_n-\theta_n\right\|^2=O\left(\frac{K}{n}\right),
$$

and by ChebyshevÂ’s inequality,

$$
\left\|\widehat{U}_n-\theta_n\right\|=O_p\left(\sqrt{\frac{K}{n}}\right), \tag{A.4.26}
$$

Therefore, given (A.4.21) and (A.4.26), we have $\left\|L_{0,n,1}^{-1}L_{1,n,1}\right\|=O_p\left(\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right)$ and then (A.4.15) holds.

(ii) By (A.4.2)

$$
L_{1,n,2}
$$

$$
\begin{aligned}
=&\frac{1}{n^2h_2}\sum_{j=2}^{n}\sum_{i<j}^{n} g\left(X_j\right)\triangle P_{i,j}^{(-1)}\partial H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\left[\bar{E}\left(Y\left|X_j\right.\right)-E\left(Y\left|X_j\right.\right)\right]\\
=&\frac{1}{n^2h_2}\sum_{j=2}^{n}\sum_{i<j}^{n} g\left(X_j\right)\triangle P_{i,j}^{(-1)}\partial H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\left[\bar{E}\left(Y\left|X_j\right.\right)-E\left(Y\left|X_j\right.\right)\right]\\
=&\frac{1}{n^2(n-1)h_2}\sum_{j=2}^{n}\sum_{i<j}^{n} g\left(X_j\right)\triangle P_{i,j}^{(-1)}\partial H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\\
&\left[\frac{\sum_{\iota=1,\neq j}^{n}T_{h_1}\left(\iota,j\right)}{n-1}\begin{pmatrix}1 & \frac{X_{1,\iota}-X_{1,j}}{h_{1,\iota}} & \cdots & \left(\frac{X_{d_x,\iota}-X_{d_x,j}}{h_{1,d_x}}\right)^R\\
\frac{X_{1,\iota}-X_{1,j}}{h_{1,1}} & \left(\frac{X_{1,\iota}-X_{1,j}}{h_{1,1}}\right)^2 & \cdots & \left(\frac{X_{1,\iota}-X_{1,j}}{h_{1,1}}\right)\left(\frac{X_{d_x,\iota}-X_{d_x,j}}{h_{1,d_x}}\right)^R\\
\vdots & \vdots & & \vdots\\
\left(\frac{X_{d_x,\iota}-X_{d_x,j}}{h_{1,d_x}}\right)^R & \left(\frac{X_{d_x,\iota}-X_{d_x,j}}{h_{1,d_x}}\right)^R\left(\frac{X_{1,\iota}-X_{1,j}}{h_{1,1}}\right) & \cdots & \left(\frac{X_{d_x,\iota}-X_{d_x,j}}{h_{1,d_x}}\right)^{2R}\end{pmatrix}\right]^{-1}\\
&\times T_{h_1}\left(\iota,j\right)\begin{pmatrix}1 & \frac{X_{1,\iota}-X_{1,j}}{h_{1,1}} & \cdots & \frac{X_{d_x,\iota}-X_{d_x,j}}{h_{1,d_x}} & \cdots & \left(\frac{X_{1,\iota}-X_{1,j}}{h_{1,1}}\right)^R & \cdots & \left(\frac{X_{d_x,\iota}-X_{d_x,j}}{h_{1,d_x}}\right)^R\end{pmatrix}'\\
&\times\left[e_\iota+\partial^{R+1}E\left(Y|\,X_j\right)\cdot\left(\triangle X_{\iota j}\right)^{\otimes R+1}+\partial^{R+2}E\left(Y|\,X_j\right)\cdot\left(\triangle X_{\iota j}\right)^{\otimes R+2}+R_{n,5,\iota,j}\right];
\end{aligned}
$$

although the summation is over $j\in\{1,\cdots,n\}$, $i<j\in\{1,\cdots,n\}$ and $\iota\neq j\in\{1,\cdots,n\}$, the asymptotic properties should be the same if the summation is taken over $j\in\{1,\cdots,n\}$, $i<j\in\{1,\cdots,n\}$ and $\iota\neq i,j\in\{1,\cdots,n\}$ and then it is a U-statistic. Similar with the analysis of (A.4.15), by asymptotic properties of U-statistic, (A.4.16) holds.

(iii) It is trivial to prove that (A.4.17) holds and

$$
\begin{aligned}
&\left\|L_{1,n,4}\right\|\\
\leq&\left(h_2^{-1}\delta_{h,n}\right)^2\frac{1}{n^2}\sum_{j=2}^{n}\sum_{i<j}^{n}\left\|g\left(X_j\right)\triangle P_{i,j}^{(-1)}\partial^2H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right\|\\
=&\left(h_2^{-1}\delta_{h,n}\right)^2,
\end{aligned}
$$

which shows (A.4.18) holds.

(iv) Thus, by ((A.4.15)), ((A.4.16)) and ((A.4.17)), ((A.4.18)) , (A.4.19) holds.

(v) Similar with (A.4.21),

$$
\begin{aligned}
&\left\|L_{0,n,1}^{-1}L_{2,n,1}\right\|^2\\
=&tr\left[\left(vector_2'vector_2\right)^{-1}vector_2'vectoh_{1,1}vectoh_{1,1}'vector_2\left(vector_2'vector_2\right)^{-1}\right]\\
\leq&\lambda_{\min}\left(vector_2'vector_2\right)tr\left(vector_2'vectoh_{1,1}vectoh_{1,1}'vector_2\left(vector_2'vector_2\right)^{-1}\right)\\
=&\lambda_{\min}^{-1}\left(vector_2'vector_2\right)tr\left(vectoh_{1,1}vectoh_{1,1}'\right)\\
=&\lambda_{\min}^{-1}\left(vector_2'vector_2\right)\left\|vectoh_{1,1}\right\|^2
\end{aligned}
$$

$$=\lambda_{\min}^{-1}\left(E\left\{\triangle P_{i,j}^{(-1)}\triangle P_{i,j}^{(-1)\prime}H_{h_2}\left[E\left(Y\left|X_j\right.\right)-E\left(Y\left|X_i\right.\right)\right]\right\}\right)O_p\left(K^{-2\alpha}\right). \tag{A.4.27}$$

It is trivial to prove that $\left\|L_{0,n,1}^{-1}L_{2,n,2}\right\|^2=o_p\left(\left\|L_{0,n,1}^{-1}L_{2,n,1}\right\|^2\right)$ and $\left\|L_{0,n,1}^{-1}L_{2,n,3}\right\|^2=o_p\left(\left\|L_{0,n,1}^{-1}L_{2,n,1}\right\|^2\right)$. Thus, $\left\|L_{0,n,1}^{-1}L_{2,n,1}\right\|=O_p\left(K^{-\alpha}\right)$ and (A.4.20) holds. ■

A.4.2 Lemmas for the proof of Theorem 2.3.2

Lemma A.4.4 *Suppose that Assumptions 1 - 4, 5 (1) and (3) and 6 (1)-(3) hold. Then*

(i)

$$B_{2,1}=O_p\left(K^{-2\alpha}+\frac{K}{n}+Kh_2^{2a_H}\right). \tag{A.4.28}$$

(ii)

$$\left|B_{2,2}\right|=O_p\left(K^{-2\alpha}+\frac{K}{n}+Kh_2^{2a_H}\right), \tag{A.4.29}$$

$$\left|B_{2,3}\right|,\left|B_{2,4}\right|,\left|B_{2,5}\right|=O_p\left(K^{-2\alpha}+\frac{K}{n}+Kh_2^{2a_H}\right) \tag{A.4.30}$$

and

$$\left|B_{2,6}\right|=o_p\left(K^{-2\alpha}+\frac{K}{n}+Kh_2^{2a_H}\right). \tag{A.4.31}$$

(iii)

$$B_2=O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right), \tag{A.4.32}$$

and

$$B_3=o_p\left(B_2\right)=o_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right), \tag{A.4.33}$$

holds uniformly over $i\in\{1,\cdots,n\}$.

Proof. Here, we show that

$$B_2=O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right)$$

and before it, we firstly show the corresponding Bernstein Inequality, whose proof is based on Exercise 5.4.15 in Vershynin (2018).

As Exercise 5.4.15 in Vershynin (2018) states, let $X_1, \cdots, X_n$ be independent mean zero $D_1 \times D_2$ random matrices such that $\|X_i\|_{op} \le M$ almost surely for all i and then we have, for $t \ge 0$,

$$P\left\{\left\|\sum_{i=1}^{n} X_i\right\|_{op} \ge t\right\} \le 2(D_1 + D_2)\exp\left(-\frac{t^2/2}{\sigma^2 + Mt/3}\right), \tag{A.4.34}$$

where $\sigma^2 = \max\left(\|\sum_{i=1}^{n} EX_i'X_i\|_{op}, \|\sum_{i=1}^{n} EX_iX_i'\|_{op}\right)$ and $\|\cdot\|_{op}$ is the operator norm. Here we extend the above Bernstein Inequality for operation norm to Euclidean norm. Given the inequality that,

$$\left\|\sum_{i=1}^{n} X_i\right\|^2 = tr\left\{\left[\sum_{i=1}^{n} X_i\right]\left[\sum_{i=1}^{n} X_i\right]'\right\} \le \max(D_1, D_2)\,\lambda_{\max}\left\{\left[\sum_{i=1}^{n} X_i\right]\left[\sum_{i=1}^{n} X_i\right]'\right\} = \max(D_1, D_2)\left\|\sum_{i=1}^{n} X_i\right\|_{op}^2,$$

we have,

$$P\left\{\left\|\sum_{i=1}^{n} X_i\right\|_{op} \ge t\right\} = P\left\{\sqrt{\max(D_1, D_2)}\left\|\sum_{i=1}^{n} X_i\right\|_{op} \ge \sqrt{\max(D_1, D_2)}t\right\} \ge P\left\{\left\|\sum_{i=1}^{n} X_i\right\| \ge \sqrt{\max(D_1, D_2)}t\right\}. \tag{A.4.35}$$

We substitute (A.4.35) into (A.4.34),

$$P\left\{\left\|\sum_{i=1}^{n} X_i\right\| \ge \sqrt{\max(D_1, D_2)}t\right\} \le 2(D_1 + D_2)\exp\left(-\frac{\left(\sqrt{\max(D_1, D_2)}t\right)^2}{2\max(D_1, D_2)\left(\sigma^2 + \frac{M\sqrt{\max(D_1, D_2)}t}{3\sqrt{\max(D_1, D_2)}}\right)}\right),$$

and finally, we get the Bernstein Inequality for Euclidean norm,

$$P\left\{\left\|\sum_{i=1}^{n} X_i\right\| \ge t\right\} \le 2(D_1 + D_2)\exp\left(-\frac{t^2}{2\max(D_1 + D_2)\left(\sigma^2 + \frac{Mt}{3\sqrt{\max(D_1 + D_2)}}\right)}\right).$$

If $X_1, \cdots, X_n$ are not only independent but also identically distributed, $\sigma^2 = \max\left(n\|EX_i'X_i\|_{op}, n\|EX_iX_i'\|_{op}\right)$. Furthermore, if $X_1, \cdots, X_n$ are $D_1 \times D_1$ matrices or $D_1 \times 1$ vectors, $\sigma^2 = n\|EX_iX_i'\|_{op}$. Then, we have

$$P\left\{\left\|\frac{1}{n}\sum_{i=1}^{n} X_i\right\| \ge t\right\} \le 4D_1\exp\left(-\frac{nt^2}{2D_1\left(\|EX_iX_i'\|_{op} + \frac{Mt}{3\sqrt{D_1}}\right)}\right). \tag{A.4.36}$$

(i)

$B_{2,1}$

$$
\begin{aligned}
=&\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\bar{g}(X_j)-g(X_j)\right]^2 \\
=&\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[P^K(X_j)'\bar{\beta}-P^K(X_j)'\beta_0+P^K(X_j)'\beta_0-g(X_j)\right]^2 \\
\leq&2\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[P^K(X_j)'\bar{\beta}-P^K(X_j)'\beta_0\right]^2+2\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[P^K(X_j)'\beta_0-g(X_j)\right]^2 \\
\leq&\left(\bar{\beta}-\beta_0\right)'\frac{2}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\left(\bar{\beta}-\beta_0\right)+O\left(K^{-2\alpha}\right)\frac{2}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) \\
\leq&2\left\|\bar{\beta}-\beta_0\right\|^2\lambda_{\max}\left(\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\right)+O\left(K^{-2\alpha}\right). \qquad \text{(A.4.37)}
\end{aligned}
$$

$$
\begin{aligned}
&E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\middle| g\left(X_i\right)\right] \\
=&E\left\{T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)E\left[P^K(X_j)P^K(X_j)'\middle| g\left(X_j\right)\right]\middle| g\left(X_i\right)\right\} \\
=&\int T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\Big\{E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]+\partial E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]\left[g\left(X_j\right)-g\left(X_i\right)\right] \\
&+\partial^2 E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]\left[g\left(X_j\right)-g\left(X_i\right)\right]^2+\cdots\Big\}\Big\{f_{g(X)}\left[g(X_i)\right]+\partial f_{g(X)}\left[g(X_i)\left[g\left(X_j\right)-g\left(X_i\right)\right]\right] \\
&+\partial^2 f_{g(X)}\left[g(X_i)\left[g\left(X_j\right)-g\left(X_i\right)\right]\right]^2+\cdots\Big\} \\
=&E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]f_{g(X)}\left[g(X_i)\right]+h_3^{S+1}\partial E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]\partial f_{g(X)}\left[g(X_i)\right] \\
&+h_3^{S+1}\partial^2 E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]f_{g(X)}\left[g(X_i)\right]+h_3^{S+1}E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]\partial^2 f_{g(X)}\left[g(X_i)\right]+\cdots. \\
&\qquad \text{(A.4.38)}
\end{aligned}
$$

Given Assumptions 5 (2) and 6 (1) and (3),

$$
\left\|E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\middle| g\left(X_i\right)\right]-E\left[P^K(X_i)P^K(X_i)'\middle| g\left(X_i\right)\right]f_{g(X)}\left[g(X_i)\right]\right\|=O\left(\sqrt{K}h_3^{S+1}\right), \qquad \text{(A.4.39)}
$$

holds uniformly over $i\in\{1,2,\cdots,n\}$.

Given the Assumptions 5 (2) and 6 (1), all eigenvalues of $E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\middle| g\left(X_i\right)\right]$ are bounded and bounded away from zero.

$$
\begin{aligned}
&\Big\|E\Big(\left\{T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\middle| g\left(X_i\right)\right]\right\} \\
\times&\left\{T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\middle| g\left(X_i\right)\right]\right\}\Big| g\left(X_i\right)\Big)\Big\|_{op}
\end{aligned}
$$

$$
\begin{aligned}
=&\left\| E\left\{ T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)^2 P^K(X_j)P^K(X_j)'P^K(X_j)P^K(X_j)' \middle| g\left(X_i\right)\right\}\right\|_{op} \\
+&\left\| E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) P^K(X_j)P^K(X_j)' \middle| g\left(X_i\right)\right] E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) P^K(X_j)P^K(X_j)' \middle| g\left(X_i\right)\right]\right\|_{op} \\
\leq&\left\| E\left\{ T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)^2 E\left[P^K(X_j)P^K(X_j)'P^K(X_j)P^K(X_j)' \middle| g\left(X_j\right)\right] \middle| g\left(X_i\right)\right\}\right\|_{op} \\
+&\left\| E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) P^K(X_j)P^K(X_j)' \middle| g\left(X_i\right)\right]\right\|_{op}^2 \\
=&\left\| \int T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)^2 E\left[P^K(X_j)P^K(X_j)'P^K(X_j)P^K(X_j)' \middle| g\left(X_j\right)\right] f_{g(X)}\left[g\left(X_j\right)\right] dg\left(X_j\right)\right\|_{op} + O(1) \\
\leq& h_3^{-1}\left\| E\left[P^K(X_i)P^K(X_i)'P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right] f_{g(X)}\left[g\left(X_i\right)\right]\right\|_{op} \\
+&O(h_3)\left\| \partial E\left[P^K(X_i)P^K(X_i)'P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right] \partial f_{g(X)}\left[g\left(X_i\right)\right]\right\|_{op} \\
+&O(h_3)\left\| \partial^2 E\left[P^K(X_i)P^K(X_i)'P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right] f_{g(X)}\left[g\left(X_i\right)\right]\right\|_{op} \\
+&O(h_3)\left\| E\left[P^K(X_i)P^K(X_i)'P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right] \partial^2 f_{g(X)}\left[g\left(X_i\right)\right]\right\|_{op} + O(1) \\
=& h_3^{-1}A_{2,1} + O(h_3)\left(A_{2,2}+A_{2,3}+A_{2,4}\right).
\end{aligned}
\tag{A.4.40}
$$

Since $E\left[\left\|P^K(X_i)\right\|^2 P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right]$ are semi-positive definite and given the Assumptions 5 (2) and 6 (1),

$$
\begin{aligned}
&A_{2,1} \\
=& f_{g(X)}\left[g\left(X_i\right)\right]\left\| E\left[P^K(X_i)P^K(X_i)'P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right]\right\|_{op} \\
=& f_{g(X)}\left[g\left(X_i\right)\right]\left\| E\left[\left\|P^K(X_i)\right\|^2 P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right]\right\|_{op} \\
=& f_{g(X)}\left[g\left(X_i\right)\right] \max_{\|\omega\|=1} \omega' E\left[P^K(X_i)P^K(X_i)'P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right] \omega \\
=& f_{g(X)}\left[g\left(X_i\right)\right] \max_{\|\omega\|=1} E\left\{ \left[\omega' P^K(X_i)\right]^2 \left\|P^K(X_i)\right\|^2 \middle| g\left(X_i\right)\right\} \\
\leq& f_{g(X)}\left[g\left(X_i\right)\right] \sup_{x\in\mathcal{X}} \left\|P^K(X)\right\|^2 \left\| E\left[P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right]\right\|_{op} \\
\leq& K f_{g(X)}\left[g\left(X_i\right)\right] \left\| E\left[P^K(X_i)P^K(X_i)' \middle| g\left(X_i\right)\right]\right\|_{op} \\
=& O(K).
\end{aligned}
\tag{A.4.41}
$$

Substitute ((A.4.41)) into ((A.4.40)) and we have

$$
\begin{aligned}
&\Big\| E\Big(\left\{ T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) P^K(X_j)P^K(X_j)' - E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) P^K(X_j)P^K(X_j)' \middle| g\left(X_i\right)\right]\right\} \\
\times&\ \left\{ T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) P^K(X_j)P^K(X_j)' - E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right) P^K(X_j)P^K(X_j)' \middle| g\left(X_i\right)\right]\right\} \Big| g\left(X_i\right)\Big)\Big\|_{op}
\end{aligned}
$$

$$=O\left(h_3^{-1}K\right). \tag{A.4.42}$$

$$\begin{aligned}&\left\|T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\right\|_{op}\\=&O(h_3^{-1})\left\|T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\right\|_{op}\\\leq&O(h_3^{-1})\left\|P^K(X_j)P^K(X_j)'\right\|\\\leq&O(h_3^{-1}K).\end{aligned} \tag{A.4.43}$$

By Bernstein Inequality ((A.4.36)), there exist positive constants C_1 and C_2 such that

$$\begin{aligned}&P\left\{\left\|\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\right|g\left(X_i\right)\right]\right\|\geq t\right\}\\\leq&C_1K\exp\left(-\frac{C_2nt^2}{2K\left(h_3^{-1}K+\frac{h_3^{-1}Kt}{3\sqrt{K}}\right)}\right)\\\leq&C_1K\exp\left(-C_2nh_3K^{-2}t^2\right).\end{aligned}$$

Thus,

$$\begin{aligned}&P\left\{\max_{i\in\{1,\cdots,n\}}\left\|\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\right|g\left(X_i\right)\right]\right\|\geq t\right\}\\\leq&C_1nK\exp\left(-C_2nh_3K^{-2}t^2\right),\end{aligned} \tag{A.4.44}$$

and then

$$\left\|\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\right|g\left(X_i\right)\right]\right\|=O\left(K\sqrt{\frac{\log n}{nh_3}}\right), \tag{A.4.45}$$

holds uniformly over $i\in\{1,\cdots,n\}$.

Afterwards, substitute ((A.4.39)) into ((A.4.45)),

$$\begin{aligned}&\left\|\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[P^K(X)P^K(X)'\right|g\left(X_i\right)\right]f_{g(X)}\left[g(X_i)\right]\right\|\\=&O\left(K\sqrt{\frac{\log n}{nh_3}}+\sqrt{K}h_3^{S+1}\right)=o_p(1),\end{aligned}$$

holds uniformly over $i \in \{1, 2, \cdots, n\}$. It follows that from the definition of maximum eigenvalue that

$$
\begin{aligned}
&\lambda_{\max}\left(\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'\right)\\
=&\max_{\|\omega\|=1}\left\{\omega' E\left[\left.P^K(X)P^K(X)'\right| g\left(X_i\right)\right] f_{g(X)}\left[g(X_i)\right]\omega\right.\\
&\left.+\,\omega\left(\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[\left.P^K(X)P^K(X)'\right| g\left(X_i\right)\right] f_{g(X)}\left[g(X_i)\right]\right)\omega\right\}\\
\leq&\lambda_{\max}\left\{E\left[\left.P^K(X)P^K(X)'\right| g\left(X_i\right)\right] f_{g(X)}\left[g(X_i)\right]\right\}\\
&+\max_{i\in\{1,\cdots,n\}}\left\|\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)P^K(X_j)P^K(X_j)'-E\left[\left.P^K(X)P^K(X)'\right| g\left(X_i\right)\right] f_{g(X)}\left[g(X_i)\right]\right\|\\
=&o_p(1). \qquad \text{(A.4.46)}
\end{aligned}
$$

By ((A.4.37)) and ((A.4.46)),

$$
B_{2,1} = O_p\left(K^{-2\alpha} + \frac{K}{n} + Kh_2^{2a_H}\right),
$$

and (A.4.28) holds.

(ii) Given the Assumption 6 (3),

$$
\begin{aligned}
&\left|B_{2,2}\right|\\
=&h_3^{-1}\frac{1}{n}\sum_{j=1,\neq i}^{n}\left|\partial T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\right|\left[\bar{g}\left(X_j\right)-g\left(X_j\right)\right]^2\left|\bar{g}\left(X_j\right)-g\left(X_j\right)\right|\\
\leq&h_3^{-1}\max_{x\in\mathcal{X}}\left|\bar{g}\left(x\right)-g\left(x\right)\right|\frac{1}{n}\sum_{j=1,\neq i}^{n}\left|\partial T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\right|\left[\bar{g}\left(X_j\right)-g\left(X_j\right)\right]^2\\
=&O_p\left[h_3^{-1}\sqrt{K}\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right)\right]O_p\left(K^{-2\alpha}+\frac{K}{n}+Kh_2^{2a_H}\right)\\
=&O_p\left(K^{-2\alpha}+\frac{K}{n}+Kh_2^{2a_H}\right).
\end{aligned}
$$

and thus (A.4.29) holds. It is trivial to prove (A.4.30), (A.4.31) and (A.4.32). ■

Lemma A.4.5 *Suppose that Assumptions 1 - 4, 5 (1) and (3) and 6 (1)-(3) hold. Then all eigenvalues of A_1 are bounded and bounded away from zero and it holds uniformly over $i \in \{1, \cdots, n\}$.*

Proof. Let $A_{1,d} = \dfrac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left(\dfrac{\triangle \bar{g}_{i,j}}{h_3}\right)^d$, for $d \in \{0, \cdots, 2S\}$ and then

$$A_1 = \begin{pmatrix} A_{1,0} & A_{1,1} & \cdots & A_{1,S} \\ A_{1,1} & A_{1,2} & \cdots & A_{1,S+1} \\ \vdots & \vdots & & \vdots \\ A_{1,S+1} & A_{1,S+2} & \cdots & A_{1,2S}. \end{pmatrix}. \tag{A.4.47}$$

$$\begin{aligned}
&E\left[T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\middle| g\left(X_i\right)\right] \\
=&\int T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\Big\{f_{g(X)}\left[g(X_i)\right]+\partial f_{g(X)}\left[g(X_i)\right]\left[g(X_j)-g(X_i)\right]+\frac{1}{2}\partial^2 f_{g(X)}\left[g(X_i)\right]\left[g(X_j)-g(X_i)\right]^2 \\
+&\frac{1}{6}\partial^3 f_{g(X)}\left[g(X_i)\right]\left[g(X_j)-g(X_i)\right]^3+\frac{1}{24}\partial^4 f_{g(X)}\left[g(X_i)\right]\left[g(X_j)-g(X_i)\right]^4+\cdots\Big\}dg(X_j) \\
=&f_{g(X)}\left[g(X_i)\right]+h_3^{S+1}\frac{1}{2}\partial^2 f_{g(X)}\left[g(X_i)\right]\int \omega^2 T_{h_3}(\omega)d\omega.
\end{aligned} \tag{A.4.48}$$

$$\begin{aligned}
&E\left\{T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]\middle| g\left(X_i\right)\right\} \\
=&\int T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]\Big\{f_{g(X)}\left[g(X_i)\right]+h_3\partial f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right] \\
+&h_3^{S+1}\frac{1}{2}\partial^2 f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^2+h_3^3\frac{1}{6}\partial^3 f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^3 \\
+&h_3^4\frac{1}{24}\partial^4 f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^4+\cdots\Big\}dg(X_j) \\
=&h_3\partial f_{g(X)}\left[g(X_i)\right]\int \omega^2 T_{h_3}(\omega)d\omega+h_3^3\frac{1}{6}\partial^3 f_{g(X)}\left[g(X_i)\right]\int \omega^4 T_{h_3}(\omega)d\omega.
\end{aligned}$$

$$\begin{aligned}
&E\left\{T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^2\middle| g\left(X_i\right)\right\} \\
=&\int T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^2\Big\{f_{g(X)}\left[g(X_i)\right]+h_3\partial f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right] \\
+&h_3^{S+1}\frac{1}{2}\partial^2 f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^2+h_3^3\frac{1}{6}\partial^3 f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^3 \\
+&h_3^4\frac{1}{24}\partial^4 f_{g(X)}\left[g(X_i)\right]\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^4+\cdots\Big\}dg(X_j) \\
=&f_{g(X)}\left[g(X_i)\right]\int \omega^2 T_{h_3}(\omega)d\omega+h_3^{S+1}\frac{1}{2}\partial^2 f_{g(X)}\left[g(X_i)\right]\int \omega^4 T_{h_3}(\omega)d\omega.
\end{aligned}$$

For $d = 1, 2, \cdots,$

$$E\left\{T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^{2d}\middle| g\left(X_i\right)\right\}$$

$$=f_{g(X)}\left[g(X_i)\right]\int \omega^{2d}T_{h_3}(\omega)d\omega+h_3^2\frac{1}{2}\partial^2 f_{g(X)}\left[g(X_i)\right]\int \omega^{2d+2}T_{h_3}(\omega)d\omega. \tag{A.4.49}$$

and

$$E\left\{T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^{2d+1}\middle| g\left(X_i\right)\right\}$$
$$=h_3\partial f_{g(X)}\left[g(X_i)\right]\int \omega^{2d+2}T_{h_3}(\omega)d\omega+h_3^3\frac{1}{6}\partial^3 f_{g(X)}\left[g(X_i)\right]\int \omega^{2d+4}T_{h_3}(\omega)d\omega. \tag{A.4.50}$$

Given Assumption 4, by ((A.4.49)) and ((A.4.50)),

$$\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(g\left(X_j\right)-g\left(X_i\right)\right)\begin{pmatrix} 1 & \frac{g\left(X_j\right)-g\left(X_i\right)}{h_3} & \cdots & \left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^S \\ \frac{g\left(X_j\right)-g\left(X_i\right)}{h_3} & \left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^2 & \cdots, & \left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^{S+1} \\ \vdots & \vdots & & \vdots \\ \left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^S & \left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^{S+1} & \cdots & \left[\frac{g\left(X_j\right)-g\left(X_i\right)}{h_3}\right]^{2S} \end{pmatrix}$$

$$=\begin{pmatrix} f_{g(X)}\left[g(X_i)\right] & h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^2T_{h_3}(\omega)d\omega & \cdots & h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^{S+1}T_{h_3}(\omega)d\omega \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^2T_{h_3}(\omega)d\omega & f_{g(X)}\left[g(X_i)\right]\int\omega^2T_{h_3}(\omega)d\omega & \cdots & f_{g(X)}\left[g(X_i)\right]\int\omega^{S+1}T_{h_3}(\omega)d\omega \\ f_{g(X)}\left[g(X_i)\right]\int\omega^2T_{h_3}(\omega)d\omega & h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^4T_{h_3}(\omega)d\omega & \cdots & h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^{S+3}T_{h_3}(\omega)d\omega \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^4T_{h_3}(\omega)d\omega & f_{g(X)}\left[g(X_i)\right]\int\omega^4T_{h_3}(\omega)d\omega & \cdots & f_{g(X)}\left[g(X_i)\right]\int\omega^{S+3}T_{h_3}(\omega)d\omega \\ \vdots & \vdots & & \vdots \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^{S-1}T_{h_3}(\omega)d\omega & f_{g(X)}\left[g(X_i)\right]\int\omega^{S-1}T_{h_3}(\omega)d\omega & \cdots & f_{g(X)}\left[g(X_i)\right]\int\omega^{2S-2}T_{h_3}(\omega)d\omega \\ f_{g(X)}\left[g(X_i)\right]\int\omega^{S-1}T_{h_3}(\omega)d\omega & h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^{S+1}T_{h_3}(\omega)d\omega & \cdots & h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^{2S}T_{h_3}(\omega)d\omega \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int\omega^{S+1}T_{h_3}(\omega)d\omega & f_{g(X)}\left[g(X_i)\right]\int\omega^{S+1}T_{h_3}(\omega)d\omega & \cdots & f_{g(X)}\left[g(X_i)\right]\int\omega^{2S}T_{h_3}(\omega)d\omega \end{pmatrix}$$

$$+O_p(h_3^{S+1})\begin{pmatrix} 1 & h_3 & \cdots & h_3 \\ h_3 & 1 & \cdots & 1 \\ 1 & h_3 & \cdots & h_3 \\ h_3 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ h_3 & 1 & \cdots & 1 \\ 1 & h_3 & \cdots & h_3 \\ h_3 & 1 & \cdots & 1 \end{pmatrix}+\sqrt{\frac{\log n}{nh_3}};$$

and thus,

$$A_1 = f_{g(X)}\left[g(X_i)\right] \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \int \omega^2 T_{h_3}(\omega)d\omega & \cdots & \int \omega^{S+1} T_{h_3}(\omega)d\omega \\ \int \omega^2 T_{h_3}(\omega)d\omega & 0 & \cdots & 0 \\ 0 & \int \omega^4 T_{h_3}(\omega)d\omega & \cdots & \int \omega^{S+3} T_{h_3}(\omega)d\omega \\ \vdots & \vdots & & \vdots \\ 0 & \int \omega^{S-1} T_{h_3}(\omega)d\omega & \cdots & \int \omega^{2S-2} T_{h_3}(\omega)d\omega \\ \int \omega^{S-1} T_{h_3}(\omega)d\omega & 0 & \cdots & 0 \\ 0 & \int \omega^{S+1} T_{h_3}(\omega)d\omega & \cdots & \int \omega^{2S} T_{h_3}(\omega)d\omega \end{pmatrix} + o_p(1), \qquad \text{(A.4.51)}$$

hold uniformly over $i \in \{1, \cdots, n\}$.

Since

$$A_1 = f_{g(X)}\left[g(X_i)\right] \int T_{h_3}(\omega) \begin{pmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \vdots \\ \omega^{h_3} \end{pmatrix} \begin{pmatrix} 1 & \omega & \omega^2 & \omega^3 & \vdots & \omega^{h_3} \end{pmatrix} d\omega + o_p(1),$$

is positive definite and $\forall |\lambda| = 1$, $\int \omega^{2S} T_{h_3}(\omega)d\omega \leq \lambda' A_1 \lambda \leq \int T_{h_3}(\omega)d\omega$, all eigenvalues of A_1 are bounded and bounded away from zero and it holds uniformly over $i \in \{1, \cdots, n\}$. ■

Lemma A.4.6 *Suppose that Assumptions 1 - 6, 6 (1), (2) and (3) hold. Then*

(i)

$$|B_{1,1}| = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H}\right), \qquad \text{(A.4.52)}$$

holds uniformly over $i \in \{1, \cdots, n\}$.

(ii)

$$B_{1,2} = \partial F_i \left[\bar{g}(X_i) - g(X_i)\right] \begin{pmatrix} f_{g(X)}\left[g(X_i)\right] \\ h_3 \partial f_{g(X)}\left[g(X_i)\right] \int \omega^2 T_{h_3}(\omega) d\omega \\ f_{g(X)}\left[g(X_i)\right] \int \omega^3 T_{h_3}(\omega) d\omega \\ h_3 \partial f_{g(X)}\left[g(X_i)\right] \int \omega^4 T_{h_3}(\omega) d\omega \\ \cdots \\ h_3 \partial f_{g(X)}\left[g(X_i)\right] \int \omega^{S-1} T_{h_3}(\omega) d\omega \\ f_{g(X)}\left[g(X_i)\right] \int \omega^{S-1} T_{h_3}(\omega) d\omega \\ h_3 \partial f_{g(X)}\left[g(X_i)\right] \int \omega^{S+1} T_{h_3}(\omega) d\omega \end{pmatrix} + O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H}\right), \tag{A.4.53}$$

holds for each $i \in \{1, \cdots, n\}$.

(iii)

$$|B_{1,3}| = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K} h_2^{a_H}\right), \tag{A.4.54}$$

holds uniformly over $i \in \{1, \cdots, n\}$.

(iv)

$$|B_{1,4}| = O\left(\left[\bar{g}(X_i) - g(X_i)\right]^2\right), \tag{A.4.55}$$

holds for each $i \in \{1, \cdots, n\}$.

(v)

$$|B_{1,5}| = O_p(h_3^{S+1}), \tag{A.4.56}$$

holds uniformly over $i \in \{1, \cdots, n\}$.

(vi)

$$|B_{1,6}| = O_p(h_3^{S+2}), \tag{A.4.57}$$

holds uniformly over $i \in \{1, \cdots, n\}$.

(vii)

$$|B_{1,7}| = \frac{1}{n} \sum_{j=1, \neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right) \begin{pmatrix} 1 \\ \frac{\triangle \bar{g}_{i,j}}{h_3} \\ \vdots \\ \left(\frac{\triangle \bar{g}_{i,j}}{h_3}\right)^{h_3} \end{pmatrix} e_i = O_p\left(\sqrt{\frac{\log n}{n h_3}}\right), \tag{A.4.58}$$

holds uniformly over $i \in \{1, \cdots, n\}$.

Proof. (i)

$$
\begin{aligned}
&|B_{1,1}|\\
=&\left|\partial F_i \begin{pmatrix} \frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left[\bar{g}(X_j) - g(X_j)\right] \\ \frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\frac{\triangle \bar{g}_{i,j}}{h_3}\left[\bar{g}(X_j) - g(X_j)\right] \\ \vdots \\ \frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left(\frac{\triangle \bar{g}_{i,j}}{h_3}\right)^{h_3}\left[\bar{g}(X_j) - g(X_j)\right] \end{pmatrix}\right|\\
\leq& O(1)\begin{pmatrix} \left\{\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\right\}^{1/2}\left\{\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left[\bar{g}(X_j) - g(X_j)\right]^2\right\}^{1/2} \\ \left\{\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left(\frac{\triangle \bar{g}_{i,j}}{h_3}\right)^2\right\}^{1/2}\left\{\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left[\bar{g}(X_j) - g(X_j)\right]^2\right\}^{1/2} \\ \vdots \\ \left\{\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left(\frac{\triangle \bar{g}_{i,j}}{h_3}\right)^{2S}\right\}^{1/2}\left\{\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\left[\bar{g}(X_j) - g(X_j)\right]^2\right\}^{1/2} \end{pmatrix}\\
=&O(1)\begin{pmatrix} A_{1,0}^{1/2}B_2^{1/2} \\ A_{1,2}^{1/2}B_2^{1/2} \\ \vdots \\ A_{1,2S}^{1/2}B_2^{1/2} \end{pmatrix} = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H}\right),
\end{aligned}
$$

and (A.4.52) hold uniformly over $i \in \{1, \cdots, n\}$.

(ii) Similar with the proof of Lemma (A.4.5), we have

$$
\begin{aligned}
B_{1,2} =& \partial F_i\left[\bar{g}(X_i) - g(X_i)\right]\frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle \bar{g}_{i,j}\right)\begin{pmatrix} 1 \\ \frac{\triangle \bar{g}_{i,j}}{h_3} \\ \vdots \\ \left(\frac{\triangle \bar{g}_{i,j}}{h_3}\right)^{h_3} \end{pmatrix}\\
=& \partial F_i\left[\bar{g}(X_i) - g(X_i)\right]\begin{pmatrix} A_{1,0} \\ A_{1,1} \\ \vdots \\ A_{1,S} \end{pmatrix}
\end{aligned}
$$

$$
=\partial F_i\left[\bar{g}(X_i)-g(X_i)\right]\begin{pmatrix} f_{g(X)}\left[g(X_i)\right] \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int \omega^2 T_{h_3}(\omega)d\omega \\ f_{g(X)}\left[g(X_i)\right]\int \omega^2 T_{h_3}(\omega)d\omega \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int \omega^4 T_{h_3}(\omega)d\omega \\ \cdots \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int \omega^{S-1} T_{h_3}(\omega)d\omega \\ f_{g(X)}\left[g(X_i)\right]\int \omega^{S-1} T_{h_3}(\omega)d\omega \\ h_3\partial f_{g(X)}\left[g(X_i)\right]\int \omega^{S+1} T_{h_3}(\omega)d\omega \end{pmatrix}+R_{n,3,i},
$$

and by the Assumption 6 (3),

$$
\begin{aligned}
R_{n,3,i}=&\left(h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}\right)O_p\left(\sup_{x\in\mathcal{X}}\bar{g}(x)-g(x)\right)=\left(h_3^{S+1}+\sqrt{\frac{\log n}{nh_3}}\right)\sqrt{K}O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right)\\
=&O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right),
\end{aligned}
$$

and then (A.4.53) hold for each $i\in\{1,\cdots,n\}$.

(iii)

$$
\begin{aligned}
&|B_{1,3}|\\
=&\left|\partial^2 F_i\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots \\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}\left[\bar{g}(X_j)-g(X_j)\right]^2\right|\leq O(1)\begin{pmatrix}A_{1,0}^{1/2}B_3^{1/2}\\ A_{1,2}^{1/2}B_3^{1/2}\\ \vdots \\ A_{1,2S}^{1/2}B_3^{1/2}\end{pmatrix}\\
=&O_p\left(K^{-\alpha}+\sqrt{\frac{K}{n}}+\sqrt{K}h_2^{a_H}\right),
\end{aligned}
$$

and then (A.4.54) hold uniformly over $i\in\{1,\cdots,n\}$.

(iv)

$$
\begin{aligned}
&|B_{1,4}|\\
=&\left|\partial^2 F_i\left[\bar{g}(X_i)-g(X_i)\right]^2\frac{1}{n}\sum_{j=1,\neq i}^{n}T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots \\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}\right|=\left|\partial^2 F_i\left[\bar{g}(X_i)-g(X_i)\right]^2\right|\begin{pmatrix}A_{1,0}\\ A_{1,1}\\ \vdots \\ A_{1,S}\end{pmatrix}
\end{aligned}
$$

$$=O\left([\bar{g}(X_i)-g(X_i)]^2\right),$$

and then (A.4.55) hold for each $i \in \{1,\cdots,n\}$.

(v)

$$|B_{1,5}| = \left|\partial^{S+1}F_i \frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots \\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}[g(X_j)-g(X_i)]^{S+1}\right| = O_p(h_3^{S+1}),$$

and then (A.4.56) hold uniformly over $i \in \{1,\cdots,n\}$.

(vi)

$$|B_{1,6}| = \left|\partial^{S+2}F_i \frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots \\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix}[g(X_j)-g(X_i)]^{S+2}\right| = O_p(h_3^{S+2}),$$

and then (A.4.57) hold uniformly over $i \in \{1,\cdots,n\}$.

(vii)

$$|B_{1,7}| = \frac{1}{n}\sum_{j=1,\neq i}^{n} T_{h_3}\left(\triangle\bar{g}_{i,j}\right)\begin{pmatrix}1\\ \frac{\triangle\bar{g}_{i,j}}{h_3}\\ \vdots \\ \left(\frac{\triangle\bar{g}_{i,j}}{h_3}\right)^{h_3}\end{pmatrix} e_i = O_p\left(\sqrt{\frac{\log n}{nh_3}}\right),$$

and then (A.4.58) hold uniformly over $i \in \{1,\cdots,n\}$. ■

A.4.3 Lemmas for the proof of Theorem 2.3.3

Lemma A.4.7 *Suppose that Assumptions Assumptions 1 - 6 hold. Then*

(i)

$$A_4 = E\left\{\partial F\left[g\left(X\right)\right]^2 \middle| X_l = x_l\right\} f_{X_l}(x_l)\begin{pmatrix}\int T(\omega)d\omega & 0\\ 0 & \int \omega^2 T(\omega)d\omega\end{pmatrix} + o_p(1).$$

(ii)

$$\sqrt{nt}B_{4,1,1} \xrightarrow{d} N\left(0, E\left\{\partial F\left[g\left(X\right)\right]^2 \Big| X_l = x_l\right\} E\left(e^2 \big| X_l = x_l\right) f_{X_l}(x_l) \begin{pmatrix} \kappa_{02} & 0 \\ 0 & \int \omega^2 T(\omega)^2 d\omega \end{pmatrix}\right). \quad \text{(A.4.59)}$$

(iii)

$$B_{4,2} = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H} + h_3^{S+1} + \sqrt{\frac{\log n}{nh_3}}\right). \quad \text{(A.4.60)}$$

(iv)

$$B_{4,3} = O_p\left(B_2\right) = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H}\right). \quad \text{(A.4.61)}$$

(v)

$$B_{4,4} = \begin{pmatrix} h_{4,l}^2 \frac{1}{2}\partial^2 g_l E\left\{\partial F\left[g\left(X\right)\right]^2 \Big| X_l = x_l\right\} f_{X_l}(x_l)\kappa_{21} \\ 0 \end{pmatrix} + o_p(h_{4,l}^2). \quad \text{(A.4.62)}$$

(vi)

$$B_{4,5} = O_p(h_{4,l}^3). \quad \text{(A.4.63)}$$

Proof. Given that

$$\begin{aligned}
& E\left[\left(\partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i} - x_l\right)\right] \\
=& E\left(E\left\{\left[\partial F\left(\sum_{j=1,\neq l}^{d_x} g(X_j)\left(X_{j,i}\right)\right) + g_l(X_{l,i})\right]^2 \Bigg| X_{l,i}\right\} T_{h_{4,l}}\left(X_{l,i} - x_l\right)\right) \\
=& E\left\{\left[\partial F\left(\sum_{j=1,\neq l}^{d_x} g(X_j)\left(X_{j,i}\right) + g_l(x_l)\right)\right]^2 \Bigg| x_l\right\} E\left[T_{h_{4,l}}\left(X_{l,i} - x_l\right)\right] + o_p(1) \\
=& E\left\{\left[\partial F\left(\sum_{j=1,\neq l}^{d_x} g(X_j)\left(X_{j,i}\right) + g_l(x_l)\right)\right]^2 \Bigg| x_l\right\} f_{X_l}(x_l) \int T(\omega) d\omega + o_p(1),
\end{aligned}$$

$$\begin{aligned}
& A_4 \\
=& \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \left(1, \frac{X_{l,i} - x_l}{h_{4,l}}\right) \left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i} - x_l\right) \\
=& \frac{1}{n}\sum_{i=1}^n \begin{pmatrix} 1 \\ \frac{X_{l,i} - x_l}{h_{4,l}} \end{pmatrix} \left(1, \frac{X_{l,i} - x_l}{h_{4,l}}\right) \left(\partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i} - x_l\right) + o_p(1)
\end{aligned}$$

$$=\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\left(\partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right) & \frac{1}{n}\sum_{i=1}^{n}\frac{X_{l,i}-x_l}{h_{4,l}}\left(\partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right) \\ \frac{1}{n}\sum_{i=1}^{n}\frac{X_{l,i}-x_l}{h_{4,l}}\left(\partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right) & \frac{1}{n}\sum_{i=1}^{n}\left(\frac{X_{l,i}-x_l}{h_{4,l}}\right)^2\left(\partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right) \end{pmatrix}+o_p(1)$$

$$=E\left\{\left[\partial F\left(\sum_{\iota=1,\neq l}^{d_x} g_\iota\left(X_{\iota,i}\right)+g_l(x_l)\right)\right]^2\middle| x_l\right\} f_{X_l}(x_l)\begin{pmatrix}\int T(\omega)d\omega & 0 \\ 0 & \int \omega^2 T(\omega)d\omega\end{pmatrix}+o_p(1).$$

(ii)

$$\begin{aligned}
&B_{4,1}\\
=&\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\widehat{\partial F}_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)e_i\\
=&\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\partial F_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)e_i+\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left(\widehat{\partial F}_i-\partial F_i\right)T_{h_{4,l}}\left(X_{l,i}-x_l\right)e_i\\
=&B_{4,1,1}+B_{4,1,1}.
\end{aligned}$$

Given that

$$E\left[\begin{pmatrix}1\\ \frac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\partial F_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)e_i\right]=0,$$

and

$$\begin{aligned}
&E\left[\begin{pmatrix}1 & \frac{X_{l,i}-x_l}{h_{4,l}}\\ \frac{X_{l,i}-x_l}{h_{4,l}} & \left(\frac{X_{l,i}-x_l}{h_{4,l}}\right)^2\end{pmatrix}\partial F_i^2 T_{h_{4,l}}^2\left(X_{l,i}-x_l\right)e_i^2\right]\\
=&E\left[\begin{pmatrix}1 & \frac{X_{l,i}-x_l}{h_{4,l}}\\ \frac{X_{l,i}-x_l}{h_{4,l}} & \left(\frac{X_{l,i}-x_l}{h_{4,l}}\right)^2\end{pmatrix}E\left(e_i^2\partial F_i^2\middle| X_{1,i}\right)T_{h_{4,l}}^2\left(X_{l,i}-x_l\right)\right]\\
=&E\left\{E(e^2|X)\partial F\left[g\left(X\right)\right]^2\middle| X_l=x_l\right\}f_{X_l}(x_l)\begin{pmatrix}\kappa_{02} & 0\\ 0 & \int\omega^2 T(\omega)^2 d\omega\end{pmatrix}+o_p(1).
\end{aligned}$$

Thus, by central limit theorem for i.i.d random variables,

$$\sqrt{nt}B_{4,1,1}\xrightarrow{d}N\left(0,E\left\{E(e^2|X)\partial F\left[g\left(X\right)\right]^2\middle| X_l=x_l\right\}f_{X_l}(x_l)\begin{pmatrix}\kappa_{02} & 0\\ 0 & \int\omega^2 T(\omega)^2 d\omega\end{pmatrix}\right).$$

and thus (A.4.59) holds.

In $B_{4,1,2}$, given the Assumption that $\widehat{F}$ and $\widehat{\partial F}$ do not involve the i th observation, we have, for $i \neq j \in \{1, \cdots, n\}$,

$$COV\left(\begin{pmatrix} 1 \\ \frac{X_{l,i}-x_l}{h_{4,l}} \end{pmatrix}\left(\widehat{\partial F}_i - \partial F_i\right) T_{h_{4,l}}\left(X_{l,i}-x_l\right) e_i, \begin{pmatrix} 1 \\ \frac{X_{l,j}-x_l}{h_{4,l}} \end{pmatrix}\left(\widehat{\partial F}_j - \partial F_j\right) T_{h_{4,l}}\left(X_{l,j}-x_l\right) e_j\right) = 0.$$

Further given that,

$$E\left[\begin{pmatrix} 1 & \frac{X_{l,i}-x_l}{h_{4,l}} \\ \frac{X_{l,i}-x_l}{h_{4,l}} & \left(\frac{X_{l,i}-x_l}{h_{4,l}}\right)^2 \end{pmatrix}\left(\widehat{\partial F}_i - \partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right) e_i^2\right] = O\left(\max_{i\in\{1,\cdots,n\}}\left|\widehat{\partial F}_i - \partial F_i\right|\right)^2 = o_p(1).$$

Thus, by the central limit theorem for uncorrelated sequence, compared with $B_{4,1,1}$, $B_{4,1,2}$ is negelactable.

(iii)

$$B_{4,2} = \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \frac{X_{l,i}-x_l}{h_{4,l}} \end{pmatrix}\widehat{\partial F}_i T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(F_i - \widehat{F}_i\right) = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H} + h_3^{S+1} + \sqrt{\frac{\log n}{nh_3}}\right),$$

and then (A.4.60) holds.

(iv)

$$\begin{aligned} B_{4,3} &= \frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \frac{X_{l,i}-x_l}{h_{4,l}} \end{pmatrix}\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left[\bar{g}_l\left(X_{l,i}\right) - g_l\left(X_{l,i}\right)\right] \\ &= O_p\left(\frac{1}{n}\sum_{i=1}^{n} T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(\bar{g}_l\left(X_{l,i}\right) - g_l\left(X_{l,i}\right)\right)^2\right) = O_p\left(B_2\right) = O_p\left(K^{-\alpha} + \sqrt{\frac{K}{n}} + \sqrt{K}h_2^{a_H}\right), \end{aligned}$$

and (A.4.61) holds.

(v) Given that

$$\begin{aligned} &E\left\{\begin{pmatrix} 1 \\ \frac{X_{l,i}-x_l}{h_{4,l}} \end{pmatrix}\partial F_i^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(X_{l,i}-x_l\right)^2\right\} \\ =&\begin{pmatrix} h_{4,l}^2 E\left\{\partial F\left[g\left(X\right)\right]^2\middle| X_l = x_l\right\} f_{X_l}(x_l)\kappa_{21} \\ 0 \end{pmatrix} + o_p(h_{4,l}^2), \end{aligned}$$

$$\begin{aligned} &B_{4,4} \\ =&\frac{1}{2}\partial^2 g_l\left(x_l\right)\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix} 1 \\ \frac{X_{l,i}-x_l}{h_{4,l}} \end{pmatrix}\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(X_{l,i}-x_l\right)^2 \end{aligned}$$

$$
\begin{aligned}
=&\frac{1}{2}\partial^2 g_l\left(x_l\right)\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \dfrac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left[\left(\widehat{\partial F}_i-\partial F_i\right)+\partial F_i\right]^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(X_{l,i}-x_l\right)^2\\
=&\frac{1}{2}\partial^2 g_l\left(x_l\right)\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \dfrac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\partial F_i^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(X_{l,i}-x_l\right)^2\\
+&\frac{1}{2}\partial^2 g_l\left(x_l\right)\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \dfrac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left(\widehat{\partial F}_i-\partial F_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(X_{l,i}-x_l\right)^2\\
+&\partial^2 g_l\left(x_l\right)\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \dfrac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\partial F_i\left(\widehat{\partial F}_i-\partial F_i\right) T_{h_{4,l}}\left(X_{l,i}-x_l\right)\left(X_{l,i}-x_l\right)^2\\
=&\begin{pmatrix}h_{4,l}^2\frac{1}{2}\partial^2 g_l E\left\{\partial F\left[g\left(X\right)\right]\middle| X_l=x_l\right\} f_{X_l}(x_l)\kappa_{21}\\ 0\end{pmatrix}+o_p(h_{4,l}^2),
\end{aligned}
$$

and thus (A.4.62) holds.

(vi)

$$
B_{4,5}=\frac{1}{n}\sum_{i=1}^{n}\begin{pmatrix}1\\ \dfrac{X_{l,i}-x_l}{h_{4,l}}\end{pmatrix}\left(\widehat{\partial F}_i\right)^2 T_{h_{4,l}}\left(X_{l,i}-x_l\right)\frac{1}{6}\partial^2 g_l\left(\bar{X}_{l,i}\right)\left(X_{l,i}-x_l\right)^3=O_p(h_{4,l}^3),
$$

and thus (A.4.63) holds. ■

Appendix B

Appendix for Chapter 3

B.1 Proofs of the Main Results

In this section we prove Theorems 1–3 in the paper.

B.1.1 Proof of Theorem 3.3.1

Convergence Rate of $\bar{E}\left(D|X=x\right)$

Recall that $g\left(x_t\right) = \sum_{l=1}^{d} g_l(x_{l,t})$ and $\Delta g\left(x\right) = g\left(x_2\right) - g\left(x_1\right)$. Recall that $LF\left(\cdot\right) = L^{-1}\left(F\left(\cdot\right)\right)$ and $\Delta g\left(X_i\right) = g\left(X_{i2}\right) - g\left(X_{i1}\right)$. By (3.2.7) and the definition of $R\left(\cdot\right)$, we have

$$\begin{aligned} E\left(D_i|X_i\right) &= F\left(\Delta g\left(X_i\right)\right) = L(LF(\Delta g\left(X_i\right)) \\ &= L\left(R(X_i)'\pi^0 + r(X_i)\right) = L\left(R(X_i)'\pi^0\right) + \dot{L}\left(R_i^*\right) r(X_i) \\ &\equiv L\left(R(X_i)'\pi^0\right) + r_L(X_i), \end{aligned} \tag{B.1.1}$$

where R_i^* lies between $R(X_i)'\pi^0 + r(X_i)$ and $R(X_i)'\pi^0$, and $r_L(X_i) \equiv \dot{L}\left(R_i^*\right) r(X_i)$ signifies the error for the logit sieve approximation of $E\left(D_i|X_i\right)$ by $L\left(R(X_i)'\pi^0\right)$. By uniform boundedness of $\dot{L}\left(\cdot\right)$, we see that $r_L(X_i)$ behaves similarly to $r(X_i)$ in that $\sup_{x=(x_1',x_2')'\in\mathcal{X}^{\otimes 2}} |r(x)| = O\left(K^{-\gamma}\right)$ under Assumptions 9 and 13.

Let $\eta_{1Kn} = \sqrt{s_{\pi_1}\log(K^R \vee n)/n} + K^{-\gamma}$. Under Assumptions 7– 13, one can follow the proof of Theorem 6.2 in Belloni et al. (2017) hold and obtain the following result: result, we obtain

$$\frac{1}{n}\sum_{i=1}^{n}\left[R(X_i)'\left(\bar{\pi}-\pi^0\right)\right]^2 = O_p\left(\eta_{1Kn}^2\right). \tag{B.1.2}$$

Under Assumption 10(1), we can show that

$$\frac{1}{n}\sum_{i=1}^{n}\left[R(X_i)'\left(\bar{\pi}-\pi^0\right)\right]^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\bar{\pi}-\pi^0\right)' R(X_i)R(X_i)'\left(\bar{\pi}-\pi^0\right)$$

$$\geq \left(\bar{\pi}-\pi^0\right)' \lambda_{\min}\left(\frac{1}{n}\sum_{i=1}^{n} R(X_i)R(X_i)'\right)\left(\bar{\pi}-\pi^0\right)$$
$$\geq C_1 \left\|\bar{\pi}-\pi^0\right\|^2/2 \text{ w.p.a.1.} \tag{B.1.3}$$

Combining (B.1.2) and (B.1.3) yields

$$\left\|\bar{\pi}-\pi^0\right\| = O_p\left(\eta_{1Kn}\right). \tag{B.1.4}$$

Next,

$$\begin{aligned}
& \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|R(x)'\bar{\pi} - LF\left(\sum_{l=1}^{d} g_l(x_{l,t_1}) - \sum_{l=1}^{d} g_l(x_{l,t_2})\right)\right| \\
\leq & \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|R(x)'\left(\bar{\pi}-\pi^0\right)\right| + \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|R(x)'\pi^0 - LF\left(\sum_{l=1}^{d} g_l(x_{l,t_1}) - \sum_{l=1}^{d} g_l(x_{l,t_2})\right)\right| \\
\leq & \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \|R(x)\| \left\|\bar{\pi}-\pi^0\right\| + O\left(K^{-\gamma}\right) \\
= & \zeta_{0K} O_p\left(\eta_{1Kn}\right) + O\left(K^{-\gamma}\right) = O_p\left(\zeta_{0K}\eta_{1Kn}\right).
\end{aligned} \tag{B.1.5}$$

By (B.1.1) and the uniform boundedness of the first derivative of $L(\cdot)$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left[\bar{E}\left(D_i|X_i\right) - E\left(D_i|X_i\right)\right]^2 &\leq \frac{2}{n}\sum_{i=1}^{n}\left[\bar{E}\left(D_i|X_i\right) - L\left(R(X_i)'\pi^0\right)\right]^2 + \frac{2}{n}\sum_{i=1}^{n}\left[r_L\left(X_i\right)\right]^2 \\
&= \frac{2}{n}\sum_{i=1}^{n}\left[L\left(R(X_i)'\bar{\pi}\right) - L\left(R(x)'\pi^0\right)\right]^2 + \frac{2}{n}\sum_{i=1}^{n}\left[r_L\left(X_i\right)\right]^2 \\
&\lesssim \frac{2}{n}\sum_{i=1}^{n}\left[R(X_i)'\left(\bar{\pi}-\pi^0\right)\right]^2 + \frac{2}{n}\sum_{i=1}^{n}\left[r_L\left(X_i\right)\right]^2 \\
&= O_p\left(\eta_{1Kn}^2\right) + O_p\left(K^{-2\gamma}\right) = O_p\left(\eta_{1Kn}^2\right),
\end{aligned} \tag{B.1.6}$$

and

$$\begin{aligned}
& \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|\bar{E}\left(D_i|X_i=x\right) - E\left(D_i|X_i=x\right)\right| \\
\leq & \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|\bar{E}\left(D_i|X_i=x\right) - L\left(R(x)'\pi^0\right)\right| + \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|r_L\left(x\right)\right| \\
\lesssim & \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|R(x)'\left(\bar{\pi}-\pi^0\right)\right| + \sup_{x=\left(x_1',x_2'\right)'\in\mathcal{X}^{\otimes 2}} \left|r_L\left(x\right)\right| \\
= & \zeta_{0K} O_p\left(\eta_{1Kn}\right) + O_p\left(K^{-\gamma}\right) = O_p\left(K^{1/2}\eta_{1Kn}\right).
\end{aligned} \tag{B.1.7}$$

Convergence Rate of $\bar{g}_l(\cdot)$

Noting that $\triangle P_{i,j}^{1,K} = \triangle g_{i,j} - \theta_0{}'\triangle P_{i,j}^{K-1,K} + \left\{\triangle P_{i,j}^{K}{}'\beta_0 - \triangle g_{i,j}\right\}$ and recalling that $\bar{H}_{1h_1,ji} = H_{1h_1}\left[\bar{E}\left(D_j\,|X_j\right) - \bar{E}\left(D_i\,|X_i\right)\right]$, by (3.2.12) we have

$$\bar{\theta}-\theta_0=-\left\{\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}\bar{H}_{1h_1,ji}\right\}^{-1}\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}\bar{H}_{1h_1,ji}$$
$$-\left\{\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}\bar{H}_{1h_1,ji}\right\}^{-1}\frac{1}{n^2}\sum_{1\le i\neq j\le n}\left\{\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right\}\triangle P_{i,j}^{K-1,K}\bar{H}_{1h_1,ji}$$
$$\equiv -L_{0,n}^{-1}L_{1,n}-L_{0,n}^{-1}L_{2,n}, \quad \text{(B.1.8)}$$

where, e.g., $L_{0,n}\equiv\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}\bar{H}_{1h_1,ji}$. Noting that $\triangle P_{i,j}^{K-1,K}=-\triangle P_{j,i}^{K-1,K}$, we have

$$L_{1,n}=\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle g_j\triangle P_{i,j}^{K-1,K}\bar{H}_{1h_1,ji}-\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle g_i\triangle P_{i,j}^{K-1,K}\bar{H}_{1h_1,ji}$$
$$=\frac{2}{n^2}\sum_{1\le i\neq j\le n}\triangle g_j\triangle P_{i,j}^{K-1,K}\bar{H}_{1h_1,ji}. \quad \text{(B.1.9)}$$

First, we study the asymptotic properties of $L_{0,n}$. Recall that $\bar{H}_{1h_1,ji}=H_{1h_1}\left[\bar{E}\left(D_j|X_j\right)-\bar{E}\left(D_j|X_i\right)\right]$ and $\bar{m}_i=\bar{E}\left(D_j|X_j\right)$. Let $\bar{m}_{ji}=\bar{E}\left(D_j|X_j\right)-\bar{E}\left(D_j|X_i\right)$, $m_j=E\left(D_j|X_j\right)$, $m_{ji}=E\left(D_j|X_j\right)-E\left(D_j|X_i\right)$, and $H_{1h_1,ji}=H_{1h_1}\left(m_{ji}\right)=H_{1h_1}\left[E\left(D_j|X_j\right)-E\left(D_j|X_i\right)\right]$. For $i\neq j\in\{1,\cdots,n\}$

$$\bar{H}_{1h_1,ji}-H_{1h_1,ji}$$
$$=H_{h_1}\left[\bar{E}\left(D_j|X_j\right)-\bar{E}\left(D_j|X_i\right)\right]-H_{1h_1}\left[E\left(D_j|X_j\right)-E\left(D_j|X_i\right)\right]$$
$$=h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_{ji}-m_{ji}\right)+\frac{1}{2}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_{ji}-m_{ji}\right)^2+\frac{1}{6}h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)\left(\bar{m}_{ji}-m_{ji}\right)^3$$
$$=h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j-m_j\right)-h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i-m_i\right)$$
$$-h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i-m_i\right)\left(\bar{m}_j-m_i\right)+\frac{1}{2}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j-m_j\right)^2$$
$$+\frac{1}{2}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i-m_i\right)^2+h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)\left(\bar{m}_{ji}-m_{ji}\right)^3, \quad \text{(B.1.10)}$$

where m_{ji}^* is between $\bar{m}_{ji}$ and m_{ji}. It follows that

$$L_{0,n}=\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}H_{1h_1,ji}$$
$$+\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j-m_j\right)$$
$$-\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}h_1^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i-m_i\right)$$
$$-\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_i-m_i\right)\left(\bar{m}_j-m_i\right)$$

$$
\begin{aligned}
&+\frac{1}{2n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_j-m_j)^2\\
&+\frac{1}{2n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_i-m_i)^2\\
&+\frac{1}{6n^2}\sum_{1\le i\neq j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)(\bar{m}_{ji}-m_{ji})^3\\
&\equiv L_{0,n1}+L_{0,n2}+....+L_{0,n7}. \qquad\qquad (B.1.11)
\end{aligned}
$$

By B.2.1, $\sum_{\ell=2}^{7}\|L_{0,n\ell}\|_{op}=O_p\left(K^{1/2}h_1^{-1}\eta_{1Kn}\right)=o_p(1)$ and $\lambda_{\min}(L_{0,n})\ge c_{L_0}/2$ w.p.a.1.

Next, we derive the asymptotic properties of $L_{1,n}$. By (B.1.10), the symmetry of the kernel function $H(\cdot)$, and the fact that $\triangle P_{i,j}^{K-1,K}=-\triangle P_{j,i}^{K-1,K}$,

$$
\begin{aligned}
L_{1,n}=&\frac{1}{n^2}\sum_{1\le i\neq j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}H_{1h_1,ji}+\frac{2}{n^2}\sum_{1\le i\neq j\le n}\triangle g_j\triangle P_{i,j}^{K-1,K}h_1^{-1}\dot{H}_{1h_1}(m_{ji})(\bar{m}_j-m_j)\\
&-\frac{2}{n^2}\sum_{1\le i\neq j\le n}\triangle g_j\triangle P_{i,j}^{K-1,K}h_1^{-1}\dot{H}_{1h_1}(m_{ji})(\bar{m}_i-m_i)\\
&-\frac{2}{n^2}\sum_{1\le i\neq j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_i-m_i)(\bar{m}_j-m_i)\\
&+\frac{1}{2n^2}\sum_{1\le i\neq j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_j-m_j)^2\\
&+\frac{1}{2n^2}\sum_{1\le i\neq j\le n}\triangle g_j\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}(m_{ji})(\bar{m}_i-m_i)^2\\
&+\frac{1}{6n^2}\sum_{1\le i\neq j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)(\bar{m}_{ji}-m_{ji})^3\\
&\equiv\sum_{\ell=1}^{7}L_{1,n\ell}. \qquad\qquad (B.1.12)
\end{aligned}
$$

By Lemma B.2.2, $\sum_{\ell=2}^{7}\|L_{1,n\ell}\|=O_p(\eta_{1Kn})$ and $\|L_{1,n}\|=O_p(\eta_{1Kn}+\sqrt{K}h_1^{a_1})$.

Next, we study $L_{2,n}$. By (B.1.10),

$$
\begin{aligned}
L_{2,n}=&\frac{1}{n^2}\sum_{1\le i\neq j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}H_{1h_1,ji}\\
&+\frac{1}{n^2}\sum_{1\le i\neq j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}h_1^{-1}\dot{H}_{1h_1}(m_{ji})[\bar{m}_j-m_j]\\
&-\frac{1}{n^2}\sum_{1\le i\neq j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}h_1^{-1}\dot{H}_{1h_1}(m_{ji})[\bar{m}_i-m_i]\\
&-\frac{1}{n^2}\sum_{1\le i\neq j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}(m_{ji})[\bar{m}_i-m_i][\bar{m}_j-m_i]\\
&+\frac{1}{2n^2}\sum_{1\le i\neq j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}(m_{ji})[\bar{m}_j-m_j]^2
\end{aligned}
$$

$$+\frac{1}{2n^2}\sum_{1\le i\ne j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left[\bar{m}_i-m_i\right]^2$$
$$+\frac{1}{6n^2}\sum_{1\le i\ne j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}h_1^{-3}\dddot{H}_{1h_1}\left(m_{ji}^*\right)\left(\bar{m}_{ji}-m_{ji}\right)^3$$
$$\equiv\sum_{\ell=1}^{7}L_{2,n\ell}. \tag{B.1.13}$$

By Lemma B.2.3, $\sum_{\ell=2}^{7}\|L_{2,n\ell}\|=O_p\left(K^{-\gamma}\right)$ and $\|L_{2,n}\|=O_p(K^{-\gamma+1/2})$.

By the above results, we have

$$\bar{\theta}-\theta_0=-\left\{\frac{1}{n^2}\sum_{1\le i\ne j\le n}\triangle P_{i,j}^{K-1,K}\triangle P_{i,j}^{K-1,K\prime}H_{1h_1,ji}\right\}^{-1}$$
$$\times\left\{\frac{1}{n^2}\sum_{1\le i\ne j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}H_{1h_1,ji}+\frac{1}{n^2}\sum_{1\le i\ne j\le n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}H_{1h_1,ji}\right\}$$
$$+R_{1n}, \tag{B.1.14}$$

where $\|R_{1n}\|=O_p\left(\eta_{1Kn}+K^{-\gamma}\right)$.

Given the result in (B.1.14) and using the results in B.2.1-B.2.3, we can readily show that

$$\left\|\bar{\theta}-\theta_0\right\|=O_p\left(\eta_{1Kn}+\sqrt{K}h_1^{a_1}+K^{-\gamma+1/2}\right).$$

Then following the arguments as used in the derivation of (B.1.6)-(B.1.7), we have

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\left[\bar{g}_l\left(X_{l,it}\right)-g_l\left(X_{l,it}\right)\right]^2 &\le \frac{2}{n}\sum_{i=1}^{n}\left[p_K\left(X_{l,it}\right)'\left(\bar{\beta}^{x_l}-\beta_0^{x_l}\right)\right]^2+\frac{2}{n}\sum_{i=1}^{n}\left[g_l\left(X_{l,it}\right)-p_K\left(X_{l,it}\right)'\beta_0^{x_l}\right)\right]^2\\ &\lesssim \left\|\bar{\beta}^{x_l}-\beta_0^{x_l}\right\|^2+\sup_{x_l}\left|g_l\left(x_l\right)-p_K\left(x_l\right)'\beta_0^{x_l}\right)\right|^2\\ &= O_p\left(\left(\eta_{1Kn}+\sqrt{K}h_1^{a_1}+K^{-\gamma+1/2}\right)^2\right)+O_p\left(K^{-2\gamma}\right)\\ &= O_p\left(\eta_{2Kn}^2\right)\ \text{ for } l\in[d],\end{aligned}$$

and

$$\begin{aligned}\sup_{x_l\in\mathcal{X}_l}\left|\bar{g}_l\left(x_l\right)-g_l\left(x_l\right)\right| &\le \sup_{x_l\in\mathcal{X}_l}\left|p_K\left(x_l\right)'\left(\bar{\beta}^{x_l}-\beta_0^{x_l}\right)\right|+\sup_{x_l\in\mathcal{X}_l}\left[g_l\left(X_{l,it}\right)-p_K\left(X_{l,it}\right)'\beta_0^{x_l}\right)\right]^2\\ &\lesssim \sup_{x_l\in\mathcal{X}_l}\left\|p_K\left(x_l\right)\right\|\left\|\bar{\beta}^{x_l}-\beta_0^{x_l}\right\|+\sup_{x_l\in\mathcal{X}_l}\left|g_l\left(x_l\right)-p_K\left(x_l\right)'\beta_0^{x_l}\right)\right|^2\\ &= \sqrt{K}O_p\left(\eta_{1Kn}+\sqrt{K}h_1^{a_1}+K^{-\gamma+1/2}\right)+O_p\left(K^{-\gamma}\right)\\ &= O_p\left(\sqrt{K}\eta_{2Kn}\right)\ \text{ for } l\in[d].\end{aligned}$$

where $\eta_{2Kn}=\eta_{1Kn}+\sqrt{K}h_1^{a_1}+K^{-\gamma+1/2}$.

B.1.2 Proof of Theorem 3.3.2

Let

$$\widehat{U}_{\triangle g_j} = \left(\widehat{LF}(\triangle g_j), h_2\widehat{\partial LF}(\triangle g_j), \cdots, h_2^{a_2}\widehat{\partial^{a_2} LF}(\triangle g_j)\right)' - (LF(\triangle g_j), h_2\partial LF(\triangle g_j), \cdots, h_2^{a_2}\partial^{a_2} LF(\triangle g_j))'.$$

Let $\vartheta_{i,j}(b) = b_0 + \sum_{l=1}^{a_2} \frac{1}{h_2^l}(\triangle\bar{g}_{i,j})^l b_l$, where $b = (b_0, b_1, ..., b_{a_2})'$. Noting that $L(z) = \exp(z)/(1+\exp(z))$, we have

$$\begin{aligned}
&Q_n(\triangle\bar{g}_j, b) \\
&= \frac{-1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\{D_i \ln[L(\vartheta_{i,j}(b))] + (1-D_i)\ln[1-L(\vartheta_{i,j}(b))]\} \\
&= \frac{-1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\{D_i\vartheta_{i,j}(b) - D_i\ln[1+\exp(\vartheta_{i,j}(b))] - (1-D_i)\ln[1+\exp(\vartheta_{i,j}(b))]\} \\
&= \frac{1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\{\ln[1+\exp(\vartheta_{i,j}(b))] - D_i\vartheta_{i,j}(b)\}.
\end{aligned}$$

For an arbitrary $U_{\triangle g_j} \in \mathbb{R}^{a_2+1}$ and $\tau \in \mathbb{R}$, let

$$l_{i,j}(\tau) = H_{2h_2}(\triangle\bar{g}_{i,j})\ln\left(1+\exp\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}(\triangle\bar{g}_{i,j})^l\partial^l LF(\triangle g_j) + \tau\varsigma_{1i,j}' U_{\triangle g_j}\right)\right).$$

where $\varsigma_{1i,j} = \left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2} a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)'$. Then

$$\begin{aligned}
l_{i,j}'(\tau) &= H_{2h_2}(\triangle\bar{g}_{i,j})\varsigma_{1i,j}' U_{\triangle g_j} L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j) + \tau\varsigma_{1i,j}' U_{\triangle g_j}\right) \text{ and} \\
l_{i,j}''(\tau) &= H_{2h_2}(\triangle\bar{g}_{i,j})\left[\varsigma_{1i,j}' U_{\triangle g_j}\right]^2 L'\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j) + \tau\varsigma_{1i,j}' U_{\triangle g_j}\right).
\end{aligned}$$

It is easy to see that $\left|l_{i,j}'''(\tau)\right| \le \left|l_{i,j}''(\tau)\right|\left|\varsigma_{1i,j}' U_{\triangle g_j}\right|$. Define

$$\begin{aligned}
\widehat{U}_{\triangle g_i} &= \arg\max_{U_{\triangle g_j}}\left\{Q_n\left(\triangle\bar{g}_j, \left(\left(\widehat{LF}(\triangle g_j), h_2\widehat{\partial LF}(\triangle g_j), \cdots, h_2^{a_2}\widehat{\partial^{a_2} LF}(\triangle g_j)\right)' + U_{\triangle g_j}\right)\right)\right. \\
&\quad \left. - Q_n(\triangle\bar{g}_j, (LF(\triangle g_j), \partial LF(\triangle g_j), \cdots, \partial^{a_2} LF(\triangle g_j))')\right\} \\
&= \arg\max_{U_{\triangle g_j}}\frac{1}{n}\sum_{i=1}^{N}[l_{i,j}(1) - l_{i,j}(0)] - \frac{1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\varsigma_{1i,j}' U_{\triangle g_j} D_i.
\end{aligned}$$

We calculate the first order derivative with respect to τ:

$$\begin{aligned}
&\partial_\tau Q_n\left(\triangle\bar{g}_j, (LF(\triangle g_j), \partial LF(\triangle g_j), \cdots, \partial^{a_2} LF(\triangle g_j))' + \tau U_{\triangle g_j}\right) \\
&= \frac{1}{n}\sum_{i=1}^{N} H_{2h_2}(\triangle\bar{g}_{i,j})\varsigma_{1i,j}' U_{\triangle g_j}\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j) + \tau\varsigma_{1i,j}' U_{\triangle g_j}\right) - D_i\right\}.
\end{aligned}$$

Evaluating the above derivative at τ= 0 yields

$$\begin{aligned}
&\partial_\tau Q_n \left(\triangle\bar{g}_j, (LF(\triangle g_j), \partial LF(\triangle g_j), \cdots, \partial^{a_2} LF(\triangle g_j))' + \tau U_{\triangle g_j}\right)\big|_{\tau=0} \\
=&\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma'_{1i,j}U_{\triangle g_j}\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle g_j)\right) - D_i\right\} \\
=&\frac{1}{n}\sum_{i=1}^{N} l'_{i,j}(0) - \frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma'_{1i,j}U_{\triangle g_j}D_i.
\end{aligned}$$

Let

$$\begin{aligned}
G_n\left(U_{\triangle g_j}\right) =& Q_n\left(\triangle\bar{g}_j, \left(\left(\widehat{LF}(\triangle g_j), h_2\widehat{\partial LF}(\triangle g_j), \cdots, h_2^{a_2}\widehat{\partial^{a_2} LF}(\triangle g_j)\right)' + U_{\triangle g_j}\right)\right) \\
&- Q_n\left(\triangle\bar{g}_j, (LF(\triangle g_j), \partial LF(\triangle g_j), \cdots, \partial^{a_2} LF(\triangle g_j))'\right) \\
&- \partial_\tau Q_n\left(\triangle\bar{g}_j, (LF(\triangle g_j), \partial LF(\triangle g_j), \cdots, \partial^{a_2} LF(\triangle g_j))' + \tau U_{\triangle g_j}\right)\big|_{\tau=0}.
\end{aligned}$$

Let $\varsigma_{i,j} = \varsigma'_{1i,j}U_{\triangle g_j}$. Noting that $L'(x) = L(x)\left[1 - L(x)\right]$ and $F(\triangle g_i) \in (0,1)$, there exists a positive constant $\underline{c}> 0$ such that

$$\begin{aligned}
G_n\left(U_{\triangle g_j}\right) &= \frac{1}{n}\sum_{i=1}^{N}\left[l_{i,j}(1) - l_{i,j}(0) - l'_{ij}(0)\right] \\
&\geq \frac{1}{n}\sum_{i=1}^{N}\frac{l_i''(0)}{\varsigma_{i,j}^2}\left[\exp\left(-\left|\varsigma_{i,j}\right|\right) + \left|\varsigma_{i,j}\right| - 1\right] \\
&= \frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right) L'\left(\sum_{l=0}^{a_2}\triangle\bar{g}_{i,j}^l\partial^l LF\left(\triangle g_j\right)/l!\right)\left[\exp\left(-\left|\varsigma_{i,j}\right|\right) + \left|\varsigma_{i,j}\right| - 1\right] \\
&\geq \frac{\underline{c}}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left[\exp\left(-\left|\varsigma_{i,j}\right|\right) + \left|\varsigma_{i,j}\right| - 1\right] \\
&\geq \frac{\underline{c}}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left(\frac{\varsigma_{i,j}^2}{2} - \frac{\left|\varsigma_{i,j}\right|^3}{6}\right),
\end{aligned}$$

where the first inequality holds by Lemma 1 in **?** and the last inequality follows from the fact that

$$e^{-x} + x - 1 \geq \frac{x^2}{2} - \frac{x^3}{6}\ \forall x > 0.$$

By Step 1 in the proof of Theorem 5.6 in **?**, there exists a positive constant $\underline{c}^*$ such that

$$\begin{aligned}
G_n\left(\widehat{U}_{\triangle g_j}\right) &\geq \frac{\underline{c}^*}{3}\min\left(\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{i,j}^2, \bar{l}\left[\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{i,j}^2\right]^{1/2}\right) \\
&\geq \frac{\underline{c}^*}{3}\min\left(\left\|\widehat{U}_{\triangle g_j}\right\|^2, \bar{l}\left\|\widehat{U}_{\triangle g_j}\right\|\right), \qquad \text{(B.1.15)}
\end{aligned}$$

where

$$\bar{l} = \inf_{U\in\mathbb{R}^{a_2+1}}\frac{\left\{\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\varsigma_{i,j}^2\right\}^{3/2}}{\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left|\varsigma_{i,j}\right|^3}.$$

Noting that

$$\bar{l} \geq \inf_{U \in \mathbb{R}^{a_2+1}} \frac{\left\{\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{i,j}^2\right\}^{1/2}}{\max\limits_{i \neq j \in \{1,\cdots,n\}} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\|\varsigma_{1i,j}\|\left\|U_{\triangle g_j}\right\|} \geq O_p\left(h_2\right),$$

we have

$$\bar{l}\left(h_2^{a_2+1} + \sqrt{\ln\left(n\right)/\left(nh_2\right)}\right)^{-1} \xrightarrow{p} \infty. \tag{B.1.16}$$

In addition, by construction and the submultiplicative and triangle inequalities

$$\begin{aligned}
G_n\left(\widehat{U}_{\triangle g_j}\right) &\leq \left|\partial_t Q_n\left(\triangle \bar{g}_j, \left(LF(\triangle g_j), \partial LF(\triangle g_j), \cdots, \partial^{a_2} LF(\triangle g_j)\right)' + \tau U_{\triangle g_j}\right)\big|_{\tau=0}\right| \\
&= \left|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}'\widehat{U}_{\triangle g_j}\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l \partial^l LF(\triangle g_j)\right) - D_i\right\}\right| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l \partial^l LF(\triangle g_j)\right) - D_i\right\}\right\|\left\|\widehat{U}_{\triangle g_j}\right\| \\
&\leq \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}'\left\{L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l \partial^l LF(\triangle g_j)\right) - F(\triangle g_i)\right\}\right\|\left\|\widehat{U}_{\triangle g_j}\right\| \\
&+ \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}'\left[F(\triangle g_i) - D_i\right]\right\|\left\|\widehat{U}_{\triangle g_j}\right\|.
\end{aligned} \tag{B.1.17}$$

Noting that $L\left(LF(\triangle \bar{g}_i)\right) = F(\triangle \bar{g}_i)$, by Taylor expansions we have

$$\begin{aligned}
&L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l \partial^l LF(\triangle g_j)\right) - F(\triangle g_i) \\
&= \left[F(\triangle \bar{g}_i) - F(\triangle g_i)\right] + \left[L\left(\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l \partial^l LF(\triangle g_j) - LF(\triangle \bar{g}_i) + LF(\triangle \bar{g}_i)\right) - F(\triangle \bar{g}_i)\right] \\
&= \left[F(\triangle \bar{g}_i) - F(\triangle g_i)\right] + L'\left(LF(\triangle \bar{g}_i)\right)\chi_{i,j} + \frac{1}{2}L''\left(LF(\triangle \bar{\bar{g}}_{i,j})\right)\chi_{i,j}^2.
\end{aligned} \tag{B.1.18}$$

where $\triangle \bar{\bar{g}}_{i,j}$ is between $\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l \partial^l LF(\triangle g_j)$ and $LF(\triangle \bar{g}_i)$, and

$$\chi_{i,j} = \sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l\left[\partial^l LF(\triangle g_j) - \partial^l LF(\triangle \bar{g}_j)\right]/l! - \sum_{l=a_2+1}^{\infty}\frac{1}{h_2^l l!}\triangle \bar{g}_{i,j}^l \partial^l LF(\triangle \bar{g}_j)/l!.$$

Substituting (B.1.18) into (B.1.17) yields

$$\begin{aligned}
G_n(\widehat{U}_{\triangle g_j}) \leq &\left\{\left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}L'\left(LF(\triangle \bar{g}_i)\right)\chi_{i,j}\right\| + \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}\left[F(\triangle \bar{g}_i) - F(\triangle g_i)\right]\right\|\right. \\
&+ \left.\left\|\frac{1}{2n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}L''\left(LF(\triangle \bar{\bar{g}}_{i,j})\right)\chi_{i,j}^2\right\| + \left\|\frac{1}{n}\sum_{i=1}^{N} H_{2h_2}\left(\triangle \bar{g}_{i,j}\right)\varsigma_{1i,j}\left[F(\triangle g_i) - D_i\right]\right\|\right\} \\
&\times \left\|\widehat{U}_{\triangle g_j}\right\| \\
&\equiv \left\{\|G_{1,nj}\| + \|G_{2,nj}\| + \|G_{3,nj}\| + \|G_{4,nj}\|\right\}\left\|\widehat{U}_{\triangle g_j}\right\|,
\end{aligned} \tag{B.1.19}$$

where the definitions of $G_{\ell,nj}$, $\ell = 1, ..., 4$, are self-evident. By Lemma B.2.4, we have uniformly over $j \in \{1, \cdots, n\}$,

$$\left|G_n\left(\widehat{U}_{\triangle g_j}\right)\right| \le c_F \left|\triangle g_j - \triangle \bar{g}_j\right| \left\|\widehat{U}_{\triangle g_j}\right\| + O_p\left(h_2^{a_2+1} + \sqrt{\log(n)/(nh_2)} + \eta_{2Kn}\right)\left\|\widehat{U}_{\triangle g_j}\right\|. \quad \text{(B.1.20)}$$

Combining (B.1.15) and (B.1.20), we have uniformly over $j \in \{1, \cdots, n\}$

$$\frac{c^*}{3} \min\left(\left\|\widehat{U}_{\triangle g_j}\right\|, \bar{l}\right) \lesssim \left|\triangle g_j - \triangle \bar{g}_j\right| + O_p\left(h_2^{a_2+1} + \sqrt{\log(n)/(nh_2)} + \eta_{2Kn}\right).$$

which, in conjunction with (B.1.16) implies that

$$\left\|\widehat{U}_{\triangle g_j}\right\| \lesssim \left|\triangle g_j - \triangle \bar{g}_j\right| + O_p\left(h_2^{a_2+1} + \sqrt{\log(n)/(nh_2)} + \eta_{2Kn}\right)$$

uniformly over $j \in \{1, \cdots, n\}$. The completes the proof of (i).

Given the above uniform rate, (ii) follows automatically.

B.1.3 Proof of Theorem 3.3.3

Convergence Rate of $\left(\widehat{g}_l(x_l),\ \widehat{\dot{g}}_l(x_l)\right)'$

Let $\widehat{U}_{x_l} = \left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)' - \left(g_l(x_l),\ h_3\dot{g}_l(x_l)\right)'$. Let $c = (c_0, c_1)$ and $H_{3h_3,itx_l} = H_{3h_3}\left(X_{l,it} - x_l\right)$ for $t = 1, 2$. Define

$$\begin{aligned} W_{1,nx_l}(c) = \frac{1}{n}\sum_{i=1}^{N} H_{3h_3,i1x_l}\Bigg\{&\ln\left(1 + \exp\left(\widehat{LF}_i - \widehat{\dot{LF}}_i \cdot \left[c_0 + c_1\frac{1}{h_3}\left(X_{l,i1} - x_l\right) - \bar{g}_l\left(X_{l,i1}\right)\right]\right)\right) \\ &+ D_i\widehat{\dot{LF}}_i \cdot \left[c_0 + c_1\frac{1}{h_3}\left(X_{l,i1} - x_l\right)\right] - D_i\left(\widehat{LF}_i + \widehat{\dot{LF}}_i\bar{g}_l\left(X_{l,i1}\right)\right)\Bigg\}, \end{aligned}$$

and

$$\begin{aligned} W_{2,nx_l}(c) = \frac{1}{n}\sum_{i=1}^{N} H_{3h_3,i2x_l}\Bigg\{&\ln\left(1 + \exp\left(\widehat{LF}_i + \widehat{\dot{LF}}_i \cdot \left[c_0 + c_1\frac{1}{h_3}\left(X_{l,i2} - x_l\right) - \bar{g}_l\left(X_{l,i2}\right)\right]\right)\right) \\ &- D_i\widehat{\dot{LF}}_i \cdot \left[c_0 + c_1\frac{1}{h_3}\left(X_{l,i2} - x_l\right)\right] - D_i\left(\widehat{LF}_i - \widehat{\dot{LF}}_i\bar{g}_l\left(X_{l,i2}\right)\right)\Bigg\}. \end{aligned}$$

Then $W_{nx_l}(c) = W_{1,nx_l}(c) + W_{2,nx_l}(c)$. Let $U_{x_l} = \left(c_0 - g_l(x_l),\ c_1 - h_3\dot{g}_l(x_l)\right)' \in \mathbb{R}^2$. Then

$$\begin{aligned} &\widehat{LF}_i + \widehat{\dot{LF}}_i \cdot \left[c_0 + c_1\frac{1}{h_3}\left(X_{l,i2} - x_l\right) - \bar{g}_l\left(X_{l,i2}\right)\right] \\ &= \widehat{LF}_i + \widehat{\dot{LF}}_i\left[\left(c_0 - g_l(x_l)\right) + \left(c_1 - h_3\dot{g}_l(x_l)\right)\frac{1}{h_3}\left(X_{l,i2} - x_l\right) + g_l(x_l) + \dot{g}_l(x_l)h_3\frac{1}{h_3}\left(X_{l,i2} - x_l\right) - \bar{g}_l\left(X_{l,i2}\right)\right] \end{aligned}$$

$$= \widehat{LF}_i - \widehat{\dot{LF}}_i \left[\bar{g}_l \left(X_{l,i2} \right) - g_l(x_l) - \dot{g}_l(x_l) h_3 \frac{1}{h_3} \left(X_{l,i2} - x_l \right) \right] + \widehat{\dot{LF}}_i \left(1, \frac{1}{h_3} \left(X_{l,i2} - x_l \right) \right) U_{x_l}$$
$$= \widehat{LF}_i - \widehat{\dot{LF}}_i \cdot r_{l,i2} + \widehat{\dot{LF}}_i u_{x_l,i2},$$

and similarly,

$$\begin{aligned}
&\widehat{LF}_i - \widehat{\dot{LF}}_i \cdot \left[c_0 + c_1 \frac{1}{h_3} \left(X_{l,i1} - x_l \right) - \bar{g}_l \left(X_{l,i1} \right) \right] \\
&= \widehat{LF}_i + \widehat{\dot{LF}}_i \left[\bar{g}_l \left(X_{l,i1} \right) - g_l(x_l) - \dot{g}_l(x_l) h_3 \frac{1}{h_3} \left(X_{l,i1} - x_l \right) \right] - \widehat{\dot{LF}}_i \left(1, \frac{1}{h_3} \left(X_{l,i1} - x_l \right) \right)' U_{x_l} \\
&= \widehat{LF}_i + \widehat{\dot{LF}}_i \cdot r_{l,i1} - \widehat{\dot{LF}}_i u_{x_l,i1},
\end{aligned}$$

where $r_{x_l,it} \equiv \bar{g}_l \left(X_{l,it} \right) - g_l(x_l) - \dot{g}_l(x_l) h_3 \frac{1}{h_3} \left(X_{l,it} - x_l \right)$ and $u_{x_l,it} = \left(1, \frac{1}{h_3} \left(X_{l,it} - x_l \right) \right)' U_{x_l}$ for $t = 1, 2$. Further define

$$\begin{aligned}
l_{l,i1}(\tau) &= H_{3h_3,i1x_l} \ln \left(1 + \exp \left(\widehat{LF}_i + \widehat{\dot{LF}}_i \cdot r_{x_l,i1} - \tau \widehat{\dot{LF}}_i \cdot u_{x_l,i1} \right) \right) \text{ and} \\
l_{l,i2}(\tau) &= H_{3h_3,i2x_l} \ln \left(1 + \exp \left(\widehat{LF}_i - \widehat{\dot{LF}}_i \cdot r_{x_l,i2} + \tau \widehat{\dot{LF}}_i \cdot u_{x_l,i2} \right) \right),
\end{aligned}$$

Then we have

$$\begin{aligned}
l'_{l,i1}(\tau) &= -\widehat{\dot{LF}}_i \cdot u_{x_l,i1} H_{3h_3,i1x_l} L \left(\widehat{LF}_i + \widehat{\dot{LF}}_i \cdot r_{x_l,i1} - \tau \widehat{\dot{LF}}_i \cdot u_{x_l,i1} \right), \\
l'_{l,i2}(\tau) &= \widehat{\dot{LF}}_i \cdot u_{x_l,i2} H_{3h_3,i2x_l} L \left(\widehat{LF}_i - \widehat{\dot{LF}}_i \cdot r_{x_l,i2} + \tau \widehat{\dot{LF}}_i \cdot u_{x_l,i2} \right), \\
l''_{l,i1}(\tau) &= \widehat{\dot{LF}}_i^2 \cdot u_{x_l,i1}^2 H_{3h_3,i1x_l} L' \left(\widehat{LF}_i + \widehat{\dot{LF}}_i \cdot r_{x_l,i1} - \tau \widehat{\dot{LF}}_i \cdot u_{x_l,i1} \right), \text{ and} \\
l''_{l,i2}(\tau) &= \widehat{\dot{LF}}_i^2 \cdot u_{x_l,i2}^2 H_{3h_3,i2x_l} L' \left(\widehat{LF}_i - \widehat{\dot{LF}}_i \cdot r_{x_l,i2} + \tau \widehat{\dot{LF}}_i \cdot u_{x_l,i2} \right).
\end{aligned}$$

It is straightforward to show that

$$\left| l'''_{l,it}(\tau) \right| \leq l''_{l,it}(\tau) \left| \widehat{\dot{LF}}_i \cdot u_{x_l,it} \right| \text{ for } t = 1, 2.$$

Define

$$\begin{aligned}
\widehat{U}_{\triangle g_i} = \arg\max_{U_{x_l}} &\left\{ W_{1,nx_l} \left((g_l(x_l), \dot{g}_l(x_l)) + U'_{x_l} \right) + W_{2,nx_l} \left((g_l(x_l), \dot{g}_l(x_l)) + U'_{x_l} \right) \right. \\
&\left. - W_{1,nx_l} \left(g_l(x_l), \dot{g}_l(x_l) \right) - W_{2,nx_l} \left(g_l(x_l), \dot{g}_l(x_l) \right) \right\} \\
= \arg\max_{U_{x_l}} &\left\{ \frac{1}{n} \sum_{i=1}^{N} \left[l_{l,i1}(1) - l_{l,i1}(0) \right] + \frac{1}{n} \sum_{i=1}^{N} \left[l_{l,i2}(1) - l_{l,i2}(0) \right] + \frac{1}{n} \sum_{i=1}^{N} D_i H_{3h_3,i1x_l} \widehat{\dot{LF}}_i \cdot u_{x_l,i1} \right. \\
&\left. - \frac{1}{n} \sum_{i=1}^{N} D_i H_{3h_3,i2x_l} \widehat{\dot{LF}}_i \cdot u_{x_l,i2} \right\}.
\end{aligned}$$

We calculate the first order derivative at τ,

$$\partial_\tau W_{1,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)=-\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i1x_l}u_{x_l,i1}\left\{L\left(\widehat{LF}_i+\widehat{\dot{LF}}_i r_{x_l,i1}-\tau\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right)-D_i\right\},$$

and

$$\partial_\tau W_{2,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)=\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}u_{x_l,i2}\left\{L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i r_{x_l,i2}+\tau\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right)-D_i\right\}.$$

Evaluating the above derivatives at $\tau=0$ yields

$$\begin{aligned}\partial_\tau W_{1,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)\big|_{\tau=0} &= -\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i1x_l}u_{x_l,i1}\left\{L\left(\widehat{LF}_i+\widehat{\dot{LF}}_i r_{x_l,i1}\right)-D_i\right\},\\ \partial_\tau W_{2,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)\big|_{\tau=0} &= \frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}u_{x_l,i2}\left\{L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i r_{x_l,i2}\right)-D_i\right\}.\end{aligned}$$

Let

$$\begin{aligned}G_n\left(U_{x_l}\right)=&\,W_{1,nx_l}\left((g_l(x),\dot{g}_l(x))+U'_{x_l}\right)+W_{2,nx_l}\left((g_l(x),\dot{g}_l(x))+U'_{x_l}\right)-W_{1,nx_l}\left(g_l(x),\dot{g}_l(x)\right)\\ &-W_{2,nx_l}\left(g_l(x),\dot{g}_l(x)\right)-\partial_\tau W_{1,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)\big|_{\tau=0}-\partial_\tau W_{2,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U'_{x_l}\right)\big|_{\tau=0}.\end{aligned}$$

Noting that $L'(x)=L(x)\left[1-L(x)\right]$ and $F(\triangle g_i)\in(0,1)$. There exist a positive constant C_1 such that

$$\begin{aligned}G_n\left(U_{x_l}\right)&=\frac{1}{n}\sum_{i=1}^{N}\left[l_{l,i2}(1)-l_{l,i2}(0)-{l_{l,i2}}'(0)\right]+\frac{1}{n}\sum_{i=1}^{N}\left[l_{1,i}(1)-l_{1,i}(0)-{l_{1,i}}'(0)\right]\\ &\geq\frac{1}{n}\sum_{i=1}^{N}\frac{{l_i}''(0)}{\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right|^2}\left[\exp\left(-\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right|\right)+\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right|-1\right]\\ &+\frac{1}{n}\sum_{i=1}^{N}\frac{{l_i}''(0)}{\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right|^2}\left[\exp\left(-\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right|\right)+\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right|-1\right]\\ &=\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i2x_l}L'\left(\widehat{LF}_i-\widehat{\dot{LF}}_i\cdot r_{x_l,i2}\right)\left[\exp\left(-\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right|\right)+\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right|-1\right]\\ &+\frac{1}{n}\sum_{i=1}^{N}T_{h_3}\left(X_{,l,i1}-x_l\right)L'\left(\widehat{LF}_i+\widehat{\dot{LF}}_i\cdot r_{x_l,i1}\right)\left[\exp\left(-\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right|\right)+\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right|-1\right]\\ &\geq C_1\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i2x_l}\left(\frac{1}{2}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right]^2-\frac{1}{6}\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right|^3\right)\\ &+C_1\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i1x_l}\left(\frac{1}{2}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right]^2-\frac{1}{6}\left|\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right|^3\right),\end{aligned}$$

where the first inequality holds by Lemma 1 in **?** and the last inequality holds because

$$e^{-x}+x-1\geq\frac{x^2}{2}-\frac{x^3}{6}\ \forall\, x>0.$$

By Step 1 in the proof of Theorem 5.6 in **?**, there exist some positive constant $\underline{C}_1$ and $\underline{C}$ such that

$$\begin{aligned}G_n\left(\widehat{U}_{x_l}\right) \geq & \frac{1}{3}\min\Bigg(\underline{C}_1\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i1x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right]^2+\underline{C}_1\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i2x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right]^2, \\ & \underline{C}_1\bar{l}_1\left[\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i1x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i1}\right]^2\right]^{1/2}+\underline{C}_1\bar{l}_2\left[\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,i2x_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,i2}\right]^2\right]^{1/2}\Bigg) \\ \geq & \frac{\underline{C}}{3}\min\left(\left\|\widehat{U}_{x_l}\right\|^2,\left(\bar{l}_1+\bar{l}_2\right)\left\|\widehat{U}_{x_l}\right\|\right),\end{aligned} \tag{B.1.21}$$

where

$$\bar{l}_t=\inf_{U_{x_l}\in\mathbb{R}^{a_2+1}}\frac{\left\{\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,itx_l}\left[\widehat{\dot{LF}}_i\cdot u_{x_l,it}\right]^2\right\}^{3/2}}{\frac{1}{n}\sum_{i=1}^{N}H_{3h_3,itx_l}\left|\widehat{\dot{LF}}_i\cdot u_{x_l,it}\right|^3}\text{ for } t=1,2.$$

As in (B.1.16), we have

$$\left(\bar{l}_1+\bar{l}_2\right)\left(h_3^2+\sqrt{1/(nh_3)}\right)^{-1}\xrightarrow{p}\infty. \tag{B.1.22}$$

In addition, by construction,

$$\begin{aligned}G_n\left(\widehat{U}_{x_l}\right) & \leq\left|\partial_\tau W_{1,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U_{x_l}'\right)\big|_{\tau=0}+\partial_\tau W_{2,nx_l}\left((g_l(x),\dot{g}_l(x))+\tau U_{x_l}'\right)\big|_{\tau=0}\right| \\ & \leq\left|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_iH_{3h_3,i1x_l}u_{x_l,i1}\left\{L\left(\widehat{LF}_i+\widehat{\dot{LF}}_i\cdot r_{x_l,i1}\right)-D_i\right\}\right| \\ & +\left|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_iH_{3h_3,i2x_l}u_{x_l,i2}\left\{L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i\cdot r_{x_l,i2}\right)-D_i\right\}\right| \\ & \leq\|U_{x_l}\|\left\{\left\|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_iH_{3h_3,i1x_l}\mu_{x_l,it}\left\{L\left(\widehat{LF}_i+\widehat{\dot{LF}}_i\cdot r_{x_l,i1}\right)-D_i\right\}\right\|\right. \\ & \left.+\left\|\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_iH_{3h_3,i2x_l}\mu_{x_l,it}\left\{L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i\cdot r_{x_l,i2}\right)-D_i\right\}\right\|\right\} \\ & \equiv\|U_{x_l}\|\left\{\|D_{1n}(x_l)\|+\|D_{2n}(x_l)\|\right\},\end{aligned} \tag{B.1.23}$$

where $\mu_{x_l,it}=\left(1,\frac{1}{h_3}\left(X_{l,it}-x_l\right)\right)'$.

Note that $L\left(LF_i\right)=F_i$. By Taylor expansions,

$$\begin{aligned}& L\left(\widehat{LF}_i-\widehat{\dot{LF}}_i\cdot r_{x_l,i2}\right)-D_i \\ & =L\left(LF_i+\widehat{LF}_i-LF_i-\widehat{\dot{LF}}_i\left[\bar{g}_l\left(X_{l,i2}\right)-g_l\left(X_{l,i2}\right)\right]+\frac{1}{2}\widehat{\dot{LF}}_i\ddot{g}_l(\bar{x}_{li})\left(X_{l,i2}-x_l\right)^2\right)-D_i \\ & =L\left(LF_i+\delta_{x_l,i2}\right)-D_i \\ & =F_i-D_i+L'\left(LF_i\right)\delta_{x_l,i2}+\frac{1}{2}L''\left(\bar{LF}_i\right)\delta_{x_l,i2}^2,\end{aligned} \tag{B.1.24}$$

where $\delta_{x_l,i2} = \widehat{LF}_i - LF_i - \widehat{\dot{LF}}_i \left[\bar{g}_l\left(X_{l,i2}\right) - g_l\left(X_{l,i2}\right)\right] + \frac{1}{2}\widehat{\dot{LF}}_i \ddot{g}_l(\bar{x}_{l,i2})\left(X_{l,i2} - x_l\right)^2$, $\bar{x}_{l,i2}$ is between $X_{l,i2}$ and x_l, and $\bar{LF}_i$ is between LF_i and $LF_i + \delta_{x_l,i2}$. Similarly,

$$\begin{aligned}
& L\left(\widehat{LF}_i + \widehat{\dot{LF}}_i \cdot r_{x_l,i1}\right) - D_i \\
&= L\left(LF_i + \widehat{LF}_i - LF_i + \widehat{\dot{LF}}_i \left[\bar{g}_l\left(X_{l,i1}\right) - g_l\left(X_{l,i1}\right)\right] - \frac{1}{2}\widehat{\dot{LF}}_i \ddot{g}_l(\bar{x}_{li})\left(X_{l,i1} - x_l\right)^2\right) - D_i \\
&= L\left(LF_i + \delta_{x_l,i1}\right) - D_i \\
&= F_i - D_i + L'\left(LF_i\right)\delta_{x_l,i1} + \frac{1}{2}L''\left(\bar{LF}_i\right)\delta^2_{x_l,i1}, \qquad \text{(B.1.25)}
\end{aligned}$$

where $\delta_{x_l,i1} = \widehat{LF}_i - LF_i + \widehat{\dot{LF}}_i \left[\bar{g}_l\left(X_{l,i1}\right) - g_l\left(X_{l,i1}\right)\right] - \frac{1}{2}\widehat{\dot{LF}}_i \ddot{g}_l(\bar{x}_{l,i1})\left(X_{l,i1} - x_l\right)^2$, $\bar{x}_{l,i1}$ is between $X_{l,i1}$ and x_l, and $\bar{LF}_i$ is between LF_i and $LF_i + \delta_{x_l,i1}$. Let

Then by (B.1.24),

$$\begin{aligned}
D_{1n}\left(x_l\right) &\equiv \frac{1}{n}\sum_{i=1}^{n}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\mu_{x_l,it}\left\{L\left(\widehat{LF}_i - \widehat{\dot{LF}}_i \cdot r_{x_l,i2}\right) - D_i\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}\dot{LF}_i H_{3h_3,i2x_l}\mu_{x_l,it}\left(F_i - D_i\right) + \frac{1}{N}\sum_{i=1}^{n}(\widehat{\dot{LF}}_i - \dot{LF}_i)H_{3h_3,i2x_l}\mu_{x_l,it}\left(F_i - D_i\right) \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\mu_{x_l,it}L'\left(LF_i\right)\delta_{x_l,i2} + \frac{1}{2n}\sum_{i=1}^{n}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\mu_{x_l,it}L''\left(\bar{LF}_i\right)\delta^2_{x_l,i2} \\
&\equiv D_{1n,1}\left(x_l\right) + D_{1n,2}\left(x_l\right) + D_{1n,3}\left(x_l\right) + D_{1n,4}\left(x_l\right).
\end{aligned}$$

It is standard to show that

$$\begin{aligned}
\left\|D_{1n,1}\left(x_l\right)\right\| &= O_p\left(\sqrt{1/(nh_3)}\right) \text{ for each } x_l \in \mathcal{X}_l \text{ and} \\
\max_{x_l\in\mathcal{X}_l}\left\|D_{1n,1}\left(x_l\right)\right\| &= O_p\left(\sqrt{\log(n)/(nh_3)}\right).
\end{aligned}$$

In addition, we can show that $D_{1n,2}\left(x_l\right) = O_p\left(\eta_{3Kn}\right)$, $D_{1n,3}\left(x_l\right) = O_p\left(h_3^2 + \eta_{3Kn}\right)$, and $D_{1n,4}\left(x_l\right) = O_p\left(h_3^4 + \eta^2_{3Kn}\right)$ uniformly in $x_l \in \mathcal{X}_l$ by using Theorem 3.3.2. It follows that

$$\begin{aligned}
D_{1n}\left(x_l\right) &= O_p\left(h_3^2 + \sqrt{1/(nh_3)} + \eta_{3Kn}\right) \text{ for each } x_l \in \mathcal{X}_l \text{ and} \\
\max_{x_l\in\mathcal{X}_l}\left\|D_{1n}\left(x_l\right)\right\| &= O_p\left(h_3^2 + \sqrt{\log(n)/(nh_3)} + \eta_{3Kn}\right).
\end{aligned}$$

The same conclusion holds for $D_{2n}\left(x_l\right)$. Consequently, by (B.1.23)

$$G_n\left(\widehat{U}_{x_l}\right) \le O_p\left(h_3^2 + \sqrt{1/(nh_3)} + \eta_{3Kn}\right)\left\|\widehat{U}_{x_l}\right\|. \qquad \text{(B.1.26)}$$

Combining (B.1.21) and (B.1.26), we have

$$\frac{c}{3}\min\left(\left\|\widehat{U}_{x_l}\right\|, \bar{l}\right) \le O_p\left(h_3^2 + \sqrt{1/(nh_3)} + \eta_{3Kn}\right).$$

This result, in conjunction with (B.1.22), implies that

$$\left\|\widehat{U}_{x_l}\right\| = O_p\left(h_3^2 + \sqrt{1/(nh_3)} + \eta_{3Kn}\right).$$

In addition, our conditions ensure that $\eta_{3Kn} = o\left(h_3^2 + \sqrt{h_3/n}\right)$. It follows that

$$\left\|\left(\widehat{g}_l(x_l),\ \frac{1}{h_3}\widehat{\dot{g}}_l(x_l)\right) - \left(g_l(x_l),\ \frac{1}{h_3}\dot{g}_l(x_l)\right)\right\| = \left\|\widehat{U}_{x_l}\right\| = O_p\left(h_3^2 + \sqrt{1/(nh_3)}\right).$$

The above results can be made to hold uniformly in x_l with little modification: $\max_{x_l\in\mathcal{X}_l}\left\|\widehat{U}_{x_l}\right\| = O_p\left(h_3^2 + \sqrt{\log(n)/(nh_3)}\right)$.

Asymptotic Distribution of $\left(\widehat{g}_l(x_l),\ \widehat{\dot{g}}_l(x_l)\right)$

Noting that $\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)' = \arg\min_{c_0,c_1} W_{n,x_l}(c_0,c_1)$, we have

$$\begin{aligned}
&\left.\frac{\partial W_{n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(c_0,c_1)=\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)} \\
=\ &\left.\frac{\partial W_{1,nx_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(c_0,c_1)=\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)} + \left.\frac{\partial W_{2,nx_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(c_0,c_1)=\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)'} = 0.
\end{aligned}$$

Since we have already proved that $\left(\widehat{g}_l(x_l),\ \widehat{\dot{g}}_l(x_l)\right)' \xrightarrow{p} (g_l(x_l),\ \dot{g}_l(x_l))'$, $\left(\widehat{g}_l(x_l),\ \widehat{\dot{g}}_l(x_l)\right)'$ is close to $(g_l(x_l),\ \dot{g}_l(x_l))'$ for sufficiently large n and we only need to examine the minimization of $W_{n,x_l}(c_0,c_1)$ around $(g_l(x_l),\ h_3\dot{g}_l(x_l))'$. By the first order Taylor expansion, we have

$$\begin{aligned}
0 &= \left.\frac{\partial W_{n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(a,b)=\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)} \\
&= \left.\frac{\partial W_{n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(a,b)=(g_l(x_l),\ h_3\dot{g}_l(x_l))} \\
&+ \left.\frac{\partial^2 W_{n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'\,\partial (c_0,c_1)}\right|_{(c_0,c_1)=\left(g_l^*(x_l),\ h_3\dot{g}_l^*(x_l)\right)} \left[\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)' - (g_l(x_l),\ h_3\dot{g}_l(x_l))'\right],
\end{aligned}$$

where $(g_l^*(x_l),\ h_3\dot{g}_l^*(x_l))$ lies between $\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)$ and $(g_l(x_l),\ h_3\dot{g}_l(x_l))$.

By the Taylor expansions in (B.1.25) and (B.1.24), we have

$$\begin{aligned}
&\left.\frac{\partial W_{1n,x_l}(c_0,c_1)}{\partial (c_0,c_1)'}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))} \\
&= -\frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i1x_l}\left(1, \frac{1}{h_3}(X_{l,i1}-x_l)\right)(F_i - D_i) \\
&+ \frac{1}{2n}\sum_{i=1}^{N} L'(LF_i)\,\widehat{\dot{LF}}_i^2\,\ddot{g}_l(\bar{x}_{li})H_{3h_3,i1x_l}\left((X_{l,i1}-x_l)^2, \frac{1}{h_3}(X_{l,i1}-x_l)^3\right)
\end{aligned}$$

$$- \frac{1}{n}\sum_{i=1}^{N} L'\left(LF_i\right)\widehat{\dot{LF}}_i H_{3h_3,i1x_l}\left(1, \frac{1}{h_3}\left(X_{l,i1}-x_l\right)\right)\left\{\widehat{LF}_i - LF_i + \widehat{\dot{LF}}_i\left[\bar{g}_l\left(X_{l,i1}\right) - g_l\left(X_{l,i1}\right)\right]\right\}$$

$$- \frac{1}{2n}\sum_{i=1}^{N} L''\left(\bar{LF}_i\right)\widehat{\dot{LF}}_i H_{3h_3,i1x_l}\left(1, \frac{1}{h_3}\left(X_{l,i1}-x_l\right)\right)\delta^2_{x_l,i1},$$

and

$$\left.\frac{\partial W_{2n,x_l}\left(c_0,c_1\right)}{\partial\left(c_0,c_1\right)'}\right|_{(a,b)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$$

$$= \frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}\left(X_{l,i2}-x_l\right)\right)\left(F_i - D_i\right)$$

$$+ \frac{1}{2n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}\left(X_{l,i2}-x_l\right)\right)L'\left(LF_i\right)\widehat{\dot{LF}}_i\ddot{g}_l(\bar{x}_{li})\left(X_{l,i2}-x_l\right)^2$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}\left(X_{l,i2}-x_l\right)\right)L'\left(LF_i\right)\left\{\widehat{LF}_i - LF_i - \widehat{\dot{LF}}_i\left[\bar{g}_l\left(X_{l,i2}\right) - g_l\left(X_{l,i2}\right)\right]\right\}$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\widehat{\dot{LF}}_i H_{3h_3,i2x_l}\left(1, \frac{1}{h_3}\left(X_{l,i2}-x_l\right)\right) + \frac{1}{2}L''\left(\bar{LF}_i\right)\delta^2_{x_l,i2}.$$

Given the above results, we can show that

$$\sqrt{nh_3}\left(\frac{\partial W_{n,x_l}\left(g_l(x_l), h_3\dot{g}_l(x_l)\right)}{\partial\left(c_0,c_1\right)'} - E\left[L'\left(LF(\triangle g(X_i))\right)\dot{LF}^2(\triangle g(X))\middle| X_{l,it}=x_l\right]\ddot{g}_l(x_l)\begin{pmatrix}h_3^2\kappa_{21}\\ 0\end{pmatrix}\right)$$

$$\xrightarrow{d} N\left(0, \left\{E\left[\dot{LF}^2(\triangle g(X_i))F(\triangle g(X_i))\left[1-F(\triangle g(X_i))\right]\middle| X_{l,i1}=x_l\right]\right.\right.$$

$$\left.\left.+\ E\left[\dot{LF}^2(\triangle g(X_i))F(\triangle g(X_i))\left[1-F(\triangle g(X_i))\right]\middle| X_{l,i2}=x_l\right]\right\}\begin{pmatrix}\kappa_{02} & 0\\ 0 & \kappa_{22}\end{pmatrix}\right).$$

where $\frac{\partial W_{n,x_l}(g_l(x_l),\ h_3\dot{g}_l(x_l))}{\partial(c_0,c_1)'} = \left.\frac{\partial W_{n,x_l}(c_0,c_1)}{\partial(c_0,c_1)'}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$, and $\kappa_{ab} = \int u^a\left[H_3(u)\right]^b du$.

To derive the linear expression and asymptotic distribution of $\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)'$, we calculate the second order derivative:

$$\left.\frac{\partial^2 W_{n,x_l}\left(c_0,c_1\right)}{\partial\left(c_0,c_1\right)'\partial\left(c_0,c_1\right)}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$$

$$= \left.\frac{\partial^2 W_{1,nx_l}\left(c_0,c_1\right)}{\partial\left(c_0,c_1\right)'\partial\left(c_0,c_1\right)}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))} + \left.\frac{\partial^2 W_{2,nx_l}\left(c_0,c_1\right)}{\partial\left(c_0,c_1\right)'\partial\left(c_0,c_1\right)}\right|_{(c_0,c_1)=(g_l(x_l),\ h_3\dot{g}_l(x_l))}$$

$$= \frac{1}{n}\sum_{i=1}^{N}\left(\widehat{\dot{LF}}_i\right)^2 H_{3h_3,i1x_l}\mu_{x_l,i1}L'\left(\widehat{LF}_i - \widehat{\dot{LF}}_i\cdot\left[g_l(x_l) + h_3\dot{g}_l(x_l)\frac{1}{h_3}\left(X_{l,i1}-x_l\right) - \bar{g}_l\left(X_{l,i1}\right)\right]\right)$$

$$+ \frac{1}{n}\sum_{i=1}^{N}\left(\widehat{\dot{LF}}_i\right)^2 H_{3h_3,i2x_l}\mu_{x_l,i2}L'\left(\widehat{LF}_i + \widehat{\dot{LF}}_i\cdot\left[g_l(x_l) + h_3\dot{g}_l(x_l)\frac{1}{h_3}\left(X_{l,i2}-x_l\right) - \bar{g}_l\left(X_{l,i2}\right)\right]\right)$$

$$\xrightarrow{p} 2E\left[L'\left(LF(\triangle g(X))\right)\dot{LF}^2(\triangle g(X))\middle| X_l = x_l\right]\begin{pmatrix}1 & 0\\ 0 & \kappa_{21}\end{pmatrix}$$

$$= \left\{ E\left[L'\left(LF(\triangle g(X))\right) \dot{LF}^2(\triangle g(X)) \middle| X_{l,i1} = x_l \right] + E\left[L'\left(LF(\triangle g(X))\right) \dot{LF}^2(\triangle g(X)) \middle| X_{l,i2} = x_l \right] \right\}$$
$$\times \begin{pmatrix} 1 & 0 \\ 0 & \kappa_{21} \end{pmatrix}$$
$$= \left\{ E\left[\dot{LF}^2(\triangle g(X)) F(\triangle g(X_i)) \left[1 - F(\triangle g(X_i))\right] \middle| X_{l,i1} = x_l \right] \right.$$
$$\left. + E\left[\dot{LF}^2(\triangle g(X)) F(\triangle g(X_i)) \left[1 - F(\triangle g(X_i))\right] \middle| X_{l,i2} = x_l \right] \right\} \begin{pmatrix} 1 & 0 \\ 0 & \kappa_{21} \end{pmatrix},$$

where $\mu_{x_l,it} = \left(1, \frac{1}{h_3}\left(X_{l,it} - x_l\right)\right)' \left(1, \frac{1}{h_3}\left(X_{l,it} - x_l\right)\right)$ for $t = 1, 2$.

It follows that

$$\left(\widehat{g}_l(x_l),\ h_3\widehat{\dot{g}}_l(x_l)\right)' - \left(g_l(x_l),\ h_3\dot{g}_l(x_l)\right)'$$
$$= \left(\frac{\partial^2 W_{n,x_l}(c_0, c_1)}{\partial (c_0, c_1)' \partial (c_0, c_1)} \Bigg|_{(c_0,c_1) = \left(g_l^*(x_l),\ h_3\dot{g}_l^*(x_l)\right)'} \right)^{-1} \frac{\partial W_{n,x_l}(c_0, c_1)}{\partial (c_0, c_1)'} \Bigg|_{(c_0,c_1) = (g_l(x_l),\ h_3\dot{g}_l(x_l))'}$$
$$= \left(\frac{1}{n} \sum_{i=1}^{N} \left(\dot{LF}_i\right)^2 H_{3h_3,i1x_l} \mu_{x_l,it} L'\left(LF_i\right) \right)^{-1}$$
$$\times \left\{ -\frac{1}{n} \sum_{i=1}^{N} \dot{LF}_i H_{3h_3,i1x_l} \left(1, \frac{1}{h_3}\left(X_{l,i1} - x_l\right)\right) (F_i - D_i) \right.$$
$$+ \frac{1}{2n} \sum_{i=1}^{N} L'\left(LF_i\right) \dot{LF}_i^2 \ddot{g}_l(\bar{x}_{li}) H_{3h_3,i1x_l} \left(\left(X_{l,i1} - x_l\right)^2, \frac{1}{h_3}\left(X_{l,i1} - x_l\right)^3 \right)$$
$$+ \frac{1}{n} \sum_{i=1}^{N} \dot{LF}_i H_{3h_3,i2x_l} \left(1, \frac{1}{h_3}\left(X_{l,i2} - x_l\right)\right) (F_i - D_i)$$
$$\left. + \frac{1}{2n} \sum_{i=1}^{N} \dot{LF}_i H_{3h_3,i2x_l} \left(1, \frac{1}{h_3}\left(X_{l,i2} - x_l\right)\right) L'\left(LF_i\right) \dot{LF}_i \ddot{g}_l(\bar{x}_{li}) \left(X_{l,i2} - x_l\right)^2 \right\} + R_{3n},$$

where $\|R_{3n}\| = O_p\left(h_3^2 + \sqrt{h_3/n} + \eta_{3Kn}\right)$. Then

$$\sqrt{nh_3} \left(\begin{pmatrix} 1 & 0 \\ 0 & h_3 \end{pmatrix} \left(\begin{pmatrix} \widehat{g}_l(x_l) \\ \widehat{\dot{g}}_l(x_l) \end{pmatrix} - \begin{pmatrix} g_l(x_l) \\ \dot{g}_l(x_l) \end{pmatrix} \right) - \frac{1}{2}\ddot{g}_l(x_l) \begin{pmatrix} h_3^2 \kappa_{21} \\ 0 \end{pmatrix} \right)$$
$$\xrightarrow{d} N\left(0, \left\{ E\left[\dot{LF}^2(\triangle g(X)) F(\triangle g(X)) \left[1 - F(\triangle g(X))\right] \middle| X_{l,i1} = x_l \right] \right.\right.$$
$$\left.\left. + E\left[\dot{LF}^2(\triangle g(X)) F(\triangle g(X)) \left[1 - F(\triangle g(X))\right] \middle| X_{l,i2} = x_l \right] \right\}^{-1} \begin{pmatrix} \kappa_{02} & 0 \\ 0 & \frac{\kappa_{22}}{\kappa_{21}^2} \end{pmatrix} \right),$$

Noting that $LF(\cdot) = L^{-1}(F(\cdot))$, we have

$$\dot{LF}(\triangle g(X)) = \frac{\partial L^{-1}\left(F(\triangle g(X))\right)}{\partial x} = \frac{\partial \left[\ln\left(F(\triangle g(X))\right) - \ln\left(1 - F(\triangle g(X))\right)\right]}{\partial x}$$

$$= \frac{\dot{F}(\triangle g(X))}{F(\triangle g(X))} + \frac{\dot{F}(\triangle g(X))}{1 - F(\triangle g(X))} = \frac{\dot{F}(\triangle g(X))}{F(\triangle g(X))\left[1 - F(\triangle g(X))\right]}$$

and then we have the final asymptotic distribution of $\left(\widehat{g}_l(x_l)\ \widehat{\dot{g}}_l(x_l)\right)$,

$$\begin{pmatrix} \sqrt{nh_3} & 0 \\ 0 & \sqrt{nh_3^3} \end{pmatrix} \left[\begin{pmatrix} \widehat{g}_l(x_l) \\ \widehat{\dot{g}}_l(x_l) \end{pmatrix} - \begin{pmatrix} g_l(x_l) \\ \dot{g}_l(x_l) \end{pmatrix} - \frac{1}{2}\ddot{g}_l(x_l) \begin{pmatrix} h_3^2 \kappa_{21} \\ 0 \end{pmatrix} \right]$$

$$\xrightarrow{d} N\left(0, \left\{ E\left[\frac{\dot{F}^2(\triangle g(X_i))}{F(\triangle g(X_i))\left[1 - F(\triangle g(X_i))\right]} \middle| X_{l,i1} = x_l \right] + E\left[\frac{\dot{F}^2(\triangle g(X_i))}{F(\triangle g(X_i))\left[1 - F(\triangle g(X_i))\right]} \middle| X_{l,i2} = x_l \right] \right\}^{-1} \right.$$

$$\left. \times \begin{pmatrix} \kappa_{02} & 0 \\ 0 & \frac{\kappa_{22}}{\kappa_{21}^2} \end{pmatrix} \right).$$

This completes the proof of the theorem.

B.2 Technical Lemmas

In this appendix we state some technical lemmas that are used in the proofs of the main results and then prove them.

Recall that $m_i = E\left(D_i | X_i\right), m_{ji} = m_j - m_i,\ H_{1h_1,ji} = H_{1h_1}(m_j - m_i)$, and $f_m(\cdot)$ denotes the PDF of m_i. Let $\eta(m_i) = E\left[\triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} | m_i)\right]$ for $j \neq i$.

Lemma B.2.1 *Let $L_{0,n}$ and $L_{0,n1}$ be as defined in the proof of Theorem 3.3.1. Suppose that the conditions in Theorem 3.3.1 are satisfied. Then*

(i)$\|L_{0,n1} - E\left(L_{0,n1}\right)\| = O_p\left(\sqrt{K/n}\right), \|E\left(L_{0,n1}\right) - E\left[\eta(m_i) f_m\left(m_i\right)\right]\| = O\left(K h^{a_1}\right) = o\left(1\right),$ *and* $\lambda_{\min}\left(L_{0,n1}\right) \geq C_{1L}/2$ *w.p.a.1.*;

(ii) $\|L_{0,n\ell}\|_{op} = O_p\left(K^{1/2} h_1^{-1} \eta_{1Kn}\right) = o_p(1)$ *for* $\ell = 2, ..., 7$;

(iii) $\|L_{0,n} - E\left[\eta(m_i) f_m\left(m_i\right)\right]\|_{op} = o_p(1)$ *and* $\lambda_{\min}\left(L_{0,n}\right) \geq C_{1L}/2$ *w.p.a.1.*

Proof. (i) By the variance calculation and Chebyshev inequality, it is standard to show that

$$\left\| \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} \left\{ \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1,ji} - E\left[\triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1,ji} \right] \right\} \right\| = O_p\left(\sqrt{K/n}\right).$$

By Taylor expansions and the i.i.d. condition on $\{X_i\}$, for any $j \neq i$

$$\begin{aligned} E\left[\triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1}(m_j - m_i)\right] &= E\left\{\eta(m_i) H_{1h_1}(m_j - m_i)\right\} \\ &= E\left[\eta(m_i) \int \frac{1}{h_1} H_1\left(\frac{m - m_i}{h_1}\right) f_m\left(m\right) dm\right] \end{aligned}$$

$$= E\left[\eta(m_i)\int H_1(u) f_m\left(m_i + hu\right) du\right]$$
$$= E\left[\eta(m_i) f_m\left(m_i\right)\right] + O\left(h_1^{a_1}\right) E\left[\eta(m_i) f_m^{(a_1)}\left(m_i\right)\right],$$

Noting that $\left\|E\left[\eta(m_i) f_m^{(a_1)}\left(m_i\right)\right]\right\| = O\left(K\right)$, the second part of (i) follows. By the Weyl's inequality,

$$\begin{aligned}
\lambda_{\min}\left(L_{0,n1}\right) &\geq \lambda_{\min}\left(E\left(L_{0,n1}\right)\right) - \left\|L_{0,n1} - E\left(L_{0,n1}\right)\right\| \\
&\geq \lambda_{\min}\left(E\left(L_{0,n1}\right)\right) - O_p\left(\sqrt{K/n}\right) \\
&\geq \frac{1}{2}\lambda_{\min}\left(E\left[\eta(m_i) f_m\left(m_i\right)\right]\right) - O\left(K h^{a_1}\right) - O_p\left(\sqrt{K/n}\right) \\
&\geq C_{1L}/2 \text{ w.p.a.1.}
\end{aligned}$$

(ii) As in part (i), we can readily show that $\left\|L_{0,n\ell}\right\|_{\text{op}} = O_p\left(K^{1/2} h_1^{-1}\eta_{1Kn}\right) = o_p\left(1\right)$ for $\ell = 2, ..., 7$. For example, for $L_{0,n2}$, we have

$$\begin{aligned}
\left\|L_{0,n2}\right\|_{op} &= \left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} {h_1}^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\left(\bar{m}_j - m_j\right)\right\|_{op} \\
&\leq \max_{1\leq j\leq n}\left|\bar{m}_j - m_j\right| \frac{1}{n^2}\sum_{j=1}^{n}\left\|\sum_{i=1, i\neq j}^{n} \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} {h_1}^{-1}\dot{H}_{1h_1}\left(m_{ji}\right)\right\|_{op} \\
&= O_p\left(K^{1/2}\eta_{1Kn}\right) O_p\left(h_1^{-1}\right) = O_p\left(K^{1/2} h_1^{-1}\eta_{1Kn}\right) = o_p\left(1\right).
\end{aligned}$$

(iii) The result follows from (i)-(ii) and the Weyl's inequality. ■

Lemma B.2.2 *Let $L_{1,n}$, $L_{1,n1}$ and $L_{1,n2}$ be as defined in the proof of Theorem 3.3.1. Suppose that the conditions in Theorem 3.3.1 are satisfied. Then*

(i) $\left\|L_{1,n1}\right\| = O_p\left(\sqrt{K/n} + \sqrt{K} h_1^{a_1}\right)$;

(ii) $\left\|L_{1,n\ell}\right\| = O_p\left(\eta_{1Kn}\right)$ *for* $\ell = 2, 3$;

(iii) $\left\|L_{1,n\ell}\right\| = O_p\left(\eta_{1Kn}\right)$ *for* $\ell = 4, 5, 6, 7$;

(iv) $\left\|L_{1,n}\right\| = O_p\left(\eta_{1Kn} + \sqrt{K} h_1^{a_1}\right)$.

Proof. (i) Note that $m_i = E\left(D_i|X_i\right) = F\left(\triangle g_i\right)$ under Assumption 2. First, notice that

$$L_{1,n1} = \frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left(\triangle g_i - \triangle g_j\right)\triangle P_{i,j}^{K-1,K} H_{1h_1}\left(m_j - m_i\right) = \varphi_n U_{1n},$$

where

$$U_{1n} = \binom{n}{2}^{-1}\sum_{1\leq i\neq j\leq n}\left(\triangle g_i - \triangle g_j\right)\triangle P_{i,j}^{K-1,K} H_{1h_1}\left(m_j - m_i\right) = \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n} q_{1n}(X_i, X_j),$$

$\varphi_n = \binom{n}{2}/n^2 \to 1/2$ as $n \to \infty$, and $q_{1n}(X_i, X_j) = \triangle g_{ij} \triangle P_{i,j}^{K-1,K} H_{1h_1}(m_{ji})$. Note that $q_{1n}(\cdot,\cdot)$ is symmetric in its two arguments. Let

$$r_{1n}(X_j) = E\left[q_{1n}(X_i, X_j)|X_j\right] \text{ for } j \neq i, \text{ and } \theta_{1n} = E\left[r_{1n}(X_j)\right],$$

By the Hoeffding decomposition (see, Theorem 1 in Section 1.6 of Lee (1990), we have $U_{1n} = \theta_{1n} + \mathbb{U}_{1n}^{(1)} + \mathbb{U}_{1n}^{(2)}$, where

$$\begin{aligned}
\mathbb{U}_{1n}^{(1)} &= \frac{1}{n}\sum_{i=1}^{n}\left[r_{1n}(X_j) - \theta_{1n}\right], \\
\mathbb{U}_{1n}^{(2)} &= \binom{n}{2}^{-1}\sum_{1\leq i<j\leq n}\left[q_{1n}(X_i, X_j) - r_{1n}(X_i) - r_{1n}(X_j) + \theta_{1n}\right].
\end{aligned}$$

Note that for $j \neq i$,

$$E\left\|q_{1n}(X_i, X_j)\right\|^2 = E\ \left[\operatorname{tr}\left\{(\triangle g_{i,j})^2 \triangle P_{i,j}^{K-1,K} \triangle P_{i,j}^{K-1,K\prime} H_{1h_1}^2(m_{ji})\right\}\right] = O(Kh^{-1}) = o(n).$$

Then by Lemma 3.1 in Powell et al. (1989), $E\left\|\mathbb{U}_n^{(2)}\right\|^2 = o(n^{-1})$ and thus $\mathbb{U}_n^{(2)} = o_p(n^{-1/2})$. It remains to study θ_n and $\mathbb{U}_n^{(1)}$.

Let $\rho_P(m_i) = E\left[\triangle P_i^{K-1,K}|m_i\right]$, $\rho_g(m_i) = E\left[\triangle g_i|m_i\right]$ and $\rho(m_i) = E\left[\triangle g_i \triangle P_i^{K-1,K}|m_i\right]$ Note that $r_{1n}(X_j) = E\left[(\triangle g_i - \triangle g_j)\triangle P_{i,j}^{K-1,K} H_{1h_1}(m_{ji})|X_j\right] = E\left[\triangle g_i \triangle P_{i,j}^{K-1,K} H_{1h_1}(m_{ji})|X_j\right] - \triangle g_j E\left[\triangle P_{i,j}^{K-1,K} H_{1h_1}(m_{ji})|X_j\right]$ $r_{1n,1}(X_j) - r_{1n,2}(X_j)$. By straightforward moment calculations and the independence of $\{X_i\}$, we have

$$\begin{aligned}
r_{1n,1}(X_j) &= E\left\{\triangle g_i\left[\triangle P_j^{K-1,K} - \triangle P_i^{K-1,K}\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle P_j^{K-1,K} E\left[E\left(\triangle g_i|m_i, X_j\right) H_{1h_1}(m_{ji})|X_j\right] \\
&- E\left\{E\left[\triangle g_i \triangle P_i^{K-1,K}(X_i)|m_i, X_j\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle P_j^{K-1,K} E\left[\rho_g(m_i) H_{1h_1}(m_{ji})|X_j\right] - E\left\{\rho(m_i) H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle P_j^{K-1,K}\int \rho_g(m)\frac{1}{h_1}H_1\left(\frac{m-m_j}{h_1}\right) f_m(m)\,dm - \int \rho(m)\frac{1}{h_1}H_1\left(\frac{m-m_j}{h_1}\right) f_m(m)\,dm \\
&= \triangle P_j^{K-1,K}\left[\rho_g(m_j) f_m(m_j) + \frac{h_1^{a_1}}{a_1!}\,\partial^{(a_1)}\left[\rho_g(m) f_m(m)\right]\Big|_{m=m_j} + o(h_1^{a_1})\right] \\
&- \left[\rho(m_j) f_m(m_j) + \frac{h_1^{a_1}}{a_1!}\,\partial^{(a_1)}\left[\rho(m) f_m(m)\right]\Big|_{m=m_j} + o(h_1^{a_1})\right].
\end{aligned}$$

and

$$\begin{aligned}
r_{1n,2}(X_j) &= \triangle g_j E\left\{\left[\triangle P_j^{K-1,K} - \triangle P_i^{K-1,K}\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle g_j \triangle P_j^{K-1,K} E\left[H_{1h_1}(m_{ji})|X_j\right] - \triangle g_j E\left\{E\left[\triangle P_i^{K-1,K}|m_i, X_j\right] H_{1h_1}(m_{ji})|X_j\right\} \\
&= \triangle g_j \triangle P_j^{K-1,K} E\left[H_{1h_1}(m_{ji})|X_j\right] - \triangle g_j E\left\{\rho_P(m_i) H_{1h_1}(m_{ji})|X_j\right\}
\end{aligned}$$

$$= \triangle g_j \triangle P_j^{K-1,K} \int \frac{1}{h_1} H_1 \left(\frac{m - m_j}{h_1} \right) f_m (m)\, dm - \triangle g_j \int \rho_P (m) \frac{1}{h_1} H_1 \left(\frac{m - m_j}{h_1} \right) f_m (m)\, dm$$
$$= \triangle g_j \triangle P_j^{K-1,K} \left[f_m (m_j) + \frac{h_1^{a_1}}{a_1!} f_m^{(a_1)} (m_j) + o(h_1^{a_1}) \right]$$
$$- \triangle g_j \left[\rho (m_j) f_m (m_j) + \frac{h_1^{a_1}}{a_1!} \partial^{(a_1)} \left[\rho_P (m) f_m (m) \right] \Big|_{m=m_j} + o(h_1^{a_1}) \right].$$

Then it is easy to show that

$$\|\theta_{1n}\| = \| E \left[r_{1n,1} (X_j) \right] - E \left[r_{1n,2} (X_j) \right] \| = O \left(\sqrt{K} h_1^{a_1} \right)$$

where we use the fact that

$$\begin{aligned}
& E \left\{ \triangle g_j \triangle P_j^{K-1,K} f_m (m_j) - \triangle g_j \left[\rho (m_j) f_m (m_j) \right] \right\} \\
& -E \left\{ \triangle P_j^{K-1,K} \rho_g (m_j) f_m (m_j) - \rho (m_j) f_m (m_j) \right\} \\
= & \left\{ E \left[\triangle g_j \triangle P_j^{K-1,K} f_m (m_j) \right] - E \left[\rho (m_j) f_m (m_j) \right] \right\} \\
& +E \left\{ \triangle P_j^{K-1,K} \rho_g (m_j) f_m (m_j) - \triangle g_j \left[\rho (m_j) f_m (m_j) \right] \right\} \\
= & \left\{ E \left[E \left(\triangle g_j \triangle P_j^{K-1,K} | m_j \right) f_m (m_j) \right] - E \left[\rho (m_j) f_m (m_j) \right] \right\} \\
& + \left[\left\{ E \left(\triangle P_j^{K-1,K} | m_j \right) \rho_g (m_j) f_m (m_j) \right\} - E \left\{ E \left(\triangle g_j | m_j \right) \left[\rho (m_j) f_m (m_j) \right] \right\} \right] \\
= & \ 0 + 0 = 0
\end{aligned}$$

by the repeated use of the law of iterated expectations. In addition, we can readily show that $E \| r_{1n} (X_j) \|^2 = O(K)$ and $E \left\| \mathbb{U}_{1n}^{(1)} \right\|^2 = O(K/n)$. Then $\left\| \mathbb{U}_{1n}^{(1)} \right\| = O_p(\sqrt{K/n})$. Consequently, we have

$$\begin{aligned}
\|U_{1n}\| & \leq \|\theta_{1n}\| + 2 \left\| \mathbb{U}_{1n}^{(1)} \right\| + \left\| \mathbb{U}_{1n}^{(2)} \right\| \\
& = O_p \left(\sqrt{K} h_1^{a_1} \right) + O_p \left(\sqrt{K/n} \right) + o_p(n^{-1/2}) = O_p \left(\sqrt{K} h_1^{a_1} + \sqrt{K/n} \right).
\end{aligned}$$

Then the result in (i) follows.

(ii) Recall that $L_{1,n2} = \frac{2}{n^2} \sum_{1 \leq i \neq j \leq n} \triangle g_j \triangle P_{i,j}^{K-1,K} h_1^{-1} \dot{H}_{1h_1} (m_{ji}) \left[\bar{m}_j - m_j \right]$, where $\bar{m}_j = L \left(R(X_j)' \bar{\pi} \right)$. It is easy to see that

$$\begin{aligned}
L_{1,n2} & = \frac{2}{n^2} \sum_{1 \leq i \neq j \leq n} \triangle g_j \triangle P_{i,j}^{K-1,K} h_1^{-1} \dot{H}_{1h_1} (m_{ji}) \left[\bar{m}_j - m_j \right] \\
& = \frac{2}{n^2} \sum_{1 \leq i < j \leq n} q_{2n} \left((X_i, X_j), \bar{\pi} \right) = 2 \varphi_n U_{2n} (\bar{\pi})
\end{aligned}$$

where

$$q_{2n} \left((X_i, X_j), \pi \right) = h_1^{-1} \triangle g_j \triangle P_{i,j}^{K-1,K} \dot{H}_{1h_1} (m_{ji}) \left[L \left(R(X_j)' \pi \right) - m_j \right]$$

$$
\begin{aligned}
& +h_1^{-1}\triangle g_i \triangle P_{j,i}^{K-1,K}\dot{H}_{1h_1}(m_{ij})\left[L\left(R(X_i)'\pi\right)-m_i\right] \\
&= \left\{\triangle g_j\left[L\left(R(X_j)'\pi\right)-m_j\right]+\triangle g_i\left[L\left(R(X_i)'\pi\right)-m_i\right]\right\}h_1^{-1}\triangle P_{i,j}^{K-1,K}\dot{H}_{1h_1}(m_{ij}), \\
U_{2n}(\pi) &= \frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n} q_n\left(\left(X_i,X_j\right),\pi\right),
\end{aligned}
$$

and $\varphi_n = \binom{n}{2}/n^2 \to 1$ as $n\to\infty$. Here we use the fact that $\dot{H}_{1h_1}(m) = -\dot{H}_{1h_1}(-m)$ by the symmetry of H_1 and $\triangle P_{i,j}^{K-1,K} = -\triangle P_{j,i}^{K-1,K}$. By construction, $q_{2n}\left(\left(X_i,X_j\right),\pi\right)$ is symmetric in $\left(X_i,X_j\right)$.

It suffices to determine the probability order of $U_{2n}(\bar{\pi})$ by studying the U-process $\{U_n(\pi)\}$. Let $r_{2n}(X_j,\pi) = E\left[q_{2n}\left(\left(X_i,X_j\right),\pi\right)|X_j\right]$ and $\theta_{2n}(\pi) = E\left[r_{2n}(X_j,\pi)\right]$. Then we have the following Hoeffding decomposition:

$$
U_{2n}(\pi) = \theta_{2n}(\pi) + 2\mathbb{U}_{2n}^{(1)}(\pi) + \mathbb{U}_{2n}^{(2)}(\pi),
$$

where

$$
\begin{aligned}
\mathbb{U}_{2n}^{(1)}(\pi) &= \frac{1}{n}\sum_{i=1}^{n}\left[r_{2n}(X_j,\pi)-\theta_{2n}(\pi)\right] \\
\mathbb{U}_{2n}^{(2)}(\pi) &= \frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}\left[q_{2n}\left(\left(X_i,X_j\right),\pi\right)-r_{2n}(X_j,\pi)-r_{2n}(X_j,\pi)+\theta_{2n}(\pi)\right].
\end{aligned}
$$

Let $\delta_j \equiv \delta_{j,\pi} \equiv L\left(R(X_j)'\pi\right)-m_j$, where we frequently suppress the dependence of δ_j on π. Note that

$$
\begin{aligned}
r_{2n}(X_j,\pi) &= h_1^{-1}E\left[\left\{\triangle g_j\left[L\left(R(X_j)'\pi\right)-m_j\right]+\triangle g_i\left[L\left(R(X_i)'\pi\right)-m_i\right]\right\}\triangle P_{i,j}^{K-1,K}\dot{H}_{1h_1}(m_{ij})|X_j\right] \\
&= h_1^{-1}\triangle g_j\delta_j E\left[\triangle P_{i,j}^{K-1,K}\dot{H}_{1h_1}(m_{ij})|X_j\right] + h_1^{-1}E\left[\triangle g_i\delta_i\triangle P_{i,j}^{K-1,K}\dot{H}_{1h_1}(m_{ij})|X_j\right] \\
&\equiv r_{2n,1}(X_j,\pi)+r_{2n,2}(X_j,\pi).
\end{aligned}
$$

Note that

$$
\begin{aligned}
r_{2n,1}(X_j,\pi) &= h_1^{-1}\triangle g_j\delta_j E\left[\triangle P_i^{K-1,K}\dot{H}_{1h_1}(m_{ij})|X_j\right] - h_1^{-1}\triangle g_j\delta_j\triangle P_j^{K-1,K}E\left[\dot{H}_{1h_1}(m_{ij})|X_j\right] \\
&= h_1^{-1}\triangle g_j\delta_j E\left[\rho_P(m_i)\dot{H}_{1h_1}(m_{ij})|X_j\right] - h_1^{-1}\triangle g_j\delta_j\triangle P_j^{K-1,K}E\left[\dot{H}_{1h_1}(m_{ij})|X_j\right] \\
&= h_1^{-1}\triangle g_j\delta_j\int\dot{H}_1(u)\rho_P(m_j+hu)f_m(m_j+hu)\,du \\
&\quad -h_1^{-1}\triangle g_j\delta_j\triangle P_j^{K-1,K}\int\dot{H}_1(u)f_m(m_j+hu)\,du \\
&= -\triangle g_j\delta_j\partial\left[\rho_P(m_j)f_m(m_j)\right]+\triangle g_j\delta_j\triangle P_j^{K-1,K}\partial f_m(m_j)\,du + r_{2n1,a}(X_j,\pi)
\end{aligned}
$$

and

$$
\begin{aligned}
r_{2n,2}(X_j,\pi) &= h_1^{-1}E\left[\triangle g_i\delta_i\triangle P_i^{K-1,K}\dot{H}_{1h_1}(m_{ij})|X_j\right] - h_1^{-1}\triangle P_j^{K-1,K}E\left[\triangle g_i\delta_i\dot{H}_{1h_1}(m_{ij})|X_j\right] \\
&= h_1^{-1}E\left[\rho_\delta(m_i)\dot{H}_{1h_1}(m_{ij})|X_j\right] - h_1^{-1}\triangle P_j^{K-1,K}E\left[\rho_{\delta g}(m_i)\dot{H}_{1h_1}(m_{ij})|X_j\right]
\end{aligned}
$$

$$
\begin{aligned}
&= h_1^{-1}\int \dot{H}_1(u)\,\rho_\delta(m_j+hu)f_m(m_j+hu)\,du \\
&\quad -h_1^{-1}\triangle P_j^{K-1,K}\int \dot{H}_1(u)\,\rho_{\delta g}(m_j+hu)\,f_m(m_j+hu)\,du \\
&= -\partial\left[\rho_\delta(m_j)f_m(m_j)\right]+\triangle P_j^{K-1,K}\partial\left[\rho_{\delta g}(m_j)\,f_m(m_j)\right]+r_{2n2,a}(X_j,\pi),
\end{aligned}
$$

where $\rho_\delta(m_i)=E\left(\triangle g_i\delta_i\triangle P_i^{K-1,K}|m_i\right)$, $\rho_{\delta g}(m_i)=E(\triangle g_i\delta_i|m_i)$, $r_{2n1,a}(X_j,\pi)$ and $r_{2n2,a}(X_j,\pi)$ denote the remainder terms in the first order Taylor expansions, we use the fact that $\int \dot{H}_1(u)\,du=0$ and $\int \dot{H}_1(u)\,udu=-1$. Note that

$$
\begin{aligned}
&[\rho_{\delta g}(m_j)\,\rho_P(m_j+hu)-\rho_P(m_j)\,\rho_{\delta g}(m_j+hu)-\rho_\delta(m_j)+\rho_\delta(m_j+hu)]f_m(m_j+hu) \\
= \;& hu[\rho_{\delta g}(m_j)\,\rho_P^{(1)}(m_j)-\rho_P(m_j)\,\rho_{\delta g}^{(1)}(m_j)+\rho_\delta^{(1)}(m_j)]f_m(m_j+hu) \\
&+\frac{1}{2}h_1^2u^2[\rho_{\delta g}(m_j)\,\rho_P^{(2)}(m_j^*)-\rho_P(m_j)\,\rho_{\delta g}^{(2)}\left(m_j^*\right)+\rho_\delta^{(2)}(m_j^*)]f_m(m_j+hu) \\
\equiv \;& hu\psi_{1j}f_m(m_j+hu)+\frac{1}{2}h_1^2u\psi_{2j}f_m(m_j+hu)
\end{aligned}
$$

where m_j^* lies between m_j and m_j+hu, $\psi_{1j}=\rho_{\delta g}(m_j)\,\rho_P^{(1)}(m_j)-\rho_P(m_j)\,\rho_{\delta g}^{(1)}(m_j)+\rho_\delta^{(1)}(m_j)$ and $\psi_{2j}=\rho_{\delta g}(m_j)\,\rho_P^{(2)}(m_j^*)\;-\rho_P(m_j)\,\rho_{\delta g}^{(2)}\left(m_j^*\right)+\rho_\delta^{(2)}(m_j^*)$. Then

$$
\begin{aligned}
\theta_{2n}(\pi) &= E\left[r_{2n,1}(X_j,\pi)+r_{2n,2}(X_j,\pi)\right] \\
&= h_1^{-1}\int \dot{H}_1(u)\,E\{[\rho_{\delta g}(m_j)\,\rho_P(m_j+hu)-\rho_P(m_j)\,\rho_{\delta g}(m_j+hu) \\
&\quad -\rho_\delta(m_j)+\rho_\delta(m_j+hu)]\}f_m(m_j+hu)\,du \\
&= \int \dot{H}_1(u)\,uE\left[f_m(m_j+hu)\,\psi_{1j}\right]du+\frac{h_1}{2}\int \dot{H}_1(u)\,u^2E\left[f_m(m_j+hu)\,\psi_{2j}\right]du \\
&\equiv \theta_{2n,1}(\pi)+\theta_{2n,2}(\pi).
\end{aligned}
$$

Noting that $m_j=E(D_j|X_j)=F(\triangle g_j)=L(LF(\triangle g_j))$, we have by Taylor expansions,

$$
\begin{aligned}
\delta_j &= L\left(R(X_j)'\pi\right)-E(D_j|X_j) \\
&= L\left(R(X_j)'\pi\right)-L\left(LF(\triangle g_j)\right) \\
&= \dot{L}\left(LF(\triangle g_j)\right)\left[R(X_j)'\pi-LF(\triangle g_j)\right]+\frac{1}{2}\ddot{L}\left(LF(\triangle\bar{\bar{g}}_j)\right)\left[R(X_j)'\pi-LF(\triangle g_j)\right]^2 \\
&= \dot{L}\left(LF(\triangle g_j)\right)R(X_j)'(\pi-\pi^0)+\dot{L}\left(LF(\triangle g_j)\right)\left[R(X_j)'\pi^0-LF(\triangle g_j)\right] \\
&\quad +\frac{1}{2}\ddot{L}\left(LF(\triangle\bar{\bar{g}}_j)\right)\left\{R(X_j)'\left(\pi-\pi^0\right)+\left[R(X_j)'\pi^0-LF(\triangle g_j)\right]\right\}^2 \\
&\equiv \dot{L}\left(LF(\triangle g_j)\right)R(X_j)'\left(\pi-\pi^0\right)+\dot{L}\left(LF(\triangle g_j)\right)\left[R(X_j)'\pi^0-LF(\triangle g_j)\right]+\delta_{j,2},
\end{aligned}
$$

where $\triangle\bar{\bar{g}}_j$ is between $R(X_j)'\pi$ and $LF(\triangle g_j)$. Note that

$$
\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}}\max_{1\le j\le n}|\delta_{j,2}| \lesssim \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}}\left\|R(X_j)'\left(\pi-\pi^0\right)\right\|^2
$$

$$
\begin{aligned}
& + \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left\|R(X_j)'\pi^0 - LF(\triangle g_j)\right\|^2 \\
\le \ & K\left\|\pi-\pi^0\right\|^2 + O\left(K^{-2\gamma}\right) = O\left(K\eta_{1Kn}^2\right),
\end{aligned}
$$

$$
\begin{aligned}
\max_{1\le j\le n}\left|R(X_j)'\pi^0 - LF(\triangle g_j)\right| &= O\left(K^{-\gamma}\right), \text{ and} \\
\sup_{\|\pi-\pi^0\|\le C\eta_{Kn}} \max_{1\le j\le n}\left|LF(\triangle \bar{\bar{g}}_j) - LF(\triangle g_j)\right| &\le O\left(\sqrt{K}\eta_{1Kn}\right).
\end{aligned}
$$

In addition,

$$
\begin{aligned}
\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} E\left(\delta_j^2\right) &\lesssim \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} E\left|R(X_j)'\left(\pi-\pi^0\right)\right|^2 + E\left[R(X_j)'\pi^0 - LF(\triangle g_j)\right]^2 \\
&\quad + \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \max_{1\le j\le n}\left|\delta_{j,2}\right|^2 \\
&= O\left(\eta_{1Kn}^2\right) + O\left(K^{-2\gamma_F}\right) + O\left(K^2\eta_{1Kn}^4\right) = O\left(\eta_{1Kn}^2\right).
\end{aligned}
$$

By these results and the uniform boundedness of $\dot{L}$ and $\ddot{L}$, we can readily show that for $l = 1, 2$,

$$
\begin{aligned}
E\left[f_m(m_j)\psi_{1j}\right] &= E\left\{\left[\rho_{\delta g}(m_j)\rho_P^{(1)}(m_j) - \rho_P(m_j)\rho_{\delta g}^{(1)}(m_j) + \rho_\delta^{(1)}(m_j)\right] f_m(m_j)\right\} = O\left(\eta_{1Kn}\right), \\
\left|E\left[f_m^{(l)}\left(m_j^*\right)\psi_{1j}\right]\right| &\lesssim E\left\{\left|\rho_{\delta g}(m_j)\rho_P^{(1)}(m_j) - \rho_P(m_j)\rho_{\delta g}^{(1)}(m_j) + \rho_\delta^{(1)}(m_j)\right|\right\} = O\left(K^{1/2}\eta_{1Kn} + K\eta_{1Kn}^2\right),
\end{aligned}
$$

uniformly in π with $\left\|\pi-\pi^0\right\| \le C\eta_{1Kn}$. Then

$$
\begin{aligned}
& \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left|\theta_{2n,1}(\pi)\right| \\
= \ & \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left|\int \dot{H}_1(u)\,uE\left[f_m(m_j+hu)\psi_{1j}\right]du\right| \\
= \ & \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left|-E\left[f_m(m_j)\psi_{1j}\right] + h_1\int \dot{H}_1(u)\,u^2E\left[f_m^{(1)}(m_j)\psi_{1j}\right]du\right. \\
& \left. + \frac{h_1^2}{2}\int \dot{H}_1(u)\,u^3du E\left[f_m^{(2)}\left(m_j^*\right)\psi_{1j}\right]\right| \\
\lesssim \ & \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left|E\left[f_m(m_j)\psi_{1j}\right]\right| + \frac{h_1^2}{2}\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left|E\left[f_m^{(2)}\left(m_j^*\right)\psi_{1j}\right]\right| \\
= \ & O\left(\eta_{1Kn}\right) + h_1^2 O\left(K^{1/2}\eta_{1Kn} + K\eta_{1Kn}^2\right) = O\left(\eta_{1Kn}\right),
\end{aligned}
$$

where the second and third equalities hold by the second order Taylor expansions and the fact that $\int \dot{H}_1(u)\,udu = -1$ and $\int \dot{H}_1(u)\,u^2du = 0$ by the symmetry of $H_1(\cdot)$. Analogously, we have uniformly in π with $\left\|\pi-\pi^0\right\| \le C\eta_{1Kn}$,

$$
\begin{aligned}
\sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left|\theta_{2n,2}(\pi)\right| &= \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \left|\frac{h_1}{2}\int \dot{H}_1(u)\,u^2E\left[f_m(m_j+hu)\psi_{2j}\right]du\right| \\
&\le \sup_{\|\pi-\pi^0\|\le C\eta_{1Kn}} \frac{h_1^2}{2}\left|\int \dot{H}_1(u)\,u^3du E\left[f_m^{(1)}\left(m_j^*\right)\psi_{2j}\right]\right| \\
&= h_1^2 O\left(K^{1/2}\eta_{1Kn} + K\eta_{1Kn}^2\right) = O\left(\eta_{1Kn}\right).
\end{aligned}
$$

It follows that

$$\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \|\theta_{2n}(\pi)\| = O(\eta_{1Kn}).$$

Similarly, we can show that

$$E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|r_{2n}\left(X_1,\pi^0\right)\right\|^2\right] \lesssim K^2\eta_{1Kn}^2 = o(K)$$

and

$$\begin{aligned}
& E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|q_{2n}\left(\left(X_1,X_2\right),\pi^0\right)\right\|^2\right] \\
\lesssim\ & E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\triangle g_1\left[L\left(R(X_1)'\pi^0\right)-m_1\right]h_1^{-1}\triangle P_{1,2}^{K-1,K}\dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
\lesssim\ & E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\left\{\triangle g_1\left[L\left(R(X_1)'\pi^0\right)+L\left(R(X_1)'\pi^0\right)\right]\right\}h_1^{-1}\triangle P_{1,2}^{K-1,K}\dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
& +E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\left\{\triangle g_1\left[L\left(R(X_1)'\pi^0\right)-m_1\right]\right\}h_1^{-1}\triangle P_{1,2}^{K-1,K}\dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
\lesssim\ & E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\left\{\triangle g_1 R(X_1)'\left(\pi^0-\pi^0\right)\right\}h_1^{-1}\triangle P_{1,2}^{K-1,K}\dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
& +E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\left\{\triangle g_1\left[L\left(R(X_1)'\pi^0\right)-m_1\right]\right\}h_1^{-1}\triangle P_{1,2}^{K-1,K}\dot{H}_{1h_1}\left(m_{1,2}\right)\right\|^2\right] \\
\lesssim\ & {h_1}^{-3}K\eta_{1Kn}^2.
\end{aligned}$$

Then by Corollary 5.3 in **?**, we have

$$\sup_{\|\pi-\pi^0\|\leq C\eta_{Kn}} \left\|\mathbb{U}_{2n}^{(2)}(\pi)\right\| \lesssim n^{-1}\left({h_1}^{-3}K\eta_{1Kn}^2\right)^{1/2}\ln(n) = o_p(\eta_{1Kn}).$$

By the empirical process theory, we have

$$\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\mathbb{U}_{2n}^{(1)}(\pi)\right\| = O_p\left(n^{-1/2}K^{1/2}\ln(n)\right) = O_p(\eta_{1Kn}).$$

Consequently, we have

$$\|U_n(\bar{\pi})\| \lesssim \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}} \left\|\theta_{2n}(\pi)+2\mathbb{U}_{2n}^{(1)}(\pi)+\mathbb{U}_{2n}^{(2)}(\pi)\right\| = O_p(\eta_{1Kn})$$

and $\|L_{1,n2}\| = O_p(\eta_{1Kn})$. Analogously, we can show that $\|L_{1,n3}\| = O_p(\eta_{1Kn})$.

(iii) It suffices to obtain the rough probability bound for $\|L_{1,n\ell}\|$ with $\ell = 4,5,6,7$. For example,

$$\|L_{1,n5}\| \lesssim \left\|h_1^{-2}\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\triangle g_j\triangle P_{i,j}^{K-1,K}\ddot{H}_{1h_1,ji}\left|L\left(R(X_j)'\bar{\pi}\right)-E\left(D_j|X_j\right)\right|^2\right\|$$

$$
\begin{aligned}
&\lesssim h_1^{-2}\frac{1}{n^2}\sum_{1\le i\ne j\le n}\left\|\triangle g_j \triangle P_{i,j}^{K-1,K}\ddot{H}_{1h_1,ji}\right\|\left|\dot{L}\left(LF\left(\triangle g_j\right)\right)R(X_j)'\left(\bar{\pi}-\pi^0\right)\right|^2 \\
&+h_1^{-2}\frac{1}{n^2}\sum_{1\le i\ne j\le n}\left\|\triangle g_j \triangle P_{i,j}^{K-1,K}\ddot{H}_{1h_1,ji}\right\|\left|\left[R(X_j)'\pi^0-LF\left(\triangle g_j\right)\right]\right|^2 \\
&+h_1^{-2}\frac{1}{n^2}\sum_{1\le i\ne j\le n}\left\|\triangle g_j \triangle P_{i,j}^{K-1,K}\ddot{H}_{1h_1,ji}\right\|\left|\delta_{j,2}\right|^2 \\
&=h_1^{-2}O_p\left(K^{1/2}\eta_{1Kn}^2\right)+h_1^{-2}O_p\left(K^{-2\gamma}\right)+h_1^{-2}O_p\left(K^2\eta_{1Kn}^4\right)=O_p\left(\eta_{1Kn}\right).
\end{aligned}
$$

Similarly, $\|L_{1,n\ell}\| = O_p\left(\eta_{1Kn}\right)$ for $\ell = 5, 6, 7$. Alternatively, we can use the arguments as used in (ii).

Note that $L_{1,n4} = \frac{1}{2n^2}\sum_{1\le i\ne j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left[\bar{m}_i-m_i\right]\left[\bar{m}_j-m_j\right]$. It is easy to see that

$$
\begin{aligned}
L_{1,n4} &= \frac{1}{2n^2}\sum_{1\le i\ne j\le n}\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left[\bar{m}_i-m_i\right]\left[\bar{m}_j-m_j\right] \\
&= \frac{1}{2n^2}\sum_{1\le i<j\le n}q_{3n}\left(\left(X_i,X_j\right),\bar{\pi}\right)=\frac{1}{2}\varphi_n U_{3n}\left(\bar{\pi}\right)
\end{aligned}
$$

where

$$
\begin{aligned}
q_{3n}\left(\left(X_i,X_j\right),\pi\right) &= \triangle g_{i,j}\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{h_1}\left(m_{ji}\right)\left[L\left(R(X_i)'\pi\right)-m_i\right]\left[L\left(R(X_j)'\pi\right)-m_j\right] \\
U_{3n}\left(\pi\right) &= \frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}q_{3n}\left(\left(X_i,X_j\right),\pi\right),
\end{aligned}
$$

and $\varphi_n = \binom{n}{2}/n^2 \to 1$ as $n\to\infty$. It suffices to determine the probability order of $U_{3n}\left(\bar{\pi}\right)$ by studying the U-process $\left\{U_{3n}\left(\pi\right)\right\}$. Let $r_{3n}(X_j,\pi) = E\left[q_{3n}\left(\left(X_i,X_j\right),\pi\right)|X_j\right]$ and $\theta_{3n}\left(\pi\right) = E\left[r_{3n}(X_j,\pi)\right]$. Then we have the following Hoeffding decomposition:

$$
U_{3n}\left(\pi\right)=\theta_{3n}\left(\pi\right)+2\mathbb{U}_{3n}^{(1)}\left(\pi\right)+\mathbb{U}_{3n}^{(2)}\left(\pi\right),
$$

where

$$
\begin{aligned}
\mathbb{U}_{3n}^{(1)}\left(\pi\right) &= \frac{1}{n}\sum_{i=1}^{n}\left[r_{3n}(X_j,\pi)-\theta_{3n}\left(\pi\right)\right] \\
\mathbb{U}_{3n}^{(2)}\left(\pi\right) &= \frac{1}{\binom{n}{2}}\sum_{1\le i<j\le n}\left[q_{3n}\left(\left(X_i,X_j\right),\pi\right)-r_{3n}(X_j,\pi)-r_{3n}(X_j,\pi)+\theta_{3n}\left(\pi\right)\right].
\end{aligned}
$$

Note that

$$
\begin{aligned}
r_{3n}(X_j,\pi) &= E\left\{\triangle g_{i,j}\triangle P_{i,j}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}\left(m_{ji}\right)\left[L\left(R(X_i)'\pi\right)-m_i\right]\left[L\left(R(X_j)'\pi\right)-m_j\right]|X_j\right\} \\
&= h_1^{-2}\delta_j E\left[\triangle g_{i,j}\delta_i\triangle P_{i,j}^{K-1,K}\ddot{H}_{1h_1}\left(m_{ji}\right)|X_j\right] \\
&= h_1^{-2}\delta_j E\left[\triangle g_i\delta_i\triangle P_{i,j}^{K-1,K}\ddot{H}_{1h_1}\left(m_{ji}\right)|X_j\right]-h_1^{-2}\delta_j\triangle g_j E\left[\delta_i\triangle P_{i,j}^{K-1,K}\ddot{H}_{1h_1}\left(m_{ji}\right)|X_j\right]
\end{aligned}
$$

$$\equiv r_{3n,1}(X_j,\pi)-r_{3n,2}(X_j,\pi).$$

Note that

$$\begin{aligned}
r_{3n,1}(X_j,\pi) &= h_1^{-2}\delta_j E\left[\triangle g_i\delta_i\triangle P_i^{K-1,K}\ddot{H}_{1h_1}(m_{ji})|X_j\right]-h_1^{-2}\delta_j\triangle P_j^{K-1,K}E\left[\triangle g_i\delta_i\ddot{H}_{1h_1}(m_{ji})|X_j\right] \\
&= h_1^{-2}\delta_j E\left[\rho(m_i)\ddot{H}_{1h_1}(m_{ij})|X_j\right]-h_1^{-2}\delta_j\triangle P_j^{K-1,K}E\left[\rho_{g\delta}(m_i)\ddot{H}_{1h_1}(m_{ji})|X_j\right] \\
&= h_1^{-2}\delta_j\int\ddot{H}_1(u)\rho(m_j+hu)f_m(m_j+hu)\,du \\
&\quad -h_1^{-2}\delta_j\triangle P_j^{K-1,K}\int\ddot{H}_1(u)\rho_{g\delta}(m_j+hu)f_m(m_j+hu)\,du \\
&= \frac{1}{2}\delta_j\partial^2\left[\rho(m_j)f_m(m_j)\right]+\frac{1}{2}\delta_j\triangle P_j^{K-1,K}\partial^2\left[\rho_{g\delta}(m_j)f_m(m_j)\right]+r_{3n1,a}(X_j,\pi),
\end{aligned}$$

and

$$\begin{aligned}
r_{3n,2}(X_j,\pi) &= h_1^{-2}\delta_j\triangle g_j E\left[\delta_i\triangle P_i^{K-1,K}\ddot{H}_{1h_1}(m_{ji})|X_j\right]-h_1^{-2}\delta_j\triangle g_j\triangle P_j^{K-1,K}E\left[\delta_i\ddot{H}_{1h_1}(m_{ji})|X_j\right] \\
&= h_1^{-2}\delta_j\triangle g_j E\left[\rho_{\delta P}(m_i)\ddot{H}_{1h_1}(m_{ij})|X_j\right]-h_1^{-2}\delta_j\triangle g_j\triangle P_j^{K-1,K}E\left[\rho_{\delta}(m_i)\ddot{H}_{1h_1}(m_{ji})|X_j\right] \\
&= h_1^{-2}\delta_j\triangle g_j\int\ddot{H}_1(u)\rho_{\delta P}(m_j+hu)f_m(m_j+hu)\,du \\
&\quad -h_1^{-2}\delta_j\triangle g_j\triangle P_j^{K-1,K}\int\ddot{H}_1(u)\rho_{\delta}(m_j+hu)f_m(m_j+hu)\,du \\
&= \frac{1}{2}\delta_j\triangle g_j\partial^2\left[\rho_{\delta P}(m_j)f_m(m_j)\right]+\frac{1}{2}\delta_j\triangle g_j\triangle P_j^{K-1,K}\partial^2\left[\rho_{\delta}(m_j)f_m(m_j)\right]+r_{3n2,a}(X_j,\pi),
\end{aligned}$$

where $r_{3n1,a}(X_j,\pi)$ and $r_{3n2,a}(X_j,\pi)$ denote the remainder terms in the second order Taylor expansions, and we use the fact that $\int\ddot{H}_1(u)\,du=0$ and $\int\ddot{H}_1(u)\,udu=0$. With the above results, we can readily show that

$$\begin{aligned}
|\theta_{3n}(\pi)| &= |E\left[r_{3n,1}(X_j,\pi)+r_{3n,2}(X_j,\pi)\right]| \\
&\lesssim \frac{1}{2}\left|E\left\{\delta_j\triangle g_j\partial^2\left[\rho_{\delta P}(m_j)f_m(m_j)\right]\right\}\right|+\frac{1}{2}\left|E\left\{\delta_j\triangle g_j\triangle P_j^{K-1,K}\partial^2\left[\rho_{\delta}(m_j)f_m(m_j)\right]\right\}\right| \\
&\equiv \theta_{3n,1}(\pi)+\theta_{3n,2}(\pi).
\end{aligned}$$

uniformly in π with $\left\|\pi-\pi^0\right\|\leq C\eta_{1Kn}$. Then

$$\begin{aligned}
\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}|\theta_{3n,1}(\pi)| &= \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left|E\left\{\delta_j\triangle g_j\partial^2\left[\rho_{\delta P}(m_j)f_m(m_j)\right]\right\}\right| \\
&\lesssim \sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left[E\left(\delta_j^2\right)\right]^{1/2}\left\{E\left\|\partial^2\left[\rho_{\delta P}(m_j)f_m(m_j)\right]\right\|^2\right\}^{1/2} \\
&= O(\eta_{1Kn})\,O\left(K^{1/2}\eta_{1Kn}\right)=o(\eta_{1Kn}),
\end{aligned}$$

where we use the fact that $\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}E\left(\delta_j^2\right)=O\left(\eta_{1Kn}^2\right)$ and $\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}E\left\|\partial^2\left[\rho_{\delta P}(m_j)f_m(m_j)\right]\right\|^2$ $=O\left(K\eta_{1Kn}^2\right)$. Similarly, $\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}|\theta_{3n,2}(\pi)|=o(\eta_{1Kn})$.

Similarly, we can show that

$$E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\|r_{3n}(X_1,\pi)\|^2\right]\lesssim K^2\eta_{Kn}^2=o(K)$$

and

$$\begin{aligned}
& E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\|q_{3n}((X_1,X_2),\pi)\|^2\right] \\
\lesssim\; & E\left[\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left\|\triangle g_{1,2}\triangle P_{1,2}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}(m_{21})\left[L\left(R(X_1)'\pi\right)-m_1\right]\left[L\left(R(X_2)'\pi\right)-m_2\right]\right\|^2\right] \\
\lesssim\; & \max_{1\leq i\leq n}\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\|L\left(R(X_i)'\pi\right)-m_i\|^4 E\left[\left\|\triangle g_{1,2}\triangle P_{1,2}^{K-1,K}h_1^{-2}\ddot{H}_{1h_1}(m_{21})\right\|^2\right] \\
=\; & O\left(K^2\eta_{1Kn}^4\right)O\left(Kh^{-5}\right)=O\left(h_1^{-5}K^3\eta_{1Kn}^4\right).
\end{aligned}$$

Then by Corollary 5.3 in **?**, we have

$$\sup_{\|\pi-\pi^0\|\leq C\eta_{Kn}}\left\|\mathbb{U}_{3n}^{(2)}(\pi)\right\|\lesssim n^{-1}\left({h_1}^{-5}K^3\eta_{1Kn}^4\right)^{1/2}\ln(n)=o_p(\eta_{1Kn}).$$

By the empirical process theory, we have

$$\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left\|\mathbb{U}_{3n}^{(1)}(\pi)\right\|=O_p\left(n^{-1/2}K^{1/2}\ln(n)\right)O\left(K^{1/2}\eta_{1Kn}\right)=o_p(\eta_{1Kn}).$$

Consequently, we have

$$\|U_{3n}(\bar{\pi})\|\lesssim\sup_{\|\pi-\pi^0\|\leq C\eta_{1Kn}}\left\|\theta_{3n}(\pi)+2\mathbb{U}_{3n}^{(1)}(\pi)+\mathbb{U}_{3n}^{(2)}(\pi)\right\|=o_p(\eta_{1Kn})$$

and $\|L_{1,n4}\|=o_p(\eta_{1Kn})$. Analogously, we can show that $\|L_{1,n\ell}\|=o_p(\eta_{1Kn})$ for $\ell=5,6,7$.

(iv) The result follows from (i)-(iii). ■

Lemma B.2.3 *Let $L_{2,n}$, $L_{2,n1},...,L_{2,n7}$ be as defined in the proof of Theorem 3.3.1. Suppose that the conditions in Theorem 3.3.1 are satisfied. Then*

(i) $\|L_{2,n1}\|=O_p\left(K^{-\gamma+1/2}\right)$;

(ii) $\|L_{2,n\ell}\|=o_p\left(K^{-\gamma}\right)$ *for* $\ell=2,...,7$;

(iii) $\|L_{2,n}\|=O_p\left(K^{-\gamma+1/2}\right)$.

Proof. (i) Note that

$$\|L_{2,n1}\|=\left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left(\triangle P_{i,j}^{K\prime}\beta_0-\triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K}H_{1h_1,ji}\right\|$$

$$\leq 2\max_i \left|\triangle P_i^{K\prime}\beta_0 - \triangle g_i\right| \frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left\|\triangle P_{i,j}^{K-1,K}\right\| \left|H_{1h_1,ji}\right|$$
$$= O_p\left(K^{-\gamma}\right) O_p\left(K^{1/2}\right) = O_p\left(K^{-\gamma+1/2}\right)$$

(ii) Note that

$$\begin{aligned}
\left\|L_{2,n2}\right\| &= \left\|\frac{1}{n^2}\sum_{1\leq i\neq j\leq n}\left(\triangle P_{i,j}^{K\prime}\beta_0 - \triangle g_{i,j}\right)\triangle P_{i,j}^{K-1,K} h_1^{-1}\dot{H}_{1h_1,ji}\left(\bar{m}_j - m_j\right)\right\| \\
&\leq 2\max_i\left|\triangle P_i^{K\prime}\beta_0 - \triangle g_i\right|\max_j\left|\bar{m}_j - m_j\right| \frac{h_1^{-1}}{n^2}\sum_{1\leq i\neq j\leq n}\left\|\triangle P_{i,j}^{K-1,K}\right\|\left|\dot{H}_{1h_1,ji}\right| \\
&= O_p\left(K^{-\gamma}\right) O_p\left(K^{1/2}\eta_{1Kn}\right) O_p\left(h_1^{-1}\right) = o_p\left(K^{-\gamma}\right).
\end{aligned}$$

Similarly, we can show that $\|L_{2,n\ell}\| = o_p\left(K^{-\gamma}\right)$ for $\ell = 3, ..., 7$.

(iii) This follows from (i)-(ii). ■

Lemma B.2.4 *Let $G_{1,nj}, ..., G_{4,nj}$ be as defined in (B.1.19) in the proof of Theorem 3.3.2. Suppose that the conditions in Theorem 3.3.2 are satisfied. Then*

(i) There exists a positive constant c_F such that $\|G_{1,nj}\| \leq c_F\left|\triangle g_j - \triangle\bar{g}_j\right| + O_p(h_2^{a_2+1})$ uniformly in $j = 1, 2, ..., n$;

(ii) There exists a positive constant c_F such that $\|G_{2,nj}\| \leq c_F\left|\triangle g_j - \triangle\bar{g}_j\right| + O_p(h_2^{a_2+1})$ uniformly in $j = 1, 2, ..., n$;

(iii) $\max_{1\leq j\leq n}\|G_{3,nj}\| \leq O_p\left(\eta_{2Kn}\right)$;

(iv) $\max_{1\leq j\leq n}\|G_{4,nj}\| \leq O_p\left(\sqrt{\ln(n)/(nh_2)}\right)$.

Proof. (i) Recall that $\varsigma_{1i,j} = \left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)'$ and

$$\chi_{i,j} = \sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\left[\partial^l LF(\triangle g_j) - \partial^l LF(\triangle\bar{g}_j)\right]/l! - \sum_{l=a_2+1}^{\infty}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle\bar{g}_j)/l!.$$

Then

$$\begin{aligned}
&\|G_{1,nj}\| \\
&\leq \left|\frac{1}{n}\sum_{i=1}^N H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left(1, \frac{1}{h_2}\triangle\bar{g}_{i,j}, \cdots, \frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)'\right. \\
&\left.\times L'\left(LF(\triangle\bar{g}_i)\right)\left\{\sum_{l=0}^{a_2}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\left[\partial^l LF(\triangle g_j) - \partial^l LF(\triangle\bar{g}_j)\right]/l! - \sum_{l=a_2+1}^{\infty}\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\partial^l LF(\triangle\bar{g}_j)/l!\right\}\right| \\
&\leq \sum_{l=0}^{a_2}\left|\partial^l LF(\triangle g_j) - \partial^l LF(\triangle\bar{g}_j)\right|/l!
\end{aligned}$$

$$\times\left\|\frac{1}{n}\sum_{i=1}^{N}H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left(1,\frac{1}{h_2}\triangle\bar{g}_{i,j},\cdots,\frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)'L'\left(LF(\triangle\bar{g}_i)\right)\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l/l!\right\|$$

$$+\left\|\sum_{l=a_2+1}^{\infty}\left|\partial^l LF(\triangle\bar{g}_j)\right|/l!\left|\frac{1}{n}\sum_{i=1}^{N}H_{h_2}\left(\triangle\bar{g}_{i,j}\right)\left(1,\frac{1}{h_2}\triangle\bar{g}_{i,j},\cdots,\frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)'L'\left(LF(\triangle\bar{g}_i)\right)\frac{1}{h_2^l l!}\triangle\bar{g}_{i,j}^l\right|\right\|$$

$$\equiv G_{1,nj,1}+G_{1,nj,2}.$$

By the uniform boundedness of all finite order derivatives of $L(\cdot)$,

$$\begin{aligned}\left\|G_{1,nj,1}\right\| &\leq C\left|\triangle g_j-\triangle\bar{g}_j\right|\sum_{l=0}^{a_2}\left\|\frac{1}{n}\sum_{i=1}^{N}H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left(1,\frac{1}{h_2}\triangle\bar{g}_{i,j},\cdots,\frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)\left(\frac{\triangle\bar{g}_{i,j}}{h_2}\right)^l\right\| \\ &\leq C\left|\triangle g_j-\triangle\bar{g}_j\right|\sum_{l=0}^{a_2}\frac{1}{n}\sum_{i=1}^{N}\left\|H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\right\|\leq C\left|\triangle g_j-\triangle\bar{g}_j\right|\end{aligned}$$

where recall that C can vary over places. For $G_{1,nj,2}$, we have

$$\begin{aligned}\left\|G_{1,nj,2}\right\| &\lesssim \left\|\sum_{l=a_2+1}^{\infty}\left|\frac{1}{n}\sum_{i=1}^{N}\left|H_{h_2}\left(\triangle\bar{g}_{i,j}\right)\right|\right\|\left(1,\frac{1}{h_2}\triangle\bar{g}_{i,j},\cdots,\frac{1}{h_2^{a_2}a_2!}\triangle\bar{g}_{i,j}^{a_2}\right)\right\|\left(\frac{\triangle\bar{g}_{i,j}}{h_2}\right)^l\right\| \\ &= O_p(h_2^{a_2+1})\end{aligned}$$

where we use the fact that $\max_i\left|\triangle\bar{g}_i\right|=O_p\left(h_2^2\right)$. It follows that $G_{1,nj}\leq c_F\left|\triangle g_j-\triangle\bar{g}_j\right|+O_p(h_2^{a_2+1})$ uniformly in $j\in[n]$.

(ii) Note that $G_{2,nj}=\left(G_{2,nj,0},....,G_{2,nj,a_{h_2}}\right)$, where

$$G_{2,nj,l}=\frac{1}{n}\sum_{i=1}^{N}H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left(\frac{1}{h_2}\triangle\bar{g}_{i,j}\right)^l\left[F\left(\triangle\bar{g}_i\right)-F\left(\triangle g_i\right)\right]\text{ for }l=0,1,...,a_{h_2}.$$

Note that

$$\begin{aligned}G_{2,nj,0} &= \frac{1}{n}\sum_{i=1}^{N}H_{2h_2}\left(\triangle\bar{g}_{i,j}\right)\left\{F\left(\triangle\bar{g}_i\right)-F\left(\triangle g_i\right)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{N}H_{2h_2}\left(\triangle g_{i,j}\right)\left\{F\left(\triangle\bar{g}_i\right)-F\left(\triangle g_i\right)\right\}+\frac{1}{n}\sum_{i=1}^{N}\dot{H}_{2h_2}\left(\triangle\bar{g}_{i,j}^*\right)\left(\triangle\bar{g}_{i,j}-\triangle g_{i,j}\right)\left\{F\left(\triangle\bar{g}_i\right)-F\left(\triangle g_i\right)\right\} \\ &\equiv G_{2,nj,01}+G_{2,nj,02},\end{aligned}$$

where $\triangle\bar{g}_{i,j}^*$ lies between $\triangle\bar{g}_{i,j}$ and $\triangle g_{i,j}$. For $G_{2,nj,02}$, we have by Theorem 3.3.1(iv),

$$\begin{aligned}\left|G_{2,nj,02}\right| &\lesssim h_2^{-1}\max_{i,j}\left|\triangle\bar{g}_{i,j}-\triangle g_{i,j}\right|\max_i\left|\triangle\bar{g}_i-\triangle g_i\right| \\ &= O_p\left(h_2^{-1}\left(K^{1/2}\eta_{2Kn}\right)^2\right)=O_p\left(\eta_{2Kn}\right)\text{ uniformly in }j.\end{aligned}$$

For $G_{2,nj,01}$, we can show that

$$\left|G_{2,nj,01}\right|\lesssim\left|\frac{1}{n}\sum_{i=1}^{N}H_{2h_2}\left(\triangle g_{i,j}\right)\left\{F\left(\triangle\bar{g}_i\right)-F\left(\triangle g_i\right)\right\}\right|$$

$$\lesssim \left| \frac{1}{n} \sum_{i=1}^{N} \left| H_{2h_2} \left(\triangle g_{i,j} \right) F' \left(\triangle g_i \right) \right| \left(\triangle \bar{g}_i - \triangle g_i \right) \right| = O_p \left(\eta_{2Kn} \right).$$

Then $|G_{2,nj,0}| = O_p(\eta_{2Kn})$. Similarly, we can show that $|G_{2,nj,0\ell}| = O_p(\eta_{2Kn})$ for $\ell = 1, ..., a_2$. It follows that $|G_{2,nj}| = O_p(\eta_{2Kn})$.

(iii) The proof is analogous to that of (i) and thus omitted.

(iv) Recall that $\triangle \bar{g}_i = \bar{g}(X_{i2}) - \bar{g}(X_{i1})$ and $\triangle \bar{g}_{i,j} = \triangle \bar{g}_i - \triangle \bar{g}_j$. Let $\triangle g_i = g(X_{i2}) - g(X_{i1})$ and $\triangle g_{i,j} = \triangle g_i - \triangle g_j$. Note that $G_{4,nj} = (G_{4,nj,0}, ..., G_{4,na_S})$, where

$$G^0_{4,nj,l} = \frac{1}{n} \sum_{i=1}^{N} H_{2h_2} \left(\triangle \bar{g}_{i,j} \right) \left(\frac{1}{h_2} \triangle \bar{g}_{i,j} \right)^l \left[F(\triangle g_i) - D_i \right].$$

Let $G^0_{4,nj,l} = \frac{1}{n} \sum_{i=1, i \neq j}^{N} H_{h_2} \left(\triangle g_{i,j} \right) \left(\frac{1}{h_2} \triangle g_{i,j} \right)^l [F(\triangle g_i) - D_i]$ for $l = 0, 1, ..., a_2$ where 0^0 is defined to be 1. Noting that $E(D_i | X_i) = F(\triangle g_i)$, we can apply Bernstein exponential inequality to show that

$$\max_{1 \leq j \leq n} \left\| G^0_{4,nj,l} \right\| = O_p \left(\sqrt{\log(n)/(nh_2)} \right) \text{ for } l = 0, 1, ..., a_2.$$

Next,

$$G_{4,nj,0} - G^0_{4,nj,0} = \frac{1}{n} \sum_{i=1, i \neq j}^{N} \left[H_{2h_2} \left(\triangle \bar{g}_{i,j} \right) - H_{2h_2} \left(\triangle g_{i,j} \right) \right] \left[F(\triangle g_i) - D_i \right] - \frac{1}{n} H_{2h_2}(0) \left[F(\triangle g_j) - D_j \right].$$

It is easy to see that the second term on the right hand side of the last equation is $O_p \left(n^{-1} h_2^{-1} \right)$ uniformly in j. For the first term we can readily apply the arguments as used in the proof of Lemma B.2.2 and show it is $o_p \left(\sqrt{\log(n)/(nh_2)} \right)$ uniformly in j. Similarly, for $l = 1, ..., a_2$, we have

$$G_{4,nj,l} - G^0_{4,nj,l} = \frac{1}{n} \sum_{i=1, i \neq j}^{N} \left[H_{2h_2} \left(\triangle \bar{g}_{i,j} \right) \left(\frac{1}{h_2} \triangle \bar{g}_{i,j} \right)^l - H_{2h_2} \left(\triangle g_{i,j} \right) \left(\frac{1}{h_2} \triangle g_{i,j} \right)^l \right] \left[F(\triangle g_i) - D_i \right],$$

and we can use the arguments as used in the proof of Lemma B.2.2 and show it is $o_p \left(\sqrt{\log(n)/(nh_2)} \right)$ uniformly in j. ■